C.A.N.L Duynstee
M.J. Haayen
D. Kyritsis
L.M. Ortega-Córdova
S. N. N. Samat

Synthesis Project Dordrecht Final Report

Smart City Dordrecht – Identification of Pedestrian Movement Patterns with Wi-Fi Tracking Sensors

TUDelft

Final Report

GEO-1101 Synthesis Project Dordrecht
Smart City Dordrecht – Identification of Pedestrian Movement Patterns with Wi-Fi Tracking Sensors

By

C.A.N.L Duynstee
M.J. Haayen
D. Kyritsis
L.M. Ortega-Córdova
S. N. N. Samat

In partial fulfilment of the requirements for the degree of

Master of Science
Geomatics

At the Delft University of Technology,
To be presented on Monday 13th June, 2016

Supervisor:	dr. ir. E. Verbree	TU Delft
Thesis committee:	dr. ir. S. van der Spek	TU Delft

Preface

The Synthesis Project (GEO1101) is a graduation requirement of the MSc. Geomatics of TU Delft. The topic of the Synthesis Project of Spring 2015-2016 is ‘Monitoring flows and occupation patterns with Wi-Fi’ and the project is divided in two groups which are ‘Rhythm of the Campus (TU Delft) and ‘Smart City Dordrecht’.

Three topics was identified under ‘Rhythm of the Campus (TU Delft)’ which were ‘Occupation & Exploitation’, ‘Trajectories: Movement Patterns’, and ‘Activities & Activity Patterns’. Our team, known as ‘The D-Team’, was given the opportunity to work on the ‘Smart City Dordrecht’ initiative.

C.A.N.L Duynstee, M.J. Haayen, D. Kyritsis, L. Ortega-Cordova, S.N.N. Samat
Delft, June 2016

Contents

List of Figures

List of Tables

List of Acronyms

AP	Access Point
CEG/CiTG	Faculty of Civil Engineering and Geosciences
CTS	Clear To Send
EEMCS/EWI	Faculty of Electrical Engineering, Mathematics and Computer Science
GSP	Geomatics Synthesis Project
MAC Address	Media Access Control address
RFID	Radio Frequency Identification
RotC	Project Rhythm of the Campus
RSSI	Received Signal Strength Indicator
SSID	Service Set Identifier
WLS	Wireless Location

Abstract

Wi-Fi tracking technology has entered an age of advancement and is in demand for identifying places of usage demand, discover movement patterns, reconstruct flows and identify parabolic patterns flow and density. For this project, passive Wi-Fi tracking was used to track visitors of the City of Dordrecht and gain insight and perspective of their movement patterns as requesed by the Municipality of Dordrecht for their 'Smart City Dordrecht' initiative. The focus area of our research is the City of Dordrecht which includeds the streets Sarisgang, Kolfstraat, Voorstraat, and Visstraat. These streets are all commercial and mostly pedestrain only streets. The research question that was addressed is'What pedestrian movement patterns could be recognized by the use of Wi-Fi tracking sensors in the city centre of Dordrecht?' In addition to the sensor data, camera data was provided through the Municipality of Dordrecht with two weeks of data collection of people counts. Another data used to answer the sub research questions is our questionnaire done twice during the research period. After filtering the static devices, the above mentioned datasets where analysed individually and then confronted with each other providing validation and insights. As result of the analysis various movement patterns were identified, as well as: 'hot' periods, different patterns between days, and in relation to the opening hours of shops. Charts and appropriate maps and animated visualisations are provided in order to show these results.

Acknowledgements

Though only our names appear on the cover of this report, a great many people have contributed towards making this happen. We owe our gratitude to all those people who have made this Synthesis project possible and because whom our graduation experience has been one that we will cherish.

Our deepest gratitude is to our Synthesis Project Coach, Assistant Professor Edward Verbree who gave us the freedom to explore and at the same time, the guidance. Assistant Professor Edward Verbree taught us to question our thoughts and express ideas. His patience and support helped us during critical situations.

Our specialist expert, drs. Wilko Quak always made himself available to listen and give advice. We are deeply grateful to him for the long discussions that helped us sort out the technical details of our work. We would like to thank him for his willingness on helping us re-configuring the sensors even though he was on vacation. Another gratitude note goes to Lorenzo Dalla Corte for his valuable assistance regarding privacy issues.

We are also pleased to acknowledge the cooperation and support of Cunera Smit, Niek de Wit and Fedde Kuilman from the Municipality of Dordrecht for their interest, support and feedback.

Last but not least, we would like to thank the tutors and mentors of the Synthesis Project, dr. ir. Stefan van der Spek, dr. ir. Martijn Meijers and everyone else involved directly or indirectly with our synthesis project.

Executive Summary

1. Introduction

Wi-Fi tracking technology has entered an age of advancement and is in demand for identifying places of usage demand, discover movement patterns, reconstruct flows and identify parabolic patterns flow and density. To achieve this goal five students, teamed up to integrate and apply all knowledge acquired during the first year courses of the MSc. Geomatics program.

For this project, passive Wi-Fi tracking was used to track visitors of the City of Dordrecht and gain insight and perspective of their movement patterns as requested by the Municipality of Dordrecht for their ‘Smart City Dordrecht’ initiative.

The research question that was addressed is *‘What pedestrian movement patterns could be recognized by the use of Wi-Fi tracking sensors in the city centre of Dordrecht?’* The focus area of our research is the City of Dordrecht which includeds the streets Sarisgang, Kolfstraat, Voorstraat, and Visstraat. These streets are all commercial and mostly pedestrain only streets. This project aims to support the implementation of Smart City concepts in the City of Dordrecht.

In addition to the sensor data, camera data was provided through the Municipality of Dordrecht with two weeks of data collection of people counts. Another data used to answer the sub research questions is our questionnaire done twice during the research period.

After filtering the static devices, the above mentioned datasets where analyzed individually and then confronted with each other providing validation and insights. As result of the analysis various movement patterns were identified, as well as: ‘hot’ periods, different patterns between days, and in relation to the opening hours of shops. Charts and appropriate maps and animated visualizations are provided in order to show these results.

2. Data collection

Sensor Equipment

The equipment used in this project are four Libelium Meshlium Xtreme Sensors which detect and capture wireless signals transmitted from smartphones, tablets and/or laptops. These signals contain a unique identification number known as the MAC address of the device, which is capture by the sensors to be used as signal identifiers. Other devices used for this project are laptops for data retrieval, and smartphones of the team members to validate the data collection. The collected data is stored in TU Delft’s data network and processed using PostgreSQL. Tests were conducted to assess the range of the sensor’s signal detection in comparison to its manufacture’s specifications, and to determine its capacity to track devices indoor and outdoors.

Protection of Collected Personal Data

Due to privacy issues, the MAC addresses collected by the sensors were hidden in our results and visualizations where applicable to avoid the misuse of this information. This information is considered private personal information by the European Personal Data Protection directive. Collected data was not disclosed to other parties in the project except to school officials, because the data was collected for the specific purpose of research. Measures mandated by the Personal Data Protection directive, as well as those given by the 2002/58/EC Directive which safeguards personal data in electronic communications were complied with in the processing of the collected data.by the sensors.

3. Collected Data

The overall data collected from the City of Dordrecht comes from three sources: First, the responses of pedestrians on streets within our research area to the team’s survey questions.

Second, the count of people passing by surveillance cameras. Third, among other attributes, the time and MAC addresses of detected signals collected by the sensors.

Questionnaire

The pedestrian surveys were given as a validation method to the sensor data. All team members surveyed pedestrians during Friday, open market's day. The first survey was done in the afternoon of a market-day and during the weekend festival ''Dordt in Stoom' on Friday, May 27. The second survey was conducted during the morning of the next market day. Random pedestrians passing by the same sensor locations were surveyed by the team on both periods. The questionnaire focused on the use of Wi-Fi devices, the routes and frequency of visits to the city center and if they were local or not.

Camera Data

Weekly surveillance camera data was provided by the research centre 'Onderzoekcentrum Drechtsteden'. Only three locations had cameras placed on streets within the research area. The data contained daily and hourly flow, and inflow and outflow of people to and away from Statenplein. Camera data was processed and analyzed for pattern validation to the sensor data and to complement the research in areas where the sensor data was limited. Because of side-streets along the streets on which the cameras were placed, their count could only be assigned to those who walked/cycled along specific segments. The distance covered in these segments was measured and used in subsequent density analysis. Daily and hourly flow, directional flow and density patterns could be identified.

Sensor Data

After a physical walk through the city centre, sensor locations and arrangements for rotating the four sensors for maximum and most reliable detections of the Wi-Fi signals were considered. Data collection was delayed due to delays in obtaining permission from external influences. The data collection setup was re-examined to obtain the most coverage within the remaining project time. While the initial setup included rotating the sensors' placement within a three week period, after this period was cut to two weeks due to the mentioned delays, the final sensor placement was changed to only detect Wi-Fi signals in the busy areas of the city centre as the objective was to understand the visitors' movement patterns. The sensors were turned on, and monitored via frequent visits to download the data throughout the entire observed period of two weeks.

4. Data Analysis

Sensor Data Analysis Flow Breakdown

First the data collected from all sensors was stored in a table for each sensor. Second, we combined the sensor tables to one single database table, which contain all data for a specific period for all sensors. Third, we made separate tables for each day ordered by MAC address and timestamp. We removed the static devices from these tables by ordering how many hours the device is seen in that specific day. We also remove values that are not useful, like MAC addresses that only contain zeros and timestamps that only contain zeros.

In order to remove the outliers that are only seen by one sensor, we made separate tables for flows between the sensors hourly and the devices that are only seen by 2, 3, or 4 sensors on daily bases. By running the appropriate queries to extract the moving devices which were detected by the different sensors, we were able to identify movement patterns. The found patterns were visualized in order to create a better understanding of the findings.

Validation

One significant factor that can affect the results of the research is the correlation between the number of devices detected and the total amount of people that were in the area, as well as the general coefficient level between the different data sources. In order to do it, the regression method was applied to each pair of datasets and based on the outcomes and the statistic results we defined the relative equation and the coefficient of determination. The high correlation between the filtered devices and the total amount of people counted by the cameras system can lead to the conclusion that the final outcomes, based on the sensors system, can be characterized as reliable with respect to the general movement behavior of people.

5. Results

Questionnaire

The responses of the two surveys conducted, one on a Friday which was part of the festival and also the traditional market day, and the second Friday was a regular, market day reflect the following: As expected, during a festival, there were more outside visitors than locals who rarely came to the city center, stayed anywhere from three to 6 hours in town, and preferred to take different routes as they walked through town. They also carried at least one Wi-Fi device. On a regular market-Friday, most visitors are locals who come every day to every other day, stay from one to two hours, do not often carry a Wi-Fi device with them and tend to use the same route when coming in and out of town.

Cameras

Based on cameras placed at three out of the four streets within the research area, pedestrians flow; direction and density patterns were identified. Given that the camera view sheds were limited to the continuous street segments of their placement, their counts cannot be said to be the same for the rest of the given street because of other intersecting side streets, but the relative differences between the streets can still be noted.

The hour with highest flux of pedestrians in all three streets is around 2:00PM including the market days on Friday but excluding late-night shopping Thursdays. On Thursdays there are two different peak hours of flow and density: around 1:00PM and 7:00PM with the highest values at 7:00PM. This may indicate that late shopping night was a popular day.

The direction in which most people travelled resulted with the traffic on Sarisgang street (closest to train station) of almost twice as much going towards Statenplein than away from this city square. Voorstraat was higher in the outward direction even though the market and popular shops are located in the opposite direction, near Statenplein. Even on market day, the outflow on Voorstraat is higher.

Pedestrian flow on Voorstraat was consistently the highest, followed by Kolfstraat and then Sarisgang even on Fridays-market days. On average, Voorstraat had about 7.5K per day on a regular week, and its flow increased by about 2K or 27% during the weekend festival. In terms of density, Kolfstraat had the highest on a daily basis indicating that this street gets the most crowded at peak shopping time perhaps because it is narrower than Voorstraat. When looking at the aggregated flow of the three streets for each of a week, Voorstraat had at least 50% of the aggregated flow; maybe because of the assorted restaurants, cafes and small shops coupled with a street and that it is not as narrow as Kolfstraat. Friday, Saturday and Sundays had the highest flow and density when compared to the other weekdays. When compared with the festival weekend, the festival brought an increase of visitors to Voorstraat street of about 30% on Friday and Saturday, and almost doubled on Sunday.

Sensors

Based on the data analysis of sensors data, a high flow can be identified every day during the lunch time as well as a higher amount of pedestrians on Saturday 21/05/2016. During this time as well as the whole period of observations, Visstraat and Voorstraat streets are those with the most concentrated flows. Furthermore, unlike Friday night (begin of 21/05), on Saturday night (begin of 22/05) there is a significant increase of the total amount of tracked devices only at these streets. This increase can be explained by the fact that people are used to go out at bars on Saturday night till the first hours of the next day.

Despite the expectations based on previous researches, a low flow can be identified on Sarisgang and Kolfstaat throughout the observation period, compared to the Voorstraat and Visstraat, which contain almost the half of the total amount of devices were tracked . This outcome could also was verified by the patterns of two or three sensors combination.

The majority of the pedestrian movements (≈90%) belongs in the category of patterns between two sensors while a percentage of around 10% contains patterns between three sensors.

In all days we can see that the route Sarisgang – Kolfstraat - Voorstraat is the most frequent used pattern for devices that are scanned by four sensors, unlike the outcomes from the combination of two and three sensors. However, it is important to refer that this pattern can be the most frequently used from this kind of sensor combinations but it consists only the 0.1% of the total amount of movements for the whole day.

6. Visualisation

In order to provide the results of the analysis to the client, in addition to the project report that is submitted to both, the client and TU Delft evaluators, visualizations such as flow and density maps, movement patterns, etc. are used to visualize the results making them more understandable.

7. Limitations

During the project we faced a few limitations. First off all we had a limitation of four sensors to do the total research, this influenced the size of the research area and also the time for a sensor could track from one location. The Meslium sensors also had there technical limitations and the amount of the data we got from the sensors considering the time error should also be taken into account.

Considering the time planned for the project we couldn't capture enough data to compare all days of the week. So we were also not able to see the influence of the weather on the amount of tracked pedestrians. This was also influenced by the time it took to get an approval to place the sensors at the wanted locations. For the validation of the results we used the camera data which where placed on three locations within our research area, there was no camera on the Visstraat, so we were not able to validate this street.

8. Conclusions

When we look at the analysis and the results of our project we can answer our research question *'What pedestrian movement patterns could be recognized by the use of Wi-Fi tracking sensors in the city centre of Dordrecht?'*

Various movement patterns, as well as difference between days and 'hot' periods were identified within the city centre of Dordrecht. They are mostly concentrated on the Visstraat and Voorstraat. There is a strong relationship between the sensor data and the camera data, so the Wi-Fi monitoring system can lead to precise assessment of the movement behavior of the pedestrians within the city center.

9. Recommendations

For further research and for the Municipality of Dordrecht we would recommend to install a free Wi-Fi network within the city centre of Dordrecht. This will make it possible to get a very good representative insight on the movement patterns of the city center. We strongly recommend to capture more data, because the results of the patterns get influenced by many factors, like the weather, the opening times of the shops and restaurants and organized festivals within the city centre. It will also give more data of different days to get insight of differences between all the days within a week, but also weeks within months and difference between the four seasons of a year etc.

1. Introduction

1.1. Context / Geographical Area

The city centre of Dordrecht (***Figure 1*** and ***Figure 2***) mainly the streets of Visstraat & Bagijnhof, Sarisgang, Kolfstraat & Statenplein, and Voorstraat.

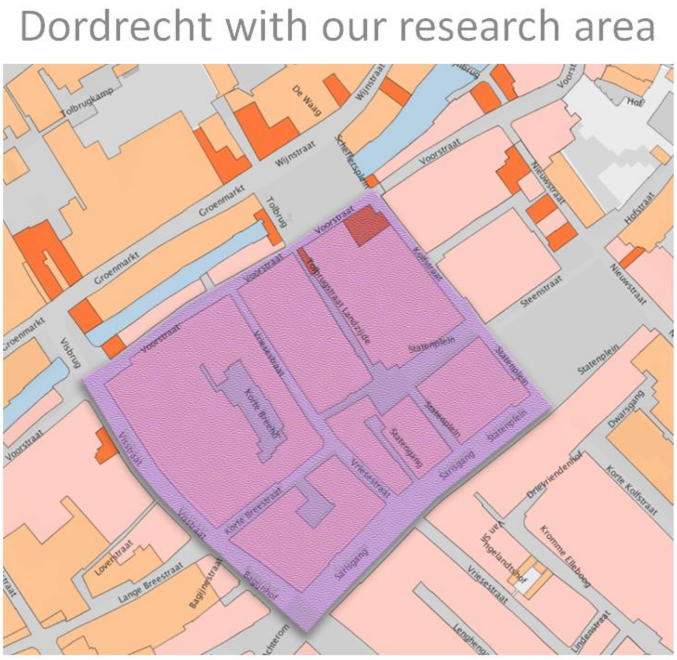


Figure 1, Area of interest shaded in purple

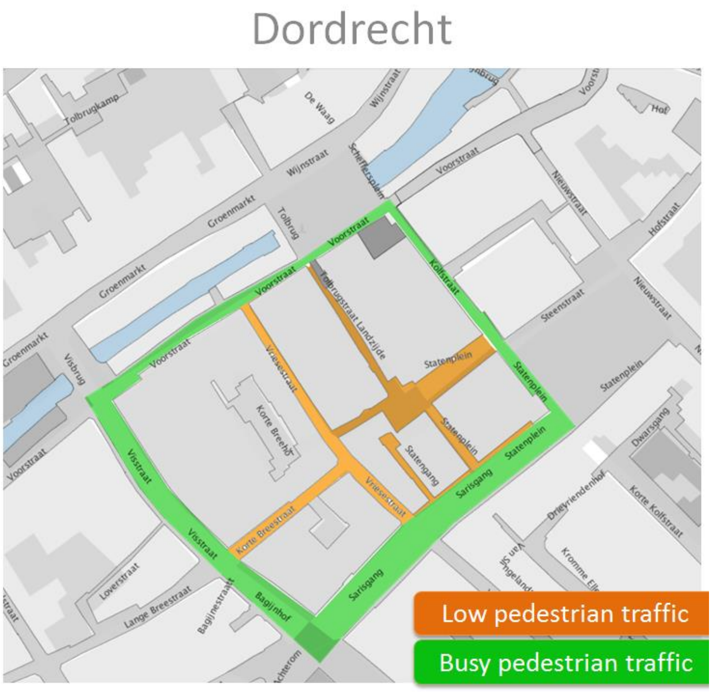


Figure 2, Low and busy pedestrian traffic streets

1.2. Problem definition

Dordrecht is a modern city with a rich past in the western Netherlands. It is the oldest city in the Holland area and has a rich history and culture with 900 national and 160 municipal monuments as well as another 400 characteristic buildings. However, its position near to Rotterdam and other big cities of the country has increased significantly its population and changed its land uses. All the above-mentioned make the center of the city quite distinctive as it combines historical buildings with many new shops such as super markets, shopping malls, coffee bars, and restaurants. However, this combination makes the urban planning and the decision making procedure very difficult as many times the protection of monuments prohibits the development of the area.

One of the objectives of the municipality of Dordrecht is to gain more insight in the movement behavior of pedestrians in the city center. As they want to investigate the existence of hotspots and basically to promote the walking in the area by different roads as a mean to take advantage of the rich past of the city, the knowledge of the Geomatics group can contribute to this research. One approach to address this objective is identifying a way to collect the data from the movement of pedestrians and bicycles in order to have a better view on their traffic flows in the city center. In this research the focus lies on the possibility of using Wi-Fi tracking sensors for that purpose where no free Wi-Fi network exists in the area.

1.1. Research question

Aim and scope of the project

Root definition
The generic topic of the GEO1101 Synthesis Project 2016 is the sensing of movement patterns of individual people and bicyclists, over time in the city centre of Dordrecht by Wi-Fi-monitoring.

Project Goal
Patterns identification for the moving people, pedestrian through the city centre of Dordrecht. The focus area is on a specific part of the historical centre.

Research Question
What pedestrian movement patterns could be recognized by the use of Wi-Fi tracking sensors in the city centre of Dordrecht?

Research Sub-Questions:
- Do the users return by the same way?
- Is it possible to identify “hot” periods?
- Is it possible to identify changes per day?
- Is it possible to identify different patterns of tourists and daily visitors?

Patterns of the movement of pedestrians and bicyclists through the city centre of Dordrecht.

2. Process, Project Organisation and Analysis Requirements

2.1. Process

2.1.1. Users and stakeholders

- The following stakeholders and experts are involved in our project:
- Project Supervisors: Stefan van der Spek and Edward Verbree
- Group Coach: Edward Verbree (TU Delft) and
- Stakeholders: Cunera Smit, Niek de Wit and Fedde Kuilman (Dordrecht)
- Experts: Wilko Quak (Database and SQL)
- D-team: Dimitris Kyritsis, Charlotte Duynstee, Syarifah Nurul Nadwah, Lessie Ortega-Cordova, Jade Haayen
- The stakeholders are the TU Delft and the municipality of Dordrecht.

2.1.2. Meeting with stakeholders, supervisors and experts

First meeting with Stakeholders: 18th April amongst the 5 group members together with stakeholders Cunera Smit, Niek de Wit and Fedde Kuilman. The Dordrecht team explained their expectations towards the project and after discussion came up with the following points:

The goals of the Dordrecht project of the municipality:
- Making Dordrecht a more attractive city to visit, work, live and play.
- The analysis of the project will be used to identify which locations are best to change the traffic movement of pedestrians and bicycle. Direct people towards the centre.
- Patterns from the movements will be used to convince the commercial side of investment

Proposals from both sides on what can be done:
- Placing 4 sensors in strategic locations identified by the Stakeholders and suggestions from team members.
- Doing manual surveys from pedestrians using Google Forms and tablets.
- Combining the data from Dordrecht (cameras and surveys) for the analysis.

After weighing in regarding time and facilities, more specific upshots were done:
- Placing the 4 sensors in the old city.
- Dordrecht is interested to know the pattern of pedestrian movement. The result of the data analysis will be presented to Dordrecht and TU Delft as the final project of the project.
- Dordrecht agrees to provide the team with the necessary data (maps, previous surveys, camera data) and help getting the permits/permission to place the sensors from shop owners and assign a contact person (Fedde Kuilman).

Project Management aspects:
- IAD project management using the MoSCoW method for efficient and effective planning.
- Project Planning. Plan the detailed workload of tasks for the current week and next week. Have daily meetings to keep each other updated.
- Communication is the key in Project Management hence make sure all the questions are addressed and answered accordingly.

Technical details:

- Seek Wilko's help for technical setup of the database used to host the data.
- See as working with only 4 sensors as an opportunity. Have a test run in the campus before bringing it on site to Dordrecht. Assigned 2 people responsible for the sensor.

After weighing in regarding time and facilities, more specific upshots were done:

- Define our scope and what we will not do (make the list as clear as possible). Must work within our Geomatics scope and given time frame.
- Concentrate on the Geomatics techniques / parts of the project and no other areas e.g urban planning. Important to translate the movements and patterns into questions.

2.2. Project Organisation

Our project plan summarizes the objective and scope of the project in order to implement a research of traffic flow of pedestrians in the city centre of Dordrecht. The overall approach of the project from the managerial and technical perspectives are described, as well as, the scope, limitations and risks of the project. An organizational chart of the team which includes the project supervisors and stakeholders, a Work Breakdown Structure (WBS) of the plan, and a Work Package Descriptions (WPD) are included. A Work plan or Time Schedule, and a project Gantt chart are used to help tracking the progress of the project. Team and individual log files are also implemented in order to track the time spent in team activities and the time spent by individuals.

2.2.1. Objective

The objective of the project plan is to serve as a guide to the team and customer, the municipality of Dordrecht, of the overall technical and management activities that are to be executed within the allotted time frame and with available resources in order to achieve the goals of the project as presented in the Specification Requirements document.

2.2.2. Overall approach

There are managerial and technical activities to be implemented within the project period which starts on 18th April 18 and lasts until 17th June.

From the managerial aspect, the team's organization, activities, and the execution of these activities in collaboration with our customer and with other area experts are as follows:

- The team members will fulfill the different roles identified in the organizational breakdown structure (OBS), please refer to the *Appendix A*.
- An overall breakdown of the work to be done by the team is illustrated in *Appendix A.*
- The work plan and schedule of the project is presented in the Ghantt chart (refer to *Appendix B*).
- In terms of communication and project tracking, there are daily team meetings at school and/or online meetings via Skype besides the planned weekly meetings with our customer and supervisor. All meetings are documented with the content of the discussion and agreed upon tasks along with the team member responsible for the task(s). Links to the minutes of the meetings are accessible via the Team log file where the times spent during team discussions are also logged. Contact with the municipality of Dordrecht members is also logged.
- The time spent by individual team members in the execution of tasks is logged in the Individual log file; therefore, we have two logs, the group's activities and individual logs.
- During team discussions, the progress and steps to be done for the present week and the following week are tackled. Through the meetings with our coach and stakeholders, coordination and discussion on possible steps are also planned.

- As a general project tracking tool, the team can refer to the Gantt chart or to the work plan and schedule table.
- All logs and documents are placed and shared with the team via Google drive.

From the technical perspective, the activities to be done to satisfy the technical requirements outlined in the Requirements Specification document. During the planning phase (first phase) of the project, a trial setup, collection and retrieval of the data capture with a single sensor is also scheduled. This will allow the team to get a head start on the technical aspects of the project. The schedule and tasks to be executed in this regard are illustrated in the table and in the Ghantt chart in Appendix Ghantt Chart.

2.2.3. Scope, limitation and risks

Scope:
The focus of the research is in capturing the flow of pedestrians and bicyclists in the city center of Dordrecht using Wi-Fi sensors. The area in which these streets are located is described in detail in the Requirements Specifications document. The team is to deliver flow-patterns and other related analysis results. The team will not include recommendations on decisions to be made as a result of the research findings.

Limitations:
The key technical limitation to this project is the low number of sensors available. Only four sensors can be used. At least two sensors are needed to capture a trajectory so only two streets can be sensed at a time, and the plan is to track pedestrians in more than two streets. Another limitation is a delay of two weeks in the start of the data collection. This is due to the two week holiday period in the city of Dordrecht. This delay has a significant impact on the project since it limits further the amount of data that can be collected. It also makes our project timeline more aggressive. The earliest time data collection can start is sometime on week 3 of the project. The impact to the project is elaborated further on in the Gantt chart section.

There is a risk that the data collection may start even later; due to internal dynamics within the municipality of Dordrecht in order to procure approvals for the team to be able to place the sensors on public space or on business owner’s premises, or to be able to use part of their premises. Should there be delays due to this, then the data analysis period is educed further. Best efforts will be done by the team to still generate good quality results.

2.3. Analysis Requirements

Our project plan summarizes the objective and scope of the project in order to implement a research of traffic flow of pedestrians in the city centre of Dordrecht. The overall approach of the project from the managerial and technical perspectives are described, as well as, the scope, limitations and risks of the project. An organizational chart of the team which includes the project supervisors and stakeholders, a Work Breakdown Structure (WBS) of the plan, and a Work Package Descriptions (WPD) are included. A Work plan or Time Schedule, and a project Gantt chart are used to help tracking the progress of the project. Team and individual log files are also implemented in order to track the time spent in team activities and the time spent by individuals.

2.3.1. Known and unknown in the beginning of the project

Known:
In Dordrecht two times per year one company count for the municipality the pedestrian flows **(*Figure 3*).** In the city center of Dordrecht there are six infrared cameras which count the pedestrian flows.
Pedestrian Intensity - old estimates **(*Figure 4 and Figure 5*)**
Effects on the structure:
- Shift development city center.

- Clustering pedestrian flows: strengthening main structure.
- Recognition pedestrian flows for consumers.
- Connect amplified from central station.

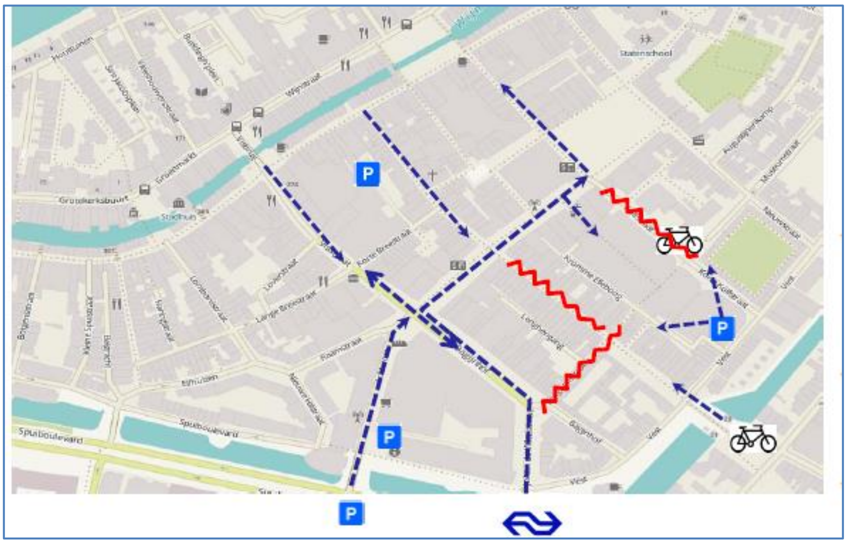


Figure 3, The basic streets in the city centre of Dordrecht

Inquiries from the municipality:

- Establish pedestrian flows (passers numbers).
- Analysis current pedestrian flows: missing links / supply / source points.
- Guidance on future pedestrian flows options?
- Advice on education / compact main shopping circuit.

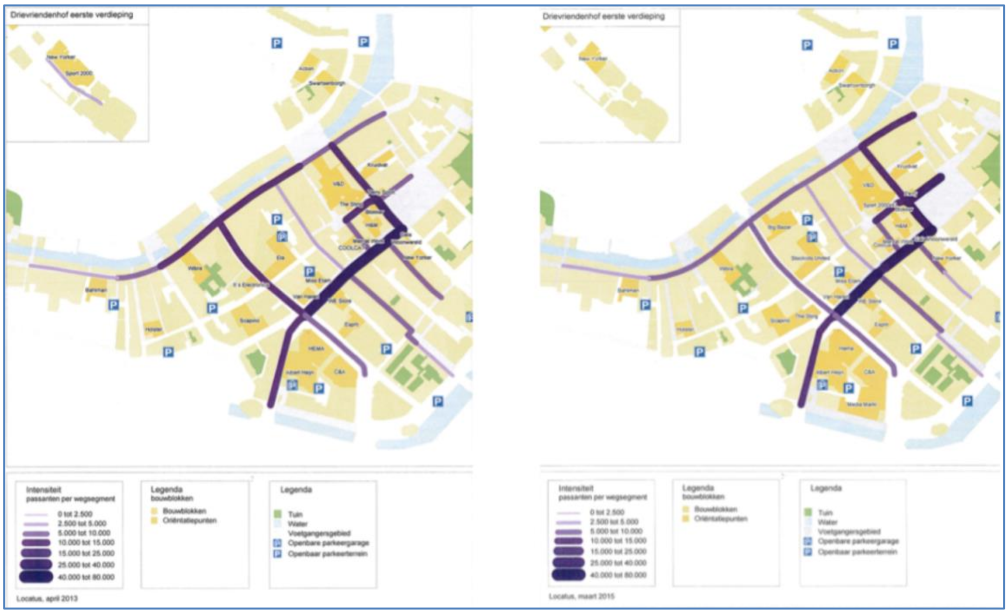


Figure 4, Examples of intensity maps provided from the municipality. This map is a result of manual counting affected by bad weather and prior to Primark's opening. Week day is unknown. Year: 2015.

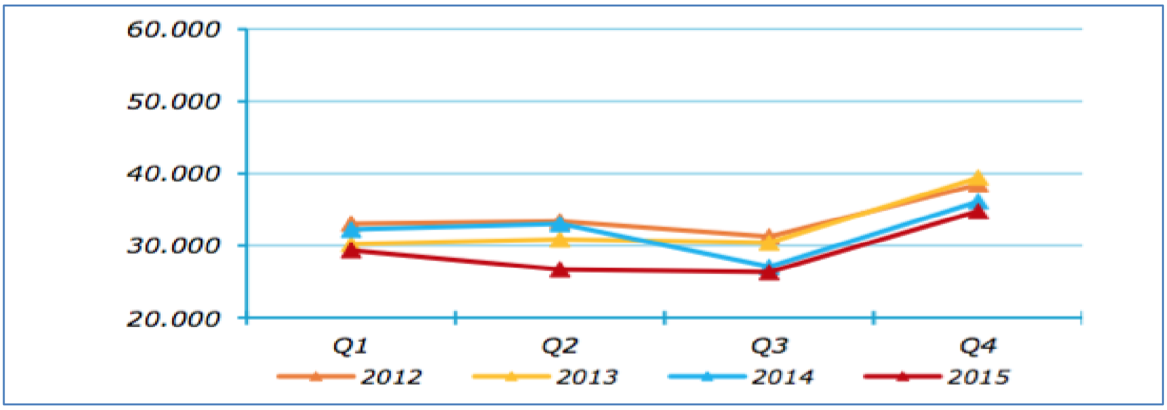


Figure 5, Example of results chart from the pedestrians counting efforts: Average weekly totals per quarter for the period 2012-2015 (provided by the municipality)

Unknown:

- Movement patterns, they only counted (unreliable).
- No background information (about the people).
- No info about route.
- No info about the destination.
- No info about duration and the activities.

2.3.2. MoSCoW analysis

For this project we use the MoSCoW method which is mainly used to achieve a common understanding between the stakeholders and the team that is assigned to this project. By this way, we can order the priorities of the project as well as tasks and deadlines can be managed by an efficient way. The MoSCoW rules include the following structure (wikipedia, 2016):

- Requirements labeled as MUST are critical for the succeed result of the project and thus should be done during the available time defined by the project limitations. If even one MUST requirement is not included, the implementation of the project can be defined as a failure.
- Denoted as SHOULD requirements are the ones which can lead to the success of the project but are not needed to be done within a certain time period. Although SHOULD requirements can be as important as MUST, they can be held back in case of unexpected changes in the plan of the project and can be replaced by another way to satisfy the requirement or can be fulfilled in future timeboxes.
- Requirements labeled as COULD are the ones which helps the project completion but are not critical for its success. They improve user experience or customer satisfaction for little development cost and decision about them is based on the project progress and the time permission.
- Denoted as WON'T Requirements are cases that the project team had already decided to not research and deliver within this schedule timebox. However, they can be delivered in later time boxes or remained as ideas for future research.

If we look to our project we can use the MoSCoW rules in the following way, see ***Table 1.***

MoSCoW

Must:	**Should:**
• Municipality timeliness-- setup tools & biz. approvals • Reliable traffic data • Available sensors • Working data management environment • Analyzed data • 2D visualization	• Other 2D visualizations • 3D visualizations • Keep separate the research data and not combine them
Could:	**Won't:**
• Use camera data • Use past pedestrian surveys • Use our survey • Animation of flow • Use data from the parkings of the area • Take advantage of the already existed application from the municipality for the visitors	• Solution for changing the flow on Voorstraat and Visstraat streets • Consider the use of other tracking methods apart from the already chosen. • Research the access streets

Table 1, The application of MoSCoW rules in our project

2.3.3. Identifications of killer requirements

- Data collection: enough data should be collected in order to manage to do the analysis.
- Security of the Libelium sensors in Dordrecht: They should be fixed and locked in a suitable place in order to not be able to be stolen.
- Location of sensors: they should be located in appropriate places in order to identify pedestrian patterns during the data analysis.
- Number of sensors: The total number of the scanners, four, provided for this project might not be enough to cover the whole research area.
- Electricity power: the shop owners where we will place the sensors should agree to provide us access to the electricity power 24 continuously during the data collection.
- Privacy issues: encryption of MAC address.
- Time limitations: the whole project should be finish in eight weeks.
- Holiday period: During the following period till the deadline of this project there is a holiday period of two weeks. Thus, we cannot collect the data during this period as the project focus on the daily behavior of the users and the data will not be representative. Also, it is possible to have communication problems with the municipality during this period due to holidays.

2.3.4. Rich picture

The following rich picture provides a visual overview of the actors and stakeholders of the Dordrecht project. Those who predominantly have something to gain or lose, or can affect decisions and/or the implementation of decisions are deemed as stakeholders, and those who predominantly are part of the system but do not influence decisions or their implementation are deemed to be actors.

We have identified four stakeholder groups:

- Us, the students who will execute the research activities of this project.
- The Dordrecht municipality team whose city is the site on which our research will take place. The municipality is our customer for whom the research is made
- The business owners who at the moment experience a low flow of people passing by their businesses

- The business owners who currently experience a high flow of people passing by their businesses

As actors, we have identified three groups:
- The TU Delft synthesis project team who guide, facilitate and evaluate the project
- Dordrecht municipality residents
- Pedestrians and bicyclists who visit the city center of Dordrecht

The Dordrecht project aims at collecting and analyzing data on the flow of pedestrians and bicycles in order to identify flow patterns within the city center of Dordrecht. This information is for the municipality of Dordrecht officials for them to have a better view of the flow of visitors in this area. At the moment the flow is high in some streets and low in other streets. Due to this imbalance, there is motivation to use the measured flow and found patterns to help the municipality find ways to better distribute the flow of visitors throughout the city center. The flow information may also help those business owners whose shops or restaurants have a low flow of visitors given that some business owners are willing to invest on improvements of their infrastructures but provided they get more visitors. The Rich Picture below ***(Figure 6)*** intends to visualize the Dordrecht project within this context.

The project is at the core of co-centric spheres which represent the Wants, Impacts and Questions of identified stakeholders and actors. Each group has different wants or needs with respect to the current situation. Some of these *wants* are presented within each group slice in the green sphere that surrounds the project. The results of the team's research will assist in the decision making process of both the municipality officials and possibly of business owners within the low traffic streets of the city center. The impact of these decisions may be positive (+) and negative (-) to each group. Examples of possible impacts are shown on the *Impact* sphere. As in every situation that involves many stakeholders and actors, *questions* or concerns about what is perceived as risks are also present. Examples of such questions are presented in the blue sphere which surrounds all other spheres. In general, the rich picture is not static. As decisions and implementations of those decisions start occurring, the wants, impacts and questions will expand and also change especially among the stakeholders because the roles and actions are interrelated. A larger image of the Rich Picture can be found in the attached pdf file.

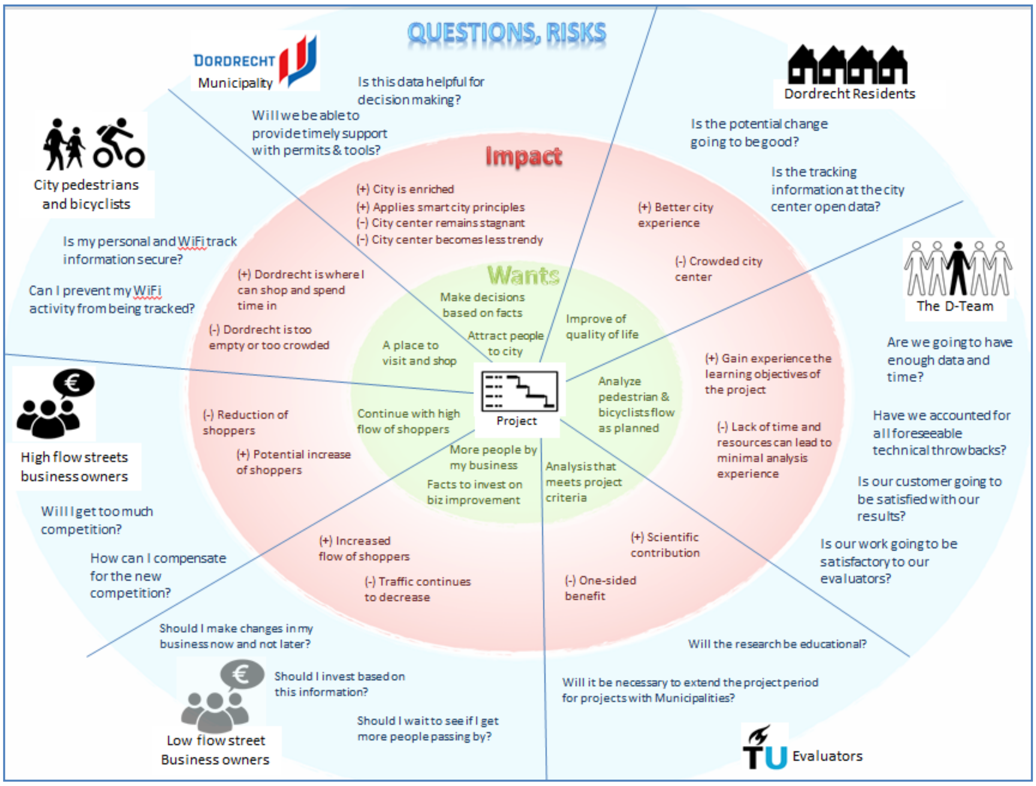


Figure 6, Dordrecht Project Rich Picture

3. Literature Review

3.1. Location tracking techniques

With technology advancement, tracking enters a new era and is revolutionized making the need for WiFi technology tracking. Over time, the use of WiFi tracking has changed from from ‘thread trailing’ and ‘mark and recapture’ approaches to low cost, almost continuous capture of individual trajectories with possibly sub-second sampling rates. Hence, helping WiFi tracking gain popularity, whether it is indoor or outdoor positioning (Deak, Curran, & Condell, 2011).

These location tracking techniques are known as (Deak, Curran, & Condell, 2011)

- Active tracking system – Requires the participants to carry out a GPS device which sends information to the system.
- Passive tracking system – Position of participants are tracked using estimation of the variance of measured signal, this is also known as passive localisation. WiFi scanners are usually the tool for this tracking system.

Active localisation systems for multi-person tracking is easy and can be done using electronic devices, for examples sensors or tags (GPS tracker, RFID tags, membership club cards, etc).
In a wireless network, location distinction is used to detect one or more receivers and when a transmitter’s position is changed. When the position is changed, measurements are made at one or more receivers, it could be vice verse as well. Location distinction is different compared to localisation or location estimation as it detects when its location is different from past locations. Fundamentally, comparisons can be made from below (Li, Wenyuan, Xu, Miller, & Wade, 2006):

- Multipath channel blocks signals hence not enough accuracy for localisation.
- Localisation uses coordinates whereby location distinction does not.
- Localisation system needs three access points or more however location distinction requires less coverage making it possible to determine a new location with only one access point.

Most of the techniques and methods in tracking of localised devices are active system such as RADAR fingerprinting, outdoors WiFi localisation, Bluetooth transmissions and more. Before technology advancement, it is suggested using channel measurements gathered between a single transmitter and multiple receivers in order to perform location distinction. Methods based on wireless link characteristics are growing more popular day by day. Examples are as below (Zhang, Firooz, Patwari, & Kasera, 2008):

- Received Signal Strength Indicator (RSSI): Received radio signal measures the power and it has information regarding a certain link and espcially useful when it comes to using multiple measurements at different receivers. RSSI can also be useful to detect transmitter’s movements. However, this technology is more prone to “spoofing” and putting the information at risks.
- Temporal Channel Impulse Response: Different location of transmitters affects path delays and different multipath characteristics. Each temporal channel impluse link presents a single path in the link multipath.
- Channel Gains of Multi-Tonal Probes: The link received from the transmitter to the receiver composes of paths which are caused by reflections and scattering of radio waves. Multipath is what this is called and it is different depending on time and location, surrounding environment also factors in.

3.2. Passive Wi-Fi tracking

Wifi Tracking uses a multiple number of WiFi monitoring sensors which are then connected to a central tracking server. The device is detected by its emitted WiFi transmissions in which contains the information of the unique MAC address of the devices and does not change the settings of phones and laptops. A WiFi device searches for a known network then it will follow two types of approaches (Leber, 2014):

- Laptops or non-smartphone devices scans for Beacon Frames which are packets broadcast by WiFi routers to let their presence known. The device tries to connect to a network which is familiar or has been used before. In other words, it is searching for its former established connection.
- Probe Request is the second technique and mostly used by smartphones. According to Musa and Eriksson, trajectory estimation of moving devices can be obtained from passive WiFi tracking (Musa & Eriksson, 2012). This is done through beacon frames from the monitor that functions as an access point and connections are tracked between devices and access point. At times, it might use probe requests but most of the time, trying to get more packets from phones and not device detections. Better accuracy means more packets are received.

Having the phone on standby mode is a problem as a fast connection is required between the phone and access points and also how to receive more packets for a higher accuracy (Gudmundsson, Laube, & Wolle, 2008).

3.3. Movement patterns

In order to study spatial behaviours amongst human, animals and the surroundings through tracking technologies it is crucial to have a before and after measurements which are known as spatio-temporal data. Conventional methods lacks validity and accuracy of the data (Spek). Aligned with the advancement of technology, tracking enters a new era and is evolutionized from 'thread trailing' and 'mark and recapture' approaches to low cost, almost continuous capture of individual trajectories with possibly sub-second sampling rates (Gudmundsson, Laube, & Wolle, 2008)

These days, GIS face challenges in handling dynamic and ever constantly changing data hence there is a major reason for this study. Spatio-temporal data is any information relating to space and time which specifically considers involving point objects moving over time. A point object is know as and entity in which the movements are represented as trajectories (Laube & Purves, 2011). As an example ***(Figure 7)***, below is an analysis illustrating the trajectories of four moving entities over 20 steps. The information is: a flock of three entities over five time-steps, a periodic pattern where an entity shows the same Spatio-temporal pattern with periodicity, a meeting place where three entities meet for four steps, and finally, a frequently visited location which is a region where a single entity spends a lot of time (Gudmundsson, Laube, & Wolle, 2008)

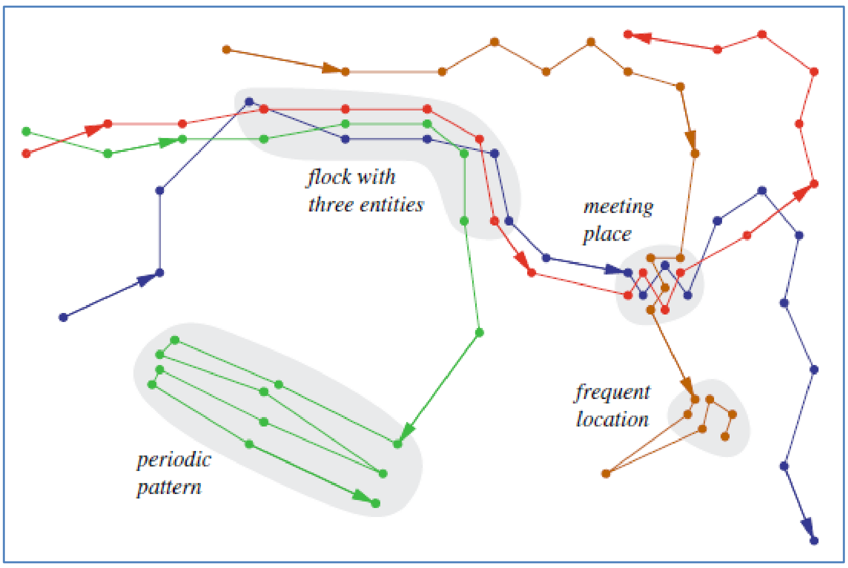


Figure 7, Patterns of Trajectory Movements (Gudmundsson, Laube, & Wolle, 2008)

In GIS context, spatio-temporal data is all about exploratory data analysis and visualization. Patterns are used in various surroundings especially when movements are needed to be addressed. Salient movements are derived from movement patterns and it can emerge from different factors such as (Gudmundsson, Laube, & Wolle, 2008):

- 2D maps of fixes aligned in trajectories – basic movements are derived from simple plotting of movement trajectories on a 2D maps. Trajectories narrow, directed bottlenecks = use corridors. Less focused trajectory / scattered presents more arbitary movement e.g grazing animals, visitors strolling at sports events.
- Effective approach: GIS analysis tools on points and lines representing moving entities. E.g GIS tools for generalization, interpolation and surface generation applied to support detection of movement patterns in trajectory data, use regularly sampled vector field to illustrate the overall pricture of animals moving in their habitat, each vector coding in orientation and size for mean azimuth and mean speed at the very location. Dykes and mountains use a continuous density surface and a 'spotlight' metaphor for the detaction of activity patterns. Common GIS algorithm tools for DTM can be adopted to search salient movement patterns e.g peaks of frequent visitations and ridges of busy corridors.
- Movie-like animated maps – suited to uncover specific movement behaviours of individuals and groups. Constant moving time window in dynamic view uncovers speed patterns of individuals. Flocking / converging are more complex patterns of coordination in groups. Such group patterns are striking in animations.
- 3D representations of movement, if time is used as a 3rd, orthogonal axis – produces a very powerful tool for uncovering movement patterns. Specific geometry in 3d space-time aquarium episodes of immobility and certain speed behaviours produce distinctive patterns of vertical and inclined time lines, respectively. Patterns of spatio-temporal collocation can be identified from vertical bottleneck structures in sets of time lines.

Spatio-temporal data can be use for multiple situation hence the tehnology advancement and gaining popularity day by day. Below are some of the key applications (Gudmundsson, Laube, & Wolle, 2008):

- Animal Behaviour – Crucial to understand their behaviour aspects in terms of popular places, investigation of social structure within a species and also social behaviour in a group for example flock of sheeps, pack of wolves, etc.
- Human Movement – Movement of people can be collected actively and passively. It is good for urban planning, optimizing Location Based System and more. An example of analysis is to estimate current activities or distances or track big events and religious purposes.
- Traffic Management – By collecting data and patterns, traffic monitorings can be done for efficient traffic management in traffic jams or airplane course conflicts. Other parameters can also be observed such as speed and movement direction.
- Surveillance and Security – Through surveillance, a more detailed data sets can be captured hence higher accuracy in terms of analysis. Through surveillance, for example it can prevent a suspect from committing crime. Another good example is the monitoring of fishing boats in sea around Australia. Accidents can be reported fast and illegal fishing or piracy can be regconized almost instantly. Other application includes rerouting of crowds in emergency disaster events.
- Military and Battlefied – Digital battlefield using analysis of spatio-temporal data helps military perform better in the battlefield as predictions, possibilities and probability can be estiated beforehand. In a way, this reduces the risk of fatality in the battlefield.
- Sports Scene Analysis
- Movement in Abstract Spaces

Meshlium sensors are used and it does not function as an Access Point but more into actively scanning the WiFi signals sent out by devices of passerbys, making the detection fast.

3.4. Traffic, Flow and Density Theory

In traffic theory, flow (q) is the rate at which objects pass a fixed point and density (k) is the concentration of objects (N) over a stretch (L) in units of objects per kilometer (WikiBooks, 2015). The hourly and daily flow of individuals is estimated with the following equations:

Flow:

$$q = \frac{pedestrians\ and\ cyclists}{hour} \text{ or } \frac{3600\,(seconds)N}{hour}$$

The density (k) along the covered street segments is estimated with the following equation:

Density:

$$k = \frac{N}{L}$$

Where:

N = number of pedestrians and cyclists occupying a segment of length L (Km)
q = hourly flow
L = length of travelled segment
k = density

In addition to traffic flow and density, the space mean speed ($\overline{v_s}$) measure helps understanding specifics of a flow under investigation. The space mean speed (v_n) is the mean of speeds (same direction) passing a specific point within a time period. It is also known as the average speed over a distance, and it is given by the following equation:

$$\overline{v_s} = \frac{N}{\sum_{n=1}^{N} \frac{1}{v_n}}$$

The individual speed of both pedestrians and cyclists is unknown since no individual tracking is done with the cameras. Such tracking would have to be very precise since all street segments have shops, and/or restaurants so visitors do not walk on a continuously. They stop, shop, or eat and then resume their journey. In theory, the relationship of flow, density, and space mean speed is represented by second flow equation $q = k\overline{v_s}$:which when plotted, provides more insights on the traffic under study.

This second equation below describes Density (q) impacts flow (k). Both are related to each other through ($\overline{v_s}$) described at the beginning of this section:

$$q = k\overline{v_s}$$

The first curve on ***Figure 8*** outlines this relationship with different variables: Q for flow, D for density and V for he the space mean speed. When observing a given speed, a traveler goes faster when density is low, and he or she goes slower when density increases. This is because the negative slope of the flow-density curve is the space mean speed. In general there is a Parabolic relationship: When there is no density (0), flow is also zero given that there is no object moving. When density is low, there is free flow given that a maximum travel speed can be achieved. As density increases, flow starts to decrease becoming bounded (limited), and at maximum density, flow reaches a congested status--when the sidewalks are so full that one barely walk forward.

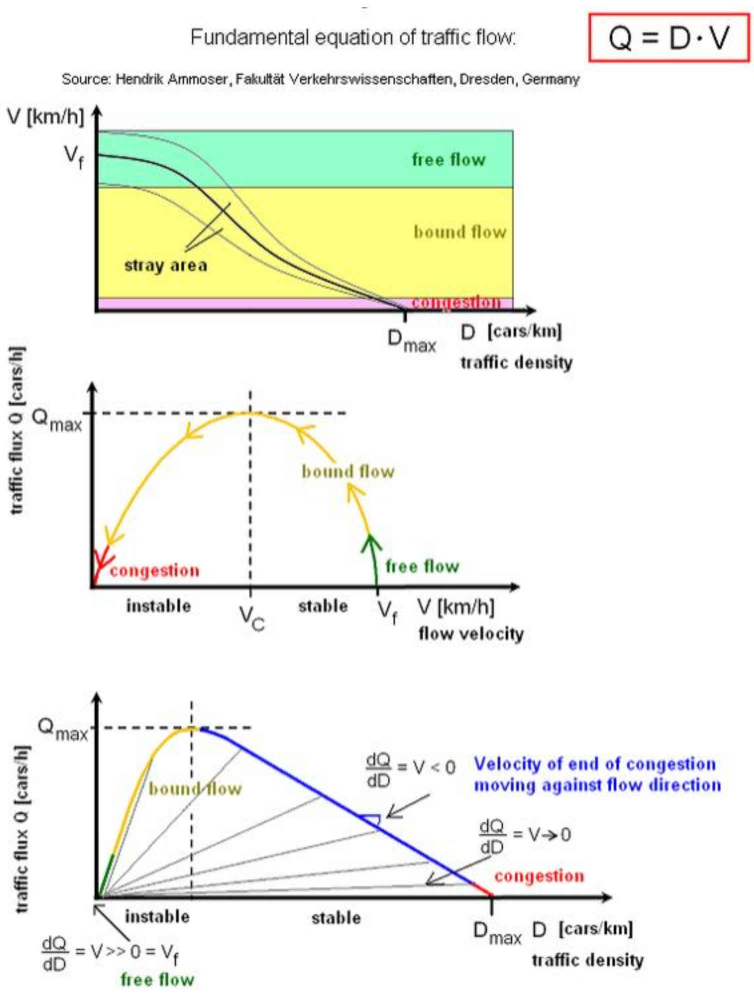


Figure 8, Flow-density and speed curves in a continuous speed model (Wikimedia Commons, 2007)

3.5. Data protection and privacy issues

With the advent of the Public Sector Information (PSI) regulations with the goal of encouraging government transparency and the creation of new opportunities via the growth of information industries, there is a concern that the growing release of all sorts of public data on the internet and its re-use can lead to the breaching of the Data Protection Directive. This is argued in the ‘Brave New Open Data World?’ written by Stefan Kulk, Bastiaan van Loenen (2012). According to these authors, the spread of open data, its re-use and further implementation of open data policies may be obstructed by the Data Protection Directive. This directive provides the rules for the legal collection and processing of personal data in the European Union.

The obstruction concept is raised because data available as open-data can become personal-data when it is combined with other public data or when it is de-anonymized. When this happens, “open data policies can be in conflict with the individual’s right to privacy as protected by the Data Protection Directive” (Kulk & Loenen, 2012). An example of this combination is when people share their current location (geographic data) on the Internet, while It can then be used to develop further location based services, but gathering the geographical information can compromise the protection of personal data even if made open data as aggregated or anonymized information. This is because it is relatively easy to associate such information with individuals given the abundance and long lasting storage of information in the Internet. When this association occurs, the information as seen by the Data Protection Directive becomes illegal because under this directive, personal data can only be collected and processed when there is a legitimate purpose, and in open data initiatives there is no explicit or specified purpose. For this reason the Data Protection Directive supervisor has recommended to forbid re-identification of data-subjects (Kulk & Loenen, 2012).

There are efforts in reconciling regulations between open government data initiatives also referred to as the Public Sector Information (PSI) and the safeguards of personal data provided by the Data Protection Directive. According to Hans Graux (2011) in his report titled ‘Open government data: reconciling PSI re-use rights and privacy concerns’, it is misleading to say that there can be a balance between the re-use of public information and the right to privacy-protection of personal data because their current legal framework, both theoretically and in practice, call for compliance to their set of rules and not for a balance between the two.

Through a couple of application examples, the Slovak case and the UK case, the author argues that within the PSI context, there is no specific rule or exemption for data protection compliance. In the Slovak case, the core functionally of an application involved the processing and publication of personal data, but it brought up the issue of re-using the personal data. Is that still covered by the Data Protection Directive? Is there a legal purposes for this re-use of such personal data? In the UK example, the application design focused on the challenging minimizing privacy issues risks by cutting out as much as possible the processing of personal data including anonymizing techniques and means of remedying in the event of an incident. Both examples illustrate some of the challenges in data protection in designing and offering PSI applications from the two given perspectives. The Slovak example shows how difficult is to measure the legitimate purpose personal data of re-users against the privacy interest of the individual impacted by the open data initiative. It is clear that the PSI re-users face a challenge in complying with data protection rules which continuously change, yet, the legal balance of both perspectives is important to ensure future legitimacy and a positive public perception of PSI re-use (Graux, 2011).

The ever increasing supply of the information added to the Internet by government actors, private sector companies or individuals also increases the supply of information that can be represented in geographical terms. Information that has and geographical component has an added value because it brings insight of on social, political, environmental, demographic and other phenomena. The diverse information and publishers pose a challenge for data protection because those that look after the safety of personal data need to do so without hindering innovative data applications or the flow of public information. So, is argued in the research titled ‘Geographical Information as “Personal Information”’ by Scassa (2010). The author focuses on personal data that has been anonymized prior to being placed in a geographical context. In his view, this personal data is still not secure because of the continuous growth of computerized data from a growing range of data sources which, in the course of time, can make it possible for previously de-identified data sets to be re-identified. Additionally, there are terms and phrases in the personal data protection legislation that needs further definition or clarification. For example ‘reasonable expectation’, ‘serious possibility’, or in the ‘information about an identifiable individual’, the author calls for more concrete meaning of “information about” and ‘identifiable individual’ given that in some cases it is interpreted as ‘inferred’ information as opposed to ‘actual ‘ information. Scassa’s research is based on the on Canadian data protection legislation, but it is based on the Organisation for Economic Co-operation and Development’s Guidelines on the Protection of Privacy and Transborder Flows of Personal Data, “which form the backbone of many data protection statutes worldwide” (Scassa, 2010).

There is another directive that also safeguards personal data and it is the EU-Directive on privacy and electronic communications (2002/58/EC, OJ L 201) which regulates the transmission of a communication which includes the ‘location data’, which is considered to be the geographic position of the terminal equipment, i.e., a mobile phone. An example is the use the processing of any data for location based services (LBS). These services are only allowed if the supplier has prior, informed consent from the user” (Dr. ir. van Loenen, B.; Prof. mr. de Jong, J. ; Mr. dr. ir. Zevenbergen, 2008). In general, the location information of an individual is also protected by the EU Directive of Data Protection and the EU-Directive on privacy and electronic communications (2002/58/EC, OJ L 201) complements the directive by covering the electronic data from services and also the traffic.

4. Practical of Literature Review

4.1. Location tracking techniques

Without having any option, the tracking techbique we use is the passive tracking technique whereby the passer-bys are tracked by the MAC address using WiFi scanners. This method are getting more and more popular by day especially in the commercial side where business owners use WiFi scanners to monitor the flow of their customers so there is efficiency in managing the business and customers. However, in our context, we are deriving an analysis from the results in order to understand the movement patterns in the city of Dordrecht.

4.2. Passive Wi-Fi tracking

Using passive WiFi tracking as our method means trajectory estimation of moving devices, which in our case are the mobile phones can be done with WiFi scanners. . At times, it might use probe requests but most of the time, trying to get more packets from phones and not device detections. Better accuracy means more packets are received. We might incur some problems along the way but because our actual tests are not done yet, none can be reported so far. One of the anticipated problem is that the mobile is on standby mode or the WiFi switched off which will lead to less results and less accuracy in the data.

4.3. Movement patterns

As GIS technology faces the challenge of handling trajectory movements, technology advancement in a way, helps. More research are also being done towards improving tracking technologies and the data can be defined in specific terms. In our case, we collect the data to understand the behavioural human in which the outcome can influence the commercial side of the city area.

For our project of collecting information on passerby in the city of Dordrecht, the purpose is to detect people movement in a short period of time. Using the other method of counting people, is unrealiable as only the numbers are known and not the movement patterns itself (van der Spek, 2010).

Meshlium sensors are used and it does not function as an Access Point but more into actively scanning the WiFi signals sent out by devices of passerbys, making the detection fast which is the purpose of the project.

4.4. Traffic, Flow and Density theory

The traffic flow theory was used to calculate the camera data in which correlated with the sensor data.

4.5. Privacy issues

In order to identify pedestrian and bicyclists traffic the city center of Dordrecht, we need to be compliant to the Data Protection Directive of the European Union to which the Netherlands is a member of. The compliance to this directive is imperative to our project for legal reasons and to address any potential privacy issue that could be raised. While the final report of our project may become public via the TU Delft's website, and therefore deemed as open data, the data that is considered 'personal data' which is protected by the Data Protection Directive is not to be explicitly provided in the report, nor will it be available for de-anonymization or for re-identification of concerned individuals. This is further explained below when discussing our actions in compliance to Article 5 and 6 of the Data Protection Directive.

To fulfil the objectives of our project, we collect an identifier referred to as the Media Access Control (MAC) address which is a unique number hard coded into a device by its manufacturer. The devices in question are mostly mobile phones, tablets, and laptops possessed by a pedestrian or bicyclist. When such device has their WiFI service turned on it can then connect to a WiFi network through their interface with IEEE 802.11 radios signals. Since the MAC address is included in every transmission any WiFi sensor within range (50-200m) that is tuned to the right frequency can detect it (Cormack, 2013). In our case, we are using Meshlium routers.

In Europe, a MAC address is considered personal data. In Opinion 9/2014 on device fingerprinting from the Article 29 Working Party (WP29), accessing the MAC address of a WiFi device is considered to be covered by Article 5 of the ePrivacy Directive. This is because even though the MAC address itself does not provide information about an individual, it is permanent to a given device and because it is easy to intercept using a WiFi network adapter, router; or a simple sensor. Therefore, a MAC address can be used for tracking the movement of a person as his or her MAC address is detected at different sensor points (European Digital Rights, 2015). In our project, a person can be identified especially if his trajectory is coupled with other information (Kulk & Loenen, 2012) such as a timed and geotagged photo.

In general, the collection, processing and movement (transfer) of personal data is regulated by the Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995. Its Article 2 defines 'personal data' as *“any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity*” (European Parliament, Council of October 24, 1995)

As per the directive’s definitions provided in its Article 2, the team members are both controllers and processors of the data. The data subjects are the visitors of the city of Dordrecht whose MAC addresses are detected and stored by our sensors. In our project, there is no ‘recipient’, nor a ‘third party’ to whom the data is given. Both our TU Delft evaluators and the staff from the Municipality of Dordrecht will receive only results, charts and visualizations where the data collected is aggregated, not including individual MAC addresses.

The ‘personal data filing system’ also defined in Article 2 is the built-in storage in the team members’ personal computers. The sensors that we use have the capability of encrypting the MAC addresses as it collects the data. This is done by activating their anonymization feature (Libelium, 2016); however, doing so will not allow us to determine the flow patterns of people for more than one day. Our research focuses on recognizing patterns of people that visit the city center every day (like local citizens), so using the built-in encryption feature of the Meshlium sensors is not optimal; therefore, when the project is complete, all data collected will be deleted, as well as, any copies made for processing. If any data is to be provided or kept by anyone, then team will encrypt all MAC address using other method(s) outside the sensor’s feature.

The data collection and processing as described in article 3 and 4 of this directive is within the scope of this directive and the national law in this regard is applicable.

Article 5 and 6 of the directive provide the conditions under which the collection, processing and move of personal data can be done lawfully. Article 6 specifies that personal data must be collected for specified, explicit and legitimate purposes including historical, statistical or scientific purposes provided the appropriate safeguards are in place. The data collection in our project is solely for statistical and scientific research as stated in the research question of the project. The data is collected and processed by the team members. The storage is secured and not accessible by others, and it is not shared with anyone. It is important to note that, at any time during and after the completion of the project, the team will not to provide the raw data collected to the municipality of Dordrecht (our client) because they do not have a legitimate reason for such collection.

Article 6 also specifies that the data collection should be not in excess and relevant to the purpose, kept accurate and up-to-date if needed, and if kept in a form in which the data subject is identifiable, then it should be kept as such for no longer than necessary. Even though data is collected for statistical research, the amount of processing is not in excess given than a limited amount of querying is done. The data is only taken once and there is no need for it to be updated. Additionally, a person is not identifiable in a direct way given that there is no other personal data linked or referred to in our research. Yet, the data will be kept only as long as duration of the project. The report of our project is to include information of the data subjects only in an aggregated, graphical way. The raw data which contains the MAC addresses from which the visualizations are to be created is not to be included in the report. This will then ensure that MAC address of individuals cannot be re-used, de-anonymized, or re-identified. The visualizations of traffic patterns are to be presented within a geographical context but without any accompanying source data.

Article 7 concerns the processing of personal data. It stipulates that the data subject must consent or the processing of the data is lawful if it is part of a legal action or to protect vital interests of the data subject or of the public or by a third party with a legitimate purpose. In our case, the trial data was collected at TU Delft campus where students are already aware of the WiFi monitor by the school while on campus and they can choose to turn off their WiFi devices if they wish not to be tracked. At the city center of Dordrecht, there will be a sign posted near the sensors stating of collection of WiFi signals and informing by-passers that they can turn off their WiFi devices if they wish not to be tracked. Consent is therefore, given if the by-passer chooses not to turn off their device(s) WiFi service. With regards to the processing of special categories of data addressed in Article 8 such as racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and the processing of data concerning health or sex life is not part of our research so such data is not collected. Article 9 concerned about the freedom of speech and journalism does not apply to our project. Article 10 is applicable and it refers to the information that controllers must provide data subjects. The information to be provided is mainly the identity of the controller or representative, the purpose of the collection, and any further information depending on the circumstances in which the data it collected. In our case, we inform of our identity via a sign, and any other information shall be provided when the opportunity arises. The remaining articles in the directive are either not applicable to our project or have already been addressed in this report.

Meshlium sensors are used and it does not function as an Access Point but more into actively scanning the WiFi signals sent out by devices of passerbys, making the detection fast which is the purpose of the project.

5. Data Collection

The research area of interest is the old city centre of Dordrecht shown on Figure 1 and all the streets with this area are pedestrian streets except Voorstraat. To get to this research area it is important to know that there is a train station on the south end and boat transportation on the north end of the city centre. The research area mainly contains shops, and the late shopping nights is on Thursdays. On Fridays there is a weekly market on Sarisgang street up to Statenplein. Statenplein is a landmark plaza where during market days, groceries items are sold. Within the 15 days of tracking, a weekend-long festival started of Friday, May 27 (Wk2).

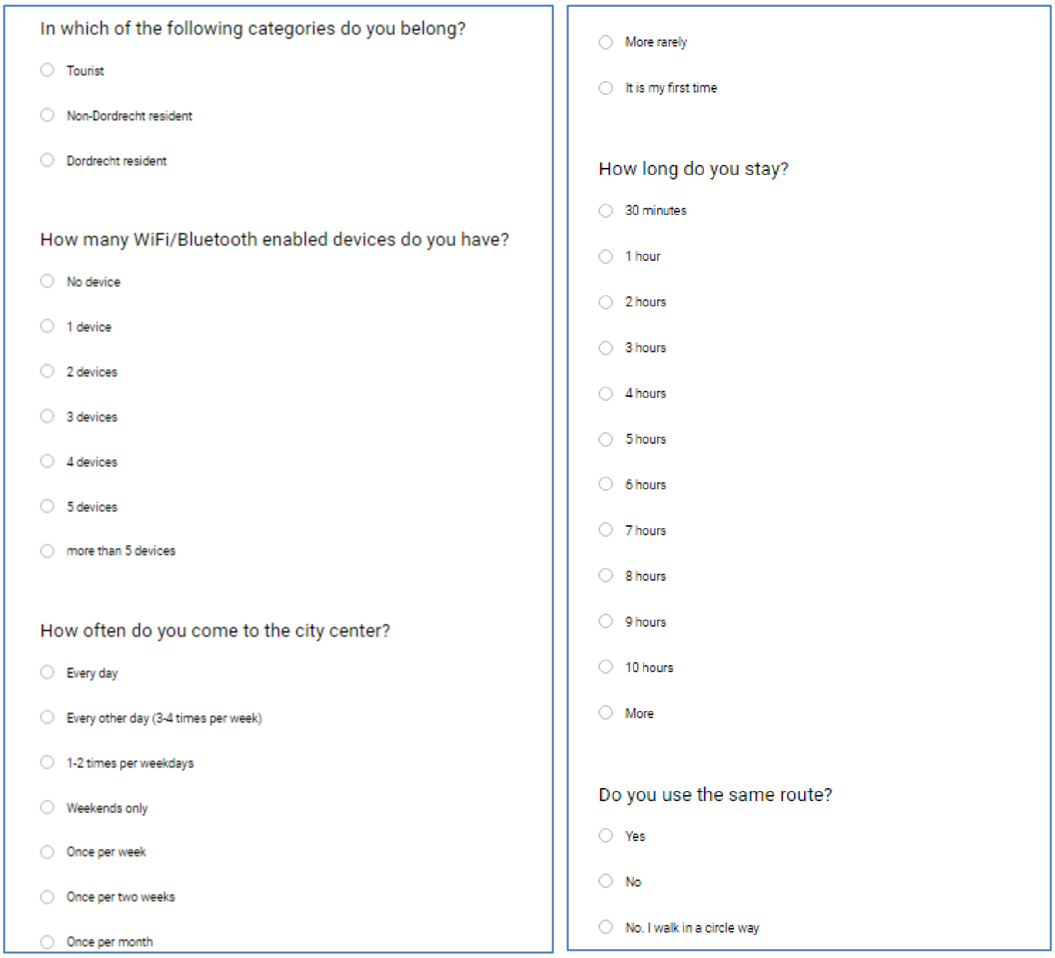
5.1. Questionnaire

The first phase of data collection of the questionnaire was on the 27th May 2016 from 16.00u to 18.15u. This date was chosen because of the weekly market day in the City of Dordrecht in conjunction with the festival ‘Dordt in Stoom’ that took place in that weekend. The target participants were random pedestrians passing by the four sensor locations.

The second phase was done on the 3rd June 2016 from 10.00am – 11.30am, also during the weekly market day but without any event going on. The locations of the survey were the same as in the previous survey, near the four sensor locations.

To conduct the surveys, smartphones with an internet connection were used. This was needed to submit the answers given into the google-base questionnaire. The online questionnaires responses are available through the web link below and the questions asked are shown in ***Figure 9.***

https://docs.google.com/forms/d/1rJFhXWtS_iRrKVb_IiRnf5VLqQejJLxlaLtzKLrt3eY/edit?c=0&w=1


In which of the following categories do you belong?

- Tourist
- Non-Dordrecht resident
- Dordrecht resident

How many WiFi/Bluetooth enabled devices do you have?

- No device
- 1 device
- 2 devices
- 3 devices
- 4 devices
- 5 devices
- more than 5 devices

How often do you come to the city center?

- Every day
- Every other day (3-4 times per week)
- 1-2 times per weekdays
- Weekends only
- Once per week
- Once per two weeks
- Once per month
- More rarely
- It is my first time

How long do you stay?

- 30 minutes
- 1 hour
- 2 hours
- 3 hours
- 4 hours
- 5 hours
- 6 hours
- 7 hours
- 8 hours
- 9 hours
- 10 hours
- More

Do you use the same route?

- Yes
- No
- No. I walk in a circle way

Figure 9, Survey conducted during market day at the city center of Dordrecht

5.2. Camera

5.2.1. Source Data and Coverage

Camera data was provided by the research center 'Onderzoekcentrum Drechtsteden' as daily and hourly counts of pedestrians and cyclists going into (IN) the city centre and coming out (OUT) from the city centre. The IN direction is for flow that goes toward the Statenplein (Dordrecht landmark) while the OUT direction goes away from it. The data covers 15 days: Week 1 (Wk1) starts on Friday, May 20 and ends on Thursday, May 26 (7 days) and week 2 (Wk2) starts on Friday, May 27 to Sunday, June 3 (8 days).

Only 4 out of 7 cameras ***Table 2*** for which traffic count were provided were on streets within the research area of interest. ***Figure 10*** shows their placement on a map of the city centre of Dordrecht. All street segments monitored by cameras are pedestrian-only streets. Given that there is no camera on Visstraat, the traffic of only three out of the four streets that loop around the research area can be analyzed with camera data. Further, the cameras on Sarisgang street-- cameras 3 and 4 were combined by the data provider because both cover the same street. Per the data provider, about a third of the traffic or 33% of its flow is not seen by these cameras except on market day. On non-market days, the cameras do not see those that walk or cycle at the very centre of this street. On market days the market booths setup at the centre of the street so people are forced to move within the range of the cameras.

Figure 10, Cameras within the research area, City Centre of Dordrecht

	Camera ID	Location	Note
1	Kolfstraat	Kolfstraat*	
2	Voorstraat wcd 21	Voorstraat West	Outside research area of interest
3	Voorstraat wcd 32	Voorstraat	Outside research area of interest
4	Voorstraat wcd 23	Voorstraat Mid*	Within Research area of interest
5	Sarisgang	Sarisgang*	Incomplete count on non-market days
6	Sarisgang	Sarisgang*	Data for camera 5 and 6 was combined
7	Vriestraat	Vriestraat	Outside research area of interest

Table 2, Cameras location for which data was provided. A '*' indicates placement within research area.

Because of side-streets along the streets on which the cameras were placed, their count could only be assigned to those who walked/cycled along specific segments. ***Figure 11*** displays the camera locations (yellow pins) and respective street segments (orange lines). ***Table 3*** lists the estimated length of these segments. The distance covered in these segments was measured using Google Earth's ruler feature. These distances are used in subsequent density analysis.

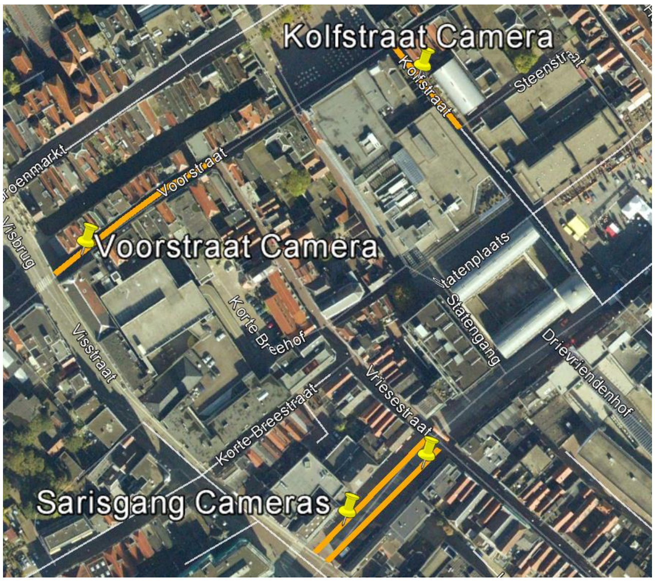


Figure 11, Camera locations (yellow pins) and paths covered (orange lines)

Camera ID	Coverage distance (Km)
Voorstraat	.09
Kolfstraat	.05
Sarisgang 1	.07
Sarisgang 2	.07

Table 3, Segment distance for which the camera counts can be used for density analysis.

5.3. Sensor Data

5.3.1. Sensor Findings

From the previous Test Zero, we managed to test the sensor and determined its capability. As per our observation of placing inside the Civil Engineering and EWI building, it shows that the sensors are able to track devices inside and as well as outside despite the glass or wall barrier. The placement of the sensor in CiTG and EWI shows a wide range of detection hence placing the sensor in the narrow streets of Dordrecht would not be a problem. Additionally, the manufacturer and model of the tracked devices influence the signal strength as the result shows variation in terms of different smartphones models.

5.3.2. Choosing appropriate locations for the sensors

In order to set-up the locations for the sensors we went through a lot of different options. We discussed the positive and negative sides of each set up option we have created, in order to choose our final set-up for the locations of the sensors within the city centre of Dordrecht.

First we looked into our first setup, see ***Figure 12***. In this option we will have 12 locations, so a great coverage of the research area and we will measure within the low and high pedestrian traffic areas. When we went to the research area in Dordrecht we find out that the scale of our research area is not as big as we thought so the movement of the sensors from one phase location to another was not a very big change in distance. Besides having to arrange 12 locations in a very short amount of time, it would also increase the legal issues for the municipality.

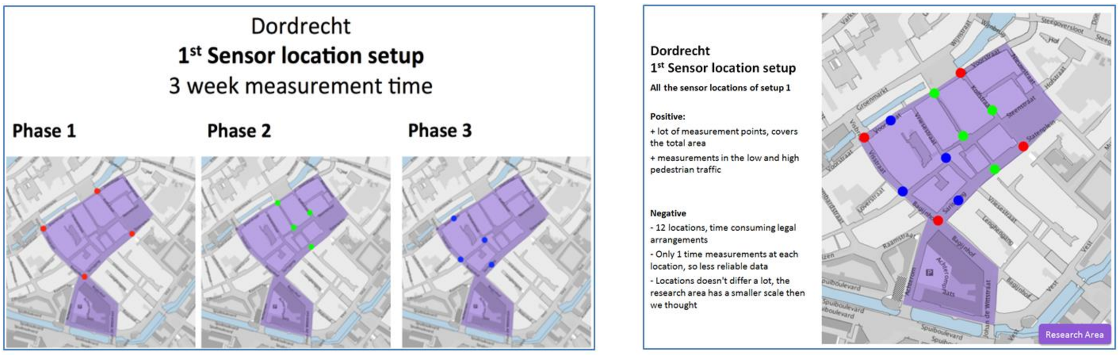


Figure 12, First location setup for 3 week measurement time

In order to get approval to place the sensors at 12 different locations is to much of a risk for this project, considering the time. So we decided to create an option with less sensor location points, so we would have more time to measure. We also took the reliability of the measured data in consideration, and we came to the conclusion that 2 time measuring on the same location creates more reliable data, instead for moving the sensors around several times. After discussing all kind of different options (see attached file to report) we came down to 2 final setups.

One final setup for a measurement time of 3 weeks has 6 locations, see ***Figure 13***. We created a rectangular setup at the high pedestrian traffic streets. The high pedestrian streets will measure more people and will create a global view on how people move within our research area, because this is the main question from the client. With this setup each sensor location would have a 2 week measurement time, so the data would be more reliable than with the 1 week measurement option.

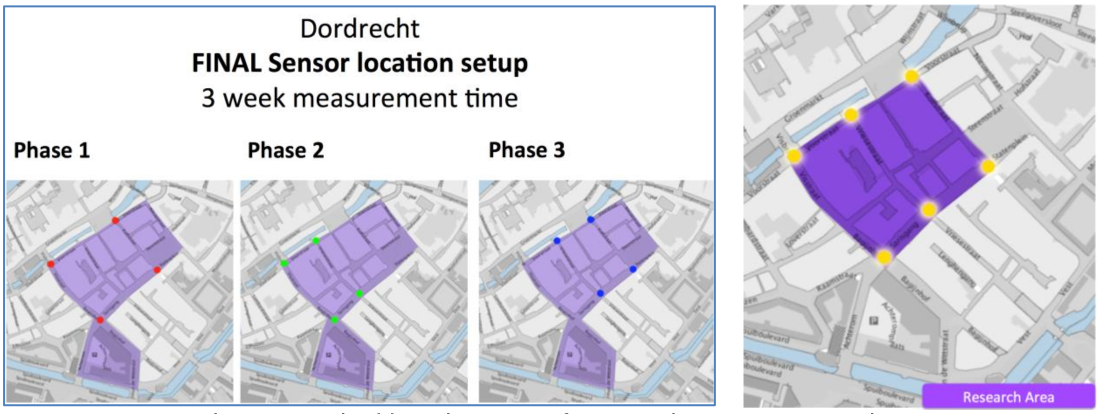


Figure 14, Final location setup for 3 week measurement time

The other final upset is created for a measurement time of 2 weeks ***(Figure 15).*** We created this option in case there is no more time left to do the 3 week measurement. Also for the 2 week measurement time we consider several options, refer to *Appendix*, but we have chosen to go for the option with the 4 sensor locations, because to get approval from the municipality/shop owners to place the sensors is more time consuming than we could have ever imagined. We also choose the 4 location option, because the longer the measurements time at the same location the more reliable the data. Note that we prefer the 3 week measurement time setup, but if we cannot get the approval for placing the sensors in time from the municipality we cannot do anything else then starting with the 2 week measurement time setup because of the time we have left for the project.

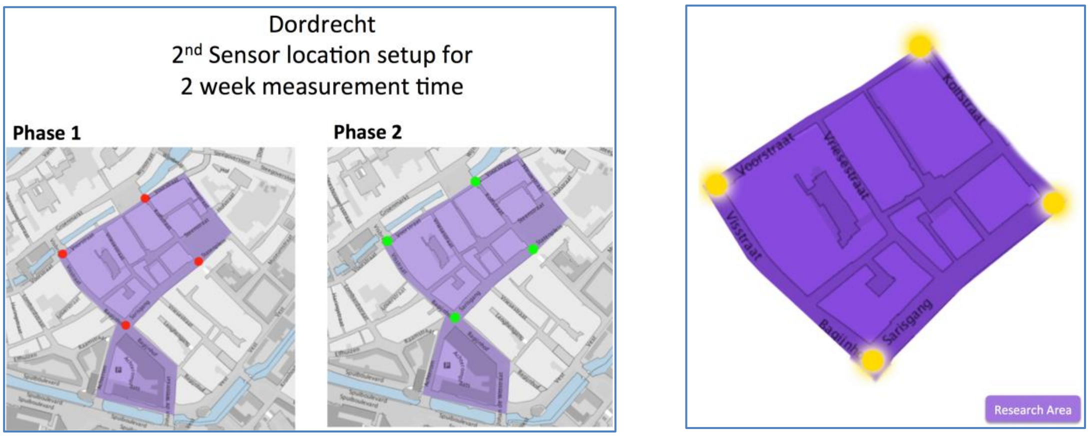


Figure 15, Final location setup for 2 week measurement time

As you can see in the final setup images we created a new research area. We decided to do this because we want to have high reliability with the 4 sensors we have. So we prefer to measure two times at the same location instead of more measurement points. We prefer quality instead of quantity. See ***Figure 16***. The finalised location with sensor's name is in ***Figure 17 and Figure 18.***

Figure 16, Left is the old research area and right the new research area

Figure 17, Final setup of sensors

Figure 18, Sensor locations on buildings

For additional setup details on the sensor location options, please see Appendix Sensor Locations.

6. Data Analysis

6.1. Questionnaire

The respondents that participated in the survey were in total 202 individuals: 119 individuals from Phase 1 and 83 from Phase 2. The responses of the survey are visualized in pie charts in ***Figure 19***. In this analysis, responses given on each survey are compared and overall conclusions are derived, keeping in mind that both surveys were conducted in Friday-market days.

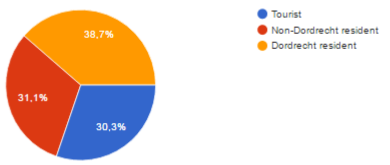
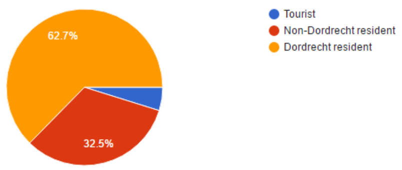
Comparing the answers to the first question: *To which of the following categories do you belong to? Tourist or (non) Dordrecht resident?* One notice that there were more (two thirds) outsiders than locals (one third) visiting the town during the first Friday. Further, the split between tourists--first time visitors or those who rarely come to Dordrecht was almost even (one third each). This could be because of the festival ‘Dordt in Stoom’ that was held during the weekend of phase 1. The reduction of tourists is significant on the second Friday confirming the higher presence of tourists in the first Friday was mostly due to the festival.

When asked in question 2 *How many Wi-Fi/Bluetooth enabled devices do you have?,* at least half of the surveyed people had one device, but this answers included people who had smart phones Wi-Fi/BT disabled. The significantly higher number of people carrying no device at all in the second Friday could be because there were more Dordrecht residents on this day. One can speculate that outsiders always look for Wi-Fi signals hence they have the Wi-Fi turned on all the time because they do not have mobile data.

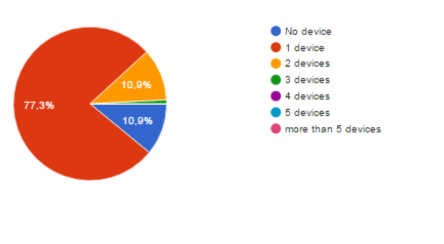
In Question 3 *How often do you come to the city center?* The most significant outcome was that responders of the first Friday rarely came to the city. This coincides with the outcome from the second Friday survey which mostly reflected the presence of local visitors who came to the city center more often.

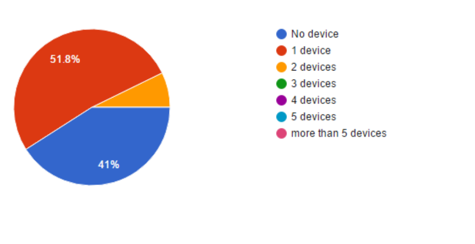
Regarding *how long they stayed in the city center* question, in question 4, during the festival Friday, people mostly stayed between 3 to 6 hours. This correlated to those visitors being mostly from out of town. In the second Friday, the most predominant stay is between one to two hours coinciding with the fact that most visitors are local or residents from Dordrecht itself. Locals stayed for shorter periods perhaps on this day (market day) because going into town means buying groceries or making a quick stop at one of the shops.

In terms of taking *what kind of route* as asked in Question 5, Festival Friday responders avoid taking the same route. In the second Friday, most of the locals are familiar with the surroundings area so they go in and out of the city following their common route.

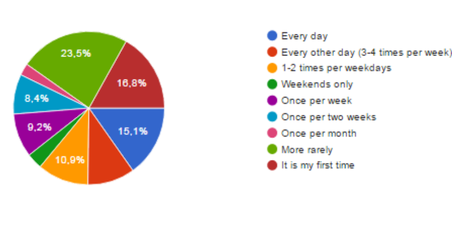
Questions	**Phase 1** (119 respondents)	**Phase 2** (83 respondents)
1. In which of the following categories do you belong to?		

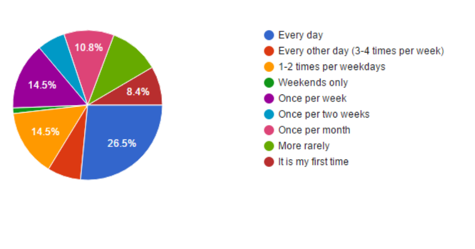
2. How many WiFi / Bluetooth enabled devices do you have?



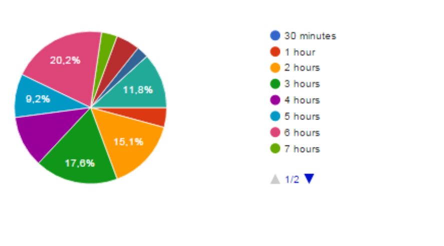


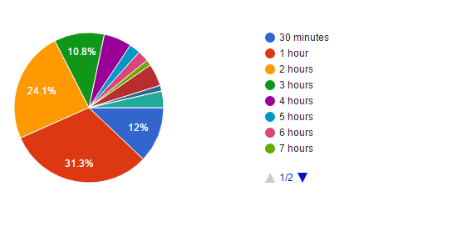
3. How often do you come to the city centre?



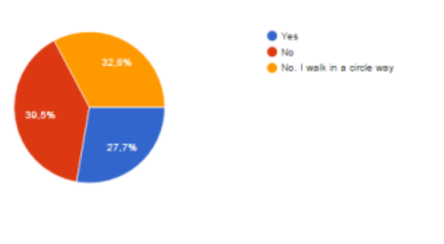


4. How long do you stay?





5. Do you use the same route?



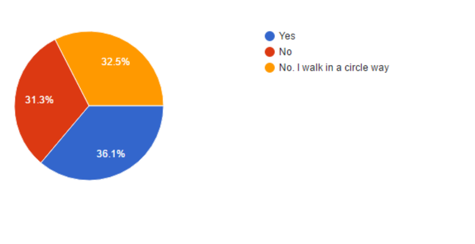


Figure 19, Questionnaire results

6.2. Camera Data

In this section camera data for 14 consecutive days is analyzed. The focus in the analysis is on the identification of patterns and trends and not much on the numerical results.

6.2.1. Scope and limitations

The analysis of pedestrian and cyclists along the street described in the previous section is limited to the specified segments (yellow paths shown on ***Figure 24***. It is also limited in other ways. First, the count provided by the cameras is not 100% accurate. The data provider could only state the accuracy of the cameras is a bit lower when there are a lot of people. Also, as stated in ***Table 2***, there are coverage gaps on the Sarisgang street. Secondly, the time each passerby took to walk or cycle is not measured, so individual speed cannot be calculated. Additionally, there is no way of separating the counted pedestrians from cyclists. Finally, because the street segments covered by the cameras are not connected so individual trajectories cannot be assessed because there is no data to track individual movement from one camera to the other. Yet, in spite of these limitations, flow, directional flow and traffic density along the covered segments could be estimated.

6.2.2. Daily flow

The first set of data, Wk1, spans for seven days. The second set of data, Wk2, data spans for 8 days and includes the festival weekend of the city Dordrecht. The daily flow of each street segments are charted for each week in ***Figure 20.*** The curves are color coded to their location markers. During Wk1, the street with the highest flow and lowest flow are Voorstraat and Sarisgang, respectively, while the Kolfstraat street segment flow is in between the other two. Additionally, we can observe that the daily flow started to increase on Tuesday to reach its maximum on Saturday. It then drops drastically the next day, Sunday, to reach the lowest flow on Mondays. This pattern repeats for Wk2. Full size charts are available in a separate file also submitted with the report.

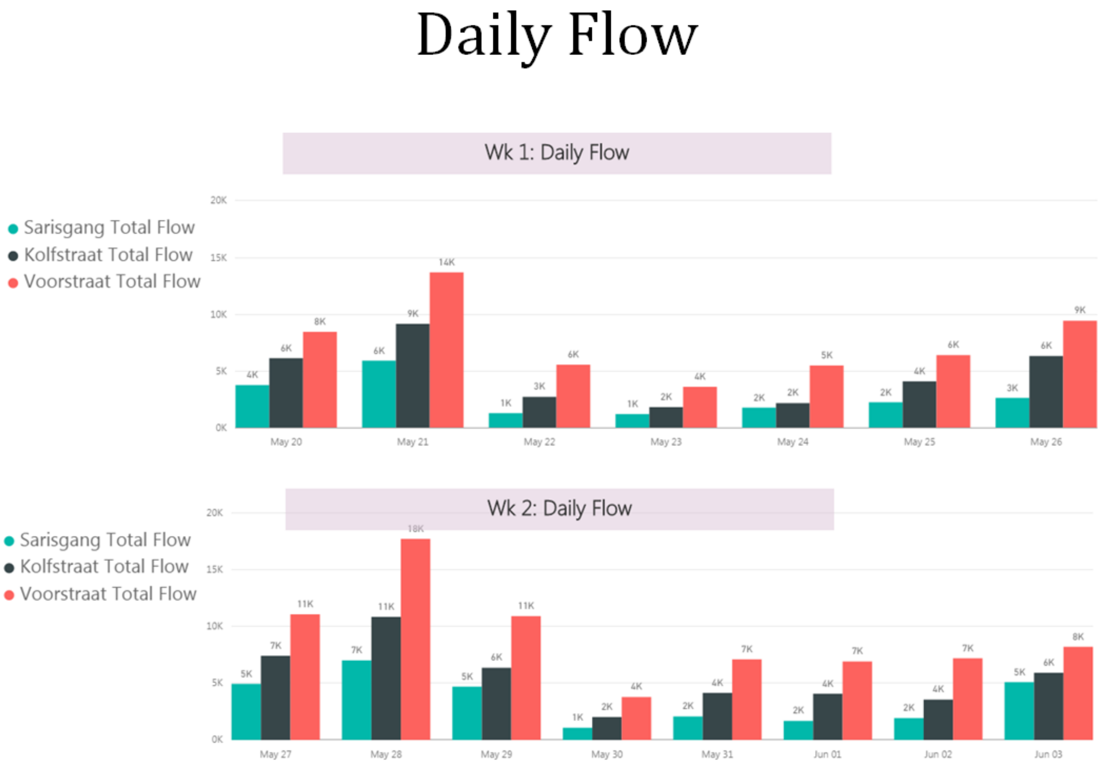


Figure 20, Daily flow of Pedestrians and cyclists from Friday, May 20 to Thursday, May 26.
WK2: Friday, May 27 to Sunday, June 3 (total of 8 days)

As expected, wk2 had a higher average ***(Figure 21)*** than wk1 because of the weekend festival. Generalizing, the Voorstraat street segment enjoyed a flow of about 7.5K people during a regular week, and during the festival weekend week, in average, its flow increased by 1.5K. Similar deductions can be done for the other two street segments.

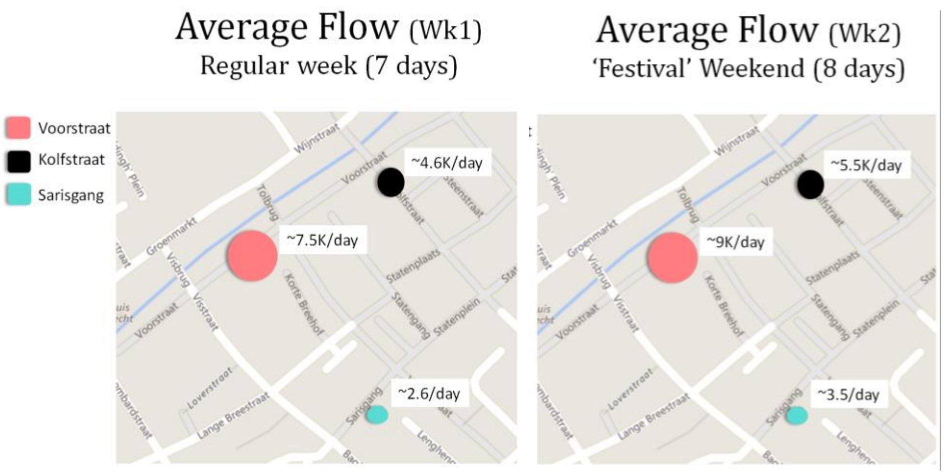
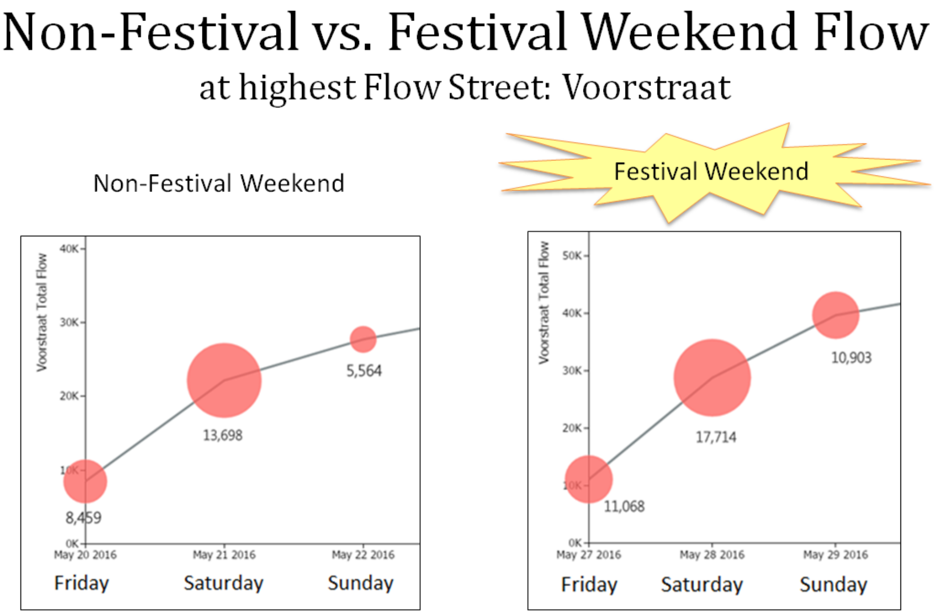


Figure 21, Daily average flow

In terms of weekends only, the increase of visitors that went by the Voorstraat street segment is shown on ***Figure 22.*** The festival brought an increase of flow of about 30% on Friday and Saturday and on Sunday, the flow almost doubles. Similar increase pattern results for the other two street segments.



Percent Increase of Flow due to Festival		
Friday	Saturday	Sunday
31%	29%	96%

Figure 22, Magnitude and percentage increase on Voorstraat segment due to the weekend festival that started on Friday, May 27.

When comparing the street flows to each other in terms of percentage, one can see that in both weeks, Wk1 and Wk2, about 50% or more of the daily pedestrian and cyclists flow is on the Voorstraat street segment ***Figure 23***, while the remaining flow is split between Kolfstraat and Sarisgang. Wk2 charts are available in the Appendix Camera Data Visualisation.

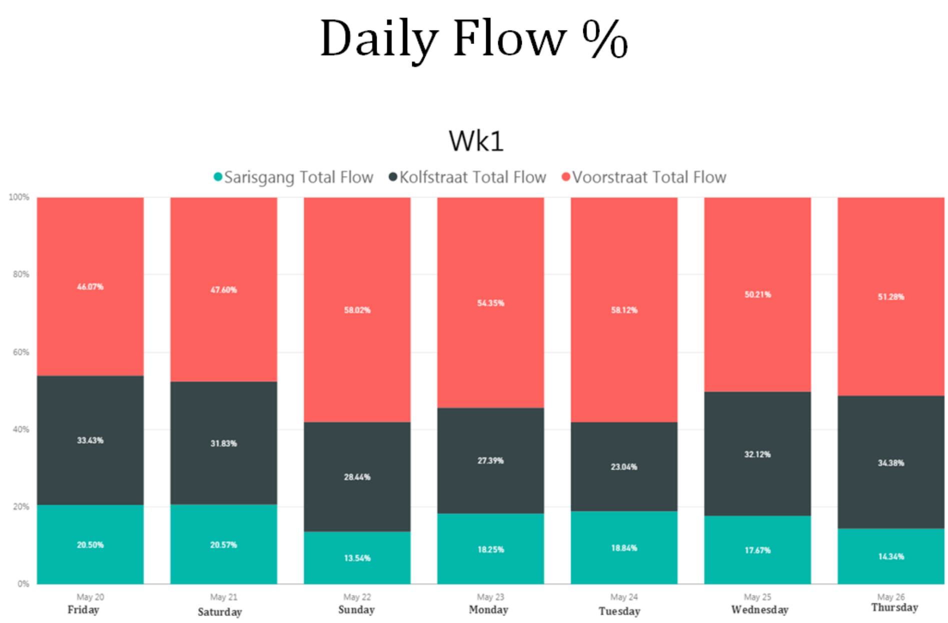


Figure 23, Comparison of daily flow on the three street segments as percentage

6.2.3. Directional flow

In this section we examine the split of the flow in terms of IN and OUT flows. IN represents the flow that goes towards Satenplijn. When looking at the daily flow in this directional way and on the highest traffic days, Friday (market day) and Saturday ***(Figure 24 and Figure 25),*** the inflow of Sarisgang (towards Statenplein) is almost double its outflow. The inflow of Kolsfstraat is slightly higher than its outflow, and the Voorstraat street segment outward flow is higher than its inward flow. Given that Sarisgang is nearest to the train station in the south side, it is justifiable that this street would experience a high inflow, especially on Friday-market day when people went to the groceries side of the market near Statenplein. Similarly, given that boat transport arrives in the north side near Kolfstraat, the inflow rises on this street, as people also go to the market site. Additionally popular shops like Blokker, Kruitvat and Action are on the way from both, Kolfstraat and Sarisgang, to Statenplein. The Voorstraat street segment offer of assorted cafes and restaurants somewhat its higher outward flow given that it is common to have a snack after shopping before leaving town. The trajectory back to the train station could have been somewhat verified if there was a camera on Visstraat, but that was not the case. The same directional flow pattern is found on the second week. The only difference is in their magnitude due to the festival weekend and described in the previous section.

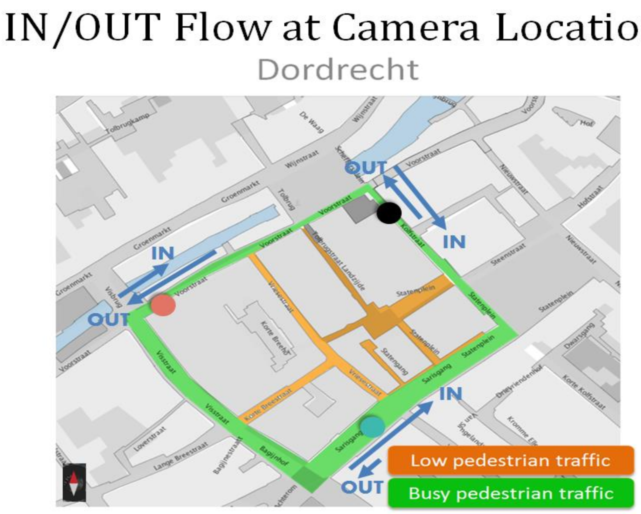


Figure 24, In and outflow from Statenplein as seen from the camera dataset

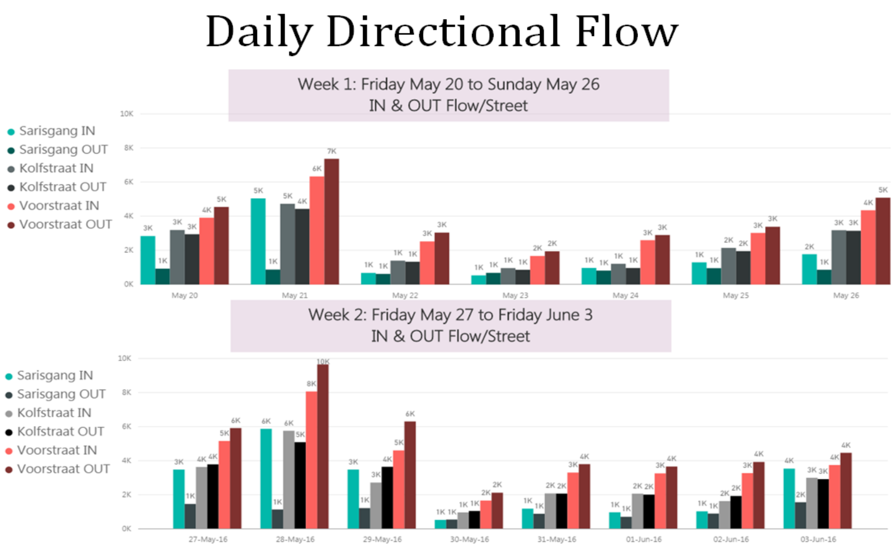


Figure 25, In and outflow magnitudes of each street segment under observation

6.2.4. Hourly flow

Next, the hourly flow of each day ***(Figure 26 and Figure 27),*** Friday through Thursday is observed. Their flow is compared against the second weeks. In general, the same pattern as in the daily flow is exhibited: the Voorstraat street segment had the highest flow throughout the day, every day, and Sarisgang had the lowest flow throughout the day every day except for Tuesday when Sarisgang and Kolfstraat hourly flows were similar. The flow of pedestrians and cyclists peaked every day in all street segments around 2:00PM except for Thursday indicating that most business and services enjoy the highest flow almost every day in early afternoons after lunch

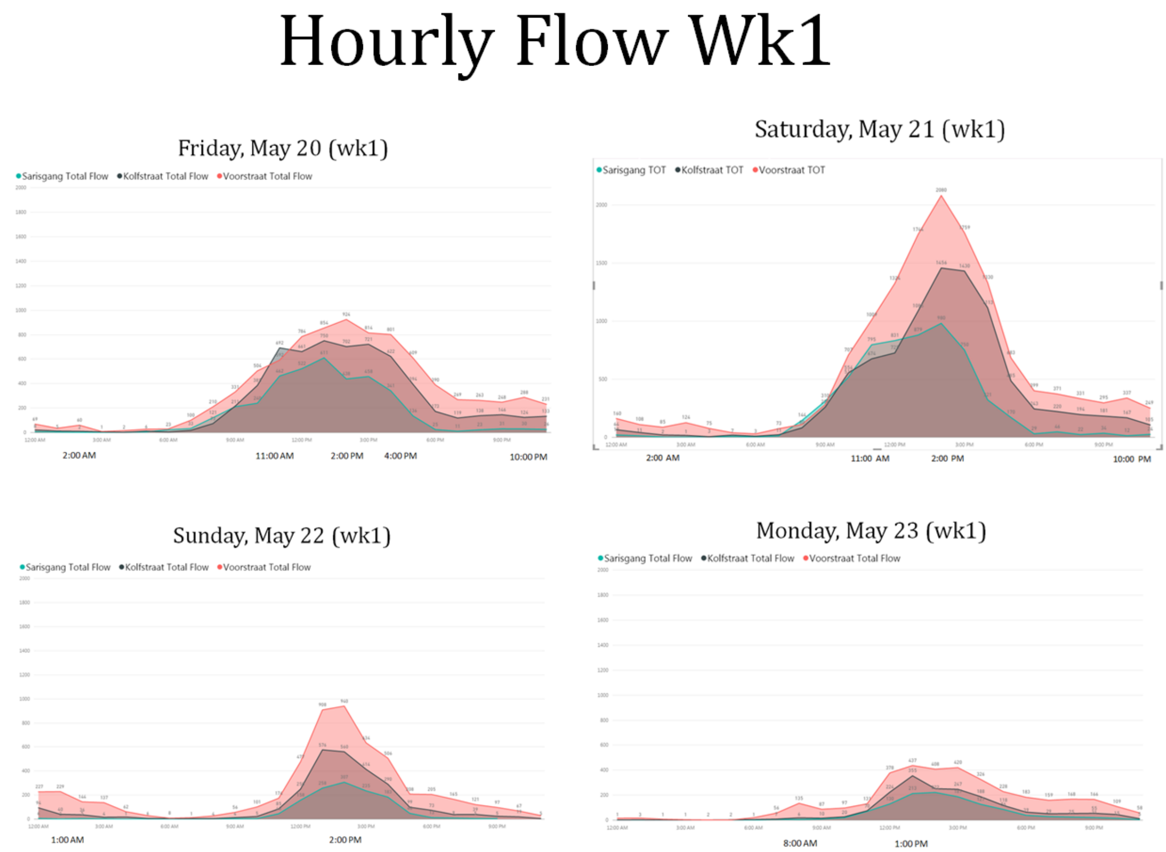


Figure 26, Hourly flow of the weekend including Monday (wk1) patterns

When looking at evening hourly flows, as expected, both Friday and Saturday had the highest flows when compared to other days of the week. Saturday's night flow being the highest and bleeding into Sunday's early morning. The lowest evening flow is on Sunday night.

On Thursday of Wk1 ***(Figure 27)***, the three street segments peaked twice: around 1:00PM and around 7:00PM indicating that Thursday's flow was predominantly food related. On this day, for Voorstraat and Kolsftraat, the evening peak was higher than the earlier peak indicating that more people were present around dinner time than during lunch time. For the segment in Sarisgang street, the peak flow at the evening dinning hour (7:00PM) was lower than the earlier lunch peak. This may be due to the fact that less eating businesses are present on such segment.

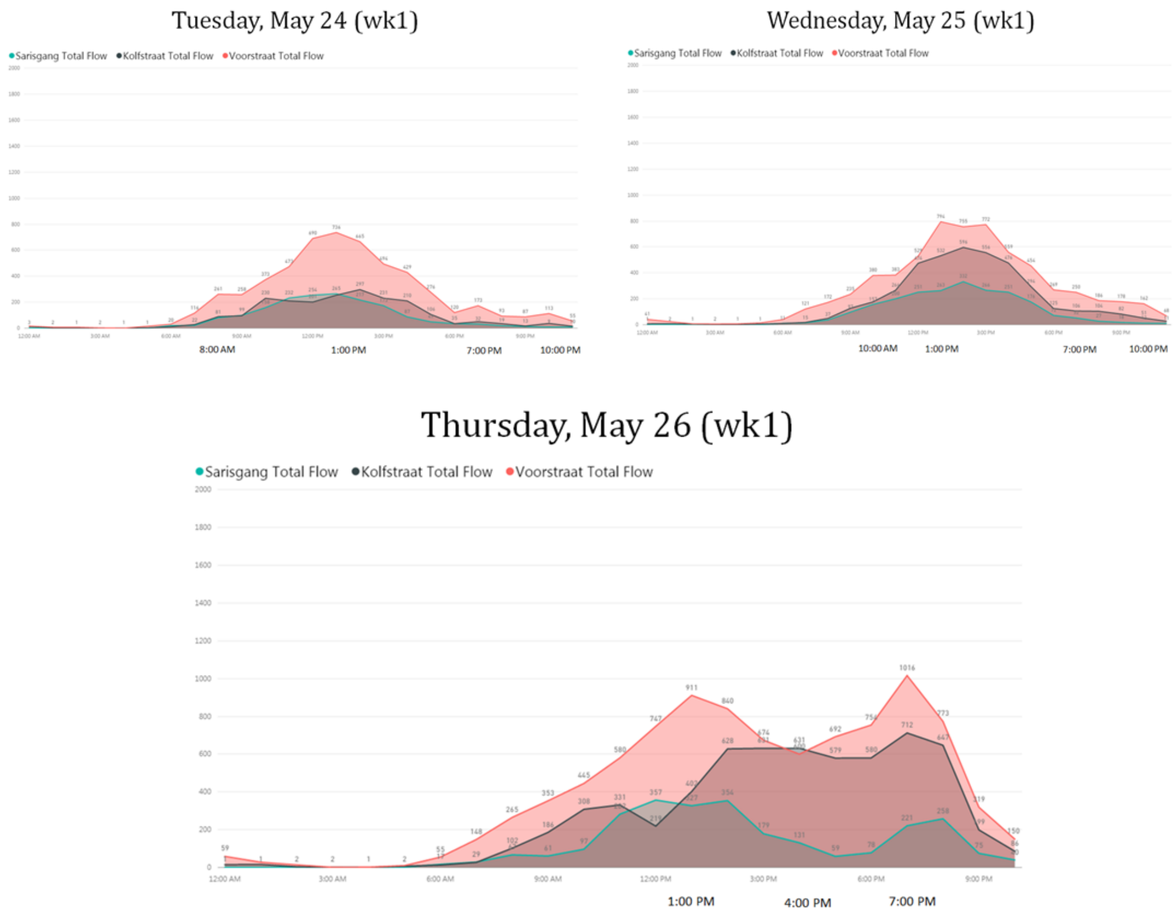


Figure 27, Hourly flow of Tuesday, Wednesday and Thursday (wk1) patterns

Thursday is late shopping night in the city center of Dordrecht, justifying the change in the flow pattern of this day when compared to the other days of the week. The same pattern is found during Wk2 (charts can be found in appendix).

6.2.5. Density

As mentioned earlier in literature review section, density (k) is the concentration of objects (N) over a stretch (L) in units of objects per kilometer (WikiBooks, 2015). In this section, we examine the densities within the three street segments of interest for both weeks. The street segments lengths and definition were described in the Camera Coverage section. Based on ***Figure 28***, the street segment with higher density is Kolfstraat, followed by Voorstraat and then Sarisgang. The magnitudes reflect the limited space in Kolfstraat since it is the narrowest street. Voorstraat does have the highest average flow, but it is also wider so its lower density is justifiable. The daily density pattern is the same on the second week.

Daily Density (wk1)
Pedestrian and cyclists/Km

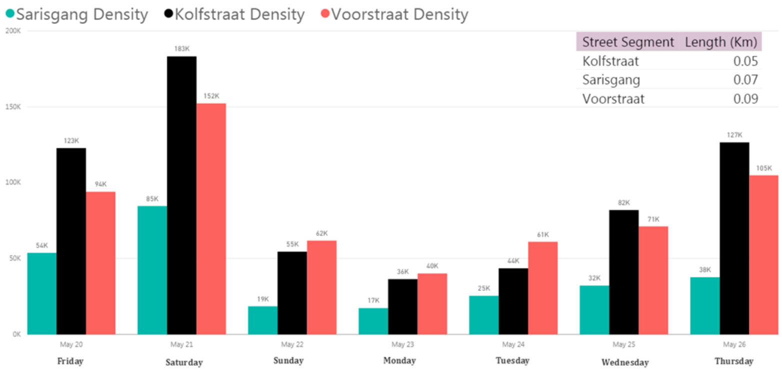


Figure 28, Daily density of each of the three street segments on wk1

Daily Density (wk2)
Pedestrian and cyclists/Km

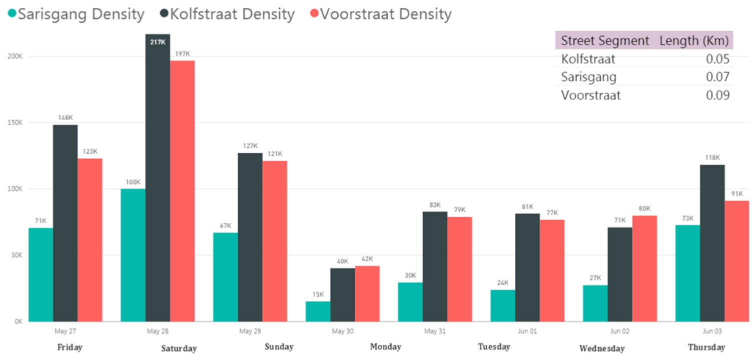


Figure 29, Daily density of each of the three street segments on wk2

The hourly density pattern is consistent with their hourly flow counterparts. The time in which people are closer to each other on the three street segments is at the peak of their flow, around 2:00 in the afternoon.

Date	Hour	Voorstraat IN	Voorstraat OUT	Voorstraat Total Flow	Voorstraat Density
20-May-16	0:00	32	37	69	767
20-May-16	1:00	16	19	35	389
20-May-16	2:00	42	18	60	667
20-May-16	3:00	5	4	9	100
20-May-16	4:00	10	6	16	178
20-May-16	5:00	9	19	28	311
20-May-16	6:00	17	13	30	333
20-May-16	7:00	42	58	100	1,111
20-May-16	8:00	131	79	210	2,333

Table 4, Example of density calculations or camera data for Voorstraat. Distance shown on was used.

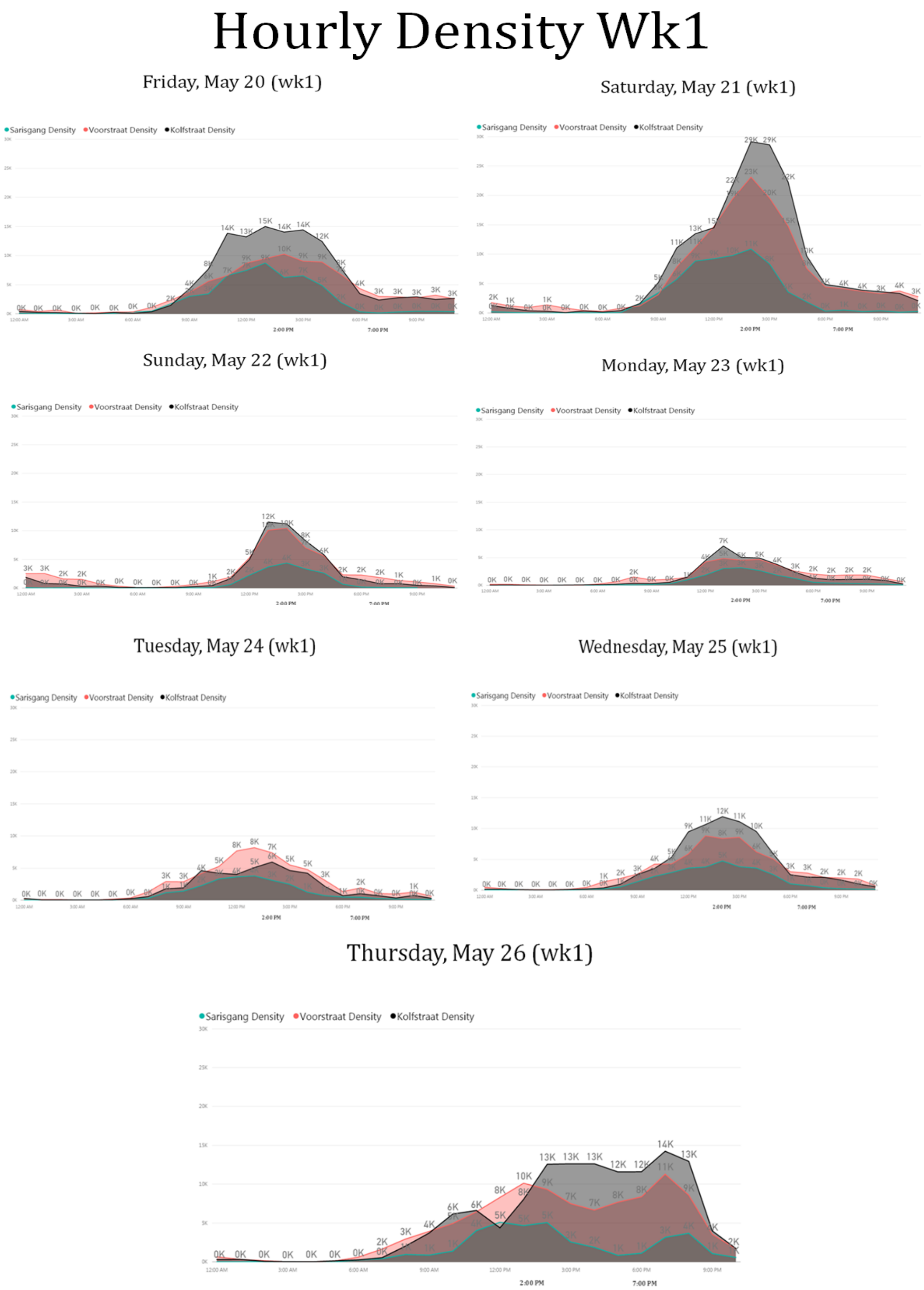


Figure 30, Density patterns of each street for the first week

A similar overall pattern is present in the hourly density of wk2; however, during the festival weekend, the hourly density increases significantly in the weekend as expected, and in Voorstraat street, the density increase much sharply around 2:00PM. Past the weekend, the hourly density comes back to the same level as in wk1. Charts for Wk2 can be found in ***Appendix Camera Data Visualisation.***

6.3. Sensor Data

All the sensor data collected during the periods of 20th May 2016 to 23rd May 2016 have to be filtered before the analysis part. During the filtering procedure, as many outliers as possible should be removed from the dataset. There are two main categories of devices which can be characterized as outliers in this research. First of all, static devices have to be removed. Nowadays, there are more and more devices which work by the use of Wi-Fi such as printers, Wi-Fi routers, parking payment devices etc. or devices that belongs to shop owners or employees like payment devices in restaurants and cafes. As these ongoing records will not be useful for the basic research questions of this project, they can be filtered in order to reduce the size of data and make easier the part of analysis. In order to do this separation, we taking into account the total amount of records for each unique device (unique MAC address) per day. The basic idea is that the number of records for static devices will be much higher than the other ones.

First of all, for the identification and removing of static devices we took into account that a device that is only seen by one sensor, many timed and continuously, it can be characterized as static. However, although the distances between the sensors were quite big, there were devices which were tracked almost simultaneously by two different sensors ***(Figure 31).***

2099	6C:******:DF	2016-05-22 23:58:04	2	Azurewave Technologies, Inc.
2100	6C:******:DF	2016-05-22 23:58:05	1	Azurewave Technologies, Inc.
2101	6C:******:DF	2016-05-22 23:58:24	2	Azurewave Technologies, Inc.
2102	6C:****** :DF	2016-05-22 23:59:51	1	Azurewave Technologies, Inc.

Figure 31, Example of static device which was tracked simultaneously from sensors 1 and 2 despite the fact that the relative distances between them was around 200meters.

Phenomenon like this cannot be characterized as unique as there were many static devices with different MAC address having the same problem. Furthermore, if we check the signal strength of the static devices, theoretically it should remain stable. However, in our cases it is not true as the signal strength was changed quite often ***(Figure 32).***

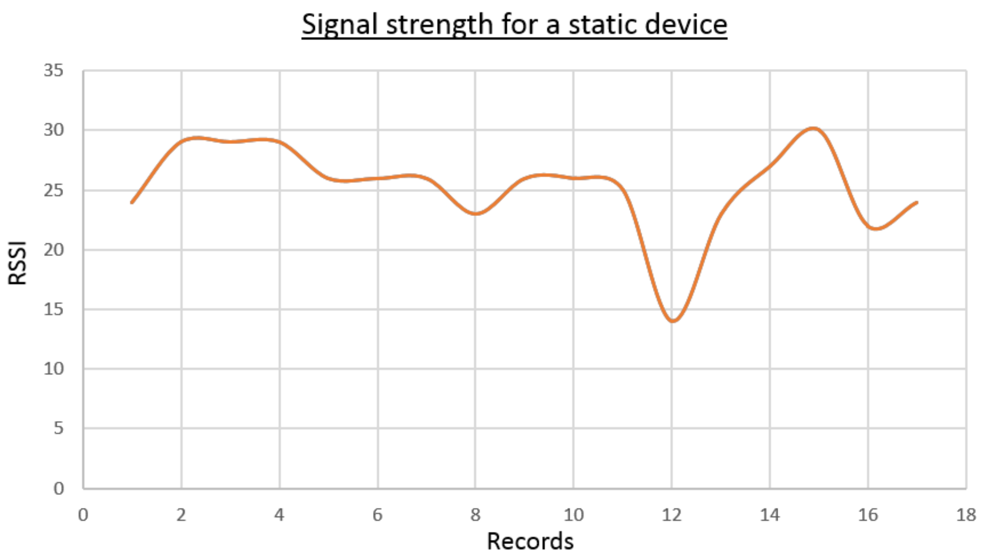


Figure 32, Change of signal strength for a static device

After these outcomes we need to change our definition of a static device. These static devices can be seen by more than one sensor and don't have the same signal strength for longer period. After this results we need to change plan for the filtering. The basic idea of our new filtering procedure is that the number of hours that the static device is scanned should be much higher than others. In ***Table 5*** all unique devices are ordered by how many hours they are tracked on a day. On 20 may we don't have a full day of scanning so the maximum of hours is 14 hour.

Number of hours	20 May	21 May	22 May	23 May
1	15276	25139	11431	9135
2	2365	4089	2041	1504
3	810	1468	654	440
4	301	548	226	208
5	174	245	109	87
6	106	111	55	65
7	69	64	47	49
8	67	82	35	46
9	41	52	37	45
10	44	39	21	41
11	39	36	26	17
12	79	31	11	24
13	113	18	22	25
14	101	18	17	27
15		20	17	10
16		18	15	15
17		13	9	14
18		23	13	11
19		15	11	12
20		18	11	11
21		16	16	10
22		12	20	14
23		20	18	25
24		248	246	245
Total	19.585	32.343	15.108	12.080

Table 5, Amount of unique devices ordered by how many hours they are scanned

As we can see on the above table, the number of devices which were tracked by the sensors reduced exponentially as the duration of the check increases. A peak is visible unique devices that are scanned all 24 hours in the day. This should be the result of static devices that are tracked the whole day. Also, it is clear than the majority of devices were remained in the city center for a period between one to four hours, which is similar to our expectations. We made the assumption that if one unique device was tracked for a period bigger than 12 hours, then this device should be filtered out, this result is shown in ***Figure 33.***

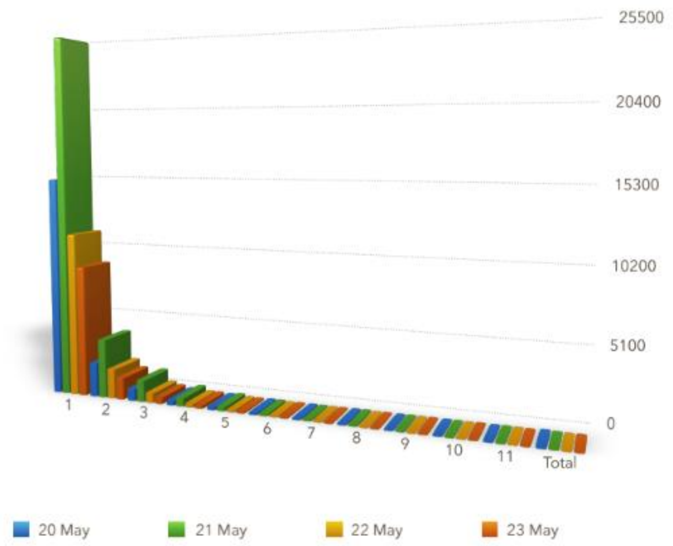


Figure 33, Amount of devices tracked by sensors for each hourly duration

After the filtering we checked how many data is still captured by the different sensors. In ***Table 6*** the results of this test are shown.

	20-May		21-May		22-May		23-May	
sensor	total scans	unique devices	total scans	unique devices	total scans	unique devices	total scans	unique devices
1	42092	11307	54170	16000	31094	8105	25059	6552
2	43225	6993	53628	12241	21179	5269	26201	4775
3	58430	4308	37658	9723	15547	3936	9450	2296
309	11895	983	4930	2268	2041	1058	1840	676

Table 6, Data after Filtering

If we look in to this table, we discovered big differences between sensors. Sensor 309 captured really less data. In most cases sensor 309 has 90% less data than other sensors. This problem is seen before the filtering too. But in the first case we thought that sensor 309 has less data because it was standing in a complete empty building without static devices and that we can solve this problem after the filtering. But if we look at the filtered data this problem still appears. This difference can cause by 2 different reasons:

- The amounts of devices that are in the area of sensor 309 are less than other sensors. This means that there were fewer pedestrians around the Statenplein.
- Sensor 309 had problems with capturing all the devices that were on the Statenplein.

In the first case we can say that the analyses we did on the data is correct. If the second case is the actual case, then it will influence our analysis. The patterns and counting of the pedestrians of the sensor data will not showing the real amount of pedestrians on the streets around sensor 309. Because there was a camera on the Sarisgang and the Kolfstraat we can validate the results of the sensor data with the camera data.

After the filtering of static devices and due to the fact that the main goal of this project is to identify movement patterns, we had to filter the devices which was tracked by only one sensor per day. These kinds of devices belong to users who walked in a different way than our observation network or just passed through one sensor and did not get inside to the city center. Thus, by taking into account the number of sensors which was tracked each unique MAC address, we managed to do also this filtering and reduce significantly the size of the dataset which will be used for the identification of patterns. As an index of the reduction, we can refer that for the whole day of 23rd May 2016 the total amount of unique devices which were tracked, after the filtering of static devices, was 11661 and after removing the devices which tracked only by one sensor the relative value was 2109. Thus the size of the dataset was reduced by around 80%.

The final plan of filtering static devices and pattern recognition is shown in the data flow breakdown ***Figure 34***.

Data Analysis - Flow Chart

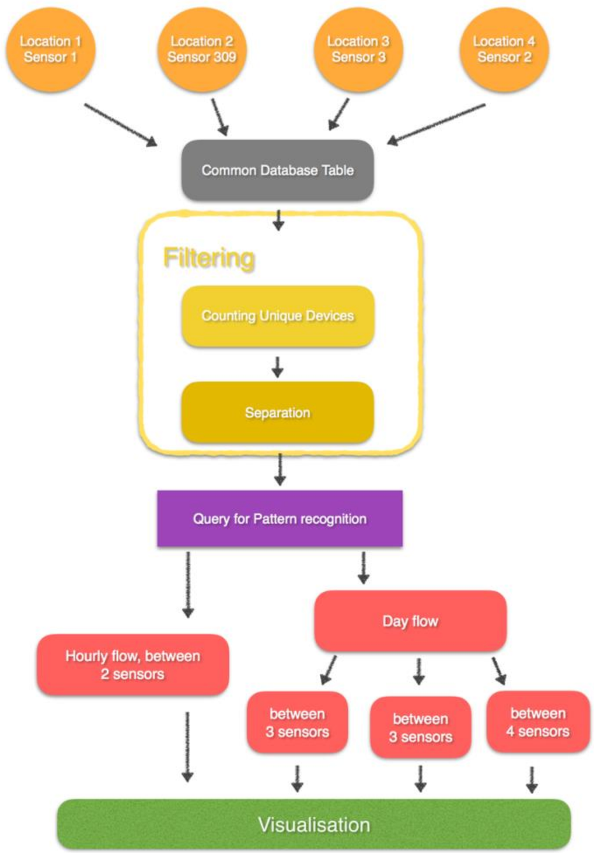


Figure 34, Data analysis breakdown flow

First the data collected from all sensors was stored in a table for each sensor. We added a column sensor to these tables, so that we afterwards can see from which sensor the data is collected. After this we combined the sensor tables to one single database table, which contain all data for a specific period for all sensors. After this we make separate tables for each day ordered by Mac and timestamp. To these tables we add an id and next id to be able to do the queries for the pattern recognition where we apply the filtering queries. We removed the static devices from these tables by ordering how many hours the device is seen in that specific day. We also remove values that are not useful, like Mac addresses that are only contains zero's and timestamps that only contains zeros.

To remove the outliers that only seen by one sensor we make separate tables for flows between sensors hourly and devices that are only seen by 2, 3, or 4 sensors.

From this point, only the moving devices data was further analyzed in order to answer our research question: What pedestrian movement patterns could be recognized by the use of Wi-Fi tracking sensors in the city center of Dordrecht? By querying for devices that were detected by the different sensors, movement patterns were identified, which were visualized along with results in order to create a better understanding of the findings.

7. Sensor Problems

The following ***Table 7*** illustrates the whole period of data collection, 20th May 2016 to 3rd June 2016, in the city center of Dordrecht. It shows the days we visited the area and emptied the data from the sensors, which are visualized with colors.

Date	Meshlium1	Meshlium2	Meshlium3	Meshlium309
20/05/2016				
21/05/2016				
22/05/2016				
23/05/2016	OK	OK	OK	
24/05/2016				+0^h : 23
25/05/2016				
26/05/2016				
27/05/2016	OK	OK	+0^h : 18	+1^h : 54
28/05/2016			-	
29/05/2016		-	-	
30/05/2016	OK	OK	+1^h : 57 RESTART	+1^h : 41
31/05/2016	OK	OK	OK	+0^h : 30
01/06/2016				
02/06/2016				
03/06/2016	OK	OK	+0^h : 46	+2^h : 11

Symbol	Meaning
	Duration of data collection between downloads
	No data
	Download without problems
	Download with problems
	Fixed time problem
	Can be used based on assumption
	Periods which can be used for patterns

Table 7, Procedure of data collection and the relative time error problems

Based on the diagram it is clear that there were no problems for the sensors Meshlium1 and Meshlium2, apart from the 29th May 2016 where we do not have data from sensor Meshlium2. This error can be explained by the fact that the Wi-Fi tracking was disable for some reason (electricity power outage). However, the other two sensors had significant problems with the time synchronization. During the placement of sensors on 20th May 2016, we fixed and synchronized the time zone of all of them. However, when we checked at the first time the sensor Meshlium309 we noted that the time of the device was wrong as it was 23 minutes later than the real time ***(Figure 35)***. After the downloading of the data, the time was corrected but we had the same problem also during the next efforts with various time differences from the real time.

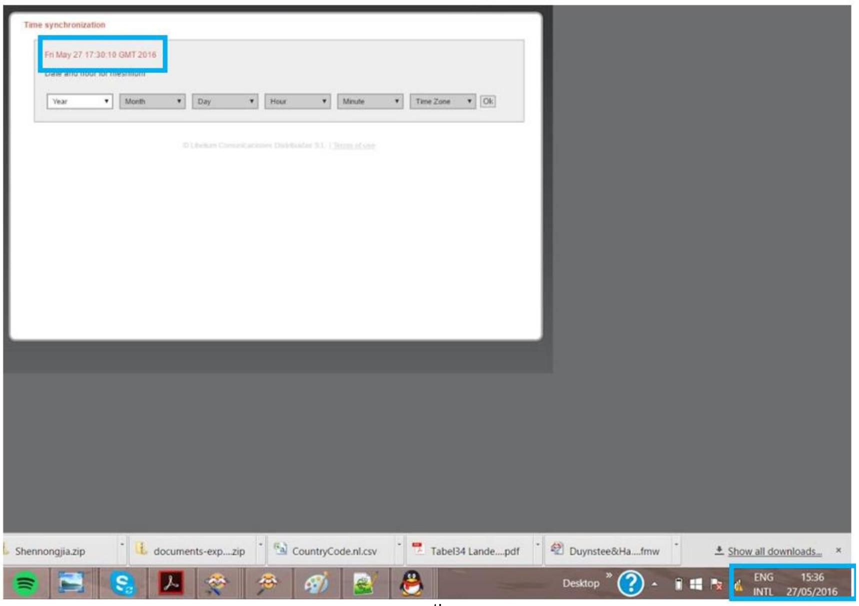


Figure 35, Time problem on 27th May 2016 for Meshlium309

Despite the fact that sensor Meshlium3 worked correct till 23rd June 2016, we had similar problems with the setting of time during the following days ***(Figure 36).***

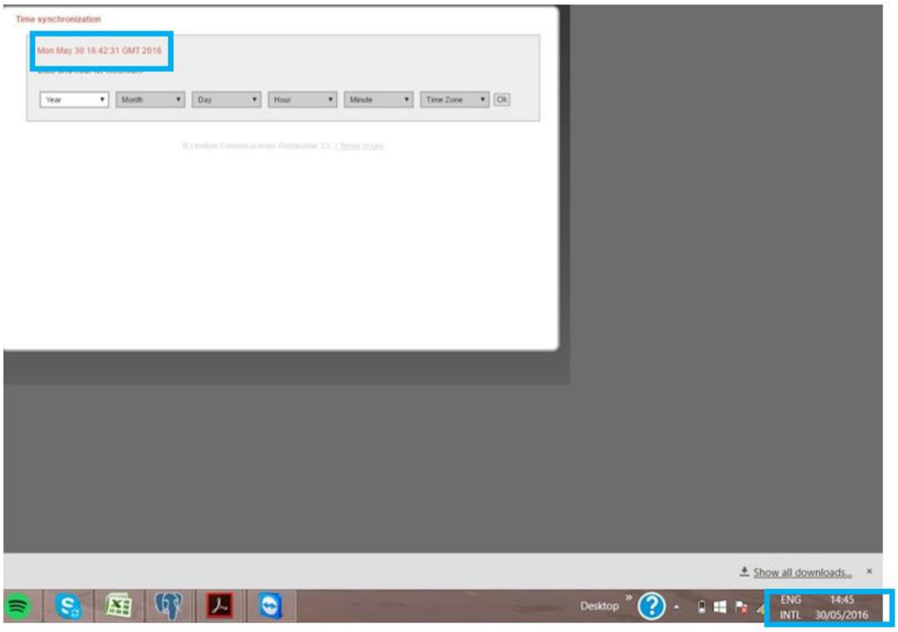


Figure 36, Time problem on 30/05/2016 for Meshlium3

Furthermore, on 30th May 2016, it was impossible to connect to the sensor and get the data as we get different kind of error messages when we tried to have access to the local database of the device. As an alternative way, we tried the use of SQL query in the relative software of the sensor in order to explore also this way for the download of the data. However, a similar error message was appeared. The following figures illustrate some of them ***(Figure 37, 38, 39 and 40).***

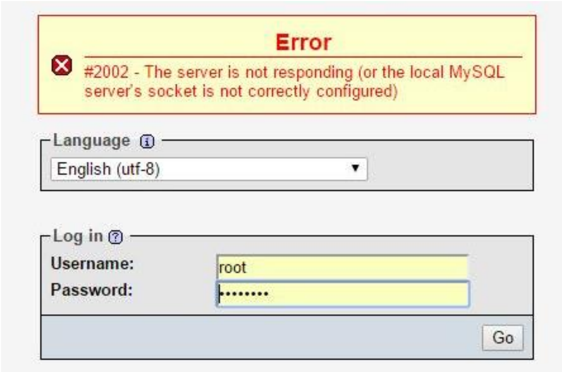


Figure 37, First connection problem on 27th May 2016 for Meshlium3

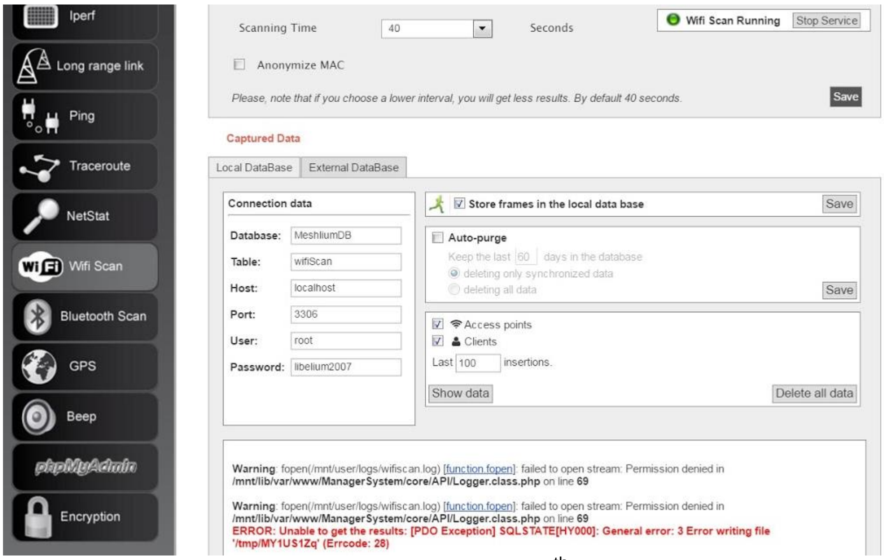


Figure 38, Second connection problem on 27th May 2016 for Meshlium3

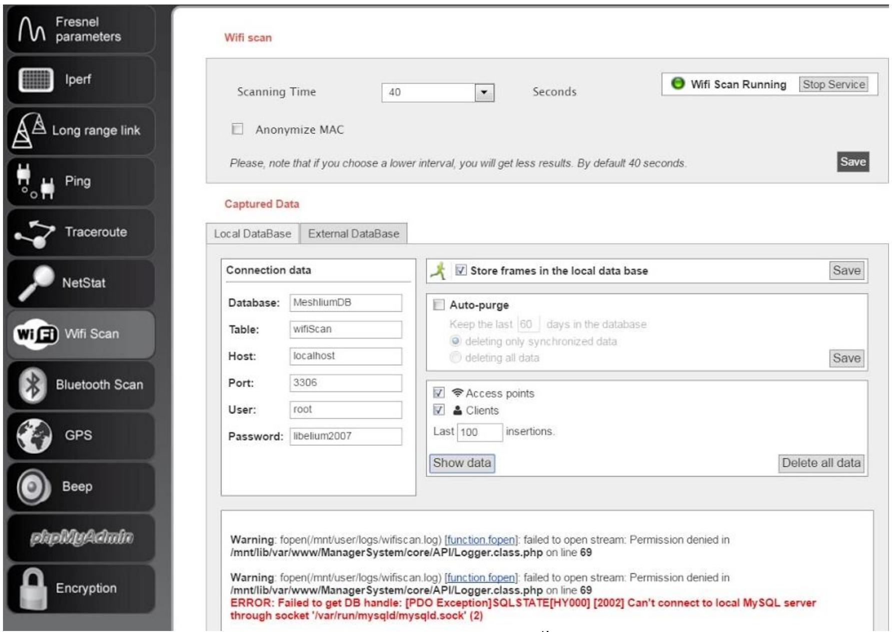


Figure 39, Third connection problem on 27th May 2016 for Meshlium3

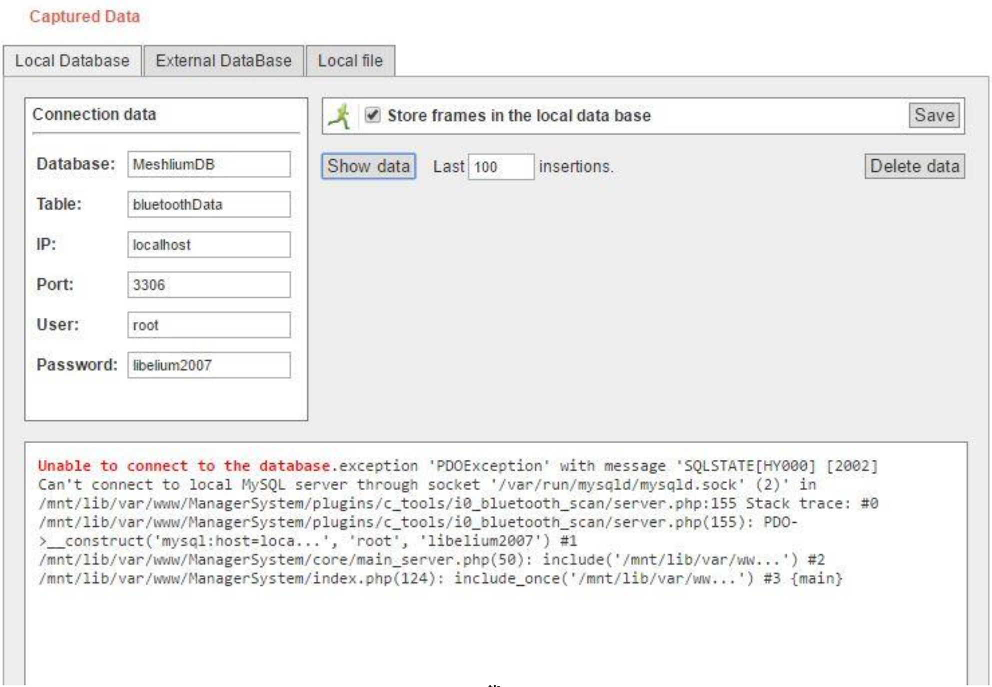
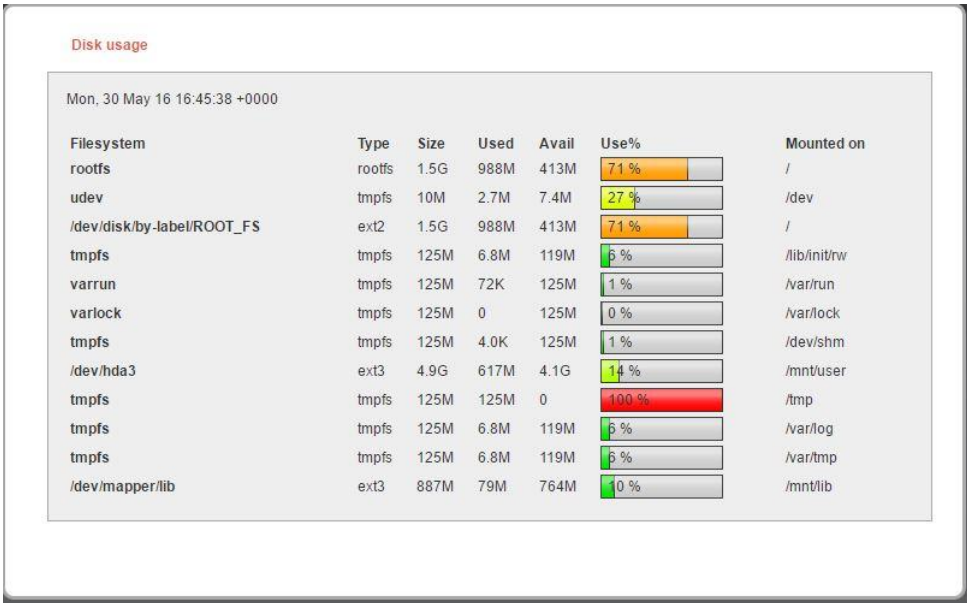


Figure 40, Fourth connection problem on 27th May 2016 for Meshlium3 by the use of SQL query

Finally, on 30th May 2016 we managed to connect to the sensor and after a research into the disk usage we noted that one disk called "tmpfs" was full and we were not able again to download again the data ***(Figure 41)***. Thus, we assume this should be the reason why the sensor did not store the tracking devices during the weekend (28th to 29th May 2016). However, after a restart of the sensor the relative disk was 1% filled and we managed to get the stored data till 27th may 2016 where it was the last record.


Disk usage

Mon, 30 May 16 16:45:38 +0000

Filesystem	Type	Size	Used	Avail	Use%	Mounted on
rootfs	rootfs	1.5G	988M	413M	71 %	/
udev	tmpfs	10M	2.7M	7.4M	27 %	/dev
/dev/disk/by-label/ROOT_FS	ext2	1.5G	988M	413M	71 %	/
tmpfs	tmpfs	125M	6.8M	119M	6 %	/lib/init/rw
varrun	tmpfs	125M	72K	125M	1 %	/var/run
varlock	tmpfs	125M	0	125M	0 %	/var/lock
tmpfs	tmpfs	125M	4.0K	125M	1 %	/dev/shm
/dev/hda3	ext3	4.9G	617M	4.1G	14 %	/mnt/user
tmpfs	tmpfs	125M	125M	0	100 %	/tmp
tmpfs	tmpfs	125M	6.8M	119M	6 %	/var/log
tmpfs	tmpfs	125M	6.8M	119M	6 %	/var/tmp
/dev/mapper/lib	ext3	887M	79M	764M	10 %	/mnt/lib

Figure 41, Visualization of the disk usage of the sensor and the full "tmpfs" disk

Based on all the previous, it is clear that the time error on some of the sensors affects significantly the analysis procedure. Despite the fact that it will be possible to have the tracking of the devices per day, the relative order of movement in the area will be wrong due to the time errors. In order to solve this problem and after the support of Wilko Quak, we use the Putty software in order to connect to the sensors and create a log file, by the use of a script, where we store per minute two different kind of times. The first is the system time while the second one is the hardware clock. By this way, we are able to compute the difference between the wrong and correct time of the device per minute and thus explore the option of a systematic error ***(Figure 42 and Figure 43)***.

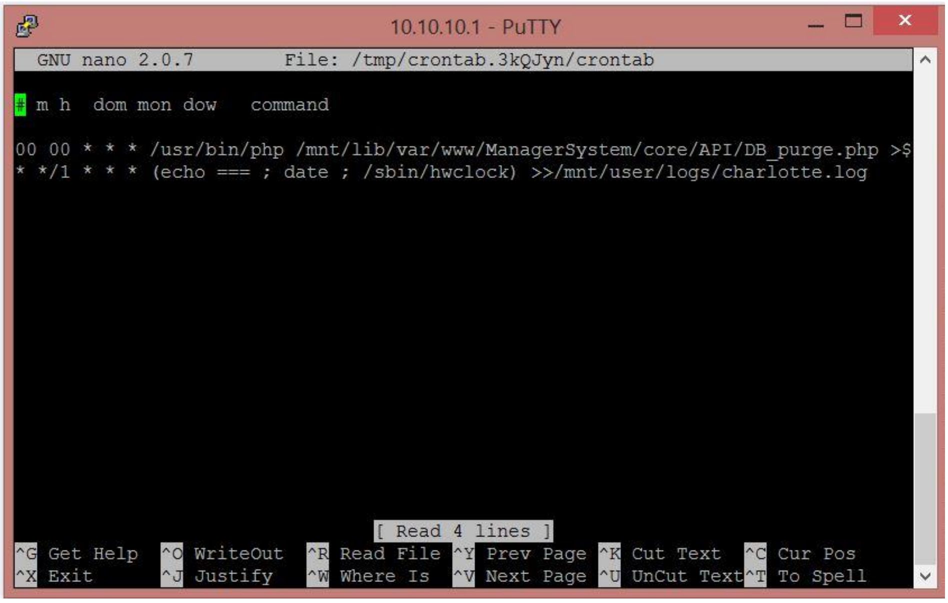


Figure 42, The script based on which we store per minute the system and hardware clock

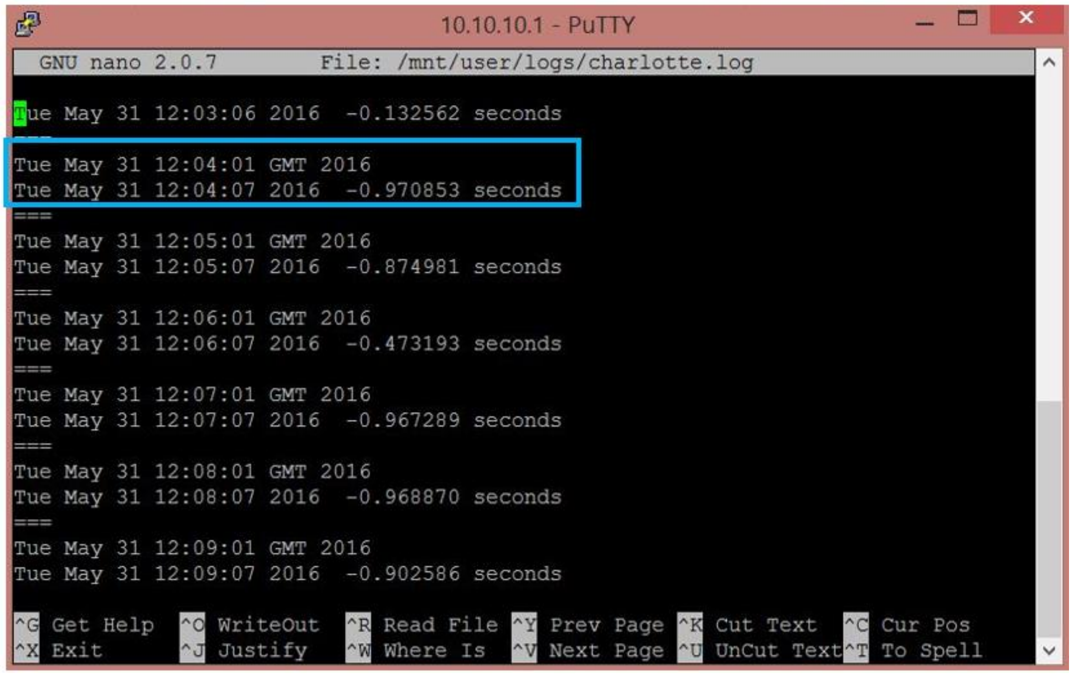


Figure 43, Example of the log file where we compute and store the time error

On 3rd June 2016 we got the last part of data for the period 31st May 2016 to 3rd June 2016 as well as the relative data from the log file about the two different kinds of times for sensors 3 and 309. About a computation of the time differences between the system and the hardware times, we identified that there was a linear systematic error on the system clock, as it is illustrated in ***Figure 44***.

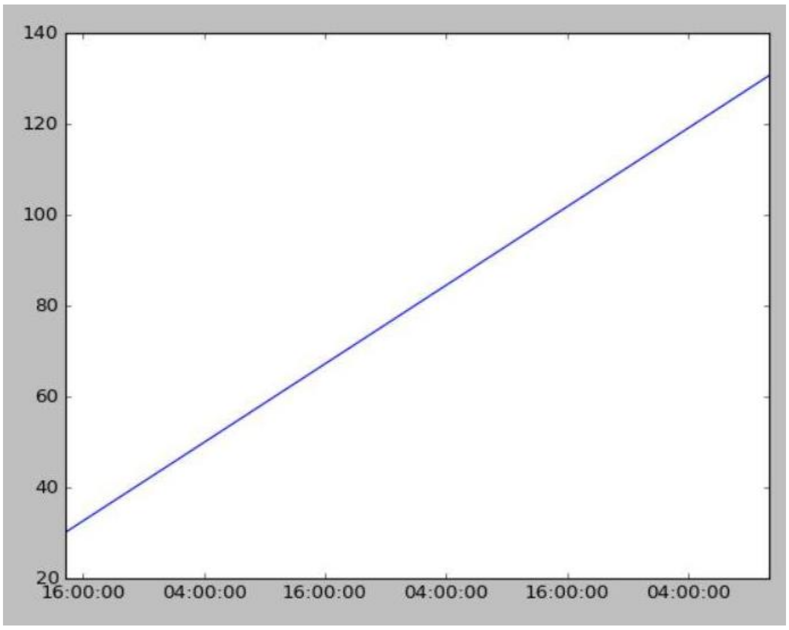
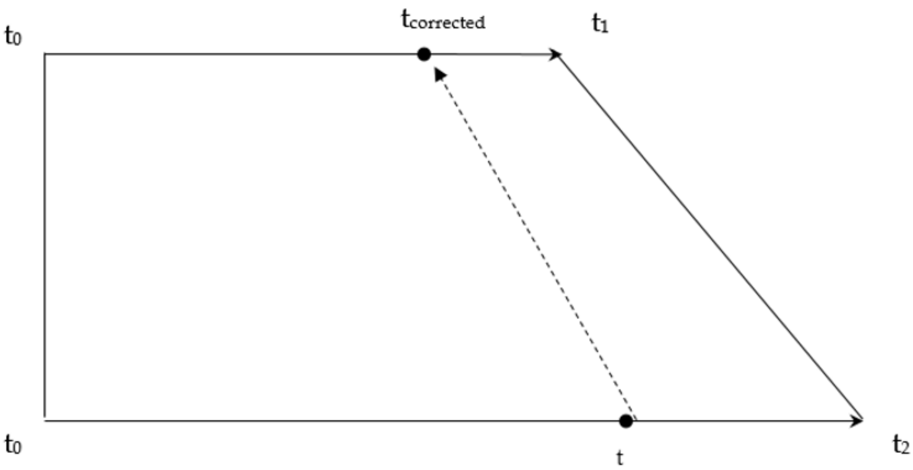


Figure 44, Visualization of time difference between system and hardware clocks for sensor 309. Vertical and horizontal axes illustrate the time difference in minutes and the timestamp of the tracking period respectively.

Based on this linear error, we tried to identify the relative equation in order to correct each time ***t*** to the correct time ***t***$\boldsymbol{t_{corrected}}$. The following figure shows the procedure for this correction ***(Figure 45).***



$$\frac{t_c - t_0}{t_1 - t_0} = \frac{t - t_0}{t_2 - t_0} \Rightarrow \boxed{t_c = t_0 + (t - t_0)\frac{t_1 - t_0}{t_2 - t_0}}$$

$$\boxed{t_c = t_0 + (t - t_0) * \text{factor}}$$

Figure 45, Identification of necessary formula to correct each time t to t corrected. The initial time t0 are the same as we set the correct time when we installed the relative script. t1 and t2 represent the final time of the system and hardware clock respectively.

The fraction $(t_1$-$t_0)/(t_2$-$t_0)$ remains stable for the sensor and after calculations it was computed equal to 0.976562409. Applying this correction formula we managed to fix the time error problem for sensor 309 for the periods 30th May 2016 to 3rd June. However, when we applied it to the data of 20th May 2016 to 24th May 2016 it does not work with this value of factor but for value equal to 0.995662409.

Unlike sensor 309, sensor 3 did not have a linear way of time change but it works correct for a short period and afterwards it had a linear rhythm of change ***(Figure 46).*** Based on that, it is impossible to create a formula to correct the data records for this sensor about the previous periods. However, taking advantage of the log file with the relative time differences we managed to fix the record time for the last period 30th May 2016 to 3rd June.

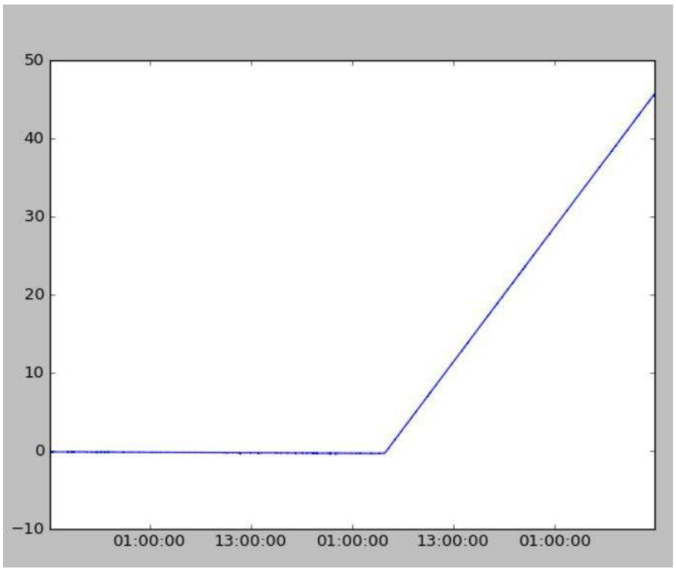


Figure 46, Visualization of time difference between system and hardware clocks for sensor 3. Vertical and horizontal axes illustrate the time difference in minutes and the timestamp of the tracking period respectively.

8. Data Validation and Correlation

A very important part of the project is the data validation and the computation of correlation level between the different data sources. This step will help significantly to determine the strength of the relationship between the various data and also to act as an indicator about how reliable can be the final outcomes.

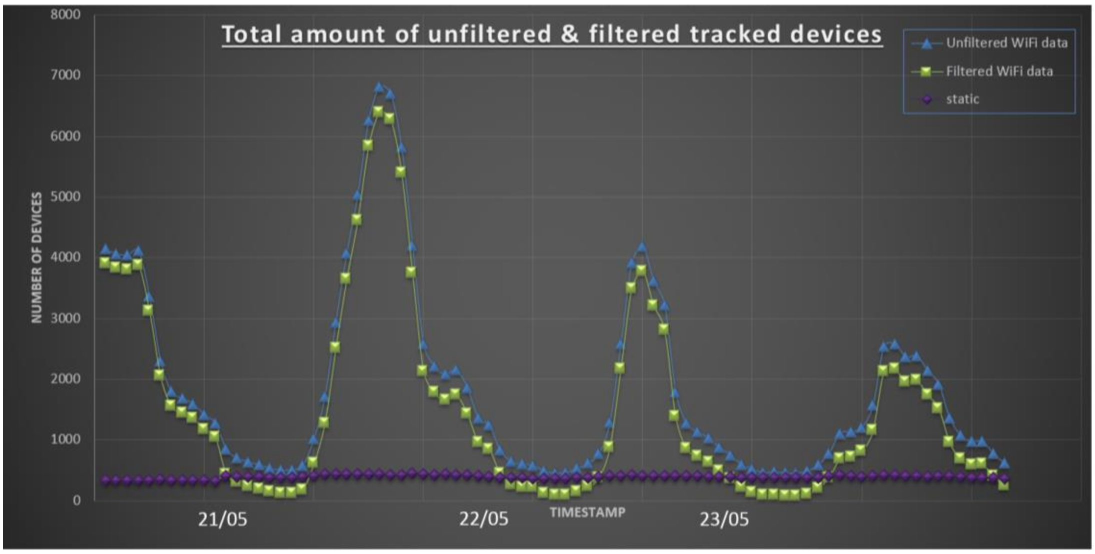


Figure 47, Total amount per hour of unfiltered tracked devices as well as the relative ones after the removing of static devices

Figure 47 illustrates the total amount per hour of unfiltered tracked devices in the research area as well as the relative amount after the removing of static devices. As we expected, there is a close relationship between the initial dataset and the filtered one. Also, as it is showed in the figure, the total amount of static devices remain stable comparing to the other two datasets.

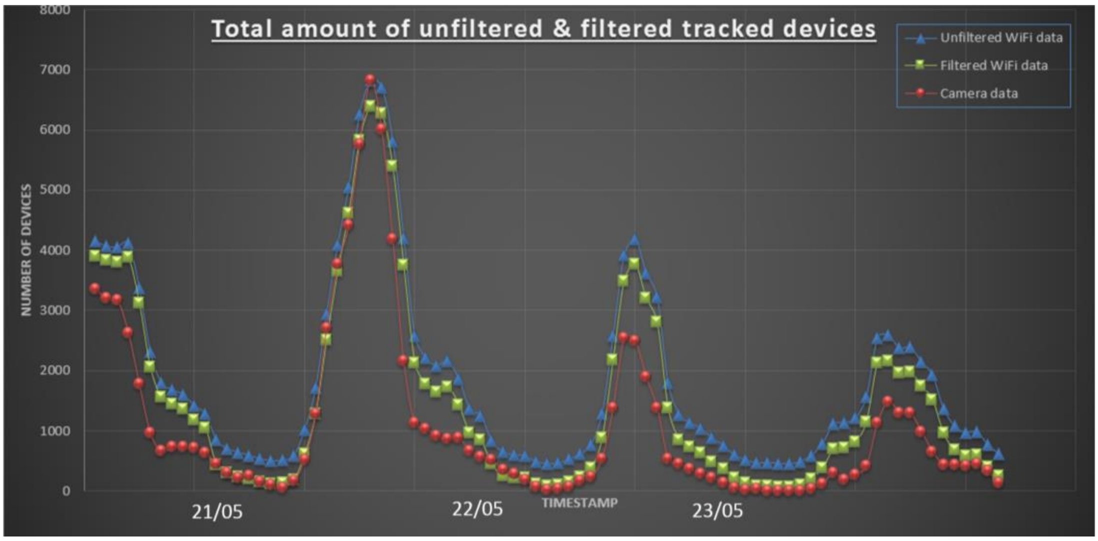


Figure 48, Total amount of unfiltered and filtered (from static devices) tracked devices as well as the total amount of pedestrians counted by the existed cameras system

In ***Figure 48*** we can see the relationship between the unfiltered, filtered as well as the counting dataset from the cameras system. Also in this case, the large coefficient between the datasets is clear during throughout the first period of observations in the city centre. However, it is very important to explain the fact that the total amount of pedestrians many times is smaller than the other two datasets, as at first sight it can be characterized as unexpected. As it is illustrated in the following figure (***Figure 49),*** the range of the existed cameras in the research area is quite small and only parallel to the street. On the other hand, sensors have a bigger range of cover and their placement in the corner of the buildings allows the tracking of devices which exist to the one or the other street. Based on that, we can interpret this relationship of the datasets.

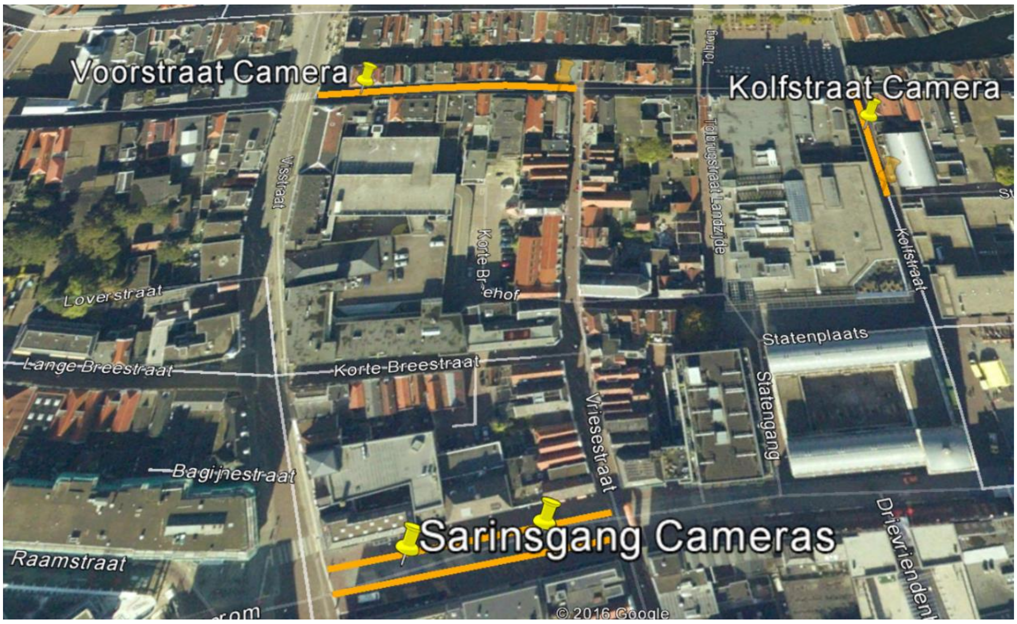


Figure 49, Range of the existed cameras in the research area (with orange line)

Due to the high interrelation, we tried to identify the correlate equation between these three datasets. In order to do it we applied the regression method to each pair of datasets and based on the outcome and the statistic results we defined the relative equation and the coefficient of determination. The following figure (***Figure 50)*** proves the linear relationship between the datasets, visualizing their combination per hour, the equation which represent their correlation as well as the statistic index of R^2.

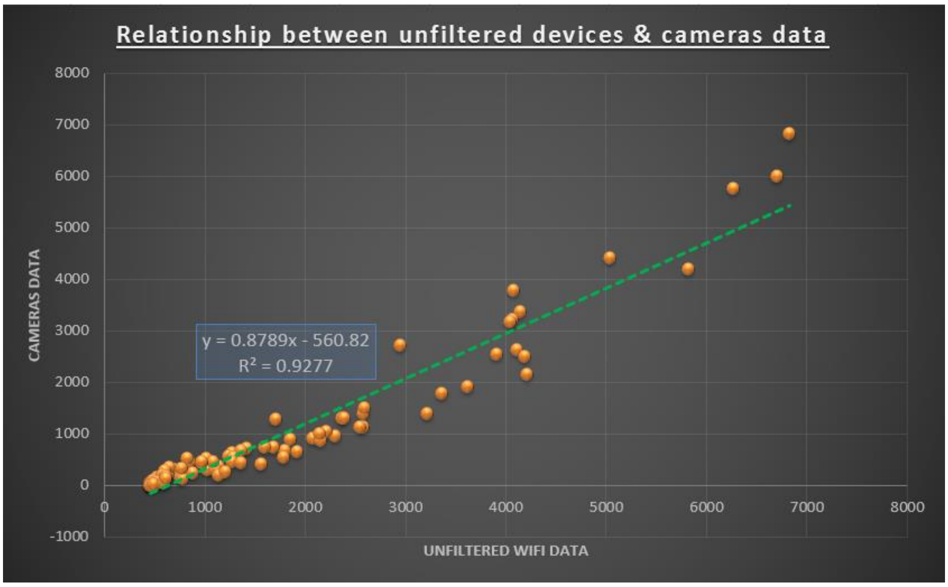


Figure 50, Correlation between unfiltered devices and number of pedestrians

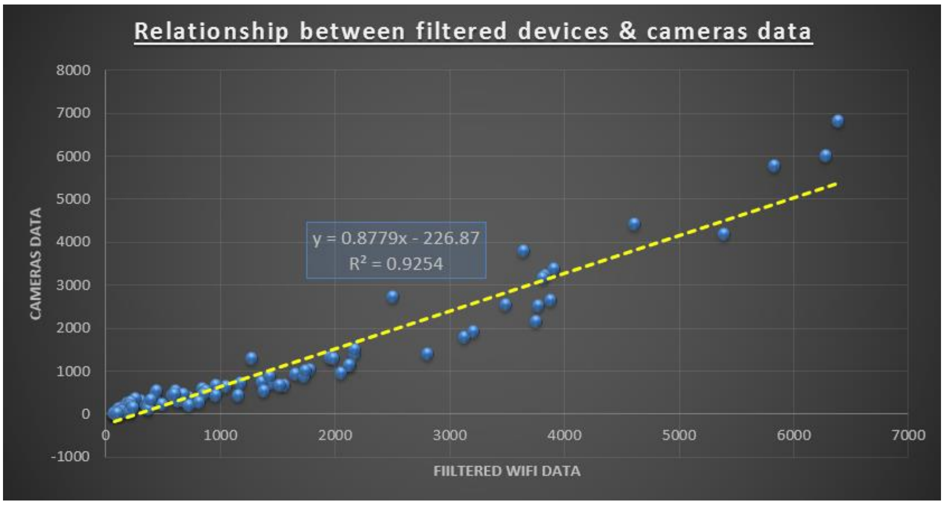


Figure 51, Correlation between filtered devices and number of pedestrians

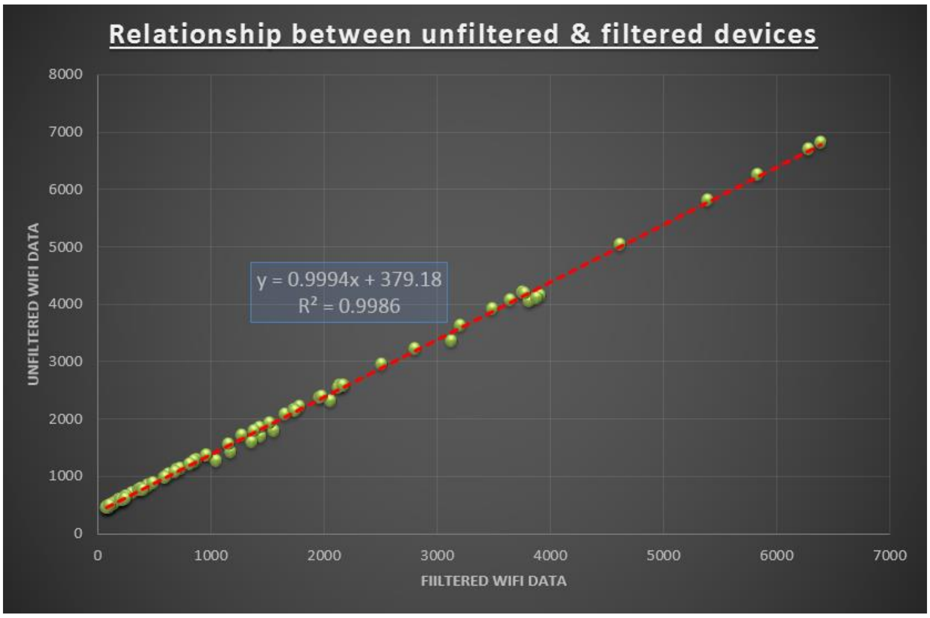


Figure 52, Correlation between unfiltered and filtered devices

As it is illustrated on the above ***Figure 51*** and ***Figure 52***, the calculated linear equation, which connect the datasets, is fixed very well, having a very high value of the R^2 index. Furthermore, the X variable about the relationship between the number of pedestrians and filtered dataset is almost equal to 1, matching very well to the statistic outcomes from the questionnaire. Finally, the high correlation between the filtered devices and the total amount of people counted by the cameras system ***(Figure 52)*** can lead to the conclusion that the final outcomes, based on sensors system, can be characterized as reliable with respect to the general movement behavior of people.

9. Results

9.1. Questionnaire

The responses of the two surveys conducted, one on a Friday which was part of the festival and also the traditional market day, and the second Friday was a regular, market day reflect the following: As expected, during a festival, there were more outside visitors than locals who rarely came to the city center, stayed anywhere from three to 6 hours in town, and preferred to take different routes as they walked through town. They also carried at least one Wi-Fi device. On a regular market-Friday, most visitors are locals who come every day to every other day, stay from one to two hours, do not often carry a Wi-Fi device with them and tend to use the same route when coming in and out of town.

9.2. Camera Data

Based on cameras placed at three out of the four streets within the research area, flow and density patterns of pedestrians were identified. Given that the view sheds of cameras are limited to the continuous street segments of their placement, their counts cannot be said to be the same for the rest of the given street because of intersecting side streets, but the relative differences between the streets can still be noted.

Voorstraat consistently had the highest daily flow of pedestrians followed by Kolfstraat and then Sarisgang even on Friday-market days (on Sarisgang). On average, Voorstraat had about 7.5K pedestrians per day during a regular week, and its flow increased by about 2K or 27% during the festival weekend. In terms of density, Kolfstraat had the highest density on a daily basis indicating that this street gets the most crowded at peak shopping time. Perhaps,because it is narrower than Voorstraat.

Voorstraat had at least 50% of all of the pedestrian traffic when looking at the aggregated flow of the three streets for each of a week; maybe because of the assorted restaurants, cafes and small shops coupled with a street and that it is not as narrow as Kolfstraat.

Friday, Saturday and Sundays had the highest flow and density in all three street segments when compared to the other weekdays. When compared with the festival weekend, the festival brought an increase of visitors to Voorstraat street of about 30% on Friday and Saturday, and almost doubled on Sunday.

The highest flow-hour of the three streets is around 2:00PM including the market days on Friday but excluding late-night shopping Thursdays. Thursdays which are the late-night shopping days have different peak hours of flow and density. These are around 1:00PM and 7:00PM with the highest flow and density at 7:00PM indicating that late shopping night was a popular day for combining dining or lunch and shopping in a comfortable and speedier way given that both the flow and density during this period seems to be consistently less than Saturday's.

Because the cameras do not cover a continuously the four streets that loop the research area, one cannot discern if visitors walk around the city center in a loop or not solely from camera data. Yet, the direction in which most people travelled along the three street segments can be observed: the flow on Sarisgang street was almost twice as much going towards Statenplein than away from this city square. As elaborated in the Directional and Hourly Flow Patterns section, the location of public transport access points, popular shops and of the open market have a significant influence on the direction in which visitors go. An interesting observation is that the flow on the segment monitored by the camera on Voorstraat was higher in the outward direction even though the market and popular big shops xare located in the opposite direction, near Statenplein. Even on market day, the outflow on Voorstraat is higher; perhaps indicating a stronger preference of the mix of small shops and places to sit down, eat and drink than the streets were mostly shops are present. In general, on a daily basis, most people went to towards Statenplein' via Sarisgang and Kolfstraat, and most went away from it via Voorstraat.

9.3. Sensor Data

9.3.1. Sensor Data Patterns

After the filtering part of the data, various patterns were investigated by the use of sql queries, graphs and charts. This chapter aim s to present the main outcomes of the research and the more frequently used patterns which were identified. Thus, the derived patterns are only the more interested ones and do not represent the full analysis of the data. Detailed graphs and tables about the whole set of outcomes and patterns can be viewed in the Appendix Sensor Data Patterns.

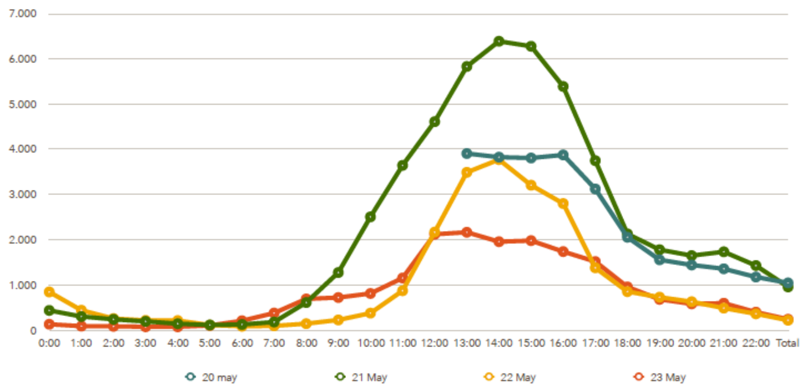


Figure 53, Different days

During the analysis part, moving flows between sensors 1, 2, 3, and 309 were computed per hour. After these computations, data were visualized in order to understand easier the flows of the different streets. The different cases of visualization are descried in the flowing flowchart (***Figure 54)***.

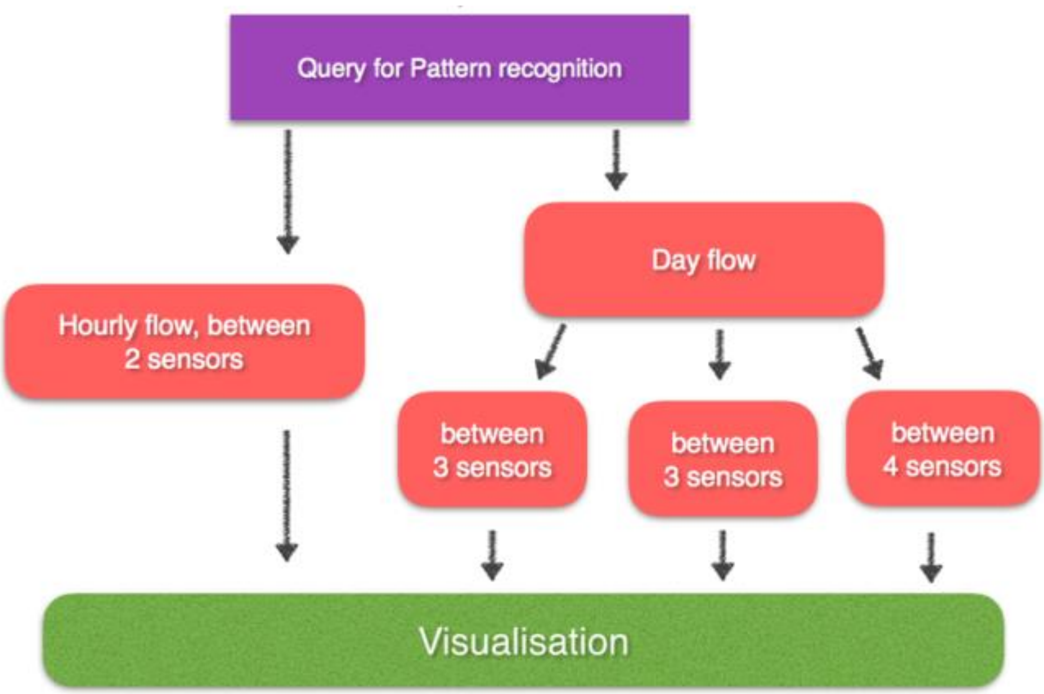


Figure 54, The data analysis flowchart

In this chapter we will refer to the sensor names as they are shown on the sensor map (**Figure 55**).

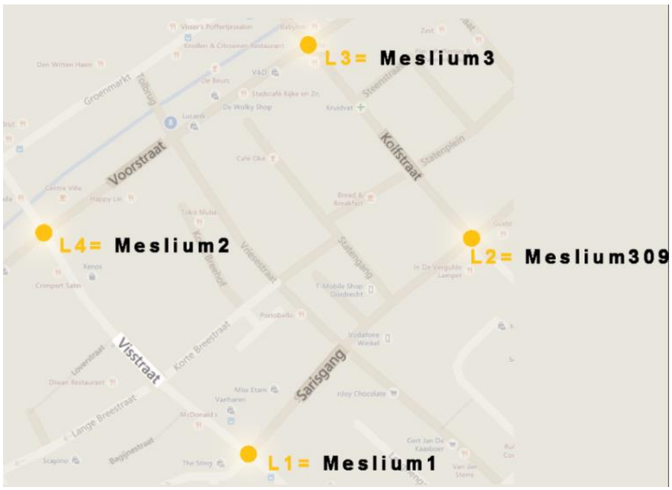


Figure 55, Overview of the research area and the place where each sensor was located

In this chapter all patterns between two sensors in every hour will be shown. These include the devices that are seen by two or more sensors per day.

Figure 56 illustrates the daily flow per hour for the period 20th to 23rd May 2016 (The data collection was started at the middle of the 20th May 2016). A high flow can be identified every day during lunch time as well as a higher amount of pedestrians on Saturday 21st May 2016. ***Figure 57*** visualizes the hourly flows in the research area for the day of 22nd May 2016. As we can see, Visstraat and Voorstraat streets are those with the most concentrated flows and especially during to the “hot” period of lunch time.

Figure 56, Hourly flows for 22nd May 2016

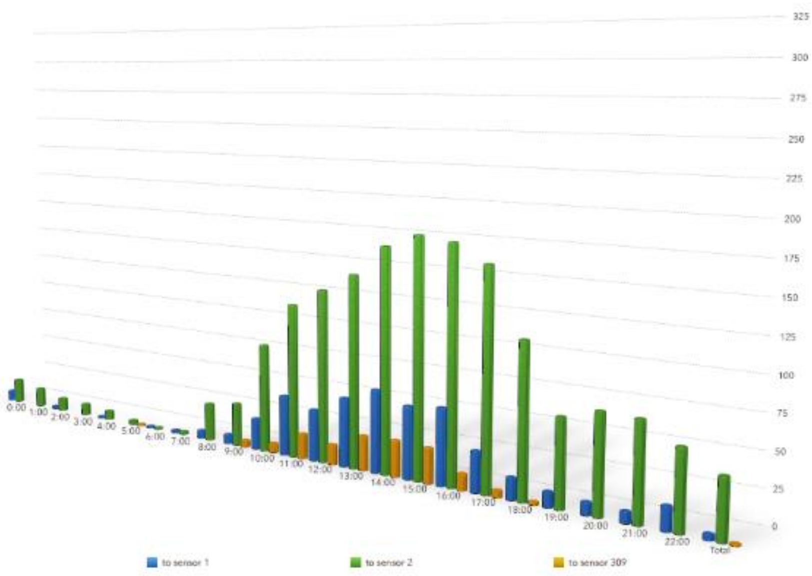


Figure 57, Movements from sensor 3 to the other sensors on 21st May 2016.

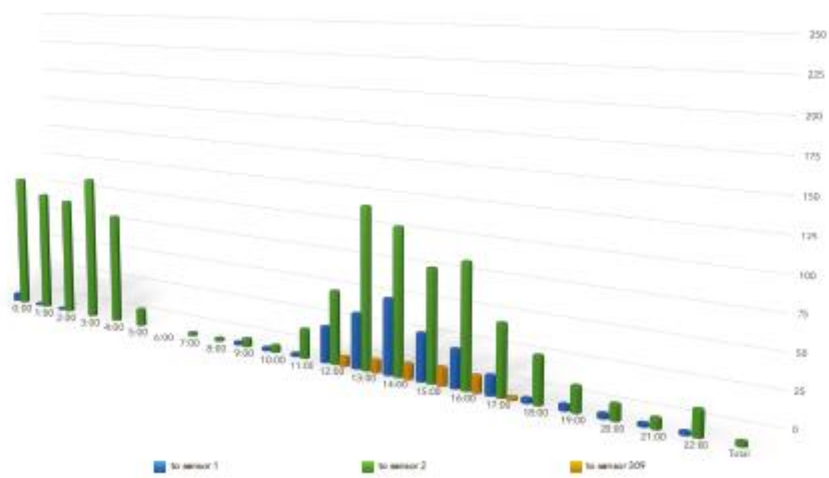
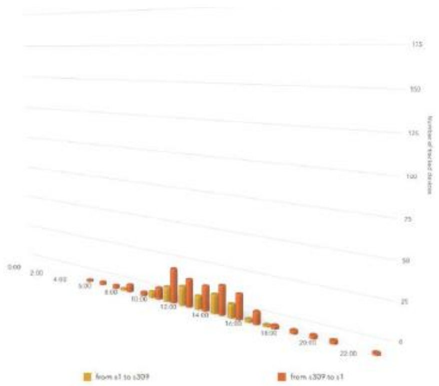
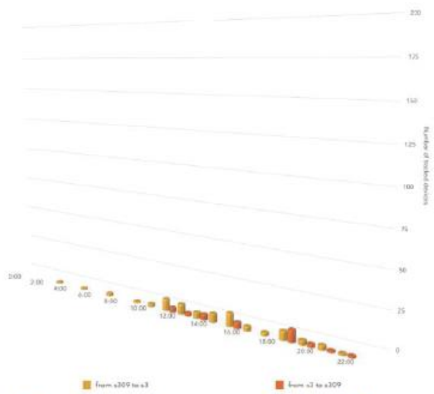


Figure 58, Movements from sensor 3 to the other sensors on 22nd May 2016.

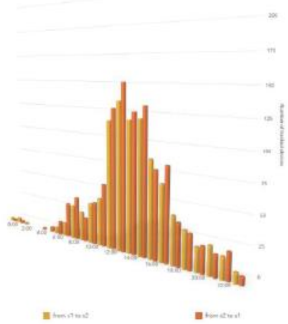
Figures 57 and Figure 58 show the total amount of hourly movements from sensor 3 to the others for 21 and 22 of May respectively. As it is clear, unlike Friday night 21st May 2016, on Saturday night beginning of 22nd May there is a significant increase of the total amount of tracked devices only between sensors 3 and 2 (Voorstraat). This increase can be explained by the fact that people use on Saturday night to go out at bars till the first hours of the Sunday morning. Furthermore, as it is illustrated in the following charts ***(Figure 59),*** an unexpected low flow can be identified on Sarisgang and Kolfstaat compared to the Voorstraat and Visstraat.



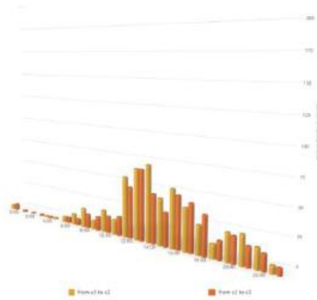
Total amount of movements in Sarisgang on 23/05/2016



Total amount of movements in Kolfstraat on 23/05/2016



Total amount of movements in Visstraat on 23/05/2016



Total amount of movements in on 23/05/2016Voorstraat

Figure 59, Total amount of movements for the four main streets of the research areas on 23rd May 2016

As a separate step for the analysis part, we divide the datasets into three main categories:

- Devices which were tracked by only two sensors during the day.
- Devices which were tracked by only three sensors during the day.
- Devices which were tracked by all the four sensors during the day.

For each of the above subcategories, we explored all the possible combinations during the whole day. As the duration of the pattern could be completed in a 24-hours period, a time threshold of eight hours was applied. Finally, taking into account the ability of the observation system to record the MAC address of each device, we computed the relative percentage of unique movements for each pattern, as it is illustrated in the ***Table 8***.

Patterns	Total Devices	Total Unique Devices	Percentage of Unique Devices	Percentage from the total amount of devices
Patterns 2	3617	3056	84.5	91.4%
Patterns 3	325	278	85.5	8.2%
Patterns 4	14	14	100.0	0.4%
Total	**3956**	**3348**	**84.6**	**100.0%**

Table 8, Relationship between flows and unique devices

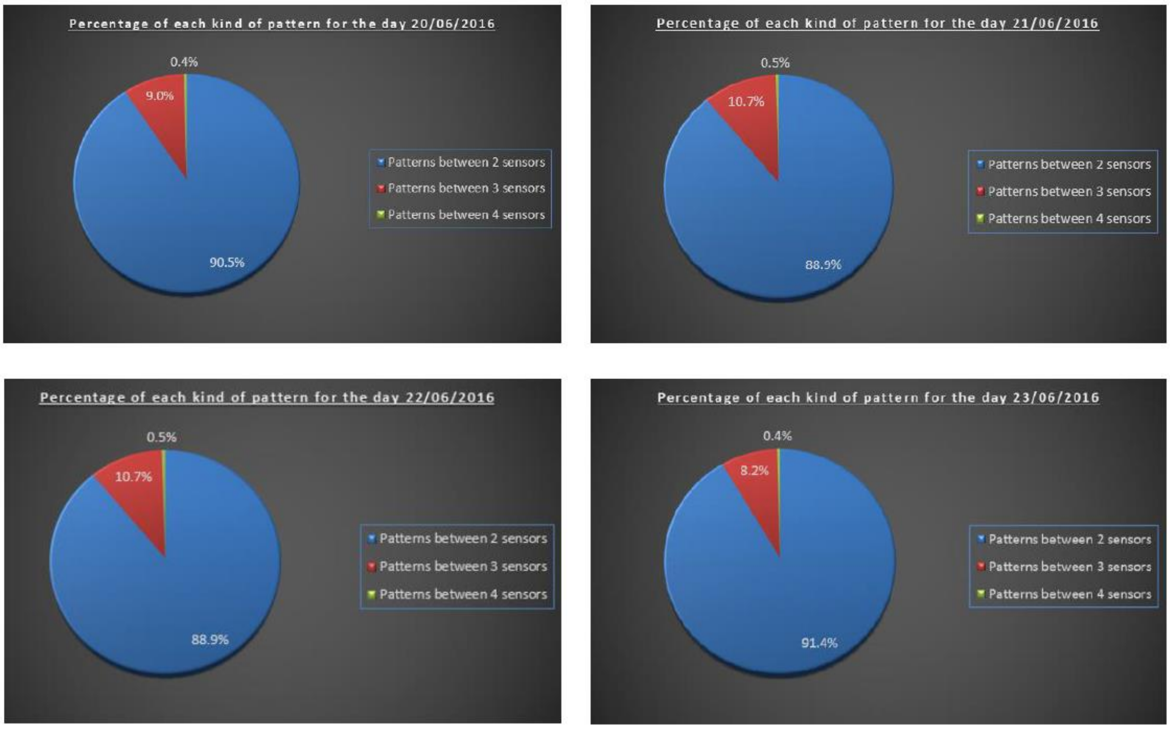


Figure 60, Total percentage for each kind of pattern per day

As it is clear from the above figure ***(Figure 60),*** the majority of the pedestrian movements (≈90%) belong in the category of patterns between two sensors while a percentage of around 10% contains patterns between three sensors.

Route	Devices	In a period of 8 hours	Percentage of this kind	Percentage of total
2-1	1037	1028	28.7%	26.2%
1-2	996	983	27.5%	25.2%
3-2	488	485	13.5%	12.3%
2-3	452	443	12.5%	11.4%
309-1	166	166	4.6%	4.2%
3-1	159	154	4.4%	4.0%
1-3	146	143	4.0%	3.7%
1-309	89	85	2.5%	2.2%
309-3	40	40	1.1%	1.0%
309-2	18	18	0.5%	0.5%
3-309	18	18	0.5%	0.5%
2-309	8	7	0.2%	0.2%
Total	**3617**	**3570 (98.7%)**	**100.0%**	**91.4%**

Table 9, Identified patterns of combination of two sensors ordered by their frequency for 23/05/2016

In ***Tables 9, 10,and 11*** we can see the computation of the relative percentage of each pattern from the total amount of movements as well as from the category that it belongs. As it is clear from table 2, Voorstraat and Visstraat are again the most frequently used streets unlike

Sarisgang and Kolfstaat which comes with quite lower percentage, verifying the outcome from ***Figure 61.***

Route	Devices	In a period of 8 hours	Percentage of this kind	Percentage of total
1-309-3-2	3	3	21.4%	0.1%
2-309-1-3	2	2	14.3%	0.1%
3-2-1-309	2	2	14.3%	0.1%
309-1-3-2	2	2	14.3%	0.1%
1-2-3-309	1	1	7.1%	0.0%
1-2-309-3	1	1	7.1%	0.0%
1-309-2-3	1	1	7.1%	0.0%
2-3-309-1	1	1	7.1%	0.0%
2-309-3-1	1	0	7.1%	0.0%
Total	**14**	**13 (92.9%)**	**100.0%**	**0.4%**

Table 10, Identified patterns of combination of three sensors ordered by their frequency for 23/05/2016

Figure 61, Patterns of devices that are seen by 4 sensors Order by appearance on 23 may 2016

There are a lot of combinations possible for devices are seen by four sensors. In ***Figure 62*** are shown the patterns that are used on 23 of May. In all days we can see that the route 1-309-3-2 is the most frequent used pattern, unlike the outcomes from the combination of two sensors. However, it is important to refer that this pattern can be the most frequently used from this kind of sensor combinations but it consists only the 0.1% of the total amount of movements for the whole day.

Route	Devices	In a period of 8 hours	Percentage of this kind	Percentage of total
3-2-1	63	62	19.4%	1.6%
1-2-3	45	42	13.8%	1.1%
2-3-1	43	40	13.2%	1.1%
1-3-2	34	34	10.5%	0.9%
2-1-3	27	23	8.3%	0.7%
309-2-1	14	14	4.3%	0.4%
309-3-1	14	14	4.3%	0.4%
3-1-2	13	12	4.0%	0.3%
309-3-2	9	9	2.8%	0.2%
2-309-1	8	8	2.5%	0.2%
1-309-3	7	7	2.2%	0.2%
2-1-309	6	6	1.8%	0.2%
2-3-309	6	6	1.8%	0.2%
309-1-3	6	6	1.8%	0.2%
1-2-309	5	5	1.5%	0.1%
3-309-1	5	5	1.5%	0.1%
309-1-2	5	5	1.5%	0.1%
1-3-309	3	3	0.9%	0.1%
2-309-3	3	3	0.9%	0.1%
3-2-309	3	3	0.9%	0.1%
309-2-3	3	3	0.9%	0.1%
1-309-2	1	1	0.3%	0.0%
3-1-309	1	1	0.3%	0.0%
3-309-2	1	1	0.3%	0.0%
Total	**325**	**313 (96.3%)**	**100.0%**	**8.2%**

Table 11, Identified patterns of combination of four sensors ordered by their frequency for 23/05/2016

Figure 62, Visualization of the most frequently used patterns of combination of 3 sensors

As we can see in ***Figure 63***, the patterns of combination of 3 sensors which were used more during the research period, are the movements 3→2→1 ***(Figure 62.1)*** and 1→2→3 ***(Figure 62.3)***. These patterns can support the previous referred outcomes. Furthermore, the direct connection, also in these combinations, between sensor 3 and 1, proves the preference of pedestrians for the intermediate streets or the wrong use of sensors 309.

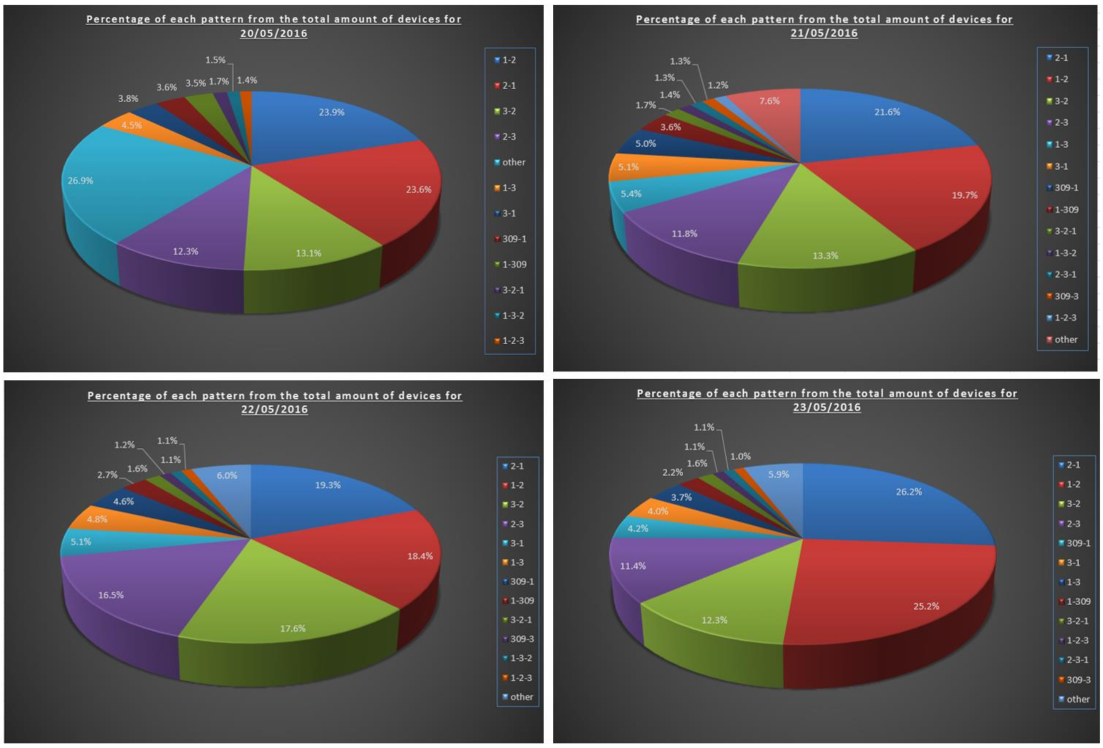


Figure 63, The highest frequently used patterns from all the categories for each day

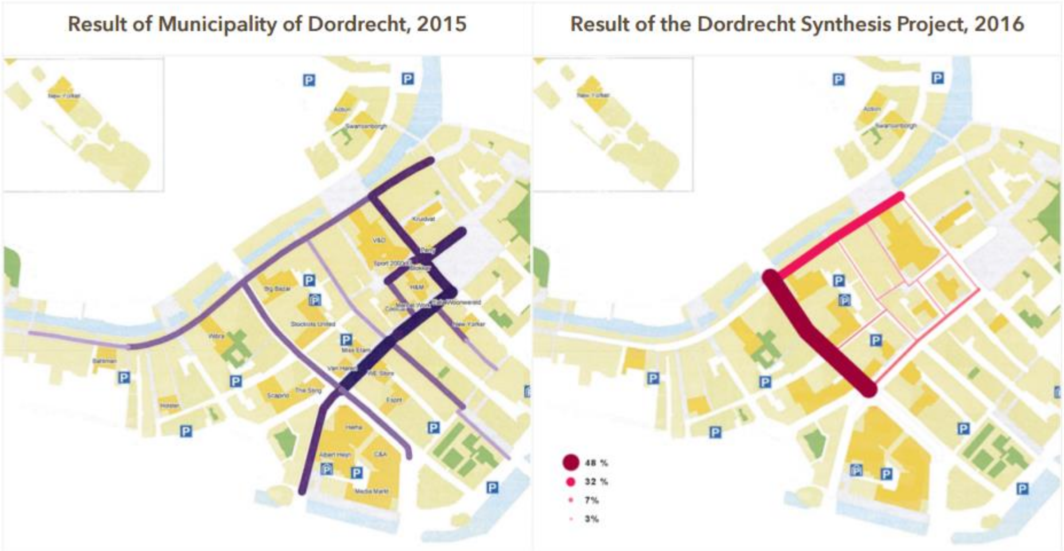


Figure 64, Overview of the previous research outcomes (left) and outcomes of this research about the most frequently used streets based on WI-FI tracking data

Despite the expectations based on the previous research, the outcome of this research does not match with the existed patterns. As it is clear from the overall charts for each day ***(Figure 63)*** as well as from the comparison in ***Figure 64***, the street which connects the sensors 1 and 2 (Visstraat) is continuously the most frequently used, especially during the week days, as there is a short reduction of the relative amount of devices during the weekend. Almost the half of the total amount of devices was tracked at Visstraat while Voorstraat (connection of sensors 2 and 3) comes second with a total percentage of 34% of movements per day.

The movement between sensors 1 and 3 was continuously the third most frequently used pattern during the whole period of observations. As an effort to explain this pattern, two options can be referred:

- Due to the limited number of tracking devices, not all the streets of the research area were covered. Thus, it is possible pedestrians prefer to use some of the intermediate streets of the research area such as Vriesestraat.

- Despite the fact that the range of sensor 309 was checked during its placement, it is possible to have technical problems during the tracking procedure apart from the error about the record time. This problem is discussed in Chapter 6 data after filtering.

Maybe the second scenario, which was mentioned above, can explain also the fact that the amount of pedestrian movements between sensor 1 and sensor 309 (Sarisgang) was unexpectedly low throughout the observation period. However, it is significant to refer that apart from Friday 20th May 2016, there is notable difference between the amounts of people of the relative flows. Thus, the amount of people which walked from location 2 (sensor 309) to location 1 (sensor 1) is quite higher than the opposite.

10. Conclusions and Recommendations

10.1. Conclusions

This project investigated what kind of pedestrian movement patterns can be recognized by the use of Wi-Fi tracking sensors in the city centre of Dordrecht. In order to answer the main research question, we have to answer the research sub-questions:

- Is it possible to identify "hot" periods?
- Is it possible to identify changes per day?
- Do the users return by the same way?
- Is it possible to identify different patterns of tourists and daily visitors?

It was possible to identify the "hot" periods in the research area. Based on camera and sensor data, the peak was around lunch time, starting at noon until 3:00PM on Friday through Wednesday. On Thursdays, based only on camera data, there were two hot periods, one around 2:00PM and the second one around 7:00PM. This can be explained by the fact that Thursday is late night shopping day.

From the sensor data analysis, daily changes can be recognized. The highest amount of pedestrians was identified on Saturdays and the lowest, was on Mondays. The flow magnitude was influenced by the weather and opening hours of shops. The highest night time flow of Saturdays, trickles into Sunday especially on Voorstraat. Since the shops open later on Sundays, the flow start is delayed, as well as the relative flow reduction on Friday and Saturday evening. Based on camera data, when comparing a regular weekend with the festival weekend, the flow increased significantly. On Fridays and Saturdays, it increased by 30% and on Sundays it doubled. Finally, Visstraat had the highest flow consistently from all the other streets.

Based on the analysis of the questionnaire, only locals who are familiar with the area, return by the same way, whilst tourists and visitors tend to avoid taking the same routes. More than 50% of the tourists and visitors take different routes possibly because they are unfamiliar with the routes or would want to see other parts of the city centre.

It is possible to identify patterns of tourists and daily visitors using our survey data, which asked "*To which of the following categories do you belong to? Tourist or (non) Dordrecht resident?"* One notices that there were more (two thirds) outsiders than locals (one third) visiting the town during the first Friday. Further, the split between tourists--first time visitors or those who rarely come to Dordrecht was almost even (one third each). This could be because of the festival 'Dordt in Stoom'. The reduction of tourists is significant on the second Friday confirming the higher presence of tourists in the first Friday was mostly due to the festival.

To answer our main research question, based on this research, we can conclude that we can use the Wi-Fi tracking system as a smart way to identify pedestrian movements. This is supported through our ability to identify various pedestrian movement patterns, which differ hourly and daily, as well as 'hot' periods within the city centre of Dordrecht. They are mostly concentrated on the Visstraat and Voorstraat. There is a strong correlation between the sensor data and the camera data, so the Wi-Fi monitoring system can lead to precise assessment of the movement behavior of the pedestrians within the city center.

10.2. Recommendations

For further research and for the Municipality of Dordrecht we would recommend installing a free Wi-Fi network within the city centre of Dordrecht. This will make it possible to get a very good representative insight on the movement patterns of the city center. We strongly recommend capturing more data, because the results of the patterns get influenced by many factors, like the weather, the opening times of the shops and restaurants and organized festivals within the city centre. It will also give more data of different days to get insight of differences between all the days within a week, but also weeks within months and difference between the four seasons of a year etc.

In this project, not enough sensors were available to place in the city centre as desired. In an ideal scenario to properly detect the flow of visitors within the City Centre of Dordrecht with the appropriate street coverage, at least 18 sensors would be needed ***(Figure 65).*** The team had only four sensors to work with and made the best use of them.

Figure 65, Ideal sensor amount and placement

In addition to the very limited amount of sensors, half of them had hardware issues and yielded data with skewed time making the data processing more complicated and causing the loss of a few days' worth of data. As a result, the data was not sufficient as planned for tracking the overall people's movement due to the limited coverage of the research area. Fortunately, the surveillance camera data provided by the Municipality of Dordrecht brought in an additional data for the analysis to be done in the project.

The Meslium sensors also had technical limitations, reducing the amount of the data because they had time errors. The opportunity to choose different sensor manufacturers or fix the hardware of the current sensor is recommended. In any case, the same type of sensors should be used for the project, concurrently with the same calibration.

The time required to get the approval to place the sensors at the wanted locations took too long. This reduced the data collection period. Thus, bureaucracy issues regarding permits and device setup need to be addressed in advance or prior to the start of the project. This would save valuable time throughout the project and facilitate the data collection and execution of the project more efficient given its fixed duration and deadlines.

For the validation of the results we used the camera data which were placed on three locations within our research area; however, there was no camera on the Visstraat. Based on that, we recommend the placement of an additional camera on this street in order to improve validation if further research is to be done.

As further recommendations for better use of the sensors and for a better experience for future research, we recommend the following:

- Ideally the sensors are connected directly to the internet and therefore the data is stored directly into a server. This would reduce the amount of time for downloading and storing the data collected.
- Explicit testing to determine the actual sensor coverage based on the equipment model and manufacturer of smartphones.

The following observations may help in future decision making for the Municipality of the City of Dordrecht;

- From the business owner’s perspective, Voorstraat is the ideal street for more potential sales.
- From the shopper’s perspective, shopping or visiting these street segments at around 2:00PM is the busiest, so most crowded time. The most densed street at this time is Kolfstraat.
- Thursday late-night shopping presents a convenient and more comfortable visit experience. Based on the flow and density observations, and recalling the fundamental traffic flow diagram, avoiding the high density hour, around 2:00PM, would yield a speedier shopping experience. This seems to be applied on Thursdays.
- The data also supports the concept that having weekend festivals is an effective way to increase the flow of pedestrians and cyclists with the advantage of additional sales on Sundays from the almost double flow of people on this day due to the festival.

References

Cormack, A. (2013). Bins, MACs and Privacy Law | Jisc community. Retrieved May 13, 2016, from https://community.jisc.ac.uk/blogs/regulatory-developments/article/bins-macs-and-privacy-law

Dr. ir. van Loenen, B.; Prof. mr. de Jong, J. ; Mr. dr. ir. Zevenbergen, J. A. (2008). *Locating mobile devices*.

European Digital Rights. (2015). WiFi tracking and the ePrivacy Directive in Denmark. Retrieved May 13, 2016, from https://edri.org/wifi-tracking-eprivacy-directive-denmark/

European Parliament, Council of October 24, 1995. Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data, Official Journal L 281 0031 – 0050 (1995). OPOCE. Retrieved from http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:31995L0046

Graux, H. (2011). Open government data : reconciling PSI re-use rights and privacy concerns. In *European Public Sector Information Platform:*

Kulk, S., & Loenen, B. Van. (2012). Brave New Open Data World? *International Journal of Spatial Data Infrastructures Research*, *7*, 196–206. http://doi.org/10.2902/1725-0463.2012.07.art10

Libelium. (2016). Meshlium X treme Technical Guide.

Scassa, T. (2010). Geographical Information As “Personal Information”. *Oxford University Commonwealth Law Journal*, *10*(2), 185–214. http://doi.org/10.5235/147293410794895322

Musa, A., & Eriksson, J. (November, 2012). Tracking Unmodified Smartphones Using Wi-Fi Monitors. *10th ACM Conference on Embedded Network Sensor Systems*

van der Ham, M., Kalogianni, E., Lam, M., Sileryte, R., & Kaixuan, Z. (2014). *Rythm of the Campus.* Delft: Geomatics Synthesis Project

van der Spek, S. (2010). Tracking Tourists in Historic City Centres. *Information and Communication Technologies in Tourism.* Lugano: Springer

Zhang, J., Firooz, M., Patwari, N., & Kasera, S. K. (2008). Advancing Wireless Link Signature for Location Distinction. *14th ACM International Conference on Mobile Computing and Networking*

WikiBooks (2015). Fundamentals of Transportation/Traffic Flow - Wikibooks, open books for an open world. Retrieved June 4, 2016, from https://en.wikibooks.org/wiki/Fundamentals_of_Transportation/Traffic_Flow

Wikimedia Commons, (2007), Fundamental diagram of traffic flow, source: Ammoser, H. Retrieved June 8, 2016, from https://en.wikipedia.org/wiki/Fundamental_diagram_of_traffic_flow#/media/File:Fundamental_Diagram.PNG

Appendices

The following files are submitted with the report as appendices

1 Work Plan and Schedule
2 Project Schedule
3 Ghantt Chart
4 Rich Picture
5 Test Zero
6 Sensor Locations Options
7 Camera Data Visualisation
8 Sensor Data Visualisation Flows
9 Sensor Data Visualisation Patterns
10 SQL Queries
11 Python Scripts
12 Dordrecht Team Group Activity Log
13 Dordrecht Team Individual Time Log