PREDICTION OF SEVERITY OF INJURIES DURING TRAFFIC CRASHES ON CHICAGO DATA

INTRODUCTION

Traffic accidents are responsible for most of the deaths. Especially, in USA there are millions of deaths reported every year.

Injuries during the accident depends on the several aspects like road conditions, weather conditions, road defects, speed limit, device conditions and light conditions in particular street.

So, I am working on the dataset of Chicago for prediction of severity of injuries and no injuries during traffic accidents due to various conditions in the year 2022.

BUSINESS PROBLEM

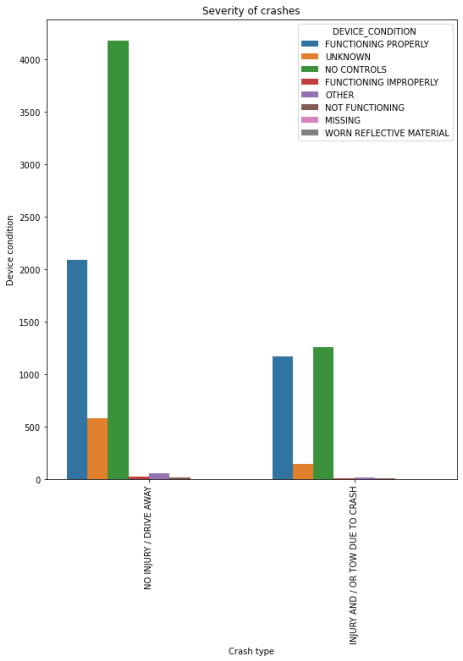
- Every year, the crashes reported in the city of Chicago are increasing, especially on weekends.
- Among them, 87.3% reported no injuries during the crashes and rest of them reported injuries and 0.1% fatal.
- In the city of Chicago, Western Ave street reported majority of crashes in which the primary reasons are unknown.
- The injuries are experienced by workers, passengers, pedestrians and cyclists.
- Predicting the reasons for the injury.

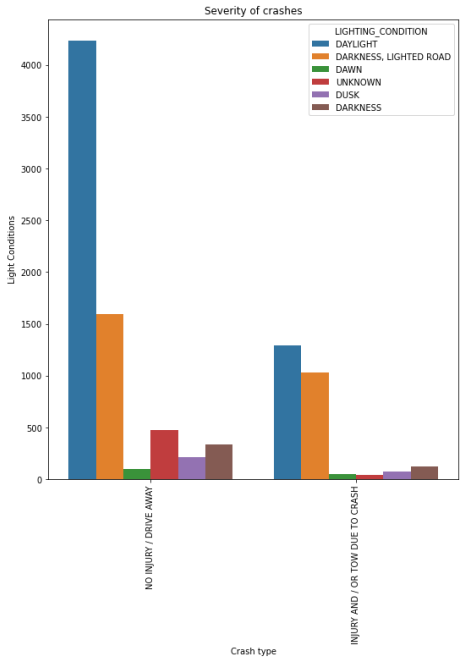
ANALYSIS

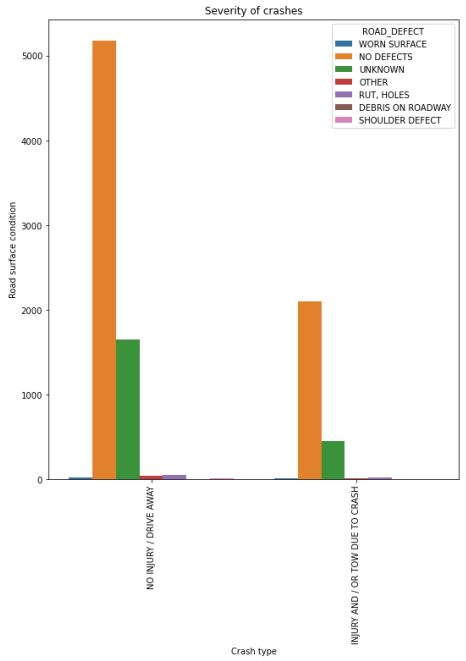
- Initially I got the data from Chicago data portal (Source given in Ref.) which consists of 9568 rows with 49 columns.
- Dropped few columns and rows by noticing skewness, outliers and null values.. . Standardization of column names and categorical values, detected outliers, checked correlation of data and removed columns with high correlation, removed columns with null values more than 90% that is not providing relevant data for the modeling and checked duplicate values for effective modeling.
- From the analysis, severity in most of the crashes is found to be "no injury" which means most of the crashes doesn't lead to injuries.

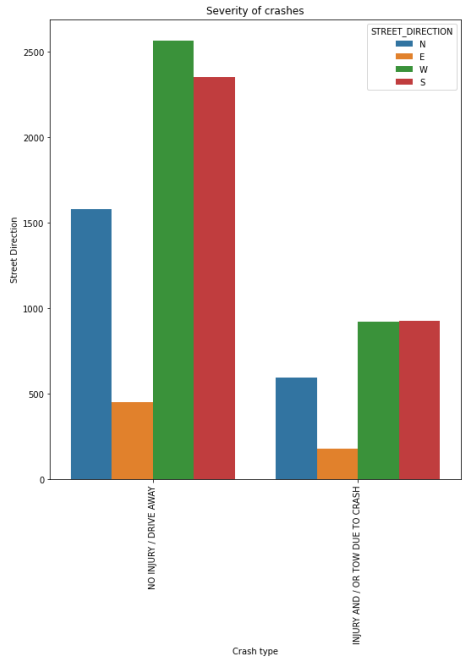
PROBLEM

- Severity of traffic crashes has been rapidly increasing and there are various reasons of this traffic crashes.

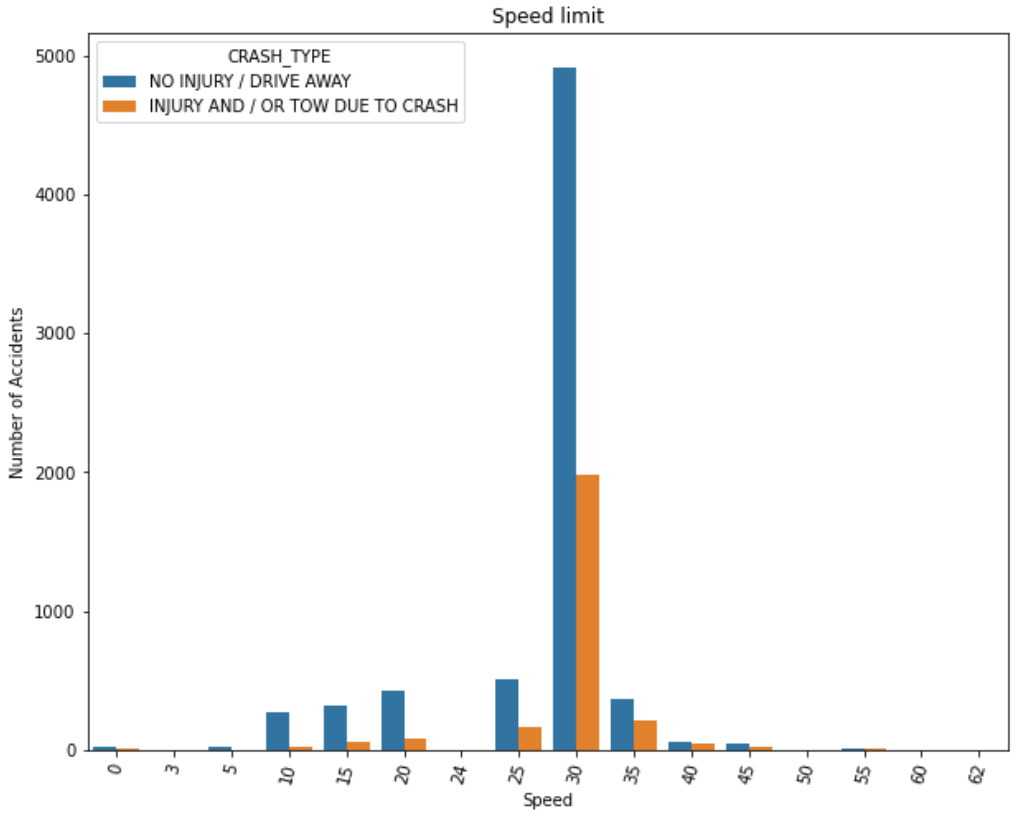


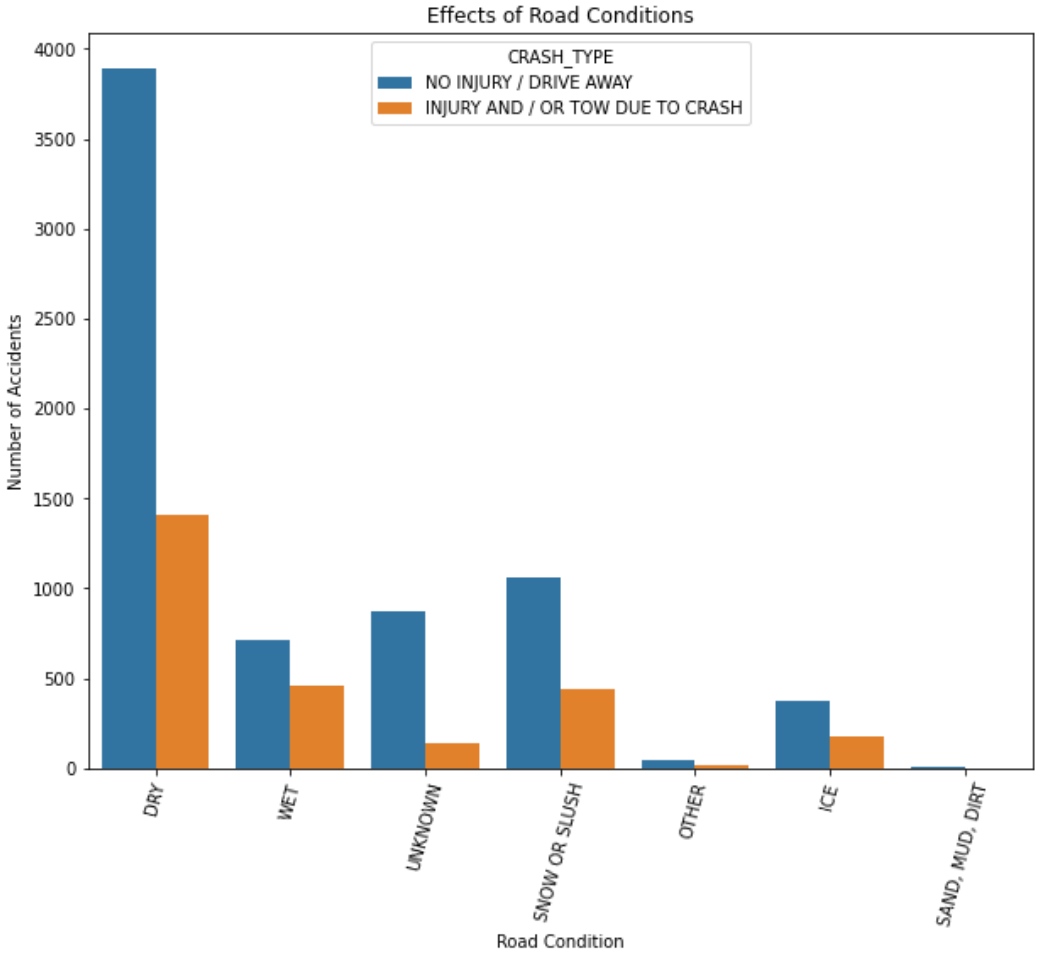






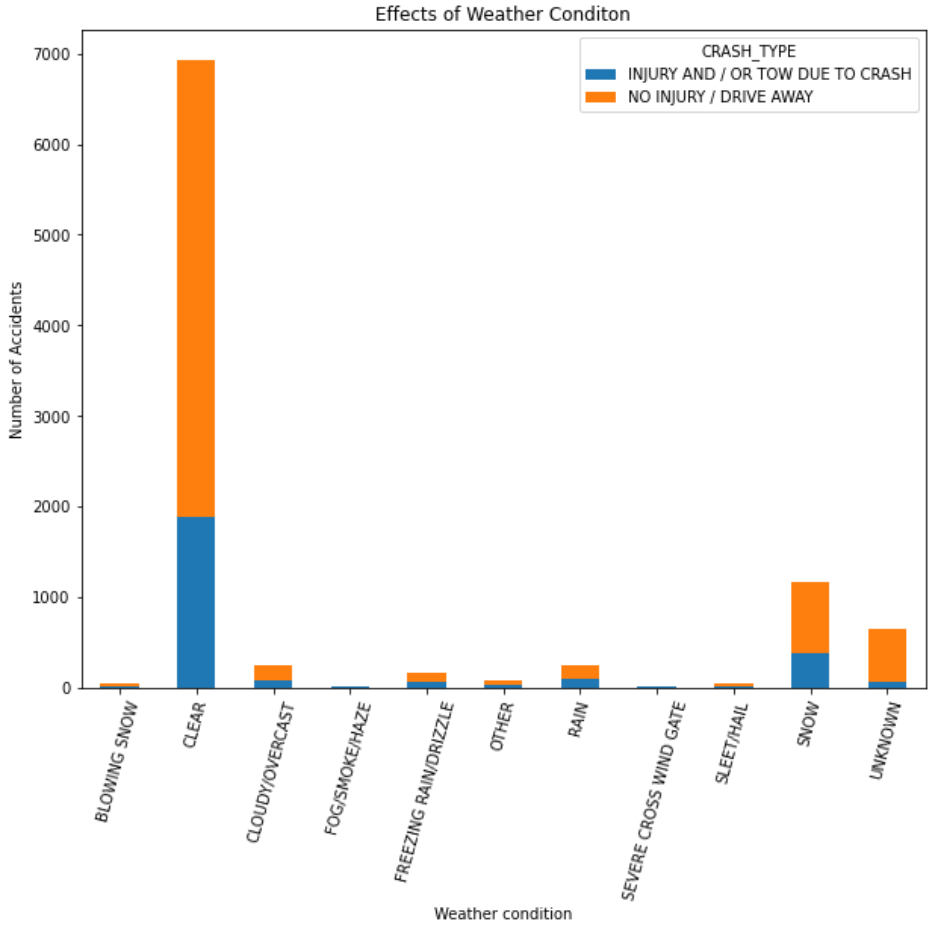
EFFECTS OF EXTERNAL FACTORS:





EFFECTS OF EXTERNAL FACTORS:

- Most of the crashes happened in the year 2022 is in clear weather followed by snow when the condition of road is “DRY”.
- The speed limit of 30 is more prone to crashes in the year 2022.
- The primary and secondary reasons for crashes are unknown.



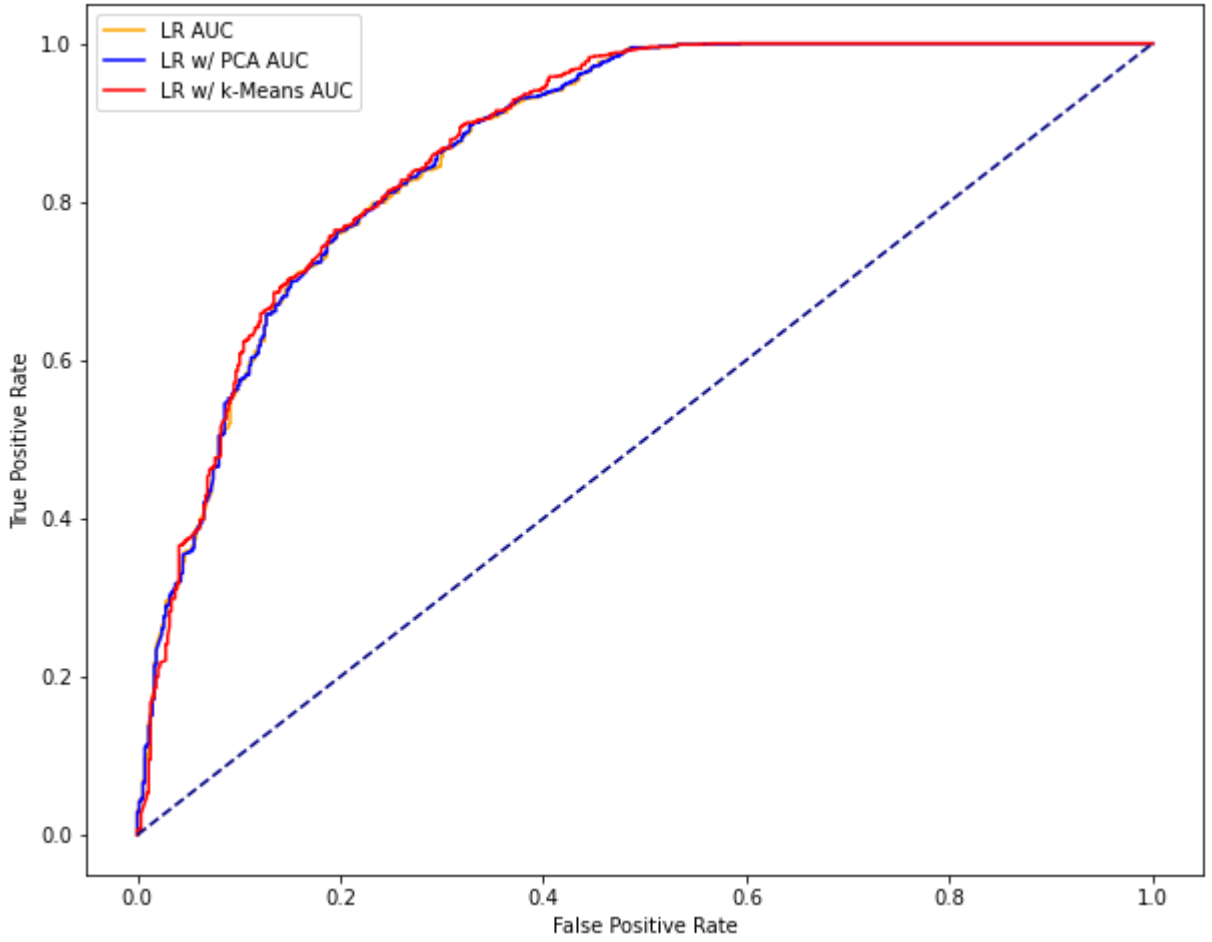
MODELING: LOGISTIC REGRESSION

Logistic regression performed with an accuracy of 85.53% and recall of 75.89% in test data

MODELING: DECISION TREES

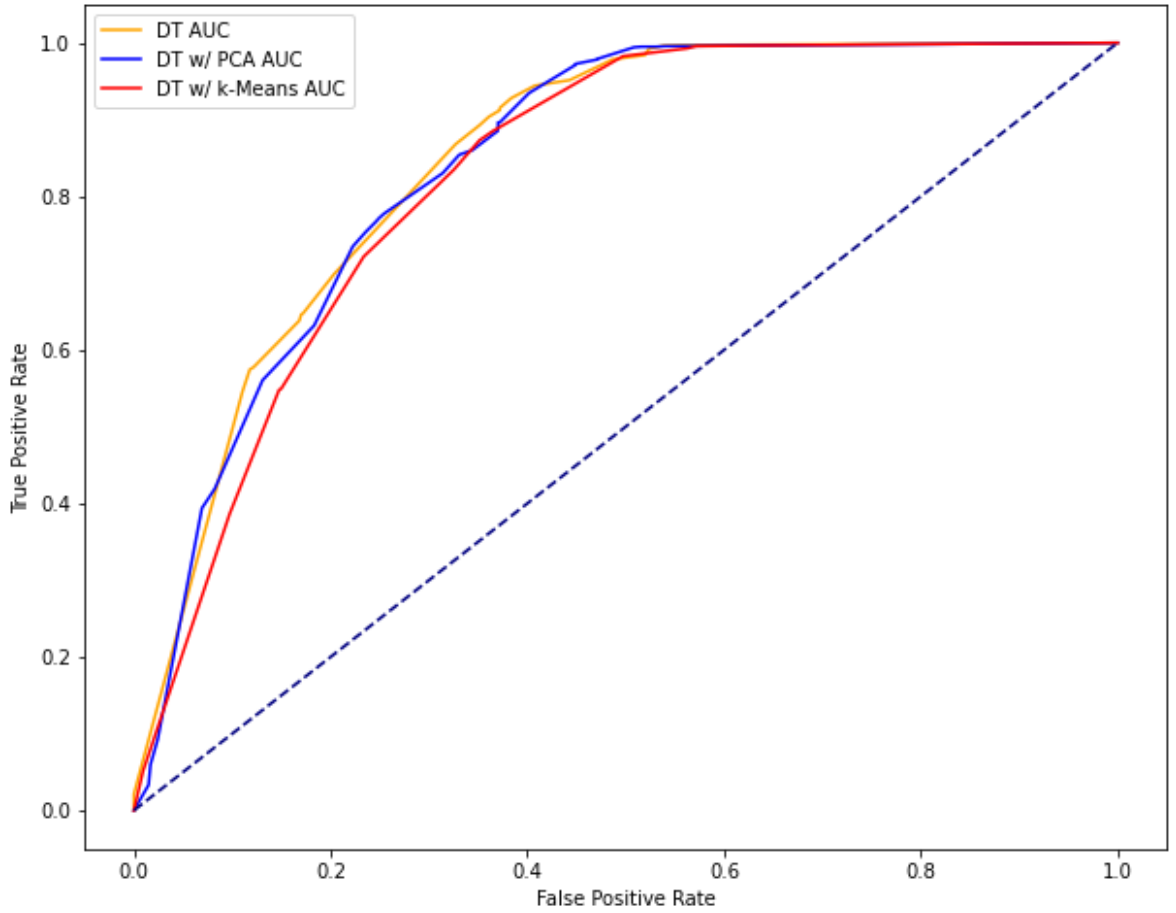
Decision trees performed with an accuracy of 81.56% and recall of 75.42% in test data

COMPARISON ON TEST DATA OF LOGISTIC REGRESSION



Logistic Regression AUC: 87.01%
Logistic Regression with PCA AUC: 87.06%
Logistic Regression with k-Means AUC: 87.51%

COMPARISON ON TEST DATA OF DECISION TREES



Decision Trees AUC: 84.71%

Decision Trees with PCA AUC: 84.34%

Decision Trees with k-Means AUC: 81.53%

RESULTS

- Logistic Regression with an accuracy of 85%, precision of 87%, recall of 75% and roc_auc of 87.51% test data is considered to be the better model

FUTURE SCOPE

- Reduce the severity of crashes and injuries during crashes.
- Safety measures.
- Reduces usage of defective devices.

References: Traffic Crashes - Crashes | City of Chicago | Data Portal

THANK YOU