**Deliverable 1: Project Topic**

Cheng Yuan Lee  
Cross Zeigler  
Ewin Hong  
Sagar D. Goswami  
Suchada Hapikul

George Mason University

AIT 614: Big Data Essentials

Dr. Liao

February 27, 2022

**Deliverable 1: Project Topic**

New York City (NYC) suffers heavily from traffic jams and congestion. Knowing about them in advance with accurate predictions can help plan trips ahead of time, thereby saving on time otherwise lost in traffic during the commute. Our project will utilize datasets from NYC open data and focus on traffic in the city to discover patterns of transportation to resolve congestion amongst other traffic issues within NYC. The project's primary goal is to provide valuable insights into the city's traffic system to improve transportation within NYC for increased satisfaction, safety, and decreased travel times. Some potential focus areas will be on specific points and regions within NYC, specifically those dealing with high volumes of traffic at specific times of day and giving valuable advice to certain audiences. Creating a machine learning model based on the previous NYC Average Traffic Speed Data sourced from the New York City Department of Transportation can help determine times/dates and locations that observe a relatively high amount of congestion. This analysis can benefit residents of New York City, emergency services, stakeholders in the urban transportation industry, and businesses in determining an ideal schedule for shifts based on the location, time of the day, day of the week, day of the year, etc. Transportation agencies can use this result to improve the efficiency of transportation operations. Moreover, travelers/passengers will be safer and more convenient because the risk of accidents and traffic problems can be reduced by using a system that can recognize traffic situations in advance.

The amount of traffic can vary due to many features like time of the day, day of the week (working day/holiday), seasonal holidays, annual events, major decennial events (pandemic, elections), etc. Therefore, creating a machine learning model based on the previous New York City (NYC) average Traffic Speed Data sourced from the NYC Department of Transportation can help determine times/dates and locations that observe a relatively high amount of congestion.

We will utilize several tools to complete the project. One of them will be Python or R to clean the data amongst any analyses we may run. Another tool may be used for this project is Power Query to explore the data and provide specific insights into traffic within NYC. We are currently debating on what NoSQL database and Big Data Engine to run the project on, but we're considering the ones given in class and will decide upon further research.

Works Cited

NYC OpenData. (2017, July 13). *Dot traffic speeds NBE: NYC open data*. DOT Traffic Speeds NBE | NYC Open Data. Retrieved February 28, 2022, from https://data.cityofnewyork.us/Transportation/DOT-Traffic-Speeds-NBE/i4gi-tjb9