

# **Data Analysis Report on the Supermarket Sales**

## **Overview**

Supermarkets are expanding in most populated cities, and market competition is fierce. The main focus of every business is to make profit. Using the existing data of the supermarket, we will analyse and provide information on which aspects bring more sales to the supermarket and how the rate of sales can be increased in future.

## **Data**

The dataset is one of the historical sales of a supermarket company that was recorded in three different branches over a three-month period.

### **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 a member card and Normal for those without a 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)

## **Results**

Different data analysis methods were performed on the data in order to come up with a meaningful insight, which are stated below:

Data cleaning and preparation: The data was cleaned and prepared in order to use at the model building stage. Certain data cleaning models were done such as outlier detection and removal, features conversion e.t.c

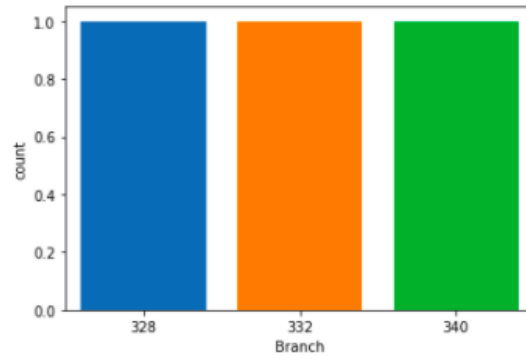
Data visualisation: Below are some plots of the data based on the data. They include

- Number of Branches: the supermarket has 3 branches namely A, B, and C. figure A shows the total number of branches that the supermarket has in all the cities.

```
In [80]: branches = market.Branch.value_counts()  
branches
```

```
Out[80]: A    340  
        B    332  
        C    328  
        Name: Branch, dtype: int64
```

```
In [81]: sns.countplot(data=market,x=branches, saturation=True);
```



as we see, 3 branches are approximately equal

Figure A

- Rating based on branch

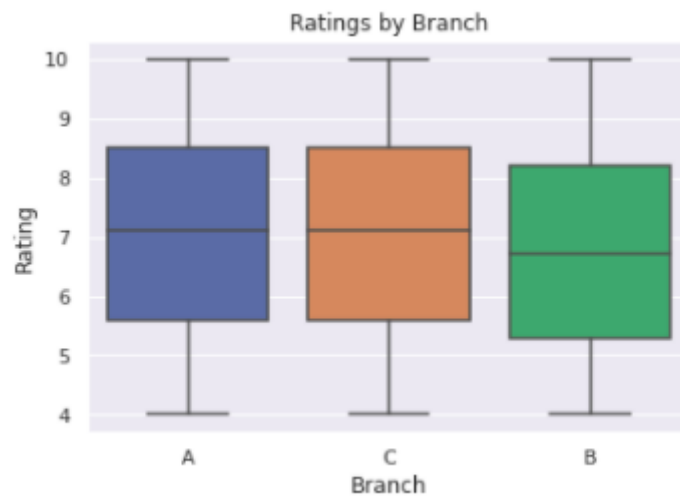


Figure B

- Customer\_type based on Gender: the supermarket has 2 types of customers patronising them. They are people with membership and

normal people. Figure C shows the number of customer types based on their gender.

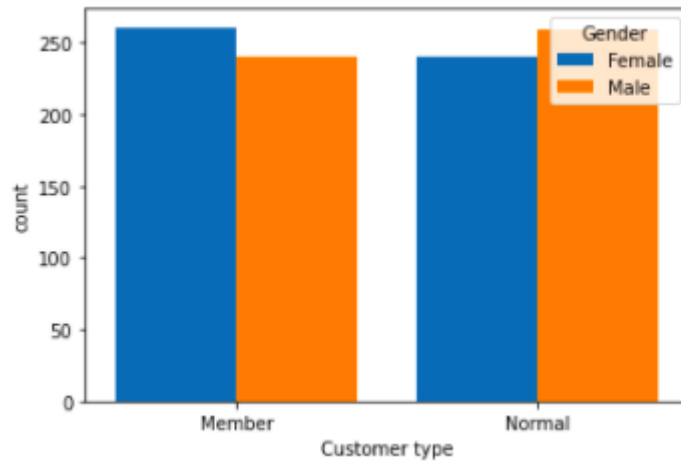


Figure C

- Product Sales per Hour: Sales by the hour in the company Most of the item were sold around 14:00 hrs local time



Figure D

- Top sales product line: The image below shows the top product line item type sold in the given dataset. Fashion Accessories is the highest while Health and beauty is the lowest

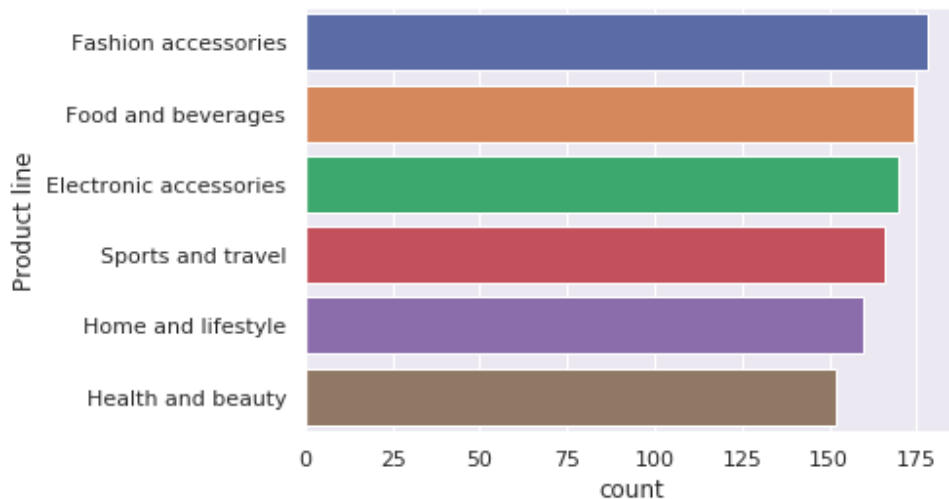


Figure E

- Product line sales: From the below visual, Health and Beauty, Electronic accessories, Home and lifestyle, Sports and travel have a better average quantity sales than food and beverages as well as Fashion accessories.

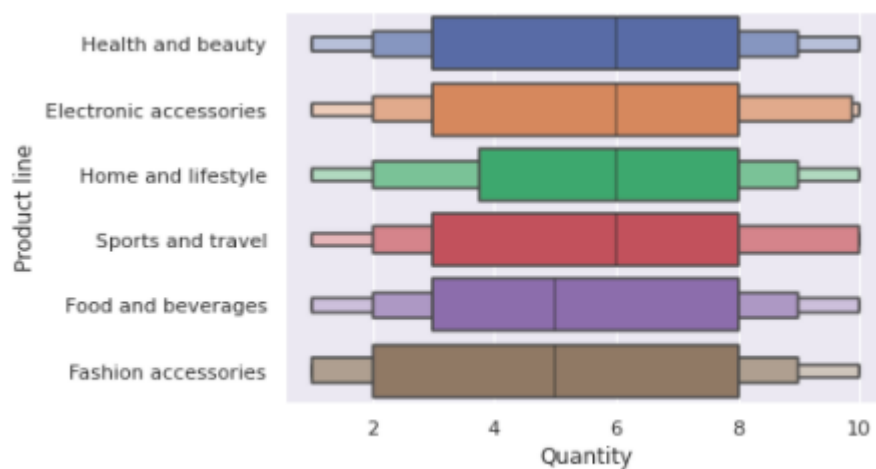


Figure F

- Rating per Product: Food and Beverages have the highest average rating while sports and travel the lowest

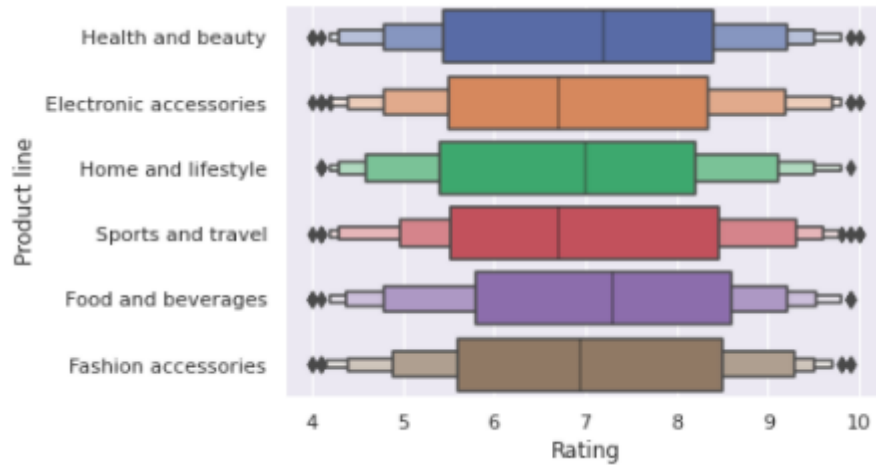
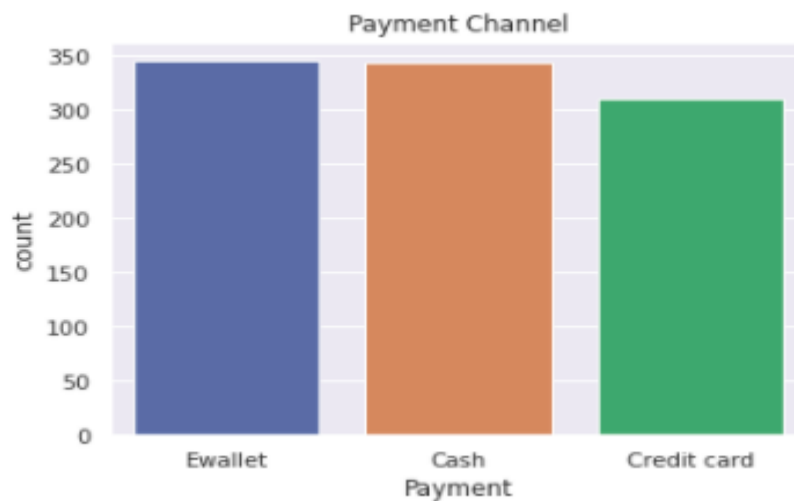
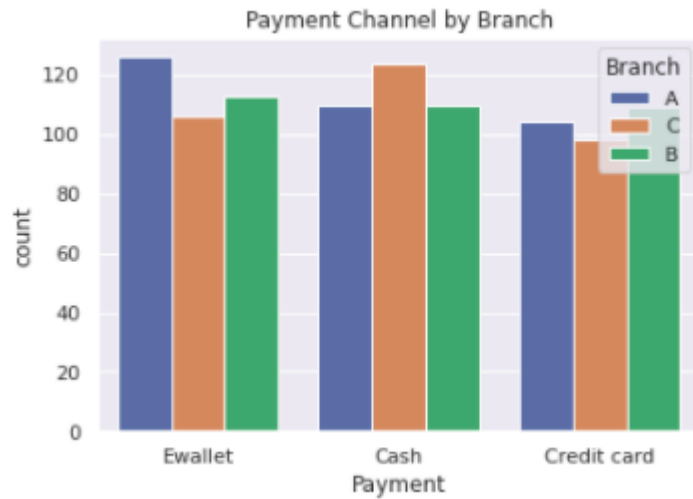


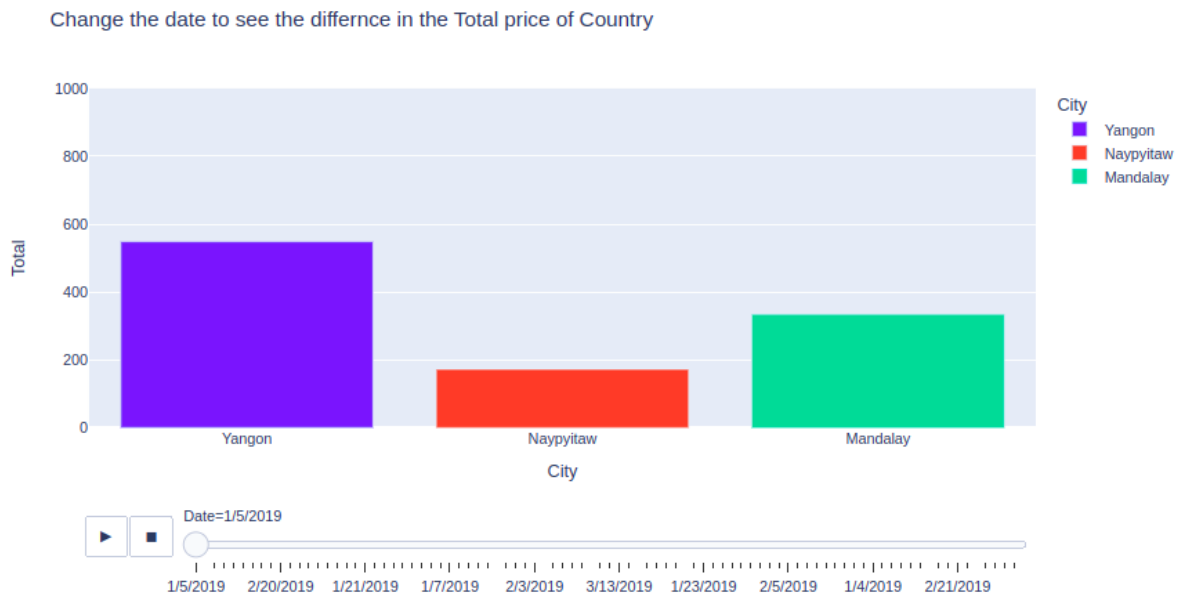
Figure G

- Payment channel i.e how customers make payment in general and also across the three branches: Most of the customers pay through the Ewallet and Cash Payment while under 40 percent of them pay with their credit card.





- Total price in each city with changes in date:



## Conclusion

Various factors contributed to the amount of sales made in the supermarket across all the branches. Further in this study a machine learning model will be built to determine the quantity of sales in future in the supermarket.