JOURNAL OF ICIVILTECH
INNOVATIONS IN CIVIL ENGINNERING AND TECHNOLOGY

e-ISSN 2687-2129
jiciviltech, **2024**, *6*(2), 131-142

Research Article

Application of Hierarchical Clustering Algorithm to Classify Docked Bike-Sharing System Stations

*[1]Oruc ALTINTASI, [2]Dila GUZEL, [3]Sila OVGU KORKUT

[1] *Izmir Katip Celebi University, Department of Civil Engineering, Faculty of Engineering and Architecture, İzmir, Türkiye, oruc.altintasi@ikcu.edu.tr, ORCID ID: https://orcid.org/0000-0002-4217-1890*
[2] *Izmir Katip Celebi University, Department of Civil Engineering, Faculty of Engineering and Architecture, İzmir, Türkiye, dilaguze@gmail.com, ORCID ID: https://orcid.org/0009-0000-1506-1787*
[3] *Izmir Katip Celebi University, Department of Engineering Sciences, Faculty of Engineering and Architecture, İzmir, Türkiye, silaovgu.korkut@ikcu.edu.tr, ORCID ID: https://orcid.org/0000-0003-4784-2013*

Recieved: 12.10.2024; Accepted: 20.11.2024

Abstract

This study applied the hierarchical clustering algorithm to categorize bike-sharing system stations operating in İzmir, Türkiye, based on their temporal usage patterns. By distinguishing between transportation- and leisure-oriented trips, the research provides insights into the operational dynamics of the system. The findings highlighted spatial and temporal distinctions, with Konak İskele Station consistently emerging as a separate cluster for transportation-oriented trips, indicating its crucial role due to its central location and proximity to public transportation hubs. In the case of leisure-oriented trips, the analysis revealed three clusters on weekdays, with Konak İskele Station maintaining its prominence, particularly in the afternoon and late evening hours. However, weekend trips were characterized by the identification of two main clusters. These results emphasize the importance of tailored management strategies for bike-sharing systems, suggesting that transportation-oriented trips may benefit from more strategic station placements and enhanced connectivity to public transit.

Keywords: *Bike-Sharing System, Hierarchical Clustering, Temporal Usage, Leisure Trips*

*[1]Corresponding author:

To cite this article
Altintasi, O., Guzel, D. & Ovgu Korkut, S. (2024). Application of Hierarchical Clustering Algorithm to Classify Docked Bike-Sharing System Stations. *Journal of Innovations in Civil Engineering and Technology (JICIVILTECH)*, *6*(2), 131-142. https://doi.org/10.60093/jiciviltech.1565857

Bisiklet Paylaşım Sistemi İstasyonlarının Sınıflandırılmasında Hiyerarşik Kümeleme Algoritmasının Uygulanması

Öz

Bu çalışmada, İzmir'de faaliyet gösteren bisiklet paylaşım sistemi istasyonları, zamansal kullanım sıklığına göre hiyerarşik kümeleme algoritması ile sınıflandırılmıştır. Araştırma, ulaşım ve eğlence amaçlı yolculuklar arasında bir ayrım yaparak sistemin operasyonel dinamikleri hakkında değerli bilgiler sunmaktadır. Bulgular, Konak İskele İstasyonu'nun ulaşım amaçlı yolculuklar için sürekli olarak ayrı bir küme oluşturduğunu ve bu istasyonun merkezi konumu ile toplu taşıma merkezlerine yakınlığının ona önemli bir rol kazandırdığını vurgulamaktadır. Eğlence amaçlı yolculuklar incelendiğinde, yapılan analiz hafta içi üç ayrı küme ortaya koymuştur. Konak İskele İstasyonu'nun özellikle öğleden sonra ve akşam geç saatlerde yoğun bir şekilde kullanıldığı sonucuna varılmıştır. Öte yandan, hafta sonu yolculukları için istasyonlar iki ayrı küme halinde toplanmıştır. Bu sonuçlar, bisiklet paylaşım sistemleri için özel yönetim stratejilerinin önemini ön plana çıkarmakta ve ulaşım odaklı yolculukların daha stratejik istasyon yerleşimlerinden ve toplu taşımaya daha iyi bağlantılardan faydalanabileceğini göstermektedir.

Anahtar kelimeler: *Bisiklet Paylaşım Sistemi, Hiyerarşik Kümeleme, Zamansal Kullanım, Eğlence Yolculukları*

1. Introduction

Bike-sharing systems (BSS) have become a key element of sustainable urban transportation, providing a flexible and eco-friendly alternative for short-distance travels (Bullock et al., 2017). By offering an alternative to car usage, BSS helps reduce traffic congestion and improve air quality and supporting cities' efforts towards sustainable mobility goals (Fishman, 2016). While numerous studies have examined the factors influencing BSS usage, recent research highlights the potential of using advanced data analytics, such as clustering algorithms, to better understand travel behavior patterns of bike-sharing system users (Zhou, 2015; Du et al., 2019; Liu and Lin, 2019; Ma et al., 2019; Chen et al., 2022; Zhu and Diao, 2020). These studies frequently employ clustering algorithms to classify stations based on temporal usage patterns over a day. For example, Ma et al. (2019) investigated the effects of spatiotemporal factors on travel behavior of BSS users, comparing three clustering methods and ultimately using the k-means algorithm to segment 477 BSS stations into seven distinct clusters. Likewise et al. (2019) conducted two clustering analyses based on the temporal station usage intensity and trip distances, using a Multinomial Logistic Regression model to explore relationships between BSS station usage and land use structure. Zhou (2015) analyzed the spatiotemporal characteristics of bike usage in Chicago during 2013 and 2014, applying a hierarchical clustering method to categorize bike stations into five clusters according to usage intensity and time of day. This approach facilitated the analysis of weekday and weekend travel behavior across both years. Zhu and Diao (2020) adopted a fuzzy c-means clustering algorithm to classify BSS stations in Hangzhou, China, into five clusters based on temporal variations in daily usage. Collectively, these studies reveal that travel behavior patterns derived from clustering analyses are often correlated with demographic characteristics and land-use patterns. In a related study, Du et al. (2019) examined the spatiotemporal usage patterns of the BSS in Shanghai. By employing a hierarchical clustering analysis, they identified temporal usage characteristics of stations, grouping them into three clusters based on peak usage times: morning, evening, and both morning and evening. Additionally, the study employed a random forest algorithm to identify key factors influencing station usage frequency. Chen et al. (2022) investigated factors affecting BSS usage at rail stations in Nanjing, China, utilizing k-means clustering to divide 90 stations into five clusters based on peak bike rental and return times. Beck et al. (2023) found five different clusters via K-medoid analysis and associated the built environment and land use variables with each cluster.

In addition to the clustering algorithms-based studies, travel characteristics of bike-sharing trips and designing bike routes have been focused on many studies. For example, Pekdemir et al. (2024) examined the factors influencing the frequency of bike trips for different purposes using the one-year BSS trips data for İzmir, Türkiye. The results showed that the presence of car parks and rail transit stations within a 600 m catchment area of BSS stations was associated with a higher frequency of bike trips for transportation purposes. In another study by Pekdemir et al. (2021) the variations in the frequency of bike trips were examined for different seasons for the same city. Li et al. (2021) reported similar results indicating that the proximity to metro station was positively associated with the bike usage. Saplıoğlu and Aydın (2018) proposed a GIS-based multicriteria decision support system to identify safe routes for connecting bike paths with public bus stops. Accident prone areas, bus lanes and car parks were identified as the most influential variables. On the other hand, Yemişcioğlu et al. (2024) found that slope was found the most important criteria for designing the bicycle routes. Another study by Saplıoğlu and Günay (2016) identified key parameters for designing bicycle routes on urban roads through survey-based research.

The existing literature suggests that clustering algorithms have primarily been applied to large-scale BSSs trips, where bicycles are predominantly used for transportation purposes. However, travel characteristics in BSSs designed for recreational use, such as the İzmir BSS system in Türkiye, have not been extensively explored. This gap in literature forms the primary focus of this research. This study examines the unique travel characteristics associated with a smaller-scale, predominantly leisure-oriented BSS. A distinctive feature of this research is its focus on analyzing usage patterns through a temporal frequency perspective, which helps reveal the distinct dynamics of recreational BSS system. By separating travels for transportation and leisure, this study offers new insights into the operational dynamics of such systems. In this study, a hierarchical clustering algorithm was employed to analyze BSS trip data, categorizing BPS stations based on their daily temporal usage patterns. The BSS trip data covers trips recorded from 2021 to 2022 across 60 stations in İzmir, Türkiye, a period which captures post-pandemic shifts in recreational mobility. The findings clarify the travel patterns associated with BSS trips, providing insights into usage trends and behaviors. This study makes a strategic contribution to urban transportation planning by identifying distinct usage trends in leisure-oriented BSS.

2. Materials and Methods

2.1 Study area and Data Preprocessing

Data was obtained from İzmir Metropolitan Municipality for 60 BSS stations, covering a two-year period of BSS trip records from 2021 to 2022. The location of the BSS is illustrated in Figure 1. This dataset includes information for each trip, such as bicycle number, rental station, rental time, drop-off station, drop-off time, and trip duration. The initial step was to eliminate noisy trip data from the dataset (Pekdemir et al., 2024; Guzel et al., 2025). Noisy trips were characterized as those that returned to the starting station within three minutes or had a duration of 150 minutes or more (Pekdemir, 2024; Guzel et al., 2025). Furthermore, official holidays within the designated period were identified, and trips occurring on these days were excluded from the analysis.

Figure 1. BSS stations in İzmir, Türkiye.

2.2 Implementation of Hierarchical Clustering Algorithm

Before applying the clustering algorithm, it is essential to categorize BSS trips based on their purpose. This necessity arises from preliminary analyses of bike trips that indicated an unbalanced relationship between travel times and trip distances, as highlighted in authors' previous studies (Pekdemir et al., 2024; Guzel et al., 2025). To prevent any conflicts of interest, the methodology for classifying BSS trips was not discussed in this research but detailed information can be found in Guzel et al. (2025). Following the classification of BSS trips into categories of cycling for transportation and cycling for leisure, a hierarchical clustering algorithm was applied to each trip purpose, separately.

The fundamental principle of the hierarchical model is grouping the object into clusters based on their similarity. The hierarchical approach was chosen because it does not require a predefined number of clusters, labels or categories. The application of the hierarchical clustering algorithm started by the selection of choice of distance metric. A suitable distance metric was selected as Euclidean distance as similar in study (Du et al., 2019). We applied Ward's method, a hierarchical technique that reduces within-cluster variance. This approach progressively combines the two nearest clusters at each step, ensuring that the resulting clusters are as consistent and homogeneous as possible (Shalizi, 2009). More precisely,

$$d(X \cup Y, Z) = \sqrt{\frac{n_X + n_Z}{n_X + n_Y + n_Z}\|X - Z\|^2 + \frac{n_Y + n_Z}{n_X + n_Y + n_Z}\|Y - Z\|^2 - \frac{n_Z}{n_X + n_Y + n_Z}\|X - Y\|^2}$$

where X, Y, and Z denote the clusters and n_X, n_Y, and n_Z represent the number of elements within those clusters, respectively. Ward's method offers notable advantages. By aiming to minimize total variance within clusters, Ward's approach yields more balanced, well-defined, and homogeneous clusters. Additionally, applying Ward's method enhances the interpretability of trip patterns across various temporal and purpose-based categories. For further details on hierarchical clustering methods and their applications, we refer the interested reader to the works of Shalizi (2009) and Müllner (2011).

Based on the focus of the study, the hierarchical clustering algorithm was applied to stations exhibiting similar time-use patterns. To determine the optimal number of clusters, an experimental evaluation was performed using the Silhouette score, which measures the cohesion within clusters and separation between clusters. This score assesses how well each station fits within its assigned cluster compared to other clusters, with values ranging from -1 to 1. A score close to 1 indicates that the station is well-clustered and distinct from neighboring clusters, while a score near -1 suggests poor assignment, meaning the station may belong to a different cluster. The Silhouette score can be expressed as follows:

$$\max_{k}\left\{\frac{1}{N_k}\sum_j \frac{b(j)-a(j)}{\max\{a(j),b(j)\}}\right\} \quad (1)$$

where N_k denotes the number of elements over the cluster k. Moreover, $a(i)$ is the average distance of i^{th} element to other points in the same cluster whereas $b(i)$ is the average distance to the nearest other cluster. The calculation of those values is as follows:

$$a(i) = \frac{1}{|K_i|-1}\sum_{j\in K_i, i\neq j} d(i,j) \quad (2)$$

$$b(i) = \min_{m\neq i}\frac{1}{|K_m|}\sum_{j\in K_m} d(i,j) \quad (3)$$

where $|K_i|$ stands for the number of points of cluster involving i.

After categorizing the BSS trips by purpose, clustering analyses were carried out based on the temporal usage patterns of the stations. These analyses were conducted separately for weekday and weekend trips, resulting in a total of four distinct clustering analyses.

3. Results

The hierarchical structure of clusters for transportation-oriented trips is shown in Figures 2 and 3, representing weekdays and weekends, respectively. In these figures, each horizontal line corresponds to a cluster, with the length of the line indicating the distance between clusters. The dendrogram demonstrates that BSS stations can be grouped into several clusters based on varying levels of similarity. Accordingly, the Silhouette analysis assesses the quality of the clustering, where higher silhouette scores reflect better cluster solutions. The plot suggests that the optimal number of clusters falls between 2 and 3, as the silhouette coefficient reaches its peak within this range. This indicates the presence of two primary groups of BSS stations, with distinct usage patterns. The heatmap or hierarchical clustering dendrogram results for the BSS stations are depicted in Figure 4. It can be observed that Konak İskele Station forms its own cluster, while the remaining 59 BSS stations fall into a separate cluster. Additionally, Figure 4 highlights that Konak İskele Station was selected significantly more often than the other stations. The clustering results suggest a clear spatial and temporal distinction between Konak İskele Station and the other stations, possibly reflecting its central location, proximity to the central business district, and ease of access to public transportation facilities. In other words, the fact that Konak consistently forms a distinct cluster suggests a strong station-specific pattern that requires further investigation.

Regarding the clustering results for leisure-oriented trips, Figures 5 and 6 display the hierarchical clustering dendrograms for weekday and weekend trips, respectively. Unlike transportation-oriented cycling, three distinct clusters were identified for weekdays based on the silhouette score. Once again, Konak İskele Station formed its own cluster, 20 BSS stations grouped into a second cluster, and the remaining 39 stations comprised the third cluster.

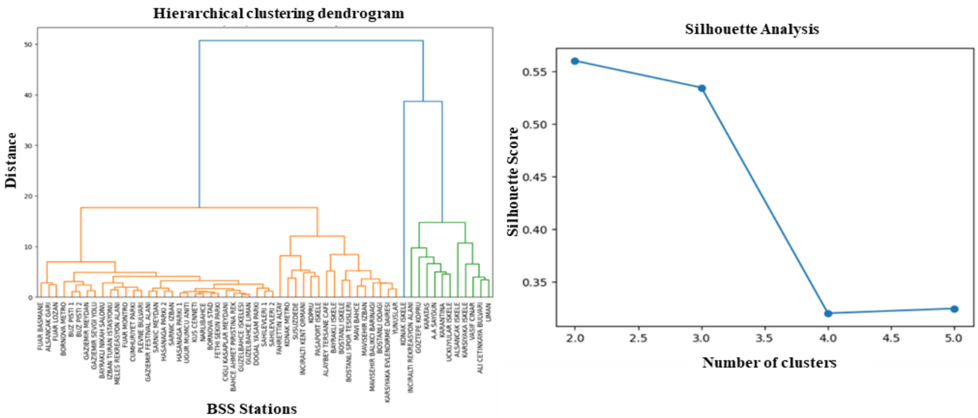

Figure 2. Clustering analysis results for transportation trips for weekdays.

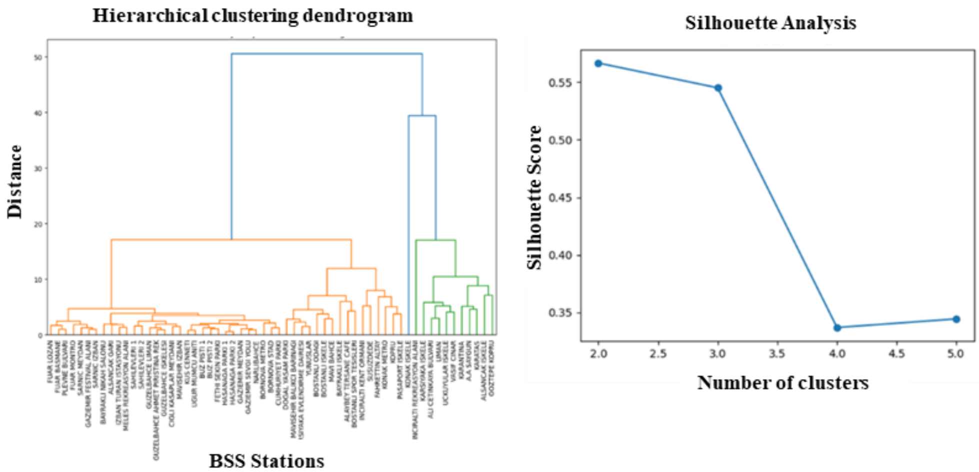

Figure 3. Clustering analysis results for transportation trips for weekends.

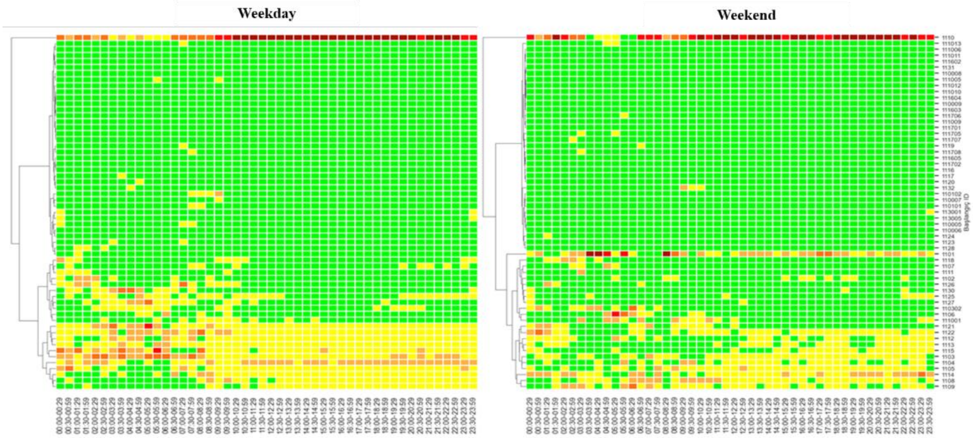

Figure 4. Hierarchical clustering dendrogram results for transportation trips.

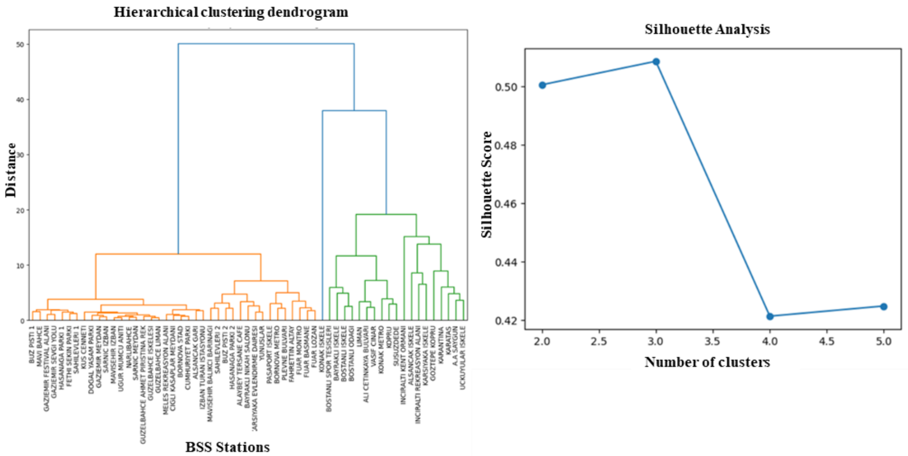

Figure 5. Clustering analysis results for cycling for leisure trips for weekdays.

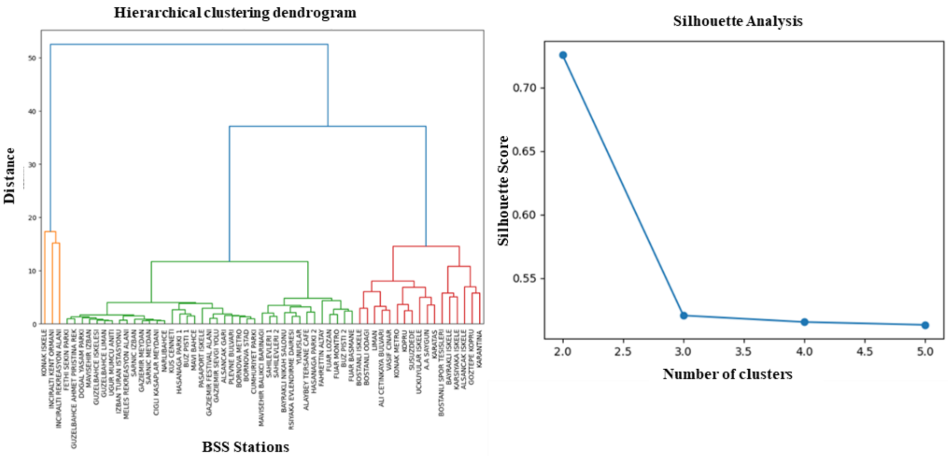

Figure 6. Clustering analysis results for cycling for leisure trips for weekends.

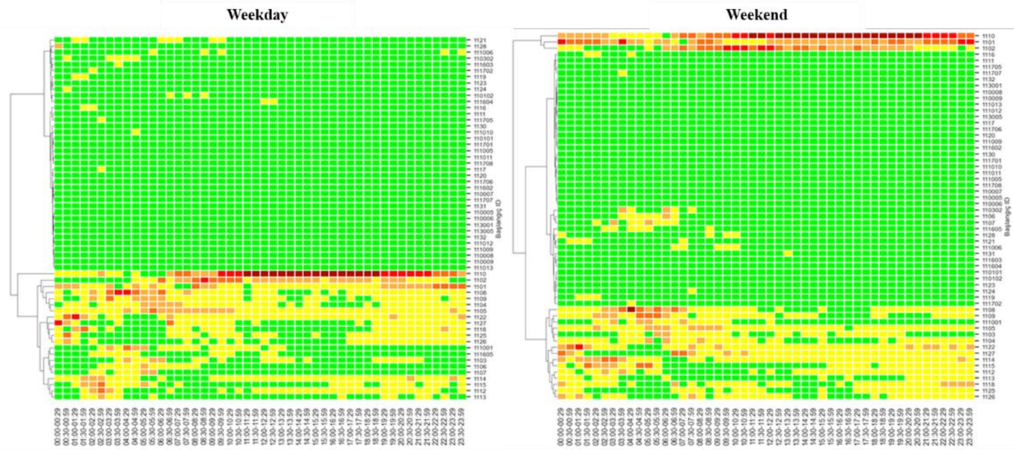

Figure 7. Hierarchical clustering dendrogram results of BSS stations for leisure trips.

For leisure-oriented trips, Konak İskele Station is primarily used in the afternoon and continues into the late evening (Figure 7). Analysis of Cluster 2 (Figure 7) reveals that several stations are preferred during morning peak hours but also experience frequent usage throughout the day. Cluster 3, highlighted in green on the heatmap, comprises stations that are less popular and used infrequently during the day. The temporal and spatial usage patterns of these clusters suggest varying levels of demand for leisure cycling, potentially influenced by factors such as closeness to recreational areas or access to public facilities. In contrast, station usage on weekends follows a different pattern, as shown in Figure 7. The Silhouette analysis indicated the presence of two distinct clusters. The hierarchical clustering dendrogram (Figure 7) revealed that three stations formed one cluster, while the remaining stations were grouped into a separate cluster. The stations in the first cluster (a total of three) are primarily located in recreational areas of Izmir, reflecting the leisure-oriented nature of their usage. This distinct clustering on weekends highlights a preference for certain stations in leisure-oriented locations, underscoring the impact of station location on trip behavior. It can be inferred that the distinctions between the usage frequencies of stations on weekdays and weekends were more observable, particularly for leisure trips (see Figure 7). This may be because leisure trips are more likely to be affected by factors such as recreational preferences, seasonal activities, and free time availability, which vary between weekdays and weekends. Leisure trips tend to increase on weekends, leading to a concentration on trips at stations near parks, or main attractions. On the other hand, the number of clustered stations for transportation-related cycling trips remains consistent on both weekdays and weekends. This likely reflects the regularity and routine nature of transportation-oriented trips. Consequently, station usage for transportation trips is less affected by the day of the week, as users often rely on predictable stations close to residential or work areas, creating stable clustering patterns across both weekdays and weekends.

4. Discussion and Conclusions

This study provides a comprehensive analysis of the travel behavior patterns associated with the İzmir Bike-Sharing System (BSS) by employing a hierarchical clustering algorithm to categorize stations according to their temporal usage patterns. The results yield valuable insights into the dynamics of the system, particularly in distinguishing between transportation- and leisure-oriented trips. The clustering analysis for transportation-oriented trips revealed a distinct spatial and temporal separation, with Konak İskele Station consistently emerging as a separate cluster. This distinct segregation highlights Konak İskele Station's strategic role within the system, which may be attributed to its central location, accessibility, and integration with the city's public transportation network. This finding underscores the station's importance, likely attributable to its central location, proximity to the central business district, and integration with public transportation. On the other hand, the remaining stations were gathered into one cluster, exhibiting similar temporal usage characteristics. This may be due to their locations in residential or mixed-use areas with less direct access to major transit hubs.

In contrast, the investigation into leisure-oriented trips uncovered different clustering patterns, especially on weekdays where three unique clusters were identified. Again, Konak İskele Station formed a unique cluster, indicating its central role in leisure trips, especially during the afternoon and late evening. Other stations in Cluster 2 demonstrated frequent use throughout the day, while those in Cluster 3, were less frequently used. The analysis of weekend trips, however, indicated only two main clusters, with three stations forming a separate group. This separation on weekends may indicate that these stations are situated near popular leisure destinations. These findings underscore the need for a strategic approach in managing bike-sharing systems by aligning station placement and service availability with observed demand patterns based on trip purposes and temporal variations. For instance, while transportation-oriented trips may benefit from more strategic placement of stations and improved connectivity with public transit, leisure-oriented stations might require enhancements in availability during peak leisure times, particularly on weekends. Future research should investigate the demographic and land-use factors influencing the observed clustering patterns.

Future research could explore the application of alternative clustering algorithms, such as k-means or C-fuzzy clustering, to compare their effectiveness in grouping trips based on purpose. While this study focused on hierarchical clustering with Ward's method, investigating how these other algorithms perform in terms of cluster homogeneity and interpretability could provide valuable insights. Additionally, incorporating a broader range of external factors, such as seasonal variations, might further refine the clustering process and offer a more comprehensive understanding of trip behaviors within bike-sharing systems, which is the focus of the future research.

Declaration of Ethical Standards

The authors declare that they comply with all ethical standards.

Credit Authorship Contribution Statement

Author 1: Resources, Research, Writing – original draft, Conceptualization, Methodology / Study design, Writing – review and editing, Supervision, Validation, Data curation,

Author 2: Methodology / Study design, Validation, Visualization, Writing – original draft, Data curation, Writing – review and editing

Author 3: Resources, Research, Writing – original draft, Conceptualization, Methodology / Study design, Writing – review and editing, Supervision, Data curation

Declaration of Competing Interest

The authors have no conflicts of interest to declare regarding the content of this article.

Data Availability

All data generated or analyzed during this study can be shared upon request.

Acknowledgement

This research has been supported by the National Research Project of Izmir Katip Celebi University, grant number of 2023-ARC-MÜMF-0002. The authors also express their thanks to Izmir Metropolitan Municipality for sharing bike-sharing trip data.

5. References

Beck, B., Winters, M., Nelson, T., Pettit, C., Leao, S. Z., Saberi, M., ... & Stevenson, M. (2023). Developing urban biking typologies: Quantifying the complex interactions of bicycle ridership, bicycle network and built environment characteristics. *Environment and Planning B: Urban Analytics and City Science, 50*(1), 7-23.

Bullock, C., Brereton, F., Bailey, S. (2017). The economic contribution of public bike-share to the sustainability and efficient functioning of cities. *Sustainable Cities and Society 28,* 76–87. https://doi.org/10.1016/j.scs.2016.08.024.

Chen, W., Chen, X., Chen, J., Cheng, L. (2022). What factors influence ridership of station-based bike sharing and free-floating bike sharing at rail transit stations? *International Journal of Sustainable Transportation, 16*(4), 357-373. https://doi.org/10.1080/15568318.2021.1872121.

Du, Y., Deng, F., Liao, F. (2019). A model framework for discovering the spatio-temporal usage patterns of public free-floating bike-sharing system. *Transportation Research Part C: Emerging Technologies, 103,* 39-55. https://doi.org/10.1016/j.trc.2019.04.006

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

Guzel, D., Altintasi, O., Korkut, S.O. (2025). Assessment of weather-driven travel behavior on a small-scale docked bike-sharing system usage. *Travel Behaviour and Society. 38* https://doi.org/10.1016/j.tbs.2024.100927.

Li, W., Chen, S., Dong, J., & Wu, J. (2021). Exploring the spatial variations of transfer distances between dockless bike-sharing systems and metros. *J Transp Geogr., 92, 103032.*

Liu, H. C., Lin, J. J. (2019). Associations of built environments with spatiotemporal patterns of public bicycle use. *Journal of Transport Geography, 74,* 299-312. https://doi.org/10.1016/j.jtrangeo.2018.12.010.

Ma, X., Cao, R., Jin, Y. (2019). Spatiotemporal clustering analysis of bicycle sharing system with data mining approach. *Information (Switzerland), 10*(5). https://doi.org/10.3390/info10050163.

Müllner, D. (2011). Modern hierarchical, agglomerative clustering algorithms. *arXiv preprint arXiv*:1109.2378.

Pekdemir, M. İ., Altintasi O., & Özen, M. (2021). Bisiklet Paylaşım Sistemi Kullanıcıların Mevsimsel Farklılıklarının İncelenmesi: Bisim İzmir Örneği. *2nd International Conference on Intelligent Transportation Systems, BANU-ITSC'21* October 22-24, 2021 Bandırma, Turkey.

Pekdemir, M.I., Altintasi, O., Ozen, M. (2024). Assessing the impact of public transportation, bicycle infrastructure, and land use parameters on a small-scale bikesharing system: a case study of Izmir, Türkiye. *Sustain. Cities Soc. 101.* https://doi.org/10.1016/j.scs.2023.105085

Saplıoğlu, M., & Aydın, M. M. (2018). Choosing safe and suitable bicycle routes to integrate cycling and public transport systems. *Journal of Transport & Health*, *10*, 236-252.

Saplioglu, M., & Günay, E. Y. (2016). Investigating the effective parameters of safe bicycle route by using a survey study. *Sigma, 7*(1), 89-96.

Shalizi, C. (2009). *Distances between clustering, hierarchical clustering*. Retrieved November 20, 2024, from https://www.stat.cmu.edu/~cshalizi/350/lectures/08/lecture-08.pdf

Yemişçioğlu, Ş., Çivici, T., & Yıldız, Y. (2024). Analiz araçları yardımıyla sürdürülebilir bisiklet yolları seçimi üzerine bir çalışma. *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi, 39*(4), 2011-2022.

Zhou, X. (2015). Understanding spatiotemporal patterns of biking behavior by analyzing massive bike sharing data in Chicago *PLoS ONE*, *10*(10). https://doi.org/10.1371/journal.pone.0137922.

Zhu, Y., Diao, M. (2020). Understanding the spatiotemporal patterns of public bicycle usage: A case study of Hangzhou, China. International *Journal of Sustainable Transportation,* *14*(3), 163-176. https://doi.org/10.1080/15568318.2018.1538400