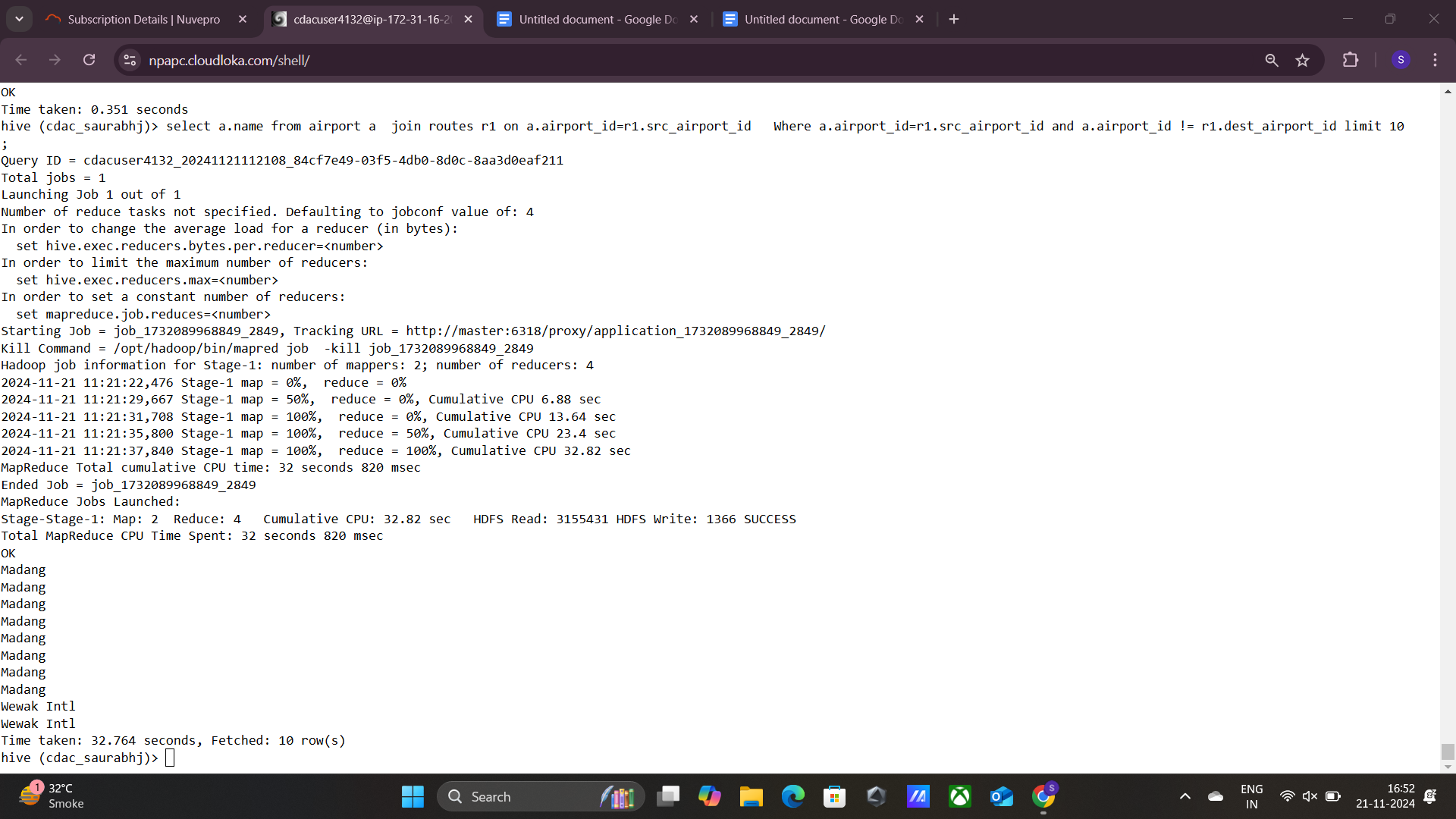
**HIVE**

**Q.1**

**1**. **select a.name from airport a**

**join routes r1 on a.airport\_id=r1.src\_airport\_id**

**Where a.airport\_id=r1.src\_airport\_id and a.airport\_id != r1.dest\_airport\_id limit 10;**



**2.**

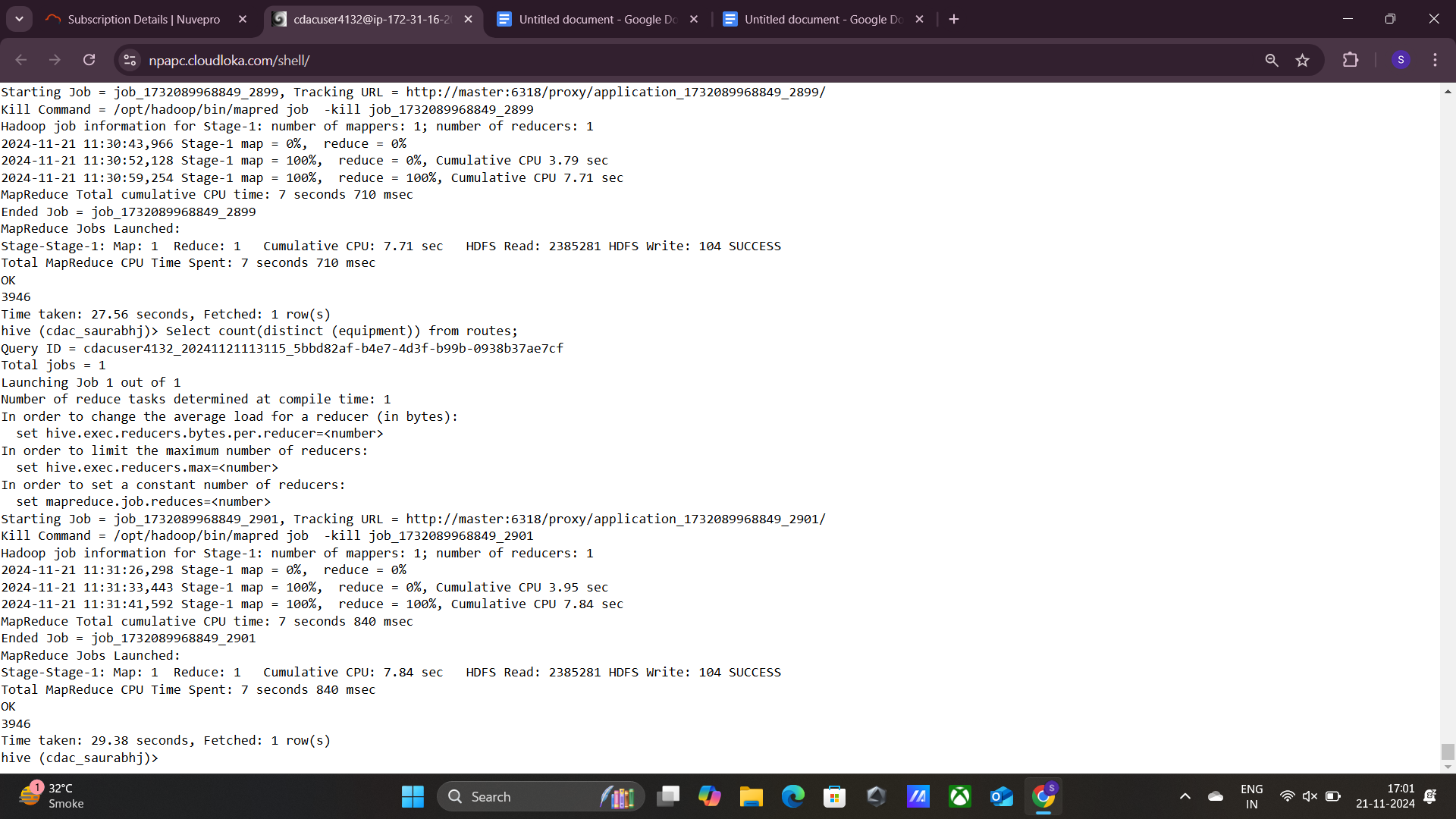
Select al.name from airlines al

join routes r on al.airline\_id=r.airline\_id

Where count(al.name)> count(distinct r.airline\_id)limit 10;

**3.**

**Select count(distinct (equipment)) from routes;**

****

**Q.2**

**1.**

**Set hive.exec.dynamic.partition.mode=nonstrict;**

**Set hive.exec.dynamic.partition=true;**

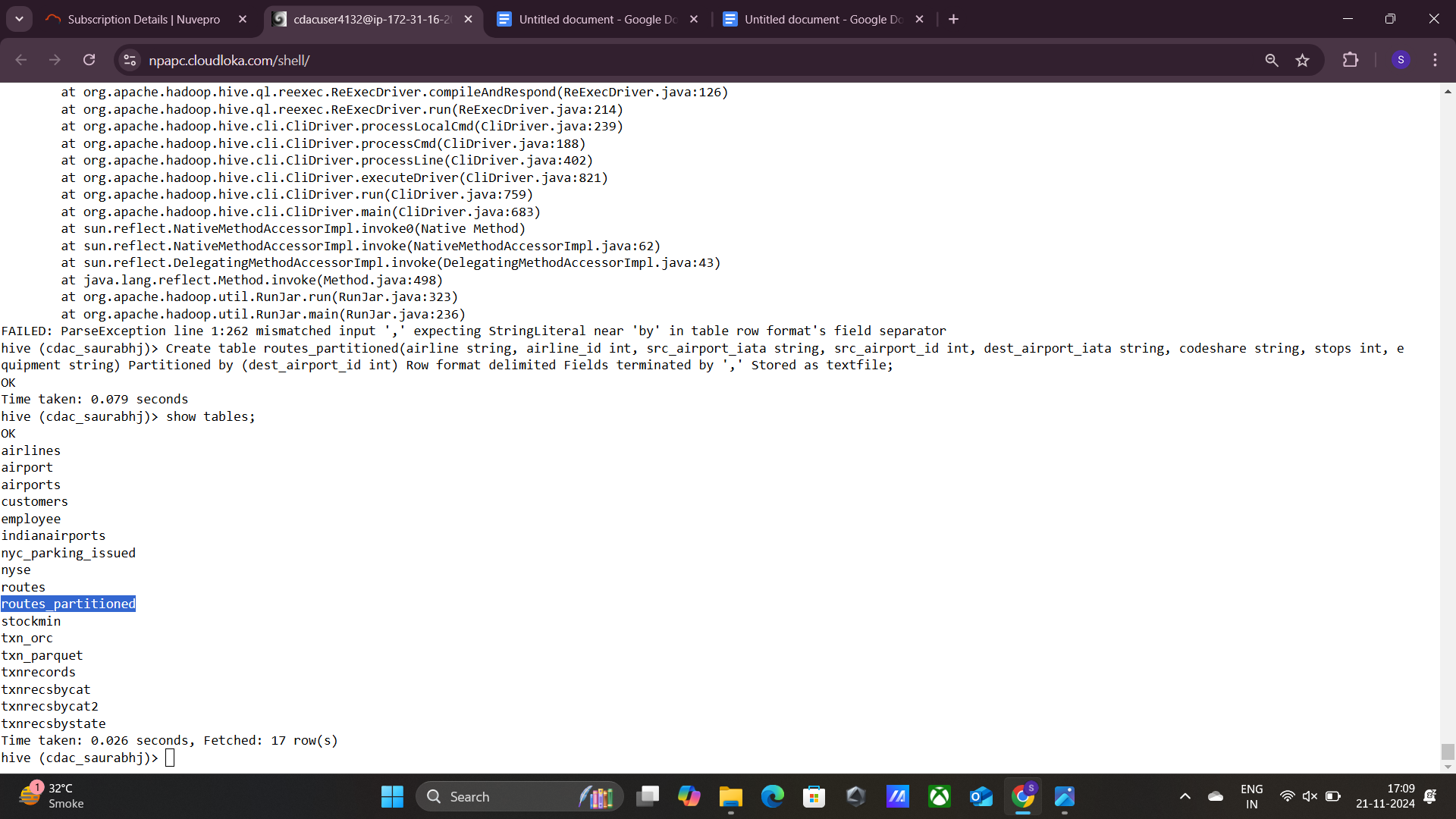
**Create table routes\_partitioned(airline string, airline\_id int, src\_airport\_iata string, src\_airport\_id int, dest\_airport\_iata string, codeshare string, stops int, equipment string)**

**Partitioned by (dest\_airport\_id int)**

**Row format delimited**

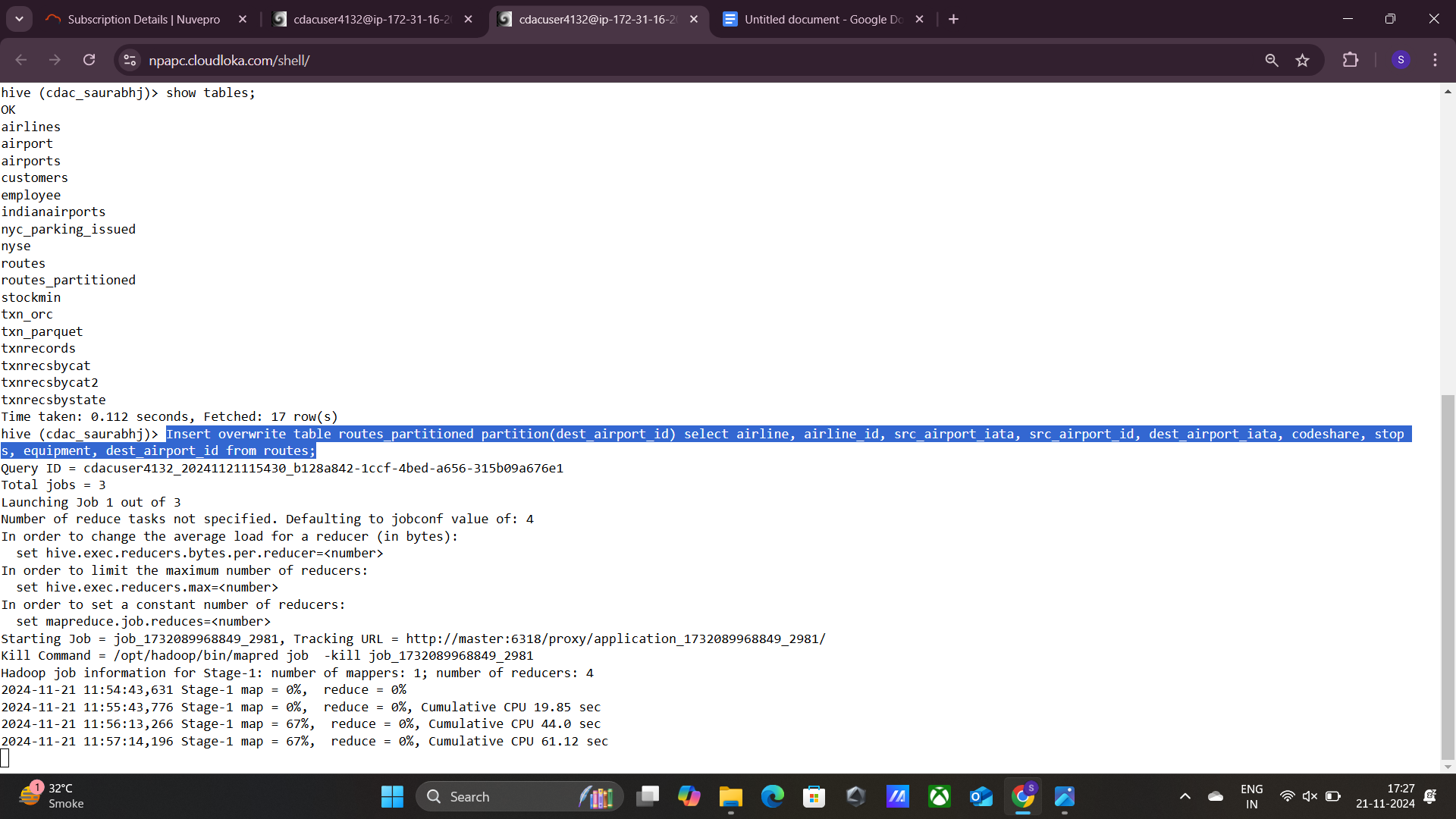
**Fields terminated by ‘,’**

**Stored as textfile;**

****

**2.**

**Insert overwrite table routes\_partitioned partition(dest\_airport\_id) select airline, airline\_id, src\_airport\_iata, src\_airport\_id, dest\_airport\_iata, codeshare, stops, equipment, dest\_airport\_id from routes;**

****

**3.**

**Select \* from routes\_partitioned where dest\_airport\_id =’ORD’ limit 10;**

**###\*Last (insert) query didn’t run due to lag, can’t perform operations\***

**4.**

**SPARK**

**Q.2**

**#loading data..**

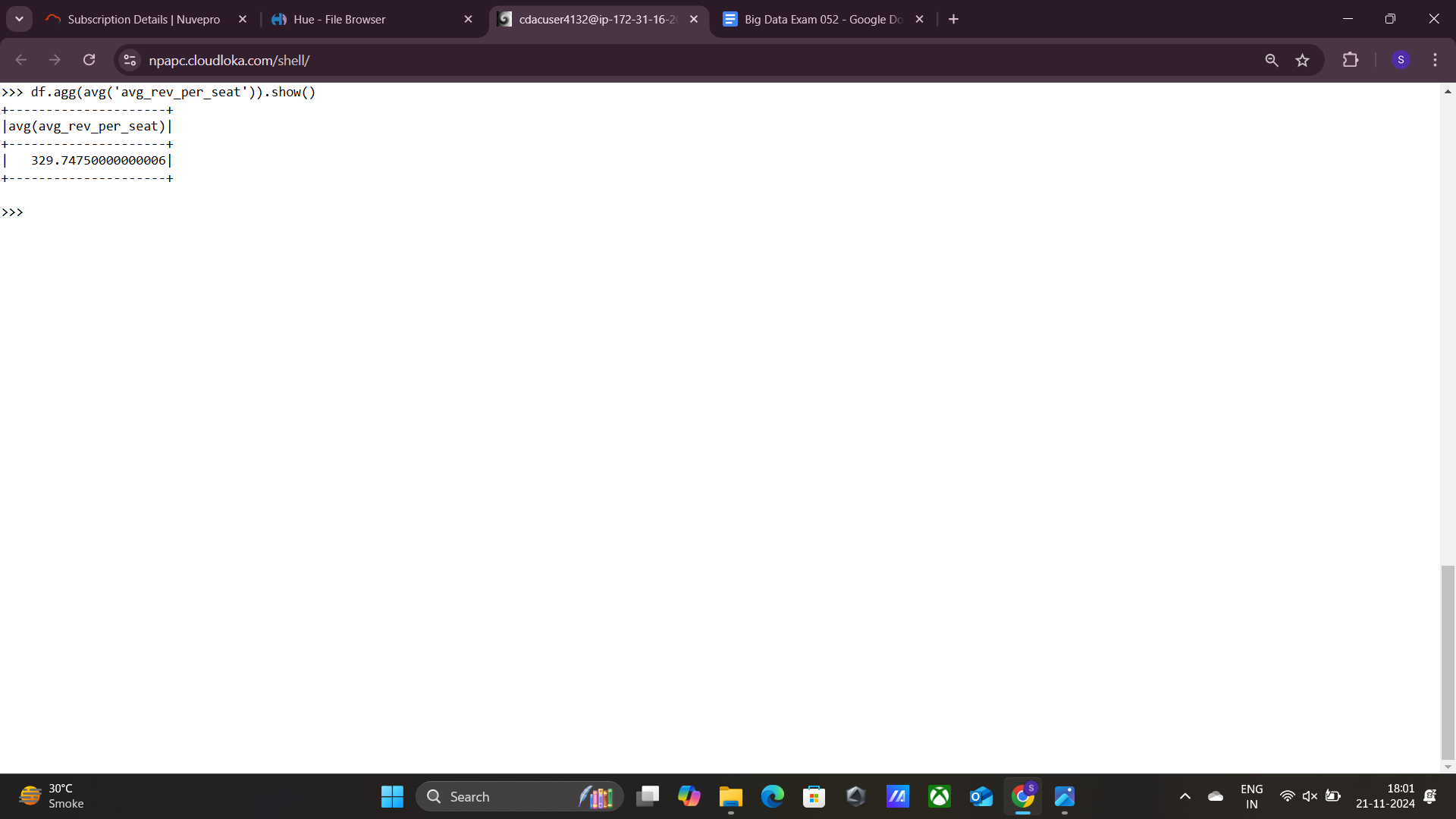
**from pyspark.sql.functions import sum,col,avg,max**

**df=spark.read.format("csv").option("header",True).option("inferSchema",True).load("/user/cdacuser4132/airlinesrdd")**

**1.**

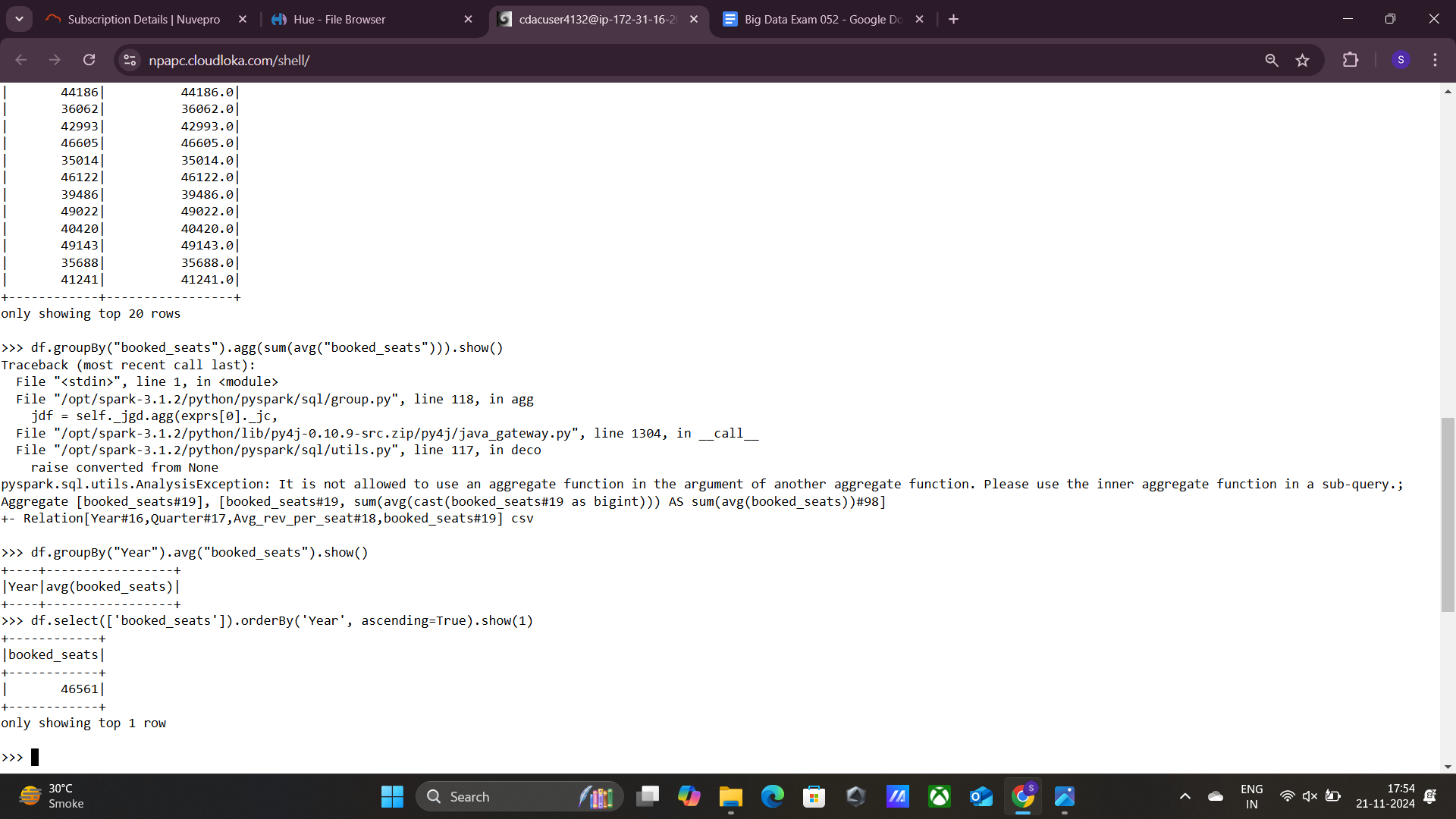
**avg:**

**df.agg(avg('avg\_rev\_per\_seat')).show()**

****

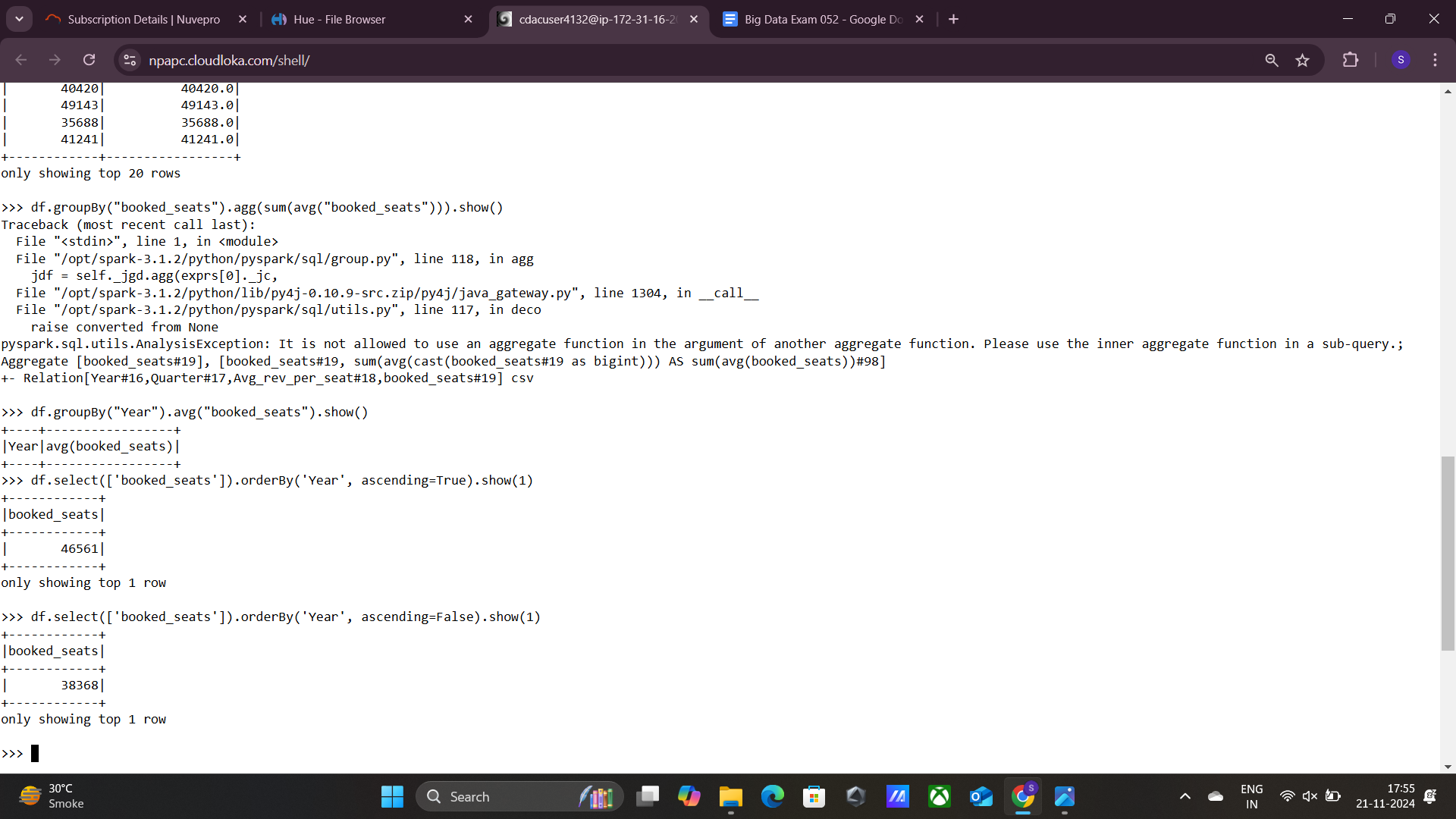
**Max:**

**df.select(['booked\_seats']).orderBy('Year', ascending=True).show(1)**

****

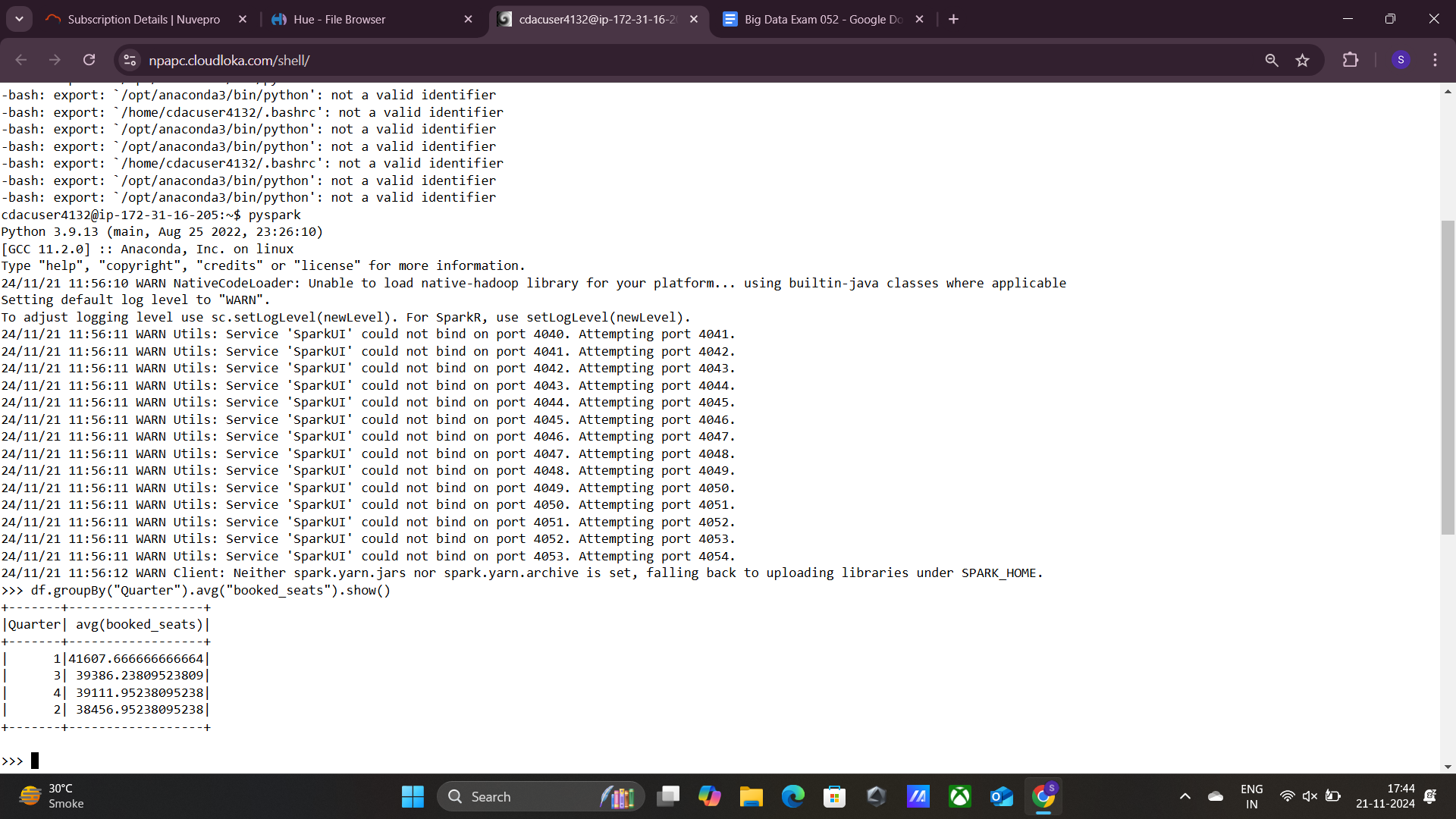
**Min:**

**df.select(['booked\_seats']).orderBy('Year', ascending=False).show(1)**

****

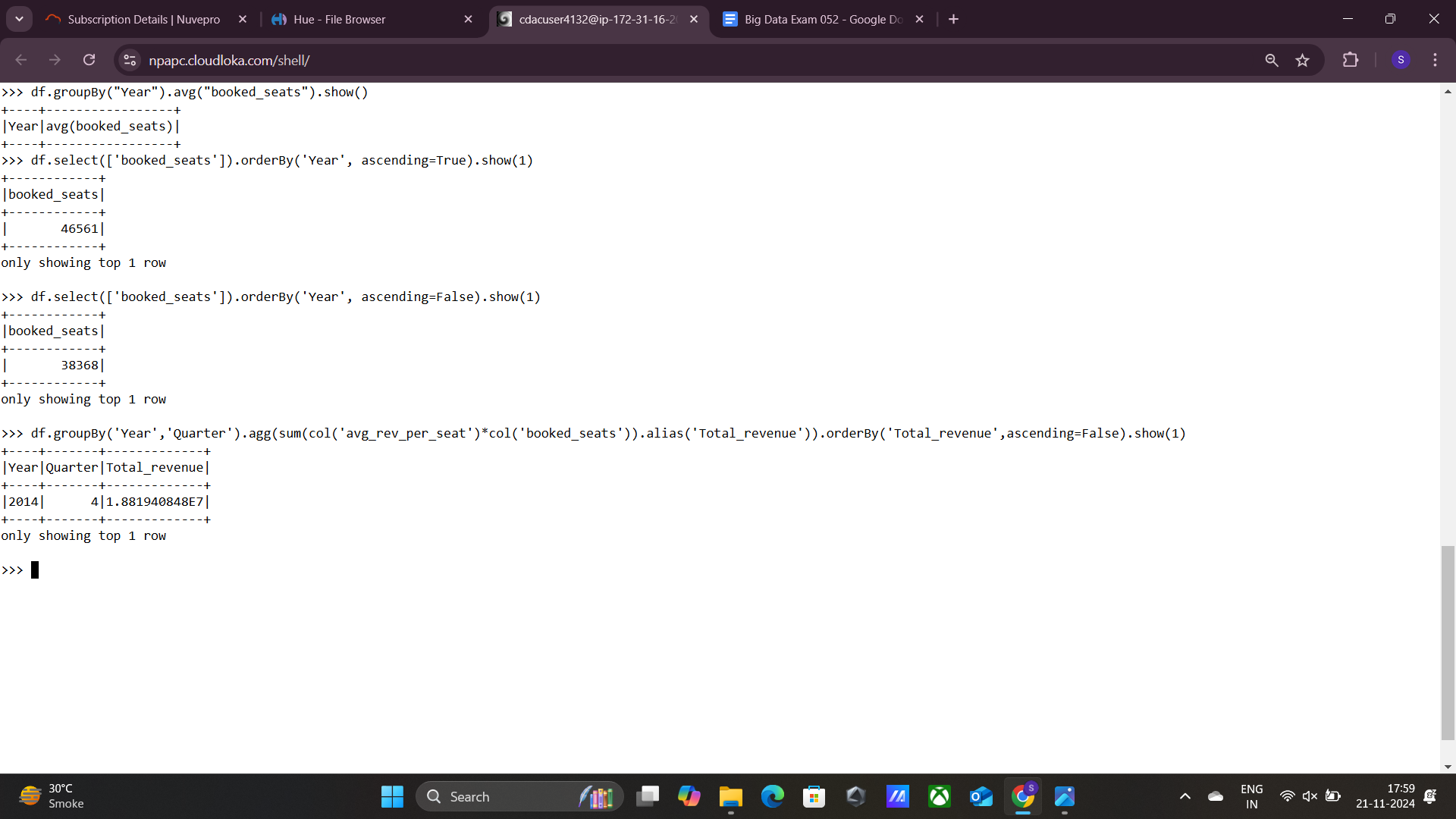
**3.**

**df.groupBy("Quarter").avg("booked\_seats").show()**

****

**5.**

**df.groupBy('Year','Quarter').agg(sum(col('avg\_rev\_per\_seat')\*col('booked\_seats')).alias('Total\_revenue')).orderBy('Total\_revenue',ascending=False).show(1)**

****