

## Create Aggregate file for number of bookings daywise

### <Command to run the python file>

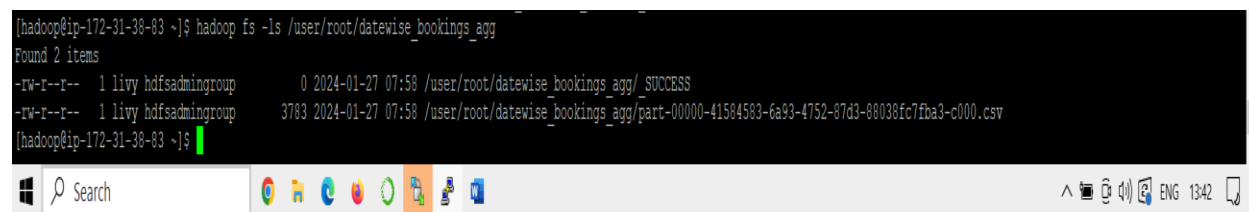
```
export SPARK_KAFKA_VERSION=0.10
spark-submit --packages org.apache.spark:spark-sql-kafka-0-10_2.12:3.1.2
datewise_bookings_aggregates_spark.py
```

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

```
agg_df.coalesce(1).write.format('csv').mode('overwrite').save('/user/root/datewise_bookings_ag
g',header='true')
```

## we could place it any place as needed

### <Screenshot of the file in HDFS>



```
[hadoop@ip-172-31-38-83 ~]$ hadoop fs -ls /user/root/datewise_bookings_agg
Found 2 items
-rw-r--r--  1 livy hdfsadmingroup      0 2024-01-27 07:58 /user/root/datewise_bookings_agg/_SUCCESS
-rw-r--r--  1 livy hdfsadmingroup 3783 2024-01-27 07:58 /user/root/datewise_bookings_agg/part-00000-41584583-6a93-4752-87d3-88038fc7fba3-c000.csv
[hadoop@ip-172-31-38-83 ~]$
```