



Big Data and Business Intelligence

Supermarket Insights: Analyzing Sales, Products, and Customer Feedback

Name: Ogundeko Ayo-Oluwa Aduragbemi

Student ID: S3258780

1. Executive summary.....	3
2. Body	4
2.1 Introduction	4
2.1.1 Data source.....	4
2.1.2 BI Questions.....	6
2.2 Finding based on analysis and evaluation	7
Q1 What is the average customer rating, and what factors influence those ratings?	7
Q2 How do total sales and profit break down across the three months, and how are they distributed across the three branches?	13
Q3 What is the predicted sales performance for April, and how can the company prepare for this?	16
Q4 Which product line has the highest impact on profitability, and how does this contribute to overall performance?	17
Q5 Which gender generated the most revenue, and how does this inform marketing strategies?	23
Q6 How are total sales distributed across different payment methods, and what trends can be observed regarding customer preferences?.....	25
3. Conclusion	30
3.1. Recommendations.....	30
3.2. Personal conclusion	31
4. Appendix: BI Design	32
4.1 Data Pre-Processing and data cleaning	32
4.2 BI Data Modelling via Star Schema - Facts and Dimensions.....	42
4.3 DAX and M Language	45
DAX Measures Created	45
4.4 Dashboard	51
5 Self-assessment	56

1. Executive summary

Introduction

This report analyzes sales, satisfaction, and profitability data from a retail business, identifying key trends and insights. It explores customer satisfaction, sales distribution, and product line profitability. This report analyzes key business metrics to address the following questions:

- What is the average customer rating, and what factors influence these ratings?
- How are sales and profits distributed across months and branches?
- What is the sales forecast for April, and how should the company prepare?
- Which product line contributes most to profitability?
- Which gender generates the most revenue, and what does this mean for marketing?
- How are sales distributed across payment methods, and what trends emerge?

Key Findings

A key highlight is that 84.7% of customers are satisfied, as shown in the pie chart. However, the 15.3% dissatisfaction rate also requires investigation to identify areas for improvement



Other findings include:

- Food and Beverages emerged as the most profitable product line, contributing 17.38% to total profit.
- January recorded the highest revenue at £121.83k, indicating peak performance.

Recommendations

- Investigate the factors contributing to January's high revenue and implement similar strategies in other months.
- Promote and expand the Food and Beverages product line to capitalize on its profitability.

2. Body

2.1 Introduction

This report analyzes key performance metrics for a supermarket chain, focusing on factors that impact customer satisfaction, sales, and profitability. The goal is to extract insights from the data that can help shape future business strategies. The analysis covers various aspects such as customer ratings, sales trends, product performance, and payment methods, with the aim of identifying patterns that can guide the company's growth and improve customer experience.

The report aims to answer several important questions related to customer satisfaction, sales performance, and business development. By exploring these questions, we seek to uncover actionable insights that can inform better decision-making and enhance overall business performance. The following key questions are addressed in the report:

- What is the average customer rating, and what factors influence those ratings?
- How do total sales and profit break down across the three months, and how are they distributed across the three branches?
- What is the predicted sales performance for April, and how can the company prepare for this?
- Which product line has the highest impact on profitability, and how does this contribute to overall performance?
- Which gender generated the most revenue, and how does this inform marketing strategies?
- How are total sales distributed across different payment methods, and what trends can be observed regarding customer preferences?

2.1.1 Data source

The dataset used for this project is from Kaggle, a popular platform for datasets. It is called "Supermarket Sales" and contains 17 columns and 1,000 rows. The data covers sales transactions from a supermarket chain with three branches over three months (January to March 2019). It includes details on customer behavior, product categories, sales, and financial metrics, making it ideal for Business Intelligence (BI) and analytics.

Key details:

- **Source:** [Kaggle - Supermarket Sales Dataset](#)
- **Format:** CSV file

- **Purpose:** Analyzing sales, customer behaviors, and product performance.

Database and Table Description

The dataset was divided into multiple tables for easier analysis in the BI solution. Below is a summary of the tables and their attributes:

Table Name	Columns
Sales_Fact Table	Invoice ID, Total, Quantity, Tax 5%, Gross Income, COGS
Product Table	Product Line, Unit Price
Customer Table	Customer Type, Gender, Rating
Branch Table	Branch, City
Date Table	Date, Time

Data Characteristics

- **Invoice ID:** Computer-generated sales slip invoice identification number.
- **Branch:** Branch of the supercenter (3 branches identified by A, B, and C).
- **City:** Location of the supercenters.
- **Customer Type:** Type of customers, recorded as Members (using a member card) or Normal (without a member card).
- **Gender:** Gender type of customer.
- **Product Line:** General item categorization groups, such as Electronic Accessories, Fashion Accessories, Food and Beverages, Health and Beauty, Home and Lifestyle, Sports and Travel.
- **Unit Price:** Price of each product in dollars.
- **Quantity:** Number of products purchased by the customer.
- **Tax:** 5% tax fee for customer purchases.
- **Total:** Total price including tax.
- **Date:** Date of purchase (records available from January 2019 to March 2019).
- **Time:** Purchase time (from 10 AM to 9 PM).
- **Payment:** Payment method used by the customer (3 methods available – Cash, Credit Card, and E-wallet).
- **COGS (Cost of Goods Sold):** Cost of goods sold for the transaction.
- **Gross Margin Percentage:** Gross margin percentage for the product.
- **Gross Income:** Gross income derived from the sales.

- **Rating:** Customer satisfaction rating based on their overall shopping experience (on a scale from 1 to 10).

2.1.2 BI Questions

1. What is the average customer rating, and what factors influence those ratings?
 - a. What is the average customer rating, and what factors influence those ratings?
 - b. Customer Satisfaction Breakdown: Analysis of the distribution of customer satisfaction ratings across different score ranges (e.g., 1-3, 4-6, 7-10).
 - c. What is the average customer rating by gender
2. How do total sales and profit break down across the three months, and how are they distributed across the three branches?
 - a. Monthly Sales and Profit Distribution: Comparison of sales and profit trends over three months (January to March).
 - b. How are sales and profits distributed among the three branches (A, B, and C)?
3. What is the predicted sales performance for April, and how can the company prepare for this?
 - a) Sales predictions are based on trends observed from January to March.
 - b) Risk Assessment: What risks are involved in relying on historical data for forecasting April's sales?
4. Which product line has the highest impact on profitability, and how does this contribute to overall performance?
 - a) Identification of the product line with the highest profit contribution.
 - b) Product Line Preference by Gender: How do male and female customers prefer different product lines, and how does this affect sales and profitability?
 - c) Product Line Contribution: What role does each product line play in overall sales performance, including the total cost of goods sold, the total volume of goods sold, and the average unit price?
5. Which gender generated the most revenue, and how does this inform marketing strategies?
 - a. Revenue Comparison by Gender: Breakdown of total sales by gender to identify which gender generated more revenue.
 - b. Average Product Line Rating: What is the average customer satisfaction rating for different product lines, and how does this vary between male and female customers?

6. How are total sales distributed across different payment methods, and what trends can be observed regarding customer preferences?

- a) **Payment Method Breakdown:** Distribution of total sales based on the payment method (Cash, Credit Card, Ewallet).
- b) **Customer Preferences:** Which payment methods are most preferred by customers?
- c) **Implications for Payment Systems:** How should the company adapt to changing payment preferences (e.g., promoting eWallet or offering discounts for certain methods)?

2.2 Finding based on analysis and evaluation

This section explores the critical insights derived from the Power BI dashboards. Each subsection delves into specific metrics, providing a detailed explanation of the data, the visuals used, and the rationale for including them in the analysis.

All formulas and calculations used in this project, including those for creating these measures and columns, will be explained in detail in the Appendix section of this project.

Q1 What is the average customer rating, and what factors influence those ratings?

The average customer rating, average female rating, and average male rating were calculated using Dax formulas and stored in the Customer table for future reference and analysis. Additionally, a rating category column was created in Power BI using M language to categorize the ratings into ranges for enhanced breakdown and analysis.

a. What is the average customer rating, and what factors influence those ratings



Figure 1: Displays the average customer rating

Title	Average Customer rating
Chart type	Infographic Designer Chart
Reason	To visually represent the average rating with an engaging icon (smiling face), making it easier to communicate the general satisfaction level.
Findings	The average customer rating is 6.97 , approximately 7/10 , representing a good score. Hovering over the chart displays the precise value of 6.97 .

← Rating Category is more likely to be Average when Product line is Sports and travel than otherwise (on average).

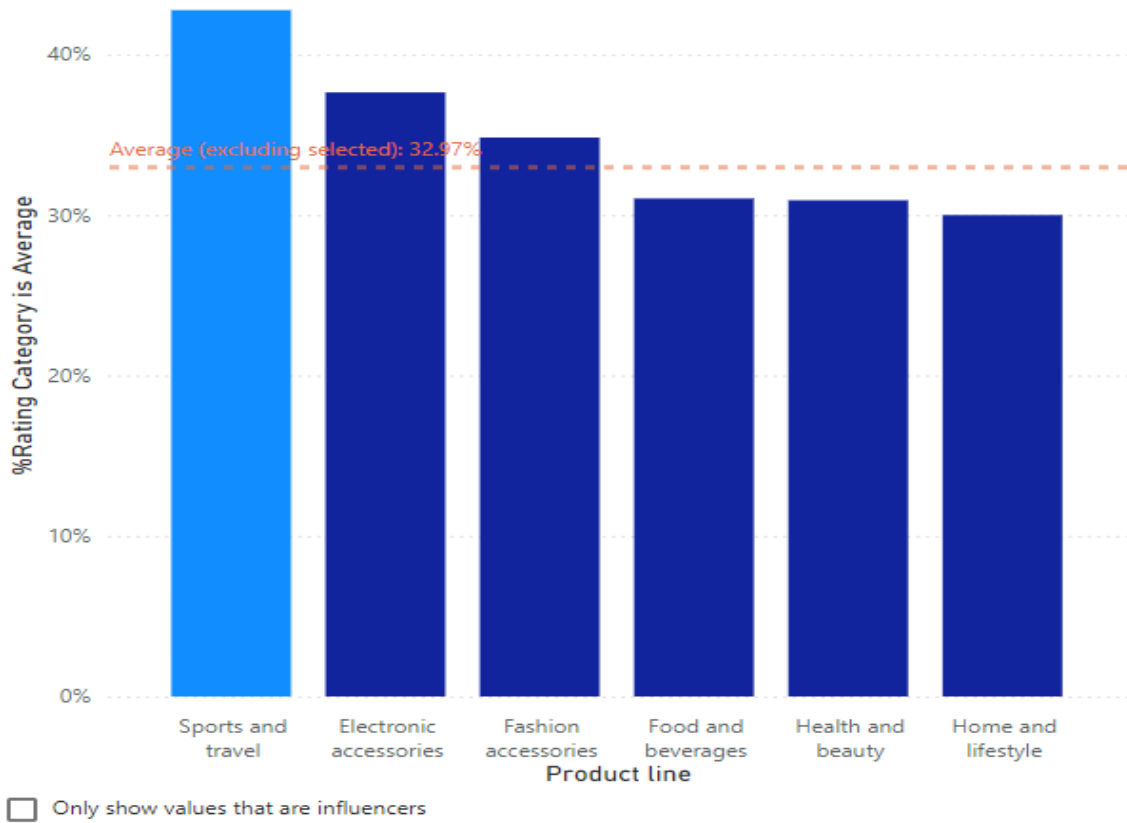
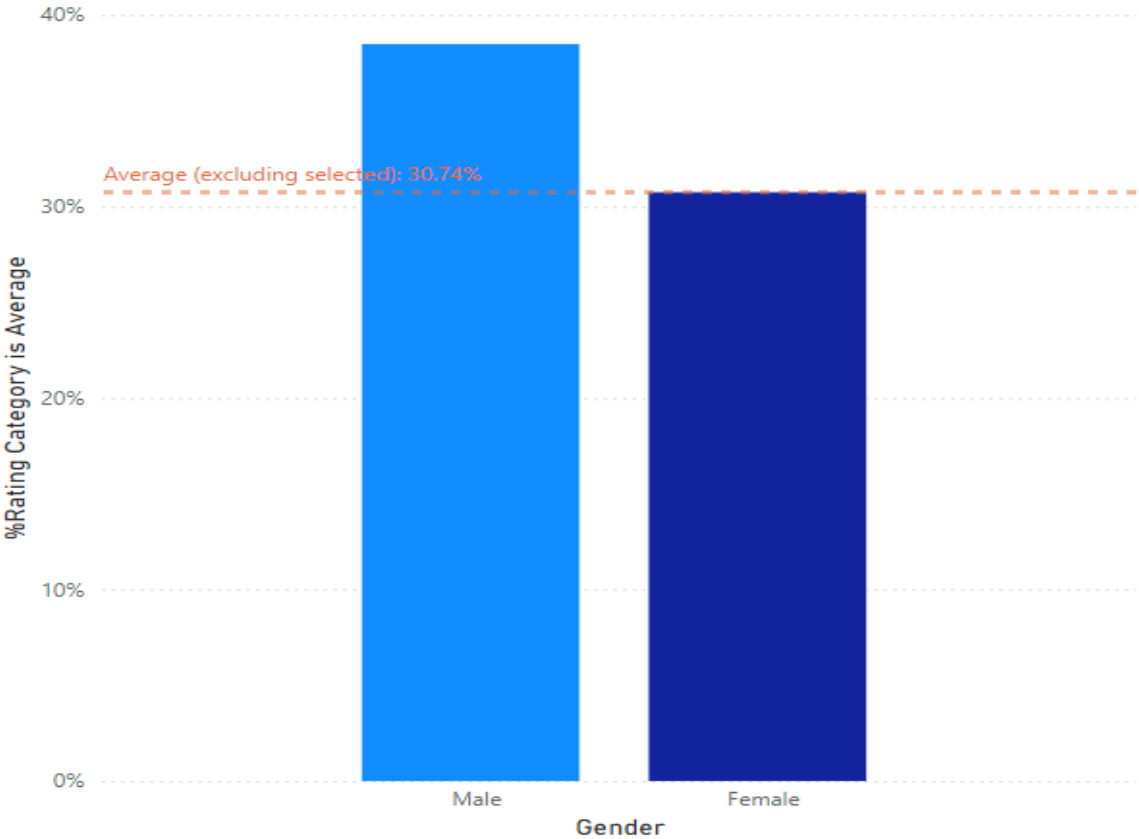


Figure 2: Customer satisfaction factors on product line

Title	Customer Satisfaction Factors
Chart type	Key Influencer Chart
Reason	To analyze the factors influencing the rating categories and identify key insights

	related to customer satisfaction on product line.
Findings	Ratings are more likely to be average when the product line is Sports and Travel (on average).

← Rating Category is more likely to be Average when Gender is Male than otherwise (on average).



☐ Only show values that are influencers

Figure 2: Customer satisfaction factors on gender

Title	Customer Satisfaction Factors
Chart type	Key Influencer Chart
Reason	To visually represent the average rating with an engaging icon (smiling face), making it easier to communicate the general satisfaction level on gender.
Findings	Ratings are more likely to be average when the gender is Male (on average).

b. Customer Satisfaction Breakdown: Analysis of the distribution of customer satisfaction ratings across different score ranges (e.g., 1-3, 4-6, 7-10).



Figure 3: Breakdown of customer satisfaction.

Title	Breakdown of Customer Satisfaction Rating.
Chart type	Pie Chart
Reason	To visualize the distribution of customer satisfaction ratings across defined score ranges, highlighting the percentage of responses categorized as good or poor.
Findings	<ul style="list-style-type: none">Based on a total of 1,000 customer responses, 15.3% rated their experience below 5, which is considered poor.The remaining 84.7% provided a rating of 5 or above, which is classified as good, indicating a largely positive customer experience overall.

c. What is the average customer rating by gender

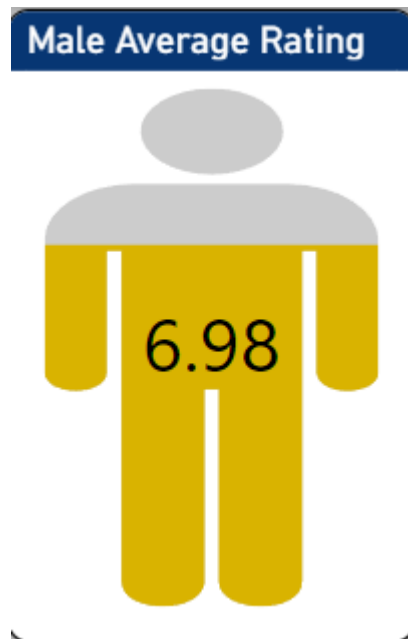


Figure 4: Average male rating.

Title	Male Average rating
Chart type	Infographic Designer Chart (with Male Icon)
Reason	To visually represent the average rating given by male customers using a recognizable icon for clarity and appeal.
Findings	The average male rating is 6.98, indicating that male customers tend to provide slightly higher satisfaction scores.

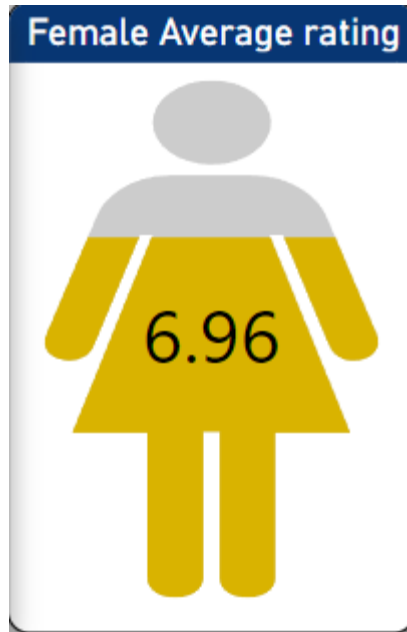


Figure 5: Average Female rating.

Title	Male Average rating
Chart type	Infographic Designer Chart (with Female Icon)
Reason	To visually represent the average rating given by female customers using a gender-specific icon for ease of interpretation.
Findings	The average female rating is 6.96, which is closely aligned with the male rating but marginally lower.

Q2 How do total sales and profit break down across the three months, and how are they distributed across the three branches?

The Total Sales measure was calculated and stored in Power BI to aggregate sales data from the "Total" column, providing a clear understanding of overall sales performance. This measure is essential for analyzing sales trends and comparing them across different branches and time periods.

The Gross Margin Percentage was derived to assess profitability, as the dataset required transforming the provided gross margin data into a usable format. Additionally, the Average Sales Across Cities was calculated and stored in the Branch table to evaluate and compare the performance of each branch. These metrics are critical for understanding sales and profit distribution and will be further explained in the appendix section.

a. Monthly Sales and Profit Distribution: Comparison of sales and profit trends over three months (January to March).

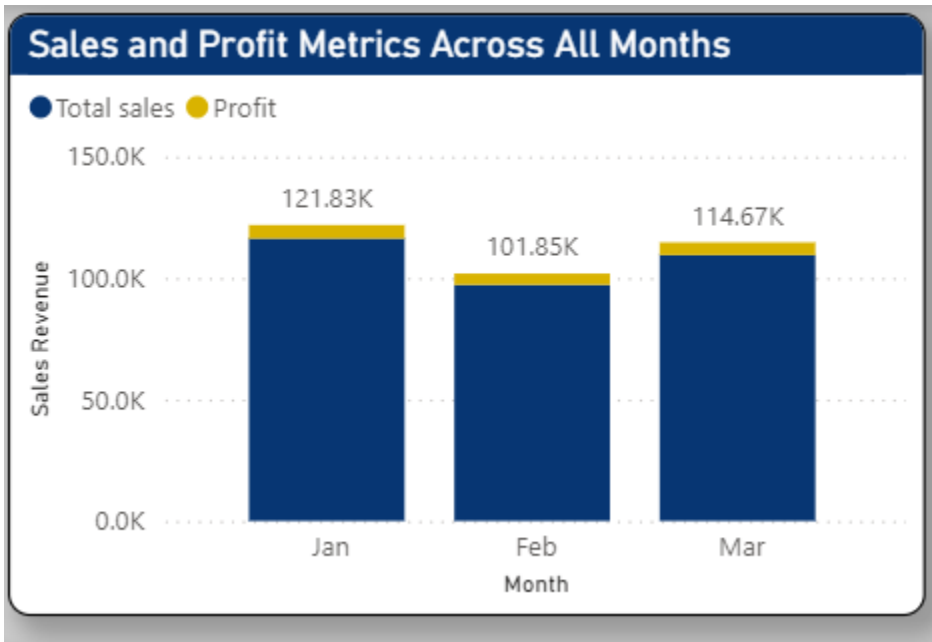


Figure 6: Stacked Column Chart (Financial Performance Overview)

Title	Sales and Profit Metrics Across All Months
Chart type	Stacked Column Chart
Reason	To visually compare the total sales and profit across the three months, providing insights into monthly revenue trends and profitability.

Findings	<ul style="list-style-type: none">January recorded the highest total revenue, reaching £121.83k, indicating a strong start to the quarter.February had the lowest total revenue at £101.85k, reflecting a dip in sales performance.March's revenue was moderate, falling between January and February, showcasing consistent but not peak performance.
----------	--

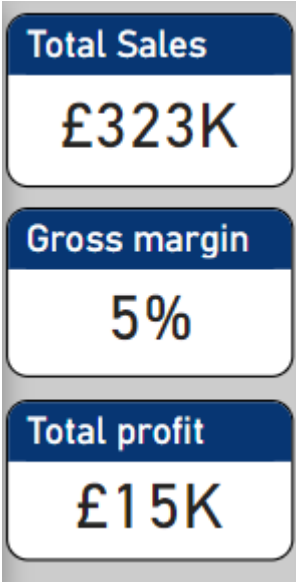


Figure 7: Sales and Profit Metrics Across All Months using card visuals

Title	Total Sales, Gross Margin Percentage, and Total Gross Income
Chart type	Single-Value Cards
Reason	To provide a quick and focused visualization of key financial metrics for better understanding and decision-making.
Findings	<ul style="list-style-type: none">Total Sales: The total sales across the dataset amount to approximately £323k, reflecting overall revenue generated by the supermarket chain.Total Gross Income: The gross income stands at approximately £15k, indicating the total profit, after accounting for the cost of goods sold.Gross Margin Percentage: The gross margin percentage is calculated at 4.76%, which can be

	approximated to 5% , showing a modest profit margin.
--	---

b. How are sales and profits distributed among the three branches (A, B, and C)

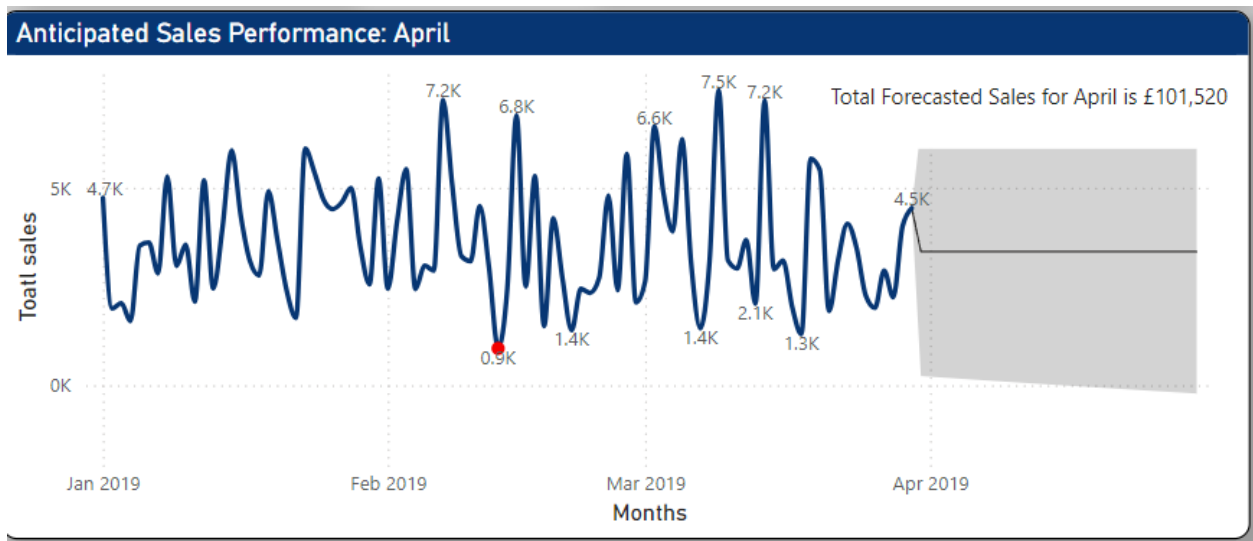


Figure 8: Map Visualization (Sales Distribution Across Branches and Cities)

Title	Sales Distribution Across Cities
Chart type	Map Chart
Reason	To visually display the sales and profit distribution across the three branches, providing geographic context and clarity on branch performance.
Findings	<ul style="list-style-type: none">• Average Sales Across Cities: The average sales across all three branches amount to £108k, reflecting the overall sales performance of the supermarket chain across its locations.• Branch C Performance: Branch C leads in sales with £110.56k in revenue and £5,265 in profit, outperforming the other branches in both total sales and profitability.

Q3 What is the predicted sales performance for April, and how can the company prepare for this?

a. Sales predictions are based on trends observed from January to March



Title	Anticipated sales performance: April
Chart type	Line Chart
Reason	To visualize sales trends observed from January to March and predict future sales for April based on past data.
Findings	<ul style="list-style-type: none">• Highest Sales Day: The highest sales recorded during the analysis period was 7.5K on March 9th. This represents the peak sales day within the three-month period.• Lowest Sales Day: The lowest sales observed was 934 on March 13th, marking the day with the least sales during the period.• Forecasted Average Daily Sales for April: Based on the trends from January to March, the forecasted average sales per day for April is estimated at 3,384. This value is calculated from the historical data trends.• Total Forecasted Sales for April: Using the daily forecast value of 3,384, the total forecasted sales for April amount to 101,520 (3,384 * 30 days).

b. Risk Assessment: What risks are involved in relying on historical data for forecasting April's sales?

Relying only on historical data for forecasting April's sales can pose risks due to various external factors that may not be reflected in past trends. While data from January to March provides valuable insights, it does not account for seasonal changes, holidays, economic shifts, or unforeseen events that could impact sales. Customer behavior and market conditions may also change, making previous trends less reliable. Additionally, inaccuracies in the data, such as missing or incorrect information, could lead to misleading forecasts. To improve forecast accuracy, it's important to consider these external influences and combine historical data with broader market analysis and expert judgment.

Q4 Which product line has the highest impact on profitability, and how does this contribute to overall performance?

To analyze the product line's impact on profitability, DAX measures were used to calculate each product line's contribution to profit. This involved aggregating the gross income for each product line to assess profitability. Additionally, the top product line by profit was identified by comparing the total profit of each product line. These calculations provide insights into which product lines drive the supermarket's overall performance.

a. Identification of the product line with the highest profit contribution.

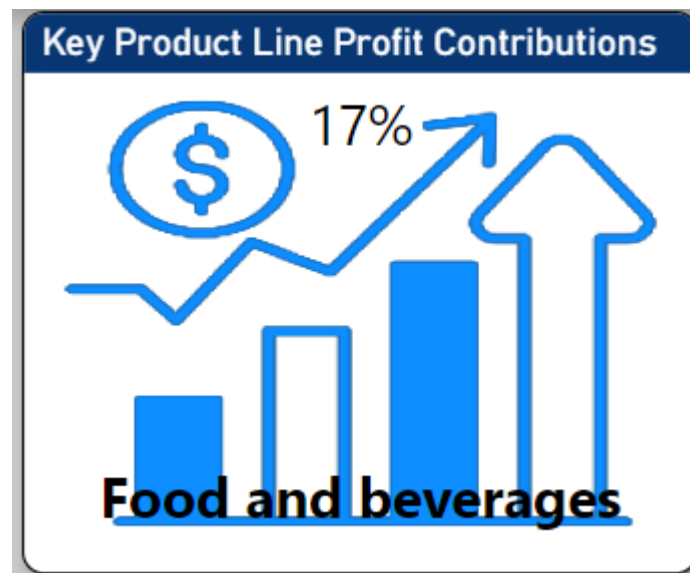


Figure 10: Top product line

Title	Key product line profit contributions
Chart type	Infographic Designer Chart
Reason	To visually highlight the product line with the highest profit contribution, using an infographic chart for easy interpretation.
Findings	The "Food and Beverage" product line emerged as the top performer, contributing approximately 17.38% to the overall profit, which rounds to 17%. This indicates that Food and Beverage plays a significant role in driving the supermarket's profitability.

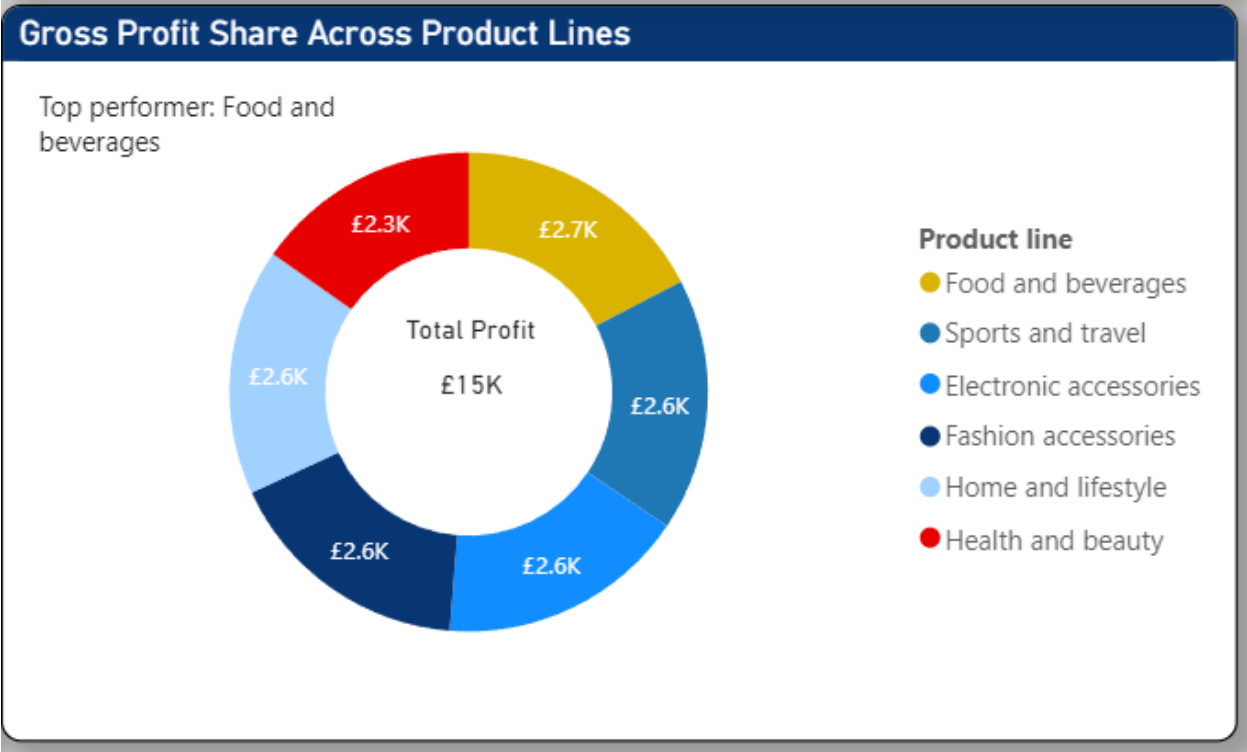


Figure 11: Product Line Preference by Gender Stacked Bar Chart

Title	Gross profit shared across product line
Chart type	Donut Chart
Reason	The donut chart was used to visually represent the contribution of each product line to overall profit, offering a clear and interactive comparison.

Findings	<ul style="list-style-type: none">• "Food and Beverage" led with the highest profit of 2.7k pounds, making the largest contribution to overall profitability.• "Health and Beauty" had the lowest profit contribution at 2.3k pounds, accounting for 15.23% of the total profit. This highlights the disparity in profitability between the different product lines.
----------	---

b. How do male and female customers prefer different product lines, and how does this affect sales and profitability?

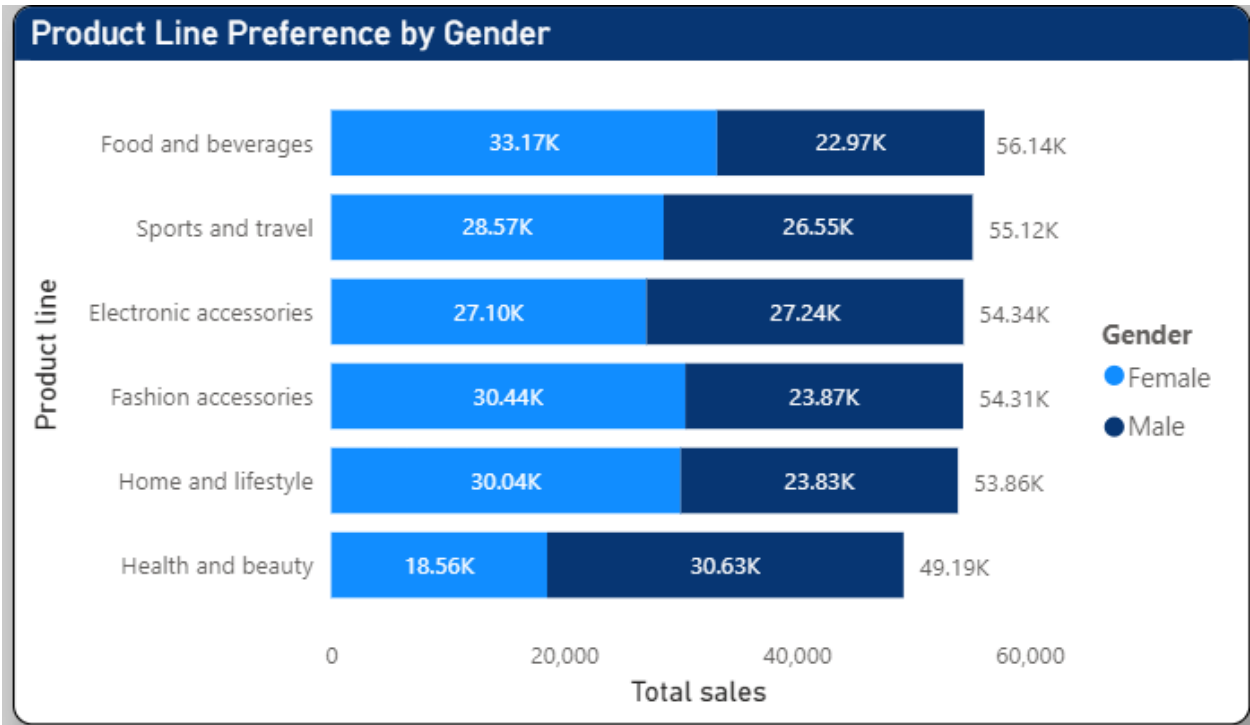


Figure 12: Product Line Preference by Gender Stacked Bar Chart

Title	Product Line Preference by Gender
Chart type	Stacked Bar Chart
Reason	The stacked bar chart was used to visually compare male and female customers' preferences across different product lines, allowing easy identification of sales trends by gender.
Findings	<ul style="list-style-type: none">• "Food and Beverages" generated the highest total sales across all product lines.• Female customers contributed more sales to "Food and Beverages" than any other product line.

	<ul style="list-style-type: none">Male customers generated the highest sales in "Health and Beauty," indicating a preference for this product line compared to others.
--	--

A Product Line slicer was added to the dashboard for better filtering and visualization of sales by product category. Additionally, a Gender slicer and a Date slicer were included to compare male and female customer preferences over different periods. These enhancements allow for a more detailed analysis of how product lines, gender, and time affect sales and profitability.

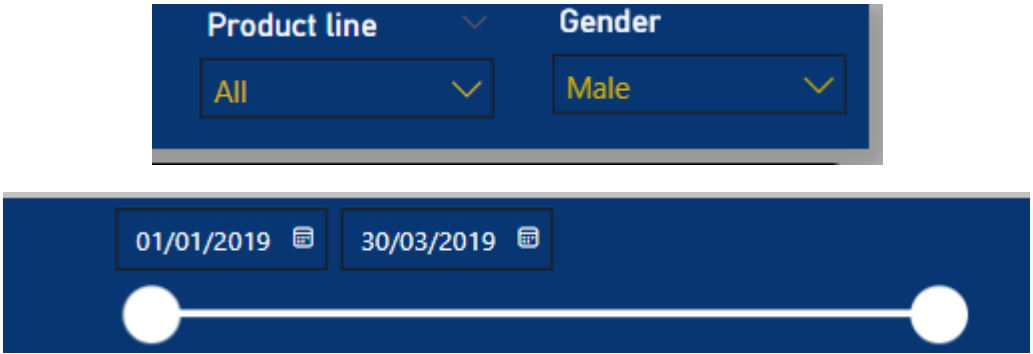


Figure 13 and 14: Necessary slicers added

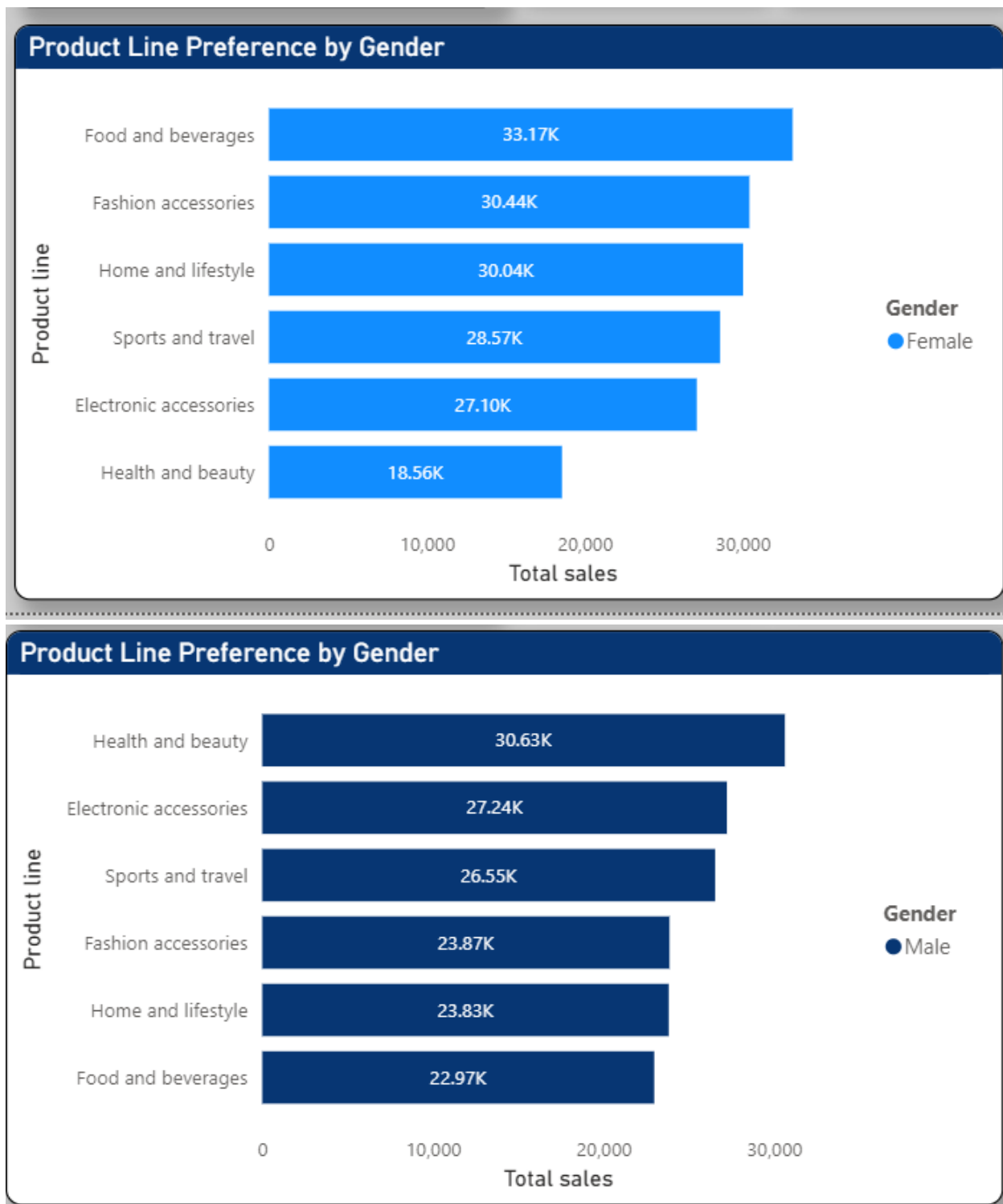


Figure 13 and 14: Preference by Gender Stacked Bar Chart when gender slicer is used.

- c. What role does each product line play in overall sales performance, including the total cost of goods sold, the total volume of goods sold, and the average unit price?

To evaluate each product line's impact on overall performance, cards were used to display metrics such as Total Cost of Goods Sold, Total Volume Sold, Average Unit Price, and Total Transactions. The Product Line slicer enables easy filtering for detailed analysis. This helps assess the contribution of each product line to the overall sales performance.

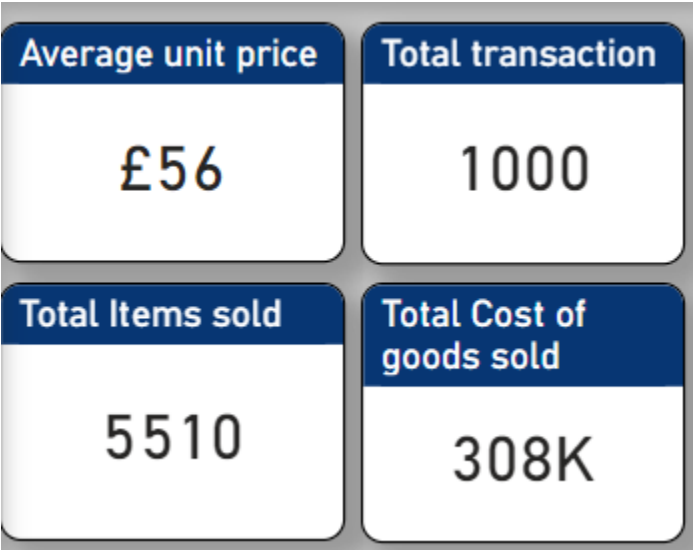


Figure15 Product Line Role in Overall Sales Performance

Title	Contribution of Product Lines to Sales Metrics
Chart type	Card Visual
Reason	Card visuals were employed to summarize essential metrics such as total cost of goods sold, total items sold, average unit price, and transaction count, providing a clear overview of product line contributions to overall performance.
Findings	<ul style="list-style-type: none">• The average unit price across product lines is 56 pounds.• Total transactions recorded are 1,000.• A total of 5,510 items were sold.• The total cost of goods sold reached 308k.

Q5 Which gender generated the most revenue, and how does this inform marketing strategies?

a. Revenue Comparison by Gender: Breakdown of total sales by gender to identify which gender generated more revenue.

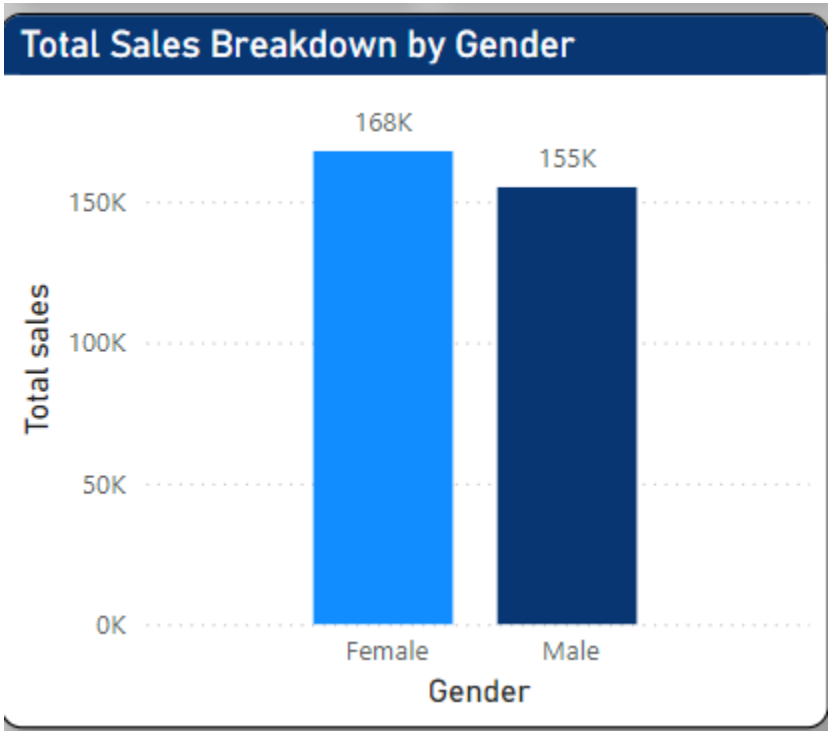
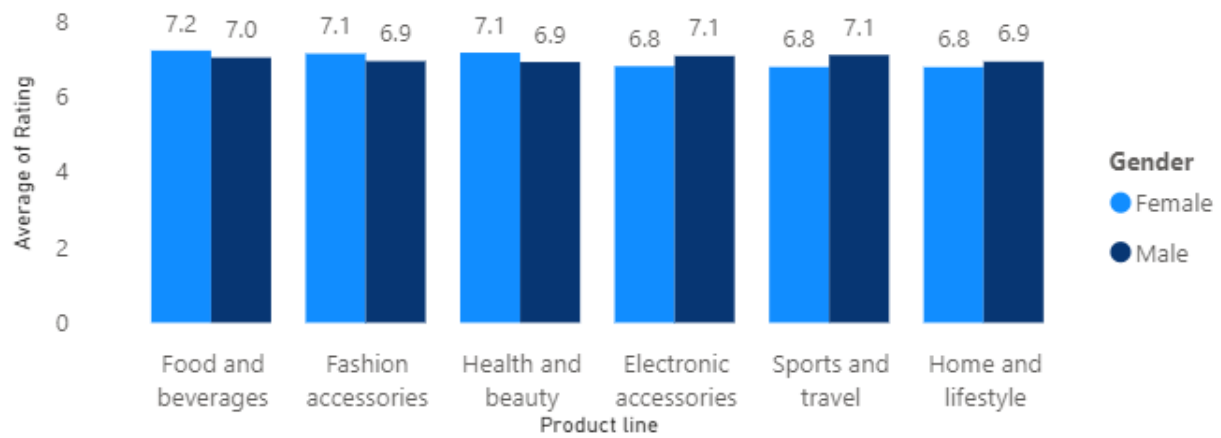


Figure16 Comparison by Gender

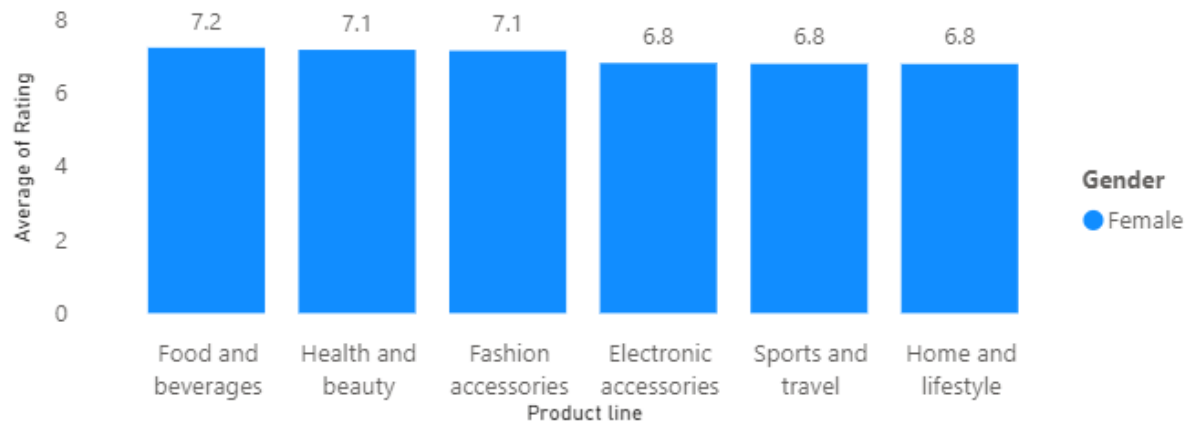
Title	Total sales Breakdown by Gender
Chart type	Stacked Column Chart
Reason	The stacked column chart effectively visualizes and compares total sales revenue by gender, providing a clear breakdown of contributions.
Findings	<ul style="list-style-type: none">Female customers generated the highest sales revenue, totaling £168k.Male customers, while slightly behind, contributed £155k to total sales revenue.

b. Average Product Line Rating: What is the average customer satisfaction rating for different product lines, and how does this vary between male and female customers?

Average Customer Satisfaction by Product Line



Average Customer Satisfaction by Product Line



Average Customer Satisfaction by Product Line

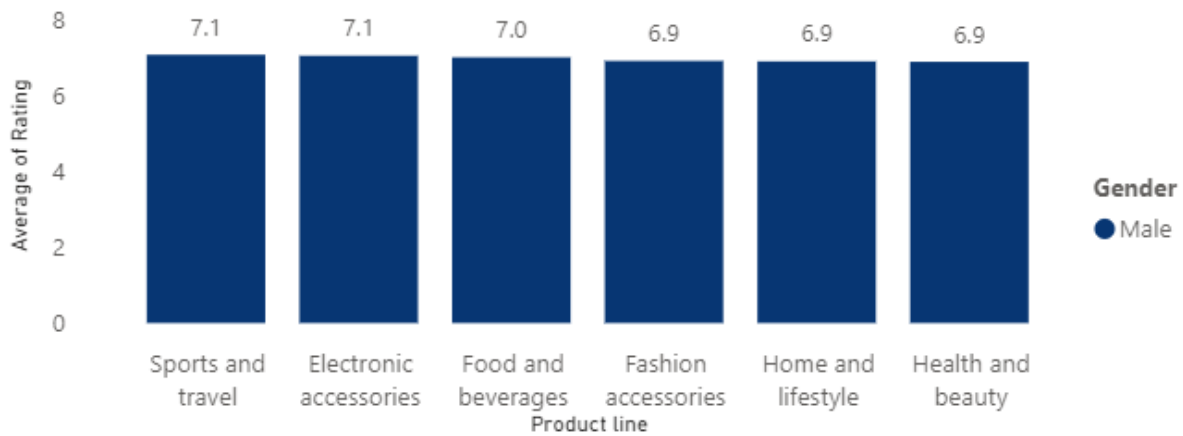


Figure17,18 and 19: Product Line Rating by Gender

Title	Total sales Breakdown by Gender
Chart type	Clustered Column Chart (with Gender Slicer)
Reason	The clustered column chart provides a clear visualization of average customer satisfaction ratings across product lines, segmented by gender. The addition of a gender slicer enables dynamic filtering for focused analysis.
Findings	<ul style="list-style-type: none">• "Food and Beverages" received the highest average rating of 7.1 across all customers.• "Home and Lifestyle" was rated the lowest, with an average rating of 6.8.• Female customers rated "Food and Beverages" the highest, with an average score of 7.2.• Male customers gave their highest rating to "Sports and Travel," averaging 7.1.

Q6 How are total sales distributed across different payment methods, and what trends can be observed regarding customer preferences?

a. **Payment Method Breakdown: Distribution of total sales based on the payment method (Cash, Credit Card, EWallet).**

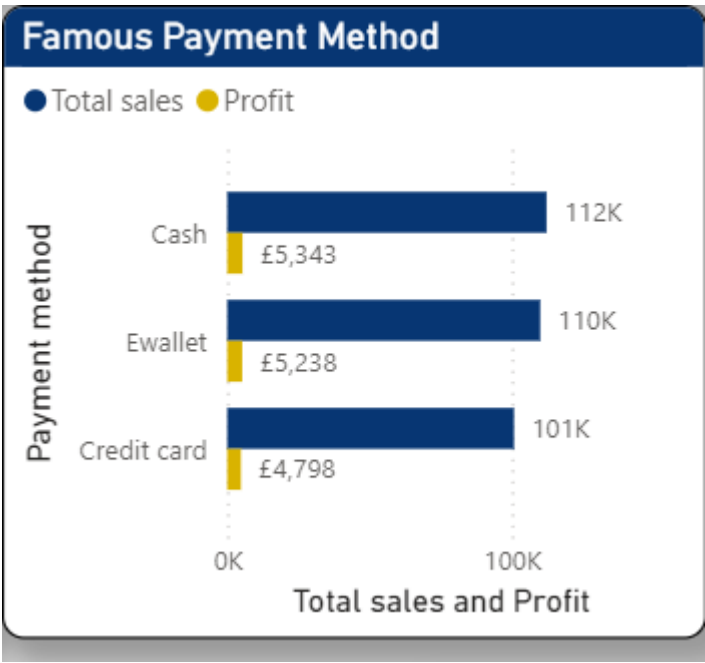


Figure 20: Payment Method Breakdown

Title	Famous payment method
Chart type	Stacked Bar Chart

Reason	The stacked bar chart was used to display the distribution of total sales by payment method, allowing for a clear comparison of sales performance across Cash, Credit Card, and Ewallet options.
Findings	<ul style="list-style-type: none"> • Cash was the most utilized payment method, generating the highest sales with a total of 112k. • This insight highlights a customer preference for cash transactions over other payment methods.

b. Customer Preferences: Which payment methods are most preferred by customers?

From the observed data, Cash is the most preferred payment method, contributing the highest sales overall. To encourage the use of digital payment options like E-wallets, the company could introduce strategies such as offering discounts or rewards. This approach could increase E-Wallet adoption, improve customer convenience, and align with growing trends in digital payments.

Identifying Poor Rating Dates for Further Investigation

This section analyzes customer ratings on specific dates to identify patterns in satisfaction. The 9th of February was found to have the lowest average rating of 5.91. Visual tools, such as cards and infographic designer visuals, help highlight key dates for further investigation.

Understanding these trends can help improve customer service and address issues that may have led to lower satisfaction on certain days.

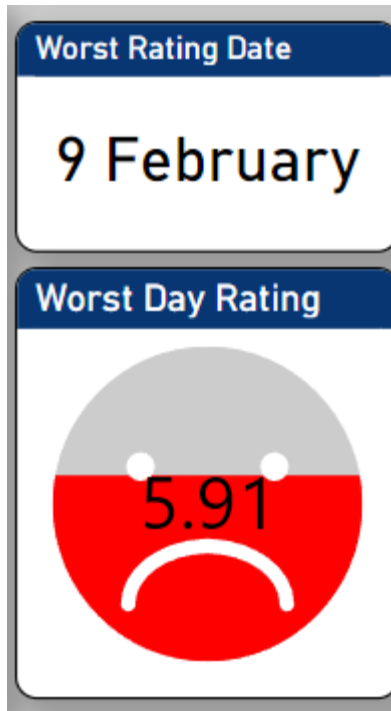


Figure 21: Worst rating investigation.

Title	Key Dates and Ratings
Chart type	Card Visual and Infographic Designer Visual
Reason	The card visual was used to highlight the specific date with the lowest rating, while the infographic designer visuals helped illustrate and pinpoint the overall trend of customer satisfaction for better clarity and focus. A sad face icon was used to represent a low rating, emphasizing the need for attention on this particular date.
Findings	<ul style="list-style-type: none">• The worst-rated date was 9th February, with an average rating of 5.91.• This date stands out as an area of concern, indicating the need for further analysis to address potential causes of customer dissatisfaction.

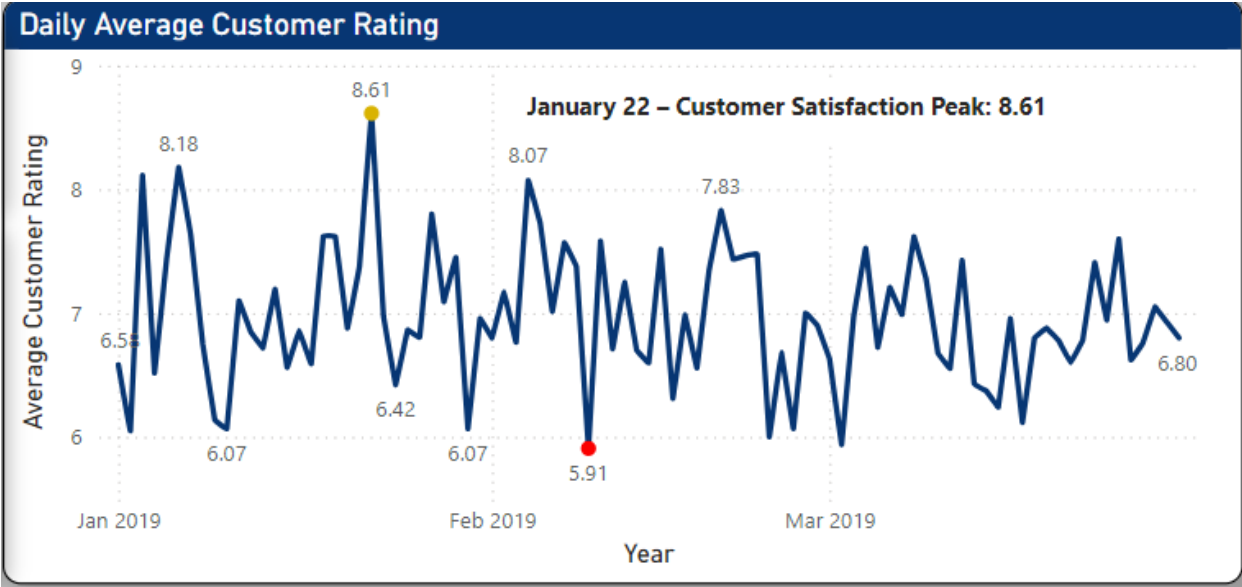


Figure 22: Average Customer rating each day

Title	Daily Average Customer Rating
Chart type	Line Graph
Reason	The line graph was used to display the daily average customer rating, providing a clear visualization of trends over time and identifying key peaks and troughs in customer satisfaction.
Findings	<ul style="list-style-type: none">January 22nd had the highest average rating of 8.61, indicating a peak in customer satisfaction on this specific day.The graph helped identify the date with the most positive feedback, highlighting a period of strong customer approval.

A decomposition tree was used to break down sales data by branch, product line, and customer type, offering a detailed view of sales performance and validating insights from other visuals. Additionally, a Q&A feature was added to enhance interactivity, allowing users to explore specific questions and gain a deeper understanding of the factors driving sales.

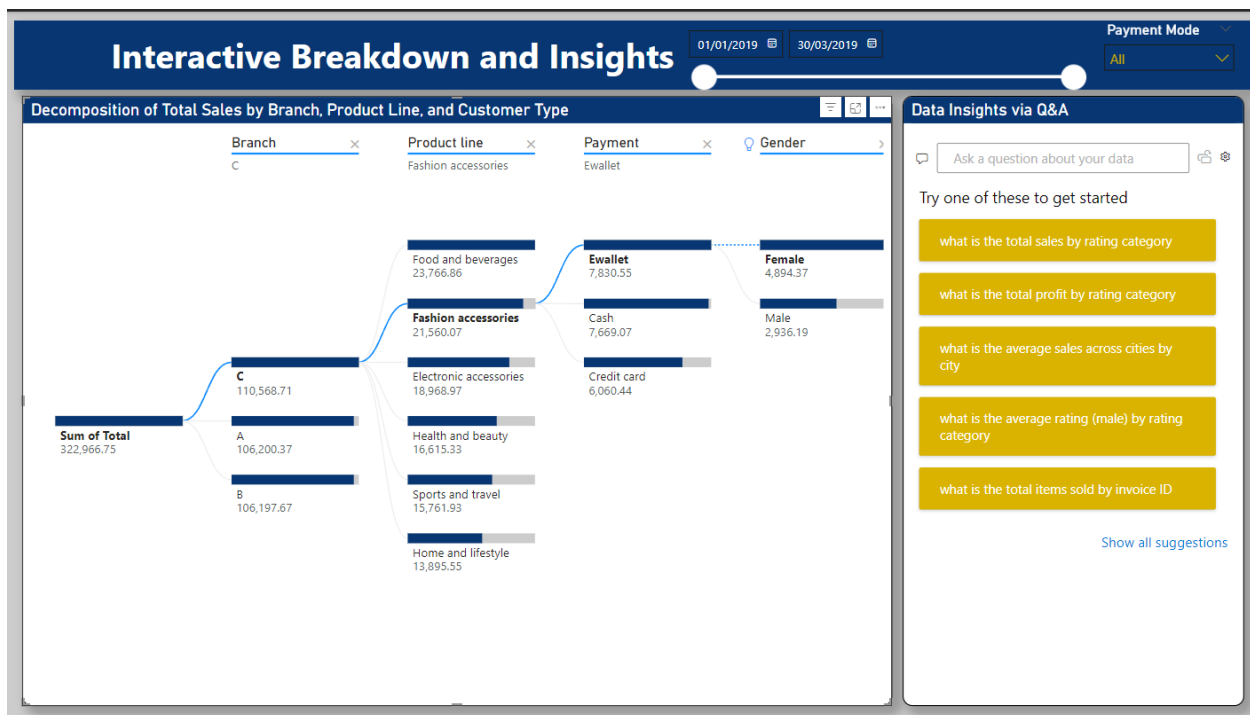


Figure 22: Decomposition tree for validation

3. Conclusion

Based on the comprehensive analysis of the data, several significant findings have been identified. Customer satisfaction shows room for improvement, with an overall average rating of 6.97. Key dates were highlighted, including January 22, which had the highest satisfaction score of 8.61, and February 9, which recorded the lowest at 5.91. Financially, the total sales for the analyzed period amounted to £322,000, with a profit of £15,000, and January stood out as the most profitable month, contributing £121.83k in revenue. Among the branches, Branch C led with sales of £110.56k, significantly contributing to overall performance.

Product line analysis revealed that Food and Beverages is the most profitable category, contributing 17% of total profits, while Health and Beauty recorded the lowest contribution at 15.23%. Gender analysis showed that female customers generated higher sales (£168k) than males (£155k), particularly favoring the Food and Beverages line. Conversely, males demonstrated a preference for Health and Beauty. Notably, average ratings for Food and Beverages were the highest at 7.1, while Home and Lifestyle had the lowest average rating of 6.8. Furthermore, payment preferences showed Cash as the dominant payment method, accounting for £112k of sales.

3.1.Recommendations

- investigate the reasons behind low satisfaction on February 9 to address service gaps and improve customer experiences on critical dates.
- Replicate strategies that drove January's peak sales to capitalize on seasonal trends and boost revenue during similar periods.
- Expand successful practices from Branch C to other branches to improve performance across locations.
- Focus marketing efforts on high-performing product lines, particularly Food and Beverages, while developing strategies to enhance the performance of underperforming categories like Health and Beauty.
- Promote eWallet usage through discounts or incentives to adapt to shifting payment preferences and diversify revenue channels.
- Develop targeted engagement strategies for male customers to address their lower satisfaction ratings and improve their experience with the business.

- Use the product line slicer to monitor individual category performance and adjust strategies accordingly for sustained profitability.

3.2. Personal conclusion

Before this course, I had no experience in data analysis using Power BI, I had some basic familiarity with Python. However, this project provided an opportunity to gain valuable skills in data analysis and visualization, especially with Power BI. While there were challenges during the data cleaning process, these difficulties offered important learning experiences and improved problem-solving skills. Overall, this has been a rewarding experience, leading to greater confidence in working with data and using Power BI effectively.

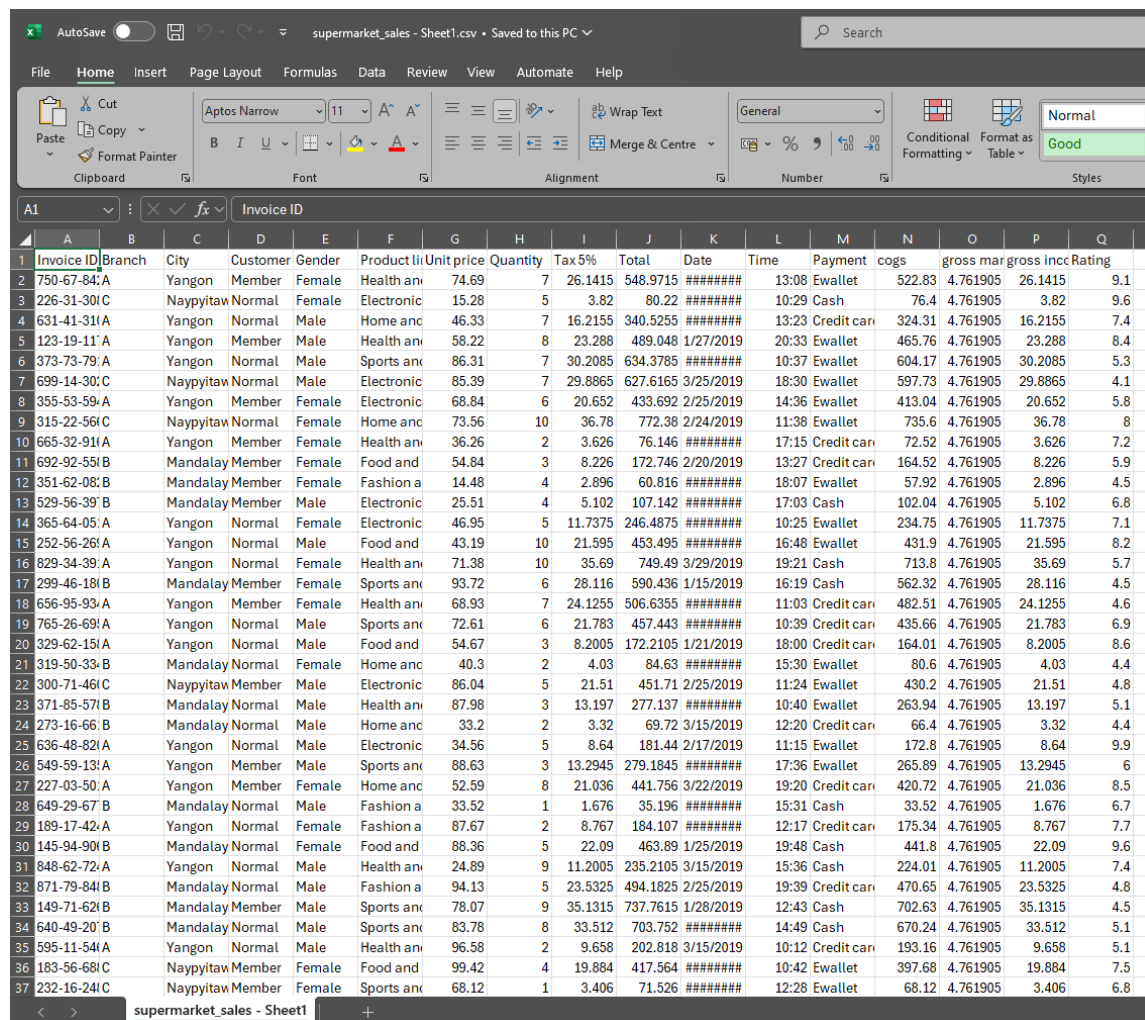
4. Appendix: Bi Design

4.1 Data Pre-Processing and data cleaning

The data pre-processing and cleaning phase was critical in preparing the dataset for analysis and visualization. Below is a detailed explanation of each step taken during this process:

Data Loading

- The dataset was downloaded in CSV format from the provided data source.



Invoice ID	Branch	City	Customer	Gender	Product line	Unit price	Quantity	Tax 5%	Total	Date	Time	Payment	cogs	gross mar	gross inc	Rating
750-67-84	A	Yangon	Member	Female	Health an	74.69	7	26.1415	548.9715	#####	13:08	Ewallet	522.83	4.761905	26.1415	9.1
226-31-30	C	Naypyitaw	Normal	Female	Electronic	15.28	5	3.82	80.22	#####	10:29	Cash	76.4	4.761905	3.82	9.6
631-41-31	A	Yangon	Normal	Male	Home anc	46.33	7	16.2155	340.5255	#####	13:23	Credit car	324.31	4.761905	16.2155	7.4
123-19-11	A	Yangon	Member	Male	Health an	58.22	8	23.288	489.048	1/27/2019	20:33	Ewallet	465.76	4.761905	23.288	8.4
373-73-79	A	Yangon	Normal	Male	Sports an	86.31	7	30.2085	634.3785	#####	10:37	Ewallet	604.17	4.761905	30.2085	5.3
699-14-30	C	Naypyitaw	Normal	Male	Electronic	85.39	7	29.8865	627.6165	3/25/2019	18:30	Ewallet	597.73	4.761905	29.8865	4.1
355-53-59	A	Yangon	Member	Female	Electronic	68.84	6	20.652	433.692	2/25/2019	14:36	Ewallet	413.04	4.761905	20.652	5.8
315-22-56	C	Naypyitaw	Normal	Female	Home anc	73.56	10	36.78	772.38	2/24/2019	11:38	Ewallet	735.6	4.761905	36.78	8
665-32-91	A	Yangon	Member	Female	Health an	36.26	2	3.626	76.146	#####	17:15	Credit car	72.52	4.761905	3.626	7.2
692-92-55	B	Mandalay	Member	Female	Food and	54.84	3	8.226	172.746	2/20/2019	13:27	Credit car	164.52	4.761905	8.226	5.9
351-62-08	B	Mandalay	Member	Female	Fashion a	14.48	4	2.896	60.816	#####	18:07	Ewallet	57.92	4.761905	2.896	4.5
529-56-39	B	Mandalay	Member	Male	Electronic	25.51	4	5.102	107.142	#####	17:03	Cash	102.04	4.761905	5.102	6.8
365-64-05	A	Yangon	Normal	Female	Electronic	46.95	5	11.7375	246.4875	#####	10:25	Ewallet	234.75	4.761905	11.7375	7.1
252-56-26	A	Yangon	Normal	Male	Food and	43.19	10	21.595	453.495	#####	16:48	Ewallet	431.9	4.761905	21.595	8.2
829-34-39	A	Yangon	Normal	Female	Health an	71.38	10	35.69	749.49	3/29/2019	19:21	Cash	713.8	4.761905	35.69	5.7
299-46-18	B	Mandalay	Member	Female	Sports an	93.72	6	28.116	590.436	1/15/2019	16:19	Cash	562.32	4.761905	28.116	4.5
656-95-93	A	Yangon	Member	Female	Health an	68.93	7	24.1255	506.6355	#####	11:03	Credit car	482.51	4.761905	24.1255	4.6
765-26-69	A	Yangon	Normal	Male	Sports an	72.61	6	21.783	457.443	#####	10:39	Credit car	435.66	4.761905	21.783	6.9
329-62-15	A	Yangon	Normal	Male	Food and	54.67	3	8.2005	172.2105	1/21/2019	18:00	Credit car	164.01	4.761905	8.2005	8.6
319-50-33	B	Mandalay	Normal	Female	Home anc	40.3	2	4.03	84.63	#####	15:30	Ewallet	80.6	4.761905	4.03	4.4
300-71-46	C	Naypyitaw	Member	Male	Electronic	86.04	5	21.51	451.71	2/25/2019	11:24	Ewallet	430.2	4.761905	21.51	4.8
371-85-57	B	Mandalay	Normal	Male	Health an	87.98	3	13.197	277.137	#####	10:40	Ewallet	263.94	4.761905	13.197	5.1
273-16-66	B	Mandalay	Normal	Male	Home anc	33.2	2	3.32	69.72	3/15/2019	12:20	Credit car	66.4	4.761905	3.32	4.4
636-48-82	A	Yangon	Normal	Male	Electronic	34.56	5	8.64	181.44	2/17/2019	11:15	Ewallet	172.8	4.761905	8.64	9.9
549-59-13	A	Yangon	Member	Male	Sports an	88.63	3	13.2945	279.1845	#####	17:36	Ewallet	265.89	4.761905	13.2945	6
227-03-50	A	Yangon	Member	Female	Home anc	52.59	8	21.036	441.756	3/22/2019	19:20	Credit car	420.72	4.761905	21.036	8.5
649-29-67	B	Mandalay	Normal	Male	Fashion a	33.52	1	1.676	35.196	#####	15:31	Cash	33.52	4.761905	1.676	6.7
189-17-42	A	Yangon	Normal	Female	Fashion a	87.67	2	8.767	184.107	#####	12:17	Credit car	175.34	4.761905	8.767	7.7
145-94-90	B	Mandalay	Normal	Female	Food and	88.36	5	22.09	463.89	1/25/2019	19:48	Cash	441.8	4.761905	22.09	9.6
848-62-72	A	Yangon	Normal	Male	Health an	24.89	9	11.2005	235.2105	3/15/2019	15:36	Cash	224.01	4.761905	11.2005	7.4
871-79-84	B	Mandalay	Normal	Male	Fashion a	94.13	5	23.5325	494.1825	2/25/2019	19:39	Credit car	470.65	4.761905	23.5325	4.8
149-71-62	B	Mandalay	Member	Male	Sports an	78.07	9	35.1315	737.7615	1/28/2019	12:43	Cash	702.63	4.761905	35.1315	4.5
640-49-20	B	Mandalay	Normal	Male	Sports an	83.78	8	33.512	703.752	#####	14:49	Cash	670.24	4.761905	33.512	5.1
595-11-54	A	Yangon	Normal	Male	Health an	96.58	2	9.658	202.818	3/15/2019	10:12	Credit car	193.16	4.761905	9.658	5.1
183-56-68	C	Naypyitaw	Member	Female	Food and	99.42	4	19.884	417.564	#####	10:42	Ewallet	397.68	4.761905	19.884	7.5
232-16-24	C	Naypyitaw	Member	Female	Sports an	68.12	1	3.406	71.526	#####	12:28	Ewallet	68.12	4.761905	3.406	6.8

Figure 23: Data set in Micro soft excel.

- It was imported into the Power Query Editor in Power BI using the **Get Data** option and selecting the **Text/CSV** option from the dropdown menu.

fx = Table.TransformColumnTypes(#"Promoted Headers",{{"Invoice ID", type text}, {"Branch", type text}, {"City", type text}, {"Customer type", type text}, {"Gender", type text}, {"Product line", type text}, {"Unit price", type number}, {"Quantity", type number}})

	Invoice ID	Branch	City	Customer type	Gender	Product line	Unit price	Quantity
1	750-67-8428	A	Yangon	Member	Female	Health and beauty	74.69	
2	226-31-3081	C	Naypyitaw	Normal	Female	Electronic accessories	15.28	
3	631-41-3108	A	Yangon	Normal	Male	Home and lifestyle	46.33	
4	123-19-1176	A	Yangon	Member	Male	Health and beauty	58.22	
5	373-73-7910	A	Yangon	Normal	Male	Sports and travel	86.31	
6	699-14-3026	C	Naypyitaw	Normal	Male	Electronic accessories	85.39	
7	355-53-5943	A	Yangon	Member	Female	Electronic accessories	68.84	
8	315-22-5665	C	Naypyitaw	Normal	Female	Home and lifestyle	73.56	
9	665-32-9167	A	Yangon	Member	Female	Health and beauty	36.26	
10	692-92-5582	B	Mandalay	Member	Female	Food and beverages	54.84	
11	351-62-0822	B	Mandalay	Member	Female	Fashion accessories	14.48	
12	529-56-3974	B	Mandalay	Member	Male	Electronic accessories	25.51	
13	365-64-0515	A	Yangon	Normal	Female	Electronic accessories	46.95	
14	252-56-2699	A	Yangon	Normal	Male	Food and beverages	43.19	
15	829-34-3910	A	Yangon	Normal	Female	Health and beauty	71.38	
16	299-46-1805	B	Mandalay	Member	Female	Sports and travel	93.72	
17	656-95-9349	A	Yangon	Member	Female	Health and beauty	68.93	
18	765-26-6951	A	Yangon	Normal	Male	Sports and travel	72.61	
19	329-62-1586	A	Yangon	Normal	Male	Food and beverages	54.67	
20	319-50-3348	B	Mandalay	Normal	Female	Home and lifestyle	40.3	
21	300-71-4605	C	Naypyitaw	Member	Male	Electronic accessories	86.04	
22	371-85-5789	B	Mandalay	Normal	Male	Health and beauty	87.98	
23	273-16-6619	B	Mandalay	Normal	Male	Home and lifestyle	33.2	
24	636-48-8204	A	Yangon	Normal	Male	Electronic accessories	34.56	
25	549-59-1358	A	Yangon	Member	Male	Sports and travel	88.63	
26	227-03-5010	A	Yangon	Member	Female	Home and lifestyle	52.59	
27	649-29-6775	B	Mandalay	Normal	Male	Fashion accessories	33.52	
28	189-17-4241	A	Yangon	Normal	Female	Fashion accessories	87.67	
29	145-94-9061	B	Mandalay	Normal	Female	Food and beverages	88.36	

Power Query Editor showing 29 rows of data. The table has 9 columns: Invoice ID, Branch, City, Customer type, Gender, Product line, Unit price, and Quantity. The data is filtered based on the top 1000 rows.

Figure 24: Data set in Power Bi.

Initial Exploration

- The dataset contained **17 columns** and **1,000 rows**.
- On initial inspection in Microsoft Excel, it was observed that there were no missing values. However, errors were identified in the **Date** column.
- The data was checked for duplicates, but none were found. This was done by selecting all columns, right-clicking, and choosing **Remove Duplicates**.

Data Type Adjustments

- Each column's data type was changed to match its contents (e.g., **Text**, **Whole Number**, **Date**, etc.) by right clicking the column header and selecting **Change Type**.
- An error occurred when attempting to set the **Date** column to the **Date** data type, resulting in a **58% error rate**.

Date
05/01/2019
03/08/2019
03/03/2019
Error
02/08/2019
Error
Error
Error
01/10/2019
Error
02/06/2019
03/09/2019
02/12/2019
02/07/2019
Error
Error
03/11/2019
01/01/2019
Error
03/11/2019
Error
03/05/2019
Error
Error
03/02/2019
Error
02/08/2019
03/10/2019

Figure 25: Data set in Power Bi.

Resolving Date Column Errors

- Due to challenges in resolving the issue within Power BI, the dataset was exported to **Jupyter Notebook** for correction using Python:

```
date_before = data['Date'].head(10)
data['Date'] = pd.to_datetime(data['Date'],
format='%m/%d/%Y')
date_after = data['Date'].head(10)
date_before, date_after
```

The code ensures the **Date** column is in a standard and usable datetime format, essential for further data processing and analysis.

- The corrected dataset was saved , named "supermarket_sales" and re-imported into Power BI.

A ^B _C Date	Date_1
05/01/2019	05/01/2019
03/08/2019	08/03/2019
03/03/2019	03/03/2019
1/27/2019	27/01/2019
02/08/2019	08/02/2019
3/25/2019	25/03/2019
2/25/2019	25/02/2019
2/24/2019	24/02/2019
01/10/2019	10/01/2019
2/20/2019	20/02/2019
02/06/2019	06/02/2019
03/09/2019	09/03/2019
02/12/2019	12/02/2019
02/07/2019	07/02/2019
3/29/2019	29/03/2019
1/15/2019	15/01/2019
03/11/2019	11/03/2019
01/01/2019	01/01/2019
1/21/2019	21/01/2019
03/11/2019	11/03/2019
2/25/2019	25/02/2019
03/05/2019	05/03/2019
3/15/2019	15/03/2019
2/17/2019	17/02/2019
03/02/2019	02/03/2019
3/22/2019	22/03/2019
02/08/2019	08/02/2019
03/10/2019	10/03/2019
1/25/2019	25/01/2019

Figure 26: Comparison between the corrected date column and the previous date column.

- The new **Date** column was successfully updated to the **Date** data type. The original **Date** column was deleted by right clicking the column and selecting **Remove**.
- It was later noted that DAX or M Language could have been used to resolve the issue directly in Power BI, showing potential for future improvements.

Renaming Columns

- Columns were renamed for better clarity by right-clicking on the column header and selecting **Rename**:
 - **COG** → **Cost of Goods**
 - Corrected **Date_1** → **Date**

Indexing

- Primary keys were created to facilitate data modeling by adding **Index Columns**:
 - **Product_Line**, **Customer_ID**, and **Date_ID**.
 - This was done by selecting **Add Column** from the menu and choosing the **Index Column** option.

1 ² ₃ Customer ID	1 ² ₃ Product ID	1 ² ₃ Date ID
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29

Figure 27: Primary key added to the table

Creating Dimensional Tables

- The dataset was duplicated and split into four dimensional tables to adhere to a star schema design:

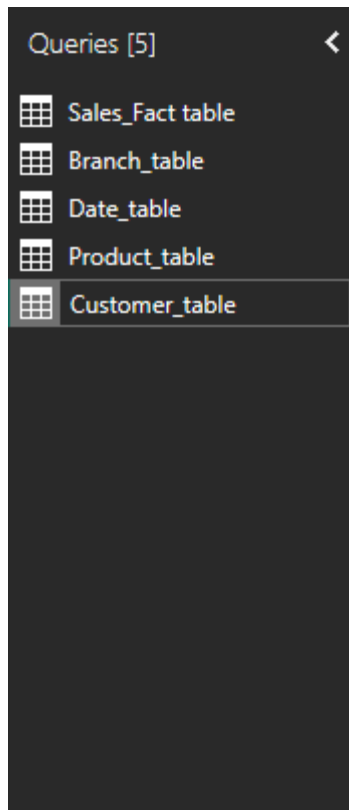


Figure 28: Dimensional tables

a. Branch Table

- Only the **Branch** and **City** columns were retained by right-clicking on other columns and selecting **Remove Columns**.
- Duplicates were removed to show unique branches (A, B, and C) and their corresponding cities using the **Remove Duplicates** option.

<div> <div>✕</div> <div>✓</div> <div>f_x</div> </div> <div>= Table.Sort("#Removed Columns2",{{"Bran</div>	
<div> <div> <div></div> <div></div> </div> <div> <div>A^BC</div> <div>Branch</div> </div> <div> <div> <div>↕</div> <div>↑</div> </div> </div> </div>	<div> <div> <div>A^BC</div> <div>City</div> </div> <div>▼</div> </div>
1	A
2	B
3	C

Figure 29: Final view of the branch table

b. Date Table

- Retained only **Date**, **Time**, and **Date_ID** columns by removing others.
- New columns were created to extract **Day**, **Month**, **Year**, and **Month Name** from the Date column by using the **Add Column** menu and selecting the **Date** options.

= Table.TransformColumns(#"Inserted Month Name", {{"Month Name", each Text.Start(_, 3), type text}})							
1 ² Date ID	Time	Date	1 ² Day	1 ² Month	1 ² Year	1 ² Month Name	
1	13:08:00	05/01/2019	5	1	2019	Jan	
2	10:29:00	08/03/2019	8	3	2019	Mar	
3	13:23:00	03/03/2019	3	3	2019	Mar	
4	20:33:00	27/01/2019	27	1	2019	Jan	
5	10:37:00	08/02/2019	8	2	2019	Feb	
6	18:30:00	25/03/2019	25	3	2019	Mar	
7	14:36:00	25/02/2019	25	2	2019	Feb	
8	11:38:00	24/02/2019	24	2	2019	Feb	
9	17:15:00	10/01/2019	10	1	2019	Jan	
10	13:27:00	20/02/2019	20	2	2019	Feb	
11	18:07:00	06/02/2019	6	2	2019	Feb	
12	17:03:00	09/03/2019	9	3	2019	Mar	
13	10:25:00	12/02/2019	12	2	2019	Feb	
14	16:48:00	07/02/2019	7	2	2019	Feb	
15	19:21:00	29/03/2019	29	3	2019	Mar	
16	16:19:00	15/01/2019	15	1	2019	Jan	
17	11:03:00	11/03/2019	11	3	2019	Mar	
18	10:39:00	01/01/2019	1	1	2019	Jan	
19	18:00:00	21/01/2019	21	1	2019	Jan	
20	15:30:00	11/03/2019	11	3	2019	Mar	
21	11:24:00	25/02/2019	25	2	2019	Feb	
22	10:40:00	05/03/2019	5	3	2019	Mar	
23	12:20:00	15/03/2019	15	3	2019	Mar	
24	11:15:00	17/02/2019	17	2	2019	Feb	
25	17:36:00	02/03/2019	2	3	2019	Mar	
26	19:20:00	22/03/2019	22	3	2019	Mar	
27	15:31:00	08/02/2019	8	2	2019	Feb	
28	12:17:00	10/03/2019	10	3	2019	Mar	
29	19:48:00	25/01/2019	25	1	2019	Jan	
30	15:36:00	15/03/2019	15	3	2019	Mar	

Figure 30: Final view of the Date table

c. Product Table

- Retained columns related to product line analysis, such as **Product_Line** and relevant sales metrics, by removing unrelated columns.

	1 ² ₃ Product ID	A ^B _C Product line	1.2 Unit price	1 ² ₃ Quantity
1	1	Health and beauty	74.69	7
2	2	Electronic accessories	15.28	5
3	3	Home and lifestyle	46.33	7
4	4	Health and beauty	58.22	8
5	5	Sports and travel	86.31	7
6	6	Electronic accessories	85.39	7
7	7	Electronic accessories	68.84	6
8	8	Home and lifestyle	73.56	10
9	9	Health and beauty	36.26	2
10	10	Food and beverages	54.84	3
11	11	Fashion accessories	14.48	4
12	12	Electronic accessories	25.51	4
13	13	Electronic accessories	46.95	5
14	14	Food and beverages	43.19	10
15	15	Health and beauty	71.38	10
16	16	Sports and travel	93.72	6
17	17	Health and beauty	68.93	7
18	18	Sports and travel	72.61	6
19	19	Food and beverages	54.67	3
20	20	Home and lifestyle	40.3	2
21	21	Electronic accessories	86.04	5
22	22	Health and beauty	87.98	3
23	23	Home and lifestyle	33.2	2
24	24	Electronic accessories	34.56	5
25	25	Sports and travel	88.63	3
26	26	Home and lifestyle	52.59	8
27	27	Fashion accessories	33.52	1
28	28	Fashion accessories	87.67	2
29	29	Food and beverages	88.36	5
30	30	Health and beauty	74.89	9

profiling based on top 1000 rows

Figure 31: Final view of the Product line table

d. Customer Table

- Retained columns related to customer analysis, such as **Customer_ID**, **Gender**, **Rating**, and other demographic details.
- A new column, **Rating Category**, was added using M Language. This will be explained in a subsequent section.

= Table.AddColumn("#Reordered Columns1", "Rating Category", each						
1.2 Customer ID	1.0 Payment	1.0 Gender	1.0 Customer type	1.2 Rating	1.0 Rating Category	
1	1 Ewallet	Female	Member		9.1 Excellent	
2	2 Cash	Female	Normal		9.6 Excellent	
3	3 Credit card	Male	Normal		7.4 Good	
4	4 Ewallet	Male	Member		8.4 Good	
5	5 Ewallet	Male	Normal		5.3 Average	
6	6 Ewallet	Male	Normal		4.1 Poor	
7	7 Ewallet	Female	Member		5.8 Average	
8	8 Ewallet	Female	Normal		8 Good	
9	9 Credit card	Female	Member		7.2 Good	
10	10 Credit card	Female	Member		5.9 Average	
11	11 Ewallet	Female	Member		4.5 Poor	
12	12 Cash	Male	Member		6.8 Average	
13	13 Ewallet	Female	Normal		7.1 Good	
14	14 Ewallet	Male	Normal		8.2 Good	
15	15 Cash	Female	Normal		5.7 Average	
16	16 Cash	Female	Member		4.5 Poor	
17	17 Credit card	Female	Member		4.6 Poor	
18	18 Credit card	Male	Normal		6.9 Average	
19	19 Credit card	Male	Normal		8.6 Good	
20	20 Ewallet	Female	Normal		4.4 Poor	
21	21 Ewallet	Male	Member		4.8 Poor	
22	22 Ewallet	Male	Normal		5.1 Average	
23	23 Credit card	Male	Normal		4.4 Poor	
24	24 Ewallet	Male	Normal		9.9 Excellent	
25	25 Ewallet	Male	Member		6 Average	
26	26 Credit card	Female	Member		8.5 Good	
27	27 Cash	Male	Normal		6.7 Average	
28	28 Credit card	Female	Normal		7.7 Good	
29	29 Cash	Female	Normal		9.6 Excellent	
30	30 Cash	Male	Normal		7.4 Good	

Figure 32: Final view of the Customer table

These steps ensured the dataset was cleaned, organized, and ready for efficient modeling and visualization. Each process was performed systematically to maintain data integrity and support insightful analysis.

4.2 BI Data Modelling via Star Schema - Facts and Dimensions.

The dataset was organized using a Star Schema to make it more efficient and easier to analyze. The Star Schema includes a central fact table connected to several dimension tables, allowing for clear and straightforward data relationships.

Process and Relationships

The following steps and relationships were established to create the model:

1. Unnecessary columns were removed, reducing the **Fact Table** to 10 essential columns.

= Table.RemoveColumns(#Renamed Columns6,{"Quantity", "Unit price"})										
Invoice ID	Branch	Customer ID	Tax	Date ID	Total	Cost of goods sold	1.2 gross margin percentage	1.2 gross income	Product ID	
1 0-67-8428	A	1	26.14	1	548.97	522.83	4.761904762	26.1415	26.1415	
2 6-31-3081	C	2	3.82	2	80.22	76.40	4.761904762	3.82	3.82	
3 1-41-3108	A	3	16.22	3	340.53	324.31	4.761904762	16.2155	16.2155	
4 3-19-1176	A	4	23.29	4	489.05	465.76	4.761904762	23.288	23.288	
5 3-75-7910	A	5	30.21	5	634.38	604.17	4.761904762	30.2085	30.2085	
6 9-14-3026	C	6	29.89	6	627.62	597.73	4.761904762	29.8865	29.8865	
7 5-53-5943	A	7	20.65	7	433.69	413.04	4.761904762	20.652	20.652	
8 5-22-5665	C	8	36.78	8	772.38	735.60	4.761904762	36.78	36.78	
9 5-32-9167	A	9	3.63	9	76.15	72.52	4.761904762	3.626	3.626	
10 2-92-5582	B	10	8.23	10	172.75	164.52	4.761904762	8.226	8.226	
11 1-62-0822	B	11	2.90	11	60.82	57.92	4.761904762	2.896	2.896	
12 9-56-3974	B	12	5.10	12	107.14	102.04	4.761904762	5.102	5.102	
13 5-64-0515	A	13	11.74	13	246.49	234.75	4.761904762	11.7375	11.7375	
14 2-56-2699	A	14	21.60	14	453.50	431.90	4.761904762	21.595	21.595	
15 9-34-3910	A	15	35.69	15	749.49	713.80	4.761904762	35.69	35.69	
16 9-46-1805	B	16	28.12	16	590.44	562.32	4.761904762	28.116	28.116	
17 6-95-9349	A	17	24.13	17	506.64	482.51	4.761904762	24.1255	24.1255	
18 5-26-6951	A	18	21.78	18	457.44	435.66	4.761904762	21.783	21.783	
19 9-62-1586	A	19	8.20	19	172.21	164.01	4.761904762	8.2005	8.2005	
20 9-50-3348	B	20	4.03	20	84.63	80.60	4.761904762	4.03	4.03	
21 0-71-4605	C	21	21.51	21	451.71	430.20	4.761904762	21.51	21.51	
22 1-85-5789	B	22	13.20	22	277.14	263.94	4.761904762	13.197	13.197	
23 3-16-6619	B	23	3.32	23	69.72	66.40	4.761904762	3.32	3.32	
24 6-48-8204	A	24	8.64	24	181.44	172.80	4.761904762	8.64	8.64	
25 9-59-1358	A	25	13.29	25	279.18	265.89	4.761904762	13.2945	13.2945	
26 7-03-5010	A	26	21.04	26	441.76	420.72	4.761904762	21.036	21.036	
27 9-29-6775	B	27	1.68	27	35.20	33.52	4.761904762	1.676	1.676	
28 9-17-4041	A	28	8.77	28	184.11	175.34	4.761904762	8.767	8.767	
29 5-94-9061	B	29	22.09	29	463.89	441.80	4.761904762	22.09	22.09	
30 8-63-7242	A	30	11.20	30	235.21	224.01	4.761904762	11.2005	11.2005	
31 1-79-6483	B	31	23.53	31	494.18	470.65	4.761904762	23.5325	23.5325	
32 9-71-6366	B	32	35.13	32	737.76	702.63	4.761904762	35.1315	35.1315	
33 0-49-2076	B	33	33.51	33	703.75	670.24	4.761904762	33.512	33.512	
34 5-11-5460	A	34	9.66	34	202.82	193.16	4.761904762	9.658	9.658	
35 3-56-6882	C	35	19.88	35	417.56	397.68	4.761904762	19.884	19.884	
36 2-16-2483	C	36	3.41	36	71.53	68.12	4.761904762	3.406	3.406	

Figure 33: Final view of the Fact table

- The **Branch** column in the fact table connects to the **Branch Table** using a **1-to-Many relationship**.
- The **Date ID** column links the fact table to the **Date Table** in a **1-to-1 relationship**.
- The **Product ID** serves as the key between the fact table and the **Product Line Table**, creating a **1-to-1 relationship**.
- The **Customer ID** acts as the key connecting the fact table to the **Customer Table**, also in a **1-to-1 relationship**.

Star Schema Structure

The result is a Star Schema, where the fact table sits at the center and connects to the dimension tables. This layout makes it easier to analyze the data and understand relationships between sales, branches, dates, products, and customers.

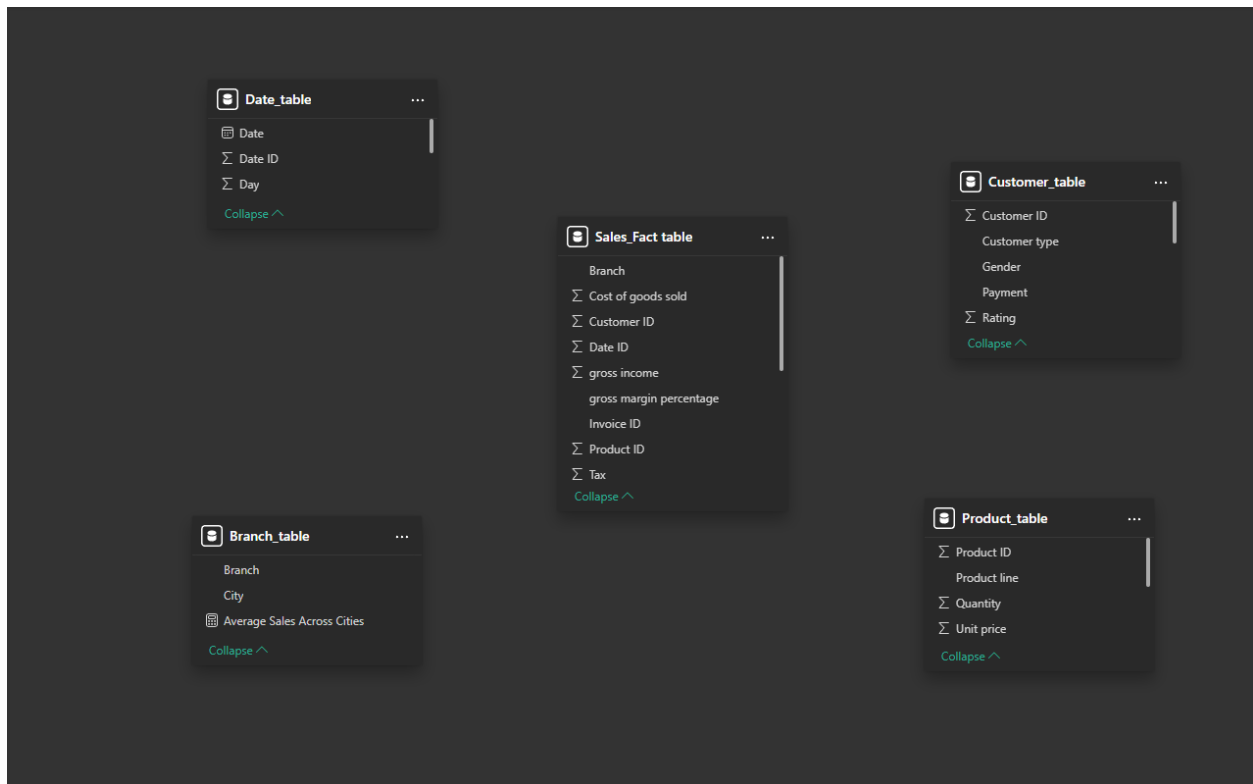


Figure 33: View of the Fact Table and Dimensional Table Before Connection

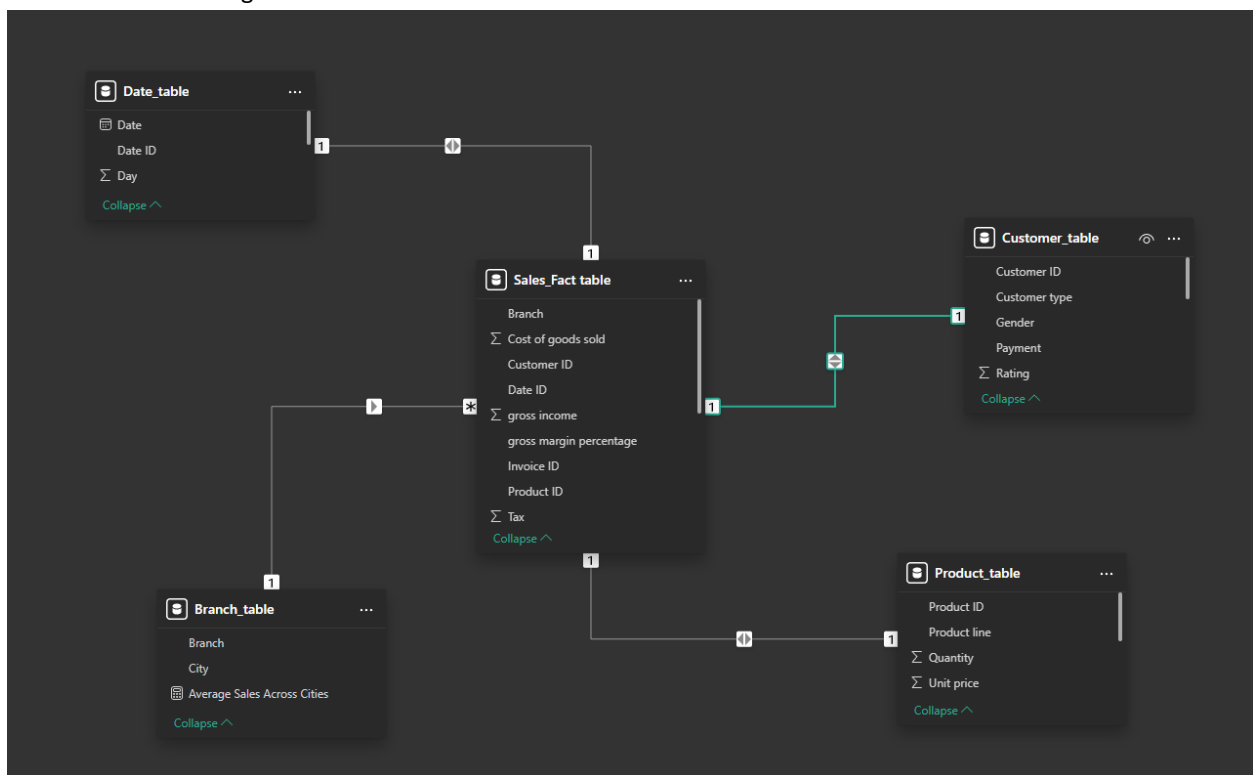


Figure 33: View of the Fact Table and Dimensional Table After Connection

4.3 DAX and M Language

DAX Measures Created

1. Average Customer Rating

```
1 Average Customer Rating = AVERAGE('Customer_table'[Rating])
2
```

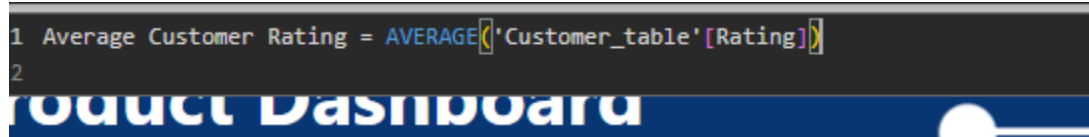
A screenshot of a software interface showing a DAX formula bar. The formula is 'Average Customer Rating = AVERAGE('Customer_table'[Rating])'. Below the formula bar, there is a blue header with the text 'Product Dashboard'.

Figure 34: Average Customer Rating Dax formula

- **Function Used:** AVERAGE
- **Justification:** This measure calculates the average customer satisfaction rating from the dataset, providing insight into overall customer sentiment.

2. Total Sales

```
structure      Formatting      Pr
1 Total sales = SUM('Sales_Fact table'[Total])
2
```

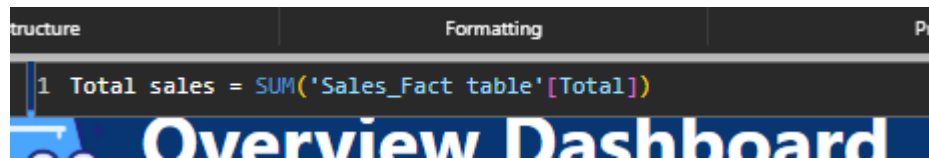
A screenshot of a software interface showing a DAX formula bar. The formula is 'Total sales = SUM('Sales_Fact table'[Total])'. Above the formula bar, there are three tabs: 'structure', 'Formatting', and 'Pr'. Below the formula bar, there is a blue header with the text 'Overview Dashboard'.

Figure 34: Total sales Dax formula,

- **Function Used:** SUM
- **Justification:** Summing up the Total column provides a clear understanding of the overall revenue generated across all transactions.

3. Profit Percentage

```
1 Profit Percentage = 'Sales_Fact table'[Total Profit]/ SUM('Sales_Fact table'[Total])
2
3
```

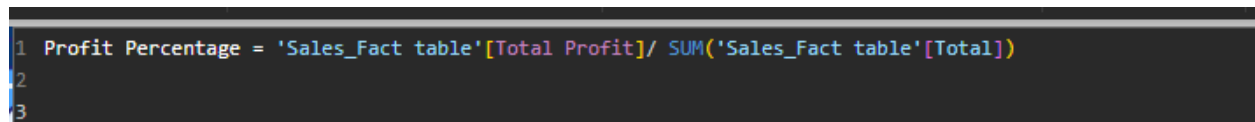
A screenshot of a software interface showing a DAX formula bar. The formula is 'Profit Percentage = 'Sales_Fact table'[Total Profit]/ SUM('Sales_Fact table'[Total])'. The formula bar is numbered 1, 2, and 3.

Figure 34: Profit Percentage Dax formula

- **Functions Used:** Division operator / and SUM
- **Justification:** This measure computes the profit as a percentage of total sales, helping gauge profitability.

4. Total Profit

```
1 Total Profit = SUM('Sales_Fact table'[gross income])
```

Figure 35: Total Profit Dax formula

- **Function Used:** SUM
- **Justification:** This measure adds up all gross income values to provide the overall profit from sales.

5. Average Sales Across Cities

```
1 Average Sales Across Cities =  
2 AVERAGEX(  
3     SUMMARIZE(  
4         'Branch_table',  
5         'Branch_table'[City],  
6         "Total Sales", SUM('Sales_Fact table'[Total])  
7     ),  
8     [Total Sales]  
9 )  
10
```

Figure 36: Average Sales Across Cities Dax formula

- **Functions Used:** AVERAGEX, SUMMARIZE, and SUM
- **Justification:** This measure calculates the average total sales across different cities, offering insights into regional performance.

6. Product Line Contribution to Profit

```

1 Product Line Contribution to Profit =
2     DIVIDE(
3         MAXX(
4             SUMMARIZE(
5                 'Product_table',
6                 'Product_table'[Product line],
7                 "Total Profit", SUM('Sales_Fact table'[gross income])
8             ),
9             [Total Profit]
10        ),
11        SUM('Sales_Fact table'[gross income])
12    )
13
14

```

Figure 37: Product Line Contribution to Profit Dax formula

- **Functions Used:** DIVIDE, MAXX, SUMMARIZE, and SUM
- **Justification:** This measure identifies the percentage contribution of each product line to the total profit, highlighting top-performing categories.

7. Top Product Line by Profit Name

```

1 Top Product Line by Profit Name =
2     SELECTCOLUMNS(
3         TOPN(
4             1,
5             SUMMARIZE(
6                 'Product_table',
7                 'Product_table'[Product line],
8                 "Total Profit", SUM('Sales_Fact table'[gross income])
9             ),
10            [Total Profit],
11            DESC
12        ),
13        "Top Product Line", 'Product_table'[Product line]
14    )
15

```

Figure 38: Top Product Line by Profit Name Dax formula

- **Functions Used:** SELECTCOLUMNS, TOPN, SUMMARIZE, and SUM
- **Justification:** Identifies the product line with the highest profit, simplifying decision-making for product strategies.

8. Total Transactions

```
1 Total Transactions = COUNT('Sales_Fact table'[Invoice ID])
```

Figure 39: Total Transactions Dax formula

- **Function Used:** COUNT
- **Justification:** Counts the total number of sales transactions, offering a metric for sales volume.

9. Total Items Sold

```
1 Total Items Sold = SUM('Product_table'[Quantity])
2
```

Figure 40: Total Items Sold Dax formula

- **Function Used:** SUM
- **Justification:** Calculates the total quantity of products sold, aiding inventory management.

10. Average Rating (Male)

```
1 Average Rating (Male) =
2 AVERAGEX(
3     SUMMARIZE(
4         FILTER(
5             'Customer_table',
6             'Customer_table'[Gender] = "Male"
7         ),
8         'Customer_table'[Gender],
9         "Average Rating", AVERAGE('Customer_table'[Rating])
10    ),
11    [Average Rating]
12 )
13
```

Figure 41: Average Rating (Male) Dax formula

- **Functions Used:** AVERAGEX, SUMMARIZE, FILTER, and AVERAGE
- **Justification:** Calculates the average rating given by male customers, identifying gender-based satisfaction trends.

11. Average Rating (Female)


```

1 Average Rating (Female) =
2 AVERAGEX(
3     SUMMARIZE(
4         FILTER(
5             'Customer_table',
6             'Customer_table'[Gender] = "Female" -- Filter for female gender
7         ),
8         'Customer_table'[Gender],
9         "Average Rating", AVERAGE('Customer_table'[Rating])
10    ),
11    [Average Rating]
12 )

```

Figure 41: Average Rating (Female) Dax formula

- **Functions Used:** AVERAGEX, SUMMARIZE, FILTER, and AVERAGE
- **Justification:** Similar to the male rating measure, but focused on female customers, enabling comparison between genders.

12. Date of Lowest Average Rating

```

DateOfLowestAverageRating =
VAR LowestAvgRating =
    MINX(
        SUMMARIZE(
            Date_table,
            Date_table[Date],
            "DailyAvgRating", AVERAGE(Customer_table[Rating])
        ),
        [DailyAvgRating]
    )
RETURN
    CALCULATE(
        MIN(Date_table[Date]),
        FILTER(
            SUMMARIZE(
                Date_table,
                Date_table[Date],
                "DailyAvgRating", AVERAGE(Customer_table[Rating])
            ),
            [DailyAvgRating] = LowestAvgRating
        )
    )

```

Figure 41: Date of Lowest Average Rating Dax formula

- **Functions Used:** VAR, MINX, SUMMARIZE, AVERAGE, and CALCULATE
- **Justification:** Identifies the specific date with the lowest customer rating, pinpointing areas for improvement.

13. Lowest Average Rating by Day

```
1 LowestAverageRatingByDay =
2 MINX(
3     SUMMARIZE(
4         Date_table,
5         Date_table[Date],
6         "DailyAvgRating", AVERAGE(Customer_table[Rating])
7     ),
8     [DailyAvgRating]
9 )
```

Figure 42: Lowest Average Rating by Day Dax formula

- **Functions Used:** MINX, SUMMARIZE, and AVERAGE
- **Justification:** Determines the lowest daily average customer rating, complementing the previous measure.

Each DAX measure was designed to extract critical insights from the dataset, enhancing the ability to analyze and present data meaningfully.

M language Used

```
= Table.AddColumn("#Reordered Columns1", "Rating Category", each
    if [Rating] >= 9 then "Excellent"
    else if [Rating] >= 7 then "Good"
    else if [Rating] >= 5 then "Average"
    else "Poor",
    ., .)
```

Figure 43: Rating category by Day Dax formula

Functions Used:

- Table.AddColumn: Adds a new column to the table.
- each: Defines the logic to categorize each rating.
- Conditional Statements (if, else if, else): Used to define the different rating categories based on the value of the rating.

Justification: This M code categorizes customer ratings into four categories: "Excellent," "Good," "Average," and "Poor." This transformation helps group ratings, making it easier to analyze and identify trends in customer satisfaction.

4.4 Dashboard

Overview of the Power BI Dashboard

The Overview Dashboard provides a clear summary of key business metrics and insights. It is designed for easy understanding, with key figures like average customer rating, total sales, total profit, and gross margin displayed prominently at the top. These give a quick snapshot of overall performance.

In the center, a stacked column chart shows sales and profit trends across January, February, and March. Alongside, a map visualization displays sales distribution across cities, highlighting the best-performing branch (Branch C) and the lowest-performing branch (Branch B). At the bottom, more detailed analytics are included, such as a line chart forecasting April sales and a key influencers chart showing factors like product lines and gender that affect customer satisfaction.

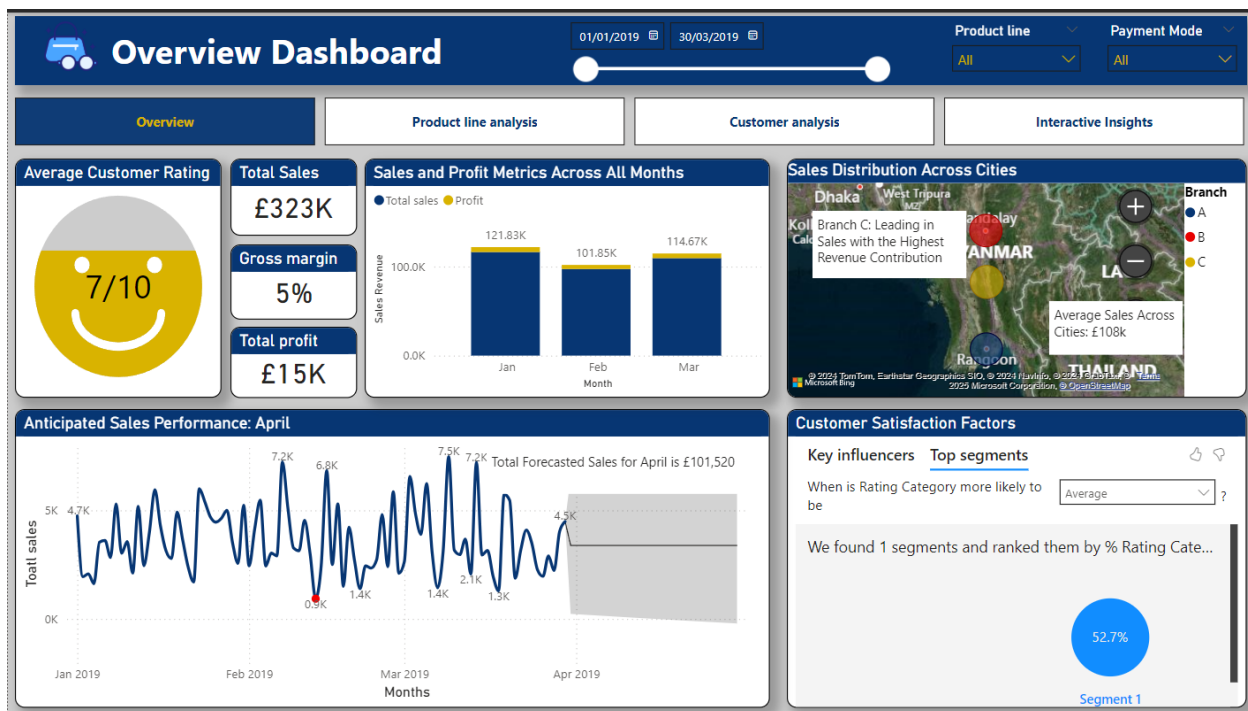


Figure 44: Overview dashboard

The dashboard uses color strategically to enhance readability. Shades of blue maintain a cohesive and professional appearance. Gold highlights both the highest sales branch (Branch C), drawing attention to top performance. and key metrics like the average customer rating Red is used sparingly to indicate areas of concern, such as Branch B's sales and the lowest sales day, February 13. This thoughtful use of color ensures the dashboard is visually appealing while

making it easy to identify significant insights. Additionally, a page navigator was included on all pages, allowing for seamless navigation between different sections of the dashboard, enhancing user experience and accessibility.

Product Dashboard

The Product Dashboard offers a comprehensive overview of product performance, profitability, and customer preferences. It is designed to provide key insights in a user-friendly format, with essential metrics displayed clearly to support decision-making.

At the top left, the Key Product Line Profit Contributions visualization highlights **Food and Beverages** as the top-performing product line, contributing **17%** to total profitability. The use of a blue bar chart enhances this insight, with blue tones maintaining consistency with the dashboard's overall design.

To the right, summary cards provide critical metrics such as **Average Unit Price (£56)**, **Total Transactions (1,000)**, **Total Items Sold (5,510)**, and **Total Cost of Goods Sold (£308K)**. These figures deliver a quick snapshot of sales volume and pricing trends, supporting a clear understanding of product performance.

In the center, the **Average Customer Satisfaction by Product Line** bar chart compares satisfaction ratings for each product line across genders. **Food and Beverages** and **Fashion Accessories** maintain high average ratings, while other product lines show slightly lower scores. Gender-specific shading, with varying shades of blue, visually distinguishes male and female responses, making it easy to identify trends.

The **Product Line Preference by Gender** visualization provides further insights into customer preferences, displayed using a stacked bar chart. **Food and Beverages** is the most preferred product line, with strong contributions from both male and female customers. Other popular categories include **Sports and Travel** and **Electronic Accessories**, showcasing diverse customer interests across genders.

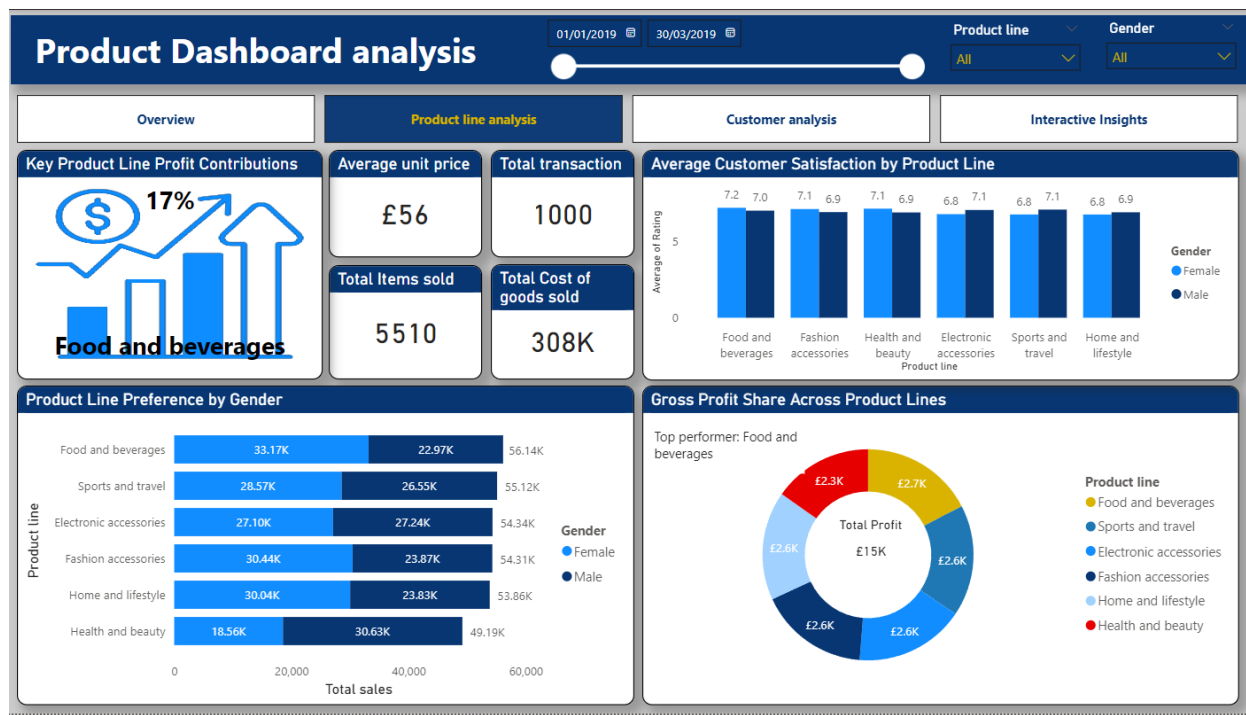


Figure 45: Product dashboard analysis

Finally, the **Gross Profit Share Across Product Lines** donut chart at the bottom right reinforces profit distribution among all product lines. **Food and Beverages** leads with **£2.7K** in profit, while other lines such as **Health and Beauty** contribute less significantly. Gold and blue tones are used to highlight high-performing segments, while red is sparingly used to indicate the least profitable product lines.

Customer Dashboard Analysis

The Customer Dashboard is well-organized and provides a clear view of customer satisfaction metrics and related insights. At the top left, the male average rating of 6.98 is displayed using an infographic design in gold, emphasizing good performance. Beside it, the female average rating of 6.96 is shown with the same style and color for consistency. In the center top, a clustered bar chart highlights the most popular payment method, with gold used to emphasize profit. On the top right, a pie chart illustrates customer satisfaction, showing 84.7% satisfaction and 15.3% dissatisfaction. 15.3% is highlighted in red to draw attention to areas for improvement.

At the bottom left, a stacked column chart breaks down revenue by gender, with light blue representing female customers, who generated the highest revenue of £175.88k, and darker blue for male customers. In the bottom center, a card displays the worst rating date, 9th

February, alongside an infographic showing the average worst day rating of 5.91. This is marked in red with a sad face emoji to highlight poor performance.

The bottom right features a line chart tracking customer satisfaction trend over time. The peak satisfaction on 22nd January, with a rating of 8.61, is highlighted in gold to show success, while the lowest rating date is marked in red to signify a concern.



Figure 46: Customer dashboard

The use of color is consistent and purposeful throughout the dashboard. Gold highlights high performance, such as profits and peak ratings, while red is used to signal areas needing improvement, like dissatisfaction and low ratings. This clear structure and thoughtful design make the dashboard easy to understand and visually impactful.

The "Interactive Insights" dashboard provides a dynamic approach to exploring sales data, offering a decomposition tree and Q&A visual to enable detailed data exploration. The decomposition tree breaks down total sales by branch, product line, and customer type, offering a clear hierarchy that helps identify which segments contribute most significantly to overall performance. This tool also serves as a validation check for related visuals, confirming their accuracy and consistency across different analytical views.

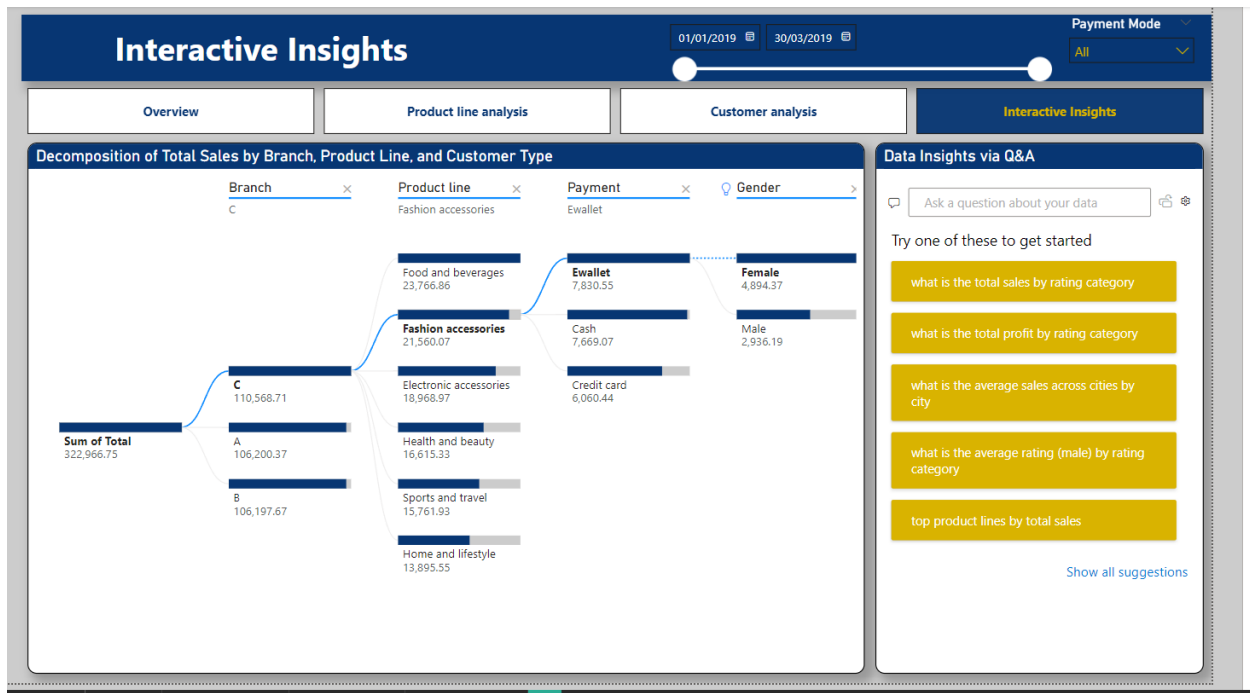


Figure 47: Interactive Insights dashboard

The Q&A visual enhances interactivity by allowing users to pose specific questions and receive instant, tailored insights. This feature empowers stakeholders to explore the data independently, uncovering trends and relationships that might otherwise go unnoticed. Together, these visuals provide a flexible and interactive experience, making it easier to delve deeper into the data and extract actionable insights.

5 Self-assessment

Report Section	Description	Grade your work from 0 to 100
Report Structure	The report is well-written, and it contains all the relevant sections	90
Data Pre-processing and Data Modelling	Many pre-processing steps have been applied. The data model is well-structured	95
Dax and M language	Both DAX and M Language have been extensively used in the report	88
Dashboard Design	The dashboard contains a variety of charts, including advanced ones not covered in the module.	90
Average		91