Diwali Sales Analysis using Python Libraries (Pandas, NumPy, Matplotlib, Seaborn---

Objective:

The primary objective of the Diwali Sales Analysis project is to conduct an in-depth examination of consumer purchasing behaviors during the Diwali festival. This analysis aims to identify and understand key demographic trends, regional preferences, and occupational impacts on sales performance. By leveraging these insights, the project seeks to optimize future sales campaigns, enhance customer targeting strategies, and drive overall business growth during the festive season.

Dataset Used:

Diwali Sales Data

Dataset Descriptions…

- **User ID**: Unique identifier for each user.
- **Cust Name**: Name of the customer.
- **Product ID**: Unique identifier for the product purchased.
- **Gender**: Gender of the customer.
- **Age Group**: Age group of the customer (e.g., 18-25, 26-35).
- **Age**: Actual age of the customer.
- **Marital Status**: Marital status of the customer (e.g., Single, Married).
- **State**: State where the customer resides.
- **Zone**: Geographical zone of the customer (e.g., East, West).
- **Occupation**: Occupation of the customer.
- **Product Category**: Category of the product purchased (e.g., Electronics, Apparel).
- **Orders**: Number of orders placed by the customer.
- **Amounts**: Total amount spent by the customer on purchases

Importing Python Libraries and CSV files

Importing Python Libraries using the following query…

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

To import the CSV file containing Diwali Sales Data…

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

Data Cleaning and Transformations:

Deleting Null values-

To find no of rows with null values.

```
pd.isnull(df).sum()
```

```
User_ID                0
Cust_name              0
Product_ID             0
Gender                 0
Age Group              0
Age                    0
Marital_Status         0
State                  0
Zone                   0
Occupation             0
Product_Category       0
Orders                 0
Amount                12
Status             11251
unnamed1           11251
dtype: int64
```

Deleting null values from all rows

```
[10]: df.dropna(inplace = True)

[11]: pd.isnull(df).sum()

[11]: User_ID              0
      Cust_name            0
      Product_ID           0
      Gender               0
      Age Group            0
      Age                  0
      Marital_Status       0
      State                0
      Zone                 0
      Occupation           0
      Product_Category     0
      Orders               0
      Amount               0
      Status               0
      unnamed1             0
      dtype: int64
```

Deleting unwanted column

Column ‘Status’ and ‘Unnamed1’ are not required so deleting it…

Using following query

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

Changing Datatype

Changing datatype of Amount column from float to integer…

```
[57]: df['Amount'] = df['Amount'].astype('int')

[58]: df['Amount'].dtypes

[58]: dtype('int64')
```

Renaming Column

Checking the name of column and renaming the required column

```
df.columns
```

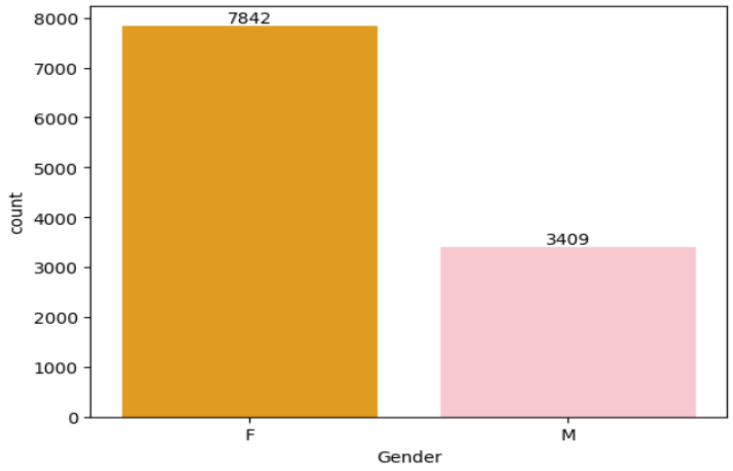
```
Index(['User_ID', 'Cust_name', 'Product_ID', 'Gender', 'Age Group', 'Age',
       'marital_status', 'State', 'Zone', 'Occupation', 'Product_Category',
       'Orders', 'Amount'],
      dtype='object')
```

```
df.rename(columns = {'Cust_name':'Customer_Name'}, inplace = True)
```

Diwali Sales Analysis: Gender wise Breakdown

1. Gender-Wise Distribution of Members

```
[6]: ax = sns.countplot(x ='Gender', data = df, palette = ['orange','pink'], dodge = False, legend = False, hue = 'Gender')
for bars in ax.containers:
    ax.bar_label(bars)
plt.show()
```

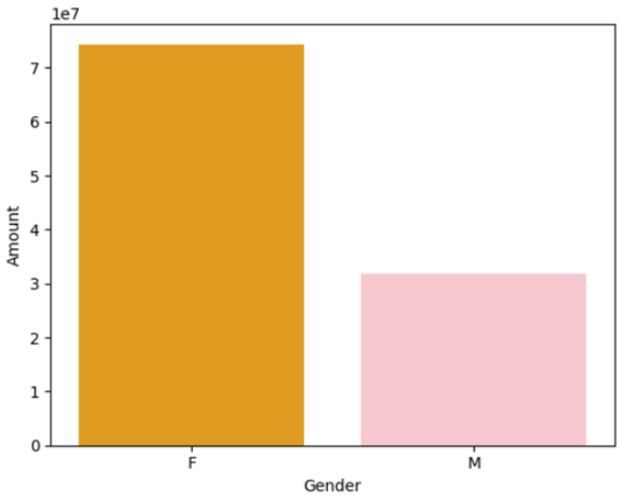


2. Gender-Wise Sales Amount Analysis

```
[9]: sales_gen = df.groupby(['Gender'], as_index = False)['Amount'].sum().sort_values(by = 'Amount', ascending = False)

[10]: sns.barplot(x='Gender', y ='Amount', data = sales_gen, palette = ['orange','pink'], legend = False , hue = 'Gender')

[10]: <Axes: xlabel='Gender', ylabel='Amount'>
```



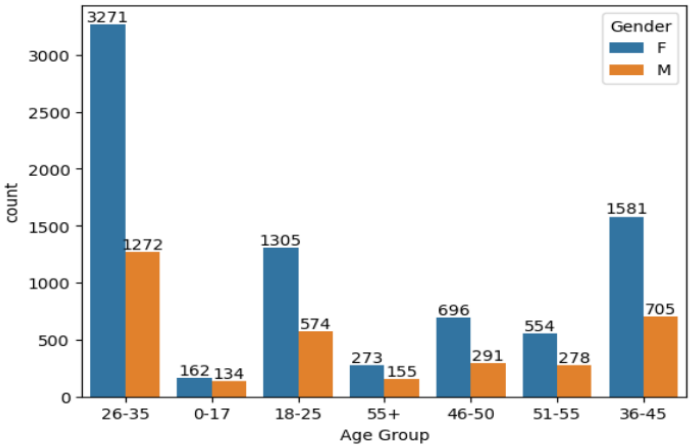
From the above two graph, the analysis reveals that females represent the majority of buyers and exhibit higher purchasing power compared to males. This indicates a strong engagement and greater financial contribution from female members.

Diwali Sales Analysis: Age wise Breakdown

1. Age wise distribution of members

```
[13]: ax = sns.countplot(x = 'Age Group', data = df, hue = 'Gender')

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

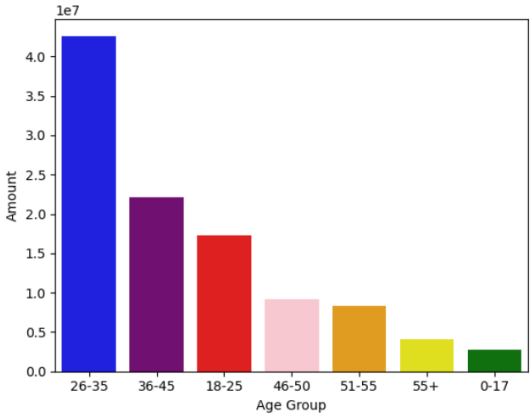


2. Age wise sales amount analysis

```
[4]: sales_age = df.groupby(['Age Group'], as_index = False)['Amount'].sum().sort_values(by = 'Amount', ascending = False)

[9]: sns.barplot(x= 'Age Group', y = 'Amount', data = sales_age, palette = ['blue', 'purple', 'red' , 'pink' , 'orange', 'yellow', 'green'],
                 hue = 'Age Group', legend = False)

[9]: <Axes: xlabel='Age Group', ylabel='Amount'>
```



Based on the two graphs, it is evident that the majority of buyers fall within the age group of 26-35 years and are predominantly female. This demographic trend highlights a key target audience for our Diwali Sales project.

Diwali Sales Analysis: State-wise Breakdown

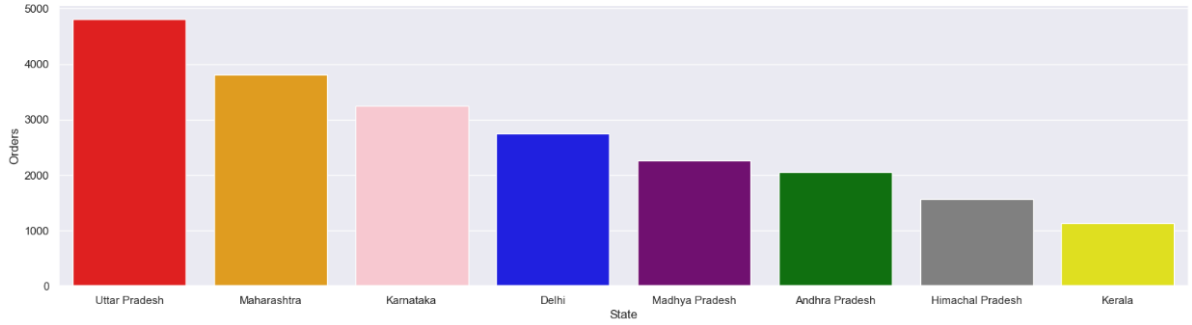
State wise distribution of members

Diwali Sales Analysis: State-wise Breakdown

```
[20]: sales_data = df.groupby(['State'], as_index = False)['Orders'].sum().sort_values(by = 'Orders' ,ascending = False).head(8)
```

```
[27]: sns.set(rc= {'figure.figsize' : (20,5)})
sns.barplot( x = 'State', y = 'Orders', data = sales_data,
            palette=['Red','Orange','Pink','Blue','Purple', 'Green', 'Grey', 'Yellow'],
            hue = 'State', legend= False)
```

```
[27]: <Axes: xlabel='State', ylabel='Orders'>
```



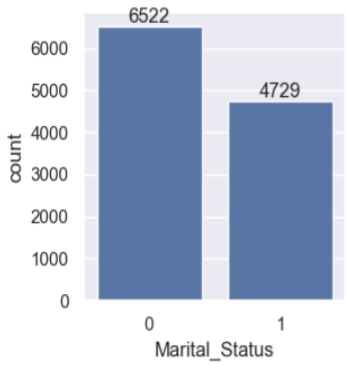
Based on the analysis, it is evident that the majority of orders originate from Uttar Pradesh, Maharashtra, and Karnataka. This insight highlights key regions contributing significantly to our Diwali sales.

Diwali Sales Analysis: Marital Status Insights

1. No of buyers who are married & unmarried

```
[34]: ax = sns.countplot(x = 'Marital_Status', data = df)

sns.set(rc = {'figure.figsize' : (4,3)})
for bars in ax.containers:
    ax.bar_label(bars)
```

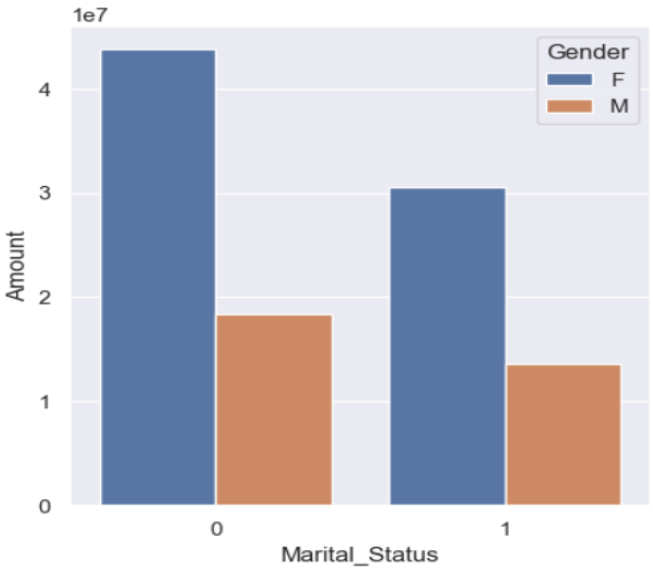


2. Sales Amount analysis with respect to Marital Status

```
[35]: data_couple = df.groupby(['Marital_Status','Gender'],
      as_index = False)['Amount'].sum().sort_values(by = 'Amount', ascending = False)

[40]: sns.barplot(x= 'Marital_Status', y= 'Amount', data = data_couple, hue = 'Gender')

[40]: <Axes: xlabel='Marital_Status', ylabel='Amount'>
```

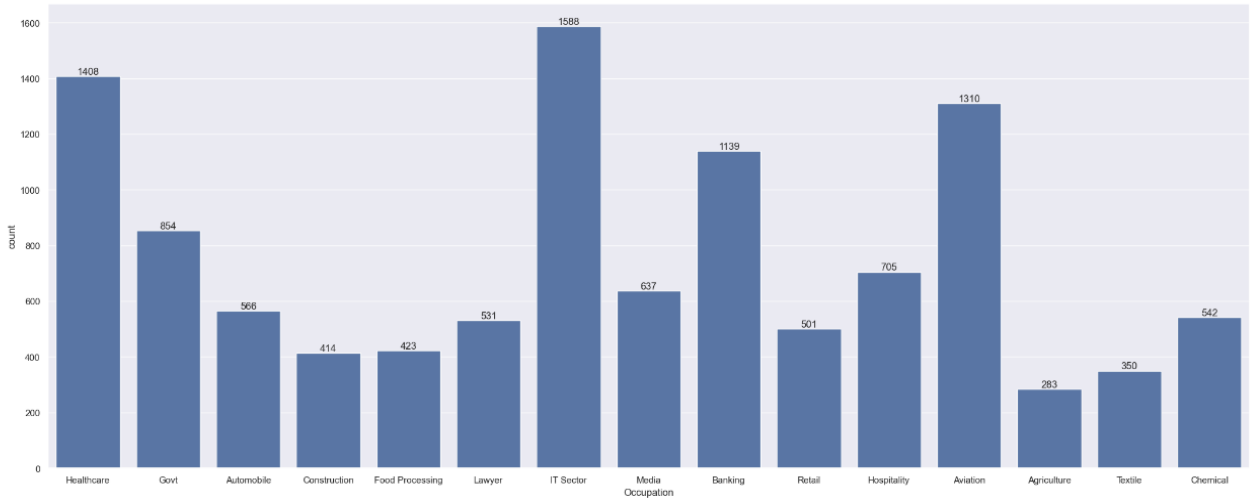


From the above two graphs, we can analyze that the majority of buyers are married women. This insight is crucial for tailoring our Diwali sales strategies to better target this demographic.

Diwali Sales Analysis: Occupation-wise Insights

1. No of orders by each occupation

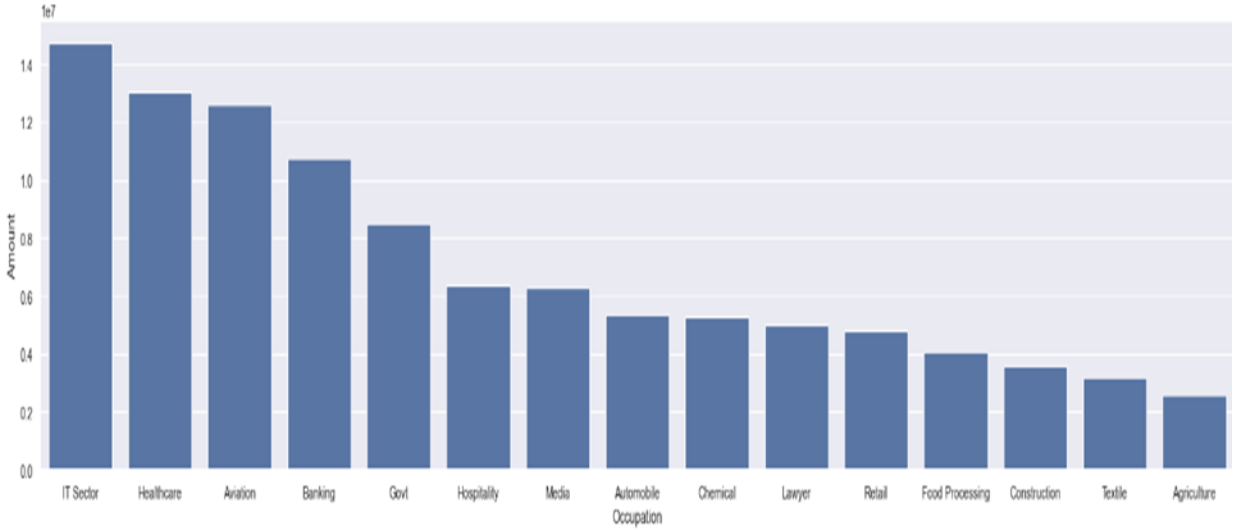
```
[58]: ax = sns.countplot(x = 'Occupation', data = df)
      sns.set(rc = {'figure.figsize':(26,5)})
      for bars in ax.containers:
          ax.bar_label(bars)
```



2. Amount of sales by each occupation

```
[49]: sales_occupation = df.groupby(['Occupation'], as_index = False)['Amount'].sum().sort_values('Amount', ascending= False)
```

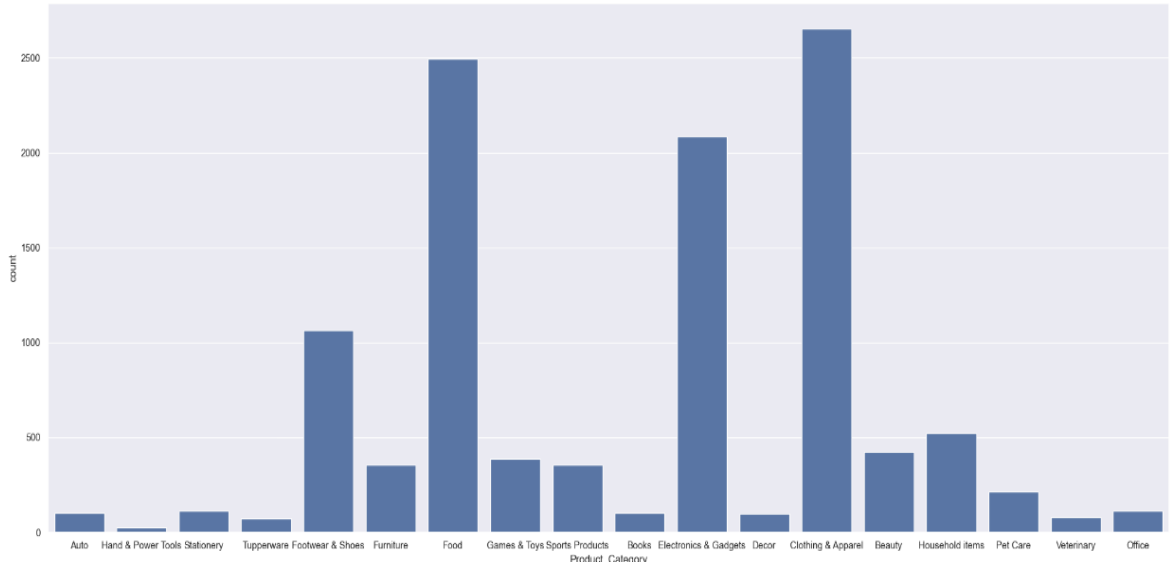
```
[59]: sns.barplot(x= 'Occupation' , y = 'Amount' , data = sales_occupation)
      sns.set(rc = {'figure.figsize' :(26,5)})
```



From the above two graphs, we can see that most buyers are from the IT, Healthcare, and Aviation sectors, whereas Agriculture and Textile have the fewest buyers. This insight highlights key industries driving our Diwali sales.

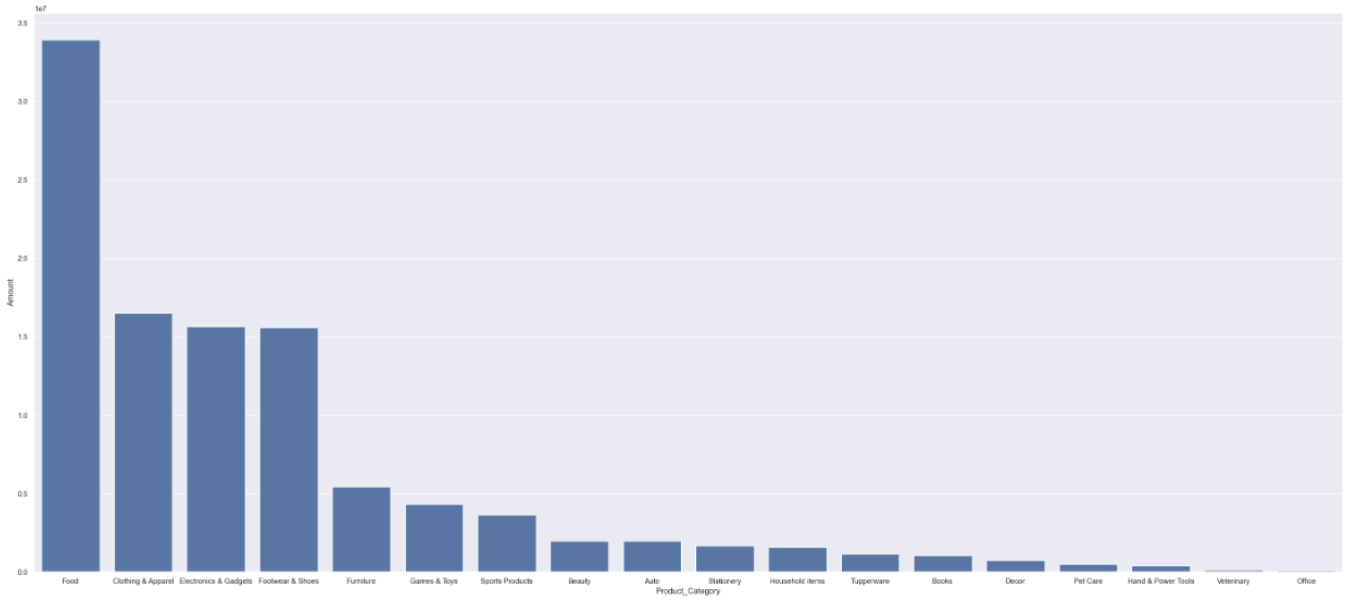
Diwali Sales Analysis: Product Category-wise Insights

```
[72]: ax = sns.countplot(x = 'Product_Category', data = df)
      sns.set(rc= {'figure.figsize':(30,10)})
```




```
product_data = df.groupby(['Product_Category'], as_index = False)['Amount'].sum().sort_values('Amount', ascending = False)
```

```
sns.barplot(x = 'Product_Category', y = 'Amount', data = product_data)
sns.set(rc= {'figure.figsize' : (35,15)})
```



From the above data, it is evident that the majority of Diwali sales are driven by purchases in the food, clothing, and electronics categories. This trend underscores the high consumer demand for these essential and desirable items during the festive period.

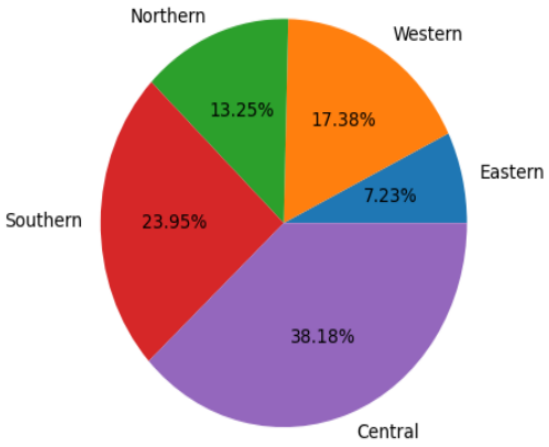
Diwali Sales Analysis: Zone-wise Insights

```
[36]: Eastern = df.loc[(df['Zone'] == "Eastern")].count()
      Western = df.loc[(df['Zone'] == "Western")].count()
      Northern = df.loc[(df['Zone'] == "Northern")].count()
      Southern = df.loc[(df['Zone'] == "Southern")].count()
      Central = df.loc[(df['Zone'] == "Central")].count()

[37]: l = ["Eastern", "Western", "Northern", "Southern", "Central"]

[38]: mlist = [Eastern["Zone"], Western["Zone"], Northern["Zone"], Southern["Zone"], Central["Zone"]]

[39]: plt.pie(mlist, labels = l, autopct = "%1.2f%%")
      plt.show()
```



For the Diwali sales analysis project, it is evident that the majority of sales are concentrated in the central and southern zones, whereas the eastern zone shows the least purchasing activity.

Conclusions:

1. Consumer Demographics: The analysis reveals a notable engagement from female consumers, who not only represent a significant portion of the buyer demographic but also demonstrate higher purchasing power compared to their male counterparts. This suggests that tailored marketing strategies focusing on female customers can maximize sales potential during the Diwali period.

2. Age Group Trends: The predominant age group contributing to Diwali sales is between 26-35 years. This demographic trend indicates that marketing campaigns tailored to the preferences and purchasing behaviors of this age group can significantly enhance sales outcomes.

3. Regional Insights: The states of Uttar Pradesh, Maharashtra, and Karnataka emerge as top contributors to Diwali sales. Targeted marketing and promotional activities in these regions, customized to local cultural preferences, can further boost sales performance in these high-potential markets.

4. Marital Status Impact: Married individuals, particularly married women, constitute the majority of the customer base. This insight is crucial for designing marketing strategies that cater to the needs and preferences of married consumers, thereby enhancing their engagement and purchase frequency.

5. Occupational Influence: Professionals in the IT, Healthcare, and Aviation sectors are identified as leading contributors to Diwali sales, while the Agriculture and Textile sectors exhibit lower purchasing activity. Focusing marketing efforts on high-contributing occupational sectors can drive better sales results.

6. Product Category Preferences: Food, clothing, and electronics are identified as the most popular product categories during the Diwali season. Ensuring adequate inventory and attractive offers in these categories can meet consumer demand and drive substantial sales growth.

7. Geographical Zone Analysis: The central and southern zones show the highest concentration of sales, whereas the eastern zone exhibits lower purchasing activity. Strengthening marketing efforts in the central and southern zones and exploring strategies to engage consumers in the eastern zone can optimize sales performance.