RIDE SHARING ANALYTICS DASHBOARD
PROGRESS FINAL REPORT

VAMSI KRISHNA YADAV LOYA

2. Explanation of all Visualized data (github/bike and e-scooter analysis, divvy bike 2018&2023)

And the dataset used for below exploratory data analysis is for the year 2018 & 2023 Bike and E-scooter

Drive Link: https://drive.google.com/drive/u/1/folders/166RVZX8pm9_rqz45iXNkCrLCT2UA2WvG

GitHub Link: https://github.com/Bindiyaa5/SER517team29/tree/main/Bike%20and%20E-scooter%20Analysis

https://github.com/Bindiyaa5/SER517team29/tree/main/Divvy%20Bike%202018%20%26%202023

- ***ride counts by seasons for bike, escooter***

Figure:A

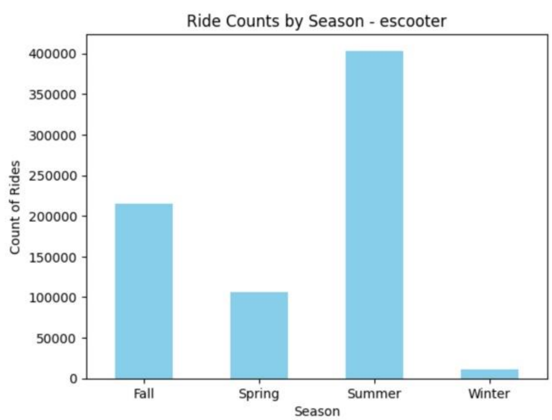
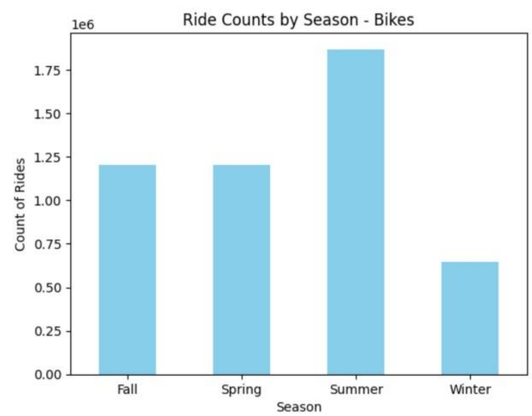


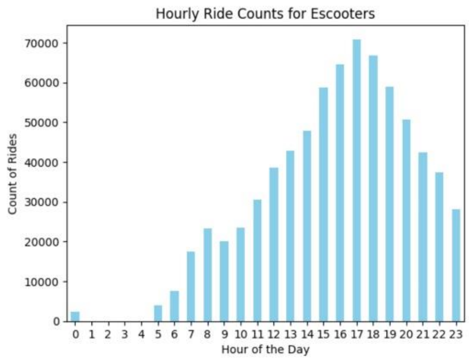
Figure:B



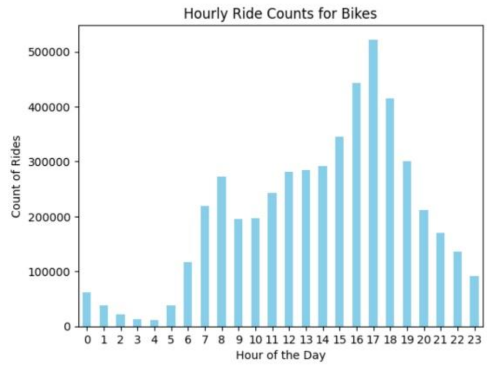
“Ride Counts by Season” plots above **Figure:A** which shows the number of scooter rides taken in each of the four seasons: fall, spring, summer, and winter. The y-axis shows the count of rides, and the x-axis shows the season. There are four data points plotted on the graph, one for each season. **The highest data point is for summer, which shows that there were more scooter rides taken in the summer than in any other season.** The number of rides taken in the spring and fall were similar, and there were the fewest rides taken in the winter.

Figure:B The graph shows a seasonal trend in bike ridership. The highest data point is for summer, which shows that there were more bike rides taken in the summer than in any other season. The number of rides taken in the spring and fall were similar, and there were the fewest rides taken in the winter.

- ***hourly ride counts for bike, escooter***



A: Hourly ride counts e-scooter



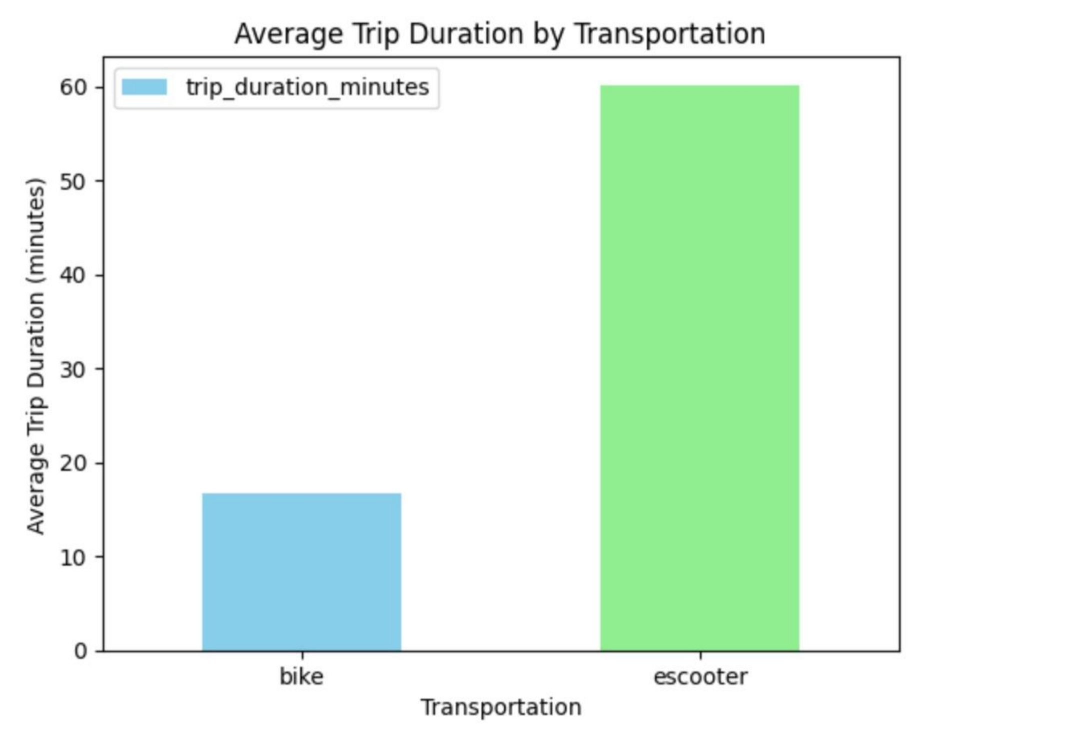
B: Hourly ride counts Bike

- ***average trip duration***

	ride_id	Ride_start_timestamp	Ride_end_timestamp	trip_duration_minutes	transportation	preference
1471049	3960004BE8F6B206	2022-06-24 09:46:19	2022-06-24 09:55:13	8.900000	bike	bike
3339312	CCFDE14886043243	2022-10-10 16:54:26	2022-10-10 17:58:23	63.950000	bike	escooter
309603	17909190	2018-03-18 17:52:43	2018-03-18 17:59:30	6.783333	bike	bike
3380246	1EA2CD794E7AF61F	2022-10-10 14:54:22	2022-10-10 14:59:38	5.266667	bike	bike
114443	8a193a65-a92d-41f0-b183-2a83685436f6	2022-08-05 22:00:00	2022-08-05 23:00:00	60.000000	escooter	escooter
...	...	...	...	...	...	...
1067587	019B0A4E49E44EF8	2022-05-15 14:07:08	2022-05-15 14:10:57	3.816667	bike	bike
225379	92B197E7C211A552	2023-02-14 08:19:29	2023-02-14 08:48:01	28.533333	bike	bike
4308201	1EABDFE87B2C7BCB	2022-09-15 07:18:54	2022-09-15 07:43:21	24.450000	bike	bike
2925298	44D207C204D4E977	2022-08-26 22:52:36	2022-08-26 23:33:03	40.450000	bike	escooter
1305468	29B72BE8D63349FD	2022-05-18 13:01:30	2022-05-18 13:17:53	16.383333	bike	bike

5657648 rows × 6 columns

Preference: Preference is basically what type of transportation a rider choose based on season, duration and distance by using poison regression. I have performed one hot encoding and assigned 0 as bike and 1 as e-scooter. Since the Poisson regression gives predictions, I have kept a threshold of 0.7, if the prediction values are greater than 0.7 then I assumed that the rider prefer e-scooter otherwise bike.



A: For comparison between Bike vs E-scooter

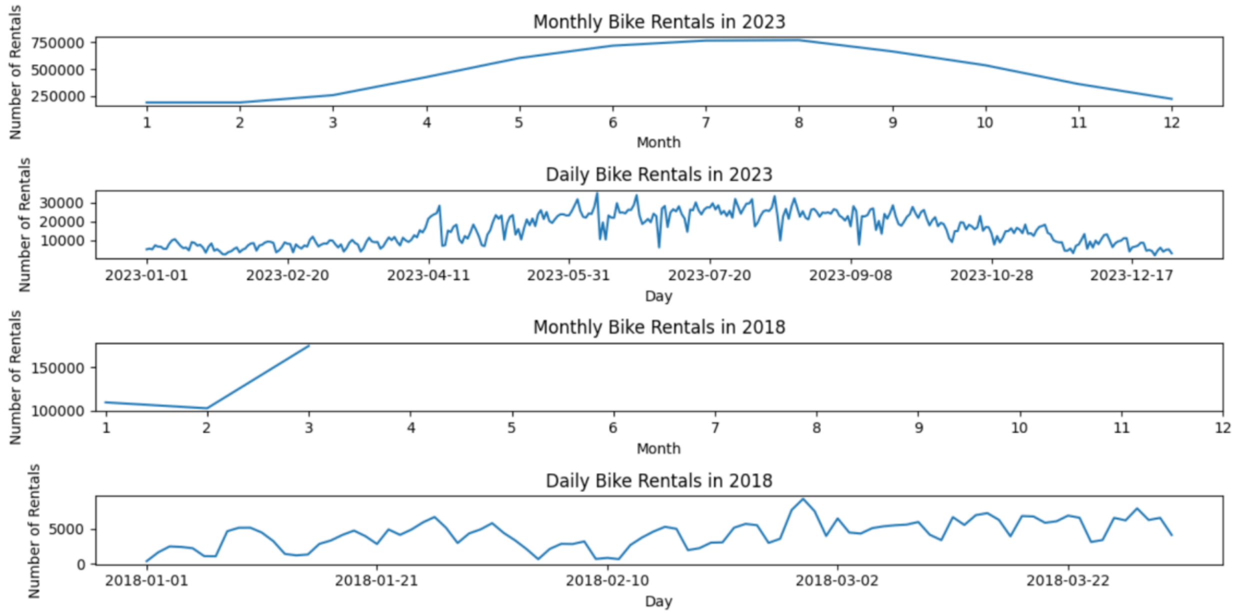
Fig A: depicts bicycles have the shortest average trip duration at around 10 minutes. This is likely because bicycles are commonly used for short trips around town, such as commuting to work or running errands. E-scooters have a slightly longer average trip duration than bicycles, at around 15 minutes. This could be because e-scooters are used for similar trips as bicycles, but they may also be used for slightly longer recreational trips.

```
# Calculate average trip duration for each transportation type
avg_trip_duration_escooter = merged_df[merged_df['transportation'] == 'escooter']['trip_duration_minutes'].mean()
avg_trip_duration_bike = merged_df[merged_df['transportation'] == 'bike']['trip_duration_minutes'].mean()

# Use the average trip duration as the benchmark
benchmark_duration = (avg_trip_duration_escooter + avg_trip_duration_bike) / 2

print("Average trip duration for escooter rides:", avg_trip_duration_escooter)
print("Average trip duration for bike rides:", avg_trip_duration_bike)
print("Benchmark duration:", benchmark_duration)
```

```
Average trip duration for escooter rides: 60.14126196191738
Average trip duration for bike rides: 16.61618633797343
Benchmark duration: 38.378724149945405
```

B: Average Trip Duration

- monthly trends 2018 vs 2023



The top graph shows daily bike rentals in 2023 and the bottom graph shows daily bike rentals in 2018. The y-axis of both graphs shows the number of rentals. The x-axis of both graphs shows the day, with tick marks every two months. There are many more daily rentals in 2023 than in 2018. For example, in 2023, the highest number of daily rentals is around 30,000, whereas in 2018, the highest number of daily rentals is around 5,000.

- summary statistics

summary statistics

```
In [55]:
# Summary statistics for trip durations in 2018
statistics_2018 = df_2018[df_2018['year'] == 2018]['trip_duration'].describe()

# Summary statistics for trip durations in 2023
statistics_2023 = rides_df[rides_df['year'] == 2023]['trip_duration'].describe()
```

```
In [56]:
statistics_2018
```

```
Out[56]: count                        387145
mean     0 days 00:17:16.236864740
std      0 days 14:38:20.493509768
min                0 days 00:01:01
25%                0 days 00:05:22
50%                0 days 00:08:31
75%                0 days 00:13:57
max              165 days 23:20:41
Name: trip_duration, dtype: object
```

```
In [57]:
statistics_2023
```

```
Out[57]: count                       5712887
mean     0 days 00:15:08.597128912
std      0 days 00:36:29.173668223
min             -12 days +10:23:29
25%                0 days 00:05:25
50%                0 days 00:09:31
75%                0 days 00:16:53
max                8 days 10:16:18
Name: trip_duration, dtype: object
```

The code first filters the dataframes df_2018 and rides_df to select only trips that occurred in the year 2018 and 2023 respectively. Then, it uses the .describe() method to calculate summary statistics for the 'trip_duration' column. The .describe() method typically outputs various statistics including the count, mean, standard deviation, minimum, quartiles (25th and 75th percentiles), and maximum values.

For instance, an increase in the average trip duration could indicate that people are using bikes for longer commutes. On the other hand, a decrease in standard deviation might suggest a shift towards a more standardized use of bikes, perhaps for shorter trips.

3. Regression results for bike and e-scooter. show as a table with significance mark *

Poisson Regression Model E-scooter:

Independent variables: 'TMIN','TMAX','Lime', 'Lyft', 'Spin','Trip Distance (miles)','Trip Duration (minutes)', 'Friday', 'Monday', 'Saturday', 'Sunday','Thursday', 'Tuesday', 'Wednesday']

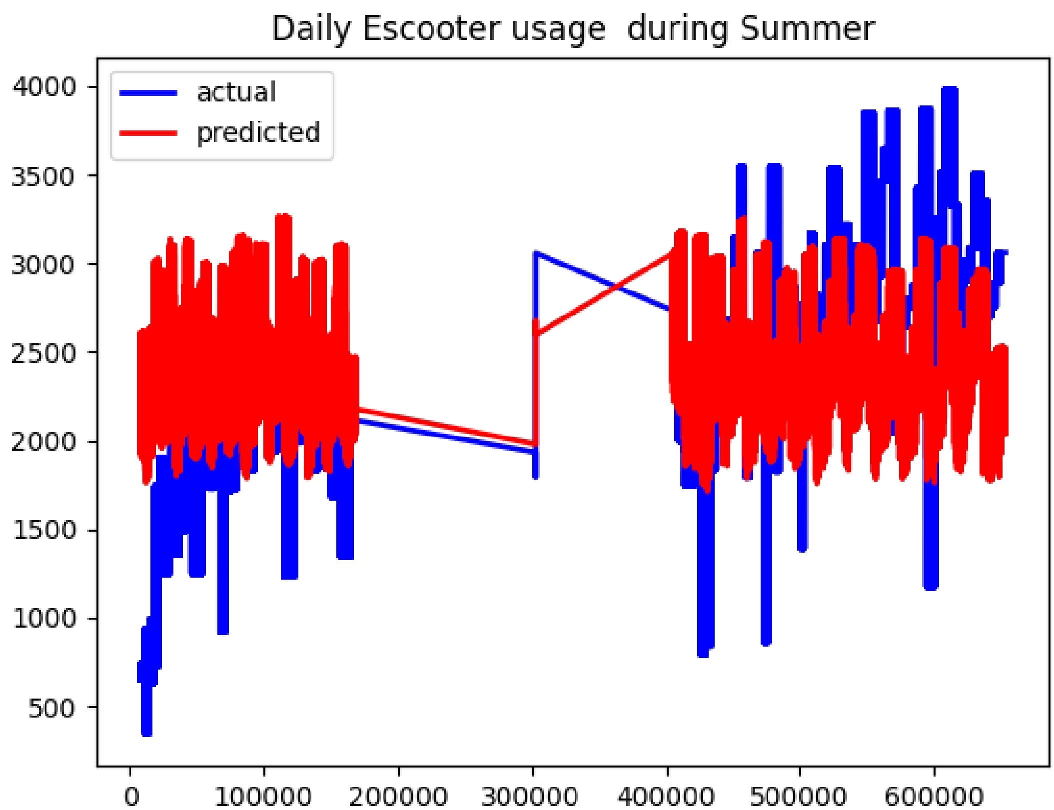
Dependent variable: Daily_usage_count

```
Start Date: 2022-05-10 09:00:00
End Date: 2023-09-30 19:00:00
```

Summer:

```
Pm params values and coefficients for Summer
6.444027610148412*** (p-value: 0.0)
-0.0006061633251404911*** (p-value: 0.0)
0.0055525782163132744*** (p-value: 0.0)
0.0826437333864797*** (p-value: 0.0)
-0.03578486557415667*** (p-value: 0.0)
-0.009314616119533416*** (p-value: 0.0)
0.008945646430521927*** (p-value: 0.0)
-0.00206209636380601*** (p-value: 0.0)
1.038760609786434*** (p-value: 0.0)
0.7823795064795924*** (p-value: 0.0)
1.0865897409524474*** (p-value: 0.0)
0.9202402870107558*** (p-value: 0.0)
0.9197145474730851*** (p-value: 0.0)
0.8521268043352546*** (p-value: 0.0)
0.8442161141108463*** (p-value: 0.0)

Coefficient Values for Summer:
const                      6.444028
TMIN                      -0.000606
TMAX                       0.005553
Lime                       0.082644
Lyft                      -0.035785
Spin                      -0.009315
Trip Distance (miles)      0.008946
Trip Duration (minutes)   -0.002062
Friday                     1.038761
```

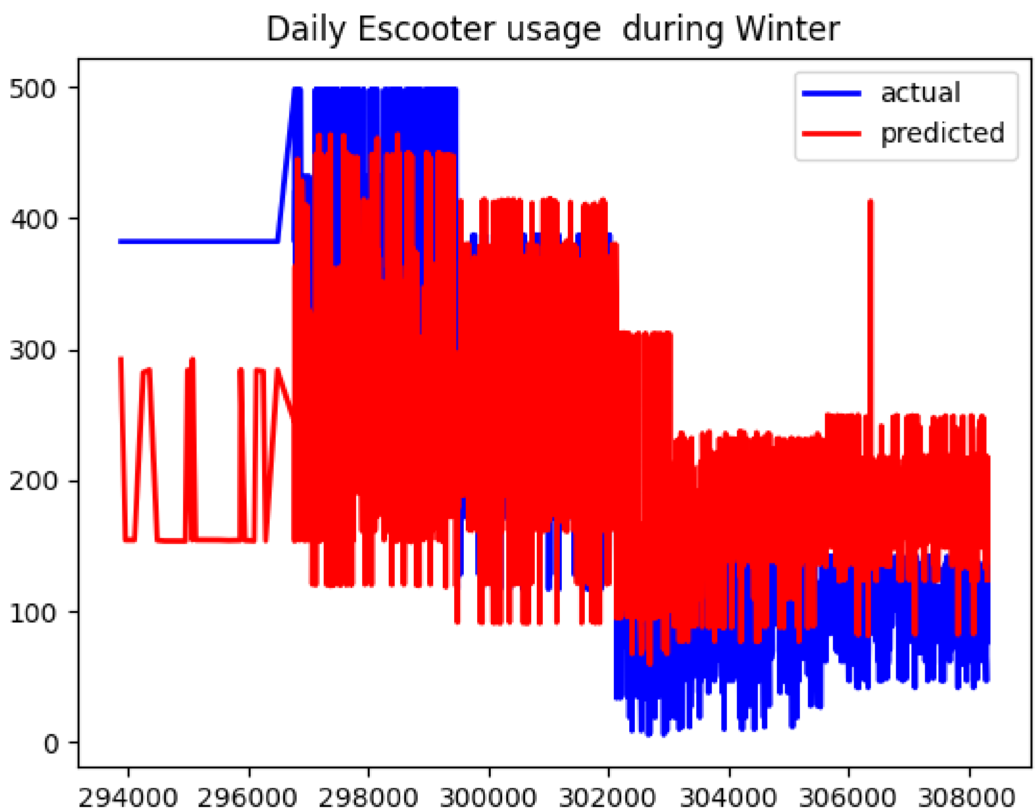
```
Monday                         0.782380
Saturday                       1.086590
Sunday                         0.920240
Thursday                       0.919715
Tuesday                        0.852127
Wednesday                      0.844216
```



Winter:

```
Pm params values and coefficients for Winter
3.1136801419958653*** (p-value: 0.0)
0.019273463302698884*** (p-value: 0.0)
0.0098577450146950066*** (p-value: 0.0)
1.2262758772672977*** (p-value: 0.0)
0.617084812562942*** (p-value: 0.0)
1.2703194521656203*** (p-value: 0.0)
-0.00014696496367674 2 (p-value: 0.831983934047788)
-9.978335628341726e-05 (p-value: 0.14635494798033666)
0.5533749686540863*** (p-value: 0.0)
0.44420874449789105*** (p-value: 0.0)
0.42642110116660187*** (p-value: 0.0)
0.30877373702607813*** (p-value: 0.0)
0.5857853286640436*** (p-value: 0.0)
0.42614223983139915*** (p-value: 0.0)
0.36897402215576663*** (p-value: 0.0)
```

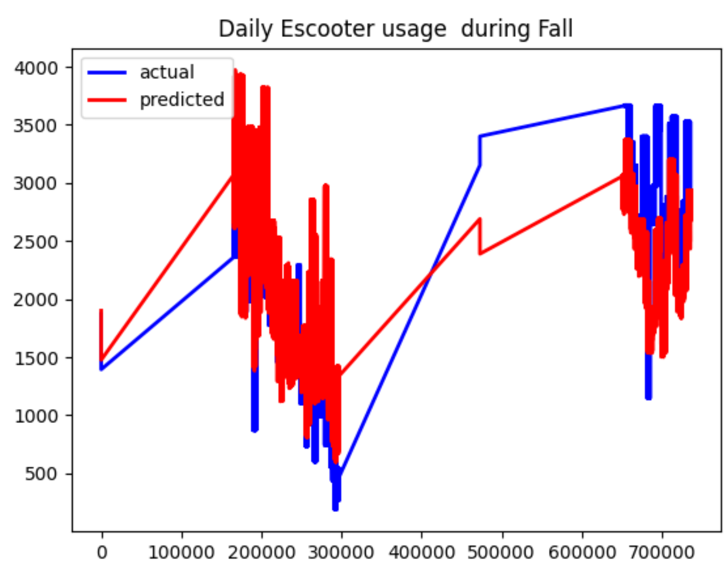
```
Coefficient Values for Winter:
const                          3.113680
TMIN                           0.019273
TMAX                           0.009858
Lime                           1.226276
Lyft                           0.617085
Spin                           1.270319
Trip Distance (miles)         -0.000147
Trip Duration (minutes)       -0.000100
Friday                         0.553375
Monday                         0.444209
Saturday                       0.426421
Sunday                         0.308774
Thursday                       0.585785
Tuesday                        0.426142
   Wednesday                      0.368974
```



Fall:

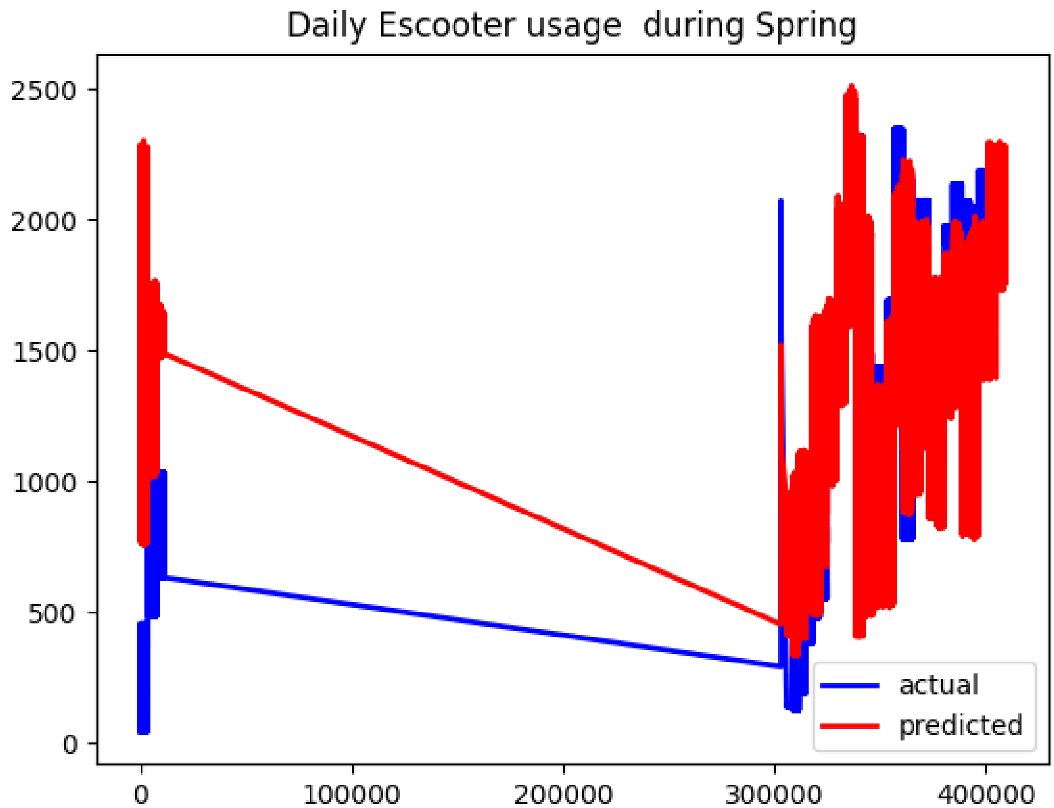
```
Pm params values and coefficients for Fall
5.05727983671337*** (p-value: 0.0)
0.018055613769844O9*** (p-value: 0.0)
0.009439838653787997*** (p-value: 0.0)
0.16929281750915884*** (p-value: 0.0)
0.13663616128727277*** (p-value: 0.0)
```

```
0.09539250668123628*** (p-value: 0.0)
0.0033060541624647327*** (p-value: 0.0)
-0.0007452759200672967*** (p-value: 0.0)
0.934012314266846*** (p-value: 0.0)
0.5636999263853814*** (p-value: 0.0)
0.9093516372344972*** (p-value: 0.0)
0.7004074980417213*** (p-value: 0.0)
0.7494560283094924*** (p-value: 0.0)
0.5297502553171125*** (p-value: 0.0)
0.6706021771583176*** (p-value: 0.0)

Coefficient Values for Fall:
const                         5.057280
TMIN                          0.018056
TMAX                          0.009440
Lime                          0.169293
Lyft                          0.136636
Spin                          0.095393
Trip Distance (miles)         0.003306
Trip Duration (minutes)      -0.000745
Friday                        0.934012
Monday                        0.563700
Saturday                      0.909352
Sunday                        0.700407
Thursday                      0.749456
Tuesday                       0.529750
Wednesday                     0.670602
```



Spring:

```
Pm params values and coefficients for Spring
4.660516742756279*** (p-value: 0.0)
-0.00845403520265825*** (p-value: 0.0)
0.03352646256641555*** (p-value: 0.0)
-0.02002677922759274*** (p-value: 0.0)
-0.2569166427142478*** (p-value: 0.0)
-0.02070919192641227*** (p-value: 0.0)
0.006213793760214917*** (p-value: 0.0)
-0.0006488520426752369*** (p-value: 0.0)
0.8061538551286288*** (p-value: 0.0)
0.5405569724194472*** (p-value: 0.0)
0.8465272214390837*** (p-value: 0.0)
0.6933119793729983*** (p-value: 0.0)
0.6281122701119982*** (p-value: 0.0)
0.5334942070576221*** (p-value: 0.0)
0.6123602372264966*** (p-value: 0.0)

Coefficient Values for Spring:
const                     4.660517
TMIN                     -0.008454
TMAX                      0.033526
Lime                     -0.020027
Lyft                     -0.256917
Spin                     -0.020709
Trip Distance (miles)     0.006214
Trip Duration (minutes)  -0.000649
Friday                    0.806154
Monday                    0.540557
Saturday                  0.846527
Sunday                    0.693312
Thursday                  0.628112
Tuesday                   0.533494
Wednesday                 0.612360
```



Poisson Regression Model Bike:

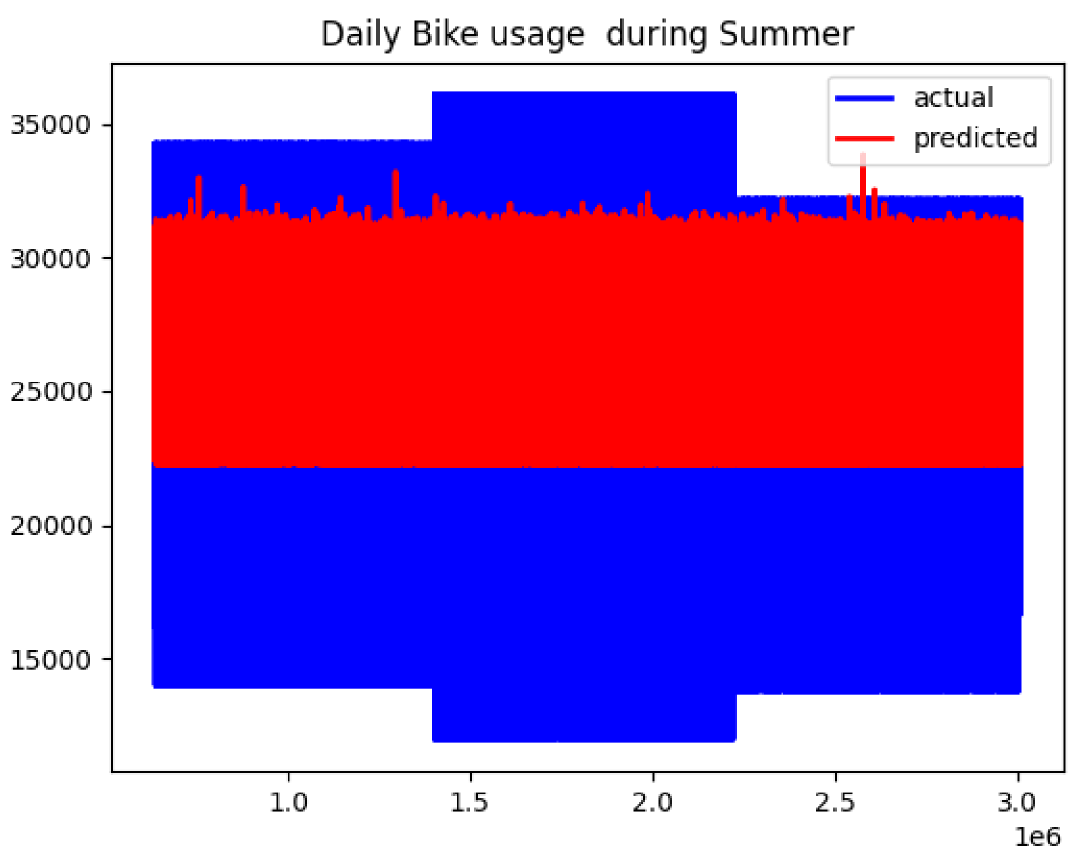
Start Date: 2022-05-01 00:00:06
End Date: 2023-04-30 23:59:05

Independent_variable: 'casual', 'member','classic_bike','docked_bike', 'electric_bike','distance','duration', 'Friday', 'Monday', 'Saturday', 'Sunday','Thursday', 'Tuesday', 'Wednesday'.
Dependent Variables: 'daily_ride_count'

Summer:

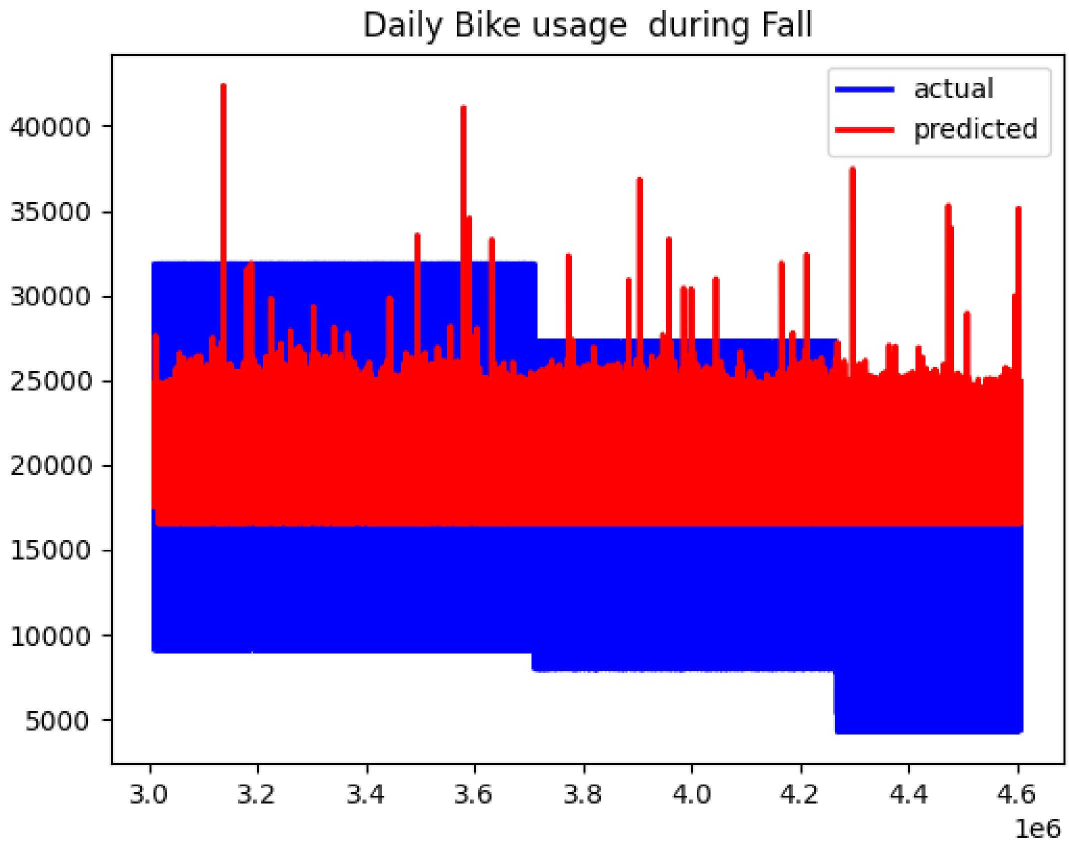
```
Bike - Pm params values and coefficients for Summer
5.1465362141454145*** (p-value: 0.0)
2.577230366201101*** (p-value: 0.0)
2.5693058479443347*** (p-value: 0.0)
1.717381972702155*** (p-value: 0.0)
1.7167151740595568*** (p-value: 0.0)
1.7124390673836998*** (p-value: 0.0)
0.0012313265231745035*** (p-value: 0.0)
6.861006638103914e-05*** (p-value: 0.0)
0.7863834669906017*** (p-value: 0.0)
```

```
0.5804779969252961*** (p-value: 0.0)
0.9015694200676407*** (p-value: 0.0)
0.7627722997752777*** (p-value: 0.0)
0.7381631480673415*** (p-value: 0.0)
0.6813928201112591*** (p-value: 0.0)
0.6957770622079957*** (p-value: 0.0)

Coefficient Values for Summer:
const             5.146536
casual            2.577230
member            2.569306
classic_bike      1.717382
docked_bike       1.716715
electric_bike     1.712439
distance          0.001231
duration          0.000069
Friday            0.786383
Monday            0.580478
Saturday          0.901569
Sunday            0.762772
Thursday          0.738163
Tuesday           0.681393
Wednesday         0.695777
```



Fall:

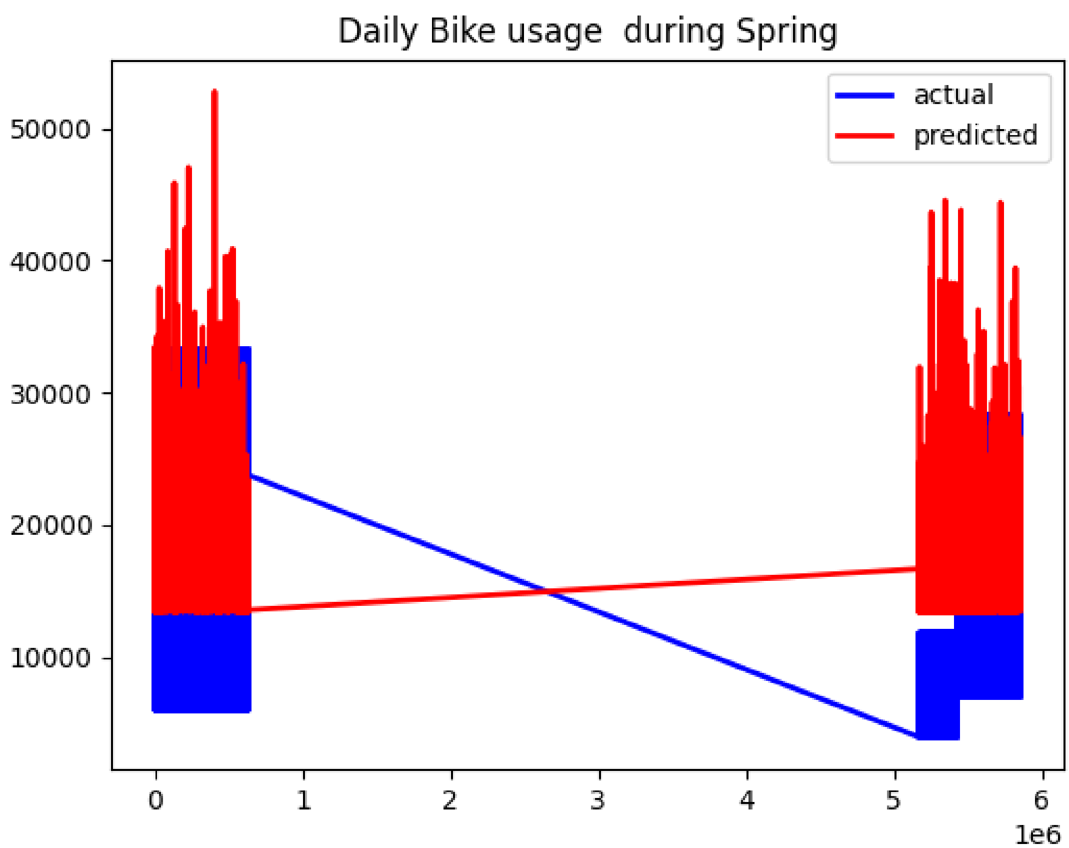
```
Bike - Pm params values and coefficients for Fall
5.008189789477836*** (p-value: 0.0)
2.53283649743067*** (p-value: 0.0)
2.475353292047169*** (p-value: 0.0)
1.6800794478728736*** (p-value: 0.0)
1.6705020311552923*** (p-value: 0.0)
1.6576083104496635*** (p-value: 0.0)
4.3661786645179525e-05*** (p-value: 0.0)
0.0003911066413391712*** (p-value: 0.0)
0.7887689586447679*** (p-value: 0.0)
0.5731000982465457*** (p-value: 0.0)
0.9035633298298623*** (p-value: 0.0)
0.6382899154480821*** (p-value: 0.0)
0.7856963407218975*** (p-value: 0.0)
0.6080896942145722*** (p-value: 0.0)
0.710681452372106*** (p-value: 0.0)

Coefficient Values for Fall:
const            5.008190
casual           2.532836
member           2.475353
classic_bike     1.680079
docked_bike      1.670502
electric_bike    1.657608
distance         0.000044
duration         0.000391
Friday           0.788769
Monday           0.573100
Saturday         0.903563
Sunday           0.638290
Thursday         0.785696
Tuesday          0.608090
Wednesday        0.710681
```



Spring:

```
Bike - Pm params values and coefficients for Spring
4.951446636522401*** (p-value: 0.0)
2.553005617731559*** (p-value: 0.0)
2.3984410187908445*** (p-value: 0.0)
1.6840650656256662*** (p-value: 0.0)
1.694465652712321*** (p-value: 0.0)
1.572915918184421*** (p-value: 0.0)
0.017428288037815*** (p-value: 0.0)
0.0005041142640115825*** (p-value: 0.0)
0.695368666811402*** (p-value: 0.0)
0.720880387005963*** (p-value: 0.0)
0.8468495634380805*** (p-value: 0.0)
0.6612598126440307*** (p-value: 0.0)
0.7576799146677214*** (p-value: 0.0)
0.689981686361476*** (p-value: 0.0)
0.5794266055937308*** (p-value: 0.0)

Coefficient Values for Spring:
const              4.951447
```

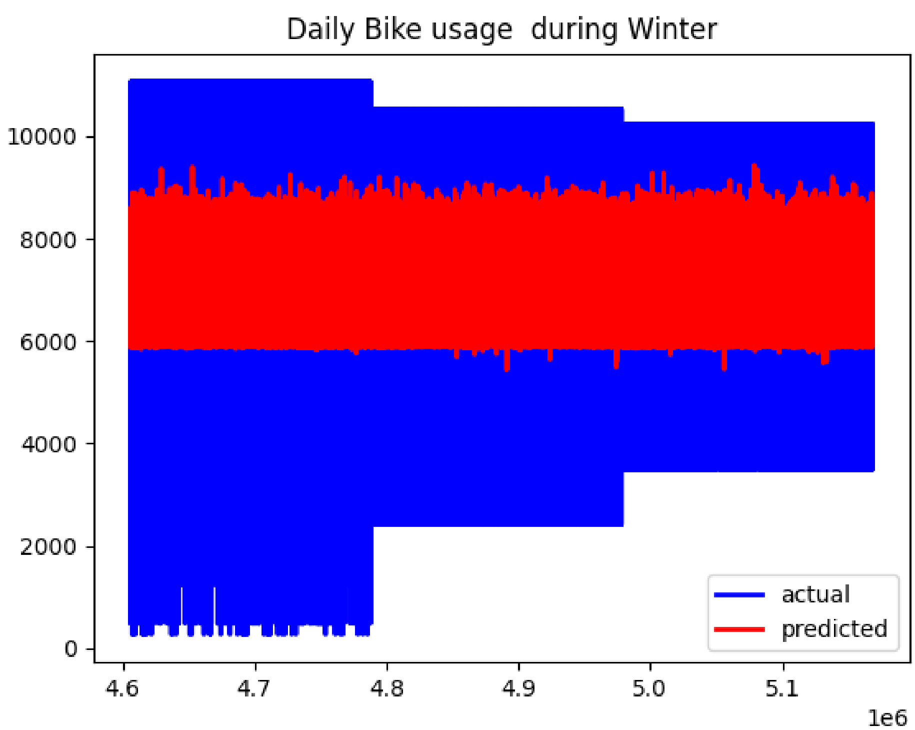
```
casual           2.553006
member           2.398441
classic_bike     1.684065
docked_bike      1.694466
electric_bike    1.572916
distance         0.017428
duration         0.000504
Friday           0.695369
Monday           0.720880
Saturday         0.846850
Sunday           0.661260
Thursday         0.757680
Tuesday          0.689982
Wednesday        0.579427
```



Winter:

```
Bike - Pm params values and coefficients for Winter
4.481397174185188*** (p-value: 0.0)
2.2529641706056576*** (p-value: 0.0)
2.228433003579533*** (p-value: 0.0)
1.4904405966155831*** (p-value: 0.0)
```

```
1.50231773833323*** (p-value: 0.0)
1.4886388392363725*** (p-value: 0.0)
0.007987083464319263*** (p-value: 0.0)
-6.033225364364872 5e-05*** (p-value: 0.0)
0.5798257000621494*** (p-value: 0.0)
0.6601351718696283*** (p-value: 0.0)
0.47914717827087566*** (p-value: 0.0)
0.5403547791333281*** (p-value: 0.0)
0.664610100636078*** (p-value: 0.0)
0.8189380160836476*** (p-value: 0.0)
0.7383862281294805*** (p-value: 0.0)
Coefficient Values for Winter:
const            4.481397
casual           2.252964
member           2.228433
classic_bike     1.490441
docked_bike      1.502318
electric_bike    1.488639
distance         0.007987
duration        -0.000060
Friday           0.579826
Monday           0.660135
Saturday         0.479147
Sunday           0.540355
Thursday         0.664610
Tuesday          0.818938
Wednesday        0.738386
```



Poisson Regression:

Poisson regression is a type of regression analysis used to model count data where the dependent variable is a count of occurrences within a fixed time or space interval. It's commonly used when the dependent variable represents counts of events or occurrences and follows a Poisson distribution. Poisson regression models the relationship between the independent variables and the expected counts of the dependent variable.

Function poisson_regression(df, season_type): This function performs Poisson regression analysis. Here's what it does:

Data Preprocessing:

It preprocesses the input DataFrame df to prepare it for regression analysis. This includes creating dummy variables for categorical variables like day_of_week and Vendor.

Model Fitting:

It fits a Poisson regression model using the sm.GLM() function from the statsmodels library. The dependent variable y is daily_ride_count, and the independent variables x include weather-related variables (TMIN, TMAX), e-scooter company variables (Lime, Lyft, Spin), trip characteristics (Trip Distance (miles), Trip Duration (minutes)), and day of the week variables.

Model Summary and Coefficients: It prints the summary of the fitted Poisson regression model, including the coefficients, standard errors, z-scores, and p-values. It prints the coefficients and their significance levels based on p-values.

Root Mean Squared Error (RMSE): It calculates and prints the RMSE value of the Poisson regression model, which measures the difference between predicted and actual counts of daily rides.

Function pr_predict(x, y, weather):

This function is responsible for predicting daily e-scooter usage based on the Poisson regression model. Here's what it does:

Train-Test Split: It splits the data into training and testing sets using an 80:20 ratio.

Model Fitting and Evaluation: It fits a Poisson regression model to the training data using sm.GLM() and evaluates the model performance on both the training and testing datasets. It calculates and prints the RMSE values for both the training and testing datasets.

Visualization: It visualizes the actual and predicted daily e-scooter usage using a line plot. It saves the plot as an image file.

Note: The provided functions perform Poisson regression analysis to estimate daily ride counts for both bikes and e-scooters based on various independent variables such as weather conditions, trip characteristics, and day of the week. The analysis helps understand the factors influencing ride sharing and predicts daily e-scooter usage based on the fitted regression models.

Performance of the Model based on Coefficients and Significance:

Coefficient Values:

- The coefficient values represent the change in the daily ride count for a one-unit change in the corresponding independent variable, holding all other variables constant.
- For example, a one-unit increase in casual daily bike rentals is associated with an increase of approximately 2.53 rides per day, holding other variables constant.
- Similarly, a one-unit increase in distance is associated with an increase of approximately 0.000044 rides per day, holding other variables constant.

Significance (P-values):

- The coefficient values represent the change in the daily ride count for a one-unit change in the corresponding independent variable, holding all other variables constant.
- The p-values indicate the significance of each coefficient. A low p-value (typically less than 0.05) suggests that the coefficient is statistically significant, meaning that the variable has a significant effect on the daily ride count.
- In this case, all p-values are reported as 0.0, which implies that all coefficients are statistically significant at conventional levels of significance.

Interpretation for Bikes:

Based on the coefficients and their significance (Spring):

The number of casual and member bike rentals (casual, member) have significant positive effects on the daily ride count. This suggests that increasing the number of casual and member bike rentals leads to more daily rides.

Similarly, the type of bike (classic_bike, docked_bike, electric_bike) also has significant positive effects on the daily ride count.

Variables related to the day of the week (Friday, Monday, Saturday, Sunday, Thursday, Tuesday, Wednesday) also have significant effects on the daily ride count, with different magnitudes depending on the day.

Interpretation for E-scooter (Spring):

The presence of e-scooter companies such as Lime, Lyft, and Spin (Lime, Lyft, Spin) has significant effects on the daily ride count. Negative coefficients suggest that an increase in the number of scooters from these companies leads to fewer daily rides.

Temperature variables (TMIN, TMAX) also have significant effects on the daily ride count, with negative coefficients suggesting that higher temperatures may lead to fewer daily rides.

Variables related to the day of the week (Friday, Monday, Saturday, Sunday, Thursday, Tuesday, Wednesday) also have significant effects on the daily ride count, with positive coefficients indicating higher ride counts on certain days compared to others.