Preventing the car accident in **Seattle**

Xuhuantao Bei

IBM Data Analysis

Preventing car accident is important

- Government will save the money wasted by car accident to improve the development of city from other aspects like HealthCare, Construction.

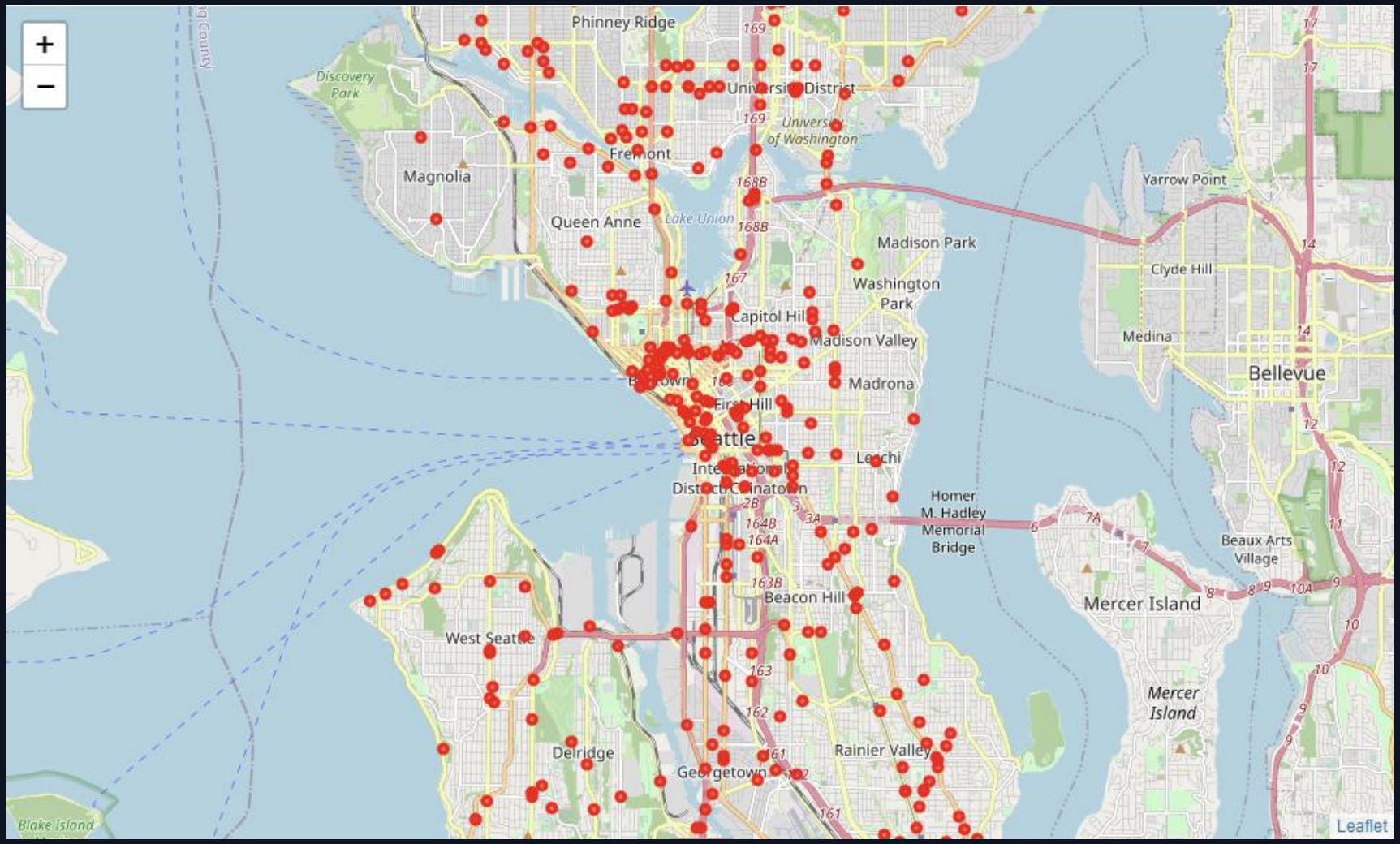
- Preventing accidents help insurance companies to increase profit on their products.

- Residents have better safety protection.

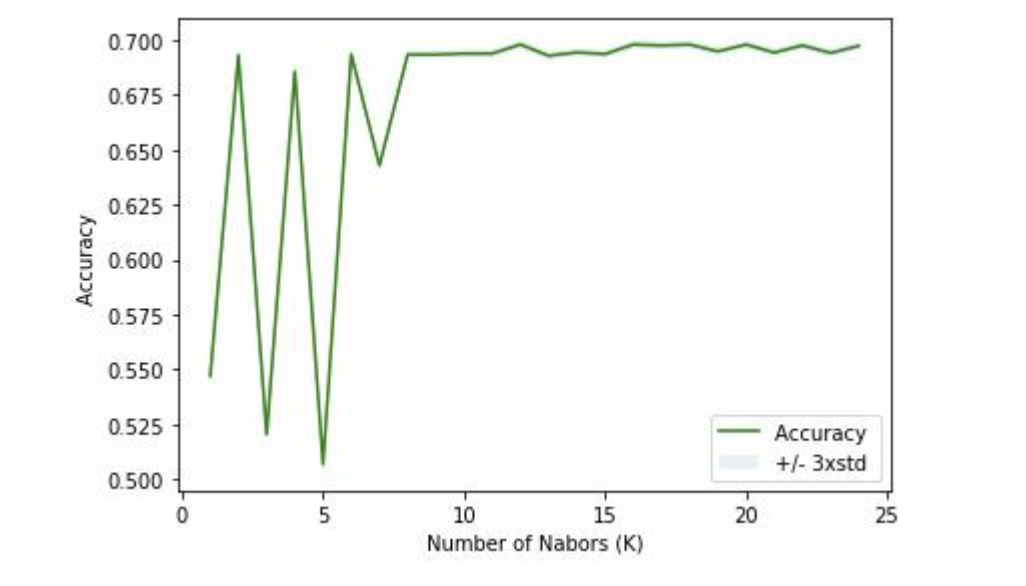
Data Acquisition and Cleaning

- Data-Collisions from IBM Data Analysis on Coursera.
- Irrelevent columns are dropped to improve the visualization.
- Rows contain NaN value are dropped to build model.
- Data is balanced based on the SEVERITYCODE.
- Categorical columns are transformed into numerical columns for machine learning analysis purpose.

Using folium libruary to map the accidents



KNN, Decision tree, Logistic Regression



The best accuracy for Knn is 0.698056406464561 1 with k= 12
Best depth for Decision Tree is 5 and the accuracy is 0.6979155663532974
Accuracy of logistic regression classifier on test set: 0.70

Evaluation

	Model Name	Jaccard Score	F1 Score	Log Loss
0	KNN	0.697	0.575	NaN
1	Decision Tree	0.698	0.574	NaN
2	Logistic Regression	0.698	0.574	0.604

Conclusion

- The model can be used to predict the Severity Level of Accident.
- Accident records with various severity level are required to make improvement.
- The areas of International District and Belltown are required to make improvement to reduce accident.
- Light Condition may be one of the reasons for high damage accident(The accident involve more than four cars).
- Critical information may dropped because of the rebalance of unbalanced data.