USTACKY

USTACKY MICRO DEGREE COURSE

DATA SCIENCE: PANDAS ANALYTICS

Executive Summary - Pandas Data Analytics

Company XYZ management

NAME: ADEBANJO ESTHER CHIZOBA

(Student)

4th, February, 2023

Contents

Introduction

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, to help the company understand sales trends and determine its growth, as the rise of supermarket competition is seen to increase.

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.

Description

Invoice ID: Customer Identification Number

Branch: Supermarket Branches across the country (A, B, C)

A - Lagos Branch,   B - Abuja Branch,   C - Port Harcourt Branch

City: Supermarket Location

Customer Type: Type of customers, Members - Returning customer with a membership card, Normal - Customer without membership (could be returning, first-time or walk-in customer)

Gender: Customer Gender Information

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

Unit Price: The price of each product in Naira

Quantity: Number of products purchased by the customer

Tax: 5% tax fee for customers buying

Total: Total price including tax

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

Time: Purchase time (Supermarket Hours - 10 am to 9 pm)

Payment: Payment used by the customer for the purchase (3 methods are available – Cash, Card, and Epay)

COGS: Cost of goods sold

Gross margin percentage: Gross margin percentage

Gross income: Gross income

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

Requirements

Step 1 - Loading Datasets

Step 2 - Data Exploration

Step 3 - Dealing with DateTime Features

Step 4 - Unique Values in Columns

Step 5 - Aggregation with GroupBy

Step 6 - Data Visualization

Analysis

Analysis was carried out to determine:

- The statistical summary of the combined dataset.
- Various datetime features
- The unique value of some columns
- The city with the highest gross income
- The percentage of normal customers Vs member customers
- The percentage of female Vs male
- The gender that spent the highest number time
- The branch with the highest sales record using countplot,
- The highest & lowest sold product line, using Countplot.
- The payment channel used by customers to pay for each product line.
- The payment channel for each branch.
- The branch with the lowest rating.

Further visualization analysis was carried out to:

- Generate visualization for the "product line" per gender.
- Generate visualization for the "product line" per "Total" sale.
- Plot Product line per unit price, and Product line per Quantity.

Insight

- There are 3 different datasets (csv files): Dataset for Abuja_Branch, Lagos Branch, and Port Harcourt Branch
  - Lagos branch (A) consists of 340 rows and 17 columns
  - Abuja branch (B) consists of 332 rows and 17 columns
  - Port Harcourt branch (C) consists of 328 rows and 17 columns
- The datasets were combined to generate a single dataframe (dataset) which consists of 1000 rows and 17 columns
- The statistical summary shows the Standard Deviation (STD) of the unit price and gross income to be 9538.066205 and 4215.177173 respectively and also, the mean value of the unit price and the gross income is 20041.966800 and 5536.572840 respectively.
- The dataset contains no null values
- The dataset contains records of 501 female customers and 499 male customers
- Port-Harcourt (C) records the highest revenue(gross income) while Lagos (A) follows next and then Abuja (B) as shown in Figure 1
- There are more member customers than normal customers which correspond to females and males in the gender count as shown in Figure 6 and Figure 7 respectively.
- The female spent more time in the City of Port Harcourt as shown in Figure 8
- 'A' has the highest no of customers/sales followed by 'B' and the least is 'C' as can be seen in Figure 2.
- The most used payment method, in general, is EBAY and the least is Card
- The most sold product line in all 3 branches is 'food and beverages' in 'C', next is 'Home and lifestyle' in 'A' and the 3rd most sold product line is 'fashion and accessories' in 'C'. However, in 'B', the most sold is 'Travel and sport' as can be seen in figure 3.
- The most used payment method for 'Product line' is Cash
- The branch with the lowest rating is "B" while 'C' has the highest as can be seen in Figure 4.

- It is observed that the most frequent customers are of the female gender and the most bought product is the 'Home and lifestyle' as can be seen in Figure 5.
- It is also observed that the most expensive 'product line' is the 'food and beverages' and it is mainly bought by the female gender.
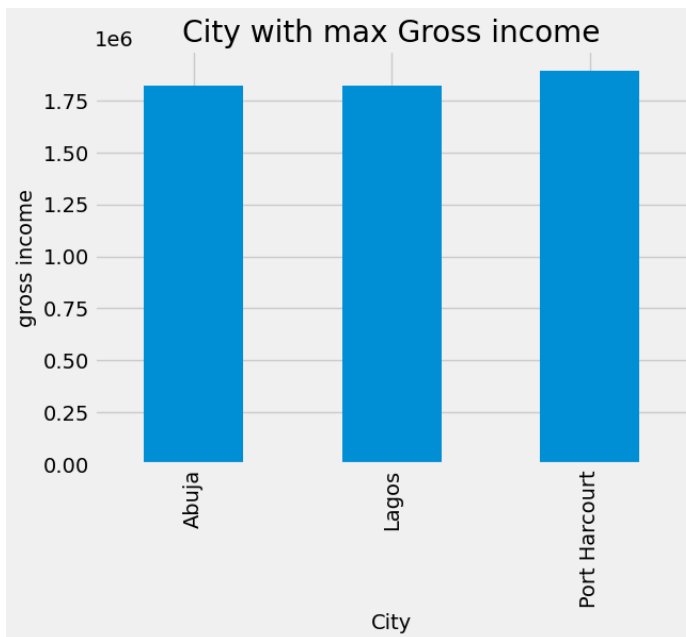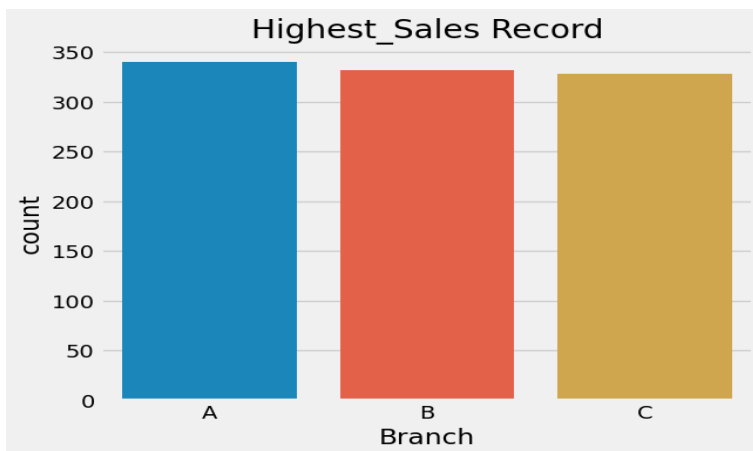

Figure 1: City with maximum gross income
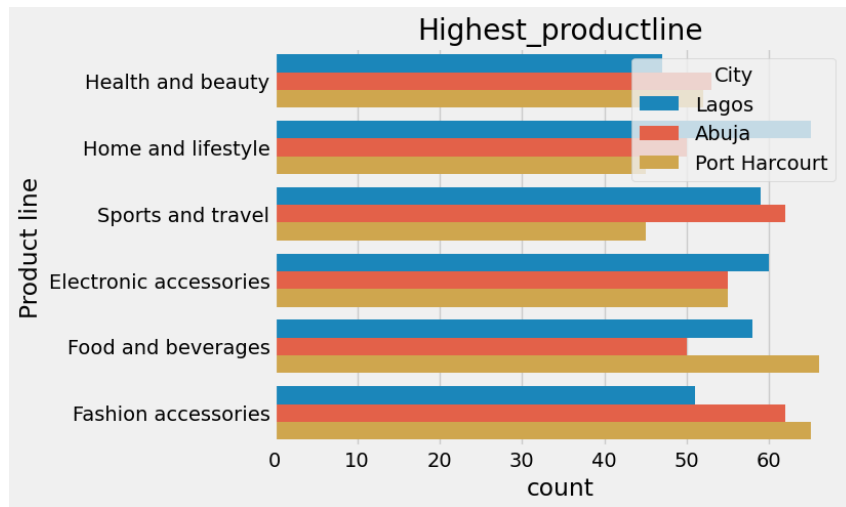

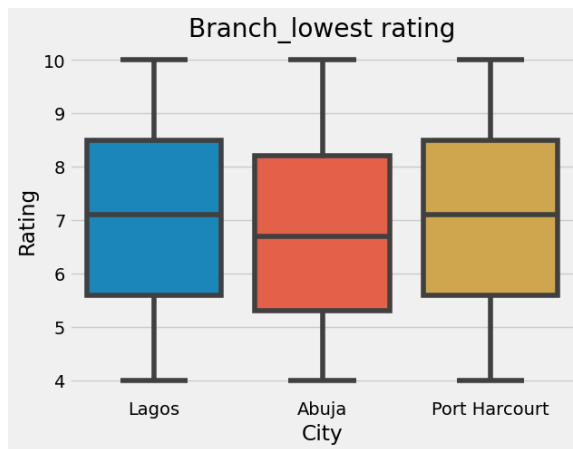*Figure 2: Branch with the highest sales*
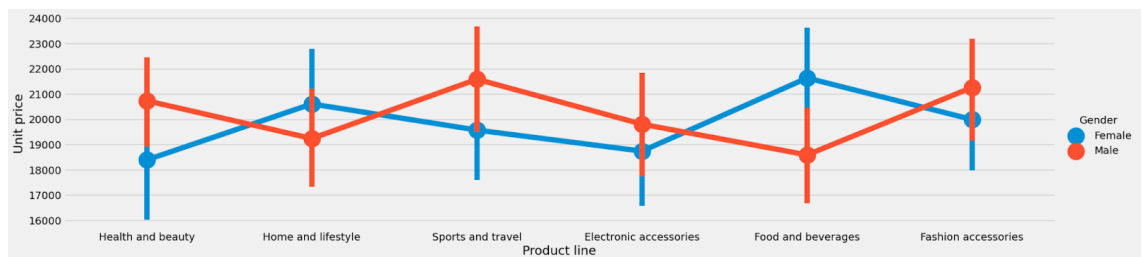
Figure 3: Most sold product line

Figure 4: Ratings

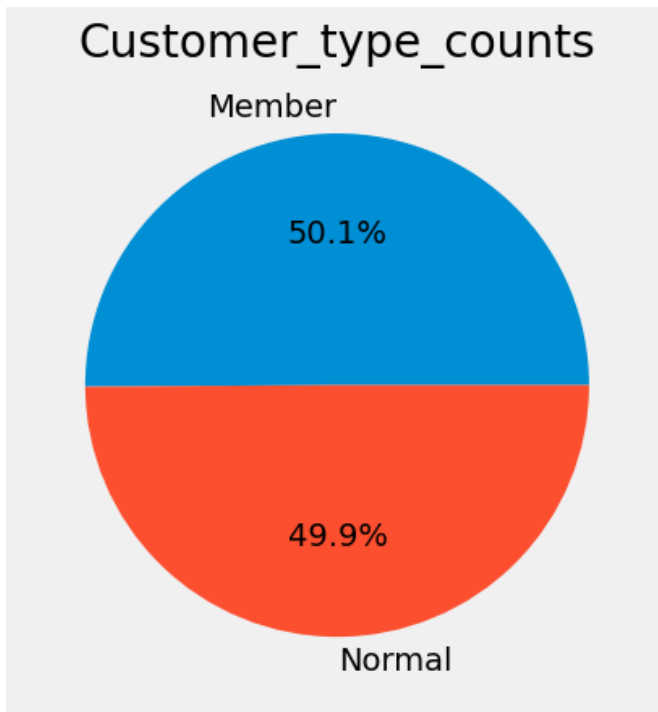*Figure 5: Most sold product line and most frequent gender*
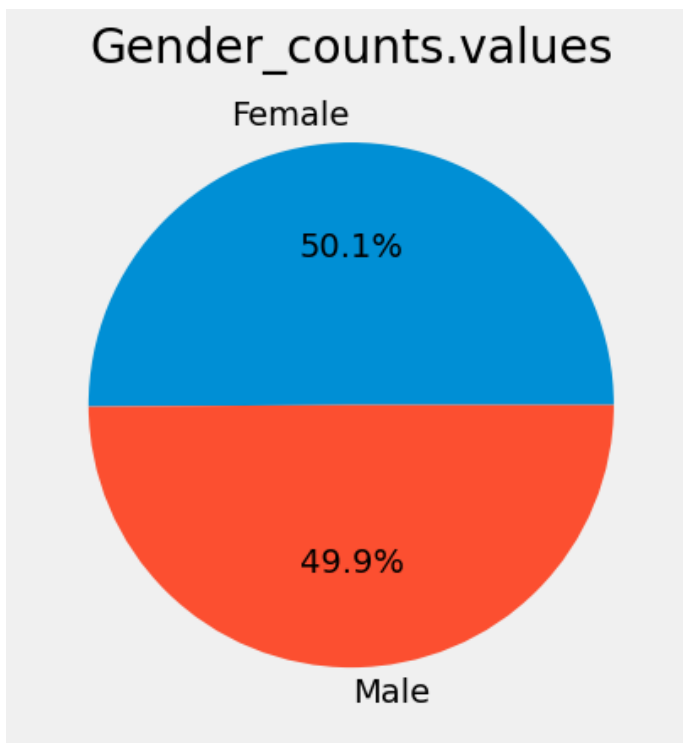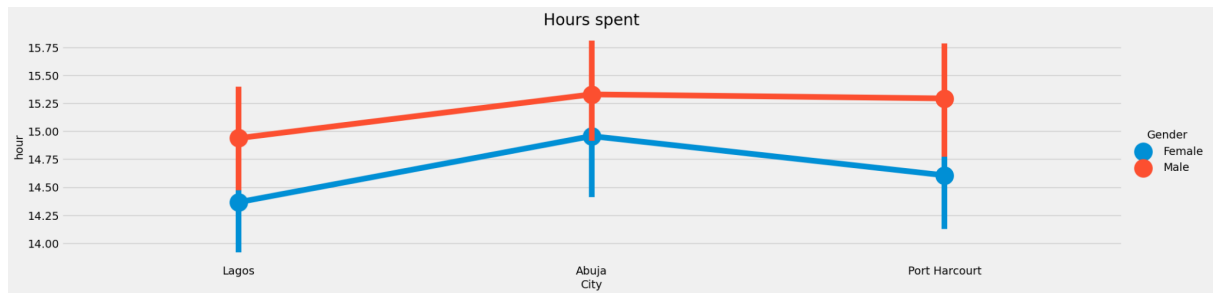
Figure 6: Customer type

Figure 7: Gender counts

Figure 8: Hours spent in city and by gender

Recommendations

Carry out further analysis on the types of customers (members, returning, and first-time) to ensure that the supermarket is making progress monthly.