CHICAGO TRAFFIC ACCIDENTS

CONSULTANTS
Harrison Gu
and
Jojo de Leon

Business Case:

Chicago has been experiencing more accidents lately. With limited funds for ambulance services, the city has hired us to predict the conditions to be aware of so as to allocate their funds more efficiently when an accident occurs.

Project objective:

Our aim is to investigate the factors of reported accidents in the city of Chicago to ascertain which features are more likely to lead to an accident that involves injury.

Overview

THE DATA

Our Chicago Traffic Crashes datasets came from the city of Chicago's data portal. We combined two datasets, one that was referenced by the vehicle while the other regarded the person(s) involved in the accident. The records in the dataset date from 2015 to present day. The Crash Record ID which was present in both datasets was used to combine the two. Initially, there were over 1.6 million accidents to explore but our final dataset has roughly 162,000 entries with 17 unique features.

https://data.cityofchicago.org/Transportation/Traffic-Crashes-Vehicles/68nd-jvt3

https://data.cityofchicago.org/Transportation/Traffic-Crashes-People/u6pd-qa9d

Understanding the problem

Step 1

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

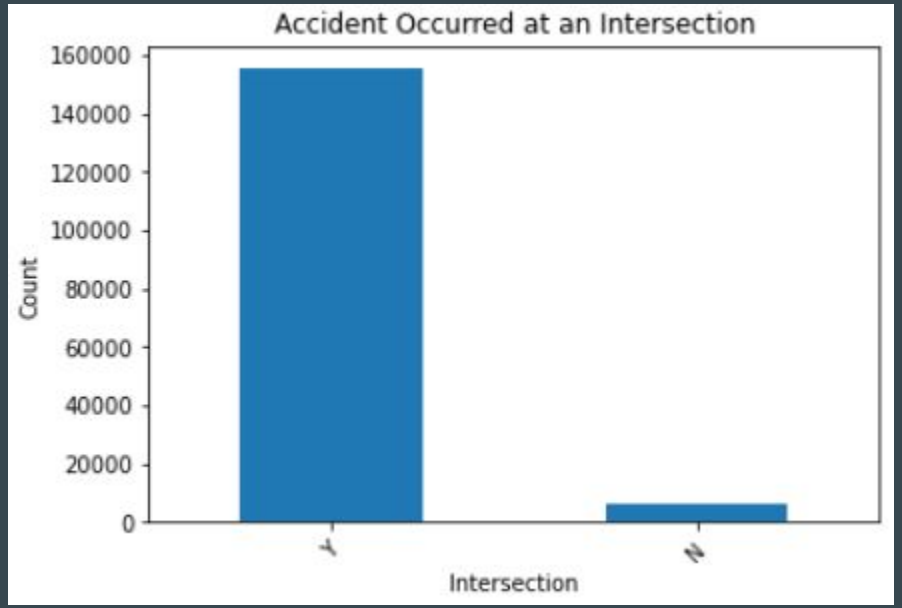
Step 2

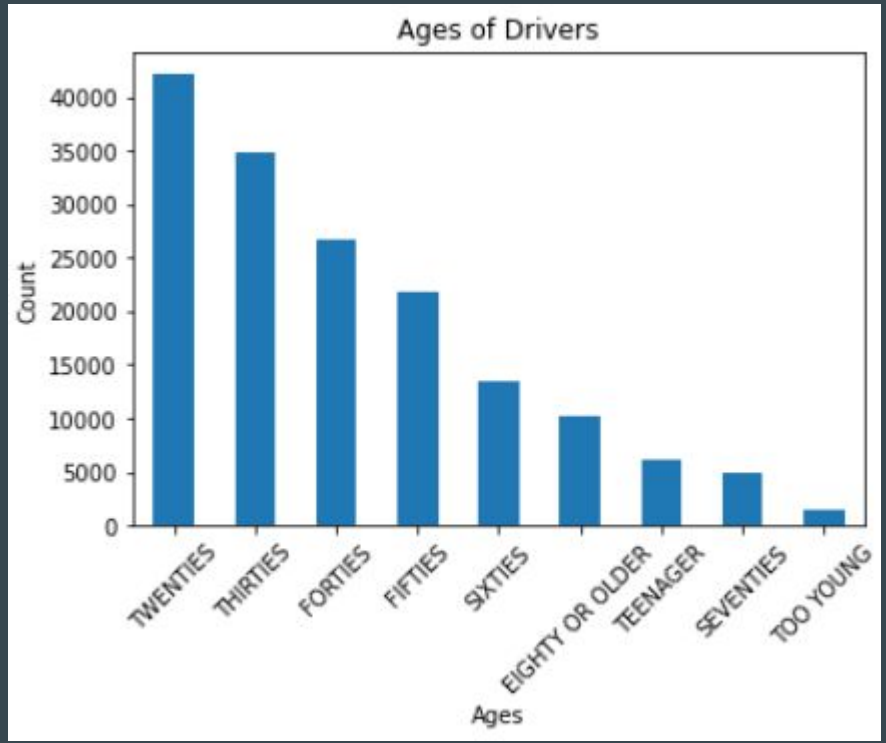
Run various models in order to select which classification models will be the best predictors for our dataset

Step 3

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

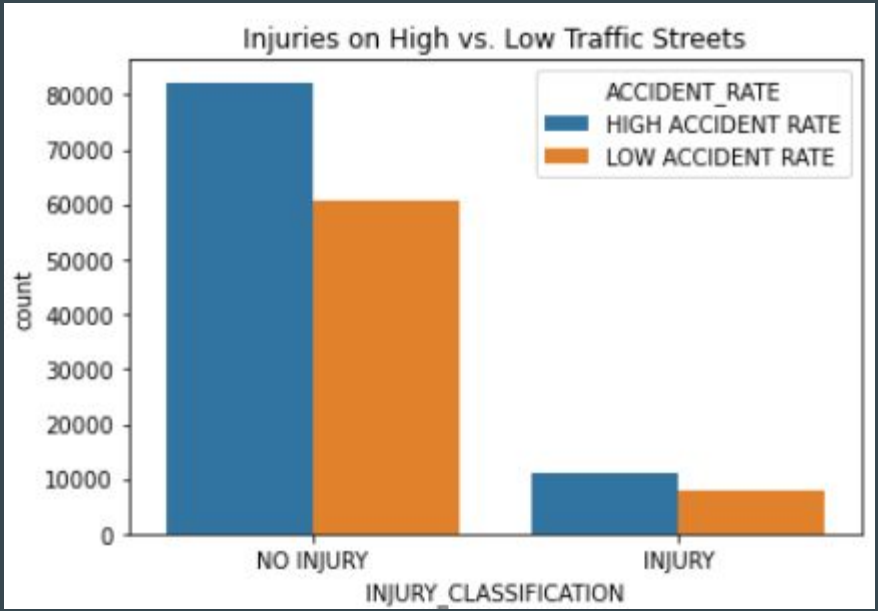
RELEVANT FEATURES Observations During EDA

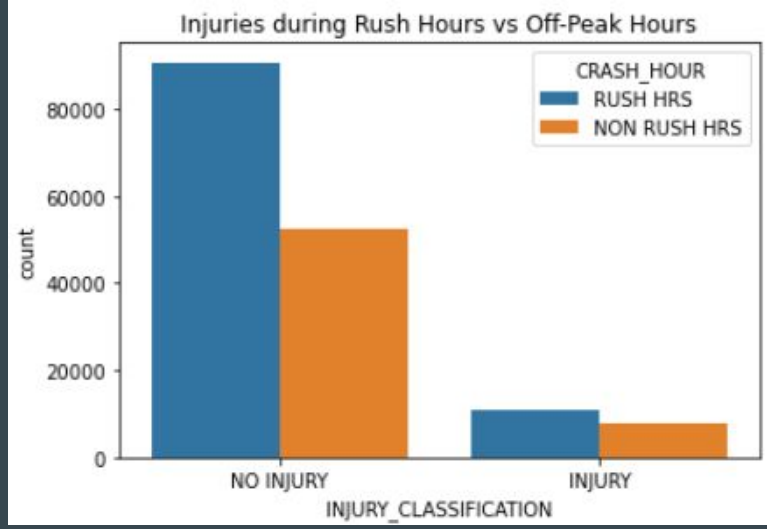


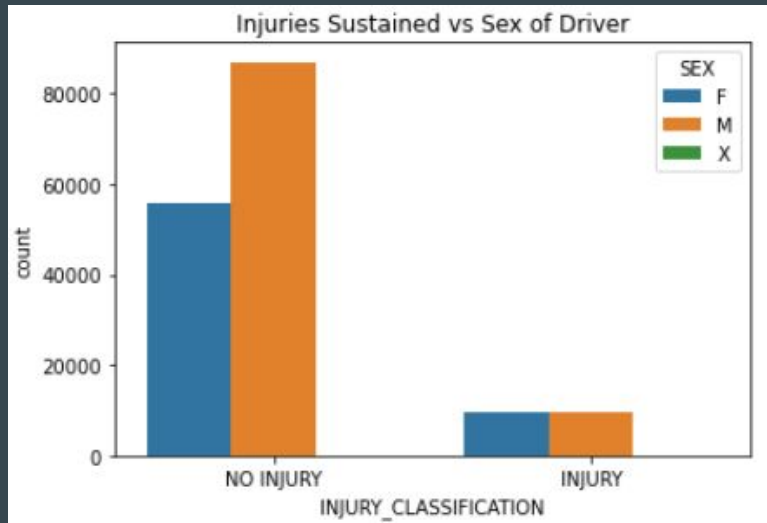


RELEVANT FEATURES

More Observations During EDA

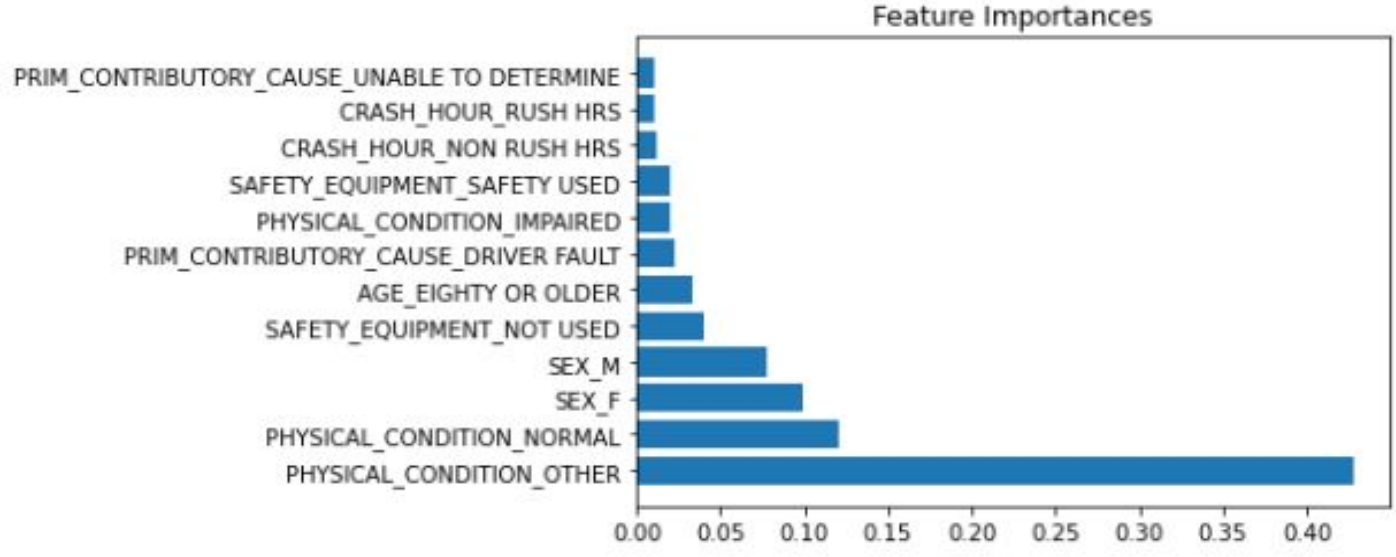






RELEVANT FEATURES

Observations from Baseline Models



Next Step

Type II errors:

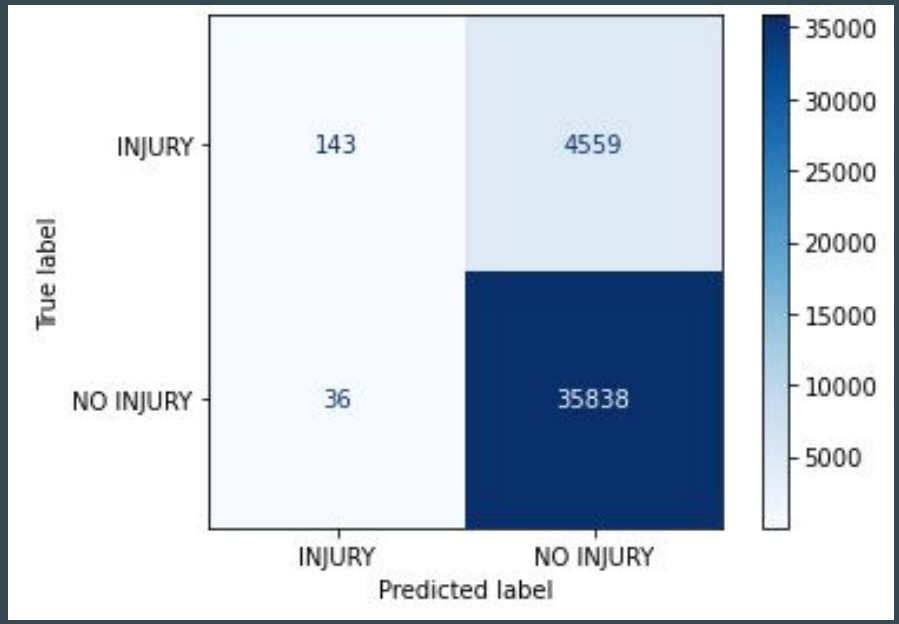
Our current models make a lot of type II errors

This means:

- Accidents that result in injury are often misclassified as no injury
- Some people who are injured will not get the help that they need

We will:

- Explore methods such as SMOTE while training our future models. This will make the classes more balanced, and improve the model's ability to identify injuries



THANK YOU

For your time and attention.

Project Repository: Github repo link

Authors today can be reached using the following information:

Harrison Gu: *harrison.s.gu@gmail.com*
Jojo de Leon: *trudell1977@gmail.com*