

Analysis Review

1. Logistic Regression:

Categorical Features: *member_casual* , *season*

Continuous Features: *TMAX*

Target Variable: *rideable_type*

NaN Values: dropped all the rows with null values

Logistic Regression Accuracy: 0.64

Logistic Regression Confusion Matrix: $\begin{bmatrix} 1661458 & 0 \\ 948894 & 0 \end{bmatrix}$

The model correctly predicted the negative class (possibly 'non-bike') 1,661,458 times. The model incorrectly predicted the negative class 948,894 times when it should have predicted the positive class. The model did not correctly predict any true positive cases (possibly 'bike'), indicated by the 0 in this cell.

Categorical Features: *member_casual* , *season*

Continuous Features: *TMAX*

Target Variable: *rideable_type*

NaN Values: Handled NaN Values.

Logistic Regression Accuracy: 0.53

Logistic Regression Confusion Matrix: $\begin{bmatrix} 660353 & 1005389 \\ 615642 & 1134883 \end{bmatrix}$

The model has a fairly balanced distribution of predictions across the positive and negative classes but with substantial error rates. The model struggles to do so accurately, indicating potential issues with feature selection.

Categorical Features: *member_casual* , *day_of_week*

Continuous Features: *TMAX*

Target Variable: *rideable_type*

Logistic Regression Accuracy: 0.64

Logistic Regression Confusion Matrix: $\begin{bmatrix} 1661307 & 151 \\ 948781 & 113 \end{bmatrix}$

This model primarily predicts the negative class, with an extremely low number of true positives, which suggests a significant bias towards the negative class. This could be indicative of class imbalance or an inability of the chosen features (`member_casual`, `day_of_week`) to adequately differentiate between the classes.

Categorical Features: *member_casual* , *day_of_week*, *season*

Contínuos Features: *TMAX*, *Elevation_Change*

Target Variable: *rideable_type*

Logistic Regression Accuracy: 0.64

Logistic Regression Confusion Matrix: $\begin{bmatrix} 1661032 & 426 \\ 948390 & 504 \end{bmatrix}$

Still correctly predicting the negative class most of the time. Slightly higher than the previous model, indicating a marginal increase in incorrect positive predictions. An increase in correct positive predictions, but still very low.

Coefficients from the logistic regression model on the *rideable_type* prediction:

Coefficient

member_casual_casual	-0.041679
member_casual_member	0.031800
season_Fall	-0.807996
season_Spring	-0.075316
season_Summer	-0.033264
season_Winter	0.035987
day_of_week_Friday	-0.339277
day_of_week_Monday	-0.034578
day_of_week_Saturday	-0.046600
day_of_week_Sunday	-0.047978
day_of_week_Thursday	-0.174134
day_of_week_Tuesday	0.000109
day_of_week_Wednesday	-0.041014
Elevation_Change	-0.150065
Distance	-0.146041
trip_duration	0.022835
TMAX	-0.011860
TMIN	0.022744

```
{0: 'electric', 1: 'classic'}
```

The coefficients from the logistic regression model elucidate the relationship between various features and the log-odds of predicting the rideable type, assumed to be the positive class ('1').

Here's a concise overview: Member type coefficients suggest that casual riders are less likely than members to choose the target rideable type, as indicated by a negative coefficient for casual riders and a positive one for members.

Seasonal analysis shows a strong decrease in the likelihood of choosing the positive class during fall, with minor decreases in spring and summer, and a slight increase in winter.

Daily variations reveal that Fridays, Mondays, Saturdays, Sundays, and Thursdays all see decreases in the likelihood of selecting the positive class, with Tuesday showing no significant change and Wednesday a minor decrease.

Other factors such as elevation change and distance show negative impacts on the choice probability, indicating that greater distances and elevation changes decrease the likelihood of choosing the positive class, whereas trip duration and higher minimum temperatures slightly increase it.