**Sales of a Supermarket**

**About Dataset**

**Context**

The growth of supermarkets in most populated cities is 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 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 E-wallet).

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).

**Tasks for portfolio project**

1. ***Data Exploration and Cleaning:***

Explore the structure of the dataset and identify the number of rows and columns in each table.

Check for any missing or duplicate values in the dataset and handle them appropriately.

Ensure data types are appropriate for each column.

1. ***Basic Analysis:***

Calculate the total sales revenue for each branch.

Determine the number of transactions made by each customer type (Member vs. Normal).

1. ***Customer Insights:***

Segment customers based on their purchasing behavior (e.g., frequency of visits, total spending).

Calculate the average rating for each customer segment.

Identify the top-paying customers.

1. ***Product Analysis:***

Determine the best-selling product line.

Calculate the total quantity sold and revenue generated for each product line.

Identify the products with the highest and lowest gross margin percentages.

1. ***Sales Trends:***

Analyze sales trends over time (monthly or quarterly) for each branch.

Visualize sales performance using line charts or bar graphs.

1. ***Payment Analysis:***

Determine the most common payment method used by customers.

Calculate the average transaction amount for each payment method.

1. ***Market Competition:***

Compare sales performance between different branches.

Analyze the market share of the supermarket company in each city.

1. ***Predictive Analytics (Optional):***

Use historical sales data to predict future sales for each branch or product line.

Build a model to predict customer satisfaction rating based on various factors such as product line, payment method, etc.

1. ***Data Visualization (Optional):***

Create visualizations to present key insights and trends using tools like Tableau, Power BI, or matplotlib in Python.

Provide explanations for any assumptions made during the analysis process.