Seattle Accidents Severity Prediction

Sarveswara Rao Basa

Predicting Weather impact on Severity & Visually locating High incident Loc

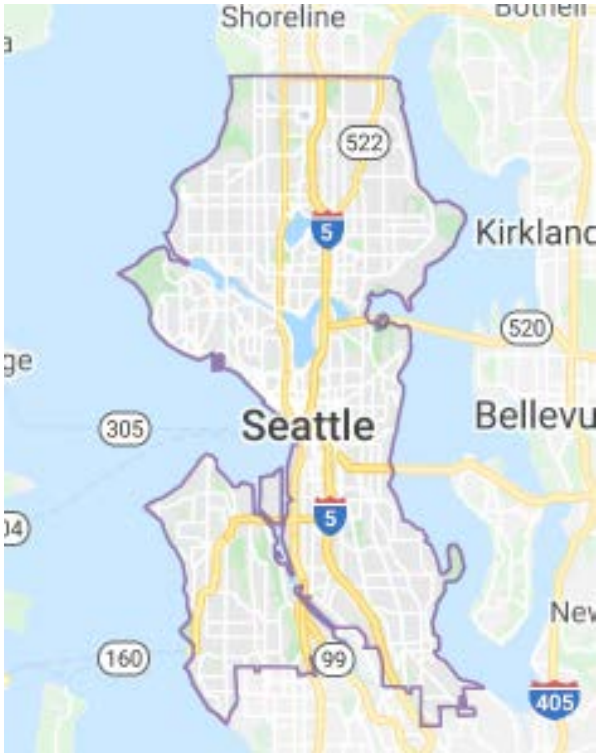
- Predicting the severity of an accident due to weather impact is useful for 911 dispatchers
- Useful in mobilizing additional resources proactively using predicted model will save lives

- Town Planners will be able to locate the high incidents zones for additional infrastructure consideration

Data Acquisition and cleaning

- Sample dataset provided for Seattle Car accidents with severity
- Total 194,673 rows and 38 features in the raw data set
- Only 2020, 2019 years data is selected for modeling
- After cleaning and dropping the not required features , 9896 rows and 14 features left in the dataset

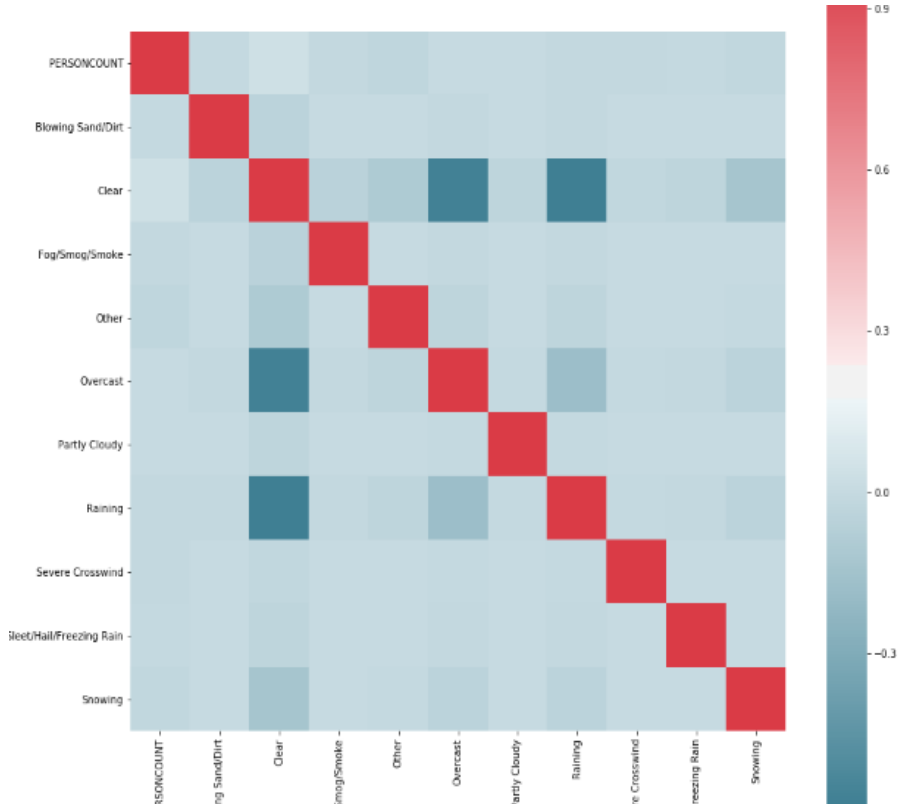
Data source:
https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Data-Collisions.csv



Data Feature Selection

- Weather feature was used to create a further features using one hot encoding
- Correlation Heat map was created

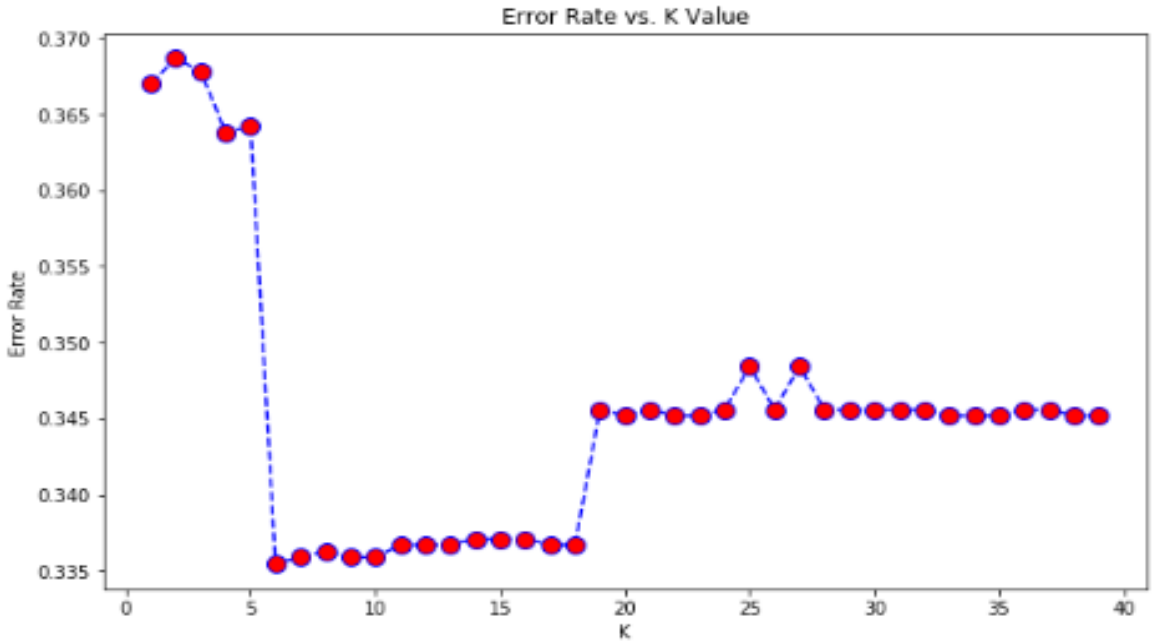
Conclusion:
Did not find any highly
Correlated features hence nothing dropped



Modeling using KNN

Find Optimum K for the data set:

```
Minimum error:- 0.3354890864995958 at K = 5
```

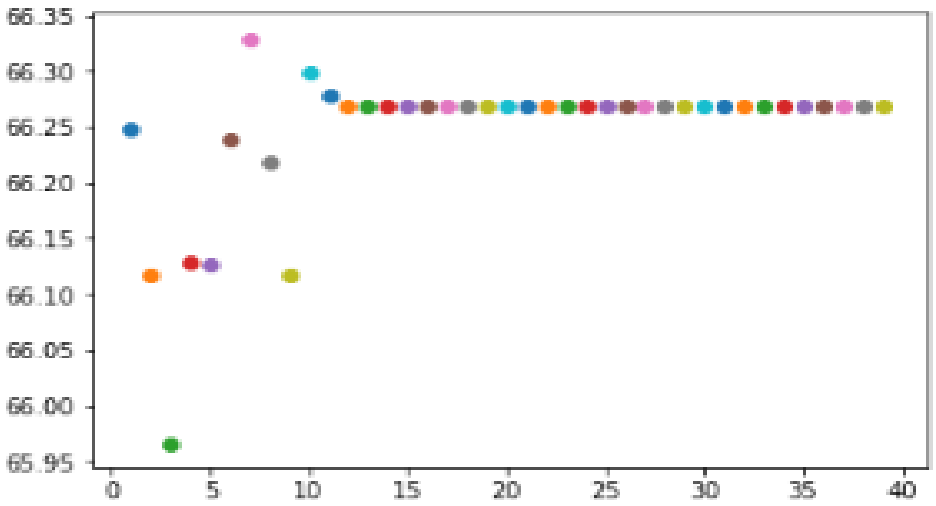


Based on Optimum K=5, Test data scores:

KNN Jaccard index: 0.64
KNN F1-score: 0.58

Modeling using Decision Tree

Find Max Tree depth for Best accuracy:



Max depth for best accuracy is: 6

Based on Max Depth = 6, Test data scores:

DT Jaccard index: 0.67
DT F1-score: 0.57

Modeling using SVM & Logistic Regression

Based on SVM modeling, Test data scores:

SVM Jaccard index: 0.66
SVM F1-score: 0.53

Based on LR modeling, Test data scores:

LR Jaccard index: 0.66
LR F1-score: 0.53
LR LogLoss: 0.63

LR Coefficients

PERSONCOUNT	Blowing Sand/Dirt	Clear	Fog/Smog/Smoke	Other	Overcast	Partly Cloudy	Raining	Severe Crosswind	Sleet/Hail/Freezing Rain	Snowing
0.256242	-0.0742473	0.253212	0.0254702	0.0133844	0.175831	0.0262019	0.180772	0.0189025	-0.0338915	-0.0223147

Weather Severity Conclusion

Comparison of all scores:

Algorithm	Jaccard	F1-score	LogLos
KNN	0.64	0.58	NA
Decision Tree	0.67	0.57	NA
SVM	0.66	0.53	NA
LogisticRegression	0.67	0.55	0.63

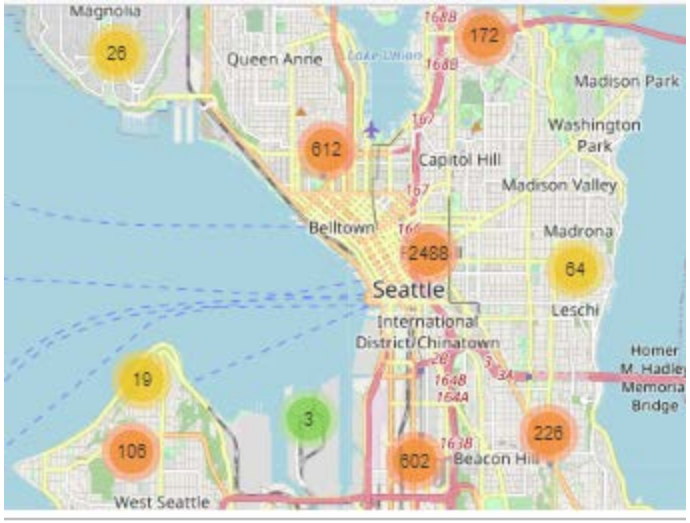
Conclusion:
All algorithms are very close in their Prediction.

Log Regression Coefficients:

PERSONCOUNT	Blowing Sand/Dirt	Clear	Fog/Smog/Smoke	Other	Overcast	Partly Cloudy	Raining	Severe Crosswind	Sleet/Hail/Freezing Rain	Snowing
0.256242	-0.0742473	0.253212	0.0254702	0.0133844	0.175831	0.0262019	0.180772	0.0189025	-0.0338915	-0.0223147

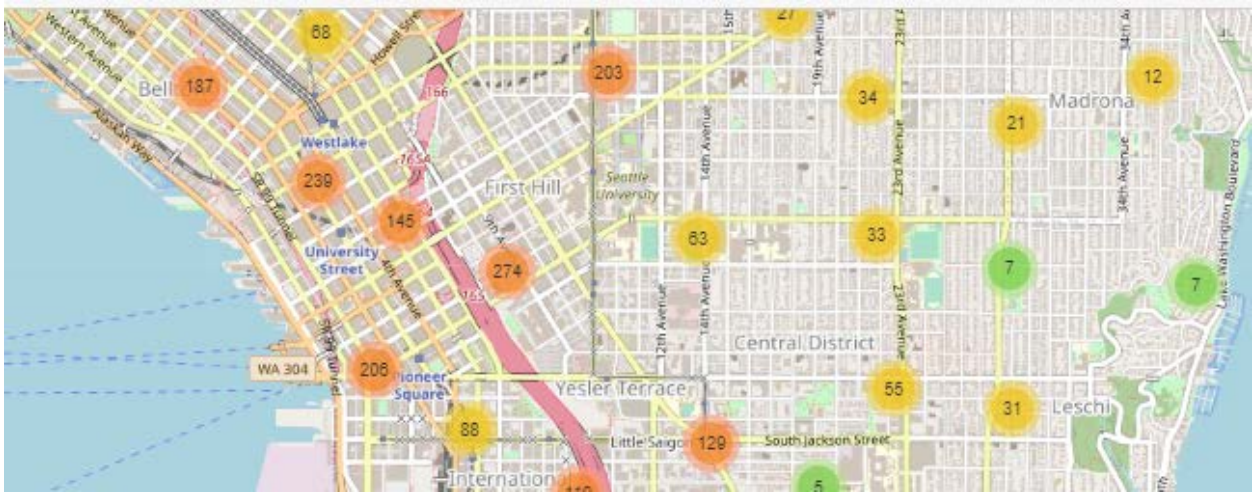
Conclusion:
From LR Coefficients, we can conclude that
Raining, Overcast, Clear & PersonCount
have significant contribution to severity of the accident

High Incident Location – Visual

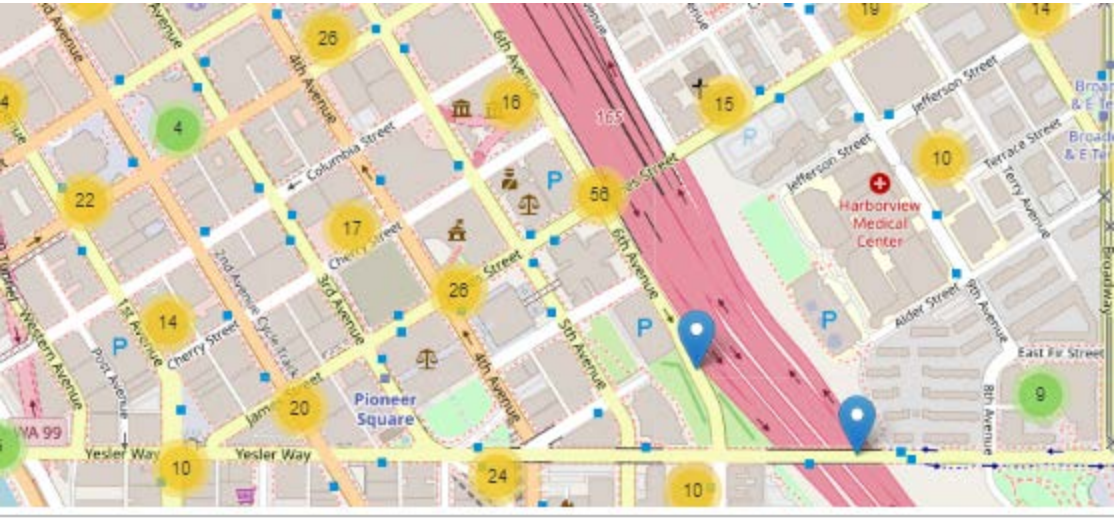


- Used Folium Library
- Marker Cluster Plug in

On zooming high number cluster



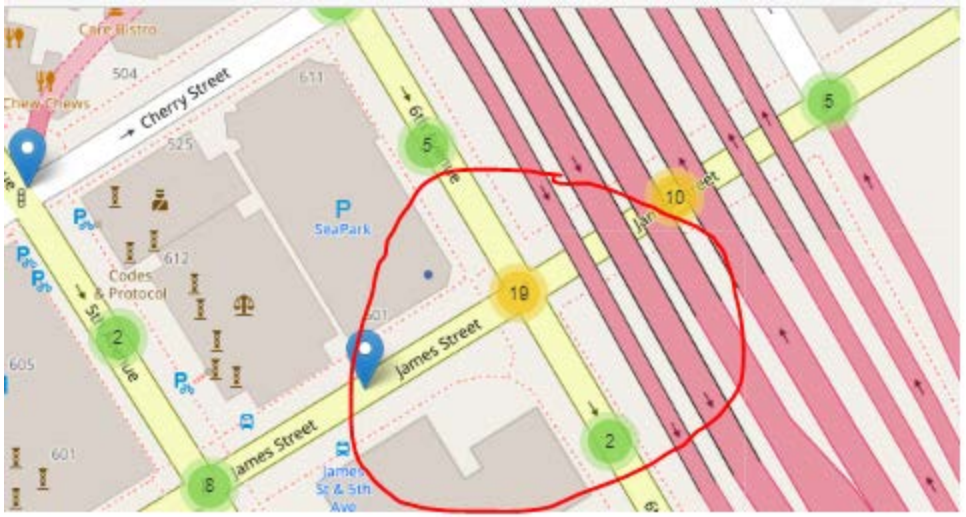
High Incident Location – Conclusion



- one of the high incident intersection

James street/6th Avenue

Zoom to High incident intersection



Interactive zoom video

- You can play the video of interactive from the Github location.
- VisualMapAnalysis.mp4

Future directions

- To make the tool more useful

- Build the interactive graphic model
- Build classification algorithm on the graphic model cluster
- Provide the predictive tool to 911 dispatchers