

Accident Severity Project

Coursera Project

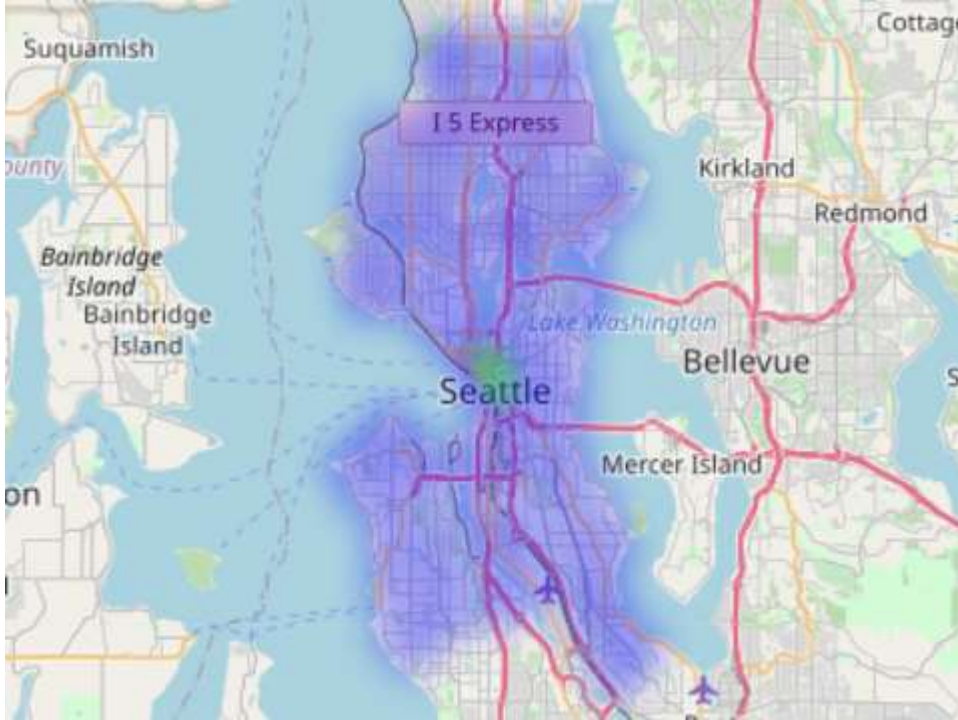
Predicting accident severity is important to save lives

- Generally first responders have to respond to calls made by civilians and at times the severity of the accident can be unclear which can lead in adequate amount of help reaching to the victims.
- Predicting accident severity can: -
 - Help first responders be more prepared to deal with the situation.
 - Help understand various factors that effect the severity of the accident.
- Besides that another objective of this report was to identify areas where more precautions can be taken

Data Acquisition And Cleaning

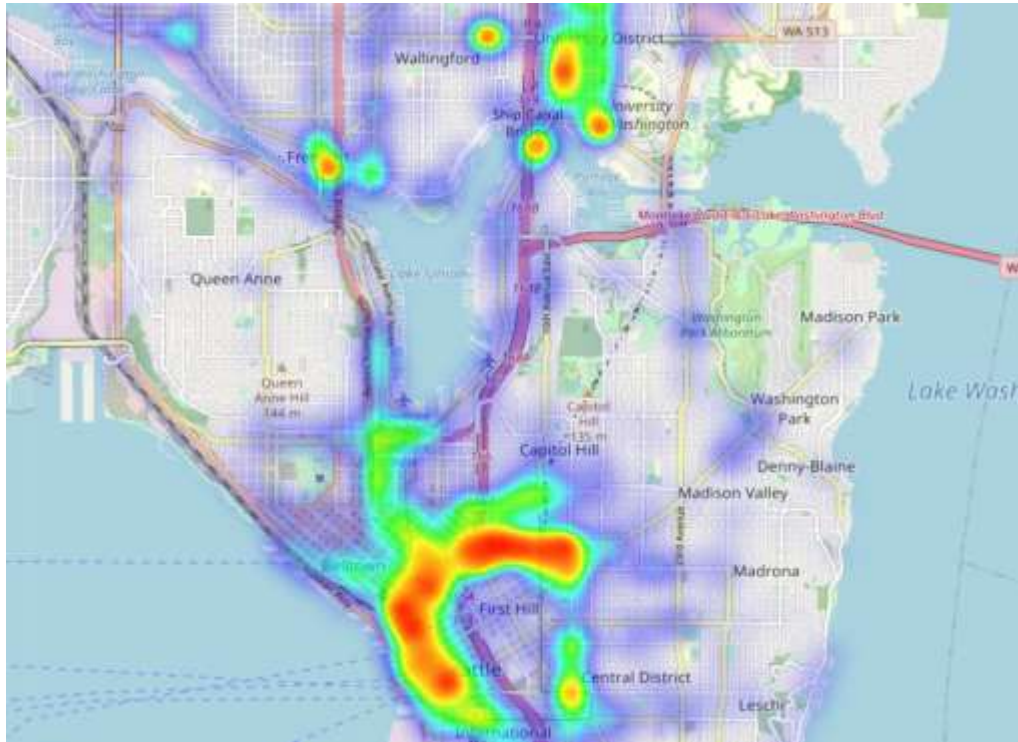
- The entire dataset was downloaded from the Seattle Geodata website and can be downloaded from [here](#).
- The raw data contained 221,525 rows and 40 columns
- Variable with unique values and duplicate values were dropped
- Finally, we were left with 188,255 rows and 17 independent variables for the algorithm development.

Area With High Density Of Accidents



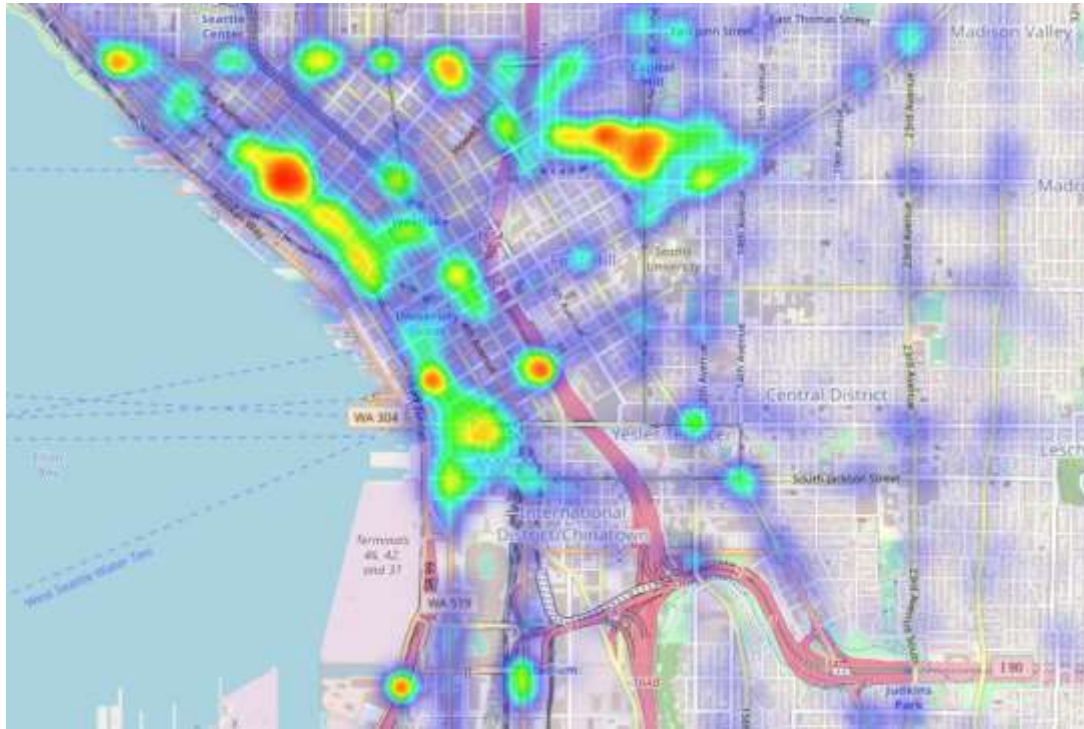
The map is a heatmap for collisions in Seattle for the year in 2019 and is similar to all such heatmaps for every year from 2004 to 2019 which indicates that to reduce the number of collisions in Seattle, the local government will have to put in more resources in the Downtown area.

Areas With High Density Of Collisions Involving Cyclist



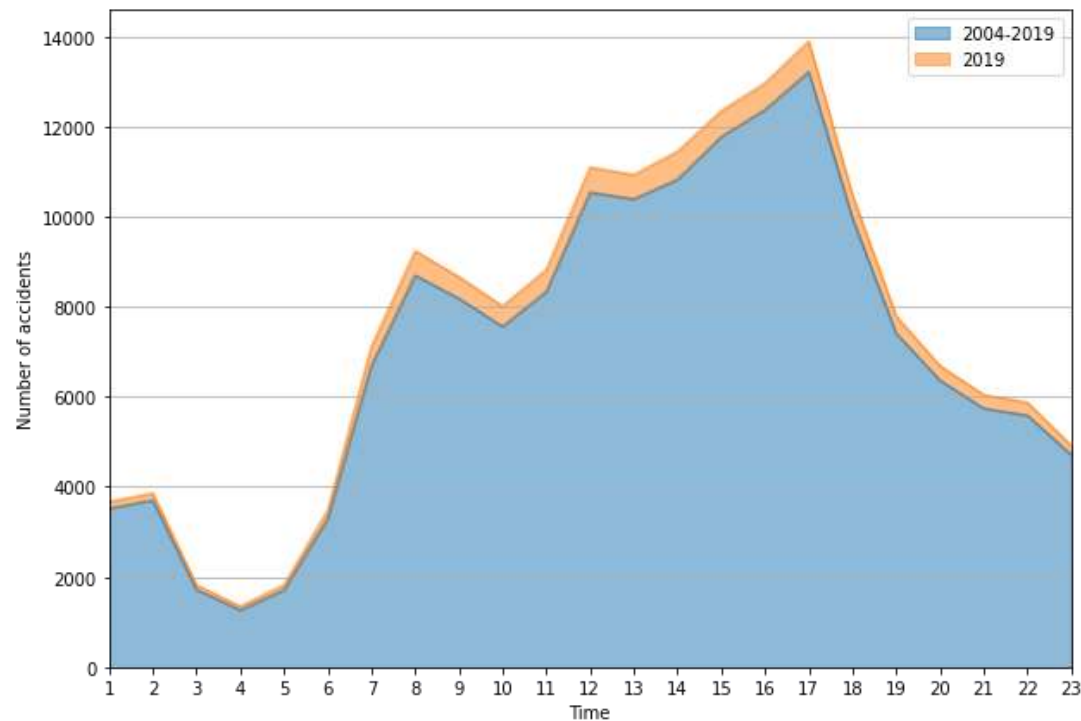
The map is a heatmap showing areas with most cycle related collisions from the year 2004 to 2019. According to the Seattle's government website the arc in the downtown area has protected bike lanes, so this raises the question whether these lanes are enough ?

Areas With High Density Of Accidents Where Intoxication Was A Factor



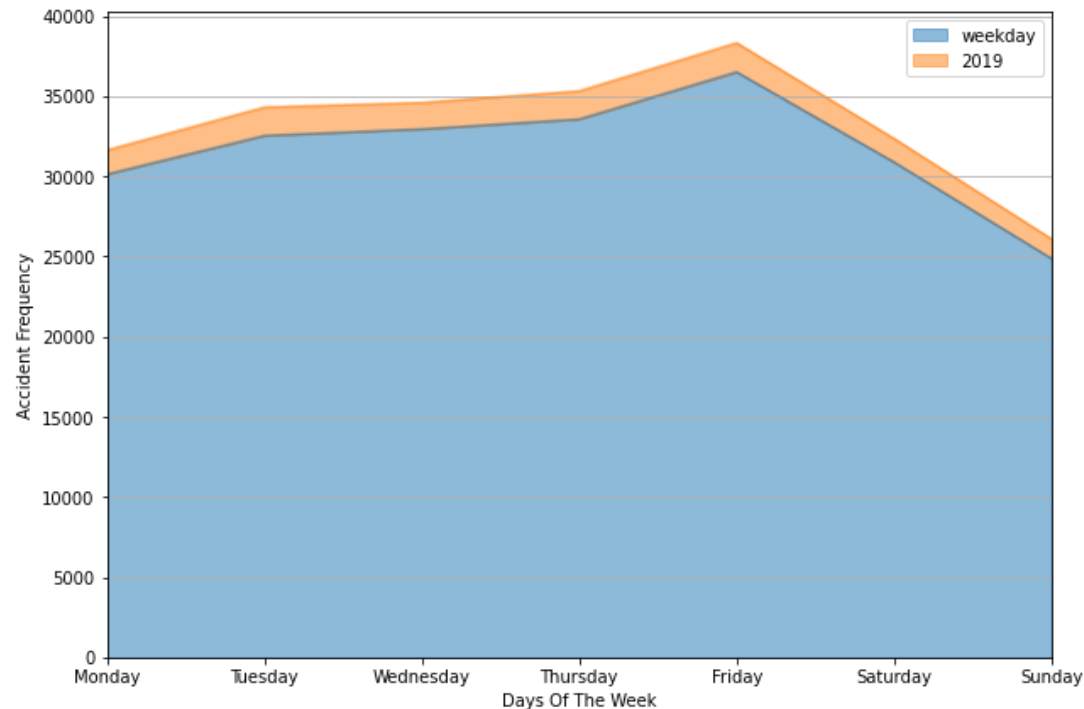
The map is a heatmap showing areas where intoxication was a factor that led to the collision and a quick search on the internet will also show that these areas have a high density of bars and restaurants as well. So deploying more resources here in the evening will help reducing these numbers.

Relationship Between Accidents And Time



According to NHTSA (National Highway Traffic Administration), most accidents tend to occur between 3 pm to 6 pm. This is often because of the high volume of vehicles on the road. The following graph for the city of Seattle shows a similar trend with accidents peaking around 5pm.

Relationship Between Accidents And Day Of The Week



According to NHTSA Saturday is considered to be the most dangerous day to drive in USA but according to the graph shown below people are more likely to get into a road accident on weekdays with Friday and not Saturday being the most dangerous day.

Result Of The Prediction Models

Algorithm	F1 Score(weighted)	Jaccard Score
Decision Tree Classifier	0.682782	0.735279
K Nearest Neighbors	0.681620	0.715094
Naïve Bayes	0.596810	0.654910

Conclusion

The Weighted accuracy of our models, lies in the range of 0.58 to 0.69.
the steps that can be taken to increase it: -

- Collecting more data.
- Better data collection to decrease the number of missing values as multiple columns had to be removed just because they had a high number of missing values.
- More extensive hyper parameter tuning.