**NYC Motor Vehicle Collisions | September 2018 – August 2020**

The recent increase in motor collisions being reported in New York City was the catalyst for us selecting the Motor Vehicle Collisions – Crashes dataset from the NYC Open Data website. The source of this data is a compilation of reporting from the NYPD Officers who were on the scene of motor collisions where a person(s) were either injured and/or killed, or that has total damage worth more than one-thousand dollars. For the scope of our analysis we used data that was captured from 9/1/18 through 8/31/20. The data included: Dates, Times, Locations, Number of Persons Injured and Killed by Category, and Contributing Factors. Our goal is to raise awareness to this public safety issue by reporting where, when, and why these motor collisions are occurring. Furthermore, we would like to share this report with the Mayor’s Office in an effort to supplement the current ‘Vision Zero’ safety plan initiative which strives to eliminate all traffic deaths by 2024.

The data sources that were used for this analysis included the following: ‘Motor Vehicle Collisions – Crashes’ dataset from the NYC Open Data website; an article from the New York Times titled ‘More Pedestrians and Cyclists Are Dying in NYC Drivers are Often to Blame’; an article from the website [www.amny.com](http://www.amny.com) titled “Cyclists Deaths in NYC: A month-by-month Breakdown”; and a statistic from an article on the website, ny.curbed.com, titled ‘New York’s Traffic has Decreased, but More Drivers are Speeding’. After completing the data cleaning, there were a total of 237,381 rows and 16 columns remaining from the original dataset that was referenced in this study.

The platforms that were used for this analysis were Google Colab and Tableau. In Google Colab the data cleaning was done via coding. The cleaning process included: The removal of thirteen columns; the removal of null values within the ‘Borough’ column; renaming of one of the columns as ‘Crash\_Date’; and changing the format of the data within the ‘Crash\_Date’ column from ‘object’ to ‘date time’. Coding was also entered to display the following graph types in Google Colab: Bar graphs displaying the total number of Collisions, Number of Persons Injured, and Number of Persons Killed by Borough; Lastly, a line graph was used to display the number of collisions over the time that was chosen for the scope of this analysis, which ranges from 9/1/18 through 8/31/20.

After completing the cleaning process, the data was uploaded into Tableau. In Tableau a dashboard was created using an array of data visualizations tools for the Capstone presentation. The dashboard included the following stories: Introduction, Summary Citywide, ‘What Happened in April 2020?’, Contributing Factors, Injured and Killed by Category, ‘What’s with Brooklyn?’, and the Conclusion. The following techniques and tools were applied in Tableau for our presentation: Bar Graphs, Line Graphs, Dual Axis, Heat Maps and Tables.

Based on our analysis, the following findings were identified: A significant drop in the Aggregate Number of Collisions in April 2020; Contributing factors that lead to the highest total Number of Persons Injured and Killed; Outliers existing among each of the collision categories and the specific time factors for each; And lastly, Brooklyn being the outlier borough topping essentially every statistical category.

In April 2020, the aggregate Number of Collisions dropped to its lowest amount at 2,598. During this time, New York City was declared the epicenter of the COVID-19 pandemic. Subsequently, Governor Andrew Cuomo implemented and extended the stay-at-home order during the month of April. This aligns with our assumed causation of the drop which resulted in traffic dipping by thirty-five to fifty percent, as per the data reported by Sam Schwartz Engineering. Based on our reporting, with the exception of the ‘Unspecified’ category listed, the main contributing factor of all the aggregate Total Number of Collisions was Driver Inattention/Distraction, which accounted for twenty-five percent. We found this to be very interesting. The consensus in our group was that driving under the influence (DUI) would have been the most prominent cause.An analysis was also conducted correlating specific time factors that were associated with the Total Number of Collisions, Persons Injured, and Persons Killed by Month, Day of the Week, and Hour of the Day. Some of the notable findings from this analysis included: The highest number of cyclists were killed most usually in the month of September; The highest weekday for the number of motorists killed were Saturdays; And the most number of pedestrian fatalities occurred at Six o’clock in the evening. Lastly, of all of the boroughs Brooklyn ranked thehighest in Total Number of Collisions, Total Number of Persons Injured and Killed. An analysis was done by each zip code in the borough and we discovered that the neighborhood of Canarsie ranked the highest for the total Number of Persons killed and was the second highest for total Number of Persons Injured. The neighborhood of East New York ranked the highest for total Number of Persons Injured, however it ranked tenth overall for total Number of Persons Killed. We also found that there were a significant Number of Cyclists Killed in Brooklyn. A whopping sixty percent of all motorist collisions involving cyclists ended in fatalities. The assumptive cause for the increase in cyclist fatalities were due to the emerging popularity and rising number to bicycles on the road. This alternative mode of transportation is particularly preferred in trendy neighborhoods of Brooklyn like Williamsburg, where the demographic includes a younger age range of residents. This assumption is also consistent with the implementation of the bicycle sharing plan known as Citi Bike and the shared scooters plan via Revel, thus resulting in more bicycles and scooters on the streets of Brooklyn. The issue of the increase in cyclist fatalities prompted New York City’s Mayor Bill de Blasio to implement the “Green Wave” plan that will add new bike lanes, redesign intersections and step up enforcement of traffic laws. This plan was also rolled into the overall Vision Zero safety initiative.

Some of the limitations that were noted in our findings included the following: Inconsistency of data entry; Preliminary statuses of data reported; and the challenge of finding data to corroborate our findings with the level of granularity and specifics relative to Brooklyn being the outlier. Overall, we were able to navigate through these challenges and materialize a comprehensive study that can be used as a reference in supplementing Mayor de Blasio’s Vision Zero safety initiative.

**Work Cited**

Brown, Nicole, ed. “Cyclist Deaths in NYC: A Month-by-Month Breakdown - AmNewYork.” amNewYork. N.p., 24 Sept. 2019. Web. 17 Sept. 2020.

Data.cityofnewyork.us. 2020. *Motor Vehicle Collisions - Crashes*. [online] Available at: <https://data.cityofnewyork.us/Public-Safety/Motor-Vehicle-Collisions-Crashes/h9gi-nx95> [Accessed 17 September 2020].

Fitzsimmons, Emma G. “More Pedestrians and Cyclists Are Dying in N.Y.C. Drivers are Often to Blame”. New York Times, March 10, 2020,

Plitt, A., 2020. *New York’s Traffic Has Decreased, But More Drivers Are Speeding*. [online] Curbed NY. Available at: <https://ny.curbed.com/2020/4/7/21211911/new-york-streets-speeding-increase-coronavirus> [Accessed 17 September 2020].