**Anilkumar Palavelli**

**Introduction:**

Most towns with a lot of people have more and more stores, and market competition is also very high. The dataset is a record of one of the company's past sales, which has been kept in 3 different locations for 3 months. This information makes it easy to use predictive data analytics methods. This project provides a brief overview of the Supermarket sales of the one company, which has been kept in 3 different places for 3 months.

**Domain:**

Supermarket Sales

**Data File:**

****

**Data Source:**

[**https://www.kaggle.com/datasets/aungpyaeap/supermarket-sales**](https://www.kaggle.com/datasets/aungpyaeap/supermarket-sales)

**Number of Rows:** 1000

**Number of Columns**: 17

Here are the meanings of each column in the Supermarket Sales dataset:

* Invoice id: A computer-made sales slip with an invoice number
* Branch: A branch of a supercenter. There are three branches, marked A, B, and C.
* City: Where supercenters are
* Customer Type: Type of customer, which is kept track of by Members for customers with a member card and by Normal for customers without a member card.
* Gender: Type of customer by gender
* Product line: Electronic accessories, fashion accessories, food and drinks, health and beauty, home and living, sports and travel, and other general groups of items.
* Unit price: How much each item costs in $
* Quantity: number of items a customer bought
* Tax: 5% tax on customer purchases
* Total: Price total, including tax
* Date: Date of buying. Record is available from January 2019 to March 2019.
* Time: 10 a.m. to 9 p.m.
* Payment: There are three ways for a customer to pay for an item: cash, a credit card, or an e-wallet.
* COGS: Cost of goods sold
* Gross margin percent: Gross margin percent
* Gross income: Total money made.
* Rating: Customers rate their shopping experience generally on a range from 1 to 10.

**Visualization tools selected:**

I have used Excel and Tableau to visualize the charts for the goals prepared.

**Data Cleaning Strategies:**

The dataset which was found I the source has some unwanted data; I have removed that and renamed the columns for better understanding and have displayed the data in the sorted order of states.

**Cleaned Data Set:**