An interpretable LightGBM-SHAP analysis of self-balance in Barcelona's DBS stations

Xuxilu Zhang*[1], Francesc Soriguera[2], and Lingqi Gu[3]

[1]PhD candidate, Barcelona Innovative Transportation, Department of Civil and Environmental Engineering, UPC BarcelonaTech, Spain.
[2]Associate professor, Barcelona Innovative Transportation, Department of Civil and Environmental Engineering, UPC BarcelonaTech, Spain.
[3]Associate professor, School of Mathematics and Statistics, Fujian Normal University, China.

SHORT SUMMARY

Docked bike-sharing systems (DBS) play a pivotal role in promoting sustainable urban mobility by facilitating active travel and reducing motorized congestion. However, their long-term efficiency relies on each station's ability to self-balance under temporally and spatially uneven demand. This study investigates the intrinsic self-balancing behavior of Barcelona's Bicing network throughout 2024, encompassing 412 stations at a daily resolution. We introduce a station self-balancing index (SSI)—a normalized indicator capturing day-to-day dock-availability variability—and employ a LightGBM–SHAP framework to quantify nonlinear associations among spatial, climatic, and temporal factors. Results show that spatial attributes, particularly residential and recreational points of interest (POIs) and building density, account for roughly two-thirds of the model-explained variance in SSI. SHAP interaction analyses further reveal that weekday work-oriented flows are associated with higher imbalance levels, whereas leisure-oriented urban morphologies help stabilize weekend operations. Overall, the findings provide interpretable, data-driven insights for enhancing DBS planning and operations, while offering a generalizable framework for assessing the self-regulatory capacity of urban micromobility systems.
Keywords: Bike-sharing systems, self-balancing, nonlinear effects, LightGBM, SHAP.

1 INTRODUCTION

Bike-sharing systems (BSSs) have emerged as a sustainable transport mode that complements public transit and promotes low-carbon mobility (Soriguera et al., 2018; Torres et al., 2024). Despite these benefits, BSSs face persistent operational challenges stemming from spatiotemporal imbalances across stations (Jiménez & Soriguera, 2024; Jiménez-Meroño & Soriguera, 2024). Frequent occurrences of empty or full stations undermine service reliability and user satisfaction, while manual redistribution remains costly and energy-intensive (X. Zhang et al., 2025). Understanding how stations autonomously maintain equilibrium through endogenous dynamics is thus central to enhancing the long-term resilience and operational efficiency of BSSs.

The concept of self-balancing in BBSs—defined as the intrinsic ability of stations to sustain stable inventory levels through user-driven flows without external intervention—has gained increasing scholarly attention(Soriguera & Jiménez-Meroño, 2020; Y. Song et al., 2024). This phenomenon reflects the natural equilibration of bicycle inflows and outflows induced by recurring travel routines (e.g., home–work–home commutes) and adaptive user responses to demand conditions. When cumulative net transfers at a station approximate zero over extended periods, self-balancing mitigates imbalance propagation and enhances overall system stability (Dell'Amico et al., 2014). Crucially, when stations can achieve self-balancing through natural user behavior, it significantly reduces the need for manual, costly, and energy-intensive vehicle redistribution efforts by the operator, thereby enhancing operational efficiency and long-term service reliability (Haider et al., 2018; Dell'Amico et al., 2018). Hua et al. (2025) formally conceptualized and quantified this mechanism by defining a metric based on net pickup–dropoff differences, demonstrating that over 70% of stations in Nanjing and New York exhibited self-balancing behavior. Similarly, X. Zhang et al. (2025) showed that user-driven retrieval and docking behaviors can autonomously stabilize certain stations in Xiamen, although their strength varies with temporal demand cycles. However, most existing studies emphasize the spatial identification of self-balancing stations rather than examining the underlying drivers—particularly the interaction between geographic (e.g., topography, land use) and climatic (e.g.,

temperature, humidity) factors—that govern these dynamics.

This study investigates the nonlinear factors influencing self-balancing in Barcelona's Bicing system. Building on prior measures, we define a station self-balancing index (SSI) as the daily standard deviation of available docks, normalized by station capacity, to quantify inventory stability. We then analyze the impact of spatial-environmental factors (elevation, transit access, land-use diversity, points of interest) and climatic-temporal characteristics (temperature, precipitation, solar radiation, daily/weekly cycles) on this index. The research objectives are threefold: (1) to determine the relative importance and direction of geographic and functional drivers of SSI; (2) to explore nonlinear and interaction effects among climatic, temporal, and spatial variables; and (3) to create a comparative modeling framework using both linear and machine learning methods to provide actionable insights for Bicing system management and planning.

2 MATERIALS AND METHODS

Study Area

The empirical analysis focuses on Barcelona, Spain, whose public bike-sharing system, Bicing, serves over 164,000 registered users and operates more than 500 docking stations across the metropolitan area. According to official statistics, the system recorded 1,792,755 trips in June 2025. Bicing is deeply integrated into Barcelona's multimodal transport network and spans diverse land-use and topographic conditions, providing an ideal setting for investigating long-term, station-level self-balancing dynamics.

Data

Operational data were obtained from the OpenData Barcelona platform, covering the period from January to December 2024. The dataset includes high-frequency station-level availability and trip records updated approximately every five minutes. To ensure temporal consistency, only stations that operated continuously throughout the year and reported both pickup and return data on every day were retained, yielding a final sample of 412 stations.

To characterize environmental and contextual conditions, multiple supplementary datasets were integrated. Public transport stop locations (bus, metro, tram, and rail) were sourced from OpenData Barcelona, while street geometry and network topology were extracted from OpenStreetMap (OSM) to compute metrics such as road density, junction density, and network connectivity. Topographic elevation data were obtained from the Open-Elevation API. For points of interest (POIs), 51,073 OSM entries were categorized into five functional groups—traffic, life, recreation, work, and residential—using a rule-based mapping framework. Hourly meteorological data, including temperature, relative humidity, precipitation, wind speed, UV index, and solar radiation, were retrieved from the VisualCrossing API.

Variables

Dependent Variable

Previous studies typically evaluate station self-balancing using the net difference between pickups and drop-offs (Hua et al., 2025). However, this static measure fails to reflect temporal fluctuations in station inventory—an intrinsic characteristic of docked systems with finite capacity constraints.

To more accurately capture a station's intrinsic stability, we introduce the station self-balancing index (SSI), which measures the normalized intra-day variability in empty dock availability:

$$SSI_{i,d} = \frac{\mathrm{std}(D_{i,d})}{C_i} = \frac{\mathrm{std}(D_{i,d})}{D_{i,d,t} + B_{i,d,t}}, \tag{1}$$

where $D_{i,d,t}$ and $B_{i,d,t}$ denote, respectively, the number of empty docks and available bikes observed at time t on day d at station i. The total docking capacity of station i is denoted by C_i. In principle, $D_{i,d,t} + B_{i,d,t} = C_i$ at all times. In practice, however, telemetry delays and sensor errors can cause slight deviations. To obtain a robust estimate of C_i, we define it as the annual mode of observed total slots, thereby reducing the impact of outliers.

The numerator $\mathrm{std}(D_{i,d})$, representing the standard deviation of dock availability within a day, captures intra-day fluctuations in dock availability, while normalization by C_i ensures comparability across stations

of varying sizes and utilization levels. By definition, $SSI_{i,d}$ ranges from 0 to 0.5, where 0 denotes a perfectly stable inventory (no intra-day variation) and 0.5 represents maximal oscillation between full and empty states.

Lower SSI values imply stronger natural self-balancing, while higher values indicate dependency on external redistribution. Aggregating the index at the daily level filters transient noise yet retains recurring behavioral patterns (e.g., commuting peaks), thereby facilitating analysis of persistent equilibrium dynamics in docked systems.

Explanatory variables

Following established literature and data availability, explanatory variables are grouped into three domains: spatial environment, temporal context, and climatic conditions. Spatially, four dimensions are defined: (1) Density is captured by building density (BD), junction density (JD), and road density (RD), representing built-environment intensity (K. Song & Diao, 2025; L. Zhang et al., 2025). (2) Diversity is measured by land-use diversity (LUD) using Shannon entropy within a 400m network buffer, approximating walking distance to bike-sharing stations (Liu et al., 2024). (3) POI distribution is represented by category-specific POI counts (POI_c, where $c \in \{\text{tra, life, rec, work, res}\}$) (Cai et al., 2025). (4) Accessibility and topography are accounted for by distance to the nearest bus stop (DB), distance to the nearest rail stop (DR), the number of DBS stations within the buffer (NS) and elevation (E), capturing multimodal connectivity and terrain effects (Böcker et al., 2020). Temporally, a public-holiday indicator (IsH) and day of the week (DoW) (Peláez-Rodríguez et al., 2024) capture cyclical and institutional effects. Climatically, hourly weather data—temperature (T), precipitation (P), wind speed (W), solar radiation (SR), relative humidity (H), feels-like temperature (FT), and UV index (UV)—capture environmental influences (Dai et al., 2023). All variables are summarized in Table 1.

Methods

A Light Gradient Boosting Machine (LightGBM) was employed to quantify the joint influence of climatic, temporal, and spatial factors on station self-balancing (SSI). LightGBM is a tree-based ensemble algorithm that aggregates multiple weak learners into a strong predictive model through gradient boosting (Ke et al., 2017). Compared to traditional regression models, it effectively captures nonlinear relationships and higher-order feature interactions while maintaining computational efficiency, making it particularly suitable for large-scale, heterogeneous BSS datasets with intertwined spatial–temporal structures.

To prevent overfitting and enhance model generalization, Bayesian optimization via the Optuna framework was used to automatically tune hyperparameters (Shekhar et al., 2021). The optimization minimized cross-validated mean squared error using a k-fold strategy stratified by station ID to avoid spatial data leakage.

To interpret model behavior, SHAP (Shapley Additive Explanation) analysis was applied to decompose predictions into additive feature contributions. Based on cooperative game theory (Shapley et al., 1953), the SHAP value ϕ_i for feature i measures its average marginal contribution across all feature subsets:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(M - |S| - 1)!}{M!} [f(S \cup \{i\}) - f(S)], \tag{2}$$

where S is a subset of the full feature set N with cardinality M, and $f(\cdot)$ denotes the model's prediction function. SHAP provides both global and local interpretability, enabling systematic analysis of nonlinear effects and feature interactions that drive spatial–temporal variability in SSI.

3 RESULTS AND DISCUSSION

Feature preprocessing and selection

A two-stage feature selection procedure was used to improve model stability, interpretability, and computing efficiency. First, multicollinearity among predictors ($|r| > 0.8$) was found using pairwise Pearson correlation analysis, which resulted in the elimination of FT and UV. Second, all retained features showed acceptable multicollinearity (VIF < 10) (O'brien, 2007), according to a variance inflation factor (VIF) study. This process produced a statistically orthogonal feature set that maintained the explanatory variables' information content while guaranteeing model robustness.

Table 1: Variable definitions, descriptions, and summary statistics.

Category	Variable	Description	Min	Max	Mean	Std
Independent Variable						
	SSI (dimensionless)	Station self-balancing index	0.00	0.48	0.19	0.09
Spatial Environment						
	BD (km^2/km^2)	Building density	0.06	1.40	0.80	0.30
	RD (km/km^2)	Road density	0.01	0.14	0.08	0.01
	JD ($count/km^2$)	Junction density	0.00	103.45	48.33	20.59
	LUD (dimensionless)	Land-use diversity (Shannon index)	0.51	1.50	1.13	0.18
	POI_tra (count)	Transportation POIs	17.00	278.00	147.44	46.85
	POI_life (count)	Lifestyle POIs	0.00	97.00	27.88	17.86
	POI_rec (count)	Recreation POIs	1.00	250.00	58.01	42.59
	POI_work (count)	Workplace POIs	0.00	574.00	14.23	35.03
	POI_res (count)	Residential POIs	0.00	940.00	232.39	197.22
	DR (m)	Distance of DBS stations to nearest rail station	0.00	1100.00	296.51	202.01
	DB (m)	Distance of DBS stations to nearest bus station	0.00	600.00	96.74	78.25
	NS (count)	Number of DBS stations within the buffer area	0.00	8.00	3.19	1.71
	E (m)	Topographic elevation	0.00	158.00	46.79	33.74
Meteorological Conditions						
	T (°F)	Air temperature	41.50	85.00	62.70	10.30
	FT (°F)	Perceived temperature	37.20	85.60	62.44	11.30
	H (%)	Relative humidity	38.10	92.20	70.88	9.50
	P (mm)	Precipitation	0.00	5.71	0.08	0.37
	W (km/h)	Wind speed	6.20	28.50	13.51	3.78
	SR (W/m^2)	Solar radiation	15.60	319.70	159.61	84.22
	UV (dimensionless)	UV index	1.00	10.00	6.17	2.18
Temporal Context						
	IsH (binary)	Public holiday indicator	0.00	1.00	0.04	0.20
	DoW (ordinal)	Day of week (0=Mon,6=Sun)	0.00	6.00	2.99	2.00

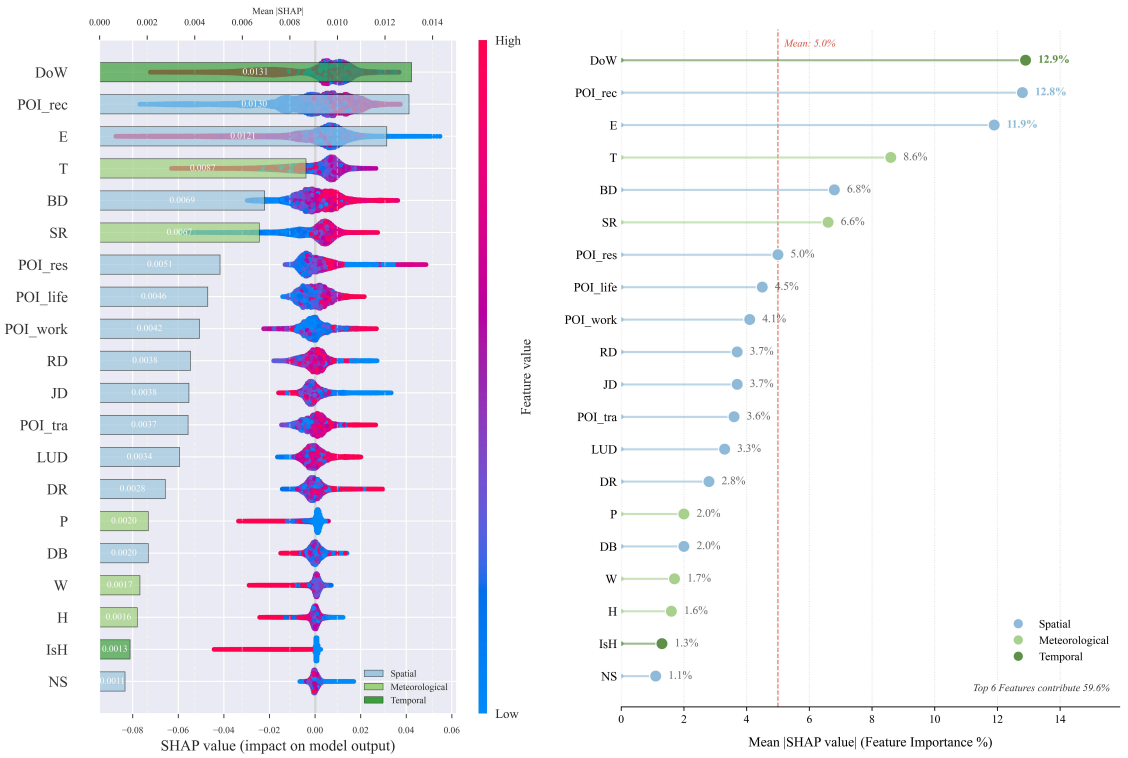
Relative importance of features

All continuous variables were standardized prior to model fitting (Marquardt, 1980). Specifically, in this study, all variables except IsH and DoW were subjected to standardization, ensuring that subsequent analyses involving continuous predictors were based on standardized values. The optimized LightGBM configuration employed a learning rate of 0.0101, 1,350 estimators, a maximum depth of 21,107 leaves, a minimum of 38 samples per leaf, a subsample ratio of 0.909, a colsample-by-tree ratio of 0.630, an L1 regularization term of 1.792, and an L2 regularization term of 0.579, achieving an R^2 of 0.6745, an RMSE of 0.0509, and an MAE of 0.0405. For benchmarking, an ordinary least squares (OLS) model and two ensemble models (GBDT and XGBoost) were fitted using identical standardized features (Table 2). Comparative results affirm LightGBM's superior predictive accuracy and reliability for assessing station-level self-balancing relative to linear and other ensemble baselines.

Table 2: Performance comparison of different models.

Model	R^2	**RMSE**	**MAE**
OLS	0.1601	0.0817	0.0674
GBDT	0.5590	0.0592	0.0476
XGBoost	0.6214	0.0548	0.0436
LightGBM	**0.6745**	**0.0509**	**0.0405**

Figure 1 presents the global ranking and relative contributions of explanatory variables derived from mean absolute SHAP values. As shown in Figure 1a, the six most influential features—DoW, POI_rec, E, T, BD, and SR—collectively explain 59.6% of total mean absolute contributions (Figure 1b). Spatial-environment variables dominate with a combined 65.5% contribution, followed by meteorological (20.4%) and temporal variables (14.2%), highlighting the preeminent role of spatial and morphological determinants in shaping global SSI dynamics.



(a) (b)

Figure 1: Feature importance analysis with SHAP. (a) Global feature importance (b) Distribution of feature importance.

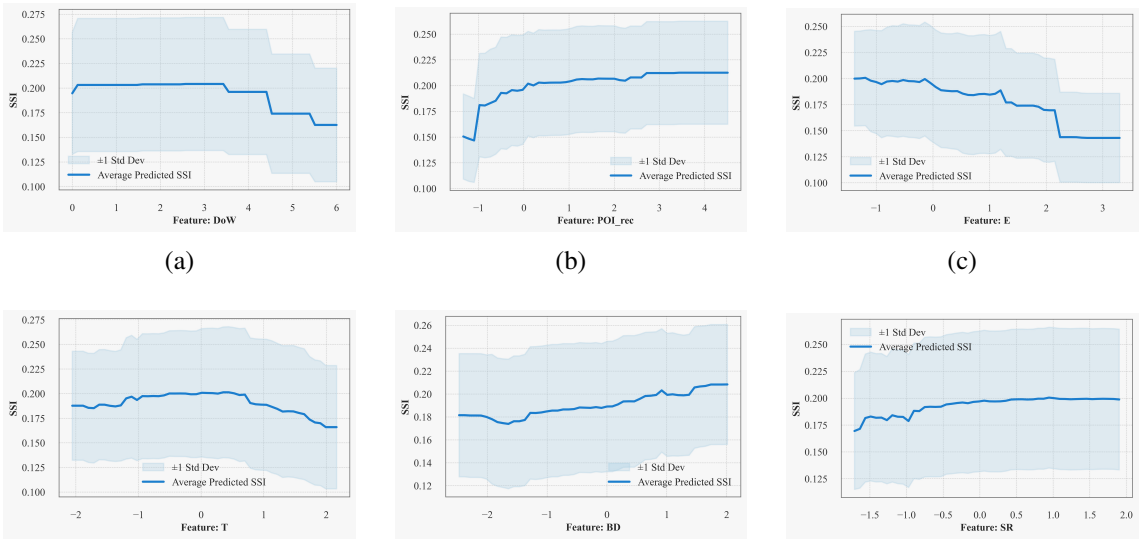
The local SHAP dependence plots (Figure 1a) illustrate the directionality of feature effects. Each point is color-coded by its normalized feature value (red = high, blue = low), while the x-axis denotes its marginal contribution to SSI. Positive SHAP values indicate higher SSI (i.e., reduced self-balancing), whereas negative values indicate enhanced self-balancing capacity. The plots reveal that higher levels of POI_rec, BD, SR, POI_life, POI_tra, LUD, and DR predominantly yield negative SHAP impacts, thus strengthening station self-balancing. Conversely, variables such as DoW, E, T, P, W, H, IsH, and NS produce positive SHAP effects, implying that extreme or elevated values may hinder the endogenous equilibrium of DBS stations. Some variables—particularly POI_res, POI_work, JD, RD, and DB—display nonlinear effects, reflecting threshold-dependent or context-sensitive influences on self-balancing performance.

Nonlinear association of variables

Partial dependence plots (PDPs) for the six dominant features (Figure 2) elucidate the nonlinear relationships between these variables and SSI, where higher SSI values indicate weaker intrinsic self-balancing in the system.

In the spatial domain, recreation-related points of interest (POI_rec, Figure 2b) exhibit a monotonic increase in predicted SSI with rising POI density, plateauing at higher intensities. This pattern suggests that areas with greater recreational amenities experience diminished natural equilibrium, necessitating enhanced active redistribution. Similarly, building density (BD, Figure 2e) displays an ascending trend in SSI, implying that more compact urban environments are associated with reduced self-balancing capacity and thus demand greater intervention for stability. In contrast, elevation (E, Figure 2c) yields a threshold response: SSI rises modestly at low elevations but declines sharply beyond a mid-range value. This inversion at higher altitudes points to enhanced self-balancing, potentially attributable to reduced cycling demand or altered trip distributions in topographically challenging terrains.

For temporal and meteorological influences, the day-of-week effect (DoW, Figure 2a; coded 0–6 for Monday–Sunday) remains largely invariant across weekdays before declining toward the weekend. Solar radiation (SR, Figure 2f) correlates positively with SSI, such that clearer conditions align with attenuated self-balancing, possibly due to heightened recreational cycling volumes. Temperature (T, Figure 2d) peaks with SSI at moderate levels, suggesting a disruption in the natural equilibrium under thermally optimal conditions. At thermal extremes, however, SSI diminishes, reflecting bolstered self-balancing that may stem from curtailed trip volumes or shifts toward essential, shorter journeys.



(a) (b) (c)

(d) (e) (f)

Figure 2: Partial dependence of the top 6 high-contribution features on SSI: (a) DoW, (b) POI_rec, (c) E_z, (d) T_z, (e) BD_z, (f) SR. Note: continuous variables such as POI_rec, E, and so on were standardized before training the LightGBM model to ensure consistent scales, which explains the presence of negative values in the plots.

Interaction effects among variables

The long-term self-balancing capability of docked bike-sharing stations (SSI) appears to be associated with the interplay of spatial, temporal, and environmental factors. SHAP dependence plots were used to explore pairwise feature interactions and their contextual patterns (Figure 3).

Spatial–temporal interactions reveal that commuter-dominated contexts are linked to diminished intrinsic self-balancing. The interaction between day-of-week (DoW) and workplace POI density (POI_work) (Figure 3a) demonstrates that, during weekdays ($DoW = 0$–4), stations in high-POI_work areas (red hues) yield positive SHAP values, elevating SSI and thus indicating weaker natural equilibrium—consistent with pronounced unidirectional commuter flows. In contrast, on weekends ($DoW = 5$–6), these same locations exhibit subdued or negative SHAP values (blue hues), resulting in lower SSI and enhanced self-balancing, as work-related trips diminish. A parallel pattern emerges in the DoW–BD interaction (Figure 3e): Low-density building areas during weekdays show positive SHAP contributions, indicating commuter-induced imbalances. In contrast, high-density urban cores on weekends often produce attenuated or negative SHAP values, reflecting a shift from commuting to a more balanced leisure-oriented mobility.

Interactions among POI types further highlight complementary demand dynamics. The POI_rec–POI_res interplay (Figure 3b) indicates that regions with elevated intensities of both recreational and residential POIs drive positive SHAP values, thereby increasing SSI and weakening self-balancing; this likely arises from intensified trip generation and attraction in mixed-use locales. Conversely, areas with low values across both features typically produce negligible or negative SHAP contributions, fostering lower predicted SSI and greater intrinsic balance.

Environmental interactions exhibit nonlinear, threshold-dependent behaviors. The E–T interaction (Figure 3c) reveals that low-elevation sites under moderate temperatures (mid-range T colors) consistently yield positive SHAP values, amplifying SSI and eroding self-balancing. However, high-elevation areas paired with thermal extremes generate negative SHAP values, reducing SSI and bolstering equilibrium—plausibly due to suppressed cycling under such constraints. Temperature's marginal effect reinforces an inverted-U profile in SHAP dependence plots (Figure 3d), mirroring PDP trends: near-mean temperatures (optimally comfortable) correlate with positive SHAP values and heightened SSI (weaker self-balancing), whereas deviations to cold or hot extremes elicit negative SHAP values and diminished SSI (stronger self-balancing), attributable to reduced overall trip volumes or shifts toward shorter, essential journeys. Finally, solar radiation (SR) (Figure 3f) predominantly yields positive SHAP values at elevated levels, associating sunnier conditions with increased SSI (weaker natural balancing); low SR, by comparison, attenuates SSI, suggesting subdued activity in overcast scenarios.

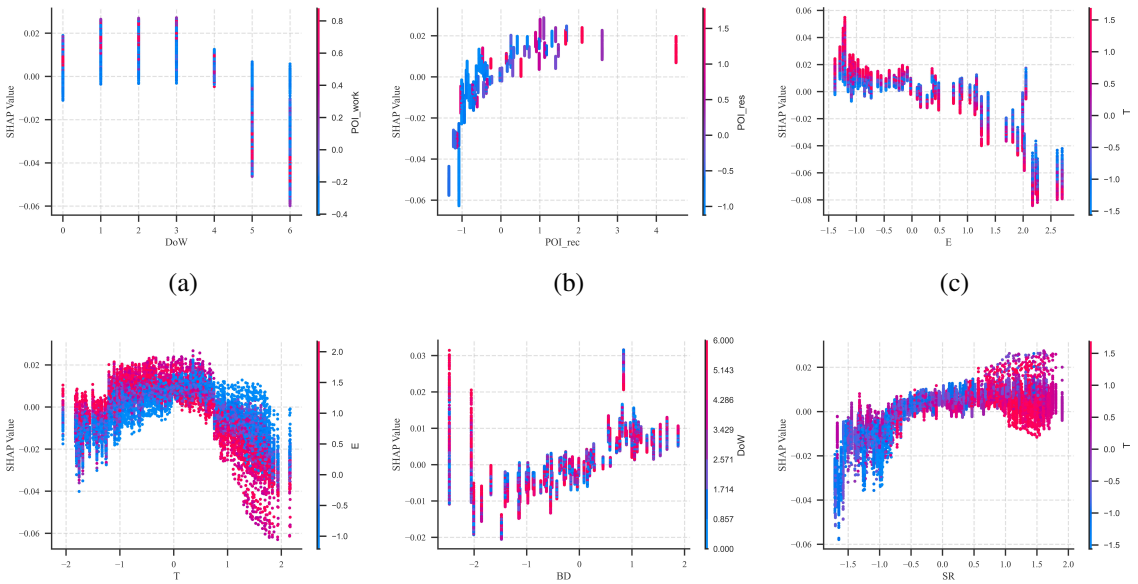


Figure 3: Two-factor SHAP dependence plots for the top 6 high-contribution features based on LightGBM: (a) DoW, (b) POI_rec, (c) E, (d) T, (e) BD, (f) SR.

4 CONCLUSIONS AND FUTURE RESEARCH

This study contributes to the understanding of long-term self-balancing in docked bike-sharing systems by introducing the station self-balancing index (SSI) and examining its potential determinants through a LightGBM–SHAP framework using Barcelona's Bicing system. Our findings show that differences in SSI are substantially correlated with spatial-environmental characteristics, especially building density and recreational and residential POIs, which account for more than two-thirds of the observed variance. Nonlinear analysis indicate threshold effects, such as temperature exhibiting a potential inverted-U connection and elevation being associated with decreased stability beyond modest levels. Further patterns of co-occurrence are shown by SHAP interaction analyses; for instance, weekend leisure in densely populated locations is associated with enhanced station self-balance, while weekday work-related flows tend to correspond with higher equilibrium. Overall, these results demonstrate the strong relationship between self-balancing patterns in BSS networks and behavioral flows that are influenced by urban form and weather–temporal contexts.

Future research should extend the framework to multi-city analyses, integrating real-time, high-resolution data and agent-based simulations to forecast dynamic imbalances. Moreover, incorporating socio-economic variables or applying causal inference methods would enhance policy relevance and foster data-driven strategies for designing scalable, resilient, and sustainable micromobility systems.

ACKNOWLEDGEMENTS

This research has been partially funded by the Spanish Ministry of Science, Innovation and Universities (Ministerio de Ciencia, Innovación y Universidades, Gobierno de España), Grant No. PID2024-159028OB-I00.

REFERENCES

Böcker, L., Anderson, E., Uteng, T. P., & Throndsen, T. (2020). Bike sharing use in conjunction to public transport: Exploring spatiotemporal, age and gender dimensions in oslo, norway. *Transportation research part A: policy and practice*, *138*, 389–401. (https://doi.org/10.1016/j.tra.2020.06.009)

Cai, X., Gu, X., Silm, S., Hadachi, A., Jin, T., & Witlox, F. (2025). Differences in bike-sharing usage and its associations with station-surrounding characteristics: A multi-group analysis using machine learning techniques. *Journal of Transport Geography*, *125*, 104201. (https://doi.org/10.1016/j.jtrangeo.2025.104201)

Dai, S., Zhao, W., Wang, Y., Huang, X., Chen, Z., Lei, J., ... Jia, P. (2023). Assessing spatiotemporal bikeability using multi-source geospatial big data: A case study of xiamen, china. *International Journal of Applied Earth Observation and Geoinformation*, *125*, 103539. (https://doi.org/10.1016/j.jag.2023.103539)

Dell'Amico, M., Hadjicostantinou, E., Iori, M., & Novellani, S. (2014). The bike sharing rebalancing problem: Mathematical formulations and benchmark instances. *Omega*, *45*, 7–19. (https://doi.org/10.1016/j.omega.2013.12.001)

Dell'Amico, M., Iori, M., Novellani, S., & Subramanian, A. (2018). The bike sharing rebalancing problem with stochastic demands. *Transportation research part B: methodological*, *118*, 362–380. (https://doi.org/10.1016/j.trb.2018.10.015)

Haider, Z., Nikolaev, A., Kang, J. E., & Kwon, C. (2018). Inventory rebalancing through pricing in public bike sharing systems. *European Journal of Operational Research*, *270*(1), 103–117. (https://doi.org/10.1016/j.ejor.2018.02.053)

Hua, M., Yu, X., Chen, X., Chen, J., & Cheng, L. (2025). Can bike sharing achieve self-balancing distribution? evidence from dockless and station-based cases. *Travel Behaviour and Society*, *38*, 100879. (https://doi.org/10.1016/j.tbs.2024.100879)

Jiménez, E., & Soriguera, F. (2024). A continuous approximation model for the optimal design of mixed free-floating and station-based car-sharing systems. *Sustainable Cities and Society*, *113*,

105673. (https://doi-org.recursos.biblioteca.upc.edu/10.1016/j.scs.2024.105673)

Jiménez-Meroño, E., & Soriguera, F. (2024). Optimization of bike-sharing repositioning operations: A reactive real-time approach. *EURO Journal on Transportation and Logistics*, *13*, 100138. (http://dx.doi.org/10.1016/j.ejtl.2024.100138) doi: 10.1016/j.ejtl.2024.100138

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, *30*.

Liu, Z., Fang, C., Li, H., Wu, J., Zhou, L., & Werner, M. (2024). Efficiency and equality of the multimodal travel between public transit and bike-sharing accounting for multiscale. *Sustainable Cities and Society*, *101*, 105096. (https://doi.org/10.1016/j.scs.2023.105096)

Marquardt, D. W. (1980). Comment: You should standardize the predictor variables in your regression models. *Journal of the American Statistical Association*, *75*(369), 87–91. (https://doi.org/10.1080/01621459.1980.10477430)

O'brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & quantity*, *41*, 673–690. (https://doi.org/10.1007/s11135-006-9018-6)

Peláez-Rodríguez, C., Pérez-Aracil, J., Fister, D., Torres-López, R., & Salcedo-Sanz, S. (2024). Bike sharing and cable car demand forecasting using machine learning and deep learning multivariate time series approaches. *Expert Systems with Applications*, *238*, 122264. (https://doi.org/10.1016/j.eswa.2023.122264)

Shapley, L. S., et al. (1953). A value for n-person games.

Shekhar, S., Bansode, A., & Salim, A. (2021). A comparative study of hyper-parameter optimization tools. In *2021 ieee asia-pacific conference on computer science and data engineering (csde)* (pp. 1–6). doi: 10.1109/CSDE53843.2021.9718485

Song, K., & Diao, M. (2025). Nonlinear and spatially-varying effects of the built environment on dockless bike-sharing usage. *Transportation Research Part D: Transport and Environment*, *145*, 104807. (urlhttps://doi.org/10.1016/j.trd.2025.104807)

Song, Y., Zhang, L., Luo, K., Wang, C., Yu, C., Shen, Y., & Yu, Q. (2024). Self-loop analysis based on dockless bike-sharing system via bike mobility chain: empirical evidence from shanghai. *Transportation*, 1–25. (https://doi.org/10.1007/s11116-024-10500-w)

Soriguera, F., Casado, V., & Jiménez, E. (2018, Jan). A simulation model for public bike-sharing systems. *Transportation Research Procedia*, *33*, 139–146. (http://dx.doi.org/10.1016/j.trpro.2018.10.086) doi: 10.1016/j.trpro.2018.10.086

Soriguera, F., & Jiménez-Meroño, E. (2020, Jan). A continuous approximation model for the optimal design of public bike-sharing systems. *Sustainable Cities and Society*, *52*, 101826. (http://dx.doi.org/10.1016/j.scs.2019.101826) doi: 10.1016/j.scs.2019.101826

Torres, J., Jiménez-Meroño, E., & Soriguera, F. (2024). Forecasting the usage of bike-sharing systems through machine learning techniques to foster sustainable urban mobility. *Sustainability*, *16*(16), 6910. (https://doi.org/10.3390/su16166910)

Zhang, L., Lieske, S. N., Pojani, D., Buning, R. J., & Corcoran, J. (2025). Hybrid machine learning-based approaches for modeling bikeability. *Journal of Transport Geography*, *123*, 104150. (https://doi.org/10.1016/j.jtrangeo.2025.104150)

Zhang, X., Gu, L., Soriguera, F., & Zhao, N. (2025). Evaluating parking load in geofenced bike-sharing systems: A case study of xiamen island. *Transport Policy*. (https://doi.org/10.1016/j.tranpol.2025.07.008)