UNIVERSITY OF TARTU
Faculty of Science and Technology
Institute of Ecology and Earth Sciences
Department of Geography

Master's thesis in Geoinformatics for Urbanised Society (30 ECTS)

Differences between population groups residing in Tallinn in terms of staying at important locations during the COVID-19 pandemic

Petr Anashin

Supervisors: Amnir Hadachi, PhD
Siiri Silm, PhD
Serio Agriesti, MSc

Tartu 2023

Differences between population groups residing in Tallinn in terms of staying at important locations during the COVID-19 pandemic

Abstract:

This thesis examines the problem of segregation between occupation, ethnic and income groups residing in Tallinn during the COVID-19 pandemic in terms of staying at important locations. While previously mobility was considered as an asset vital for social integration and interaction, this paradigm has changed during the pandemic when the possibility of staying at home became an advantage for some population groups because people without the possibility of teleworking had to continue to visit their physical workplaces what created an additional threat to their health. In this thesis with the help of spatial clustering of mobile positioning data using information on addresses of important locations there were identified GPS points related to home, work, hobby and other locations. Further aggregated measures were calculated to analyse how much time respondents from each population group spent at different important locations, when respondents visited these locations during the day, and what percentage of respondents from each group visited each type of location throughout the investigated period. Also, with the help of these measures it was investigated what lifestyle representatives of different population groups had before and after new COVID restrictions.

Keywords:

Mobile positioning, COVID-19 pandemic, segregation, spatial clustering

CERCS code:

S230 – Social geography

Contents

1 Introduction

Spatial mobility crucially impacts our daily lives. It plays a significant role starting from commuting, engagement in leisure and recreational activities and finishing with migration to the other country. Our movement in time and space that is determined by spatial mobility is an outcome of the complex interrelation of various individual factors (socioeconomic status, personal preferences, gender, age etc.), interaction with other individuals, and influence of external factors, for example, social values, state policy and surrounding environment (Järv et al., 2015).

Mobility is considered as a new form of capital independent from the other forms of capital, including economic, social and cultural capital (Flamm and Kaufmann, 2006). Nowadays, mobility has become a vital asset in terms of social integration and interaction; it gives a feeling of personal freedom and promotes the formation of social status. Thus, differences in mobility between population groups may become a reason that ultimately contributes to segregation.

However, the COVID-19 pandemic significantly altered the usual state of things. Governmental and people's voluntary restrictions aimed to slow down the spread of the virus seriously changed mobility behaviour. For instance, people started to use public transport or taxi less frequently because their usage was associated with the potential risk of being infected. At the same time, individual transportation means like private cars, bicycles or walking were assumed as the safest options (Beck and Hensher, 2020). Studies reveal that during the COVID period people tried to avoid public places such as shops and attempted to use online services (Bhaduri et al., 2020). Also, educational organizations and many office employees whose work did not require physical attendance switched to remote working mode. It led to a decrease of study and work-related trips.

However, different population groups adapted differently to new conditions, and it strengthened segregation. Those who could not stay at home were more vulnerable towards COVID-19, significantly altering the previous paradigm of mobility as an asset. For example, research carried out in King County in the U.S. state of Washington demonstrated that residents with lower socioeconomic and educational level reduced their work-related number of trips much less than people with higher socioeconomic status and educational level, because the latter could perform their work duties remotely (Brough, Freedman, and Phillips, 2021). The other studies demonstrate that spatial segregation and isolation of ethnic minorities also has been growing throughout the pandemic. For example, it happened to Chinese migrants in Seoul, South Korea, who mainly worked in low-paid jobs (Shin, 2022).

Thus, socioeconomic and ethnic segregation that has accelerated during the pandemic can significantly impact already segregated urban communities. Talking about Tallinn, even before COVID-19 the city had high levels of socioeconomic and ethnic segregation. It was reported that it could end up as the most segregated European capital in 25-30 years (Tammaru, 2015). Considering the fact that representatives of the Russian-speaking minority predominantly worked as blue collars without the possibility of remote work (United Nations Development Programme, 2017), it could result in lower opportunities to stay at home and follow governmental recommendations, what in its turn might result in higher infectious rate in this population group.

Currently, there are no studies that investigate how the pandemic affected segregation in Tallinn in the context of mobility and differences between population groups in terms of staying and visiting important locations. However, this topic requires thorough examination because these differences could worsen existing social disparities.

Thus, the following research questions can be formulated:

1) What were the differences between ethnic, income and occupation groups in terms of time spent at important locations such as home, work and others and percentage of population group members visited these locations per day?
2) At what time of the day did population group representatives visit important locations, and what were the main differences between groups?
3) How different population groups responded to newly imposed restrictions?

2 Theoretical overview

2.1 Spatial segregation around the clock

The vast number of papers in the past focused primarily on the bias and disadvantages of segregation indexes (Williams, 1948; Taeuber and Taeuber, 1976), however, the majority of publications relied on residential segregation in the city's neighbourhoods and did not take into account spatial and temporal distribution of population groups during the whole day. The time that people spend at home represents only one life domain, and daily activity space is not limited only to residential neighbourhoods (Van kempen and Wissink, 2014).

More recent scientific literature suggests shifting from the residential domain to the other spheres of human activity using work and leisure time anchor points to investigate mobility (Tammaru et al., 2021). Interestingly, socioeconomic and ethnic segregation have different patterns in the case of workplace, leisure and residential segregation. Overall, ethnic segregation or racial discrimination affects workplace segregation less than residential segregation (Strömgren et al., 2014). At the same time, socioeconomic segregation decreases opportunities for interaction between different socioeconomic classes at work, while possible communication at their place of living is higher (Hu, Moayyed, and Frank, 2022). These conclusions have been confirmed by research where segregation around the clock was analysed. For instance, segregation between social groups in mono-ethnic Beijing was higher on weekends and lower on weekdays (Zhang, Wang, and Kan, 2022). On the contrary, segregation between ethnic groups in Tallinn which is mainly ethnically segregated, was lower on weekdays and higher on weekdays (Silm and Ahas, 2014).

Talking specifically about Tallinn, ethnic segregation in the city has a clear temporal rhythm. In the research conducted by Silm and Ahas, segregation indexes of Russian and Estonian-speaking minorities received from CDR data – Call Detail Records – revealed that Tallinn had the highest segregation in the afternoon at 13:00-16:00 and the lowest at night at 22:00-01.00 according to Figure 1 based on the index of dissimilarity, while isolation followed the same temporal patterns. As already mentioned, segregation on weekends was higher compared to weekdays. The modified isolation index revealed that the Russian-speaking minority tended to spend their weekends in areas inhabited by residents from the same ethnic group. In terms of seasonality, segregation was lower in winter and higher in summer; however, isolation of the Russian-speaking minority, even though the difference was not large, had reversed pattern.

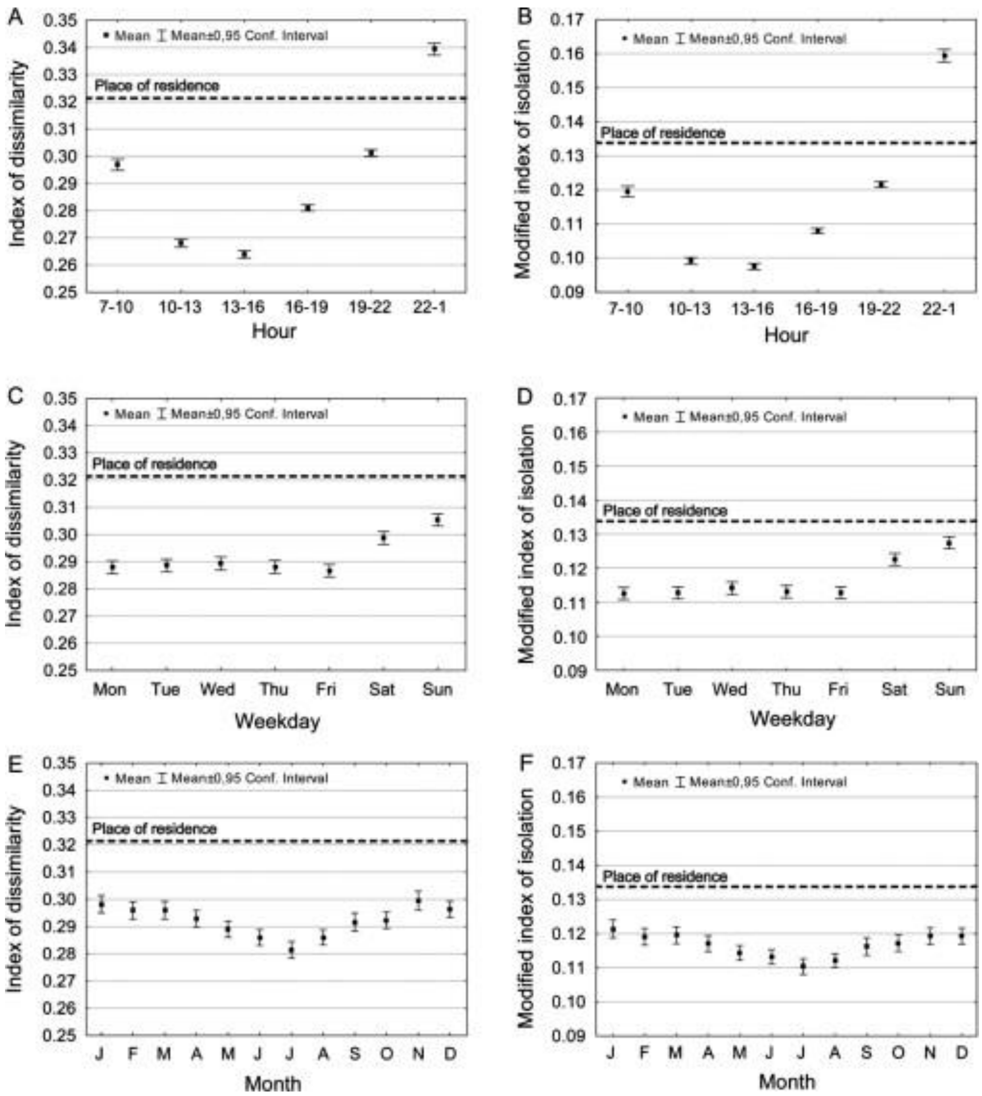


Figure 1. The index of dissimilarity (ID) and the modified index of isolation (MII) according to mobile-phone positioning data during the day (A and B), week (C and D) and year (E and F) compared to the places of residence by census data (Silm and Ahas 2014)

2.2 Activity space differences between Russian and Estonian-speaking population

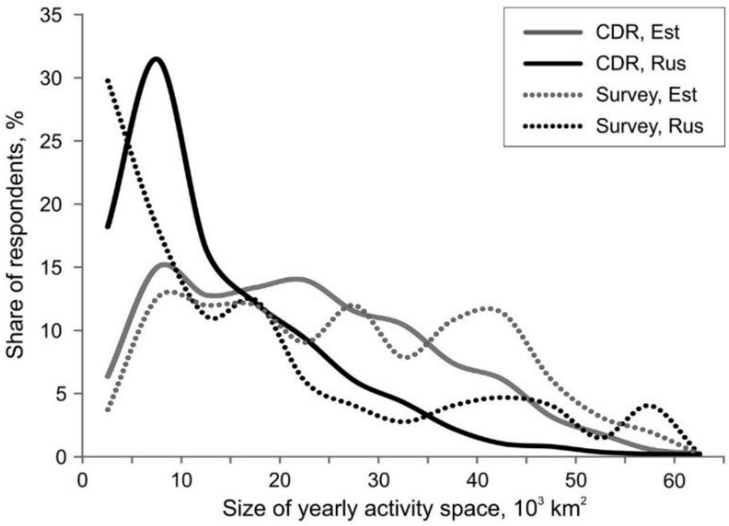
Previous studies carried out before the pandemic demonstrated that the mobility of Russian-speaking minority and Estonian speakers in Estonia significantly differs. For example, the authors of the previously quoted paper found out that the location quotient – LQ – of Russian speakers during nighttime was higher in the Lasnamäe neighbourhood, where the share of residing Russian-speaking minority is the largest, whereas in daytime LQ values grew in almost all neighbourhoods. The reason of the uneven distribution of the Russian-speaking minority at night which flattened later in daytime, is that members of the minority used more extended urban space during the day.

Weekdays and weekends also had spatial patterns. On working days, LQ values were higher in North-Tallinn and Nõmme, but almost all neighbourhoods were classified as “weekend areas“.

Järv in the other research work where differences in activity space between Russian and Estonian speaking population groups have been analysed also stated that the activity space of the Russian-speaking minority was limited to locations that were populated by Russian speakers (Järv et al., 2015). For example, Russian residents from Tallinn in the study sample travelled mainly within Tallinn, including the Lasnamäe neighbourhood and to the Ida-Viru County in North-Eastarn Estonia. On the contrary, the distribution of locations visited by Estonian speakers was more equal across the country, and it was not limited only to specific areas. However, representatives of the majority group also did not visit areas populated predominantly by Russian speakers. Additionally, the number of visited places and activity locations of the Russian-speaking population was much fewer compared to Estonian speakers. Moreover, differences in mobility behaviour between these two ethnic groups became more explicit when the study period was extended. In other words, the daily activity space of Russian and Estonian speakers differed less than the monthly and annual activity spaces.

Finally, Järv in the research dedicated to the link between ethnic segregation and socioeconomic status where both CDR and mobility survey data on Tallinn’s residents from Russian and Estonian-speaking population groups were used demonstrated that numerically median individual activity space for Estonian speakers was 23 223 square km (survey data) and 18 575 square km (CDR data), whereas activity space of Russian residents was more than twice less: about 7 500 square km based on both data sources.

Furthermore, distributions of the size of individual activity space had significant differences between the two ethnic groups. Estonian speakers had relatively equal distribution across the whole range, but the distribution of Russian speakers had a large spike in the lower range that rapidly reduced afterwards based in Figure 2. It indicates that only a small share of Russian-speaking residents had large activity space, the most of them used very limited activity space.



		CDR	Survey
Mean	Est	19,926	23,966
	Rus	11,067	14,892
SE. of mean	Est	211	1,099
	Rus	163	1,254
Std. Dev.	Est	12,705	14,324
	Rus	9,578	15,656
Kurtosis	Est	0,65	0,98
	Rus	1,19	0,34
Skewness	Est	0,43	0,22
	Rus	1,27	1,22
Q1	Est	9,084	11,944
	Rus	3,449	1,924
Median	Est	18,575	23,223
	Rus	7,543	7,544
Q3	Est	29,156	36,264
	Rus	16,590	20,973
N	Est	3,614	170
	Rus	3,442	156

Figure 2. The comparison of distribution and descriptive statistics on the spatial extent of individual yearly activity spaces between Russian and Estonian-speaking population groups and two data sources – CDR and survey data (Järv et al. 2021)

2.3 The general change in mobility behaviour and habits due to COVID.

The pandemic seriously affected all people's daily activities, especially in the first months. Google mobility report results for Harju county, visualised in Figure 3, reveals that retail and recreation, grocery and pharmacy, transit stations and workplaces have seen a significant drop in visitors from late February to June. The largest decrease was in retail and recreation (-60%) and transit stations (-60%), whereas grocery stores and pharmacies and workplaces were less affected. At the same time, people stayed much more time in residential areas, while the number of park visitors was dependent more on seasonal factor rather than on COVID.

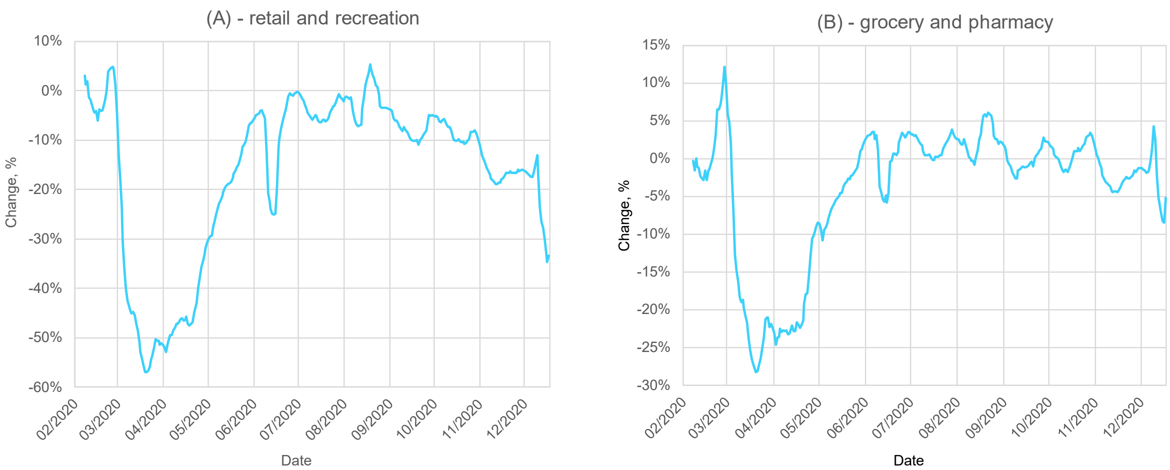


Figure 3. Percentage change in the number of people visiting: (A) – retail and recreation services, (B) – grocery stores and pharmacies, (C) – parks, (D) – transit stations, (E) – workplaces, (F) – residential areas in Harju county during 15.02.2020-31.12.2020. Moving average based on a 7-day period (Google, 2022)

A detailed survey on changes in mobility behaviour carried out in Australia (Beck and Hensher, 2020) showed that COVID had the greatest impact on outdoor leisure activities such as meeting friends (80%), visiting restaurants (76%), going to movies (55%) and bars (50%). In contrast, more vital activities that were harder to postpone, such as doctor's appointments (36%), schooling and childcare (26%) and work functions (23%), were less affected, what is demonstrated in Figure 4.

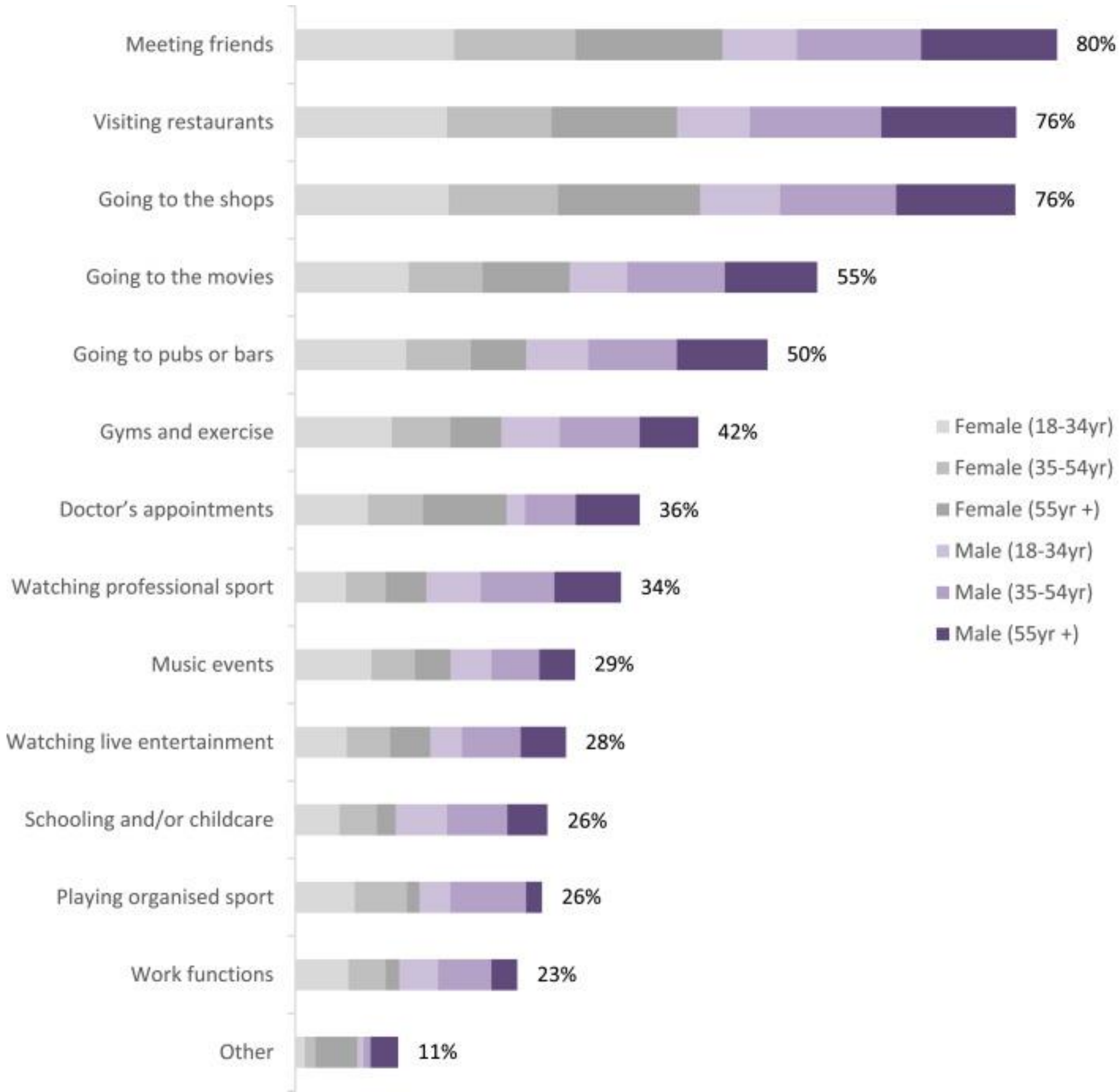


Figure 4. Interruption to normal activities due to COVID-19 (Beck and Hensher, 2020)

Regarding income differences, lower-income respondents replied less frequently that their attendance of restaurants, bars, gyms, watching or playing sports and work functions have been disrupted by the pandemic. The more incomes the households had, the more interruption in a wider list of activities has been reported. Middle-income households acknowledged that visiting restaurants, bars and pubs was disrupted, while high-income households also listed work functions, physical activities and watching or playing sports.

According to the same research, there was also a significant shift in preferred modes of transport. For instance, the vast majority of respondents claimed that individual transportation means were the most comfortable: private cars (84%) and walking or cycling (11%) had the highest positive perception during COVID and the lowest negative attitude (1% and 3% respectively). Meanwhile, public transport was considered as a risk area: among the least convenient modes of transport have been listed bus (42%), train (33%), taxi (12%) and ferry (7%). Overall, 58% of respondents were extremely concerned about hygiene in public transport.

In a situation when people perceived public transport and public spaces as more dangerous due to the higher risk of being infected, changes in the urban habits of residents have been noticed. The research carried out in Stockholm with the help of a combination of georeferenced tweets and a survey revealed a significant increase of activity in outdoor recreation areas that could provide quietness and solitude, such as parks and hiking areas (Legeby et al., 2023). However, beaches, event areas and outdoor seating areas were less visited. Sports facilities have demonstrated a similar trend. As for retail activity, grocery stores have seen an increase in Tweets because they were not closed. Overall, the less essential service was, the sharper decrease it had. Non-essential shops like book, furniture or clothing shops were visited less, as well as the other indoor recreation services including restaurants and especially night clubs, pubs and bars. Healthcare services were the only category that maintained the same level of activity. According to the survey results, the most common reason people continued to use specific services was the ability to access them because some kinds of services have been closed.

Additionally, the study showed that access to green spaces and services played a crucial role in compliance with COVID requirements. Local services and green areas became more popular in the neighbourhoods where they were more accessible, while residents of districts that did not have both types of amenities nearby travelled longer distances to reach them in the other places within the city. Therefore, the pandemic accelerated inequality because some neighbourhoods with access to services and green areas were more prepared to respond to the pandemic than others, and their residents received more benefits due to their living location.

Statistics Estonia mobility analysis results revealed that after the declaration of the emergency in Estonia, the mobility of people in the country substantially decreased (Statistics Estonia, 2020). Estonian residents spent on average 20 hours in their main location, the total daily travelled distance decreased from 27.5 km to 17.6 km, and the number of trips from the main location fell from 1.8 to 1.5 per day. Additionally, the population number in Tallinn and Tartu, the two largest Estonian cities, decreased during that time because people moved to their summer houses or other locations outside the city.

2.4 Mobility differences between population groups in the COVID and post-COVID periods.

Currently, there are no studies dedicated to the investigation of changes in the mobility between different population groups in Tallinn during the pandemic. Thus, researches carried out in other locations will be taken into account.

Paper in which segregation of Chinese migrants in Seoul, South Korea (Shin, 2022) was examined stated that segregation and isolation of Chinese migrants has increased in the first phase of the pandemic and continued to grow afterwards. For instance, the dissimilarity index between Chinese immigrants and native-born Koreans grew by 5.6% in the first stage and at the end of 2020 it was already 9%. Segregation of non-Chinese migrants initially also was greater than before COVID-19, but it was smaller compared to Chinese migrants and it almost vanished in half a year.

Regardless of the pandemic, both segregation and isolation of Chinese immigrants during working hours of the day were lower than during the night. It confirms findings from other studies that ethnic segregation is less pronounced in the labour market. Also, Chinese immigrant segregation and isolation were lower during weekdays than on weekends.

Interestingly, Chinese immigrants' mobility was more restricted to their residining neighbourhoods both during daytime and nighttime, what worsened their spatial segregation even further.

Authors of other paper devoted their research to socioeconomic disparities in travel behaviour during the COVID-19 pandemic in King County in the U.S. state of Washington (Brough, Freedman, and Phillips, 2021). Data from public transport cards was used in this study, but the authors claimed that transportation substitution (e.g. from public transport to private cars) took place only in the first phase of the pandemic.

The primary outcome of the research was that people from lower-income and educational groups did not reduce number the of trips so much compared to higher educational and income groups. It was confirmed by a more pronounced day and weekly travel cycle of lower educational and income groups: the number of trips was considerably higher during working hours and on the weekdays.

The main reason was that people from high educational and income groups could switch to remote work, while people with lower incomes and education level had jobs that could not be done remotely. Authors demonstrated that when COVID restrictions were eased, it led to immediate growth in trips number among people from low-income and low-educational groups, but those who could work from home further continued to do so. Therefore, teleworking was a crucial factor affecting mobility both in the short and long term.

Results of the most recent paper, where authors analysed segregation during daytime in greater Stockholm ,(Müürisepp et al., 2023) showed that even in a city which did not impose strict mobility restrictions during the first phase of the pandemic, there was a significant impact on daytime social diversity. The number of neighbourhoods providing a venue for interaction between various socioeconomic and ethnic groups has decreased. The analysis revealed that diversity fell more in high-income neighbourhoods compared to low-income and more in minority neighbourhoods compared to majority districts. Places with a higher concentration of workplaces and commercial services suffered more heavily.

2.5 Factors affecting differences in urban mobility in Estonia

Previous research where differences in levels of urban mobility in Estonia have been analysed stated that language was the primary factor which explained variance of the activity space (Järv et al., 2015). In a more recent paper devoted to the link between ethnic segregation and socioeconomic status, it also was found that language was the most critical factor, while education and income did not explain the extent of activity space (Järv et al., 2021). However, a deeper analysis revealed that variation in spatial mobility might be explained by the interaction between self-estimated social status and ethnicity, so differences in spatial mobility happen not only due to ethnic inequalities but rather because of individual positioning in society.

The most considerable deviations in the extent of activity space have been detected between Russian and Estonian speakers who defined themselves as affiliated with the middle and middle-high social class, as shown in Figure 5. Surprisingly, Russian-speaking residents with middle or middle-high social status were not interested in investigating society beyond their segregated activity space even though they had more resources compared to people from low-middle and low social classes. The substantial growth of yearly activity space has been observed only among Russian speakers having high social status. Therefore, in the author’s opinion, there was no direct correlation between the social status among representatives of the Russian-speaking minority and activity space.

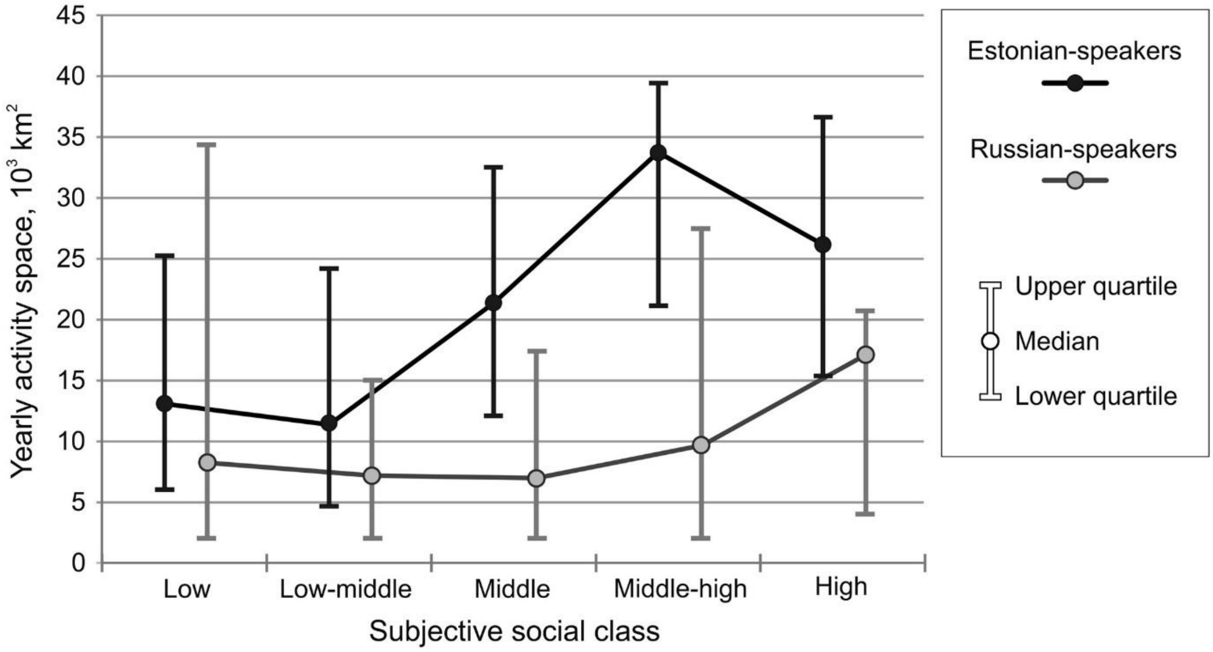


Figure 5. The spatial extent of individual yearly activity spaces derived from the survey data by ethnicity and subjective social status (Järv et al., 2021)

However, the other important factor affecting daily mobility appeared during the pandemic that could seriously alter the previous state of things in Estonian society. Studies that considered remote work indicated significant deviations in urban mobility between socioeconomic groups during COVID (Brough, Freedman, and Phillips, 2021). Most blue collars could not perform their work tasks remotely, unlike white collars, who preferred remote job even after lifted restrictions.

In Estonia, there is a division between the Estonian-speaking population working more often as white collars and the Russian-speaking minority, which members are more frequently employed as blue collars. For example, in Figure 6 it is shown that the share of non-Estonians in 2015 was larger in such occupational groups as skilled workers and unskilled workers, while the percentage of Estonians was higher among professionals and managers.

Figure 6. Distribution of ethnic groups by occupational groups, 2015 (United Nations Development Programme, 2017)

Figure 7 shows the percentage of employees by broad occupation groups who started to work remotely during the COVID-19 pandemic. The survey on which the figure is based was carried out in 6 different countries: Japan, China, USA, Italy, South Korea and the United Kingdom. Respondents employed in such occupations as managers and professionals had the highest levels of teleworkability, about 55% of employees from these occupation groups switched to remote work after the beginning of the pandemic. Associate professionals and clerks revealed lower levels of being able to telework, but they were still higher than among blue-collars: about 40-42%. At the same time, people employed as services and sales workers or in other blue-collar occupations had the lowest levels of teleworkability, only 24-26% of employees started to work remotely.

These statistics tell us about the potential inability of the Russian-speaking population to work remotely since their occupational groups are related to low teleworkable groups. At the same time, Estonian speakers have the opposite situation. Thus, it could seriously impact these population groups' mobility during COVID.

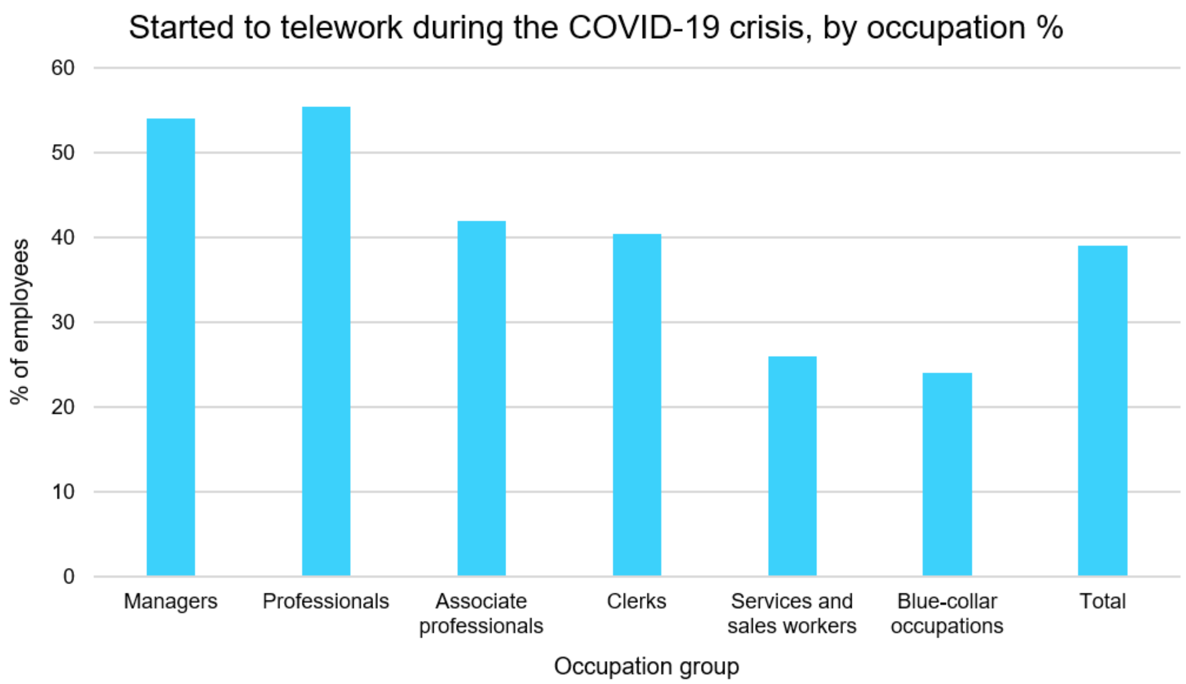


Figure 7. Percentage of employees by occupation group started to work remotely (Sostero M. et al., 2020)

3 Data

3.1 Study area

The study area of this research is Tallinn, the capital of Estonia. According to the Statistical Yearbook of Tallinn 2021, in 2021 in Tallinn lived 445 thousand residents, for 49.5% of the city's population mother tongue was Estonian and for 44.1% – Russian language (Tallinn city government, 2021).

Before the Soviet occupation of Estonia, the country and capital in particular had a rather homogenous population consisting primarily of the indigenous population. However, a large-scale migration started after World War II from the USSR states, which resulted in significant changes in population's composition. The share of the Estonian population decreased to less than two-thirds before the collapse of the Soviet Union. Migrants were allocated in newly built urban districts near industrial areas or in neighbourhoods severely damaged due to the war. They did not have enough opportunities to interact with the Estonian population in these segregated areas, which created prerequisites for further ethnic segregation.

Nowadays, there are several neighbourhoods within the city with a high share of minorities. Visualisation of ethnic composition is displayed in Figure 8. Lasnamäe is a city district which is the most populated by Russian speakers. In some subdistricts percentage of minority population reaches more than 60%. The other district with a high share of Russian-speaking population is Põhja-Tallinn or Nothern Tallinn, where the city's port is located. The average share of non-Estonians there is about 50%. Mustamäe is the last area with a higher concentration of minorities, which percentage is 30-50%. The spatial distribution of Estonians and Russians in Tallinn may be found in Figure 8.

Data for this work has been collected mainly from residents of 2 subdistricts: Priisle, located in Lasnamäe near the western border of the city, and Kalamaja, which is situated near the city centre. These two subdistricts have opposite population structures: based on Figure 8, 64% of residents in Priisle in 2021 were Russians and only 22% were Estonians, while in Kalamaja there were 76% Estonians and 13% Russians. Both subdistricts have comparable total population: 12 652 people lived in Kalamaja in 2021 and 10 976 in Priisle.

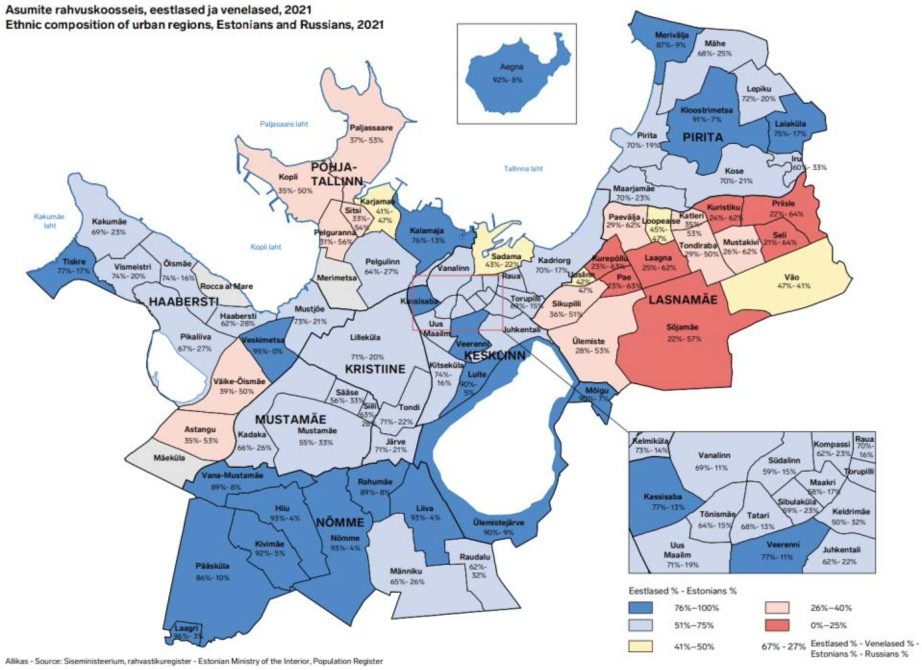


Figure 8. A map on Ethnic composition of neighbourhoods in Tallinn, 2021 (Tallinn city government, 2021)

Comparing the average monthly gross incomes per employee in these 2 subdistricts in 2020 using Figure 9, it is possible to say that the average incomes of employees living in Kalamaja are almost 1.5 times higher than incomes of Priisle's residents. Moreover, taking into account the previous map, districts populated predominantly by Russians are on average poorer than neighbourhoods where the majority of the population consists of Estonians.

As described in the previous part, non-Estonians more frequently work as blue collars like (un)skilled workers or in customer service. In contrast, Estonians are more often employed as white collars working as associate professionals, professionals or managers. As the economy of Estonia shifted from an industrial to a market economy after the collapse of the USSR and residents who came from the previous Soviet states had poor local language knowledge, they became a population group who suffered the most due to these changes. They could not afford to change their apartments in ageing residential development to newly built houses in the city centre or in the suburbs like the Estonian middle class. As a result, not only ethnical segregation but also economic one has emerged, and there is a significant divide in incomes between Estonian and Russian speakers (Silm and Ahas 2014).

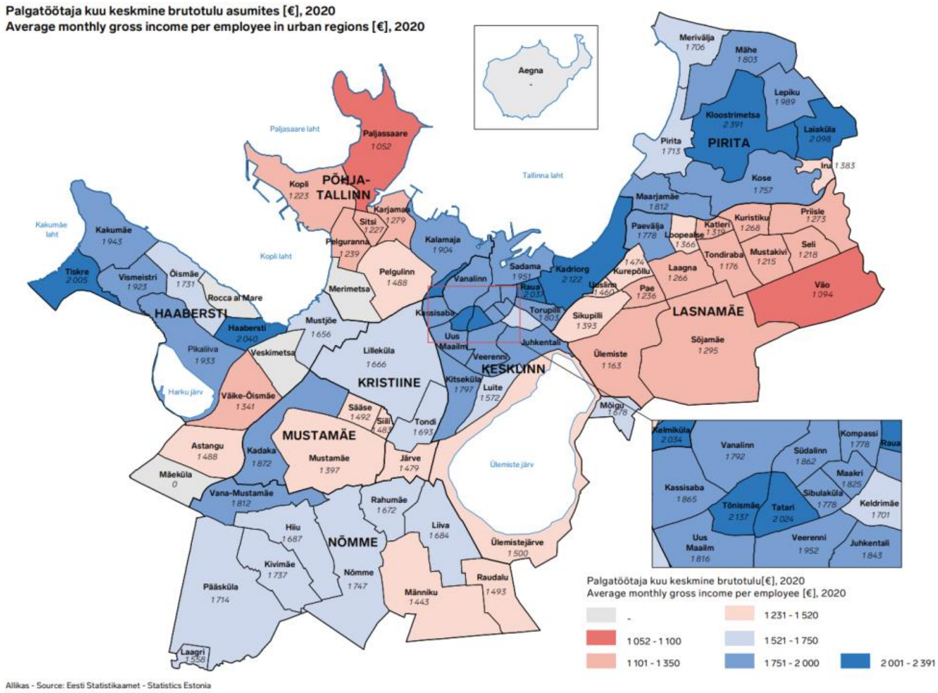


Figure 9. Average monthly gross income per employee in euro in Tallinn's subdistricts, 2020 (Tallinn city government, 2021)

Talking about the number of positive COVID cases registered in the study area, Figure 10 demonstrates a relatively mild first wave from the middle of February until the middle of April, but the second wave, which started in September and continued in the following months, can be characterized by much more rapid increase in a number of infected. In 2020 the fastest growth has been observed in November and December.

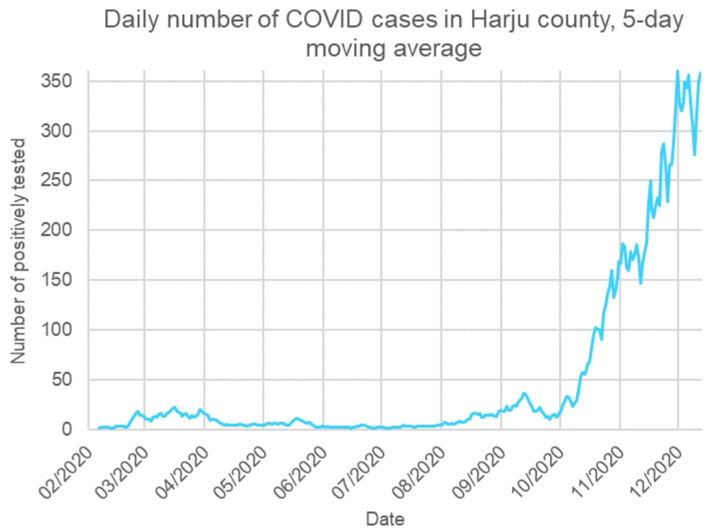


Figure 10. Number of positive COVID tests registered in Harju county, 5-day moving average (Health and Welfare Information Systems Centre of Estonia)

Since the study period in the analysis was November 2020, only those COVID restrictions that were introduced in November 2020 will be considered. As Figure 10 demonstrates, a number of new cases grew rapidly in November. Therefore to tackle this situation, the Estonian government introduced COVID restrictions on the 10-th of November that entered into force on the 16-th of November (Republic of Estonia Government, 2020a). These restrictions included:

1) All employees were recommended to work remotely if it was possible. It was advised to cancel all joint events.
2) Wearing a mask became obligatory in public places and transport.
3) 2 + 2 rule in service halls and public areas of commercial enterprises had been established. No more than two people could move together keeping a distance of 2 meters with others, except for families and if this could not be reasonably guaranteed.
4) A two-metre distance rule between groups of people with a maximum of 10 people per group in catering and entertainment establishments was set.
5) Opening hours of catering and entertainment establishments were limited from midnight to 6 a.m. to avoid large people's gatherings.

Further restrictions were announced on 23rd November and enacted on 28th November (Republic of Estonia Government, 2020b). Restrictions were stricter in Ida-Viru and Harju counties because coronavirus spread faster in these regions. There were the following restrictions in these counties:

1) A 50% occupancy limit in indoor halls with stationary seating (theatres, cinemas, concert venues) and at public events.
2) The maximum number of participants in indoor places with stationary seating was set to 400, and 250 elsewhere. For outdoor events, the maximum number was 500 participants.
3) Outside the stationary seating area, it was allowed to move according to the 2 + 2 rule instead of groups of maximum ten people.
4) Maximum ten people could be in a group attending hobby education or activities, indoor trainings and refresher courses. Different groups must not contact each other.
5) Indoor sports group trainings could be performed in groups of maximum ten people and different groups could not interact with each other.
6) Up to two spectators could move together at indoor sporting or exercise events, keeping a distance of at least two metres from the others.

Thus, these restrictions seriously affected public events like exhibitions and public meetings, sports and hobby activities, attendance of theatres, cinemas, museums, and usage of transport.

3.2 Data description

Analysed data consisted of two main datasets: a survey which included information on each respondent and recorded GPS points of respondents. The rest of the data was collected in tables consisting of important locations and their addresses.

The GPS points dataset had 15 columns, including respondent ID corresponding to the GPS point, point ID, time when the GPS point has been recorded, GPS point coordinates in well-known text format stored in WGS 84 coordinate system, accuracy and altitude in meters and speed in meters per second. There were recorded more than 10 million GPS points from 10.05.2020 until 01.02.2021.

The survey, in its turn, had 11 columns, including respondent ID, place of residence (neighbourhood), age, gender, mother tongue, occupation and occupation group, education and income levels, type of house, and main language in the area.

There were collected data from 186 respondents, the majority of which were female (104 people), according to Figure 11.

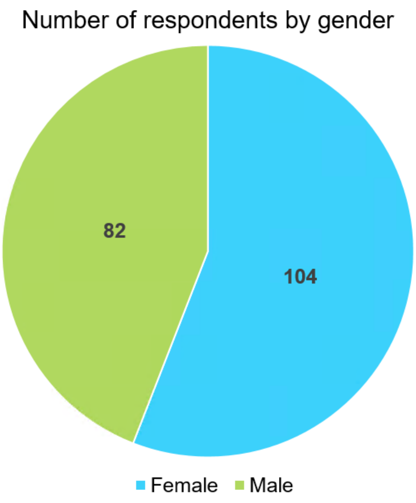


Figure 11. Number of respondents by gender

Looking at the age density plot in Figure 12, a conclusion can be made that the most frequent age among respondents was from 30 to 35 years, and the frequency of people being older is gradually falling after this age interval.

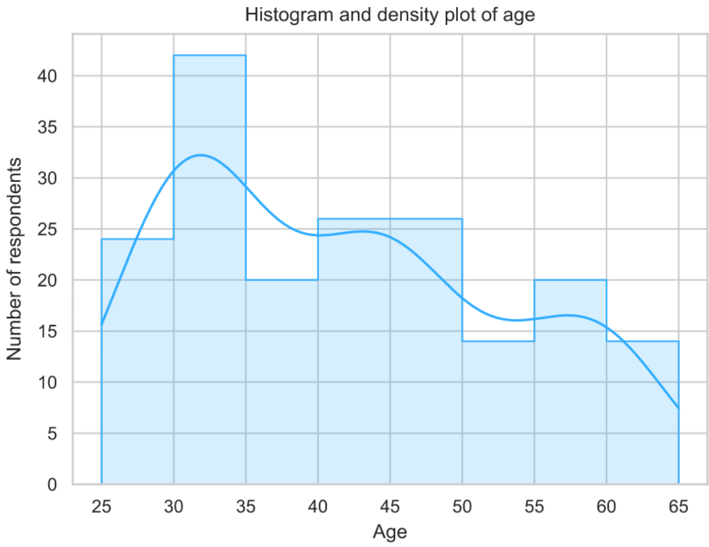


Figure 12. Histogram and density plot of age

As for the place of residents, Figure 13 demonstrates that an almost equal number of residents lived in neighbourhoods mainly populated by representatives of one ethnic group: Estonians in Kalamaja and North Tallinn on the one hand and Russians in Priisle and Lasnamäe on the other. The proportion of Estonian and Russian speakers in the neighbourhoods is similar to the real data shown in Figure 8.

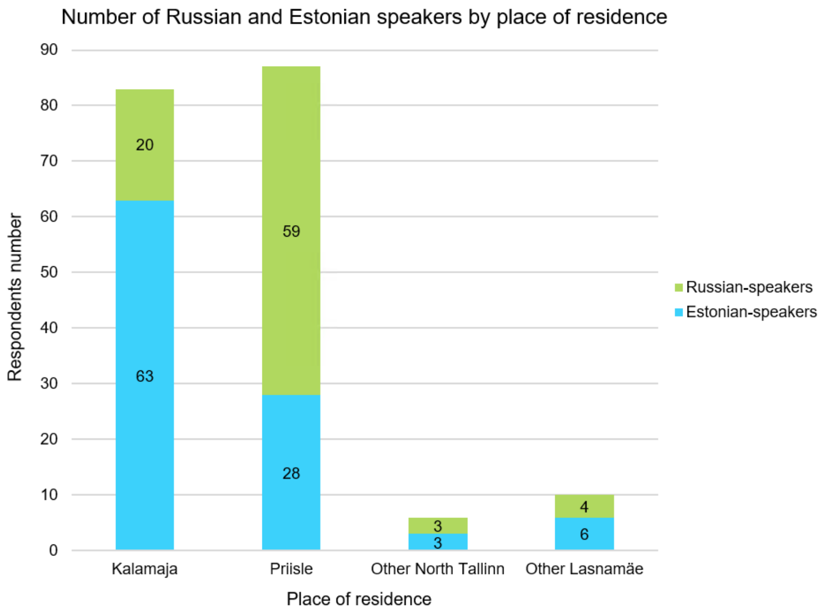


Figure 13. Number of Estonian and Russian speakers by place of residence

In Figure 14, respondents were grouped by mother tongue and income level below 2000 EUR per month and equal or more than 2000 EUR per month. 2000 EUR per month was slightly more than the average salary in Tallinn in the period when data has been collected. The number of Estonian speakers and Russian speakers earning 2000 EUR per month or more is almost the same, but there were approximately two times more Russian speakers getting less than 2000 EUR per month.

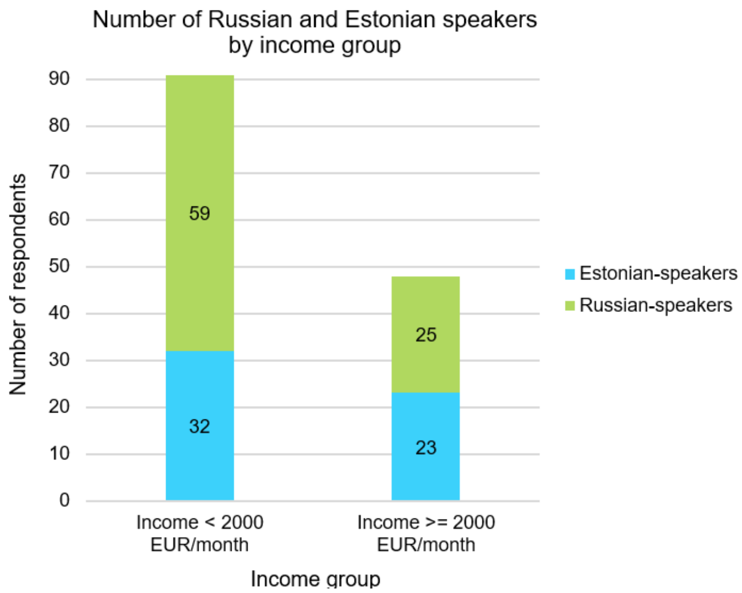


Figure 14. Number of Estonian and Russian speakers by income group

Based on Figure 15, mid-level and top specialists were the most common occupation groups among respondents, while assemblers, officials and manual workers were the least popular occupations.

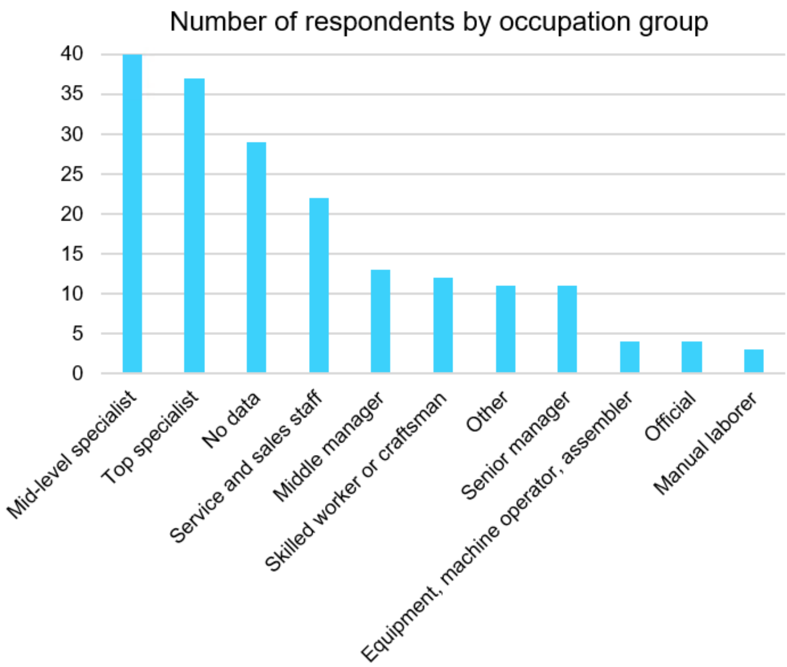


Figure 15. Number of respondents by occupation group

As for important locations, there were provided the following types of locations: home, work, work other, study, study other, hobby, kids, other and second home. Work other and study other represented locations connected with work and study. Among hobby locations, there were listed indoor and outdoor sports areas, other places of regular hobby activities like language schools, singing choirs, student clubs etc., places related to culture and other places related to hobbies and entertainment like nightclubs, camping grounds and others. Kids' location types included kindergartens, schools, hobby clubs or locations for doing sports, playgrounds and other places. Other locations involved a broad range of places, for instance: places of visit (home of parents, a friend, a relative etc.), grocery stores, other speciality stores (clothing store, construction store etc.), places related to catering and outdoor activities (cafe, restaurant, pub), places connected with personal services (beauty salon, hairdresser, therapist), institutions (bank, post office, police station), transport-related places (bus or railway station, airport), places of other services (shoemaker, tailor) and other regularly visited locations (partner's workplace etc.).

3.3 Data cleaning

Firstly, all features have been converted to appropriate data formats. Columns containing information on time were changed from string to datetime format, and the column containing information on coordinates of GPS point was used as a geometry column to build a geometry dataframe applying a well-known text format. As the study area was Tallinn, Estonia, the coordinate reference system of the geometry dataframe was changed to the Estonian Coordinate System of 1997 – EPSG:3301 from WGS 84 – EPSG:4326.

Afterwards, a new column was created to measure time spent at each point as a difference between the time when the person will be at a new point and the time of the current point.

Further, points have been filtered by stop duration from the previously received column. Since the aim of this work is to analyse time spent at important locations, it was important to extract points related to stops. Frequencies of the first 30 unique values of stop duration have been received in descending order to identify the optimal threshold for stop.

A number of points where the time difference was 1 second was the greatest (8.2 million records) because if a person moves, then information on his location is recorded every second, but in case of stop, recording starts when the person begins moving. The next most frequently encountered stop duration equals only 18 seconds, and was found only in 258 thousand records.

As Figure 16 shows, most of the 30 most frequently encountered values of stop duration were from 15 to 40 seconds, while fewer stops with lower duration were encountered. Therefore, it was decided to set a 15-second stop duration threshold to capture points related to stops. So, all records where the stop duration was less than 15 seconds have been deleted. As a result, the size of the dataset after filtering decreased by 5.1 times, but the total time of all stops decreased only by 2.25%.

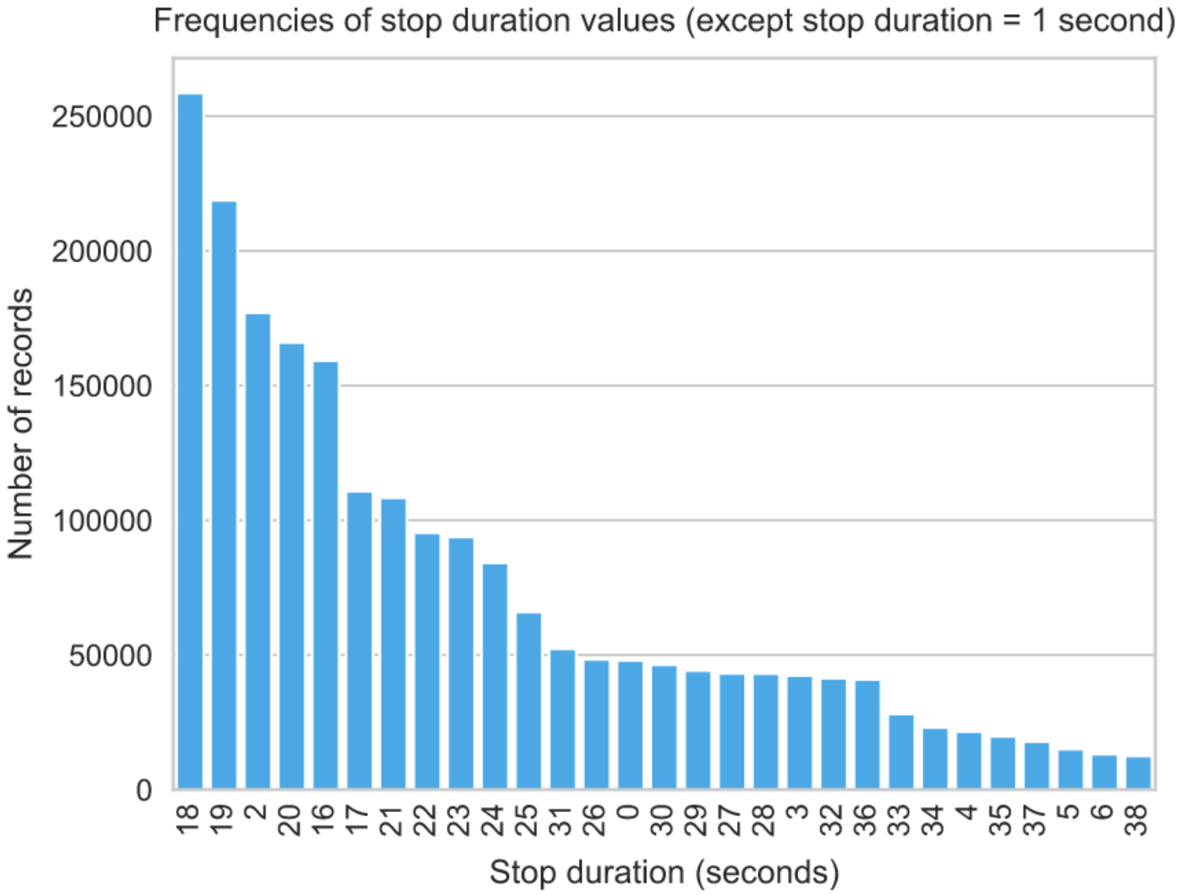
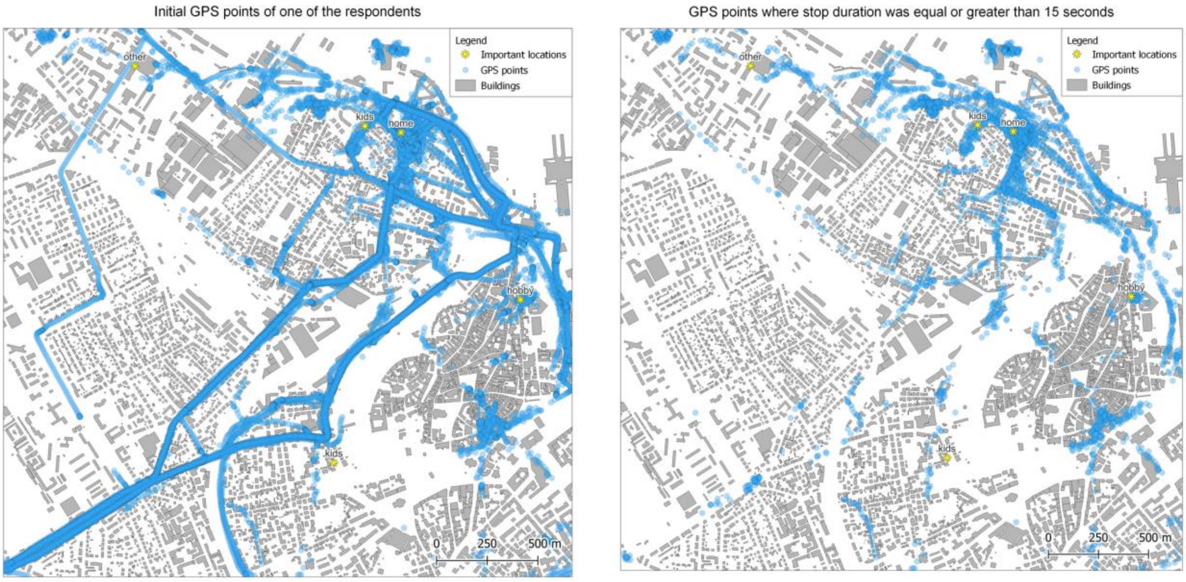


Figure 16. Frequencies of stop duration values (except 1-second stops)

It is clearly seen from Figure 17 how filtering significantly decreased the number of GPS points lying along the respondent's movement paths, for example along roads. Even though some points located within the area of important locations could be removed after the filtering, they contained much less valuable information than remained points – only 2.2% of the total stop duration time. Moreover, this operation improved computational time because of the reduced dataset. In the next, step it also ensured better quality of clusterization because the share of points that can be classified as stops has increased around important locations, whereas amount of points that could be classified as movement shrunk.



(a) initial GPS points of one of the respondents

(b) GPS points after filtering by time duration equal to or greater than 15 seconds

Figure 17. GPS points before and after filtering by time duration

4 Methodology

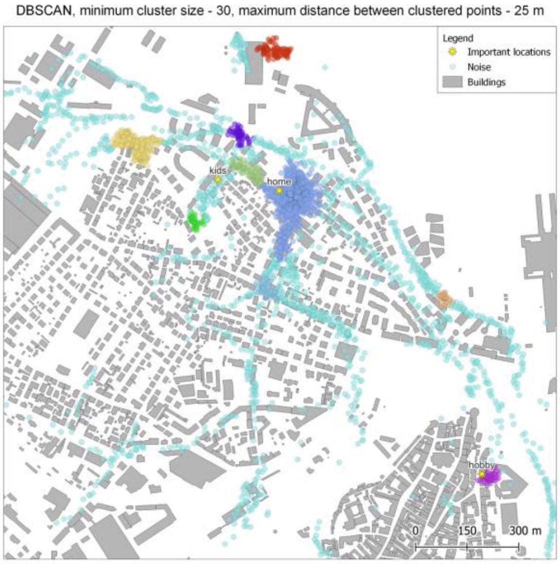
4.1 Clusterization

The next step was to carry out a clusterization to split clouds of points into groups and assign important locations of each respondent to these groups. One of the most commonly used clusterization methods applicable for spatial data and implemented in GIS software is DBSCAN – density-based spatial clustering of applications with noise. The core idea behind this algorithm is that dense regions of data points are separated by regions of lower density, thus it allows to identify clusters of different sizes and shapes. The advantage of DBSCAN compared to the other popular methods, such as k-means clustering, is that DBSCAN does not require a number of clusters as an input parameter. The other advantages of DBSCAN are resistance to outliers and scalability, which allows it to deal with large-scale datasets..

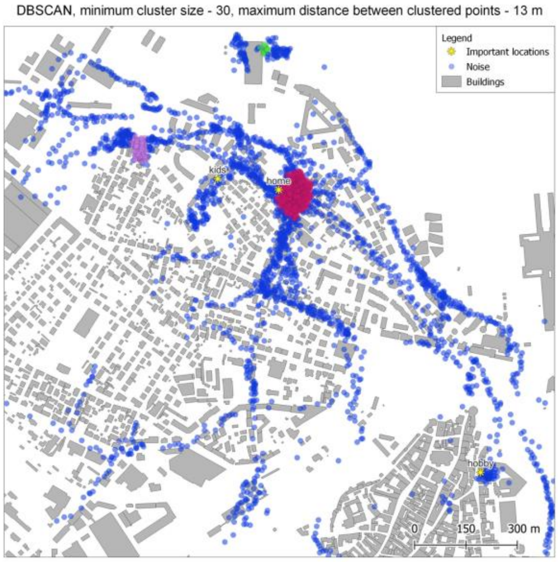
However, this method is quite susceptible towards input parameters: $minPts$ (minimum points) – minimum number of points required to form a dense region (cluster) and eps (epsilon) – a maximum bound on the distances of points within a cluster.

Figure 18 demonstrates how DBSCAN clusterization works with various input parameters using data of one of the respondents. Figure 18a seems to be close to the optimal clusterization. However, the cluster around the home location captures too many points nearby. As a result, the cluster turns out to be larger than in reality. Since in the provided example respondent lives in the apartment in the city centre, GPS points related to the home location should be located within the building and around a small area around it, while this condition is not met in the observed case.

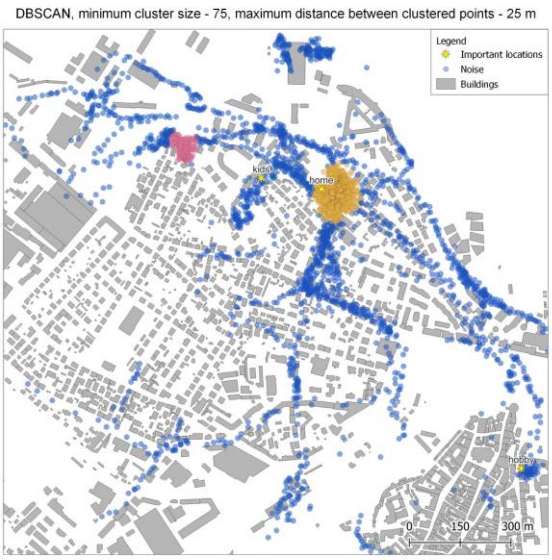
In order to solve this problem, input parameters may be changed in two directions: decrease the eps value to stop cluster extension or increase the number of points to extract more dense clusters. Although points around the home location are better clusterized in Figures 18b and 18c, there is a similar issue arising from such approach: GPS points near some locations, like the hobby location on the bottom right corner, are marked as noise, and they will not get in further analysis. Figure 18d demonstrates clusterization when both parameters have been increased, but it did not solve the problem with clusterization of points around the home location and also did not allocate points related to the hobby location to a separate cluster.



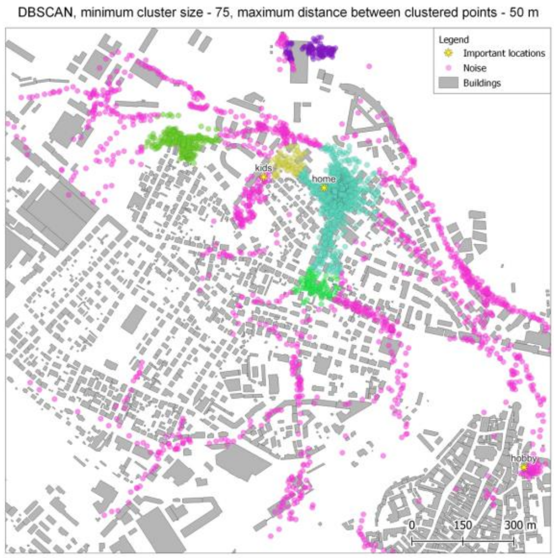
(a) DBSCAN with $eps = 25\ meters, minPts = 30$



(b) DBSCAN with $eps = 13\ meters, minPts = 30$



(c) DBSCAN with $eps = 25\ meters, minPts = 75$



(d) DBSCAN with $eps = 50\ meters, minPts = 75$

Figure 18. DBSCAN clusterization of GPS points with different input parameters

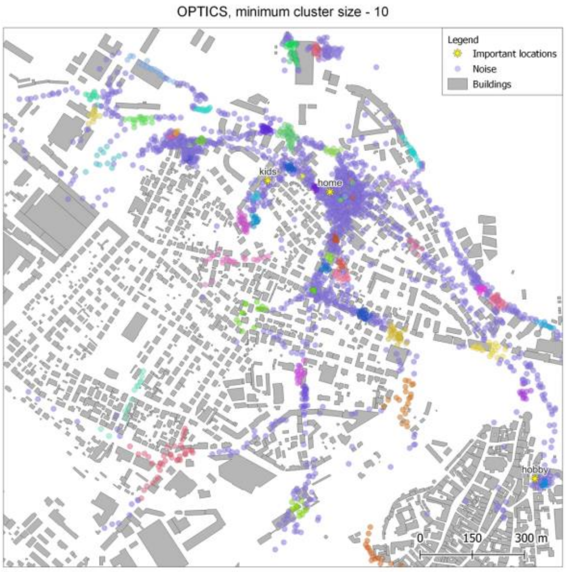
Furthermore, the input parameters should be tuned individually for each user, and it is required to design a measure of errors that would be relevant for the specifics of this dataset. Considering the complex of problems associated with DBSCAN, it was decided to review the alternative methods.

The scientific literature provides other methods to find clusters in data varying in density. For example, one of them is the OPTICS algorithm – Ordering Points to Identify Clustering Structure that serves as a hierarchical extension of the DBSCAN algorithm, resolving its weakness in the identification of clusters with varying densities. The core concept behind OPTICS is so-called “reachability distance”, which measures the density-based connectivity between data points. OPTICS algorithm defines a neighbourhood around each data point and orders points by reachability distance. The reachability distance of a point shows how densely the other points surround it. As a result, unlike DBSCAN, the OPTICS algorithm allows to explore all possible density levels.

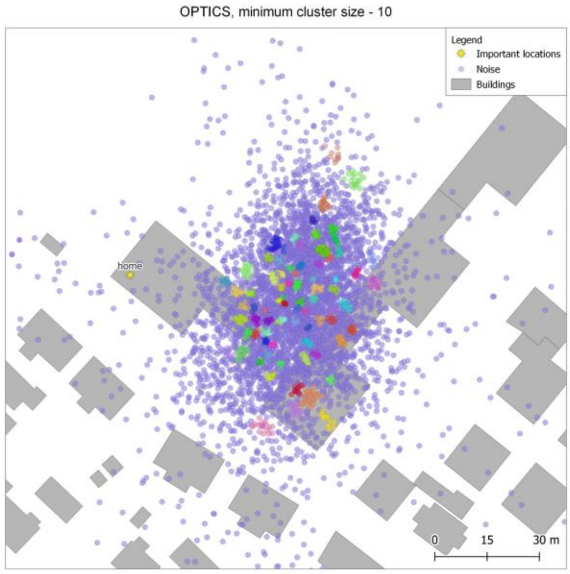
OPTICS algorithm requires only minimum cluster size as an input parameter, but the method is also quite sensitive to its choice. Figure 19 reveals how the OPTICS algorithm has been applied to the same sample as DBSCAN. Setting an input parameter $minPts = 10$, clusterization fails to distinguish large clusters, which is shown in Figure 19b, where the home location is zoomed in. A large cluster related to the home location is split into many small dense clusters. When $minPts$ was increased to 40, like in Figure 19c, the OPTICS algorithm classified all points near the home location as noise (Figure 19d). However, in both cases the algorithm could handle smaller clusters like clusters around the hobby location.

Consequently, the OPTICS clusterization algorithm is also irrelevant for the clusterization of GPS points in this work. Drawing a conclusion after application of these two algorithms, a method that is less susceptible to input parameters and can handle data of varied densities is needed. The algorithm should not divide large clusters into several small ones like OPTICS, but it should be capable to limit the cluster size if the density drops, which is the major drawback of DBSCAN.

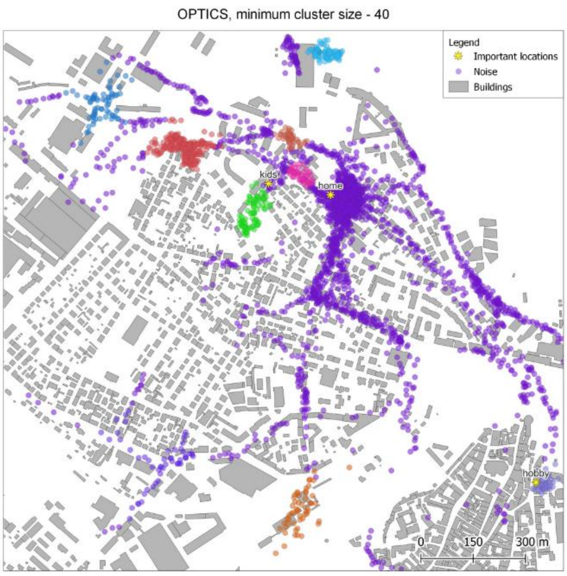
Malzer and Baum proposed a hybrid approach to hierarchical density-based cluster selection that combines the advantages of DBSCAN and hierarchical clusterization algorithms (Malzer and Baum, 2020). In this article, authors introduced HDBSCAN($\hat{\epsilon}$) – a combination between DBSCAN* and HDBSCAN – hierarchical DBSCAN that has been applied to a collection of GPS points of varied densities.



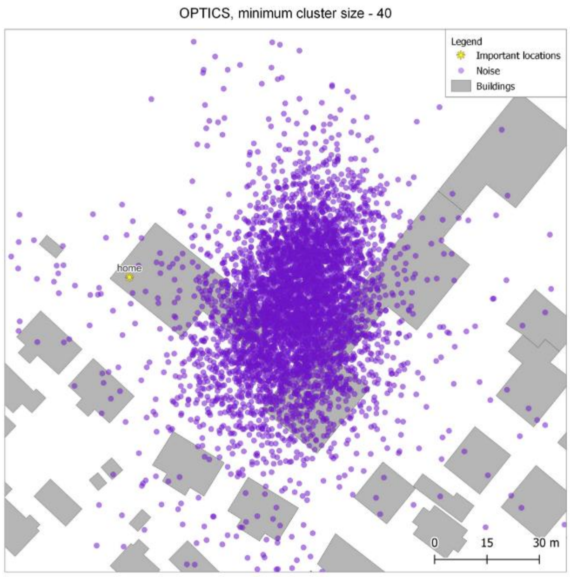
(a) OPTICS, $minPts = 10$



(b) OPTICS, $minPts = 10$, clusters around home location



(c) OPTICS, $minPts = 40$



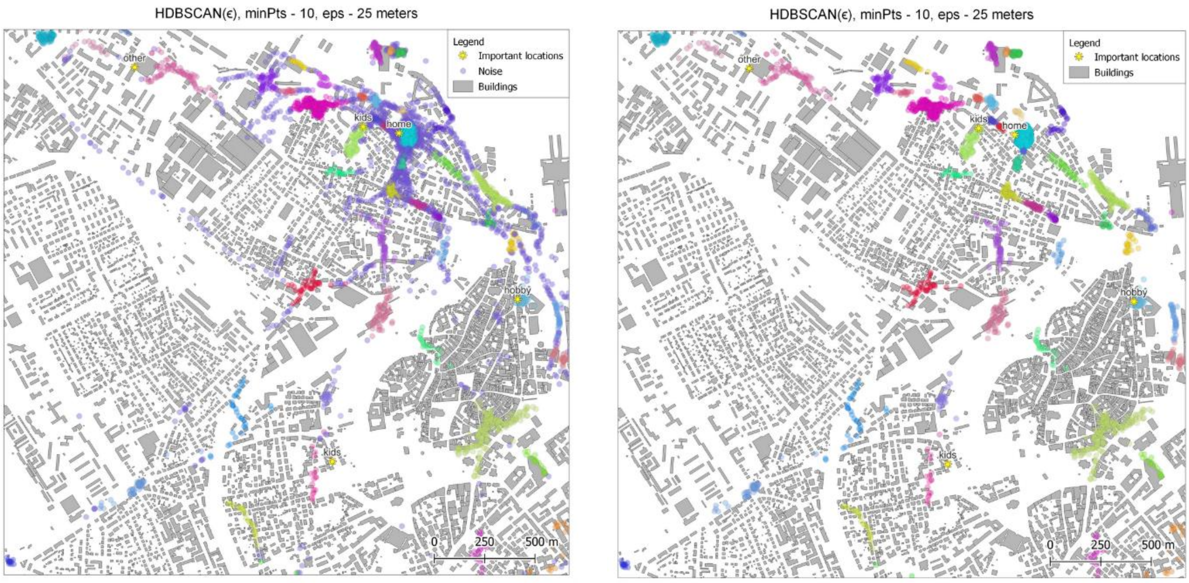
(d) OPTICS, $minPts = 40$, clusters around home location

Figure 19. OPTICS clusterization of GPS points with different input parameters

It turned out that HDBSCAN($\hat{\epsilon}$) outperformed the other clusterization algorithms such as DBSCAN, HDBSCAN, OPTICS and AUTO-HDS. The algorithm applies a distance threshold to HDBSCAN's hierarchy and it serves as a hybrid between DBSCAN* and HDBSCAN: for data partitions that have been affected by the given threshold $\hat{\epsilon}$ results of DBSCAN* are extracted, while HDBSCAN clusterization results are used for the remained points.

Figure 20a demonstrates the results of HDBSCAN($\hat{\epsilon}$) clusterization with $eps = 25\, meters, minPts = 10$. Different values of eps almost did not change the results of the clusterization, while it was decided that 10 GPS points would be enough to form a cluster because some respondents had rather a small number of points around important locations. It could happen because the period of data gathering is different for each respondent.

HDBSCAN($\hat{\epsilon}$) extracted clusters around both home and hobby locations, and unlike DBSCAN, it selected only a dense cloud of points near the home location, so the cluster did not capture a large number of points outside this location. The results of the algorithm also have been randomly checked for the other users. Overall, the algorithm managed to perform a qualitative clusterization in terms of distinguishing data of varied densities and identifying noise, however in some situations, when a density of points stayed even on a large area and points went far beyond the square of the important location, the algorithm was unable to cut the cluster.



(a) HDBSCAN($\hat{\epsilon}$) with $eps = 25\ meters, minPts = 10$

(b) HDBSCAN($\hat{\epsilon}$) with $eps = 25\ meters, minPts = 10$, without noise

Figure 20. HDBSCAN($\hat{\epsilon}$) clusterization of GPS points

After clusterization, data on the survey with important locations has been imported. Geometry column was built using longitude and latitude coordinates of points provided in WGS 84, then coordinate reference system was changed to the Estonian Coordinate System of 1997 – EPSG:3301.

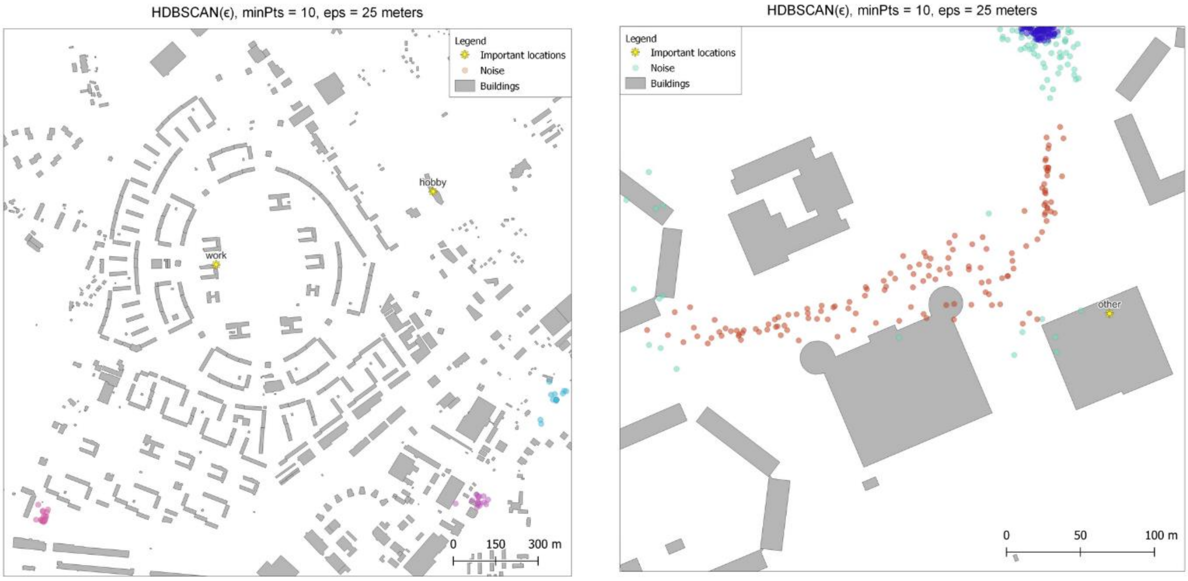
The geometry column of the survey geodataframe has been checked on empty values, and then such rows were removed. Duplicated records were also deleted by simultaneous checking of user id, location type and column containing information on address and coordinates.

Further, the number of clusters and points for each user was measured with the help of a geodataframe containing GPS coordinates. It turned out that after clusterization, all points of 9 respondents were clusterized as 1 cluster. It means that all of them were identified as noise. Table 1 reveals that the number of points for respondents with only one cluster varied from 3 to 126. Additionally, these points were not concentrated only in one area, therefore the algorithm could not find dense groups of points which could be extracted to a separate cluster, so all of them were classified as noise. Rows associated with respondents who had only 1 cluster were removed both from the survey and tables with important locations, and from the geodataframe containing GPS points.

Table 1. Number of respondents' points that had only 1 cluster

Number of points	Number of clusters
55	1
16	1
13	1
82	1
113	1
126	1
3	1
38	1
5	1

Later, it was needed to connect the received clusters and important locations of each respondent. A cluster should be associated with the closest important location. However, in some cases, no clusters were near important locations, as demonstrated in Figure 21a. The opposite situation was when the cluster went far beyond the actual space of the location, and it is shown in Figure 21b. These circumstances had to be considered in order not to assign clusters to important locations in these situations.



(a) Important location without clusters nearby

(b) Important location with close cluster going beyond space of the actual location

Figure 21. Examples of important locations where there are no close clusters or clusters cover much more space than the location

Additionally, there were some important locations that had several clusters nearby, such an example is provided in Figure 22. In this situation, both clusters are connected with the important location.

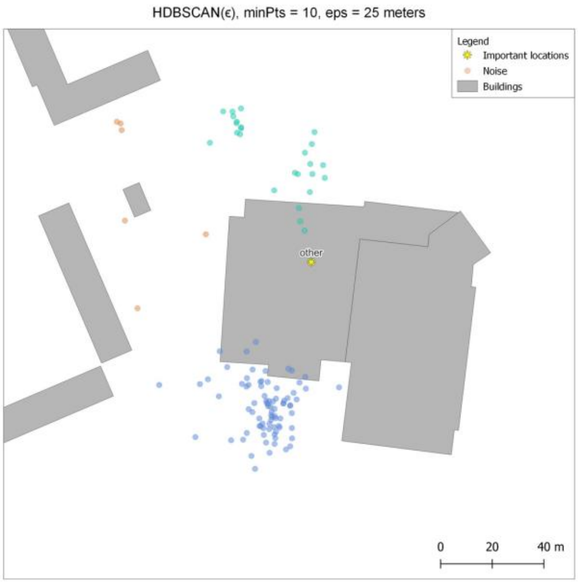


Figure 22. Several clusters were found near the important location

In order to take into account all the abovementioned specifics, thresholds on the distance from the important location to the cluster's centroids and their closest points have been applied. It was decided that a cluster would be assigned to the important location if a distance from the important location to the closest point of the cluster would not exceed 30 meters and the distance to the cluster's centroid would not exceed 100 meters. Further, important locations without assigned clusters have been removed from the geodataframe containing important locations.

However, after assigning clusters to the important locations, it turned out that in an insignificant number of cases, HDBSCAN with given parameters split large clusters into several smaller clusters like in Figure 23. Thus, to eliminate this problem, GPS points of users with more than 3 clusters near important locations have been clustered with the help of HDBSCAN($\hat{\epsilon}$) using larger amount of points required to form the cluster: $eps = 25, minPts = 30$.

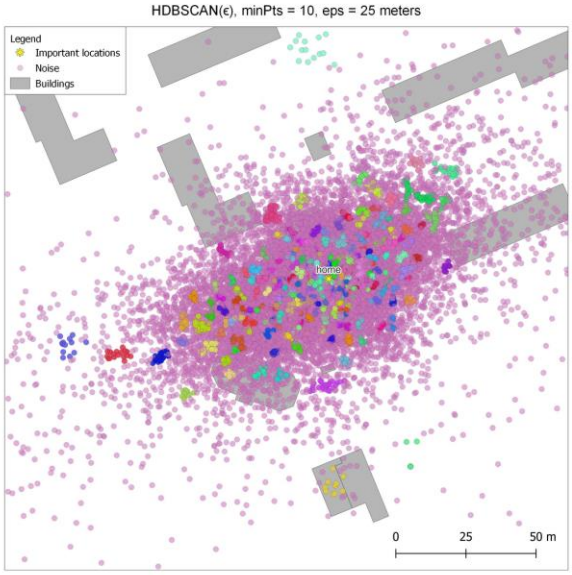


Figure 23. GPS points around the home location that were improperly clusterized

The next step was to remove duplicates in the important locations geodataframe. It was checked whether there were rows with duplicated respondent IDs and cluster IDs. 79 rows satisfied these criteria. 62 rows out of these 79 had duplicated addresses and user IDs, so these records contained different important locations with the same address. There was created a Table 2 to look at combinations of important location types.

Table 2. Number of rows corresponding to combinations of important locations having duplicated addresses and respondent IDs, and a new location type assigned to these locations

Combination of important locations	Number of rows	New location type
Home, work	19	Home
Hobby, other	3	Hobby

Kids, work	3	Work
Other, other	2	Other
Hobby, work	1	Remove rows
Home, kids	1	Home
Kids, other	1	Kids
Other, work	1	Remove rows

Home and work were the most common combination among the other important locations with the same addresses. Most probably, these people worked remotely from home, that is why they indicated their home and work location as the same place. The other variants were much less frequent.

In order not to exclude respondents with duplicated addresses of important locations that could result in the loss of a sufficient information, types of the important locations have been changed according to Table 2. If several important locations of the respondent with different types had the same address, they were either removed or changed to one location type based on Table 2.

For the remaining important locations related to the same clusters, the distances between them were calculated. All of these important locations were rather close to each other: the distance did not exceed 100 meters. The check showed that the density of GPS points did not drop between these important locations. Thus the algorithm could not assign points around these locations to different clusters.

If the distance between two important locations was not greater than 20 meters, then their types have been changed according to Table 3 because they were situated in the same building or apartment.

Table 3. New location types assigned to important locations situated not further than 20 meters away from each other with the same cluster ID and respondent ID

Combination of important locations	New location type	Number of occurrences
Home, other	Home	1
Other, other	Other	1
Other, work	Remove rows	1

The other locations were removed because it was impossible to assign one location type for such locations because GPS points related to these locations have been found in different buildings.

Finally, the list of important locations without duplicated respondent IDs and cluster IDs was joined with a geodataframe containing GPS points of respondents based on a combination of the respondent and cluster IDs in both tables. In this way, a relationship between GPS points and important locations was established.

4.2 Preparing data for the analysis

Some respondents had long time breaks between GPS records, which could affect analysis results, distorting time spent at important locations. A pivot table consisting of stop duration in days and a number of records has been created to handle this issue. The pivot table is shown in Table 4. It is clearly seen how the number of records sharply drops from records where the stop duration is 0-1 day to records with longer breaks. The pause between some GPS records could even reach 20 or more days. Probably, some respondents who agreed to take part in the survey did not use GPS trackers properly. Thus, records with long pauses between records can be assumed as outliers. Based on table 4, all records where stop duration exceeded 4 days were deleted.

Table 4. Pivot table with duration of stops in days and number of corresponding GPS records

Duration of stop, days	Number of records
0	2080808
1	3798
2	76
3	21
4	16
More than 4	57

The main time unit of the analysis was one calendar day, and thus if the pause between records either exceeded one day or a stop started in one day and finished the next day, then the records had to be split into several records by calendar days in order to measure the stop duration for each day. Figure 24 demonstrates how the stop time was distributed between calendar days if the stop lasted several days and if a stop happened between two calendar days.

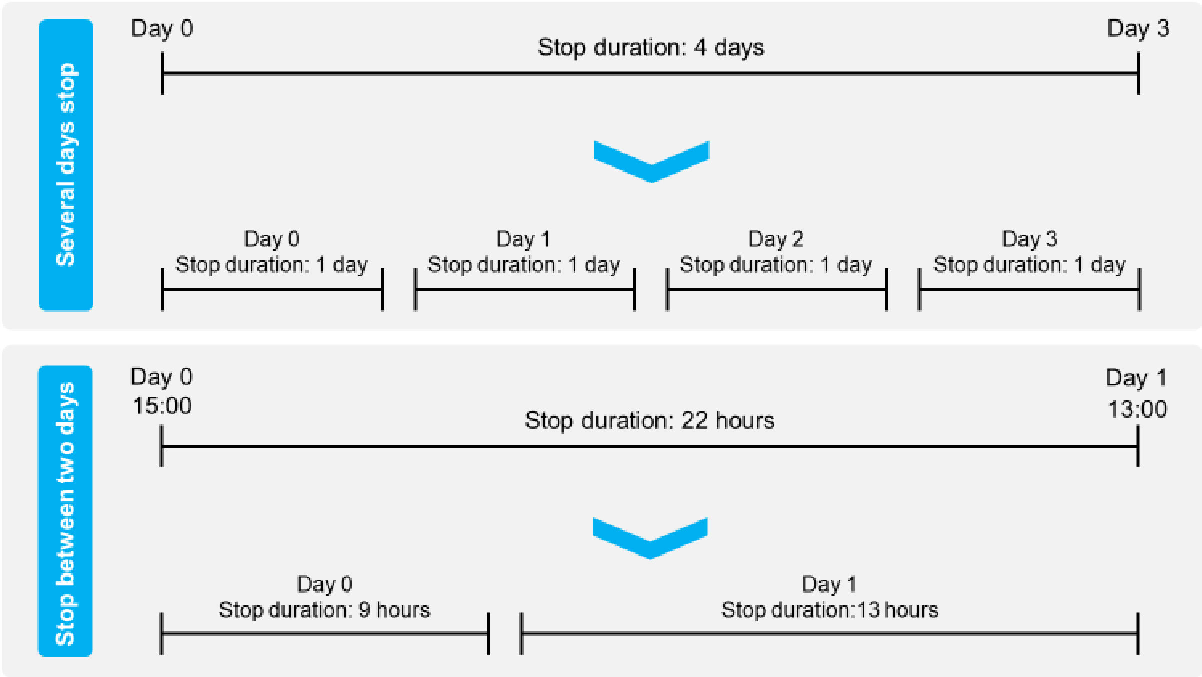


Figure 24. Distribution of stop time between calendar days if the stop lasted several days and if a stop happened between two calendar days

Afterwards, the number of unique respondents per day was received and visualised with the help of a line chart shown in Figure 25. Even though the data was collected between October 2020 and February 2021, the number of respondents on each day significantly varied. We can observe a massive spike in November when the number of unique respondents exceeded 100 people, but before and after November, amount of users rarely was higher than 20.

This research aimed to examine differences between socioeconomic and ethnic population groups, so if we split less than 20 respondents by groups then the data will not be representative enough, and the composition of respondents can be different in November and the other periods. It can lead to the distortion of the analysis and its further outcomes. Considering these facts, only data for November has been examined.

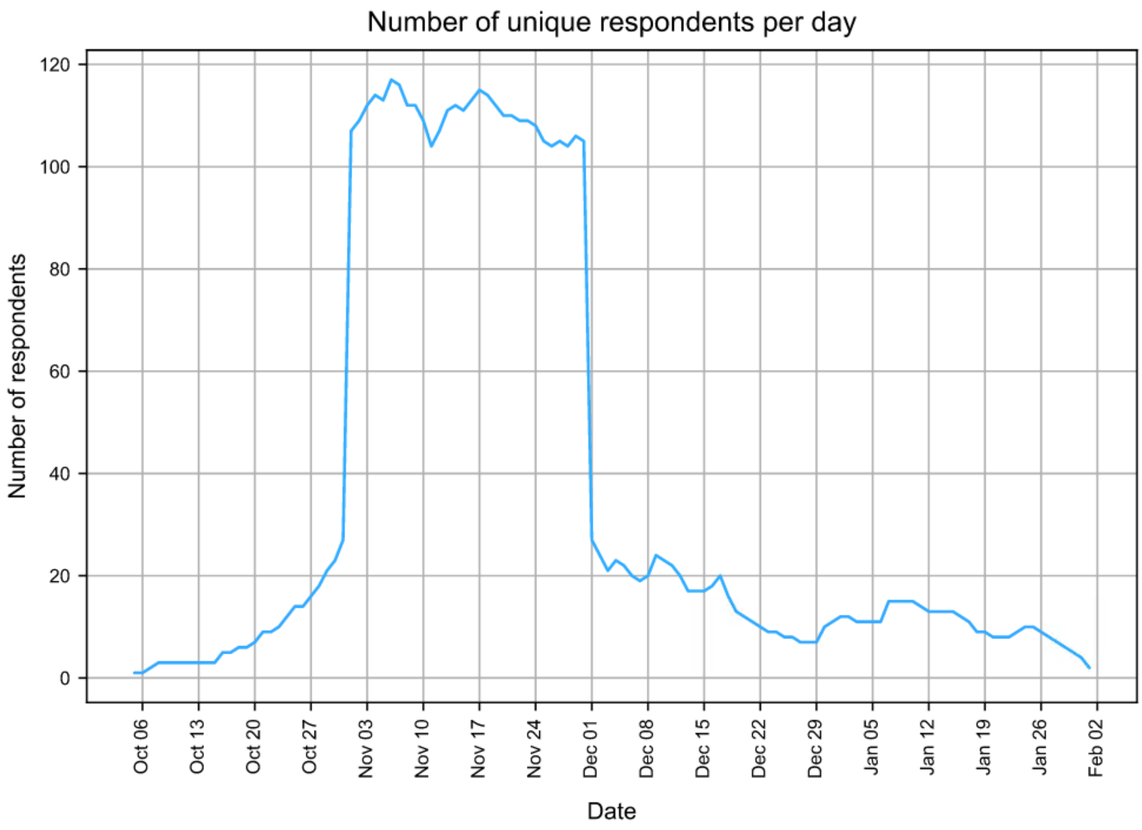


Figure 25. Number of unique respondents per each day

Finally, locations related to hobbies and kids have been replaced by the "other" location type. This decision was taken because when the analysis was carried out, there were too few respondents after splitting the data into population groups. Also, there were overlaps between these categories when some locations, like restaurants or parks, could be in multiple categories simultaneously.

Also, location type "work other" was combined with work location because "work other" locations were connected with the regular job.

4.3 Analysis methods

There were written three different Python functions to measure differences between population groups in terms of staying at important locations. There were calculated the following metrics:

1) The mean amount of time that one respondent from the selected group spent at the particular location type per day.
2) Share of people from the particular population group who visited selected important location per day.

3) Relative (normalised) number of people from the selected population group that were at specific important location within 30-minute time intervals on weekdays and weekends.

These functions took location type and a list of respondent IDs as input parameters. The last function also considered weekdays and weekends because the distribution of people staying at specific locations like home or work can significantly differ for weekdays and weekends.

For the first metric, a geodataframe containing GPS points was filtered by a list of users related to the selected population group. Then it was additionally filtered by the location type. The same procedure has been applied to the other functions at the beginning of the algorithm. Afterwards, the filtered geodataframe was grouped by respondent IDs and date. Grouped dataframe was summed by the stop duration column, so the total time spent at the important locations per day by each user has been received.

In order to count a number of people who visited a specific kind of location per day, the number of stops made by each user at the considered location per day was calculated. It has been done by grouping the geodataframe by respondent IDs and date, and aggregation via the count function that took stop IDs. Then if the user had at least one detected stop at the location, it was assumed that this person visited this location on this day. If there were no detected stops at this location, the user did not visit it. In other words, a number of unique users who visited the location per day was identified by the presence of respondent's stops at this location. For that purpose, a table was created grouped by respondent IDs and date, and value 0 was assigned if the user did have stops at the location on this day, and 1 was assigned if there was at least one stop at the location.

The final result was calculated as a division of total time spent at the selected location by all users from the population group per day by the number of unique users from the investigated population group that had stops at this location on the same days. It gave us a mean time spent by each user from the population group at the investigated location type per day.

The second function was based on the algorithm similar to the first function. After geodataframe filtering, a table was created which was grouped by user IDs and dates consisting of 0 and 1 corresponding to the presence of stops of the particular user on this day. The result was measured as a division of a number of people from the population group who visited the investigated kind of location on this day by a number of respondents from the same population group who visited this kind of location throughout the whole period and had records on the same day. Applying such rules, we avoid the situation when we include respondents who did not have any stops at the selected type of location within the considered period, and we take into account the fact that the number of respondents was different each day.

To apply the last function, a separate geodataframe has been created. Based on the initial geodataframe, there was used a function that returned the range of equally spaced time points so that all of them would satisfy the following condition: $start\ time \leq\ x \leq end\ time$, where the first one and the last one are, respectively, the first and last time points in that range that fall on the boundary of 30-minute time interval (Pandas, 2023). In other words, if a time interval exceeded 30 minutes, then there were added additional time points between start and end time points that would be equally spaced, and the number of time intervals would be equal to the rounded-up number of 30-minute time intervals falling in this time range. Further, additional rows were created. The number of these rows corresponded to the number of created time points. Time of each day has been extracted to the new column and rounded down by half an hour. An example of this process is shown in Figure 26.

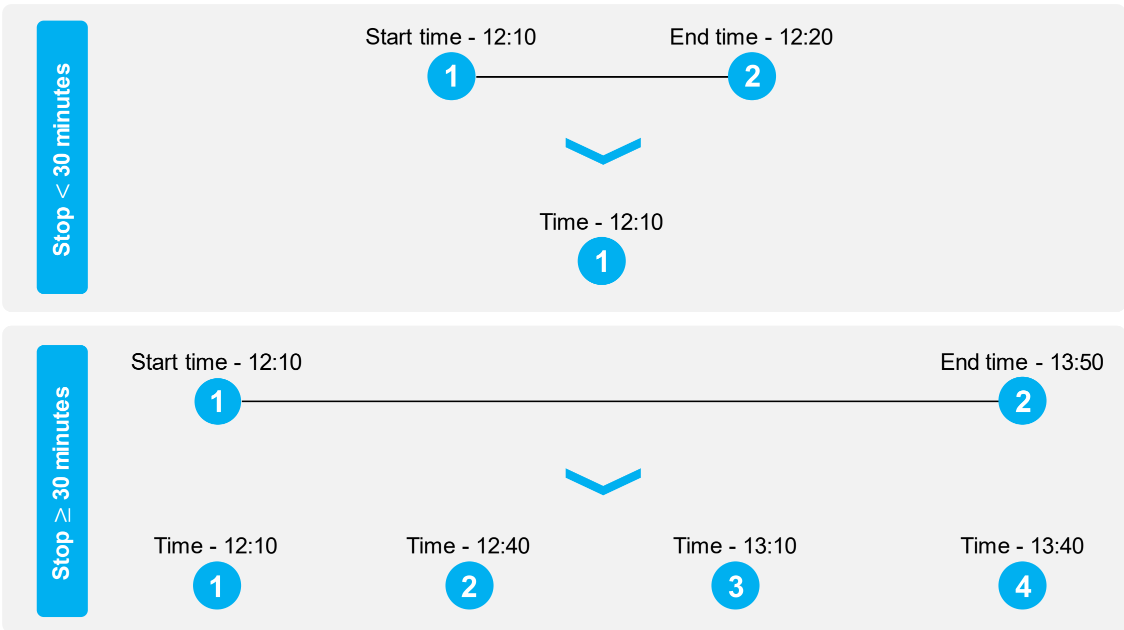


Figure 26. Scheme of splitting stop duration time by 30-minute timestamps

In this function, specially created geodataframe was also filtered by a list of users related to the specific population group and by the location type. Then all remaining records were additionally filtered by weekday and weekend. The number of location visits (or the number of stops by a user detected at the location) was calculated by grouping the dataframe by user IDs and a column containing date and time rounded down to half an hour and aggregation by the count function. The time has been extracted to the new column, and the result was calculated by grouping the previously received table by the extracted time, which was rounded down by half an hour, and the number of unique users for each timestamp has been measured. So, we got a number of unique users related to the specific population group that visited the investigated kind of location in each 30-minute timestamp within a day on weekdays and weekends during the whole observation period.

Outputs of all functions have been visualised with the help of line charts. Time units of the mean amount of time spent at the important location and percentage of people visited the location were days, while the relative (normalised) number of people that were at specific important location within a weekday or weekend was half an hour.

All location types were considered, but some population groups did not have enough records required to build a chart because there were too few stops at these locations, so plots for some locations could not be created. Also, in the case of the last function, if there was only one respondent from the population group at some point of time, then the line chart for this population group was skipped because such data was not representative.

For better results representation, the outputs of the two first functions were smoothed using a 5-day moving average. Due to the usage of 5-day moving average data for 1-4 November would be skipped. In order to avoid that, the minimum number of observations in the window required to have a value was set to 1 day. In other words, if there were not enough values to calculate mean based on five days, then the mean would be measured using data for available days. As for the last function, the number of unique users at each timestamp varied depending on the size of the population group. To eliminate the influence of the population's group size, output values were normalised separately for each group.

All three functions have been applied to respondents that were split into population groups by the following categories:

1) Ethnic by mother tongue, two groups: Russian speakers and Estonian speakers.
2) Economic by salary, two groups: respondents whose income was equal to or greater than 2000 euros and respondents that earned less than 2000 euros.
3) Combination of ethnic and economic categories: Estonian and Russian speakers whose salary was equal to or more than 2000 euros, Estonian and Russian speakers who earned less than 2000 euros.

4) Occupation, number of groups varied depending on whether the minimum number of respondents in the group was greater than 1. There were nine occupational groups (equipment and machine operator or assembler, manual labourer, middle manager, mid-level specialist, official, senior manager of an institution or company, service and sales staff, skilled worker or craftsman, top specialist) excluding variant “other” and empty values.

Besides functions showing aggregated statistics of staying at important locations, there were calculated correlations between different features using information on each respondent from the survey. Additional columns were added indicating location types that users listed. If the respondent listed one or more location types that he or she visited, then the value in the column corresponding to the location type for this user was 1, else if the respondent did not list such kind of location, then this location type and the user had an empty value, however later they were replaced by 0. The remaining empty values in the survey table containing respondent information have been replaced by “No data”.

As some columns contained categorical data, dummy variables were received for such columns. Dummy variables are additional columns created for each unique categorical value consisting of 0 and 1, indicating the absence or presence of categorical value. Further, Pearson correlation coefficient values have been received and have been left only values that in absolute value were equal to or greater than 0.3, which corresponds to a weak correlation.

Additionally, there were measured values of the Pearson correlation coefficient between the mean amount of time spent at home and work locations per day and the other features. The mean amount of time spent at the location by each user was measured as the total time spent at the location by the user throughout the whole period divided by the number of days when the person had records. Respondents were not considered if their mean amount of time at the location was an empty value.

Unfortunately, attempts to build regression models where the mean amount of time spent at home or work location were dependent variables and the other features were independent in order to examine which factors had the highest impact on time spent at these locations failed because none of the models including linear regression model, Lasso regression, Random Forest and Gradient Boosting models was unable to demonstrate good performance on test data, because some or all values of determination coefficient in cross-validation were negative what shows that a model can not be used for data prediction.

5 Analysis results

5.1 Correlation analysis

Correlation analysis allows us to identify key relationships in data and confirm or deny assumptions. In Table 5, the main outputs of the correlation analysis have been collected. According to the table, a mean positive relationship exists between place of residence and mother tongue. Correlation between residency in Kalamaja and Estonian as a mother tongue is 0.4, while a correlation between residency in Priisle and Russian as a mother tongue is almost the same and is equal to 0.41. The correlation is stronger if the mother tongue and the main language in the area were the same, about 0.5. Finally, the strongest correlation has been observed between the place of residence and the main language in the area. Place of residence in Kalamaja has a rather strong relationship with Estonian as the main language in the area, the similar situation was in Priisle, where the correlation with the Russian language as the main language in the area is 0.6. It confirms that Estonian and Russian speakers tend to live in neighbourhoods populated by people speaking the same language.

The other correlation, close to moderate, was detected between mother tongue and income level. The correlation coefficient is positive between income level less than 2000 euros and Russian as the mother tongue but negative when Estonian was the mother tongue. So, Russian speakers earned lower salaries more frequently, while Estonian got higher salaries.

Analogical opposite relationships between Estonian speakers and Russian speakers have been found in the case of mean time spent at work. Respondents whose mother tongue was Russian tended to spend more time at work, while Estonian speakers spent less time at work. Moreover, the correlation was even stronger between Russian speakers whose incomes were less than 2000 euros per month and their mean time at work, which was equal to 0.5, while such a relationship without consideration of the mother tongue was lower – only 0.45. So, Russian speakers earning lower salaries spent more time at work than the other population groups. Interestingly, there was some positive relationship between people earning a higher salary in range 2501-3000 euros per month and mean time at work.

The other interesting relationship has been found between time spent at the other location types and occupation status. So, people employed in the information and communication technology sector and entrepreneurs spent more time at the other location types than people from the rest of the occupation groups.

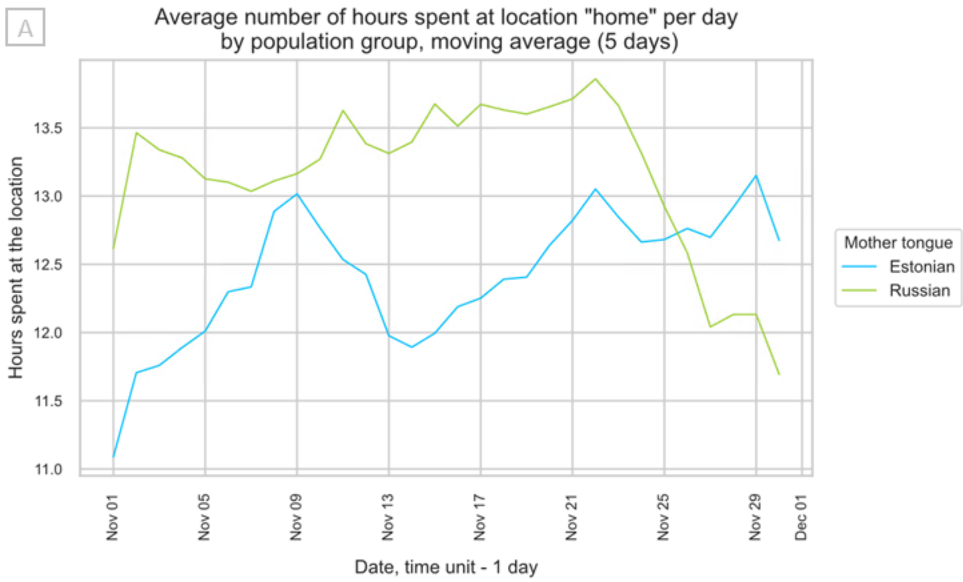
Table 5. Correlation analysis, main outputs.

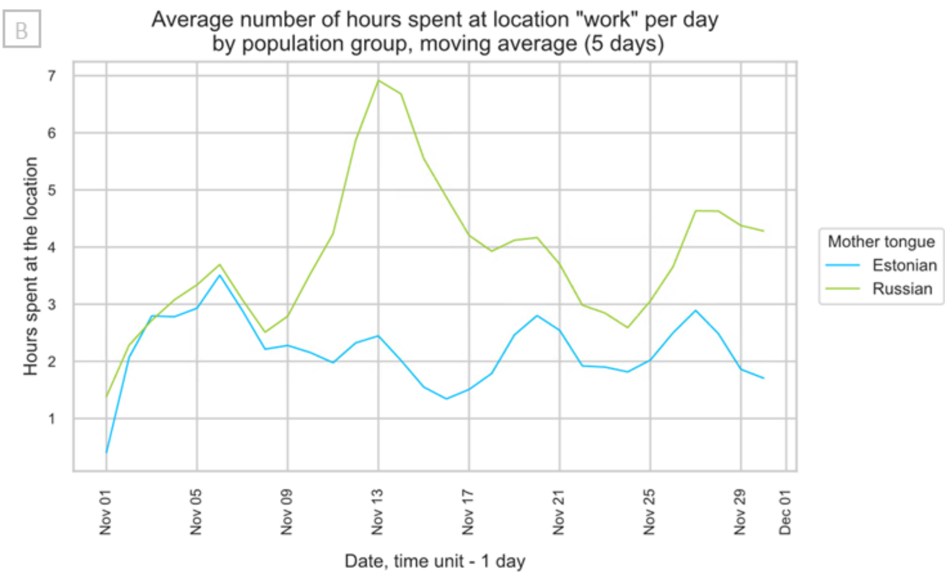
Category	Feature 1	Feature 2	Correlation coefficient value
Language in the area	Residency – Kalamaja	Mother tongue – Estonian	0.4
	Residency – Priisle	Mother tongue – Russian	0.41
	Main language in the area – Estonian	Mother tongue – Estonian	0.5
	Main language in the area – Russian	Mother tongue – Russian	0.52
	Residency – Kalamaja	Main language in the area – Estonian	0.73
	Residency – Priisle	Main language in the area – Russian	0.6
Income and mother tongue	Income is less than 2000 euros per month	Mother tongue – Russian	0.37
	Income is less than 2000 euros per month	Mother tongue – Estonian	-0.37
Mean time spent at work and income or tongue	Mean time at work	Income 2501-3000 euros per month	0.31
	Mean time at work	Mother tongue – Russian	0.36
	Mean time at work	Income is less than 2000 euros per month	0.45
	Mean time at work	Income is less than 2000 euros per month and mother tongue – Russian	0.5
	Mean time at work	Mother tongue – Estonian	-0.36
Mean time spent at other locations and occupation status	Mean time at other locations	Information and communication technology	0.33
	Mean time at other locations	Entrepreneur with paid labor	0.39

5.2 Aggregated measures of staying at important locations

Aggregated measures of staying at important locations were the central part of the analysis that enabled us to find out what differences were between people from different economic, ethnic and occupation groups in terms of staying or visiting various location types such as home, work and other locations. It is important to note that some plots where differences between population groups could not be observed explicitly have been discarded.

Figure 27 demonstrates the average number of hours spent at home, work and other locations by Estonian and Russian speakers. According to Figure 27a, it is well seen how Russian speakers spent more time at home, but at the end of the month, situation changed in favour of Estonian speakers. Russian and Estonian speakers spent equal amount of time at work only at the beginning of November, while further the gap has increased because Russian speakers stayed at work more than Estonian speakers. At other locations, the gap between the two groups was very substantial, and the situation has remained the same across the whole period: Estonian speakers spent from 1.7 to 2.5 hours per day at other locations. For Russian speakers, this indicator fluctuated between 0.3 and 1 hour per day.





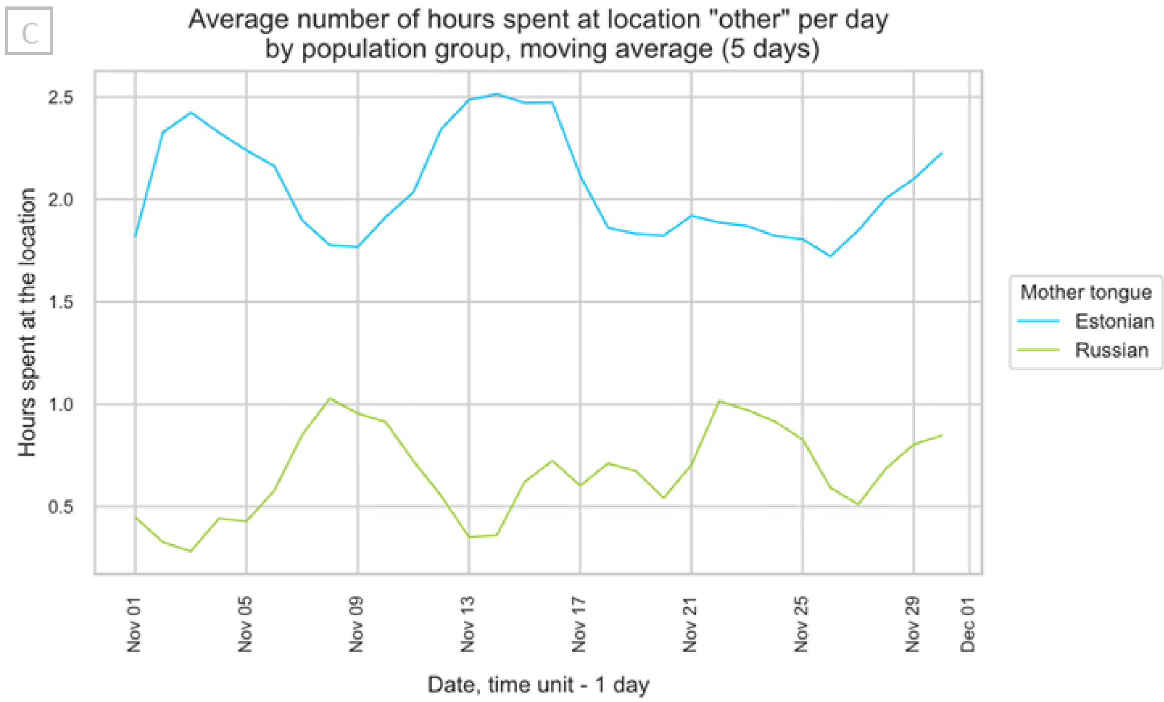


Figure 27. Average number of hours spent by Estonian and Russian speakers at: (A) – home location, (B) – work location, (C) – other location per day.

Figure 28 demonstrates the difference in the number of hours spent at important locations per day between the two economic groups. Significant changes in the second half of the month, when new COVID restrictions were imposed in Estonia, can be noticed in all locations. Based on Figure 28a, respondents with income equal to or greater than 2000 EUR spent much less time at home until the 21-st of November, when a considerable spike was observed that aligned the two groups. Figure 28b reveals how people with higher incomes stayed less at work location in the first half of the month, but in the second half of November, this difference grew even further due to reduced time that the higher income group spent at work. Figure 28c shows a reversed picture of Figure 28a when the high-income group spent more time at other locations. However, after a significant drop two groups became equal, but after 27-th November higher income group demonstrated a new upward trend.

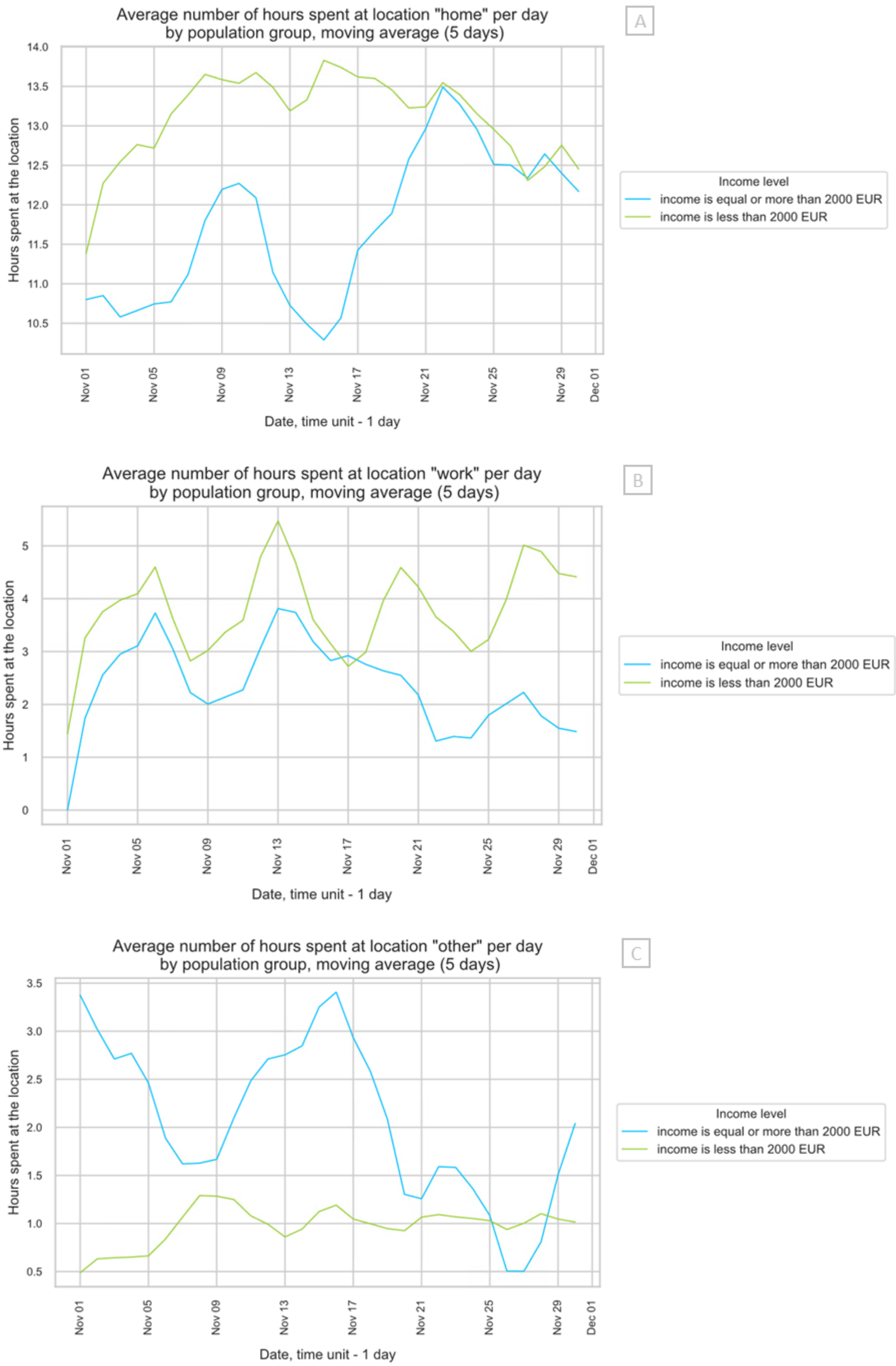


Figure 28. Average number of hours spent by people with income equal to or more than 2000 EUR per month and people whose income was less than 2000 EUR at: (A) – home location, (B) – work location, (C) – other location per day.

Figure 29 represents the combination of two previous groups of plots considering income and mother tongue at the same time. Figure 29a shows how Estonian speakers with incomes equal to or more than 2000 EUR consistently increased their time spent at home throughout the month. Time changes of Estonian speakers earning less than 2000 EUR were smaller, but at the end of the period, they spent the same number of hours at home as people from the same ethnic group earning higher salaries. As for Russian speakers, people with lower incomes from this group most of the observed period spent the greatest amount of time at home, but in the last days, duration of staying at home decreased. Russian speakers with higher incomes spent less time at home than those with low incomes but more than Estonian speakers with high incomes. However, after 22nd November, they became a group that spent the least time at home.

The following plot, Figure 29b, shows that Russian speakers stayed more at work than Estonian speakers, while the same specifics are inherent in people with higher incomes that stayed less at work and low-income population groups. So, Russian speakers who earned less than 2000 EUR per month spent the most time at work, whilst Estonian speakers with a salary of 2000 EUR or more stayed the least at their workplace.

Figure 29c displays specifics which are opposite in relation to the workplace. Russian-speaking population with lower incomes spent the least time at other locations, but Estonian speakers with the same incomes spent 1-1.5 hours more at such kinds of locations, which was similar to Russian speakers earning 2000 EUR or more except the higher fluctuations characteristic of the group. Estonian speakers with high incomes spent the greatest amount of time at other locations, but after imposed restrictions, there was observed a dramatic drop which shows that people from this group almost did not stay at other locations. The upward trend could be seen only in the end of November.

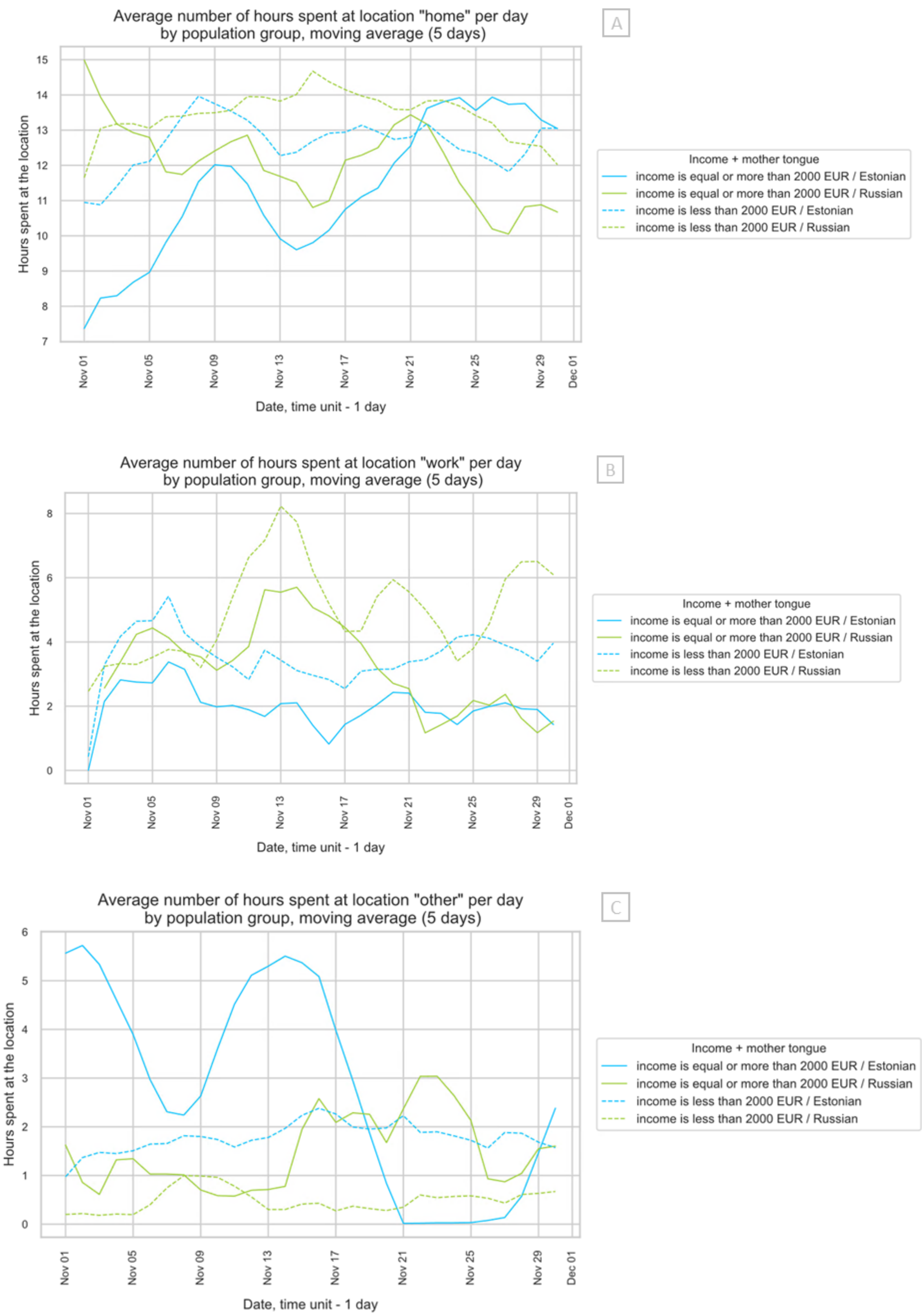


Figure 29. The average number of hours spent by people from Estonian and Russian language groups combined with income levels (below 2000 EUR and equal to or more than 2000 EUR) at: (A) – home location, (B) – work location, (C) – other location per day.

Talking about differences between occupation groups, mid-level specialists, skilled workers or craftsmen, top specialists, and middle managers (except some drop in the middle of November) spent more time at home than people employed as senior managers of an institution or company and service and sales staff based on Figure 30.

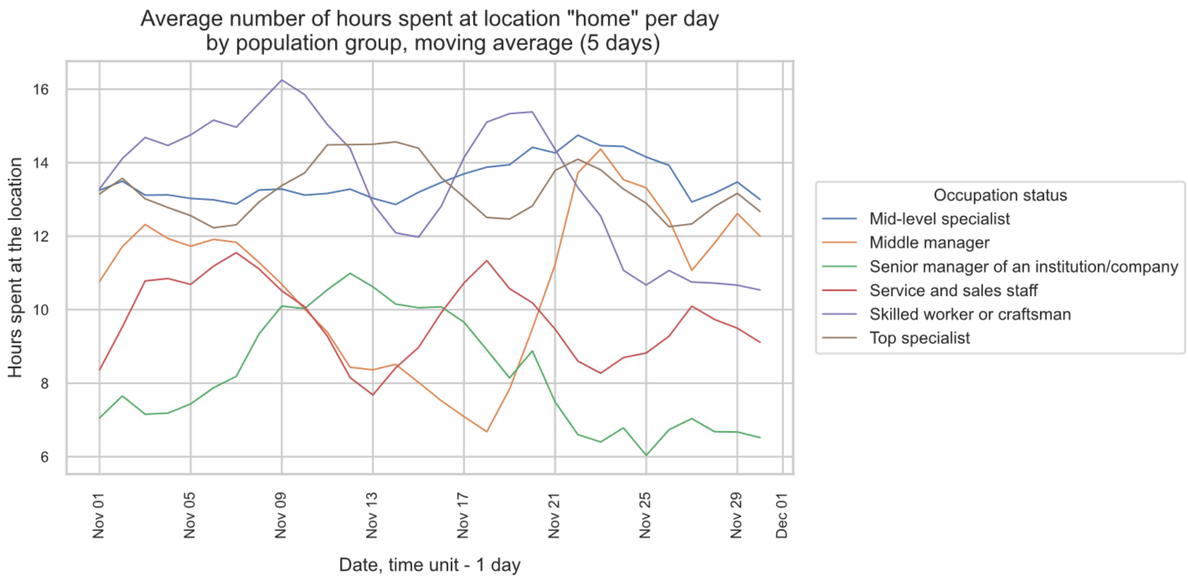


Figure 30. The average number of hours spent by people from different occupation groups at home location per day.

The next list of plots is identical to the previous one, but instead of the number of hours spent at the important location, a share of people from the population group visited investigated kind of location was used.

Figure 31 shows differences between ethnic groups. Estonian speakers visited their home locations less frequently, according to Figure 31a, but the Russian-speaking population group also reduced its attendance at the end of November. Attendance of work location was even higher among Estonian speakers at the beginning of the month, but later share of work visitors shortened. The percentage of Estonian-speaking people attending other locations was much larger in the early phase. Nevertheless, after the 25th of November, both groups fluctuated at the same level.

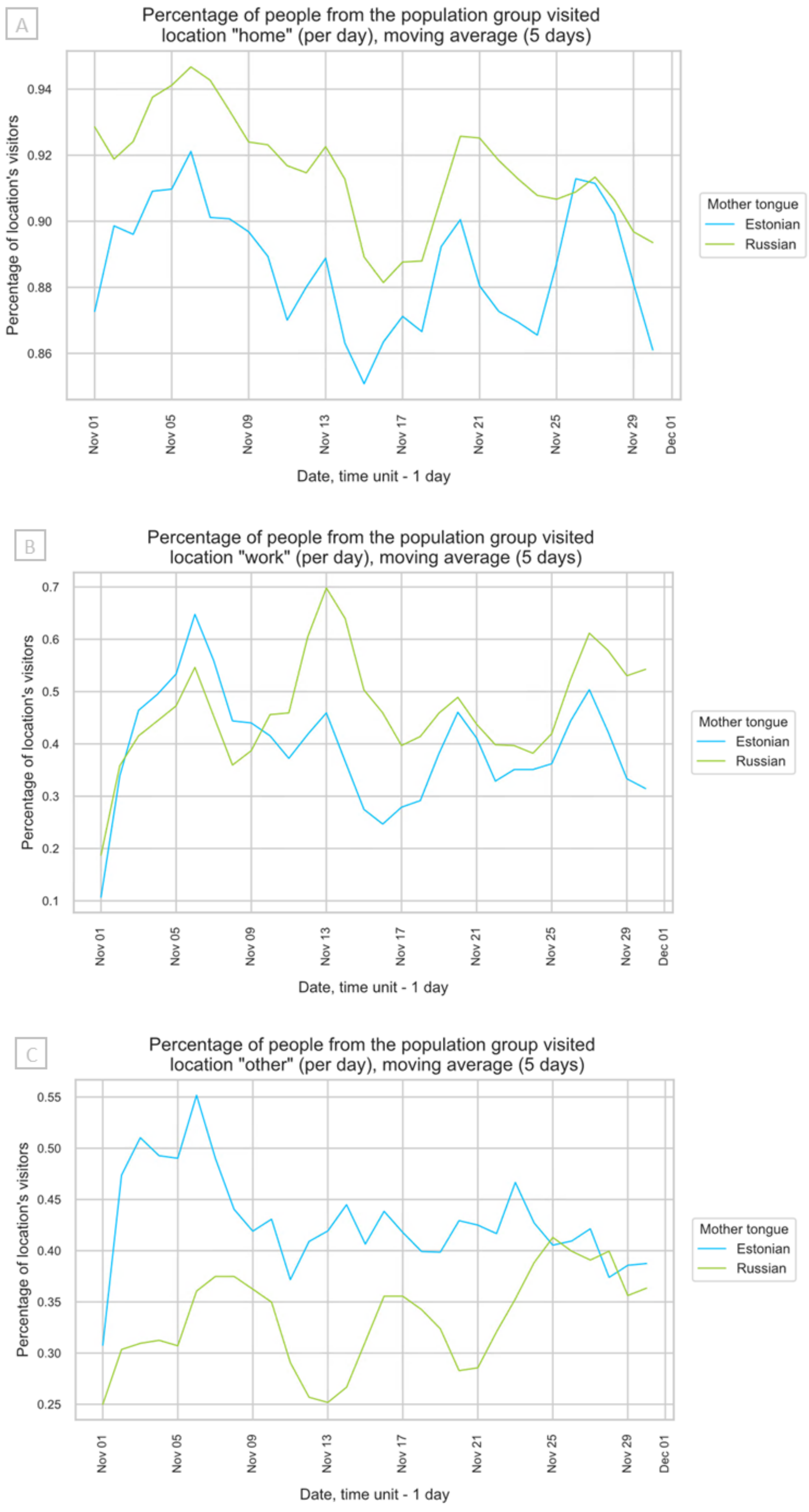
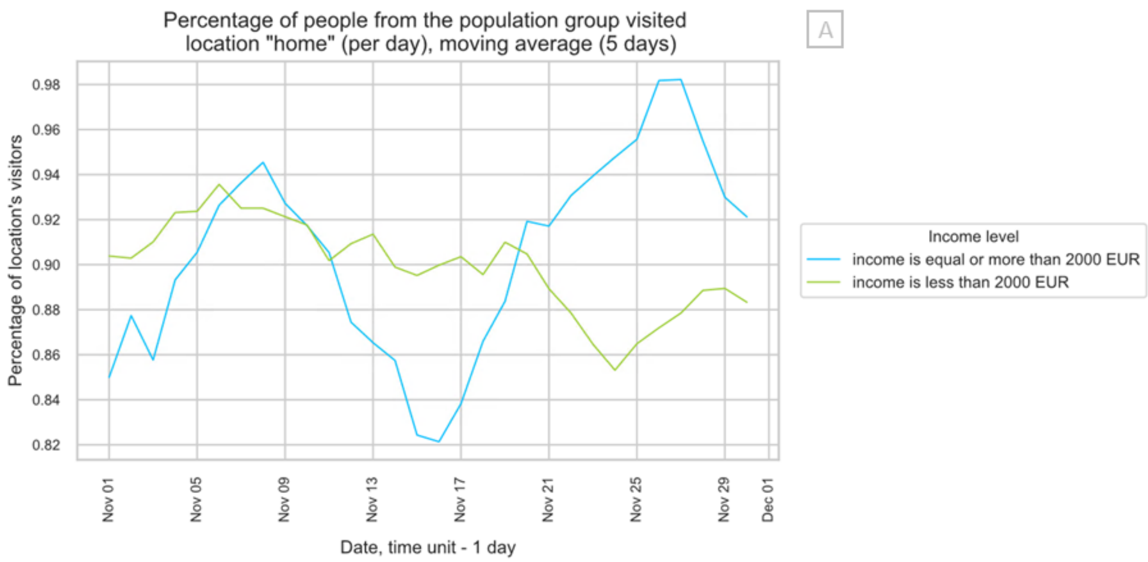
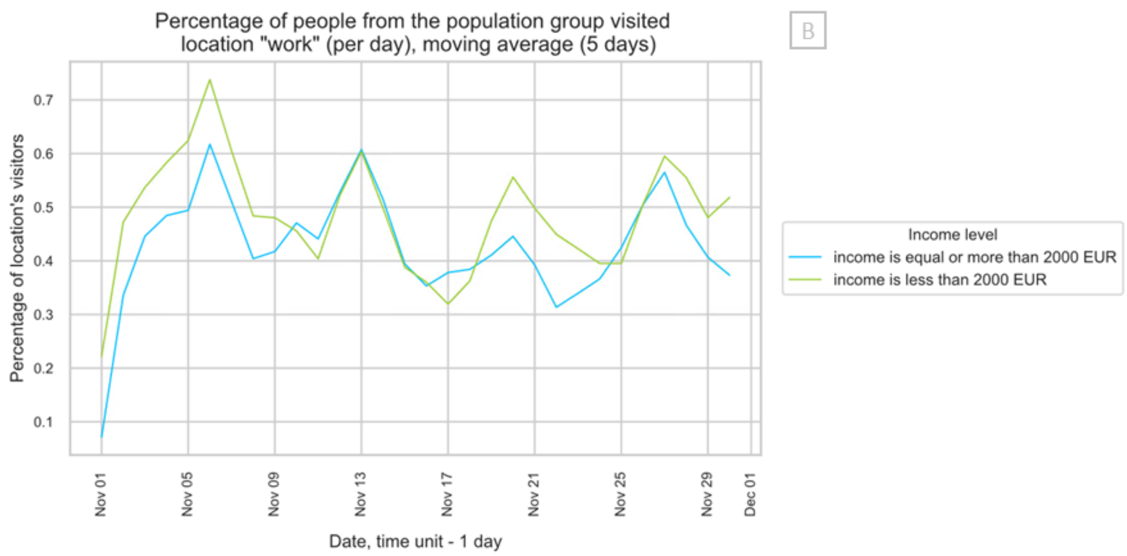


Figure 31. Share of people from the Russian-speaking and Estonian-speaking population groups visited: (A) – home location, (B) – work location, (C) – other location per day.

The other set of figures is devoted to differences between income groups. Figure 32a reveals a steady decline in the percentage of home location attendants among respondents from the low-income group, whereas respondents earning 2000 EUR per month or more showed significant fluctuations, but after 20th November, 90% and more people from this group visited their home locations, and this value was much greater compared to the low-income group.

The percentage of people from the high-income group was almost constantly lower compared with the percentage of respondents from the low-income group who visited work locations, which is shown in Figure 32b. The gap between groups was greater in Figure 32c, where a higher share of respondents who had an income of 2000 EUR per month or more attended other locations. Like in the previously described figures, at the end of November the behaviour of people from this group has changed and at that time share of people visiting other locations was the same in both groups.





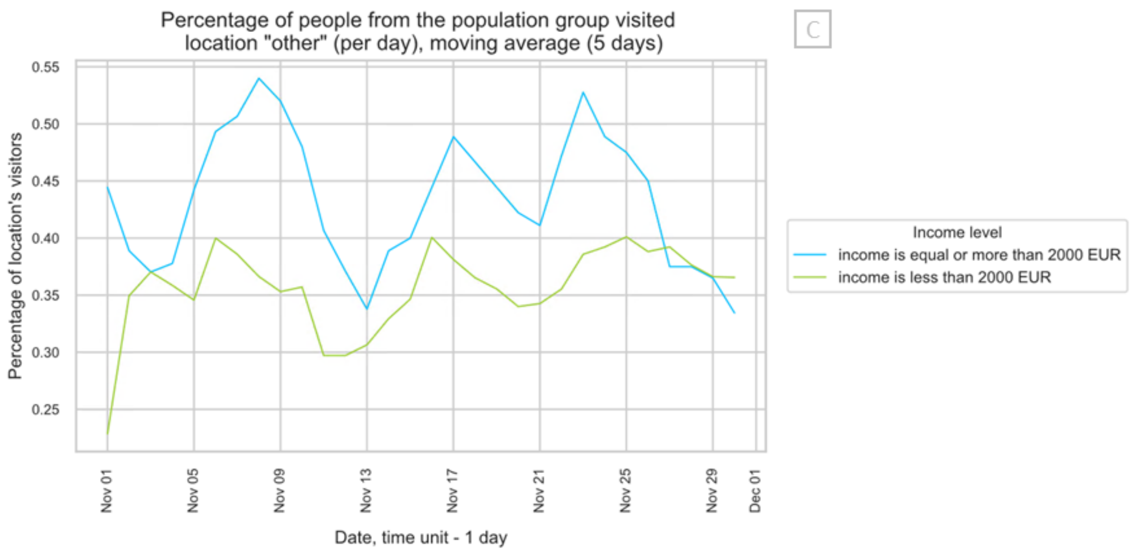
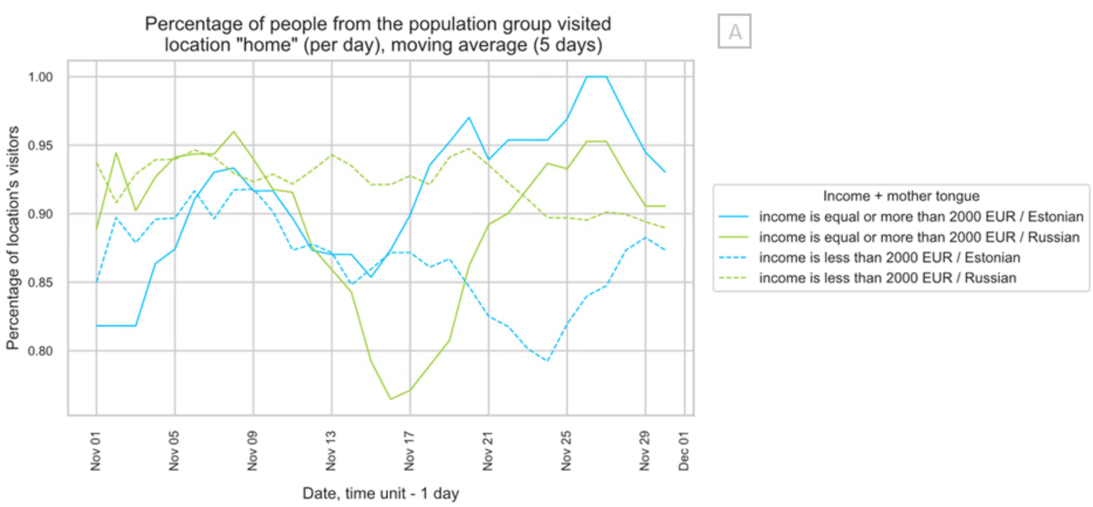


Figure 32. Share of people with income equal to or more than 2000 EUR per month and people whose income was less than 2000 EUR that visited: (A) – home location, (B) – work location, (C) – other location per day.

Figure 33 takes into consideration both mother tongue and income level. In Figure 33a, differences between population groups were less noticeable in early November as at the end of the month when a greater percentage of respondents from both high-income groups visited home rather than representatives of low-income groups.

Regarding visiting work location, people with higher incomes, on the contrary, attended it less frequently compared to low-income groups almost regardless of the mother tongue factor, as seen in Figure 33b. Finally, the percentage of people visiting other locations was the lowest in the Russian-speaking population group with low income level based on Figure 33c. Despite large fluctuations in the other population groups, the percentage of other locations visitors reduced at the end of the period.



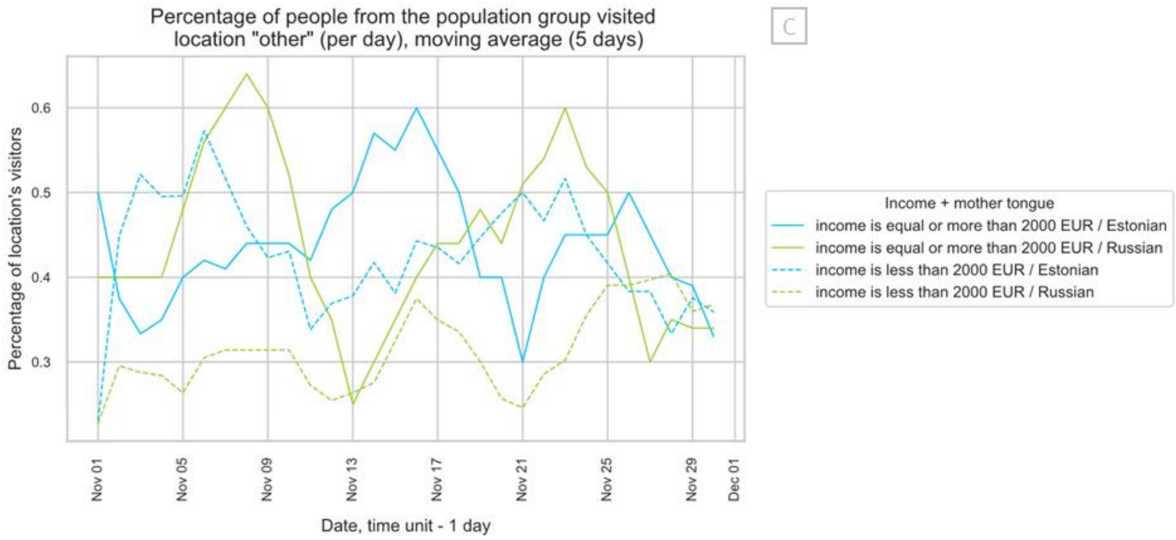


Figure 33. Share of people from the Estonian and Russian language groups combined with income levels (below 2000 EUR and equal to or more than 2000 EUR) visited: (A) – home location, (B) – work location, (C) – other location per day.

Figure 34 repeats Figure 30, where a number of hours spent at an important location was used. Service and sales staff and senior managers of an institute or company attended home least frequently, whereas mid-level specialists, middle managers, skilled workers or craftsmen and top specialists (except the last few days) visited home more often.

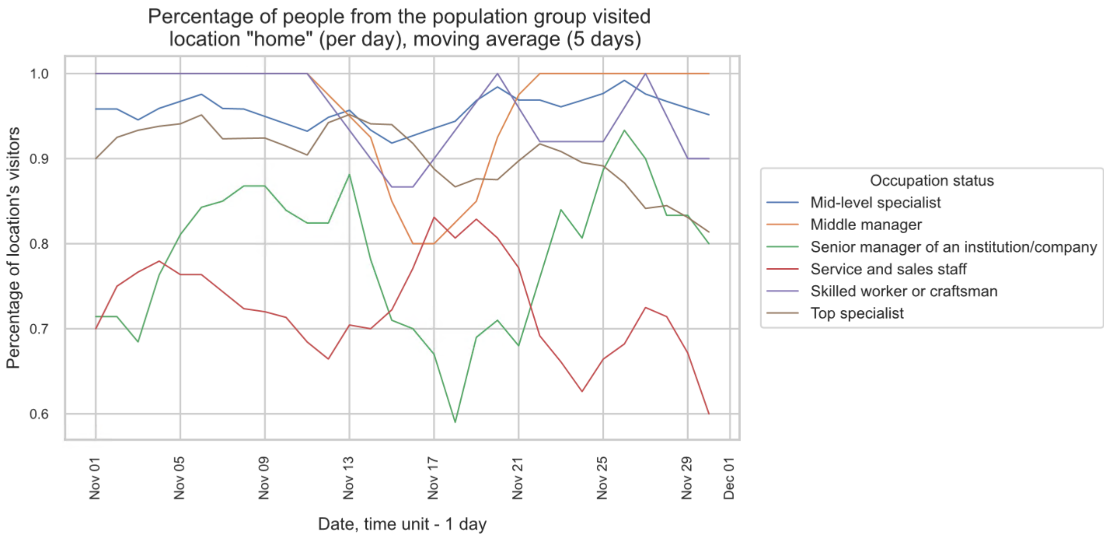


Figure 34. Percentage of people from different occupation groups visited home location per day.

The last set of plots displays time distribution showing at what time of weekday respondents from ethnic, income and occupation population groups attended their workplaces more often. The other location types and weekend data did not have such an explicit picture. Figures 35 and Figure 36 reveal differences between ethnic groups and income groups and follow almost identical trends. Russian speakers and part of the sample earning less than 2000 EUR per month visited their workplace at night and in the early morning or late evening more often than Estonian speakers and people who earned 2000 EUR per month or more, who more frequently work during standard working hours.

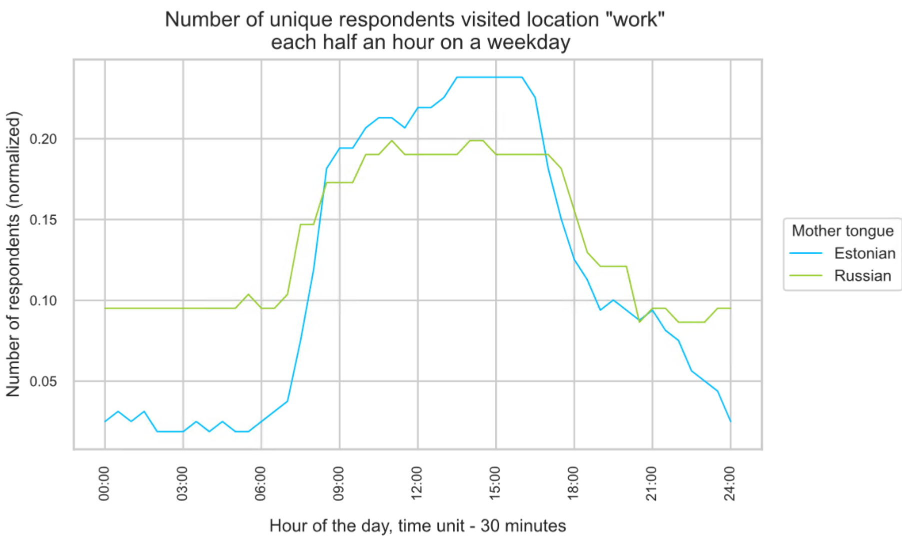


Figure 35. Distribution of respondents from the Russian-speaking and Estonian-speaking population groups visiting their work locations during weekdays.

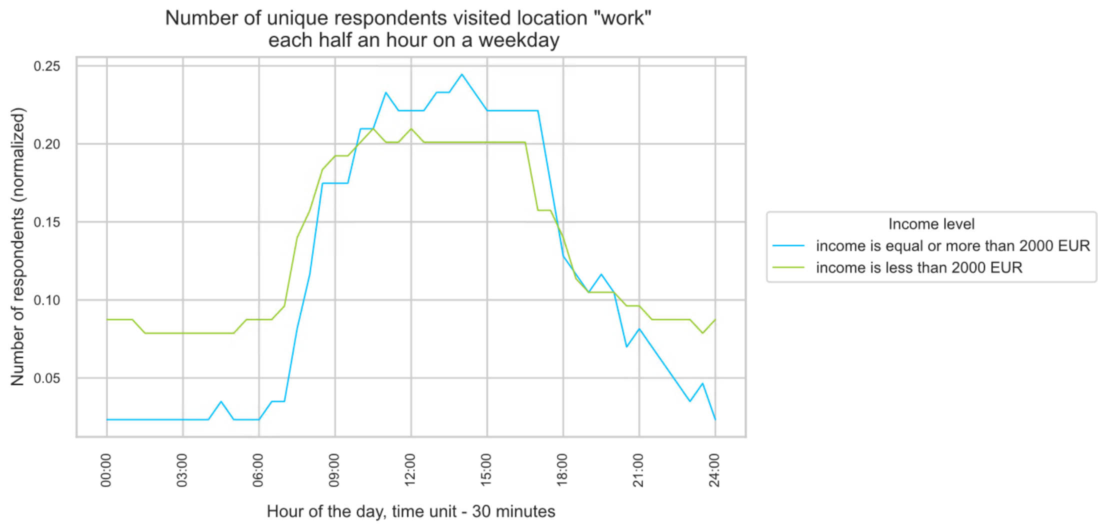


Figure 36. Distribution of respondents with income equal to or more than 2000 EUR and people whose income was less than 2000 EUR that visited their work locations during weekdays.

Figure 37 provides a more detailed look at the time of staying at the workplace on a weekday between population groups. Russian-speaking group earning less than 2000 EUR per month excels the other groups because people from this group more often have night shifts and work during non-standard working hours. Russian speakers whose salary was 2000 EUR per month or more also stood from both Estonian-speaking groups, but much less. At the same time, there was no difference within the Estonian-speaking population getting different salaries.

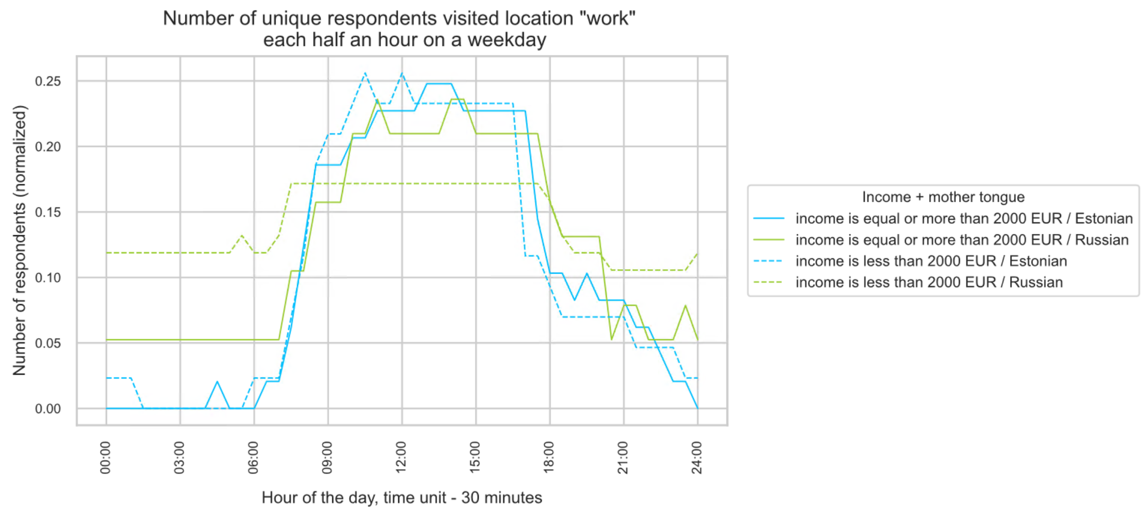


Figure 37. Distribution of respondents from Estonian and Russian language groups combined with income levels (below 2000 EUR and equal to or more than 2000 EUR) visited work location during weekdays.

Based on Figure 38, night shifts were pretty common for people employed as service and sales staff. Also, the plot shows that senior managers of an institution or company could leave their workplaces later than employees from the other occupation groups, but they also more rarely appeared at their workplaces at night or in the early morning.

Figure 38. A distribution of respondents from different occupation groups visited the work location on weekdays.

6 Discussion and conclusions

Results of the analysis revealed that there are significant differences between ethnic and income groups in the investigated study area. Correlation analysis confirmed that the considered sample had specifics described in the literature review. For instance, it was found that there was a positive moderate correlation between the Russian-speaking population and salary level below 2000 EUR per month, while Estonian speakers had the same value of correlation but with the opposite sign. Also, a positive correlation was between time spent at work and the Russian-speaking population, especially from the low-income group, while Estonian speakers tended to spend less time at their workplaces. Besides that, people preferred to live in neighbourhoods where the majority of residents spoke in their native language, which resulted in ethnic segregation.

Further analysis with the help of multiple aggregation measures allowed us to understand differences between ethnic, income and occupation groups in terms of staying or visiting important locations in more detail. In the literature overview, there has been assumed that Russian speakers and people with lower incomes who are usually employed as blue-collar workers would not be able to switch to remote work during the acute phases of the pandemic, unlike people who earn higher salaries and whose mother tongue is Estonian. Indeed, Russian-speaking respondents earning wages lower than 2000 EUR per month more frequently visited their workplaces and spent more time there. The gap between language and income groups became bigger after new COVID restrictions were imposed in the second half of November, when white-collar workers most probably started to work remotely.

Nevertheless, Russian speakers and representatives of low-income groups spent a lot of time at home even though they had fewer opportunities to work remotely. Respondents from the other groups started to visit and stay at home more only after mid-November restrictions.

Probably, the previous fact that Russian speakers and people getting lower salaries spent more time at home before new COVID restrictions compared to the other groups can be explained by lower attendance rates and less time spent at other locations because the previous research pointed out that the activity space, especially among low-income Russian speakers, was very limited to their place of residence and areas inhabited by minorities speaking in the same language. Estonian speakers and people earning 2000 EUR per month or more visited other locations much more often and stayed longer, but after restrictions, they shortened their activity.

Also, a distribution showing when representatives of each group stayed more frequently at their workplace on weekdays showed that night shifts were much more common for Russian speakers with lower salaries than for the rest of the groups. It also can be explained by the fact that this part of the population more often works as blue collars, which also more frequently implies non-standard working hours.

There were also some differences between occupation groups. Apparently, service and sales staff employees stayed at home less time and could visit it less frequently. Additionally, a larger share of such workers had non-standard working hours. The second occupation group that stood out from the others were senior managers of an institution or a company. Most probably, senior managers still had to visit their offices and they spent more time to organize work during the COVID period, thus they also had to stay longer at work.

Notable, the behaviour of Russian speakers earning 2000 EUR per month or more was quite similar to Estonian speakers. Most probably, it happened because they were employed as white collars with the possibility of remote work, and were more integrated into Estonian society, having more resources rather than respondents from the same language group getting lower salaries.

Additionally, people who potentially could work remotely as white collars demonstrated high fluctuations during the observed period. Perhaps they attempted to quickly adapt and change their lifestyle after the new COVID restrictions because previously, in the first half of November, they spent less time at home location and visited it less frequently, but they very often attended other locations and visited their workplaces sometimes even more than representatives of potential blue collars. It seriously changed in the second half of November when they started to spend more time at home and reduced visits to work and other places. At the same time, indicators of potential blue collars were much more stable, which shows that they did not change their lifestyle so much after the new COVID restrictions.

Unfortunately, due to limits in available data, there was no opportunity to analyse what effect differences in staying and visiting different kinds of locations had on each population group, primarily in case of vulnerability to being infected, which can be considered as a direction for future research.

Summary

Differences between population groups residing in Tallinn in terms of staying at important locations during the COVID-19 pandemic

Petr Anashin

Nowadays, mobility is recognised as a new separate form of capital along with economic, social and cultural capital (Flamm and Kaufmann, 2006). Mobility gives a feeling of personal freedom and promotes the formation of social status. However, differences in mobility between population groups may result in increased segregation because a lack of spatial mobility leads to fewer opportunities for social integration and interaction.

Tallinn is one of the cities which is highly economically and ethnically segregated, where differences in mobility between population groups were well studied. It has been proved that the activity space of the Russian-speaking minority was limited to locations mainly populated by Russian speakers. They include the country's capital and North-Eastern Estonia (Järv et al., 2015). On the contrary, the distribution of locations visited by Estonian speakers was more equal across the country. Also, the activity space of Russian speakers was more than twice less compared to Estonian speakers, and it did not correlate with self-perceived social class (Järv et al., 2021).

However, the COVID-19 pandemic could significantly alter the previous state of things in Estonian society. Several research reported segregation increase in locations which have already been segregated either economically or ethnically (Brough, Freedman, and Phillips, 2021; Shin, 2022). The possibility of remote work became a new factor dividing population groups because those who could not perform their work duties remotely had to visit their physical workplaces more often, less stay at home and have more daily trips, which made such people more vulnerable towards COVID-19.

In Estonia, representatives of the Russian-speaking minority are more often employed as blue collars (National Human Development Programme, 2017) with less oppostunities for remote work. Therefore, Russian speakers and people earning lower salaries, or people who potentially had fewer possibilities for remote work, could suffer more due to the consequences of COVID-19 because they could less stay at home. To check this assumption, the following research questions have been formulated:

1) What were the differences between ethnic, income and occupation groups in terms of time spent at important locations such as home, work and others, and the percentage of population group members visited these locations per day?
2) At what time of the day did population group representatives visit important locations, and what were the main differences between groups?
3) How different population groups responded to newly imposed restrictions?

The analysis has been carried out with the help of mobile positioning data collected from people living in Tallinn in two subdistricts, one of which is populated mainly by Estonian speakers and the other by Russian speakers. Also, information on each respondent from the survey was used and data on important locations of respondents related to home, work, hobbies, study, kids etc.

A dataframe containing GPS points was preprocessed, and further points were clusterised by HDBSCAN($\hat{\epsilon}$) algorithm. Then the received clusters and a list of respondents' important locations were joined. As a result, a list of GPS points belonging to specific respondents' locations was received.

The performed analysis included correlation analysis and a list of aggregated measures: the mean amount of time that representatives of the selected group spent at the particular location type per day, the share of people from the particular population group who visited selected important location per day, and relative (normalised) number of people from the selected population group who visited the particular location type on weekdays and weekends split into 30-minute time intervals.

Correlation analysis confirmed that population groups in the considered sample had the same traits as the real population living in Tallinn: Russian speakers more likely got lower salaries, unlike Estonians, and representatives of all ethnic groups preferred to live in neighbourhoods where residents spoke in their native language.

Both aggregated measures and correlation analysis showed that Russian speakers from the low-income group spent more time at work. The difference from the other population groups became larger after imposed COVID restrictions. However, Russian speakers with low incomes spent quite a lot of time at home. Estonian speakers and Russian speakers with higher incomes, on the other hand, visited locations different from home and work more often, especially before the restrictions, but afterwards, they significantly reduced their activity and started to stay more at their homes.

References

Beck, Matthew J., and David A. Hensher. 2020a. 'Insights into the Impact of COVID-19 on Household Travel and Activities in Australia – The Early Days under Restrictions'. *Transport Policy* 96 (September): 76–93. https://doi.org/10.1016/j.tranpol.2020.07.001.

———. 2020b. 'Insights into the Impact of COVID-19 on Household Travel and Activities in Australia – The Early Days under Restrictions'. *Transport Policy* 96 (September): 76–93. https://doi.org/10.1016/j.tranpol.2020.07.001.

Bhaduri, Eeshan, B.S. Manoj, Zia Wadud, Arkopal K. Goswami, and Charisma F. Choudhury. 2020. 'Modelling the Effects of COVID-19 on Travel Mode Choice Behaviour in India'. *Transportation Research Interdisciplinary Perspectives* 8 (November): 100273. https://doi.org/10.1016/j.trip.2020.100273.

Brough, Rebecca, Matthew Freedman, and David C. Phillips. 2021. 'Understanding Socioeconomic Disparities in Travel Behavior during the COVID-19 Pandemic'. *Journal of Regional Science* 61 (4): 753–74. https://doi.org/10.1111/jors.12527.

Flamm, Michael, and Vincent Kaufmann. 2006. 'Operationalising the Concept of Motility: A Qualitative Study'. *Mobilities* 1 (2): 167–89. https://doi.org/10.1080/17450100600726563.

Google. 2022. 'COVID-19 Community Mobility Reports'. Https://Www.Google.Com/Covid19/Mobility/Index.Html?Hl=en. 17 October 2022. https://www.google.com/covid19/mobility/index.html?hl=en.

Health and Welfare Information Systems Centre of Estonia. n.d. 'Estonian Covid-19 Open-Data Portal'. Accessed 26 July 2023. https://opendata.digilugu.ee/docs/#/en/readme?id=estonian-covid-19-open-data-portal.

Hu, Lingqian, Hamideh Moayyed, and Nancy Frank. 2022. 'Racial/Ethnic and Class Segregation at Workplace and Residence: A Los Angeles Case Study'. *Journal of Urban Affairs* 44 (7): 907–25. https://doi.org/10.1080/07352166.2020.1782224.

Järv, Olle, Anu Masso, Siiri Silm, and Rein Ahas. 2021. 'The Link Between Ethnic Segregation and Socio-Economic Status: An Activity Space Approach'. *Tijdschrift Voor Economische En Sociale Geografie* 112 (3): 319–35. https://doi.org/10.1111/tesg.12465.

Järv, Olle, Kerli Müürisepp, Rein Ahas, Ben Derudder, and Frank Witlox. 2015. 'Ethnic Differences in Activity Spaces as a Characteristic of Segregation: A Study Based on Mobile Phone Usage in Tallinn, Estonia'. *Urban Studies* 52 (14): 2680–98. https://doi.org/10.1177/0042098014550459.

kempen, Ronald Van, and Bart Wissink. 2014. 'Between Places and Flows: Towards a New Agenda for Neighbourhood Research in an Age of Mobility'. *Geografiska Annaler: Series B, Human Geography* 96 (2): 95–108. https://doi.org/10.1111/geob.12039.

Legeby, Ann, Daniel Koch, Fábio Duarte, Cate Heine, Tom Benson, Umberto Fugiglando, and Carlo Ratti. 2023. 'New Urban Habits in Stockholm Following COVID-19'. *Urban Studies* 60 (8): 1448–64. https://doi.org/10.1177/00420980211070677.

Malzer, Claudia, and Marcus Baum. 2020. ‘A Hybrid Approach To Hierarchical Density-Based Cluster Selection’. In *2020 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI),* 223–28. IEEE. https://doi.org/10.1109/MFI49285.2020.9235263.

Müürisepp, Kerli, Olle Järv, Feliks Sjöblom, Marina Toger, and John Östh. 2023. ‘Segregation and the Pandemic: The Dynamics of Daytime Social Diversity during COVID-19 in Greater Stockholm’. *Applied Geography* 154 (May): 102926. https://doi.org/10.1016/j.apgeog.2023.102926.

Pandas. 2023. ‘Pandas.Date_range’. 2023. https://pandas.pydata.org/docs/reference/api/pandas.date_range.html.

Programme), UNDP (United Nations Development. 2017. ‘National Human Development Report Estonia: 2016/2017’. *UNDP (United Nations Development Programme)*. https://www.2017.inimareng.ee/en.

Republic of Estonia Government. 2020a. ‘The Government Announced New Measures to Stop the Spread of the Coronavirus’. 10 November 2020. https://www.valitsus.ee/en/news/government-announced-new-measures-stop-spread-coronavirus.

———. 2020b. ‘The Government Approved the Restrictions to Prevent the Spread of the Coronavirus’. 24 November 2020. The government approved the restrictions to prevent the spread of the coronavirus.

Shin, Eun Jin. 2022. ‘Spatial Segregation of Chinese Immigrants in Seoul, South Korea, during the COVID-19 Pandemic: Evidence from Population Data Derived from Mobile Phone Signals’. *The Social Science Journal,* January, 1–22. https://doi.org/10.1080/03623319.2021.2014207.

Silm, Siiri, and Rein Ahas. 2014. ‘The Temporal Variation of Ethnic Segregation in a City: Evidence from a Mobile Phone Use Dataset’. *Social Science Research* 47 (September): 30–43. https://doi.org/10.1016/j.ssresearch.2014.03.011.

Sostero M., Milasi S., Hurley J., Fernández-Macías E., and Bisello M. 2020. ‘Teleworkability and the COVID-19 Crisis: A New Digital Divide?’ JRC121193. Seville.

Statistics Estonia. 2020. ‘Statistics Estonia: People Stay in One Location 20 Hours per Day on Average’. Https://Www.Stat.Ee/En/Uudised/News-Release-2020-047. 9 April 2020. https://www.stat.ee/en/uudised/news-release-2020-047.

Strömgren, Magnus, Tiit Tammaru, Alexander M. Danzer, Maarten van Ham, Szymon Marcińczak, Olof Stjernström, and Urban Lindgren. 2014. ‘Factors Shaping Workplace Segregation Between Natives and Immigrants’. *Demography* 51 (2): 645–71. https://doi.org/10.1007/s13524-013-0271-8.

Taeuber, Karl E., and Alma F. Taeuber. 1976. ‘A Practitioner’s Perspective on the Index of Dissimilarity’. *American Sociological Review* 41 (5): 884. https://doi.org/10.2307/2094734.

Tallinn city government. 2021. *Statistical Yearbook of Tallinn 2021*. Edited by Peeter Kuulpak. Tallinn: Tallinn Strategic Management Office.

Tammaru, Tiit. 2015. *Socio-Economic Segregation in European Capital Cities*. Routledge. https://doi.org/10.4324/9781315758879.

Tammaru, Tiit, David Knapp, Siiri Silm, Maarten Van Ham, and Frank Witlox. 2021. 'Spatial Underpinnings of Social Inequalities: A Vicious Circles of Segregation Approach'. *Social Inclusion* 9 (2): 65–76. https://doi.org/10.17645/si.v9i2.4345.

Williams, Josephine J. 1948. 'Another Commentary on So-Called Segregation Indices'. *American Sociological Review* 13 (3): 298. https://doi.org/10.2307/2086569.

Zhang, Yanji, Jiejing Wang, and Changcheng Kan. 2022. 'Temporal Variation in Activity-Space-Based Segregation: A Case Study of Beijing Using Location-Based Service Data'. *Journal of Transport Geography* 98 (January): 103239. https://doi.org/10.1016/j.jtrangeo.2021.103239.

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

I, Petr Anashin,

1. grant the University of Tartu a free permit (non-exclusive licence) to

reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright, my thesis “Differences between population groups residing in Tallinn in terms of staying at important locations during the COVID-19 pandemic” supervised by Amnir Hadachi (PhD), Siiri Silm (PhD), Serio Agriesti (MSc).

2. I grant the University of Tartu a permit to make the thesis specified in point 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 4.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.

3. I am aware of the fact that the author retains the rights specified in points 1 and 2.

4. I confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Petr Anashin

01/10/2023