Load data from AWS RDS to Hadoop

<Command to run the python file>

export SPARK_KAFKA_VERSION=0.10
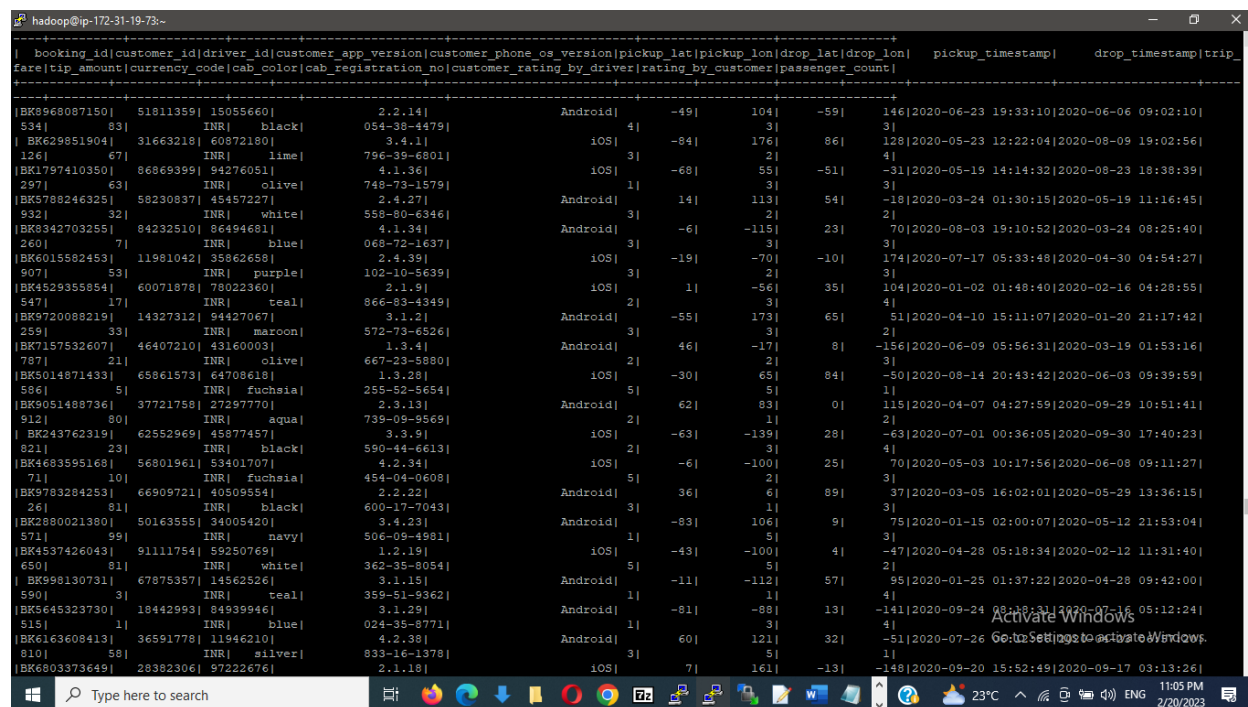spark-submit --packages org.apache.spark:spark-sql-kafka-0-10_2.11:2.4.5
datewise_bookings_aggregates_spark.py

<Command to move the csv file to HDFS>
hadoop fs -put /home/hadoop/datewise_booking /user/capstone/carride2/aggregated1/

<Screenshot of the file in HDFS>

```
23/02/20 17:31:21 INFO DAGScheduler: Job 3 finished: showStri
+-------------------+--------------+
|        pickup_date|total_bookings|
+-------------------+--------------+
|2020-03-31 00:00:00|             6|
|2020-08-19 00:00:00|             2|
|2020-03-15 00:00:00|             3|
|2020-08-23 00:00:00|             3|
|2020-02-07 00:00:00|             3|
|2020-06-14 00:00:00|             4|
|2020-02-20 00:00:00|             3|
|2020-02-12 00:00:00|             3|
|2020-04-06 00:00:00|             1|
|2020-09-10 00:00:00|             3|
|2020-06-22 00:00:00|             2|
|2020-02-06 00:00:00|             3|
|2020-05-20 00:00:00|             3|
|2020-09-18 00:00:00|             2|
|2020-08-18 00:00:00|             5|
|2020-02-24 00:00:00|             6|
|2020-02-23 00:00:00|             1|
|2020-10-22 00:00:00|             3|
|2020-10-16 00:00:00|             4|
|2020-04-29 00:00:00|             5|
+-------------------+--------------+
only showing top 20 rows
```