**Members: Madison Melton**

**Repository:** <https://github.com/madisonbliss24/6100-semester-project>

**Data set Title: New York City Taxi and Limousine Commission (TLC) Trip Record Data**

**Link to data:** <https://registry.opendata.aws/nyc-tlc-trip-records-pds/>

**Objective and Research Questions:**

**This data set contains various data from trips taken by taxis and for-hire vehicles in New York City. The data they have collected goes back to 2009 which happens to be the year Uber was founded. I think it would be interesting to look for trends in the average distance traveled to the passenger’s destination. The data set documentation notes “The TLC publishes base trip record data as submitted by the bases, and we cannot guarantee or confirm their accuracy or completeness. Therefore, this may not represent the total amount of trips dispatched by all TLC-licensed bases.” In light of this, analyzing the number of trips may not provide an accurate story but we could still look at the overall trend. In order to condense the amount of data analyzed, I will just be looking at data from the yellow cab company**

Another possible objective is to note the increase in non-cash payment methods since 2009. The payment methods are labeled as a number, so I’m hoping I’ll be able to find documentation on their website or GitHub that explains the labeling system. Another consideration is looking at how traffic congestion changed throughout the pandemic since they have data through July of 2021. There’s so much to analyze in this dataset, so I’m hoping to narrow my objectives as the project progresses. I will come back to update this document after cleaning and processing the data more in preparation for Deliverable 2.