Python Project
Subject code: DA514

ANALYZING DIWALI SALES DATA

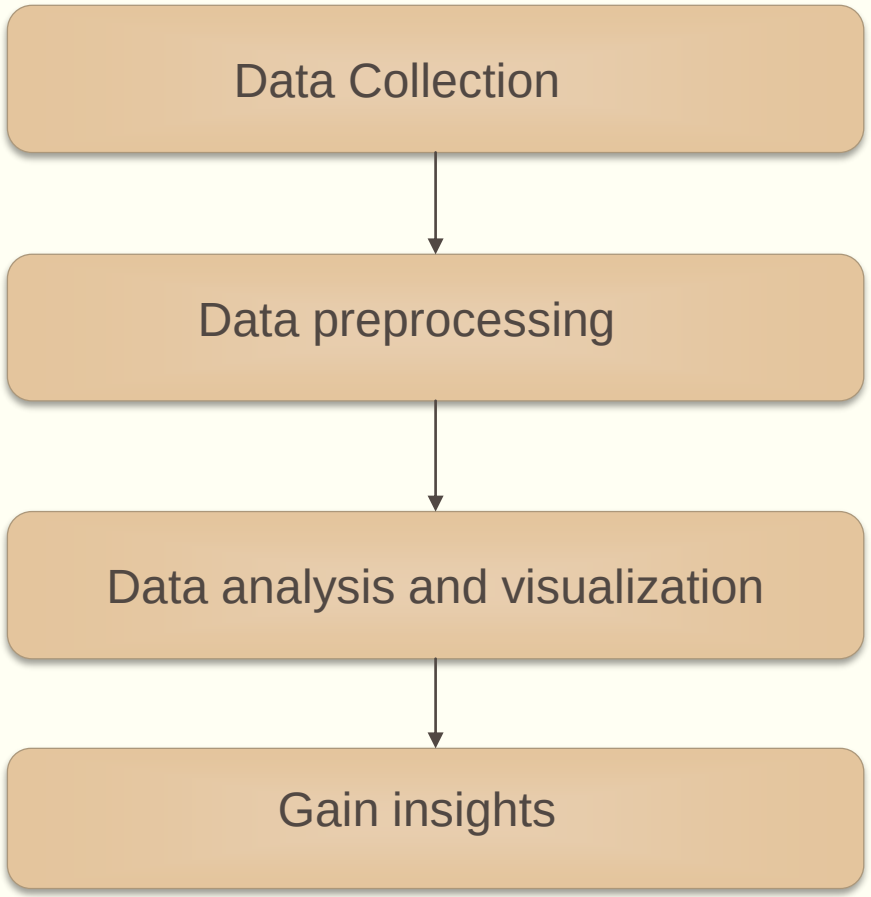
Presentation by krishnapal Yadav

Course Instructor - Neeraj Kumar Sharma

Motivation

Understanding customer purchasing behavior during Diwali To analyze Diwali sales data and gain insights into customer demographics, purchasing behavior, and sales trends to inform strategic business decisions

Block diagram



Pseudo-code

Import libraries
Read dataset
Cleanse dataset

```
def plot_gender_vs_total_amount()

def plot_age_distribution()

def plot_state_wise_sales()

def plot_product_category_vs_amount()

def plot_marital_status_vs_amount()

def plot_stacked_bar_and_pie()

def plot_zone_vs_total_amount()

def plot_gender_vs_avg_order_value()

def plot_age_vs_amount()
```

```
def main():

    plot_gender_vs_total_amount()
    plot_age_distribution()
    plot_state_wise_sales()
    plot_product_category_vs_amount()
    plot_marital_status_vs_amount()
    plot_stacked_bar_and_pie()
    plot_zone_vs_total_amount()
    plot_gender_vs_avg_order_value()
    plot_age_vs_amount()
    plot_top_10_product_categories()
    plot_order_frequency_by_age_group()
    plot_occupation_vs_zone_sales()
    plot_repeat_orders()
    plot_avg_order_value_by_age_group()
    plot_occupation_vs_state_area()



if __name__ == "__main__":

    main()
```

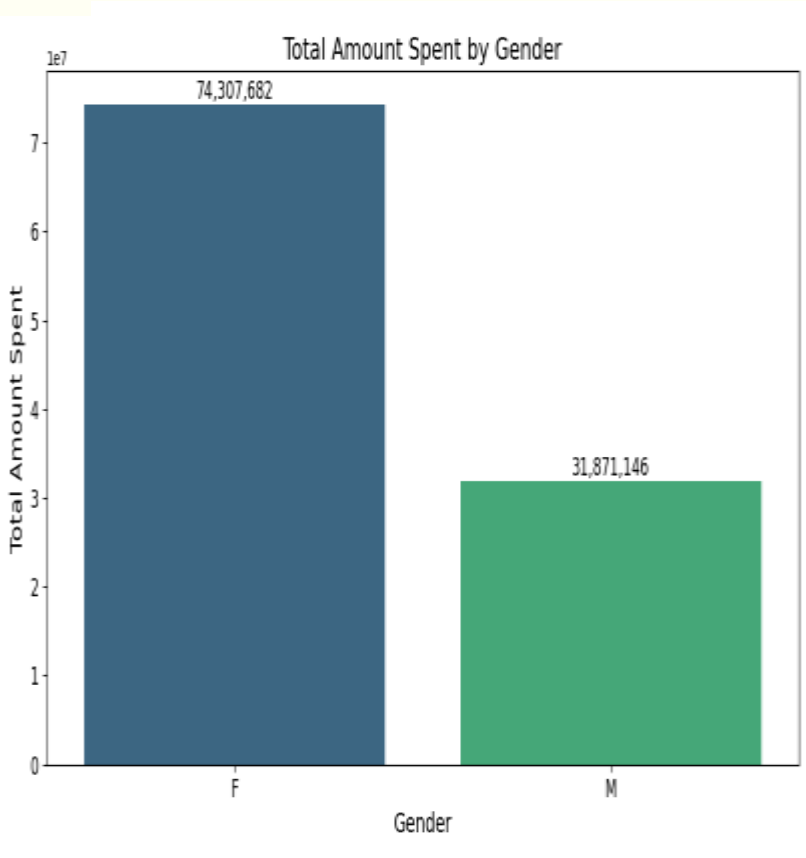
Snapshots

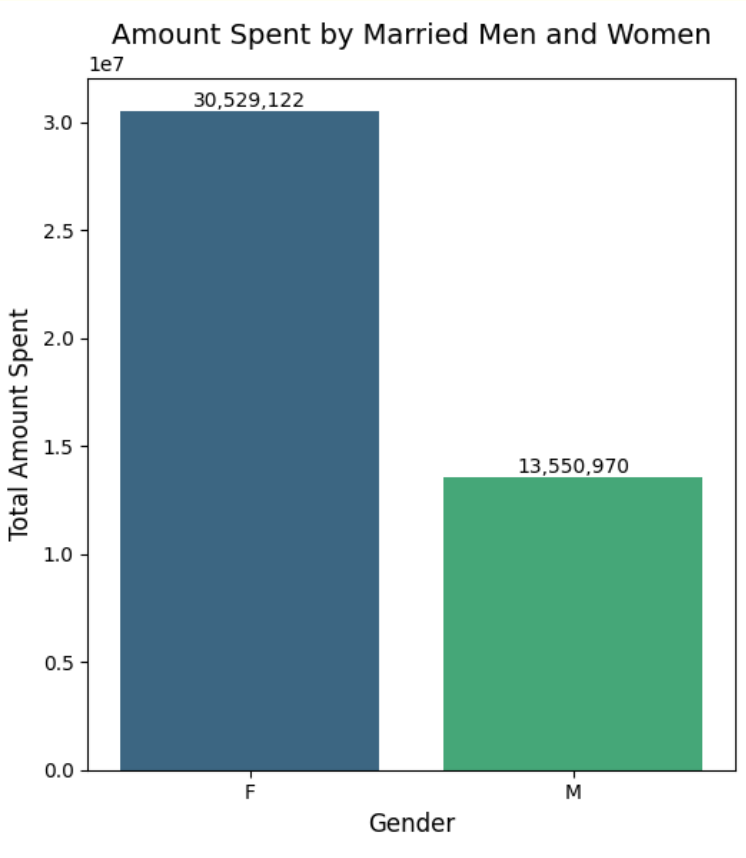
```python
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

def read_dataset(filepath="diwali_sales_data.csv"):

    try:
        df = pd.read_csv(filepath)
        # dropping the 'User_ID' column
        df = df.drop(['User_ID'], axis=1)

        df = df.drop_duplicates()
        print(f"The dimensions of the imported dataset: {df.shape}")
        return df
```

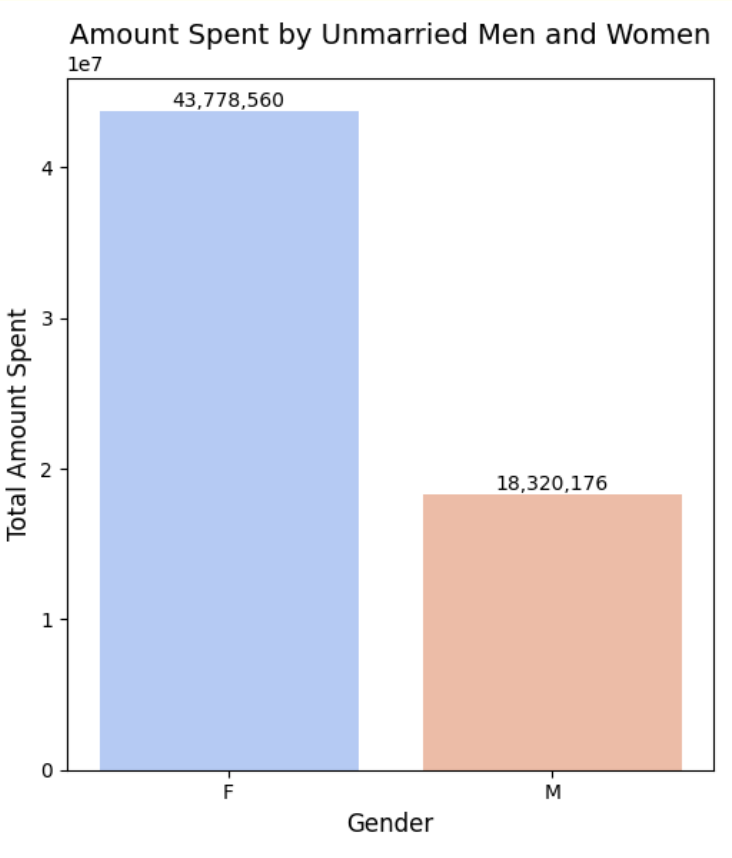
```python
    median_age = np.median(df['Age'])
    df['Age'] = np.where(df['Age'] > 100, np.nan, df['Age'])
    df['Age'].fillna(median_age, inplace=True)

    final_row_count = df.shape[0]
    print(f"Rows dropped during cleansing: {initial_row_count - final_row_count}")

    return df
```

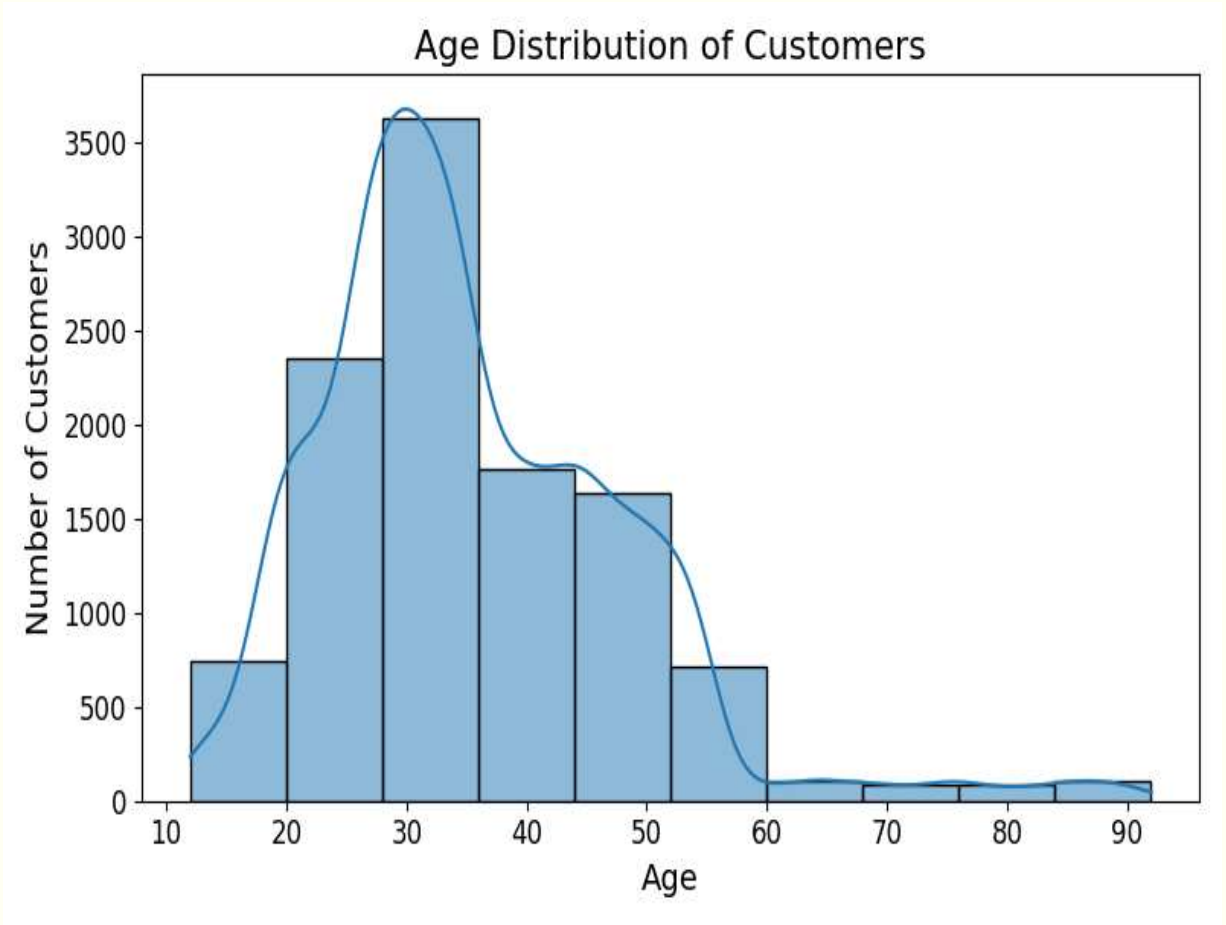
```python
def plot_age_distribution(df):
    plt.figure(figsize=(8, 5))
    sns.histplot(df['Age'], bins=10, kde=True)

    plt.title('Age Distribution of Customers', fontsize=16)
    plt.xlabel('Age', fontsize=14)
    plt.ylabel('Number of Customers', fontsize=14)
    plt.xticks(fontsize=12)
    plt.yticks(fontsize=12)
    plt.tight_layout()
    plt.show()
```

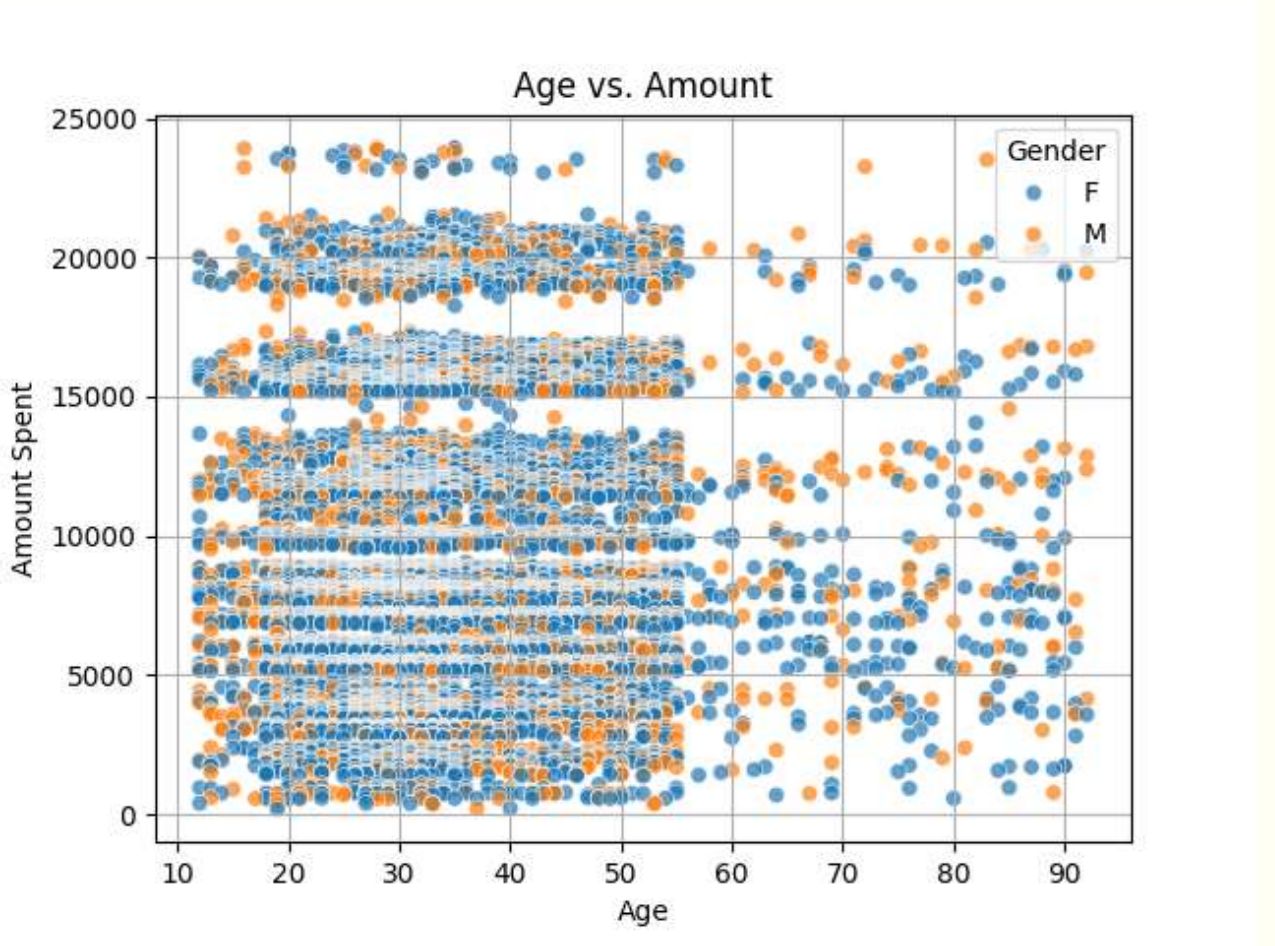
Results

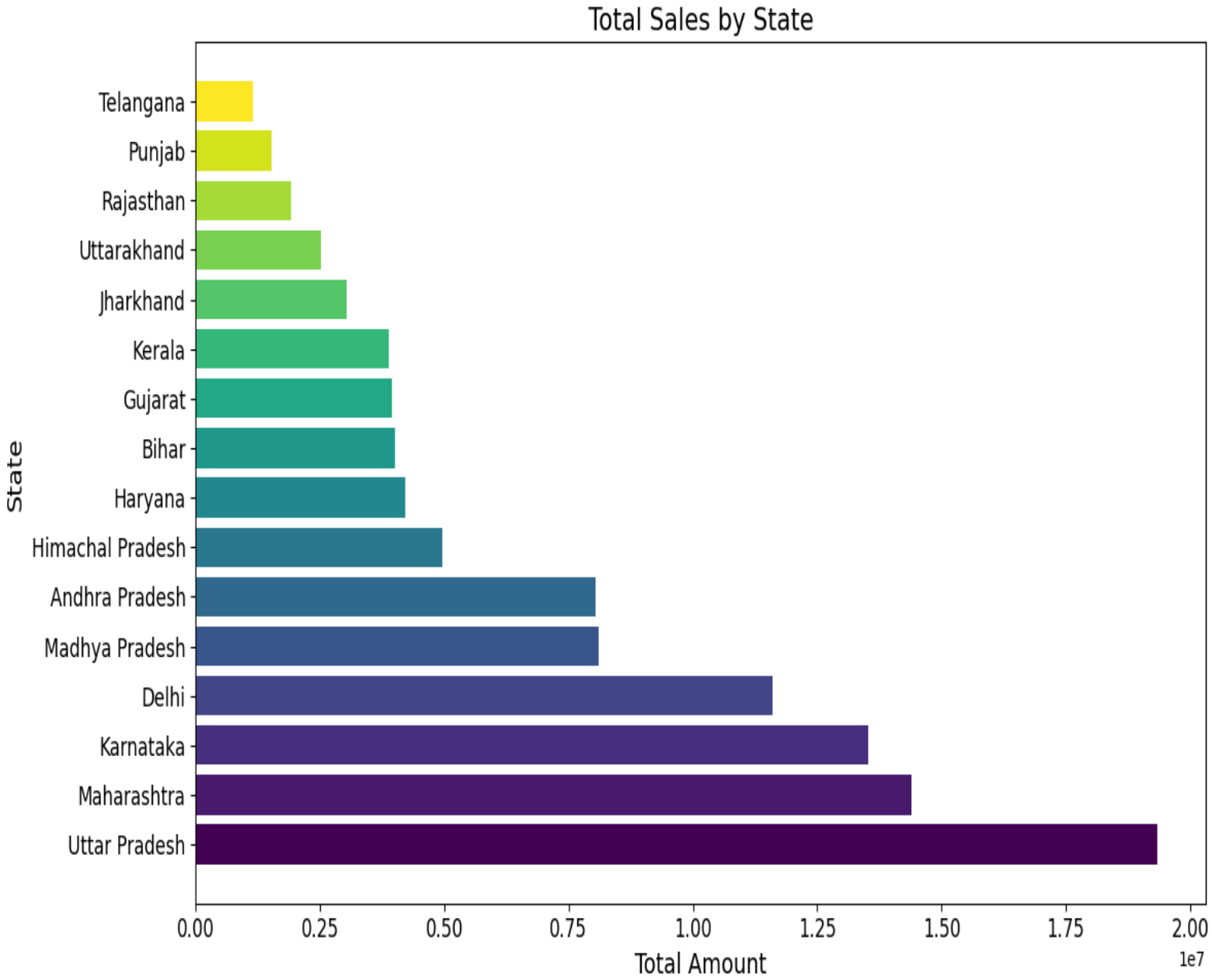

Total Amount Spent by Gender
1e7
74,307,682
31,871,146
Total Amount Spent
Gender
F
M

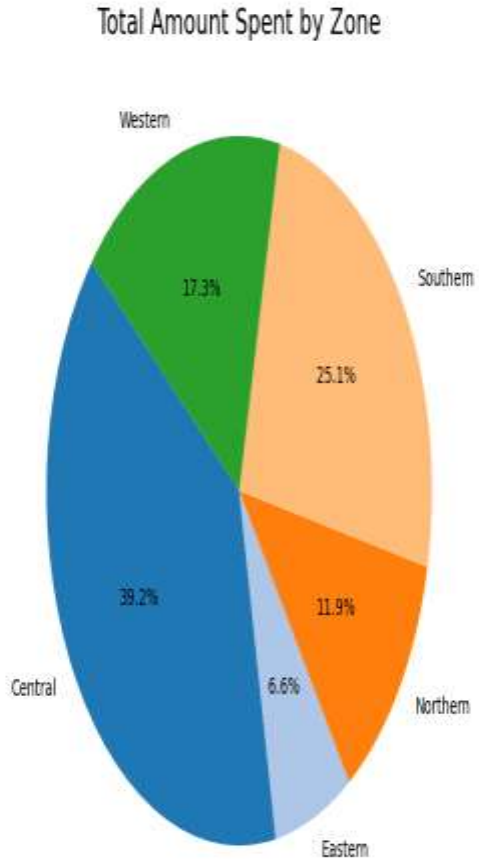

Amount Spent by Married Men and Women
1e7
30,529,122
13,550,970
Total Amount Spent
Gender
F
M

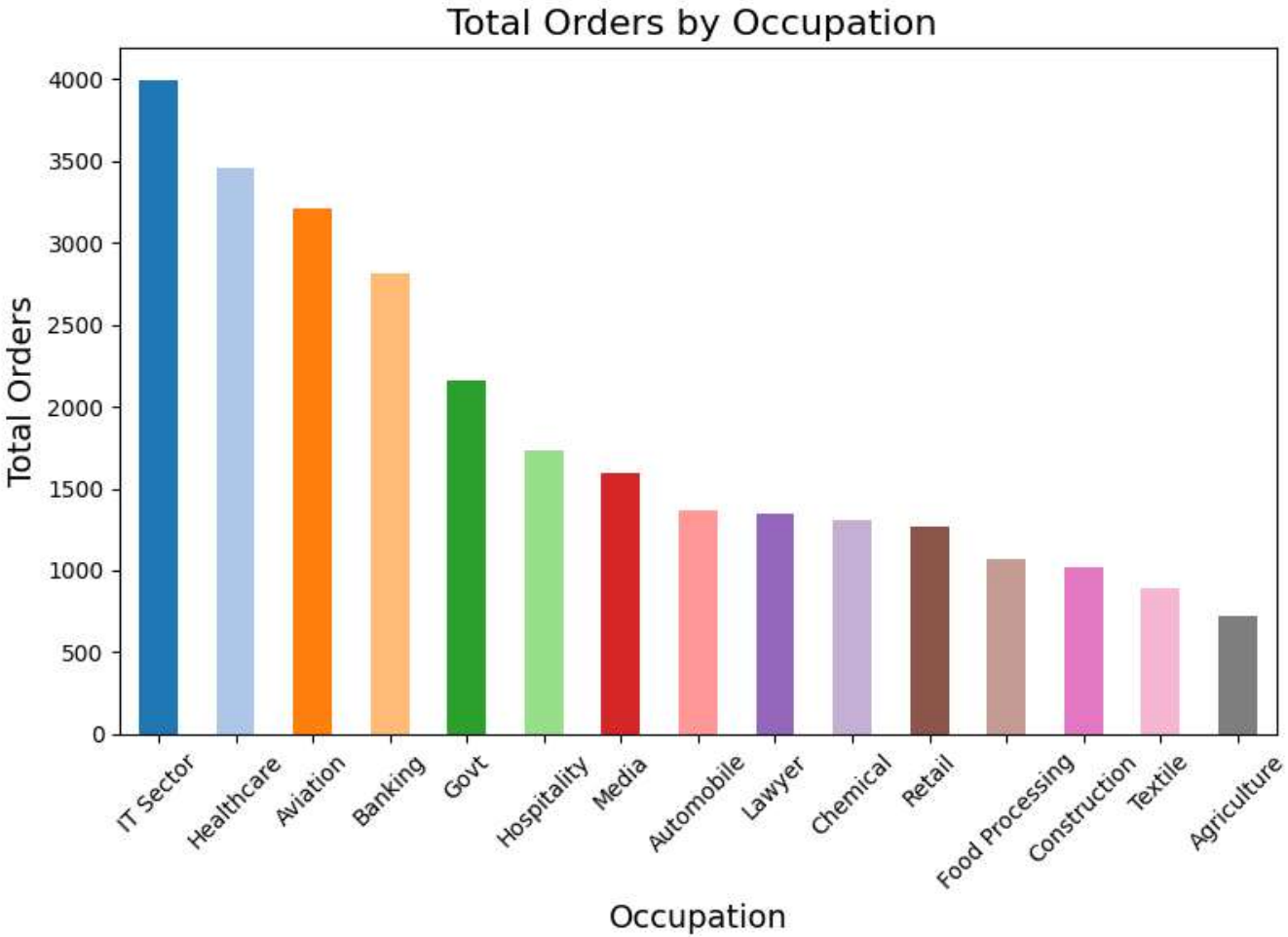

Amount Spent by Unmarried Men and Women
1e7
43,778,560
18,320,176
Total Amount Spent
Gender
F
M

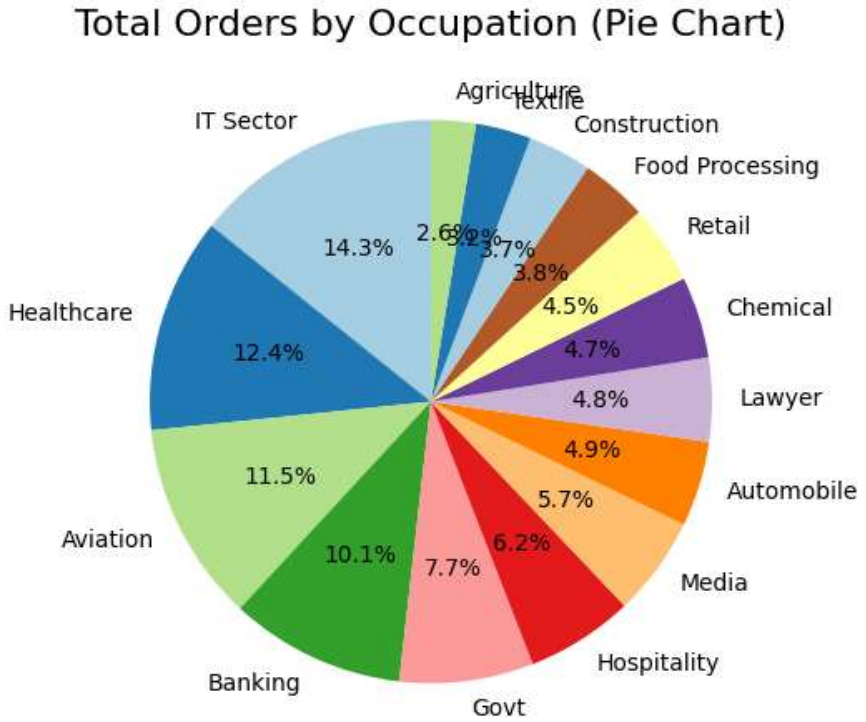

Age Distribution of Customers
Number of Customers
Age
0
500
1000
1500
2000
2500
3000
3500
10
20
30
40
50
60
70
80
90

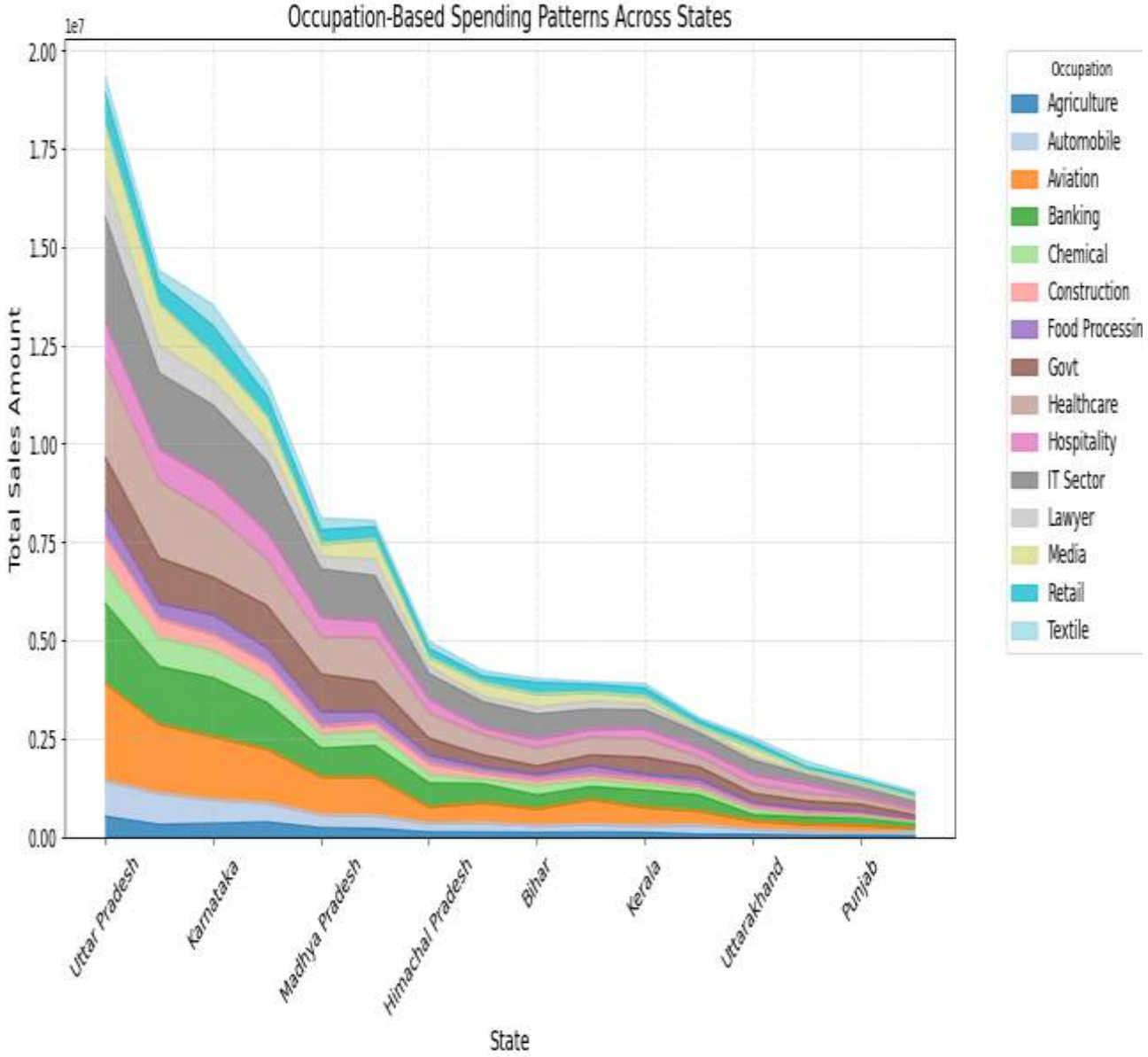

Age vs. Amount
Amount Spent
Age
Gender
F
M
0
5000
10000
15000
20000
25000
10
20
30
40
50
60
70
80
90

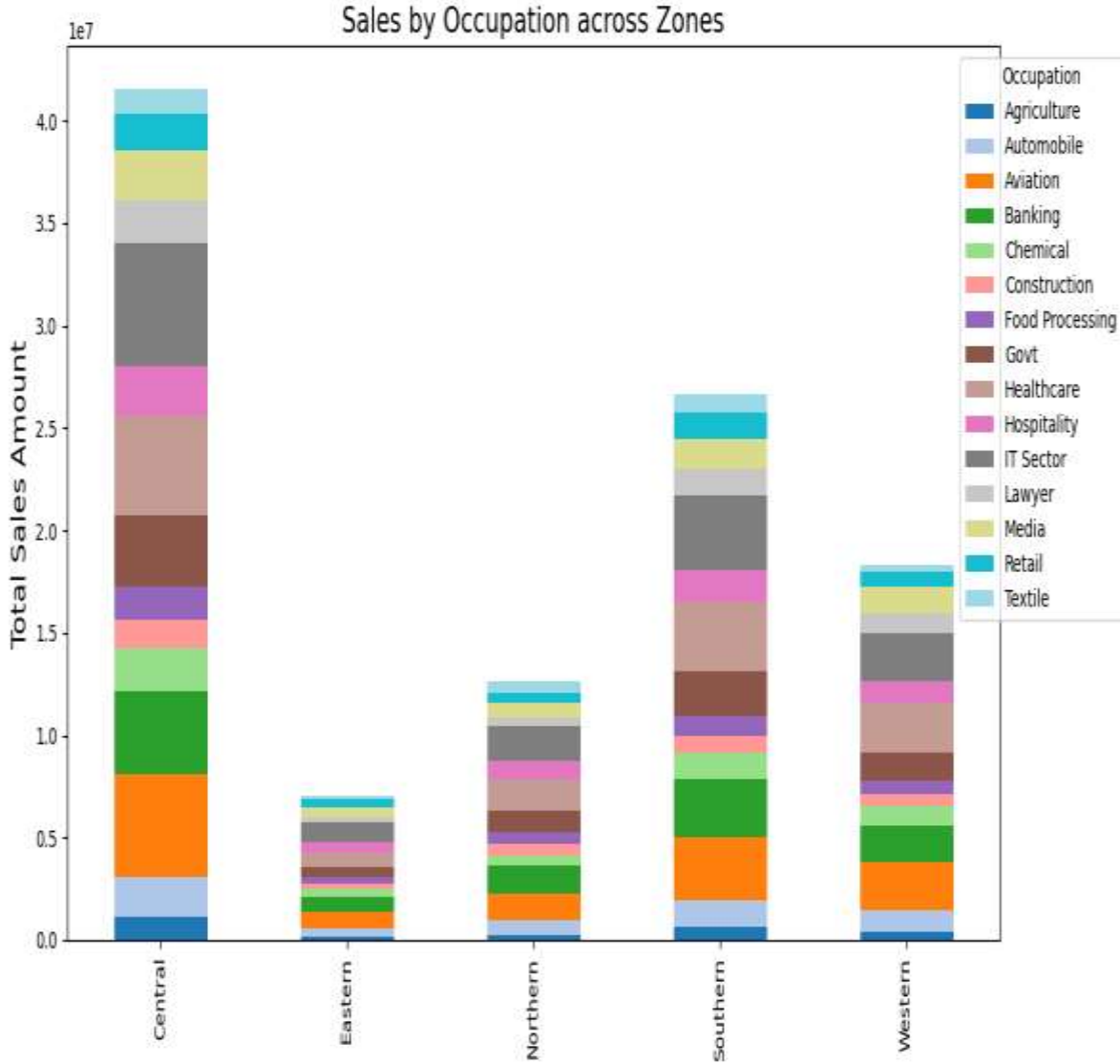

Total Sales by State
Telangana
Punjab
Rajasthan
Uttarakhand
Jharkhand
Kerala
Gujarat
Bihar
Haryana
Himachal Pradesh
Andhra Pradesh
Madhya Pradesh
Delhi
Karnataka
Maharashtra
Uttar Pradesh
State
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
Total Amount
1e7

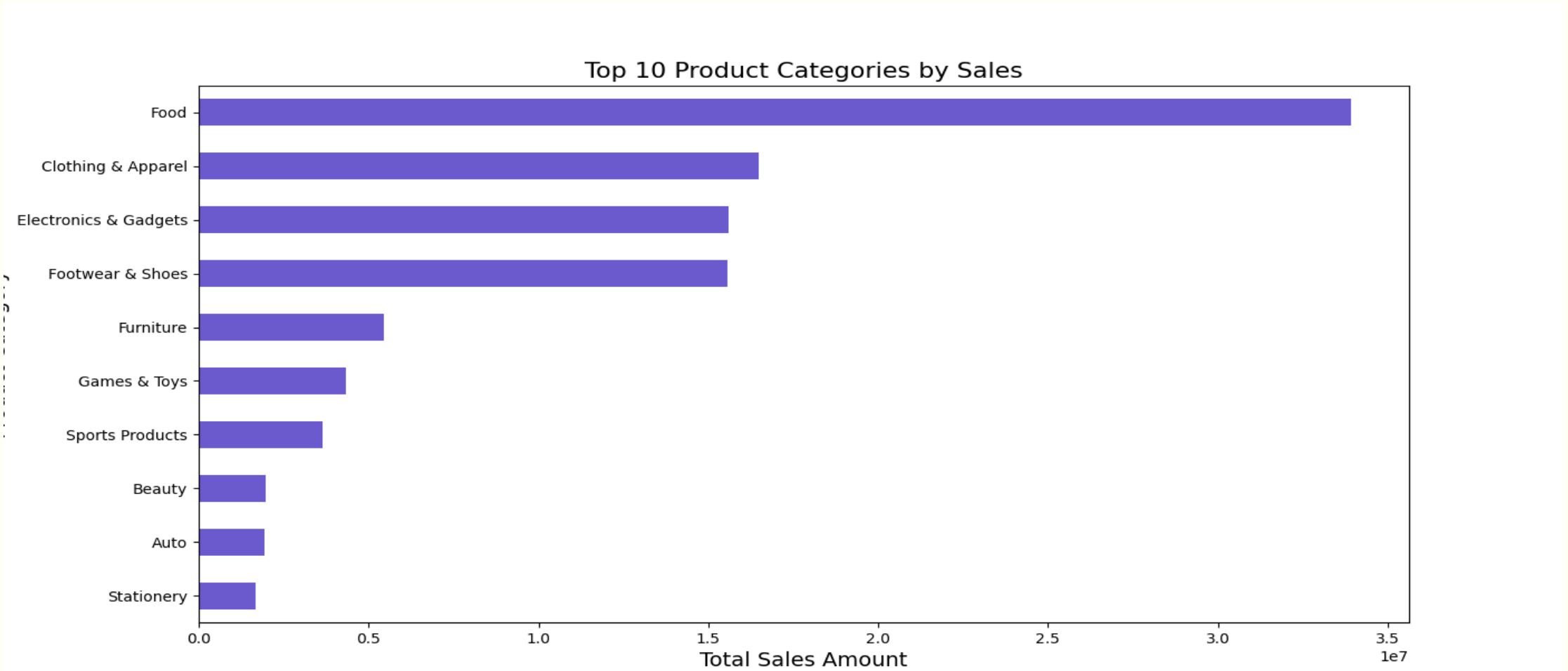

Total Amount Spent by Zone
Western
17.3%
Southern
25.1%
11.9%
Northern
6.6%
Eastern
39.2%
Central


Total Orders by Occupation
Total Orders
Occupation
0
500
1000
1500
2000
2500
3000
3500
4000
IT Sector
Healthcare
Aviation
Banking
Govt
Hospitality
Media
Automobile
Lawyer
Chemical
Retail
Food Processing
Construction
Textile
Agriculture


Total Orders by Occupation (Pie Chart)
IT Sector
Agriculture
Textile
Construction
Food Processing
Retail
Chemical
Lawyer
Automobile
Media
Hospitality
Govt
Banking
Aviation
Healthcare
14.3%
2.6%
3.2%
3.7%
3.8%
4.5%
4.7%
4.8%
4.9%
5.7%
6.2%
7.7%
10.1%
11.5%
12.4%


Occupation-Based Spending Patterns Across States
1e7
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Total Sales Amount
Uttar Pradesh
Karnataka
Madhya Pradesh
Himachal Pradesh
Bihar
Kerala
Uttarakhand
Punjab
State
Occupation
Agriculture
Automobile
Aviation
Banking
Chemical
Construction
Food Processin
Govt
Healthcare
Hospitality
IT Sector
Lawyer
Media
Retail
Textile


Sales by Occupation across Zones
1e7
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Total Sales Amount
Central
Eastern
Northern
Southern
Western
Occupation
Agriculture
Automobile
Aviation
Banking
Chemical
Construction
Food Processing
Govt
Healthcare
Hospitality
IT Sector
Lawyer
Media
Retail
Textile


Top 10 Product Categories by Sales
Food
Clothing & Apparel
Electronics & Gadgets
Footwear & Shoes
Furniture
Games & Toys
Sports Products
Beauty
Auto
Stationery
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
1e7
Total Sales Amount

conclusion

with this data analysis business decisions for Diwali sales may tend to target marketing which focuses on females who are between 25 to 35 years old working in IT sector and belonging to central zone of the country .

Thank You