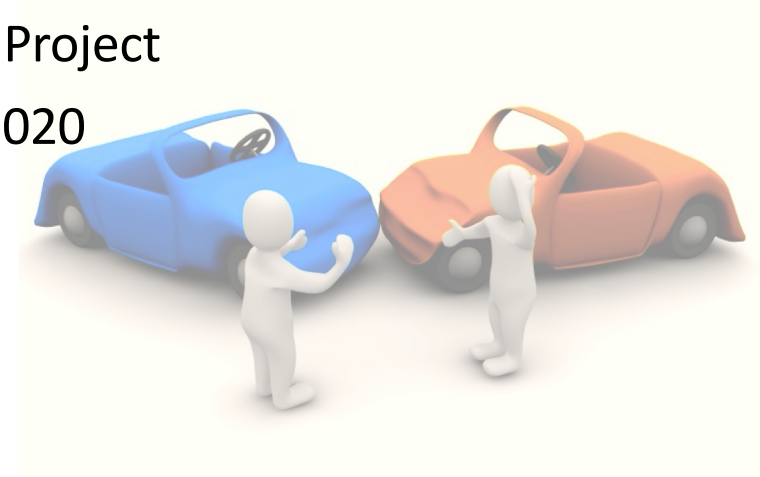
# Predicting Collision Severity

With Machine Learning

IBM – Capstone Project

September 2020



# Introduction: Predicting Road Traffic Collisions in Seattle

- Reasons and factors of high severity of an accident can be determined using machine learning algorithms
- Can these algorithms be used to decrease risks and severity?
- Can city planners use these to increase security?

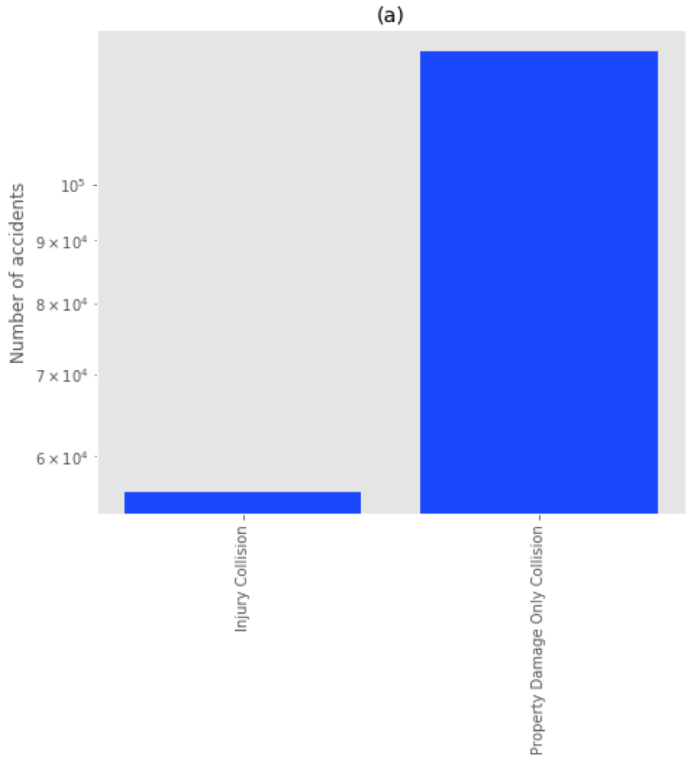
# Dataset and pre-processing

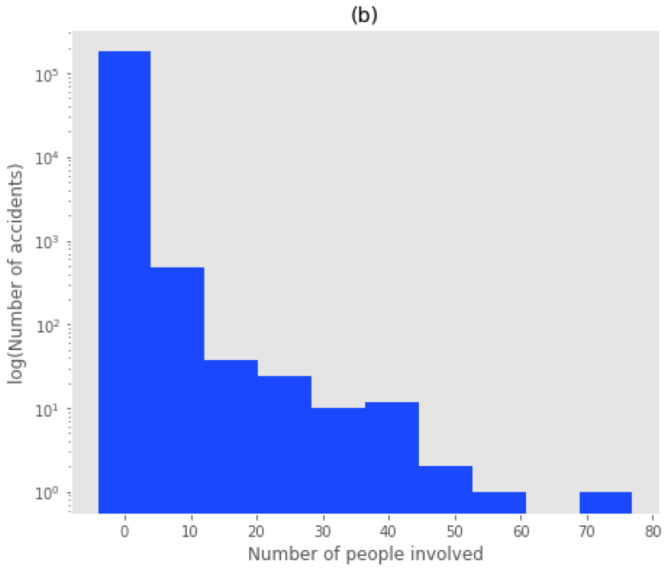
- Seattle collision data was provided by Coursera, containing entries from 2004 to present day (link), with a link to the metadata descriptions
- There were 194.673 entries of collision data, and 37 features in the original dataset.
- Irrelevent features were dropped and the cleaned data contained 29 features.

# Exploring the data

- There are correlations between the day of the week, month of the year as well as several different factors such as road conditions, light conditions, weather conditions and location that influence the severity of an accident.

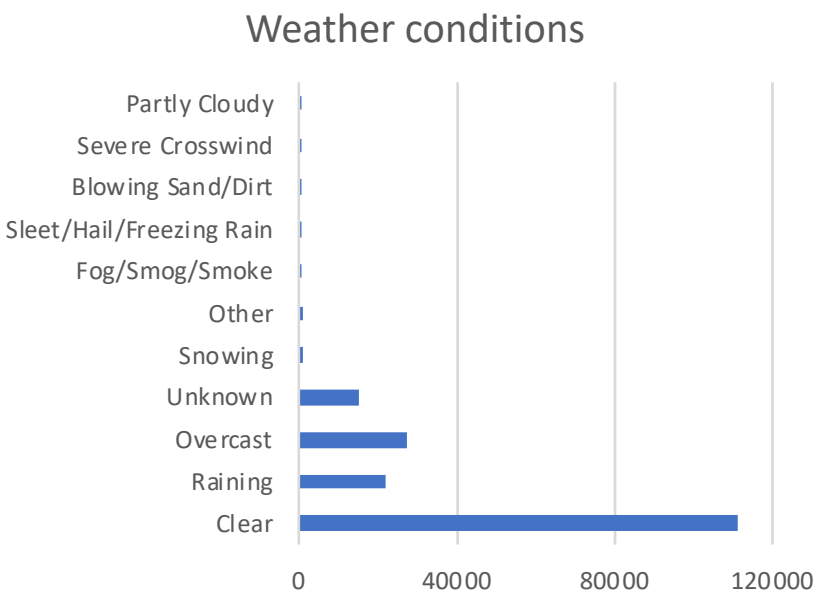
# Collision severity and number of people involved

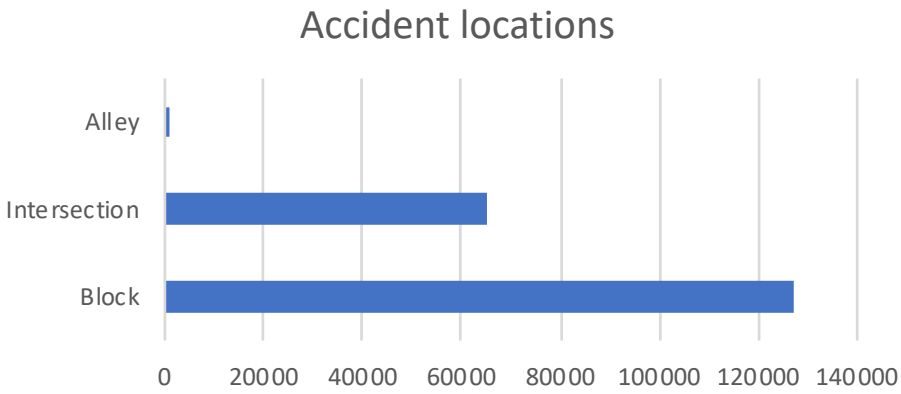


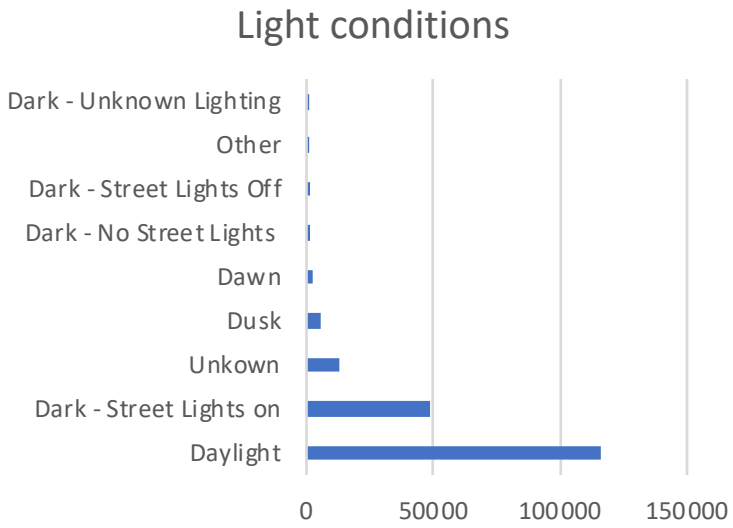



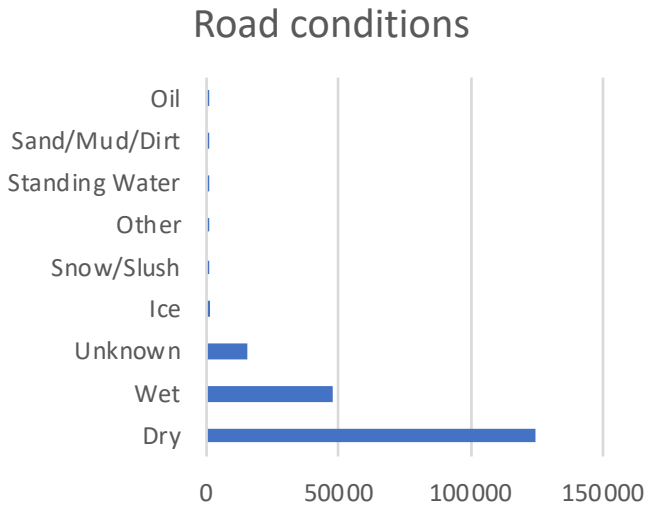

Most accidents were material, and not involving injuries, however there were several accidents that did

# Collisions overview

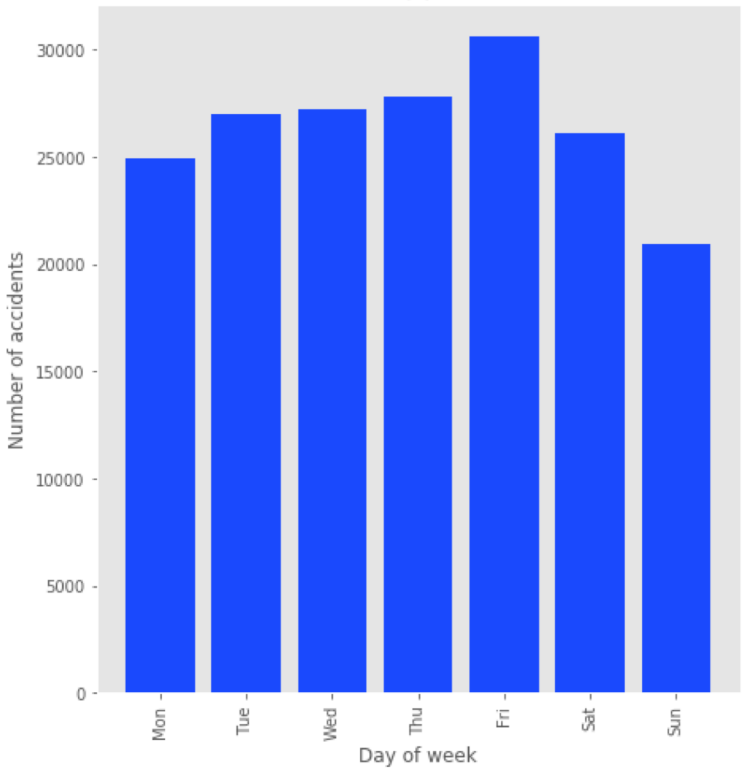


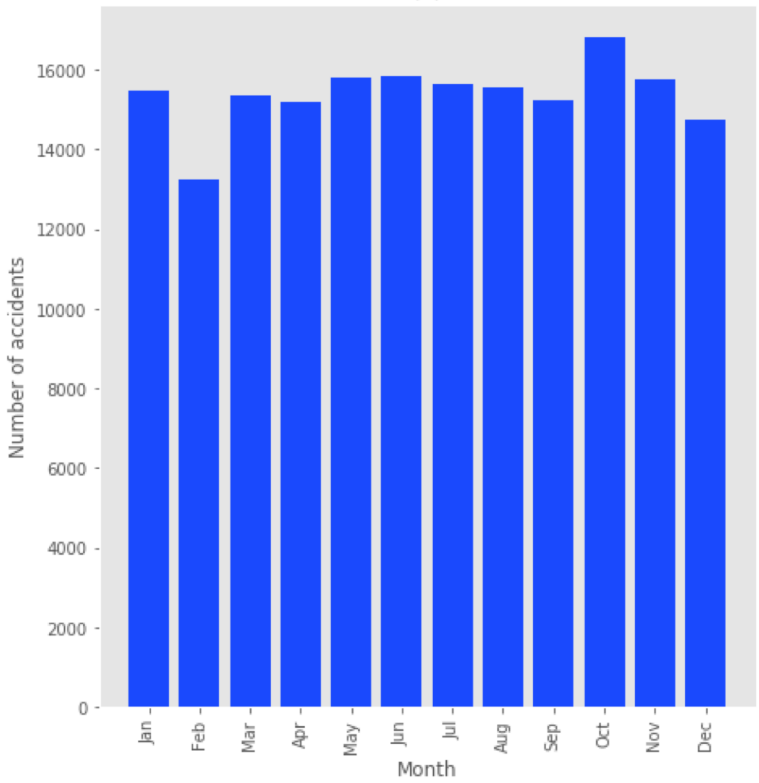



# Collisions overview – day of week, month of year



# Model development and evaluation

- 3 methods were trained and evaluated
  - Decision Tree model
  - K-Nearest Neighbors (kNN) model
  - Logistic Regression model
- Decision Tree algorithm produced the highest accuracy

| | Algorithm | Jaccard | F1-score | Precision |
|---|---|---|---|---|
| 0 | KNN | 0.73 | 0.68 | 0.71 |
| 1 | Logistic Regression | 0.7 | 0.58 | 0.67 |
| 2 | Decision Tree | 0.75 | 0.69 | 0.77 |

# Conclusions and future directions

- This work highlights that machine learning techniques can be applied to historical data to be able to make reliable predictions about the outcome of road traffic accidents.
- The model can be extended to include new features so that city planners can gain insight into the conditions associated with high accident severity and use this to improve road or traffic flow design.