**IBM Capstone Report**

**Introduction:**

Toronto is one of the largest and important cities in Canada. Compared to all other Canadian cities, Toronto has the largest contribution to Canada's GDP. It is also a major hub for immigration due to its vast array of options for employment and education. However, with such explosive growth comes a major issue. The problem of growing traffic. Even though Toronto is a well-planned city, the traffic seems to be worsening for this city. So this has prompted me to do an analysis of the traffic data of Toronto. This analysis could be of interest to the law enforcement of the city and the civic authorities. Using an analysis of this data they can determine in which areas they would need to upgrade the infrastructure or deploy more traffic police to better manage the traffic.

**Data:**

The most important part of any analytics project is the data. The data I'm going to use in this case is from the Toronto Police Service web portal(http://data.torontopolice.on.ca/datasets/ksi/data). Using this data I can find out the areas of the cities where most accidents take place and where the most fatal accidents take place. These areas can be brought into the focus of the concerned authorities who can take the required steps to deal with the problem. The data is in CSV format and has 57 columns and 12,244 rows. Many of these columns will need to be dropped as they are not relevant to the analysis. Then some preprocessing of the data will need to be done to make it suitable for analysis. This will be used along with location data from Foursquare to draw relevant insights.

**Methodology:**

I used IBM’s Watson Studio to create a Jupyter Notebook. I used the pandas library in Python to load the data into a data frame and then filter out the unnecessary data. This dataset contains a lot of information and in this analysis, I’m using THE following parameters: LATITUDE, LONGITUDE, ACCLASS , District and YEAR. ACCLASS refers to the class of the accident, whether the accident resulted in any fatalities or not. I then used Folium (another Pyth to make a map of Toronto and plotted the Fatal and Non-Fatal accidents on 2 different maps of Folium. The number on the marker represents how many accidents have taken place in a particular area. Zooming in can make the marker appear at the exact location of the accident. I then used the seaborn (based on matplotlib) library to make 2 bar plots showing the distribution of accidents by boroughs per year. This helps in visualizing the given data and helps draw insights into the traffic problem in Toronto.

**Results:**

The results can be seen in the ipynb file uploaded on my Github as it consists of Folium maps which are interactive HTML files.

**Discussion:**

From the map and the visualizations, we can draw a few insights. The first, Toronto and East York borough is an area very prone to accidents, both fatal and non-fatal ones. This could be due to the presence of an airport and other popular places in the area. The second thing we notice is even though Scarborough has fewer non-fatal accidents than the Toronto and East York area, it seems to more fatal accidents than the latter. This means that a large percentage of accidents that occur in the Scarborough area tend to be fatal. This could be due to a variety of reasons such as poor police presence or lack of good infrastructure.

**Conclusion:**

From the above insights, we can draw a few conclusions. Scarborough area needs to be brought to the focus of the police or civic authorities. Focusing on this area in the form of regular patrolling, stricter enforcement of traffic regulations and building better infrastructure should help bring down the fatalities in the borough.