



Chicago Car Crashes

By Olaide Kashimawo



Outline

- Business Understanding
- Data
- Exploratory Data Analysis
- Result
- Recommendation

Business Understanding



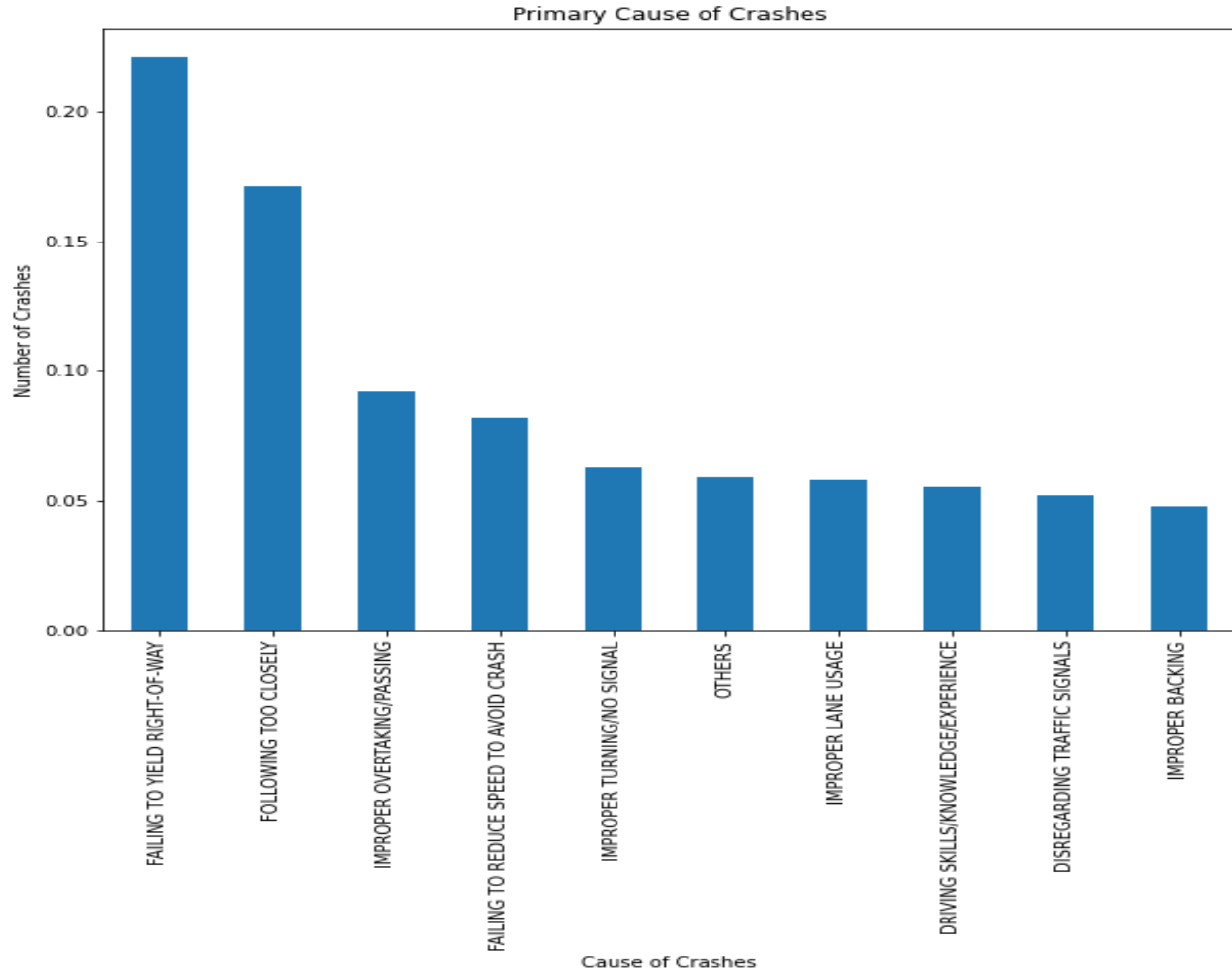
- This project seeks to create a multi-classification model to accurately identify the primary contributory cause of car crashes in Chicago
- I will analyze the data of car crashes to determine which factors are most likely to lead to injury and build a predictive model to alert help emergency medical services dispatchers of the potential injury.

Data

- According to statistics from the state of Chicago department of transportation in 2022
- with an average of 70,000 reported each week and 20,000 crashes caused injuries
- Chicago department of transportation Crash Dataset contains almost 500,000
- There were three datasets that we observed, "Traffic Crashes - Crashes", "Traffic Crashes - Vehicle", and "Traffic Crashes - People" which were merged into one data.



Primary Causes of Crashes

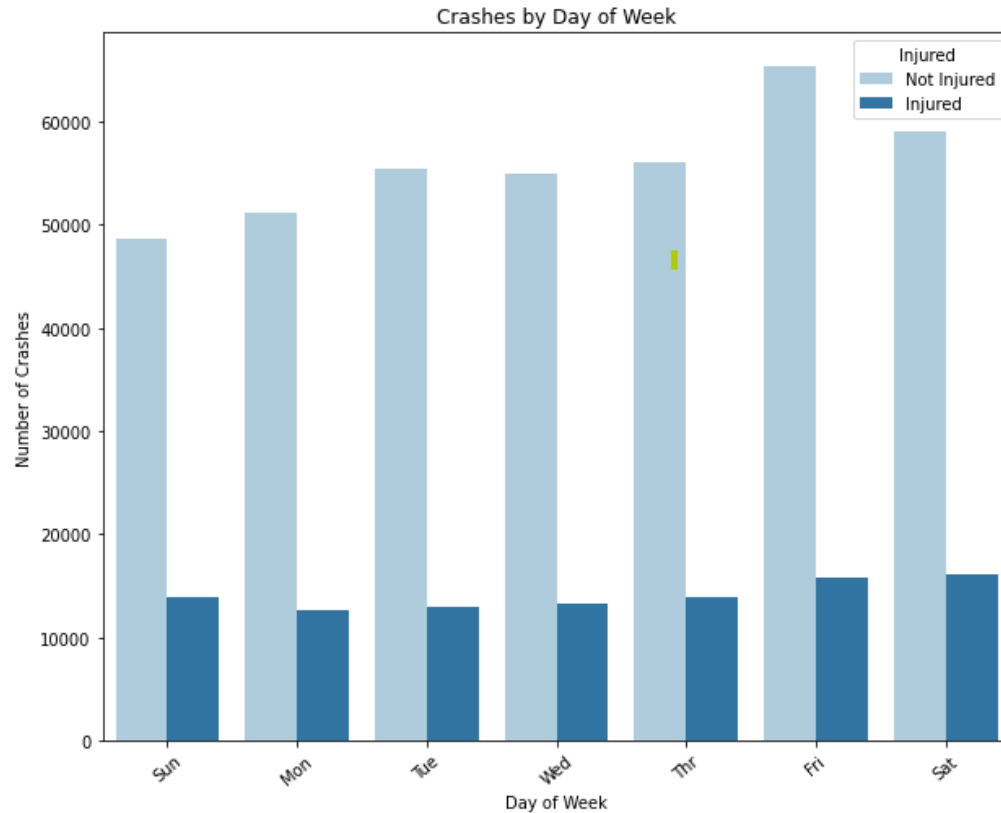


Some major causes of crashes are;

- failing to yield the right of way
- following too closely
- improper overtaking/passing
- failing to reduce speed to avoid a crash

Crash day of the week

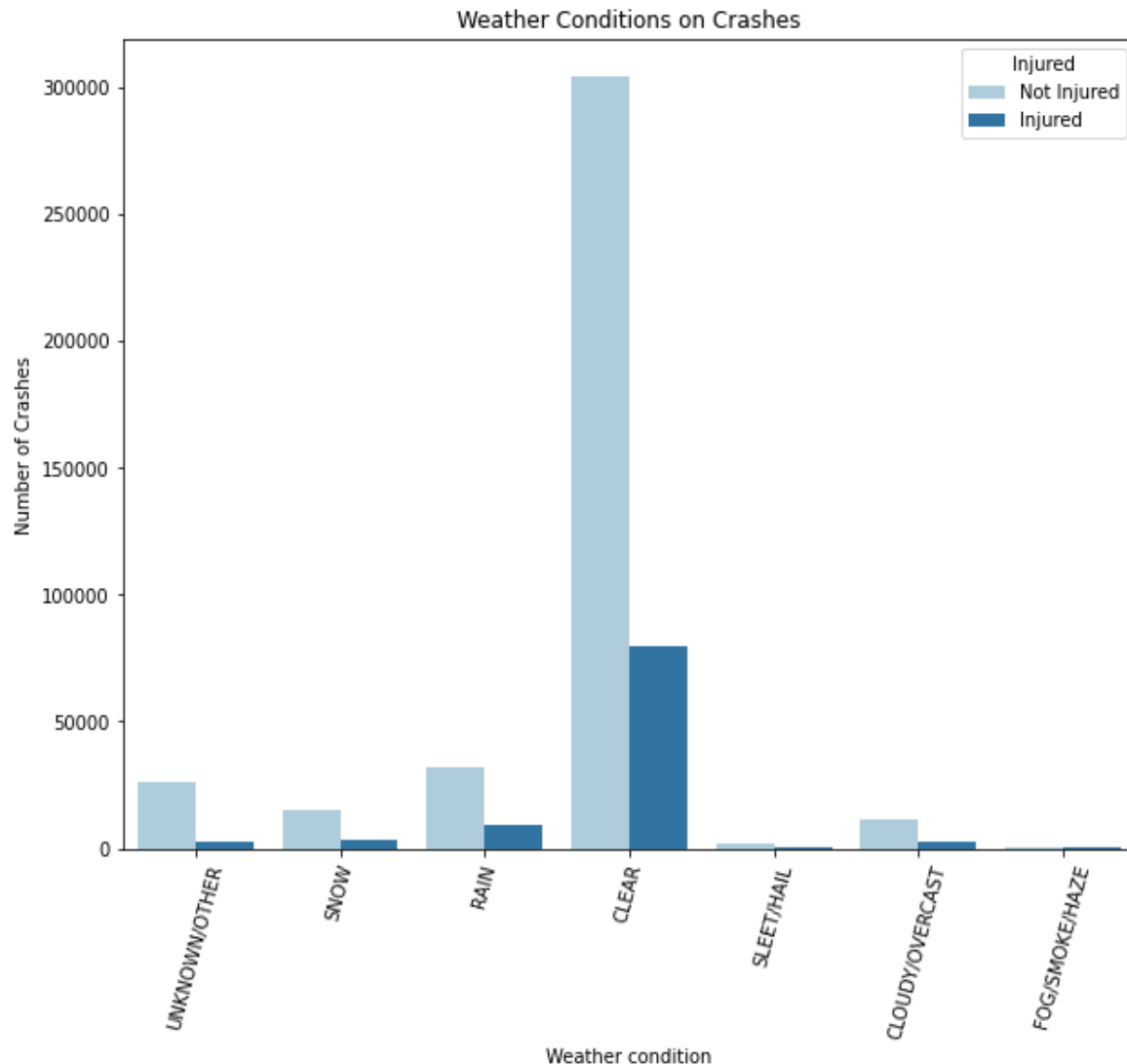
Relationship between crashes and injury



Most accidents occurred on weekends mainly (on Fridays and Saturdays)

Weather Conditions

Relationship between Weather conditions and



From the visualization above we can see most crashes happens in broad daylight



RESULT

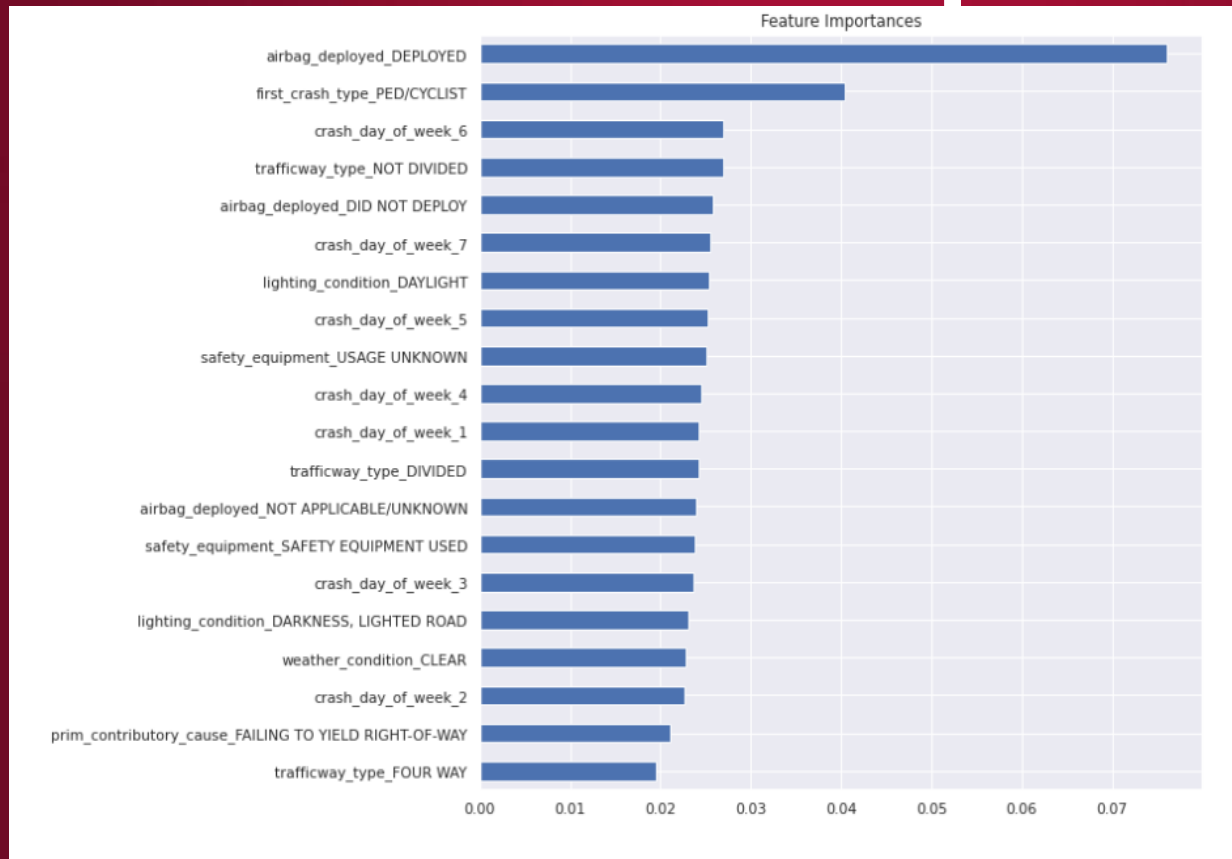
Model Result

MODEL	ACCURACY TESTING	SMOTE-ACCURACY	F1-SCORE 0 TESTING	F1-SCORE 1 TESTING	SMOTE F1-SCORE 1 TRAINING
Logistic regression	0.71	0.71	0.79	0.52	0.52
KNearest Neighbors	0.80	0.88	0.88	0.46	0.88
Decision Tree	0.81	0.96	0.89	0.44	0.96
Random Forest	0.83	0.96	0.90	0.57	0.96

Overall, it seems that the resampling technique using SMOTE has significantly improved the performance of the Decision Tree and Random Forest models, as shown by their high F1 scores on the resampled data. I recommend the following because the Random classifier has a 96% smote accuracy rate;



Feature Importance's



- The airbag deployed
- Pedestrians/Cyclist, and
- Traffic ways not divided have the highest coefficients with
- injuries occur mostly on Saturdays



Recommendations

- Increase enforcement of traffic laws and regulations, particularly with respect to distracted driving and failure to maintain a proper lane.
- Installation of physical barriers between opposing lanes of traffic, wider lanes, and better marking of lanes and road edges.
- Encourage the use of advanced driver assistance systems (ADAS) for instance blind spot detection, lane departure warning, and automatic emergency braking.



Next Step

To further analyze to better understand the factors that contribute to crashes;

- Analyze driver behavior and roadway design.
- Binning data to find crash locations can suggest lowering a speed limit or adding a traffic signal soon.



Thank you!

GitHub link: [HTTPS://github.com/olaidekashimawo/](https://github.com/olaidekashimawo/)
Email: goldprint3@outlook.com