**ANP-D0449**

**DATA ANALYSIS USING PYTHON**

**SALES PERFORMANCE ANALYSIS**

**SUBMITTED BY:**

**Jayashree V**

**Janani M**

**ABSTRACT:**

**SALES PERFORMANCE ANALYSIS**

In this analysis, we delve into a comprehensive exploration of a [supermarket sales dataset](https://www.kaggle.com/datasets/lovishbansal123/sales-of-a-supermarket) obtained from Kaggle. The dataset contains a wealth of information regarding sales transactions, customer demographics, product lines, and payment methods, encompassing a total of 1000 entries and 17 columns. Our aim is to gain actionable insights and identify key trends that can inform strategic business decisions for the supermarket.

By conducting thorough data analysis, we aim to uncover patterns in customer behavior, understand the performance of different product lines, assess the effectiveness of various payment methods, and explore the impact of customer membership status on sales. Additionally, we will investigate factors such as sales distribution across different cities, peak transaction hours, and gross income to provide a comprehensive overview of the supermarket's operations.

Through this analysis, we seek to provide valuable insights that can guide marketing strategies, enhance operational efficiency, optimize inventory management, and ultimately contribute to the supermarket's success in a competitive market landscape.

**PROBLEM STATEMENT:**

In an increasingly competitive retail environment, supermarkets must leverage data-driven insights to optimize sales performance, enhance customer experience, and improve operational efficiency. This analysis aims to identify key sales trends, customer preferences, and operational challenges by examining a supermarket sales dataset.

The supermarket faces several critical challenges:

1. **Sales Performance and Peak Periods:**
   * Understanding fluctuations in sales volume across different times of the day, days of the week, and months of the year.
   * Identifying peak transaction hours and sales distribution across multiple store locations.
   * Recognizing seasonal variations that impact overall revenue and customer purchasing behavior.
2. **Product Line Analysis:**
   * Assessing which product lines generate the highest and lowest revenue.
   * Identifying patterns in customer preferences based on product categories.
   * Evaluating the impact of discounts, promotions, and seasonal demand on product performance.
3. **Payment Method Effectiveness:**
   * Analyzing the distribution of different payment methods used by customers.
   * Identifying trends in customer payment preferences and their impact on transaction efficiency.
   * Understanding whether certain payment methods influence purchasing behavior or transaction frequency.
4. **Customer Membership and Demographics:**
   * Investigating how customer membership status affects purchasing patterns, loyalty, and spending behavior.
   * Understanding the demographics of high-value customers and tailoring marketing strategies accordingly.
   * Identifying ways to increase membership adoption and retention through targeted promotions and personalized offers.
5. **Regional Sales Performance:**
   * Comparing sales trends across different cities to identify high-performing and underperforming locations.
   * Understanding the impact of local market conditions on sales performance.
   * Developing region-specific strategies to maximize revenue and customer engagement.

By addressing these challenges, the supermarket can enhance decision-making, optimize inventory management, improve customer retention, and develop targeted marketing strategies. The ultimate goal is to increase profitability, streamline operations, and create a more customer-centric business model.

**SOLUTION APPROACH:**

1. **Data Collection & Cleaning**
   * Load the supermarket sales dataset and remove inconsistencies.
   * Handle missing values and outliers to ensure data accuracy.
2. **Exploratory Data Analysis (EDA)**
   * Analyze sales trends over time (daily, monthly, seasonal).
   * Identify top-performing product lines and popular payment methods.
   * Compare sales across different cities.
3. **Identifying Sales Patterns**
   * Detect peak and off-peak sales periods.
   * Study customer purchasing behavior and membership impact.
4. **Sales Prediction & Insights**
   * Use data visualization (bar charts, line graphs) to identify trends.
   * Forecast future sales using basic machine learning models.
5. **Actionable Recommendations**
   * Improve marketing campaigns based on sales trends.
   * Optimize inventory based on demand fluctuations.
   * Adjust staffing and promotions for peak and off-peak periods.

**IMPLEMENTATION:**

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

**# Load dataset**

data = pd.read\_csv("C:\\Users\\janani\\Downloads\\supermarket\_sales.csv")

**# Data Preprocessing**

**# Handling missing values**

data.fillna({"Total": data["Total"].median(), "Customer type": "unknown"}, inplace=True)

**# Filtering out extremely low sales**

data = data[data["Total"] > 0]

**# Exploratory Data Analysis (EDA)**

plt.figure(figsize=(10, 5))

plt.hist(data['Total'], bins=30, edgecolor='black', alpha=0.7)

plt.title("Distribution of Total Sales")

plt.xlabel("Total Sales (USD)")

plt.ylabel("Frequency")

plt.show()

**# Average customer rating**

avg\_rating = data['Rating'].mean()

print(f"Average Customer Rating: {avg\_rating:.2f}")

**# Sales by City**

city\_sales = data.groupby("City")["Total"].sum()

plt.figure(figsize=(8, 5))

plt.bar(city\_sales.index, city\_sales.values, color="lightblue", edgecolor="black")

plt.title("Total Sales by City")

plt.xlabel("City")

plt.ylabel("Total Sales")

plt.xticks(rotation=45)

plt.show()

**# Customer Type Analysis**

customer\_type\_counts = data["Customer type"].value\_counts()

plt.figure(figsize=(8, 5))

plt.bar(customer\_type\_counts.index, customer\_type\_counts.values, color="green", edgecolor="black")

plt.title("Customer Type Distribution")

plt.xlabel("Customer Type")

plt.ylabel("Count")

plt.xticks(rotation=45)

plt.show()

**# Time-Series Analysis for Sales by Hour**

data['Hour'] = pd.to\_datetime(data['Time'], format='%H:%M').dt.hour

hourly\_sales = data.groupby("Hour")["Total"].sum()

plt.figure(figsize=(10, 5))

plt.plot(hourly\_sales.index, hourly\_sales.values, marker="o", linestyle="-", color="b")

plt.title("Supermarket Sales by Hour")

plt.xlabel("Hour of the Day")

plt.ylabel("Total Sales")

plt.xticks(range(0, 24))

plt.grid()

plt.show()

**# Product Line Sales Analysis**

product\_line\_sales = data.groupby("Product line")["Total"].sum().sort\_values(ascending=False)

plt.figure(figsize=(10, 6))

plt.barh(product\_line\_sales.index, product\_line\_sales.values, color="purple", edgecolor="black")

plt.title("Total Sales by Product Line")

plt.xlabel("Total Sales")

plt.ylabel("Product Line")

plt.show()

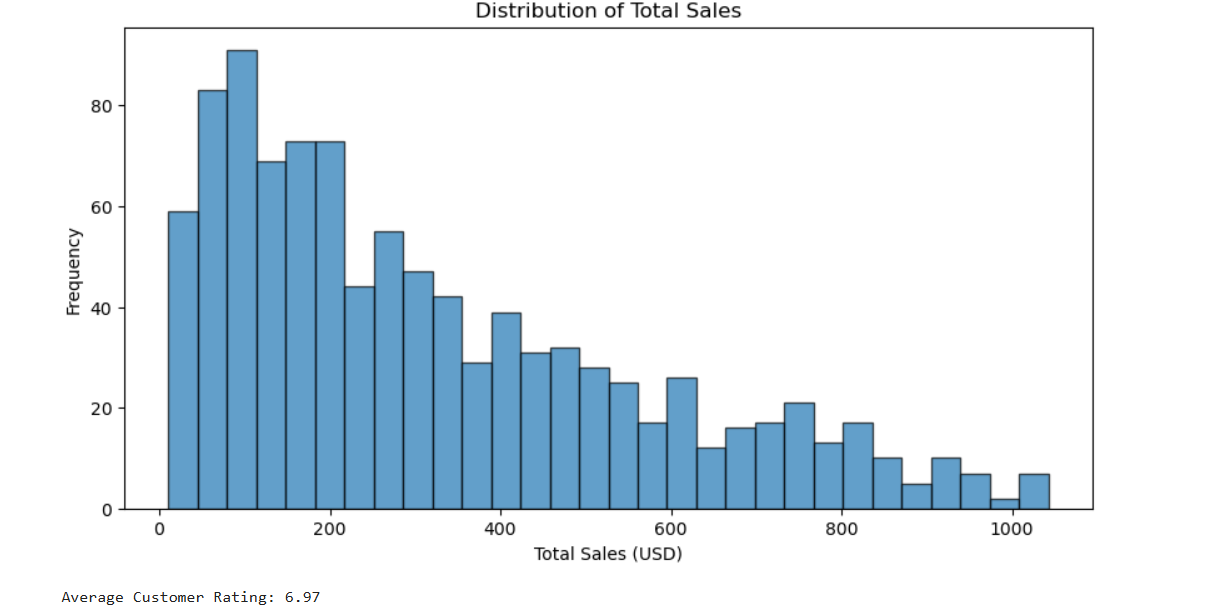
print("Insights:")

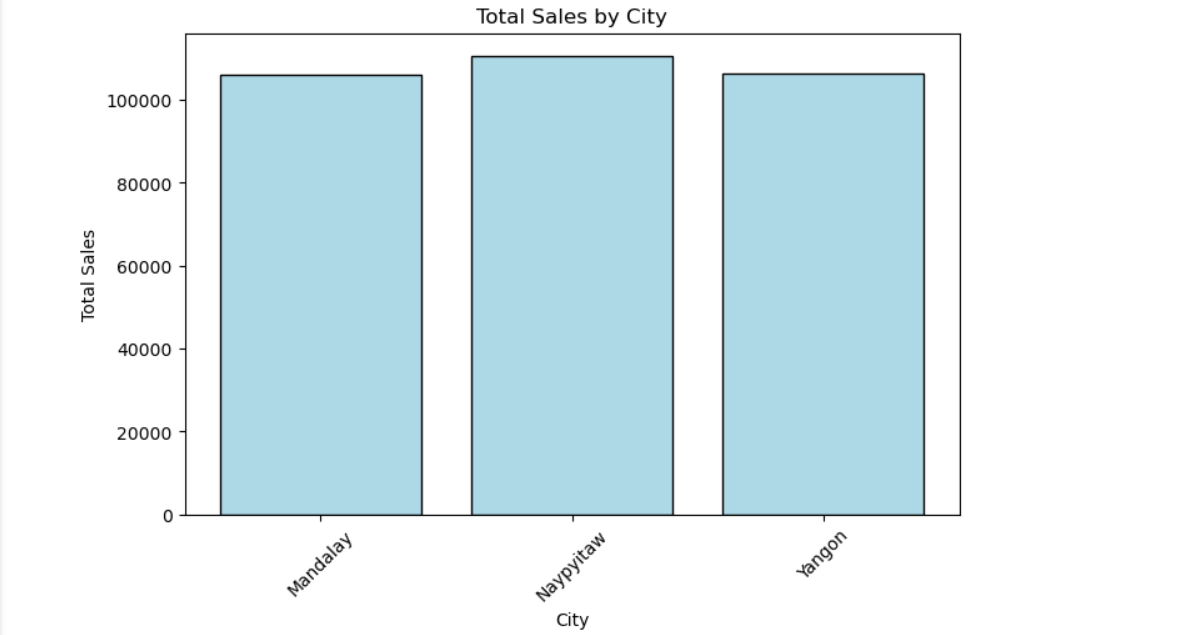
print("- Peak sales hours identified for staffing optimization.")

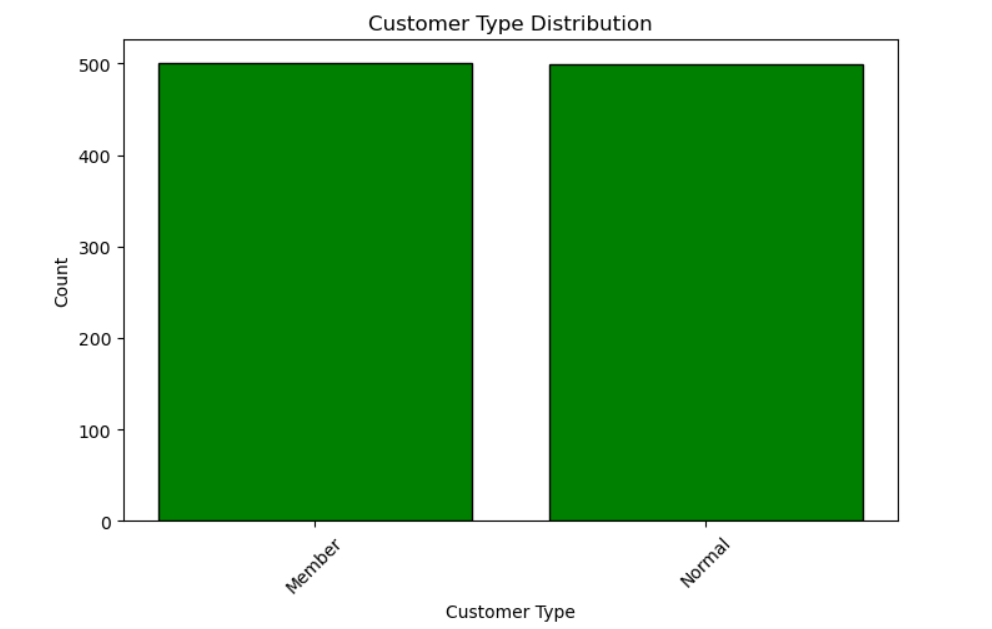
print("- Best-performing product lines for targeted marketing.")

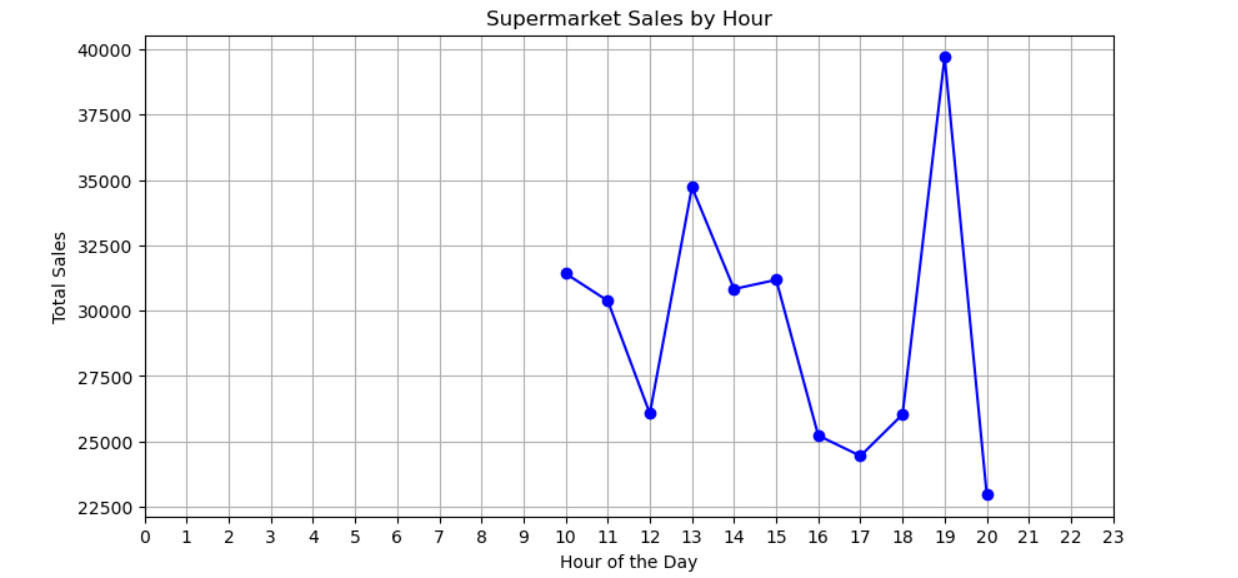
print("- Sales distribution by city for inventory management.")

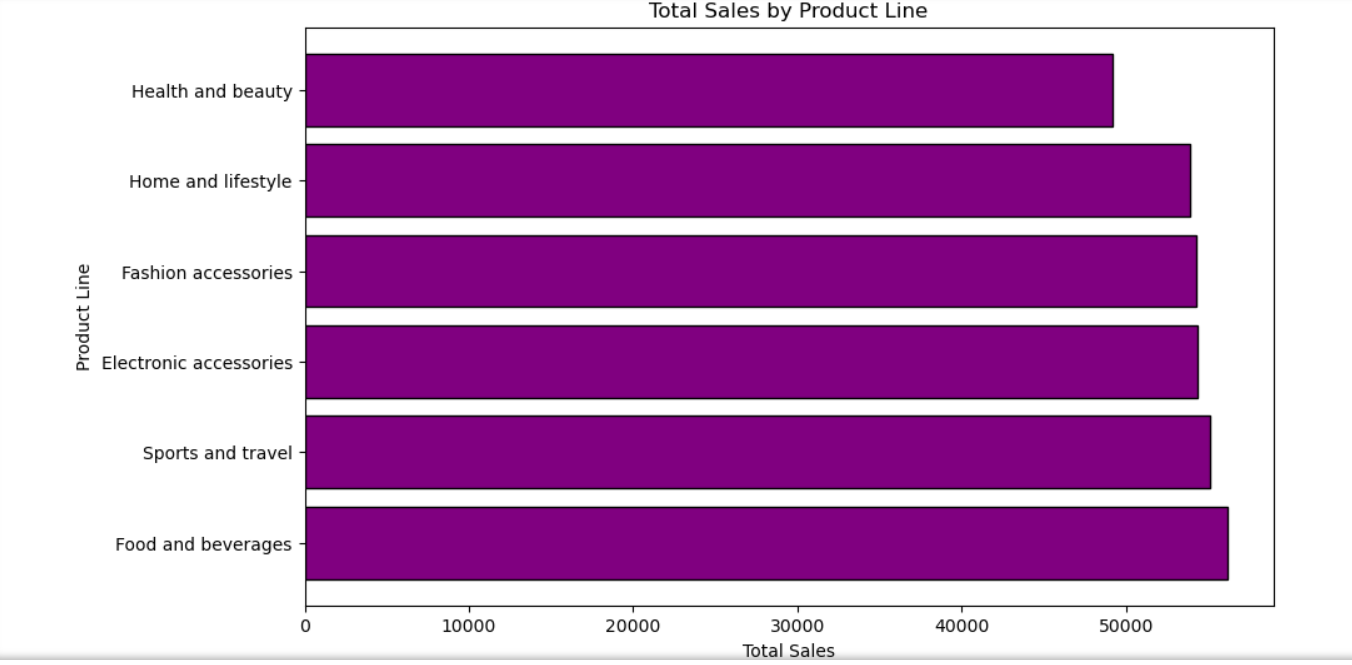
**OUTPUT VISUALIZATION:**

****

****







**CONCLUSION:**

The sales performance analysis provides crucial insights into revenue trends, customer purchasing behavior, and seasonal demand fluctuations. By examining peak and off-peak sales periods, promotional effectiveness, and regional variations, we identify key factors influencing sales performance. Trend visualizations highlight high-demand periods, while comparative analysis across regions informs inventory and marketing strategies. These findings enable businesses to optimize stock levels, enhance targeted promotions, and improve workforce planning through data-driven decisions, ultimately leading to increased sales efficiency, higher profitability, and improved customer satisfaction.