

Predicting Vehicle Collision Severity

Capstone Project



Problem & Target Audience

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

Predicting Seattle Collision Severity

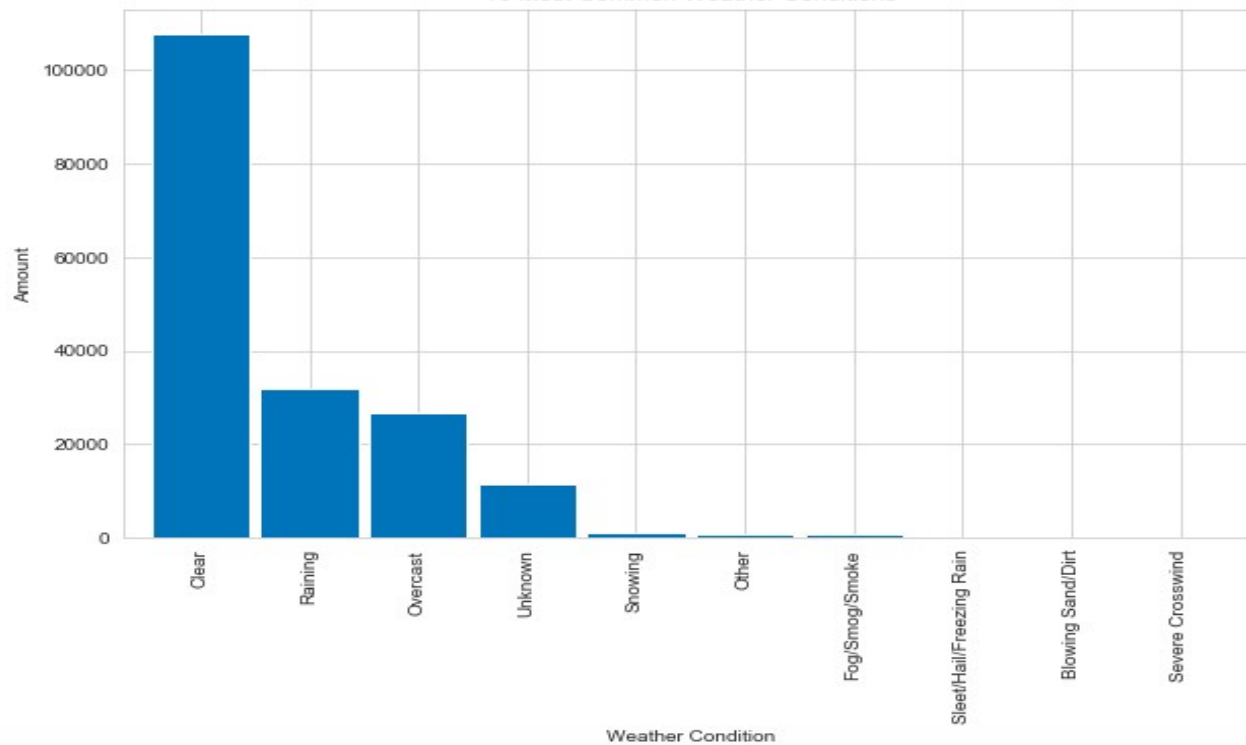
- A model that predicts vehicle collision to a reasonable degree under given conditions.
- Businesses & city transportation can have interest in this problem to maintain customer and employee safety.
- Given the parameters, the goal is to provide them with trends and similar situations that have higher accident probabilities.

Data Acquisition & Cleaning

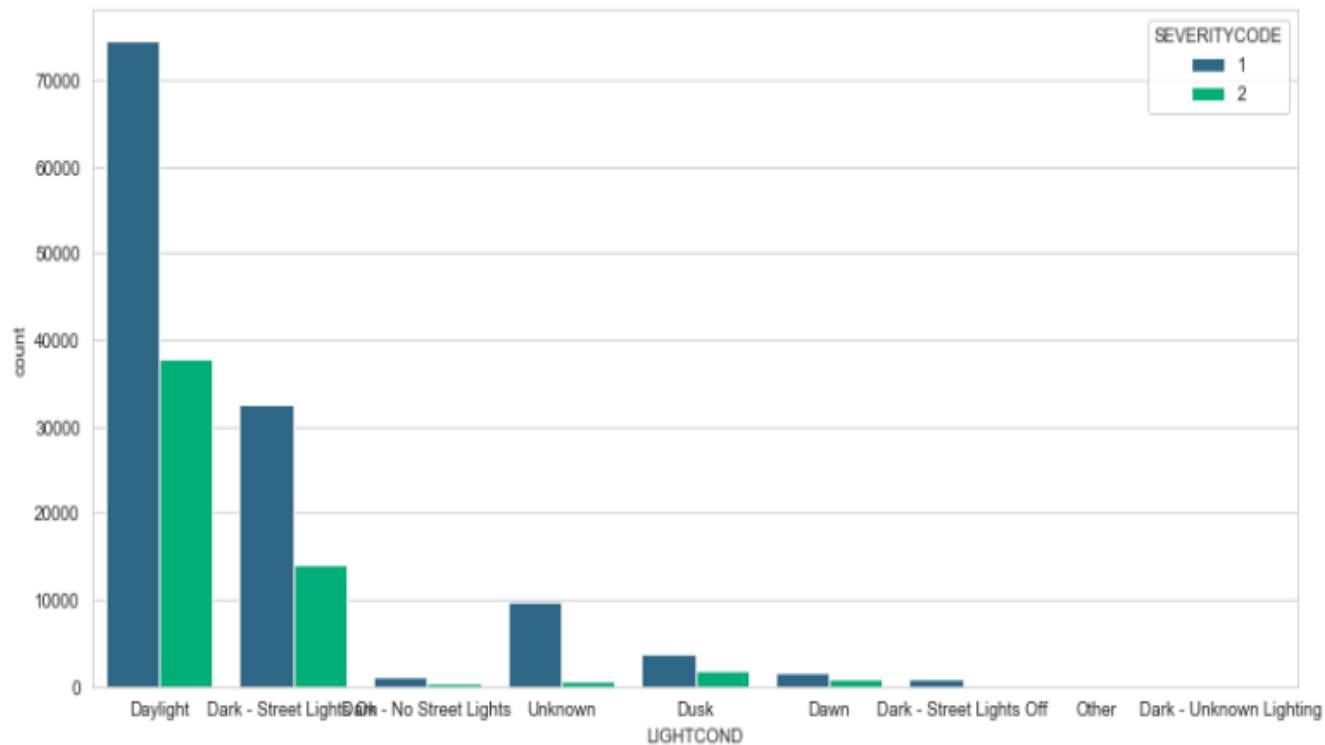
- Collision data was obtained from the Seattle Police Department and Traffic Records from 2004 – Present.
- A total of 194,673 rows and 38 columns was obtained from the raw data.
- Columns with excessive amounts of missing data & did not directly pertain to the problem were deleted.

.

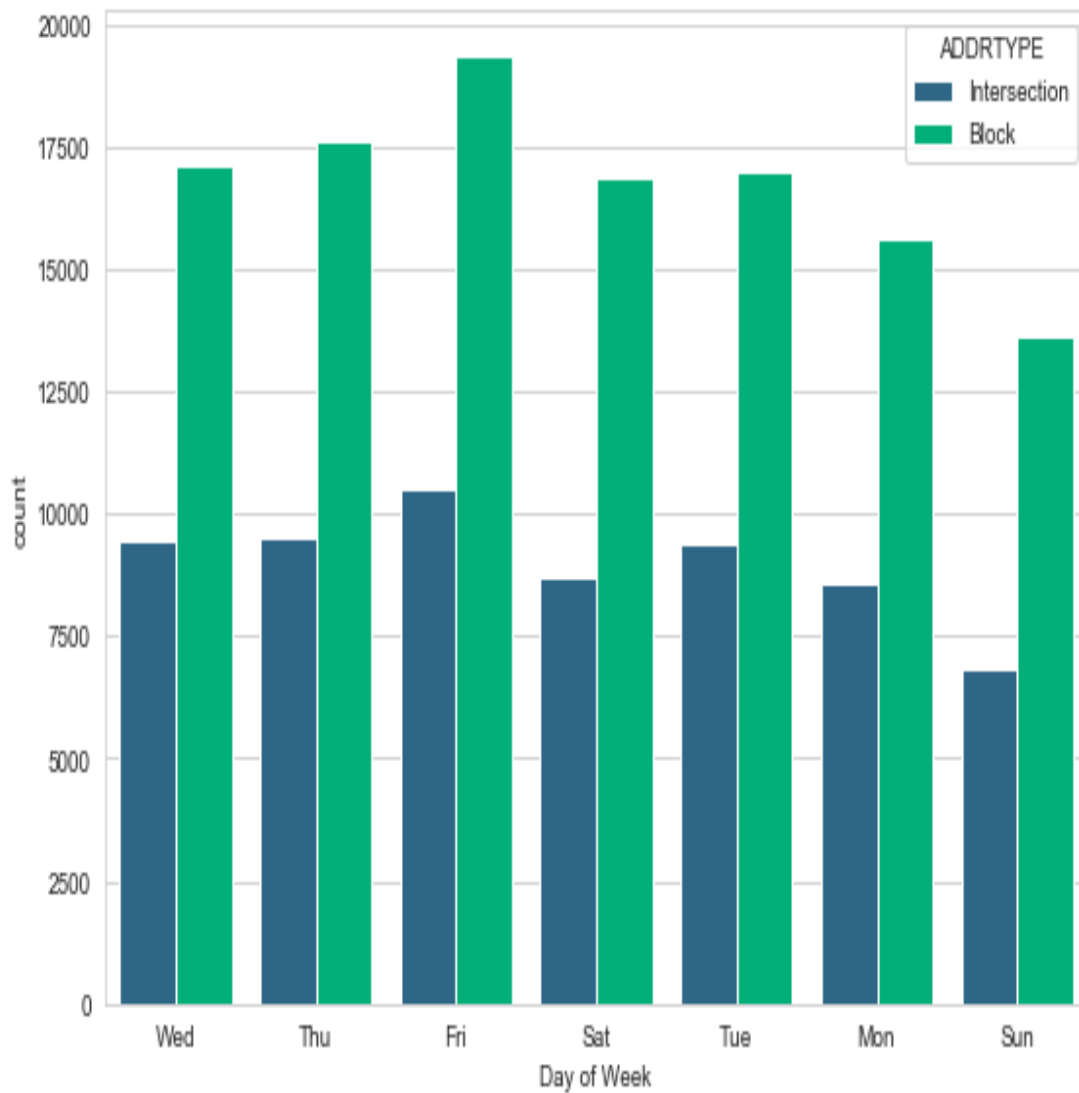
10 Most Common Weather Conditions



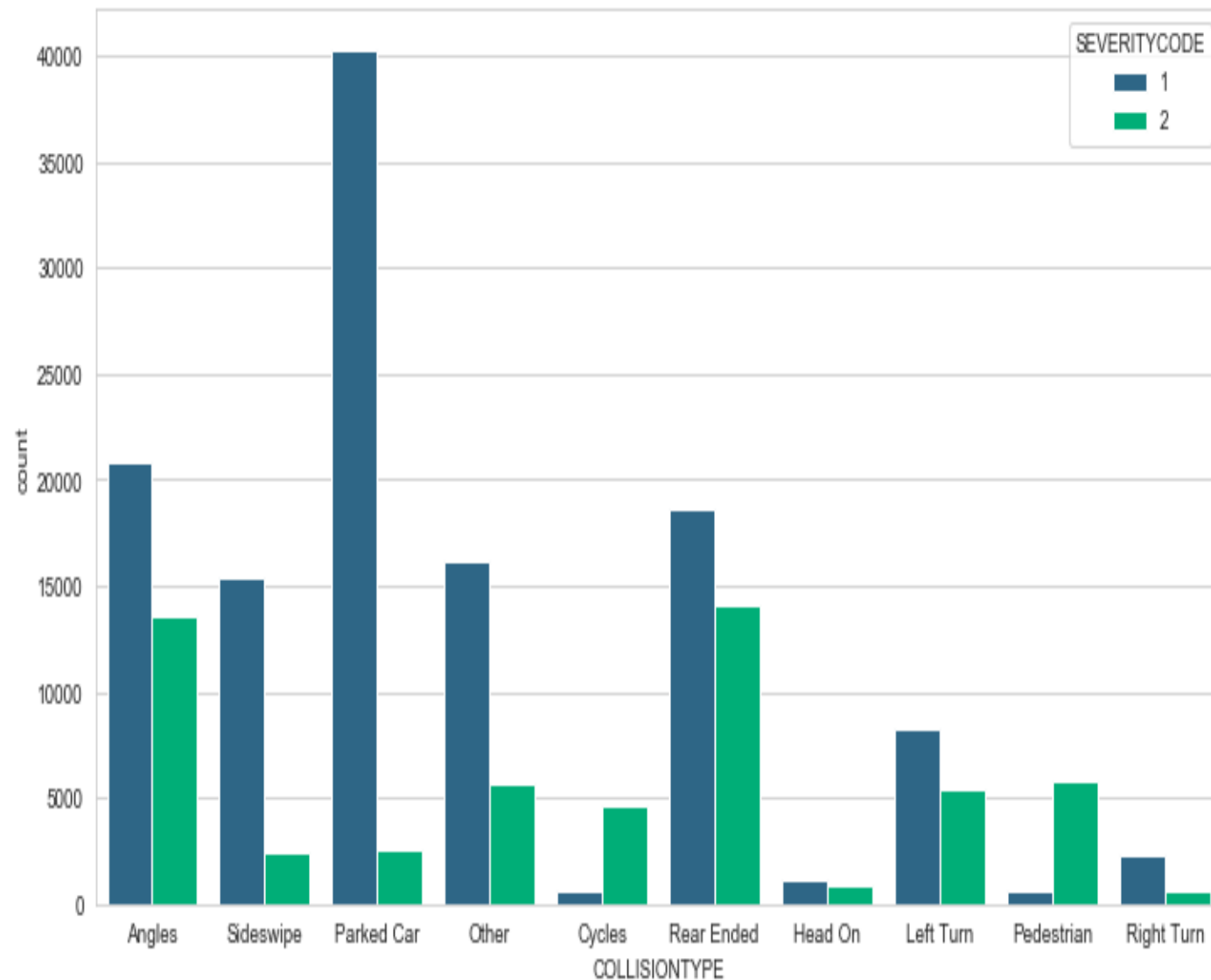
- Here we are able to see that weather does not play a major role since most prominent weather condition and not a low visibility situation



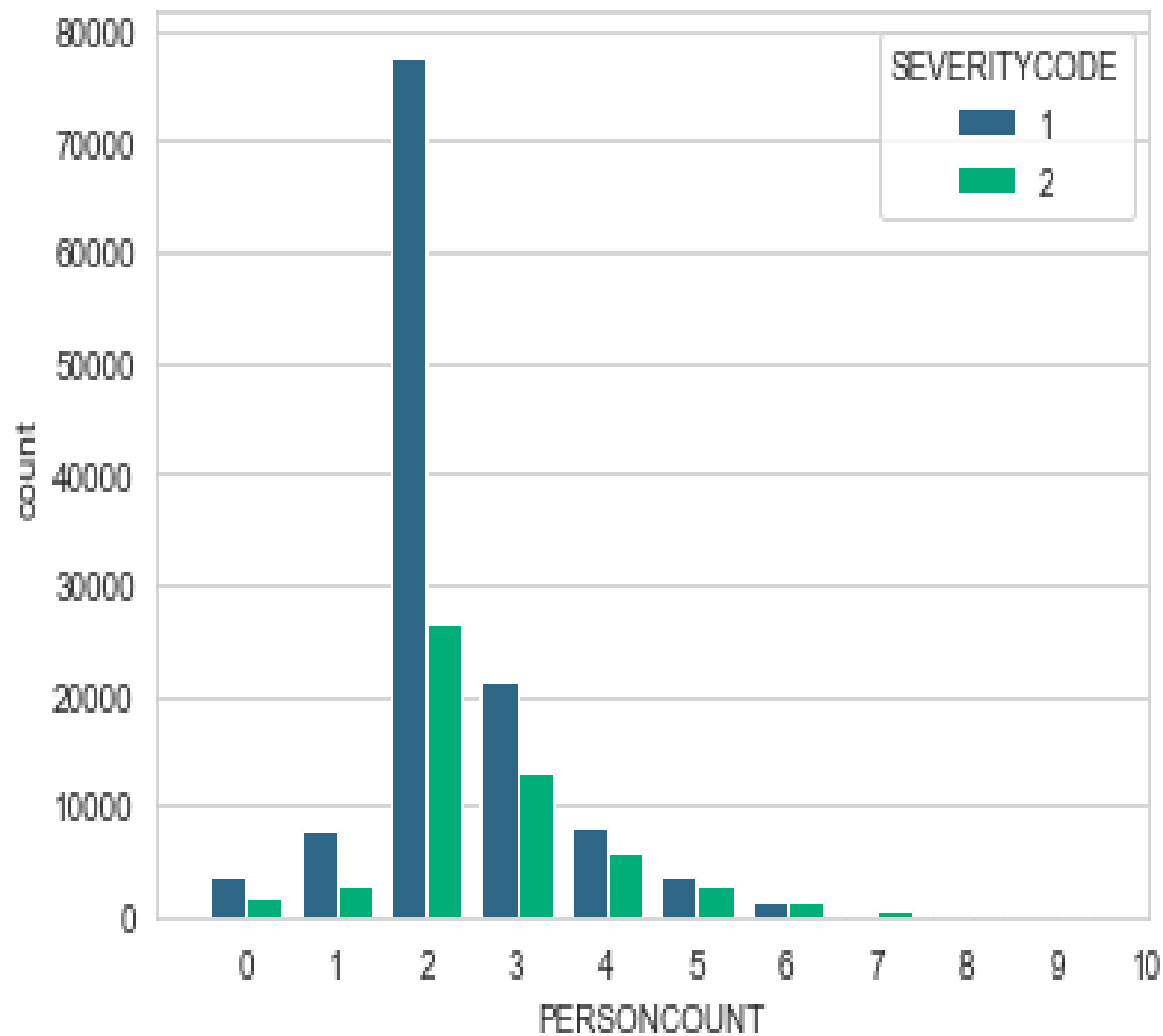
- However, nighttime is the second highest cause of collisions weather wise. In this case, nighttime with street lights on came in second. Again, no adverse weather conditions causing a higher rate of collisions.



- In this case, we are able to see the amount of collision per day with an emphasis on address type given by intersection or block.
- The graph shows that both Thursday and Friday have the highest number of collisions which does align with accepted days for social gatherings that raise the population density within a given area.



- We can conclude that this graph also supports the previous that most collisions occur within a block.
- 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 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 the 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 logistic regression model created a score of 0.74
 - With a recall of avg of 0.74.
 - Precision avg of 0.75
- The model worked reasonably well predicting the majority of the cases as code 1. This is due to the fact that there were 2.38 times more code 1 cases than code 2.

Discussion & Conclusion

- Using the results shown, the most probably cases come from daylight weather, followed by night with street lights on. This is correlated to the high amount 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 weary 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