In [1]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

In [2]:
```
df = pd.read_csv('C:/Users/monty/OneDrive/Desktop/Diwali Sales Data.csv',encoding='latin1')
df.head()
```

Out[2]:

	User_ID	Cust_name	Product_ID	Gender	Age Group	Age	Marital_Status	State	Zone	Occupation	Product_Category	Orders	Amount	Status	unnamed1
0	1002903	Sanskriti	P00125942	F	26-35	28	0	Maharashtra	Western	Healthcare	Auto	1	23952.0	NaN	NaN
1	1000732	Kartik	P00110942	F	26-35	35	1	Andhra Pradesh	Southern	Govt	Auto	3	23934.0	NaN	NaN
2	1001990	Bindu	P00118542	F	26-35	35	1	Uttar Pradesh	Central	Automobile	Auto	3	23924.0	NaN	NaN
3	1001425	Sudevi	P00237842	M	0-17	16	0	Karnataka	Southern	Construction	Auto	2	23912.0	NaN	NaN
4	1000588	Joni	P00057942	M	26-35	28	1	Gujarat	Western	Food Processing	Auto	2	23877.0	NaN	NaN

In [3]:
```
df.info()
```

Out[3]:
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11251 entries, 0 to 11250
Data columns (total 15 columns):
 #   Column          Non-Null Count  Dtype  
---  ------          --------------  -----  
 0   User_ID         11251 non-null  int64  
 1   Cust_name       11251 non-null  object 
 2   Product_ID      11251 non-null  object 
 3   Gender          11251 non-null  object 
 4   Age Group       11251 non-null  object 
 5   Age             11251 non-null  int64  
 6   Marital_Status  11251 non-null  int64  
 7   State           11251 non-null  object 
 8   Zone            11251 non-null  object 
 9   Occupation      11251 non-null  object 
 10  Product_Category  11251 non-null  object 
 11  Orders          11251 non-null  int64  
 12  Amount          11239 non-null  float64
 13  Status          0 non-null      float64
 14  unnamed1        0 non-null      float64
dtypes: float64(3), int64(4), object(8)
memory usage: 1.3+ MB
```

In [4]:
```
df.drop(['Status', 'unnamed1'], axis=1, inplace=True)
```

In [5]:
```
df.info()
```

Out[5]:
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11251 entries, 0 to 11250
Data columns (total 13 columns):
 #   Column          Non-Null Count  Dtype  
---  ------          --------------  -----  
 0   User_ID         11251 non-null  int64  
 1   Cust_name       11251 non-null  object 
 2   Product_ID      11251 non-null  object 
 3   Gender          11251 non-null  object 
 4   Age Group       11251 non-null  object 
 5   Age             11251 non-null  int64  
 6   Marital_Status  11251 non-null  int64  
 7   State           11251 non-null  object 
 8   Zone            11251 non-null  object 
 9   Occupation      11251 non-null  object 
 10  Product_Category  11251 non-null  object 
 11  Orders          11251 non-null  int64  
 12  Amount          11239 non-null  float64
dtypes: float64(1), int64(4), object(8)
memory usage: 1.1+ MB
```

In [6]:
```
pd.isnull(df)
```

Out[6]:

	User_ID	Cust_name	Product_ID	Gender	Age Group	Age	Marital_Status	State	Zone	Occupation	Product_Category	Orders	Amount
0	False	False	False	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...	...	...	...	...	...
11246	False	False	False	False	False	False	False	False	False	False	False	False	False
11247	False	False	False	False	False	False	False	False	False	False	False	False	False
11248	False	False	False	False	False	False	False	False	False	False	False	False	False
11249	False	False	False	False	False	False	False	False	False	False	False	False	False
11250	False	False	False	False	False	False	False	False	False	False	False	False	False

11251 rows × 13 columns

In [7]:
```
pd.isnull(df).sum()
```

Out[7]:
```
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
dtype: int64
```

In [8]:
```
df.shape
```

Out[8]:
```
(11251, 13)
```

In [9]:
```
df.dropna(inplace=True)
```

In [10]:
```
df.shape
```

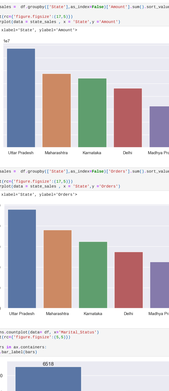
Out[10]:
```
(11239, 13)
```

In [11]:
```
df.info()
```

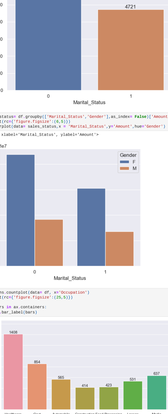
Out[11]:
```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 11239 entries, 0 to 11250
Data columns (total 13 columns):
 #   Column          Non-Null Count  Dtype  
---  ------          --------------  -----  
 0   User_ID         11239 non-null  int64  
 1   Cust_name       11239 non-null  object 
 2   Product_ID      11239 non-null  object 
 3   Gender          11239 non-null  object 
 4   Age Group       11239 non-null  object 
 5   Age             11239 non-null  int64  
 6   Marital_Status  11239 non-null  int64  
 7   State           11239 non-null  object 
 8   Zone            11239 non-null  object 
 9   Occupation      11239 non-null  object 
 10  Product_Category  11239 non-null  object 
 11  Orders          11239 non-null  int64  
 12  Amount          11239 non-null  float64
dtypes: float64(1), int64(4), object(8)
memory usage: 1.2+ MB
```

In [12]:
```
sns.countplot(df, x='Age Group',hue="Gender")
```

Out[12]:
```
<Axes: label='Age Group', xlabel='count'>
```

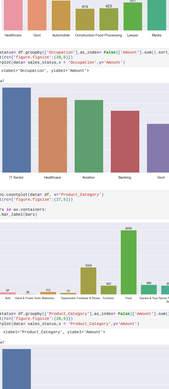


In [13]:
```
ax = sns.countplot(data=df,x='Age Group',hue="Gender")
for bars in ax.containers:
    ax.bar_label(bars)
```



In [14]:
```
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)
```

Out[14]:
```
<Axes: label='Age Group', xlabel='Amount'>
```



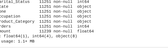
In [15]:
```
state_sales = df.groupby(['State'],as_index=False)['Amount'].sum().sort_values(by='Amount',ascending= False).head(10)
sns.set(rc={'figure.figsize':(17,5)})
sns.barplot(data = state_sales , x = 'State',y ='Amount')
```

Out[15]:
```
<Axes: xlabel='State', ylabel='Amount'>
```


In [16]:
```
state_sales = df.groupby(['State'],as_index=False)['Orders'].sum().sort_values(by='Orders',ascending= False).head(10)
sns.set(rc={'figure.figsize':(17,5)})
sns.barplot(data = state_sales , x = 'State',y ='Orders')
```

Out[16]:
```
<Axes: xlabel='State', ylabel='Orders'>
```

In [17]:
```
ax = sns.countplot(data=df, x='Marital_Status')
sns.set(rc={'figure.figsize':(5,5)})

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



In [18]:
```
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')
```

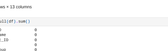
Out[18]:
```
<Axes: xlabel="Marital_Status", ylabel="Amount">
```

In [19]:
```
ax = sns.countplot(data=df, x='Occupation')
sns.set(rc={'figure.figsize':(20,5)})

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

In [20]:
```
sales_state= df.groupby(['Occupation'],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 = 'Occupation',y='Amount')
```

Out[20]:
```
<Axes: xlabel="Occupation", ylabel="Amount">
```

In [21]:
```
ax = sns.countplot(data=df, x='Product_Category')
sns.set(rc={'figure.figsize':(27,5)})

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

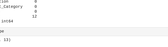


In [22]:
```
sales_state= df.groupby(['Product_Category'],as_index= False)['Amount'].sum().sort_values(by='Amount',ascending= False).head(10)
sns.set(rc={'figure.figsize':(20,5)})
sns.barplot(data=sales_state,x = 'Product_Category',y='Amount')
```

Out[22]:
```
<Axes: xlabel="Product_Category", ylabel="Amount">
```

In [23]:
```
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')
```

Out[23]:
```
<Axes: xlabel="Product_ID", ylabel="Orders">
```



In [ ]:
```

```