Chicago crashes project

BY - Jimcollins Wamae

Overview

Our objective was to develop inferential classification models for the Vehicle Safety Board of Chicago to advice them on two things.

1. What are the major causess of car accidents?
2. Were the accidents avoidable or unavoidable?

Our goal was to categorize the causes of car accidents into two groups: avoidable and unnavoidable.

Business understanding

The Vehicle Safety Board of Chicago (our stakeholder) has entrusted us with the task of conducting a comprehensive analysis of car crash data to enhance their understanding of the causes and factors influencing avoidable and unavoidable accidents in the region. Our project aims to delve deep into the available data, spanning from 2016 to 2020, to identify the primary contributory causes of crashes and uncover valuable insights that can drive actionable recommendations.

Data Understanding

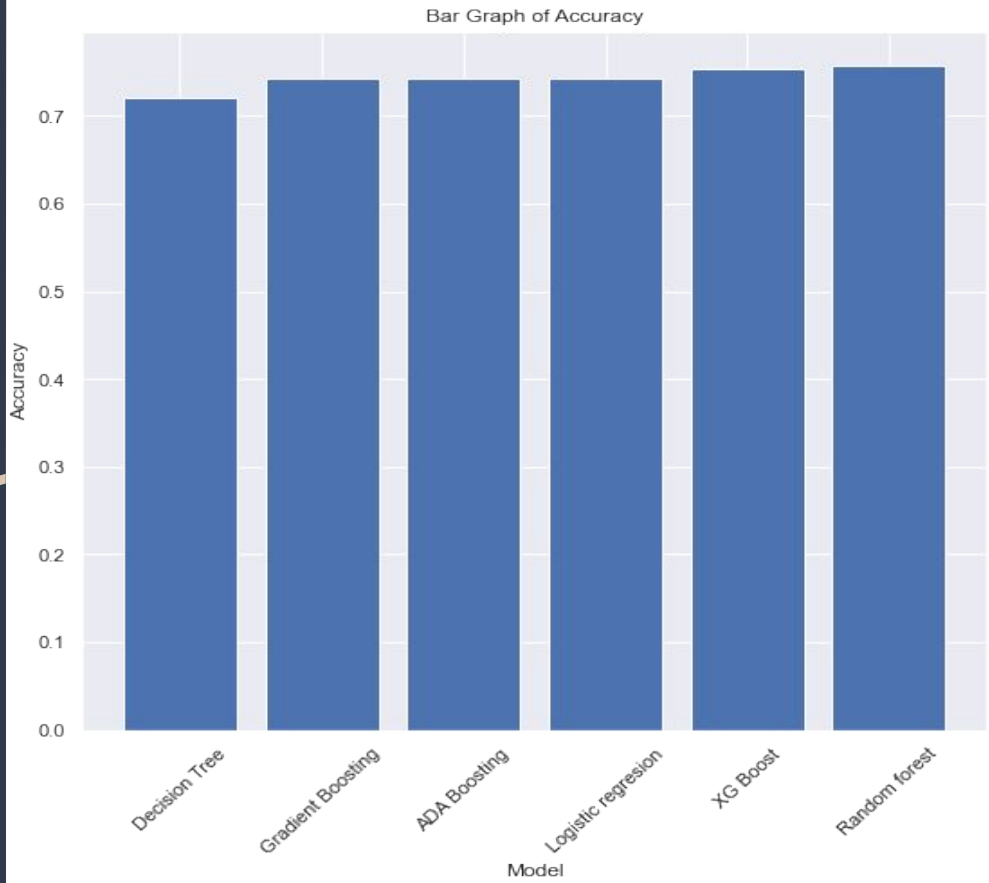
We used data from the City of Chicago Data Portal, which contains information about Chicago Car Crashes from January 2016 to December 2020.

We combined 3 different datasets that we found on the City of Chicago website. Those datasets were crashes, people, and vehicles.

The data has many categorical columns thus the need to use classification model metrics.

Modelling

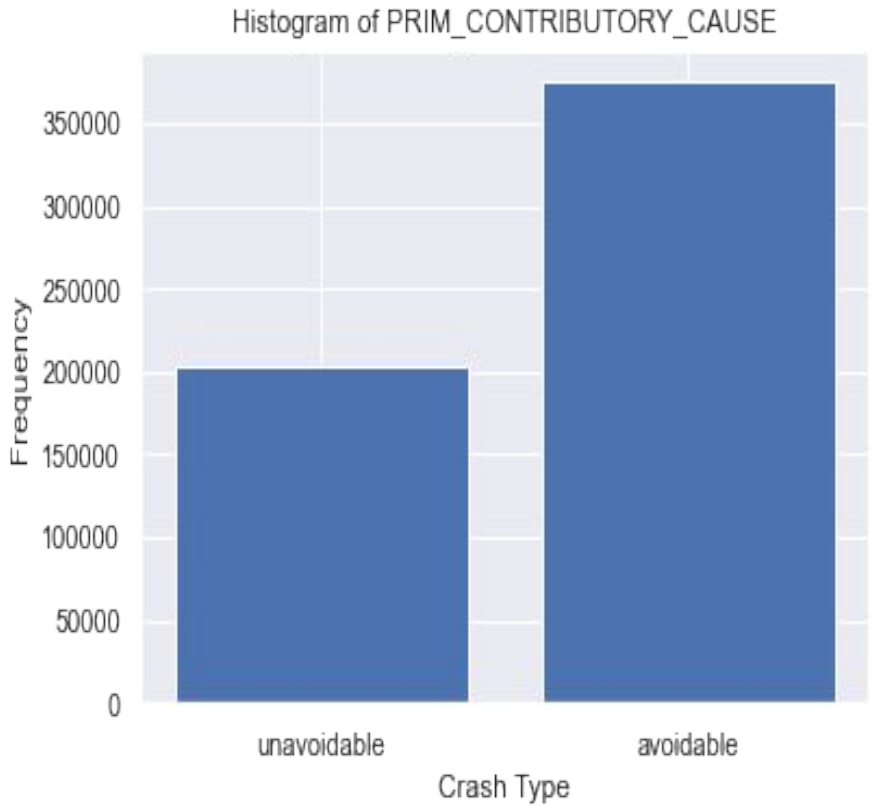
We deployed six models to predict our target variable the primary contributor cause against various predictors. We used the decision tree classifier as our best model and our best model was the random forest. Model performance was calculated using classification metrics such as accuracy.



Findings

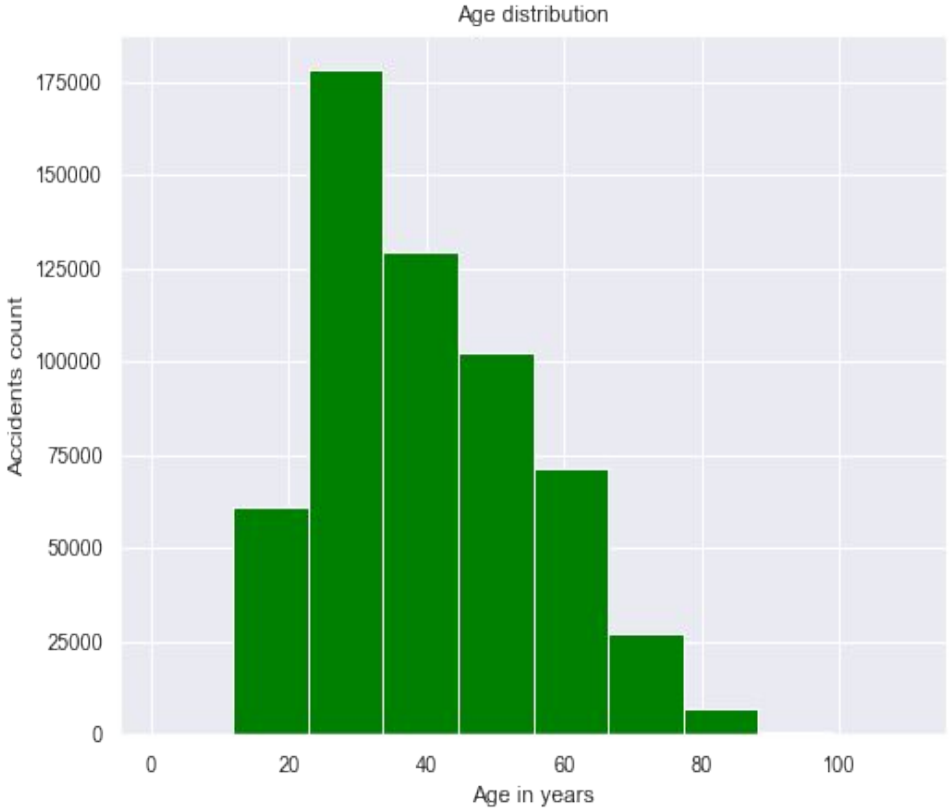
Avoidable vs Unavoidable

67 % of the accidents were preventable where only 37 were less preventable



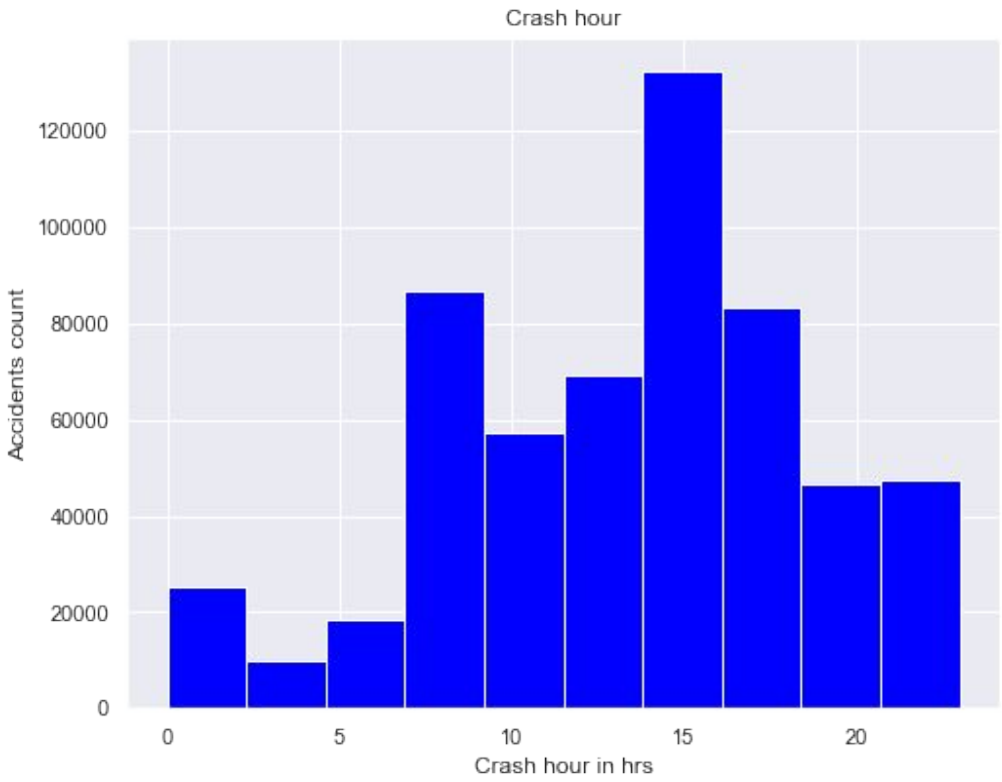
Age of Drivers

From our model we can identify that majority of the drivers in the accidents are aged 25-35 years.



Crash Hour

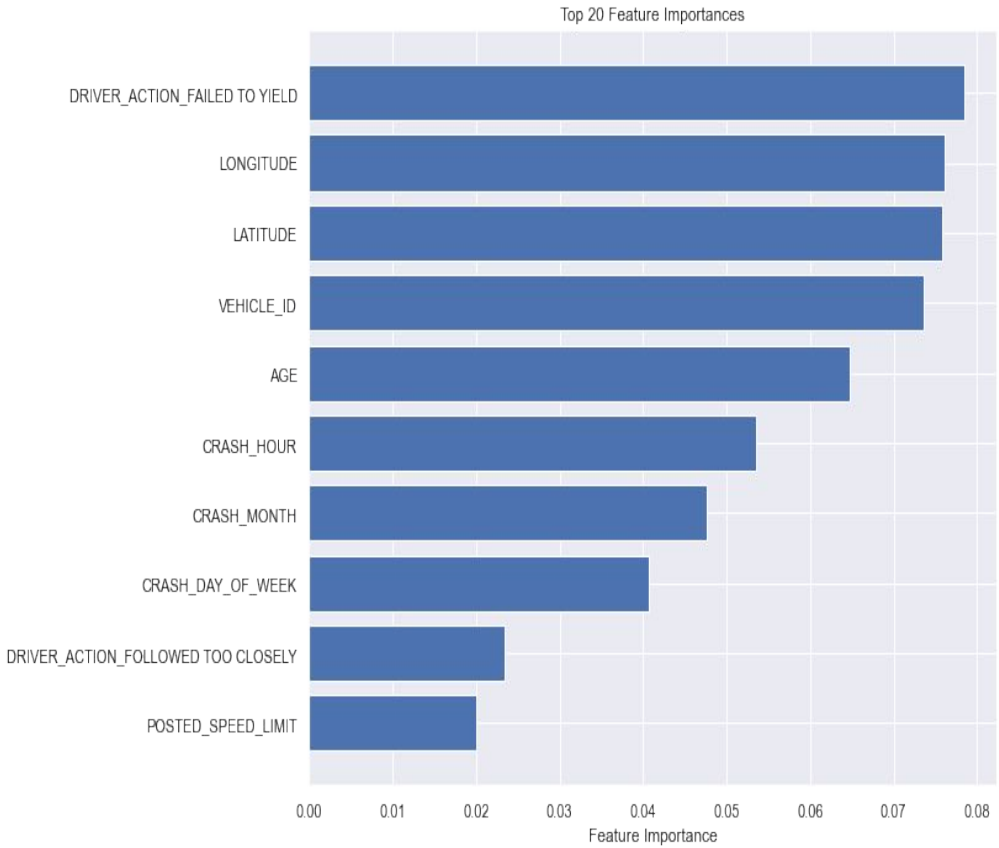
From our model most of the accidents occur from 12 noon to 4pm . Equally around 7am to 10am there are more accidents.



Feature Importance

we can identify that the most correlated feature to our target is Drivers action followed by location parameters.

This helps us identify that majority of the accidents are driver lead mistakes and hotspots in the city.



Conclusion

Avoidable vs Unavoidable crashes

Majority of accidents in our dataset were preventable, accounting for approximately 67%, less preventable accidents constituted around 37%

Based on this information, it is evident that focusing on preventive measures such as driver education, improved road conditions, and traffic law enforcement could greatly contribute to reducing the occurrence of preventable accidents.

Age of drivers

Understanding the age distribution of drivers involved in accidents (25-35) is crucial for developing targeted interventions and initiatives to improve road safety. By focusing on this specific age group, we can tailor driver education programs, awareness campaigns, and enforcement efforts to address the unique challenges and risk factors associated with drivers in this demographic.

Crash Hour

From the timeframes that most accidents occur (7am -10am) and (12 noon - 4pm) it mainly indicates that most motorists are in rush hour and the traffic congestion is higher at that time considering most people work from 8-5. However the underlying condition can be mitigated by more police control during these hours.

RECOMMENDATIONS.

1. Fix damaged or defective roads: We recommend prioritizing the repair of roads in hot spot areas that have a higher frequency of non-preventable crashes. By addressing infrastructure issues, such as potholes or inadequate signage, we can create safer road conditions and minimize the occurrence of such incidents.
2. Invest in an online driver and behavior education campaign: Implementing an online driver education program can have a significant impact on reducing preventable crashes. This approach is not only affordable but also easily accessible to a wide audience. By providing educational resources and promoting safe driving practices, we can enhance driver awareness and decision-making skills.

RECOMMENDATIONS.

3. Target the younger audience (age 20-39): Our analysis indicates that drivers within the age range of 20-39 accounted for a significant proportion of preventable crashes. Therefore, we recommend tailoring the driver education campaign to specifically target this demographic. By focusing on this age group, we can effectively address their unique driving behaviors and contribute to a substantial reduction in preventable accidents.
4. Increase traffic policing during the hours for 2 pm - 6pm and moreso on weekends. Most accidence seem to happen during this time period. Targeting peak traffic hours: By focusing on the time period of 2 pm - 6 pm, when traffic congestion is typically high, increased traffic policing can help manage traffic flow more efficiently. This can reduce the likelihood of accidents caused by reckless driving, speeding, or aggressive behavior during peak hours.

NEXT STEPS

1. Targeted Driver Education: Develop and implement targeted driver education programs for the age group of 25-35 years, focusing on addressing common driver behaviors and promoting safe driving practices.
2. Enhanced Traffic Policing: Increase traffic policing during high-risk hours, particularly from 12 noon to 4 pm and 7 am to 10 am, to enforce traffic laws, deter negligent driving behaviors, and raise awareness about responsible driving.
3. Road Safety Improvements: Advocate for road safety improvements in areas with high accident rates, addressing issues such as road defects, traffic flow management, and visibility to create safer driving environments.
4. Monitoring and Evaluation: Continuously monitor and evaluate the impact of implemented interventions and programs, collecting and analyzing data on accident rates, driver behaviors, and road conditions to assess effectiveness and make necessary adjustments.

THANK YOU

Questions

- ……
- ……
- ……
- ……

- JIMCOLLINS WAMAE
- Email - jimwamaef12@gmail.com
- Linkedin – jim wamae
- Phone - +254-722614538.