

## Create aggregates for finding date-wise total bookings

### <Command to run the python file>

1. `sudo -i`
2. `vi datewise_bookings_aggregates_spark.py`
3. `spark2-submit --jars "spark-sql-kafka-0-10_2.11-2.3.0.jar" datewise_bookings_aggregates_spark.py`

### <Command to move the csv file to HDFS>

`hdfs dfs -put home/datewise_aggregation /user/root/clickstream_data_flatten`

### <Screenshot of the file in HDFS>

```
[hdfs@ip-10-0-0-218 ~]$ hadoop fs -ls /user/root/datewise_aggregation
Found 2 items
-rw-r--r--  3 root supergroup      0 2021-03-12 07:19 /user/root/datewise_aggregation/_SUCCESS
-rw-r--r--  3 root supergroup 3769 2021-03-12 07:19 /user/root/datewise_aggregation/part-00000-fd22c33b-ef88-49a5-a2cc-721bd83e426c-c000.csv
[hdfs@ip-10-0-0-218 ~]$ hadoop fs -cat /user/root/datewise_aggregation/part-00000-fd22c33b-ef88-49a5-a2cc-721bd83e426c-c000.csv | wc -l
290
[hdfs@ip-10-0-0-218 ~]$
```