

Report/Presentation of Coursera Capstone Project

Accident Severity Prediction

Business Problem Introduction

The problem is that there are too many car accidents. I would like to find out using data science what the main causes of the problem are, how likely is it for someone to get in a car accident and how severe it can be.

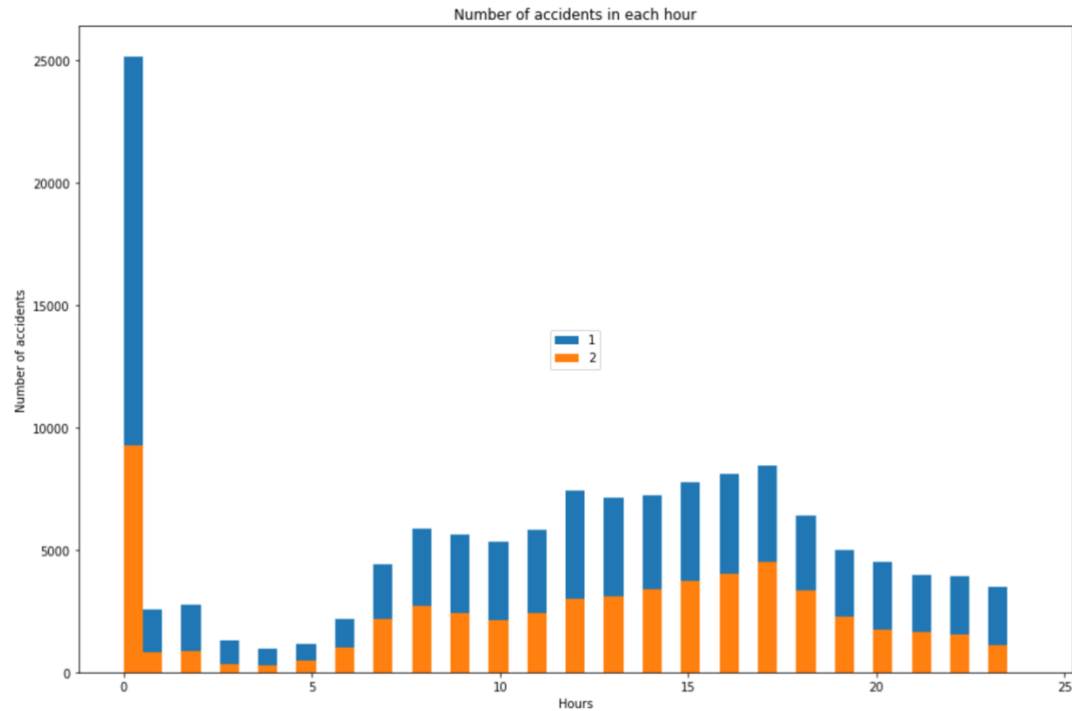
The analysis aims to help people that need to travel by car because they could estimate the probability of an accident based on factors such as weather condition, time of day, day of the week, season etc. Knowing the probability would help them decide if the travel is worth it or they can use other types of transit/choose not to travel.

The government agency responsible for the roads would also benefit because the analysis might help in preventing accidents by uncovering places where it is most likely that an accident will happen. With this knowledge it would be easier to implement safety measures to reduce/prevent accidents from happening or even just reduce their severity in the future

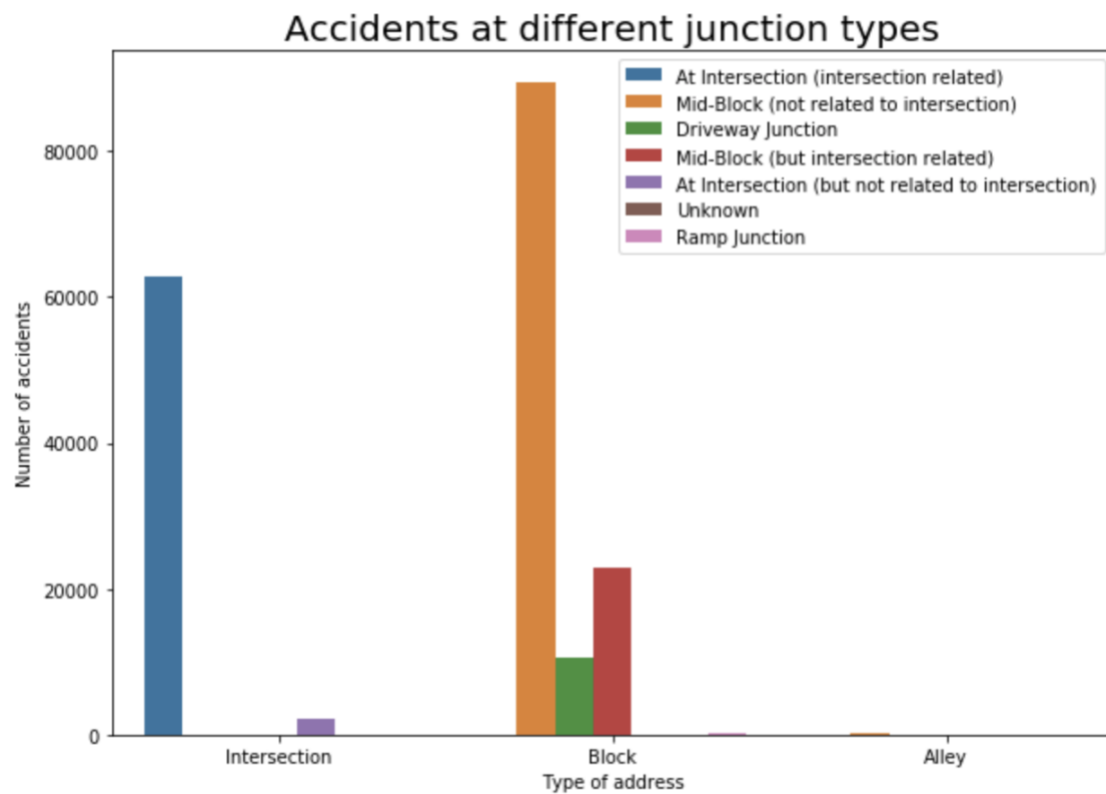
Data

I will be using the dataset provided in .csv format by the course's teachers. The data will be read into a dataset using python 3 on Jupyter notebook. Initial analysis will make clear which attributes of the dataset will be most useful in predicting accident severity. My initial guess is that the type/location of accidents, car speed, weather, time of day, day of the week attributes will be the most useful but I will

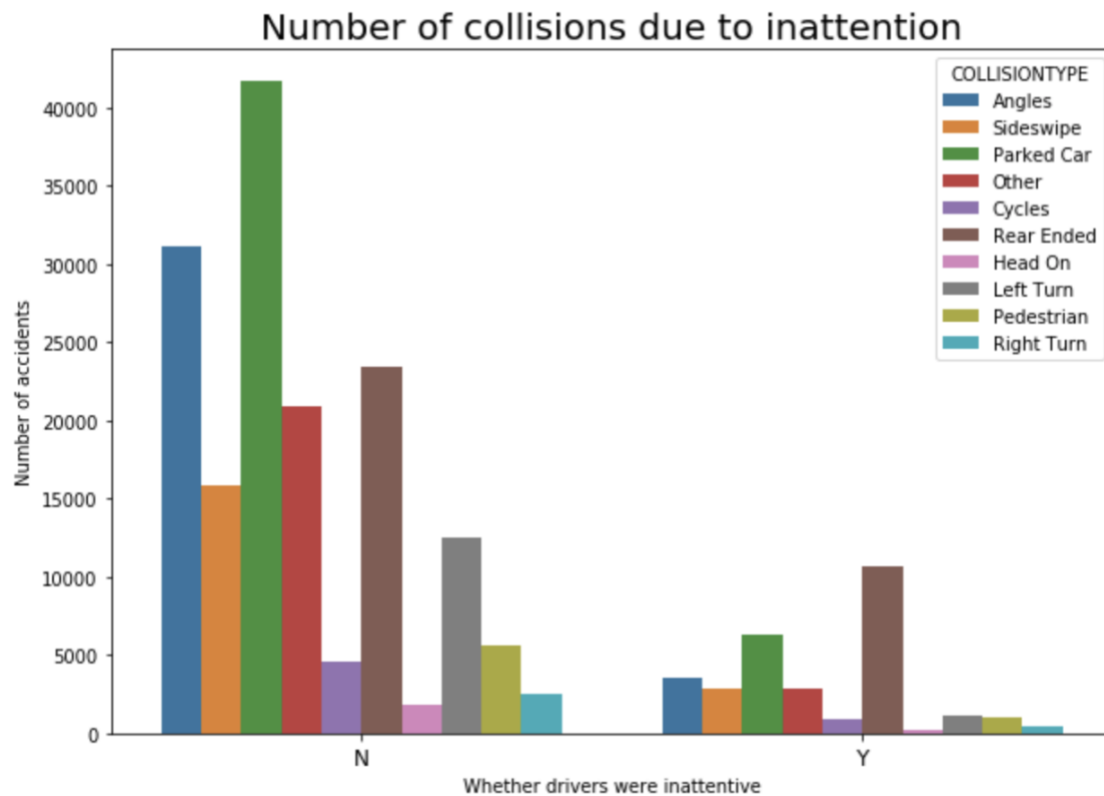
verify the correlation of each attribute by using regression functions.



This bar chart shows that most of the accidents happen at midnight.



Intersection and Mid-Block accident are the most common.



Accidents due to inattention vary and there is no major pattern to be seen.

Summary

The main observations discovered during data exploration:

1. Most accidents happen mid-block
2. The second most number of accidents happen at intersections
3. Virtually no accidents happen in alleys
4. When not attentive, drivers mostly drive into rear end of other cars
5. When attentive, drivers mostly drive into parked cars
6. Almost no pedestrians are hit when drivers are inattentive
7. Accidents are most likely to occur at 00:00 hour.
8. The time slots where it is the least likely that an accident happens is 01:00-06:00 and 20:00-23:00.

Conclusion

The aim of this project was to build a model that can predict the severity of an accident with the help of the given attributes. An accuracy of 75% was achieved which means that this model could be used for its intended purpose. However, there is always room to improve and with more data and work it could be made more precise. This model can be helpful for different stake holders such as the relevant authorities and drivers that want to evaluate their choices before sitting in the car.