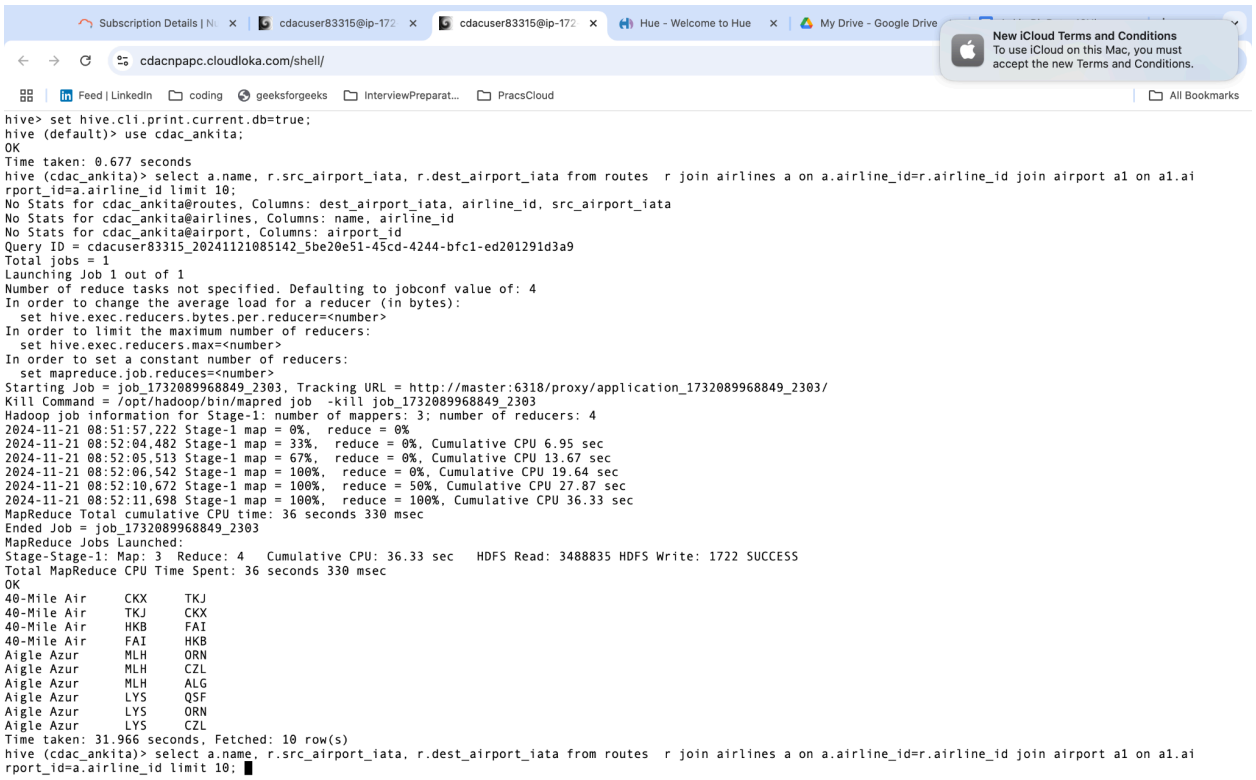
Hive
Ques1.
1.ans

```
select a.name, r.src_airport_iata, r.dest_airport_iata from routes  r join airlines a on a.airline_id=r.airline_id join airport a1 on a1.airport_id=a.airline_id limit 10;
```



2.

3.

```
select max(airline_iata), count(*) as no_routes from routes;
```

Subscription Details | Nuvep… | cdacuser83315@ip-172-31-9… | cdacuser83315@ip-172-31-9… | Hue - File Browser | My D…

cdacnpapc.cloudloka.com/shell/

Feed | LinkedIn | coding | geeksforgeeks | InterviewPreparat… | PracsCloud

Updates Available
Do you want to install the updates now or try tonight?

New iCloud Terms and Conditions
To use iCloud on this Mac, you must accept the new Terms and Conditions.

```
hive (cdac_ankita)> select distinct(a.name) from routes r join airline a on a.airline_id=r.airline_id where trim(upper(r.equi
FAILED: SemanticException [Error 10001]: Line 1:43 Table not found 'airline'
hive (cdac_ankita)> select distinct(a.name),max(equipment) from routes r join airline a on a.alirline_id=r.airline_id;
FAILED: SemanticException [Error 10001]: Line 1:58 Table not found 'airline'
hive (cdac_ankita)> select distinct(a.name),max(equipment) from routes r join airlines a on a.alirline_id=r.airline_id;
FAILED: SemanticException [Error 10002]: Line 1:74 Invalid column reference 'alirline_id'
hive (cdac_ankita)> select distinct(a.name),max(equipment) from routes r join airlines a on a.airline_id=r.airline_id;
FAILED: SemanticException [Error 10128]: Line 1:24 Not yet supported place for UDAF 'max'
hive (cdac_ankita)> select distinct(a.name),max(equipment) from routes r join airlines a on a.airline_id=r.airline_id group by airline_iata;
FAILED: SemanticException 1:107 SELECT DISTINCT and GROUP BY can not be in the same query. Error encountered near token 'airline_iata'
hive (cdac_ankita)> select distinct(a.name),max(equipment) from routes r join airlines a on a.airline_id=r.airline_id group by name;
FAILED: SemanticException 1:107 SELECT DISTINCT and GROUP BY can not be in the same query. Error encountered near token 'name'
hive (cdac_ankita)> select distinct(a.name),max(equipment) from routes r join airlines a on a.airline_id=r.airline_id group by airline_id;
FAILED: SemanticException 1:107 SELECT DISTINCT and GROUP BY can not be in the same query. Error encountered near token 'airline_id'
hive (cdac_ankita)> select distinct(a.name),max(equipment) from routes r join airlines a on a.airline_id=r.airline_id;
FAILED: SemanticException [Error 10128]: Line 1:24 Not yet supported place for UDAF 'max'
hive (cdac_ankita)> select a.name,max(equipment) from routes r join airlines a on a.airline_id=r.airline_id group by airline_id;
FAILED: SemanticException Column airline_id Found in more than One Tables/Subqueries
FAILED: SemanticException 1:107 SELECT DISTINCT and GROUP BY can not be in the same query. Error encountered near token 'airline_iata'_iata;
hive (cdac_ankita)> select distinct(a.name),max(equipment) from routes r join airlines a on a.airline_id=r.airline_id ;
FAILED: SemanticException [Error 10128]: Line 1:24 Not yet supported place for UDAF 'max'
hive (cdac_ankita)> select max(airline_iata), count(*) as no_routes from routes;
Query ID = cdacuser83315_20241121101922_5fa23171-7d89-40fa-a4fb-1b5608415bfd
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1732089968849_2661, Tracking URL = http://master:6318/proxy/application_1732089968849_2661/
Kill Command = /opt/hadoop/bin/mapred job  -kill job_1732089968849_2661
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2024-11-21 10:19:32,970 Stage-1 map = 0%,  reduce = 0%
2024-11-21 10:19:41,168 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.31 sec
2024-11-21 10:19:49,370 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 6.69 sec
MapReduce Total cumulative CPU time: 6 seconds 690 msec
Ended Job = job_1732089968849_2661
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 6.69 sec   HDFS Read: 2390588 HDFS Write: 108 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 690 msec
OK
ZM      67663
Time taken: 29.453 seconds, Fetched: 1 row(s)
hive (cdac_ankita)>
```

Spark:

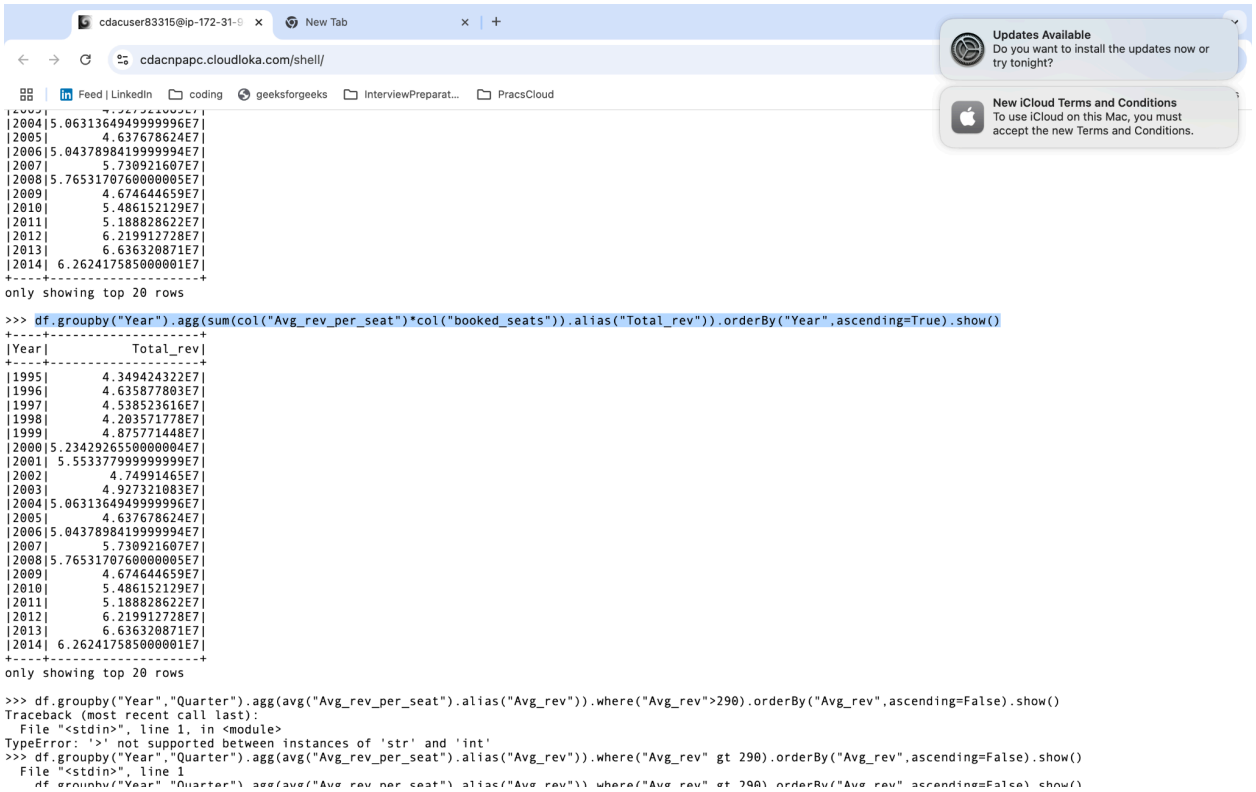
<u>Ques1</u>

1.Ans
df.groupby(“Year”,”booked_seat”).where(“booked_seats”>40000).show()

2. Ans

Using df show distinct years data

```
df.groupby("Year").agg(sum(col("Avg_rev_per_seat")*col("booked_seats"
)).alias("Total_rev")).orderBy("Year",ascending=True).show()
```

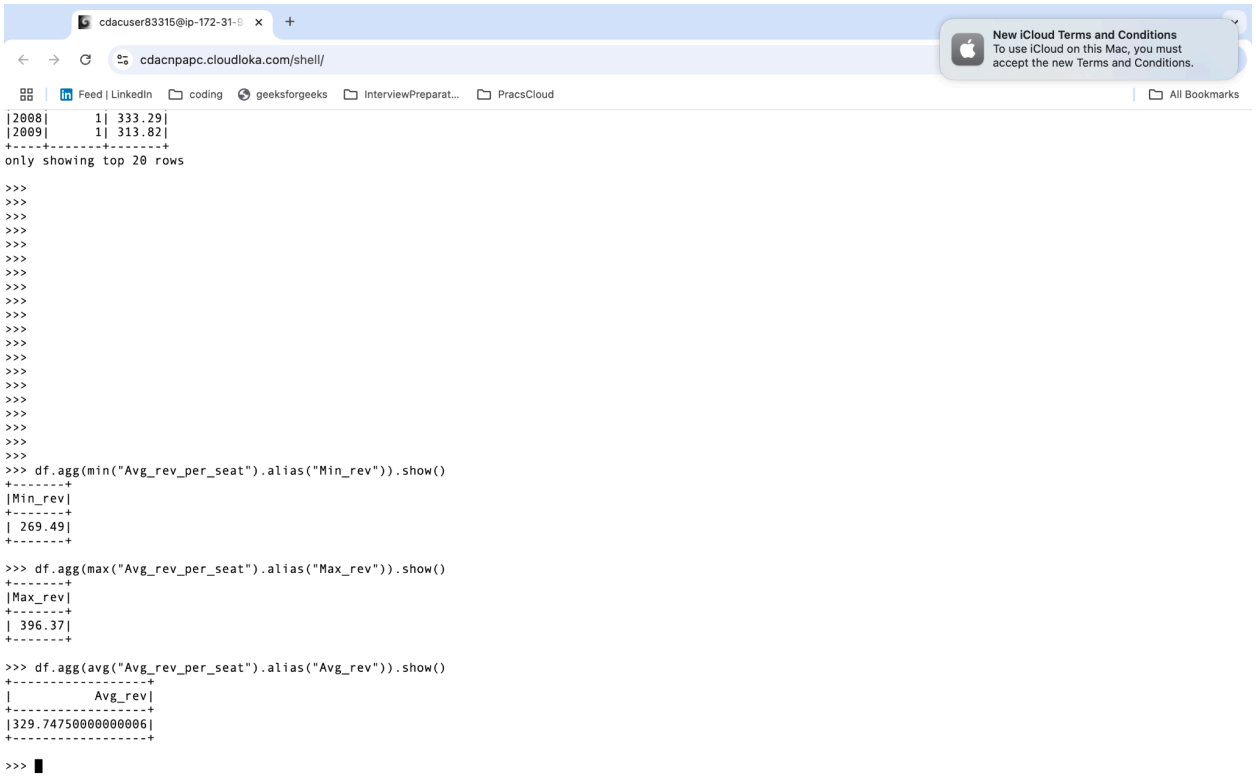


Ques2:

1 ans

```
df.agg(min("Avg_rev_per_seat").alias("Min_rev")).show()
df.agg(max("Avg_rev_per_seat").alias("Max_rev")).show()
```

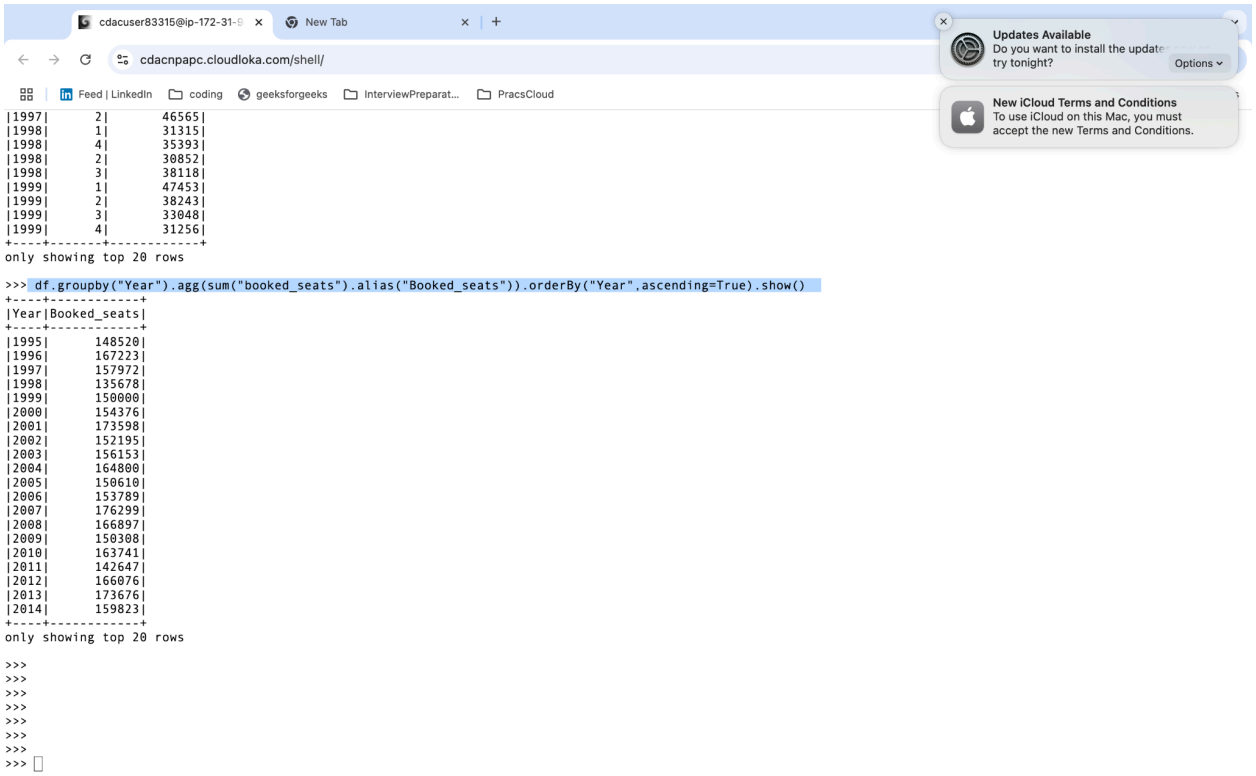
```
df.agg(avg("Avg_rev_per_seat").alias("Avg_rev")).show()
```



2.

3.

```
df.groupby("Year").agg(sum("booked_seats").alias("Booked_seats")).orderBy("Year",ascending=True).show()
```



4.

```
df.groupby("Year").agg(sum("booked_seats").alias("Booked_seats")).orderBy("Year",ascending=True).show()
```

cdacuser83315@ip-172-31-6 | New Tab

cdacnpapc.cloudloka.com/shell/

Feed | LinkedIn | coding | geeksforgeeks | InterviewPreparat... | PracsCloud

Updates Available
Do you want to install the updates now or try tonight?

New iCloud Terms and Conditions
To use iCloud on this Mac, you must accept the new Terms and Conditions.

```
only showing top 20 rows

>>> df.groupby("Year").orderBy("Year",ascending=True).show()
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
AttributeError: 'GroupedData' object has no attribute 'orderBy'
>>> df.groupby("Year").orderby("Year",ascending=True).show()
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
AttributeError: 'GroupedData' object has no attribute 'orderby'
>>> df.groupby("Year").agg(sum("booked_seats").alias("Booked_seats"))orderBy("Year",ascending=True).show()
  File "<stdin>", line 1
    df.groupby("Year").agg(sum("booked_seats").alias("Booked_seats"))orderBy("Year",ascending=True).show()
                                                                     ^
SyntaxError: invalid syntax
>>> df.groupby("Year").agg(sum("booked_seats").alias("Booked_seats")).orderBy("Year",ascending=True).show()
+----+------------+
|Year|Booked_seats|
+----+------------+
|1995|      148520|
|1996|      167223|
|1997|      157972|
|1998|      135678|
|1999|      150000|
|2000|      154376|
|2001|      173598|
|2002|      152195|
|2003|      156153|
|2004|      164800|
|2005|      150610|
|2006|      153789|
|2007|      176299|
|2008|      166897|
|2009|      150308|
|2010|      163741|
|2011|      142647|
|2012|      166076|
|2013|      173676|
|2014|      159823|
+----+------------+
only showing top 20 rows

>>> df.groupby("Year").show()
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
AttributeError: 'GroupedData' object has no attribute 'show'
>>> 
```

5.

```
df.groupby("Year").agg(sum(col("Avg_rev_per_seat")*col("booked_seats"
)).alias("Total_rev")).orderBy("Year",ascending=True).show()
```

cdacuser83315@ip-172-31-6 | New Tab

cdacnpapc.cloudloka.com/shell/

Feed | LinkedIn | coding | geeksforgeeks | InterviewPreparat... | PracsCloud

Updates Available
Do you want to install the updates now or try tonight?

New iCloud Terms and Conditions
To use iCloud on this Mac, you must accept the new Terms and Conditions.

```
|2000|5.2342926550000004E7|
|2001| 5.553377999999999E7|
|2002|        4.74991465E7|
|2003|       4.927321083E7|
|2004|5.0631364949999996E7|
|2005|       4.637678624E7|
|2006|5.0437898419999994E7|
|2007|       5.730921607E7|
|2008|5.7653170760000005E7|
|2009|       4.674644659E7|
|2010|       5.486152129E7|
|2011|       5.188828622E7|
|2012|       6.219912728E7|
|2013|       6.636320871E7|
|2014| 6.262417585000001E7|
+----+--------------------+
only showing top 20 rows

>>> df.groupby("Year").agg(sum(col("Avg_rev_per_seat")*col("booked_seats")).alias("Total_rev")).orderBy("Year",ascending=True).show()
+----+--------------------+
|Year|           Total_rev|
+----+--------------------+
|1995|       4.349424322E7|
|1996|       4.635877803E7|
|1997|       4.538523616E7|
|1998|       4.203571778E7|
|1999|       4.875771448E7|
|2000|5.2342926550000004E7|
|2001| 5.553377999999999E7|
|2002|        4.74991465E7|
|2003|       4.927321083E7|
|2004|5.0631364949999996E7|
|2005|       4.637678624E7|
|2006|5.0437898419999994E7|
|2007|       5.730921607E7|
|2008|5.7653170760000005E7|
|2009|       4.674644659E7|
|2010|       5.486152129E7|
|2011|       5.188828622E7|
|2012|       6.219912728E7|
|2013|       6.636320871E7|
|2014| 6.262417585000001E7|
+----+--------------------+
only showing top 20 rows

>>> 
```