QUERY1.

```
hive> CREATE TABLE telecom_data (
    > customerID STRING,
    > gender STRING,
    > SeniorCitizen INT,
    > Partner STRING,
    > Dependents STRING,
    > tenure INT,
    > PhoneService STRING,
    > MultipleLines STRING,
    > InternetService STRING,
    > OnlineSecurity STRING,
    > OnlineBackup STRING,
    > DeviceProtection STRING,
    > TechSupport STRING,
    > StreamingTV STRING,
    > StreamingMovies STRING,
    > Contract STRING,
    > PaperlessBilling STRING,
    > PaymentMethod STRING,
    > MonthlyCharges FLOAT,
    > TotalCharges FLOAT,
    > Churn STRING
    > )
    > ROW FORMAT DELIMITED
    > FIELDS TERMINATED BY ','
    > STORED AS TEXTFILE;
OK
Time taken: 1.732 seconds
hive> set hive.cli.print.header=true;
hive> load data local inpath 'file:////home/darshanas0510/Telecom_customer_churn_data.csv' into table telecom_data;
Loading data to table default.telecom_data
OK
Time taken: 0.534 seconds
```

QUERY2.

```
hive> SELECT * FROM telecom_data LIMIT 10;
OK
telecom_data.customerid telecom_data.gender     telecom_data.seniorcitizen      telecom_data.partner    telecom_data.d
ependents       telecom_data.tenure     telecom_data.phoneservice       telecom_data.multiplelines      telecom_data.i
nternetservice  telecom_data.onlinesecurity     telecom_data.onlinebackup       telecom_data.deviceprotection   teleco
m_data.techsupport      telecom_data.streamingtv        telecom_data.streamingmovies    telecom_data.contract   teleco
m_data.paperlessbilling telecom_data.paymentmethod      telecom_data.monthlycharges     telecom_data.totalcharges    t
elecom_data.churn
7590-VHVEG      Female  0       Yes     No      1       No      No phone service        DSL     No      Yes     No   N
o       No      No      Month-to-month  Yes     Electronic check        29.85   29.85   No
5575-GNVDE      Male    0       No      No      34      Yes     No      DSL     Yes     No      Yes     No      No   N
o       One year        No      Mailed check    56.95   1889.5  No
3668-QPYBK      Male    0       No      No      2       Yes     No      DSL     Yes     Yes     No      No      No   N
o       Month-to-month  Yes     Mailed check    53.85   108.15  Yes
7795-CFOCW      Male    0       No      No      45      No      No phone service        DSL     Yes     No      Yes  Y
es      No      One year        No      Bank transfer (automatic)       42.3    1840.75 No
9237-HQITU      Female  0       No      No      2       Yes     No      Fiber optic     No      No      No      No   N
o       No      Month-to-month  Yes     Electronic check        70.7    151.65  Yes
9305-CDSKC      Female  0       No      No      8       Yes     Yes     Fiber optic     No      No      Yes     No   Y
es      Yes     Month-to-month  Yes     Electronic check        99.65   820.5   Yes
1452-KIOVK      Male    0       No      Yes     22      Yes     Yes     Fiber optic     No      Yes     No      No   Y
es      No      Month-to-month  Yes     Credit card (automatic) 89.1    1949.4  No
6713-OKOMC      Female  0       No      No      10      No      No phone service        DSL     Yes     No      No   N
o       No      No      Month-to-month  No      Mailed check    29.75   301.9   No
7892-POOKP      Female  0       Yes     No      28      Yes     Yes     Fiber optic     No      No      Yes     Yes  Y
es      Yes     Month-to-month  Yes     Electronic check        104.8   3046.05 Yes
6388-TABGU      Male    0       No      Yes     62      Yes     No      DSL     Yes     Yes     No      No      No   N
o       One year        No      Bank transfer (automatic)       56.15   3487.95 No
Time taken: 0.253 seconds, Fetched: 10 row(s)
```

QUERY3.

```
hive> SELECT COUNT(*) FROM telecom_data;
Query ID = darshanas0510_20231207070750_553e8c59-5580-4032-a476-8e7267c7cced
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1701930939803_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: 10.43 s
--------------------------------------------------------------------------------------------
OK
_c0
7043
Time taken: 13.292 seconds, Fetched: 1 row(s)
```

QUERY4.

```
hive> SELECT COUNT(*) FROM telecom_data WHERE Churn = 'Yes';
Query ID = darshanas0510_20231207070835_6f1517f6-fb16-48f4-9a41-8543be0d1be3
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1701930939803_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: 9.48 s
--------------------------------------------------------------------------------------
OK
_c0
1869
Time taken: 10.283 seconds, Fetched: 1 row(s)
```

QUERY5.

```
hive> SELECT gender, SeniorCitizen, COUNT(*)
    > FROM telecom_data
    > GROUP BY gender, SeniorCitizen;
Query ID = darshanas0510_20231207070943_9c50345f-22e2-490f-bc96-3e7f8ecc0f93
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1701930939803_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: 9.24 s
--------------------------------------------------------------------------------------
OK
gender   seniorcitizen   _c2
Female   0        2920
Female   1        568
Male     0        2981
Male     1        574
Time taken: 9.864 seconds, Fetched: 4 row(s)
```

QUERY6.

```
hive> SELECT SUM(TotalCharges)
    > FROM telecom_data
    > WHERE Churn = 'Yes';
Query ID = darshanas0510_20231207071024_ae484a48-0e22-420d-a054-9b6802c15eae
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1701930939803_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: 10.02 s
--------------------------------------------------------------------------------------
OK
_c0
2862926.896551132
Time taken: 10.577 seconds, Fetched: 1 row(s)
```

QUERY7.

```
hive> SELECT Contract, COUNT(*)
    > FROM telecom_data
    > WHERE Churn = 'Yes'
    > GROUP BY Contract;
Query ID = darshanas0510_20231207071132_8458046e-d05c-48b5-a71d-bed1e8d4b3f4
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1701930939803_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: 10.35 s
--------------------------------------------------------------------------------
OK
contract       _c1
Month-to-month 1655
One year        166
Two year         48
Time taken: 10.898 seconds, Fetched: 3 row(s)
```

QUERY8.

```
hive> SELECT Churn, AVG(MonthlyCharges)
    > FROM telecom_data
    > GROUP BY Churn;
Query ID = darshanas0510_20231207071226_de2c94ef-dffc-4a47-a3be-d1f308edd0a3
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1701930939803_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: 10.12 s
--------------------------------------------------------------------------------
OK
churn   _c1
No      61.26512372083631
Yes     74.44133227059011
Time taken: 10.697 seconds, Fetched: 2 row(s)
```

QUERY9.

```
hive> SELECT MAX(tenure), MIN(tenure), AVG(tenure)
    > FROM telecom_data;
Query ID = darshanas0510_20231207071304_1069e2aa-d3d0-4d07-a1fd-be51b17b50aa
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1701930939803_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: 9.50 s
--------------------------------------------------------------------------------
OK
_c0     _c1     _c2
72      0       32.37114865824223
Time taken: 10.139 seconds, Fetched: 1 row(s)
```

QUERY10.

```
hive> SELECT PaymentMethod, COUNT(*)
    > FROM telecom_data
    > GROUP BY PaymentMethod
    > ORDER BY COUNT(*) DESC LIMIT 1;
Query ID = darshanas0510_20231207071353_f6a4606a-c8e7-4239-9aeb-89d51e08190e
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1701930939803_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
Reducer 3 ...... container       SUCCEEDED      1         1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 03/03  [==========================>>] 100%  ELAPSED TIME: 9.48 s
--------------------------------------------------------------------------------
OK
paymentmethod    _c1
Electronic check         2365
Time taken: 10.046 seconds, Fetched: 1 row(s)
```

QUERY11.

```
hive> SELECT PaperlessBilling, COUNT(*)
    > FROM telecom_data
    > WHERE Churn = 'Yes'
    > GROUP BY PaperlessBilling;
Query ID = darshanas0510_20231207071430_eadf8301-f47a-4063-80cb-b06d9779018c
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1701930939803_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: 10.05 s
--------------------------------------------------------------------------------
OK
paperlessbilling        _c1
No      469
Yes     1400
Time taken: 10.468 seconds, Fetched: 2 row(s)
```

QUERY12.

```
hive> INSERT OVERWRITE TABLE telecom_data_partitioned
    > PARTITION(Contract)
    > SELECT * FROM telecom_data;
Query ID = darshanas0510_20231207071844_90886067-7212-4c05-82c5-0f174b6e9c4c
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1701930939803_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: 10.80 s
--------------------------------------------------------------------------------
Loading data to table default.telecom_data_partitioned partition (contract=null)

Loaded : 2/2 partitions.
        Time taken to load dynamic partitions: 0.548 seconds
        Time taken for adding to write entity : 0.002 seconds
OK
_col0   _col1   _col2   _col3   _col4   _col5   _col6   _col7   _col8   _col9   _col10  _col11  _col12  _col13  _col14
_col15  _col16  _col17  _col18  _col19  _col20
Time taken: 13.473 seconds
```

QUERY13.

```
hive> SELECT Contract, COUNT(*)
    > FROM telecom_data_partitioned
    > WHERE Churn = 'Yes'
    > GROUP BY Contract;
Query ID = darshanas0510_20231207072206_828b35c0-46a5-48e2-b335-64a96ecde08a
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1701930939803_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: 10.06 s
----------------------------------------------------------------------------------------
OK
contract         _c1
Time taken: 10.613 seconds
```

QUERY14.

```
hive> SELECT Contract, MAX(tenure)
    > FROM telecom_data_partitioned
    > GROUP BY Contract;
Query ID = darshanas0510_20231207073250_a9361ae6-e640-496f-9a22-cb2fc31b2673
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_1701930939803_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: 10.86 s
----------------------------------------------------------------------------------------
OK
contract         _c1
No       72
Yes      72
Time taken: 23.77 seconds, Fetched: 2 row(s)
```

QUERY15.

```
hive> CREATE TABLE telecom_data_bucketed (
    > customerID STRING,
    > gender STRING,
    > SeniorCitizen INT,
    > Partner STRING,
    > Dependents STRING,
    > tenure INT,
    > PhoneService STRING,
    > MultipleLines STRING,
    > InternetService STRING,
    > OnlineSecurity STRING,
    > OnlineBackup STRING,
    > DeviceProtection STRING,
    > TechSupport STRING,
    > StreamingTV STRING,
    > StreamingMovies STRING,
    > Contract STRING,
    > PaperlessBilling STRING,
    > PaymentMethod STRING,
    > MonthlyCharges FLOAT,
    > TotalCharges FLOAT,
    > Churn STRING
    > )
    > CLUSTERED BY (tenure) INTO 6 BUCKETS;
OK
Time taken: 0.148 seconds
```

QUERY16.

```
hive> INSERT OVERWRITE TABLE telecom_data_bucketed
    > SELECT * FROM telecom_data;
Query ID = darshanas0510_20231207073826_ea2b169c-3320-4af7-9251-535323d21855
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1701930939803_0003)

--------------------------------------------------------------------------------
        VERTICES        MODE        STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
--------------------------------------------------------------------------------
Map 1 .......... container   SUCCEEDED     1         1        0        0       0       0
Reducer 2 ...... container   SUCCEEDED     6         6        0        0       0       0
Reducer 3 ...... container   SUCCEEDED     1         1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 03/03  [==========================>>] 100%  ELAPSED TIME: 21.22 s
--------------------------------------------------------------------------------
Loading data to table default.telecom_data_bucketed
OK
telecom_data.customerid telecom_data.gender     telecom_data.seniorcitizen      telecom_data.partner    telecom_data.d
ependents       telecom_data.tenure     telecom_data.phoneservice       telecom_data.multiplelines      telecom_data.i
nternetservice  telecom_data.onlinesecurity     telecom_data.onlinebackup       telecom_data.deviceprotection   teleco
m_data.techsupport      telecom_data.streamingtv        telecom_data.streamingmovies    telecom_data.contract   teleco
m_data.paperlessbilling telecom_data.paymentmethod      telecom_data.monthlycharges     telecom_data.totalcharges    t
elecom_data.churn
Time taken: 22.418 seconds
```

QUERY17.

```
hive> SELECT 'Bucket 1' AS Bucket, MAX(TotalCharges) as
    > MaxCharge
    > FROM telecom_data_bucketed
    > WHERE tenure BETWEEN 1 AND 10
    > UNION ALL
    > SELECT 'Bucket 2' AS Bucket, MAX(TotalCharges) as
    > MaxCharge
    > FROM telecom_data_bucketed
    > WHERE tenure BETWEEN 11 AND 20
    > UNION ALL
    > SELECT 'Bucket 3' AS Bucket, MAX(TotalCharges) as
    > MaxCharge
    > FROM telecom_data_bucketed
    > WHERE tenure BETWEEN 21 AND 30
    > UNION ALL
    > SELECT 'Bucket 4' AS Bucket, MAX(TotalCharges) as
    > MaxCharge
    > FROM telecom_data_bucketed
    > WHERE tenure BETWEEN 31 AND 40
    > UNION ALL
    > SELECT 'Bucket 5' AS Bucket, MAX(TotalCharges) as
    > MaxCharge
    > FROM telecom_data_bucketed
    > WHERE tenure BETWEEN 41 AND 50
    > UNION ALL
    > SELECT 'Bucket 6' AS Bucket, MAX(TotalCharges) as
    > MaxCharge
    > FROM telecom_data_bucketed
    > WHERE tenure BETWEEN 51 AND 60;
Query ID = darshanas0510_20231207080311_9424c2f9-b2c7-4adb-80df-173879bed9ff
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1701935912682_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
Map 4 .......... container   SUCCEEDED     1         1        0        0       0       0
Reducer 5 ...... container   SUCCEEDED     1         1        0        0       0       0
Map 6 .......... container   SUCCEEDED     1         1        0        0       0       0
Reducer 7 ...... container   SUCCEEDED     1         1        0        0       0       0
Map 8 .......... container   SUCCEEDED     1         1        0        0       0       0
Reducer 9 ...... container   SUCCEEDED     1         1        0        0       0       0
Map 10 ........ container    SUCCEEDED     1         1        0        0       0       0
Reducer 11 ..... container   SUCCEEDED     1         1        0        0       0       0
Map 12 ........ container    SUCCEEDED     1         1        0        0       0       0
Reducer 13 ..... container   SUCCEEDED     1         1        0        0       0       0
--------------------------------------------------------------------------------
VERTICES: 12/12  [==========================>>] 100%  ELAPSED TIME: 29.81 s
--------------------------------------------------------------------------------
OK
Bucket 2        2239.65
Bucket 1        1081.45
Bucket 3        3327.05
Bucket 4        4469.1
Bucket 5        5791.1
Bucket 6        7049.5
Time taken: 49.869 seconds, Fetched: 6 row(s)
```