Chicago Car Crashes Prevention

By Boula Akladyous

Problem Statement

Car Accidents Prevention system.

- The main objective of this project is to design a predictive model capable of analyzing and predicting the primary causes of car crashes that can be used as a detection system to predict road accidents in real time.

- Provide statistical analysis as a prevention tool aimed at reducing road accidents

Methodology

- Data Collection
- Scrub
- Exploratory Data Analysis
- Feature Engineering
- Model Building
- Model Evaluation
- Interpretation

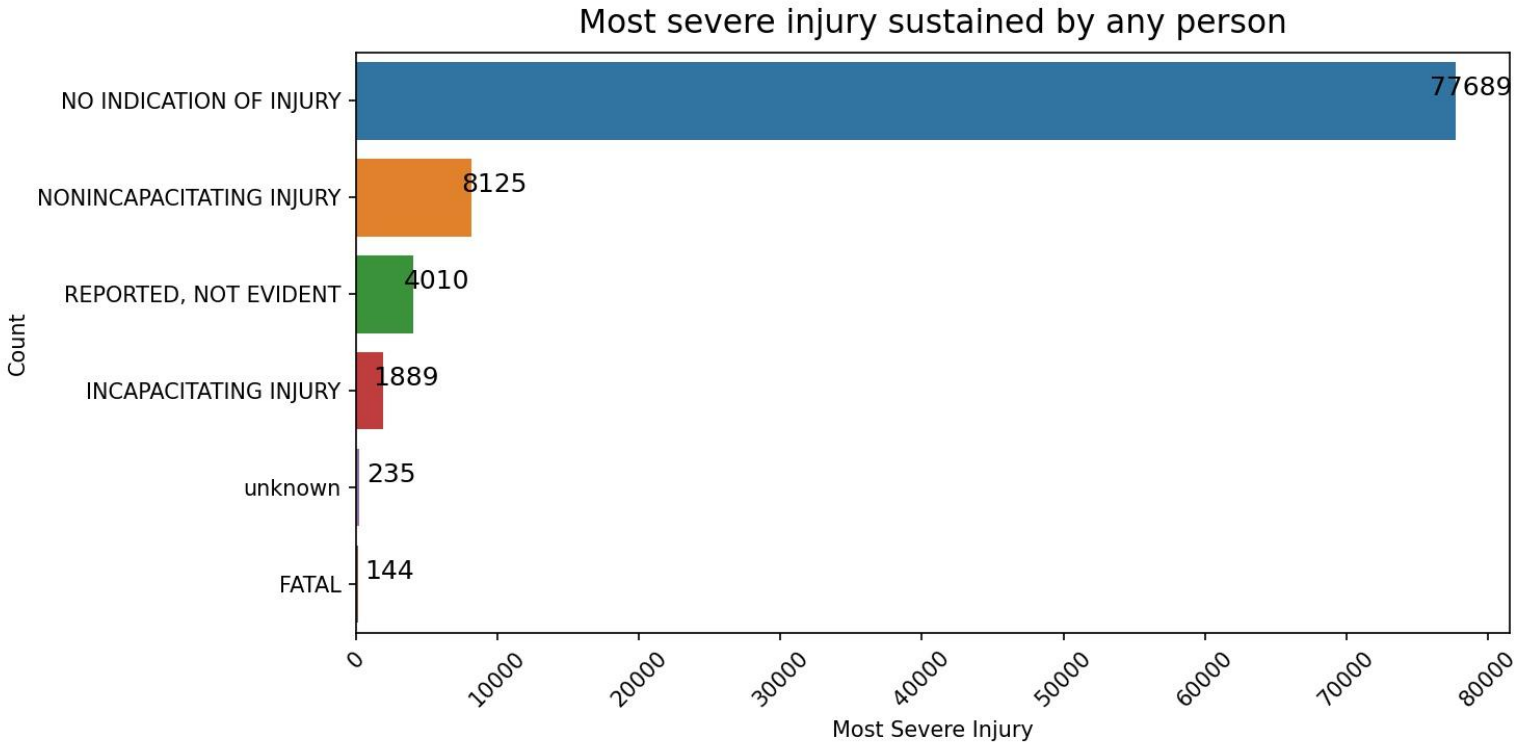
Road Accident Statistics

According to statistics from the Illinois Department of Transportation (IDOT) in the 2020 in city of Chicago there were:

- 92,092 car crashes of these, 144 accidents resulted in fatalities and 19,755 caused injuries and 12,526 involved visible injuries that did not incapacitate the victim.
- 72.7% of car crashes caused by speeding where the posted speed limit is 30 mph.
- 62.5% of car crashes occurs during daylight.
- 30,502 involved in HIT-AND-RUN road accidents.

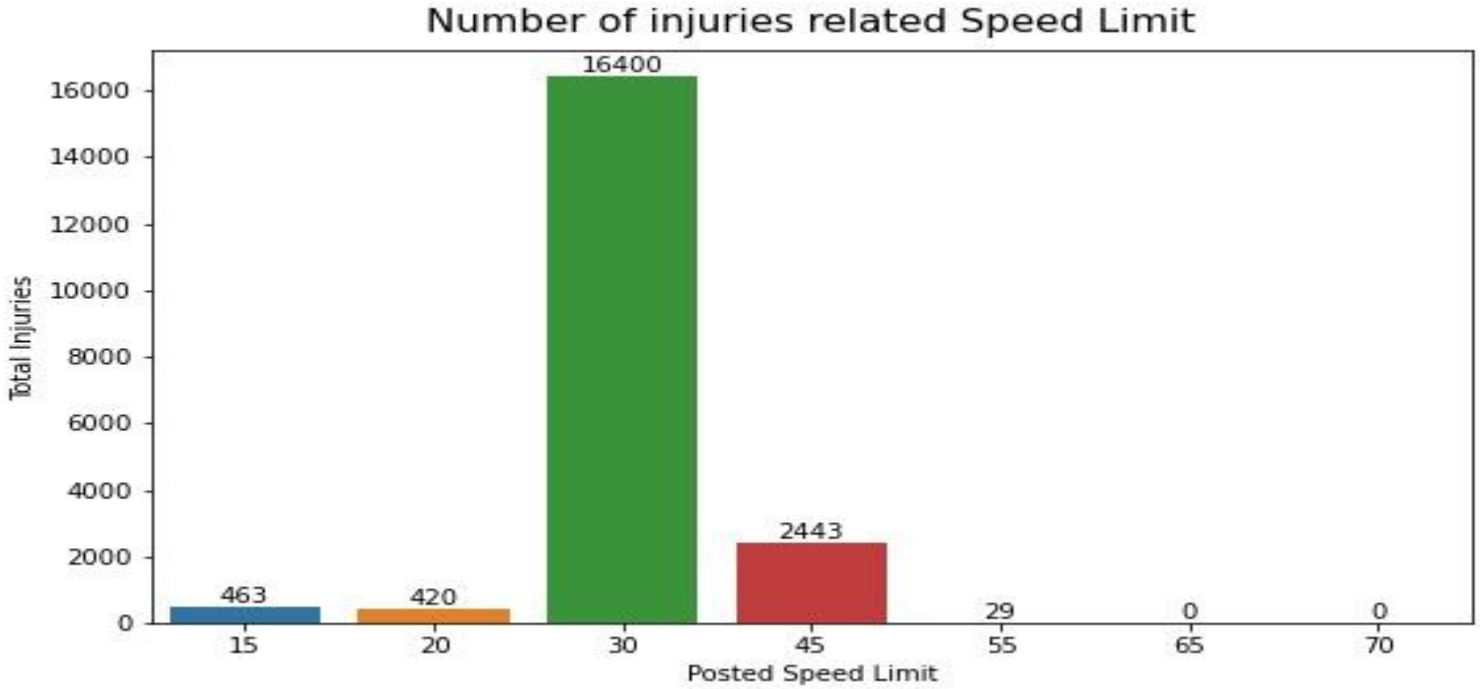
MOST SEVERE INJURY

Number of severe injuries involved in car accident



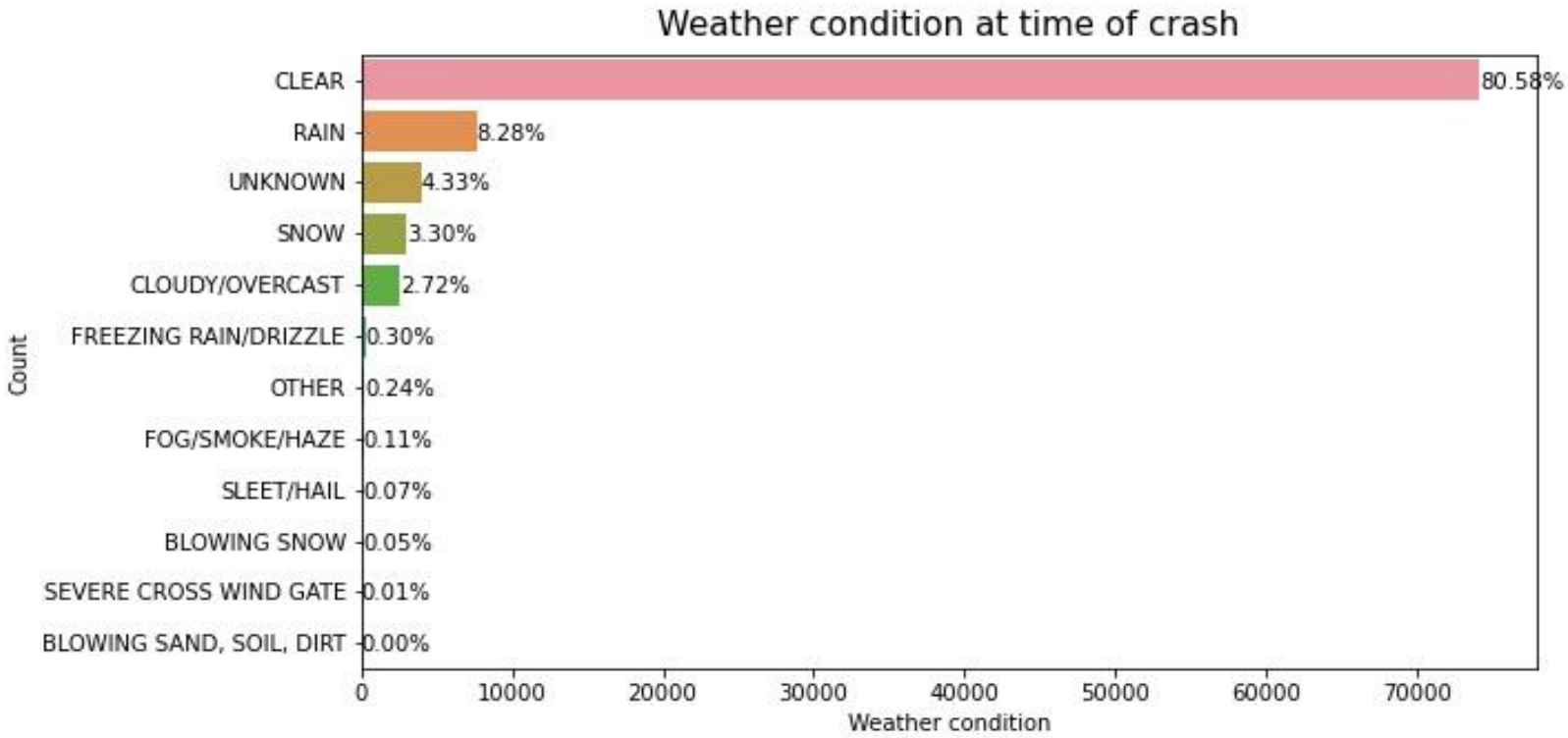
INJURY ANALYSIS

Number of injuries directly related to a driver exceeding the posted speed limit?



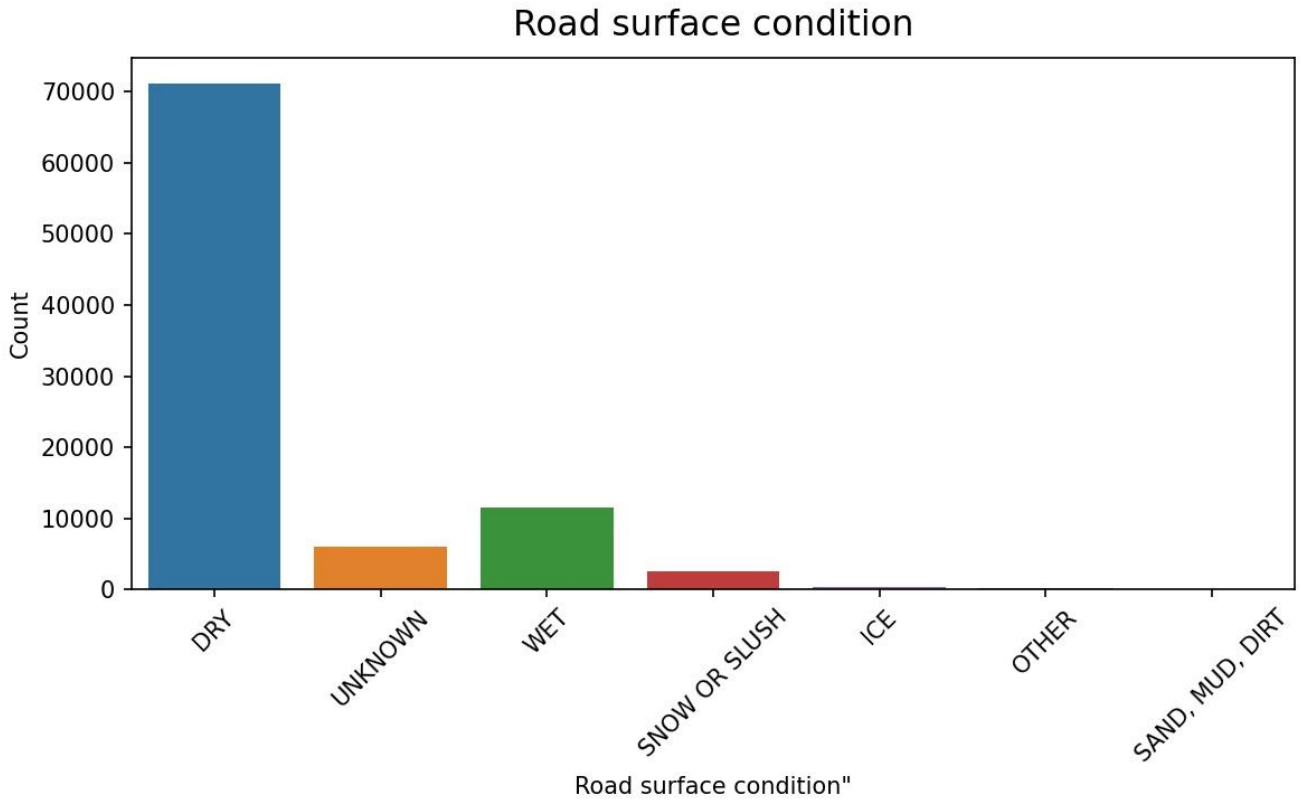
WEATHER CONDITION

Weather condition impact in road accident



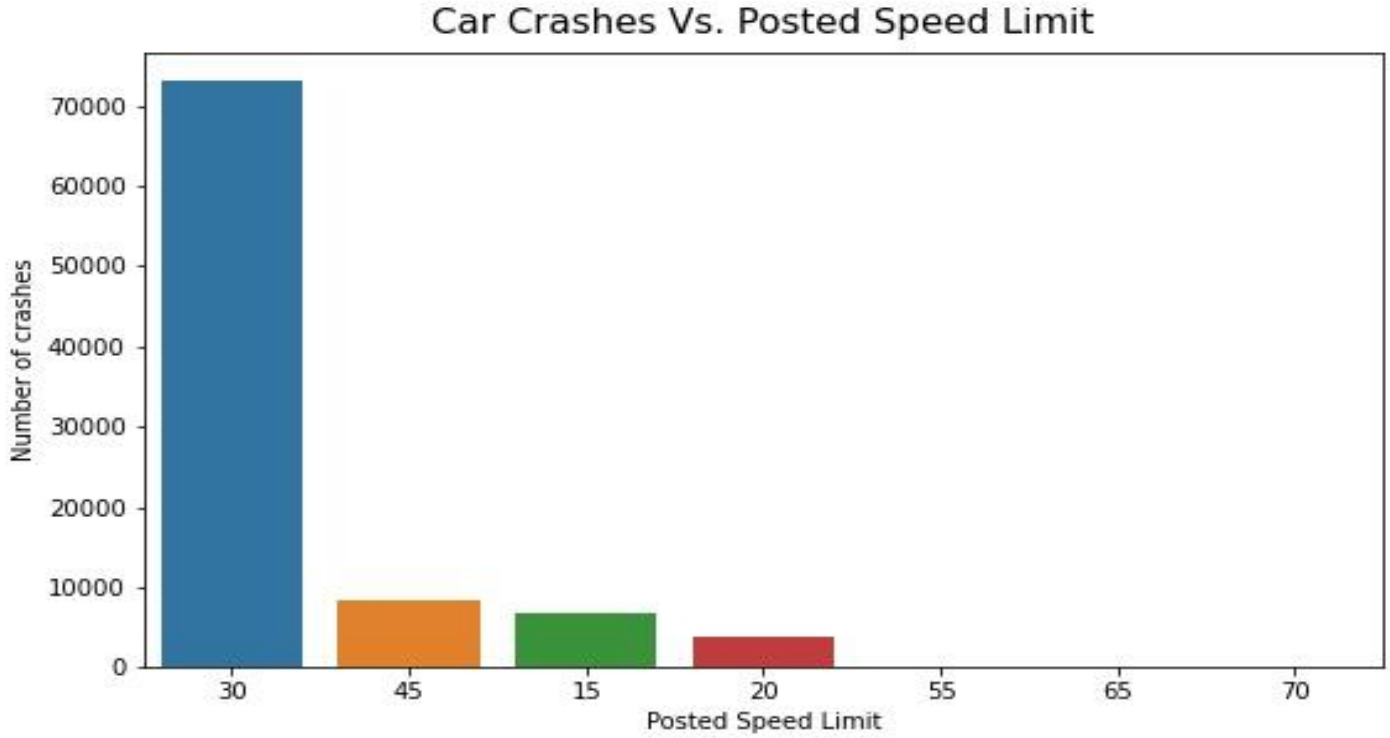
ROAD SURFACE CONDITION

Relative accident risk of different road conditions



Number of crashes related to Posted Speed Limit

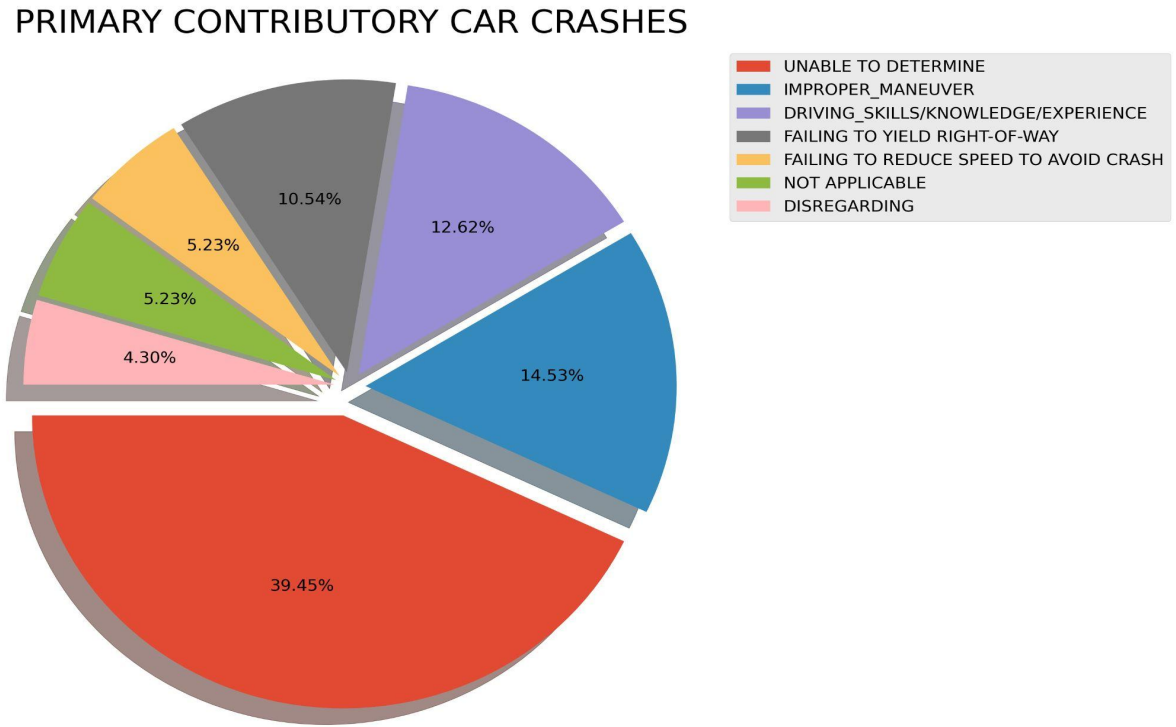
Relationship between posted speed limit and road accident



PRIMARY CONTRIBUTORY CAUSES

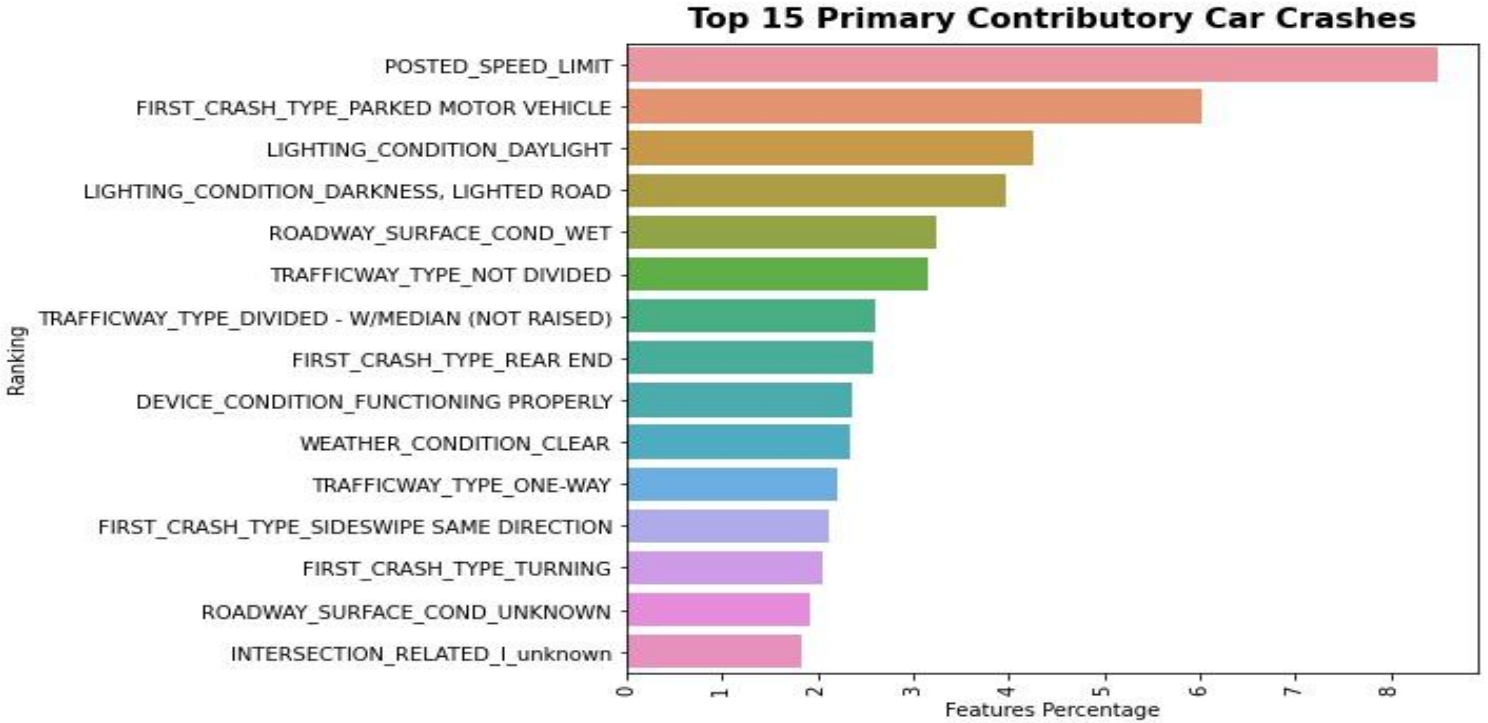
Most significant factors in causing the accident in 2020

- FAILING TO YIELD RIGHT-OF-WAY
- UNABLE TO DETERMINE
- OPERATING VEHICLE IN ERRATIC
- IMPROPER_MANEUVER DISREGARDING
- NOT APPLICABLE
- DRIVING_SKILLS/KNOWLEDGE/EXPERIENCE
- ROAD_CONDITION
- WEATHER
- PHYSICAL CONDITION OF DRIVER
- FAILING TO REDUCE SPEED TO AVOID CRASH
- ANIMAL
- ALCOHOL_DRUGS
- DISTRACTION
- EQUIPMENT - VEHICLE CONDITION
- TURNING RIGHT ON RED RELATED TO BUS STOP
- PASSING STOPPED SCHOOL BUS
- BICYCLE ADVANCING LEGALLY ON RED LIGHT
- MOTORCYCLE ADVANCING LEGALLY ON RED LIGHT



INTERPRETATION

Top 15 predicted features which represent the contributory causes of road accidents.



CONCLUSION

- Based on the data, 40% of accidents in the year 2020 were "UNABLE TO DETERMINE", followed by "FAILING TO YIELD THE RIGHT-OF-WAY" and "FOLLOWING TO CLOSELY" which represent 10% and 9% respectively.
- According to our classification model, the road accidents that seem to occur the most are caused by the speeding where the posted speed limit is 30mph.
- Random Forest Model predicts car crashes with 78% accuracy even dealing with a multiclass imbalanced dataset. the accuracy score between the 2 models (training 74.23% - test 78.03%) are too close which indicates that we avoided over-fitting.

FUTURE WORK

- improve modeling prediction by incorporating other datasets, namely the People and Vehicle dataset. These datasets contain additional information related to incidents, such as the make of the car and the physical condition of the drivers / passengers. Adding more relevant information would improve the accuracy of our models.
- investigate why traffic control devices are not present or not working
- Road Type Division: Segregate the different types of streets/roads to understand the unique properties of accidents that occurs in each