Coursework submission for Transport Data Science (TRAN5340M)

Shared e-Bikes in London: Analysing Usage Trends and Patterns

Student 201798764

1 Introduction

Shared bicycle systems globally have revolutionized urban mobility, significantly altering how city residents commute and connect with public transport systems. In London, the success of shared bikes, particularly with the integration of e-bikes, has been noteworthy. To sustain and build on this progress, Transport for London (TfL) launched its first docked e-bikes in October 2022. This initiative caters to the growing demand for efficient and sustainable transportation options within the metropolitan area, introducing a novel dimension to the city's shared mobility landscape.

This research investigates the adoption and impact of shared e-bikes in London, focusing on usage patterns, and spatial and temporal distribution in high-demand areas like commercial hubs and commuting hotspots. By analyzing these elements, the study aims to identify trends that can guide the strategic placement and promotion of e-bike stations to enhance their accessibility and operational efficiency. The report, "Lime in London (2023)," underscores the significant benefits of shared e-bike services in reducing traffic congestion and improving urban mobility. It also provides detailed recommendations for optimizing bike-share operations, especially during peak commuting times and in high-traffic areas. This research is particularly timely as it aligns with TfL's 2023 initiatives that seek to increase cycling trips significantly, reflecting a strategic shift towards more sustainable urban transport solutions.

1.1 Scope

The primary goal of this report is to analyze Santander shared e-bike usage in London for 2023, identifying usage trends and developing strategies to enhance the bike-sharing program's operations. The analysis focuses on evaluating extensive trip data and docking station patterns, as well as understanding user behaviors across various times and locations within the city. By leveraging detailed datasets of trip records and station information, this report aims to provide actionable insights to support policy decisions and operational enhancements. The findings are presented in sections covering data understanding and preprocessing (Section 2), thorough temporal and geospatial analyses (Sections 3 and 4), predictive modeling to forecast usage trends (Section 5), and a detailed examination of usage patterns at the highest usage e-bike station (Section 6). This will culminate in strategic policy recommendations (Section 7) designed to improve the efficiency and accessibility of London's e-bike sharing system.

1.2 Area of Study

The geographical scope of this study is focused on the city of London, with particular emphasis on the regions served by the Santander Cycle Scheme. The analysis covers various boroughs within London, examining both central and peripheral regions to understand how e-bike usage varies across different urban densities and neighborhood characteristics. Special attention is given to areas with high commuter traffic and regions that have shown significant uptake of shared mobility solutions.

1.3 Datasets

The primary datasets employed in this study were the shared bike trip (OD) data, accessible from the public TFL data repository. TFL periodically updates the shared bike trip data, starting from 2014, with the latest available records extending up to December 2023. Dock location coordinates were extracted using an API. Subsequently, these two datasets were merged to facilitate further geographical analysis. The main dataset was compiled by consolidating 38 CSV files, encompassing all trip data from 792 bike dock stations in London throughout 2023.

2 Data Understanding and Pre-Processing

Before analyzing the data, it is essential to thoroughly understand the raw dataset and prepare it appropriately for further analysis. This section focuses on comprehending the structure and content of the original dataset.

2.1 Understanding the Datasets

The raw dataset comprises 8,563,356 entries across 11 columns. Utilizing the `info()` function helps in comprehending the column names and their respective data types. Each entry represents a singular trip, encompassing details such as starting time, end time, starting station, end station, bike number, bike model (classic or e-bike), and total travel duration.

2.2 Data Preparation

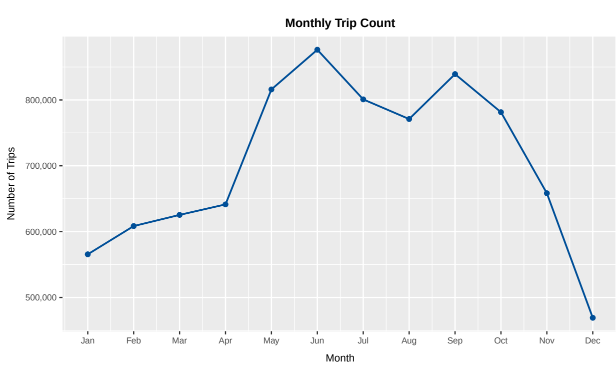
After combining and loading the dataset, the data type of the 'date' column needed to be converted from an object to datetime type for further analysis. Post-conversion, new columns such as "Month," "Weekday," "Date," and "Hour" were created using the .dt accessor to categorize trips. This research excludes trips spanning from December 31, 2022, to 2023, retaining only the data within 2023. Trips that start and end at the same station with a total travel time of less than one minute, and trips that exceed 24 hours in duration, are also removed from the dataset. Additionally, the total duration in milliseconds was converted to minutes for better comprehension.

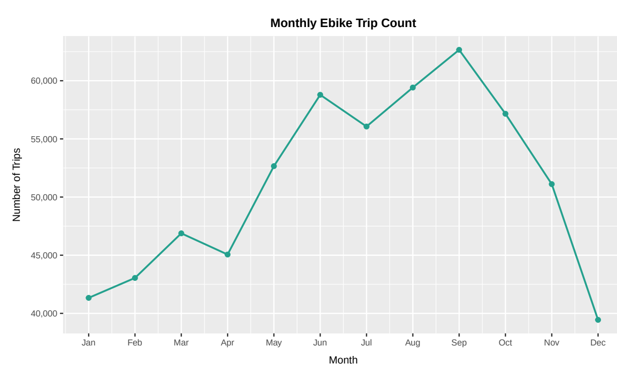
2.3 Exploratory Data Analysis

A key aspect of exploratory data analysis (EDA) is to extract summary statistics and visualize general trends from the dataset. The describe() function is employed to summarize key numerical variables, such as "duration time (mins)." Additionally, we explore the overall proportion of e-bike usage and its variation across different months.

2.3.1 Monthly Ride Trend

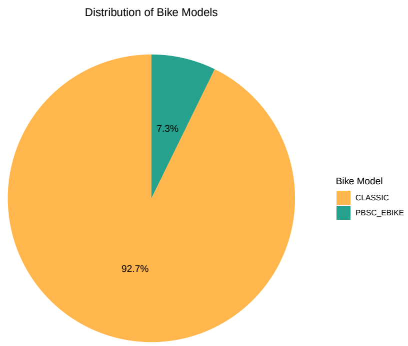
The usage of e-bikes follows the overall trend observed. June and September are the peak months for rides, reflecting the highest popularity. Conversely, the winter months exhibit the lowest riding activity, indicating a significant seasonal impact on usage patterns.

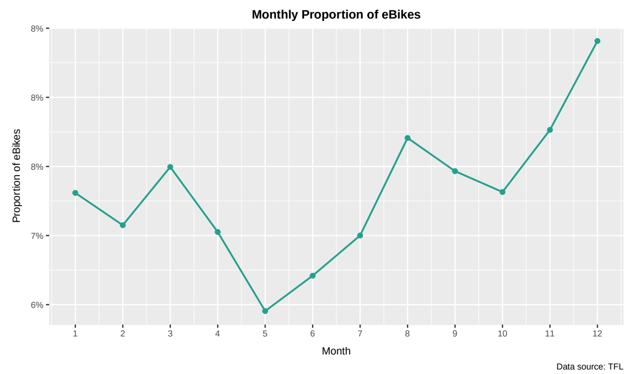




2.3.2 Proportion Analysis

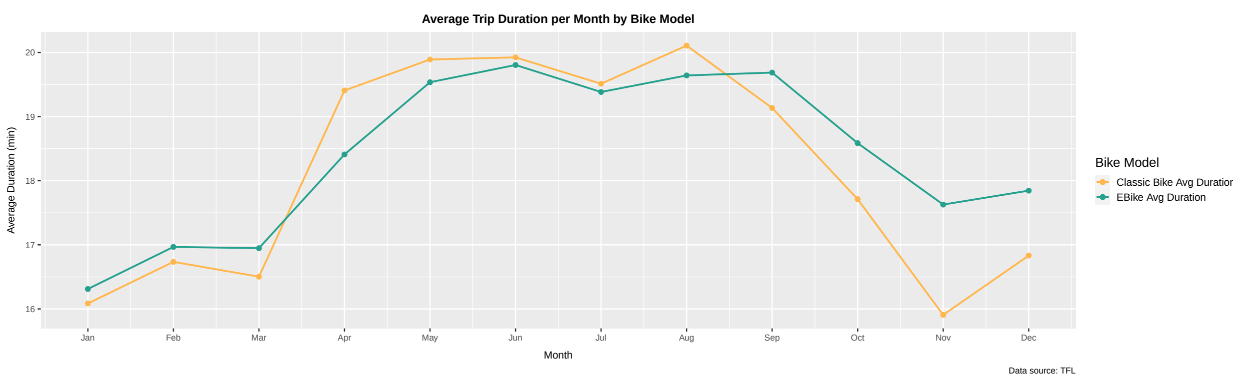
In the overall bike usage, e-bike trips constitute approximately 7% of all rides, with the proportion fluctuating between 6.5% and 8.5% throughout the year. Notably, there is an increase in the percentage of e-bike usage during the winter months compared to the summer. This trend suggests that people may prefer using e-bikes in colder weather, potentially due to the added ease and comfort provided by electric assistance.





2.3.3 Average Monthly Ride Duration

Overall, both classic bikes and e-bikes experience longer ride durations in the summer, averaging around 19.5 minutes. Notably, from April to August, the duration of rides on classic bikes exceeds that of e-bikes. However, from September to March, e-bikes show longer ride durations. This trend suggests that in colder weather, people tend to use e-bikes for longer trips.



3 The 'When': Temporal Distribution of eBike Usage

Riding patterns can be influenced by riders' daily routines and are likely to vary throughout the year. This section focuses on the temporal distribution of both classic bikes and e-bikes, aiming to explore and highlight their commonalities and differences. By examining these usage patterns, we seek to understand how factors such as time of day and seasonality impact bike usage.

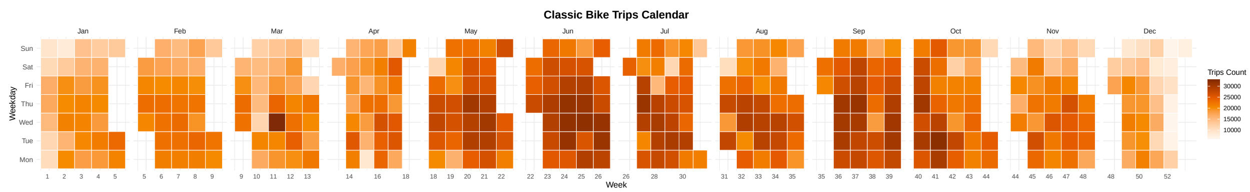
3.1 Methodology

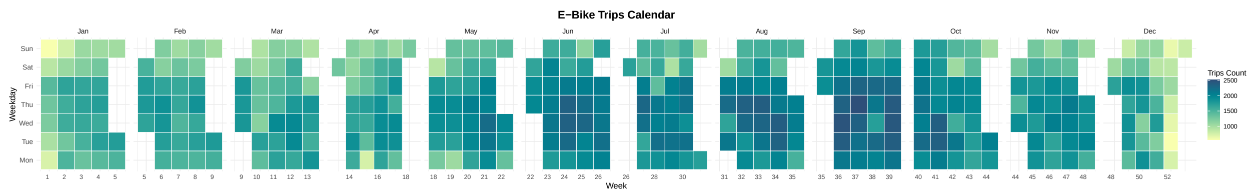
To more effectively compare temporal distributions, heatmaps are utilized to highlight bike usage patterns. The dataset is divided into two categories: classic bikes and e-bikes, which helps visualise both annual and weekly patterns. For year-long trends, calplot() is employed, while grouping by 'weekday' and 'hour' facilitates the creation of hourly data for analysing weekly and hourly patterns. These visualizations make it easier to observe variations in bike usage over different timescales.

3.2 Results and Discussions

3.2.1 Annual and Weekly Pattern

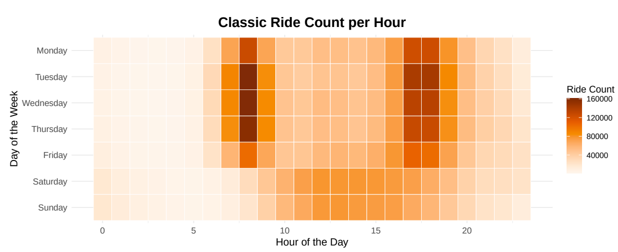
Both bike modes exhibit increased popularity during the summer, with classic bikes peaking in June and e-bikes peaking in August and September. Additionally, both types are more frequently used on weekdays compared to weekends. Notably, the usage rate of e-bikes declines significantly during weekends, suggesting that e-bikes are predominantly used for commuting purposes.

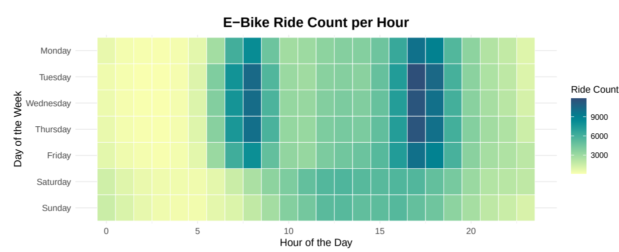




3.2.2 Weekly and Hourly Pattern

Both types of bikes exhibit higher usage during weekdays, particularly during commuting hours. Tuesday, Wednesday, and Thursday are the most popular days of the week, with peak usage times aligning with typical commuting periods. Specifically, the classic bikes show the highest usage during the morning commute at 8 AM, while e-bikes are most frequently used during the evening commute between 5 and 6 PM.





4 The 'Where': Geospatial Distribution of eBike Usage

Following the exploration of the temporal distribution of e-bike usage, this section delves into a geographical analysis. It identifies popular routes, key starting points, and frequent end stations to enhance understanding of potential user profiles and their travel purposes.

4.1 Methodology

The initial step involves integrating bike dock latitude and longitude data from the TfL API with the main trip dataset using station names to facilitate the merge. After incorporating the latitude and longitude information, each station is visualized on a map. The dataset is divided into two categories: classic bikes and e-bikes, to discern usage patterns between the two modes. To identify popular routes, the analysis employs a group_by() function to aggregate data by "Start Station" and "End Station," extracting the top 10 most popular routes. On the map, the top 5000 routes for classic bikes and the top 2000 routes for e-bikes are plotted, with the thickness of each route line indicating its popularity. Routes that begin and end at the same station are visualized as circles, with the size of the circle reflecting the frequency of such trips. The results are also saved as an HTML file for better viewing.

4.2 Results and Discussions

4.2.1 Popular Route

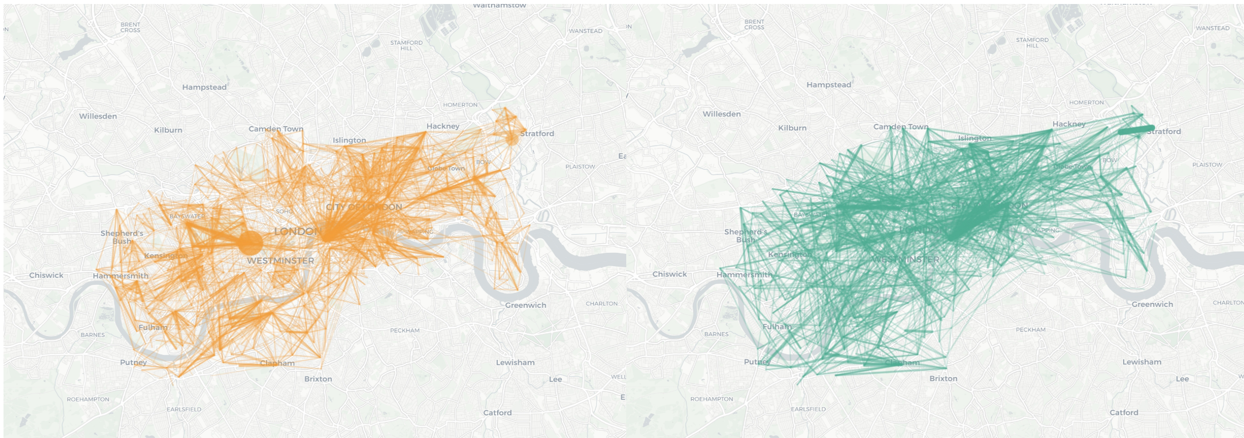
The map visualization of popular routes for classic and e-bikes reveals distinct patterns. Classic bikes predominantly feature circular routes that begin and end at the same location, primarily around parks and green spaces like Hyde Park, Queen Elizabeth Olympic Park, and Kensington Gardens. This suggests that classic bike usage is heavily oriented towards leisure activities within these areas. Conversely, e-bike routes display a more dispersed and interconnecting pattern, indicating a strong commuting function. For instance, the most frequented routes between Stratford Station and Monier Road highlight a significant demand for e-bikes, suggesting their utilization for practical commuting purposes. Additional routes, such as those from Wormwood Street to Waterloo Station and Liverpool Road to Eagle Wharf Road, primarily connect commercial areas, further supporting the commuting usage of e-bikes. These observations underscore the differing roles that classic bikes and e-bikes play in urban mobility, with classic bikes serving more recreational purposes and e-bikes fulfilling daily commute needs. This analysis assists in understanding the specific needs and behaviors of different user groups, providing valuable insights for urban planning and transportation management.

Top 5 Popular Classic Bike Routes

Rank	Start Stations	End Station	Counts
1	Hyde Park Corner, Hyde Park	Hyde Park Corner, Hyde Park	17,906
2	Podium, Queen Elizabeth Olympic Park	Podium, Queen Elizabeth Olympic Park	9,051
3	Albert Gate, Hyde Park	Albert Gate, Hyde Park	7,309
4	Black Lion Gate, Kensington Gardens	Black Lion Gate, Kensington Gardens	6,972
5	Triangle Car Park, Hyde Park	Triangle Car Park, Hyde Park	6,157

Top 5 Popular E-bike Routes

Rank	Start Stations	End Station	Counts
1	Stratford Station, Stratford	Monier Road, Hackney Wick	291
2	London Street, Paddington	Little Argyll Street, West End	260
3	Monier Road, Hackney Wick	Stratford Station, Stratford	222
4	Wormwood Street, Liverpool Street	Waterloo Station 3, Waterloo	193
5	Podium, Queen Elizabeth Olympic Park	Podium, Queen Elizabeth Olympic Park	176



4.2.2 Popular Stations

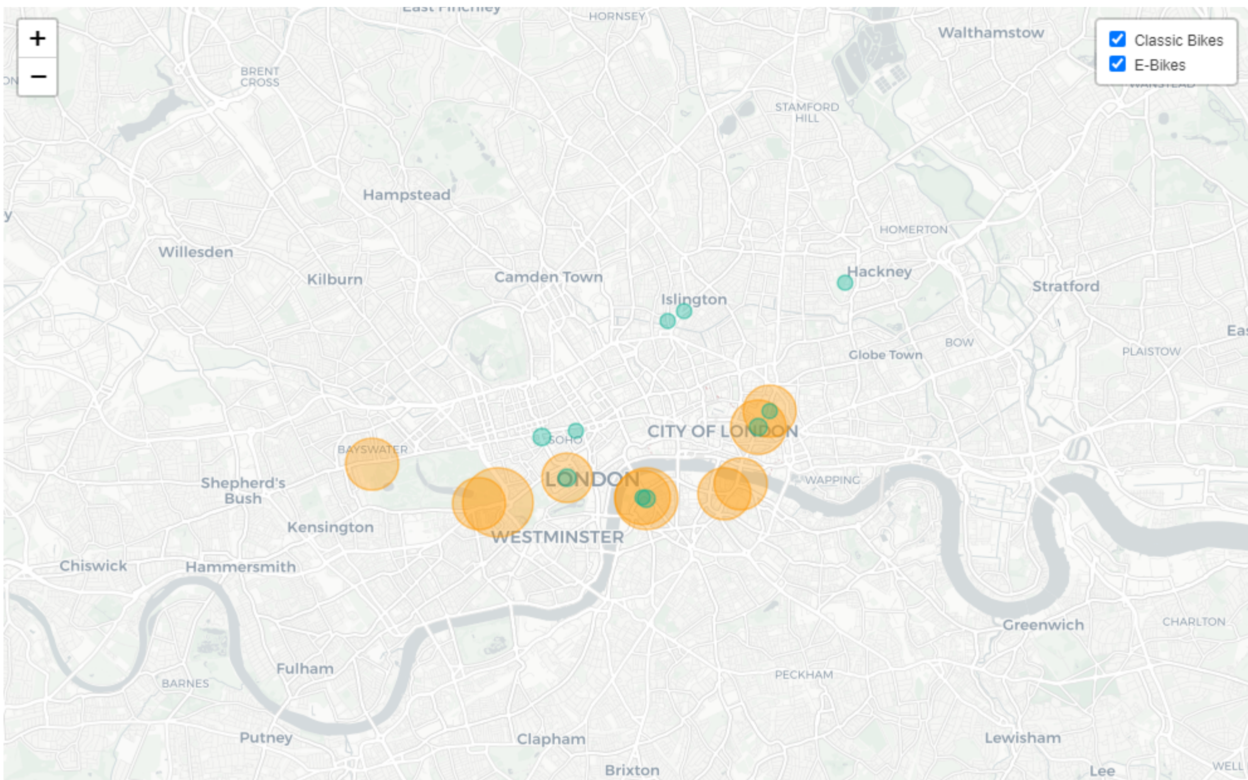
Stations located near major transport hubs such as Waterloo and Liverpool Street Station exhibit high usage rates due to their extensive transportation connections. Classic bikes demonstrate a more recreational usage pattern, with Hyde Park being the most frequented location. St. James's Square records the highest usage of e-bikes. Stations situated in Soho, Islington, and Hackney also show significant e-bike usage, indicating a distinct preference for e-bikes in these areas.

Top 10 Bike Stations Comparison

Rank	Classic Bike Station	Classic Counts	E-Bike Station	E-Bike Counts
1	Hyde Park Corner, Hyde Park	108,954	St. James's Square, St. James's	7,098
2	Waterloo Station 3, Waterloo	92,052	Waterloo Station 3, Waterloo	6,937
3	Wormwood Street, Liverpool Street	70,025	Wormwood Street, Liverpool Street	6,551
4	Waterloo Station 1, Waterloo	69,451	Little Argyll Street, West End	6,393
5	Hop Exchange, The Borough	65,916	Islington Green, Angel	5,955
6	Brushfield Street, Liverpool Street	65,775	Soho Square, Soho	5,220
7	Black Lion Gate, Kensington Gardens	65,573	London Fields, Hackney Central	4,890
8	Duke Street Hill, London Bridge	62,888	Brushfield Street, Liverpool Street	4,890
9	Albert Gate, Hyde Park	62,178	Liverpool Road (N1 Centre), Angel	4,701
10	St. James's Square, St. James's	58,293	Waterloo Station 1, Waterloo	4,686

Table 4: Top 10 Bike Stations Comparison

Rank	Classic_Bike_Station	Classic_Counts	E_Bike_Station	E_Bike_Counts
1	Hyde Park Corner, Hyde Park	108954	St. James's Square, St. James's	7098
2	Waterloo Station 3, Waterloo	92052	Waterloo Station 3, Waterloo	6937
3	Wormwood Street, Liverpool Street	70025	Wormwood Street, Liverpool Street	6551
4	Waterloo Station 1, Waterloo	69451	Little Argyll Street, West End	6393
5	Hop Exchange, The Borough	65916	Islington Green, Angel	5955
6	Brushfield Street, Liverpool Street	65775	Soho Square, Soho	5220
7	Black Lion Gate, Kensington Gardens	65573	London Fields, Hackney Central	4890
8	Duke Street Hill, London Bridge	62888	Brushfield Street, Liverpool Street	4890
9	Albert Gate, Hyde Park	62178	Liverpool Road (N1 Centre), Angel	4701
10	St. James's Square, St. James's	58293	Waterloo Station 1, Waterloo	4686



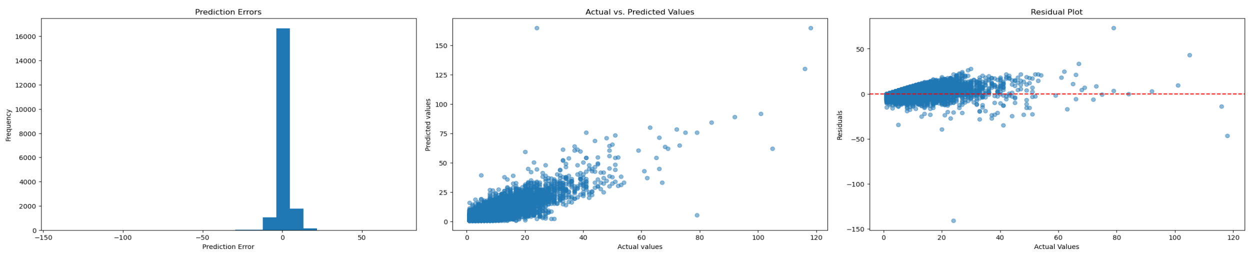
5 The 'How': Predicting Usage of Specific Stations by Weekday and Time

5.1 Methodology

To predict ebike usage across various stations on different weekdays and times, only ebike trip data was selected for analysis. The data was organized using the groupby() function to group "Start Station," "Weekday," and "Hour" into one row, resulting in 98,375 entries. The Random Forest algorithm was employed due to its robustness in handling multiple features without the risk of overfitting. The dataset was split in an 80:20 ratio for training and testing, respectively.

5.2 Results and Discussions

The final model achieved a Mean Squared Error (MSE) of 14.66, which is relatively low given the total number of entries. The residual plot shows that most data points are close to the y=0 line, indicating that the model's predictions are generally accurate. However, inaccuracies tend to increase with larger usage values, although most predictions remain reliable. Using this prediction model is straightforward: simply input the station name, weekday, and time, and the model will generate the predicted number of e-bike uses. For example, if you want to know the demand for LSBU (Borough Road) at 19:00 on Thursday, the output would be: "Predicted demand for LSBU (Borough Road), Elephant & Castle on Thursday at 19:00 is 6.45."



6 A Case Study of station "St. James's Square, St. James's"

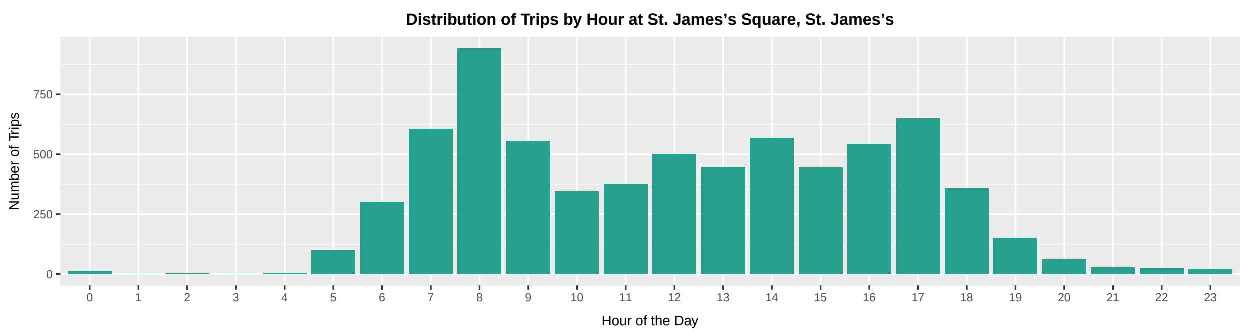
St. James's Square is the station with the highest e-bike usage. This section examines the travel patterns in this area.

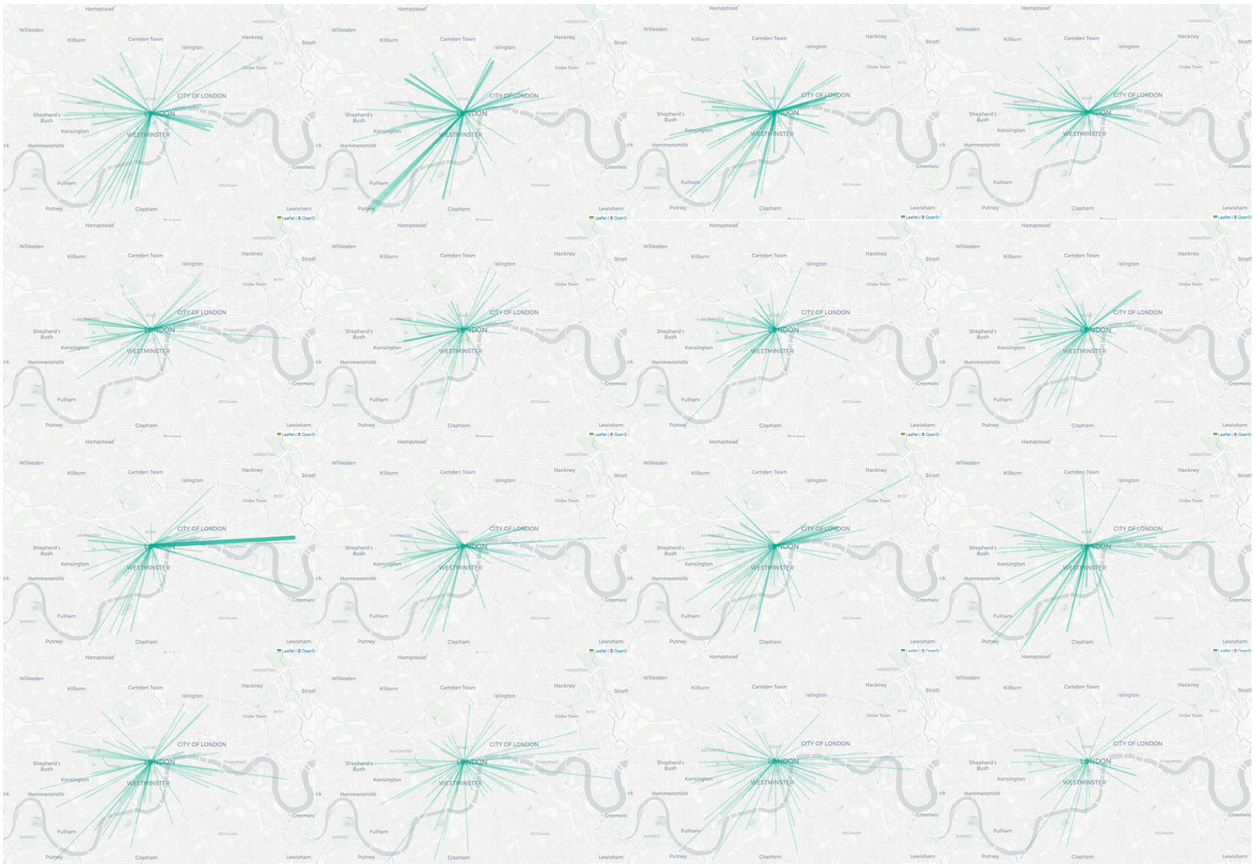
6.1 Methodology

The trip data, which includes trips that start or end at the station "St. James's Square, St. James's" and are made using e-bikes, was filtered from the main dataset. The plots below show the trips from 6am to 9pm.

6.2 Results and Discussions

The peak time is at 8am, with a high demand for trips from Putney. This location also shows significant demand after work hours. Notably, at 2pm, there is a high demand for trips between Chrisp Street Market, Poplar. More e-bikes should be considered for these locations to meet the demand.





7 Policy Recommendation

Based on the analyses conducted above, it is evident that there are notable differences between the usage patterns of classic bikes and e-bikes. In light of these distinctions, we propose strategic recommendations to promote e-bike stations effectively. Firstly, stations with high pedestrian traffic, particularly those situated in commercial and commuting-centric areas, are more likely to exhibit significant demand for bicycles. Therefore, prioritizing the placement of e-bike stations in these locations could capitalize on existing traffic flows and enhance accessibility for commuters and urban dwellers. Secondly, it is imperative to consider the temporal dynamics of bike usage, especially during peak commuting hours. Increasing the availability of e-bikes during these periods, particularly in commercial areas where commuters heavily rely on alternative transportation modes, could stimulate e-bike adoption and usage. Additionally, ensuring that e-bikes are adequately charged and maintained, particularly during peak hours, is essential to guarantee a seamless user experience. Stations located in commercial hubs and residential areas should be equipped with sufficient charging infrastructure and capacity to meet the demands of users returning e-bikes after their commute. By implementing these recommendations, transportation authorities and bike-share operators can effectively leverage the unique advantages of e-bikes and tailor their services to meet the evolving needs of urban commuters and residents. This strategic approach not only promotes sustainable transportation alternatives but also enhances the overall accessibility and usability of bike-share systems in urban environments.

8 Limitations of Analyses

Availability of e-Bikes and Parking: The current dataset does not provide information on the availability of e-bikes for rent or the adequacy of parking facilities at each station. This limitation means that fluctuations in usage patterns may be influenced by the unavailability of e-bikes or insufficient parking, rather than a lack of user demand. Instances where potential users have opted not to use e-bikes could mistakenly be interpreted as low demand rather than being attributed to these logistical constraints.

Assumptions About Trip Purposes: The purpose of each trip has been inferred based on the time of the trip and the locations of the start and end stations. This assumption might not accurately reflect the true intentions of the users. For example, a trip taken during typical commuting hours might be assumed to be for commuting purposes, but it could also be for other reasons such as leisure or personal errands. This assumption could skew the understanding of why different demographics choose to use or avoid using e-bikes, potentially leading to misinformed recommendations.

9 Conclusions

E-bike sharing has proven to enhance urban mobility. While both classic bikes and e-bikes are components of shared bike systems, this research clearly demonstrates distinct usage patterns between them. A detailed analysis of these patterns within London's Santander Cycle Scheme has revealed significant trends and behaviors in urban mobility. The findings indicate that although overall bike usage is seasonally affected, the proportion of e-bike usage increases during the colder months. This suggests a preference for e-bikes, attributed to their electric assist feature, which provides an advantage in less favorable weather conditions. Additionally, e-bike usage is predominantly concentrated during peak commuting times, particularly during off-work hours, and is geographically focused around business districts, major transit hubs, and residential areas such as Stratford, which also shows high demand. Since the implementation of e-bikes in October 2022, the collected trip data through December 2023 has been substantial and has enabled the initial development of a random forest model. This model captures usage patterns at specific stations, on particular weekdays and times, with a low error rate. It can assist Transport for London (TfL) in identifying which stations are likely to experience high demand for e-bikes, thereby optimizing resource allocation and enhancing service efficiency.

10 Reference

Lovelace, R., Beecham, R., Heinen, E., Vidal Tortosa, E., Yang, Y., Slade, C., Roberts, A., 2020. Is the London Cycle Hire Scheme becoming more inclusive? An evaluation of the shifting spatial distribution of uptake based on 70 million trips. Transportation Research Part A: Policy and Practice, 140, pp.1–15. Available at: https://doi.org/10.1016/j.tra.2020.07.017 [Accessed 30 Apr 2024].

Bishop, J., Hauru, M., Nanni, F. and Rangel Smith, C., 2023. Unveiling London's mobility patterns with Boris Bikes. [online] The Alan Turing Institute. Available at: https://alan-turing-institute.github.io/TuringDataStories/stories/2022-10-06-Boris-Bikes/2023-07-07-Boris-Bikes.html [Accessed 15 Apr 2024].

Transport for London (TfL) (2023) TfL sets out vision to further boost cycling by making it more diverse than ever. 15 June. Available at: https://www.tfl.gov.uk (Accessed: 20 May 2024).

Lime in London: Assessing the benefits of shared e-bike services and recommendations for future regulation (2023) July. Available at: https://uk.steergroup.com/sites/default/files/2023-07/Steer_Lime_Report.pdf (Accessed: 22 May 2024).