

SALES AND RETURNS DATA ANALYSIS

Amisha Nagiya

CONTENTS

- Agenda
- Data Overview
- Data Preparation
- SQL and Excel Approaches
- Key Insights
- Summary

AGENDA

This presentation aims to provide a comprehensive analysis of the company's sales and return data using both SQL and Excel approaches. By uncovering key metrics, trends, and patterns, it will help identify areas for improvement, enhance customer satisfaction, and drive strategic business decisions. The insights gained will offer actionable recommendations to optimize sales performance and minimize return rates, ultimately contributing to the company's growth and profitability.

DATA OVERVIEW

SALES

- CustomerID
- OrderID
- Sales
- TransactionDate

RETURNS

- CustomerID
- OrderID
- ReturnSales
- ReturnDate

Raw Data Details

Sales Table – The dataset contains transaction records.

- **No. of rows: 33752**
- **CustomerID:** Unique identifier for each customer.
- **OrderID:** Unique identifier for each order.
- **Sales:** Monetary value of the sales transaction.
- **TransactionDate:** Date when the sale was made.

Returns Table - The dataset contains details of returned orders.

- **No. of rows: 15829**
- **CustomerID:** Unique identifier for each customer.
- **OrderID:** Unique identifier for the order being returned.
- **ReturnSales:** Monetary value of the returned sales.
- **ReturnDate:** Date when the return was made.

Data Relationships

- The 'Sales' and 'Returns' table are related through the 'OrderID' column.

DATA PREPARATION AND PROCESSING

Data Cleaning and Transformation (Power Query Editor in Excel)

- 1.Removed Duplicates: Ensured no duplicate entries in both tables.
- 2.Checked for Null Values: Identified and handled any null values.
- 3.Detected Data Types: Verified and corrected data types for all columns.
- 4.Formatted Columns:
 - TransactionDate: Ensured correct date format.
 - ReturnDate: Ensured correct date format.

Result After Cleaning and Transformation

- Sales Table: 33,752 rows (unchanged)
- Returns Table: 15,755 rows

Data Import and Analysis in MySQL

- 1.Created Database: Set up a new database for analysis.
- 2.Created Table Structures: Defined the schema for both Sales and Returns tables with appropriate data types.
- 3.Imported Cleaned Data: Loaded the cleaned data into MySQL tables.
- 4.Wrote Efficient Queries: Performed analysis using SQL queries to ensure quick and efficient data retrieval.

Analysis in Excel

- 1.Data Loading:
 - Preferred loading data into worksheets rather than a data model in Power Pivot.
- 2.Computations:
 - Utilized relative and absolute referencing.
 - Applied various formulas for data manipulation and analysis.
- 3.Pivot Tables: Created Pivot Tables for summarizing and analyzing data.
- 4.Visualizations: Developed charts and graphs for visual insights.
- 5.Similar Analysis as SQL: Conducted similar analysis in Excel as performed in SQL for consistency.



DIFFERENT APPROACHES FOR DATA ANALYSIS

% of Sales Resulting in a Return

Approach 1: It calculates the percentage of orders that resulted in a return based on the count of sales and return orders.

```
SELECT  
ROUND((COUNT(DISTINCT r.OrderID) * 100 / COUNT(DISTINCT s.OrderID)),2) as Return_Percentage  
FROM Sales s  
LEFT JOIN Returns r  
ON s.OrderID = r.OrderID;
```

Output: 4.23%

Excel

1. In sales table, VLOOKUP function was used to connect the two tables on OrderID column and retrieved the OrderID from Returns table in Sales table. Also used IFERROR function to handle errors.
2. Inserted a pivot table to organize and count the total number of orders in both Sales and Returns tables. Formatted the numbers and used ROUND function too.
3. Based on these data, divided the number of orders with a return by the total sales orders. Multiplying this result by 100 gives us the percentage of sales that resulted in a return.

Total No. of Sales	Total No. of Returns	Return Percentage
33,751	1,428	4.23

INSIGHTS

- **Low return rate**

A return rate of 4.23% states that a relatively small proportion of sales are being returned. This indicates high customer satisfaction with the products and services.

- **Customer behavior and satisfaction**

Customers are not facing significant issues with their purchases and the products are meeting their expectations. This would definitely improve customer retention and build their loyalty towards the company.

- **Operational efficiency**

Fewer returns leads to lower operational costs involved in return policies, ultimately improving company's overall operational efficiency and profitability.

- **Areas of improvement**

Even though the return rate is low, certain areas still need investigation such as on the products returned and the cause behind it. Based on this further analysis, the quality of products or services can be improved accordingly.

Return % of Sales



% of Sales Resulting in a Return

Approach 2: It calculates the percentage of orders that resulted in a return based on the total sales and return values.

```
Select  
ROUND(SUM(r.ReturnSales)*100/SUM(s.SALES),2) as Return_Percentage  
from Sales s  
LEFT JOIN Returns r  
on s.OrderID = r.OrderID;
```

Output: 2.59%

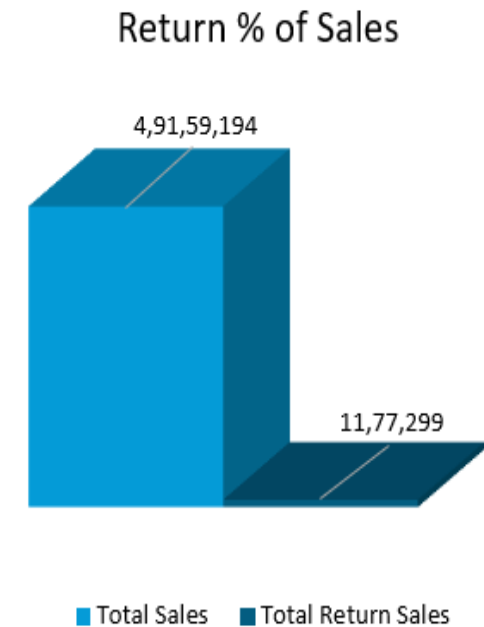
Excel

1. In sales table, used a VLOOKUP function to connect the two tables on OrderID column and retrieved the ReturnSales from Returns table in Sales table. Also used IFERROR function to handle errors.
2. Inserted a pivot table to organize and sum up the total sales and total return sales. Formatted the numbers and used ROUND function too.
3. Based on these data, divided the total return sales by total sales. Multiplying this result by 100 gives us the percentage of sales that resulted in a return.

Total Sales	Total Return Sales	Return Percentage
4,91,59,194	11,77,299	2.39

INSIGHTS

- **Slight discrepancy**
There's a slight difference in SQL and Excel results which might occur due to the different data handling of both the tools.
- **Low return rate**
Both results indicates relatively low return rate (below 3%), suggesting good product quality and customer satisfaction.
- **Revenue protection**
With low return rate, there's a direct implication that the total return sales doesn't cause high impact on the overall sales value. Lesser money is being lost which can be compensated with further sales.



Full Returns % of Total Returns

Approach 1: It calculates the percentage of full returns on the basis of total number of full returns and returns.

```
SELECT
    ROUND((FR.TotalFullReturns * 100.0 / TR.TotalReturns), 2)
    AS FullReturnPercentage
FROM
    (SELECT COUNT(*) AS TotalFullReturns
     FROM Returns r
     INNER JOIN Sales s
     ON r.OrderID = s.OrderID
     AND r.ReturnSales = s.Sales) AS FR,
    (SELECT COUNT(OrderID) AS TotalReturns
     FROM Returns) AS TR;
```

Output: 0.34%

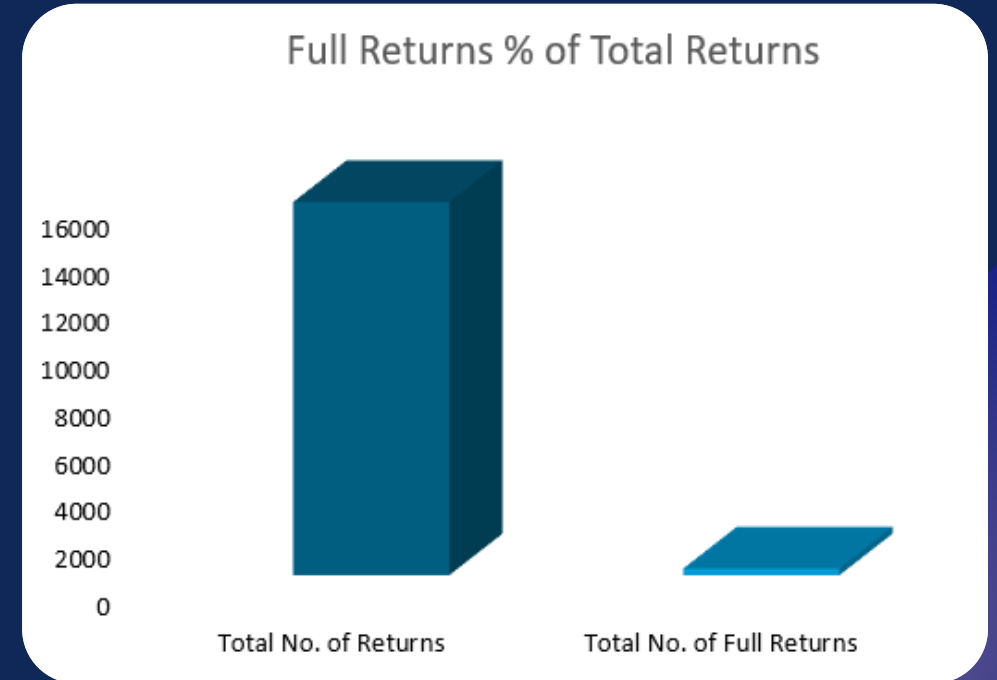
Excel

1. Used VLOOKUP to find corresponding Sales amount for each Return based on OrderID column and IFERROR function to handle potential errors if matching value not found.
2. Employed IF function to compare Sales and Return Sales value to find Full Returns.
3. Used COUNTIF function to count the total number of Sales and total number of Return Sales. Based on these calculated the percentage of full returns in the total returns.

Total No. of Returns	Total No. of Full Returns	Full Returns Percentage
15754	283	1.8%

INSIGHTS

- **Slight discrepancy**
There's a slight difference in SQL and Excel results which might occur due to the different data handling of both the tools. In SQL, its 0.34% and in excel, its 1.8%.
- **Full return rate**
Both results indicates that the full returns are rare. A very less percentage of returns are full returns and most returns are partial.
- **Customer retention**
Customers are mostly satisfied with most of the products purchased which means they would most likely want to make further purchases in future.
- **Return policy**
It also indicates that the company's return policy is flexible for the customers which makes it easier for them to return the specific products as per their choice.



Full Returns % of Total Returns

Approach 2: It calculates the percentage of full returns on the basis of total full returns and total returns.

```
SELECT
    ROUND((FR.TotalFullReturns * 100.0 / TR.TotalReturns), 2) AS FullReturnPercentage
FROM
    (SELECT SUM(r.ReturnSales) AS TotalFullReturns
     FROM Returns r
     INNER JOIN Sales s
     ON r.OrderID = s.OrderID
     AND r.ReturnSales = s.Sales) AS FR,
    (SELECT SUM(ReturnSales) AS TotalReturns
     FROM Returns) AS TR;
```

Output: 0.31%

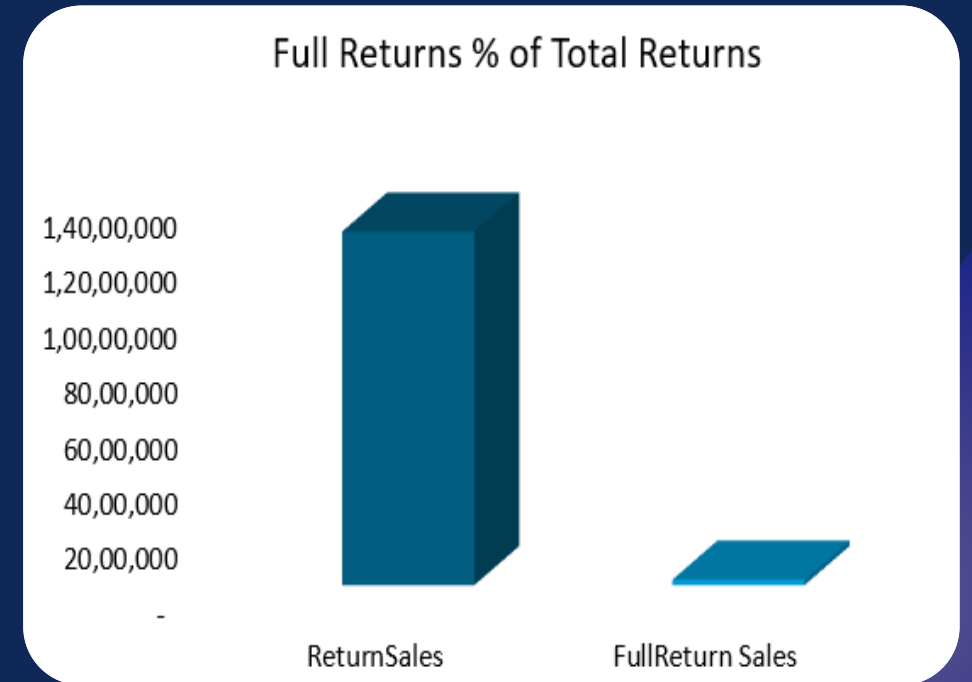
Excel

1. Used VLOOKUP to find corresponding Sales amount for each Return based on OrderID column and IFERROR function to handle potential errors if matching value not found.
2. Employed IF function to compare Sales and Return Sales value to find Full Returns.
3. Used COUNTIF function to count the total Sales and total Return Sales value. Based on these calculated the percentage of full returns in the total returns.

ReturnSales	FullReturn Sales	FullReturns Percentage
1,27,36,404	1,98,777	1.6%

INSIGHTS

- **Slight discrepancy**
As before, there's a notable difference in SQL and Excel results. In SQL, it's 0.31% and in Excel, it's 1.6%.
- **Full return rate**
Both percentages are relatively low. Full returns are very uncommon.
- **Product quality**
Clearly, customers are satisfied with most of the products and services which is why there are mostly partial returns.
- **Analysis required**
Though there are low full returns, there's still a requirement for finding out the issues with products which are being returned.



Average Return % Amount of Original Sale

SQL

```
SELECT  
    ROUND(AVG((r.ReturnSales / s.Sales) * 100), 2) AS AvgReturnPercentage  
FROM Returns r  
LEFT JOIN Sales s  
ON r.OrderID = s.OrderID;
```

Output: 52.88%

Excel

1. Used VLOOKUP and IFERROR to retrieve and handle potential errors in the Sales amount data for each Return based on the OrderID.
2. Leveraged pivot tables to efficiently organize the return amount and sales amount for each OrderID.
3. Using separate formulas, we calculated the return percentage for each order by dividing the return amount by the corresponding sales amount and multiplying by 100. Then rounded it off to enhance readability.
4. Used AVERAGE function, to calculate the average of all individual return percentages obtained. This provides a single value representing the average percentage of the original that's being returned.

Average Return %
57.36%

INSIGHTS

- **High return percentage**

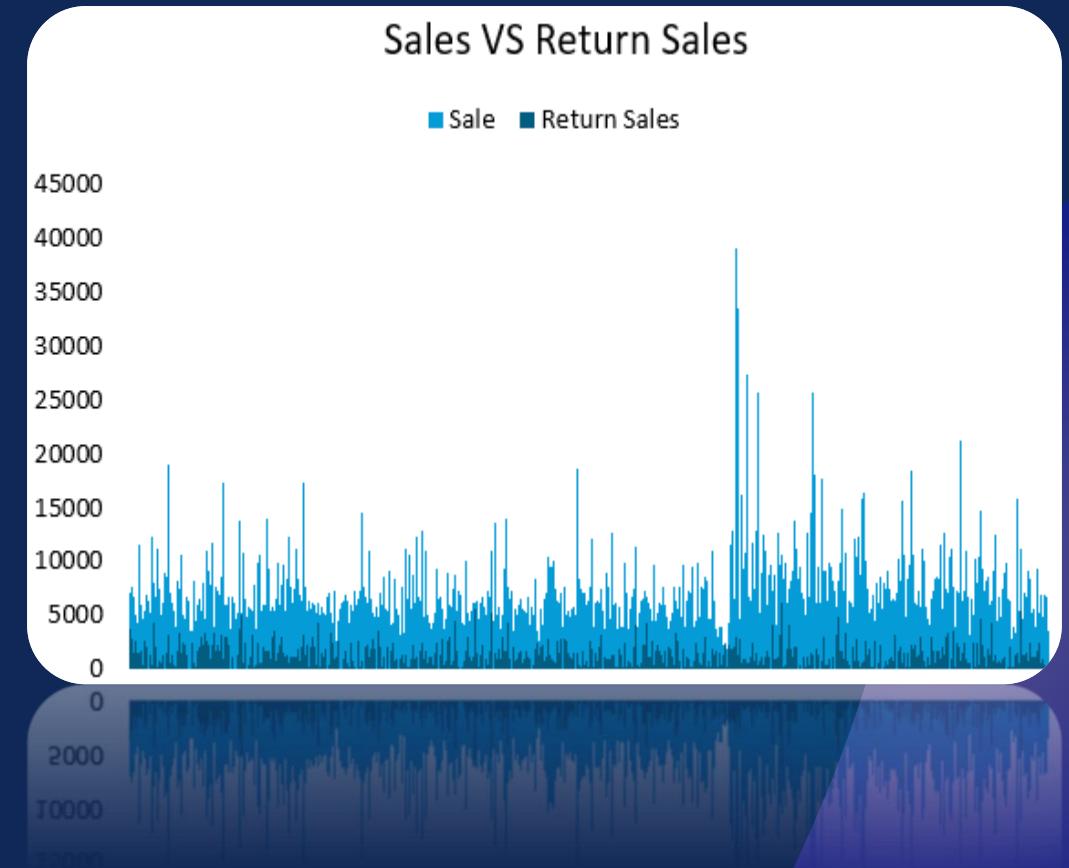
Both of the percentages are relatively high, indicating that over half of the sale amount is being returned on average. This is a significant finding and suggests potential issues that needs to be addressed.

- **Financial Impact**

High return percentages can have a substantial financial impact on the company. This includes not only the direct loss of sales revenue but also the costs associated with processing returns, restocking, and potential loss of customer trust and future sales.

- **Return policy influence**

Lenient return policy encourages customers to return the orders when they don't need it anymore even if they are satisfied with the products. There might be a need to reconsider the return policy.



% of Returns Occurring Within 7 Days of Sale

SQL

```
SELECT  
ROUND((SUM(CASE WHEN DATEDIFF(r.ReturnDate, s.TransactionDate) <= 7 THEN 1 ELSE 0 END)/ COUNT(*)* 100.0 ),2) AS Percentage  
FROM Returns r  
JOIN Sales s  
ON r.OrderID = s.OrderID;
```

Output: 40.13%

Excel

1. In Sales table, used VLOOKUP and IFERROR function to retrieve corresponding return date for each OrderID.
2. Employed an IF function along with ISNUMBER to calculate the difference between return date and transaction date if a return date exists. This gives days since purchase.
3. Used another IF function to identify the returns that occurred within 7 days of purchase.
4. Counted the total number of relevant returns using COUNTIF and summed up the number of returns occurring within 7 days using SUMIF function.
5. Base on this, derived the percentage of returns occurring within 7 days of original sale and rounded it off to improve readability.

DaysSincePurchase<=7	Total Count	Return % within 7 days
601	1428	42.09%

INSIGHTS

- **High early return rate**
Both percentages indicate that a significant portion of returns (around 40-42%) occur within the first week of purchase. This is an important finding.
- **Customer behavior**
The results indicates that the customers quickly regret their purchases and return items. This could be related to impulse buying, unmet expectations, or product dissatisfaction.
- **Product quality**
Early returns can indicate that products do not meet customer expectations upon delivery which means it needs further analysis to find insights on this.
- **Operational efficiency**
This can be beneficial for quick restocking, reselling and reducing inventory holding costs.



Average Number of Days for a Return to Occur

SQL

```
SELECT  
ROUND(AVG(DATEDIFF(r.ReturnDate, s.TransactionDate))) AS AverageDaysToReturn  
FROM Returns r  
JOIN Sales s  
ON r.OrderID = s.OrderID;
```

Output: 79

Excel

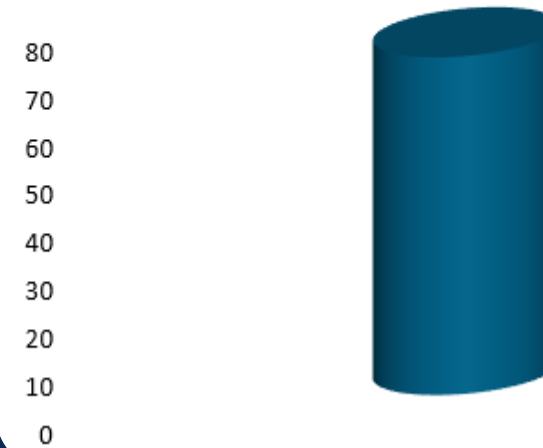
1. Previously, "Days Since Purchase" column was already created by calculating the difference between transaction date and return date.
2. For this problem, average of these number of days for a return to occur is being needed. This was done by inserting a pivot table based on Sales table to find the average of "Days Since Purchase" values.
3. Formatted the outcome in whole number to represent the average number of days.

Average Days
70

INSIGHTS

- **Customer behavior**
Customers generally take about 2 to 2.5 months to return products.
- **Return policy**
A longer return period should not be in the policy as it may encourage customers to return the products anytime they wish with maximum leniency, return used/damaged items and it may also impact operational efficiency.
- **Area of improvement**
There's a requirement for balancing customer satisfaction and operational efficiency too by reconsidering the return policy as it maybe be flexible for the customers but it still needs some enhancements.

Average Days for a Return to Occur



Most Valuable Customers

SQL

```
WITH SalesData AS (  
    SELECT CustomerID, ROUND(SUM(Sales),2) AS TotalSales, COUNT(OrderID) AS TotalOrders  
    FROM Sales  
    GROUP BY CustomerID  
)  
,  
ReturnsData AS (  
    SELECT CustomerID, ROUND(SUM(ReturnSales),2) AS TotalReturns  
    FROM Returns  
    GROUP BY CustomerID  
)  
,  
CustomerMetrics AS (  
    SELECT s.CustomerID, s.TotalSales,  
        COALESCE(r.TotalReturns, 0) AS TotalReturns,  
        ROUND((s.TotalSales - COALESCE(r.TotalReturns, 0)),2) AS NetSales,  
        s.TotalOrders, ROUND((s.TotalSales / s.TotalOrders),2) AS AverageOrderValue  
    FROM SalesData s  
    LEFT JOIN ReturnsData r  
    ON s.CustomerID = r.CustomerID  
)  
SELECT CustomerID, TotalSales, TotalReturns,  
    NetSales, TotalOrders, AverageOrderValue  
FROM CustomerMetrics  
ORDER BY  
    NetSales DESC,  
    AverageOrderValue DESC  
LIMIT 10;
```


Output:

CustomerID	TotalSales	TotalReturns	NetSales	TotalOrders	AverageOrderValue
RIVES87271	86441.8	0	86441.8	4	21610.45
HQARS21556	40840.8	699.95	40140.85	3	13613.6
POLLF17399	28606.5	0	28606.5	2	14303.25
BISSP26380	25659.7	0	25659.7	1	25659.7
RODJO44719	25798.9	2886.86	22912.04	1	25798.9
JUSUS89531	21581.6	0	21581.6	3	7193.87
DIALM39591	21335.8	753.15	20582.65	1	21335.8
KEENP37511	20413.5	0	20413.5	4	5103.38
ABDAH58684	19395	0	19395	2	9697.5
MARTE89584	19024.5	0	19024.5	1	19024.5

Most Valuable Customers

Excel

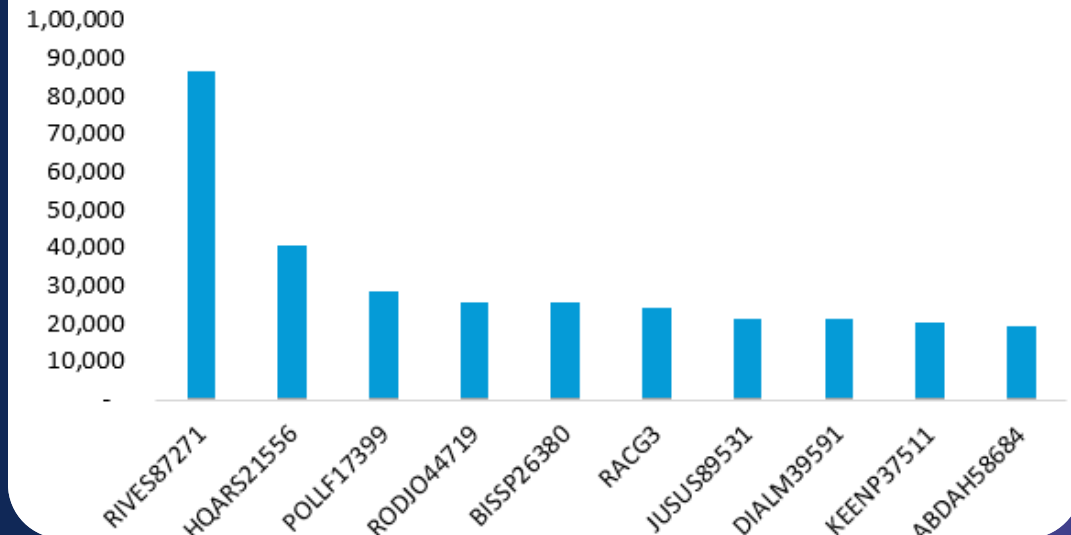
1. Calculated Net Sales for each customer by subtracting the total return amount from their total sales amount along with IFERROR function to handle errors.
2. Created a pivot table to organize and summarize the data of each customers. It includes Total Sales, Net Sales and Total Orders.
3. Based on the available data, derived Average Order Value too.
4. Sorted the pivot table and prioritized the top 10 customers based on their Net Sales.
5. This list represents the customers who have contributed the most in amplifying business growth.

Customer ID 	Total Sales	Net Sales	Total Orders	Average Order Value
RIVES87271	86,442	86,442	4	21,610.4
HQARS21556	40,841	40,841	3	13,613.6
POLLF17399	28,606	28,606	2	14,303.2
RODJO44719	25,799	25,799	1	25,798.9
BISSP26380	25,660	25,660	1	25,659.7
RACG3	24,420	24,420	20	1,221.0
JUSUS89531	21,582	21,582	3	7,193.9
DIALM39591	21,336	21,336	1	21,335.8
KEENP37511	20,413	20,413	4	5,103.4
ABDAH58684	19,395	19,395	2	9,697.5

INSIGHTS

- The most valuable customer, considering the metrics, appears to be **RIVES87271** due to their highest Net Sales, high Average Order Value, and no returns.
- Other valuable customers like **HQARS21556** and **POLLF17399** also show strong performance but with slightly lower overall sales or some returns.
- **BISSP26380** and **RODJO44719** are notable for their high average order values but have fewer transactions or significant returns affecting their net value.
- This multi-faceted approach helps in identifying not just the top spender but also the consistency and reliability of revenue from each customer.

Top 10 Customers based on their Net Sales



SUMMARY

Key Findings

- **High Customer Satisfaction:** Low overall return rate (around 4%) and low full return rate indicate high customer satisfaction with products and services.
- **High Early Return Rates:** A significant portion of returns (around 40%) occur within the first week of purchase, suggesting potential product-related issues or unmet customer expectations.
- **Lenient Return Policy:** The current return policy might be encouraging unnecessary returns, impacting finances and operational efficiency. Customers take a long time (2-2.5 months) to return products, potentially abusing the policy.

Areas for Improvement

- **Product Quality:** Reasons behind returned products and early returns needs to be analyzed and addressed to identify potential quality issues.
 - **Marketing and Customer Education:** Implement targeted marketing campaigns to address potential customer concerns and ensure that customers have accurate information about the products they are purchasing and its return policies too.
 - **Return Policy Optimization:** Re-evaluate the return policy to balance customer satisfaction with operational efficiency. Consider a shorter return period for non-defective items or stricter criteria and flexible at the same time, for returns.
- Overall, the data analysis suggests a positive customer experience with some key areas for improvement. By implementing the some optimizations, the company can further enhance customer satisfaction, minimize unnecessary returns, and enhance operational efficiency.
 - The company is doing great as there's customer satisfaction and retention, loyal customer base, and regular revenue generation where return sales doesn't impact much on overall sales of the company.



THANK YOU!

-Amisha Nagiya