Since github has limit max 100 MB. Following are files that need to run on this homework :

- Fhv_tripdata_2019-01.csv
- Spark.ipynb
- Output file : output.json and output.parquet