Haley Candia Perez

September 28, 2023

DAD-220

Module 5 Major Activity

Haley's starting point as per email (I lost all of my databases last week so this is what I have redone so far)

```
mysql> show databases;
+--------------------+
| Database           |
+--------------------+
| information_schema |
| QuantigrationRMA   |
| candiaperez        |
| mysql              |
| performance_schema |
+--------------------+
5 rows in set (0.00 sec)

mysql> use candiaperez;
Database changed
mysql>
```

1. Write SQL commands that capture specific, usable data that can be used in your analysis.

2. Analyze the results of queries to identify specific information that can be presented in your summary.

A. Sales by region:

- Analyze sales data by state to determine where the company has the largest customer base.

```
mysql> SELECT State, COUNT(*) FROM Customers GROUP BY State ORDER BY COUNT(*) DESC;
+---------------+----------+
| State         | COUNT(*) |
+---------------+----------+
| Massachusetts |      982 |
| Arkansas      |      854 |
| West Virginia |      843 |
| Oregon        |      842 |
| Alabama       |      838 |
| Idaho         |      838 |
| Mississippi   |      834 |
| Tennessee     |      831 |
| Connecticut   |      830 |
| Delaware      |      824 |
```

The state with the largest customer base is Massachusetts.

- Analyze the data to determine the top three products sold in the United States.

```
mysql> SELECT Orders.SKU, COUNT(*) FROM Orders GROUP BY Orders.SKU ORDER BY COUNT(*) DESC;
+------------+----------+
| SKU        | COUNT(*) |
+------------+----------+
| BAS-48-1 C |     8385 |
| ENT-48-40F |     6186 |
| ENT-48-10F |     4329 |
| BAS-08-1 C |     4285 |
| ENT-24-10F |     4275 |
| ADV-24-10C |     4178 |
| ADV-48-10F |     4174 |
| ENT-24-40F |     2152 |
| BAS-24-1 C |       34 |
+------------+----------+
9 rows in set (0.21 sec)
```

The top three products in the US are:

BAS-48-1 C

ENT-48-40F

ENT-48-10F

- Analyze the data to determine the top three products sold in the southeastern region of the United States.
    - Southeastern states to include in your analysis: Virginia, North Carolina, South Carolina, and Georgia

```
mysql> SELECT Orders.SKU, COUNT(*) FROM Orders INNER JOIN Customers ON Orders.CustomerID = Customers.CustomerID WHERE Customers
.State IN ('North Carolina', 'South Carolina', 'Georgia', 'Virginia') GROUP BY Orders.SKU ORDER BY COUNT(*) DESC;
+------------+----------+
| SKU        | COUNT(*) |
+------------+----------+
| BAS-48-1 C |      504 |
| ENT-48-40F |      337 |
| BAS-08-1 C |      257 |
| ADV-48-10F |      255 |
| ENT-48-10F |      247 |
| ADV-24-10C |      243 |
| ENT-24-10F |      235 |
| ENT-24-40F |      143 |
| BAS-24-1 C |        1 |
+------------+----------+
9 rows in set (0.20 sec)
```

The top three products sold in the southeastern region of the US:

BAS-48-1 C

ENT-48-40F

BAS-08-1 C

B. Returns by region:

- Analyze the data to determine the top three products returned in the United States.

```
mysql> SELECT Orders.SKU, COUNT(*) FROM Orders INNER JOIN RMA ON RMA.OrderID = Orders.OrderID GROUP BY Orders.SKU ORDER BY COUN
T(*) DESC;
+------------+----------+
| SKU        | COUNT(*) |
+------------+----------+
| BAS-48-1 C |     8282 |
| ENT-48-40F |     6118 |
| ENT-48-10F |     4287 |
| BAS-08-1 C |     4248 |
| ENT-24-10F |     4231 |
| ADV-48-10F |     4124 |
| ADV-24-10C |     4122 |
| ENT-24-40F |     2121 |
| BAS-24-1 C |       33 |
+------------+----------+
9 rows in set (0.09 sec)
```

The top three products returned by region are:

BAS-48-1 C

ENT-48-40F

ENT-48-10F

- Analyze the data to determine the top three products returned in the northwestern region of the United States.
  - Northwestern states to include in your analysis: Washington, Oregon, Idaho, and Montana

```
mysql> SELECT Orders.SKU, COUNT(*) FROM Customers INNER JOIN Orders ON Customers.CustomerID = Orders.CustomerID INNER JOIN RMA
ON Orders.OrderID = RMA.OrderID WHERE Customers.State IN ('Washington', 'Oregon', 'Idaho', 'Montana') GROUP BY Orders.SKU ORDER
 BY COUNT(*) DESC;
+------------+----------+
| SKU        | COUNT(*) |
+------------+----------+
| BAS-48-1 C |      697 |
| ENT-48-40F |      534 |
| BAS-08-1 C |      379 |
| ENT-24-10F |      374 |
| ENT-48-10F |      362 |
| ADV-48-10F |      357 |
| ADV-24-10C |      350 |
| ENT-24-40F |      198 |
+------------+----------+
8 rows in set (0.04 sec)
```

The top three products returned in the northwestern region of the US are:

BAS-48-1 C

ENT-48-40F

BAS-08-1 C

3. Write a report to the Quantigration product manager that explains your findings in a way nontechnical stakeholders can digest and use.

- This report should include an effective summary of the analysis of the captured data.
  - Sales data by region: Provide a well-written summary of your analysis on Part A.
  - Returns data by region: Provide a well-written summary of your analysis of Part B.

Summary:

Sales data by region Part A:

For my analysis, I began by requesting information for the most sales by State, so that we could see what states the company is performing the best in and determine where the largest customer base is and sorted them from highest to lowest. I then requested to see which products are performing the best across the United States, so that we can see what most customers are keen to

buy, and again sorted from highest to lowest. Finally, I queried to see which are the most sold products in the southeastern United States, where we will be able to see what kind of products consumers are interested in regionally.

Returns data by region Part B:

As far as the return data, I started out my queries by checking to see which are the three most returned products across the United States and sorted this from most returned to least returned. With this information we can see which products are not performing well, and make adjustments as such. Afterwards, I made a query to see which are the top three products returned in the northwestern region of the United States from most returned in the northwest region to least returned. This information can display which products this region is more keen on, and which ones are not performing as well as we would like.