AND-DO449

DATA ANALYSIS USING PYTHON

SUPERMARKET SALES ANALYSIS: UNDERSTANDING BUYING PATTERNS

Submitted by:

Kaviya..S (AF0477607)

Archana.M(AF0477647)

ABSTRACT:

This study focuses on analyzing supermarket sales data to identify trends, patterns, and factors that influence consumer purchasing behavior. By examining historical sales records, the analysis aims to uncover correlations between variables such as product categories, pricing strategies, promotional offers, and seasonality. Advanced data mining and statistical techniques, including regression analysis and time series forecasting, are applied to predict future sales and optimize inventory management. The findings provide valuable insights for supermarket managers to enhance product assortment, tailor promotions, and improve customer targeting, ultimately increasing sales and profitability. Furthermore, the research highlights the impact of external factors such as economic conditions, competition, and consumer preferences on sales performance, offering a comprehensive understanding of supermarket operations.

PROBLEM STATEMENT:

Supermarkets face numerous challenges in accurately analyzing sales data to drive operational decisions and enhance profitability. One key challenge is the vast volume and complexity of data generated across different product categories, store locations, and time periods. This data is often fragmented and unstructured, making it difficult to extract actionable insights. Additionally, external factors such as shifting consumer preferences, economic fluctuations, and competition complicate the analysis further. The absence of robust predictive models to account for seasonal demand, price sensitivity, and promotional effectiveness leads to inefficiencies in inventory management, stockouts, and overstocking. Furthermore, supermarket managers often struggle with implementing data-driven strategies due to limited integration of sales data with other operational systems. The challenge lies in developing accurate, scalable, and actionable analytics to enhance decision-making processes, optimize sales strategies, and ultimately improve the customer experience.

The key challenges in supermarket sales analysis include:

1. **Data Volume and Complexity**: Supermarkets generate vast amounts of sales data across multiple product categories, store locations, and time periods. Managing and analyzing such large datasets can be overwhelming, especially when the data is unstructured or inconsistent.
2. **Data Integration**: Sales data is often siloed across various systems (inventory, customer relationship management, etc.), making it difficult to integrate and derive holistic insights. Lack of seamless integration between data sources hinders effective decision-making.
3. **Seasonality and Trends**: Predicting sales fluctuations due to seasonal demand, holidays, and changing consumer behavior poses a challenge. Without accurate forecasting, supermarkets may face stockouts or overstocking, impacting customer satisfaction and profitability.

4. **External Factors**: Economic shifts, regional trends, and competitive actions are difficult to quantify but significantly influence supermarket sales. Accounting for these variables in predictive models adds another layer of complexity.
5. **Promotional Effectiveness**: Assessing the success of discounts, offers, and in-store promotions is often difficult. Supermarkets struggle to measure the long-term impact of promotions on both short-term sales and customer loyalty.

SOLUTION APPROACH:

To address the challenges in supermarket sales analysis, a comprehensive and strategic solution approach can be employed, integrating data analytics, advanced technologies, and process improvements. Here are key solutions:

1. Data Integration and Centralization

- Solution: Implement a centralized data warehouse or cloud-based system that consolidates sales data, inventory data, customer insights, and promotional information from various sources. This integrated system ensures all data is easily accessible and can be analyzed in real-time.
- Benefit: Improves the accuracy and efficiency of data analysis by breaking down silos and enabling a holistic view of operations.

2. Advanced Data Analytics & Predictive Modeling

- Solution: Leverage machine learning models and advanced statistical techniques (e.g., time series forecasting, regression analysis) to predict future sales trends, customer behavior, and demand fluctuations. Models can factor in seasonality, promotions, and external factors like economic changes.
- Benefit: Enhances the accuracy of sales forecasts, enabling better inventory planning and promotional strategies, leading to reduced stockouts and overstocking.

3. Real-Time Analytics

- Solution: Utilize real-time data processing tools and dashboards to track sales performance, inventory levels, and customer behavior in real-time. This allows for immediate response to trends, such as adjusting promotions or reordering stock when necessary.
- Benefit: Ensures agile decision-making, helps supermarkets respond quickly to market changes, and reduces the risk of lost sales or excess inventory.

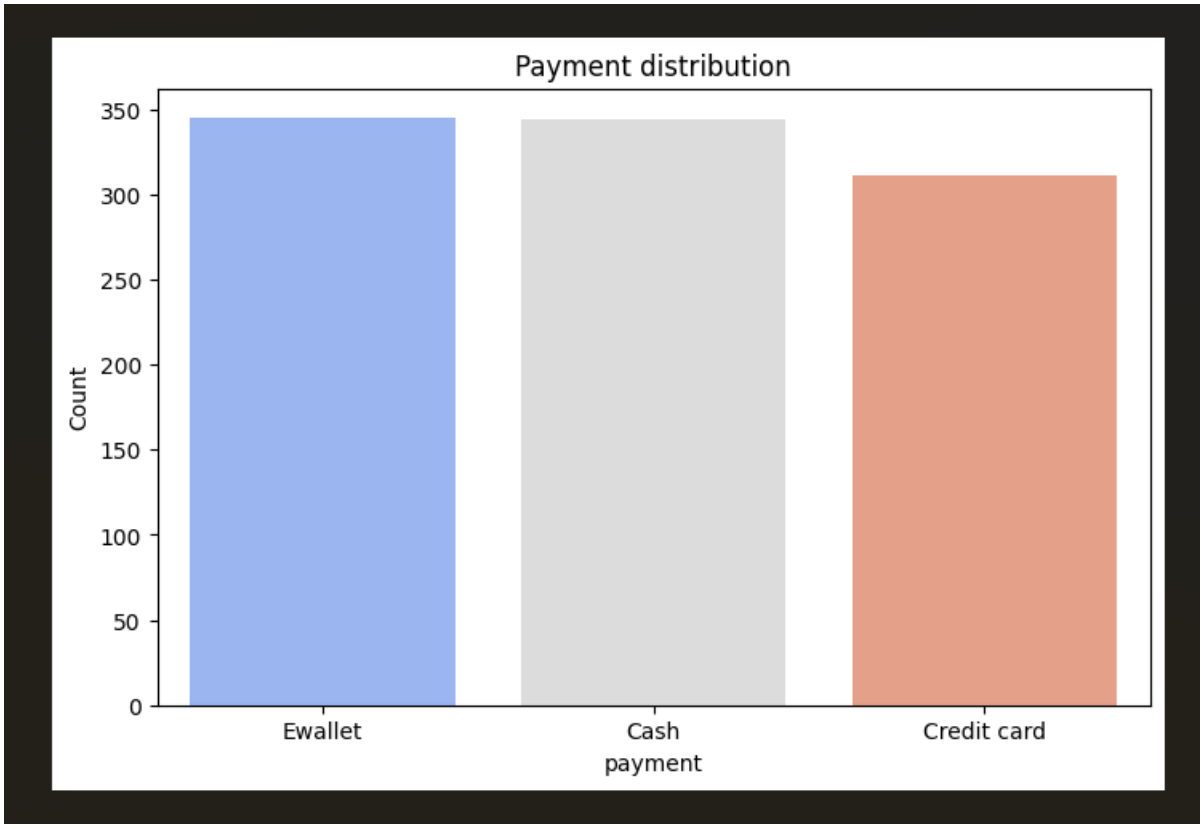
IMPLEMENTATION:

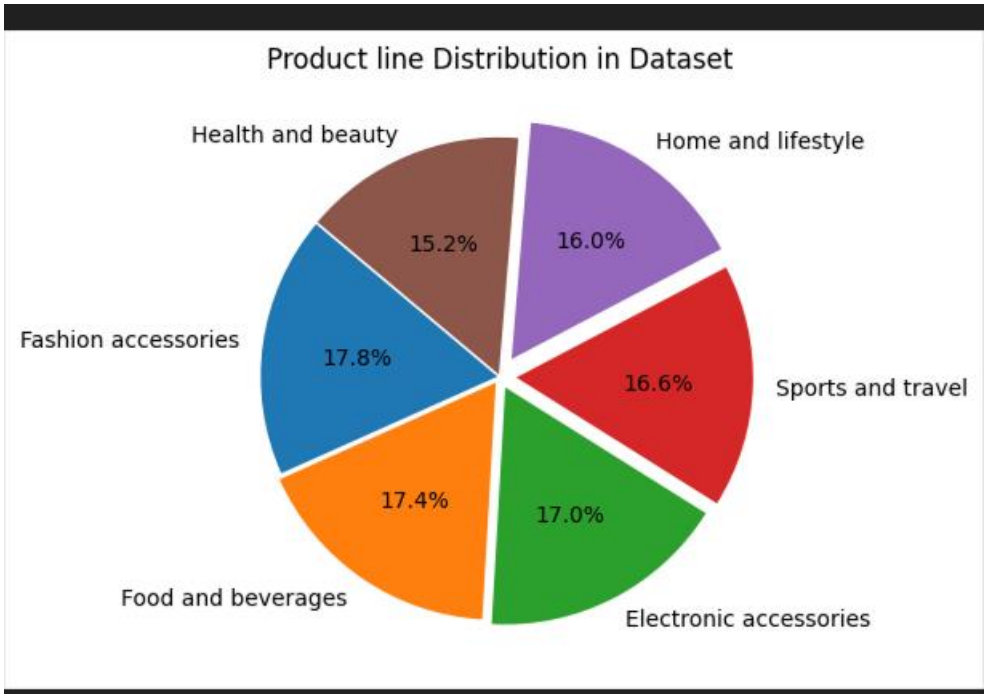
```
 import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

data=pd.read_csv('/content/supermarket_sales - Sheet1.csv')
print(data)
 print(data.head())
data=pd.read_csv('/content/supermarket_sales - Sheet1.csv')
print(data)
 print(data.tail(3))
 data.isnull().sum() #checking the null values
data.columns #check the column names
 counts = data['Gender'].value_counts()# Replace 'Category' with the actual column name
myexplode=[0.05,0.02]

plt.pie(counts, labels=counts.index, autopct='%1.1f%%', startangle=140, colors=['red',
'blue'],explode=myexplode)
plt.title("Gender Distribution in Dataset")
plt.legend(title='Gender Distribution',loc='upper left')
plt.show()
 plt.figure(figsize=(8, 5))
sns.countplot(x='Payment', data=data, palette='coolwarm')  # Replace with actual column
name
plt.title("Payment distribution")
plt.xlabel("payment")
plt.ylabel("Count")
plt.show()
```

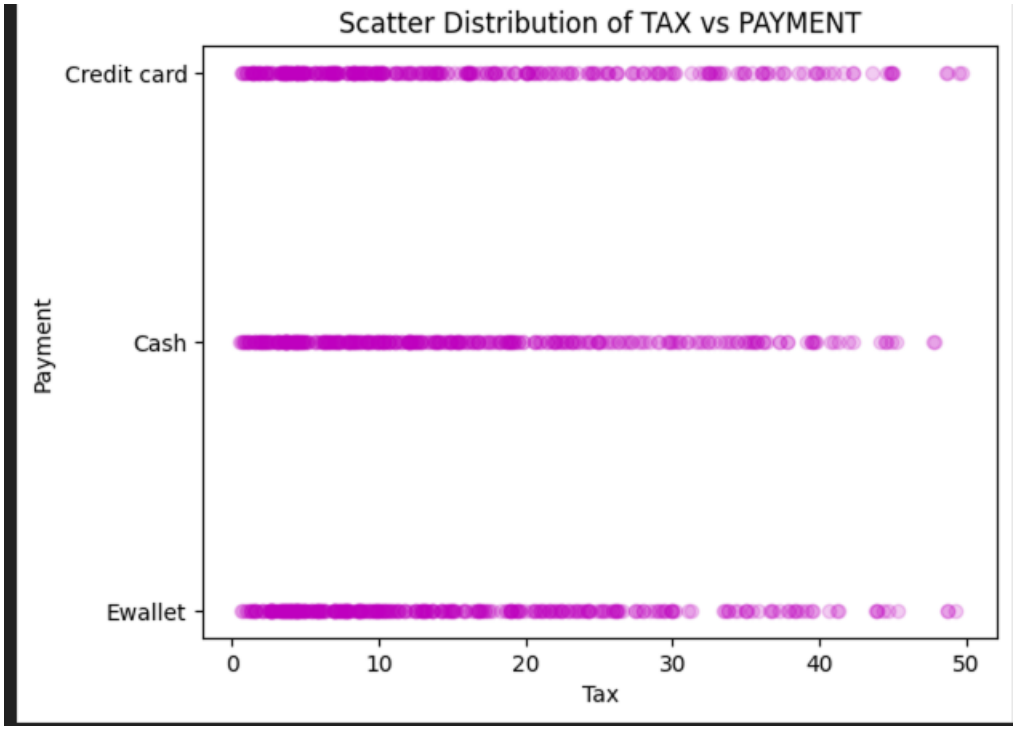
```
 counts = data['Product line'].value_counts()# Replace 'Category' with the actual column name

myexplode=[0.01,0.03,0.05,0.07,0.09,0.011]

plt.pie(counts, labels=counts.index, autopct='%1.1f%%', startangle=140,explode=myexplode)

plt.title("Product line Distribution in Dataset")

plt.show()

 x=data['Tax 5%']

y=data['Payment']

plt.scatter(x,y,color='m',alpha=0.2)

plt.xlabel('Tax')

plt.ylabel('Payment')

plt.title('Scatter Distribution of TAX vs PAYMENT')

plt.show()

 print(data.duplicated().sum())

 plt.figure(figsize=(10,20))

plt.hist(data['Product line'], bins=20, color='blue', edgecolor='black', alpha=0.7)  # Adjust bins as needed

plt.xlabel("Product Line")

plt.ylabel("Frequency")

plt.title("Product Line distribution")

plt.show()

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

plt.hist(data['Customer type'],bins=10, color='green', edgecolor='black', alpha=0.7)

plt.xlabel("Customer type")

plt.title("Distribution of customer type")

plt.show(

import seaborn as sns


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

sns.violinplot(y=data['Gender'], color='purple')
```

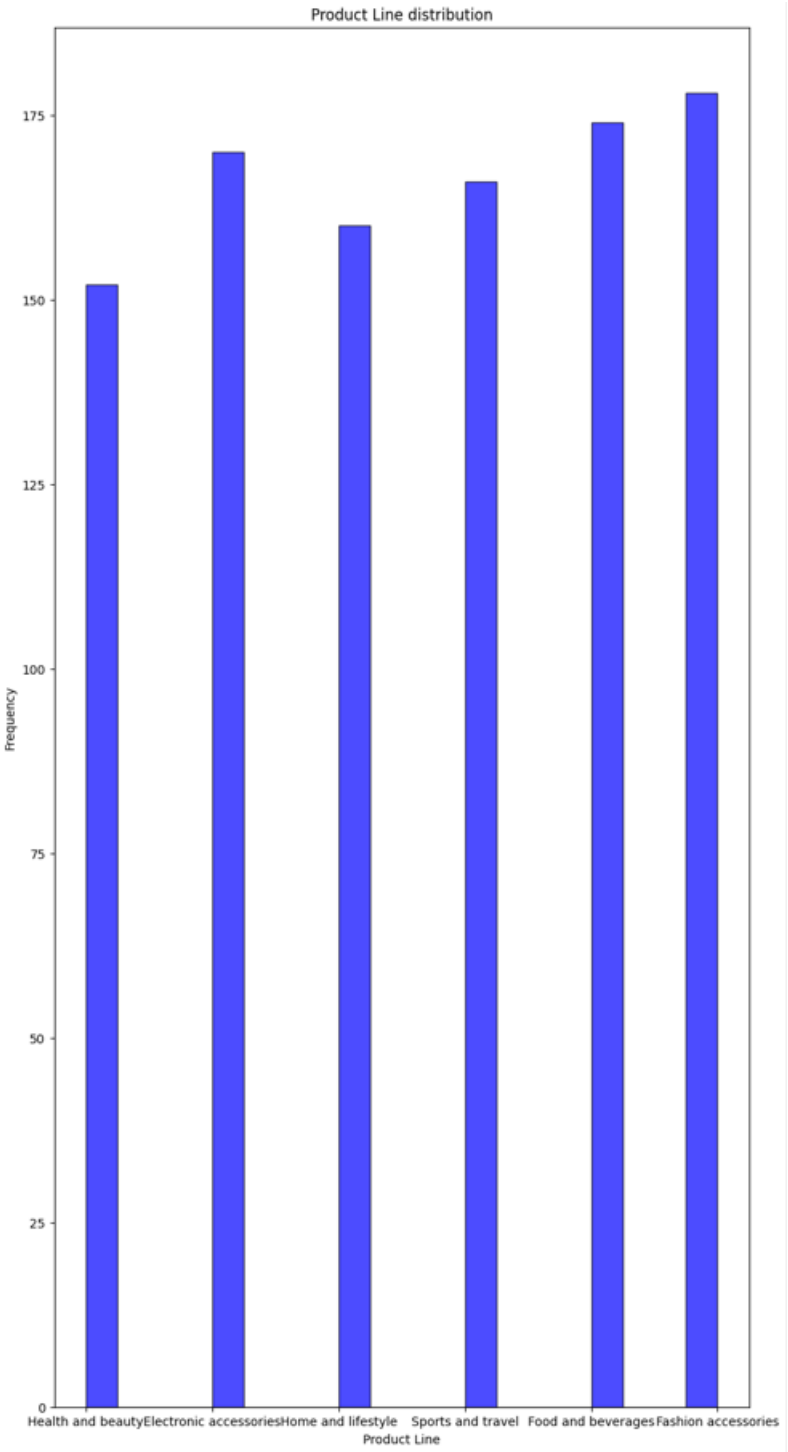
```
# Add labels and title
plt.title("Violin Plot for Age Distribution")
plt.ylabel("Age")

# Show plot
plt.show()
 Dummy=data.drop(columns=['cogs'])
Dummy
 sns.lineplot(data=data,x='Quantity',y='Payment')
 a=np.random.randint(100,size=(100))
b=np.random.randint(100,size=(100))
colors=np.random.randint(100,size=(100))
sizes=10*np.random.randint(100,size=(100))

plt.scatter(a,b,c=colors,cmap='nipy_spectral',s=sizes,alpha=0.6)
plt.xlabel('Monthly Revenue')
plt.ylabel('Subscription Type')
plt.colorbar()
plt.title('Monthly revenue VS Subscription Type')
plt.show()
datasets=sns.get_dataset_names()
print(datasets)
 data_sets=sns.load_dataset('diamonds')
data_sets
 pivot_table = data.pivot_table(values='Tax 5%', index='Gender', columns='Customer type',
aggfunc='mean')

# Print Pivot Table
print(pivot_table)
```

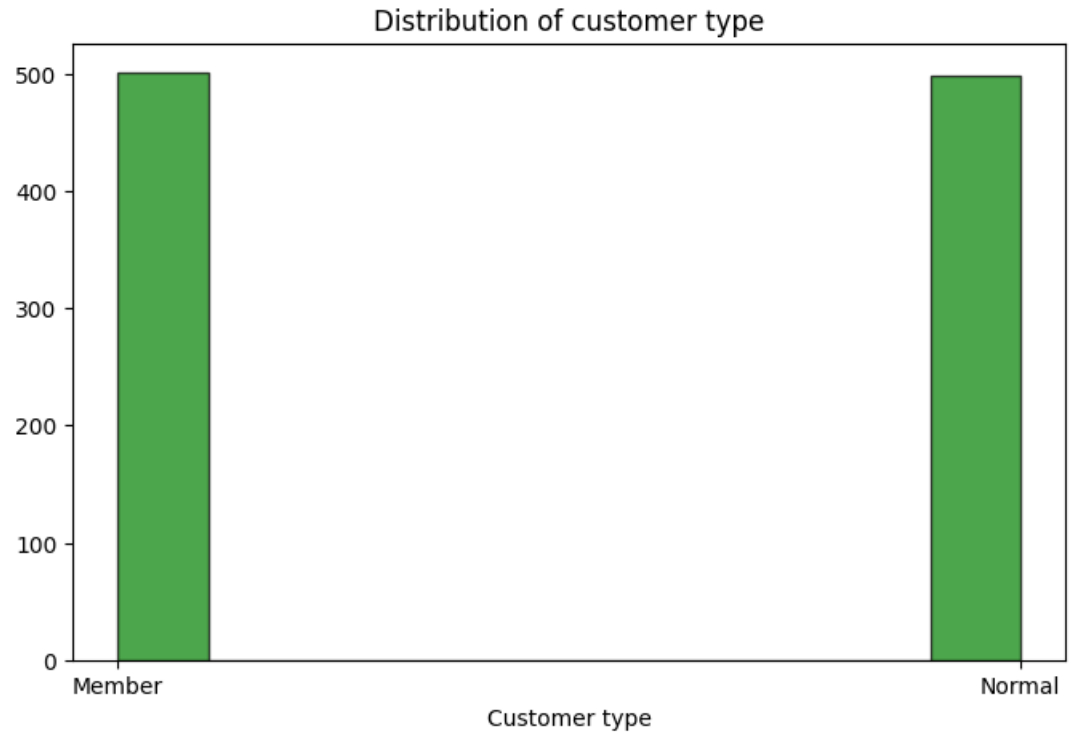
OUTPUT VISUALIZATION:

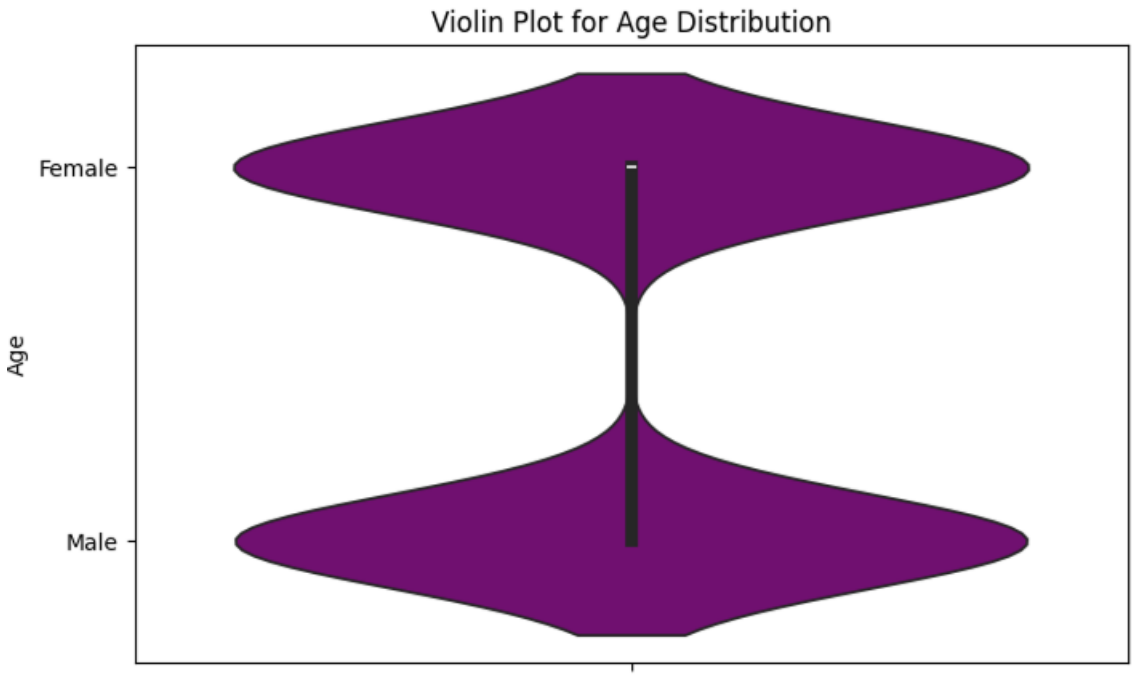


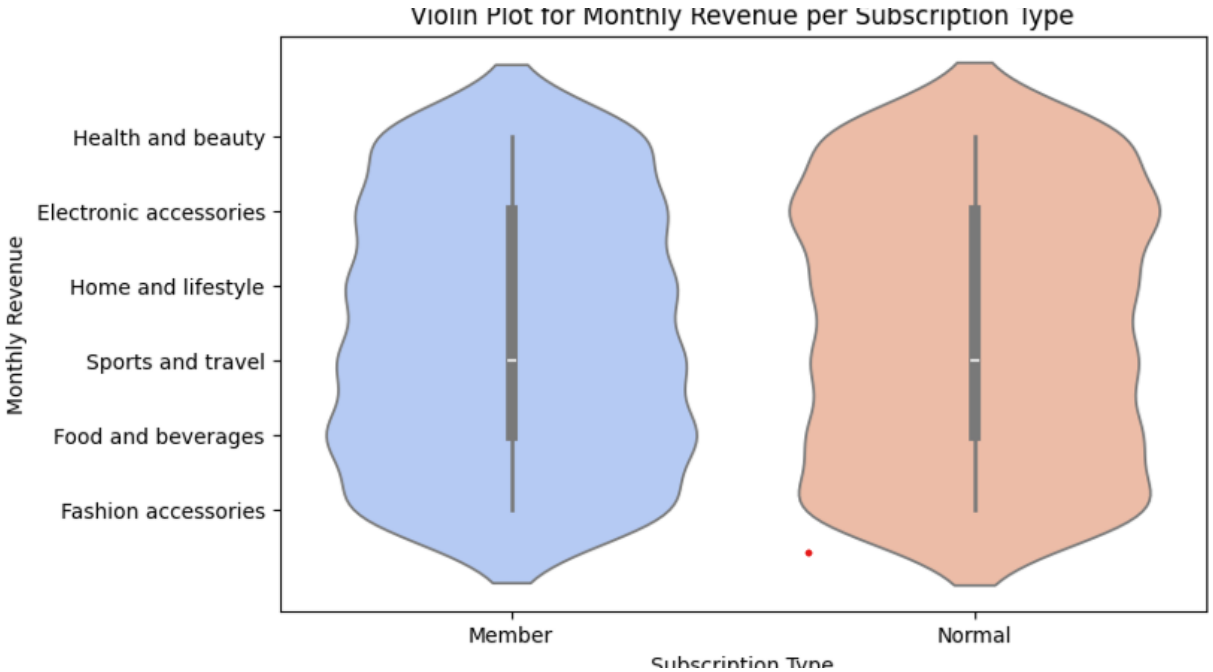


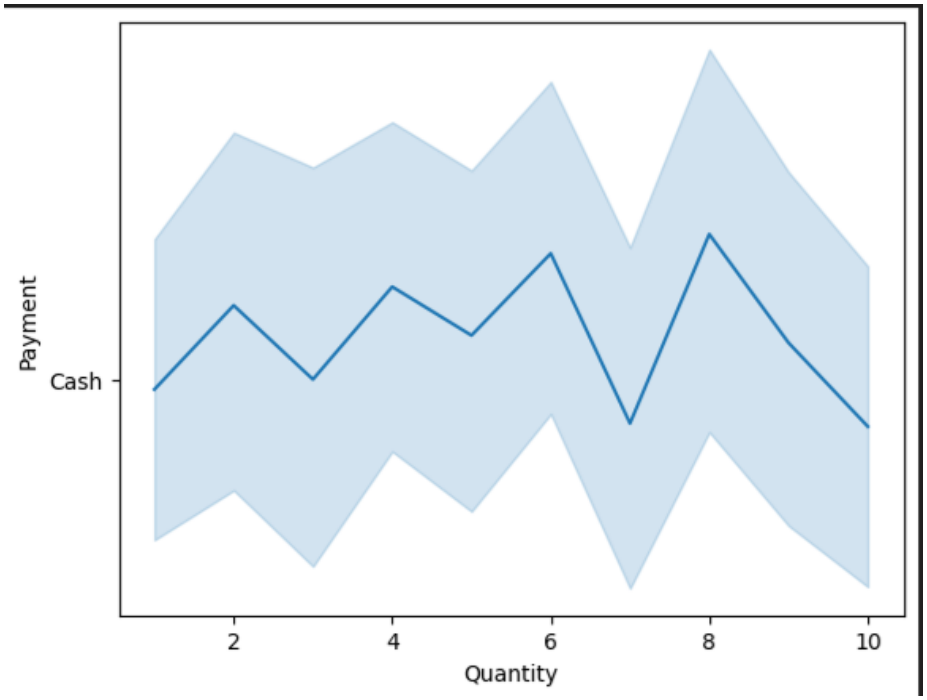

Scatter Distribution of TAX vs PAYMENT
Credit card
Cash
Ewallet
Payment
0
10
20
30
40
50
Tax

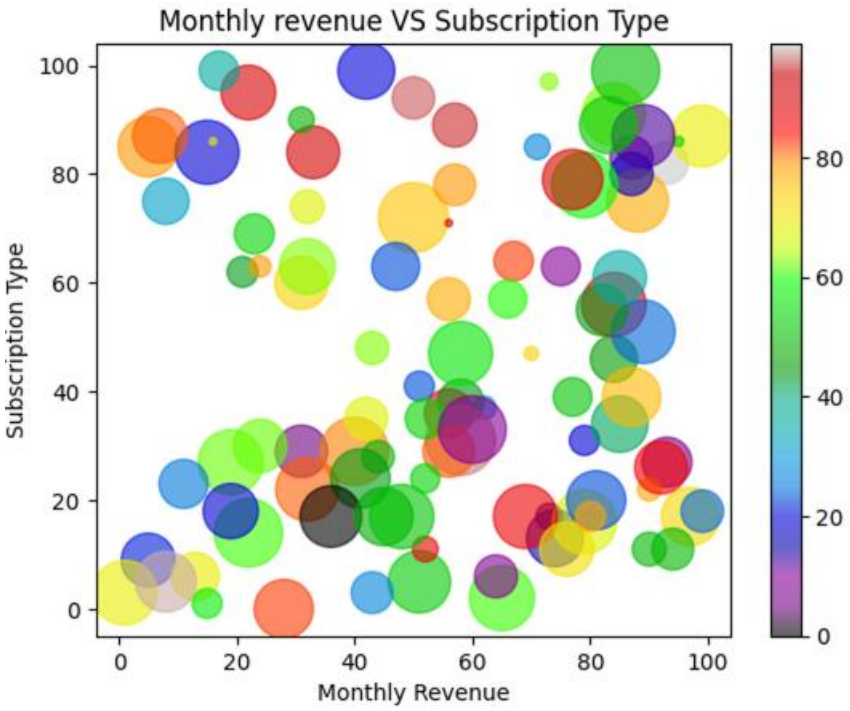

Product Line distribution
Frequency
0
25
50
75
100
125
150
175
Health and beauty
Electronic accessories
Home and lifestyle
Sports and travel
Food and beverages
Fashion accessories
Product Line

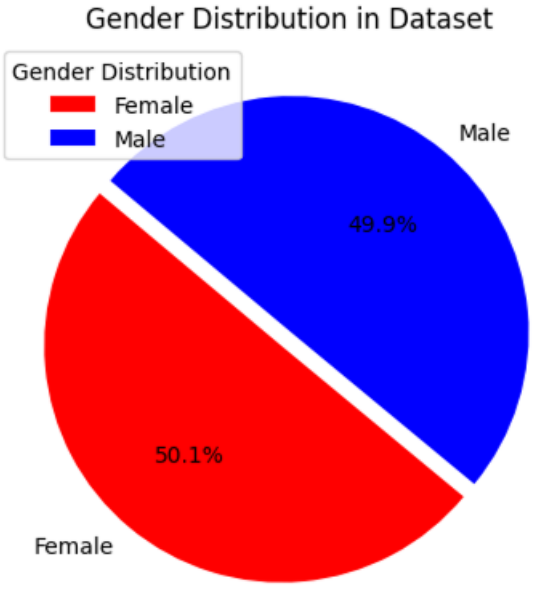

Distribution of customer type
500
400
300
200
100
0
Member
Normal
Customer type


Violin Plot for Age Distribution
Female
Male
Age


Violin Plot for Monthly Revenue per Subscription Type
Monthly Revenue
Health and beauty
Electronic accessories
Home and lifestyle
Sports and travel
Food and beverages
Fashion accessories
Member
Normal
Subscription Type


Payment
Cash
2
4
6
8
10
Quantity


Monthly revenue VS Subscription Type
Subscription Type
Monthly Revenue
0
20
40
60
80
100


Gender Distribution in Dataset
Gender Distribution
Female
Male
49.9%
50.1%

CONCLUSION:

Supermarket sales analysis presents a complex challenge due to the vast amount of data, external factors, and the need for real-time decision-making. However, by adopting a comprehensive approach that integrates advanced data analytics, predictive modeling, and real-time insights, supermarkets can unlock significant operational efficiencies and improve profitability. Leveraging technologies such as machine learning, AI, and cloud-based analytics tools allows supermarkets to optimize inventory management, enhance promotional strategies, and better understand customer preferences.

Furthermore, integrating various data sources into a unified system, optimizing sales forecasting, and adopting personalized marketing strategies help supermarkets navigate the intricacies of consumer behavior and market trends. By addressing the challenges through effective solutions, supermarkets can not only increase sales but also create a more personalized, efficient, and agile shopping experience for customers, ultimately driving long-term success in an increasingly competitive market.