Testing the relationship between socio-economic inequalities and accessibility in the Paris Urban Area

Author: Yvain REDON

Date: February 10, 2026

TU Delft

Testing the relationship between socio-economic inequalities and accessibility in the Paris Urban Area

Using GTFS data to determine accessibility to a selection of key services and highlighting associations with socio-economic variables

By

Yvain REDON

In partial fulfilment of the requirements for the degree of:

Master of Science
in Transport, Infrastructure and Logistics

at the Delft University of Technology,
to be defended publicly on Tuesday the 24th, February 2026

Supervisor:	Dr. S. Balakrishnan	
Thesis committee:	Dr. J.A. Annema,	TU Delft
	Dr. C. Maat,	TU Delft
	Dr. S. Balakrishnan,	TU Delft

TU Delft

Acknowledgements

This thesis aimed to accurately study the links between accessibility by public transport and socio-economic inequalities. Indeed, while this thematic was quite commonly brought up during my engineering studies in France, I thought that new research using accessibility could bring a new outlook on this issue.

When I started this thesis project, almost a year ago, I had no idea that it would be such a challenging work. But while it was more difficult than I had anticipated, it also proved to be one of the most stimulating and interesting experiences that I encountered during my studies. This process allowed me to realize the many steps and all the work that is necessary to produce new scientific knowledge. Despite the many shortcomings of this thesis, I am genuinely grateful for having this opportunity to really focus on a specific subject, and I am proud to have contributed ever so slightly to this topic.

I encountered a lot of difficulties in making this thesis. At first, the project proposal was probably too ambitious to carry out entirely. Thus, the project was reorganized several times. In addition to these technical problems, I was also confronted with mental health issues that severely impacted the planning of the thesis. However, I am still happy to have been able to finally complete this project, even if I still have some regrets about what could have been improved or done better.

I would like to thank all the members of my supervising committee. While this project encountered many difficulties, I always felt supported by my supervisors, who never put me under too much pressure and worked to find a way forward that would make the process as stress-free as possible. I would like to thank Dr. S. Balakrishnan, who gave much of his time to this thesis, meeting with me weekly and always providing interesting and relevant insight and advice, that I unfortunately was not always able to follow. I do not exaggerate when I say that this thesis would not have been the same without his valuable input. I would also like to thank Dr. J.A. Annema, for all his valuable advice about research approach in general during our meetings, and for his involvement in the latter steps of the thesis, especially in providing the framework and the way forward to successfully make progress on the subject. I am grateful to both of them for providing the subject of the thesis and allowing me to work on it. I am also grateful to Dr. C. Maat, who despite arriving later in the supervising team was very interested in the thesis, and provided me with extensive and detailed feedback on the thesis deliverables and was an invaluable help in improving the original project proposal. They all helped tremendously despite the many problems and changes that happened during the thesis, and I will forever be grateful for their incredible support and patience.

Likewise, I would like to thank my friends and family for supporting me through the difficult phases of this thesis, and for helping me recover during the periods where I was at my lowest mentally. I am especially grateful to my mother, for always finding the right words to motivate me and make me persevere on this work. I am also forever thankful to my friend Louis, who was both a great help in terms of research and programming advice, but also supported throughout the last months of the thesis, helping me both work and relax. I hope I will be able to support him in the same way during his PhD.

Once again, I am infinitely grateful to all of you, I am very lucky to have been supported through tough times by such wonderful people.

Yvain REDON
Paris, February 2026

Table of Contents

Summary

Introduction

Since the development of industrial cities, inequalities tend to be geographically clustered in European cities, with many cities displaying a east/west divide, where poorer populations live to the east of the centre and the wealthier populations live to the west. In France, in particular, other spatial disparities could be worth studying. Indeed, after the second world war and due to rapid population growth, many districts were built, characterised by a high proportion of social housing. Several studies and census data have emphasised the differences between these districts often dubbed 'grands ensembles' or 'cités', and other suburbs. The literature has pointed out that these districts are often poorer, and face high unemployment, and are seen as having poor access to facilities.

This study aims to assess if inequalities of access to several types of facilities are indeed related to income and if 'cités' face specific challenges in access to these facilities. To that end, an accessibility metric is used, in order to consider both the impact of the transport system and the impact of the land-use system. The Paris Urban Area is chosen as a case study. The following research question is considered:

To what extent are inequalities in accessibility of multimodal transport networks related to existing inequalities in the Paris Urban Area?

Literature review

The literature review of this study mostly focuses on accessibility as a metric, and more specifically how it has been used to study inequalities. Studies on distributive justice in transportation have highlighted that accessibility is a very relevant indicator when studying inequalities. However, many different types of accessibility measures have been theorised and/or used in the literature, most notably cumulative accessibility measures and gravity-based accessibility measures. Insights from the literature show that choosing an accessibility measure to use involves striking a balance between interpretability and low data use on one hand, and accuracy and complexity on the other hand. Thus, some studies ignore effects such as competition effects, which can have an impact on results.

Studies also vary in scope, with some studies focusing on specific cities whereas other articles encompass entire countries in their analysis. Furthermore, studies in the literature have led to different findings depending on the case study, which seems to indicate that patterns of accessibility in cities are highly dependent on the context of the city.

Research gaps explored in the present study include taking into account competition effect to more accurately represent access to commodities such as jobs, considering both travel time and monetary travel cost, as well as studying access to multiple types of facilities at the same type.

Methodology

In this study, access to employment is considered, as well as access to primary and secondary schools, to higher education facilities, to general practitioners and hospitals, as well as access to pharmacies, convenience stores and supermarkets.

Accessibility and other variables are considered at the ‘iris’ scale. Iris are census tracts at the sub-city level designed to be homogeneous in terms of population size.

GTFS data obtained from the open data platform of French public institutions are used to model the network. This dataset contains information about the location of public transport stations, the timetable of all service lines at each station, and the type of service line (i.e. bus, metro, tramway, heavy rail, …).

Information about the characteristics of facilities is obtained through multiple datasets accessed using the public open data platform. These datasets are then filtered to focus on specific types of facilities. Socio-economic and demographic variables are also considered and are obtained through census data available on the INSEE website.

Based on the GTFS data, a travel-time matrix is computed using the r5py library on Python. This matrix contains the travel-time between each pair of iris area (the origin and destination are considered to be the centroid of the areas). Travel-times are computed during a weekday, with a departure time at 8AM.

The accessibility metric used in this study is Shen’s accessibility. This measure is gravity-based and thus needs to be calibrated. It is calibrated using travel survey data from the 2018-2020 period. This measure also considers competition effects. Competition effects are taken into account to avoid overestimating (in areas where there is high competition over access to facilities) or underestimating accessibility (in areas where competition is low). To take into account competition effects, indicators must be used to represent demand and supply. In this study, demand is generally the population that could use the facility, whereas supply is more specific to the type of facility.

One of the analysis methods applied to the results is k-means clustering. The number of clusters is determined using the elbow method. Different sets of clusters are built, with approximately one set of clusters being computed for each type of facility considered, though for some types of facilities multiple sets of clusters are created. Clusters are computed using several variables: accessibility metrics on one hand, and socio-economic variables on the other hand. These socio-economic variables are median available income, the ratio of social housing in the area, the ratio of households that own at least one car in the area, and population density.

The other analysis method applied to the results is spatial regression. It is done using a spatial-lag model. For spatial regression, less facilities are considered: only jobs and primary and secondary education. The spatial weights scheme used is the k-nearest neighbours scheme, with k being set to ten. For each spatial regression model, the dependant variable is accessibility, and the independent variables are the same socio-economic variables as for clustering.

Descriptive analysis

Maps of the Paris urban area are created, in order to show the shape of the public transport network, with one map showing heavy rail lines and one map showing subway and tramway service lines. These maps also display socio-economic variables such as the median available income or the poverty rate.

Further analysis is conducted. These analyses are performed to test the potential correlation between income and access to the public transport system. Two indicators are computed for each 'iris' area: the distance to the closest public transport station and the number of public transport stations within 2km of the area centroid. All public transport stations are considered for these indicators. Then, scatterplots are created comparing the value of these variables to median available income and poverty rate.

These analyses show large disparities: linear regressions performed show low R-squared values. Though income seems to be correlated with access, this relationship is not strictly linear, especially for average income areas.

Results

The results obtained from the different sets of clusters are quite coherent. Overall, several archetypes have been identified. First, two types of central areas: average income central areas and wealthy central areas. These areas have overall the best access of all clusters, with the wealthiest areas having even better access. These wealthy central areas are located mostly to the west of Paris itself and in some of the cities directly next to Paris. Average income central areas are mostly located in the northeast of Paris itself, and in some cities directly to the east of Paris. Areas in both of these clusters are characterized by very high population density and low car ownership.

Then, there is commonly two clusters corresponding to suburbs. Overall, one archetype describes below average income areas, and another corresponds to above average income areas. These areas are characterized by a high proportion of car ownership, average accessibility and average to low population density. Again, suburbs with below average income and more social housing tend to have less accessibility than more affluent suburbs. Areas with lower income are located to the northeast, east and south of Paris, whereas more affluent suburbs are mostly located to the west of Paris.

Then two clusters of areas have the lowest accessibility across most types of facilities. First, low density areas with average income. These areas are often located further from the centre than areas in other clusters. These areas have very high car ownership, low density, and poor access to all types of facilities. Another cluster of areas with poor accessibility seems to correspond to 'cités'. These areas are characterised by a very high proportion of social housing, the lowest income of all clusters, a quite high population density, and average car ownership. These areas tend to be more spread out than areas in other clusters, though they are mostly located in the northeast and parts of the south of the Paris Urban Area.

Results from the spatial regression models indicate that income positively contributes to accessibility to jobs and primary and secondary education. Conversely, the higher car ownership and the ratio of social housing, the lower the accessibility. Population density was found to be significant for jobs but not for education. In the case of jobs, it is associated with higher accessibility. Furthermore, spatial regression results confirm that accessibility is spatially correlated.

Conclusion and discussion

Results seem to indicate that generally, higher income levels are correlated with higher accessibility. Poorer areas tend to have lower accessibility compared to comparable areas with higher income levels. In general, areas with the lowest income tend to have low accessibility, and areas with the highest income levels tend to have very good accessibility. However, areas with about average income display more diverse situations. For less dense, more remote areas, accessibility is poor, whereas more central areas have comparable income figures, but much higher population density and accessibility.

There seems to be a specific challenge faced by 'cités' districts, as these neighbourhoods have very low accessibility even though they have quite high population density. Furthermore, car ownership is low in these areas compared to other types of areas with similar accessibility. This means that households in 'cités' cannot benefit from good accessibility through the public transport system and are less able to compensate bad public transport access with car travel.

However, results from this study are subject to several limitations. The calibration method used for the accessibility measure could lead to underestimating some inequalities. Furthermore, detailed travel survey data could be used to calibrate parameters for the accessibility measure more accurately.

For some facilities, more detailed datasets describing more characteristics of the facilities could lead to a more accurate representation of supply, for example for jobs and stores.

Furthermore, more detailed analysis would be necessary to identify precisely if poor accessibility in some areas is due to a lack of access to the public transport system, or a lack of facilities in the land-use system.

Besides, some characteristics of the transport system are ignored, such as the reliability of service lines. If some service lines are less reliable and are more prone to delay and cancellations, this could lead to other inequalities of access ignored in the study. Moreover, different departure times may be considered, to better represent access to part-time jobs or to jobs with nigh shifts.

Policy wise, the inequalities of access highlighted in this study should be taken into account when designing future public transport service lines. In an egalitarian perspective on distributive justice, authorities should aim for a more equal repartition of accessibility through the urban area. One objective could be to even out accessibility levels so that areas with comparable population density have comparable accessibility.

It remains to be seen if future planned extensions of the network such as ‘Grand Paris Express’ will address these inequalities of access, or if on the other hand they will reinforce inequalities.

1. Introduction

1.1. Background

In many French cities, inequality in terms of wealth and unemployment, education, and access to health and culture is highly geographically clustered. Indeed, since the development of industrial cities, already existing inequalities have started to be more and more by spatial separation, and districts have become more homogeneous in terms of income, professional categories and other indicators. For example, as a result of industrialization in Europe, rich populations in many European cities have chosen to the west of the cities, where pollution was less severe, whereas working class populations lived closer to factories.

In France, due to the destructions caused by the second world war, there was a need to reconstruct the district and cities that had been destroyed. Another reason for the fast construction of new housing and districts was the population growth that took place between 1946 and 1974 in France, known as the ‘baby boom’. This also increased the need for public services, especially education, as both the population and the mean duration of education increased. Finally, another factor that led to building of new cities and districts was decolonization, especially the end of the Algerian independence war: this led to the arrival of one million ‘pieds-noir’ (French population of Algeria) over a very short period. Thus, due to different factors, the French metropolitan population increased by a little more than 10 million between 1950 and 1974 (INED).

Though these cities and districts built in the 30 years after the war were supposed to be planned with access to public services in mind, however this was not always the case in reality, as is explained by Lévêque (2023) in the case of the city of Vaulx-en-Velin.

Over the years, these urban spaces, often dubbed ‘*banlieues*’, ‘*cités*’ or ‘*grands ensembles*’, have gained a reputation for poor living conditions. Indeed, taking as an example the *département* of Seine Saint-Denis, north of Paris, can help in representing the characteristics of these *banlieues*. Using census data from INSEE, it is noticeable that this territory had a poverty rate of 28,4 % in 2021, almost twice as high as the general poverty rate in metropolitan France. Furthermore, the unemployment rate is also higher, at 16,4%. In addition, the fact that cities in this area developed over the last century is shown by the fact that only 3.5% of housing units were built before 1919. These inequalities may be correlated with inequalities in accessibility to key services, compared to other parts of an urban area (which can be the centre of the urban area, or other cities in the urban area). Thus, it can be interesting to study exactly how accessibility (to a selection of key services) is distributed spatially in a French urban area, and if there is a significant correlation between the income and poverty of an area, and its access.

1.2. Aim of the study

In this context, in order to have a better grasp on the spatial distribution of inequalities in the context of French urban areas, this study aims to evaluate accessibility to a selection of key services and see if the accessibility levels are correlated with income and other socio-economic indicators.

To conduct this study, the city of Paris and its urban area was chosen.

This research aims to contribute to knowledge that can counter inequalities in accessibility to (1) work locations, and (2) daily facilities (e.g. supermarkets and convenience stores, pharmacy, medical facilities). Results about accessibility will be useful in determining if the transport network contributes to the reduction of inequalities or if on the other hand it reinforces these inequalities. Indeed, this question is important, as depending on the context of a city, the literature has shown differing results. Indeed, in some cases low-income households benefit from higher accessibility levels than high-income households, whereas in other cases accessibility and income are positively correlated. Furthermore, another aim is to see if the spatial repartition of inequalities and transport access in Paris is similar to other cities worldwide.

1.3. Urban area used as a case study

The urban area which is studied in order to achieve the research aims is the urban area of Paris.

There are multiple advantages in using Paris for this study. First of all, the transport network of the Paris urban area is the most developed in France, meaning that it is the best suited in France to study effects linked to the presence of a large choice of public transport modes.

Another perk of studying Paris is that it allows studying a population with a lot of heterogeneity regarding socio-economic variables: there are households among the richest in France, as well as very poor households for example, different types of housing, and employment wise, though the service sector is dominating, other fields in the secondary sector such as construction work are represented.

Finally, since only accessibility by walk, bike and public transport is considered in the study, Paris is an adequate city as these modes represent a large part of trips performed in the urban area, a part which of course varies depending on the exact location and type of trip. For example, public transport is the most used mode when commuting to work.

1.4. Research question

Given the research gaps identified, the approach of this thesis aims to consider competition effects and the impact of monetary cost. Thus, in this thesis, accessibility is described with generalised travel time as the cost of access to a location. Generalized travel time encompasses the disutility of travel time, taking into account the monetary cost of travel such as fares. GTFS data is used to represent the public transport network and to estimate the travel time between locations. The outcomes are evaluated on the criteria of egalitarianism. Results are analysed using clustering and spatial regression methods.

To this end, the following question has been drawn up:

Are socio-economic inequalities in the Paris Urban Area compounded by additional accessibility inequalities?

Three sub-questions are formulated:

- How are inequalities of access distributed, both geographically and among the population?
- Are inequalities of access associated with existing inequalities in terms of income, poverty rate, unemployment, etc ?
- Do the land-use and transport systems in the Paris urban area reduce or amplify existing inequalities?

The general approach is the following: GTFS data is used to represent the multimodal transport network of Paris and its urban area. Using this representation, accessibility to employment and key locations, such as supermarkets and convenience stores, primary and secondary education, higher education and hospitals are calculated. Accessibility is then compared with existing socio-economic characteristics, using clustering methods and spatial regression.

1.5. Outline of the report

This report aims to present this research. First, after this introduction, the first part of this report is a literature review were different accessibility metrics and use cases across the literature are compared, and a research gap is identified. The second part of the report then describes briefly the aims of the research, as well as the main and sub research questions. Then, the third part of the report is dedicated to the methodology of the study, including the data used, and the type of accessibility metric used. The fourth part of the report shows some of the preliminary results obtained before computing accessibility. Then, the fifth part contains the main results and analysis of the study. Finally, these results are then used for the conclusion of the paper and are discussed.

2. Literature review

2.1. Methodology

This literature review was conducted with several objectives. First, to identify how accessibility is defined by various authors. Secondly, to find and compare the different forms that accessibility indicators can take. Finally, this literature review also aims to establish how accessibility can be used to better represent equity issues related to the transportation system. In particular, in the context of the study, urban public transport systems are focused on specifically in this literature review. Indeed, focusing on cities allow to compare socio-economic variables in a similar context, which allow a more accurate representation of inequalities.

This literature review is composed of articles that focus on the theory of accessibility, whereas other articles are more practical studies, usually focusing on a specific city and using an accessibility indicator. Most of the articles were published after the year 2000. Indeed, studies that focused specifically on public transport were rare before this date.

Most of the articles presented in this literature review were selected after searching for articles related to accessibility in public transport networks on websites such as Scopus. Furthermore, particular attention was given to articles citing or cited by already selected articles. In addition, some articles were suggested directly by supervisors of this research project.

In total, 22 articles have been reviewed in this literature review, as well as one book chapter. The selected articles are the following: Geurs and Van Wee (2004), Ermagun and Tilahun (2020), Mavoa et al. (2012), Fan et al. (2010), Verduzco Torres and McArthur (2024), Bouzouina et al. (2021), El-Geneidy et al. (2016), Fransen et al. (2019), Levinson and Wu (2020), Foth et al. (2013), Ford et al. (2015), Pereira et al. (2017), Bocarejo S. and Oviedo H. (2012), Radzimski (2023), Shen (1998), Hu et al. (2017), Giannotti et al. (2022), Adorno et al. (2025), van Wee and de Jong (2023), Aberle and Getz (2025), Lucas et al. (2016), Lucas (2012). The book chapter that was selected is Martens et al. (2019).

2.2. Presentation and analysis

2.2.1. Ethical underpinnings

According to Pereira et al. (2017), even though a lot of articles about transport planning consider the importance of equity and justice in transportation, few researchers explicitly state what theory of justice they are using in their study, or if there is a underlying theory of justice guiding the direction of their research. In the opinion of the authors, ‘This lack of conceptual clarity makes it difficult to compare the findings from different studies and to obtain insights that could inform policy decisions’.

Indeed, among the articles that were reviewed, few explicitly state which theory of justice or underlying ethical principles are used to evaluate their findings and to estimate how (un)equal accessibility is.

Though not explicitly linked to a specific theory of justice, some articles still present their understanding of equity in the context of the transport system and accessibility. For example, in Ermagun and Tilahun (2020), the authors study inequities of accessibility in the case of Chicago. In the paper, the authors seem to have an egalitarian conception of justice: to them, the inequalities of access between different socio-economic groups highlighted by their study need to be addressed by transit agencies. Though this may seem natural, this assertion would not be in line with, for example, a libertarian vision of justice.

One of the challenges when working based on a specific theory of justice, is that it may be hard to find relevant indicators that can measure 'equity' and 'justice' efficiently. In the case of egalitarianism and sufficientarianism, Lucas et al. (2016) showed that the gini index can be used to evaluate the distribution of accessibility. Furthermore, since the gini index is widely used in many statistical studies, it allows for comparison with the repartition of income, poverty, and other indicators. Thus, accessibility can be analysed in a more global, egalitarian perspective.

Though using an already existing and explicitly stated theory of justice does not affect the validity of the results, it can help in multiple aspects. It informs readers about how the researchers have interpreted their results and allows for easier comparison between articles and their conclusions.

2.2.2. Definition of accessibility

In order to assess inequalities in the transport system, accessibility has often been used. Accessibility is generally agreed to be linked to both the transport infrastructure and the land-use system. However, several different definitions have been proposed over the year. One of the first proposed definitions, cited by multiple articles considered here, is the definition of Hansen (1959), where accessibility is 'the potential of opportunities for interactions'. Since then, other definitions have been proposed, for example in Geurs and Van Wee (2004) : 'accessibility measures are seen as indicators for the impact of land-use and transport developments and policy plans on the functioning of society in general'. In Levinson and Wu (2020), accessibility is described more simply as 'the ease of reaching valued destinations'. Thus, accessibility is widely used as an indicator that can describe interactions between the transport infrastructure on one hand and the land-use distribution on the other hand.

In their article, Geurs and Van Wee (2004) identify four different components of accessibility. In addition to the already mentioned 'land-use component' and 'transportation component', the authors also identify the 'temporal component', and the 'individual component'. The 'temporal component' denotes the differences in availability over time of activities, whereas the 'individual component' are the characteristics of individuals that influences their access to certain activities, jobs and modes of transportation. This shows that accessibility is not limited to measuring the ease of travelling from one location to another in order to access activities using a transport system. Indeed, it can also be extended to measure inequalities between individuals.

In their article, Geurs et Van Wee (2004) point out the limitations of ‘Infrastructure based accessibility measures’, as these measures are often easily interpretable but lack accuracy and completeness. Indeed, other articles have argued for more generalized approaches to accessibility, such as, of course, Levinson and Wu (2020) who, in a theoretical approach, conceptualize ‘general access’ which encompasses ‘all places, all modes, all purposes, at all times, over the lifecycle of a project’. According to Pereira et al. (2017), accessibility should be conceptualized as a ‘human capability’ in order to be an accurate indicator to represent equity in transportation.

Thus, theoretical definitions of accessibility allow for many use cases of this indicator, not limited to only a performance measure.

2.2.3. Types of indicators

Across all the articles reviewed, different accessibility indicators have been used. Most indicators used in these articles belong to the ‘Location-based accessibility measured’ as defined by Geurs and Van Wee (2004). One commonly used measure is the ‘contour measure’ or ‘cumulative accessibility measure’. This measure describes the number of opportunities of a certain activity type reachable given a certain constraint. This constraint can be distance, travel time, or another variable. This indicator is widely used in studies studying accessibility to jobs, such as Verduzco Torres and McArthur (2024) (for some activities), Ermagun and Tilahun (2020), El-Geneidy et al. (2016), Fan et al. (2010) or Radzimski (2023). This indicator has the advantage of being easy to interpret, being the number of jobs reachable from a certain location given a constraint.

However, several criticisms have been made regarding this indicator. First of all, as mentioned by Geurs and Van Wee (2004) and Foth et al. (2013) for example, is that such indicators do not represent the fact that individuals may prefer opportunities that are closer to heir home to an opportunity that is farther away, even if both opportunities are under the threshold. Another criticism, particularly important for studies on accessibility to employment, is that this indicator does not represent the fact that different individuals may have different capabilities in which jobs they can access, as explained by Fransen et al. (2019).

Thus, in order to represent the preference for closer locations, ‘potential accessibility measures’ or ‘gravity based’ measures are used. In these measures, the cost function generally takes the form of a negative exponential:

$$f(c_{ij}) = e^{-\beta c_{ij}}$$

Where β is a coefficient, and c_{ij} the cost of accessing an opportunity in zone *j* from zone *i*. Thus, for this type of indicator, an opportunity that needs a larger cost to access (generally travel time) will contribute less to accessibility. It is used by Foth et al. (2013), Bouzouina et al. (2021), Bocarejo S. and Oviedo H. (2012).This indicator is useful in showing differences in access cost that may have been hidden by simply using a threshold. However, it is less easily interpretable and still has other disadvantages of location-based accessibility measures. Indeed, it still does not take into account the fact that some opportunities may be unavailable for some individuals, depending on their characteristics. Furthermore, Giannotti et al. (2022) has pointed out that, depending on the process chosen for the calibration of the parameters of the exponential decay curve, inequalities of access can be underestimated, and results can be inaccurate, especially when using parameters specific to a some groups.

In Fransen et al. (2019), the authors develop a more complex accessibility indicator in order to take into account the specific characteristics of individuals (such as education level) when studying accessibility to employment in the context of Flanders in Belgium. Though this kind of indicators are more accurate than other mentioned above, a key limitation pointed out by Geurs and Van Wee (2004) is that such approaches require much more data than other approaches. Another potential limitation pointed out in Mavoa et al. (2012) is the high computational requirements of studies using accessibility measures on the individual level.

One of the key weaknesses of potential accessibility measures is that competition effects are not well represented. This can lead to inaccuracies, as pointed out in Geurs and Van Wee (2004), especially when studying access to jobs. Indeed, one individual may have access to a lot of job opportunities in terms of travel time, but if these opportunities are in high demand, then the real access to jobs is lower than would be expected without competition effects. To take these effects into account, Shen (1998) introduced a modified potential accessibility measure which takes into account the interaction of supply and demand, this measure being referred as Shen's accessibility in the literature. Hu et al. (2017) for example have used Shen's accessibility to measure the evolution of accessibility to jobs in Beijing. However, it is still susceptible to the problems highlighted in Giannotti et al. (2022).

Thus, even though many accessibility measures have been theorised, many articles use the same kind of indicators, as more complex measures often lack interpretability and have high requirements in terms of data.

2.2.4. Uses of accessibility measures

Due to its nature, accessibility can be used in different contexts and to study different topics. One of the ways in which accessibility has been used frequently is to study inequalities related to the transportation system. In fact, in both Martens et al. (2019) and Pereira et al. (2017), accessibility is highlighted as a powerful measure to study this topic.

Inequalities are the main topic of several studies on the subject. In most of these articles, the assumption is that the public transport system should in priority benefit the most economically and socially disadvantaged categories of the population. In Foth et al. (2013), use a potential accessibility measure as well as a social indicator composed of several variables (such as median income, unemployment, percentage of income spent on rent and the number of recent immigrants) to evaluate the public transit system in Toronto. In Ermagun and Tilahun (2020) a cumulative accessibility measure is used to assess the inequalities both in terms of spatial location, but also taking into account the income and racial composition of neighbourhoods, in the context of Chicago in the USA. From a more economical point of view, Fransen et al. (2019) have studied how inequalities in accessibility to employment influences inequalities in long term employment, and how accessibility benefits different population groups in this regard. Thus, depending on the context different variables for explaining inequalities of access are considered.

However, when studying inequalities in transportation, some authors have found that only considering travel time to the destination can be limiting. Thus, both El-Geneidy et al. (2016) and Bocarejo S. and Oviedo H. (2012) add an additional variable to the cost function, which the monetary cost of the public transport fare. Indeed, it is argued in both these articles that although it does not have a significant impact on the accessibility of high-income categories of the population, it can be very limiting for more disadvantaged categories of the population. Aberle and Gertz (2025) use a 'fare accessibility' measure, which represents the 'amenity destinations reachable within a €2.30 ticket'. Their study show that fare accessibility is not equally distributed and that some people cannot reach amenities with a ticket, showing the importance of taking into account fare cost when discussing inequalities of access.

Another use of accessibility measures in the literature has been to evaluate the effectiveness and impact of new infrastructure projects. Some studies evaluate the expected effects of future infrastructure projects, whereas other studies study the observed impacts of already implemented transportation projects. In Ford et al. (2015), the authors use a more complex than average accessibility measure, as well as a geographic information system (GIS), to evaluate the expected effects on accessibility of several planned transportation projects in the London area.

This approach can of course be combined with a focus on equity in transportation. Both Fan et al. (2010) and Bocarejo S. and Oviedo H. (2012) consider the effects of transportation projects on accessibility. In Fan et al. (2010), the authors aim to evaluate the impact of a new rail line on accessibility, while also studying the distribution of the benefits among different categories of the population. Indeed, from a transport planning perspective, many transport projects have been justified as a mean to reduce inequalities regarding transportation. In Bocarejo S. and Oviedo H. (2012), one of the considerations is the impact of the cost of projects in Bogota on accessibility for disadvantaged people, as in this context the cost of projects has resulted in an increase in fares, potentially limiting the access of some people to the transport network.

In other contexts, accessibility has been used to assess the performance of the transport network, such as in Mavoa et al. (2012). In some cases, such as in Ford et al. (2015), the evaluation of the public transport system has the aim of comparing its performance to accessibility by car, in order to assess if mode shift can be profitable in terms of accessibility for the users, in the perspective of the reduction of greenhouse gas emissions.

2.2.5. Scope of studies

The studies considered in this literature review differ by the scope of what they consider.

From a spatial point of view, most studies focus on a single city or urban area, with the notable exception of Verduzco Torres and McArthur (2024). Indeed, in this study the authors create an accessibility database for the United Kingdom, which can be used to consider effects that are not properly highlighted when only considering a single urban area.

Regarding the different modes of transport available, the multimodal nature of some transport networks is sometimes hard to represent. For example, in Ford et al. (2015), even though multiple modes are considered, when calculating accessibility only one mode per trip is considered, which can be considered unrealistic as transfers between different modes (for example bicycle and light rail) may be possible. Similarly, in Fan et al. (2010), only one transfer is allowed in the model, and only 400m may be travelled by walking, which once again could be unrealistic depending on the context. Thus, even though transport networks are often multimodal, representing accurately these networks remains a subject of research.

From the perspective of time, most studies focus only on accessibility at peak hours. While it may be relevant for accessibility to jobs, this may not be accurate for other activities. Furthermore, only estimating accessibility at peak hours may lead to overestimation of accessibility to night jobs, for example. Some studies however consider both peak hours and off-peak hours, such as Mavoa et al. (2012). Furthermore, in this study the authors highlight the significant differences of accessibility of activities between peak hours and off-peak hours, though in this study frequency of public transport is considered separately.

2.2.6. Analysis performed on the results

Across all articles studied, there are a lot of variety regarding the statistical analysis performed on the empirical results obtained after estimating the accessibility.

Of course, since accessibility is linked to the land-use system and to the structure of the transport network, results are often represented in the form of maps, representing the accessibility to a selection of destinations from each zone considered in the study. Such visualisations are used for example in Bocarejo S. and Oviedo H. (2012), Radzimski (2023), Foth et al. (2013), and many other articles.

Statical analysis often performed on accessibility also include, for example, scatterplots and linear regression (Foth et al., 2013), linear models (Ermagun and Tilahun, 2020), box plots (Radzimski, 2023), or cumulative curves (Ford et al., 2015).

Another analysis method which can be quite insightful is the use of the gini index and Lorenz curves. This method has been used, for example, in Lucas et al. (2016), van Wee and de Jong (2023) or Aberle and Gertz (2025). The Lorenz curve is the cumulative curve of distribution of a commodity, representing the share of a commodity owned by a certain percentage of a population. From this curve the gini index can be calculated. One advantage of this analysis is that, since the gini index is widely used for assessing equality, it allows for comparison between the equity of the distribution of accessibility and the distribution of other indicators such as income.

Other methods focus on the spatial nature of accessibility. For example, Adorno et al. (2025), in their analysis of accessibility to green spaces, use two analysis methods: spatial regression and spatial clustering. These methods have multiple advantages: they take into account the spatial nature of accessibility and the data used to calculate it, and they allow for a more nuanced representation of access. Indeed, it allows to explore the complex relationship between people’s income, gender, age, and the region of a city in which they inhabit.

Thus, it appears that there is no clear dominating method when analysing results, although some approaches beyond the basic statistical analysis can be very insightful.

2.2.7. Results and findings

Regarding studies on specific cities, it is noticeable that the articles mentioned in this literature review cover some countries more than other. Countries such as the United Kingdom, the USA and Canada are well represented. On the other hand, some other countries still have been studied, but less than the USA for example. One notable absence is Asia, which may be due to a lack of studies in English language, or by a lesser influence on literature compared to studies from English speaking countries.

Thus, given the lack of studies on some areas of the world, it is no surprise that there are no global findings which could be applied to most cities in the world. In addition, studies in different contexts have shown very different outcomes, even when comparing similar countries, like the USA and Canada.

When comparing the results of studies, it is important to be careful about the spatial scale at which the study was conducted. Indeed, van Wee and de Jong (2023) have shown in their article, using the example of accessibility to health facilities, that results on accessibility are sensitive to the spatial scale chosen. Indeed, the same method at different spatial scales can lead to different results and conclusions. Thus, a well chosen spatial scale is important when discussing accessibility.

One example of this is ‘spatial mismatch’, i.e. the difference between the worker’s location and employment location, especially observed in the USA. Indeed, in the case of the studies of Fan et al. (2010) and Ermagun and Tilahun (2020), the researchers observed that some population groups have lower access than others, showing clear inequalities in access. In the case of the study of Fan et al. (2010), the authors explicitly observe spatial mismatch in the ‘Twin Cities’ of Minneapolis and Saint-Paul. However, in studies about cities in Canada, namely Toronto studied by Foth et al. (2016) and Montreal studied by El-Geneidy et al. (2016), the outcomes are quite different. Indeed, in their study Foth et al. do not observe a spatial mismatch in Toronto, and highlight that the public transit system of Toronto is quite equitable, as it benefits primarily the most disadvantaged. Similar findings are present in El-Geneidy et al. (2016), even as the study also takes into account the impact of public transport fares on accessibility to jobs.

In their study, Bouzouina et al. (2021) take a different approach. In this study, the focus is given to the distinction between renters and owners, and the trade-off between housing prices and accessibility to employment. They point out that owners generally prefer lower housing prices over better accessibility, the opposite being true for renters. Thus, residential self-selection is also something that influences accessibility inequalities.

In another context, Bocarejo S. and Oviedo H. (2012) point out that, in Bogota, high-income populations generally have better access to jobs than low-income groups. Furthermore, in order to access jobs, low-income populations have to spend a much larger part of their income on transportation than high-income populations. The conclusion of the article is that infrastructure projects are less effective in improving accessibility of disadvantaged groups than fare subsidies. Similarly, Fransen et al. (2019), also point out that accessibility improvements in Flanders benefit those who are already advantaged regarding employment.

Thus, it seems a distinction could be made between advantaged population groups, which have more agency in which level of accessibility is sufficient for them, and disadvantaged population groups, which are less likely to be able to improve their accessibility if they suffer from poor accessibility.

These different outcomes could be somewhat coherent if a distinction would be made between ‘chosen’ accessibility disadvantages and ‘imposed’ accessibility disadvantages. In the first case, advantaged populations chose a poorer accessibility in order to have other advantages regarding their residential location. Conversely, in the second case, some disadvantaged population have to suffer low accessibility levels because residential areas with a higher accessibility to jobs are unavailable to them.

2.3. Conclusion

Accessibility is a very versatile indicator, that has been used in a variety of use cases, such as evaluating the performance of transportation systems, but can also be a valuable indicator to highlight inequalities regarding transportation, according to Pereira et al. (2017). In turn, it can be used to show how inequalities in transportation translate directly into disadvantages in other fields, such as employment, access to health and education and social inclusion. It is also an indicator well-suited to the evaluation of transportation projects -as seen in Ford et al. (2015)- which may be used in addition to measures such as the number of trips or the financial profitability.

Many different accessibility measures are used in the literature, depending on the context of the study and the specific aims of the study. However, more complex and general accessibility measures conceptualized in the theoretical literature have seen much fewer use, most likely due to being harder to interpret, and because these measures require much more data than simpler ones. As pointed out in Geurs and Van Wee (2004), cumulative accessibility has been widely used, for example in Foth et al. (2013) or Verduzco Torres and McArthur (2024), as this measure is easy to interpret and requires only a threshold as a parameter. The inclusion of competition effects has lead to the creation of accessibility measures such as Shen’s accessibility (Shen, 1998).

From what can be deduced from the literature, the findings from the studies considered depend strongly on the context and history of the urban areas that have been considered. Even though some networks appear more equitable than others, it is hard to know if this is due to transport policies, or if it can be explained otherwise, for example by considering residential self-selection and the historic urban development of the area.

2.4. Discussion

Several research gaps persist in the studied literature. One such research gap is that for some areas in the world, few studies have been conducted regarding accessibility and inequalities of access. This can lead to extrapolating findings that are only valid for a restricted geographic or cultural area. Thus, producing more studies in countries that have been neglected thus far could prove very interesting and may provide additional insights specific to these areas.

Regarding accessibility measures, if the context and data available allows it, it may be worth considering using more complex and accurate accessibility measures. Indeed, some computational and data limitations may be less relevant compared to when these measures where first theorised. In particular, the inclusion of competition effects may lead to more accurate results, as pointed out in Shen (1998).

Considering the type of analysis applied to accessibility, it may be interesting to consider using other methods. Gini index methods -used in Lucas et al. (2016) and Van Wee and De Jong (2023)- allow for comparison between accessibility and other indicators and can also be useful to allow an easier comparison between different case studies. Meanwhile, spatial regression and clustering methods used in Adorno et al. (2025) may allow to explore more in depths the relationship between accessibility and other socio-economic variables such as income, neighbourhood, age or household composition for example.

Thus, it would be worth studying urban areas in relatively less studied countries, while focusing on an accurate representation of multimodal transport networks. Further accuracy could be brought in the form of a larger time span considered. In order to account for individual capabilities, future studies should consider taking into account the types of jobs offered to the population. To that end, it may be relevant to improve already used accessibility indicators or try to apply already theorised accessibility measures to practical situations.

3. Methodology

3.1. Trips characteristics

3.1.1. Destinations considered in the study

The access to the following points of interests and commodities is calculated:

- Jobs (divided by socio-professional category)
- Hospitals
- General practitioners
- Pharmacists
- Primary and secondary education establishments
- Higher education establishments
- Supermarkets
- Convenience stores

These locations are considered key services in our study. This selection is quite similar to that of Verduzco Torres and McArthur (2024).

3.1.2. Transport modes

No public transport mode was excluded in the study. This means that all available public transport modes available in the urban area were considered. This includes not only the transport networks showcased in the descriptive analysis such as subway, tramway and regional trains, but also other modes, most notably bus service lines.

In addition to public transport modes, walking and cycling are considered to access public transport stops. In some cases, where the trip is very short, an itinerary may only have cycling or walking components.

3.2. Data

3.2.1. Census tracts

There are multiple administrative and urban divisions that exist for Paris and its neighbouring cities: the 'Ile-de-France' region, the metropolitan area, the urban area and other divisions. This study will focus on the Paris urban area, which is the area where there is a built environment continuity from Paris.

The spatial scale selected for the study are the 'iris' (INSEE). These zones are designed to provide detailed data below the city level, for cities which exceed at least 5000 inhabitants. The advantage of using these zones over grid data, is that these zones are designed to be somewhat homogeneous in terms of population and type of built environment (housing, activities or other). In this study, the iris geographic limits from 2022 were used.

Figure 1 presents a map of both the boundaries of iris zones and the outline of the study area.

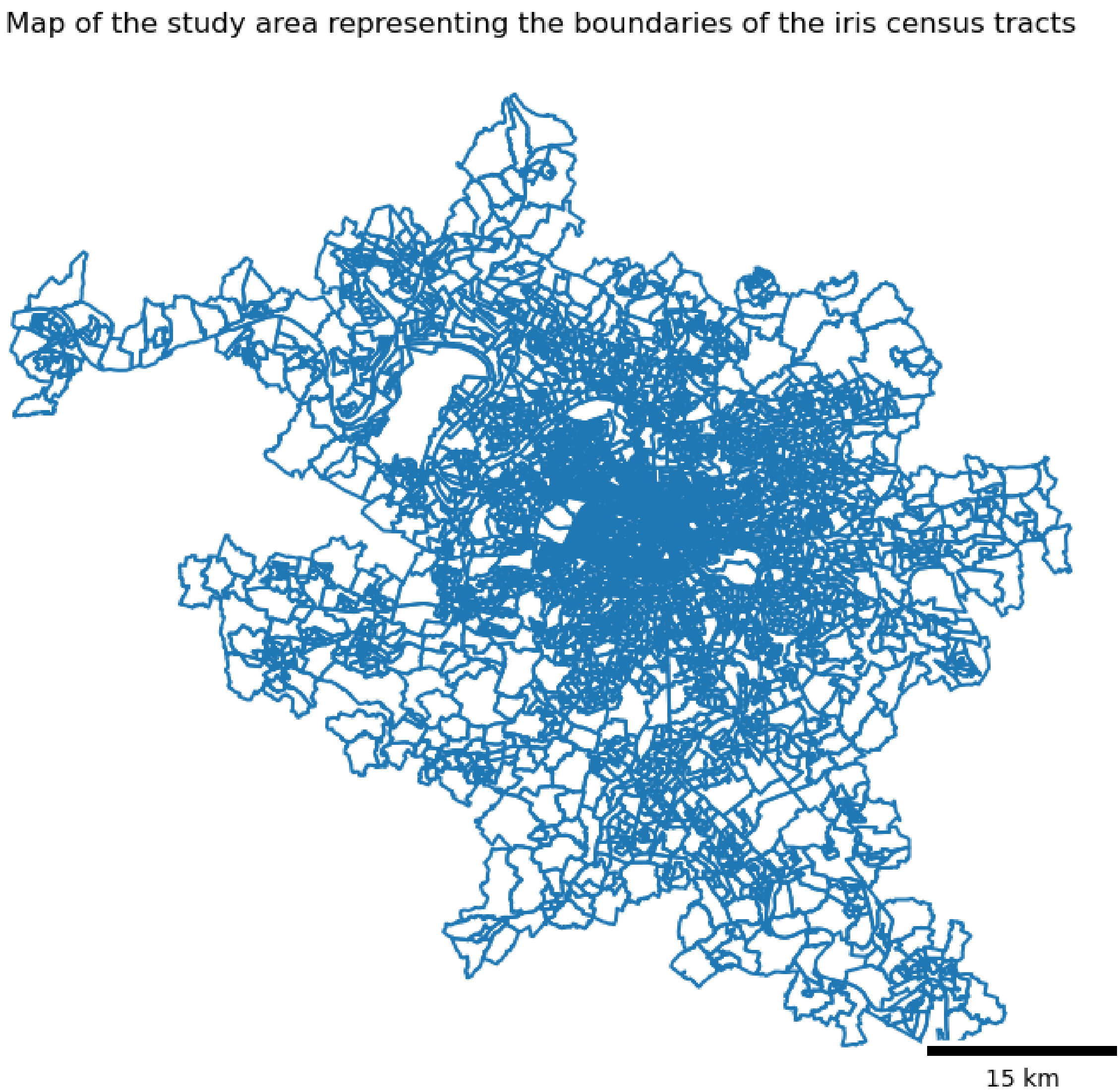

Figure 1: Map of iris census tracts boundaries in the study area

3.2.2. Datasets used

Table 1 shows the datasets used in the study. Socio-economic and demographic data about the population of the iris zones come from two main databases. The first one is the Filosofi data, which mostly provides data about the income structure and poverty rate of the iris areas. This database includes both the income before and after taxes and redistribution. In our study we use the available income as a reference to better represent the influence of fares on the budget of individuals and thus we use income after redistribution. The second database is the census data at the iris level. This data is quite extensive and contains the type of housing, the size of households, the socio-professional categories of individuals, among other indicators.

Data about jobs is extracted from the siret establishment list, which contains information about all companies in France. This include the number of employees in the company, though the exact number of employees is unknown, instead a code is given which corresponds to a certain range of number of employees (for example, the code '01' corresponds to a number of employees between 1 and 2). The location of the companies is also included in the dataset. This dataset is also used to estimate access to general practitioners, as they correspond to a certain type of company registered in the dataset.

The health facilities dataset is used to access information about hospitals. Each data entry contains information about the type of the facility and its location.

To collect data about convenience stores and supermarkets, we use open street maps extracts.

For education establishments, datasets are available which contains the location as well as the size of the establishments.

Finally, the GTFS data is obtained from the 'Ile-de-France mobilités' open data (IDFM), IDFM being the transit agency organising public transports in the Ile-de-France region. Since the Paris urban area is entirely contained within this region, there is no need for additional GTFS data. More details about the GTFS data format are given in appendix C.

Title	**Format**	**Description**	**Source**
Filosofi data – iris level	.csv	Filosofi data from 2021. This dataset contains information about the income of the population at the iris level. This dataset also contains information about income both before and after redistribution of wealth	INSEE
Census data – iris level	.csv	Census data from 2021. This dataset contains a large amount of information about the population.	INSEE
Iris geographical data	shapefile	Contains the geographical information of the iris zones	Geoservices
Siret establishments list	.csv	Information about all entities which are in the siret registry (this registry encompasses all companies in France)	sirene
Siret establishments geolocation	.csv	Dataset containing the coordinates of siret establishments, as well as the iris to which they belong	Datagouv.fr
Health facilities dataset	.csv	Dataset containing the type, coordinates and other information about health facilities (including hospitals)	Datagouv.fr
Medical Staff dataset	.csv	Dataset containing information about medical staff at the national scale	Datagouv.fr
Stores dataset	.geojson	OSM extract of stores operating in France, containing information such as the coordinates, names, brands and the type of store	Datagouv.fr

Primary and secondary education directory	.geojson	Dataset containing the coordinates, names and types of primary and secondary education establishments	Datagouv.fr
Higher education dataset	.geojson	Dataset containing the coordinates, names and types of higher education establishments	Datagouv.fr
IDFM GTFS	gtfs	Gtfs data from 'Ile-de-France Mobilités', which is the administrative entity in charge of organising public transports in the Ile-de-France region. This dataset encompasses all public transport lines of the region.	Datagouv.fr

Table 1: Datasets used in the study

3.2.3. Data filtering

Table 2 describes the different types of filtering applied to the datasets containing the points of interest used for the study.

This filtering is done to only keep relevant information (for example, in the case of stores, not accounting for hairdressers when we only want to study accessibility).

Furthermore, when filtering the data, points of interests of similar size and nature were grouped together (for example, avoiding mixing together large hospitals with smaller medical private cabinets). This is done in order to keep the points of interest of a certain dataset comparable with one another. Indeed, when computing accessibility, all points of interest of a certain type are considered 'equal', which is why it is important to be careful when grouping together points of interest.

Naturally, in addition to the filtering described in Table 2, data that was at the national scale was filtered so that only data entries in the Paris urban area are taken into account.

Type of point of interest	Dataset	Sub-datasets created
Education	Primary and secondary education	Regular schools (primary schools, middle schools, high schools)
		Special education institutions (medico-social education facilities and adapted schools for disabled people)
	Higher education	Universities (including specific 'teaching and research' units)
		Selective schools
		'Grandes écoles' (selective higher

		education schools, considered very prestigious)
		Specialized education (for example nurse schools)
		High schools that have higher education programs (professional high schools, Agricultural high schools, etc)
Employment	Registry of companies (siret establisments)	Only companies that have employees (so no volunteer work for example), possibility to filter by type of activity
	Geolocation of companies	Location of companies, which have an id number called ‘siret’
Health	Health facilities dataset	Active Hospitals (since the notion of ‘hospital’ can vary from country to country, what was specifically selected were ‘hospitals centers’, which are quite large)
	Medical staff datasets	Medical staff with the status of doctor and a specialty in ‘Médecine Générale’ (General Medecine)
Stores	National dataset of open shops	Convenience Stores
		Supermarkets
		Pharmacies

Table 2: Filtering performed on the datasets

3.3. Creation of a travel time matrix

The travel time matrix is calculated based on the GTFS data available for the urban area. GTFS data are widely used in transportation and accessibility studies. GTFS data is convenient because it contains in one format much of the data necessary to analyse transport systems: it contains the location of service lines, public transport stations, as well as timetables and frequency. It also contains information about the type of public transport line (bus, train ,tramway, subway, etc). The calculation of travel time is done on python, using the r5py library.

The origins and destinations considered in the travel time matrix are the centroids of the iris areas. When calculating travel time from an iris zone to an activity, the centroid of the origin zone is used as the start location of the trip and the centroid of the destination iris is used as the location of the activity. In case iris centroids happen to be located at a point unreachable by the network, the option ‘snap-to-network’ is activated allowing for the origin to be relocated to the nearest point connected to the network.

All public transport modes are considered when calculating the travel time matrix, as well as walk and cycling. Modes can be combined along a given trip to account for the multimodal nature of the network. The calculation is done for the 25th of September 2025, which is a Thursday, outside of school holidays. The departure time is set to 8AM, with the departure time window being set to 30 minutes to avoid inaccuracies in the calculation, since some of the modes involved such as buses can have a low frequency and thus could be neglected if using a smaller departure time window.

3.4. Accessibility metric

Accessibility is a measure well-suited to this research for several reasons. As mentioned by Geurs and Van Wee (2004), accessibility allows to consider the interactions between the land-use system and the transport system. It is important in the case of this study as some places may be deprived of access to transport to certain places but benefit from local facilities that are sufficient. On the other hand, some areas may be deprived of both access to transport and be located far from key facilities, whereas some areas may benefit from good access to both.

Furthermore, accessibility as a measure has been widely used to study inequalities (such as in El-Geneidy et al. (2016) or Bocarejo S. and Oviedo H. (2012) for example), which makes comparison between case studies easier, though one needs to be aware of the type of accessibility metric used. Moreover, from an ethical perspective, Pereira et al. (2017) claim that accessibility 'stands out as the most promising focal variable of distributive justice'. Thus, accessibility was chosen in order to properly address the subject of inequalities in this study.

Shen's accessibility is used (Shen, 1998). This accessibility measure is chosen for multiple reasons. First, since it is based on potential accessibility measures, it supersedes the need for an arbitrary travel-time or travel-cost threshold which may impact the results. Moreover, shen's accessibility takes into account the competition effects, which is good for all types of activities but especially for jobs.

To take into account the cost of fares in accessibility, we use generalized time travel as the cost function for accessibility. Generalized time travel is defined as follows in the context of this thesis:

$$C_{ij} = t_{ij} + \frac{F_{ij}}{w_i}$$

Where C_{ij} is the cost of travel from the zone i to zone j, t_{ij} is the travel time from i to j, F_{ij} is the fare to travel from i to j, and w_i is the median wage in zone i. The form of this cost formula is similar to the one from El-Geneidy et al. (2016).

Another cost formulas is considered, by using only travel time as the impedance,

Thus, the general formula of accessibility used in this study is (with small changes from Shen, 1998:

$$A_i = \sum_j \frac{O_j e^{-\beta c_{ij}}}{D_j}$$

$$D_j = \sum_k P_k e^{-\beta c_{ij}}$$

Where:

- A_i is the accessibility from zone i to a certain type of activities
- O_j is the number of opportunities for the activity in zone j
- β is the exponential decay curve parameter for the activity
- C_{ij} is the cost of traveling from i to j
- D_j is the demand potential for zone j
- P_k is the number of people in k seeking for the opportunities to perform the given activity

The cost function chosen is a negative exponential, as seen above. This is because, according to Geurs and van Wee (2004), 'the negative exponential function is the most often used and also the most closely tied to travel behaviour theory'. This allows for easier comparison across articles.

The parameter of the negative exponential function β is calibrated based on the most recent travel survey in the Ile-de-France region (OMNIL).

3.5. Estimation of the demand and number of opportunities for competition effects

3.5.1. Jobs

For jobs, the number of opportunities (that is to say, the number of jobs in an area) is not known at the iris level. To accommodate for this fact, we estimate the weight of all iris in a given city by summing the number of employments for each company in the iris. Then, this number is divided by the total estimated number of jobs in the city, obtained by making the sum of jobs over all iris in the city. By dividing the estimated number of jobs in an iris and the estimated number of jobs in the city, the relative weight of the iris in the employment of the city is obtained. Then, we multiply the total number of jobs at the city level obtained from the 2021 census by this weight, to obtain an estimation of the number of jobs in a given iris.

$$J_i = \frac{J_{i,estimated}}{J_{c,estimated}} \times J_c$$

Where:

- J_i is the number of jobs in iris i which belongs to city c
- J_c is the number of jobs in city c
- $J_{estimated}$ is the estimated number of jobs in a zone based on employment in companies that belong to this zone in the siret registry

We consider on one hand the total number of jobs as opportunities and the total working age population as the demand, and on the other hand the jobs and working age population are divided by socio-professional category (INSEE). The general composition of these socio-professional categories is shown in Table 3.

Number of the category	Fields represented by the category
1	Agriculture
2	Business owners
3	Executives, intellectual occupations
4	Intermediary occupations
5	Employees
6	Workers

Table 3: Composition of socio-professional categories

3.5.2. Education

For primary and secondary education facilities, the demand is the population that corresponds to the age range of the type of education (primary or secondary) and that is in education. The number of opportunities for primary and secondary education is represented by the number of students the school can accommodate.

For higher education, the situation is more complex as not all the population in age of attending higher education will do so. Thus, the demand is considered to be all people aged 18 to 24, and in education. Regarding the supply side of higher education, it is considered to be the total number of students enrolled in each facility.

3.5.3. Hospitals and health facilities

For hospitals, there are many potential indicators such as the number of beds, or the number of surgical operating rooms. In this study, the chosen indicator is the size of the medical staff working at the hospital, as other measures such as the number of beds may neglect other factors in the importance of a specific facility (such as diagnosis equipment). Thus, the number of opportunities in an area is the size of the staff in all hospitals of the area. Regarding the demand for hospitals, virtually all of the population can be in need of access to an hospital. Thus, we consider the entire population of the zones as the demand.

3.5.4. General practitioners

For general practitioners, the supply side is considered to be the number of medical offices in an area. Regarding the demand, it is, as for hospitals, the entire population.

3.5.5. Stores

Stores are in a similar situation as hospitals: there is no clear indicator of opportunities available. Thus, it is considered that each store count as one opportunity, and that all stores in a given category are equal in terms of opportunity. To compensate for this problem, supermarkets and convenience stores are separated to help ensure that stores of similar sizes are considered. Pharmacists are also considered separately from convenience stores and supermarkets.

The demand for stores is considered to be all of the population above a minimum age threshold.

3.6. Clustering

In order to analyse the association of accessibility with other variables, a clustering method was used in this study.

The clustering method chosen for this thesis was the k-means clustering methods. Though there are more suited methods for spatial data such as spatial clustering (used for example in Adorno et al. (2025)), they were not used in this study, both for technical reasons and time constraints.

Clustering has several perks in the context of this study. It allows to show association between variables that do not have a linear relationship. Indeed, the assumption was that clustering would be able to display information and associations between variables that remained hidden in the more “linear” descriptive analysis. Furthermore, in the context of our study where it is expected that data has a spatial structure, clustering results can be easily communicated on a map. This also makes it easier to identify patterns in the data and the repartition of both accessibility and socio-economic variables. Since clustering is an exploratory data analysis method, it may also be helpful in identifying unexpected relationships between variables.

In order to determine the number of clusters for each clustering model, the elbow method was used.

In addition to accessibility data, multiple socio-economic indicators were used to form the clusters. First, the yearly median available income in the area, ‘available income’ meaning that this data takes into account the loss of income from taxes, and the gain of income through social subsidies and other redistribution. Second, the ratio of social housing over all housing in the area. Though this indicator may be correlated with income, it could help identify specific neighbourhoods, especially the ‘*cités*’ mentioned previously. Indeed, this neighbourhoods are often characterised by a high proportion of social housing, as mentioned in Castel (2006). Furthermore, the ratio of households that own at least one car over the total of households in an area is also considered. Indeed, households that own a car may be less reliant on public transportation, and may travel by car by choice, though of course some households are also reliant on car travel by necessity. Finally, the population density of the area was considered. This is done in order to separate the central parts of the urban area, which are more densely populated, and the areas further to the periphery of the urban area. At first, the poverty rate of the area was also considered in the clustering process, but due to the large amount of missing or unavailable data for this variable it was not used.

Moreover, more global clusters were created using all types of accessibility, as well as socio-economic data. This was done in order to see if there were irregularities in accessibility (for example, if some clusters would have good accessibility to jobs but bad accessibility to other facilities) or if, on the other hand, areas with good accessibility for one type of service tend to have good access to other facilities, and conversely for bad access.

Then, clusters were created using the data mentioned above and the accessibility to key services belonging to specific sectors. Table x summaries the different sets of clusters that were created and shows what accessibility measure were used in creating the sets of clusters.

Type of facility	**Accessibility data considered for creating clusters**
Employment	Accessibility to all jobs
	Accessibility to jobs for each socio-professional category
Education	Accessibility to primary schools, middle schools and high schools
	Accessibility to each type of higher education facilities
Health	Accessibility to hospitals and accessibility to general practitioners
Stores	Accessibility to convenience stores and accessibility to supermarkets
	Accessibility to pharmacies
All types	Accessibility to all jobs Accessibility to all primary and secondary education facilities (aggregated) Accessibility to all higher education facilities (aggregated) Accessibility to hospitals Accessibility to general practitioners Accessibility to convenience stores Accessibility to supermarkets

Table 4: Summary of clustering analysis performed

Thus, there are 8 sets of clusters in total. The detailed analysis of the characteristics of each cluster is presented below, in the 'results' section.

3.7. Spatial regression

It was decided to use a spatial regression model in addition to clustering to analyse the data. Indeed, while clustering is helpful in identifying patterns in the data, it cannot provide precise information about the statistical link between specific variables and accessibility. Thus, a regression model was also considered necessary to this study.

Furthermore, spatial regression is useful for testing the assumptions that were made about the data, as it provides a similar framework to a linear regression model. This proximity in terms of structure is also helpful in order to compare results, for example if one wants to see the differences between a standard linear regression model and a spatial regression model. A spatial regression model was chosen, as it was considered more fitted to the type of data studied. Indeed, due to the data studied, it seems likely that spatial factors could influence the results. More specifically, the spatial lag model was used.

The spatial weights scheme chosen is the k-nearest neighbours, with k's value being set to 10. Independent variables were normalized.

Spatial regression was performed on accessibility to jobs and accessibility to primary and secondary education. These facilities were chosen to represent different types of facilities and to identify potential differences in the models. All accessibility data obtained was not used for spatial regression. Indeed, because good accessibility to one facility often translates to good access to all types of facilities, it was not deemed necessary to include all measures.

4. Descriptive analysis

4.1. Insights from the travel survey

The ‘Enquête Globale Transport’ is the travel survey conducted in the Ile-de-France region (OMNIL). The latest edition of this survey took place between 2018 and 2020. Though there may have been some changes in the transportation choices of inhabitants after the COVID 19 pandemic, this travel survey still represents valuable insight to better understand transport in the region and to interpret results.

Regarding the modal split in the region, in Paris itself, only 6% of trips are performed by car, with PT representing 31% of trips, and walking and cycling representing 59% of trips. The low use of car is consistent with the fact that two thirds of households in Paris don’t own a car. Residents of Paris also perform more trips per day compared to other departements, since they perform 4.34 trips/person/day on average. As expected, Paris ‘attracts’ more trips than it ‘emits’.

Comparing these results to the three ‘*départements’* of the Paris inner urban area, there are notable differences. In these areas, individual vehicles and PT are on par in terms of the number of trips performed. Furthermore, there are less trips per person per day on average compared to Paris. Besides, in all three of these ‘*départements’*, about one third of households don’t own a car, which is lower than Paris and could show that people need to rely more on their car.

The picture drawn by this data shows that the centre of the urban area relies primarily on PT, whereas beyond the limits of Paris proper reliance on travel by car tends to increase. This could indicate that it is harder to access to activities in these zones by PT compared to Paris.

4.2. Maps of the network

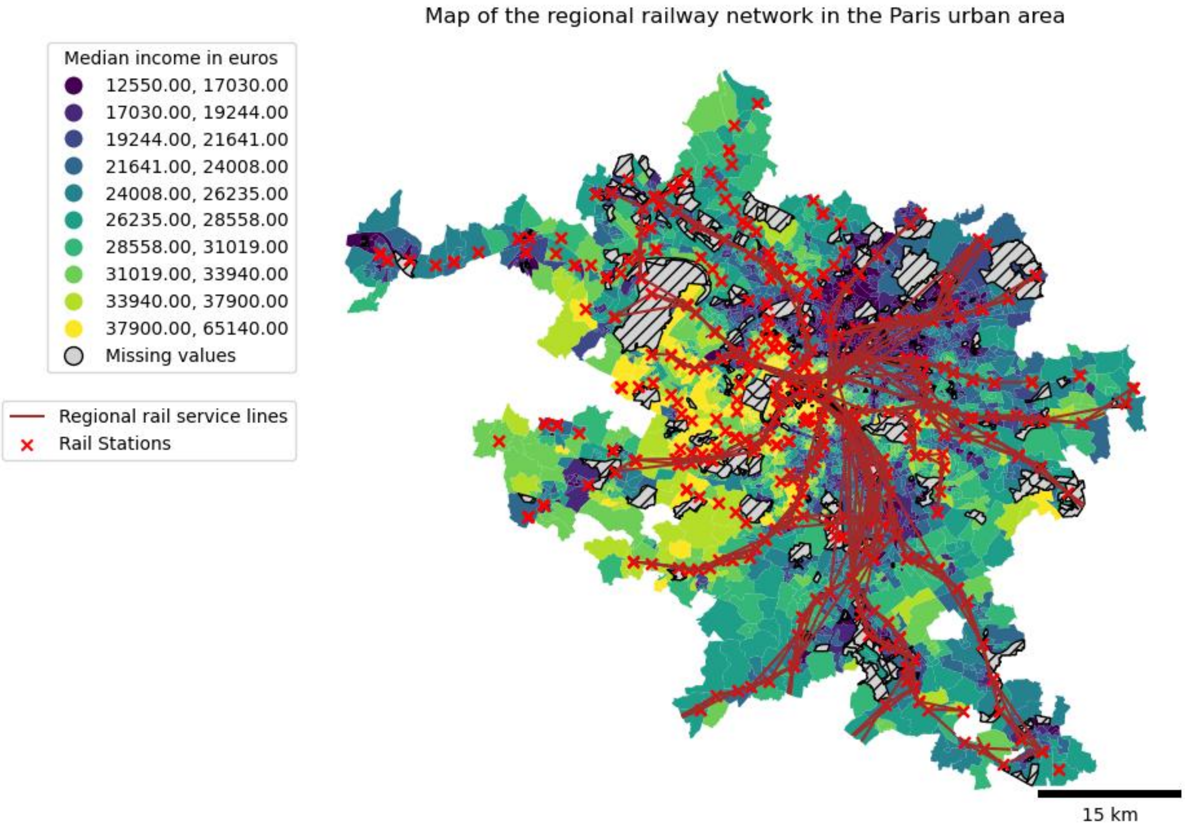

Figure 2: Map of the train network in the Paris urban area

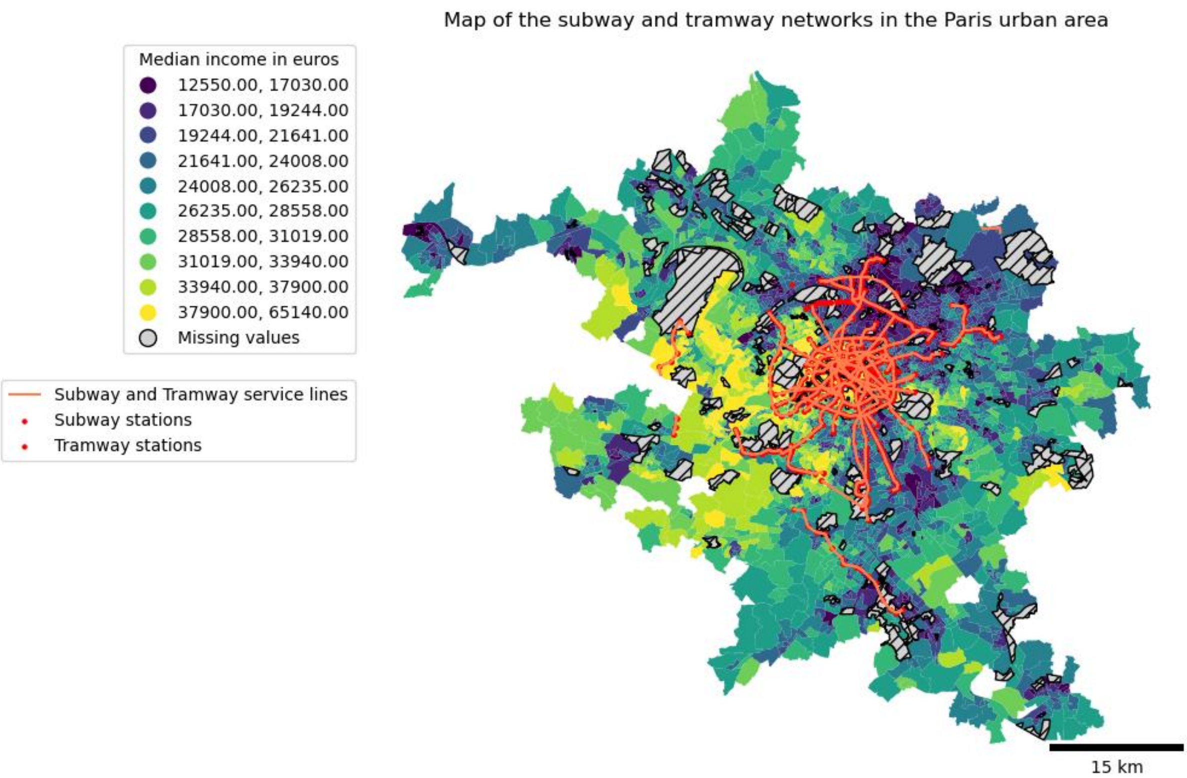

Figure 3: Map of the subway and tramway networks in the Paris Urban Area

Figure 2 shows the repartition of regional train lines and stations, compared to the median income of areas. Figure 3 represents the subway and tramway networks in the study area. Missing values often indicates iris that correspond to non-residential zones or natural zones such as forests. As could be expected, the network is denser in the centre of the urban area. Further from the centre, the network is less dense and some iris are left with few access to the train network. What is interesting is that, for cities in the outer urban area, it seems that areas where train stations are located often have less median income than neighbouring areas. One explanation to this observation could be that in peri-urban areas, cities that are the core of a smaller agglomeration are more likely to have both poorer households (one of the causes could be easier access to social housing compared to neighbouring, less urbanized cities) and transport infrastructure (due to being the centre of the agglomeration and having a higher density of population).

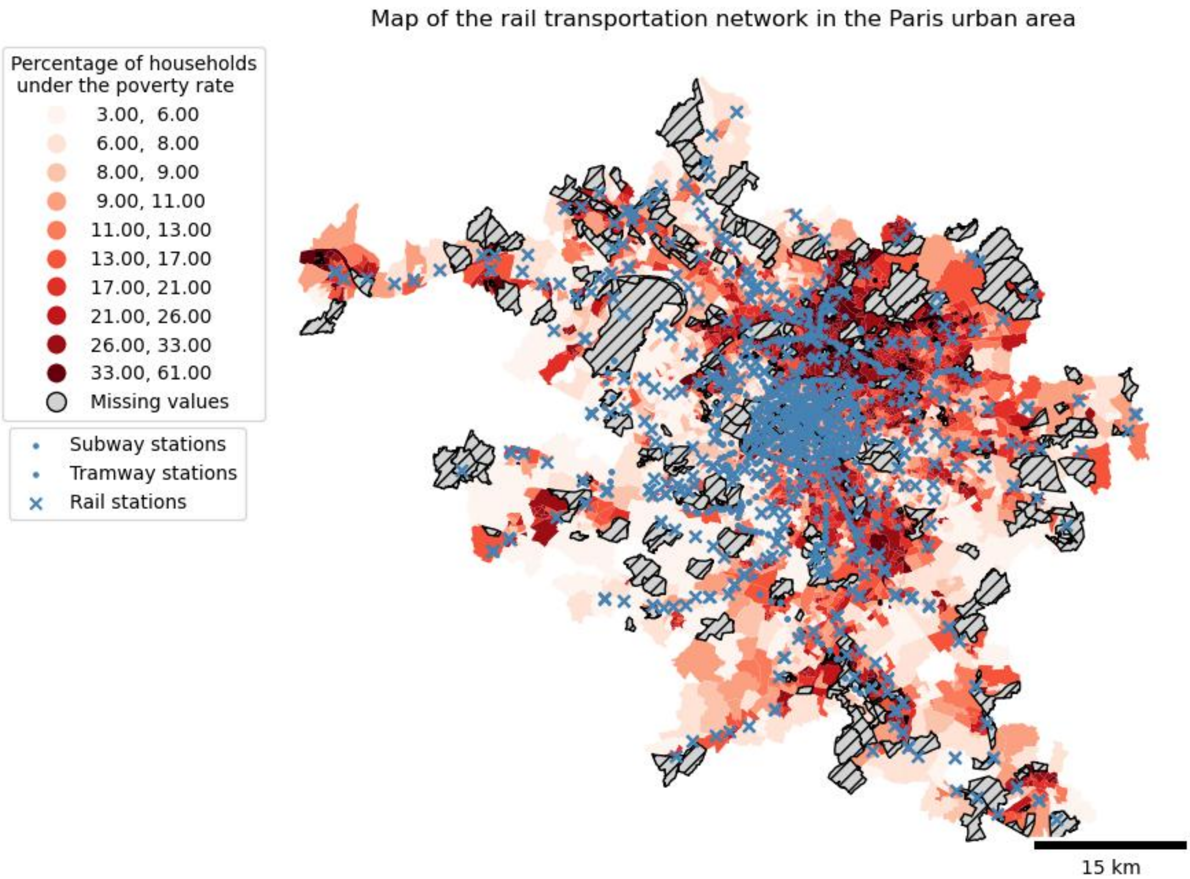

Figure 4: Map of rail transportation stations compared to poverty rate

Figure 4 is a map representing in blue metro, tramway and train stations, and in reds the poverty rate in the area (in France, households are considered ‘poor’ if the income of the household is less or equal than 60% of the median income). Of course, most stations are concentrated towards the centre, which includes both affluent and more poor areas. Further from the centre, as mentioned above stations seem to be in relatively poorer areas. However, since bus stations, which are more important for public transportation in the outer urban area, are not represented, it would be wrong to assume areas without stations on this map are completely deprived of local access to public transport.

4.3. Statistics about access to PT stations

For this section, two indicators are calculated for each census tract in the study area: the distance to the nearest public transport station, and the number of public transport stations in a 2km radius. These indicators are calculated from the centroid of each census tract. For nodes with multiple transport modes available, each transport mode present at this node represents a station in the data. For example, if a transport node contains both a train station and a transfer to a subway station, it will be counted as two stations. All public stations are considered in these analyses, including bus stations that were previously not included in maps of the network. Using these indicators, a simple linear regression is performed, and a linear regression prediction is presented alongside a scatter plot on a graph.

4.3.1. Distance to the nearest station

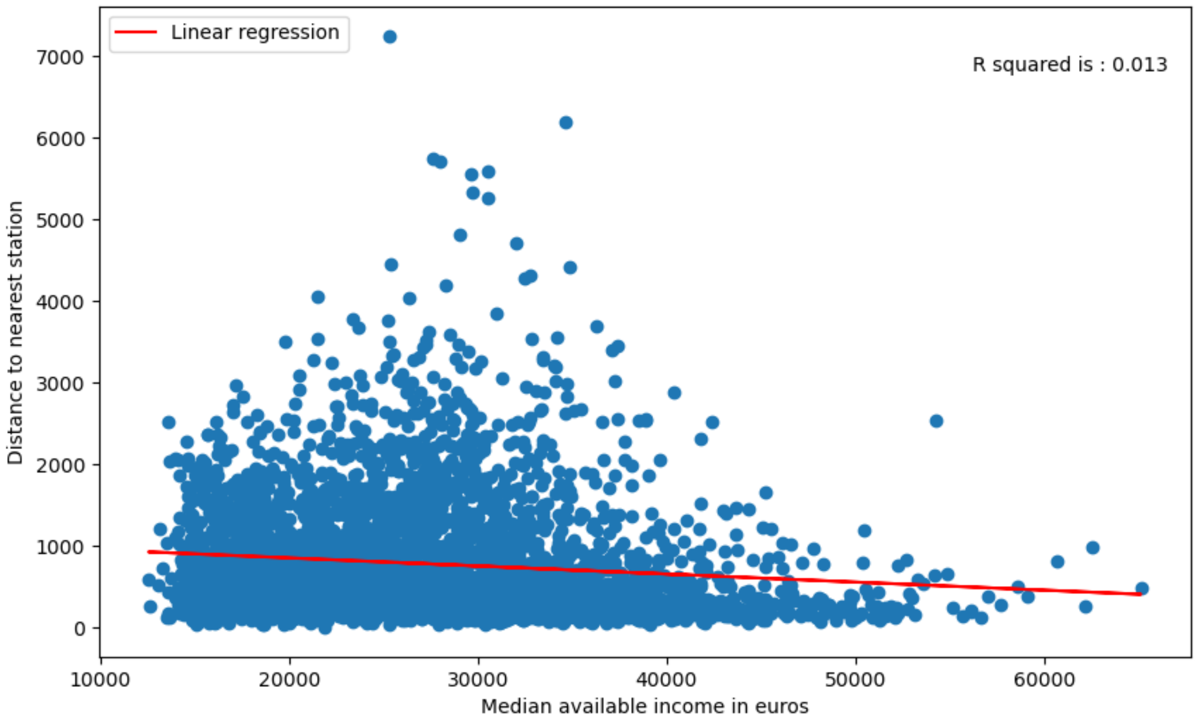

Figure 5: Distance to the nearest PT station compared to the median available income

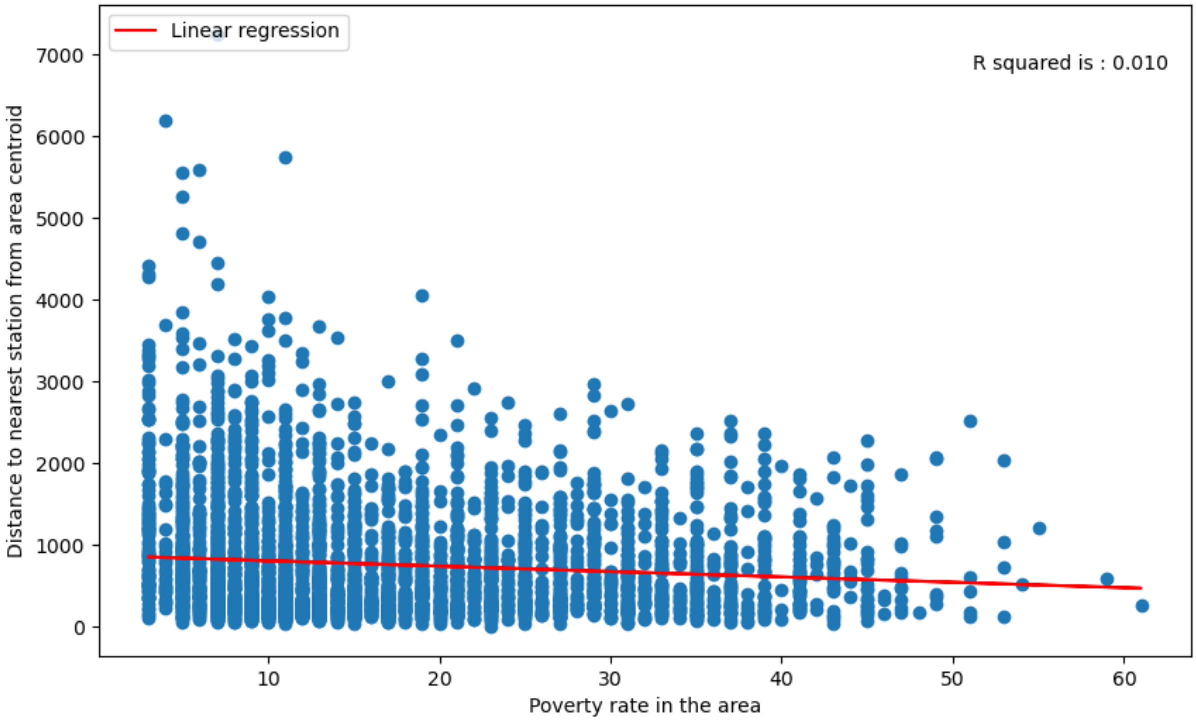

Figure 6: Distance to the nearest PT station compared to the poverty rate

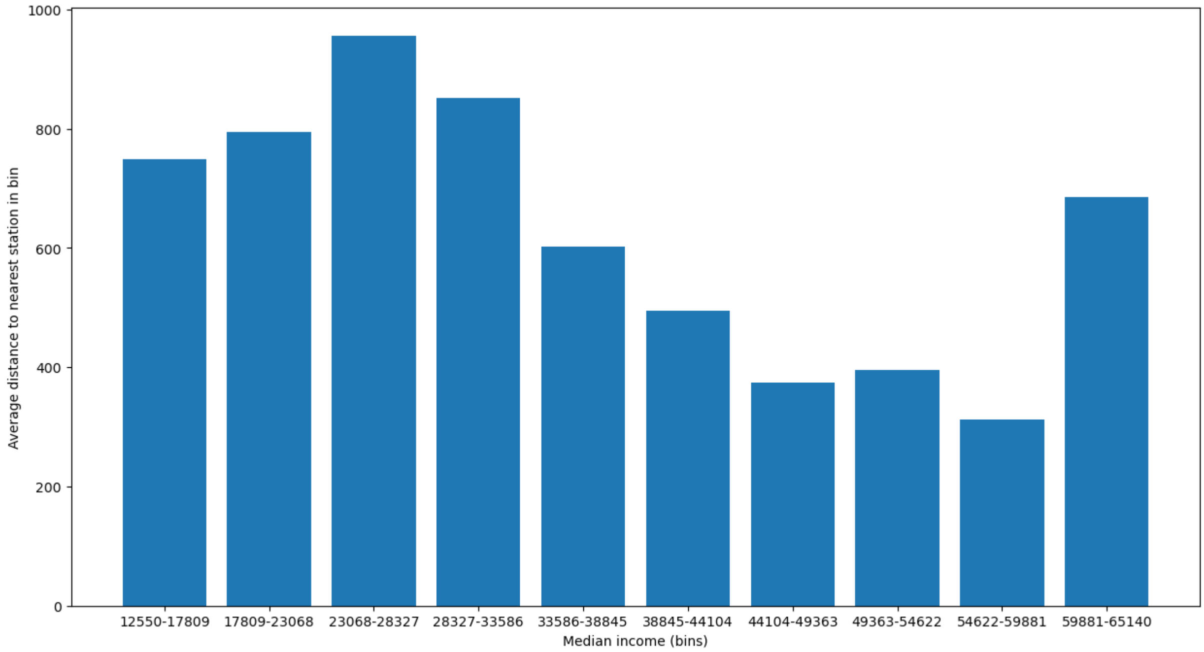

Figure 7: Bar plot representing distance to the nearest PT station compared to the median available income

The distance to the nearest stations decreases as the median available income increases, as shown in Figure 5, which could indicate, as was expected, a link between access to the PT network and income.

However, in the case of poverty, distance to the nearest station decreases as the poverty rate increases as shown in Figure 6, which seems to indicate a more complex relationship between income, poverty and distance to stations. To note is that there are a wide variety of situations for low poverty areas (suggesting the existence of both highly connected affluent areas and areas that are more isolated from the public transport network), whereas very poor neighbourhoods are generally close to stations. Also to keep in mind when reading Figure 6 is that the poverty rate is 14.5% nationwide (see INSEE) meaning that areas that are above 40% poverty rate are very vulnerable areas.

Furthermore, it is important to keep in mind that while poor people often live where they can, affluent people have more liberty to choose where they live. Indeed, areas that are often considered affluent include areas close to public transportation (some districts of Paris, the west of Paris in the inner urban area), whereas some other are located further from public transports (these areas have other perks that attract affluent households, such as large green spaces and forests, proximity to prestigious domains and castles, and so on).

Using Figure 7 to observe the average distance to the nearest station for bins of areas (areas are grouped based on median available income), some observations can be made. First of all, the first four bins have a higher distance to the nearest station on average than all other bins. The bin with the worst access to the nearest station is the 3rd bin, which represents the bin that is roughly around the national median income (which was 23 160€ in 2021 according to INSEE). The most affluent areas also have a high distance to their nearest station. For households above the average income, the distance to the nearest station decreases as income increases.

4.3.2. Number of stations within 2km of the area centroid

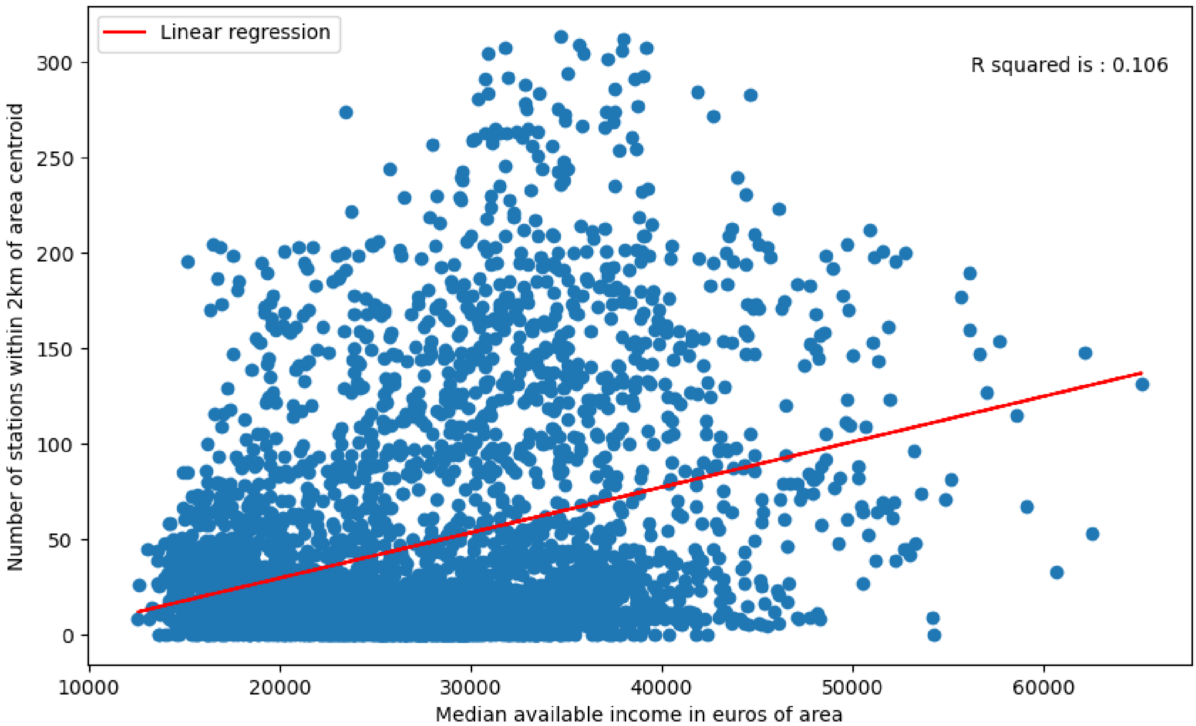

Figure 8: Number of PT stations reachable within 2km compared to the median income

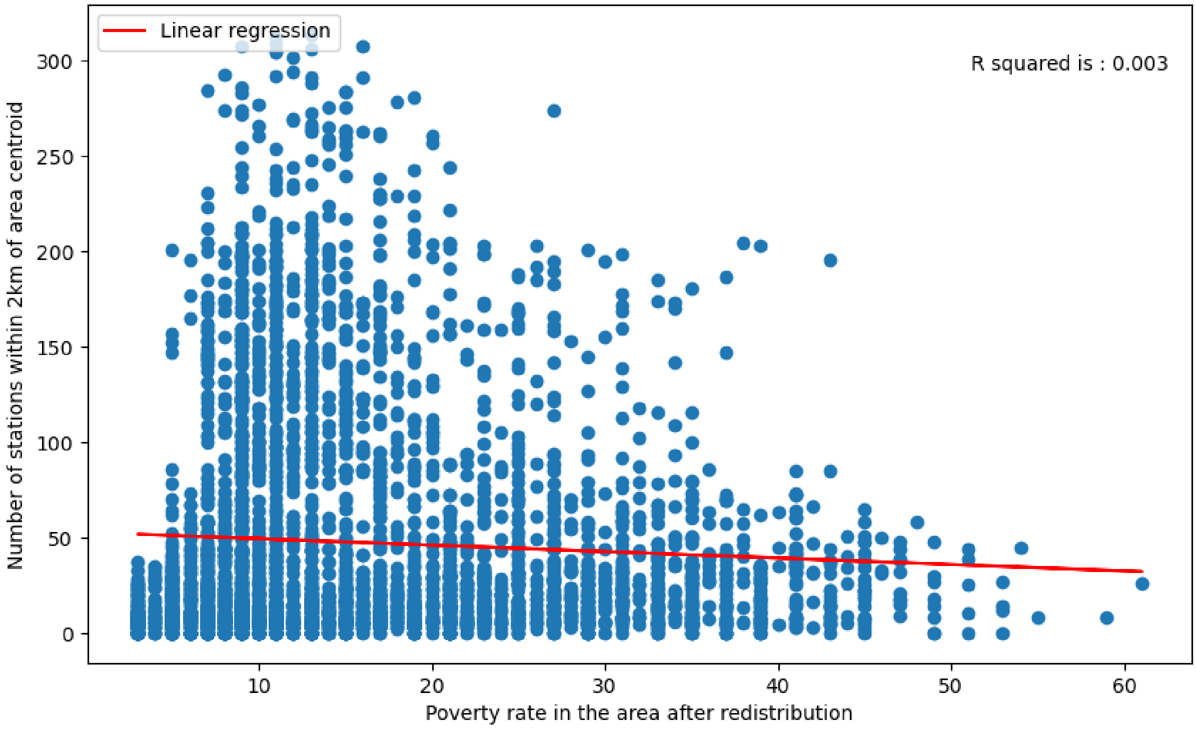

Figure 9: Number of PT stations reachable within 2km compared to the median income

When considering not only the nearest stations but rather the number of stations within a certain radius, though the scatterplots' points are more spread, the general relationship remains true for median income as seen in Figure 8, and for poverty rate the relationship changes as shown in Figure 9. This is probably due to some affluent areas being located in the very centre of Paris where a lot of stations are accessible within 2km. In fact, when looking at the 'less poor' areas, they have access to a comparable number of stations as the poorest areas.

This is interesting because while the poorest areas (poverty rate over 40%) often have a nearest station closer to them than very affluent areas, these areas still don't benefit from a lot of choice in their journey by public transport.

Overall, R-squared values are very low, both when considering the nearest public transport station and the number of stations within 2km of the area centroid. This seems to indicate that while income is not irrelevant for explaining access to the public transport network, it is far from the only factor that is at play. Furthermore, while the indicators in this section take into account the transport system and infrastructure, it does not consider the land-use system and the location of facilities in the urban area. Thus, there is a need for an indicator that can consider interactions between both the transport infrastructure system and the land-use system.

5. Results and analysis

5.1. Clustering results and analysis

As mentioned above in section 4.7, eight sets of clusters were computed, using the k-means clustering method, with the number of clusters for each set being determined using the elbow method.

5.1.1. Employment data

Two sets of clusters were computed for employment data: one with accessibility to jobs without separating the jobs and population into socio-professional categories, and another set of clusters where accessibility to jobs is considered separately depending on socio-professional categories.

		Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Mean	Population density	8820,08	3938,05	11158,57	19684,49	26756,57	41582,17
	Accessibility to jobs	1,00	0,68	0,79	0,73	1,28	1,12
	Ratio of social housing	0,09	0,11	0,33	0,74	0,04	0,19
	Median available income	33546,62	26598,60	20712,76	17312,28	41178,60	28400,25
	Ratio of households that own at least one car	0,75	0,87	0,65	0,55	0,44	0,31
Std error	Population density	5382,62	2989,16	6592,11	11496,81	9330,64	12249,02
	Accessibility to jobs	0,09	0,14	0,13	0,14	0,08	0,11
	Ratio of social housing	0,09	0,10	0,14	0,15	0,05	0,16
	Median available income	4915,05	3978,03	3112,28	2205,03	6138,08	4831,24
	Ratio of households that own at least one car	0,11	0,06	0,13	0,13	0,14	0,11
Size		687	887	833	530	415	566

Table 5: Structure of the clusters created from accessibility to jobs

The general accessibility to jobs is considered first. For this accessibility measures, the socio-professional category to which a person or a job belongs is not considered. Table 5 describes the mean values and standard errors of variables for each cluster. Meanwhile, figure 10 shows the spatial shape of the clusters in the urban area.

Cluster 0 is characterized by a decent accessibility, having the third highest accessibility to jobs. It also has the second lowest percentage of social housing. It has the second highest income but has the second lowest population density among the studied clusters. Furthermore, it has a high percentage of households owning a car. This cluster seems to be characterized by above average income (high middle class or higher) but also represents less dense areas of the urban area, with a high reliance on car, though accessibility by public transport is still above the average of all clusters. It seems to correspond to affluent suburbs.

Cluster 1 has the lowest population density, quite average income (since the average in the urban area is approximately 25898 € per year), and a high reliance on car, since accessibility by public transport is low and car ownership is high. Taking into account the spatial distribution of areas in this cluster, it is quite clear that these areas correspond to the areas far from the centre and from major public transport lines.

Cluster 2 has below average income and above average car ownership. It also has the second highest percentage of social housing at 33%. Accessibility to jobs is poor. This cluster seems to represent low-income suburbs, as population density is quite high but nowhere near that of the city centre.

Cluster 3 seems typical of “grands ensembles” and ‘*cités*’, that is to say districts built in the second half of the 20th century that are mostly made up of social housing apartments as mentioned previously. Furthermore, accessibility is the second lowest among all clusters, income is the lowest, and social housing accounts for 77% of housing. What is also noticeable is the high population density. Looking at the map, some of these areas are very close and other very far from the city centre. Most of these areas are located to the east, and to the north-east specifically.

Cluster 4 represents highly central areas that are very affluent. It also has the highest accessibility among all clusters. It seems to correspond to the rich neighbourhoods of the centre of the urban area. Social housing is almost non-existent.

Finally, cluster 5 has a similar accessibility to cluster 4 but is characterised by a lower income, though the mean income in this cluster is still above the median income in the area. In the case of cluster 4 and 5, since these represent neighbourhoods close to the centre, car ownership is quite low. Cluster 5 also has the highest population density among all clusters and has the third highest percentage of social housing.

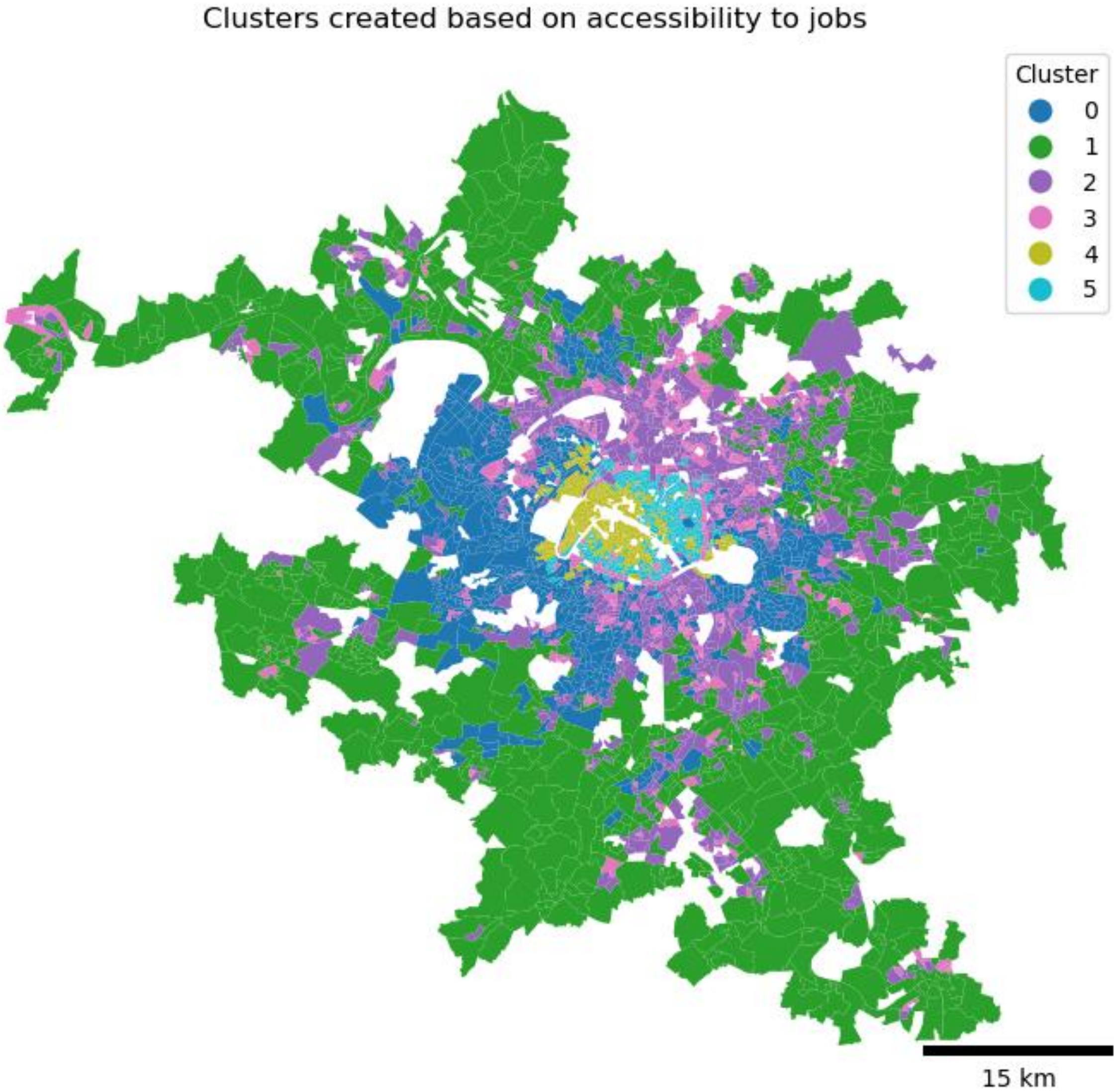

Figure 10: Map of the urban area showing the cluster to which each area belongs

		Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Mean	Population density	31118,36	17041,09	4392,93	39259,78	6648,62	10592,01
	Ratio of social housing	0,05	0,63	0,21	0,29	0,15	0,11
	Median available income	38794,78	17630,72	24182,94	25596,62	25588,10	33330,80
	Ratio of households that own at least one car	0,38	0,58	0,85	0,35	0,79	0,71
	Accessibility to CS1 jobs	1,93	1,07	0,79	1,56	1,17	1,50
	Accessibility to CS2 jobs	1,45	0,84	0,63	1,20	0,92	1,16
	Accessibility to CS3 jobs	1,37	0,79	0,61	1,13	0,87	1,11
	Accessibility to CS4 jobs	1,51	0,87	0,65	1,25	0,95	1,21

	Accessibility to CS5 jobs	1,46	0,85	0,62	1,21	0,92	1,17
	Accessibility to CS6 jobs	1,52	0,87	0,63	1,25	0,95	1,21
Std error	Population density	11420,79	10335,88	4537,80	14549,66	4797,57	6270,12
	Ratio of social housing	0,07	0,20	0,19	0,20	0,12	0,12
	Median available income	6546,71	2201,94	5265,48	4192,46	4674,74	5452,85
	Ratio of households that own at least one car	0,15	0,13	0,10	0,11	0,12	0,12
	Accessibility to CS1 jobs	0,13	0,16	0,16	0,14	0,13	0,14
	Accessibility to CS2 jobs	0,08	0,11	0,12	0,09	0,08	0,08
	Accessibility to CS3 jobs	0,08	0,10	0,12	0,09	0,07	0,08
	Accessibility to CS4 jobs	0,08	0,12	0,13	0,10	0,08	0,08
	Accessibility to CS5 jobs	0,08	0,11	0,13	0,09	0,08	0,08
	Accessibility to CS6 jobs	0,09	0,12	0,14	0,10	0,09	0,09
Size		567	727	520	462	996	646

Table 6: Summaries of the clusters created using accessibility to jobs per category

Then, the clusters created based on accessibility to jobs per category are considered. In this case, access to employment depends on both the category of the position and the socio-professional category of the person for competition effects. Table 6 details the characteristics of the clusters, and figure 11 shows their location in the urban area. Using the elbow method, the analysis was performed with six clusters.

Cluster 0 seems to correspond to wealthy, dense urban areas. Indeed, it has the highest median income across all clusters in this analysis. The map further supports this analysis, since areas belonging to cluster 0 are located in the west of Paris proper, and in the western inner urban areas, which are both dense and quite wealthy areas. This cluster also has the highest accessibility to jobs of all categories, though since the income of the cluster is high, population inside the areas of the cluster probably belong mostly to CS2 and CS3 (business owners and executives respectively). As could be expected, the ratio of social housing is quite low. Overall, this cluster is quite similar to cluster 4 from the previous analysis.

Cluster 1 has a quite high population density, and most importantly the lowest income across all clusters as well as the highest ratio of social housing. It seems to correspond to ‘grands ensembles’ or ‘cités’, similarly, to cluster 3 in the previous analysis. This interpretation can be supported by the fact that areas belonging to this cluster are quite spread on the map. It has the second lowest accessibility to jobs of all categories.

Cluster 2 has the lowest population density of all clusters, as well as the lowest accessibility to jobs overall. Thus, reliance on car would be necessary, which is shown by the fact that approximately 85% of households own at least one car. Median income is a bit below the median of the studied region which is approximately 25898 €. The ratio of social housing is rather average. This cluster corresponds to areas further from the centre that are less urbanized of the urban area, as seen on the map.

Cluster 3's mean income is quite similar to that of Cluster 2: it is almost the same of the median income of the entire study area. However, this cluster is very close to the centre (most areas are located directly in Paris proper as seen on figure **X**). It is the cluster with the highest population density. Due to its centrality, it has the second highest accessibility to jobs of all categories. The ratio of social housing is quite high at 29%. As with cluster 1, most households don't own a car.

Cluster 4's structure is quite similar to that of cluster 2, with mean income and population density being a bit higher, and the ratio of social housing being a bit lower. Accessibility however is higher than Cluster 2's, though it is still the third lowest accessibility among all clusters. Conversely, car ownership in this cluster is the second highest among all clusters. This cluster seems to correspond to areas that are further from the centre, but a bit more connected than areas in cluster 2.

Finally, Cluster 5 represents wealthy areas, but these areas are further from the centre than areas in Cluster 0. The most notable differences with areas in Cluster 0 is a higher ownership of cars, a lower population density and lower accessibility, though accessibility is still the third highest among all clusters. This cluster seems to represent wealthy suburbs.

Overall, accessibility to all categories of jobs seem to be correlated: clusters often have good or bad accessibility to all job categories at the same time. Overall, areas located near the centre have better accessibility, as expected. Low-income areas have poor accessibility, whereas for areas with average income, there are many diverse situations: some have quite good accessibility whereas others have some of the lowest accessibility. This seems to confirm that the relationship between income and accessibility is quite complex.

Comparing this set of clusters to the previous one where socio-professional categories were ignored, there are some differences, most notably for areas with average income. Indeed, for central areas, and affluent or poor neighbourhoods, the clusters' structures are mostly consistent. However, for average income areas, distance to the centre seems to become a more important factor when job category is considered. Indeed, when considering general accessibility to jobs a lot of areas are grouped in cluster 1, whereas when taking into account job category, these areas are split between clusters 2 and 4, the latter also encompassing areas belonging to cluster 2 in the analysis of general access to jobs.

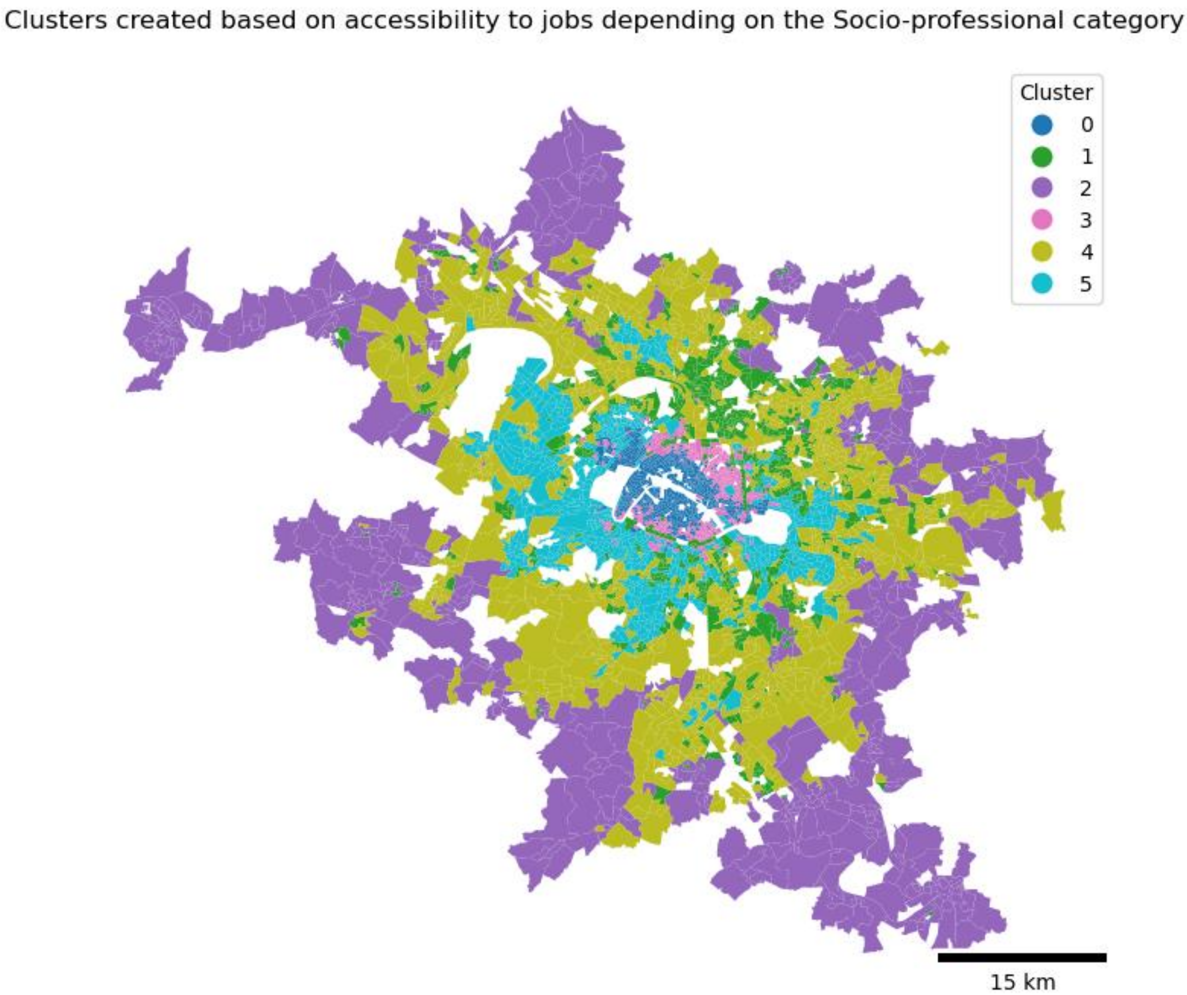

Figure 11: Spatial distribution of clusters made by accessibility to jobs per category

5.1.2. Education data

Two sets of clusters were created: one based on access to primary schools, middle schools and high schools, and another based on accessibility to various higher education facilities.

Regarding primary and secondary education, the clusters are described in table 7 and are shown on figure 12. It is noticeable that the clusters in this analysis resemble those obtained from accessibility to jobs by category. Roughly speaking, between those two sets, clusters 2 are mostly the same, cluster 5 in the previous analysis corresponds to cluster 3 for primary and secondary education, clusters 4 are similar, as well as clusters 1. Finally, cluster 0 for jobs is similar to cluster 5 for primary and secondary education, and cluster 3 for jobs is almost equivalent to cluster 0. Of course, there are some differences in the exact makeup of each cluster, but the same general archetypes seem to be present.

Thus, cluster 0 is representative of average income but central areas. Cluster 1 represents ‘cités’ i.e. densely populated areas with low income and a high proportion of social housing. Cluster 2 represents areas with average income and low density, which are

		Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Mean	Population density	39830,32	16570,37	3674,97	10167,37	7366,37	30925,26
	Accessibility to middle schools	1,31	0,87	0,66	1,28	1,00	1,62
	Accessibility to primary schools	1,28	0,86	0,65	1,26	0,99	1,59
	Accessibility to high schools	1,29	0,87	0,67	1,29	1,01	1,60
	Ratio of social housing	0,27	0,64	0,17	0,10	0,16	0,05
	Median available income	25686,37	17645,24	25402,66	34295,41	24832,58	38960,99
	Ratio of households that own at least one car	0,35	0,58	0,87	0,73	0,77	0,38
Std error	Population density	14333,55	9898,74	3428,14	6216,92	4927,81	10940,22
	Accessibility to middle schools	0,13	0,15	0,15	0,11	0,09	0,10
	Accessibility to primary schools	0,12	0,15	0,15	0,11	0,09	0,10
	Accessibility to high schools	0,12	0,15	0,15	0,10	0,09	0,10
	Ratio of social housing	0,20	0,19	0,14	0,10	0,13	0,06
	Median available income	4357,71	2248,53	5117,83	5030,88	4246,00	6514,45
	Ratio of households that own at least one car	0,12	0,13	0,09	0,12	0,12	0,14
Size		482	746	556	619	969	546

Table 7: Summary of clusters for primary and secondary education

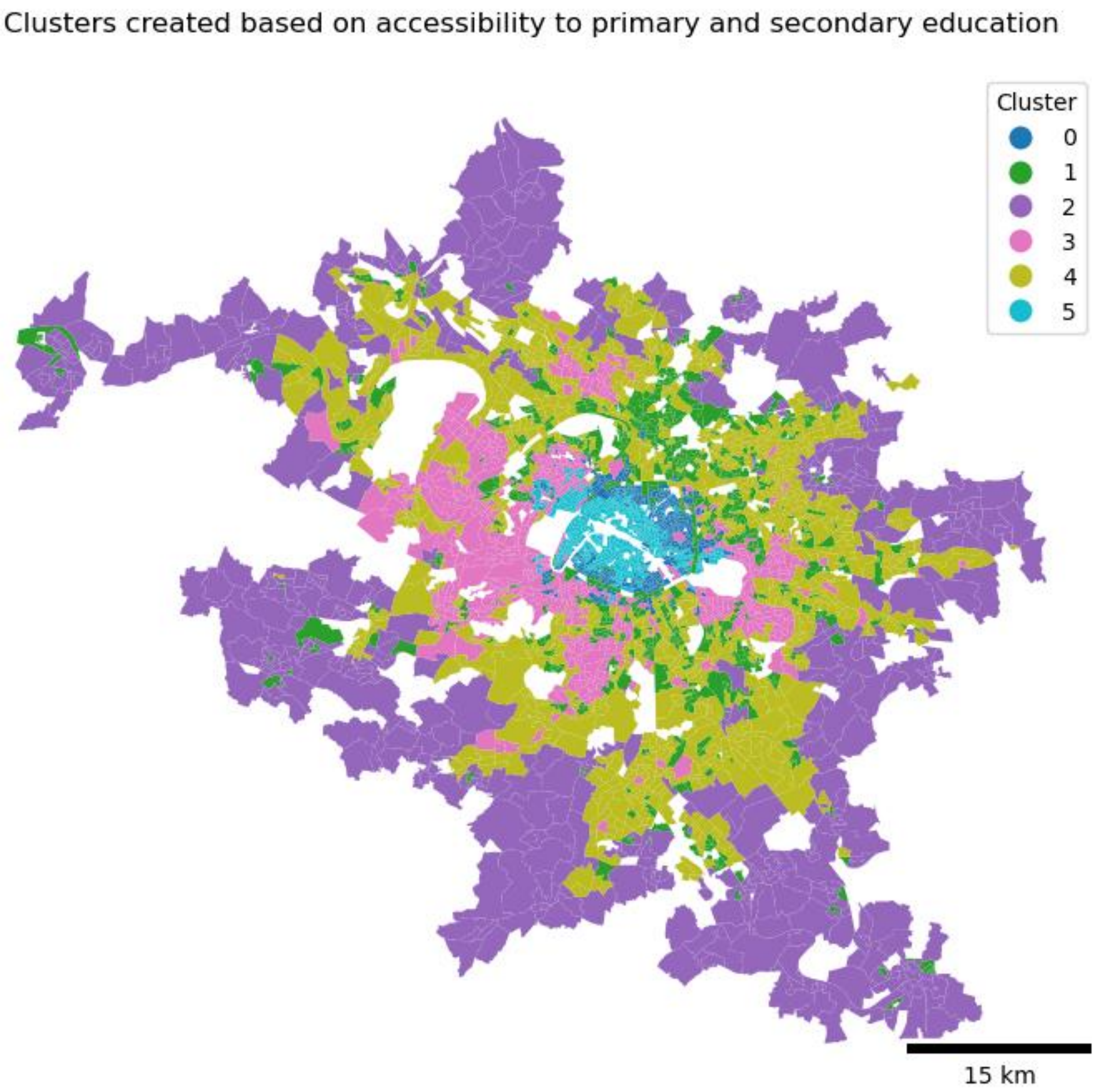

Figure 12: Map showing the spatial structure of clusters for primary and secondary education

mostly areas that are the furthest from the centre of the urban area. Cluster 3 represents suburbs with above-average income. Cluster 4 represents the 'middle' areas that are neither too far nor too close to the centre of the urban areas, and is also characterized by average income, accessibility and high car ownership. Finally, cluster 5 represents wealthy and central neighbourhoods, as it has both the highest income and the highest accessibility on average.

		Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Mean	Population density	40953,16	4981,09	23130,81	8617,13	14314,23	26468,59
	Accessibility to universities	0,60	0,36	0,47	0,52	0,34	0,70
	Accessibility to 'Grandes Ecoles'	0,24	0,13	0,18	0,21	0,13	0,28
	Accessibility to specialized education	0,12	0,07	0,09	0,10	0,07	0,14
	Accessibility to higher education programs in high schools	0,10	0,06	0,08	0,09	0,06	0,12
	Ratio of social housing	0,15	0,15	0,47	0,11	0,65	0,04
	Median available income	29749,70	24937,55	20553,23	32213,96	17160,14	41281,58
	Ratio of households that own at least one car	0,33	0,83	0,48	0,76	0,63	0,43
Std error	Population density	11899,91	3927,65	14842,37	5349,74	8367,98	9830,20
	Accessibility to universities	0,05	0,07	0,05	0,05	0,05	0,04
	Accessibility to 'Grandes Ecoles'	0,02	0,03	0,02	0,02	0,02	0,02
	Accessibility to specialized education	0,01	0,01	0,01	0,01	0,01	0,01
	Accessibility to higher education programs in high schools	0,01	0,01	0,01	0,01	0,01	0,01
	Ratio of social housing	0,13	0,12	0,21	0,10	0,19	0,05
	Median available income	3997,29	4442,60	2883,26	5286,93	1950,97	6179,93
	Ratio of households that own at least one car	0,12	0,10	0,14	0,11	0,11	0,15
Size		502	1092	499	834	579	412

Table 8: Summary of clusters for higher education

Regarding higher education some patterns are similar to previous analysis. Cluster 0 corresponds to average to high income central areas, though income in this cluster is a bit higher than its equivalents in other analyses, cluster 1 represents less densely populated areas with average income and a high reliance on car. Furthermore, cluster 2 represents low-income suburbs with dense population, cluster 4 represents 'cités' and cluster 5 represents wealthy, well connected areas, mostly in Paris and to its west.

One notable difference though is the shape of cluster 3 on the map. This cluster seems to represent above-average income areas (middle class and higher); however, it expands further south than its counterparts in previous analysis. This is probably due to the fact that a lot of higher education facilities have moved from Paris to the southern urban areas in previous years, including many prestigious schools. Thus, more areas in the south have good access to higher education than for other facilities.

Another notable insight from this analysis is that 'cités' areas (in cluster 4) have the worst access to higher education, which wasn't the case for accessibility to other facilities such as jobs or primary and secondary education. Indeed, while for some services ‘cités’ have better access than areas further from the centre in the outer urban areas, that is not the case for higher education. This seems to indicate that it is not only a problem of access to the transport system, but also a lack of higher education facilities near these areas.

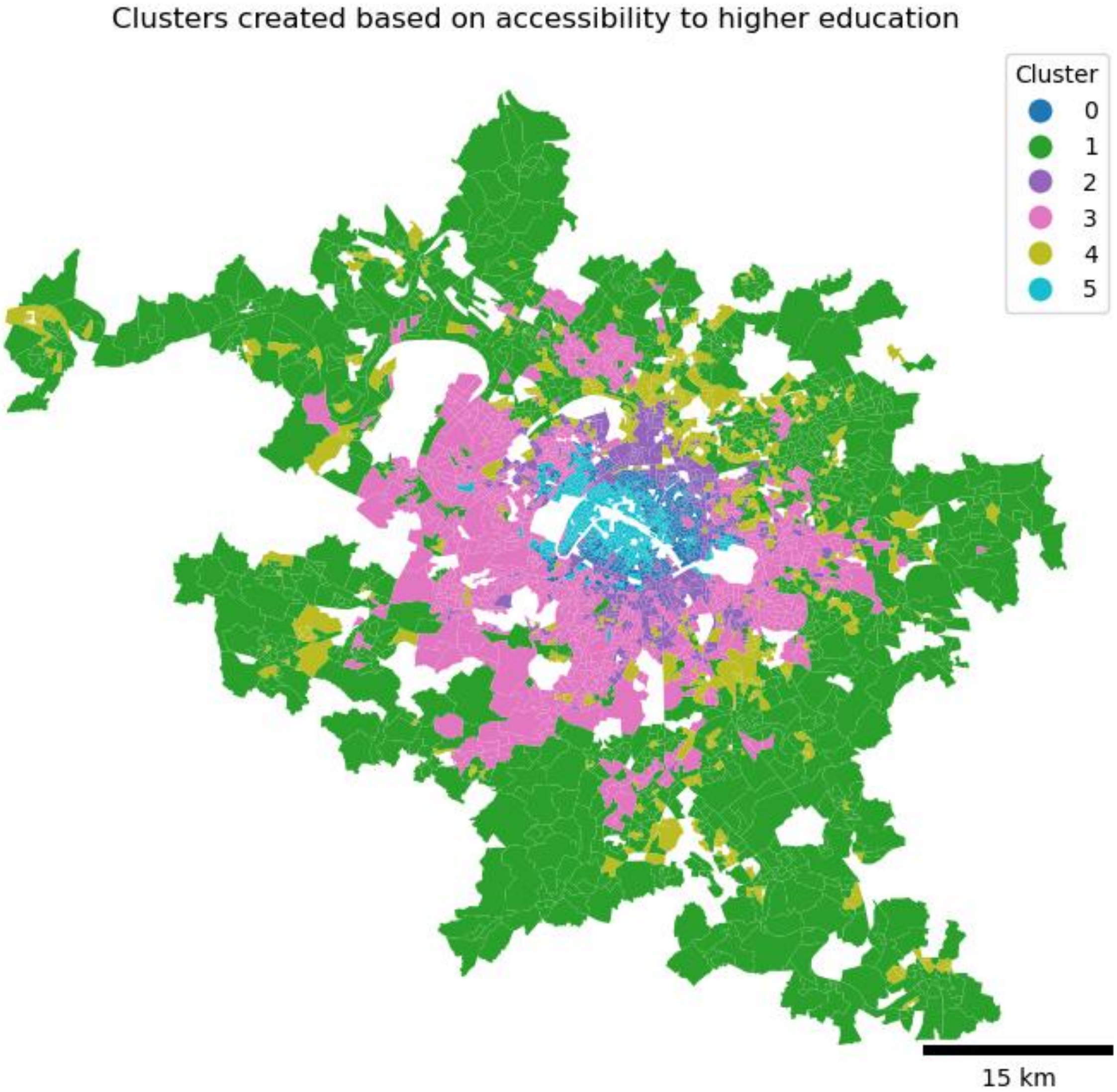

Figure 13: Spatial structure of clusters for higher education

Thus, for education, while access to primary and secondary education displays similar patterns as, for example, access to jobs, access to higher education seems to be more dependent on the location of services. The fact that opportunities are more concentrated in specific locations (universities and scientific campus) means that inequalities are more likely to arise if these locations are not easily accessible by public transport.

5.1.3. Health data

For health data, both access to general practitioners and access to hospitals were taken into account when creating the set of clusters.

At first glance, the structure of this set of clusters is similar to the structure of clusters built using the general accessibility to jobs. As in most analyses performed previously,

Table 9: Summary of clusters for accessibility to hospitals and general practitioners

the cluster corresponding to wealthy, central areas (cluster 5 in this case) also has the highest accessibility to both general practitioners and hospitals. The second highest accessibility is associated to the cluster containing areas of close to average income in the centre of the urban area (cluster 0 in this case). Cluster 1 is reminiscent of cluster 2 from the analysis performed on general job accessibility, with below-average income, a quite high proportion of social housing, and with a high proportion of households having access to at least one car, which seems to indicate that this cluster contains mostly below-average income suburbs.

Cluster 2 represents less dense, less connected areas, with the worst accessibility to both general practitioners and hospitals of all clusters. Cluster 3 represents above-average income suburbs, with a high proportion of households owning a car and a low ratio of social housing. Finally, cluster 4 corresponds to ‘cités’, as indicated by the low income and the high ratio of social housing in areas belonging to the cluster.

Overall, the conclusions that can be drawn from this set of clusters is quite similar to analyses above. Areas with the worst accessibility are the less dense, further from the centre areas. Then comes the poor neighbourhoods with a high proportion of social housing and a high density of population. After that, below-average income suburbs, followed by above-average income suburbs. Finally, central areas have the best access to hospitals and general practitioners, with the wealthiest of these areas having the best accessibility among all clusters.

		Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Mean	Population density	42266,88	12054,66	3998,29	8447,22	17649,05	28551,86
	Access to general practitioners	8,60E-04	6,31E-04	4,89E-04	7,75E-04	5,20E-04	1,03E-03
	Access to hospitals	2,59E-03	1,88E-03	1,42E-03	2,28E-03	1,55E-03	3,08E-03
	Ratio of social housing	0,23	0,30	0,14	0,09	0,72	0,05
	Median available income	27067,87	21128,91	25893,63	33119,91	17218,66	39637,36
	Ratio of households that own at least one car	0,33	0,63	0,87	0,77	0,58	0,41
Std error	Population density	12803,24	6813,97	3134,60	5288,10	11059,84	10031,43
	Access to general practitioners	8,74E-05	8,35E-05	1,09E-04	7,96E-05	1,06E-04	7,33E-05
	Access to hospitals	2,63E-04	2,59E-04	3,23E-04	2,44E-04	3,28E-04	2,04E-04
	Ratio of social housing	0,19	0,15	0,12	0,09	0,16	0,06
	Median available income	4528,08	3189,70	4384,25	5012,22	1963,71	6326,76
	Ratio of households that own at least one car	0,11	0,13	0,07	0,11	0,13	0,15
Size		488	714	878	751	569	518

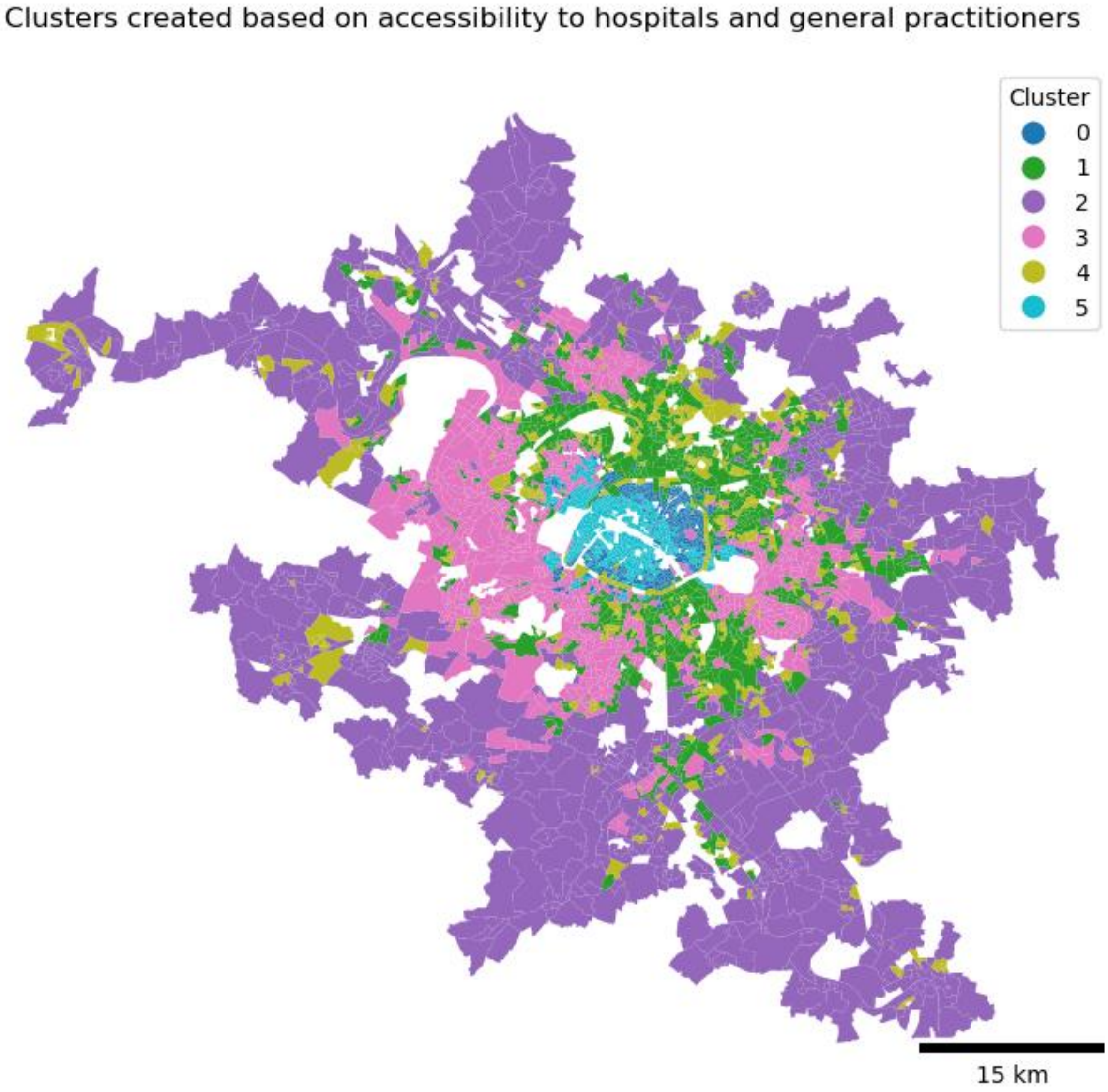

Figure 14: Map of clusters for accessibility to hospitals and general practitioners

5.1.4. Stores data

5.1.4.1. Convenience stores and supermarkets

In this set of clusters, the accessibility data considered are accessibility to convenience stores as well as accessibility to supermarkets. A map of the clusters is provided in figure 15 and table 10 presents a summary of the characteristics of each cluster in the set.

Cluster 0 presents indicator values characteristic of below-average income suburbs. Its main features are medium accessibility both for convenience stores and supermarkets and below-average median income. Furthermore, population density is quite high, as well as car ownership. There is a quite high ratio of social housing among all housing, as social housing represents 30% of housing on average in these areas. Areas belonging to this cluster are concentrated near Paris, mostly to the north, northeast and south of Paris.

Cluster 1 on the other hand is characterized by a low population density worst accessibility both to convenience stores and supermarkets. The median income is almost

Table 10: Summary of clusters for accessibility to supermarkets and convenience stores

equal to the average of the study area. Car ownership is high, which can be explained by the lack of access to public transportation. There are few social housings as it makes up only 15% of housing in areas belonging to this cluster. The position of areas belonging to this cluster on the map suggests that they represent less urbanized areas and are thus further from the centre.

Cluster 2 represents ‘cités’, similarly to previous clustering analysis for accessibility to other services.

Similarly, Cluster 3 represents wealthy central areas, similarly to other clusters in previous analysis. Once again, this type of cluster has the highest accessibility to services among all other clusters.

Cluster 4 represents suburbs characterized by above average income. It has the third best accessibility and high car ownership.

Cluster 5 represents the less affluent districts of Paris and its suburbs, though the income across these areas is still slightly above average. It has the second highest accessibility due to its centrality and has both low car ownership and a medium ratio of social housing.

Similarly to accessibility to other services, accessibility to GPs and to hospitals is correlated, which is not very surprising. The shape of the clusters and their characteristics resembles closely that of other clusters analysis, such as those for accessibility to jobs or accessibility to hospitals and general practitioners.

		Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Mean	Population density	13273,17	4303,70	17402,52	27933,50	8144,01	43078,23
	Accessibility to convenience stores	3,65E-04	2,80E-04	2,92E-04	5,82E-04	4,29E-04	4,93E-04
	Accessibility to supermarkets	1,76E-04	1,34E-04	1,40E-04	2,82E-04	2,08E-04	2,36E-04
	Ratio of social housing	0,31	0,15	0,71	0,05	0,09	0,22
	Median available income	21239,11	25443,85	17196,78	39726,46	33131,13	27421,59
	Ratio of households that own at least one car	0,60	0,86	0,59	0,42	0,77	0,32
Std error	Population density	7161,01	3306,45	11012,96	9734,34	5142,36	12287,99
	Accessibility to convenience stores	4,85E-05	5,88E-05	5,83E-05	4,23E-05	4,41E-05	5,04E-05
	Accessibility to supermarkets	2,44E-05	2,91E-05	2,89E-05	2,07E-05	2,28E-05	2,50E-05
	Ratio of social housing	0,15	0,13	0,16	0,06	0,09	0,18
	Median available income	3288,17	4402,80	1939,37	6356,67	4979,02	4648,28
	Ratio of households that own at least one car	0,13	0,08	0,13	0,15	0,11	0,11
Size		652	949	572	509	751	485

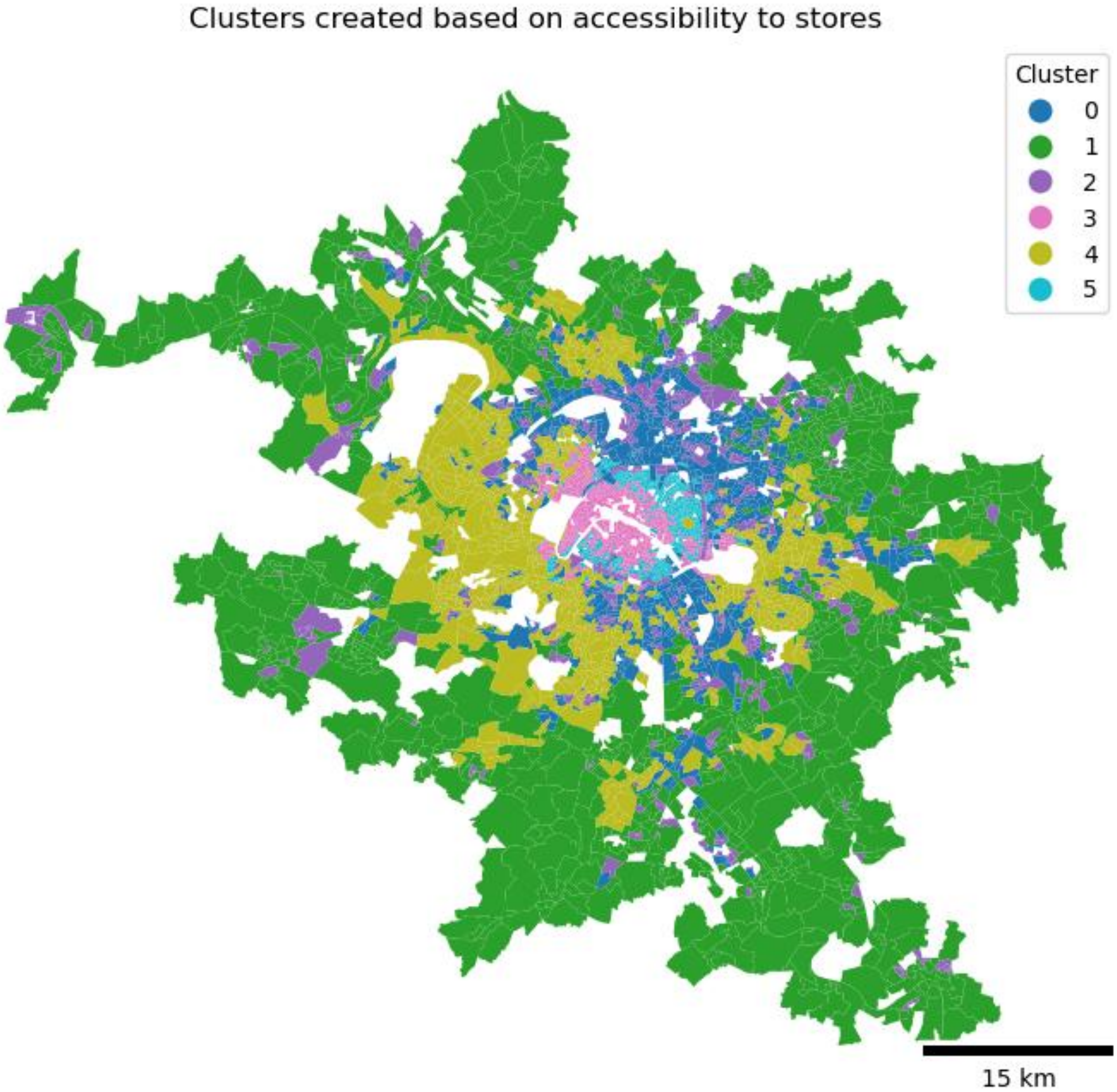

Figure 15: Map of clusters for accessibility to supermarkets and convenience stores

5.1.4.2. Pharmacies

		Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Mean	Population density	40296,78	10387,16	4617,05	17518,10	19994,17
	Accessibility to pharmacies	3,68E-04	2,33E-04	2,36E-04	3,77E-04	2,20E-04
	Ratio of social housing	0,15	0,31	0,09	0,06	0,72
	Median available income	30154,21	21025,87	28738,86	39338,39	17480,19
	Ratio of households that own at least one car	0,31	0,67	0,86	0,60	0,55
Std error	Population density	11839,75	6365,46	3479,39	9595,54	11734,20
	Accessibility to pharmacies	4,24E-05	4,90E-05	5,67E-05	4,82E-05	4,86E-05
	Ratio of social housing	0,15	0,14	0,09	0,08	0,15
	Median available income	5532,43	3164,25	4247,97	6764,73	2374,81
	Ratio of households that own at least one car	0,10	0,13	0,07	0,15	0,13
Size		672	930	1148	595	573

Table 11: Summary of clusters for accessibility to pharmacies

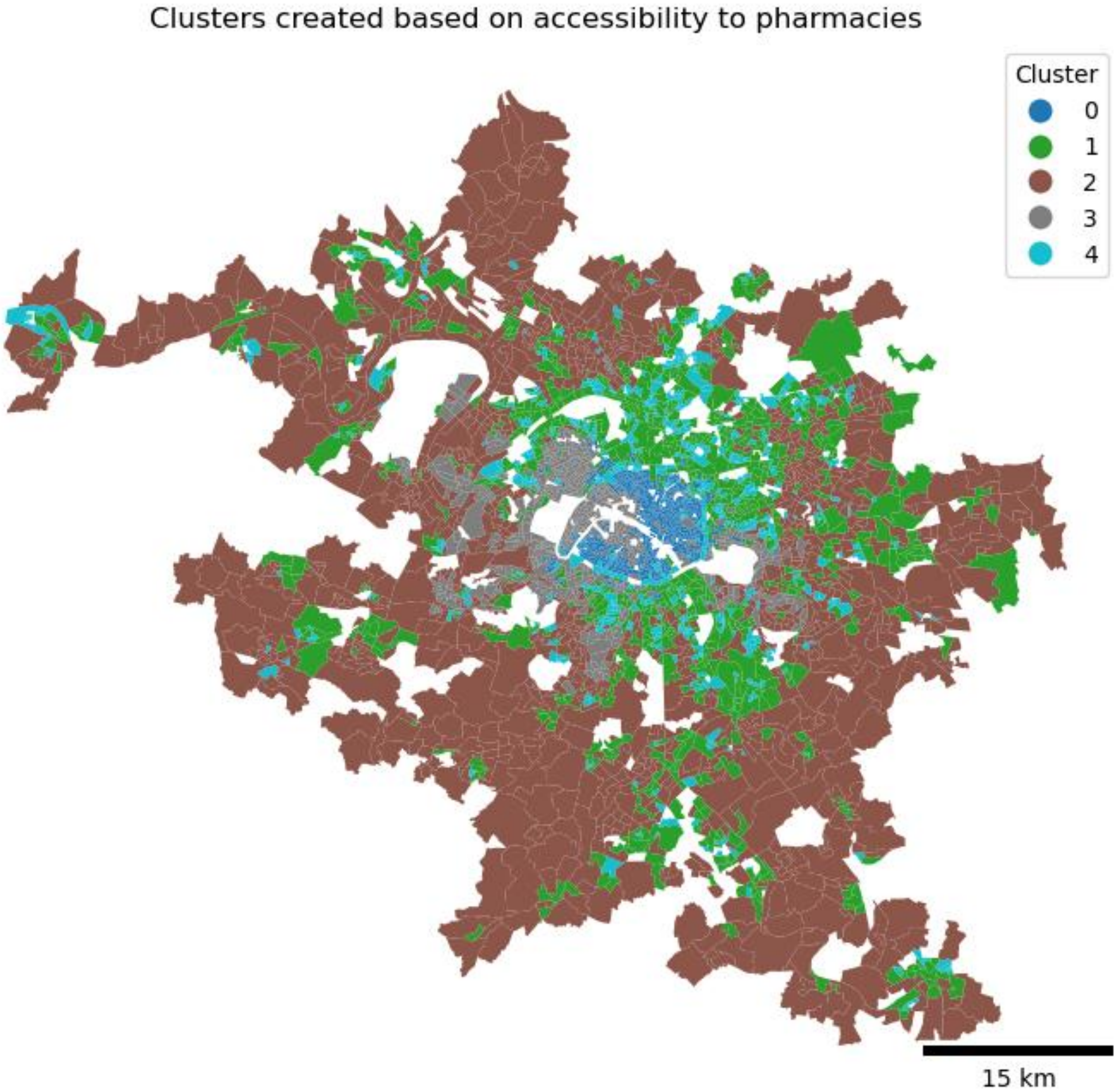

Figure 16: Map of clusters for accessibility to pharmacies

The following set of clusters is done only considering accessibility to pharmacies in addition to socio-economic variables. In this case using the elbow method only five clusters were created.

Figure 16 is a map of the clusters obtained, whereas table 11 summarises the main characteristics of the clusters in the set.

Clusters’ structures are similar to clustering analyses performed on other accessibility data. Namely, cluster 0 corresponds to central areas, cluster 3 corresponds to wealthy areas, where some areas are further from the centre, and also has higher car ownership. Cluster 3 roughly corresponds to some areas of cluster 3 and cluster 4 in the previous clustering.

Cluster 2 corresponds to less dense areas, with above average income and high reliance on car.

Cluster 1 corresponds to below average income suburbs, with quite low accessibility and the second highest car ownership.

Cluster 4 corresponds to 'cités' or 'grands ensembles': neighbourhoods with a high ratio of social housing, low income and in this case the lowest accessibility to pharmacists overall.

There are however some notable differences in the spatial repartition of these clusters. In this analysis there is only one cluster encompassing wealthy suburbs and central areas, whereas previously these areas were in two separate clusters. Furthermore, areas in cluster 1 are more spread out compared to clusters in previous analyses displaying similar socio-economic characteristics. This of course is probably because of the difference in the total number of clusters.

Table 12: Summary of clusters for accessibility to all types of services

5.1.5. Clustering multiple accessibility data types

		Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Mean	Population density	37266,11	16596,93	10571,29	4591,87	32346,65	6648,03
	Ratio of social housing	0,33	0,63	0,11	0,21	0,06	0,15
	Median available income	24556,66	17603,45	33609,57	23964,12	38041,43	25761,56
	Ratio of households that own at least one car	0,36	0,58	0,72	0,84	0,37	0,79
	Accessibility to hospitals	2,43E-03	1,62E-03	2,40E-03	1,18E-03	3,06E-03	1,82E-03
	Accessibility to general practitioners	8,05E-04	5,45E-04	8,12E-04	4,05E-04	1,02E-03	6,20E-04
	Accessibility to convenience stores	4,58E-04	3,07E-04	4,53E-04	2,32E-04	5,78E-04	3,48E-04
	Accessibility to supermarkets	2,19E-04	1,48E-04	2,20E-04	1,11E-04	2,79E-04	1,68E-04
	Accessibility to higher education	9,59E-01	6,55E-01	9,68E-01	5,13E-01	1,21E+00	7,42E-01
	Accessibility to primary and secondary education	1,26E+00	8,74E-01	1,27E+00	6,37E-01	1,58E+00	9,84E-01
	Accessibility to jobs	1,03E+00	7,34E-01	1,03E+00	5,58E-01	1,27E+00	8,12E-01
Std error	Population density	15499,93	10140,45	6310,96	4622,81	11766,67	4816,69
	Ratio of social housing	0,21	0,21	0,11	0,18	0,07	0,13
	Median available income	3915,21	2202,61	5280,44	5280,65	6699,56	4640,12
	Ratio of households that own at least one car	0,12	0,13	0,12	0,10	0,14	0,12
	Accessibility to hospitals	2,21E-04	2,40E-04	1,85E-04	2,65E-04	1,91E-04	1,74E-04
	Accessibility to general practitioners	7,24E-05	7,80E-05	6,13E-05	8,76E-05	6,97E-05	5,64E-05
	Accessibility to convenience stores	4,11E-05	4,38E-05	3,37E-05	4,92E-05	3,85E-05	3,06E-05
	Accessibility to supermarkets	2,00E-05	2,17E-05	1,80E-05	2,41E-05	1,93E-05	1,63E-05
	Accessibility to higher education	8,61E-02	9,22E-02	7,03E-02	1,06E-01	7,95E-02	6,66E-02
	Accessibility to primary and secondary education	1,08E-01	1,20E-01	9,73E-02	1,46E-01	1,06E-01	9,11E-02
	Accessibility to jobs	7,92E-02	9,29E-02	7,07E-02	1,13E-01	7,48E-02	6,64E-02
Size		440	707	627	549	624	971

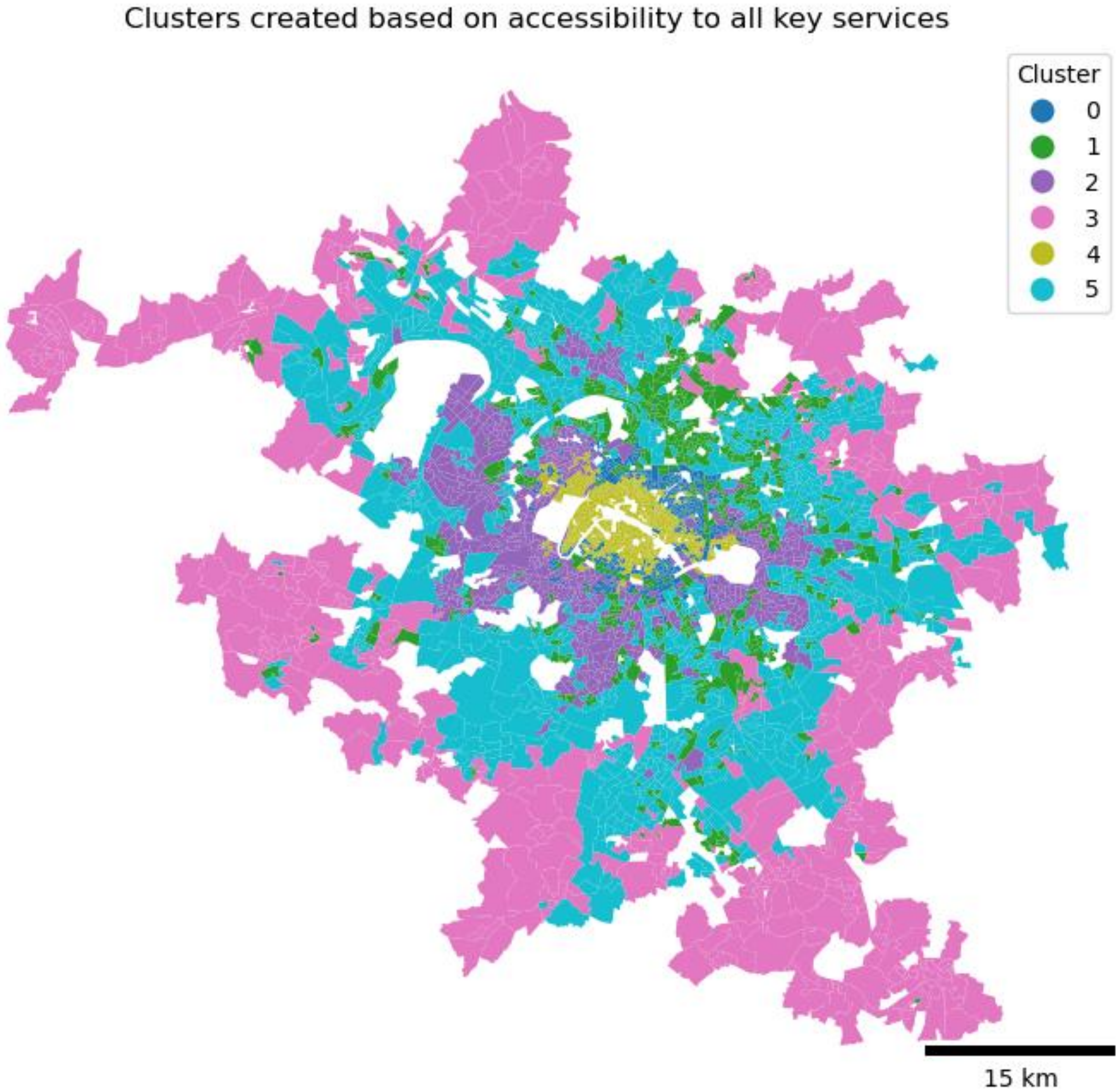

Figure 17: Map of clusters for accessibility to all types of services

This clustering analysis was performed on a selection of accessibility measures of different types. Table 12 summarises the characteristics of the clusters and figure 17 presents a map of the clusters in the Paris urban area.

Overall, accessibilities to services seems to be correlated across services. That is to say, areas of a cluster have good access to all services or bad access to all services. As expected, areas in the centre of the urban area (cluster 4) have high accessibility to services. Indeed, the centre of the urban area benefits to access to most service lines, since service lines often converge towards Paris.

On the other hand, areas further from the centre in cluster 3 suffer from poor accessibility to all services. Indeed, these areas are reliant on travel by car, probably due to the lack of public transport infrastructure in these zones.

In between the inner suburbs and the outer urban areas, areas that belong to cluster 5 have medium accessibility, and a high ratio of car ownership.

Cluster 2 seems to represent wealthy suburbs. These areas have quite high accessibility overall, and are located mostly to the west of the centre.

Cluster 1 represents 'cités', with a high proportion of social housing and a poor access to key services overall.

Finally, cluster 0 represents suburbs directly next to Paris and some areas near Paris. It has an overall quite good accessibility, very similar to cluster 2 for some services.

5.1.6. Insights from clustering

Overall, though there are some variations across the different analyses that were performed, the results obtained from the different sets of clusters are consistent with each other.

For most accessibility measures, two archetypes of areas have the lowest value. First, relatively less densely populated areas, that are close to the average income in the urban area. These areas are often located further from the centre of the urban area and thus have less access to transport infrastructure. Poor accessibility in these areas may explain the high proportion of households that own a car, since in the absence of convenient transit via public transport, car transport becomes much more necessary to access key services. The second archetype of areas that have bad access consistently across analyses are 'cités' or 'grands ensembles'. Though many of these areas are sometimes close to the centre, they still suffer from bad access. This type of cluster also has the lowest income of all clusters in all analyses.

Then, some clusters represent the 'suburbs' archetypes, often with one cluster representing the more well-off suburbs (with above-average income, lower population density and higher accessibility) and poorer suburbs (with below-average income, higher population density and lower accessibility). Finally, central areas are divided into two types. One of this type is characterised by income slightly above the average of the urban area, very good access, a quite high proportion of social housing and the highest population density across all types of areas. The other archetype of these central areas are very wealthy central areas, which have the highest accessibility and income, as well as high population density but a very low ratio of social housing.

Thus, when comparing accessibility and income using clustering, it appears that the poorest areas often have low accessibility, whereas areas with high income always have at least good accessibility compared to other clusters. There are more diverse situations for areas with average income however: some of these areas are located near the centre of the urban area and thus benefit from very good access to services, whereas other areas have comparable income but suffer from much lower accessibility, and are more reliant on travel by car, since these areas are located further from the centre, and are less densely populated.

5.2. Spatial regression

Spatial regression results are summarized in the tables below. Table 13 details the results of the spatial regression for accessibility to jobs, whereas table 14 shows the outputs for the model using accessibility to primary and secondary education facilities.

Variable	**Coefficient**	**Std.Error**	**z-Statistic**	**Probability**
CONSTANT	0,46098	0,01739	26,50175	0,00000
Population density (thousands of inhabitants/km^2)	0,00492	0,00012	2,68426	0,00727
Ratio of social housing	-0,01940	0,00163	-11,90161	0,00000
Median available income (in thousands of euros)	0,08377	0,00322	26,04563	0,00000
Ratio of households that own at least one car	-0,06758	0,00313	-21,59593	0,00000
W_accessibility_jobs_gtt	0,04825	0,00193	24,95327	0,00000
SPATIAL LAG MODEL IMPACTS				
Variable	**Direct**	**Indirect**	**Total**	
Population density (thousands of inhabitants/km^2)	0,0049	0,0002	0,0052	
Ratio of social housing	-0,0194	-0,0010	-0,0204	
Median available income (in thousands of euros)	0,0838	0,0042	0,0880	
Ratio of households that own at least one car	-0,0676	-0,0034	-0,0710	
Pseudo R-squared	0,9112			

Table 13: Spatial regression model with accessibility to jobs as the dependent variable

Variable	**Coefficient**	**Std.Error**	**z-Statistic**	**Probability**
CONSTANT	0,59591	0,02315	25,74328	0,0000
Population density (thousands of inhabitants/km^2)	0,00462	0,00261	1,77041	0,07666
Ratio of social housing	-0,02811	0,00231	-12,15824	0,0000
Median available income (in thousands of euros)	0,11484	0,00454	25,27336	0,0000
Ratio of households that own at least one car	-0,09663	0,00454	-21,28468	0,0000
W_accessibility_primary_secondary	0,04497	0,00211	21,27426	0,0000
SPATIAL LAG MODEL IMPACTS				
Variable	**Direct**	**Indirect**	**Total**	
Population density (thousand of inhabitants/km^2)	0,0046	0,0002	0,0048	
Ratio of social housing	-0,0281	-0,0013	-0,0294	
Median available income (in thousands of euros)	0,1148	0,0054	0,1202	
Ratio of households that own at least one car	-0,0966	-0,0046	-0,1012	
Pseudo R-squared	0,897			

Table 14: Spatial regression model with accessibility primary and secondary education as the dependent variable

	Population density (thousands of inhabitants/km^2)	Ratio of social housing	Median available income (in thousands of euros)	Ratio of households that own at least one car
Population density (thousands of inhabitants/km^2)	1,00	0,08	0,09	-0,77
Ratio of social housing	0,08	1,00	-0,66	-0,14
Median available income (in thousands of euros)	0,09	-0,66	1,00	-0,01
Ratio of households that own at least one car	-0,77	-0,14	-0,01	1,00

Table 15: Correlation matrix

For both model, most of the variations are captured by the models, with both pseudo R-squared values being above 0.89.

Both models display similar results in some aspects. In both cases, the median available income contributes positively to accessibility. This is expected as clustering analyses have shown that areas with high income often had better accessibility than other areas with otherwise similar characteristics. Furthermore, a higher proportion of social housing negatively impacts accessibility, which could support the idea that neighbourhoods with a high proportion of social housing (which are often ‘cités’) suffer from a specific lack of access. Finally, a high proportion of households that have access to a car is also associated with less access.

Regarding population density, it is a significative value for the job accessibility model, but not for the primary and secondary education model. In the former case, it is associated with higher accessibility, which is not surprising since more densely populated areas tend to be located at the centre of the urban area, where more service lines and stations are located.

In both models, indirect effects associated with variables tend to be low.

Overall, the insights from these models are coherent with those obtained from clustering: income is an important factor in the level of accessibility of an area, but other factors contribute as well.

The correlation matrix shows strong correlation between income and the ratio of social housing, and population density and car ownership. Overall, these relationships are expected: social housing primarily serves people with below average income, and people living in more dense areas are less dependent on car transportation due to being closer to daily life facilities.

The detailed summaries of the spatial regression models are available in appendix B.

6. Discussion

6.1. Data used

All data used in this study is directly available on different public data websites for France. However, restricted access data was not used. Future studies could improve the approach of this paper, for example by benefitting from a better calibration of model parameters by using detailed trips data from the travel survey.

Furthermore, the lack of data for employment at the 'iris' scale means that jobs data in our study were estimated using rough employment figures from company registries. Though these estimations were not too far off the real figures (when comparing data at the city level), the estimation method used in this study may lead to inaccuracies. For example, it may lead to underestimating the importance of self-employment, as self-employed individuals are not always considered in companies registries for confidentiality purposes.

Finally, indicators chosen as supply and demand for each type of service may also be subject to change in future studies. If data is available, using a metric such as surface area to estimate the number of opportunities for stores could be helpful, as in our study all stores of a category are considered to represent the same number of opportunities, even though, in a same category, the range of products offered in a store can vary greatly. Regarding hospitals, other metrics could be considered for the number of opportunities such as the number of hospital beds. All general practitioners are considered to represent one opportunity each, but to differentiate between different general practitioners, the number of opening hours could be used, since some doctors work full time in one medical office whereas others work at multiple offices.

6.2. Travel-time calculations

In this study, one travel-time matrix was computed using r5py, using a weekday morning as the departure time. While this choice is relevant for accessibility to jobs and to education, it may be less suited to study access to other services considered in this study, such as health services and stores. Furthermore, it also means that this study doesn't take into account atypical working hours, such as night shifts, or part-time work. This, in particular, could lead to a less accurate representation of accessibility to low-income jobs, which are more often part-time jobs.

Another limitation of the travel-time computation method is that, while it was possible to take into account timetables and frequency of service lines using GTFS data, this study did not take into account the reliability of public transport. Indeed, recurring delays or cancelation on certain service lines can dramatically decrease the accessibility of an area, even if it theoretically has good access. Furthermore, since some transport service lines may be less reliable than others, it could create potential inequalities in addition to accessibility inequalities that were already observed.

In this study, the data used to make the travel-time matrix was only the travel-time from the origin to the destination. Due to computational power limitations, it wasn't possible to use the full, detailed breakdown of the trips from one zone to another. This means that while the travel time is accurate, this study cannot account for transfers or check precisely which public transport modes are used when making the travel time calculations. Future studies making use of detailed travel information may be able to conduct more detailed analysis about reliability of service lines involved (as mentioned above), the number of transfers necessary to reach a destination and the number of transport legs.

6.3. Clustering and spatial regression methods

For technical reasons and time constraints, the k-means clustering method was used, even if other methods could have been more relevant. For example, a spatial clustering method was used in Adorno et al. (2025), in combination with spatial regression models. Thus, future studies could benefit from using spatial regression models in order to properly take into account the spatial characteristic of the data studied.

While the spatial regression model used in this study is a spatial lag model, there are a wide variety of spatial regression models that could potentially be considered when studying accessibility data and inequalities. One possibly interesting method could be to build spatial clusters and use theses clusters as the basis for a spatial regression model using spatial regimes.

6.4. Results

Using Shen's accessibility (Shen, 1998), this study properly considers competition effects when calculating accessibility to services. However, as pointed out in Giannotti et al. (2022), results obtained using Shen's accessibility can lead to underestimating inequalities of access, since the calibration of the parameters of this measure assumes that travel surveys reveal unconstrained preferences. Thus, whereas the results of our study are interesting in that they reveal inequalities and consider competition effects, one must be aware that some inequalities of access could be underestimated. Other accessibility studies in the Paris urban area could benefit from using different accessibility measures to verify that results from this study are accurate.

The findings of this study hint at a positive correlation of income and population density with accessibility, and a negative correlation of the ratio of social housing and the ratio of car ownership with accessibility. However, the facts that these variables have statistical links does not necessarily indicate a causality link, or the direction of this causality.

For example, public transport accessibility may be poor in areas with high car ownership because decision makers consider that these areas can rely on car travel in the absence of public transport. However, another possible explanation is that people having access to a car are more likely to move to areas further from the public transport network, since they have an alternative to public transport. In this study, no causal link between variables can be rigorously established, and our observations are limited to correlations and associations between variables.

In order to keep the analyses in this study easily interpretable and understandable, not all available socio-economic variables were considered. Other variables could be considered in future studies, such as the number of renters compared to the number of homeowners in an area, the average construction year of housing in the area, the size of households or the size of residences in the area.

Poor accessibility figures observed in some areas could be explained by a lack of facilities near or in these areas, a lack of public transport infrastructure, or both. For higher education in particular, the concentration of facilities to the south of Paris (in campuses such as 'Paris-Saclay') has a visible impact on accessibility compared to other types of services, since areas to the northeast of Paris had worse accessibility to higher education, compared to other facilities. It can be expected that areas with the lowest accessibility values suffer from both a lack of local facilities and a lack of public transport infrastructure. Further analyses could be done, focusing on one specific type of facilities, and comparing accessibility to these services and their repartition in the urban area.

Focusing on the two archetypes of areas with the poorest accessibility, different insights can be drawn. Regarding less densely populated areas with average income and high car ownership, the lack of access through public transport can be explained by the fact that over less dense areas, public transport infrastructure is often less efficient, resulting in lower accessibility. Considering 'cités' neighbourhoods, these areas are often densely populated, meaning that facilities and transport infrastructure could be expected to be efficient. However, these areas still suffer from very low access, which could be explained by a lack of facilities (compared to the size of the population) and by a lack of public transport infrastructure suited to the needs of the population.

From a sufficientarian perspective, it is hard to draw conclusions from this study as the abstract nature of the accessibility measure chosen renders choosing a threshold for 'sufficient' access hard. However, from an egalitarian perspective, there are clear inequalities in the distribution of accessibility among the population. In particular, as mentioned above, when areas have comparable figures in terms of location in the urban area, population density, and car ownership, more wealthy areas perform better than lower income areas.

Comparing the results of this study with other articles from the literature, it appears that the accessibility in the Paris urban area is less equal than some other cities in the literature. For example, in Foth et al. (2013), it was found that in Toronto the public transport system tends to benefit disadvantaged areas, which is not the case in the case of the Paris urban area. In our case, high-income areas often benefit more from the land-use and transport systems. Thus, currently urban planning and transportation planning in the Paris urban area fails in correcting existing economic inequalities.

7. Conclusion

This study aimed to investigate potential links between accessibility to several key services via the public transit system, and socio-economic variables in the Paris urban area. The goal was to identify potential differences in access between the centre of the urban area, the inner urban area and the outer urban area. Furthermore, it was considered possible that 'cités' districts -neighbourhoods built during the second half of the 20th century, characterised by a high proportion of social housing, low income and high unemployment- might display specific patterns in terms of accessibility. Several types of facilities were considered to account for potential differences in the location of facilities in the urban area. Accessibility was chosen as an indicator to account for interactions between the land-use system and the transport system.

Since competition effects were considered significant for some services, Shen's accessibility was chosen as an accessibility metric. Furthermore, this measure avoids the need to arbitrarily determine a travel-time threshold. The parameters for this accessibility formula were calibrated using travel survey data from 2020. Data on key services and facilities were collected mostly on the French governmental open data platform. Depending on the type of service considered, different indicators were used to determine the number of opportunities available for each type of service in each area. To establish the demand for these services, census data was used. Iris census tracts were considered, as these are mostly homogeneous in terms of population size.

Preliminary analyses revealed a potential statistical link between the median income in an area and the availability of public transport in said area. However, these preliminary results also seemed to indicate that there were a wide variety of situations regarding availability of public transport, especially for areas that were located around the average, income wise.

The travel time matrix for the urban area was computed using the r5py library. Two types of analyses were then performed on the accessibility data obtained: clustering and spatial regression.

Clustering results provided several insights. Areas that have above-average income tend to have high accessibility to all types of services. On the other hand, areas that have below-average income tend to have lower accessibility, compared to other areas with similar characteristics. Interestingly, for areas that have about average incomes, accessibility can vary greatly: less densely populated further from the city centre have very low access to all types of services, whereas some areas have similar levels of income but better access due to their location closer to the centre of the urban area. Across the different types of services, patterns in accessibility were quite similar, though there were some notable differences in some cases. For example, in the case of higher education, areas to the south of the urban areas tended to have better access than for other services, which is in all likeliness due to the spatial repartition of opportunities in the urban area. Clustering results also seem to confirm the specificity of 'cités' neighbourhoods: these districts are quite spread out over the urban area, and have very low levels of accessibility on average, even though some of these neighbourhoods are quite close to the centre of the urban area and are densely populated. Thus, this could indicate that these areas face specific disadvantages in terms of accessibility.

Spatial regression results further supported this analysis, by showing that the higher the ratio of social housing in an area, the lower the accessibility. Spatial regression results also confirmed that income, generally, is associated with higher accessibility, whereas high car ownership is associated with lower accessibility. Population density was not found to be a significative variable for all types of accessibility.

Thus, the results of this study seem to indicate that while income is positively correlated with accessibility to different type of facilities, it is not the only variable that has to be taken into account. Specifically, for low- and high-income ranges, accessibility tends to be very low and very high respectively, but for average income ranges there are more diverse situations. The ratio of social housing and the ratio of car-owning households also are important variables correlated with low access.

When focusing on districts which correspond to the characteristics of 'cités', these districts seem to face specific challenges, even compared to other low-income suburbs. 'Cités' have lower income, a much higher proportion of social housing, and low car ownership (between 50% and 60%) when taking into account the lack of access through the public transport system. Thus, in addition to facing poor accessibility to services via public transport, inhabitants of these neighbourhoods may not be able to always rely on car travel to make up for this lack of access through the public transport network.

In conclusion, this study has highlighted inequalities in terms of accessibility to a range of different services. In general, lower income areas have worse access than higher income areas. Areas with average income have more diverse situations depending on their location in the urban area. When comparing areas with otherwise similar characteristics, it was found that areas with higher income were better off in terms of accessibility. Finally, there seems to be a specific deprivation for 'cités', as these neighbourhoods' poor accessibility figures even compared to poor suburbs areas next to them.

It remains to be seen if new extensions to the network will affect this relationship. Planned subway service lines linking Paris to several destinations in the 'Seine Saint-Denis' area could lead to an improvement in access in those areas. Indeed, this administrative area is renowned for its large number of 'cités' cities and districts.

Hopefully, this study can prompt more in-depths research focusing on specific types of services and provide useful insight for decision makers and city planners in order to address inequalities in the Paris urban area.

References

Ined - Institut national d'études démographiques. 'Population de la France-séries longues - Bases de données en ligne - Les chiffres'. Accessed 5 November 2025. https://www.ined.fr/fr/tout-savoir-population/chiffres/bases-donnees/population-de-la-france-series-longues/.

Bocarejo S., Juan Pablo, and Daniel Ricardo Oviedo H. 'Transport Accessibility and Social Inequities: A Tool for Identification of Mobility Needs and Evaluation of Transport Investments'. *Journal of Transport Geography*, Special Section on Theoretical Perspectives on Climate Change Mitigation in Transport, vol. 24 (September 2012): 142–54. https://doi.org/10.1016/j.jtrangeo.2011.12.004.

Bouzouina, Louafi, Ioannis Baraklianos, Patrick Bonnel, and Hind Aissaoui. 'Renters vs Owners: The Impact of Accessibility on Residential Location Choice. Evidence from Lyon Urban Area, France (1999–2013)'. *Transport Policy* 109 (August 2021): 72–84. https://doi.org/10.1016/j.tranpol.2021.05.022.

Castel, Robert. 'La discrimination négative :Le déficit de citoyenneté des jeunes de banlieue'. Sociologie. *Annales. Histoire, Sciences Sociales* 61, no. 4 (2006): 777–808.

El-Geneidy, Ahmed, David Levinson, Ehab Diab, Genevieve Boisjoly, David Verbich, and Charis Loong. 'The Cost of Equity: Assessing Transit Accessibility and Social Disparity Using Total Travel Cost'. *Transportation Research Part A: Policy and Practice* 91 (September 2016): 302–16. https://doi.org/10.1016/j.tra.2016.07.003.

Ermagun, Alireza, and Nebiyou Tilahun. 'Equity of Transit Accessibility across Chicago'. *Transportation Research Part D: Transport and Environment* 86 (September 2020): 102461. https://doi.org/10.1016/j.trd.2020.102461.

Fan, Yingling, Andrew Guthrie, and David Levinson. 'Impact of Light Rail Implementation on Labor Market Accessibility: A Transportation Equity Perspective'. *University of Minnesota: Nexus Research Group, Working Papers* 5 (January 2010). https://doi.org/10.5198/jtlu.v5i3.240.

Ford, Alistair C., Stuart L. Barr, Richard J. Dawson, and Philip James. 'Transport Accessibility Analysis Using GIS: Assessing Sustainable Transport in London'. *ISPRS International Journal of Geo-Information* 4, no. 1 (2015): 1. https://doi.org/10.3390/ijgi4010124.

Foth, Nicole, Kevin Manaugh, and Ahmed M. El-Geneidy. 'Towards Equitable Transit: Examining Transit Accessibility and Social Need in Toronto, Canada, 1996–2006'. *Journal of Transport Geography* 29 (May 2013): 1–10. https://doi.org/10.1016/j.jtrangeo.2012.12.008.

Fransen, Koos, Kobe Boussauw, Greta Deruyter, and Philippe De Maeyer. 'The Relationship between Transport Disadvantage and Employability: Predicting Long-Term Unemployment Based on Job Seekers' Access to Suitable Job Openings in

Flanders, Belgium'. *Transportation Research Part A: Policy and Practice* 125 (July 2019): 268–79. https://doi.org/10.1016/j.tra.2018.01.023.

Garbin, David, and Gareth Millington. 'Territorial Stigma and the Politics of Resistance in a Parisian <em>Banlieue</Em>: La Courneuve and Beyond'. *Urban Studies* 49, no. 10 (2012): 2067–83.

Geurs, Karst T., and Bert van Wee. 'Accessibility Evaluation of Land-Use and Transport Strategies: Review and Research Directions'. *Journal of Transport Geography* 12, no. 2 (2004): 127–40. https://doi.org/10.1016/j.jtrangeo.2003.10.005.

Hansen, Walter G. 'How Accessibility Shapes Land Use'. *Journal of the American Institute of Planners* 25, no. 2 (1959): 73–76. https://doi.org/10.1080/01944365908978307.

Lévêque, Antoine. 'Une double absence dans les quartiers populaires ?:Ancrage politique en déclin et relégation dans l'action publique'. *Politix* 143, no. 3 (2023): 111–47. https://doi.org/10.3917/pox.143.0111.

Levinson, David, and Hao Wu. 'Towards a General Theory of Access'. *Journal of Transport and Land Use* 13 (June 2020): 129–58. https://doi.org/10.5198/jtlu.2020.1660.

Lucas, Karen. 'Transport and Social Exclusion: Where Are We Now?' *Transport Policy*, URBAN TRANSPORT INITIATIVES, vol. 20 (March 2012): 105–13. https://doi.org/10.1016/j.tranpol.2012.01.013.

Lucas, Karen, Bert van Wee, and Kees Maat. 'A Method to Evaluate Equitable Accessibility: Combining Ethical Theories and Accessibility-Based Approaches'. *Transportation* 43, no. 3 (2016): 473–90. https://doi.org/10.1007/s11116-015-9585-2.

Martens, Karel, Jeroen Bastiaanssen, and Karen Lucas. '2 - Measuring Transport Equity: Key Components, Framings and Metrics'. In *Measuring Transport Equity*, edited by Karen Lucas, Karel Martens, Floridea Di Ciommo, and Ariane Dupont-Kieffer. Elsevier, 2019. https://doi.org/10.1016/B978-0-12-814818-1.00002-0.

Mavoa, Suzanne, Karen Witten, Tim McCreanor, and David O'Sullivan. 'GIS Based Destination Accessibility via Public Transit and Walking in Auckland, New Zealand'. *Journal of Transport Geography*, Special Section On Child & Youth Mobility, vol. 20, no. 1 (2012): 15–22. https://doi.org/10.1016/j.jtrangeo.2011.10.001.

Pereira, Rafael H. M., Tim Schwanen, and David Banister. 'Distributive Justice and Equity in Transportation'. *Transport Reviews* 37, no. 2 (2017): 170–91. https://doi.org/10.1080/01441647.2016.1257660.

Verduzco Torres, J. Rafael, and David Philip McArthur. 'Public Transport Accessibility Indicators to Urban and Regional Services in Great Britain'. *Scientific Data* 11, no. 1 (2024): 53. https://doi.org/10.1038/s41597-023-02890-w.

Wacquant, Loïc. 'Territorial Stigmatization in the Age of Advanced Marginality'. *Thesis Eleven* 91, no. 1 (2007): 66–77. https://doi.org/10.1177/0725513607082003.

‘Dossier Complet − Département de La Seine-Saint-Denis (93) | Insee’. Accessed 25 April 2025. https://www.insee.fr/fr/statistiques/2011101?geo=DEP-93.

Shen, Qing. ‘Location Characteristics of Inner-City Neighborhoods and Employment Accessibility of Low-Wage Workers’. *Environment and Planning B: Planning and Design* 25 (February 1998): 345–65. https://doi.org/10.1068/b250345.

Hu, Lingqian, Yingling Fan, and Tieshan Sun. ‘Spatial or Socioeconomic Inequality? Job Accessibility Changes for Low- and High-Education Population in Beijing, China’. *Cities* 66 (June 2017): 23–33. https://doi.org/10.1016/j.cities.2017.03.003.

Giannotti, Mariana, Diego B. Tomasiello, and Taina A. Bittencourt. ‘The Bias in Estimating Accessibility Inequalities Using Gravity-Based Metrics’. *Journal of Transport Geography* 101 (May 2022): 103337. https://doi.org/10.1016/j.jtrangeo.2022.103337.

Adorno, Bruno Vargas, Rafael H. M. Pereira, and Silvana Amaral. ‘Combining Spatial Clustering and Spatial Regression Models to Understand Distributional Inequities in Access to Urban Green Spaces’. *Landscape and Urban Planning* 256 (April 2025): 105297. https://doi.org/10.1016/j.landurbplan.2025.105297.

Wee, Bert van, and Tom de Jong. ‘Differences in Levels of Accessibility: The Importance of Spatial Scale When Measuring Distributions of the Accessibility of Health and Emergency Services’. *Journal of Transport Geography* 106 (January 2023): 103511. https://doi.org/10.1016/j.jtrangeo.2022.103511.

Aberle, Christoph, and Carsten Gertz. ‘Does a Short Journey Get Me to the Food Bank? An Empirical Study on Fare-Based Public Transport Accessibility and Its Implications for Social Equity’. *Journal of Transport Geography* 128 (October 2025): 104348. https://doi.org/10.1016/j.jtrangeo.2025.104348.

Radzimski, Adam. ‘Accessibility of Social Housing by Sustainable Transport Modes: A Study in Poznań, Poland’. *Journal of Transport Geography* 111 (July 2023): 103648. https://doi.org/10.1016/j.jtrangeo.2023.103648.

‘Définition - IRIS | Insee’. Accessed 6 November 2025. https://www.insee.fr/fr/metadonnees/definition/c1523.

OMNIL. ‘EGT H2020’. Accessed 6 November 2025. https://www.omnil.fr/egt-h2020.

‘La PCS 2020 | Insee’. Accessed 6 November 2025. https://www.insee.fr/fr/information/6205305.

‘Taux de Pauvreté Selon Le Seuil de Pauvreté | Insee’. Accessed 6 November 2025. https://www.insee.fr/fr/statistiques/2408282#tableau-figure1.

‘Niveau de Vie Des Personnes et Pauvreté − Les Revenus et Le Patrimoine Des Ménages | Insee’. Accessed 6 November 2025. https://www.insee.fr/fr/statistiques/7941375?sommaire=7941491#tableau-figure1.

‘What Is GTFS? - General Transit Feed Specification’. Accessed 10 February 2026. https://gtfs.org/getting-started/what-is-GTFS/.

References related to r5py

Conway, Matthew Wigginton, Andrew Byrd, and Michael van Eggermond. 'Accounting for Uncertainty and Variation in Accessibility Metrics for Public Transport Sketch Planning'. *Journal of Transport and Land Use* 11, no. 1 (2018). https://doi.org/10.5198/jtlu.2018.1074.

Conway, Matthew Wigginton, Andrew Byrd, and Marco van der Linden. 'Evidence-Based Transit and Land Use Sketch Planning Using Interactive Accessibility Methods on Combined Schedule and Headway-Based Networks'. *Transportation Research Record* 2653, no. 1 (2017): 45–53. https://doi.org/10.3141/2653-06.

Conway, Matthew Wigginton, and Anson F. Stewart. 'Getting Charlie off the MTA: A Multiobjective Optimization Method to Account for Cost Constraints in Public Transit Accessibility Metrics'. *International Journal of Geographical Information Science* 33, no. 9 (2019): 1759–87. https://doi.org/10.1080/13658816.2019.1605075.

Fink, Christoph, Willem Klumpenhouwer, Marcus Saraiva, Rafael Pereira, and Henrikki Tenkanen. *R5py: Rapid Realistic Routing with R5 in Python*. Zenodo, released 9 September 2022. https://doi.org/10.5281/zenodo.7060438.

Appendix A: Scientific article

Testing the relationship between accessibility to jobs and socio-economic inequalities in the Paris Urban Area

Yvain REDON
Transport, Infrastructure and Logistics
Delft University of Technology

Abstract: Accessibility has long been used in transportation studies as a way to compare the utility derived by individuals from the transport system. Furthermore, accessibility is also a helpful indicator in studying distributive justice issues in the field of transportation. On the other hand, a vast literature has studied inequalities, in terms of income and employment among other variables, between 'cités' neighbourhoods, in France as a whole and in the Paris Urban Area more specifically. In this study we aim to identify how these inequalities could be associated with poor accessibility to jobs in these areas. GTFS data from the Paris Urban area is used to estimate time travels between census tracts. Using this data as input, accessibility to jobs from an area is then computed using both travel time and fare cost. These accessibility results are then analysed, using a clustering method and a spatial regression model. Results obtained show that in general areas with lower income tend to perform more poorly accessibility-wise, but that a variety of situations exist for areas with average income.

1. Introduction

In many French cities, inequality in terms of wealth and unemployment, education, and access to health and culture is highly geographically clustered. Indeed, due to several factors, namely the destructions caused by the second world war, fast demographic growth and the return of French citizens to metropolitan France from then newly independent countries, led to government programs aiming at building a lot of housing units in a quite short time. Though these cities and districts built in the 30 years after 1945 were supposed to be planned with access to public services in mind, this was not always the case in reality, as explained by Lévêque (2023) in the case of the city of Vaulx-en-Velin.

Over the years, these urban spaces, often dubbed '*banlieues*', '*cités*' or '*grands ensembles*', have gained a reputation for poor living conditions. Indeed, taking as an example the *département* of Seine Saint-Denis, north of Paris, can help in representing the characteristics of these *banlieues*. Using census data from INSEE, it is noticeable that this territory had a poverty rate of 28,4 % in 2021, almost twice as high as the general poverty rate in metropolitan France. Furthermore, the unemployment rate is also higher, at 16,4%. In addition, the fact that cities in this area developed over the last century is shown by the fact that only 3.5% of housing units were built before 1919. These inequalities may be correlated with inequalities in accessibility to key services, especially jobs, compared to other parts of an urban area (which can be the centre of the urban area, or other cities in the urban area). Thus, it can be interesting to study exactly how accessibility (to a selection of key services) is distributed spatially in a French urban area, and if there is a significant correlation between the income and poverty of an area, and its access to jobs.

In this context, in order to have a better grasp on the spatial distribution of inequalities in the context of French urban areas, this study aims to evaluate accessibility to jobs and see if it is associated with income and other socio-economic indicators. To conduct this study, the city of Paris and its urban area was chosen.

This research aims to contribute to knowledge that can counter inequalities in accessibility to work locations. Results about accessibility will be useful in determining if the transport network contributes to the reduction of inequalities or if on the other hand it reinforces these inequalities. Indeed, this question is important, as depending on the context of a city, the literature has shown differing results. Indeed, in some cases low-income households benefit from higher accessibility levels than high-income households, whereas in other cases accessibility and income are positively correlated. Furthermore, another aim is to see if the spatial repartition of inequalities and transport access in Paris is similar to other cities worldwide. Thus this article aims to test the assumption of a link between accessibility to jobs and other socio-economic variables including income, in the Paris Urban Area.

2. Literature review

In order to assess inequalities in the transport system, accessibility has often been used. Accessibility is generally agreed to be linked to both the transport infrastructure and the land-use system. However, several different definitions have been proposed over the year. One of the first proposed definitions, cited by multiple articles considered here, is the definition of Hansen (1959), where accessibility is 'the potential of opportunities for interactions'. Since then, other definitions have been proposed, for example in Geurs and Van Wee (2004) : 'accessibility measures are seen as indicators for the impact of land-use and transport developments and policy plans on the functioning of society in general'. In Levinson and Wu (2020), accessibility is described more simply as 'the ease of reaching valued destinations. Thus, accessibility is widely used as an indicator that can describe interactions between the transport infrastructure on one hand and the land-use distribution on the other hand.

Among the many accessibility indicators that exist, two have been especially widely used: 'contour' measures (or cumulative accessibility measures) and 'potential' accessibility measures (or 'gravity based' measures). While contour measures only consider destinations below a certain threshold of distance or travel time, potential accessibility measures can consider all destinations in a given area.

Potential accessibility measures have been used in multiple studies such as by Foth et al. (2013), Bouzouina et al. (2021), Bocarejo S. and Oviedo H. (2012). This indicator is useful in showing differences in access cost that may have been hidden by simply using a threshold. However, it is less easily interpretable and still has other disadvantages of location-based accessibility measures. Indeed, it does not take into account the fact that some opportunities may be unavailable for some individuals, depending on their characteristics. Furthermore, Giannotti et al. (2022) has pointed out that, depending on the process chosen for the calibration of the parameters of the cost function, inequalities of access can be underestimated, and results can be inaccurate, especially when using parameters specific to a some groups.

One of the key weaknesses of potential accessibility measures is that competition effects are not well represented. This can lead to inaccuracies, as pointed out in Geurs and Van Wee (2004), especially when studying access to jobs. Indeed, one individual may have access to a lot of job opportunities in terms of travel time, but if these opportunities are in high demand, then the real access to jobs is lower than would be expected without competition effects. To take these effects into account, Shen (1998) introduced a modified potential accessibility measure which takes into account the interaction of supply and demand, this measure being referred as Shen's accessibility in the literature. Hu et al. (2017) for example have used Shen's accessibility to measure the evolution of accessibility to jobs in Beijing. However, it is still susceptible to the problems highlighted in Giannotti et al. (2022).

Various methods have been applied to help interpreting the results of accessibility analyses. While some authors such as Foth et al. (2013) and Ermagun and Tilahun (2020) have used linear methods, others such as Lucas et al. (2016) and Adorno et al. (2025) have used other methods that can provide meaningful insights when discussing accessibility, respectively a combination of the Lorenz curve and Gini index (also used by , and a combination of spatial clustering and spatial regression.

3. Methodology

3.1. Scope of the study

In this study, accessibility is considered for jobs and work locations. It is considered at the aggregate level but also at a more detailed level by separating both the working age population and the job positions by socio-professional category (SPC).

No public transport mode was excluded in the study. This means that all available public transport modes available in the urban area were considered. This includes not only the transport networks showcased in the descriptive analysis such as subway, tramway and regional trains, but also other modes, most notably bus service lines.

In addition to public transport modes, walking and cycling are considered to access public transport stops. In some cases, where the trip is very short, an itinerary may only have cycling or walking components.

There are multiple administrative and urban divisions that exist for Paris and its neighbouring cities: the 'Ile-de-France' region, the metropolitan area, the urban area and other divisions. This study will focus on the Paris urban area, which is the area where there is a built environment continuity from Paris.

The spatial scale selected for the study are the 'iris' (INSEE). These zones are designed to provide detailed data below the city level, for cities which exceed at least 5000 inhabitants. The advantage of using these zones over grid data, is that these zones are designed to be somewhat homogeneous in terms of population and type of built environment (housing, activities or other). In this study, the iris geographic limits from 2022 were used.

Figure 1 presents a map of both the boundaries of iris zones and the outline of the study area.

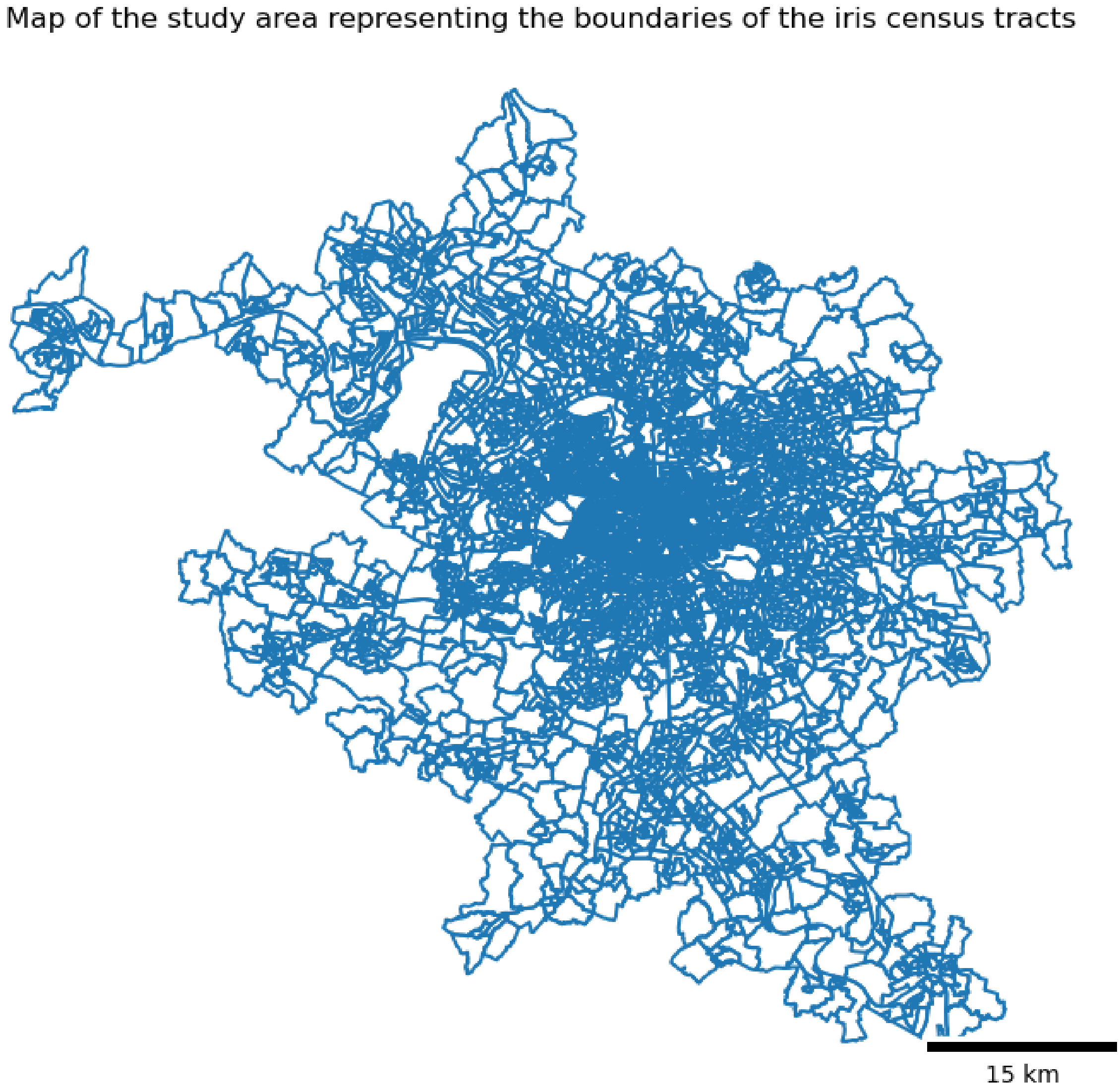

Figure 1: Map of iris census tracts boundaries in the study area

3.2. Data

Socio-economic and demographic data about the population of the iris zones come from two main databases. The first one is the Filosofi data, which mostly provides data about the income structure and poverty rate of the iris areas. This database includes both the income before and after taxes and redistribution. In our study we use the available income as a reference to better represent the influence of fares on the budget of individuals and thus we use income after redistribution. The second database is the census data at the iris level. This data is quite extensive and contains the type of housing, the size of households, the socio-professional categories of individuals, among other indicators.

Data about jobs is extracted from the siret establishment list, which contains information about all companies in France. This include the number of employees in the company, though the exact number of employees is unknown, instead a code is given which corresponds to a certain range of number of employees (for example, the code '01' corresponds to a number of employees between 1 and 2). The location of the companies is also included in the dataset. This dataset is also used to estimate access to general practitioners, as they correspond to a certain type of company registered in the dataset.

The GTFS data is obtained from the 'Ile-de-France mobilités' open data (IDFM), IDFM being the transit agency organising public transports in the Ile-de-France region. Since the Paris urban area is entirely contained within this region, there is no need for additional gtfs data.

3.3. Computation of the travel-time matrix

The travel time matrix is calculated based on the GTFS data available for the urban area. GTFS data are widely used in transportation and accessibility studies. GTFS data is convenient because it contains in one format much of the data necessary to analyse transport systems: it contains the location of service lines, public transport stations, as well as timetables and frequency. It also contains information about the type of public transport line (bus, train, tramway, subway, etc). The calculation of travel time is done on python, using the r5py library.

The origins and destinations considered in the travel time matrix are the centroids of the iris areas. When calculating travel time from an iris zone to an activity, the centroid of the origin zone is used as the start location of the trip and the centroid of the destination iris is used as the location of the activity. In case iris centroids happen to be located at a point unreachable by the network, the option 'snap-to-network' is activated allowing for the origin to be relocated to the nearest point connected to the network.

All public transport modes are considered when calculating the travel time matrix, as well as walk and cycling. Modes can be combined along a given trip to account for the multimodal nature of the network. The calculation is done for the 25th of September 2025, which is a Thursday, outside of school holidays. The departure time is set to 8AM, with the departure time window being set to 30 minutes to avoid inaccuracies in the calculation, since some of the modes involved such as buses can have a low frequency and thus could be neglected if using a smaller departure time window.

3.4. Defining the accessibility metric

Accessibility is a measure well-suited to this research for several reasons. As mentioned by Geurs and Van Wee (2004), accessibility allows to consider the interactions between the land-use system and the transport system. It is important in the case of this study as some places may be deprived of access to transport to certain places but benefit from local facilities that are sufficient. On the other hand, some areas may be deprived of both access to transport and be located far from key facilities, whereas some areas may benefit from good access to both.

Furthermore, accessibility as a measure has been widely used to study inequalities (such as in El-Geneidy et al. (2016) or Bocarejo S. and Oviedo H. (2012) for example), which makes comparison between case studies easier, though one needs to be aware of the type of accessibility metric used. Moreover, from an ethical perspective, Pereira et al. (2017) claim that accessibility 'stands out as the most promising focal variable of distributive justice'. Thus, accessibility was chosen in order to properly address the subject of inequalities in this study.

Shen's accessibility is used (Shen, 1998). This accessibility measure is chosen for multiple reasons. First, since it is based on potential accessibility measures, it supersedes the need for an arbitrary travel-time or travel-cost threshold which may impact the results. Moreover, shen's accessibility takes into account the competition effects, which is good for all types of activities but especially for jobs.

To take into account the cost of fares in accessibility, we use generalized time travel as the cost function for accessibility. Generalized time travel is defined as follows in the context of this thesis:

$$C_{ij} = t_{ij} + \frac{F_{ij}}{w_i}$$

Where C_{ij} is the cost of travel from the zone i to zone j, t_{ij} is the travel time from i to j, F_{ij} is the fare to travel from i to j, and w_i is the median wage in zone i. The form of this cost formula is similar to the one from El-Geneidy et al. (2016).

Another cost formulas is considered, by using only travel time as the impedance,

Thus, the general formula of accessibility used in this study is (with small changes from Shen, 1998:

$$A_i = \sum_j \frac{O_j e^{-\beta c_{ij}}}{D_j}$$

$$D_j = \sum_k P_k e^{-\beta c_{ij}}$$

Where:

- A_i is the accessibility from zone i to a certain type of activities
- O_j is the number of opportunities for the activity in zone j
- β is the exponential decay curve parameter for the activity
- C_{ij} is the cost of traveling from i to j
- D_j is the demand potential for zone j
- P_k is the number of people in k seeking for the opportunities to perform the given activity

The cost function chosen is a negative exponential, as seen above. This is because, according to Geurs and van Wee (2004), 'the negative exponential function is the most often used and also the most closely tied to travel behaviour theory'. This allows for easier comparison across articles.

The parameter of the negative exponential function β is calibrated based on the most recent travel survey in the Ile-de-France region (OMNIL).

3.5. Competition effects

The number of opportunities (that is to say, the number of jobs in an area) is not known at the iris level. To accommodate for this fact, we estimate the weight of all iris in a given city by summing the number of employments for each company in the iris. Then, this number is divided by the total estimated number of jobs in the city, obtained by making the sum of jobs over all iris in the city. By dividing the estimated number of jobs in an iris and the estimated number of jobs in the city, the relative weight of the iris in the employment of the city is obtained. Then, we multiply the total number of jobs at the city level obtained from the 2021 census by this weight, to obtain an estimation of the number of jobs in a given iris.

$$J_i = \frac{J_{i,estimated}}{J_{c,estimated}} \times J_c$$

Where:

- J_i is the number of jobs in iris i which belongs to city c
- J_c is the number of jobs in city c
- $J_{estimated}$ is the estimated number of jobs in a zone based on employment in companies that belong to this zone in the siret registry

We consider on one hand the total number of jobs as opportunities and the total working age population as the demand, and on the other hand the jobs and working age population are divided by socio-professional category (INSEE). The general composition of these socio-professional categories is shown in Table 3.

Number of the category	**Fields represented by the category**
1	Agriculture
2	Business owners
3	Executives, intellectual occupations
4	Intermediary occupations
5	Employees
6	Workers

Table 1: Composition of socio-professional categories

3.6. Clustering

In order to analyse the association of accessibility with other variables, a clustering method was used in this study.

The clustering method chosen for this thesis was the k-means clustering methods. Though there are more suited methods for spatial data such as spatial clustering (used for example in Adorno et al. (2025)), they were not used in this study, both for technical reasons and time constraints.

In order to determine the number of clusters for each clustering model, the elbow method was used.

In addition to accessibility data, multiple socio-economic indicators were used to form the clusters. First, the yearly median available income in the area, 'available income' meaning that this data takes into account the loss of income from taxes, and the gain of income through social subsidies and other redistribution. Second, the ratio of social housing over all housing in the area. Though this indicator may be correlated with income, it could help identify specific neighbourhoods, especially the '*cités*' mentioned previously. Indeed, this neighbourhoods are often characterised by a high proportion of social housing, as mentioned in Castel (2006). Furthermore, the ratio of households that own at least one car over the total of households in an area is also considered. Indeed, households that own a car may be less reliant on public transportation, and may travel by car by choice, though of course some households are also reliant on car travel by necessity. Finally, the population density of the area was considered. This is done in order to separate the central parts of the urban area, which are more densely populated, and the areas further to the periphery of the urban area. At first, the poverty rate of the area was also considered in the clustering process, but due to the large amount of missing or unavailable data for this variable it was not used.

3.7. Spatial regression

It was decided to use a spatial regression model in addition to clustering to analyse the data. Indeed, while clustering is helpful in identifying patterns in the data, it cannot provide precise information about the statistical link between specific variables and accessibility. Thus, a regression model was also considered necessary to this study.

Furthermore, spatial regression is useful for testing the assumptions that were made about the data, as it provides a similar framework to a linear regression model. This proximity in terms of structure is also helpful in order to compare results, for example if one wants to see the differences between a standard linear regression model and a spatial regression model. A spatial regression model was chosen, as it was considered more fitted to the type of data studied. Indeed, due to the data studied, it seems likely that spatial factors could influence the results. More specifically, the spatial lag model was used.

The spatial weights scheme chosen is the k-nearest neighbours, with k's value being set to 10.

Spatial regression was performed on accessibility to jobs and accessibility to primary and secondary education. These facilities were chosen to represent different types of facilities and to identify potential differences in the models. All accessibility data obtained was not used for spatial regression. Indeed, because good accessibility to one facility often translates to good access to all types of facilities, it was not deemed necessary to include all measures.

4. Results and analysis

4.1. Clustering

		Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Mean	Population density	8820,08	3938,05	11158,57	19684,49	26756,57	41582,17
	Accessibility to jobs	1,00	0,68	0,79	0,73	1,28	1,12
	Ratio of social housing	0,09	0,11	0,33	0,74	0,04	0,19
	Median available income	33546,62	26598,60	20712,76	17312,28	41178,60	28400,25
	Ratio of households that own at least one car	0,75	0,87	0,65	0,55	0,44	0,31
Std error	Population density	5382,62	2989,16	6592,11	11496,81	9330,64	12249,02
	Accessibility to jobs	0,09	0,14	0,13	0,14	0,08	0,11
	Ratio of social housing	0,09	0,10	0,14	0,15	0,05	0,16
	Median available income	4915,05	3978,03	3112,28	2205,03	6138,08	4831,24
	Ratio of households that own at least one car	0,11	0,06	0,13	0,13	0,14	0,11
Size		687	887	833	530	415	566

Table 2: Structure of the clusters created from accessibility to jobs

Two sets of clusters were computed: one with accessibility to jobs without separating the jobs and population into socio-professional categories, and another set of clusters where accessibility to jobs is considered separately depending on socio-professional categories.

The general accessibility to jobs is considered first. For this accessibility measures, the socio-professional category to which a person or a job belongs is not considered. Table 5 describes the mean values and standard errors of variables for each cluster. Meanwhile, figure 10 shows the spatial shape of the clusters in the urban area.

Cluster 0 is characterized by a decent accessibility, having the third highest accessibility to jobs. It also has the second lowest percentage of social housing. It has the second highest income but has the second lowest population density among the studied clusters. Furthermore, it has a high percentage of households owning a car. This cluster seems to be characterized by above average income (high middle class or higher) but also represents less dense areas of the urban area, with a high reliance on car, though accessibility by public transport is still above the average of all clusters. It seems to correspond to affluent suburbs.

Cluster 1 has the lowest population density, quite average income (since the average in the urban area is approximately 25898 € per year), and a high reliance on car, since accessibility by public transport is low and car ownership is high. Taking into account the spatial distribution of areas in this cluster, it is quite clear that these areas correspond to the areas far from the centre and from major public transport lines.

Cluster 2 has below average income and above average car ownership. It also has the second highest percentage of social housing at 33%. Accessibility to jobs is poor. This cluster seems to represent low-income suburbs, as population density is quite high but nowhere near that of the city centre.

Cluster 3 seems typical of "grands ensembles" and '*cités*', that is to say districts built in the second half of the 20th century that are mostly made up of social housing apartments as mentioned previously. Furthermore, accessibility is the second lowest among all clusters, income is the lowest, and social housing accounts for 77% of housing. What is also noticeable is the high population density. Looking at the map, some of these areas are very close and other very far from the city centre. Most of these areas are located to the east, and to the north-east specifically.

Cluster 4 represents highly central areas that are very affluent. It also has the highest accessibility among all clusters. It seems to correspond to the rich neighbourhoods of the centre of the urban area. Social housing is almost non-existent.

Finally, cluster 5 has a similar accessibility to cluster 4 but is characterised by a lower income, though the mean income in this cluster is still above the median income in the area. In the case of cluster 4 and 5, since these represent neighbourhoods close to the centre, car ownership is quite low. Cluster 5 also has the highest population density among all clusters and has the third highest percentage of social housing.

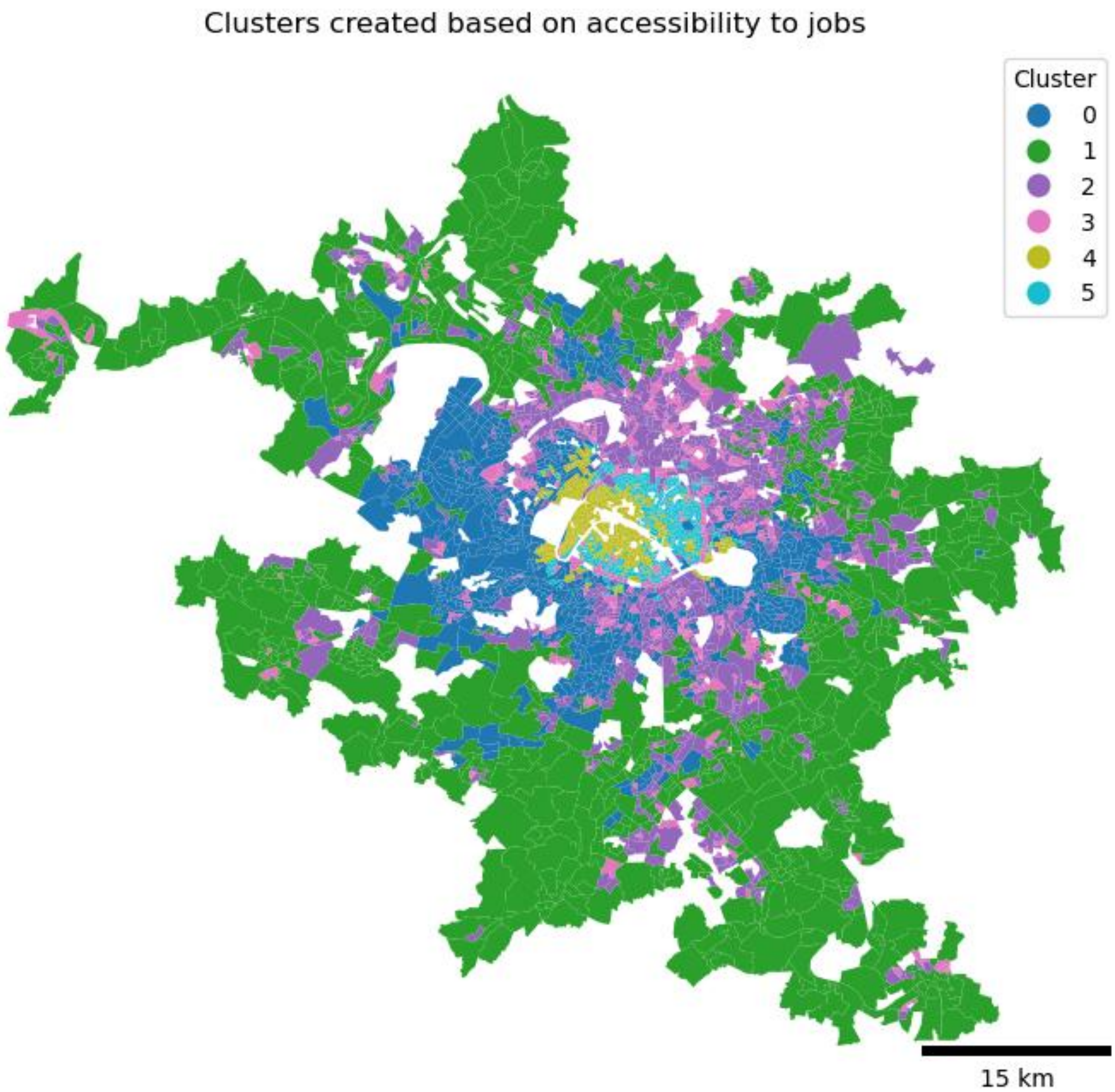

Figure 2: Map of the urban area showing the cluster to which each area belongs

		Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Mean	Population density	31118,36	17041,09	4392,93	39259,78	6648,62	10592,01
	Ratio of social housing	0,05	0,63	0,21	0,29	0,15	0,11
	Median available income	38794,78	17630,72	24182,94	25596,62	25588,10	33330,80
	Ratio of households that own at least one car	0,38	0,58	0,85	0,35	0,79	0,71
	Accessibility to CS1 jobs	1,93	1,07	0,79	1,56	1,17	1,50
	Accessibility to CS2 jobs	1,45	0,84	0,63	1,20	0,92	1,16
	Accessibility to CS3 jobs	1,37	0,79	0,61	1,13	0,87	1,11
	Accessibility to CS4 jobs	1,51	0,87	0,65	1,25	0,95	1,21

	Accessibility to CS5 jobs	1,46	0,85	0,62	1,21	0,92	1,17
	Accessibility to CS6 jobs	1,52	0,87	0,63	1,25	0,95	1,21
Std error	Population density	11420,79	10335,88	4537,80	14549,66	4797,57	6270,12
	Ratio of social housing	0,07	0,20	0,19	0,20	0,12	0,12
	Median available income	6546,71	2201,94	5265,48	4192,46	4674,74	5452,85
	Ratio of households that own at least one car	0,15	0,13	0,10	0,11	0,12	0,12
	Accessibility to CS1 jobs	0,13	0,16	0,16	0,14	0,13	0,14
	Accessibility to CS2 jobs	0,08	0,11	0,12	0,09	0,08	0,08
	Accessibility to CS3 jobs	0,08	0,10	0,12	0,09	0,07	0,08
	Accessibility to CS4 jobs	0,08	0,12	0,13	0,10	0,08	0,08
	Accessibility to CS5 jobs	0,08	0,11	0,13	0,09	0,08	0,08
	Accessibility to CS6 jobs	0,09	0,12	0,14	0,10	0,09	0,09
Size		567	727	520	462	996	646

Table 3: Summaries of the clusters created using accessibility to jobs per category

Then, the clusters created based on accessibility to jobs per category are considered. In this case, access to employment depends on both the category of the position and the socio-professional category of the person for competition effects. Table 6 details the characteristics of the clusters, and figure 11 shows their location in the urban area. Using the elbow method, the analysis was performed with six clusters.

Cluster 0 seems to correspond to wealthy, dense urban areas. Indeed, it has the highest median income across all clusters in this analysis. The map further supports this analysis, since areas belonging to cluster 0 are located in the west of Paris proper, and in the western inner urban areas, which are both dense and quite wealthy areas. This cluster also has the highest accessibility to jobs of all categories, though since the income of the cluster is high, population inside the areas of the cluster probably belong mostly to CS2 and CS3 (business owners and executives respectively). As could be expected, the ratio of social housing is quite low. Overall, this cluster is quite similar to cluster 4 from the previous analysis.

Cluster 1 has a quite high population density, and most importantly the lowest income across all clusters as well as the highest ratio of social housing. It seems to correspond to 'grands ensembles' or 'cités', similarly, to cluster 3 in the previous analysis. This interpretation can be supported by the fact that areas belonging to this cluster are quite spread on the map. It has the second lowest accessibility to jobs of all categories.

Cluster 2 has the lowest population density of all clusters, as well as the lowest accessibility to jobs overall. Thus, reliance on car would be necessary, which is shown by the fact that approximately 85% of households own at least one car. Median income is a bit below the median of the studied region which is approximately 25898 €. The ratio of social housing is rather average. This cluster corresponds to areas further from the centre that are less urbanized of the urban area, as seen on the map.

Cluster 3's mean income is quite similar to that of Cluster 2: it is almost the same of the median income of the entire study area. However, this cluster is very close to the centre (most areas are located directly in Paris proper as seen on figure **X**). It is the cluster with the highest population density. Due to its centrality, it has the second highest accessibility to jobs of all categories. The ratio of social housing is quite high at 29%. As with cluster 1, most households don't own a car.

Cluster 4's structure is quite similar to that of cluster 2, with mean income and population density being a bit higher, and the ratio of social housing being a bit lower. Accessibility however is higher than Cluster 2's, though it is still the third lowest accessibility among all clusters. Conversely, car ownership in this cluster is the second highest among all clusters. This cluster seems to correspond to areas that are further from the centre, but a bit more connected than areas in cluster 2.

Finally, Cluster 5 represents wealthy areas, but these areas are further from the centre than areas in Cluster 0. The most notable differences with areas in Cluster 0 is a higher ownership of cars, a lower population density and lower accessibility, though accessibility is still the third highest among all clusters. This cluster seems to represent wealthy suburbs.

Overall, accessibility to all categories of jobs seem to be correlated: clusters often have good or bad accessibility to all job categories at the same time. Overall, areas located near the centre have better accessibility, as expected. Low-income areas have poor accessibility, whereas for areas with average income, there are many diverse situations: some have quite good accessibility whereas others have some of the lowest accessibility. This seems to confirm that the relationship between income and accessibility is quite complex.

Comparing this set of clusters to the previous one where socio-professional categories were ignored, there are some differences, most notably for areas with average income. Indeed, for central areas, and affluent or poor neighbourhoods, the clusters' structures are mostly consistent. However, for average income areas, distance to the centre seems to become a more important factor when job category is considered. Indeed, when considering general accessibility to jobs a lot of areas are grouped in cluster 1, whereas when taking into account job category, these areas are split between clusters 2 and 4, the latter also encompassing areas belonging to cluster 2 in the analysis of general access to jobs.

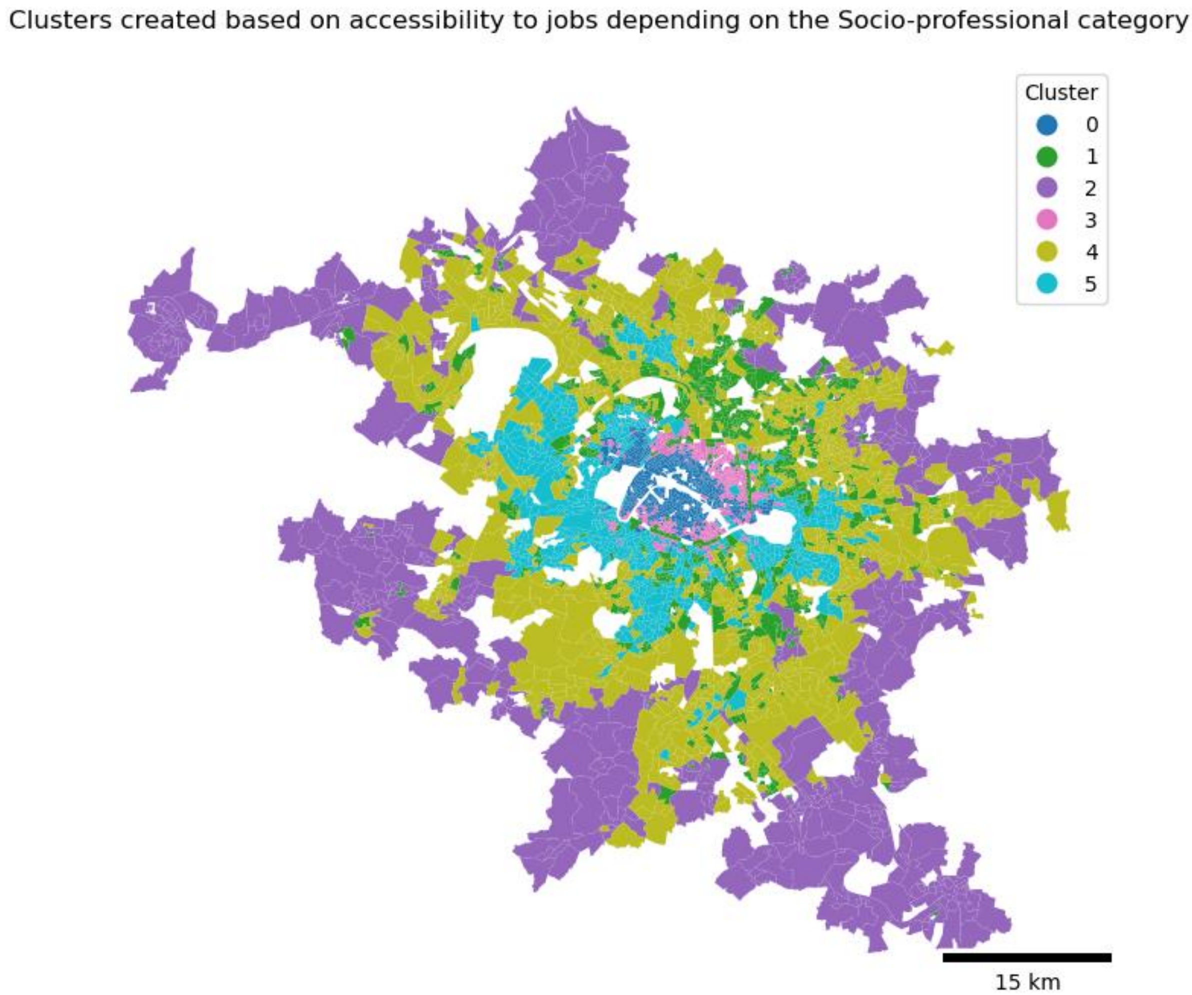

Figure 3: Spatial distribution of clusters made by accessibility to jobs per category

4.2. Spatial regression

Variable	**Coefficient**	**Std.Error**	**z-Statistic**	**Probability**
CONSTANT	0,46098	0,01739	26,50175	0,00000
Population density (thousands of inhabitants/km²)	0,00492	0,00012	2,68426	0,00727
Ratio of social housing	-0,01940	0,00163	-11,90161	0,00000
Median available income (in thousands of euros)	0,08377	0,00322	26,04563	0,00000
Ratio of households that own at least one car	-0,06758	0,00313	-21,59593	0,00000
W_accessibility_jobs_gtt	0,04825	0,00193	24,95327	0,00000
SPATIAL LAG MODEL IMPACTS				
Variable	**Direct**	**Indirect**	**Total**	
Population density (thousands of inhabitants/km²)	0,0049	0,0002	0,0052	

Ratio of social housing	-0,0194	-0,0010	-0,0204
Median available income (in thousands of euros)	0,0838	0,0042	0,0880
Ratio of households that own at least one car	-0,0676	-0,0034	-0,0710
Pseudo R-squared	0,9112		

Table 4: Spatial regression model with accessibility to jobs as the dependent variable

	Population density (thousands of inhabitants/km²)	**Ratio of social housing**	**Median available income (in thousands of euros)**	**Ratio of households that own at least one car**
Population density (thousands of inhabitants/km²)	1,00	0,08	0,09	-0,77
Ratio of social housing	0,08	1,00	-0,66	-0,14
Median available income (in thousands of euros)	0,09	-0,66	1,00	-0,01
Ratio of households that own at least one car	-0,77	-0,14	-0,01	1,00

Table 5: Correlation matrix

Most of the variations are captured by the model, with the pseudo R-squared value being above 0.9.

The median available income contributes positively to accessibility. This is expected as the clustering analysis has shown that areas with high income often had better accessibility to jobs than other areas with otherwise similar characteristics. Furthermore, a higher proportion of social housing negatively impacts accessibility, which could support the idea that neighbourhoods with a high proportion of social housing (which are often 'cités') suffer from a specific lack of access. Finally, a high proportion of households that have access to a car is also associated with less access. Regarding population density, it is associated with higher accessibility, which is not surprising since more densely populated areas tend to be located at the centre of the urban area, where more service lines and stations are located.

Indirect effects associated with variables tend to be low.

Overall, the insight from this model is coherent with those obtained from clustering: income is an important factor in the level of accessibility of an area, but other factors contribute as well.

The correlation matrix shows strong correlation between income and the ratio of social housing, and population density and car ownership. Overall, these relationships are expected: social housing primarily serves people with below average income, and people living in more dense areas are less dependent on car transportation due to being closer to daily life facilities.

5. Discussion and conclusion

5.1. Data used

All data used in this study is directly available on different public data websites for France. However, restricted access data was not used. Future studies could improve the approach of this paper, for example by benefitting from a better calibration of model parameters by using detailed trips data from the travel survey.

Furthermore, the lack of data for employment at the 'iris' scale means that jobs data in our study were estimated using rough employment figures from company registries. Though these estimations were not too far off the real figures (when comparing data at the city level), the estimation method used in this study may lead to inaccuracies. For example, it may lead to underestimating the importance of self-employment, as self-employed individuals are not always considered in companies registries for confidentiality purposes.

5.2. Travel-time calculations

In this study, one travel-time matrix was computed using r5py, using a weekday morning as the departure time, which means that this study doesn't take into account atypical working hours, such as night shifts, or part-time work. This, in particular, could lead to a less accurate representation of accessibility to low-income jobs, which are more often part-time jobs.

Another limitation of the travel-time computation method is that, while it was possible to take into account timetables and frequency of service lines using GTFS data, this study did not take into account the reliability of public transport. Indeed, recurring delays or cancelation on certain service lines can dramatically decrease the accessibility of an area, even if it theoretically has good access. Furthermore, since some transport service lines may be less reliable than others, it could create potential inequalities in addition to accessibility inequalities that were already observed.

In this study, the data used to make the travel-time matrix was only the travel-time from the origin to the destination. Due to computational power limitations, it wasn't possible to use the full, detailed breakdown of the trips from one zone to another. This means that while the travel time is accurate, this study cannot account for transfers or check precisely which public transport modes are used when making the travel time calculations. Future studies making use of detailed travel information may be able to conduct more detailed analysis about reliability of service lines involved (as mentioned above), the number of transfers necessary to reach a destination and the number of transport legs.

5.3. Results

The results of this study seem to indicate that while income is positively correlated with accessibility to different type of facilities, it is not the only variable that has to be taken into account. Specifically, for low- and high-income ranges, accessibility tends to be very low and very high respectively, but for average income ranges there are more diverse situations. The ratio of social housing and the ratio of car-owning households also are important variables correlated with low access.

When focusing on districts which correspond to the characteristics of 'cités', these districts seem to face specific challenges, even compared to other low-income suburbs. 'Cités' have lower income, a much higher proportion of social housing, and low car ownership (between 50% and 60%) when taking into account the lack of access through the public transport system. Thus, in addition to facing poor accessibility to services via public transport, inhabitants of these neighbourhoods may not be able to always rely on car travel to make up for this lack of access through the public transport network.

Using Shen's accessibility (Shen, 1998), this study properly considers competition effects when calculating accessibility to services. However, as pointed out in Giannotti et al. (2022), results obtained using Shen's accessibility can lead to underestimating inequalities of access, since the calibration of the parameters of this measure assumes that travel surveys reveal unconstrained preferences. Thus, whereas the results of our study are interesting in that they reveal inequalities and consider competition effects, one must be aware that some inequalities of access could be underestimated. Other accessibility studies in the Paris urban area could benefit from using different accessibility measures to verify that results from this study are accurate.

The findings of this study hint at a positive correlation of income and population density with accessibility, and a negative correlation of the ratio of social housing and the ratio of car ownership with accessibility. However, the facts that these variables have statistical links does not necessarily indicate a causality link, or the direction of this causality. In particular, this study cannot indicate if this association exists because bad access to jobs leads to a degradation of the socio-economic condition of individuals, or if existing deficits in income and other variables lead to an area having a worse access to employment positions.

Comparing the results of this study with other articles from the literature, it appears that the accessibility in the Paris urban area is less equal than some other cities in the literature. For example, in Foth et al. (2013), it was found that in Toronto the public transport system tends to benefit disadvantaged areas, which is not the case in the case of the Paris urban area. In our case, high-income areas often benefit more from the land-use and transport systems. Thus, currently urban planning and transportation planning in the Paris urban area fails in correcting existing economic inequalities.

Hopefully, this study can prompt more in-depths research focusing on specific types of services and provide useful insight for decision makers and city planners in order to address inequalities in the Paris urban area.

Appendix B: Summaries of spatial regression models

B.1. Summary of the spatial regression for accessibility to jobs

```
REGRESSION RESULTS
------------------

SUMMARY OF OUTPUT: SPATIAL TWO STAGE LEAST SQUARES
--------------------------------------------------
Data set            :     unknown
Weights matrix      :     unknown
Dependent Variable  :accessibility_jobs_gtt                Number of Observations:        3918
Mean dependent var  :      0.8941                Number of Variables   :           6
S.D. dependent var  :      0.2370                Degrees of Freedom    :        3912
Pseudo R-squared    :      0.9112
Spatial Pseudo R-squared:  0.8168

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT         0.46098         0.01739        26.50175         0.00000
       pop_density_k         0.00492         0.00183         2.68426         0.00727
      percentage_HLM        -0.01940         0.00163       -11.90161         0.00000
              MED21K         0.08377         0.00322        26.04563         0.00000
      percentage_car        -0.06758         0.00313       -21.59593         0.00000
W_accessibility_jobs_gtt         0.04825         0.00193        24.95327         0.00000
------------------------------------------------------------------------------------
Instrumented: W_accessibility_jobs_gtt
Instruments: W_MED21K, W_percentage_HLM, W_percentage_car, W_pop_density_k

DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST                              DF         VALUE           PROB
Anselin-Kelejian Test              1       239.957           0.0000

SPATIAL LAG MODEL IMPACTS
Impacts computed using the 'simple' method.
            Variable         Direct        Indirect          Total
       pop_density_k         0.0049          0.0002          0.0052
      percentage_HLM        -0.0194         -0.0010         -0.0204
              MED21K         0.0838          0.0042          0.0880
      percentage_car        -0.0676         -0.0034         -0.0710
================================ END OF REPORT =====================================
```

B.2. Summary of the spatial regression for accessibility to primary and secondary education

```
REGRESSION RESULTS
------------------

SUMMARY OF OUTPUT: SPATIAL TWO STAGE LEAST SQUARES
--------------------------------------------------
Data set            :     unknown
Weights matrix      :     unknown
Dependent Variable  :accessibility_high_schools_gtt                Number of Observations:        3918
Mean dependent var  :      1.0872                Number of Variables    :        6
S.D. dependent var  :      0.3144                Degrees of Freedom     :     3912
Pseudo R-squared    :      0.8970
Spatial Pseudo R-squared:  0.7971

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT         0.59591         0.02315        25.74328         0.00000
       pop_density_k         0.00462         0.00261         1.77041         0.07666
      percentage_HLM        -0.02811         0.00231       -12.15824         0.00000
              MED21K         0.11484         0.00454        25.27336         0.00000
      percentage_car        -0.09663         0.00454       -21.28468         0.00000
W_accessibility_high_schools_gtt         0.04497         0.00211        21.27426         0.00000
------------------------------------------------------------------------------------
Instrumented: W_accessibility_high_schools_gtt
Instruments: W_MED21K, W_percentage_HLM, W_percentage_car, W_pop_density_k

DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST                              DF         VALUE           PROB
Anselin-Kelejian Test              1        229.514          0.0000

SPATIAL LAG MODEL IMPACTS
Impacts computed using the 'simple' method.
            Variable         Direct        Indirect          Total
       pop_density_k         0.0046          0.0002         0.0048
      percentage_HLM        -0.0281         -0.0013        -0.0294
              MED21K         0.1148          0.0054         0.1202
      percentage_car        -0.0966         -0.0046        -0.1012
================================ END OF REPORT =====================================
```

Appendix C: The GTFS data format

The GTFS (General Transit Feed Specification) data format is widely used for public transportation studies. Indeed, it is used in most articles used as reference in this study. Most of the time, it is produced by public transport agencies or local authorities, and for many areas this data is open and available online. This allows for better comparisons and reproducibility across studies. Most notably, GTFS data are used by trip planners applications.

The GTFS data for a specific area is most often made available in the form of a compressed (.zip) file. This file contains information about:

- The transport agency which operates in the public transport network of the area
- The public transport stops, including:
    - An id
    - The name of the stop
    - Its position (latitude and longitude)
    - The type of location (which can be: 0 for "platform", 1 for "station", 2 for "entrance/exit", 3 for "generic node" and 4 for "boarding area")
- Routes, including:
    - An id
    - The name of the route
    - A description of the route
- Trips, including:
    - Id of the route on which the trip is performed
    - Id of the service
    - Id of the trip
    - "headsign" (most often the last stop in the direction of the trip)
- Stop times, including:
    - The id of the trip stopping at the location
    - The id of the stop location
    - The arrival and departure times at the location of the trip
- The calendar (type of service), including:
    - The days of the week when the service takes place
- The fares
- The frequencies
- The transfers available between stops

In the case of the public transport network in the Ile-de-France region, the GTFS data file contains information about the calendar, the routes, the stops and stop times, the transfers and the trips.

Thus, combining the data from the GTFS file, it is possible to simulate trips between origin-destination pairs, which take into account transfer times, the frequency and the timetables at the stops (thus taking into account waiting times) and obtain the expected travel-time for a chosen departure time.