For my part, I was responsible for the data ingestion portion of the project. With James, I focused on an efficient way of taking in the data and storing it. We decided on Apache Hive/HDFS platform for the Data Warehouse/Storage and Apache Spark for data processing. We both worked in parallel to find creative ways to process the data and came together when we found new discoveries. A lot of the data we dealt with was from a Database that contains data from Ubers, Taxis, and Lyfts. Once this data was ingested we sent it off to Gaofeng and Yann to have k-means clustering and the Page Rank algorithms ran on it. I also focused on the challenges portion of the project report.