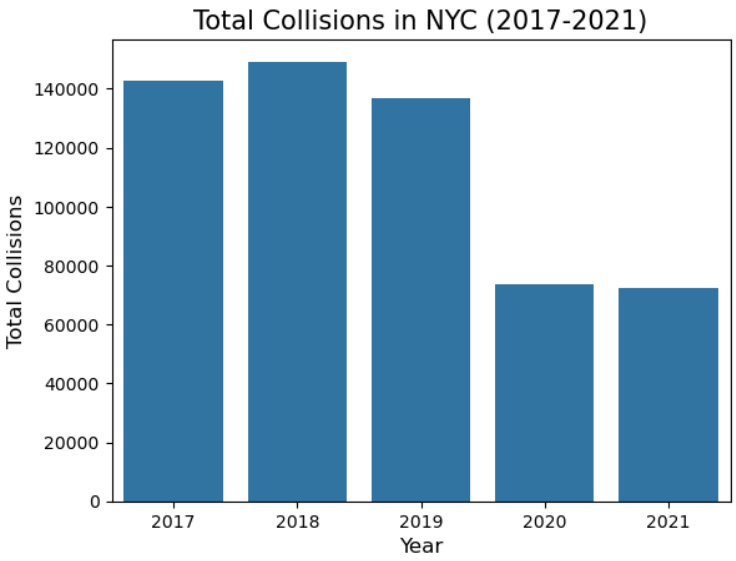
Elias Illescas

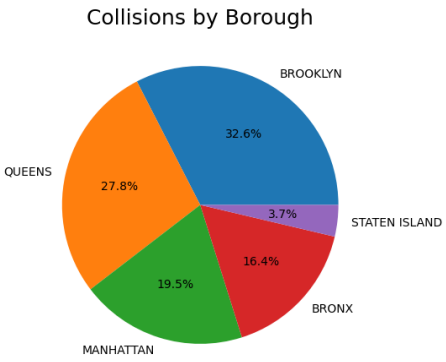
CUS 625

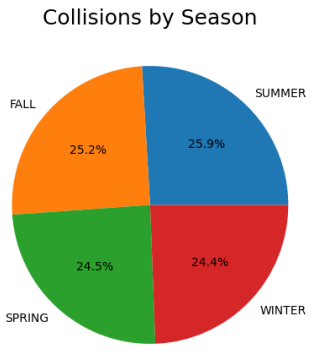
Final Project Plots

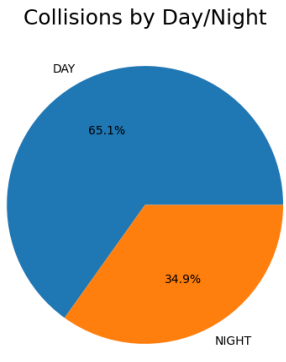


1. Bar Chart

I wanted to use a bar chart to visual the total amount of collisions that happened each year in the boroughs of New York City. I wanted to see if COVID-19 had any impact on motor vehicle collisions. The bar chart clearly shows a drop in motor vehicle collisions, possibly due to everyone being home during quarantine. Since the dataset only contains collisions, I grouped the data by year and just use the size (how often that year occurred in the dataset) of that year to create this plot. I needed to convert the 'CRASH DATE' attribute to date time data type, then I made a new column that only had the year.

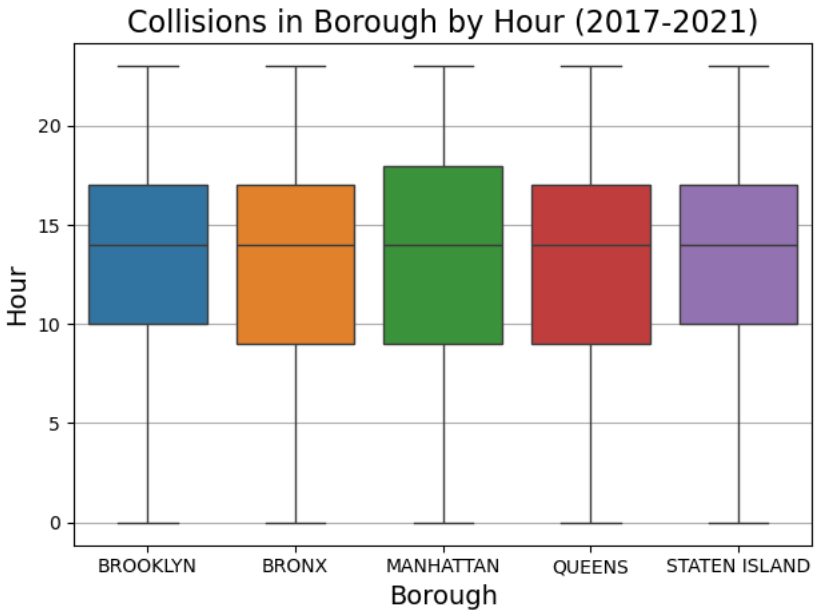






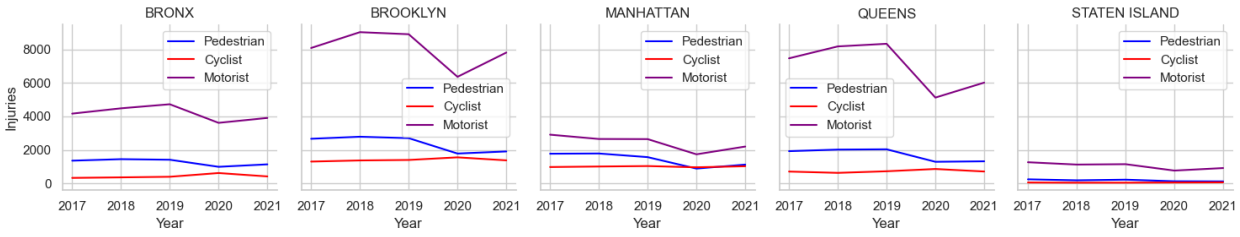
2. Pie Charts

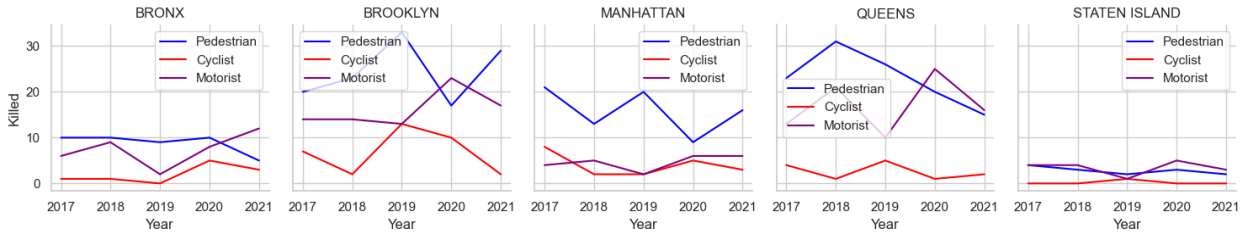
I wanted to see if there were any patterns of when and where a collision would occur. More specifically, I wanted to see if weather and/or visibility was a major factor. I created count variables that grouped the data by borough, season and day/night. I had to create a new column for season and day/night. For season, I classified the dates as followed: December - February: Winter, March - May: Spring, June - August: Summer, September - November: Fall. For day/night I distinguished day as anywhere between 6AM and 6PM, else it was classified as night. This gave me a more generalization of the season and time of day.



3. Box Plot

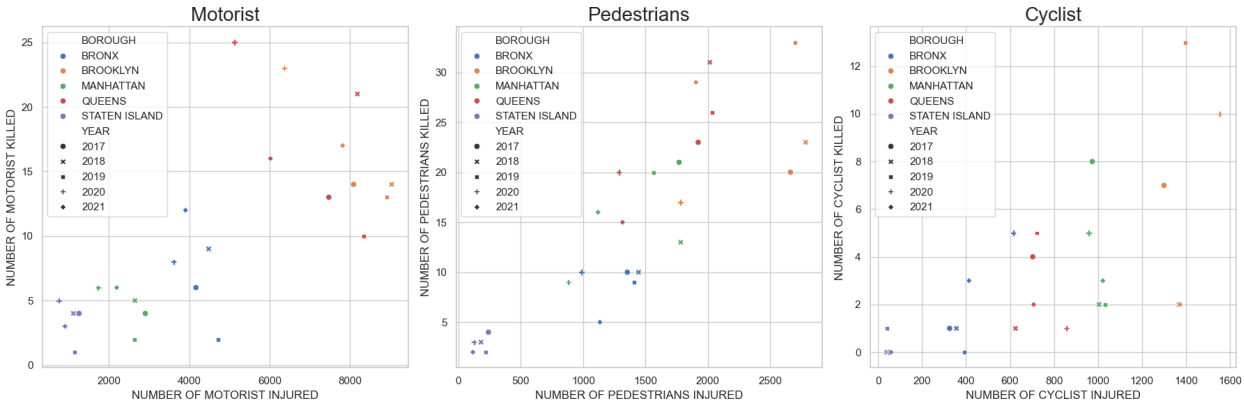
I created this box plot to allow me to see the distribution of motor collisions by the hour, by the borough. I had to create a new column that change the 'CRASH TIME' attribute to just the hour to simplify the plotting. It helped me get a general idea of what the median time was for collision occurrences.





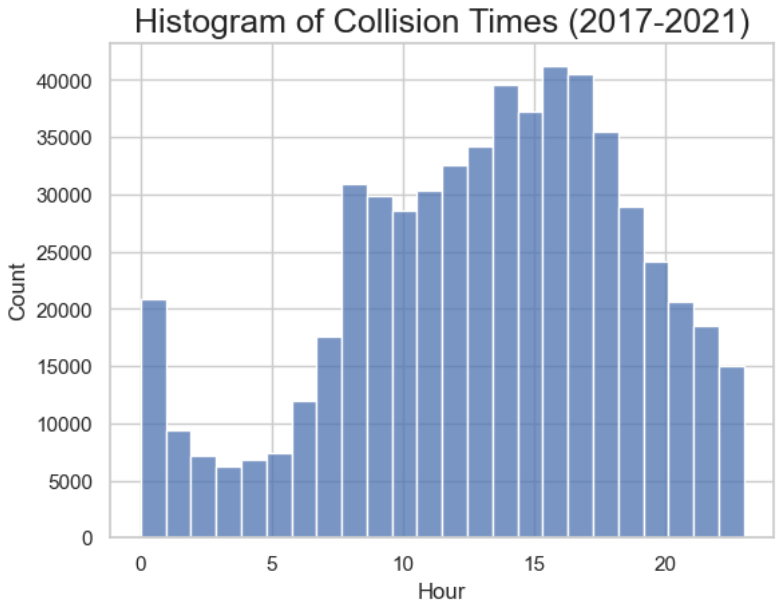
4. Line Chart (Facet Grid)

I wanted to see more details about the type of people who were being injured or killed during these collisions, and which borough is had the most frequent occurrences of injury or deaths. I used the facet grid to separate the columns by borough. Then I mapped each person type to a line plot, and color coded them to make it easy to differentiate between them. I made the injuries and killed plots separately because the numbers for those killed were significantly lower and would not have been useful if plotted together with the injuries.



5. Scatter Plot (Sub Plots)

To compare the total amounts of injured or killed by year, I created a scatterplot. I separated the type of person involved in the collision by using sub plots. I used colors to identify the borough, and shape to identify the year it took place.



6. Histogram

The histogram shows the frequency of occurrences for a given hour. I used the 'CRASH HOUR' attribute I created earlier to plot this. I used 24 bins to account for each hour of the day. This allows me to see what hours had the lowest and highest occurrences of motor collisions.