DAD 220 Project 2: Analyzing Databases

Darrell Walker II

12/9/2023

Prof. Andrew August

In this report we will be providing a summary report for the analysis of the return merchandise authorizations (RMA’s) that have been received via State.

**Our Goal:**

Is to summarize the data we’ve been working with and to identify key information that will help the company streamline operations.

* The first thing my team would like to share with you is our analysis of the number of returns by state. We do this by using SQL to access our database and data sets and in this case we will be using the database called “QuantigrationRMA“. Once we are in our IDE we enter the command ‘mysql’ to start the SQL environment and then enter the command ‘USE QuantigrationRMA’ to navigate to the desired database where we can analyze our data.

A screen shot of a computer

Description automatically generated

A black screen with white lines

Description automatically generated

The Command Used:

SELECT Customers.State AS STATE, COUNT(\*) AS RETURNS

FROM Orders INNER JOIN RMA ON Orders.OrderID=RMA.OrderID INNER JOIN Customers ON Customers.CustomerID = Orders.CustomerID

GROUP BY STATE

ORDER BY RETURNS DESC;

* As you can see above with the screenshots, we used a specific command that retrieves and summarizes data related to returns (RMAs) by analyzing the number of returns based on the states of customers who made those returns. It combines information from the Orders, RMA, and Customers tables through INNER JOIN operations, linking orders to returns and associating them with customer details. The SELECT statement outputs the state names as "STATE" and counts the occurrences of returns, aliasing it as "RETURNS." The GROUP BY clause organizes the results by state, and the ORDER BY clause arranges the output in descending order of return counts. Overall, this query aims to provide insights into the distribution of returns across different states, allowing for a focused analysis of customer return patterns.
* With this data we can clearly see that Massachusetts has the greatest number of returns with a total of 988.
* The next thing we will do in this report is analyze the percentage of returns by product type.

A screenshot of a computer

Description automatically generated

A black background with a black square

Description automatically generated with medium confidence

The Command Used:

SELECT Customers.State AS STATE, ( COUNT(\*) \* 100 / (SELECT COUNT(\*)

FROM Orders INNER JOIN RMA ON Orders.OrderID = RMA.OrderID)) AS RETURN\_PERCENTAGE

FROM Orders INNER JOIN RMA ON Orders.OrderID = RMA.OrderID INNER JOIN Customers ON Customers.CustomerID = Orders.CustomerID

GROUP BY STATE

ORDER BY RETURN\_PERCENTAGE DESC;

* The screenshots above and the specific command calculates the percentage of returns for each state by dividing the count of returns in each state by the total count of returns across all states and multiplying the result by 100. The query utilizes data from the Orders, RMA, and Customers tables, linking orders to returns and associating them with customer details. The SELECT statement outputs the state names as "STATE" and computes the return percentage, while the GROUP BY clause organizes the results by state. The ORDER BY clause arranges the output in descending order of return percentages, facilitating the identification of states with the highest proportions of returns relative to the total. This query is useful for understanding regional variations in return rates and can aid in targeted strategies to address specific customer satisfaction issues.
* Now we need to find which products have the highest return rate based on the total amount returned.

A screenshot of a computer program

Description automatically generated

The Command Used:

SELECT SKU AS SKU, description AS DESCRIPTION, COUNT(\*) AS RETURN\_FREQUENCY

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

GROUP BY SKU

ORDER BY RETURN\_FREQUENCY DESC;

A screenshot of a computer program

Description automatically generated

The Command Used:

SELECT SKU AS SKU, description AS DESCRIPTIN, (COUNT(\*) \* 100 / (SELECT COUNT(\*)

FROM Orders INNER JOIN RMA ON Orders.OrderID = RMA.OrderID)) AS RETURN\_PERCENTAGE

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

GROUP BY SKU

ORDER BY RETURN\_PERCENTAGE DESC;

* The first SQL command above retrieves data on product returns by SKU, counting the occurrences of each SKU in the Orders and RMA tables and presenting the results with the SKU, product description, and return frequency, ordered by return frequency in descending order. The second command calculates the percentage of returns for each SKU, expressing the return frequency as a proportion of the total returns across all SKUs. The output includes SKU, product description, and return percentage, with results grouped by SKU and ordered by return percentage in descending order. Together, these commands offer insights into both the raw return frequency and the normalized return percentage for each product SKU, enabling a comprehensive analysis of product return patterns. This information is valuable for identifying products with high return rates or frequencies, aiding in targeted strategies to improve product quality and customer satisfaction.

**Summary of the analysis for stakeholders:**

1. How does the data provide the product manager with usable information?
   1. The data provides the product manager with actionable insights in two key ways. Firstly, the analysis of return frequency by SKU identifies products with consistently high return rates, enabling targeted investigations into potential quality, design, or customer communication issues. This information is crucial for implementing specific improvements and optimizing product offerings. Secondly, the return percentage by SKU offers a normalized view, helping prioritize products with disproportionately high return rates compared to the overall average. This allows the product manager to strategically allocate resources to address systemic issues and enhance overall customer satisfaction. By leveraging these insights, the product manager can make informed decisions, streamline operations, and implement targeted strategies to reduce returns and enhance the company's product and service offerings.
2. What are the potential flaws in the data that has been presented?
   1. The presented data may have several potential flaws. Incomplete data, stemming from missing or incomplete records, could compromise the accuracy of return frequencies and percentages. The analysis assumes causation for returns without providing insights into the specific reasons behind them, potentially leading to misguided conclusions. Lack of contextual information, such as customer feedback or external factors influencing returns, limits the depth of understanding. The analysis focuses on return frequencies and percentages, neglecting additional dimensions that could provide a more comprehensive view. Assumptions about uniform SKU definitions and the absence of considerations for external factors or data aging may impact the reliability of the findings. Furthermore, sampling bias, where the data may not represent the entire customer base or product range, could introduce inaccuracies in the derived insights. Addressing these potential flaws requires a holistic approach, combining quantitative data analysis with qualitative insights and considering a range of contextual factors to make well-informed decisions.
3. Are there any limitations on your conclusions, or any other ways of looking at it that you haven’t considered?
   1. While the data analysis provides valuable insights, it's important to acknowledge certain limitations and consider alternative perspectives. The analysis focuses on the frequency and percentage of returns but lacks information on the specific reasons behind each return. Understanding customer feedback and the context of returns could provide a more comprehensive picture. Additionally, external factors such as economic changes or industry trends might influence return rates, but these are not accounted for in the current analysis. The assumption of a uniform definition for product codes (SKUs) may oversimplify the data, and variations in labeling could impact the accuracy of conclusions. Moreover, the analysis does not explore changes over time, and relying on historical data may not capture the evolving dynamics of the market or customer preferences. To enhance the robustness of our conclusions, stakeholders should consider incorporating qualitative insights, exploring additional dimensions of the data, and regularly updating analyses to reflect the current business environment. This approach ensures a more well-rounded understanding and informs more effective decision-making processes.