Accident Severity Prediction
In
Seattle, WA

Coursera Capstone Project

Abhineet Sharma

Business Understanding

Seattle DMV is interested in knowing possibility of an accident given weather and other road conditions.

Problem Statement

At any given day, predict the severity of accident given,

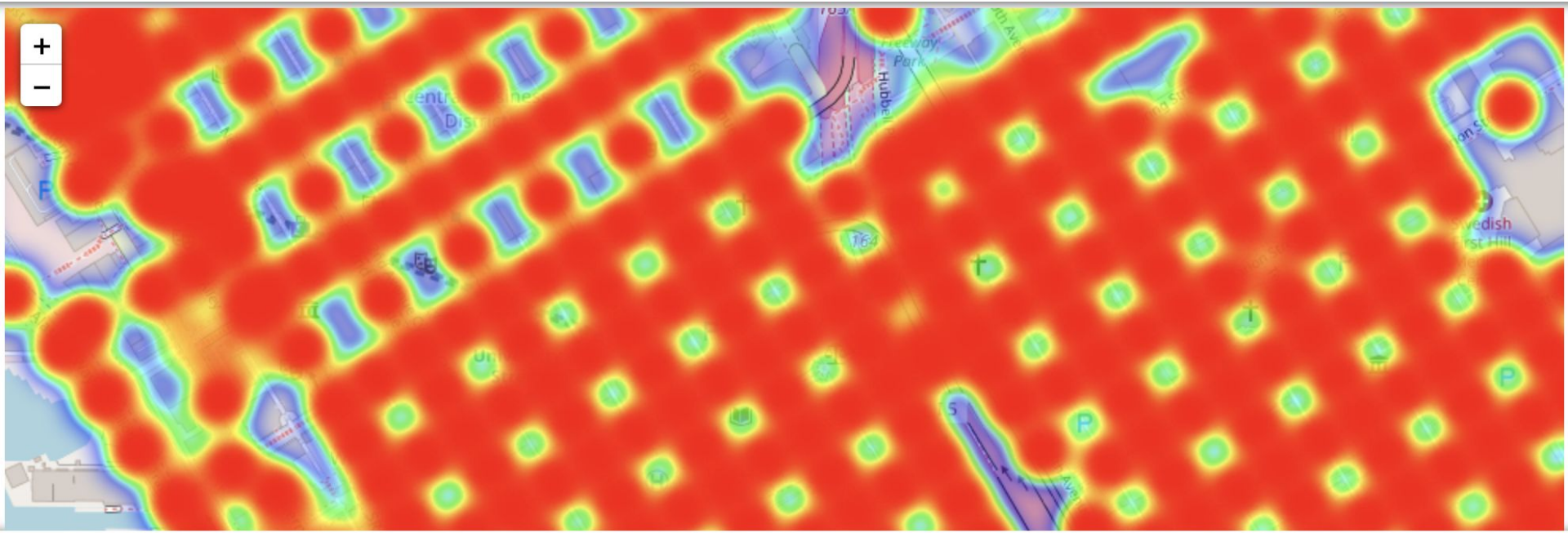
- Weather Conditions
- Road Conditions
- Light Conditions
- Mode of Travel
- Area of Travel

Data Understanding

- Available Data: Accident Data collected
- Format: CSV
- Data Size: 194673 rows x 38 columns
- Redundant or Not useful Data: 24 columns

- 70.11% Severity 1 Accidents
- 29.89% Severity 2 Accidents

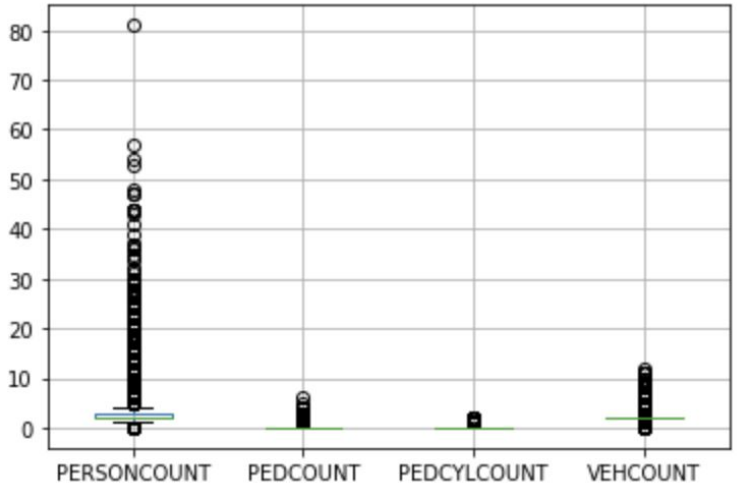
Location Heat Map



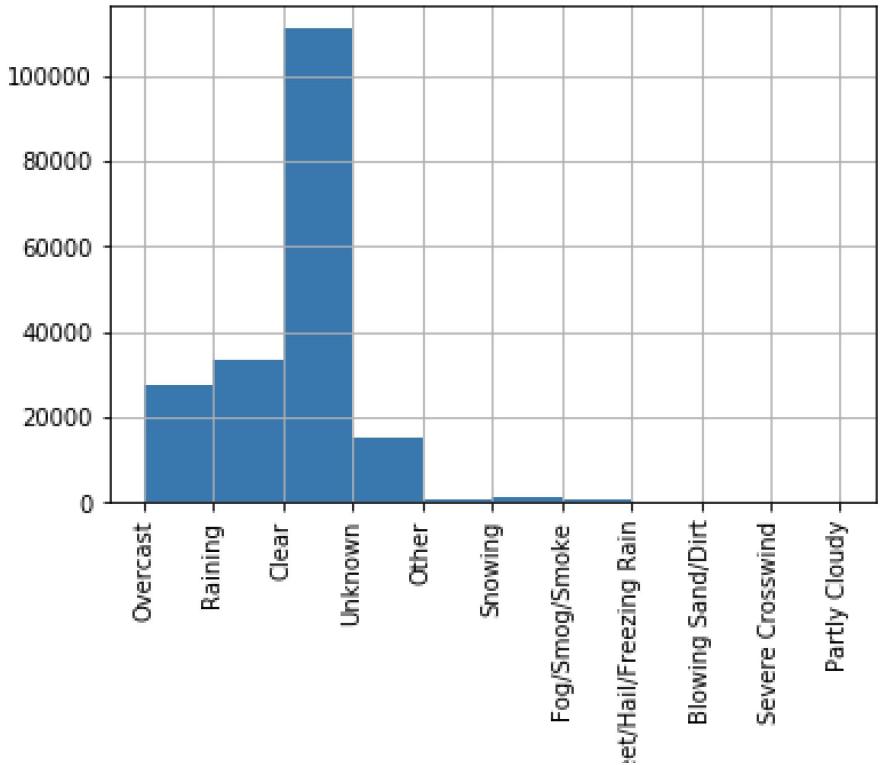
Address Type Data Segregation

- Block: 65.85%
- Intersection: 33.76%
- Alley: 0.39%

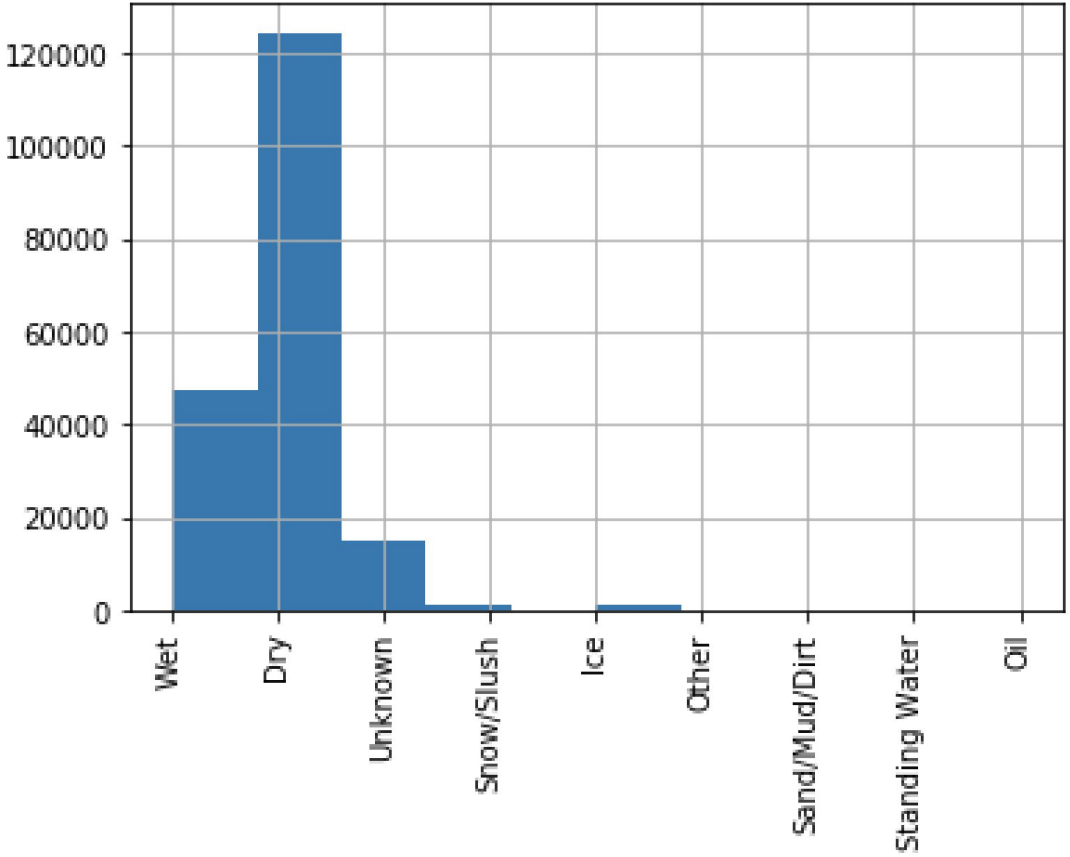
Box Plot: Person, Pedestrian, Cyclist & Vehicles Count



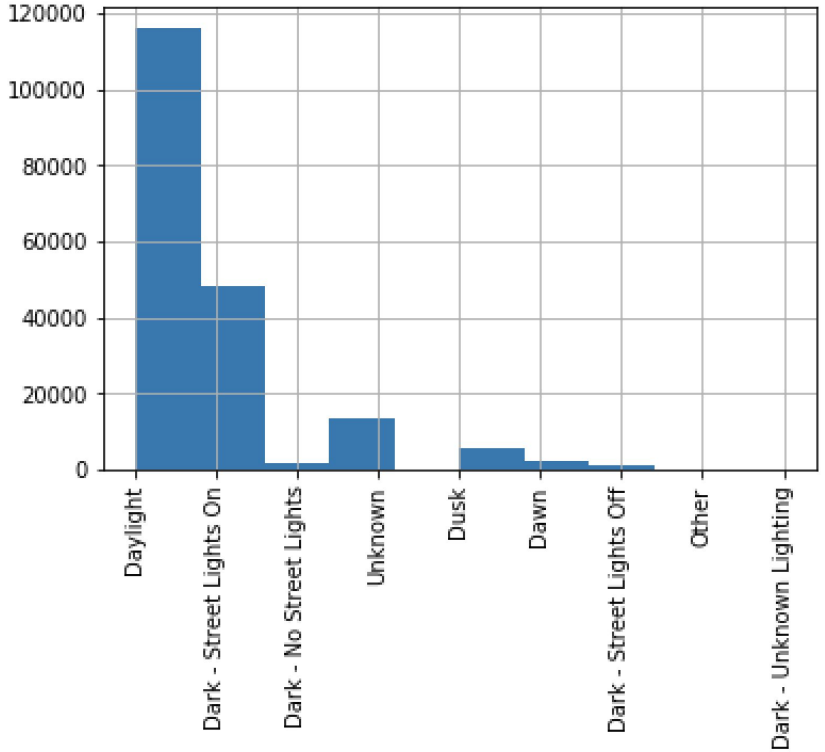
Distribution By Weather



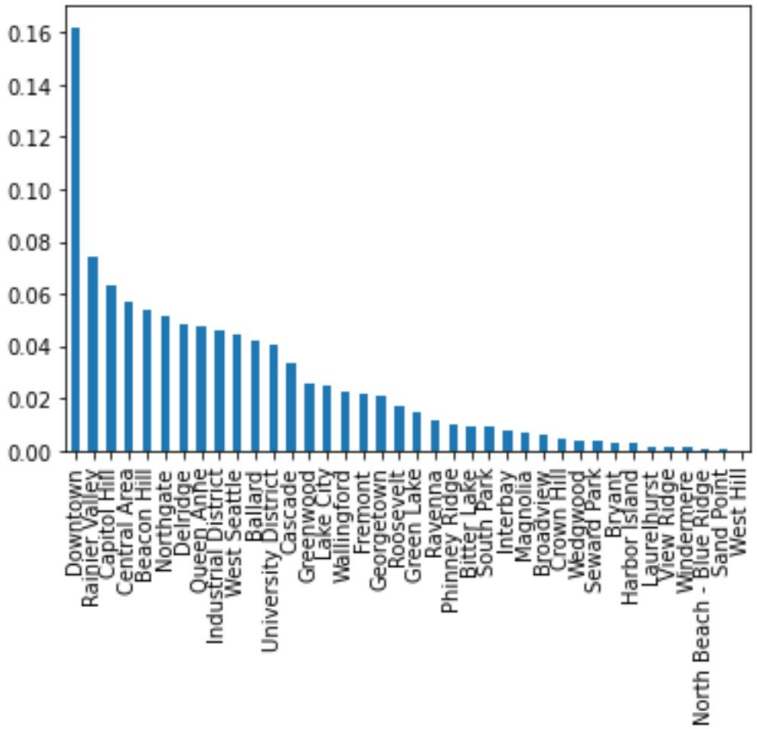
By Road Conditions



By Light Conditions



By Neighborhood



Data Modelling

Target Variable: SEVERITYCODE

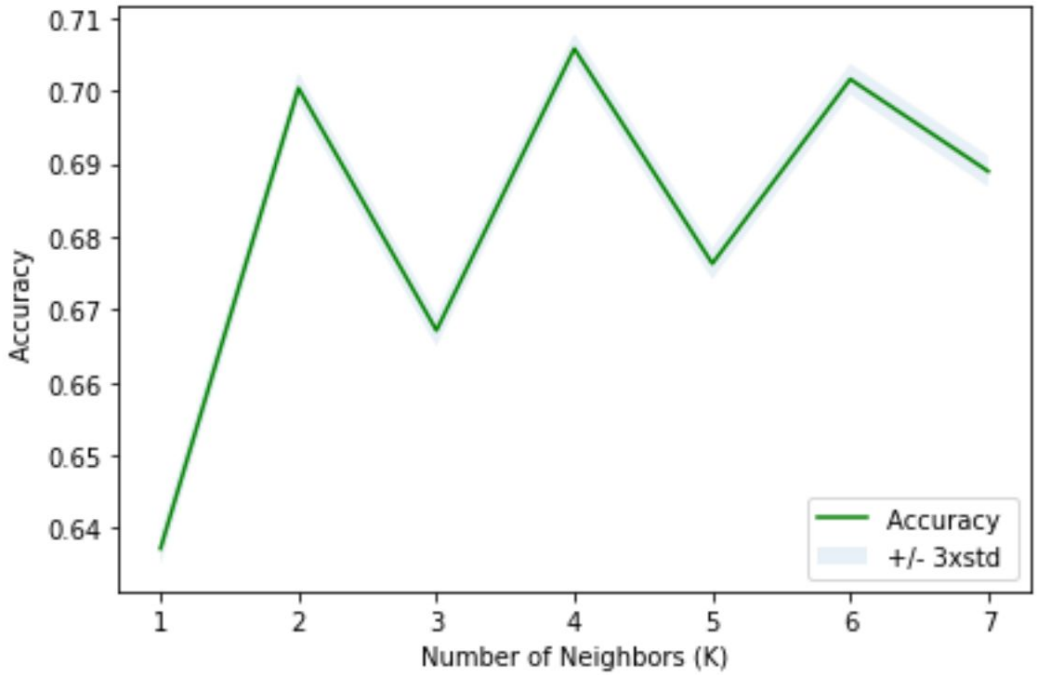
Independent Variables:

- WEATHER
- ROADCOND
- LIGHTCOND
- HITPARKEDCAR
- NEIGHBORHOOD
- PED
- CYCLIST
- VEHICLE
- DAY_PART
- SEASON

Using 70-30 train-test split

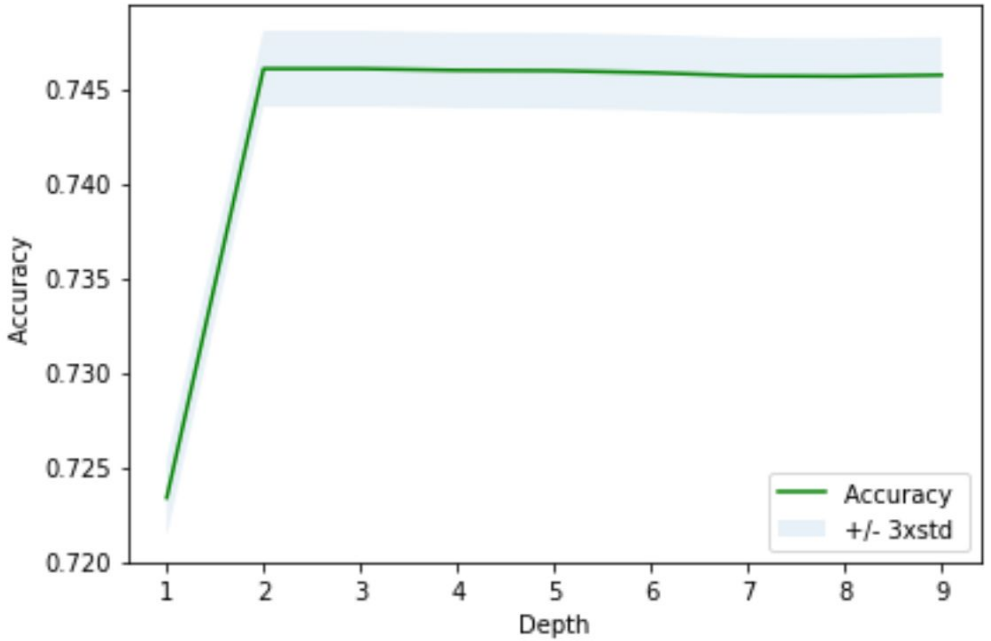
KNN

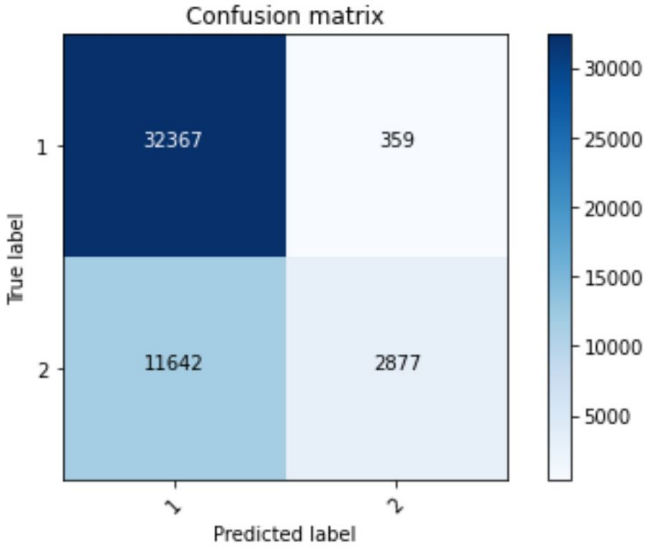
Optimal Nearest Neighbors, k = 4. Accuracy: 70.58%



Decision Tree

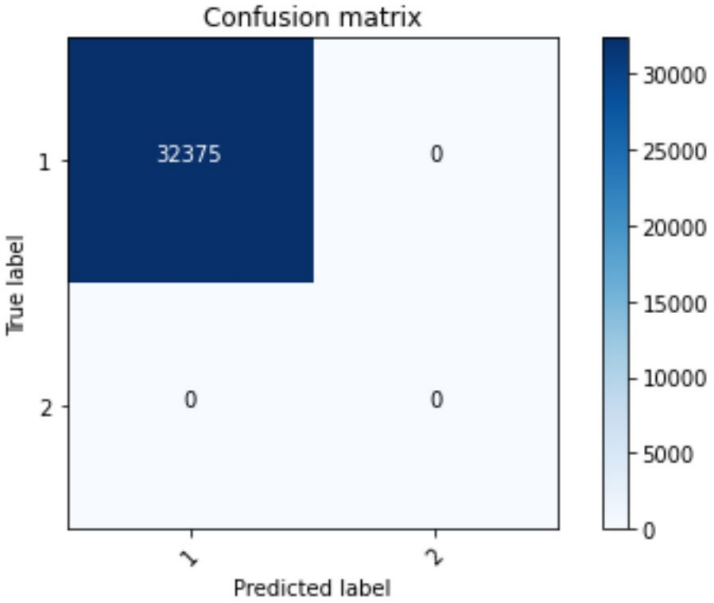
Optimal Depth, d=2. Accuracy: 70.58%





Logistic Regression

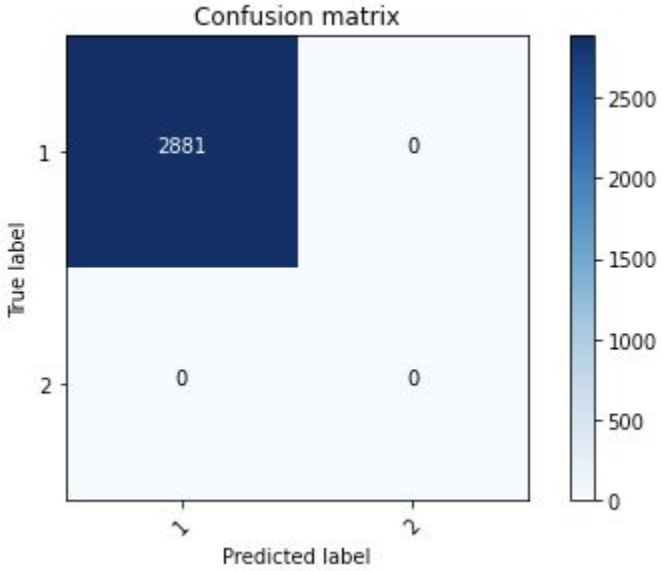
Accuracy Score: 74.54%



SVM

Kernel: RBF

Accuracy: 74.60%



Model Evaluation

Model	Accuracy	Jaccard Similarity score	F1-Score	Log Loss Score
KNN	70.58%	66.21%	79.67%	-
Decision Tree	70.58%	72.95%	84.36%	-
Logistic Regression	74.54%	72.92%	84.33%	53.86%
SVM	74.60%	72.95%	84.36%	-

Best accuracy is achieved on SVM. SVM with RBF kernel is the model of choice.