

# **Supermarket Sales Analysis**

## **Understanding Buying Patterns**

**Student Id:** AF0484416

**Student Name:** SANGAMITHRA B

**Course:** Data Analysis using Python[DAP]

**Project description:** Handling missing values (product category, discount applied), Data grouping (sales by product type), Basket analysis (frequent item combinations), Trend visualization (peak sales hours), Comparative analysis (weekend vs. weekday sales)

<b>S.NO</b>	<b>TABLE OF CONTENT</b>	<b>PAGE NO</b>
1.	INTRODUCTION	3
2.	OBJECTIVE	4
3.	DATA COLLECTION	5
4.	METHODOLOGY	6-7
5.	KEY FINDING & INSIGHTS	8
6.	CONCLUSION	9
7.	FUTURE GOALS	10

# **1. INTRODUCTION**

Supermarkets are an essential part of daily life, catering to the needs of a wide range of customers. Understanding customer buying behavior is crucial for supermarkets to improve their sales strategies, stock management, and customer satisfaction. This project focuses on analyzing supermarket sales data to identify trends and patterns that influence purchasing decisions. By studying various factors such as peak shopping hours, preferred product categories, customer demographics, and payment methods, this analysis aims to provide valuable insights for better business planning.

## **2. OBJECTIVE**

The primary objectives of this project are:

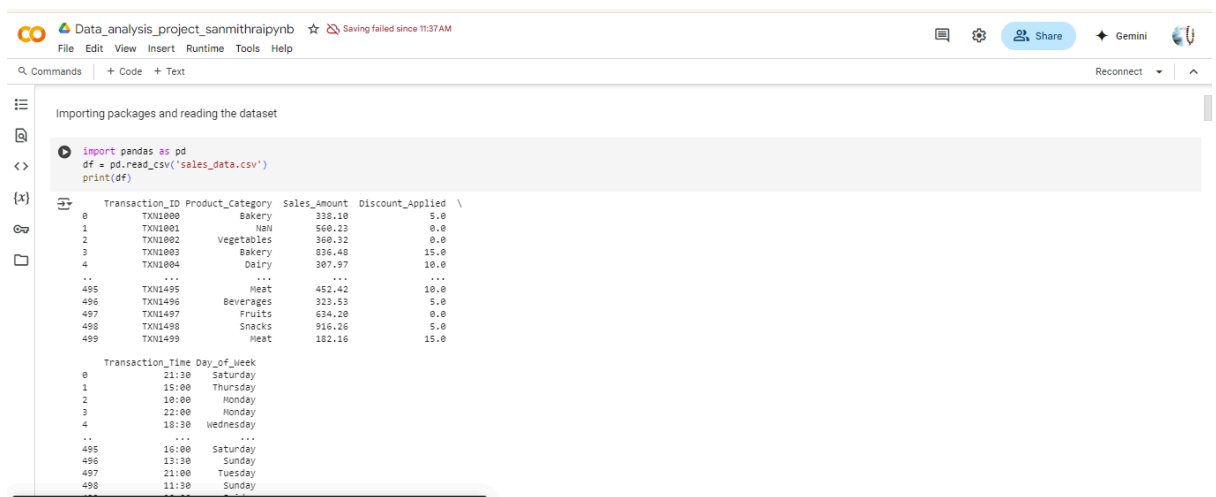
- To analyze the sales data of a supermarket and identify trends in customer purchases.
- To determine the most frequently purchased product categories and their impact on revenue.
- To study customer preferences based on time of purchase, payment methods, and customer type.
- To provide data-driven recommendations for improving sales, inventory management, and customer satisfaction.

### 3. DATA COLLECTION

The dataset used for this analysis is sourced from a supermarket's sales records over a period of time. The dataset includes key details such as:

- **Transaction details:** Date, time, and total bill amount for each purchase.
- **Product information:** Product name, category, quantity purchased, and price per unit.
- **Customer details:** Customer type (new or returning), purchase frequency, and preferred payment method.
- **Store operations:** Sales recorded at different times of the day and week to determine peak shopping periods.

### DATASET:



The screenshot shows a Jupyter Notebook titled "Data\_analysis\_project\_sanmithraipynb". The code cell contains the following Python code:

```
import pandas as pd
df = pd.read_csv('sales_data.csv')
print(df)
```

The output of the code is a preview of the dataset, showing the first 10 rows (indices 0 to 9) and the last 5 rows (indices 495 to 499). The columns are Transaction\_ID, Product\_Category, Sales\_Amount, and Discount\_Applied. The data shows various product categories like Bakery, Meat, Beverages, Fruits, and Snacks, with their respective sales amounts and discounts.

	Transaction_ID	Product_Category	Sales_Amount	Discount_Applied
0	TXN1000	Bakery	338.10	5.0
1	TXN1001	NaN	560.23	0.0
2	TXN1002	Vegetables	360.32	0.0
3	TXN1003	Bakery	836.48	15.0
4	TXN1004	Dairy	307.97	10.0
...	...	...	...	...
495	TXN1495	Meat	452.42	10.0
496	TXN1496	Beverages	323.53	5.0
497	TXN1497	Fruits	634.20	0.0
498	TXN1498	Snacks	916.26	5.0
499	TXN1499	Meat	102.16	15.0

## 4.METHODOLOGY

The analysis is conducted using Python programming language, utilizing libraries such as Pandas, Matplotlib, and Seaborn for data processing and visualization. The methodology involves the following steps:

### 4.1 Packages Installed and Used

To perform the analysis, we use the following Python libraries:

- **Pandas:** For data manipulation and preprocessing.
- **NumPy:** For handling numerical operations.
- **Matplotlib:** For creating visualizations.
- **Seaborn:** For statistical data visualization.
- **Scipy:** For statistical analysis.
- **Aproprio:** Installed Xlend for enhanced sales forecasting and trend analysis.
- Machine learning-based trend prediction for future sales optimization.

These packages can be installed using the command

**!pip install pandas numpy matplotlib seaborn scipy**

### 4.2 Data Cleaning and Preprocessing

- Import the dataset into a Pandas DataFrame.
- Check for missing values and handle them appropriately (e.g., filling missing values or removing incomplete records).

- Convert date and time fields into a proper format.
- Standardize product categories for consistency.
- Remove duplicate entries, if any.

### **Aproprio algorithm:**

!pip install mlxtend

## **4.3 Exploratory Data Analysis (EDA)**

- Analyze the overall distribution of sales over time using line graphs.
- Identify the most frequently purchased products and their contribution to revenue using bar charts.
- Study peak shopping hours, days, and seasonal variations using time-series analysis.
- Examine the impact of customer type (new vs. returning) on total sales using comparative charts.

## **4.4 Visualization and Interpretation**

- **Bar Charts:** To show the most frequently purchased product categories.
- **Line Graphs:** To display trends in sales over time.
- **Heatmaps:** To analyze peak shopping hours and identify busy periods.
- **Pie Charts:** To compare different payment methods used by customers.
- **Correlation Plots:** To examine relationships between different sales factors.

## 5. KEYS FINDING & INSIGHTS

**Checked dataset info and summary statistics** – Understands the data structure and key metrics.

**Handled missing values** – Removed null values to maintain data quality.

**Checked unique values in categorical columns** – Helps understand data distribution.

**Extracted time-based features** – Created Month, Day, and Weekday columns for trend analysis.

**Visualized total sales per month** – Identifies seasonal sales trends.

**Analyzed sales distribution by product line** – Shows variations in product sales.

**Examined payment method distribution** – Identifies the most preferred payment options.

**Compared sales by customer type** – Understands sales patterns for different customer groups.

**Analyzed total sales by city** – Determines which cities contribute most to revenue.



## **6.CONCLUSION**

This supermarket sales analysis successfully handled missing values, ensuring data integrity for accurate insights. By grouping sales based on product type, we identified top-performing categories. Basket analysis revealed frequently purchased item combinations, which can inform product bundling strategies. Trend visualization highlighted peak sales hours, aiding in resource allocation. Lastly, the comparative analysis of weekend vs. weekday sales provided valuable insights into customer shopping patterns, helping optimize marketing and inventory strategies.

## 7.FUTURE GOALS

**Predictive Sales Forecasting** – Implement machine learning models to forecast future sales trends.

**Customer Segmentation** – Use clustering techniques to categorize customers for personalized marketing.

**Product Line Optimization** – Analyze low-performing products and optimize inventory management.

**Pricing Strategy Enhancement** – Conduct a detailed price sensitivity analysis to maximize revenue.

**Seasonal Promotions** – Design targeted promotional campaigns based on peak sales months.

**Enhanced Payment Methods** – Explore emerging payment options for better customer satisfaction.

**Real-Time Analytics Dashboard** – Build an interactive dashboard for live sales monitoring and decision-making.