PREDICTING PRIMARY CAUSES FOR CAR CRASHES IN CHICAGO

MICHAEL FLORES

FLATIRON SCHOOL

INTRODUCTION

- As part of an effort to eliminate car crash fatalities and injuries, the city of Chicago has initiated the Vision Zero commitment.

- Traffic crashes **ARE** preventable and predictable.

- Designing predictive models will help us to analyze causes for crashes and prevent them from happening.

OUTLINE FOR PRESENTATION

- Where does our data come from and how was it filtered?
- What observations do we come across upon analyzing the dataset?
- What were the most effective models for predicting causes for car crashes?
- How can we apply our results towards improving traffic safety?

OBTAINING AND CLEANING THE DATA

OBTAINING THE DATA

- The data comes from the Traffic Crashes dataset, readily accessible from the Chicago Data Portal.
- Also linked with it are datasets corresponding to vehicles and people related to the crash.

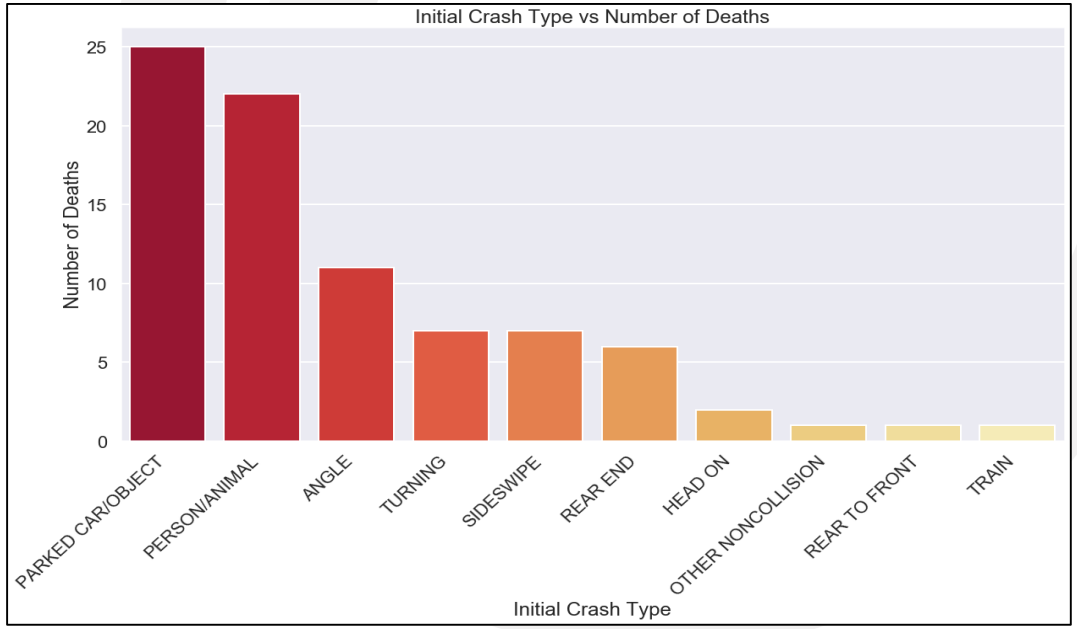
CHICAGO DATA PORTAL

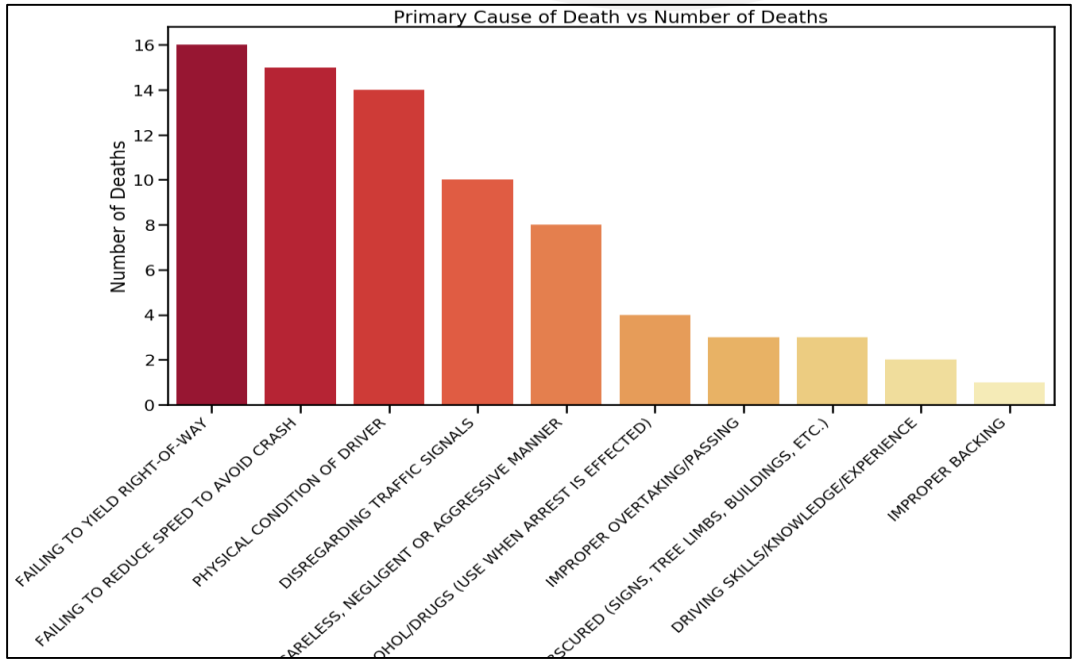
CLEANING THE DATA

- I looked at car crashes from July 2019 – July 2020.
- We only analyzed incidents in which we had a definitive cause.
- We ended up with nearly 65,000 crash incidents for our model.

EXPLORING THE DATA

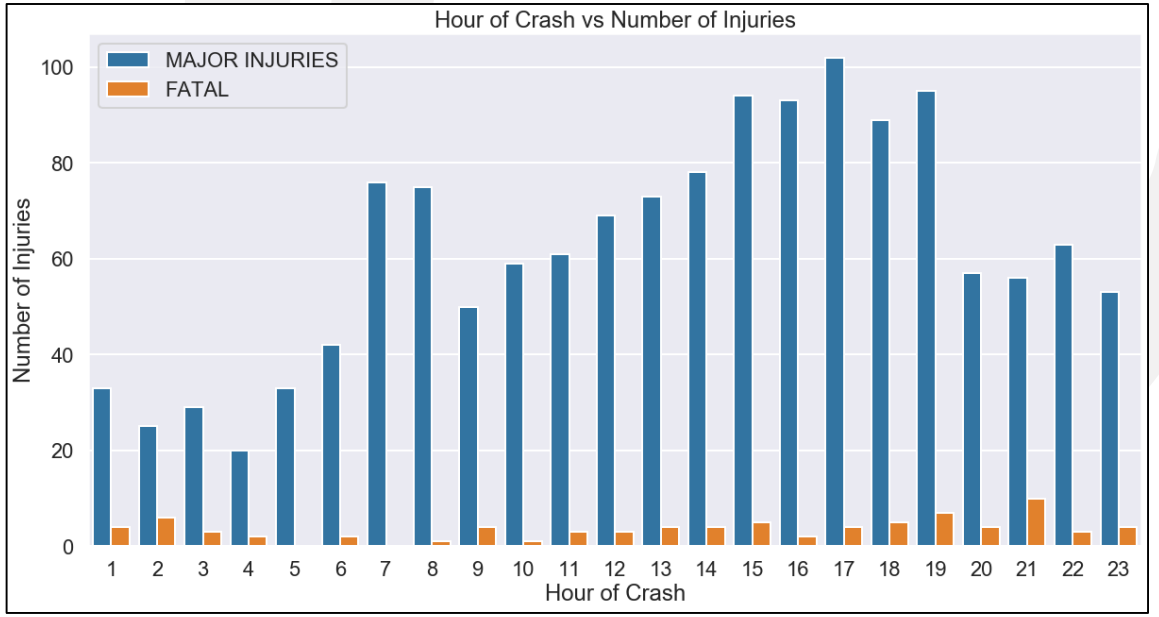
- I wanted to first examine the incidents that resulted in fatal injuries to determine if there were any trends.
- Crashes into stationary cars and objects as well as certain driving maneuvers resulted in deaths.
- Primary causes for crashes that resulted in deaths were related to the driver's condition and failing to act in a safe manner.

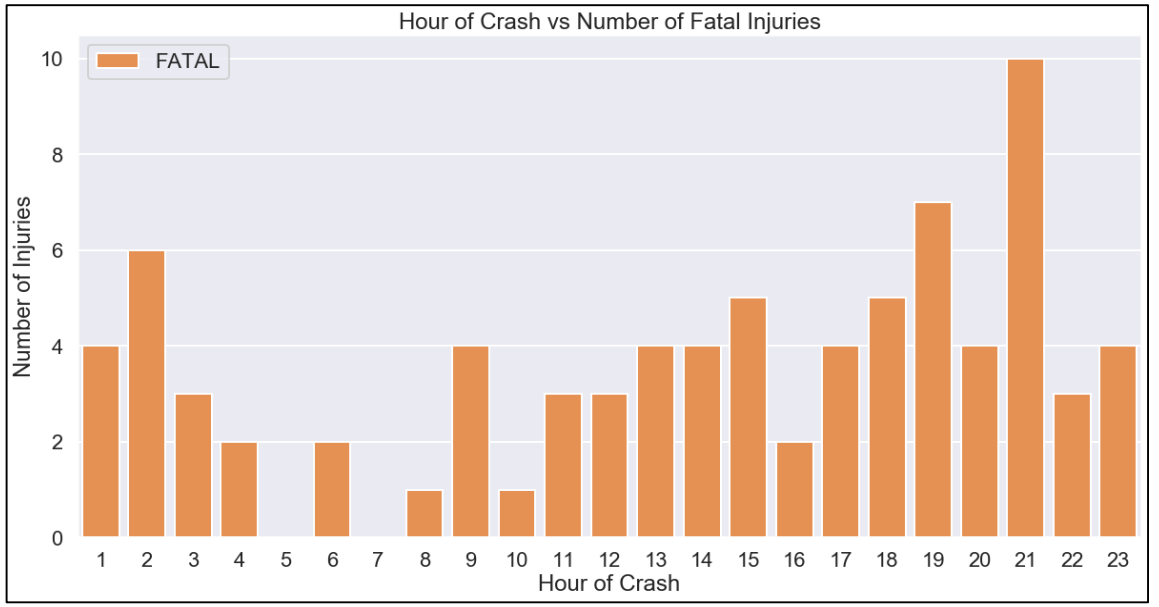




EXPLORING THE DATA (CONT.)

- Next, I wanted to observe any kind of trends with regards to the time of the incidents.
- More serious car crash incidents occur during peak commuter hours and late at night.
- Fatal crashes were most prevalent around 7pm-3am.



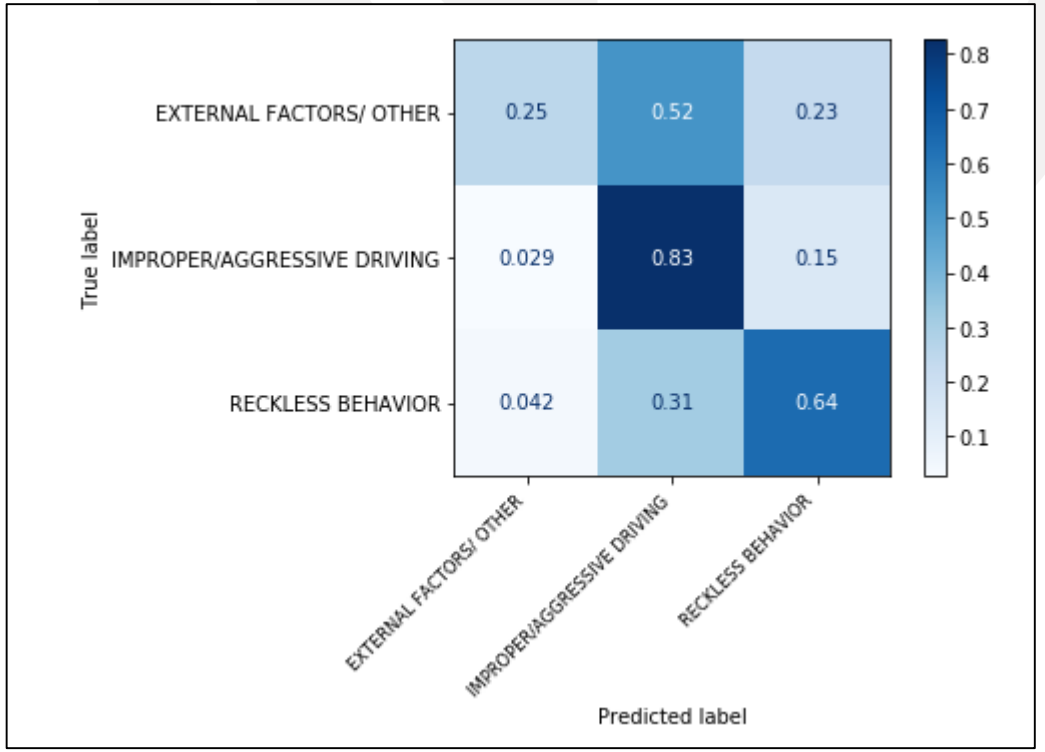


MODELING THE DATA

- Our means of classifying causes for crashes were as follows: Improper/Aggressive Driving, Reckless Behavior, External Factors/Other.

- We tested multiple models for predicting the primary cause such as Decision Trees and K-Nearest Neighbors.

- Our final model with XGBoost had an accuracy of roughly 74%.



RECOMMENDATIONS AND FUTURE WORK

- Further promotion of driving at slower speeds and less aggressive driving.
- Further promotion of driver safety, especially those who are active late at night.
- Continuing to test our model's accuracy with historical and recent data.
- Collaborating with other major cities involved in Vision Zero such as NYC, San Francisco, and LA.

THANK YOU!

Michael Flores

Flatiron School