TRAFFIC ACCIDENT SEVERITY PREDICTION

Applied Data Science Capstone Project

INTRODUCTION

Traffic Accidents are:

- Cause of 1.35 million deaths globally.
- Predicted to become the 7th leading cause of death by 2030.

Prediction of Accident Severity will be beneficial to multiple stakeholders included below:

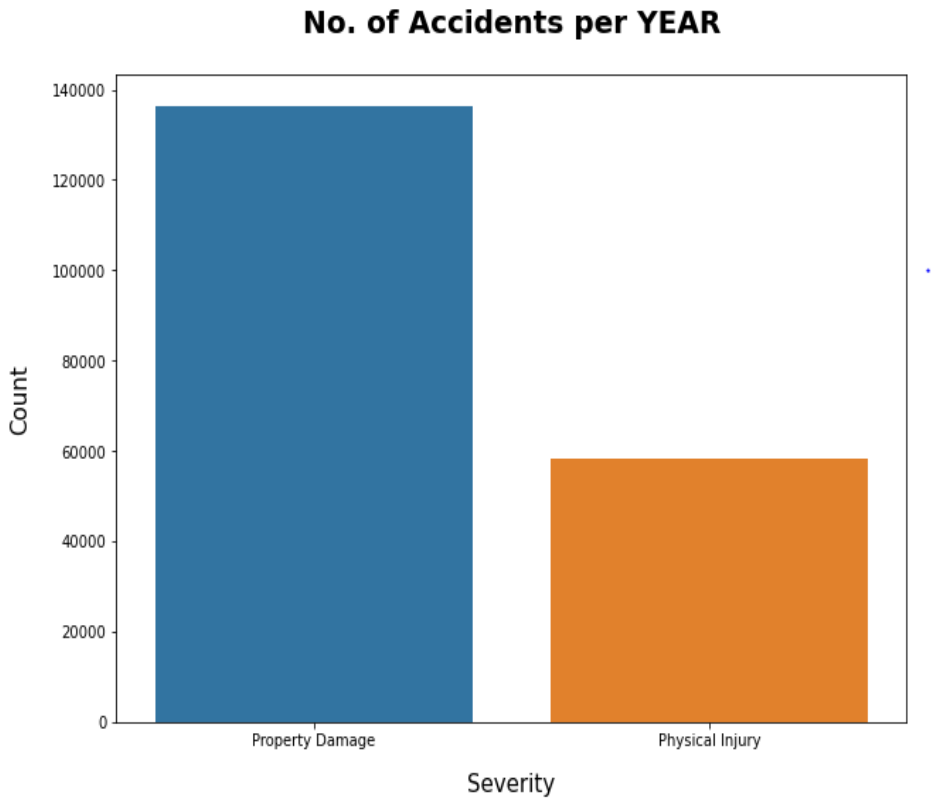
- Public Development Authority of Seattle (PDAS)
- Citizens
- Insurance / Healthcare Agency

TARGET VARIABLE

The target variable feature a binary classifier

- Property Damage
- Physical Injury

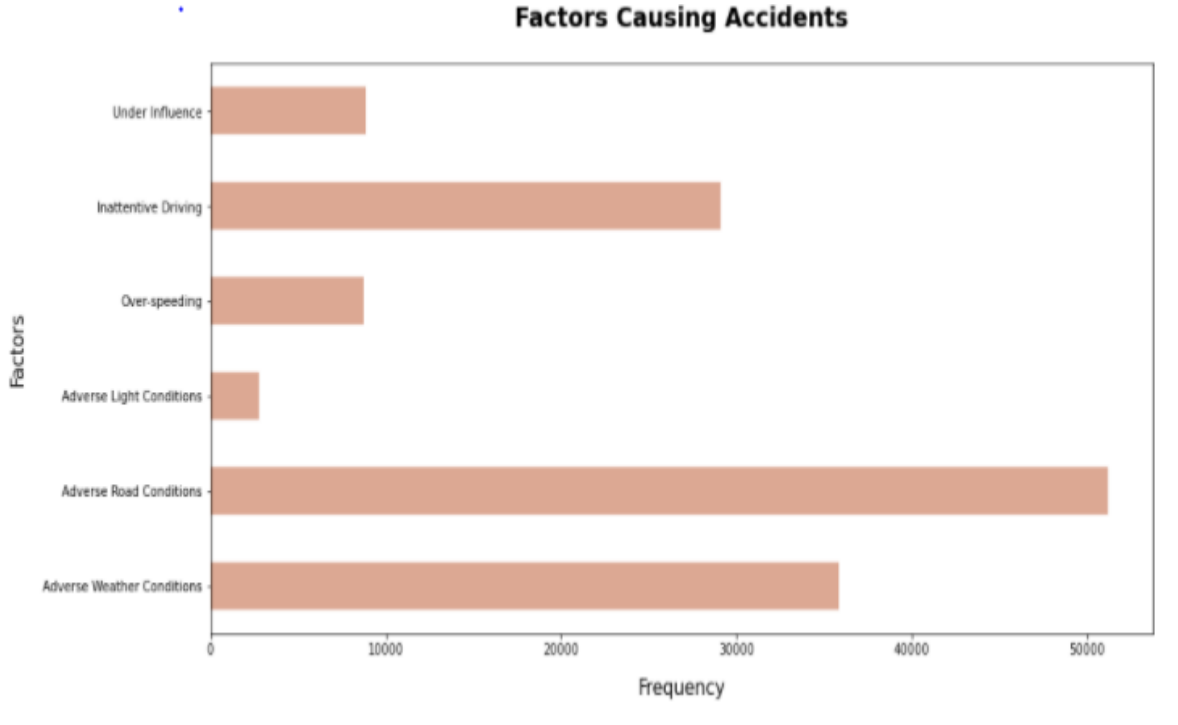
The dataset is not balanced well as can be seen from the figure on the right.



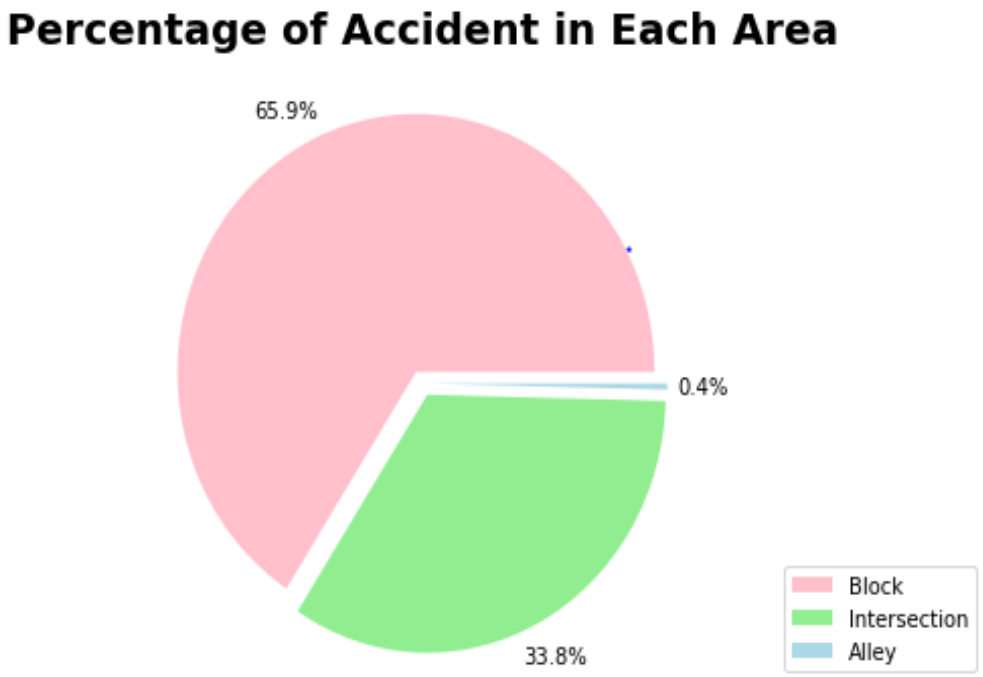
FACTORS CAUSING ACCIDENTS

Ranking (Highest to Lowest)

- Adverse Road Conditions
- Adverse Weather Conditions
- Inattentive Driving
- Under Influence
- Over-speeding
- Adverse Light Condition

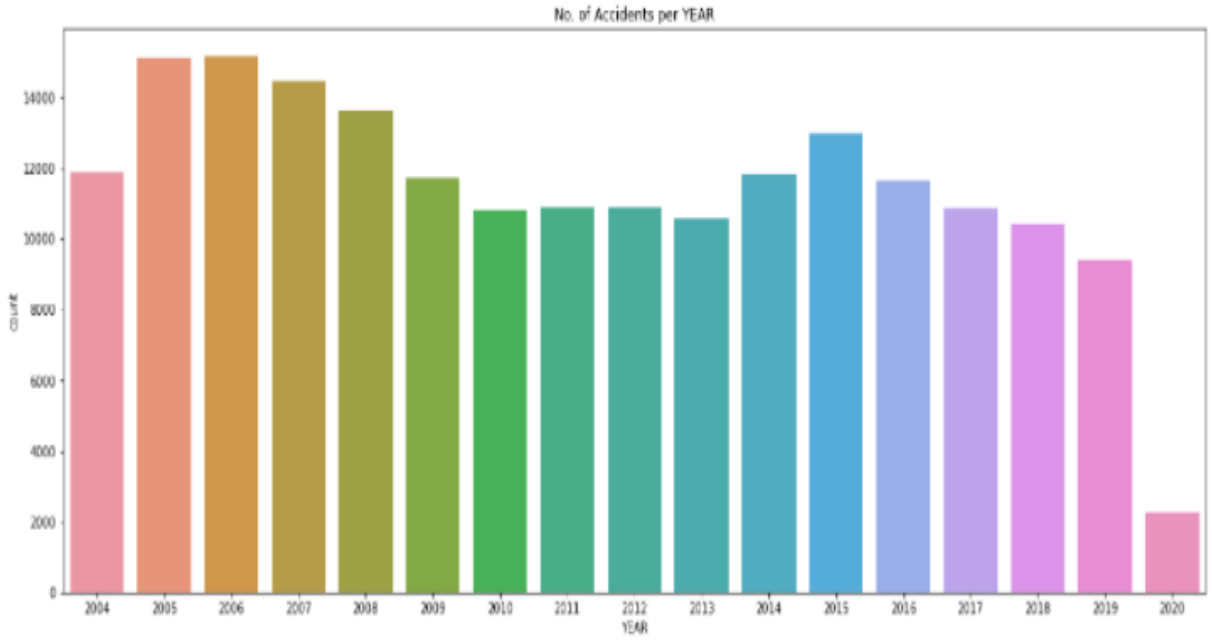


ACCIDENTS PRONE AREAS



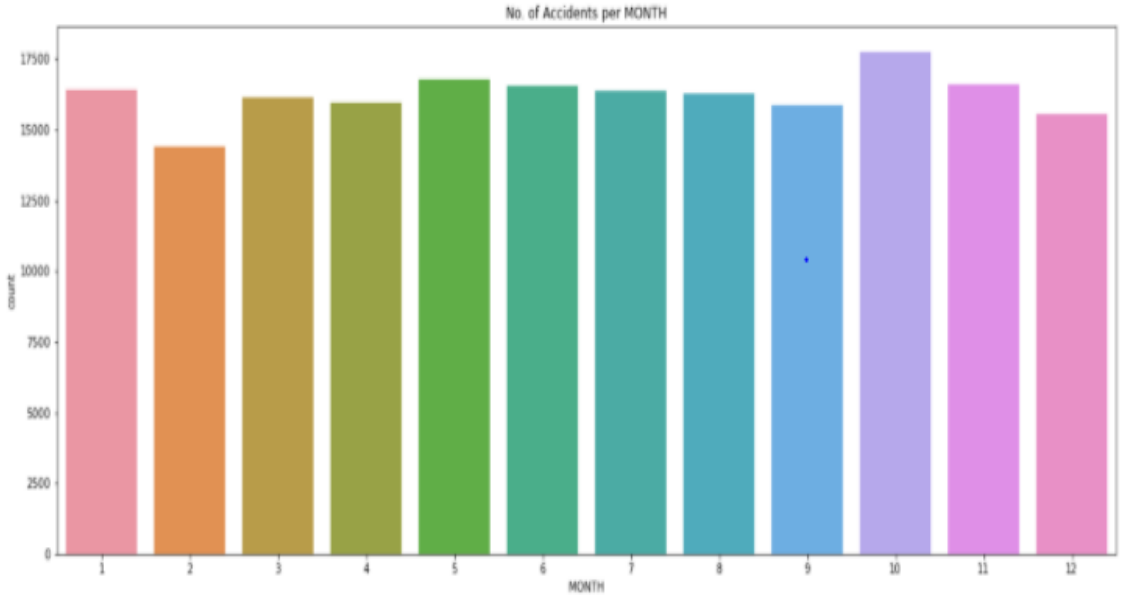
- Accidents are prone to happen at Block and at Intersection
- More attention should be place at these place to reduces occurance

ACCIDENTS SEASONALITY - YEARLY



Downtrend of accidents happening yearly. The number of accidents occurred peaked at 2005-2006.

ACCIDENTS SEASONALITY - MONTHLY



Number of accidents increase from February to May, then again in September. It decreases at the end of the year.

CLASSIFICATION MODELS

- Decision Tree
  - Criterion = Entropy
  - Max-Depth = 6
- Linear Regression
  - C = 0.1
- k-Nearest Neighbour
  - K = 4

RESULTS

Algorithm	Average F1-score	Property Damage (1) vs Injury (2)	Precision	Recall
Decision Tree	0.60	1	0.72	0.68
		2	0.35	0.39
Logistic regression	0.60	1	0.72	0.69
		2	0.35	0.39
k-Nearest Neighbour	0.61	1	0.70	0.88
		2	0.35	0.15

Logistic Regression would be the best model as it shows a well-balanced recall and also a good F1-score.

CONCLUSION

- Model could have performed well if few more criteria are present:
  - A balanced dataset for the target variable
  - More instances recorded for accidents taken place in Seattle, Washington
  - Less missing values within the dataset
  - More factors to be evaluated