UNIVERSITY OF TARTU
Faculty of Science and Technology
Institute of Computer Science
Computer Science Curriculum

Emma Belinda Semilarski

# Understanding Mobility Patterns through GPS Data

Bachelor's Thesis (9 ECTS)

Supervisor(s): Rahul Goel, MSc
Anto Aasa, PhD
Rajesh Sharma, PhD

Tartu 2023

## Understanding Mobility Patterns through GPS Data

**Abstract:**
The Global Positioning System (GPS) has enabled to collect locational data from humans. One way to use GPS data is for exploring mobility patterns. The aim of this bachelor thesis is to unravel the mobility patterns and identify points of interest to gain more insight of spatiotemporal patterns of humans in Tartumaa. To achieve this, we use exploratory analysis techniques such as data preprocessing, visualisation, and statistical tests. Outcomes of this thesis describe the temporal and spatial patterns of people moving in Tartu and the surrounding area, and identify the most visited places from the data. The findings have the potential to contribute to the field of human mobility analysis and can be useful for city officials and policy makers in developing efficient urban planning strategies.

**Keywords:**
GPS data, Mobility patterns, Estonia, Descriptive analysis, Data analytics

**CERCS:**: P170 Computer science, numerical analysis, systems, control

## Liikumismustrite mõistmine GPS-andmete kaudu

**Lühikokkuvõte:**
Globaalne positsioneerimissüsteem (GPS) on võimaldanud koguda inimeste asukohaandmeid. Üks võimalik GPS-andmete kasutusviis on liikumismustrite uurimine. Käesoleva bakalaureusetöö eesmärk on liikumismustrite avastamine ja huvipunktide tuvastamine, et saada ülevaadet inimeste ajalis-ruumilistest mustritest Tartumaal. Eesmärgi saavutamiseks kasutatakse uurimusliku analüüsi võtteid, nagu andmete eeltöötlus, visualiseerimine ja statistilised testid. Bakalaureusetöö tulemused kirjeldavad Tartus ja selle lähiümbruses liikuvate inimeste ajalis- ja ruumilisi mustreid ning toovad andmete põhjal välja enim külastatud kohad. Saadud tulemused panustavad inimeste liikuvuse analüüsi valdkonda ning neid saavad ära kasutada linnaametnikud ja poliitikakujundajad linnaplaneerimise strateegiate väljatöötamisel.

**Võtmesõnad:**
GPS-andmed, Liikuvusmustrid, Eesti, Kirjeldav analüüs, Andmeanalüütika

**CERCS:**P170 Computer science, numerical analysis, systems, control

# Contents

# 1 Introduction

The widespread use of Global Positioning System (GPS) technology has provided great amounts of data over the recent years. GPS can offer valuable information about various aspects of human behaviour. This thesis examines mobility patterns across different days of the week and hours of day using an exploratory analysis of GPS data. By unravelling the spatiotemporal dynamics of human mobility, this research aims to shed light on the underlying patterns and preferences that govern our daily movements.

Several researches have been conducted in the past exploring the use of GPS data in discovering mobility patterns. Some researchers have attempted to identify frequently visited locations by individuals and analyse the routes taken to reach those locations [1], [2]. Others have focused on extracting temporal mobility patterns from GPS data [3], [4]. Many researchers have examined the factors that influence human mobility [5], [6]. Additionally, a study exploring the spatiotemporal behaviour of students and academics, using data from the same study area as this thesis, was also conducted a few years ago [7].

The main goal of this thesis is to examine movement across different days of the week and times of the day in order to identify the most frequently visited locations, specially in Tartu, second largest city of Estonia. To achieve this, we conduct an exploratory analysis using GPS data, provided by the Mobility Lab at the University of Tartu. The dataset consisted of roughly 23.04 Million records from approximately 200 individuals over the course of a little over two months spanning from February 1, 2017, till April 5, 2017. The dataset has 13 different features, some of them being user ID, GPS coordinates, times, speed, altitude, bearing, and accuracy. Considering that the data has both temporal and spatial information of users, we formulated two research questions as follows:

1. What are the temporal patterns of human mobility across different days of the week and times of the day?
2. Which locations are the most frequently visited by individuals and when are they visited the most?

This thesis leverages advanced exploratory analysis techniques including data preprocessing, visualisation, and statistical tests. In this work, we adopt a data-driven approach and extract meaningful information from the GPS data such as discovering individuals' mobility patterns. The findings of this thesis give an overview of people's temporal habits in Tartumaa and reveal their most frequently visited places in Tartu. The major results of this study are as follows:

1. We find the daily and hourly patterns of mobility, including peak times and slower periods, and propose the factors that might influence them.

2. We detect top visited places and offer reasoning for visitations, as well as discover the visitation patterns of different locations.

The findings of this thesis have the potential to contribute to the field of human mobility analysis. By providing a detailed understanding of movement patterns across different temporal contexts and identifying the most frequently visited locations, the analysis of the thesis can potentially be used by city officials and policy makers in Tartu to develop more efficient urban planning. Information about citizens' mobility habits, including peak hours and hotspots within the city, could lead to identifying areas that experience high movement frequency, determining optimal locations for infrastructure development and optimizing transportation systems within the city.

This thesis is structured as follows. Chapter 2 gives an overview of the background of GPS data and the area of focus of this study. Chapter 3 provide insights into related works covering various mobility research that have used GPS data with spatial and temporal information. Next, we describe the GPS dataset, its key features, preprocessing steps, and implementation tools used in Chapter 4. Chapter 5 and 6 outline the methodology and findings of temporal and spatial analysis respectively. We conclude with a discussion of limitations and future directions in Chapter 7.

# 2 Background

In this section we cover the background information relevant to our work. As previously stated, this thesis analyses GPS data of Tartu. Therefore, we explained the background of GPS data in Section 2.1. In addition, we covered our area of study details that is Tartu, Estonia in Section 2.2.

## 2.1 GPS data

Over the years, location-based data collection has become increasingly common and has greatly benefited various areas of life. The Global Positioning System (GPS), which was developed by the United States and is still owned and operated by them has played a significant role in this advancement [8]. GPS is widely used for navigation and positioning, as seen in map applications and transportation systems [9].

At present, there are at least 24 satellites orbiting the earth that help determine the location of an object or a person on earth by continuously sending out radio signals [8]. Technical devices equipped with GPS, such as mobile phones, must receive the signals and measure the latitude, longitude, and time spent in motion. These satellites are distributed in six different circular orbits to ensure coverage of the earth at all times.

To determine its own location, a GPS receiver, like a mobile phone, measures the time it takes for signals from a minimum of four satellites to reach the receiver [10]. Radio waves propagate at a constant speed, enabling the receiver to calculate its distance from each satellite based on time measurements. However, four satellites alone do not provide a position with an accuracy of one metre. Instead, positioning accuracy is enhanced by the inclusion of more satellites within the signal transmission radius. Using seven satellites, for instance, can yield a position with an accuracy of up to 10 metres [11].

Due to its accuracy and universality, GPS has become an integral part of many people's daily lives, enabling humanity to accomplish tasks that were previously impossible [12]. For example, the Global Positioning System enables people to navigate their way through tough routes, such as unexplored nature or oceans [12].

## 2.2 Study area

Estonia is a country in the northeastern part of Europe, spreaded out in a total area of 45 339 km² [13]. With a population of 1.36M as of 1st of January, 2023 [14], it is considered an innovative digital society.

Tartu is the second largest city in Estonia and is located in the southern part of the country, in Tartu County (Tartumaa). The area of Tarty County is 2993 km², which is about 7% of all the territory in Estonia [15]. The city is situated on the banks of the river Emajõgi. The total surface area of Tartu is 154 km² and it has a population of 94 663 as of 1st of Januray, 2022 [16]. The oldest and most renowned university in Estonia, University of Tartu, is located in the city, therefore Tartu is considered a university town and the intellectual capital of Estonia. More than a fifth of the population consists of people who study at higher educational institutions [16].

As the river Emajõgi flows through Tartu, it also splits the city into two parts. There are a total of 7 bridges that connect the two parts of the city. Tartu is divided into 17 districts, whereas Kesklinn is considered the central district of the city. Another notable district is Maarjamõisa, which is considered as the medical campus in Tartu [17]. In addition to clinics, it also has numerous educational buildings mainly associated with the University of Tartu [17].

# 3 Related Works

Through the years, there have been many researches that have conducted mobility analysis to understand human behaviour. The understanding of mobility can contribute to many different domains. One way to understand human mobility is through GPS data. This section reviews some of the works that have explored the application of GPS data in mobility analysis. In this related works section, we will review some of the works that have explored the application of GPS data in mobility analysis and identify how this thesis contributes to the existing knowledge in this field.

GPS technology has been used extensively to study human mobility patterns in various contexts. Some studies have focused on the analysis of GPS data to identify the locations that people visit and the routes they take to reach these locations. A study conducted by Ashbrook and Starner [1] used GPS data collected from Atlanta, Georgia, to create clusters of places using a variant of k-means clustering algorithm. For every cluster they found, they also determined if there were any sublocations that were connected to the initial location. Based on this they were able to create a Markov model for each of the locations. A similar research was done by Khetarpaul et al [2], where they had a GPS trajectory dataset, collected from people in China, mainly in Beijing. Firstly they detected stay points of each user by using thresholds. Their next step was to perform a bag union operation to identify dense areas visited by many users. Lastly, they applied mode operation to find high frequency locations visited by multiple users. Their approach detected 42 locations as interesting locations.

Several studies have tried to extract mobility patterns from GPS data. Mohareb and Omar [3] conducted a study, where they analysed the movements of Beirut Arab University in the city of Tripoli, Lebanon. The results of their study revealed that male students report longer trip durations and cover more distance than female students, although the number of male students was lower in their study than female participants. Furthermore, they discovered that the mobility patterns at weekends were not locked to the same destination, and students move more freely to their own individual sources and destination attractors. Xia et al [4] explored the mobility patterns in Shanghai, combining taxi GPS data with a dataset consisting of subway smart card transactions. The authors used heat maps to show hotspots and busy routes for taxis, and pinpointed which locations are busy on weekends, with more taxi trips in the morning rush hours. They also identified hotspots that were busy during the evening rush hours on the weekends. Finally, they analysed the differences in human mobility patterns between weekdays and weekends and found that more people travel by taxi on weekends, and the distribution by subway is more regular than that by taxi.

Other studies have focused on identifying factors that have influenced human mobility. A study conducted in the Netherlands, Meijles et al [5] conducted a case study, where they investigated visitors roaming a national park. They used Classification and Regression Tree (CART) analysis to identify factors influencing visitor distributions. By analysing the data, the authors find that walking speed, trip time, and spatial distributions vary among different visitor groups. In another study conducted in Canada, Hirsch et al [6] looked into the mobility habits of older adults (age 65+) and generated activity spaces. Their results showed that factors such as age, neighbourhood walkability, possession of a valid driver's licence, access to a vehicle, and physical support for going outside the home influenced the size of the activity spaces.

There has also been a similar study that used data from the same study area as this thesis. Daiga Paršova [7] wrote a thesis that explored the spatiotemporal behaviour of students and academics at the University of Tartu based on the socio-economic characteristics of respondents and the location of the campus. The study analysed the space and time use, daily and weekly temporal rhythms of activities, and differences regarding time use and location of stops depending on an individual's academic role. The results found that the main work or study place did not prove to be a statistically significant factor affecting the number or duration of stops, but ownership of a car was proven to have a positive effect on the number of stops per day, since it provides easier mobility.

The aforementioned studies have showcased how GPS data is useful in mobility analysis. Researchers have used various methods to get an understanding of human behaviour, including clustering, CART analysis, identification of significant places, classification of behaviour, and the examination of factors. These studies have contributed valuable insights into understanding human mobility patterns that are useful for urban planning, transportation management, and public health.

By applying appropriate analytical methods to the collected GPS data, our research intends to contribute to a deeper understanding of human mobility patterns and provide insights that can inform decision-making processes in urban planning, transportation management, and related fields.

# 4 Dataset

## 4.1 Data description

The dataset used in the thesis was provided by the Mobility Lab at the University of Tartu. This dataset contains mobility data from approximately 200 individuals spanning from February 1, 2017, till April 5, 2017. The data was collected via a mobile application developed by the Mobility Lab, which captured GPS data after every second whenever the mobile was turned on. The dataset consists of GPS locations mostly in Estonia, but also from abroad, if the user left the country. The collected dataset consists of the following features:

- **user_id:** A unique identifier to distinguish users.
- **id:** A unique identifier to distinguish data records.
- **restart:** Technical parameter.
- **counter:** Counter to count data records.
- **time_system:** Time when the location was saved according to the time on the phone.
- **time_gps:** Time when the location was saved according to the time of the GPS.
- **time_system_ts:**Timestamp of the time of when the location was saved according to the time on the phone.
- **time_gps_ts:** Timestamp of the time of when the location was saved according to the time on the phone.
- **accuracy:** Accuracy of the GPS.
- **altitude:** Altitude value compared to the last location.
- **bearing:** Bearing value compared to the last location.
- **speed:** Speed value of the mobile.
- **header_id:** A unique identifier to distinguish headers.
- **series_id:** A unique identifier to distinguish series.
- **X:** Longitude value of the location.
- **Y:** Latitude value of the location.

Figure 1 shows the distribution of the number of records per user in the GPS data. Here, the x-axis represents users and the y-axis represents the number of data records of the user. A higher bar means more records for the user. In the data, we observe that there are only a handful of users with over 300,000 records, while there are the majority of users with less than 150,000 records.

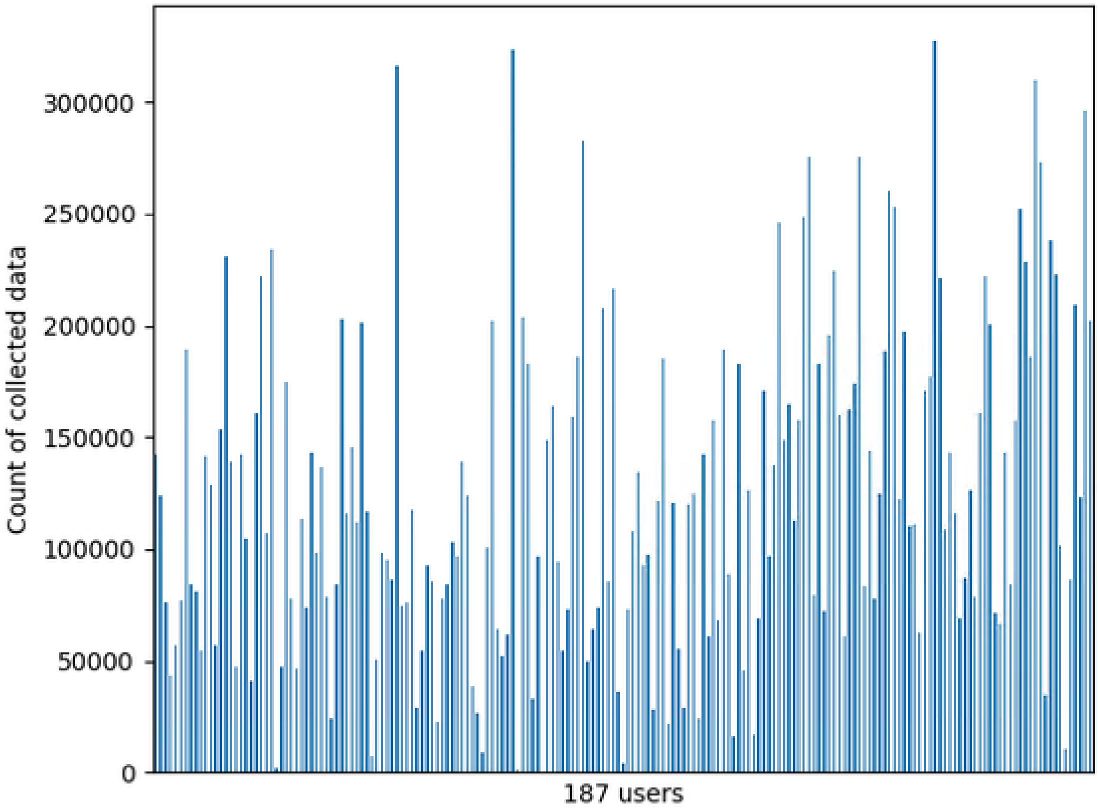


Figure 1. Number of Data Records Collected per User.

Figure 2 shows the distribution of the number of records with respect to dates. Here, the x-axis represents the date in the data and the y-axis represents the number of records. In the figure, we can observe the weekly pattern, where there are more records of data from the end of the week compared to the beginning. Besides this, the pattern is consistent through the two month period and the distribution between days similar. From the figure we can also tell that the number of records collected from April is very minimal and out of proportion compared to other days. This leads us to the conclusion that the two days worth of data recorded from April might be anomalies.

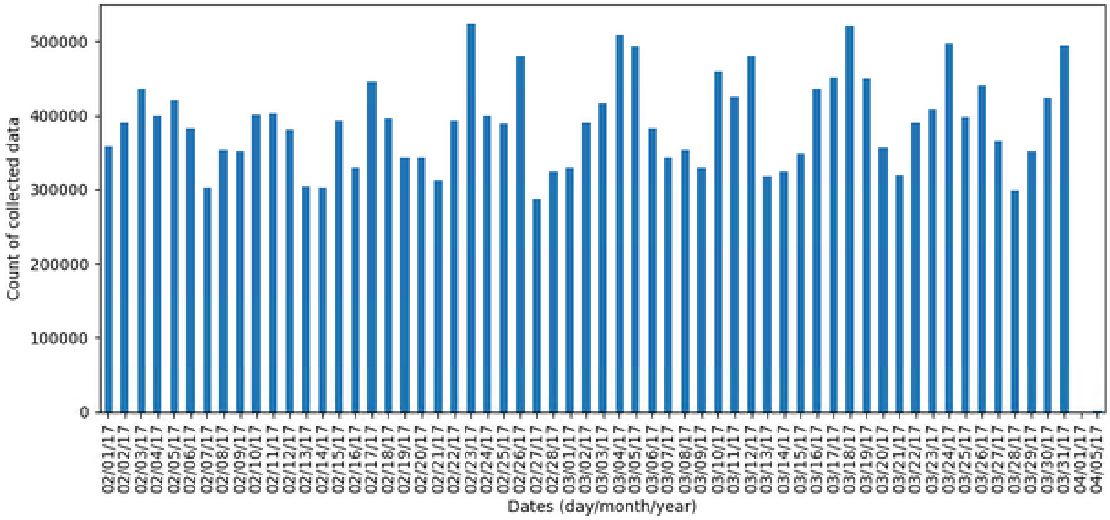


Figure 2. Number of Data Records Collected per Dates.

## 4.2 Implementation and tools

### 4.2.1 Python

Python is one of the most popular open-source programming languages in the world, since it is very versatile and flexible. It is widely used in many application domains, some of them being web development, scientific computing, data analysis, and machine learning. Python is a popular choice for data analysis and manipulation due its ease of use and large standard and third-party libraries available.

### 4.2.2 Jupyter Notebook

Jupyter Notebook is a web application, which allows users to create and share documents that contain live code, equations, visualisations, and narrative text. It supports multiple programming languages, including Python, R, and Julia. Jupyter Notebook allows the code to be executed and the result displayed in real-time, which makes it easy to explore and analyse data. This also makes it a popular tool among data scientists and researchers.

### 4.2.3 QGIS

QGIS is an open-source geographic information system (GIS) software that is used for analysing and visualising data. It is a powerful tool in geoinformatics used by geographers, GIS professionals and many more. QGIS is able to import and manage a wide range of geospatial data formats and provides an extensive data analysis toolbox. It is customizable, which makes it an ideal tool for researchers.

### 4.2.4 Code availability

All of the source code used for data processing and analysis is gathered in a private GitHub repository. To obtain access to the code, please reach out to the author.

## 4.3 Data preparation

The given dataset initially consisted of roughly 23.04 Million records collected over the period of two months (February and March) and two additional days (in April). Since the data from the two days in April was minimal and seemed anomaly, therefore this data was removed. All the data that was collected while the user was abroad, were extracted due the inexistence of some base maps that were needed for processing the data. The leftover data was then plotted to a map, which revealed that most of the gps locations were clustered in Tartumaa. Therefore, all the locations outside of Tartumaa were excluded.

After removing these rows, roughly 14.17 M records remained. Since this thesis focuses on the spatiotemporal patterns of human mobility, the following features were redundant and removed: 'id', 'counter', 'restart', 'header_id', 'series_id', 'time_system', 'time_system_ts', 'altitude', 'bearing'. The data was already cleaned by The Mobility Lab, therefore there were no missing or undetermined records in it and the final number of records remained the same number reported previously. The number of users decreased by a few people.

Next step was adding two additional columns that were useful for later work. One of the added columns shows speed in kilometres per hour. This was derived from the initial speed column, which showcased speed in metres per second. In order to get speed in kilometres per hour the given speed column values were multiplied by 3.6. The second added column provides information about the distance travelled between the current and previous GPS points. To achieve this, the haversine formula was used from the haversine library [18]. The latitude and longitude values of two consecutive GPS points were given to this formula, which then calculated the distance between the two points and returned the distance in kilometres. Using this approach on the whole dataset, a new column of distances was created.

## 4.4 Detecting movement records

In order to understand mobility patterns, movements needed to be detected from geolocation points. In order to achieve this, a threshold based algorithm was written, which takes all of the GPS locations of one user and detects when a movement has begun and when it ends. In addition, the algorithm calculates the total distance travelled in one movement.

At each record user id, time, timestamp of time, accuracy, speed in metres, latitude, longitude, speed in kilometres and distance is stored.

For movement end detection, two parameters are used: time and speed in metres. Movements are considered ended when the time elapsed between the last and observable record is more than 300000 milliseconds (5 minutes) and the speed of the observable record is under 1 m/s. The idea and the thresholds were inspired by the literature review of existing methods by Gong et al [19]. The algorithm was tested before on a smaller sample and the results checked manually by visualizing all the geolocation data and checking movements from there.

Once the movement ending was detected, all the records in one movement were monitored to detect a starting point. The parameter used for movement end detection was speed in metres. A movement has started when three consecutive records have the speed of over 1 m/s. All the records before this were removed from the movement. With the start and end of a movement detected, the movement is complete and distance of the all the movement is calculated based on the distance parameter in the movement records. The movement data is then added to a dataset, where all the movements from all users are stored. This dataset consists of the following features:

- **datapoints:** The number of all individual records that make up the movement.
- **start_time:** Time when the movement has begun.
- **end_time:** Time when the movement has ended.
- **start_time_unix:** Timestamp of the time of when the movement has begun.
- **end_time_unix:** Timestamp of the time of when the movement has ended.
- **start_X:** Longitude value of the location where the movement started.
- **start_Y:** Latitude value of the location where the movement started.
- **end_X:** Longitude value of the location where the movement ended.
- **end_Y:** Latitude value of the location where the movement ended.
- **total_distance:** The total distance travelled within the movement.
- **user_id:** The id of the user associated with the movement.

After the movement is detected and added to the all movements dataset, the algorithm takes the next data records and repeats the same process on all the users. As a result, we had a dataset of all movement records, the final number of records being approximately 40000.

# 5 Temporal analysis

## 5.1 Research goal

The goal for the temporal analysis is to examine mobility across different days of the week and times of the day. In particular, the following research question is stated: **what are the temporal patterns of human mobility across different days of the week and times of the day?** By examining how individuals move and navigate their surroundings during different temporal contexts, we aim to uncover any recurring patterns or variations. This analysis will enable us to understand the dynamics of human mobility and explore potential factors influencing movement behaviours.

## 5.2 Detecting movements daily and hourly

Next step was to understand when the movements were made. For this two new columns were created, that store the weekday and hour of when the trip was made. The base for obtaining this information is the starting time of the movement.

Another goal was to divide the movements into clusters based on the distance travelled to observe the distance travelled in different days and time. In this thesis, we differentiated three clusters: movements with distance less than 500 metres, movements with greater equals than 500 metres and less than 3000 metres, and movements greater equals of 3000 metres. This allowed us to classify these movements into short, medium and long length movements respectively. The range of these three clusters were decided taking into account the size and measurements of the county and the city of Tartu.

## 5.3 Results

We have split the data into movements for every user in the original dataset. We start the analysis by looking into the weekly patterns of movements and later hourly trends. Since our data is collected from a two month period, from the beginning of February until the end of March, the behaviour can be associated with only late winter and early spring.

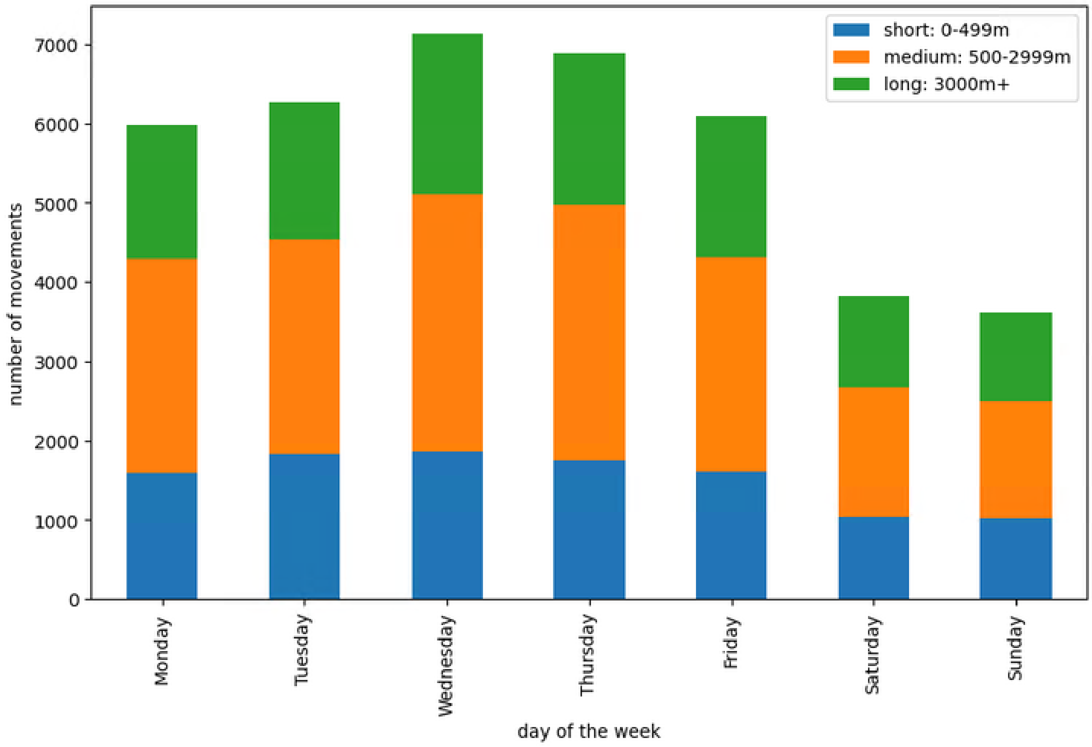


Figure 3. Number of Movements Aggregated by Days of the Week and Clustered by Length.

The total number of movements during the two month period was approximately 40000, out of which 82% were done during weekdays (Monday to Friday). From Figure 3 we can see that the number of movements peak on Wednesday and it gradually drops down, with a big drop during the weekend. The least amount of movements were made on Sunday, where only 9% of movements were recorded. This can be explained by the fact that Monday to Friday are considered workdays while the weekends are left for resting and people tend to not move around so much. Figure 3 also shows that every day the most movements done were considered medium length, with the distance being between 500 - 2999 metres. While the number of movements differ between days, the proportions of movements lengths have a tendency to remain very similar.

Next we looked into the distances travelled within movements aggregated by days of the week. From Figure 4 we can see that Wednesday through Friday are the days when more distances are travelled altogether. Although less kilometres are associated with the weekend, the difference is not as drastic as with the number of movements made, as shown in Figure 1. This means that although the users make less movements on the weekends, the distances travelled are usually longer than during the week. One way to explain this is that people use the weekend more to visit places further from their daily trajectories.

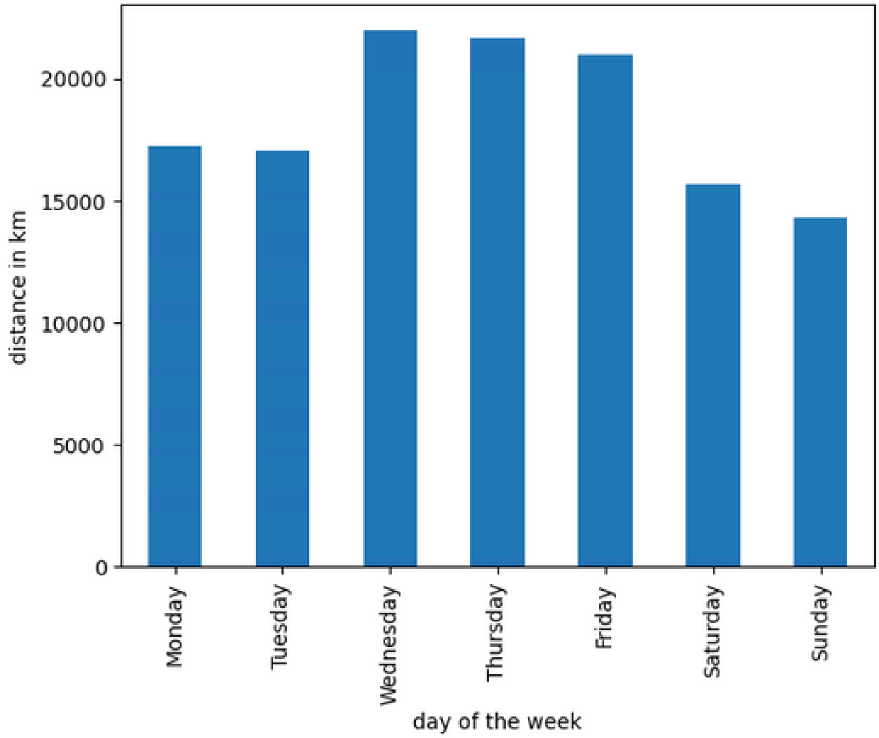


Figure 4. Distance Travelled Within Movements Aggregated by Days of the Week.

Figure 5 shows that the highest number of movements were made at 5 PM, with 4 PM and 6 PM also ranking high. Other noteworthy peak hours include 12 PM, 13 PM and 8 AM. This can be explained by the fact that the usual workday starts around 8 AM and ends around 5 PM, meaning the users most likely commuted to and from work during these times. High number of movements being made at midday suggests that people might take breaks for lunch at this time. When looking at the lengths of the movements made during different times we noticed that almost all of the long movements were made during daytime. Furthermore, longer movements make up more of the total during the 8 AM and 5 PM peak hours, while at 12 PM and 13 PM most of the movements are considered less than 3000 metres. This suggests that during the daytime people tend to move around in the same area and longer distances are more travelled in the mornings or evenings.

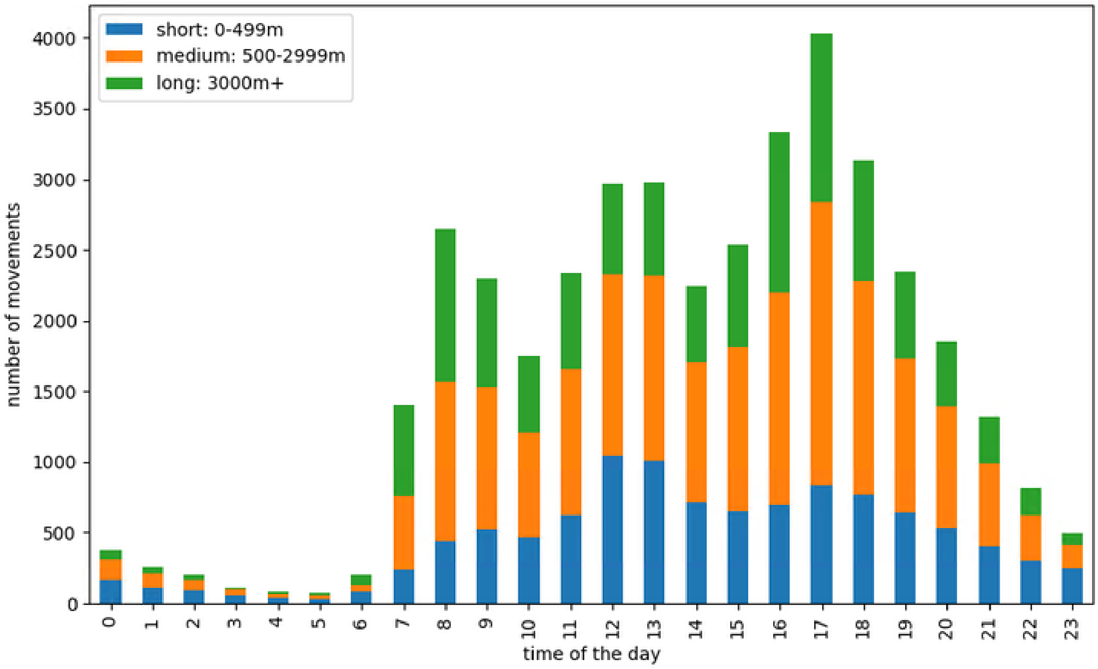


Figure 5. Number of Movements Aggregated Hourly and Clustered by Length.

In addition, we looked at the distances involved hourly. Figure 6 shows the peak hours of 8 AM and 5 PM more clearly, meaning that people moved the most during the stated times. After 5 PM the distances are gradually getting smaller, which indicates people are not as active in the evening as during the day.

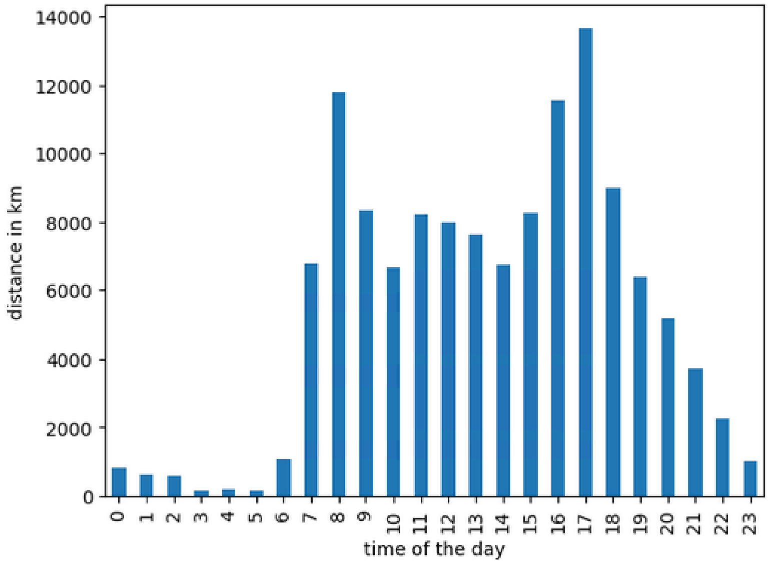


Figure 6. Distance Travelled Within Movements Aggregated Hourly.

From Figure 7 we can conclude that most movements are from 5 PM to 6 PM during weekdays, with Wednesday having the largest count. The smallest number of movements are happening at night time. It is also worth noting that people are making a lot less movements during the weekend and if they are moving, the movements take place later than on the weekdays.

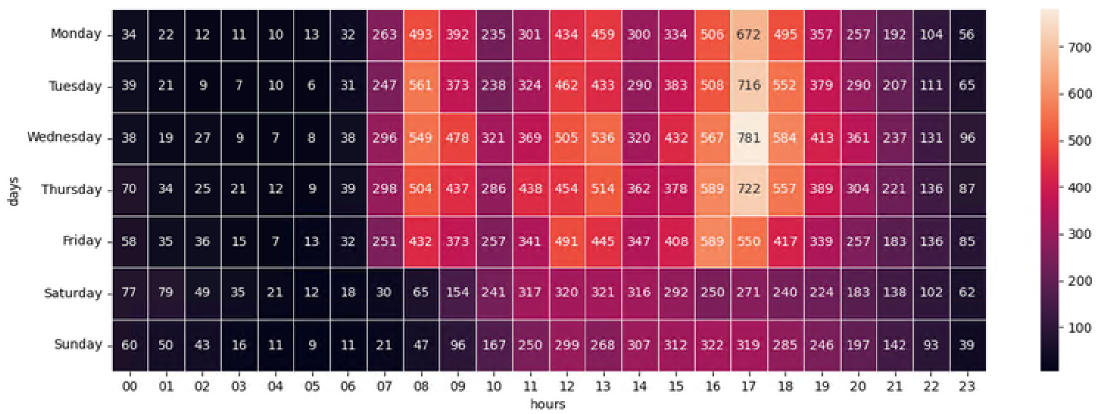


Figure 7. Distance Travelled Within Movements Aggregated Hourly.

# 6 Spatial analysis

## 6.1 Research goal

The goal for the spatial analysis is to identify the key locations that attract people and where movements take place. For this, the following research question is stated: **which locations are the most frequently visited by individuals, and when are they visited the most?** By dissecting the temporal dimension, we can uncover when these locations experience peak visitation, providing insights into the temporal preferences of individuals and potential factors driving their visitation patterns.

## 6.2 Detecting points of interest (POIs)

The movements dataset is also used for identifying important places within the users. To achieve this, the movements end points are retracted into a new dataset, where their GPS location and time are stored. This dataset, consisting of approximately 40000 records, was then added to QGIS as a layer with a map of all the buildings in the study area. An algorithm provided by QGIS, called intersection [20], is then used to only get movements end points, which are inside a building. This resulted in a new dataset, which also included addresses from the buildings map and matched them with the geolocation provided by us. The outcome was then processed further in Jupyter Notebook. A total of approximately 1500 different addresses were visited over the two month period. All the movement endings are then counted and grouped by address to get an insight of the most visited places. Most of the top 25 locations are considered public locations, after the 25th location private addresses started to emerge more frequently. Therefore, we decided to extract the top 25 places from the result to execute further analysis on most visited places.

Further action included extracting the hour from the time to a new column, to gain insight on visited hours. Once this was extracted, a simple algorithm was written that counted how much each user visited the extracted places and at what hours.

To get context on the visited places we manually checked the locations using the Estonian Land Board data [7] and divided the locations into the following categories: university building, museum/archive, shopping centre, residential building, grocery store, and hospital. The reason for choosing these categories lies in the fact that we looked at public indoor locations only. All addresses that were presumed to be home locations were removed due to privacy reasons. This left us with the final number of POIs being 23 out of 25. All other locations can be classified under one of the categories listed above.

## 6.3 Mapping important places

For visualisation of the important places 2 maps (See Figures 8 & 9) were created using QGIS. The Mobility Lab provided some maps by the Estonian Land Board, which were used as the basis for creating a map for showcasing POIs. This included a map of buildings, road networks, watercourses and stagnant water bodies.

## 6.4 Results

First, the locations were mapped with their addresses and divided into six categories. This resulted in a visual that revealed that all of the most visited places are in Tartu, as seen in Figure 8. The total of 23 locations consists of nine university buildings, five shopping centres, three museums and/or archives, three residential buildings, two grocery stores, and one hospital. With making up more than a third of the locations, university buildings were the most visited during the two month period, which indicates that most of the participants were university students or workers. When looking at the map, it is clear that roughly half of these POIs are located in the city centre. Another district, where 7 of the locations are clustered, is Maarjamõisa.

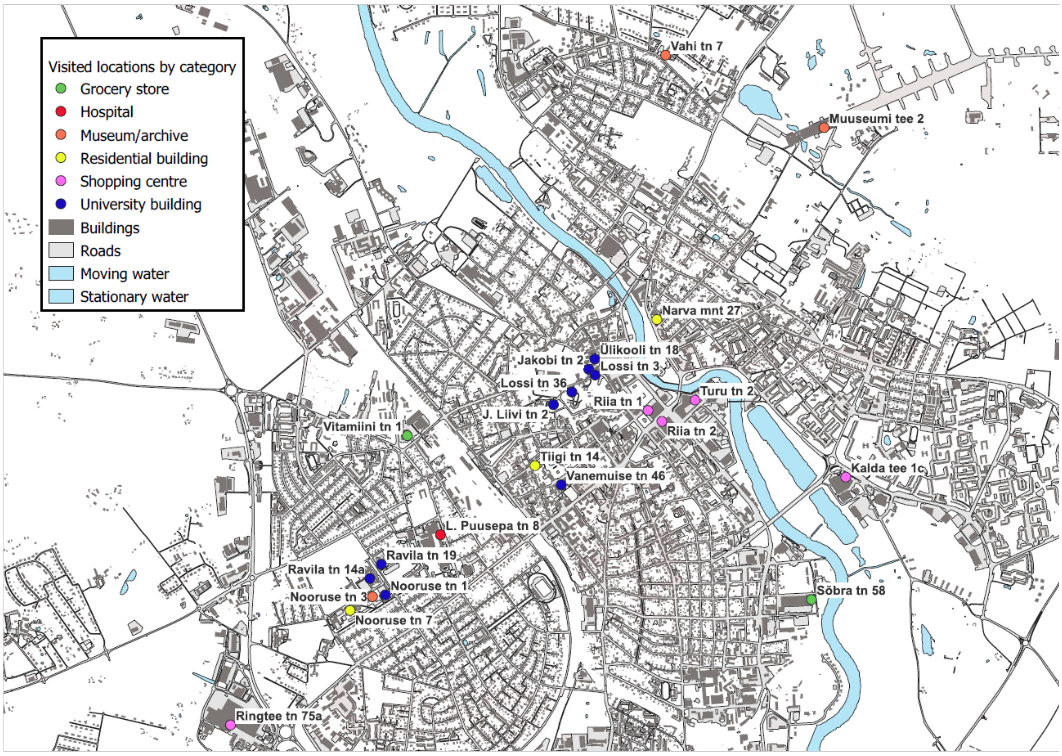


Figure 8. Map of 23 Most Visited Places with Addresses.

Figure 9 shows the same locations with their category on a gradual scale, ranked by visitation counts. The darkest locations, meaning the most visited, are museums/archives, university buildings, and shopping centres. These locations are all over the city and do not cluster in one specific region. An interesting find is that museums/archives are the most visited places and the reason for that might be that these museums/archives were opened only a few months before the data was collected, therefore they were visited a lot during the first few months.

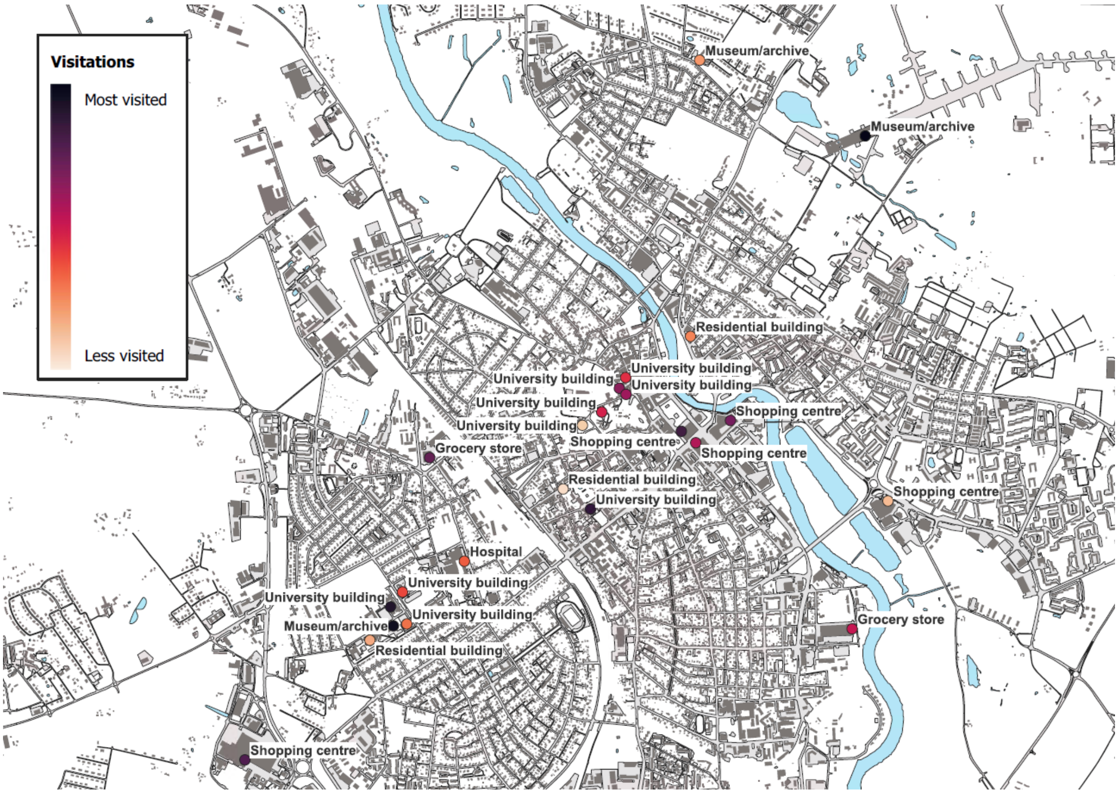


Figure 9. Heat Map of 23 Most Visited Places with Categories.

From Figure 10 we can conclude that university buildings are mostly visited during the early hours of the day, from 7AM to 10AM. Grocery stores and shopping centres seem to be popular later in the day, with their peak hours being from 3PM to 6PM. Museums/archives tend to have the same pattern of visitations as university buildings, which might indicate that they are used for studies as well.

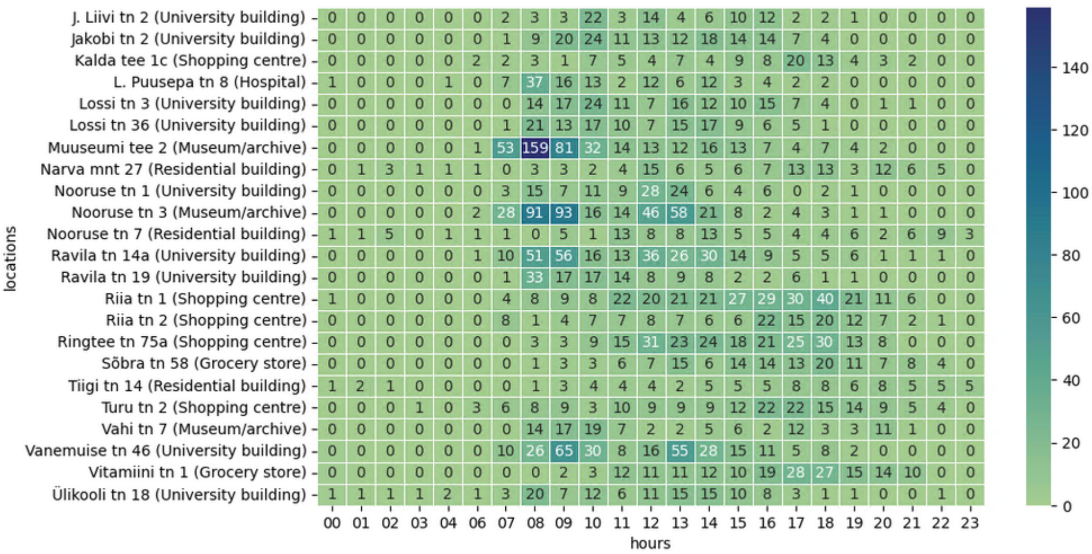

Figure 10. Heat Map of 23 Most Visited Places with Visit Counts Hourly.

# 7 Conclusion

The goal of this thesis was to conduct an exploratory analysis of GPS data from Tartumaa to examine mobility patterns and identify most frequently visited locations to get a better understanding of spatiotemporal patterns of humans.

The dataset used in this thesis was obtained from the Mobility Lab at the University of Tartu. The dataset comprised of roughly 23.04 Million records, which were collected over the period of slightly more than two months, from February 1, 2017, to April 5, 2017, and gathered from approximately 200 individuals. It contained 13 distinct features, including user ID, GPS coordinates, timestamps, speed, altitude, bearing, and accuracy.

Based on the analysis of movement data, derived from the aforementioned dataset, interesting patterns of weekly and hourly mobility were found. The majority of movements were made during weekdays, with the highest count being on Wednesdays, while the least amount of movements occurred on Sundays. The distances travelled on weekends were longer on average, which suggests that people used their free time for more leisurely activities. The peak hours for movements were at 8 AM and 5 PM, which can be explained by the typical workday schedule. Additionally, longer movements were more likely to occur in the morning or evening, while shorter movements were more common during midday. Finally, people made fewer movements at nighttime and during the weekends. In addition to these results, it was discovered that the top places visited were university buildings, following were shopping centres, museums/archives, residential buildings, grocery stores, and hospitals. Some of the POIs were clustered in the city centre and Maarjamõisa. Looking at the patterns of visitations hourly, university buildings were mostly visited during the early hours of the day, while grocery stores and shopping centres were popular later in the day. Museums/archives had a similar pattern of visitation to university buildings, which suggests that they may be used for academic purposes as well.

The results of this thesis offer an insight into spatiotemporal patterns of people roaming around Tartumaa, and specifically Tartu. This work could be useful for the people responsible of the urban planning in Tartu to develop optimal infrastructure and create better transportation systems.

**Limitations:** This thesis analysed the mobility patterns based on GPS data only and did not consider factors, such as weather or special events, that could affect the patterns. Furthermore, we used mobile phone data to analyse mobility patterns, which means that the data was not collected if the mobile was not with the user or it was turned off. This means that not all movements were represented, therefore it could have an impact on the mobility patterns.

**Future scope:** For future work, several directions could be exploited. First, the factors affecting the mobility patterns can be deeper examined as this thesis did not focus on this topic. Secondly, the relationship between the points of interest and mobility patterns could be further explored to build Markov models. Lastly, a proper framework to analyse GPS data can be developed to explore the spatiotemporal patterns.

# References


[1] Daniel Ashbrook and Thad Starner. Learning significant locations and predicting user movement with gps. *Procs. of the 6th IEEE International Symposium on Wearable Computers*, 7:101–108, 2002.

[2] Sonia Khetarpaul, Rashmi Chauhan, S. K. Gupta, L. Venkata Subramaniam, and Ullas Nambiar. Mining gps data to determine interesting locations. In *Proceedings of the 8th International Workshop on Information Integration on the Web: In Conjunction with WWW 2011*, pages 1–6, 2011.

[3] Nabil Mohareb and Ossama Omar. Monitoring daily mobility patterns for university students using gps tracking: Tripoli as a case study. *BAU Journal - Health and Wellbeing*, 1(3):2, 2018.

[4] Feng Xia, Jinzhong Wang, Xiangjie Kong, zhibo wang, Jianxin Li, and Chengfei Liu. Exploring human mobility patterns in urban scenarios: A trajectory data perspective. *IEEE Communications Magazine*, 56:142–149, 2018.

[5] E.W. Meijles, M. de Bakker, P.D. Groote, and R. Barske. Analysing hiker movement patterns using gps data: Implications for park management. *Computers, Environment and Urban Systems*, 47:44–57, 2014.

[6] Jana Hirsch, Meghan Winters, Philippa Clarke, and Heather McKay. Generating gps activity spaces that shed light upon the mobility habits of older adults: A descriptive analysis. *International journal of health geographics*, 13:51, 2014.

[7] Daiga Paršova. Mobility patterns in university campuses: an example of the University of Tartu. Master's thesis, University of Tartu Department of Geography, 2019. `https://dspace.ut.ee/handle/10062/65040` (accessed: 06.05.2023).

[8] GPS.gov. GPS overview. `https://www.gps.gov/systems/gps/`, 2021. (accessed: 07.05.2023).

[9] Library of Congress Science Reference Section. What is a GPS? How does it work? `https://www.loc.gov/everyday-mysteries/technology/item/what-is-gps-how-does-it-work/`, 2019. (accessed: 07.05.2023).

[10] National Geographic. GPS. `https://education.nationalgeographic.org/resource/gps/`, 2022. (accessed: 07.05.2023).

[11] Hellotracks Editorial Team. Common problems with GPS. `https://hellotracks.com/en/blog/How-to-Improve-your-GPS-Accuracy/`, 2023. (accessed: 07.05.2022).

[12] David Hambling. What would the world do without GPS? https://www.bbc.com/future/article/20201002-would-the-world-cope-without-gps-satellite-navigation, 2020. (accessed: 07.05.2023).

[13] The Information Architects of Encyclopaedia Britannica. Estonia. https://www.britannica.com/facts/Estonia, 2023. (accessed: 07.05.2023).

[14] Statistikaamet. Rahvaarv. https://www.stat.ee/et/avasta-statistikat/valdkonnad/rahvastik/rahvaarv, 2023. (accessed: 07.05.2023).

[15] Eesti Entsüklopeedia. Tartu maakond. http://entsyklopeedia.ee/artikkel/tartumaa4, 2011. (accessed: 07.05.2023).

[16] Tartu City Government. Tartu in figures 2021/2022. https://tartu.ee/sites/default/files/uploads/Tartu%20linn/Statistika/Tartu_arvudes_2022_ENG-uus.pdf, 2022. (accessed: 07.05.2023).

[17] Tartu Linn. Maarjamõisa linnaosa. https://tartu.ee/et/linnaosad#maarjamoisa, 2022. (accessed: 07.05.2023).

[18] Balthazar Rouberol. Haversine. https://pypi.org/project/haversine/#description, 2023. (accessed: 02.05.2023).

[19] Lei Gong, Takayuki Morikawa, Toshiyuki Yamamoto, and Hitomi Sato. Deriving personal trip data from gps data: A literature review on the existing methodologies. *Procedia - Social and Behavioral Sciences*, 138:557–565, 2014.

[20] QGIS Documentation 3.28. 27.1.19. vector overlay. https://docs.qgis.org/3.28/en/docs/user_manual/processing_algs/qgis/vectoroverlay.html?highlight=intersection#intersection, 2023. (accessed: 07.05.2023).

# Appendix

# I. Licence

## Non-exclusive licence to reproduce thesis and make thesis public