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

# importing dataset
data = pd.read_csv("Diwali Sales Data.csv", encoding='unicode_escape')

# top 5 and bottom 5 rows of dataset
data.head(5)
```

```
   User_ID  Cust_name Product_ID Gender Age Group  Age  Marital_Status
\
0  1002903  Sanskriti  P00125942      F      26-35   28               0

1  1000732     Kartik  P00110942      F      26-35   35               1

2  1001990      Bindu  P00118542      F      26-35   35               1

3  1001425     Sudevi  P00237842      M       0-17   16               0

4  1000588       Joni  P00057942      M      26-35   28               1


            State      Zone       Occupation Product_Category  Orders
\
0     Maharashtra   Western       Healthcare             Auto       1

1  Andhra Pradesh  Southern             Govt             Auto       3

2   Uttar Pradesh   Central       Automobile             Auto       3

3       Karnataka  Southern     Construction             Auto       2

4         Gujarat   Western  Food Processing             Auto       2


    Amount  Status  unnamed1
0  23952.0     NaN       NaN
1  23934.0     NaN       NaN
2  23924.0     NaN       NaN
3  23912.0     NaN       NaN
4  23877.0     NaN       NaN
```

```python
data.tail(5)
```

```
       User_ID    Cust_name Product_ID Gender Age Group  Age
Marital_Status  \
11246  1000695      Manning  P00296942      M     18-25   19
1
```

```
11247  1004089  Reichenbach  P00171342      M     26-35   33
0
11248  1001209        Oshin  P00201342      F     36-45   40
0
11249  1004023       Noonan  P00059442      M     36-45   37
0
11250  1002744      Brumley  P00281742      F     18-25   19
0

                State      Zone   Occupation Product_Category  Orders
Amount  \
11246     Maharashtra   Western     Chemical           Office       4
370.0
11247         Haryana  Northern   Healthcare       Veterinary       3
367.0
11248  Madhya Pradesh   Central      Textile           Office       4
213.0
11249       Karnataka  Southern  Agriculture           Office       3
206.0
11250     Maharashtra   Western   Healthcare           Office       3
188.0

       Status  unnamed1
11246     NaN       NaN
11247     NaN       NaN
11248     NaN       NaN
11249     NaN       NaN
11250     NaN       NaN
```

```
# no of rows and columns
data.shape
print(f'Number of rows in dataset:{data.shape[0]}')
print(f'Number of columns in dataset:{data.shape[1]}')
```

```
Number of rows in dataset:11251
Number of columns in dataset:15
```

```
# Display columns
data.columns
```

```
Index(['User_ID', 'Cust_name', 'Product_ID', 'Gender', 'Age Group',
'Age',
       'Marital_Status', 'State', 'Zone', 'Occupation',
'Product_Category',
       'Orders', 'Amount', 'Status', 'unnamed1'],
      dtype='object')
```

```
# stats about dataset
data.describe(include='all')
```

```
             User_ID Cust_name Product_ID Gender Age Group
Age  \
count   1.125100e+04     11251      11251  11251     11251
11251.000000
unique           NaN      1250       2351      2         7
NaN
top              NaN  Vishakha  P00265242      F     26-35
NaN
freq             NaN        42         53   7842      4543
NaN
mean    1.003004e+06       NaN        NaN    NaN       NaN
35.421207
std     1.716125e+03       NaN        NaN    NaN       NaN
12.754122
min     1.000001e+06       NaN        NaN    NaN       NaN
12.000000
25%     1.001492e+06       NaN        NaN    NaN       NaN
27.000000
50%     1.003065e+06       NaN        NaN    NaN       NaN
33.000000
75%     1.004430e+06       NaN        NaN    NaN       NaN
43.000000
max     1.006040e+06       NaN        NaN    NaN       NaN
92.000000

        Marital_Status          State     Zone Occupation
Product_Category  \
count     11251.000000          11251    11251      11251
11251
unique             NaN             16        5         15
18
top                NaN  Uttar Pradesh  Central  IT Sector  Clothing &
Apparel
freq               NaN           1946     4296       1588
2655
mean          0.420318            NaN      NaN        NaN
NaN
std           0.493632            NaN      NaN        NaN
NaN
min           0.000000            NaN      NaN        NaN
NaN
25%           0.000000            NaN      NaN        NaN
NaN
50%           0.000000            NaN      NaN        NaN
NaN
75%           1.000000            NaN      NaN        NaN
NaN
max           1.000000            NaN      NaN        NaN
NaN
```

```
              Orders        Amount  Status  unnamed1
count   11251.000000  11239.000000     0.0       0.0
unique           NaN           NaN     NaN       NaN
top              NaN           NaN     NaN       NaN
freq             NaN           NaN     NaN       NaN
mean        2.489290   9453.610858     NaN       NaN
std         1.115047   5222.355869     NaN       NaN
min         1.000000    188.000000     NaN       NaN
25%         1.500000   5443.000000     NaN       NaN
50%         2.000000   8109.000000     NaN       NaN
75%         3.000000  12675.000000     NaN       NaN
max         4.000000  23952.000000     NaN       NaN
```

```
# infromation about dataset
data.info()
```

```
<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
```

```
# checking if any column is null
data.isnull().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
```

```
# handling null vaues
average = data.Amount.mean()
average
```

```
np.float64(9453.610857727557)
```

```
data.Amount.fillna(round(average,0),inplace=True)
```

```
# remove Status and unnamed1 columns
data.drop(columns=['Status','unnamed1'],inplace=True)
```

```
data.columns
```

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

```
# check whether duplicate values or not
data.duplicated().sum()
```

```
np.int64(8)
```

```
# handling dupicate values
data[data.duplicated()==True]
```

```
       User_ID   Cust_name Product_ID Gender Age Group  Age
Marital_Status  \
19     1001883     Praneet  P00029842      M     51-55   54
1
4404   1004725     Jackson  P00150842      F     36-45   37
1
5703   1003208      Bowman  P00171642      F     26-35   31
0
5908   1001260     Dheeraj  P00344042      M     26-35   28
0
6173   1001325       Reese  P00111742      F     26-35   27
1
8651   1000083        Gute  P00242842      M     26-35   35
```

```
0
8941   1001476     Anudeep  P00036842      M     18-25   25
0
10571  1004404  Rittenbach  P00150142      F     26-35   28
1

               State      Zone    Occupation       Product_Category
Orders  \
19     Uttar Pradesh   Central   Hospitality                   Auto
1
4404     Maharashtra   Western   Hospitality  Electronics & Gadgets
4
5703           Bihar   Eastern   Agriculture  Electronics & Gadgets
4
5908     Maharashtra   Western     IT Sector  Electronics & Gadgets
4
6173         Gujarat   Western  Construction  Electronics & Gadgets
3
8651   Uttar Pradesh   Central   Hospitality     Clothing & Apparel
3
8941     Maharashtra   Western     IT Sector     Clothing & Apparel
4
10571        Haryana  Northern      Aviation  Electronics & Gadgets
3

        Amount
19     23568.0
4404    9859.0
5703    8088.0
5908    8015.0
6173    7923.0
8651    5345.0
8941    5202.0
10571   2304.0
```

```
data.drop_duplicates()
```

```
       User_ID    Cust_name Product_ID Gender Age Group  Age
Marital_Status  \
0      1002903    Sanskriti  P00125942      F     26-35   28
0
1      1000732       Kartik  P00110942      F     26-35   35
1
2      1001990        Bindu  P00118542      F     26-35   35
1
3      1001425       Sudevi  P00237842      M      0-17   16
0
4      1000588         Joni  P00057942      M     26-35   28
1
...        ...          ...        ...    ...       ...  ...
```

```
...
11246  1000695      Manning  P00296942      M     18-25   19
1
11247  1004089  Reichenbach  P00171342      M     26-35   33
0
11248  1001209        Oshin  P00201342      F     36-45   40
0
11249  1004023       Noonan  P00059442      M     36-45   37
0
11250  1002744      Brumley  P00281742      F     18-25   19
0

                State      Zone       Occupation Product_Category
Orders  \
0         Maharashtra   Western       Healthcare             Auto
1
1      Andhra Pradesh  Southern             Govt             Auto
3
2       Uttar Pradesh   Central       Automobile             Auto
3
3           Karnataka  Southern     Construction             Auto
2
4             Gujarat   Western  Food Processing             Auto
2
...               ...       ...              ...              ...
...
11246     Maharashtra   Western         Chemical           Office
4
11247         Haryana  Northern       Healthcare       Veterinary
3
11248  Madhya Pradesh   Central          Textile           Office
4
11249       Karnataka  Southern      Agriculture           Office
3
11250     Maharashtra   Western       Healthcare           Office
3

        Amount
0      23952.0
1      23934.0
2      23924.0
3      23912.0
4      23877.0
...        ...
11246    370.0
11247    367.0
11248    213.0
11249    206.0
11250    188.0
```

```
[11243 rows x 13 columns]
```

```
# datatypes of columns
data.dtypes
```

```
User_ID                      int64
Cust_name                   object
Product_ID                  object
Gender                      object
Age Group                   object
Age                          int64
Marital_Status               int64
State                       object
Zone                        object
Occupation                  object
Product_Category            object
Orders                       int64
Amount                     float64
dtype: object
```

```
# Amount datatype convert from float64 to int64
data.Amount=data.Amount.astype('int64')
```

```
data.dtypes
```

```
User_ID                     int64
Cust_name                  object
Product_ID                 object
Gender                     object
Age Group                  object
Age                         int64
Marital_Status              int64
State                      object
Zone                       object
Occupation                 object
Product_Category           object
Orders                      int64
Amount                      int64
dtype: object
```

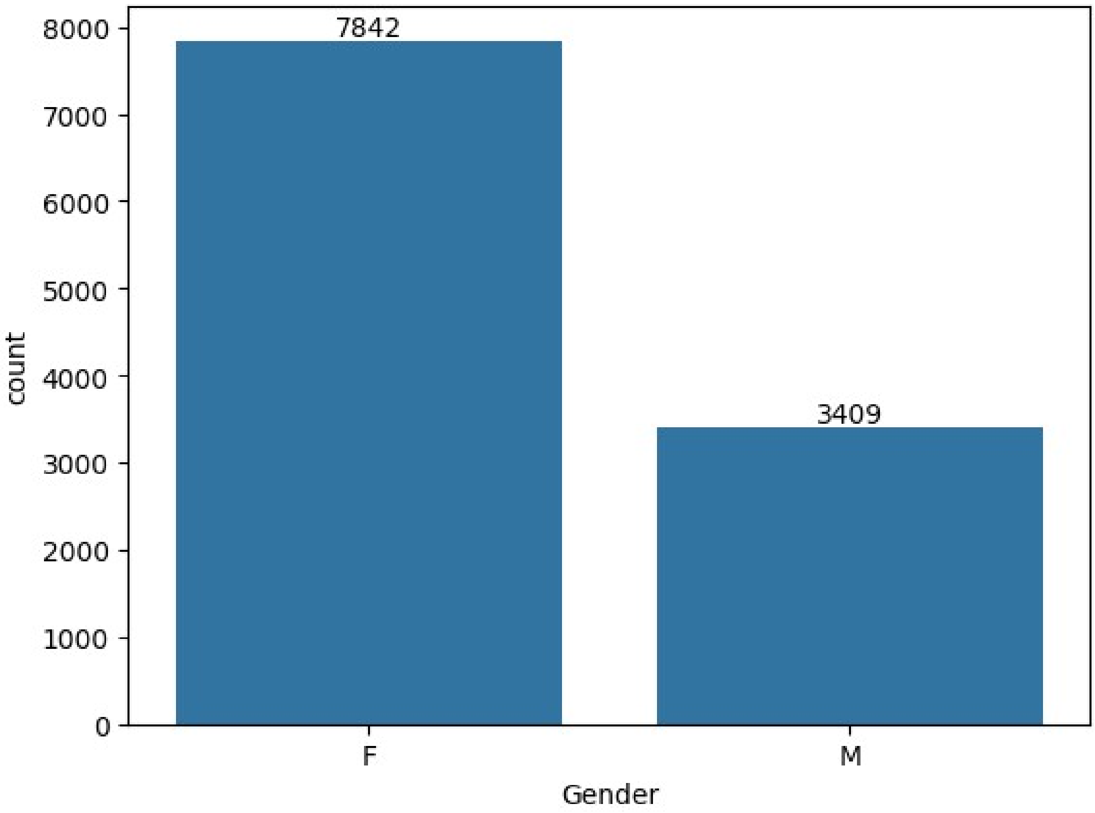
```
# rename columns
data.rename(columns={'Marital_Status':'Married_Or_Not'},inplace=True)
data.columns
```

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

Exploratory Data Analysis

Gender

```
count_plot = sns.countplot(x='Gender',data=data,stat='count')
for bars in count_plot.containers:
    count_plot.bar_label(bars)
```

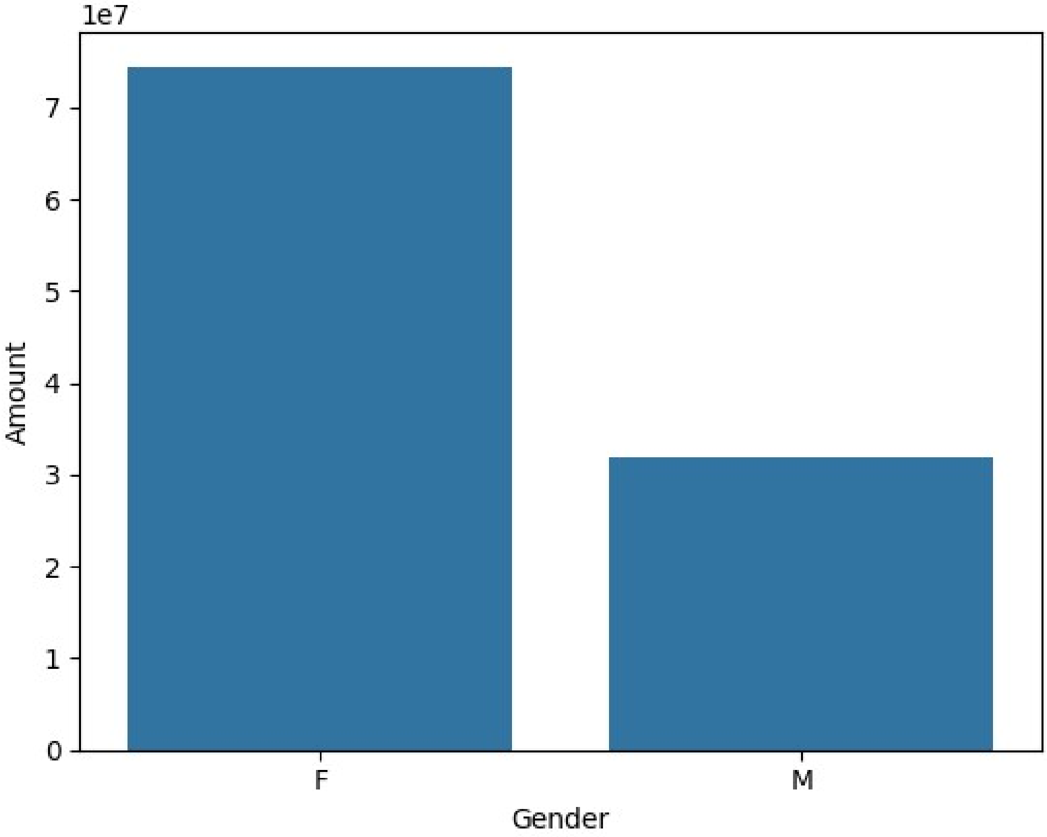


```
# genderwise amount
am =data.groupby(['Gender'],as_index=False)
['Amount'].sum().sort_values(by='Amount',ascending=False)
am
```

```
  Gender    Amount
0      F  74430393
1      M  31932184
```

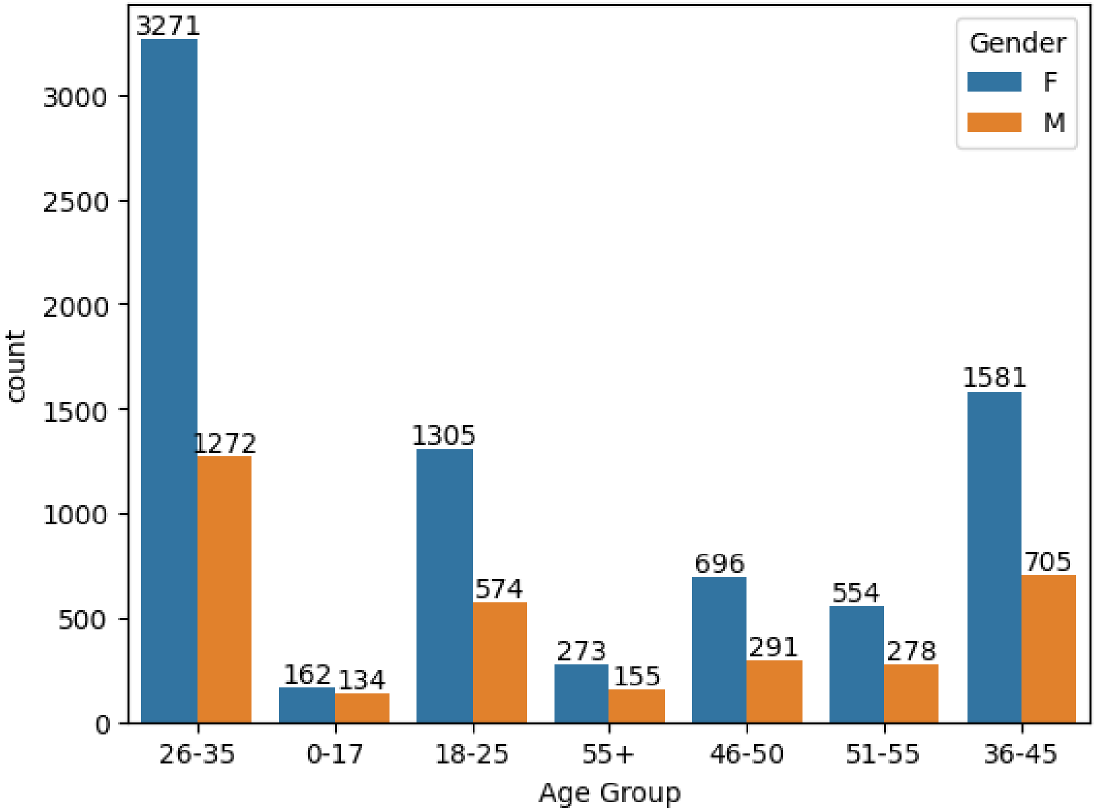
```
sns.barplot(data=am,x= 'Gender',y='Amount')
```

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



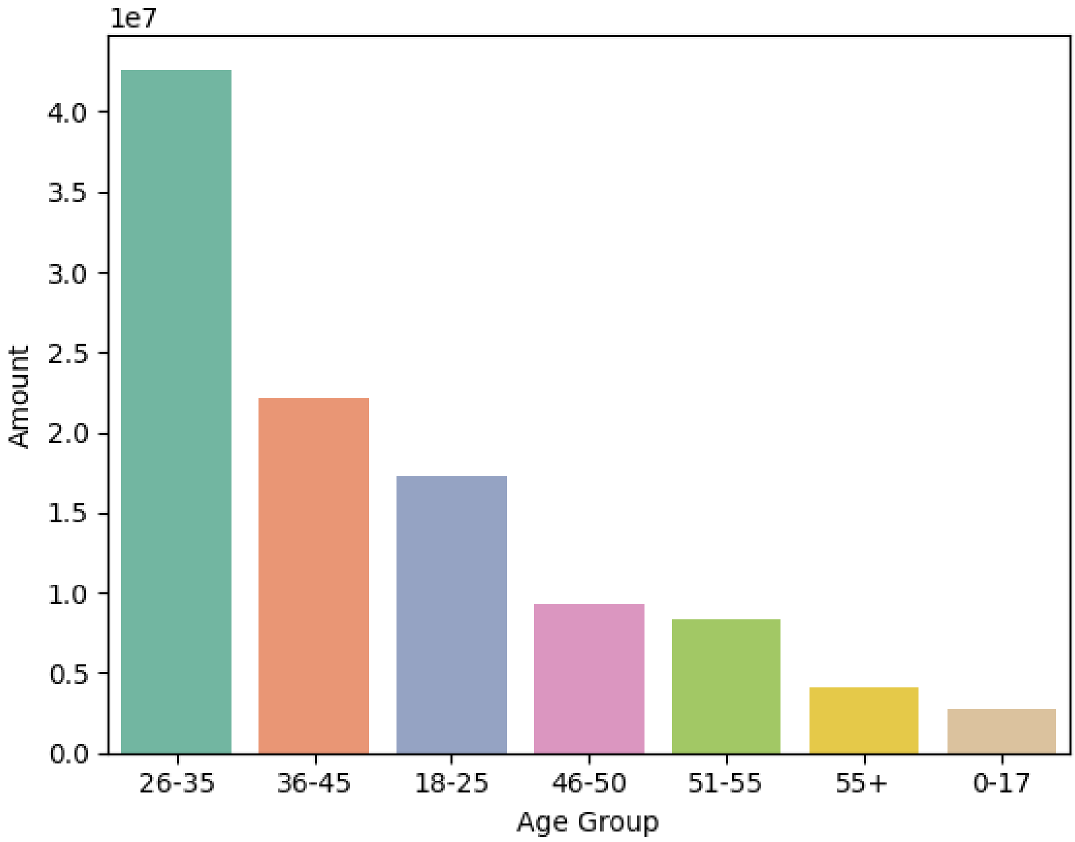
Age_Group

```
age = sns.countplot(data = data, x = 'Age Group', hue = 'Gender')

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

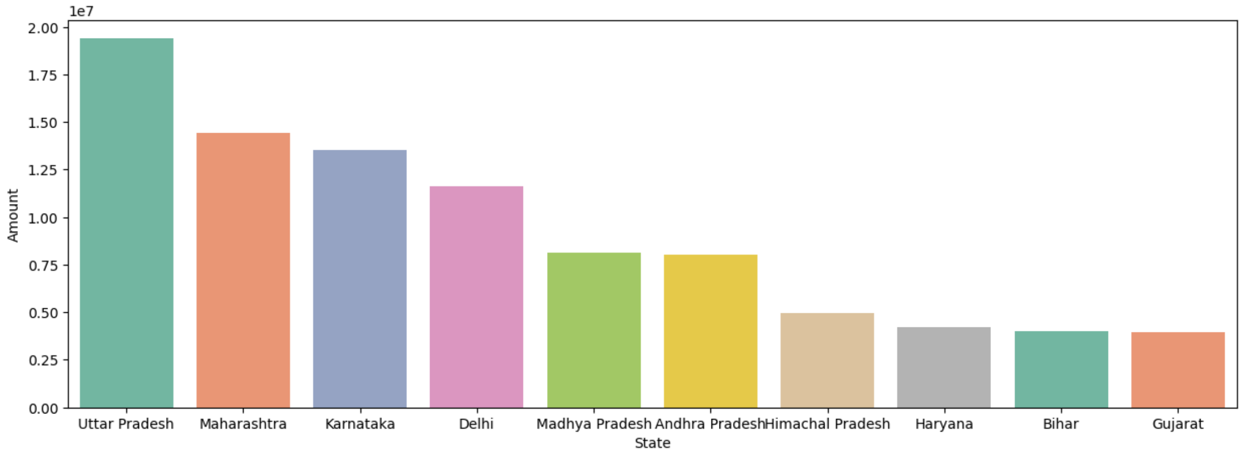


```
import warnings
warnings.filterwarnings('ignore')

# age_group wise amount
am1 = data.groupby(['Age Group'],as_index=False)
['Amount'].sum().sort_values(by='Amount',ascending=False)
sns.barplot(data=am1,x='Age Group',y= 'Amount', palette='Set2')
```

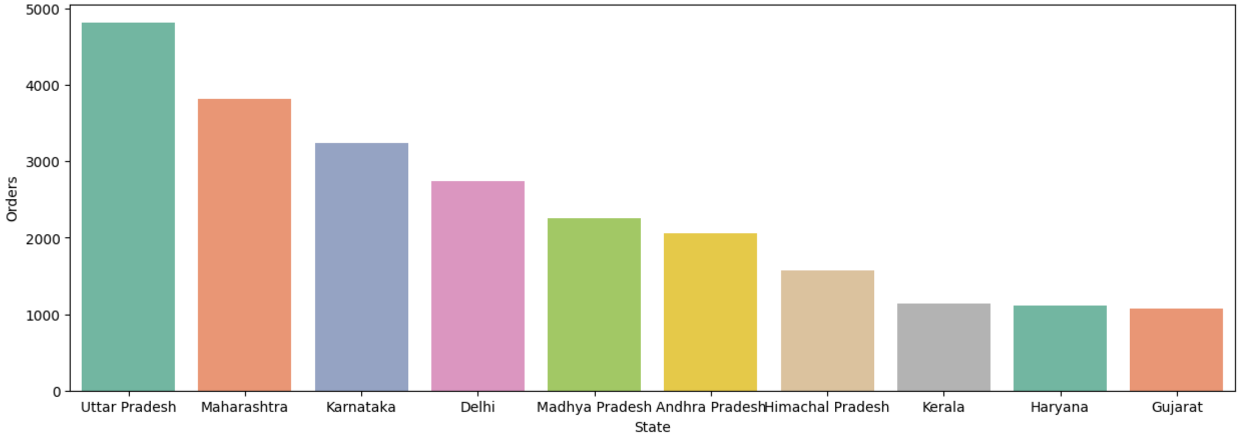
```
<Axes: xlabel='Age Group', ylabel='Amount'>
```



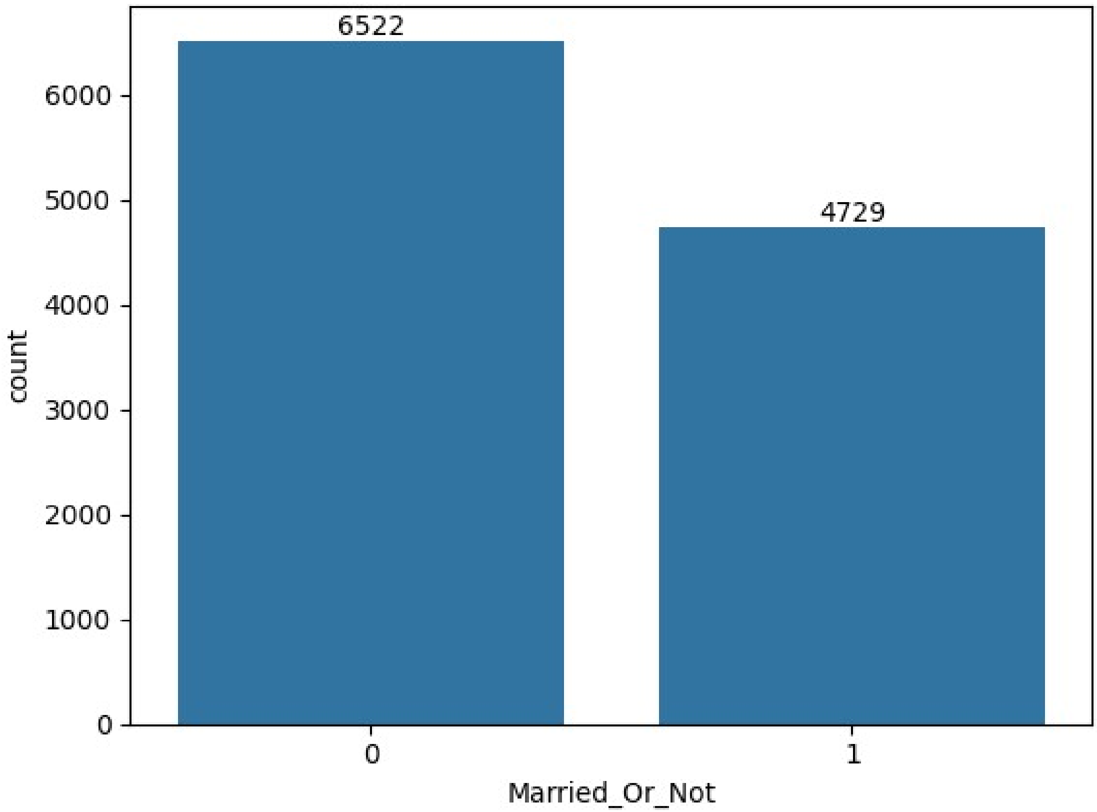
```
# top 10 states-wise total amount
states =data.groupby(['State'],as_index=False)
['Amount'].sum().sort_values(by= 'Amount',ascending=False).head(10)
plt.figure(figsize=(15,5))
sns.barplot(data= states,x='State',y='Amount',palette='Set2')
plt.show()
```



```
# top 10 state wise orders
# top 10 states
states =data.groupby(['State'],as_index=False)
['Orders'].sum().sort_values(by= 'Orders',ascending=False).head(10)
plt.figure(figsize=(15,5))
sns.barplot(data= states,x='State',y='Orders',palette='Set2')
plt.show()
```



```
married =sns.countplot(data = data, x = 'Married_Or_Not')
plt.figure(figsize=(4,5))
for bars in married.containers:
    married.bar_label(bars)
```



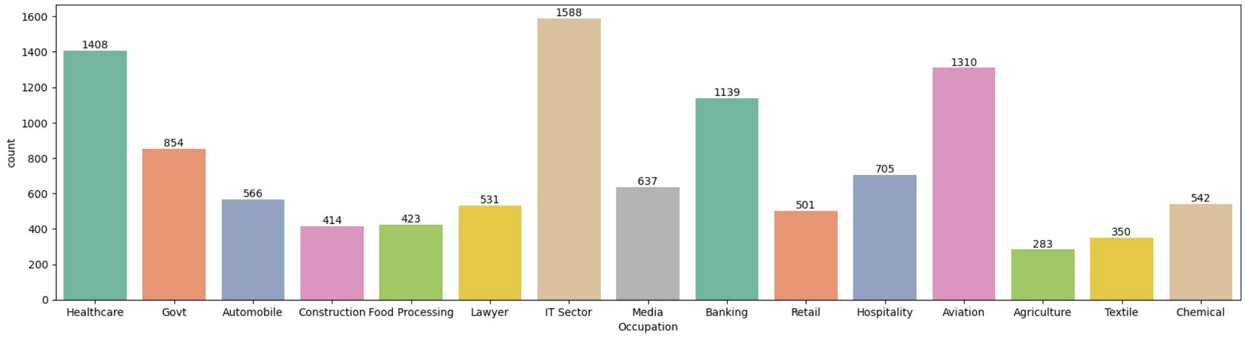
```
<Figure size 400x500 with 0 Axes>
```

```
data.Married_Or_Not.unique()
```

```
array([0, 1])
```

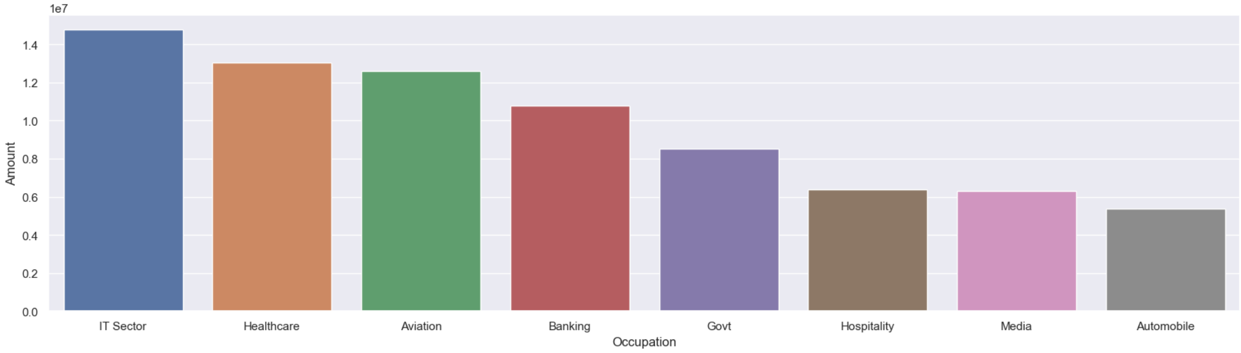
Occupation

```
plt.figure(figsize=(20,5))
ax = sns.countplot(data = data, x = 'Occupation',palette='Set2')

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



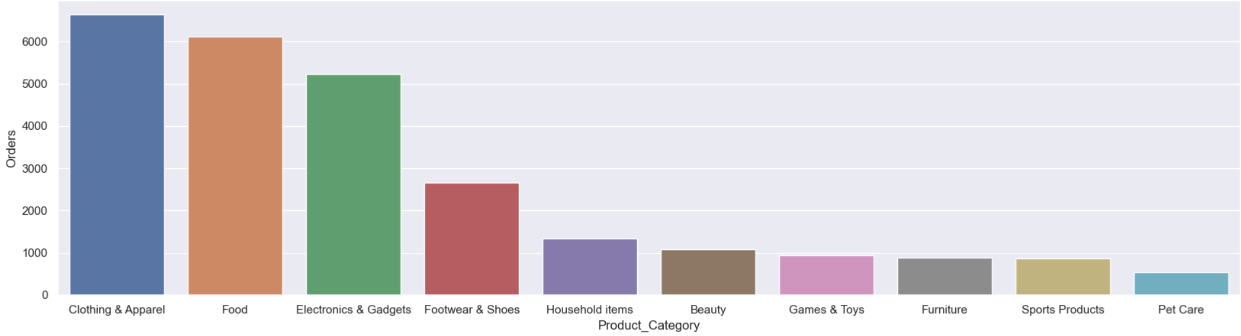
```python
# top 8 performing occupation-wise total amount
top_states = data.groupby(['Occupation'], as_index=False)
['Amount'].sum().sort_values(by='Amount', ascending=False).head(8)

sns.set(rc={'figure.figsize':(20,5)})
sns.barplot(data = top_states, x = 'Occupation',y=
'Amount',palette='deep')
```

```
<Axes: xlabel='Occupation', ylabel='Amount'>
```



```python
# top 10 most sold products
top_products = data.groupby(['Product_Category'], as_index=False)
['Orders'].sum().sort_values(by='Orders', ascending=False).head(10)

sns.set(rc={'figure.figsize':(20,5)})
sns.barplot(data = top_products, x = 'Product_Category',y=
'Orders',palette='deep')
```

```
<Axes: xlabel='Product_Category', ylabel='Orders'>
```



Conclusion

During Dwali Sales, Married women age group 26-35 yrs from UP, Maharastra and Karnataka working in IT, Healthcare and Aviation are more likely to buy products from Food, Clothing and Electronics category