SUPERMARKET STORE SALES DATASET

Description:

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. Predictive data analytics methods are easy to apply with this dataset.

Attribute information:

Invoice id: Computer generated sales slip invoice identification number

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

City: Location of supermarkets

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

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

Kaggle Link:

https://www.kaggle.com/aungpyaeap/supermarket-sales

(Don’t use data from this link. Only for Reference).

Problem Statements:

1. Describe the Data in your own words.
2. Check Datatypes of each column and Bring Date time together in one column.
3. Find the Null Values and Treat them Appropriately.
4. Count the No. of Branches and Cities.
5. Create a pie chart of gender.
6. Calculate Average Rating for each product line.
7. No. of products in each category.
8. Total Amount collected in each product line
9. Find out highest percentage of payment method
10. Find out the category with highest Rating.
11. State the conclusion of the project based on the anal.

Note – Creating Related Charts for each problem statement is a plus point.