**Super Market Sales Analytics**

**Introduction**

# A Supermarket is a self-service shop offering a wide variety of food, beverages and household products, organized into sections. The growth of super markets in most populated cities is increasing and market competitions are also high. The Super Market Sales Analytics is used to analyze the sales of different products in a super market located in different cities.

**Objective**

Follow the Data Science Methodology that we studied in Module 3. Analyze the data set, investigate and evaluate the result and predict the overall performance.

**Dataset**

The dataset is *Supermarket\_Sales\_Dataset.csv*. It has been uploaded to canvas.

This dataset contains various attributes of supermarket like Invoice Id, Branch, City, Customer Type, Gender, Product Type, Unit Price, Quantity, Tax, Selling Price, Date, Time, Payment Type, Cost Price, Gross Income, Rating.

**Tasks in this assignment**

1. Write a Data Science Proposal for achieving the objective mentioned.

2. Perform exploratory analysis on the data and describe your understanding of the data.

3. Perform data wrangling / pre-processing.

a. E.g., missing data, normalization, discretization, etc.,

4. Apply any two feature selection engineering techniques

5. Compare the two selected feature engineering techniques.

6. Plot top 5, 6, and 8 features.

7. Provide a high-level description of Machine Learning models – association rules and random forest to predict.

8. Compare the performance of the two classifiers – association rules and random forest.

9. Present the conclusions/results in the format shared.

**Expected Submissions**

Two files are expected as the assignment submission.

1. The summary of the work in the template provided. (you may fill only the boxes relevant to this problem statement)

2. The executed ipynb file with clear subdivision of the codes and brief description of the purpose of respective code. All the executed tables or graphs and results should be present in the ipynb file. The ipynb file maybe submitted as a single .pdf file