

Basic SQL analysis

Connect to RDS

```
mysql -h mlc-testcapstone.cyaiehc9bmnf.us-east-1.rds.amazonaws.com -u student -p
```

```
STUDENT123
```

```
Use mlctest;
```

```
Show tables;
```

Count of unique device ids in the train table

Query: Select count(distinct(device_id)) from train;

Identified 74645 unique device ids

```
MySQL [mlctest]> Select count(distinct(device_id)) from train;
+-----+
| count(distinct(device_id)) |
+-----+
| 74645 |
+-----+
1 row in set (0.38 sec)
```

Check whether there are any duplicate device ids present in the brand_device table. If yes, how many duplicates?

Query: Select device_id, count(1) cnt from brand_device group by device_id having cnt>1;

There are duplicates

```
MySQL [mlctest]> Select device_id, count(1) cnt from brand_device group by device_id having cnt>1;
+-----+-----+
| device_id | cnt |
+-----+-----+
| -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 |
| -1788272684905940000 | 2 |
| -1831358856417310000 | 2 |
| -1913074978634530000 | 2 |
| -1919970327849700000 | 2 |
| -1922527040000505000 | 2 |
```

Query: Select count(device_id) from (Select device_id, count(1) cnt from brand_device group by device_id having cnt>1) a;

Identified 532 duplicates

```
MySQL [mlctest]> Select count(device_id) from (Select device_id, count(1) cnt from brand_device group by device_id having cnt>1) a;
+-----+
| count(device_id) |
+-----+
|          532 |
+-----+
1 row in set (1.05 sec)
```

Number of unique phone brands from the brand_device table

Query: Select count(distinct(phone_brand)) from brand_device;

There are distinct 97 phone brands

```
MySQL [mlctest]> Select count(distinct(phone_brand)) from brand_device;
+-----+
| count(distinct(phone_brand)) |
+-----+
|                97 |
+-----+
1 row in set (0.08 sec)
```

Count of device ids where the latitude and longitude detail are zero, from the events table

Query: Select count(device_id) from events where latitude=0 and longitude=0;

There are 968675 rows with lat and long =0

```
MySQL [mlctest]> Select count(device_id) from events where latitude=0 and longitude=0;
+-----+
| count(device_id) |
+-----+
|        968675 |
+-----+
1 row in set (1.38 sec)
```