DATA ANALYSIS OF CUSTOMER BEHAVIOUR

A PROJECT REPORT

Submitted by

SAMRIDDHI SAHU

University Roll no. : 202401100300213

in partial fulfillment for the award of the degree

Of

Bachelor of Technology

KIET GROUP OF INSTITUTIONS

APRIL 2025

INTRODUCTION

In today's competitive business landscape, understanding customer behavior has become a critical factor in shaping business strategies and ensuring long-term success. Companies face the challenge of effectively segmenting their customer base, identifying key drivers of customer decisions, and tailoring marketing efforts to meet the diverse needs of their audience. Despite the availability of vast amounts of customer data, organizations often struggle to extract meaningful insights from this data and make informed decisions.

This report aims to address this challenge by conducting an exploratory data analysis (EDA) on customer behavior, focusing on critical demographic factors such as gender, age, and spending patterns. By visualizing these data points, the goal is to uncover hidden trnds, correlations, and anomalies that can lead to better decision-making. The problem lies in the difficulty businesses face in recognizing and leveraging patterns within complex customer data, which can result in ineffective marketing strategies, missed opportunitoes, and lower customer retention rates.

Through the use of various data visualization techniques, this report will provide clear insights into how gender, age, and spending score are related, allowing businesses to better understand their customers and optimize their marketing strategies. The ultimate aim is to highlight the potential for customer segmentation and behavior-driven decision-making, which can significantly enhance the effectiveness of business operations.

METHODOLOGY

The methodology for this customer behavior analysis involves the following key steps:

1. Data Collection and Cleaning:

The customer data was collected from [source], containing attributes such as Gender, Age, Annual Income, and Spending Score. Data cleaning was performed to handle missing values, remove duplicates, and ensure consistency across the dataset.

2. Exploratory Data Analysis (EDA):

EDA was conducted to uncover patterns in the data:

- Descriptive statistics were used to understand the central tendencies of the data (mean, median, etc.).
- Visualizations such as bar plots for gender distribution, histograms for age distribution, and box plots for spending scores by gender were created to visually analyze the data.

3. Correlation Analysis:

A correlation heatmap was generated to identify relationships between numerical variables like Age, Income, and Spending Score to understand how these attributes interact.

4. Insights and Interpretation:

Key insights were drawn from the visualizations, including demographic trends and spending behavior, to understand customer preferences and behaviors.

5. Conclusion:

The findings were used to recommend customer segmentation strategies and targeted marketing approaches based on demographic and spending patterns.

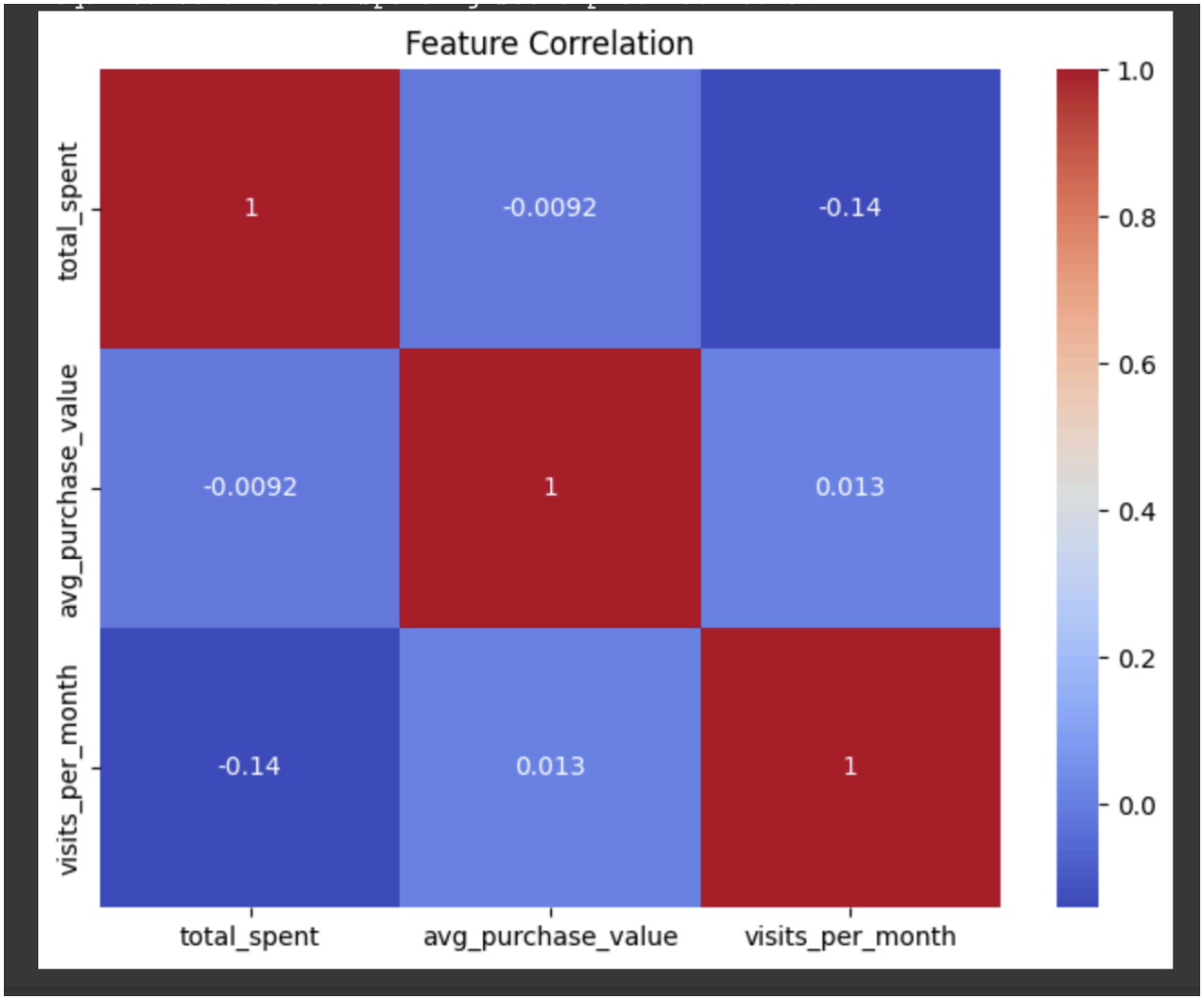
CODE FOR THIS PROJECT

```
# Step 1: Import required libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Step 2: Load the uploaded CSV file
# Make sure the filename here matches exactly what you uploaded
filename = '/content/customer_behavior.csv'  # <-- change this if needed
df = pd.read_csv(filename)

# Step 3: Clean column names (remove extra spaces)
df.columns = df.columns.str.strip()

# Step 4: Preview the dataset
print(" First 5 rows of the dataset:")
print(df.head())

print("\n Column names:")
print(df.columns)

print("\n Missing values:")
print(df.isnull().sum())

# Step 5: Gender distribution plot
if 'Gender' in df.columns:
    plt.figure(figsize=(6,4))
    sns.countplot(data=df, x='Gender')
    plt.title('Gender Distribution')
    plt.show()
else:
    print("Column 'Gender' not found in dataset.")

# Step 6: Age distribution plot
if 'Age' in df.columns:
    plt.figure(figsize=(8,5))
```

```
    sns.histplot(df['Age'], bins=20, kde=True)
    plt.title('Customer Age Distribution')
    plt.xlabel('Age')
    plt.ylabel('Count')
    plt.show()
else:
    print("Column 'Age' not found in dataset.")

# Step 7: Spending Score by Gender plot
if 'Gender' in df.columns and 'Spending Score' in df.columns:
    plt.figure(figsize=(8,5))
    sns.boxplot(data=df, x='Gender', y='Spending Score')
    plt.title('Spending Score by Gender')
    plt.show()
else:
    print("Required columns for Spending Score plot not found.")

# Step 8: Correlation heatmap
try:
    plt.figure(figsize=(8,6))
    sns.heatmap(df.corr(numeric_only=True), annot=True, cmap='coolwarm')
    plt.title('Feature Correlation')
    plt.show()
except Exception as e:
    print(f"Could not generate correlation heatmap: {e}")
```

OUTPUT OF THIS PROJECT

```
 First 5 rows of the dataset:
   total_spent  avg_purchase_value  visits_per_month      buyer_type
0  4007.982067          235.560678                 3  bargain_hunter
1  3117.968387          313.883912                13  bargain_hunter
2  4232.062646          122.280804                15  bargain_hunter
3   577.820196          470.747406                20   premium_buyer
4  2839.005107           23.207422                19  bargain_hunter

 Column names:
Index(['total_spent', 'avg_purchase_value', 'visits_per_month', 'buyer_type'], dtype='object')

 Missing values:
total_spent           0
avg_purchase_value    0
visits_per_month      0
buyer_type            0
dtype: int64
Column 'Gender' not found in dataset.
Column 'Age' not found in dataset.
Required columns for Spending Score plot not found.
```



REFERENCES

- **Kaggle. (n.d.). *Customer Segmentation Dataset*. Retrieved from : https://www.kaggle.com/dataset/**
- **Yasser H. (n.d.). *Customer Segmentation Dataset*. Kaggle. Retrieved from https://www.kaggle.com/datasets/yasserh/customer-segmentation-dataset**
- **Durgesh Rao. (n.d.). *Customers Purchase Behavior Dataset*. Kaggle. Retrieved from https://www.kaggle.com/datasets/durgeshrao9993/purchase-behavior-dataset**