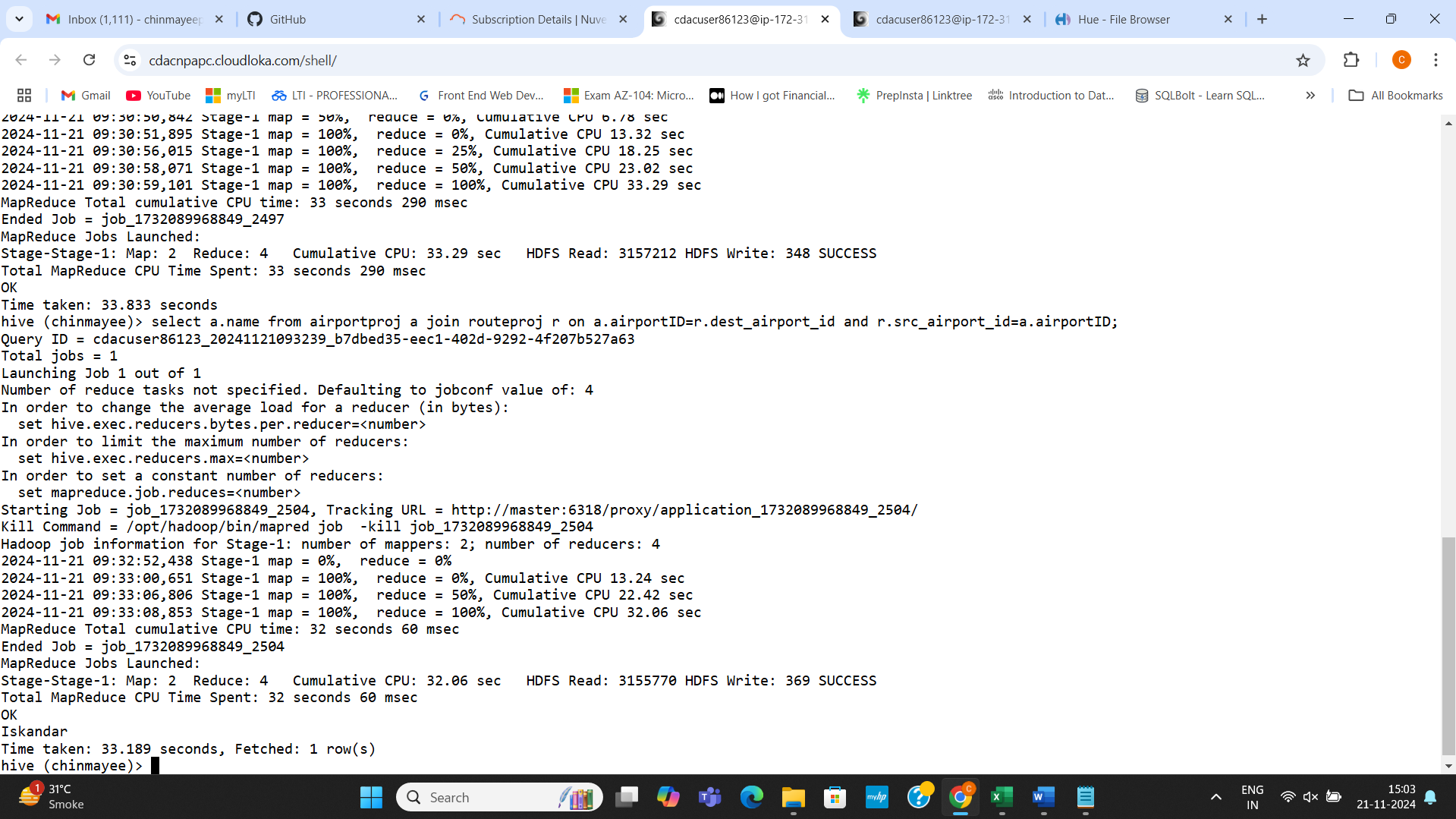
**240840325020**

**HIVE:**

Question 1:

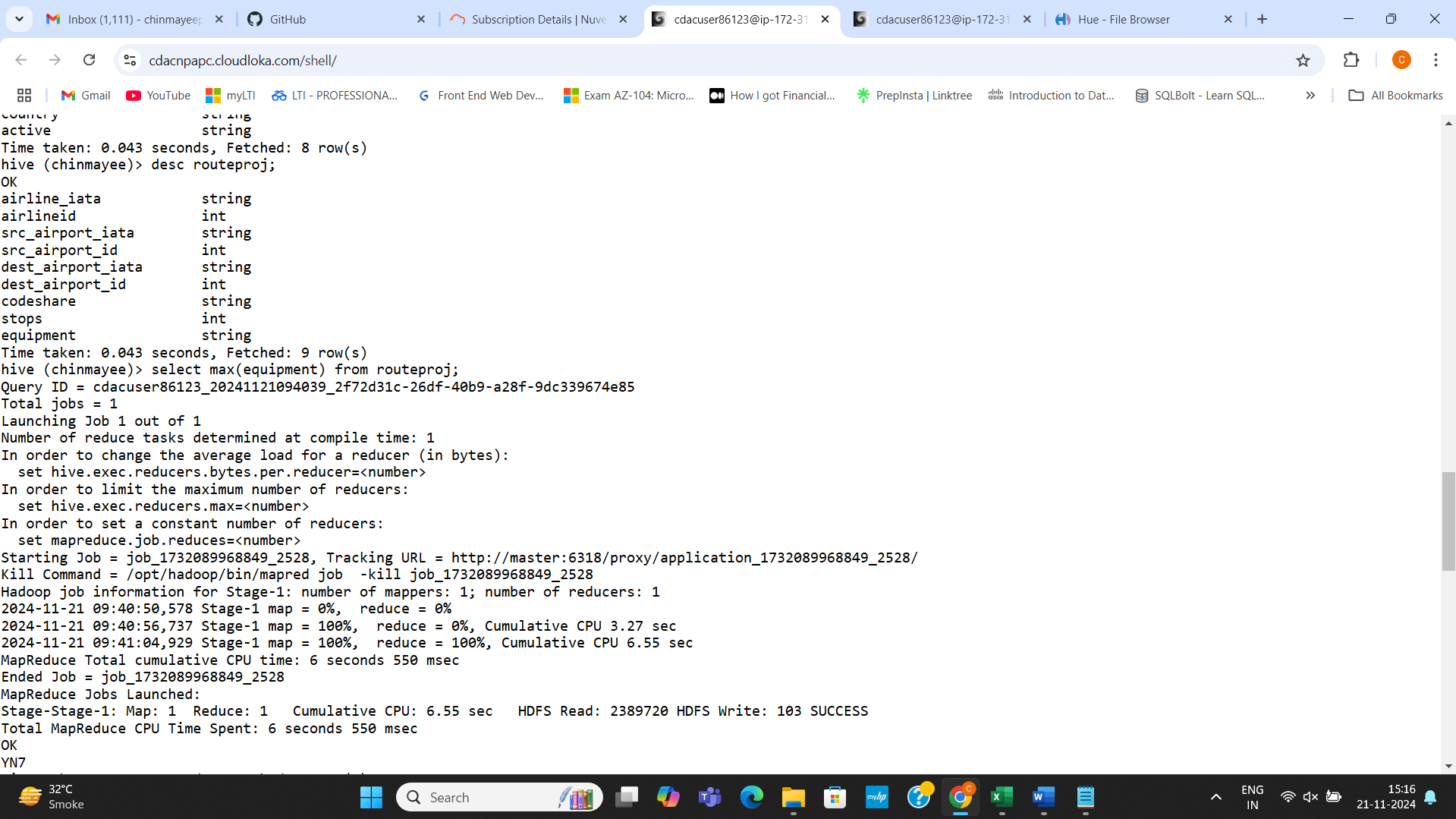
1. Select a.name from airportproj a join routeproj r on a.airportID=r.airlineID where r.dest\_airport\_id=a.airportID and r.src\_airport\_id=a.airportID;

select a.name from airportproj a join routeproj r on a.airportID=r.dest\_airport\_id where r.dest\_airport\_id is null limit 15;



1. Equipment

Select distinct equipment from routeproj limit 15;



Question 2:

1. create table Source\_Airport(airline\_iata String, airlineid int,src\_airport\_iata string, src\_airport\_id int,dest\_airport\_iata string,dest\_airport\_id int,codeshare string,stops int,equipment string)

row format delimited

fields terminated by ','

stored as textfile;

hadoop fs -put routes.dat /user/hive/warehouse/Chinmayee.db/Source\_Airport

set hive.exec.dynamic.partition.mode=nonstrict;

set hive.exec.dynamic.partition=true;

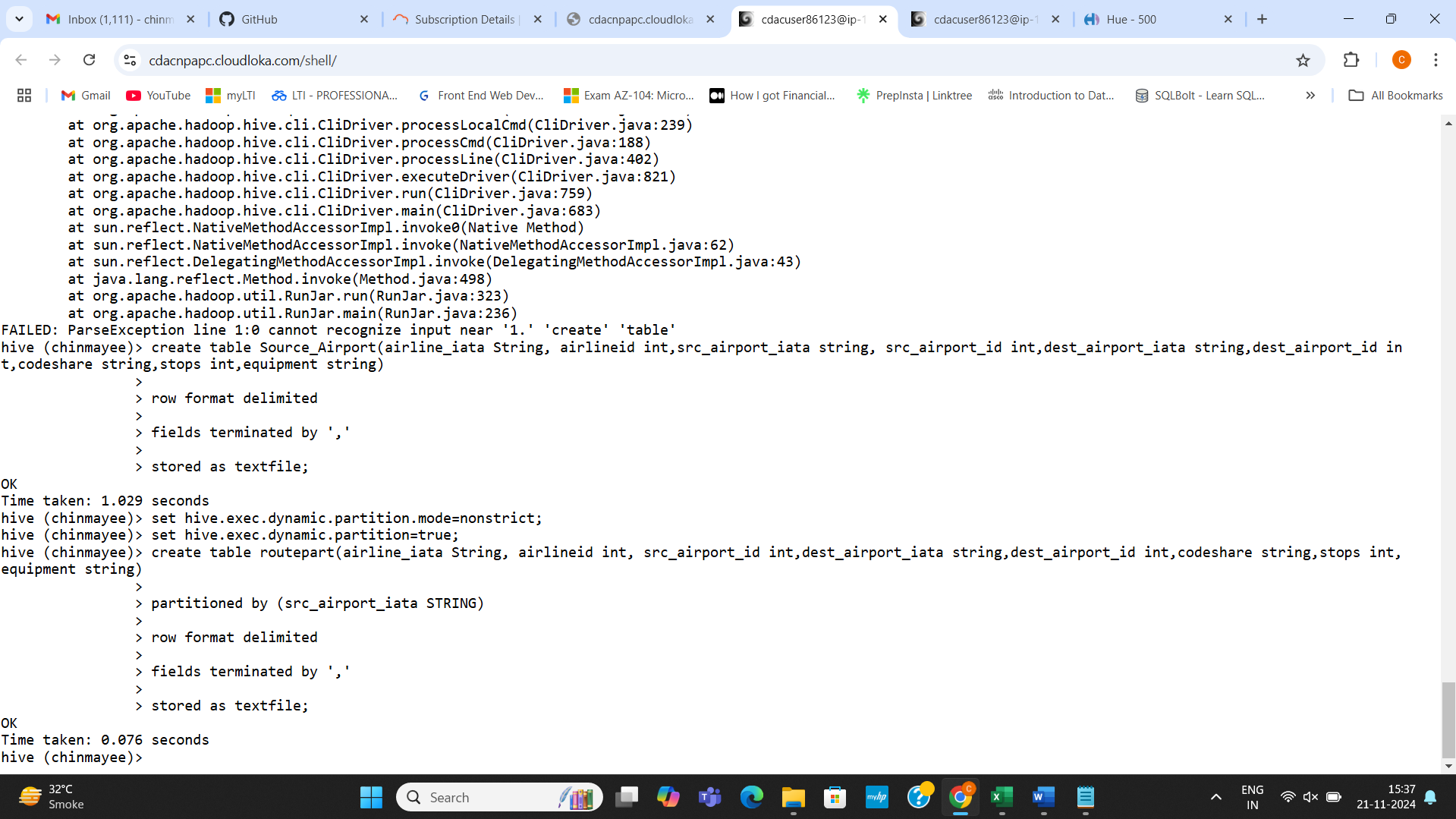
1. create table routepart(airline\_iata String, airlineid int, , src\_airport\_id int,dest\_airport\_iata string,dest\_airport\_id int,codeshare string,stops int,equipment string)

partitioned by (src\_airport\_iata STRING)

row format delimited

fields terminated by ','

stored as textfile;



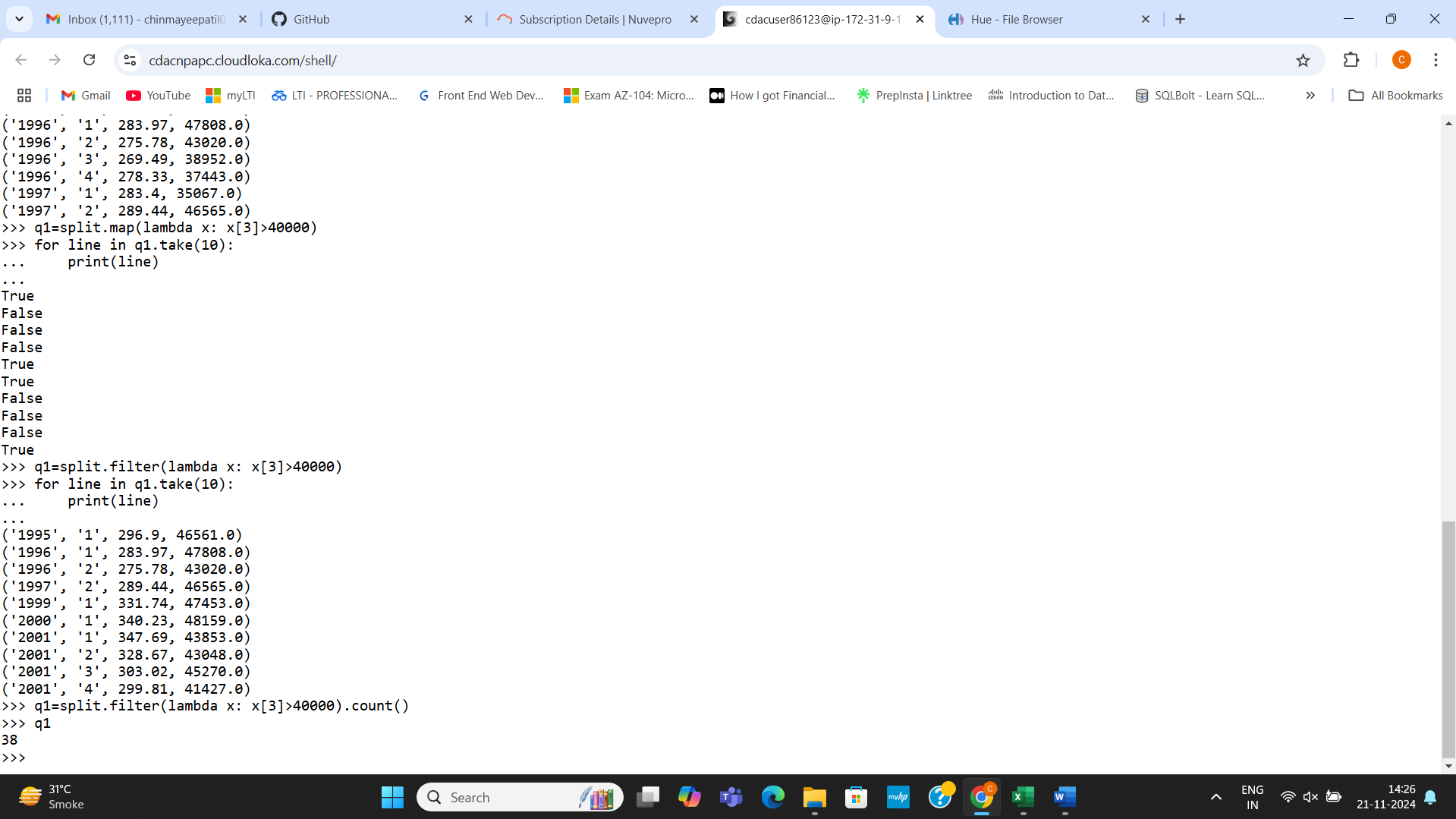
1. select src\_airport\_id, dest\_airport\_id from routepart where src\_airport\_iata='LAX';

hive was not working session closed error

**SPARK**

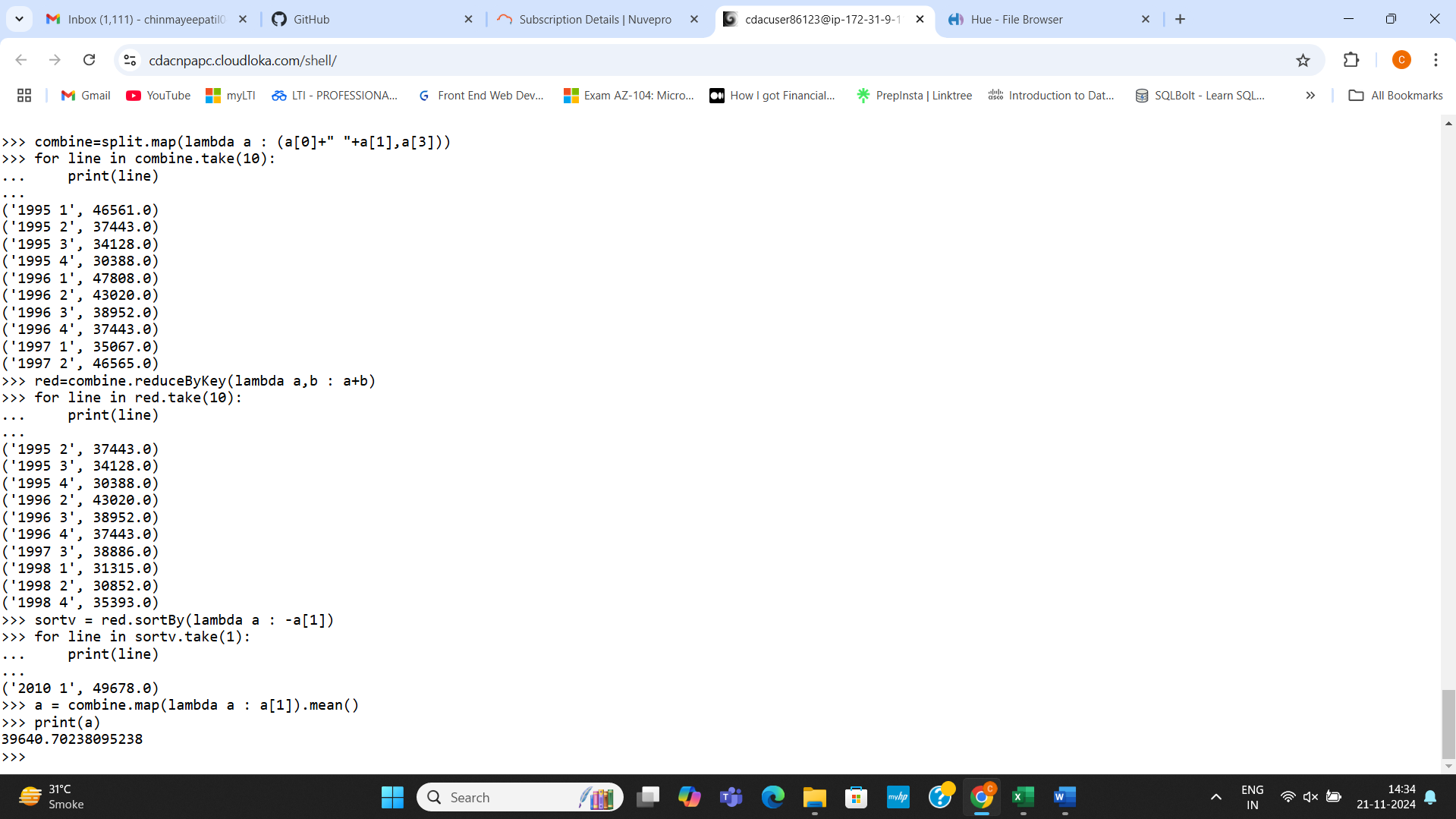
**1.**





Question 2:

1. Min,max and average



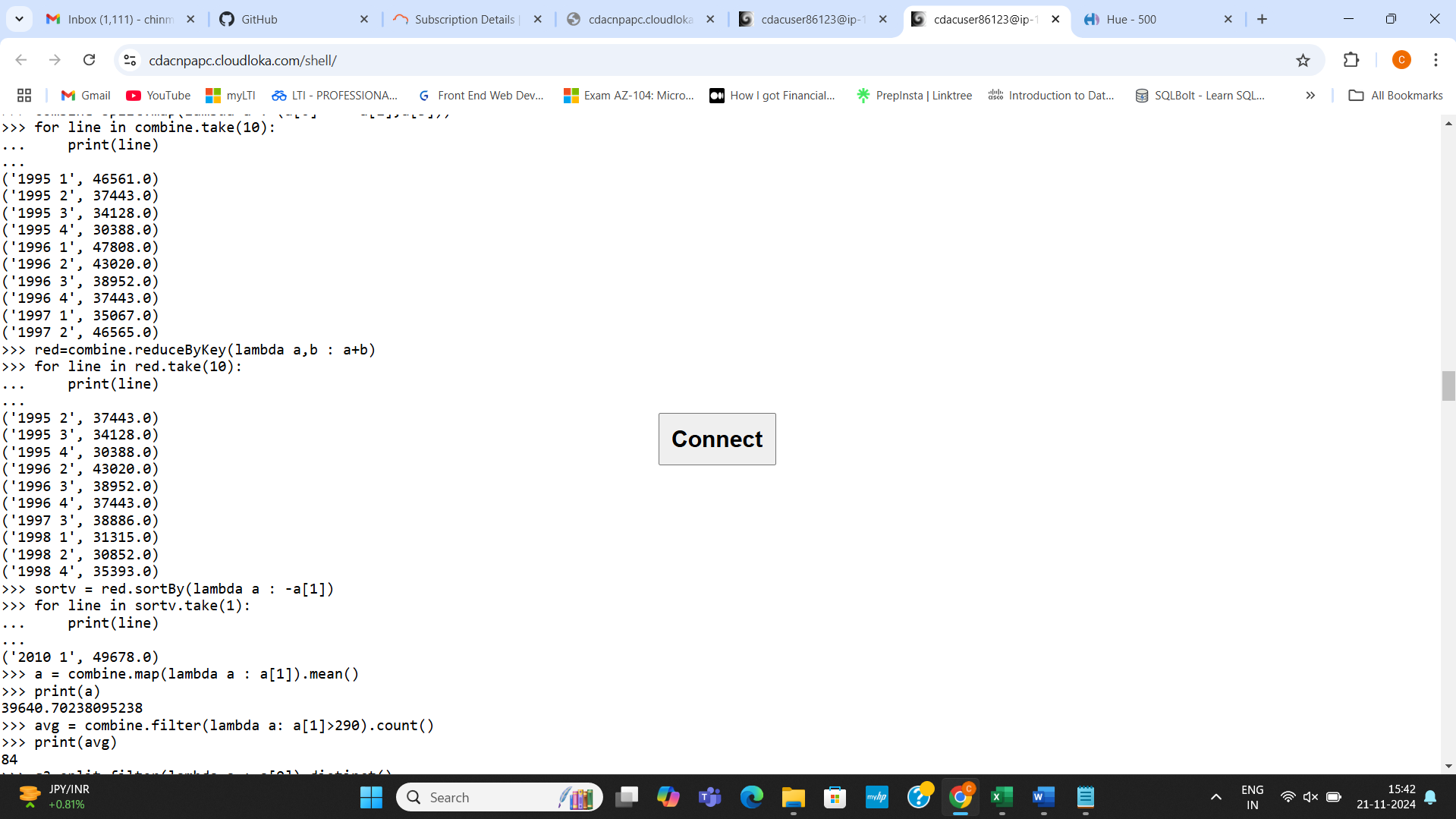
b=combine.map(lambda a: a[1].max()

print(b)

c=combine.map(lambda a:a[1].min()

print(c)

1. More than 290



1. combine=split.map(lambda a : (a[1],a[3]))

print(combine.take(4))

1. combine=split.map(lambda x : (x[0],x[1]))

c=combine.reduceByKey(lambda a,b : a+b)

for l in c.collect():

print(l)