A.J.T.H. van Wezel

Weathering the storm towards sustainable mobility

TU Delft ASML

Weathering the storm towards sustainable mobility

A case study on how policy interventions can encourage sustainable travel choices in everyday commuting

By

A.J.T.H. van Wezel
Student number: 4717716

in partial fulfilment of the requirements for the degree of

Master of Science
in Engineering & Policy Analysis

at the Delft University of Technology,
to be defended publicly on March 21, 2025

Thesis Committee:
Dr.ir. M. Kroesen (First supervisor & Chair, TU Delft)
Dr. O. Oviedo-Trespalacios (Second supervisor, TU Delft)
P. de Weerd (External supervisor, ASML)

Cover photo by Michel de Heer – Power Pictures (ASML | Picturepark)

An electronic version of this thesis is afterwards available at http://repository.tudelft.nl/.

TU Delft

Preface

This thesis is the end result of my research project to complete the program MSc Engineering and Policy Analysis at Delft University of Technology. I am thankful that I could finish my time at Delft University of Technology with a thesis regarding sustainable mobility. I once chose to study at the faculty of Technology, Policy and Management because of the broad program and I was not yet sure where my interest lay. Now looking back in retrospect with the chosen domain of Transport & Logistics during my Bachelor of Science, a minor called 'Airport of the future' with as purpose the development of sustainable airports, and a bachelor's thesis with the aim to identify the ideal city layout for promoting bicycle usage I think I can say that sustainable mobility has attracted me all these years.

That is why I am grateful that I had the opportunity to work on this subject within the Access & Mobility team at ASML, where in addition to applying theoretical knowledge, I also had the unique opportunity to work with real data and think along with how to achieve sustainable commuting behaviour in a congested region. I would like to mention Pim de Weerd as he was part of the thesis committee and the program manager making him the bridge between the academic side and the case at the company. The many conversations and insightful meetings have led to a better understanding of the struggle towards sustainable commuting.

I would like to thank Oscar Oviedo-Trespalacios for thinking along and the critical but valuable feedback during the plenary sessions, which has been much appreciated and contributed to (hopefully) better highlighting the relevance of sustainable mobility in addition to performing 'just' a data analysis. As chair and first supervisor, Maarten Kroesen cannot be missing among the people I would like to thank. The bi-weekly sparring sessions have helped me tremendously towards this thesis as a final result and the further guidance during the thesis was also helpful resulting in a pleasant cooperation.

In addition to the people, I collaborated with during the thesis, I would like to end with mentioning my parents and girlfriend. Not only for the support the past six months, but also for supporting me throughout my time at Delft University of Technology. In terms of my study, but certainly also beyond that.

Antoine van Wezel
Delft, March 2025

Summary

Shifting toward sustainable mobility is a key element to reduce global emissions and mitigate climate change, while the transport sector is responsible for more than a third of all CO_2 emissions from end-use sectors. The car has a much larger carbon footprint than alternatives as the train, cycling, or the bus. In the Netherlands more than half of the total kilometres driven in a year are for business traffic and commuting. This causes significant CO_2 emissions. Employers therefore have a major role in a shift towards sustainable mobility by stimulating employees to *travel less*, *travel more efficient*, *or travel different*. Sustainable mobility is defined as transport that fulfils its economic and social role while containing the harmful effects of transport on the environment. Not only a shift towards sustainable contributes to reducing emissions but also decreases congestion on roads and the well-being of people. A universal solution to achieve sustainable mobility is lacking: there is no "one-size-fits-all" solution, and there is a lack of research measuring the effects after policies are implemented.

In the Netherlands, ASML implemented various interventions to stimulate more sustainable commuting of employees. However, ASML realizes that the current incentives, services, and facilities are not yet optimal. There are concerns regarding mobility as ASML is a fast-growing company in the busy top technology Brainport region. The goal of ASML is to keep the campus accessible and let the employees experience a safe, seamless, and sustainable commute. This case ties in with this research that aims to contribute to a shift towards sustainable mobility. As stated, employees can stimulate employees to travel less, more efficient, or different. Various factors influence the travel behaviour of commuters. With a focus on the desire of employees travelling less and different, Figure A depicts a framework with the determinants of travel behaviour and how these have sustainability impacts.

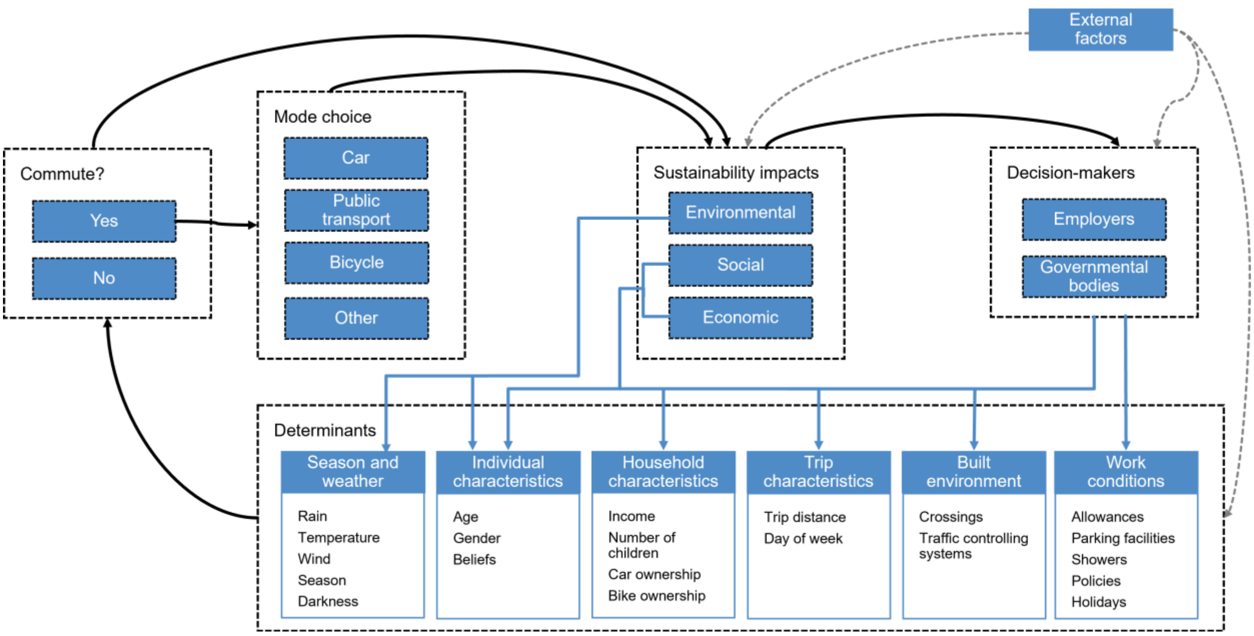


Figure A: Framework synthesized from the literature review

There is a lot of day-to-day variability in travel behaviour, which should be considered in researching long term changes in commuting behaviour. On a daily level, beside work conditions, weather conditions are the most important additional factor of influence on travel behaviour of commuters. But, for specific journey purposes as commuting, day-to-day changing factors remain understudied in relation to travel behaviour. Many studies state that seasonality is also an important influence on travel behaviour, however incorporating seasonality is difficult in practice due to a limited period of observations or a too large interval between observations. In the case of ASML, data on commuting behaviour of employees is collected for every day in a long continuous period of two years. This enables investigating the relationships between seasonality and travel behaviour. Alongside with the lack of research after policies are implemented and the need to research travel behaviour on a daily level for day-to-day variations, the following research question is the result:

"To what extent are weather conditions, seasonality, and implemented policies related to the day-to-day variations in commuting behaviour of ASML employees?"

This question is answered with a quantitative case study approach that follows the steps of a Data Science Process. This includes data collection, data preparation, Exploratory Data Analysis (EDA), modelling, visualizing, and reporting. EDA identified the patterns in the data and comparing the data at hand with findings in the literature regarding variables. The combined insights of EDA and the literature review contribute to an understanding on which choices are based in the modelling process. Ordinary Least Squares (OLS) regression is applied to investigate the associations of weather conditions, seasonality, and policies with daily travel volumes for car, bike, bus, and employees in office. These models are connected to the *reduction approach*. For the *alteration approach*, a Multinomial Logistic Regression (MLR) model is applied. This model limits the total probability to 1, regarding the probability of choosing a given alternative. By using car commuter share, bicycle commuter share, bus commuter share, and 'other' commuter share as alternatives, the probability of 1 can be seen as a percentage of 100% for which this model can estimate the modal split. For a dynamic component, three methods are considered that together are called Time Series Analysis. The first method is decomposition, which splits an observed pattern into a trend, seasonal pattern, and a random part. A Seasonal Auto-Regressive Integrated Moving Average with Exogenous variables (SARIMAX) model is tried to see whether this captures time patterns better than static regression models. Thirdly, with Interrupted Time Series Analysis, instead of the long-term influences of policies the immediate effects of policies are assessed right after implementation.

Figure B shows how the commuter volumes and shares for each mode change over time. This is based on the provided data. With the mentioned models, these trends are captured with explanatory variables with the objective to investigate the extent of the influences of the explanatory variables.

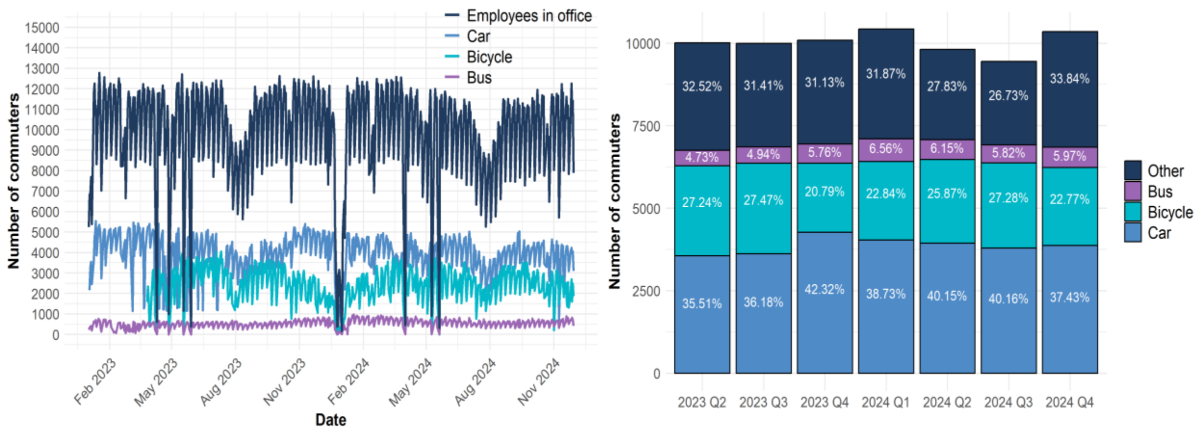


Figure B: Line graph and stacked bar plot of how commuters travel to campus

For the OLS regression this resulted in policies and interactions between weather conditions and seasonality significantly improved the model fit comparing to a model without interactions and policies. This depicts that weather conditions have different influences on travel behaviour in different seasons, and those policies (or at least moments in time) are associated with changes in travel volume of commuters. It can be seen that precipitation significantly reduces the number of bicycle commuters and also the employees in office while there is no significant effect for car and bus. If the precipitation is snow, there is a significant negative effect for car next to for bike and employees in office. From this it seems that while rain also is regarded as contributing to less safety on the roads, in this case cars are invariant to precipitation unless snow is involved. Darkness is found to play a role on commuting behaviour, with especially a negative effect on cyclist and increasing the number of bus commuters. A higher wind speed increases the commuters by car slightly while no significant effect for commuting by bike was found which is contradictory to expectations. This is probably due to the missing data in bicycle commuting data due to counting cameras not always working. Multiple Imputation by Chained Equations was utilized, but this introduces uncertainty and bias which can be an explanation for the absence of significant association. Darkness in spring has a significantly differing effect for cycling and car commuting, which depicts that after the switch to Daylight Saving Time and there is darkness again in the morning rush hours, people switch from cycling to car more than in other seasons. School holidays and Fridays decrease the number of commuters for all modes of transport and therefore also the employees in office. Using the models to predict actual values, it could be seen that all models capture the trends quite well depicting that seasonal related variables together with weather conditions can capture trends over time regarding commuting behaviour.

The MLR model shows that while for volumes Fridays are decreasing the commuters, in the setting of mode shares the commuters by sustainable modes decrease more relatively than car commuters. This shows that when there is more parking capacity available, people tend to switch back to unsustainable practices. This is evident also during school holidays. While the wind was not very present for the number of commuters, it does significantly change the modal split. A half hour of darkness already causes a lower odds ratio than snow for cycling to work. This depicts that darkness is an important negative predictor for bicycle usage. As from the travel volume model already evident, bicycle commuting is most severely impacted by adverse weather conditions.

Time Series Analysis showed that decomposition is able to make trends clearly visible and a seasonal weekly pattern can be captured for the modes. Although the regression models capture the trends quite well, the specification does not consider that the observations are linked to each other. With Heteroskedasticity- and Autocorrelation-consistent (HAC) standard errors, there is accounted for this issue. The other method to deal with the violation of the regression models is by using another model specification. A SARIMAX was fit for bus commuters, but the directions of coefficient turned out to be the same as for the static models. While the coefficients are hardly interpretable due to all seasonal, components, moving averages and autoregressive parts, SARIMAX is better for predicting but not for interpreting. While the goal is to support decision-makers and the model performance was the same it is chosen to use predominantly findings from the regression models. ITS showed impacts of two policies directly after implementation. The bus intervention seems more to have a long-term effect than a sudden effect when comparing with the decomposition of bus commuters. For the bicycle allowance, it looks like the trend is interrupted in 2024 which indicates a more sudden effect.

With the results the overarching research question can be answered:

"To what extent are weather conditions, seasonality, and implemented policies related to the day-to-day variations in commuting behaviour of ASML employees?"

Everyday weather and seasonality are associated with day-to-day variations in travel volumes and mode shares. Next to long-term trends with seasonality, everyday weather can be decisive for travel behaviour as darkness, wind speed, temperature, precipitation, all have impact. Weather conditions are also having different effects in different seasons for different modes. Holidays and Friday are associated with less commuters but also a shift towards unsustainable modes and therefore a modal split that is less desired. Policies are contributing to day-to-day variations in commuting behaviour while models improved by incorporating policies. However, isolating the effect of the policies of weather conditions, seasonality, other determinants of mode choice, and external factors is complex.

This research contributes to multiple knowledge gaps:

- While everyday weather is inadequately addressed, this research contributes to this with also incorporating wind speed (which is often overlooked).
- As no consensus in the literature how to incorporate seasonality, this research tries to incorporate seasons in multiple ways while also looking for interactions with weather conditions. Also darkness, which is rarely studied on travel behaviour, is studied and seems to be especially for commuting in the morning important.
- As assessing policy towards sustainable mobility is lacking, this research attempted to assess policies with extending the literature by confirming that there is no "one-size-fits-all" solution
- There is a lack of research focussed on commuting with cycling as an option. In this research cycling is used both in travel volume models and mode share models for commuters. Also, day-to-day changing factors for specific trip purposes is understudied, to which is contributed by looking at weather and seasonality for commuting.
- To shift towards sustainable mode choices, more research is needed in settings were cycling is common. The case of this research is in the Netherlands for commuters of a company with a considerable share of commuters by bike.

Several recommendations for policymakers are proposed:

- For assessing policies, dedicated attention is needed towards data quality if data-driven decision-making is the desire of policymakers. Careful verification and validation with acknowledging limitations can lead to better policy evaluation. With a plan how policy interventions are going to be assessed and also adding a qualitative component (e.g. surveys), decision-makers can get a more complete interview of the effects of interventions.
- While it is a shared responsibility to shift towards sustainable mobility broader collaboration is encouraged between employers, governments, transportation companies, and researchers regarding data which subsequently can provide additional insights in local contexts.
- Incentivize commuters to travel after sunrise (while darkness is found as one of the prevailing variables), which also causes less congestion and by incentivizing commuting by bike also lead to less emissions. This is an extension on working from home which is here to stay: flexible working by starting and ending the working day at home
- Lobbying for HOV lanes, making cycling in adverse weather conditions more attractive, and creating awareness on days with fewer people in office to take a sustainable mode of transport while there are more parking places available (which can potentially lead to spill-over effects on busier days) can all contribute to more sustainable transport, but policymakers need to be aware that there is no "one-size-fits-all" solution and a mix of policies is needed. Also next to these "pull" measures, also "push" measures should be considered.

Future research should also focus on the qualitative side of daily commuting, e.g. surveys regarding beliefs and attitudes towards weather conditions among commuters. While this research uses data on a daily level, a finer granularity of data during the day can be studied to see distributions during peak hours. Next to studying the economic and societal trade-offs and the first-mile of commuters back home, Doing research to everyday weather including darkness and wind in other contexts can fill gaps in knowledge and pave the way towards more sustainable mobility with less travel, more sustainable travel, and more efficient travel.

Table of Contents

Abbreviations and acronyms

Table 1: Abbreviations and acronyms

Abbreviation	Definition
A&M	Access & Mobility
AI	Artificial Intelligence
AIC	Akaike information criterion
ANOVA	Analysis of variance
API	Application Programming Interface
ASML	Advanced Semiconductor Materials Lithography
CET	Central European Time
CRAN	The Comprehensive R Archive Network
EV	Electric Vehicle
HAC	heteroskedasticity and autocorrelation
HOV	High-occupancy vehicle
i.i.d.	Independent and identically distributed
ITS	Interrupted Time Series
LRT	Likelihood-ratio test
MICE	Multiple Imputation by Chained Equations
MLR	Multinomial Logistic Regression
MLR	Multinomial Logistic Regression
Mton	Metric ton
NaN	Not a Number
OD	Origin-Destination
OLS	Ordinary least squares
OLS	Ordinary Least Squares
RMSE	Root Mean Square Error
SARIMAX	Seasonal Autoregressive Integrated Moving Average with Exogenous variables
SDG	Sustainable Development Goal
UTC	Universal Time Coordinated
WPM	work-related person mobility

1 Introduction

The transport sector, which includes passenger transport, freight transport, and business travel, is responsible for more than a third of all CO_2 emissions from end-use sectors (International Energy Agency [IEA], 2024). Therefore, the need to shift towards sustainable mobility is one of the priorities to reduce global emissions and mitigate climate change. The desired shift towards sustainable mobility revolves around transport that fulfils its economic and social role while containing the harmful effects on the environment (European Commission, 1992). This shift is essential for meeting commitments as outlined in the Paris Agreement and the 11th Sustainable Development Goal (SDG) of the United Nations regarding sustainable cities and communities (Paris Agreement, 2016; United Nations [UN], 2024).

In the Netherlands, business traffic and commuting by car together account for more than 50% of the total kilometres driven in a year, causing significant CO_2 emissions (Netherlands Enterprise Agency [RVO], 2024). The Netherlands agreed to reduce CO_2 emissions, and in the National Climate Agreement it is stated that employers have a major role in making mobility more sustainable (Government of the Netherlands, 2020). Encouraging employers to make conscious choices about mobility policy contributing to reducing emissions can help to stay below the collective ceiling of 1.5 Mton of CO_2 emissions before 2030. This can be achieved with three main pillars, being reducing travel demand, realising a modal shift, and increasing the efficiency of transport (Banister, 2008). A shift away from car travel to active modes as cycling can benefit commuters and contribute to healthier cities and a reduction in traffic congestion (Banerjee et al., 2021). Public transport can relieve the roads from too much traffic and the CO_2 emissions per travelled passenger kilometre are generally lower than for car travel (Santos et al., 2010). Although the electric vehicle is upcoming, this does not solve the traffic congestion and will take a long time before it rivals the traditional car.

1.1 Problem statement

Although many polices are proposed and implemented, there is still a struggle towards sustainable mobility due to a lack of consensus what sustainable mobility constitutes and how to achieve it (Berger et al., 2014). One of the key challenges is that the behaviour of commuters is highly influenced by various factors that change on a day-to-day basis, including weather, seasonality, and work conditions (Heinen et al., 2011). According to Chatterjee et al. (2016), a considerable minority of commuters turn out to have a day-to-day variability in commute mode choice and this should be considered in researching long term changes in commuting behaviour.

Gössling et al. (2023) found that the prevailing determinant on the depicted variability is weather, which has a higher influence in countries that have a greater cycling population. While the Netherlands is a country with a high cycling percentage, the modal split is varying on a daily basis with weather as one of the dominant factors that influences behaviour. This factor complicates implementing well-considered policies, while weather is an external factor that cannot be controlled or directly influenced by decision-makers. Although there is an existence of seasons and there are models in place that can predict the weather reasonably accurate, within seasons there is a lot of variation on a daily basis and forecasts become more inaccurate when forecasting for further into the future. This complicates efforts to achieve a consistent behavioural change towards more sustainable commuting, which is already complex due to its nature as a multi-actor problem (involving government bodies, companies, public transport operators, commuters, and unions) in which objectives do not always align. In addition, for policymakers there is an absence of a “one-size-fits-all” solution as described by Santos et al. (2010) towards sustainable mobility, which is an additional complication. A clear solution is lacking because the context dependency of weather conditions, socio-demographics, beliefs, attitudes, trip characteristics, built environment, and the policies that are in place.

1.2 Knowledge gaps

Most studies predominantly focus on extreme weather effects on travel behaviour, while weather is mentioned as having a great influence on day-to-day variability in travel demand and mode choice in countries where cycling is common. While temperature and precipitation are widely studied, wind is often overlooked in relation to travel behaviour (Böcker et al., 2013; Heinen et al., 2010). Although weather can now be predicted fairly accurately, it remains an uncontrollable factor for policymakers and can vary significantly between days. This shows that more attention is needed regarding everyday weather and the effects on travel behaviour. Additionally, in the literature it is agreed upon that seasonality plays a role but there is no consensus how to effectively integrate it. On the one side interactions between individual weather conditions and seasons, which means rain can have a different effect in summer compared with winter for example, are expected but have not been sufficiently studied (Liu et al., 2017). On the other side, the focus is often on the seasons as determinant itself, while according to Wessel (2022) research into light conditions (which are linked to seasons) on travel behaviour is rather scarce. Most research either uses short-term datasets or data that is collected with large intervals between it. Research is needed that therefore utilizes data on travel behaviour over a long timeframe to better incorporate seasonality (Heinen et al., 2010).

Commuting causes significant emissions and leads to traffic congestion. Employers have a major role in promoting sustainable commuting while they can influence working schedules, implement incentives for commuting and lobby for infrastructure and investment. Scientific research has provided insights in travel behaviour and possibilities towards sustainable mobility. However, as noted by Heinen et al. (2010), there is a lack of research that focuses on commuting with cycling as option. Research is in general either focused on the use of the bicycle or on commuting with not paying enough attention to the bicycle. This does not provide a complete picture of the changes in the modal split and travel volumes between the multiple options that people have for travelling to work.

Moreover, while policy interventions are proposed to encourage sustainable commuting, research on effectiveness after implementation is limited (Griffiths et al., 2021) and just a few interventions from a wide range are implemented (Gössling & Cohen, 2014). Without evaluation of policies, it is difficult to determine the actual impact on travel behaviour. The impact of policy interventions can also be highly context-dependent and there is an absence of a "one-size-fits-all" solution (Santos et al., 2010). More research is therefore needed into assessing policy interventions and understand under which conditions these interventions can be adapted to different urban and regional settings. According to Ton et al. (2019), literature concerning sustainable mode choices can be enhanced by doing more research in the Netherlands as a context, while most of the literature is in contexts where cycling is less common as means of transport.

From the above paragraphs it can be concluded that there is a lack of knowledge in the field of everyday weather and seasonality in relation to travel behaviour of commuters with also paying attention to the bicycle as part of the modal split for commuting. Additionally, the lack of quantitative evaluation of policies limits employers to learn lessons from other settings and develop effective strategies towards sustainable commuting of employees. Addressing these gaps is crucial for both the scientific understanding of sustainable commuting as well as for practical applications to achieve a shift towards sustainable mobility.

1.3 Case description

The exponential growing semiconductor industry, boosted by megatrends as Artificial Intelligence (AI), working from home and the rising demand for Electric Vehicles (EVs) will be a trillion-dollar industry in 2030 (Burkacky et al., 2022). As the largest technology company in Europe, ASML will as certain as it gets grow along being the world's leading supplier for the semiconductor industry. It is estimated that number of employers in the Netherlands will grow towards 30000 employees. ASML as a company embraces this growth, but it will go hand in glove with a significant increase in traffic in the Brainport region (where ASML is located together with other high-tech companies) and employees being responsible for a fair share of the CO_2 emissions.

ASML attaches great important to safe, seamless, and sustainable commuting. The Access & Mobility Program (A&M Program) is established to keep the campus and surroundings accessible and support employees to experience this safe, seamless, and more sustainable commute to and from work (ASML, 2024). The mission of the A&M Program is to reach a modal split of around 33% by car, 33% by bike and 33% by public transport by 2030. But, despite implementing various interventions, ASML realizes that the current incentives, services, and facilities are not yet optimal (ASML, 2024). The complex nature of mobility patterns is the reason that interventions are on many occasions unsuccessful or lead to inadvertent effects (Berger et al., 2014), in line with the earlier mentioned absence of a universal solution to accomplish a sustainable mobility system.

The company implemented policies to stimulate sustainable commuting but struggles with measuring the effects, and while there are goals for the long-term trend, the company is seeing significant day-to-day changes in travel behaviour. Therefore, ASML's problem fits in well with the identified problem. While the company collected data over a long period about the daily commutes of employees and has implemented policies to stimulate sustainable commuting, researching this case can contribute to addressing the stated knowledge gaps regarding how weather and seasonality are related to commuting behaviour for multiple modes of transport (including cycling) and evaluation of implemented policies towards sustainable mobility. The findings of this case can be incorporated into a discussion for broader application towards sustainable mobility and supporting other struggling employers to encourage employees to reduce the significant CO_2 emissions and congestion linked to commuting trips

1.4 Research questions

The research question guiding this study is:

To what extent are weather conditions, seasonality, and implemented policies related to the day-to-day variations in commuting behaviour of ASML employees?

This overarching research question can be broken down into several sub questions:

Sub question 1: *What are the characteristics of the data regarding weather conditions, seasonality, and travel behaviour of ASML employees?*
While a case is studied which is context-specific, it is important to have a good understanding of the conditions at hand to understand the things that are unique to the specific context as well as see whether there are similarities with other settings.

Sub question 2: *What are the (static) relationships of daily weather conditions and seasonality on the modal split and travel volume of ASML employees?*
By finding the relationships between daily weather conditions and seasonality on the modal split, knowledge can be added to the lack of knowledge regarding everyday weather and seasonality on travel behaviour of commuters with a focus on multiple modes (which is understudied).

Sub question 3: *How do polices and the interactions between daily weather and seasons relate with the modal split and travel volume of commuters?*
The previous sub questions set the scene, but this sub question goes a step further and contributes to the lack of knowledge regarding interaction between weather conditions and seasonality and assessing whether policies also contributed to a change in the day-to-day variability. The latter therefore tries to add to the scarce knowledge regarding assessing policy interventions.

Sub question 4: *How are time related patterns and the implementation of policies associated with commuting patterns over time?*
In addition to identifying relationships on daily commuting behaviour, the long-term goal is to achieve more sustainable mobility independent of day-to-day fluctuations. This sub question tries to capture temporal effects. In contrast to seasons that occur every year, the implementation of a policy is a one-time or infrequent event in time.

1.5 Research objectives

The research does not end with merely answering the research question. By answering the research question, insights are gained that can be used to support decision-makers in making informed choices towards more sustainable mobility in the future and to add to the existing literature on everyday commuting to work. Also prior to answering the research question, it is an objective to gather enough knowledge in the field to be able to conduct the research. This helps for understanding during the research, but also for having a good discussion after the results are available. The objectives can be summarized as follows:

1. Investigate what the existing research reveals about the relationship of weather conditions, seasonality, and policy interventions regarding travel behaviour and modal split targeted on commuters and assess how these relationships unfold in the case-specific research setting.
2. Contribute to filling the existing knowledge gaps in the literature depicted in this thesis regarding day-to-day commuting behaviour
3. Draw relevant findings and recommendations from the research for decision-makers to contribute to a shift towards more sustainable mobility.

1.6 Research approach

This thesis employs a (quantitative) case study approach, using data collected daily over a two-year period regarding commuting behaviour of employees. The case study is relevant while it enables the examination of day-to-day variability and effects of implemented policies over a long consecutive period which is often difficult to achieve in broader studies. While the case study represents a specific company and findings should be more indicative rather than definitive for other contexts, it can still offer valuable insights that contribute to the scientific knowledge on sustainable commuting and lessons can be learned how to move further.

As a framework, the Data Science Process by O'Neil and Schutt (2013) has been adapted and used to systematically guide the research. This process creates a feedback loop by incorporating the generated insights back into the real world which results in more data. The Data Science Process is shown in Figure 1. As can be seen there is also a feedback loop from Exploratory Data Analysis (EDA) to Raw Data Collection that is present due to the occurrence of realising the data is not clean enough during EDA which requires reconsideration of the collected data. The research methodology is underpinned by this framework and many steps directly align with the methodology. It also facilitates directions for the subsequent analysis and reporting. The framework contributes to ensuring that the research effectively addresses the research questions and achieves the research objectives.

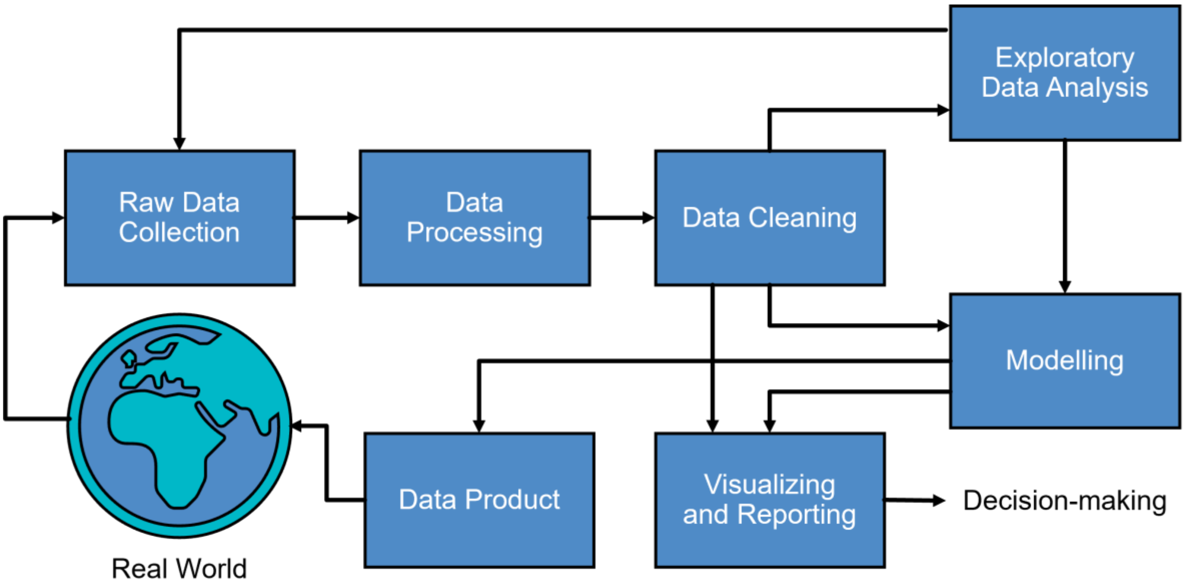


Figure 1: Data Science Process. Adapted from *Doing Data Science: Straight Talk from the Frontline* (p. 41), by C. O'Neil and R. Schutt, 2013, O'Reilly.

The framework of the Data Science Process systematically guides the research as a whole. The steps *Exploratory Data Analysis* and *Modelling* are the core and contribute most to answering the research question. This research question is broken down in the earlier mentioned four sub questions for this purpose. Sub question 1 is answered with the step *Exploratory Data Analysis* from the Data Science Process. The method that is used has the same name: Exploratory Data Analysis (EDA). The other sub questions are mainly answered within the *Modelling* step. For sub question 2, regression models identify the relationships between the determinants of mode choice and the modal split of the commuters while time series analysis can introduce time-effects to the model. Findings from the literature and insights from the EDA are used in de *Modelling* step for underlying assumptions and choices.

The research objectives are also linked to the Data Science Process. The first objective, to investigate what the existing research reveals about the relationships, is linked to understanding the *Real World*. The second objective can provide new insights on understanding the *Real World* and therefore also contributes to the gaps in knowledge in the current literature. The third objective, aimed at drawing relevant findings and recommendations from the research can be linked to *Visualizing and Reporting*. This can then support decision-makers with *Decision-making*, with an emphasis that results might not be generalizable to other cases in other settings but lessons can be learned which can help other companies and institutions understand their specific setting and getting a better understanding of the relevance.

1.7 Societal relevance

In The Netherlands, the perception about cycling is positive and infrastructure that is bike-friendly is in place, but many commuters still choose other means of transport (mostly car) while being in circumstances in which the choice for commuting by bike is very suitable (Heinen et al., 2010). As mentioned in the introduction of this chapter, it is estimated that in the Netherlands, business traffic and commuting together account for more than 50% of the kilometres driven, causing significant CO_2 emissions (Netherlands Enterprise Agency [RVO], 2024). While commuting for most people is required and certain time and location patterns are evident, commuting makes therefore disproportionate contributions to traffic congestion and emissions (Heinen et al., 2010). Time patterns as working hours result in more travel volume and rush hours in the beginning of the day, while locations with high employment levels can result in traffic jams even on traffic arteries. As employees have a major role in making mobility more sustainable, the Ministry of Infrastructure and Water Management commissioned the WPM, which is the reporting obligation for work-related mobility of persons for organizations with more than one hundred employees. The WPM was introduced on July 1, 2024, as pilot. Broader application can stimulate the need for companies to gather data regarding travel behaviour of their commuters to make it possible to do the obligatory reporting. With 23.000 employees in the Netherlands, ASML is one of the biggest companies of the country. Setting an example, can potentially contribute to more employers following. The desired shift towards sustainable mobility is critical for achieving national climate targets, improving air quality, and reducing congestions. Employers have a crucial role, especially in high density urban areas where problems are more evident. This research contributes to the broader societal goal of sustainability.

1.8 Scientific relevance

Although many policies are proposed and implemented by G20 countries, the local context plays a critical role in the successfulness of policies according to Griffiths et al. (2021). Yet, it is important to investigate how certain polices have worked out in other settings and look at similarities and difference. As Berger et al. (2014) concluded, the shift towards sustainable mobility is a highly fragmented and contentious process making sure that the pursuance of it remains of great interest for research regarding environmental policy and planning. In conventional transport planning, car-centric thinking often leads to a focus on making the system more efficient instead on alternating the modal split and travel behaviour (Banister, 2008 ; Berger et al., 2014).

As mentioned in section 1.2, research is lacking regarding day-to-day travel behaviour of commuters including also bicycle as one of the means of transport. By researching the described case, valuable insights into both travel volumes (relevant for the reduction approach) and travel shares (critical for the alteration approach) can contribute to a more complete picture of travel behaviour regarding the trip purpose of commuting. This is essential for developing effective and sustainable mobility policies. Quantifying the effects of policy and identifying in what circumstances the implemented policy are more effective, contributes to getting a better understanding of how a shift towards sustainable mobility can be achieved and demonstrate where more attention should be directed from researchers. By researching from a sustainable mobility perspective, looking beyond mobility alone is enabled. While for decades there is a struggle towards sustainable mobility, this research is relevant to add to the existing knowledge and understand the complex interplay of various determinants better.

1.9 Link to Engineering & Policy Analysis program

The research conducted in this thesis contributes to addressing the Grand Challenge of achieving sustainable mobility, which is part of the 11th Sustainable Development Goal (SDG) of the UN Global Goals regarding sustainable cities and communities (United Nations [UN], 2024). By using methodologies and techniques learned throughout the MSc Engineering & Policy Analysis program, the research aims to provide insights into relationships between weather conditions, seasonality and polices for decision-makers to support a shift towards sustainable mobility.

The two fundamental themes in the program are *1. policy & politics* and *2. analytics, modelling & simulation*. The first theme is incorporated in this thesis by exploring how policies contribute towards sustainable mobility and providing evidence-based recommendations from the case study. The second theme is covered by using the collected data for a quantitative data analysis to study travel behaviour of commuters. By the combination of these two themes, both the technological and societal dimensions of sustainable mobility are researched.

1.10 Thesis outline

In this chapter the problem is introduced, and the relevance is explained. Chapter 2 will entail on the basis of a literature review about past research regarding travel behaviour and sustainable mobility, and assess papers on policies implemented to promote sustainable mobility. The Data Science Process will be used to guide the methodology in Chapter 3. After this chapter, in the intermediate Chapter 4 it is entailed how the data in prepared for researching and answering the sub questions. The results of applying the methodology are reported in Chapter 5. This chapter is followed-up by the discussion: Chapter 6, in which the results are interpreted and limitations are highlighted. Conclusions and recommendations are reported in Chapter 7 with also a critical reflection on the research process of this thesis.

2 Literature review

The foundation of all academic research activities, regardless of discipline, is based on and related to existing knowledge (Snyder, 2019). Therefore, this chapter describes the findings of the literature review conducted for this thesis. Some findings have already been mentioned in the first chapter, but this chapter provides more context. To gather insights from the literature about sustainable mobility and the current body of research regarding travel behaviour and its determinants, the web search engine Google Scholar was used. Guided by a number of criteria, a core set of articles was compiled. Using the snowball method, subsequent articles were identified. In Appendix A, the process behind this literature review is discussed in more detail.

2.1 Sustainable mobility

Sustainable mobility, also often referred to as sustainable transport, is a relatively novel concept that has been introduced in the *1992 EC Green Paper on the Impact of Transport on the Environment* (European Commission, 1992). The paper describes sustainable mobility as transport that fulfils its economic and social role while containing the harmful effects of transport on the environment. Banister (2008) introduced the widely referenced sustainable mobility paradigm which advocates for a shift-away from car-centric (conventional) transport planning towards more sustainable practices. This entails reducing need to travel and trip lengths, encouraging a modal shift, and stimulating greater efficiency in the transport sector. Berger et al. (2014) positioned this in a people perspective: humans can travel different, travel less, and travel more efficient. This resulted in a reduction approach, alteration approach, and efficiency approach. In addition to the correspondences with Banister's paradigm, these three approaches are firmly rooted in the literature on sustainable mobility (Berger et al., 2014). People being the key in a shift towards sustainable mobility is also embraced by Holden et al. (2020), while people decide when, where, how, and with what they travel, while also electing the politicians that influence the mobility system. Santos et al. (2010) emphasized a decade earlier that besides people being the key in a shift, policymakers have a crucial role in devising policies that can trigger a change in behaviour. When coupling the reviewed literature about sustainable mobility with the problem statement, Figure 2 can be derived. The figure shows how decision-makers can stimulate commuters to change travel behaviour and on which dimensions of sustainability these changes have impact.

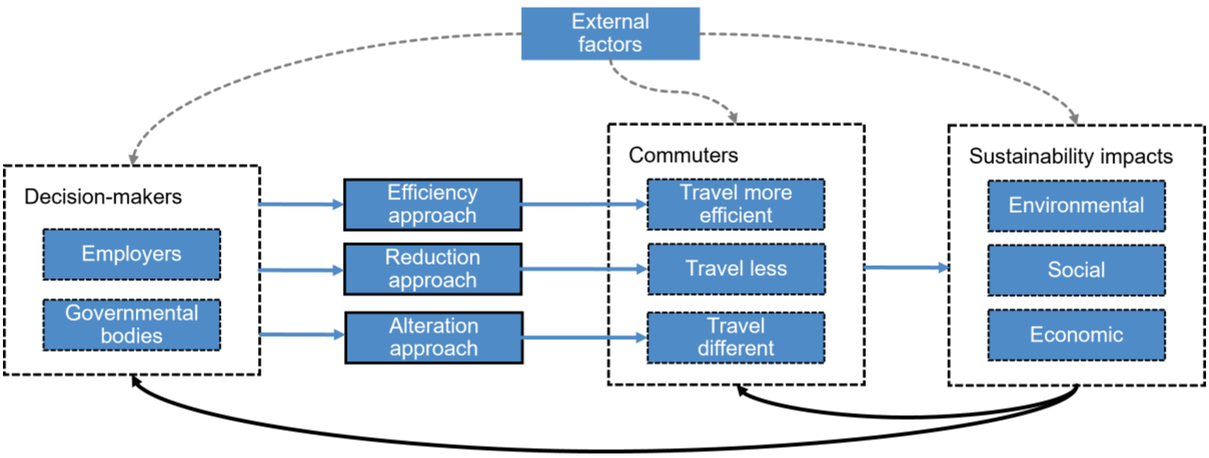


Figure 2: Sustainable commuting framework

Despite the extensive literature on the need for sustainable mobility, the current state regarding realising a shift towards sustainable mobility is below par. According to Holden et al. (2020), in the almost three decennia following the green paper there has been progress, but with the global set targets for the reduction of carbon emissions the devotion of the transport sector is insufficient. As Berger et al. (2014) indicated: there is a struggle towards sustainable mobility due to a lack of consensus on what constitutes sustainable mobility and how to achieve it. Gallo & Marinelli (2020) also indicate that there is much work to be done, yet they notice that regardless of political preference of decision-makers one of the most widespread goals in transport policy at present is the promotion of sustainable mobility.

2.2 Determinants of travel behaviour

From the previous section, it can be concluded that the right policy from policymakers can be a catalyst for behavioural change of humans, which is key for shifting towards sustainable mobility. From the alteration approach of Berger et al. (2014) people can travel different and from the reduction approach people can travel less. Therefore, it is of importance to have insights in the determinants of travel behaviour of people. Multiple scientific articles are reviewed, that each also reviewed literature regarding travel behaviour and mode choice.

2.2.1 The categories of determinants

There is a lot of varying results regarding determinants and its effects on mode choice (Ton et al., 2019). Following from Heinen et al. (2010) the occurrences of contradicting results in research calls more for an overview of determinants and not a thorough review, while it is difficult to assess which analyses are right. Moreover, looking at the context of past research is of essence to gain more insight into difference between studies. To illustrate this: Böcker et al. (2013) reviewed literature of the impact of weather across modes on a daily level, Heinen et al. (2010) had a focus on determinants for commuting to work on a bicycle, and Ton et al. (2019) compared own research in the Netherlands regarding active mode choice with other literature mainly focussed on cycling and walking. For each category of determinants identified by Ton et al. (2019), the key findings from the literature are presented.

2.2.1.1 Individual characteristics

This category of travel behaviour determinants includes socio-economics and socio-demographics as gender, age, country of origin, education level, and income of a person. Also, ownership of a means of transport and attitudinal determinants are in this category. While most research concludes that males use the bicycle more than females, there are also opposing studies (Heinen et al., 2010). It was found for example by Witlox and Tindemans (2004) that among the working population, females cycle more often than men. Ton et al., (2019) contradict the suggestion of Heinen et al. (2010) that females cycle more in countries with many cyclists by stating from their research in the Netherlands that besides age, also gender seems to be not an explanatory determinant. This ambiguity is present in the majority of individual characteristics. An exception to this is ownership of a means of transport. Car ownership is having a strongly negative effect on bicycle usage, while naturally follows that owning a bike raises the probability of using it as a means of transport (Heinen et al., 2010). Immigrants tend to cycle less in countries where cycling is quite common, with experience and cultural values regarded as having a central role (Haustein et al., 2020). It is indicated that the results should be treated with caution due to the design of the research.

Most of the individual characteristics are fairly fixed. It is given when someone’s birthday is, and gender is also fixed or in some cases will be adjusted once in a lifetime. Part of the individual characteristics category however are susceptible to change: attitudinal determinants. Attitudinal determinants (meaning, social values, and attitudes) are influential in active mode choice (Heinen et al., 2010 ; Ton et al., 2019). The problem with these determinants is that they are difficult to measure and more theory based than empirical based. Heinen et al. (2010) suggested in addition to avoid focusing too heavy on socio-economic factors, due to mixed evidence and the determinants missing clarity on the direction and causality.

2.2.1.2 Household characteristics

The category of household characteristics has some unique determinants (e.g. the number of children), but also determinants that are connected to individual characteristics (e.g. house income, and ownership of a means of transport). The structure of a household can apparently influence the probability of cycling as the literature review from Heinen et al. (2010) indicated that having no children or being divorced or widowed leads to a higher probability of cycling. In subsequent research from Heinen et al. (2011), it is mentioned that trip chaining (e.g. bringing children to school) could also be a reason to not use the bicycle for commuting which can explain the negative relationship between having children and cycling to work. Findings from Ton et al. (2019) are mostly in line with earlier research regarding household characteristics, although not finding a significant association between number of children and cycling, it was found for public transport usage instead. Not finding a significant association between income and probability of taking the bicycle from Ton et al. (2019) adds to the mentioned indecisiveness regarding this association which is evident from the review of Heinen et al. (2010). Similarly to the individual characteristics, the category of household characteristics seems also subject to ambiguity.

2.2.1.3 Season and weather

Up until the first decade of this century, relatively little attention was paid for consequences of climate change and weather conditions regarding the transport sector, but it was recognized that transport systems are affected by adverse and extreme weather conditions (Koetse & Rietveld, 2009). In research and planning, both weather conditions and climate change became key issues with growing scientific consensus surrounding climate change (Böcker et al., 2016). This has led to increased research focusing not only on extreme weather events but also on everyday weather (Böcker et al., 2013). Still, this does not mean there is enough focus on everyday weather as literature tends to focus on extreme weather (Gössling et al., 2023; Ton et al., 2019), while everyday weather which may have a substantial impact is inadequately addressed (Liu et al., 2017). Active mode users are the most sensitive to weather (Böcker et al., 2013; Liu et al., 2017), with public transport in between cycling and travelling by car according to Faber et al. (2022). Another interesting finding from Faber et al. (2022) is that people with an e-bike or a lot of cycle experience are less susceptible to adverse weather conditions.

There are many differing studies regarding weather impacts on travel behaviour, from weather impacts on different travel modes to how weather is represented (Liu et al., 2017). In the literature, weather conditions that are regarded most impactful are rain and temperature. Precipitation is by far the most important variable regarding safety on the roads according to Koetse and Rietveld (2009). Wind is often overlooked (Böcker et al., 2013; Heinen et al., 2010). Yet this factor is often mentioned having influence on travel behaviour (Faber et al., 2022; Gössling et al., 2023; Sabir, 2011). Liu et al. (2017) state that meteorological variables often are often correlated, which indicates interrelation between various variables. It is concluded by Böcker et al. (2013) that most studies deal with weather effects separately, failing to categorize the co-occurrence of relevant weather parameters.

Seasonal influences on behaviour are understudied. Most research pay attention on the impacts on the short-term, but less on the seasonal level and impacts seem to be specific to regions (Koetse & Rietveld, 2009). A concrete example has been given by Liu et al. (2017), that in many research increase in temperature is often generalized over seasons while perception of temperature can differ across seasons. Although seasonality is recognized as a crucial factor, there is no consensus on how to represent this in research (Liu et al., 2017; Sabir, 2011). This can also lead to contrary findings. In the study from Ton et al. (2019) seasonality was represented by months for September till November and weather as a subjective interpretation of respondents whether the weather conditions were extreme on a given day. This can be seen as an illustration of the findings from Liu et al. (2017) that everyday weather is often overlooked, and seasonality not adequately addressed in the literature. Yet, there are studies that report findings regarding seasonality. From the literature review of Heinen et al. (2010) comes forward that winter affects cycling in a two-fold way: there are decreases in the number of cyclists during winters and the average distance that people cycle decreases. Autumn and summer are the seasons depicted as most favourable for active mode choice, which includes cycling (Böcker et al., 2013; Heinen et al., 2010)

Next to the relationship between weather conditions and seasons, the seasons are also related to the hours of daylight with darkness having a negative influence on the number of commuters on bicycle (Heinen et al., 2010). Which might raise interest into measuring the difference between Daylight Saving Time and Standard Time on travel behaviour (Wessel, 2022). But, according to Wessel (2022) darkness itself although depicted as a variable that contributes to less bicycle commuters, is rarely mentioned or used in travel behaviour models. While there is overlap between the moment the sun rises and peak hours, especially for commuting investigating darkness can fill a gap in knowledge.

Another contrary finding from Ton et al. (2019), is that there is no significant effect of (extreme) weather on active transport use. According to the authors the reason for this might be that habitants get used to the frequent rain and mild climate in the Netherlands, where the research took place. The other possibility they mention for contradicting the existing literature is how weather is formulated in the study. Faber et al. (2022) found that commute trips are less affected by the weather than leisure trips, which is depicted by Liu et al. (2017) as being less elastic in response to weather changes. Next to this, there should be a clear explanation of how weather is represented. The extent of the effect of determinants is again context-dependent, however weather is very well measured and documented and can be traced to almost any location.

2.2.1.4 Trip characteristics

According to Banerjee et al. (2021), the distance of a trip is used in most of the existing literature regarding cycling as the sole determinant for the use of bicycle, neglecting the various other proven variables for mode choice. Travel time is less researched, but highly correlated with trip distance (Ton et al., 2019). Where the focus lies can also be mode dependent: public transport trips are expressed in time while the distance matters less but for car users distance can be of more importance due to fuel costs. Next to these two, there are other trip characteristics although being less often researched e.g. day of the week (Ton et al., 2019). The reviewed literature mainly distinguishes weekdays with weekends while on the entire transport demand trip purpose is predominantly leisure over commuting in weekends and commuting over leisure on weekdays. For a specific trip purpose, the day of the week might matter also.

2.2.1.5 Built environment

Built environment refers to man-made structures and facilities in the environment in which humans work and live. Heinen et al. (2010) distinguish two categories: the urban form and infrastructure. Travel time and travel distance are influenced by urban form determinants, as the network layout, population density, and land usage mix (Heinen et al., 2010 ; Ton et al., 2019). Logically, higher densities and a more fine-grained network leads to shorter distances which encourages the usage of bicycles, as supported by both literature reviews. The presence of dedicated bicycle infrastructure (separate lanes), traffic signage, and traffic control have an influence on the mode choice (Heinen et al., 2010). The preference of bicycle facilities can differ between experienced and unseasoned cyclists, with less experience associated to a stronger preference for bicycle dedicated facilities and infrastructure (Heinen et al., 2010). Contradicting other findings, Ton et al. (2019) concluded that infrastructure does not influence the choice for an active mode possibly caused by the fact that infrastructure regarding active modes is highly developed and advanced in the context that was researched, which was Rotterdam in the Netherlands.

2.2.1.6 Work conditions

This category is related to facilities that are present at the workplace, policy around travel allowances for commuting, and elements of the employment contract (Ton et al., 2019). It appears that cycling commuters value shower facilities and facilities to park the bike, but from the little research in this area it seems it does not have an effect on the modal split and the frequency of bicycle usage (Heinen et al., 2010). For this research, while one company is researched, for most employers the conditions are the same. It is therefore more interesting what happens when work conditions change, which within a company often is dependent on policies implemented. Statutory days off will logically lead to less travel volume caused by commuting.

2.2.2 Stated reasons to not use a sustainable mode of transport

The variety of determinants leading to a mode choice makes the predicting of this choice complex. People tend also to develop habits, leading to less rational mode choices (Heinen et al., 2010). The most prominent reasons according to the literature review of Heinen et al. (2010) regarding cycling, are depicted in the Table 2. Added in the table is also the link to one of the categories of determinants.

Table 2: Reasons to not use a bicycle as mode of transport

Reason	Category
Not safe enough	Individual characteristics / Built environment
Too congested	Trip characteristics
Weather conditions	Season and weather
Darkness, or not enough light	Season and weather
Not fit enough, too tired, effort	Individual characteristics
Difficult to chain trips	Built environment / Trip characteristics
Not being able to cycle	Individual characteristic / Trip characteristics

Not being able to cycle is considered to be due to not having a bicycle or the distance that needs to be cycled. It is unclear for which distance, and for which type of bike the reason to not cycle is the effort or how fit a person is. Darkness and safety could have some overlap, while darkness is for commuting in the Netherlands relevant because this in the winter often occurs during rush hours. Heinen & Handy (2012) state that infrastructure definitely influences the choice to cycle or not, although it could be disputed as the infrastructure is already superior in the Netherlands (Ton et al., 2019).

Although public transport is regarded a safer mode of transport in comparison with car, people stated that public transport is less convenient, less flexible, more uncomfortable, slower, less reliable, and less enjoyable (Linda, 2003). Crowding, unreliability, and waiting-times are the key drivers affecting enjoyability (Cantwell et al., 2009). Focussing on more bus usage, higher frequency, a direct bus, and a shorter travel time would increase users following from a survey with both public transport commuters and car commuters (Eriksson et al., 2008). From the same research, it comes forward that a frequent reason to not use the bus is simply that it is not possible to use the bus.

2.3 Day-to-day variability in travel behaviour

From the previous sections, it comes forward that a considerable share of the determinants is subject to day-to-day variability, leading to day-to-day fluctuations in travel behaviour. This is accompanied by stated reasons to not use a sustainable mode which can also vary by day. Understanding the variability is recognized as crucial in managing things as urban congestion (Raux et al., 2016).

Heinen et al. (2011) found evidence that the decision to cycle is for a great part influenced by factors that can be different on a daily basis, such as weather conditions, characteristics of the trip, and work characteristics. This is in line with research from Gössling et al. (2023) indicating that specifically for active transport the prevailing determinant for day-to-day variations in demand is weather. To some extent this is also evident from other research, stating that weather significantly influences mode choices (Böcker et al., 2013; Liu et al., 2017; Sabir, 2011). The different modes used by individuals because of day-to-day changing factors is understudied for specific journey purposes as commuting followed from Chatterjee et al. (2016). As mentioned by Heinen et al. (2010), the current body of literature does not provide a complete picture of the changes in the modal split and travel volumes between the multiple options that people have for travelling to work: either cycling in general is researched or commuting to work with too little attention towards the bicycle as an option. Less variable on a daily level are the built environment characteristics, while often infrastructure is built for a period stretching between a few decades towards a century (Sabir, 2011). Also, signage will not be changed overnight.

Next to choosing a mode there is also the choice to not commute at all. The choice to not commute to work can be indirectly influenced by policymakers that use the reduction approach of Berger et al. (2014), aimed at encouraging people to travel less (e.g. working from home). There are also other reasons for not commuting or not working that cause day-to-day variability in travel volume (holidays, leave) and reasons to stay home could also be due to illness or pregnancy (Heinen et al., 2011).

2.4 Policies towards sustainable mobility

There is raising interest among policymakers in promoting sustainable mobility due to the consensus in the literature of the benefits of cycling and the need for a shift towards sustainable mobility (Banerjee et al., 2021). Previously, building more roads was seen as the holy grail to combat congestion although it is now evident that it leads to higher levels of travel demand and more congestion. Banister (2008) explains this by mentioning that conventional (car-centric) analysis of transport is based on minimizing the generalised cost of travel, consisting of travel cost and travel time. More roads lead to less travel time, ultimately causing higher traffic density and more congestion around the built roads. The sustainable mobility paradigm of Banister (2008) is therefore critical in shifting away from a car-centric view that is based on optimizing derived demand and minimalizing travel time.

The difficulty that is evident from the literature, is the absence of a "one-size-fits-all" solution as described by Santos et al. (2010). Griffiths et al. (2021) phrase it as a necessity of establishing mixes of policies and Hrelja and Rye (2023) state multiple polices are key towards sustainable development. Holden et al. (2020) present three storylines, which they call *grand narratives* to compel governments, people, and employers in bringing about the necessary changes and steer policy in the right direction. They conclude that the knowledge, technology, and policies are present to shift towards a sustainable mobility system, but the problem is the lack of will-power in general. In the sustainability paradigm, this is called schizophrenic paths: There is consensus that there is a need for action, but no effective measures are taken to mitigate the problem at hand (Banister, 2008).

There seems consensus about the framework that is in place regarding a policy mix needed for achieving sustainable mobility, but it manifests itself in different ways: from a people perspective the *reduction, alteration, and efficiency approach* mentioned by Berger et al. (2014) is well established, while Griffiths et al. (2021) call for an *Avoid-Shift-Improve framework*. This framework consists of applying policy mixes that avoid unnecessary transportation volume, shifting the norms and practices regarding transportation, and improving the existing transportation systems. In both cases the goal and the means are the same and fit the sustainable mobility paradigm of Banister (2008). By coupling sustainable mobility policies to the three narratives, Griffiths et al. (2021) provided more practicality for decision-makers. The coupled policies are also divided into three categories by Griffiths et al. (2021): regulatory, economic, and information. Santos et al. (2010) have different categorization: physical, soft, and knowledge policies alongside economic policies.

To counteract the ingrained mindset in which the car is central, focus from governments and employers should be on exploiting the potential that may exist for individuals by combining their commute with exercising, allocating budgets for promoting cycling to work, and creating awareness of the psychological advantages of maintaining cycling behaviour (Handy et al., 2014). The same research showed that while many studies underline dissatisfaction and stress perceived with commuting, surveyed cycling commuters experienced their commute predominantly positive along with a better mood and stress reduction. The COVID-19 crisis made it possible for niche ideas towards sustainable mobility to become potentially widely accepted (Griffiths et al., 2021). For reducing travel demand, working from home is a major breakthrough catalysed by COVID (further discussed in the next section).

In the beginning of this section, it is stated that there is consensus that a "one-size-fits-all" solution is absent. Despite this consensus, Gössling and Cohen (2014) point out that a wide range of measures are discussed by policymakers but just a few are implemented. With more focus on achieving sustainable mobility, Griffiths et al. (2021) predict that assessing effectiveness of implemented policies designed for context-specific setting will be an important topic of much further research and with time it becomes possible to assess implemented policies which offers huge research opportunities.

2.5 ASML policies and interventions in the literature

ASML has implemented a variety of interventions to mitigate the disadvantageous consequences of unsustainable transport and is planning to make certain interventions in the short term. Some of these potential interventions are researched by Molin and Kroesen (2023) with a choice experiment among employees from ASML. This method contributes to getting an understanding of preferences of respondents, in this experiment the employees of ASML. This section reviews literature regarding past and potential interventions in other contexts, providing insights into their relevance and applicability in current research. In some cases, by drawing comparisons between findings from Molin and Kroesen (2023) and the literature insights can be gained by highlighting similarities, differences, and potential implications for putting policy interventions into practice. The policies can also be linked to the framework in Figure 2. Too which approach each policy belongs is therefore depicted in Table 3.

Table 3: Policies implemented by ASML

Policy	Reduction	Alteration	Efficiency
Bike plan		☑	
Bicycle allowance		☑	
Shared e-bikes		☑	☑
Pooling	☑		☑
Fare-free public transport	☑	☑	
Remote working policy	☑		
Frequency of public transport		☑	☑

Of course, there could be some discussion about the categorizing. For example: pooling can also lead to people change their mode choice from public transport or bike towards sharing a car. However, this is (if it happens) an unwanted side effect of the intention to reduce the number of cars travelling to work. Another thing to keep in mind is that the categorizing of policies is relative to the scope and the policy itself: an example that can be used is differentiating starting times (not included in this research as a reviewed policy). Looking at the peak hours, this policy can be seen as a reduction measure but on daily level it is more a measure for improving efficiency while the travel demand remains the same.

2.5.1 Bike Plan

In 2020, on a national level in the Netherlands the rules changed regarding the traditional bike plan with more attention for promoting e-bike possession and usage in the Netherlands by making it possible to lease a bike to reduce costs for the employee (de Haas et al., 2021). A tax scheme exists that enables employers to offer ownership of an (e-)bike to employees with reduced cost. The mentioned intervention is called the Dutch Bike Plan. Next to encouraging more cyclists, a switch from a normal bicycle to an electric bicycle can lead to 1.5 longer trips and makes it more likely that people outside ~8km radius will travel by bike (Nematchoua et al., 2020).

ASML has as policy to have a 2000 euro deduction to buy a (e-)bicycle with tax advantage. The leasing option that was nationally introduced in 2020 is also an option at ASML. The research from Molin and Kroesen (2023) showed that there is still much potential for commuters within twenty kilometres to switch to (e-)bicycle and that there is still a considerable interest in the bicycle plan.

2.5.2 Bicycle allowance

Kroesen and Handy (2014) identified a bidirectional relationship between bicycle commuting and non-work cycling, which indicates that work-related factors as receiving a travel allowance for cycling can result in spill-over effects to other domains and vice versa. This is a beneficial side effect in line with a shift towards sustainable mobility initiated by offering a travel allowance for cycling to work. Dutch research of MuConsult (2019) calculated that for companies with more than 1000 employers, an increase of 10% potentially leads to 0.3% more cycle days. For similar sectors (all sectors excluding governments, education, and health care) as ASML the effect is estimated to be 1.7% cycle days more. Much more effective than an increase in allowance is having any compensation at all company-wide: estimates of 15% for companies with more than 1000 employers and 11% for companies in the same sector. It is not clear from the Dutch research what the effects are from a more substantial increase and context-dependency should not be overlooked.

2.5.3 Shared e-bikes

In comparison with the conventional bike, it turns out that the usage of an e-bike leads to more trips and an increase in the average distance that an individual cycles (Banerjee et al., 2021). From longitudinal travel data, it turned out that in general only conventional cycling trips are reduced by e-bikes which is undesired from a policy perspective. However, when the trip purpose is commuting it turns out that the e-bike also substitutes car trips (de Haas et al., 2022). This raises the question whether this has also been the case for ASML, as the local contexts have overlap due to both being related to the Netherlands and with the trip purpose of commuting.

From a case study in the city of Delft, it was observed that due to bike sharing users, the usage of a lot of other options decreased with train as exception, which had an increase in usage (Ma et al., 2020). Mostly caused by the possibility to park and pickup shared bikes at train stations, facilitating an easy and smooth multimodal trip according to Ma et al. (2020). An advantage for the shared e-bike system of ASML, is that the parking facility is directly outside the train station. Other shared bicycle providers often have restrictions to prevent competition with the shared bicycles of the NS, which is the principal passenger railway operator in the Netherlands. In Delft, Mobike was obligated to have the shared bicycles parked at least 150 metres of train stations (Ma et al., 2020). Compared to the NS-bikes, another advantage is that you do not have to return the bicycle to the same place as where you have picked the bicycle up: employees can also choose to use a bicycle only on the outward journey or only on the return journey assuming a bike is available. Compared to all bicycle sharing systems from the case study, usage of the bike sharing system of ASML is free for commuters. The research from Molin and Kroesen (2023) concluded that time and costs are most influential in mode choice for ASML employees, making the free use of the bike sharing system probably even more attractive.

As earlier mentioned, Faber et al. (2022) concluded that people having an e-bike are better weather resistant. It could therefore be interesting to investigate whether introducing shared e-bikes also leads to more people commuting by bicycle in adverse weather conditions. In the case of ASML, e-bikes where present during the entire observation period which makes it difficult to investigate the introduction. However, there has been a switch in operator with two weeks without shared e-bikes in between.

2.5.4 Ride-pooling

Multiple options for ride-pooling are possible. Commuters can travel together (carpooling), or on-demand shuttles can be utilized that can transport multiple employees (vanpooling). Although employees that already use public transport are most interested in vanpooling, it nevertheless has much potential to reduce commutes by car for ASML (Molin & Kroesen, 2023). From the same research, it is concluded that in general vanpooling is preferred over carpooling. A study conducted in the United States, found that vanpoolers report a 21% lower stress from commuting compared to solo commuters by car (Ditmore & Deming, 2018). In the United States, among public transportation modes vanpooling was on top regarding trip growth between 2006 and 2016 (Ditmore & Deming, 2018). Carpooling for commuting is not a common practice but high-occupancy-vehicle lanes can facilitate an increase and raise economic and environmental awareness (Molina et al., 2020). A distinction must be made here between on-demand service between work locations during the day and a service that is used to travel from home to work.

2.5.5 Fare-free public transport

In Hesse, a state in Germany, the share of public transport users for commuting of state employees was before a fare-free ticket already a common practice, but it significantly increased after introducing fare-free public transport (Busch-Geertsema et al., 2021). The reduction in car use however was only noticeable in certain subgroups, e.g. employees that did not have a reduction ticket before the introduction of the fare-free ticket. Busch-Geertsema et al. (2021) conclude that on a larger scale, the potential of this fare-free tickets is not realized yet and that it might help in combination with other policies to encourage car users to use the public transport. In Chile, a fare-free pass was randomly assigned to workers for two weeks (Bull et al., 2021). The timespan was a limitation while individuals had not the opportunity to change behaviour for a longer horizon. From their research the suggestion therefore rises that car travel is not sensitive to the change in transit fare. The suggestion is made to study effects of fare-free transit on a longer time frame to assess substitution between car and public transport. Nevertheless, an increase in number of total trips was found, but most in the off-peak hour trips. Focused on bus passengers, nudging with financial instruments is not enough but works better than the provision of information according to Franssens et al. (2021) after a field experiment in Rotterdam. A limitation however regarding commuting is that 21.5% of the passengers during peak hours received the card for the experiment while peak hour passengers are responsible for 71% of the trips in a work week. Another field study in the Netherlands in Groningen by Zeiske et al. (2021), found that financial incentives result in a temporary effect, but removing the incentive led to less people engaging in the desired behaviour. ASML has fare-free public transport since January 2023.

2.5.6 Remote working policy

Remote workers of ASML receive 2.15 euro for a full day of working from home. Motivation from ASML is to give employees flexibility to improve work-life balance and spend less time commuting. For the company it means reducing the CO_2 footprint by having fewer trips to and from the office.

A large scale survey in America prognosed 20% of the full workdays to be from home post pandemic, compared to 5% before the pandemic (Barrero et al., 2021). The five reasons evident for this according to the research are positive experiences with working from home, investments in equipment and know-hows to work effectively from home, a stigma change that working from home is associated with evading work, awareness of contagiousness on the work floor, and technological advancements that make working from home more convenient. Many employers embrace working from home therefore nowadays and have policies in place.

2.5.7 Frequency of public transport

Since 22 October 2023, the frequency in the rush hours of bus line 119 was increased from four times each hour to eight times each hour. This followed from agreements between governments and ASML to make the company better accessible. This bus line is the direct bus line from Eindhoven Central Station to buildings from ASML in Veldhoven. The literature review of An et al. (2020) shows that existing studies concerning the crowding of busses during peak hour are focused on demand management or supply management. Supply management is the category under which bus frequency falls. Demand management is for example cheaper off-peak fares. While a business-card is in place and people have a working schedule , the risk of demand management is substitution towards car. This is undesirable. Supply management on the contrary by increasing frequency, can alleviate crowding effects and also reduce delays during peak-hour (An et al., 2020). Case studies in the United States, resulted in the finding that frequent routes have already the highest passengers in each vehicle, but adding more busses does not lead to a higher average in bus occupancy (Berrebi et al., 2021). For ASML however, bus occupancy is somewhat less relevant (except for overcrowded buses). More important is how many people shift from car to bus because of the frequency during rush hours.

2.6 Coherence of weather, seasonality, and policy interventions

Individual characteristics, household characteristics and weather characteristics are all subject to context-dependency and ambiguity of direction and magnitude of variables, but the day-to-day variation of the latter makes it more appealing to associate it with the day-to-day differences in the modal split. As previously depicted, day-to-day variations in variables are important to consider, due to it being the greater part of individuals decision to cycle (Heinen et al., 2011). In the light of commuting, subsequent research from Heinen et al. (2013) states that besides work-related factors the choice of an individual commuter to cycle on a day-to-day level, weather is the most important additional factor that influences choice. This significant influence on mode choice is also evident in other research (Böcker et al., 2013; Liu et al., 2017; Sabir, 2011). The prevailing determinant for day-to-day variation in demand, especially for active transport, is weather (Gössling et al., 2023). Faber et al. (2022) mention that most papers measuring the effects of weather on travel behaviour find considerable effects, on travel volume level but also on mode choice level. It can therefore be concluded that for both reduction and alteration of travel behaviour, next to implemented policies, weather as external factors influences travel behaviour significantly according to research in the field.

Weather and climate took a more prominent place regarding topics of importance in travel behaviour research and transport planning (Böcker et al., 2016). This is due that climate change is higher on the political agenda. While weather is about day-to-day variations regarding temperature, rain, wind, etcetera, climate refers to the conditions in general over a longer period on a location (Sabir, 2011). A longer period is further specified by Heinen et al. (2010) to be 30 years regarding climate. While the transport sector is a huge contributor to climate change through emissions, the sector is also directly exposed to weather. One of the ways to change travel behaviour is by making policies. Weather as significant external factor for day-to-day travel behaviour is therefore particularly important to consider in policy considerations, while it is the only category of determinants for mode choice beyond direct control of policymakers.

There is consensus that policy interventions can be the catalyst for a shift towards sustainable mobility. Sustainable mobility is needed to mitigate the effect of climate change. Sabir (2011) stated that global warming can lead to the amplifying of daily weather variations subsequently. Temperature will rise in the Netherlands due to climate change, leading potentially to more extreme precipitation, winters that are wetter, and hotter summers (Sabir et al., 2010). Regarding the case, it is deemed an important consideration for decision-makers because of this prospect. According to Sabir et al. (2010), the total demand does not have a strong variation, but trips with different transport modes have in different weather conditions which suggests the relation between weather and a modal shift. Gössling et al. (2023) however, state that weather is also the prevailing determinant for day-to-day variation in demand.

As depicted in section 2.2.1.3., weather is thoroughly studied but still the focus is often on temperature and rain, causing wind to be overlooked. Also, there is no consensus about how to represent seasonality in research although it is recognized as a crucial factor. This thesis therefore aims to fill these knowledge gaps. Generalizing the effects of all determinants from previous research to this research is difficult while it turns out the direction and impact of determinants is not always equivalent in the literature (Heinen et al., 2010 ; Ton et al., 2019). As mentioned, context-dependency plays a role, but a huge research opportunity arises by evaluating past implemented polices towards sustainable mobility.

Figure 3 results from synthesizing from the literature review regarding travel behaviour. The findings in this literature review suggest a primary focus on alteration and reduction, with keeping efficiency in mind. The figure depicts a framework of the system that illustrates the relationships between the identified determinants, travel behaviour, impacts on sustainability and the influence of decision-makers. This framework operates within the context of the sustainable mobility paradigm as outlined by Banister (2008) and also within the sustainable commuting framework of Figure 3. It tries also to represent that external factors have both an immediate effect on determinants for travel behaviour as well as longer-term consequences that can affect mode choice and travel volume through feedback loops. This implies that not only long-term trends need to be addressed by policymakers but also focus is needed on short-term variability to achieve sustainable mobility. Determinants regarding season and weather are critical to consider, as they cause significant daily variations in modal choice and unlike other determinant categories, they cannot be directly influenced by decision-makers. This underlines the importance of getting insights in the effects and take these determinants into account in policy development. According to Ton et al. (2019), active mode policy cannot change individual characteristics and household characteristics. Despite this is in general true, active mode policy can impact the ownership of means of transport for example, which is a household characteristic. Other policies besides active mode policies can also impact these characteristics e.g. salary increase and employment conditions regarding having children. Therefore, arrows are included from decision-makers towards these categories of characteristics. The categories have examples evident from the literature, it can be that not all important determinants are in the figure present.

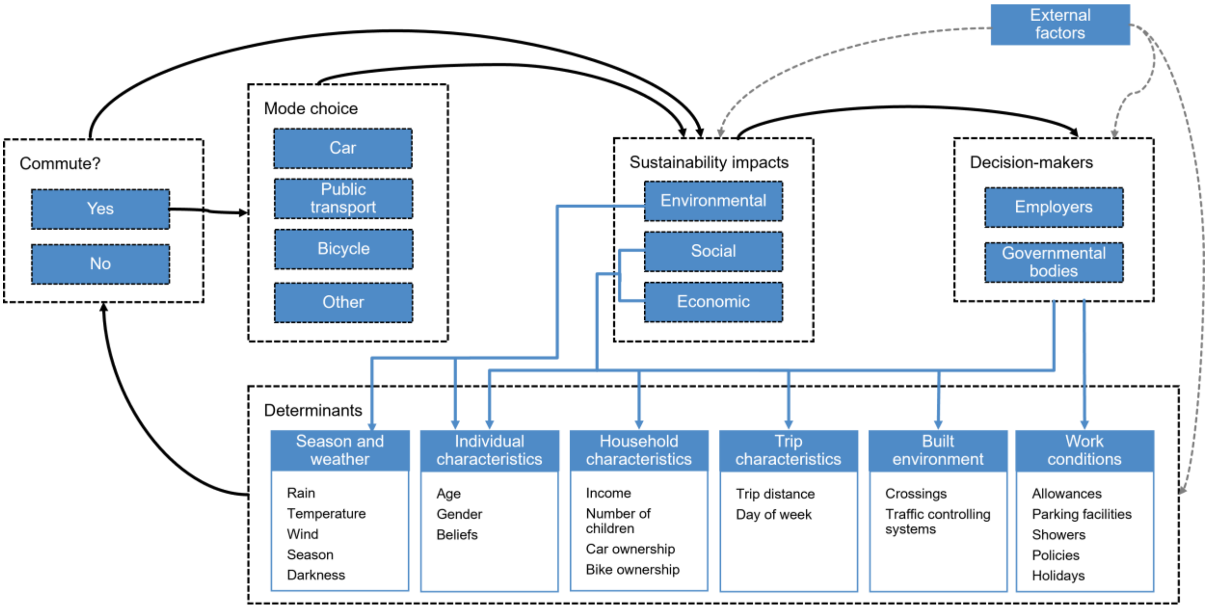


Figure 3: A framework for understanding commuter travel behaviour and the impacts towards sustainable mobility

This literature review and the urgency of a shift towards sustainable mobility as entailed in the introduction, serves as a foundation of knowledge regarding travel behaviour of commuters and why it is relevant to look into the relationships between determinants and travel behaviour to contribute towards more sustainable transport and the existing research. Building on the knowledge from the introduction and this literature review, the next chapter describes the techniques and procedures by which the subject and depicted framework is further investigated.

3 Methodology

In this chapter, the research methodology is outlined. The methodology is composed with the earlier mentioned Data Science Process as direction and foundation for the research strategy. First, the research strategy is explained, including the philosophy behind this research and the approach. This is followed by explaining the steps in the process in more detail for this research. Various research methods and techniques are used for this. To conclude this chapter, the limitations of this methodology are discussed. Following this methodology leads to being able to answer the sub questions and the overarching research question.

3.1 Research strategy

The Data Science Process, as described in Chapter 1 is used to guide this research. With the methodology, the research question can be answered. The focus is mainly on two steps of the Data Science Process: *Exploratory Data Analysis* and *Modelling*. Still, almost all steps are involved in the overall study. Figure 4 shows the Data Science Process from O'Neil and Schutt (2013) with a slight alteration. The dashed line between *Exploratory Data Analysis and Data Processing* is added, while insights from the *Exploratory Data Analysis* can lead to feature engineering (creating new variables) and encoding categorical variables, which is part of *Data Processing*.

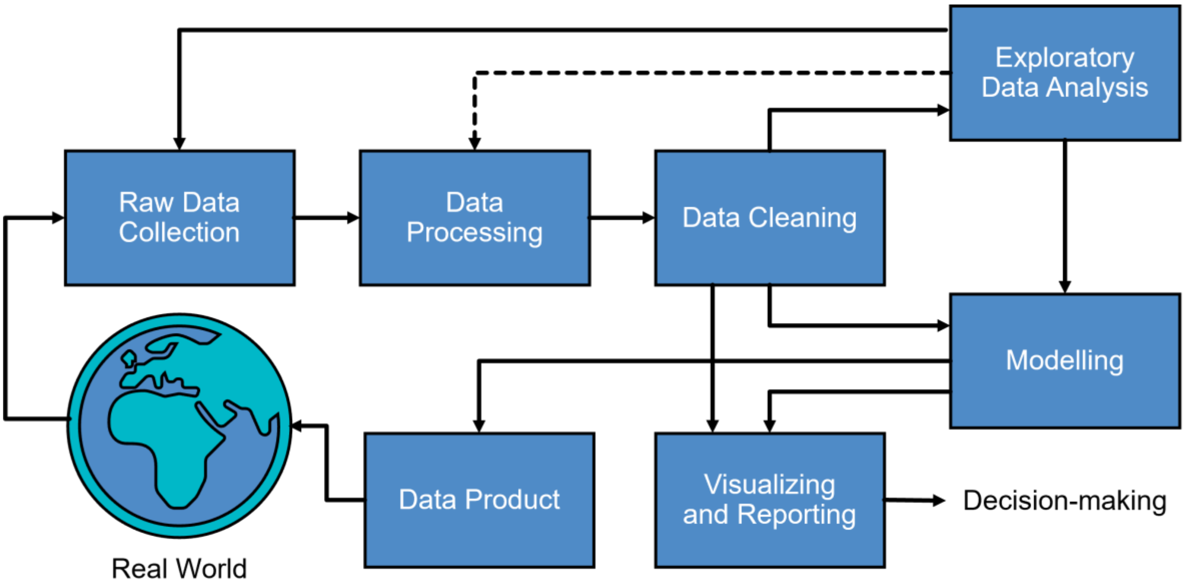


Figure 4: Data Science Process. Adapted from *Doing Data Science: Straight Talk from the Frontline* (p. 41), by C. O'Neil and R. Schutt, 2013, O'Reilly.

The philosophy behind this research is the positivism paradigm. The primary goal of positivism is to create explanatory associations that eventually contribute to predicting and mitigating the problem at hand (Park et al., 2020). This in line with the research question, to research to which extent determinants are associated with travel behaviour and the research objective to draw relevant findings that can contribute to a shift towards sustainable mobility. Park et al. (2020) also mention that from positivism a theory-verification approach is suited, facilitated by a hypothetico-deductive model. This model is based on testable hypotheses and empirically confirm or reject these hypotheses. This also fits this research, while the literature review identified associations between determinants and mode choice which can be tested whether these associations hold and to which extent within the described case.

This theory-verification approach falls in the domain of quantitative approaches, which fits best for problems calling for identifying factors that have an association with the outcome, measuring utility of interventions and comprehending of what the best predictors are for the outcome (Creswell, 2009). This once again underlines why it fits the nature of this research. As the research focuses on a specific setting to examine the extent of relationships, it can be classified as an explanatory case study. According to Yin (2018), this type of case study has the purpose to explain how or why a social phenomenon works, for which the quantitative theory-verification approach can be utilized.

The design consists of the following parts: how data for research is collected, how the data is analysed, and how the outcomes are interpreted (Creswell, 2009). These steps guide the rest of this chapter in explaining the methodology of this research accompanied by the steps from the Data Science Process.

3.2 Data collection and preparation

3.2.1 Raw Data collection

This research is primarily based on secondary data. Secondary data can come from external organizations, but also data collected within an organization and then reprocessed can be seen as secondary data. For research with limited time and resources, utilization of existing data is a practical option, while enormous amounts of data are globally collected and archived by researchers (Johnston, 2014). This research is focussed on analysing relationships between variables and policies of the past few years. Since this research has a time span reaching to dates before the research has started, data cannot be collected with a primary data method. While secondary data collection is used as method, three things need to be dealt with according to Hox & Boeije (2005): A search strategy for finding useful data, the way the data can be retrieved, and the consideration of the methodological quality of the data.

3.2.1.1 Weather conditions and seasonality

- *Search strategy:* After searching for a source with both location accurate data and time frequent data, the choice fell on OpenMeteo (Zippenfenig, 2023). Many different weather variables are monitored and there are also descriptions of these variables available. It is possible to provide coordinates and get estimates for each hour of every day.
- *Retrieve data*: OpenMeteo offers various Application Programming Interfaces (APIs). Both the Weather Forecast API and the Historical Weather API offered multiple convenient ways to retrieve the data. For this research, the data is directly converted into .XLSX and .CSV files. Also, as back-up, OpenMeteo offered a ready-made Python script to retrieve the data.
- *Methodological quality*: OpenMeteo uses combinations of weather stations, aircraft, radar, and satellite data and is based on reanalysis datasets (Hersbach et al., 2023; Muñoz Sabater, 2019; Schimanke et al., 2021). With reanalysis datasets and mathematical models, the historical weather information can be obtained for locations that do not have a weather station nearby.

3.2.1.2 Employees data and travel behaviour data

- *Search strategy:* Another search strategy is needed for this type of data Although the data is about the ASML employees, ASML does not originally own all the data. There are third parties that collect data for ASML, for example NS gathers the data about travel behaviour with the NS- business card. The required data has been requested via persons and teams within ASML with access and permissions to pass certain data for this research.
- *Retrieve data*: A favourable thing, is that ASML collected a lot of data already before the start of this research. However, retrieving internal data was a cumbersome process while not

having access directly to all mobility related data. There are a lot of regulations regarding the data and the data is not easily accessible at the level required for this research. There are some dashboards that are useful, but these do not go very deep into the data and are on an overarching level. Therefore, help from intermediaries that have access to the underlying database was needed and helpful, resulting in mainly aggregated data on a daily level.

- *Methodological quality*: While the total picture of the data is an assemblage of collected data from different parties this involves a variety of methodologies behind the data. Intermediaries who retrieve data from the database can use different queries and sources. Assumptions and rules for querying and the underlying database can therefore produce different results. This will be discussed in more detail in section 3.2.2.

3.2.1.3 Other data

The research also requires time-related data: start of Daylight Saving Time, holidays, and when policies are implemented by ASML or in the region. Much of this data is widely available on the internet and common sense or could be retrieved within ASML. To briefly place it in the context of the work of Hox & Boeije (2005): searching the internet and asking policymakers was the strategy, noting the dates down the way of retrieving and for methodological quality there is not much to argue about these factual dates of holidays and implemented policies.

3.2.1.4 Overview of datasets

Table 4 depicts the datasets used or gathered for this thesis. Beside these datasets, also data from the Spotfire Dashboard (A&M Modal Split Trend Dashboard) is used. This is further elaborated in Chapter 4.

Table 4: Overview of datasets

Name of the dataset and source	Rows x columns	Time range (MM/DD/YYYY)	file	Description of the dataset
Daily_bike_events(in) (ASML)	3560x3	03/01/2023 – 01/07-2025	.csv	For each day for each bicycle counting camera, the number of registered incoming cyclists .
NSGO_events_to_campus (ASML)	22071x6	01/02/2023 – 01/05/2025	.csv	For each day for each Origin-Destination combination the number of check-outs, coupled to Run 1000 or Run 6000.
NSGO_to_Ehv_Centraal (ASML)	20760x6	01/01/2023 – 01/05/2025	.csv	For each day for each Origin to Eindhoven Centraal the number of check-outs and train class.
PowerBI_bus_to_campus	642x6	01/01/2023 – 09/30/2024	.xlsx	For each day, the number of check-outs for Run 6000 bus stops (the exported data includes also minimum temperature, maximum temperature, and rainfall)
open-meteo-FORE_DAILY (Open Meteo)	702x23	01/02/2023 – 11/29/2024	.csv .xlsx	Forecast for each day at the coordinates of ASML Headquarters of weather conditions and seasonal variables.
open-meteo-FORE_HOUR (Open Meteo)	16755x45	01/02/2023 – 11/29/2024	.csv .xlsx	Forecast for each day for each hour at the coordinates of ASML Headquarters of weather conditions and seasonal variables.
open-meteo-HIST_DAILY (Open Meteo)	702x21	01/02/2023 – 11/29/2024	.csv .xlsx	Historical data for each day at the coordinates of ASML Headquarters of weather conditions and seasonal variables.
open-meteo-HIST_HOUR (Open Meteo)	16756x31	01/02/2023 – 11/29/2024	.csv .xlsx	Historical data for each day for each hour at the coordinates of ASML Headquarters of weather conditions and seasonal variables.

3.2.2 Data Processing and Data Cleaning

While not the main focus of this research, processing and cleaning data consumes a lot of time and is a complex process. More and more data is collected nowadays and many organizations rely on data-driven decision-making, but retrieving (dirty) data from different sources can lead to loss of quality (Ridzuan & Zainon, 2019).Data processing is the set of modifications to the collected data to organize it into a useful format for further use. Data cleaning is the set of modifications to remove or fix incorrect, duplicate, and incomplete data. The data is gathered in different files and formats and therefore processing and cleaning is needed to result in structured data This data can then be used for Exploratory Data Analysis, modelling, visualising, and reporting. Processing and cleaning are two preparatory steps that are essential as a quality control of this research. Yet, in the Exploratory Data Analysis insights can be gained that will require to reprocess parts of the data.

For data processing, the method is electronic processing. This entails using computer software to handle the data. Microsoft Excel, RStudio and Notepad are used for the purpose of electronic processing. The same software programs are used for data cleaning. Data cleaning includes rather techniques than methods. For example, the addressing of inconsistencies in variable names and handling missing data. Also, transforming, and aggregating data that is at different granularities. Techniques used for data processing and data cleaning include pivot tables, comparing available datasets, handling missing data, convert variable types. In Chapter 4, it is depicted how data is processed and cleaned and how variables are obtained. Also, an overview of the variables is added with a description. In this section, the non-standard methods are explained, which are Multivariate Imputation by Chained Equations (MICE) and encoding.

3.2.2.1 MICE imputation

Multivariate Imputation by Chained Equations (MICE) is a technique for missing data in more than one variable, in the domain of fully conditional specification approaches which can be implemented with software (Van Buuren & Groothuis-Oudshoorn, 2011). Following Azur et al. (2011), imputation of missing values is based on the observed values and the relationships with other variables. This technique is used for the cycling cameras, in which it was discovered that some cameras occasionally did not register commuters. A short example why this approach works for this kind of missing data. With this approach, both horizontal and vertical data is used: horizontal data are the observations of each individual camera on a day (row in the data), and vertical data are the observations of an individual camera over the days (column in the data). This is useful and appropriate for data that is the result of summing other data, while both the individual pattern as well as the relationships with other cameras are used for imputing missing data to result in the total number of bicycle commuters on each day.

Multiple completed datasets are created that predict the missing values, which can account partly for uncertainty and lead to smaller and more accurate standard errors (Azur et al., 2011). According to Van Buuren (2018), between 5 and 20 imputations for a moderate amount of missing data is sufficient when the primary interest is on the point estimates of coefficients. Therefore 10 imputations are chosen with method 'Predictive Mean Modelling', which is the default. Whether the algorithm is converged can be checked by plotting the imputations against the iterations. According to Van Buuren (2018), lines that are crisscrossing each other and no defying trends are visible. This is inspected in Chapter 4. All ten imputations can then be used for further use after which the estimates can be pooled. It is not recommended to simply average the imputed values, while the between-imputation variability is ignored and the uncertainty related to the underlying data cannot be fairly represented by averaging the data (Van Buuren, 2018). Figure 5 shows an adapted version of the scheme of main steps in multiple imputation for this research. As can be seen, this method has overlap between data preparation and modelling.

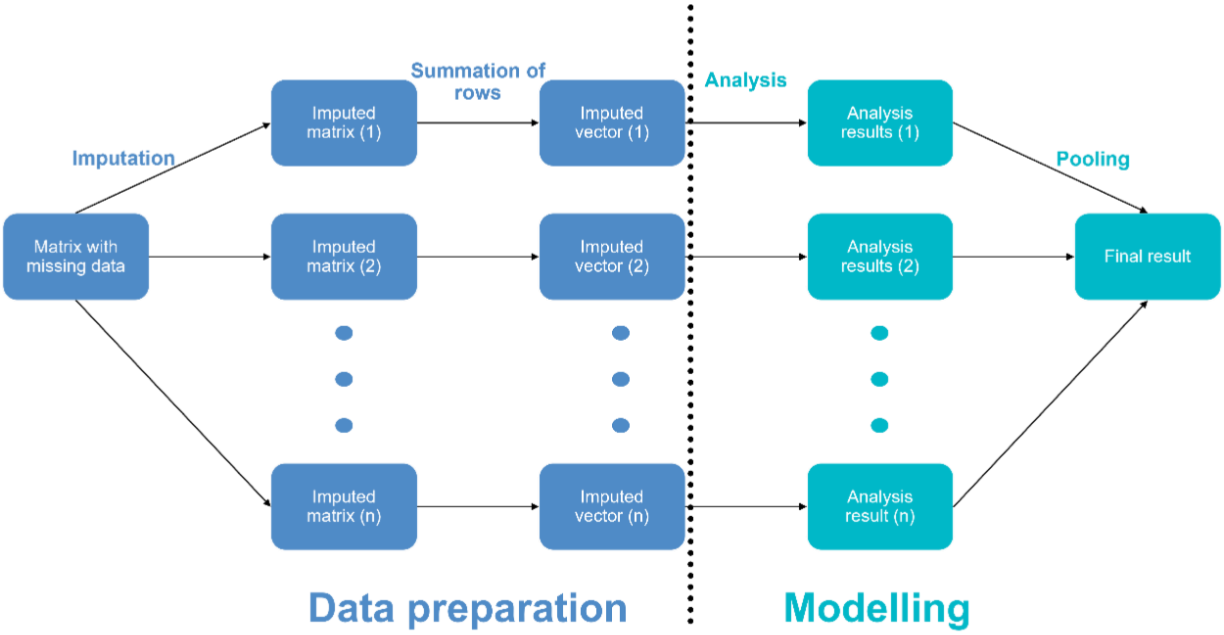


Figure 5: Scheme of main steps of Multiple Imputation by Chained Equations (MICE)

The idea of pooling is explained in this section for completeness. Rubin's rules are used, which can pool parameter estimates of regression coefficients and standard errors. The rules appeared in the book of Donald B. Rubin in 1987 called 'Multiple Imputation for Nonresponse in Surveys' (Rubin, 1987) . The formula for a mean or coefficient is straightforward and depicted in Equation 1.

$$\boldsymbol{\bar{\theta}} = \frac{\mathbf{1}}{\boldsymbol{m}}\left(\sum_{\boldsymbol{i=1}}^{\boldsymbol{m}} \boldsymbol{\theta_i}\right) \tag{1}$$

In Equation 1, the total number of imputed variables is denoted as m, and the coefficient for each imputed variable i depicted with θ_i. This makes $\bar{\theta}$ the mean of coefficients.

Regarding standard errors, applying Rubin's rules leads to properly accounting for the variability within the imputed datasets and the uncertainty between them. This leads to statistical inference that is more robust. The three equations together in Equation 2 are the Rubin's rules for standard errors.

$$Var_{within} = \frac{\sum_{i=1}^{m} SE_i^2}{m}$$

$$Var_{between} = \frac{\sum_{i=1}^{m} (\theta_i - \bar{\theta})^2}{m-1}$$

$$SE_{pooled} = \sqrt{Var_{within} + Var_{between} + \frac{Var_{between}}{m}} \tag{2}$$

As in Equation 2 can be seen, the pooled standard error is a combination of the variance between standard errors of imputations and variance within the standard errors of imputations.

3.2.2.2 Encoding

Categorical variables often have categories with non-numeric values. For algorithms to work, these variables need transformation into a numerical format. This is called encoding. Two types of encoding can be used: One-Hot Encoding and Dummy Encoding. One-Hot encoding creates n columns for n categories coded with ones and zeroes. According to Wooldridge (2010) this causes multicollinear variables, which cannot be processed by some models (also known as the Dummy Variable Trap). Dummy Encoding is almost the same, but $n - 1$ columns are created for n columns, with one category being the reference category. Next to encoding variables manually, some models handle encoding automatically. Figure 6 clarifies encoding with a fictional categorical example of mode choice.

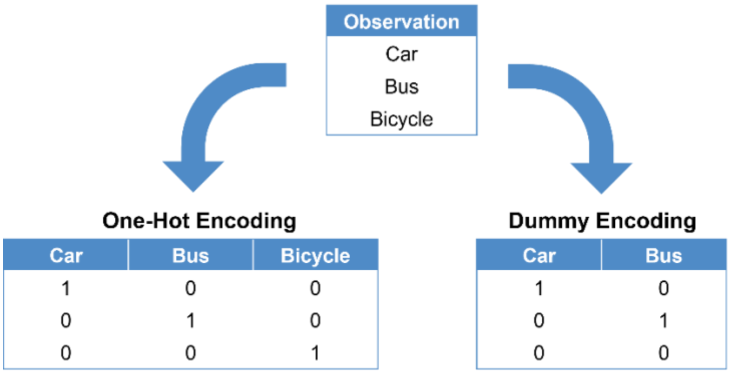


Figure 6: Encoding categorical variables

3.2.2.3 Final datasets

Processing and cleaning resulted in two final datasets: one on a daily level and one with data regarding peak hours. The final dataset that is used for the Exploratory Data Analysis consists of 500 rows and 113 columns and is a combination of the two final datasets. Rows depict daily observations ranging from 1st of January 2023 and 29th of November 2024. Columns consist of response variables, imputed vectors weather and seasonality variables for specific hour as well as for a daily level, policy dummy variables, and other variables.

3.3 Data analysis methods

RStudio, which is an integrated development environment (IDE) for programming language R, is used for data analysis (R Core Team, 2024). The version used in this research is *R version 4.3.3*. Programming language R is open-source and developed for statistical computing and graphical visualization. The primary repository for R packages is called the Comprehensive R Archive Network (CRAN). Packages can be downloaded easily from CRAN and installed within RStudio to extend the capabilities of base R. Appendix B provides a list of the used packages and for which purpose.

3.3.1 Exploratory Data Analysis

This method is used particularly to answer the first sub question: *What are the characteristics of the data regarding weather conditions, seasonality, and travel behaviour of ASML employees?"*. The goal of Exploratory Data Analysis (EDA) is to develop an understanding of the data before applying a model on the data. EDA was introduced by John Tukey in his book *Exploratory Data Analysis* (Tukey, 1970). Conducting EDA is often important to check assumptions, inspect anomalies and inform for the selection of features for modelling. Not only is EDA useful for the first sub question to understand the data, it can be also utilized for the second question to discover whether there is a linear or other relationship between dependent and independent variables as foundation for model choices. EDA was conducted to assess the characteristics and distributions of the data, identification of potential issues with data quality, and to inform modelling decisions regarding transformations and encoding of variables. Knowledge from the literature review helped with the thinking behind made choices and meaningful visualizations.

3.3.2 Modelling: overview

A model is a simplified representation of reality. As statistician George Box famously once said: "All models are wrong, but some are useful". In this step of the Data Science Process therefore, the aim is to make the best possible model given the data, time, and resources but also having in mind that it needs to interpretable for the purpose of this research. This step of the Data Science Process is of great importance for answering sub question 2, sub question 3, and sub question 4. The modelling consists of two main components: regression analysis and time-series analysis. Regression is used to establish the baseline relationships between variables and the number of commuters. This results in interpretable coefficients and main drives behind travel behaviour and is a relatively simple model to incorporate the knowledge of the literature review and EDA to see what the data unveils. With regression models, the second sub question can be answered: *"What are the (static) relationships of daily weather conditions and seasonality on the modal split and travel volume of ASML employees?".*

However, this regression methods assume that the data is time-independent and therefore may overlook patterns over time, which can also cause some violations to the assumptions on which regression is based. Time-series analysis is therefore utilized to capture the dynamic effects which provides more robust insights into the relationships of variables, especially implemented policy interventions. This enables to find an answer on the question "*How are time related patterns and the implementation of policies associated with commuting patterns over time?"*, which is the fourth sub question.

As said, the literature review and the Exploratory Data Analysis, provided useful knowledge for inclusion of variables and the expected directions (or lack thereof) of relationships. However, modelling is an iterative process and certain variables can be altered, added, or deleted from the models even if this seems contrary to the literature review or EDA. This can also be dependent between means of transport. Next to adding and deletion variables, interactions and policies are tested across modelling techniques to answer the third sub question: *"How do polices and the interactions between daily weather and seasons relate with the modal split and travel volume of commuters?".*

3.3.3 Modelling: regression analysis

Regression analysis is a method that is used to investigate relationships between variables. For this study, regression analysis is used to explore the relationships of weather conditions, seasonality, and policies on travel behaviour of commuters. A distinction is made between three transportation options (car, bicycle, and bus) for the last mile and two measures of transport, which are travel volume and mode share. Both measures of transport require another kind of regression analysis. For travel volumes, linear regression with ordinary least squares (OLS) is utilized. This model is part of the research on which variables are related to commuters travelling less (*reduction approach*). For mode share a Multinomial Logistic Regression (MLR) is used, which is part of the research on which variables are related to commuters travelling different (*alteration approach*).

The following steps in regression analysis, depicted by Chatterjee and Hadi (2015), are approximately followed

1. Statement of the problem
2. Selection of potentially relevant variables
3. Data collection
4. Model specification
5. Choice of fitting method
6. Model fitting
7. Model validation and criticism
8. Using the chosen model(s) for the solution of the posed problem

One difference is that data is collected in advance of the selection of potentially relevant variables. The posed problem in the case of this research is the research question. Also, the model specification and choice of fitting method go hand in glove together. Model selection is an iterative process, which is not included specifically in one of the steps. Selection of the model can cause the necessary revisit of step 4 up until step 7. In step 4 the model is specified, if in step 6 it is concluded that some variables need to be added or removed the specified model needs adjustment. If in model validation for example outliers are detected, the underlying data needs to be changed and the model needs to be re-fitted.

3.3.3.1 Ordinary least squares regression

The relationships between variables and travel volume are researched with linear regression models. For each transportation option, a separate model is built with the same predictors and interactions. This allows for different estimated coefficients for the distinct transportation options. For volume, the assumption is used that part of the number of employees choose to commute to the office and choose a mode of transport for that purpose. The interplay between the modes of transport is neglected and is covered in the model for the modal split. The response variables are technically discrete counts, as these representing the number of commuters on a day for either car, bus, or bike commuting. Typically, counts are modelled with Poisson regression or Negative Binomial Regression but because of the following reasons, it was chosen to perform linear regression with the procedure of ordinary least squares (OLS):

1. **Interpretability:** as the goal is to explore the static relationships, interpretability of OLS regression is more straightforward by coefficients that can be interpreted as the change in response variable (number of commuters) by a unit change of a predictor when all other variables remain constant. Poisson regression and Negative Binomial regression see the response as a logarithm of the expected value, therefore estimates of coefficients are on the log scale. By exponentiating the coefficients rate ratios are obtained which can be used to calculate the expected count. For policy analysis, OLS regression is a useful method while it helps to understand the relationships between variables and see what the extent of relationships are. With the research objective in mind to draw relevant findings for decision-makers, regression is among the category of machine learning models that are intrinsically explainable.
2. **Large counts:** The two problems with count data in OLS regression, are data along the bound of zero which can cause negative estimates and predicted values with decimals. The commute options in this research have counts of multiple hundreds till thousands. While on working days there is always travel volume an observation close to zero is extraordinary. Rounding decimal estimates to whole numbers will not make a significant difference due to the range of the number of employees and error terms: e.g. 3000 or 3001 does not really matter while a difference between 3000 and 4000 is much more interesting and the error term will likely be much larger than one.
3. **Practicality:** The book 'Regression Analysis of Count Data' (Cameron & Trivedi, 2013) acknowledges that due to potential negative values and the homoscedastic variance function the OLS estimator can be inappropriate, but in practice the results are quite similar qualitatively to Poisson and Negative Binomial estimators. Since the variance function is flawed anyway as it does not consider the time component, the Poisson and Negative Binomial would also not have the right error structure in this research. The book continues with stating that the most significant predictors in OLS are also the most significant predictors in Poisson, which makes OLS regression convenient for determination of important variables. This also ties in with the aims of this research. Researchers have to deal with an important trade-off between scientific quality and conveying the results to policymakers (Handy et al., 2014). Regression is widely used in research and the interpretability of this statistical method provides a good balance between the scientific quality and making the results explainable to policymakers.

Ordinary Least Squares (OLS) is therefore chosen as fitting method, which minimizes the sum of squared residuals. Interpretability and efficiency make this a widely used method in linear regression. Regarding regression, there is a lot of interchangeable use regarding the terms multivariate and multivariable regression, while the subtle distinction needs to be brought to attention (Ebrahimi et al., 2021) . Multivariate regression relates to regression with more than one dependent variable. This can be with one or multiple predictors. Multivariable regression is regression with multiple predictors but one outcome. With a systematic approach, Hidalgo and Goodman (2013) found that 83% of the analysed articles wrongfully used "multivariate" instead of "multivariable". This illustrates that it is rarely used correctly and therefore the intention is to adhere to the above distinction between these two in this thesis. The separate regressions performed are therefore multiple univariate multivariable regressions.

In this thesis, the dependent variables (responses) are the number of commuters by bicycle, number of commuters by public transport, and number of commuters by car for travel demand, and fractions of this compared to the number of employees in office for modal split. There are also multiple predictors, e.g. precipitation, temperature, weekday. The research units are days. To be more specific, weekdays. For a dataset with n observations, let y_i denote the response variable for observation i

($i = 1, \dots, n$). The intercept term is β_0, and β_j is the coefficient of predictor j ($j = 1, \dots, p$). The value of predictor j for observation i is denoted by $x_{i,j}$ with the error term for observation i being ϵ_i. This general multivariable regression model can be specified as follows:

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{i,j} + \epsilon_i \tag{3}$$

This specification holds for continuous variables (coefficient multiplied with the value of the continuous variable) and binary variables (the coefficient is the difference in effect for the category coded with ‘1’ in comparison with the category coded ‘0’ on the response). However, for a categorical variable with three or more categories this formula does not hold: when regarding the categorical variable as a predictor, if there are three categories there is a need of two different coefficients. This does not fit in the proposed formula while it allows for one coefficient for each predictor. Therefore, the multivariable regression model can be specified as follows for each observation and adjusted to allow for categorical variables:

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{i,j} + \sum_{c=1}^{C} \sum_{d=1}^{m_c - 1} \gamma_{c,d} D_{i,d}^{(c)} + \epsilon_i \tag{4}$$

Equation 4 extends the simple model of Equation (3 by specifying a term for including categorical variables. The number of dummy variables of categorical predictor c ($c = 1, \dots, C$) with m categories is equal to $m_c - 1$. This is in line with Dummy Encoding instead of One-Hot Encoding to prevent for the Dummy Variable Trap by having a reference category. The coefficient for the d-th category (excluding reference category) of categorical predictor c is $\gamma_{c,d}$. The value of the d-th category of categorical predictor c for observation i is denoted by $D_{i,d}^{(c)}$. The other components of the model are defined as in Equation 2.

The categorical part, $\sum_{c=1}^{C} \sum_{d=1}^{m_c-1} \gamma_{c,d} D_{i,d}^{(c)}$, is therefore the sum of the sums of each categorical predictor with its corresponding coefficients. The sum of each categorical predictor is just a series of zero’s and one value while the categorical predictor is dummy-coded by the model, which is illustrated in Table 5.

Table 5: Dummy coding

Observation i for Predictor S	$D_{i,1}^{(S)}$	$D_{i,2}^{(S)}$	$D_{i,3}^{(S)}$
Winter	0	0	0
Spring	1	0	0
Summer	0	1	0
Autumn	0	0	1

If we have for example: $\gamma_{S,1} = 100$, $\gamma_{S,2} = 300$, $\gamma_{S,3} = -100$, an observation in the summer will lead to: $\sum_{k=1}^{3} \gamma_{S,k} D_{"Summer",k}^{(S)} = 100 * 0 + 300 * 1 - 100 * 0 = 300$, which means a 300 mean difference with respect to reference category winter.

While from the literature it became evident that people can perceive a change in weather different in the summer than in the winter, the regression model can be extended with interaction terms. Interaction terms lead to a more complex model, but the advantage of creating interaction terms between weather variables and seasons, is that while seasons are categorical there are distinct contexts for weather effects that stay interpretable and can provide new insights. This leads to the final model specification, which consists of an intercept, main effects of continuous and binary predictors, main effects of categorical variables, the interactions between a subset of the predictors with the categorical variable for seasons, and the error term:

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{i,j} + \sum_{c=1}^{C} \sum_{d=1}^{m_C-1} \gamma_{c,d} D_{i,d}^{(c)} + \sum_{j\in J} \sum_{s=1}^{m_S-1} \delta_{j,s} x_{i,j} D_{i,s}^{(S)} + \epsilon_i \tag{5}$$

The first parts and error term ϵ_i are defined the same as in Equation 4. The new part is the sum over interactions between continuous or binary predictors with the categorical variable season. Each predictor j in J, where J is a subset of $\{1, \dots, p\}$ and shown in the Equation 5 as $j \in J$, has a coefficient $\delta_{j,s}$ for the interaction with the the s-th category of categorical predictor season. The value of this s-th category of predictor season for observation i is denoted as $D_{i,s}^{(S)}$.

This model is implemented with the *lm* function of package *stats* in RStudio for employees by car and employees by public transport. For model of employees by bicycle some alterations are needed. The response variable bicycle exists of ten imputed columns and as entailed, it is better to fit separate models and pool the coefficients (function *pool* from package *mice*) rather than averaging the imputations and fit a model while this neglects the variability between imputations. There is no direct way to fit and predict values for pooled coefficients in RStudio as for linear models without pooled coefficients. Therefore, a function was made as workaround to make it possible to fit and predict values for pooled coefficients.

The model summaries of the resulting models can be inspected for coefficients, standard errors, and p-values. The model with and without interactions are also assessed with a statistical test called ANOVA, which is an acronym for analysis of variance. In RStudio, the function *anova* is used from package *stats*. With this test it can be assessed whether each term significantly contributes beyond random variation. Model comparison via F-test, which is also a possibility with the *anova* function can test whether the interaction terms improve the model significantly. The next step is assessing the assumptions: linearity of the data, normality of residuals, homogeneity of residuals variance, independence of errors, lack of multicollinearity. These assumptions are assessed with various plots. The most important assumption, linear regression assuming a linear relationship, can be violated by non-linear relationships or outliers. As Draper and John (1981) point out, outliers might be influential observations but it also possible that they do not influence the model. Influential observations can impact the estimates and standard errors of regression substantially and lead to bias and incorrect inference. Therefore, it is needed to deploy diagnostics on regression models to identify outliers and influential points (Ayinde et al., 2015). Cook's distance was introduced in the paper by Cook (1977) as a measure that estimated the mentioned influence of an observation on the regression model. Cook's distance is commonly used, and a higher Cook's distance means a higher influence of the observation on the regression estimates. The following formula depicts Cook's distance:

$$D_i = \frac{\sum_{j=1}^{n} (\hat{Y}_j - \hat{Y}_{j(i)})^2}{p \times MSE} \tag{6}$$

In Equation 6, Cook's distance D_i for observation i is the sum of the squared differences between the predicted value of the dependent ($\hat{Y}_j$) and the predicted value of the dependent when a certain observation i is removed ($\hat{Y}_{j(i)}$). This is divided by the multiplication of the mean squared error (MSE) with the number of predictors p. With visual inspection and a cut-off, observations can be identified as influential points and where necessary be removed. The model is re-fitted afterwards with the dataset without removed observations to reach a better model, estimates and accordance with the assumptions of linear regression. For the regression analysis, therefore a deletion approach is utilized to deal with outliers and influential points while observations are regarded to be independent. An often used threshold for Cook's distance is $4/n$, while also three times the mean is commonly applied.

After refitting, the coefficients can be judged again, the interactions checked, and the assumptions of linear regression models assessed. One of the assumptions will still be problematic: independence of errors. As entailed, the key assumption of OLS regression are i.i.d. errors. However, it is a bold statement that while the variables are time-series that the residuals would not be autocorrelated. Autocorrelation in the residuals means that there are still patterns in the data that are unextracted. There are two common ways to deal with autocorrelation:

1. Use of heteroskedasticity and autocorrelation (HAC) robust standard errors
2. Regression with time series errors (further discussed in 3.3.4 Time Series Analysis)

The first option is to keep the same model specification and use heteroskedasticity and autocorrelation (HAC) robust standard errors instead of the conventional one. This can be implemented with package sandwich in R, which has a function that uses the HAC variance-covariance estimator from the paper of Newey and West (1968) . The travel demand models are therefore as the last step of the regression part, re-fitted with HAC robust standard errors. For the second option, another model specification is needed. This is part of section 3.3.4.

Interactions cause additional model complexity; to get sparser models therefore with a stepwise approach automatically the predictors are assessed after which unnecessary interactions can be deleted. This means, a model that does not statistically differ when removing a variable. In addition, if with HAC standard errors all interactions between the same weather condition are insignificant ($p > 0.05$), the interaction is also deleted next to the deleted variables from the stepwise approach.

3.3.3.2 Multinomial Logistic Regression

In contrast to the three transport demand models with linear regression, just one model is established for the modal share. The relationships between variables and travel volume are researched with linear regression models. While the modal split is basically a breakdown of the number of employees in the office by the three modes of transport, the modal split has the restriction that it is bounded to a total percentage of 100%. For this purpose, it works better to use a model in which a predictor takes a percentage off of one mode of transport and then adds it to the other modes of transport. With Multinomial Logistic Regression (MLR) the probability of each outcome can be modelled with predictor variables to change the log-odds with regard to the reference category resulting in changes in probabilities. For this purpose, recognition is given to the category of employees of whom it is unknown which means of transport is taken. This 'Other' category can be found by subtracting the registered number of commuters of the three means of transport from the employees in office on a day. This leads to percentages for each means of transport and the unknown means of transport. A basic Multinomial Logistic Regression model can be specified in two ways. First, the log-odds equation, which is used to implement the model and is useful for interpretation of coefficients. Second, the probability equation that show the probabilities for each mode choice are calculated. The general formulas are depicted first, starting with the log-odds formula:

$$ln \frac{\Pr\,(y_i = k)}{\Pr\,(y_i = K)} = \beta_{0,k} + \sum_{j=1}^{p} \beta_{k,j} x_{i,j} \tag{7}$$

Equation 7 is the log-odds formula, where $\Pr(y_i = k)$ is the probability of observation i for the k-th outcome variable, with $1 \leq k < K$. The outcome used as reference is K. $\Pr(y_i = k)$ is therefore the probability of observation i for reference outcome K. The intercept term for the k-th outcome variable is $\beta_{0,k}$ and coefficients for predictors j for the k-th outcome variable are depicted with $\beta_{k,j}$. The value of predictor j for observation i is denoted by $x_{i,j}$, which is equal to Equation 3 , Equation 4, and Equation 5. No error term is present in the log-odds model while the model is constrained to 100% and an additive term would make it possible to go outside the defined range. The uncertainty is therefore handled implicitly through the probabilistic framework rather than in explicit residuals as in linear regression. The log-odds equation models the relationship between predictors and the likelihood of choosing another category than the reference. The log-odds estimates can be converted to probabilities which uses a SoftMax function. For the reference outcome, the following equation can be specified:

$$\Pr(y_i = K) = \frac{1}{1 + \sum_{h=1}^{K-1} e^{\left(\beta_{0,h} + \sum_{j=1}^{p} \beta_{h,j} x_{i,j}\right)}} \tag{8}$$

The denominator in Equation 8 is the total weight of all outcome variables including reference. Exponentiating the log-odds formula gives the weight for each outcome variable. Note that h is used in the denominator instead of k. This is due to confusion that may arise in Equation 9 and consistency. The 1 in the denominator follows from the exponentiated log-odds of reference with itself, which in $\ln(1)$, and $e^{\ln(1)} = 1$.

As the equation ensures that all probabilities together sum to 1, the non-reference outcome variable probabilities can be calculated as follows:

$$\Pr(y_i = k) = \frac{e^{(\beta_{0,k} + \sum_{j=1}^{p} \beta_{k,j} x_{i,j})}}{1 + \sum_{h=1}^{K-1} e^{(\beta_{0,h} + \sum_{j=1}^{p} \beta_{h,j} x_{i,j})}} \tag{9}$$

The denominator is the same as in Equation 8. The nominator is the exponentiated function of the right-hand side of Equation 7. For the outcome variable k in question therefore, in the nominator k is used. In the denominator of Equation 9, h is used for the outcome variables and the weights of the sum over $K - 1$, which represents the same range of values as the earlier depicted $1 \leq k < K$. Equation 8 and Equation 9 will for each observation i ensure that the sum of probabilities for the modalities is equal to 1. When using the same predictors and interactions as for the linear regression, the following formulas are the result:

$$ln \frac{\Pr(y_i = k)}{\Pr(y_i = K)} = \beta_{0,k} + \sum_{j=1}^{p} \beta_{k,j} x_{i,j} + \sum_{c=1}^{C} \sum_{d=1}^{m_C - 1} \gamma_{k,c,d} D_{i,d}^{(c)} + \sum_{j \in J} \sum_{s=1}^{m_S - 1} \delta_{k,j,s} x_{i,j} D_{i,s}^{(S)} \tag{10}$$

$$\Pr(y_i = K) = \frac{1}{1 + \sum_{h=1}^{K-1} e^{\left(\beta_{0,h} + \sum_{j=1}^{p} \beta_{h,j} x_{i,j} + \sum_{c=1}^{C} \sum_{d=1}^{m_C - 1} \gamma_{h,c,d} D_{i,d}^{(c)} + \sum_{j \in J} \sum_{s=1}^{m_S - 1} \delta_{hj,s} x_{i,j} D_{i,s}^{(S)}\right)}} \tag{11}$$

$$\Pr(y_i = k) = \frac{e^{\left(\beta_{0,k} + \sum_{j=1}^{p} \beta_{k,j} x_{i,j} + \sum_{c=1}^{C} \sum_{d=1}^{m_c - 1} \gamma_{k,c,d} D_{i,d}^{(c)} + \sum_{j \in J} \sum_{s=1}^{m_S - 1} \delta_{k,j,s} x_{i,j} D_{i,s}^{(S)}\right)}}{1 + \sum_{h=1}^{K-1} e^{\left(\beta_{0,h} + \sum_{j=1}^{p} \beta_{h,j} x_{i,j} + \sum_{c=1}^{C} \sum_{d=1}^{m_c - 1} \gamma_{h,c,d} D_{i,d}^{(c)} + \sum_{j \in J} \sum_{s=1}^{m_S - 1} \delta_{hj,s} x_{i,j} D_{i,s}^{(S)}\right)}} \tag{12}$$

Equation 10 depicts the log-odds equation for a model including categorical variables and interactions. Equation 11 is the probability equation for the reference outcome variable and Equation 12 is the probability equation for the non-reference outcome variables. Most of the components are already entailed in subsection 3.3.3.1, but while some characters are different to distinguish between what is in question and the rest the following list provides an overview of the components of Equation 10, Equation 11, and Equation 12:

- $\Pr(y_i = k)$: The probability of observation i for the k-th outcome variable, with $1 \leq k < K$.
- $\Pr\,(y_i = K)$: The probability of observation i for reference outcome K.
- $\beta_{0,k}$: Intercept term for the k-th outcome variable.
- $\beta_{k,j}$: Coefficient of predictor j for the k-th outcome variable
- $x_{i,j}$: Value of predictor j for observation i.
- $m_c - 1$: Number of dummy variables of a categorical predictor c with categories d.
- $\gamma_{k,c,d}$: Coefficient for the d-th category of categorical predictor c for outcome variable k.
- $D_{i,d}^{(c)}$: value of the d-th category of categorical predictor c for observation i.
- $j \in J$: All predictors j in J, where J is a subset of $\{1, \dots, p\}$
- $m_s - 1$: Number of dummy variables of predictor season with categories s.
- $\delta_{k,j,s}$: Coefficient for the interaction between predictor j and the s-th category of categorical predictor season for outcome variable k .

The denominator uses again h instead of k. This model is implemented in RStudio with the function *multinom* from package *nnet*. Akaike Information Criterion (AIC) is used to assess the fit of the model. A lower AIC value indicates a better trade-off between the model fit and the complexity and with a Likelihood Ratio Test (LRT) nested models can be compared to support the choice between model fit and complexity.

While it is not directly possible to apply Cook's Distance, another method is needed for getting rid of outliers. A common option is to fit separate logistic regression, perform a process to identify outliers and then remove outliers and re-fit for a multinomial logistic regression. However, the approach chosen in this thesis is to use the combined influential points of the travel volume model. After re-fitting the model, the coefficients can be inspected again, the interactions checked, and the assumptions of linear regression models assessed. The model can be used to predict the past values to check whether the model captures the same patterns as the actual data. With model performance, it can be assessed how accurate the model predicts.

3.3.4 Modelling: Time Series Analysis

For Time Series Analysis, there are two areas of interest: *Time Series Analysis with exogenous variables* (section 3.3.4.1) and *Time Series Analysis focussing on only the time series itself* (section 3.3.4.2 and section 3.3.4.3).

3.3.4.1 SARIMAX

As entailed, the key assumption of OLS regression are i.i.d. errors. However, it is a bold statement that while the variables are time-series that the residuals would not be autocorrelated. Autocorrelation in the residuals means that there are still patterns in the data that are unextracted. There are two common ways to deal with autocorrelation as mentioned at the end of section 3.3.3.1:

1. Use of heteroskedasticity and autocorrelation (HAC) robust standard errors
2. Regression with time series errors

The first option is to keep the same model specification and use heteroskedasticity and autocorrelation (HAC) robust standard errors instead of the conventional one. This is what is applied in the regression models part. For the second option, another model specification is needed and is considered the more sophisticated solution although more expertise in autoregressive models is needed. This is part of this section: Time series analysis.

This method tries to overcome the limitation of regression modelling that the data is assumed to be time-independent. Although HAC robust standard errors can be easily implemented and the original model can be kept, confidence intervals could increase, and predictions could become less accurate. Time series errors could lead to a better specified model and more precise predictions, but it is more cumbersome, and the original model cannot be retained. To have covered both common ways to deal with autocorrelation, a Seasonal AutoRegressive Integrated Moving Average with Exogenous variables (SARIMAX) model is implemented on the commuting data. This can be implemented in R with package *forecast* which can fit models with ARMA errors (Hyndman & Khandakar, 2008).

SARIMAX consists of multiple parts:

- *Seasonal*: capturing repeating patterns over fixed intervals
- *AR (AutoRegressive)*: helps understanding what the influence is of past values on a current value.
- *I (Integrated)*: One of the key conditions for this model, is having stationarity. This means that the mean and the variance are constant over time. This part of the model achieves that by differencing the time series whenever necessary.
- *MA (Moving Average)*: For the influence of the errors of past forecasts on the current values.
- *X (Exogenous variables)*: Accounts for predictors that are not part of the time series but may have an impact on the series. These are most of the variables that can be used as predictors in the regression models.

Outlier alteration is utilized instead of outlier deletion. This approach is needed, while deletion of outliers breaks the cycles that are present in the data. Data can be imputed e.g. with mean, median, interpolation or regression imputation. Mean imputation is simple but ignores temporal structure. Interpolation uses neighbouring values. A missing value on Friday (regarded as a day with less commuters) can be interpolated incorrectly due to Thursday and Monday values. It is therefore chosen to use regression imputation, for which the established regression models are used. The model incorporates the known relationships, and the underlying model is already explained. Seasonality is not fully captured in the model, but it will give a reasonable approximation that is not too complicated in comparison to some other options and is better fit than some simple common options.

3.3.4.2 Decomposition

Decomposition is a time series analysis technique that can help with isolating underlying trends in time series. By defining a regular repeating pattern, seasonality can be isolated. The other two components are the trend which depicts the long-term direction of the data and a random component. The random component is what is left from the actual data after capturing the trend and seasonal component. Each component is shortly explained:

1. *Trend:* This component filters out the fluctuations on short-term to keep the overarching patterns in the data which can be used to assess the long-term directions that span over multiple defined periods.
2. *Seasonality:* The pattern that is captured by this component can be associated to a fixed interval in time. For example, workweeks. The regular pattern can therefore be subtracted from the observed values
3. *Random component:* The noise and irregularity that is not captured in either seasonality or trend is reflected in the random component. It can be used to assess how well the underlying model captures all patterns and if patterns remain in the random component it can mean that the used model does not capture the underlying patterns as well as desired.

3.3.4.3 Interrupted Time Series

To assess the implemented polices, Interrupted Time Series (ITS) Analysis is a method that can measure changes in mean and trend after a policy intervention. This enables estimating the effects of policies implemented to stimulate a shift towards sustainable mobility. ITS consists of a level change after the intervention and a slope change after an intervention. This technique is more zoomed-in and does not use exogenous variables to explain differences in trend. Equation 13 gives the time series equation in mathematical terms.

$$y = \beta_0 + \beta_1 * T + \beta_2 * D + \beta_3 * P + \epsilon \tag{13}$$

In Equation 13, y is the outcome variable, $\beta's$ are the coefficients, T depicts the time that is passed from the start of the window of observation, D is a dummy variable that indicates a zero before policy intervention and a 1 after the intervention, and P is a variable that is equal to the passed time after an interventions has been implemented.

3.4 Interpretation

Interpretation of outcomes is the last step from Creswell (2009). The step of the Data Science Process associated with the interpretation of outcomes is *Visualizing and Reporting*. *Decision-making* is not the responsibility of this research. This research aims to support decision-makers. The research does not result in a *Data Product* to deploy in the *Real World*, but the research advises in further research and policy advice towards sustainable mobility. Regarding *Visualizing and Reporting*, which is the step of the Data Science Process that enables interpretation of outcomes, no new or specific methods are used for this purpose. Visualizations result from *Data Cleaning, Exploratory Data Analysis, and Modelling* and results are reported accordingly in Chapter 5, from which a discussion and conclusion follow in Chapter 6 and Chapter 7. In this research *Visualizing and Reporting* relies therefore on steps that follow a methodology but does not have a concrete methodology itself. While the different models have all coefficients that are interpreted in different ways, Appendix F provides guidelines on how to interpret the coefficients and what should be considered.

3.5 Limitations

The quality of the collected data, which is from various sources and also have different values for the same observations, can impact the reliability of the findings of this thesis. Imputation can help with missing values but will introduce bias while the actual values cannot be retrieved. Also, the first bias can come from the measing instruments itself, which can also wrongly detect commuter for example or cannot estimate the actual weather conditions due to the existing of microclimates and the absence of measing instruments on every location.

Regression models ignore the time dependency and does not model the lagged effects or the impact of a policy over time. The observations are assumed to be independent, while it is known that the observations are ordered in time. Also, it is assumed that predictors have a relationship that in linear and additive with the responses. This is a simplification and it could be that real-world relationships are nonlinear or have interactions with unobserved variables. For regression models, while there is no time-dependency, it is possible to remove observations (days) from the model to estimate coefficients. This can be considered for statutory holidays or days with technical errors causing not expected outliers. For Time Series Analysis, this will break the cycles which means outliers cannot be removed. Outliers can skew the models and distort predictions which can have an influence on decision-making by policymakers. There are various reasons for outliers including errors or anomalies due to holidays for example.

The focus of SARIMAX models is more on prediction and forecasting rather than identifying the effects of individual determinants. There need also to be enough datapoints before and after policies. The implemented polices are mostly towards the end of the dataset and there could also be some overlap between multiple policies. A disadvantage is that the SARIMAX model cannot handle categorical variables. The variables that are categorical need to be converted to either binaries or encoding. Interpretation is different from regression in SARIMAX. A coefficient can only be interpreted conditional on the value of previous values due to the presence of lagged values, which is not very intuitive, and this is the reason that in practice, ARIMA-based models work better for forecasting than for inference purposes.

The combination of the used methods can potentially lead to complementary insights to answer the research question. Despite the limitations, the research question can be answered under assumptions and with following the methodology presented in this chapter. In subsequent chapters, the answers given to the sub question can be used to meet the research objective of giving policy recommendations to decision-makers and to compare with the research objective of investigating wat existing literature reveals about the problem domain.

4 Data preparation

The purpose of this chapter is to depict the crucial parts of how the data is prepared for analysis. Before it was known which variables would be exactly included in the analyses, data was prepared but subsequently not used in the models (e.g., origin stations of commutes towards Eindhoven Central Station by train). For completeness and for potential further research and gaining insights, the descriptive tables, and manipulations not relevant for the modelling part are included in Appendix C. Files with format .csv are first opened with Microsoft Excel. This allowed for quick inspection and with the help of Pivot Tables, the necessary data is extracted from the files. For some files, the processes of data processing and data cleaning are intertwined. Additional changes on the data are performed in RStudio.

The data is collected predominantly for De Run 6000, which is the location of the ASML Campus in Veldhoven that has around 12.000 employees in office on regular working days and is the focus of this study. During this thesis, too little data was available for the neighbouring location called De Run 1000 (approximately 3.000 employees). This location was recently renovated and expanded. These are reasons that made it impossible to conduct proper research for this location at the time of writing.

4.1 Public transport

The dataset *NSGO_events_to_campus* includes registered check-outs at certain bus stops for ASML employees with a NS-Business Card. Adding the weekday in a separate column enabled for filtering on weekdays and filtering out Saturdays and Sundays. In the file at hand, each day has for each origin-destination (OD) pair for each building a separate row. The possible buildings are 6 and 71. Building 71 corresponds to De Run 1000, while Building 6 corresponds to De Run 6000. The bus stops associated with De Run 6000 are “Veldhoven, Asml-gebouw 4”, “Veldhoven, De Run 5300”, “Veldhoven, Heerseweg”, “Veldhoven, Locht”, “Veldhoven, Mmc Veldhoven Hoofd”, “Veldhoven, Mmc Veldhoven Kempe”, “Veldhoven, Runstraat”, and “ Veldhoven, Veenstraat”. The bus stop “Veldhoven, Mmc Veldhoven Hoofd” was firstly not included by ASML, on the grounds that people will be more likely to get off at “Veldhoven, Mmc Veldhoven Kempe”. Since not all bus lines pass both stations, this station is included in this case study. Therefore, all OD pairs to “Veldhoven, Mmc Veldhoven Hoofd” are assigned to building 6. The number of destinations included is eight and the number of origins is 160. The top 5 departure stations, and the percentage of the total trips are depicted in Table 6. These percentages are based on the weekdays over all days included in the original datafile.

Table 6: Most popular departure bus stops

Origin	Percentage (%)
Eindhoven, Station	59.53
Eindhoven, Witte Dame	7.16
Eindhoven, Keizersgracht	5.20
Eindhoven, Grote Berg	4.09
Eindhoven, Mecklenburgstraat	3.82
Total of top 5	79.80

As can be seen in the table, “Eindhoven, Station” is by far the most popular departure station. Note that the research is focused on the last mile, before the bus to De Run 6000 people can come from another bus, train, home, carpool parking place, etcetera. The 8 destination stops as well as the 5 most popular origins are kept in the dataset. Four dates are missing in the dataset: 05/18/2023, 05/29/2023, 12/25/2023, and 01/09/2024.

For data validation: the grand total of origin trips matches the grand total of destination trips, but a huge increase in commuters is happening starting from April 2nd, 2024. The average over observations is 33.93 before this date and 568.51 after this date. This difference was also notable on the Spotfire Dashboard (A&M Modal Split Trend Dashboard). This dashboard provided IT-Managed Analytics on a high-level, no breakdown for stations for example. This is the reason the data was directly requested to the team responsible for the dashboard. Since the data matches, this seems to be due to how they receive and pull the data to their database. There is also another dashboard to get more in-depth details made in PowerBI. Also established by IT, but by a different department. This dashboard was discontinued in September 2024 but made available for this research. Since this was created as a side project and due to discontinuity has no data from the recent months, this data is used for validation purposes. It seems the input for this dashboard is different, while it does not have the jump starting from April 2nd, 2024, but was already measuring higher (more logical) numbers of trips before this date.

In PowerBI, not all the same bus stations are present, "Veldhoven, Heerseweg", "Veldhoven, Locht", and "Veldhoven, Runstraat" are missing. With just 45 trips over the whole period, "Veldhoven, De Run 6700" is a stop in PowerBI that geographically matches with De Run 6000. Probably a placeholder or a substitute for when there was maintenance at one of the stops. The data from April 2nd, 2024, to September 30th, 2024, has on average 2% more counts in the original dataset compared to PowerBI. Before April 2nd, 2024, the PowerBI count is on average 15 times as large as the count in the original dataset. It is chosen to multiply the PowerBI numbers with 1.02 to get new data for January 2nd, 2023, up until April 1st, 2024. Prior to April 2024, the number of trips is not distributed over the stations, therefore it should be noted that further statements about individuals bus stops are based on data from April 2024 onwards.

4.2 Car

Car data is only retrieved with the Spotfire dashboard. The data is multiple times exported, and it was not possible to look further back than 2023. The dashboard shows a maximum of two years. All data that was possible to update, is updated on January 14th, 2025. Over all dates, the difference was at most one except from the most recent months. Therefore, there are no concerns about the first two weeks of 2023 that were not included anymore on the export from January 14th, 2025. For cars, the data is aggregated, and no breakdown has been made per parking. An important remark is the definition of how car data is collected. In the retrieved data, a car is only registered as the license plate could be coupled to an employee.

4.3 Employees

Data from the first of 2023 up until the last day of 2024 is retrieved from the Spotfire Dashboard for the total employees assigned to De Run 6000 and the number of employees checked in on De Run 6000 for each day. Note: on a day there can also be check-ins from people that are stationed at another campus. Also, the other way around is possible. The office percentage therefore might differ slightly from reality.

4.4 Cycling

4.4.1 Daily_bike_events(in)

For this dataset, weekday was added to make it possible to filter out the weekends as for most of the datasets. With a Pivot Table, the possibility arose to have on each row a day which led to the columns being the bicycle counting cameras. All weekdays are present in the dataset from March 1st, 2023, to December 31st, 2024. In the dataset, one of the cameras (and therefore a column name) is called 'No camera. This is a placeholder while the cameras where not operational at the beginning. Also, there are some columns with slightly different names. These cameras are essentially the same or replace a camera in the same location. For example, one of the camera's ends with Entrance Bike Lan, which is supposed to be Entrance Bike lanes or something along those lines. See Table 7 for a detailed overview.

Table 7: Bicycle cameras

Location	Camera name	Period
5L	05.L.00.401 bicycle entrance	03/27/2023 – 12/31/2024
5L_basement	05.L.00.924 - Entrance Bike Lan	01/01/2024 – 08/02/2024
	05.L.00.924 - Entrance Bike Lanes inside	11/01/2024 – 12/31/2024
	05.L.00.924 - Entrance Bike Lanes inside - Old	08/02/2024 – 09/11/2024
	05.L.00.EXT - 19 Gate Bicycle	09/11/2024 – 10/31/2024
Maingate	Camera 1	05/12/2023 - 08/31/2023
	MainGate Bicycle Counting	09/01/2023 - 12/29/2023
	NL.VDH.Maingate_Bicycle_Countin	01/01/2024 – 08/02/2024 09/11/2024 – 10/31/2024
	NL.VDH.Maingate_Bicycle_Counting	08/02/2024 – 09/11/2024 11/01/2024 – 12/31/2024
P3	P03A oost gate bike entrance	03/27/2023 – 12/31/2024
P4	P4.PMMC.00.EXT - PMMC bike entrance	08/02/2024 – 09/11/2024 11/01/2024 – 12/31/2024
	P4.PMMC.00.EXT - PMMC.bike entr	11/01/2023 – 08/02/2024 09/11/2024 – 10/31/2024
	PMMC.00.EXT - PMMC.bike entranc	03/27/2023 – 10/31/2023

The *Daily_bike_events(in)* is validated with PowerBI data and .csv files obtained per camera per month. Until the 1st of April 2024 the three sources are identical when the placeholder is removed. For April 2024, PowerBI double counted the maingate observations. For July 2024, the PowerBI data has a lot more counts. It was found out, that PowerBI double counted the cameras of P3 and P4. The .csv and original set are the same in these periods, therefore these numbers are used. Maingate, 5L_basement and P4 have on 08/02/2024 and 09/11/2024 two measurements for these locations. There seems some fiddling with the names in the data, while these cameras are essentially the same.

Therefore, on the days itself both measurements are added, and the columns are merged. This causes a mismatch with PowerBI and .csv. Starting from the 12th of September until the 27th of September the three sources have exactly the same counts. The final day of September is the last registered in the .csv files and PowerBI. It records less than the original data, which is potentially corrected afterwards. Therefore, from this day onwards the original data is used. With the received datasets October and November could not be checked. This is a limitation while starting from April 2024, the data between different sources mismatched often while before that date, the data matched from the start of implementing the cameras. It might be worthwhile for people with access and permissions to use .csv files of each camera from these months to compare with the original data set and find out why differences occur between the ways of retrieving data. At Building 5L, a bicycle parking facility has opened in December 2023. The camera was running starting from January 2024.

4.4.2 Imputing the bicycle data

It is noticeable that cameras are sometimes turned off. This causes missing data for each camera. The chosen method to do something about this is imputation. To be more specific Multivariate Imputation by Chained Equations (MICE) as is explained in Chapter 3. For this, all zeroes and NaN (Not a Number) values in the bicycle dataset are removed, except for the zeroes in the 5L_basement column before the camera was operational (2023). This was necessary, so the zeroes in 2023 are not imputed with other values.

The imputation of values is performed in RStudio using packages *sofImpute* and *mice*. The data of the five cameras is converted to a matrix. On this matrix, both horizontal and vertical data is used for imputation, which is important while individual cameras observe a certain number and in general more travel volumes leads to more people to campus which leads to higher counts by all cameras. Some packages in R can have problems with dealing with numbers at the start of variable names or with underscores. Therefore, '5L_basement' is changed into 'basement5L' and '5L' is changed into 'old5L' In the matrix of cycling data, 10% (232 out of 2310) of the observations are missing. In Figure 7, an overview of the patterns of missing data. Blue stands for a complete observation, red for a missing observation. On the left the number of occurrences is displayed and, on the right, how many variables had a missing value. Under the columns the total of missing observations for a camera. Camera old5L has the most missing observations (99), while there are 293 days with no missing data for all cameras.

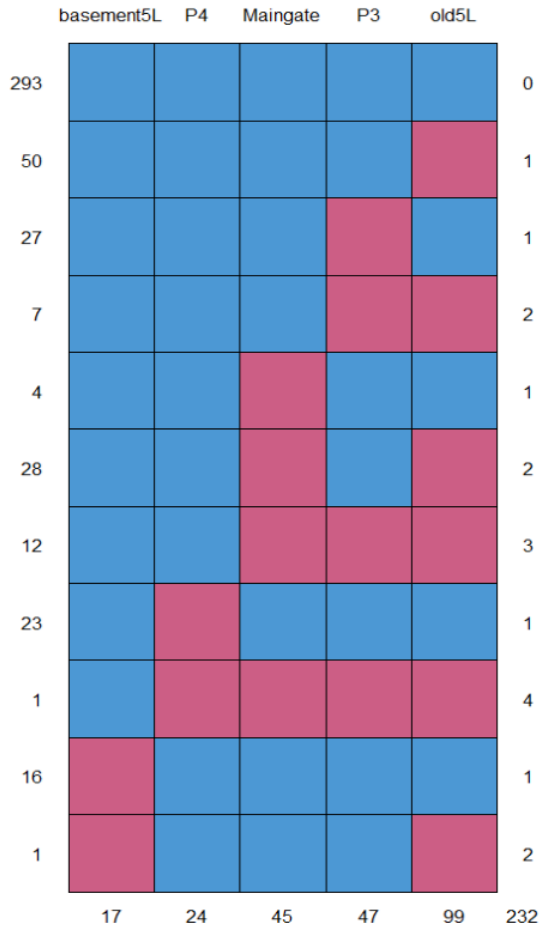


Figure 7: Missing data pattern of bicycle data

According to Van Buuren (2018), between 5 and 20 imputations for a moderate amount of missing data is sufficient when the primary interest is on the point estimates of coefficients. Therefore 10 imputations are chosen with method 'Predictive Mean Modelling', which is the default. For each imputation, 100 iterations are performed. After the final iteration, for each imputation, the cameras can be summed to have an estimate of the total number of cyclists on each day. All ten imputations can then be used for the regression model after which the estimates can be pooled, which is the second half of Figure 5 from Chapter 3. It is not recommended to simply average the imputed values, while the between-imputation variability is ignored and the uncertainty related to the underlying data cannot be fairly represented by averaging the data (Van Buuren, 2018), therefore the ten imputations are stored as separate columns in an excel file with name *camerasums_imputations.xlsx*.

Figure 8 shows healthy convergence of the MICE algorithm, while there is not a trend visible aside from a straight line and the lines for each imputation are mingling well for all camera's.

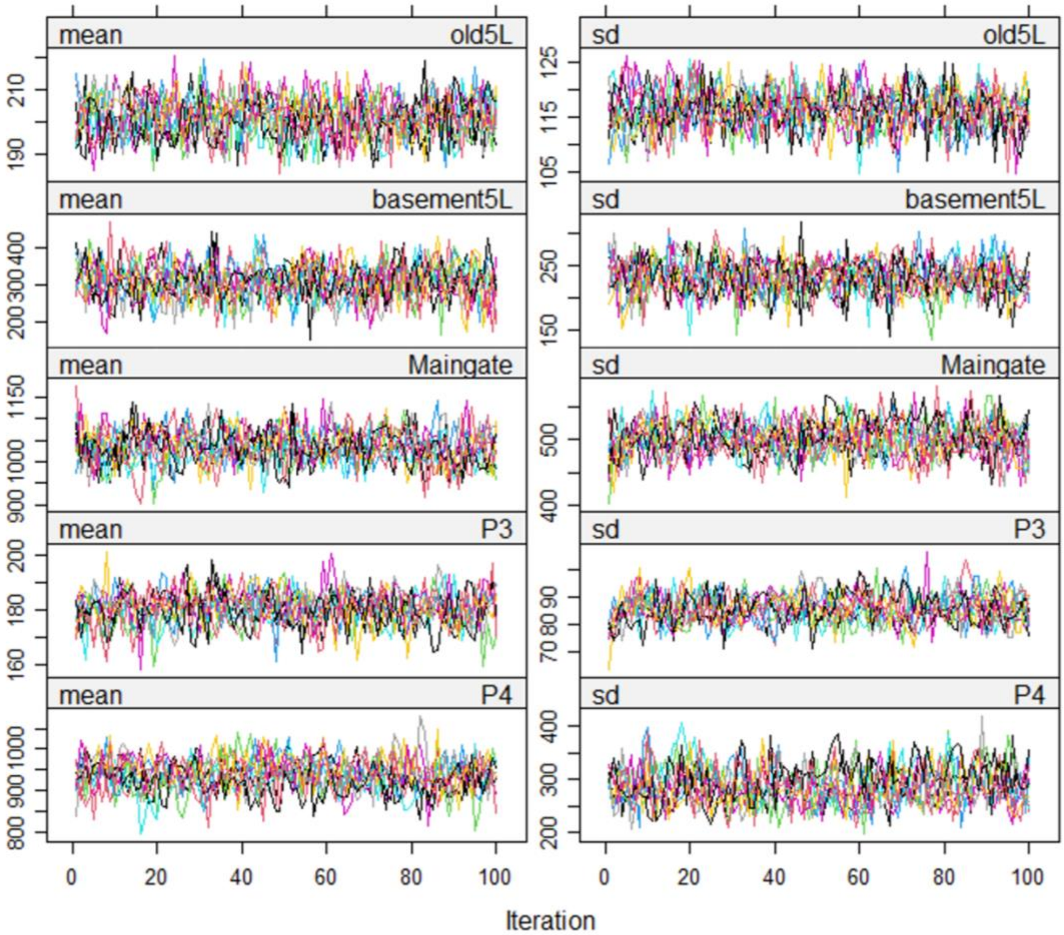


Figure 8: Convergence plots for MICE algorithm

For the final dataset, if the columns of *camerasums_imputations.xlsx* are included this leads to 16 columns regarding the number of cyclists on a day: a column for each camera, one column with the sum of cameras without imputing and 10 columns with imputing (for each imputation one).

For the regression analysis, all imputed datasets need to be modelled individually, afterwards the results can be pooled which accounts for the uncertainty introduced during the imputation process. However, in Exploratory Data Analysis the post-imputation is averaged. As Van Buuren (2018) emphasized, using this average for fitting the models is bad practice.

4.5 Interventions

The interventions that are used in the models, are interventions that are considered as large enough to show differences on an aggregated level. The first interventions is the increased bus frequency in October 2023. A huge parking place for bicycles with approximately 2000 parking places was opened in January 2024 in the basement of building 5L. While also a new camera was placed, this is used as one of the interventions. E-bikes that can be used by employees freely switched operator in August 2024 to better match the preferences of employees. There was also a period without e-bikes due to the switch between operators. The improved service was expected to increase the number of bicycle commuters. There were no e-bikes from 13th of July to 1st of August after which the shared e-bike company Drop took over from shared e-bike company Hely. Lastly, the bicycle allowance up to 20km (one-way) increased from 21 cents to 35 cents (0.23 net 0.12 gross) on October 1st, 2024. To compare, commute allowance by car is 21 eurocents for employees living at least 10.1km from their work location with a maximum of 20km compensate for, for a single trip. An intervention that is not caused by ASML, but each year occurs is Daylight Saving Time. The last Sunday of March the clock is set one hour forward, which is called Daylight Saving Time and is approximately 7 months. The Daylight Saving Time ends the last Sunday of October, when the clock is set an hour backward. This is called standard time and this time is used for 5 months in the year. This is further discussed in section 4.7.

4.6 Other variables

Statutory holidays are days that employees are entitled to as paid days off. In 2023: January 1, April 7, April 10, April 27, May 18, May 29, December 25, December 26. Dutch Liberation Day and Good Friday were not statutory holidays for ASML employees. In 2024 these days are January 1, April 1, May 9, May 20, December 25, December 26. Dutch Liberation Day and Good Friday were not statutory holidays for ASML employees. School holidays are considered while these weeks children in the Netherlands do not have school which is often linked to parents taking time-off to take care of the children. The school holidays in the Netherlands are Christmas break, Spring break, May break, Summer break, and Autumn break. In the dataset, this is a categorical variable with the weeks used that children in the southern region of the Netherlands have breaks. It can also be modelled as a binary variable with either the day being in a holiday or not.

4.7 Weather variables

Although it is not directly evident from the data received for this research (aggregated to full days), it is evident that most people arrive between 7AM and 10AM in the morning while the morning rush hours are from 7AM to 9AM. The data for this study regarding employees arriving on campus is at a daily level, nevertheless it could be worthful to link weather in specific hours to the total arriving employees due to weather being an important determinant for travel behaviour. From OpenMeteo therefore, data for each hour specific are retrieved and data on a daily basis regarding weather. This enables making various models to test which hour is most influential in relationship to the employees in office or see if the weather of the whole day is sufficient for this. There is also for both styles (hourly and daily) in addition to historical weather, also data of the predicted weather. In the literature there is no consensus which type of weather data to choose, while commuters can change their choice according to the prevailing weather at departure, at the destination or based on the forecast (Liu et al., 2017). The choice was made to use the coordinates of the ASML headquarters, as this is the destination for all employees. It is debatable whether people check the weather for Veldhoven, departure location or possible exit location of the train from where the bike can be taken. Also, different applications are used to check the weather and when. Historical data is what is actually measured and is preferred for that. Interesting to see how much models differ when forecasted data is used.

In the Netherlands, there is Daylight Saving Time. This means that during spring, Summer, and a part of autumn the clock is set one hour forwards to have darkness later during the day. This is for 7 months of the 12. The dataset uses UTC (Coordinated Universal Time), which does not have Daylight Saving Time, and offsets it by one hour to have CET (Central European Time). For the periods in 2023 and 2024 that have Daylight Saving Time, in the hourly data it is therefore necessary to do an alteration so that it has 2 hours difference with UTC. Especially for this research this is important because office hours in general not change. So, office hours are still between 8 a.m. and 5 p.m., regardless of summertime or wintertime. This makes it fairest for the research to shift the weather data also between the transitions to summertime and wintertime to align with the hours. In the daily weather data, there is a sunrise time. If this data is plotted, an approximate perfect line arises. The following figure, Figure 9 shows the time of sunrise before and after alteration.

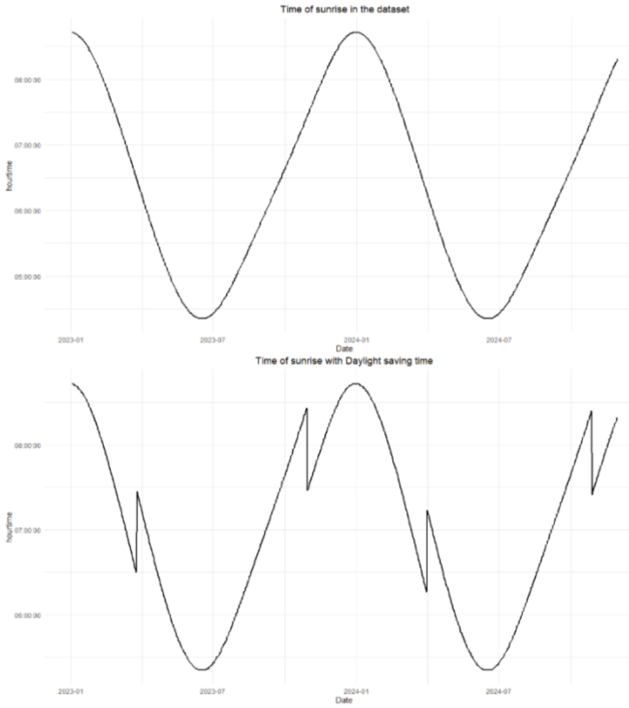


Figure 9: Alteration of Daylight Saving Time

It takes 25 days when Daylight Saving Time starts to have the same sunrise time. It takes 37 days when Daylight saving time ends to have the same sunrise time. Note that next to the Daylight Saving Time differences, the Northern Hemisphere leans more towards the sun causing more daylight hours in the summer in contrast to the winter.

Precipitation includes rain, showers, and snowfall. For the conversion from snow (cm) to the water equivalent precipitation (mm), Open-Meteo divides the snowfall by 7 (7 cm snow = 10 mm precipitation). In the data, there were just 27 days with snowfall and 22 days with snow on the ground at the end of the morning rush hours. Open-Meteo states that for snow depth, caution is needed while due to resolution of the weather models the snow depth tends to be overestimated. There were only two days with pure snow and no rain. Snowfall and rain have a very weak negative correlation of -0.02 on a daily level. While snow is in most of the literature understudied while there is a lack of observations, next to rain as a predictor also snow is included.

4.8 Descriptives

There are more than 100 variables, of which a lot of the variables are redundant. In the following tables a selection of the most important variables is depicted. This includes the response variables, policies, holiday information, and the most prominent weather variables. It can be seen in Table 8, that less than half of the employees in average is in office. Car commuters register the highest number across modes. While the morning hours are used for weather related variables, it can be seen the maximum temperature is not exceedingly high. The maximum value for precipitation is 50 times higher than average. There is always wind in and a quarter of the rush hours is dark on average with a minimum of 17 minutes of light and more than a quarter of the observations having no darkness in the dataset. Temperature and wind speed are better to center for the regression models, otherwise the interpretation of the intercept is harder. Without centering, for each variable the reference or zero value is used. As there is never zero wind and the temperature is more than 10 degrees Celsius on average, after centering the intercept can be interpreted as the scenario when all variables have their reference, with zero darkness and average temperature and wind speed.

Table 8: Descriptives of numeric variables

Variable	Min	1st Quarter	Median	Mean	3rd Quarter	Max
Employees	17699	21136	21658	21360	22055	22341
Employees in office	168	8601	10976	10066	11678	12771
Car commuters	19	3497	4146	3938	4694	5533
Bicycle commuters	0	1644	2350	2257	2944	3997
Bus commuters	7	444.8	578	551	674.5	976
Temperature	-6.4	6.2	10.4	9.978	14.6	21.8
Precipitation	0	0	0	0.1436	0	7.2
Darkness	0	0	4.5	28.62	57	103
Wind speed	0.8	8.5	12.3	13.34	17.3	37

In Table 9, binary variables are depicted, with a percentage of the observations that is coded as true in the dataset.

Table 9: Descriptives of binary variables

Variable	Count (0)	Count (1)	Percentage of Count (1) (%)
Statutory holiday	489	11	2.2%
School holiday (binary)	395	105	21%
Policy: Bus frequency	210	290	58%
Policy: Bike shed 5L	260	240	48%
Policy: Bicycle allowance	456	44	8.8%

For categorical variables, in Table 10 the descriptives are shown. The weekdays are perfectly balanced with for each weekday exactly 100 observations in the dataset. The seasons are not exactly balanced. This is due to the starting moment and end date of observations in the data set causing less winter and autumn observations. Spring and Summer are not the same while a year has 365 (or 366 days) which is not a multiple of 4.

Table 10: Descriptives of categorical variables

Level	Count	%
Weekday		
Monday	100	20%
Tuesday	100	20%
Wednesday	100	20%
Thursday	100	20%
Friday	100	20%
School holiday		
Autumn	10	2%
Christmas	15	3%
May	10	2%
Spring	10	3%
Summer	60	12%
None	395	79%
Season		
Winter	119	23.8%
Spring	134	26.8%
Summer	132	26.4%
Autumn	115	23%
Policy: Shared e-bikes		
Hely	400	80%
Drop	87	17.4%
No	13	2.6%

5 Results

This chapter presents the results of the performed analyses. As described in the methodology in Chapter 3, first an EDA is performed, which is supported with visualisations. Secondly, the main results of the regression analysis are depicted with visualisations, checks of assumptions and overview tables. The chapter ends with a section containing the main results of the Time Series Analysis. Appendix E shows raw outputs of the models, while Appendix F checks assumptions and Variance Inflation.

5.1 Exploratory Data Analysis

5.1.1 Weather conditions

In the literature it was stated that weather conditions are often correlated. Figure 10 is a correlogram of temperature, apparent temperature, precipitation, and wind speed during morning rush hours and on a daily level to check whether this holds for the case. For visibility, the correlations are shown as circles. In Appendix D, separate correlograms are added to have readable numbers for correlations.

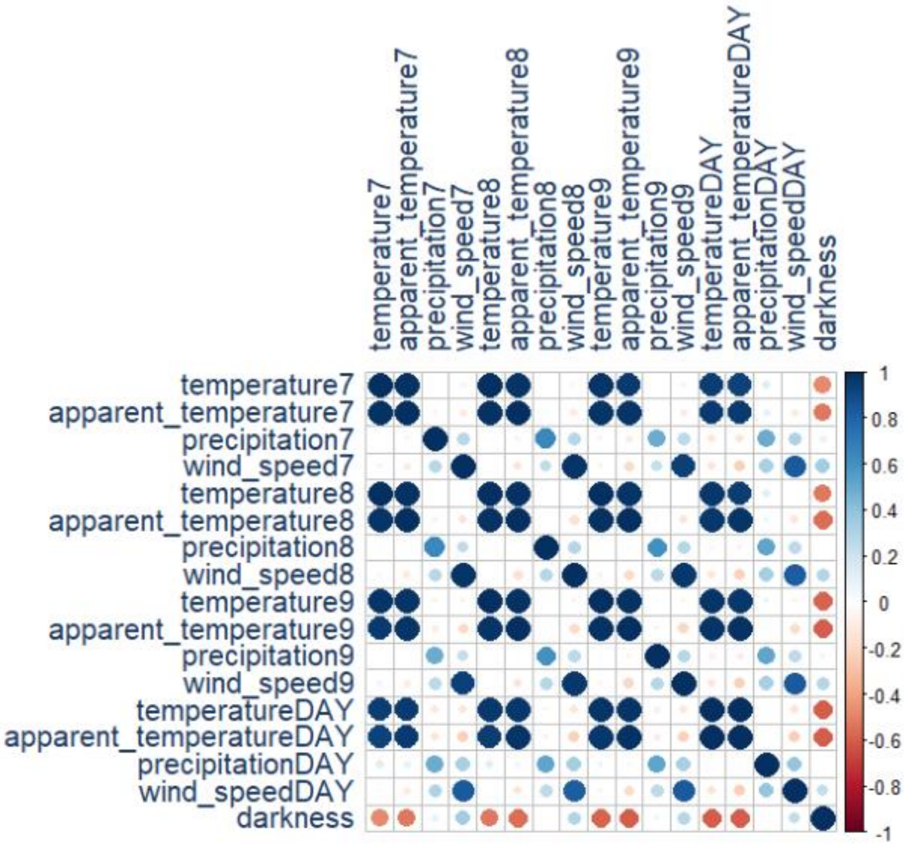


Figure 10: Correlogram weather and time of day

Temperature is only correlated in time, but not with precipitation and wind in the data. The lowest correlation between wind speeds is 0.83 between wind on a daily level and wind in the morning rush hours. Precipitation is less correlated in time; the highest correlation is 0.65 between measurements at 7 a.m. and 8 a.m. for precipitation . The lowest correlation is between daily precipitation and 7 a.m., which is 0.5. Increasing wind speed has a weak negative correlation with apparent temperature between -0.11 and -0.20 in the same hour. Wind and precipitation are also weakly correlated: within and between hours the correlation is between 0.24 and 0.28. In Appendix D, also correlations are shown between other weather related variables as humidity, cloud coverage, and surface pressure. These conditions show mostly weak to no correlations.

The choice was made to model with the weather variables of 8 a.m., as except for precipitation the values are instantaneous measured at the exact time. For precipitation, the precipitation of the preceding hour is summed. By taking 8 a.m. and 9 a.m. together, this is the precipitation between 7 a.m. and 9 a.m., with other instantaneous variables measured exactly in between. This falls in the middle of rush hour, and at day level the somewhat low correlation potentially loses accuracy. Temperature is preferred above apparent temperature, while apparent temperature is a collection of components and therefore does not purely measure the temperature. When making graphs of temperature over time and the minutes of darkness between 7 a.m. and 9 a.m., it is noticeable that the two graphs approximately have an inverse pattern of each other. The switch to standard time removes more minutes of darkness than is added when the switch happens to Daylight Saving Time. This can be seen in Figure 11.

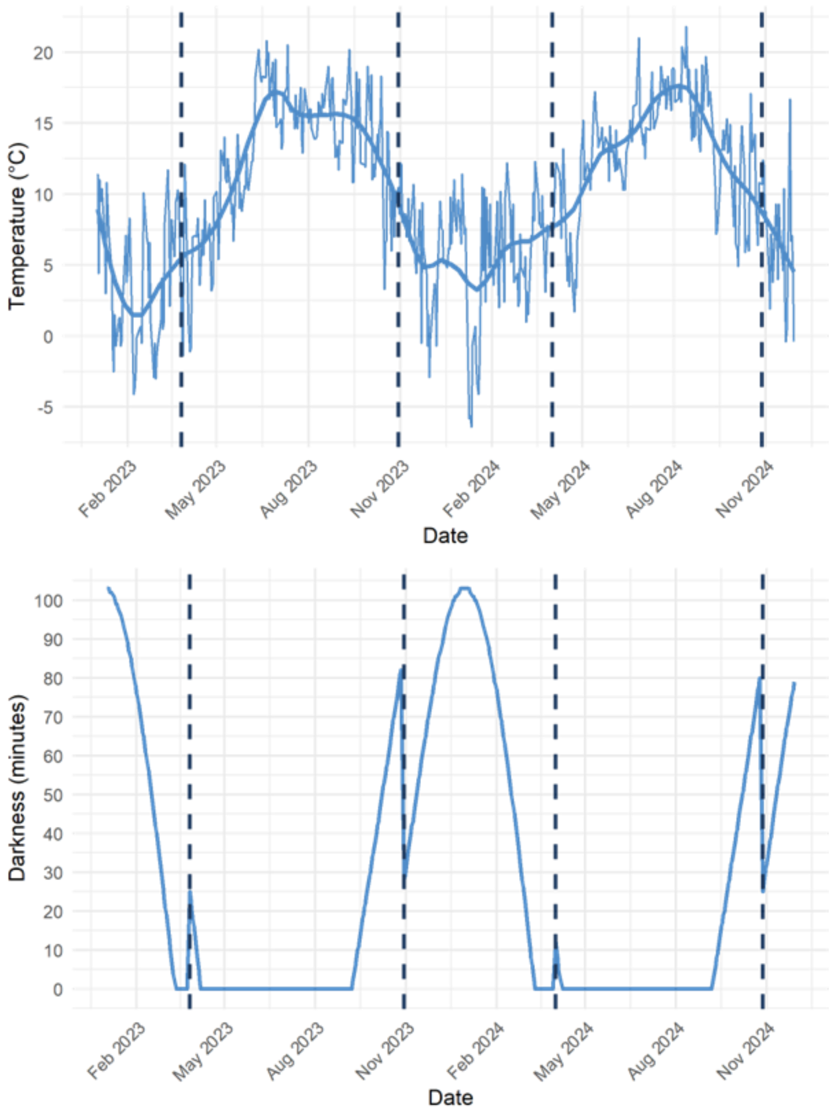


Figure 11: Line graphs of temperature and minutes of darkness at Veldhoven

Regarding precipitation, there is precipitation on 34.4% of the days when measuring in the rush hours between 7 a.m. and 9 a.m. excluding weekends. This is 128 days with precipitation and 372 without in the dataset. There are two days that stand out regarding rain in the dataset: April 24th, 2023, and June 23rd 2024 as can be seen in Figure 12.

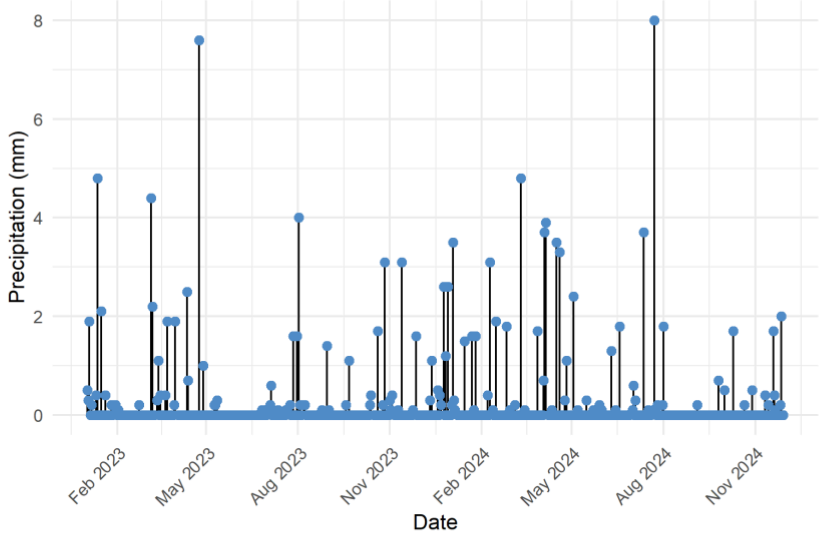


Figure 12: Lollipop graph of precipitation in Veldhoven

The wind at the headquarters of ASML is most of the time wind going from Northeast towards southwest, as can be seen in the windrose in Figure 13. For people commuting from the city Eindhoven the wind is at their back on most working days during the morning rush hour.

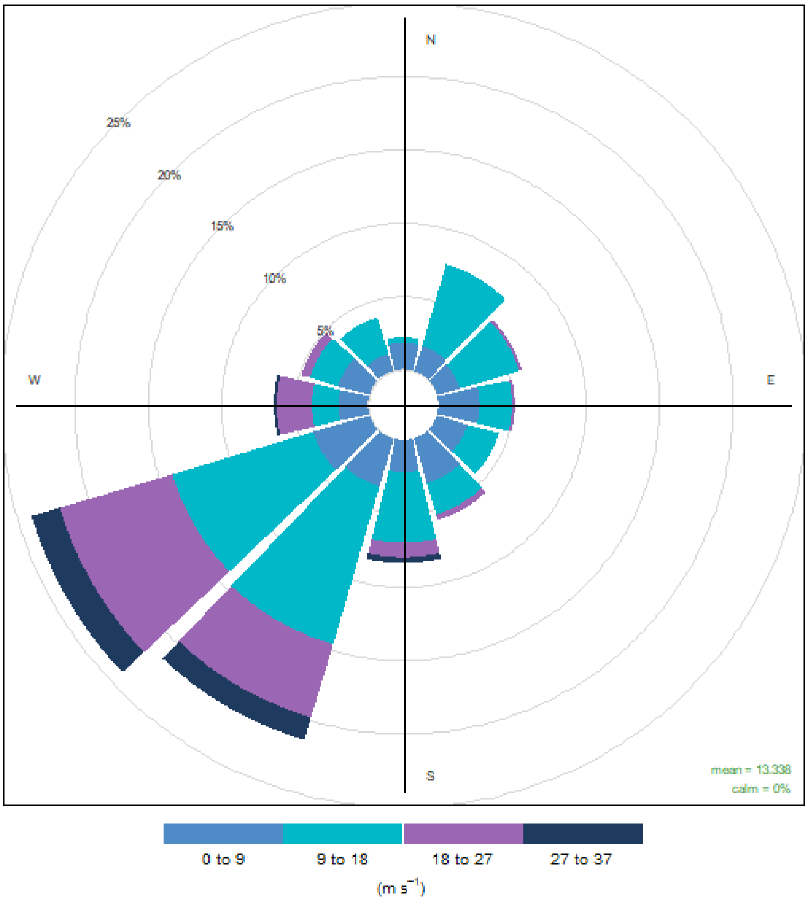


Figure 13: Windrose with frequencies of wind and direction

5.1.2 Travel behaviour

The main focus is on commuters by car, by bicycle, and by bus. There is also focus however, on which share these commuting options have in the total employees in office, and how many commuters are not captured by these three options. The first insights come from comparing commuting options with the day of week, as shown in Figure 14.

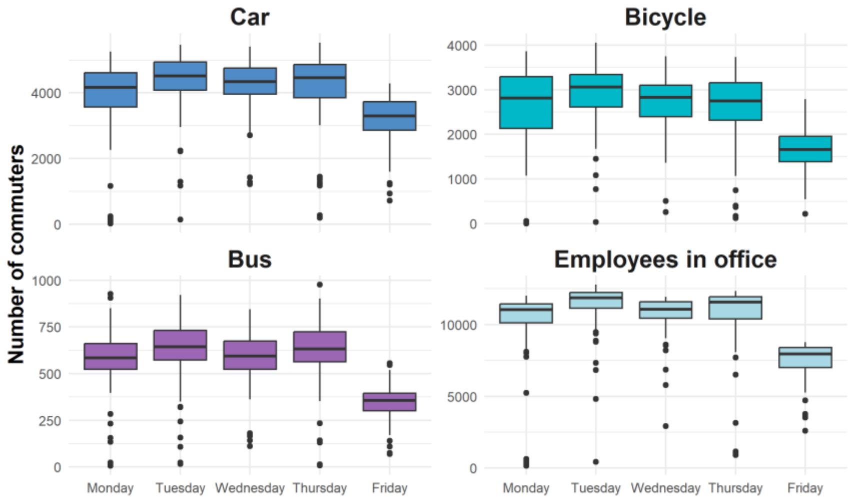


Figure 14: Boxplots between modes and weekdays

Figure 14 shows that for office presence and all commuting options the Friday is quite different from the other days. It is recognizable that there are in all graphs, outlier observations around zero. Most of these observations are statutory holidays. In cycling, more variability is visible within days while Tuesdays and Thursdays tend to have slightly higher averages in comparison with Mondays and Wednesdays. The same type of graphical representation with boxplots is made for seasons, which is depicted in Figure 15.

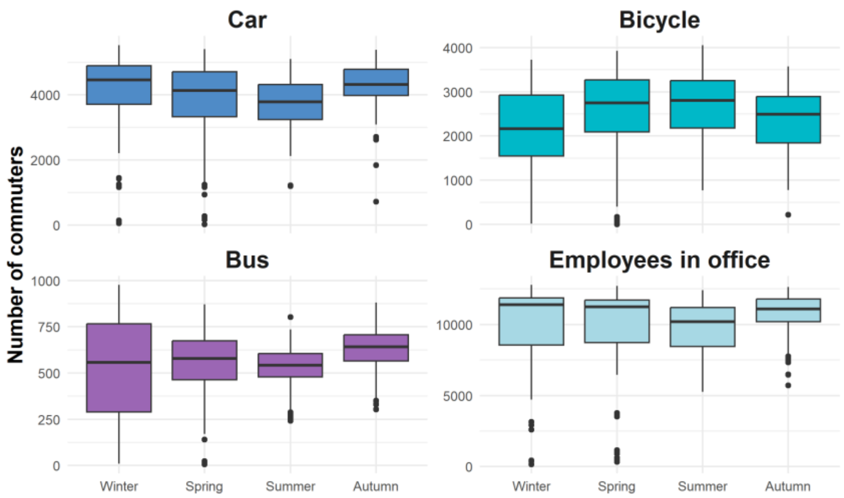


Figure 15: Boxplots between modes and seasons

The first look on Figure 15 shows a 'dip' in car usage in the summer and spring and more cyclists in these months. Also, the low outliers are mostly in winter and spring, which are the seasons with statutory holidays. When not looking in comparison with weekdays of season, it is interesting to see what the distribution is of the number of commuters in histograms. In the histograms of Figure 16, daily observations are assigned to one of the hundred bins, and it can be seen how frequent observations in the range of a bin happen.

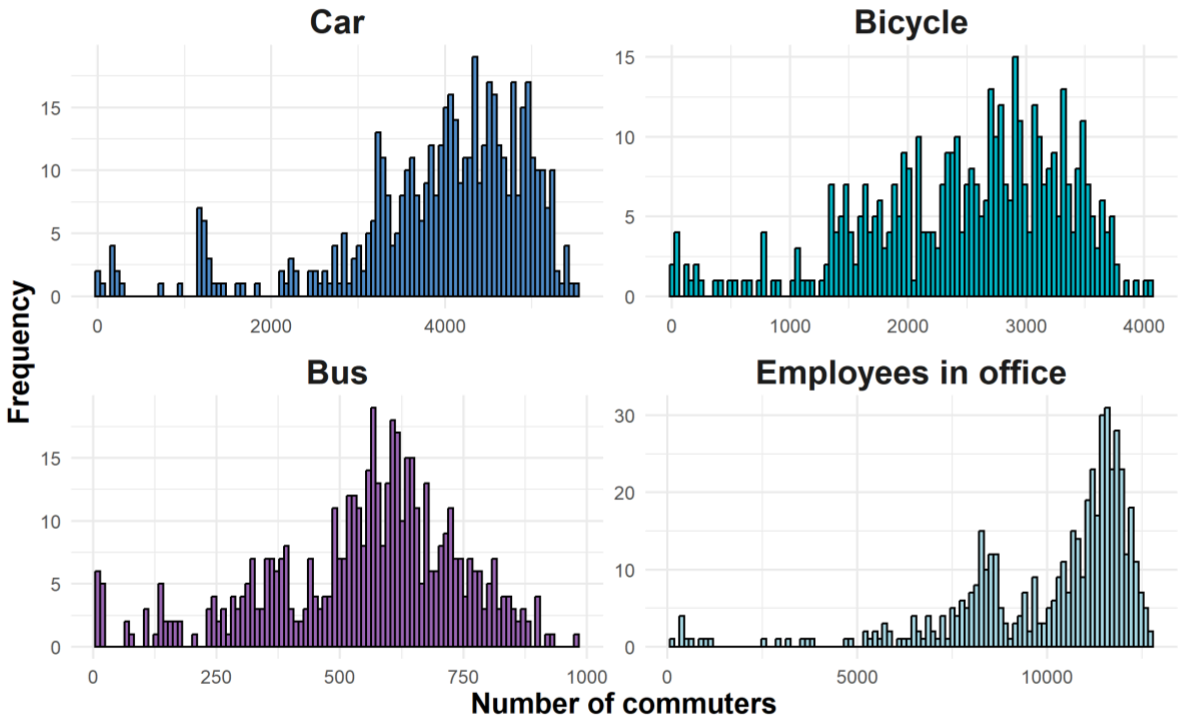


Figure 16: Histograms of modes

As can be seen, distributions are quite different between the modes. The histogram of car commuters has a more left-skewed distribution, while the public transport histogram does not seem to have a skew. The histogram of bike commuters has a bimodal distribution with two local maxima, while employees in office follows more a bimodal left-skewed distribution. Employees in office has clearly noticeable a smaller peak and a large peak. This is related to Figure 14, depicting less employees on Friday and more narrow boxes. The slight peaks on the left side are mostly associated with statutory holidays. Little peaks in between the left and the higher peaks can also potentially be outliers. The outcomes are not normally distributed, but as entailed in the methodology transforming is not necessary for outcome variables (a common misunderstanding that the outcome variable needs to be normally distributed instead of the errors for linear regression). Transformations also make it harder to interpret the analysis, which fits less with the objectives of these models. Figure 17 includes two graphs, one with daily observations over time for each mode and one with the average modal split for each quarter.

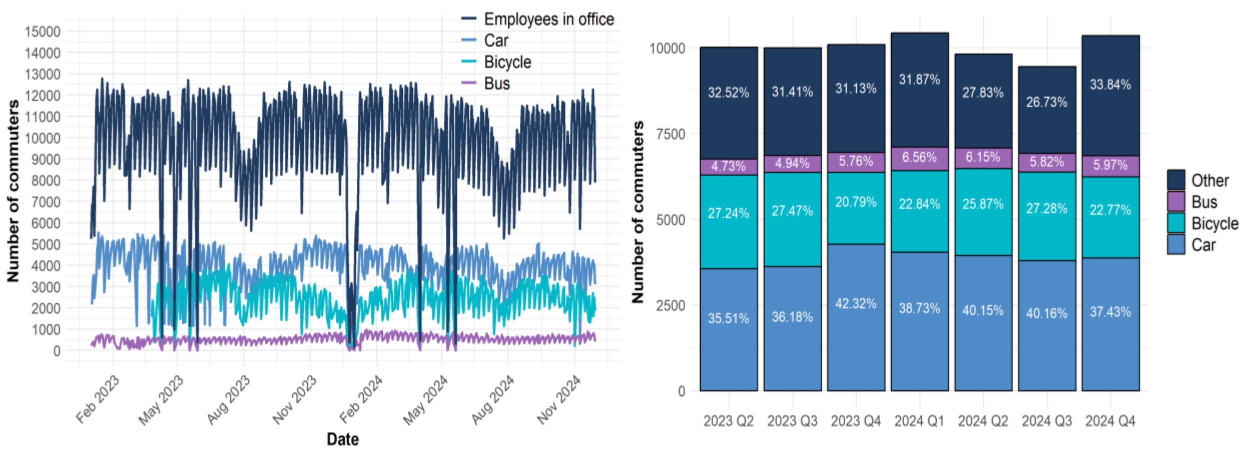


Figure 17: Line graph and stacked bar plot of modes

Note, that in Figure 17, the bicycle line starts later, while camera's were operational starting from the last week of March in 2023. This is also the reason that 2023 Q1 is not present in the modal split graph.

Employees in office and the commute mode, are thoroughly shown. But, behind the commuters in office are the total commuters registered at Run 6000. This is depicted in Figure 18, which shows the office presence over the months at Run 6000.

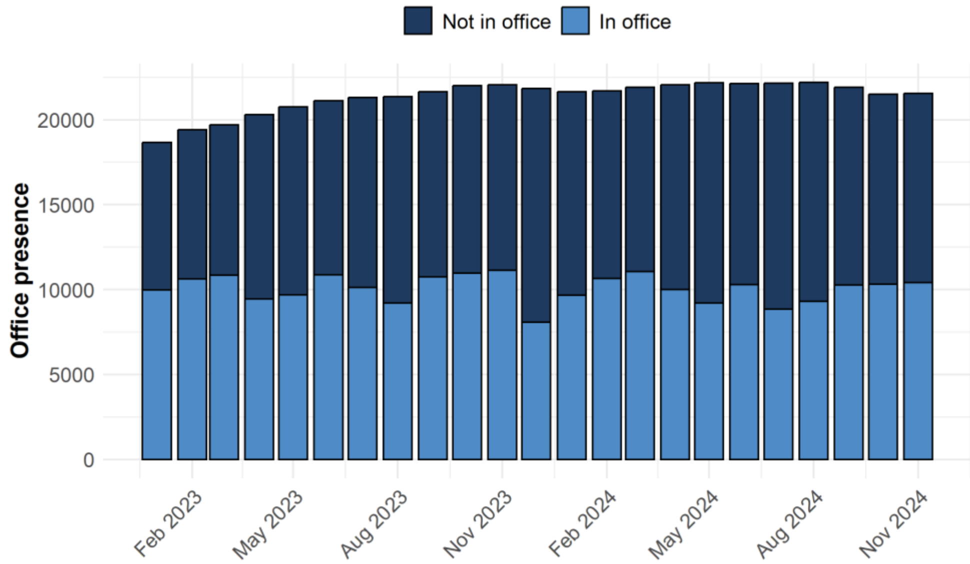


Figure 18: Run 6000 office presence

From the figure it seems the number of employees stagnated since 2024. The number of people in office did not increase with the number of employees. It rather stalled and even decreased in regard to the number of employees.

The scatterplot matrix in Figure 19, shows correlations between variables analysed in this Exploratory Data Analysis.

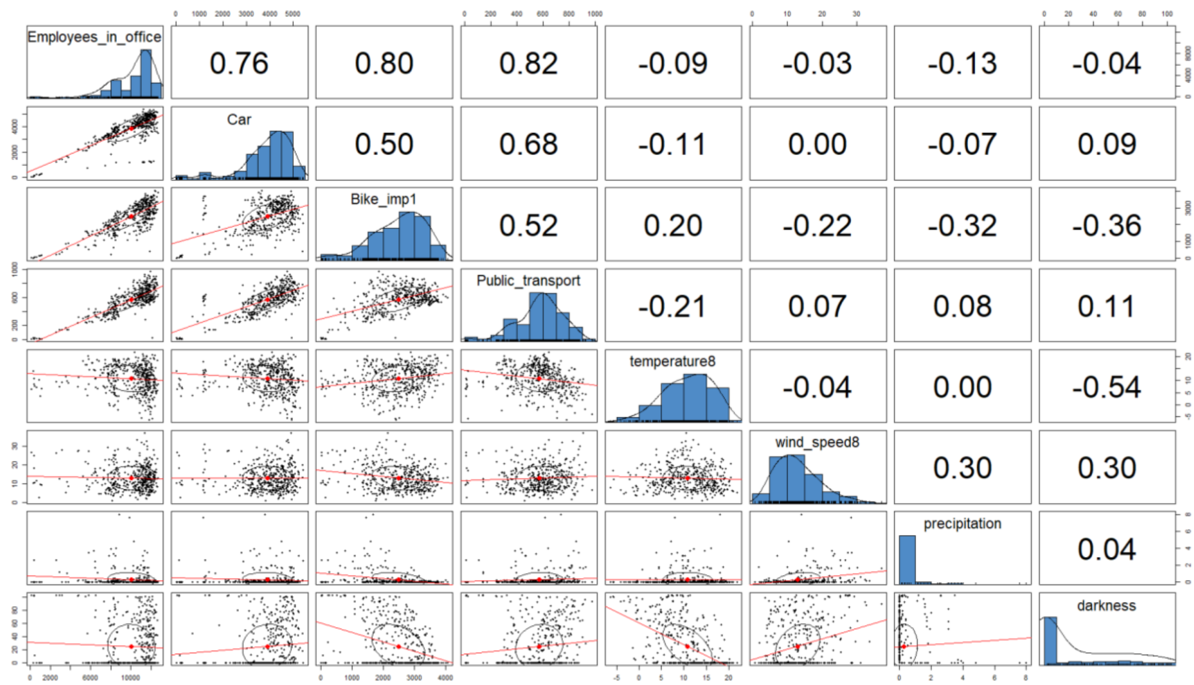


Figure 19: Scatterplot matrix key variables (mode volumes)

For commuting, there is moderate correlation between modes. While all are positive correlated with the employees in office, substitution effects between modes are not recognizable from this analysis. Commuting by bike has the strongest correlations with weather variables. Precipitation has too many observations at zero or slightly above. Therefore, it is better to encode precipitation as binary variable, which means no rain or rain. Darkness shows a clearer relationship with other variables. Both the binary option as the continuous option might work in regression. Important to notice, is that outliers and influential observations are not filtered out yet. As in the methodology explained this should only be done after fitting initial models. It is highly likely that dealing with influential observations will make relationships stronger between variables or give a clearer direction.

The same scatterplot matrix can be made, but with percentages for the commuting options. In Figure 20, the resulting scatterplot matrix is depicted.

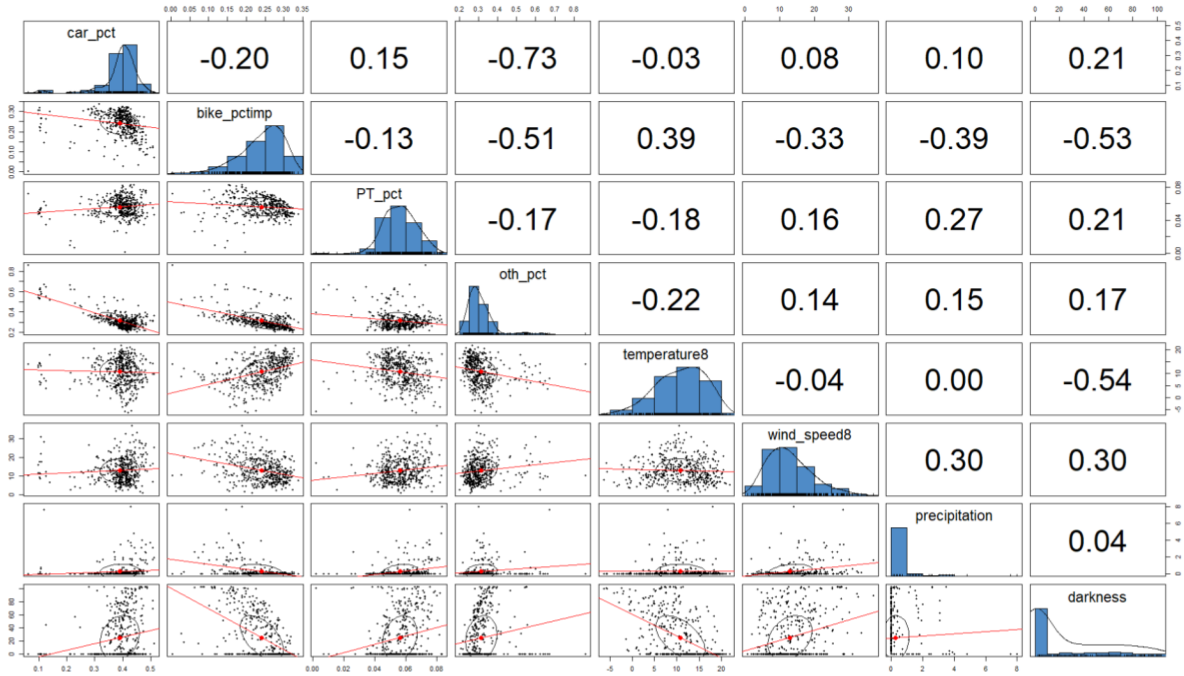


Figure 20: Scatterplot matrix key variables (mode shares)

From Figure 20, it is possible to see substitution effects. If the share of bicycle commuters increases, it is noticeable that car, bus (PT), and ‘other’ (oth) decrease. An increase in the share of car commuters also leads to an increase in bus commuters. For the mode shares, it is visible that weather conditions have a higher correlation with cycling shares, car shares, and bus shares. This is probably due that there is the constraint of the shares to be added up to 100%. It is more evident from this figure, that influential observations need to be handled, as more clear relationships are visible in the scatters but the red line not matching these scatter relationships.

5.2 Regression analysis

This section covers the results for the regression analysis. First, travel volume models are covered. After these models regarding quantities of commuters, the modal split model shows how the different modes of transport relate to each other in changing circumstances regarding weather and seasonality.

5.2.1 Travel volume models

First an overview is given for the travel volume models that briefly shows the process and gives the results of the models in tables. After this overview, for each modality further results are depicted and the results depicted in the overview tables exemplified.

5.2.1.1 Overview

To contribute to answering sub question 2, a basic model is established excluding interactions and policies. To contribute to answering sub question 3, the basic model is extended with interactions and policies.

There are outliers detected and deleted using Cook's distance. In Figure 21, for each response the threshold is plotted with the Cook's distances for the observations. For bike, 34 observations are detected as outliers. For car, 26 observations. For employees in office, 29 influential observations, and 37 for bus commuters. These influential observations are removed, together with statutory holidays.

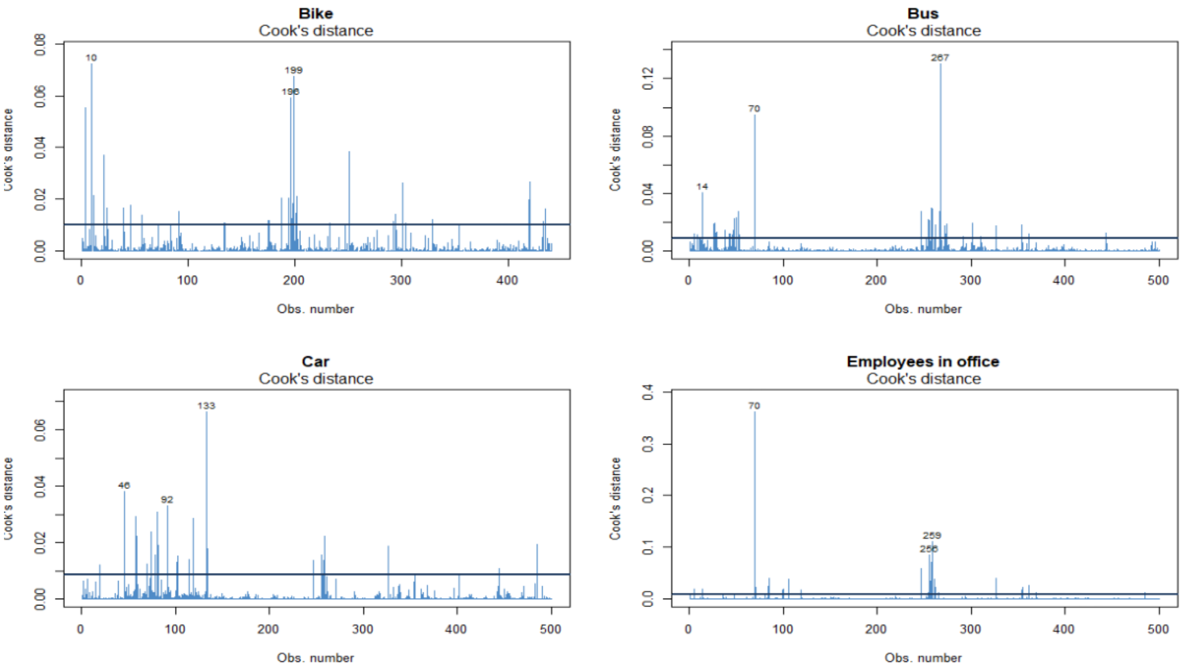


Figure 21: Detection of influential observations

Model fit statistics for the models are collected before and after removing influential observations and statutory holidays. These are documented in Table 11. The removing of influential observations leads with less observations to better explanation of variance and residual standard errors are lower, which means that the divergence from predicted values is lower for all modalities.

Table 11: Model fit statistics before and after dealing with influential observations

	Before dealing with outliers			After dealing with outliers		
	Obs.	Adjusted R2	Residual SE	Obs.	Adjusted R2	Residual SE
Bike						
Basic model	440	0.7195	450.03	405	0.7616	357.6
Advanced model	440	0.7434	430.41	405	0.7898	355.7
Bus						
Basic model	500	0.5736	125.6	457	0.7178	84.28
Advanced model	500	0.742	97.72	457	0.8745	56.2
Car						
Basic model	500	0.5273	744	467	0.6422	454.6
Advanced model	500	0.5629	715.4	467	0.7196	402.4
Employees In office						
Basic model	500	0.8624	871.7	471	0.9285	462.9
Advanced model	500	0.8713	843.1	471	0.9465	400.2

The coefficients of the different models for the basic model (after deletion) are summarized in Table 12, with the standard errors within brackets and the significance on different levels for the p-value depicted with asterisks. These coefficients have HAC standard errors. While the standard errors are pooled for the imputed bike data, it was in the software not possible to estimate HAC standard errors. In Table 12, therefore the pooled standard errors are used.

Table 12: Overview of regression models

	Dependent variable			
	Car	Bike	Bus	Employees in office
	(1)	(2)	(3)	(4)
(Intercept)	4450.84*** (160.32)	3527.54 ***(90.73)	640.67***(69.05)	11473.12 ***(96.06)
Temperature	-10.65 (10.36)	22.88*** (5.24)	-0.19 (2.07)	10.72 (9.21)
Wind speed	8.51 (4.65)	-5.29 (3.43)	0.66 (0.88)	3.59 (5.56)
Darkness	1.42 (2.15)	-9.47*** (1.22)	0.56 (0.87)	0.85 (1.43)
Precipitation (True)	52.26 (57.59)	-480*** (48.33)	48.01*** (9.94)	-245 ***(60.28)
Snow (True)	-292.4 (162.36)	-353.24** (131.02)	12.59 (43.01)	-305.79* (129.89)
Weekday				
Tuesday	293.85*** (33)	116.35* (57.51)	68.65*** (6.71)	832.68*** (36.43)
Wednesday	124.6** (42.12)	-87.4 (59.05)	10.34 (6.88)	142.46*** (40.46)
Thursday	280.62 ***(47.58)	-11.75 (59.51)	45.5 ***(8.23)	592.15 ***(47.75)
Friday	-919.85 ***(46.41)	-1092.72*** (57.72)	-255.7*** (8.56)	-2970.99*** (48.51)
School holiday				
Autumn break	-501.26* (246.33)	-250.94* (121.63)	-68.01*** (18.8)	-1578.84*** (208.97)
Christmas break	-2081.73*** (206.94)	-828.11*** (210.28)	-385.32*** (53.12)	-5215.01*** (314.98)
May break	-480.83* (205.11)	-352.94* (156.12)	-58.29 (31.56)	-1470.3 ***(126.24)
Spring break	-720.8 ***(178.66)	-520.45 **(189.92)	-143.54 (101.6)	-2138.45*** (136.05)
Summer break	-492.83** (183)	-767.7 ***(64.21)	-35.49* (14.62)	-1871.55*** (280.2)
Season				
Spring	-103.64 (187.88)	-435.72*** (91.53)	-44.29 (69.48)	-307*(133.08)
Summer	-329.66 (214.34)	-262.27** (97.69)	-85.93 (63.68)	-627.7*** (173.82)
Autumn	-173.65 (154.81)	-140.25* (70.18)	-34.73 (36.16)	-299.76** (111.95)
Observations	467	405	457	471
R2	0.6553	0.7716	0.7283	0.9311
Adjusted R2	0.6422	0.7616	0.7178	0.9285
Residual Std. Error	454.6 (df =449)	357.6 (df=387)	84.28 (df = 439)	462.9 (df = 453)
F Statistic	50.21 (df = 17 ; 449)	61.8 (df = 17 ; 387)	69.23 (df = 17 ; 439)	359.9 (df = 17 ; 453)
Note:				*p<0.05 ; **p<0.01; ***p <0.001

The basic model is extended with interactions between seasons and daily weather conditions and the most important policy interventions are added. The stepwise approach leaded to the removing of several interaction terms for the other models. The results of this process are depicted in Table 13. In the subsequent subsections, for each modality the results of the model are further depicted.

Table 13: Overview of regression models including interactions and policies

	Dependent variable			
	Car	Bike	Bus	Employees in office
	(1)	(2)	(3)	(4)
(Intercept)	4634.91*** (166.18)	3613.05 ***(118.19)	616.89*** (31.14)	11861.68*** (126.01)
Temperature	-26.59* (10.44)	17.38 ***(5.16)	11.6*** (2.79)	36.51*** (8.02)
Wind speed	6.97* (3.4)	-5.99 (3.28)	-2.64 (1.48)	-2.25 (4.94)
Darkness	-1.66 (1.69)	-7.89 ***(1.48)	1.15* (0.47)	0.94 (1.77)
Precipitation (True)	81.64 (50.53)	-457.66*** (46.66)	-0.9 (14.13)	-487.82*** (82.55)
Snow (True)	-380.73** (115.83)	-294.78 *(124.45)		-269.58* (108.76)
Weekday				
Tuesday	293.91*** (32.28)	112.64* (54.12)	63*** (5.32)	821.06*** (35.3)
Wednesday	124.77** (42.02)	-98.06 (55.86)	5.02 (5.79)	127.1** (39.63)
Thursday	276.36*** (45.8)	-27.22 (56.18)	39.95*** (7.56)	576.25*** (45.38)
Friday	-930.79 ***(48.6)	-1101.32 ***(54.46)	-260 ***(8.33)	-3003.51*** (46.04)
School holiday				
Autumn break	-365.55* (146.14)	-354.21** (126.39)	-40.9* (18.93)	-1589.43*** (70.53)
Christmas break	-1907.06*** (188.3)	-857.13*** (204.8)	-388.89*** (18.9)	-5187.01*** (238.47)
May break	-453.59 *(179.56)	-394.49 *(151.96)	-55.99 ***(15.57)	-1459.7*** (80.3)
Spring break	-715.16*** (118.63)	-502.48** (178.33)	-159.48 ***(26.06)	-2186.44*** (107.91)
Summer break	-449.74 (231.31)	-698.31 ***(67.94)	-31.45* (15.01)	-1794.94*** (239.74)
Policy				
Bus freq	252* (266.29)	-317.43** (103.16)	68.78*** (13.76)	-95.8 (150.38)
Bus freq, Shed	-218.96*** (96.72)	-204.7*** (58.71)	140.62*** (14.9)	-276.16 **(90.49)
Bus freq, Shed, No e-bike	-505.35 (150.51)	-384.74*** (108.24)	83.41*** (17.78)	-982.73***(272.01)
Bus freq , Shed, Drop	-67.76 (149.2)	-258.2*** (66.4)	70.79***(10.23)	-593.73***(121.85)
Bus freq, Shed, Drop , 35 cents	-396.44* (184)	-218.24 *(90.17)	48.72***(11.7)	-528.55***(63.58)
Season				
Spring	-233.6 (166.65)	-368.78 ***(103.81)	-84.53** (30.12)	-579.94*** (134.96)
Summer	-665.68 (380.55)	-228.05 (122.29)	-80.7* (34.38)	-861.73*** (171.95)
Autumn	-349.81 (253.14)	33.53 (158.27)	-68.71 (38.13)	-429.83*** (114.88)
Interactions				
Temperature x Spring	48.04* (19.98)		-13.6*** (2.85)	-24.95 (17.18)
Temperature x Summer	52.02 (32.41)		-14.18 ***(3.54)	-9 (19.53)
Temperature x Autumn	11.24 (9.89)		-15.11*** (2.89)	-32.59*** (8.68)
Wind speed x Spring			5.34 **(1.69)	
Wind speed x Summer			4.97** (1.7)	
Wind speed x Autumn			3.53 (2.02)	
Darkness x Spring	44.92*** (8.98)	-33.94 ***(9.45)	-2.56*** (0.72)	
Darkness x Summer	4.31 (14.12)	7.76 (6.55)	-1.68* (0.67)	
Darkness x Autumn	3.73 (2.99)	-2.26 (2.38)	-0.5 (0.63)	
Precipitation x Spring			55.49** (20.44)	400.42*** (112.74)
Precipitation x Summer			30.83 (18.73)	285.83* (127.54)
Precipitation x Autumn			57.81*** (17.43)	305.29 **(107.84)
Observations	467	405	457	471
R2	0.7365	0.8028	0.8836	0.9497
Adjusted R2	0.7196	0.7898	0.8745	0.9465
Residual Std. Error	402.4 (df =438)	335.7 (df=379)	56.2 (df = 423)	400.2 (df = 442)
F Statistic	43.72 (df = 28 ; 438)	61.8 (df = 25 ; 379)	93.31 (df = 33 ; 423)	298.2 (df = 28 ; 442)
Note:				*p<0.05 ; **p<0.01; ***p <0.001

5.2.1.2 Bicycle

The model's explained variance increased slightly by adding interactions and policies. There is one significant interaction between darkness and spring, which has a negative direction. This means that according to the model, darkness in spring decreases more bicycle commuters compared with winter if all other conditions are the same. All periods of implemented policies seem to have less bicycle commuters than baseline, but if policies are regarded in time, no shared e-bikes available has a more negative effect than the policy before, and the two subsequent policies have a relative positive effect regarding previous policies and therefore moments in time (the policies are ordered from first to last implemented).

Weather related variables have the same direction and are as significant as in the basic model. Higher temperature increases bicycle commuters. Darkness, precipitation, and snow decrease the number of commuters. For darkness, the coefficient needs to be multiplied with the minutes of darkness. With the mean of 29 minutes this equals approximately 229 commuters less on average and on the day with the most darkness (103 minutes) approximately 813 commuters, keeping all other variables constant. The effects of precipitation and snow are binary meaning the coefficient depicts the difference with days with no snow and rain. Precipitation therefore has an association with a significant decrease in commuters. If all conditions remain the same, the temperature needs to increase around 26 degrees Celsius to compensate for precipitation. Wind speed is the only weather condition that is not significant for cycling.

From all variables, Fridays have the greatest effect on the number of cyclists with a decrease of 1101.32 commuters on average compared to Mondays. Tuesday is the only day with a significant increase in commuters compared to Mondays. School holidays all decrease the number of bicycle commuters. Compared to the basic model, only spring remains significantly different from winter. Part of the effects of seasons are therefore captured by interactions and policies.

To check the model assumptions, Figure 22 can be consulted for a visual check. All assumptions are better met, than before removing outliers. The figure from before removing outliers can be found in Appendix F. Slightly deviations in the Normality of Residuals plot raises the suggestion of potential autocorrelation. Two values that are seen on the most left of the Linearity plot and the Homogeneity of Variance plot, are both Fridays during Christmas Holidays.

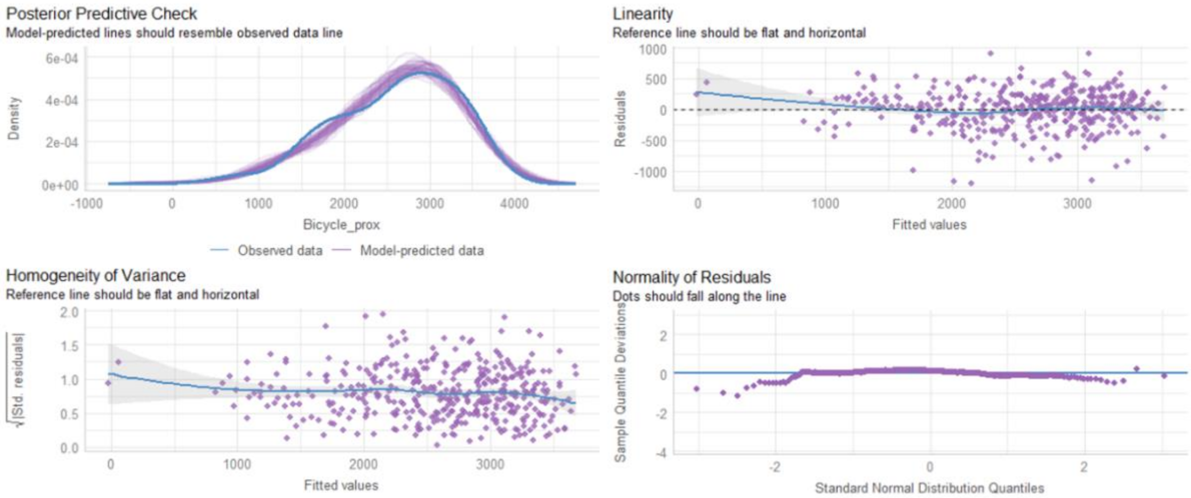


Figure 22: Visual inspection of model assumptions for bicycle

In Figure 23, the actual values and predicted values are shown for both the basic model and the advanced model. Both models capture the patterns, despite the limitations of regression, quite well. It can be noticed that the advanced model, as the residual standard error and adjusted r-squared suggested, matches the actual observations slightly better. This can be noticed for example in the first 50 observations.

Figure 23: Basic (upper) versus advanced (lower) fitted values for bicycle commuting

5.2.1.3 Public transport

Adding interactions and adding the categorical variable for policies but omitting the variable regarding snow (due to the stepwise approach) takes 16 additional degrees of freedom. Still, the explained variance is significantly increased between the basic and advance model: Adjusted R-squared from 0.7178 to 0.8745. The residual standard decreases the most of the four models relatively with a decrease of 33.3%. All interactions are significant and also all policies are significant. For policies, compared to baseline after all policies on average more bus commuters were measured, with in the period after the increased bus frequency and the bicycle shed the highest difference. According to the interactions, higher temperatures in the winter increase bus commuters more than in other seasons. Spring and summer have a slightly positive interaction with wind speed but a slightly negative interaction with darkness in comparison with winter. Precipitation in autumn and spring significantly increases bus commuters.

Coefficients for temperature and darkness became significant, which means that this relationship was hidden and the suggestion arises that temperature and darkness only are related to commuting behaviour under certain seasonal or policy-related conditions. The direct effect of both increase bus commuters with roughly the effect of ten minutes of darkness equal to one degree difference in temperature. For precipitation, a significant coefficient in the basic model indicates that precipitation in general influences commuting by bus. However, the advanced model brings nuance to this and depicts it only has a significant effect in autumn and spring while the direct effect is not significant anymore. In autumn and spring on average more than fifty commuters more take the bus in comparison with winter when it rains.

Especially Fridays decrease the number of bus commuters significantly with Tuesdays and Thursdays diverging from Mondays in a positive direction. Christmas break decreases the number of bus commuters with more than half the intercept. Also, spring break has a substantial decrease in commuters, while the other school holidays also decrease the number of commuters but in a less substantive way. The coefficients for seasons suggest that winter is the season with the most bus commuters.

In Figure 24, it can be seen that observations that have a lower fitted value in the linearity graph cause a slight deviation from the reference line. This is also visible in the Posterior Predictive Check that has a bumpy left side that is not exactly captured by the model-predicted lines. The Normality of Residuals also indicate the presence of autocorrelation.

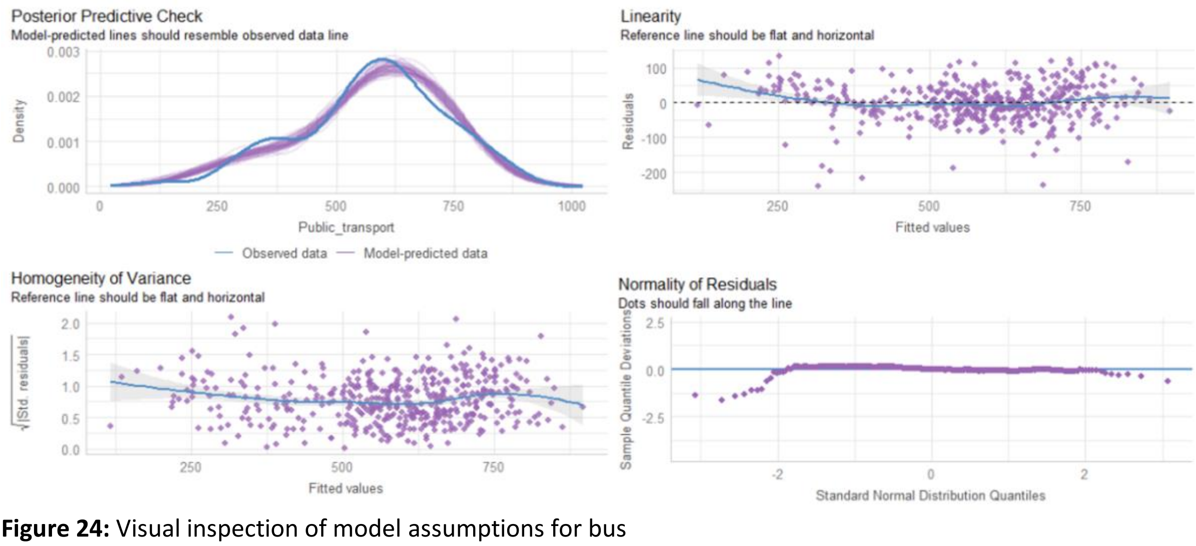


Figure 24: Visual inspection of model assumptions for bus

Figure 25 shows the predicted values by the model versus the actual values of the dataset without influential points for both models. It is noticeable that the basic plot overestimates before the bus frequency was increased and underestimates afterwards. The advanced model captures these trends better and matches significantly better with actual values. The model captures the trend of bus commuters quite well.

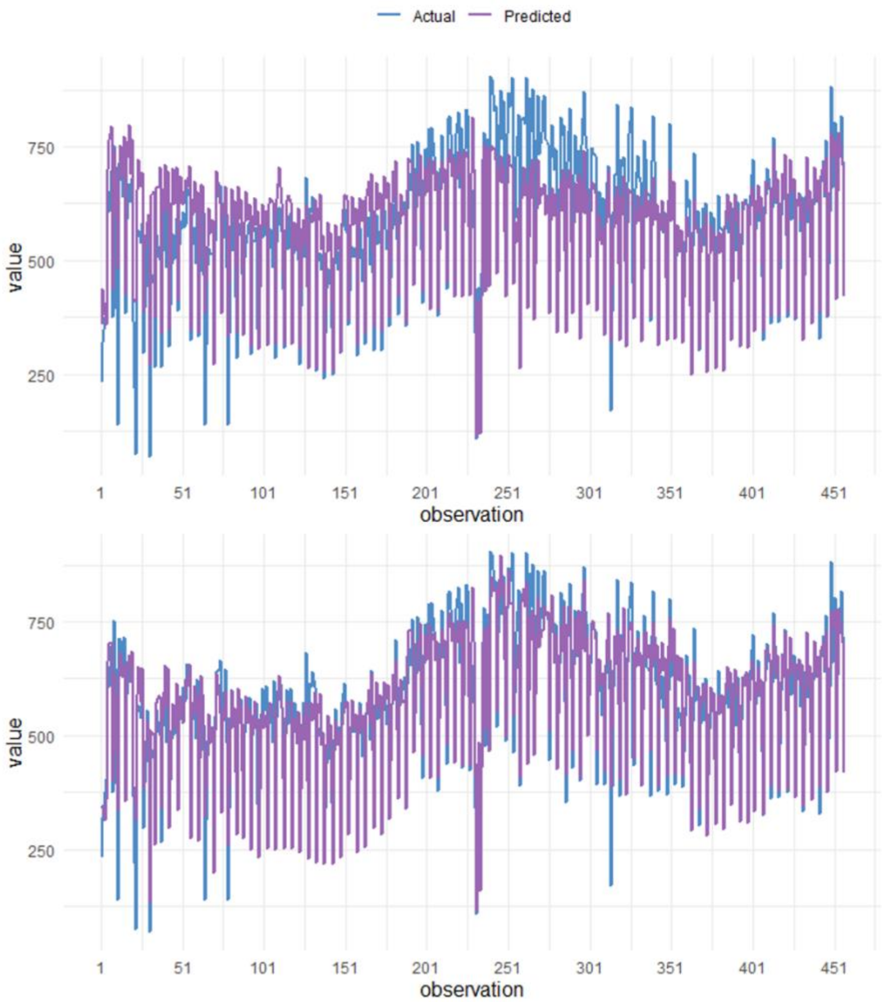


Figure 25: Basic (upper) versus advanced (lower) fitted values for bus commuting

5.2.1.4 Car

The preliminary models concerning car commuters have lower explanatory power in comparison with bus, bike, and employees in office. Dealing with influential observations did increase the explained variance, but still it is the model that has the lowest adjusted R-squared (although with 0.7196 the variance is still more than moderately explained) and in the advanced model the highest residuals. Contrasting to the basic model, the advanced model has significant coefficients for temperature, snow, and wind speed. Also, the interactions of Spring with darkness and temperature are significant. Darkness in spring leads to more car commuters than in other seasons, but relatively an increase of 44.92 car commuters is not substantial. The period in which only the bus frequency was increased, increased car commuters significantly. While the bike shed and more travel allowance for biking have a negative significant coefficient compared with the baseline before interventions were implemented.

The significant coefficient for snow, but insignificant coefficient for precipitation indicates that with rain there is no significant increase of car commuters while snow impacts the number of car commuters substantially. More wind and lower temperatures increase the commuters by car. Car is the only model in which a significant direct effect of wind speed is identified.

Tuesdays up until Thursdays increase the number of commuters by car while Fridays significantly reduce the number of car commuters. No significance of the direct effect for seasons was found, indicating that car commuters are invariant to the season they use the car.

In Figure 26, it can be seen that the fitted values follow a linear trend, with a small upward deviation for low fitted values and a small downward deviation for high fitted values. The same effect but a bit more extreme is noticeable in the Homogeneity of Variance. The Posterior predictive check resembles the observed data line. Normality of Residuals indicate autocorrelation between residuals.

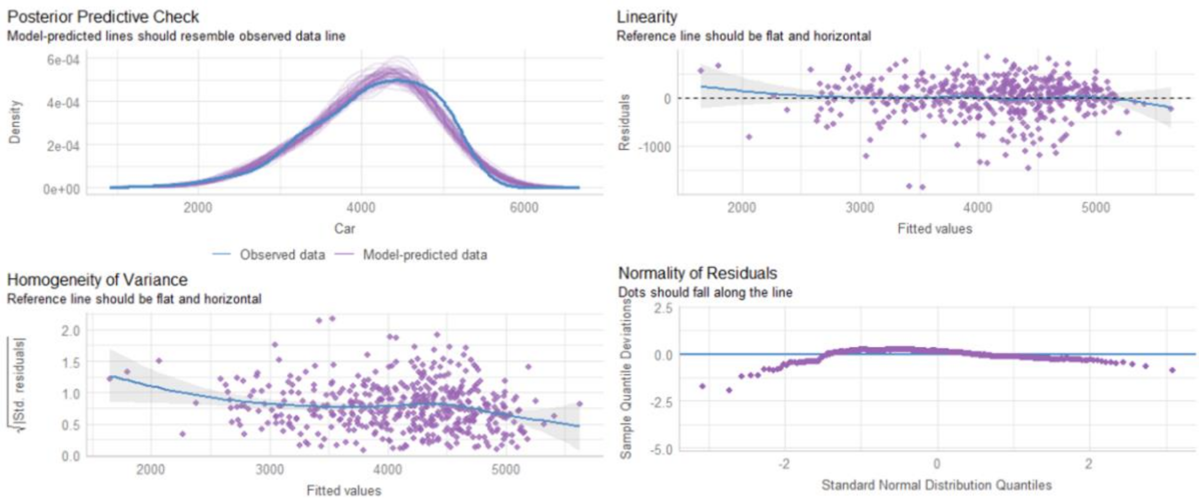


Figure 26: Visual inspection of model assumptions for car

When the model predicts the actual values, it can be seen in Figure 27 that underestimation occurs in 2023 in the basic model. This underestimation is significantly lower in the advanced model. In the basic model even, a slight gap is visible in the last 40 observations, in which the basic model overestimates the car commuters. This gap is absent in the advanced model. This depicts that interactions and policies play a role in the number of car commuters. Even for the earlier stated fact that the model of car commuters has the 'worst' fit, the trends are captured quite well by the model.

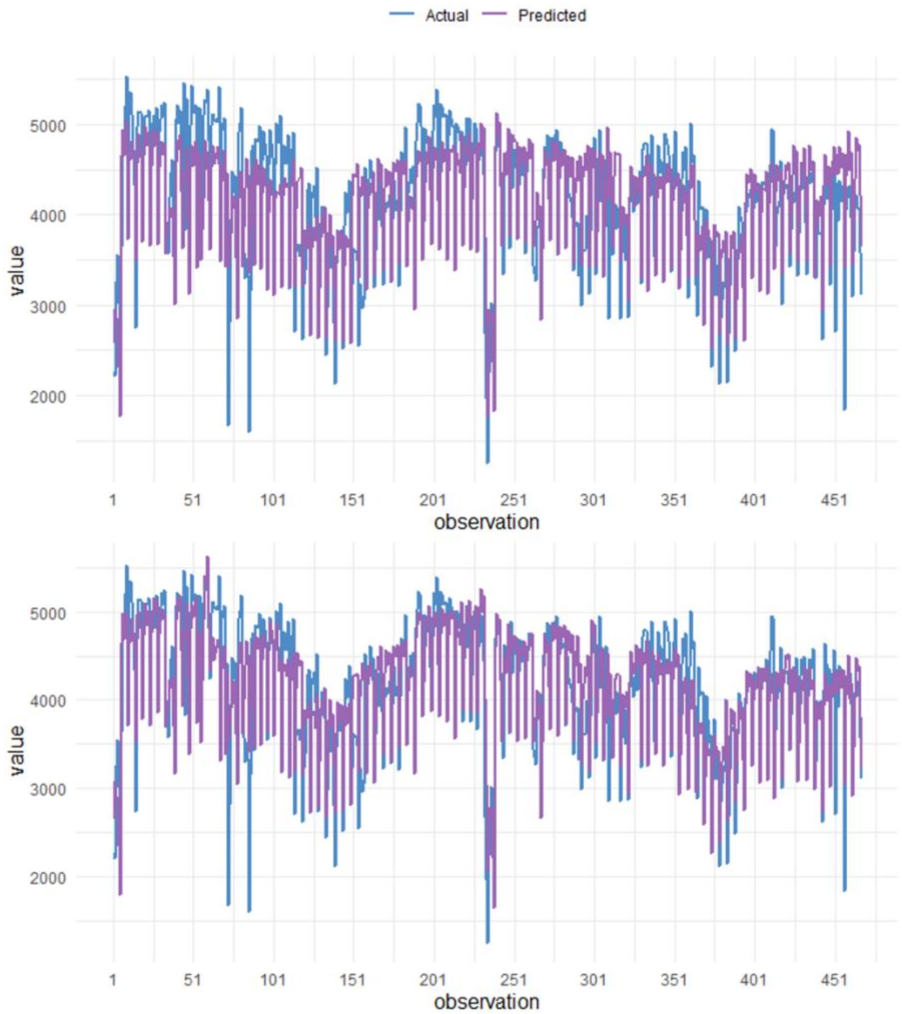


Figure 27: Basic (upper) versus advanced (lower) fitted values for car commuting

5.2.1.5 Employees in office

Employees in office has the highest explained variance compared to the other models. The residual standard error is lower than the residual standard error of the advanced car model, while there are a lot more employees in office than car commuters. The only other variable (except for interactions) that is not significant is the policy intervention of a higher bus frequency. Temperature and autumn have a significant interaction as well as all interactions regarding precipitation and season. This depicts that all seasons have a different effect on the number of employees in office when it rains.

Better temperatures increase the number of employees in office, while precipitation and snow decrease the number of employees in office. Wind speed and darkness do not change the number of employees in office. Tuesday is the busiest day in office followed by Thursday with a substantial decrease of employees on Fridays. Winter has when all other conditions are the same the highest office presence with summer the lowest. All school holidays, as for the underlying modalities decrease the number of commuters in office substantially, ranging from a decrease of 12% up to 44% of commuters.

Checking the model assumptions in Figure 28, shows the model-predicted lines resemble the observed data almost perfectly. This shows again that a response variable does not need to have a normal distribution for regression to work. One observation is impacting the Homogeneity of Variance and Linearity in the right-tail. This is the same observation that also was not detected in the models for bicycle commuters. The dots in the Normality of Residuals mostly fall along the line.

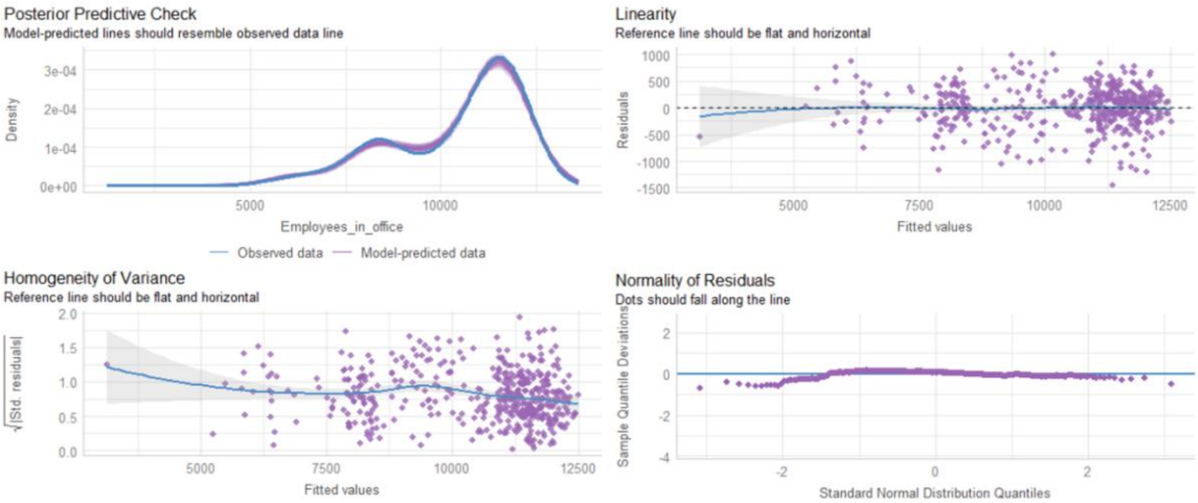


Figure 28: Visual inspection of model assumptions for employees in office

In Figure 29 it can be seen that the observation with the lowest fitted value is predicted quite close, despite the huge deviation of the other observations. In both models, the summer periods are quite noticeable and furthermore there is a slight downward trend noticeable regarding employees in office. Starting from observation 350 however, the basic model cannot follow the pattern as well as the advanced model compared to actual values.

Figure 29: Basic (upper) versus advanced (lower) fitted values for employees in office

5.2.1.6 Conclusions for travel volume models

It is for all modalities an improvement to add interactions and policies. Still, the basic models also show that weather and seasonality variables play a significant role in the number of commuters. For cycling this role is the most prevalent on a daily level and more changes between seasons are prevalent. However, the advanced models show that there are significant interactions also for car and bus between weather variables and seasons. Controlling for the policy periods, wind increases car commuters while rain has no significant effect but snow does. While car is invariant for precipitation, the employees in office substantially decrease as do the number of bike commuters. To varying degrees between holidays and between modes of transport, school holidays reduce the number of commuters. According to the models seasonality plays an important role on the number of commuters and Fridays lead to substantial decreases in commuters. Darkness is found to play a role on commuting behaviour, with especially a negative effect on cyclist and increasing the number of bus commuters. The patterns of commuting behaviour over time are matched by the models, which means the models can capture the trends very well with the used variables.

5.2.2 Modal split model

The modal split model combines the three modes of transport and adds ‘Other’ as a mode to ensure that the modal split is always equals to the number of employees in office and the chosen model ensures the total percentage of the added shares to be exactly 100% for each observation. ‘Other’ consists mainly of three things. Firstly, commuters that are not registered but checked-out or parked on places where registration tools are in place. Secondly, commuters that came by bus, bike, or car but parked or checked-out at places that do not have registration tools. Thirdly, commuters with a different means of transport (e.g. carpooled, shuttle bus, walked, dropped off at Kiss and Ride). First the basic model is depicted, followed by the advanced model. The section concludes with a conclusion regarding the modal split models. For a more detailed explanation of odds ratios and log-odds, Appendix G can be consulted.

5.2.2.1 Basic model

For this more complicated model than the models for travel volume, in Table 14 the estimates are displayed for the basic model. This model has the same variables as for the regressions. The estimates are log-odds, which are changed to odds ratio by exponentiating. The standard error of the log-odds (estimate) is depicted and the asterisks for confidence levels for different p-values.

Table 14: Overview of MLR model

	Dependent variable								
	Bicycle			Bus			Other		
	Estimate	SE	Odds ratio	Estimate	SE	Odds ratio	Estimate	SE	Odds ratio
(Intercept)	-0.111***	0.0063	0.895	-1.728***	0.0107	0.178	-0.237***	0.006	0.789
Temperature	0.009***	0.0004	1.009	-0.002**	0.0006	0.998	-0.004***	0.0004	0.996
Wind speed	-0.004***	0.0002	0.996	0	0.0004	1	0	0.0002	1
Darkness	-0.005***	0.0001	0.995	-0.001***	0.0001	0.999	-0***	0.0001	1
Precipitation (True)	-0.211***	0.0036	0.809	0.078***	0.006	1.081	0.02***	0.0033	1.02
Snow (True)	-0.103***	0.011	0.902	0.143***	0.0166	1.153	0.232***	0.009	1.261
Weekday									
Tuesday	-0.03***	0.0039	0.97	0.043***	0.0069	1.044	0.036***	0.0038	1.037
Wednesday	-0.055***	0.0039	0.947	-0.016*	0.007	0.984	0.014***	0.0038	1.014
Thursday	-0.061***	0.0041	0.941	0.019**	0.0072	1.019	0.002	0.004	1.002
Friday	-0.226***	0.0044	0.798	-0.258***	0.008	0.773	-0.055***	0.0042	0.947
School holiday									
Autumn break	0.022*	0.0091	1.023	0.03*	0.0152	1.031	-0.065***	0.0084	0.937
Christmas break	-0.38***	0.0273	0.684	-0.218***	0.0357	0.804	-0.076***	0.0194	0.927
May break	-0.034***	0.0085	0.966	-0.02	0.0156	0.98	-0.034***	0.0085	0.966
Spring break	-0.023	0.0175	0.977	0.011	0.0267	1.011	-0.029	0.0157	0.971
Summer break	-0.168***	0.0046	0.846	0.06***	0.0085	1.062	-0.087***	0.0046	0.916
Season									
Spring	-0.21***	0.0062	0.81	-0.229***	0.0106	0.795	-0.122***	0.006	0.885
Summer	-0.089***	0.0068	0.915	-0.261***	0.0118	0.77	-0.038***	0.0065	0.962
Autumn	-0.043***	0.0051	0.958	-0.178***	0.0082	0.837	-0.001	0.0046	0.999
Note:									*p<0.05 ; **p<0.01; ***p <0.001

The reference commute mode is car, with the odds ratios of the other modes therefore interpreted as an increase, decrease, or no effect when changing something compared to the baseline and references of car. Odds ratios are more intuitive to interpret than log-odds estimates, therefore in the text only odds ratios are used.

All bicycle related odds ratios are significant, except for Spring break. Only Autumn break and temperature increase the odds of taking bicycle compared to car. Friday has the most substantial negative effect on the mode share of all dummy variables. In addition to travel volume models that depict that Fridays decrease the number of commuters for all modes, the modal split model shows that the decrease is more substantial for cycling in comparison to car (as also for bus commuters). Wind speed was not significant for the number of commuters, but from the mode share model it follows that on days with a higher wind speed, more people take the car than bicycle. Autumn break is the only holiday in which odds increase for taking the bicycle in comparison with baseline.

For bus regarding weather and seasonality, only the wind speed does not have an effect compared with the share of car commuters. Of all odds ratios estimated for bus, snow has the highest odds ratio and also precipitation has a positive odds ratio. This means that compared to car, on these kinds of days there is an increase in the share of buss commuters. On Tuesdays and Thursdays, the odds of taking the bus are higher than on Mondays. On Fridays this substantially decreases for bus commuters. The effect of holidays depends. Autumn and Summer break significantly increase the odds of taking the bus compared to car, while Christmas break significantly decreases the odds and May break and Spring break do not differ significantly.

The other category, which is either another means of transport utilized or missed observations for the counting mechanisms for the other modes, has only for wind speed no significant difference regarding weather conditions. The effect of darkness is significant, but is -0.00044, which results rounded to a zero as estimate and a 1 as odds ratio. Apart from the Spring break, all holidays have decreased odds for the 'Other' category. Which means either less missed observations or more people traveling with one of the three main modes. The Spring has from all dummy-coded options the lowest odds for 'Other'.

With the model it is possible to estimate the modal split and compare it with the actual values. In Figure 30 this is graphically shown. Values are estimated for the dataset excluding outliers and influential observations. The x-axis depicts in chronological order the 377 observations.

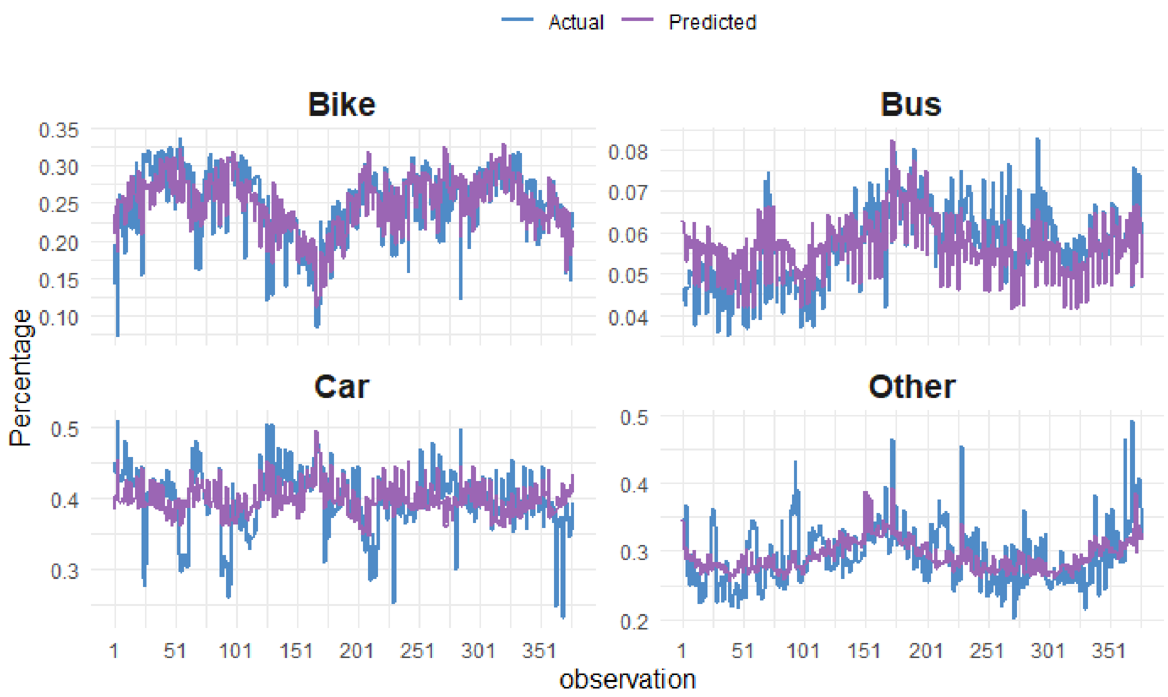


Figure 30: Modal split for basic model

The model captures the underlying trends in the modal split quite well. Still, there are some peaks noticeable in the plot which are deviations from estimated values. Despite efforts of dealing with influential points in the travel volume models and use the same outliers for this model, a small number of observations remain that significantly diverge from the estimated trend. These likely represent exceptional but plausible scenarios that are not captured by the model, but the presence of the outliers does not seem to significantly affect the overall trends that are rightly captured by the model. It is remarkable that most peaks in car, are the same but inversed for other. This leads to the suggestion that the measuring mechanisms of car are the most prevalent for substantial measuring errors. While bus commuters is a smaller percentage and for bicycle commuting a part of the measuring errors was countered by MICE this could be expected

In the bus modal split, as in the travel volume models, it is visible that first the percentage of bus commuters is overestimated, and after the moment where the bus frequency was increased the predicted values underestimate the actual values. The bus model shows an increasing trend despite seasonal variations, which is more apparent in the actual data than in the predicted data. In the car model split graph, in the last part of observations the predicted values are higher than the actual values, with also two exceptionally low spikes in comparison with the rest of the observations. While interactions make the model even more complex, the resulting graphs indicate that an effect in time is probably missing which causes over and under estimations.

While the y-axis are scaled in Figure 30, in Figure 31 the same graphs are depicted but then with the same y-axis.

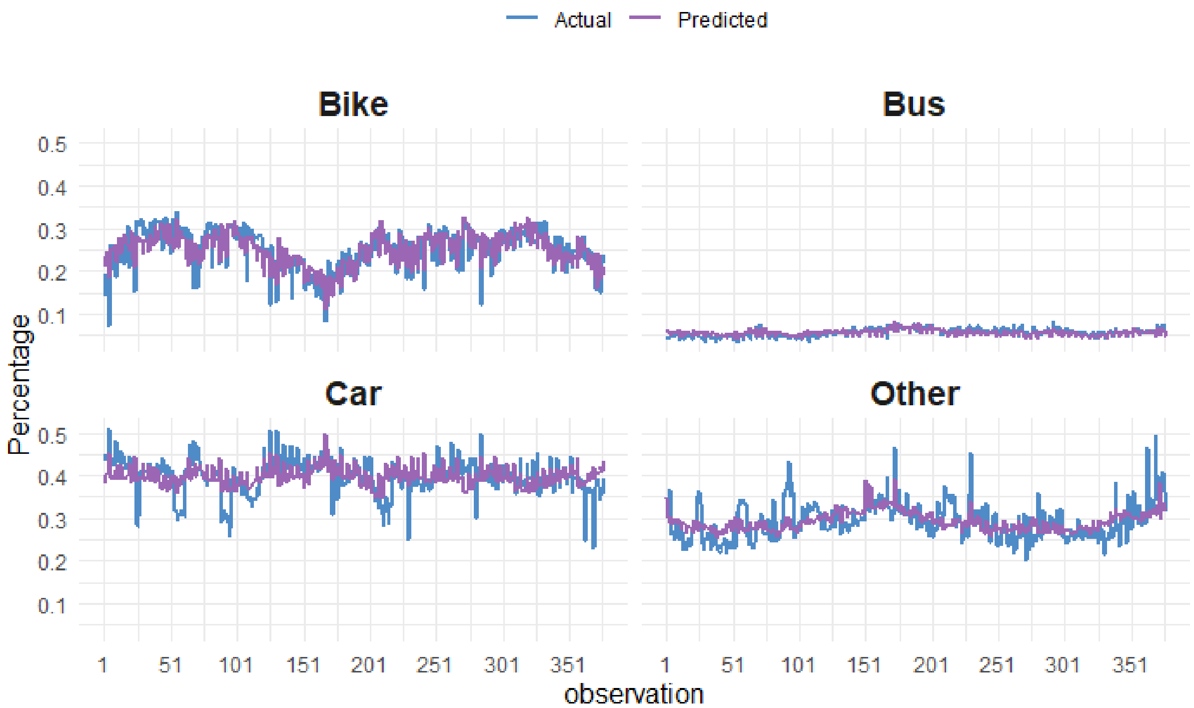


Figure 31: Modal split for basic model, same axis

5.2.2.2 Advanced model

It is assumed that policies influence the modal split. An ANOVA likelihood ratio test indicates that a model with policies significantly improves the model (LR stat. 5524.282 with a P-value of approximately zero. Table 15 provides an overview of estimates the odds ratios relative to Car and baseline.

Table 15: Overview of MLR model including policies

	Dependent variable								
	Bicycle			Bus			Other		
	Estimate	SE	Odds ratio	Estimate	SE	Odds ratio	Estimate	SE	Odds ratio
(Intercept)	-0.149***	0.0077	0.862	-1.992***	0.0134	0.136	-0.305***	0.0075	0.737
Temperature	0.006***	0.0004	1.006	-0.003***	0.0007	0.997	-0.007***	0.0004	0.993
Wind speed	-0.003***	0.0003	0.997	0.002***	0.0004	1.002	0.002***	0.0002	1.002
Darkness	-0.004***	0.0001	0.996	0	0.0002	1	0**	0.0001	1
Precipitation (True)	-0.218***	0.0037	0.804	0.047***	0.0062	1.049	-0.003	0.0034	0.997
Snow (True)	-0.094***	0.011	0.91	0.141***	0.0167	1.151	0.237***	0.009	1.267
Weekday									
Tuesday	-0.032***	0.0039	0.968	0.038***	0.0069	1.039	0.031***	0.0038	1.031
Wednesday	-0.055***	0.004	0.946	-0.013	0.007	0.987	0.014***	0.0038	1.014
Thursday	-0.059***	0.0041	0.942	0.018*	0.0072	1.018	0.004	0.004	1.004
Friday	-0.227***	0.0044	0.797	-0.255***	0.008	0.775	-0.054***	0.0042	0.948
School holiday									
Autumn break	-0.097***	0.0095	0.908	0.026	0.016	1.027	-0.169***	0.0088	0.844
Christmas break	-0.38***	0.0273	0.684	-0.184***	0.0358	0.832	-0.058**	0.0195	0.944
May break	-0.033***	0.0086	0.968	0.002	0.0156	1.002	-0.031***	0.0085	0.969
Spring break	-0.012	0.0175	0.988	0.031	0.0267	1.032	-0.01	0.0157	0.99
Summer break	-0.17***	0.005	0.844	0.033***	0.0093	1.034	-0.112***	0.005	0.894
Season									
Spring	-0.161***	0.0067	0.852	-0.109***	0.0114	0.897	-0.062***	0.0065	0.94
Summer	-0.005	0.0081	0.995	-0.062***	0.0142	0.94	0.089***	0.0079	1.093
Autumn	0.062***	0.0082	1.064	-0.006	0.0145	0.994	0.108***	0.0078	1.114
Policy									
Bus freq	-0.255***	0.0072	0.775	0.033*	0.0126	1.034	-0.194***	0.0067	0.824
Bus freq, Shed	-0.018***	0.0041	0.982	0.248***	0.0074	1.281	0.025***	0.004	1.025
Bus freq, Shed, No e-bike	-0.018*	0.0089	0.982	0.292***	0.015	1.34	0.066***	0.0086	1.068
Bus freq , Shed, Drop	-0.078***	0.0048	0.925	0.126***	0.0089	1.134	-0.13***	0.0048	0.878
Bus freq, Shed, Drop , 35 cents	-0.018**	0.0067	0.983	0.171***	0.0118	1.186	0.044***	0.0063	1.045

Note: *p<0.05 ; **p<0.01; ***p <0.001

For bike, only Spring break and Summer do not significantly differ the odds in regard to car commuting. Only temperature and Autumn have an odds ratio higher than 1. Regarding polices, compared to baseline all policies have lower odds on cycling, which may indicate an overall downward trend. If it is compared however with the bus frequency which is introduced in the second half of October, for October and November (which are both Autumn) the average odds are 26.8% more compared to a year earlier if same weather is assumed. Darkness does only change the odds regarding cycling, but no significant effect was found for other means of transport and bus commuting. Snow increases bus commuters and commuters by other means of transport. For all alternative modes, odds decrease on Friday, which means the odds are the highest when comparing days to take the car on Fridays. While for all modes the odds ratio with 35 cents compared to bus frequency is higher, it can be stated that the share of car commuters decreased between policy periods.

In Figure 32, the modal split over time for the model including policies is added. Trends are captured, but the model struggles to capture some spikes and cluster spikes with the variables used to estimate the model.

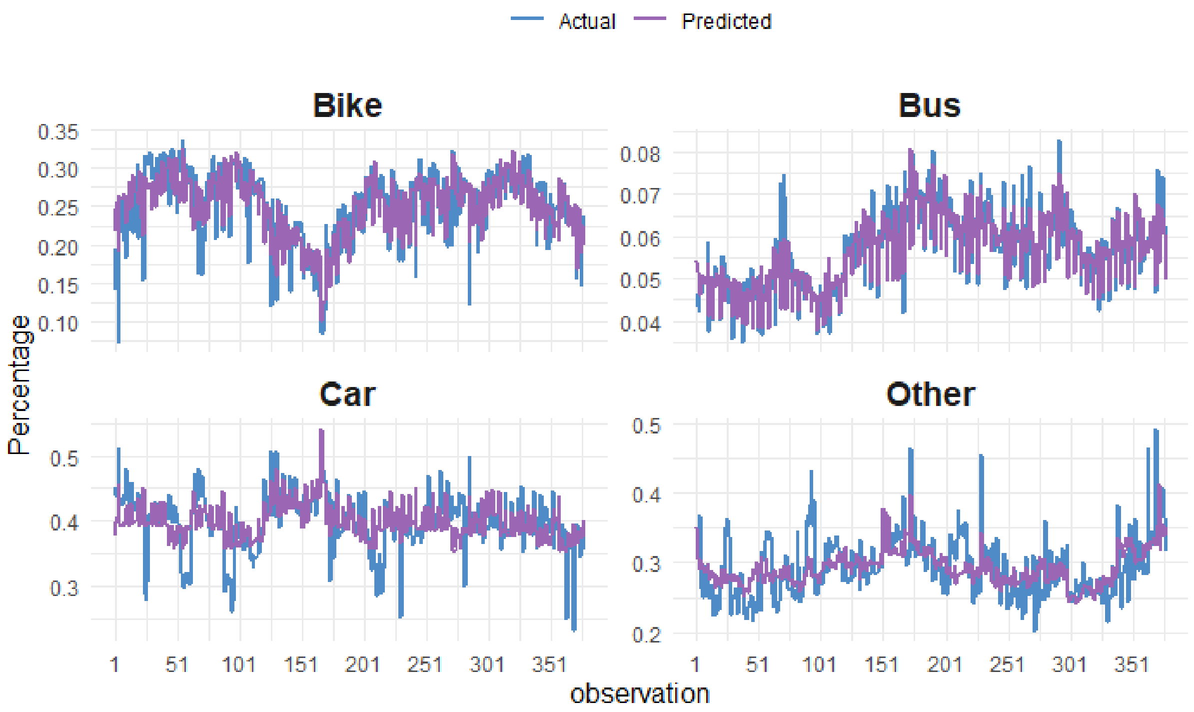


Figure 32: Modal split for advanced model

In Figure 33, the modal split is shown with the same y-axis to see that bus is, although in Figure 32 an upwards trend can be seen, bus commuters have a fairly small share of the total.

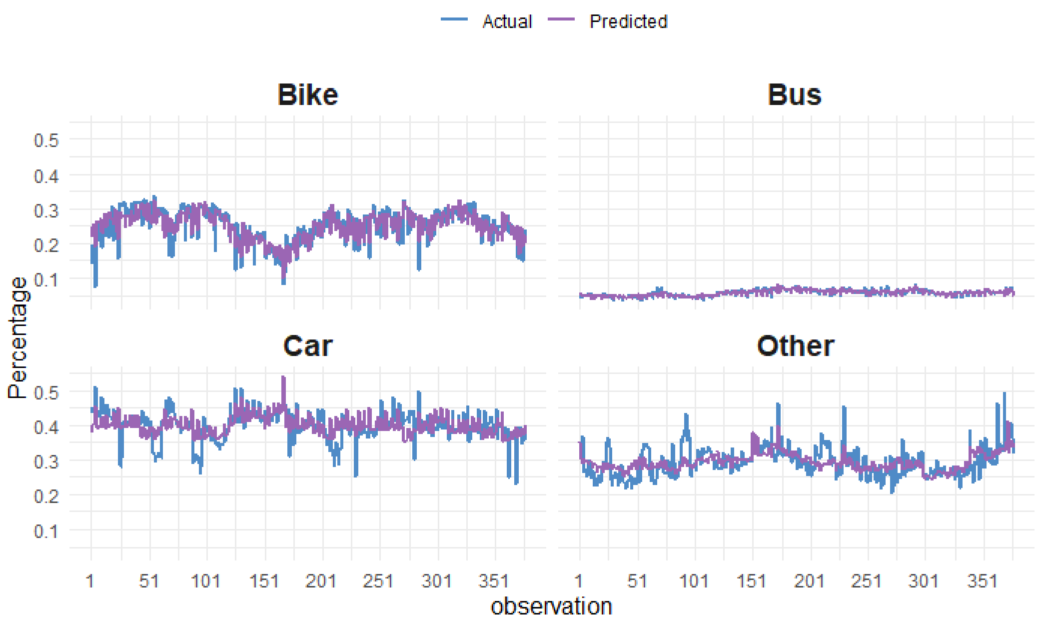


Figure 33: Modal split for advanced model, same axis

5.2.2.3 Conclusions for modal split model

One of the most notable things is that on Fridays, the modal split substantially shifts to car and therefore away from more sustainable transport modes. Also, the effect of wind which was not very present for the number of commuters does contribute significantly to changes in the modal split. While bus and other are not significantly different regarding darkness from car, but bicycle is this means people from bicycle shift to relatively to these three categories. While the odds ratio is 0.996 for each minute, when taking the average darkness of 29 minutes leads to an odds ratio of 0.89 which is a greater decrease than days with snow (when all other variables remain constant).

5.3 Time Series Analysis

5.3.1 Decompositions

For travel volume models, first the detected outliers are replaced with estimates from the static models. For each mode, the time series is decomposed. The x-axis notes the week, with week 52 being the final week of 2023. For car and bus, it starts at week 1, while bicycle starts later due to no cameras in the first part of 2023. The decomposition of the bike commuter time series is showed in Figure 34.

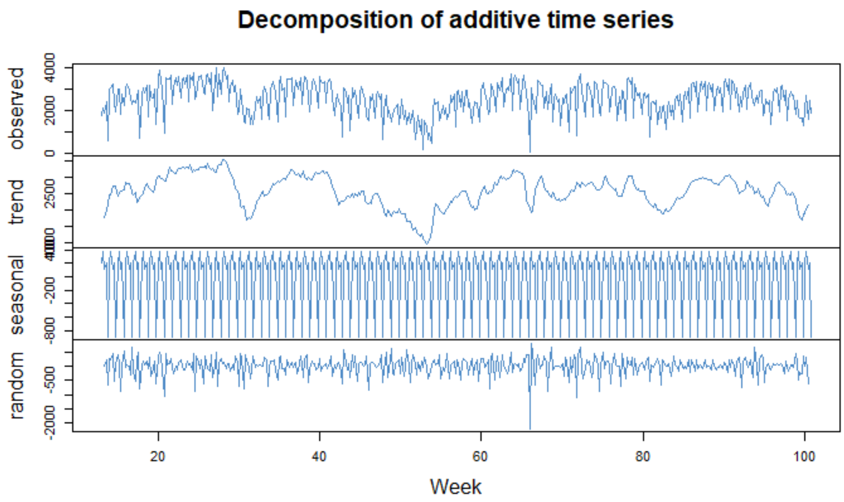


Figure 34: Bicycle decomposition

In the seasonal part, the pattern of Monday to Fridays is successfully captured. The trend starts going up until the beginning of the week 28 which is in the summer. There is a recovery but this is followed by a strong downward trend until the start of 2024. After a strong increasing trend there is not a clear up or downward trend starting from week 70 except from the summer break.

Figure 35 shows the decomposition for bus commuters.

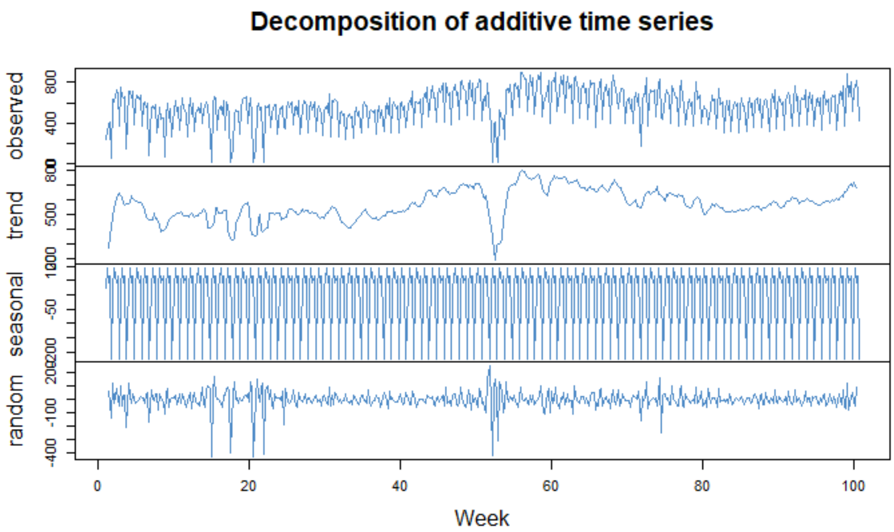


Figure 35: Bus decomposition

After decomposing the seasonal trend, which is work weeks, an overall upward trend is noticeable. With a slightly U-shaped parabolic trend in 2024 with the minimum in the summer break. Christmas break has a huge gap in the trend, which is accompanied by also significant residuals in the random part. This means the model struggles with capturing the impact of this period. There are also divergent residuals clustered around week 20.

Apart from the trend, the same things are noticeable in the decomposition of car, which is depicted in Figure 36. The trend for car however is not upward but seems to be slightly decreasing. In the same periods of 2023 and 2024 there are spikes, which is the season with many statutory holidays. These fit less in the trend, while some of them are not in a school holiday for example.

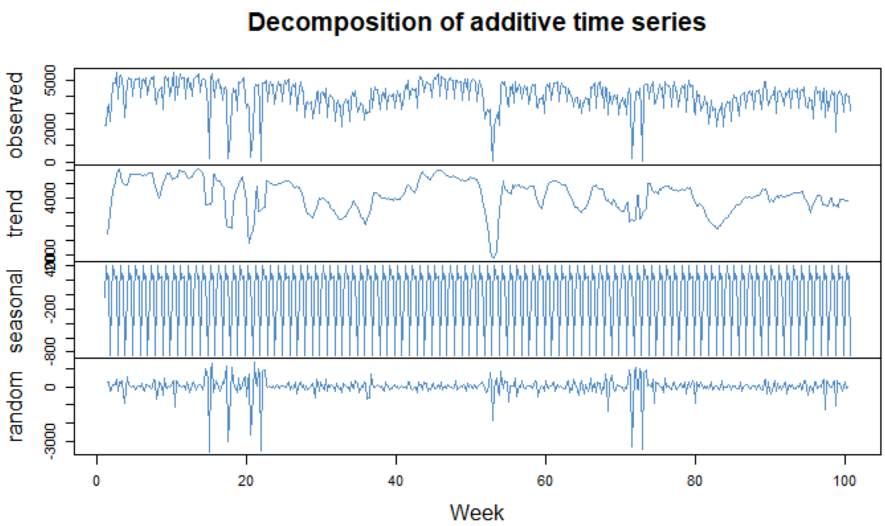


Figure 36: Car decomposition

5.3.2 SARIMAX

For bus commuters, a SARIMAX model resulted with automatic model selection for the best model in an ARIMA(0,0,2)(1,1,1) model with frequency of cycles being 5 for weekdays. Table 16 depicts the estimates for the SARIMAX model, and where possible, coefficients are compared with the static models for bus. Due to interactions in the advanced OLS, all estimates regarding seasons and weather conditions are omitted.

Table 16: Overview of SARIMAX compared to regression models for bus commuters

	Model		
	SARIMAX	Basic OLS	Advanced OLS
	Estimate	Estimate	Estimate
Temperature	0.091 (0.8990)	-0.19 (2.07)	
Wind speed	1.228(0.5543)	0.66 (0.88)	
Darkness	0.821 (0.1756)	0.56 (0.87)	
Precipitation (True)	39.683 (7.5595)	48.01*** (9.94)	
Snow (True)	-6.197 (17.0454)	12.59 (43.01)	
Weekday			
Tuesday		68.65*** (6.71)	
Wednesday		10.34 (6.88)	
Thursday		45.5 ***(8.23)	
Friday		-255.7*** (8.56)	
School holiday			
Autumn break	-35.803 (24.9469)	-68.01*** (18.8)	-40.9* (18.93)
Christmas break	-356.029(23.5837)	-385.32*** (53.12)	-388.89*** (18.9)
May break	-45.56(24.0351)	-58.29 (31.56)	-55.99 ***(15.57)
Spring break	-125.59 (23.8253)	-143.54 (101.6)	-159.48 ***(26.06)
Summer break	-44.217 (13.6258)	-35.49* (14.62)	-31.45* (15.01)
Season			
Spring		68.78*** (13.76)	
Summer		140.62*** (14.9)	
Autumn		83.41*** (17.78)	
Policy		70.79***(10.23)	
Bus freq	93.513 (19.3045)		68.78*** (13.76)
Bus freq, Shed	159.357 (20.7570)		140.62*** (14.9)
Bus freq, Shed, No e-bike	105.78 (35.4379)		83.41*** (17.78)
Bus freq , Shed, Drop	92.76 (32.3684)		70.79***(10.23)
Bus freq, Shed, Drop , 35 cents	99.594 (35.8732)		48.72***(11.7)
ARIMA terms	0.1146		
MA1	(0.0463)		
	0.1075		
MA2	(0.0470)		
	0.0265		
SAR1	(0.0666)		
	-0.9024		
SMA1	(0.0522)		
	-270.8510		
Statutory holiday	(21.3097)		

While it can be seen as comparing apples with oranges, all coefficients have the same directions and orders of magnitude (except for temperature and snow with high standard errors). The Root Mean Squared Error (RMSE) of the ARIMA model is 66.378. Without using statutory holiday as a variable, the ARIMA model changes in an ARIMA(2,0,2)(2,1,2) with RMSE 76.14. To compare, on the same data the advanced static model has a RMSE of 75.74. Autoregressive and moving average components of the ARIMA model make it more complex and give limited interpretability. Therefore, no ARIMA models are established for the other modes and focus is shifted to other methods to evaluate the effectiveness of policies.

Figure 37 shows the fitted versus actual values for the SARIMAX model. From this, it is visible that although coefficients are difficult to interpret it is a suited model for capturing trends and therefore for predicting the future it can have potential.

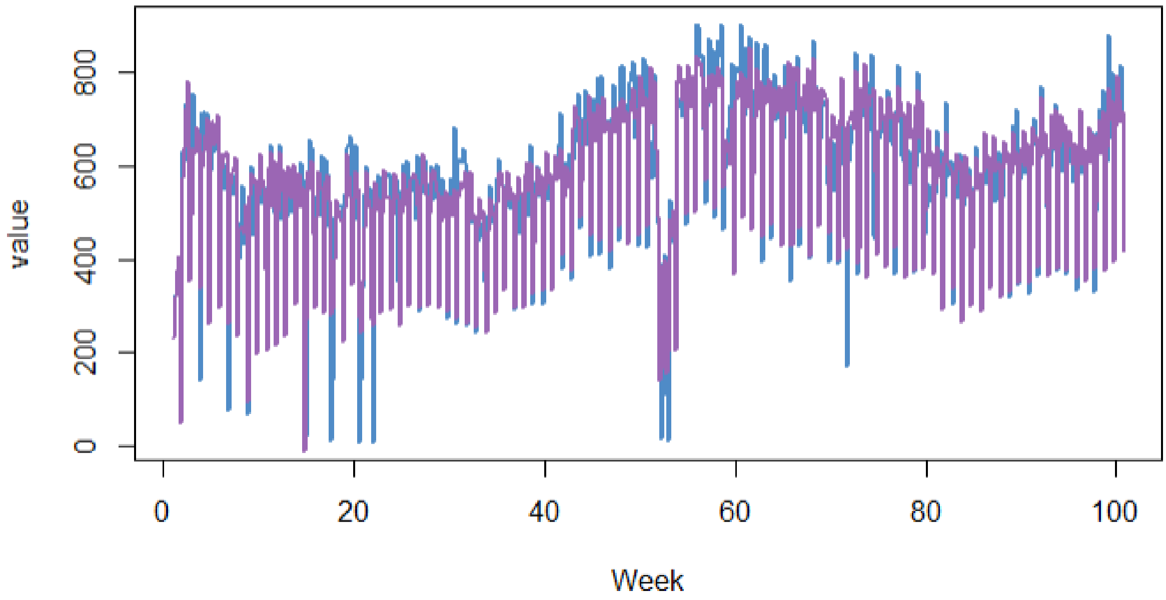


Figure 37: SARIMAX predicted vs actual values for bus commuters

5.3.3 Interrupted Time Series Analysis

Previous models helped with understanding how policies over time are related with the number of commuters for each mode and the modal split. With these models an attempt is made to account for variables that are deemed to have an impact to estimate if after policies there was a significant lasting effect. However, there could always be unseen variables that also can be the cause for a change in commuters and modal split. Interrupted Time Series (ITS) analysis is not focused on trends over the season, but on the impact right after introduction of the policy. Fridays are disregarded while Fridays having significantly less commuters which causes a bias in estimating the immediate effect after intervention. The aim is to disregard seasonality and purely look at the impact of the intervention.

As validation of the effect, for each intervention exactly one year earlier or later is used while this best represents the circumstances. While only two years of data are present it was only possible to have one validating graph for each intervention.

5.3.3.1 Increasing bus frequency

In October 2023, the frequency of direct buses to ASML locations was doubled from four to eight. Figure 38 depicts the interruption compared to the number of bus commuters. In Figure 38, on the left side the interruption is implemented. On the left side, 2024 is used validation. It can be seen that the number of commuters by bus increased more in 2023 than in 2024. The effect seems to be consumed by higher number of bus commuters towards December.

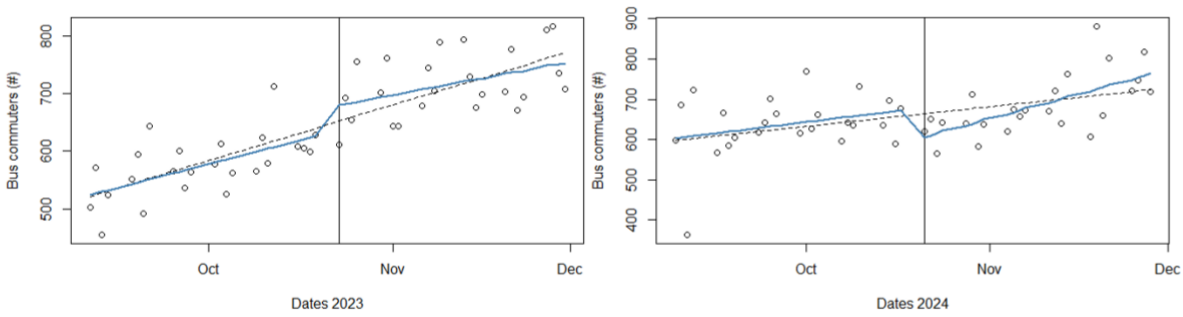


Figure 38: ITS of bus frequency (volume)

The same graph, but for the modal share of bus commuters is depicted in Figure 39. There is not an excessively change to be seen. The introduction of this policy instrument therefore does not appear to have a strong direct effect, but as the decomposition shows it could have contributed to an upwards trend for bus commuters on the long term.

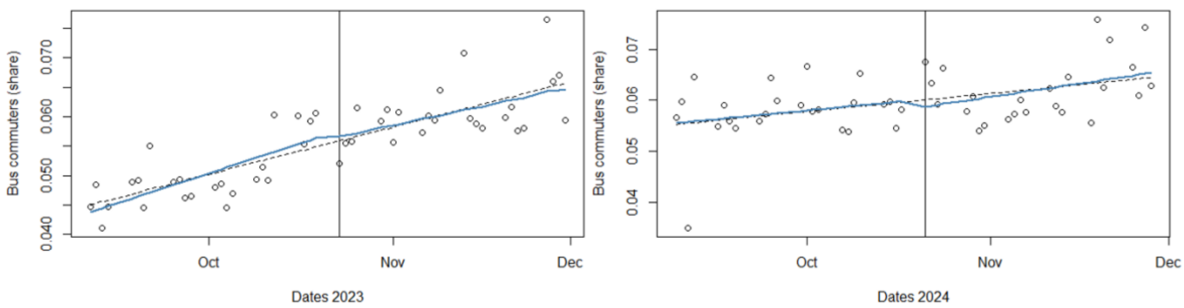


Figure 39: ITS of bus frequency (modal share)

5.3.3.2 Increasing the bicycle allowance to 35 cents

In October 2024, the bicycle allowance was increased to 35 cents. Contradictory to the bus intervention, now as validation the previous year (which is 2023) is used. It can be seen in Figure 40 for the number of commuters that in 2023 there was a decreasing trend from October onwards. The trend in 2024 is more stabilized after the start of October (for both years, a Tuesday is used for the vertical line).

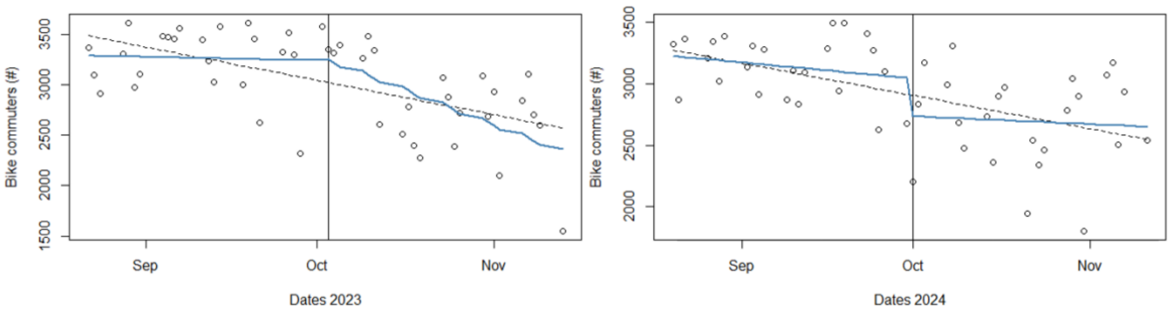


Figure 40: ITS of bicycle allowance (volume)

When looking at the figure that shows the mode shares, which is Figure 41, there are also clear patterns after the interruption. The share of cycling commuters had a decreasing trend in 2023 from September to November but it is noticeable that in 2024 the trend did not continue with decreasing and a higher share of cycling commuters sustained.

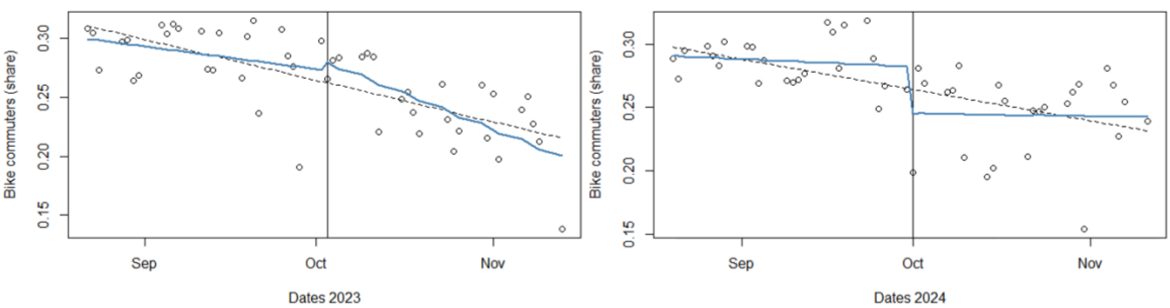


Figure 41: ITS of bicycle allowance (modal share)

5.3.4 Conclusions for Time Series Analysis

Trends are clearly visible and a seasonal weekly pattern can be captured for the modes. It can be seen that seasonality is present For interpretability and additional cost of time it has no added value with the data at hand to use a SARIMAX model, while the performance equals the performance of the static models. SARIMAX does capture trends very well and for predicting instead of explaining it still can be a good method. The ITS showed impacts of two policies directly after implementation. The bus intervention seems more to have a long-term effect than a sudden effect when comparing with the decomposition of bus commuters. For the bicycle allowance, it looks like the trend is interrupted in 2024 which indicates a more sudden effect.

6 Discussion

This thesis investigates the extent to which weather conditions, seasonality and implemented polices are related to day-to-day variations in commuting behaviour, for which as case data of commuters from ASML is used. For each main subject, first the findings are interpreted and linked to the literature.. While the findings contribute to several underexplored areas in the literature regarding travel behaviour and a shift towards sustainable mobility, this is supplemented with implications. After each topic is covered, the overall limitations of this research are stated. The chapter ends with policy recommendations and directions for future research.

6.1 Interpretation of results and implications

6.1.1 Everyday weather

As everyday weather may have substantial impact on travel behaviour, there is an increase in focus on this topic (Böcker et al., 2016; Liu et al., 2017). Active mode users are regarded the most sensitive to weather (Böcker et al., 2013; Liu et al., 2017), followed by public transport and the least affected mode being car usage (Faber et al., 2022). In line with the literature, this research found that cycling when disregarding interactions with seasons is most responsive to daily weather conditions, followed by bus commuters. Looking at the travel volume models without interactions and polices, no significant relations were found between car and daily weather, while for bus only rain significantly boosted the number of bus commuters. For cycling lower temperatures, precipitation, and snow decreased the number of commuters, which is in line with weather conditions being a stated reason to not use a bicycle from the research of Heinen et al. (2010). This is also evident according to the Multinomial Logistic Regression models, that show odds ratios smaller than 1 for precipitation, snow, and wind speed depicting a decrease in share and higher temperatures having a greater than 1 odds ratio which logically also means that the share of cyclists decreases with lower temperatures. As it was stated that weather significantly influences mode choice by several studies (Böcker et al., 2013; Liu et al., 2017; Sabir, 2011), this research affirms this while the daily weather related coefficients for bike and bus all significantly differ from the reference (car) in both MLR models.

According to the review of Liu et al. (2017), rain and temperature are regarded the most impactful regarding travel behaviour. This research found also significant relationships between these two, although in general precipitation is substantially more impactful. For the travel volume models it was found that the temperature must increase almost the whole scale measured in the observations (from -6.4 to 21.8 degrees Celsius) to compensate for the effect of precipitation. While Koetse and Rietveld (2009) found that precipitation is by far the most important variable regarding safety, it seems it does not affect the car commuters. However, snow does affect car usage significantly which suggests that for car commuting snow is regarded more important regarding safety than rain. Wind is often mentioned as having an influence on travel behaviour (Faber et al., 2022; Gössling et al., 2023; Sabir, 2011). This research affirms this especially for mode shares, while in the MLR models it is evident that the speed of the wind decreases the share for cycling compared to car and bus. For volumes, this was less evident except in the model with interactions, which is discussed in section 6.1.2. This research therefore contributes to the gap in knowledge regarding the influence of wind speed on travel behaviour, while wind is often overlooked according to Heinen et al. (2010) and Böcker et al. (2013).

While most findings form other research could be confirmed or strengthened there is also research where this is not straightforward. Faber et al. (2022) found that commute trips are less affected by the weather than leisure trips, which is depicted by Liu et al. (2017) as being less elastic in response to weather changes. While this research focuses solely on commuting, this cannot be doubted.

However, this research still shows that although being considered less elastic and affected, weather still plays a significant role in travel behaviour for commuting.

Contrary to almost all literature, Ton et al. (2019) found no significant effect of weather on active transport use in the Netherlands. Explanations of the authors are that habitants are used to the mild climate or how weather is incorporated in their study. As opposed to Ton et al. (2019), this research found substantial evidence of weather affecting active transport. Based on the results, the first reason mentioned cannot be agreed with. The second reason is valid, as the study relied on subjective interpretations from respondents regarding whether or not the weather was extreme on a given day. This research looks at everyday weather in a quantitative way. Another contrary finding is that the weather conditions are not strongly correlated, while the review from Liu et al. (2017)states that meteorological variables are often correlated. An explanation could be the variables used; another explanation could be the context-dependency: in the context of Veldhoven which has a mild Dutch climate there is no strong correlation. It could be that correlations are more evident in other climates.

Implications
While literature tends to focus on extreme weather (Gössling et al., 2023 ; Ton et al., 2019), this research contributed to the knowledge regarding everyday weather that is inadequately addressed although becoming more and more a focus in research due to climate concerns (Böcker et al., 2013; Böcker et al., 2016; Liu et al., 2017). As wind is understudied (Böcker et al., 2013; Heinen et al., 2010), this research also contributed to the knowledge regarding wind. From this study, it is evident that wind significantly contributes to changes in mode choice. This research also considers the influences of weather on travel volumes of different transport modes as well as on the interrelationships between mode choices. Therefore, the research also contributes to insights into how people will travel less (reduction) and how people will travel differently (alteration) towards more sustainable mobility (Banister, 2008; Berger et al., 2014).

6.1.2 Seasonality

While in the literature summer and autumn are mentioned as months which are the most favourable for cyclists (Böcker et al., 2013; Heinen et al., 2010), it was found that spring and summer have almost equal commuters by bicycle with less cyclists in Autumn according to the boxplots. When looking at the office presence it is noticeable that during summer less people are coming to office with one of the reasons being summer holidays. This suggests that seasons have different influences on commuting as trip purpose in comparison with other purposes e.g. leisure. Regarding autumn in comparison with spring, there is almost no darkness in the morning peak hours in spring, except immediately after implementation of Daylight Saving Time. Autumn has before and after the switch back to Standard Time, darkness in the morning peak hours. The majority of commuters travelling in the morning peak hours compared to traveling for other reasons which is not as bounded to the morning peak hours with darkness can also cause popular cycling seasons to differ from the literature.

That darkness, or the lack of daylight hours, has a negative influence on the number of bicycle commuters is briefly mentioned in the review of Heinen et al. (2010). This is in line with Wessel (2022) stating literature regarding darkness on travel behaviour is lacking. In this research, darkness is included as predictor in the travel volume models including interactions with seasons, and in the MLR models as predictor. In line with Heinen et al. (2010) it was found that darkness negatively impacts the number of commuters by bike. A significant interaction between spring and darkness depicts that people perceive darkness in the spring as more averse compared with Winter. This could be explained by the Daylight Saving Time starting in Spring. This means that first there is no darkness in the morning rush hour and then after the switch there is darkness again in the morning rush hours. The modal shares between bus and car do not significantly differ, but the odds ratio for cycling is lower than one with a substantial decrease when there is more darkness in minutes. The insignificant relationships between car, bus, and, other depict that cyclist relatively redistribute in the proportions of the other options.

On the assumptions in the literature that different weather conditions can have different effects in different seasons (Liu et al., 2017), which is briefly mentioned in the beginning of this section a small nuance needs to be made. For assessing the interactions, which turn out to indeed have significant effects, also season as categorical variable needs to be integrated which gives a direct effect. This complicates interpretation of the exact effect of season in conjunction with weather conditions. Spring has for example a negative coefficient for bicycle commuting. This is contradictory but can be explained by the underlying assumptions of regression. It depicts the change caused by a differing variable when all other conditions are kept the same. While temperature and darkness are linked to an extent with seasons, a negative coefficient for spring says if certain winter weather conditions would have been in the spring, there would be less commuters. This makes it more logical while winter conditions in spring are detrimental. For the MLR models, the following explanation can be given for counterintuitive coefficients: on first impression, the odds ratio for spring and summer less than 1 is not what was expected from boxplot. But to give a simple example, an average winter (which is the reference season) and average spring day on a Monday (which is the reference weekday) with no holiday (which is the reference school holiday) are used. Average as in, taking the mean temperature and darkness minutes and wind for a Winter and doing the same for Spring. Note, that centred temperatures and winds are used. This leads to the following odds ratio for biking on an average spring Monday versus biking on an average winter Monday, with coefficient of the basic MLR model and the numerator being spring and the denominator being winter:

$$Odds\ ratio = \frac{0.895 * \ 1.009^{0.859} * 0.996^{-2.782} * 0.995^{0.563} * 0.81}{0.895 * \ 1.009^{-2.998} * 0.996^{-0.518} * 0.995^{42.2}} = \frac{0.737}{0.708} \cong 1.04$$

The odds of cycling increase with 4%, which leads to a higher modal share of cycling. This shows, that solely looking on odds ratios might cause misinterpretation. Odds ratios are always relative to the baseline, and the baseline is zero darkness. This does not occur in the winter, and this causes the problem with direct interpretation or statements that people cycle less in spring, which is not true (for Mondays when there are no school holidays). Reasoning from this analogy is therefore important and should be kept in mind when making conclusions.

Seasonality factors as school holidays as well as Fridays lead to less commuters in office. However, the mode shares also shift more to car. This is probably due to that there is more space to park which stimulates undesired behaviour. It could be augmented for that especially on busier days a sustainable modal split is desired, but the lesson can be learned that when people see opportunities it is easy to switch back to unwanted behaviour.

Implications

It is agreed upon in the literature that seasonality plays a role, but there is no consensus how to effectively integrate it. Therefore, Heinen et al. (2010) stated that research is needed that utilizes data on travel behaviour over a long timeframe to better incorporate seasonality. This research utilized travel behaviour of 500 consecutive working days. Therefore, this study could better incorporate seasonality and assess the difference between seasons in travel behaviour. This is something that is not sufficiently studied according to Liu et al. (2017), hence this study also contributes to the quest for effectively integrating seasonality although it is still complicated. The third seasonality related lack of knowledge this research contributes to, is regarding darkness. As evident from the research from Wessel (2022), literature on light conditions and the relationships with travel behaviour is rather scarce. Although Heinen et al. (2010) mention the negative influence of darkness and Wessel (2002) looked into effects of Standard Time and Daylight Saving Time for cycling, this research contributed to the scarce literature by looking at light conditions throughout the years for multiple modes of transport. This is expressed by using the minutes of darkness during the morning rush hour in this research and also by incorporating a difference between Daylight Saving Time and Standard Time. As discussed, it seems that darkness for commuters is even more of influence than temperature and can compete with precipitation as most impactful climate related variable.

6.1.3 Assessing policies towards sustainability

From a wide range of possible interventions, just a few are implemented (Gössling & Cohen, 2014). Three approaches for a shift towards sustainable mobility are depicted: a mode shift which is an alteration of mode choice, less travellers which is a reduction of travel volume, and more efficient travels which leads to an efficient transport system (Banister, 2008; Berger et al., 2014; Griffiths et al., 2021). In this research, the investigated policies are presumably for alteration with efficiency sometimes as secondary issue. Introducing policies leaded to significant effect, although sometimes not as effective as expected:

- According to de Haas et al. (2022), introducing shared e-bikes could substitute car trips for commuters. This effect was not found in the MLR model when comparing the policy of no e-bike with the implementation of Drop e-bikes. Two logical explanations can be the reason. Firstly, the odds ratio of commuting by bus is lower after implementation of Drop e-bikes than when no shared e-bikes were available meaning that substitution could have been the case from bus commuters towards cyclists. The second explanation is that the period of no e-bikes was just two weeks during summer break, a tactical chosen moment and due to the few observations the model does not capture the effect as well as expected.
- Dutch research from MuConsult (2019) concluded a slight increase of the number of cycling commutes for each 10% increase in allowance. In this research, higher odds ratio was found after introducing the 35 cents compared to the policy phase before and in the Interrupted Time Series it seems that instead of a negative trend as in 2023, the trend stabilizes in 2024. Caution is needed while the period is relatively short and there were problems with the counting camera's.
- Fare-free public transport is found to increase public transport commuters, but a reduction in car commuters was not noticeable (Bull et al, 2021; Busch-Geertsema et al., 2021). Zeiske et al. (2021) concluded that incentives regarding public transport result in a temporary effect. In this research, it can be seen that bus odds ratios first increase but in later policy phases decrease. However, this can also be due to other implemented policies (in the beginning the bus frequency leading to an increase and in the end the bicycle incentives leading to a decrease. But, while the trend is upwards there are either more people having a fare-free card or more people using public transport (or a combination). Substitution with car is in line with the research minimal, while on the total the bus share increased from around 5% to 6.5%.
- The literature focuses more on crowding of buses. In this research, crowding could not be measured but there are indications that bus frequency helped with more bus commuters. The effect seems higher a few months later (when the new bicycle shed opened) which shows that there could be a lagged effect or that more people appreciated the higher frequency in the harsher winter weather.

The list above depicts that it is hard to isolate effects for the policies regarding alteration. One policy regarding reduction however seems to confirm the literature: As depicted by Griffiths et al. (2021), COVID-19 has provided an opportunity for changing travel behaviour and Barrero et al. (2021)

prognosed that working from home was here to stay. The research shows in the Exploratory Data Analysis that the office presence is percentual lower at the end of the observation period compared to the start of the period.

Implications

Research on effectiveness after implementation of policy interventions to encourage sustainable commuting is lacking (Griffiths et al., 2021). This research attempted to assess policies after implementation in two ways. Firstly, on immediate effects with Interrupted Time Series Analysis for the frequency of the bus line 119 to ASML and the increase in bicycle allowance. Secondly, on incorporating some of the implemented policies by ASML in the models as predictors for travel volumes and modal split. This research extends the literature regarding policies that in the context of the case there is also not a “one-size-fits-all” solution and the current mix of policies is not enough to achieve sustainable mobility.

6.1.4 Day-to-day travel behaviour of commuters

In this section, the previous sections all come together. As Heinen et al. (2010) emphasize the lack of research that focus on commuting with cycling as an option, this research contributes to this knowledge gap by incorporating various methods to assess travel behaviour for multiple transport modes including bicycle. Next to the volumes for each separate mode also the modal split over time for commuters is analysed. As Heinen et al. (2011) found evidence that the decision to cycle is for a great part influenced by factors that can be different on a daily basis, this research shows that for travel volumes, office presence and mode shares this is true. Weather conditions, seasons, darkness, day of the week (trip characteristic) and holidays all cause variations in the number of commuters on a day and the mode shares. While this is done for commuting, this research also contributes to the conclusion of Chatterjee et al. (2016) that day-to-day changing factors are understudied for specific trip purposes.

An enhancement to the literature on sustainable mode choices could be made according to Ton et al. (2019) by doing more research in the Netherlands as context, while most of the literature is in contexts where cycling is less common as means of transport. This research uses a case study of commuting behaviour of employees of ASML in Veldhoven, The Netherlands. Next to the sustainable mode choices of bicycle and bus, this research also focuses on the (unsustainable) mode choice car. This is of importance to have a clear overview regarding the case on the one hand to stimulate sustainable mode choices as on the other hand to counter unsustainable travel behaviour.

6.2 Limitations

6.2.1 Case study approach

This research uses a case to get insights into commuting behaviour, which means that findings are not directly generalizable to other contexts. Yet, this research offers valuable insights that can inform decision-makers in similar urban transport contexts. The case-study contributes to an in-depth understanding of commuting behaviour and how behaviour is related to weather conditions, seasonality, and policies. But caution must be exercised when applying the finding from this research to different settings. This is in line with the emphasis in the research of Griffiths et al. (2021) that local context play a critical role in the successfulness of policies. Despite the limited generalizability, the lessons learned from this case study can contribute to a broader policy discussion and future research regarding sustainable commuting.

6.2.2 Data quality

This research faced significant challenges related to data quality. Various reasons contribute to missing data (e.g. inaccurate measurements, incomplete coverage of all locations where commuters arrive or not all commuters being tracked by registration systems, technical issues to cameras, and data processing systems). While enormous effort is put into handling the quality of data by validating, assumptions, and imputation, it introduces bias and uncertainty in the models and corresponding outputs. This is an inherent limitation not only to case studies but to research in general, and while it is extremely rare to encounter 100% clean data, it is still important to be aware of when conducting research. Certainly, also for interpretation and analysis of results.

6.2.3 Complexity of policy interventions

This research faced challenges in isolating the effects of individual policy interventions, due to overlapping impacts on the long-term and the day-to-day variability due to weather conditions, seasonality, and other (external) factors. This complexity means that the observed changes in commuting patterns may not solely be attributable to any single intervention, but rather to a combination of policies. This fits the non-existing "one-size-fits-all" policy in the literature towards sustainable mobility.

6.2.4 Correlation does not imply causation

While significant associations are observed, the data cannot establish direct causality due to the absence of an experimental set up. Unknown external factors and other determinants of mode choice can also have influences on commuting behaviour, as also depicted in the framework synthesized from the literature review. This means results are interpreted in this research as indicative for associations rather than definitive causal relations. This limitation underscore that specific impacts of weather and policy do not imply causation while correlations are identified.

6.2.5 Attitudes and beliefs

Based on the synthesized framework from the literature review, individual characteristics as beliefs and attitudes play a role in commuting behaviour next to weather conditions, trip characteristics and work conditions. However, these factors are difficult to measure and can vary between individuals. This research primarily adopts a quantitative case study approach using aggregated data, which limits the capability of capturing the influence of differences between attitudes and beliefs of persons on commuting behaviour.

6.3 Policy recommendations

As depicted by Holden et al. (2020), people are the key in a shift towards sustainable mobility. However, as emphasized by Santos et al. (2010), policymakers have a crucial role in devising policies that can trigger a change in behaviour. Policy recommendations can be divided in two categories. The first category are recommendations to better assess policies in the future. The second category are recommendations towards sustainable commuting behaviour.

6.3.1 Assessing policies

First of all, in this era of collecting big data and an increasing interest in utilizing data for substantiating and evaluating policy interventions it is crucial to also dedicate attention to maintaining the sources of data and having the underlying assumptions clear. If data has to play a key part, a key part should also be how data is handled for data-driven decision-making. As Ridzuan and Zainon (2019) entail in their paper that is earlier mentioned, the amount of data that is available makes it time-consuming, complex, and prone to errors). The paper calls for potential application of more automated methods with the need of a domain expert for verification and validation of the data. This is in line with the ideology of this research, that for a shift towards sustainable mobility it must be ensured that organizations use good quality data for analyses which potentially lead to better policy evaluation and implementation targeted at a shift towards sustainable mobility. A concrete example from this research is that someone should immediately be notified when a camera is not working. Next to keep an eye on existing measuring instruments, for validity simple additional measuring instruments e.g. physical road sensors can help with assessing quality of measuring instruments in place and used for estimates when other instruments show defects.

Secondly, ex-ante and ex-post evaluations should be in place for commuting policies. Draw up a plan how the effects of intervention can be measured. This includes both quantitative data analysis and e.g. surveys for the perceptions of employees. These assessments should involve both short-term evaluations and long-term evaluations to understand the immediate and sustained effects of policies. In this research for measuring only quantitative data is utilized and to the quantitative side extra attention is needed when policies are implemented. To use the not-working camera example again, after introduction of a policy measure there were substantial observations with at least one of the cameras not working which can give a distorted picture of the effect. Nevertheless, while this research shows that due to day-to-day changing circumstances it is difficult to isolate the effects of policies, a combination with qualitative assessment is needed.

A third way to improve the assessment of policies is by broader collaboration regarding data. Withing organisations between departments, but also with other stakeholders in the area. In local contexts, there is not just one stakeholder that is responsible to shift towards sustainable mobility. Collaborative data-sharing partnerships between employers, transportation companies, and governmental organisations can contribute to better assessments as well as gaining more insights into travel behaviour. A concrete example is that while the data is known of people that have a fare-free card on an aggregate daily level which misses employers without a card or travelling with another type of card. E.g. the bus company can provide additional information by providing insights in the number of total check-outs at certain bus stops with also the distribution of travellers over the day. It must be noted that confidentiality and anonymity must be guaranteed, but cooperation is important to achieve common goals.

A final note for policymakers, is that while the modal split remains a standard measure for sustainable commuting, this research highlights that there is significant day-to-day variability in modal shares and also significant day-to-day variability in volumes. It is advised to not only take the long-term into account, but also be aware of how on short-term changes in the modal split are compared. It is necessary to mention that conditions on measured days can lead to significant differences and be aware that modal split is a separate thing from travel volume and therefore both should be considered.

6.3.2 Towards sustainable commuting behaviour

Building on the findings of this research, several policy recommendations can be made to promote the commuting behaviour of employees in a sustainable way.

As there are congestion records in the Netherlands, a new campaign to stimulate avoiding rush hours called 'Spitsvrij' was introduced in 2025 by the Dutch Government. While from this research it is evident that darkness impacts cycling commuters, and it is more likely to take the car, it is especially useful to implement these kinds of campaigns during the months of the year with darkness during the morning rush hours. Incentivizing commuters to cycle after sunrise by higher allowance, gamification or other kinds of rewards could therefore have potential. Especially during winter. Coming to work later doesn't have to lead to shorter working days. As evident that working from home is here to stay (Barrero et al., 2021), there could be potential in adapting flexible working. This means starting the day at home and ending the day at home by commuting after sunrise and before sunset. On the congestion side, it can also be advantageous to travel before and after rush hours. However, if this is by car this does not mitigate the emissions by car travel

Lobbying for High-Occupancy Vehicle (HOV) lanes is another recommendation to mitigate a shift from cycling to car commuting. Carpooling for commuting is not a common practice for commuting (Molina et al., 2020). Commuters that carpool will contribute to less emissions and less travel volume but still have disadvantages as less flexibility by depending on the other and still have to be on congested roads. For the same case studied in this research, Molin and Kroesen (2023) discovered that in general vanpooling is preferred over carpooling. As Ditmore and Deming (2018) found that vanpoolers had lower stress levels than commuters by car, there is potential for HOVs. As currently these vehicles still have to be on the same congested roads, it is crucial to further stimulate pooling by facilitating HOV lanes. According to Molina et al. (2020) this fulfils both economic and environmental awareness. This fits with the paper that introduced sustainability regarding transport that fulfils its economic and social role while containing the harmful effects of transport on the environment (European Commission, 1992).

Hrelja and Rye (2023) advocated for a policy mix including 'push' and 'pull' measures, this research only assesses 'pull' measures. These are measures that stimulate sustainable transport and are in the case setting solely implemented. Measures that 'push' employees away from car usage by discouraging it are rarely introduced by employers while it is undesired to destimulate employees. Yet, Molin and Kroesen (2023) showed with a stated choice experiment that 'push' measures as paid parking have potential to shift employees away from commuting by car. In the broader picture towards sustainable mobility, 'push' measures should be considered in addition to 'pull' measures.

Especially Fridays and School Holidays see a decrease in commuters, but also an increase in the share of car commuters. Of course, a sustainable modal split is desired on the busies days with the highest travel volumes. Yet a direction for policymakers could be to stimulate sustainable commuting especially on the days with less volume. As Kroesen and Handy (2014) identified a bidirectional relationship between commuting by bike and non-work cycling leading to spill-over effects when introducing a travel allowance by bike, maybe there can also be spill-over effects between days. To further elaborate; if on a day that it is easier to take the car it is stimulated to commute by bike, there could potentially be a spill-over effect on a busy day due to positive experience with cycling to work.

A final policy recommendation is while it is proven that weather and seasonality plays a significant role on commuting, attention is required to the existing facilities for cyclists. Better illuminated cycle paths, roads that drain water better and do not become slippery, a drying service at work for wet clothes, safe cycle routes out of the wind and traffic lights that are more often green for cyclists on days with bad weather are among an inexhaustible list of possibilities.

6.4 Future research

- While in the proposed frameworks, these are also determinants of mode choice (but more focussed on the long-term), a good direction for future research, is to use surveys to assess attitudes towards policies after implementation and potentially add questions to get an understanding of perceptions, beliefs and attitudes of commuters on weather conditions.
- This study uses data on a daily level regarding the number of commuters. Future research could look specifically into rush hours and the differences in the days before and after Daylight Saving Time. The switch between Daylight Saving Time and Standard Time provides a unique opportunity to see the immediate impact of more or less darkness while on other moments the difference in darkness with a day prior is limited to at most a few minutes.
- Next to investigating the last mile behaviour it could also be interesting to see on the same day if weather in the afternoon causes different first mile behaviour for the commute back to home (e.g. coming by bike but travelling back with bus or carpool due to adverse weather in the afternoon). This could then potentially be linked to whether forecasted weather later on the day also has influence on the commuting mode in the morning
- Social Cost-Benefit Analysis could be an interesting future research direction to assess next to effectiveness of the intervention whether there are economic and societal trade-offs.
- Instead of doing research into arriving at the work location investigate where people come from and how the effects of day-to-day changing factors vary for different spatial locations. This could reveal whether commuters from different regions are differently affected by weather, darkness, and policies.
- It is important to further research everyday weather with also a focus on darkness and multiple modes in other contexts, while if this case is broader applicable in more settings, as strong point can be made that darkness is one of the key factors causing congested roads during morning peak hours.

7 Conclusions

The struggle towards sustainable mobility is a complex problem which is high on the political agenda due to the expectations regarding climate change. Commuting to work is a significant contributor to unsustainable transportation which emphasizes the importance of employers to encourage employees to travel less, more efficient, and different. As shifting towards sustainable mobility starts with everyday travel behaviour, day-to-day commuting requires attention with weather and seasonality causing significant fluctuations between means of transport. To investigate day-to-day commuting behaviour and assess policies that try to change travel behaviour, the following research question has been central in this thesis:

“To what extent are weather conditions, seasonality, and implemented policies related to the day-to-day variations in commuting behaviour of ASML employees?“

The overarching research question is answered, by first answering the sub questions. Answering the overarching research question helps with achieving the research objectives. Therefore, the answer to each sub question is depicted in the next section.

7.1 Answering the research questions

Sub question 1: *What are the characteristics of the data regarding weather conditions, seasonality, and travel behaviour of ASML employees?*

Descriptives of the data and an Exploratory Data Analysis showed that temperature, wind speed, and precipitation are not strongly linked to each other. Temperature varies over the seasons with no exceptional warm temperatures measured due to measuring in the morning during rush hour. In one out of three days, there is precipitation present in the morning rush hours and the wind direction is towards the southwest which is beneficial for commuters by bike from Eindhoven.

Seasons have a different impact on different modes, with winter, spring, and autumn popular for car and bus commuting and spring and summer more for commuting by bike. There is darkness present in the morning rush hours up to approximately 80% of the morning rush hours and it is also noticeable that especially the switch to Standard Time causes a full hour of darkness less in the morning peak hours.

There are significantly less commuters on Friday compared to other days with Tuesday and Thursday the most popular working days (noted that the difference is not substantial with Mondays and Wednesdays). Distributions of the number of commuters shows skewness to the left which is due to measuring errors, statutory holidays, and the Friday versus other days. It can also depict that for employees in office and for car commuters there is a maximum capacity often reached. Over time it can be seen that during summer holidays there is a dip in employees in office and slightly less employees in office are registered over time. Car is the most popular with on average 3938 commuters, followed by bicycle commuters (2257 on average). Bus commuters are measured to be 551, however for bus only employers with a business card are tracked.

Sub question 2: *What are the (static) relationships of daily weather conditions and seasonality on the modal split and travel volume of ASML employees?*

As preliminary correlations already indicated, travellers by bike seem to be most dependent on the weather for both the number of cycling commuters as well as the share of the total commuters. The only other significant effects regarding daily weather are that on rainy days more people are commuting by bus, while less people go to office. This seems to indicate a substitution from bicycle to bus which is an alteration less desired than car to bus or bike. Although rain is not related to a desired alteration, it is related to a reduction in travel demand. It could be the case that more people work from home on rainy days. Despite the fact that employees are not free during school holidays and many people will go on holidays outside school holidays to avoid holiday crowds, many of the holidays have a significant reduction in commuters. Regarding the modal split, holidays as well as Fridays increase the average share of car commuters. This is probably due to less employees in office cause more parking availability.

There are significant results following from the regression models and time series analysis, although it should be noted that the data can be improved to more accurately and for more types of employees as it is not known for many people how they got to work. Which is why almost a third of the commuters are in the category depicted as 'other'. Precipitation seems to move people from Bike to car with precipitation not affecting the number of car commuters, however on days with snow way less people take the bike and there is also a huge decrease in the number of commuters by car. This indicates that rain does not matter that much for drivers while the risks of snow are more acknowledged by car commuters. The less good models for car, are in line with research that car commuters are more stubborn and less susceptible for weather conditions and a lot of people have ingrained patterns of car use. Darkness is found to be crucial for day-to-day variations in commuting behaviour.

Sub question 3: *How do polices and the interactions between daily weather and seasons relate with the modal split and travel volume of commuters?*

Modes of transport have different interaction terms that are significant between weather conditions and seasonality for travel volume models. This depicts that some weather conditions have a different effect across seasons on mode choice and therefore, while it is often overlooked in the literature it should be incorporated when possible. Interactions are not investigated in the modal split model due to complexity, but policies are. It could be noticed that the bus frequency increased the odds of taking the bus, but a relative more substantial effect came from the basement shed. This can for example be due to that there is a lagged effect of the implementation of policies, which is not investigated and therefore a limitation of this study. For bicycle coefficients, it is better to not compare directly with baseline, but more with the same situations. For example, when comparing with the introduction of bus frequency one year earlier the odds of taking the bike increased. Trends are however better captured when using the policies in the model, leading to predicted values that follow the same trend as the actual values. This means that additional to weather conditions and seasonality play a role on travel behaviour, interactions between weather and seasonality and policies tend to contribute to better capturing the trends of travel behaviour.

Sub question 4: *How are time related patterns and the implementation of policies associated with commuting patterns over time?*

Over time, it can be seen that the volume of bus commuters increases, car commuters slightly decrease, while commuting by bike depicts the most seasonality with also a slight decrease. There can also be a decrease spotted in the office presence of commuters. Regarding mode shares, the models can resemble the trends of the modal split, with bus commuters taking more of the share, a car share that firstly is increasing but seems to decrease a bit. This goes together unfortunately with a higher share in the 'other' category. This does not have to be an unfortunate thing, but next to people choosing another mode of transport that is not incorporated in this study, this category has also the shortcomings of how the other three modes are measured: if the car cameras do not work on a day a bigger discrepancy between employees in office which is measured in another way and the commuters registered by car can lead to more employees in the category 'other'. Decompositions of time series showed that the weekly pattern, which significantly less commuters on Fridays, can be subtracted from the observed values. The Interrupted Time Series analysis showed that policies potentially have immediate impact next to long-term changes, although some nuance could be placed on this.

Overarching research question: *To what extent are weather conditions, seasonality, and implemented policies related to the day-to-day variations in commuting behaviour of ASML employees?"*

Everyday weather and seasonality are associated with day-to-day variations in travel volumes and mode shares. Next to long-term trends with seasonality, everyday weather can be decisive for travel behaviour as darkness, wind speed, temperature, precipitation, all have impact. Weather conditions are also having different effects in different seasons for different modes. Holidays and Friday are associated with less commuters but also a shift towards unsustainable modes and therefore a modal split that is less desired. Policies are contributing to day-to-day variations in commuting behaviour while models improved by incorporating policies. However, isolating the effect of the policies of weather conditions, seasonality, other determinants of mode choice, and external factors is complex.

7.2 Concluding remarks

Although mentioned in the methodology that researchers have to deal with an important trade-off between scientific quality and conveying the results to policymakers (Handy et al., 2014), some complicated models did not lead to much better results. This indicates that it is better to stick with simpler models although this in a sense can feel contradicting while there is awareness of the flaws of the models. Therefore, as also mentioned as limitation of a case study, the generalizability of the results and decisiveness of magnitudes of associations is limited. Yet, lessons can be learned to further contribute to a shift towards sustainable commuting and mobility for researchers, decision-makers, and individuals.

Especially for commuting as trip purpose, it seems darkness plays a substantial role for commuting volumes and modal split. Decision-makers can use this to incentivize commuters to travel after sunrise, which also causes less congestion and by incentivizing commuting by bike also lead to less emissions. This is an extension on working from home which is here to stay: flexible working by starting and ending the working day at home. For research investigating darkness, which is understudied, in other contexts can contribute to a broader understanding of commuting behaviour during peak hours.

Lobbying for HOV lanes, making cycling in adverse weather conditions more attractive, and creating awareness on days with fewer people in office to take a sustainable mode of transport while there are more parking places available (which can potentially lead to spill-over effects on busier days) can all contribute to more sustainable transport, but policymakers need to be aware that there is no "one-size-fits-all" solution and a mix of policies is needed. Also next to these "pull" measures, also "push" measures should be considered.

Next to implementing policies, assessment of policies is essential. To this extent, not only quantitative but also qualitative assessment is needed. Data quality is important for the quantitative side, so decision-makers must pay attention to data verification and validation. Decision-makers can also survey their commuters ex-ante and post-ante. Not only for decision-makers, but also for researchers qualitatively investigating attitudes and beliefs can contribute to get a better understanding on day-to-day changing travel behaviour and the impact of policies.

A final remark is to join forces between different employers, transport companies, individuals, and governmental organizations. As sustainable mobility is a concept that is introduced three decades ago and a lot of research contributed towards sustainable mobility. It requires coordinated efforts. By aligning objectives and allocating resources, the way can be paved to a more sustainable future with less travel, more sustainable travel, and more efficient travel.

Bibliography

An, Q., Fu, X., Huang, D., Cheng, Q., & Liu, Z. (2020). Analysis of adding-runs strategy for peak-hour regular bus services. Transportation Research Part E: Logistics and Transportation Review, 143, 102100. https://doi.org/10.1016/j.tre.2020.102100

ASML. (2024). *Access & Mobility: The Netherlands.* myASML. https://myasml.asml.com/sites/locations/netherlands/access-mobility

Ayinde, K., Lukman, A. F., & Arowolo, O. (2015). Robust regression diagnostics of influential observations in linear regression model. Open Journal of Statistics, 5(04), 273. https://doi.org/10.4236/ojs.2015.54029

Azur, M. J., Stuart, E. A., Frangakis, C., & Leaf, P. J. (2011). Multiple imputation by chained equations: what is it and how does it work?. International journal of methods in psychiatric research, 20(1), 40-49. https://doi.org/10.1002/mpr.329

Banerjee, A., Łukawska, M., Jensen, A. F., & Haustein, S. (2021). Facilitating bicycle commuting beyond short distances: insights from existing literature. Transport reviews, 42(4), 526-550. https://doi.org/10.1080/01441647.2021.2004261

Banister, D. (2008). The sustainable mobility paradigm. Transport policy, 15(2), 73-80. https://doi.org/10.1016/j.tranpol.2007.10.005

Barrero, J. M., Bloom, N., & Davis, S. J. (2021). Why working from home will stick (No. w28731). National Bureau of Economic Research. https://doi.org/10.3386/w28731

Berger, G., Feindt, P. H., Holden, E., & Rubik, F. (2014). Sustainable mobility—challenges for a complex transition. Journal of Environmental Policy & Planning, 16(3), 303-320. https://doi.org/10.1080/1523908x.2014.954077

Berrebi, S. J., Joshi, S., & Watkins, K. E. (2021). On bus ridership and frequency. Transportation Research Part A: Policy and Practice, 148, 140-154. https://doi.org/10.1016/j.tra.2021.03.005

Böcker, L., Dijst, M., & Faber, J. (2016). Weather, transport mode choices and emotional travel experiences. Transportation Research Part A: Policy and Practice, 94, 360-373. https://doi.org/10.1016/j.tra.2016.09.021

Böcker, L., Dijst, M., & Prillwitz, J. (2013). Impact of everyday weather on individual daily travel behaviours in perspective: a literature review. Transport reviews, 33(1), 71-91. https://doi.org/10.1080/01441647.2012.747114

Bull, O., Muñoz, J. C., & Silva, H. E. (2021). The impact of fare-free public transport on travel behavior: Evidence from a randomized controlled trial. Regional Science and Urban Economics, 86, 103616. https://doi.org/10.1016/j.regsciurbeco.2020.103616

Burkacky, O., Deichmann, J., Pfingstag, P., & Werra, J. (2022). Semiconductor shortage: How the automotive industry can succeed. McKinsey & Company. https://doi.org/10.1016/j.tranpol.2021.04.007

Busch-Geertsema, A., Lanzendorf, M., & Klinner, N. (2021). Making public transport irresistible? The introduction of a free public transport ticket for state employees and its effects on mode use. Transport policy, 106, 249-261. https://doi.org/10.1016/j.tranpol.2021.04.007

Cameron, A. C., & Trivedi, P. K. (2013). *Regression analysis of count data* (No. 53). Cambridge university press. https://doi.org/10.1017/cbo9780511814365.005

Cantwell, M., Caulfield, B., & O'Mahony, M. (2009). Examining the factors that impact public transport commuting satisfaction. Journal of public transportation, 12(2), 1-21. https://doi.org/10.5038/2375-0901.12.2.1

Chatterjee, K., Clark, B., & Bartle, C. (2016). Commute mode choice dynamics: Accounting for day-to-day variability in longer term change. European Journal of Transport and Infrastructure Research, 16(4). https://doi.org/10.18757/ejtir.2016.16.4.3167

Chatterjee, S., & Hadi, A. S. (2015). Regression analysis by example. John Wiley & Sons. https://doi.org/10.1002/0470055464

Cook, R. D. (1977). Detection of Influential Observation in Linear Regression. Technometrics, 19(1), 15–18. https://doi.org/10.2307/1268249

Creswell, J. W. (2009). Research design: Qualitative, quantitative, and mixed methods approaches (3rd ed.). Sage Publications, Inc.

de Haas, M., Kroesen, M., Chorus, C., Hoogendoorn-Lanser, S., & Hoogendoorn, S. (2022). E-bike user groups and substitution effects: evidence from longitudinal travel data in the Netherlands. Transportation, 49(3), 815-840. https://doi.org/10.1007/s11116-021-10195-3

Ditmore, C. J., & Deming, D. M. (2018). Vanpooling and its effect on commuter stress. Research in Transportation Business & Management, 27, 98-106. https://doi.org/10.1016/j.rtbm.2018.11.001

Ebrahimi Kalan, M., Jebai, R., Zarafshan, E., & Bursac, Z. (2021). Distinction between two statistical terms: multivariable and multivariate logistic regression. Nicotine and Tobacco Research, 23(8), 1446-1447. https://doi.org/10.1093/ntr/ntaa055

Eriksson, L., Friman, M., & Gärling, T. (2008). Stated reasons for reducing work-commute by car. Transportation Research Part F: Traffic Psychology and Behaviour, 11(6), 427-433. https://doi.org/10.1016/j.trf.2008.04.001

European Commission. (1992). GREEN PAPER on the impact of Transport on the Environment - A Community strategy for "sustainable mobility". COM(92) 46 final. Publications Office of the European Union. https://op.europa.eu/en/publication-detail/-/publication/98dc7e2c-6a66-483a-875e-87648c1d75c8/language-en

Faber, R. M., Jonkeren, O., de Haas, M. C., Molin, E. J. E., & Kroesen, M. (2022). Inferring modality styles by revealing mode choice heterogeneity in response to weather conditions. Transportation research part A: policy and practice, 162, 282-295. https://doi.org/10.1016/j.tra.2022.06.003

Franssens, S., Botchway, E., De Swart, W., & Dewitte, S. (2021). Nudging commuters to increase public transport use: A field experiment in Rotterdam. Frontiers in psychology, 12, 633865. https://doi.org/10.3389/fpsyg.2021.633865

Gallo, M., & Marinelli, M. (2020). Sustainable mobility: A review of possible actions and policies. Sustainability, 12(18), 7499. https://doi.org/10.3390/su12187499

Gössling, S., & Cohen, S. (2014). Why sustainable transport policies will fail: EU climate policy in the light of transport taboos. *Journal of Transport Geography*, 39, 197-207. https://doi.org/10.1016/j.jtrangeo.2014.07.010

Gössling, S., Neger, C., Steiger, R., & Bell, R. (2023). Weather, climate change, and transport: a review. Natural Hazards, 118(2), 1341-1360. https://doi.org/10.1007/s11069-023-06054-2

Government of the Netherlands. (2020). *Climate policy*. Climate Change | Government.nl. https://www.government.nl/topics/climate-change/climate-policy

Griffiths, S., Del Rio, D. F., & Sovacool, B. (2021). Policy mixes to achieve sustainable mobility after the COVID-19 crisis. Renewable and Sustainable Energy Reviews, 143, 110919. https://doi.org/10.1016/j.rser.2021.110919

Handy, S., Van Wee, B., & Kroesen, M. (2014). Promoting cycling for transport: research needs and challenges. Transport reviews, 34(1), 4-24. https://doi.org/10.1080/01441647.2013.860204

Haustein, S., Kroesen, M., & Mulalic, I. (2020). Cycling culture and socialisation: modelling the effect of immigrant origin on cycling in Denmark and the Netherlands. Transportation, 47(4), 1689-1709. https://doi.org/10.1007/s11116-019-09978-6

Heinen, E., & Handy, S. (2012). Similarities in attitudes and norms and the effect on bicycle commuting: Evidence from the bicycle cities Davis and Delft. International journal of sustainable transportation, 6(5), 257-281. https://doi.org/10.1080/15568318.2011.593695

Heinen, E., Maat, K., & Van Wee, B. (2011). Day-to-day choice to commute or not by bicycle. Transportation Research Record, 2230(1), 9-18. https://doi.org/10.3141/2230-02

Heinen, E., Maat, K., & Van Wee, B. (2013). The effect of work-related factors on the bicycle commute mode choice in the Netherlands. Transportation, 40, 23-43. https://doi.org/10.1007/s11116-012-9399-4

Heinen, E., Van Wee, B., & Maat, K. (2010). Commuting by bicycle: an overview of the literature. Transport reviews, 30(1), 59-96. https://doi.org/10.1080/01441640903187001

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., Thépaut, J-N. (2023). ERA5 hourly data on single levels from 1940 to present [Data set]. ECMWF. https://doi.org/10.24381/cds.adbb2d47

Hidalgo, B., & Goodman, M. (2013). Multivariate or multivariable regression?. American journal of public health, 103(1), 39-40. https://doi.org/10.2105/ajph.2012.300897

Holden, E., Banister, D., Gössling, S., Gilpin, G., & Linnerud, K. (2020). Grand Narratives for sustainable mobility: A conceptual review. Energy Research & Social Science, 65, 101454. https://doi.org/10.1016/j.erss.2020.101454

Hox, J. J., & Boeije, H. R. (2005). Data collection, primary vs. secondary. Encyclopedia of social measurement, 1(1), 593-599. https://doi.org/10.1016/b0-12-369398-5/00041-4

Hrelja, R., & Rye, T. (2023). Decreasing the share of travel by car. Strategies for implementing 'push'or 'pull'measures in a traditionally car-centric transport and land use planning. International Journal of Sustainable Transportation, 17(5), 446-458. https://doi.org/10.1080/15568318.2022.2051098

Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for R. Journal of statistical software, 27, 1-22. https://doi.org/10.18637/jss.v027.i03

International Energy Agency [IEA]. (n.d.). Transport. https://www.iea.org/energy-system/transport

Johnston, M. P. (2014). Secondary data analysis: A method of which the time has come. Qualitative and quantitative methods in libraries, 3(3), 619-626.

Koetse, M. J., & Rietveld, P. (2009). The impact of climate change and weather on transport: An overview of empirical findings. Transportation Research Part D: Transport and Environment, 14(3), 205-221. https://doi.org/10.1016/j.trd.2008.12.004

Kroesen, M., & Handy, S. (2014). The relation between bicycle commuting and non-work cycling: results from a mobility panel. Transportation, 41, 507-527. https://doi.org/10.1007/s11116-013-9491-4

Linda, S. T. E. G. (2003). Can public transport compete with the private car?. IATSS research, 27(2), 27-35. https://doi.org/10.1016/s0386-1112(14)60141-2

Liu, C., Susilo, Y. O., & Karlström, A. (2017). Weather variability and travel behaviour–what we know and what we do not know. Transport reviews, 37(6), 715-741. https://doi.org/10.1080/01441647.2017.1293188

Ma, X., Yuan, Y., Van Oort, N., & Hoogendoorn, S. (2020). Bike-sharing systems' impact on modal shift: A case study in Delft, the Netherlands. Journal of cleaner production, 259, 120846. https://doi.org/10.1016/j.jclepro.2020.120846

Molin, E. J., & Kroesen, M. (2023). The mobility transition at ASML: a quantitative study of the effects of mobility measures on employees' commute mode choice. TU Delft.

Molina, J. A., Giménez-Nadal, J. I., & Velilla, J. (2020). Sustainable commuting: Results from a social approach and international evidence on carpooling. Sustainability, 12(22), 9587. https://doi.org/10.3390/su12229587

MuConsult. (2019). Financiële prikkels om fietsen naar het werk te stimuleren. https://muconsult.nl/wpcontent/uploads/2019/11/Rapport-Fiets.pdf

Muñoz Sabater, J. (2019). ERA5-Land hourly data from 2001 to present [Data set]. ECMWF. https://doi.org/10.24381/CDS.E2161BAC

Nematchoua, M., Deuse, C., Cools, M., & Reiter, S. (2020). Evaluation of the potential of classic and electric bicycle commuting as an impetus for the transition towards environmentally sustainable cities: A case study of the university campuses in Liege, Belgium. Renewable and Sustainable Energy Reviews, 119, 109544. https://doi.org/10.1016/j.rser.2019.109544

Netherlands Enterprise Agency [RVO]. (2024). Reporting obligation work-related mobility of persons – WPM. https://english.rvo.nl/topics/wpm

Newey, W. K., & West, K. D. (1986). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. National Bureau of Economic Research. https://doi.org/10.3386/t0055

O'Neil, C., & Schutt, R. (2013). Doing data science: Straight talk from the frontline. " O'Reilly Media, Inc.".

Paris Agreement. (2016). Official Journal, L 282, 4-18. CELEX: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:22016A1019(01)[legislation]

Park, Y. S., Konge, L., & Artino Jr, A. R. (2020). The positivism paradigm of research. Academic medicine, 95(5), 690-694. https://doi.org/10.1097/acm.0000000000003093

R Core Team (2024). _R: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

Raux, C., Ma, T. Y., & Cornelis, E. (2016). Variability in daily activity-travel patterns: the case of a one-week travel diary. European transport research review, 8, 1-14. https://doi.org/10.1007/s12544-016-0213-9

Ridzuan, F., & Zainon, W. M. N. W. (2019). A review on data cleansing methods for big data. Procedia Computer Science, 161, 731-738. https://doi.org/10.1016/j.procs.2019.11.177

Rubin, D. B. (1987). Multiple imputation for nonresponse in surveys. John Wiley & Sons. https://doi.org/10.1002/9780470316696

Sabir, M. (2011). Weather and travel behaviour.

Sabir, M., van Ommeren, J., Koetse, M. J., & Rietveld, P. (2010). Impact of weather on daily travel demand. Work Pap Dep Spat Econ VU Univ Amsterdam.

Santos, G., Behrendt, H., & Teytelboym, A. (2010). Part II: Policy instruments for sustainable road transport. Research in transportation economics, 28(1), 46-91. https://doi.org/10.1016/j.retrec.2010.03.002

Schimanke S., Ridal M., Le Moigne P., Berggren L., Undén P., Randriamampianina R., Andrea U., Bazile E., Bertelsen A., Brousseau P., Dahlgren P., Edvinsson L., El Said A., Glinton M., Hopsch S., Isaksson L., Mladek R., Olsson E., Verrelle A., Wang Z.Q. (2021). CERRA sub-daily regional reanalysis data for Europe on single levels from 1984 to present [Data set]. ECMWF. https://doi.org/10.24381/CDS.622A565A

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. Journal of business research, 104, 333-339. https://doi.org/10.1016/j.jbusres.2019.07.039

Ton, D., Duives, D. C., Cats, O., Hoogendoorn-Lanser, S., & Hoogendoorn, S. P. (2019). Cycling or walking? Determinants of mode choice in the Netherlands. Transportation research part A: policy and practice, 123, 7-23. https://doi.org/10.1016/j.tra.2018.08.023

Tukey, J.W. (1970) Exploratory Data Analysis. Limited Preliminary Edition, Addison-Wesley, Reading. https://doi.org/10.1007/bf02295986

United Nations [UN]. (2024). Goal 11: Make cities inclusive, safe, resilient and sustainable. https://www.un.org/sustainabledevelopment/cities/

Van Buuren, S. (2018). Flexible imputation of missing data. CRC press. https://doi.org/10.1201/9780429492259

Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. Journal of statistical software, 45, 1-67. https://doi.org/10.18637/jss.v045.i03

Witlox, F., & Tindemans, H. (2004). The application of rough sets analysis in activity-based modelling. Opportunities and constraints. Expert Systems with Applications, 27(4), 585-592. https://doi.org/10.1016/j.eswa.2004.06.003

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

Yin, R. K. (2018). *Case study research and applications* (6th ed.). Thousand Oaks: SAGE Publications.

Zeiske, N., van der Werff, E., & Steg, L. (2021). The effects of a financial incentive on motives and intentions to commute to work with public transport in the short and long term. Journal of Environmental Psychology, 78, 101718. https://doi.org/10.1016/j.jenvp.2021.101718

Zippenfenig, P. (2023). Open-Meteo.com Weather API [Computer software]. Zenodo. https://doi.org/10.5281/ZENODO.7970649

Appendices

A Literature review process

The foundation of all academic research activities, regardless of discipline, is based on and related to existing knowledge (Snyder, 2019). Therefore, a thorough literature review is essential for identifying existing knowledge gaps in the existing literature and defining the scope of the study. To find sources for this literature review, the web search engine Google Scholar was used. Google Scholar makes it easy to conduct a broad search for scholarly literature and with Google Scholar one can search through a wide range of disciplines and sources. Two prompts are used that align with the research aim:

Prompt 1: *"weather conditions" AND ("normal weather" OR "extreme weather" OR "daily weather") AND seasons AND ("travel behavior" OR "travel behaviour") AND ("mode choice" OR "modal split") AND ("transport mode" OR "mode of transport") AND determinants AND relationships AND Netherlands*

Prompt 2: *("sustainable transport" OR "sustainable mobility") AND ("policy interventions" OR "policy recommendations") AND ("policy objective" OR "policy objectives" OR "policy direction") AND ("travel behaviour" OR "travel behavior") AND (incentives OR measures OR pricing) AND ("decision-making" OR implementation) AND ("Netherlands" OR "Europe") AND (government OR governance OR management) AND ("modal shift" OR "mode shift" OR "shift towards")*

To further narrow down to a final set of papers, the papers needed to be on the first two pages and needed to have 10 or more citations. For both prompts, also the highest relevant review article is included and the most cited article after 2021 for also a very recent view on the literature. Also, the snowball method is applied on some of the found literature. This entails consulting publications that the target paper consulted (backward snowballing) and consulting publications that used the target paper (forward snowballing).

B R packages and code

The code is documented, due to extensiveness into multiple R Markdown (.Rmd files). These files are stored at ASML as well as the data used in this thesis. An extensive description of how to use the code and re-run the analysis is known at ASML.

This appendix mentions the packages that are used in R Studio. The citations that are used, are not in APA while the citations are used which the authors of the package want users to use. These are retrievable by using the function *citation* in RStudio for each used package (e.g. *citation("readxl")*). The citations are included while the R core development team invested a lot of time and effort in creating the programming language R and ask to give credit when used for data analysis.

- *readxl*: Importing excel files into R

Wickham H, Bryan J (2023). _readxl: Read Excel Files_. R package version 1.4.3, <https://CRAN.R-project.org/package=readxl>.

- *ggplot2*: Package based on "The Grammar of Graphics" that helps creating visualisations

H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

- *dplyr*: Package used for data manipulation and data cleaning

Wickham H, François R, Henry L, Müller K, Vaughan D (2023). _dplyr: A Grammar of Data Manipulation_. R package version 1.1.4, <https://CRAN.R-project.org/package=dplyr>.

- *tidyr*: Package with tools to make data tidier and allows for example for pivoting.

Wickham H, Vaughan D, Girlich M (2024). _tidyr: Tidy Messy Data_. R package version 1.3.1, <https://CRAN.R-project.org/package=tidyr>.

- *lubridate*: For manipulating and parsing dates

Garrett Grolemund, Hadley Wickham (2011). Dates and Times Made Easy with lubridate. Journal of Statistical Software, 40(3), 1-25. URL https://www.jstatsoft.org/v40/i03/.

- *hms*: Parsing periods for hours, minutes, and seconds

Müller K (2023). _hms: Pretty Time of Day_. R package version 1.1.3, <https://CRAN.R-project.org/package=hms>.

- *car*: "Companion to Applied Regression" for Variance Inflation Factors

Fox J, Weisberg S (2019). _An R Companion to Applied Regression_, Third edition. Sage, Thousand Oaks CA. <https://www.john-fox.ca/Companion/>.

- *corrplot*: Visual tool for correlation plotting

Taiyun Wei and Viliam Simko (2024). R package 'corrplot': Visualization of a Correlation Matrix (Version 0.95). Available from https://github.com/taiyun/corrplot

- *modelsummary*: For summarizing data and statistical models

Arel-Bundock V (2022). "modelsummary: Data and Model Summaries in R." _Journal of Statistical Software_, *103*(1), 1-23. doi:10.18637/jss.v103.i01 <https://doi.org/10.18637/jss.v103.i01>.

- *mice*: For Multivariate Imputation by chained Equations and pooling

Stef van Buuren, Karin Groothuis-Oudshoorn (2011). mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software, 45(3), 1-67. DOI 10.18637/jss.v045.i03.

- *broom*: Helping with tidying up messy data

Robinson D, Hayes A, Couch S (2024). _broom: Convert Statistical Objects into Tidy Tibbles_. R package version 1.0.7, <https://CRAN.R-project.org/package=broom>.

- *sandwich*: For robust covariance matrix (HAC)

Zeileis A (2004). “Econometric Computing with HC and HAC Covariance Matrix Estimators.” _Journal of Statistical Software_, *11*(10), 1-17. doi:10.18637/jss.v011.i10 <https://doi.org/10.18637/jss.v011.i10>.

- *lmtest*: For checking diagnostics of linear regresion

Achim Zeileis, Torsten Hothorn (2002). Diagnostic Checking in Regression Relationships. R News 2(3), 7-10. URL https://CRAN.R-project.org/doc/Rnews/

- *performance*: Visually checking the model assumptions

Lüdecke et al., (2021). performance: An R Package for Assessment, Comparison and Testing of Statistical Models. Journal of Open Source Software, 6(60), 3139. https://doi.org/10.21105/joss.03139

- *nnet*: Fitting a Multinomial Logistic Regression

Venables, W. N. & Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0

- *DescTools*: Winsorizing (which was not successful for the data at hand)

Signorell A (2025). _DescTools: Tools for Descriptive Statistics_. R package version 0.99.59, <https://CRAN.R-project.org/package=DescTools>.

- *forcast*: Ysed for SARIMAX model

Hyndman RJ, Khandakar Y (2008). “Automatic time series forecasting: the forecast package for R.” _Journal of Statistical Software_, *27*(3), 1-22. doi:10.18637/jss.v027.i03 <https://doi.org/10.18637/jss.v027.i03>.

- *tseries*: Used for decomposition of times series and making time series objects.

Trapletti A, Hornik K (2024). _tseries: Time Series Analysis and Computational Finance_. R package version 0.10-58, <https://CRAN.R-project.org/package=tseries>.

C Data preparation process

The are of interest is 'de Run 6000', which is the location of the headquarters of ASML. The reason to focus on this location is while ..% of the employees are located here. Another important location to consider is the nearby location 'de Run 1000', which facilitates place for more than employees. However, this location opened in 2024 and the required data is not yet collected well enough at this location (e.g. at the moment of writing there are no bicycle counting camera's). In Figure C1, the map of the location is presented:

Figure C1: Map

Three modalities are of interest in this research: car, bicycle, and public transport. The number of bicycles is registered by counting cameras. These cameras are located at five different places. While the research focusses on 'de Run 6000', specific bus stops are included for which check-outs with the NS-Business card are known and are most likely to have 'de Run 6000' as destination.

Weather

Daylight Saving Time in the datasets. In the historical dataset, sunrise is differing day to day nothing to a few minutes, but in the Netherlands:

March 26, 2023, +1, October 29, 2023, -1, March 31, 2024, +1, October 27, 2024, -1.

We get to curves in the data, but while working hours do not change with the Daylight Saving Time, we want to incorporate this DST to see whether it has immediate effect on the modal split.

Comparison of counts:

Public Transport: counts check-outs at certain bus stops with ASML Business cards. These Business card can be in possession by all ASML regular employees in the Netherlands and by interns that are not eligible for a student week- or weekend travel card (OV-chipcard). This is around 80% of the total employees. This means checkouts from interns with an OV-chipcard and checkouts from employees with a contract from an external supplier are not included. Hermes, a Dutch transport company that provides urban and regional transport in the Eindhoven region, can provide data of the number of check-outs at each bus station. A disadvantage of using this data is that there are also for example people check out for the hospital next to ASML, which leads to an overestimation of the number of users with ASML as their destination. The data used in this study, has therefore an underestimation of the total commuters while it does not include all commuters by bus to ASML; approximately 80% of the employees are entitled to a NS Business card.

Bicycle: Counting cameras at various locations count the number of cyclists that pass by. There is a camera for entering and leaving parking buildings. It is possible that the camera occasionally misses a cyclist or that an unauthorized is registered who cycles across the ASML site.

Car: On campus, there are parking facilities for employees, visitors, contractors, and suppliers. In the data, parking events from nine subgroups are present. The parking events are measured as employees that checked in with their badge and stayed for at least 30 minutes at the parking lot. A shortcoming of this appears to be that sometimes the barriers are open.

Total employees: Employees have a location linked to their profile. For example, De Run 6000. However, people are not always working on the same location. In the dataset, the employees in office per day are used. This is an aggregated number of all employee subgroups and is measured by the number of badge swipes at the buildings at De Run 6000. Therefore, it could include people that normally work on other places.

The following table gives an overview of which employee subgroups are present in the data for each means of transport. The number in the employees column, is the total number of employees registered at De Run 6000 on the 29th of November 2024, which is the last day used in the dataset.

Table C1: Employees numbers

Employee subgroup	Car	Public Transport	Bicycle	Employees
N1 – Flex	X			550
N4 – Outsourcing onsite	X			4445
N6 – Consultant	X			46
N7 – Students non-Payroll	X			111
P1 – Employment	X	X		14823
P2 – Employment limited	X	X		1328
P3 – Intern	X	X		243
P4 – Expatriate	X	X		19
Unknown	X		X	?

It was not possible, with the released data, to filter for all three means of transport on the same subgroups or to properly break down the number of cyclists on each group while ratios of the number of people per subgroup who came to the office can also differ per day. By including as many categories as possible, it provides the closest to the actual number as possible with the data obtained. Ultimately it is about a safe and accessible campus and by leaving out subgroups it can give a distorted picture. Therefore, it is also important to acknowledge that public transport and car probably have an underestimation and cyclists are overestimated.

Train to Eindhoven Centraal

For this data, also weekday is added to make it possible to filter out the weekends.

The data missed two dates: 04/10/2023 and 01/01/2024. The data consists of 168 origin stations, one called 'blank' for which the origin is unknown. These are contributing 0.41% to the total trips. One of the stations of origin is Eindhoven Centraal itself 22.43%. There are various reasons according to the descriptions in the dataset, e.g. a travel without a check-out corrected by NS, but almost all are categorized as journey from Eindhoven Centraal to Eindhoven Centraal. A reason can be crossing the station to avoid having to walk around. As the main interest is in journeys from other places, these counts with origin Eindhoven Centraal are subtracted. As with the data related to the bus trips, there is a sudden increase from April 2nd, 2024. Again, the PowerBI dashboard is used for validation. PowerBI (discontinued in September 2024). From June 2024 onwards, the numbers match often and the average difference is less than 0.7%. May 2024: the original data has 21.49% more trips and in April 2024 1.86% less trips. The reason could not be determined, but while from June onwards the numbers are almost the same and therefore it can be validated that it was a good idea to subtract the trips with Eindhoven Centraal as origin. To deal with inconclusiveness about April and May 2024, it is chosen to use the original data while this data provides more information about origin stations. Before April 2024, the original data has an average trip count of 21.4 while afterwards the average trip count for each day is 405.36. The overlapping data is 3.98% higher on average for the original data. While it is inconclusive what causes the fluctuations between the overlapping dates, for the period previous to April 2024 the PowerBI data is multiplied with 1.04.

After removing the counts with Eindhoven Central as origin, the following table depicts the five origin train stations with the most counts

Table C2: Top 5 stations

Origin	Percentage (%) 86427
Utrecht Centraal 11074	12.81
's-Hertogenbosch 10318	11.94
Rotterdam Centraal 5823	6.74
Tilburg 4539	5.25
Helmond 4379	5.07
Total of top 5 36133	41.81

D Exploratory Data Analysis supplements

After reducing redundant variables regarding weather and highly correlated variables, Figure D1 depicts the correlation matrix of weather conditions. The correlations between variables are less high than the suggestion in the literature review that weather variables are often highly correlated.

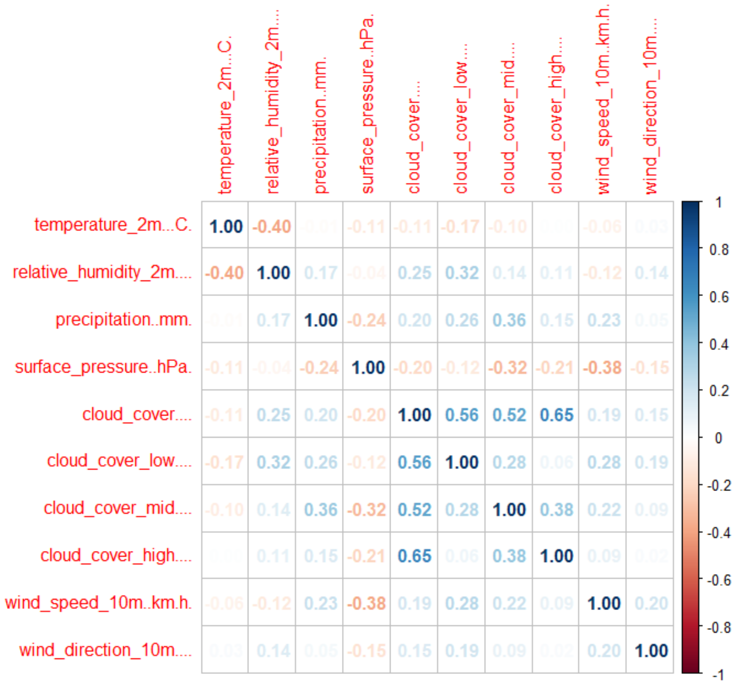


Figure D1: Correlogram various weather conditions

When focusing on weather conditions that are used in the models, also a correlogram can be made with these weather conditions but compared how the correlations are between hours and daily averages. This is depicted in Figure D2.

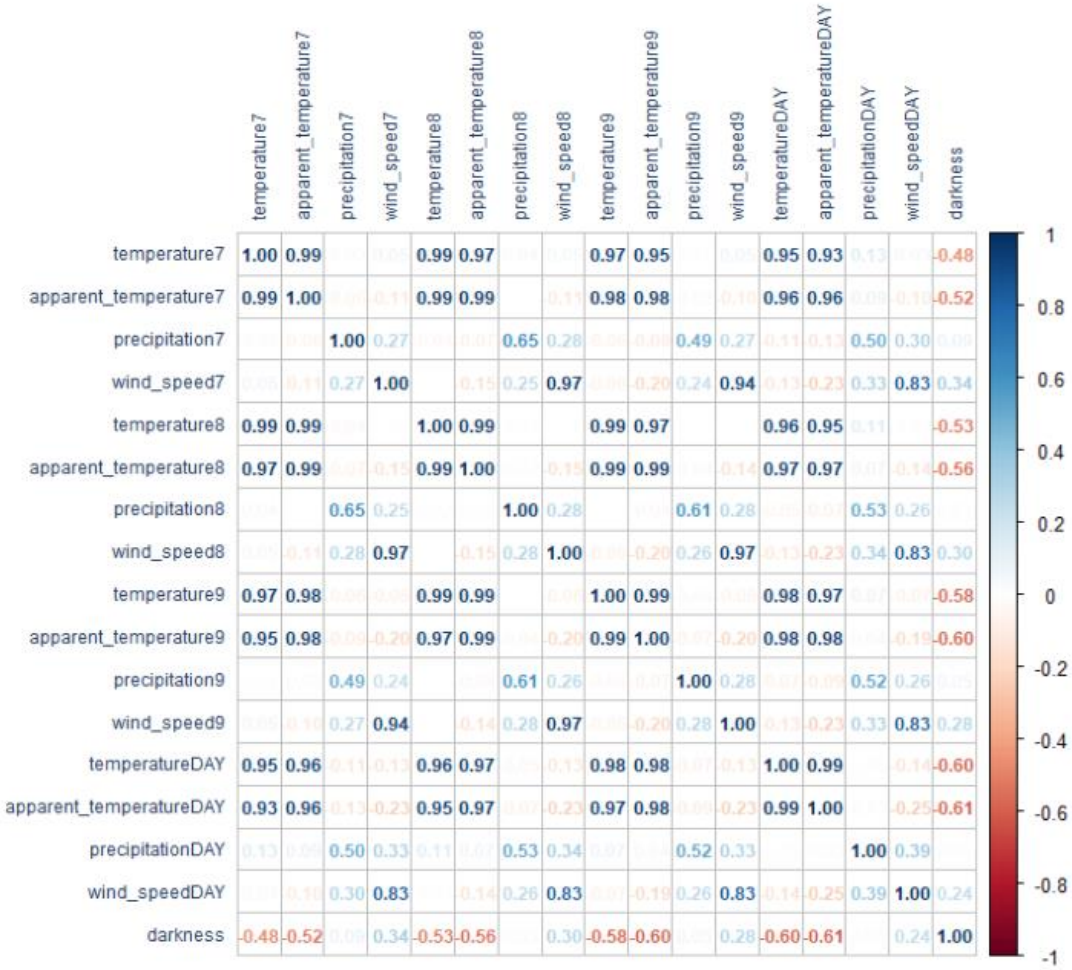


Figure D2: Correlogram main weather conditions over the day

The following correlograms can help with seeing the correlations between individual weather conditions.

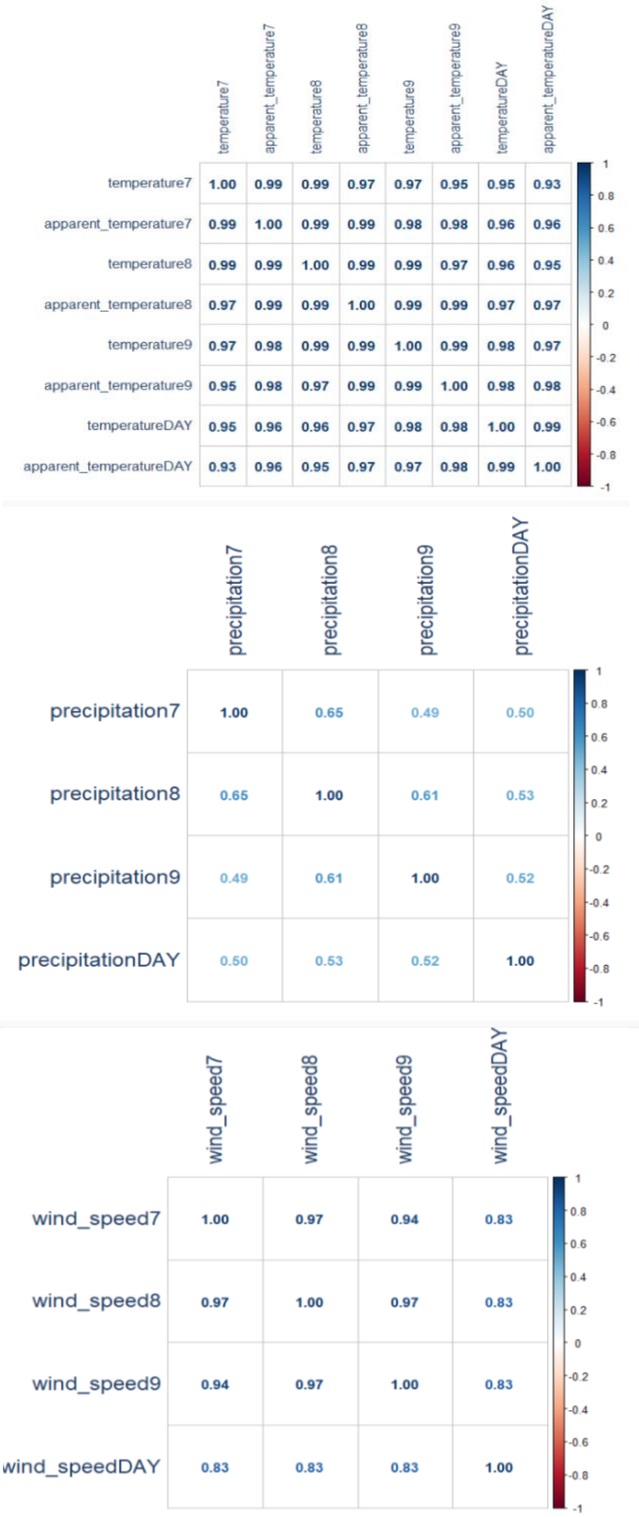


Figure D3: Correlograms main weather conditions over the day, decomposed

E Raw model outputs

In the results chapter, the final models are depicted. However, multiple models are considered and with for example ANOVA tests differences can be shown between basic and advanced models. In this appendix the raw outputs of other models, the chosen models, and intermediate tests have been added in raw format. Not all tested models are depicted, but a selection that allows do see differences between with and without influential points and full and sparse advanced models.

E.1 Bicycle

The bicycle outputs are different formatted in R in comparison to car, bus, and employees in office. This is because the results are pooled and separate code was written to retrieve model fit statistics.

Output 1: Basic model before dealing with outliers (includes statutory holidays)

```
                                 term     estimate  std.error   statistic        df       p.value
1                         (Intercept)  3459.006723 107.350466  32.2216276 411.91917 1.226208e-114
2                     temperature_cen    18.911672   6.260503   3.0207912 317.60019  2.725882e-03
3                            wind_cen    -4.866663   4.138643  -1.1759080 324.82806  2.404928e-01
4                            darkness    -9.774906   1.439210  -6.7918574 393.31625  4.103550e-11
5   precipitation_binaryPrecipitation  -490.467583  56.394832  -8.6970306 400.54502  8.897628e-17
6                     snow_binarySnow  -207.232429 143.169616  -1.4474610 402.93798  1.485454e-01
7                      WeekdayTuesday   148.119253  68.762507   2.1540700 416.33883  3.180793e-02
8                    WeekdayWednesday   -54.768749  70.500501  -0.7768562 363.19064  4.377493e-01
9                     WeekdayThursday   -54.031022  68.896683  -0.7842326 405.39107  4.333617e-01
10                      WeekdayFriday -1036.009992  69.633228 -14.8780980 385.02148  7.099994e-40
11         School_holidayAutumn break  -178.475970 152.207883  -1.1725803 415.45020  2.416360e-01
12      School_holidayChristmas break  -865.130637 190.224768  -4.5479390 304.11289  7.833749e-06
13            School_holidayMay break  -306.114664 179.899025  -1.7015916  68.41179  9.337295e-02
14         School_holidaySpring break  -545.900059 213.738708  -2.5540533 410.46873  1.100798e-02
15         School_holidaySummer break  -775.002845  79.641119  -9.7311898 415.92504  2.676311e-20
16                       seasonSpring  -424.823669 108.030841  -3.9324295 372.42447  1.002862e-04
17                       seasonSummer  -199.950527 117.408894  -1.7030271 402.45382  8.933506e-02
18                       seasonAutumn  -138.055803  83.925050  -1.6449892 371.68301  1.008174e-01
19             Statutory_holidayTRUE -2075.566184 188.597097 -11.0052923  50.44017  5.162230e-15
     Model        r2      adjr2      resSE     F_stat       df_l        df_r
 5.5000000 0.7309720  0.7194696 450.0267157 63.6285561 18.0000000 421.0000000
```

Output 2: basic model after dealing with outliers (excludes statutory holiday)

```
                                 term     estimate  std.error   statistic        df       p.value
1                         (Intercept)  3527.535905  90.727491  38.8805628 358.86797 1.021142e-130
2                     temperature_cen    22.875292   5.235074   4.3696215 272.89547  1.770351e-05
3                            wind_cen    -5.294523   3.426532  -1.5451549 331.57059  1.232627e-01
4                            darkness    -9.474107   1.223977  -7.7404300 328.92665  1.230913e-13
5   precipitation_binaryPrecipitation  -479.998036  48.325333  -9.9326379 358.39547  1.069386e-20
6                     snow_binarySnow  -353.235983 131.020323  -2.6960396 355.31204  7.350818e-03
7                      WeekdayTuesday   116.346086  57.513695   2.0229284 368.80772  4.380159e-02
8                    WeekdayWednesday   -87.396284  59.047889  -1.4800916 268.75078  1.400201e-01
9                     WeekdayThursday   -11.752109  59.505016  -0.1974978 334.06248  8.435581e-01
10                      WeekdayFriday -1092.722010  57.724375 -18.9299931 350.37610  1.510443e-55
11         School_holidayAutumn break  -250.942650 121.628222  -2.0631943 379.94087  3.977306e-02
12      School_holidayChristmas break  -828.111813 210.279333  -3.9381512 254.02545  1.061205e-04
13            School_holidayMay break  -352.936501 156.124989  -2.2606023  40.81189  2.918344e-02
14         School_holidaySpring break  -520.454675 189.917925  -2.7404189 367.45058  6.435634e-03
15         School_holidaySummer break  -767.701700  64.208417 -11.9564029 380.52583  3.432982e-28
16                       seasonSpring  -435.719063  91.526726  -4.7605665 313.48174  2.953429e-06
17                       seasonSummer  -262.271739  97.694187  -2.6846197 360.84759  7.596183e-03
18                       seasonAutumn  -140.251785  70.182398  -1.9983897 303.91321  4.656505e-02
     Model        r2      adjr2      resSE     F_stat       df_l        df_r
 5.5000000 0.7715968  0.7615636 357.5828155 76.9619489 17.0000000 387.0000000
```

Output 3: Advanced model before dealing with outliers (includes statutory holiday)

	term	estimate	std.error	statistic	df	p.value
1	(Intercept)	3530.6017032	151.538018	23.29845512	347.26488	5.784047e-73
2	temperature_cen	16.6609905	18.055571	0.92276178	336.37876	3.567928e-01
3	wind_cen	-2.4755367	10.273151	-0.24097150	351.08643	8.097179e-01
4	darkness	-7.8298555	2.133863	-3.66933357	388.79791	2.771012e-04
5	precipitation_binaryPrecipitation	-593.1919638	135.755158	-4.36957221	386.05796	1.602083e-05
6	snow_binarySnow	-155.8814888	146.127320	-1.06675116	378.10550	2.867649e-01
7	School_holidayAutumn break	-312.2713774	158.834558	-1.96601660	396.50510	4.999401e-02
8	School_holidayChristmas break	-811.4143247	206.582745	-3.92779332	347.77128	1.034018e-04
9	School_holidayMay break	-378.3899927	175.937689	-2.15070458	60.53737	3.550138e-02
10	School_holidaySpring break	-529.4056947	210.281717	-2.51760211	397.10574	1.220829e-02
11	School_holidaySummer break	-716.7113184	86.200106	-8.31450625	399.45162	1.462373e-15
12	WeekdayTuesday	145.8468302	66.174004	2.20398980	399.47235	2.809480e-02
13	WeekdayWednesday	-71.1898577	67.796776	-1.05004783	345.34752	2.944304e-01
14	WeekdayThursday	-61.7015307	66.462228	-0.92836988	390.30142	3.537895e-01
15	WeekdayFriday	-1046.5930957	67.368518	-15.53534388	366.84882	3.518487e-42
16	seasonSpring	-299.5565920	140.536638	-2.13151955	390.79366	3.366972e-02
17	seasonSummer	-224.7758722	182.302079	-1.23298579	391.62398	2.183207e-01
18	seasonAutumn	10.7523318	202.615797	0.05306759	372.06762	9.577065e-01
19	Statutory_holidayTRUE	-2102.5003977	189.571431	-11.09080827	41.53930	5.411577e-14
20	PolBus freq	-374.7905146	125.467408	-2.98715435	384.86525	2.996183e-03
21	PolBus freq + shed	-155.9010729	67.979475	-2.29335506	248.66072	2.266250e-02
22	PolBus freq + shed + No e-bike	-364.9277441	140.008520	-2.60646813	400.03030	9.489873e-03
23	PolBus freq + shed + drop	-214.0881843	83.732462	-2.55681225	377.40026	1.095411e-02
24	PolBus freq + shed + 35ct	-288.6649896	112.806357	-2.55894256	395.95782	1.087001e-02
25	temperature_cen:seasonSpring	0.4796328	20.542274	0.02334858	305.78381	9.813875e-01
26	temperature_cen:seasonSummer	-1.8604147	24.262413	-0.07667888	378.43601	9.389195e-01
27	temperature_cen:seasonAutumn	-7.7097832	21.782228	-0.35394833	368.46731	7.235801e-01
28	wind_cen:seasonSpring	10.6443377	13.492212	0.78892457	267.39856	4.308550e-01
29	wind_cen:seasonSummer	-9.8811003	12.747219	-0.77515730	377.51854	4.387320e-01
30	wind_cen:seasonAutumn	-8.9883118	12.871465	-0.69831301	338.83702	4.854604e-01
31	darkness:seasonSpring	-45.3643795	10.506297	-4.31782754	178.03047	2.611765e-05
32	darkness:seasonSummer	9.8715119	8.839773	1.11671559	398.38149	2.647890e-01
33	darkness:seasonAutumn	-1.6159260	3.222100	-0.50151337	391.83444	6.162915e-01
34	precipitation_binaryPrecipitation:seasonSpring	-41.2902861	175.101119	-0.23580824	353.08614	8.137182e-01
35	precipitation_binaryPrecipitation:seasonSummer	259.7239209	168.493635	1.54144648	391.44890	1.240161e-01
36	precipitation_binaryPrecipitation:seasonAutumn	229.9285864	171.333512	1.34199424	378.37607	1.804026e-01

Model	r2	adjr2	resSE	F_stat	df_l	df_r
5.5000000	0.7638554	0.7433973	430.4071803	37.3833279	35.0000000	404.0000000

Output 4: Advanced model after dealing with outliers (excludes statutory holiday and stepwise deleted variables)

	term	estimate	std.error	statistic	df	p.value
1	(Intercept)	3613.053400	118.194513	30.5687066	257.49497	1.171884e-87
2	temperature_cen	17.379246	5.157861	3.3694679	251.29663	8.715422e-04
3	wind_cen	-5.989562	3.284439	-1.8236179	321.03562	6.913954e-02
4	darkness	-7.888610	1.476239	-5.3437204	277.37763	1.901612e-07
5	precipitation_binaryPrecipitation	-457.655997	46.656470	-9.8090576	344.78070	3.400764e-20
6	snow_binarySnow	-294.782185	124.447321	-2.3687307	339.56567	1.840845e-02
7	School_holidayAutumn break	-354.213179	126.389598	-2.8025501	359.95690	5.344385e-03
8	School_holidayChristmas break	-857.132929	204.803599	-4.1851458	226.25114	4.080600e-05
9	School_holidayMay break	-394.494379	151.955196	-2.5961230	35.13451	1.367266e-02
10	School_holidaySpring break	-502.482769	178.327408	-2.8177540	361.00566	5.102250e-03
11	School_holidaySummer break	-698.313422	67.940996	-10.2782335	372.63535	5.544597e-22
12	WeekdayTuesday	112.642454	54.123270	2.0812204	357.85586	3.812429e-02
13	WeekdayWednesday	-98.056917	55.855181	-1.7555563	246.10309	8.040829e-02
14	WeekdayThursday	-27.216241	56.182043	-0.4844295	318.48229	6.284140e-01
15	WeekdayFriday	-1101.321409	54.455732	-20.2241594	336.32828	4.477065e-60
16	seasonSpring	-368.776295	103.811166	-3.5523760	360.89512	4.324182e-04
17	seasonSummer	-228.045830	122.287973	-1.8648263	343.76834	6.305766e-02
18	seasonAutumn	33.527742	158.273716	0.2118339	304.88543	8.323781e-01
19	PolBus freq	-317.433002	103.158412	-3.0771412	305.84101	2.279084e-03
20	PolBus freq + shed	-204.698438	58.705920	-3.4868449	129.19266	6.684480e-04
21	PolBus freq + shed + No e-bike	-384.737786	108.238241	-3.5545458	374.60034	4.270724e-04
22	PolBus freq + shed + drop	-258.195416	66.397587	-3.8886265	308.70461	1.234699e-04
23	PolBus freq + shed + drop + 35ct	-218.237503	90.174991	-2.4201555	359.40811	1.600911e-02
24	darkness:seasonSpring	-33.935550	9.447540	-3.5919984	76.82425	5.764740e-04
25	darkness:seasonSummer	7.756381	6.552798	1.1836746	367.78204	2.373064e-01
26	darkness:seasonAutumn	-2.256189	2.383743	-0.9464900	345.29688	3.445606e-01

Model	r2	adjr2	resSE	F_stat	df_l	df_r
5.5000000	0.8028143	0.7898073	335.7230086	61.7976083	25.0000000	379.0000000

Output 5: ANOVA comparing basic and advanced bicycle model

This ANOVA is applied to two proxy models, while it was not possible to do an ANOVA on the pooled models.

```
Analysis of Variance Table

Model 1: Bicycle_prox ~ temperature_cen + wind_cen + darkness + precipitation_binary +
    snow_binary + School_holiday + Weekday + season
Model 2: Bicycle_prox ~ temperature_cen + wind_cen + darkness + precipitation_binary +
    snow_binary + School_holiday + Weekday + season + Pol + season *
    darkness
  Res.Df      RSS Df Sum of Sq      F    Pr(>F)
1    387 46550677
2    379 39873244  8   6677433 7.9337 6.871e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

E.2 Public transport

Output 1: Basic model before dealing with outliers (includes statutory holidays)

```
Call:
lm(formula = Public_transport ~ temperature_cen + wind_cen + 
    darkness + precipitation_binary + snow_binary + School_holiday + 
    Weekday + season + Statutory_holiday, data = ASML_data)

Residuals:
    Min      1Q  Median      3Q     Max 
-648.95  -52.03    2.22   62.32  390.53 

Coefficients:
                                    Estimate Std. Error t value Pr(>|t|)    
(Intercept)                         562.3340    24.6262  22.835  < 2e-16 ***
temperature_cen                       3.0697     1.5579   1.970 0.049371 *  
wind_cen                              0.8645     1.0248   0.844 0.399305    
darkness                              0.9805     0.3163   3.099 0.002053 ** 
precipitation_binaryPrecipitation    49.4196    14.4022   3.431 0.000652 ***
snow_binarySnow                     -12.1703    31.9808  -0.381 0.703706    
School_holidayAutumn break          -74.4439    42.1038  -1.768 0.077677 .  
School_holidayChristmas break      -353.0337    40.8460  -8.643  < 2e-16 ***
School_holidayMay break             -56.3614    41.4447  -1.360 0.174493    
School_holidaySpring break          -92.7782    42.0842  -2.205 0.027956 *  
School_holidaySummer break          -35.2485    22.1352  -1.592 0.111948    
WeekdayTuesday                       55.9444    17.9143   3.123 0.001899 ** 
WeekdayWednesday                      2.0779    18.0046   0.115 0.908168    
WeekdayThursday                      36.6105    17.8535   2.051 0.040847 *  
WeekdayFriday                      -241.9025    17.8832 -13.527  < 2e-16 ***
seasonSpring                         32.9636    24.4640   1.347 0.178475    
seasonSummer                        -26.3426    28.8181  -0.914 0.361121    
seasonAutumn                         25.0585    19.4876   1.286 0.199107    
Statutory_holidayTRUE              -475.9854    40.6844 -11.699  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 125.6 on 481 degrees of freedom
Multiple R-squared:  0.589,	Adjusted R-squared:  0.5736 
F-statistic: 38.29 on 18 and 481 DF,  p-value: < 2.2e-16
```

Output 2: basic model after dealing with outliers (excludes statutory holiday)

```
Call:
lm(formula = Public_transport ~ temperature_cen + wind_cen +
    darkness + precipitation_binary + snow_binary + School_holiday +
    Weekday + season, data = PT_cook)

Residuals:
    Min      1Q  Median      3Q     Max
-359.60  -50.65   -0.46   56.44  197.80

Coefficients:
                                   Estimate Std. Error t value Pr(>|t|)
(Intercept)                        640.6668    18.2990  35.011  < 2e-16 ***
temperature_cen                     -0.1934     1.1087  -0.174 0.861631
wind_cen                             0.6637     0.7108   0.934 0.350975
darkness                             0.5593     0.2440   2.293 0.022331 *
precipitation_binaryPrecipitation   48.0144    10.0241   4.790 2.29e-06 ***
snow_binarySnow                     12.5934    27.4719   0.458 0.646885
School_holidayAutumn break         -68.0147    28.3738  -2.397 0.016943 *
School_holidayChristmas break     -385.3192    34.4494 -11.185  < 2e-16 ***
School_holidayMay break            -58.2925    27.8209  -2.095 0.036719 *
School_holidaySpring break        -143.5354    31.6578  -4.534 7.48e-06 ***
School_holidaySummer break         -35.4899    14.9011  -2.382 0.017658 *
WeekdayTuesday                      68.6479    12.6435   5.429 9.37e-08 ***
WeekdayWednesday                    10.3427    12.5417   0.825 0.410008
WeekdayThursday                     45.4955    12.5797   3.617 0.000333 ***
WeekdayFriday                     -255.7010    12.4051 -20.612  < 2e-16 ***
seasonSpring                       -44.2861    18.1696  -2.437 0.015191 *
seasonSummer                       -85.9288    20.7204  -4.147 4.04e-05 ***
seasonAutumn                       -34.7267    14.0332  -2.475 0.013715 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 84.28 on 439 degrees of freedom
Multiple R-squared:  0.7283,    Adjusted R-squared:  0.7178
F-statistic: 69.23 on 17 and 439 DF,  p-value: < 2.2e-16
```

Output 3: Advanced model before dealing with outliers (includes statutory holiday)

```
Call:
lm(formula = Public_transport ~ temperature_cen + wind_cen + 
    darkness + precipitation_binary + snow_binary + School_holiday + 
    Weekday + season + Statutory_holiday + Pol + season * temperature_cen + 
    season * wind_cen + season * darkness + season * precipitation_binary, 
    data = ASML_data)

Residuals:
    Min      1Q  Median      3Q     Max 
-567.81  -31.83    5.96   45.10  451.33 

Coefficients:
                                              Estimate Std. Error t value Pr(>|t|)    
(Intercept)                                   592.3015    26.8486  22.061  < 2e-16 ***
temperature_cen                                24.5638     2.9354   8.368 6.94e-16 ***
wind_cen                                       -6.1581     1.7231  -3.574 0.000389 ***
darkness                                        1.8730     0.3125   5.994 4.13e-09 ***
precipitation_binaryPrecipitation              -0.5917    22.2695  -0.027 0.978815    
snow_binarySnow                                 2.3884    26.0468   0.092 0.926979    
School_holidayAutumn break                    -49.5213    35.7168  -1.387 0.166260    
School_holidayChristmas break                -404.4189    34.9679 -11.565  < 2e-16 ***
School_holidayMay break                       -53.1794    32.4114  -1.641 0.101524    
School_holidaySpring break                   -146.3760    33.5123  -4.368 1.55e-05 ***
School_holidaySummer break                    -28.4034    19.4997  -1.457 0.145902    
WeekdayTuesday                                 54.1827    14.0095   3.868 0.000126 ***
WeekdayWednesday                               -0.1665    14.0509  -0.012 0.990552    
WeekdayThursday                                34.4602    13.9905   2.463 0.014136 *  
WeekdayFriday                                -244.3548    14.0545 -17.386  < 2e-16 ***
seasonSpring                                  -71.3350    26.5240  -2.689 0.007415 ** 
seasonSummer                                  -69.9700    36.7115  -1.906 0.057275 .  
seasonAutumn                                  -16.5037    40.9295  -0.403 0.686970    
Statutory_holidayTRUE                        -468.6831    32.0623 -14.618  < 2e-16 ***
PolBus freq                                    15.6925    26.7225   0.587 0.557330    
PolBus freq + shed                            164.6599    11.5958  14.200  < 2e-16 ***
PolBus freq + shed + No e-bike                 84.1947    31.6816   2.658 0.008143 ** 
PolBus freq + shed + drop                      65.1759    18.6470   3.495 0.000519 ***
PolBus freq + shed + drop + 35ct               21.8293    24.9552   0.875 0.382168    
temperature_cen:seasonSpring                  -27.2926     3.5722  -7.640 1.25e-13 ***
temperature_cen:seasonSummer                  -25.6642     4.7137  -5.445 8.43e-08 ***
temperature_cen:seasonAutumn                  -30.7241     3.9545  -7.769 5.11e-14 ***
wind_cen:seasonSpring                           9.6725     2.4325   3.976 8.12e-05 ***
wind_cen:seasonSummer                           8.3896     2.4295   3.453 0.000605 ***
wind_cen:seasonAutumn                           7.7788     2.4042   3.235 0.001301 ** 
darkness:seasonSpring                          -3.6829     2.1433  -1.718 0.086404 .  
darkness:seasonSummer                          -2.1402     1.9675  -1.088 0.277263    
darkness:seasonAutumn                          -1.3165     0.6361  -2.070 0.039052 *  
precipitation_binaryPrecipitation:seasonSpring 48.9858    32.2078   1.521 0.128959    
precipitation_binaryPrecipitation:seasonSummer 31.8469    31.7929   1.002 0.317012    
precipitation_binaryPrecipitation:seasonAutumn 67.0646    32.2034   2.083 0.037841 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 97.72 on 464 degrees of freedom
Multiple R-squared:  0.7601,	Adjusted R-squared:  0.742 
F-statistic: 42.01 on 35 and 464 DF,  p-value: < 2.2e-16
```

Output 4: Advanced model after dealing with outliers (excludes statutory holiday and stepwise deleted variables)

```
Call:
lm(formula = Public_transport ~ temperature_cen + wind_cen + 
    darkness + precipitation_binary + School_holiday + Weekday + 
    season + Pol + season * temperature_cen + season * wind_cen + 
    season * darkness + season * precipitation_binary, data = PT_cook)

Residuals:
     Min       1Q   Median       3Q      Max 
-239.878  -29.591    0.027   32.664  133.969 

Coefficients:
                                                Estimate Std. Error t value Pr(>|t|)    
(Intercept)                                     616.8939    17.1531  35.964  < 2e-16 ***
temperature_cen                                  11.5978     2.1237   5.461 8.09e-08 ***
wind_cen                                         -2.6387     1.1491  -2.296 0.022151 *  
darkness                                          1.1522     0.2248   5.125 4.54e-07 ***
precipitation_binaryPrecipitation                -0.8972    14.7879  -0.061 0.951649    
School_holidayAutumn break                      -40.9016    20.9175  -1.955 0.051196 .  
School_holidayChristmas break                  -388.8882    25.4984 -15.251  < 2e-16 ***
School_holidayMay break                         -55.9855    18.6570  -3.001 0.002852 ** 
School_holidaySpring break                     -159.4756    21.8363  -7.303 1.41e-12 ***
School_holidaySummer break                      -31.4470    11.2193  -2.803 0.005297 ** 
WeekdayTuesday                                   62.9997     8.4767   7.432 5.98e-13 ***
WeekdayWednesday                                  5.0190     8.4068   0.597 0.550815    
WeekdayThursday                                  39.9535     8.4521   4.727 3.11e-06 ***
WeekdayFriday                                  -260.0025     8.3718 -31.057  < 2e-16 ***
seasonSpring                                    -84.5305    16.6847  -5.066 6.07e-07 ***
seasonSummer                                    -80.6990    22.4117  -3.601 0.000355 ***
seasonAutumn                                    -68.7135    24.9037  -2.759 0.006046 ** 
PolBus freq                                      68.7751    18.2592   3.767 0.000189 ***
PolBus freq + shed                              140.6169     7.2436  19.413  < 2e-16 ***
PolBus freq + shed + No e-bike                   83.4109    18.2418   4.573 6.34e-06 ***
PolBus freq + shed + drop                        70.7901    10.8153   6.545 1.72e-10 ***
PolBus freq + shed + drop + 35ct                 48.7234    15.4638   3.151 0.001744 ** 
temperature_cen:seasonSpring                    -13.5971     2.4422  -5.568 4.60e-08 ***
temperature_cen:seasonSummer                    -14.1848     3.0205  -4.696 3.59e-06 ***
temperature_cen:seasonAutumn                    -15.1121     2.6165  -5.776 1.49e-08 ***
wind_cen:seasonSpring                             5.3392     1.5217   3.509 0.000498 ***
wind_cen:seasonSummer                             4.9668     1.5159   3.277 0.001137 ** 
wind_cen:seasonAutumn                             3.5332     1.4920   2.368 0.018326 *  
darkness:seasonSpring                            -2.5624     1.2743  -2.011 0.044974 *  
darkness:seasonSummer                            -1.6815     1.1401  -1.475 0.140983    
darkness:seasonAutumn                            -0.4952     0.4029  -1.229 0.219695    
precipitation_binaryPrecipitation:seasonSpring   55.4902    19.9560   2.781 0.005668 ** 
precipitation_binaryPrecipitation:seasonSummer   30.8316    19.7540   1.561 0.119324    
precipitation_binaryPrecipitation:seasonAutumn   57.8054    20.0261   2.887 0.004095 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 56.2 on 423 degrees of freedom
Multiple R-squared:  0.8836,	Adjusted R-squared:  0.8745 
F-statistic: 97.31 on 33 and 423 DF,  p-value: < 2.2e-16
```

Output 5: ANOVA comparing basic and advanced public transport model

```
Analysis of Variance Table

Model 1: Public_transport ~ temperature_cen + wind_cen + darkness + precipitation_binary +
    snow_binary + School_holiday + Weekday + season
Model 2: Public_transport ~ temperature_cen + wind_cen + darkness + precipitation_binary +
    School_holiday + Weekday + season + Pol + season * temperature_cen +
    season * wind_cen + season * darkness + season * precipitation_binary
  Res.Df     RSS Df Sum of Sq      F    Pr(>F)
1    439 3118357
2    423 1335915 16   1782442 35.274 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Output 6: Advanced model with HAC standard errors

```
z test of coefficients:

                                                 Estimate Std. Error  z value  Pr(>|z|)
(Intercept)                                     616.89393   31.14348  19.8081 < 2.2e-16 ***
temperature_cen                                  11.59781    2.78975   4.1573 3.220e-05 ***
wind_cen                                         -2.63869    1.48102  -1.7817 0.0748033 .
darkness                                          1.15221    0.46867   2.4585 0.0139533 *
precipitation_binaryPrecipitation                -0.89722   14.13380  -0.0635 0.9493839
School_holidayAutumn break                      -40.90158   18.93426  -2.1602 0.0307581 *
School_holidayChristmas break                  -388.88825   18.89510 -20.5814 < 2.2e-16 ***
School_holidayMay break                         -55.98548   15.56623  -3.5966 0.0003224 ***
School_holidaySpring break                     -159.47559   26.06039  -6.1195 9.389e-10 ***
School_holidaySummer break                      -31.44703   15.01316  -2.0946 0.0362038 *
WeekdayTuesday                                   62.99968    5.31890  11.8445 < 2.2e-16 ***
WeekdayWednesday                                  5.01902    5.78960   0.8669 0.3859952
WeekdayThursday                                  39.95351    7.55988   5.2849 1.257e-07 ***
WeekdayFriday                                  -260.00253    8.33134 -31.2078 < 2.2e-16 ***
seasonSpring                                    -84.53054   30.12164  -2.8063 0.0050113 **
seasonSummer                                    -80.69904   34.37710  -2.3475 0.0189016 *
seasonAutumn                                    -68.71353   38.13165  -1.8020 0.0715442 .
PolBus freq                                      68.77514   13.75918   4.9985 5.778e-07 ***
PolBus freq + shed                              140.61692   14.89957   9.4376 < 2.2e-16 ***
PolBus freq + shed + No e-bike                   83.41085   17.78381   4.6903 2.728e-06 ***
PolBus freq + shed + drop                        70.79009   10.22584   6.9227 4.432e-12 ***
PolBus freq + shed + drop + 35ct                 48.72342   11.70425   4.1629 3.143e-05 ***
temperature_cen:seasonSpring                    -13.59712    2.85462  -4.7632 1.906e-06 ***
temperature_cen:seasonSummer                    -14.18484    3.54213  -4.0046 6.212e-05 ***
temperature_cen:seasonAutumn                    -15.11207    2.88626  -5.2359 1.642e-07 ***
wind_cen:seasonSpring                             5.33918    1.69457   3.1508 0.0016284 **
wind_cen:seasonSummer                             4.96685    1.69771   2.9256 0.0034378 **
wind_cen:seasonAutumn                             3.53317    2.02483   1.7449 0.0809990 .
darkness:seasonSpring                            -2.56240    0.72486  -3.5350 0.0004077 ***
darkness:seasonSummer                            -1.68146    0.67432  -2.4936 0.0126466 *
darkness:seasonAutumn                            -0.49524    0.62529  -0.7920 0.4283468
precipitation_binaryPrecipitation:seasonSpring   55.49024   20.44317   2.7144 0.0066403 **
precipitation_binaryPrecipitation:seasonSummer   30.83160   18.73357   1.6458 0.0998062 .
precipitation_binaryPrecipitation:seasonAutumn   57.80535   17.42895   3.3166 0.0009111 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

E.3 Car

Output 1: Basic model before dealing with outliers (includes statutory holidays)

```
Call:
lm(formula = Car ~ temperature_cen + wind_cen + darkness + precipitation_binary +
    snow_binary + School_holiday + Weekday + season + Statutory_holiday,
    data = ASML_data)

Residuals:
    Min      1Q  Median      3Q     Max
-3284.8  -245.6    97.7   444.5  1279.7

Coefficients:
                                   Estimate Std. Error t value Pr(>|t|)
(Intercept)                        4385.287    145.835  30.070  < 2e-16 ***
temperature_cen                      -2.052      9.226  -0.222  0.82405
wind_cen                             -1.966      6.069  -0.324  0.74614
darkness                              2.968      1.873   1.584  0.11380
precipitation_binaryPrecipitation    51.790     85.289   0.607  0.54398
snow_binarySnow                    -181.405    189.388  -0.958  0.33862
School_holidayAutumn break         -510.939    249.336  -2.049  0.04098 *
School_holidayChristmas break     -2082.238    241.888  -8.608  < 2e-16 ***
School_holidayMay break            -104.353    245.433  -0.425  0.67090
School_holidaySpring break         -673.320    249.220  -2.702  0.00714 **
School_holidaySummer break         -423.290    131.083  -3.229  0.00133 **
WeekdayTuesday                      297.481    106.087   2.804  0.00525 **
WeekdayWednesday                    125.917    106.622   1.181  0.23820
WeekdayThursday                     137.078    105.727   1.297  0.19542
WeekdayFriday                      -847.245    105.903  -8.000 9.36e-15 ***
seasonSpring                       -426.970    144.874  -2.947  0.00336 **
seasonSummer                       -402.191    170.659  -2.357  0.01884 *
seasonAutumn                       -206.977    115.404  -1.793  0.07352 .
Statutory_holidayTRUE             -3372.934    240.930 -14.000  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 744 on 481 degrees of freedom
Multiple R-squared:  0.5443,	Adjusted R-squared:  0.5273
F-statistic: 31.92 on 18 and 481 DF,  p-value: < 2.2e-16
```

Output 2: basic model after dealing with outliers (excludes statutory holiday)

```
Call:
lm(formula = Car ~ temperature_cen + wind_cen + darkness + precipitation_binary +
    snow_binary + School_holiday + Weekday + season, data = Car_cook)

Residuals:
     Min       1Q   Median       3Q      Max
-1801.24  -229.67    40.98   336.42   938.60

Coefficients:
                                   Estimate Std. Error t value Pr(>|t|)
(Intercept)                        4450.839     91.243  48.780  < 2e-16 ***
temperature_cen                     -10.653      5.865  -1.816  0.06996 .
wind_cen                              8.509      3.922   2.169  0.03058 *
darkness                              1.417      1.168   1.213  0.22565
precipitation_binaryPrecipitation    52.262     54.047   0.967  0.33408
snow_binarySnow                    -292.402    116.167  -2.517  0.01218 *
School_holidayAutumn break         -501.262    152.653  -3.284  0.00110 **
School_holidayChristmas break     -2081.732    168.129 -12.382  < 2e-16 ***
School_holidayMay break            -480.833    150.801  -3.189  0.00153 **
School_holidaySpring break         -720.796    152.466  -4.728 3.05e-06 ***
School_holidaySummer break         -492.830     80.807  -6.099 2.31e-09 ***
WeekdayTuesday                      293.848     66.838   4.396 1.38e-05 ***
WeekdayWednesday                    124.604     66.787   1.866  0.06274 .
WeekdayThursday                     280.623     68.110   4.120 4.51e-05 ***
WeekdayFriday                      -919.850     66.436 -13.846  < 2e-16 ***
seasonSpring                       -103.638     90.743  -1.142  0.25402
seasonSummer                       -329.660    106.913  -3.083  0.00217 **
seasonAutumn                       -173.655     71.519  -2.428  0.01557 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 454.6 on 449 degrees of freedom
Multiple R-squared:  0.6553,	Adjusted R-squared:  0.6422
F-statistic: 50.21 on 17 and 449 DF,  p-value: < 2.2e-16
```

Output 3: Advanced model before dealing with outliers (includes statutory holiday)

```
Call:
lm(formula = Car ~ temperature_cen + wind_cen + darkness + precipitation_binary +
    snow_binary + School_holiday + Weekday + season + Statutory_holiday +
    Pol + season * temperature_cen + season * wind_cen + season *
    darkness + season * precipitation_binary, data = ASML_data)

Residuals:
    Min      1Q  Median      3Q     Max
-3439.3  -217.3   101.9   355.1  1561.8

Coefficients:
                                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                                   4365.774    196.565  22.210  < 2e-16 ***
temperature_cen                                -28.150     21.491  -1.310  0.19089
wind_cen                                        15.797     12.615   1.252  0.21113
darkness                                        -1.003      2.288  -0.438  0.66140
precipitation_binaryPrecipitation              -19.311    163.041  -0.118  0.90577
snow_binarySnow                               -303.058    190.695  -1.589  0.11269
School_holidayAutumn break                    -319.054    261.492  -1.220  0.22304
School_holidayChristmas break                -1920.470    256.009  -7.502 3.24e-13 ***
School_holidayMay break                        -49.909    237.292  -0.210  0.83351
School_holidaySpring break                    -632.982    245.352  -2.580  0.01019 *
School_holidaySummer break                    -319.733    142.763  -2.240  0.02559 *
WeekdayTuesday                                 314.500    102.568   3.066  0.00229 **
WeekdayWednesday                               147.451    102.870   1.433  0.15243
WeekdayThursday                                164.496    102.428   1.606  0.10896
WeekdayFriday                                 -829.819    102.897  -8.065 6.32e-15 ***
seasonSpring                                  -586.160    194.189  -3.018  0.00268 **
seasonSummer                                  -755.241    268.774  -2.810  0.00516 **
seasonAutumn                                  -109.274    299.656  -0.365  0.71553
Statutory_holidayTRUE                        -3422.383    234.736 -14.580  < 2e-16 ***
PolBus freq                                    199.520    195.643   1.020  0.30835
PolBus freq + shed                             115.209     84.896   1.357  0.17542
PolBus freq + shed + No e-bike                -507.356    231.949  -2.187  0.02921 *
PolBus freq + shed + drop                       52.296    136.520   0.383  0.70185
PolBus freq + shed + drop + 35ct              -449.134    182.703  -2.458  0.01433 *
temperature_cen:seasonSpring                    38.406     26.153   1.469  0.14264
temperature_cen:seasonSummer                    65.970     34.510   1.912  0.05654 .
temperature_cen:seasonAutumn                    14.646     28.952   0.506  0.61319
wind_cen:seasonSpring                          -40.637     17.809  -2.282  0.02296 *
wind_cen:seasonSummer                          -32.499     17.787  -1.827  0.06832 .
wind_cen:seasonAutumn                           -8.453     17.602  -0.480  0.63129
darkness:seasonSpring                           71.194     15.692   4.537 7.27e-06 ***
darkness:seasonSummer                           19.984     14.405   1.387  0.16602
darkness:seasonAutumn                            2.833      4.657   0.608  0.54328
precipitation_binaryPrecipitation:seasonSpring  62.027    235.802   0.263  0.79263
precipitation_binaryPrecipitation:seasonSummer 210.289    232.764   0.903  0.36676
precipitation_binaryPrecipitation:seasonAutumn 203.792    235.770   0.864  0.38783
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 715.4 on 464 degrees of freedom
Multiple R-squared:  0.5935,    Adjusted R-squared:  0.5629
F-statistic: 19.36 on 35 and 464 DF,  p-value: < 2.2e-16
```

Output 4: Advanced model after dealing with outliers (excludes statutory holiday and stepwise deleted variables)

```
Call:
lm(formula = Car ~ temperature_cen + wind_cen + darkness + precipitation_binary +
    snow_binary + School_holiday + Weekday + season + Pol + season *
    temperature_cen + season * darkness, data = Car_cook)

Residuals:
     Min       1Q   Median       3Q      Max
-1862.73  -179.25    61.52   233.68   860.48

Coefficients:
                                   Estimate Std. Error t value Pr(>|t|)
(Intercept)                        4634.911    101.723  45.564  < 2e-16 ***
temperature_cen                     -26.586      9.889  -2.688 0.007454 **
wind_cen                              6.974      3.700   1.885 0.060076 .
darkness                             -1.664      1.220  -1.363 0.173466
precipitation_binaryPrecipitation    81.640     48.988   1.667 0.096319 .
snow_binarySnow                    -380.726    105.118  -3.622 0.000327 ***
School_holidayAutumn break         -365.550    148.147  -2.467 0.013988 *
School_holidayChristmas break     -1907.056    158.839 -12.006  < 2e-16 ***
School_holidayMay break            -453.587    134.111  -3.382 0.000784 ***
School_holidaySpring break         -715.162    136.827  -5.227 2.68e-07 ***
School_holidaySummer break         -449.744     80.816  -5.565 4.57e-08 ***
WeekdayTuesday                      293.907     59.283   4.958 1.02e-06 ***
WeekdayWednesday                    124.772     59.204   2.108 0.035642 *
WeekdayThursday                     276.362     60.401   4.575 6.19e-06 ***
WeekdayFriday                      -930.791     58.931 -15.794  < 2e-16 ***
seasonSpring                       -233.602     98.767  -2.365 0.018456 *
seasonSummer                       -665.675    145.191  -4.585 5.94e-06 ***
seasonAutumn                       -349.811    164.920  -2.121 0.034476 *
PolBus freq                         251.995    118.988   2.118 0.034753 *
PolBus freq + shed                 -218.956     49.789  -4.398 1.38e-05 ***
PolBus freq + shed + No e-bike     -505.347    129.060  -3.916 0.000105 ***
PolBus freq + shed + drop           -67.762     76.225  -0.889 0.374501
PolBus freq + shed + drop + 35ct   -396.441    106.019  -3.739 0.000209 ***
temperature_cen:seasonSpring         48.036     13.272   3.619 0.000330 ***
temperature_cen:seasonSummer         52.024     18.782   2.770 0.005846 **
temperature_cen:seasonAutumn         11.238     13.715   0.819 0.412995
darkness:seasonSpring                44.916      9.160   4.903 1.33e-06 ***
darkness:seasonSummer                 4.310      8.038   0.536 0.592132
darkness:seasonAutumn                 3.730      2.610   1.429 0.153781
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 402.4 on 438 degrees of freedom
Multiple R-squared:  0.7365,    Adjusted R-squared:  0.7196
F-statistic: 43.72 on 28 and 438 DF,  p-value: < 2.2e-16
```

Output 5: ANOVA comparing basic and advanced car model

```
Analysis of Variance Table

Model 1: Car ~ temperature_cen + wind_cen + darkness + precipitation_binary +
    snow_binary + School_holiday + Weekday + season
Model 2: Car ~ temperature_cen + wind_cen + darkness + precipitation_binary +
    snow_binary + School_holiday + Weekday + season + Pol + season *
    temperature_cen + season * darkness
  Res.Df      RSS Df Sum of Sq      F    Pr(>F)
1    449 92780684
2    438 70930920 11  21849764 12.266 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Output 6: Advanced model with HAC standard errors

```
z test of coefficients:

                                     Estimate Std. Error  z value  Pr(>|z|)
(Intercept)                         4634.9110   166.1769  27.8914 < 2.2e-16 ***
temperature_cen                      -26.5861    10.4442  -2.5455 0.0109113 *
wind_cen                               6.9741     3.4010   2.0506 0.0403068 *
darkness                              -1.6635     1.6945  -0.9817 0.3262466
precipitation_binaryPrecipitation     81.6403    50.5300   1.6157 0.1061632
snow_binarySnow                     -380.7264   115.8251  -3.2871 0.0010123 **
School_holidayAutumn break          -365.5498   146.1385  -2.5014 0.0123706 *
School_holidayChristmas break      -1907.0559   188.3041 -10.1275 < 2.2e-16 ***
School_holidayMay break             -453.5870   179.5628  -2.5261 0.0115349 *
School_holidaySpring break          -715.1623   118.6326  -6.0284 1.656e-09 ***
School_holidaySummer break          -449.7439   231.3092  -1.9443 0.0518544 .
WeekdayTuesday                       293.9067    32.2779   9.1055 < 2.2e-16 ***
WeekdayWednesday                     124.7719    42.0170   2.9696 0.0029823 **
WeekdayThursday                      276.3625    45.8013   6.0339 1.600e-09 ***
WeekdayFriday                       -930.7906    48.5954 -19.1539 < 2.2e-16 ***
seasonSpring                        -233.6022   166.6501  -1.4018 0.1609892
seasonSummer                        -665.6755   380.5459  -1.7493 0.0802453 .
seasonAutumn                        -349.8106   253.1393  -1.3819 0.1670056
PolBus freq                          251.9954   266.2924   0.9463 0.3439901
PolBus freq + shed                  -218.9561    96.7191  -2.2638 0.0235842 *
PolBus freq + shed + No e-bike      -505.3468   150.5080  -3.3576 0.0007862 ***
PolBus freq + shed + drop            -67.7623   149.2037  -0.4542 0.6497137
PolBus freq + shed + drop + 35ct    -396.4414   183.9982  -2.1546 0.0311937 *
temperature_cen:seasonSpring          48.0363    19.9769   2.4046 0.0161907 *
temperature_cen:seasonSummer          52.0244    32.4063   1.6054 0.1084105
temperature_cen:seasonAutumn          11.2383     9.8948   1.1358 0.2560480
darkness:seasonSpring                 44.9158     8.9782   5.0027 5.652e-07 ***
darkness:seasonSummer                  4.3096    14.1157   0.3053 0.7601327
darkness:seasonAutumn                  3.7297     2.9870   1.2487 0.2117857
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

E.4 Employees in office

Output 1: Basic model before dealing with outliers (includes statutory holidays)

```
Call:
lm(formula = Employees_in_office ~ temperature_cen + wind_cen +
    darkness + precipitation_binary + snow_binary + School_holiday +
    Weekday + season + Statutory_holiday, data = ASML_data)

Residuals:
    Min      1Q  Median      3Q     Max
-4575.0  -288.5    85.3   416.1  7891.4

Coefficients:
                                   Estimate Std. Error t value Pr(>|t|)
(Intercept)                       11518.678    170.874  67.410  < 2e-16 ***
temperature_cen                      -2.253     10.810  -0.208 0.834970
wind_cen                             11.043      7.111   1.553 0.121085
darkness                             -2.595      2.195  -1.182 0.237647
precipitation_binaryPrecipitation  -251.677     99.933  -2.518 0.012110 *
snow_binarySnow                    -378.556    221.906  -1.706 0.088667 .
School_holidayAutumn break        -1433.199    292.146  -4.906 1.28e-06 ***
School_holidayChristmas break     -4976.074    283.419 -17.557  < 2e-16 ***
School_holidayMay break           -1220.401    287.573  -4.244 2.64e-05 ***
School_holidaySpring break        -2174.999    292.010  -7.448 4.41e-13 ***
School_holidaySummer break        -1860.557    153.590 -12.114  < 2e-16 ***
WeekdayTuesday                      854.434    124.302   6.874 1.95e-11 ***
WeekdayWednesday                    185.219    124.929   1.483 0.138838
WeekdayThursday                     449.657    123.880   3.630 0.000314 ***
WeekdayFriday                     -2966.206    124.087 -23.904  < 2e-16 ***
seasonSpring                       -563.486    169.749  -3.320 0.000970 ***
seasonSummer                       -563.668    199.960  -2.819 0.005018 **
seasonAutumn                       -224.383    135.219  -1.659 0.097686 .
Statutory_holidayTRUE             -8381.299    282.297 -29.690  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 871.7 on 481 degrees of freedom
Multiple R-squared:  0.8673,    Adjusted R-squared:  0.8624
F-statistic: 174.7 on 18 and 481 DF,  p-value: < 2.2e-16
```

Output 2: basic model after dealing with outliers (excludes statutory holiday)

```
Call:
lm(formula = Employees_in_office ~ temperature_cen + wind_cen +
    darkness + precipitation_binary + snow_binary + School_holiday +
    Weekday + season, data = Emp_cook)

Residuals:
     Min       1Q   Median       3Q      Max
-1311.31  -281.72     7.09   303.70  1347.78

Coefficients:
                                    Estimate Std. Error t value Pr(>|t|)
(Intercept)                        11473.120     91.772 125.017  < 2e-16 ***
temperature_cen                       10.716      5.895   1.818 0.069766 .
wind_cen                               3.594      3.897   0.922 0.356997
darkness                               0.852      1.196   0.712 0.476729
precipitation_binaryPrecipitation   -245.004     54.460  -4.499 8.70e-06 ***
snow_binarySnow                     -305.792    121.389  -2.519 0.012109 *
School_holidayAutumn break         -1578.836    155.392 -10.160  < 2e-16 ***
School_holidayChristmas break      -5215.014    209.264 -24.921  < 2e-16 ***
School_holidayMay break            -1470.302    153.139  -9.601  < 2e-16 ***
School_holidaySpring break         -2138.451    162.990 -13.120  < 2e-16 ***
School_holidaySummer break         -1871.553     81.582 -22.941  < 2e-16 ***
WeekdayTuesday                       832.679     67.742  12.292  < 2e-16 ***
WeekdayWednesday                     142.462     67.573   2.108 0.035558 *
WeekdayThursday                      592.152     68.806   8.606  < 2e-16 ***
WeekdayFriday                      -2970.987     68.521 -43.359  < 2e-16 ***
seasonSpring                        -307.001     90.998  -3.374 0.000805 ***
seasonSummer                        -627.695    106.857  -5.874 8.24e-09 ***
seasonAutumn                        -299.761     73.362  -4.086 5.19e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 462.9 on 453 degrees of freedom
Multiple R-squared:  0.9311,    Adjusted R-squared:  0.9285
F-statistic: 359.9 on 17 and 453 DF,  p-value: < 2.2e-16
```

Output 3: Advanced model before dealing with outliers (includes statutory holiday)

```
Call:
lm(formula = Employees_in_office ~ temperature_cen + wind_cen +
    darkness + precipitation_binary + snow_binary + School_holiday +
    Weekday + season + Statutory_holiday + Pol + season * temperature_cen +
    season * wind_cen + season * darkness + season * precipitation_binary,
    data = ASML_data)

Residuals:
    Min      1Q  Median      3Q     Max
-4540.0  -229.6    55.7   369.4  7869.5

Coefficients:
                                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                                 11763.1686   231.6480  50.780  < 2e-16 ***
temperature_cen                                21.9308    25.3265   0.866 0.386980
wind_cen                                       -1.0536    14.8667  -0.071 0.943533
darkness                                       -0.7650     2.6963  -0.284 0.776747
precipitation_binaryPrecipitation            -491.1763   192.1401  -2.556 0.010895 *
snow_binarySnow                              -381.3318   224.7302  -1.697 0.090397 .
School_holidayAutumn break                  -1611.0066   308.1623  -5.228 2.60e-07 ***
School_holidayChristmas break               -4831.8074   301.7012 -16.015  < 2e-16 ***
School_holidayMay break                     -1176.5212   279.6438  -4.207 3.10e-05 ***
School_holidaySpring break                  -2240.4835   289.1420  -7.749 5.90e-14 ***
School_holidaySummer break                  -1789.4836   168.2426 -10.636  < 2e-16 ***
WeekdayTuesday                                842.3455   120.8736   6.969 1.10e-11 ***
WeekdayWednesday                              166.9467   121.2304   1.377 0.169145
WeekdayThursday                               434.9140   120.7094   3.603 0.000349 ***
WeekdayFriday                               -2988.0045   121.2619 -24.641  < 2e-16 ***
seasonSpring                                 -704.6775   228.8479  -3.079 0.002198 **
seasonSummer                                 -749.7679   316.7445  -2.367 0.018337 *
seasonAutumn                                   93.3163   353.1375   0.264 0.791705
Statutory_holidayTRUE                       -8315.8719   276.6316 -30.061  < 2e-16 ***
PolBus freq                                  -599.1404   230.5606  -2.599 0.009658 **
PolBus freq + shed                           -161.7822   100.0476  -1.617 0.106547
PolBus freq + shed + No e-bike               -987.3139   273.3470  -3.612 0.000337 ***
PolBus freq + shed + drop                    -619.4144   160.8853  -3.850 0.000135 ***
PolBus freq + shed + drop + 35ct             -860.8154   215.3118  -3.998 7.43e-05 ***
temperature_cen:seasonSpring                  -15.8014    30.8206  -0.513 0.608413
temperature_cen:seasonSummer                    3.7224    40.6696   0.092 0.927113
temperature_cen:seasonAutumn                  -45.5497    34.1196  -1.335 0.182531
wind_cen:seasonSpring                          43.0092    20.9879   2.049 0.041001 *
wind_cen:seasonSummer                          -3.5519    20.9612  -0.169 0.865514
wind_cen:seasonAutumn                          -5.2231    20.7436  -0.252 0.801313
darkness:seasonSpring                          17.0548    18.4922   0.922 0.356867
darkness:seasonSummer                           0.8357    16.9757   0.049 0.960759
darkness:seasonAutumn                          -2.3574     5.4885  -0.430 0.667745
precipitation_binaryPrecipitation:seasonSpring 199.9815   277.8873   0.720 0.472103
precipitation_binaryPrecipitation:seasonSummer 280.9691   274.3075   1.024 0.306234
precipitation_binaryPrecipitation:seasonAutumn 529.0395   277.8495   1.904 0.057522 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 843.1 on 464 degrees of freedom
Multiple R-squared:  0.8803,    Adjusted R-squared:  0.8713
F-statistic: 97.49 on 35 and 464 DF,  p-value: < 2.2e-16
```

Output 4: Advanced model after dealing with outliers (excludes statutory holiday and stepwise deleted variables)

```
Call:
lm(formula = Employees_in_office ~ temperature_cen + wind_cen + 
    darkness + precipitation_binary + snow_binary + School_holiday + 
    Weekday + season + Pol + season * temperature_cen + season * 
    precipitation_binary, data = Emp_cook)

Residuals:
     Min       1Q   Median       3Q      Max 
-1462.14  -205.40    25.73   239.63  1011.82 

Coefficients:
                                             Estimate Std. Error t value Pr(>|t|)    
(Intercept)                                11861.6795   108.1889 109.639  < 2e-16 ***
temperature_cen                               36.5105    10.5000   3.477 0.000557 ***
wind_cen                                      -2.2499     3.6042  -0.624 0.532787    
darkness                                       0.9351     1.0826   0.864 0.388212    
precipitation_binaryPrecipitation           -487.8231    96.7536  -5.042 6.74e-07 ***
snow_binarySnow                             -269.5775   108.4615  -2.485 0.013307 *  
School_holidayAutumn break                 -1589.4288   140.6186 -11.303  < 2e-16 ***
School_holidayChristmas break              -5187.0094   188.9451 -27.452  < 2e-16 ***
School_holidayMay break                    -1459.6970   132.9105 -10.983  < 2e-16 ***
School_holidaySpring break                 -2186.4361   142.9051 -15.300  < 2e-16 ***
School_holidaySummer break                 -1794.9359    77.3902 -23.193  < 2e-16 ***
WeekdayTuesday                               821.0593    58.9068  13.938  < 2e-16 ***
WeekdayWednesday                             127.0975    58.6175   2.168 0.030672 *  
WeekdayThursday                              576.2493    59.9914   9.606  < 2e-16 ***
WeekdayFriday                              -3003.5098    59.9649 -50.088  < 2e-16 ***
seasonSpring                                -579.9437   105.8336  -5.480 7.17e-08 ***
seasonSummer                                -861.7257   138.8413  -6.207 1.25e-09 ***
seasonAutumn                                -429.8285   128.3758  -3.348 0.000883 ***
PolBus freq                                  -95.7978   123.8414  -0.774 0.439609    
PolBus freq + shed                          -276.1602    49.1431  -5.620 3.39e-08 ***
PolBus freq + shed + No e-bike              -982.7316   128.5907  -7.642 1.34e-13 ***
PolBus freq + shed + drop                   -593.7307    75.5255  -7.861 2.93e-14 ***
PolBus freq + shed + drop + 35ct            -528.5474   109.6284  -4.821 1.97e-06 ***
temperature_cen:seasonSpring                 -24.9465    13.2281  -1.886 0.059967 .  
temperature_cen:seasonSummer                  -8.9968    17.7744  -0.506 0.612994    
temperature_cen:seasonAutumn                 -32.5913    13.8016  -2.361 0.018638 *  
precipitation_binaryPrecipitation:seasonSpring 400.4185 132.8711   3.014 0.002730 ** 
precipitation_binaryPrecipitation:seasonSummer 285.8322 129.6513   2.205 0.027995 *  
precipitation_binaryPrecipitation:seasonAutumn 305.2938 131.2400   2.326 0.020458 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 400.2 on 442 degrees of freedom
Multiple R-squared:  0.9497,	Adjusted R-squared:  0.9465 
F-statistic: 298.2 on 28 and 442 DF,  p-value: < 2.2e-16
```

Output 5: ANOVA comparing basic and advanced employees in office model

```
Analysis of Variance Table

Model 1: Employees_in_office ~ temperature_cen + wind_cen + darkness +
    precipitation_binary + snow_binary + School_holiday + Weekday +
    season
Model 2: Employees_in_office ~ temperature_cen + wind_cen + darkness +
    precipitation_binary + snow_binary + School_holiday + Weekday +
    season + Pol + season * temperature_cen + season * precipitation_binary
  Res.Df      RSS Df Sum of Sq      F    Pr(>F)
1    453 97059500
2    442 70790154 11  26269346 14.911 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Output 6: Advanced model with HAC standard errors

```
z test of coefficients:

                                              Estimate  Std. Error  z value  Pr(>|z|)
(Intercept)                                11861.67946   126.01427  94.1297 < 2.2e-16 ***
temperature_cen                               36.51045     8.01863   4.5532 5.284e-06 ***
wind_cen                                      -2.24989     4.94011  -0.4554 0.6487973
darkness                                       0.93509     1.77337   0.5273 0.5979898
precipitation_binaryPrecipitation           -487.82307    82.54713  -5.9096 3.429e-09 ***
snow_binarySnow                             -269.57751   108.76232  -2.4786 0.0131902 *
School_holidayAutumn break                 -1589.42881    70.53263 -22.5347 < 2.2e-16 ***
School_holidayChristmas break              -5187.00937   238.46935 -21.7513 < 2.2e-16 ***
School_holidayMay break                    -1459.69703    80.29962 -18.1781 < 2.2e-16 ***
School_holidaySpring break                 -2186.43610   107.90668 -20.2623 < 2.2e-16 ***
School_holidaySummer break                 -1794.93588   239.74286  -7.4869 7.051e-14 ***
WeekdayTuesday                               821.05933    35.29595  23.2621 < 2.2e-16 ***
WeekdayWednesday                             127.09751    39.63086   3.2070 0.0013411 **
WeekdayThursday                              576.24925    45.38360  12.6973 < 2.2e-16 ***
WeekdayFriday                              -3003.50984    46.03955 -65.2376 < 2.2e-16 ***
seasonSpring                                -579.94368   134.95787  -4.2972 1.730e-05 ***
seasonSummer                                -861.72574   171.95493  -5.0113 5.405e-07 ***
seasonAutumn                                -429.82854   114.87512  -3.7417 0.0001828 ***
PolBus freq                                  -95.79780   150.37797  -0.6370 0.5240944
PolBus freq + shed                          -276.16020    90.49301  -3.0517 0.0022753 **
PolBus freq + shed + No e-bike              -982.73155   272.01356  -3.6128 0.0003029 ***
PolBus freq + shed + drop                   -593.73068   121.85035  -4.8726 1.101e-06 ***
PolBus freq + shed + drop + 35ct            -528.54745    63.57792  -8.3134 < 2.2e-16 ***
temperature_cen:seasonSpring                 -24.94649    17.17566  -1.4524 0.1463814
temperature_cen:seasonSummer                  -8.99677    19.52655  -0.4607 0.6449810
temperature_cen:seasonAutumn                 -32.59133     8.68344  -3.7533 0.0001745 ***
precipitation_binaryPrecipitation:seasonSpring 400.41851 112.74294   3.5516 0.0003829 ***
precipitation_binaryPrecipitation:seasonSummer 285.83222 127.54354   2.2411 0.0250224 *
precipitation_binaryPrecipitation:seasonAutumn 305.29381 107.83889   2.8310 0.0046400 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

F Model assumptions

Model assumptions are compared for the basic model before dealing with influential observations and after. The axis are scaled, but it can see that residuals in the linearity graph for example are higher in before compared to after. For bicycle, this is done with the proxy models, while it was not possible to retrieve graphs for pooled models. An important note: The influential observations graphs and the collinearity graphs need to be disregarded. For influential observations Cook's distance is used in this research, and because of interactions which the used R package does disregard, Variance Inflation is higher than in reality in these collinearity graphs. In F.2 the Variance Inflation Factors are discussed.

F.1 Visual inspection

Bike before

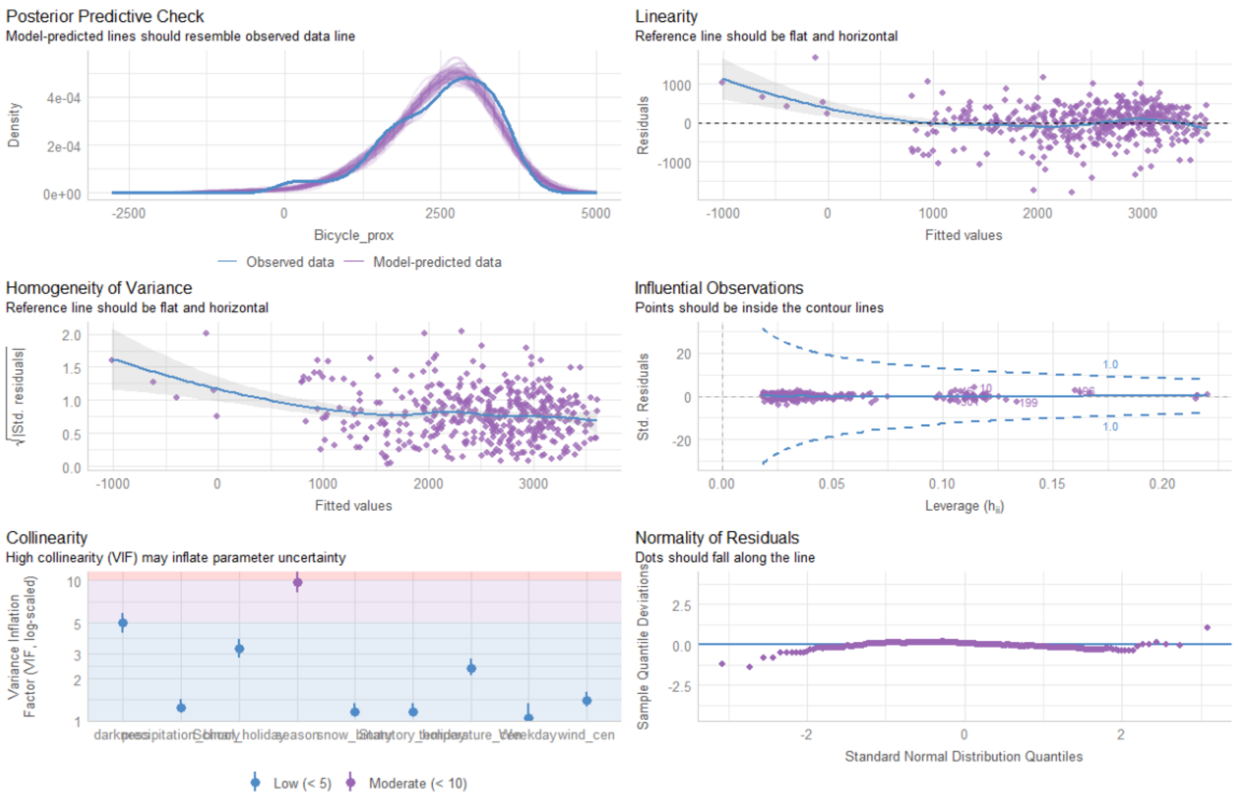


Figure F1

Bike after

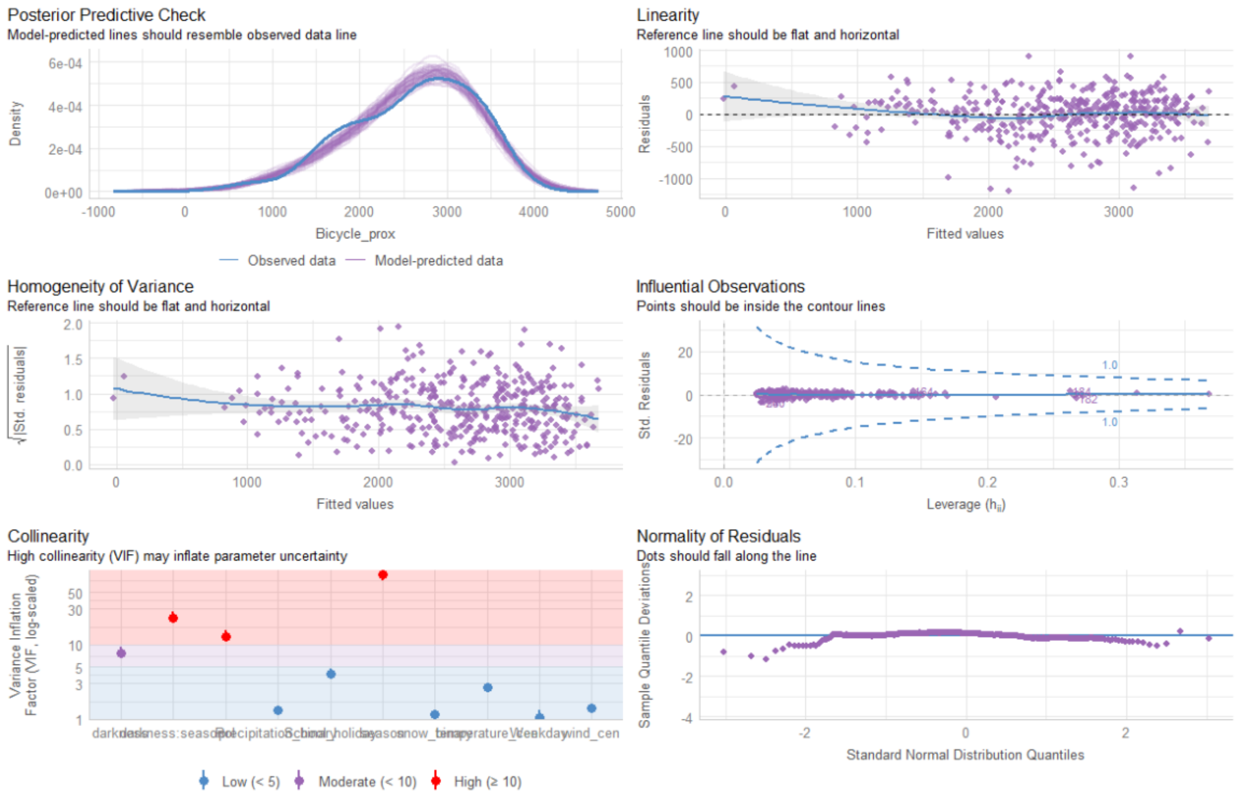


Figure F2

Public transport before

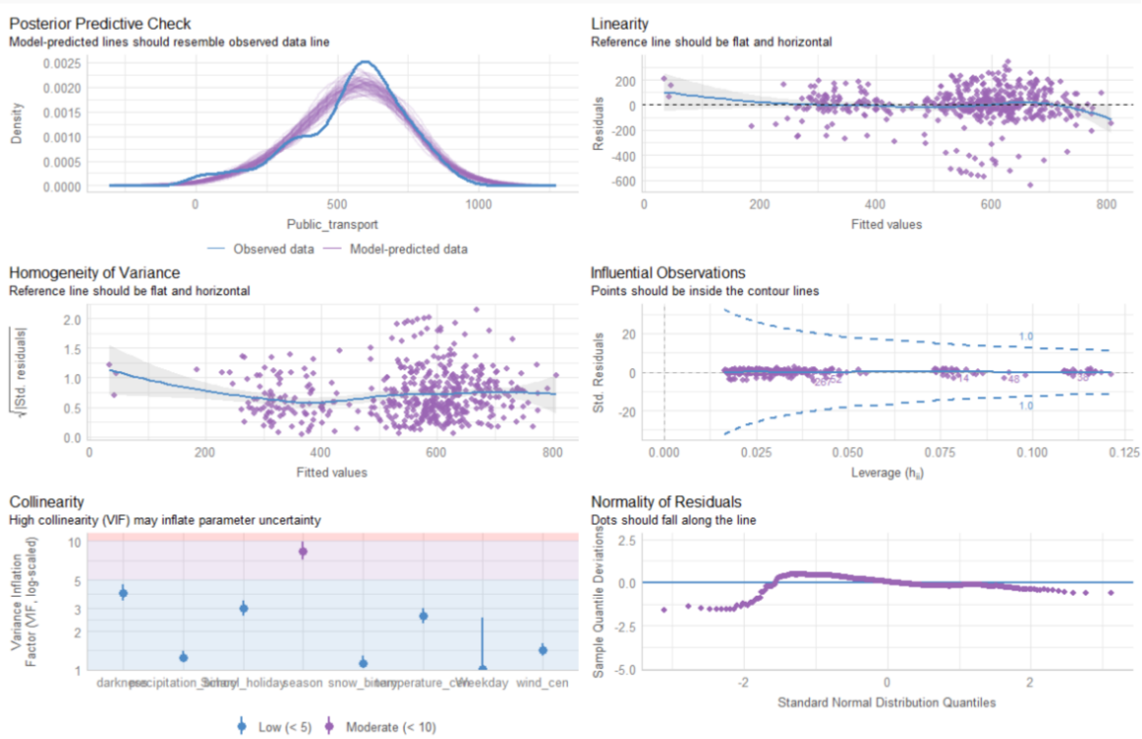


Figure F3

Public transport after

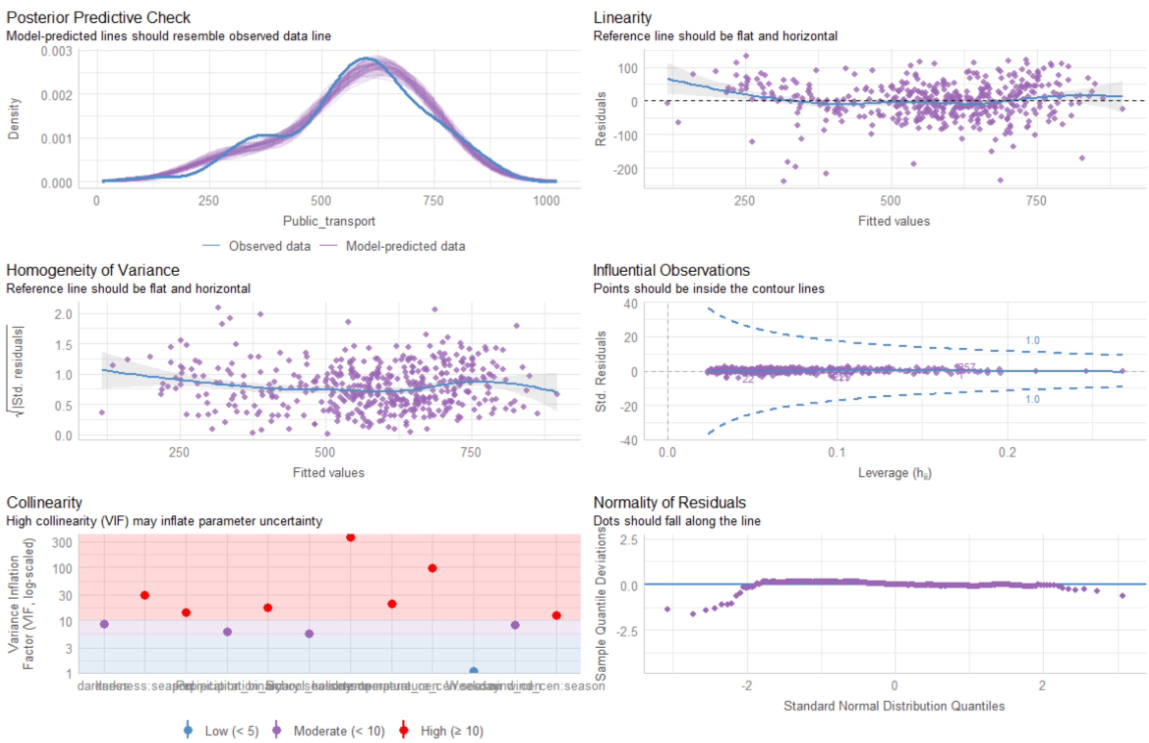


Figure F4

Car before

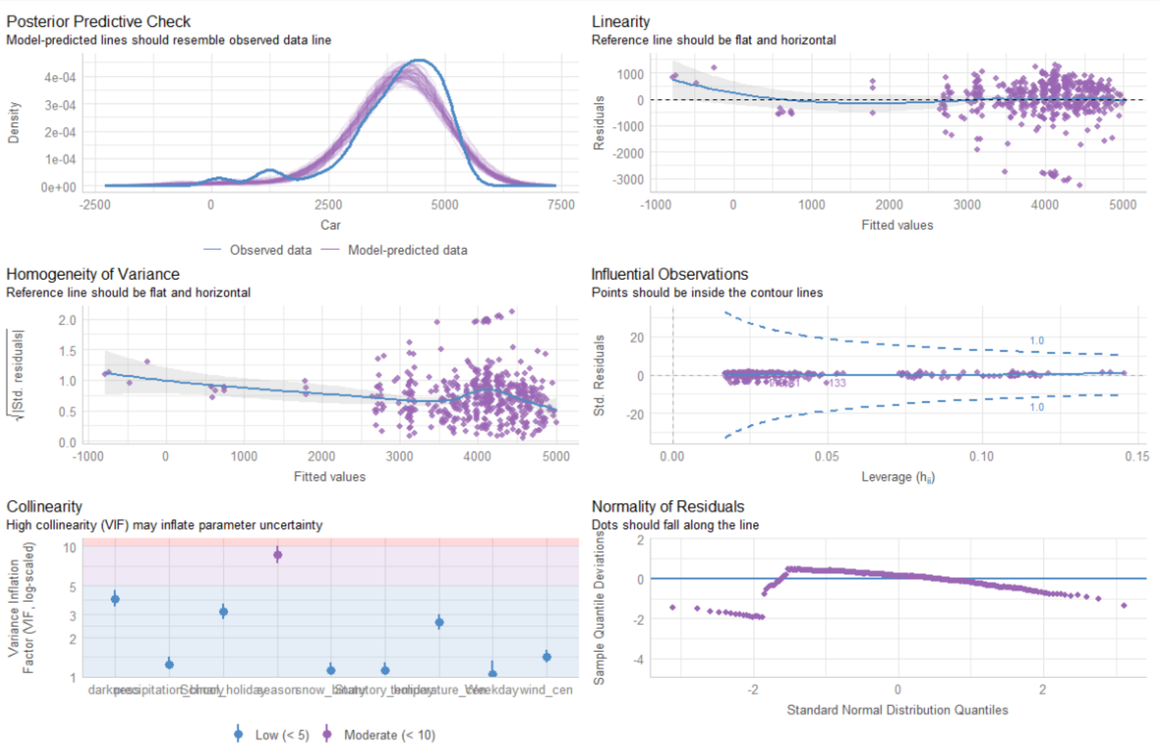


Figure F5

Car after

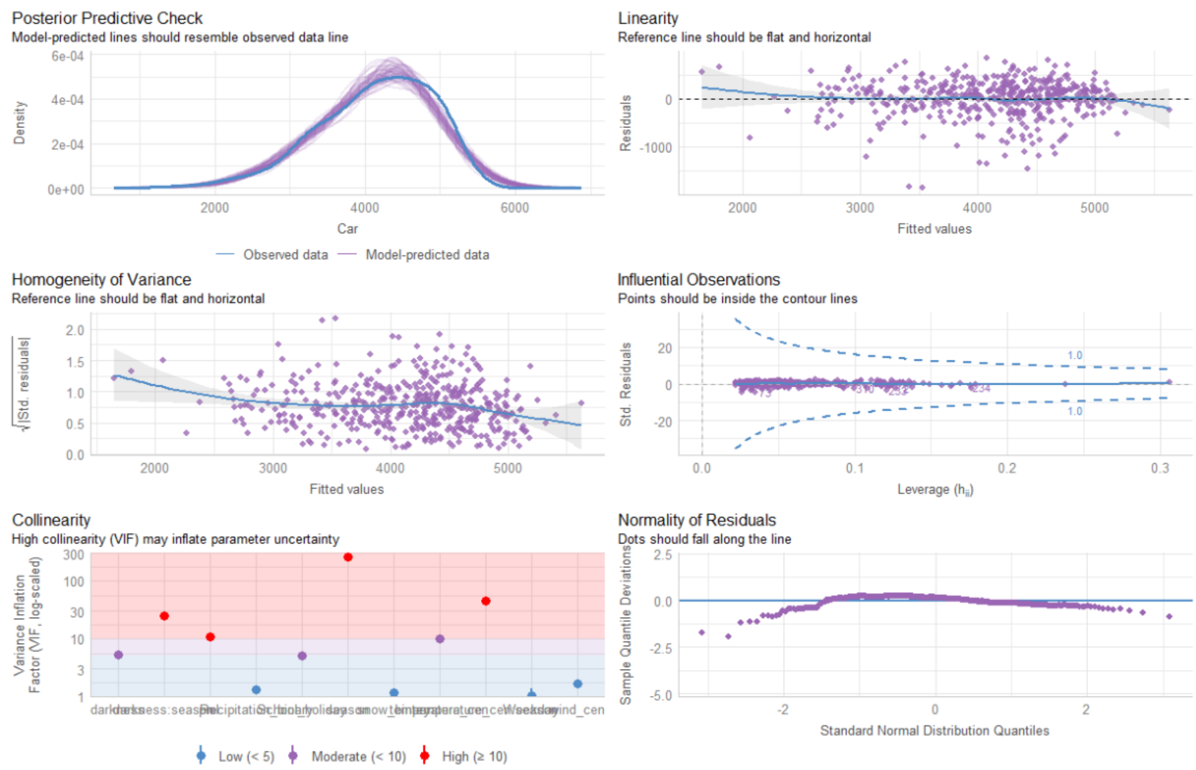


Figure F6

Employees in office before

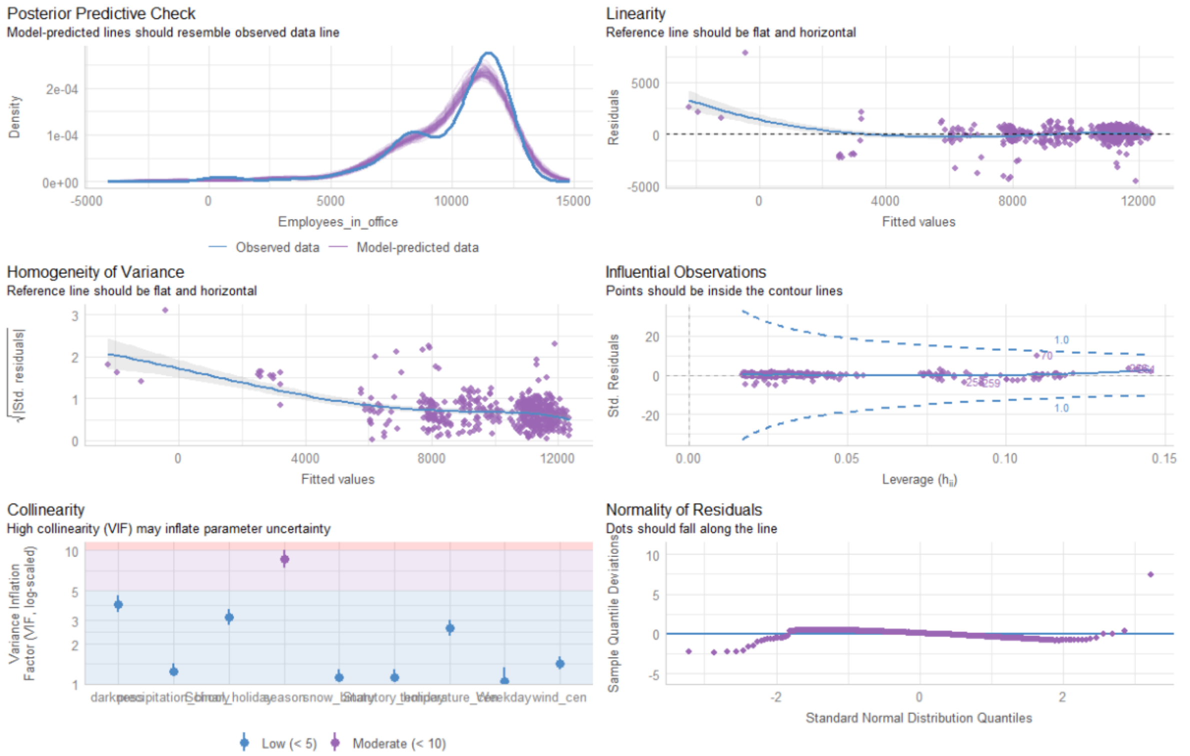


Figure F7

Employees in office after

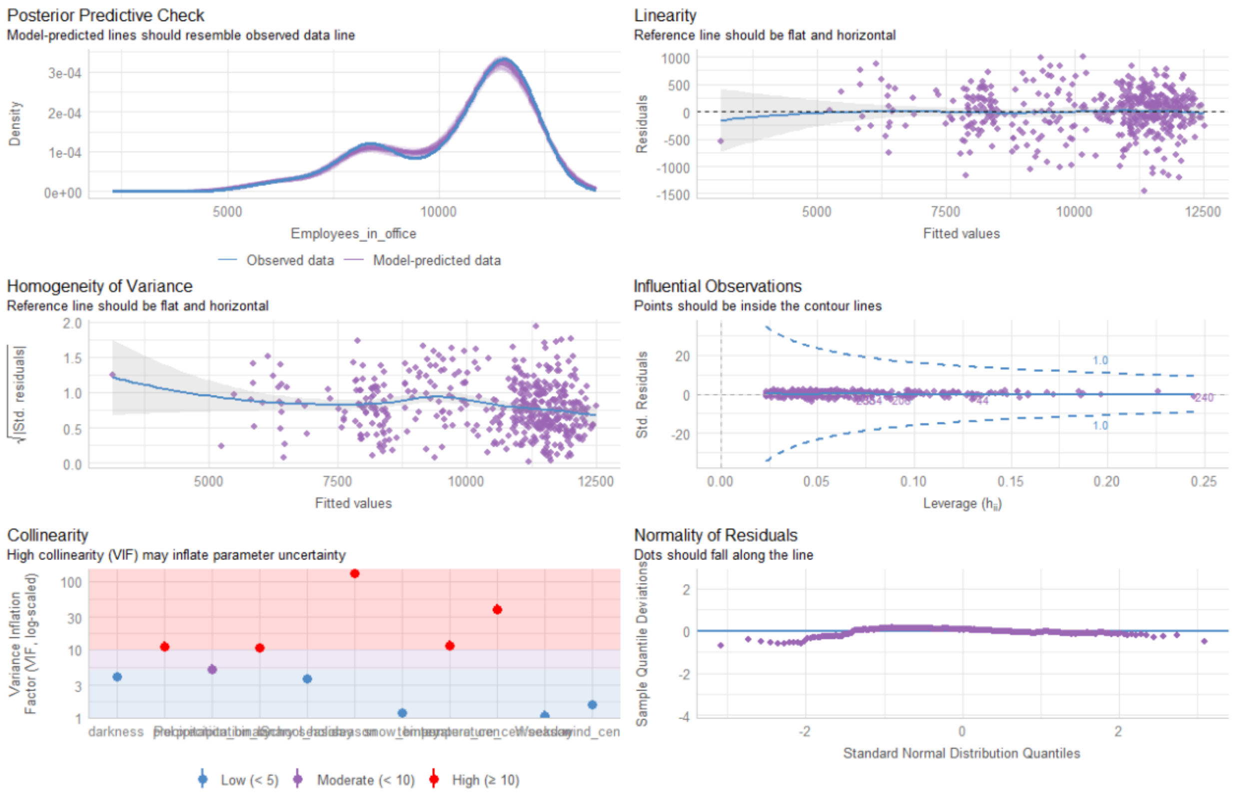


Figure F8

F.2 Variance Inflation Factors

As entailed in the thesis, the centred variable of employees was first regarded as predictor. But, while the Variance Inflation Factor (VIF) was quite high, the variable together with seasons and policies inflated the VIF. Therefore, to deal with multicollinearity a variable was needed to be deleted. Seasons are needed for interaction terms, and policies are of interest for the research. Therefore, the choice fell upon deletion of the centred employees variable.

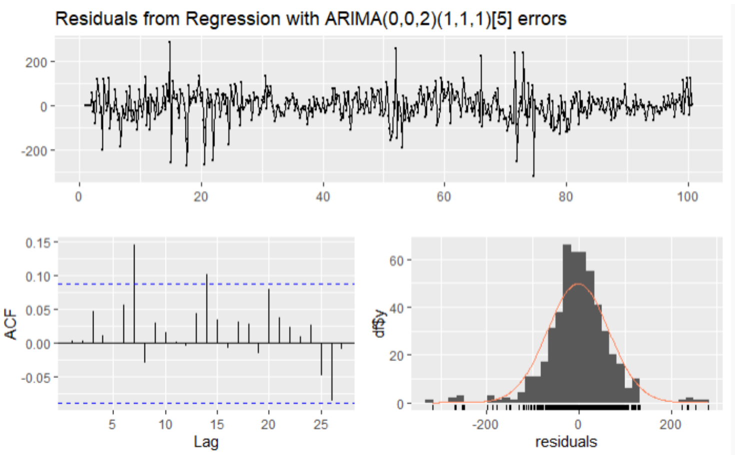
	GVIF	Df	GVIF^(1/(2*Df))	Interacts With
	<dbl>	<dbl>	<dbl>	<chr>
temperature_cen	170.915071	7	1.443724	season
wind_cen	1.559375	1	1.248749	--
darkness	4.115088	1	2.028568	--
precipitation_binary	269.377546	7	1.491411	season
snow_binary	1.204988	1	1.097719	--
School_holiday	3.782822	5	1.142304	--
Weekday	1.076684	4	1.009279	--
season	160.505986	11	1.259649	temperature_cen, precipitation_binary
Pol	47.072169	5	1.469862	--
Employees_cen	6.436513	1	2.537028	--

	GVIF	Df	GVIF^(1/(2*Df))	Interacts With
	<dbl>	<dbl>	<dbl>	<chr>
temperature_cen	81.135861	7	1.368902	season
wind_cen	1.553369	1	1.246342	--
darkness	4.006105	1	2.001526	--
precipitation_binary	162.339738	7	1.438426	season
snow_binary	1.203605	1	1.097089	--
School_holiday	3.699043	5	1.139748	--
Weekday	1.076178	4	1.009219	--
season	70.510347	11	1.213421	temperature_cen, precipitation_binary
Pol	11.108944	5	1.272235	--

Figure F9

F.3 Residuals of SARIMAX

The Autocorrelation Function (ACF) has in the ACF plot still lags surpassing the blue dotted line, meaning still not all autocorrelation is captured by the model, leading to a failing Ljung-Box test. The residuals seem reasonably well distributed, while a normal distribution can be seen.




```
	Ljung-Box test

data:  Residuals from Regression with ARIMA(0,0,2)(1,1,1)[5] errors
Q* = 14.561, df = 6, p-value = 0.02396

Model df: 4.   Total lags used: 10
```

Figure F10

G Interpretation of models

G.1 OLS regression models

Coefficients: mean change in the response variable for a unit change in a predictor variable while holding all other predictors in the model constant. A positive coefficient means that the response variable (e.g. bike commuters), increases when there is an increase in an independent variable. Vice versa a negative coefficient depicts a decrease in the response variable when the independent increase. For binary variables and categorical variables, it is change compared with the reference category. The coefficients of interaction effects depict whether there is an additional effect of a specific weather condition in a different season than baseline.

Standard errors: the standard errors are depicted with 'SE' or between parentheses is a measure of precision of the coefficients. Smaller standard errors depict more confidence in the coefficient while a larger standard error represents less confidence. Standard errors exist while a model is a simplification of reality and also the observations can lead to different estimates for the coefficients The standard error can be used to find the confidence intervals of how sure it is the estimate is within a certain range.

p-values: These values help with understanding if results are statistically significant at a certain level. This can be translated to examine whether relationships happen by change or can be real. A common applied threshold is 0.05, with lower p-values reject the hypothesis that there is no relationship and higher p-values failing to reject the hypothesis of not having a relationship between an independent variable and the response variable. If a threshold of 0.05 is used, only coefficients that have at least one '*' behind the estimates in the tables successfully reject the hypothesis of having no relationship between variable and result. A common misunderstanding is that researchers use it as a probability of the null hypothesis being true. If a person wants to see it as a probability, it is the conditional probability that when the null hypothesis were true, the p-value is the probability of obtaining results as the one found with the regression.

With the coefficients, it can be predicted what the travel behaviour would be on a particular day. A cautionary note is warranted: when wanting to compare for example winter with spring, the coefficient of spring would be the mean difference with winter if all other variables remain the same. While for example darkness is a seasonality variable to compare winter with spring it should be considered in which weather conditions. Only using the coefficient for spring is therefore the effect for example relative to winter with zero darkness, mean year-round temperature and mean year-round wind speed while these two variables are centred. In winter for example, in reality there are just 12 days of the 119 with zero darkness.

G.2 Multinomial Logistic Regression

In Multinomial Logistic Regression, next to references for individual variables, the model as a whole has also a reference category. In this research this is commuting by car. The estimates are log-odds which can be converted to odds ratios. If the odds ratio is above 1, this means the variable increases the odds for a mode of transport in comparison with taking the car. A odds ratio lower than 1 is therefore a decrease in the odds for the mode of transport in comparison with car when all other variables remain constant.

G.3 SARIMAX

Coefficients in the SARIMAX model offer insights in the direction of relationships but fully quantifying the influence is particularly challenging due to model complexity and the presence of autoregressive parts and moving averages. The coefficients therefore can be used to see the strength and the direction rather than the more precise causal effects.