**COMPANY XYZ PERFORMANCE REPORT**

**OVERVIEW**

Company XYZ owns a supermarket chain across the country. Each major branch located in 3 cities across the country recorded sales information for 3 months.

The data folder contains datasets from three different branches; Lagos, Abuja and Port Harcourt. Each data file from the branches contains the same attribute information about the branch, location, customer type, gender, products line, price, tax and rating.

Transformations and column manipulations were done using Python pandas library and our visualization was done using seaborn visualization library. Time and date were converted to datetime using pd.to\_datetime() for easy manipulation. These were some of the observations on the datasets:

Rows/Observations: There are 1000 rows/observations from different transactions with unique Invoice ID

Missing Data: There was no missing data as regards to all observations in the dataset. It was checked by isnull() and sum() method

Task:

Provide a detailed analysis of the performance of the supermarket based on relevant information. Emphasis should be made to the company to understand sales trends and determine growth. Also understand the traffic of payment channel in each branch and city.

# PROCEDURES FOR DATA PREPARATION

In response of the business task and company expectations. The time and date information data type were changed and some new columns were created.

* Time and Date: Time and date was converted to the correct data type to

be able to extract and perform necessary analysis

*Time = pd.to\_datetime(df['Time']) Date = pd.to\_datetime(df['Date'])*

These are the new columns and their designations:

* Day: This is the information about the particular day of the week in which the transactions was made. It was extracted from the date information denoted as

Day = Df[‘Date’].dt.day

* Month: This is the information about the month in which the transactions was made. It was extracted from the date information denoted as

Month = Df[‘Date’].dt.month

* Year: This is the information about the year in which the transactions was made. It was extracted from the date information denoted as

Year = Df[‘Date’].dt.year

* Hour: This is the information about the particular hour of the week in which the transactions was made. It was extracted from the time information denoted as

Hour = Df[‘Time’].dt.hour

* Rating Grade: This is based on whether the customer is satisfied or not obtained from the rating column. Any rating that is less than 5 is categorized to not satisfy while above 5 are satisfied.

# DATA EXPLORATION AND VISUALIZATION

* The general outlook of the performance of the business based on relevant metrics.

**Approach**

General information/overview about the data were extracted to understand and reveal some hidden **insights**

**Observations:**

* There were 1000 recorded observations from 3 different cites Lagos, Abuja, Port Harcourt
* # 19931.93, #5457.52, #109150.54, #114608.06 were observed for the average unit price of product, tax, cost of goods sold and total price respectively.
* There were 3 branches in 3 different cites Lagos, Abuja, Port Harcourt
* Average rating of 6.94 was observes in the supermarkets across all the cities
* There are total records 501 and 499 transactions from female and male respectively, also total records of 501 and 499 transactions from member and normal customer type respectively
* Numbers of products sold in each product line is as observed below
  + Fashion accessories 178
  + Food and beverages 174
  + Electronic accessories 170
  + Sports and travel 166
  + Home and lifestyle 160
  + Health and beauty 152
* Epay is the most used payment channel with 345 transaction payment, followed by cash and card with 344 and 311 records respectively.
* All the transactions were observed to be within 3 months in the year 2019

# Data Exploration and visualization

* Determine the branch with the highest sales record

**Approach**:

Total sales from the Total column was calculated and grouped based on branches of the supermarkets

**Observation**

Branch C has the highest sales

* Determine the numbers of transaction in each branch

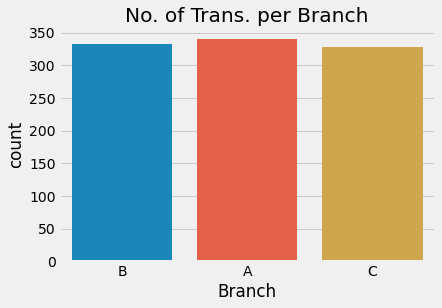
**Approach**

Number of transactions in each branches were counted by the use of value counts function in pandas library

A 340

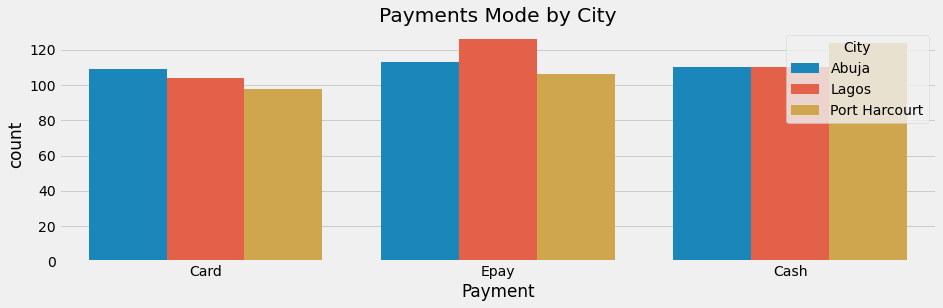
B 332

C 328

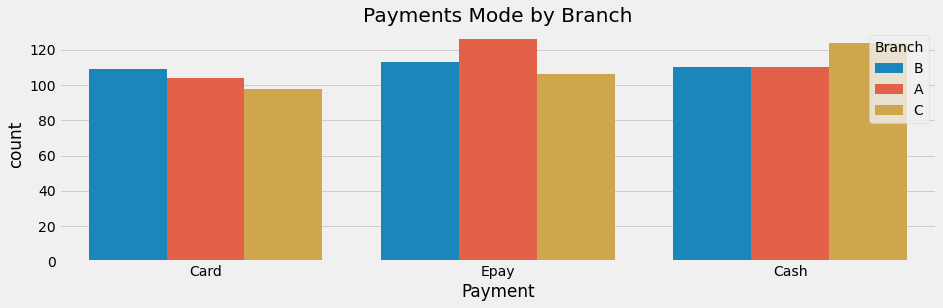


On comparing the visualization above, more transactions is carried out in Branch A

* Explore the payment channel in cities and branches of supermarket



Epay is mostly used in Lagos and Abuja while cash method is highly adopted in Port Harcourt



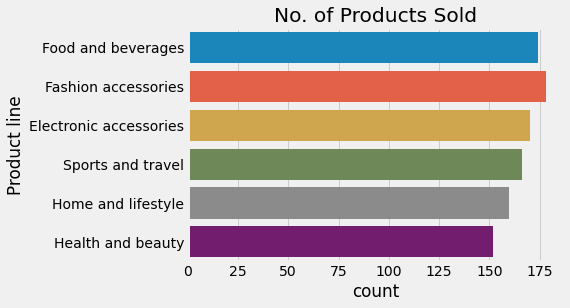
Card is mostly used in branch B, epay in branch A

* Determine the highest & lowest sold product line

**Approach**

Number of products sold in each product line were counted by the use of value counts functions in pandas library

Observations

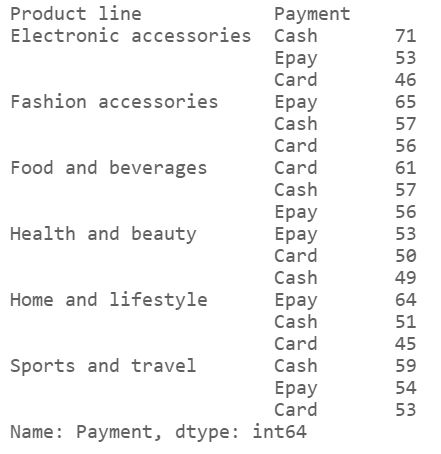


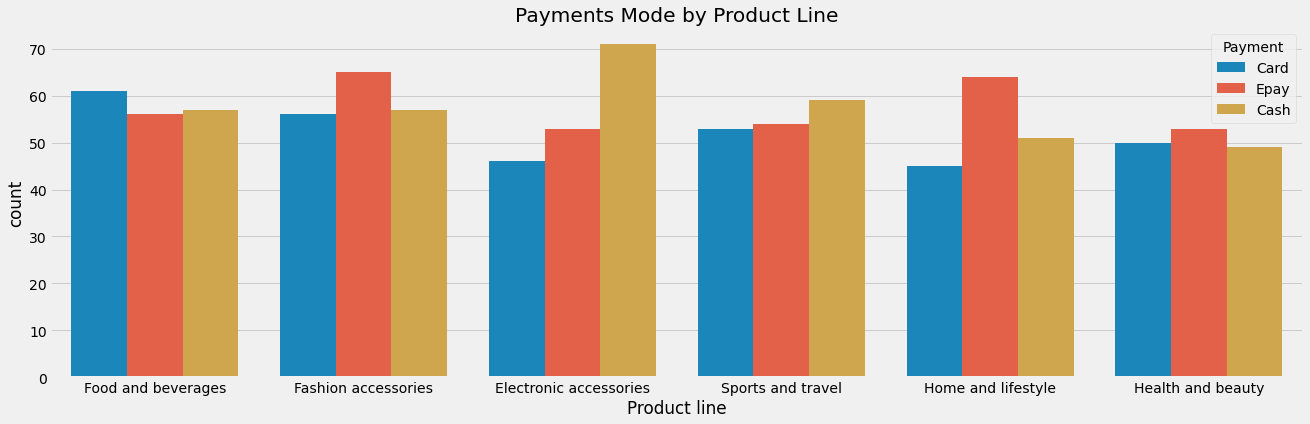
Sports and travel products are the highest products sold while Health and beauty products are the least

* Determine the Payment channel used by most customer to pay for each product line

**Approach**

Payment column was grouped based on product line available in the supermarkets



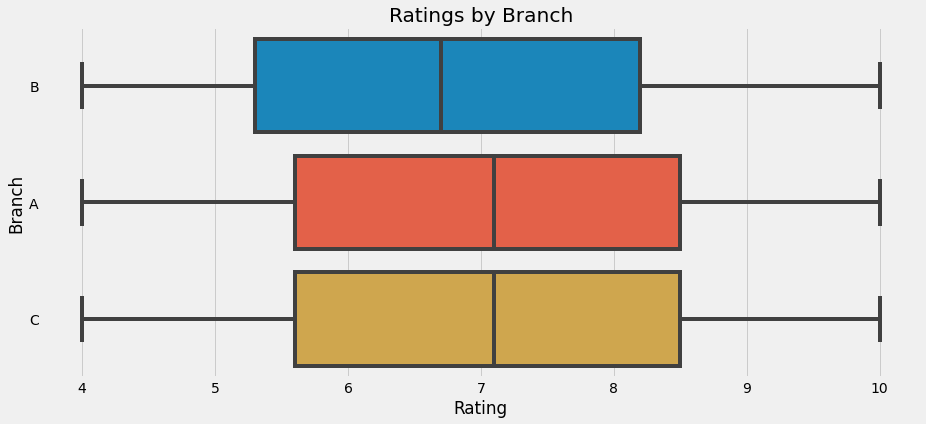


As observed from the above diagrams, Card, Epay, Cash, Cash, Epay, Epay are the most payment channel used in Food and beverages, Fashion accessories, Electronic accessories, Sports and travel, Home & lifestyle, Health and beauty respectively

* Determine the branch with the lowest rating

**Approach**

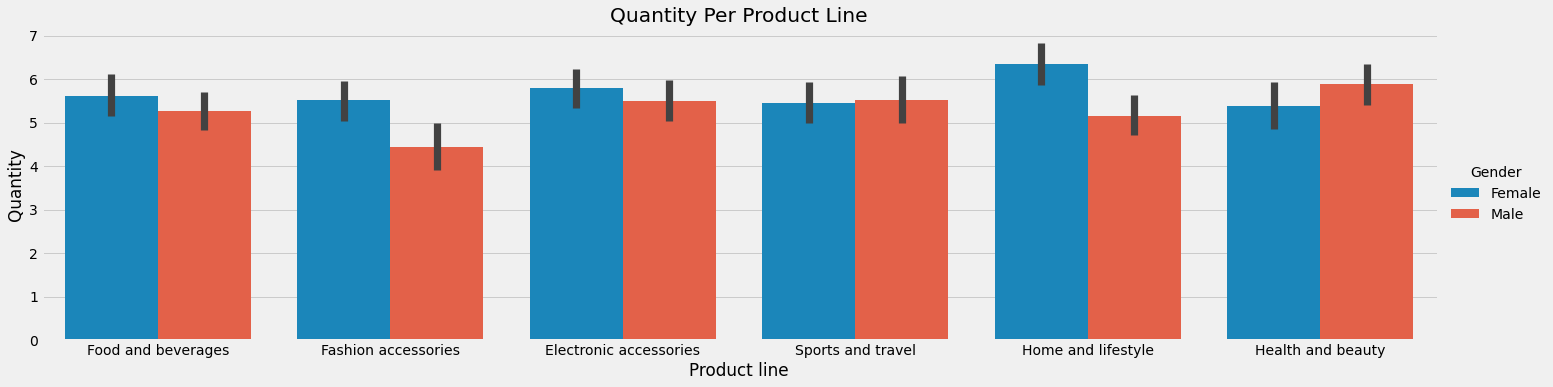
Rating was grouped based onthe 3 available branches of supermarket



Observations

Of all the branches, Branch B had the lowest rating of 4, averagely Branch B is still the lowest when compared to other branch

* Compare the gender type patronage on kind of products being purchased at the supermarket

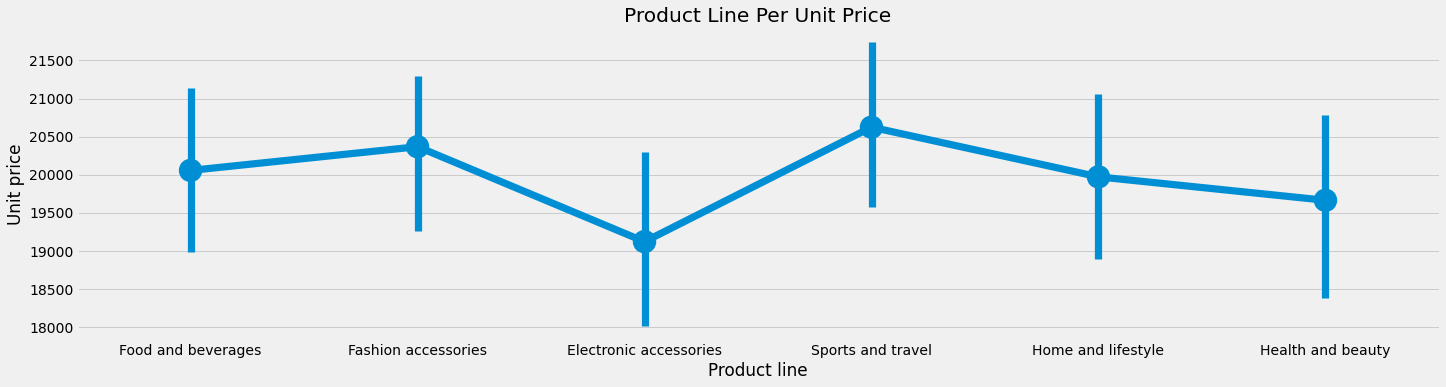


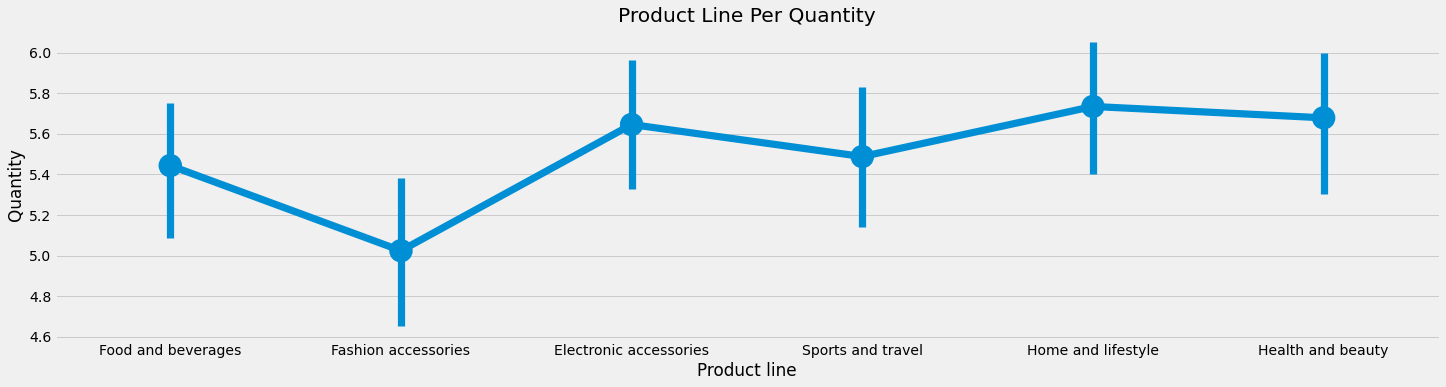
Observations

- In most product lines, females bought more items than males except for Health & Beauty and Sports & Travel

- There is wide range in quantities bought by females to males in fashion accessories and home & lifestyles than other product lines

* Explore is the interaction of Unit price on the Quantity of goods purchased





Observations

- The price of sport and travel items are the highest while electronic accessories items are the least

- Home and Lifestyle products have the highest quantities sold

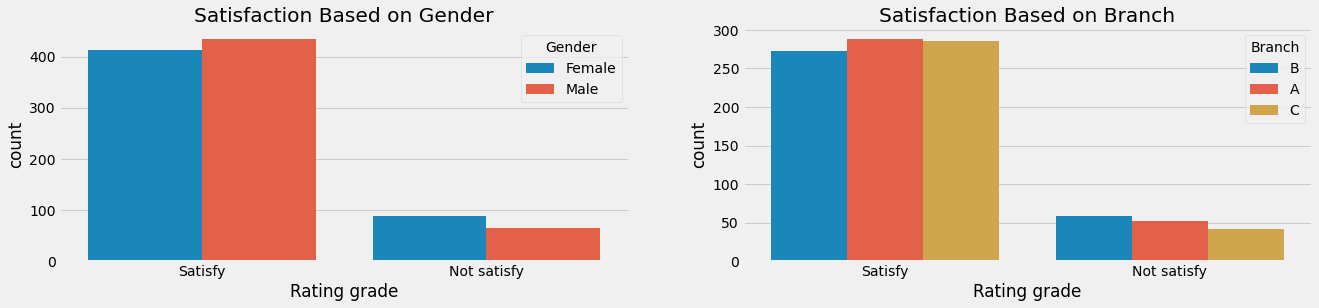
- Fashion accessories products are the lowest quantities sold

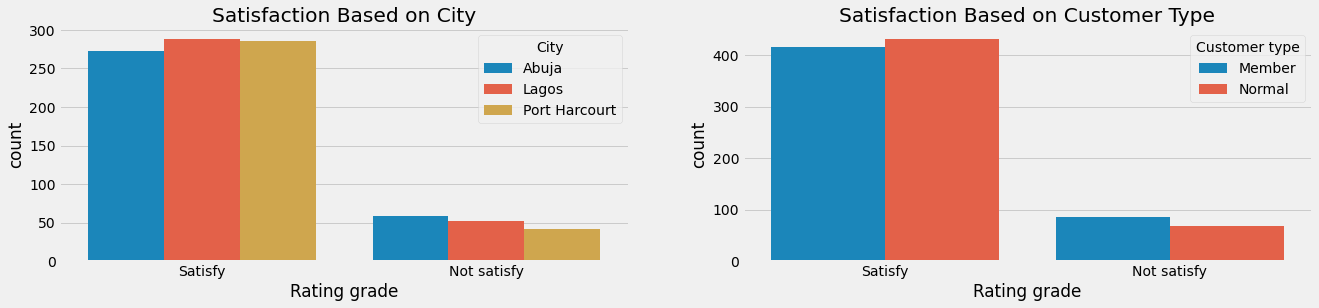
* Give a detailed exploration of satisfaction rate

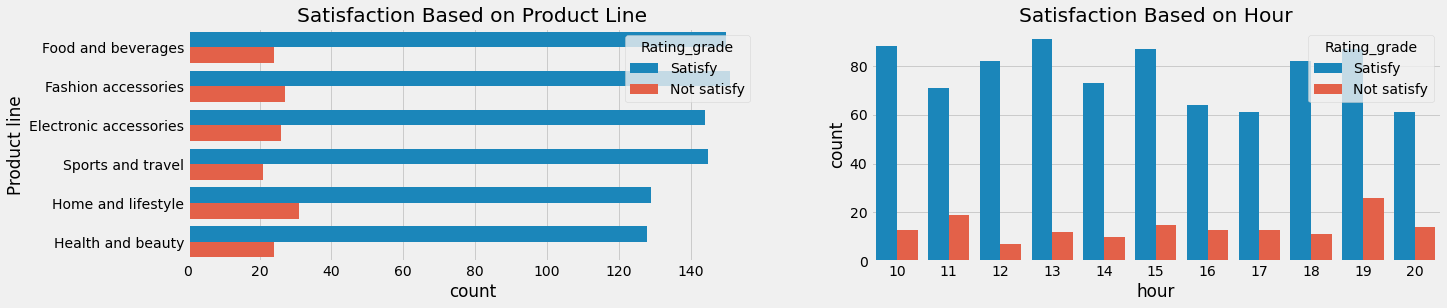
Approach

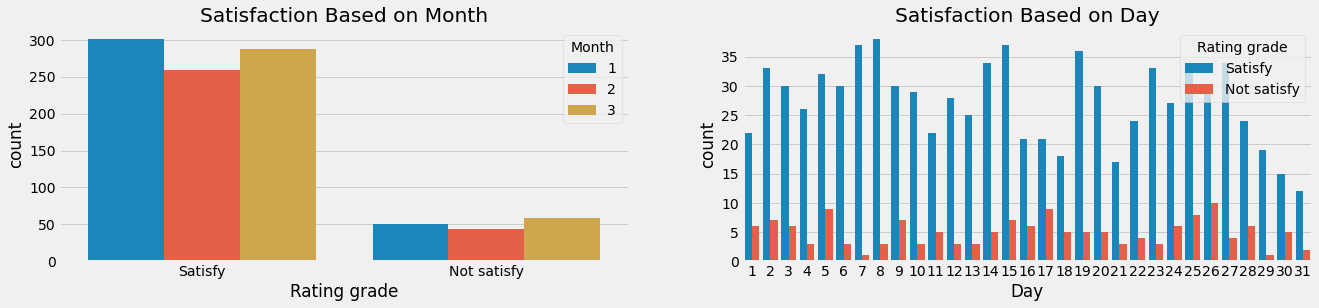
Satisfaction rate is grouped and examined against various metrics available

Observations:









**CONCLUSIONS AND RECOMMENDATIONS**

* Branch C with the highest sales amount had the lowest numbers of transaction. As compared to branch A with highest number of transaction which implies that most costly goods e.g sport and travel products are being sold in C than A. Therefore the type of product sold most in each branch should be kept supplying in abundance and available for sale and branch A should share the strategies to that earn those more customers to other branches.
* Cash payment channel services should be made available and improved in Port Harcourt city, electronic accessories section and Branch C of the supermarkets.
* Fashion accessories, food & beverages and electronic accessories should be more focused in all product lines.
* The entire team in charge of branch B should be made aware of their satisfaction rate status in comparison to other branches and they should look up to tips from them to enhance their status.
* On Health & Beauty and Sports & Travel product line, males products should be more focused and fashion accessories and home & lifestyles for females
* Generally, 85% customers are satisfied across all metrics, detailed information of days with high satisfaction could be further study to improve the performance of days with low satisfaction to increase the overall performance.