

CHICAGO TRAFFIC ACCIDENTS



PHINIDY GEORGE

BUSINESS CASE



National Problem: Traffic deaths are surging despite decreased pandemic driving. This is the worst increase since 1924.

Pedestrian Safety: Nighttime pedestrian deaths are up 67%, highlighting a need for safer crossings and better visibility.

Illinois Status: Over 1,000 lives were lost in Illinois crashes in 2019 alone. The state lacks key safety laws, receiving a "Caution" rating.

Missing Laws: Illinois lacks All-Rider Helmet, Booster Seat, Graduated Driver's License provisions for nighttime/passengers/minimum permit age/unrestricted license.

The Solution: Chicago DOT is working with Vision Zero to use crash data for targeted education, road improvements, and other safety measures



PROJECT OBJECTIVE

Our Approach: Building on these efforts, we will analyze crash data to predict crash severity based on injury outcomes.

OVERVIEW

THE DATA

Our Chicago Crashes dataset came from the City of Chicago's data portal.

The records in the dataset date from 2019 to 2022.

Initially, there were 65 columns to explain but my final dataset has roughly 24 columns



UNDERSTANDING THE PROBLEM

STEP 1

Identify which features for accidents are relevant so as to ascertain likelihood of injury in future accidents

STEP 2

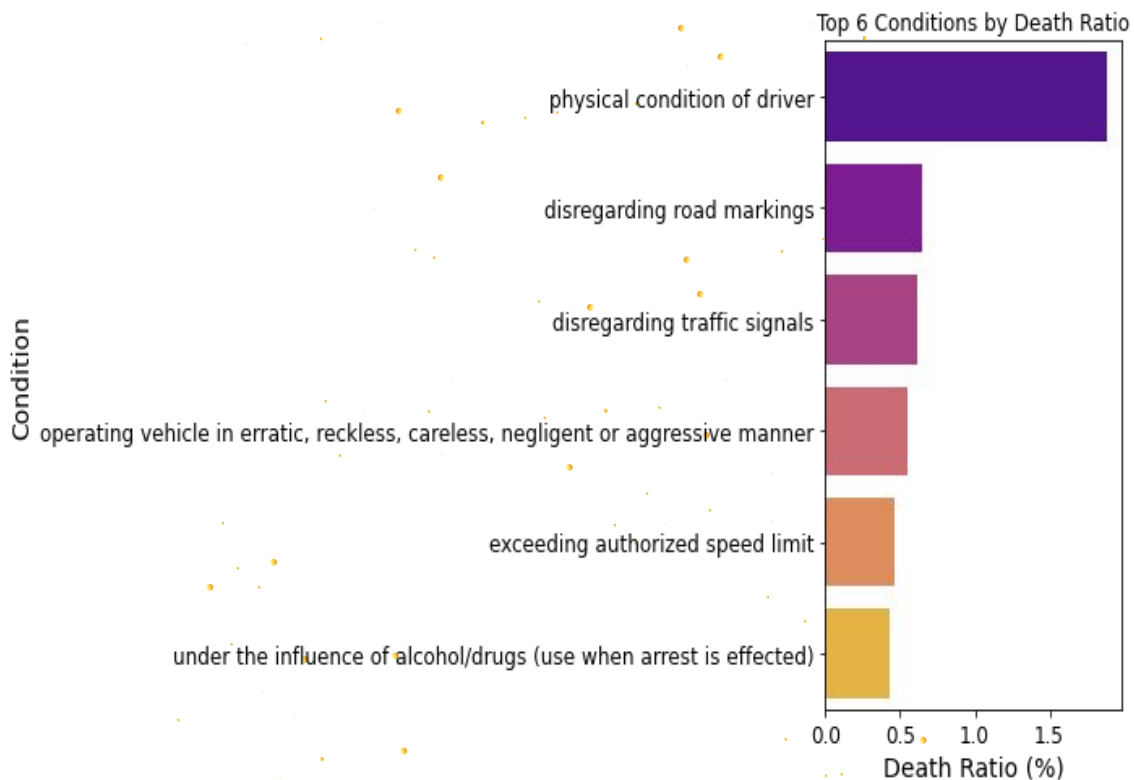
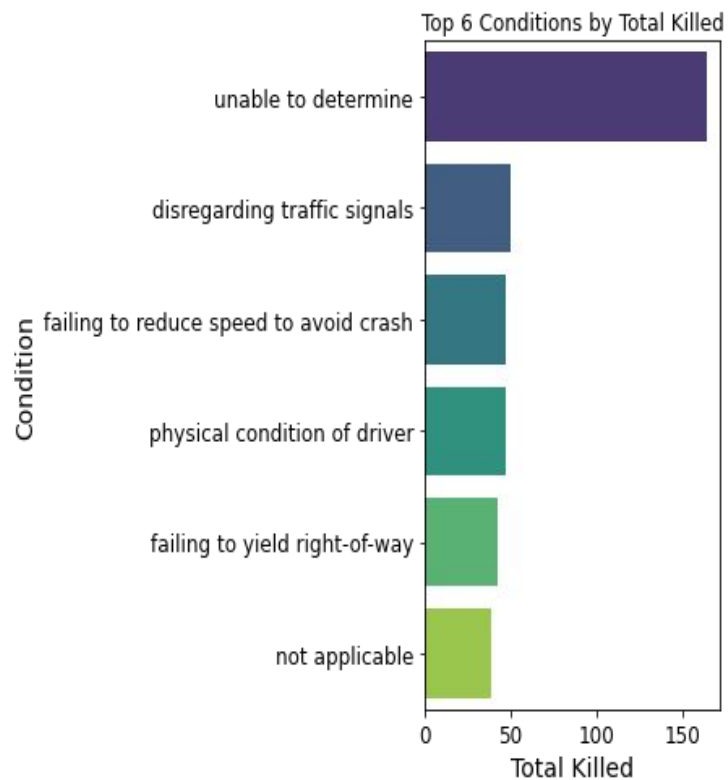
Run various models in order to select which classification models will be the best predictors for our dataset

STEP 3

Tune parameters in 2 of our best baseline models to optimize model accuracy

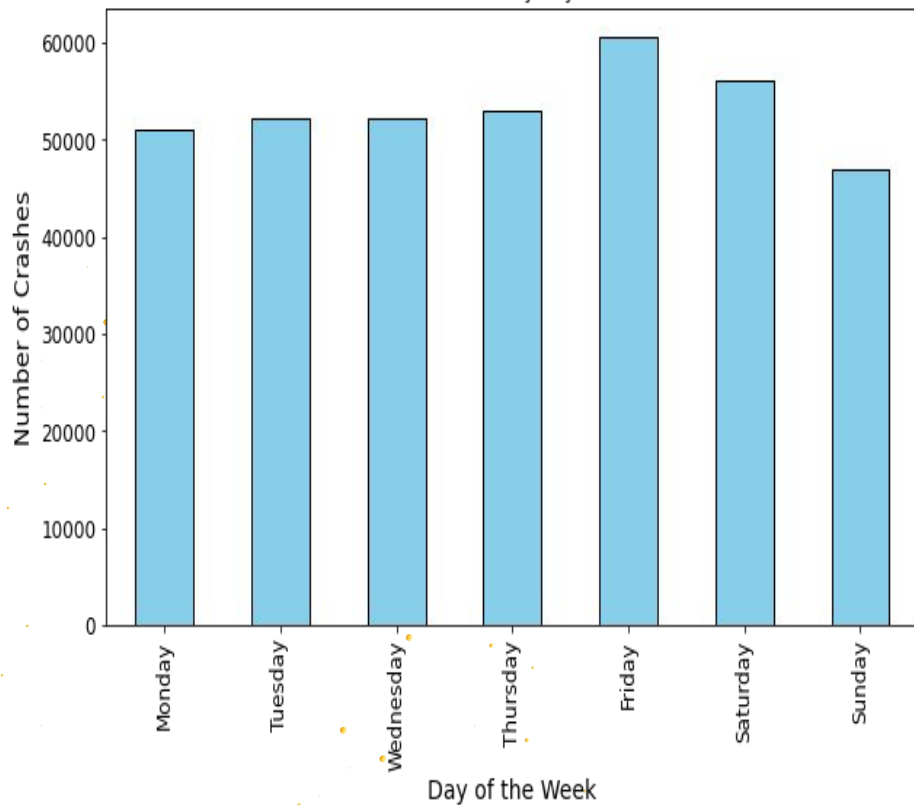


INSIGHTS GAINED DURING EDA

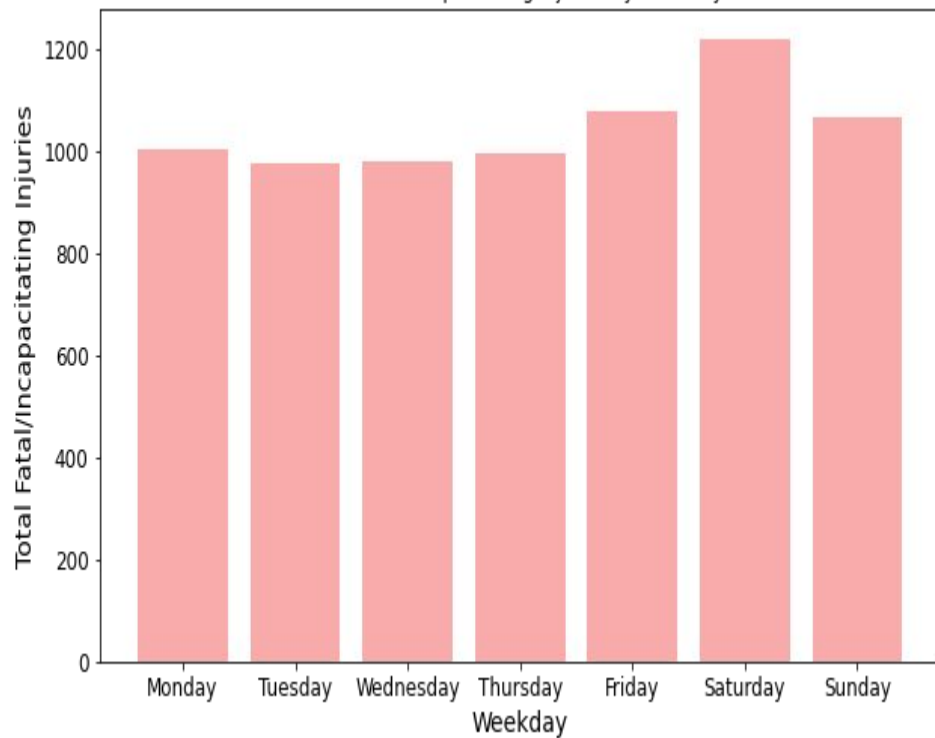


INSIGHTS GAINED DURING EDA

Number of Crashes by Day of the Week

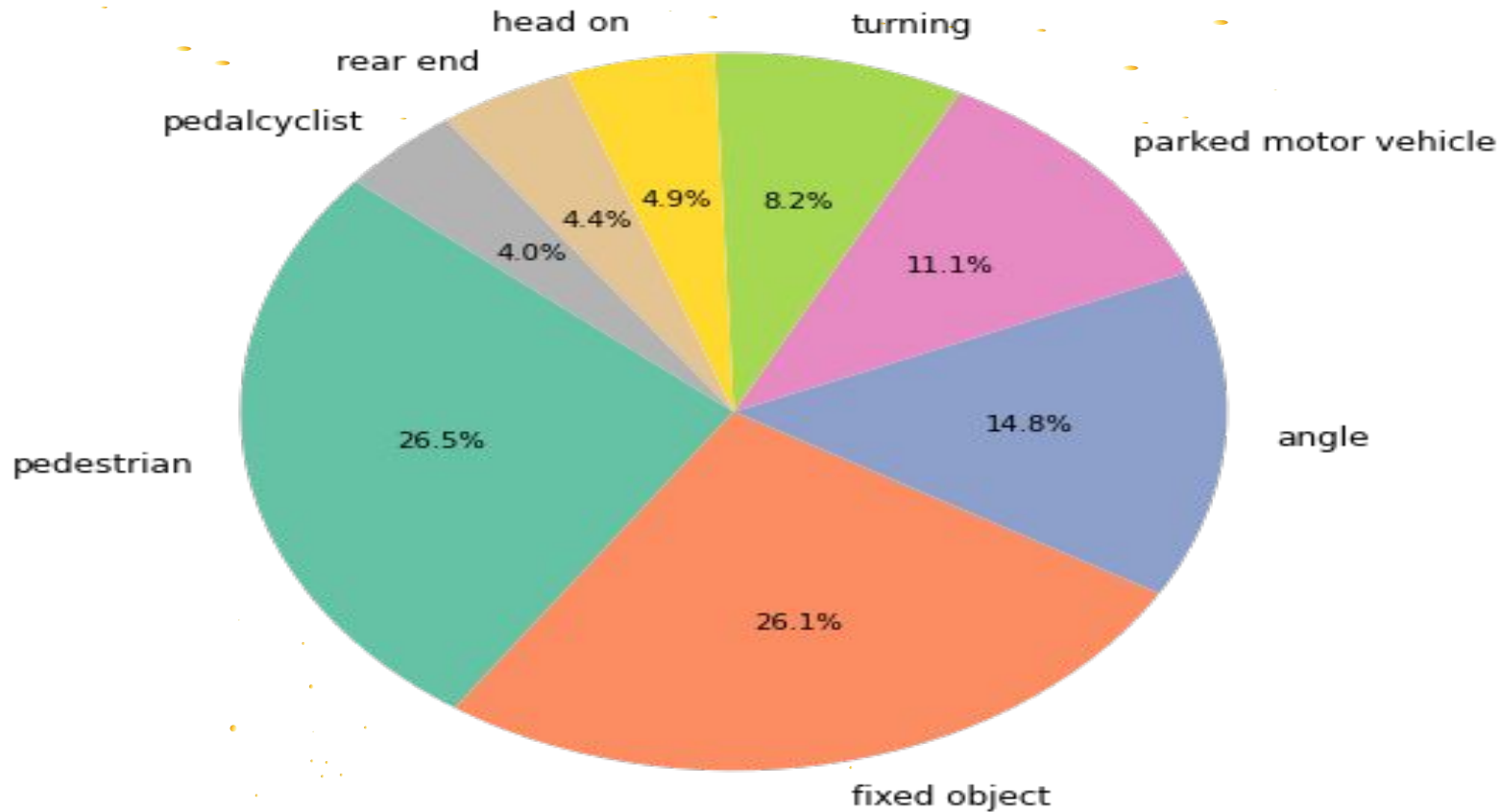


Fatal and Incapacitating Injuries by Weekday



INSIGHTS GAINED DURING EDA

Most Fatal Crash Types

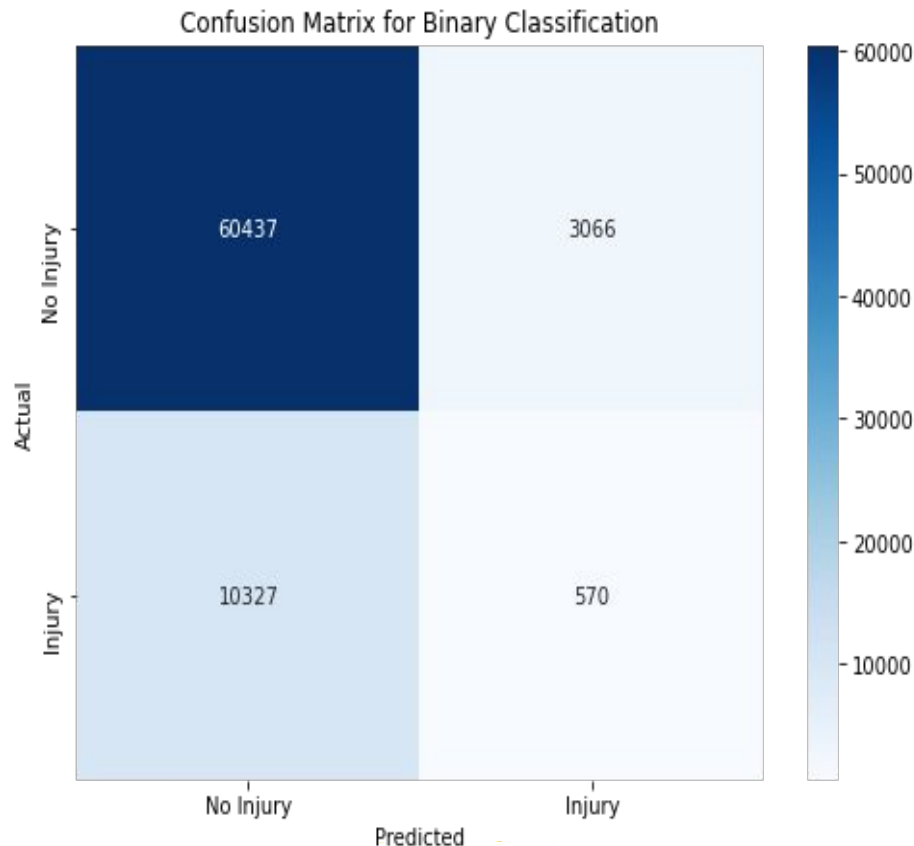


BASELINE MODEL KEY INSIGHTS

High Accuracy: Achieves 82% accuracy, largely due to correct classification of the majority class.

Class Imbalance: Significant imbalance with the majority class having 63,503 instances versus 10,897 for the minority class.

Poor Performance on Minority Class: Low precision (0.16) and recall (0.05) for severe injuries/fatalities, indicating many false negatives and positives.



NEXT STEP

TYPE II ERROR

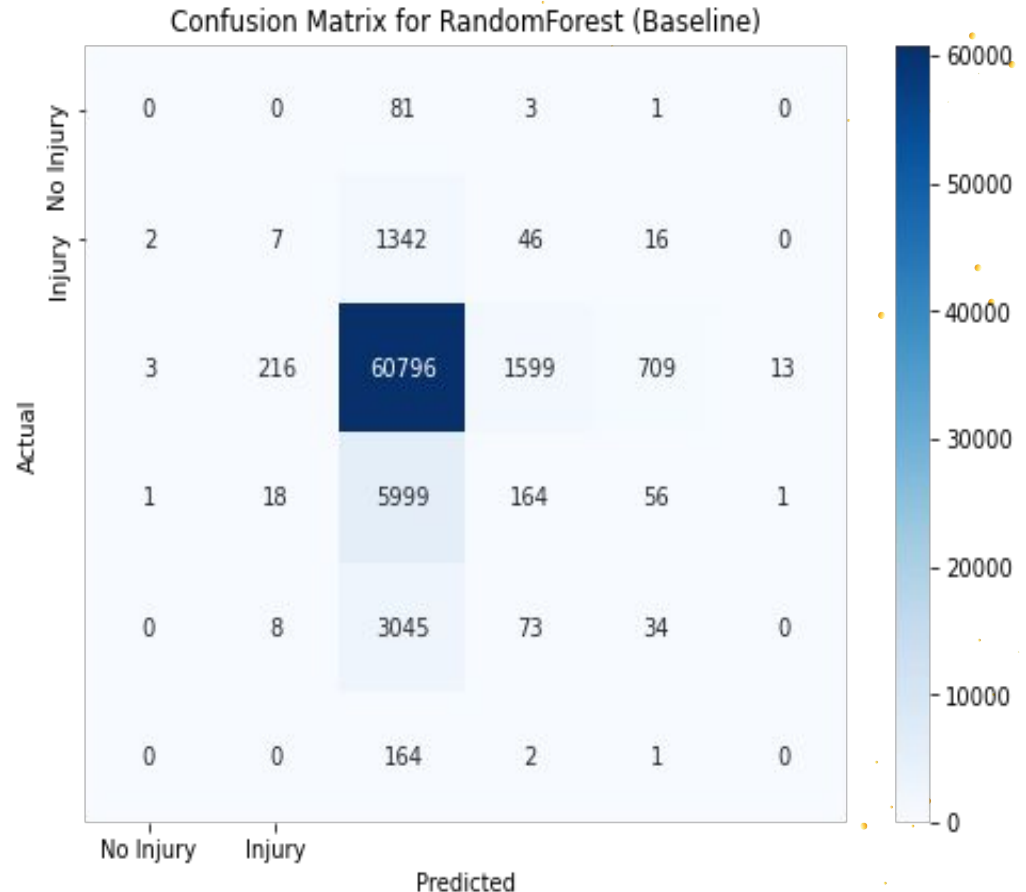
Our final model makes a lot of type 11 errors.

This means:

Predicted a less severe injury when actual class was severe

We will :

Explore methods such as SMOTE, class weighting, or undersampling the majority class and improve the Model's ability to identify injuries



THANK YOU

I look forward to hearing your thoughts and questions.

Let's connect on LinkedIn: [Phinidy George]