# Predicting Seattle Road Accident Severity

Hemalatha Murugesan

October 12, 2020

# Introduction

- Road accidents have become very common nowadays.
- Incidences of road accidents are just increasing day by day.
- As per WHO observatory data “Road traffic injuries are currently estimated to be the eighth leading cause of death across all age groups globally and are predicted to become the seventh leading cause of death by 2030”.
- Predicting the accident severity in advance could be used to save significant amount of lives each year.
- Road safety should be a prior interest for governments.

# Data

- All the recorded accidents in Seattle from 2004 to 2020.
- The data Seattle car accident comes from , Here
- Features of a metadata, Here
- In total 38 features, 14 features were pre-selected
  - Redundant and non relevant features were dropped
- On the data cleaning missing values and outliers were replaced.

# EDA-Target

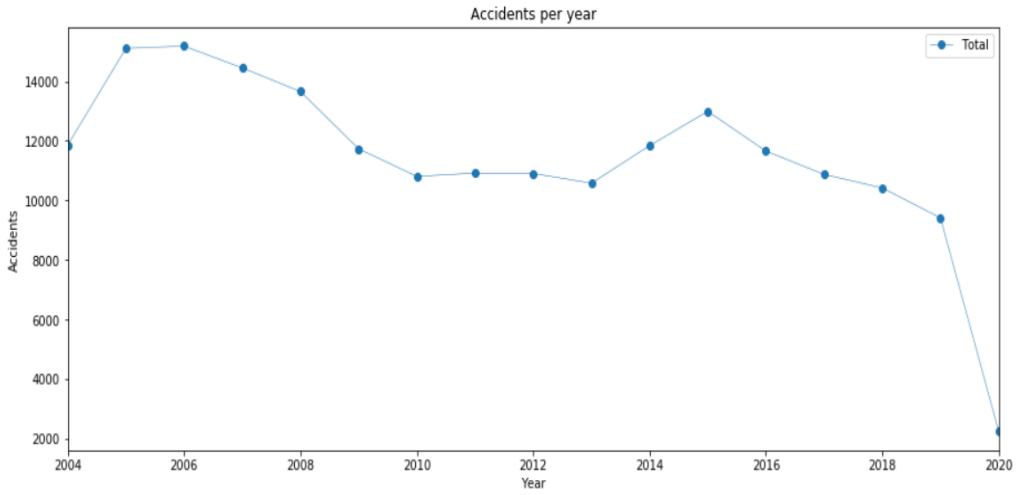
- The target feature, describing the accident severity.

  1: Property Damage

  2: Injury

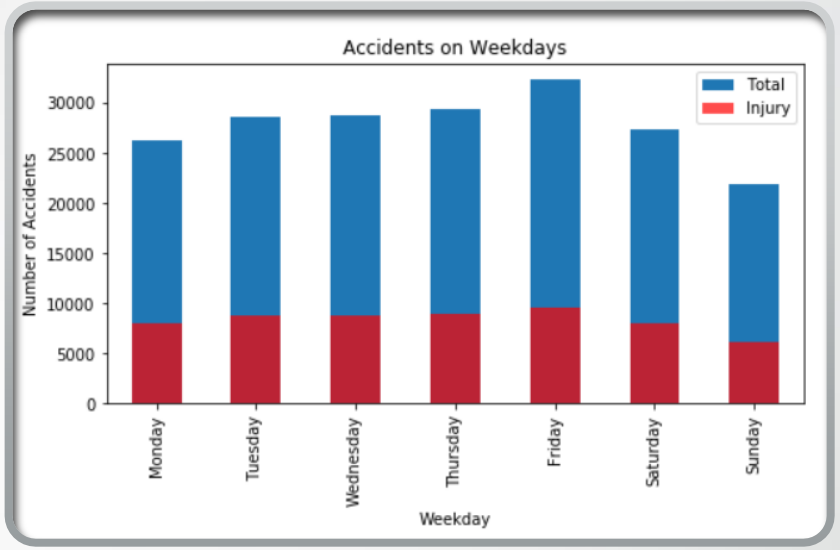
# EDA-Seasonality

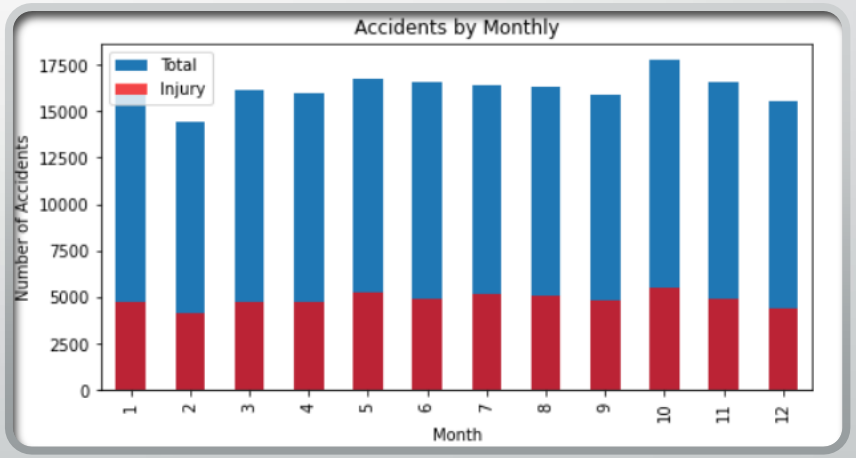
- The number of traffic accidents over the years from 2004 to 2020



# EDA-Seasonality

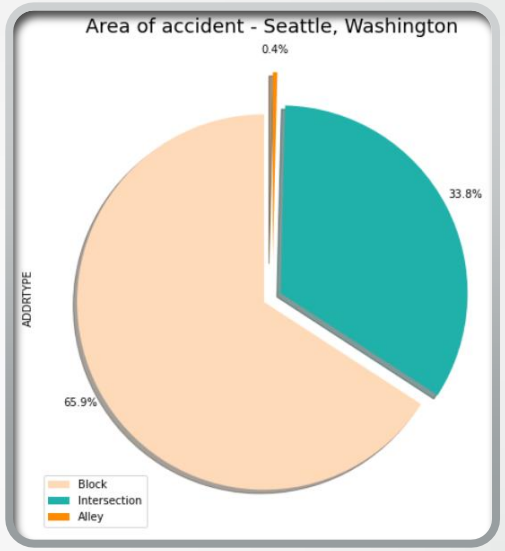
- Fridays have more accidents and less on Sundays
- Accidents high in October and low in February.
-

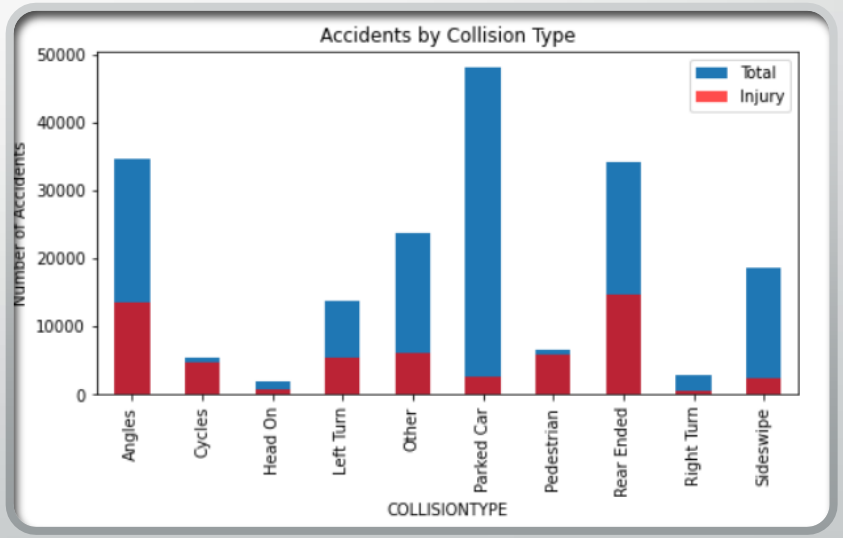


# EDA-Collision & Area

- Most accidents in the intersection
- Accidents by Collision type



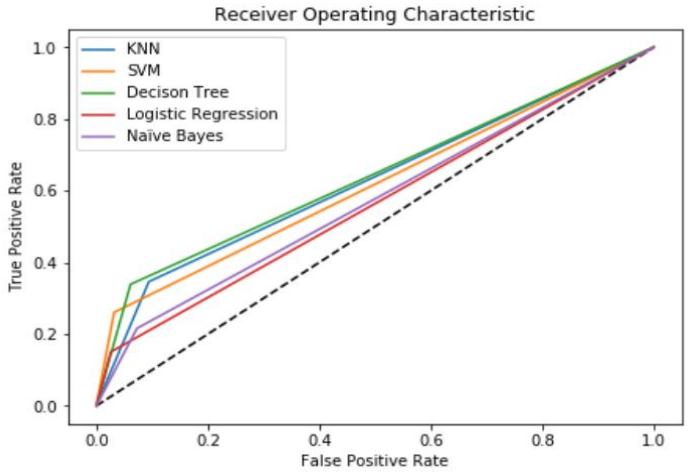
# Classification Models

- Logistic Regression
  - C=0.001
- Decision Tree
  - Max depth = 9
- K Nearest Neighbor
  - K=6
- Support Vector Machine
- Gaussian Naïve Bayes

# Results

- This table reports the results of the evaluation of each model.
- The best model is Decision Tree with 75% accuracy.
- ROC curve shows the clearly the best model.

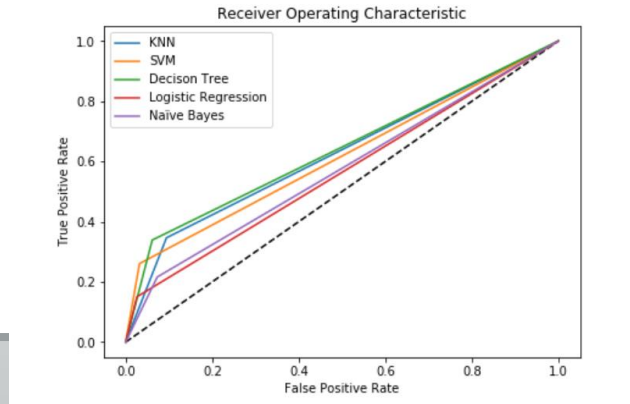
| Algorithm | Test Accuracy | Jaccard | F1-score | Precision | Recall |
|---|---|---|---|---|---|
| **Logistic Regression** | 0.72 | 0.72 | 0.65 | 0.72 | 0.72 |
| **Decision Tree** | 0.75 | 0.75 | 0.71 | 0.75 | 0.75 |
| **SVM** | 0.75 | 0.75 | 0.70 | 0.76 | 0.75 |
| **KNN** | 0.73 | 0.73 | 0.71 | 0.72 | 0.73 |
| **Gaussian Naïve Bayes** | 0.71 | 0.66 | 0.66 | 0.68 | 0.71 |



# Results

- This table reports the results of the evaluation of each model.
- The best model is Decision Tree with 75% accuracy.

| ALGORITHM | TEST ACCURACY | JACCARD | F1-SCORE | PRECISION | RECALL |
|---|---|---|---|---|---|
| **Logistic Regression** | 0.72 | 0.72 | 0.65 | 0.72 | 0.72 |
| **Decision Tree** | 0.75 | 0.75 | 0.71 | 0.75 | 0.75 |
| **SVM** | 0.75 | 0.75 | 0.70 | 0.76 | 0.75 |
| **KNN** | 0.73 | 0.73 | 0.71 | 0.72 | 0.73 |
| **Gaussian Naïve Bayes** | 0.71 | 0.66 | 0.66 | 0.68 | 0.71 |



# Conclusion

- Built useful models to predict the severity of a traffic accident
- Accuracy of the model has room for improvement.
- Add features such as vehicle speed and time of uninterrupted travelling.
- Prediction of potential accident, critical spots and time.