

Balke, Georg; Rech, Felix; Lellep, Martin; Waldner, Felix

Article — Published Version

Data-driven insights into (E-)bike-sharing: mining a large-scale dataset on usage and urban characteristics: descriptive analysis and performance modeling

Transportation

Suggested Citation: Balke, Georg; Rech, Felix; Lellep, Martin; Waldner, Felix (2025) : Data-driven insights into (E-)bike-sharing: mining a large-scale dataset on usage and urban characteristics: descriptive analysis and performance modeling, Transportation, ISSN 1572-9435, Springer US, New York, Vol. 52, Iss. 6, pp. 2433-2473,
<https://doi.org/10.1007/s11116-025-10661-2>

This Version is available at:

<https://hdl.handle.net/10419/333379>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<http://creativecommons.org/licenses/by/4.0/>



Data-driven insights into (E-)bike-sharing: mining a large-scale dataset on usage and urban characteristics: descriptive analysis and performance modeling

Felix Waldner¹ · Georg Balke¹ · Felix Rech² · Martin Lellep³

Accepted: 12 July 2025 / Published online: 7 August 2025
© The Author(s) 2025

Abstract

Bike-sharing is considered one of the most sustainable forms of urban transportation Aboulela (Trans. Res. Part A: Policy Pract. 169:103602, 2023) and, therefore, understanding what determines the efficiency of a bike-sharing system (BSS) is crucial. Various factors from different domains, like built environment or behavioral studies, have been demonstrated influential in single BSSs and small-scale datasets (from a single or few BSS). However, the multitude of potential factors requires rigorous analysis and application of statistical models on large datasets. Due to the unavailability of large-scale datasets comprising hundreds of BSSs and unified databases containing variables, interactions are only known to a limited extent. Here we show the usage patterns of different types of BSSs, and draw statistically valid conclusions on the relevance of factors from different domains by gathering over 43 Mio km open data from 267 BSSs and 108 predictors, as well as applying time-series clustering, building predictive classification models, and fitting a stepwise regressive Ordinary Least Squares (OLS) model. Results show that operational factors like operating time, design factors like operational area and share of electric bikes, sociodemographic factors like age and gender, built environment like cycling infrastructure, and economic factors like overnight stays all play an important role. Descriptive statistics reveal significant room for improvement in the operations e-bike-sharing system (electric BSS) and show how climate influences operational organizations. We anticipate this publication and the associated dataset, which we make openly available, to spark discussions about the design and operations of BSS and to be a starting point for further analyses.

Keywords Bike-sharing · Urban transport · Micromobility · e-bikes · Active mobility · Open data · Statistical models

Introduction

With Shared Micromobility (SMM) being considered a hopeful prospect for transforming urban mobility towards more sustainability on the one hand but also experiencing increased pressure in times of general financial hardship and growing competition between modes on the other hand, understanding the factors influencing the success of BSSs is more important than ever. Especially since the advent and success of E-Scooter Sharing (ESS), the question emerges how bike-sharing, which has been around for decades and is considered more sustainable (International Transport Forum 2021), can be a success.

The abundance of factors potentially influencing the success requires rigorous assessment of correlations with performance indicators and, in turn, necessitates extensive data from a multitude of BSSs. This is especially true since modern BSS vary in fundamental factors: organizational form (station based bike-sharing system (station based BSS), free-floating bike-sharing system (free floating BSS) or hybrid systems), as well as propulsion type (classic, human-power bicycles, electric BSS, or mixed systems). Success factors include, amongst others, sociodemographics, built environment, climate, policy, and many more. Their diversity affirms the need for sophisticated statistical models, allowing for valid statements to be made about correlations. While the potential of statistical analysis was shown in smaller-scale studies comprising a few to a few dozen BSSs, there are hardly any publications drawing on large datasets with hundreds of BSSs and hundreds of variables being thoroughly analyzed. Although electric BSS is becoming increasingly common as a potential next step in bike-sharing, knowledge on such electrified systems is scarce. This paper, therefore, aims to close this gap in two major steps: Firstly, by delivering a novel dataset of more than 49 Mio km from 267 of BSSs and electric BSS systems. And secondly, by providing insights into the usage patterns and the influence of external factors on the performance of these systems. The obtained findings shall serve the purpose of understanding what factors decide what to expect from potential new BSS or how to adapt existing systems to increase economic performance and sustainability. We, therefore, gather trip data from 267 international BSS (including fully electric and mixed systems), as shown in Fig. 1 and 108 predictors from different categories. We first analyze the collected data thoroughly with respect to different aspects like spatiotemporal patterns and operational details such as battery charging. Special attention is paid to electric BSS, its usage, and operations. We cluster the systems by temporal usage patterns and build different predictive classification models for cluster affiliation. Those are used to explore determinants of temporal usage patterns. Furthermore, we develop an OLS model to discover interdependencies of city-scale variables and the performance of BSSs. In the modeling process, we apply stepwise regression in order to identify relevant factors. Comparing the performance metrics of our model to those of previously published models reveals significant improvements in quality. The results of the more sophisticated OLS model were compared with those from a traditional statistical analysis, where variable selection via a Spearman correlation analysis at a 0.05 significance level revealed only weak correlations. Findings show that sociodemographic, economic, environmental, and system-specific variables all play an important role in the performance of BSSs. Insights into user behavior of different types of BSS and propulsion can serve policy makers and BSS-operators equally.

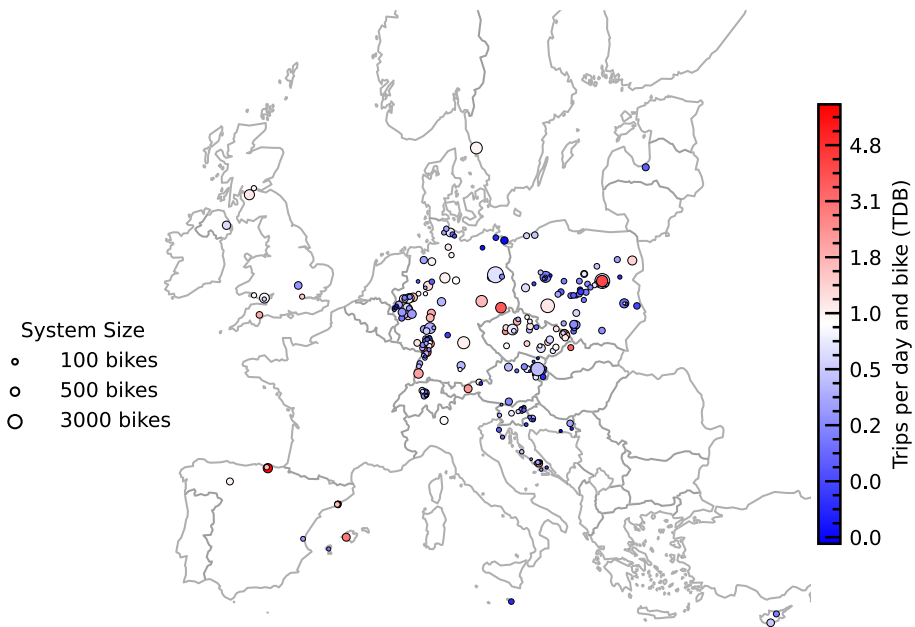


Fig. 1 Map of the 267 analyzed bike-sharing systems. The marker size corresponds to the number of bikes. The color indicates the trips per day and bike, ascending from blue to red. The largest system is Berlin with approx. 6200 bikes, while Bilbao has the highest rental frequency of 7.62 trips per day and bike. Not displayed: 3 BSS on the Canary Islands

Literature review

Although raw data of BSS, e.g., in the form of location data in smartphone applications or online reports, is theoretically available in abundance, there exist only a few medium- and hardly any large-scale analyses publications yielding thorough analyses. One reason for this might be the laborious data collection and subsequent processing (see subsection 3.2). In the following paragraph, we present previous studies on datasets of multiple BSS, publications investigating performance, usage patterns, and the influence of external variables, as well as work on statistical models applied to BSS. Due to the vastness of bike-sharing literature, we focus on the specific topics listed above. For comprehensive reviews on BSS and electric BSS, we kindly refer the reader to more general publications like Zhou et al. (2022) (BSS) or Zhou et al. (2023) (electric BSS). The literature review shows

Large scale/multi-city datasets

One of the largest BSS datasets comprises data from 75 international cities and their station based BSS (Médard de Chardon et al. 2017). Based on this information, the authors gathered station data of all systems for 14 months and estimated trip numbers, bicycle counts, and the effectiveness of repositioning efforts. The number of trips per day and bike (TDB) serves as their main Key Performance Indicator (KPI), calculated as the overall average via the aforementioned estimated trips and the maximum number of bicycles docked dur-

ing a respective month. The authors apply descriptive analysis and statistical models like robust regression or mixed models. They observe TDB values range from 0.22 to 8.4 but do not detect correlations of performance and size of systems. Results from the models are, amongst others, that station density, cycling infrastructure, and population size increase performance while mandatory helmets and unfavorable weather decrease it.

Another study on a multitude (69 systems) of BSS was conducted by Zhao et al. (2014a). Therein, the main KPI is the turnover rate, being defined as the number of rides divided by the total number of bicycles in a city. Independent variables belonging to three categories are analyzed: urban features (socioeconomic and demographic), system characteristics (e.g., number of members, bicycles, or stations), and composite indicators (connecting deposit, income, and penalties). The authors use ordinary and partial least squares models to identify drivers of the performance. They find that key variables increasing BSS performance are urban population, government expenditure, and the number of members and docking stations. Furthermore, they state there is an optimal ratio of bicycles to population of 0.2. By gathering station data, O'Brien et al. (2014) provide insights into 38 BSS. They provide exhaustive descriptive statistics on usage and spatial characteristics and are the first to introduce KPIs like system compactness (describing a system's shape) or load factor. An evaluation of the BSS of 50 cities through KPIs was presented by Pimentel de Vassimon (2015). He includes 26 parameters from the areas of system operation (e.g., opening hours), system dimension and investments, usage statistics, and demographic or economic factors. KPIs mainly focus on usage, like bicycle or station availability. The author finds that private-public partnerships tend to perform better than purely private or public-run systems, with the latter performing the worst. Contradicting previously mentioned studies, he does not find city population to be strongly influencing BSS performance. With an explicit focus on weather, Bean et al. (2021) examine BSS in 40 cities. They developed a generalized additive model predicting system usage based on time of day, day of week, climate (UTCI), and precipitation. They found the time of day to be the most influential variable, followed by precipitation. System characteristics are not part of their analysis. Most recently, Cycling Industries Europe, the trade association of the European cycling industry, published a benchmarking whitepaper comprising data from 148 BSS (Cycling Industries Europe 2023). The paper presents comprehensive data on the analyzed systems, like the ratio of bicycles to inhabitants or TDB, but does not entail explanatory statistical analyses. The largest collection of bike-sharing data is to be found in Todd et al. (2021), with 322 station based BSS on all continents. The authors cluster the systems into five bins via temporal usage and further build subgroups with distinct usage patterns, contextual indicators, and user characteristics. Mahajan and Argota Sánchez-Vaquerizo (2024) present a comparison of 40 BSSs from different countries. They developed a Bike-Share Service Accessibility Index and, using population data and urban metrics, evaluated the network performance. Similar to the presented work, they rely on publicly available data from BSSs and amend it with data from OpenStreetMap (OSM) and the Global Human Settlement Layer. The subsequent analyses focus on the degree of integration of the network and its coverage rather than an analysis of trips. The authors find four groups of city types and distinctive differences in service quality, in terms of network expansion and station connectivity. They also consider connection to public transit but do not consider a broader range of indicators. There are further studies less relevant to the present work that are left out of this review for the sake of conciseness. They either focus on other modes, mainly e-scooters (Li et al. 2022),

or comprise significantly smaller numbers (single-digit or low double-digit numbers) like Otero et al. (2018), or both like Abouelela et al. (2023), and are therefore not reviewed in the present work.

Determinants of BSS usage and performance

Many publications examine the usage and performance of BSS, especially the factors influencing both, serving as a foundation for our large-scale analysis. The influence of the built environment on BSS was analyzed by different authors and with different focuses. Duran-Rodas et al. (2019a) focus on ridership and conclude that city population, structure, and transport infrastructure are the most relevant factors by applying a data-driven methodology, encompassing statistical methods and machine learning, to datasets of six cities. Chen and Ye (2021) obtain comparable results in their study assessing the relationships between environmental elements and bike-sharing usage. They name population density (as opposed to the absolute number mentioned previously) and employment density as key variables. The influence of bicycle-oriented infrastructure was already found to be significant by El-Assi et al. (2017) in their analysis of Toronto's BSS. Besides population density, urban characteristics, especially land use or density of certain Point of Interest (POI) like restaurants or educational facilities, were found to be a major influence on BSS ridership by Faghih-Imani et al. (2014). The importance of both educational facilities and areas with a high density of workplaces is further substantiated by Tran et al. (2015). Focussing on the interrelationship between BSS and metro, Guo and He (2020) find that mixed land use, recreational areas, and again transportation infrastructure are positively connected to BSS usage. Xu and Chow (2020) quantify the importance of cycling infrastructure for bike-sharing ridership by correlating its build-up over three years to the number of trips. They find a strong local connection, showing that the numbers increase in direct proximity to the new bike lanes, but have little to no effect on a city-wide scale. This underscores the importance of integrated city planning instead of a variety of uncoordinated local measures.

Besides the built environment, there are a variety of other factors potentially influencing bike-sharing and its usage. Du and Cheng (2018) assess a multitude of socioeconomic, spatial, temporal, and urban characteristics. Besides strong temporal dependencies over the course of the day, they identify employment status, technical state of the system, and price (amongst others) as key determinants of the success of a BSS. A similar analysis was conducted by Eren and Uz (2020), yielding insights on the significant effects of weather, perceived safety, system design (station density), and demographics (especially age). The role of bike-sharing in the context of multimodal travel options in urban environments was investigated by Faghih-Imani and Eluru (2015). Besides bicycle lanes, they identify station size, land use, and user characteristics (commuters vs. casual users) as decisive factors in BSS usage. In a later publication, Faghih-Imani et al. (2017) identify station density, mean station capacity, and presence of amenities like business, recreation, and restaurants as important factors for the success of bike-sharing. Applying cluster analysis, Kim (2018) sheds light on the temporal dependencies of multiple weather-related variables and exceptional events on the usage patterns of bike-sharing. He finds that stations can be clustered into three distinct characteristics and quantifies the nonlinear connection between temperature and ridership. Soriguera and Jiménez-Meroño (2020) developed a model allowing them to obtain optimal BSS given a specific application. The models allow them to run differ-

ent analyses, leading to insights into the strategic variables of a BSS. Their results show, amongst other things, how to decide between free floating (free floating) and station based (station based) and the implications of the deployment of pedelecs. A wide set of urban characteristics and the respective impacts on bike-sharing users were investigated by Sun et al. (2017). Variables include, amongst others, population density, land use, and points of interest, safety, and convenience, as well as public transport (PT) proximity, and are applied to the Chicago bikesharing system *Divvy*. They aim to provide a basis for decisions for policymakers with their results. Applying multiple statistical models and combining usage data and weather information, Jaber and Csonka (2023) implement a model to identify members, occasional users, and visitors of BSSs. Through time-series predictions, they analyze the impact of the weather on the previously found user groups. Showing that, e.g., visitors are mostly influenced by temperature, wind, and type of day, while occasional users are mostly influenced by precipitation, they aim to provide a basis for decision-makers to predict expected usage levels for BSSs. The authors evaluate the models' performances and conclude that random forest (RF) performs best in the prediction task while a negative binomial approach outperforms Poisson Regression in modeling trip numbers. Finally, evaluation of BSS themselves is an important step towards improving them. While Caggiani et al. (2021) present a methodology to assess the efficiency of BSS stations. While they argue that their approach delivers meaningful insights for the design of BSS, they acknowledge that the choice of variables considered requires an expert with detailed knowledge. Using a data envelopment analysis framework, Hosseini et al. (2023) determine the performance of stations in BSS. They name population density and weather as the most influential and draw conclusions on the differences between BSS and electric BSS. Their results suggest pedelecs are more suited as unimodal means of transport, while human-propelled bicycles are more often used for legs of intermodal trips. Another study on the performance of stations in BSS was presented by Nogal and Jiménez (2020). By mixing objective features like walkability with subjective features like perceived security, the authors are able to draw a complete picture of the determinants of success of BSS stations.

Motivation and contribution

This paper contributes to the existing, previously outlined research, as the size of the collected dataset exceeds that of most previous publications. Furthermore, it is the first large-scale dataset on electric BSS and free floating BSS known to the authors. Using bicycle location data instead of constructing flows from dock data distinguishes our data acquisition process from that of previous studies, especially allowing analysis of free floating BSS. Drawing on this immense dataset allows for the application of previously inapplicable statistical methods due to sample size constraints. The rigorous statistical analysis combined with an in-depth descriptive evaluation generates various new insights beneficial to BSS operators and municipalities in equal terms.

Methodology

In order to understand what factors influence the usage of BSS and what determines their performance, we performed four of analyses. They comprise primary (trip data from 267 BSS) and secondary data collection (108 potential variables), descriptive analysis of the primary dataset as well as statistical modeling to draw substantial conclusions between variables and performance. The four steps are explained in the subsequent sections. The primary data is available as supplementary material online.¹

Primary data

The data pipeline employed in this paper comprises three steps: Firstly, the raw data is acquired and stored from an application programming interface (API). Afterward, the data is processed and inserted into a database system. Lastly, within the database, trips of the rental vehicles are reconstructed, constituting the primary data analyzed in this study, and the core part of the dataset published. An overview of the process is given in Fig. 2.

For step one, the raw data source is the Nextbike JSON-API documented in (nextbike by TIER 2024). It provides live information on the individual bikes (including live geographical locations), stations, the cities' rental systems (subsequently called only systems) Notably, each object is listed, including its permanent unique IDs, so its state can be traced continuously. The data is amended by information available through the Nextbike General

¹The full dataset is available open-source under the following link: <https://github.com/tumftm/european-bi>

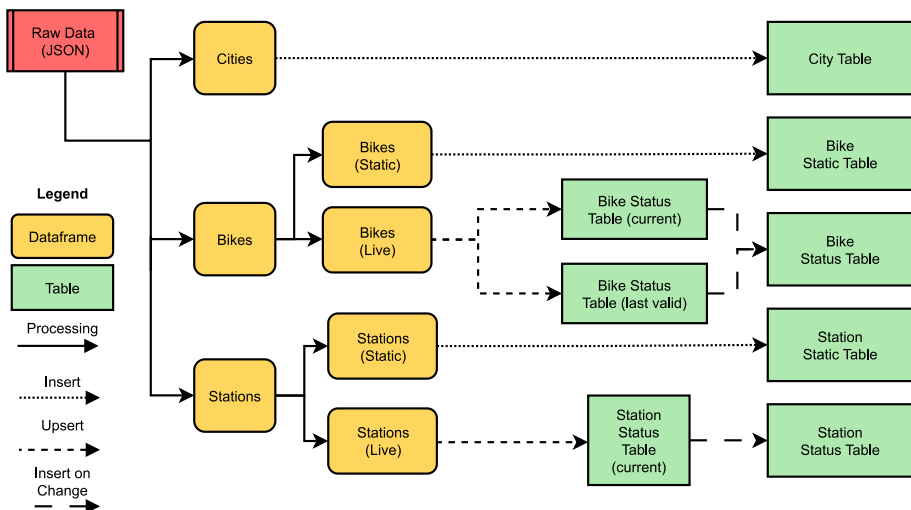


Fig. 2 Schematic overview of the data pipeline used to reduce data redundancy. Upsert ("update or insert"): Rows with new primary key values are inserted. Otherwise, the existing row is updated. Both rows from the temporary bike status tables are inserted into the main table only if the current state of the bike differs significantly from the last valid recorded state (10 m movement or changed state of charge)

Bikeshare Feed Specification (GBFS) feed (MobilityData 2024). The JSON-API was queried every minute, so bike movements could be reconstructed with high precision. Due to technical issues, on the client or server side, 4.7% of the queries failed, of which 3/4 were a single outage of 11 days. Of the successful queries, .609% were discarded as invalid. The main reason was that significant portions of the bikes were temporarily missing from the API.

Secondly, during processing, the raw data is partitioned into five distinct subsets and subsequently inserted into the database in order to create a normalized database without redundant data. Static data on cities, vehicles, and stations, such as system location, bike model, or station capacity, are stored in a table for each domain. This data is assumed to be static and thus not updated. Time-dependent data of vehicles and stations is stored in a separate table and assigned the timestamp of the scraping process. For bikes, this comprises their location and electric state of charge (SOC), while occupancy and maintenance status are stored for stations. The extracted properties are detailed in subsection A.2.

Thirdly, within the database, trips are reconstructed. The time-dependent data has a high degree of redundancy, as a single bike location and a single station status are observed multiple times, yet only state changes like rentals are of particular interest. Thus, the processing automatically removes redundant data points, reducing the required storage space and facilitating evaluation. For bikes, this is achieved by inserting dynamic data into two temporary tables. One temporary table is always updated with the last valid observation, while the second temporary table only accepts updates when a bike has been moved a minimum distance of 10 m or the SOC has changed. When this condition is met, both entries are transferred to the main table, allowing precise tracking of bike rentals' start and endpoints and conditions. For live station data, the temporary table only accepts state changes. The new data is transferred to the main table upon recognizing a change. A new observation thus implicitly defines the end of the previous operational state. The database tables for primary data have an estimated combined volume of 17 GB, a 97% decrease compared to the compressed JSON files. The total recorded distance is 43.8 Mio km of line-of-flight movements Tables 1 and 2.

Within the primary data, there is no machine-readable information on the pricing of the trips - we therefore cannot include it in the following analysis. To provide a glimpse into the pricing strategies, we assessed 10 randomly drawn BSS systems and found that across all of them, a subscription-free pay-as-you-go tariff is available (billed by 1, 15, 20, or 30 min). Commonly, monthly or yearly subscriptions are available that allow for 30 min of free riding per rental, which encompasses 87.6% of all trips according to our recorded data Tables 3 and 4.

Table 1 Statistical averages for small, medium, and large systems

Indicator	Large	Medium	Small
nr. of systems	35	95	137
Average bikes per station	17.62	7.21	5.3
Avg. bikes per rack	4.73	0.96	0.86
Avg. bikes per km ²	14.07	8.37	3.30
Avg. stations per km ²	1.68	1.37	0.72
Share of station based BSS	60%	78%	82%
Share of mixed propulsion BSS	43%	20%	25%
Share of electric BSS	2.5%	9.8%	4.5%

Table 2 Model performance metrics

Model	Accuracy	Precision	Recall	F1-score	MSE	ROC AUC
DT	0.65	0.67	0.65	0.63	n.a	n.a
GB	0.70	0.69	0.70	0.68	0.37	0.48
kM	0.49	0.75	0.49	0.32	0.51	–
kNN	0.59	0.60	0.59	0.58	0.89	0.80
SVM	0.81	0.82	0.81	0.81	0.19	n.a
NN	0.54	0.77	0.54	0.43	0.46	0.73
RF	0.78	0.76	0.73	0.70	0.35	0.87

Best values are printed in bold

Table 3 Crosstabulation of top 5 most important features across models. A check mark indicates presence among top features

Feature	DT	GB	kM	SVM	NN	RF
Housing cost	✓	✓		✓		
Household income			✓			✓
Share of pers. at risk of poverty	✓	✓		✓		✓
Trust in technology	✓	✓		✓	✓	
Use of collab. economy	✓	✓		✓		
Functional urban areas			✓			
Travel time budget			✓			
Hilliness					✓	
nr. of bike parking spots per sqkm					✓	✓
nr. of traffic barriers per sqkm					✓	
variety of traffic barriers per sqkm						✓
nr. of road accidents					✓	
bike safety index			✓			
crime index						✓
Crime rate	✓	✓	✓	✓		

Secondary data

We do not only aim to describe the performance of BSSs but also explain and predict it. In order to be able to understand and predict the influential factors of BSS performance, we collect 108 potential predictors from various additional sources, a full list of which can be found in Table 18. The European Statistical Office (Eurostat) provides statistical information on a wide range of topics for member states of the European Union, with data commonly also available for other European countries. Statistics on Eurostat [regularly include observations for members of the European Free Trade Association \(EFTA\) like Norway and EU candidate countries like Albania](#). Data on the United Kingdom is generally also available from before exiting the EU in 2020. Firstly, we access statistics on demographic information like the relative size of age groups or the shares of genders. We also utilize data on economic status, such as (un)employment and poverty rates. Moreover, further statistics related to mobility, like the number of cars per capita, economic indicators like the frequency of usage of sharing economy services like accommodation, and environmental coefficients like air pollution indicators are included, as well. Data from this source is generally

Table 4 Variables of the model that were selected through *bidi-rectional stepwise regression* for the dependent variable log(TDB)

Variable	Coef	p-value	Conf. int	PCC
Intercept	−0.555	0.000***	[−0.60, −0.51]	
Population: share < 30 (%)	0.091	0.001**	[0.04, 0.15]	0.24
Population per km ² (log)	0.088	0.031*	[0.01, 0.17]	0.49
Passenger Cars per 1000 inhabitants	0.074	0.014*	[0.02, 0.13]	0.07
Sharing econ.: accommo- dation (annual nights by foreign users, log)	0.122	0.000***	[0.06, 0.19]	0.12
Sharing econ.: accommo- dation (annual nights by domestic users)	−0.076	0.014*	[−0.14, −0.02]	0.16
EP elections: ECR (%)	0.124	0.000***	[0.07, 0.18]	0.18
EP elections: ID (%)	0.071	0.017*	[0.01, 0.13]	0.01
EP elections: RE (%)	0.171	0.000***	[0.12, 0.22]	0.31
Landuse share: forest (%)	0.069	0.019*	[0.01, 0.13]	0.03
Landuse share: industrial (%)	0.075	0.005**	[0.02, 0.13]	0.36
Landuse share: residen- tial (%)	−0.068	0.030*	[−0.13, −0.01]	0.25
Landuse share: parks (%)	−0.071	0.026*	[−0.13, −0.01]	0.29
Amenities per km ²	0.065	0.031*	[0.01, 0.12]	0.24
Bike shops per km ² (log)	0.060	0.038*	[0.00, 0.12]	0.30
BSS: operating time (s)	−0.088	0.003**	[−0.14, −0.03]	0.30
BSS: operating area (km ² , log)	0.147	0.000***	[0.08, 0.21]	0.19
BSS: bikes per km ² (log)	0.261	0.000***	[0.20, 0.33]	0.45
BSS: share electric bikes (%)	0.081	0.006**	[0.02, 0.14]	0.19

All tested variables are listed in subsection 6.3. (Coefficients, p-values, and confidence intervals for standardized explanatory variables; confidence intervals for coefficients use a confidence level of 0.95; *denotes significance at 0.05 level, **significance at 0.01 level, and ***significance at 0.001 level)

observed at multiple points in time and available in the Nomenclature of Territorial Units for Statistics (NUTS) format. The level of geographical accuracy varies between statistics and ranges from NUTS-0 (i.e., country level) to NUTS-3 levels, which correspond to districts in Germany or departments in France. Missing values are imputed by the median over present values of the same NUTS level and point in time. To merge data accessed from Eurostat with the KPIs of individual BSSs, the most recent observations on the most detailed available NUTS level are chosen, with older non-imputed values being preferred over those with originally missing observations.

OSM provides important geographical information on each BSS's operating area. The OSM dataset can be queried for available infrastructure like designated cycling lanes or public transport stops per square kilometer, but also land use, like the relative size of residential or commercial areas.

We also utilize multiple additional sources for additional predictors. Weather observations from governmental meteorological agencies are accessed via Meteostat (Lamprecht et al. 2023) and used to determine the frequency of adverse weather conditions like precipitation or temperatures outside a comfortable range. As part of the WorldPop project, Bondarenko et al. (2020) provide the count of people living within 100 m grids, which we use to compute the average population density within each BSS's operating area and to normalize other indicators, e.g., to calculate a bikes per capita predictor. We also include indicators of national bicycle helmet legislation from Statista, political tendencies based on the 2019 European Parliament election data from the European Union (Directorate-General for Communication et al. 2019) and the European NUTS-Level Election Database (EU-NED) dataset (Schraff et al. 2023) offering a more fine-grained perspective at the NUTS-3 level, the hilliness of the operating area from the Shuttle Radar Topography Mission (SRTM) observations (Earth Resources Observation And Science (EROS) Center 2011) and BSS's system information like operating hours from the Nextbike GBFS API. Some predictors are additionally computed from the primary data, like the number of bikes per square kilometer of the operating area or the percentage of electric bicycles among a BSS's fleet.

The predictors from all these sources are aggregated (respectively disaggregated) to the level of single BSS for the statistical model building.

Preprocessing and descriptive analysis

The extracted trip data contains a small number of unrealistic or otherwise invalid trips. For example, service trips for rebalancing or repair should not be counted towards the daily trips of a BSS. We therefore applied the following filters to trips before computing KPIs or, in general, conducting the analysis below:

- Minimum distance of 50 m and maximum distance of 20km. The minimum distance accounts for potential GPS jitter or reflections, while trips longer than the upper limit are likely repositioning or maintenance movements.
- Minimum duration of 90 s and maximum duration of 8 h. The lower limit poses the feasible minimum time to physically conduct a rental and return, while the upper limit accounts for maintenance rentals or workshop visits.
- Removal of trips with invalid start coordinates by checking if they are within the valid range for geographic coordinates (longitude between 180 degree, latitude between 90 degrees)
- Reasonable cycling average speed (i.e., between 0.25 km h^{-1} and 15.0 km h^{-1}) (With an adjusted speed due to the usage of direct distances instead of real traveled distance, 15 km h^{-1} is a reasonable upper limit (Willberg et al. 2021; Schröder and Gotzler 2021). This is undermined by the fact that only 0.47% of trips are faster than 15 km h^{-1} (direct distance).

Furthermore, to be able to distinguish between systems of different sizes, we define the following categorizations by the number of bikes available in the system:

- Small systems: 10–100 bicycles
- Medium systems: 101–500 bicycles

- Large system: more than 500 bicycles

After preprocessing in the above manner, we analyze the collected data thoroughly regarding spatial, temporal, and operational characteristics in a descriptive manner. We thereby focus on general system characteristics like the ratio of bicycles to the area, usage patterns like rental durations, and operational aspects like levels of charge in electric BSS. Systems are distinguished by organizational form (free floating, station based or hybrid) and propulsion type (human-powered or pedelec/e-bike). The key performance indicator used to evaluate the performance of BSS is trips per day and bike (TDB). TDB is a common metric to assess BSS (Médard de Chardon et al. 2017; Fishman 2016) as well as other forms of shared micromobility, like shared electric scooters (Li et al. 2022), and provides a meaningful benchmark since the early days of bike-sharing (Cycling Industries Europe 2023).

Time series clustering & classification

We apply time series clustering on the relative hourly rental rates to gain an understanding of characteristic diurnal usage patterns. A cluster thus denotes a characteristic profile of rental activity through the day. We differentiate between weekdays (Monday - Friday) and weekends (Saturday and Sunday). The number of clusters was determined by analyzing the silhouette coefficients for different numbers of clusters. We use the Euclidean distance to compute both cluster affiliation and barycenters. We do not apply dynamic time warping to conserve eventual shifts in otherwise similar distributions, as we consider those shifts as relevant characteristics of the diurnal usage.

Once the clusters are computed, we develop a set of state-of-the-art predictive classification models with the purpose of understanding which factors are decisive for cluster affiliation. We deploy multiple predictive models and rely on the scikit-learn Python library (Buitinck et al. 2013) to implement them: decision tree (DT), gradient boosting (GB), k means (kM), k nearest neighbours (kNN), support vector machine (SVM), neural network (NN) (Multi Layer Perceptron), and RF. DT, RF, and GB belong to the group of tree-based models, recursively splitting the data based on feature values with the goal of achieving increasingly homogeneous subsets. While they are generally versatile and able to capture non-linear relationships, they are prone to overfitting and may be hard to interpret. We also apply distance-based clustering (kM) and classification (kNN) models, relying on the distance between data points. They are relatively simple and do not assume any specific underlying distribution, which makes them applicable if knowledge of the data is limited. However, they are sensitive to outliers and require a choice of their hyperparameters, which can be laborious. With SVM, we aim to leverage the strengths of margin-based supervised learning, robustness to overfitting through regularization. The computational complexity has proven to be challenging, however. Finally, we train a neural network to exploit the ability of the multi-layer perceptron of learning complex, non-linear relationships, which comes at the price of lacking interpretability. Each classification model is trained to predict which time-series cluster a city will belong to based on the complete set of features. We apply grid search (Bergstra and Bengio 2012) to systematically find the best hyperparameters for the respective model. The respective tested and chosen values are listed in Table 5, 6, 7, 8, 9, and 10. In DT, RF, SVM, and GB, we apply 5-fold cross validation. To compare the performance of the models, we evaluate a multitude of established metrics (again, using scikit-

learn) as defined by Fawcett (2006): accuracy, precision, recall, F1-score, mean squared error (MSE), and area under the ROC (receiver operating characteristic) curve (ROC AUC) (the latter as defined by Bradley (1997)). For each model, we analyze the five most influential features in the best-performing parametrizations to understand which ones are decisive for the temporal characteristics of BSS usage.

OLS model

To understand which impact the various variables we collected and described in subsection 3.2 have on our main KPI, TDB, we fit a regression model. We treat the logarithm of the TDB as the dependent variable, as this enabled higher prediction quality R^2_{adj} later on. The $\log(\text{TDB})$ is then predicted based on the collected secondary data as independent variables.

Before we can construct any models, we first apply \log or \exp transformations to variables wherever doing so linearizes the correlation to the TDB. An example of such a transformation can be found in Fig. 3 for the case of hilliness. Variables are iteratively removed until all Variance Inflation Factor (VIF) (Marquardt 1970) scores are no larger than the common conservative limit of five to avoid multicollinearity issues (Sheather 2009). With 106 (remaining) variables and $n = 267$ BSSs in this evaluation, especially nonlinear models using all variables would strongly overfit. In order to counteract this tendency of complex models, we decided for an OLS model. Greedy model selection techniques of bidirectional (Efroymson's) stepwise regression have been successfully applied in similar research (Alcorn and Jiao 2023; Duran-Rodas et al. 2019a): It is an iterative process, with each of its steps either removing one of the model's variables or adding one of the remaining variables to the linear model, whichever action and variable improves a model performance metric beyond a certain threshold (Hocking 1976). In our case, the Akaike Information Criterion (AIC) (Akaike 1974) is utilized to identify an adequate trade-off between quality of fit and overfitting and is applied during the iterative model development to choose the respective variable to keep/discard.

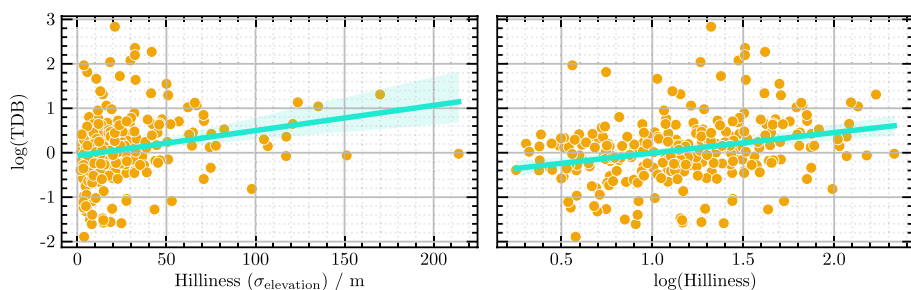


Fig. 3 Rental frequency (TDB) and hilliness computed from SRTM observations of 267 cities. We define Hilliness as the standard deviation σ of terrain elevation in the operating area according to SRTM data. Variable before (left) and after (right) logarithmic transformation. A weak positive correlation of high significance can be seen (Table 18)

Results

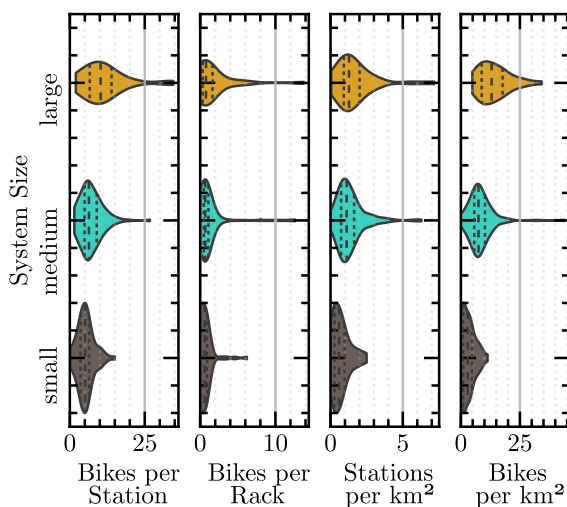
System characteristics

Before conducting advanced analyses on the dataset, we thoroughly describe the data at hand. We thereby provide a comprehensive overview of bike-sharing systems in Europe. The data set comprises 267 BSS with 94738 bicycles in total. The majority (91.6% / 86755) of recorded bicycles are classic, human-powered bikes, 6388 or 6.7% of bicycles are electric powered, 295 or 0.3% cargo-bikes and 1300 or 1.3% which could not be classified and were therefore left out of the analysis. 55% of bicycles (51738) belong to free floating BSS. In total, we recorded 26696138 trips with a line-of-flight distance of 45243349 km. Firstly, we analyze system characteristics in Table 1.

Using the classification introduced in subsection 3.3, the dataset comprises 145 small, 95 medium, and 35 large systems by our definition. In Fig. 4, we present details on statistical indicators of the system categories.

For values related to stations and/or racks, we only consider hybrid or purely station based BSS. All specific indicators (bikes per station/rack, bikes per square meter, and stations per square meter) are strictly proportional to system size. This means smaller systems exhibit less dense coverage by stations and racks as well as by bicycles. With a percentage of 60% of systems being station based, Large BSS are more likely to be hybrid or free floating compared to medium or small systems with respective likelihoods of 78% and 82% (Fig. 5). Regarding the type of propulsion of the deployed bicycles, there are distinct differences in the share of mixed propulsion BSS as well as electric BSS. With 43%, almost twice as many large systems possess a mixed fleet of bicycles as small or medium ones. However, with 9.8% and 4.5% respectively, medium and small systems are more often pure electric BSS.

Fig. 4 Statistical distribution of station occupancy and service density in the analyzed BSS. Inner lines denote the quartiles



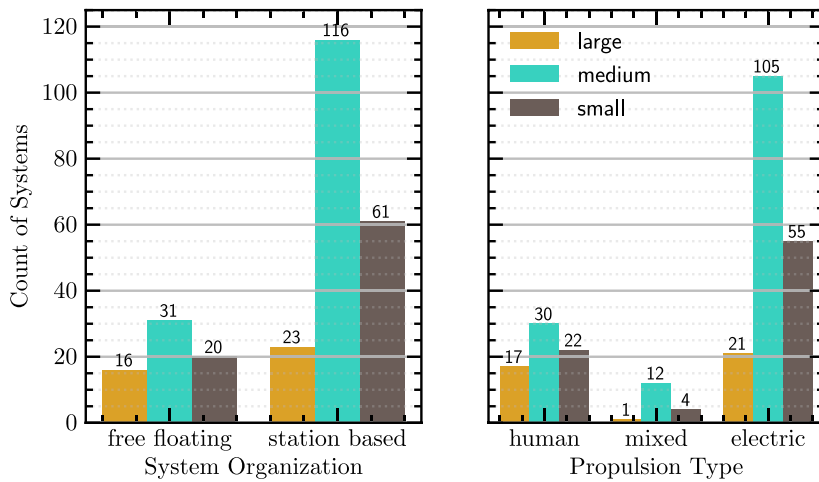


Fig. 5 Overview of BSS systems analyzed. Most systems are station-based, i.e., bikes have to be returned at fixed locations. The fleets in most systems are mixed between electric and conventional bikes. Small fleets (*less* 100 bikes) are more likely to be composed solely of conventional bikes

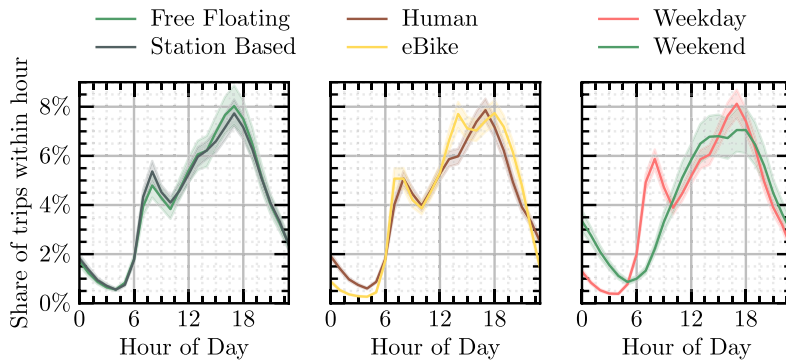


Fig. 6 Averaged temporal distribution of trip starts for operating scheme, times, and propulsion types in the 267 BSS systems. Confidence intervals are displayed as shades

Trip characteristics

This section investigates the characteristics of trips conducted in varying systems and on different kinds of bicycles. First, we take a look at temporal characteristics, which are in line with previous research on temporal patterns of urban mobility. Figure 6 shows the hourly relative shares of trip starts, each subplot for a pair of opposing characteristics. Between station based and free floating systems, there is no significant difference in the temporal distribution of trips to be found. Both system types exhibit mostly identical curves with two peaks. In station based BSS, the morning peak is slightly higher while the one in the evening is slightly lower than those of free floating BSS. The direct comparison of human-powered and electric bicycles shows only minor differences. Both types show two peaks; however, the non-electric bicycles exhibit a distinct peak at 5 p.m., while e-bikes reach a

plateau between 2 and 7 p.m. Furthermore, the nightly activity of e-bikes is lower than that of classic bicycles. Clear differences are visible when comparing weekends and weekdays. During the week, the curve exhibits two peaks, one at 9 a.m. and one at 5 p.m. In between, shares of trip departures drop during the morning but later keep rising. After the second peak, shares drop continuously with little activity during the night (midnight to 5 a.m.). On weekends, the hourly shares show how bike-sharing serves as a complement to PT and facilitates mobility needs during low service hours. While the hourly shares during the evening (7 p.m. - midnight) are similar to weekdays, there is significantly more activity during the night, when PT is inactive in most cities. Over the course of the day, we observe a steady rise during the morning, leading to a plateau between noon and evening.

We compare the rental duration in the same categorical manner in Fig. 7. In both free floating and station based BSS, most rentals take between 5 and 10 min (around 33%). However, short trips (5 min or shorter) are more common in free floating BSS (2.5% more than in station based). On the other hand, the shares of rentals taking more than 10 min are reduced. This can be explained by the detours caused by stations not being located at the desired location. The time penalty is unpleasant for the user but generates more revenue for the operator since the price is directly related to rental duration in most systems. In many systems, however, there are billing periods of 15 min (often 30 min for e-bikes), which contradicts the aforementioned argument and supports the hypothesis that the differences in rental duration are due to other effects and not strictly related to attempts to minimize user-facing costs.

A similar pattern is visible for electric and non-electric bicycles, with e-bikes being rented less often for short periods (under 5 min). Instead, the shares of 5 min to 10 min and 10 min to 15 min are higher. This may be explained by the aforementioned billing period of 30 min (mostly, 2€ for each started 30 min), making it less attractive to rent a bicycle for very short periods of time. Like in the previously discussed distribution of trip starts, the largest differences are to be found between weekends and weekdays. All bins of rental duration 15 min or less exhibit significantly smaller shares (up to 6% for 5 min to 10 min) on weekends, and all bins longer than 15 min yield higher shares. An explanation for this finding might be found in the absence of commuting trips conducted intermodally with a shared bike and PT.

Finally, we investigate average speeds for the previously introduced categories in Fig. 12. It has to be mentioned that absolute values can only serve as a qualitative indicator as we consider direct distances. Real speeds are thus expected to be higher. Furthermore, we naturally cannot identify round trips due to the methodology. Although the described errors

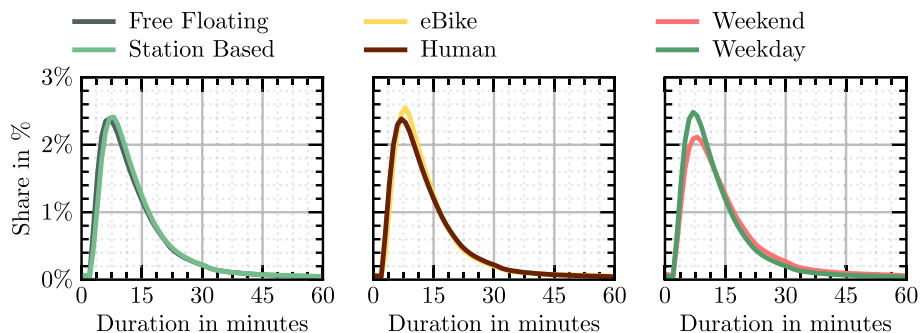


Fig. 7 Distributions of rental durations of trips in different BSS system types and operating times

induced by the methodology are equal to all categories, we consider the data rather as a qualitative indicator and, therefore, moved the graph in question to the appendix. One might expect large differences in speed between electric and purely human-powered bicycles. The measured difference in our study is 0.7 km h^{-1} or 9.4%. This transfers directly to time savings: on the most frequently traveled distance of 1.5 km to 2 km, 9.4% of speed increase leads to a time saving of 72 s on a total travel time of about 14 min. Between free floating and station based BSS, the difference in average speed amounts to 0.5 km h^{-1} (or 7.6%) in favor of free floating systems. This can be explained by the share of e-bikes, as these travel faster in free floating BSS. The traveled speed on weekdays is 4.7% (7.65 km h^{-1} and 8.01 km h^{-1}) faster than those on weekends, probably due to general work-related time pressure.

Electric BSS

As the information on electric BSS, and especially large-scale datasets are few, we investigate the technical characteristics of electric BSS. Firstly, we look into energy consumption. The mean energy consumption of shared e-bikes is 13.8 W h km^{-1} direct distance (this equals about 10.6 W h km^{-1} driven distance (Schröder and Gotzler 2021)). Our data reveals a strong correlation to temperature, as shown in Fig. 8. The displayed temperature relates to the measurement at noon of the respective day. The median value of specific energy consumption at zero degrees Celsius is 17.5 W h km^{-1} , while it decreases to 7.0 W h km^{-1} (both direct distance) when temperatures rise above 30°C . The order of magnitude of outliers is equal for all temperature categories and, therefore, does not suggest methodological errors. These results support the hypothesis that the temperature in which a BSS operates strongly influences the service in terms of recharging or battery swapping required to run the system.

Other authors have demonstrated that idling energy loss is significant in electrified shared micromobility, making up from 15% up to 55% of total energy consumption (e.g., Wang et al. (2021)). In our dataset, idle energy loss amounts to an equally significant number of 4.6 W on average (to put this in perspective: consumption is around 10 W h km^{-1} , meaning an hour of parking costs half a kilometer of driving range). Like the consumption, idle energy loss is related to the temperature. From low temperatures between 2 and 6°C and

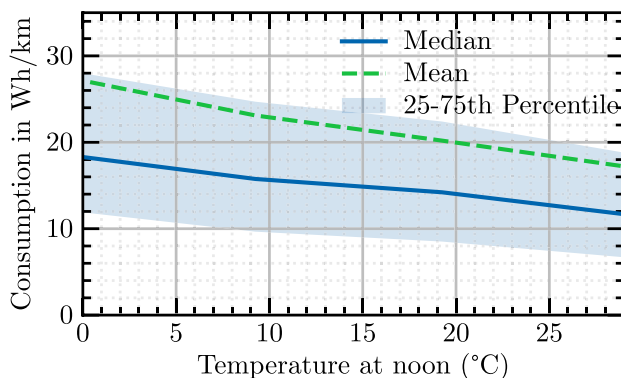


Fig. 8 Energy consumption on single trips in Wh/km over local noon temperature of rental date. The state of charge is discretized in 1% steps, corresponding to approx. 600 m of range. Short trips with a SOC decrease yield frequent outliers

up to 30 °C, we measure a continuous reduction of idle energy loss counteracting the riding consumption previously discussed. The differences are smaller and less distinctive, however, so it cannot be expected that the decreased consumption and increased idle energy loss cancel each other out. Higher energy consumption and, therefore, more frequent recharging/battery swaps are to be expected with declining temperatures.

While we can not track the operator's activities, we can gain some insights into the operation of electric BSS. The average difference in SOC when a vehicle is charged or its battery is swapped is 43.4%. This leads to the presumption that there is relevant potential left in the operational organization of electric BSS, e.g., by charging/swapping less often and therefore reducing the number of kilometers traveled by service vehicles. Looking at the detailed histogram of change in SOC when charged (Fig. 9), there is an accumulation of small amounts of energy but no clear difference for amounts of 30% SOC or more, with the exception of a spike at 90%. This might be due to the operator considering this point critical and setting out to charge/swap the respective bicycle specifically. The observed charging frequency/swapping is 4 days and 19 h on average. Figure 9 shows the distribution of time periods between charging events. It is clearly visible that the systems follow specific charging/swapping patterns as the frequencies accumulate around full days. In general, the frequency of charging processes decreases over the length of the time period, while a large share of e-bikes is being recharged daily. Bicycles are also, on average, charged (or their battery swapped) while possessing a remainder of 49.2% SOC. E-bicycles are mostly charged (or their battery swapped) at SOC levels between 10 and 90%. There is also a slight peak at 100%, meaning batteries have run completely empty. The aforementioned numbers further substantiate the assumption that less frequent service intervals might be sufficient. However, it is not possible to determine that, looking at a BSS as a whole, less recharging/swapping (and therefore fewer service kilometers) is generally desirable as a multitude of other factors like repositioning or battery lifetime, connected to the depth of discharge, influence overall system cost and sustainability.

In order to estimate the battery lifetime in electric BSS, we compute the average SOC amount that is charged per day, which is 31.15%. With an expected lifetime of 400 full charging cycles (Seyerlein 2022), this equals a total lifetime of 1244 days or roughly three years and five months.

System performance

In order to assess the usage of BSS, we define the indicator TDB as the ratio of the number of trips and the days in operation of each unique bicycle. A high utilization rate, expressed

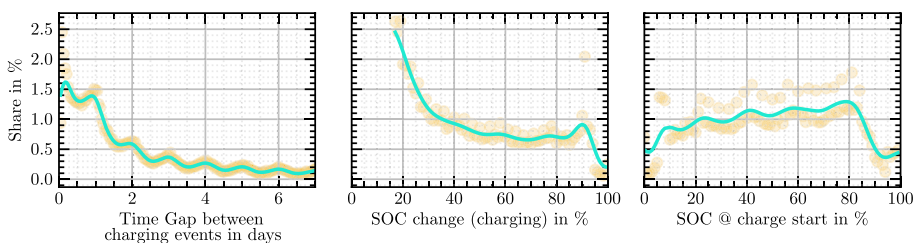


Fig. 9 Frequency of charging events, SOC increase, and initial SOC when charging (charging may also be battery swap, depending on the system)

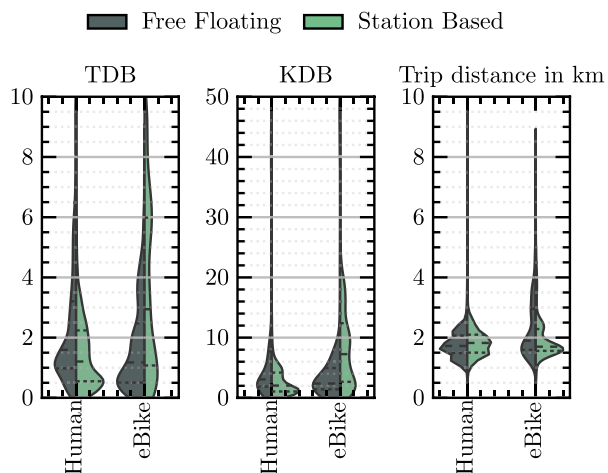
by a high TDB value, is desirable as the available assets should be used to their potential. Furthermore, we define a similar indicator, kilometres per day and bike (KDB), as the ratio of ridden kilometers (straight line distance from origin to destination) and days in operation per bicycle. We determine operational days per unique bicycle by calculating the difference between each bicycle's first and last recorded trip.² Fig. 10 includes TDB, KDB, and trip length (direct distance) for different propulsion types and organizational forms.

Analyzing TDB, both propulsion types achieve rather low values, with a majority of systems below 2. While the distribution for classic BSS (non-electric) is rather uniform around the mean value, there are more higher-ranking outliers in electric BSS. For human-propelled bike-sharing, free floating systems accomplish higher values than station based ones. The contrary is true for electric BSS, not allowing for any conclusion on the correlation between TDB and propulsion type. In line with previous findings (trip duration, Fig. 7), daily distance per bike is larger in electric BSS. In electric BSS, station based systems achieve significantly longer distances than free floating systems. Non-electric BSS show different numbers, with more short trips in station based form. As with TDB, there is no clear indication which organizational form achieves higher KDB values.

Time series clustering

To gain an understanding of the temporal usage characteristics, we cluster the diurnal distribution of trip starts for all systems. Figure 11 shows the resulting clusters for weekdays and weekends as well as BSS and electric BSS. On Weekdays, all clusters exhibit two peaks, one in the morning and one in the evening. While the first cluster has two sharp peaks, reduced activity between the peaks, and close to none during the night, cluster 2 has no sharp peaks and more activity during the day as well as in the evening/night. Cluster 3 has a significantly smaller morning peak than the other two weekday clusters and rises towards the afternoon/evening. The afternoon peak occurs earlier than in cluster 2. During the night, activities diminish. On Weekends, all clusters exhibit a singular peak or plateau over the course of 24 h. In cluster 1, the curve rises from the morning (07:00) onwards and reaches a plateau

Fig. 10 Trips per day and bike (TDB), Kilometers per day and bike (KDB), and trip distance by system properties in the analyzed 267 BSS. Inner lines denote the quartiles. The relation between TDB and KDB is specific to each city and visualized in the appendix (Fig. 13)



² Gaps in the recording are rare and mostly global across all systems as discussed in subsection 3.1

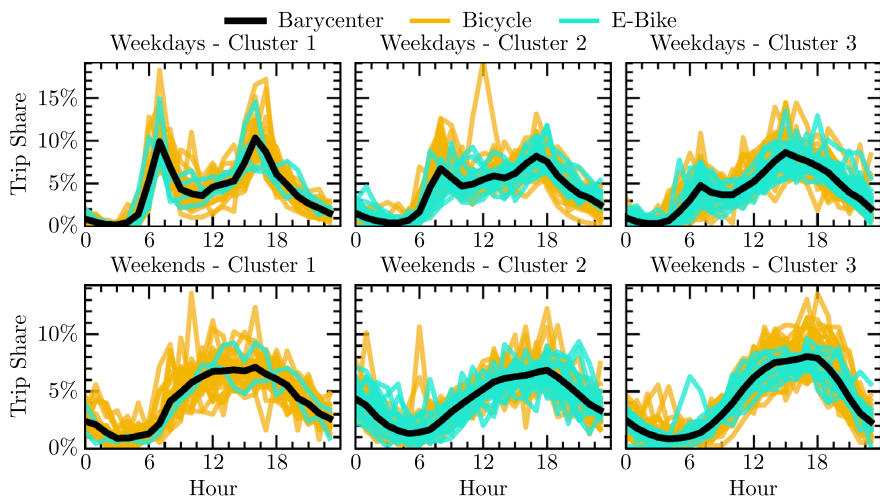


Fig. 11 Time series clusters of shares of trip starts for weekdays and weekends, by propulsion type. (Bary-center denotes the dynamic arithmetic mean.)

between 11:30 and 16:30. Significant usage during the night is visible. Cluster 2 on weekends reaches its peak at 6 pm and has the highest nocturnal activity of the three clusters, close to that during daytime. Cluster 3 shows a plateau in usage between 13:30 and 18:00 and less nocturnal activity compared to cluster 2. 97% of clusters stay in the same group between weekdays and weekends (e.g., cluster 1 for weekdays and cluster 1 for weekends), further substantiating the existence of characteristic usage patterns. Almost all electric BSS belong to clusters 2 and 3.

In order to draw conclusions on the variables' influence on usage patterns, we construct seven classification models predicting cluster affiliation, which was previously computed via time series clustering. Table 2 shows the models' performance. SVM outperforms all other models in accuracy, precision, recall, and F1-score and shows the smallest MSE. We attribute this exceptional performance to SVM's ability to deal well with high-dimensional spaces, robustness, and the extensive possibilities for hyperparameter-tuning. SVM is closely followed by RF. The good (with respect to the other models) results of RF are likely due to its robustness to imperfect data (noise, scaling, missing values, etc.) and its ability to perform feature selection inherently (Breiman 2001). Stemming from the same basic idea as RF, GB, and DT, both deliver satisfying results, too. Neither the supervised kNN nor the unsupervised kM achieve acceptable results, with accuracies around 50 % and large MSE in the case of kNN, which might be due to the model's reliance on distance metrics and the high-dimensional dataset. The NN ranges in a similar, comparatively bad prediction quality.

We list the five most important features in each model to understand which city features are relevant when determining the temporal usage patterns (Table 3). Due to the previously described differences in performance, we focus on the most promising models regarding prediction quality. SVM delivers the best results, slightly outperforming RF in accuracy, precision, recall and F1-score and a significantly lower MSE. SVM, DT, and GB yield the exact same set of indicators. The three models emphasize financial indicators (e.g., housing cost, the share of persons at risk of poverty) and openness for innovation (participation

in collaborative economy, trust in technology). RF, likewise, includes financial (household income, share of persons at risk of poverty) but additionally includes bicycle infrastructure (nr. of bike parking spots per km^2) amongst the five most important features. Overall, the prediction quality of the models is mixed, and analysis of feature importance shows how complex the interrelation of temporal characteristics and city variables is. However, the consistency of the most important features between the best-performing models suggests that the respective characteristics correlate with bike-sharing usage patterns.

Statistical analysis & OLS model

To further leverage our extensive database of 267 BSS and 108 city-level indicators, we conduct statistical correlation analysis. A straight-forward correlation analysis using Spearman Correlation (Table 18) yields few meaningful results: All statistically relevant variables (confidence level 0.05) show weak Spearman correlation with TDB. Therefore, we implement an OLS model based on the predictor selection of stepwise regression, allowing us to compute relevant determinants in a more sophisticated way. The final model with an AIC of 280 shows that BSSs are affected by both internal factors and external factors of various kinds. Our model exhibits good predictive power with an R^2 of 0.624 and an $R^2_{adj.}$ of 0.598, with comparative research commonly in the 0.4 to 0.55 range for $R^2_{adj.}$ (Zhao et al. 2014a; Médard de Chardon et al. 2017; Bielinski et al. 2019; Arias-Molinares et al. 2023). We refitted the final model after standardizing the explanatory variables. The regression coefficients thus make the impact of the variables on the TDB comparable. The non-standardized model is provided in Table 19.

Five of the model's 19 independent variables (Table 4) address the potential user base of a BSS, all of them statistically significant. The significant influence of both the share of below-thirty-aged residents aligns with the results of earlier studies of variance across gender and age group—generally declining in the early thirties. International travelers seem to have an overall positive impact on the TDB, as international short-term apartment rentals influence the TDB strongly positive, while domestic stays are moderately negative. All of these correlations are backed up by weak-to-medium positive Pearson correlation coefficient (PCC).

(Julio and Monzon 2022; Kuo et al. 2021; Zhou et al. 2019). Attitudinal characteristics of potential users also play a (statistically) significant role, with the election results of the European Parliament liberal factions (RE), conservative faction (ECR), and right-wing factions (ID) acting as their proxy. As these variables are only available on NUTS-0 (national state) level, these might also be a proxy for overarching attitudinal differences between the countries the BSS are located in.

Four BSS properties were selected for the final model, and all four yield significant p-values: Both the size of the operating area and the density of bikes within it make a BSS more accessible and convenient to use and therefore increase the TDB. A less pronounced positive effect can be observed for the share of electric bikes among the BSS's fleet. Only nine BSSs, all Spanish, are not operating during the night hours. As these nine systems have above-average TDB values, this partially explains the negative regression coefficient of operating hours in our model.

Naturally, the environment of a BSS also affects its performance, with six environmental variables included in the model, all statistically significant. In particular, existing bicycle

infrastructure increases the TDB. Land use plays a significant role, too: Industrial areas, as well as forests within a BSS's operating area have a positive impact, while the opposite is true for parks and residential areas. The latter is only true within the constraints of the OLS model – both variables have positive Pearson and Spearman correlations with the TDB (see last column of Table 19 and Table 18). We thus have to conclude that within the OLS model, the possibility to calculate a delta between the land use shares of parks and residential areas and other variables increases the predictive power. A possible hypothesis to derive and test would be that a high population density coupled with a low land use share of residential buildings indicates mixed land use that encourages bike-sharing. The testing of such hypotheses greatly overreaches the scope of this article, but we would like to encourage further research on the interaction between urban planning and bike-sharing using the open data we provide along with this article.

Many of the selected variables represent physical barriers to BSS usage. While the number of variables representing informational barriers is smaller, the large coefficients for the sharing economy indicators hint at the still strong impact of informational barriers. Additionally, it is to be mentioned that the computation of KPIs over the entire collection period leads to the elimination of some variables that show strong correlations on a more fine-grained level, for example, weather conditions like temperature or precipitation. However, variables computed over the entire collection period, like the share of hours with temperatures outside a comfortable level or the percentage of hours with precipitation or snowfall, are not selected by the stepwise regression.

Discussion & conclusion

Conclusion

This paper provides a comprehensive, data-driven overview on 267 European BSSs and their performance. The system-level data is amended by numerous city-level features, facilitating statistical studies on the interrelations of the two categories. Besides explaining the data-mining process in detail, we publish the complete dataset open source. The publication of the dataset may foster further research on bike-sharing, leading to more sustainable (ecologically, economically, and socially) urban transportation systems. Furthermore, we conduct a descriptive analysis and build models to understand the influence of the respective parameters, underlining the dataset's potential.

The collection of exhaustive BSS data allowed for a multitude of insights. Looking at system characteristics, a majority of BSS are station based, especially smaller systems. The ratio of bicycles per rack (in free floating BSS oder hybrid bike-sharing) is roughly independent of the total number of bicycles; however, we observe higher station and bicycle density for large systems than for medium and small ones. While between 25 (small systems) and 43% (large systems) include pedelecs, only a minor share of 8% of total bicycles are electric. Purely electric systems are mainly medium-sized. Those findings suggest there's extensive potential for electrification.

Our findings regarding temporal trip distribution are in line with previous publications: two peaks on working days and a longer plateau over the afternoon, combined with higher nocturnal activity. The data reveals no difference in temporal usage patterns related to pro-

pulsion type or organizational form. Although electric motors ease the physical strain of riding a bike, we only record slight differences in average speed between e-bikes and classic ones. We thus assume comfort to outweigh travel time when choosing between electric and conventional shared bikes. Our analysis shows that trip length and duration are mostly influenced by organizational form: significantly more short (both in distance and time) trips below 1 km occur in free floating BSS than in station based BSS, probably due to involuntary detours necessitated by station placement. While this benefits the operator, it is assumed to be detrimental to the usage level. E-bikes, in contrast, are less often used for very short trips, likely due to different pricing schemes, including a fixed base fee. Regarding performance and profitability through usage maximization, expressed by TDB, electric BSS achieves better performance than classic pushbikes. Consequently, KDB values are also higher at similar trip lengths. We do not observe a clear relation between organizational form and performance, as the influences are relatively small and different between propulsion types. We, therefore, do not expect major changes in performance in case BSSs change their operational form.

As analyses of electric BSS are sparse, we specifically emphasize this topic and especially shed light on the operations. Our data shows that shared e-bikes are recharged frequently and often with small amounts of energy, indicating significant potential for optimization of charging/battery swapping. On the other hand, only a few bicycles end up with empty batteries. The data shows clear proportionality between temperature and idle energy loss as well as energy consumption while driving, with the latter prevailing (close to a factor of three between cold and warm environments). While the energy cost for driving itself is of secondary importance from a financial point of view, increased consumption requires strong adaptations in operating strategy and incurs costs, therefore.

Clustering diurnal shares of trip starts, we identify three distinct temporal usage patterns in the dataset, both for weekends and weekdays. The main differences lie in nocturnal activity as well as differences in size between the morning and evening peaks, which all clusters exhibit on weekdays. On weekends, all systems are most active during the afternoon, differing between a slight peak or a plateau. As on weekdays, nocturnal activity varies between clusters. Building seven predictive classification models with mixed performance, we discuss the most relevant features of the three best-performing models. Results show that economic indicators and active mobility-oriented infrastructure possess the largest influence. However, the results of the prediction models also include potentially surprising variables such as crime index and crime rate. Previous studies have shown interrelations of wealth, crime, and innovation (included, e.g., through the variable "trust in technology") in urban areas (Bettencourt et al. 2010). This shows that such variables, which are often overlooked in urban planning or mobility analysis, are relevant factors. Being able to predict the main usage patterns of a prospective BSS allows the operator to plan the operations of the system accurately before its implementation. Especially in the context of public tenders, this knowledge is crucial and may decide whether a multiannual contract is profitable.

The OLS model resulting from stepwise regression revealed further findings. Comparison to previously published models proves superior predictive performance with an R^2 value of .66. In line with earlier publications, we identify gender and age as the major demographic factors. This strong demographic inclination towards young male persons suggests there's a need to advocate social equity in the usage of SMM. This finding supports results from prior publications (Singh 2020). Political sentiment, expressed by recent election results, is

a further statistically significant determinant, underlining the importance of political support in transforming urban mobility systems. While system characteristics are generally difficult to correlate with performance (e.g., successful systems are likely to grow - therefore, the absolute size or number of bicycles may be misleading), the model shows that relative factors like the density of bicycles per area or bicycle electrification, shedding more light on findings from (Cycling Industries Europe 2023). Similar effects, both of vehicle density and well as electrification, have been found by (Reck et al. 2022). Vehicle density being more important than absolute fleet size shows that operations, encompassing e.g., vehicle allocation, repositioning, or system design, yield significant potential for improvements in efficiency and sustainability. As sustainability is still a critical factor for SMM, and many systems do, in reality, not provide a sustainable alternative to the private car (Reck and Axhausen 2021), these findings should be considered both by decision makers and system operators. Again, in line with existing literature, our model identifies built environment, like cycling infrastructure and land use, as key factors, delivering quantitative evidence for the necessity of investments in order for bike-sharing to be a success. Economic factors like overnight stays or general familiarity with the shared economy, expressed by, e.g., accommodation rental, exhibit major influence. This is likely interconnected to the economic status of a city and, thus, to the existing infrastructure. Those findings on success determinants of BSSs can serve as a basis for future decisions, both for local politicians and for operators designing a new system or discussing the expansion of an existing one.

The comparison of stepwise regression and its OLS model to a classic Pearson correlation analysis sheds light on the strengths and weaknesses of both tools. While the model shines in automatically selecting the relevant factors from a large set, the process of stepwise regression is prone to eliminating relevant factors, like in the case of recreational areas, which show highly significant Spearman correlations but were not selected by the model. It is, therefore, crucial to combine descriptive analysis with basic and advanced methodologies as well as expert knowledge to avoid false conclusions.

Limitations

The presented paper presents a large-scale open-source dataset and includes multiple relevant findings derived from statistical modeling. However, potential shortcomings both in the process of data collection and processing, as well as in the statistical modeling, should be known to the reader when interpreting results or conducting their own analyses. Regarding primary data collection, potential errors might occur in the calculation of trips from bicycle position data, as we can not directly differentiate between actual rides and operator-based rebalancing. This distinction is achieved by processing and filtering as described in subsection 3.3, and the processed results are plausible, but the ground truth is unknown to the authors. Furthermore, accuracy is limited due to discrete intervals in data collection (i.e., querying of the API). As we only consider data from a single BSS operator, the data is potentially skewed by operator-specific strengths or weaknesses of the systems. Nextbike claims to be the European market leader in bike-sharing (nextbike by TIER 2023), and we are therefore confident in providing a conclusive view into the bike-sharing market by analyzing the entity of their systems. Regarding secondary data, whilst using consistent sources, it has to be mentioned that some of the data provided by the European Union are not available for the finest NUTS levels or the most recent years; that is why the data is, in minor parts, imputed as outlined in subsection 3.2.

The OLS model developed by applying stepwise regression delivered reliable data on the correlation of BSS performance and city-scale variables. While stepwise regression is a popular approach to model selection (Whittingham et al. 2006), it is not without its downsides: The most fundamental downside is that some spurious correlations with the dependent variable are guaranteed in large collections of variables. Any automated procedure may select such spurious correlations over causally related variables, artificially inflating the R^2 scores of resulting models. Furthermore, a large number of models were fitted to the data during the model selection procedure. This means that some of the assumptions behind the p-values and confidence intervals shown for the selected model are no longer fully met, and hence, some caution is advised when using these measures (Smith 2018). Nevertheless, both the selected model in Table 4 and the correlation analysis of all collected variables in Table 18 show that BSS are affected by a wide range of factors from multiple domains. With an R^2 of 0.66, only 34% of the variance within the data set remains unexplained. We propose further research on the remaining influencing factors, such as geographical, socio-demographic, or cultural aspects, not covered within this study.

Future work

As the presented paper shows, the potential of large-scale bike-sharing analyses is evident, and there are multiple directions for future research, with the most obvious one being the application of our open data set for various use cases within the scientific community. On the other hand, the provided primary data could be enriched further by including data from multiple operators, while secondary data might benefit from the contextualization within the larger traffic system of a city. The connection to other modes of transport, e.g., understanding interactions with the PT network, or adding routing information between start and end points, are significant gaps. More sophisticated statistical models, or the analysis on daily or even hourly resolution using deep learning methods, might generate additional insights. Regarding the application of results, direct interaction with policymakers and operators yields great potential when interpreting the findings. By publishing the complete dataset, we hope to enable further studies on bike-sharing and thus contribute to improving urban transportation systems.

Appendix A

See Tables 5, 6, 7, 8, 9, 10 and Figures 12, 13.

A.1 Model values

The following variables were used in the grid search as well as the final configurations for the respective models:

Table 5 NN (MLP) parameters

Parameter	Grid Search Values	Best Value
Hidden layer size	[(50,), (100,), (50, 50)]	(50, 50)
Activation	['logistic', 'tanh', 'relu']	logistic
Alpha	[0.0001, 0.001, 0.01]	0.001
Learning rate init	[0.001, 0.01, 0.1]	0.1
Max iter	[200, 300, 500]	200
Batch size	['auto', 16, 32]	16
Beta 1	[0.9, 0.95, 0.99]	0.95
Beta 2	[0.999, 0.9999]	0.999
Early stopping	[True]	True
n iter no change	[10]	10

Table 6 Parameters of DT

Parameter	Grid Search Values	Best Value
Criterion	['gini', 'entropy']	gini
Max depth	[None, 10, 20, 30]	None
Min samples split	[2, 5, 10]	2
Min samples leaf	[1, 2, 4]	1

Table 7 Parameters for Random Forest (RF)

Parameter	Grid Search Values	Best Value
n estimators	[100, 200, 300]	200
Max depth	[None, 10, 20, 30]	10
Min samples split	[2, 5, 10]	10
Min samples leaf	[1, 2, 4]	2
Criterion	['gini', 'entropy']	entropy
n features to select	[10, 15, 20]	20

Table 8 SVM parameters

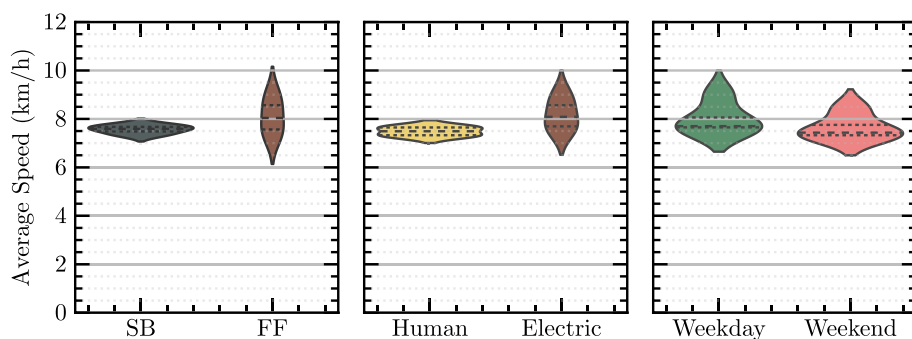
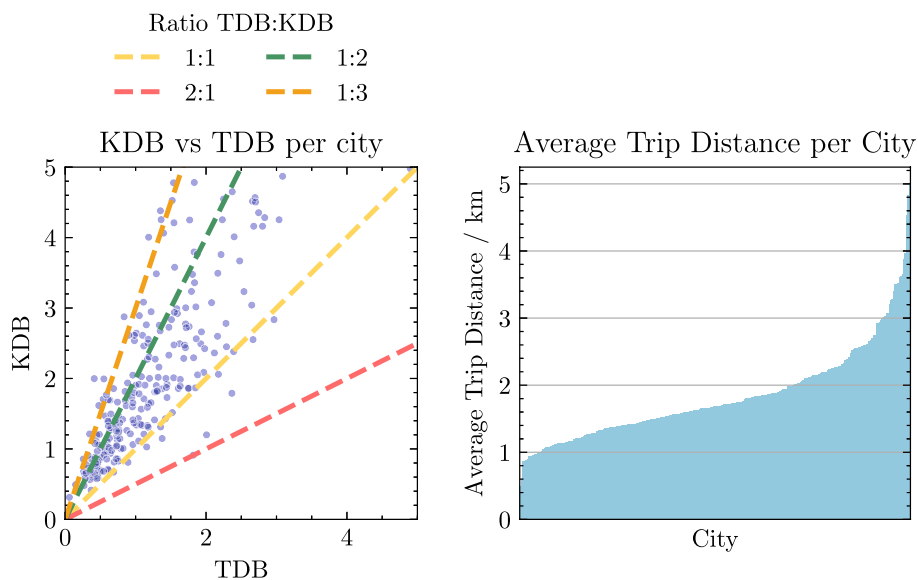
Parameter	Grid Search Values	Best Value
C	[0.1, 1, 10, 100]	0.1
Kernel	[linear, rbf, poly]	linear
Gamma	[scale, auto, 0.1, 1]	1

Table 9 Parameters of kNN

Parameter	Grid Search Values	Best Value
n neighbors	[3, 5, 7, 9]	9
Weights	['uniform', 'distance']	distance
Algorithm	['auto', 'ball tree', 'kd tree', 'brute']	auto
p	[1, 2]	1

Table 10 GB parameters

Parameter	Grid Search Values	Best Value
n estimators	[50, 100, 150]	150
Learning rate	[0.01, 0.1, 0.2]	0.1
Max depth	[3, 4, 5]	4
Min samples split	[2, 5, 10]	10
Min samples leaf	[1, 2, 4]	1
Subsample	[0.8, 0.9, 1.0]	0.8


Fig. 12 Distribution of average speeds for different BSS types. The speeds correspond to line-of-flight speeds. Inner lines denote the quartiles

Fig. 13 KDB and TDB for different cities within the dataset. The average trip distance connects KDB and TDB, it varies per city between 0.5 km and 5 km

A.2 Dataset description

The following tables describe the composition of the associated open dataset.

See Tables 11, 12, 13, 14, 15, 16 and 17.

Table 11 Description of trips. csv containing information on all reconstructed trips meeting the criteria in subsection 3.3. In total, 25210627 trips are provided. The spheroid distance is computed through PostGIS' ST_Distance method and provided for convenience

Column	Unit	Description
Bike_id	–	ID of the bike that was used for the trip, references bikes.id
City_id	–	ID of the rental city. A bike may change its city association. references cities.id
Time_start	s	Unix timestamp of rental time, coordinated universal time (UTC). Join with table cities to obtain local time zone
lon_start, lat_start, lon_end, lat_end	°	Geographical coordinates of rental and return locations in EPSG:4326
Station_id_start, station_id_end	–	Id of station the bike was associated with, NULL if free floating
Battery_start, battery_end	%	State of charge of battery before/after trip. NULL for non-electric bikes
Duration	s	Duration of trip
Distance	m	Line of flight distance between start and end point

Table 12 Description of cities. csv, containing metadata on 267 bike-sharing schemes

Column	Unit	Description
Id	–	Id of the city
Name	–	Name of the bike-sharing scheme
Lat, lon	°	Approximate geographical coordinates of city location (EPSG:4326)
Timezone	–	Time zone of the bike-sharing scheme
Country	–	Country of the bike-sharing scheme. in rare cases close to borders this differs from the geographical location
Return_to_official_only	bool	True: Bikes may only be returned to (virtual) stations. False: Bikes are free floating or hybrid

Table 13 Description of city_areas.csv, containing estimated operational areas of 267 bike-sharing schemes

Column	Description
City_id	Id of the city
Geom_ewkb	Operational area as well known binary (postgres: ST_AsEWKB), EPSG:4326
Geom_ewkt	Operational area as well known text (postgres: ST_AsEWKT), EPSG:4326
Geom_geojson	Operational area as geojson (postgres: ST_AsGeoJSON), EPSG:4326

Table 14 Description of bike_types.csv, containing information on 121 distinct bike types

Column	Unit	Description
Id	–	Id of the technical bike type
Vehicle_image	–	URL to an image of the bike
Name	–	Name of the technical bike type
Description	–	Description of the bike
Form_factor	–	Distinguishes between cargo bikes and regular bikes
Rider_capacity	–	1 for regular bikes, 2 for tandems and bikes with additional child seats
Propulsion_type	–	Distinguishes between human powered and electric bicycles
Max_range	m	Nominal maximum electric range
Battery_capacity	Wh	Capacity of traction battery in Watt-hours

Table 15 Description of bikes.csv, containing information on 88444 bicycles

Column	Unit	Description
Id	–	Id of the individual bike
Bike_type_id	–	Technical bike type. references bike_types.id
Computer_id	–	

Table 16 Description of stations.csv, containing data for 13192 stations

Column	Unit	Description
Id	–	Internal id of the station
City_id	–	ID of the city the station is placed in. references cities.id
Name	–	Short location description
App_number	–	Number as displayed in app
Terminal_type	–	Technical station type (brief name, 12 distinct type)
Place_type	–	Unknown, 23 distinct integer numbers
Bike_racks	–	Number of regular hardware bike racks. typically zero for virtual or fully-electric stations
Special_racks	–	Number of special hardware bike racks, usually electric bike charging. typically zero for regular and virtual stations
Ion,lat	°	Geographical location of the station (EPSG:4326)

Table 17 Description of station_status.csv, containing 38279885 status updates

Column	Unit	Description
Station_id	–	Internal ID of the station. references stations.id
Time	s	Unix timestamp of status, coordinated universal time (UTC). Join with table stations & cities to obtain local time zone
Bikes	–	Total number of bikes currently parked at the station
Booked_bikes	–	Number as bikes parked at the station reserved by a user
Bikes_available_to_rent	–	Number as bikes parked at the station available to rent
Free_racks	–	Free regular hardware racks, typically 0 for virtual stations
Free_special_racks	–	Free special hardware racks
Maintenance	bool	True: station is in maintenance status, false: station is online

A.3 Variable description

The variables and data sources used to predict the performance of BSS are listed in the following Table 18

Table 18 Statistical indicators used in this article

Source	Predictor	ρ	p -value	Imput	Level	Variables
Custom	Bike Index	0.09	0.158	55%	NUTS-0	'International classification of crime for statistical purposes (ICCS): 'Intentional homicide', 'Unit of measure': 'Per hundred thousand inhabitants', 'International classification of crime for statistical purposes (ICCS)'; 'Serious assault', 'Unit of measure': 'Per hundred thousand inhabitants'; 'International classification of crime for statistical purposes (ICCS)'; 'Kidnapping', 'Unit of measure': 'Per hundred thousand inhabitants'; 'International classification of crime for statistical purposes (ICCS)'; 'Sexual violence', 'Unit of measure': 'Per hundred thousand inhabitants'; 'International classification of crime for statistical purposes (ICCS)'; 'Burglary', 'Unit of measure': 'Per hundred thousand inhabitants'; 'International classification of crime for statistical purposes (ICCS)'; 'Theft', 'Unit of measure': 'Per hundred thousand inhabitants'; 'Type of household': 'Total', 'Income situation in relation to the risk of poverty threshold': 'Total', 'Affordability': 'Households with financial burden due to the housing costs', 'Unit of measure': 'Percentage'; 'Type of household': 'Total', 'Income situation in relation to the risk of poverty threshold': 'Total', 'Affordability': 'Households with heavy financial burden due to the housing costs', 'Unit of measure': 'Percentage'; 'Type of household': 'Total', 'Income situation in relation to the risk of poverty threshold': 'Total', 'Affordability': 'Households without financial burden due to the housing costs', 'Unit of measure': 'Percentage'
Custom	Crime Index	0.01	0.915	0%	NUTS-0	
Custom	Student Share	0.11	0.081	0%	NUTS-2	
Custom	Women Per 100 Men	0.02	0.730	0%	NUTS-3	
Custom	Young Share	0.25	0.000***	0%	NUTS-3	
EU-NED	Euroseptic Share	-0.18	0.004**	6%	NUTS-3	
EU-NED	Farleft Share	0.07	0.281	6%	NUTS-3	
EU-NED	Farright Share	-0.24	0.000v	6%	NUTS-3	
EU-NED	Nonvoting Share	0.02	0.781	6%	NUTS-3	
EU-NED	Populist Share	-0.20	0.001**	6%	NUTS-3	
Eurostat	CRIM_OFF_CAT	-0.10	0.116	4%	NUTS-0	
Eurostat	CRIM_OFF_CAT	0.12	0.055	4%	NUTS-0	'International classification of crime for statistical purposes (ICCS): 'Intentional homicide', 'Unit of measure': 'Per hundred thousand inhabitants', 'International classification of crime for statistical purposes (ICCS)'; 'Serious assault', 'Unit of measure': 'Per hundred thousand inhabitants'; 'International classification of crime for statistical purposes (ICCS)'; 'Kidnapping', 'Unit of measure': 'Per hundred thousand inhabitants'; 'International classification of crime for statistical purposes (ICCS)'; 'Sexual violence', 'Unit of measure': 'Per hundred thousand inhabitants'; 'International classification of crime for statistical purposes (ICCS)'; 'Burglary', 'Unit of measure': 'Per hundred thousand inhabitants'; 'International classification of crime for statistical purposes (ICCS)'; 'Theft', 'Unit of measure': 'Per hundred thousand inhabitants'; 'Type of household': 'Total', 'Income situation in relation to the risk of poverty threshold': 'Total', 'Affordability': 'Households with financial burden due to the housing costs', 'Unit of measure': 'Percentage'; 'Type of household': 'Total', 'Income situation in relation to the risk of poverty threshold': 'Total', 'Affordability': 'Households with heavy financial burden due to the housing costs', 'Unit of measure': 'Percentage'; 'Type of household': 'Total', 'Income situation in relation to the risk of poverty threshold': 'Total', 'Affordability': 'Households without financial burden due to the housing costs', 'Unit of measure': 'Percentage'
Eurostat	CRIM_OFF_CAT	0.08	0.181	4%	NUTS-0	
Eurostat	CRIM_OFF_CAT	0.03	0.660	4%	NUTS-0	
Eurostat	CRIM_OFF_CAT	-0.03	0.591	4%	NUTS-0	
Eurostat	CRIM_OFF_CAT	-0.10	0.088	4%	NUTS-0	
Eurostat	ILC_MDED04	0.26	0.000***	0%	NUTS-0	
Eurostat	ILC_MDED04	-0.03	0.679	0%	NUTS-0	
Eurostat	ILC_MDED04	-0.03	0.587	0%	NUTS-0	

Table 18 (continued)

Source	Predictor	ρ	p-value	Input	Level	Variables
Eurostat	ILC_PEPS11	-0.03	0.600	8%	NUTS-2	'Unit of measure': 'Percentage'
Eurostat	ISOC_CISCI_SP20	0.25	0.000***	5%	NUTS-0	'Individual type': 'All Individuals', 'Information society indicator'; 'Individuals use a smartphone for private purposes (3 months)', 'Unit of measure': 'Percentage of individuals'
Eurostat	ISOC_CI_CE_I	0.11	0.059	0%	NUTS-0	'Individual type': 'All Individuals', 'Information society indicator'; 'Individual used any website or app to arrange an accommodation from another individual', 'Unit of measure': 'Percentage of individuals'
Eurostat	ISOC_CI_CE_I	-0.03	0.588	0%	NUTS-0	'Individual type': 'All Individuals', 'Information society indicator'; 'Individuals used any website or app to arrange a transport service from another individual', 'Unit of measure': 'Percentage of individuals'
Eurostat	ISOC_EC_CE_I	0.18	0.003**	0%	NUTS-0	'Individual type': 'All Individuals', 'Information society indicator'; 'Online purchases (3 months) from a private person: goods, household services, transport services or rented accommodation', 'Unit of measure': 'Percentage of individuals'
Eurostat	ISOC_EC_CE_I	0.31	0.000***	7%	NUTS-0	'Individual type': 'All Individuals', 'Information society indicator'; 'Online purchases (3 months) from private persons: household services', 'Unit of measure': 'Percentage of individuals'
Eurostat	ISOC_EC_CE_I	0.18	0.003**	0%	NUTS-0	'Individual type': 'All Individuals', 'Information society indicator'; 'Online purchases (3 months) from private persons: any physical goods', 'Unit of measure': 'Percentage of individuals'
Eurostat	ISOC_EC_CE_I	0.22	0.000***	0%	NUTS-0	'Individual type': 'All Individuals', 'Information society indicator'; 'Online purchases (3 months) from a private person: rented accommodation', 'Unit of measure': 'Percentage of individuals'
Eurostat	ISOC_EC_CE_I	0.06	0.302	0%	NUTS-0	'Individual type': 'All Individuals', 'Information society indicator'; 'Online purchases (3 months) from a private person: transport service', 'Unit of measure': 'Percentage of individuals'
Eurostat	ISOC_IHOT_USE	0.18	0.003**	8%	NUTS-0	'Individual type': 'All Individuals', 'Information society indicator'; 'IoT use: for private purposes', 'Unit of measure': 'Percentage of individuals'
Eurostat	ISOC_R_BLT12_I	0.19	0.001**	60%	NUTS-2	'Information society indicator': 'Last online purchase: in the last 3 months', 'Unit of measure': 'Percentage of individuals'
Eurostat	ISOC_R_IUSE_I	0.16	0.007**	60%	NUTS-2	'Information society indicator': 'Frequency of internet access: daily', 'Unit of measure': 'Percentage of individuals'

Table 18 (continued)

Source	Predictor	ρ	p-value	Input	Level	Variables
Eurostat	ISOC_R_IUSE_I	0.20	0.001***	60%	NUTS-2	'Information society indicator'; 'Internet use: Internet banking'; 'Unit of measure'; 'Percentage of individuals'
Eurostat	NAMA_10R_2HHINC	0.02	0.770	9%	NUTS-2	'Unit of measure'; 'Euro per inhabitant'; 'Direction of flow'; 'Balance'; 'National accounts indicator (ESA 2010)'; 'Disposable income, net'; 'Unit of measure'; 'Number'
Eurostat	ROAD_EQS_CARHAB	-0.09	0.135	0%	NUTS-0	'Air pollutants and greenhouse gases'; 'Particulates < 10µm'
Eurostat	SDG_11_50	-0.14	0.024*	0%	NUTS-0	'Air pollutants and greenhouse gases'; 'Particulates < 2.5µm'
Eurostat	SDG_11_50	-0.15	0.014*	0%	NUTS-0	'Air pollutants and greenhouse gases'; 'Particulates < 2.5µm'
Eurostat	TESSI300	0.08	0.197	0%	NUTS-0	'Unit of measure'; 'Percentage'; 'Income situation in relation to the risk of poverty threshold'; 'Total'; 'Age class'; 'Total'; 'Sex'; 'Total'
Eurostat	TGS00010	-0.05	0.385	0%	NUTS-2	'International Standard Classification of Education (ISCED 2011)'; 'All ISCED 2011 levels'; 'Sex'; 'Total'; 'Age class'; '15 years or over'; 'Unit of measure'; 'Percentage'
Eurostat	TRAN_SF_ROADNU	-0.05	0.454	36%	NUTS-3	'Unit of measure'; 'Per million inhabitants'
Eurostat	TUS_00TRAVEL	-0.11	0.067	27%	NUTS-0	'Unit of measure'; 'Participation rate (%)'; 'Mode of transport'; 'Bicycles'
Eurostat	TUS_00TRAVEL	-0.03	0.567	27%	NUTS-0	'Unit of measure'; 'Time spent (hh:mm)'; 'Mode of transport'; 'Bicycles'
Eurostat	TUS_00TRAVEL	0.07	0.263	27%	NUTS-0	'Unit of measure'; 'Time spent (hh:mm)'; 'Mode of transport'; 'On foot'
Eurostat	TUS_00TRAVEL	0.08	0.216	27%	NUTS-0	'Unit of measure'; 'Time spent (hh:mm)'; 'Mode of transport'; 'Other'
Eurostat	TUS_00TRAVEL	0.17	0.005**	27%	NUTS-0	'Unit of measure'; 'Time spent (hh:mm)'; 'Mode of transport'; 'Private cars, motorcycles, vans'
Eurostat	TUS_00TRAVEL	0.10	0.118	27%	NUTS-0	'Unit of measure'; 'Time spent (hh:mm)'; 'Mode of transport'; 'Public transport'
Eurostat	URB_LTRAN	0.09	0.158	55%	NUTS-0	'Urban audit indicator'; 'Share of journeys to work by bicycle -%'
Eurostat (normalized)	TOUR_CE_OAN3	0.21	0.001***	4%	NUTS-3	'Tourism indicator'; 'Nights spent'; 'Country of residence'; 'Domestic country'; 'Unit of measure'; 'Number'
Eurostat (normalized)	TOUR_CE_OAN3	0.13	0.027*	4%	NUTS-3	'Tourism indicator'; 'Nights spent'; 'Country of residence'; 'Foreign country'; 'Unit of measure'; 'Number'
Eurostat (normalized)	TOUR_CE_OAN3	0.16	0.008**	4%	NUTS-3	'Tourism indicator'; 'Nights spent'; 'Country of residence'; 'Total'; 'Unit of measure'; 'Number'
Meteostat	Comfortable Temp Share	-0.06	0.360	0%	BSS	
Meteostat	Precipitation Share	-0.10	0.115	0%	BSS	

Table 18 (continued)

Source	Predictor	ρ	p-value	Input	Level	Variables
Nextbike	Bikes Per 1000 Capita	0.11	0.061	0%	BSS	
Nextbike	Bikes Per Sqkm	0.46	0.000***	0%	BSS	
Nextbike	Free Float	0.00	1.000	0%	BSS	
Nextbike	Hybrid	0.15	0.013*	0%	BSS	
Nextbike	Open To Members	0.00	1.000	8%	BSS	
Nextbike	Open To Non Members	0.00	1.000	8%	BSS	
Nextbike	Operating Days	0.00	1.000	8%	BSS	
Nextbike	Operating Time	-0.30	0.000***	8%	BSS	
Nextbike	Propulsion Share	0.20	0.001**	0%	BSS	{'propulsion_type':'electric_assist'}
Nextbike	Station Based	-0.15	0.013*	0%	BSS	
OSM	Amenity Per Sqkm	0.30	0.000***	0%	BSS	
OSM	Bicycle Per Sqkm	0.25	0.000***	0%	BSS	
OSM	Bike Parking N Per Sqkm	0.39	0.000***	0%	BSS	
OSM	Bike Shops Per Sqkm	0.34	0.000***	0%	BSS	
OSM	Cycleway Length Per Sqkm	0.13	0.028*	0%	BSS	
OSM	Embankment Per Sqkm	0.10	0.105	0%	BSS	
OSM	Harbour Per Sqkm	0.08	0.188	0%	BSS	
OSM	Highway Per Sqkm	0.29	0.000***	0%	BSS	
OSM	Historic Per Sqkm	0.11	0.070	0%	BSS	
OSM	Junction Per Sqkm	0.08	0.180	0%	BSS	
OSM	Landuse Area Ratio	0.21	0.000***	0%	BSS	{'landuse':'commercial'}
OSM	Landuse Area Ratio	-0.06	0.311	0%	BSS	{'landuse':'farmland'}
OSM	Landuse Area Ratio	0.02	0.713	0%	BSS	{'landuse':'farmyard'}
OSM	Landuse Area Ratio	0.07	0.282	0%	BSS	{'landuse':'forest'}
OSM	Landuse Area Ratio	0.03	0.581	0%	BSS	{'landuse':'harbour'}
OSM	Landuse Area Ratio	0.37	0.000***	0%	BSS	{'landuse':'industrial'}
OSM	Landuse Area Ratio	0.08	0.162	0%	BSS	{'landuse':'orchard'}
OSM	Landuse Area Ratio	-0.02	0.682	0%	BSS	{'landuse':'quarry'}

Table 18 (continued)

Source	Predictor	ρ	p-value	Input	Level	Variables
OSM	Landuse Area Ratio	0.01	0.930	0%	BSS	{'landuse': 'religious'}
OSM	Landuse Area Ratio	0.23	0.000***	0%	BSS	{'landuse': 'residential'}
OSM	Landuse Area Ratio	0.17	0.004**	0%	BSS	{'landuse': 'retail'}
OSM	Landuse Area Ratio	0.01	0.885	0%	BSS	{'landuse': 'vineyard'}
OSM	Leisure Per Sqkm	0.34	0.000***	0%	BSS	
OSM	Natural Per Sqkm	0.17	0.004**	0%	BSS	
OSM	Office Per Sqkm	0.23	0.000***	0%	BSS	
OSM	Park Area Ratio	0.34	0.000***	0%	BSS	
OSM	Public Transport Per Sqkm	0.35	0.000***	0%	BSS	{'stop_type': 'bus'}
OSM	Public Transport Per Sqkm	0.07	0.283	0%	BSS	{'stop_type': 'ferry'}
OSM	Public Transport Per Sqkm	0.22	0.000***	0%	BSS	{'stop_type': 'other'}
OSM	Public Transport Per Sqkm	0.24	0.000***	0%	BSS	{'stop_type': 'subway'}
OSM	Public Transport Per Sqkm	0.37	0.000***	0%	BSS	{'stop_type': 'total'}
OSM	Public Transport Per Sqkm	0.27	0.000***	0%	BSS	{'stop_type': 'train'}
OSM	Shop Per Sqkm	0.22	0.000***	0%	BSS	
OSM	Toll Per Sqkm	0.08	0.162	0%	BSS	
OSM	Tourism Per Sqkm	0.24	0.000***	0%	BSS	
OSM	Water Per Sqkm	0.03	0.650	0%	BSS	
OSM	Wetland Per Sqkm	-0.06	0.298	0%	BSS	
OSM	Wood Per Sqkm	0.00	1.000	0%	BSS	
Python-holidays	Country Holiday Share	-0.08	0.213	0%	BSS	
SRTM	Elevation Stddev	0.18	0.003**	0%	BSS	
Statista	Age Based Helmet Law	0.04	0.470	0%	NUTS-0	
Statista	Mandatory Helmet Law	-0.01	0.857	0%	NUTS-0	
Wikipedia	EP2019 Election Results	0.10	0.114	0%	NUTS-0	{'faction': 'ECR'}
Wikipedia	EP2019 Election Results	-0.40	0.000***	0%	NUTS-0	{'faction': 'EPP'}
Wikipedia	EP2019 Election Results	0.15	0.012*	0%	NUTS-0	{'faction': 'G/EFA'}

Table 18 (continued)

Source	Predictor	ρ	p-value	Input	Level	Variables
Wikipedia	EP2019 Election Results	0.26	0.000***	0%	NUTS-0	{'faction':'GUE/NGL'}
Wikipedia	EP2019 Election Results	−0.05	0.379	0%	NUTS-0	{'faction':'ID'}
Wikipedia	EP2019 Election Results	0.13	0.038*	0%	NUTS-0	{'faction':'NI'}
Wikipedia	EP2019 Election Results	0.31	0.000***	0%	NUTS-0	{'faction':'RE'}
Wikipedia	EP2019 Election Results	−0.17	0.004**	0%	NUTS-0	{'faction':'S & D'}
Worldpop	Pop Density	0.44	0.000***	0%	BSS	

Source: Data source, Predictor: Name within the data source, ρ : PCC with TDB, p-value: statistical significance of correlation, Input.: Percentage of values missing and imputed with median, Level: Geographic resolution of data, Variables: Specific value used

OLS model without variable standardization

See Table 19

Table 19 Variables of the model that were selected through *bidi-rectional stepwise regression* for the dependent variable $\log(\text{TDB})$

Variable	Coef	<i>p</i> -value	Conf. int
Intercept	−2.009	0.039*	[−3.91, −0.10]
Population: share < 30 (%)	3.342	0.001**	[1.36, 5.32]
Population per km ² (<i>log</i>)	0.227	0.031*	[0.02, 0.43]
Passenger Cars per 1000 inhabitants	0.002	0.014*	[0.00, 0.00]
Sharing econ.: accommodation (<i>annual nights by foreign users, log</i>)	0.158	0.000***	[0.07, 0.24]
Sharing econ.: accommodation (<i>annual nights by domestic users</i>)	−0.151	0.014*	[−0.27, −0.03]
EP elections: ECR (%) <i>log</i>	0.071	0.000***	[0.04, 0.10]
EP elections: ID (%)	0.863	0.017*	[0.15, 1.57]
EP elections: RE (%)	1.938	0.000***	[1.37, 2.51]
Landuse share: forest	1.534	0.019*	[0.25, 2.81]
Landuse share: industrial	2.467	0.005**	[0.73, 4.20]
Landuse share: residential	−0.583	0.030*	[−1.11, −0.06]
Landuse share: parks	−3.069	0.026*	[−5.76, −0.37]
Amenities per km ²	0.001	0.031*	[0.00, 0.00]
Bike shops per km ² (<i>log</i>)	0.030	0.038*	[0.00, 0.06]
BSS: operating time (<i>s</i>)	−0.000	0.003**	[−0.00, −0.00]
BSS: operating area (<i>km</i> ² , <i>log</i>)	0.279	0.000***	[0.15, 0.41]
BSS: bikes per km ² (<i>log</i>)	0.486	0.000***	[0.36, 0.61]
BSS: share electric bikes (%) <i>log</i>)	0.034	0.006**	[0.01, 0.06]

Predictor variables not standardized. All tested variables are listed in subsection 6.3. (Coefficients, *p*-values, and confidence intervals for standardized explanatory variables; confidence intervals for coefficients use a confidence level of 0.95; *denotes significance at 0.05 level, **significance at 0.01 level, and ***significance at 0.001 level)

Acknowledgements We want to thank the OSM contributors for their effort to create open source map data, as well as Prof. Dr.-Ing. Markus Lienkamp and Prof. Dr. Allister Loder for reviewing the paper before submission.

Author contributions **F.W.** Conceptualization, Methodology, Formal analysis, Software, Investigation, Visualization, Writing - Original draft preparation, Writing - Review & Editing, Visualization, Validation, Project administration, **G.B.** Conceptualization, Methodology, Software, Data curation, Writing - Original draft preparation, Writing - Review & Editing, Visualization, **F.R.** Data curation, Formal analysis, Writing - Original draft preparation, **M.L.** Methodology, Formal analysis, Writing - Review & Editing.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability The dataset is available under the following link: <https://syncandshare.lrz.de/getlink/fiSKzE1DtkXwB3AUWFFQgy/dataset.zip>

Declarations

Conflict of interest The authors declare no Conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aboueela, M., Chaniotakis, E., Antoniou, C.: Understanding the landscape of shared-e-scooters in north america; spatiotemporal analysis and policy insights. *Trans. Res. Part A: Policy Pract.* **169**, 103602 (2023). <https://doi.org/10.1016/j.tra.2023.103602>
- Akaike, H.: A new look at the statistical model identification. *IEEE Trans. Autom. Control* **19**(6), 716–723 (1974). <https://doi.org/10.1109/TAC.1974.1100705>
- Alcorn, L.G., Jiao, J.: Bike-sharing station usage and the surrounding built environments in major Texas cities. *J. Plan. Educ. Res.* **43**(1), 122–135 (2023). <https://doi.org/10.1177/0739456X19862854>
- Arias-Molinares, D., Xu, Y., Büttner, B., et al.: Exploring key spatial determinants for mobility hub placement based on micromobility ridership. *J. Transp. Geogr.* **110**, 103621 (2023). <https://doi.org/10.1016/j.jtrangeo.2023.103621>
- Bean, R., Pojani, D., Corcoran, J.: How does weather affect bikeshare use? a comparative analysis of forty cities across climate zones. *J. Transp. Geogr.* **95**, 103155 (2021). <https://doi.org/10.1016/j.jtrangeo.2021.103155>
- Bergstra, J., Bengio, Y.: Random search for hyper-parameter optimization. *J. Mach. Learn. Res.* **13**(10), 281–305 (2012)
- Bettencourt, L.M.A., Lobo, J., Strumsky, D., et al.: Urban scaling and its deviations: Revealing the structure of wealth, innovation and crime across cities. *PLoS ONE* **5**(11), 1–9 (2010). <https://doi.org/10.1371/journal.pone.0013541> (<https://doi.org/10.1371/journal.pone.0013541>)
- Bielinski, T., Kwapisz, A., Wazna, A.: Bike-Sharing Systems in Poland. *Sustainability* **11**(9), 2458 (2019). <https://doi.org/10.3390/su11092458>
- Bondarenko, M., Kerr, D., Sorichetta, A., et al.: Census/projection-disaggregated gridded population datasets, adjusted to match the corresponding UNPD 2020 estimates, for 183 countries in 2020 using Built-Settlement Growth Model (BSGM) outputs. <https://doi.org/10.5258/SOTON/WP00685> (2020)
- Bradley, A.P.: The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern Recogn.* **30**(7), 1145–1159 (1997). [https://doi.org/10.1016/S0031-3203\(96\)00142-2](https://doi.org/10.1016/S0031-3203(96)00142-2)
- Breiman, L.: Random forests. *Mach. Learn.* **45**(1), 5–32 (2001). <https://doi.org/10.1023/A:1010933404324>
- Buitinck, L., Louppe, G., Blondel, M., et al.: API design for machine learning software: experiences from the scikit-learn project. In: *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, pp 108–122 (2013)
- Caggiani, L., Camporeale, R., Hamidi, Z., et al.: Evaluating the efficiency of bike-sharing stations with data envelopment analysis. *Sustainability* **13**(2), 881 (2021). <https://doi.org/10.3390/su13020881>
- Chen, E., Ye, Z.: Identifying the nonlinear relationship between free-floating bike sharing usage and built environment. *J. Clean. Prod.* **280**, 124281 (2021). <https://doi.org/10.1016/j.jclepro.2020.124281>
- Cycling Industries Europe: Shared ambition: The potential for bike sharing in europe: Benchmarking 148 cities. https://cyclingindustries.com/fileadmin/CIE_Shared_Ambition%E2%80%9393_Benchmarking_Bike_Sharing_in_148_cities.pdf (2023)

- Directorate-General for Communication, Tsoulou Malakoudi, D., Büttner, M., et al.: Review of European and National Election Results (Update: September 2019). Publications Office of the European Union, (2019) <https://data.europa.eu/doi/10.2861/326863>
- Du, M., Cheng, L.: Better understanding the characteristics and influential factors of different travel patterns in free-floating bike sharing: Evidence from nanjing, china. *Sustainability* **10**(4), 1244 (2018). <https://doi.org/10.3390/su10041244>
- Duran-Rodas, D., Chaniotakis, E., Antoniou, C.: Built environment factors affecting bike sharing ridership: Data-driven approach for multiple cities. *Transp. Res. Rec.: J. Trans. Res. Board* **2673**(12), 55–68 (2019). <https://doi.org/10.1177/0361198119849908>
- Earth Resources Observation And Science (EROS) Center: Shuttle Radar Topography Mission (SRTM) 1 Arc-Second Global. <https://doi.org/10.5066/F7PR7TFT> (2011)
- El-Assi, W., Salah Mahmoud, M., Nurul Habib, K.: 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 (2017). <https://doi.org/10.1007/s11116-015-9669-z>
- Eren, E., Uz, V.E.: A review on bike-sharing: The factors affecting bike-sharing demand. *Sustain. Cities Soc.* **54**, 101882 (2020). <https://doi.org/10.1016/j.scs.2019.101882>
- Faghih-Imani, A., Eluru, N.: Analysing bicycle-sharing system user destination choice preferences: Chicago's divvy system. *J. Transp. Geogr.* **44**, 53–64 (2015). <https://doi.org/10.1016/j.jtrangeo.2015.03.005>
- Faghih-Imani, A., Eluru, N., El-Geneidy, A.M., et al.: How land-use and urban form impact bicycle flows: evidence from the bicycle-sharing system (bixi) in montreal. *J. Transp. Geogr.* **41**, 306–314 (2014). <https://doi.org/10.1016/j.jtrangeo.2014.01.013>
- Faghih-Imani, A., Hampshire, R., Marla, L., et al.: An empirical analysis of bike sharing usage and rebalancing: Evidence from barcelona and seville. *Trans. Res. Part A: Policy Pract.* **97**, 177–191 (2017). <https://doi.org/10.1016/j.tra.2016.12.007>
- Fawcett, T.: An introduction to roc analysis. *Pattern Recogn. Lett.* **27**(8), 861–874 (2006). <https://doi.org/10.1016/j.patrec.2005.10.010>
- Fishman, E.: Bikeshare: A review of recent literature. *Transp. Rev.* **36**(1), 92–113 (2016). <https://doi.org/10.1080/01441647.2015.1033036>
- Guo, Y., He, S.Y.: Built environment effects on the integration of dockless bike-sharing and the metro. *Transp. Res. Part D: Transp. Environ.* **83**, 102335 (2020). <https://doi.org/10.1016/j.trd.2020.102335>
- Hocking, R.R.: The Analysis and Selection of Variables in Linear Regression. *Biometrics* **32**(1), 1–49 (1976). <https://doi.org/10.2307/2529336>. (2529336)
- Hosseini, K., Stefaniec, A., O'Mahony, M., et al.: Optimising shared electric mobility hubs: Insights from performance analysis and factors influencing riding demand. *Case Stud. Trans. Policy* **13**, 101052 (2023). <https://doi.org/10.1016/j.cstp.2023.101052>
- International Transport Forum: Micromobility, equity and sustainability: Summary and conclusions, itf roundtable reports, no. 185. <https://www.itf-oecd.org/sites/default/files/docs/micromobility-equity-sustainability.pdf> (2021)
- Jaber, A., Csonka, B.: Investigating the temporal differences among bike-sharing users through comparative analysis based on count, time series, and data mining models. *Alex. Eng. J.* **77**, 1–13 (2023). <https://doi.org/10.1016/j.aej.2023.06.087>
- Julio, R., Monzon, A.: Long term assessment of a successful e-bike-sharing system. Key drivers and impact on travel behaviour. *Case Stud. Trans. Policy* **10**(2), 1299–1313 (2022). <https://doi.org/10.1016/j.cstp.2022.04.019>
- Kim, K.: Investigation on the effects of weather and calendar events on bike-sharing according to the trip patterns of bike rentals of stations. *J. Transp. Geogr.* **66**, 309–320 (2018). <https://doi.org/10.1016/j.jtrangeo.2018.01.001>
- Kuo, Y.W., Hsieh, C.H., Hung, Y.C.: Non-linear characteristics in switching intention to use a docked bike-sharing system. *Transportation* **48**(3), 1459–1479 (2021). <https://doi.org/10.1007/s11116-020-10102-2>
- Lamprecht, C., Franco, R., Maison, J., et al.: Meteostat Developers. <https://dev.meteostat.net/> (2023)
- Li, A., Zhao, P., Liu, X., et al.: Comprehensive comparison of e-scooter sharing mobility: Evidence from 30 european cities. *Transp. Res. Part D: Transp. Environ.* **105**, 103229 (2022). <https://doi.org/10.1016/j.tra.2022.103229>
- Mahajan, S., Argota Sánchez-Vaquerizo, J.: Global comparison of urban bike-sharing accessibility across 40 cities. *Sci. Rep.* **14**(1), 20493 (2024). <https://doi.org/10.1038/s41598-024-70706-x>. (<https://doi.org/10.1038/s41598-024-70706-x>)
- Marquardt, D.W.: Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation. *Technometrics* **12**(3), 591–612 (1970). <https://doi.org/10.1080/00401706.1970.10488699>
- Médard de Chardon, C., Caruso, G., Thomas, I.: Bicycle sharing system 'success' determinants. *Trans. Res. Part A: Policy Pract.* **100**, 202–214 (2017). <https://doi.org/10.1016/j.tra.2017.04.020>

- MobilityData: General bikeshare feed specification: Documentation for the general bikeshare feed specification, a standardized data feed for shared mobility system availability. <https://github.com/MobilityData/gbfs> (2024)
- nextbike by TIER (2023) Investor relations. <https://www.nextbike.net/en/investor-relations/>
- nextbike by TIER: official nextbike api documentation. (2024) <https://github.com/nextbike/api-doc>
- Nogal, M., Jiménez, P.: Attractiveness of bike-sharing stations from a multi-modal perspective: The role of objective and subjective features. *Sustainability* **12**(21), 9062 (2020). <https://doi.org/10.3390/su12219062>
- O'Brien, O., Cheshire, J., Batty, M.: Mining bicycle sharing data for generating insights into sustainable transport systems. *J. Transp. Geogr.* **34**, 262–273 (2014). <https://doi.org/10.1016/j.jtrangeo.2013.06.007>
- Otero, I., Nieuwenhuijsen, M.J., Rojas-Rueda, D.: Health impacts of bike sharing systems in europe. *Environ. Int.* **115**, 387–394 (2018). <https://doi.org/10.1016/j.envint.2018.04.014>
- Pimentel de Vassimon, P.: Performance evaluation for bike-sharing systems: A benchmarking amongst 50 cities. <https://repositorio.usp.br/directbitstream/9f651f36-07c5-4354-a041-77739573c9b4/PedroPimenteldeVassimon%20TCCPRO15.pdf> (2015)
- Reck, D.J., Axhausen, K.W.: Who uses shared micro-mobility services? empirical evidence from zurich, switzerland. *Transp. Res. Part D: Transp. Environ.* **94**, 102803 (2021). <https://doi.org/10.1016/j.trd.2021.102803>. (<https://www.sciencedirect.com/science/article/pii/S1361920921001073>)
- Reck, D.J., Martin, H., Axhausen, K.W.: Mode choice, substitution patterns and environmental impacts of shared and personal micro-mobility. *Transp. Res. Part D: Transp. Environ.* **102**, 103134 (2022). <https://doi.org/10.1016/j.trd.2021.103134>. (<https://www.sciencedirect.com/science/article/pii/S1361920921004296>)
- Schraff, D., Vergioglou, I., Demirci, B.B.: The European NUTS-level election dataset: A tool to map European electoral geography. *Party Politics* **29**(3), 570–579 (2023). <https://doi.org/10.1177/13540688221083553>
- Schröder, D., Gotzler, D.: Comprehensive spatial and cost assessment of urban transport options in munich. *J. Urb. Mobil.* **1**, 100007 (2021). <https://doi.org/10.1016/j.urbmob.2021.100007>
- Seyerlein, C.: Tier will die lebensdauer von e-scooter- und e-bike-akkus verlängern. <https://www.next-mobility.de/tier-will-die-lebensdauer-von-e-scooter-und-e-bike-akkus-verlaengern-a-1098527/> (2022)
- Sheather, S.: Chapter 6.4.1 - Multicollinearity and Variance Inflation Factors. In: *A Modern Approach to Regression with R*. Springer Texts in Statistics, Springer, <https://doi.org/10.1007/978-0-387-09608-7> (2009)
- Singh, Y.J.: Is smart mobility also gender-smart? *J. Gend. Stud.* **29**(7), 832–846 (2020). <https://doi.org/10.1080/09589236.2019.1650728>
- Smith, G.: Step away from stepwise. *J. Big Data* **5**(1), 32 (2018). <https://doi.org/10.1186/s40537-018-0143-6>
- Soriguera, F., Jiménez-Meroño, E.: A continuous approximation model for the optimal design of public bike-sharing systems. *Sustain. Cities Soc.* **52**, 101826 (2020). <https://doi.org/10.1016/j.scs.2019.101826>
- Sun, Y., Mobasheri, A., Hu, X., et al.: Investigating impacts of environmental factors on the cycling behavior of bicycle-sharing users. *Sustainability* **9**(6), 1060 (2017). <https://doi.org/10.3390/su9061060>
- Todd, J., O'Brien, O., Cheshire, J.: A global comparison of bicycle sharing systems. *J. Transp. Geogr.* **94**, 103119 (2021). <https://doi.org/10.1016/j.jtrangeo.2021.103119>
- Tran, T.D., Ovtracht, N., d'Arcier, B.F.: Modeling bike sharing system using built environment factors. *Procedia CIRP* **30**, 293–298 (2015). <https://doi.org/10.1016/j.procir.2015.02.156>
- Wang, Y., Wu, J., Chen, K., et al.: Are shared electric scooters energy efficient? *Commun. Trans. Res.* **1**, 100022 (2021). <https://doi.org/10.1016/j.commtr.2021.100022>
- Whittingham, M.J., Stephens, P.A., Bradbury, R.B., et al.: Why do we still use stepwise modelling in ecology and behaviour? *J. Anim. Ecol.* **75**(5), 1182–1189 (2006). <https://doi.org/10.1111/j.1365-2656.2006.01141.x>
- Willberg, E., Salonen, M., Toivonen, T.: What do trip data reveal about bike-sharing system users? *J. Transp. Geogr.* **91**, 102971 (2021). <https://doi.org/10.1016/j.jtrangeo.2021.102971>
- Xu, S.J., Chow, J.Y.J.: A longitudinal study of bike infrastructure impact on bikesharing system performance in new york city. *Int. J. Sustain. Transp.* **14**(11), 886–902 (2020). <https://doi.org/10.1080/15568318.2019.1645921>
- Zhao, J., Deng, W., Song, Y.: Ridership and effectiveness of bikesharing: The effects of urban features and system characteristics on daily use and turnover rate of public bikes in china. *Transp. Policy* **35**, 253–264 (2014). <https://doi.org/10.1016/j.tranpol.2014.06.008>
- Zhou, J., Jing, C., Hong, X., et al.: Winter Sabotage: The Three-Way Interactive Effect of Gender, Age, and Season on Public Bikesharing Usage. *Sustainability* **11**(11), 3217 (2019). <https://doi.org/10.3390/su11113217>
- Zhou, J., Guo, Y., Sun, J., et al.: Review of bike-sharing system studies using bibliometrics method. *J. Traffic Trans. Eng. (English Edition)* **9**(4), 608–630 (2022). <https://doi.org/10.1016/j.jtte.2021.08.003>

Zhou, J., Li, Z., Dong, S., et al.: Visualization and bibliometric analysis of e-bike studies: A systematic literature review (1976–2023). *Transp. Res. Part D: Transp. Environ.* **122**, 103891 (2023). <https://doi.org/10.1016/j.trd.2023.103891>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Felix Waldner holds a Master's degree in mechanical engineering and is a research associate at the Chair of Automotive Technology at the Technical University of Munich (TUM). His research focuses on smart and shared mobility, mobility data analysis and modeling, as well as the cost and pricing of mobility.

Georg Balke M.Sc. leads the research group Smart Mobility at the Institute of Automotive Technology (TU Munich). His major research interest is the application of mobility data for transportation system design.

Felix Rech holds a Master's degree in Computer Science from Technical University of Munich (TUM) and is a data scientist specializing on supply chain and logistics consultancy. His main interests include statistical and machine learning techniques across various domains.

Martin Lellep is a research assistant at the University of Edinburgh (School of Physics and Astronomy) and a Senior Staff Machine Learning Engineer doing AI security research at Palo Alto Networks. Beyond his interest in quantitative transport, he has a research background in computational fluid dynamics and machine learning, where he applies advanced numerical methods and AI to large-scale physical and other systems.

Authors and Affiliations

Felix Waldner¹  · Georg Balke¹  · Felix Rech² · Martin Lellep³ 

✉ Felix Waldner
f.waldner@tum.de

¹ Technical University of Munich, School of Engineering and Design, Institute of Automotive Technology, Boltzmannstraße 15, 85748 Garching b. München, Germany

² Technical University of Munich, School of Computation, Information and Technology, Boltzmannstraße 15, 85748 Garching b. München, Germany

³ The University of Edinburgh, School of Physics and Astronomy, Peter Guthrie Tait Road, Edinburgh EH9 3FD, UK