Predicting Vehicle Collision Severity

Capstone Project

Objectives?

❖   How can this data be used to predict the severity of the collisions occurring within Seattle?

❖   A model that predicts vehicle collision to a reasonable degree under given conditions.

❖   Businesses & city transportation can have an interest in this problem to maintain customer and employee safety.

Provided!

- Given the parameters, the goal is to provide them with trends and similar situations that have higher Accident Probabilities.

- Collision Data was obtained from the Seattle Police Department and Traffic Records from 2004 – Present.

- A total of 194,673 rows and 38 columns were obtained from the raw data.

- Columns with excessive amounts of missing data & did not directly pertain to the problem were deleted.

Factors!

Here We can see that Weather does not play a major role in terms of Collision ratio.
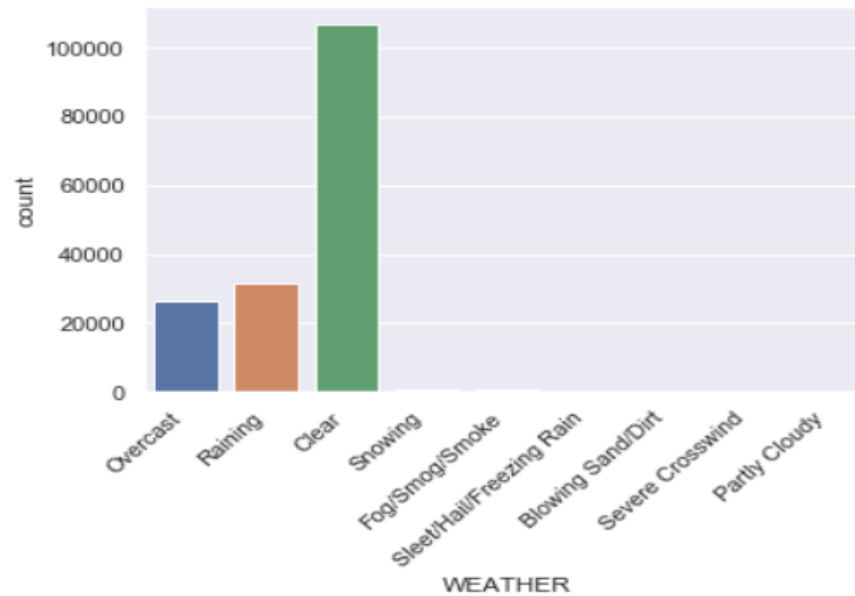
However, Night-time is the second highest cause of collisions weather-wise. In this case, Night-time with Street lights on came in second. Again, no adverse weather conditions causing a higher rate of collisions.
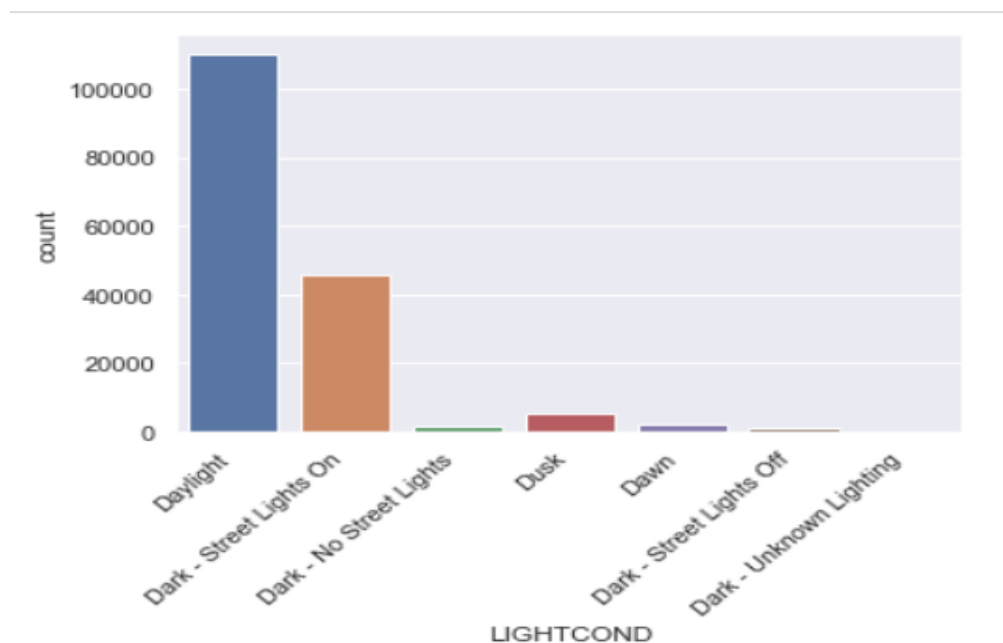
The graph shows the Different Factors of the Road like Wet, Dry, Snow, Ice and So on which Effects the Collision Ratio
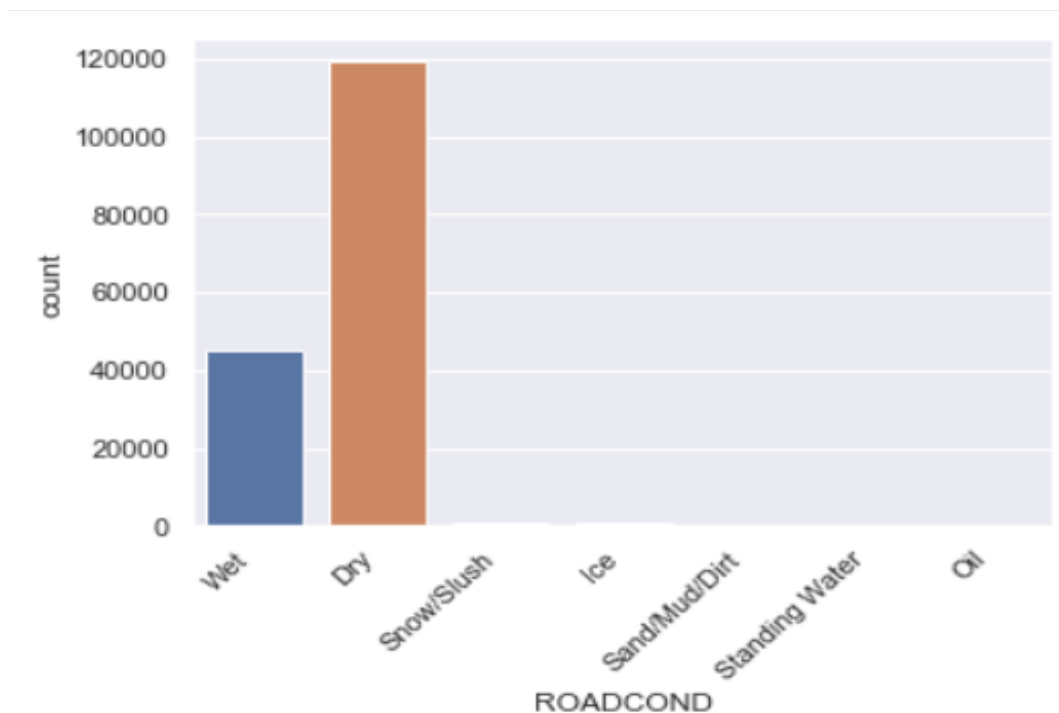
Findings.

The top three conditions parked car, angles, and rear-ended are most probable. Thus, more safety precautions should be placed within common streets.

Here we see that most collisions only involve two people with the Majority of them not sustaining a major personal injury.

Discussion

With the data given it was shown that it is possible to predict the severity code using the model.

Due to a Large number of missing data on potentially useful features that correlate with collisions, the model may be limited and may not cover the necessary situations.

A Linear regression model created a score of 0.73 – With a recall of avg of 0.73.
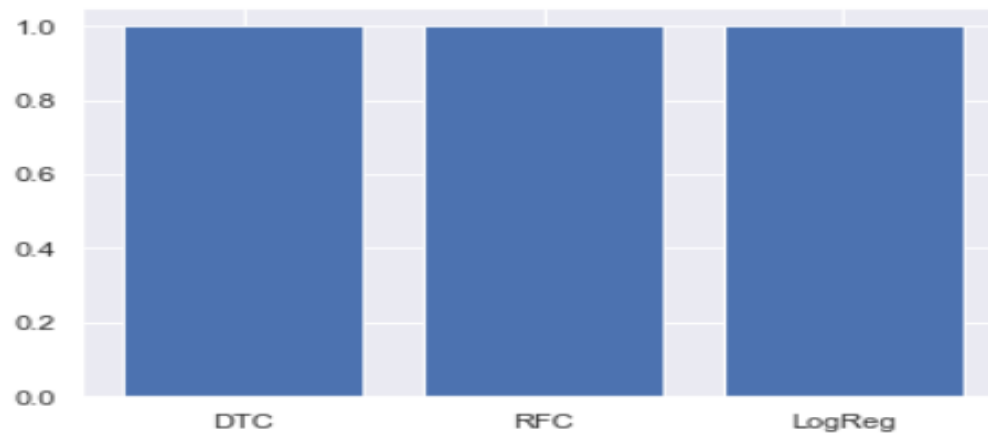
– Precision avg of 0.73

A Logistic Regression model created a score of 0.73 – With a recall of avg of 0.73.

– Precision avg of 0.75

A Random Forest Model created a score of 0.73 – With a recall of avg of 0.73.

– Precision avg of 0.75

Comparison of Models

Using the results shown, the most probable cases come from daylight weather, followed by night with Street Lights on. This is correlated to the high Ammount of parked cars hit that usually are on the shoulders of streets.

Seeing as days leading up to the weekend are the most prone to have a collision the public should be extra wary of High-Density areas during those times to maintain their safety.

Improvements for this situation would be to include more features such as:

Vehicle Malfunction

Age of those involved

Experience Driving a Vehicle

Whether or not an Infraction was committed

Population density within the City