Car Accident Severity Prediction

Akshay Surve

INTRODUCTION

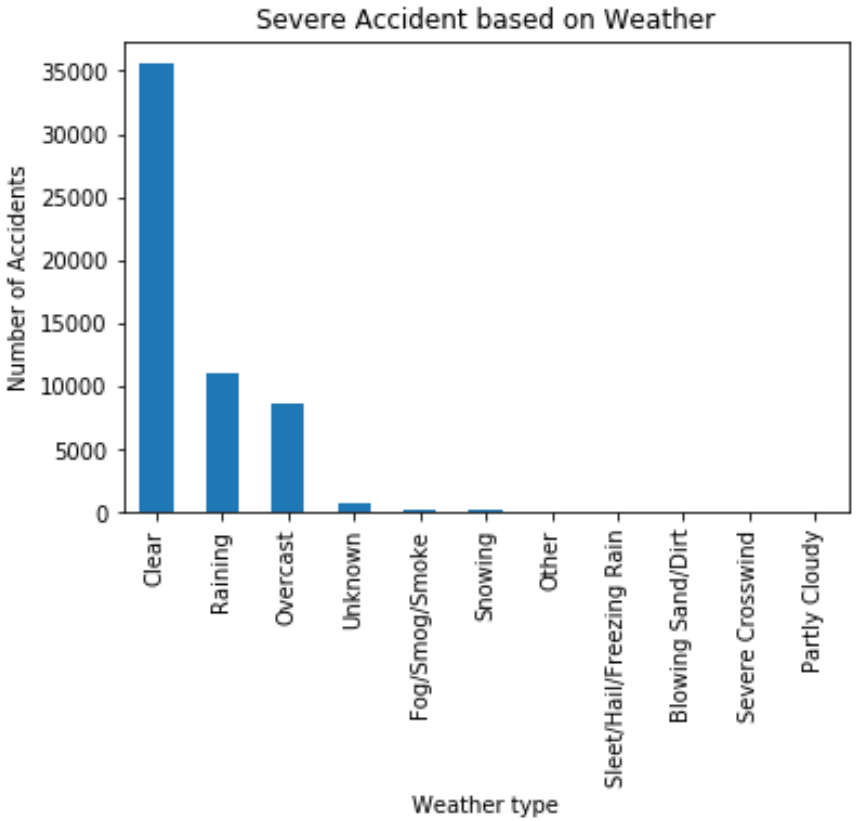
Background

In this fast-moving world, no one is safe. The whole world suffers due to car accidents. Thousands of people loses their life in car accidents due to some controllable factors like not paying enough attention during driving, abusing drugs and alcohol or driving at very high speed or some uncontrollable factors like weather, visibility, or road conditions. We are trying to solve this problem by giving warning to driver and remind them to drive carefully in critical situations. We are going to predict accident severity depending upon the factors such as current weather, road and visibility conditions which are already given to us. This project aims to reduce numbers of car accidents on highway in Seattle. The target audience of the project will be local government, police and car insurance companies.

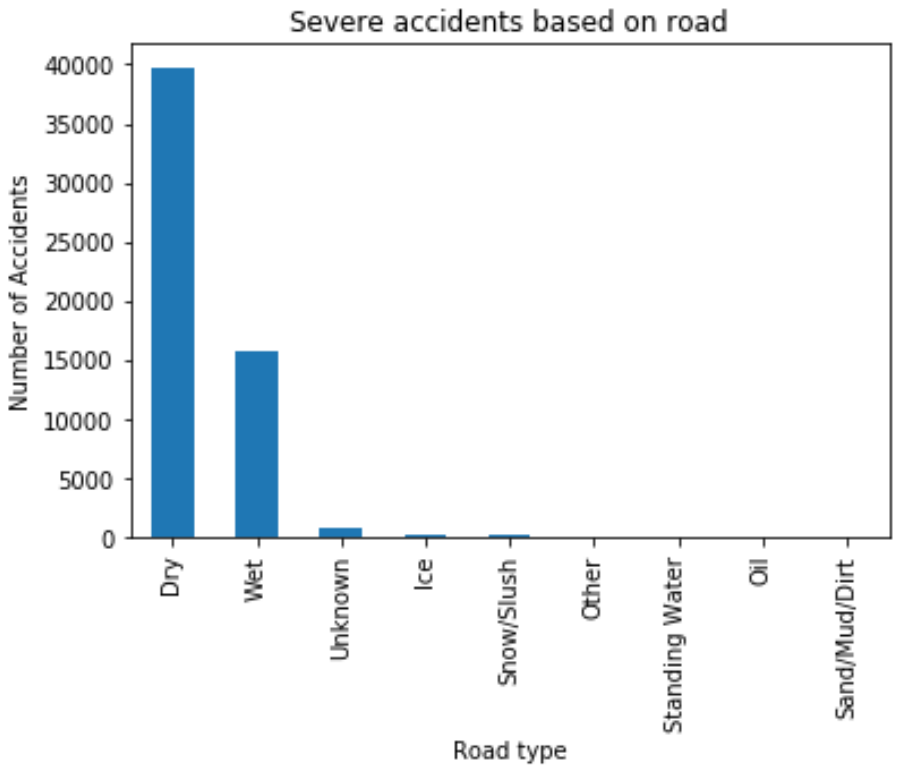
DATA

- I have considered Seattle Collision Dataset for solving this problem. I have downloaded this data set from https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Data-Collisions.csv It contains 194673 observations. It's having total 37 columns from which 37 are attributes and 1 is severity label.
- We are going to predict "SEVERITYCODE" as a target variable. Data set is already labled with 0 - Property damage only. 1- Severe Injury. We are going to use this data set for supervise learning data model. Before starting we need to clean the data as it's having many unnecessary attributes and missing values.

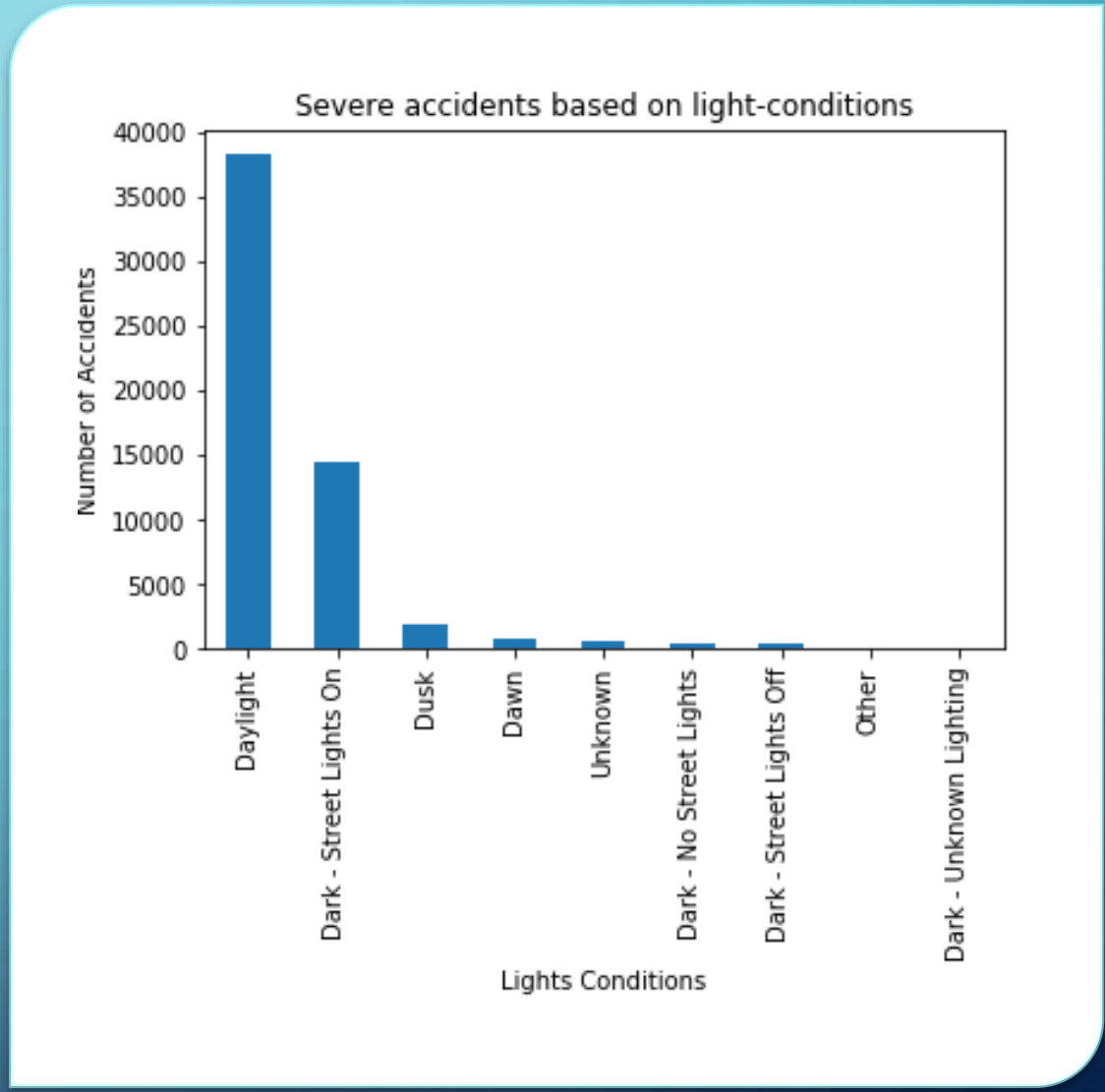
WEATHER CONDITIONS



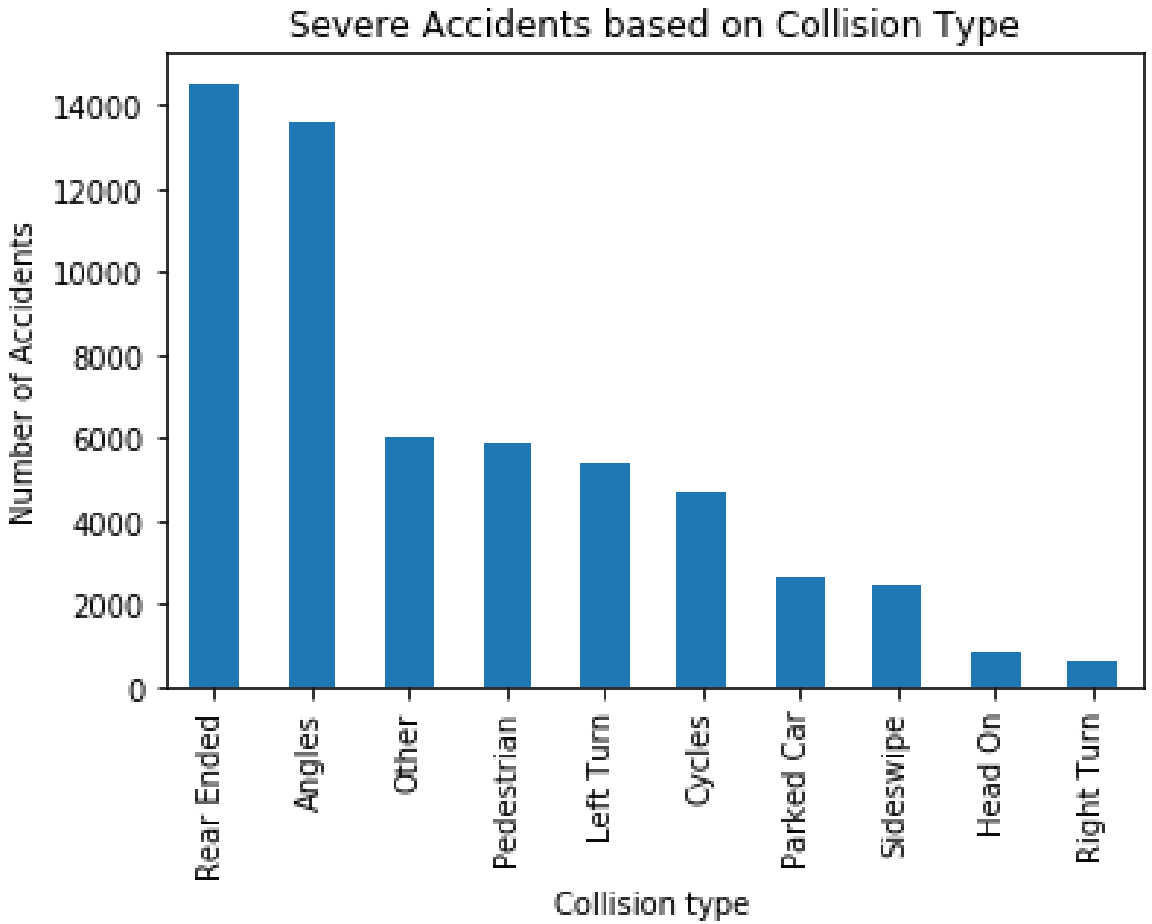
ROAD CONDITIONS

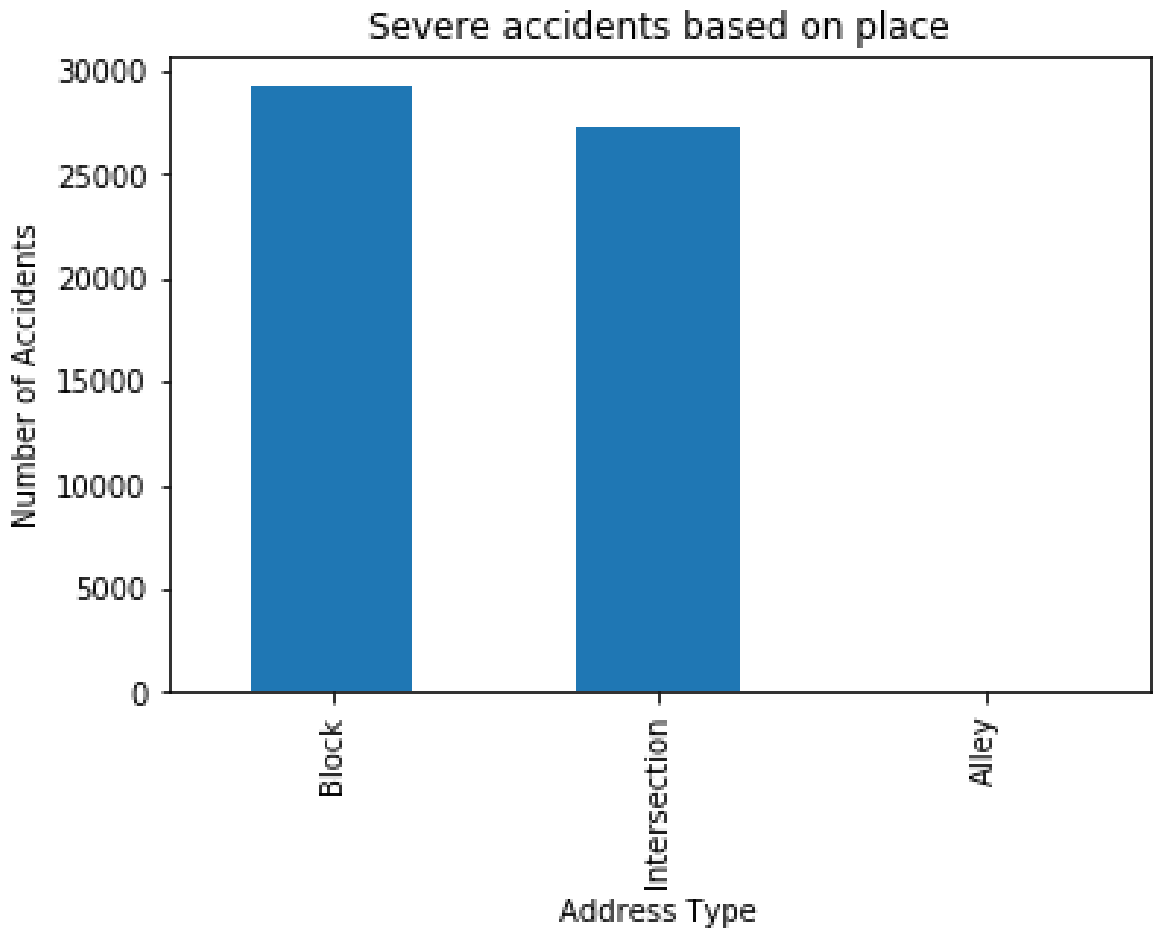


LIGHT CONDITIONS



COLLISION TYPE





LOCATION WISE

MODEL BUILDING

It is very essential to build a model that can be used to further development of the project and end up giving useful results in order to predict the severity.

In this analysis we are going to used following models as we find our categorical:

1. K Nearest Neighbor (KNN)
2. Decision Tree
3. Logistic Regression

We'll start with this importing necessary libraries and splitting the dataset into train and test dataset.

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X_rus, y_rus, test_size=0.2, random_state=4)
```

And then we implement our models accordingly.

MODEL EVALUATION

In this project F1-Score and Jaccard Score are used as evaluation metrics. And the table below depicts the result. Metrics / Models	KNN	Decision Tree	Logistic Regression
F1-Score	0.64	0.66	0.62
Jaccard Score	0.47	0.52	0.47

- Evaluation is important as it shows the clear picture of how much efficient the models were after being trained and tested.

CONCLUSION

- The exploratory analyses of the extracted dataset and the models that were built in order to develop a proper system that can predict car severity for the intended target audience mentioned in previous section of this documented report. This project is also going to help individual in determining the best model among chosen ones.