Haley Candia Perez

October 13, 2023

DAD-220

Project 2

DAD-220 Project Two: Analyzing Databases

Haley's Starting Point as per email

Scenario

The product manager of Quantigration has asked your data analytics team for a report

summarizing your analysis of the return merchandise authorizations (RMAs) that have been

received. These are the same data sets that you've already been working with. Your report should

focus on summarizing the analysis and presenting your findings to the product manager.

Directions

RMA Report

In your report, respond to the manager's requests: to summarize the data you've been working

with and to identify key information that will help the company streamline operations.

Remember, not everyone who reviews this report will have a technical background.

1. Begin by writing SQL commands to capture usable data (which you've preloaded into

   Codio) for your analysis.

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

mysql> use QuantigrationUpdates;
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A

Database changed
mysql>
```

2. Specifically, the product manager wants you to analyze the following:

- Analyze the number of returns by state and describe your findings in your report.

```
mysql> SELECT Customers.State AS State, Count(*) AS TOTAL_RETURNS FROM Orders INNER JOIN RMA ON Orders.OrderID =
    -> RMA.OrderID INNER JOIN Customers ON Customers.CollaboratorID = Orders.CollaboratorID
    -> GROUP BY State
    -> ORDER BY TOTAL_RETURNS DESC
    -> LIMIT 15;
+---------------+---------------+
| State         | TOTAL_RETURNS |
+---------------+---------------+
| Massachusetts |           972 |
| Arkansas      |           844 |
| Oregon        |           840 |
| West Virginia |           837 |
| Alabama       |           836 |
| Idaho         |           822 |
| Connecticut   |           822 |
| Mississippi   |           821 |
| Tennessee     |           819 |
| Delaware      |           811 |
| Kentucky      |           809 |
| Montana       |           808 |
| Wisconsin     |           807 |
| New Mexico    |           807 |
| Iowa          |           804 |
+---------------+---------------+
15 rows in set (0.15 sec)

mysql>
```

```
mysql> SELECT Customers.State AS STATE, COUNT(*) AS PRODUCT_SALES_NUMBER
    -> FROM Customers INNER JOIN Orders ON Customers.CollaboratorID = Orders.CollaboratorID
    -> GROUP BY STATE
    -> ORDER BY PRODUCT_SALES_NUMBER DESC
    -> LIMIT 15;
+---------------+----------------------+
| STATE         | PRODUCT_SALES_NUMBER |
+---------------+----------------------+
| Massachusetts |                  982 |
| Arkansas      |                  854 |
| West Virginia |                  843 |
| Oregon        |                  842 |
| Alabama       |                  838 |
| Idaho         |                  838 |
| Mississippi   |                  834 |
| Tennessee     |                  831 |
| Connecticut   |                  830 |
| Delaware      |                  824 |
| Wisconsin     |                  822 |
| Montana       |                  820 |
| Kentucky      |                  819 |
| New Mexico    |                  818 |
| Iowa          |                  814 |
+---------------+----------------------+
15 rows in set (0.21 sec)

mysql>
```

Here, you can see that the state with the most returns is Massachusetts with 972 returns, followed by Arkansas with 844 returns, Oregon with 840 returns, and West Virginia with 837 returns. You can see that after Massachusetts, the following states have very similar return counts. Even at that, Massachusetts does not have significantly greater returns than the others. Also, if you compare the returns by state to the sales by state as per the preceding screenshots, you can see that Massachusetts, Arkansas, West Virginia, and Oregon comparatively also have the top four sales by state.

- Analyze the percentage of returns by product type and describe your findings in your report.

```
mysql> SELECT Orders.SKU as SKU,
    -> (Count(*) * 100 / (SELECT COUNT(*) FROM
    -> Orders
    -> INNER JOIN RMA ON Orders.OrderID =
    -> RMA.OrderID)) AS Percent_Return
    -> from Orders inner join RMA on
    -> Orders.OrderID = RMA.OrderID
    -> GROUP BY SKU
    -> ORDER BY Percent_Return DESC;
+------------+----------------+
| SKU        | Percent_Return |
+------------+----------------+
| BAS-48-1 C |        22.0465 |
| ENT-48-40F |        16.2860 |
| ENT-48-10F |        11.4119 |
| BAS-08-1 C |        11.3081 |
| ENT-24-10F |        11.2628 |
| ADV-48-10F |        10.9780 |
| ADV-24-10C |        10.9727 |
| ENT-24-40F |         5.6461 |
| BAS-24-1 C |         0.0878 |
+------------+----------------+
```

```
mysql> SELECT COUNT(*) AS TOTAL_RETURNED, Orders.SKU AS PRODUCT_SKU, RMA.Reason
    -> FROM RMA INNER JOIN Orders ON Orders.OrderID = RMA.OrderID
    -> GROUP BY PRODUCT_SKU
    -> ORDER BY TOTAL_RETURNED DESC
    -> LIMIT 10;
+----------------+-------------+-----------+
| TOTAL_RETURNED | PRODUCT_SKU | Reason    |
+----------------+-------------+-----------+
|           8282 | BAS-48-1 C  | Defective |
|           6118 | ENT-48-40F  | Defective |
|           4287 | ENT-48-10F  | Defective |
|           4248 | BAS-08-1 C  | Defective |
|           4231 | ENT-24-10F  | Defective |
|           4124 | ADV-48-10F  | Defective |
|           4122 | ADV-24-10C  | Defective |
|           2121 | ENT-24-40F  | Defective |
|             33 | BAS-24-1 C  | Defective |
+----------------+-------------+-----------+
9 rows in set (0.06 sec)

mysql>
```

Here, we can see that the most purchased items were SKU BAS-48-1 C with a return rate of
22.05, the second most returned was ENT-48-40F with a return rate of 16.29, and SKU
ENT-48-10F with a return rate of 11.41. In the second screenshot, we can see that the most
common reason was "Defective," so we know that something in the product needs to be
corrected. We can speculate from the data provided that there is a defect or perhaps quality
issue in this product that is causing the returns, which may be something that should be
looked into.

3. In your report, clearly summarize your analysis of the data for stakeholders. Include screenshots of the results of each query. When summarizing results, you may want to consider the following questions:

- How does the data provide the product manager with usable information?

The data provides the product manager with useful information because they can now see what products are most likely to be returned and how they may attempt to decrease the amounts of returns that the company is receiving. Also, the product manager can view what states are returning the most to see if the issue is coming from a specific region, which it doesn't seem to be, so we can speculate that it may be an issue with the product. The product manager will also be able to see that the states that are buying the most, are also returning the most.

- What are the potential flaws in the data that has been presented?

Some potential flaws could be the way that the "reason" was recorded. A large majority has a reason of "defective" which is a vague reason that makes it hard to determine what exactly the defect is. Having a specific reason for the defect would help the company correct the issue and prevent such a high rate of returns. Also, human error in the data could affect the outcomes of these queries.

- Are there any limitations on your conclusions, or any other ways of looking at it that you haven't considered? Clearly communicate your findings to stakeholders.

I think that the main limitation I can think of would be the lack of description for the product returns. We can see that some products have a high return percentage, but it is hard to identify why a product was returned specifically or if there is a trend that could be addressed with the reasoning being so vague. For the stakeholders, the data provided shows that the best selling products from the company are also some of the most returned products. This would be a key comparison to evaluate as the company could potentially make a large profit if these products weren't so frequently returned.