IIT- M Advanced Certificate Program in Machine Learning and Cloud- upGrad Capstone Project

User Demographics Prediction using Telecom dataset

SQL Task Commands

Authors :

Mukul Pahawa

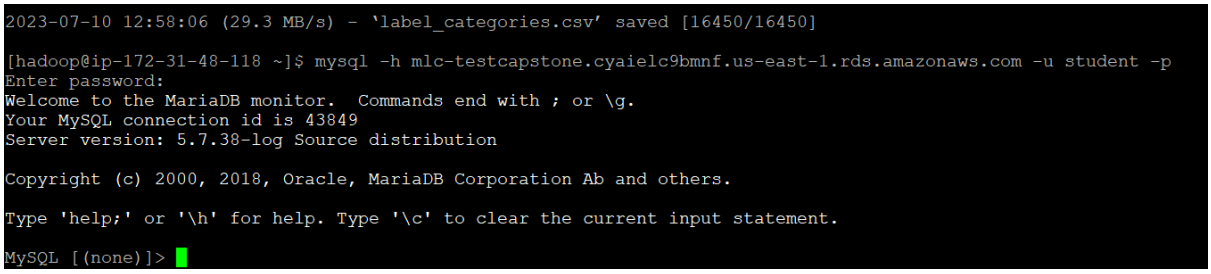
Mitesh

SQL Tasks

Connecting to RDS instance

```
RDS Information:
Endpoint/Hostname: mlc-testcapstone.cyaielc9bmnf.us-east-1.rds.amazonaws.co
m
username: student
password: STUDENT123
db: mlctest
```

mysql -h mlc-testcapstone.cyaielc9bmnf.us-east-1.rds.amazonaws.com -u student -p



Checking databases
Using - show databases;

```
2023-07-10 12:58:06 (29.3 MB/s) - 'label_categories.csv' saved [16450/16450]

[hadoop@ip-172-31-48-118 ~]$ mysql -h mlc-testcapstone.cyaielc9bmnf.us-east-1.rds.amazonaws.com -u student -p
Enter password:
Welcome to the MariaDB monitor.  Commands end with ; or \g.
Your MySQL connection id is 43849
Server version: 5.7.38-log Source distribution

Copyright (c) 2000, 2018, Oracle, MariaDB Corporation Ab and others.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

MySQL [(none)]> show databases;
+--------------------+
| Database           |
+--------------------+
| information_schema |
| mlctest            |
+--------------------+
```

Using the database

use mlctest;

Initial Verification(Sanity)
desc app_events;
desc train;
desc brand_device;
desc events;

```
Database changed
MySQL [mlctest]> desc app_events
    -> ;
+--------------+--------------+------+-----+---------+-------+
| Field        | Type         | Null | Key | Default | Extra |
+--------------+--------------+------+-----+---------+-------+
| event_id     | varchar(255) | YES  |     | NULL    |       |
| app_id       | varchar(255) | YES  |     | NULL    |       |
| is_installed | int(11)      | YES  |     | NULL    |       |
| is_active    | int(11)      | YES  |     | NULL    |       |
+--------------+--------------+------+-----+---------+-------+
4 rows in set (0.00 sec)

MySQL [mlctest]> desc train;
+-------------+--------------+------+-----+---------+-------+
| Field       | Type         | Null | Key | Default | Extra |
+-------------+--------------+------+-----+---------+-------+
| device_id   | varchar(255) | YES  |     | NULL    |       |
| gender      | varchar(255) | YES  |     | NULL    |       |
| age         | int(11)      | YES  |     | NULL    |       |
| group_train | varchar(255) | YES  |     | NULL    |       |
+-------------+--------------+------+-----+---------+-------+
4 rows in set (0.00 sec)

MySQL [mlctest]> desc brand_device;
+--------------+--------------+------+-----+---------+-------+
| Field        | Type         | Null | Key | Default | Extra |
+--------------+--------------+------+-----+---------+-------+
| device_id    | varchar(255) | YES  |     | NULL    |       |
| phone_brand  | varchar(255) | YES  |     | NULL    |       |
| device_model | varchar(255) | YES  |     | NULL    |       |
+--------------+--------------+------+-----+---------+-------+
3 rows in set (0.01 sec)

MySQL [mlctest]> desc events;
+-----------+--------------+------+-----+---------+-------+
| Field     | Type         | Null | Key | Default | Extra |
+-----------+--------------+------+-----+---------+-------+
| event_id  | int(11)      | YES  |     | NULL    |       |
| device_id | varchar(255) | YES  |     | NULL    |       |
| timestamp | datetime     | YES  |     | NULL    |       |
| longitude | varchar(225) | YES  |     | NULL    |       |
| latitude  | varchar(225) | YES  |     | NULL    |       |
```

Count Analysis
==========
select count(*) from app_events;
select count(*) from brand_device;
select count(*) from events;
select count(*) from train;

```
MySQL [mlctest]> select count(*) from app_events;
+----------+
| count(*) |
+----------+
| 32473067 |
+----------+
1 row in set (47.58 sec)

MySQL [mlctest]> select count(*) from brand_device;
+----------+
| count(*) |
+----------+
|   187245 |
+----------+
1 row in set (0.33 sec)

MySQL [mlctest]> select count(*) from events;
+----------+
| count(*) |
+----------+
|  3252950 |
+----------+
1 row in set (5.18 sec)

MySQL [mlctest]> select count(*) from train;
+----------+
| count(*) |
+----------+
|    74645 |
+----------+
1 row in set (0.16 sec)

MySQL [mlctest]>
```

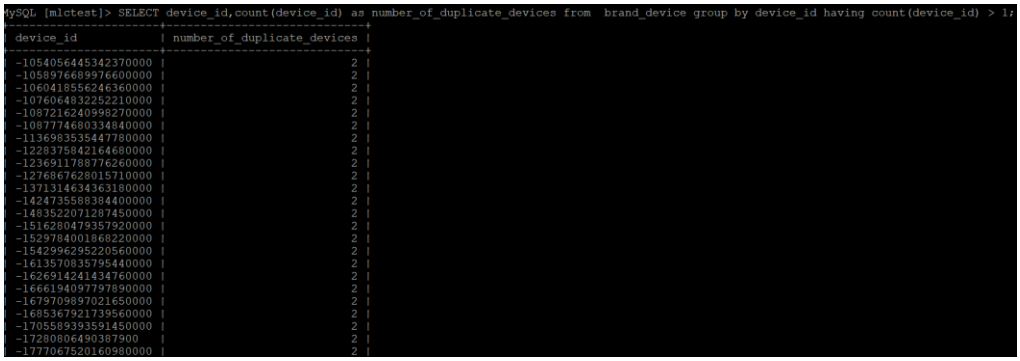
SQL Tasks

==========

1.select count(distinct(device_id)) from train;

```
MySQL [mlctest]> select count(distinct(device_id))  from train;
+---------------------------+
| count(distinct(device_id)) |
+---------------------------+
|                     74645 |
+---------------------------+
```

2 . SELECT device_id,count(device_id) as number_of_duplicate_devices from brand_device group by device_id having count(device_id) > 1;



```
MySQL [mlctest]> SELECT device_id,count(device_id) as number_of_duplicate_devices from  brand_device group by device_id having count(device_id) > 1;
+---------------------+-----------------------------+
| device_id           | number_of_duplicate_devices |
+---------------------+-----------------------------+
| -1054056445342370000 |                           2 |
| -1058976689976600000 |                           2 |
| -1060418556246360000 |                           2 |
| -1076064832252210000 |                           2 |
| -1087216240998270000 |                           2 |
| -1087774680334840000 |                           2 |
| -1136983535447780000 |                           2 |
| -1228375842164680000 |                           2 |
| -1236911788776260000 |                           2 |
| -1276867628015710000 |                           2 |
| -1371314634363180000 |                           2 |
| -1424735588384400000 |                           2 |
| -1483522071287450000 |                           2 |
| -1516280479357920000 |                           2 |
| -1529784001868220000 |                           2 |
| -1542996295220560000 |                           2 |
| -1613570835795440000 |                           2 |
| -1626914241434760000 |                           2 |
| -1666194097797890000 |                           2 |
| -1679709897021650000 |                           2 |
| -1685367921739560000 |                           2 |
| -1705589393591450000 |                           2 |
| -17280806490387900   |                           2 |
| -1777067520160980000 |                           2 |
```

3.select count(distinct(phone_brand)) from brand_device;

```
MySQL [mlctest]> select count(distinct(phone_brand)) from brand_device;
+--------------------------------+
| count(distinct(phone_brand)) |
+--------------------------------+
|                             97 |
+--------------------------------+
1 row in set (0.10 sec)
```

4.select count(device_id) from events where longitude = 0 and latitude = 0;

```
MySQL [mlctest]> SELECT count(device_id) FROM events WHERE longitude = 0 and latitude = 0;
+------------------+
| count(device_id) |
+------------------+
|           968675 |
+------------------+
1 row in set (1.74 sec)
```