Geospatial Data Science: Flow analysis of the Oslo city bikes network

Manuel Knepper & Niclas Classen

2024-05-24

Abstract

Cycling is an important mode of transport in many cities, with City Bikes being a public alternative to having to own a bike Banister (2005). This is also the case for the City of Oslo, which we base this analysis on Oslo (2024a). Identifying how people use the bike network, and which stations are frequented is valuable to many different stakeholders in the planning and operating process. We investigate usage patterns in the network, flow between stations, correlations in the station network, and identify potential locations for new stations. Here we show that traffic flow is mostly focused around the city centre, which is also supported by clustering of stations with a lot of traffic in the central districts. Stations that are located on the outskirts of the city have less traffic. Usage of the network varies between weekdays and the weekend, which might be attributed to people using bikes for their work commute. Our proposed locations for new stations increase density in the network, to make bikes more accessible to a higher number of people. The code for this analysis can be found here: `https://github.com/niclasclassen/oslo-city-bikes-network-flow-analysis`.

1 Introduction

The Oslo City Bike Network represents a well-established supplement to public transport mainly used for short distances Oslo (2024a). The aim of this paper is to answer the following research question: "How does traffic flow in the Oslo City Bikes Network, and what are typical cycling patterns?" More specifically, we look into the flow of bike rides between bike stations and districts in Oslo. Furthermore, we analyse the general usage habits. Additionally, we conduct a spatial autocorrelation analysis of where people start or end their bike rides. Finally, we propose new locations for bike stations to increase the density of the bike network.

2 Background

Cycling as a mode of transport is becoming more and more important in the cities of the 21st century. This is due to the multiple advantages it has over other forms of personal transport, most of which are motorized Banister (2005):

- no carbon emissions, leading to a reduced impact on the environment
- less noise pollution
- smaller infrastructural costs, especially compared to car infrastructure
- higher throughput of travellers, leading to less traffic

A lot of cities worldwide have recognized these advantages and have started building more bike-friendly infrastructure Szell et al. (2022), including the offering of so called "City Bikes" which can be rented at specific stations. Oslo is no exception to this development, having opened their city bike network in 2002, and changing to their current setup in 2016 Tight (2018).

Gaining insight into how these networks are used, and which stations or routes are the most popular is valuable information for the operating companies and city planners alike, whether for where to open new stations or allocate bikes, or to get data on where bike infrastructure might be needed within the city. Different geospatial methods can be of great use in this endeavour.

3 Data

3.1 Data acquisition

We acquire the data for rides in the Oslo City Bikes Network from their official website Oslo (2024b). Monthly data is available in either .csv or .json format. We conduct our analysis on data from April 2019 to April 2024, to get a better understanding how behaviour in the network might change over time and to cancel out possible effects caused by seasonality in the data. For certain parts of our analysis we limit the amount of data to the last year (May 2023-April 2024) to reduce computational complexity and get an overview over the latest usage patterns. Each row in the dataset represents one ride, for each of these rides we have information about the start- and endtime, the start- and endstations with their corresponding latitude and longitude and their associated ids, as well as the duration of the trip and descriptive information about the stations. All columns of the dataset are listed in A.

To analyse clustering and flow between districts of Oslo we use district data provided by Geonorge (2024). The data is in .geojson format, and all districts are encoded as polygons. Furthermore, we use data from OpenStreetMap by utilizing the python package OSMnx. This data is used to make sure that the proposed locations for new stations fulfill certain requirements, e.g. being on land. All datasets are derived from official data sources, which results in high quality and reliability.

3.2 Data processing

First, we aggregate the monthly ride data to one dataset. To keep track of the time period we add two new columns, one for the month and one for the year. Since the original file-format does not entail a geometry data point, we create two subsequent GeoDataFrames. The first one contains information about the station network, with each row corresponding to one station. We add a geometry column which is filled with a point geometry that we calculate with the latitude and longitude and the corresponding station id.

Secondly, we create a dataframe with all individual rides, in which we use the point geometries for the start- and endstation to create a linestring geometry that gives us the general direction of each ride. Since we only have access to the start- and endpoint of each ride we can't recreate the entire ride, but have to focus on general traffic between stations, without taking into account what routes cyclists actually took to reach their destination.

From our data of the districts in Oslo, we remove the ones that do not have at least one station in their area. An issue we encountered was that the Coordinate Reference System (CRS) with which the geometries were calculated was not provided. After some trial and error we found out that the data corresponds to European Petroleum Survey Group (EPSG) code 32633.

4 Results

4.1 General

4.1.1 Station network growth

The City Bikes bicycle network in Oslo changes over the years that were part of this project. In 2019, the network consists of 254 stations, this number grows to 268 in 2024, which is attributed to more new stations opening than old stations closing. The number of stations is derived from stations that are either the start or the end of an individual trip in the respective years. Our numbers slightly differ from the ones displayed on the website, this might be due to stations being temporarily shut down for maintenance Oslo (2024a). The changes in the network from 2019 to 2024 are visualized in Figure 1, year-to-year changes are listed in the Appendix under B.

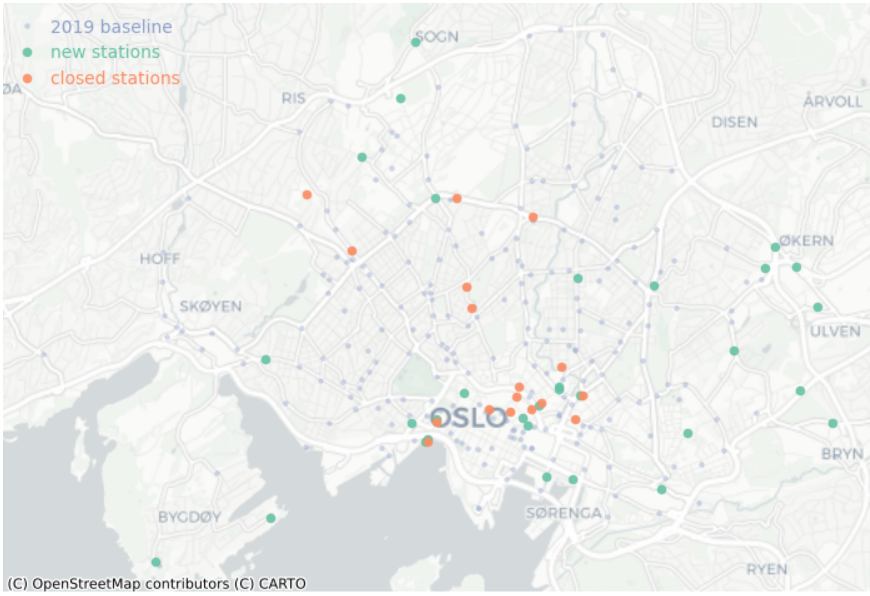

Figure 1: Changes in station network from April 2019 to April 2024

Often times new stations replace old stations in very close proximity, indicating that the changes might have been motivated by changes in the road- or bicycle-infrastructure in that respective area. We can also observe more growth away from the city centre, as the network expands further into more remote districts, particularly in the east of the city.

4.1.2 Trip duration

To get a first impression of how the Oslo City Bike Network is used, we analyze the average trip duration of all rides from May 2023 to April 2024, as shown in Figure 2. We observe that more than 75% of trips are shorter than 15 minutes, indicating that most users rent the bikes for short trips, and use motorized individual or public transport for longer trips around the city. This usage pattern is in line with the intended use of city bikes, which should be an environmentally friendly addition to the already existing forms of transportation Oslo (2024a).

In addition, to gain a more comprehensive understanding of trip durations, we look into the average duration of rides for each weekday. Our results are visualized in Figure 2. It is evident that the average trip duration increases as we approach Sunday, with the duration on Sundays being around 3 minutes longer than on Mondays.

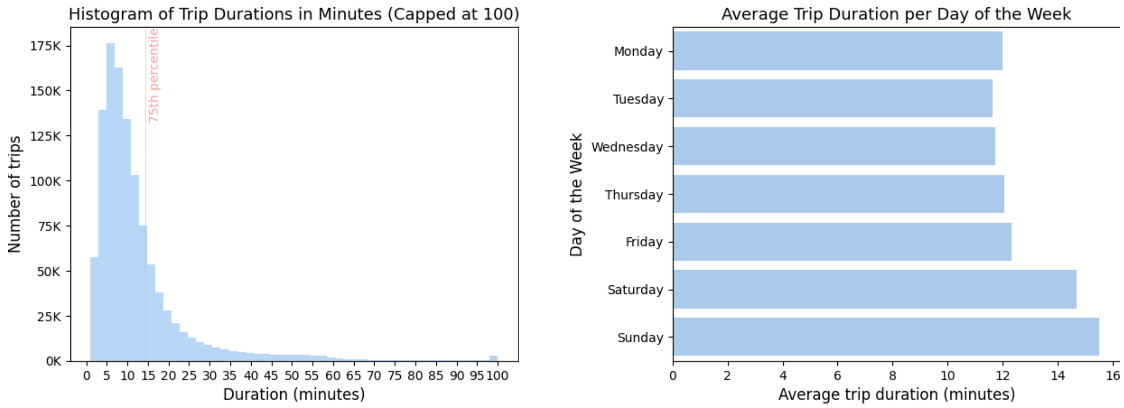

Figure 2: Distribution of trip duration and duration of rides on each day of the week May 2023-April 2024

4.1.3 Usage differences from weekdays to the weekend

Besides the average trip duration from the previous analysis, we are also interested in the actual number of rides during the day. We assume different usage patterns during the week and on weekends. Therefore, we calculate the number of trips taken for every hour of the day, split up between rides that occurred during the week or on the weekend (see Figure 3). From that, we see two spikes for the weekdays, one in the morning and one in the afternoon. This is most likely linked to people using the city bikes to commute to work and returning home. On the weekend, the usage is more evenly distributed throughout the day, with a plateau around 11 AM to 3 PM.

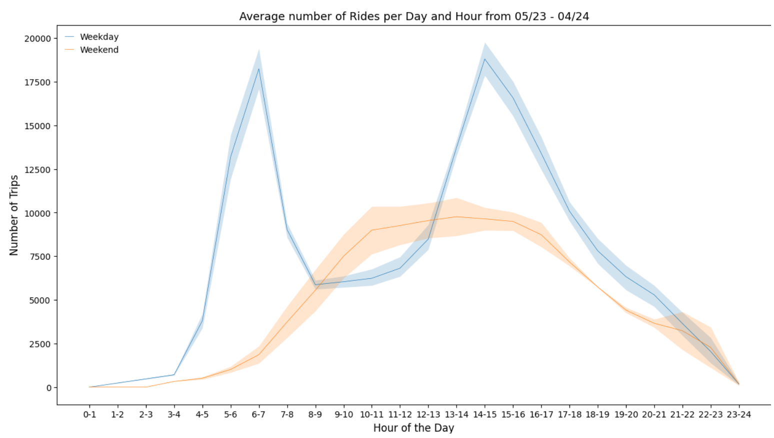

Figure 3: Distribution of number of trips for time of the day divided by weekdays and weekend

4.2 Flow

4.2.1 General Flow between stations

To gain more insight into how users move between stations we conduct analysis on the flow in the station network. For this analysis we focus on data from May 2023 to April 2024. We sum up all rides between one pair of stations, to get the absolute number of rides that occurred from station A to station B and vice versa within the scope of this analysis. The highest amount of rides between stations were taken from "Sjøsiden 1 ved trappen" to "Vippetangen vest" with 1190 rides, while there are multiple combinations of stations with only a single ride between them. To account for this discrepancy we normalize the number of rides, the resulting flow network is plotted in Figure 4, with the transparency of the lines set to the normalized number of rides for that route, to highlight frequently travelled routes.

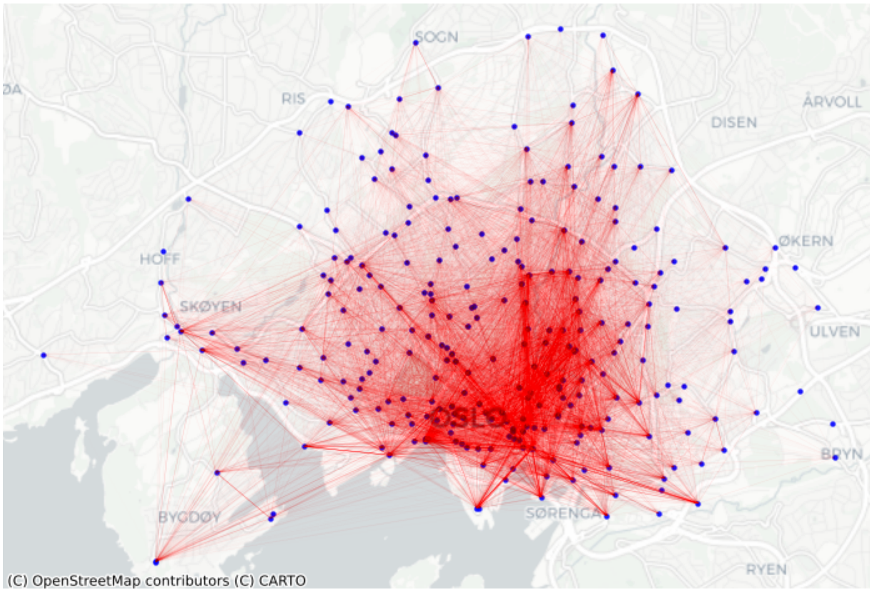

Figure 4: Flow in station network from May 2023 to April 2024

Most of the rides either start or end at a station close to the city center, while some of the more remote stations only have a very limited amount of traffic, which is expected assuming that more people live in the city center, and bikes might also be used by tourists who most likely primarily frequent central districts of cities.

4.2.2 Flow to and from new stations

We also investigate how the flow changes while the network expands. As explained in 4.1.1, we can observe an expansion of the station network over the time period that we chose to base our analysis on. We model how those new stations fit into the already existing network by plotting the flow between new stations and old stations as visualized in Figure 5.

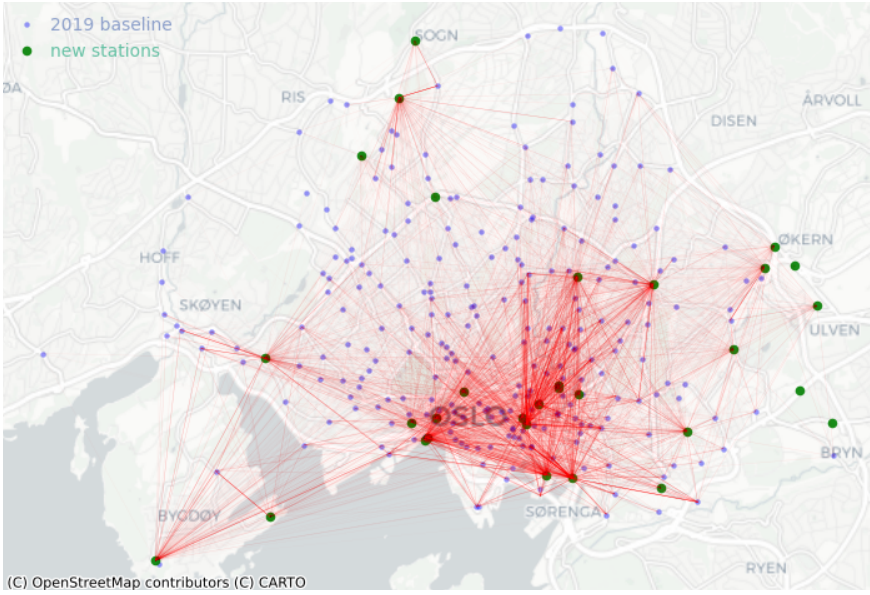

Figure 5: Flow between new and old stations

The results for this analysis are similar to the ones we got when taking all stations into account: the stations that are closer to the city center receive more traffic than the ones more on the outskirts of the city. But we can still observe that some of the more remote stations, particular in Bygdøy and the east of Sørenga, have a comparatively high amount of traffic, possibly indicating that the decision to place new stations there was correct, even though this might also be a case of induced demand after improving the biking infrastructure by adding stations to the network Marqués et al. (2015).

4.2.3 Flow between districts

We can observe from Figure 4 that there is a lot of activity between stations located in the city centre. To get a better understanding of how people move within areas in the city we use the official districts of the City of Oslo. The movement between the eight districts with the most amount of rides, and also within districts is visualized in Figure 6.

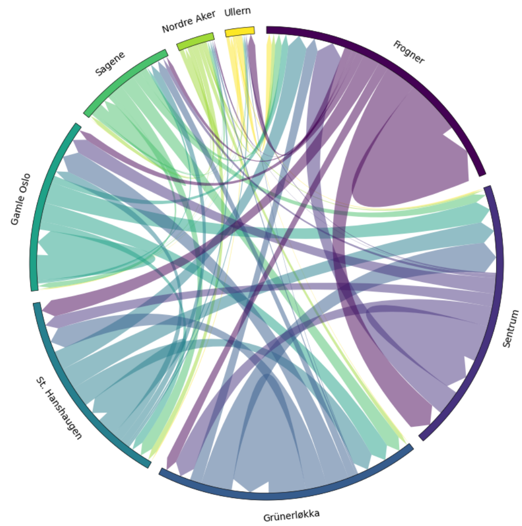

Figure 6: Flow between top 8 districts

Grünerløkka has the most people leaving the district, mostly to either Sentrum or St. Hanshaugen. Sentrum, the most central district, has the most people taking trips within the district. Overall we again observe more flow in the central districts closer to the city centre, which alligns with the results we got on station level.

4.3 Spatial autocorrelation

From our previous flow analysis, we observed that most of the rides happen around the city center. However, it is still uncertain if there are specific clusters of stations with either high or low traffic. We visualize the amount of rides for each district with choropleth maps in Appendix B.To further investigate this, we conduct a spatial autocorrelation analysis. First, we must define connectivity and weights towards neighboring stations in order to calculate the spatial weights matrix. Since we are using points as the geometry, we opted for distance-based connectivity where stations within a given radius are considered to be neighbors. To define our radius, we calculated all distances between all combinations of the two closest stations and selected the highest value to ensure that each station has at least one neighbor. Another option would have been k-nearest neighbors, but we wanted to account for the fact that some stations have more nearby stations than others. We standardize the weights so that the sum of all neighbors equals 1, allowing for comparability between areas with different numbers of neighbors.

When we look at the spatial autocorrelation for the number of ending rides per stations from a global perspective, we obtain a Moran's I of 0.27 with a p-value of 0.001, indicating a small positive correlation. However, this does not provide information about where clusters are located. Therefore, we utilized Local Indicators of Spatial Association (LISA), the results of which are shown in Figure 7.

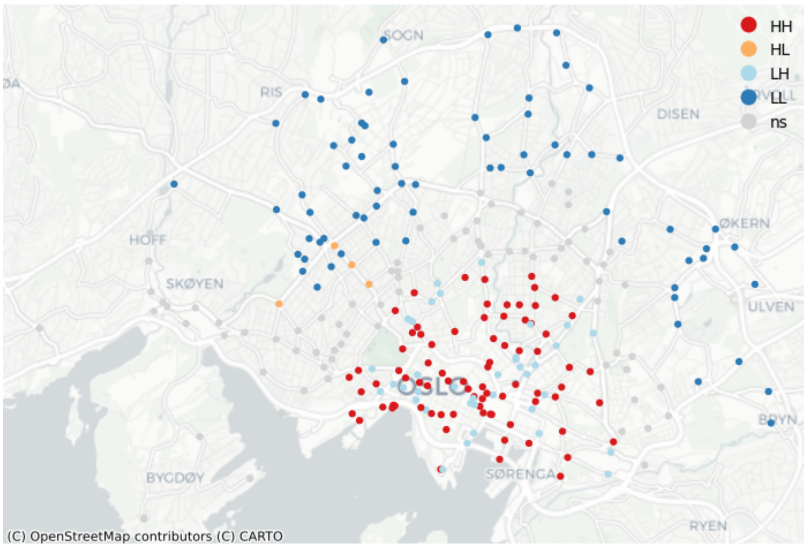

Figure 7: LISA analysis of stations ending rides May 2023 - April 2024

We observe a cluster of high-high values around the city centre, indicating that stations with a high number of ending rides are surrounded by other stations with a high number of ending rides. Interesting is that there are also some stations in the city center with few ending rides but surrounded by stations with a lot of ending rides (low-highs), which may be due to differences in accessibility. Furthermore, we see that the outer ring of the bike network presents a cluster of stations with only a few ending rides that are surrounded by other stations with few ending rides (low-lows). This is in line with the results from our flow analysis, indicating that most rides occur in the center of the city. In the intermediate area between the high-high clusters in the center and the low-low clusters on the outskirts of the city, we can observe a distinctive ring of stations with no significant correlation in either direction of the spectrum, acting as a "bufferzone" between the two clusters. We also investigate clustering when considering stations as the starting point of rides, the results mostly mirror the results obtained when stations are ending points. A visualization can be found in the Appendix under B.

4.4 Suggestion for new stations

Based on the analysis we conduct it is evident that the biggest activity in the network happens in the center of the city, where the density of the network is high. We can also observe that the network is continually growing, with new stations being added. Most new stations are located outside the city center, in areas where station coverage is not as dense. Using this information we build a simple algorithm that suggests possible locations for new stations. The goal of this algorithm is to increase the density of the network, which is one of two main factors besides the distance coverage of the network, which determine how many people are within easy reach of a public transport network Rollison (2023). We create polygons from every station with three other stations, then calculate the centroid for each of these polygons. Afterwards we propose the centroid as the next location for a station, that is the furthest away from its three closest neighbours, provided that that centroid is placed on land and not in the ocean. Whether a point is located on landmass or in the ocean is provided by OpenStreetMap. After identifying a new station we add that station to the existing ones and calculate all polygons again, to make sure we do not get suggestions that are very close to each other. We run the algorithm until we get 10 new suggested stations. The suggestions for new stations are visualized in Figure 8.

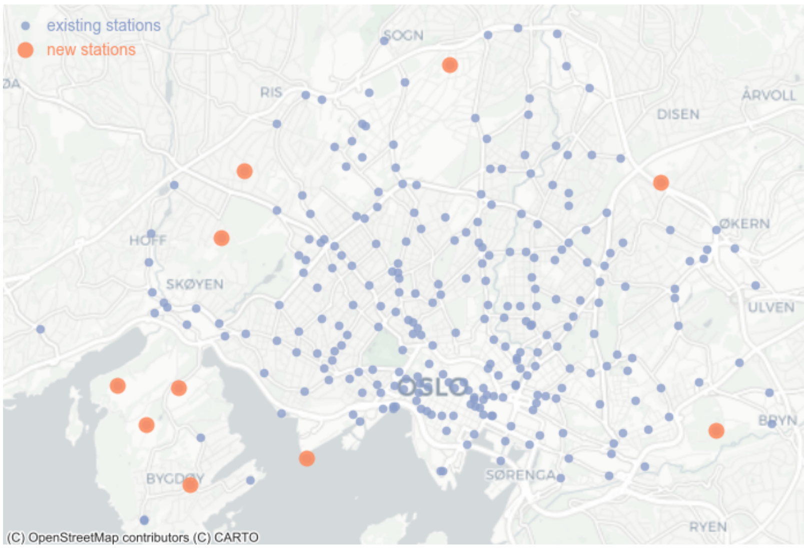

Figure 8: Suggestions for new stations

This algorithm is clearly not perfect, and the results should be interpreted as possible areas for new stations rather than a specific location. Further improvements could be made by taking population densities and land use in the area that surrounds proposed locations for stations into account. The algorithm also only increases the density of a network, and does not extend the network outside of the existing boundaries. In the case of the Oslo City Bike network, this means that new stations would not reach out into other districts of the city that are currently not served by the station network. Our suggestions should be seen as first ideas, that then need to be further investigated.

5 Discussion

Our results offer insight into how people used the City Bike network in Oslo in the timeframe of our analysis. We can show that there are clusters of stations that have more traffic than others, which seems largely due to the location of those stations, with stations in the city centre attracting more traffic than stations that are located on the outskirts of the city. This also holds true when taking flow between stations into account.

Our findings cannot be generalized for other cities, but the methods and code can be used to analyze station networks in other cities, as long as the data structure is similar. We would expect that some of our findings do hold true in a more general setting, but since we do not consider additional factors that might contribute to the usage of bikes within a city, like biking infrastructure, other transport options, or the general topology of the city, we cannot confidently make these claims.

One shortcoming with our analysis is not taking additional data sources into account, for example data from OpenStreetMap to check for areas that have a lot of bicycle infrastructure, or general topology data of the City of Oslo. Furthermore, we could look at population densities in specific areas when proposing new stations. We decided not to include these additional data sources because of time constraints for this project.

We encountered some issues with plotting some of our maps, which was the result of incorrect CRS assignments to some geodataframes. A learning was to make sure that all the CRS' that we use for the geodataframes are the same, to avoid compatibility issues further down in the analysis. We had some issues with long running times for some of the code depending on which hardware we used, which was one reason why we restricted the scope to only one year for some parts of the analysis.

6 Conclusion

Flow in the City Bikes network in Oslo is heavily clustered in the city centre. This holds true both when analysing the amount of rides on station level, as well when considering the flow between districts of the city. Districts with a lot of incoming and outgoing rides are typically closer to the city centre, which hints at more usage from tourists or people that commute to and from work. Bike rides that are taken during the week are shorter and more likely to happen during typical commute hours, while rides on the weekend are more spread out throughout the whole day, and on average a bit longer. Our propositions for new stations make the network denser, and give more people access to a station close to where they live or work.

Future work in this domain should take population densities into account, especially when suggesting locations for new stations. General data about the area of new stations, like infrastructure or amenities e.g. parks or recreational areas should also be considered. To make the network not only more dense, but also expand it into other areas of the city that are currently not served by the station network, algorithms for expansion should be considered, potentially taking existing bicycle infrastructure into account.

References

David Banister. 2005. *Unsustainable Transport: City Transport in the New Century*. Routledge.

Geonorge. 2024. Administrative enheter kommuner. `https://kartkatalog.dev.geonorge.no/metadata/administrative-enheter-kommuner/041f1e6e-bdbc-4091-b48f-8a5990f3cc5b`. Online, accessed 8th of May 2024.

R. Marqués, V. Hernández-Herrador, M. Calvo-Salazar, and J.A. García-Cebrián. 2015. How infrastructure can promote cycling in cities: Lessons from Seville. *Research in Transportation Economics*, 53:31–44.

City Bikes Oslo. 2024a. Oslo city bikes. `https://oslobysykkel.no/en`. Online, accessed 8th of May 2024.

City Bikes Oslo. 2024b. Oslo city bikes historical data. `https://oslobysykkel.no/en/open-data/historical`. Online, accessed 8th of May 2024.

Caitlin Rollison. 2023. Density: the missing piece of the uk's public transport puzzle? `https://www.centreforcities.org/blog/density-the-missing-piece-of-the-uks-public-transport-puzzle/`. Online, accessed 22ed of May 2024.

Michael Szell, Sayat Mimar, Tyler Perlman, Gourab Ghoshal, and Roberta Sinatra. 2022. Growing urban bicycle networks. *Scientific Reports*, 12(1):6765.

Tight. 2018. Hell of a ride. `https://tight.no/projects/oslo-city-bike`. Online, accessed 18th of May 2024.

A Columns in dataset

- **started_at**: Timestamp of when the trip started
- **ended_at**: Timestamp of when the trip ended
- **duration**: Trip duration in seconds
- **start_station_id**: ID of start station
- **start_station_name**: Name of the start station
- **start_station_description**: Short description of where the start station is located
- **start_station_latitude**: Latitude of start station
- **start_station_longitude**: Longitude of start station
- **end_station_id**: ID of end station
- **end_station_name**: Name of the end station
- **end_station_description**: Short description of where the end station is located
- **end_station_latitude**: Latitude of end station
- **end_station_longitude**: Longitude of end station

B Figures and plots

Figure 9: changes in station network from year to year

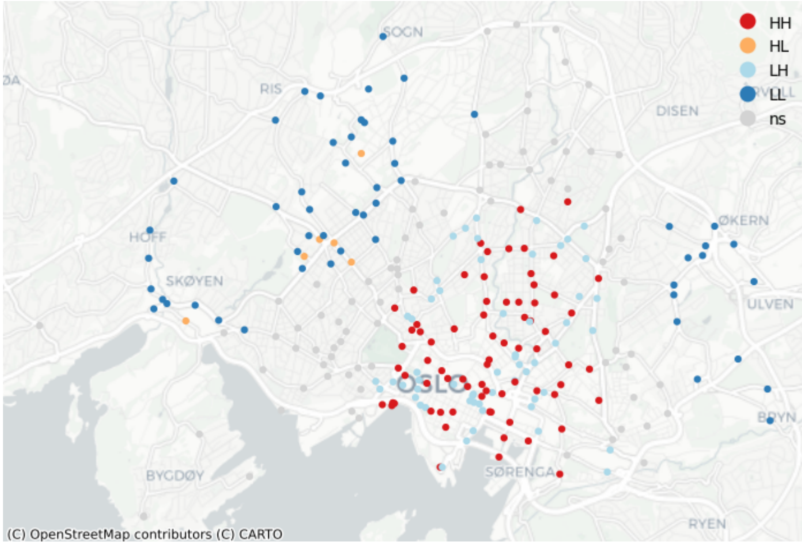

Figure 10: LISA analysis of stations starting rides May 2023 - April 2024

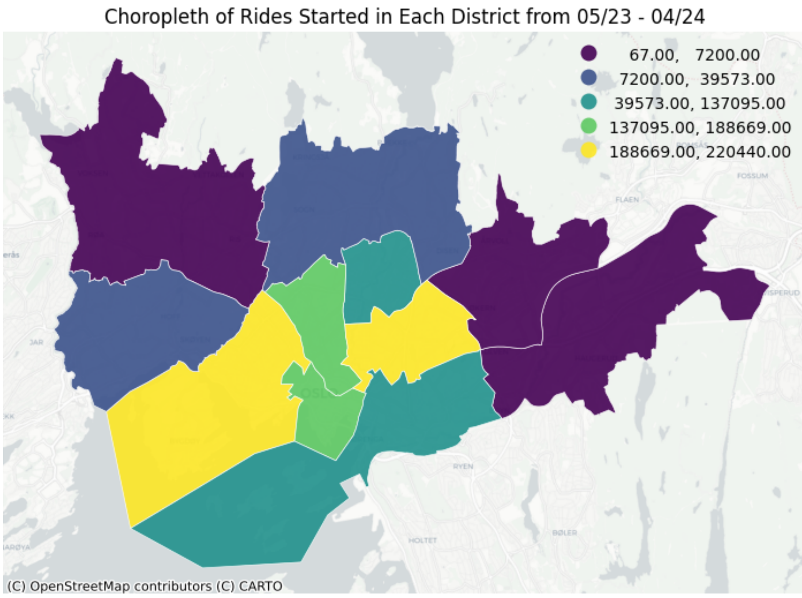

Figure 11: Choropleth of Rides Started in Each District

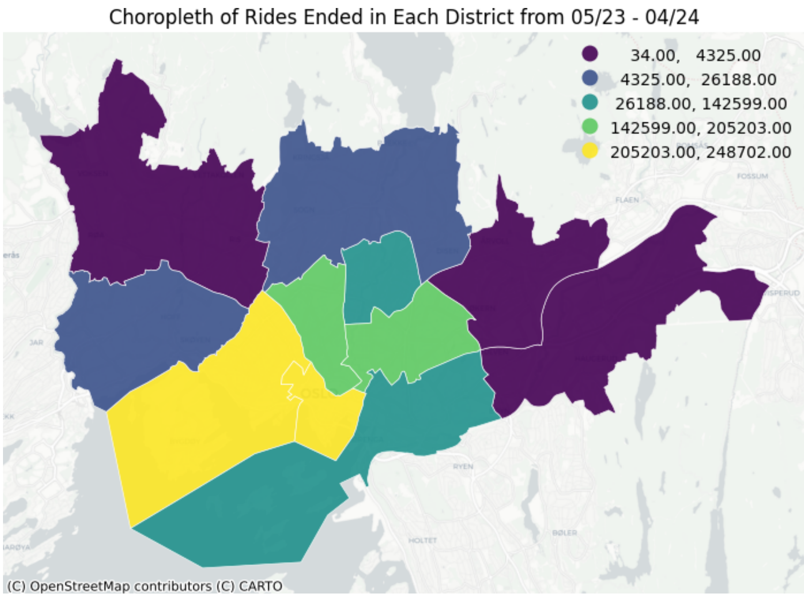

Figure 12: Choropleth of Rides Ended in Each District

C Contribution statement

Both of us (Manuel Knepper and Niclas Classen) contributed roughly the same amount of work to all parts (code and report) of this project.