Module exam

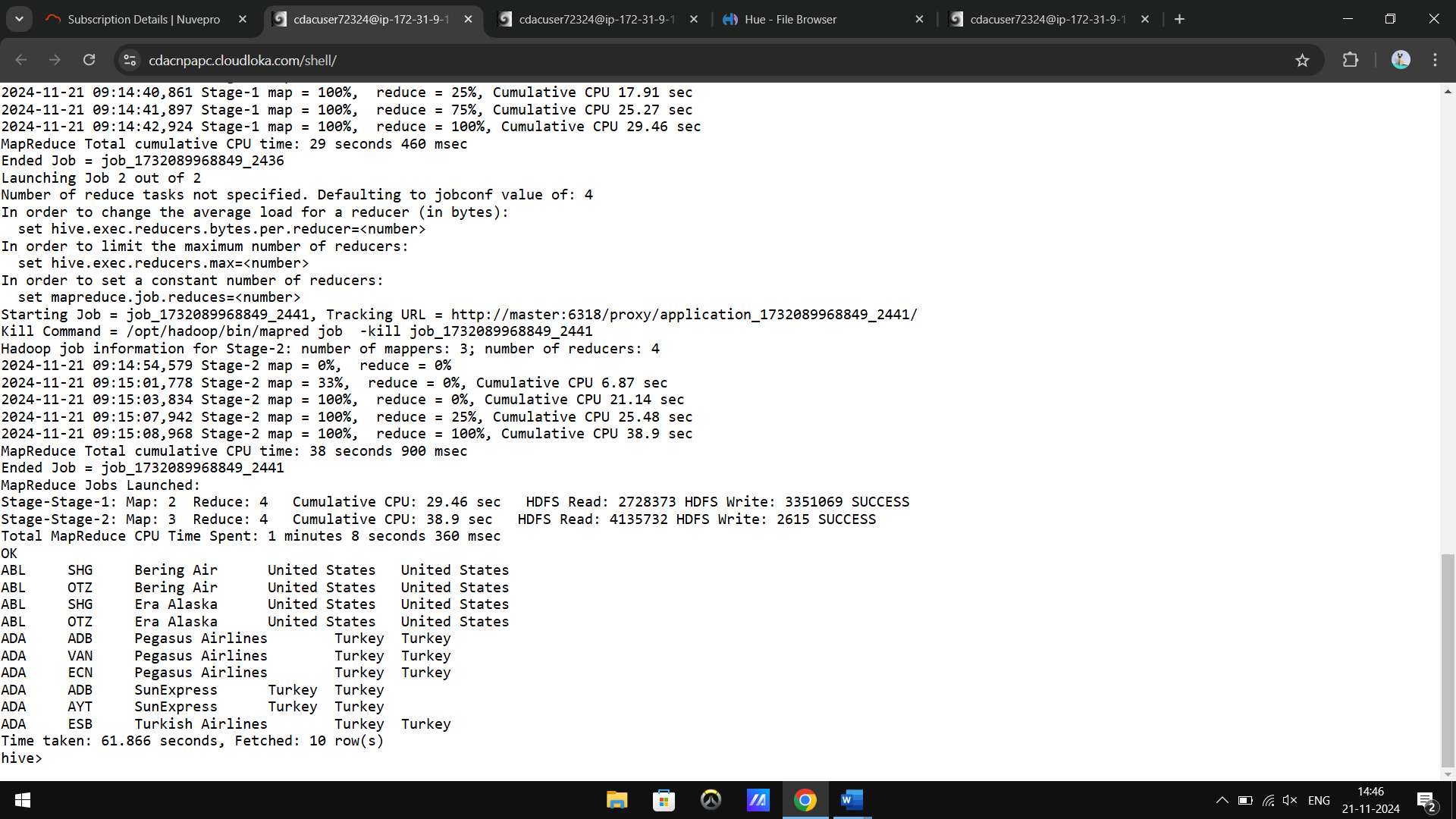
Name: Neelanjan Nandkumar Dalavi

Roll no. 240840325034

Hive

Question 1.

1. ans. select r.src\_airport\_iata,r.dest\_airport\_iata,a.name,a.country,a1.country from routes r join airlines a on r.airline\_id=a.airline\_id join airport a1 on r.src\_airport\_iata=a1.iata where a.country=a1.country limit 10;



2.

3.

Question 2.

1.

2.

3.

Spark

Question 1.

1. more = split.map(lambda a: a[3]>40000)

counter= more.count()

counter

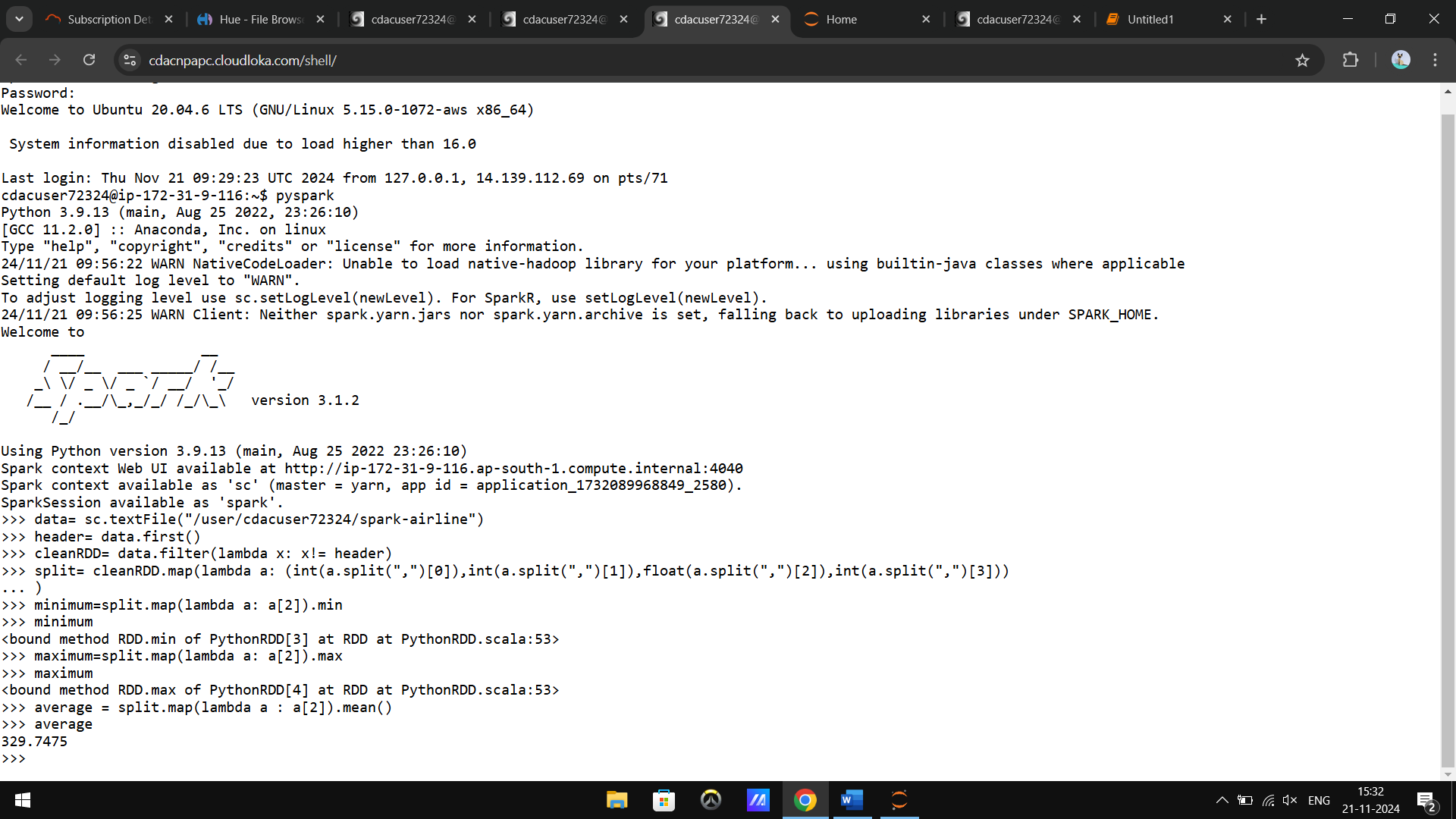
2.dist= split.map(lambda a: a[0]).distinct

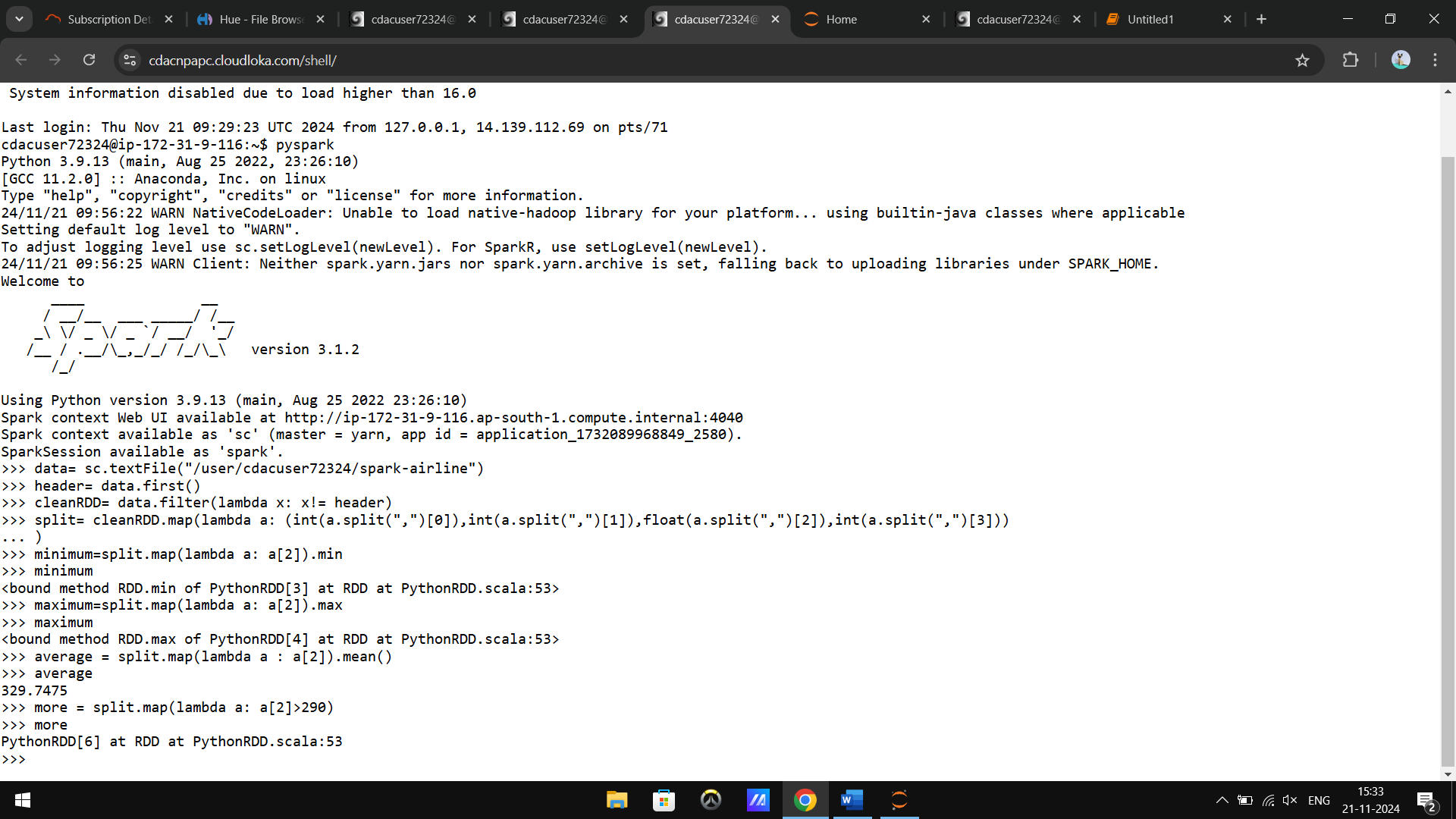
Question 2. ( every query in RDD)

1. minimum=split.map(lambda a: a[2]).min

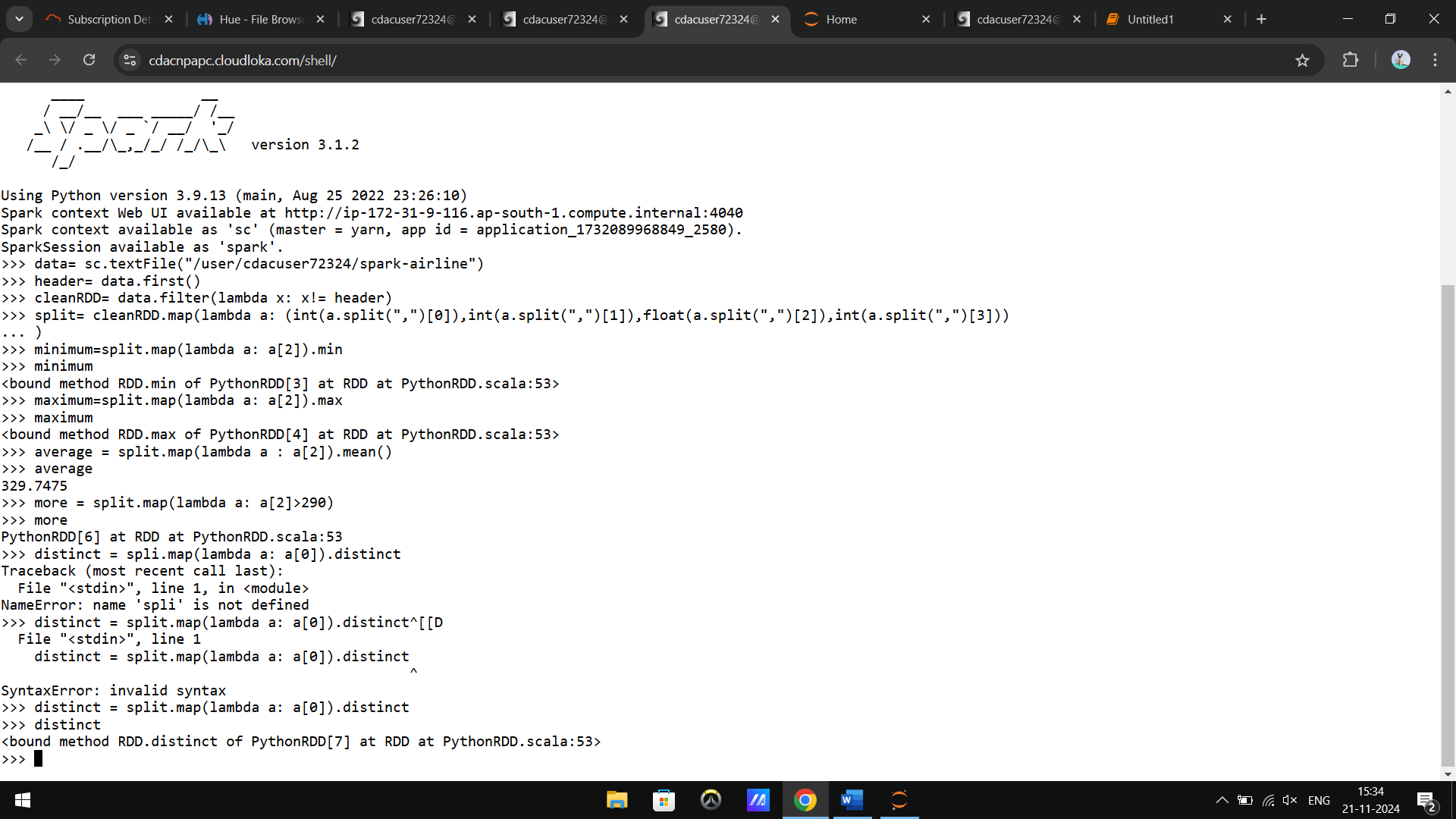
maximum=split.map(lambda a: a[2]).max

average = split.map(lambda a : a[2]).mean()



2. more = split.map(lambda a: a[2]>290) 

3.total= split.map(lambda a: a[3]).sum

4.distinct = spli.map(lambda a: a[0]).distinct 

5. combine = split.map(lambda a : (str(a[0]), a[2]))

average = combine.map(lambda a : a[1]).mean()