



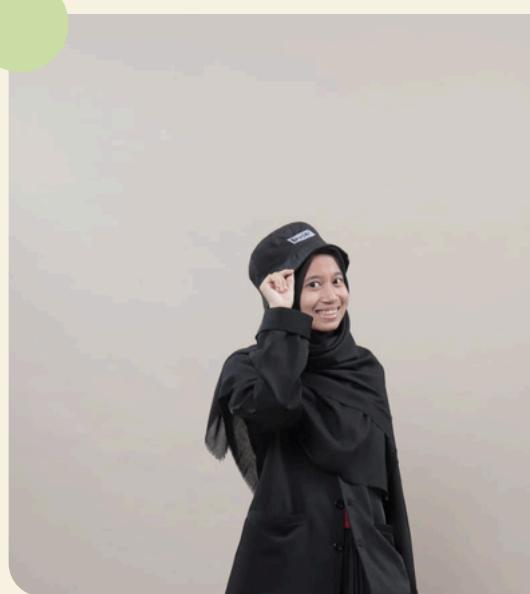
# SUPERMARKET

## CRISP-DM

GROUP 3



# TEAM MEMBERS



Alifia Qolbiyatus Syifa



Satria Putra Dharma  
Prayudha



Fadiyah Nur Ayu Ning  
Tyas



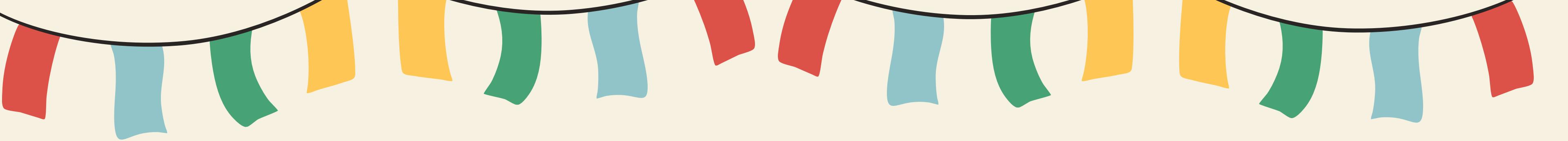
Suhartanti



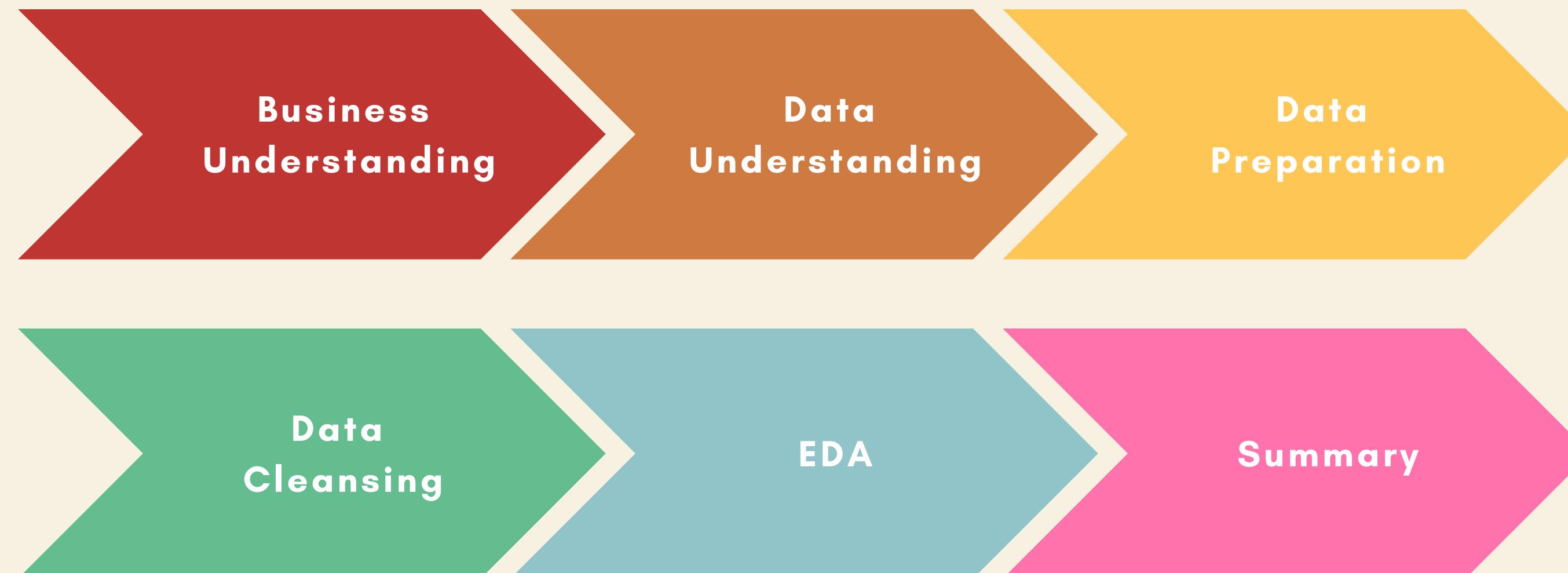
Elshaddai L. Tasane



Muhammad Dimas  
Reza Aditya



# WORK FLOW



# BUSINESS UNDERSTANDING





# BACKGROUND

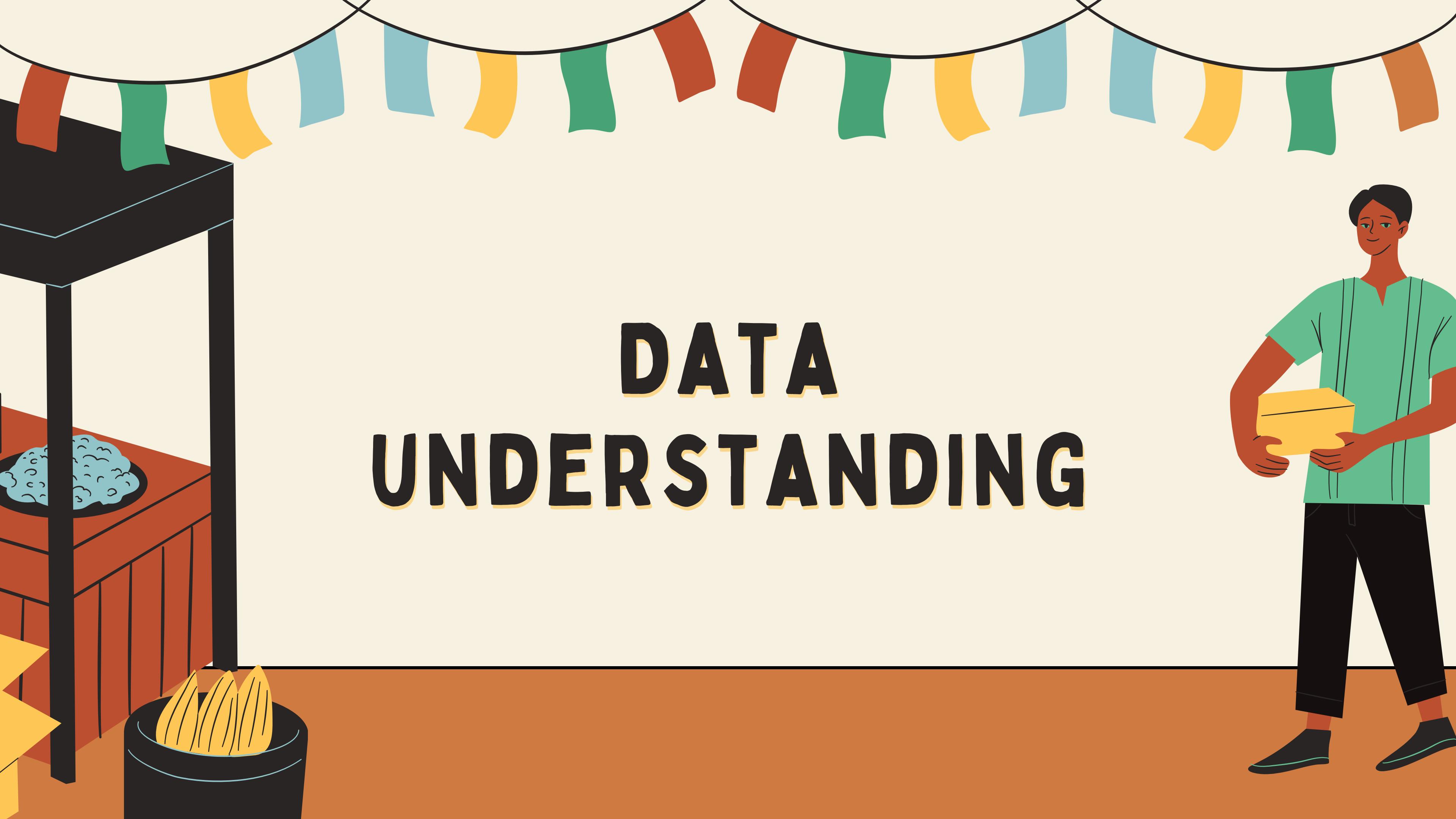
The growth of supermarkets in densely populated cities is on the rise, accompanied by intensified market competition. This phenomenon underscores the importance of analyzing historical sales data to understand market trends and competitive dynamics.

The dataset comprises three months of historical sales data from three different branches of a supermarket company. Each branch's sales data offers insights into consumer behavior, product performance, and overall market trends.



# MAIN QUESTION

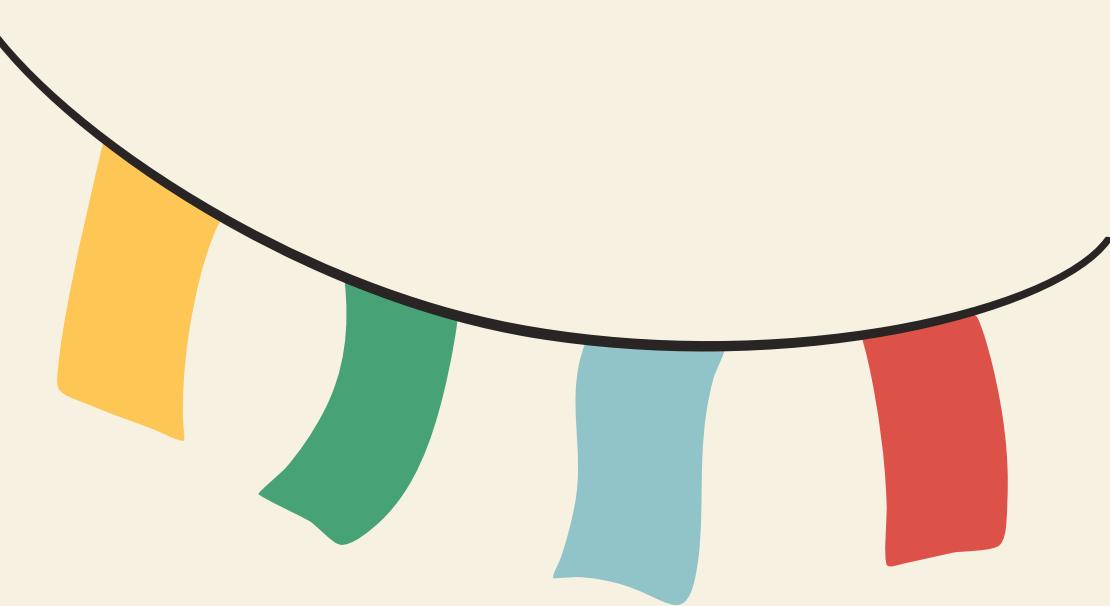
- 1 Which cities are the top performers in terms of sales?
- 2 What are the most common payment methods used by customers?
- 3 What types of products do people buy the most by gender?
- 4 Which product lines are the best-selling?
- 5 Which types of customers generate the most revenue?



# DATA UNDERSTANDING



# DATA UNDERSTANDING



## DATA

The data used is data of Supermarket transaction obtained from Kaggle : [kaggle.com/datasets/aungpyaeap/supermarket-sales](https://kaggle.com/datasets/aungpyaeap/supermarket-sales)

## TIME

The data used is Supermarket Transaction from January 2019 until March 2019

## TOTAL

The data used consists of 10 columns and 1000 rows



# DATA DICTIONARY

|                        |  |                       |   |
|------------------------|--|-----------------------|---|
| • <b>BRANCH</b>        | Branch where the transaction (City Code) took place. | • <b>QUANTITY</b>     | Number of units purchased.  |
| • <b>INVOICE ID</b>    | Unique identifier for each transaction.              | • <b>TAX</b>          | 5% tax fee for customer buying  |
| • <b>CITY</b>          | City where the branch is located.                    | • <b>TOTAL</b>        | Total price including tax   |
| • <b>CUSTOMER TYPE</b> | Type of customer in supermarket                      | • <b>DATE</b>         | Date of purchase  |
| • <b>GENDER</b>        | Gender of the customer.                              | • <b>PAYMENT</b>      | Payment used by customer for purchase   |
| • <b>PRODUCT LINE</b>  | Category of the product purchased.                   | • <b>GROSS MARGIN</b> | Gross margin percentage   |
| • <b>UNIT PRICE</b>    | Price per unit of the product.                       | • <b>COGS</b>         | Cost of goods sold  |
| • <b>PRODUCT LINE</b>  | Category of the product purchased                    | • <b>GROSS INCOME</b> | Gross income  |
|                        |  | • <b>RATING</b>       | Customer stratification rating on their overall shopping experience (On a scale of 1 to 10) |



# DATA CLEANSING



## MISSING VALUE

By checking the information from the data in the dataset used, it is known that there are no missing values , so no data imputation is performed on the dataset used.

## DATA TYPE INFORMATION

The data type of the 'Date' column was converted from object to datetime64 to facilitate time-based analysis. Additionally, the DataFrame index was set to the 'Date' column to streamline time-based operations.

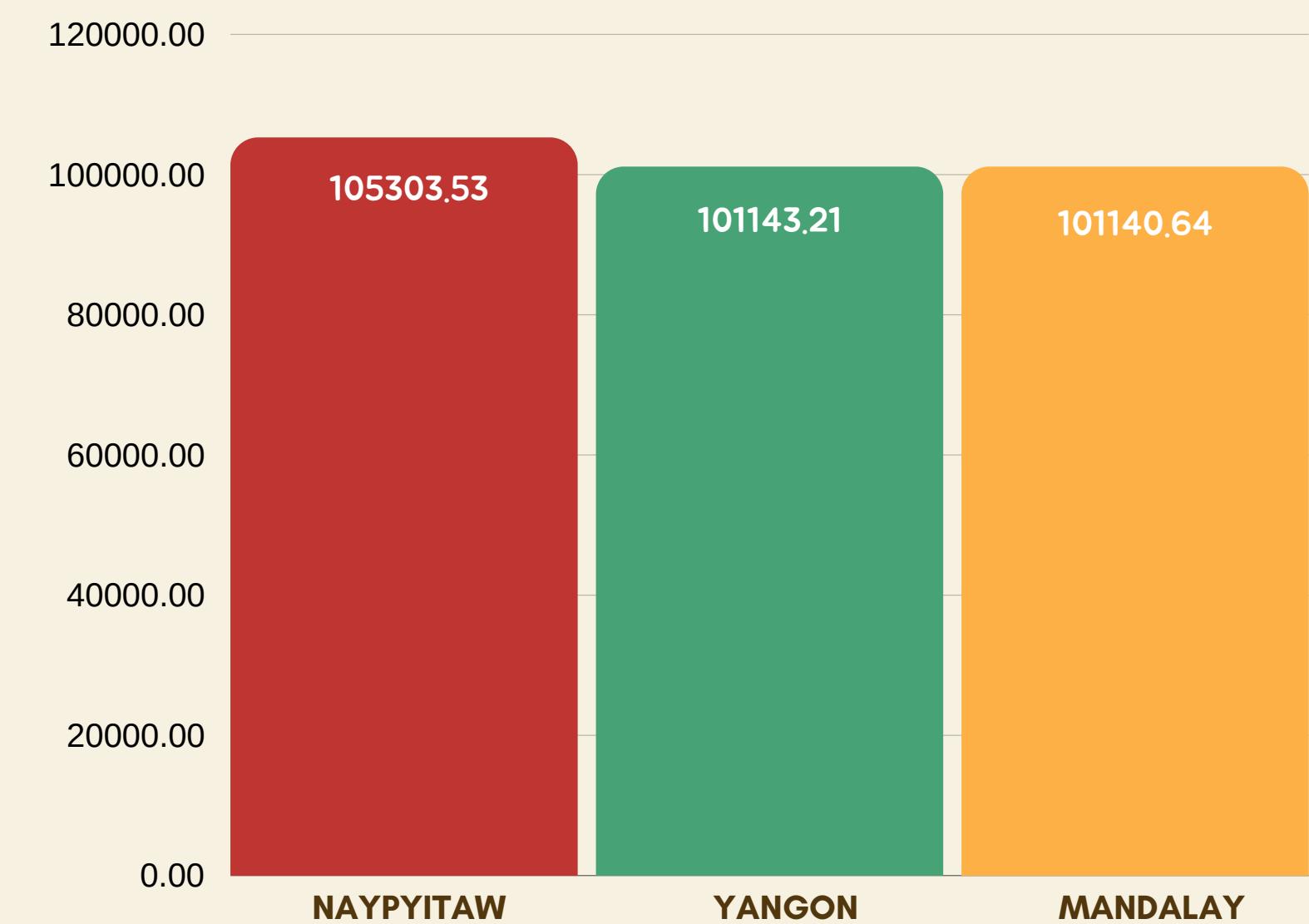


# **EXPLORATORY DATA ANALYSIS**

# WHICH CITIES ARE THE TOP PERFORMERS IN TERMS OF SALES?

From the Bar Chart, we can see that **Naypyitaw** emerged as the **leader** in sales when analyzing key metrics such as **total revenue**, average sales per transaction, and sales growth rate.

Its consistently **strong performance** in these areas confirms its position as the **city** with the **highest sales**, providing important **insights** into its economic prowess and dominance in the market.

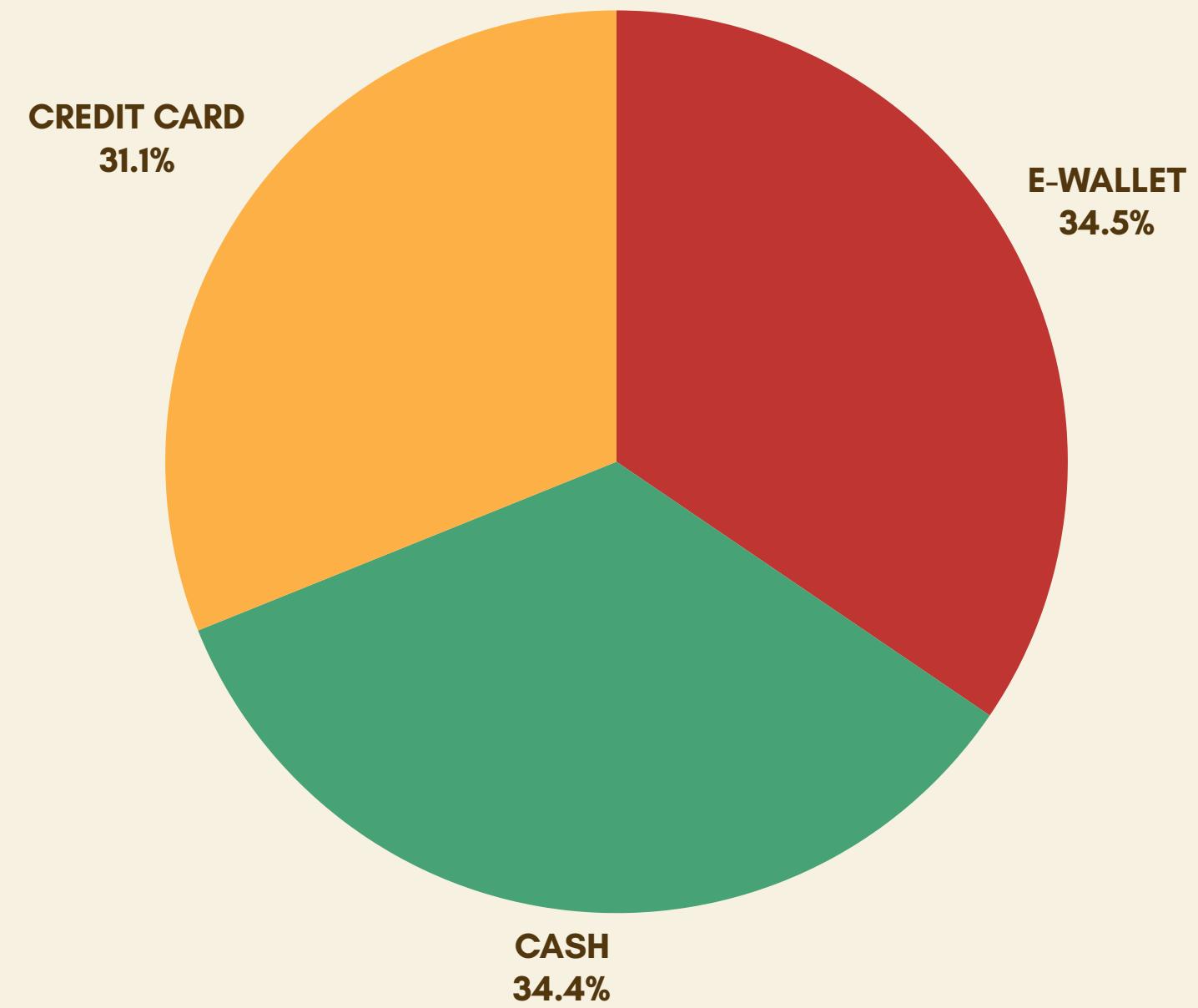




# WHAT ARE THE MOST COMMON PAYMENT METHODS USED BY CUSTOMERS?

From the Pie chart, we can see that **E-wallets** are growing in **popularity** compared to cash, indicating that more and more people prefer **digital transactions**. However, **cash** is still widely used, indicating that many people still prefer **traditional methods**. Credit card usage is not very high, suggesting that there is still room to encourage its use or make it more attractive.

This suggests that businesses can capitalize on the **popularity** of e-wallets, **accommodate** those who still prefer cash, and make credit cards more attractive to cater to different **customer preferences**.

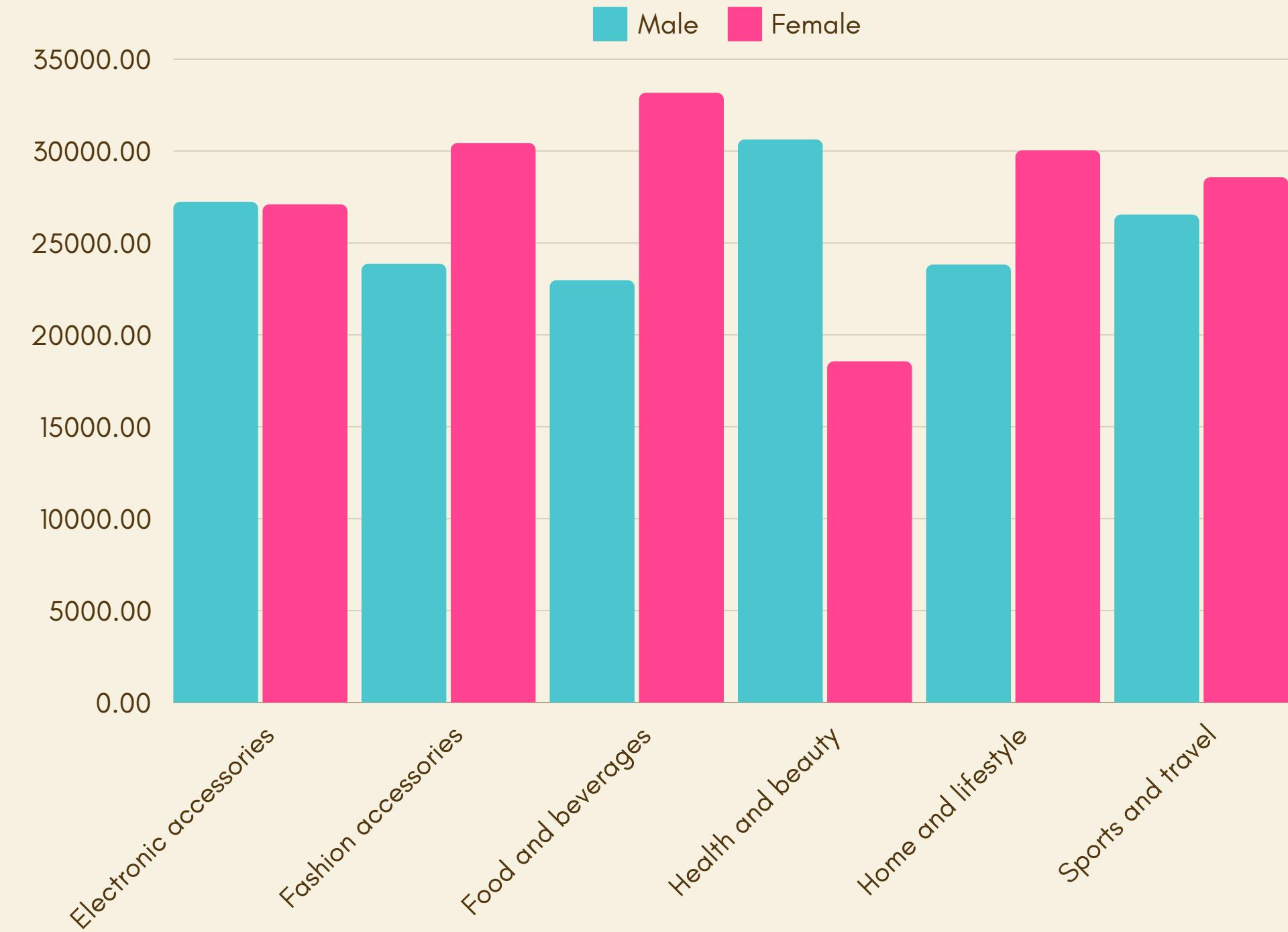




# WHAT TYPES OF PRODUCTS DO PEOPLE BUY THE MOST BY GENDER?

Based on the chart, it is clear that there are **differences** in the types of **products preferred** by different **genders**. **Men** tend to gravitate towards purchases in the "**Food and Beverage**" category, whereas **women** show a stronger inclination towards products categorized under "**Health and Beauty**".

These gendered buying patterns indicate different preferences and **priorities** among consumers, highlighting the importance of **gender-specific** marketing strategies and product **offerings** in meeting the needs and **interests** of diverse consumers.



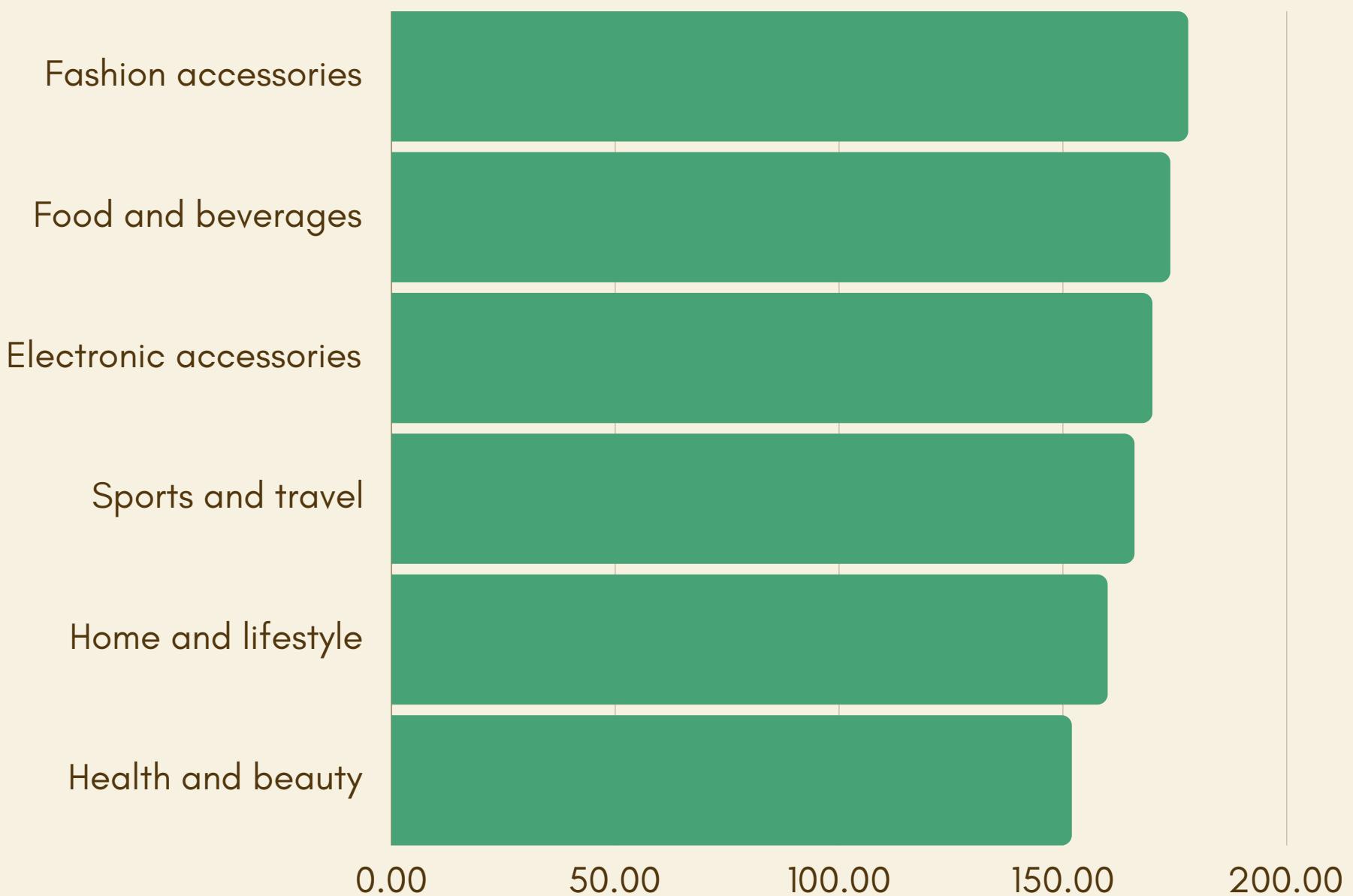


# WHICH PRODUCT LINES ARE THE BEST-SELLING?



Based on the chart, **Fashion accessories** emerge as the **best-selling** product line, with **178 transactions**, closely followed by Food and beverages with 174 transactions. Electronic accessories secured the third position with 170 transactions.

This data emphasizes the **strong demand** for fashion accessories and food and beverages, highlighting the importance of catering to **customer preferences** across various product categories.



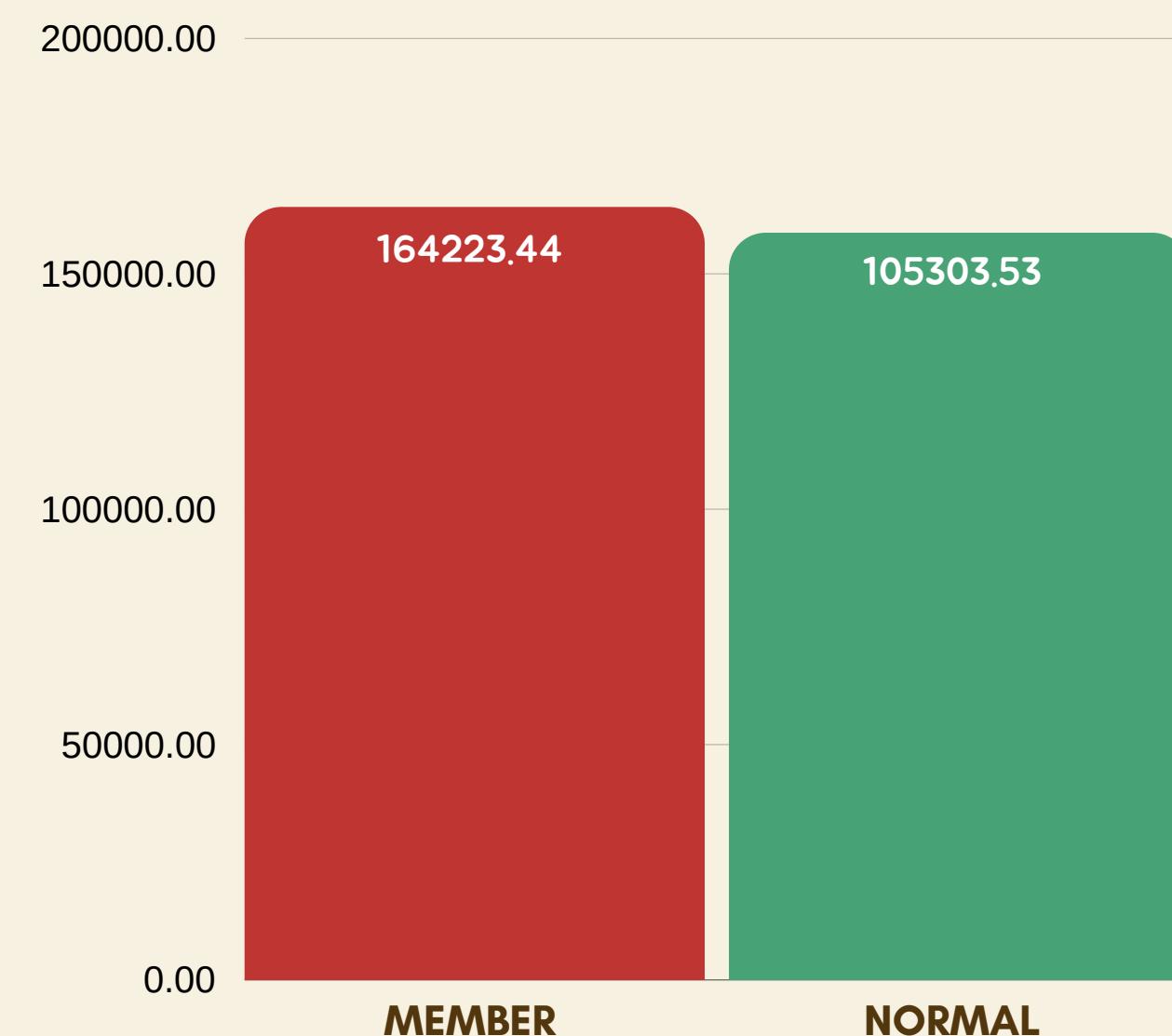


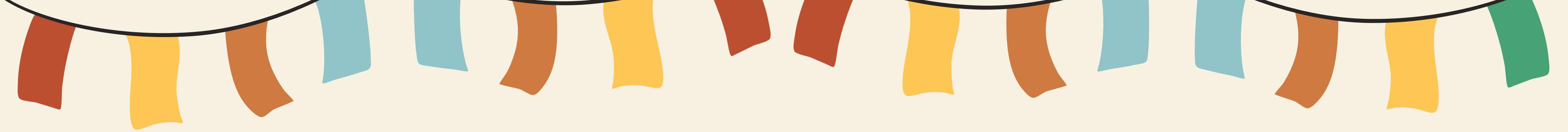
# WHICH TYPES OF CUSTOMERS GENERATE THE MOST REVENUE?

From the data presented, an analysis of customer types in the sales records unveils two distinct categories: regular **customers** and **members**.

Among these, members stand out as the primary revenue generators, boasting a substantially **higher total** spend compared to regular customers. This observation underscores the pivotal

role of **customer loyalty** programs and highlights the potential **advantages** of offering **membership** incentives to bolster both revenue generation and customer retention efforts.





# SUMMARY

- 1 **Naypyitaw** leads in sales, indicating market dominance
- 2 **E-wallets** are **growing**, **cash** remains **popular**, and **credit card** use is **low**.
- 3 Gender-specific preferences: men favor "**Food and Beverage**", women prefer "**Health and Beauty**"
- 4 **Fashion accessories** and **food/beverages** are **top sellers**
- 5 **Revenue** is primarily driven by **members**, underscoring the importance of loyalty programs



**THANK YOU**  
**GROUP 3**

**CHUR**