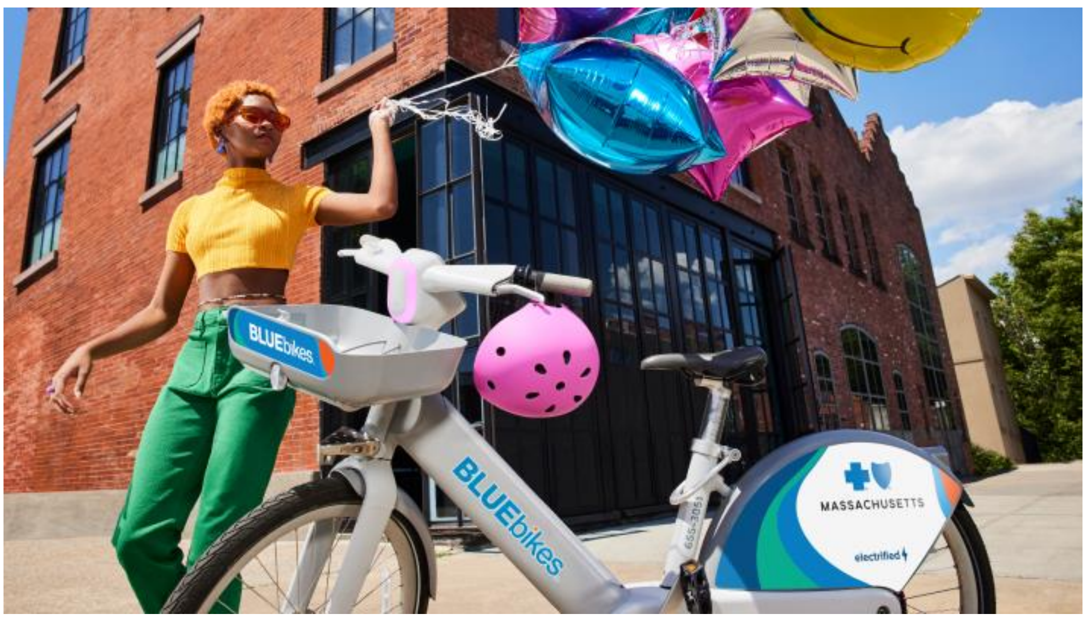
Northeastern University
Mechanical and Industrial Engineering Department
IE 6200: Engineering Probability and Statistics
Prof. Rehab Ali
Section 02 - Fall 2024

Impact of Electric Bikes on Urban Mobility in Boston

Final Report



Group 3

Nathan David Obeng-Amoako

Zeeshan Ghudusab Shaikh

Dorian Chirackal

Sailesh Ananthanarayanan

Alisha Joy Alappatt

Problem Statement

The introduction of electric bikes (e-bikes) into Boston's BlueBikes system in December 2023 marked a pivotal moment in the City's urban transportation. As cities strive to enhance mobility while reducing traffic congestion and carbon emissions, e-bikes provide a promising alternative to traditional bicycles and motor vehicles. However, a critical understanding of how these e-bikes influence trip duration, usage patterns, and overall bike-sharing dynamics is essential for optimizing urban infrastructure and improving service provision. This project aims to analyze the impact of e-bikes compared to traditional bikes on various aspects of urban mobility, focusing on trip duration, user behavior, and station utilization.

Resources

The resource for this research is already available. The dataset for this project will be sourced from the BlueBikes bike-sharing website **[1]**, which provides open access to trip data. It features details on trip duration, start and end times, start and end stations, bike type (traditional or electric), user type, and geographic coordinates of stations. The dataset, including a dictionary describing each feature, is further discussed in a later section.

Objective Functions / Performance Measures

The following objective functions will guide the analysis.

1. **Trip Duration Analysis:** Compare the average trip durations between electric and traditional bikes to determine how much further riders are willing to travel on an e-bike.

2. **Usage Patterns:** Assess differences in usage patterns by time of day, day of the week, and user type (casual vs. member) for both bike types.

3. **Station Popularity:** Identify which stations are most frequently used for electric vs. traditional bikes, providing insights into demand patterns.

Scope

This project will focus on analyzing trip duration for electric vs. traditional bikes, examining the usage patterns based on time of day and day of the week, and identifying popular stations for each bike type. Beyond the scope of this project is an analysis of electric scooter data, as this is not included in the dataset; an examination of external factors that may affect bike usage, such as weather conditions; an evaluation of demographic factors such as gender, and socio-economic status, which would require additional data; and spatial or geographic analysis of the stations.

Expected Result

The project will produce a detailed report and visualizations (graphs and charts) comparing electric and traditional bikes across various metrics. Statistical analyses will provide clear evidence of differences in trip durations, usage patterns, and station popularity. The report will summarize the key findings and provide recommendations for improving the BlueBikes system.

Project Goals

Goal Statement

The primary goal of this project is to conduct a comprehensive analysis of the impact of e-bikes on the BlueBikes system in Boston. This analysis will help clarify how e-bikes affect trip durations, user behaviors, and station usage compared to traditional bikes. Insights derived from this analysis will inform future improvements in bike-sharing systems and contribute to sustainable urban transportation planning.

Measuring Success and Expected Benefits

The success of e-bikes and their benefits can be measured by analyzing trip duration using hypothesis testing and 95% confidence intervals to check for significant changes between classic bikes and e-bikes. An Analysis of Variance (ANOVA) test will assess seasonal differences in average trip duration for both bike types. Also, a proportion test will be conducted to check whether more than a quarter of trips occur during the summer. Heatmaps of hourly and weekly trip counts will help identify peak and non-peak times, providing valuable insights for service optimization. Understanding usage patterns across stations will aid in managing docking capacity effectively, ensuring supply meets demand. Overall, this analysis aims to promote Blue Bike usage in Boston, supporting sustainable mobility and fostering a greener environment.

Data Collection

Data Source

The primary data source is the BlueBikes dataset **[1]**, which includes comprehensive information on bike trips from 2011. The data is aggregated into months. Since e-bikes were first added to the fleet in December 2023, this project will focus on and amalgamate data between **December 2023 and November 2024** (a full 12-month period). A data dictionary describing each of the features contained in the dataset have been summarized in **Table 1**.

Target Population

The target population consists of all users of the BlueBikes system in Boston during the specified one-year period. This includes both casual users (those who utilize single trips or day passes) and members (those who subscribe for annual or monthly access).

Table 1: Data dictionary for the BlueBikes dataset

Feature	Description
ride_id	IDs assigned to each ride
rideable_type	whether bike is electric or non-electric
started_at	when the ride begins (time and date)
ended_at	when the ride ends (time and date)
start_station_name	the name of the station where the ride starts
start_station_id	ID assigned to the station where the ride starts
end_station_name	the name of the station where the ride starts
end_station_id	ID assigned to the station where the ride ends
start_lat	latitude coordinate of the station where the ride starts
start_lng	longitude coordinate of the station where the ride starts
end_lat	latitude coordinate of the station where the ride ends
end_lng	longitude coordinate of the station where the ride ends
member_casual	whether the rider is a casual user (e.g., single trip or day pass) or a member (e.g., annual or monthly subscription).

Data Cleaning and Preprocessing

The BlueBikes data is aggregated monthly, and this analysis focuses on how e-bikes have influenced trip dynamics over a roughly one-year period after their introduction in December 2023. To start, the twelve (12) monthly data files (from **202312-bluebikes-tripdata.csv** to **202411-bluebikes-tripdata.csv**) were combined into a single DataFrame for easier analysis. During this process, column names were changed for clarity; specifically, **member_casual** was renamed to **rider_type**, and **rideable_type** was changed to **bike_type**.

Missing Values

This 12-month dataset contained **4,730,559** trip records prior to data cleaning. Each record represents a trip taken on a Bluebike, detailing the type of bike used, along with trip start and end times. A check for missing values was conducted across each column of the combined dataset. The results, shown in **Table 2**, display the count of missing values for columns with missing data **only**.

Missing values in the **start_station_name** and **start_station_id** columns may indicate instances where bikes were taken from a station without proper check-out. Additionally, missing values for **end_station_name**, **end_station_id**, **end_lat**, and **end_lng** suggest that bikes were not returned to a designated station. This may be due to theft, breakdown, abandonment, or issues with the tracking system. Since these trips represent only a small percentage of the total trip count, they were **removed** from the dataset.

Table 2: Count of missing values in combined dataset

Column with missing value	Count
start_station_name	1,350
start_station_id	1,350
end_station_name	8,345
end_station_id	8,633
end_lat	3,850
end_lng	3,850

Same-Station Trips

The dataset was further filtered to identify trips where the start and end stations were the same. This inspection revealed 165,028 such trips, of which 21,730 (approximately **0.46%** of the total dataset) had a duration of less than five minutes. These short trips were removed, as they are unlikely to represent realistic usage of the bike-sharing service.

Since the dataset did not include trip durations, a new column, **trip_duration_minutes**, was created by calculating the absolute difference between the **started_at** and **ended_at** fields for each trip. Short trips between stations in close proximity were not removed, as these could represent legitimate travel. For example, Massachusetts Institute of Technology (MIT) students and staff benefit from highly discounted BlueBikes memberships, which may encourage frequent **free** short rides between lecture halls, often only a few minutes apart. There's no such benefit for Northeastern University students and staff.

Validation of Station Mappings

The **validate_station_mappings** function was implemented to identify inconsistencies between station names and their corresponding IDs in the bike trip data. Several discrepancies were found; for example, the station name "Somerville Hospital" was associated with two different IDs ("S32020" and "S32052"), and the ID "A32046" corresponded to three distinct station names. To verify some of these mappings, the data dictionary from the BlueBikes website **[1]** was referenced. Other discrepancies were minor and generally involved variations in punctuation or spelling. To address these variations, a function called **update_station_names** was created. This function accepts a DataFrame, an old station name, and a new station name as arguments and updates both the start and end station names wherever the old name appears, ensuring uniformity. For instance, "Chestnut Hill Ave. at Ledgemere Road" was updated to "Chestnut Hill Ave at Ledgemere Rd." After applying these updates through an iterative process, the station names were confirmed to be consistently updated in the DataFrame. All identified inconsistencies are listed in **Table A1** in **Appendix A** and were effectively resolved.

Creation of New Columns in the DataFrame

To support a more detailed analysis of bike trip data, several new columns were created in the DataFrame: **start_hour**, **start_day_of_week**, **start_month**, and **season**. The **start_hour** column extracts the hour from the trip's start timestamp, while **start_day_of_week** indicates the day each trip began, allowing for analysis of usage patterns across different days. The **start_month** and **season** columns capture the month and season of each trip, facilitating the examination of seasonal trends. These additions enhance the dataset, enabling more granular insights into bike-sharing patterns and user behavior.

The R code for all data processing steps, including the functions described above, is available in the accompanying R Markdown file (**rides.Rmd**). This document also includes code for other minor preprocessing techniques not covered in this report. After data cleaning, only **76,381** trip records out of the original **4,730,559** (approximately **1.61%**) were removed.

Exploratory Data Analysis

The exploratory data analysis (EDA) of the BlueBikes dataset aims to uncover patterns and trends in bike-sharing usage across the city. This analysis focuses on examining trip durations by user type and bike type, identifying popular start and end stations, determining peak ride times, and observing daily and monthly variations in ridership.

Station Popularity Analysis

The popularity of bike stations was evaluated by counting the number of trips originating from each station. To visualize these popular stations, an interactive map was created using the Folium library (**Figure 1**). Additionally, horizontal bar plots of trip count for both the most and least popular stations are shown in **Figures B1** and **B2** in **Appendix B**. The most popular stations for traditional bikes are concentrated in Cambridge, particularly near MIT, where many students and faculty with annual memberships frequently use the service for short commutes between campus facilities. In contrast, the most popular electric bike stations are in neighboring cities such as Somerville, Watertown, and Brookline, suggesting that these bikes are commonly used for longer commutes to and from Boston's CBD. Their extended range and ease of use make them ideal for such trips. The least popular traditional bike stations are in areas such as Arlington, Medford, Mattapan, Hyde Park, Dorchester, Quincy, Revere, and Everett, possibly due to lower population density, limited station availability, or alternative transportation preferences. For electric bikes, the Peabody stations rank lowest, probably because of the absence of an electric bike supply in that area.

Figure 1: Map of popular bike stations by bike type: top 10 and bottom 10

Hourly and Weekly Distribution of Bike Trips

This analysis examined bike trip distribution by hour of the day and day of the week. The **start_hour** and **start_day_of_week** columns were created by extracting the start hour and weekday from the timestamps. The data was then grouped to calculate total trips for each hour and day combination, resulting in a line plot (**Figure 2**) to visualize these trends. A heatmap illustrating the same trend is provided in **Figure B3** in **Appendix B**. The analysis reveals that trip counts are lower on weekends compared to weekdays (same trend for both bike types), with peak usage times occurring on weekdays around **8:00 A.M.** and **5:00 P.M.**, suggesting that most BlueBikes trips are for commuting purposes during the workweek.

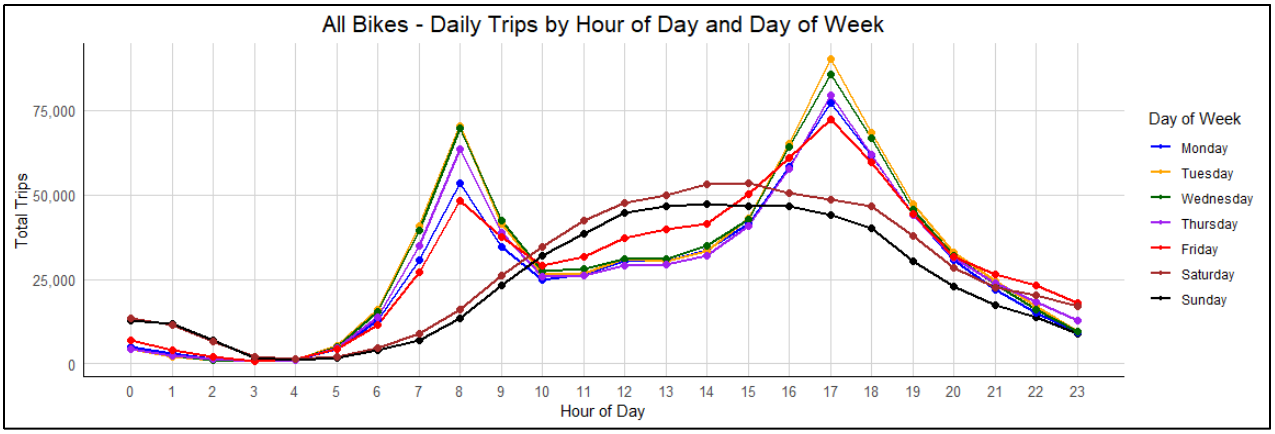


Figure 2: Line plot for total bike trips by hour of day and day of week

Analysis of Bike Usage Patterns by Hour and Bike Type

This analysis examined fluctuations in bike trips throughout the day, segmented by bike type. By organizing the data by start hour and bike type, the total number of trips for both traditional bikes (Bluebikes chooses to call them classic bikes) and electric bikes was calculated across different times of day. The resulting line plot (**Figure 3**) illustrates usage patterns for each bike type and rider type. The data shows that most commuters tend to favor classic bikes, likely due to additional charges associated with electric bikes, as well as subscription plans that generally offer free rides on classic bikes only.

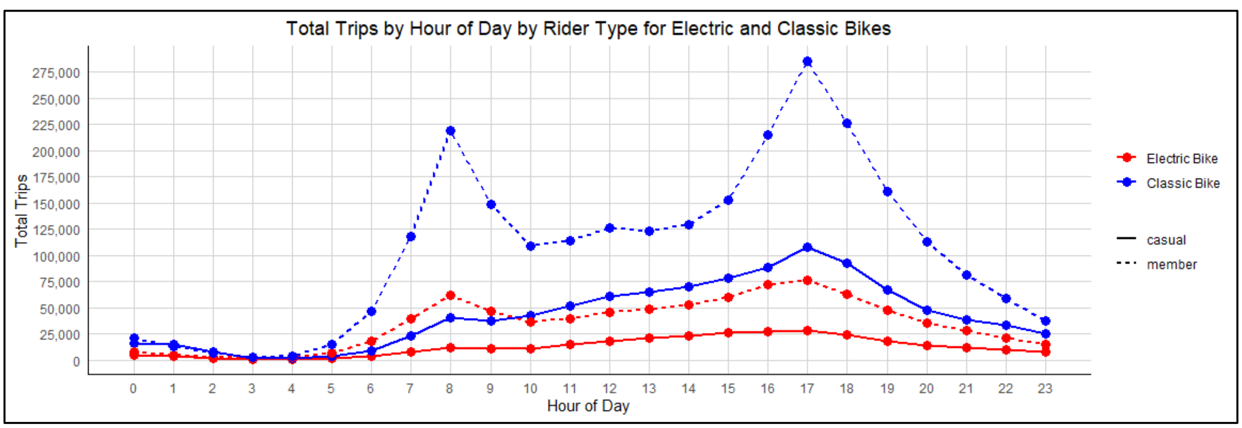


Figure 3: Line plot for total bike trips by hour of day and bike type

Top Station Pairs

In this analysis, the top 10 station pairs were calculated by grouping the data based on start and end station names. Using the **dplyr** package, the total number of trips for each station pair was summarized, sorted in descending order, and the 10 most frequently traveled pairs were selected. The output (**Table 3**) indicates that many of these station pairs are clustered around

MIT. This is likely because a substantial number of MIT students hold bike passes, allowing frequent BlueBikes use without additional costs. Consequently, there is a higher concentration of trips between nearby stations, reflecting student commuting patterns between classes, dorms, and other campus facilities.

Table 3: Top 10 station pairs

A tibble: 10 × 3

start_station_name <chr>	end_station_name <chr>	Count <int>
MIT at Mass Ave / Amherst St	Beacon St at Massachusetts Ave	5148
MIT at Mass Ave / Amherst St	MIT Vassar St	5132
Beacon St at Massachusetts Ave	MIT at Mass Ave / Amherst St	4585
MIT Vassar St	MIT at Mass Ave / Amherst St	4558
MIT Vassar St	MIT Stata Center at Vassar St / Main St	4542
MIT at Mass Ave / Amherst St	Central Square at Mass Ave / Essex St	4339
MIT Vassar St	Ames St at Main St	4326
Central Square at Mass Ave / Essex St	MIT Pacific St at Purrington St	4279
MIT Stata Center at Vassar St / Main St	MIT Vassar St	3944
MIT Pacific St at Purrington St	MIT Stata Center at Vassar St / Main St	3812

1-10 of 10 rows

Analysis of Rider Type by Bike Type

An analysis by rider type and bike type shows that members account for a significantly higher number of trips than casual riders, suggesting that membership provides incentives that encourage frequent use. Additionally, classic bikes are more popular than electric bikes among both rider types. This trend may be influenced by factors such as lower costs and greater availability of classic bikes. These distinctions are illustrated in the stacked bar plot (**Figure 4**).

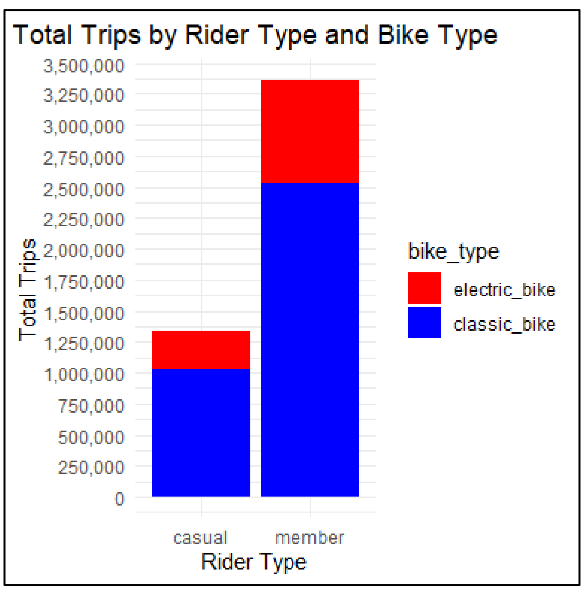


Figure 4: Rider type analysis by bike type

Comparing Trip Duration by Bike Type

Trip duration in minutes was calculated by finding the difference between each trip's start and end timestamps. To identify outliers, any trip duration falling below the first quartile minus 1.5 times the interquartile range (IQR) or above the third quartile plus 1.5 times the IQR was defined as an outlier. After filtering these outliers, a boxplot (**Figure 5**) was generated to visualize trip durations by bike type, providing insights into rider preferences and behaviors over various trip lengths. Interestingly, the average trip duration for classic bikes is higher than that for electric bikes, likely due to the per-minute charges associated with electric bikes, which may encourage shorter trips.

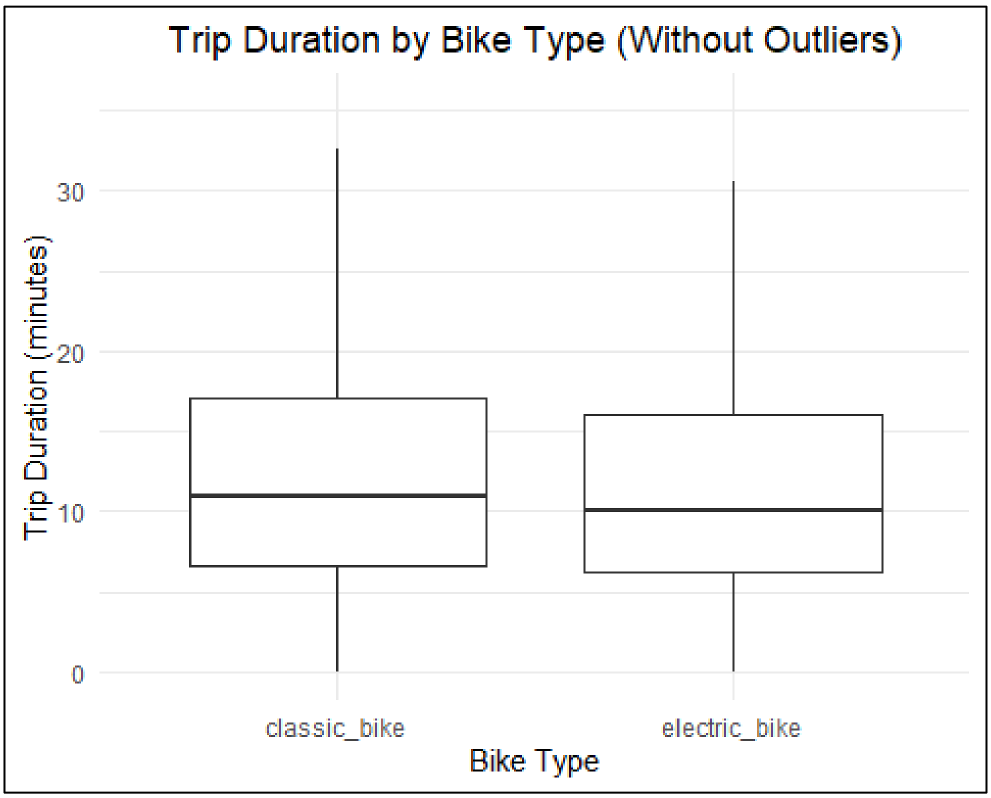


Figure 5: Box plot for trip duration by bike type

Density Distribution of Trip Durations

The density plot in **Figure 6** illustrates the distribution of filtered trip durations, providing a clear view of typical trip lengths. By employing a density estimate, the plot reveals the concentration of trip durations across a range of values. The peak of the plot (roughly 15 minutes) indicates the most common trip duration, demonstrating that many riders complete their journeys within about half an hour. The curve looks positively skewed in that there are trips that are much longer than the median trip duration, making the mean trip duration to be higher than the median.

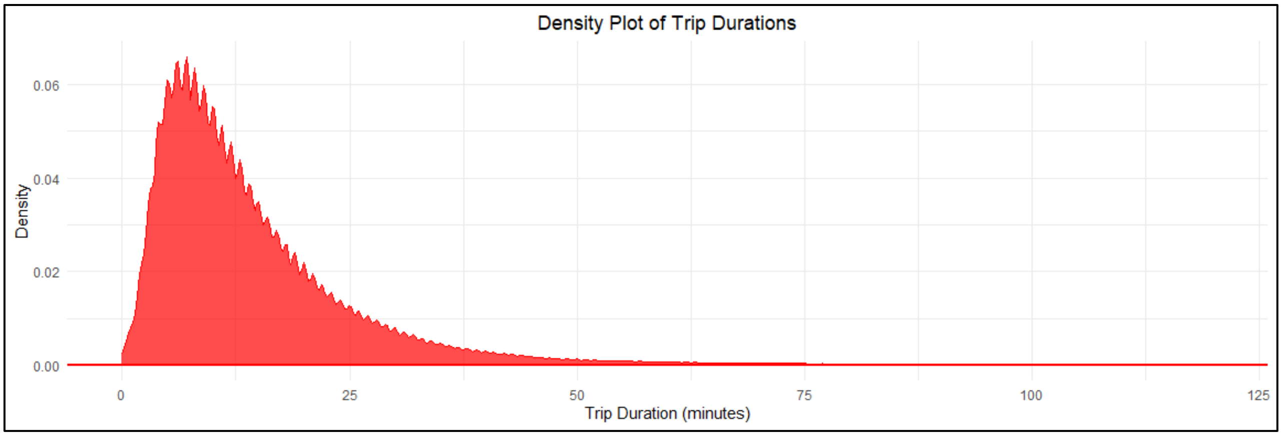


Figure 6: Density plot of trip durations

Seasonal Trends in Bike Trips

This analysis examines bike trips across different seasons by categorizing trips based on their month of occurrence. A custom function was employed to assign each trip to a season: winter (December through February), spring (March through May), summer (June through August), and fall (September through November). The data was then grouped by season to calculate the total number of trips for each. The resulting bar plot (**Figure 7**) visually represents the number of trips by season, illustrating how bike usage varies throughout the year. We see more trips during the warmer seasons and fewer trips during the colder seasons, as expected. **Figure B4** in the Appendix breaks the seasons down into months.

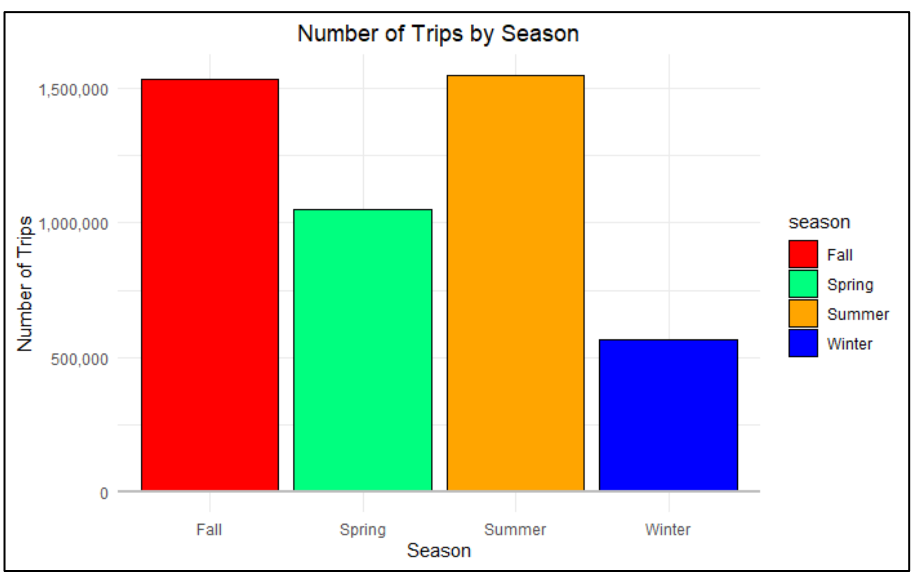


Figure 7: Bar plot showing seasonal trends in bike usage

Statistical Analyses

This section provides an inferential statistical analysis to assess the impact of e-bike integration on BlueBikes ridership patterns in Greater Boston. The analysis focuses on comparing electric and classic bike usage, particularly in terms of trip duration and trip frequency. Various statistical methods were applied, including hypothesis testing, confidence intervals, and Analysis of Variance (ANOVA). The following outlines the results and interpretations of the tests conducted.

1. Hypothesis Testing to Compare Trip Durations Between Bike Types

The first question addressed was whether the introduction of e-bikes significantly altered the average trip duration compared to classic bikes. The hypotheses tested were:

- **Null Hypothesis (H_0):** There is no significant difference in the average trip durations between e-bikes and classic bikes.

$$H_0: \mu_{electric\ bike} = \mu_{classic\ bike}$$

- **Alternative Hypothesis (H_1)**: The average trip duration for e-bikes is different from the average trip duration for traditional bikes.

$$H_1: \mu_{electric\ bike} \neq \mu_{classic\ bike}$$

A two-sample t-test was performed to compare the mean trip durations of e-bikes and classic bikes. The test examined whether the difference in mean durations between the two groups was statistically significant.

The analysis yielded a t-statistic of **-80.85** with a corresponding p-value of **approximately zero**. Since the p-value is much smaller than the significance level of 0.05 (which is the standard threshold in most scientific studies), the null hypothesis is **rejected**. This means that there is a statistically significant difference in the average trip durations between e-bikes and classic bikes, suggesting that the introduction of e-bikes has indeed had a measurable impact on trip durations in Greater Boston. The full analysis with all the necessary equations can be found in the accompanying R Markdown file (**rides.Rmd**).

2. Confidence Interval for the Difference in Mean Trip Durations

In addition to the hypothesis test, a **95%** confidence interval was calculated to estimate the range within which the true difference in mean trip durations between e-bikes and classic bikes is likely to lie. The resulting confidence interval for the difference in mean trip durations was **-1.23 minutes** to **-1.17 minutes**. Since this interval does not include zero, it provides further evidence that there is a statistically significant difference in mean trip durations between the two bike types.

At a significance level of **0.05**, the results suggest that e-bikes have an average trip duration between 1.17 and 1.23 minutes shorter than classic bikes. This finding aligns with expectations, as e-bikes allow for higher speeds due to their electric assistance, enabling riders to complete trips more quickly compared to manually pedaling classic bikes. The full analysis with all the necessary equations can be found in the accompanying R Markdown file (**rides.Rmd**).

The plot in **Figure 8** presents the average monthly trip duration for both electric and classic bikes, along with the overall average for all bike types.

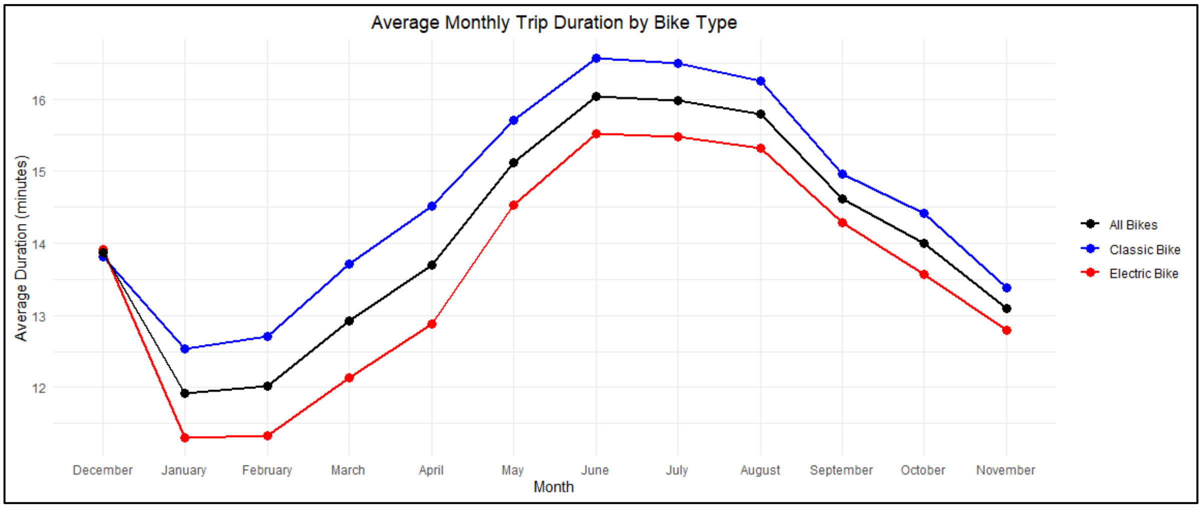


Figure 7: Average monthly trip duration by bike type

3. Impact of Seasonality on Trip Duration for E-Bikes and Traditional Bikes

The next test investigated the impact of seasonality on trip durations for both e-bikes and classic bikes. The aim was to determine whether the average trip duration varied significantly across different seasons—spring, summer, fall, and winter. For each bike type, the following set of hypotheses were formulated:

- **Null Hypothesis (H_0):** There is no significant difference in the average trip duration across seasons.

$$H_0: \mu_{winter} = \mu_{spring} = \mu_{summer} = \mu_{fall}$$

- **Alternative Hypothesis (H_1)**: There is a significant difference in the average trip durations across seasons.

$$H_1: At\ least\ the\ mean\ trip\ duration\ for\ one\ season\ is\ different$$

To examine the impact of seasonality on trip durations, an **Analysis of Variance (ANOVA)** was conducted for both e-bikes and classic bikes, comparing trip durations across the four seasons. The ANOVA results yielded extremely low p-values (close to 0) for both e-bikes and classic bikes, leading to the rejection of the null hypothesis. This indicates that seasonality significantly influences trip durations for both bike types.

The F-statistics for e-bikes (**4,061.92**) and classic bikes (**8,647.42**) were notably large, further supporting the conclusion that trip durations vary significantly across seasons. These results suggest that factors associated with different seasons—such as weather conditions and daylight hours—play a key role in influencing trip durations for both e-bikes and classic bikes.

Interestingly, the F-statistic for classic bikes was higher than that for e-bikes, indicating that the effect of seasonality on trip duration is more pronounced for classic bikes. This could be due to the electric assistance in e-bikes, which may mitigate some of the seasonal challenges, such as slower speeds during colder months, that classic bikes are more susceptible to. Classic bike riders may be more dependent on favorable weather conditions, leading to greater variability in trip durations across seasons.

4. Hypothesis Test for Proportion of Summer Trips

A one-sample t-test was conducted to test whether more than a third of all trips in the analysis period occurred during the summer. The **null hypothesis (H_0)** stated that *the mean proportion of summer trips equals one quarter*, while the **alternative hypothesis (H_1)** suggested that *the mean proportion of summer trips is greater than one quarter*. The calculated t-statistic was **366.1187**, and the p-value was approximately 0. Since the p-value is much smaller than the significance level of 0.05, we **reject** the null hypothesis. Hence, there is strong evidence to suggest that more than a quarter of trips occur during the summer. The full analysis with all the necessary equations can be found in the accompanying R Markdown file (**rides.Rmd**).

Results

1. Trip Duration

- The average duration of e-bike trips (13.84 minutes) was shorter than that of traditional bikes (15.00 minutes). Statistical tests, including a two-sample t-test, confirmed this difference as highly significant ($t = -80.85$, $p < 0.05$).

- A 95% confidence interval of (-1.23, -1.17) minutes indicated that e-bike trips were consistently shorter. The primary driver of this finding is likely the higher cost structure for e-bikes, with per-minute charges discouraging extended use, and higher speed of e-bikes making a trip on an e-bike to be completed in a shorter time, as compared to a classic bike.

2. Seasonal Variations

- A significant seasonal effect was observed, with ridership peaking in the warmer months of summer and fall and declining in winter. ANOVA testing for seasonal differences revealed highly significant results for both bike types (F = 4061.917 for e-bikes; F = 8647.422 for traditional bikes).

- Interestingly, e-bikes demonstrated slightly higher resilience to seasonal declines (evidenced in **Figure B4**). Their appeal in colder months may be attributed to reduced physical effort compared to traditional bikes.

- The proportion test confirmed that more than a quarter of all trips are taken in the summer, making summer the most popular season in terms of Bluebikes ridership.

3. Usage Patterns by Time

- Both bike types experienced peak usage during weekday commuting hours (8:00 AM and 5:00 PM), with lower demand on weekends. These patterns underscore the system's role in facilitating daily commutes.

- When segmented by bike type, traditional bikes maintained higher usage overall, possibly due to their inclusion in most membership plans. E-bikes, while utilized less frequently, showed increased preference for shorter, more specific trips.

4. Station Popularity

- Stations near MIT were consistently the most popular. This was attributed to the high concentration of annual pass holders within the University, many of whom use the system for frequent, short trips between campuses or nearby facilities.

- Conversely, stations in suburban areas, characterized by lower population densities and fewer amenities, experienced minimal demand. This disparity highlights a potential opportunity to optimize station placements or improve accessibility in less frequented areas.

Conclusions

The findings demonstrate the transformative impact of e-bikes on urban mobility within the BlueBikes system. Detailed conclusions and lessons learned include:

1. Strategic Infrastructure Investments

- Popular stations near educational institutions and urban hubs require prioritized resource allocation. Expanding docking capacity and maintenance services in these areas will better accommodate high demand.

- Suburban areas, although underutilized, represent an untapped opportunity. Targeted interventions, such as promotional campaigns or partnerships with local businesses, may increase ridership in these regions.

2. **Adapting to Seasonality**
    - Seasonal fluctuations emphasize the need for adaptive strategies. For example, winter-specific promotions or discounted e-bike rates during colder months could mitigate the decline in usage.

3. **Policy Recommendations**
    - Dynamic pricing models tailored to trip durations and peak demand periods can optimize system usage while balancing operational costs.
    - A focus on expanding e-bike fleets in areas with dense commuter populations or constrained public transit options could further enhance their impact.

Limitations

- The dataset lacked information about rider demographics, weather conditions, and socio-economic factors. These variables could have provided deeper insights into usage patterns and motivations.
- Handling a dataset of nearly four and a half million records required high-performance computational resources, which posed occasional delays during the data processing and analysis phases.

Proposed Next Steps and Future Work

- Future analyses could integrate external datasets, such as weather records and demographic profiles of riders. This would allow for a more nuanced understanding of how weather conditions and socio-economic factors influence ridership behaviors.
- A geospatial analysis of station locations relative to population density, public transit hubs, and key amenities could guide strategic expansion of the BlueBikes network, especially in underserved suburban areas.
- Comparing findings with bike-sharing systems in cities with similar urban layouts and transportation challenges (such as New York City, and the cities of Chicago and Los Angeles) could provide valuable benchmarks and insightful practices across the country.

Collaboration

Copyright

Appendix A – Data Cleaning and Preprocessing

Table A1. Station name-ID mismatches

```
Start Station Names Mapping to Multiple IDs:
Somerville Hospital: S32020, S32052 (Instances: 2)
Tremont St at Court St: A32046, A32058 (Instances: 2)

Start Station IDs Mapping to Multiple Names:
A32046: Tremont St at Court St, Canal St. at Causeway St., Canal St at Causeway St (Instances: 3)
A32058: Tremont St. at Court St., Tremont St at Court St (Instances: 2)
B32038: Chestnut Hill Ave. at Ledgemere Road, Chestnut Hill Ave at Ledgemere Rd (Instances: 2)
c32014: Damrell st at Old Colony Ave, Damrell St at Old Colony Ave (Instances: 2)
C32109: Centre St. at Allandale St., Centre St at Allandale St (Instances: 2)
E32003: Hyde Square - Barbara St at Centre St, Hyde Square - Centre St at Perkins St (Instances: 2)
L32007: Swan Pl. at Minuteman Bikeway, Swan Place at Minuteman Bikeway (Instances: 2)
M32019: CambridgeSide Galleria - CambridgeSide PL at Land Blvd, Cambridgeside Pl at Land Blvd (Instances: 2)
S32052: Summer St at Quincy St, Somerville Hospital (Instances: 2)
V32003: Everett Square (Broadway at Chelsea St), Everett Square (Broadway at Norwood St) (Instances: 2)

End Station Names Mapping to Multiple IDs:
Somerville Hospital: S32020, S32052 (Instances: 2)
Tremont St at Court St: A32046, A32058 (Instances: 2)

End Station IDs Mapping to Multiple Names:
A32046: Tremont St at Court St, Canal St. at Causeway St., Canal St at Causeway St (Instances: 3)
A32058: Tremont St. at Court St., Tremont St at Court St (Instances: 2)
B32038: Chestnut Hill Ave. at Ledgemere Road, Chestnut Hill Ave at Ledgemere Rd (Instances: 2)
c32014: Damrell st at Old Colony Ave, Damrell St at Old Colony Ave (Instances: 2)
C32109: Centre St. at Allandale St., Centre St at Allandale St (Instances: 2)
E32003: Hyde Square - Barbara St at Centre St, Hyde Square - Centre St at Perkins St (Instances: 2)
L32007: Swan Pl. at Minuteman Bikeway, Swan Place at Minuteman Bikeway (Instances: 2)
M32019: CambridgeSide Galleria - CambridgeSide PL at Land Blvd, Cambridgeside Pl at Land Blvd (Instances: 2)
S32052: Summer St at Quincy St, Somerville Hospital (Instances: 2)
V32003: Everett Square (Broadway at Chelsea St), Everett Square (Broadway at Norwood St) (Instances: 2)
```

Appendix B – Exploratory Data Analysis

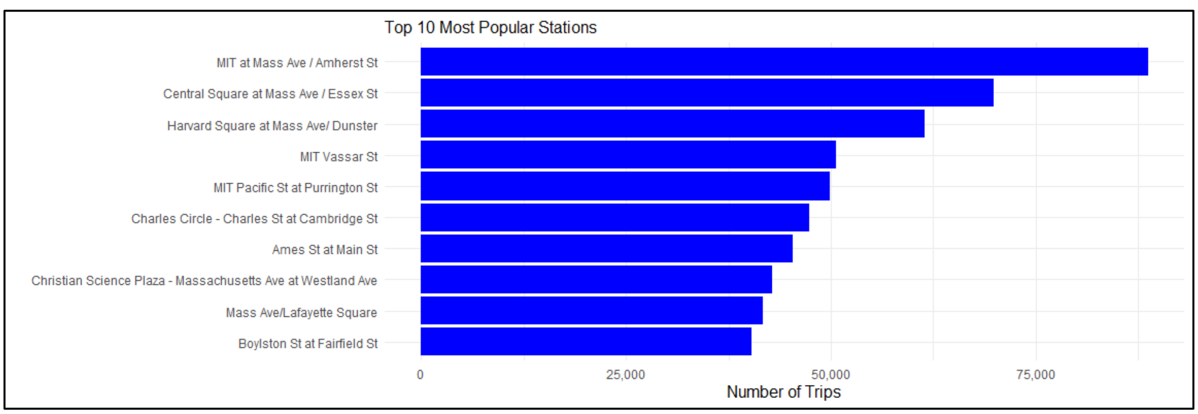


Figure B1. Top 10 most popular stations

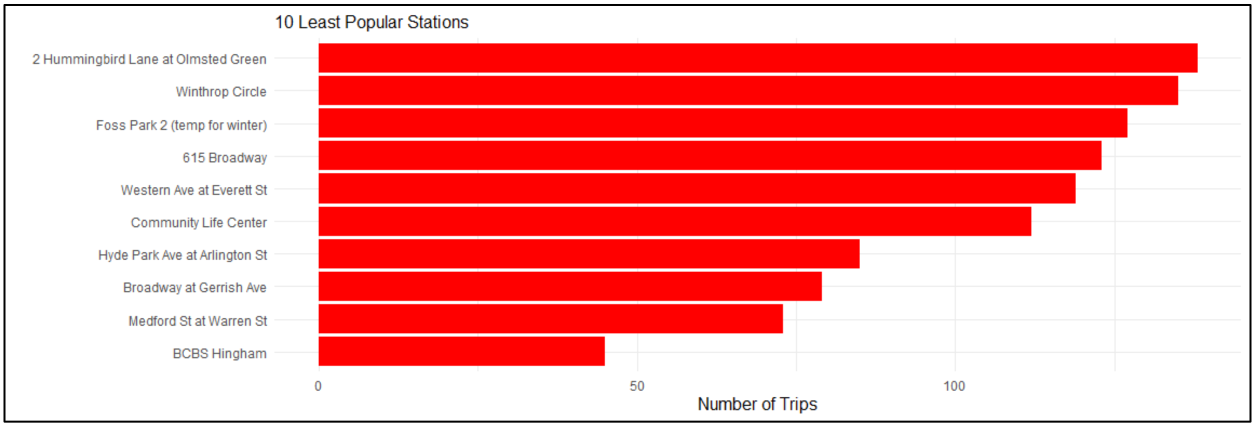


Figure B2. Least 10 most popular stations

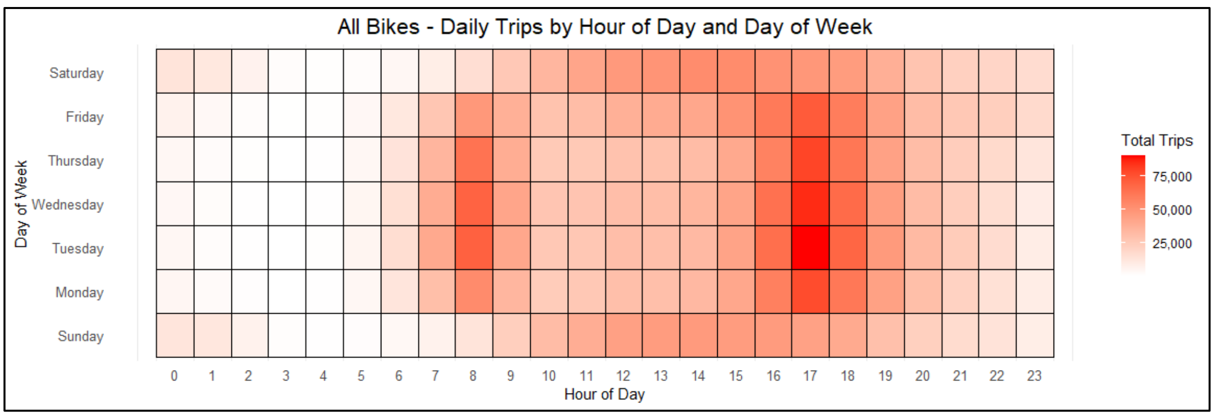


Figure B3. Heatmap for total bike trips by hour of day and day of week

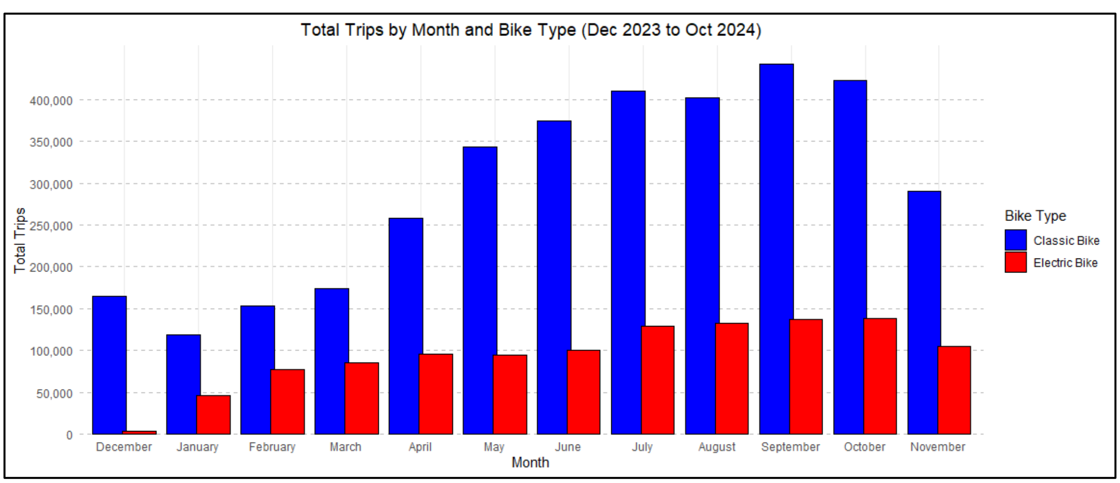


Figure B4. Bar plot showing monthly trends in bike usage

Bibliography

[1] BLUEBikes, “Blue Bikes Comprehensive Trip Histories,” Accessed: Oct. 31, 2024. [Online]. Available: https://bluebikes.com/system-data