Internship Report: Data Analysis

Introduction:

During my internship as a data analyst, I had the opportunity to work with a large dataset and extract valuable insights from it. My primary responsibilities involved cleaning and formatting the data, combining related information from different sources, grouping and aggregating data, and conducting thorough analysis. In this report, I will provide an overview of each task and explain the technical aspects in a way that is understandable for both technical and non-technical individuals.

Cleaning and Formatting Data:

The first step in data analysis is ensuring that the dataset is accurate and consistent. To achieve this, I focused on cleaning and formatting the data. This process involved identifying and correcting any errors, removing duplicate entries, and organizing the data in a structured manner. By cleaning and formatting the data, we ensure its reliability and suitability for further analysis.

Joining Tables:

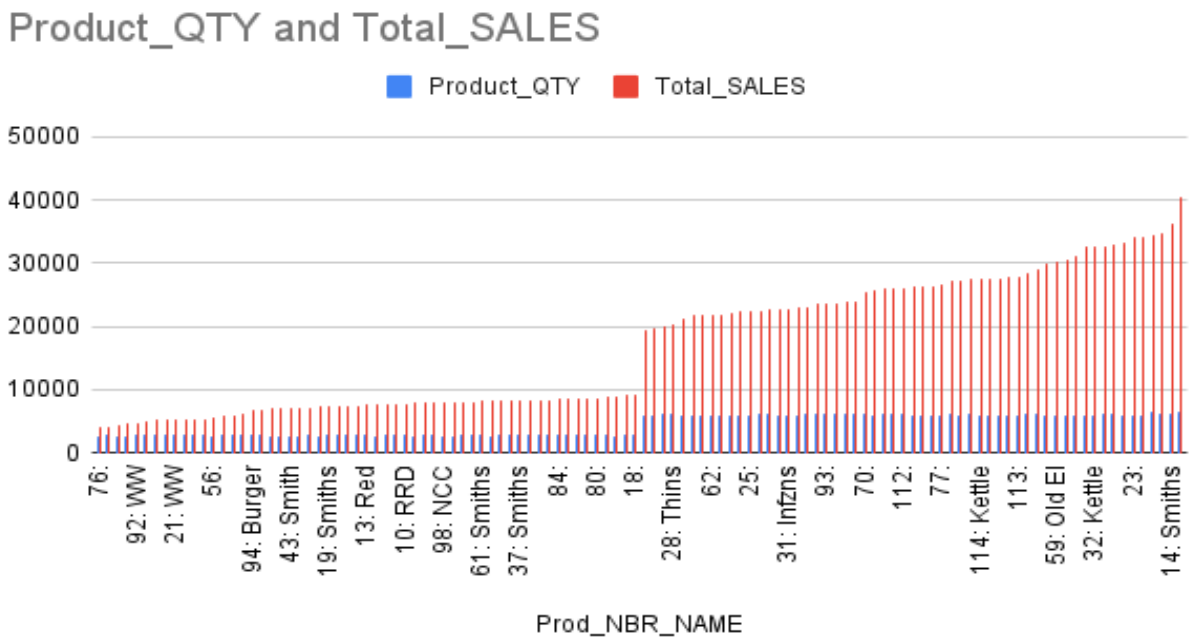
To gain a comprehensive understanding of the data, it was necessary to combine information from different tables and columns. By joining tables and columns, I merged relevant data based on common identifiers, such as customer IDs. This integration allowed me to create a unified dataset that contained consolidated information, facilitating more insightful analysis. This process involved utilizing techniques like SQL joins to combine the data effectively.

Select * From internship1-384609.Intern1.QVI_purchase_behaviour purchase44 INNER JOIN internship1-384609.Intern1.transaction_data transaction44 ON purchase44.LYLTY_CARD_NBR =transaction44.LYLTY_CARD_NBR	SELECT CONCAT(PROD_NBR, ": ", PROD_NAME) AS Prod_NBR_NAME, SUM(PROD_QTY) AS Product_QTY, SUM(TOT_SALES) AS Total_SALES FROM `internship1-384609.Intern1.Whole_datasheet` GROUP BY Prod_NBR_NAME ORDER BY Total_SALES ASC

Grouping and Aggregating:

To better understand the purchasing behavior of different customer segments, I grouped the data based on premium customer and lifestage categories. By aggregating the data within each segment, I obtained the total product quantity and total sales for each category. This grouping and aggregating process allowed me to identify patterns and trends specific to different customer groups. Analyzing data at this level of granularity provided valuable insights for business decision-making.

SELECT PREMIUM_CUSTOMER, SUM(PROD_QTY) AS Product_QTY, SUM(TOT_SALES) AS Total_SALES FROM `internship1-384609.Intern1.Whole_datasheet` GROUP BY PREMIUM_CUSTOMER ORDER BY Total_SALES ASC	SELECT LIFESTAGE, SUM(PROD_QTY) AS Product_QTY, SUM(TOT_SALES) AS Total_SALES FROM `internship1-384609.Intern1.Whole_datasheet` GROUP BY LIFESTAGE ORDER BY Total_SALES ASC



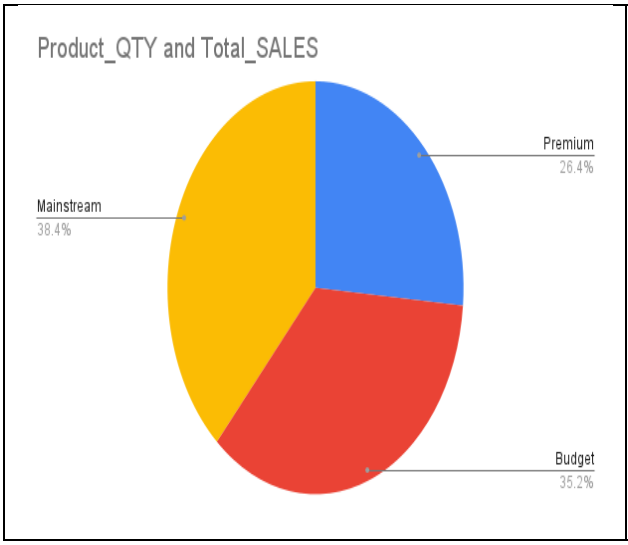
Analysis:

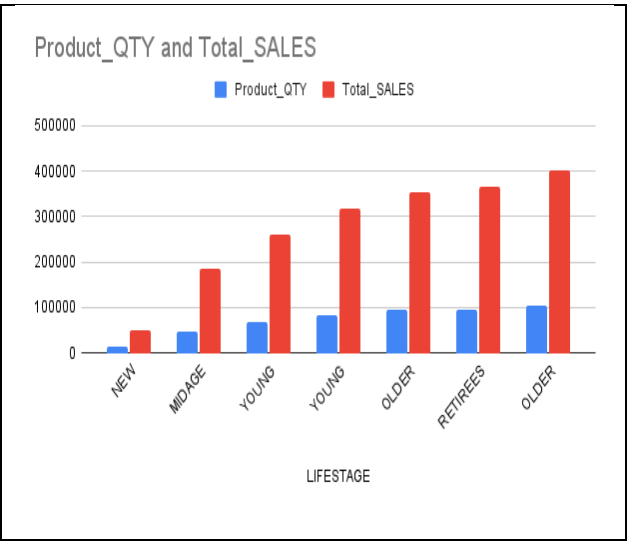
In order to delve deeper into the dataset, I conducted analysis to classify products based on their sales performance. By categorizing products into different sales levels, such as low sales, medium sales, high sales, and very high sales, I could identify the top-performing products as well as those in need of improvement. This analysis was essential for understanding which products were driving revenue and which ones required attention.

```
SELECT
CONCAT(PROD_NBR, ": ", PROD_NAME) AS Prod_NBR_NAME,
SUM(PROD_QTY) AS Product_QTY,
SUM(TOT_SALES) AS Total_SALES,
CASE
WHEN SUM(TOT_SALES) < 10000 THEN 'Low Sales'
WHEN SUM(TOT_SALES) >= 10000 AND SUM(TOT_SALES) < 20000 THEN 'Medium Sales'
WHEN SUM(TOT_SALES) >= 20000 AND SUM(TOT_SALES) < 30000 THEN 'High Sales'
WHEN SUM(TOT_SALES) >= 30000 THEN 'Very High Sales'
END AS Sales_Level
FROM `internship1-384609.Intern1.Whole_datasheet`
GROUP BY Prod_NBR_NAME
ORDER BY Total_SALES ASC
```

Visualization:

During my internship as a data analyst, I focused not only on analyzing and deriving insights from the dataset but also on effectively visualizing the findings. Two impactful visualizations were created to enhance understanding. The pie chart showcased the distribution of customers across premium, budget, and mainstream segments, providing stakeholders with a quick grasp of the customer base composition and guiding resource allocation and marketing strategies. Additionally, the histogram portrayed the distribution of product quantity and total sales within various lifestages, enabling decision-makers to identify patterns, trends, and potential areas for improvement or targeted marketing efforts. These visual representations played a pivotal role in conveying complex data insights, empowering decision-makers to make informed choices based on the visually compelling and easily understandable information.





Conclusion:

During my internship as a data analyst, I gained hands-on experience in working with large datasets and performing various analytical tasks. The process involved cleaning and formatting data to ensure its reliability, joining tables and columns to create a comprehensive dataset, grouping and aggregating data to understand customer segments, and conducting in-depth analysis to identify sales performance levels. These skills are crucial for extracting meaningful insights from data and supporting data-driven decision-making processes.