[supermarket sales]

Data Wrangling Project

BY: ABDELAZIM MOHAMED

Introduction

Real-world sales data rarely comes clean. Using **Python** and its libraries, I gathered data from a supermarket sales dataset, assessed its quality and tidiness, and cleaned it to make it ready for analysis. This process is called **data wrangling**.

I documented the entire wrangling process in a **Colab Notebook** included in the project folder and showcased the results through analyses and visualizations using **Pandas, Matplotlib, and Seaborn**.

The dataset that I wrangled (and analyzed and visualized) contains transactional sales data from a supermarket. It includes details such as **Invoice ID, Branch, City, Customer Type, Gender, Product Line, Unit Price, Quantity, Tax, Total, Date, Time, Payment Method, and Rating**.

The goal of this project is to clean and transform the data, then perform exploratory data analysis (EDA) to answer key business questions, discover sales trends, compare branches and product lines, analyze customer behavior, and ultimately extract insights that can support **data-driven decision-making**.

BY:
ABDELAZIM
MOHAMED

Data Visualization

0- Key Performance Indicators (KPI Cards)

| Total Sales	Average Sales	Total Quantity
$322,966.75	$322.97	5,510

To provide a quick overview of the dataset, three KPI cards were created:

- **Total Sales: Shows the overall sales revenue generated by all branches.**
- **Average Sales: Displays the average value of sales transactions.**
- **Total Quantity: Represents the total number of products sold.**

These KPI cards give management a fast and clear summary before moving into detailed visualizations.
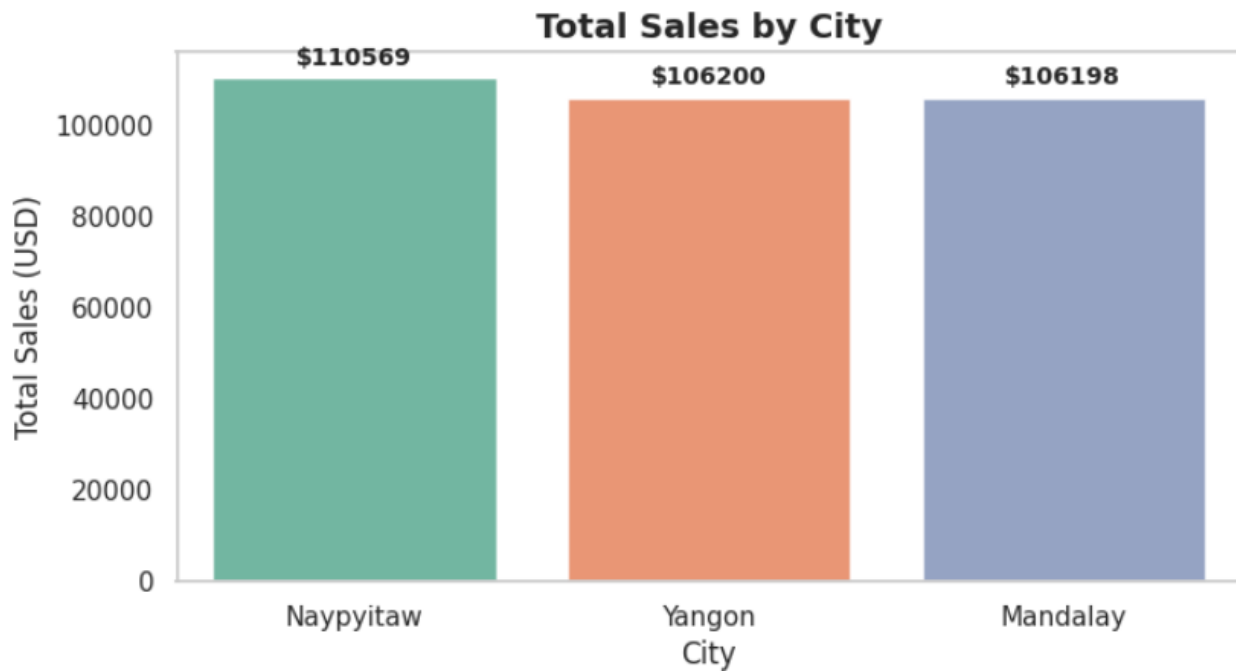
1- Sales Insights

1-1- Total Sales by Branch

Total Sales by Branch

A ($106,200)   32.9%

C ($110,569)   34.2%

B ($106,198)   32.9%
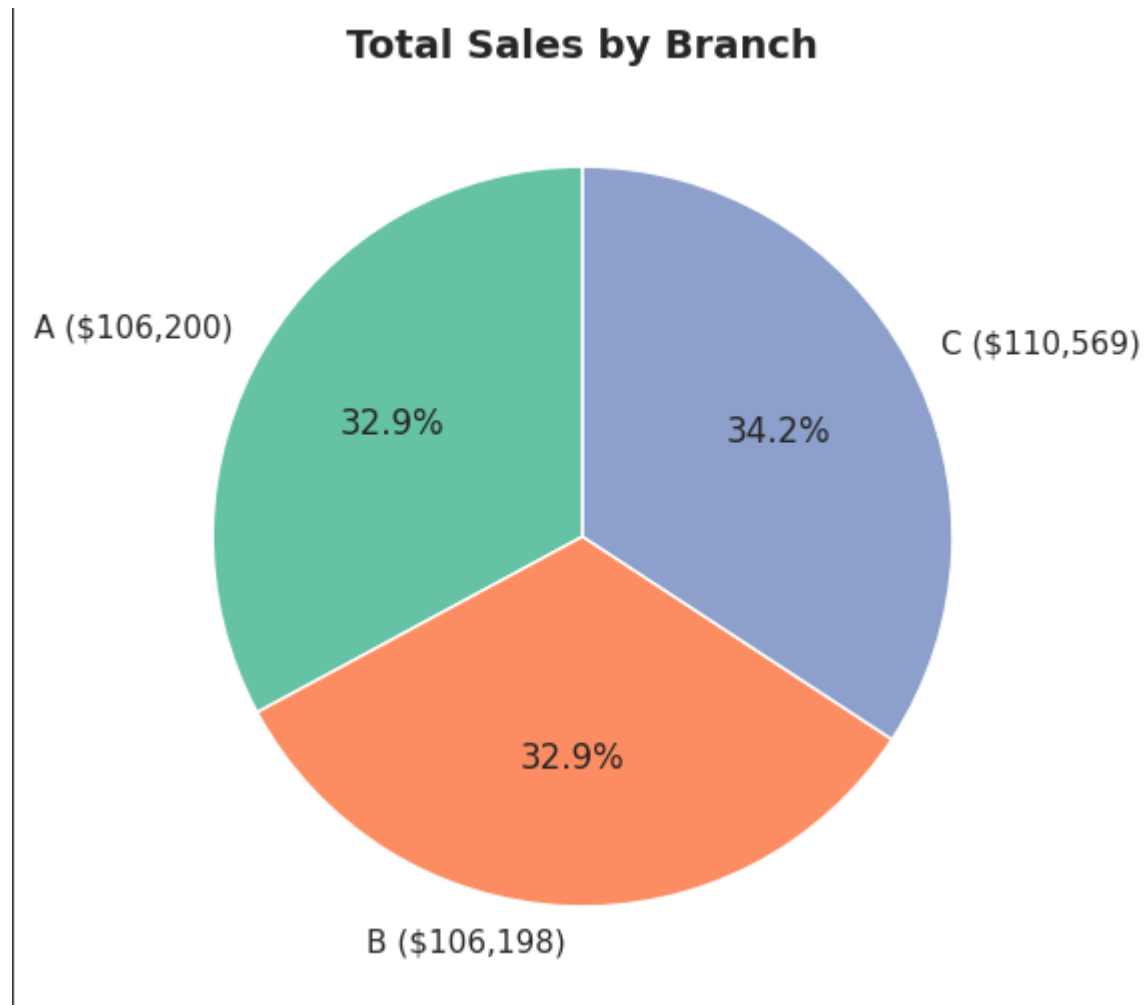
This bar chart shows the total sales achieved by each supermarket branch.
As we can see, there are clear differences between branches: Branch C generates the highest revenue, followed by Branch B, while Branch A records the lowest.
This indicates that Branch C is the top performer in terms of sales volume.
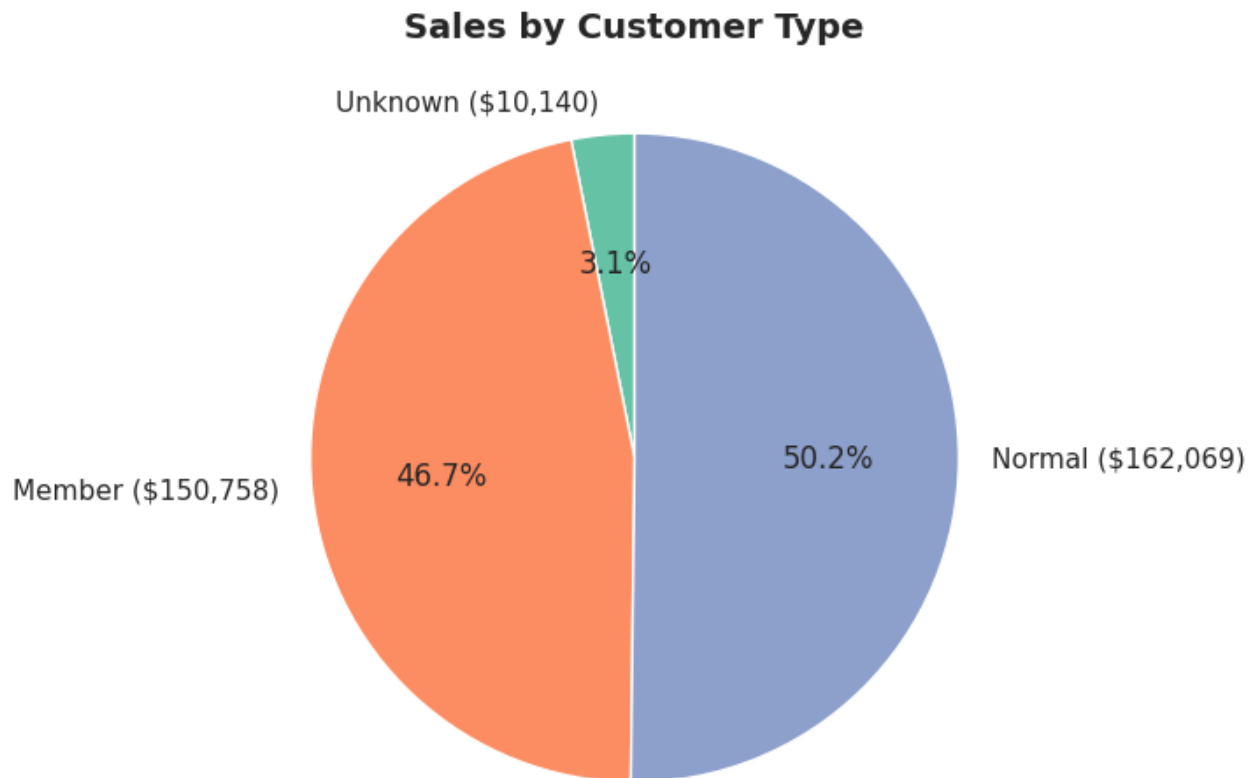
1-2- Total Sales by City

This bar chart presents the total sales distribution across cities.
It highlights which city contributes the most to overall revenue, giving insight into
geographic performance.

2- Customer Insights

2-1- Sales by Customer Type

Sales by Customer Type

Unknown ($10,140)

3.1%
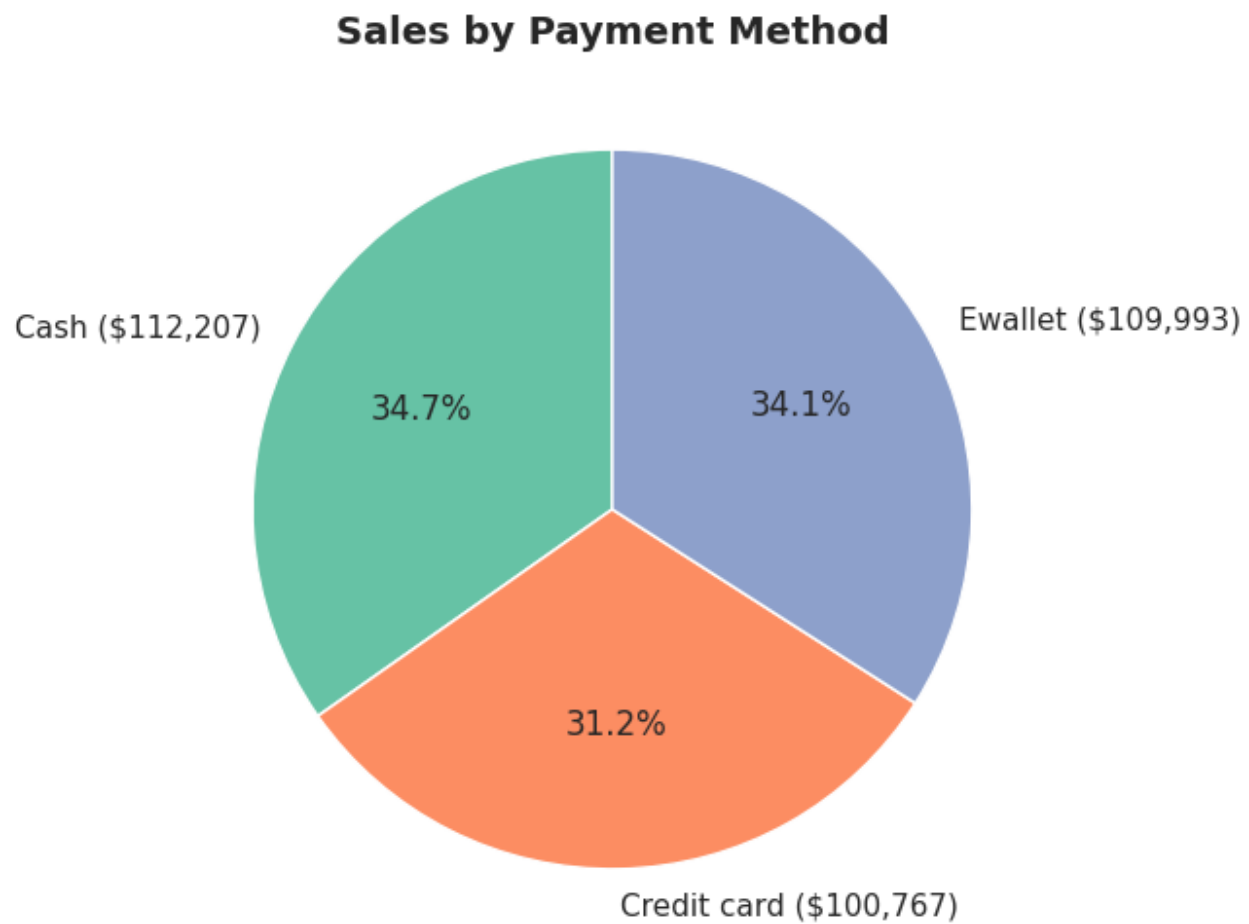
46.7%

50.2%   Normal ($162,069)

Member ($150,758)

This pie chart shows the share of sales between customer types (Normal, Member, Unknown).
As we see, Normal customers slightly contribute more to total sales compared to Members, while Unknown accounts are very minimal.

2-2- Sales by Payment Method

Sales by Payment Method

Cash ($112,207)

34.7%

Ewallet ($109,993)

34.1%
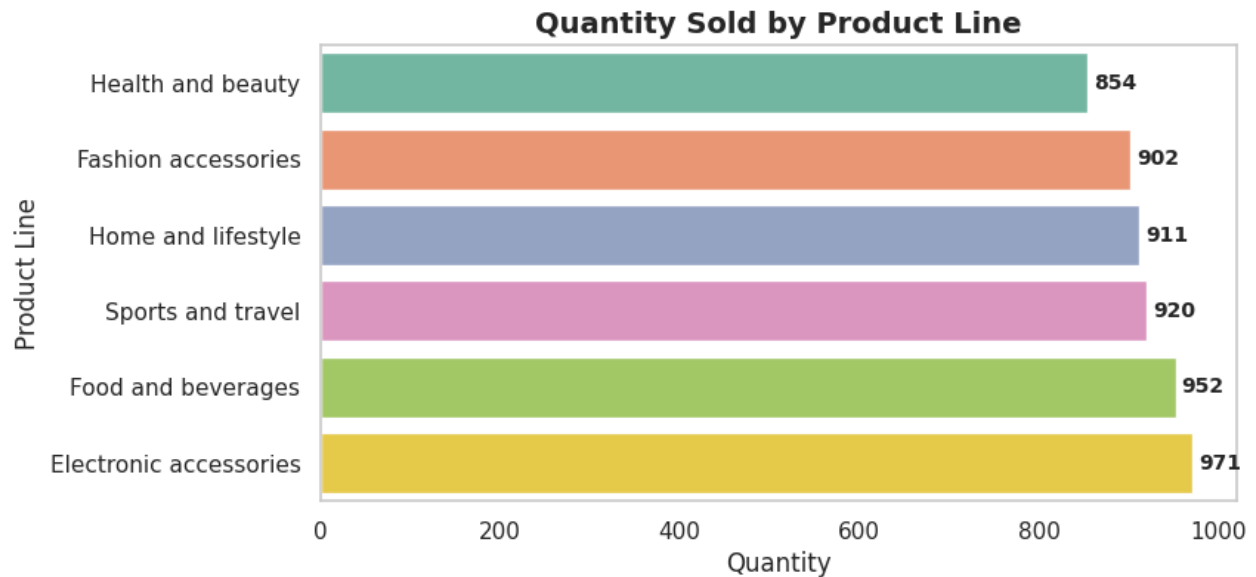
31.2%

Credit card ($100,767)

This pie chart illustrates the distribution of sales by payment method.
We can observe that Credit Card and Cash dominate as the preferred methods, while E-Wallet contributes less.
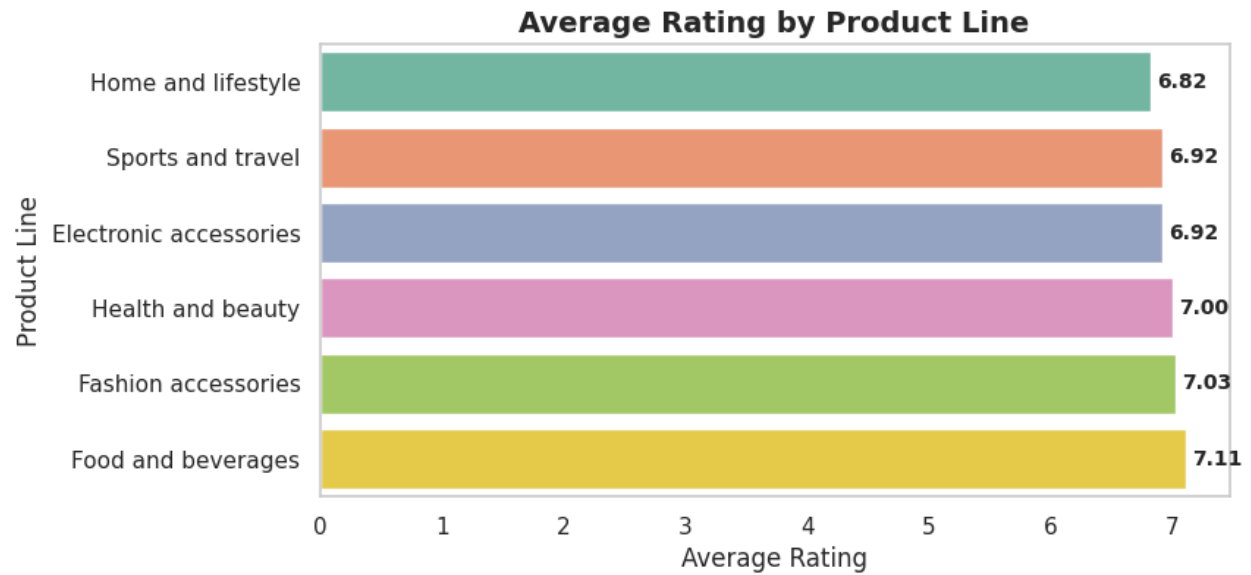
3- Product Insights

3-1- Quantity Sold by Product Line

**This bar chart shows how many units were sold from each product line.
As we see, some product lines like Food and Beverages contribute more significantly in terms of sales volume.**

3-2- Average Rating by Product Line

Average Rating by Product Line

**This bar chart highlights the average customer rating for each product line.
The chart shows which product lines customers rated more positively, providing insight into customer satisfaction.**

3-3- Distribution of Sales (Boxplot by Product Line)

**Distribution of Total Sales per Product Line**
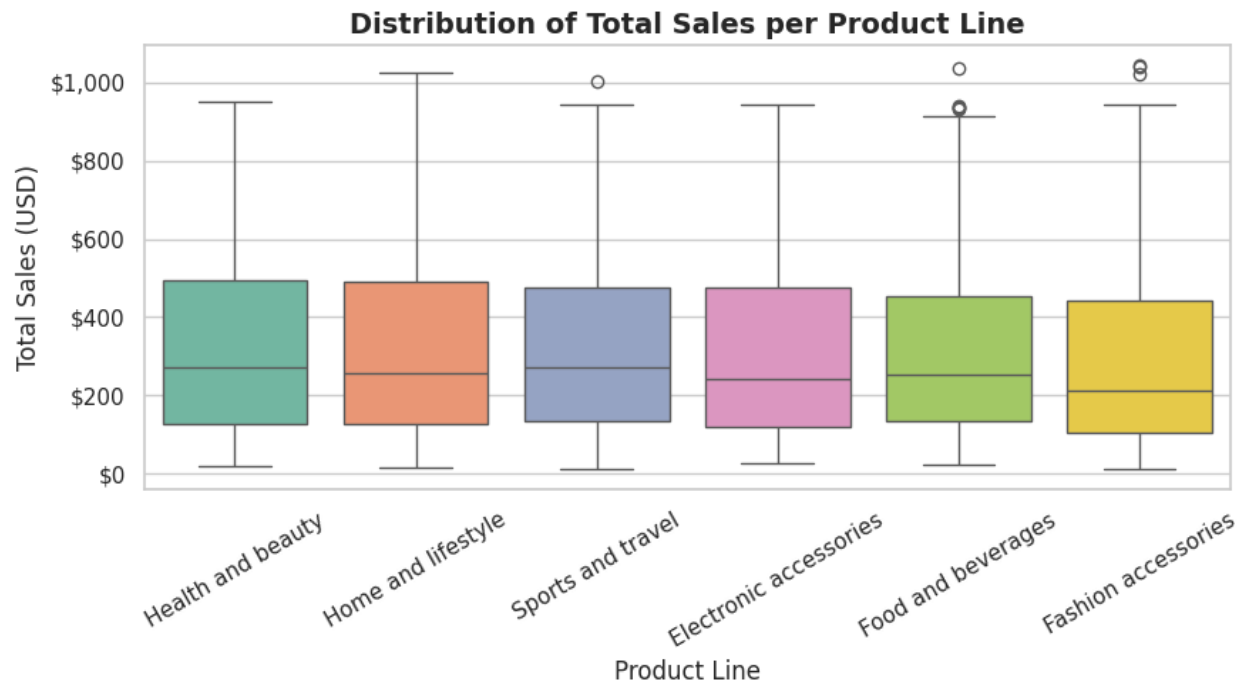
**This boxplot shows the distribution of total sales across different product lines.
It highlights not only the median sales values but also the spread and outliers in each product category.
From this chart, we can identify product lines with higher variability in sales, meaning some transactions generate very high totals compared to others.**

BY:
ABDELAZIM
MOHAMED

4- Sales Trends

4-1- Daily Total Sales
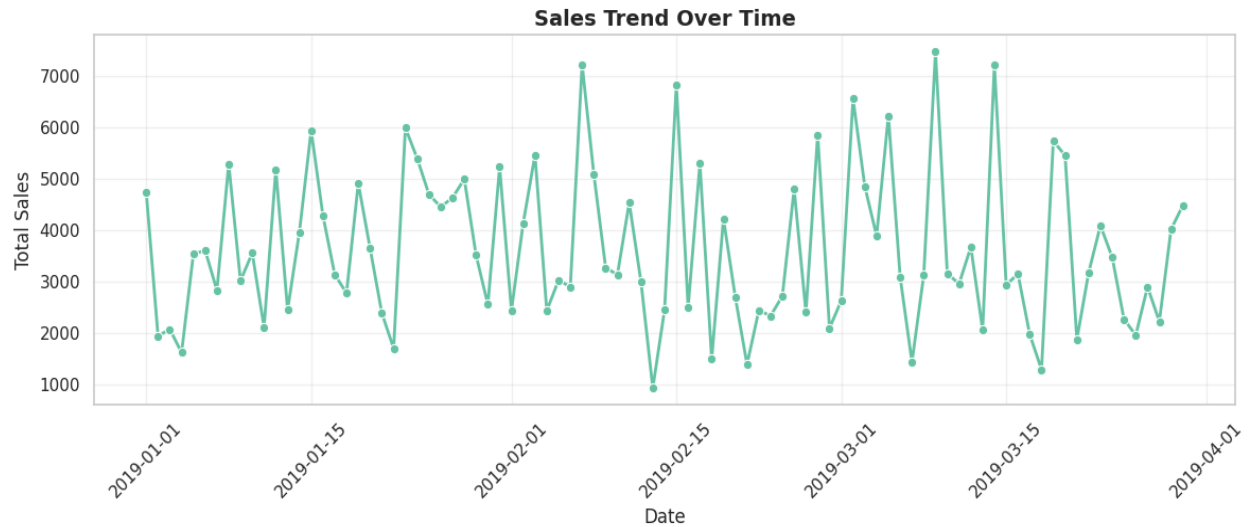
**This line chart illustrates sales trends across different dates.
As we can see, sales fluctuate daily but follow a general trend that can help identify peak days.**

Summary Insight

From these visualizations, we can conclude that **Branch C and its city lead in revenue**, **Normal customers contribute slightly more than Members**, **Credit Card and Cash are the dominant payment methods**, and **Food & Beverage-related product lines show strong performance**.

The KPI cards also emphasize that overall sales revenue is strong, with a healthy average transaction value and significant product quantities sold.

Thank You