Coursera Capstone Project – Data Science

Accident Severity Model

Ray Julin

8/22/2020

**Table of Contents**

[1. Executive Summary 3](#_Toc48991136)

[2. Introduction 3](#_Toc48991137)

[3. Methodology 4](#_Toc48991138)

[4. Results 6](#_Toc48991139)

[5. Discussion 6](#_Toc48991140)

[6. Conclusion 7](#_Toc48991141)

[7. Acknowledgements 7](#_Toc48991142)

[8. References 7](#_Toc48991143)

[9. Appendices 7](#_Toc48991144)

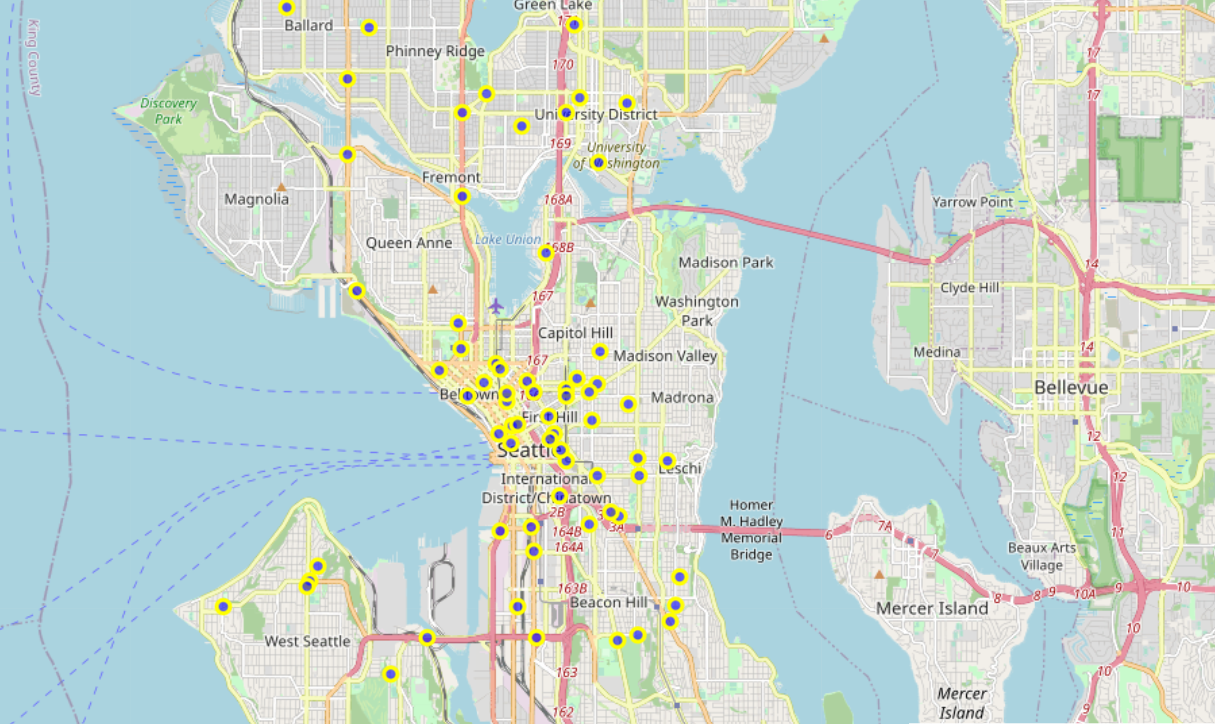
# Executive Summary

This project will use the weather, road conditions, lighting conditions, and accident codes to predict the different accidents' severity. Predict accident severity at each output area given datetime and driving conditions. I will build a prediction model with the following objective: Given date, time, weather, light and road conditions: predict accident severity within Seattle areas.

# Introduction

The biggest challenge with trying to manage severe accidents is they are typically infrequent and appear to be random. However, many are not random at all, but a natural culmination of a series of subtle indicators that can be detected and addressed well in advance of an accident. Using this information, first responders can be better equipped in advance as conditions present themselves. The data set used is from the SDOT Traffic Management Division for Collisions—All Years. from the Traffic Records Group. SDOT GIS Analyst.

A map of the locations of some of the accidents is below:



# Data

Data acquisition and cleaning

**data that will be used:**

Is the "Collisions—All Years" dataset which includes all types of collisions at the intersection or mid-block of a segment, 2004 to Present the features we will be mainly using are :

**INCDATE** - the data of the incident trying to understand is there a certain day which has a higher probability to have an accident

**WEATHER** - the weather at time of the accident, trying to understand does wether effect accident probability

**ROADCOND** - road condition, does the road condition affect the probability to have an accident

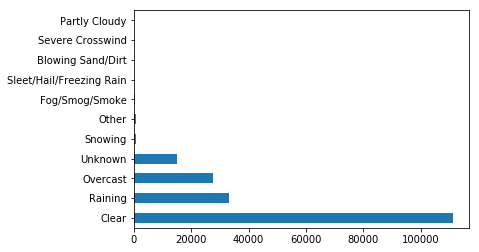
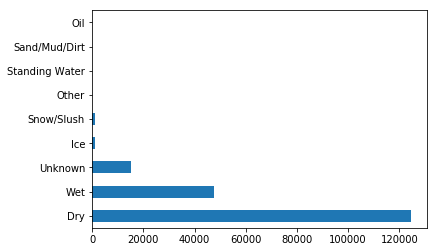
**LIGHTCOND** - light condition, does the lighting in the time of the accident affect the probability to an accident

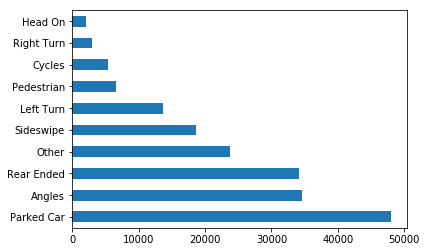
**VEHCOUNT** - the number of vehicles involved

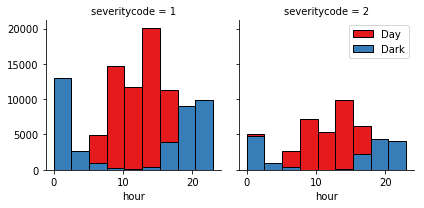
**SEVERITYCODE** - the severity of an injury this will be or target prediction

# Methodology

The data was investigated with several methods and looked at several ways to better understand the types of data that was being presented and how it fit together. From the data we can see that

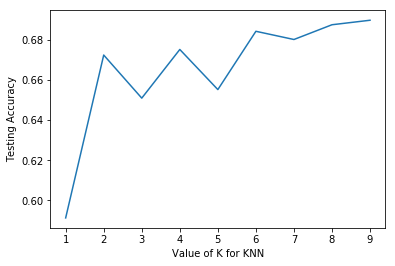
 







# Model Selection and Evaluation

A function was created to check which k gave the best result for the knn model.  
The goal at this stage was to find to optimal k that will give us the highest f1 score without overfitting. We can see the results for each k and the corresponding f1 scores in the plot below.

From this a k of, 6 was selected for the model

Additionally, a decision tree was investigated and showed that a depth of 3 was a good solution to the problem.

| **Depth** | **F1-score** | **Jacard** |
| --- | --- | --- |
| **d=3** | 0.650625 | 0.698430 |
| **d=4** | 0.636617 | 0.699429 |
| **d=5** | 0.642695 | 0.700102 |

# Discussion

We can see here that if an accident occurs at night it is more likely to be on a weekend Friday, Saturday or Sunday, it is least likely to be on a Monday

Fridays have a high occurrence of daytime and nighttime accidents but Sunday is the least dangerous day for daytime accidents which makes sense since most people stay home on Sundays and go out on weekends

# Conclusion

The model does a reasonably good job of predicting injuries or property damage only accidents. Unfortunately the data does not include further classification of fatalities or types of injuries incurred and by who. With this data we could better prepare first responders to react to accidents give a time of day, day of week, weather, and road conditions.

# Acknowledgements

Thanks to Coursera instructors for working with us on developing these skills to understand Data Science and to Alex Aklson, Ph.D., Data Scientist, IBM Developer Skills Network, for his work in leading this capstone assignment.

# References

Data: Collisions—All Years. SDOT Traffic Management Division, Traffic Records Group. SDOT GIS Analyst.

# Appendices

Location of data and Jupyter Notebook: <https://github.com/rdjulin/Coursera_Capstone>