Diwali sales data analysis

Project Overview

Objective: *The primary objective of this project is to analyze the Diwali sales data of an e-commerce company to solve various problem statements ,uncover patterns, trends, and taking data-driven decisions to improve future sales strategies. By leveraging python and its data analysis libraries like Pandas , Numpy, Matplotlib, Seaborn I aim to provide a detailed examination of sales performance during the Diwali period.*

Problem Statements:

i. An*alyse the distribution of buyers by age and gender to identify the primary demographic group making purchases during Diwali.*

ii. Determine which states have the highest number of orders and total sales to understand regional sales performance during Diwali.

iii. Examine the marital status of buyers to identify if there is a significant trend.

iv. Analyse the top selling sub categories of the products to optimize the inventory.

v. Identify the sectors to which buyers belong to tailor marketing strategies and for better understanding of buyers.

Tools and Technologies:

✓*Python:* *Primary Programming Language used for data processing and analysis*

✓*Pandas:* *Data Manipulation (loading, cleaning and transforming data).*

✓*Matplotlib And Seaborn:* *Data Visualization , for creating visuals to represent data insights clearly.*

Methodology

- ✓ *Loading data: imported the dataset into Jupyter Notebook.*
- ✓ *Data cleaning: checked and handled null values, removed blank columns, and addressed inconsistencies.*
- ✓ *Exploratory data analysis(EDA): Conducted to understand data distribution and identify patterns.*
- ✓ *Visualization: created various plots like count plot, bar plot, etc. to visually represent data.*
- ✓ *Insights and recommendations: summarized findings and provided actionable business recommendations.*

Data Loading

- *Importing python libraries like numpy for numerical operations and efficient computation, pandas for data loading, cleaning , Matplotlib and seaborn for data visualization. into Jupyter Notebook*

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

- *Importing data set into Jupyter Notebook*

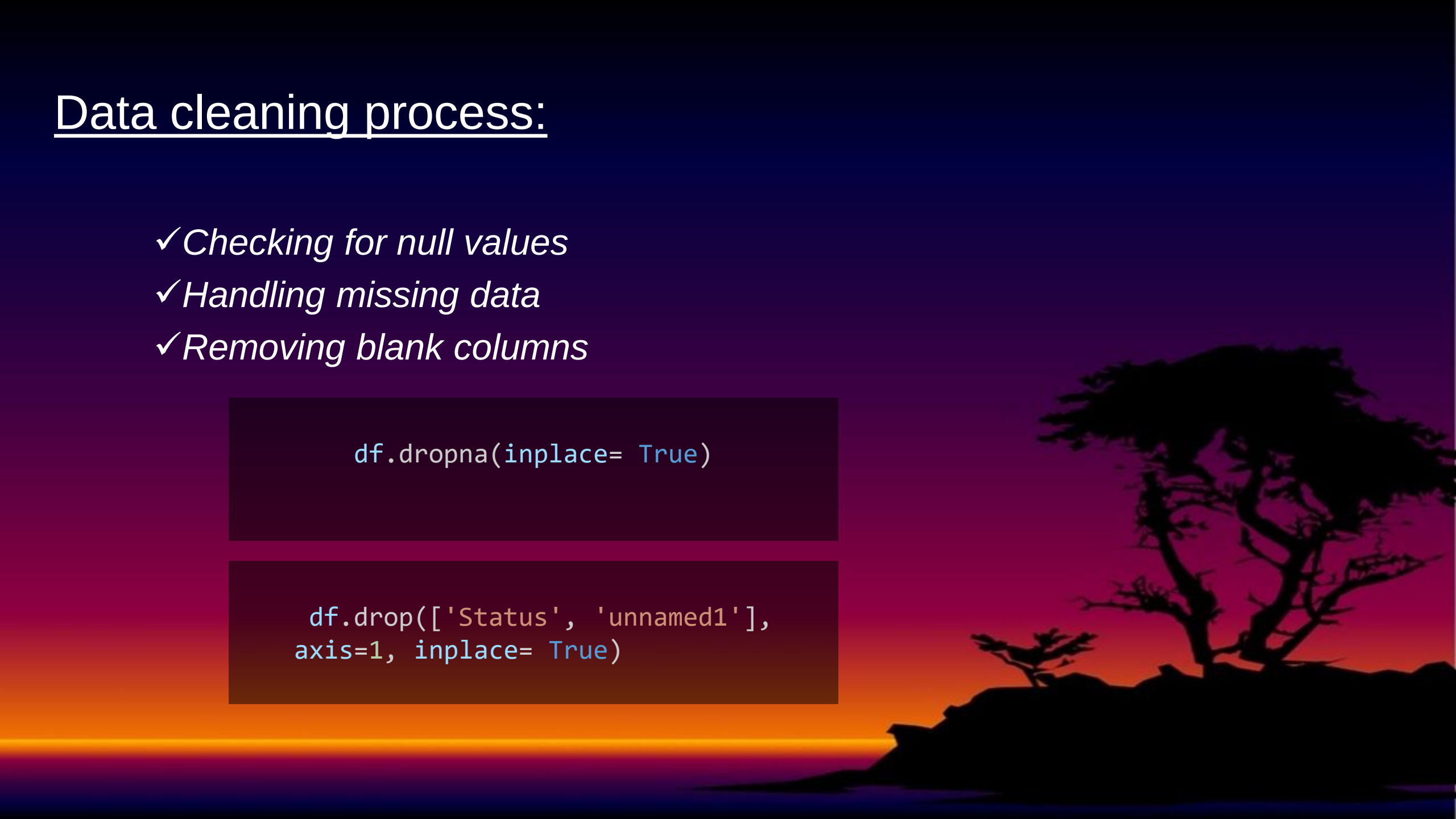
```
df=
pd.read_csv("C:/Users/mdami/Downloads/Python_Diwali_Sales_Analysis
-main/Python_Diwali_Sales_Analysis-main/Diwali Sales
Data.csv",encoding='unicode escape’)
```

Data cleaning process:

- ✓*Checking for null values*
- ✓*Handling missing data*
- ✓*Removing blank columns*

```
df.dropna(inplace= True)
```

```
df.drop(['Status', 'unnamed1'],
axis=1, inplace= True)
```

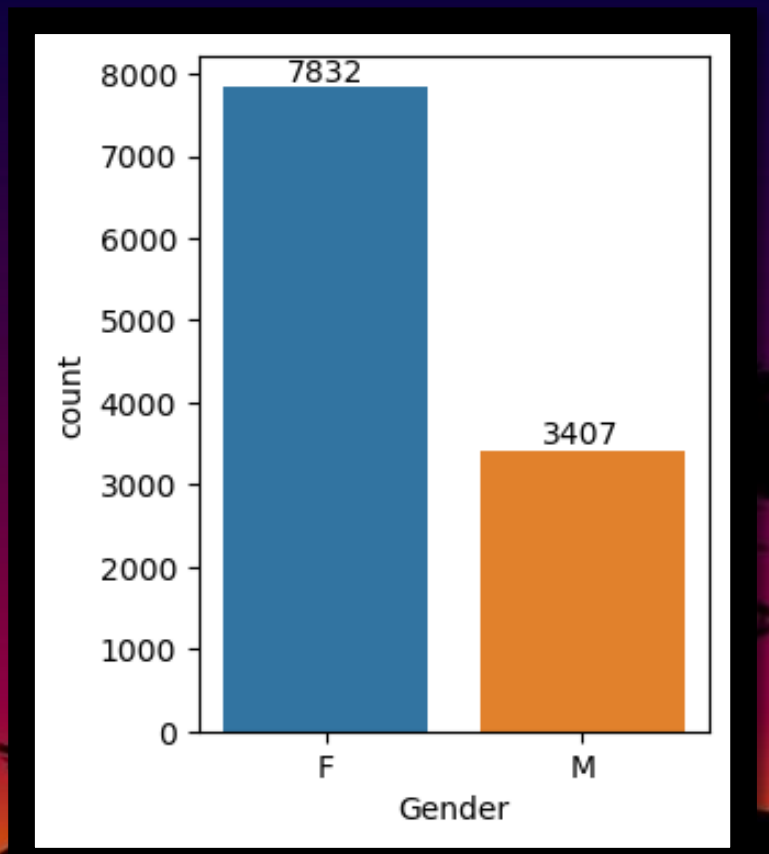


Exploratory Data Analysis

- *Age and gender distribution of buyers*

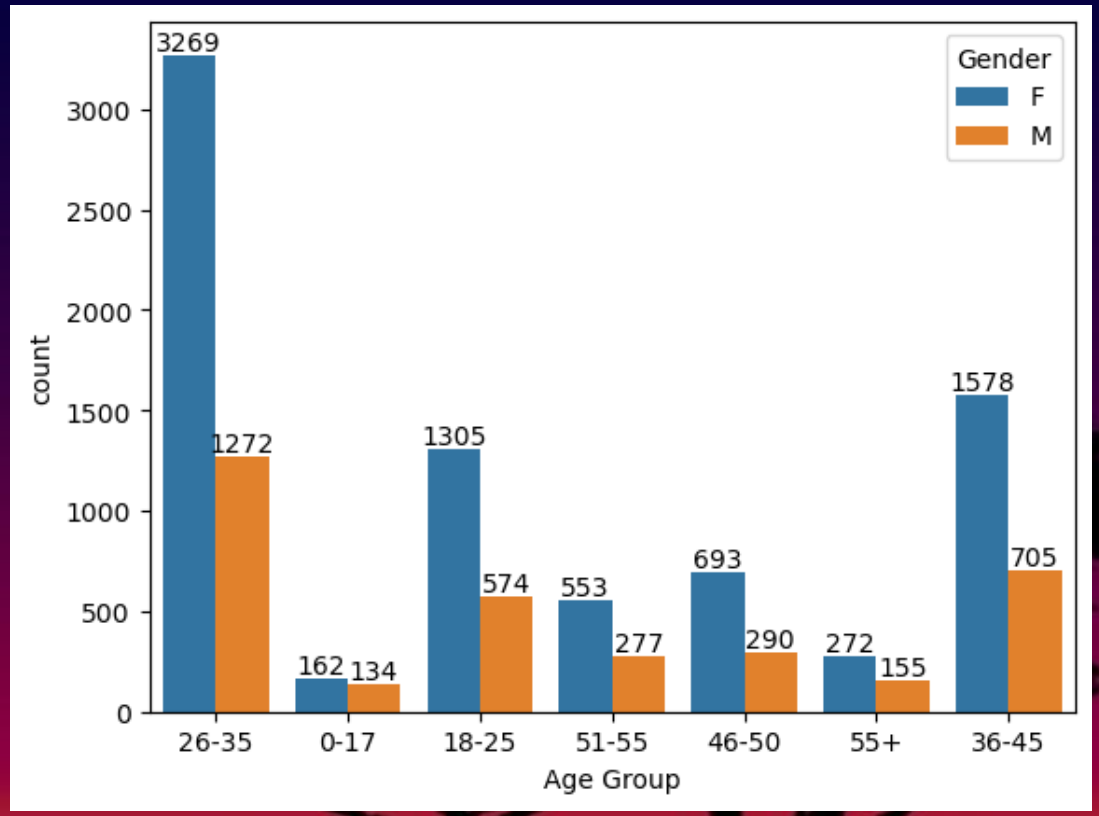
```
plt.figure(figsize=(3,4))

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

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

```
# From above graph we can see that women
have bought almost 2 times more than men.
```



➢ Count of buyers of different Age group

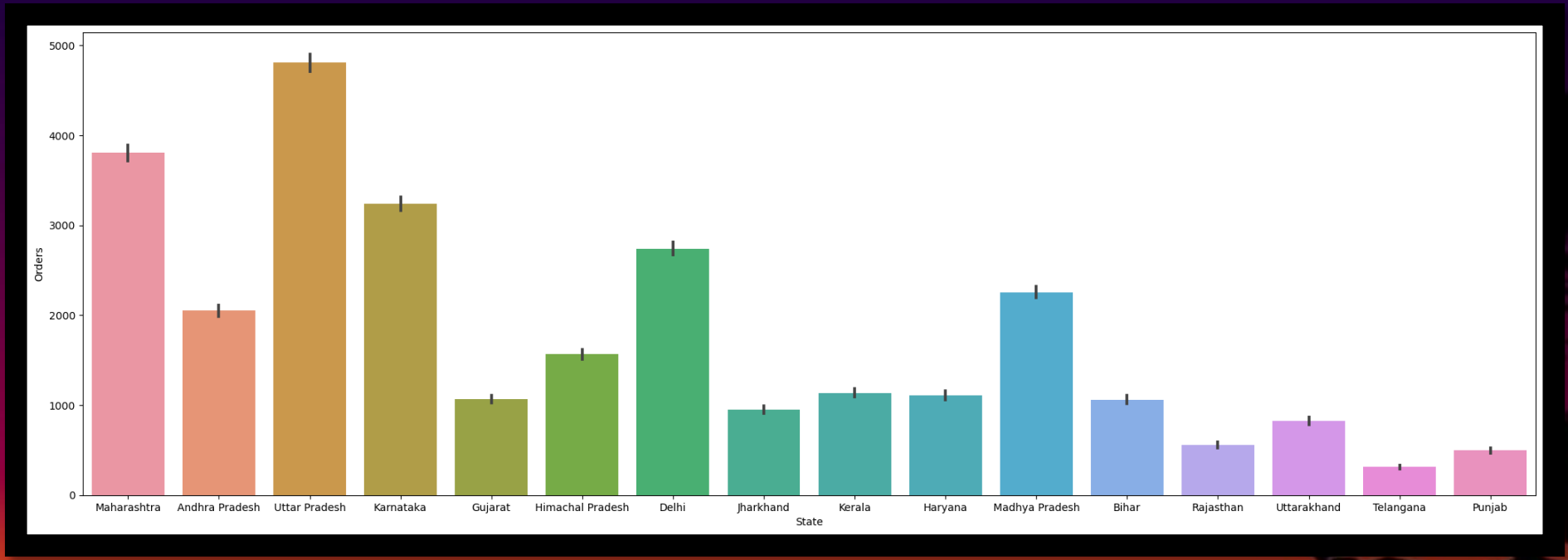
```
ax= sns.countplot(x='Age Group',
data= df, hue='Gender')
for bars in ax.containers:
    ax.bar_label(bars)
```




```
# From above graph we can see that most no of buyers are of age group 26-35 and those are women.
```

State-wise Orders and Total Sales

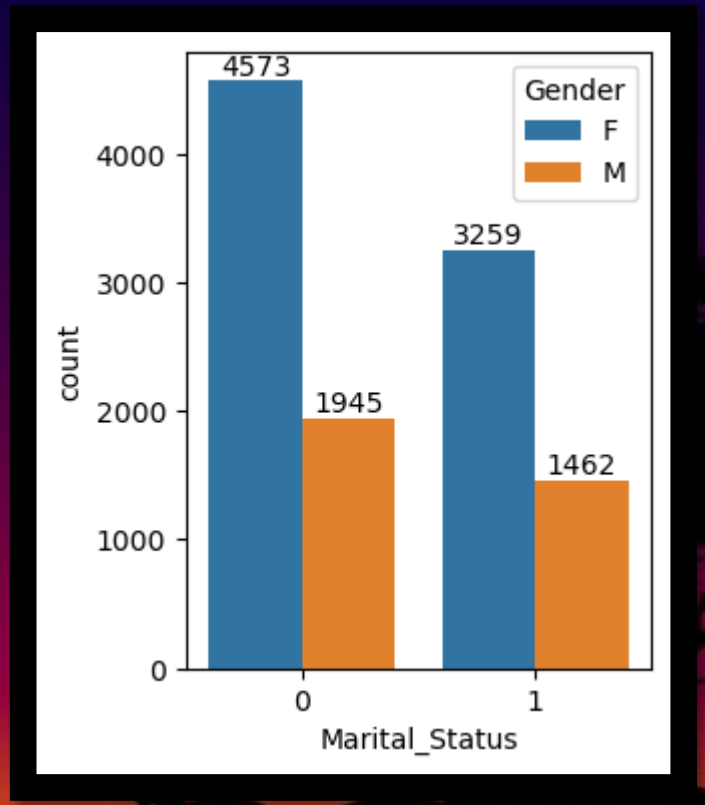
```
plt.figure(figsize=(25,8))
sns.barplot(x='State',y='Orders’,
estimator= 'sum',data=df)
```




```
# From the above graphs we can see that most of the orders and total sales
are from Uttar Pradesh, Maharashtra and Karnataka respectively.
```

➢ *Marital status Distribution of Buyers*

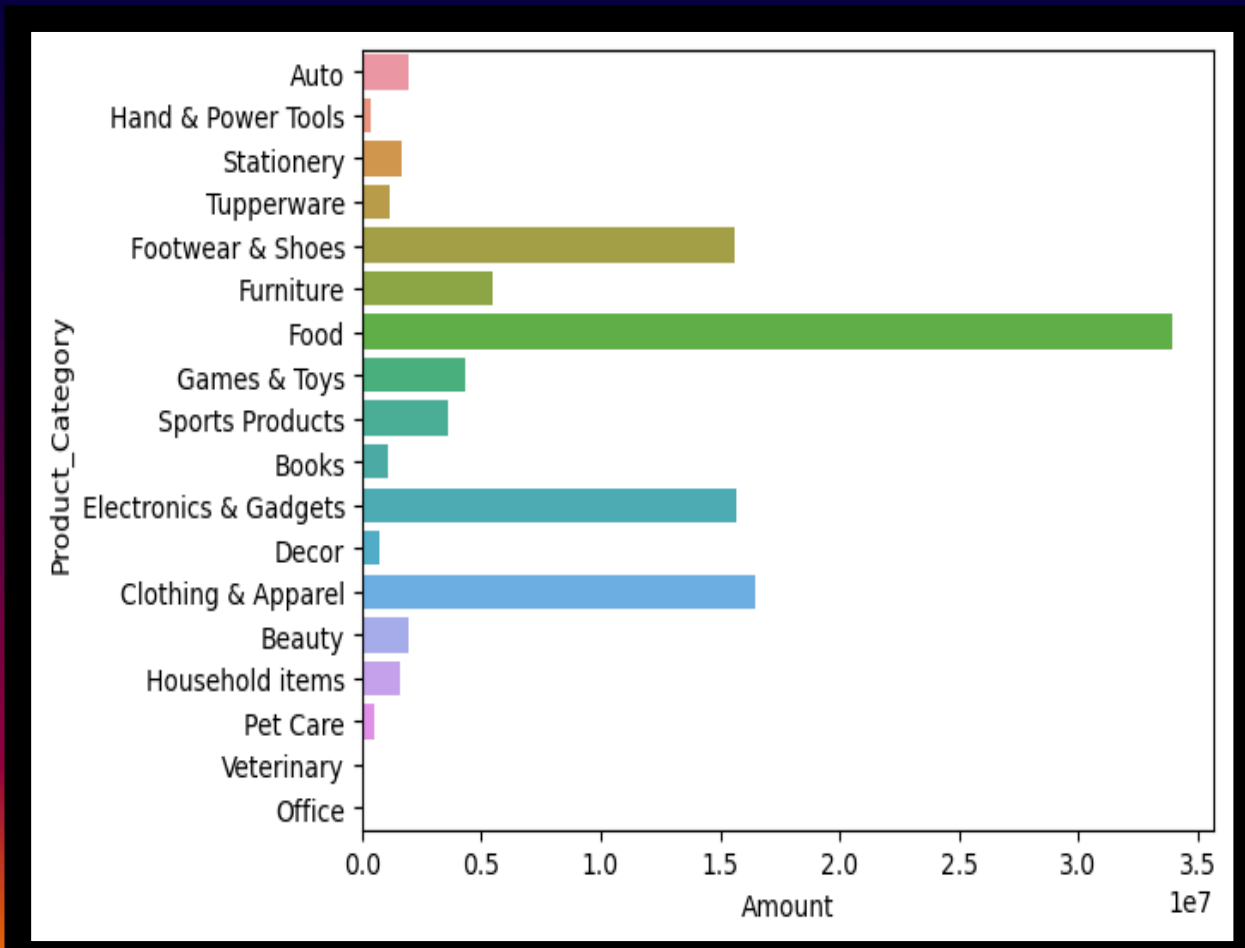
```
plt.figure(figsize=(3,4))
ax= sns.countplot(x='Marital_Status', hue=
'Gender',data=df)

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




```
# Most of the buyers are married women
```

➢ Category wise sales

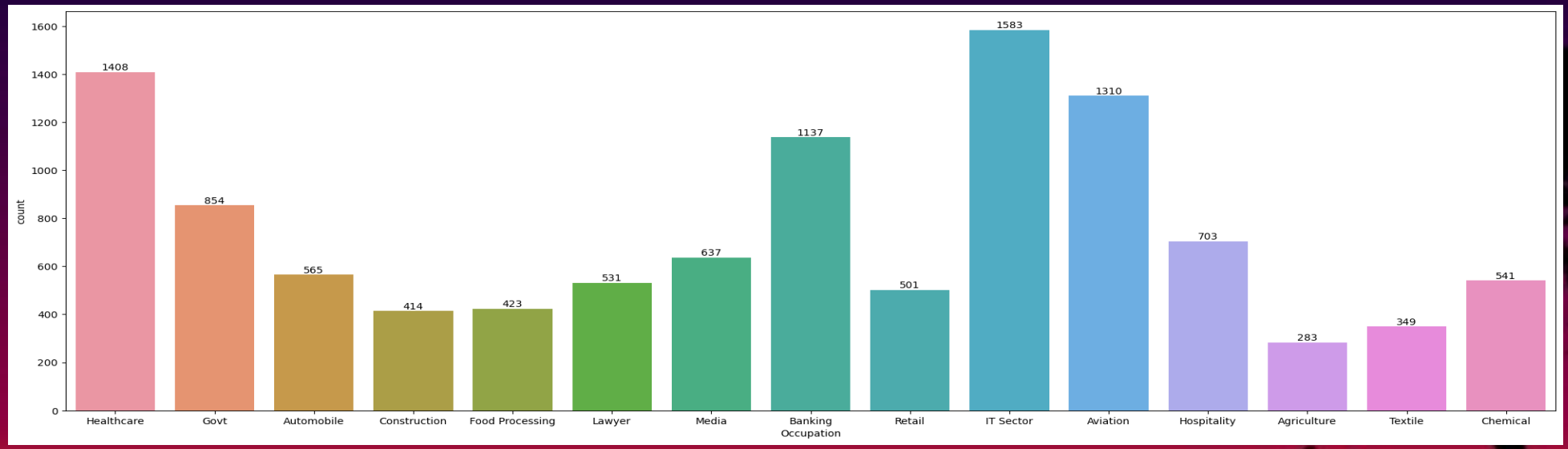
```
sns.barplot(x='Amount',y='Product_Category',
estimator= 'sum', errorbar= None, data=df)
```




```
# Most of the sold products belongs to Food,
Clothing and then Electronics & Gadgets.
```

➢ Distribution of buyers based on their occupation

```
plt.figure(figsize=(25,8)
ax=sns.countplot(x='Occupation', data=df)
for bars in ax.containers:
    ax.bar_label(bars)
```




```
#From the above graph we can see that most buyers are from IT sector then
Healthcare then Aviation followed by other sectors.
```

Key Findings and Insights :

- ✓ *Most buyers are aged 26-35.*
- ✓ *Majority of buyers are women.*
- ✓ *Most buyers are married women.*
- ✓ *Buyers predominantly come from IT, healthcare, aviation sectors.*
- ✓ *Top-selling products include food, clothing, electronics, gadgets.*
- ✓ *Highest sales are in Uttar Pradesh, Maharashtra, Karnataka highlighting these regions as key markets.*

Recommendations and data-driven decisions

- ✓ *Demographics - Target marketing to 26-35 age group using social media and influencers.*
- ✓ *Gender - Tailor content and products for women; collaborate with female influencers.*
- ✓ *Region - Improve logistics and run localized marketing campaigns in UP, Maharashtra, and Karnataka.*
- ✓ *Marital Status - Create family-oriented promotions and product bundles.*
- ✓ *Sectors - Offer discounts for IT, healthcare, and aviation sector employees; partner with companies for exclusive deals.*
- ✓ *Product Categories - Maintain strong inventory and promote bundles and discounts for food, clothing, electronics, gadgets.*

Thank You

-Mohd Amish

Seeking oppurtunities as a data analyst