INTRODUCTION

Road accidents over the years have led to the loss of numerous lives both locally and internationally. These traffic or road accidents aside claiming of lives has also led to the destruction of properties which in many cases can have negative impact on the economy of a country. Though, there has been numerous campaigns and advocacy to help reduce the number of accidents. However, the number continue to rise, hence, it is imperative to target the root cause of these accidents as this would help to largely reduce the number of the traffic accidents.

As our target is to decipher the factors that could lead to a possible road accidents, and define a pattern to them. We will be using the data provided by the IBM coursera platform to develop our project. The data was sourced from the Seattle police Department which presents collisions and the severity of the incident. The data records showed accidents dating from 2004 to 2020. The data also contained a number of information that can be used to determine the severity of road accidents and the causes of such. Some of these include weather, road condition, type of collision etc. With this information, our target is to find the exact factors that have led to loss of lives and with the outcome, such factors can be addressed. Hence, our target audience will be the government bodies who are in charge of these roads or individual who might want to ply some of these roads as to be aware of events that have happened in the past and drive with caution when travelling along such roads.

DESCRIPTION OF DATA TO BE USED

The data provided by coursera was collected from the Seattle police department showing the level of severity of accidents and factors that led to the accidents. The dataset contains about 194,674 numbers of accidents in the state of seattle from 2004 to present. The data also contains a number of columns which will be used to decipher the major factors of accidents in the area. The attribute I will be using for my analysis will be:

- Location (X and Y coordinate)
- Severity code
- Collision type
- Severity description
- Person count
- Vehicle count
- Junction type
- Weather condition
- Road condition
- Light condtion
- State coding designation

The outlined factors above will be correlated and from the result, we can ascertain the factors that mostly led to traffic accidents.

METHODOLOGY

I started with jupyter notebook where I conducted all of my analysis upon importing the required libraries such as pandas, numpy, seaborn and matplotlib.

After close examination of the data, it was discovered that most of it are categorical involving more than three values. Hence, I did graphical analysis to check the correlation between the features and the target variable which is 'severity code'.

After observing the data, only selected columns were chosen for our analysis as the others are irrelevant to the analysis. I went further to look closely at the missing data in each column and then dropped the cells that contain empty features. However, the 'Speeding' column having missing values was replaced with 'Y' signifying 'Yes'.

Upon completion of the data cleaning, each of the columns was visualized and the relationship between the columns and the severity code was observed

RESULTS AND DISCUSSION

Most of my analysis was done with matlotlib and seaborn libraries. I observed that the dataframe composed of 194673 rows and 38 columns. However, the number of columns was reduced to 11 bringing out the most important details that are related to accident. It was observed that most of the accidents occur at intersections. However, looking at factors that could normally cause accidents such as weather, light condition, and road condition, it was surprising to see that these account for only little percentage of the accidents in the area.

I went further to check the severity of the accidents using the variables provided. It was concluded that the severity of accidents is more in level 2. And these level 2 accidents occur majorly at the intersections.