

University of Tartu
Faculty of Science and Technology
Institute of Ecology and Earth Sciences
Department of Geography

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

**Differences in bike-sharing usage among varying demographic groups and its
association with station-surrounding characteristics in Tartu, Estonia**

Xiao Cai

Supervisors: Siiri Silm, PhD
Amnir Hadachi, PhD

Tartu 2024

Abstract

Differences in bike-sharing usage among varying demographic groups and its association with station-surrounding characteristics in Tartu, Estonia

The bike share system offers a wide range of benefits to promote human mobility. However, many bike share systems favor particular demographic groups (e.g., younger people and males), suggesting that the resulting benefits are not equally distributed among the public. To this end, the goal of this thesis is to empirically examine the differences in bike-sharing usage among varying demographic groups and its association with station-surrounding characteristics (i.e., land use, transport infrastructure, and population distribution) in Tartu, Estonia. For doing so, non-parametric tests (i.e., the Mann-Whitney U test and the Kruskal-Wallis H test) and the hot spot analysis were conducted to investigate the gender-based and age-based differences in bike-sharing usage; an explainable machine learning regression algorithm (i.e., gradient boosting decision trees) was used to examine the complex relationships between station-surrounding characteristics and bike-sharing usage by eight demographic groups. The findings suggest that bike-sharing usage varied among eight demographic groups in general; bike-sharing usage by varying demographic groups was differently influenced by various station-surrounding characteristics. The floor area ratio and the proportion of commercial land played the most important role in increasing bike-sharing usage. Other relative important characteristics included the proportion of public land, bike lanes, the number of foreigner residents, and population composition. Especially, it found that bike lanes were of particular concern to female users due to road safety issues; the population distribution of senior residents around bike stations played a particularly important role in increasing their use of the bike share system when compared to other age groups. These findings can provide the bike share system owner with significant insights into interventions for increasing bike-sharing usage of groups with specific demographics, so that inclusivity and equity in urban transportation could be promoted.

Keywords: bike-sharing, built environment, machine learning, non-linearity, social equity, sustainable mobility

CERCS code: S230 – Social geography

Annotatsioon

Rattaringluse kasutuse erinevused erinevate demograafiliste rühmade puhul ja seosed rattaringluse jaamade ümbruse omadustega Tartus, Eestis

Rattaringlus pakub mitmeid eeliseid inimeste liikuvuse edendamiseks. Rattaringluse süsteemid eelistavad tihti teatud demograafilisi rühmi (nt. noori ja mehi), mis viitab sellele, et sellest tulenev kasu ei ole võrdselt jaotunud. Käesoleva magistritöö eesmärk on empiiriliselt uurida erinevusi rattaringluse kasutamises erinevate demograafiliste rühmade vahel ja selle seoseid rattaringluse jaamasid ümbritseva keskkonna tunnustega (nt. maakasutus, transpordi taristu ja rahvastiku jaotus) Tartus, Eestis. Selleks on läbi viidud mitte-parametriselised testid (Mann-Whitney U-test ja Kruskal-Wallis H-test) ning kuumkoha analüüs, et uurida soolisi ja vanuselisi erinevusi rattaringluse kasutuses. Rattaringluse jaamasid ümbritseva keskkonna tunnuste ja 8 demograafilise rühma rattaringluse kasutamise kompleksete seoste uurimiseks on kasutatud selgitavat masinõppe regressioon algoritmi. Mitteparametriseliste testide tulemused näitasid soopõhiseid erinevusi rattaringluse kasutuses ning need erinevused varieerusid vanuserühmade lõikes. Kesk- ja vanemaealiste meeste ja naiste rattaringluse kasutuses tulid välja erinevused erinevates linnapiirkondades, samas kui teismeliste ja noorte täiskasvanute seas ei täheldatud ruumilist erinevust meeste ja naiste vahel. Regressioonimudelite tulemused näitasid, et erinevate demograafiliste rühmade rattaringluse kasutamist mõjutasid rattajaamasid ümbritsevad tunnused erinevalt. Liiklusohutusprobleemide tõttu valmistasid jalgrattateed eriti muret naissoost kasutajatele, samas kui eakate elanike jaotus rattajaamade ümber mängis eriti olulist rolli nende rattaringluse kasutamise suurendamise osas. Need tulemused annavad rattaringluse süsteemi rajajatele ülevaate sekkumistest, mis suurendavad teatud demograafiliste rühmade rattaringluse kasutamist, nii et oleks võimalik edendada linnatranspordi planeerimisel kaasamist ja võrdsust.

Märksõnad: rattaringlus, ehitatud keskkond, masinõpe, mittelineaarsus, sotsiaalne võrdsus, jätkusuutlik mobiilsus

CERCS Kood: S230 – Sotsiaalne geograafia

Table of contents

| | |
|--|----|
| 1. Introduction..... | 6 |
| 2. Literature review..... | 8 |
| 2.1. Factors affecting bike-sharing usage and their measurements..... | 8 |
| 2.2. Demographic differences in bike-sharing usage..... | 9 |
| 2.3. Approaches to examining associations of station-surrounding characteristics with bike-sharing usage..... | 10 |
| 3. Data and variables..... | 12 |
| 3.1. Study area..... | 12 |
| 3.2. Data sources and pre-processing..... | 13 |
| 3.2.1. Bike-sharing usage data..... | 13 |
| 3.2.2. Built environment data..... | 15 |
| 3.2.3. Demographic data..... | 18 |
| 3.3. Variables..... | 20 |
| 4. Research methodology..... | 26 |
| 4.1. Non-parametric tests..... | 26 |
| 4.2. Hot spot analysis..... | 26 |
| 4.3. Regression analysis: Gradient boosting decision trees..... | 27 |
| 5. Results..... | 29 |
| 5.1. Differences in bike-sharing usage among varying demographic groups..... | 29 |
| 5.2. Relationships between various station-surrounding characteristics and bike-sharing usage of varying demographic groups..... | 32 |
| 5.2.1. The impact of land use on bike-sharing usage..... | 34 |
| 5.2.2. The impact of transport infrastructure on bike-sharing usage..... | 36 |
| 5.2.3. The impact of population distribution on bike-sharing usage..... | 37 |
| 6. Discussion..... | 39 |

| | |
|-----------------------|----|
| 7. Conclusion..... | 41 |
| Summary..... | 43 |
| Kokkuvõte..... | 45 |
| Acknowledgements..... | 47 |
| References..... | 48 |

1. Introduction

Existing shared mobility practices have generated tremendous benefits for addressing major challenges faced by urban transportation systems, such as air pollution, traffic congestion, and limited accessibility (Machado et al., 2018; Sun et al., 2021). Compared to applications such as ride-hailing and scooters programs, the bike share system stands out by its affordability and positive effects on users' physical health (Cheng et al., 2022; Mizdrak et al., 2023; Teixeira et al., 2021). However, it is widely reported that the bike share system was in favor of particular demographic characteristics (Fishman, 2016). For instance, Goodman and Cheshire (2014) found that the number of trips made by female users merely accounted for less than one fifth of total bike-sharing trips in London. Such an usage imbalance issue was also observed in an Irish bike share system – the number of registered male bike-sharing users were about four times higher than that of the female counterpart (Murphy & Usher, 2015). Besides, Hossain et al. (2023) found that a larger number of bike-sharing trips were generated by population with younger age profiles, while the proportion of seniors was negatively related to bike-sharing usage as indicated by Ma et al. (2020).

These reported statistics suggest that the resulting benefits from the bike share system are not equally distributed among the public (Ricci, 2015; Woodcock et al., 2014). To construct more equally utilized bike share systems, it is important to understand factors affecting bike-sharing usage of varying demographic groups, especially for those under-represented. For doing so, existing studies have extensively investigated the associations of bike-sharing usage with various built and demographic environments (Eren & Uz, 2020; Guo et al., 2022; Sun et al., 2021; Zhang & Kamargianni, 2023; Zhu et al., 2022). However, few studies took the demographics of bike-sharing users (e.g., age and gender) into account because most bike share system holders do not have such information in their datasets. To this end, it remains under-explored in terms of how the bike-sharing usage of varying demographic groups is influenced by various characteristics. Furthermore, earlier empirical studies were mostly based on the traditional regression models (El-Assi et al., 2017; Faghih-Imani et al., 2017; Lee et al., 2022; Meng & Zacharias, 2021). These approaches fail to disclose the relative importance of various explanatory variables of bike-sharing

usage. Therefore, policy implications gained from these studies could be limited for informing decision making of relevant stakeholders.

Given the above, this study aims to empirically examine the associations of various station-surrounding characteristics with demographic-specific bike-sharing usage in Tartu (Estonia) by an explainable machine learning approach (i.e., the gradient boosting decision tree regression algorithm). Specifically, the following two research questions were answered by this study:

- **Is there a difference in bike-sharing usage by various demographic groups, and what is the difference (especially from the spatial perspective)?**
- **How do various station-surrounding characteristics (i.e., land use, transport infrastructure, and population distribution) influence bike-sharing usage of different demographic groups?**

To sum up, this study contributes to existing literature mainly in three ways. First, previous studies seldom examined the disparities in effects of various station-surrounding characteristics on demographic-specific bike-sharing usage due to data unavailability. This study filled such a knowledge gap by a large-scale bike-sharing dataset containing the information of bike-sharing users' gender and age. Second, it remains unclear in the relationship between station-surrounding characteristics and bike-sharing usage, especially regarding non-linearity and relative importance of explanatory variables. This was handled by this study using an explainable machine learning regression algorithm, and therefore a more comprehensive understanding of their associations was informed. Third, few scholars considered the distance decay effect when quantifying explanatory variables of bike-sharing usage. It could overestimate the effect of the station-surrounding characteristics that are relatively far from the location of bike-sharing trips. To address this problem, this study applied the Gaussian function to generate a series of coefficients to weigh various station-surrounding characteristics based on their distance to bike stations. By doing so, the estimates of the effect of station-surrounding characteristics on bike-sharing usage could be more realistic.

2. Literature review

2.1. Factors affecting bike-sharing usage and their measurements

A large number of studies have empirically examined the effects of various built and demographic environments on bike-sharing usage (Eren & Uz, 2020; Guo et al., 2022; Sun et al., 2021; Zhang & Kamargianni, 2023; Zhu et al., 2022). The influencing factors can be classified into three categories, i.e., 1) land use, 2) transport infrastructure, and 3) population distribution. First of all, bike-sharing usage could vary across urban areas due to the spatial distribution of land uses. This is because that cyclists' travel behavior is closely related to the diverse use of lands where bike-sharing trips occur (Ross-Perez et al., 2022; R. X. Wang et al., 2022; Xing et al., 2020). For instance, individuals riding to commercial areas are most likely to go there for shopping, while those riding to urban green spaces are largely for recreational purposes. To this end, earlier studies quantified the land use factor by calculating proportions of diverse land uses as well as land use mix (He et al., 2022; Lee et al., 2022; Li et al., 2020). Some studies also took the floor area ratio into account since it indicates the land development intensity (Meng & Zacharias, 2021; Tu et al., 2019).

Transport infrastructure is another vital factor influencing bike-sharing usage (Kim, 2023; Wu et al., 2021; Xu et al., 2019). Road networks (especially the bike lane system) enable individuals to cycle from one location to another. Bike stations located in areas with higher road connectivity are more likely to be frequently used (Yang et al., 2020). Therefore, most existing studies qualified this factor by the road density (Gao et al., 2022; Gao et al., 2021) or the road intersection density (Hossain et al., 2023; Yang et al., 2020). Other public transits such as bus/metro stops are also important determinants for shaping bike-sharing usage because the bike share system serves as the first-and-last mile solution in many cities. Existing studies generally quantified this variable by the bus/metro density (El-Assi et al., 2017; Gao et al., 2023; Ma et al., 2020; Xiao & Wei, 2023).

Bike-sharing usage could be also affected by the station-surrounding population distribution (Schimohr et al., 2023). For instance, Feng and Li (2016) found that gender was the most important factor affecting people's willingness to use the bike share system, compared to other variables; Hossain et al. (2023) discovered that areas with younger age profiles were likely to generate a larger number of bike-sharing trips. Therefore, these variables were typically quantified by the density/proportion of

varying demographic groups (Bao et al., 2018; Gao et al., 2022; Hossain et al., 2023; Ma et al., 2020).

Regarding the measurement of variables, it is widely recognized that the use of share bikes could be largely influenced by the surrounding characteristics of where bike sharing trips take place (Fishman, 2016). Therefore, all of the above variables need to be measured based on a specific bike-sharing catchment area. However, few studies have taken into account the distance decay effect when measuring variables; instead, they have treated all characteristics within the catchment area as the same, no matter how far they are from the location of bike-sharing trips. This could lead to problematic estimations, as the characteristic that is closer to the bike station is likely to have a greater impact on bike-sharing trip generation and therefore should be weighted more heavily. Given this, it is required to optimize the existing approach to measuring explanatory variables of bike-sharing usage in the future studies.

2.2. Demographic differences in bike-sharing usage

The demographics of bike-sharing users serve as one of the significant factors affecting their travel behavior (Eren & Uz, 2020; Wang et al., 2018). It mainly involved in two aspects, i.e., gender and age. To begin with, the difference in bike-sharing usage between genders might be due to various factors. Road safety is one of main concerns – it is widely reported that females were more sensitive to road safety when cycling on the road (AitBihiOuali & Klingen, 2022; Akar et al., 2013; Le et al., 2019; Pearson et al., 2023). The bike-sharing usage of different genders also varies in frequencies and travel distances – as reported by (Goodman & Cheshire, 2014; Murphy & Usher, 2015), male users generated longer and a larger number of bike-sharing trips than female users. In addition, such a difference could be caused by unequal access to bike stations by gender (Fortes et al., 2024; Hail & McQuaid, 2021; Jin & Sui, 2024).

Besides, generational differences in bike-sharing behavior may exist because of two reasons: age effects (Wang et al., 2018) and cohort effects (de Haas et al., 2022). Age effects indicate that different age groups may have different needs when using the bike share system. For instance, seniors typically have higher requirements on road aesthetics when compared to younger people, as they cycle mostly for recreational

and exercise purposes (Böcker et al., 2020; Fishman et al., 2015; Kim & Cho, 2023). Cohort effects could be another important indicator determining demographic differences in travel behavior (Heinen et al., 2018). To illustrate, individual travel behavior can be largely affected by its current socio-economic and techno-cultural contexts. For example, younger generations typically drive less and use multiple other modes of transport (e.g., public transport and share bikes) than previous generations due to the influence of lifestyles and adoption of information and communication technologies (de Haas et al., 2022; Wang et al., 2018).

In recent years, a few bike share system owners have collected the gender and age of users and included this information into their datasets. This allows for comparative analyses of bike-sharing usage by different gender and age groups (Böcker et al., 2020; Kim & Cho, 2023). However, existing studies only focus on comparisons between gender groups or age groups; few studies combined both age and gender to gain a more comprehensive understanding of bike-sharing usage by varying demographic groups. Therefore, the difference in bike-sharing usage by different gender and age groups remains under-explored.

2.3. Approaches to examining associations of station-surrounding characteristics with bike-sharing usage

Scholars have been dedicated to measuring associations of various factors with bike-sharing usage by multiple regression models. First of all, traditional non-spatial regression models have been applied to measure their linear relationships in earlier studies (El-Assi et al., 2017; Faghieh-Imani et al., 2017; Lee et al., 2022; Meng & Zacharias, 2021). Afterwards, some studies considering spatial heterogeneity emerged, holding that bike-sharing usage can be affected by various characteristics in a different way in different regions across the city (Bao et al., 2018; Gao et al., 2022; Hossain et al., 2023; Li et al., 2020; Ma et al., 2020). Nevertheless, such spatial phenomena may not be observed from small-scale bike-sharing systems with a limited number of bike stations (Yang et al., 2020).

According to these studies, it enables to be aware of how various characteristics affect bike-sharing usage in a linear way. However, these traditional approaches fail to compare the relative importance of station-surrounding characteristics for affecting

bike-sharing usage. For instance, several studies empirically verified the positive association of both commercial and residential facilities with the number of bike-sharing trips (Lee et al., 2022); however, which characteristic is more important than another for boosting bike-sharing usage? Understanding such differences can provide bike share system owners with justifications for finding optimal locations for installing bike stations. More importantly, the relationships between various factors and active travel behaviors are likely to be more sophisticated than linearity (Chen & Ye, 2021; Ji et al., 2023; Y. C. Wang et al., 2022; Xiao & Wei, 2023). To this end, an increasing number of studies have attempted machine learning techniques to fill this knowledge gap in the most recent years (Du et al., 2019; Gao et al., 2023; Lin et al., 2020; Schimohr et al., 2023; Zhuang et al., 2022).

These emerging studies provide an overall understanding of how various built and demographic environments affect bike-sharing usage by measuring trip records of all users as an integrated bike-sharing demand. So far, however, merely limited attention has been paid to demographic-specific bike-sharing usage - residents of different ages or genders may react differently to a specific characteristic, and thus differentiated bike-sharing usage may occur. Therefore, it remains unclear regarding discrepancies in the effects of various station-surrounding characteristics on bike-sharing usage among varying demographic groups.

3. Data and variables

3.1. Study area

Bike share systems can generate tremendous benefits in building a sustainable and healthy urban environment (Cheng et al., 2022; Teixeira et al., 2021), thus having gained wide attention from relevant stakeholders. It reported a rapid increase in new bike share system deployments from 20 in 2008 to 337 in 2018 - Europe has the highest number of new bike share systems compared with North America and Asia (Galatoulas et al., 2020). Therefore, more studies should be conducted to explore how these newly built European bike share systems promote urban mobility.

Given this, this study selected the Tartu bike share system as a case study, and the Tartu urban area was chosen as the study area, occupying an area of 38.82 square kilometers (Figure 1). Tartu is the second largest city in Estonia, with about one hundred thousand population residing in urban areas until 2023, according to Statistics Estonia. The local government established the bike share system in June 2019 and till now it has been a vital component of the local public transportation network. The bike share system in Tartu consists of 750 bikes and 100 bike stations. However, this study only took bike stations within the Tartu urban area (n=80) into account because it is where the share bikes are intensively used.

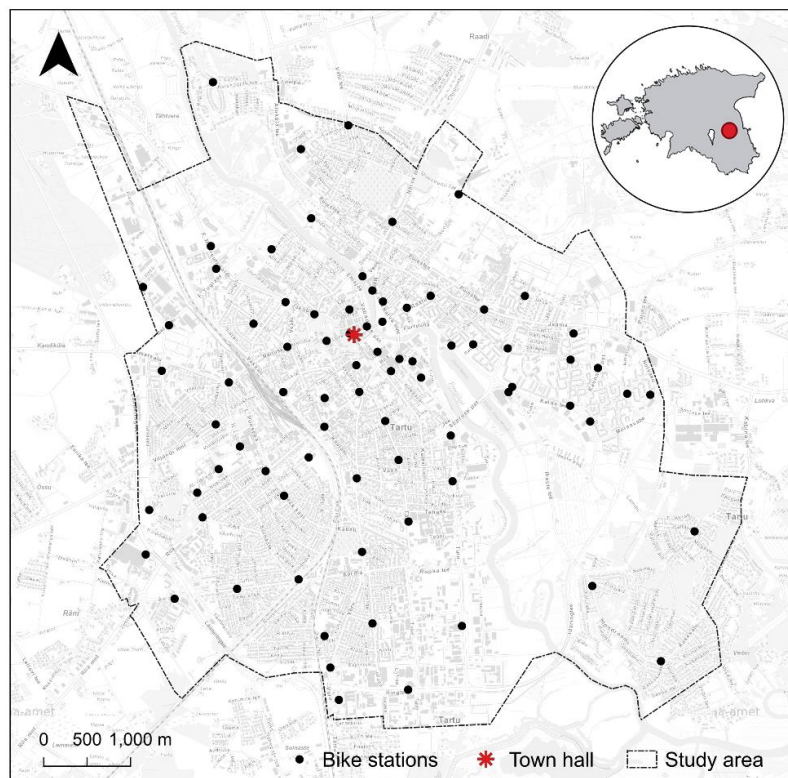


Figure 1. The study area and the bike share system.

3.2. Data sources and pre-processing

3.2.1. Bike-sharing data

Bike-sharing data was obtained from the Tartu city government, consisting of two parts. The first part is the geographic location (i.e., latitude and longitude) of bike stations, and the second part contains two-month bike-sharing trip records (n=250,897) from the first day of May to the last day of June in 2021. This study selected trips in May and June because the bike share system in Tartu had the most frequent usage during this period throughout the year. Each trip record contains two types of features: 1) route-related features: route code, unlock date and time, lock date and time, name of start and end stations; 2) user-related features: user identification (ID), membership types, and the first three digits of users' personal ID code (for inferring users' age and gender)¹. Some examples of bike-sharing trip records are presented in Table 1.

Table 1. Examples of bike-sharing trip records.

| Route code | User ID | Unlock date | Unlock time | Lock date | Lock time | Start station | End station | Membership | Personal ID code |
|------------|---------|-------------|-------------|------------|-----------|-----------------|-------------|---------------------------------|------------------|
| 161xxx | 48fxxx | 2021-05-01 | 23:58:51 | 2021-05-02 | 00:05:40 | Ülejõe park | Ujula | Free membership (with bus card) | 502 |
| 162xxx | 2cbxxx | 2021-05-16 | 00:00:30 | 2021-05-16 | 00:06:12 | K. E. von Baeri | Tulbi | One-year membership | 505 |
| 162xxx | 83axxx | 2021-06-26 | 07:46:18 | 2021-06-26 | 07:49:19 | E-Kaubamaja | Kuu | Free membership (with bus card) | 493 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |

¹ The first three digits of personal ID codes in Estonia contain the information related to ID holders' age and gender. The first digit could be from 1 to 6 – if it is odd, it refers to males, otherwise it refers to females. Furthermore, if it is 1 or 2, it indicates that the ID holder was born in the 19th century; if it is 3 or 4, it indicates that the ID holder was born in the 20th century; if it is 5 or 6, it indicates that the ID holder was born in the 21st century. The second and third digits refer to the year in the corresponding century. Taking 502 as an example, according the first digit “5”, it can first infer that the ID holder is male and was born in the 21st century; combined with the second and third digits “02”, it can further infer that the ID holder was born in 2002. Given that the trip occurred in 2021, it can finally infer that this is a 19-year-old male user when using the bike share system.

A few steps were conducted to clean the data. This study first removed bike-sharing trip records without personal ID code information (n=53,840) because the gender and age of generators of these trips cannot be acquired for further analyses. Then, the bike-sharing trip records with their start stations or end stations located outside of the Tartu urban area (i.e., the study area) (n=12,931) were removed since this study aims to measure local residents' urban mobility via the bike share system, and therefore these excluded trips (occurring in neighboring counties rather than urban areas) were out of the research scope. Plus, the Tartu urban area contains a majority of bike stations (80 out of 100 stations) and is where the share bikes are predominantly used (accounting for around 93.44% of total trips). Furthermore, trip records generated by short-term membership² holders (n=2,553) were removed as these trips were largely generated by visitors or tourists rather than local residents. Lastly, this study excluded the bike-sharing trip records with the same start and end stations as well as their duration less than two minutes (n=4,037) because these trip records occurred largely due to users' operation mistakes, but they did not actually use the bike.

After the above-mentioned steps, a total of 177,536 trips records were eligible for the following analysis (occupying around 70.76% of total trip records). These trips were generated by 6,325 unique users, with an average of about 28 trips per user. Among these trips, there are 3,074 male users generating 56.97% of total trips (n=101,146) and 3,251 female users generating remaining trips (n=76,390), occupying 6.97% and 6.50% of total male and female residents respectively in the study area by 2021 according to Statistics Estonia. This indicates that male residents were more likely to be the bike-sharing users and used share bikes much more frequently than female counterparts. Besides, the age of bike-sharing users ranges from 8 to 91; starting from 14 years old, bike-sharing usage significantly surged, especially from users aged 15 to 19 (Figure 2).

² There are six types of membership for using the bike share system in Tartu, i.e., 1) free membership (with bus card), 2) one-year membership, 3) one-month membership, 4) one-week membership, 5) one-day membership, and 6) one-hour membership. This study regards the last four types as the short-term membership.

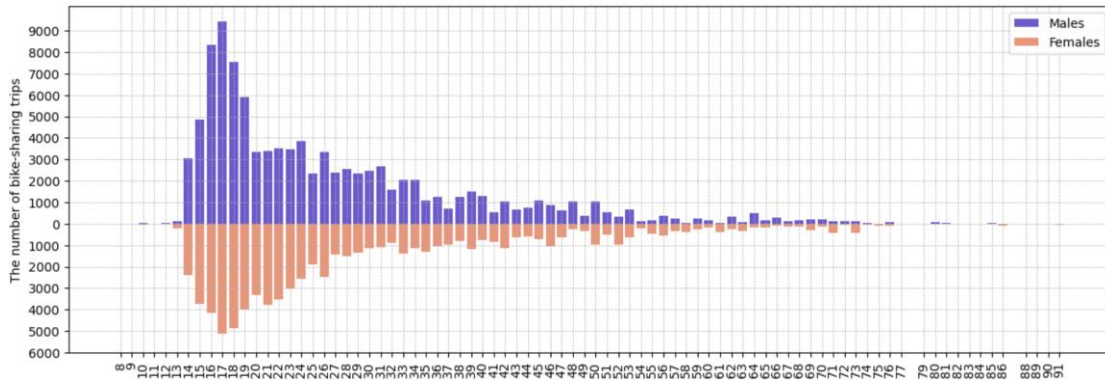


Figure 2. Distribution of bike-sharing trips by gender and age.

Furthermore, these trips were grouped based on the gender and age of bike-sharing users. Specifically, this study considered two gender types (i.e., male and female) and four age range types, including 1) 13-19 years old (i.e., representatives of teenagers), 2) 20-34 years old (i.e., representatives of young adults), 3) 35-64 years old (i.e., representatives of middle-aged adults), and 4) 65 years old and above (i.e., representatives of seniors). As a result, eight demographic groups were obtained for further analyses, and their detailed bike-sharing usage is presented in Table 2.

Table 2. Description of bike-sharing usage by eight demographic groups.

| Gender | Age | Representative | Total trips | Number of unique users | Trips by user |
|--------|-------|---------------------------|-------------|------------------------|---------------|
| Male | 13-19 | Male teenagers | 39,183 | 957 | 41 |
| | 20-34 | Male young adults | 41,389 | 1,370 | 30 |
| | 35-64 | Male middle-aged adults | 18,884 | 676 | 28 |
| | >= 65 | Male seniors | 1,627 | 66 | 25 |
| Female | 13-19 | Female teenagers | 24,535 | 898 | 27 |
| | 20-34 | Female young adults | 30,546 | 1,494 | 20 |
| | 35-64 | Female middle-aged adults | 18,945 | 767 | 25 |
| | >= 65 | Female seniors | 2,282 | 87 | 26 |

3.2.2. Built environment data

Four built environment datasets were included in this study, i.e., 1) land use, 2) buildings, 3) bus stops, and 4) bike lanes. The first three datasets were acquired from the Tartu city government, while the last dataset was obtained from OpenStreetMap. In terms of data formats, land use and building data were stored in polygon formats; bus stop and bike lane data were stored in point and linestring formats respectively. Land use data contains various types of land parcels, i.e., commercial, residential, institutional, public land, transportation, industrial, and other lands. Building data contains attributes of buildings' boundary and height, used to calculate the floor area

ratio around bike stations. Besides, 134 bus stops and about 65-kilometer bike lanes around bike stations consist of bus stop and bike lane datasets respectively.

Noticeably, it is a challenge to measure the distance of built environments to bike stations when these data are not stored in point formats, such as linestring or polygon formats. A common way for doing so is based on the centroid of the linestring or the polygon data to identify the location of built environments. However, it could be problematic when there is a large variation in the polygon size or the linestring length. For instance, a large land parcel or a long bike lane could have some parts very close to bike stations, serving as a critical factor affecting bike-sharing users' travel behaviors; however, it may be excluded from evaluations because its centroid is beyond the maximum distance threshold. To handle this problem, this study partitioned all land parcels into identical fine-scale grids (30m*30m) as applied by Shi et al. (2020), and the grid centroid was used to locate a specific land use type for calculating its distance to the bike station (Figure 3a). The same way was applied to bike lanes – every bike-lane was converted into a series of points with an identical distance interval of 30 meters (Figure 3b). Additionally, the centroid of building boundaries was directly applied to locate the buildings given that they were stored already in fine-scale polygon formats (Figure 3b). After these pre-processing steps, all built environment features around bike stations were stored in the point format (Figure 3) so that their distance decay effects on bike-sharing usage could be handled in the following analysis (specified in Section 3.3).

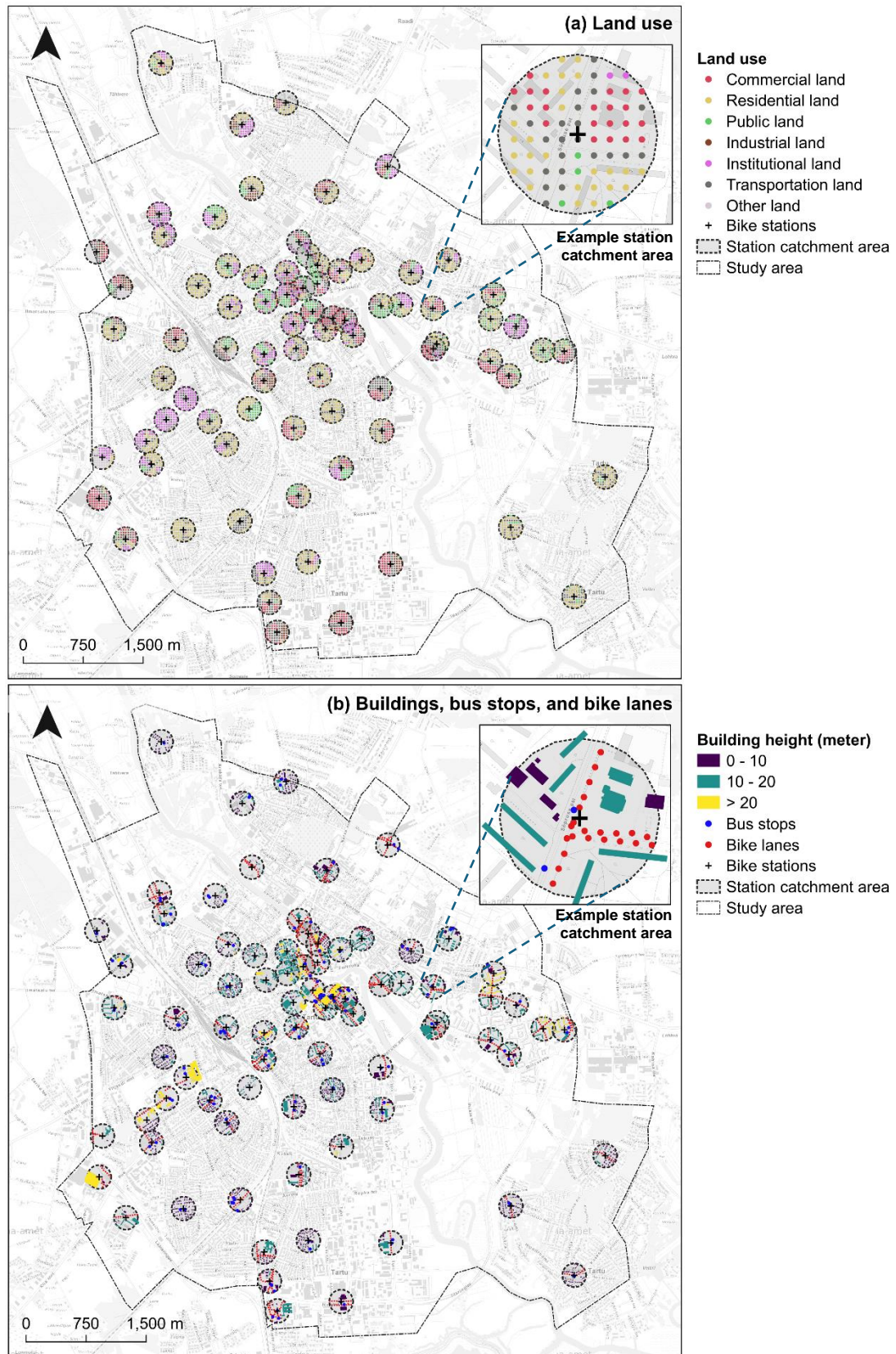


Figure 3. Spatial distribution of built environments (i.e., land use, buildings, bus stops, and bike lanes) within the catchment area of bike stations.

3.2.3. Demographic data

Demographic data was acquired from Statistics Estonia, consisting of two attributes. The first attribute is the ethnicity³, i.e., Estonian, Russian, and foreigners. The spatial distribution of three ethnicities is presented in Figure 4. Another attribute contains residents' gender and age, based on which eight demographic groups (as mentioned earlier) were classified. The spatial distribution of eight groups in the study area is presented in Figure 5. Both attributes are stored at the grid of 100m by 100m. The centroid of these grids was generated for calculating their distance to the nearby bike stations. The census moment for both datasets is by 31 December 2021; thus, the data has a consistent temporal range with the bike-sharing data.

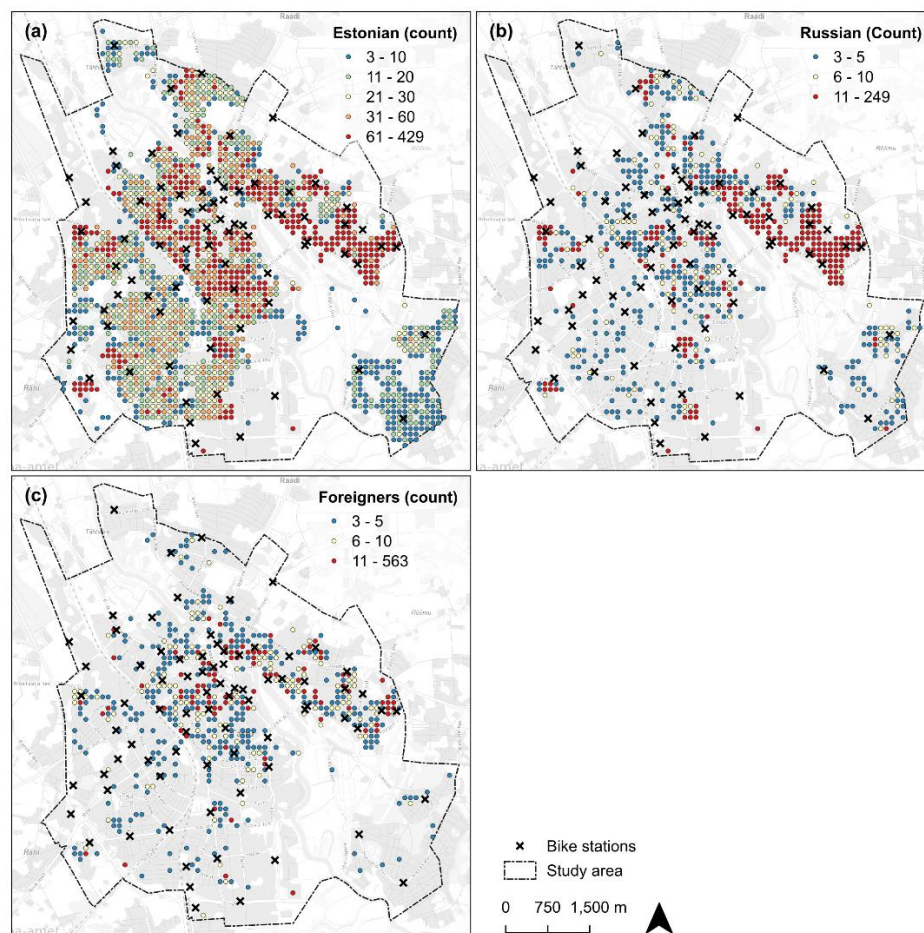


Figure 4. Spatial ethnicity distribution: (a) Estonian, (b) Russian, and (c) Foreigners.

³ The ethnicity of residents is classified based on their native languages. Estonian refer to residents who speak Estonian as their native language; Russian refer to residents who speak Russian as their native language; Foreigners refer to residents who do not speak Estonian or Russian as their native language.

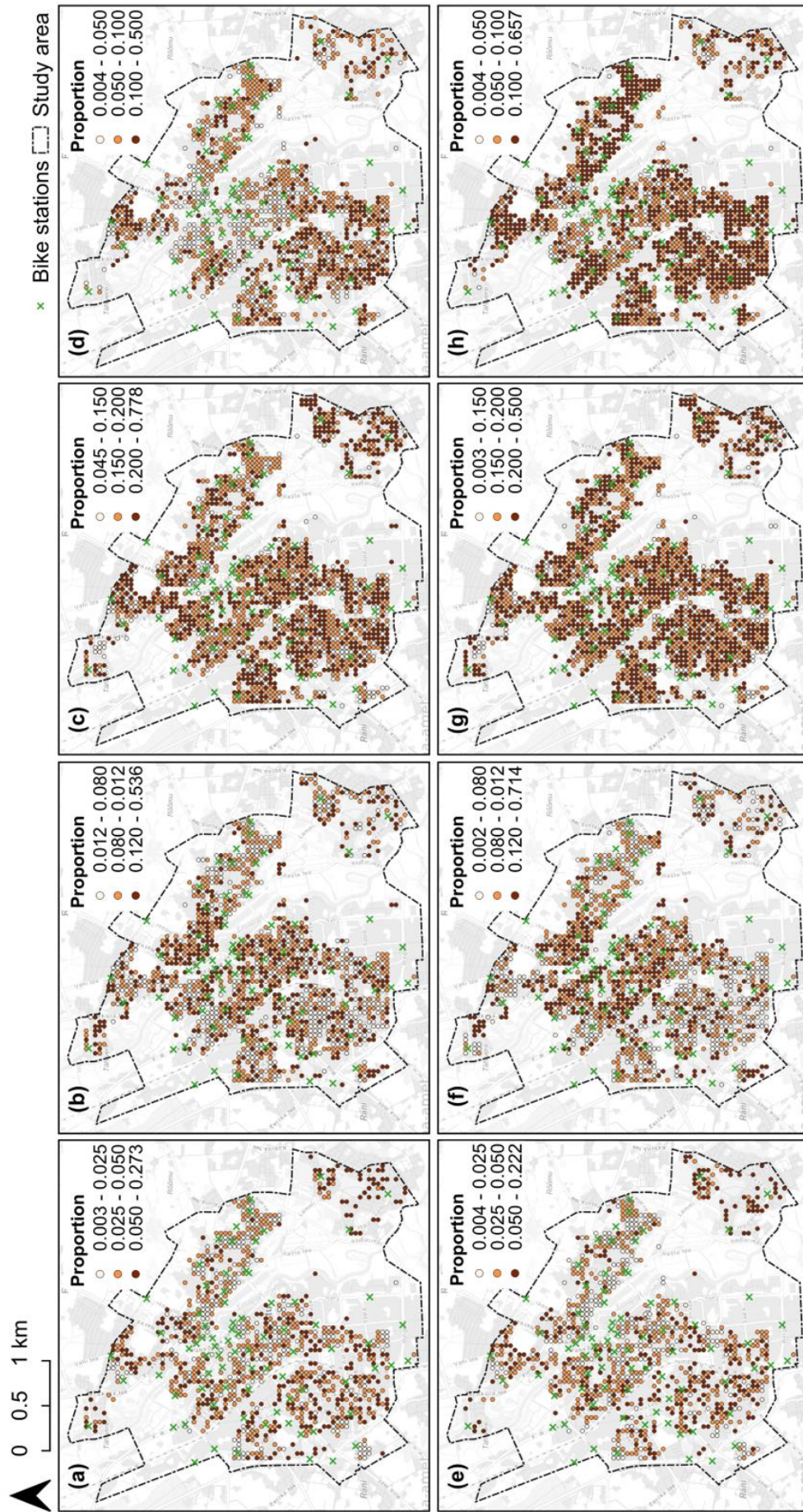


Figure 5. Spatial distribution of eight demographic groups: (a) male teenagers, (b) male young adults, (c) male middle-aged adults, (d) male seniors, (e) female teenagers, (f) female young adults, (g) female middle-aged adults, and (h) female seniors.

3.3. Variables

One goal of this study is to examine how station-surrounding environments affect bike-sharing usage of different demographic groups. Given it, the station-level average daily trips by each demographic group were calculated as the dependent variable. The descriptive statistics are presented in Table 3. It found that teenagers and young adults dominated bike-sharing usage, where male users generated evidently more trips than female users. However, bike-sharing usage by female middle-aged adults and seniors slightly surpassed that by their male counterparts. It also observed that the senior users were underrepresented, whose daily average trips were far less than one trip per bike station.

Table 3. Descriptive statistics of the dependent variable (i.e., the station-level average daily trips by each demographic groups).

| Gender | Age | Representative | Avg. | Std. | Min. | Med. | Max. |
|---------------|------------|---------------------------|-------------|-------------|-------------|-------------|-------------|
| Male | 13-19 | Male teenagers | 8.03 | 5.69 | 0.18 | 6.67 | 34.28 |
| | 20-34 | Male young adults | 8.48 | 5.81 | 0.34 | 7.80 | 33.64 |
| | 35-64 | Male middle-aged adults | 3.87 | 2.72 | 0.03 | 3.59 | 16.26 |
| | >= 65 | Male seniors | 0.33 | 0.34 | 0.00 | 0.20 | 1.87 |
| Female | 13-19 | Female teenagers | 5.03 | 3.98 | 0.05 | 3.94 | 22.49 |
| | 20-34 | Female young adults | 6.26 | 4.48 | 0.03 | 5.67 | 27.30 |
| | 35-64 | Female middle-aged adults | 3.88 | 2.79 | 0.05 | 3.43 | 16.75 |
| | >= 65 | Female seniors | 0.47 | 0.51 | 0.00 | 0.28 | 2.90 |

To understand what influences bike-sharing usage, this study selected a series of explanatory variables through a comprehensive literature review (Eren & Uz, 2020; Guo et al., 2022; Sun et al., 2021; Zhang & Kamargianni, 2023; Zhu et al., 2022). These variables are further classified into three categories, i.e., 1) land use, 2) transport infrastructure, and 3) population distribution. First of all, land use influences bike-sharing users' travel behaviors by providing various services (e.g., recreation, shopping, dining, and others) from different types of land use. This would attract people to visit and then leave from where they visit. The transport infrastructure is also closely related to people's usage of the bike share system. For instance, road networks enable people to access their destinations; bike lanes ensure cyclists' safety; other transport amenities (e.g., bus/metro stops) provide people with convenience for transferring. Besides, share bikes are highly likely to be used by people residing around them, especially by those without their own transport vehicles. Therefore, the

population size and demographics around bike stations are non-ignorable for understanding determinants of bike-sharing usage.

Given the above, this study selected a total of fifteen variables falling into aforementioned three categories for explaining bike-sharing usage by varying demographic groups. Specifically, it includes 1) **land use**: the proportions of residential land, commercial land, institutional land, industrial land, and public land, as well as land use mix and the floor area ratio; 2) **transport infrastructure**: proximity to roads, bike lanes, bus stops, and downtown; and 3) **population distribution**: the number of Estonian, Russian, and foreigner residents, as well as the proportion of residents having the same gender and age range with the bike-sharing users of the specific demographic group.

Before regression analyses, it is important to test if there is multicollinearity among the selected variables. Therefore, this study calculated the variance inflation factor (VIF) for each variable among eight models (specified in Section 4.3). It found that the four variables (i.e., proximity to roads, land use mix, the number of Estonian residents, and proximity to downtown) had their VIFs above 10, indicating a high possibility of collinearity among these variables with others. As a result, this study removed these variables and kept the remaining variables for further regression analyses. Consequently, VIFs of all eleven variables in eight models became below 10 (Table 4), indicating that no evident multicollinearity issue was observed among these variables. The description of eleven explanatory variables is presented in Table 5.

Table 4. VIFs of variables in eight models in this study.

| Variables | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 | Model 7 | Model 8 |
|------------------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| Residential land | 3.273 | 4.161 | 5.797 | 4.098 | 3.213 | 3.656 | 5.689 | 4.753 |
| Commercial land | 3.858 | 3.791 | 3.997 | 3.768 | 3.761 | 3.831 | 3.942 | 3.832 |
| Institutional land | 1.936 | 1.902 | 1.961 | 1.907 | 1.897 | 1.912 | 1.985 | 1.906 |
| Industrial land | 1.292 | 1.293 | 1.283 | 1.287 | 1.360 | 1.284 | 1.262 | 1.269 |
| Public land | 1.639 | 1.784 | 1.93 | 1.568 | 1.738 | 1.753 | 1.902 | 1.600 |
| Floor area ratio | 4.072 | 4.099 | 4.051 | 4.063 | 4.054 | 4.089 | 4.046 | 4.054 |
| Bike lanes | 2.478 | 2.778 | 2.486 | 2.561 | 2.514 | 2.554 | 2.526 | 2.586 |
| Bus stops | 2.822 | 2.771 | 2.801 | 2.783 | 2.824 | 2.771 | 2.773 | 2.797 |
| Russian residents | 1.434 | 1.536 | 1.433 | 1.471 | 1.463 | 1.544 | 1.434 | 1.550 |
| Foreigner residents | 1.163 | 2.671 | 1.272 | 1.257 | 1.145 | 2.086 | 1.310 | 1.298 |
| Population composition | 2.677 | 8.185 | 7.738 | 3.859 | 2.409 | 5.746 | 7.363 | 5.214 |

Table 5. Description of explanatory variables.

| Variable | Description | Avg. | Std. | Min. | Med. | Max. |
|---------------------------------------|---|-------------|-------------|-------------|-------------|-------------|
| Land use (n=6) | | | | | | |
| Residential land | The proportion of weighted residential land | 0.283 | 0.218 | 0.000 | 0.280 | 0.749 |
| Commercial land | The proportion of weighted commercial land | 0.140 | 0.164 | 0.000 | 0.076 | 0.771 |
| Institutional land | The proportion of weighted general institutional land (e.g., schools, hospitals, governments, and others) | 0.139 | 0.189 | 0.000 | 0.039 | 0.789 |
| Industrial land | The proportion of weighted industrial land | 0.033 | 0.078 | 0.000 | 0.000 | 0.346 |
| Public land | The proportion of weighted public land (e.g., urban green space, playgrounds, and others) | 0.099 | 0.133 | 0.000 | 0.054 | 0.490 |
| Floor area ratio | The weighted floor area ratio | 0.942 | 0.818 | 0.018 | 0.681 | 4.826 |
| Transport Infrastructure (n=2) | | | | | | |
| Bike lanes | The number of weighted bike lane points | 7.565 | 6.526 | 0.000 | 6.372 | 27.18 |
| Bus stops | The number of weighted bus stops | 1.120 | 0.949 | 0.000 | 1.169 | 4.381 |
| Population distribution (n=3) | | | | | | |
| Russian residents | The number of weighted residents whose native language is Russian | 32 | 76 | 0 | 4 | 392 |
| Foreigner residents | The number of weighted residents whose native language is neither Russian nor Estonian | 20 | 66 | 0 | 4 | 482 |
| Population composition | | | | | | |
| Male teenager residents | The proportion of weighted male residents aged 15-19 | 0.025 | 0.023 | 0.000 | 0.023 | 0.125 |
| Male young residents | The proportion of weighted male residents aged 20-34 | 0.114 | 0.088 | 0.000 | 0.110 | 0.481 |
| Male middle-aged residents | The proportion of weighted male residents aged 35-64 | 0.175 | 0.097 | 0.000 | 0.206 | 0.347 |
| Male senior residents | The proportion of weighted male residents aged >= 65 | 0.055 | 0.045 | 0.000 | 0.060 | 0.222 |
| Female teenager residents | The proportion of weighted female residents aged 15-19 | 0.029 | 0.030 | 0.000 | 0.027 | 0.222 |
| Female young residents | The proportion of weighted female residents aged 20-34 | 0.113 | 0.090 | 0.000 | 0.103 | 0.493 |
| Female middle-aged residents | The proportion of weighted female residents aged 35-64 | 0.195 | 0.107 | 0.000 | 0.235 | 0.383 |
| Female senior residents | The proportion of weighted female residents aged >= 65 | 0.106 | 0.075 | 0.000 | 0.110 | 0.241 |

To reflect the distance decay effect, a series of coefficients are generated by the Gaussian function (*Eq. 1*) to weigh station-surrounding characteristics based on their distance to bike stations. Coefficients range between 0 and 1 – the shorter the distance from the characteristic to the bike station, the larger the coefficient attached to that characteristic. One challenge for considering distance decay effects is to identify an appropriate distance threshold. This study selected 150m as the service radius of bike stations, even though most earlier studies set the distance threshold between 200-300m (Bao et al., 2017; He et al., 2022; Ma et al., 2019). Some may doubt if the distance of 150m is adequate for the station-based bike share system because the user opts to park the bike at a nearest available station that might be far beyond 150m. However, in this way, people may adopt other modes of transport (e.g., by bus) instead of shared bikes for travelling because the bus stop density in Tartu (about nine bus stops per square kilometer within the study area) is quite high. Plus, Tartu is a small city with its urban areas occupying no more than 40 square kilometers. As a result, the travel distance within the urban area is rather short – if it requires people to walk a long distance to/from a bike station, they are more likely to walk to their destinations directly. Therefore, both above-mentioned aspects indicate that the distance threshold of 200-300m, applied by earlier studies with large cities as case studies (Bao et al., 2017; He et al., 2022; Ma et al., 2019), might be over-estimated for this study. Given that, the threshold distance of 150m was selected, as supported by Lam et al. (2022) – they examined that 150m is the most appropriate distance that people are willing to walk to a public transit station when comparing with 500m and 1000m.

Specifically, the proportion of weighted residential, commercial, institutional, industrial, and public land was calculated by *Eq. 2*. The weighted floor area ratio was calculated by *Eq. 3*. Proximity to bike lanes and bus stops were calculated by *Eq. 4* and *Eq. 5*, respectively. The number of weighted Russian and foreigner residents was calculated by *Eq. 6*. The population composition variable was calculated by *Eq. 7*.

$$G(d_{ik}) = \begin{cases} \frac{e^{-\frac{1}{2}\left(\frac{d_{ik}}{d_{max}}\right)^2} - e^{-\frac{1}{2}}}{1 - e^{-\frac{1}{2}}}, & d_{ik} < d_{max} \\ 0, & d_{ik} \geq d_{max} \end{cases} \quad Eq. 1$$

Where i is the index of bike stations, k is the index of built environment features (i.e., land use, buildings, bike lanes, bus stops, and population grids), d_{ik} is the distance between the station I and the built environment feature k , d_{max} is the distance threshold beyond which few effect of built environment features on bike-sharing usage would be observed.

$$land_prop_{il} = \frac{\sum_{l=1}^p G(d_{il})}{\sum_{k=1}^m G(d_{ik})} \quad Eq. 2$$

Where i is the index of bike stations, k is the index of land grids, l is the index of the land grids of a specific land use type, m is the total number of all land grids, p is the total number of land grids of a specific land use type.

$$floor_area_ratio_i = \frac{\sum_{z=1}^b G(d_{iz}) FA_z}{\sum_{k=1}^m 900 * G(d_{ik})} \quad Eq. 3$$

Where i is the index of bike stations, z is the index of buildings, k is the index of land grids, d_{iz} is the distance between the station i and the building z , d_{ik} is the distance between the building i and the land grid k , FA_z is total floor areas of the building z , m is the total number of all land grids, b is the total number of buildings.

$$proximity_bike_lane_i = \sum_{c=1}^a G(d_{ic}) \quad Eq. 4$$

Where i is the index of bike stations, c is the index of bike lane points, d_{ic} is the distance between the station i and the bike lane point c , a is the total number of bike lane points.

$$proximity_bus_stop_i = \sum_{s=1}^S G(d_{is}) \quad Eq. 5$$

Where i is the index of bike stations, s is the index of bus stops, d_{is} is the distance between the station i and the bus stop s , S is the total number of bus stops.

$$pop_ethnicity_i = \sum_{o=1}^O Pop_o G(d_{io}) \quad Eq. 6$$

Where i is the index of bike stations, o is the index of population grids, Pop_o is the number of residents living in the population grid o , d_{io} is the distance between the station i and the population grid o , O is the total number of population grids.

$$pop_composition_i = \frac{\sum_{f=1}^F Pop_f G(d_{if})}{\sum_{h=1}^H \sum_{u=1}^U Pop_u G(d_{iu})} \quad Eq. 7$$

Where i is the index of bike stations, f and u are the indices of population grids, h is the index of demographic groups, Pop_f and Pop_u is the number of residents living in the population grids f and u respectively, d_{if} is the distance between the station i and the population grid f , d_{iu} is the distance between the station i and the population grid u , F and U are the total number of population grids in a single demographic group, H is the number of demographic groups.

4. Research methodology

4.1. Non-parametric tests

This study conducted non-parametric tests to identify if there is a difference in bike-sharing usage (i.e., the station-level average daily trips) among eight demographic groups. Non-parametric tests were chosen because the station-level average daily trips of eight demographic groups were not normally distributed based on the results of both Shapiro-Wilk and Kolmogorov-Smirnov tests (Table 6). Specifically, the Mann-Whitney U test was first applied to test the gender-based difference in bike-sharing usage among four age groups respectively. The null hypothesis of this test is that there is no difference in the station-level daily average trips (in terms of central tendency) between genders. Then, the Kruskal-Wallis H test was carried out to test the age-based difference in bike-sharing usage. The null hypothesis is that all age groups have the same central tendency of the station-level daily average trips. Finally, the Mann-Whitney U test was conducted to test if there is a difference in bike-sharing usage within each age group pair for male and female users respectively. The null hypothesis is that there is no difference in the station-level daily average trips (in terms of central tendency) between the two age groups of the same gender.

Table 6. Normality test results of the station-level daily average trips by eight groups.

| Groups | Shapiro-Wilk | | Kolmogorov-Smirnov | |
|---------------------------|--------------|---------|--------------------|---------|
| | Statistic | p-value | Statistic | p-value |
| Male teenagers | 0.861 | 0.000 | 0.896 | 0.000 |
| Male young adults | 0.919 | 0.000 | 0.884 | 0.000 |
| Male middle-aged adults | 0.870 | 0.000 | 0.825 | 0.000 |
| Male seniors | 0.817 | 0.000 | 0.500 | 0.000 |
| Female teenagers | 0.836 | 0.000 | 0.798 | 0.000 |
| Female young adults | 0.877 | 0.000 | 0.847 | 0.000 |
| Female middle-aged adults | 0.870 | 0.000 | 0.770 | 0.000 |
| Female seniors | 0.791 | 0.000 | 0.500 | 0.000 |

4.2. Hot spot analysis

The hot spot analysis is a spatial technique for identifying the clustering of spatial phenomena. In recent years, several studies have applied it to evaluate the spatial pattern of bike-sharing usage (Choi et al., 2024; Choi et al., 2023). To this end, this method was selected to examine where higher and lower daily average trips of eight demographic groups were clustered in the study area respectively. The k-nearest neighbors algorithm (k=5) was used to create the spatial weight matrix. The outcome

of this analysis contains three parts: 1) hot spot: it indicates a statistically significant spatial cluster of higher bike-sharing usage, 2) cold spot: it indicates a statistically significant spatial cluster of lower bike-sharing usage, and 3) not significant: it means that no statistically significant spatial cluster is observed.

4.3. Regression analysis: Gradient boosting decision trees

The gradient boosting decision trees (GBDT) algorithm was selected to examine the non-linear associations of various station-surrounding characteristics with bike-sharing usage, as applied by (Chen & Ye, 2021; Lin et al., 2020; Zhuang et al., 2022). This is an ensemble machine learning algorithm that creates a strong learner by combining multiple weak learners (i.e., decision trees). To be more specific, it works by multiple decision trees, and every subsequent decision tree is trained to predict the error made by the previous decision tree until the final prediction is accurate enough, but no evident overfitting is observed.

Eight GBDT regression models were built to examine the non-linear associations of station-surrounding environments on bike-sharing usage by eight demographic groups respectively. Seventy percent of data served as the training set, while the remaining thirty percent of data served as the test set. The optimal parameters for training each model were identified based on the GridSearchCV method (Pedregosa et al., 2011), and the five-fold cross-validation method was applied. The parameter grid was defined as follows: {"n_estimators": [100, 150, 200, 250, 300], "learning_rate": [0.05, 0.055, 0.06, 0.065, 0.07, 0.075, 0.08], "max_depth": [3, 4, 5, 6], "max_features": [0.4, 0.5], "min_samples_leaf": [3, 4], "subsample": [0.5, 0.6]}. After 5,600 iterations, eight models with the optimal performance were acquired. Table 7 shows the R-squared, the mean squared error, and the mean absolute error of both training and test sets in eight models, indicating a good fit to the data for all models.

Furthermore, this study calculated the relative importance (RI) of eleven explanatory variables in each model. By comparing RIs among GBDT regression models, it could identify the station-surrounding characteristics that play an important role in affecting bike-sharing usage by each demographic group. Besides, the partial dependence plot (PDP) was drawn to display the non-linear associations of various station-surrounding characteristics with demographic-specific bike-sharing usage.

Table 7. Model performance.

| Model No. | Representative | Training sets | | | Test sets | | |
|------------------|---------------------------|----------------------|-------------|------------|----------------------|-------------|------------|
| | | R² | RMSE | MAE | R² | RMSE | MAE |
| Model 1 | Male teenagers | 0.988 | 0.624 | 0.266 | 0.634 | 3.306 | 2.621 |
| Model 2 | Male young adults | 0.962 | 1.165 | 0.671 | 0.742 | 2.685 | 2.247 |
| Model 3 | Male middle-aged adults | 0.959 | 0.597 | 0.346 | 0.731 | 1.029 | 0.751 |
| Model 4 | Male seniors | 0.950 | 0.079 | 0.049 | 0.702 | 0.169 | 0.139 |
| Model 5 | Female teenagers | 0.959 | 0.86 | 0.536 | 0.637 | 1.852 | 1.558 |
| Model 6 | Female young adults | 0.928 | 1.293 | 0.935 | 0.718 | 1.785 | 1.531 |
| Model 7 | Female middle-aged adults | 0.957 | 0.613 | 0.459 | 0.709 | 1.232 | 0.949 |
| Model 8 | Female seniors | 0.971 | 0.090 | 0.050 | 0.737 | 0.231 | 0.199 |

Note: RMSE refers to the root mean squared error; MAE refers to the mean absolute error.

5. Results

5.1. Differences in bike-sharing usage among varying demographic groups

Figure 6 presents the results of the Mann-Whitney U test for examining gender-based differences in the station-level average daily trips within four age groups respectively. It found that p-values for both teenager and young adult groups were below 0.05. This indicates that bike-sharing usage by male users was significantly different from that by female users within these two age groups. Moreover, it observed a higher mean rank in male groups, indicating that male teenagers and young adults used the bike share system more frequently than their female counterparts. Such a gender-based difference in bike-sharing usage was more evident within teenager groups because there was a larger gap of mean ranks between male and female users (difference=31.72) than young adult groups (difference=19.48).

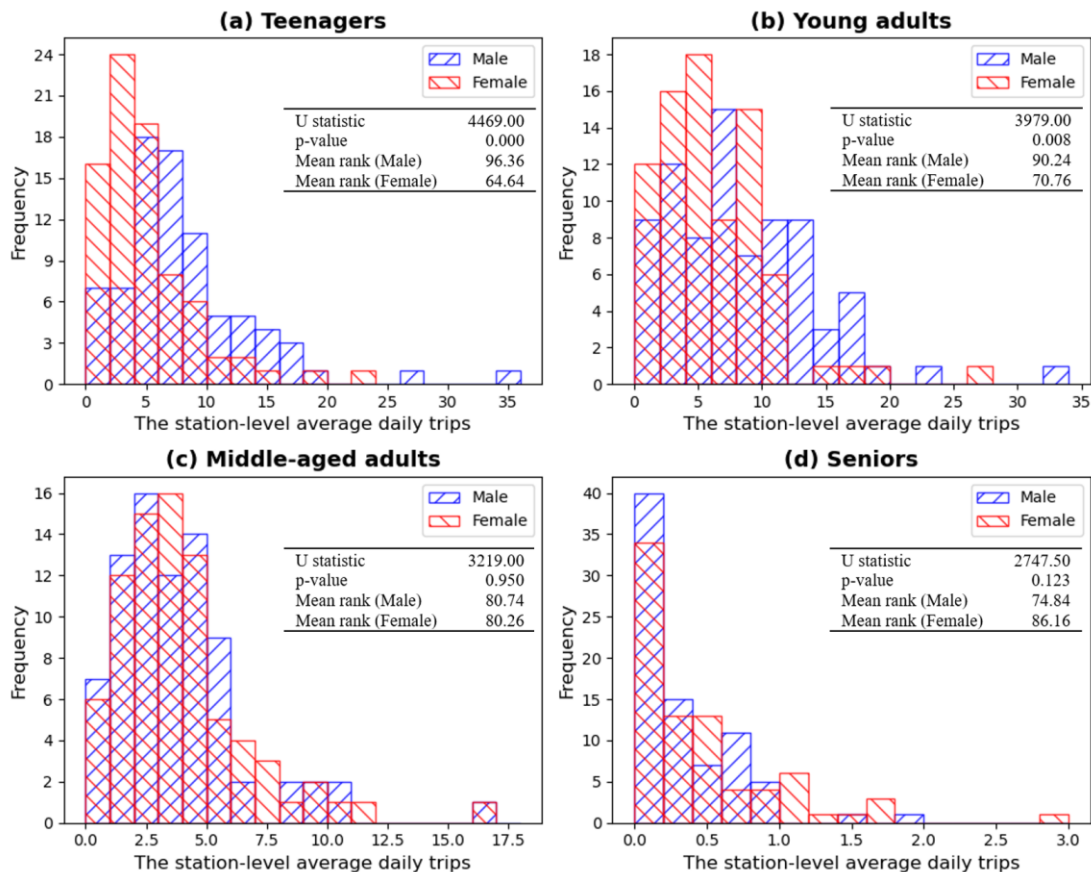


Figure 6. Gender-based differences in the station-level average daily trips.

However, no significant gender-based difference in bike-sharing usage was detected within middle-aged adult and senior groups as both p-values were above 0.05. By comparing Figure 6c and 6d, it clearly shows that the distribution of bike-sharing

usage was quite similar between male and female users, especially within middle-aged adult groups. Unlike other three age groups, it observed that female senior users had a slightly higher bike-sharing usage than male ones as a lower mean rank was found in male senior groups. A possible reason for this may be that the existing bike share system was more accessible to female than male seniors – as shown in Table 5, the average proportion of weighted male senior residents around bike stations was only at 0.055, about half that of female counterparts (at 0.106).

Table 8 shows the results of the Kruskal-Wallis H test for examining age-based differences in the station-level average daily trips within two gender groups respectively. Both groups had their p-values below 0.05, suggesting that at least one age group had its bike-sharing usage different with other age groups. Young adults and teenagers used the bike share system the most frequently, with their mean ranks around two hundred or above; it was followed by middle-aged users, while seniors experienced the least bike-sharing usage (the mean rank is only around fifty). A larger difference of mean ranks among four age groups was observed in male users (minimum=45.67, maximum=222.69), compared to female users (minimum=50.58, maximum=220.14). This indicates that age-based bike-sharing usage varied at a greater level among male users than female users.

Table 8. Age-based differences in the station-level average daily trips.

| Age groups | Male groups | | Female groups | |
|--------------------|---------------|------------|---------------|------------|
| | Median (IQR) | Mean ranks | Median (IQR) | Mean ranks |
| Teenagers | 6.672 (5.197) | 219.13 | 3.943 (3.295) | 196.26 |
| Young adults | 7.795 (8.152) | 222.69 | 5.672 (5.730) | 220.14 |
| Middle-aged adults | 3.590 (2.836) | 154.51 | 3.426 (2.582) | 175.03 |
| Seniors | 0.205 (0.414) | 45.67 | 0.279 (0.484) | 50.58 |
| H statistic | 191.849 | | 160.103 | |
| Degrees of freedom | 3 | | 3 | |
| p-value | 0.000 | | 0.000 | |

Note: IQR refers to the interquartile range.

For understanding detailed differences in bike-sharing usage among age groups, this study further carried out Mann-Whitney U tests for all age group pairs (n=6) by gender. Figure 7 presents the mean rank ratio of six age group pairs among male and female users respectively. Only one age group pair had its p-value above 0.05, i.e., male teenagers and young adults (Figure 7a). It suggests that there was no evident difference in bike-sharing usage between these two age groups. However, all other

age group pairs had their p-values below 0.05, indicating a common existence of age-based bike-sharing usage disparities. In addition, it observed quite large mean rank ratios of seniors to other groups (standing at close to 3). This further confirmed that seniors were highly under-represented in bike-sharing usage. It also found that mean rank ratios of all male age group pairs were higher than those of female counterparts. This indicates that there was a larger discrepancy in bike-sharing usage between male age groups than female ones.

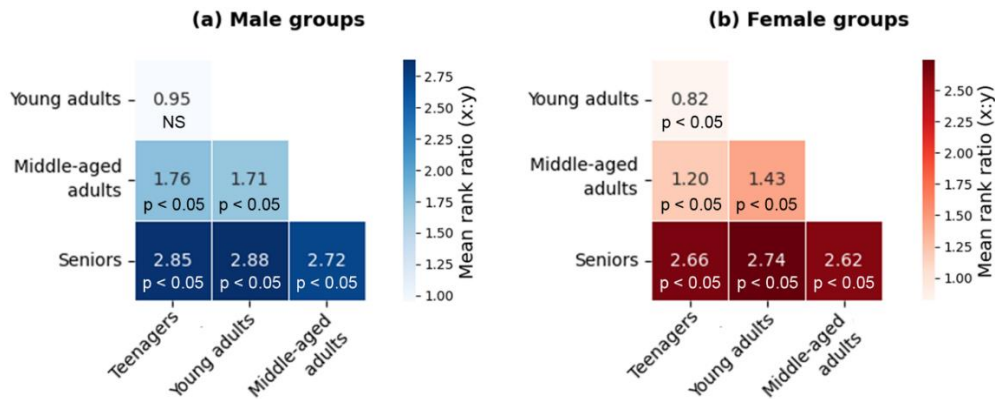


Figure 7. Differences in the station-level average daily trips among age group pairs.

Besides, a hot spot analysis was conducted to examine the spatial disparities of bike-sharing usage among varying demographic groups. Figure 8 illustrates the clusters of station-level daily average daily trips from eight groups. It did not observe an evident spatial disparity between genders for teenager and young adult groups by comparing Figure 8a with 8e and Figure 8b with 8f, even though earlier analyses suggest a gender-based difference in bike-sharing usage within these two age groups based on the outcome of the Mann-Whitney U test. Teenagers of both genders had their bike-sharing trips clustered in the urban center and eastern suburbs but experienced low bike-sharing usage in the northwestern and southwestern inner suburbs (within the 3km-buffer area). Young adults of both genders also had their high bike-sharing usage clustered in the urban center. Surprisingly, they were under-represented even in some urban central areas (within the 1km-buffer area); however, most of their lower bike-sharing usage was clustered in outer suburbs (outside of the 3km-buffer area).

By contrast, the spatial disparity in bike-sharing usage between genders was much more evident among middle-aged adults and seniors. For middle-aged adults, the main gender-based difference was that males had relatively lower bike-sharing usage in the southern and southeastern outer suburbs (Figure 8c), while females had

additional high bike-sharing usage clusters in the eastern suburb besides central clusters (Figure 8g). Moreover, male seniors had some bike-sharing trips clustered in the urban center, but the range of hot spots was much less widespread than other age groups (Figure 8d). No hot spots were observed within the 1km-buffer area for female seniors (Figure 8h). This indicates that they seldom used the bike share system in the urban center; instead, their bike-sharing trips were widely clustered in the direction of southeastern suburbs (around 2km- and 3km-buffer areas).

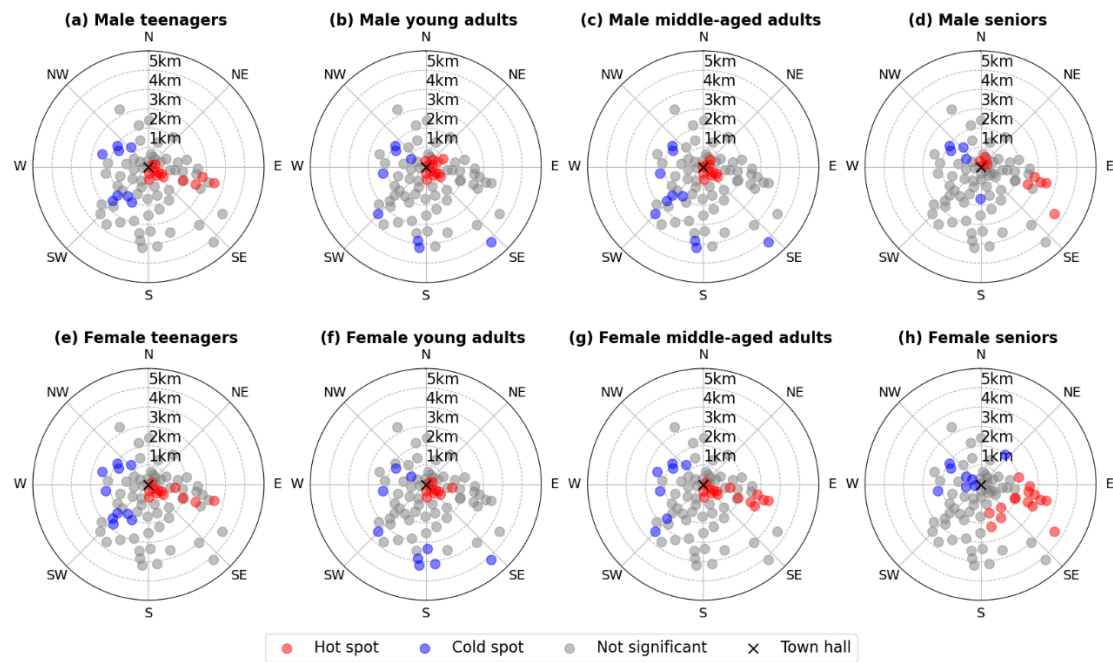


Figure 8. Clusters of the station-level daily average trips by eight groups.

5.2. Relationships between station-surrounding characteristics and bike-sharing usage of varying demographic groups

This study built eight GBDT regression models to empirically examine how station-surrounding characteristics affect bike-sharing usage of eight demographic groups respectively. The indicator of RI and PDPs were used to interpret their hidden complex associations. Table 9 presents RIs of eleven explanatory variables and their ranks in each model – the variable with a larger RI or a higher rank indicates its relatively more important role than other variables in affecting bike-sharing usage for that specific demographic group. Figures 9-11 illustrate the dependence between the station-level daily average trips and three categories of explanatory variables (i.e., land use, road infrastructure, and population distribution), respectively. Details of these results are presented as follows.

Table 9. Relative importance (RI) of explanatory variables influencing bike-sharing usage by eight demographic groups.

| Variable | Male teenagers | | Male young adults | | Male middle-aged adults | | Male seniors | | Female teenagers | | Female young adults | | Female middle-aged adults | | Female seniors | |
|---------------------------------------|----------------|------|-------------------|------|-------------------------|------|--------------|------|------------------|------|---------------------|------|---------------------------|------|----------------|------|
| | RI | Rank | RI | Rank | RI | Rank | RI | Rank | RI | Rank | RI | Rank | RI | Rank | RI | Rank |
| Land use (n=6) | | | | | | | | | | | | | | | | |
| Residential land | 0.118 | 3 | 0.063 | 8 | 0.111 | 8 | 0.055 | 6 | 0.091 | 9 | 0.053 | 4 | 0.085 | 8 | 0.077 | 5 |
| Commercial land | 0.146 | 2 | 0.116 | 5 | 0.147 | 3 | 0.238 | 3 | 0.141 | 1 | 0.101 | 2 | 0.149 | 4 | 0.101 | 2 |
| Institutional land | 0.047 | 9 | 0.034 | 10 | 0.028 | 10 | 0.044 | 10 | 0.059 | 10 | 0.043 | 9 | 0.059 | 10 | 0.073 | 9 |
| Industrial land | 0.046 | 10 | 0.021 | 11 | 0.013 | 11 | 0.024 | 11 | 0.036 | 11 | 0.031 | 11 | 0.038 | 11 | 0.019 | 11 |
| Public land | 0.097 | 5 | 0.119 | 4 | 0.130 | 4 | 0.114 | 4 | 0.087 | 3 | 0.105 | 5 | 0.094 | 3 | 0.080 | 4 |
| Floor area ratio | 0.179 | 1 | 0.173 | 1 | 0.163 | 2 | 0.109 | 2 | 0.220 | 4 | 0.269 | 1 | 0.127 | 1 | 0.066 | 3 |
| Transport Infrastructure (n=2) | | | | | | | | | | | | | | | | |
| Bike lanes | 0.107 | 4 | 0.079 | 7 | 0.111 | 5 | 0.075 | 5 | 0.113 | 6 | 0.085 | 3 | 0.079 | 5 | 0.127 | 7 |
| Bus stops | 0.041 | 11 | 0.059 | 9 | 0.049 | 7 | 0.093 | 7 | 0.062 | 5 | 0.064 | 8 | 0.041 | 7 | 0.047 | 10 |
| Population distribution (n=3) | | | | | | | | | | | | | | | | |
| Russian residents | 0.085 | 6 | 0.080 | 6 | 0.036 | 9 | 0.060 | 9 | 0.070 | 8 | 0.049 | 7 | 0.079 | 9 | 0.155 | 6 |
| Foreigner residents | 0.068 | 7 | 0.134 | 2 | 0.165 | 1 | 0.070 | 1 | 0.070 | 7 | 0.133 | 6 | 0.184 | 2 | 0.088 | 1 |
| Population composition | 0.065 | 8 | 0.123 | 3 | 0.046 | 8 | 0.119 | 8 | 0.052 | 2 | 0.065 | 10 | 0.066 | 6 | 0.166 | 8 |
| Sum | 1.000 | | 1.000 | | 1.000 | | 1.000 | | 1.000 | | 1.000 | | 1.000 | | 1.000 | |

5.2.1. The impact of land use on bike-sharing usage

Among six land use variables, the floor area ratio had the most important effect on bike-sharing usage because its RIs ranked very high in all models except the Model 8 (Table 9). According to Figure 9f, its overall positive association with bike-sharing usage was observed. Both findings indicate that the bike share system was used intensively in the area with the high area floor ratio (e.g., areas occupied by high-rise apartment or office buildings, or shopping malls). This positive impact was especially reflected within teenager and young adult groups as the RI of the floor area ratio ranked first in all four models (i.e., Models 1, 2, 5, and 6) (Table 9). It also detected a gender-based difference within these two groups – the RIs for female groups were evidently higher than male groups. This indicates that female teenagers and young adults were more subject to the floor area ratio than their male counterparts. By contrast, it ranked ninth in the Model 8, suggesting that the floor area ratio was not an important determinant for female seniors to use the bike share system.

It is followed by the commercial land – all its RIs were above 0.1 in eight models (Table 9), and Figure 9b illustrates its positive impact on bike-sharing usage. Especially, a very high RI (i.e., 0.238) was observed in the Model 4, suggesting that male seniors frequently use the bike share system in commercial areas. Besides, its RIs (close to 0.15) were also quite high for teenagers and middle-aged adults. Even though it found that the commercial land was important for bike-sharing usage of both groups, their travel purposes might be different. For instance, teenagers probably use share bikes in commercial areas for recreational purposes, while middle-aged adults are more likely to cycle in commercial areas for commuting from their workplaces. By contrast, the positive impact of commercial land was relatively weak for young adult groups as their RIs (at around 0.1) ranked relatively lower than other groups (Table 9).

The public land also played a relatively important role in bike-sharing usage. Figure 9e shows an apparent positive relationship between public land and bike-sharing usage, and its impact was quite equal among all demographic groups, with its RIs ranging from 0.08 to 0.13 (Table 9). One possible reason is that public land (e.g., urban green space) can bring a wide range of benefits accommodating the needs of varying groups – for instance, it can provide spaces for physical exercises,

recreational or social activities - and therefore it is valued by almost all groups no matter what their age or gender is.

By contrast, the impact of residential land on bike-sharing usage was much more diverse among different groups - teenager groups had an evidently higher RI ranks than senior groups (Table 9). In addition, Figure 9a illustrates that the proportion of residential land had an overall negative effect on bike-sharing usage. This suggests that bike-sharing trips were less likely to be generated in areas where the residential land dominated. This might be due to the fact that several bike stations were located in sparsely populated residential areas in suburbs where residents were largely dependent on private vehicles/buses for commuting to urban central areas given their farther travel distance demands. It also found that this variable was especially important for teenagers as its high-rank RIs in the Model 1 and Model 5 (Table 9). This might be because teenagers were most likely to reside with their parents in suburban houses, and therefore nearby bike stations could only serve a limited number of teenager residents.

Besides, this study did not observe an evident impact of the institutional and industrial land on bike-sharing usage as their RIs ranked very low in all models (Table 9). Especially, the industrial land ranked second to the last in the Model 1 and had the least RIs in all remaining models (Table 9). Figure 9c and 9d also display almost horizontal lines for most groups, indicating that bike-sharing usage remained stable no matter how high/low the proportion of the institutional or industrial land was. However, it still detected a slight positive impact of the institutional land on bike-sharing usage of male teenagers and young adults (Figure 9c). This suggests that share bikes located around schools or university buildings served as the important transport vehicle for accommodating male students' commuting needs.

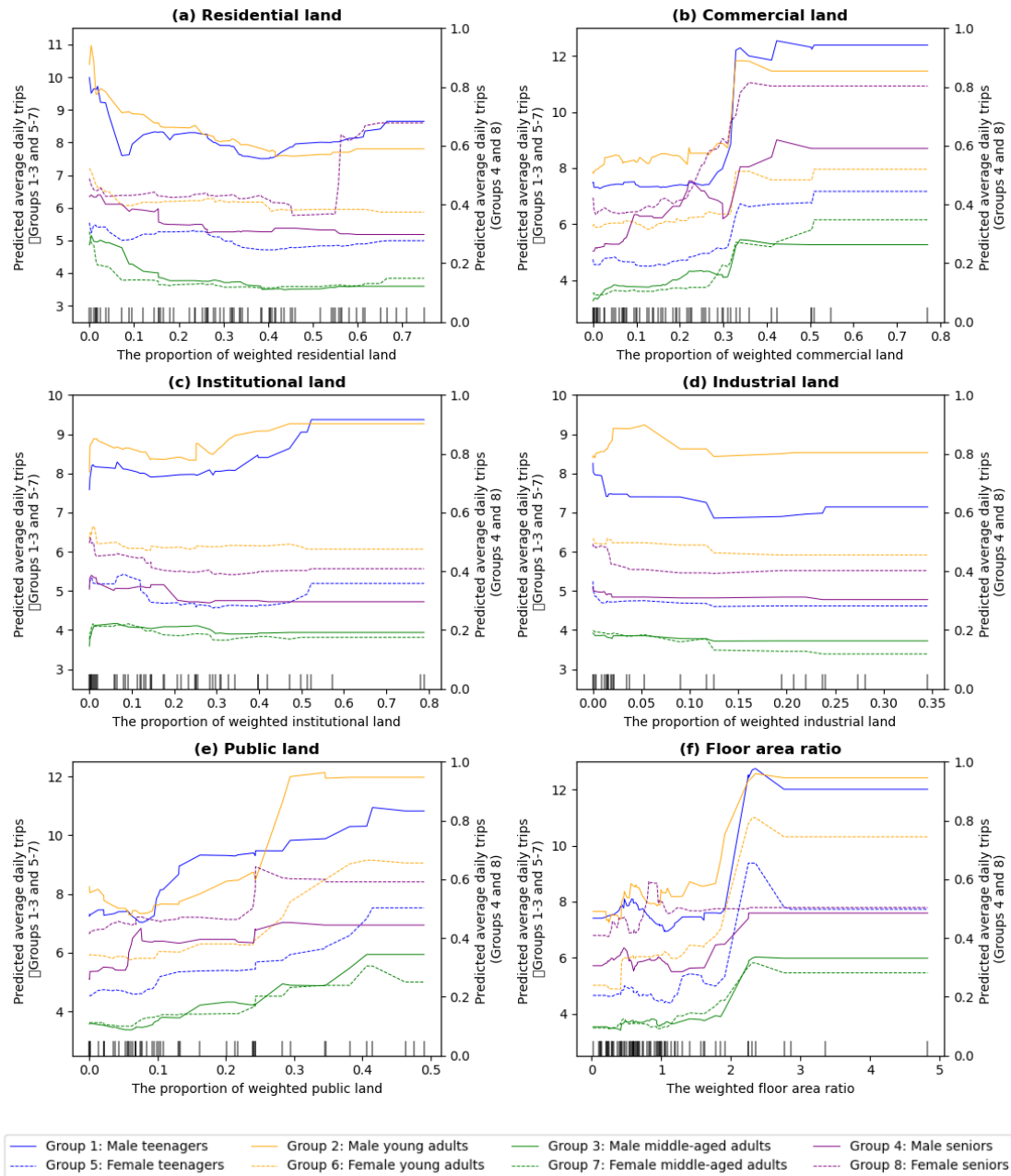


Figure 9. Non-linear associations of station-surrounding land use with bike-sharing usage: (a) residential land, (b) commercial land, (c) institutional land, (d) industrial land, (e) public land, and (f) floor area ratio.

5.2.2. The impact of transport infrastructure on bike-sharing usage

The transport infrastructure factor consists of two variables, i.e., bike lanes and bus stops. Figure 10a illustrates the positive association of bike lanes with bike-sharing usage for all groups. When combined this with RIs in Table 9, it can be seen that bike lanes played a fairly important role in increasing bike-sharing usage of female teenagers and seniors, as the RIs of bike lanes ranked third in both Model 5 and

Model 8. Moreover, the RIs for both groups were slightly higher than those for their male counterparts, suggesting a gender difference. One possible explanation is that female users were more sensitive to road safety than male users – bike lanes could largely reduce the threat from motorized vehicles on the road, and therefore increase the likelihood of bike-sharing usage by female groups. In addition, this study did not observe a clear correlation between proximity to bus stops and bike-sharing usage since their RIs ranked quite low in most models (Table 9). Figure 10b also shows that the number of bike-sharing trips fluctuated at a specific horizontal level as the number of weighted bus stops changed.

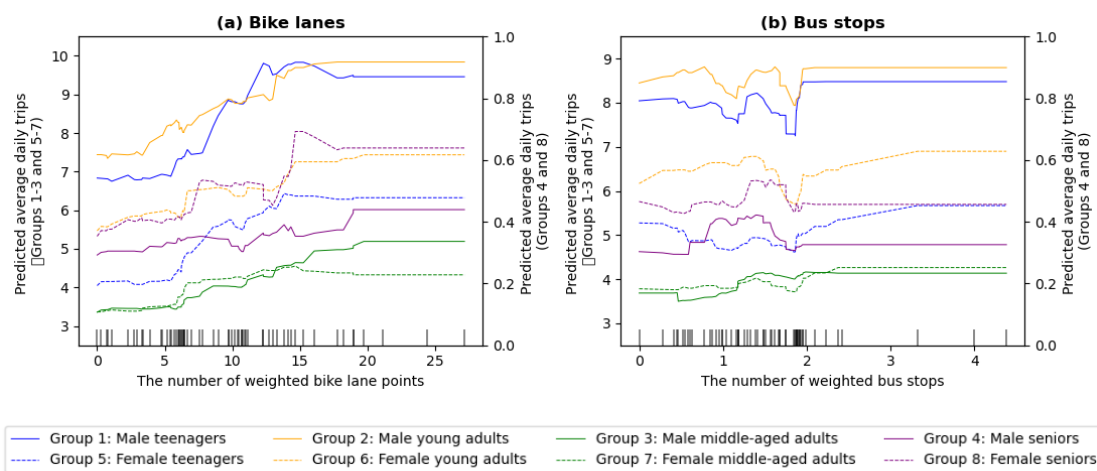


Figure 10. Non-linear associations of station-surrounding transport infrastructure with bike-sharing usage: (a) bike lanes and (b) bus stops.

5.2.3. The impact of population distribution on bike-sharing usage

The population distribution factor consists of three variables, i.e., Russian residents, foreigner residents, and population composition. First of all, it found that the variable of foreigner residents played an important role in bike-sharing usage of young adults and middle-aged adults, as it had very high RIs from corresponding models (i.e., Model 2, Model 3, Model 6, and Model 7), ranking second and first respectively (Table 9). According to Figure 11b, it observed an evident positive relationship between the number of foreigner residents and the daily average trips of these age groups. Tartu is a university town, and its main foreigner source is international students and graduates of the University of Tartu. As shown in Figure 4c, foreigners were mainly residing in the urban central area where major university buildings and companies are located. Therefore, the share bike is the most likely to be the choice of these foreigners for commuting to schools or companies in urban central areas.

Meanwhile, according to Figure 11b, male young adults used the bike share system much more frequently than their female counterparts; however, no evident gender-based difference existed within middle-aged groups. By contrast, the number of Russian residents was not considered to be an important determinant of bike-sharing usage in general, as its RIs were quite low in most models (Table 9). It only had a high RI (at 0.155, ranking second) in model 8 (Table 9), and Figure 11a shows its overall positive associations with bike-sharing usage. This suggests that Russian female seniors are largely likely to be the potential bike-sharing users.

In addition, population composition played an important role in affecting bike-sharing usage by seniors, as it ranked second and first in Model 4 and Model 8 respectively (Table 9). According to Figure 11c, a positive association was observed between population composition and bike-sharing usage of seniors. Both results indicate that share bikes located in areas with dense senior residents gained high usage by them. Such an interpretation was also applied for male young adults according to the high RI of population composition (ranking third among eleven variables) in the Model 2 (Table 9) and its PDP (indicated by the orange solid line) in Figure 11c.

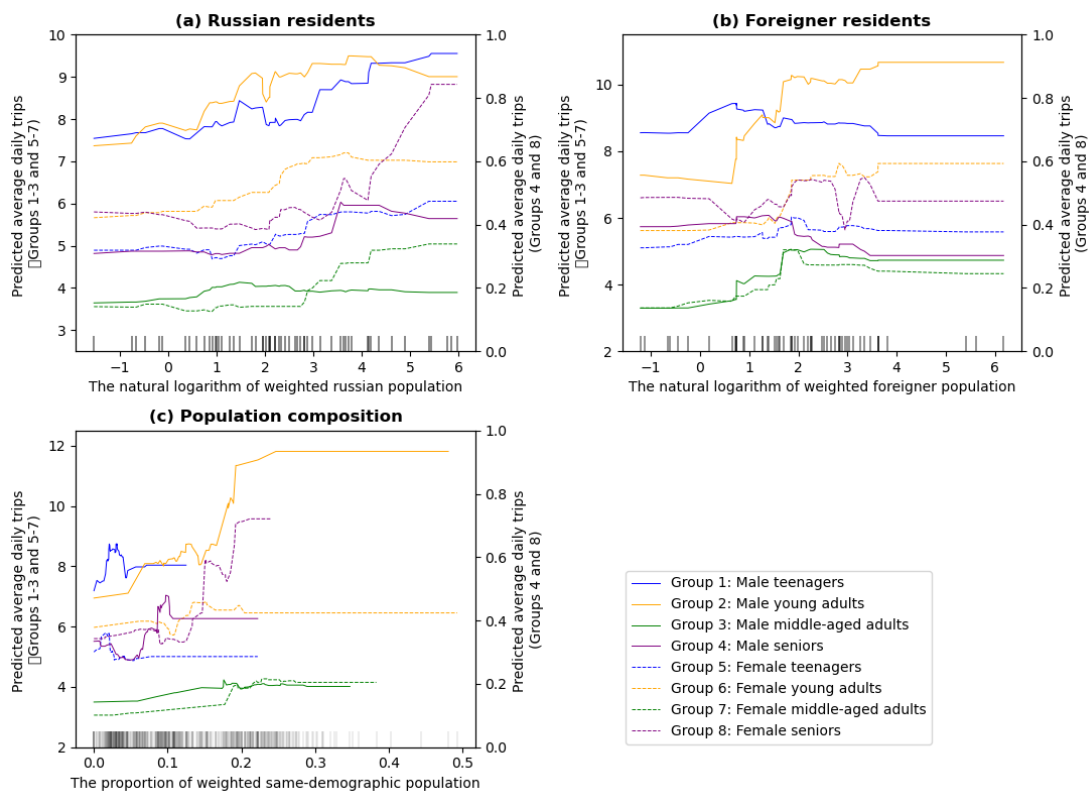


Figure 11. Non-linear associations of station-surrounding demographics with bike-sharing usage: (a) Russian residents, (b) foreign residents, and (c) population composition.

6. Discussion

The results of the non-parametric tests revealed the gender-based difference in bike-sharing usage, and such a difference varied between age groups. For the younger generation, male users generated more bike-sharing trips than female counterparts; however, the bike share system was more frequently used by female than male seniors. This finding complemented understandings from previous studies (Goodman & Cheshire, 2014; Murphy & Usher, 2015), which concluded that male users consistently dominated bike-sharing usage without considering the influence of age.

The floor area ratio was regarded as the most important determinant of bike-sharing usage in this study. A large number of earlier studies have also found its positive impact on bike-sharing usage (Meng & Zacharias, 2021). However, the new insight is that this study found this characteristic especially important for teenagers and young adults. This may be because of two reasons. First, these groups are actively in the commercial center (with the high floor area ratio) because it offers various amenities (i.e., various shops, restaurants, bars, or cafes) to accommodate their various daily needs; it is also where lots of young people work. As a result, bike-sharing trips could be stimulated by them for visiting/returning from the commercial center. The second reason is due to the accommodation preference of the younger generation - they prefer to live in multi-story apartment buildings given its affordability. These buildings can accommodate a quite larger number of residents than houses, thus potentially bringing more bike-sharing trips. In addition, the first reason can also be used to explain the importance of the commercial land identified in this study, supported by existing empirical evidence (Lee et al., 2022).

However, the negative impact of the residential land on bike-sharing usage was observed, inconsistent with previous studies (Lee et al., 2022; Xiao & Wei, 2023). This is largely because the discrepancy in population density – if there is a relatively even population distribution, a higher proportion of the residential land around bike stations in general indicates a larger number of residents there, potentially increasing the bike-sharing usage. However, this phenomenon is not applied to Tartu because several bike stations were located around sparsely populated suburbs, occupying a larger proportion of residential land. As result, these stations can only serve a few nearby residents but have limited attractiveness to generate trips by visitors.

Meanwhile, this study did not observe the impact of bus stops on bike-sharing usage, although a large number of existing studies have emphasized the role of the bike share system in connecting with other transports such as metros and buses (Kim, 2023; Wu et al., 2021; Xu et al., 2019); it is because that these bike share systems are located in metropolises where people are required to travel longer distances and thus necessarily transfer different means of transport to their destination. By contrast, Tartu is a small city with an urban area of no more than 40 square kilometers. Especially, it has densely distributed bus stops and bike stations, allowing residents to reach their destinations by a single trip. Therefore, a combination of different public transit systems would largely not happen in Tartu, Estonia. Besides, a quite equal positive associations of the public land with bike-sharing usage were detected among public, while institutional land was only positively related to bike-sharing usage of male teenagers and young adults due to their schooling purposes.

Bike lanes were of particular concern to female users due to road safety issues. This is also supported by (AitBihiOuali & Klingen, 2022; Akar et al., 2013; Le et al., 2019; Pearson et al., 2023). However, it can be seen that the existing bike lanes around bike stations were under-developed in Tartu (Figure 3b) – more than one fifth of bike stations (n=17) were not connected by bike lanes. Therefore, local governments could be dedicated to constructing a more accessible bike lane system to encourage bike-sharing usage by female groups. Relevant stakeholders could also consider installing new bike stations in areas with a high density of senior residents, as this study found that the population distribution of senior residents around bike stations played a particularly important role in increasing their use of the bike share system when compared to other age groups.

7. Conclusion

This study is set to understand the differences in bike-sharing usage among varying demographic groups and its associations with station-surrounding characteristics. To conclude, bike-sharing usage varied among eight demographic groups in general; bike-sharing usage by varying demographic groups was differently influenced by various station-surrounding characteristics.

To answer the first research question: The non-parametric tests discovered the gender-based difference in bike-sharing usage among teenagers and young adults, but such a difference was not detected among middle-aged adults and seniors; the age-based difference in bike-sharing usage commonly existed, but it varied between genders. The hot spot analysis indicated no evident spatial disparity between genders among teenagers and young adults. Male and female middle-aged adults had some clusters of higher and lower bike-sharing usage in different areas. Senior groups were under-represented in the urban center. Especially, there was no cluster of high bike-sharing trips by female seniors in the urban center; male seniors had some bike-sharing trips clustered in the urban center, but the range of hot spots was much less widespread than other age groups.

To answer the second research question: The outcome of GBDT regression models disclosed that the floor area ratio around stations was the most important characteristic and had a positive effect on bike-sharing usage; however, such an important effect was not observed for the female elderly users. The proportion of commercial land also played a very important role in and was positively associated with bike-sharing usage, especially for male seniors. Other relative important characteristics include the proportion of public land, bike lanes, the number of foreigner residents, and population composition; all of them had a positive impact on bike-sharing usage. However, their relative importance and non-linear associations with bike-sharing usage were different among varying demographic groups. For instance, bike lanes were of particular concern to female users due to road safety issues; the population distribution of senior residents around bike stations played a particularly important role in increasing their use of the bike share system when compared to other age groups.

These findings have several implications. First, this study provides valuable insights into what demographic groups are using the bike share system or are underrepresented. Such information can guide efforts to promote inclusivity and equity in urban transportation, ensuring that the bike share system is appealing to a diverse range of users. Second, the most important built or demographic environments identified from this study can inform bike share system owners of interventions to increase bike-sharing usage by particular groups. Third, it also enables urban planners to know what station-surrounding characteristics are more likely to boost cycling activities, making cities more sustainable and reducing traffic congestion and air pollution.

By means of a machine learning approach, this study identified the most important station-surrounding characteristics affecting bike-sharing usage for varying demographic groups and extracted their non-linear associations. It made up for the shortcomings of existing studies on linearity between bike-sharing usage and its factors using the traditional statistical methods as well as failing to address the demographic-specific variation in bike-sharing usage. However, it still has some limitations. First, this study did not contain the trip records occurring in the winter, and the weather factor was also not taken into account. This might fail to provide more in-depth information because the climate and weather conditions might exist as a determinant of bike-sharing usage for different demographic groups. For instance, males might be more resilient to extreme climate and weather than females, and the younger groups could also do better than the elderly. Second, given the data unavailability, this study did not include some socio-economic attributes of residents as independent variables, e.g., occupation, incomes, family composition, and whether possessing a car/bike or not. These attributes can help identify if a resident is a potential bike-sharing user and thus provide more targeted information regarding how to accommodate these potential users to increase their bike-sharing usage. Third, this study fails to provide evidence of causal relationships between built or demographic environments and bike-sharing usage. Future work could contribute to the above-mentioned aspects.

Differences in bike-sharing usage among varying demographic groups and its association with station-surrounding characteristics in Tartu, Estonia

Xiao Cai

Summary

The bike share system offers a wide range of benefits to promote human mobility. However, many bike share systems favor particular demographic groups (e.g., younger people and males), suggesting that the resulting benefits are not equally distributed among the public. To this end, the goal of this thesis is to empirically examine the differences in bike-sharing usage among varying demographic groups and its association with station-surrounding characteristics in Tartu, Estonia. Specifically, the following two research questions were answered:

- **Is there a difference in bike-sharing usage by various demographic groups, and what is the difference (especially from the spatial perspective)?**
- **How do various station-surrounding characteristics (i.e., land use, transport infrastructure, and population distribution) influence bike-sharing usage of different demographic groups?**

For answering these questions, this study firstly grouped two-month bike-sharing trip records (n=177,536) into eight categories based on the gender and age of bike-sharing users. Specifically, bike-sharing users of eight categories include 1) male teenagers, 2) male young adults, 3) male middle-aged adults, 4) male seniors, 5) female teenagers, 6) female young adults, 7) female middle-aged adults, and 8) female seniors. Then, the station-level daily average trips of eight groups were calculated, as an indicator of bike-sharing usage. Further, this study carried out non-parametric tests (i.e., the Mann-Whitney U test and the Kruskal-Wallis H test) to examine the gender-based and age-based differences in bike-sharing usage. This study also conducted the hot spot analysis to compare the spatial patterns of bike-sharing usage among eight demographic groups. Finally, eight gradient boosting decision tree (GBDT) regression models were built to examine the impact of station-surrounding characteristics (including eleven explanatory variables) on demographic-specific bike-sharing usage. In detail, the indicator of relative importance (RI) was calculated for understanding how important the role of each variable in affecting bike-sharing usage is, while the partial dependence plot (PDP) was drawn to reflect their non-linear associations. By means of these two techniques, the relationships between station-surrounding

characteristics and bike-sharing usage were therefore compared among eight demographic groups.

The non-parametric tests discovered the gender-based difference in bike-sharing usage among teenagers and young adults, but such a difference was not detected among middle-aged adults and seniors; the age-based difference in bike-sharing usage commonly existed, but it varied between genders. The hot spot analysis indicated no evident spatial disparity between genders among teenagers and young adults. Male and female middle-aged adults had some clusters of higher and lower bike-sharing usage in different areas. Senior groups were under-represented in the urban center. Especially, there was no cluster of high bike-sharing trips by female seniors in the urban center; male seniors had some bike-sharing trips clustered in the urban center, but the range of hot spots was much less widespread than other age groups.

The outcome of GBDT regression models disclosed that the floor area ratio around stations was the most important characteristic and had a positive effect on bike-sharing usage; however, such an important effect was not observed for the female elderly users. The proportion of commercial land also played a very important role in and was positively associated with bike-sharing usage, especially for male seniors. Other relative important characteristics include the proportion of public land, bike lanes, the number of foreigner residents, and population composition; all of them had a positive impact on bike-sharing usage. However, their relative importance and non-linear associations with bike-sharing usage were different among varying demographic groups. For instance, bike lanes were of particular concern to female users due to road safety issues; the population distribution of senior residents around bike stations played a particularly important role in increasing their use of the bike share system when compared to other age groups.

In conclusion, bike-sharing usage varied among eight demographic groups in general; bike-sharing usage by varying demographic groups was differently influenced by various station-surrounding characteristics. The findings from this study can provide the bike share system owner with significant insights into interventions for increasing bike-sharing usage of groups with specific demographics, so that inclusivity and equity in urban transportation could be promoted.

Rattaringluse kasutuse erinevused erinevate demograafiliste rühmade puhul ja seosed rattaringluse jaamade ümbruse omadustega Tartus, Eestis

Xiao Cai

Kokkuvõte

Rattaringlus pakub mitmeid eeliseid inimeste liikuvuse edendamiseks. Rattaringluse süsteemid eelistavad tihti teatud demograafilisi rühmi (nt. noori ja mehi), mis viitab sellele, et sellest tulenev kasu ei ole võrdselt jaotunud. Käesoleva magistritöö eesmärk on empiiriliselt uurida erinevusi rattaringluse kasutamises erinevate demograafiliste rühmade puhul ja selle seoseid rattaringluse jaamasid ümbritseva keskkonna tunnustega Tartus, Eestis. Konkreetselt saavad vastuse järgnevad kaks uurimisküsimust:

- Kas erinevate demograafiliste rühmade rattaringluse kasutuses on erinevusi ning millised need erinevused on (eriti ruumilisest vaatenurgast)?
- Kuidas erinevad rattajaamasid ümbritsevad tunnused (nt maakasutus, transpordi taristu ja rahvastiku jaotus) mõjutavad rattaringluse kasutust erinevates demograafilistes rühmades?

Nendele küsimustele vastamiseks grupeeriti esmalt kahe kuu rattaringluse reise kirjeld (n = 177 536) kaheksasse kategooriasse põhinedes reisi tegijate soo ja vanuse tunnustele. Täpsemalt, reisi tegijate kaheksa kategooriat on järgnevad: 1) teismelised mehed, 2) noored täiskasvanud mehed, 3) keskealised mehed, 4) vanemaealised mehed, 5) teismelised naised, 6) noored täiskasvanud naised, 7) keskealised naised ja 8) vanemaealised naised. Seejärel arvutati rattaringluse kasutamise indikaatoriks kaheksa rühma kohta keskmine reise hulk päevas rattajaamade kohta. Läbi on viidud mitte-parameetrilised testid (Mann-Whitney U-test ja Kruskal-Wallis H-test), et uurida soolisi ja vanuselisi erinevusi rattaringluse kasutuses. Uuringus viidi läbi ka kuumkoha analüüs, et võrrelda kaheksa demograafilise rühma rattaringluse kasutuse ruumilisi mustreid. Lõpuks koostati kaheksa gradient-võimendustüübiga otsustuspuu (GBDT) regressioonimudelit, et uurida rattajaama ümbruse tunnuste (sealhulgas üheteist selgitava muutuja) mõju demograafiliselt rattaringluse kasutamisele. Arvutati suhtelise tähtsuse näitaja (RI), et mõista, kui oluline on iga muutuja roll rattaringluse kasutamisel, samas kui osalise sõltuvuse joonis (partial dependence plot) (PDP) koostati, et kajastada mittelinearseid seoseid. Nende kahe tehnika abil saab võrrelda

kaheksa demograafilise rühma vahelisi seoseid rattajaama ümbritsevate tunnuste ja rattaringluse kasutamise vahel.

Mitteparameetriliste testide tulemused näitasid soopõhiseid erinevusi rattaringluse kasutuses ning need erinevused varieerusid vanuserühmade lõikes. Samuti leiti, et keskealised täiskasvanud ja vanemaealised inimesed kasutasid sageli rattaringluse süsteemi erinevates linnapiirkondades lähtuvalt soost, samas kui teismeliste ja noorte täiskasvanute seas ei täheldatud ruumilist erinevust meeste ja naiste vahel. Regressioonimudelite tulemused näitasid, et rattajaamade ümbruse ehitistega kaetud pindala osakaalu suhe oli kõige olulisem tunnus, millel oli positiivne mõju rattaringluse kasutamisele; see oluline mõju ei tulnud esile aga vanemaealiste naiste puhul. Ärimaa osakaal mängis samuti väga olulist rolli ning oli positiivselt seotud rattaringluse kasutamisega, eriti vanemaealiste meeste puhul. Teised olulised näitajad olid avaliku maa osakaal, jalgrattateed, välismaalastest elanike arv ja rahvastiku koosseis. Kõigil neil oli positiivne mõju rattaringluse kasutamisele, kuid nende mittelineaarsus oli veidi erinev.

Kokkuvõtteks võib öelda, et rattaringluse kasutamine varieerub kaheksas demograafilises rühmas ning erinevate demograafiliste rühmade rattaringluse kasutamine oli mõjutatud rattajaamasid ümbritsevatest tunnustest erinevalt. Need tulemused võivad anda rattaringluse süsteemi rajajatele olulisi teadmisi sekkumistest, mille abil suurendada teatud demograafiliste rühmade rattaringluse kasutusaktiivsust, mis aitaks edendada kaasamist ja võrdsust linnatranspordi planeerimisel.

Acknowledgements

I would like to thank my supervisors: Dr. Siiri Silm and Dr. Amnir Hadachi. They provided me with constructive feedback for making this thesis more than better.

I would like to thank Prof. Dr. Frank Witlox and his student (Dr. Tanhua Jin). The research project (conducted during my stay at the Economic Geography Research group, Ghent University) supervised by them served as a significant part of this thesis. I also would like to thank Xinyue Gu (a PhD student from the Hong Kong Polytechnic University) for her contributions to the research project.

I would like to thank Prof. Dr. Tuuli Toivonen and Dr. Elias Willberg as well as other Digital Geography Lab members at the University of Helsinki. They proposed several useful suggestions for me to interpret the findings of this thesis.

I would like to thank two reviewers (Prof. Dr. Tõnu Oja and Martin Haamer) at the Department of Geography, University of Tartu. Their comments helped a lot to improve the quality of this thesis.

I would like to thank all lecturers of my courses at the University of Tartu. Because of them, I enhanced my spatial data analysis skills and acquired machine learning techniques, so that I can conduct this thesis.

I would like to thank the University of Tartu for offering me this study opportunity with the tuition-waiver scholarship.

I would like to thank the Erasmus+ program for financially supporting my mobility at the Ghent University and the University of Helsinki.

The most importantly, I would like to thank my parents; they always unconditionally support me both emotionally and financially, and are proud of my every baby step forward, so that I can achieve who I am now.

References

- AitBihiOuali, L., & Klingen, J. (2022). Inclusive roads in NYC: Gender differences in responses to cycling infrastructure. *Cities*, 127, Article 103719. <https://doi.org/10.1016/j.cities.2022.103719>
- Akar, G., Fischer, N., & Namgung, M. (2013). Bicycling Choice and Gender Case Study: The Ohio State University. *International Journal of Sustainable Transportation*, 7(5), 347-365. <https://doi.org/10.1080/15568318.2012.673694>
- Bao, J., Shi, X. M., & Zhang, H. (2018). Spatial analysis of bikeshare ridership with smart card and POI data using geographically weighted regression method. *Ieee Access*, 6, 76049-76059. <https://doi.org/10.1109/access.2018.2883462>
- Böcker, L., Anderson, E., Uteng, T. P., & Throndsen, T. (2020). Bike sharing use in conjunction to public transport: Exploring spatiotemporal, age and gender dimensions in Oslo, Norway. *Transportation Research Part a-Policy and Practice*, 138, 389-401. <https://doi.org/10.1016/j.tra.2020.06.009>
- Chen, E. H., & Ye, Z. R. (2021). Identifying the nonlinear relationship between free-floating bike sharing usage and built environment. *Journal of Cleaner Production*, 280, Article 124281. <https://doi.org/10.1016/j.jclepro.2020.124281>
- Cheng, L., Mi, Z. F., Coffman, D., Meng, J., Liu, D. N., & Chang, D. F. (2022). The role of bike sharing in promoting transport resilience. *Networks & Spatial Economics*, 22(3), 567-585. <https://doi.org/10.1007/s11067-021-09518-9>
- Choi, S. J., Jiao, J. F., & Karner, A. (2024). Tracing the effects of COVID-19 on short and long bike-sharing trips using machine learning. *Travel Behaviour and Society*, 35, Article 100738. <https://doi.org/10.1016/j.tbs.2024.100738>
- Choi, S. J., Jiao, J. F., Lee, H. K., & Farahi, A. (2023). Combatting the mismatch: Modeling bike-sharing rental and return machine learning classification forecast in Seoul, South Korea. *Journal of Transport Geography*, 109, Article 103587. <https://doi.org/10.1016/j.jtrangeo.2023.103587>
- de Haas, M., Kroesen, M., Chorus, C., Hoogendoorn-Lanser, S., & Hoogendoorn, S. (2022). E-bike user groups and substitution effects: evidence from longitudinal travel data in the Netherlands. *Transportation*, 49(3), 815-840. <https://doi.org/10.1007/s11116-021-10195-3>

- Du, Y. C., Deng, F. W., & Liao, F. X. (2019). A model framework for discovering the spatio-temporal usage patterns of public free-floating bike-sharing system. *Transportation Research Part C-Emerging Technologies*, 103, 39-55. <https://doi.org/10.1016/j.trc.2019.04.006>
- El-Assi, W., Mahmoud, M. S., & Habib, K. N. (2017). Effects of built environment and weather on bike sharing demand: a station level analysis of commercial bike sharing in Toronto. *Transportation*, 44(3), 589-613. <https://doi.org/10.1007/s11116-015-9669-z>
- Eren, E., & Uz, V. E. (2020). A review on bike-sharing: The factors affecting bike-sharing demand. *Sustainable Cities and Society*, 54, Article 101882. <https://doi.org/10.1016/j.scs.2019.101882>
- Faghih-Imani, A., Hampshire, R., Marla, L., & Eluru, N. (2017). An empirical analysis of bike sharing usage and rebalancing: Evidence from Barcelona and Seville. *Transportation Research Part a-Policy and Practice*, 97, 177-191. <https://doi.org/10.1016/j.tra.2016.12.007>
- Feng, P. Y., & Li, W. Q. (2016). Willingness to use a public bicycle system: An example in Nanjing city. *Journal of Public Transportation*, 19(1), 84-96. <https://doi.org/10.5038/2375-0901.19.1.6>
- Fishman, E. (2016). Bikeshare: A review of recent literature. *Transport Reviews*, 36(1), 92-113. <https://doi.org/10.1080/01441647.2015.1033036>
- Fishman, E., Washington, S., Haworth, N., & Watson, A. (2015). Factors influencing bike share membership: An analysis of Melbourne and Brisbane. *Transportation Research Part a-Policy and Practice*, 71, 17-30. <https://doi.org/10.1016/j.tra.2014.10.021>
- Fortes, L. M., Giannotti, M., & de Freitas, F. S. (2024). Gender, class and race uneven access to bike systems across five Brazilian cities. *Cities*, 148, Article 104822. <https://doi.org/10.1016/j.cities.2024.104822>
- Galatoulas, N. F., Genikomsakis, K. N., & Ioakimidis, C. S. (2020). Spatio-temporal trends of E-bike sharing system deployment: A review in Europe, North America and Asia. *Sustainability*, 12(11), Article 4611. <https://doi.org/10.3390/su12114611>

- Gao, F., Tang, J. J., & Li, Z. T. (2022). Effects of spatial units and travel modes on urban commuting demand modeling. *Transportation*, 49(6), 1549-1575. <https://doi.org/10.1007/s11116-021-10219-y>
- Gao, K., Li, A. Y., Liu, Y., Gil, J., & Bie, Y. M. (2023). Unraveling the mode substitution of dockless bike-sharing systems and its determinants: A trip level data-driven interpretation. *Sustainable Cities and Society*, 98, Article 104820. <https://doi.org/10.1016/j.scs.2023.104820>
- Gao, K., Yang, Y., Li, A. Y., & Qu, X. B. (2021). Spatial heterogeneity in distance decay of using bike sharing: An empirical large-scale analysis in Shanghai. *Transportation Research Part D-Transport and Environment*, 94, Article 102814. <https://doi.org/10.1016/j.trd.2021.102814>
- Goodman, A., & Cheshire, J. (2014). Inequalities in the London bicycle sharing system revisited: impacts of extending the scheme to poorer areas but then doubling prices. *Journal of Transport Geography*, 41, 272-279. <https://doi.org/10.1016/j.jtrangeo.2014.04.004>
- Guo, Y. Y., Yang, L. C., & Chen, Y. (2022). Bike share usage and the built environment: A review. *Frontiers in Public Health*, 10, Article 848169. <https://doi.org/10.3389/fpubh.2022.848169>
- Hail, Y., & McQuaid, R. (2021). The Concept of Fairness in Relation to Women Transport Users. *Sustainability*, 13(5), Article 2919. <https://doi.org/10.3390/su13052919>
- He, M. J., Ma, X. W., Wang, J. B., & Zhu, M. Q. (2022). Geographically weighted multinomial logit models for modelling the spatial heterogeneity in the bike-sharing renting-returning imbalance: A case study on Nanjing, China. *Sustainable Cities and Society*, 83, Article 103967. <https://doi.org/10.1016/j.scs.2022.103967>
- Heinen, E., Kamruzzaman, M., & Turrell, G. (2018). The public bicycle-sharing scheme in Brisbane, Australia: Evaluating the influence of its introduction on changes in time spent cycling amongst a middle- and older-age population. *Journal of Transport & Health*, 10, 56-73. <https://doi.org/10.1016/j.jth.2018.07.003>

- Hossain, S., Loa, P., Ong, F., & Habib, K. N. (2023). Exploring the spatiotemporal factors affecting bicycle-sharing demand during the COVID-19 pandemic. *Transportation*. <https://doi.org/10.1007/s11116-023-10378-0>
- Ji, S. J., Heinen, E., & Wang, Y. Q. (2023). Non-linear effects of street patterns and land use on the bike-share usage. *Transportation Research Part D-Transport and Environment*, 116, Article 103630. <https://doi.org/10.1016/j.trd.2023.103630>
- Jin, S. T., & Sui, D. Z. (2024). Bikesharing and equity: A nationwide study of bikesharing accessibility in the U.S. *Transportation Research Part a-Policy and Practice*, 181, Article 103983. <https://doi.org/10.1016/j.tra.2024.103983>
- Kim, K. (2023). Discovering spatiotemporal usage patterns of a bike-sharing system by type of pass: a case study from Seoul. *Transportation*. <https://doi.org/10.1007/s11116-023-10371-7>
- Kim, M., & Cho, G. H. (2023). The effects of bike-share users' socio-demographics and trip features on the bike-transit relationships. *International Journal of Sustainable Transportation*, 17(8), 897-910. <https://doi.org/10.1080/15568318.2022.2119903>
- Le, H. T. K., West, A., Quinn, F., & Hankey, S. (2019). Advancing cycling among women: An exploratory study of North American cyclists. *Journal of Transport and Land Use*, 12(1), 355-374. <https://doi.org/10.5198/jtlu.2019.1461>
- Lee, M., Hwang, S., Park, Y., & Choi, B. (2022). Factors affecting bike-sharing system demand by inferred trip purpose: Integration of clustering of travel patterns and geospatial data analysis. *International Journal of Sustainable Transportation*, 16(9), 847-860. <https://doi.org/10.1080/15568318.2021.1943076>
- Li, A. Y., Zhao, P. X., Huang, Y. Z., Gao, K., & Axhausen, K. W. (2020). An empirical analysis of dockless bike-sharing utilization and its explanatory factors: Case study from Shanghai, China. *Journal of Transport Geography*, 88, Article 102828. <https://doi.org/10.1016/j.jtrangeo.2020.102828>
- Lin, P. F., Weng, J. C., Hu, S., Alivanistos, D., Li, X., & Yin, B. C. (2020). Revealing Spatio-Temporal Patterns and Influencing Factors of Dockless Bike Sharing

- Demand. *Ieee Access*, 8, 66139-66149. <https://doi.org/10.1109/access.2020.2985329>
- Ma, X. W., Ji, Y. J., Yuan, Y. F., Van Oort, N., Jin, Y. C., & Hoogendoorn, S. (2020). A comparison in travel patterns and determinants of user demand between docked and dockless bike-sharing systems using multi-sourced data. *Transportation Research Part a-Policy and Practice*, 139, 148-173. <https://doi.org/10.1016/j.tra.2020.06.022>
- Machado, C. A. S., Hue, N., Berssaneti, F. T., & Quintanilha, J. A. (2018). An overview of shared mobility. *Sustainability*, 10(12), Article 4342. <https://doi.org/10.3390/su10124342>
- Meng, S. A., & Zacharias, J. (2021). Street morphology and travel by dockless shared bicycles in Beijing, China. *International Journal of Sustainable Transportation*, 15(10), 788-798. <https://doi.org/10.1080/15568318.2020.1813853>
- Mizdrak, A., Tatah, L., Mueller, N., Shaw, C., & Woodcock, J. (2023). Assessing the health impacts of changes in active transport: An updated systematic review. *Journal of Transport & Health*, 33, Article 101702. <https://doi.org/10.1016/j.jth.2023.101702>
- Murphy, E., & Usher, J. (2015). The role of bicycle-sharing in the City: Analysis of the Irish experience. *International Journal of Sustainable Transportation*, 9(2), 116-125. <https://doi.org/10.1080/15568318.2012.748855>
- Pearson, L., Reeder, S., Gabbe, B., & Beck, B. (2023). What a girl wants: A mixed-methods study of gender differences in the barriers to and enablers of riding a bike in Australia. *Transportation Research Part F-Traffic Psychology and Behaviour*, 94, 453-465. <https://doi.org/10.1016/j.trf.2023.03.010>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12, 2825-2830.
- Ricci, M. (2015). Bike sharing: A review of evidence on impacts and processes of implementation and operation. *Research in Transportation Business and Management*, 15, 28-38. <https://doi.org/10.1016/j.rtbm.2015.03.003>

- Ross-Perez, A., Walton, N., & Pinto, N. (2022). Identifying trip purpose from a dockless bike-sharing system in Manchester. *Journal of Transport Geography*, 99, Article 103293. <https://doi.org/10.1016/j.jtrangeo.2022.103293>
- Schimohr, K., Doeblner, P., & Scheiner, J. (2023). Prediction of bike-sharing trip counts: Comparing parametric spatial regression models to a geographically weighted XGBoost algorithm. *Geographical Analysis*, 55(4), 651-684. <https://doi.org/10.1111/gean.12354>
- Shi, L., Halik, Ü., Abliz, A., Mamat, Z., & Welp, M. (2020). Urban Green Space Accessibility and Distribution Equity in an Arid Oasis City: Urumqi, China. *Forests*, 11(6), Article 690. <https://doi.org/10.3390/f11060690>
- Sun, Z. Y., Wang, Y. C., Zhou, H. Y., Jiao, J. J., & Overstreet, R. E. (2021). Travel behaviours, user characteristics, and social-economic impacts of shared transportation: a comprehensive review. *International Journal of Logistics-Research and Applications*, 24(1), 51-78. <https://doi.org/10.1080/13675567.2019.1663162>
- Teixeira, J. F., Silva, C., & Sa, F. M. E. (2021). Empirical evidence on the impacts of bikesharing: a literature review. *Transport Reviews*, 41(3), 329-351. <https://doi.org/10.1080/01441647.2020.1841328>
- Tu, Y. J., Chen, P., Gao, X., Yang, J. W., & Chen, X. H. (2019). How to make dockless bikeshare good for cities: Curbing oversupplied bikes. *Transportation Research Record*, 2673(6), 618-627. <https://doi.org/10.1177/0361198119837963>
- Wang, K. L., Akar, G., & Chen, Y. J. (2018). Bike sharing differences among Millennials, Gen Xers, and Baby Boomers: Lessons learnt from New York City's bike share. *Transportation Research Part a-Policy and Practice*, 116, 1-14. <https://doi.org/10.1016/j.tra.2018.06.001>
- Wang, R. X., Wu, J. P., & Qi, G. Q. (2022). Exploring regional sustainable commuting patterns based on dockless bike-sharing data and POI data. *Journal of Transport Geography*, 102, Article 103395. <https://doi.org/10.1016/j.jtrangeo.2022.103395>
- Wang, Y. C., Zhan, Z. L., Mi, Y. H., Sobhani, A., & Zhou, H. Y. (2022). Nonlinear effects of factors on dockless bike-sharing usage considering grid-based

- spatiotemporal heterogeneity. *Transportation Research Part D-Transport and Environment*, 104, Article 103194. <https://doi.org/10.1016/j.trd.2022.103194>
- Woodcock, J., Tainio, M., Cheshire, J., O'Brien, O., & Goodman, A. (2014). Health effects of the London bicycle sharing system: health impact modelling study. *Bmj-British Medical Journal*, 348, Article g425. <https://doi.org/10.1136/bmj.g425>
- Wu, C. L., Kim, I., & Chung, H. (2021). The effects of built environment spatial variation on bike-sharing usage: A case study of Suzhou, China. *Cities*, 110, Article 103063. <https://doi.org/10.1016/j.cities.2020.103063>
- Xiao, W. Y., & Wei, Y. D. (2023). Assess the non-linear relationship between built environment and active travel around light-rail transit stations. *Applied Geography*, 151, Article 102862. <https://doi.org/10.1016/j.apgeog.2022.102862>
- Xing, Y. Y., Wang, K., & Lu, J. J. (2020). Exploring travel patterns and trip purposes of dockless bike-sharing by analyzing massive bike-sharing data in Shanghai, China. *Journal of Transport Geography*, 87, Article 102787. <https://doi.org/10.1016/j.jtrangeo.2020.102787>
- Xu, Y., Chen, D. C., Zhang, X. H., Tu, W., Chen, Y. Y., Shen, Y., & Ratti, C. (2019). Unravel the landscape and pulses of cycling activities from a dockless bike-sharing system. *Computers Environment and Urban Systems*, 75, 184-203. <https://doi.org/10.1016/j.compenvurbsys.2019.02.002>
- Yang, H. T., Zhang, Y. B., Zhong, L. Z., Zhang, X. J., & Ling, Z. W. (2020). Exploring spatial variation of bike sharing trip production and attraction: A study based on Chicago's Divvy system. *Applied Geography*, 115, Article 102130. <https://doi.org/10.1016/j.apgeog.2019.102130>
- Zhang, Y. R., & Kamargianni, M. (2023). A review on the factors influencing the adoption of new mobility technologies and services: autonomous vehicle, drone, micromobility and mobility as a service. *Transport Reviews*, 43(3), 407-429. <https://doi.org/10.1080/01441647.2022.2119297>
- Zhu, L. D., Ali, M., Macioszek, E., Aghaabbasi, M., & Jan, A. (2022). Approaching sustainable bike-sharing development: A systematic review of the influence of built environment features on bike-sharing ridership. *Sustainability*, 14(10), Article 5795. <https://doi.org/10.3390/su14105795>

Zhuang, C. G., Li, S. Y., Tan, Z. Z., Feng, G., & Wu, Z. F. (2022). Nonlinear and threshold effects of traffic condition and built environment on dockless bike sharing at street level. *Journal of Transport Geography*, 102, Article 103375. <https://doi.org/10.1016/j.jtrangeo.2022.103375>

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

I, Xiao Cai,
(*author's name*)

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 in bike-sharing usage among varying demographic groups and its
association with station-surrounding characteristics in Tartu, Estonia,
(*title of thesis*)

supervised by

Siiri Silm, Amnir Hadachi.
(*supervisor's name*)

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.

Xiao Cai
14/06/2024