Exploring socioeconomic factors' impact on human mobility during the COVID-19 pandemic

A case study of Västra Götaland region of Sweden

Master's thesis in Computer science and engineering

PETER GÄRDENÄS, ERIK MAGNUSSON

Department of Computer Science and Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG
Gothenburg, Sweden 2021

Master's thesis 2021

Exploring socioeconomic factors' impact on human mobility during the COVID-19 pandemic

A case study of Västra Götaland region of Sweden

PETER GÄRDENÄS, ERIK MAGNUSSON

UNIVERSITY OF GOTHENBURG

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Computer Science and Engineering
Chalmers University of Technology
University of Gothenburg
Gothenburg, Sweden 2021

Exploring socioeconomic factors' impact on human mobility during the COVID-19 pandemic
A case study of Västra Götaland region of Sweden
PETER GÄRDENÄS, ERIK MAGNUSSON

Supervisor: Sonia Yeh, Space, Earth and Environment
Supervisor: Yuan Liao, Space, Earth and Environment
Supervisor: Jorge Gil, Architecture and Civil Engineering
Examiner: Aarne Ranta, Computer Science and Engineering

Master's Thesis 2021
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Typeset in LaTeX
Gothenburg, Sweden 2021

Exploring socioeconomic factors' impact on human mobility during the COVID-19 pandemic
A case study of Västra Götaland region of Sweden
PETER GÄRDENÄS, ERIK MAGNUSSON
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg

Abstract

Human mobility decreased worldwide during COVID-19. Most countries issued policies limiting or banning non-essential travel. Sweden issued no such policies and instead relied on non-binding recommendations. Sweden's unique response makes it an interesting country for mobility studies. International studies have found that mobility has not reduced equally between socioeconomic groups during COVID-19. This inequality is potentially harmful since mobility during COVID-19 has been linked to an increased risk of infection and death.

In this thesis, we explore how mobility has changed during COVID-19 for different socioeconomic groups in Västra Götaland region, Sweden. This is achieved by building a mobility model from 250 million geolocation records generated from phone apps. The data was collected during October–November for both 2019 and 2020 Five mobility metrics are calculated from the mobility model, namely: radius of gyration, trip distance, trips between regions, visitation frequency, and location temporal profile. Additionally, five socioeconomic clusters are created by clustering 1000 areas based on their official socioeconomic data, including income and education. The socioeconomic groups are then defined as all users living within a cluster. Finally, for each metric, the mobility change between the years is compared for the socioeconomic groups.

We found that mobility decreased during the pandemic in Sweden but to a lesser degree than in countries that issued lockdowns. Consistent with the literature, our study also observed differences in pre-pandemic mobility between the different socioeconomic clusters. Those differences, however, decreased during October-November 2020 compared to 2019. We conclude that the reduction in mobility is primarily driven by individuals with high pre-pandemic mobility, as most of the reduction was observed among the more mobile clusters and at the 50th and 75th percentile of the metrics. We also found that the most socially disadvantaged cluster had the lowest reduction in mobility, while the wealthiest cluster had the largest. This finding is consistent with our earlier conclusion, since the socially disadvantaged cluster had low pre-pandemic mobility, while the wealthy cluster had high mobility. We conclude that socioeconomic factors had a larger effect on pre-pandemic mobility than on the reductions in mobility during the observed months of the pandemic.

Keywords: Human mobility, COVID-19, socioeconomic.

Acknowledgements

We would like to thank our supervisors Sonia Yeh, Yuan Liao, and Jorge Gil for their generosity with their time. Providing continuous feedback, support, and expertise. Without them this project would not have been possible.

Peter Gärdenäs & Erik Magnusson, Gothenburg, June 2021

Contents

List of Figures

List of Tables

1
Introduction

The COVID-19 pandemic caused a significant reduction in human mobility worldwide [3, 4]. To curb the spread of the virus most countries issued nationwide lockdowns while Sweden took a different approach, relying mostly on individuals practising self-responsibility by following recommendations [5].

During the pandemic, some socioeconomic groups were more affected by the virus than others. There were significant differences in risk of infection and mortality depending on socioeconomic attributes such as income, education, and country of birth [6, 7, 8, 9]. Research has found a correlation between infection and mobility during previous epidemics as well as the COVID-19 pandemic [8, 10, 11]. Mobility could be a partial explanation for the discrepancy in risk of infection between socioeconomic groups since some groups might have been unable or unwilling to reduce their mobility. Therefore, in this thesis, we seek to explore *the relationship between groups of different socioeconomic status and the change in human mobility during the COVID-19 pandemic in Sweden.* Understanding if there is a link between socioeconomic factors and change of mobility patterns would allow for more efficient and equal policies, potentially saving lives. Sweden makes an interesting area of study due to its unique response to the pandemic. This study will focus on the Västra Götaland Region (VGR), the second most populated region in Sweden, and examines the period surrounding the outbreak of the second wave of infections, October and November. A timeline number of cases can be seen in Figure 1.1.

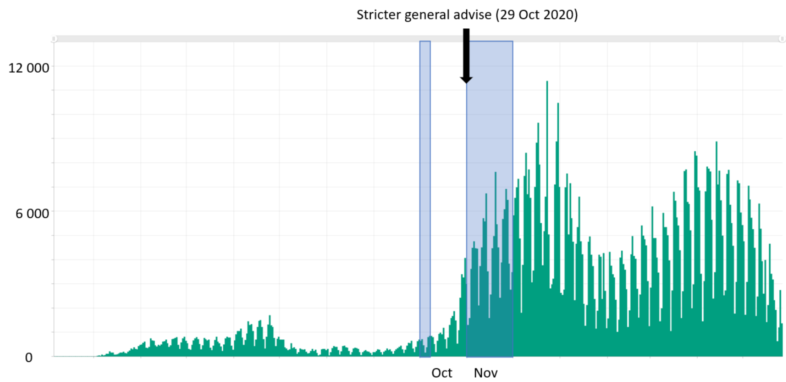


Figure 1.1: Number of detected COIVD-19 cases in Sweden. Highlighted areas represent time period data was collected. The 2019 data was collected during the same time the year. Source, Public Health Agency of Sweden [1]. Figure edited by authors.

To find a possible link between socioeconomic groups and mobility, a dataset of aggregated geolocation records generated by mobile phone app usage is used to build a mobility model. The records were collected during the first week in October as well as the entire month of November during both 2019 and 2020. The datasets include 100 million records from 42 000 devices in 2019 and 150 million records from 63 000 devices in 2020, with all devices visiting VGR at least once during the time period. For both years, a model is constructed by first inferring trips and home location from the records. For each device, we calculate mobility metrics including the radius of gyration, frequency of visits to each location, and inter-regional trips. An official socioeconomic dataset, DeSO, which divides VGR into 1000 areas, is then used to create socioeconomic clusters based on selected socioeconomic attributes and their inferred home locations. The mobility metrics were then aggregated for all users in the same cluster. We then compare the mobility change for each cluster to determine the effects of socioeconomic status on mobility during COVID-19.

1.1 Background

Human mobility refers to the movement of humans in space and time. Understanding human mobility can be useful in a wide range of applications, from estimating migratory flows, traffic forecasting, urban planning, to epidemic modelling [12]. At the time of writing, the relationship between the COVID-19 pandemic and mobility is still being actively studied. These studies have different aims such as understanding the relationship between human mobility and infection rates to make informed reopening recommendations after lockdown or which policies have the greatest effect on lowering infection rates [8, 13]. Other studies have focused on the demand for public transport during COVID-19, and how socioeconomic factors can attribute to these changes in demand [14, 15]. The Swedish health agency uses mobility data to understand the population's movement, finding insights to curb the spread of the virus [16].

The Swedish health agency considers socioeconomic data when crafting policies, such as the order of vaccination. Taking socioeconomic factors into consideration when deciding the order of vaccination is motivated by the correlation between mortality and socioeconomic factors [17]. Attributes such as country of birth, income level, and education are all shown to affect the risk of mortality [7, 17]. This is not unique to COVID-19, as socioeconomic factors are known to influence overall health, with groups of higher incomes having a greater life expectancy [7].

Research has also found a connection between infection and mobility. One such connection is the correlations between visits to points of interest such as restaurants or hotels and increased rates of infection [8]. Moreover, mobility has also been used to predict the spread of infection both during COVID-19 as well as previous epidemics. Early in the pandemic, mobility out of Wuhan had such an effect on infection growth rates outside the city that it alone was a strong predictor of the total number of cases [11]. In a Spanish study a single mobility metric, radius of gyration, was sufficient to predict if the COVID-19 mortality would increase or decrease in the upcoming weeks [18].

Socioeconomic attributes such as level of education and income level are shown to affect mobility during the pandemic [8, 14, 15, 19, 20]. Public transport usage is known to drop drastically during epidemics or pandemics [21] and COVID-19 is no exception [14, 15, 22]. The decrease in public transport ridership was found to be smaller among disadvantaged socioeconomic groups in the US. Likely due to a greater dependency on public transport and a greater inability to work from home [15]. Another study predicted higher rates of infection among disadvantaged racial and socioeconomic groups in the US due to lower decreases in mobility [8]. In Colombia, the wealth of a city was found to have a greater effect on mobility than the strictness of the lockdown, with wealthier citizens reducing their mobility to a greater extent [19]. In Sweden, two studies looking at different mobility metrics have found different results. The first study found that the reduction in the use of public transport among different groups was strongly tied to socioeconomic factors [14]. However, the other study found no differences in maximum distance travelled from home between different socioeconomic factors in Stockholm [23].

1.2 Thesis goal

As outlined in Section 1.1, some socioeconomic groups have a higher risk than the others in terms of mortality and infection during the COVID-19 pandemic. It is also known that mobility behaviour changes have a significant impact on the virus spread. Therefore, it is important to explore the impact of socioeconomic factors on the mobility change during the pandemic, which contributes to better policymaking curbing virus spread.

Therefore, our goal in this thesis is to explore *How socioeconomic factors impact the change of mobility during the COVID-19 pandemic*? We aim to do this by presenting how different socioeconomic groups have changed their mobility during COVID-19. We do not aim to infer the connection between infection and mobility.

To achieve this goal, the following sub-tasks are implemented:

- Build a mobility model based on aggregated geolocation records.
- Quantify mobility metrics that may have been affected by the pandemic using the mobility model.
- Generate socioeconomic clusters by clustering local areas based on the socioeconomic attributes of the residents who live in these areas.
- Compare the mobility change between 2019 and 2020 for the identified mobility metrics for each cluster.

1.3 Ethical considerations

Location data generated from mobile phone apps is privacy-intrusive. A study suggests that as little as four data points can be sufficient to uniquely identify a trace among millions of others. The trace can then in some cases be used to identify the individual [24]. To mitigate the privacy concerns, no individual traces will be presented. Furthermore, inferred home locations will be presented and analysed on an area level, where each area has at least 500 residents.

2

Related work

The following chapter will introduce how mobility models and mobility metrics have been used and calculated in literature. The chapter will also introduce how socioeconomic data have been applied to mobility research.

2.1 Mobility data

Commonly used mobility data sources include traffic data, mode-specific trip data, social media data, household travel surveys, Call detail records (CDRs), and smartphone app data [12, 25]. The two latter, both stemming from cell phones have revolutionised the field of human mobility with their vast timestamped location datasets. As these datasets allow for studying mobility on the individual level, with a spatio-temporal resolution that was not possible before [12]. The timestamped location data often comes in form of a geolocation record. The attributes of a simple record of geolocation can be seen in Table 2.1.

Attribute	Description
Device id	Unique id of the device.
Longitude	Longitude coordinate.
Latitude	Latitude coordinate.
Timestamp	Timestamp of the activity generating the geolocation

Table 2.1: The attributes of a simple geolocation record.

CDRs and *smartphone app data* are the most common sources of mobility traces from cell phones.

Call detail records (CDRs) are obtained from mobile operators and turned into mobility data by approximating cell phone locations based on which cell phone tower they are connected to. This type of approximation can in urban areas be quite precise where a tower can cover as little as tens of meters but gets much worse in rural areas where a single cell tower can service several kilometres [12]. Issues with this type of data include irregular sample rate since it depends on individuals' frequency of texting and calling, and the vastly varying location precision mentioned before.

Smartphone app data on the other hand is data collected from various smartphone applications. The geolocations of this data can be gathered in various ways such as from WiFi and cell tower connections, the phone's Global Positioning System (GPS), or a fusion of multiple sources. The quality varies but the main benefit of this data is that in most cases it utilises the phone's GPS which has a median precision of at least 10 metres [26]. Locations from fused sources are the most precise as GPS does not function well indoors. This is the type of data that we will be using, with all the different location sources mentioned above.

2.2 Mobility model

Mobility models are useful to calculate mobility metrics. In this section, four important components of a mobility model will be introduced: stops, trips, home location, and work location.

2.2.1 Stops

A stop refers to a location where a user is stationary for some time. Stops are a central building piece in the mobility model since they can be used to infer trips, home location, and work location as well as used to calculate mobility metrics.

According to Horanont (2013) stops are inferred from a series of traces, L_u.

$$L_u = \{l(1)_u, l(2)_u, ..., l(n)_u\}$$

Were $l(i)_u$ is the i^{th} location trace for user u. A subset of L, S, is then defined as all traces which are in close spatial proximity for a sufficiently long time. I.e for every subset S:

$$\begin{aligned}
&S(x)_u = \{l(a)_u, l(a+1)_u, ..., l(b)_u\} \text{ then} \\
&time(l(b)_u - l(a)_u) > min\ time \text{ and} \\
&l(i)_u, l(j)_u \in S(x)_u \text{ then } dist(l(i)_u, l(j)_u) < max\ distance
\end{aligned}$$

Stops are then defined as the first and last locations within the subset S [27]. This method can be adapted by instead defining a stop as the centroid of S [28].

2.2.2 Trips

Trips are a central concept in mobility studies since many mobility metrics are generated from them. They describe a journey a user takes between geolocation records. A simple approach to infer trips is simply defining it as the displacement between two consecutive location traces from the same user [29]. Another approach is to first infer stops and then defining a trip as the displacement between two consecutive stops, with the possibility of including all location traces in between [27, 28].

2.2.3 Home and work location

Home and work location can be inferred from the most visited night-time and day-time location, respectively. Where a visit can either be a geolocation record or a stop [27, 30, 31]. Another approach of inferring home and work location is by looking at the most and second most frequently visited locations since they have been found to correlate with home and work location respectively [32, 33].

2.3 Origin-destination matrix

An Origin-destination matrix (OD-matrix) is an aggregation that is useful for analysing movements of an entire population. It provides an estimated number of people travelling between nodes, over a given period of time, these nodes are often administrative regions or similar [12]. Let $\boldsymbol{T}$ be a $n \times m$ OD-matrix, n is the number of origin nodes, m the number of destinations, then T_{ij} is the number of people travelling from node i to node j. Badr (2020) used two OD-matrices, one from before COVID-19 and one during to calculate a mobility metric which could be used as a proxy for social distancing [13].

2.4 Mobility metrics

Mobility metrics are used to characterise mobility and applied to quantify the impact on mobility during COVID-19. The following section will introduce a few mobility metrics used in the thesis and how they have been used in literature.

2.4.1 Radius of gyration

Radius of gyration is a common metric within the field of human mobility which indicates the characteristic travel distance an individual travels from the centre of mass of their stops [12].
It is defined as:

$$r_g = \frac{1}{N}\sqrt{\sum_{i=1}^{N}(r_i - r_{cm})^2}$$

where N is the amount of stops made by the individual during the time frame, r_i is the coordinates of stop i, and r_{cm} is the coordinates of the centre of mass, which is defined as $r_{cm} = \sum_{i=1}^{N} r_i/N$.
Averaging individuals' radius of gyration can be used as a measure of the extent of individual movements [30]. A study of Spain's aggregated radius of gyration was able to predict if COVID-19 deaths would increase or decrease with a 3-week offset [18].

2.4.2 Visitation frequency

Visitation frequency describes how frequently an individual visits a location. When ordering locations in terms of most frequently visited first, the frequency of visits is expected to decay with Zipf's law, except for the most and second most visited location [32].

The visitation frequency for each rank can then be aggregated [32, 33]. The aggregated visitation frequency can then provide insights on how frequently the population visits their home and work location, since they tend to be the most visited and second most visited location respectively as discussed in Section 2.2.3 [32]. Furthermore, this metric can also provide insights into different sub-populations, since they can have distinct patterns in their visitation frequency, such as some sub-populations tend to visit a few locations very frequently whereas others spread out their visits to a higher number of locations [33].

2.4.3 Location temporal profile

The location temporal profile describes when a user is active or when activities occur at a given location. The metric is calculated by first dividing a time unit into time intervals, such as dividing a week into time intervals of one hour, resulting in intervals such as Monday 5 pm to 6 pm and Friday 7 pm to 8 pm. Then the number of visits occurring at a location during each interval is counted. The visits for each interval can then be compared to approximate when a location is typically visited. The temporal profile can also be aggregated for categories of locations, such as creating a gym category by aggregating the temporal profile of all gyms. Different categories of locations can have different temporal profiles, as shown in Figure 2.1 [2].

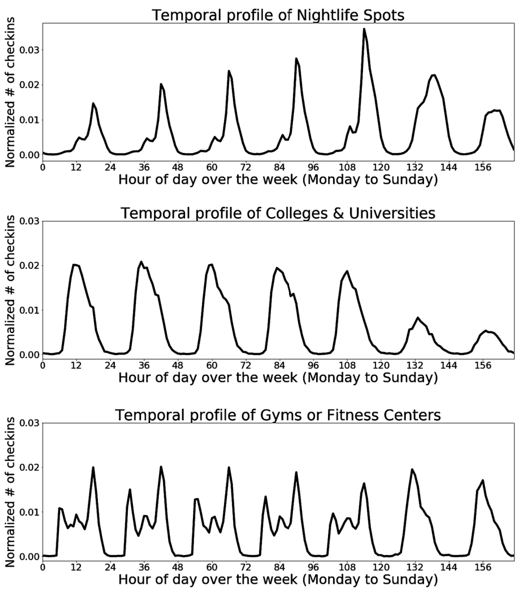


Figure 2.1: Example of temporal profile for three different types of locations. Source: D'Silva (2018) [2].

2.5 Socioeconomic factors in mobility studies

Location data extracted from CDRs, phone apps, and social media lack any socioeconomic annotation due to the protection of privacy. A common way to resolve this issue is to infer the socioeconomic information using census data from a user's inferred home location, either on municipality level or a more local area level [8, 14, 23, 31, 34]. The mobility can then be aggregated for all individuals with a home location in an area, and the aggregated mobility can then be compared between the areas to see if there exist any correlation between the socioeconomic factors and mobility.

Another approach is to divide the areas into groups and compare the aggregated mobility between the groups instead. The groups can either be created by dividing them based on some thresholds [23, 31] or by clustering the areas based on their socioeconomic attributes [14, 35] which is the approach adopted in the thesis.

3
Methods

This chapter describes the data from which the model is built in Section 3.1 and how it was pre-processed in Section 3.2. Sections 3.3 describes how trips were inferred. Section 3.4 describes how the most visited locations were found, a necessary step for defining home locations, as described in Section 3.5. Section 3.6 describes how mobility metrics were calculated. Lastly, Section 3.7 goes through the method used to cluster areas on socioeconomic factors.

3.1 Data sources

For the thesis, two different data sources were used. Firstly mobile data was used to construct a mobility model. Secondly, the socioeconomic dataset *Demographical Statistical Regions of Sweden* (DeSO), was used to divide VGR into smaller areas with socioeconomic information.

3.1.1 Mobile data

The mobile data was provided in an already aggregated format, similar to stops used in literature. The stops had been constructed by aggregating devices' consecutive mobility traces if they were within 100 m and 1 h of each other and had a displacement speed of less than 3 km/h. A description of a stop's attributes can be seen in Table 3.1.

Attribute	Description
Device id	Unique id of the device
X	Longitude, centriod of geolocations. EPSG:3006 coordinate encoding.
Y	Latitude, centriod of geolocations. EPSG:3006 coordinate encoding.
Arrive time	Timestamp of the first geolocations
Depart time	Timestamp of the last geolocations
Duration	Depart time - arrive time, NaN if a single point

Table 3.1: Description of a stop, the data type used to build the mobility model.

The mobility traces used to construct the stops had been generated by app usage from individuals aged 18 and above. The data was collected during the first week in October and the entire month of November for both 2019 and 2020. The data was generated by mobile apps which used either: GPS, Wifi, Celluar towers, or a fusion of them to approximate the location of the device. All users in the dataset had visited the Västra Götaland Region (VGR) in Sweden at least once. What is referred to as users are in fact devices, meaning that a single individual could own multiple devices. However, for the sake of simplicity, we will refer to them as users.

While the data for both 2019 and 2020 were collected by the same provider, the data exhibit significant differences both in the method used to approximate location and sparsity, as seen in Table 3.2 and Table 3.3. Indicating that the data provider themselves have changed their collection method. The 2019 data was significantly sparser than that of 2020. Furthermore, the location during 2020 was approximated by more precise methods.

Year/Method	Celluar towers	Fused	Gps	Wifi
2019	0.10	0.06	0.05	0.78
2020	0.01	0.31	0.67	0.01

Table 3.2: The method used to approximate the location changed from 2019 to 2020.

Year	Users	Stops (million)	Days with activity	Median stops of active day	Median duration (h) of active day
2019	42 000	1.9	295 000	4	1.59
2020	63 000	19.5	799 000	14	14.19

Table 3.3: Available data before pre-processing. An active day is a day containing at least one stop.

3.1.2 Demographic Statistical Regions of Sweden - DeSO

DeSO, a dataset provided by Statistics Sweden, divides VGR into almost 1000 areas, with a population of 700–2700 people in each area. Each area includes socioeconomic data about the population, such as age, gender, level of education, median income, work industry, and much more [36]. The DeSO data from 2019 was used.

3.2 Pre-processing

Pre-processing was done to 1) find a valid subset of users and stops and 2) adjust for some of the sparsity differences between 2019 and 2020. However, the aim was NOT to completely equalise the sparsity between the years, because the sparsity difference might not only be due to a change in collection method but could also be due to behavioural change both in mobility as well as in-app usage. Therefore, we applied the below filters to create more comparable datasets for 2019 and 2020 than directly using the raw data.

The pre-processing can be summarised in the following steps, which are explained more in-depth below:

1. Remove invalid points
2. Remove stops with short or no duration
3. Remove all users with home location outside VGR
4. Remove stops outside Sweden
5. Remove static users
6. Remove days with little activity
7. Sample 10 days per user

3.2.1 Invalid points

For all users, displacement speed between consecutive stops is defined as the distance between the two stops divided by the time difference between the departure time from the first stop, and the arrival time at the second. If the displacement speed between any two stops was larger than 1000 km/h (the speed of an aeroplane), both points were removed, similar as done in other studies [29].

3.2.2 Stops with short or no duration

The 2020 data had almost four times more stops per day than 2019 data, as seen in Table 3.3. One explanation for the difference in data sparsity is the large number of stops with a short duration in 2020 as seen in Appendix A.1. To adjust for this difference stops with a duration shorter than five minutes were filtered away since data sparsity affect the resulting mobility metrics. Five minutes was selected as a threshold since it has precedence as the minimum time duration of a stop [30]. Points with no duration were also removed due to the difficulty of interpreting them.

3.2.3 Users outside of VGR

Users in the dataset do not necessarily reside in VGR. A requirement to be included in the dataset is to at least visit the region once during the collection time period. Users residing outside VGR could be unrepresentative since they had at least one inter-regional trip from their home region to VGR. Therefore, all users with a home location outside VGR were removed. This was achieved by calculating the home location as described in Section 3.5 with the stops left after the two first processing steps.

3.2.4 Stops outside of Sweden

All stops outside of Sweden were removed since the coverage outside of Sweden was unknown.

3.2.5 Static users

As mentioned in Section 3.1, what is referred to as users are in fact devices. Devices, such as tablets, might never leave the house but short trips could still be incorrectly inferred due to noise in the location data. All devices with no stops 250 meters away from the home location were filtered away in an effort to remove these static devices.

One drawback of this filtering is that it also removes valid devices that are just inactive for the entire period, such users could be sick, in quarantine, or strictly following the COVID-19 guidelines. However, this step was kept since a higher share of users in 2019 than in 2020 exhibited this stationary behaviour. To ensure that the higher share of stationary users in 2019 was not due to sparsity differences, this step was tested as the last pre-processing step. The test replicated the earlier result, with 7% of the fully pre-processed users in 2019 deemed stationary while only 2% in 2020.

3.2.6 Days with little activity

Days with only a few short stops create a sparse picture of a user's day, making it difficult to accurately estimate mobility [37]. To ensure a minimum coverage of each day all days with less than five stops and less than 3 h of total duration were removed, the user day was kept if either of these conditions were met. This step also reduces the sparsity difference between the two years' data.

3.2.7 Sample 10 days per user

To make the two years more comparable, the number of days used to calculate the metrics was the same for all users. This was achieved by sampling all stops from 10 randomly selected days for each user, and users with fewer days were removed. To allow for trips over datelines, the trips were first constructed from all stops. The dates were then sampled from the stops and all trips starting on any of the sampled dates were kept. The dates were sampled randomly rather than consecutively since it should mitigate some of the effects of weather, which has been shown to affect mobility [27]. Consecutively sampling dates would risk sampling entire rainy or sunny weeks while the random sampling more likely catches some of both.

This step removed almost 80% of all remaining users for 2019, a lower threshold would allow for more users to remain. However, the selection of 10 days was deemed necessary to get a more complete picture of the actual mobility which contributes to a more reliable calculation of mobility metrics.

3.3 Trip construction

A trip was defined as a displacement between consecutive stops which were within eight hours of each other and at least 100 m apart. Eight hours were selected as a threshold due to it being the 99th percentile in travel time in VGR [38]. The threshold of 100 meters was selected for two reasons. Firstly it makes it consistent with the spatial threshold on which the stops data had been aggregated as described in Section 3.1. Secondly, 100 meters strikes a balance between locations approximated by highly accurate GPS and the less accurate cell towers. A higher threshold would lose correctly identified trips short trips and using a lower threshold could increase the risk of incorrectly identify trips for noisy cellular data.

3.4 Locations' visitation frequency

The locations a user had visited were calculated by clustering together stops using a spatial clustering algorithm, DBSCAN, with epsilon set to 100 m [39]. The locations were then defined as the centroid of each cluster. Lastly, the location's visitation frequency was defined as the number of stops in each cluster. The epsilon size of 100 meters was motivated for the two same reasons as described in Section 3.3.

3.5 Inferring home location

The home location was calculated with the stops remaining following the two first steps of pre-processing, meaning that invalid points and stops with a duration of less than five minutes had already been removed.

Of the two methods to infer home location described in Section 2.2.3, the first method using the most visited night-time location was selected. This method was expanded to also include the entire weekend since some users had none or very few night-time stops. Finally, the most visited location was calculated as the location with the highest visitation frequency. The visitation frequency was calculated as described in Section 3.4, using stops that occurred either during the weekend or between 17:00 and 08:00 on weekdays.

3.6 Mobility metrics

This section describes the calculation of the mobility metrics: trip distance, trips between regions, and temporal profile of locations. Two additional mobility metrics were calculated, radius of gyration and visitation frequency. Radius of gyration was calculated as described in the related work, Section 2.4.1, and the visitation frequency was calculated as described in Section 3.4.

3.6.1 Trip distance

The straight line distance between the start and stop location of each trip was calculated. Trip distances are underestimates as in reality, the route is likely not straight. The mean, median, and max trip distance were calculated for all users.

3.6.2 Trips between regions

The frequency of trips within VGR and between VGR and neighbouring regions, and non-neighbouring regions were calculated by building a OD-matrix with Swedish regions as nodes. A neighbouring region was defined as a region that shares a land border with VGR, a map of this can be seen in Figure 3.1.

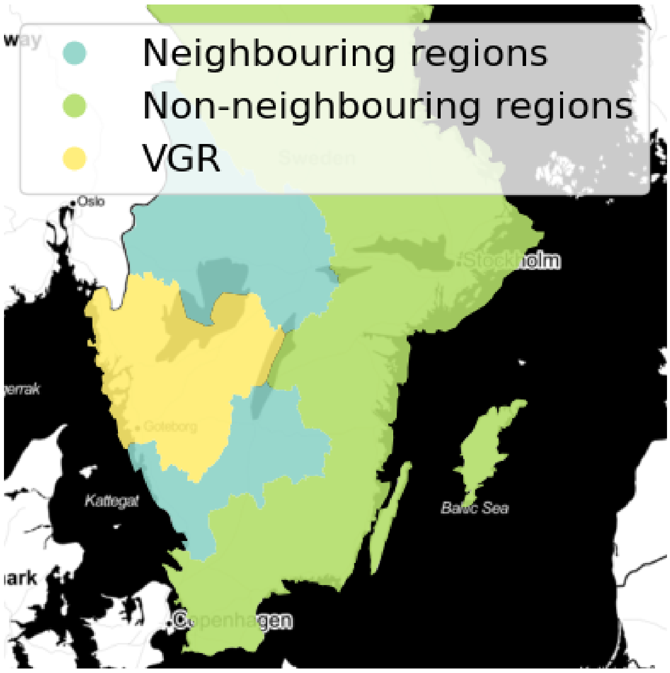


Figure 3.1: Map over regional categorisations.

3.6.3 Location temporal profile

The location temporal profile, described in Section 2.4.3, was calculated separately for the two most visited locations. For each location, the temporal profile was calculated with for the weekdays and the weekend separately, a one hour time interval was used. The weekday and weekend profile was then normalised, so they together summed to one. Furthermore, the temporal profile of the home and work location was calculated from the Swedish National Travel Survey for the years 2011–2016 [38]. Lastly, the temporal profile of the most and second most visited location was compared to the temporal profile of home and work location respectively [33].

3.7 Socioeconomic clustering

In order to associate socioeconomic attributes with the users' mobility metrics, the users' home locations were used to assign each user to a DeSO area [8, 14, 23]. Since our data were too sparse to be analysed on a DeSO level, the DeSO areas had to be clustered into larger groups. The clustering method chosen was K-means, which clusters objects into k clusters trying to maximise how similar the objects are within each cluster [40].

The DeSO areas of VGR were clustered based on socioeconomic properties such as age distributions, employment sectors of the population, and median income. All of the attributes with a description can be seen in Table 3.4.

Attribute group	**Attributes**	**Comment**
Age	0–4 years old, 5–9 years old, ..., 75–79 years old, 80- years old	
Sex	Male, Female	
Income	Median income	Only attribute that is not a percentage, rather the median income for the entire area.
Housing type	Tenant-ownership, Owned housing, Rented housing	Tenant-ownership refers to the Swedish term "bostadsrätt".
Education	Less than upper secondary school, Upper secondary school, Less than 3 years at University, More than 3 years at university	Upper secondary school refers to the Swedish educational institution "gymnasium".
Economic standard	Low economic standard	Low economic standard is defined by SCB as households with an economic standard lower than 60 percent of the median for Sweden's entire population.
Country of birth	Born in Sweden	
Employment	Not working	Percentage of population in ages 16–64 that are not working. Note that this is not the same as unemployment.
Work sector	Farming, forestry and fishing, Manufacturing and extraction, Energy and environment, Construction, Commerce, Transport and warehousing, Hotel and restaurant, Information and communication, Finance and insurance, Real estate, Business services, Public administration and Defence, Education, Health, social care and Social services, Cultural and personal services etc.	Percentages of the population in ages 16–74 working in a given sector.

Table 3.4: DeSO attributes for clustering the areas into groups. Unless otherwise is stated all attributes are percentages of the area's entire population.

All features were standardised by subtracting the mean and scaling to unit variance in order for each parameter to have equal weight in the clustering. Many of these variables suffer from being strongly correlated, which is an issue when performing K-means as two variables that are highly correlated would in effect get a larger weight. To address this principal component analysis (PCA), which is a method of changing the basis of data, was performed with the goal of removing any correlation between components while keeping 99% of the explained variance. The result was a decrease in dimensions from 45 parameters to 33 components, which is an added bonus as K-means works poorly in high dimensions. The resulting components from PCA were then used in the clustering process.

The initialisation of the K-means centroids can have a substantial effect on the quality of the outcoming clusters. With an unfortunate initialisation, the algorithm could finish at an unsatisfying local minimum. Two steps were taken to avoid this, the first is using the K-means++ initialisation algorithm which guarantees clusters that are $O(\log k)$-competitive of the optimal K-means solution [41]. The second step is that a few thousand different seeds for the random generator within K-means++ were tried to further optimise the solution slightly, selecting the seed that granted the best cluster cohesion score.

K-means clustering requires the parameter k or the number of clusters to be provided. In order to find the best k-value five commonly used methods were used. Each method consists of running K-means for different k-values and calculating a score measuring the quality of the cluster solution, in our case we tested $k = [2, ..., 14]$. The results of the five methods can be seen in Appendix Figure A.1, in combination they suggest that clustering on $k = 5$ clusters works well for our data. Although some scores suggest a lower k like $k = 3$ would be better, those were decided against as their clusters were deemed too broad both socioeconomically and geographically. For example, $k = 3$ would yield a countryside cluster and two urban clusters. So $k = 5$ was used for clustering.

4
Data description

This chapter describes and visualises the processed data and inferred trips and home location. The number of users and stops left after pre-processing can be seen in Table 4.1 as well as descriptive statistics about the remaining data. The difference in sparsity between the 2019 and 2020 data is significantly lower than before pre-processing, with the difference in the median number of stops shrinking from ten more stops in 2020 to just two more. A similar change can be seen regarding the median duration for each day, which was nine times longer in 2020 before pre-processing but less than twice as long after.

Year	Users	Stops	Trips	Median duration (h) of active day	Median stops of active day	Median trips of active day
2019	2 140	118 000	55 000	8.79	5.0	2.0
2020	11 960	950 000	369 000	15.87	7.0	2.0

Table 4.1: Overall model statistics after pre-processing and trip construction. An active day is a day containing at least one stop.

4.1 Trips

The distribution of total trips per user can be seen in Figure 4.1, there are more users in 2019 with very few trips, while 2020 has a larger share of users with 70 or more trips.

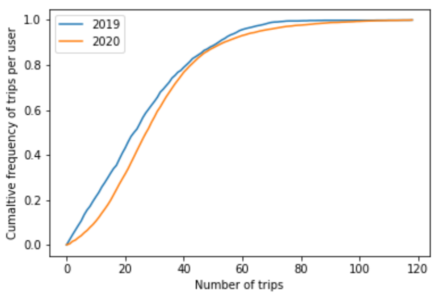
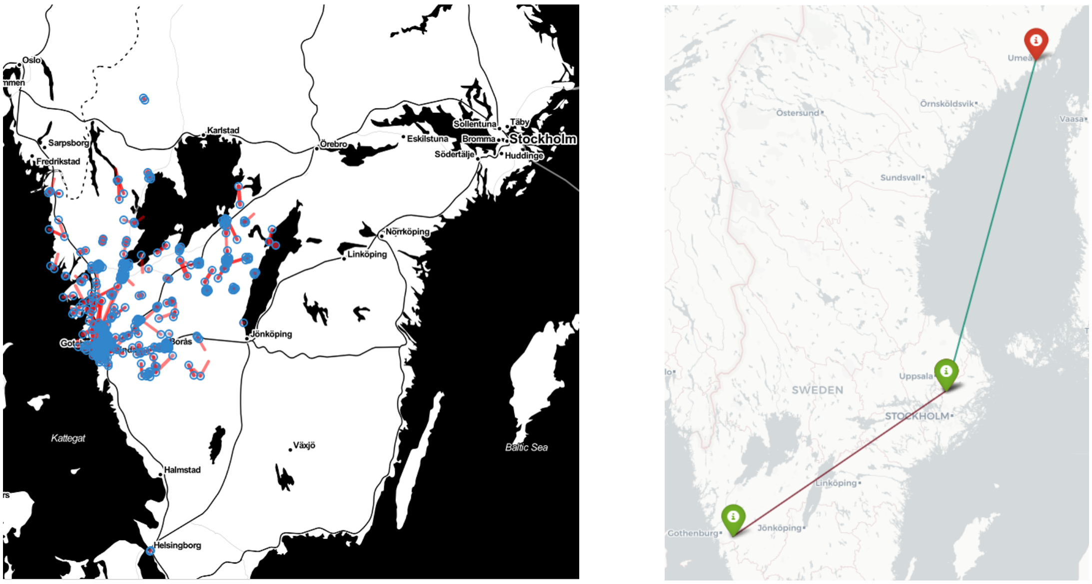


Figure 4.1: The cumulative frequency of trips per users.

Almost all trips are within VGR, as indicated by Figure 4.2 (a). Figure 4.2 (b) displays one of the limitations of the model, it interprets trips with transits longer than five minutes as multiple trips. The figure displays an additional limitation of the model, it infers the route of a trip as a straight line from the origin to destination. In reality, the route is likely not straight and therefore distance will be underestimated.



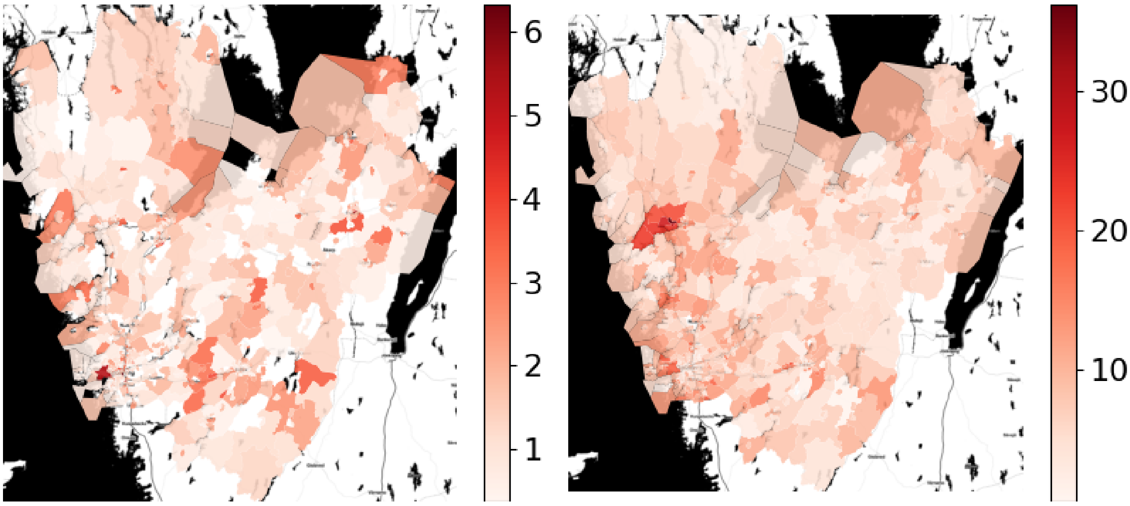
(a) OD-matrix with DeSO areas as nodes for 2019, only the 5% most popular edges are displayed

(b) Syntetic flight between Gothenburg - Umeå, with tranisit in Stockholm

Figure 4.2: Visualisation of trips and their locations **(a)** OD-matrix of 2019, almost all the 5% most popular edges are within VGR. **(b)** Synthetic example, a user flies between Gothenburg and Umeå with a transit in Stockholm

4.2 Home location

Figure 4.3 shows that the users are fairly evenly spread throughout the region except for some hotspots. Table 4.2 shows descriptive statistics of the relationship between the number of devices in each area and the actual population. Since many areas especially in 2019 have so few users further aggregation in the form of clustering was done.



(a) 2019 (b) 2020

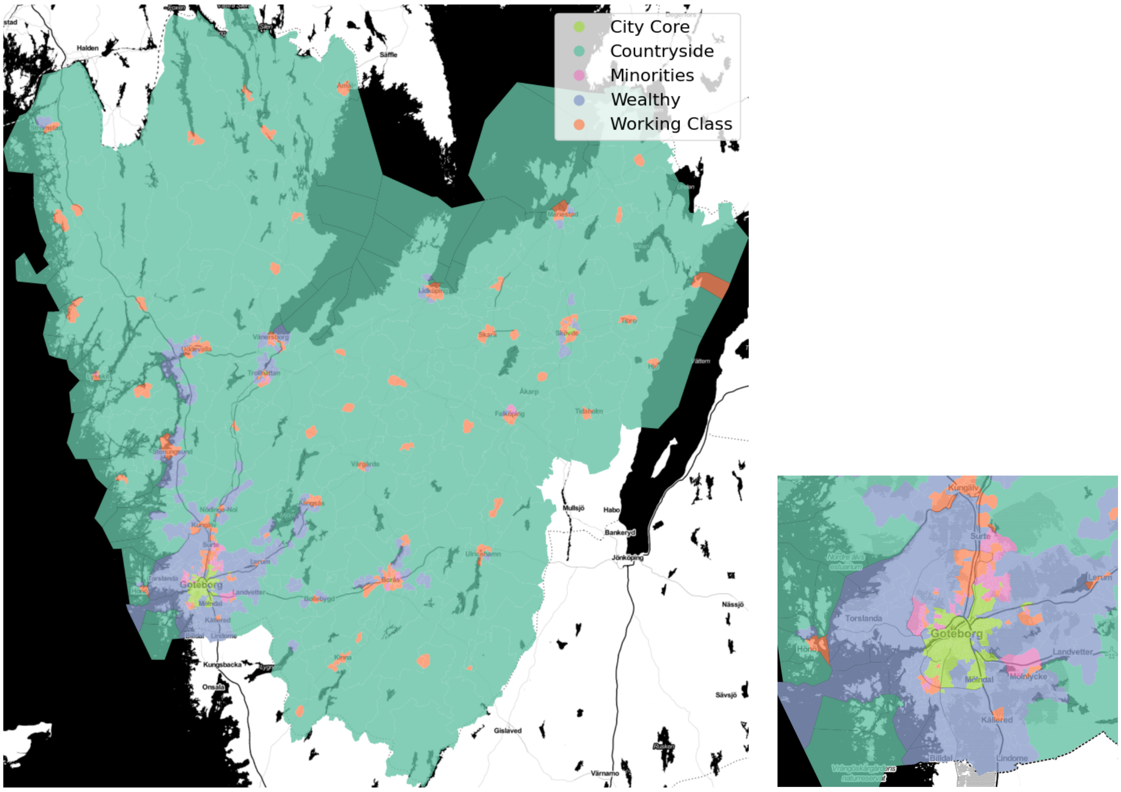
Figure 4.3: Users per 1000 area population for each DeSO area 2019 to the left and 2020 to the right.

Year		Users	Stops	Trips	Population	Users / 1000 Population
2019	min	1.0	14.0	0.0	767	0.37
	25%	1.0	68.0	28.0	1464	0.78
	50%	2.0	120.0	55.0	1753	1.28
	75%	3.0	188.0	89.0	2054	2.00
	max	14.0	866.0	467.0	3746	6.33
2020	min	1.0	33.0	4.0	767	0.49
	25%	7.0	538.8	209.8	1444	4.24
	50%	11.0	852.0	330.5	1727	6.36
	75%	16.0	1211.2	484.2	2036	8.82
	max	47.0	4611.0	1823.0	3746	36.20

Table 4.2: Descriptive statistics on mobility model for individual DeSO areas. 25%, 50%, 75% refers to percentiles.

4.3 Descriptive analysis of socioeconomic clusters

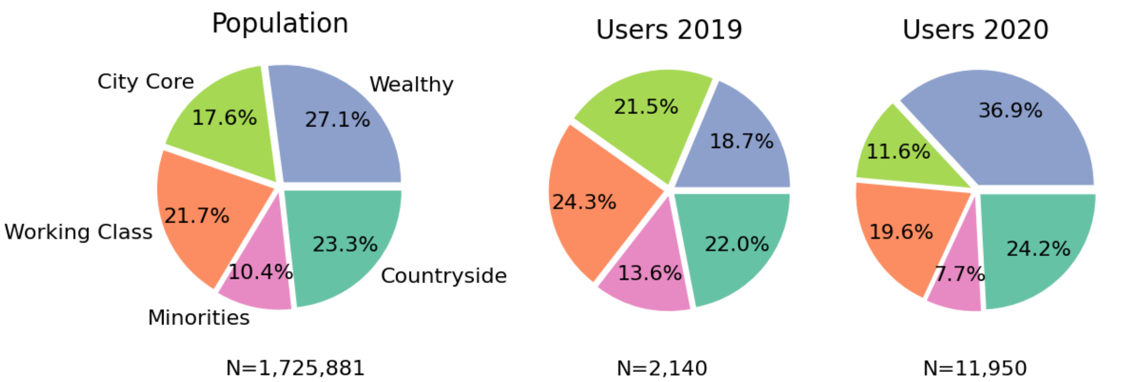
The five clusters resulting from the K-Means clustering is mapped in Figure 4.4. Though no geographic attributes were used for clustering, there are clear geographic concentrations of the clusters. Many clusters are mostly present in or around Gothenburg, the largest city and main urban centre in the Västra Götaland Region. The clusters were named based on how their populations' socioeconomic attributes are distributed (introduced below) and their geographic placement. That said the names are based on the authors and supervisors interpretation and are generalisations, everyone living in these clusters will not necessarily fit the name of the cluster.



(a) **(b)**

Figure 4.4: Map over DeSO areas in VGR coloured after the five clusters. All of VGR in figure (a), closeup of Gothenburg in (b).

The users are fairly evenly spread out among the clusters (Figure 4.5) following the true population fairly well. However, we observe one exception: in 2020 there are more users in the Wealthy cluster than expected both comparing to 2019 and the true population. A further breakdown of the mobility models descriptive numbers into clusters can be seen in Appendix A.2



(a) **(b)** **(c)**

Figure 4.5: Distribution of population (a) and users (b & c) among the five clusters.

Countryside is geographically the largest cluster and is mostly situated outside urban centres. Most people working within farming, forestry and fishing live in this cluster as can be seen in Appendix A.2. The Countryside cluster also has the largest share of people between 50 and 74 years old as can be seen in Appendix A.3.

Minorities is geographically mostly surrounding Gothenburg but also appears in some other urban areas (e.g. Borås, Trollhättan, Falköping). They stand out from the other clusters: it is the only cluster where most people are not born in Sweden and it is economically the most vulnerable with decisively more people living with low economic standard (Figure 4.6). The median income in the Minorities cluster is also significantly below the other clusters, which can be seen in Figure 4.7.

Wealthy is the largest cluster by population (Table 4.5), most people in this cluster live in houses they own (see Living in 'äganderätt' in Figure 4.6) close to cities. Their median income is the greatest amongst our clusters (Figure 4.7), they are the second most educated cluster (Figure 4.6), and they have more kids and people between 40-54 years old than most clusters which can be seen in Figure A.3.

Working Class is located in the suburbs of Gothenburg and in the centre of smaller towns and cities. The most distinctive feature of this cluster is that its median income lies between that of the Minorities cluster and the other clusters (Figure 4.7). Additionally, this cluster has the greatest share of people over the age of 80 (Appendix A.3).

City Core is almost exclusively located in the centre of Gothenburg except for some areas in a few other cities. This cluster has the greatest share of people with university education and people living in apartments (Figure 4.6). There are more people between 20–39 years in this cluster and fewer kids than the other clusters, which can be seen in Appendix A.3.

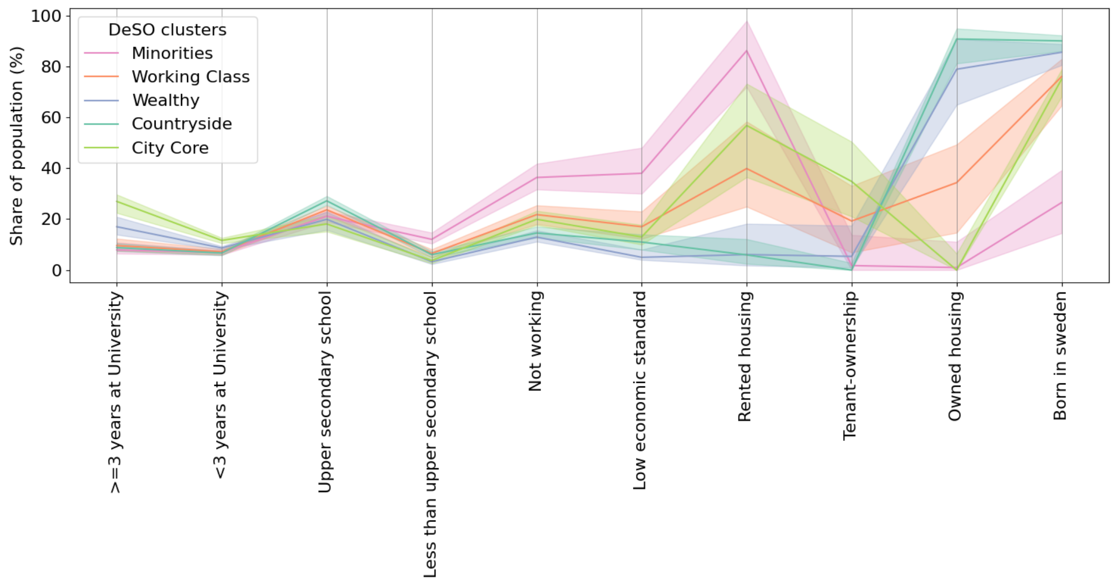


Figure 4.6: Comparison of cluster populations over a selection of the socioeconomic attributes. Centre line and the shaded area represent the three main quartiles (25th, 50th, and 75th percentiles). The attributes relating to education refer to highest level of education achieved.

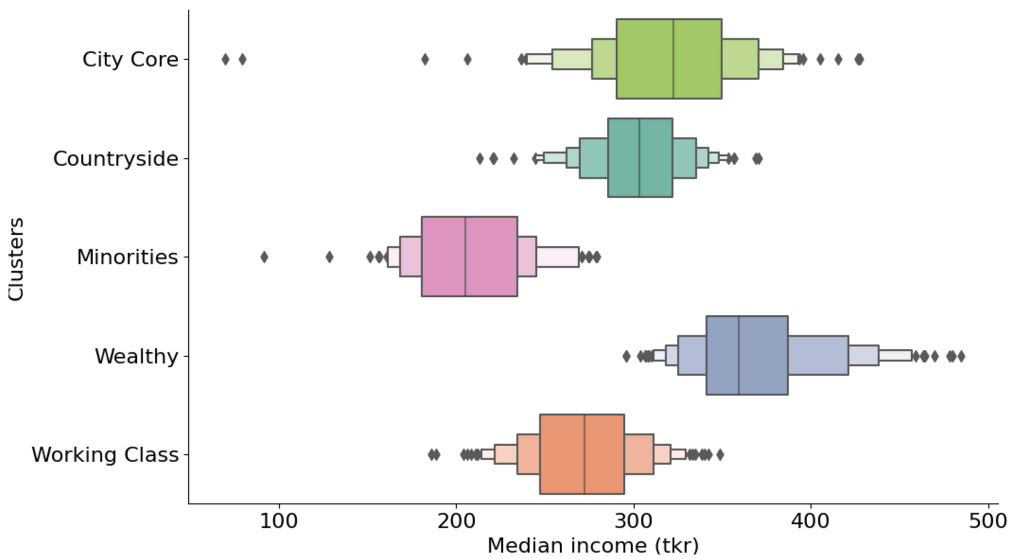


Figure 4.7: Median income distributions of the five clusters. Centre box is the same as in a box plot representing the three main quartiles, each smaller box contain half of the remaining data.

5

Results

In this chapter, Section 5.1 presents how the mobility metrics have changed between the years for the entire userbase. Section 5.2 goes one step further and presents how the mobility metrics have changed between socioeconomic clusters over the years. The socioeconomic clusters were introduced in Section 4.3.

5.1 Mobility metrics change - The whole population

This section presents how the mobility metrics have changed between 2019 and 2020 for all users collectively. Descriptive statistics of the results presented can be found in Appendix A.3.

5.1.1 Radius of gyration

Users radius of gyration changed by less than 0.5 km at the 25th or 50th percentile between the years as shown in Figure 5.1. However, the mean and 75th percentile radius of gyration decreased by 6.8 km and 6 km respectively, a 35% reduction for both.

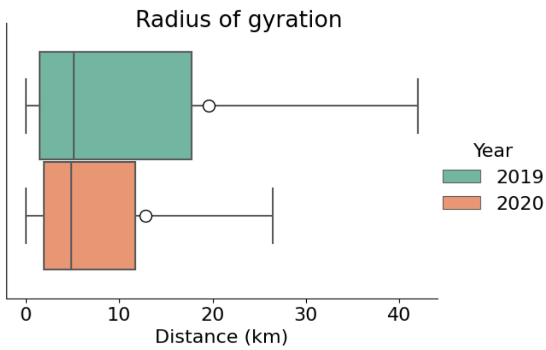


Figure 5.1: Distribution of radius of gyration (km) for the entire userbase displayed in a boxplot.

5.1.2 Trip distance

The distribution of mean and median trip distance for each individual can be seen in Figure 5.2. The median trip distance increased at all percentiles. The mean trip distance increased at the 25th and 50th percentile while decreased at the 75th and mean.

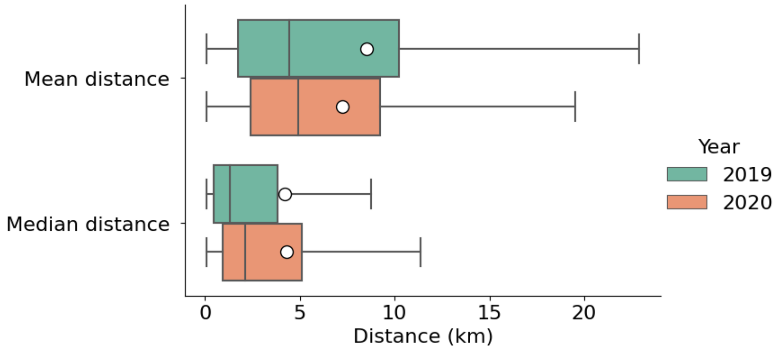
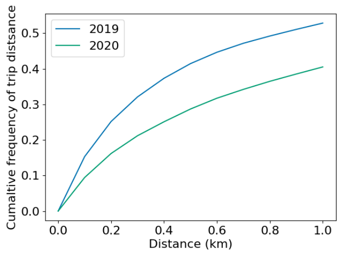
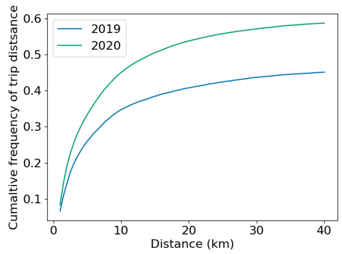


Figure 5.2: Distribution of all users mean and median trip distances.

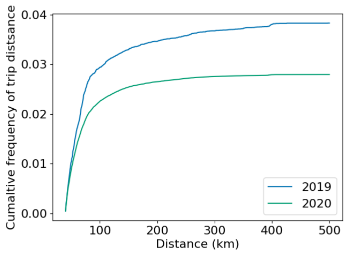
Three different trends are visible when looking at the cumulative frequency of trip distances in Figure 5.3. There are more very short trips (0.1–1km) in 2019, while short and medium (1–40km) trips are more frequent in 2020. Trips longer than 40km are more frequent in 2019, with the difference in cumulative frequency is continuously growing with the trip distance.



(a) Distance range: 100m - 1km



(b) Distance range: 1km - 40km.



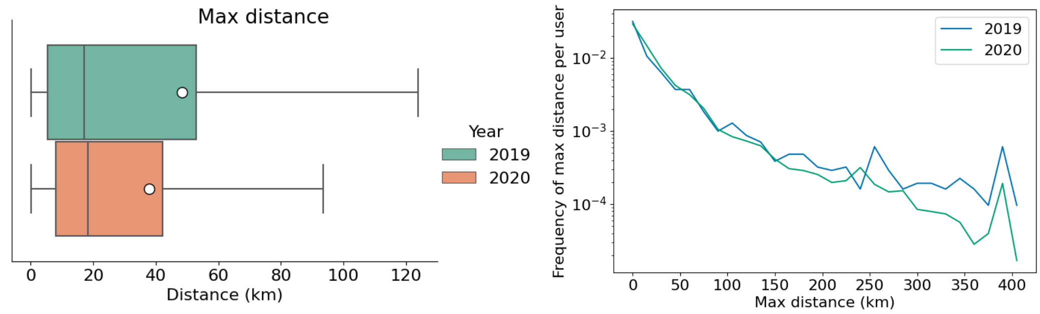
(c) Distance range: 40km - 500km.

Figure 5.3: Cumulative frequency of all trip distances. Each figure shows a different range of distances. The cumulative starts at 0 for each range. The frequency is calculated on all trips, not only the trips within the range.

The distribution of the user's maximum trip distance can be seen in Figure 5.4, there is a substantial decrease at the 75th. Furthermore, Two spikes are visible Figure 5.4 (b). These spikes correspond to the distance between Gothenburg–Stockholm (397km) and Gothenburg–Malmö (243km), the largest and third-largest city in Sweden respectively. The spikes decrease during 2020, indicating that fewer users are travelling from Gothenburg to the two other large cities in Sweden, Stockholm and Malmö.

5.1.3 Trips between regions

Trips between VGR and other regions decreased during 2020 as seen in Figure 5.5. The decrease is larger for VGR–non-neighbouring (63%) regions than VGR -



(a) Boxplot of users' maximum trip distance

(b) Frequency of users' maximum trip distance.

Figure 5.4: **(a)** shows a distribution of users' maximum trip distance for both 2019 and 2020. **(b)** displays the frequency of users' maximum trip distance.

neighbouring regions (15%).

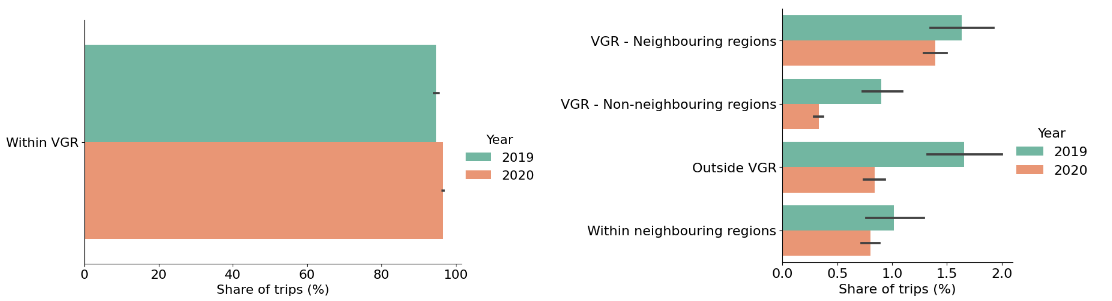


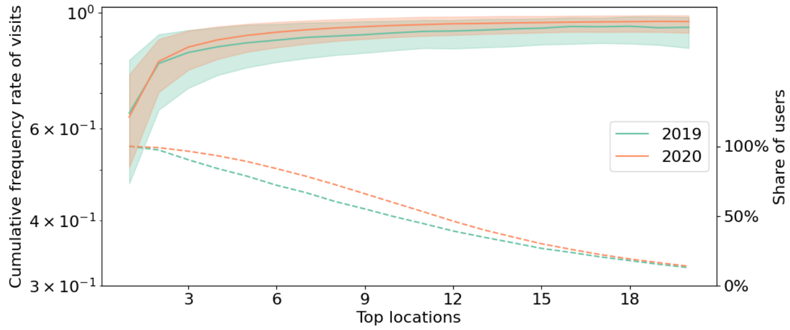
Figure 5.5: Trips between VGR and other regions have decreased, with the biggest decrease being between VGR and non-neighbouring regions.

5.1.4 Visitation frequency

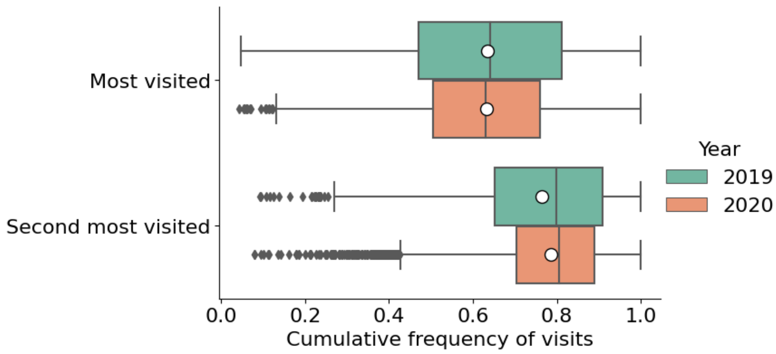
Figure 5.6 (a) shows the cumulative visitation frequency of 20 locations ordered by their rank. The cumulative visitation frequency increased at all percentiles, indicating that users in 2020 have more concentrated visits to a limited number of locations. The difference between the years is small, usually within 5 percentage points. Users in 2020 tend to visit more unique locations in total, increasing the median number of locations visited from 9 to 11. Figure 5.6 (b) displays the visits to the two most visited locations more closely. These two locations are an exception to the trend mentioned above since the 75th percentile is larger for 2019 and there is no difference at the 50th percentile.

5.1.5 Location temporal profile

The temporal profile for the most visited and second most visited location can be seen in Figure 5.7. For both locations, the weekday temporal profile follows



(a) Cumulative frequency of 20 most visited locations, dotes lines indicate share of user visiting at least x locations. Shaded area represents 25th and 75th percentile.



(b) Distribution of cumulative frequency of visits to top two locations.

Figure 5.6: Cumulative frequency of most visited locations.

the shape of the baseline which was calculated from a travel survey, introduced in Section 3.6.3. However, this is not the case for the weekend. There is only a small difference for the most visited location between the years, the year 2020 follows the baseline more closely 00–15 for weekdays while 2019 find a better fit after that. For the second most visited location, the year 2020 has the most activity around 06–16. The temporal profile of all location can be seen in Appendix A.8, 2020 have a flatter curve with about the same amount of activity between 06–22 while 2019 had more activities in the evening and less early morning.

5.2 Mobility metrics change - Socioeconomic breakdown

In this section, the results from Section 5.1 are broken down for each socioeconomic group introduced in Section 4.3. Descriptive statistics for the changes described can be seen in Appendix A.4.

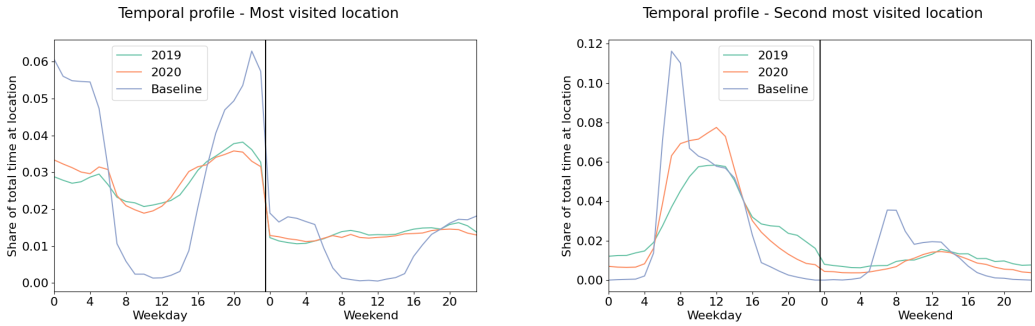


Figure 5.7: Temporal profile of most and second most visited location respectively.

5.2.1 Radius of gyration

As seen in Figure 5.8, the median radius of gyration decreased for all clusters except Minorities and City Core, while the 75th percentile decreased for all clusters except Minorities. The decrease was found to be stronger at the 75th percentile, decreasing 30–43% while the median shrank with 10–30%. The radius of gyration increased at the 25th percentile for most clusters, but by no more than 0.5 km.

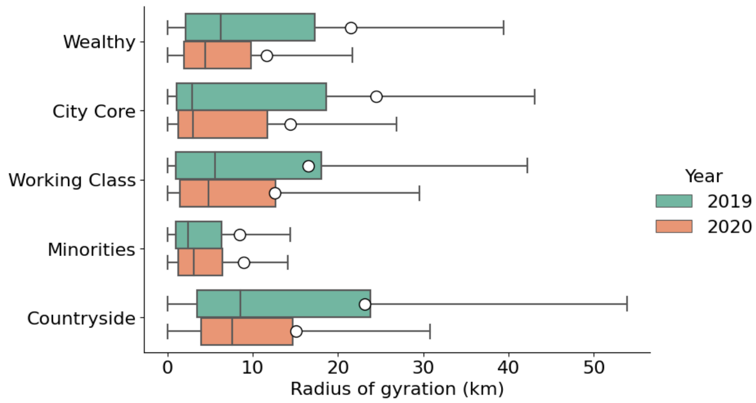
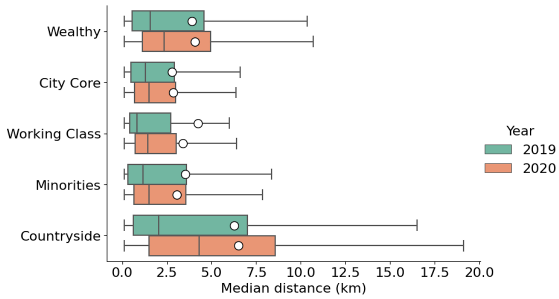


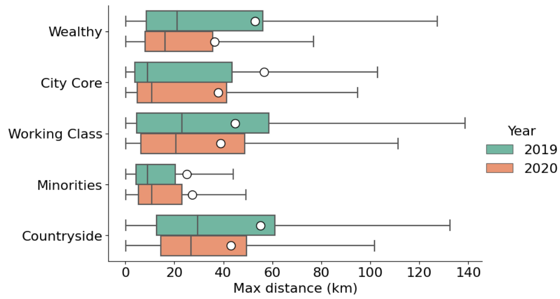
Figure 5.8: Box plot of radius of gyration for each cluster. Outliers are removed from the plot.

5.2.2 Trip distance

The median trip distance experienced an overall increase at 25th and 50th percentile for all clusters during 2020, and an increase at the 75th percentile for all clusters except City Core and Minorities as seen in Figure 5.9 (a). The Wealthy cluster saw a reduction in maximum trip distance at the 50th and 75th percentile while both the Working class cluster and Countryside saw a smaller reduction as seen in Figure 5.9 (b). The Minorities cluster saw a small increase in maximum trip distance travelled. Mean trip distance can be seen in Appendix A.10.



(a) Median trip distance

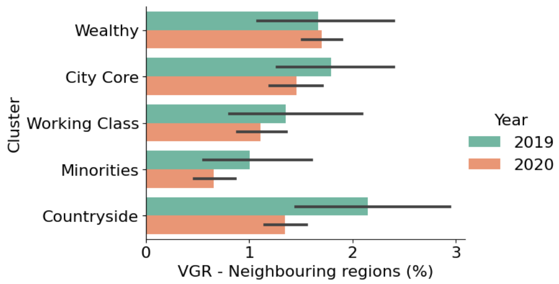


(b) Max trip distance

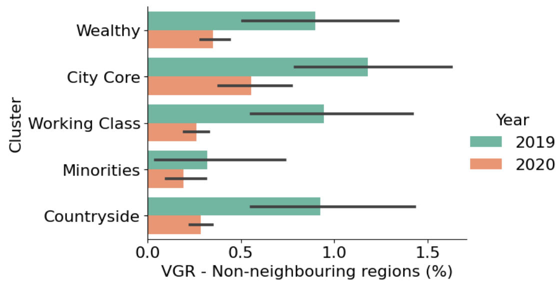
Figure 5.9: Distribution of users median (a) and max (b) trip distance for each cluster.

5.2.3 Trips between regions

Trips between VGR and other regions can be seen in Figure 5.10. There is a clear distinction between VGR trips with neighbouring regions and trips with non-neighbouring regions, as there is a substantially sharper decline in trips with non-neighbouring regions than trips between VGR and neighbouring regions. The reduction in travel between VGR and neighbouring regions are within the margin of error for all clusters. The Wealthy cluster sees no decline in trips between VGR and neighbouring regions. Figures for trips within VGR, within neighbouring regions, and outside VGR can be seen in the Appendix A.11.



(a) Share of trips between VGR and neighbouring regions



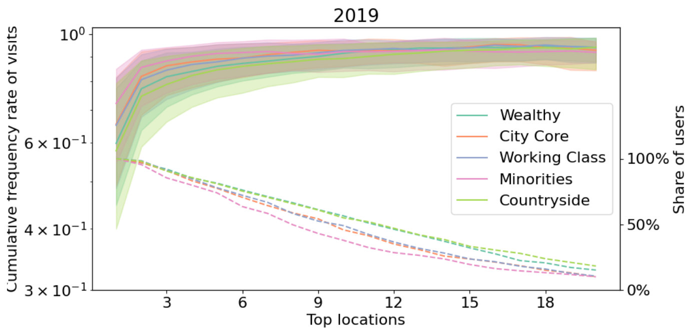
(b) Share of trips between VGR and non-neighbouring regions

Figure 5.10: Trips going between VGR and other regions.

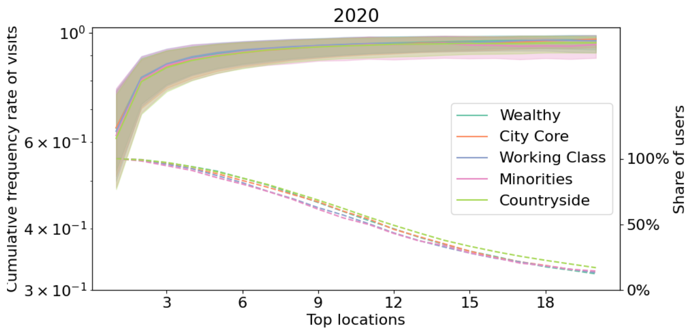
5.2.4 Visitation frequency

The cumulative visitation frequency homogenised between the clusters during 2020 as seen in Figure 5.11. The minority cluster changes from having the highest share of their visits to the most visited location to the lowest share, as displayed in Figure 5.11 (c). The homogeneity becomes evident in Figure 5.11 (d), there is a very low variance at 25th, 50th, and 75th percentile between clusters during 2020.

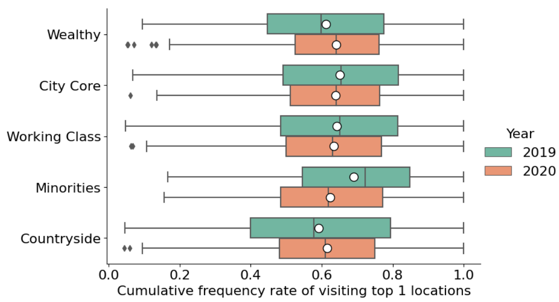
The change in cumulative visitation frequency for the Minorities and the Wealthy cluster can be seen in Figure 5.12, the same figure for all cluster can be seen in Appendix A.12. The Wealthy and Countryside cluster have the greatest reduction in visits to their less visited locations. Users from these clusters also have the lowest increase in the number of unique locations visited. The Minorities cluster have the lowest reduction in visits to their less visited locations and the greatest increase in the number of unique locations visited.



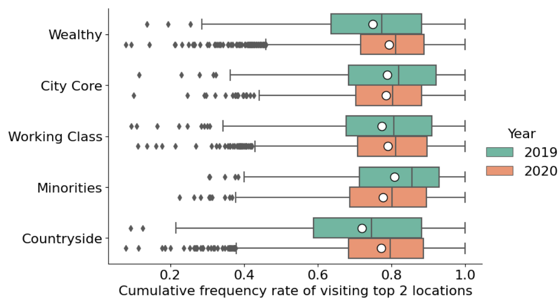
(a) Top 20 locations 2019



(b) Top 20 locations 2020



(c) Most visited location



(d) Top 2 locations

Figure 5.11: The cumulative frequency of each users visits to the 20 most visited location for 2019 (a) and 2020 (b). The dotes line represent the share of users who have visited at least x locations. (c) Boxplot of frequency of visits to most visited location. (d) Boxplot of the cumulative frequency of the two most visited locations.

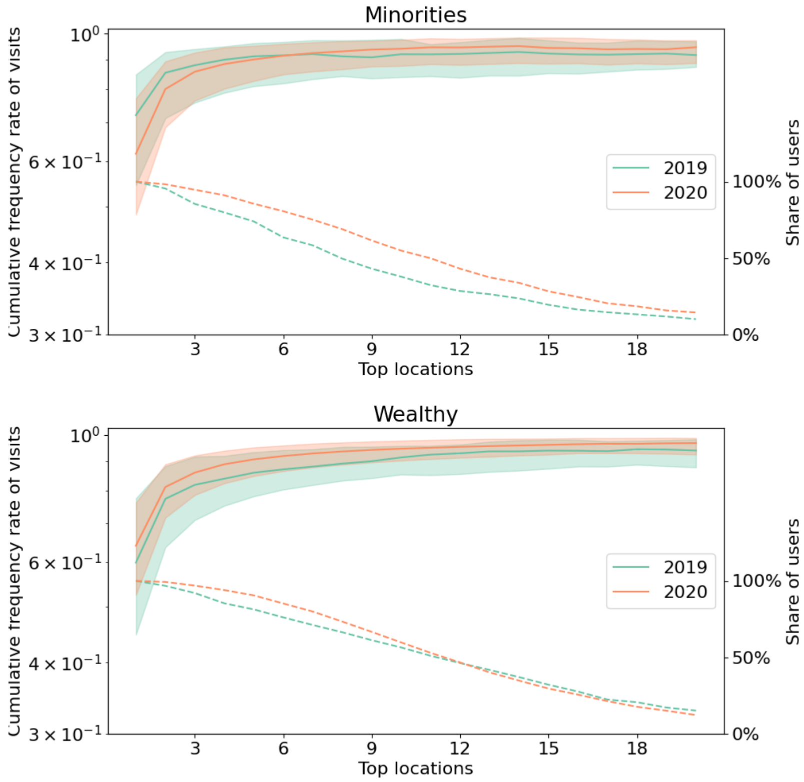


Figure 5.12: The cumulative frequency of visits for the minorities and wealthy cluster.

5.2.5 Location temporal profile

The temporal profile for all clusters for the most visited and second most visited location can be seen in Figure 5.13. Most of the difference between the clusters can be seen during the weekdays, there are only small differences between the clusters during the weekend. The differences between the clusters reduced during 2020. The wealthy cluster has the strongest dip between 6–16 for their most visited location while having the strongest peak at the same for the second most visited location.

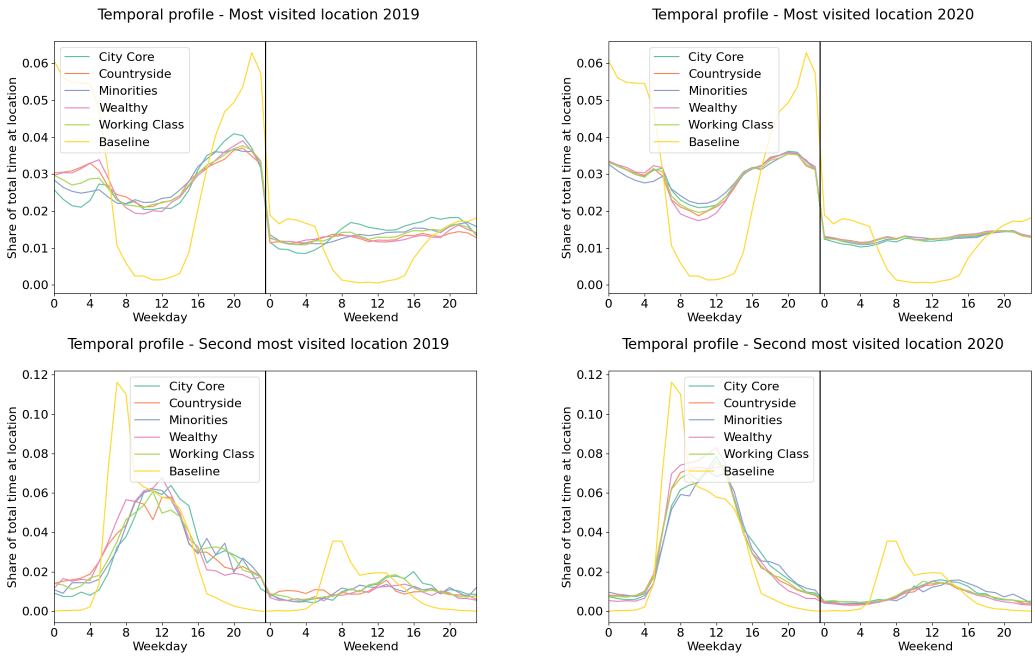


Figure 5.13: The temporal profile for the most visited and second most visited location for all clusters.

6
Discussion

The limitations of the mobile app and socioeconomic data are discussed in Section 6.1. Section 6.2, discuss how the socioeconomic clusters relate to cluster found in another Swedish study which used the same socioeconomic dataset. Section 6.3, covers the results compare to previous studies. Lastly, the validity of the result is discussed in Section 6.4.

6.1 Limitations

The dataset might contain underlying biases for several reasons. First, it was collected based on the usage of a variety of undisclosed apps, where the only notation of a user is an anonymised id. It is unknown whether the collection of geolocations associated with app use are representative of the mobility of the whole population. This potential bias issue has not been handled in the thesis.

While the same data provider is used for both the 2019 and 2020 data, the data has significant differences in the method of collection between the years as we discussed in Section 3.1. As a result of different data collection methods, the 2019 data is significantly sparser. Preprocessing and mobility metrics were adapted to mitigate these differences to some degree. However, it is unknown whether the change in collection method is solely responsible for the difference in the observed sparsity or if some of it was due to a change in human behaviour caused by the COVID-19 pandemic (e.g. more frequent app uses during the pandemic). Hence, the preprocessing method we introduced here does not aim to completely equalise the sparsity between the years, since it would risk removing new behaviours caused by the pandemic. In the end, this uncertainty could make the results presented here less conclusive. However, since these sparsity differences affect all clusters, the relative changes between the cluster should still be valid.

Lastly, the available attributes in the socioeconomic data are limited, thus, making it difficult to determine which attributes affect mobility. If an attribute is found to affect mobility, it could be due to an unknown attribute that is correlated with it. For instance, one such correlation could be university education and the possibility to work from home. It is likely the latter that has a direct impact on mobility change during the pandemic. The use of indirect variables can introduce additional uncertainty and reduce the validity of our results.

6.2 Socioeconomic Clustering

Our study uses five socioeconomic clusters, which incidentally has the same number of clusters as Almlöf (2020) [14]. Their study of Stockholm that performed similar clustering on DeSO attributes used a different clustering method namely two-step clustering from IBM's SPSS tools. A big difference from K-means is that it determines how many clusters to find on its own. Since we are using the same dataset we clustered largely on the same attributes, with some differences like the employment sector in our study and population density in theirs.

Their five clusters and ours are comparable in their socioeconomic attributes with just some minor differences like they describe their Rural cluster as having more children which is not the case for our Countryside cluster, our City Core cluster seems to be even younger than their Central cluster with the greatest share of 20–39-year-old versus their greatest share of 30–39-year-old, their Garden suburb is the most educated cluster while our otherwise very comparable Wealthy cluster is not. But these differences are likely due to regional differences, like Stockholm County being more urban with a population density of 367 people/km^2 versus VGR's 73 people/km^2 [42].

6.3 Mobility changes

The results indicate a substantial decrease in mobility during 2020. For most measured metrics the decrease is primarily observed at the 50th percentile and higher. Furthermore, the clusters with the highest mobility in 2019 made larger reductions in mobility, both in absolute and relative terms. These two factors combined indicate that it is primarily individuals who were highly mobile pre-pandemic who reduced their mobility. How this conclusion is reached as well as how the results relate to other studies is discussed below.

6.3.1 Mobility change by mobility metric

In this section, the results of each mobility metric are discussed.

Radius of gyration
The 5% reduction in median radius of gyration during 2020 is substantially lower than the 47% reduction found during lockdown in Italy [43]. However, the mean reduction of 35% is similar to the 30–40% reduction observed in Spain during November 2020 [18]. A possible explanation for the differences is that lockdowns heavily affect the average individual while only individuals with high pre-pandemic mobility reduce their mobility during softer rules. This argument is further supported when investigating which socioeconomic groups reduce their radius of gyration. The clusters with larger radius of gyration reduce it to a greater extent and the reduction is larger at the 75th percentile as seen in Figure 5.8. The cluster with the lowest pre-pandemic radius of gyration, Minorities, shows no reduction and City Core, which had a low median radius of gyration, only shows a reduction at the 75th percentile.

The Minorities cluster have a lower radius of gyration at the 50th and 75th percentile than the other clusters, even without reducing their radius during 2020. The large difference at the 75th percentile between the City Core and the Minorities cluster is curious since the clusters appear in the same cities. Radius of gyration has been found in the other studies to differ within cities, where the less wealthy are found to either have the highest or lowest radius of gyration. However, the study found a difference of about 50% at the 75th percentile between the poorest and wealthiest neighbourhoods while it is close to 300% between the Minorities and City Core cluster [34].

The reduction in radius of gyration is likely driven by the reduction of longer-distance trips. Long trips create a stop far away from a user's centre of mass and have a large effect on the metric since the radius is calculated with a power term. The reduction of longer-distance trips is evident in Figure 5.3.

Trip distance
The max trip distance decreased while the median trip distance increased during the pandemic, as displayed in Figure 5.9. The mean trip distance shows only small changes at the 50th percentile but decreased at the 75th percentile, except for the Minorities cluster which had a small increase as seen in Appendix A.10.

As discussed above in *radius of gyrations*, long trips and radius of gyration are related. The observed decrease in maximum trip distance is consistent with the decrease in radius of gyration observed in this study as well as other studies [18, 43]. Similarly, a French study conducted during lockdown also show a reduction in number of trips longer than 100 km [44], which is also observable both in Figure 5.3 and 5.9. Consistent with previous results in the thesis, the reduction is larger for the study conducted during lockdown at 85% compared the 40% reduction in trips over 100 km seen in this study.

The increase in median trip distance is explained by a decrease in very short trips during 2020, with about 30% of trips being 250 meters or less in 2019 while only 20% in 2020. One possible explanation could be the decrease in visits to local restaurants and transit stations [4, 45]. However, this decrease could also be a consequence of using the same spatial threshold for both years. The 2019 data likely has more noise in the location data, since the 2019 location data approximated by less accurate sources, Wifi and cellular networks while the 2020 data is inferred almost exclusively from the more accurate GPS and fused locations. Therefore, 2019 data is more likely to infer very short trips from noise in the location data.

Lastly, some of the implementations of the model lower the accuracy regarding the lengths of trips. As mention in Section 4.1, the model can infer a single trip into multiple trips if a transit is longer than five minutes. Also, the lengths of trips are underestimated since the length of trips are calculated as a straight line distance. Furthermore, the static maximum time of inferring a trip regardless of the distance of the trip makes it impossible to infer trips with a travel time of more than 8 hours. It also makes it difficult to infer multiple hour-long trips in general, since a user must not necessarily use their device directly at the end of the trip. All these three factors contribute to more uncertain aggregated trip distances.

Trip between regions
Trips between VGR and other regions declined during 2020, as seen in Figure 5.5 and Figure 5.10. The decrease of 23% in trips between regions is smaller than observed in Italy and France during lockdown where the inter-regional trips decreased by 50% and 65% respectively [43, 44]. The decrease was larger for the share of trips between VGR and non-neighbouring regions which decreased by 80% than for between VGR and neighbouring regions which only was a reduction of 15%. A possible reason for this difference could be the greater reduction in non-commuting trips than commuting trips and more users commuting between VGR and neighbouring regions and fewer commuting between VGR and non-neighbouring regions.

Another reason for the sharp decrease in trips between VGR and non-neighbouring regions could be previous government recommendations to travel by car, and no longer than 2 hours. Few cities in non-neighbouring regions are reachable from the population centres of VGR within 2 hours. While this recommendation was no longer in effect during November 2020, it could still affect the behaviours of individuals [46] as the pandemic was still was ongoing.

Visitation frequency
In general, the visits were more concentrated on a few top locations during 2020 while the visits were more spread out during 2019, as displayed in Figure 5.6. This makes sense given the government's recommendations. Furthermore, the clusters exhibited differences in cumulative visitation frequency during 2019. The Countryside and Wealthy cluster visited more locations and spread out their visits to different locations to a larger extent, as seen in Figure 5.11. However, these differences were almost erased during 2020, with almost identical shares of visits to the two most visited locations, and small differences at the less frequently visited locations. This change was mainly driven by the Countryside and Wealthy clusters which reduced their visits to their less frequently visited locations to a greater extent than the other clusters, as seen in Figure 5.12 and Appendix A.12. The opposite is true for the number of unique locations visited, the clusters where the users visited the fewest unique locations in 2019 see the largest increase during 2020.

However, some of the changes during 2020 go against our common understanding of what's being reported in the literature. In particular, the number of unique location visited by individuals increases during 2020, as well as some cluster decrease their share of visits to the most visited location. A possible explanation for these disparities is the data sparsity difference. With more data points in 2020 the likelihood of finding more unique locations increases. Furthermore, a sparse dataset might overestimate the visits to a location where an individual visits for a long duration, since a stop can be inferred even if the geolocation records a far apart in time. Individuals are found to stay at the most visited location for a long duration, with 14 hours as the most common stay time [32]. Hence, making it easy to find with a sparse dataset.

Temporal profile
During the pandemic, the Swedish health agency issued a recommendation to work from home and stay home as much as possible [47]. Furthermore, both temporally and permanent layoffs were prevalent [48]. With this in mind, the hypothesis was that the temporal profile would change for the most visited and second most visited location since they are typically home and work locations, respectively. The hypothesis is that these changes would make the temporal profile of the most visited location flatter since individuals are more likely to be at home during the daytime. Furthermore, the second most visited location would no longer be the workplace for individuals who reduced or stopped their visits to their workplace. Therefore, the resulting temporal profile of the second most visited location would be a mixture of the temporal profile from both workplaces as well as other commonly visited locations such as grocery stores, bus stops and gyms. An additional hypothesis was that these changes would be more prominent for some socioeconomic groups, such as the Working Class which could have been more heavily affected by the lay-offs or the higher educated clusters since it is more likely that their jobs could be performed from home.

However, our results do not confirm this hypothesis. The results instead indicate small differences between the clusters and decreased shares of visits to the home location during office hours and an increase of visits to the work location during the same hours.

The increasing share of visits occurring during office hours for the second most visited location is difficult to interpret since it could have different causes. One possible interpretation is that the reduction of workplace visits have mostly affected non-office hours. Another possibility could be that the second most visited location naturally becomes a daytime weekday activity. This is since the temporal profile of the home location have a large dip in visits during 6–17 weekdays, meaning that the user is away during these hours. Since this is the largest dip the most visited location during this dip often becomes the most visited location except for the home location, I.e. the second most visited location.

6.3.2 Mobility change by socioeconomic cluster

In the following section, we briefly summarise the general mobility trends observed during 2020 for each socioeconomic cluster as well as any trends seen between multiple clusters. The results indicate a difference in mobility between the clusters before the pandemic. Our study suggests that these differences seem to decrease during the pandemic.

The results from US and Colombian studies indicated that wealthier areas had lower mobility than poorer areas during the pandemic [8, 19]. This result was not replicated in our thesis. The least wealthy cluster, Minorities, still had the lowest mobility during the pandemic. There are a few possible explanations. Firstly, our observed differences in mobility pre-pandemic are larger than in the US study, where the mobility differed by about 10 percentage points. Secondly, the studies were conducted in the countries that issued lockdowns and even banned leisure travel in some cases. No bans on leisure travel were issued in Sweden, hence our result contains both essential and leisure travel while other studies might contain only essential travel.

As mentioned at the start of Section 6.3, the reduction in mobility seemingly correlates with the pre-pandemic mobility for each cluster. With few exceptions, the cluster with the greatest pre-pandemic mobility had the greatest reduction, and the upper percentiles show greater reductions than the lower. If the clusters would be ranked in order of most mobile for each metric, few clusters would change order during 2020. One exception to this rule is the Wealthy cluster, which has the largest reduction in radius of gyration and max trip distance, while only having the second or third largest pre-pandemic mobility. However, the reduction only differs by a few percentage points. Therefore, our main conclusion is that socioeconomic factors have a greater impact on mobility than on the changes of mobility during the pandemic. This argument is further supported when observing median trips per user in Table A.2, the differences between the cluster decrease during 2020.

Wealthy
The Wealthy cluster was one of the more mobile clusters before the pandemic. In accordance with other studies, the cluster saw a larger reduction in mobility during COVID-19 [8, 19].

City Core
The City Core cluster show some similarities with the Minorities cluster at the 50th and lower percentile for radius of gyration and the trip distance metrics. However, the City Core had higher mobility at the upper percentiles.

Interestingly, while the cluster reduced the radius of gyration and mean trip distance at the 75th percentile, the median and max trip distance stayed the same at the 75th percentile. This would indicate that many users still travel far but they do it less frequently.

Working Class
The Working Class cluster is difficult to analyse due to its spread. Some of it is located around the suburbs or centres of major cities, whereas other parts of the cluster are located around minor population centres in the countryside. The cluster reduced their mobility for every metric to a similar extent as the other metrics.

Minorities
The Minorities cluster showed the smallest reduction in mobility, which is consistent with other studies [8]. However, the cluster still had the lowest mobility during 2020,

since it had significantly lower mobility than the other cluster before the pandemic.

Countryside
The Countryside was the most mobile cluster in 2019, with the highest mobility in almost every metric. A study found that the states with the lowest population density in the US have the most miles travelled per day, both before the pandemic, as well as during its first weeks [49]. While total distance travel during a day is not measured in the thesis, the high mean trip distance and highest median trips per user indicates that is the case. The Countryside cluster had the most significant increase in median trip distance, indicating that the cluster reduced trips within its local community.

6.4 Model validity

While no formal validation steps were performed due to time concerns, the result shows promise. As discussed in Section 6.3, much of the result is in line with previous studies. Our results are also consistent with some of the basic expectations, including the mobility reduction during 2020 and that the Countryside cluster has a large radius of gyration and a long mean trip distance.

However, some mobility metrics indicate an increase in mobility for at least some clusters, which goes against both the authors' expectation and other studies. These increases, as discussed in Section 6.3.1, could potentially be a result of the sparsity differences between the years.

The general conclusions regarding the observed reductions between the years and the observed differences in relative mobility change between the clusters should still be valid. However, one should be careful when interpreting the exact values of mobility change.

7
Conclusion

The COVID-19 pandemic caused a reduction in mobility worldwide. While many countries relied on lockdowns, Sweden relied on recommendations to curb the spread of the virus. This makes Sweden a unique study area to study mobility changes during the pandemic. The pandemic did not affect all socioeconomic groups equally, with some groups having a higher risk of infection and mortality. In some cases, this increased risk could be predicted by mobility.

Five mobility metrics were calculated using a big dataset of geolocations generated by app usage in the Västra Götaland region in Sweden during October and November for both 2019 and 2020. The following metrics were calculated for both years: radius of gyration, trip distance, trips between regions, visitation frequency, and temporal profile. Moreover, five clusters of areas representing socioeconomic groups were generated by clustering 1000 areas based on their socioeconomic attributes. The socioeconomic information was collected from an official socioeconomic dataset, DeSO. Clustering was performed with K-means on socioeconomic attributes such as age, income, and education with a k-value selected based on cluster evaluation scores. This resulted in the five clusters: Wealthy, City Core, Minorities, Working Class, and Countryside. Users were then assigned to the clusters based on their inferred home location and the mobility metrics were aggregated for all users within the cluster.

The results show reductions in mobility during the pandemic, including the radius of gyration, long trips, and trips to other regions for the entire population as a whole. The results also indicate that users tend to concentrate their visits to fewer locations during the pandemic. However, the number of unique visited locations have increased. This increase could be due to the difference in data sparsity, where 2020 contained 45% more stops after pre-processing. Additionally, the temporal profiles of the most and second most visited locations provided no conclusive insights. The reduction in mobility was substantially smaller than the reduction found in other countries during lockdown, the median radius of gyration decreased by 5% compared to 47% and trips longer than 100 km reduced by 40% compared to 85%.

We found that mobility decreased in Sweden during the pandemic, however to a lesser degree than in countries that issued lockdowns. Our study also observes differences in pre-pandemic mobility between the different socioeconomic clusters. The wealthier and rural clusters were more mobile, visiting more locations, making more trips, and travelling further. Those differences shrunk during October–November

2020 compared to 2019, since the more mobile clusters reduced their mobility while the travel behaviours of the least mobile clusters were almost unchanged. We also that found that the upper percentile reduced their mobility more than the lower percentiles, both in relative and absolute terms. Therefore, we conclude that the reduction in mobility is primarily driven by individuals with high pre-pandemic mobility. Consistent with the literature, we found that the most socially disadvantaged cluster, Minorities, had the lowest reduction in mobility while the wealthiest cluster, Wealthy, had the largest. However, this observation is consistent with our earlier conclusion, the Minorities cluster has the lowest pre-pandemic mobility while the Wealthy cluster has among the highest and only has a slightly larger reduction in mobility than the other mobile clusters. We, therefore, conclude that socioeconomic factors have a larger effect on pre-pandemic mobility than on the reductions in the observed months during the pandemic.

7.1 Future work

While this thesis measure the extent users travel, how and why the users travel is unknown. Inferring information such as the purpose of the trips, mode of transportation, and category of location visited would enable a deeper understanding of which mobility behaviours cause the difference in mobility between the socioeconomic clusters. Moreover, this information would increase the understanding of which groups have an increased risk of infection due to their mobility, since not all mobility increase the risk of infection equally. Studies determining the purpose of the trip and the type of location visited have been conducted during COVID-19, however, no such studies have been done in Sweden to our knowledge. Conducting such a study in Sweden could provide insight into which types of travel individuals voluntary reduce or stop during the pandemic.

The clustering provides socioeconomic groups, making it easier to interpret and compare the results. However, the clusters are large and might encompass multiple sub-populations, such as students and pensioners, with different mobility patterns. These sub-populations might only occupy a few areas and share many attributes with the cluster, making them difficult to find when clustering. As is the case with students in City Core. A possible solution could be to cluster the areas within a cluster. We also believe this step could be applied to separate clusters such as Working Class, which appear both in cities as well as in the countryside. We believe this step would allow for more well-defined clusters, making the results easier to interpret, as well as more detailed insights.

The locations temporal profiles discussed in Section 6.3.1 provided no conclusive results. It was theorised that the second most visited location became a merger of multiple different locations temporal profile. To explore if this is the case, similar temporal profiles could be grouped together. These groups could then potentially be identified with help of baselines. Lastly, it could be measured how the size of each group has shifted between the years. Allowing for a partial measure if users have changed their second most visited location from workplace to something else.

Bibliography

[1] Public Health Agency of Sweden, "Bekräftade fall i Sverige — Folkhälsomyndigheten." [Online]. Available: https://www.folkhalsomyndigheten.se/smittskydd-beredskap/utbrott/aktuella-utbrott/covid-19/statistik-och-analyser/bekraftade-fall-i-sverige/

[2] K. D'Silva, A. Noulas, M. Musolesi, C. Mascolo, and M. Sklar, "Predicting the temporal activity patterns of new venues," *EPJ Data Science*, vol. 7, no. 1, pp. 1–17, 12 2018. [Online]. Available: https://doi.org/10.1140/epjds/s13688-018-0142-z

[3] Apple, "COVID-19 - Mobility Trends Reports - Apple." [Online]. Available: https://covid19.apple.com/mobility

[4] Google, "COVID-19 Community Mobility Reports." [Online]. Available: https://www.google.com/covid19/mobility/

[5] M. Savage, "Lockdown, what lockdown? Sweden's unusual response to coronavirus - BBC News." [Online]. Available: https://www.bbc.com/news/world-europe-52076293

[6] V. Abedi, O. Olulana, V. Avula, D. Chaudhary, A. Khan, S. Shahjouei, J. Li, and R. Zand, "Racial, Economic, and Health Inequality and COVID-19 Infection in the United States." [Online]. Available: https://doi.org/10.1007/s40615-020-00833-4

[7] V. Bartelink, P. Tynelius, A. Walander, B. Burström, A. Ponce de Leon, J. Nederby Öhd, M. Hergens, and A. Lager, *Socioekonomiska faktorer och covid-19 i Stockholms län*, 2020. [Online]. Available: www.folkhalsoguiden.se

[8] S. Chang, E. Pierson, P. W. Koh, J. Gerardin, B. Redbird, D. Grusky, and J. Leskovec, "Mobility network models of COVID-19 explain inequities and inform reopening," *Nature*, vol. 589, no. 7840, pp. 82–87, 1 2021.

[9] M. Rostila, A. Cederström, M. Wallace, M. Brandén, B. Malmberg, and G. Andersson, "Disparities in covid-19 deaths by country of birth in Stockholm, Sweden: A total population based cohort study," Tech. Rep., 8 2020. [Online]. Available: /articles/preprint/Disparities_in_covid-19_deaths_by_country_of_birth_in_Stockholm_Sweden_A_total_population_based_cohort_study/12852854/1

[10] S. Engebretsen, K. Engø-Monsen, M. A. Aleem, E. S. Gurley, A. Frigessi, and

B. F. De Blasio, "Time-aggregated mobile phone mobility data are sufficient for modelling influenza spread: The case of Bangladesh," *Journal of the Royal Society Interface*, vol. 17, no. 167, pp. 1–25, 2020.

[11] M. U. G. Kraemer, C.-H. Yang, B. Gutierrez, C.-H. Wu, B. Klein, D. M. Pigott, L. du Plessis, N. R. Faria, R. Li, W. P. Hanage, J. S. Brownstein, M. Layan, A. Vespignani, H. Tian, C. Dye, O. G. Pybus, and S. V. Scarpino, "The effect of human mobility and control measures on the COVID-19 epidemic in China," *Science*, vol. 368, no. 6490, 5 2020.

[12] H. Barbosa, M. Barthelemy, G. Ghoshal, C. R. James, M. Lenormand, T. Louail, R. Menezes, J. J. Ramasco, F. Simini, and M. Tomasini, "Human mobility: Models and applications," pp. 1–74, 3 2018.

[13] H. S. Badr, H. Du, M. Marshall, E. Dong, M. M. Squire, and L. M. Gardner, "Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study," *The Lancet Infectious Diseases*, vol. 20, no. 11, pp. 1247–1254, 11 2020.

[14] E. Almlöf, I. Rubensson, M. Cebecauer, and E. Jenelius, "Who Is Still Travelling by Public Transport during COVID-19? Socioeconomic Factors Explaining Travel Behaviour in Stockholm Based on Smart Card Data," *SSRN Electronic Journal*, 9 2020.

[15] L. Liu, H. J. Miller, and J. Scheff, "The impacts of COVID-19 pandemic on public transit demand in the United States," *PLoS ONE*, vol. 15, no. 11 November, 11 2020.

[16] Telia, "Covid-19 mobilitetsanalys via Telia Crowd Insights - Telia.se." [Online]. Available: https://www.telia.se/privat/aktuellt/hemma-i-folknatet/covid-19-mobilitetsanalys

[17] Folkhälsomyndigheten, "Nationell plan för vaccination mot covid-19," Tech. Rep. [Online]. Available: www.folkhalsomyndigheten.se/publicerat-material/.

[18] A. Hernando, D. Mateo, J. Bayer, and I. Barrios, "Radius of Gyration as predictor of COVID-19 deaths trend with three-weeks offset," *medRxiv*, p. 2021.01.30.21250708, 2 2021. [Online]. Available: https://doi.org/10.1101/2021.01.30.21250708

[19] S. Heroy, I. Loaiza, A. Pentland, and N. O'Clery, "COVID-19 policy analysis: labour structure dictates lockdown mobility behaviour," *Journal of The Royal Society Interface*, vol. 18, no. 176, p. rsif.2020.1035, 3 2021. [Online]. Available: https://royalsocietypublishing.org/doi/10.1098/rsif.2020.1035

[20] J. Kim and M.-P. Kwan, "The impact of the COVID-19 pandemic on people's mobility: A longitudinal study of the U.S. from march to september of 2020," *Journal of Transport Geography*, p. 103039, 3 2021. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0966692321000922

[21] K.-Y. Wang, "How Change of Public Transportation Usage Reveals Fear of the SARS Virus in a City," *PLoS ONE*, vol. 9, no. 3, 3 2014.

[22] E. Jenelius and M. Cebecauer, "Impacts of COVID-19 on public transport ridership in Sweden: Analysis of ticket validations, sales and passenger counts," *Transportation Research Interdisciplinary Perspectives*, vol. 8, p. 100242, 11 2020.

[23] M. Dahlberg, P.-A. Edin, E. Grönqvist, J. Lyhagen, J. Östh, A. Siretskiy, and M. Toger, "Effects of the COVID-19 Pandemic on Population Mobility under Mild Policies: Causal Evidence from Sweden," 2020.

[24] Y. A. De Montjoye, C. A. Hidalgo, M. Verleysen, and V. D. Blondel, "Unique in the Crowd: The privacy bounds of human mobility," *Scientific Reports*, vol. 3, 2013.

[25] Y. Liao, "THESIS FOR THE DEGREE OF LICENTIATE OF ENGINEERING Understanding Human Mobility with Emerging Data Sources: Validation, spatiotemporal patterns, and transport modal disparity," Tech. Rep., 2020. [Online]. Available: https://research.chalmers.se/en/publication/515718

[26] P. A. Zandbergen and S. J. Barbeau, "Positional Accuracy of Assisted GPS Data from High-Sensitivity GPS-enabled Mobile Phones THE JOURNAL OF NAVIGATION," *THE JOURNAL OF NAVIGATION*, vol. 64, pp. 381–399, 2011.

[27] T. Horanont, S. Phithakkitnukoon, T. W. Leong, Y. Sekimoto, and R. Shibasaki, "Weather effects on the patterns of people's everyday activities: A study using GPS traces of mobile phone users," *PLoS ONE*, vol. 8, no. 12, 12 2013.

[28] F. Calabrese, G. Di Lorenzo, L. Liu, and C. Ratti, "Estimating origin-destination flows using mobile phone location data," *IEEE Pervasive Computing*, vol. 10, no. 4, pp. 36–44, 10 2011.

[29] B. Hawelka, I. Sitko, E. Beinat, S. Sobolevsky, P. Kazakopoulos, and C. Ratti, "Geo-located Twitter as proxy for global mobility patterns," *Cartography and Geographic Information Science*, vol. 41, no. 3, 5 2014.

[30] E. Pepe, P. Bajardi, L. Gauvin, F. Privitera, B. Lake, C. Cattuto, and M. Tizzoni, "COVID-19 outbreak response, a dataset to assess mobility changes in Italy following national lockdown," *Scientific Data*, vol. 7, no. 1, 12 2020.

[31] Q. Wang, N. E. Phillips, M. L. Small, and R. J. Sampson, "Urban mobility and neighborhood isolation in America's 50 largest cities," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 115, no. 30, pp. 7735–7740, 7 2018. [Online]. Available: www.pnas.org/cgi/doi/10.1073/pnas.1802537115

[32] S. Hasan, C. M. Schneider, S. V. Ukkusuri, and M. C. González, "Spatiotemporal Patterns of Urban Human Mobility," *Journal of Statistical Physics*, vol. 151, no. 1-2, pp. 304–318, 4 2013.

[33] Y. Liao, S. Yeh, and G. S. Jeuken, "From individual to collective behaviours: exploring population heterogeneity of human mobility based on social media

data," *EPJ Data Science*, vol. 8, no. 1, p. 34, 12 2019. [Online]. Available: https://epjdatascience.springeropen.com/articles/10.1140/epjds/s13688-019-0212-x

[34] Y. Xu, A. Belyi, I. Bojic, and C. Ratti, "Human mobility and socioeconomic status: Analysis of Singapore and Boston," *Computers, Environment and Urban Systems*, vol. 72, pp. 51–67, 11 2018.

[35] C. Cottineau and M. Vanhoof, "Mobile phone indicators and their relation to the socioeconomic organisation of cities," *ISPRS International Journal of Geo-Information*, vol. 8, no. 1, p. 19, 1 2019. [Online]. Available: www.mdpi.com/journal/ijgi

[36] Statistics Sweden, "DeSO – Demografiska statistikområden," 2021. [Online]. Available: https://scb.se/hitta-statistik/regional-statistik-och-kartor/regionala-indelningar/deso---demografiska-statistikomraden

[37] G. Chen, A. Carneiro Viana, M. Fiore, and C. Sarraute, "Complete trajectory reconstruction from sparse mobile phone data," 2019. [Online]. Available: https://doi.org/10.1140/epjds/s13688-019-0206-8

[38] Transport Analysis, "Swedish National Travel Survey (RVU Sweden) 2011–2016," 2016. [Online]. Available: https://www.trafa.se/en/travel-survey/travel-survey/

[39] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," Tech. Rep., 1996. [Online]. Available: www.aaai.org

[40] J. Macqueen, "SOME METHODS FOR CLASSIFICATION AND ANALYSIS OF MULTIVARIATE OBSERVATIONS," Tech. Rep.

[41] D. Arthur and S. Vassilvitskii, "k-means++: The Advantages of Careful Seeding," Tech. Rep., 2006.

[42] Statistics Sweden, "Invånare per kvadratkilometer efter region och år." [Online]. Available: https://www.statistikdatabasen.scb.se/sq/108321

[43] E. Pepe, P. Bajardi, L. Gauvin, F. Privitera, B. Lake, C. Cattuto, and M. Tizzoni, "COVID-19 outbreak response: A first assessment of mobility changes in Italy following national lockdown," p. 2020.03.22.20039933, 4 2020. [Online]. Available: https://doi.org/10.1101/2020.03.22.20039933

[44] G. Pullano, E. Valdano, N. Scarpa, S. Rubrichi, and V. Colizza, "Evaluating the impact of demographic, socioeconomic factors, and risk aversion on mobility during COVID-19 epidemic in France under lockdown: a population-based study," p. 2020.05.29.20097097, 9 2020. [Online]. Available: https://doi.org/10.1101/2020.05.29.20097097

[45] Statistics Sweden, "Fortsatt svag restaurangförsäljning tredje kvartalet," 2020. [Online]. Available: https://www.scb.se/hitta-statistik/statistik-efter-amne/handel-med-varor-och-tjanster/

inrikeshandel/omsattning-inom-tjanstesektorn/pong/statistiknyhet/restaurangindex-kvartal-3-2020/

[46] Government office of Sweden, "Pressträff med statsministern 4 juni." [Online]. Available: https://www.regeringen.se/pressmeddelanden/2020/06/presstraff-med-statsministern-4-juni/

[47] The Public Health Agency of Sweden, "Föreskrifter och allmänna råd – covid-19 — Folkhälsomyndigheten." [Online]. Available: https://www.folkhalsomyndigheten.se/smittskydd-beredskap/utbrott/aktuella-utbrott/covid-19/foreskrifter-och-allmanna-rad

[48] Statistics Sweden, "Fortsatt dämpad arbetsmarknad." [Online]. Available: https://www.scb.se/hitta-statistik/statistik-efter-amne/arbetsmarknad/arbetskraftsundersokningar/arbetskraftsundersokningarna-aku/pong/statistiknyhet/arbetskraftsundersokningarna-aku-november-2020/

[49] M. Lee, J. Zhao, Q. Sun, Y. Pan, W. Zhou, C. Xiong, and L. Zhang, "Human Mobility Trends during the COVID-19 Pandemic in the United States," *PLoS ONE*, vol. 15, no. 11 November, 5 2020. [Online]. Available: http://arxiv.org/abs/2005.01215http://dx.doi.org/10.1371/journal.pone.0241468

A

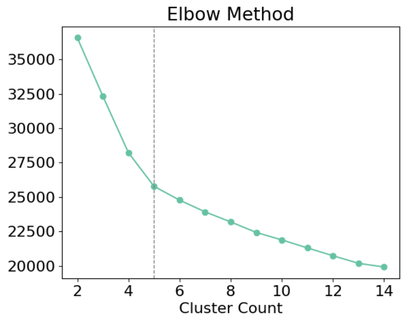
Appendix 1

A.1 Data

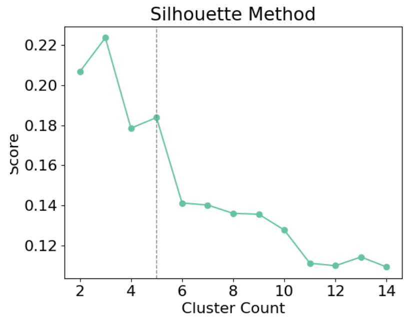
Year/Duration	mean	std	min	25%	50%	75%	max
2019	3024	7421	1	52	650	3209	1582941
2020	1991	6576	1	10	24	616	2012003

Table A.1: Descriptive statistics of duration of stops before preprocessing for both years. The duration is in seconds.

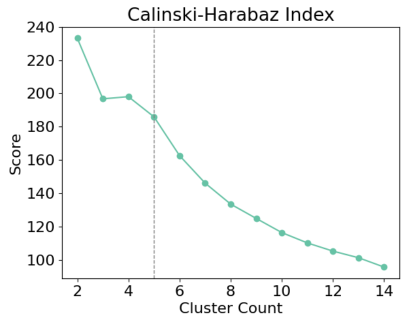
A.2 Clustering



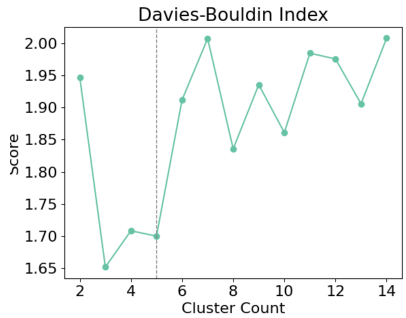
(a) Elbow method, best value found at "elbow"



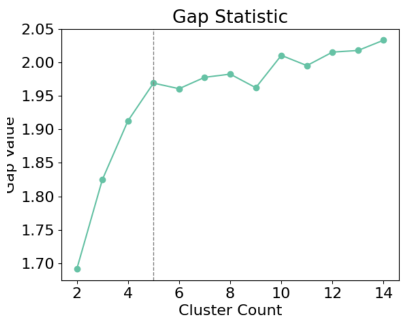
(b) Silhouette Method, the closer to 1 the better



(c) Calinski-Harabaz Index, higher score - better separated clusters



(d) Davies-Bouldin Index, minimum score is the best



(e) Gap Statistic, maximum Gap value is the best

Figure A.1: Five different methods of finding a good k value, $k = 5$ is marked in each plot.

Year	Cluster	Population	Users	Users/ 1000 Pop	Stops	Trips	Median trips/user
2019	Wealthy	461900	400	0.87	22300	11200	27
	City Core	303300	460	1.52	25600	10500	20
	Working Class	385000	530	1.38	29100	13800	24
	Minorities	179300	290	1.62	15300	7000	21.5
	Countryside	396500	460	1.16	25800	12500	25
2020	Wealthy	461900	4380	9.48	345500	132500	27
	City Core	303300	1390	4.58	105100	41100	26
	Working Class	385000	2420	6.29	193900	75100	28
	Minorities	179300	920	5.13	70300	28600	27
	Countryside	396500	2850	7.19	235200	91600	29

Table A.2: Descriptive numbers of the mobility model for each cluster.

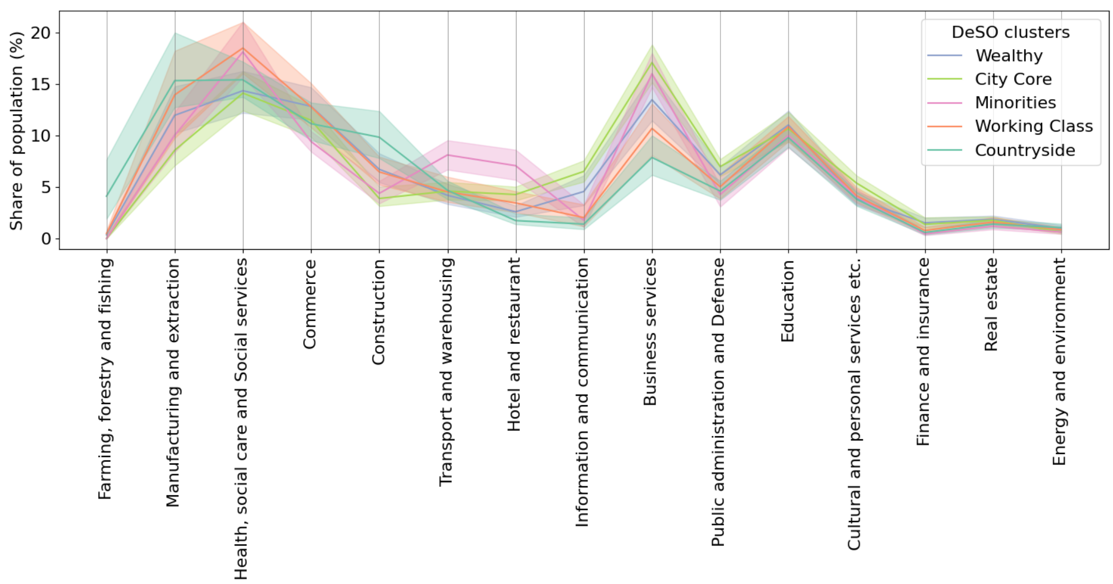


Figure A.2: Parallel coordinate graph over work industry of cluster population. Centre line and the shaded area represent the three main quartiles. The population in this graph refers to people between the ages 16-74.

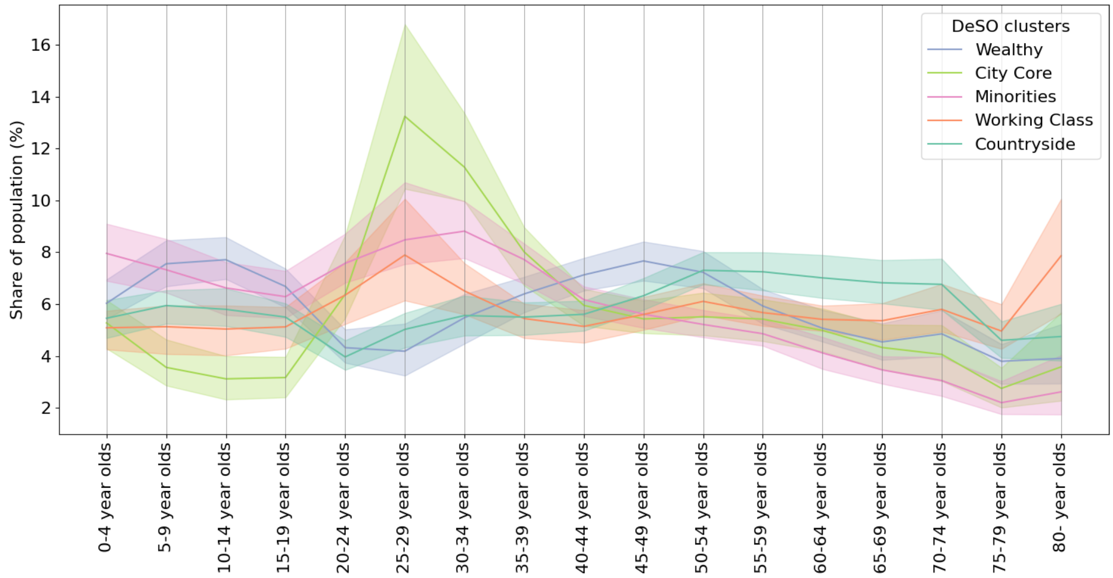


Figure A.3: Parallel coordinate graph over ages of cluster population. Centre line and the shaded area represent the three main quartiles.

A.3 Mobility metrics - Whole population

Radius of gyration	25-th	50-th	75-th	Mean
2019	1.46	5.16	17.71	19.54
2020	1.97	4.88	11.75	12.79
Difference	-0.51	0.28	5.96	6.75
Change (%)	35%	-5%	-34%	-35%

Table A.3: Change in radius of gyration, year on year

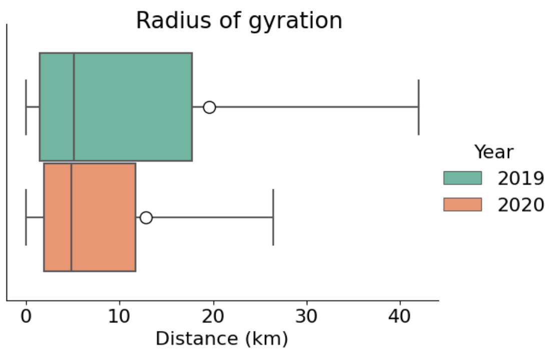


Figure A.4: Box plot of radius of gyration

Mean trip distance	25-th	50-th	75-th	Mean
2019	1.74	4.47	10.22	8.5
2020	2.41	4.93	9.25	7.26
Difference	-0.67	-0.46	0.97	1.24
Change (%)	39%	10%	-9%	-15%

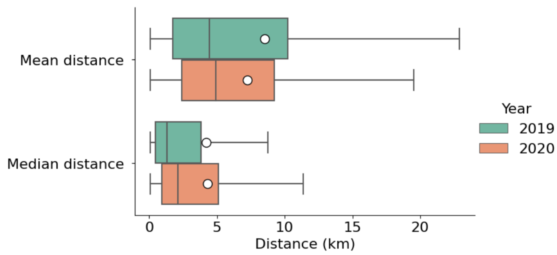
Table A.4: Change in mean trip distance year on year

Median trip distance	25-th	50-th	75-th	Mean
2019	0.48	1.31	3.83	4.21
2020	0.95	2.13	5.12	4.3
Difference	-0.47	-0.82	-1.29	-0.09
Change (%)	98%	63%	34%	2%

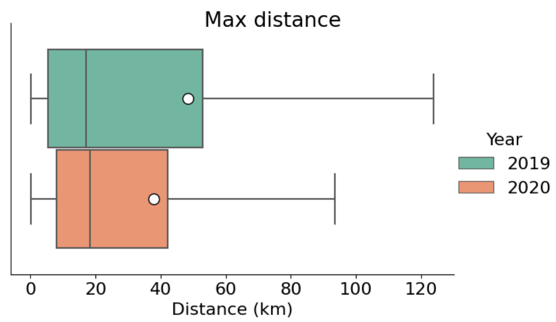
Table A.5: Change in median trip distance year on year

Max trip distance	25-th	50-th	75-th	Mean
2019	5.48	17.08	52.88	48.48
2020	7.97	18.26	42.21	37.96
Difference	-2.49	-1.18	10.67	10.52
Change (%)	45%	7%	-20%	-22%

Table A.6: Change in max trip distance year on year



(a)

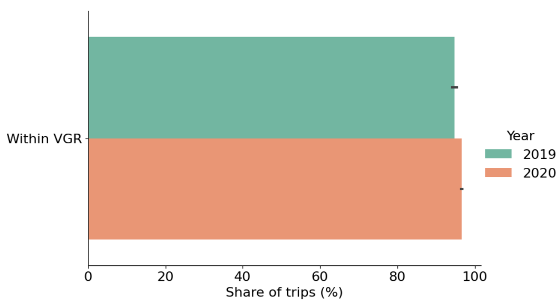


(b)

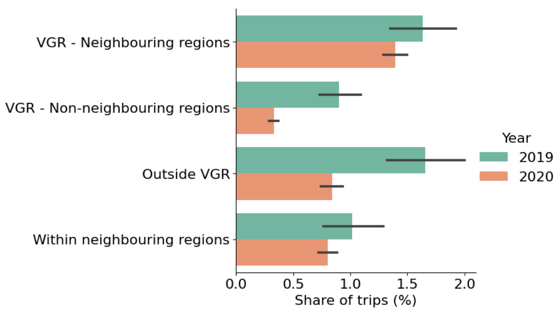
Figure A.5: Trip distance breakdown on mean and median (a) and max (b).

Trips between regions	2019	2020	Difference	Change (%)
Within VGR (%)	94.79	96.64	-1.85	2.0%
VGR - Neighbouring regions (%)	1.63	1.39	0.24	-15.0%
VGR - Non-neighbouring regions (%)	0.9	0.33	0.57	-63.0%
Outside VGR (%)	1.66	0.84	0.82	-49.0%
Within neighbouring regions (%)	1.02	0.8	0.22	-22.0%

Table A.7: Change in travel between regions



(a)



(b)

Figure A.6: Trips to neighbouring regions.

Unique locations visted	25-th	50-th	75-th	Mean
2019	5	9	15	11.18
2020	7	11	16	12.37
Difference	-2	-2	-1	-1.19
Change (%)	40.0%	22.0%	7%	11.0%

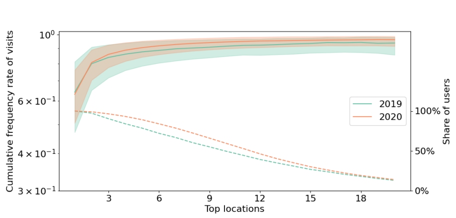
Table A.8: Change in locations visited

Location rank	1	2	5	10	20
25-th	0.47	0.65	0.78	0.85	0.85
50-th	0.64	0.8	0.88	0.91	0.94
75-th	0.81	0.91	0.95	0.96	0.98
Mean	0.63	0.76	0.85	0.89	0.9
Share of users	100%	97%	79%	50%	13%

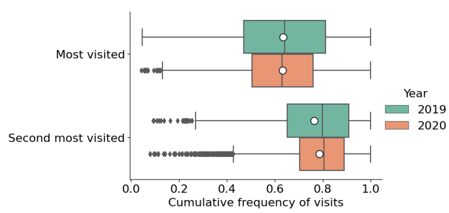
Table A.9: Cumulative visitation frequency for 2019

Location rank	1	2	5	10	20
25-th	0.51	0.7	0.84	0.9	0.91
50-th	0.63	0.81	0.9	0.94	0.96
75-th	0.76	0.89	0.95	0.98	0.99
Mean	0.63	0.79	0.88	0.93	0.94
Share of users	100%	99%	89%	59%	14%

Table A.10: Cumulative visitation frequency for 2020

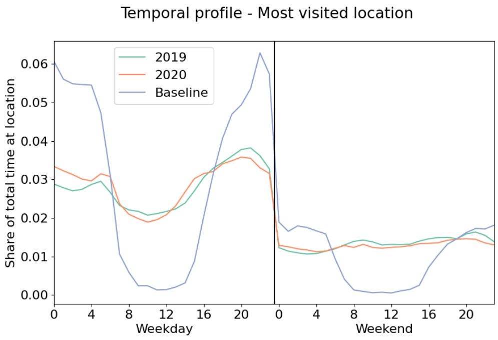


(a) Cumulative frequency of 20 most visited locations, dotes lines indicate share of user visiting at least x locations.

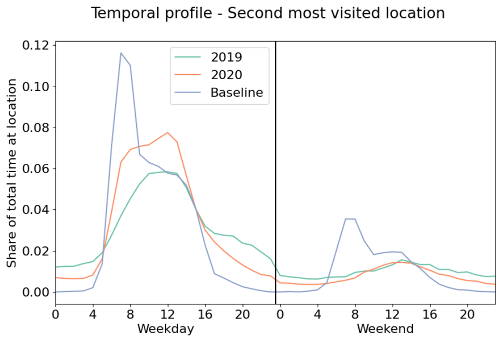


(b) Boxplot of cumulative frequency of visits to top two locations.

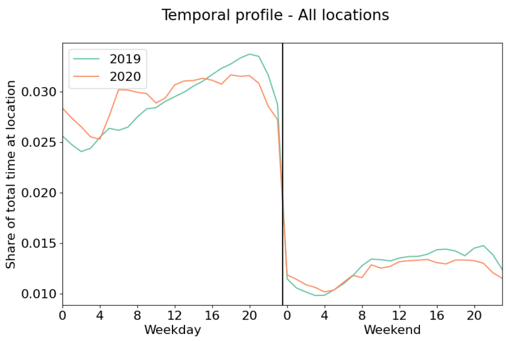
Figure A.7: Cumulative frequency of most visited locations.



(a)



(b)



(c)

Figure A.8: Temporal profile of (a) most visited location, (b) second most visited location, and (c) all locations.

A.4 Mobility metrics - Socioeconomic clusters

Radius of gyration (km)	25-Percentile	Median	75-Percentile	Mean
City Core	0.23 (21%)	0.08 (2%)	-6.89 (-37%)	-10.0 (-40%)
Countryside	0.45 (12%)	-0.89 (-10%)	-9.07 (-38%)	-8.03 (-34%)
Minorities	0.33 (33%)	0.63 (25%)	0.09 (1%)	0.42 (4%)
Wealthy	-0.19 (-8%)	-1.88 (-29%)	-7.46 (-43%)	-9.82 (-45%)
Working Class	0.48 (49%)	-0.78 (-14%)	-5.37 (-29%)	-3.92 (-23%)

Table A.11: Difference in radius of gyration

Median trip distance (km)	25-Percentile	Median	75-Percentile	Mean
City Core	0.21 (42%)	0.17 (13%)	0.04 (1%)	0.06 (2%)
Countryside	0.87 (139%)	2.25 (109%)	1.57 (22%)	0.22 (3%)
Minorities	0.33 (100%)	0.34 (29%)	-0.05 (-1%)	-0.47 (-13%)
Wealthy	0.57 (103%)	0.76 (47%)	0.38 (8%)	0.16 (4%)
Working Class	0.3 (70%)	0.62 (76%)	0.3 (10%)	-0.83 (-19%)

Table A.12: Difference in median trip distance

Mean trip distance (km)	25-Percentile	Median	75-Percentile	Mean
City Core	0.28 (20%)	0.12 (3%)	-0.95 (-13%)	-1.6 (-21%)
Countryside	1.02 (29%)	0.41 (5%)	-1.23 (-8%)	-1.1 (-10%)
Minorities	0.36 (28%)	0.24 (7%)	0.36 (6%)	-0.67 (-11%)
Wealthy	0.27 (11%)	-0.55 (-10%)	-1.91 (-18%)	-1.58 (-18%)
Working Class	0.5 (36%)	-0.2 (-4%)	-1.5 (-14%)	-2.17 (-24%)

Table A.13: Difference in mean trip distance

Maximum trip distance (km)	25-Percentile	Median	75-Percentile	Mean
City Core	1.18 (30%)	1.75 (19%)	-2.27 (-5%)	-18.76 (-33%)
Countryside	1.76 (13%)	-2.77 (-9%)	-11.57 (-18%)	-12.03 (-21%)
Minorities	0.99 (22%)	1.79 (19%)	2.7 (13%)	2.25 (9%)
Wealthy	-0.37 (-4%)	-4.86 (-23%)	-20.49 (-36%)	-16.61 (-31%)
Working Class	1.66 (35%)	-2.27 (-9%)	-9.92 (-16%)	-5.79 (-12%)

Table A.14: Difference in maximum trip distance

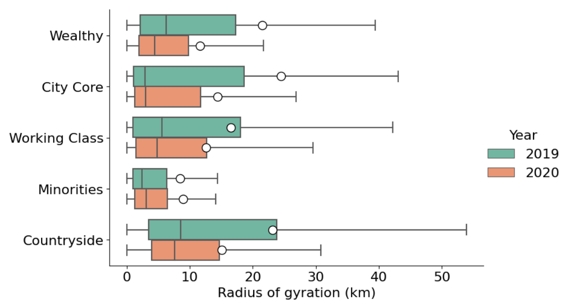
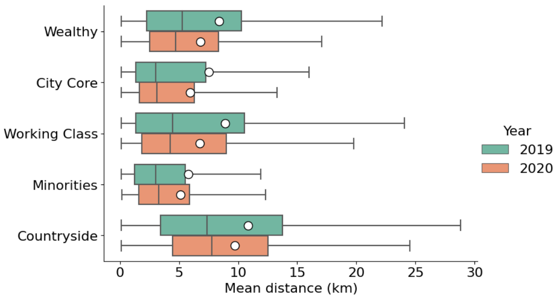
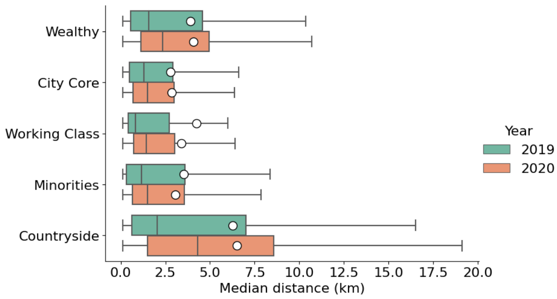


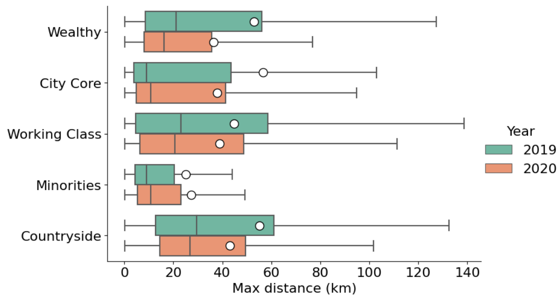
Figure A.9: Box plot of radius of gyration



(a)



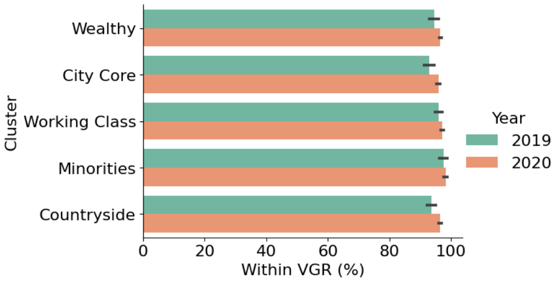
(b)



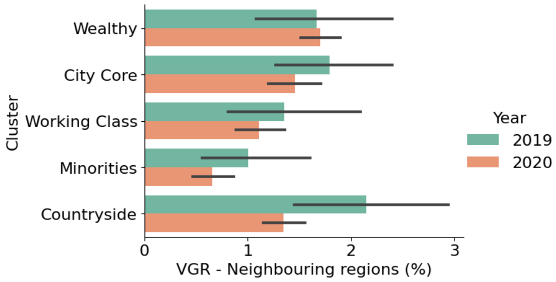
(c)

Figure A.10: Trip distance breakdown on mean (a), median(b) and max (c).

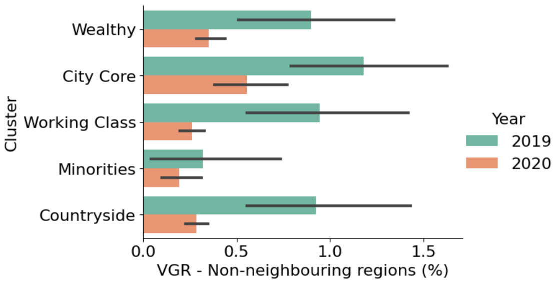
Trips between regions	Within VGR	VGR - Neighbouring	VGR - Non-neighbouring	Outside VGR	Within neighbouring
City Core	2.96 (3%)	-0.34 (-18%)	-0.62 (-52%)	-1.24 (-50%)	-0.76 (-47%)
Countryside	2.74 (2%)	-0.8 (-37%)	-0.64 (-69%)	-1.34 (-57%)	0.04 (3%)
Minorities	0.52 (0%)	-0.35 (-34%)	-0.13 (-40%)	0.23 (77%)	-0.28 (-41%)
Wealthy	1.87 (1%)	0.04 (2%)	-0.55 (-60%)	-0.97 (-58%)	-0.38 (-33%)
Working Class	1.06 (1%)	-0.24 (-17%)	-0.68 (-72%)	-0.26 (-23%)	0.12 (20%)



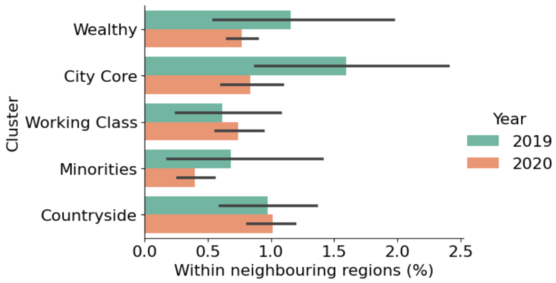
(a)



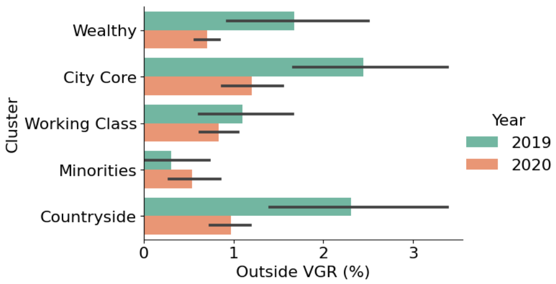
(b)



(c)



(d)



(e)

Figure A.11: Trips breakdown on a regional level.

City Core/Location rank	1	2	5	10	20
Share of users	0.0 (0%)	2.0 (2%)	12.0 (15%)	14.0 (30%)	4.0 (40%)
Mean	-0.01 (-1%)	0.0 (0%)	0.02 (2%)	0.03 (3%)	0.04 (4%)
25-th	0.02 (4%)	0.02 (2%)	0.04 (4%)	0.04 (4%)	0.09 (10%)
50-th	-0.01 (-1%)	-0.02 (-2%)	0.01 (1%)	0.02 (2%)	0.04 (4%)
75-th	-0.06 (-7%)	-0.04 (-4%)	0.0 (0%)	0.01 (1%)	0.03 (3%)

Table A.15: Change in cumulative frequency of visits, City Core cluster

Working Class/Location rank	1	2	5	10	20
Share of users	0.0 (0%)	1.0 (1%)	9.0 (11%)	8.0 (16%)	2.0 (18%)
Mean	-0.01 (-1%)	0.02 (2%)	0.04 (4%)	0.03 (3%)	0.04 (4%)
25-th	0.02 (4%)	0.03 (4%)	0.04 (4%)	0.04 (4%)	0.04 (4%)
50-th	-0.02 (-3%)	0.0 (0%)	0.03 (3%)	0.02 (2%)	0.02 (2%)
75-th	-0.04 (-4%)	-0.01 (-1%)	0.01 (1%)	0.01 (1%)	0.01 (1%)

Table A.16: Change in cumulative frequency of visits, Working Class cluster

Minorities/Location rank	1	2	5	10	20
Share of users	0.0 (0%)	3.0 (3%)	12.0 (16%)	17.0 (44%)	4.0 (40%)
Mean	-0.07 (-10%)	-0.03 (-3%)	-0.01 (-1%)	0.02 (2%)	0.02 (2%)
25-th	-0.07 (-12%)	-0.02 (-2%)	0.02 (2%)	0.04 (4%)	0.01 (1%)
50-th	-0.1 (-13%)	-0.06 (-6%)	-0.01 (-1%)	0.02 (2%)	0.03 (3%)
75-th	-0.08 (-9%)	-0.04 (-4%)	-0.01 (-1%)	0.0 (0%)	0.0 (0%)

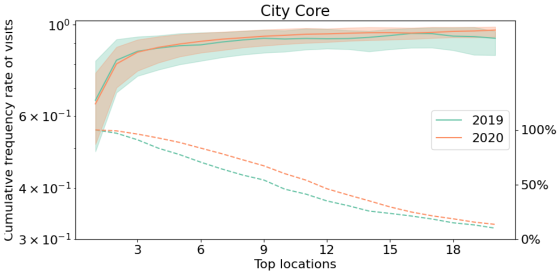
Table A.17: Change in cumulative frequency of visits, Minorities cluster

Countryside/Location rank	1	2	5	10	20
Share of users	0.0 (0%)	1.0 (1%)	9.0 (11%)	7.0 (12%)	-1.0 (-5%)
Mean	0.02 (3%)	0.05 (6%)	0.06 (7%)	0.06 (6%)	0.03 (3%)
25-th	0.08 (19%)	0.09 (15%)	0.09 (12%)	0.07 (8%)	0.06 (7%)
50-th	0.03 (5%)	0.05 (6%)	0.05 (5%)	0.05 (5%)	0.01 (1%)
75-th	-0.04 (-5%)	0.01 (1%)	0.02 (2%)	0.03 (3%)	0.0 (0%)

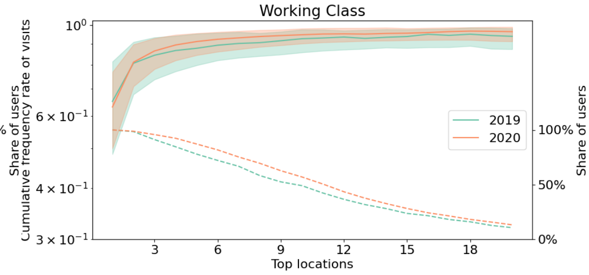
Table A.18: Change in cumulative frequency of visits, Countryside cluster

Wealthy/Location rank	1	2	5	10	20
Share of users	0.0 (0%)	2.0 (2%)	10.0 (12%)	4.0 (7%)	-3.0 (-20%)
Mean	0.03 (4%)	0.04 (5%)	0.05 (5%)	0.04 (4%)	0.03 (3%)
25-th	0.08 (17%)	0.08 (12%)	0.07 (8%)	0.05 (5%)	0.04 (4%)
50-th	0.04 (6%)	0.04 (5%)	0.05 (5%)	0.04 (4%)	0.03 (3%)
75-th	-0.02 (-2%)	0.01 (1%)	0.02 (2%)	0.02 (2%)	0.01 (1%)

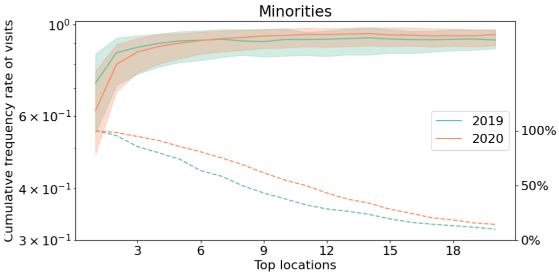
Table A.19: Change in cumulative frequency of visits, Wealthy cluster



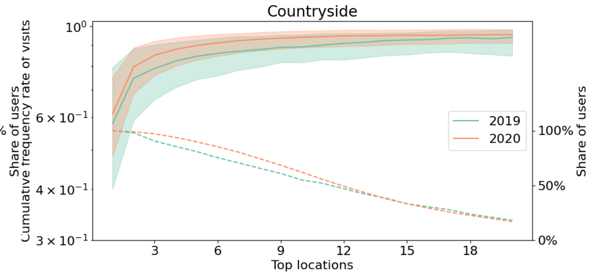
(a)



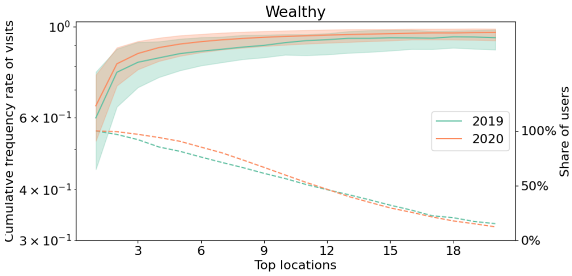
(b)



(c)

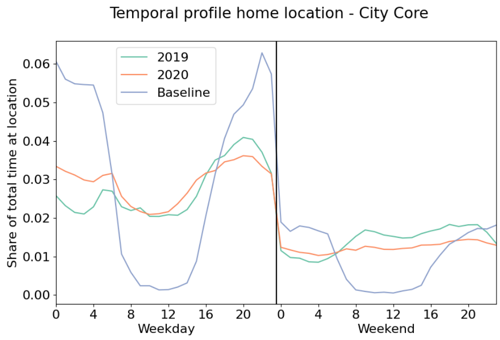


(d)

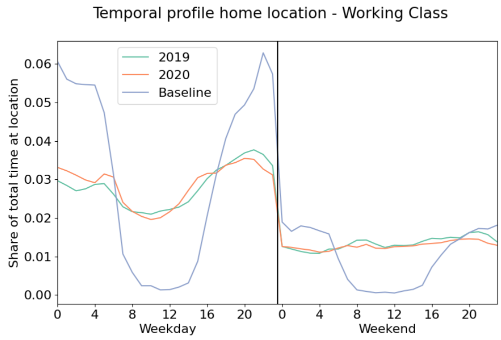


(e)

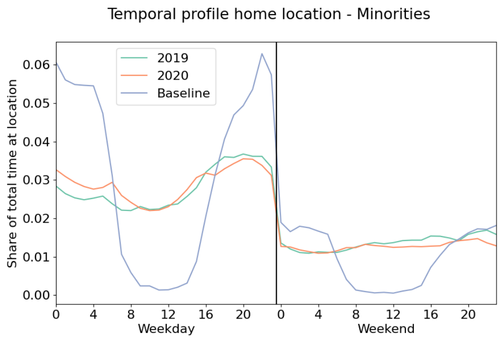
Figure A.12: Cumulative visitation frequency for all clusters.



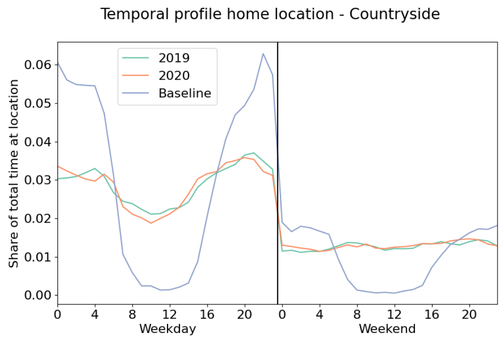
(a)



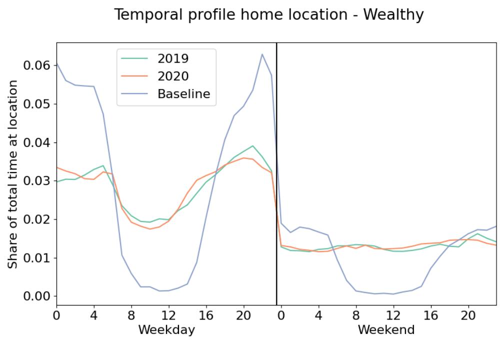
(b)



(c)

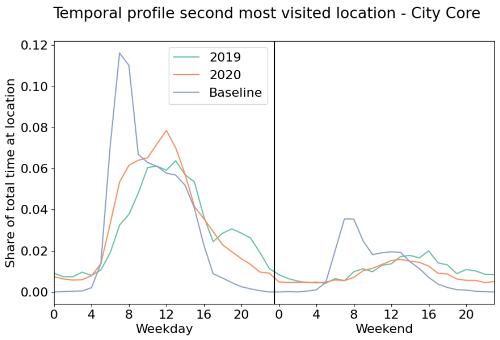


(d)

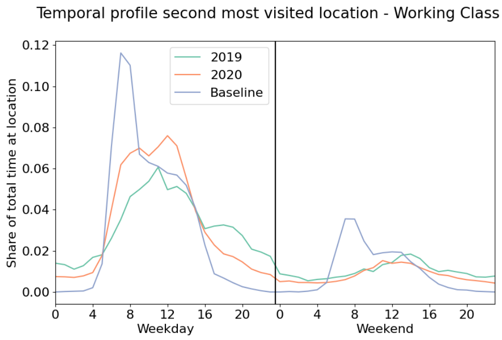


(e)

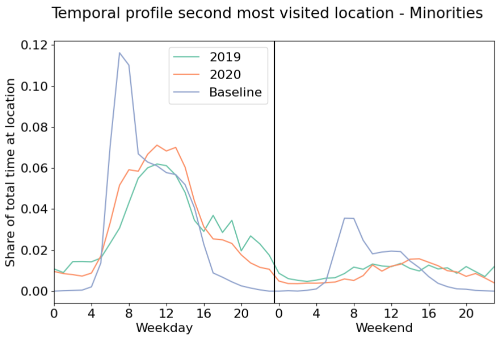
Figure A.13: Temporal profile of the most visited location for all clusters.



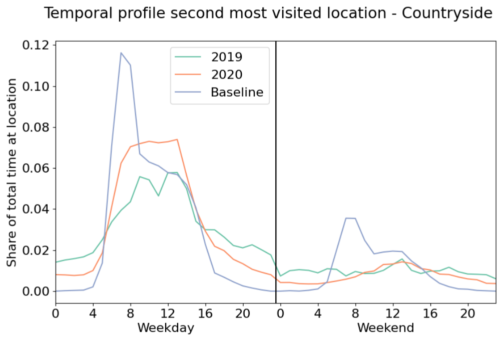
(a)



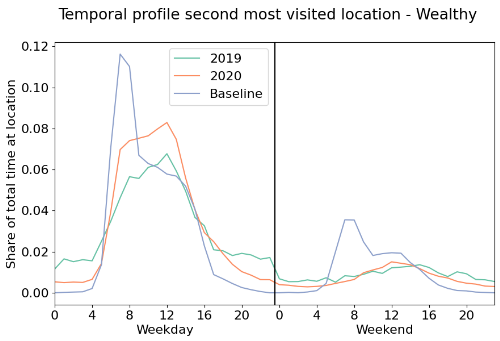
(b)



(c)

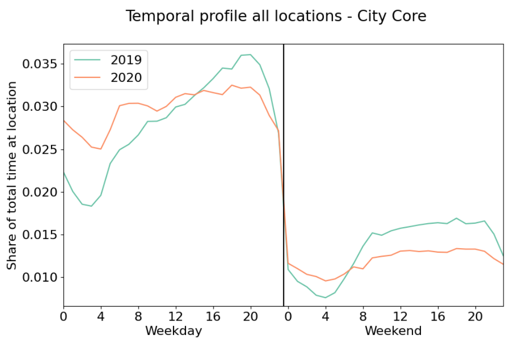


(d)

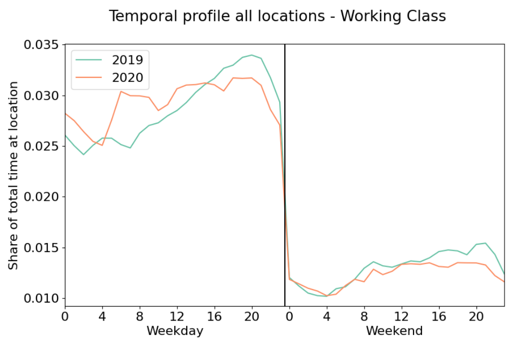


(e)

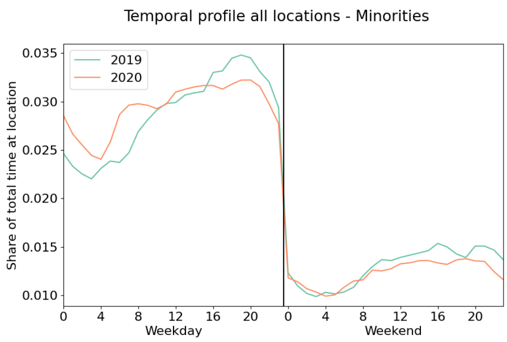
Figure A.14: Temporal profile of the second most visited location for all clusters.



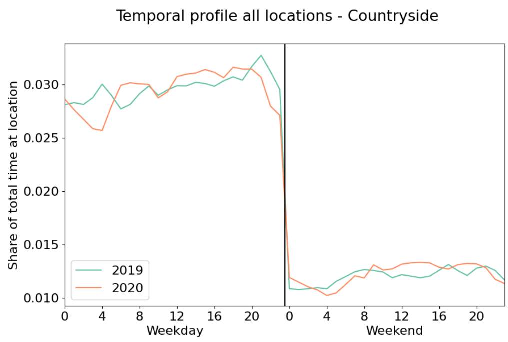
(a)



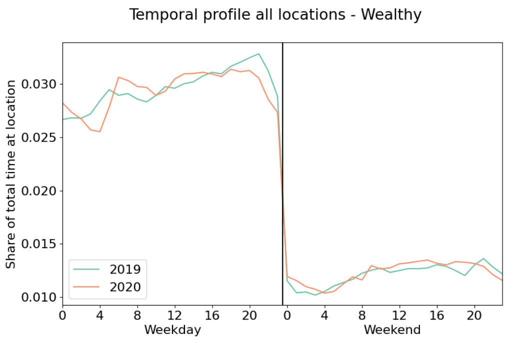
(b)



(c)



(d)



(e)

Figure A.15: Temporal profile of all locations for all clusters.