IST 707 AML: HW 2 Harshit Joshi

Supermarket Sales Analysis:

HW2: Tell a Data Story

Growth of supermarkets in most populated cities is increasing and market competition is high. The following dataset contains historical sales of a supermarket across 3 different branches for a period of 3 months.

Assume you are working as a data scientist for the supermarket. **Tell the story** of this data by using appropriate data exploration and transformation techniques.

You are required to provide insights into sales data across branches. For example, what’s the gross income distribution over different branches? Furthermore, are there gender differences in each branch?

Generally, you need to explore all highlighted variables below and tell the story (or stories) in this data. Don’t forget relationships between or combinations of variables.

**Find the story, tell it visually and, above all, truthfully.**

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)

Source: <https://www.kaggle.com/aungpyaeap/supermarket-sales?select=supermarket_sales+-+Sheet1.csv>

Which Store has highest Gross Income:

Just by looking at the barplot we can see which branch has highest gross income of all, It is C. We can look at what Store C is doing in terms of targeting customer. We can also see at what location of city is store located. We can compare those aspects with aspects of A and B and find out what things our store brand must consider when opening a new store in a city.

Chart, bar chart

Description automatically generated

Graphical user interface, application

Description automatically generated

Store C has gross income of $5265

Store A has gross income of $5057

Store B also has gross income of $5057

The proportion of members and normal customers is equal. This means that our store is doing a good job of retaining employees as more than half of customers are member of the stores.Graphical user interface, text, application, email

Description automatically generated

--

The following histogram suggests that more transactions have been done in the morning than the whole day. It shows number of hours with their level of frequency based on transaction time. the hours of 10-11AM are popular among customers to visit the stores in a day. So managers must make sure that there is no employee shortage during these hours especially at the checker counters cause its crowded. For precaution, we must have standby employees during mornings in case of no shows.

Chart, histogram

Description automatically generated

--

The following table shows gross income of product lines across all 3 branches:Graphical user interface, text, application, email

Description automatically generated

This shows that Food and beverages have the highest gross income out of all product lines. We can use this to make sure that we have enough stock the day before to serve customers.

--

The following table suggests that Food and Beverages have highest rating compared to all product lines. But Home and lifestyle has rating of 6.83 which is lowest. Executives of the company can focus more on Home lifestyle products as this is the lowest rated product of all.

Graphical user interface, application, Teams

Description automatically generated

--

Text

Description automatically generated

The above table shows correlation of genders of our customers with their membership status. Female members have 261 transactions in total as compared to 240 female normal customers (members higher than normal customers). Also Male members have 240 transactions in total as compared to 259 Male normal customers (Members Lower than normal customers).

This can be used to introspect our brands’ marketing strategy that targets every genders. We can certainly focus our male customers and use appropriate strategy to convert more male customers to become a member. One thing can be done is by offering free Men Care products if they sign up to become a member.

--

This graph can be used to infer that:

Cash is most popular method of payment among C Branch shoppers (Pink color)

Credit card is more popular form of payment in B branch (Green color)

E-Wallet is most widely used in A branch (Blue color)

Chart, bar chart

Description automatically generated

--

We can see that A has more quantity of product stored as compared to B and C. But in C, female buy more in bulk as compared to females of other branches. So we can focus on introducing more supply of feminine product in A and B, or We can learn from branch C about what are they doing to attract more female customers.Chart, bar chart

Description automatically generated