**Dataset: Supermart Grocery Sales - Retail Analytics Dataset**

**Source:** Kaggle - **Supermart Grocery Sales Dataset** <https://www.kaggle.com/datasets/mohamedharris/supermart-grocery-sales-retail-analytics-dataset>

**Overview:**  
This dataset contains detailed sales data from a grocery retail store, making it ideal for analyzing sales performance, customer segmentation, and trend analysis.

**Key Features:**

* **Number of Rows:** Approximately 600
* **Number of Columns:** 11
* **Columns Include:**
  + Order ID: Unique identifier for each order
  + Customer Name: Name of the customer
  + Gender: Gender of the customer
  + Product Category: Category of the purchased product
  + Product: Specific product name
  + Quantity: Number of items purchased
  + Price: Price per item
  + Tax: Tax applied to the purchase
  + Total: Total amount spent
  + Date: Date of purchase
  + City: City where the purchase was made

**Project Title: Grocery Sales Performance and Customer Segmentation Analysis**

**Project Description:**

This project involves analyzing sales data from a **grocery retail business** to understand **sales performance, customer segmentation, and sales trends**. The dataset includes various attributes related to product categories, sales, discounts, profits, customer regions, and demographics. The goal is to gain insights into **customer purchasing behaviors**, **identify key trends**, and **segment customers to optimize sales strategies**.

The analysis will utilize statistical techniques and data visualization to uncover patterns that can help in better decision-making. This will provide valuable insights to improve marketing campaigns, optimize inventory, and enhance customer targeting.

**Problem Statement:**

The objective is to analyze the sales data from a grocery retail store to derive actionable insights for improving sales strategies and business performance. Specifically, we aim to:

1. Understand the performance of different product categories and sub-categories.
2. Identify customer segmentation based on purchase behaviors, regions, and other demographic factors.
3. Analyze the impact of discounts on sales and profit.
4. Examine sales trends and seasonality to forecast future sales.
5. Provide strategic recommendations for enhancing business performance.

**Problem Solving Approach**

To address the problem, the following approach will be used:

1. **Data Preprocessing**: Cleaning the dataset by handling missing values, converting data types, and removing any irrelevant information.
2. **Exploratory Data Analysis (EDA)**: Performing univariate, bivariate, and multivariate analysis to understand key relationships between features like category, sales, discounts, and profit.
3. **Customer Segmentation**: Grouping customers based on various criteria like region, sales, and discount behavior.
4. **Trend and Seasonality Analysis**: Identifying sales trends over time, along with any seasonal variations.
5. **Visualization**: Creating charts and graphs to clearly communicate findings.

**Dataset Insights:**

Here are five potential insights based on the dataset:

1. **Top-Performing Categories and Sub-Categories**: Identify which product categories and sub-categories generate the highest sales and profits.
2. **Regional Sales Differences**: Compare sales performance across different regions to highlight regional preferences and performance disparities.
3. **Discount Impact**: Analyze how discounts correlate with sales and profits, revealing if larger discounts lead to higher sales or if they erode profit margins.
4. **Sales Trends Over Time**: Analyze seasonal trends in sales to understand which months or periods are the highest for grocery sales.
5. **Customer Segmentation Based on Sales**: Segment customers into different groups based on their purchase volume, sales frequency, and product preferences, which can help tailor marketing strategies.

This approach will provide a comprehensive analysis to improve decision-making and optimize business operations.