Matthew Keaton

Professor Renetta English

DAD 220

February 24, 2024

**DAD 220 Project 2**

1. Begin by writing SQL commands to**capture usable data** (which you’ve preloaded into Codio) for your analysis.A computer screen with white text

   Description automatically generated

Code used:

Show databases;

Use QuantigrationUpdates;

Show tables;

1. Specifically, the product manager wants you to analyze the following:  
   * **Analyze** the **number of returns** **by state** and describe your findings in your report.

The following query joins all three tables in the database QuantigrationUpdates to pull the data for the amount products returned in each state. The data shows that the leading state with the most products returned is Massachusetts with 972 and the lowest amount of products being returned in South Carolina at 707.

A computer screen with white text

Description automatically generated

A computer screen shot of a black screen

Description automatically generated

Code used:

SELECT Customers.State AS State, COUNT(\*) AS Product\_RMA

FROM RMA

INNER JOIN Orders ON Orders.OrderID = RMA.OrderID

INNER JOIN Customers ON Customers.CustomerID = Orders.CustomerID

GROUP BY State

ORDER BY Product\_RMA

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

The first image shows the total amount of products sold compared to the total amount of products returned, which has a 98.9% return on all products. As a company this would be very concerning as almost every product being sold is being returned. The second image describes the product that is sold as well as the return percentage. This can show us what product is affecting the overall return percentage (out of 100%) the most.

A computer screen shot of white text

Description automatically generated

Code used:

SELECT

(SELECT COUNT(\*) FROM Orders) AS Orders,

COUNT(\*) AS RMA,

COUNT(\*)/(SELECT COUNT(\*) FROM Orders)\*100 AS Percentage\_Orders\_Returned

FROM Customers

INNER JOIN Orders ON Customers.CustomerID = Orders.CustomerID

INNER JOIN RMA ON Orders.OrderID = RMA.OrderID;

A computer screen shot of a computer program

Description automatically generated

Code used:

SELECT

SKU AS Product\_SKU,

Description AS Product\_Description,

COUNT(\*) AS Total,

(COUNT(\*) / (SELECT COUNT(\*) FROM RMA)) \* 100 AS Return\_Percentage

FROM RMA

INNER JOIN Orders ON Orders.OrderID = RMA.OrderID

GROUP BY Product\_SKU, Description

ORDER BY Return\_Percentage DESC LIMIT 10;

1. 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 shown can provide the product manager with usable information by allowing them to see exact data points to where in the country they are seeing the most products being returned. This helps the product manager by showing the exact states where they are seeing the highest returns on their products, allowing them to adjust production in specific states or regions to minimize overall return percentages. The data provided also shows exactly what product is having the highest effect on the overall return percentages. The product manager can see that the product BAS 48 1 C has the highest impact on the overall return percentage at 22%. This data can help the product manager detect patterns in products and locations to make informed decisions and ensure business success.

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

A potential flaw can be found in listing the returns for the states. The command only brought up a total of 48 states. This could mean that the two missing states have no products returned or are just missing, which we are unsure of. Having the two missing states can help us make informed decisions depending on their data. If the two states have no products being returned, then the product manager would want to know why they are having so much success in those states and can implement that in other states.

Another potential flaw in the data is the percentage of the products’ return rate may be biased. The product BAS 48 1 C has the highest impact on return rates at 22%. However, the data doesn’t show the number of products sold. This will help show why the percentage is extremely high. For example, the image below shows that the BAS 48 1 C is the highest product sold so it will have more returns, leading to a higher percentage on return rates. As for the bottom product, BAS 24 1 C has a percentage of .08 which may seem impressive, but compared to the number of products sold it is only a total of 34 with 33 being returned.

A computer screen with white text

Description automatically generated

Code used:

SELECT

SKU,

COUNT(distinct Orders.OrderID) AS Orders,

COUNT(distinct RMA.OrderID) AS RMA,

(COUNT(RMA.OrderID) / COUNT(Orders.OrderID) \* 100) AS ReturnPercentage

FROM Orders LEFT JOIN RMA ON RMA.OrderID = Orders.OrderID

GROUP BY SKU

ORDER BY Orders DESC;

* + 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.

A limitation that can be found in the data shown is under the “Reasons” from returns. Currently the reasons for returns are listed as defective, incorrect, or other. Having more specific reasons rather than “other” will help the company find a solution to why the products are being returned. Instead, the reason of “other” can be a list of multiple reasons the item was returned, which doesn’t help in understanding the issue for stakeholders to generate solutions. Another limitation that can be found in the data is there are no time stamps. For example, the product BAS 24 1 C has been ordered only 34 times with 33 being returned. This could be seen as a failed product if it were on the market for a while. However, if the product is brand new and has just been released, it would need time for the company to really understand the impact this product has. Having time stamps will help the stakeholders understand the data a little more for each product to be able to make informed decisions.