# ABOUT DATASET

The growth of supermarkets in most populated cities are increasing and market competitions are also high. The dataset is one of the historical sales of supermarket company which has recorded in 3 different branches for 3 months data. The data source come from:
https://www.kaggle.com/datasets/aungpyaeap/supermarket-sales

**Attribute information**:

Invoice id: Computer generated sales slip invoice identification number

Branch: Branch of supercenter (3 branches are available identified by A, B and C).

City: Location of supercenters

Customer type: Type of customers, recorded by Members for customers using member card and Normal for without member card.

Gender: Gender type of customer

Product line: General item categorization groups - Electronic accessories, Fashion accessories, Food and beverages, Health and beauty, Home and lifestyle, Sports and travel

Unit price: Price of each product in $

Quantity: Number of products purchased by customer

Tax: 5% tax fee for customer buying

Total: Total price including tax

Date: Date of purchase (Record available from January 2019 to March 2019)

Time: Purchase time (10am to 9pm)

Payment: Payment used by customer for purchase (3 methods are available – Cash, Credit card and Ewallet)

COGS: Cost of goods sold

Gross margin percentage: Gross margin percentage

Gross income: Gross income

Rating: Customer stratification rating on their overall shopping experience (On a scale of 1 to 10)

**Data Integrity:**

The data is just collected with 3 months from January to March in 2019, the period is too short, so analysis results may be bias.

**Goals**

To find out:

- Which type of customer generate highest gross income, sale

- Which method of payment preferred by customer type

- Products which are bought more by male, female?

- Sale and gross income change over years

- Which day of week contribute most gross income

From there, I will provide some recommendations to increase sales and profit for supermarket.

# DATA CLEANING

I used SQL for cleaning process. In this process, I will check duplication, null values, accuracy and remove unnecessary columns. File detail about cleaning process using SQL, I have attached in supermarket_sale folder.
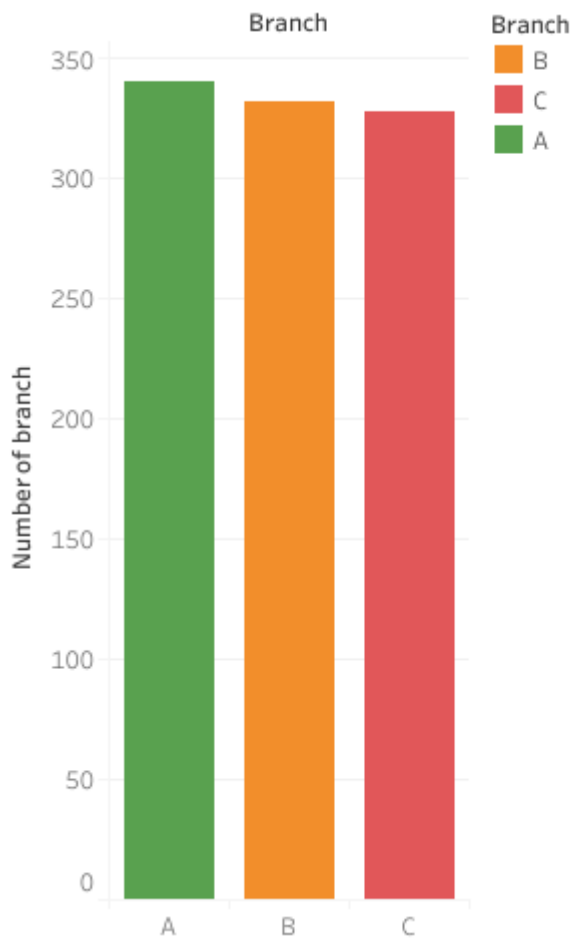
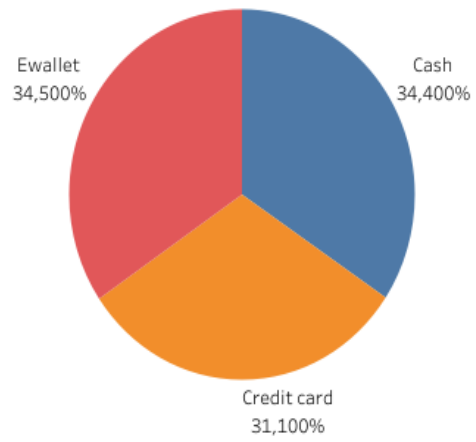# Analyze

- Customer Type:
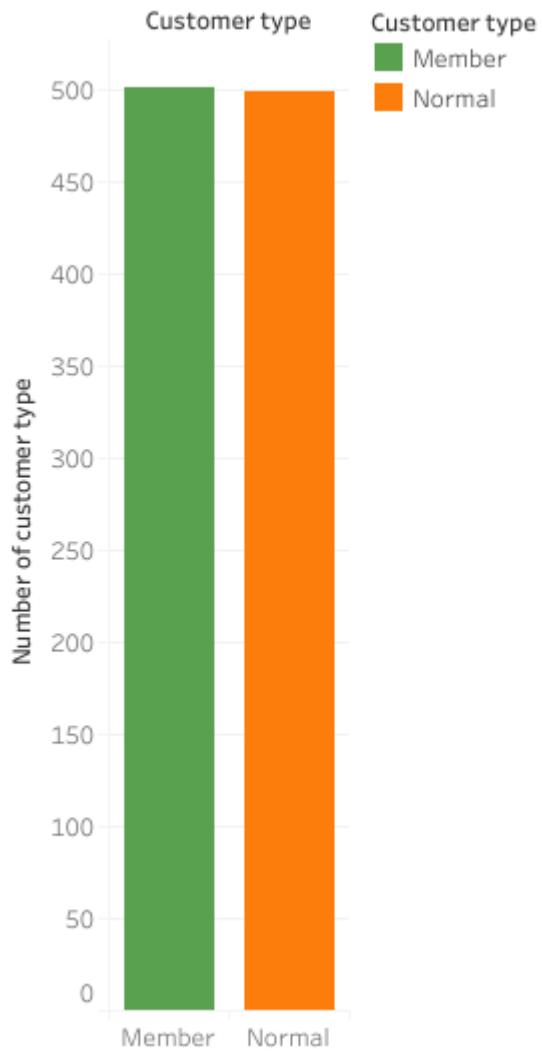


Number of member is slightly higher than Normal.
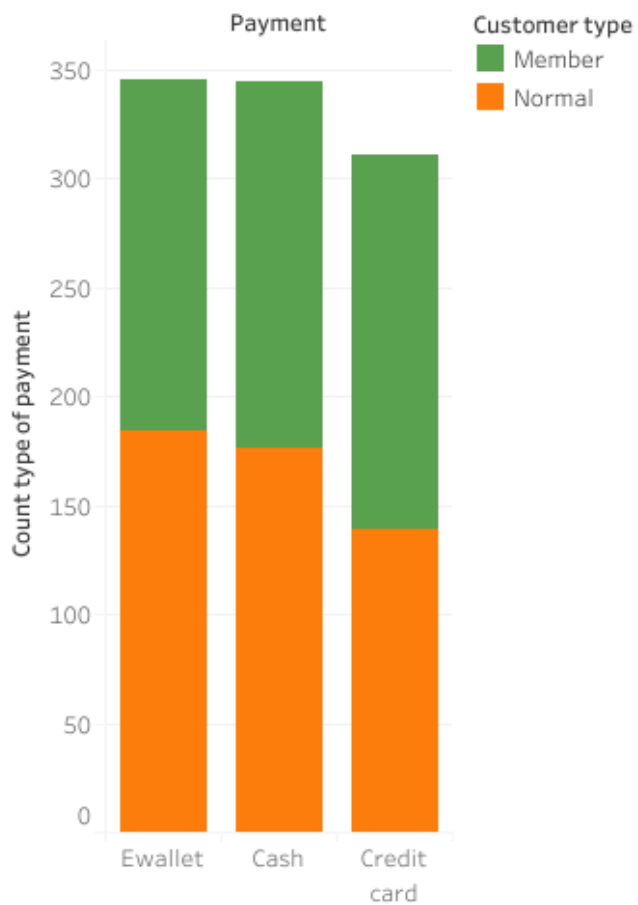
- Branch:



Through graph, we can see that branch A is more popular than B and C, but the difference is not too much.

- Payment:



Payment by Ewallet and cash is more popular than Credit Card.
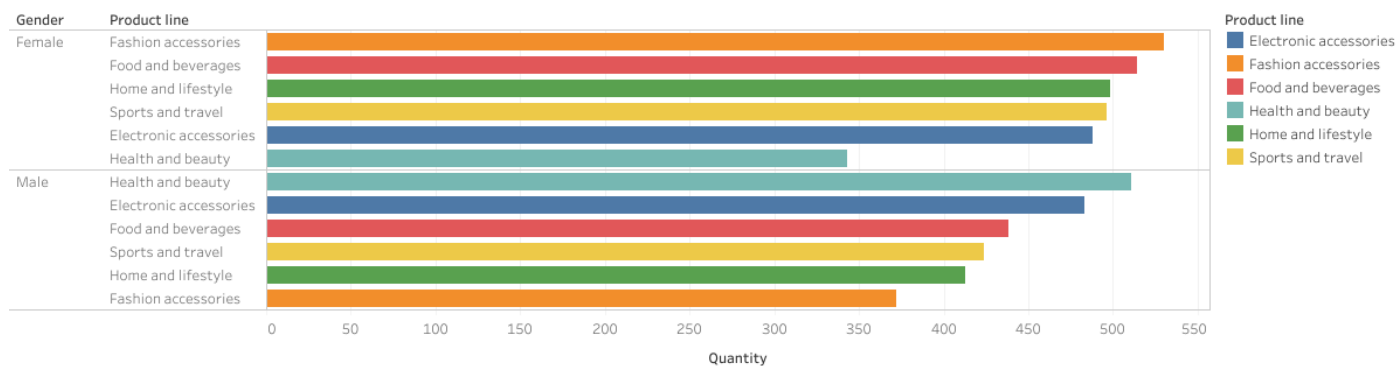
- Next we will find out which payment method is more preferred by each type of customer:



Member used credit card for payment more than Normal.
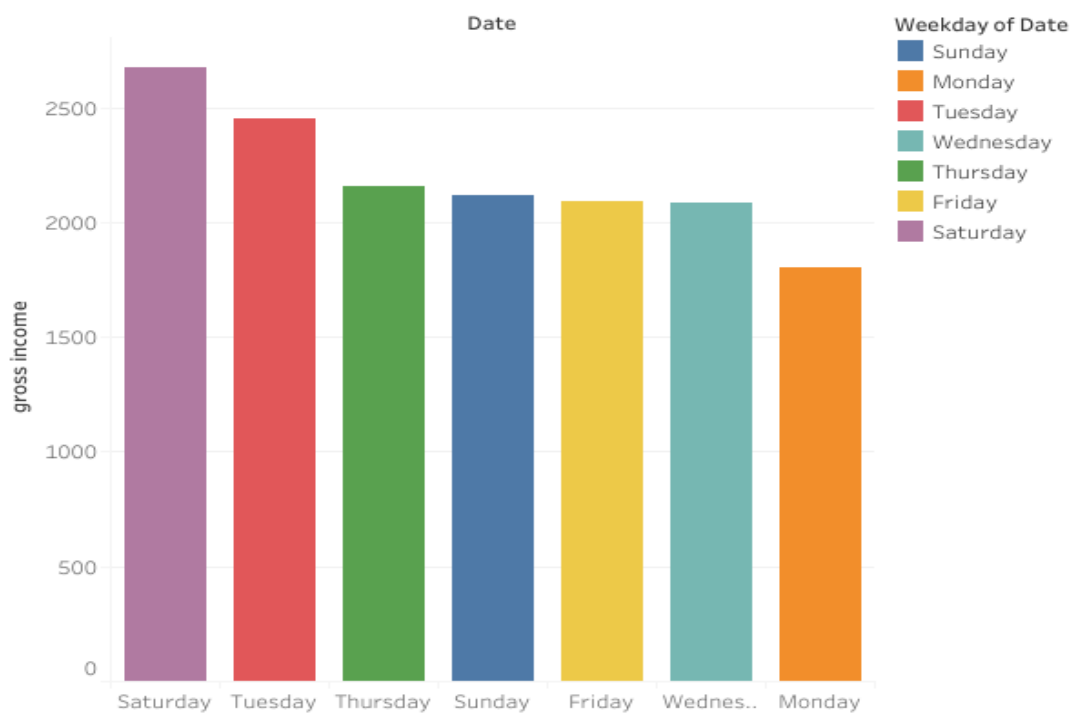
- Product Line by gender:



There is a contrast between men and women. While women prefer Fashion accessories, food and beverages and less interested in Healthy and beauty, Electronic accessories, men is interested in health and beauty, electronic accessories.

- Sale and Gross income change over time:
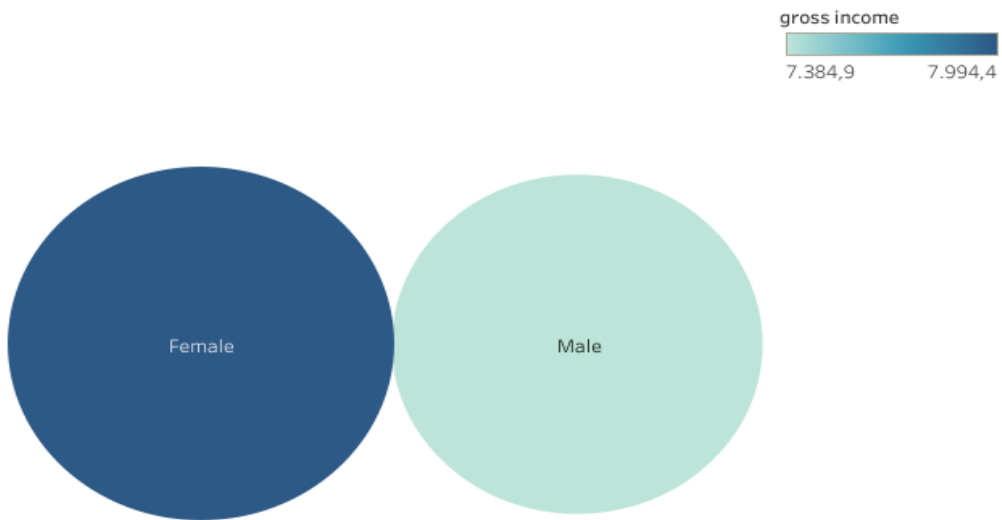
Sale and Gross Income change
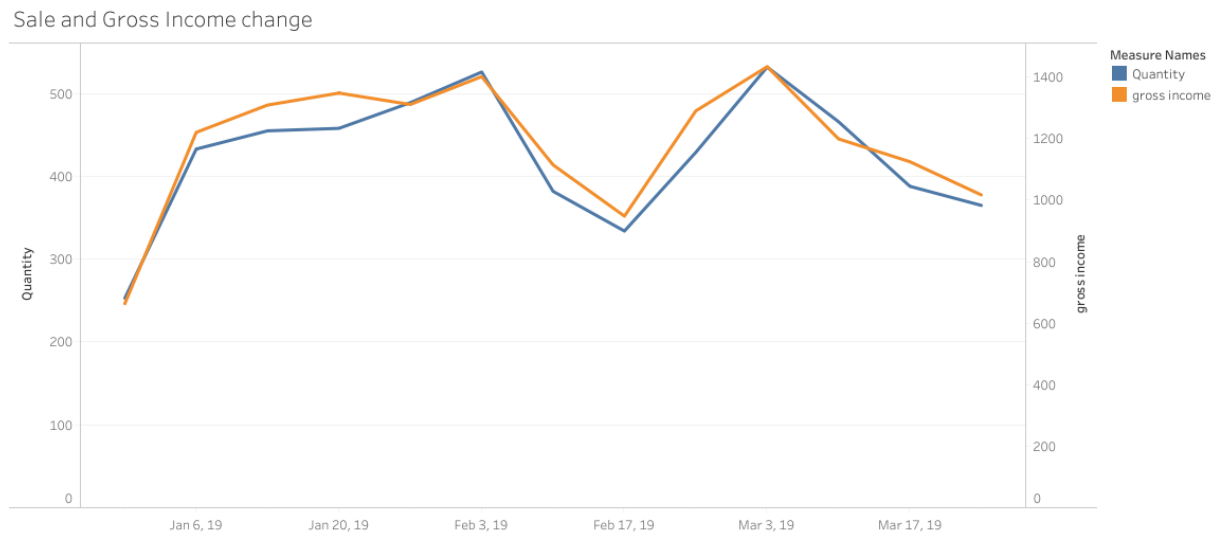


Sale and Gross income is high on January, decrease on February.

- Gross income by weekday:



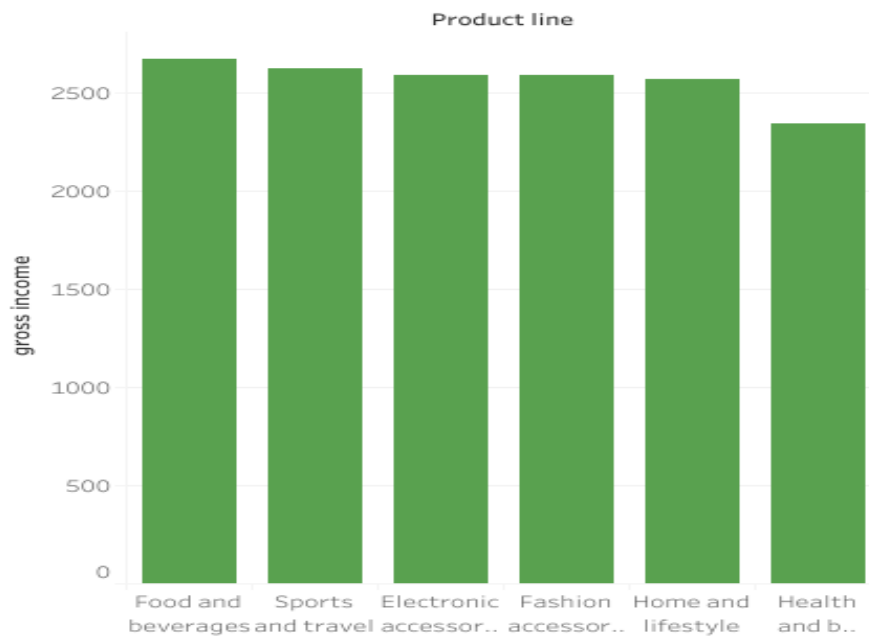Saturday and Tuesday are two days of week generate high profit, while Monday has the least profit.

- Gender that generate the most profit and sale:

gross income

7.384,9            7.994,4



Female contribute more to sales and gross income.

- Profit by product line:

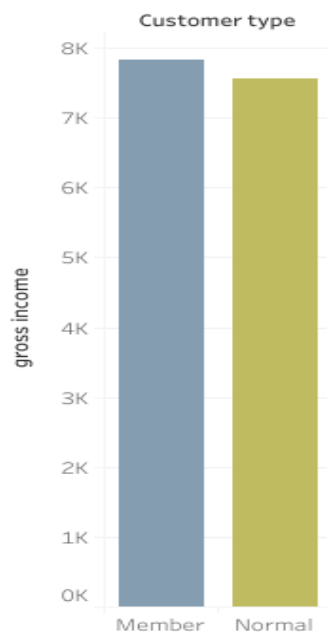**Product line**



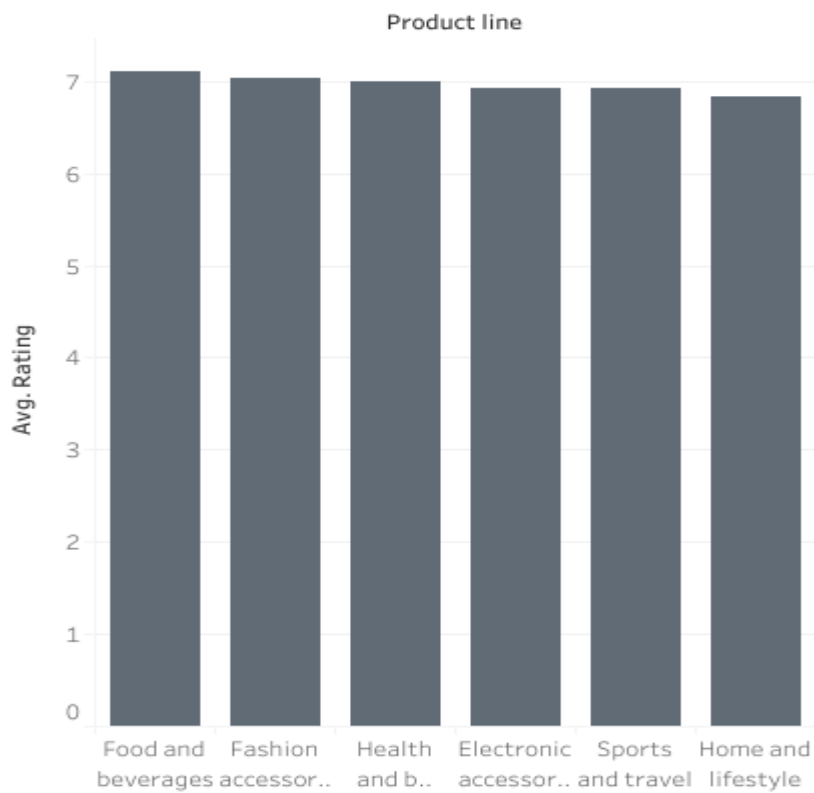Food and beverages generate greatest gross income.

- Profit by customer type:



Customer who are members will contribute to gross income more than normal.

- Rating for product line:



Similar to profit, Food and beverages continue to have the highest rating with average rating = 7.11.

# CONCLUSIONS

- Branch A is more popular than B, C

- Customers use ewallet and cash for payment more than credit cart

- In three types of payment, Credit card is used for member payment more than normal

- Women go to the supermarket to buy a lot of Fashion accessories, while men mainly buy Health and Beauty products

- Saturday is the most profit day of supermarket

- Female spend more money for shopping than male

- Food and beverages is the most profitable product-line for supermarket and has the highest rating

# RECOMMENDATIONS

- Create loyalty programs:

    - Discount: discount on first purchase when become member

    - Accumulate points: points will be added after each purchase. The larger the order, the higher the point. Depend on points obtained, customer will have different offers.

- Celebrate women's day: on these day, supermarket can offer discount on fashion accessories, food and beverages products for women

- Offer more discounts and promotions on weekend, especially on Saturday

- Display Food and beverages, Fashion accessories where they are most visible such as: as near the entrance, near the cashier,…