

CAR ACCIDENT SEVERITY PREDICTION PROJECT

Coursera Capstone

- ▶ Traffic accidents have a huge economic, social and health impact. Every year millions are spent and even though one can never avoid them 100% there are many factor that we can take in consideration to avoid participating in a car accident.
- ▶ The approach focuses on preventing potential unsafe road conditions from occurring in the first place. For the effective implementation of this approach, accident prediction and severity prediction are critical. If we can identify the patterns of how these serious accidents happen and the key factors, we might be able to implement well-informed actions and better allocate financial and human resources.

MOTIVATION

- ▶ The objective of this project is to evaluate the conditions that increase the risk in participating in car accident as well as predict the severity.
- ▶ For a given accident, without any detailed information about itself, like driver attributes or vehicle type, this model will be able to predict the likelihood of this accident being severe or not.
- ▶ This is a basic predictor model, however it has the potential to be developed further to perform real-time predictions

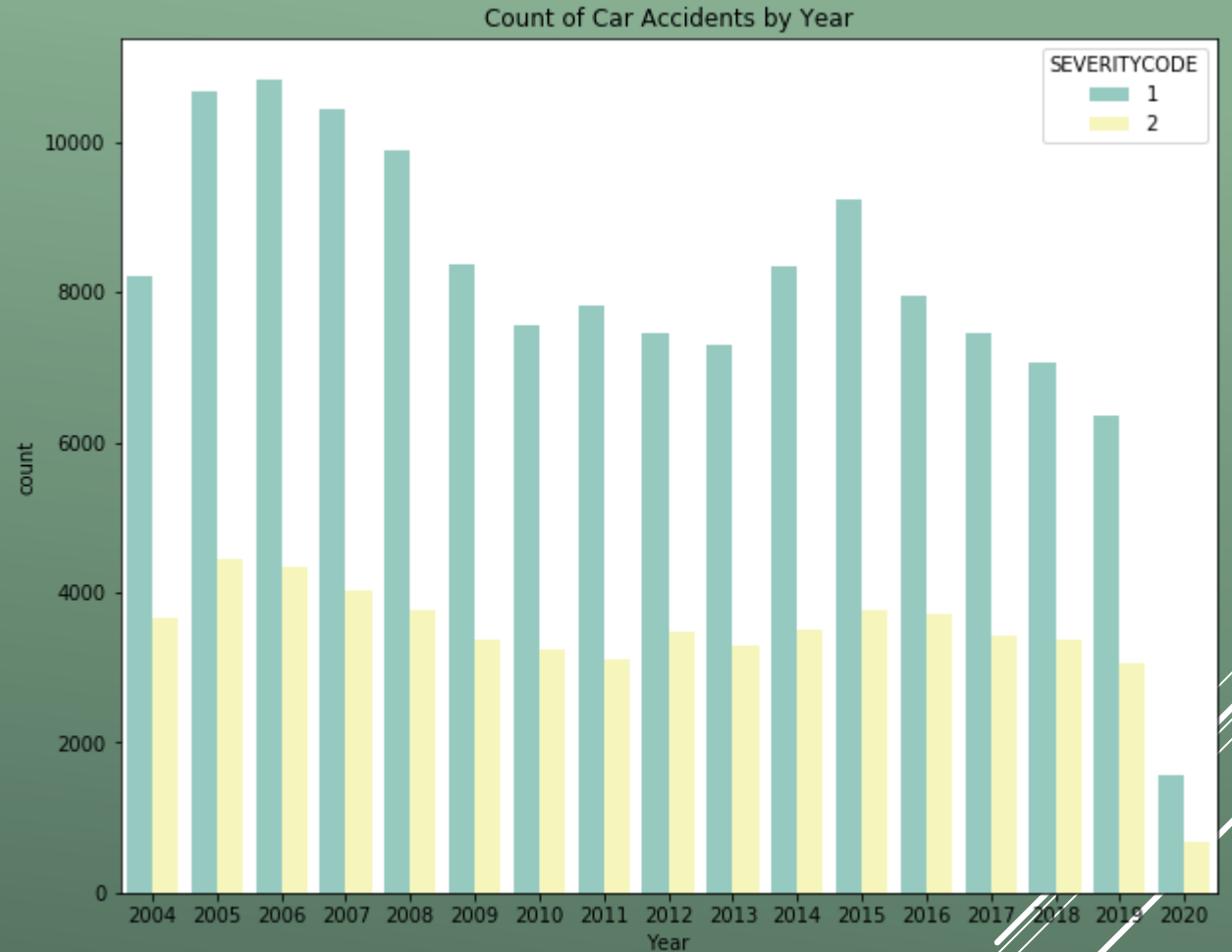
OBJECTIVE

- ▶ Data cleaning was the first step performed to detect and handle corrupt or missing records. EDA (Exploratory Data Analysis) and feature engineering were then performed on several features. Finally, Logistic regression, were used to develop the predictive model

PROCESS

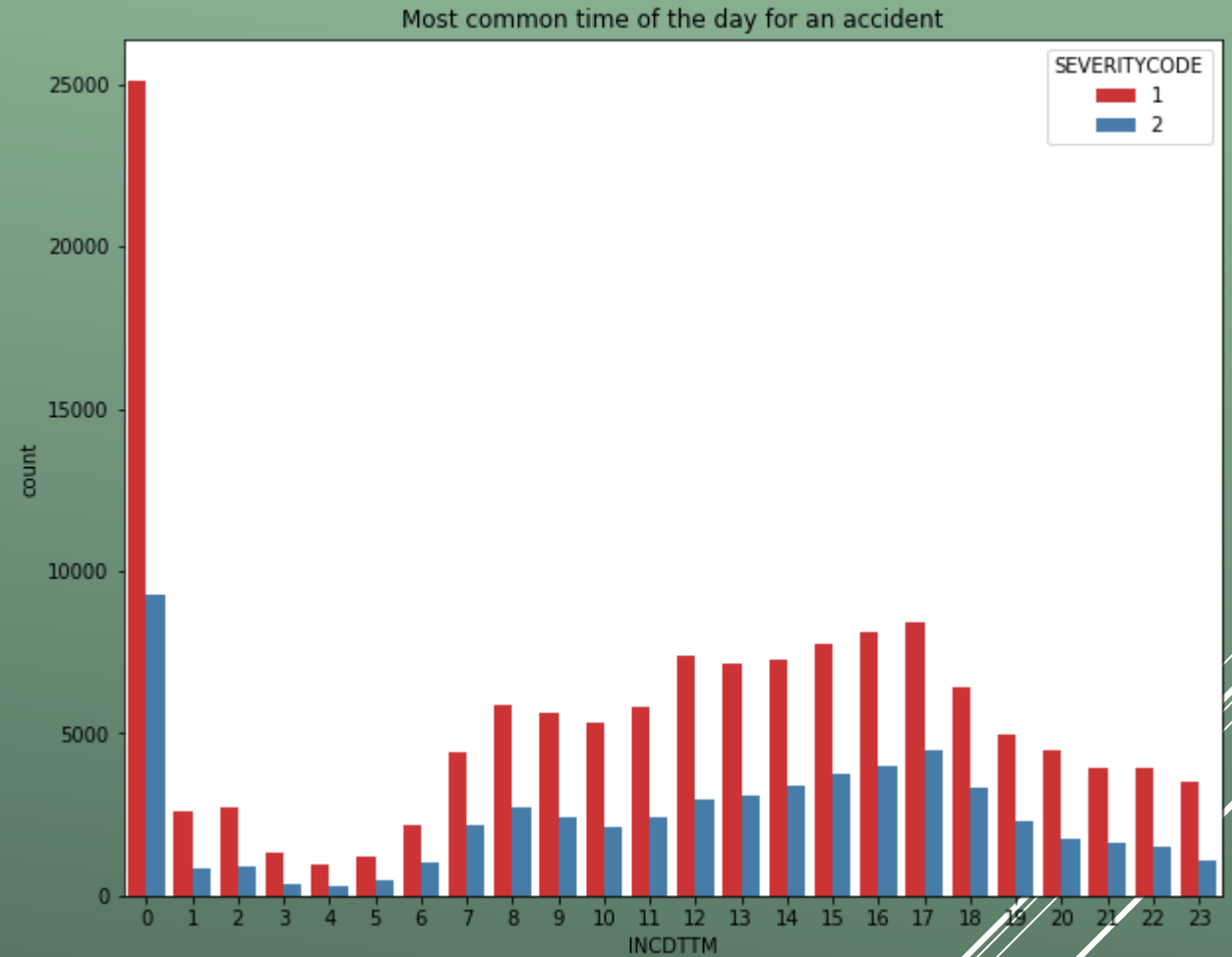
- ▶ Visualization that shows histogram on count of car accidents by severity since 2004. There is a small trend downwards over time, it would be interesting to predict how much decreases in whole 2020 year considering COVID19 lockdown. That will be done in future versions of this model.
- ▶ *Severity 1 = no injured passengers
- ▶ Severity2 = Injured passengers

DATA EXPLORATION



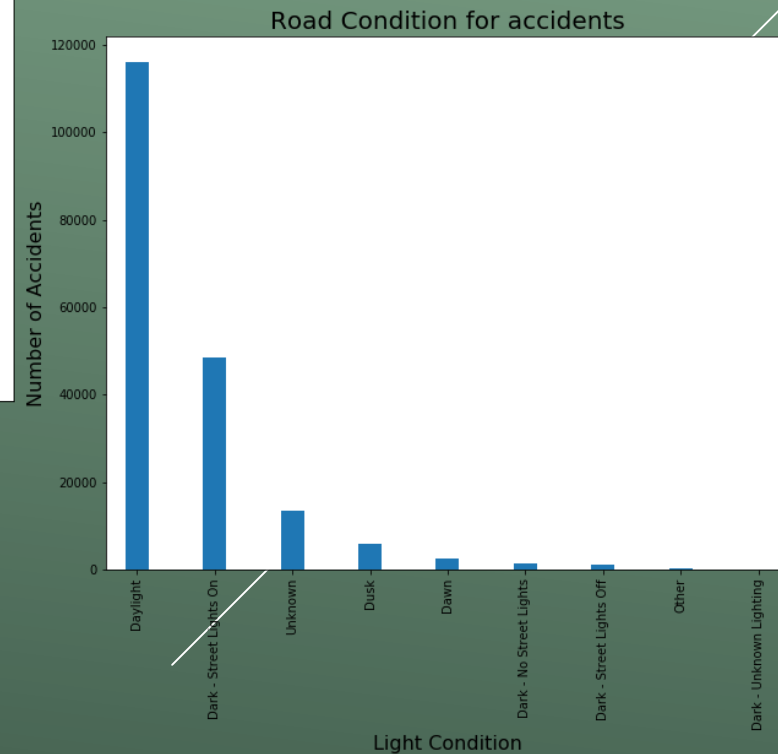
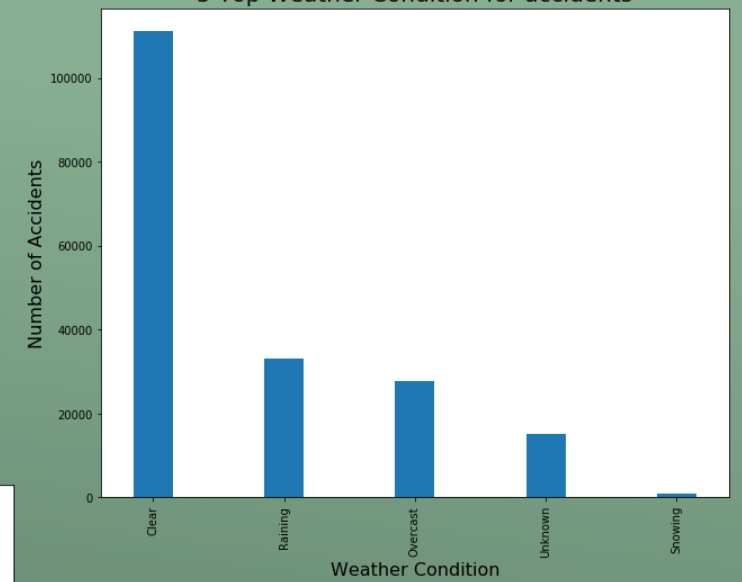
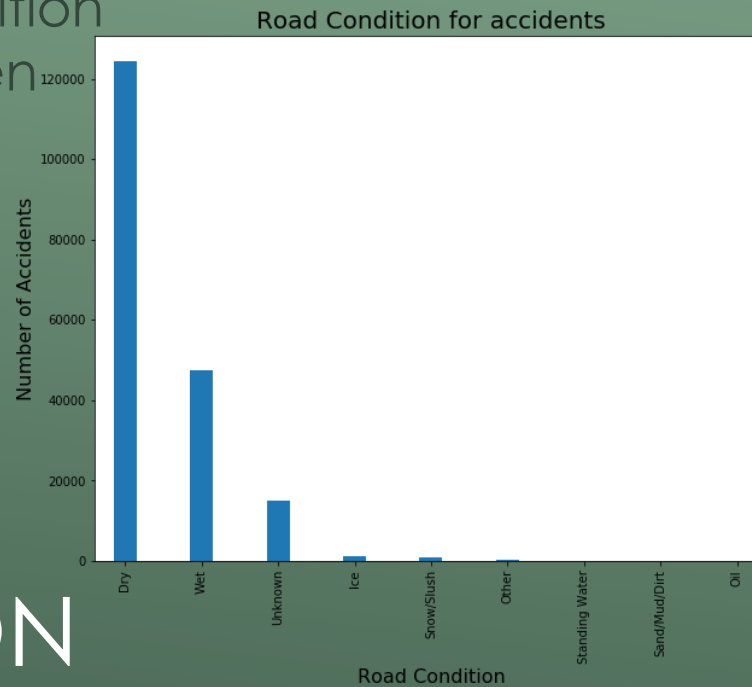
- In this visualization we can see the most common hour for an accident.

DATA EXPLORATION

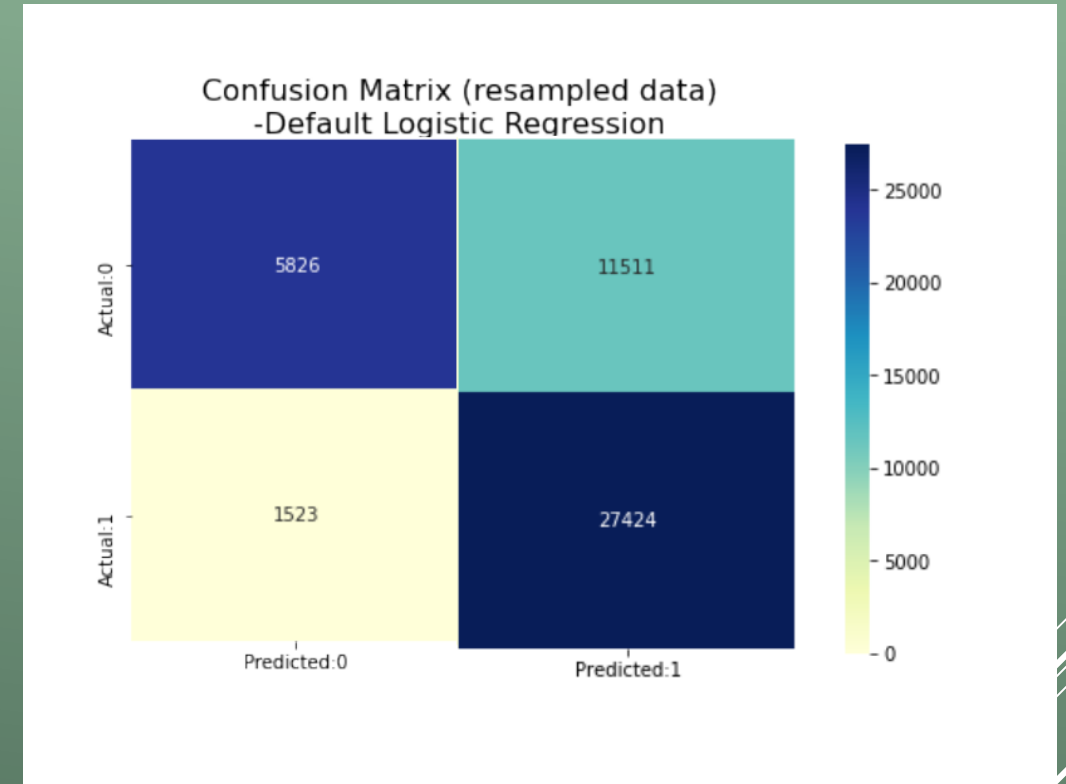


- I have plotted the top weather condition for accidents, as well as road condition and light condition to find any correlation between these conditions

DATA EXPLORATION



- ▶ Logistic regression was used as a baseline to perform binary classification
- ▶ Train Accuracy: 70 %
- ▶ Test Accuracy 70.4%



MODELING

- ▶ My initial assumption was that wet or foggy conditions were going to play a big role on this model, however i came to find that there was not much influence.
- ▶ Most accident happened at night with clear weather and dry conditions and at around 5pm.

KEY FINDINGS

- ▶ Since this is a basic prediction model, once it is developed at a much higher detail, I believe it can be used in a realtime database and connect to a maps API and show the best alternative routes

FUTURE WORK