GDAŃSK UNIVERSITY OF TECHNOLOGY

FACULTY OF ELECTRONICS, TELECOMMUNICATIONS AND INFORMATICS

Student's name and surname: Olga Paszkiewicz
Cycle of studies: bachelor's degree studies
Mode of study: full-time studies
Interfaculty field of study: Data Engineering
realized at: Faculty of Electronics,Telecommunications and Informatics; Faculty of Management and Economics
Profile: Intelligent data processing

Student's name and surname: Aleksandra Susmarska
Cycle of studies: bachelor's degree studies
Mode of study: full-time studies
Interfaculty field of study: Data Engineering
realized at: Faculty of Electronics,Telecommunications and Informatics; Faculty of Management and Economics
Profile: Intelligent data processing

ENGINEERING DIPLOMA THESIS

Title of the thesis: Analysis of Factors Affecting the Number of Rentals in the MEVO Metropolitan Bicycle System

Title of the thesis (in Polish): Analiza czynników mających wpływ na ilość wypożyczeń w Systemie Rowerów Metropolitalnych MEVO

Supervisor: dr inż. Joanna Raczek

Abstract

The aim of this paper is to help in the development of the MEVO bike-sharing system by analysing user behaviour and identifying factors influencing station usage, as well as by proposing optimal locations for new stations. We have analysed the available data using Python to examine station flows, temporal trends and the influence of chosen factors on station popularity. Based on this analysis we studied the impact of nearby infrastructure on the popularity of stations and chose 336 potential new MEVO station locations, increasing the availability of MEVO in residential areas by 13.29%. The complicated nature of the problem and outdated data sources limit the accuracy of the results. However, it can still serve as a framework for more detailed studies in the future and support strategic system expansion.

Keywords: bike-sharing system, MEVO, urban mobility, bicycle rentals, transport, spatial analysis, station location optimisation, electric bicycles

Field of science and technology according to OECD requirements: Engineering and Technology, Transport Engineering

Streszczenie

Celem niniejszej pracy jest pomoc w rozwoju systemu rowerów miejskich MEVO poprzez analizę zachowań użytkowników i identyfikację czynników wpływających na korzystanie ze stacji, a także poprzez zaproponowanie optymalnych lokalizacji nowych stacji. Przeanalizowaliśmy dostępne dane za pomocą języka Python, aby zbadać przepływy na stacjach, trendy czasowe i wpływ wybranych czynników na popularność stacji. Na podstawie tej analizy zbadaliśmy wpływ pobliskiej infrastruktury na popularność stacji i wybraliśmy 336 potencjalnych lokalizacji nowych stacji MEVO, zwiększając dostępność MEVO na obszarach mieszkalnych o 13.29 %. Skomplikowany charakter problemu i nieaktualne źródła danych ograniczają dokładność wyników. Mogą one jednak nadal służyć jako podstawa do bardziej szczegółowych badań w przyszłości i wspierać strategiczną rozbudowę systemu.

Contents

Abstract **2**

Contents **3**

List of Important Symbols and Abbreviations **6**

1 Introduction **7**

1.1 Topic 7

1.2 MEVO System 8

1.3 Problem Statement 8

1.4 Aim and Objectives 8

1.5 Scope and Limitations 9

1.5.1 Area of Interest 9

1.5.2 Characteristics of the Area 9

1.5.3 Variables Included and Excluded 9

1.6 Thesis Structure 10

2 Theoretical Foundations and Methodology **11**

2.1 Choosing Factors for Analysis 11

2.1.1 Location Characteristics 11

2.1.2 Temporal and Weather Influence 11

2.2 Points of Interest (POIs) 12

2.2.1 Deciding on Categorisation 12

2.2.2 Effect of Distance to POIs 12

2.3 Data Analysis Methods and Visualisation 13

2.4 Regression Models 13

2.4.1 Overview of Different Regression Models 13

2.4.2 Equations and Further Explanation of Chosen Regression Models 14

2.5 Tools and Environment 15

3 Data **16**

3.1 MEVO Data 16

3.1.1 Raw Data Description 16

3.1.2 Data Processing and Cleaning 16

3.2 External Data Sources 20

3.2.1 List of Data Sources 20

3.3 Points of Interest Data 20

3.3.1 Getting POIs for Analysis 20

3.4 Population Density 21

3.5 Data Processing and Cleaning 21

4 Data Analysis - Exploratory Analysis and Regression Modelling **22**

4.1 Information about Average Number of Bike Rentals per Station 22

4.2 Route Analysis 23

4.2.1 Internal Bike Trips 23

4.2.2 Relationship with Nearby Cities 29

4.3 Average Number of Bikes by Gdańsk Districts 31

4.4 Bike Relocation by Gdańsk Districts 33

4.5 Temporal Analysis and Weather 37

4.6 POI Analysis 39

4.6.1 Correlations 39

4.6.2 Regression 42

5 Geographic Information Systems (GIS) Analysis **49**

5.1 Influence Distance Decay of Chosen Factors 49

5.2 Factor Weight Assignment 51

5.3 Spatial Suitability Score 53

5.4 Locations for New MEVO Stations 55

5.5 MEVO Availability 57

6 Conclusions **59**

Bibliography **60**

A OSM Subcategory Mapping **62**

List of Important Symbols and Abbreviations

Abbreviation	Explanation
OLS	Ordinary Least Squares
GWR	Geographically weighted regression
SVC	Spatially varying coefficient
SAR	Spatial Autoregression
OSM	OpenStreetMap
OMGGS	Metropolitan Area Gdańsk Gdynia Sopot
BSS	Bike-sharing system
POI	Point of Interest
GIS	Geographic Information System

1. Introduction

Author: Olga Paszkiewicz

1.1 Topic

In recent years, there has been a noticeable development in the area of shared mobility services, notably within larger cities [11]. This trend can be attributed to the growing need for green alternatives in both public and individual transport. While research shows that bike-share often replaces walking or public transport trips rather than car trips [8], studies also indicate that neighbourhood-level implementations of bike-sharing systems (BSS) can reduce traffic congestion by up to 4% [2].

Bike-sharing systems can be particularly appealing to users, as they often offer both traditional (non-electric) and electric vehicles. Electric bikes allow users to travel farther with less effort, while traditional bikes offer a means of physical activity. Both types contribute to the general health of residents [1, 8]. In addition, micromobility transport, including bike-sharing systems, makes it easier to reach places where travel has been obstructed.

All of this contributes to the success of BSS in many cities across Europe, as well as in Poland. Growth has become especially prominent in large Polish cities, such as Warsaw, Kraków, and the Tricity metropolitan area [1] [3]. The MEVO Metropolitan Bicycle System, based in Tricity, is one of the largest and most developed systems. It features a network of docking stations, electric bikes, and a mobile application that simplifies bike rental.

The development of BSS also brings several challenges. These include optimising bike-station locations, ensuring adequate bicycle availability, maintaining user safety, and preserving city aesthetics by preventing bikes from being left in inappropriate places. Studies suggest that a 10% increase in bike availability can raise ridership by approximately 12.2% [5]. To plan, operate, and integrate these systems effectively with other forms of transport, it is necessary to examine the factors that affect the functioning of bike stations.

In this thesis, we study the factors that affect the number of rentals in the MEVO Metropolitan Bicycle System. Based on prior research we assume that station location, population density, and distance to points of interest (POIs) influence how often bikes are rented. We also look at cycling activity inside Gdańsk and to nearby cities, as well as weather and time factors.

[1] The Tricity metropolitan area consists primarily of three cities: Gdańsk, Gdynia, and Sopot.

1.2 MEVO System

MEVO is one of the largest and most advanced bike-sharing systems in Central and Eastern Europe. It is also the first system in Poland that works all year round [4]. Its goal is to give residents and visitors a convenient, flexible and environmentally friendly way of transport.

MEVO uses bike stations with different amounts of bikes, many of which have electric motors and batteries of specific capacity. Additionally bikes are equipped with a location module, tracking their movement. Location module allows for monitoring of vehicle availability and efficient planning related to relocation and recharging the bikes.

The systems operate in a station-based and semi-station-based style. Bicycles can be returned in designated zones (bike stations) or outside of them (FreeFloat), albeit with a penalty for a user.

1.3 Problem Statement

The MEVO system have several issues that hinder their effectiveness, and may raise their operating costs. These include:

- uneven distribution of stations throughout the city, causing inefficiency,
- insufficient number of bicycles in stations, and too high number at others, compared to demand,
- difficulties in relocating bikes in accordance with need.

As a result, the residents' ability to rent bicycles is negatively impacted. This limits how much the system can be used and may slow down its wider adoption in the city.

1.4 Aim and Objectives

The aim of this study is to analyse the factors that affect the number of rentals in the MEVO system. Consequently, the focus is on how bikes are distributed across stations.

The following are the main tasks:

- examining the current MEVO station network and determining the factors that affect station performance,
- identifying stations with inadequate number of bikes and the routes which contribute to that.

The obtained result will:

- show areas with high demand and the factors that contribute to it,
- lay groundwork towards improving the bike relocation process.

The goal of this work is to give a clear analysis of the current MEVO system, which can be used for further research and further system development.

1.5 Scope and Limitations

1.5.1 Area of Interest

The main area of focus is the city of Gdańsk with 388 stations. We also examine trips that leave the area, particularly to Sopot and ones without designated final station ("FreeFloating").

The data covers the period from April 1st to October 31st in 2024. This period excludes the coldest months, instead focusing on the months that constitute Poland's main cycling season. The selected time range accounts for the fact that more people will be using services during the spring and summer, thanks to:

- longer daylight hours and favourable weather conditions,
- the tourist season, when the number of city events increases.

1.5.2 Characteristics of the Area

Gdańsk is a large metropolitan centre with varied urban structure, which allows for detailed analysis. The structure of the city includes:

- densely built busy residential areas, packed with office and business districts,
- parks and recreational spaces,
- an extensive system of public transportation.

1.5.3 Variables Included and Excluded

The main focus of the analysis is examining the relationship between the number of rentals and the following factors:

- population density,
- length and availability of cycling infrastructure,
- availability of public transport (tram, bus, train),
- presence of Points of Interest (POI) in the area of the station - the choice of such, explained in Section 3.3.

The analysis does not consider individual or operational factors, such as user type, personal behaviour, or equipment failures. Therefore, rather than concentrating on user preferences, the work looks on infrastructure factors.

1.6 Thesis Structure

This thesis is structured as follows:

Chapter	**Content**
Abstract	Short summary of goals, methods, and main results.
Contents	List of all chapters and sections.
List of Symbols and Abbreviations	Description of all abbreviations and symbols used.
Chapter 1: Introduction	Research context, MEVO system overview, aims, scope, and thesis outline.
Chapter 2: Theoretical Foundations and Methodology	Theory background and explanation of analytical methods.
Chapter 3: Data	Data sources, preprocessing, cleaning, and preparation steps.
Chapter 4: Data Analysis- Exploratory Analysis and Regression Modelling	Presentation and discussion of exploratory data analysis results and regression modelling results.
Chapter 5: GIS Analysis	Spatial analysis and maps based on GIS tools.
Chapter 6: Conclusions	Summary of findings and recommendations.
Bibliography	List of all references used, and links to external data sources.
Appendix	Additional data - explanation which tags in OSM correspond to our POI categories.

2. Theoretical Foundations and Methodology

Author: Olga Paszkiewicz, Aleksandra Susmarska

In this chapter, we describe the main concepts used in thesis. Not all factors listed in the literature are analysed in detail in this thesis. We focus only on those most relevant to MEVO and the available data.

2.1 Choosing Factors for Analysis

2.1.1 Location Characteristics

We based our factors choice on a review by Ezgi Eren and Volkan Emre Uz [10]. The authors gathered together findings from multiple empirical studies on BSS. The work finds that the number of rentals is influenced by:

- The land structure - steeper areas have lower bike use, slope is one of the most important natural factors,
- Population and employment density, mixed land use areas - places with more residents or workplaces have higher bike-share demand, additionally mixed land use also increases demand (depending on the local land structure),
- Proximity to major POIs - stations near shopping centres, schools or hotels experience difference in number of rentals compared to stations, which location is not close to such places,
- Urban features - bicycle paths, green areas density, office space density, street lighting, and natural features such as lakesides or seaside,
- Station characteristics - station capacity, or number of docks and racks.

These findings guided our selection of variables used in this thesis.

2.1.2 Temporal and Weather Influence

Studies indicate differences between the weekday and weekend usages of BSS [8]. Therefore, we briefly analyse the usage for these periods separately.

In addition previous studies have shown that weather conditions have significant influence on bike use. In particular, moderate temperatures (depending on the study ranging from 10–30, 20–30, or 30–35 °C) [5, 10] are associated with the highest ridership levels. Temperatures above 30 °C shows either positive or negative effects depending on the local climate, while very low temperatures reduce demand [10].

Adverse conditions such as rain, snow, high humidity, or strong winds reduce number of rentals even more. Precipitation is leading factor to the strongest short-term declines. [10].

Because of this, we include basic weather information in our analysis.

2.2 Points of Interest (POIs)

2.2.1 Deciding on Categorisation

To find locations where demand for rentals will be higher it is important to acquire and categorise relevant factors into separate groups [7, 10, 11, 13].

Previous studies have analysed categories of POIs such as parks, shopping centres, schools, universities, sports facilities, museums, and hotels affect bike-sharing demand [10].

As proposed in the paper [7], we take both activity-based groups and residential buildings into consideration. As study suggests we split residential buildings into smaller (houses) and large residential categories (apartments). We decide on the activity POIs based on different trip purposes and on existing literature.

During deciding on POI categories, we consulted the OpenStreetMap wiki [16].

2.2.2 Effect of Distance to POIs

To study how nearby locations influence bike rentals it is important to decide how the distance is taken into consideration.

Some studies use distance-decay functions [5], while others use Thiessen polygons (also called Voronoi tessellation) [11]. In our analysis, w decided to focus on distance-decay function and buffers around each station. This allow some POIs to overlap between stations.

We base our definition of influence function on a study of BSS in different cities [5]. It looked at POIs in different categories, and found that:

- For distances below 300 m likelihood of using a bike gets decreased by 0.194 % for every meter. About 80 % of rentals come from this range.
- For distances between 300-500 m, the likelihood decreases by 1.307 % per meter.

We decided to cut-off all the POIs above the distance of 500 m.

Other research [9] supports the 500 m cut-off, showing that residents living within 500 m of a station are 3.2 times more likely to use bike-share.

Thus, we define the influence function as:

$$I_{\text{POI}}(d) = \begin{cases} (1 - 0.00194)^d, & d \leq 300, \\ (1 - 0.00194)^{300}(1 - 0.01307)^{(d-300)}, & 300 < d \leq 500, \\ 0, & d > 500, \end{cases} \tag{2.1}$$

where:

- $I_{\text{POI}}(d)$ - the influence of a point of interest at distance d,
- d - the distance in meters from the POI to the bike station.

We chose to use different influence function for cycleways. While analysing the effect of distance to cycleways, it has been found [6] that in Montreal and Vancouver each 1 km decrease to a Lakeway was associated with a 3.92-4.44 times higher number of bike commuters.

Therefore, basing on the above we take the influence of cycleways as:

$$I_{\text{cycleway}}(d) = \left(\frac{1}{3.91}\right)^{d/1000}, \tag{2.2}$$

where:

- $I_{\text{cycleway}}(d)$ - the influence of a cycleway at distance d,
- d - the distance in meters from the cycleway to the bike station.

2.3 Data Analysis Methods and Visualisation

To study the factors affecting bike-sharing demand, we used several exploratory analyses:

- **Distributions:** We used histograms, boxplots and scatterplots to look at bike demand.
- **Correlations:** We calculated Pearson and Spearman correlations to see how distance to POIs relate to demand.
- **Maps:** We made maps and spatial plots to show how demand and bike routes are spread across the network. We also used them to visualise POIs, their influence, and population density.

2.4 Regression Models

2.4.1 Overview of Different Regression Models

Regression modelling was used in numerous studies to examine the relationship of different factors to bicycle demand.

The main types of models applied in the literature [11, 12] include:

- Global linear regression - estimates one set of coefficients for all stations. It assumes the relationship between factors and bike rentals is the same everywhere (model does not consider space). Example of such is Ordinary Least Squares (OLS).
- Spatial models (SAR) - includes demand at nearby stations (additional predictor) when calculating the demand for a station. It still uses the same coefficients for factors globally (one set).
- Geographically Weighted Regression (GWR) - an example of Spatially Varying Coefficients (SVC) model. It calculates local coefficients for each station using distance-based weights. So, the model captures how factors affect each station differently. The limitation of such method is scalability.
- Graph-regularized SVC regression - calculates set of coefficients for each station. Nearby stations are linked through graph, which applies spatial smoothing. Such method is scalable, and allows prediction of coefficients for new stations.

In our research, we use the GWR model and compare its results to those from OLS, although if possible in future research it would be interesting to apply Graph-regulized SVC regression.

2.4.2 Equations and Further Explanation of Chosen Regression Models

Ordinary Least Squares (OLS) is a basic regression method, it estimates the linear relationship [13].

The OLS model can be written as:

$$y_i = \beta_0 + \sum_{k=1}^{K} \beta_k x_{ik} + \varepsilon_i, \tag{2.3}$$

where:

- y_i is the observed bike-sharing demand at station i,
- x_{ik} represents the k-th explanatory variable,
- β_k are the global regression coefficients (constant across the study area), and
- ε_i is the error term.

Geographically Weighted Regression (GWR) is a regression method that considers space. It allows the effect of each factor to change from one location to another [12, 13].

The GWR model can be written as:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{K} \beta_k(u_i, v_i) x_{ik} + \varepsilon_i, \tag{2.4}$$

where:

- y_i is the observed bike-sharing demand at station i,
- x_{ik} represents the k-th explanatory variable, and
- (u_i, v_i) denote the geographic coordinates of station i,
- the coefficients $\beta_k(u_i, v_i)$ vary smoothly over space (local relationships) [12, 13],
- ε_i is the error term.

2.5 Tools and Environment

Below are list tools we used for the analysis:

- QGIS 3.44.1: for general analysis of geographic data, with GRASS 8.4.1 used to find optimal locations for new stations
- Python 3.13.9 - for data processing, statistical analysis, and visualisation. Key Python libraries included:
  - **GeoPandas**, **Shapely**, **OSMnx**, **folium** - for using spatial data,
  - **Pandas**, **NumPy** - for data manipulation and numerical calculations.
  - **Matplotlib**, **Seaborn** - for creating plots and maps,
  - **SciPy**, **Statsmodels**, **scikit-learn**, **mgwr** - for statistical tests and regression modelling,
  - **Holidays**, **Requests** – for calendar data, and data retrieval from APIs.

3. Data

Author: Olga Paszkiewicz, Aleksandra Susmarska

3.1 MEVO Data

3.1.1 Raw Data Description

The data we obtained from the Metropolitan Area Gdańsk Gdynia Sopot (OMGGS) is made up of 13 CSV files, each capturing different aspects of bike trips. Table 3.1 summarises the raw datasets.

Inside the data there is information about 388 stations inside Gdańsk.

For routes there is 388 origin Gdańsk stations, and destination numbers are as described in Table 3.2.

Two stations in Gdańsk had bikes taken from them, but 0 returned.

Additional destination taken into consideration is FreeFloat (meaning not returned to the station).

Not all stations contained in the dataset were considered. The locations (stations) excluded from the dataset were: OMGGS location,"Ed Sheeran 1" and "Ed Sheeran 2" stations, "Stacja Iron Man", and "Opener Gdynia".

The raw data varied in structure and level of detail, including per-bike trip logs, aggregated trips per station pair, and daily or hourly summaries.

The files connected to specific bike routes contained overlapping information, which required careful examination. Some of the datasets were structured in a way that needed to be changed for further analysis.

All data is from the time period described previously: April 1st to October 31st in 2024.

3.1.2 Data Processing and Cleaning

From the raw dataset, we constructed several data frames, prepared specifically for further analysis. They were prepared by performing the following:

- **Data cleaning and validation**: We identified missing values, data was analysed to find what it described and validated for its usefulness. Header rows that did not store actual observations were removed, and string values were converted to numeric types if applicable.
- **Column renaming and redefining**: We renamed the columns so that they described the

Table 3.1: Overview of the raw datasets.

File	Description
Czestotliwosc_wypozyczania_rowerow.csv	Hourly rental frequency per station (bikes rentals per hour), includes both trips within Gdańsk and rentals from outside
Dystans_na_podroz_dziennie.csv	Total kilometers traveled by bike per day in Gdańsk (year, month, day → total distance)
Dzienne_wycieczki.csv	Total number of trips made daily for Gdańsk (year, month, day → number of trips)
Liczba_podrozy_na_użytkownika_rozklad_procentowy.csv	Hourly percentage of trips per user over the day (hours are additionally split into bins), for Gdańsk
Liczba_podrozy_na_uzytkownika.csv	Shows how users distribute their trips throughout the day (in Gdańsk) for each station – number of trips per user for specific hour
Liczba_rowerow_wypozyczonych_na_stacji.csv	Bikes rented (picked up) from each station – total over the time period, for Gdansk
Liczba_wypozyczonych_rowerow_ze_stacji_na_godzine_procentowo.csv	Hourly percentage of bike rentals over the day (hours are additionally split into bins), for Gdansk
Liczba_wypozyczonych_rowerow_ze_stacji_rozklad_godzinowy.csv	Hour-by-hour rental counts per station, for Gdańsk
Liczba_zwroconych_rowerow_na_stacji.csv	Number of bikes returned to each station – total over the time period, for Gdansk
Podroze_na_rower_na_dzien.csv	Daily trips per bike (year, month, day → number of bike trips per bike), for Gdansk
Podroze_wedlug_tras_i_kierunkow_liczba.csv	Trips by route, contains the loops, only Gdansk → Gdansk trips (start_station → end_station (path), number_of_trips)
Podroze_wedlug_tras_i_kierunkow.csv	Trips by route, contains overlaping data with above, but additionally trips outside of Gdansk and FreeFloating (start_station → end_station (path), number_of_trips)
Przejazdy_wedlug_stacji_poczatkowej_i_koncowej.csv	Origin–destination adjacency matrix, rows represent departure points and columns store aggregated arrivals
Stacja_poczatkowa_i_docelowa_kazdej_wycieczki_rowerowej.csv	Individual bike trip logs (bike ID, path, distance, trip count)

actual information. Some columns were split so that they will be used more easily.

- **Dates transformation**: We changed data format for datasets storing daily information, from year/month/day fields into a single `date` field using datetime conversion.

  Feature engineering:

- **Station statistics merging**: The datasets with information about pickup and return counts were combined into a single data frame through relational joins (by station name). In the resulting table net bicycle flow per station was calculated. This dataset was expanded by adding hourly rental averages and geographic data (latitude, longitude, address attributes).

- **Route dataset changes**: Two datasets with overlapping data about number of trips made

Table 3.2: Numbers of destination stations.

destination	number of stations
Gdańsk	386
Gdynia	133
Pruszcz Gdański	29
Sopot	28
Rumia	12
Tczew	10
Kosakowo	9
Reda	6
Kolbudy	5
Władysławowo	5
Żukowo	5
Kartuzy	4
Puck	3
Somonino	2
Stegna	1

on specific path were analysed for what data they specifically contained, and joined to created data frame. We added additional columns such as a column describing which paths were loops, or if the destination was outside or inside Gdańsk.

The resulting data frames are summarised in Table 3.3 and were constructed as described below:

- `df_hours_trip_per_person`: Hourly user trip counts by station, got from `Liczba_podrozy_na_uzytkownika.csv`, cleaned and numeric-typed.
- `df_hours_rented_vehicles`: Number of rented bikes per hour, got from `Liczba_wypozyczonych_rowerow_ze_stacji_rozklad_godzinowy.csv`, cleaned and numeric-typed.
- `df_station_statistics`: Station-level dataset combining pickups, returns, net flow, hourly averages, and geographic metadata, got from `Liczba_rowerow_wypozyczonych_na_stacji.csv`, `Liczba_zwroconych_rowerow_na_stacji.csv`, and `Czestotliwosc_wypozyczania_rowerow.csv`. Geographic data was obtrained from the external `station_information.json` [19] described in the external data sources section.
- `df_trip_from_station_to_station`: Cleaned origin–destination adjacency matrix of trips between stations, got from `Przejazdy_wedlug_stacji_poczatkowej_i_koncowej.csv`.
- `df_daily_values`: Daily activity indicators including distance, total trips, and trips per bike, got from `Dystans_na_podroz_dziennie.csv`, `Dzienne_wycieczki.csv`, and `Podroze_na_rower_na_dzien.csv`.
- `df_hourly_bins_trips_rentals`: Hourly distributions of trips per user and bike rentals per station, joined using hour bins, got from `Liczba_podrozy_na_użytkownika_rozklad_`

procentowy.csv and Liczba_wypozyczonych_rowerow_ze_stacji_na_godzine_procentowo.csv.

- df_all_paths: Trip paths dataset got by merging Podroze_wedlug_tras_i_kierunkow.csv and Podroze_wedlug_tras_i_kierunkow_liczba.csv, including loops and trips outside Gdańsk.
- df_every_bike_trip: Individual bike trip logs, got from Stacja_poczatkowa_i_docelowa_kazdej_wycieczki_rowerowej.csv.

Table 3.3: Overview of the processed and resulting datasets used in the analysis.

Dataset	**Key Columns / Features**	**Description / Notes**
df_hours_trip_per_person	name, hours 00–23, Total	Hourly trips per person per station
df_hours_rented_vehicles	name, hours 00–23, Total	Hourly bike rentals per station
df_station_statistics	name, picked_up, returned, net_flow, average_bikes_per_hour, lat, lon, ...	Flows with average number of bikes per hour. Includes geographic coordinates, activity metrics, and peak hour metrics
df_trip_from_station_to_station	station name, stations GDA001–GDA388	Station-to-station flow matrix
df_daily_values	date, distance_km, trips, trips_per_bike, trips_7d_avg	Daily usage statistics
df_hourly_bins_trips_rentals	hour, hour_bin, trips_per_user, bikes_rented_pct	Hourly distribution of trips and rentals
df_all_paths	tps_path, count, trip_type, origin, destination, lat/lon	All bike paths
df_every_bike_trip	Bike_ID, Trip, Trips, Distance_km	Individual bike trip logs

After further analysis, the dataframe df_every_bike_trip was fully excluded from the study, as it contained non-informative data due to the small number of records and uncertainty about which stations they pertained to.

3.2 External Data Sources

3.2.1 List of Data Sources

External data sources used to prepare the analysis:

- **Gdańsk in Numbers (Gdańsk w liczbach)** [17]:
  Data that we used, concerning population of Gdańsk residents by administrative division (districts) was obtained from the file *04.1.1.D. Stan ludności - liczba mieszkańców oraz gęstość zaludnienia dzielnic Gdańska.*

- **OpenStreetMap (OSM)** [18]:
  OpenStreetMap was used to get geographic information about Points of Interest (POIs), their location and metadata. We also used OSM to obtain spatial data about cycleways and public transportation stops. Additionally geographical data missing from MEVO station information was added.

- **MEVO Open Data - Stations** [19]:
  Source used to obtain necessary information about all MEVO stations that was missing from internal data, including geolocation, addresses, and metadata. The station information was obtained through the endpoint `station_information.json`.

- **Gdańsk District Boundaries (Shapefile)** [20]:
  A shapefile containing the boundaries of Gdańsk districts.

- **Weather Data** [21]:
  We obtained the historical weather data (daily maximum temperature and total precipitation) for specified period from the Open-Meteo API for Gdańsk.

- **Geoportal gov** [22]:
  We used geoportal to obtain spatial data about types of roads and buildings, as well as voivodeship and city boundaries for clear data visualisation.

3.3 Points of Interest Data

3.3.1 Getting POIs for Analysis

POIs were extracted from OpenStreetMap (OSM) using the Python library `osmnx`. The city boundary was buffered by 1 km to include nearby areas. The queries were made for different OSM categories: shops, amenities, buildings, offices, leisure facilities, tourism sites, and public transport stations.

Each OSM feature was assigned a value for `subcategory`, which was used as the distinguishing feature for POIs. It was done according to a predefined mapping.

The mapping of OSM tags to these subcategories is provided in Appendix A.

Centroids were calculated for features represented by polygons or lines, and used as the POIs geometry. Invalid or empty geometries were removed.

The resulting dataset contains POIs categorised with their geographical locations, so that the analysis connected to their distance to stations can be conducted, it is stored in both CSV and GeoJSON formats.

3.4 Population Density

The population density was obtained from *Gdańsk w liczbach*, published by the City of Gdańsk in periodic reports. The provided datasets include information for each district of Gdańsk, for the year 2024 (status for: 31.12.2024).

3.5 Data Processing and Cleaning

Data was cleaned from missing values, and POIs were checked for overlap.

The initially created categories were updated to more descriptive ones, after primary analysis.

4. Data Analysis - Exploratory Analysis and Regression Modelling

Author: Olga Paszkiewicz

4.1 Information about Average Number of Bike Rentals per Station

In this section, we conducted an analysis only on the data connected to the average number of bike rentals for each station. It was done in order to examine the distribution of the primary variable.

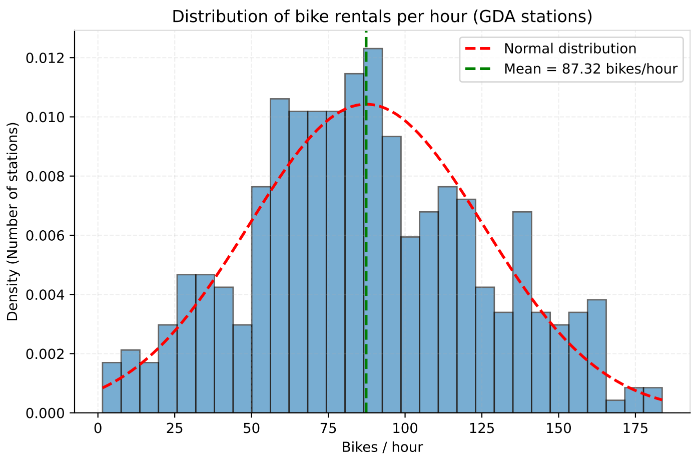

Figure 4.1: Histogram of bike trips in Gdańsk.

As shown at Figure 4.1 the data is fairly uniform, without an extreme skew towards very high or low station averages. The skewness is 0.154 and kurtosis -0.474, which supports this observation and suggests only light tails.

The majority of stations show moderate usage - 60.9, 85.8, and 113.4 bikes/hour for 25th, 50th, and 75th percentiles (Mean being 87.3 bikes per hour and a standard deviation of 38.3).

Formal normality tests we performed reject the null hypothesis of a normal distribution. The Shapiro-Wilk test yields W = 0.990, p = 0.013, and the Anderson-Darling test statistic is 0.839 (higher than 5 % critical value: 0.779).

We also checked whether the data could be approximated by a Student's t-distribution. Our results of a Kolmogorov-Smirnov test (KS = 0.048 and p = 0.333) do not reject the null hypothesis. This indicates that a t-distribution could be a fit for the data.

Despite the slightly non-normal raw data, the Central Limit Theorem implies that the sampling distribution of the mean will be approximately normal because the sample size is large (388).

Overall, the distribution indicates that we do not need to focus on correcting the few extreme stations. The data is stable enough to be used for further analysis and to explore general patterns.

4.2 Route Analysis

4.2.1 Internal Bike Trips

We analysed the number of bike trips made from station to station. It was done taking into consideration only Gdańsk stations, excluding both FreeFloating and external stations from the destination field.

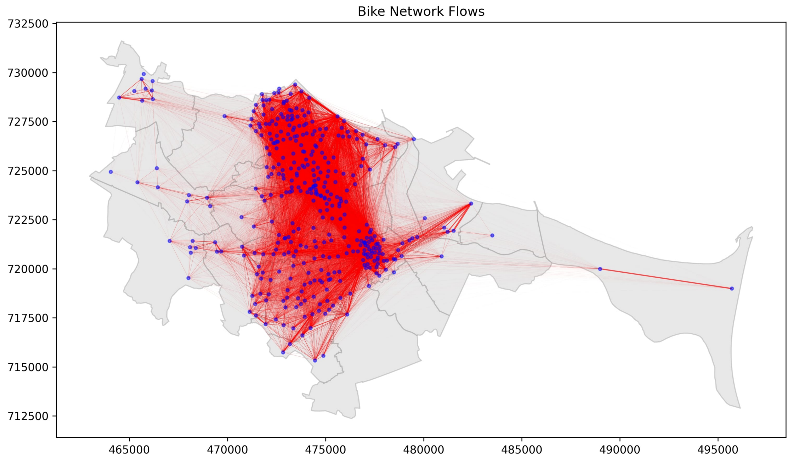

Figure 4.2: Bike flow for MEVO in Gdańsk.

Figure 4.2 illustrates how the bikes move through Gdańsk. The map was made while specifying that the lines between stations vary in width based on number of trips made (the more visible the line the more trips made). Figure 4.3 shows only the most popular trips. The most visible on the map being trips close to the beach Plaża Brzeźno and the beach Plaża Jelitkowo, showing that BSS users have an interest in this area.

When it comes to only internal trips, while in 92.98 % cases the bike went to another station, in the substantial number of trips the bike returned to its starting location (loop), as shown by Figure 4.4.

Figure 4.5 shows how loop and non-loop trips compare in stations which have the highest counts. The number of loop vs non-loop trips made vary between stations in different locations.

As example the most prominent routes overall (those with the highest trip counts) were: GDA071 → GDA071 and GDA072 → GDA072. Both of which are loops, and the stations are located close to the beach Plaża Brzeźno and near each other. It might suggest that users in this location enjoy recreational trips, rather than using bikes as a form of transport.

The net flow (difference between number of returned and picked up bikes) is shown on Figure 4.6

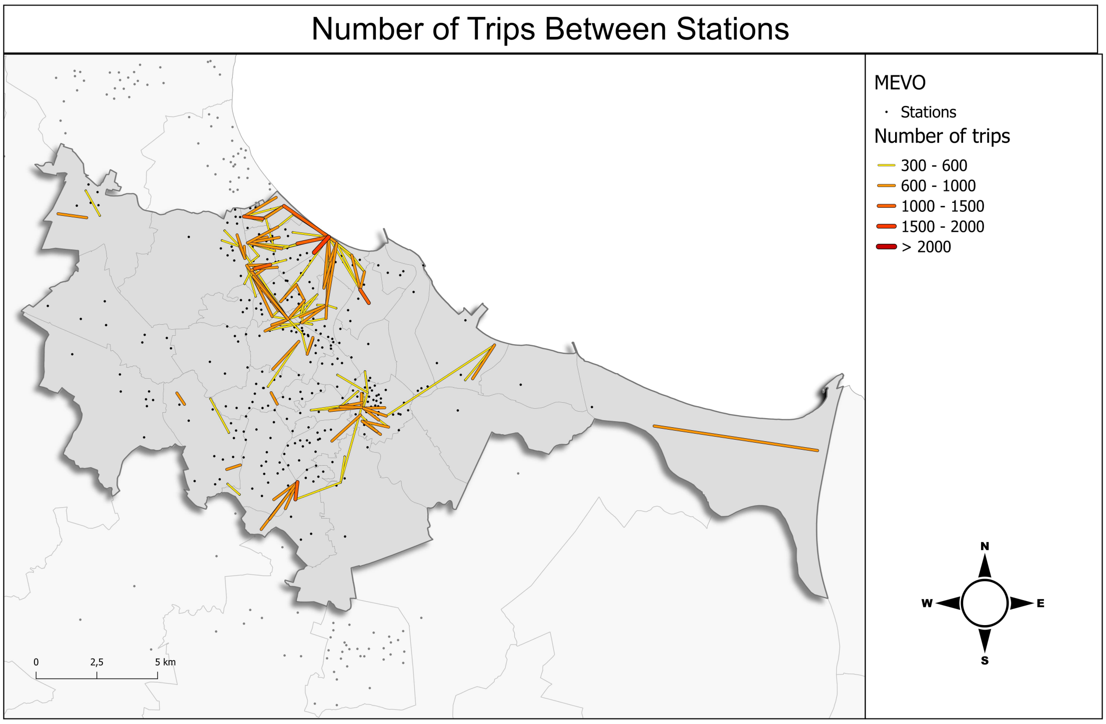

Figure 4.3: Number of trips between the most popular pairs of stations.

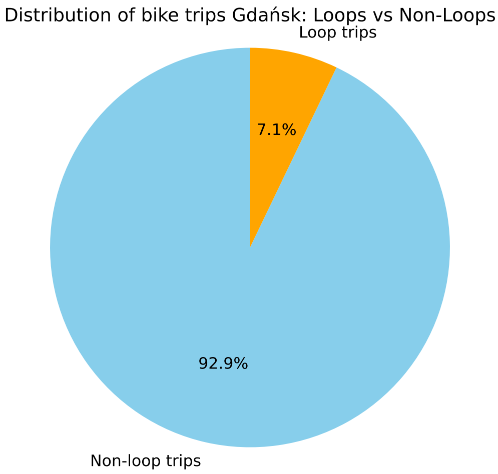

Figure 4.4: Distribution of bike trips in Gdańsk - loops (meaning returning to the same station) vs non-loops.

and Table 4.1, while skewness and kurtosis of number of picked up and returned bikes is in Table 4.2.

The net flow is centred around zero, with most stations showing small imbalance. The distribution of it has a positive skew (0.77) and heavy tails (kurtosis 18.2). This shows that a few stations experience a substantial difference between bikes rented and returned.

Figure 4.8 shows all stations and their average number of bikes rented. It helps distinguish potential areas to study. We can distinguish that stations with highest numbers of rentals are located closer

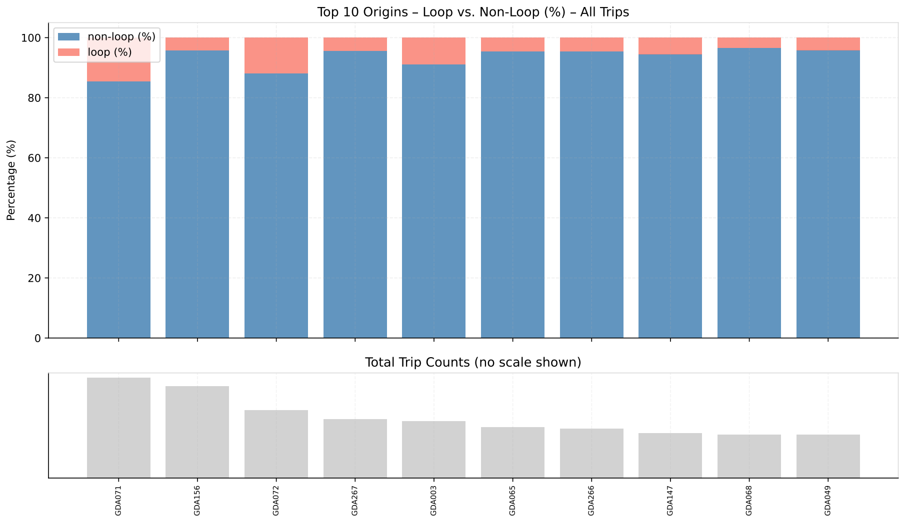

Figure 4.5: Top 10 routes by number of trips.

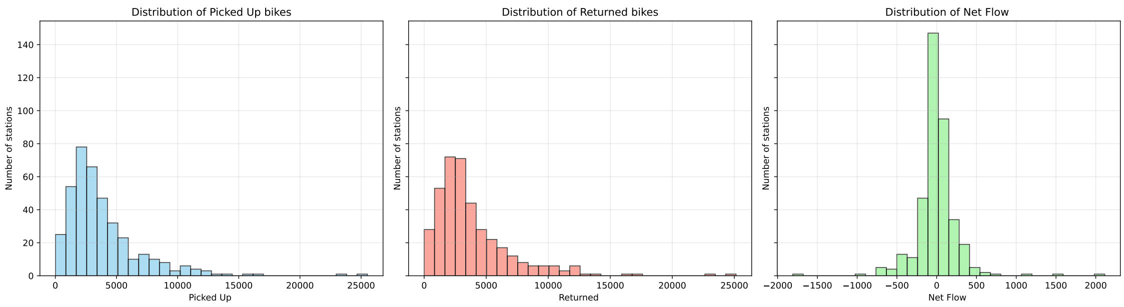

Figure 4.6: Histograms of the distributions for all stations: bikes picked up, bikes returned, and net flow (returned - picked up).

to cycleways and the sea. Additionally, many stations with high bike rentals are located in the Śródmieście area. This map functions as a way to show areas with high activity, providing an overview of bike rentals.

The stations with highest numbers of rentals and the stations with the lowest can also be seen on Figure 4.7.

Similarly Figure 4.9 and Figure 4.9 gives an overview how net flow looks trough the are of Gdańsk, and highlights places in which there are high differences between the number of bikes picked up and returned. This helps identify which stations require frequent bike relocation to ensure their availability. It also shows the need to track where bikes travel beyond the boundaries of Gdańsk, as accounting all the trips the net flow for most of Gdańsk stations is negative.

Table 4.1: Summary statistics of net flow in Gdańsk

	net flow
mean	0.000000
median	-4.000000
std	267.680080
min	-1809.000000
25%	-84.000000
75%	99.500000
max	2113.000000
skewness	0.770691
kurtosis	18.192420

Table 4.2: Skewness and kurtosis of the bike flow in Gdańsk

	picked up	returned
skewness	2.503754	2.415433
kurtosis	10.485168	9.429410

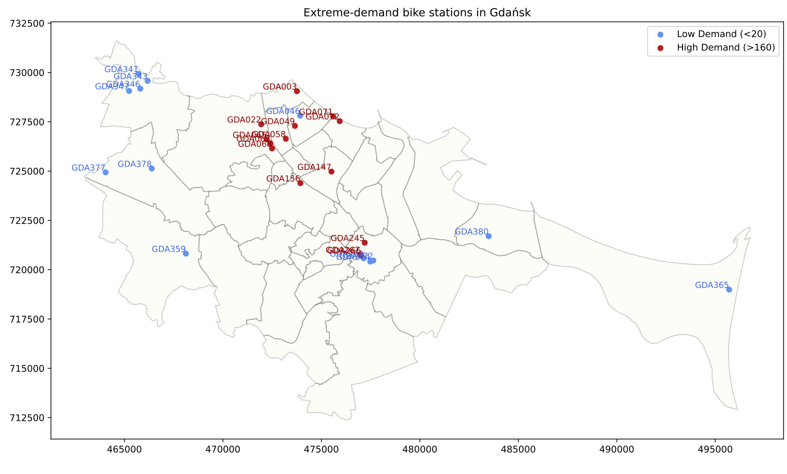

Figure 4.7: Stations with very low and high number of average bikes rented per hour.

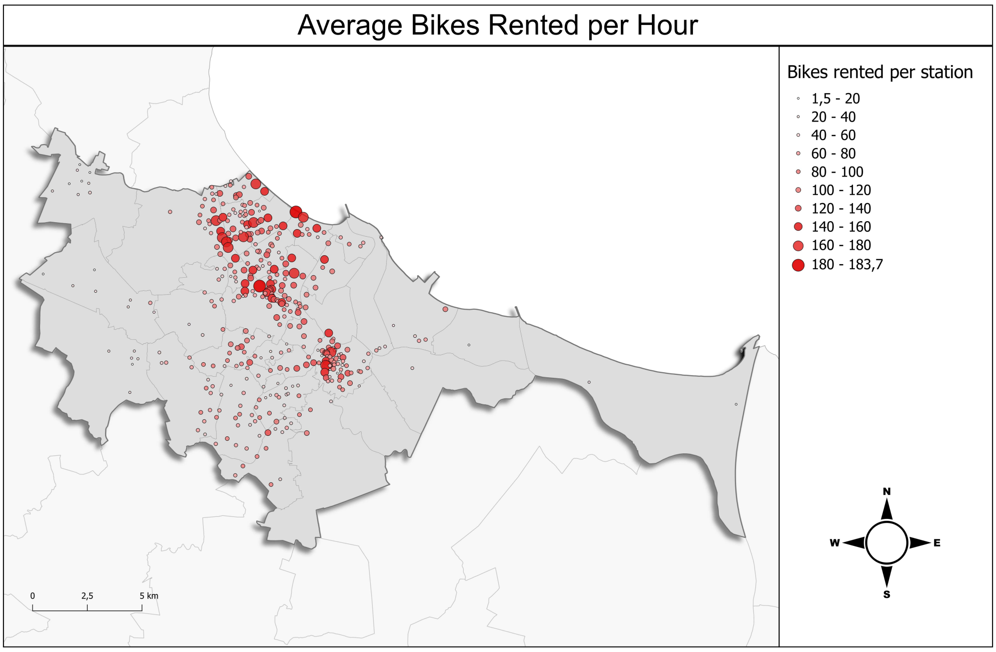

Figure 4.8: Average number of bikes rented per hour from each station.

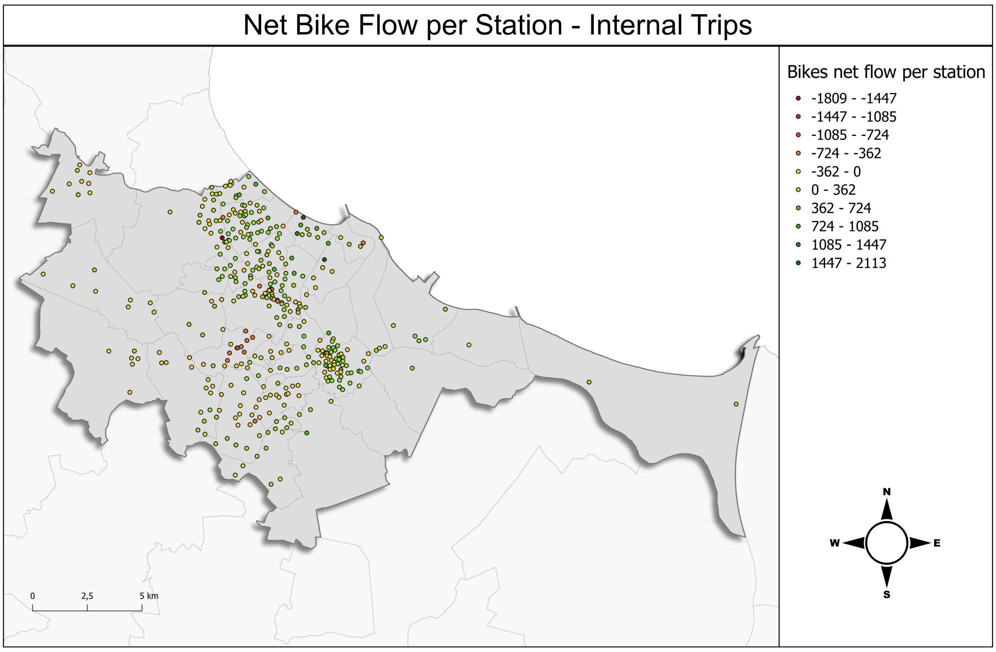

Figure 4.9: Net flow of bikes per station including only internal trips.

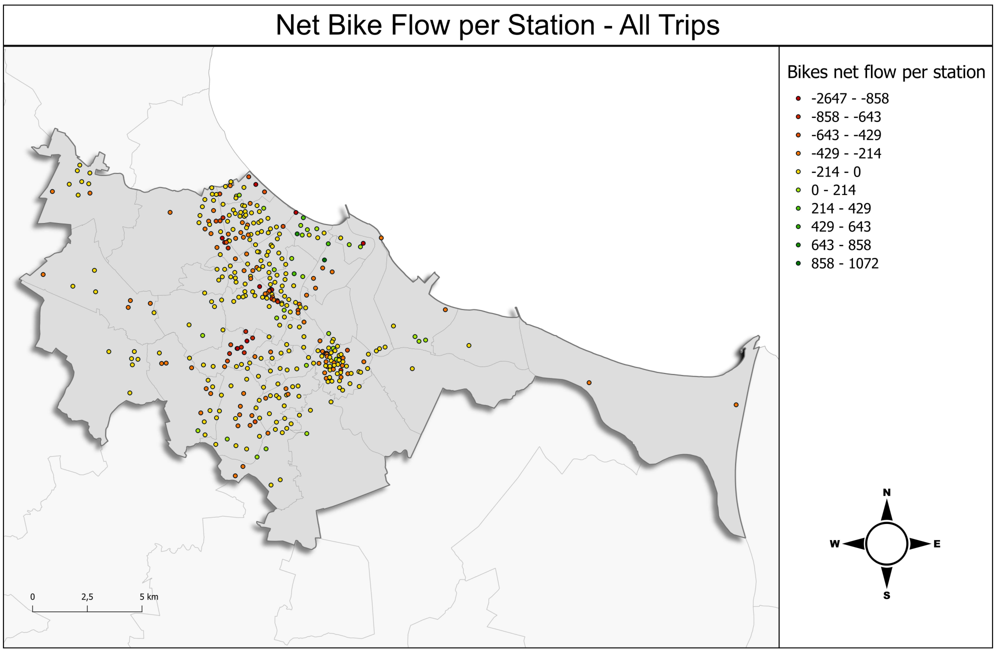

Figure 4.10: Net flow of bikes per station including all trips.

4.2.2 Relationship with Nearby Cities

To find out how many bikes leave the city of Gdańsk, we also conducted an analysis that includes trips made outside of Gdańsk, but only in the direction of the external city. Trips made to the FreeFloating locations are also being considered.

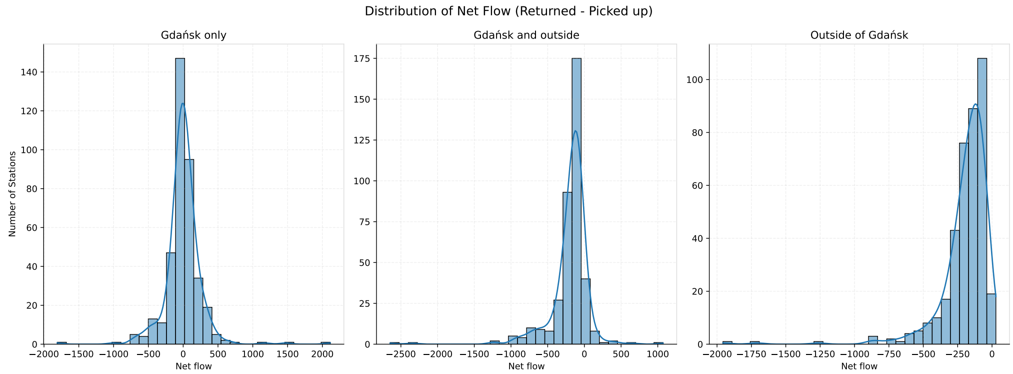

Figure 4.11: Histogram of netflows - between only Gdańsk stations, between Gdańsk stations and any stations, and between Gdańsk stations and external stations.

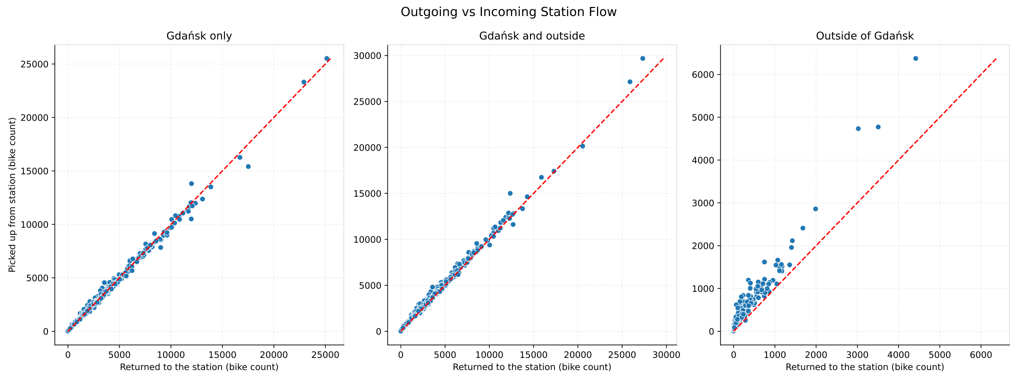

Figure 4.12: Scatterplots of netflows - between only Gdańsk stations, between Gdańsk stations and any stations, and between Gdańsk stations and external stations.

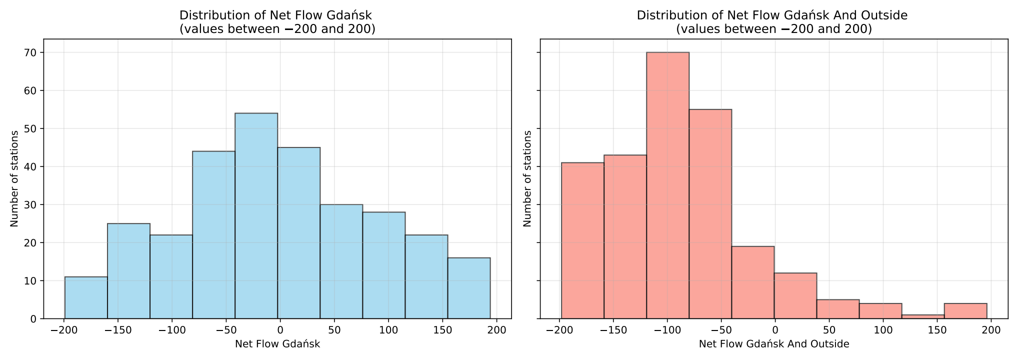

Figure 4.13: Closer (only values between -200 and 200) look at histograms of net flows only inside Gdańsk, and taking places outside of Gdańsk into consideration.

The net flow of bikes depending on the area scope is summarised in Table 4.3, and shown as histograms and scatterplots on Figure 4.11 and Figure 4.12. Figure 4.13 shows that in close up of the histograms of all trips and only Gdańsk trips.

As stated previously net flow for stations inside Gdańsk is centered around zero, indicating that

Table 4.3: Summary statistics of the different new flows

	Gdańsk	Gdańsk and outside	Outside of Gdańsk
mean	0.000000	-194.636598	-194.636598
median	-4.000000	-133.000000	-143.500000
std	267.680080	277.533553	190.993728
min	-1809.000000	-2647.000000	-1957.000000
25%	-84.000000	-239.000000	-236.000000
75%	99.500000	-75.750000	-89.750000
max	2113.000000	1072.000000	28.000000
skewness	0.770691	-3.399783	-4.467942
kurtosis	18.192420	26.424058	31.194899

pickups and returns are largely balanced. Such cannot be said about the distribution of net flow with the area outside of Gdańsk. The distribution is substantially negatively skewed with much heavier tails when journeys to and from outside of Gdańsk are included. Additionally, the mean net flow is negative, indicating that many stations lose more bikes than they acquire.

Considering only trips that contain stations outside of Gdańsk [1] makes this pattern more evident. This provides us with information that bikes that leave Gdańsk more often do not return.

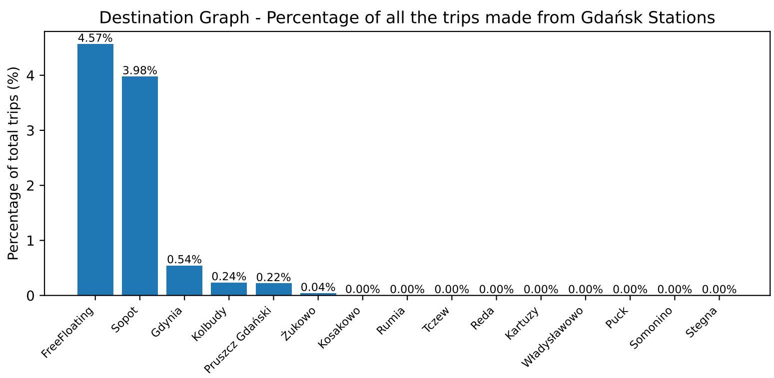

Figure 4.14: The number of trips made outside of the Gdańsk.

Gdańsk was excluded from Figure 4.14, due to high number of internal trips: 90.41 % of total trips made from Gdańsk stations for the whole time period, counting the loops (the ones made from the same station to the same station).

FreeFloating is taken as separate category, but still may contain the trips within Gdańsk. Overall the trips not made to any station make 4.57 % of all the trips.

The aggregated destination distribution shows that the highest number of trips is being made to Sopot. Even Gdynia, while being the second city with the largest bike inflow, is falling far behind. It shows that the highest number of bikes that leaves Gdańsk will be found in Sopot,

[1] This only includes trips from Gdańsk stations to external stations and from external stations to Gdańsk

and suggest that when considering relocation of the bikes, this city should be especially taken into consideration. Possibly seeing if such relationship exists the other way around should be also examined.

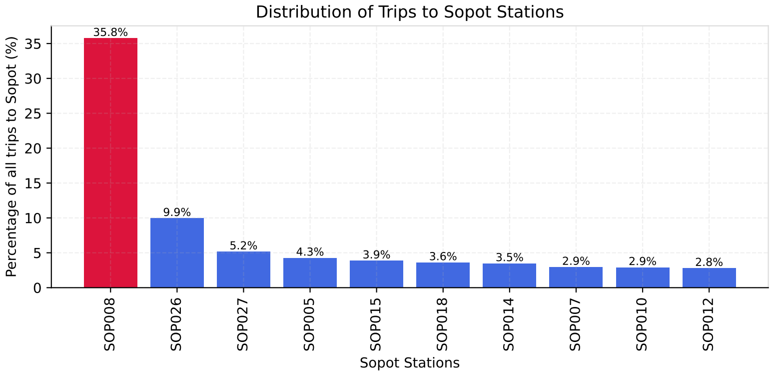

Figure 4.15: The distribution of trips that end in Sopot. Stations with percent below 2.5 % got excluded (18 Sopot stations). Only trips to Sopot are considered for the figure.

The station that most often ends up with the bikes from Gdańsk is SOP008 (visible on Figure 4.16) with 35.75 % of all the bikes that leave for Sopot, and as shown on Figure 4.15 the difference to other Sopot stations is substantial. The same Sopot station accounts for 28.30 % of all external trips, next station ordered by number of trips contributes way less - 7.87 %.

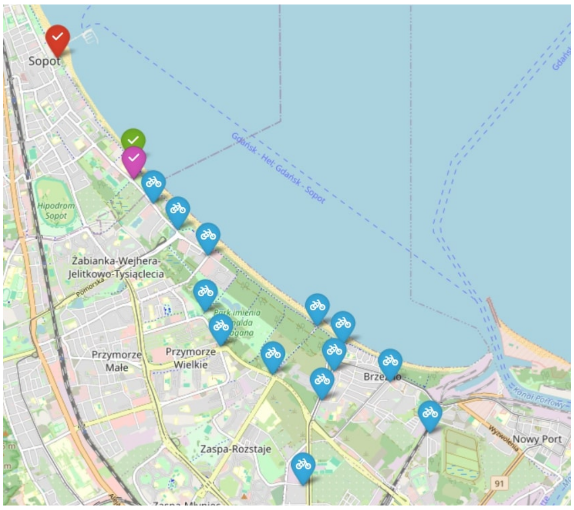

Figure 4.16: Three Sopot stations: Red - SOP008, Green - SOP026, Purple - SOP027, and in Blue: all Gdańsk stations that have over 400 trips to station SOP008

4.3 Average Number of Bikes by Gdańsk Districts

Since we got data concerning population split by districts, we wanted to see its relationship to bike rentals. From data available we calculated statistic for bike sharing activity for each district. We analysed such, in hopes of helping with planning bike relocation between districts.

Table 4.4 shows relationships between district characteristics (population statistics and bike-sharing activity).

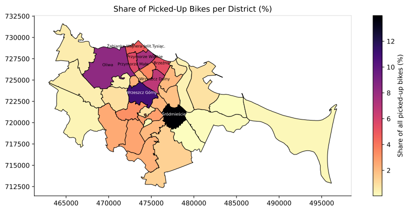

Figure 4.17: Share of picked up bikes by district.

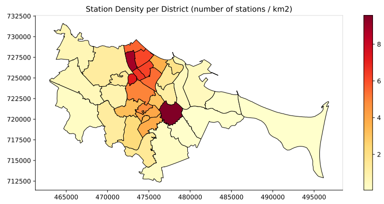

Figure 4.18: Station density in each district.

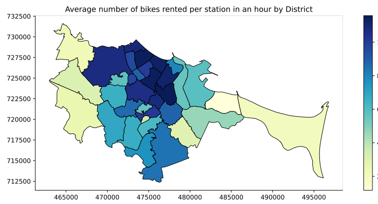

Figure 4.19: Average bike rentals per hour for each district.

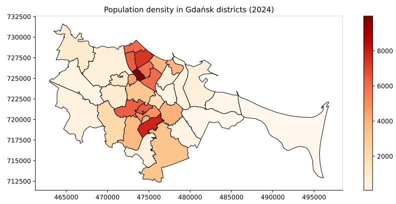

Figure 4.20: Population density in each district.

Table 4.4: Correlations between districts

	avg bikes per hour	total picked up bikes	station density (station / km2)
population density (person/km2)	0.458409	0.336453	0.782878
population number	0.303780	0.510301	0.359004
avg bikes per hour	1.000000	0.613919	0.564832
station density (station / km2)	0.564832	0.659393	1.000000
total picked up bikes	0.613919	1.000000	0.659393

As shown in the Table 4.4, districts with higher population density tend to have more stations per square kilometre (0.78), and they also show moderately higher average bike usage (0.46). Station density itself correlates strongly with both the total number of bikes picked up (0.66) and the average number of bikes rented per hour (0.56), which suggests that areas with more concentrated station placement see higher activity. Overall, station density appears to be the strongest structural factor linked to bike activity.

The picked factors are shown in following figures: Figure 4.17 (share of all picked up bikes), Figure 4.18 (station density), Figure 4.19 (average bikes per hour), and Figure 4.20 (population density). In visualisation Śródmieście clearly stands out with very high bike rentals. In addition, it can also be seen how the population density is related to bike usage. However, some districts with similar population density have different rental levels, which suggests to us that other factors also affect BSS activity.

The boxplots (Figure 4.21 and Figure 4.22) further shows that population density does not fully explain bike rentals. Interestingly, while having low population density Młyniska have high average bike rentals.

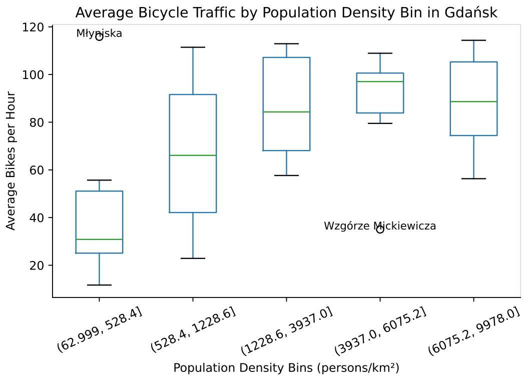

Figure 4.21: Box plot showing the average bikes per hour for different population density bins for districts in Gdańsk.

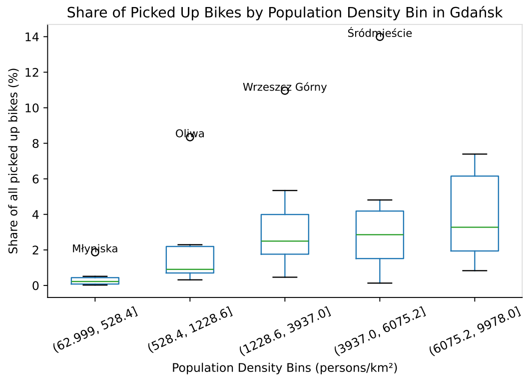

Figure 4.22: Box plot showing the share of all picked up bikes for different population density bins for districts in Gdańsk.

4.4 Bike Relocation by Gdańsk Districts

We wanted to understand how bikes move between different districts in Gdańsk. In particular, we looked at which city or district receives the highest number of bikes from each origin district.

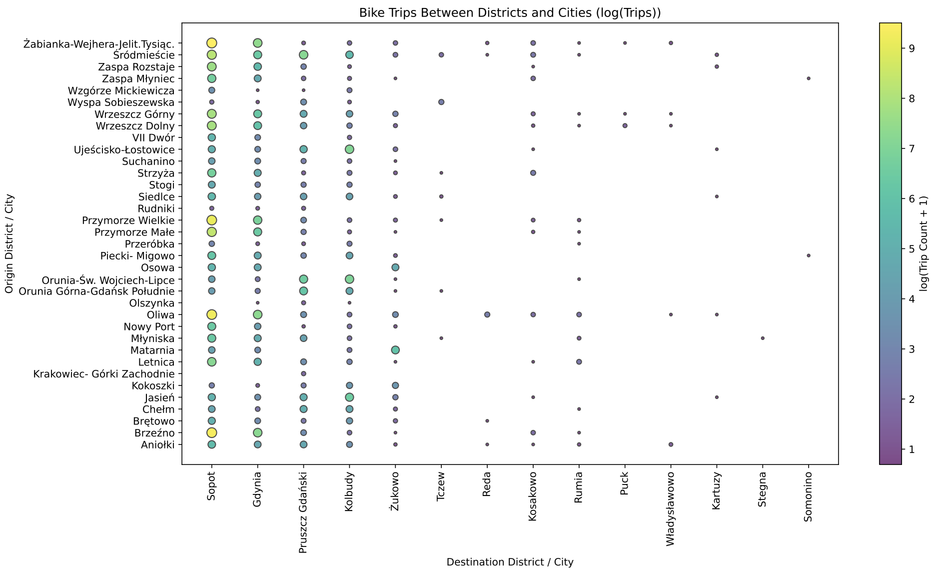

Figure 4.23: Scatterplot showing the number of trips made from each origin station to each destination station (grouped by district within Gdańsk and by city outside Gdańsk). Marker sizes were scaled using the logarithm of the trip counts. Showing only the external destinations.

Figure 4.23 shows the number of trips from each origin station to each destination station outside Gdańsk. The size of the markers corresponds to the logarithm of the number of trips. From this figure, it is clear that most bikes rented outside the city go to Sopot, as already seen previously. The districts that most of trips to Sopot are made from are Żabianka-Wejhera-Jelitkowo-Tysiąclecia, Przymorze Wielkie, Oliwa, and Brzeźno. Such information makes us understand on which districts to focus when it comes to relocating the bikes that get out of Gdańsk.

Figure 4.24 shows trips within Gdańsk. Here we can see that Śródmieście and Wrzeszcz Górny have the highest number of trips. The scatterplot shows that trips with the same origin and destination

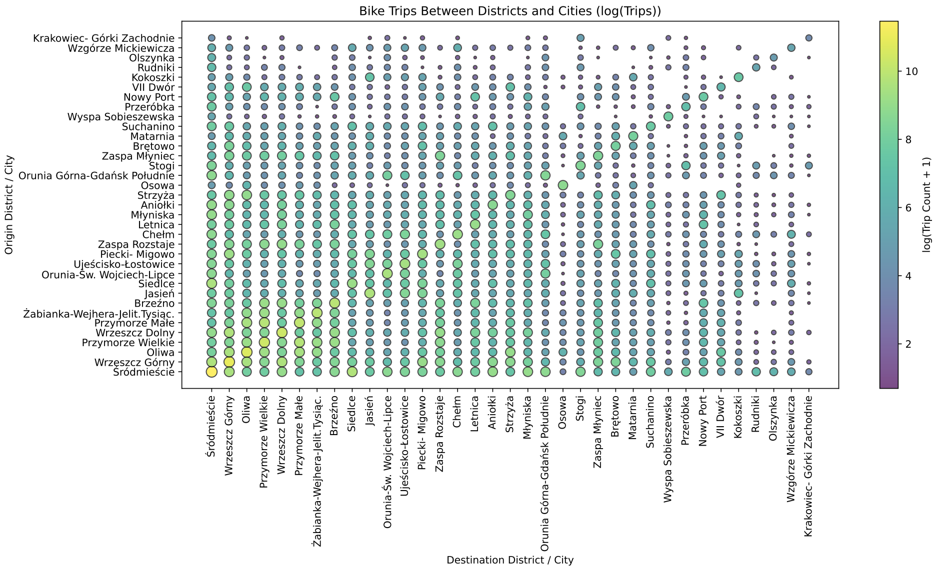

Figure 4.24: Scatterplot showing the number of trips made from each origin station to each destination station (grouped by district within Gdańsk and by city outside Gdańsk). Marker sizes were scaled using the logarithm of the trip counts. Showing only the internal destinations.

districts are the most frequent, as reflected by the largest circles. This suggests that a large part of bike trips are local, staying inside the origin district.

To measure the imbalance in bike relocation between districts, we created heatmaps. Figure 4.25 shows the ten districts with the largest imbalances in bike flows. Here, the imbalance is calculated as the difference between the number of bikes leaving and entering each district.

When considering all districts, including cities outside Gdańsk, the average imbalance between districts is about 169 bikes. The median imbalance is 17 bikes. This shows that, on average, some districts lose or gain a substantial number of bikes relative to others.

Figure 4.26 shows the full heatmap of imbalances for all districts. Many gray areas on this heatmap are a result, of low number of bikes being returned and picked up inside some districts. Considering that, we decided to include the smaller heatmap. The full heatmap still shows which districts can be focused on.

During the studied time period, the districts between which the highest imbalance occurs are:

- Śródmieście → Siedlce: >1600 bikes
- Wrzeszcz Górny → Brzeźno: >1100 bikes
- Śródmieście → Brzeźno: >1100 bikes
- Piecki-Migowo → Śródmieście: >1100 bikes
- Piecki-Migowo → Wrzeszcz Górny: >850 bikes
- Oliwa → Przymorze Wielkie: >850 bikes
- Ujeścisko-Łostowice → Śródmieście: >850 bikes
- Oliwa → Wrzeszcz Górny: >850 bikes

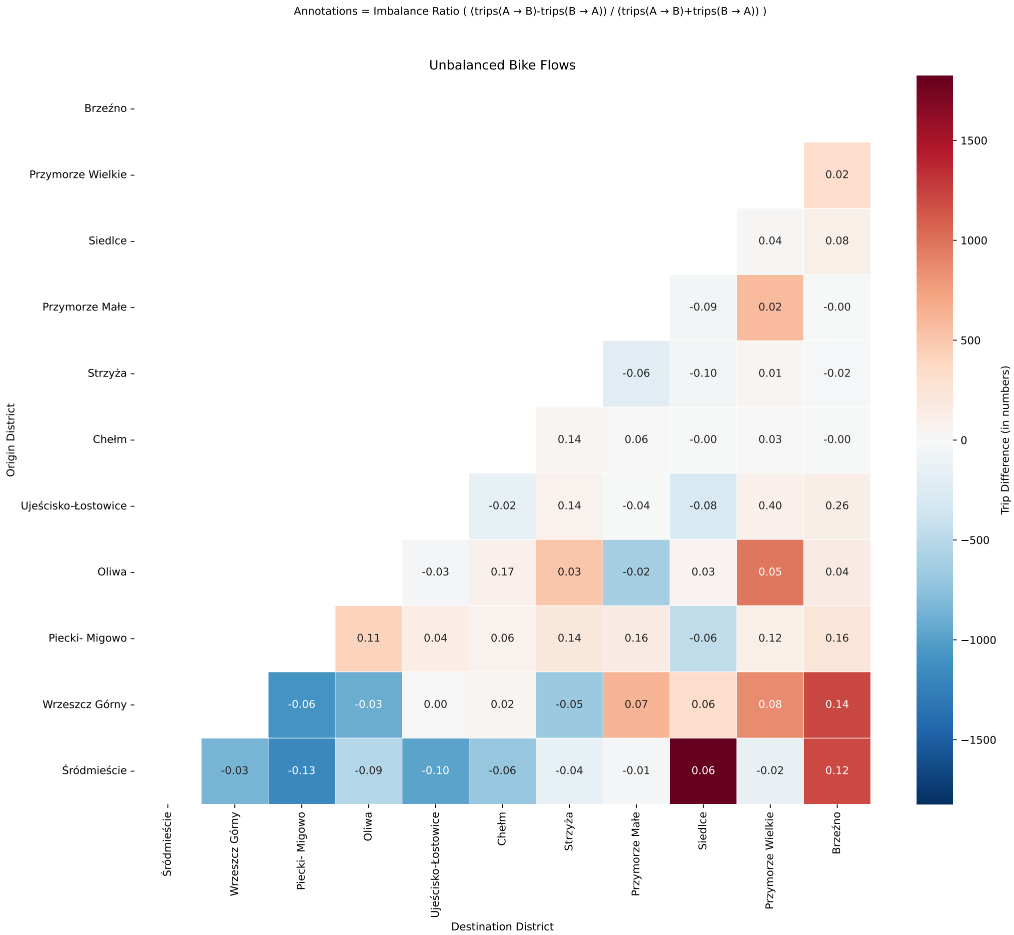

Figure 4.25: Heatmap showing the imbalance for number of bikes relocated from one district to another (by BSS users). Only 10 districts which experience the highest imbalances are shown.

- Wrzeszcz Górny → Przymorze Wielkie: >850 bikes
- Wrzeszcz Górny → Śródmieście: >600 bikes
- Chełm → Śródmieście: >600 bikes
- Strzyża → Wrzeszcz Górny: >600 bikes
- Przymorze Małe → Oliwa: >600 bikes
- Wrzeszcz Górny → Przymorze Małe: >600 bikes

These trips are highlighted on Figure 4.27, which shows the ten most imbalanced routes between districts. Each arrow points from the district where bikes are usually rented to the district where they are usually returned.

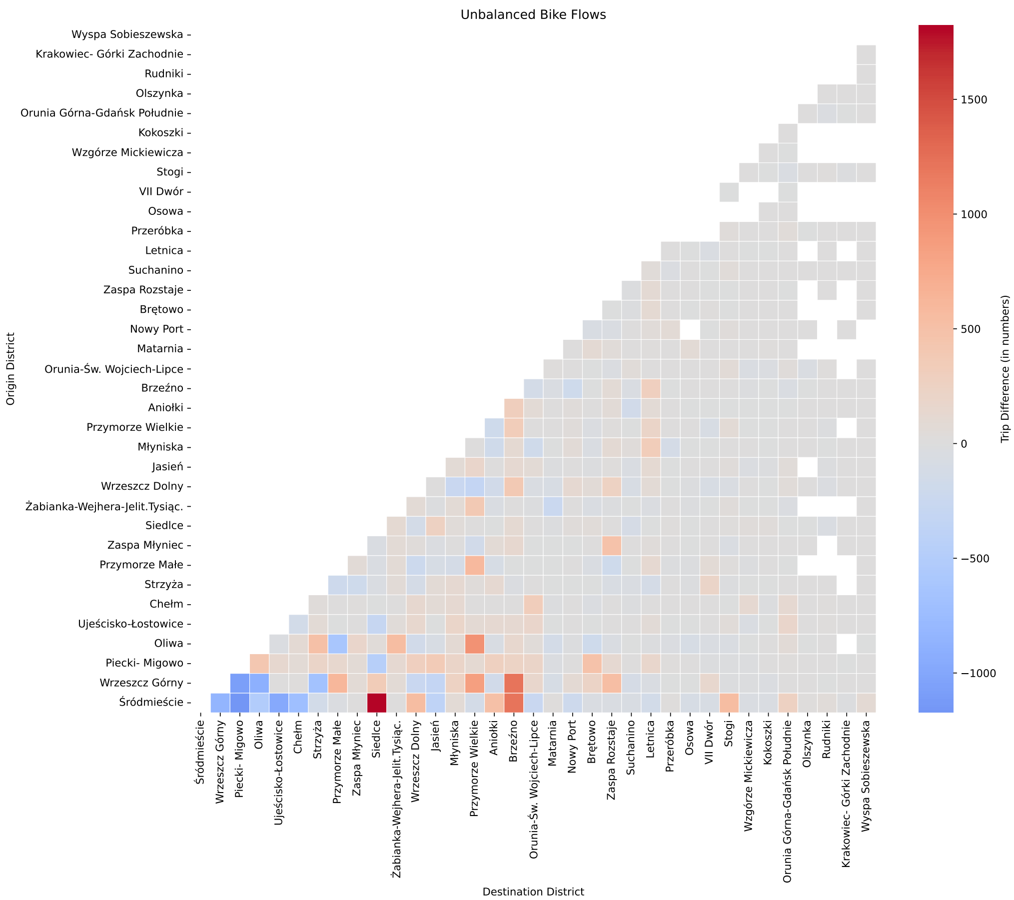

Figure 4.26: Heatmap showing the imbalance for number of bikes relocated from one district to another (by BSS users).

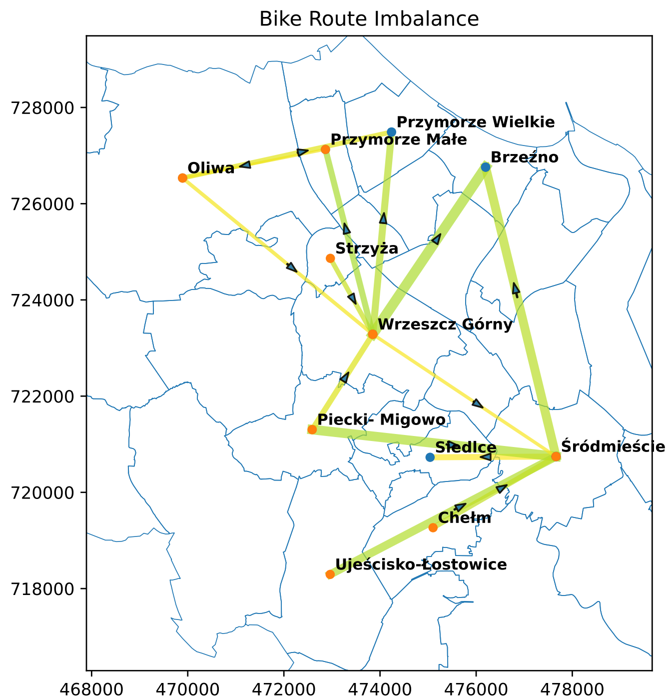

Figure 4.27: Map showing the ten most imbalanced routes between districts. Each arrow points from the district where bikes are usually taken to the district where they are usually returned.

4.5 Temporal Analysis and Weather

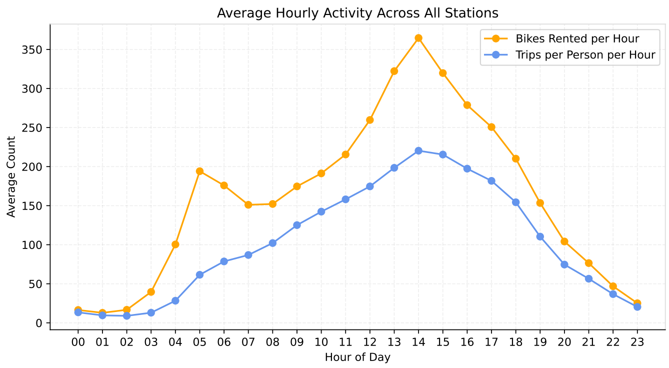

Figure 4.28: Average number of bike rented from the stations and trips per person made from the station at each hour.

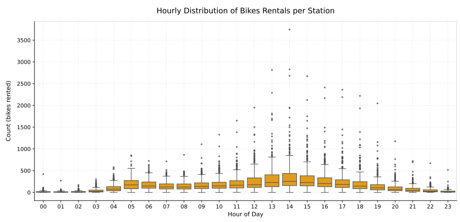

Figure 4.29: Average number of bike rented from the stations at each hour.

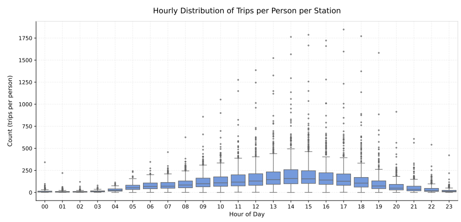

Figure 4.30: Average number of trips per person made from the station at each hour.

Figure 4.28 shows the average number of bikes rented from all stations at each hour of the day, along with the average number of trips per person. Similarly, the boxplots in Figures 4.29 and 4.30 show how they distribute at each hour.

The line plots and boxplots expose that most rentals occur during peak hours, around 5:00 and 14:00. During these hours, there are more extreme values (many outliers), meaning some stations have unusually high rental numbers.

Trips per person also peak around 14:00, but the distribution is smoother, suggesting that most people make a similar number of trips, with fewer extreme cases.

Figure 4.31 shows the daily rental trends over the year, smoothed with a 7-day moving average. Rentals are lower in April and September, while the distance travelled per trip is higher at the start of May. This shows that bike demand changes over the year, possibly because of seasonal factors.

Figure 4.32 and 4.33 show the average number of trips by weekday and by type of day (workday, weekend, holiday). Fewer trips occur on weekends and holidays, especially Sundays. A t-test confirms that weekday and weekend trips are significantly different (T-statistic = 4.32, p-value = 0.000). Holidays have fewer data points, so the results are less statistically reliable, but they behave similarly to weekends.

Temperature also affects bike rentals. Figure 4.35 shows that rentals increase with higher temperatures. The correlation between daily maximum temperature and rentals is 0.68, while precipitation has a smaller negative correlation of -0.22. This indicates that temperature is a stronger driver of

Figure 4.31: Daily bike rentals statistics through the time period with 7 day moving average.

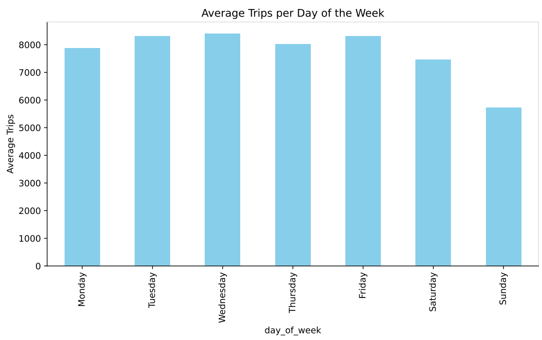

Figure 4.32: Average number of trips by weekday.

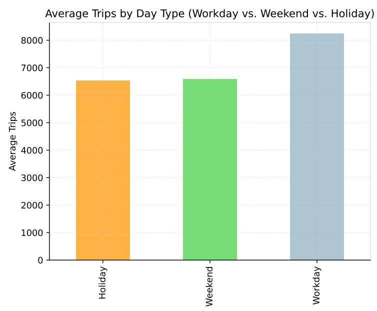

Figure 4.33: Average number of trips by day type.

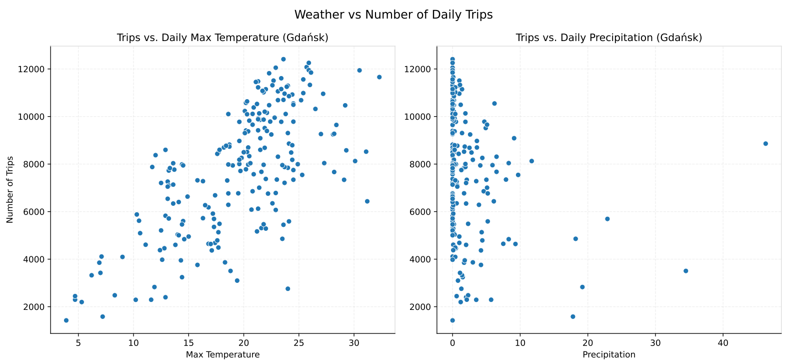

Figure 4.34: Temperature analysis scatterplot.

demand than rainfall, but other factors also influence usage.

Overall, when analysing bike rentals by date, weather needs to be taken into account, as it has a

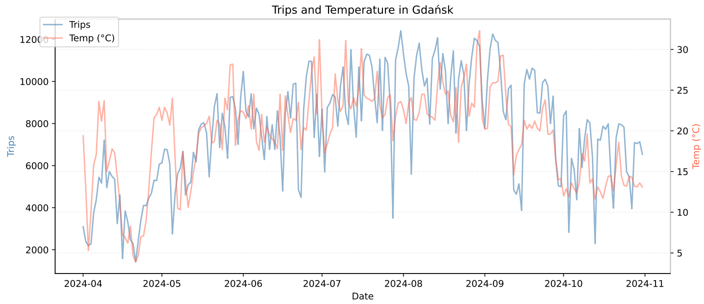

Figure 4.35: Relationship between temperature and bike rentals in a day.

strong influence on demand. Additionally, if we want to predict rentals for a specific day, seasonal patterns should be taken into consideration, since demand changes throughout the year.

4.6 POI Analysis

4.6.1 Correlations

To study how the location of points of interest (POIs) affects bike rentals, we calculated Pearson and Spearman correlations. Additionally, the results function as baseline to later considerations.

Pearson correlation measures linear relationship and Spearman correlation measures monotonic relationship. Spearman correlation helped us distinguish the POIs that show non-linear relationship, or cases where some stations have many outliers within certain POI categories.

The correlations were calculated based on:

- Average bikes rented per hour at each station,
- Distance-decay function (Equation 2.1) for POIs (up to 500 m), an example of the buffer is shown on Figure 4.36
- Number of POIs within different distance bands (150 m, 300 m, 1000 m), the
- Distance to the nearest POI.

All distances were measured as straight-line distances. The correlations are sorted by distance decay, showing that closer POIs have a stronger influence on bike rentals.

Figures 4.37 and 4.38 show correlations for POIs grouped into a smaller number of categories, while Figures 4.39 and 4.40 show correlations with more detailed POI categories. In the end, we decided to use more granular categories.

The results helped us identify which types of POIs are most closely associated with higher bike usage. For example, shops and places related to transport (taken as more general categories) tend to have stronger correlations with bike rentals. Based on calculated correlations we can distinguish what kind of places to focus on when deciding on locations for bike stations, and possibly predict the bike rentals for the station.

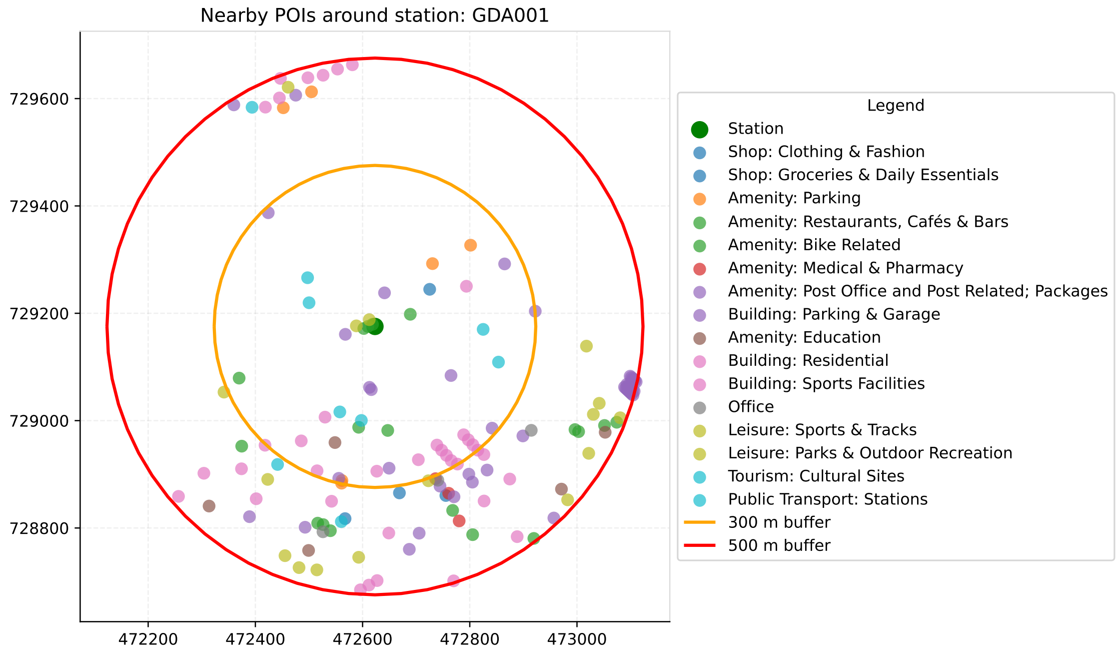

Figure 4.36: The example for how the area around station looks for calculating distance-decay.

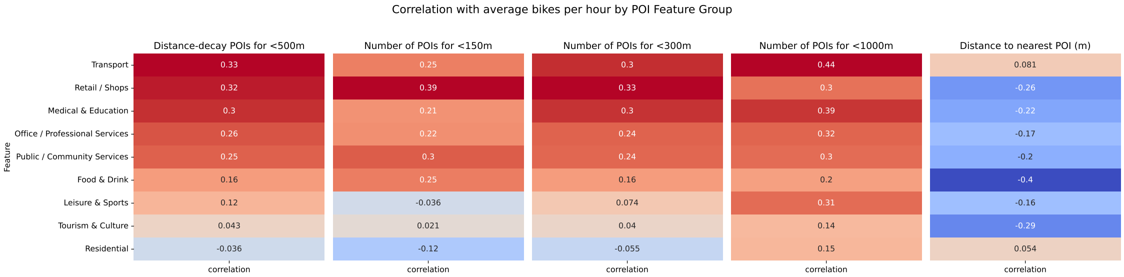

Figure 4.37: Pearson correlations with POIs grouped into smaller number of categories.

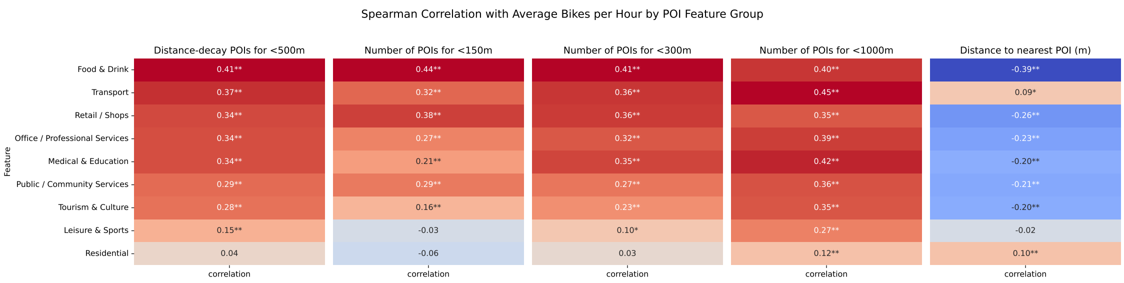

Figure 4.38: Spearman correlations with POIs grouped into smaller number of categories.

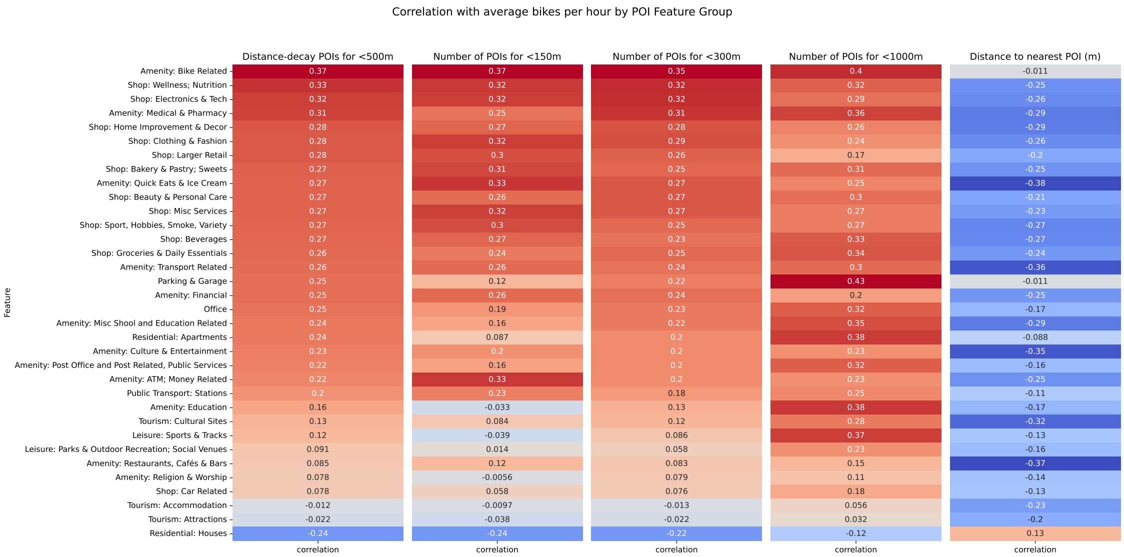

Figure 4.39: Pearson correlations with POIs grouped into bigger number of categories.

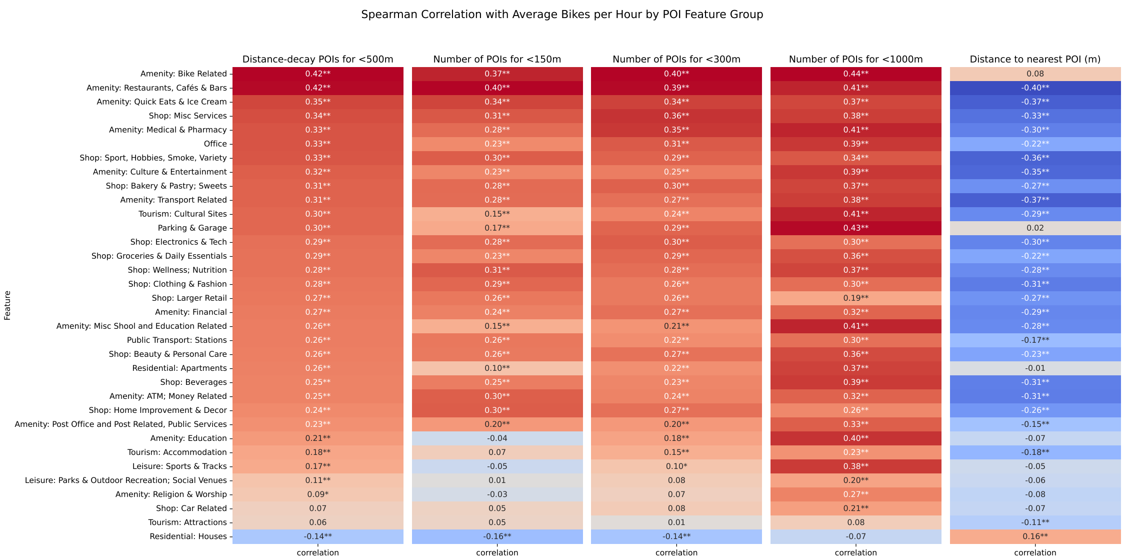

Figure 4.40: Spearman correlations with POIs grouped into bigger number of categories.

4.6.2 Regression

We used two regression approaches: GWR and OLS. We included it in our analysis to see if considering more local coefficients instead of global ones, will change the result significantly.

GWR is our main model, because it captures spatial variation in the relationships. We ran an OLS model as well, mainly to compare the global coefficients with GWR and to understand how much taking the space (local area) into consideration improves the fit.

We estimated both models twice. The starting list of predictors were all the distance-decay values for POIs and population density for districts.

First Run

In the first run, we included all POI categories that showed at least 0.2 correlation (Pearson or Spearman) with bike rentals. This gave us 29 predictors without filtering. We scaled both predictors and target values using Min-Max scaling (`sklearn.preprocessing.MinMaxScaler`).

The results are in Table 4.5.

Table 4.5: Comparison of OLS and GWR models for bike rentals in Gdańsk

Metric	OLS	GWR
Method	Least Squares	Geographically Weighted Regression
RMSE	0.166	0.138
RMSE (rescaled)	30.28	25.13
R^2	0.372	0.568
Moran's I	0.162	-0.017
Moran's I p-value	0.001	0.342
Residual sum of squares	10.719	7.379
Residual sum of squares (rescaled)	355830.382	244964.253
Log-likelihood	145.719	218.147
AIC	-231.4	-304.484
BIC	-112.6	-43.432
Adjusted R^2	0.321	0.480

We found that GWR fits the data better than OLS. Which means that different areas of Gdańsk react differently to the same POI categories. It also suggests that we may still miss some important factors, such as land structure, or closeness to the sea.

The R^2 for OLS is 0.372, while GWR achieves 0.568 (accounting for spatial variation improves model fit). Similarly, RMSE is lower for GWR than for OLS, indicating smaller prediction errors.

For OLS, the strongest positive predictors of bike rentals are Bike-related amenities, Electronics and Tech shops, Quick Eats & Ice Cream, Medical & Pharmacy amenities, and Finance-related amenities. The strongest negative predictors are ATMs, Beauty & Personal Care shops, Clothing

& Fashion shops, miscellaneous shops in categories such as Sport, Hobbies, Smoke, and Variety Shops, and Offices.

For GWR, the intercept ranges from 0.25 to 0.54 and several predictors change depending on location. Moran's I = -0.017 (p = 0.342) shows that the model errors do not depend on location (model successfully accounts for spatial variation). The Moran's I is also way higher for OLS, which suggests that GWR captures local patterns that OLS misses.

For GWR, the strongest positive predictors of bike rentals are Misc Services shops, Medical & Pharmacy amenities, Quick Eats & Ice Cream, Bike-related amenities, and Clothing & Fashion shops. The strongest negative predictors are ATMs, Beauty & Personal Care shops, Post Office and Postal Services, Education amenities, and Financial amenities.

Overall, both OLS and GWR highlight similar key predictors, such as ATMs and Beauty & Personal Care shops as negative factors and Medical & Pharmacy, Quick Eats, and Bike-related amenities as positive factors.

For GWR the effects depending on location can be seen in the Figure 4.42. Comparison observed and predicted values can be seen by looking at Figure 4.41. And comparison between coefficients for different factors in Figure 4.43, the graph shows both GWR and OLS coefficients.

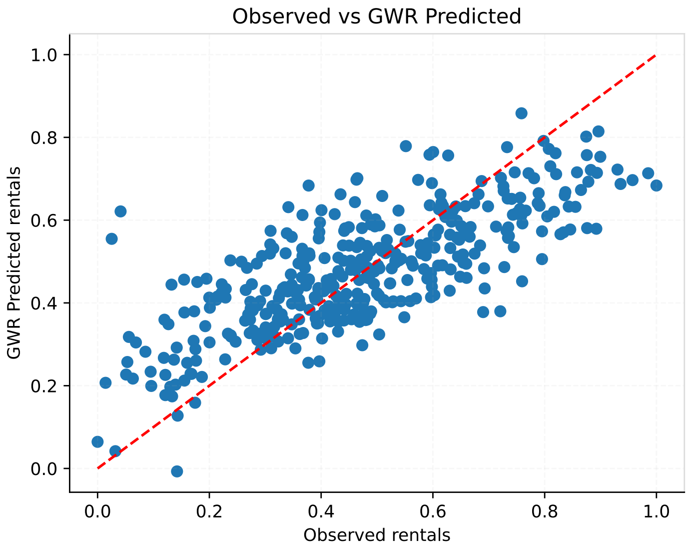

Figure 4.41: Observed vs predicted for all POIs with Spearman or Pearson correlation above 0.2.

Overall, GWR reduces the error, and explains more variation, so the effect of nearby POIs on rentals varies across the city. However, the results indicate that both the OLS and GWR models explain a moderate portion of the differences in bike rental patterns.

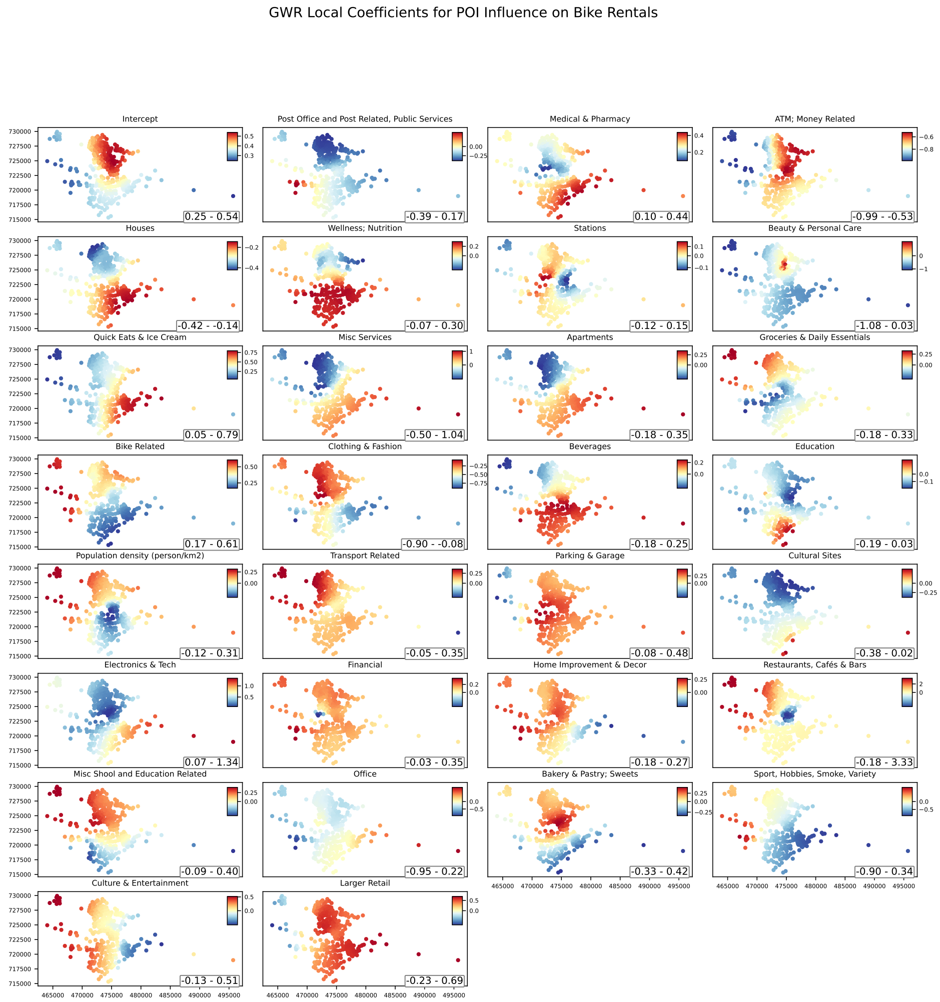

Figure 4.42: Plots showing how different POIs influence the rentals depending on location of the station.

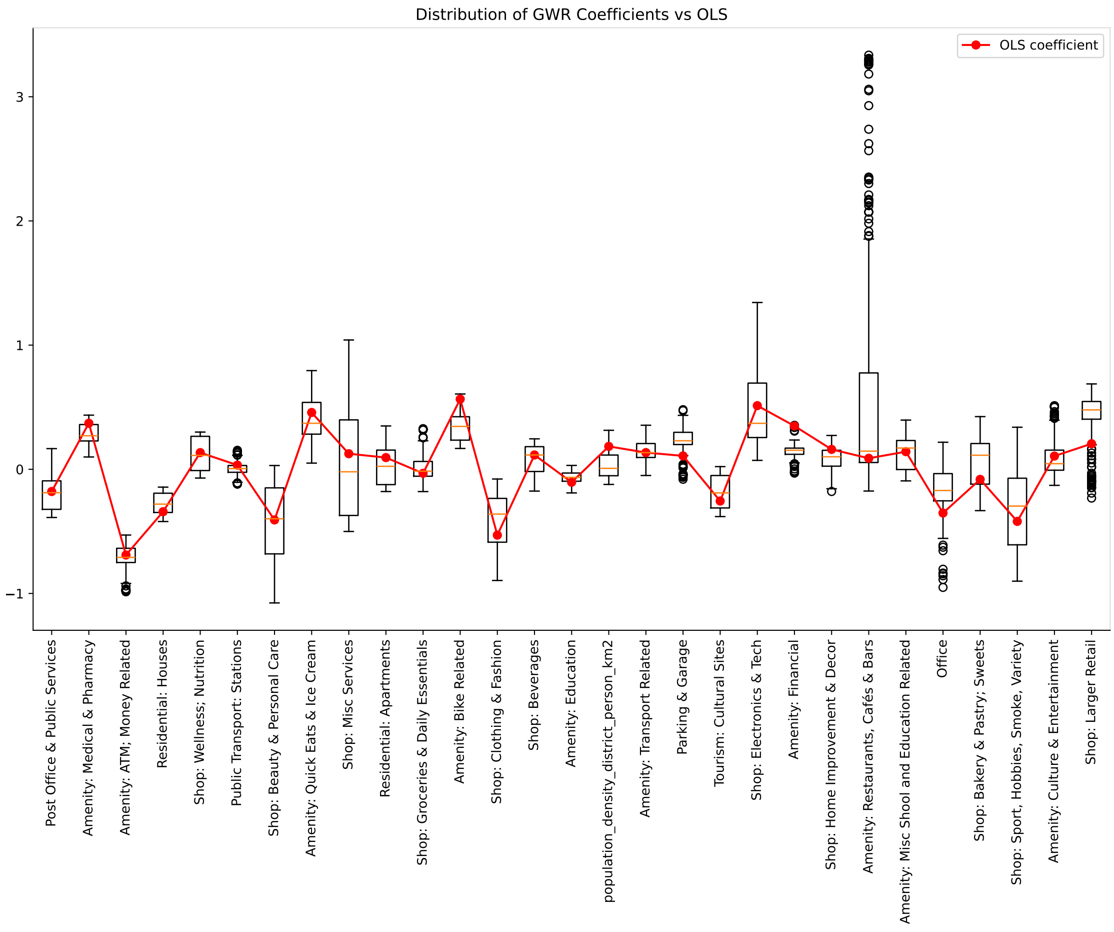

Figure 4.43: Distribution for coefficients for POIS.

Second Run

In the second run, we removed features with very high multicollinearity. We followed two steps:

- We checked feature variance.
  None of the predictors had near-zero variance. This meant that POI distances varied between stations for all categories.
- Then we removed highly collinear predictors.
  We added features one by one and checked the condition number each loop. If the condition number stayed below the threshold (70), we kept the feature. We computed the condition number with `numpy.linalg.cond`.

A high condition number means that some predictors are almost linear combinations of others. The chosen threshold let us preserve relevant information, while not cutting too many predictors (the chosen value is not too strict).

It made the number of predictors go down from 29 to 13. We did not scaled predictors in this case.

Table 4.6: Comparison of OLS and GWR models for bike rentals in Gdańsk

Metric	**OLS**	**GWR**
Method	Least Squares	Geographically Weighted Regression
RMSE	33.10	25.03
R^2	0.250	0.571
Moran's I	0.291	-0.038
Moran's I p-value	0.001	0.153
Residual sum of squares	425050.37	243088.34
Log-likelihood	-1908.35	-1799.94
AIC	3844.69	3739.09
BIC	422820.96	4014.79
Adjusted R^2	0.224	0.478

GWR still fits the data much, it reduces the error and explains more variation in bike rentals.

For OLS the strongest positive effects comes from Medical & Pharmacy amenities, and the strongest negative from ATMs.

For GWR intercept ranges from 34 to 140. Several POI categories show strong regional effects, for example, the strongest positive effect comes from Shop Misc Services, and the strongest negative from ATMs.

The effects depending on location can be seen in the Figure 4.45. Comparison observed and predicted values can be seen by looking at Figure 4.44. And comparison between coefficients for different factors in Figure 4.46.

Reducing the number of predictors in the configuration we implemented does not improve the effect significantly. It may require stricter threshold, or other selection methods (for example multicollinearity checks based on Variance Inflation Factor).

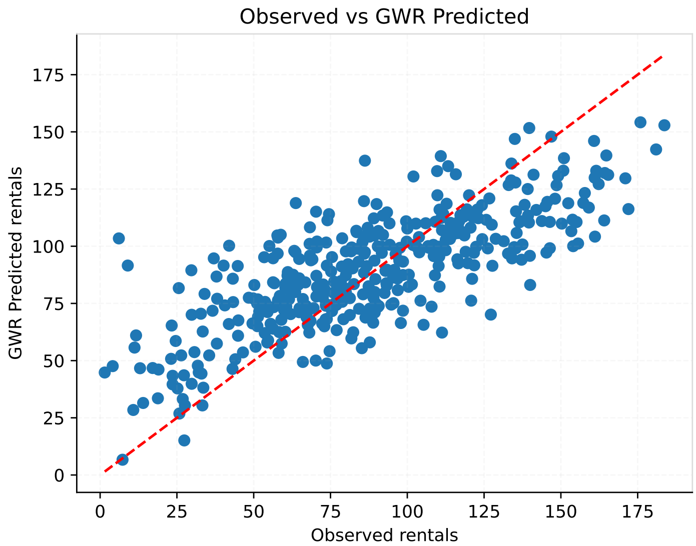

Figure 4.44: Observed vs predicted for fewer predictors.

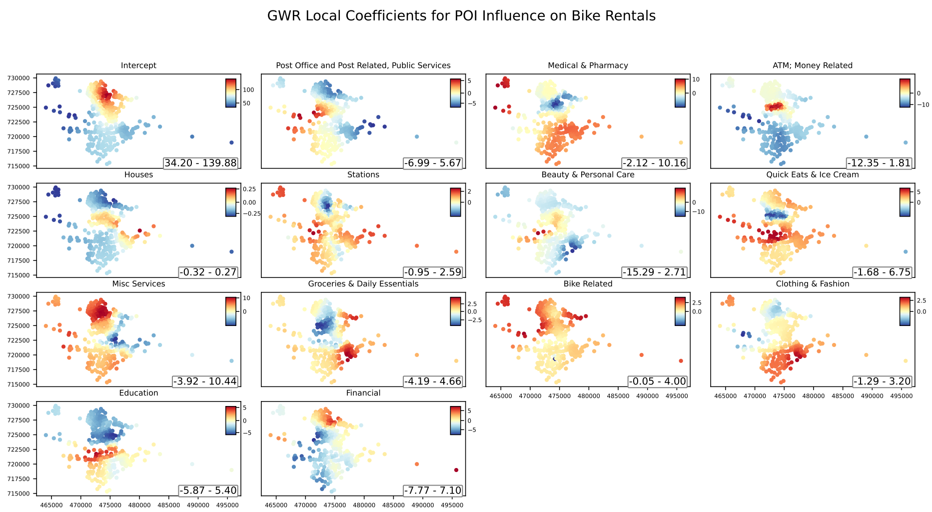

Figure 4.45: Plots showing how different POIs influence the rentals depending on location of the station. Used fewer predictors.

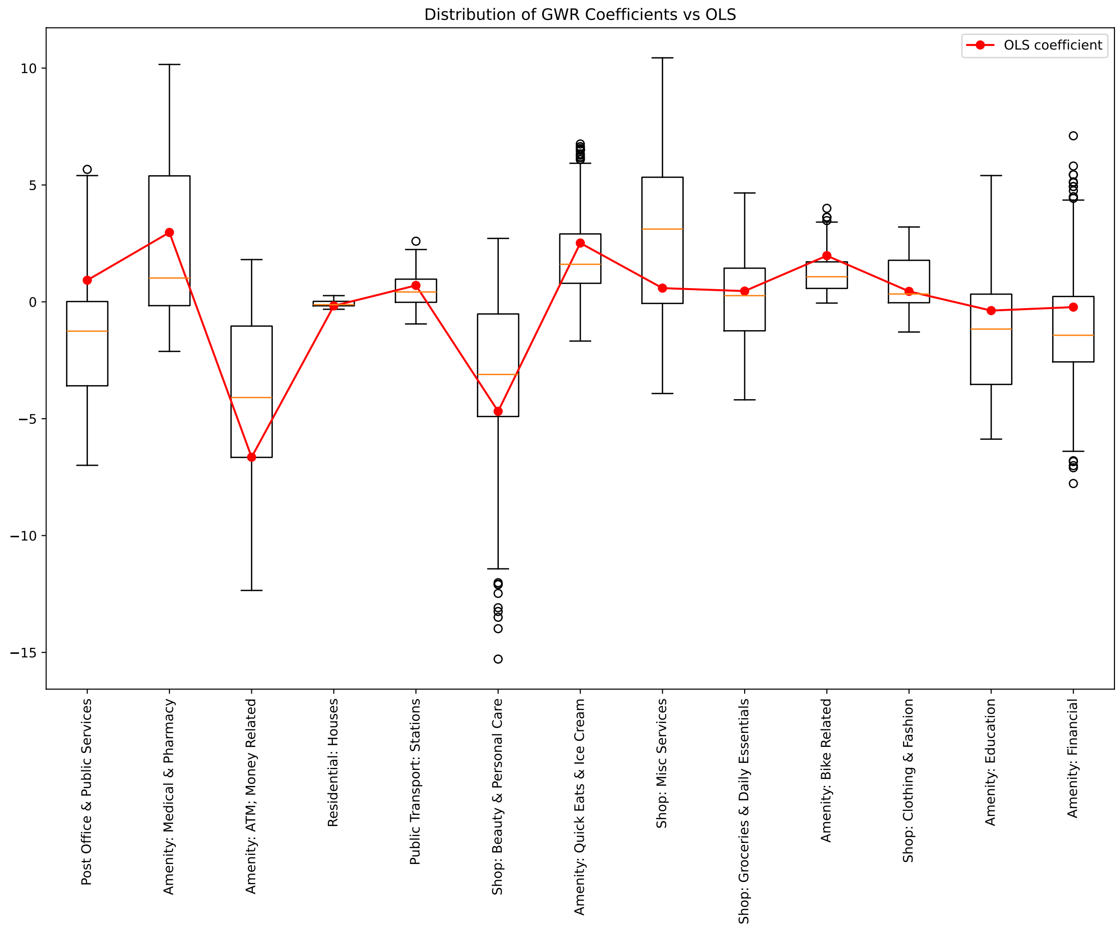

Figure 4.46: Distribution for coefficients for fewer number of POIS.

5. Geographic Information Systems (GIS) Analysis

Author: Aleksandra Susmarska

5.1 Influence Distance Decay of Chosen Factors

We conducted a spatial analysis of the distance decay of our chosen factors to calculate their influence on the potential locations of new MEVO stations.

In order to calculate the decay for each factor, first we had to convert the spatial vector data to raster. We chose a grid of pixels, 10x10 meters each, to balance data accuracy and computation time.

For each of our chosen points of interest, public transport stops, cycleways, residential areas and existing MEVO stations, we calculated their influence for each pixel as a value between 1 and 0.

Figure 5.1 shows the influence values calculated from the distance decay model. The influence value for pixels more than 3 kilometres from the nearest cycleway was rounded to 0. Pixels overlapping with cycleway segments have the maximum influence value of 1.

Figure 5.2 shows the influence values calculated using the POI influence function.

Using the same method, we calculated the influence distance decay for other public transport stops, points of interest and residential buildings.

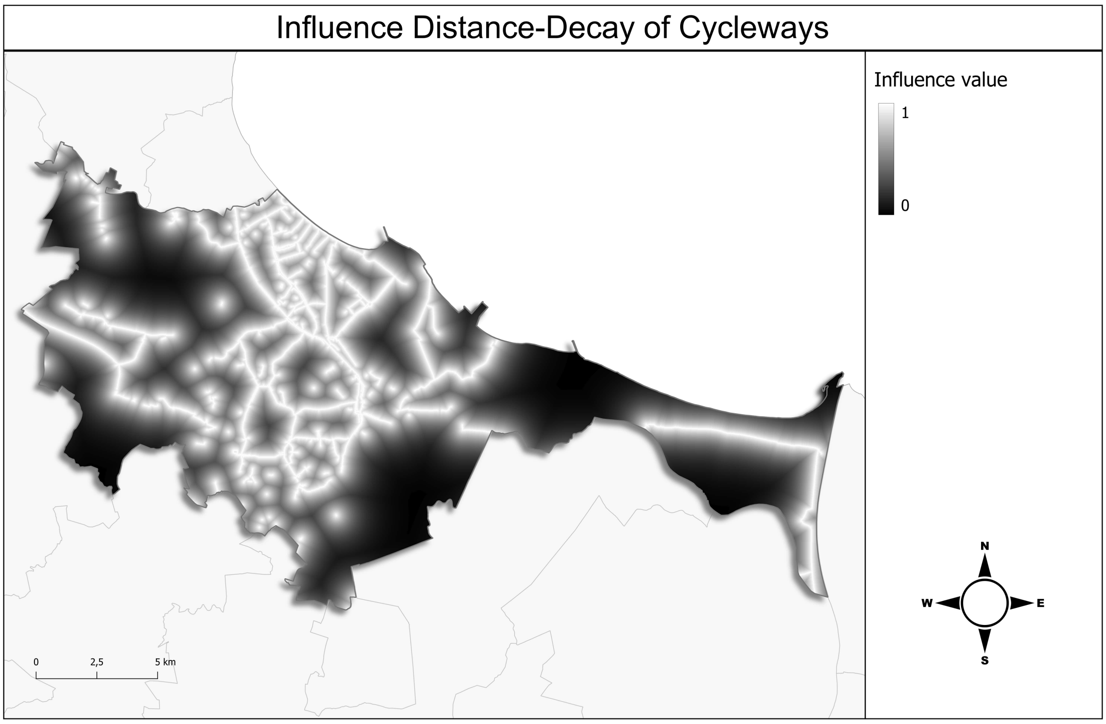

Figure 5.1: Distance decay of the influence of cycleways.

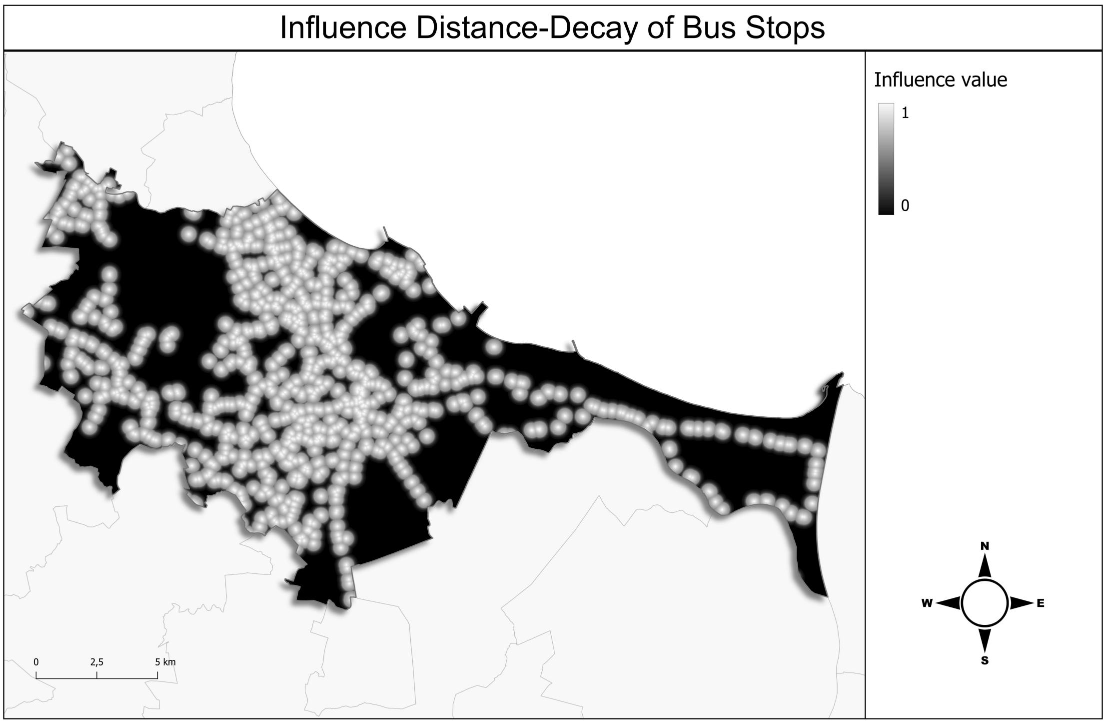

Figure 5.2: Distance decay of the influence of bus stops.

5.2 Factor Weight Assignment

Using the calculated POI distance-decay correlations we assigned weights to each of the factors included in the spatial analysis.

The table 5.2 lists the points of interest that were available in spatial format and had a positive correlation coefficient. We calculated the weights by dividing the coefficient values by 0.4 to scale them within the [0,1] range.

Similarly, the weights for multi-family residential buildings and cycleways were derived from the correlation coefficients of the Residential: Apartments and Amenity: Bike Related categories.

Existing research shows that railway and tram stops have a stronger influence on bike sharing systems than bus stops [14]. To account for these differences, we adjusted the weights for train and bus stops by multiplying them by 1.2 and 0.9, respectively.

Factor name	Category	Distance-decay correlation	Assigned weight
apartments	Residential: Apartments	0.24	0.60
train stops	Amenity: Transport Related	0.26	0.78
bus stops	Amenity: Transport Related	0.26	0.59
tram stops	Amenity: Transport Related	0.26	0.65
cycleways	Amenity: Bike Related	0.37	0.93

Table 5.1: Assigned weights for other factor categories.

POI name	Category	Distance-decay correlation	Assigned weight
museum	Tourism: Cultural Sites	0.13	0.33
bakery	Shop: Bakery & Pastry; Sweets	0.27	0.68
butcher	Shop: Groceries & Daily Essentials	0.26	0.65
clothes	Shop: Clothing & Fashion	0.28	0.70
convenience	Shop: Groceries & Daily Essentials	0.26	0.65
department_store	Shop: Larger Retail	0.28	0.70
doityourself	Shop: Home Improvement & Decor	0.28	0.70
general	Shop: Groceries & Daily Essentials	0.26	0.65
greengrocer	Shop: Groceries & Daily Essentials	0.26	0.65
hairdresser	Shop: Beauty & Personal Care	0.27	0.68
kiosk	Shop: Misc Services	0.27	0.68
laundry	Shop: Misc Services	0.27	0.68
mall	Shop: Larger Retail	0.28	0.70
optician	Shop: Misc Services	0.27	0.68
stationery	Shop: Misc Services	0.27	0.68
supermarket	Shop: Groceries & Daily Essentials	0.26	0.65
park	Leisure: Parks & Outdoor Recreation	0.08	0.19
pitch	Leisure: Sports & Tracks	0.12	0.30
playground	Leisure: Fun & Social Venues	0.18	0.45
sports_centre	Leisure: Sports & Tracks	0.12	0.30
swimming_pool	Leisure: Sports & Tracks	0.12	0.30
track	Leisure: Sports & Tracks	0.12	0.30
arts_centre	Amenity: Culture & Entertainment	0.23	0.58
atm	Amenity: ATM; Money Related	0.22	0.55
bank	Amenity: Financial	0.25	0.63
bar	Amenity: Restaurants, Cafés & Bars	0.09	0.21
cafe	Amenity: Restaurants, Cafés & Bars	0.09	0.21
car_wash	Shop: Car Related	0.08	0.20
cinema	Amenity: Culture & Entertainment	0.23	0.58
clinic	Amenity: Medical & Pharmacy	0.31	0.78
college	Amenity: Education + Small Schools	0.22	0.55
community_centre	Amenity: Post Office and Post Related, Public Services	0.22	0.55
dentist	Amenity: Medical & Pharmacy	0.31	0.78
doctors	Amenity: Medical & Pharmacy	0.31	0.78
fast_food	Amenity: Quick Eats & Ice Cream	0.27	0.68
food_court	Amenity: Restaurants, Cafés & Bars	0.09	0.21
fountain	Amenity: Culture & Entertainment	0.23	0.58
kindergarten	Amenity: Education + Small Schools	0.22	0.55
library	Amenity: Culture & Entertainment	0.23	0.58
nightclub	Amenity: Culture & Entertainment	0.23	0.58
pharmacy	Amenity: Medical & Pharmacy	0.31	0.78
police	Amenity: Post Office and Post Related, Public Services	0.22	0.55
post_box	Amenity: Post Office and Post Related, Public Services	0.22	0.55
post_office	Amenity: Post Office and Post Related, Public Services	0.22	0.55
pub	Amenity: Restaurants, Cafés & Bars	0.09	0.21
restaurant	Amenity: Restaurants, Cafés & Bars	0.09	0.21
school	Amenity: Education + Small Schools	0.22	0.55
theatre	Amenity: Culture & Entertainment	0.23	0.58
university	Amenity: Education + Small Schools	0.22	0.55
veterinary	Amenity: Medical & Pharmacy	0.31	0.78

Table 5.2: Assigned weights for chosen points of interest.

5.3 Spatial Suitability Score

We calculated the total suitability score for each pixel by summing the weighted influence values of all factors and scaling the result to a [0,1] range.

Figure 5.3 shows the highest suitability score values for dense urban areas with high concentration of points of interest, such as the city centre and areas along the train tracks, including Aleja Grunwaldzka.

Current MEVO stations are located in areas with high suitability score as shown on 5.4.

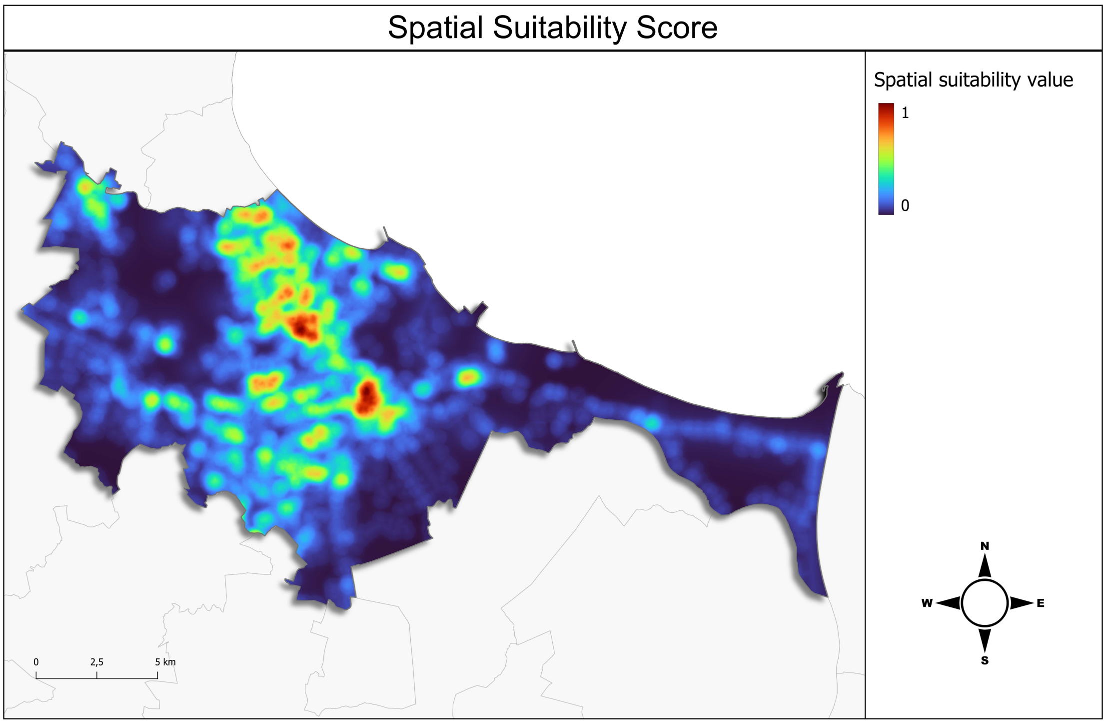

Figure 5.3: Spatial suitability score for MEVO stations.

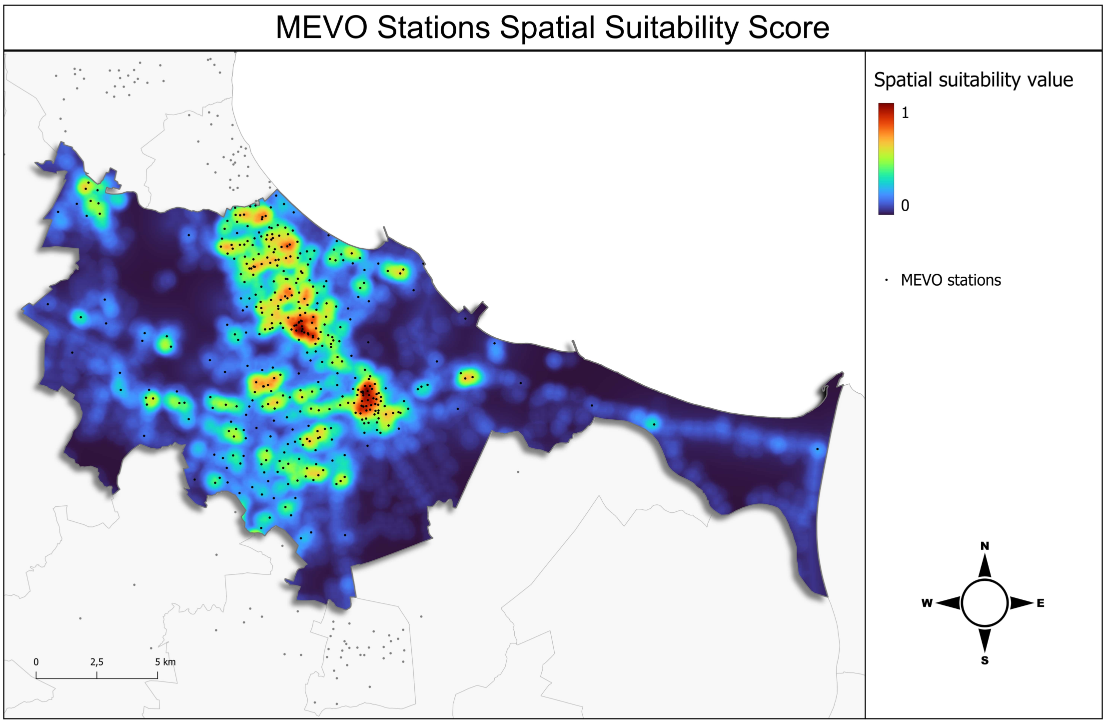

Figure 5.4: Overlay of existing MEVO stations on the spatial suitability score.

5.4 Locations for New MEVO Stations

To find the optimal locations for new MEVO stations we have to consider the placement of existing ones. According to existing research, the optimal distance between bike sharing stations is approximately 300 meters [15]. Therefore, we avoid choosing locations for new stations closer than 200 meters to existing ones, maintaining the distance between them close to optimal. At the same time, demand does not decrease for locations more than 300 meters away from existing stations. New stations should be placed in all suitable locations lacking coverage.

Figure 5.5 shows the demand for new MEVO stations, expressed as a value in the range [0,1]. Locations within 20 meters of existing MEVO stations are assigned the value of 0. This value increases linearly to 1 at a distance of 200 meters and stays constant beyond that distance.

Figure 5.6 shows the spatial suitability score multiplied by the calculated demand values. As a result areas within 200 meters of current stations have their scores greatly reduced. High scores highlight locations with strong suitability and high demand for new MEVO stations.

To calculate optimal new locations for MEVO, we first created buffers around roads and cycleways - 70 meters for roads (to account for varying road widths) and 20 meters for cycleways, allowing stations to be placed in areas without direct road access.

To identify the local maxima, we used the GRASS `r.neighbors` function set to maximum. This divided the buffered area into regions defined by the highest pixel value within each neighbourhood. We found the optimal parameters to be a circle with a radius of 310 meters. This radius prevents crowding of the areas with high suitability score and maintains a balance between high- and low-score regions.

Then we selected the pixels whose suitability score matched the output of `r.neighbours` function, treating the centres of these clusters as proposed station locations. Points with a suitability value below 0.2 were removed to avoid placing stations in low-score areas.

By extracting local maxima of the suitability map within the defined buffers, we were able to determine optimal MEVO station locations, as shown in figure 5.7.

The suggested locations account for areas covered by buildings. However, we did not verify if the selected spaces are available and unobstructed by other infrastructure, so the chosen locations should be treated as a suggestion.

In total we added 336 proposed locations, which corresponds to 51.85% of the number of current MEVO stations in Gdańsk.

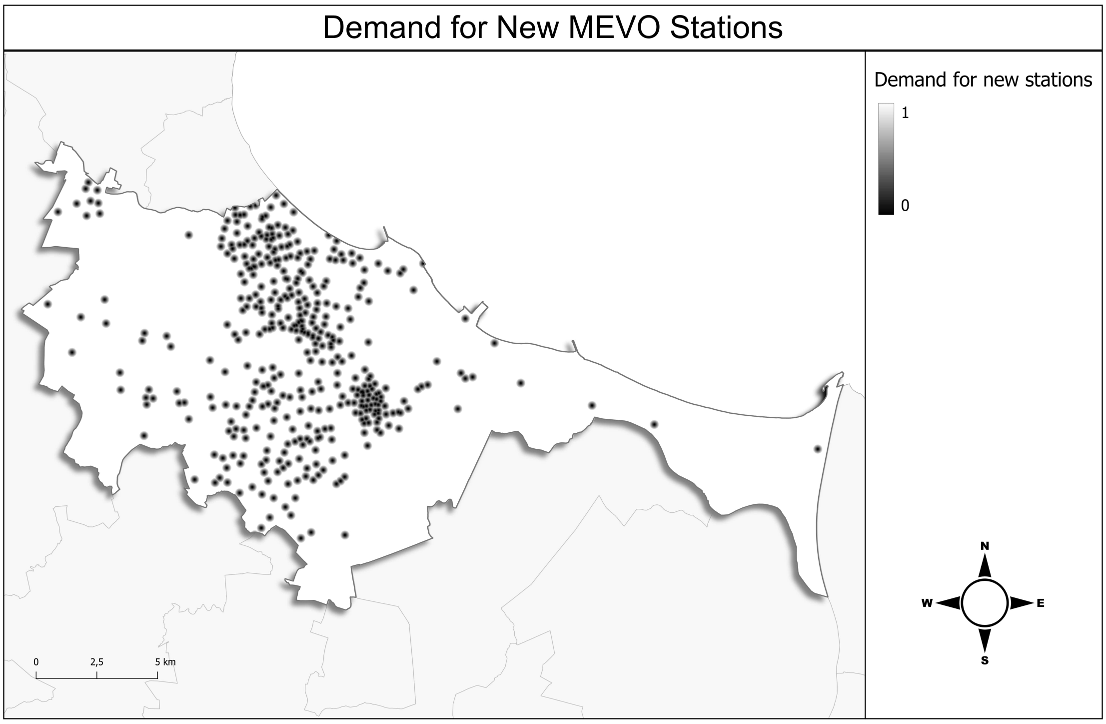

Figure 5.5: Demand for new MEVO stations.

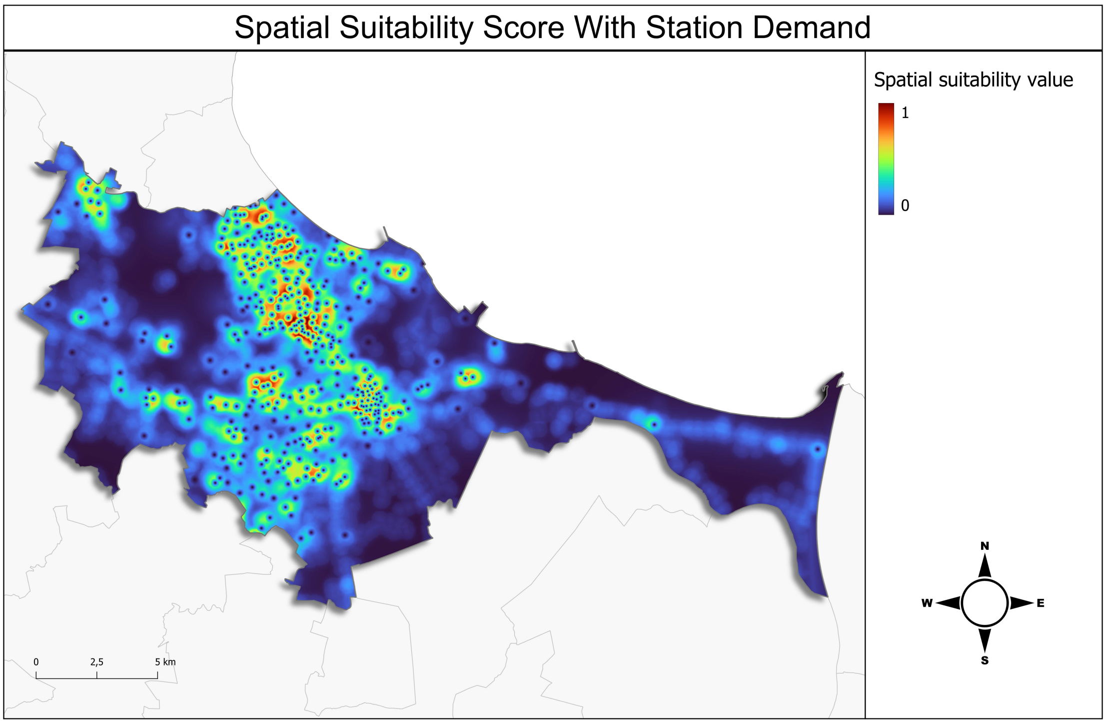

Figure 5.6: Spatial Suitability Score with new station demand.

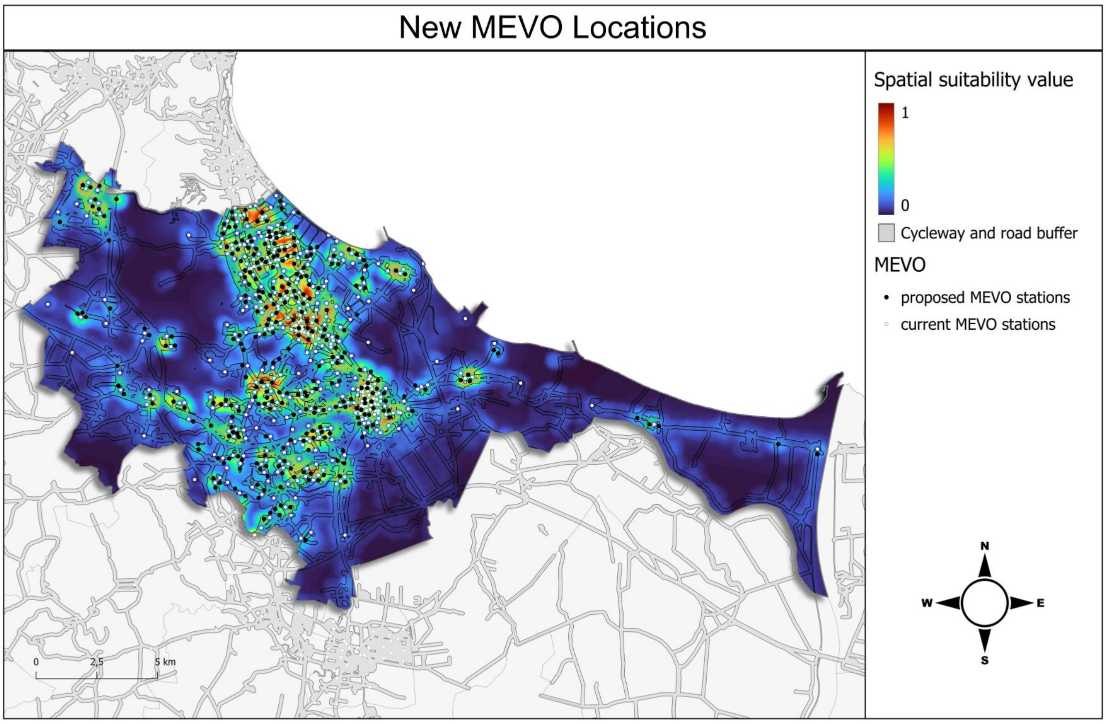

Figure 5.7: Proposed new MEVO locations.

5.5 MEVO Availability

We measure the availability of MEVO stations as the fraction of residential buildings located within 300 meters of the nearest MEVO station.

Figure 5.8 shows the residential areas located within 300 meters of MEVO stations. The total area covered by current MEVO stations corresponds to 61.82% of the total area of residential buildings in Gdańsk.

If we include our new proposed station locations, the total area covered by stations increases by 13.29%, and now corresponds to 70.04% of the total area of residential buildings, as shown on 5.9.

Overall the proposed new locations had a relatively small impact on MEVO availability, despite the station count increasing by over 50%. The is likely because our analysis prioritises points of interest, transport infrastructure and densely populated areas. As a result, low-population areas far from the city centre receive low suitability score.

To further increase MEVO availability, we need to compromise between placing stations in high-demand locations and increasing coverage in low-population areas currently lacking access to MEVO.

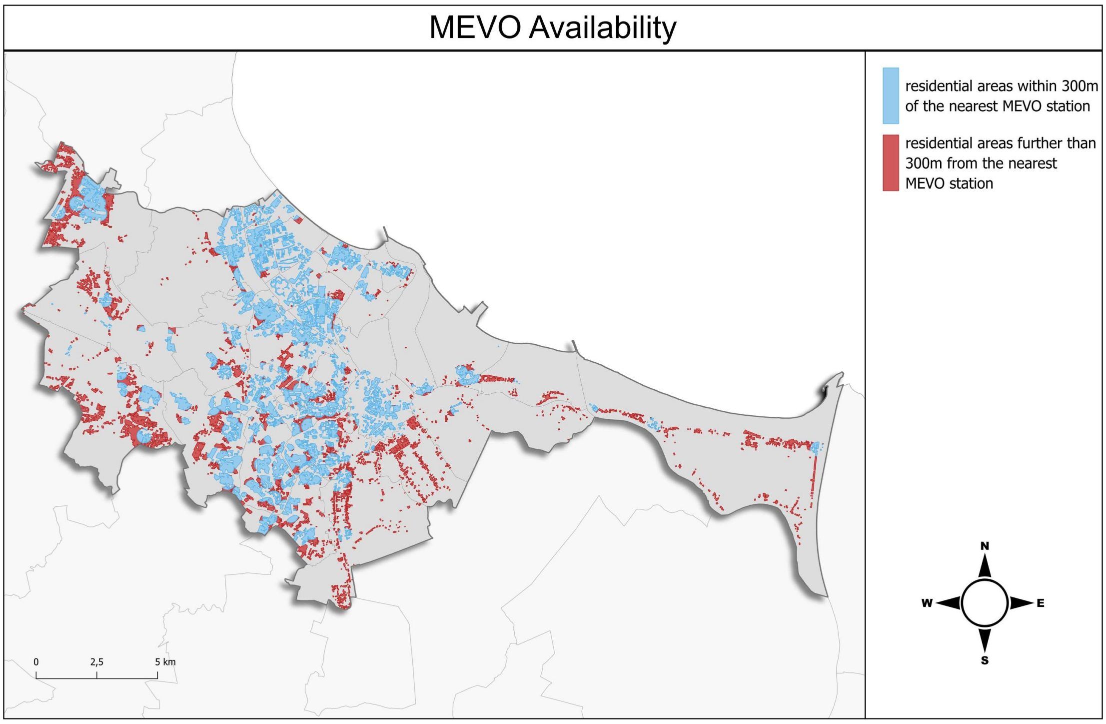

Figure 5.8: MEVO availability.

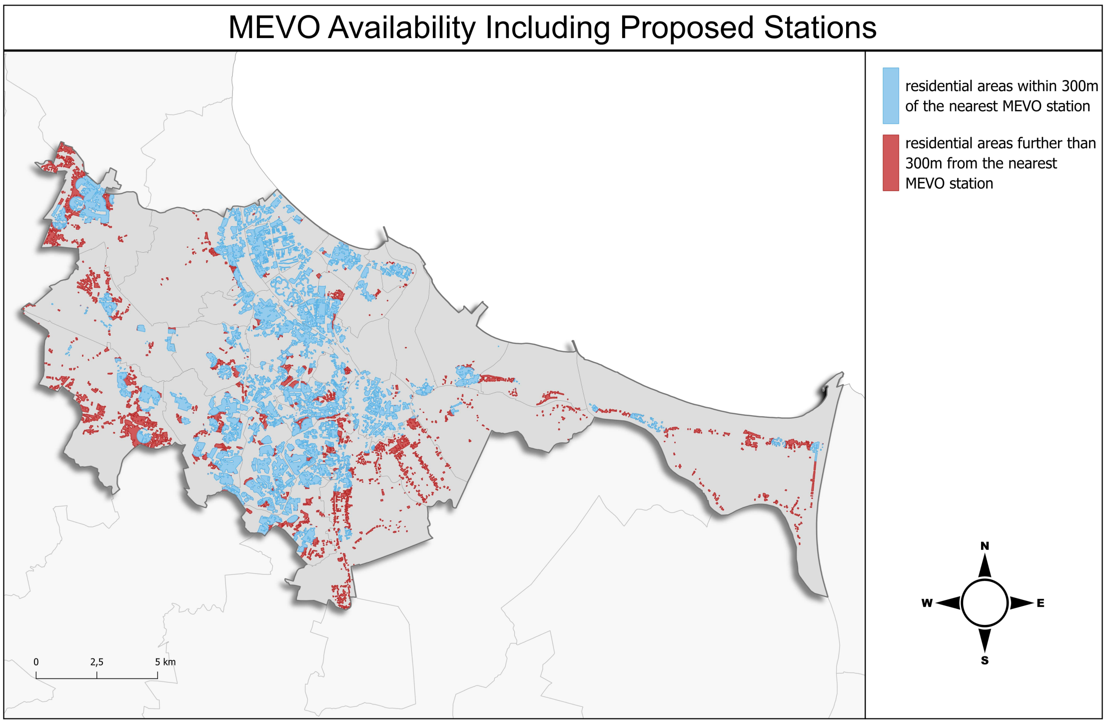

Figure 5.9: MEVO availability with proposed new locations.

6. Conclusions

The aim of this study was to support the development of the MEVO bike-sharing system in Gdańsk. The work combined two complementary approaches: data analysis of system usage and GIS-based spatial modelling.

In the analytical part, we examined station flows, temporal activity trends, and the patterns in bike rentals. The analysis included regression modelling (OLS and GWR) to quantify the effects of nearby POIs and other variables on bike rentals. During, which we found that including spatial variation with GWR provided a better model fit than OLS.

In the GIS analysis part, we evaluated spatial suitability for new station placement and analysed the impact of points of interest and surrounding infrastructure. We proposed 336 new station locations, improving station availability in residential areas by 13.29%.

The study was limited by outdated and inconsistent data sources that do not reflect the current conditions in Gdańsk, as well as the lack of detailed demographic data such as precise population density. We were also unable to verify the physical availability of proposed locations of new stations, which reduces the precision of the recommendations.

Our proposed approach is the integration of usage data analysis with GIS-based spatial analysis, while balancing demand in high-usage areas with the coverage of peripheral areas with low access to the MEVO system.

Future work should rely on updated datasets, include further verification of land availability, and incorporate more detailed analysis of factors affecting station popularity. Further improvements should also focus on optimisation of bike relocation and monitoring bike availability. The analysis should also be extended to all areas covered by the MEVO system and, in the long term, expanded to introduce MEVO to the whole Pomeranian Voivodeship.

Bibliography

[1] Clockston R. L. M., Rojas-Rueda D., *Health impacts of bike-sharing systems in the U.S.*, Environmental Research, 2021.

[2] Hamilton T. L., Wichman C. J., *Bicycle infrastructure and traffic congestion: Evidence from DC's Capital Bikeshare*, 2016.

[3] Bieliński T., Kwapisz A., Ważna A., *Bike-sharing systems in Poland*, Sustainability, 2019.

[4] EU Urban Mobility Observatory, *One of EU's largest electric bike-sharing systems launched in Gdańsk–Gdynia–Sopot Metropolitan Area*, online: `https://urban-mobility-observatory.transport.ec.europa.eu/news-events/news/one-eus-largest-electric-bike-sharing-systems-launched-gdansk-gdynia-sopot-metropolitan-area-2019-04-11_en`, accessed: 01.11.2025.

[5] Kabra A., Belavina E., Girotra K., *Bike-share systems: Accessibility and availability*, DOI: `https://doi.org/10.1287/mnsc.2019.3407`, 2019.

[6] Teschke K., Chinn A., Brauer M., *Proximity to four bikeway types and neighborhood-level cycling mode share of male and female commuters", Journal of Transport and Land Use, Vol. 10, No. 1*, DOI: `https://doi.org/10.5198/jtlu.2017.943`, 2017.

[7] Notelaers L., Verstraete J., Vansteenwegen P., *A travel demand modeling framework based on OpenStreetMap*, 2024.

[8] Fishman E., *Bikeshare: A Review of Recent Literature*, Transport Reviews, Vol. 36, No. 1, pp. 92–113, DOI: `https://doi.org/10.1080/01441647.2015.1033036`, 2016.

[9] Bachand-Marleau J., Lee B.H.Y., El-Geneidy A.M., *Better understanding of factors influencing likelihood of using shared bicycle systems and frequency of use*, DOI: `https://doi.org/10.3141/2314-09`, 2012.

[10] Eren E., Uz V. E., *A Review on Bike-Sharing: The Factors Affecting Bike-Sharing Demand*, Sustainable Cities and Society, DOI: `https://doi.org/10.1016/j.scs.2019.101882`, 2019.

[11] Wang X., Cheng Z., Trepanier M., Sun L., *Modeling bike-sharing demand using a regression model with spatially varying coefficients*, Preprint submitted to Elsevier, 2021.

[12] Lu B., Charlton M., Harris P., Fotheringham A. S., *Geographically weighted regression with a non-Euclidean distance metric: a case study using hedonic house price data*, International Journal of Geographical Information Science, DOI: `https://doi.org/10.1080/13658816.2013.865739`, 2014.

[13] Li A., Zhao P., Huang Y., Gao K., Axhausen K. W., *An empirical analysis of dockless bike-sharing utilization and its explanatory factors: Case study from Shanghai, China*, Journal of Transport Geography, Vol. 88, DOI: `https://doi.org/10.1016/j.jtrangeo.2020.102828`, 2020.

[14] Jaber A., Abu Baker L., Csonka B., *The Influence of Public Transportation Stops on Bike-Sharing Destination Trips: Spatial Analysis of Budapest City*, DOI: `https://doi.org/10.3390/futuretransp2030038`, 2022.

[15] Banerjee S., Kabir M. M., Khadem N. K., Chavis C., *Optimal locations for bikeshare stations: A new GIS based spatial approach* , DOI: `https://doi.org/10.1016/j.trip.2020.100101`, 2020.

[16] OpenStreetMap community. *Map features*. OpenStreetMap Wiki, online: `https://wiki.openstreetmap.org/wiki/Map_features`, accessed: 19.11.2025.

[17] Miasto Gdańsk, *Gdańsk w liczbach*, online: `https://www.gdansk.pl/gdansk-w-liczbach/mieszkancy,a,108046`, accessed: 23.11.2025.

[18] OpenStreetMap, *OpenStreetMap*, online: `https://www.openstreetmap.org`, accessed: 23.11.2025.

[19] MEVO, *Station Information – Open Data*, online: `https://rowermevo.pl/en/open-data/realtime`, endpoint: `station_information.json`, accessed: 09.09.2025.

[20] Gdańsk Open Data, *Granice dzielnic w Gdańsku - Shapefile*, online: `https://dane.gov.pl/pl/dataset/1821`, accessed: 23.11.2025.

[21] Open-Meteo, *Historical Weather Data API*, online: `https://open-meteo.com`, accessed: 23.11.2025.

[22] Geoportal gov, Państwowy Rejestr Granic (PRG), Baza danych obiektów topograficznych (BDOT10k), online: `https://www.geoportal.gov.pl/`, accessed: 30.07.2025.

A. OSM Subcategory Mapping

This appendix lists the mapping of OpenStreetMap tags to the subcategories used in the analysis of Points of Interest (POIs) in Gdańsk.

They are structured in a way:
Category we are using
{ osm tag category : osm tags list }

```
Shop: Groceries & Daily Essentials
{'shop': ['general', 'convenience', 'greengrocer', 'butcher', 'seafood', '
    deli', 'frozen_food', 'kiosk', 'dairy']}

Shop: Larger Retail
{'shop': ['wholesale', 'supermarket', 'department_store', 'mall']}

Shop: Shopping Mall
{'shop': ['mall']}

Shop: Variety Store
{'shop': ['variety_store']}

Shop: Bakery & Pastry; Sweets
{'shop': ['bakery', 'pastry', 'confectionery', 'ice_cream', 'chocolate']}

Shop: Beverages
{'shop': ['coffee', 'tea', 'wine', 'alcohol', 'beverages', 'water']}

Shop: Clothing & Fashion
{'shop': ['clothes', 'shoes', 'bag', 'fashion_accessories', 'leather', '
    tailor', 'fashion', 'jewelry', 'watches', 'second_hand']}

Shop: Beauty & Personal Care
{'shop': ['cosmetics', 'perfumer', 'tattoo', 'hairdresser', '
    hairdresser_supply']}

Shop: Wellness; Nutrition
{'shop': ['herbalist', 'nutrition_supplements', 'medical_supply', 'chemist
    ', 'health_food']}

Shop: Electronics & Tech
{'shop': ['electronics', 'computer', 'mobile_phone', 'printer_ink', '
    optical_instruments', 'hifi']}

Shop: Home Improvement & Decor
{'shop': ['doityourself', 'hardware', 'houseware', 'electrical', 'paint', '
    doors', 'flooring', 'kitchen', 'interior_decoration', 'curtain', '
```

```
    fireplace', 'furniture', 'trade', 'bathroom_furnishing']}

Shop: Hobbies & Entertainment
{'shop': ['music', 'musical_instrument', 'games', 'craft', 'video_games', '
    art', 'toys']}

Shop: Tobacco & Smoke Related
{'shop': ['tobacco', 'e-cigarette', 'cannabis']}

Shop: Sport Related
{'shop': ['bicycle', 'running', 'sports']}

Shop: Misc Services
{'shop': ['travel_agency', 'pawnbroker', 'money_lender', 'copyshop', '
    ticket', 'lottery', 'laundry', 'massage', 'skate', 'security', 'party',
     'outpost', 'gift', 'stationery', 'florist', 'photo']}

Shop: Car Related
{'shop': ['car', 'car_parts', 'car_repair', 'motorcycle']}

Amenity: Restaurants, Cafes & Bars
{'amenity': ['cafe', 'restaurant', 'bar', 'pub']}

Amenity: Quick Eats & Ice Cream
{'amenity': ['fast_food', 'food_court', 'ice_cream']}

Amenity: Transport Related
{'amenity': ['taxi', 'ferry_terminal', 'vehicle_inspection']}

Amenity: Parking
{'amenity': ['parking', 'parking_entrance', 'parking_space']}

Amenity: Bike Related
{'amenity': ['bicycle_parking', 'bicycle_repair_station', 'bicycle_rental
    ']}

Amenity: Medical & Pharmacy
{'amenity': ['clinic', 'doctors', 'dentist', 'pharmacy', 'veterinary', '
    hospital', 'social_facility']}

Amenity: Public Services & Safety
{'amenity': ['police', 'courthouse', 'townhall']}

Amenity: Post Office and Post Related; Packages
{'amenity': ['post_office', 'parcel_locker', 'post_box']}

Amenity: Education
{'amenity': ['school', 'college', 'university', 'kindergarten']}

Amenity: Misc Shool and Education Related
{'amenity': ['dancing_school', 'driving_school', 'language_school', '
```

```
    library', 'music_school']}

Amenity: Culture & Entertainment
{'amenity': ['arts_centre', 'cinema', 'community_centre', 'events_venue', '
    gambling', 'nightclub', 'studio', 'theatre', 'music_venue', '
    planetarium']}

Amenity: Financial
{'amenity': ['bank', 'bureau_de_change']}

Amenity: ATM; Money Related
{'amenity': ['atm', 'money_transfer']}

Amenity: Religion & Worship
{'amenity': ['place_of_worship']}

Amenity: Misc Places
{'amenity': ['fountain', 'marketplace']}

Building: Residential
{'building': ['apartments', 'dormitory', 'detached', 'house', 'residential
    ', 'semidetached_house']}

Building: Parking & Garage
{'building': ['garage', 'parking']}

Office
{'office': ['government', 'company', 'lawyer', 'estate_agent', 'insurance',
     'notary', 'accountant', 'ngo', 'association', 'foundation', 'it', '
    employment_agency', 'architect', 'telecommunication', '
    advertising_agency', 'property_management', 'tax_advisor', 'diplomatic
    ', 'security', 'visa', 'translator', 'educational_institution', '
    consulting', 'coworking', 'financial', 'engineer', 'financial_advisor',
     'construction_company', 'health_insurance', 'therapist', 'physician',
    'publisher', 'politician', 'political_party', 'newspaper']}

Leisure: Parks & Outdoor Recreation
{'leisure': ['park', 'playground', 'picnic_site']}

Leisure: Fun & Social Venues
{'leisure': ['amusement_arcade', 'bowling_alley', 'swimming_pool']}

Leisure: Sports & Tracks
{'leisure': ['track', 'pitch', 'sports_centre', 'sports_hall'], 'building':
     ['stadium', 'sports_hall', 'sports_centre']}

Tourism: Accommodation
{'tourism': ['hotel', 'guest_house', 'hostel', 'apartment']}

Tourism: Cultural Sites
{'tourism': ['museum', 'gallery', 'artwork']}
```

```
Tourism: Attractions
{'tourism': ['viewpoint', 'attraction']}

Public Transport: Stations
{'public_transport': ['station', 'platform']}

Shop: Sport, Hobbies, Smoke, Variety
{'shop': ['bicycle', 'running', 'sports', 'music', 'musical_instrument', '
    games', 'craft', 'video_games', 'art', 'toys', 'variety_store', '
    tobacco', 'e-cigarette', 'cannabis', 'music', 'musical_instrument', '
    games', 'craft', 'video_games', 'art', 'toys', 'variety_store', '
    tobacco', 'e-cigarette', 'cannabis', 'variety_store', 'tobacco', 'e-
    cigarette', 'cannabis']}

Amenity: Post Office and Post Related, Public Services
{'amenity': ['police', 'courthouse', 'townhall', 'post_office', '
    parcel_locker', 'post_box', 'post_office', 'parcel_locker', 'post_box
    ']}

Leisure: Parks & Outdoor Recreation; Social Venues
{'leisure': ['park', 'playground', 'picnic_site', 'amusement_arcade', '
    bowling_alley', 'swimming_pool', 'fountain', 'marketplace', '
    amusement_arcade', 'bowling_alley', 'swimming_pool', 'fountain', '
    marketplace']}

Parking & Garage
{'building': ['garage', 'parking'], 'amenity': ['parking', '
    parking_entrance', 'parking_space']}

Shop: Hobbies, Smoke, Variety
{'shop': ['music', 'musical_instrument', 'games', 'craft', 'video_games', '
    art', 'toys', 'variety_store', 'tobacco', 'e-cigarette', 'cannabis', '
    variety_store', 'tobacco', 'e-cigarette', 'cannabis']}

Amenity: Education + Small Schools
{'amenity': ['school', 'college', 'university', 'kindergarten', '
    dancing_school', 'driving_school', 'language_school', 'library', '
    music_school', 'dancing_school', 'driving_school', 'language_school', '
    library', 'music_school']}

Building: Residential
{'building': ['apartments', 'dormitory', 'residential']}

Residential: Houses
{'building': ['house', 'detached', 'semidetached_house']}

Retail / Shops
{'shop': ['art', 'bag', 'bathroom_furnishing', 'bicycle', 'butcher', '
    cannabis', 'car', 'car_parts', 'car_repair', 'chemist', 'clothes', '
    computer', 'convenience', 'copyshop', 'cosmetics', 'craft', 'curtain',
```

```
    'dairy', 'deli', 'department_store', 'doityourself', 'doors', 'e-
    cigarette', 'electrical', 'electronics', 'fashion', '
    fashion_accessories', 'fireplace', 'flooring', 'florist', 'frozen_food
    ', 'furniture', 'games', 'general', 'gift', 'greengrocer', 'hairdresser
    ', 'hairdresser_supply', 'hardware', 'health_food', 'herbalist', 'hifi
    ', 'houseware', 'interior_decoration', 'jewelry', 'kiosk', 'kitchen', '
    laundry', 'leather', 'lottery', 'mall', 'massage', 'medical_supply', '
    mobile_phone', 'money_lender', 'motorcycle', 'music', '
    musical_instrument', 'nutrition_supplements', 'optical_instruments', '
    outpost', 'paint', 'party', 'pawnbroker', 'perfumer', 'photo', '
    printer_ink', 'running', 'seafood', 'second_hand', 'security', 'shoes',
     'skate', 'sports', 'stationery', 'supermarket', 'tailor', 'tattoo', '
    ticket', 'tobacco', 'toys', 'trade', 'travel_agency', 'variety_store',
    'video_games', 'watches', 'wholesale']}
```

```
Food & Drink
{'shop': ['alcohol', 'bakery', 'beverages', 'chocolate', 'coffee', '
    confectionery', 'ice_cream', 'pastry', 'tea', 'water', 'wine'], '
    amenity': ['bar', 'cafe', 'fast_food', 'food_court', 'ice_cream', 'pub
    ', 'restaurant']}
```

```
Transport
{'amenity': ['bicycle_parking', 'bicycle_rental', 'bicycle_repair_station',
     'ferry_terminal', 'parking', 'parking_entrance', 'parking_space', '
    taxi', 'vehicle_inspection'], 'building': ['garage', 'parking'], '
    public_transport': ['platform', 'station']}
```

```
Medical & Education
{'amenity': ['clinic', 'college', 'dancing_school', 'dentist', 'doctors', '
    driving_school', 'hospital', 'kindergarten', 'language_school', '
    library', 'music_school', 'pharmacy', 'school', 'social_facility', '
    university', 'veterinary']}
```

```
Public / Community Services
{'amenity': ['arts_centre', 'atm', 'cinema', 'community_centre', '
    courthouse', 'events_venue', 'fountain', 'gambling', 'marketplace', '
    money_transfer', 'music_venue', 'nightclub', 'parcel_locker', '
    place_of_worship', 'planetarium', 'police', 'post_box', 'post_office',
    'studio', 'theatre', 'townhall']}
```

```
Office / Professional Services
{'amenity': ['bank', 'bureau_de_change'], 'office': ['accountant', '
    advertising_agency', 'architect', 'association', 'company', '
    construction_company', 'consulting', 'coworking', 'diplomatic', '
    educational_institution', 'employment_agency', 'engineer', '
    estate_agent', 'financial', 'financial_advisor', 'foundation', '
    government', 'health_insurance', 'insurance', 'it', 'lawyer', '
    newspaper', 'ngo', 'notary', 'physician', 'political_party', '
    politician', 'property_management', 'publisher', 'security', '
    tax_advisor', 'telecommunication', 'therapist', 'translator', 'visa']}
```

```
Residential
{'building': ['apartments', 'detached', 'dormitory', 'house', 'residential
    ', 'semidetached_house']}

Leisure & Sports
{'building': ['sports_centre', 'sports_hall', 'stadium'], 'leisure': ['
    amusement_arcade', 'bowling_alley', 'park', 'picnic_site', 'pitch', '
    playground', 'sports_centre', 'sports_hall', 'swimming_pool', 'track']}

Tourism & Culture
{'tourism': ['apartment', 'artwork', 'attraction', 'gallery', 'guest_house
    ', 'hostel', 'hotel', 'museum', 'viewpoint']}
```