Travel Behaviour and Society 34 (2024) 100705

Contents lists available at ScienceDirect

Travel Behaviour and Society

journal homepage: www.elsevier.com/locate/tbs

Unveiling the roles of public bike systems: From leisure to multimodal transportation

Xuan Li [a], Jaehyun Ha [b], Sugie Lee [a,*]

[a] Dept. of Urban Planning and Engineering, Hanyang University, Seoul, Korea
[b] Price School of Public Policy, University of Southern California, CA, United States

ARTICLE INFO

Keywords:
Bike sharing
Modal shift
Shared mobility
DBSCAN

ABSTRACT

As bike-sharing systems gain popularity, understanding their roles in urban transportation, such as the first/last mile problem and the replacement of public transit, is crucial to comprehensively promote urban transport systems. Due to the Internet of Things (IoT) technique, copious data from public bike trips worldwide are available to urban researchers. In this study, we examine how public bikes function not just as a means of transportation, but also for leisure. The leisure function was ignored in previous empirical studies using data mining techniques, an oversight that could skew results in places like Seoul where leisure trips constitute a significant portion of the trips. We investigate when, where, and to what degree bike-sharing serves as a leisure tool and substitutes for, integrates with, or complements the public transit system. We propose an integrated data-mining approach to identify the above four categories of public bike trips. Our method combines multiple sources of real-world data, the distributed density-based spatial clustering of applications with noise (DBSCAN), and spatial feature engineering. The results reveal that the distribution of each category shows substantial spatial–temporal heterogeneity. Longitudinal analysis reveals that the use of public bicycles as a transportation mode primarily contributes to the growth observed from 2018 to 2021, compared to use for leisure. The determining factors and non-linear functions of each type of bike trip were interpreted by explainable machine learning methods, Shapley additive explanations (SHAP), and partial dependence plot (PDP). We discovered that lower slopes, higher residential densities, and more commercial neighborhoods consistently attract users across all categories. Additionally, rivers and green areas attract more leisure trips, while areas with denser transit routes see more substitution trips and fewer complementary trips. By understanding the characteristics, spatial–temporal distribution, and determining factors of public bike trips across these four categories, city planners and operators can tailor services to meet the diverse needs of citizens, fostering a more convenient, sustainable, and accessible urban transport system.

1. Introduction

Internet of Things (IoT) techniques for analysis of big data have improved safety and management of public bike systems, which have grown in prevalence and popularity in more than 800 cities around the world since the turn of the 21st century (DeMaio, 2009; Li et al., 2020; Shaheen et al., 2010). Citizens have embraced bicycles as an easy-to-use and flexible transportation mode in their daily lives (Pucher & Buehler, 2007; Fraser & Lock, 2011; Fishman et al., 2013; Qiu & He, 2018). Bike sharing has advantages over other modes of transportation in terms of flexibility and physical fitness (Fraser & Lock, 2011; Qiu & He, 2018). Based on the broader perspective of energy saving and carbon emission reduction, city governments have taken the lead in promoting bike sharing (Lumsdon & Tolley, 2001; Zhang et al., 2015).

Therefore, the potential for integrating public bike-share systems with mass transit has attracted renewed attention from researchers, planners, and policymakers. Understanding the interactions between bike sharing and the rest of the transportation system is the key to systematically improving the urban transportation system as a whole. As a mode of transportation in cities, bike sharing can compete (substitute), complement, or integrate with other modes, from public transit to private cars (Kong et al., 2020). Despite their competitive edge being limited to shorter distances, bike sharing can still compete with or complement other modes because of cost, accessibility, frequency, and

* Corresponding author.
E-mail addresses: lixuan@hanyang.ac.kr (X. Li), jaehyunh@usc.edu (J. Ha), sugielee@hanyang.ac.kr (S. Lee).

https://doi.org/10.1016/j.tbs.2023.100705
Received 21 January 2023; Received in revised form 12 September 2023; Accepted 24 October 2023
Available online 16 November 2023
2214-367X/

comfort (Kapuku et al., 2022). In areas that transit services cannot cover well, bike sharing can complement or be integrated with the transit system as a mode of public transportation (Liu et al., 2012; Yang et al., 2019). In particular, more flexible dockless bike-sharing systems are regarded as a decent solution to the first- and last-mile problem in Chinese Megacities (Liu et al., 2012; Fan et al., 2019; Guo et al., 2021). However, the new generation of flexible dockless bike systems also bring problems to city management because of oversupply and parking chaos (Tu et al., 2019).

Even though some studies have examined the interaction between bikes and transit, the Seoul case is valuable. First, the docked bike-sharing system in Seoul is still developing. This docked system combines the advantages, such as flexibility and infectiveness, of the Mobility as a service (MaaS) bike systems and avoids problems caused by the mismanagement of dockless bikes. The method used in this study of Seoul offers a framework for understanding other cities where the government hopes to develop better docked bike-sharing systems. Second, Seoul already has a well-running, highly accessible transit system that is one of the biggest in the world (Lee et al., 2008). Our research will evaluate the bike-transit interaction under these conditions, which can be a good reference for other megacities with developed transit systems. Third, the large number of leisure trips made by bike in Seoul make it a unique case and a good sample for boosting citizens' physical activity for health benefits and tourism development (Lee and Noland, 2021).

In this paper, we reveal when and where bike-share trips serve as leisure tools and substitute for, integrate with, and complement public transit. To do so, we designed a data-mining approach with machine learning clustering to categorize each individual bike trip and utilized data including trip length, the shortest route, and the spatial relationships between docks and the transit stations. Statistical data from the Korea National Travel Survey were used to validate our results. After displaying temporal and spatial distributions of each type of bike trip, we explored the influence of built and natural environments on each type of bike trips using explainable machine learning.

2. Literature review

2.1. Bike sharing scheme and analysis

The evolution of bike-sharing systems has passed through four generations (Ricci, 2015). The combination of Bike Share Schemes (BSS) and transit systems has been shown to reduce carbon emissions (Zhou et al., 2023). The emergence and prevalence of bike sharing have reconstructed urban mobility via shifts from other transit modes and mutual promotion, which has attracted the interest of researchers. From the perspective of planners, bike-sharing plays a mixed role in cities. It can act as a dependent mode of transportation (Castillo-Manzano et al., 2016), be a feeder mode of metros (Guo et al., 2021), or a leisure tool to promote physical activity (Chen & Chancellor, 2020) with each function requiring different management and services for citizens to have better experiences.

The purposes of general transportation modes, such as flights, subway, bus, and taxi, were defined as home, work, dining, shopping, recreation, schooling, and life service (Habib et al., 2014). However, as an active mode of transportation, the purpose of cycling includes "leisure." In some studies it is labeled "physical activity" or "workout" (Buck et al., 2013; Habib et al., 2014; Chen & Chancellor, 2020).

Many individuals use bike sharing for leisure purposes(Habib et al., 2014; Chen & Chancellor, 2020), although BSSs such as in Washington, DC were originally designed as a form of public transportation (Gao & Lee, 2019). In Dublin, Ireland, for example, 48.3 % of users ride bikes only for leisure during off-peak hours (Murphy & Usher, 2015). In Brisbane, Australia, 65 % of unsubscribed users stated that their last bike-sharing trip was for leisure or sightseeing (Fishman, 2016). A recent survey about public bikes in Seoul suggested that 24.4 % of subscribers used bike-sharing mainly for leisure on weekdays, while 51 % of weekend users were leisure riders (Park et al., 2023).

There are two commonly accepted approaches to inferring travel purpose in the existing literature on bike-sharing. One is conventional surveys (Fishman et al., 2013), and the other is inferring trip purposes using data mining (Bordagaray et al., 2016). Surveys require considerable time and effort and carry risks for potential errors and biases associated with small samples. Data mining methods utilize the distribution of Point of interest (POI) by type to infer travel purpose and differ in how they represent data, from learning embedding to Bayes probability estimation regarding the distance decay function to every POI (Yan et al., 2017; Li et al., 2021; Ross-Perez et al., 2022). However, most empirical studies using data mining techniques ignore or oversimplify leisure trips as a mode of physical activity (Ross-Perez et al., 2022; Shen et al., 2018). For instance, Lee and Noland (2021) identified trips that start from parks and amusement parks as "bike leisure." However, a user may rent a bike to travel from where they live and take a detour for leisure before returning to the dock where the bike was originally rented. Therefore, some researchers identify "'loop trips" in which bikes are rented from and returned to the same dock as "leisure trips" (Lee and Noland, 2021). This would leave out leisure trips where bikes are rented and returned at closely located docks, but not the exact same one.

To address these shortcomings, some researchers have applied a clustering method taking into account real travel distance, shortest distance, and roaming distance of shared bikes to detect leisure trips. This method has been used to successfully detect leisure hotspots along the Han River (Lee et al., 2021a), but the results of this study indicate an abnormal morning peak of leisure trips. We hypothesize that this result is a systematic false positive caused by the relatively naive clustering method, K-means, that was used.

2.2. Interactions between bike-sharing and other modes of transportation

Various interaction effects between bike sharing and pre-existing transportation networks, integration (connection), substitution (competition), and complementation, have been discussed in the literature (Wan et al., 2018; Liu et al., 2012; Romm et al., 2022).

Substitution reflects mode shifting from other public mobility services and is frequently addressed in research and policies. Users are generally willing to make the switch to bike sharing, especially for short trips (Politis et al., 2020). Bike sharing was found to reduce the use of cars, ride-sourcing, rail, subway, and buses (Campbell and Brakewood, 2017; Fishman, 2016; Fuller et al., 2013), which can be considered generalized substitution effects that occur between bike sharing and other transit modes.

Integration occurs when bikes facilitate first- and last-mile connections to and from public transit networks or reduce the number of transfers for public transit riders. Specifically, bike-sharing systems can improve the connectivity of the train, metro, bus rapid transit, and light rail transit in megacities (Shelat et al., 2018; van Mil et al., 2021). As a footnote, because of their high availability and similar transport niches in terms of speed and suitable trip distance, the combination of buses and bikes is generally negligible (Guo et al., 2021; Xue & He, 2009).

Complementation refers to offering a convenient and flexible bike-share service where public transit is unavailable or has poor coverage. Bike sharing fills a blank in those areas, especially dockless bike systems, which provide high flexibility and availability (Li et al., 2021). However, few quantitative papers have considered the complementary function of bike sharing because it has not been a hot political concern. Considering the equity and suitability goals of BSSs (Jin et al., 2019), the complementary function deserves to be valued and measured properly.

The multifaceted role of bike sharing within urban transportation systems has been examined. Research increasingly recognizes that bike-sharing can encompass more than a single type of interaction (Martin & Xu, 2022). Kong et al. (2020) first systematically provided a theoretical framework and empirical evidence to consider the three types of bike-transit interactions simultaneously, followed by other case studies using the same methodology (Wu et al., 2022).

There is abundant evidence that several environmental attributes (bike infrastructure, built environment, land use, and urban form, and public transportation) have significant impacts on bike share usage. Generally, higher residence density, restaurants, retail POIs, and better bike station networks are related to higher public bike demand (Guo et al., 2022). The effects also vary across trip purposes (Faghih-Imani et al., 2014; Guo et al., 2022). For instance, proximity to transit stations is related to greater numbers of users for transportation purposes, but not leisure (Lee et al., 2021b). These factors also make a difference in transportation trips. The presence of high-quality bus service in metro catchment areas increases competition between bike share and buses and reduces the integration between bikes and subways (Guo & He, 2020), and vice versa. Martin and Shaheen (2014) found that bike shares served prominently as first/last-mile facilitators in areas with low–intensity transit networks and replaced many short transit trips in high-density, transit-rich areas.

2.3. Research gaps

This study fills three gaps in the existing literature. First, the use of BSSs for leisure trips has been included in surveys and some big data analyses, but no empirical studies of bike-transit relationships (Substitution, Integration, and Complementation) have processed leisure trips in advance, which could lead to inaccurate conclusions in further analyses, especially when and where the study area has large proportions of users who tend to use public bikes only for physical activity rather than as transportation. We suggest that it is necessary to separate bike trips for leisure purposes (as shown in Fig. 1), because such trips have significant differences in statistical features and policy applications from trips for transportation purposes. Second, existing classifications and descriptions of leisure trips have flaws related to small samples and inadequate methodology. In this study, we used a methodology that can accurately infer leisure trips, and validated our results using surveys and data features. This will help to reveal and track the real roles of public bikes in cities. Third, in terms of multiple bike–transit interactions, few studies have made longitudinal comparisons or discussed the nonlinear effects of the environment on the bike–transit relationship. Our analysis in this article provides a more comprehensive and detailed discussion of these gaps by labeling bike trip data and analyzing patterns and determining factors with a fusion machine learning approach.

3. Data and methodology

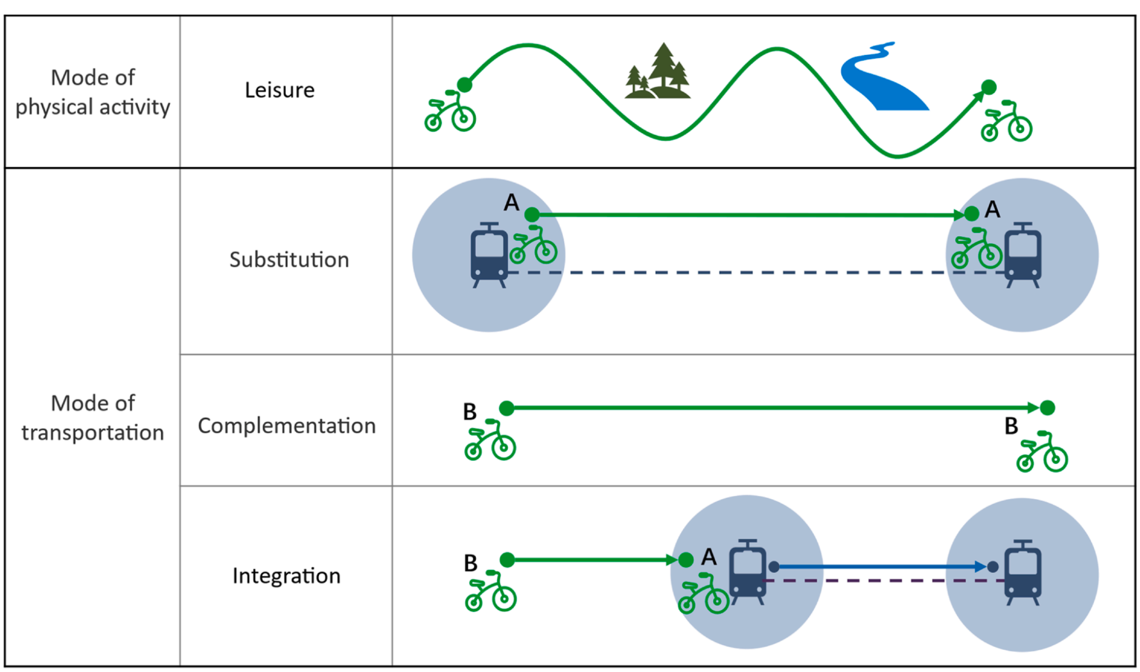
3.1. Study area

The Seoul Metropolitan Government has promoted a bike-sharing service called Seoul Bike (Tareungyi in Korean) since October 2015. As of 2021, more than 2,500 rental docks have been distributed throughout Seoul to offer bike-sharing rental and return services at fixed locations. Users are able to return bikes even if the racks are full, as long as they are within a virtual boundary, which overcomes one common problem of conventional docked public bike systems. Even during the COVID-19 outbreak, BSS ridership maintained strong growth, as Appendix Fig. 1. The average daily bike-sharing use in 2021 was 1.2 times higher than in 2020 (Kim, 2021; Ku et al., 2021; Lee et al., 2021c). Beyond the BSS, Seoul has fast and convenient train, subway, and bus systems that play different roles in the city for trips of different lengths. The study area and spatial distribution of the relevant systems are displayed in Fig. 2. As of 2019, Seoul's public transportation share reached 65 % of all modes (Well Intergrated Transport, 2019). At the time of data collection in December 2021, there was a higher density of bus stations compared to public bike docks, with a total of 11,592 bus stations and 2,519 bike docks. More than 95 % of docks were located within 200 m of subway stations (calculated using spatial data).

3.2. Data sources

3.2.1. Bike-sharing data

Bike-share ridership data were provided by the Seoul Open Data Platform and consist of name, ID, latitude and longitude of origin and destination bike-share stations, and timestamp of borrow and return. Because the raw data contained redundant information and errors, we ran the following preprocessing steps. First, we retained only data from



* **A** stands for public bike docks which are located in the 200-meter buffer of transit stations (including bus and subway)
B stands for docks which are not in any 200-meter buffer of transit stations

Fig. 1. Conceptual diagram of bike trips for Leisure, Substitution, Complementation, and Integration.

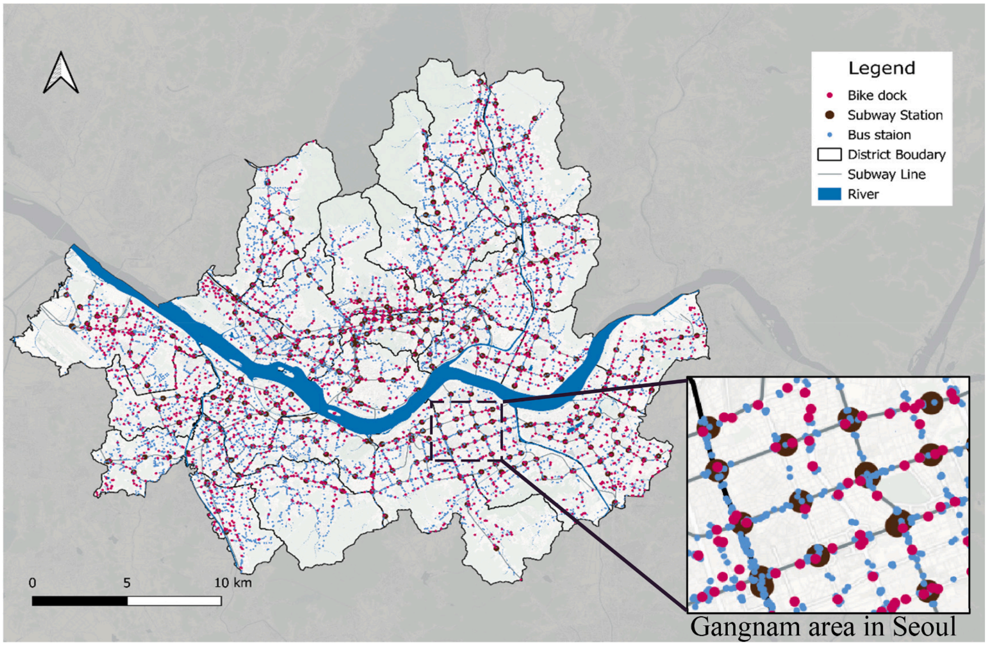


Fig. 2. Study area.

non-rainy days when the average temperature was between 15 and 25 degrees Celsius, because weather can cause major variations in bike-sharing use and affect demand for active transportation (Miranda-Moreno & Nosal, 2011; Wei & Liu, 2022). The distribution of sample days and general trends are displayed in Appendix Fig. 1 and Appendix Table 1. We analyzed data for a total of 73 days. Second, some abnormal records show rentals for 0-km distance, durations shorter than 1 min, longer than 6 h, shorter distance than the straight-line distance to the return dock, or speeds greater than 50 km/hour due to the resolution and sensitivity of the data-collecting devices or GPS error. We cleaned those outliers, as is customary (Shen et al., 2018). After applying the preprocessing filters, we retained data for 5,060,576 trips that involved 2519 bike-rental docks in Seoul, comprising 72.7 % of the original raw data, which contained 6,960,815 trips.

3.2.2. Public transit data

The public transit data reflect 11,592 bus stations, 1193 bus routes, and 738 subway stations in Seoul in 2021. Stations were set as nodes, and routes were set as edges. We made a 2-layer transit network using two modes of transportation, subway and bus. The spatial proximity analysis was based on the spatial relationships between bike docks and transit stations, and the transfer analysis was based on the network relationships among the layers by transit route.

3.2.3. Supporting data

We used multiple temporal and spatial data sources to support our analysis in two dimensions. Temporal data (weather data and holidays) were used to control for variation and were acquired from the Korean National Meteorological Agency and a Python package named Holiday, respectively. Spatial data were used to explain the social and natural environments. For social environment data, we collected employee density, points of interest (POIs), and household information from an open dataset that can be downloaded from Seoul Open Data Plaza or Kakao maps. For natural environment data, we collected information about river areas, green areas, and road slopes in Seoul. The slopes were estimated using a digital elevation model (90 m $\times$ 90 m) provided by the Korea National Spatial Data Infrastructure Portal.

3.2.4. Validation data

National Travel Survey data (The Korea Transport Institute, 2021) were used to validate our classification. The Korean National Travel Survey is conducted every-five years. The one closest to our chosen time range was in 2021. The National Travel Survey is one of the most reliable surveys on travel behavior in Korea at the national level (Ha et al., 2020; Sung et al., 2015), making it suitable to validate our results. This survey investigated the distribution of bike trip durations when people use bikes for transportation (excluding leisure trips).

3.3. Methodology

We designed an integrated approach to decipher the roles of bike sharing in Seoul. The overall structure of this approach is illustrated in Fig. 3, and the details are elaborated below.

3.3.1. Identifying trip purposes through distributed DBSCAN analysis

Bikes are regarded primarily as either a mode of transportation to go visiting, shopping, to work, or as a tool for recreation (Xing et al., 2010). The Seoul bike-sharing system has many users who do not use public bicycles as a mode of transport but instead for recreation and fitness. The most popular routes for leisure occurred along the Han River, and a large amount of them had the same rent and return location (Lee and Noland, 2021), which is strong evidence to suggest that many bike-sharing trips do not interact with public transit at all but rather meet a distinctive need (recreation and leisure) that the transit system cannot fulfill. Therefore, leisure trips should be distinguished by purpose before we identify the relationship between bike trips and the transit system as substitution, integration, or complementation.

Travelers who use bikes as tools for transportation tend to choose the shortest possible routes (Aultman-Hall et al., 1997). Hills and areas with high traffic volumes (Broach et al., 2012) could impede travelers from choosing the shortest path, so the detour rate for transportation trips is near one rather than always equal to one. In addition, given the advantageous range (Curran, 2008) of bikes as a transportation tool, the trip distance and trip duration are typically short.

It would thus be inappropriate to set a rigid cutoff for distinguishing trips for transportation and leisure by trip distance or duration. Instead, we considered unsupervised clustering models for three normalized indicators: trip distance, trip duration, and detour rate. Among these, the density-based spatial clustering of applications with noise (DBSCAN) model is sensitive to centralized and decentralized points and performs

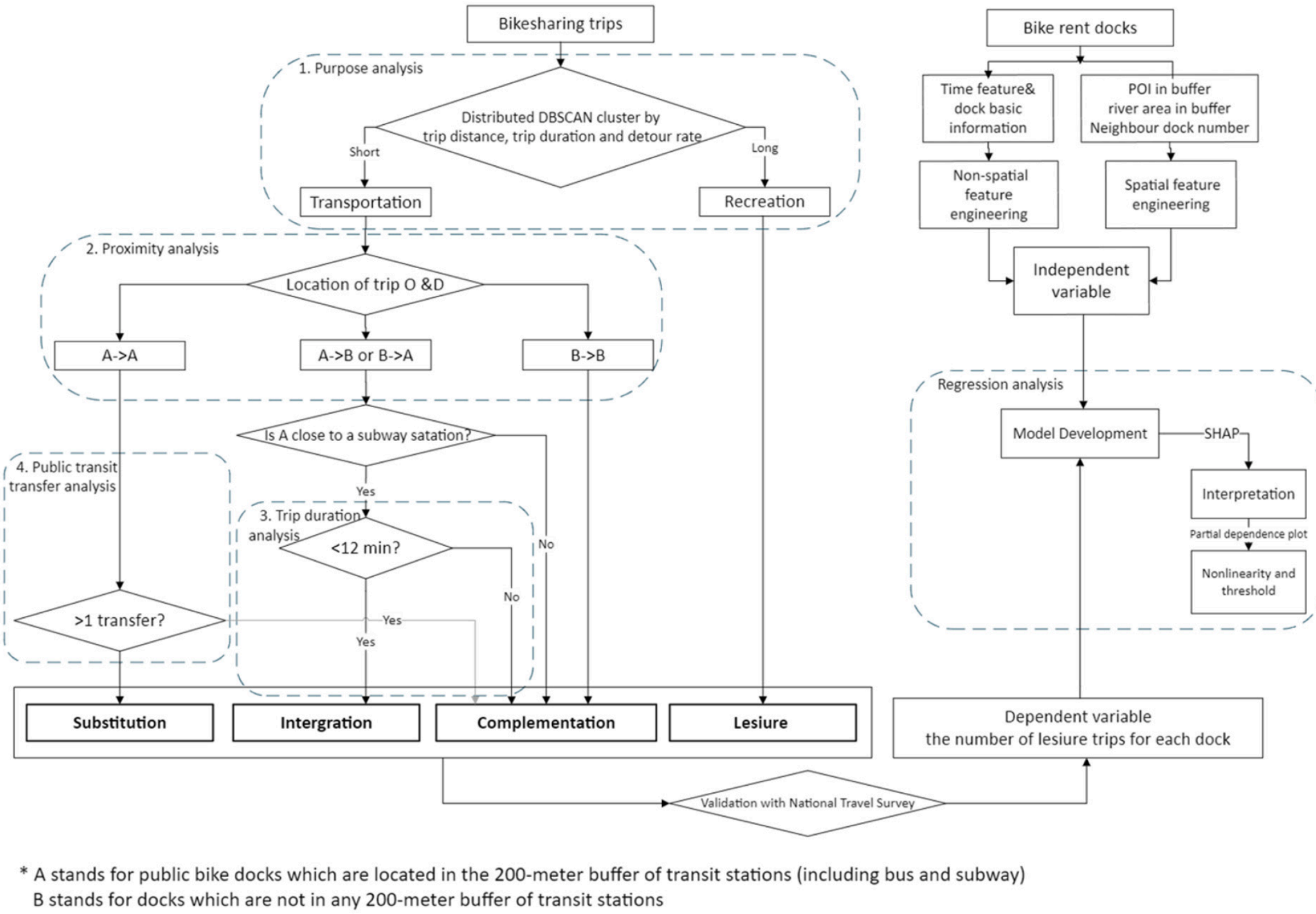


Fig. 3. Research framework.

better than other clustering models (Ester et al., 1996; Schubert et al., 2017). However, a centralized clustering technique such as DBSCAN is ineffective for big data sets because analyzing and clustering large data sets carries a very large time and memory overhead (Forero et al., 2011; Wang et al., 2016). Therefore, we applied distributed DBSCAN, which can rapidly and accurately label big datasets, and we validated the labels recurrently-six times, which means that each single data point was categorized into a subset six times and received a label from each clustering process. The framework for the distributed DBSCAN is shown in Fig. 4. Because the two parameters (trip duration and detour rate) have different scales, we normalized the entire data set before clustering to ensure that the data were evenly distributed across the three dimensions, which is a common process used for unsupervised learning. After clustering, only 50,165 points were clustered to different labels after six iterations, accounting for 0.99 % of the whole raw data set, which contained 5,060,576 points. We assigned labels for those wavering data points by rounding the average values of the labels in the six repetitions.

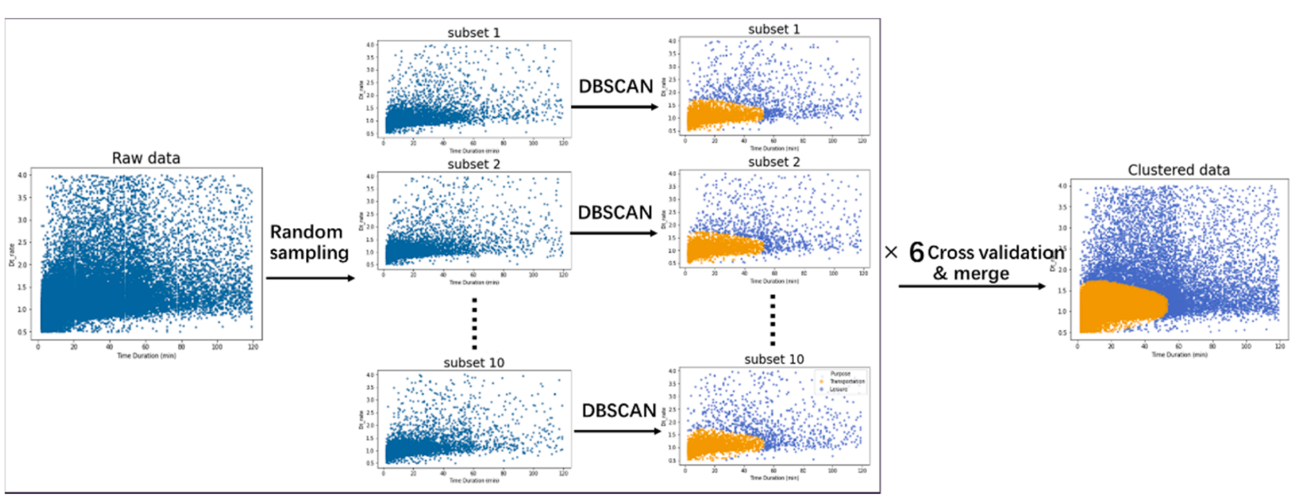


Fig. 4. Conceptual diagram of the distributed DBSCAN.

3.3.2. Identifying the relationships between bike sharing and public transit

After identifying bike trips with potential relationships with the public transit system (transportation), we used proximity analysis to consider the spatial relationships between the origins and destinations (OD) of bike trips and transit service range. Basically, if a bike rental dock had at least one public transit station nearby, we regarded trips to or from it as having the potential to interact with the public transit system. Following previous studies (Jin et al., 2019; Kong et al., 2020) and coverage capacity, 100 m was set as the threshold to identify the conditions in which riders might use shared bikes in conjunction with public transit. In Seoul, about 2000 docks are located within the 100-meter buffers around transit stations.

We used the following workflow by conducting a purpose analysis, proximity analysis, trip duration analysis, and public transit transfer analysis, as shown in Fig. 3.

Substitution means that a bike-sharing trip replaced a public transit trip. If a particular bike-sharing trip had public transit stations near both the origin and destination and the route of transit did not include transfers, then the user could have used public transit to complete the trip while minimizing walking or transfers. Using the proximity analysis shown in Fig. 3, we identified bike trips with origin and destination both within 100-m of a transit station. Then, in the transfer analysis, we calculated how many transit options without any transfers (0 transfers) were offered from those origins to those destinations.

Complementation means that riders use shared bikes in areas not well covered by public transit. In other words, a bike-share trip complements public transit if its origin, its destination, or both are far away from a transit station or if both its origin and destination are near a transit station, but the trip requires many transfers.

Integration describes a situation in which riders use shared bikes to connect to the subway and train system. This definition requires that either the origin or destination be very close to a transit station. In addition, the duration of the bike-share part of the trip was expected to be short (Ma et al., 2018). We set a 12-minute threshold, which is the average time required for an adult to ride 2 miles and the median duration of all transportation trips. The sensitivity test of this threshold can be found in Appendix Fig. 2

The three categories are all included under the heading “transportation trip” in Fig. 3.

3.3.3. Validation

Given the lack of ground-truth labels for trip type in the big data set and related State Preference survey, we could not directly validate our four categories. However, because we wanted to use solid methodology and analysis, we used the Korean National Travel Survey (KNTS) () to indirectly validate our categories (The Korea Transport Institute, 2021). A detailed explanation of the validity of KNTS can be found in Appendix B. KNTS recorded the trip durations of citizens using private bikes as a mode of transportation. We utilized this distribution to validate our classification of bike trips for different purposes and types of interactions with the public transit system.

Researchers working in physics and mathematics have shown that trips taken using different transportation modes have distinctive mobility patterns (Brockmann et al., 2006; Ma et al., 2018), such as a gamma distribution, power law distribution, or lognormal distribution. Therefore, if our classification approach is accurate, the distribution of data labeled as transportation trips in this paper should be more similar to the data in the national survey than the leisure trips. The error metrics MD (min), R2, MAE, MSE, and RMSE can be found in Appendix B.

3.3.4. Machine learning analysis and interpretation

Our objectives are to promote synergy between public bikes and existing transit systems, and to improve services for leisure riders due to the numerous societal benefits of cycling, notably public health. Having classified bike trips and analyzed their distribution patterns, we are particularly interested in understanding how different natural and built environments influence each type of bike trip. We used interpretable machine learning to answer this question. We tested the linear regression, random forest, XGBoost, gradient boosting, and gradient boosting regressor with their basic parameter settings to find the one most suitable for our data. Because of the complexity of our data, the machine learning model algorithms fit better than the linear model according to R^2, which indicated that the potential driving factors have nonlinear relationships with each type of bike–transit interaction. To avoid uneven distribution and overfitting problems, we used 5-fold cross-validation. The cross-validation scores conducted on R^2 were used to measure the average accuracy.

Among the models we tested the best model, the gradient boosting regressor, gives a prediction in the form of an ensemble of weak prediction models, which are typically decision trees. It exhibited the best quality in our dataset, with decreased running time and higher accuracy than the other models.

To interpret machine learning models, Shapley additive explanations (SHAP) provides local explanations for individual predictions (Lundberg et al., 2020; Lundberg & Lee, 2017). SHAP values are based on cooperative game theory and are used to increase the transparency and interpretability of machine learning models. In our research, we used them to interpret the contributions of different factors to the numbers of bike trips by type.

All indicators for these models are listed in Appendix Table 3.

4. Results

4.1. Validation

Because we used anonymous data, we did not have precise information about why and how people used bike sharing at an individual level. We could only infer their purpose and relationship with the transit system on a large scale based on existing theory and data processing methods. Therefore, it was necessary to validate and justify our inferences to ensure that our research classifications were appropriate. The Korean National Travel Survey was used to support our inferences from another perspective. The KNTS found that riders use personally owned bikes for transportation use on rides with an 18.5-minute mean trip duration (The Korea Transport Institute, 2021). A subsample of bike-share trips that fit the scenario of the KNTS should display a distribution pattern similar to that in the survey, according to the sampling distribution in statistical inference (Lipson, 2003). The distribution for bike trip durations is displayed in Fig. 5 (a), with each color indicating a specific type of bike trip. They have different distributions: Integration trips are skewed left, and leisure trips are skewed right. Quantile-quantile plots (Q-Q plot) in Fig. 5 (b) were used to compare the distribution of the bike trips in KNTS to the distribution of each type of public trip labeled following our classification.

Substitution trips are left-skewed compared with the KNTS (Fig. 5). Because bike-sharing is more available and convenient than buying and riding a private bike, many people use bike-sharing for shorter trips, which echoes a previous study conducted using survey data (Castillo-Manzano et al., 2016). The error metrics show that the distribution of complementary trips had the lowest error value and can thus be regarded as a counterpart to the KNTS.

4.2. Descriptive analysis

Using the DBSCAN algorithm described in section 3.3.1, we successfully identified trip purposes (leisure or transportation). We randomly selected 10 % of the data points plotted in Fig. 6(a) as a sample to display the data distribution. The detour rate of transportation trips is around one, as predicted. According to Fig. 6(b), transportation trips are mainly distributed around 12 min in duration (median values are 12 and 11 min for 2018 and 2021, respectively), whereas leisure trips do not have a clear peak in their duration distribution, and the median is

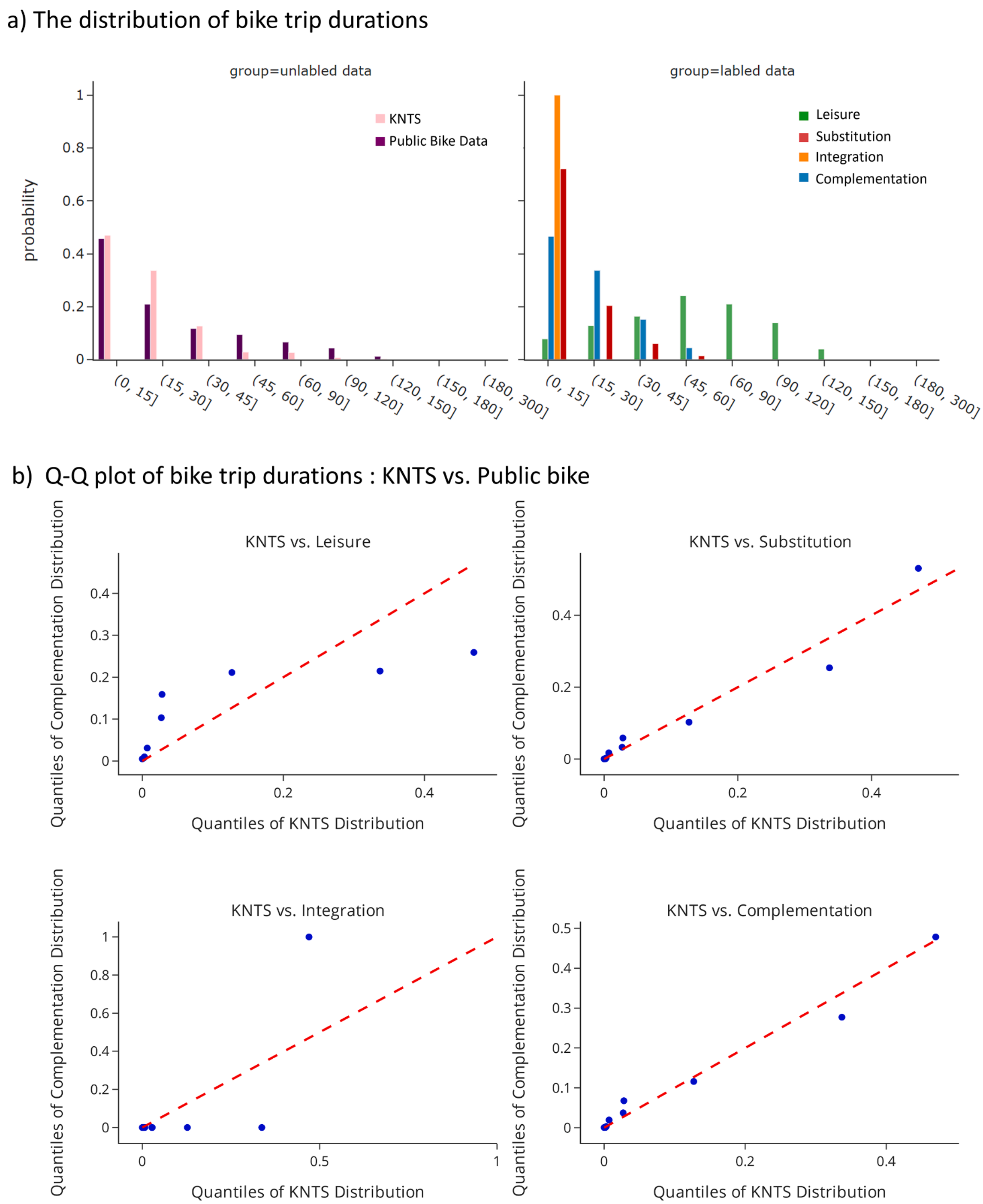


Fig. 5. Comparison of the distribution of bike trip durations.

around 50 min (median values are 53 min for both 2018 and 2021). In 2021, the growth of short-time transportation trips was significant.

Ridership varies between business days and non-business days. The absence of commuter peaks for leisure trips, both weekday and non-weekday, indicates that our classification is accurate. In some studies, researchers simply considered trips made on non-working days to be recreational trips and trips made on working days to be transportation trips (Kim et al., 2021), but we found that recreational trips accounted for only half of the overall trips on non-working days, as shown in Fig. 6 (c), and the gap between the ridership for those two purposes increased gradually from 2018 to 2021.

Fig. 7 displays the hourly average bike-sharing ridership by interaction type. Trips for transportation (blue, orange, and red) have obvious morning and evening peaks, with the morning peak lower than the evening peak. Trips for leisure (green) do not have an obvious morning or evening peak. We identified many more substitution trips in 2021 than in 2018. Complementation trips at evening peak times also increased significantly during that period. For integration trips, the ridership displays a similar commuting peak pattern as public transit, with two peaks having similar levels of ridership.

Aggregating each trip to its departure dock allowed us to identify the types of trips that accounted for the largest number of trips from each dock, as shown in Fig. 8. Each concentric circle displays the number of trips by type, while the biggest circle displays the main interaction category. Leisure trips (red circles) were concentrated and distributed more commonly in waterfront and green areas. Substitution trips

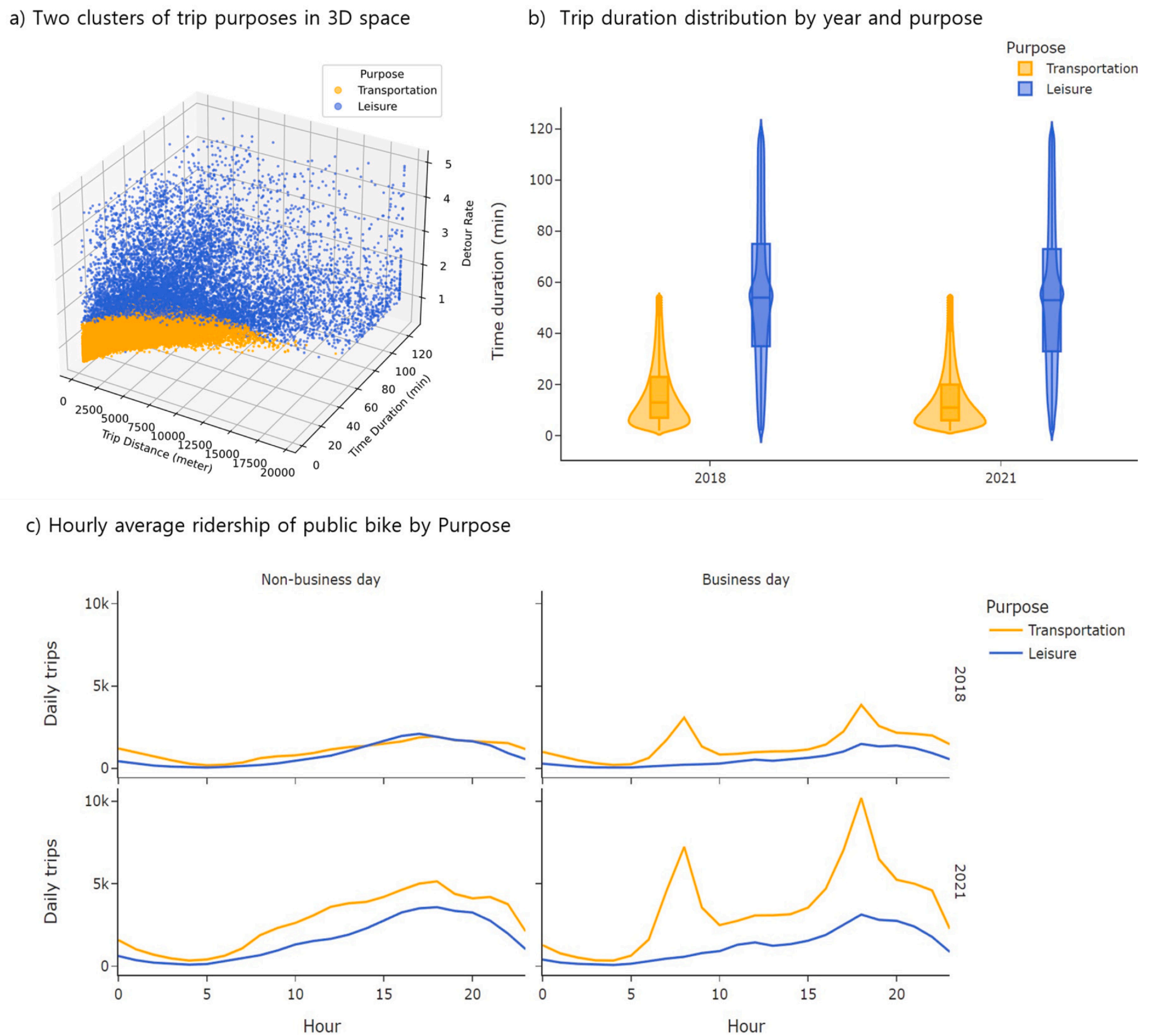


Fig. 6. Cluster results of trips by purpose.

(yellow dots) were located in densely populated areas. The blank area at the bottom right is the Gangnam district; we hypothesize that the lack of bicycle trips there is due to the steep terrain.

4.3. Machine learning models and evaluation

To validate the insights derived from our temporal and spatial analyses, and to delve deeper into the factors influencing the integration of bike sharing with public transit, we employed machine learning models. The variables are described in Table 1.

When we tested the data using the linear model, XGBoost, random forest, and gradient boosting regressor with basic parameters, the gradient boosting regressor showed the best accuracy (Appendix Table 4). We therefore chose the gradient boosting model and optimized its hyperparameters through the grid research method. The model performances are presented in Table 2.

For the four types of public trips, the cross-validation scores of the gradient boosting models are all higher than 0.665, which indicates that models explain more than 65 % of the heterogeneity.

4.4. Interpretation of the regression with explainable machine learning

Because the gradient boosting model had the highest accuracy, we used it with the SHAP explainer to understand the effects of the built and natural environments on each type of bike-share trip. A SHAP value plot is an aggregation of scatters. The colors of the scatters display the feature values (dependent variables). A red scatter represents an analysis unit with a high value. In our case, a red scatter means that the daily usage for a dock is high. In contrast, blue indicates a low feature value. The position of the scatter represents the value of the independent variable that corresponds to this one unit of analysis. Therefore, the SHAP value plot shows how each independent variable affects the feature value. Fig. 9 visualizes the relative importance of the determining factors and shows how the values of each dock contribute to its usage for each interaction type. The top 10 variables are sorted in descending order by their global importance. Each scatter displays local explanations.

The positive or negative influences, such as slope and Neighborhood Commercial, are generally universal. However, in Seoul, non-linearly varying thresholds can give planners clear information about improving public bicycle infrastructure services. The dependence plot (Fig. 10) shows that the usage of a bike dock grows with the number of

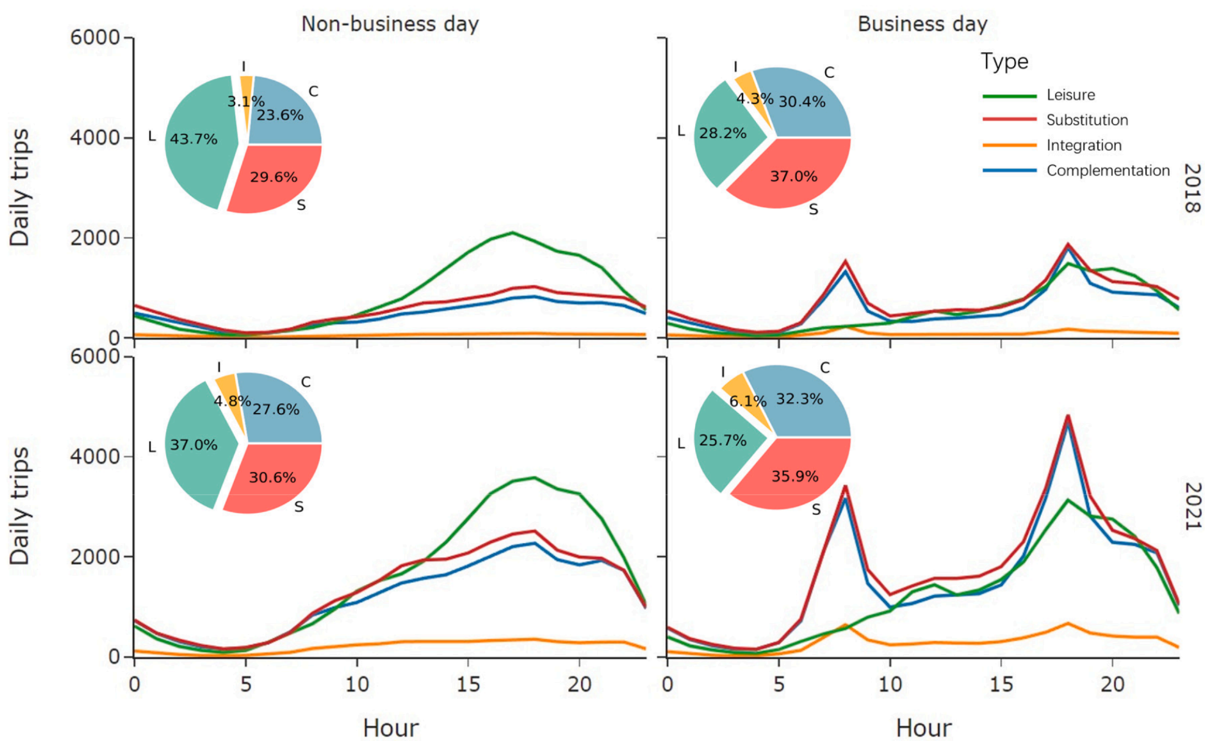


Fig. 7. Distribution of public bike trips by interaction type.

neighbor docks within a 500-meter buffer until 12–15 neighbors are present, and areas with a mean slope lower than 30 are deemed bikeable by Seoul citizens.

4.4.1. *Leisure trips*

For the number of leisure trips, the most important factor is the slope (Slope_mean, 1st). We used the mean slope value within a 500-meter buffer of each dock, excluding the river area. Leisure travel is more likely to occur along rivers (river_area, 3rd) than elsewhere, so long as sufficient bikes are available (bike_num,5th) and neighborhood commercial facilities are present (Neighborhood_Commercial, 8th). Green areas (green_area, 7th) also promoted leisure usage, but the effect was not as significant as river access (river_area, 2nd). Public transportation accessibility (Den_bus_count, Den_train) is not very important for leisure trips, which means that people tend to choose areas with nice natural environments rather than areas from which they can easily return home.

4.4.2. *Substitution trips*

The most important factors for the number of bike-share uses to substitute the public transit are the slope (Slope_mean, 1st) and the number of transit routes (Transit_routes, 2nd). The denser the transit routes are, the more people chose to use public bikes. Business_day (4th) is a dummy variable, where 1 represents a business day, and 0 represents a non-business day. As shown in the SHAP plot for substitution trips (Fig. 9 (b)), substitution trips occurred more often on business days than non-business days for all docks. The positive effects of BikePath_len (8th) and Bike_num (9th) suggest that substitution trips usually occur where bike infrastructures are available. The indicator of affluence (Car_own, 10th) had a negative effect on bike sharing, which indicates that bike sharing has not generally been accepted as a substitution for public transit by citizens of middle class and higher socioeconomic status.

4.4.3. *Integration and Complementation trips*

The important factors for these two trip types are similar. The distance to the nearest transit station ranked first and fifth in integration and complementation trips, respectively, which might be due to the criteria that we used to classify them. Therefore, these results should be interpreted with caution. The density of foreign residents (fifth most significant factor affecting Integration), which is not significant for other types of trips, indicates that foreigners are less likely to accept public bikes as integration with public transit, but use public bikes for Leisure, Substitution and Complementation similarly to Korean citizens. In addition to listing the top 10 factors for each category, we display SHAP value plots for all 25 independent variables in Appendix Fig. 3 as a reference.

5. Discussion and conclusions

5.1. *Discussion*

The main contribution of this study is its analysis and discussion of the role of bike sharing and the spatiotemporal distributions of transportation and leisure functions in Seoul using fine-resolution methodology powered by machine leaning algorithms.

The methodology proposed in this paper overcomes three limitations common in the existing literature about bike–transit connections. The first limitation is the cost of data and computation. Previously, researchers in Korea introduced the Google Map Direction API to accessibility research (Ha and Lee, 2016), which saved the time required to build transit networks and eliminated the need for massive calculations. However, Google no longer allows users to process the data further, and the method is computationally expensive, especially when the trip durations of millions of OD pairs need to be calculated (Directions API Overview, 2022). Therefore, we applied a simple path algorithm to calculate the transfer options for each OD pair based on the transit network. Second, the distributed clustering method we applied is appropriate for any set of real-world big data because commonly used clustering algorithms such as K-means and DBSCAN have high computational costs when used with big data (Shirkhorshidi et al., 2014). Overall, when considering the potential application of our methods and results to practical planning scenarios, using an effective algorithm with

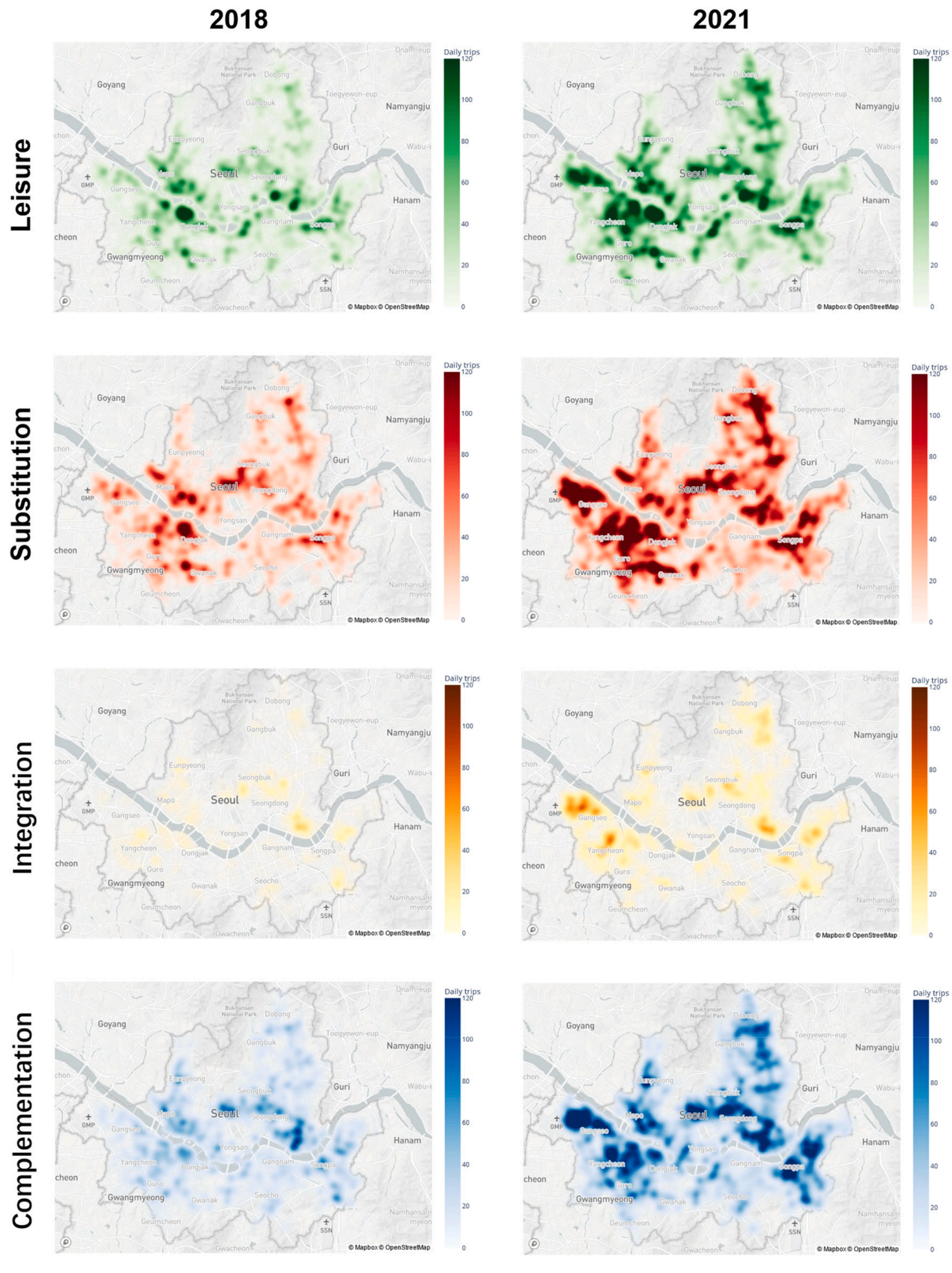


Fig. 8. Spatial pattern of public bike trips by interaction type.

local data can protect information security and offer stable service.

We used empirical evidence about bike and transit relationships in Seoul to classify the type of each bike trip. Existing research had limitations in protocols for classification, assuming that all trips with the same OD points were leisure trips (Lee and Noland, 2021) or wrongly classifying part of the morning peak commuting flow as leisure trips (Lee et al., 2021b). In terms of interpretability, our model of independent variables can explain the presence of various interaction types more accurately than the 20 % accuracy reported in previous studies (Kong et al., 2020; Lee et al., 2021b), which allows us to offer a better understanding of public bike usage than has previously been available.

Based on the empirical case study in this paper, we make the following specific recommendations and suggestions, by which the city of Seoul can effectively promote bike sharing.

First, we suggest that the optimal number of bike docks for Seoul is 6000–8000 (there were 2590 docks in 2022). This number is based on the results of our nonlinear relationship analysis. According to our data, the usage of a bike dock in Seoul grows with the number of neighbor docks within a 500-meter buffer until 12–15 neighbors are present. A dock-based public bike system maintains the cleanliness of a city but sacrifices flexibility. Ensuring a sufficient number of bike docks solves that problem and make biking more competitive with dock-less systems (Laa & Emberger, 2020).

Second, we suggest building a bike-friendly city taking into account factors that encourage leisure trips and increase the attraction of transportation trips. Places with better bike infrastructure, higher density of Neighborhood Commercial areas, and lower slopes consistently attract bike riders for any type of bike-sharing trip. In other words, citizens prefer to ride bikes in hotspots that have these attributes. For instance, other than building more bike infrastructure, e-bikes could be

Table 1
Descriptions of independent variables.

Category	Variable name	Description
Dependent Variables	Leisure	Daily average number of leisure trips for each dock
	Substitution	Daily average number of substitution trips for each dock
	Integration	Daily average number of integration trips for each dock
	Complementation	Daily average number of complementation trips for each dock
Temporal Variables	Year	2018/2021
	Business_day	0/1
Natural Environment	River_area	Area of river within 500-meter buffer of the rental dock
	Slope_mean	Mean value of slope within 500-meter buffer excluding the river area
	Green_area	Area of green land within 500-meter buffer of the rental dock
Land Use	LUM	The land use mix (LUM) of the following 11 categories of land use type within 500-meter buffer from the dock, measured by Shannon entropy.
	Educational	Area of educational facilities within 500-meter buffer of the bike dock (m^2)
	Entertainment	Area of entertainment facilities within 500-meter buffer of the bike dock (m^2)
	Residential	Area of residential land within 500-meter buffer of the bike dock (m^2)
	Medical	Area of medical facilities within 500-meter buffer of the bike dock (m^2)
	Neighborhood_Commercial	Area of neighborhood commercial facilities within 500-meter buffer of the bike dock (m^2)
	Office	Area of offices within 500-meter buffer of the dock (m^2)
	Religious	Area of religious facilities within 500-meter buffer of the bike dock (m^2)
	Retail	Area of retail facilities within 500-meter buffer of the bike dock (m^2)
	Senior & Childcare	Area of senior and childcare facilities within 500-meter buffer of the bike dock (m^2)
	Sports	Area of sports facilities within 500-meter buffer of the bike dock (m^2)
	Tourism	Area of tourism facilities within 500-meter buffer of the bike dock (m^2)
Bicycle Infrastructure	Bike_num	Number of bike rakes
	Exits_days	How many days this dock has existed
	BikePath_len	Length of bike path within 500-meter buffer the bike dock
	Near_dock	Number of bike docks in 500-meter buffer of the bike dock
Accessibility of Transportation Facilities	Distances_to_trans	Distance to the nearest station from the bike dock
	Den_bus_count	Number of buses arriving in the area per hour
	Den_train	Number of subways arriving in the area per hour
	Transit_routes	Number of transit routes (subway and bus) within the 200-meter buffer of the dock
Industrial Distribution	I-Consu&Serv	Percentage of employees in Accommodation, Food Service Activities, and Education
	I-PubAdmin	Percentage of employees in Public Administration and Defense and Compulsory Social Security
	I-Health&SocialWork	Percentage of employees in Human Health and Social Work Activities
	I-Manuf&Stor	Percentage of employees in Manufacturing, Transportation, and Storage
	I-Fina&Tech	Percentage of employees in Financial and Insurance Activities, Financial and Insurance Activities, and Information and Communication
Socioeconomic	Car_own	Car ownership per household for the smallest administrative area where the rental dock is located
	ReDensity	Total population divided by the area for the smallest administrative area where the rental dock is located (thousands/km2)
	%Foreigner	Percentage of foreign residents of the smallest administrative area where the rental dock is located
	%Female_employee	Percentage of female employees of the smallest administrative area where the rental dock is located

used to address the problems associated by hilly terrain and better assist public transportation in Seoul (Huo et al., 2021).

Third, we suggest that planners pay attention to residential areas in which transit is not available or poor to enhance the supplementary function of bike sharing. Considering the distribution of bike docks in Seoul, we believe that scarcity of docks limits bike usage. Currently, 95.7 % of docks are within 200 m of a transit station, as per our proximity analysis. Therefore, in areas where transit is unavailable, bike dock services are also not available as a supplement for complementation or integration. A previous critique in other parts of the world demonstrated that low-income households are less likely than wealthier ones to have both trip origin and destination be within bike-sharing coverage, and are unlikely to have adequate cycling infrastructure (Médard de Chardon, 2019), which might also be a problem in Seoul. Our results show that only a few bike-sharing trips (around 5 %) in Seoul were used for the last/first mile to transit as integration, with even fewer for foreigners, which is different from other megacities in Asia, where bike sharing plays an important connecting role (Guo et al., 2021; Liu et al., 2012; Yang et al., 2019). After reviewing the literature and analyzing existing findings, we found that cities with better connectivity between public transit and bikes applied dockless system (Guo & He,

Table 2
Model performances with the gradient boosting regressor.

Model	Training data			Test data			Cross-validation score
	R^2	MSE	MAE	R^2	MSE	MAE	
Leisure	0.976	9.037	2.184	0.804	64.780	4.295	0.754
Substitution	0.978	6.364	1.908	0.817	51.033	4.628	0.748
Integration	0.983	0.283	0.393	0.725	5.025	1.059	0.747
Complementation	0.967	5.307	1.721	0.778	36.863	3.903	0.697

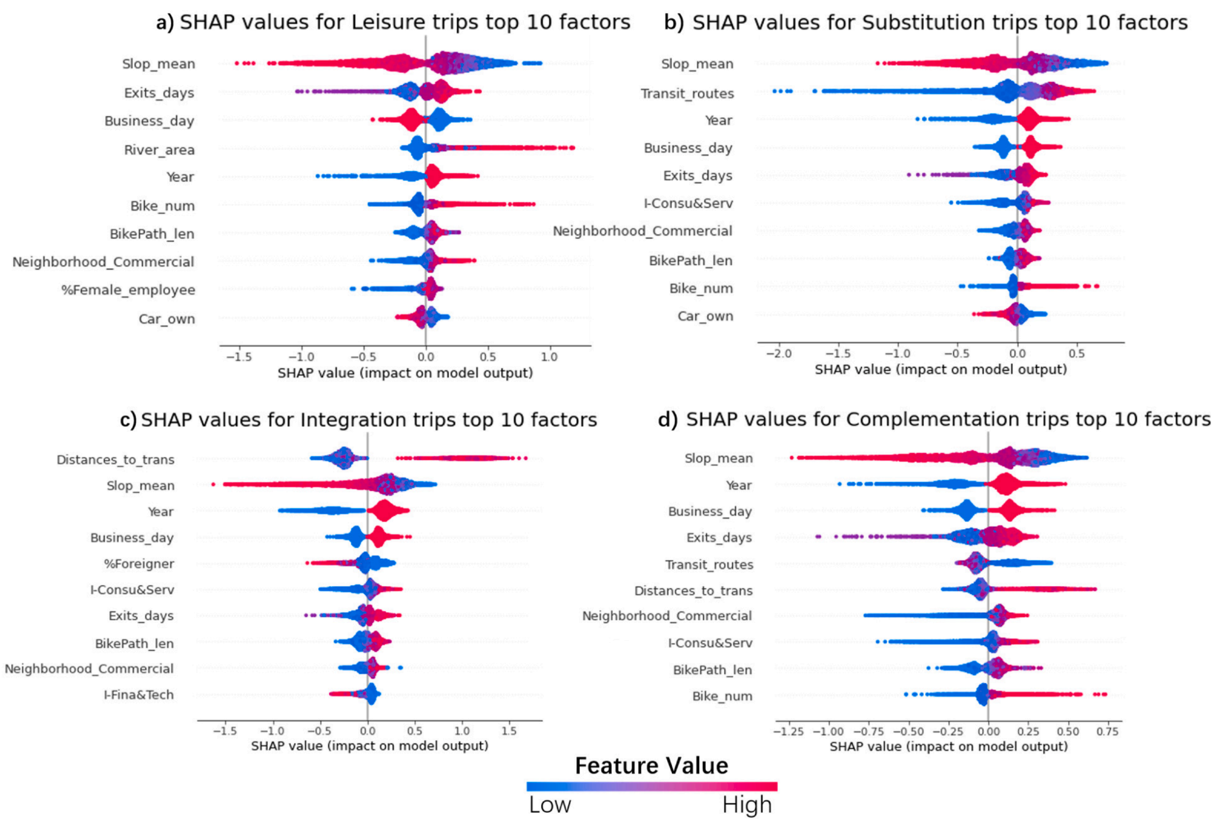


Fig. 9. SHAP summary plots.

2020; Li et al., 2021). Therefore, considering that bike docks are already more commonly installed near transit stations, residence areas should be better covered.

On the other hand, from the perspective of mass public transit, shared bikes comprise too small a share to challenge the dominant modes. In terms of the total usage of these two modes in Seoul, transit serves 5 million daily passengers, while a maximum of 0.12 million daily bike riders use public bikes during the summer peak (Appendix Fig. 1). However, we believe that BSS has the potential to be a mode of transportation in city life and improve the accessibility and resilience of the urban transportation system (Li et al., 2021; Martin & Shaheen, 2014).

As previously noted (Batty, 2013), big data directly generated by travelers cannot be used without good inference and estimation with the support of other independent datasets; thus, analyses are prone to error. Detective estimation is the main contribution of this study. However, error is unavoidable; we can only infer rider purposes and relationships when using the transit system based on existing theory, datasets, and data mining processes. The results can only be validated on a larger aggregated scale. We have provided some practical suggestions for further studies using suitable datasets as follows. We considered only spatial coverage of the transfers between transit and bikes when we distinguished Integration and Complementation trips without taking temporal coverage into consideration. Temporal coverage, especially when the headway between trains is longer than usual, might blur the classification boundary between Substitution and Complementation and cause more error using our methods. Therefore, researchers can conduct more accurate studies in the future when Google Transit Feed Specification (GTFS) data are available, as in some US cities (McHugh, 2013). In addition, the boundary between Leisure and Transportation is be blurred when and where users prioritize physical health. It is worth exploring how many people intentionally choose to take bikes and even detour during Transportation trips in the future.

5.2. Conclusions

Our study provides a framework for analyzing the purposes of public bike trips and the relationships between bike sharing and other transportation modes with a fast, accurate, and integrated data-mining approach. We emphasized the leisure and transportation functions of public bikes in Seoul and affirmed the role of the transportation function, which has become more important with the construction of better infrastructure. The spatiotemporal distributions of each type of public bike use offers useful information for real-world management of BSSs. We successfully identified the smooth ridership wave of leisure trips per day, most spatially located near the Han River, using DBSCAN, a method shown to have better performance identifying bike trips for leisure in

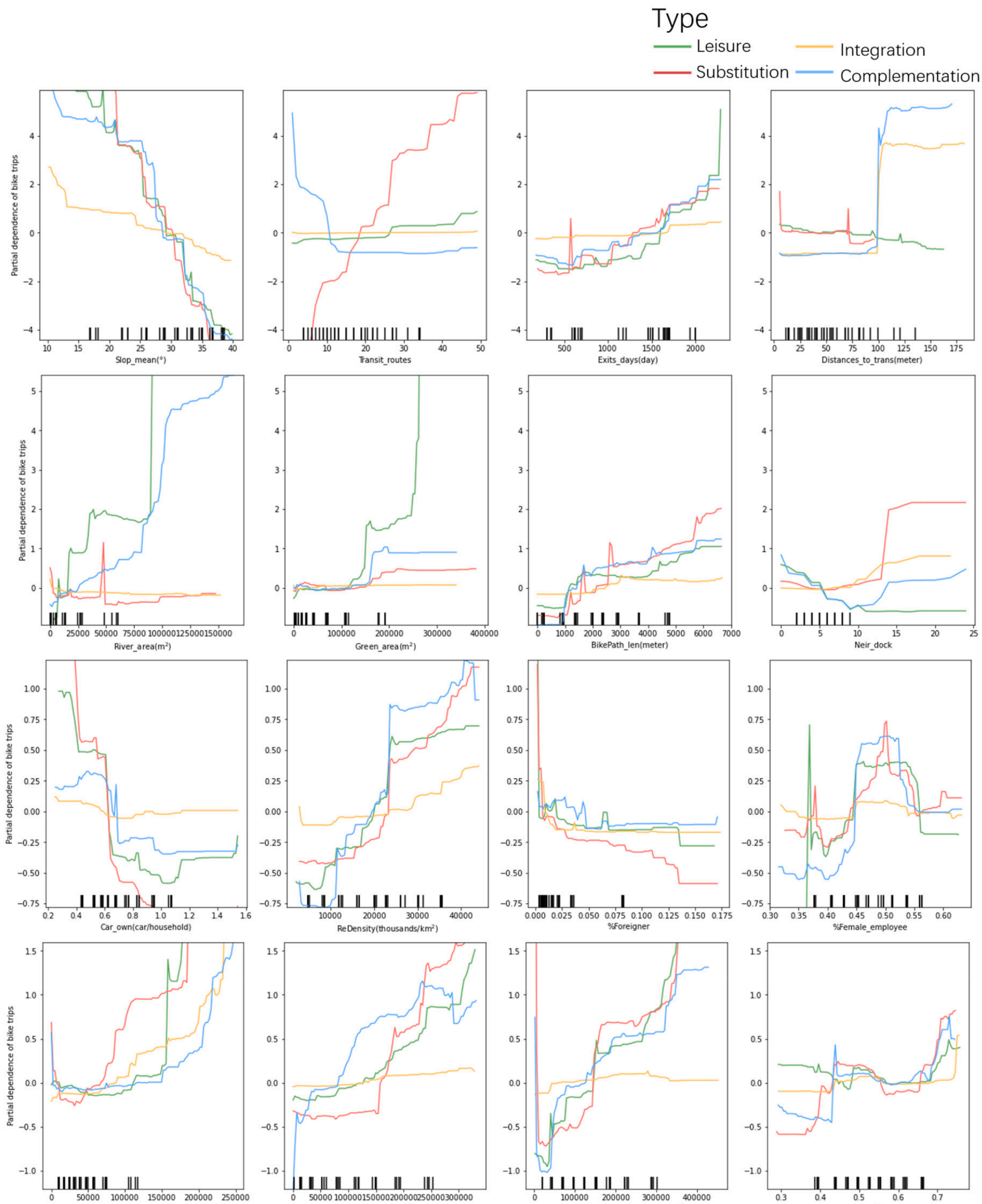


Fig. 10. Partial dependence plots for environmental factors.

Seoul. Unlike other cities, in Seoul, bikes have a competitive advantage where transit service is sufficient, but the integration function is not significant, and only accounts for 5 % of total public bike usage. The public bike system has the potential to be comprehensively improved by building better infrastructure for bikes, offering e-bikes, and installing more docks in residential areas.

Because Seoul's public bicycle infrastructure is still developing, our study offers an opportunity for planners and operators to monitor the system's use status from the perspective of two functions, transportation and physical activity. With dynamic monitoring, planners will be able to further develop bike sharing in the city and to build more effective, accessible, and comfortable services for users who have different needs. Moreover, the methodology presented here is applicable to other cities where public bike use for leisure trips needs to be taken into consideration.

CRediT authorship contribution statement

Xuan Li: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Visualization, Writing – original draft. **Jaehyun Ha:** Conceptualization, Validation, Writing – review & editing. **Sugie Lee:** Conceptualization, Methodology, Validation, Writing – review & editing, Funding acquisition, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was supported by the National Research Foundation of Korea (NRF-2023S1A5B1076175).

Appendices A–E. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.tbs.2023.100705.

References

Aultman-Hall, L., Hall, F.L., Baetz, B.B., 1997. Analysis of bicycle commuter routes using geographic information systems: implications for bicycle planning. Transp. Res. Rec. 1578 (1), 102–110. https://doi.org/10.3141/1578-13.

Batty, M., 2013. Big data, smart cities and city planning. Dialogues Hum. Geogr. 3 (3), 274–279. https://doi.org/10.1177/2043820613513390.

Bordagaray, M., dell'Olio, L., Fonzone, A., Ibeas, Á., 2016. Capturing the conditions that introduce systematic variation in bike-sharing travel behavior using data mining techniques. Transp. Res. Part C Emerg. Technol. 71, 231–248. https://doi.org/10.1016/j.trc.2016.07.009.

Broach, J., Dill, J., Gliebe, J., 2012. Where do cyclists ride? A route choice model developed with revealed preference GPS data. Transp. Res. A Policy Pract. 46 (10), 1730–1740. https://doi.org/10.1016/j.tra.2012.07.005.

Brockmann, D., Hufnagel, L., Geisel, T., 2006 Jan. The scaling laws of human travel. Nature 439 (7075), 462–465.

Buck, D., Buehler, R., Happ, P., Rawls, B., Chung, P., Borecki, N., 2013. Are bikeshare users different from regular cyclists?: A first look at short-term users, annual members, and area cyclists in the Washington, D.C., Region. Transp. Res. Rec. 2387 (1), 112–119. https://doi.org/10.3141/2387-13.

Campbell, K.B., Brakewood, C., 2017. Sharing riders: How bikesharing impacts bus ridership in New York City. Transp. Res. A Policy Pract. 100, 264–282. https://doi.org/10.1016/j.tra.2017.04.017.

Castillo-Manzano, J.I., López-Valpuesta, L., Sánchez-Braza, A., 2016. Going a long way? On your bike! Comparing the distances for which public bicycle sharing system and private bicycles are used. Appl. Geogr. 71, 95–105. https://doi.org/10.1016/j.apgeog.2016.04.003.

Chen, L.-H., Chancellor, H.C., 2020. Examining the leisure use of a bicycle share program: a case study of YouBike in Taipei. J. Leis. Res. 51 (2), 183–205. https://doi.org/10.1080/00222216.2019.1660598.

Curran, A., 2008. Translink Public Bike System Feasibility Study. Quay Communications Inc., Vancouver.

DeMaio, P., 2009. Bike-sharing: history, impacts, models of provision, and future. J. Public Transp. 12 (4), 41–56.

Directions API overview, 2022. December 31). Google Developers. https://developers.google.com/maps/documentation/directions/overview.

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. Kdd 96 (34), 226–231.

Faghih-Imani, A., Eluru, N., El-Geneidy, A.M., Rabbat, M., Haq, U., 2014. How land-use and urban form impact bicycle flows: evidence from the bicycle-sharing system (BIXI) in Montreal. J. Transp. Geogr. 41, 306–314. https://doi.org/10.1016/j.jtrangeo.2014.01.013.

Fan, A., Chen, X., Wan, T., 2019. How Have travelers changed mode choices for first/last mile trips after the introduction of bicycle-sharing systems: an empirical study in Beijing, China. J. Adv. Transp. 2019, e5426080. https://doi.org/10.1155/2019/5426080.

Fishman, E., 2016. Bikeshare: a review of recent literature. Transp. Rev. 36 (1), 92–113. https://doi.org/10.1080/01441647.2015.1033036.

Fishman, E., Washington, S., Haworth, N., 2013. Bike share: a synthesis of the literature. Transp. Rev. 33 (2), 148–165. https://doi.org/10.1080/01441647.2013.775612.

Forero, P.A., Cano, A., Giannakis, G.B., 2011. Distributed clustering using wireless sensor networks. IEEE J. Sel. Top. Signal Process. 5 (4), 707–724. https://doi.org/10.1109/JSTSP.2011.2114324.

Fraser, S.D.S., Lock, K., 2011. Cycling for transport and public health: a systematic review of the effect of the environment on cycling. Eur. J. Pub. Health 21 (6), 738–743. https://doi.org/10.1093/eurpub/ckq145.

Fuller, D., Gauvin, L., Kestens, Y., Morency, P., Drouin, L., 2013. The potential modal shift and health benefits of implementing a public bicycle share program in Montreal, Canada. Int. J. Behav. Nutri. Phys. Act. 10 (1), 66. https://doi.org/10.1186/1479-5868-10-66.

Gao, X., Lee, G.M., 2019. Moment-based rental prediction for bicycle-sharing transportation systems using a hybrid genetic algorithm and machine learning. Comput. Ind. Eng. 128, 60–69. https://doi.org/10.1016/j.cie.2018.12.023.

Guo, Y., He, S.Y., 2020. Built environment effects on the integration of dockless bike-sharing and the metro. Transp. Res. Part D: Transp. Environ. 83, 102335 https://doi.org/10.1016/j.trd.2020.102335.

Guo, Y., Yang, L., Lu, Y., Zhao, R., 2021. Dockless bike-sharing as a feeder mode of metro commute? The role of the feeder-related built environment: analytical framework and empirical evidence. Sustain. Cities Soc. 65, 102594 https://doi.org/10.1016/j.scs.2020.102594.

Guo, Y., Yang, L., Chen, Y., 2022. Bike share usage and the built environment: a review. Front. Public Health 10. https://www.frontiersin.org/articles/10.3389/fpubh.2022.848169.

Ha, J., Lee, S., 2016. An analysis of vulnerable areas for public transit services using API route guide information—focused on the mobility to major employment centers in Seoul, Korea. Korea Plan. Assoc. 51 (5), 163–181.

Ha, J., Lee, S., Ko, J., 2020. Unraveling the impact of travel time, cost, and transit burdens on commute mode choice for different income and age groups. Transp. Res. A Policy Pract. 141, 147–166. https://doi.org/10.1016/j.tra.2020.07.020.

Habib, K.N., Mann, J., Mahmoud, M., Weiss, A., 2014. Synopsis of bicycle demand in the City of Toronto: investigating the effects of perception, consciousness and comfortability on the purpose of biking and bike ownership. Transp. Res. A Policy Pract. 70, 67–80. https://doi.org/10.1016/j.tra.2014.09.012.

Huo, J., Yang, H., Li, C., Zheng, R., Yang, L., Wen, Y., 2021. Influence of the built environment on E-scooter sharing ridership: a tale of five cities. J. Transp. Geogr. 93, 103084 https://doi.org/10.1016/j.jtrangeo.2021.103084.

Jin, S.T., Kong, H., Sui, D.Z., 2019. Uber, public transit, and urban transportation equity: a case study in New York City. Prof. Geogr. 71 (2), 315–330. https://doi.org/10.1080/00330124.2018.1531038.

Kapuku, C., Kho, S.-Y., Kim, D.-K., Cho, S.-H., 2022. Modeling the competitiveness of a bike-sharing system using bicycle GPS and transit smartcard data. Transp. Lett. 14 (4), 347–351. https://doi.org/10.1080/19427867.2020.1758389.

Kim, K., 2021. Impact of COVID-19 on usage patterns of a bike-sharing system: case study of Seoul. J. Transp. Eng. Part A Syst. 147 (10), 05021006. https://doi.org/10.1061/JTEPBS.0000591.

Kim, J., Ki, D., Lee, S., 2021. Analysis of travel mode choice change by the spread of COVID-19: the case of Seoul, Korea. J. Korea Plan. Assoc. 56 (3), 113–129. https://doi.org/10.17208/jkpa.2021.06.56.3.113.

Kong, H., Jin, S.T., Sui, D.Z., 2020. Deciphering the relationship between bikesharing and public transit: modal substitution, integration, and complementation. Transp. Res. Part D: Transp. Environ. 85, 102392 https://doi.org/10.1016/j.trd.2020.102392.

Ku, D.-G., Um, J.-S., Byon, Y.-J., Kim, J.-Y., Lee, S.-J., 2021. Changes in passengers' travel behavior due to COVID-19. Sustainability 13 (14), Article 14. https://doi.org/10.3390/su13147974.

Laa, B., Emberger, G., 2020. Bike sharing: regulatory options for conflicting interests – case study Vienna. Transp. Policy 98, 148–157. https://doi.org/10.1016/j.tranpol.2020.03.009.

Lee, M., Hwang, S., Park, Y., Choi, B., 2021a. Factors affecting bike-sharing system demand by inferred trip purpose: integration of clustering of travel patterns and geospatial data analysis. Int. J. Sustain. Transp. 1–14. https://doi.org/10.1080/15568318.2021.1943076.

Lee, K., Jung, W.-S., Park, J.S., Choi, M.Y., 2008. Statistical analysis of the Metropolitan Seoul Subway System: network structure and passenger flows. Physica A 387 (24), 6231–6234. https://doi.org/10.1016/j.physa.2008.06.035.

Lee, S., Noland, R.B., 2021. Bikeshare trips in Seoul, South Korea. Case Stud. Transp. Policy 9 (2), 743–755. https://doi.org/10.1016/j.cstp.2021.03.011.

Lee, S., Shin, S., Nam, D., Kim, J., Park, J., 2021c. The analysis correlation subway and bike sharing ridership before and during COVID-19 pandemic in Seoul. J. Korea Inst. Intelligent. Transp. Syst. 20 (6), 14–25. https://doi.org/10.12815/kits.2021.20.6.14.

Lee, J., Yu, K., Kim, J., 2021b. Public bike trip purpose inference using point-of-interest data. ISPRS Int. J. Geo Inf. 10 (5), Article 5. https://doi.org/10.3390/ijgi10050352.

Li, C., Xiao, W., Zhang, D., Ji, Q., 2021. Low-carbon transformation of cities: understanding the demand for dockless bike sharing in China. Energy Policy 159, 112631. https://doi.org/10.1016/j.enpol.2021.112631.

Li, A., Zhao, P., Huang, Y., Gao, K., Axhausen, K.W., 2020. An empirical analysis of dockless bike-sharing utilization and its explanatory factors: case study from Shanghai, China. J. Transp. Geogr. 88, 102828 https://doi.org/10.1016/j.jtrangeo.2020.102828.

Li, S., Zhuang, C., Tan, Z., Gao, F., Lai, Z., Wu, Z., 2021. Inferring the trip purposes and uncovering spatio-temporal activity patterns from dockless shared bike dataset in Shenzhen, China. J. Transp. Geogr. 91, 102974 https://doi.org/10.1016/j.jtrangeo.2021.102974.

Lipson, K., 2003. The role of the sampling distribution in understanding statistical inference. Math. Educ. Res. J. 15 (3), 270–287. https://doi.org/10.1007/BF03217383.

Liu, Z., Jia, X., Cheng, W., 2012. Solving the last mile problem: ensure the success of public bicycle system in Beijing. Procedia. Soc. Behav. Sci. 43, 73–78. https://doi.org/10.1016/j.sbspro.2012.04.079.

Lumsdon, L., Tolley, R., 2001. The National Cycle Strategy in the UK: to what extent have local authorities adopted its model strategy approach? J. Transp. Geogr. 9 (4), 293–301. https://doi.org/10.1016/S0966-6923(01)00022-9.

Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.-I., 2020. From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. 2 (1), Article 1. https://doi.org/10.1038/s42256-019-0138-9.

Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. Adv. Neural Inf. Proces. Syst. 30. In: https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html.

Ma, X., Ji, Y., Yang, M., Jin, Y., Tan, X., 2018. Understanding bikeshare mode as a feeder to metro by isolating metro-bikeshare transfers from smart card data. Transp. Policy 71, 57–69. https://doi.org/10.1016/j.tranpol.2018.07.008.

Martin, E.W., Shaheen, S.A., 2014. Evaluating public transit modal shift dynamics in response to bikesharing: a tale of two U.S. cities. J. Transp. Geogr. 41, 315–324. https://doi.org/10.1016/j.jtrangeo.2014.06.026.

Martin, R., Xu, Y., 2022. Is tech-enhanced bikeshare a substitute or complement for public transit? Transp. Res. A Policy Pract. 155, 63–78. https://doi.org/10.1016/j.tra.2021.11.007.

McHugh, B., 2013. Pioneering open data standards: the GTFS Story. In: Beyond Transparency: Open Data and the Future of Civic Innovation, 125–135.

Médard de Chardon, C., 2019. The contradictions of bike-share benefits, purposes and outcomes. Transp. Res. A Policy Pract. 121, 401–419. https://doi.org/10.1016/j.tra.2019.01.031.

Miranda-Moreno, L.F., Nosal, T., 2011. Weather or not to cycle: temporal trends and impact of weather on cycling in an urban environment. Transp. Res. Rec. 2247 (1), 42–52. https://doi.org/10.3141/2247-06.

Murphy, E., Usher, J., 2015. The role of bicycle-sharing in the city: analysis of the irish experience. Int. J. Sustain. Transp. 9 (2), 116–125. https://doi.org/10.1080/15568318.2012.748855.

Park, J., Namkung, O.S., Ko, J., 2023. Changes in public bike usage after the COVID-19 outbreak: a survey of Seoul public bike sharing users. Sustain. Cities Soc. 96, 104716 https://doi.org/10.1016/j.scs.2023.104716.

Politis, I., Fyrogenis, I., Papadopoulos, E., Nikolaidou, A., Verani, E., 2020. Shifting to shared wheels: factors affecting dockless bike-sharing choice for short and long trips. Sustainability 12 (19), Article 19. https://doi.org/10.3390/su12198205.

Pucher, J., Buehler, R., 2007. At the frontiers of cycling. Policy innovations in the Netherlands, Denmark, and Germany. https://repository.difu.de/jspui/handle/difu/126085.

Qiu, L.-Y., He, L.-Y., 2018. Bike sharing and the economy, the environment, and health-related externalities. Sustainability 10 (4), Article 4. https://doi.org/10.3390/su10041145.

Ricci, M., 2015. Bike sharing: a review of evidence on impacts and processes of implementation and operation. Res. Transp. Bus. Manag. 15, 28–38. https://doi.org/10.1016/j.rtbm.2015.03.003.

Romm, D., Verma, P., Karpinski, E., Sanders, T.L., McKenzie, G., 2022. Differences in first-mile and last-mile behaviour in candidate multi-modal Boston bike-share micromobility trips. J. Transp. Geogr. 102, 103370 https://doi.org/10.1016/j.jtrangeo.2022.103370.

Ross-Perez, A., Walton, N., Pinto, N., 2022. Identifying trip purpose from a dockless bike-sharing system in Manchester. J. Transp. Geogr. 99, 103293 https://doi.org/10.1016/j.jtrangeo.2022.103293.

Schubert, E., Sander, J., Ester, M., Kriegel, H.P., Xu, X., 2017. DBSCAN revisited, revisited. ACM Trans. Database Syst. (TODS). 42 (3), 1–21. https://doi.org/10.1145/3068335.

Shaheen, S.A., Guzman, S., Zhang, H., 2010. Bikesharing in Europe, the Americas, and Asia: past, present, and future. Transp. Res. Rec. 2143 (1), 159–167. https://doi.org/10.3141/2143-20.

Shelat, S., Huisman, R., van Oort, N., 2018. Analysing the trip and user characteristics of the combined bicycle and transit mode. Res. Transp. Econ. 69, 68–76. https://doi.org/10.1016/j.retrec.2018.07.017.

Shen, Y., Zhang, X., Zhao, J., 2018. Understanding the usage of dockless bike sharing in Singapore. Int. J. Sustain. Transp. 12 (9), 686–700. https://doi.org/10.1080/15568318.2018.1429696.

Shirkhorshidi, A.S., Aghabozorgi, S., Wah, T.Y., Herawan, T., 2014. Big data clustering: a review. In: Murgante, B., Misra, S., Rocha, A.M.A.C., Torre, C., Rocha, J.G., Falcão, M.I., Taniar, D., Apduhan, B.O., Gervasi, O. (Eds.), Computational Science and Its Applications – ICCSA 2014. Springer International Publishing, pp. 707–720. https://doi.org/10.1007/978-3-319-09156-3_49.

Sung, H., Lee, S., Cheon, S., 2015. Operationalizing Jane Jacobs's urban design theory: empirical verification from the great City of Seoul, Korea. J. Plan. Educ. Res. 35 (2), 117–130. https://doi.org/10.1177/0739456X14568021.

The Korea Transport Institute, Final Report of 2021 Korean National Travel Survey, 2021, pp. 111–395. https://www.ktdb.go.kr/www/contents.do?key=1324.

Tu, Y., Chen, P., Gao, X., Yang, J., Chen, X., 2019. How to make dockless bikeshare good for cities: curbing oversupplied bikes. Transp. Res. Rec. 2673 (6), 618–627. https://doi.org/10.1177/0361198119837963.

van Mil, J.F.P., Leferink, T.S., Annema, J.A., van Oort, N., 2021. Insights into factors affecting the combined bicycle-transit mode. Public Transp. 13 (3), 649–673. https://doi.org/10.1007/s12469-020-00240-2.

Wan, L., Gao, S., Wu, C., Jin, Y., Mao, M., Yang, L., 2018. Big data and urban system model - substitutes or complements? A case study of modelling commuting patterns in Beijing. Comput. Environ. Urban Syst. 68, 64–77. https://doi.org/10.1016/j.compenvurbsys.2017.10.004.

Wang, Y., Chen, Q., Kang, C., Xia, Q., 2016. Clustering of electricity consumption behavior dynamics toward big data applications. IEEE Trans. Smart Grid 7 (5), 2437–2447. https://doi.org/10.1109/TSG.2016.2548565.

Wei, M., Liu, X., 2022. How wet is too wet? Modelling the influence of weather condition on urban transit ridership. Travel Behav. Soc. 27, 117–127. https://doi.org/10.1016/j.tbs.2022.01.001.

Well-Intergrated Transport, 2019, September 4. Seoul Urban Solutions Agency - SUSA. https://susa.or.kr/en/content/well-intergrated-transport.

Wu, Y., Li, W., Yu, Q., Li, J., Li, W., 2022. Analysis of the relationship between dockless bicycle-sharing and the metro: connection, competition, and complementation. J. Adv. Transp. 2022, e5664004. https://doi.org/10.1155/2022/5664004.

Xing, Y., Handy, S.L., Mokhtarian, P.L., 2010. Factors associated with proportions and miles of bicycling for transportation and recreation in six small US cities. Transp. Res. Part D: Transp. Environ. 15 (2), 73–81. https://doi.org/10.1016/j.trd.2009.09.004.

Xue, X., He, Q., 2009. Urban public transportation ecological niche marginal distance base on passenger trip cost analysis. In: 2009 Second International Conference on Intelligent Computation Technology and Automation, vol. 3, 872–875. doi: 10.1109/ICICTA.2009.676.

Yan, B., Janowicz, K., Mai, G., Gao, S., 2017. From ITDL to Place2Vec: reasoning about place type similarity and relatedness by learning embeddings from augmented spatial contexts. In: Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 1–10. doi: 10.1145/3139958.3140054.

Yang, Y., Heppenstall, A., Turner, A., Comber, A., 2019. A spatiotemporal and graph-based analysis of dockless bike sharing patterns to understand urban flows over the last mile. Comput. Environ. Urban Syst. 77, 101361 https://doi.org/10.1016/j.compenvurbsys.2019.101361.

Zhang, L., Zhang, J., Duan, Z., Bryde, D., 2015. Sustainable bike-sharing systems: characteristics and commonalities across cases in urban China. J. Clean. Prod. 97, 124–133. https://doi.org/10.1016/j.jclepro.2014.04.006.

Zhou, Y., Yu, Y., Wang, Y., He, B., Yang, L., 2023. Mode substitution and carbon emission impacts of electric bike sharing systems. Sustain. Cities Soc. 89, 104312 https://doi.org/10.1016/j.scs.2022.104312.