This readme if for the analytics portion of our project. We conducted our analytic with scala, using the data sets found in ana\_code/input files, run entirely on NYU’s Dataproc. To run our code, add both of the datasets as well as main.scala to HDFS, open a spark shell, and load the main.scala in the shell. (See screenshots/mittal\_screenshots/Spark Code Running/ IMG\_1744\_Compressed.mov for an example of our code running).

Ultimately, the output is written to ana\_code/FINAL OUTPUT/part-00000.