BEM/Ec 150 Final Assignment

Dataset: NYPD Motor Vehicle Collisions
Author: Bianca Yang
Date: 6/8/2017

The purpose of this series of data analysis is to try to identify trends in motor vehicle collisions from July 2012 - June 2017 in New York City. The dataset has over 1 million instances and 29 features. The features includes: type of vehicle involved, contributing factor to accident, borough, and number of people injured. I have created three sets of time series graphs which show trends in total number of people injured by borough, trends in contributing factors to number of accidents, and trends in the type of motor vehicles causing accidents.

Below are some details explaining the insights from each set of graphs.

Graph 1: Number of people injured per borough per time

There is some cyclicality in the data, with peak number of people injured in the summer. This could be due to increased tourist traffic in the summer months. Tourists may not be familiar with traffic patterns in New York City and cause disruptions to normal traffic flow which confuse other drivers. This trend could also be explained by increased travel frequency over the summer. Families tend to travel more over the summer, since kids are out of school. This trend appears to be the strongest in Brooklyn.
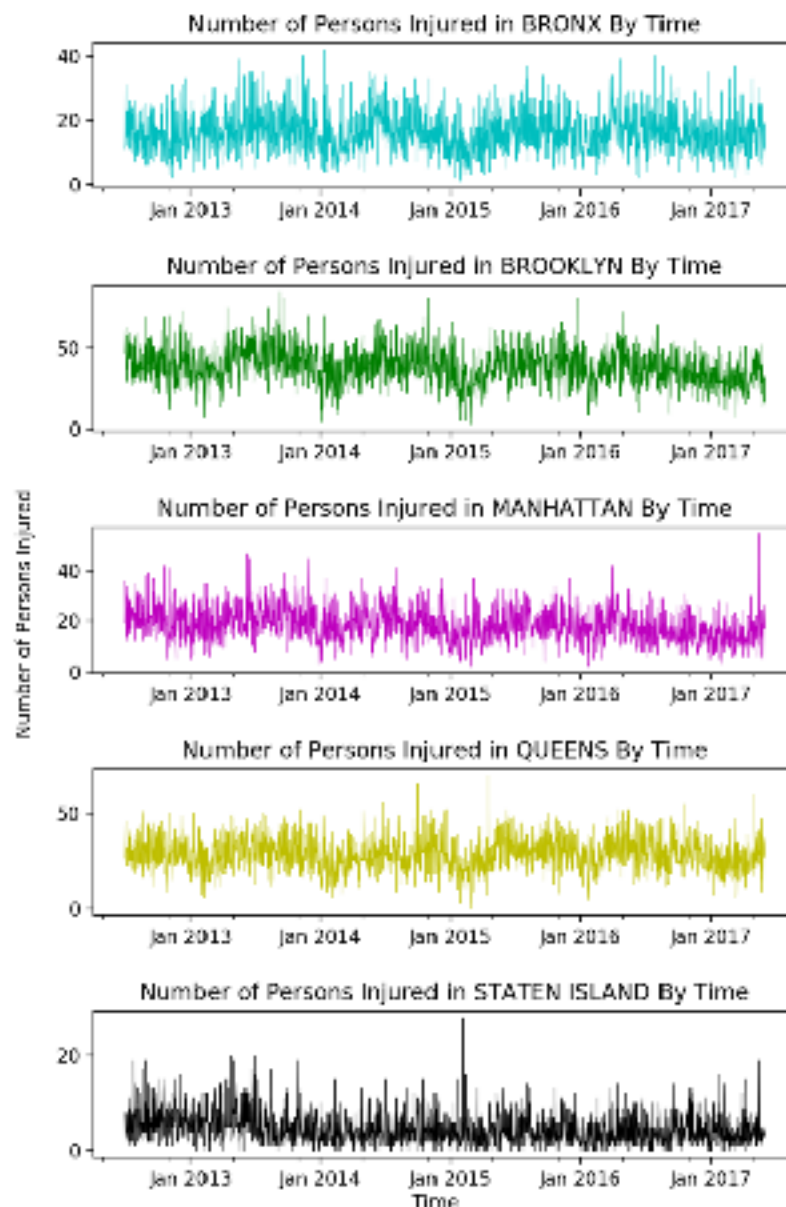
Total Number of Persons Injured Over Time
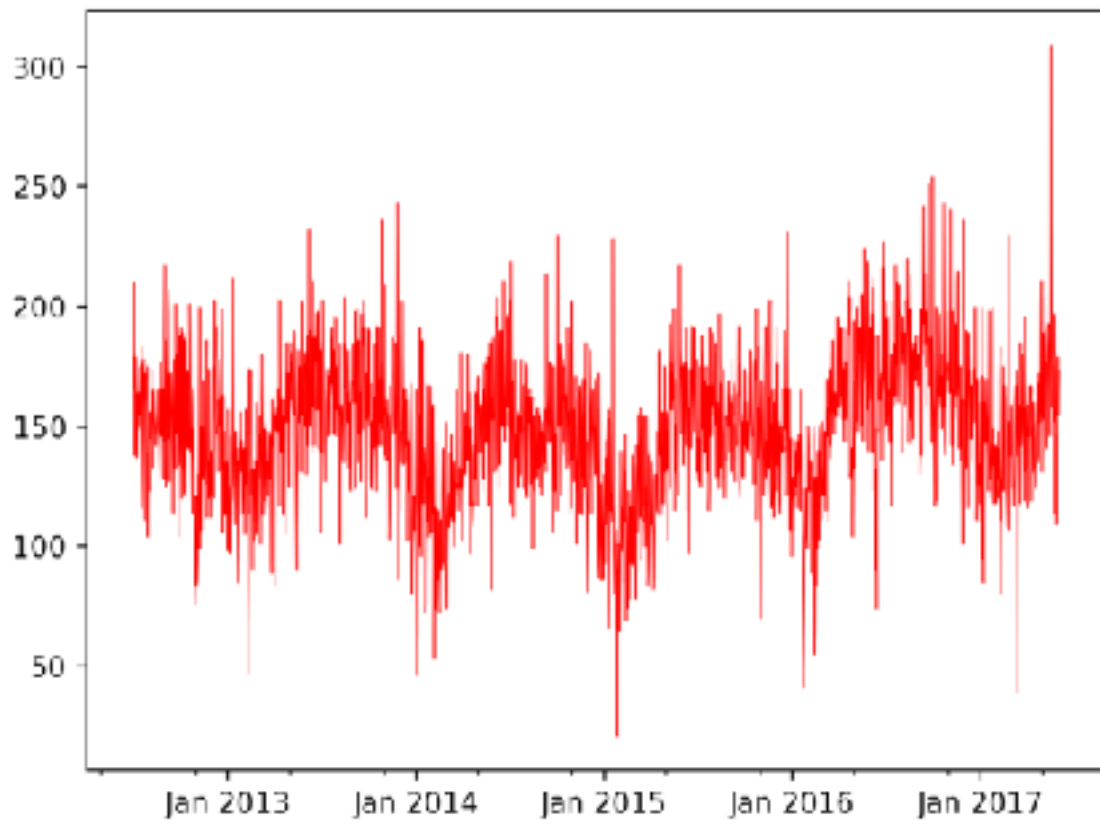
Graph 2: Contributing factors by time

The contributing factors are ranked at time 1. The top 5 factors' frequencies are plotted over the rest of the time series. There is an area in 2016 where the frequency drops to zero. This is because there is no information on contributing factors for that period. The top contributing factor for accidents is driver inattention or distraction. The second factor up until mid-2016 is driver fatigue. These are both behavioral changes that are hard to change, so perhaps the best is to hope that driverless cars or other transportation solutions can be made more convenient so people who are not at peak mental state will not need to drive.

The third contributing factor could be behavioral or a failure of the driver training system. One interpretation is that the driver was excessively aggressive. Another interpretation is that he wasn't aware who had the right of way. This factor could also co-occur with distraction and fatigue because a driver may not be paying sufficient attention to notice other cars.

The final two factors are likely failures of the driver evaluation and training system. To avoid these problems, the driver test could made more strict so drivers will develop better skill in backing up and turning.
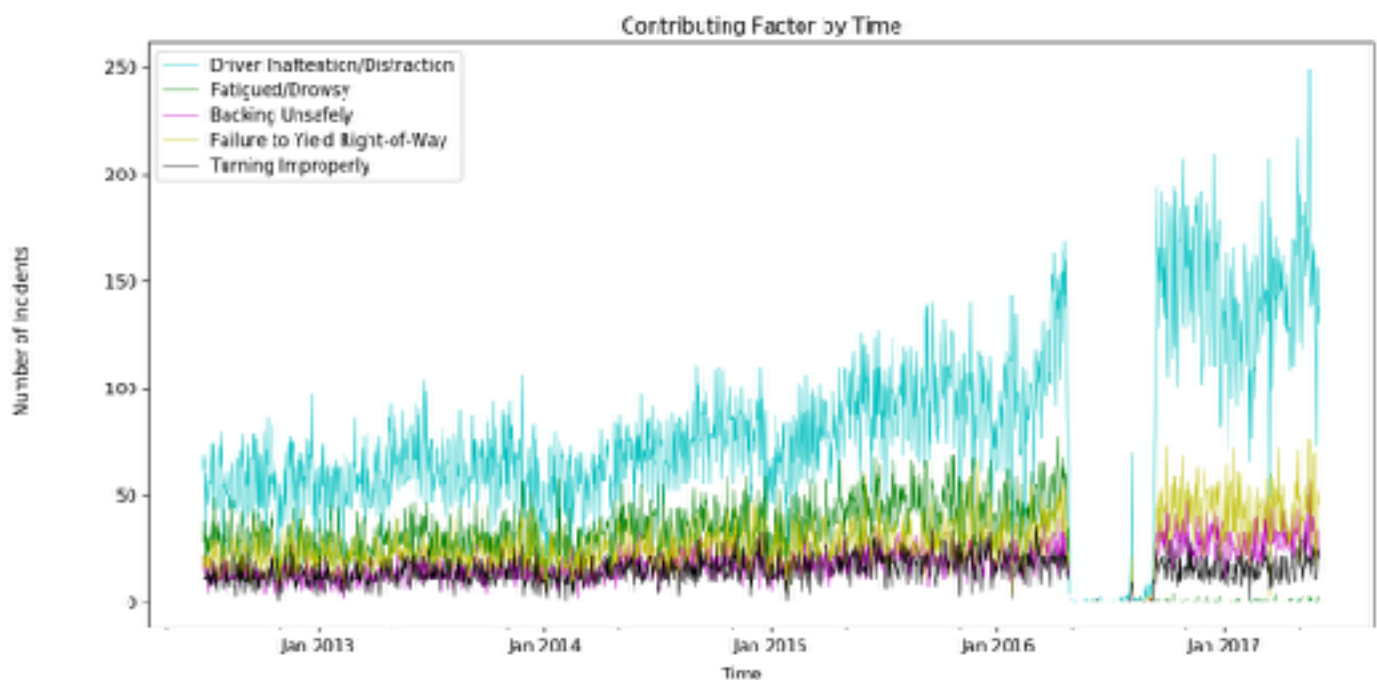
Graph 3: Incidents by vehicle type over time.

It is commonly assumed that SUVs are safer than sedans because they are able to absorb more force than SUVs. These graphs confirms that assumption. We can clearly see that passenger vehicles have been involved in the most incidents by time and have resulted in the most overall injuries. If one is going to travel through NYC by automobile, taxis are the safest option. This is like because taxi drivers have the most driving experience. This graph could also simply be a reflection of the relative proportions of cars in the city.
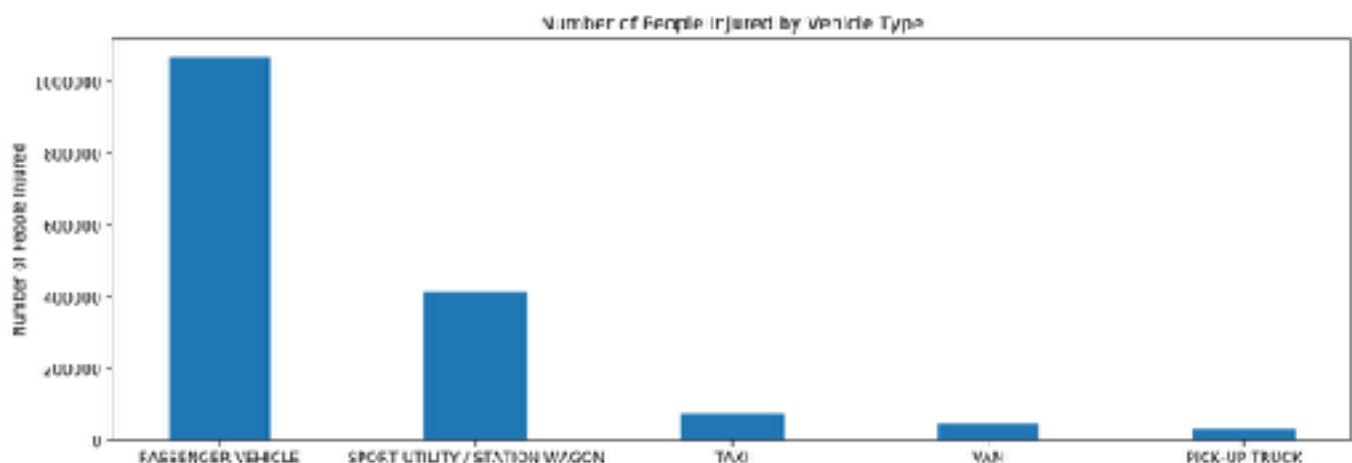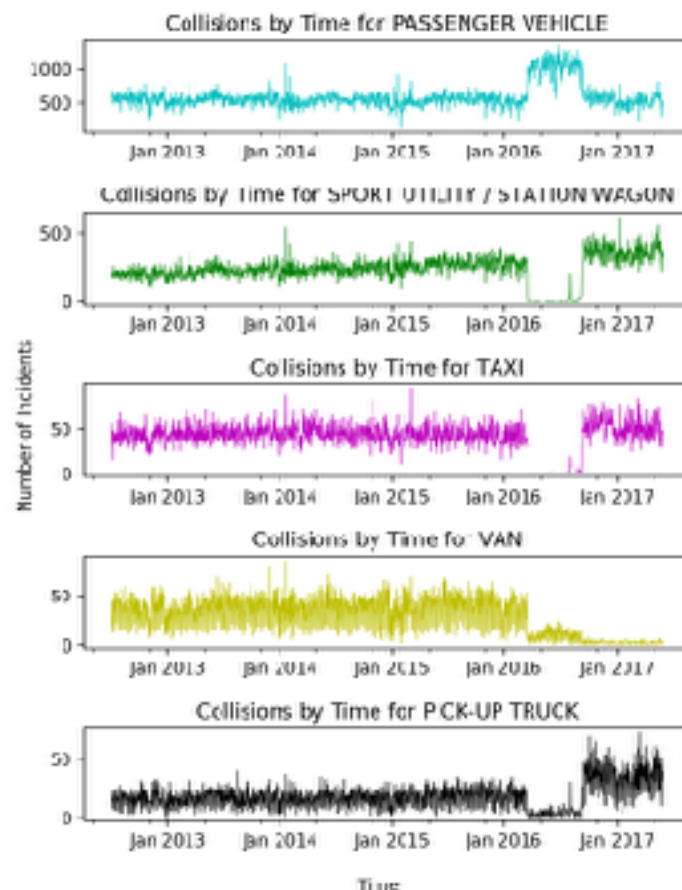
Conclusion:
These graphs help to confirm some intuition about collision patterns in New York. The major factors in automotive collisions are caused by human error. These factors cause significant number of incidents, which should encourage further research into streamlining and developing alternative transportation solutions.