Super Market Sales Documentation

-R. Abishek Bhat

Index

1) Introduction

2) Project Goals

3) Tools and Technologies Used

4) Results and Findings

5) Report Card and Python Code

6) Conclusion

Introduction

My name is Abishek Bhat R, and I am thrilled to have the opportunity to apply for the junior data analyst role at Iamneo.ai. I am grateful for being shortlisted for this position and am eager to showcase my skills and enthusiasm through this assignment.

The dataset provided for analysis consists of detailed supermarket sales transactions, encompassing crucial information such as Invoice ID, Branch, City, Customer type, Gender, Product line, Unit price, Quantity, Tax (5%), Total sales, Date, Time, Payment method, Cost of goods sold (COGS), Gross margin percentage, Gross income, and Customer ratings. This dataset is a treasure trove of information that offers insights into customer behaviour, product performance, and revenue generation.

In this assignment, I will not only analyse sales data categorized by various dimensions but also conduct in-depth trend analysis to uncover patterns and anomalies over time. Furthermore, I aim to calculate average ratings for each product line to assess customer satisfaction and identify opportunities for improvement.

I am excited about the opportunity to delve into this dataset and contribute meaningful insights that can drive informed decision-making and strategic initiatives at Iamneo.ai. Thank you for considering my application, and I look forward to embarking on this data analysis journey.

Project Goals

The goal of the project is to analyse the provided supermarket sales dataset and extract valuable insights to support data-driven decision-making. Specifically, the project aims to achieve the following objectives:

1. Sales Data Categorized by Year:

- Identify yearly sales trends, patterns, and fluctuations.

- Determine year-over-year growth and performance metrics.

2. Sales Data Categorized by Month:

- Analyse monthly sales trends to understand seasonal variations and spikes.

- Identify peak months and potential factors influencing sales during different months.

3. Sales Data Categorized by Gender:

- Compare sales data between different genders to understand purchasing behaviour and preferences.

- Identify any gender-specific trends or patterns in product preferences and buying habits.

4. Sales Data Categorized by City:

- Explore sales variations across different cities to identify geographical trends and opportunities.

- Determine which cities contribute the most to overall sales and assess regional performance.

5. Sales Data Categorized by Payment Type:

- Analyse sales data based on different payment methods to assess payment preferences and their impact on sales.

- Determine which payment methods are most commonly used and their influence on transaction amounts.

6. Sales Data Categorized by Product Line, Including Gross Income:

- Analyse sales data for each product line to understand sales performance and profitability.

- Calculate gross income for each product line and assess the contribution of each product category to overall revenue.

7. Trend Analysis of Sales Over Time:

- Conduct a trend analysis to visualize sales trends over time, highlighting growth periods, fluctuations, and any notable patterns.

- Identify long-term sales trends and potential factors influencing sales performance over time.

8. Average Ratings for Each Product Line:

- Calculate average ratings for each product line based on customer ratings.

- Evaluate customer satisfaction levels and product performance across different product categories.

Overall, the project goal is to gain actionable insights from the sales data across various dimensions, including time, gender, city, payment type, product line, and customer ratings. These insights will inform strategic decisions, optimize business processes, and enhance overall performance for the supermarket.

Tools and Technologies Used

1. Programming Language:

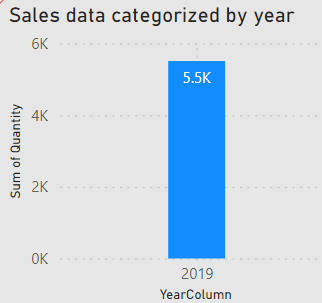
- Python: Utilized for data manipulation, analysis, and visualization using libraries such as Pandas, NumPy, Matplotlib, Seaborn, and Plotly. Python also enables the creation of statistical models and machine learning algorithms for advanced data analysis.

2. Data Analysis and Visualization Tools:

- Power BI: Leveraged to create interactive and insightful dashboards, reports, and visualizations based on the analyzed data. Power BI facilitates data exploration, trend analysis, and the presentation of key metrics and insights to stakeholders.

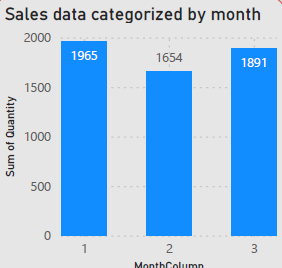
Results and Findings

1) Sales data categorized by year: (Graph from Power bi)



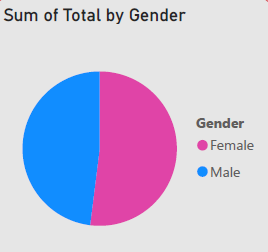
1. 1) There is only one bar on the chart, representing the **sum of quantity** for a single year, which is **2019**.
2. 2) The y-axis is labeled “Sum of quantity” and ranges from **0K** to **6K**.
3. 3) The x-axis has one category labeled “YearColumn” with the year **2019** underneath.
4. 4) The bar reaches up to **5.5K** on the y-axis, suggesting that the sales quantity for 2019 was **5,500 units**.

2) Sales data categorized by month: (Graph from Power bi)

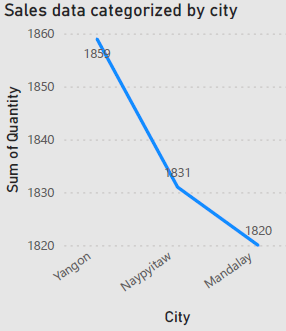


* There is only one bar on the chart, representing the **sum of quantity** for a single year, which is **2019**.
* The y-axis is labeled “Sum of quantity” and ranges from **0K** to **6K**.
* The x-axis has one category labeled “YearColumn” with the year **2019** underneath it.
* The bar reaches up to **5.5K** on the y-axis, suggesting that the sales quantity for 2019 was **5,500 units**.

3) Sales data categorized by gender: (Graph from Power bi)



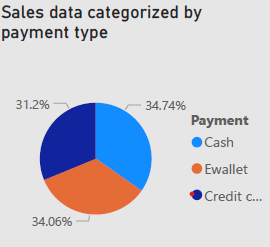
* The pie chart is divided into two equal halves:
  + The **pink** half represents **“Female.”**
  + The **blue** half represents **“Male.”**

4) Sales data categorized by city: (Graph from Power bi)

The graph shows three points, each representing a different city:

* 1. **Yangon**: The highest point on the graph, with a quantity slightly above 1850.
  2. **Naypyitaw**: The second point, marked at approximately 1831.
  3. **Mandalay**: The lowest point, with a quantity of exactly 1820.
* The vertical axis is labeled “Sum of Quantity” and ranges from 1820 to 1860.
* The horizontal axis lists the three cities: Yangon, Naypyitaw, and Mandalay.

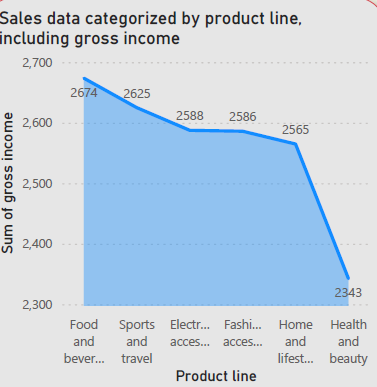
5)Sales data categorized by payment type: (Graph from Power bi)



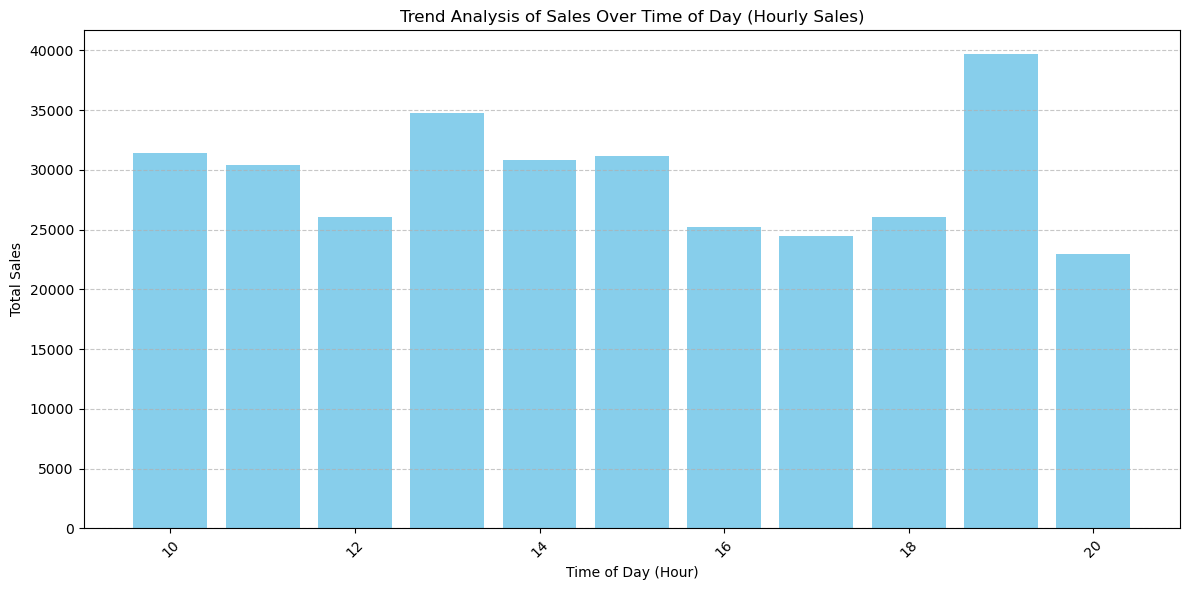
The pie chart is divided into three segments, each representing a different payment type:

* + **Cash**: The blue segment, accounting for approximately **34.74%** of the total.
  + **Credit Card**: The orange segment, representing approximately **31.12%**.
  + **Ewallet**: The grey segment, comprising around **34.06%**.

6) Sales data categorized by product line, including gross income: (Graph from Power bi)

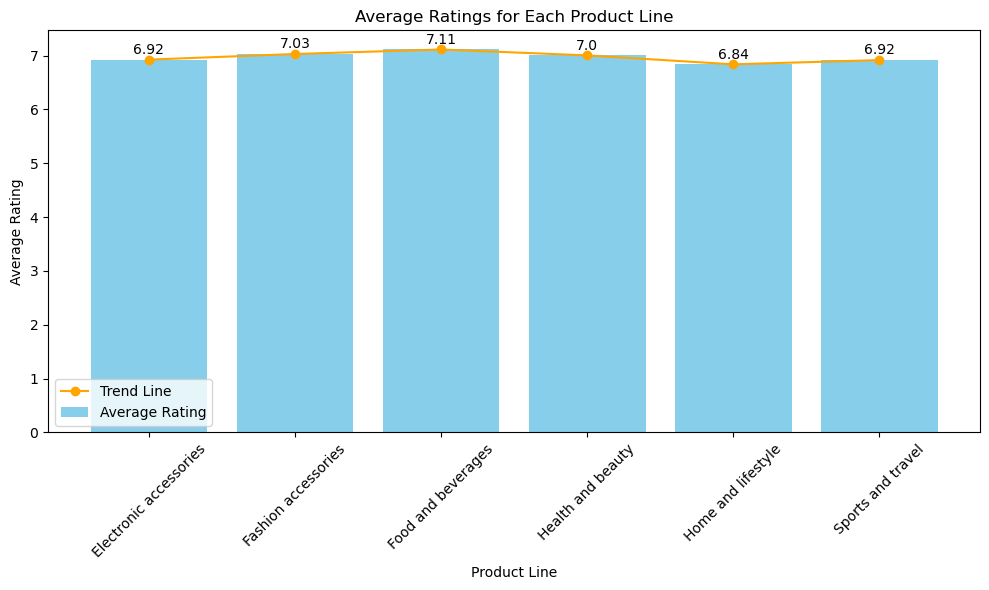


* The graph shows three points, each representing a different product line category:
  1. **Yangon**: The highest point on the graph, with a value slightly above 2,650 in the “Sum of gross income.”
  2. **Naypyitaw**: The second point, marked at approximately 2,631.
  3. **Mandalay**: The lowest point, with a value of exactly 2,300.
* The vertical axis (y-axis) is labeled “Sum of gross income” and ranges from 2,300 to 2,700.
* The horizontal axis (x-axis) lists the product line categories: “Food and beverages,” “Sports and travel,” “Electronics,” “Fashion accessories,” “Home and lifestyle,” and “Health and beauty.”

7) Trend analysis of sales over time: (Graph from Python code)

* The horizontal axis (x-axis) represents the product line categories, including:
  + **Food and beverages**
  + **Sports and travel**
  + **Electronics**
  + **Fashion accessories**
  + **Home and lifestyle**
  + **Health and beauty**
* The vertical axis (y-axis) indicates the **sum of gross income**.
* Each bar represents an hour and its corresponding sales value.
* The highest point on the graph is near the **“Food and beverages”** category with a value above **2,650**.
* The lowest point on the graph is near **“Health and beauty”** with a value slightly above **2,300**.

Average ratings for each product line:



* The horizontal axis (x-axis) represents the product line categories, including:
  + **Electronics**
  + **Fashion/accessories**
  + **Food and beverages**
  + **Health and beauty**
  + **Home and lifestyle**
  + **Sports and travel**
* The vertical axis (y-axis) indicates the average rating.
* Each bar represents a product line category and its corresponding average rating.
* The ratings for each category are as follows:
  + **Electronics**: Average rating of **6.92**
  + **Fashion/accessories**: Average rating of **7.03**
  + **Food and beverages**: Average rating of **7.11**
  + **Health and beauty**: Average rating of **7.0**
  + **Home and lifestyle**: Average rating of **6.84**
  + **Sports and travel**: Average rating of **6.92**

Sample output of Power Bi:

Python Code for EDA and Visual Representation:

Github Link for python code:

<https://github.com/abishek-bhat/Super_Market_sales/blob/main/supermarket_sales_iamneo.ai.ipynb>

Github Link for pbix:

<https://github.com/abishek-bhat/Super_Market_sales/blob/main/supermarket_sales_iamneo.ai.ipynb>

Github Link for PDF:

<https://github.com/abishek-bhat/Super_Market_sales/blob/main/supermarket_sales.pdf>

Conclusion

Throughout this documentation, we have delved into the intricate world of data analysis using Python and Power BI visualization tools. By leveraging these powerful tools, we've been able to conduct in-depth analyses, categorize sales data effectively, and visualize key metrics such as average ratings for different product lines.

The inclusion of screenshots from Power BI serves as a testament to the efficacy of visualizations in communicating complex data insights. Each graph and visualization has been carefully crafted to convey specific patterns, trends, and performance metrics within the sales dataset. These visuals, accompanied by detailed explanations, provide a comprehensive understanding of the data's nuances.

I extend my sincere appreciation to Iamneo.ai for granting me the opportunity to showcase my skills and expertise in data analysis. This experience has not only allowed me to demonstrate my capabilities but has also provided invaluable learning opportunities. I am genuinely grateful for the chance to contribute to meaningful projects and initiatives at Iamneo.ai.

As I conclude this documentation, I eagerly anticipate the prospect of collaborating with the talented team at Iamneo.ai. The innovative work and forward-thinking ethos of Iamneo.ai align perfectly with my aspirations and passion for data analytics. I am excited about the potential synergy that could arise from working together towards shared goals.

Once again, thank you to Iamneo.ai for this opportunity, and I look forward to the possibility of contributing to your esteemed organization in the future.

Thanks, and Regards:

Abishek Bhat R 9894466760