**Project Title: Supermarket Campaign Analysis**

**Dataset Link:** [**https://www.kaggle.com/datasets/ahsan81/superstore-marketing-campaign-dataset**](https://www.kaggle.com/datasets/ahsan81/superstore-marketing-campaign-dataset)

**Problem Statement:** All You Need Supermarket is preparing for a year-end sale and plans to introduce a new offer exclusively for existing customers. The offer, called gold membership, provides a 20% discount on all purchases for a price of $499, which is usually sold at $999. To promote the offer, the supermarket plans to conduct a phone call campaign targeting its existing customers. The management believes that building a predictive model to identify potential customers who are more likely to purchase the offer can help reduce campaign costs. They have provided data gathered from last year's campaign to build the predictive model. This involves analyzing the provided data and building a predictive model using machine learning techniques. Additionally, the model needs to be production-ready using pipelines. The end goal is to provide the supermarket with insights that can help them identify potential customers who are likely to purchase the gold membership offer, reducing campaign costs and increasing sales.

**Aim:** The objective of this project is to identify the factors that influence a customer's response to the gold membership offer and predict the probability of a customer giving a positive response.

**Research Questions and Milestones:**

1. Analyze the factors that would influence customers response
2. Do customers have any complaints for the company?
3. Can the company target customers who buy premium products?
4. Can the company provide offers to any specific category of customers?
5. Building model to predict the probability of a customer giving positive response

## **Flow:**

## **Part1:** Data Ingestion and Preprocessing - Loading the data and checking the common statistics

## **Part 2:** Data Preprocessing - Cleaning data, removing noise, outliers, duplicates and null values.

## **Part 3:** Feature Imputation- Data transformation and Making data ready for modeling

## **Part 4:** Model Training- Training different ML models on the processed data like RF, DT etc.

## **Part 5**: Model Evaluation and Predicting - Evaluating models using different metrics and generating predictions. **Tools and Technologies:** Pyspark and Google Colab.