Bike-Ride Sharing Data Analysis

Vedasree Bodavula
Ira A. Fulton Schools of Engineering
Arizona State University
Tempe, Arizona
vbodavul@asu.edu

Bindiya Vundavalli
Ira A. Fulton Schools of Engineering
Arizona State University
Tempe, Arizona
bvundav1@asu.edu

Sai Venkat Naresh Kasaragadda
Ira A. Fulton Schools of Engineering
Arizona State University
Tempe, Arizona
skasara1@asu.edu

Mohana Narasimha Reddy Attunuru
Ira A. Fulton Schools of Engineering
Arizona State University
Tempe, Arizona
mattunur@asu.edu

Vamsi Krishna Yadav Loya
Ira A. Fulton Schools of Engineering
Arizona State University
Tempe, Arizona
vloya1@asu.edu

Jayachandra Poluri
Ira A. Fulton Schools of Engineering
Arizona State University
Tempe, Arizona
jpoluri@asu.edu

***Abstract*—This paper presents a Ride Share Analytics Dashboard designed to analyze ride-sharing dynamics in urban environments. Utilizing data from classic and electric bike platforms in Chicago, the project focuses on developing predictive models to determine factors that influence ride-sharing tendencies among users. Techniques such as linear, logistic, multi linear, Poisson, random forest, and gradient boosting regression models are employed to analyze variables including user type, ride type, time of day, day of the week, elevation, minimum and maximum temperature, season, and trip distance and duration, start station ID and name. end station ID and name, start date/time and end date/time. The dashboard integrates data preprocessing, exploratory data analysis, and machine learning modeling to provide stakeholders with insights for traffic management, infrastructure improvement, and sustainability initiatives.**

***Index Terms*—Ride sharing, Regression analysis, Linear regression, Logistic regression, Poisson regression, Random forest regression, Gradient boosting regression, Machine learning, Data analysis, Data visualization, Dashboard, City planning, Sustainability initiatives.**

I. LITERATURE REVIEW

In our exploration of multiple research papers on sustainable smart city transportation and ride-sharing systems, we've synthesized findings that highlight the critical role of technological integration in enhancing urban mobility. Studies show that while single-passenger vehicles remain prevalent, they are not the most efficient due to poor resource allocation and environmental impacts. Bike Sharing Systems (BSS) are gaining popularity for their convenience and positive environmental and health impacts. Furthermore, the integration of environmental factors, such as the analysis of elevation differences in BSS usage, informs our approach to developing a Ride Share Analytics Dashboard. This dashboard leverages dynamic data integration, predictive analytics, and visualization tools to support comprehensive urban transportation planning and promote sustainability. Our research specifically addresses gaps identified in prior studies by delving deeper into the determinants of BSS usage. By scrutinizing variables such as elevation, route distances, and seasonal fluctuations, we employ a diverse array of analytical methodologies, including Poisson regression, multiple linear regression (MLR), gradient boosting regression, and random forest regression to enhance predictive accuracy. This approach not only enhances the prediction accuracy of user behavior and system efficiency but also supports the development of more nuanced urban mobility solutions that consider topographical and environmental variables.

.

II. REGRESSION ANALYSIS REPORT

A. Data Description

For our regression analysis, we utilized a comprehensive dataset encompassing ridesharing data from 2022 to 2023 [9]. This dataset includes detailed records for both classic and elctric bikes with variables such as ride_id, rideable_type, started_at, ended_at, start_station_name, start_station_id, end_station_name, end_station_id, member_casual, start_lat, start_lng, end_lat, end_lat, end_lng, start_altitude, end_altitude. Factors that we derived/calculated - Elevation_Change, Trip Distance, season, day_of_week, time_of_the_Day trip_duration, TMAX (maximum temperature) and TMIN (minimum temperature)

B. Data Preprocessing

Prior to analysis, data preprocessing steps were undertaken to ensure data quality and usefulness.
1. Elevation Data: Elevation information was incorporated into the dataset using GPS visualizer, enriching the dataset with topographical features to potentially enhance the analysis.
2. Weather data - TMIN and TMIN were calculated by merging the chicago city's weather dataset and analysing it.
3. Feature Engineering: Additional features such as trip duration and distance were calculated based on the start and end

timestamps and station coordinates. This allowed for a more comprehensive analysis of ride-sharing patterns.
4. Handling Missing Values: Any missing values in the dataset, particularly in the weather-related features, were addressed through appropriate imputation techniques. Categorical variables were also encoded for analysis [5].

C. Exploratory Data Analysis

1. Ride Counts by Season: Visualizations were generated to illustrate the distribution of ride counts across different seasons, providing insights into seasonal variations in ride-sharing activity. This plot shows the distribution of ride counts by season for electric and classic bikes. It helps in understanding the seasonal variations in bike usage based on their types [6].

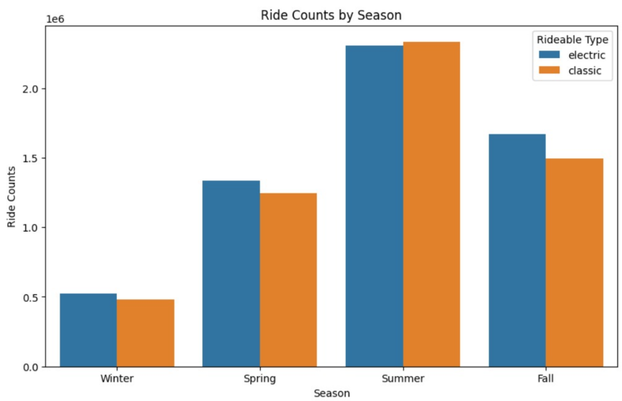


Fig. 1. Seasonal variation in ride counts for electric and classic bikes.

Insights from the above graph [Fig 1]:

- Electric rideables are more popular than classic rideables in all seasons except fall.
- The highest number of rides for both electric and classic rideables occur in the summer.
- The lowest number of rides for electric rideables is in the winter, while for classic rideables it is in the spring.
- There is a significant drop in the number of rides for classic rideables from summer to fall, whereas electric rideables only show a slight decrease.

2. Hourly Ride Counts: Hourly ride counts were analyzed to understand the temporal patterns of ride-sharing usage throughout the day, revealing peak hours of activity. This plot displays the hourly ride counts for both bikes and escooters. It provides insights into the usage patterns throughout the day, helping to identify peak hours and trends.
Insights from the above graph [Fig 2]:

- For both types of rideables, the least popular hours are from midnight to 5 am, with very low ride counts.
- There is a noticeable increase in rides starting around 5 am, peaking between 7 am and 9 am, which could correspond with the morning rush hour.

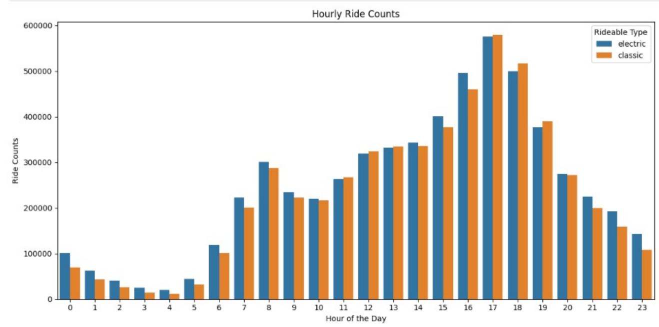


Fig. 2. Hourly variation in ride counts for electric and classic bikes.

- After a slight decrease midday, there is another increase in ride counts starting around 14:00, with the peak hours for rides between 16:00 and 18:00, likely coinciding with the evening rush hour.
- Electric rideables consistently show higher ride counts than classic ones throughout the day.
- Ride counts begin to decrease after 19:00 and continue to drop significantly as it gets later in the night.

3. Average Trip Duration: Monthly trends in average trip duration were examined to identify any seasonal patterns or anomalies in ride durations. This line plot illustrates the average trip duration for each month, categorized by the type of rideable (electric bike or classic bike). It helps in understanding how the trip duration varies over different months for each rideable type.

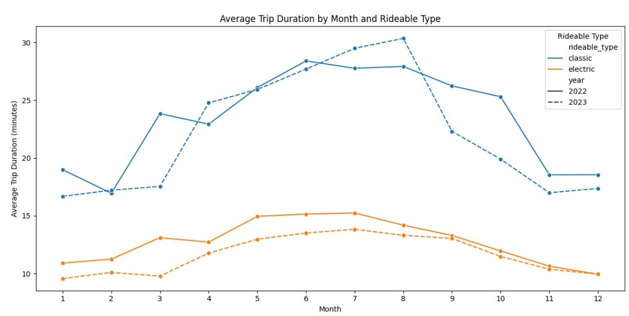


Fig. 3. Average trip duration by month for electric and classic bikes.

Insights from the above graph[Fig 3]:

- The average trip duration for classic rideables peaks around May and June for both years, indicating that during these months, trips tend to be longer.
- Electric rideables have a shorter average trip duration compared to classic rideables across most months.
- For both types of rideables, there is a significant drop in average trip duration in the later months of the year, particularly in November and December. This could be related to colder weather or holiday periods.
- In 2023, the average trip duration for electric rideables appears to increase slightly compared to 2022, while for

classic rideables, the durations decrease or remain similar to the previous year.
- The longest trips for classic rideables occur around the middle of the year (May to July), which might correspond with summer weather and potentially more recreational usage.

4. Monthly Trends: Monthly ride trends for each rideable type (electric bikes and classic bikes) were explored to detect any longitudinal shifts or changes in ride-sharing behavior over time This line plot depicts the monthly ride trends for each rideable type (electric bike or classic bike). It allows us to observe the variations in ride counts over the months, facilitating the identification of seasonal trends and patterns.

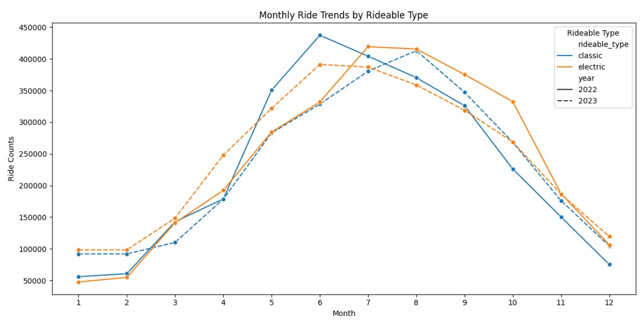


Fig. 4. Monthly ride trends by rideable type for electric and classic bikes.

Insights from the above graph [Fig 4]:

- For both types of rideables, there is a growth in rides starting from the beginning of the year, peaking around the middle months, which is likely due to more favorable weather conditions.
- The peak ride counts for both classic and electric rideables occur during the warmer months, roughly from May to September.
- After the peak, there is a decline in ride counts as the year progresses towards winter. Comparing the two years, there is a noticeable trend where the number of rides for electric rideables in 2023 generally surpasses the counts for the same months in 2022. This could indicate an increasing popularity or availability of electric rideables.
- The trend for classic rideables is less consistent, with some months in 2023 showing more rides than in 2022 and others showing fewer.

4. Count of Rides by Member Type and Rideable Type This bar plot presents the count of rides categorized by both member type (member or casual) and rideable type (electric bike or classic bike). It provides insights into the preferences of different member types for each rideable type.

Insights from the graph below[Fig 5]:

- Members have taken more rides than casual riders for both electric and classic rideables.
- For both member types, electric rideables seem to be slightly more popular than classic rideables.

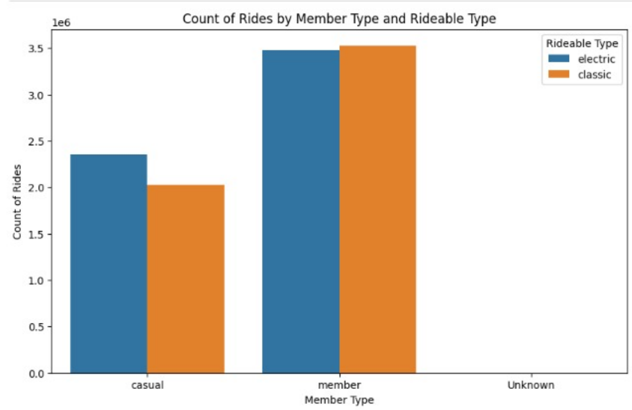


Fig. 5. Count of rides by member type and rideable type.

5. Count of rides by Member type, Month The bar plots present the count of rides categorized by both member type (member or casual) and month for the years 2022 and 2023. They provide insights into the frequency of rides taken by different member type's, based on all the months in the years 2022 and 2023.

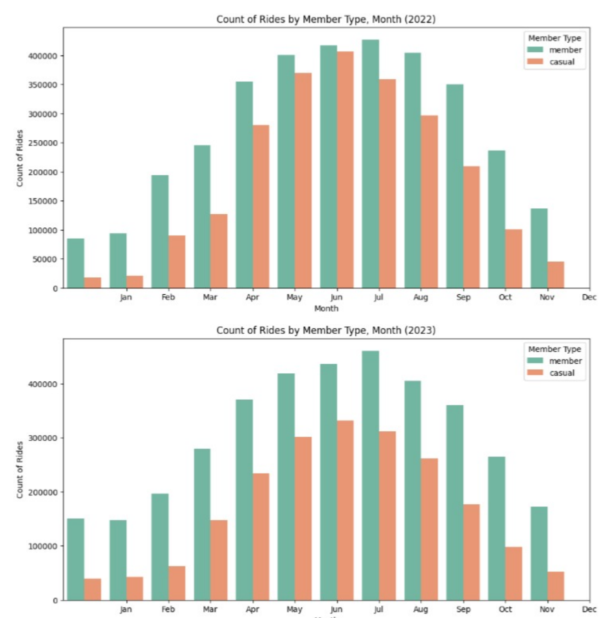


Fig. 6. Count of Rides by member type and month [2022, 2023]

Insights from the above graph [Fig 6]:

- In both 2022 and 2023, the number of rides for both members and casual users peaks during the warmer months (from around May to September), suggesting seasonal influences on ride usage.
- For most months in both years, members seem to take more rides than casual users.
- There's a sharp increase in rides starting from spring and peaking in summer, followed by a decrease as the months

get colder, showing less activity in the winter season.
- Comparing both years, there are similar patterns of ride usage, indicating consistent behavior across the years despite any potential changes in service or population.

5. Map Visualization

Interactive map visualization for both classic and electric bikes, where the points are showed as numbered clusteres initially, and on zooming in, exact map locations of classic and electric bikes are shown [7].

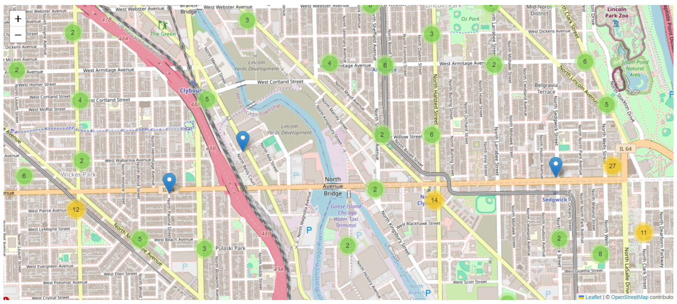


Fig. 7. Interactive Map of Electric bikes

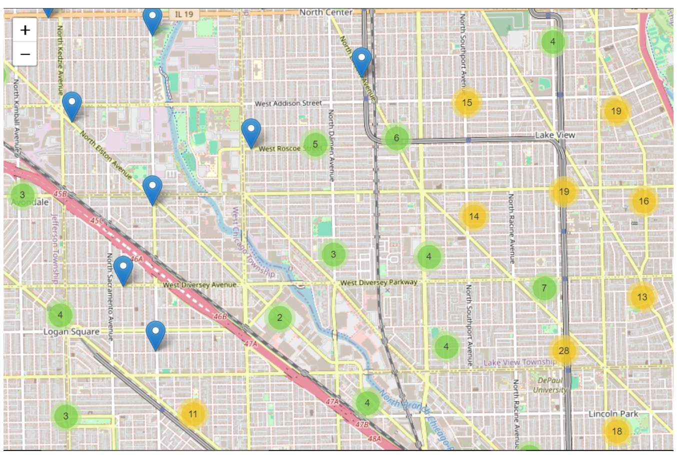


Fig. 8. Interactive Map of Classic bikes

D. Regression Analysis Methodology

We implemented Linear Regression, Logistic regression, Poisson regression, Random Forest Regression and Gradient Boosting regression models. These models were chosen to compare their effectiveness in predicting daily usage counts based on various factors including distance, season, day of the week, altitude difference, and trip duration.

Independent Variables:

- ride_id: A unique identifier assigned to each individual ride. This helps differentiate each trip from others in the dataset.
- started_at: The date and time when the ride started. This timestamp allows analysis of ride patterns over time and can be used to determine the busiest times of day or week.
- ended_at: The date and time when the ride ended. Similar to 'started_at', this helps in understanding the duration of each ride and the distribution of ride times throughout the day or the year.
- start_station_name: The name of the station where the ride originated. This is useful for geographic analysis and understanding popular starting points within the system.
- start_station_id: A unique identifier for the start station. This can be used in conjunction with the station name to accurately identify locations and avoid confusion in cases where multiple stations might have similar names.
- end_station_name: The name of the station where the ride concluded. This provides insights into popular destinations within the ride-sharing network.
- end_station_id: A unique identifier for the end station, used similarly to the start_station_id to uniquely identify stations within the system.
- member_casual: Indicates whether the rider is a registered member or a casual (non-member) user of the service. This helps in demographic and usage pattern analysis, distinguishing between different user engagement levels.
- Elevation_Change: The difference in elevation (altitude) between the start and end points of the ride. This can be used to study the physical effort required for the ride and to understand how topography influences ride preferences.
- Distance: The total distance traveled during the ride. This is key for analyzing the average length of rides and can be used to compute things like fuel or energy usage, if applicable.
- season: Categorizes the ride based on the time of year it took place (e.g., Spring, Summer, Fall, Winter). This allows for seasonal analysis of ride-sharing usage, such as identifying seasonal trends or the effect of weather conditions on ride frequency.
- day_of_week: The day of the week on which the ride occurred. This variable is useful for weekly pattern analysis, such as determining which days are busiest or comparing weekday usage to weekend usage.
- trip_duration: The total duration of the ride, usually calculated from the 'started_at' and 'ended_at' timestamps. This measure is crucial for understanding how long rides typically last, which can inform operational and pricing strategies.
- TMAX: The maximum temperature on the day of the ride. This environmental variable can be used to analyze the impact of weather conditions on ride-sharing behavior.
- TMIN: The minimum temperature on the day of the ride. Similar to TMAX, this provides insights into how extreme temperatures might affect rider choices and ride frequency. Dependent variable:
- rideable_type: Specifies the type of vehicle used for the ride- classic or electric bike ride share. This variable helps categorize the data by the vehicle type.

III. Regression Models

A. *Random Forest Regression Model*

Feature Importance:
The top three important features influencing the prediction of usage count for both classic bikes and electric bikes are trip_duration, Distance, and Elevation_Change. Other significant features include TMAX and TMIN (temperature), member_casual_member and member_casual_casual (member type), as well as season and day_of_week.

Performance Metrics:
For classic bikes, the precision, recall, and F1-score are around 0.65 to 0.69, indicating a reasonably good predictive performance. Similarly, for electric bikes, the precision, recall, and F1-score range from 0.63 to 0.69, suggesting a comparable performance [8].

TABLE I
Comparison of Coefficient Values between Random Forest and Multivariate Linear Regression Models

Feature	Random Forest	MLR
Elevation_Change	-0.0002161199	-0.0002161199**
Distance	-0.00002003583	-0.00002003583
trip_duration	-0.00009359789	-0.00009359789**
TMAX	-0.001588950	-0.001588950**
TMIN	-0.0002694059	-0.0002694059**
member_casual_casual	0.02735762	0.02735762***
member_casual_member	-0.02735762	-0.02735762***
season_Fall	0.01525531	0.01525531*
season_Spring	0.001849525	0.001849525
season_Summer	0.008342118	0.008342118
season_Winter	-0.02544695	-0.02544695**
day_of_week_Friday	0.01708638	0.01708638*
day_of_week_Monday	-0.005980562	-0.005980562
day_of_week_Saturday	-0.02203042	-0.02203042**
day_of_week_Sunday	-0.02604875	-0.02604875***
day_of_week_Thursday	0.01682715	0.01682715*
day_of_week_Tuesday	0.005087729	0.005087729
day_of_week_Wednesday	0.01505847	0.01505847*

B. *Linear Regression Model*

Coefficient Analysis:
The coefficients indicate the direction and magnitude of the effect of each independent variable on the usage count. When we take All variables except rideable_type as independent variables, then, similar to the Random Forest model, trip_duration, Distance, and Elevation_Change exhibit significant coefficients in influencing the usage count. Additionally, variables such as TMAX and TMIN, member_casual_member and member_casual_casual, and season and day_of_week also contribute to the prediction.

Performance Metrics:
When we take All variables except rideable_type as independent variables, the recall and F1-score for classic bikes are approximately 0.41 to 0.64, indicating a fair predictive performance. Likewise, for electric bikes, the recall and F1-score range from 0.53 to 0.64, suggesting a comparable performance.

When we take elevation, trip duration and distance as independent variables and rideable_type as dependent variable, then the recall and F1-score for classic bikes are approximately 0.41 to 0.64, indicating a fair predictive performance. Likewise, for electric bikes, the recall and F1-score range from 0.53 to 0.64, suggesting a comparable performance [3].

C. *Gradient Boosting Regression Model*

Independent variables - All factors except rideable_type
Dependent variable - rideable_type

Feature Importance:
The top three important features influencing the prediction of usage count for classic and electric bikes are Distance , trip_duration, , and Elevation_Change in the decresing order. Other significant features include , member_casual_casual andmember_casual_member (member type), TMAX and TMIN (temperature) as well as season and day_of_week.

	Feature	Importance
1	Distance	0.498346
2	trip_duration	0.416795
0	Elevation_Change	0.050170
6	member_casual_casual	0.019550
7	member_casual_member	0.010294
4	TMIN	0.002744
3	TMAX	0.000775
8	season_Fall	0.000307
10	season_Summer	0.000297
9	season_Spring	0.000266
12	day_of_week_Friday	0.000125
15	day_of_week_Sunday	0.000122
11	season_Winter	0.000104
14	day_of_week_Saturday	0.000092
13	day_of_week_Monday	0.000013
5	member_casual_Unknown	0.000000
16	day_of_week_Thursday	0.000000
17	day_of_week_Tuesday	0.000000
18	day_of_week_Wednesday	0.000000

Fig. 9. GBR - Important Features

Performance Metrics:
The Mean Squared error is approximately 0.195 which means that the the squared differences between the predicted and actual values are around 0.20
Mean Absolute Error error is 0.411 which meand that on average, the model's predictions are around 0.41 units away from the actual values.
Root Mean Squared Error is 0.442 which means the same as MSE but in the original units of the target variable [1].

D. *Poisson Regression Model*

Coefficient Analysis:
Coefficients from the Poisson regression model provide insights into the impact of each independent variable on the count of usage events. Variables such as trip_duration, Distance, and Elevation_Change demonstrate significant coefficients, aligning with findings from other models. Additional factors like TMAX and TMIN (temperature), member_casual_member and member_casual_casual (member

type), and season and day_of_week also exhibit notable coefficients. These coefficients help quantify the influence of each independent variable on the count of usage events, providing valuable insights into the factors driving bike and electric scooter usage.

Feature Importance:
In the Poisson regression model used to predict ride-sharing trip counts, we delve into the significance of several features. Notably, the feature of distance shows a positive correlation with the event count for both bikes and e-scooters, though the coefficients are small. Seasonality also emerges as an influential factor, particularly affecting e-scooter usage, with a pronounced dip during winter months. Days of the week further exhibit variations, with weekends typically experiencing a decrease in ridership for bikes, with Tuesday as the reference day. In contrast, e-scooter usage displays a consistent decline across all days.

Performance Metrics:
The model's performance metrics, specifically the Root Mean Squared Error (RMSE), for the features 'Season' and 'Duration (Minutes)', indicate that 'Season' has a relatively low RMSE value of 0.62, suggesting a weaker correlation with transportation preference, while 'Duration' has an RMSE value of 0.6498, which is similar to that of 'Season'. 'Distance' has an RMSE value of 0.47, revealing a clearer correlation and better model accuracy for predicting ride-sharing usage [2].

E. *Logistic Regression Model*

Independent variables - Elevation_Change, season, trip_duration, Distance, day_of_week.
Dependent variable - rideable_type.

Coefficient Analysis:
Elevation_Change has a coefficient of -0.000671, while Distance has a more significant coefficient of 0.000095, suggesting a positive correlation with the probability of selecting an electric bike over a classic one. The model also accounts for time-related variables such as trip duration and categorical factors like season and day of the week, with each day having its coefficient, showing how these factors influence daily usage patterns.

Performance Metrics:
The Mean Squared Error (MSE) stands at 0.4424835646628323, suggesting that the predicted probabilities deviate from the actual observed frequencies. Additionally, the R squared score is -0.771025042377862, indicating that the model does not explain the variability of the response data around its mean effectively.
Moreover, the model has a recall of 0.7923400122820297 and an F1-Score of 0.6472793947448984, reflecting its ability to identify the true positives and its balance between precision and recall, respectively [4].

Mean Squared Error: 0.4424835646628323
R² Score: -0.7710250423778622

	Coefficient
Elevation_Change	-0.000671
Distance	0.000095
trip_duration	-0.014597
season_Fall	0.083438
season_Spring	0.069393
season_Summer	0.001189
season_Winter	0.043723
day_of_week_Friday	0.094066
day_of_week_Monday	-0.004058
day_of_week_Saturday	-0.002810
day_of_week_Sunday	-0.015210
day_of_week_Thursday	0.065395
day_of_week_Tuesday	0.009808
day_of_week_Wednesday	0.050551

Fig. 10. Logistic Regression

Recall: 0.7923400122820297
F1-Score: 0.6472793947448984

	precision	recall	f1-score	support
0	0.59	0.31	0.41	1665742
1	0.55	0.79	0.65	1750525
accuracy			0.56	3416267
macro avg	0.57	0.55	0.53	3416267
weighted avg	0.57	0.56	0.53	3416267

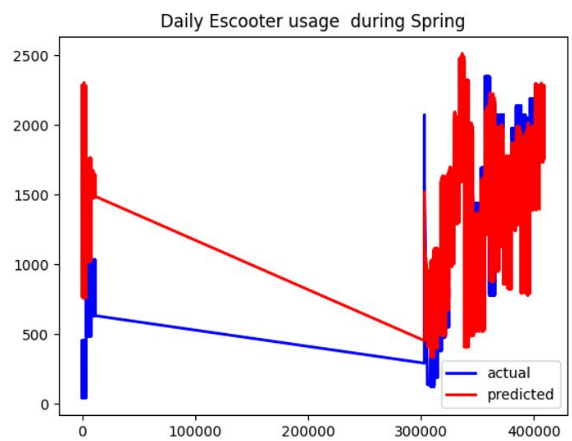
Fig. 11. Logistic Regression - Recall, F1 Score

IV. Roles

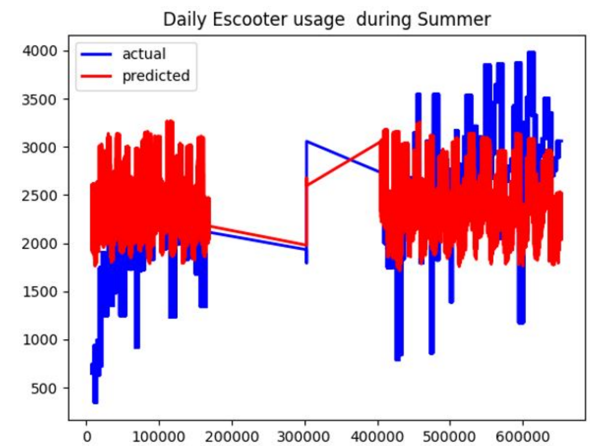
- **Sai Venkat Naresh Kasaragadda** - Naresh worked as both Backend developer and Tester - He played an integral part in data collection and preprocessing, particularly in integrating elevation data for bike and e-scooter stations, to advanced feature engineering. . His contributions to feature engineering were crucial, as he developed several key predictive variables that were used in the regression analysis. He also participated actively in EDA, identifying peak hours and seasonal usage patterns that informed the development of more accurate models in the regression analysis phase. He tested the regression models to ensure their accuracy and reliability before they were deployed in the dashboard.
- **Vedasree Bodavula** : Vedasree worked as Frontend and Backend developer - She was pivotal in gathering and preprocessing the data, particularly focusing on the accuracy of the GPS and elevation data for e-scooters. In regression analysis, she played a crucial role, especially in refining

TABLE II
THE POISSON REGRESSION MODELS FOR THE DAILY E-SCOOTER USAGE IN SPRING, SUMMER, FALL, AND WINTER.

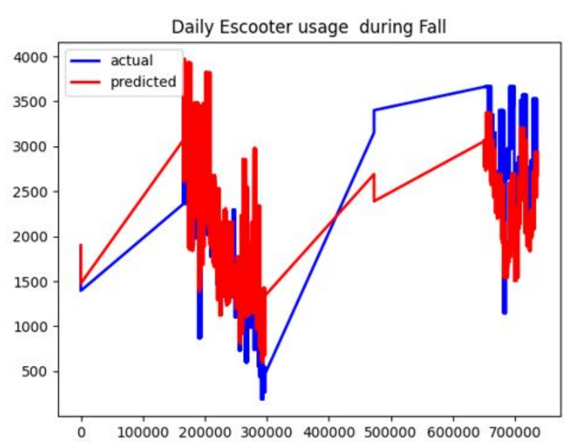
Variable	**coef_spring**	**coef_summer**	**coef_fall**	**coef_winter**
Intercept	4.660517***	6.444028***	5.057280***	3.113680***
TMIN	-0.008454***	-0.000606***	0.018056***	0.019273***
TMAX	0.033526***	0.005553***	0.009440***	0.009858***
Lime	-0.020027***	0.082644***	0.169293***	1.226276***
Lyft	-0.256917***	-0.035785***	0.136636***	0.617085***
Spin	-0.020709***	-0.009315***	0.095393***	1.270319***
Trip Distance (miles)	0.006214***	0.008946***	0.003306***	-0.000147
Trip Duration (minutes)	-0.000649***	-0.002062***	-0.000745***	-0.000100
Friday	0.806154***	1.038761***	0.934012***	0.553375***
Monday	0.540557***	0.782380***	0.563700***	0.444209***
Saturday	0.846527***	1.086590***	0.909352***	0.426421***
Sunday	0.693312***	0.920240***	0.700407***	0.308774***
Thursday	0.628112***	0.919715***	0.749456***	0.585785***
Tuesday	0.533494***	0.852127***	0.529750***	0.426142***
Wednesday	0.612360***	0.844216***	0.670602***	0.368974***



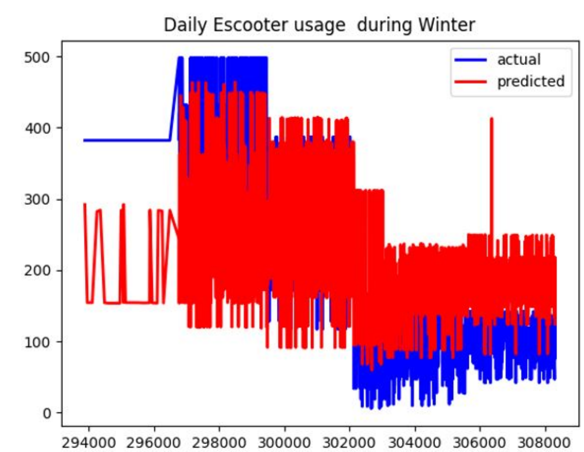
(a) Daily e-scooter usage in Spring.



(b) Daily e-scooter usage in Summer.



(c) Daily e-scooter usage in Fall.



(d) Daily e-scooter usage in Winter.

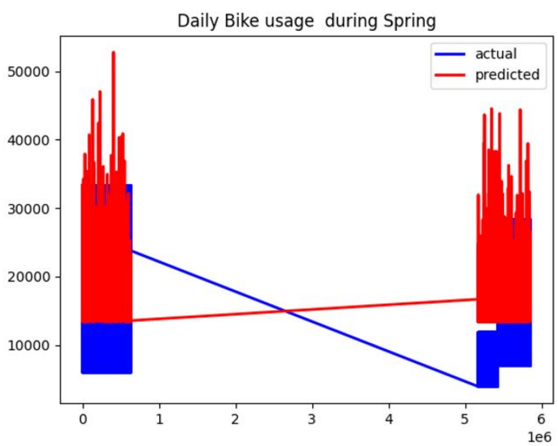
Fig. 12. Seasonal analysis of daily e-scooter usage showing the actual versus predicted counts, highlighting variations and trends across different seasons.

the variables that significantly impact ride counts using Multivariate Linear Regression model. Additionally, Vedasree took the lead in designing the user interface (UI) of the dashboard, ensuring that the visualizations were both functional and aesthetically pleasing, enhancing user engagement and data accessibility.

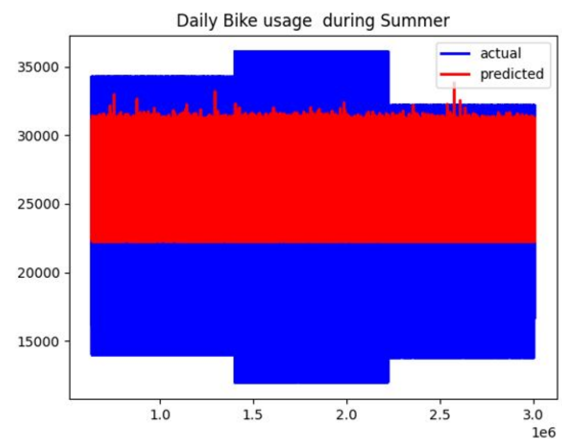
- **Bindiya Vundavalli** - Bindiya worked as Scrum Master and Backend developer - She was deeply involved in

TABLE III
THE POISSON REGRESSION MODEL COEFFICIENTS FOR THE DAILY BIKE USAGE IN SPRING, SUMMER, FALL, AND WINTER.

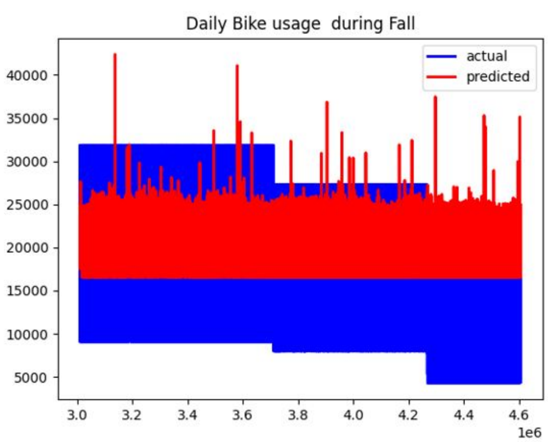
Variable	**coef_spring**	**coef_summer**	**coef_fall**	**coef_winter**
Intercept	4.951447***	5.146536***	5.008190***	4.481397***
casual	2.553006***	2.577230***	2.532836***	2.252964***
member	2.398441***	2.569306***	2.475353***	2.228433***
classic_bike	1.684065***	1.717382***	1.680079***	1.490441***
docked_bike	1.694466***	1.716715***	1.670502***	1.502318***
electric_bike	1.572916***	1.712439***	1.657608***	1.488639***
distance	0.017428***	0.001231***	0.000044***	0.007987***
duration	0.000504***	0.000069***	0.000391***	-0.000060***
Friday	0.695369***	0.786383***	0.788769***	0.579826***
Monday	0.720880***	0.580478***	0.573100***	0.660135***
Saturday	0.846850***	0.901569***	0.903563***	0.479147***
Sunday	0.661260***	0.762772***	0.638290***	0.540355***
Thursday	0.757680***	0.738163***	0.785696***	0.664610***
Tuesday	0.689982***	0.681393***	0.608090***	0.818938***
Wednesday	0.579427***	0.695777***	0.710681***	0.738386***



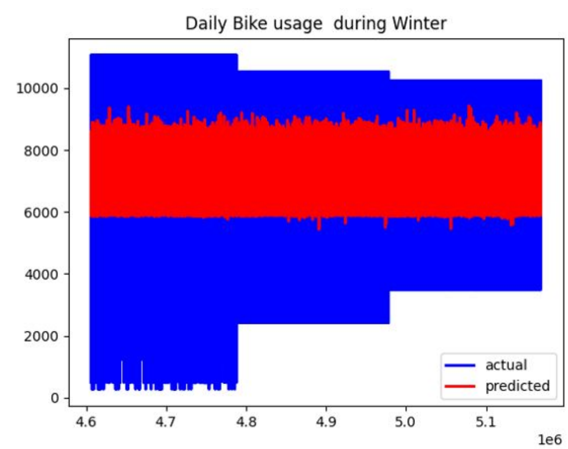
(a) Daily bike usage in Spring.



(b) Daily bike usage in Summer.



(c) Daily bike usage in Fall.



(d) Daily bike usage in Winter.

Fig. 13. Seasonal analysis of daily bike usage showing the actual versus predicted counts, highlighting variations and trends across different seasons.

the initial stages of data collection and preprocessing, addressing missing values with sophisticated imputation techniques and preparing the dataset for deeper analysis. She collaborated in feature engineering, creating variables that were crucial for regression models, especially using the Gradient Boosting Regression model. In the EDA phase, she analyzed the impact of weather conditions on ride counts. Her insights were critical in shaping the regression analysis strategies and she used Gradient Boosting Regression model to predict how various factors impact the ride sharing tendency of users. She also worked on creating an interactive map to visualize the geographic positions of both electric and classic bikes.

- **Mohana Narasimha Reddy Attunuru** : Narasimha worked as a both Frontend and backend developer - He was particularly invloved in EDA, where his analysis of seasonal ride trends helped to identify variables for regression analysis. He worked closely with the team on the regression models, analyzing the interaction between ride durations and seasonal trends. His efforts were also key in the dashboard development, where he ensured that the data visualizations were insightful and effectively communicated the findings.
- **Vamsi Krishna Yadav Loya** : Vamsi Krishna worked as both Backend developer and Tester - His work spanned from initial data collection to detailed regression analysis. He was particularly influential in feature engineering, where he calculated trip distances and developed features that significantly enhanced the regression models. His extensive work with Poisson Regression model helped to predict ride counts with high accuracy, and he played a major role in evaluating the performance and accuracy of the regression models under various data scenarios.
- **Jayachandra Poluri** - Jayachandra worked as both Frontend and Backend developer - He was actively involved in the entire lifecycle of the project, from data collection and preprocessing to dashboard development. He played a significant role in feature engineering and was instrumental in the regression analysis phase, where he fine-tuned the models to improve their predictive power. His technical skills were crucial in the integration of complex regression outputs into the dashboard, allowing for interactive exploration of predictive models

V. RESULTS

The logistic regression model reveals several significant features influencing the usage of bikes and electric scooters. Features such as `member_casual_casual`, `season_Fall`, and `day_of_week_Friday` show positive coefficients, indicating a strong positive association with the likelihood of choosing bikes or electric scooters.

Poisson regression analysis for electric bikes highlights notable coefficients across different seasons. During the Summer, predictors like `Lime` and `day_of_week_Saturday` exhibit positive coefficients, suggesting a higher usage on Saturdays and when Lime scooters are available. In contrast, the Winter season features `Lime`, `Spin`, and `day_of_week_Thursday` as influential factors. The Fall season shows `Lime` and `day_of_week_Saturday` as significant, while Spring emphasizes the importance of `Lime`, `day_of_week_Saturday`, and `TMAX`.

The multi-linear regression model demonstrates a moderate level of predictive performance, with recall and F1-scores ranging from 0.41 to 0.64 for classic bikes and from 0.53 to 0.64 for electric bikes. The model identifies `trip_duration`, `Distance`, and `Elevation_Change` as key predictors.

Furthermore, Gradient Boosting regression and random forest models corroborate the findings of the logistic and multi-linear regression analyses. They highlight the importance of features such as `trip_duration`, `Distance`, and `Elevation_Change` as top predictors for both types of bikes. Additional variables including `TMAX`, `TMIN`, `member_casual_member`, `member_casual_casual`, as well as `season` and `day_of_week`, also contribute significantly to the prediction of usage counts.

These comprehensive analyses offer valuable insights into the factors that shape urban transportation usage patterns, providing a foundation for enhancing transportation services and city planning.

REFERENCES

[1] Gradient Boosting Regression Model .
[2] Poisson Regression Model
[3] Linear Regression Model
[4] Logistic Regression Model
[5] Data Preparation
[6] Exploratory Data Analysis
[7] Map Visualization
[8] Random Forest Regression Model
[9] Divvy 2022-23 Dataset