Report on Analysis of Used Car Sales in Germany and Czech Republic

Submitted By:

Karthikeyan Suresh Kumar

200489370

BDAT

Time Table A

Contents

Overview

This report provides an in-depth analysis of the used car sales in Germany and Czech Republic since 2015. The analysis gives necessary insights into the used car sales business and provides us with the key understandings of the market trends available for this business.

Through this data and its outcomes, the management would be able to take key decisions about venturing into the used car sales business.

Dataset Details

The dataset that has been used for the analysis is the data scraped from various websites and surveys. The dataset has a count of 3.5 million rows.

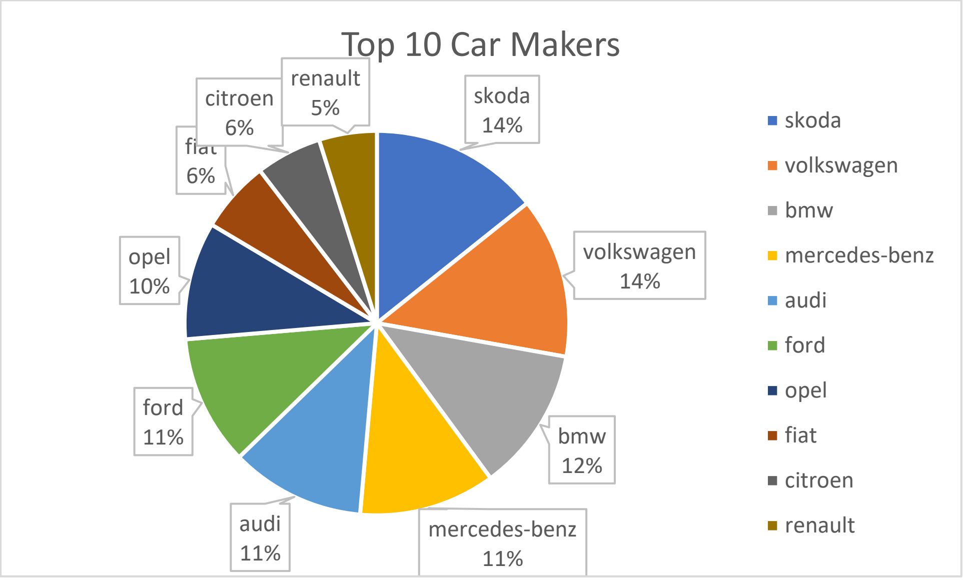
The dataset contains the following columns:

- maker
- model
- mileage
- manufacture_year
- engine_displacement
- engine_power
- color_slug
- stk_year
- transmission
- door_count
- seat_count
- fuel_type
- date_created
- datelastseen
- price_eur

Analysis

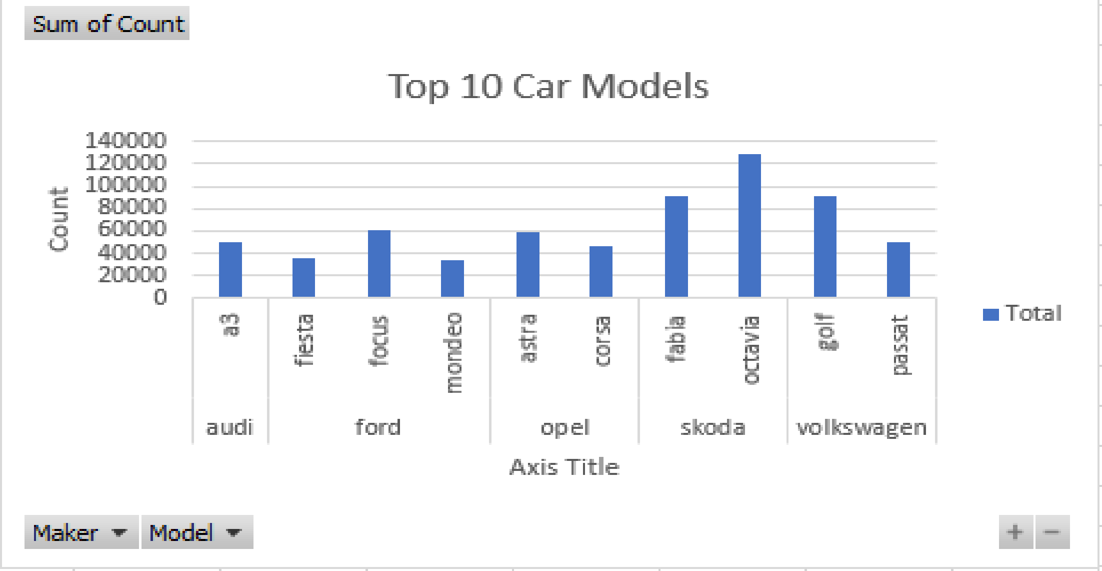
Analysis has been performed on the variables of the data to find the relation between the variables. The following analysis questions were answered using the data.

1) Which are the top 10 car makers in terms of numbers?



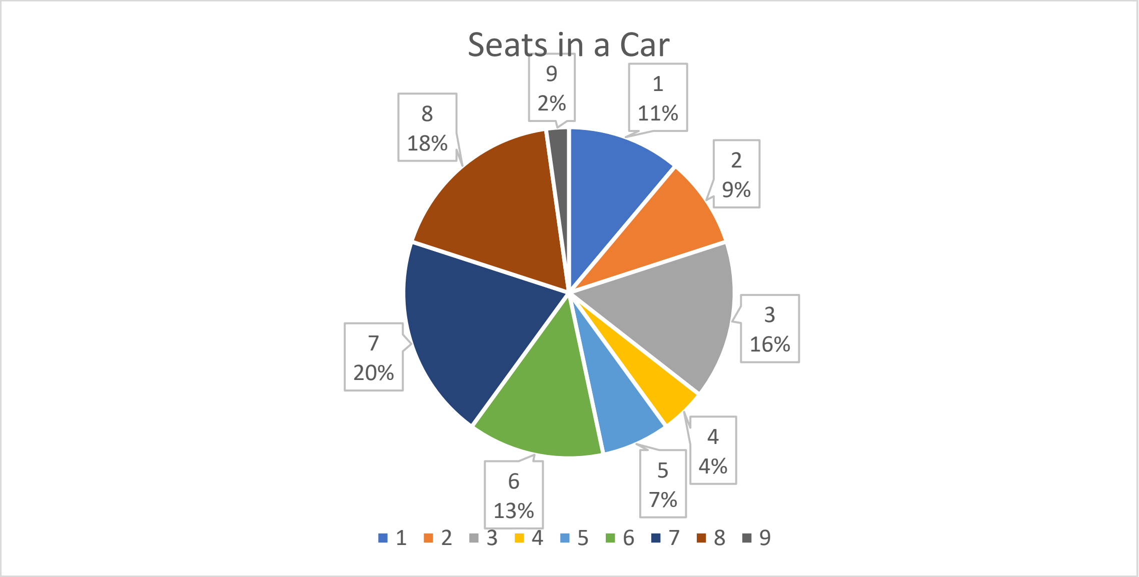
Upon the analysis of the data, it has been found that the Top 10 car makers vehicles sold are as in the above pie chart. From the chart, it can be seen that Skoda and Volkswagen top the list with equal share percentage.

2) Which are the top 5 models in terms of numbers?



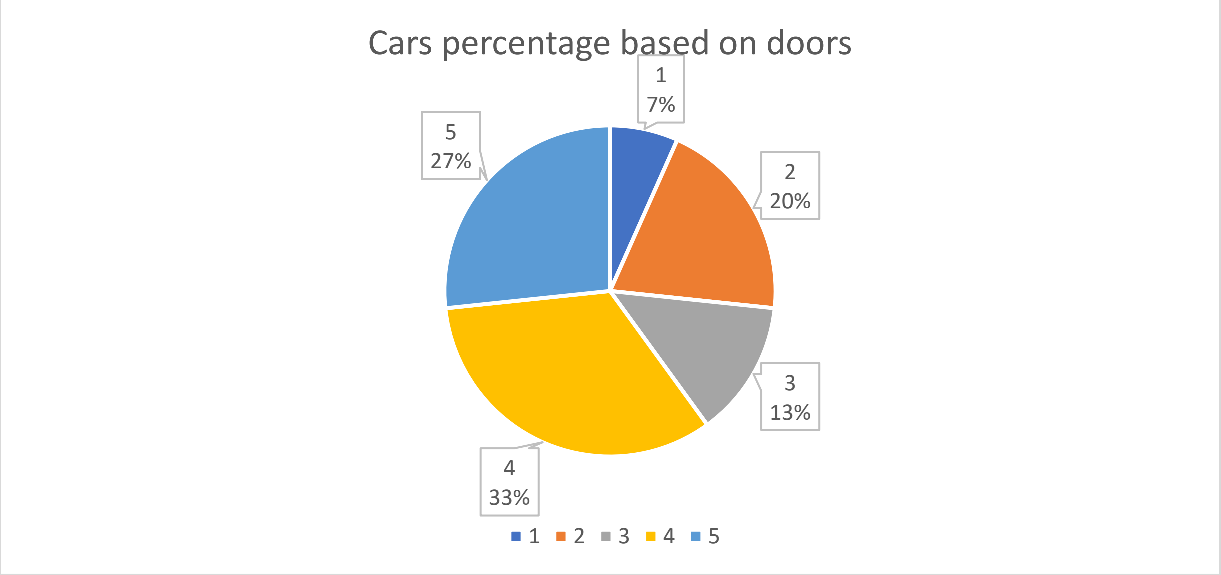
From the data, it was clearly seen that Fabia and Octavia models of the Skoda company are the most sold cars. Other models of the cars such as Golf, Passat, Focus, etc., are equally sold in large numbers as the Fabia and Octavia.

3) What is the most common preferred seat capacity for a car?



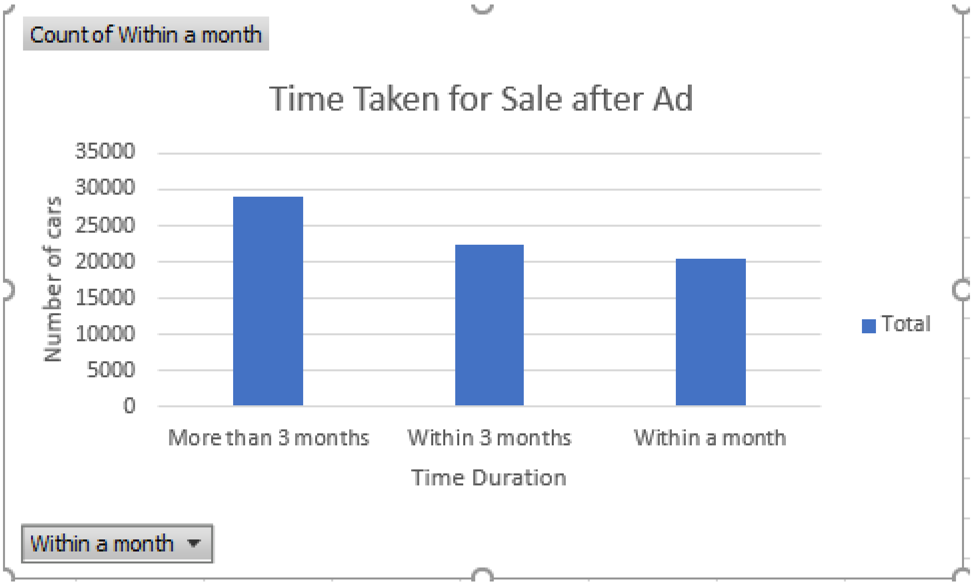
From the above pie chart, it is quite visible that the most common preference for the car is 5 seats. This shows the consumer interest in the buying of a car. Based on this we can take business decisions like the priority buying and selling of 5-seater car in large numbers.

4) What is the most common preferred door capacity for a car?



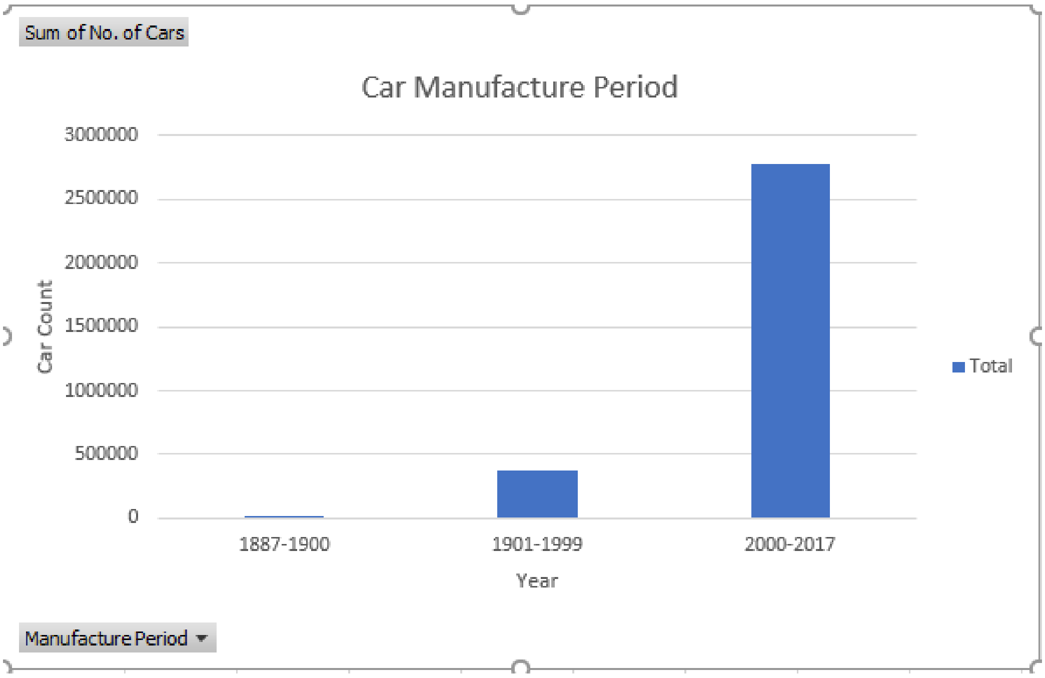
People have preference for a 4-door car rather than the other models. The majority of the cars listed for sale are 4-door cars, with the second majority being 5-door cars.

5) **Time taken for a car to be sold after ad post.**



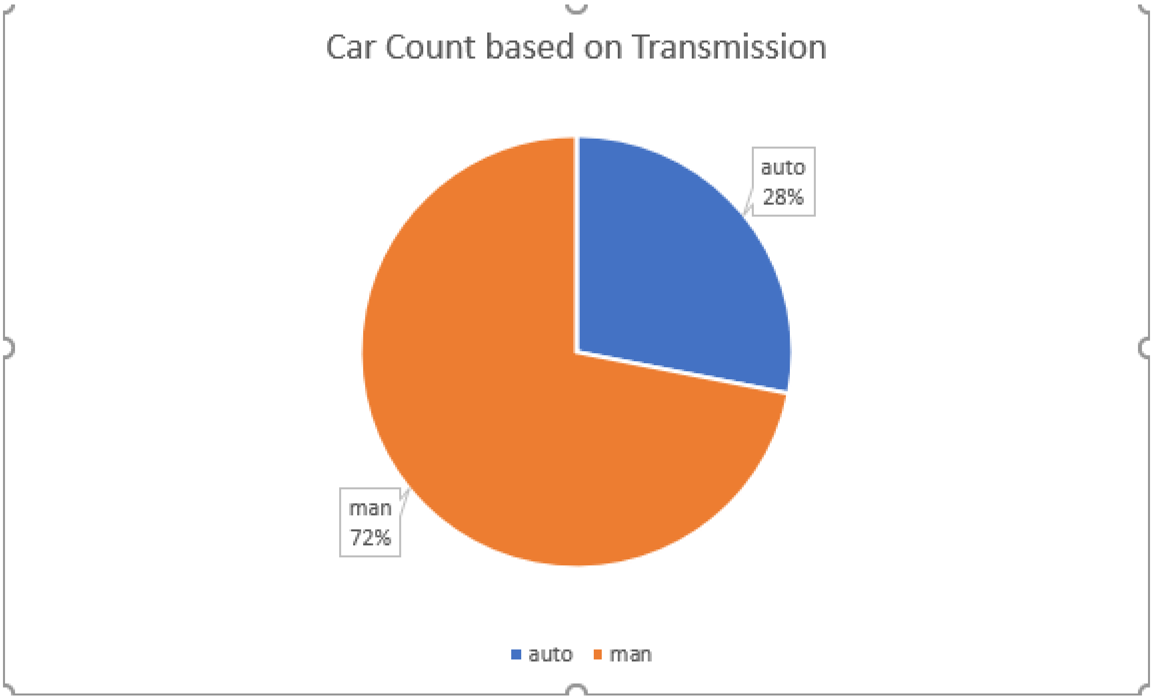
From the visual, it is evident most of the cars are sold within 3 months. The standalone majority of the cars being sold more than three months.

6) Time Period of Manufacture of cars



The car ads being listed in the portal with respect to the manufacturing year are shown in the above visual. We can observe most of the cars sold dates only few years back. From this we can observe that the trends show large number of selling of the used cars.

7) Classification of cars based on transmission



Cars usually have two types of transmission: manual and automatic. Based on the analysis for the type of transmission most sold, It is evident that manual type of transmission contributes a large chunk.

Conclusion

Upon the analysis of the used car sales data, we can observe that the trends show that the people are on a selling spree of the cars this decade compared to those of the previous decade. Also, the cars posted for sales are sold within a span of three months for most of the time.

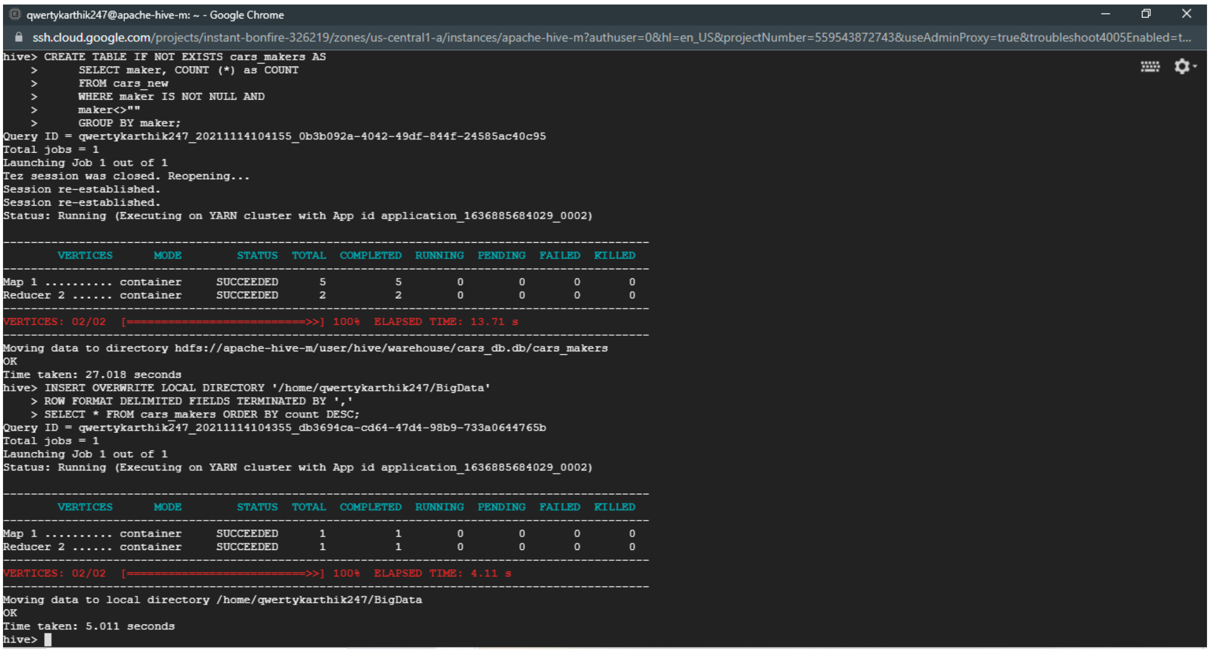
So we can safely conclude that the market for used car sales is expanding and it is a safe investment with high returns for the company given the current market trends are continuing.

Dataset Source

https://www.kaggle.com/mirosval/personal-cars-classifieds

Queries Used for Analysis

Question 1

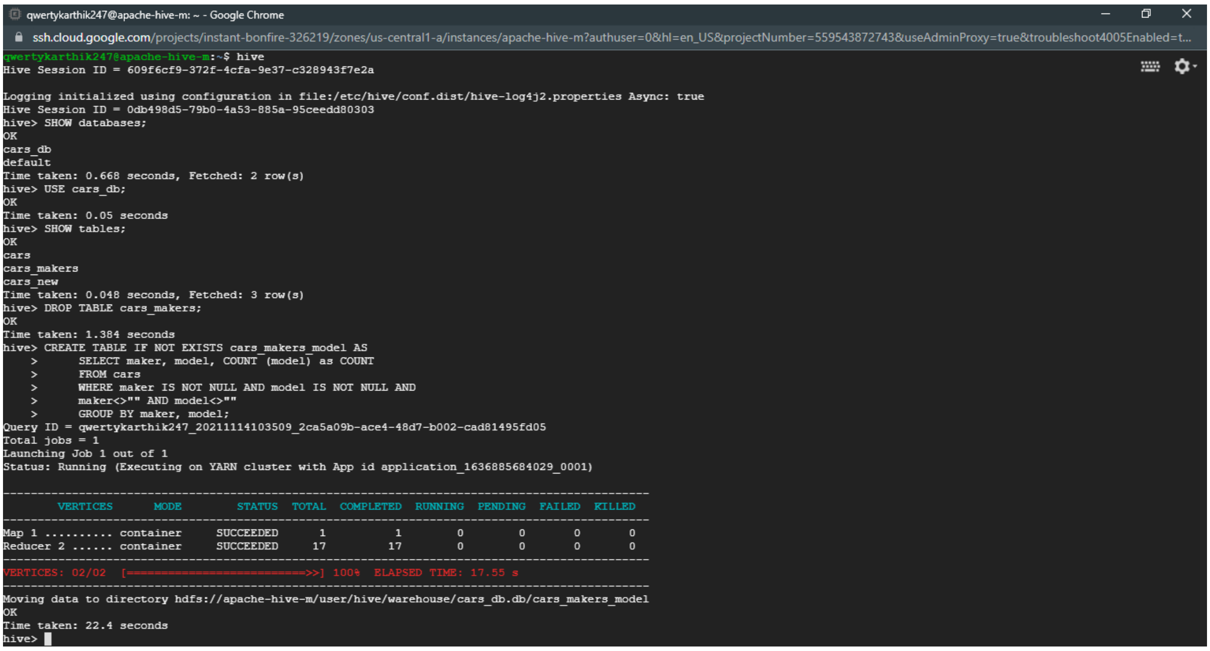

qwertykarthik247@apache-hive-m: ~ - Google Chrome
ssh.cloud.google.com/projects/instant-bonfire-326219/zones/us-central1-a/instances/apache-hive-m?authuser=0&hl=en_US&projectNumber=559543872743&useAdminProxy=true&troubleshoot4005Enabled=t...

```
hive> CREATE TABLE IF NOT EXISTS cars_makers AS
    >      SELECT maker, COUNT (*) as COUNT
    >      FROM cars_new
    >      WHERE maker IS NOT NULL AND
    >      maker<>""
    >      GROUP BY maker;
Query ID = qwertykarthik247_20211114104155_0b3b092a-4042-49df-844f-24585ac40c95
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.
Session re-established.
Status: Running (Executing on YARN cluster with App id application_1636885684029_0002)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED      5          5        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      2          2        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 13.71 s
----------------------------------------------------------------------------------------------
Moving data to directory hdfs://apache-hive-m/user/hive/warehouse/cars_db.db/cars_makers
OK
Time taken: 27.018 seconds
hive> INSERT OVERWRITE LOCAL DIRECTORY '/home/qwertykarthik247/BigData'
    > ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
    > SELECT * FROM cars_makers ORDER BY count DESC;
Query ID = qwertykarthik247_20211114104355_db3694ca-cd64-47d4-98b9-733a0644765b
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1636885684029_0002)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED      1          1        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      1          1        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 4.11 s
----------------------------------------------------------------------------------------------
Moving data to local directory /home/qwertykarthik247/BigData
OK
Time taken: 5.011 seconds
hive> 
```

Question 2

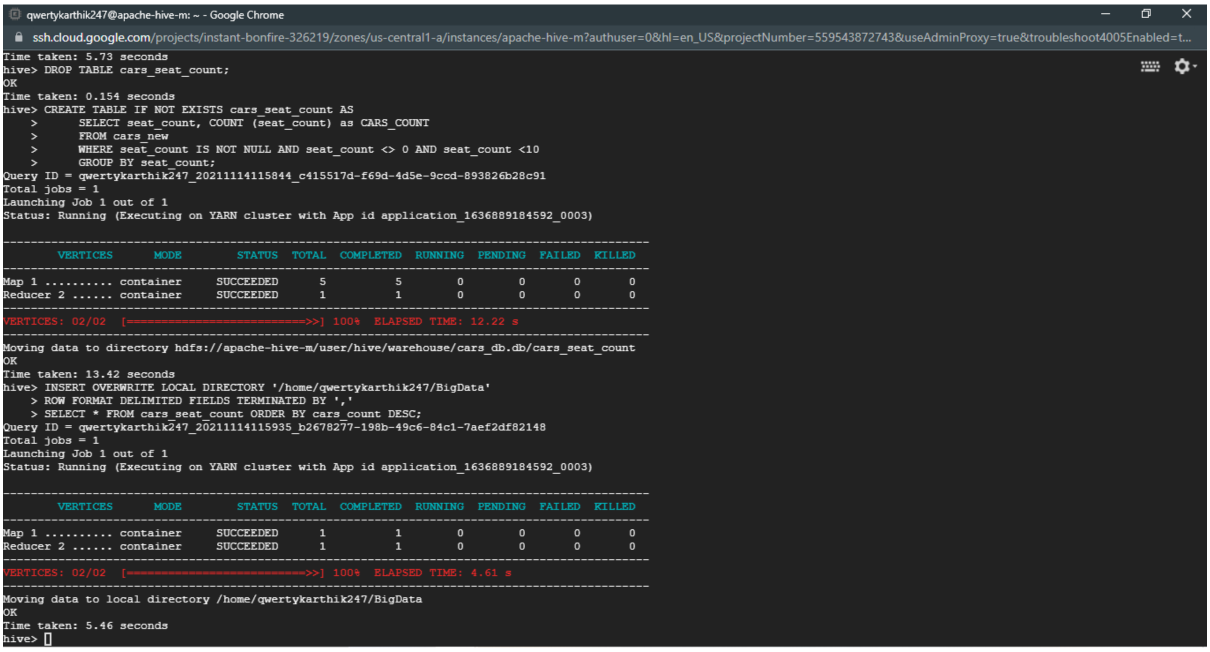

qwertykarthik247@apache-hive-m: ~ - Google Chrome
ssh.cloud.google.com/projects/instant-bonfire-326219/zones/us-central1-a/instances/apache-hive-m?authuser=0&hl=en_US&projectNumber=559543872743&useAdminProxy=true&troubleshoot4005Enabled=t...

```
qwertykarthik247@apache-hive-m:~$ hive
Hive Session ID = 609f6cf9-372f-4cfa-9e37-c328943f7e2a

Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j2.properties Async: true
Hive Session ID = 0db498d5-79b0-4a53-885a-95ceedd80303
hive> SHOW databases;
OK
cars_db
default
Time taken: 0.668 seconds, Fetched: 2 row(s)
hive> USE cars_db;
OK
Time taken: 0.05 seconds
hive> SHOW tables;
OK
cars
cars_makers
cars_new
Time taken: 0.048 seconds, Fetched: 3 row(s)
hive> DROP TABLE cars_makers;
OK
Time taken: 1.384 seconds
hive> CREATE TABLE IF NOT EXISTS cars_makers_model AS
    >      SELECT maker, model, COUNT (model) as COUNT
    >      FROM cars
    >      WHERE maker IS NOT NULL AND model IS NOT NULL AND
    >      maker<>"" AND model<>""
    >      GROUP BY maker, model;
Query ID = qwertykarthik247_20211114103509_2ca5a09b-ace4-48d7-b002-cad81495fd05
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1636885684029_0001)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED      1          1        0        0       0       0
Reducer 2 ...... container     SUCCEEDED     17         17        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 17.55 s
----------------------------------------------------------------------------------------------
Moving data to directory hdfs://apache-hive-m/user/hive/warehouse/cars_db.db/cars_makers_model
OK
Time taken: 22.4 seconds
hive> 
```

Question 3

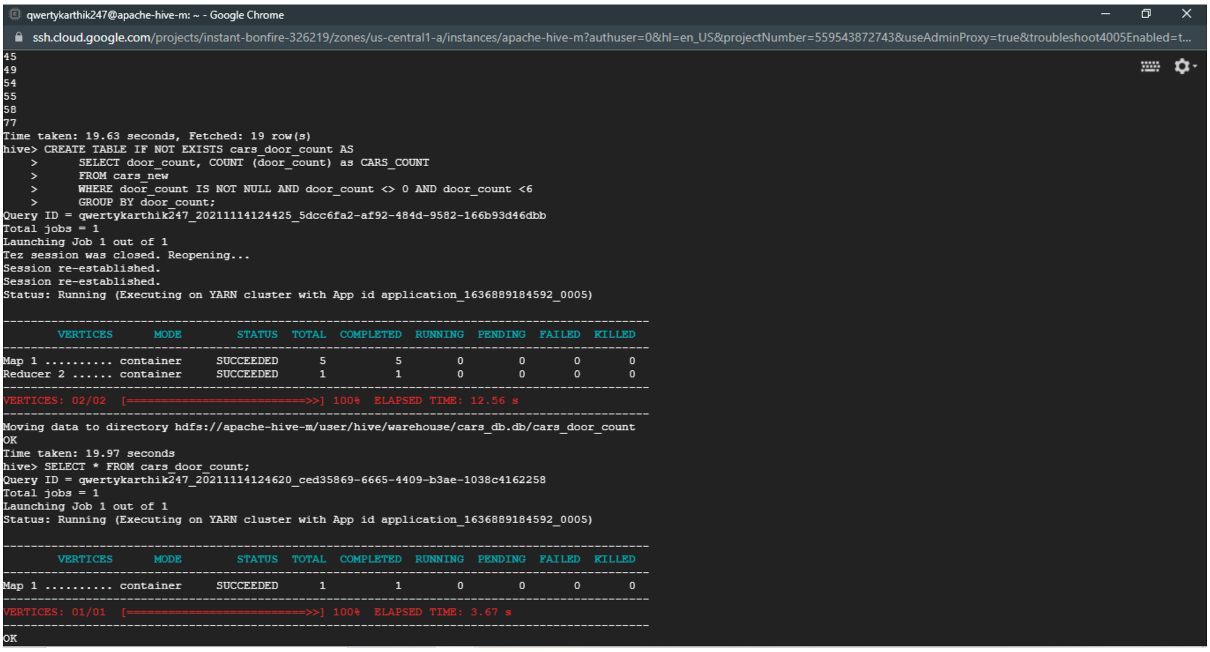

qwertykarthik247@apache-hive-m: ~ - Google Chrome
ssh.cloud.google.com/projects/instant-bonfire-326219/zones/us-central1-a/instances/apache-hive-m?authuser=0&hl=en_US&projectNumber=559543872743&useAdminProxy=true&troubleshoot4005Enabled=t...

```
Time taken: 5.73 seconds
hive> DROP TABLE cars_seat_count;
OK
Time taken: 0.154 seconds
hive> CREATE TABLE IF NOT EXISTS cars_seat_count AS
    >      SELECT seat_count, COUNT (seat_count) as CARS_COUNT
    >      FROM cars_new
    >      WHERE seat_count IS NOT NULL AND seat_count <> 0 AND seat_count <10
    >      GROUP BY seat_count;
Query ID = qwertykarthik247_20211114115844_c415517d-f69d-4d5e-9ccd-893826b28c91
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1636889184592_0003)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED      5          5        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      1          1        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 12.22 s
----------------------------------------------------------------------------------------------
Moving data to directory hdfs://apache-hive-m/user/hive/warehouse/cars_db.db/cars_seat_count
OK
Time taken: 13.42 seconds
hive> INSERT OVERWRITE LOCAL DIRECTORY '/home/qwertykarthik247/BigData'
    > ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
    > SELECT * FROM cars_seat_count ORDER BY cars_count DESC;
Query ID = qwertykarthik247_20211114115935_b2678277-198b-49c6-84c1-7aef2df82148
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1636889184592_0003)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED      1          1        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      1          1        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 4.61 s
----------------------------------------------------------------------------------------------
Moving data to local directory /home/qwertykarthik247/BigData
OK
Time taken: 5.46 seconds
hive> 
```

Question 4

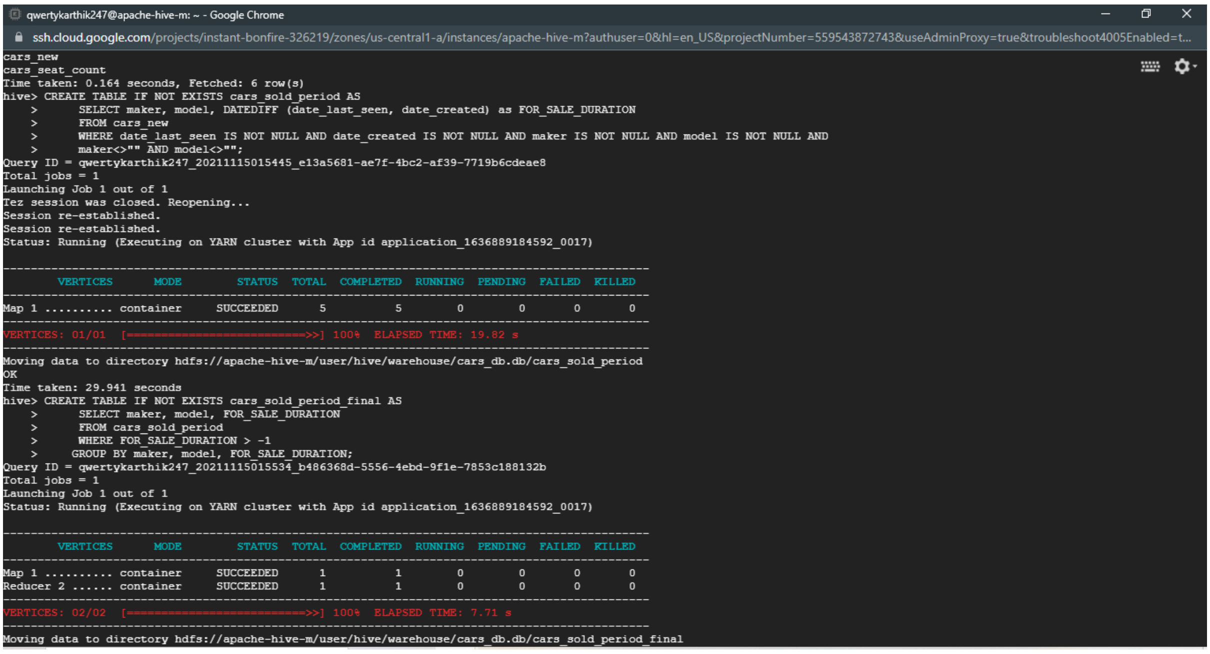

qwertykarthik247@apache-hive-m: ~ - Google Chrome
ssh.cloud.google.com/projects/instant-bonfire-326219/zones/us-central1-a/instances/apache-hive-m?authuser=0&hl=en_US&projectNumber=559543872743&useAdminProxy=true&troubleshoot4005Enabled=t...

```
45
49
54
55
58
77
Time taken: 19.63 seconds, Fetched: 19 row(s)
hive> CREATE TABLE IF NOT EXISTS cars_door_count AS
    >      SELECT door_count, COUNT (door_count) as CARS_COUNT
    >      FROM cars_new
    >      WHERE door_count IS NOT NULL AND door_count <> 0 AND door_count <6
    >      GROUP BY door_count;
Query ID = qwertykarthik247_20211114124425_5dcc6fa2-af92-484d-9582-166b93d46dbb
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.
Session re-established.
Status: Running (Executing on YARN cluster with App id application_1636889184592_0005)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED      5          5        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      1          1        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 12.56 s
----------------------------------------------------------------------------------------------
Moving data to directory hdfs://apache-hive-m/user/hive/warehouse/cars_db.db/cars_door_count
OK
Time taken: 19.97 seconds
hive> SELECT * FROM cars_door_count;
Query ID = qwertykarthik247_20211114124620_ced35869-6665-4409-b3ae-1038c4162258
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1636889184592_0005)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED      1          1        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 01/01  [==========================>>] 100%  ELAPSED TIME: 3.67 s
----------------------------------------------------------------------------------------------
OK
```

Question 5

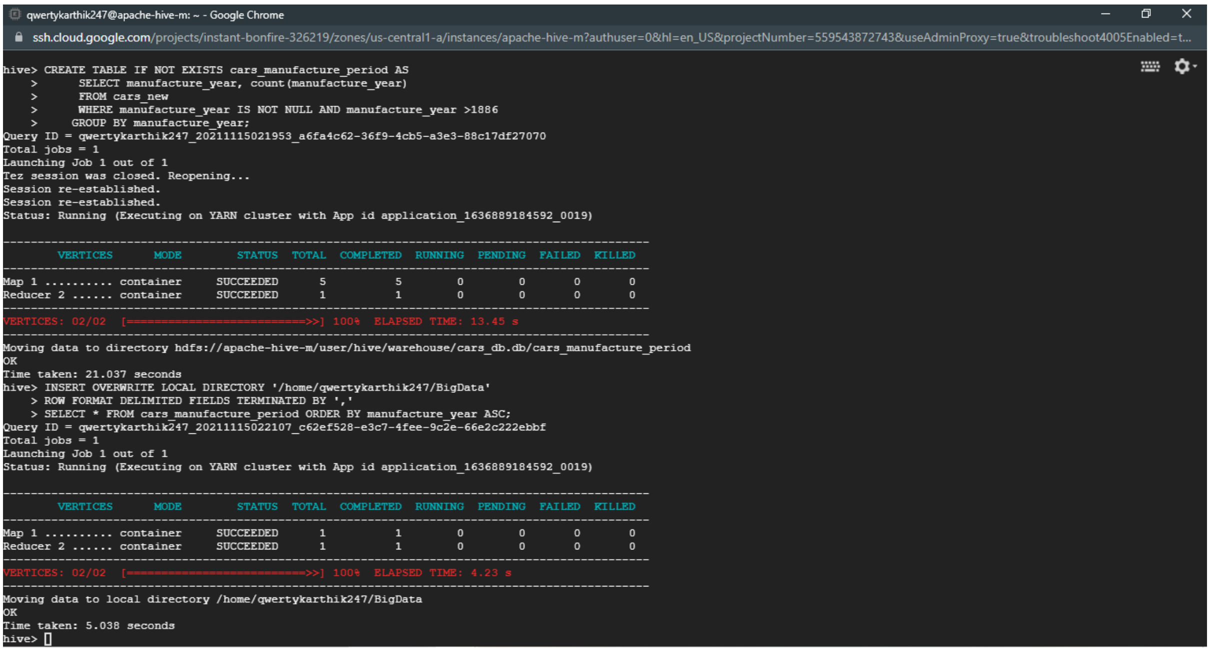

qwertykarthik247@apache-hive-m: ~ - Google Chrome
ssh.cloud.google.com/projects/instant-bonfire-326219/zones/us-central1-a/instances/apache-hive-m?authuser=0&hl=en_US&projectNumber=559543872743&useAdminProxy=true&troubleshoot4005Enabled=t...

```
cars_new
cars_seat_count
Time taken: 0.164 seconds, Fetched: 6 row(s)
hive> CREATE TABLE IF NOT EXISTS cars_sold_period AS
    >       SELECT maker, model, DATEDIFF (date_last_seen, date_created) as FOR_SALE_DURATION
    >       FROM cars_new
    >       WHERE date_last_seen IS NOT NULL AND date_created IS NOT NULL AND maker IS NOT NULL AND model IS NOT NULL AND
    >       maker<>"" AND model<>"";
Query ID = qwertykarthik247_20211115015445_e13a5681-ae7f-4bc2-af39-7719b6cdeae8
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.
Session re-established.
Status: Running (Executing on YARN cluster with App id application_1636889184592_0017)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED      5          5        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 01/01  [==========================>>] 100%  ELAPSED TIME: 19.82 s
----------------------------------------------------------------------------------------------
Moving data to directory hdfs://apache-hive-m/user/hive/warehouse/cars_db.db/cars_sold_period
OK
Time taken: 29.941 seconds
hive> CREATE TABLE IF NOT EXISTS cars_sold_period_final AS
    >       SELECT maker, model, FOR_SALE_DURATION
    >       FROM cars_sold_period
    >       WHERE FOR_SALE_DURATION > -1
    >      GROUP BY maker, model, FOR_SALE_DURATION;
Query ID = qwertykarthik247_20211115015534_b486368d-5556-4ebd-9f1e-7853c188132b
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1636889184592_0017)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED      1          1        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      1          1        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 7.71 s
----------------------------------------------------------------------------------------------
Moving data to directory hdfs://apache-hive-m/user/hive/warehouse/cars_db.db/cars_sold_period_final
```

Question 6

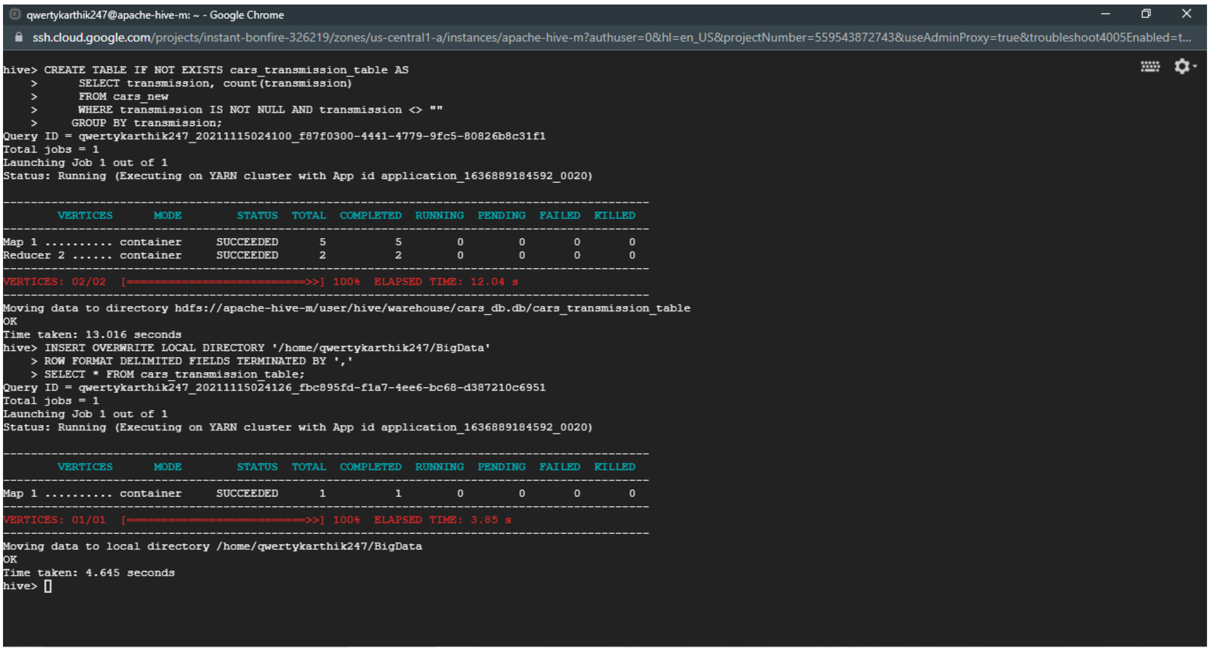

qwertykarthik247@apache-hive-m: ~ - Google Chrome
ssh.cloud.google.com/projects/instant-bonfire-326219/zones/us-central1-a/instances/apache-hive-m?authuser=0&hl=en_US&projectNumber=559543872743&useAdminProxy=true&troubleshoot4005Enabled=t...

```
hive> CREATE TABLE IF NOT EXISTS cars_manufacture_period AS
    >      SELECT manufacture_year, count(manufacture_year)
    >      FROM cars_new
    >      WHERE manufacture_year IS NOT NULL AND manufacture_year >1886
    >     GROUP BY manufacture_year;
Query ID = qwertykarthik247_20211115021953_a6fa4c62-36f9-4cb5-a3e3-88c17df27070
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.
Session re-established.
Status: Running (Executing on YARN cluster with App id application_1636889184592_0019)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED      5          5        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      1          1        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 13.45 s
----------------------------------------------------------------------------------------------
Moving data to directory hdfs://apache-hive-m/user/hive/warehouse/cars_db.db/cars_manufacture_period
OK
Time taken: 21.037 seconds
hive> INSERT OVERWRITE LOCAL DIRECTORY '/home/qwertykarthik247/BigData'
    > ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
    > SELECT * FROM cars_manufacture_period ORDER BY manufacture_year ASC;
Query ID = qwertykarthik247_20211115022107_c62ef528-e3c7-4fee-9c2e-66e2c222ebbf
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1636889184592_0019)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED      1          1        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      1          1        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 4.23 s
----------------------------------------------------------------------------------------------
Moving data to local directory /home/qwertykarthik247/BigData
OK
Time taken: 5.038 seconds
hive> 
```

Question 7


qwertykarthik247@apache-hive-m: ~ - Google Chrome
ssh.cloud.google.com/projects/instant-bonfire-326219/zones/us-central1-a/instances/apache-hive-m?authuser=0&hl=en_US&projectNumber=559543872743&useAdminProxy=true&troubleshoot4005Enabled=t...

```
hive> CREATE TABLE IF NOT EXISTS cars_transmission_table AS
    >      SELECT transmission, count(transmission)
    >      FROM cars_new
    >      WHERE transmission IS NOT NULL AND transmission <> ""
    >     GROUP BY transmission;
Query ID = qwertykarthik247_20211115024100_f87f0300-4441-4779-9fc5-80826b8c31f1
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1636889184592_0020)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED      5          5        0        0       0       0
Reducer 2 ...... container     SUCCEEDED      2          2        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 02/02  [==========================>>] 100%  ELAPSED TIME: 12.04 s
----------------------------------------------------------------------------------------------
Moving data to directory hdfs://apache-hive-m/user/hive/warehouse/cars_db.db/cars_transmission_table
OK
Time taken: 13.016 seconds
hive> INSERT OVERWRITE LOCAL DIRECTORY '/home/qwertykarthik247/BigData'
    > ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
    > SELECT * FROM cars_transmission_table;
Query ID = qwertykarthik247_20211115024126_fbc895fd-f1a7-4ee6-bc68-d387210c6951
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1636889184592_0020)

----------------------------------------------------------------------------------------------
        VERTICES      MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
----------------------------------------------------------------------------------------------
Map 1 .......... container     SUCCEEDED      1          1        0        0       0       0
----------------------------------------------------------------------------------------------
VERTICES: 01/01  [==========================>>] 100%  ELAPSED TIME: 3.85 s
----------------------------------------------------------------------------------------------
Moving data to local directory /home/qwertykarthik247/BigData
OK
Time taken: 4.645 seconds
hive> 
```