



Data Fireworks: Uncovering Diwali Sales Insights with Python



By Prerna Vashisht



Agenda



DESCRIPTION



APPROACH



TECH-STACK



INSIGHTS



CONCLUSION



DESCRIPTION

In this project, I analyze Diwali sales data using Python to uncover key trends and insights. Utilizing powerful libraries like Pandas, Matplotlib, and Seaborn, I process and visualize the data to understand consumer behavior, identify top-selling products, and evaluate sales performance during the festive season. My analysis includes identifying potential customers across different states, occupations, genders, and age groups, with a focus on understanding which groups have shopped more, comparing male and female customer behavior, and performing extensive exploratory data analysis (EDA). This comprehensive analysis provides valuable insights for strategic decision-making and optimizing future Diwali sales campaigns.



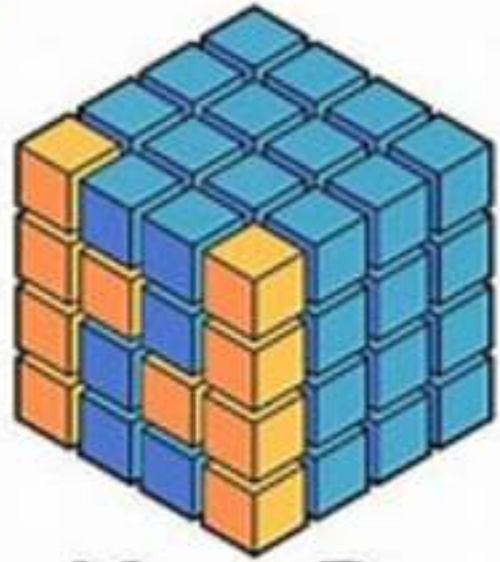


APPROACH

In this project, I dive into Diwali sales data using Python to uncover meaningful insights and trends. I begin by gathering and loading the data into a Pandas DataFrame, then clean it to address any missing values and inconsistencies. With the data ready, I use Pandas, Matplotlib, and Seaborn to perform exploratory data analysis (EDA), visualizing trends and patterns. This includes examining sales across different states, occupations, genders, and age groups, and identifying top-selling products. I focus on understanding customer behavior, comparing spending patterns between male and female shoppers, and analyzing different demographic groups. By evaluating overall sales performance and the effectiveness of marketing efforts, I provide actionable insights and recommendations to enhance future Diwali sales strategies. The analysis wraps up with key findings and suggestions for optimizing sales and identifying new growth opportunities.



TECH-STACK



NumPy

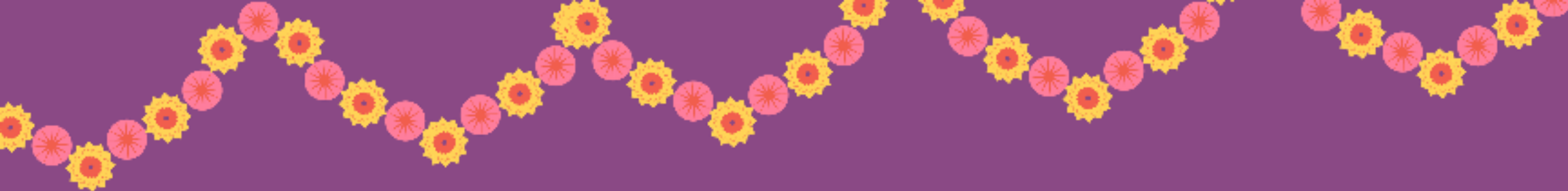


Pandas



INSIGHTS





Data Cleaning and Processing

The first step in analyzing Diwali sales data is to import the necessary Python libraries, including Pandas, Matplotlib, and Seaborn. Once the libraries are imported, the next step is to load the sales data into a Pandas DataFrame. After loading the data, it's crucial to inspect the dataset for any inconsistencies, null values, or NaN values. These issues are addressed through data cleaning and processing, which may involve dropping rows with null values using `dropna()` or filling in missing data using appropriate techniques. This ensures the dataset is clean and ready for further analysis.



Data Cleaning and Processing

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

```
#check for null values
pd.isnull(df).sum()
```

```
# import csv file
df = pd.read_csv('Diwali Sales Data.csv', encoding= 'unicode_escape')
```

```
# drop null values
df.dropna(inplace=True)
```

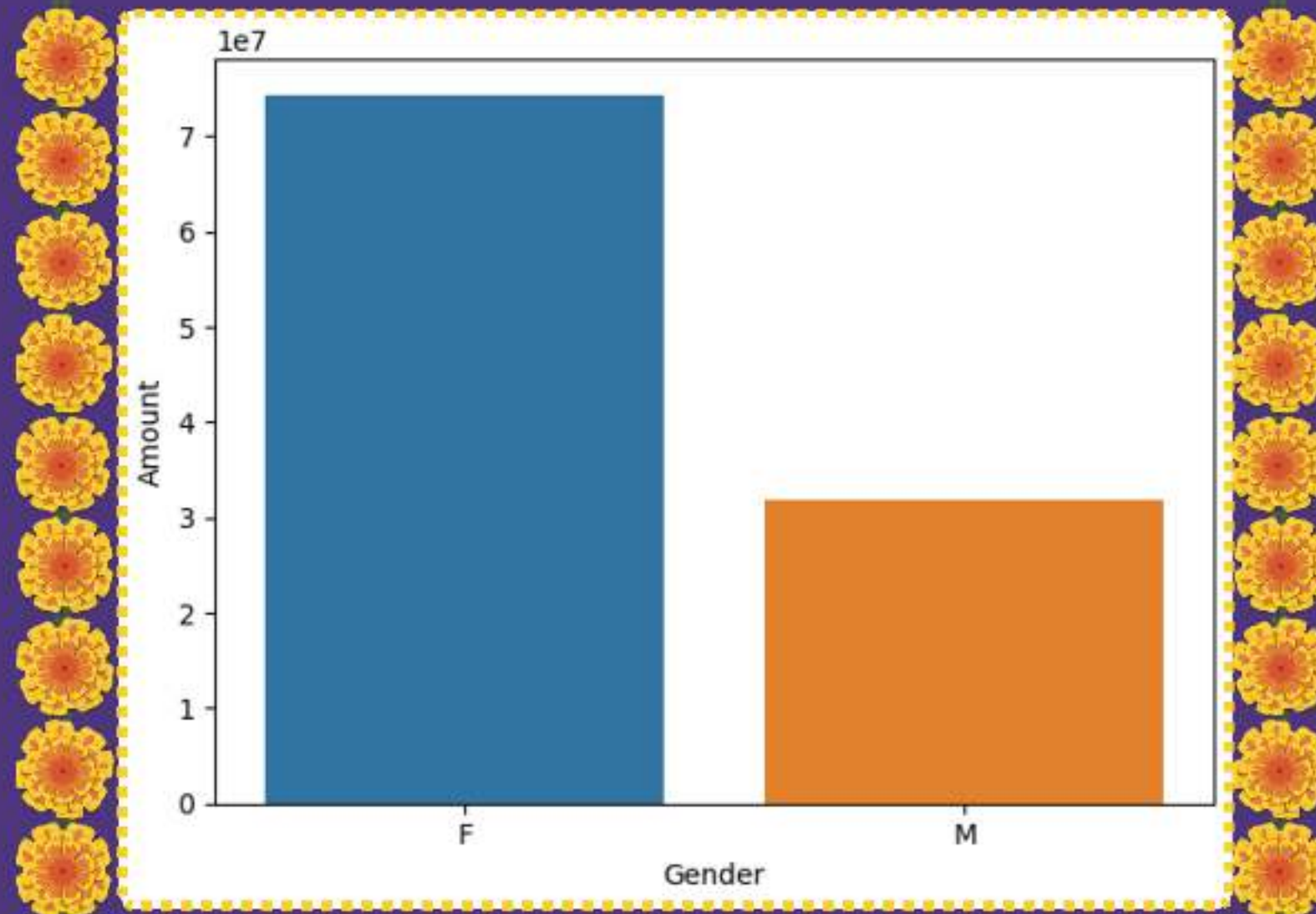
```
# change data type
df['Amount'] = df['Amount'].astype('int')
```


Gender-Based Spending Analysis

```
ax = sns.countplot(x = 'Gender', data = df)

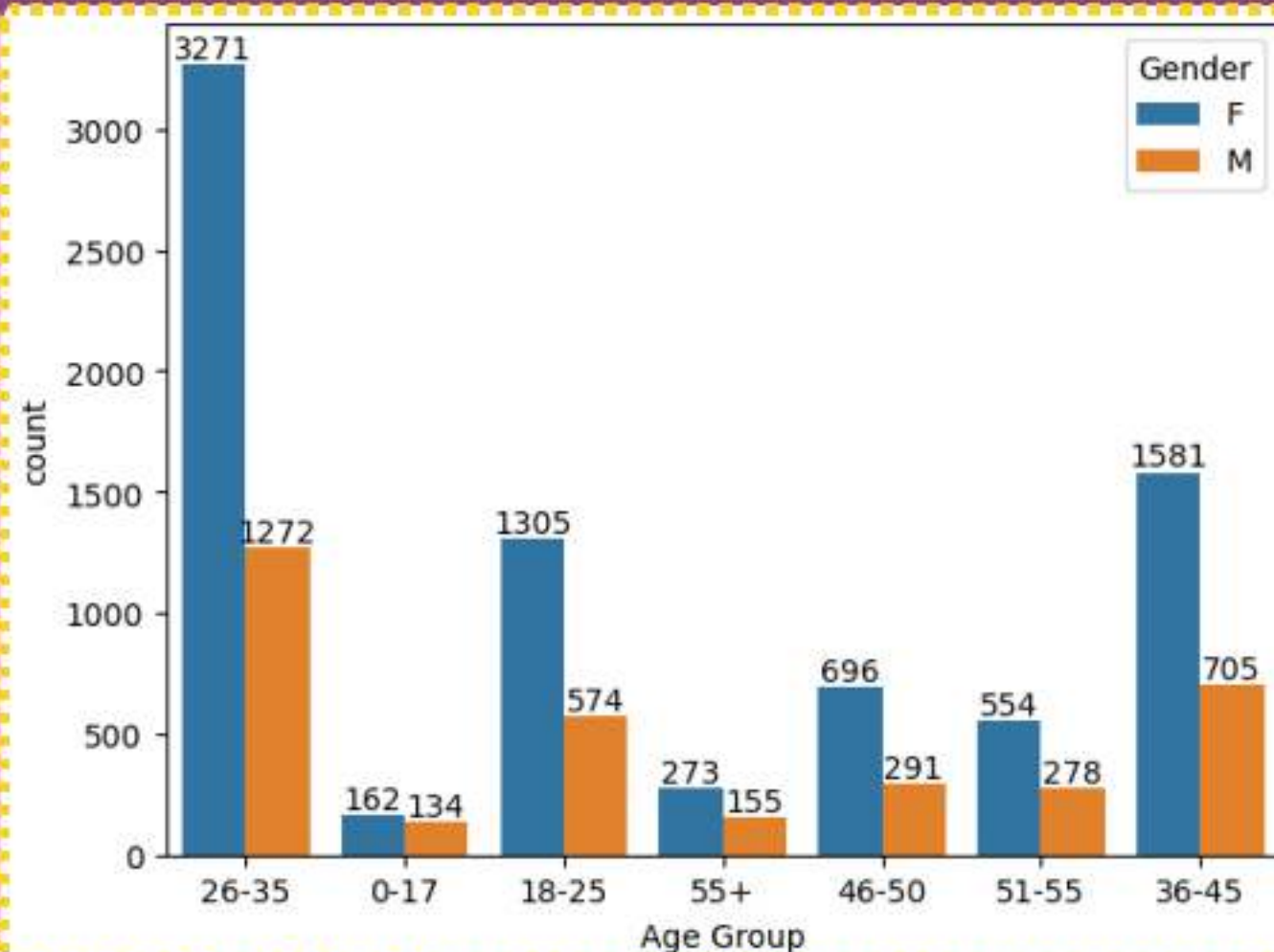
for bars in ax.containers:
    ax.bar_label(bars)
```

```
sales_gen = df.groupby(['Gender'], as_index=False)
['Amount'].sum().sort_values(by='Amount', ascending=False)
sns.barplot(x = 'Gender', y = 'Amount', data = sales_gen)
```

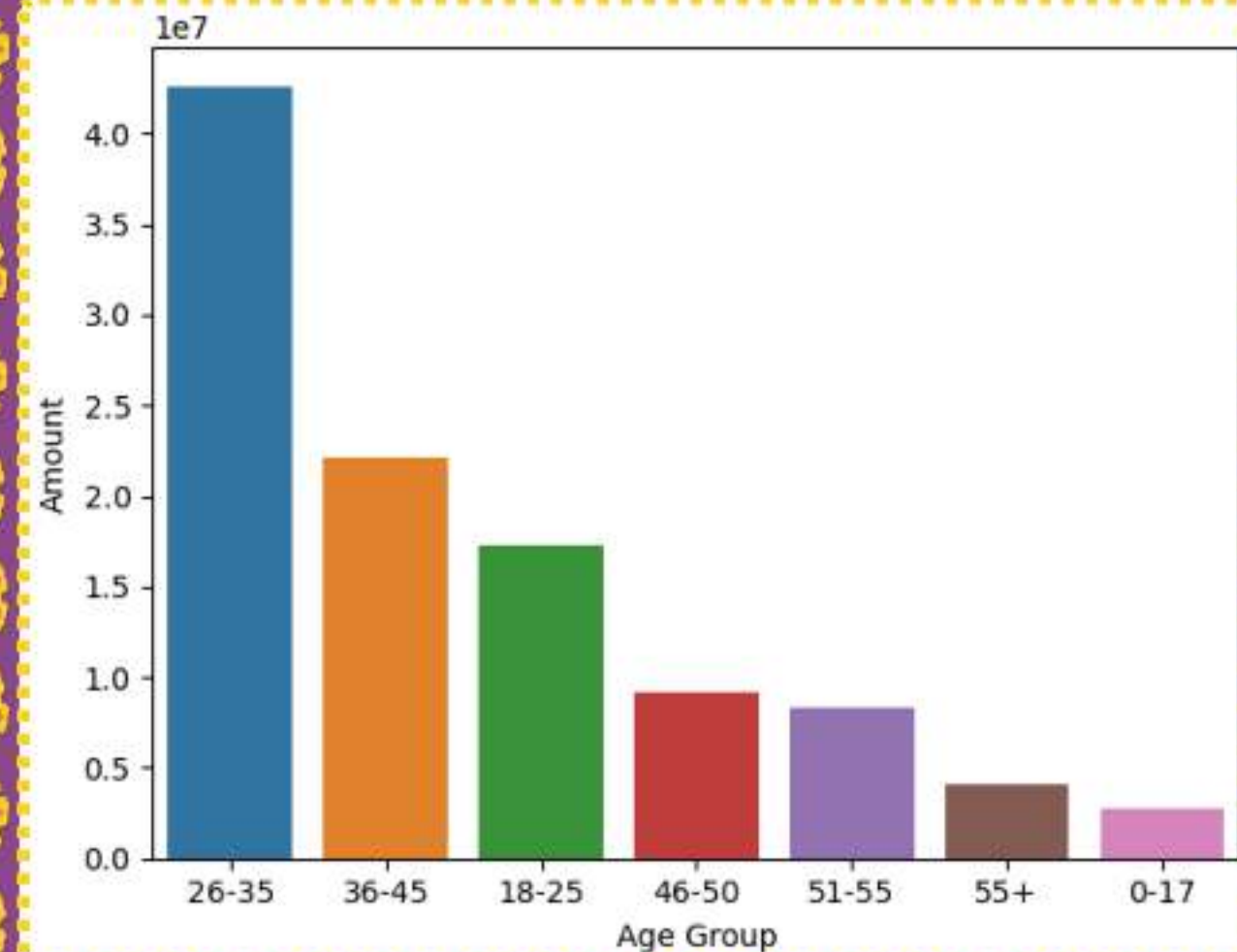


Age Group-Based Spending Analysis

```
ax = sns.countplot(data = df, x = 'Age Group', hue = 'Gender')  
  
for bars in ax.containers:  
    ax.bar_label(bars)
```



```
sales_age = df.groupby(['Age Group'], as_index=False)  
['Amount'].sum().sort_values(by='Amount', ascending=False)  
sns.barplot(x = 'Age Group', y = 'Amount', data = sales_age, hue = 'Age Group')
```

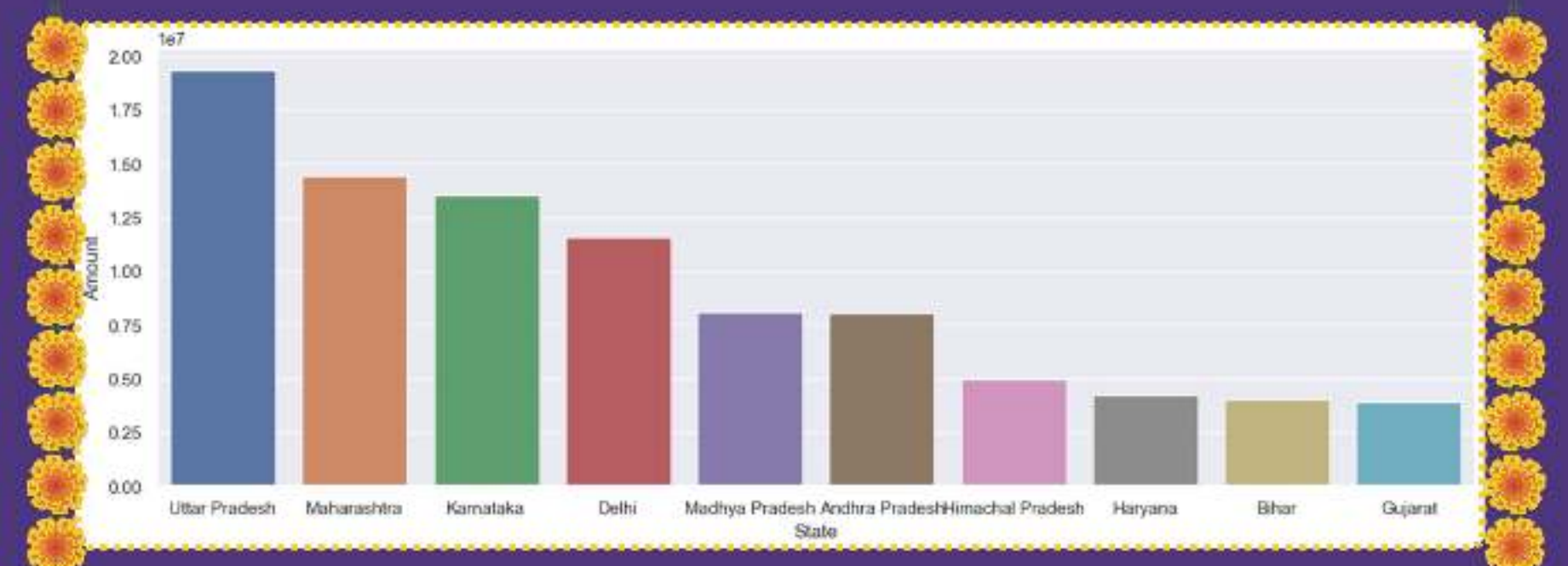


Top 10 States by Total Number of Sales

```
sales_state = df.groupby(['State'], as_index=False)
['Orders'].sum().sort_values(by='Orders', ascending=False).head(10)
sns.set(rc={'figure.figsize':(15,5)})
sns.barplot(data = sales_state, x = 'State',y= 'Orders')
```

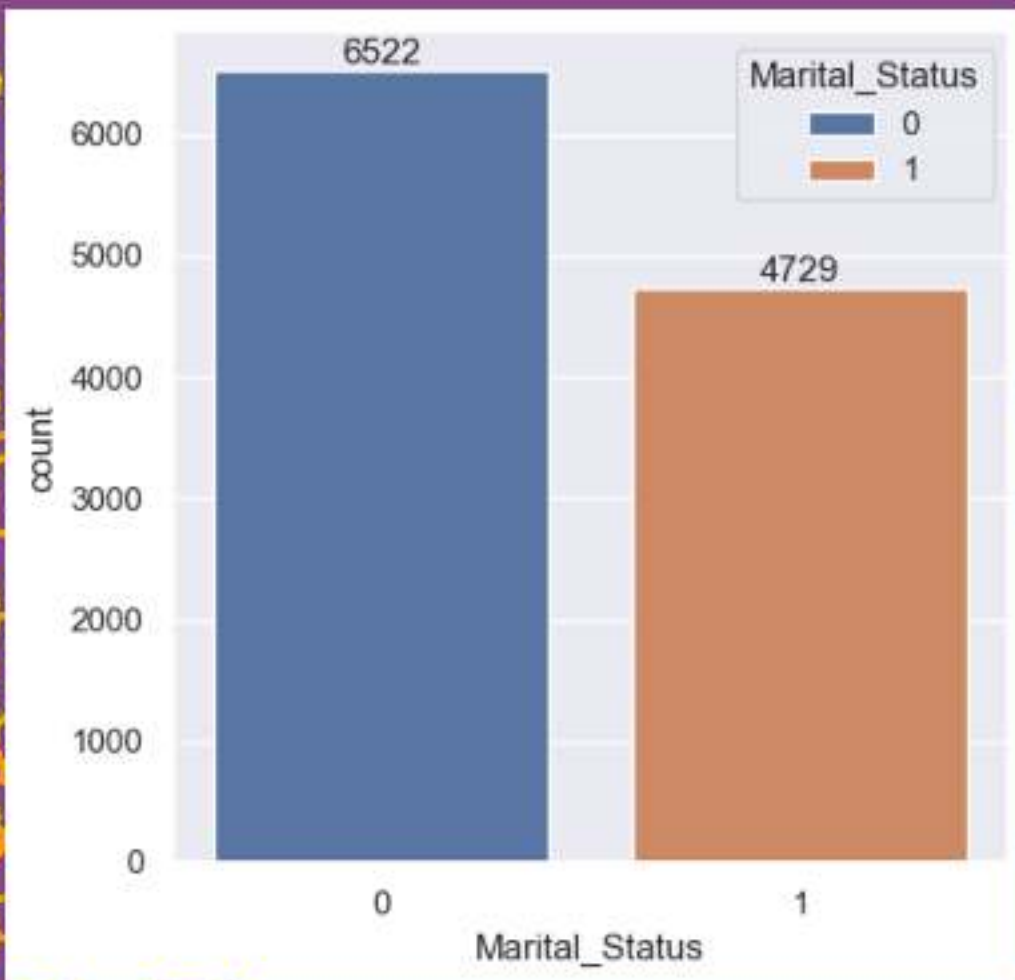


```
sales_state = df.groupby(['State'], as_index=False)
['Amount'].sum().sort_values(by='Amount', ascending=False).head(10)
sns.set(rc={'figure.figsize':(15,5)})
sns.barplot(data = sales_state, x = 'State',y= 'Amount', hue='State')
```

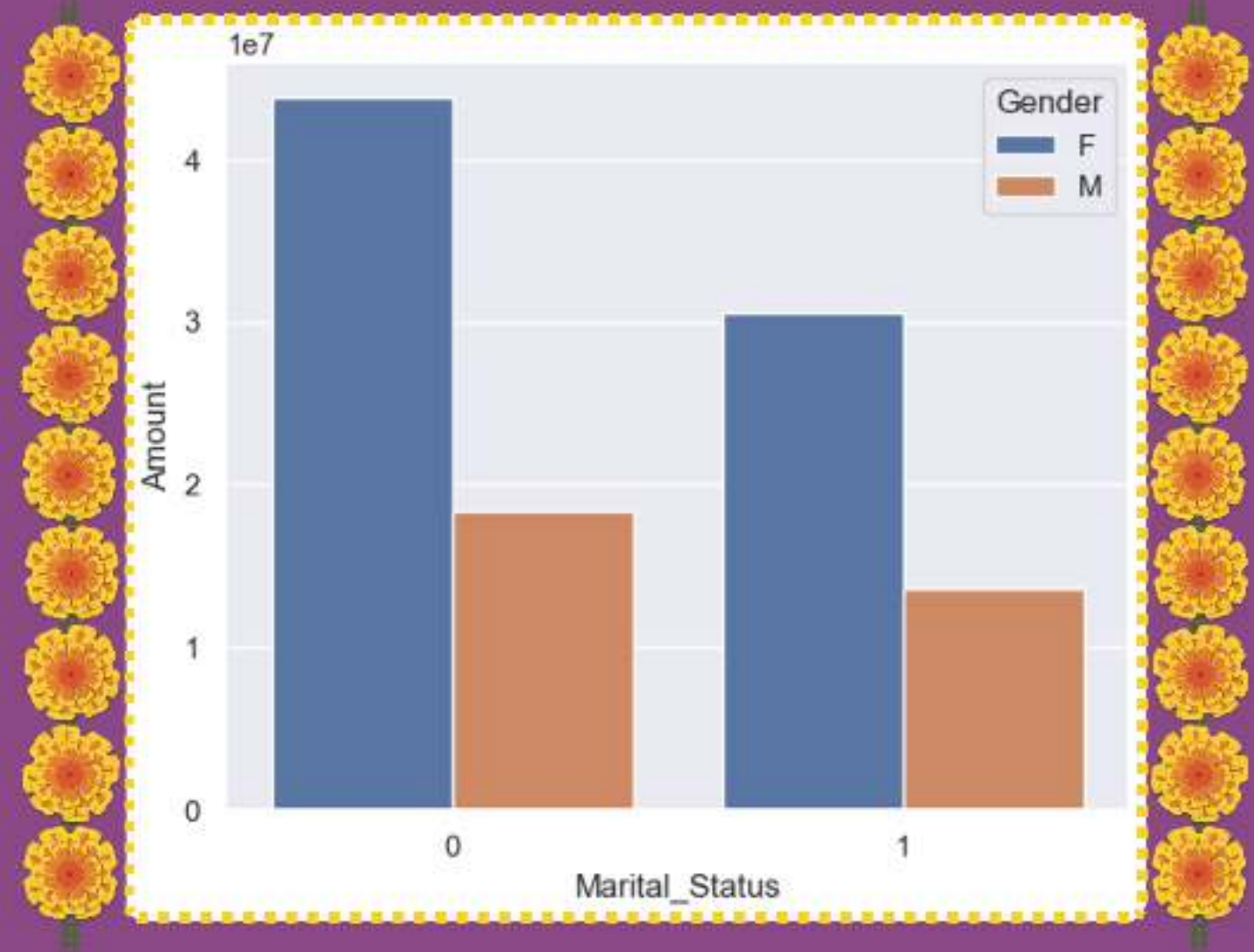


Marital Status-Based Spending Analysis

```
ax = sns.countplot(data = df, x = 'Marital_Status')
sns.set(rc={'figure.figsize':(7,5)})
for bars in ax.containers:
    ax.bar_label(bars)
```



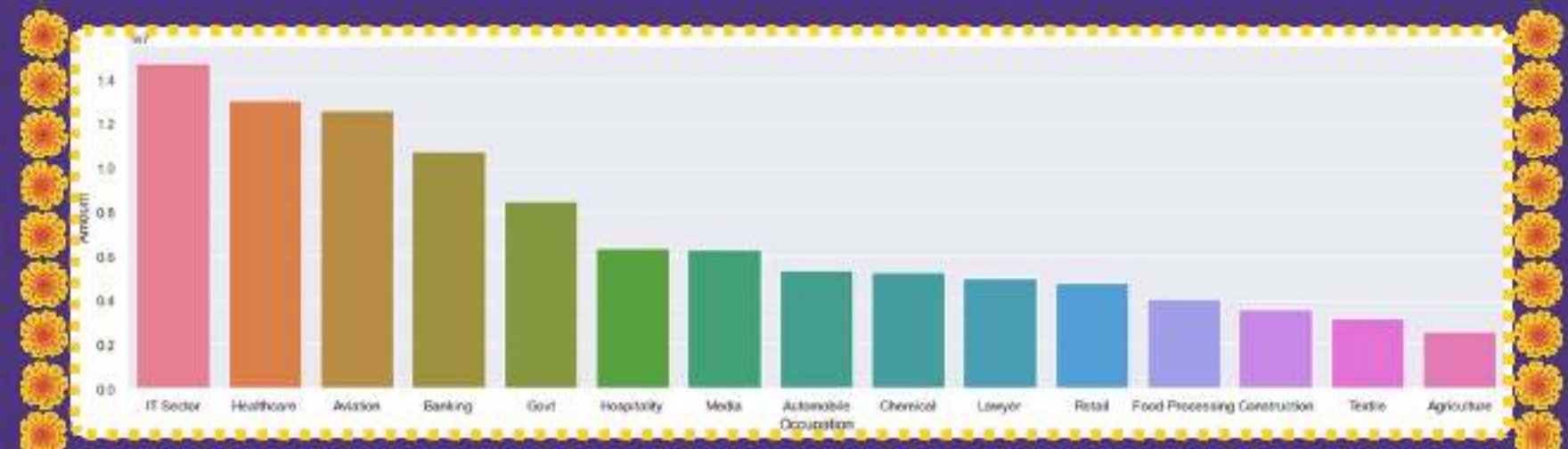
```
sales_state = df.groupby(['Marital_Status', 'Gender'], as_index=False)
['Amount'].sum().sort_values(by='Amount', ascending=False)
sns.set(rc={'figure.figsize':(6,5)})
sns.barplot(data = sales_state, x = 'Marital_Status', y = 'Amount', hue='Gender')
```



Profession-Based Spending Analysis

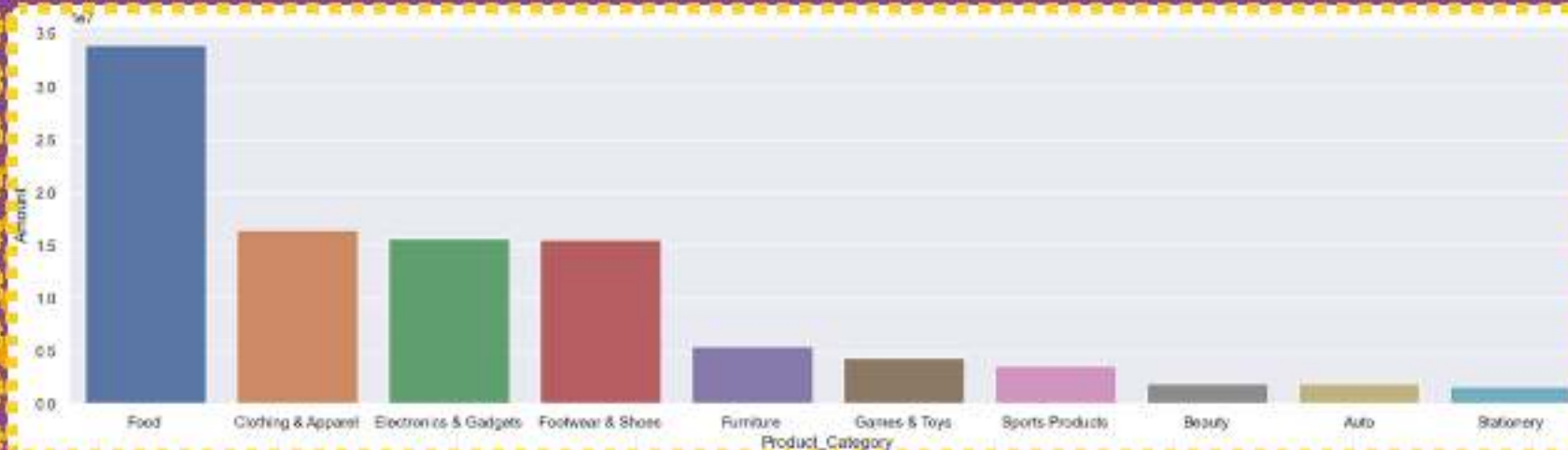
```
sns.set(rc={'figure.figsize':(20,5)})  
ax = sns.countplot(data = df, x = 'Occupation')  
for bars in ax.containers:  
    ax.bar_label(bars)
```

```
sales_state = df.groupby(['Occupation'], as_index=False)  
['Amount'].sum().sort_values(by='Amount', ascending=False)  
sns.set(rc={'figure.figsize':(20,5)})  
sns.barplot(data = sales_state, x = 'Occupation', y = 'Amount')
```

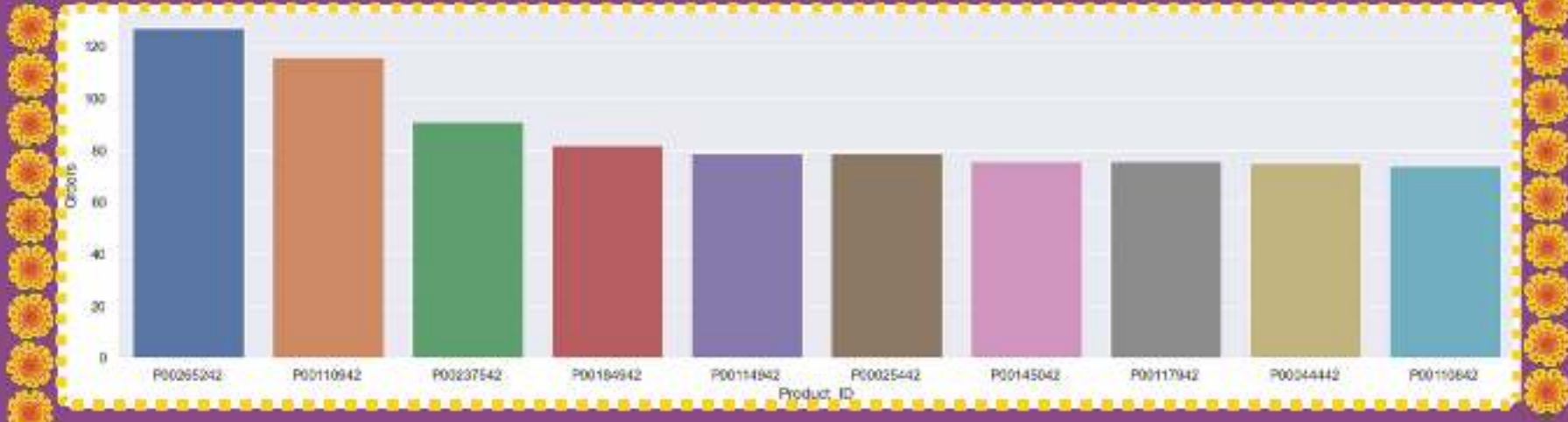


Product Category-Based Spending Analysis

```
sales_state = df.groupby(['Product_Category'], as_index=False)
['Amount'].sum().sort_values(by='Amount', ascending=False).head(10)
sns.set(rc={'figure.figsize':(20,5)})
sns.barplot(data = sales_state, x = 'Product_Category',y= 'Amount')
```



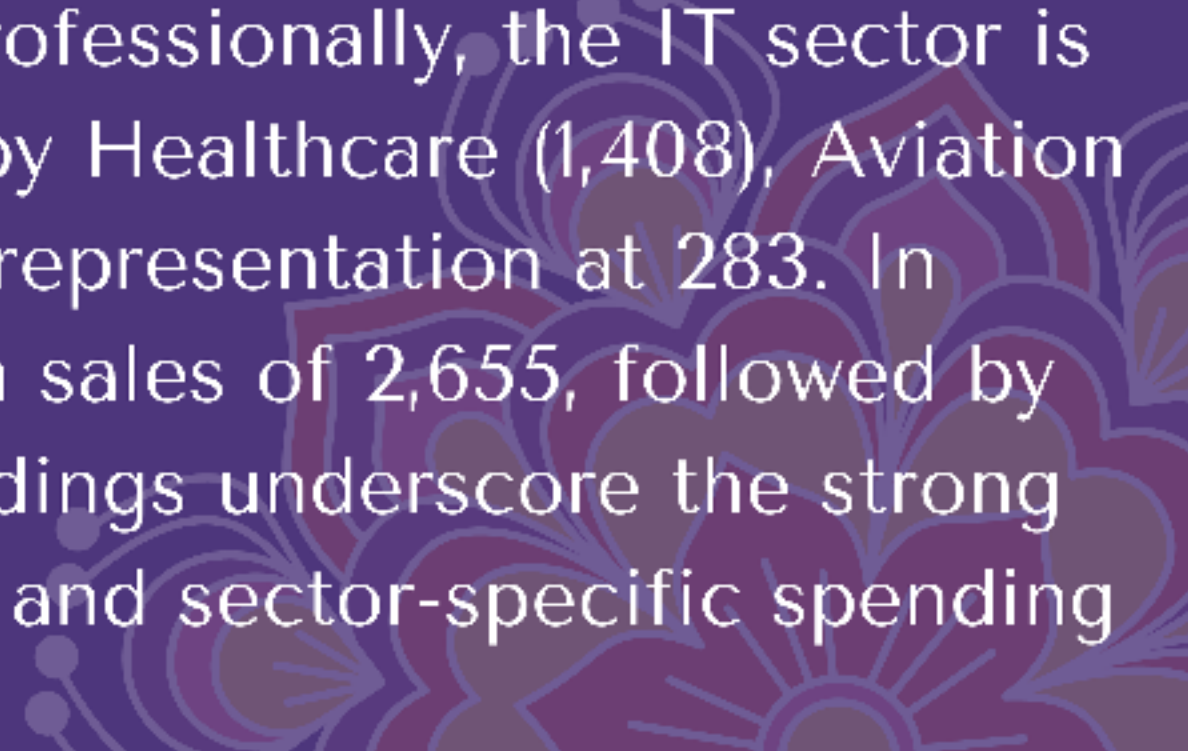
```
sales_state = df.groupby(['Product_ID'], as_index=False)
['Orders'].sum().sort_values(by='Orders', ascending=False).head(10)
sns.set(rc={'figure.figsize':(20,5)})
sns.barplot(data = sales_state, x = 'Product_ID',y= 'Orders')
```



A decorative border at the top left of the slide, featuring a string of small, colorful flowers (yellow, orange, and pink) arranged in a wavy line.

CONCLUSION

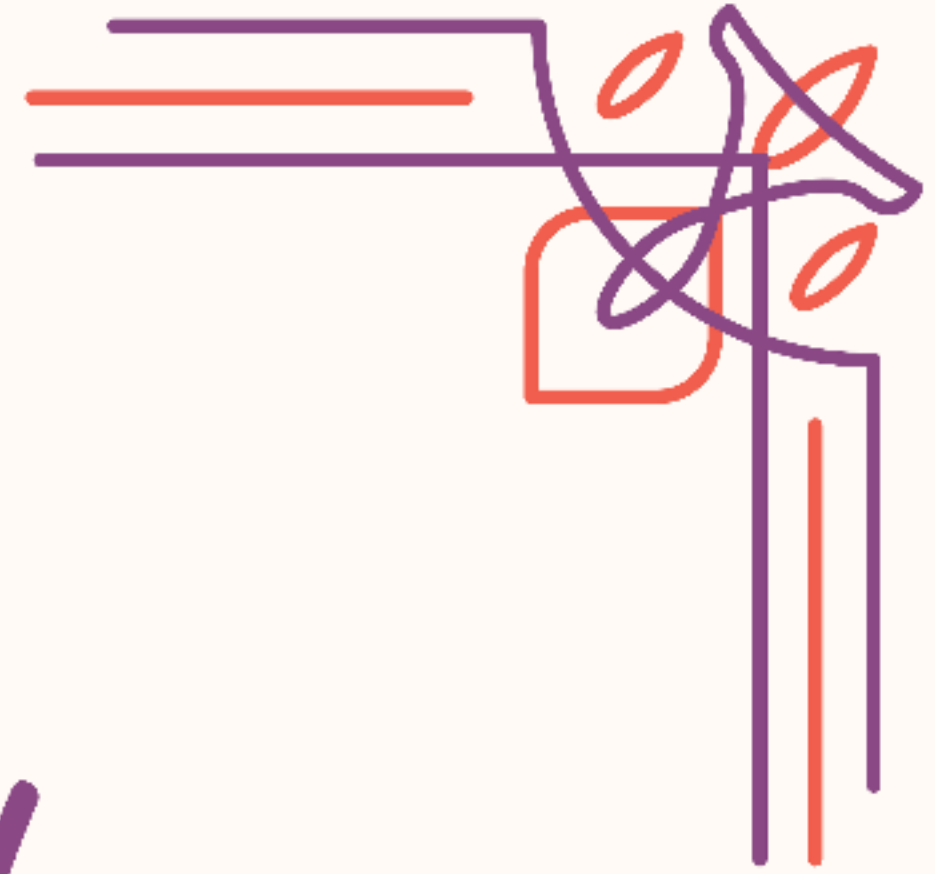
The Diwali sales analysis reveals that females are the predominant buyers, with 7,832 females demonstrating higher purchasing power compared to 3,407 males. In the 26-35 age group, females (3,269) outnumber males (1,272). Sales data shows Uttar Pradesh in the Central Zone leading in the number of orders, followed by Maharashtra, Karnataka, and Delhi, while Gujarat and Haryana from the Western and Northern Zones have the lowest order volumes. Marital status analysis indicates that married women are the highest spenders, totaling 6,518, compared to 4,721 spent by men. Professionally, the IT sector is the largest contributor to sales with 1,583 buyers, followed by Healthcare (1,408), Aviation (1,310), and Banking (1,137), with Agriculture having the least representation at 283. In terms of product categories, Clothing and Apparel lead with sales of 2,655, followed by Food (2,490) and Electronics and Gadgets (2,087). These findings underscore the strong female purchasing power, varied regional sales distribution, and sector-specific spending patterns.

A large, intricate, light-colored floral pattern in the bottom right corner, featuring multiple layers of petals and a central circular motif.

Based on the Diwali sales analysis, several strategic actions can enhance future sales. Focus marketing campaigns on female buyers and the 26-35 age group. Boost sales in low-performing regions like Gujarat and Haryana, and strengthen efforts in high-performing states such as Uttar Pradesh, Maharashtra, Karnataka, and Delhi. Promote offers tailored for married women and men. Create exclusive deals for IT and healthcare professionals, and attract buyers from less represented sectors like agriculture. Continue promoting clothing, apparel, and food products while launching new electronics and gadgets lines. Utilize seasonal promotions during peak periods like Diwali. Conduct customer surveys for feedback and develop engagement initiatives like loyalty programs to foster brand loyalty and drive repeat purchases. These strategies will target key consumer segments and enhance regional sales for future growth.

Strategic Action Plan





Thank you!

