On the Influence of Land Use and Weather in a Station-based Bike Sharing System

Lucas Schuhmacher*[1], Gabriel Wilkes[1], Pia Tulodetzki[1], Martin Kagerbauer[1], and Peter Vortisch[1]

[1]Institute for Transport Studies, Karlsruhe Institute of Technology (KIT), Germany

Short summary

This paper investigates the impact of weather conditions and land use on the station-based bike sharing system *StadtRAD* in Hamburg, Germany. Using *k-means* clustering, the study identifies five distinct station clusters, revealing different daily usage patterns. Results indicate that land use and proximity to transit stations significantly influence station clustering. Weather impact analysis reveals a decrease in bike demand during precipitation and increased demand on warmer days. The impact of weather is higher at stations associated with recreational areas and on weekends. Overall, the study provides valuable insights into the interplay between weather, land use, and bike sharing patterns, consistent with existing literature.

Keywords: bike sharing, cluster analysis, cycling behaviour, land use, weather

1 Introduction

As cities face the challenges of congestion, environmental sustainability, and last mile connectivity, bike sharing systems have attracted increasing attention from policymakers and researchers. Research's interest is to better understand how people incorporate shared bikes into their everyday mobility and what environmental conditions influence usage patterns. This study performs a cluster analysis based on demand profiles and examines how land use and weather conditions influence the number of bike rentals and returns. It uses data from a large-scale station-based bike sharing system in Hamburg, Germany, from the year 2023.

Many studies examined the relationship between weather and mobility behavior and found that cycling is the most weather-dependent mode of transportation Liu et al. (2015); Rudloff et al. (2015). Bike sharing systems offer a flexible way to ride a bike through ease of use and real-time digital bookings Shaheen et al. (2010), however, being a specially flexible mode ambient factors such as weather have an even greater potential to influence bike sharing use Shaheen et al. (2010). Previous studies found that bike sharing users are often less experienced with bike use in general, especially in contrast to people for whom the bike is a frequently used means of transport Bachand-Marleau et al. (2012). Frequent cyclists are less sensitive to adverse weather conditions than occasional cyclists Heinen et al. (2011), which suggests that bike sharing is at least as affected by weather as regular cycling. This is confirmed by various studies, such as Saneinejad et al. (2012) and Gebhart & Noland (2014), for the cities of Toronto and New York, where reduced ridership correlated with cold temperatures and precipitation. In addition, Ashqar et al. (2019) finds that temperature is a reliable predictor of bike sharing usage in the San Francisco Bay Area. These results can be confirmed for bike sharing in Cologne Schimohr & Scheiner (2021).

Land use, population density and employment density are important influencing factors for bike sharing demand, especially since they are directly related to travel purposes Rixey (2013). El-Assi et al. (2017) found that the correlation between bike sharing trips and employment density exists only on weekdays in Toronto. At the same time, the presence of a train station next to a bike sharing station shows a strong association with the number of trips, which they explain by regular trips such as commuting. Schimohr & Scheiner (2021) confirm this spatial relationship between transit stations and bike sharing usage for the bike sharing system in Cologne. Using cluster analysis, some studies have already examined the interrelated influence of weather and land use on bike sharing. For Vienna, Daejeon (Korea) and Washington D.C., the effect of weather is related to the purpose of use in the clusters Gehrke & Welch (2019); Kim (2018); Vogel & Mattfeld

(2011). All these studies identify stations in residential areas as clusters, with typical morning and afternoon peaks, assuming that people use bike sharing for commuting. Other clusters include commercial or recreational areas, leading to higher changes in usage patterns under adverse weather conditions. This study will analyse both land use and weather conditions by clustering the bike sharing stations in Hamburg based on their average rental and return pattern over the day. After giving an overview on the data used for this paper, results regarding the weather influence and the built environment for each cluster will be presented.

2 Data

This study analyzes open-source bike sharing data from Hamburg's station-based *StadtRAD* service, including 295 stations and 3,700 bikes over the course of a year. The dataset, sourced from the City of Hamburg's *Urban Data Platform* BVM (2023), captures the number of bikes available at each station throughout 2023. Each data point consists of a station ID, a timestamp, and the number of bikes available, and is created only when a bike is picked up or dropped off at a station. The data is used to calculate bike rentals and returns separately. We aggregated this information to 30-minute intervals per station, resulting in a granular representation of station activity throughout the year. Some stations did not contribute complete data for the entire year, possibly for a number of reasons. Cases of missing data were addressed by excluding periods where information was unavailable for more than one day for individual stations. Since only changes at stations are tracked, there is no information about the person who rented the bike or the trip for which it was used.
As a second data source, *CORINE Land Cover* provided by the European Environment Agency (2019) is used to obtain information on land use. We analyzed the predominant land use within a 200 meter radius around each station. The radius was chosen with respect to the proximity between stations to reflect the context of the station's land use without too much overlap.
Finally, we use weather data for Hamburg for the year 2023, provided by the German Weather Service (2023) at a temporal resolution of ten minutes. We used the daily maximum temperature and precipitation on an hourly level, as these variables have shown high explanatory in preliminary work.
Based on these data sources, a set of indicators was calculated to describe bike rentals and returns over the course of the day at each station, attempting to capture the individual influence of weather and seasonality.
As a base to calculate these indicators, we define five time periods: weekday morning peak (5 - 9 AM), afternoon peak (3 - 7 PM), and night (8 - 4 AM), as well as weekend day (12 - 5 PM) and night (8 PM - 4 AM).
The first indicators are based on the share of demand in each of the time periods. It is calculated using the number of rentals in each time period divided by the total number of rentals.
To account for temperature, the daily maximum temperature was used to distinguish cold ($< 10°$C), medium (between 10°C and 20°C), and hot ($> 20°$C) days. We calculated the shares in each of the five time periods for each temperature group. As further indicator, we distinguished whether precipitation was observed. As last, all time periods were calculated with respect to the meteorological seasons. All these indicators were calculated for both rentals and returns. Finally, the ratio of rentals to returns was calculated for these time periods.

3 Methodology

The resulting data set was used to perform a cluster analysis using the *k-means* algorithm MacQueen et al. (1967), carried out in R. The overall goal of the algorithm is to group data points, in our case bike sharing stations, into different clusters. The stations within a cluster should be as homogeneous as possible, while the stations in different clusters should be as heterogeneous as possible in terms of observed behavior.
To initialize the algorithm, a number of clusters k is specified, and k stations are randomly selected as initial cluster centers. To determine an appropriate number of clusters, the elbow method can be applied Marutho et al. (2018). In this approach, the sum of squared errors (SSE) is calculated for different values of k. The number of clusters where the reduction in SSE is the largest is selected. Once the number of clusters and the initial centers are defined, stations are assigned to clusters based on their proximity to each center. Next, the centers of each cluster are updated by calculating

the mean of the assigned stations. This process continues until the algorithm converges when the assignment of the stations to the clusters reaches a stable state MacQueen et al. (1967).
In the following, we analyze the resulting clusters in terms of spatial composition, observed usage patterns, and sensitivity to temperature and precipitation.

4 Results

The performed *k-means* algorithm resulted in five clusters, outlined in Table 1. *Cluster 1* consists of only five stations and therefore has limited explanatory power. *Cluster 2* and *3* contain 53 and 49 stations respectively, while *Cluster 4* is by far the largest cluster with 163 stations. *Cluster 5* includes 25 stations. The spatial distribution of the clusters is shown in Figure 1.

Table 1: Key properties of the clusters identified

cluster	size	stations at transit stop *	stations by land use **			rentals per day ***	
			commercial	residential	recreational	weekdays	weekends
1	5	0 %	40 %	0 %	0 %	5.0	1.2
2	53	19 %	19 %	72 %	6 %	9.6	9.5
3	49	22 %	65 %	20 %	6 %	11.2	5.7
4	163	42 %	23 %	62 %	9 %	13.8	13.6
5	25	24 %	32 %	52 %	12 %	11.8	11.1

* Share among all stations in the clusters. Missing to 100%: no public transit stop nearby.
** Share among all stations in the clusters. Missing to 100%: other land uses.
*** As the number of returns and rentals are nearly identical if daily values are analyzed for every cluster, only rentals are presented.

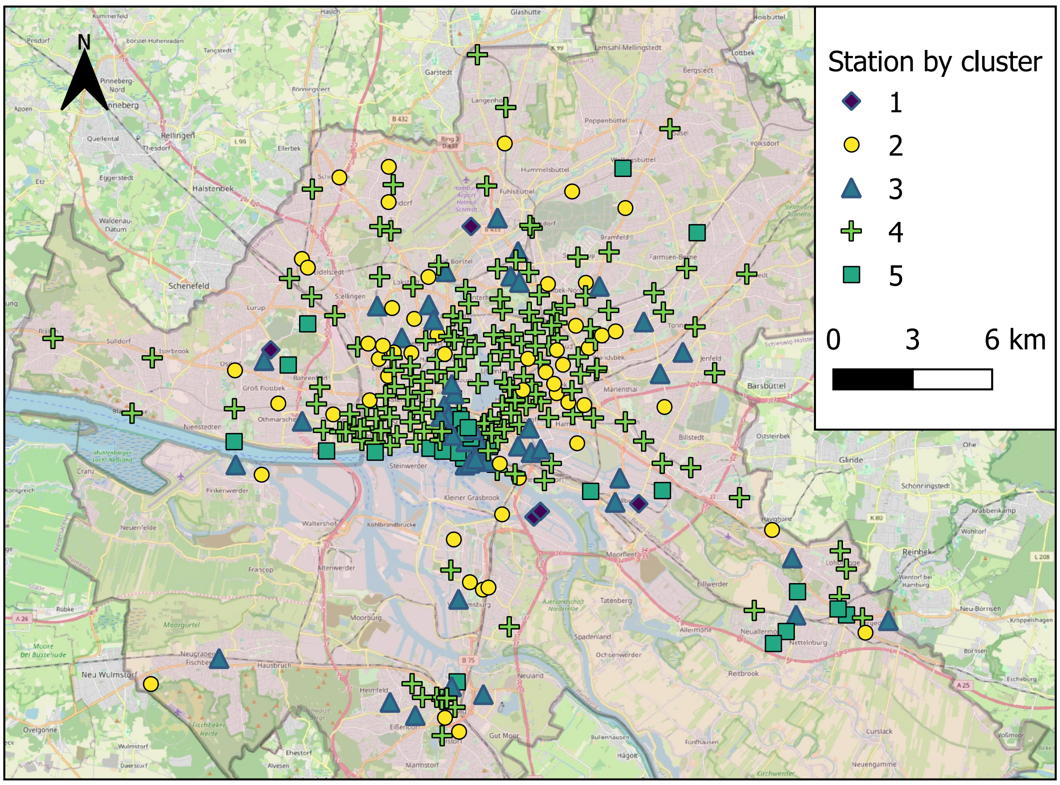

Figure 1: Visual presentation of the distribution of the bike sharing stations including the clusters

According to a Wilcoxon rank sum test, *Cluster 2* and *Cluster 4* are similar in terms of land use. Both are dominated by residential areas. However, *Cluster 4* has 42 % of its stations directly connected to either the subway or suburban trains, while *Cluster 2* has only 19 %. According to Fisher's exact test, we do not find a significant association between land use and being close to a public transportation stop. Thus, these two variables can be regarded separately in our analysis. *Cluster 3* is dominated by industrial and commercial land use (65 % of stations). *Cluster 5* has the highest proportion of stations being nearby recreational areas.
Figure 2 shows the average bike rental and return behavior differentiated by weekdays and weekends over a day. The average behavior of all stations is compared with that of the individual clusters. Considering the average of all stations, a typical pattern is found for weekdays and weekends, with very balanced relation between rentals and returns, dominant morning and evening peaks during weekdays, and a lower but longer demand plateau at weekends. However, if we consider the individual clusters, we can see fundamental differences.
Cluster 3 on weekdays has a clear morning peak for bike returns, but no peak for rentals. In the afternoon, however, the opposite can be observed. Also the stations are frequented below average on weekends. This is consistent with the observation that most stations are associated with commercial areas, which often are workplaces, indicating that bike sharing is used for commuting to and from work.
On weekdays *Cluster 2* behaves contrary to *Cluster 3*. This can be well explained by the fact that Cluster 2 mainly includes residential areas. Thus, people commute from their residential areas to their workplaces associated with *Cluster 3*. On weekends, bikes are increasingly rented in the morning and returned in the afternoon.
In terms of land use, one might expect similar behavior in *Cluster 4*. However, Noland et al. (2019) found for New York that people use bike sharing more often in the morning to get to subway stations, while in the afternoon bike sharing trips tend to start at subway stations and are used to get home. This can explain the increase of returns in the morning, as well as the increase of rentals in the afternoon compared to *Cluster 2*. Compared to all other clusters, *Cluster 5* has a high proportion of bikes rented at night.

Next, we analyze the impact of weather on each cluster. Table 2 shows the decrease in rentals and returns per cluster when precipitation was observed during the characteristic time periods. Overall, a decrease between 19 % and 45 % can be observed during precipitation. However, there is a high correlation between weekday/weekend and precipitation sensitivity of 0.75 for rentals and 0.77 for returns. This indicates that people are more sensitive to precipitation on weekends, as bike sharing may then mainly be used for leisure activities and not for commuting. This is underlined by the fact that *Cluster 5* is 31 % more precipitation sensitive for rentals and returns compared to all other clusters. All other clusters show a fairly homogeneous behavior in terms of precipitation sensitivity.

Table 2: Number of rentals and returns at hours with precipitation compared to hours without precipitation by cluster and time of day

	rentals					returns				
	weekday			weekend		weekday			weekend	
	m	a	n	d	n	m	a	n	d	n
1	-21 %	-21 %	-21 %	-30 %	-29 %	-23 %	-19 %	-19 %	-31 %	-36 %
2	-26 %	-25 %	-24 %	-33 %	-23 %	-25 %	-26 %	-24 %	-32 %	-33 %
3	-23 %	-23 %	-22 %	-35 %	-37 %	-24 %	-22 %	-20 %	-33 %	-34 %
4	-28 %	-27 %	-26 %	-34 %	-34 %	-28 %	-28 %	-26 %	-34 %	-34 %
5	-32 %	-32 %	-30 %	-43 %	-43 %	-33 %	-30 %	-30 %	-45 %	-43 %

m - morning, d - day, a - afternoon, n - night

For temperature, the impact of different daily maximum temperatures on demand is analyzed on entire days (see Table 3). We only present rentals, because returns are nearly identical to these at this level of aggregation. Comparing cold days (daily maximum temperature below 10°C) with medium days (between 10°C and 20°C), bike sharing usage increases by 31 % on days of medium temperature. However, comparing cold days with warm days (daily maximum temperature above

20°C), the average usage increases by 98 % on warm days.
Consistent with the findings for precipitation, *Cluster 5* is more sensitive to higher temperatures than all other clusters. For temperatures above 20°C, the increase in rentals is 65 % higher, for bike returns even 73 % compared to the other clusters. This also supports the hypothesis that leisure travel is more sensitive to weather.
Days with a maximum temperature between 10°C and 20°C show only a low correlation between weekday/weekend and temperature sensitivity of 0.08 for rentals and 0.09 for returns. This changes drastically for warm days (> 20°C). The corresponding correlation is 0.64 for rentals and 0.66 for returns. Thus, bike sharing is more sensitive to higher temperatures on weekends, which could be explained by increased outdoor activities.

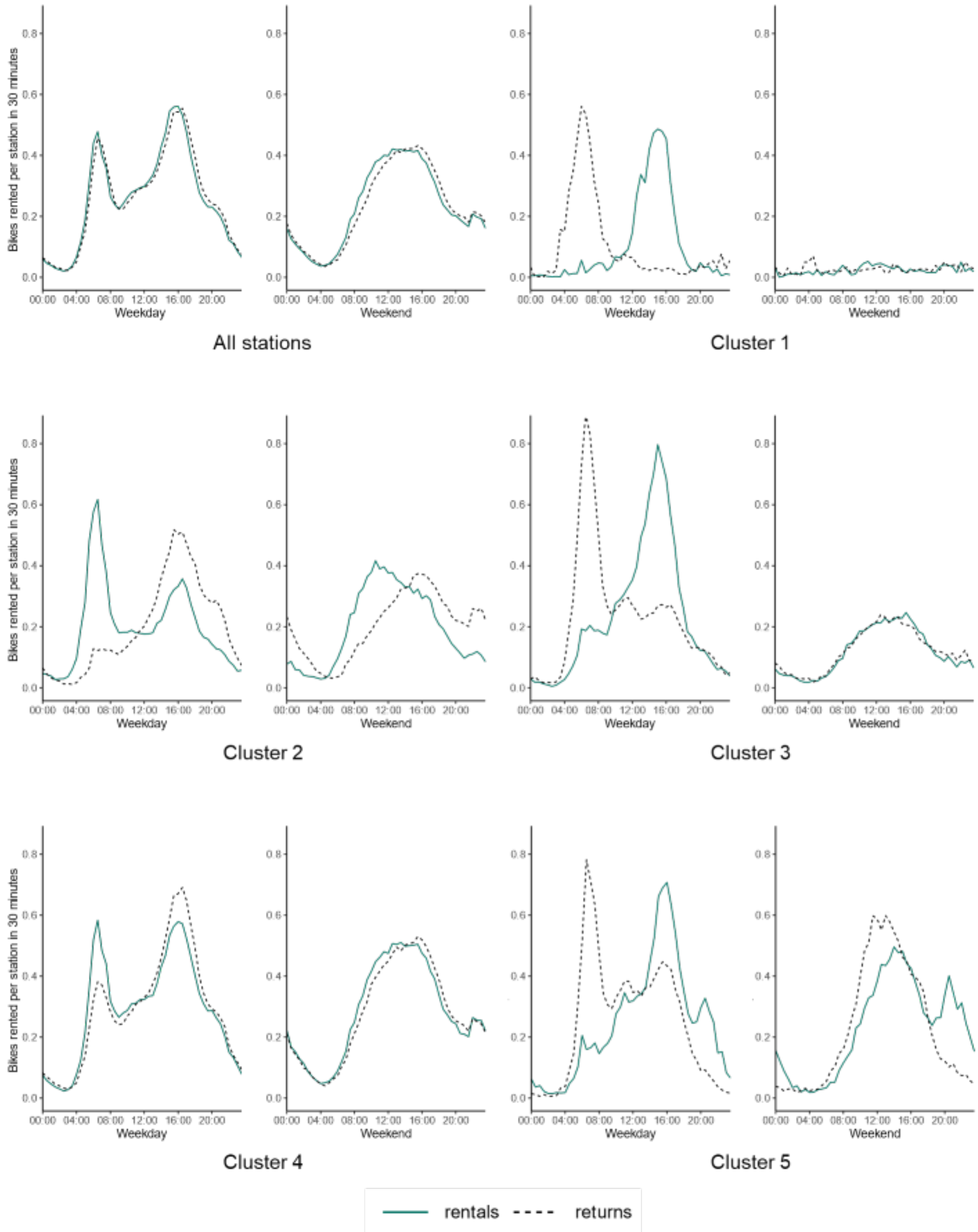

Figure 2: Average bike rentals and returns on weekdays and weekends over one day

Table 3: Effect of temperature in the different clusters

	weekdays			weekends		
	number of rentals *	rel. diff. to L **		number of rentals *	rel. diff. to L **	
cluster	L **	M **	H **	L **	M **	H **
1	4.2	12%	45%	0.9	-3%	85%
2	6.9	34%	82%	5.9	39%	136%
3	8.4	30%	66%	3.4	44%	156%
4	9.7	37%	87%	8.0	43%	159%
5	7.7	46%	109%	5.3	66%	244%

* Absolute number of rentals per day and station.
** L = low maximum daily temperature (below 10°C),
M = medium maximum daily temperature (between 10°C and 20°C),
H = high maximum daily temperature (> 20°C)

5 Conclusion

Using *k-means* clustering we were able to identify different usage patterns in station based bike sharing by combining station demand profiles, including weather data. It was found that the clusters can be described well and land use and proximity to transit stations are important explanatory variables for differences between clusters. Although information on trips made was unavailable, different trip purposes could be identified in the demand patterns. The influence of weather also led to plausible results regarding a negative influence of precipitation and a positive influence of rising temperatures on the frequency of use. Additionally, it can be seen that the influence of weather is significantly greater for leisure trips and on weekends. The results are thus consistent with the effects observed in the literature, both in terms of land use and weather.

Acknowledgements

This study is part of the project #transmove at the DigiLab of the State Agency for Roads, Bridges, and Waterways (LSBG) for the Free and Hanseatic City of Hamburg. It is funded under the "Digitalization of Municipal Transport Systems" grant program by the Federal Ministry for Digital and Transport (BMDV) (Grant Reference Number: 16DKV41086).

References

Ashqar, H. I., Elhenawy, M., & Rakha, H. A. (2019). Modeling bike counts in a bike-sharing system considering the effect of weather conditions. *Case Studies on Transport Policy*, *7*(2), 261–268. doi: 10.1016/j.cstp.2019.02.011

Bachand-Marleau, J., Lee, B. H. Y., & El-Geneidy, A. M. (2012). Better understanding of factors influencing likelihood of using shared bicycle systems and frequency of use. *Transportation Research Record: Journal of the Transportation Research Board*, *2314*(1), 66–71. doi: 10.3141/2314-09

BVM. (2023). *Stadtrad stations dataset.* Behörde für Verkehr und Mobilitätswende. Retrieved 01/26/2024, from `https://suche.transparenz.hamburg.de/dataset/stadtrad-stationen-hamburg29`

El-Assi, W., Salah Mahmoud, M., & Nurul Habib, K. (2017). Effects of built environment and weather on bike sharing demand: a station level analysis of commercial bike sharing in Toronto. *Transportation*, *44*(3), 589–613. doi: 10.1007/s11116-015-9669-z

European Environment Agency. (2019). *Corine land cover 2018 (raster 100 m), europe, 6-yearly - version 2020_20u1, may 2020.* European Environment Agency. Retrieved 2024-01-

06, from https://sdi.eea.europa.eu/catalogue/copernicus/api/records/960998c1-1870-4e82-8051-6485205ebbac?language=all

Gebhart, K., & Noland, R. B. (2014). The impact of weather conditions on bikeshare trips in washington, dc. *Transportation*, *41*(6), 1205–1225. doi: 10.1007/s11116-014-9540-7

Gehrke, S. R., & Welch, T. F. (2019). A bikeshare station area typology to forecast the station-level ridership of system expansion. *Journal of Transport and Land Use*, *12*(1). (Number: 1) doi: 10.5198/jtlu.2019.1395

German Weather Service. (2023). *Climate data center.* Retrieved 01/26/2024, from https://www.dwd.de/EN/climate_environment/cdc/cdc_node_en.html

Heinen, E., Maat, K., & Van Wee, B. (2011). Day-to-Day Choice to Commute or Not by Bicycle. *Transportation Research Record: Journal of the Transportation Research Board*, *2230*(1), 9–18. doi: 10.3141/2230-02

Kim, K. (2018). Investigation on the effects of weather and calendar events on bike-sharing according to the trip patterns of bike rentals of stations. *Journal of Transport Geography*, *66*, 309–320. doi: 10.1016/j.jtrangeo.2018.01.001

Liu, C., Susilo, Y. O., & Karlström, A. (2015). The influence of weather characteristics variability on individual's travel mode choice in different seasons and regions in Sweden. *Transport Policy*, *41*, 147–158. doi: 10.1016/j.tranpol.2015.01.001

MacQueen, J., et al. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth berkeley symposium on mathematical statistics and probability* (Vol. 1, pp. 281–297).

Marutho, D., Hendra Handaka, S., Wijaya, E., & Muljono. (2018). The determination of cluster number at k-mean using elbow method and purity evaluation on headline news. In *2018 international seminar on application for technology of information and communication* (p. 533-538). doi: 10.1109/ISEMANTIC.2018.8549751

Noland, R. B., Smart, M. J., & Guo, Z. (2019). Bikesharing trip patterns in new york city: Associations with land use, subways, and bicycle lanes. *International Journal of Sustainable Transportation*, *13*(9), 664-674. doi: 10.1080/15568318.2018.1501520

Rixey, A. R. (2013). *Station-Level Forecasting of Bikesharing Ridership: Station Network Effects in Three U.S. Systems* (Vol. 2387) (No. 1). doi: 10.3141/2387-06

Rudloff, C., Leodolter, M., Bauer, D., Auer, R., Brög, W., & Kehnscherper, K. (2015). Influence of Weather on Transport Demand: Case Study from the Vienna, Austria, Region. *Transportation Research Record: Journal of the Transportation Research Board*, *2482*(1), 110–116. doi: 10.3141/2482-14

Saneinejad, S., Roorda, M., & Kennedy, C. (2012). Modelling the impact of weather conditions on active transportation travel behaviour. *Transportation Research Part D - Transport and Environment*, *17*. doi: 10.1016/j.trd.2011.09.005

Schimohr, K., & Scheiner, J. (2021). Spatial and temporal analysis of bike-sharing use in Cologne taking into account a public transit disruption. *Journal of Transport Geography*, *92*, 103017. doi: 10.1016/j.jtrangeo.2021.103017

Shaheen, S., Guzman, S., & Zhang, H. (2010). *Bikesharing in Europe, the Americas, and Asia: Past, Present, and Future.* doi: 10.3141/2143-20

Vogel, P., & Mattfeld, D. C. (2011). Strategic and Operational Planning of Bike-Sharing Systems by Data Mining – A Case Study. In J. W. Böse, H. Hu, C. Jahn, X. Shi, R. Stahlbock, & S. Voß (Eds.), *Computational Logistics* (Vol. 6971, pp. 127–141). Springer. doi: 10.1007/978-3-642-24264-9_10