In [1]:

```
import numpy as np
import seaborn as sns
import pandas as pd
import os
```

In [2]:

```
df = pd.read_csv("C:\\Users\\ASUS\\OneDrive\\Desktop\\Darshna\\Project\\Diwali Sales Dat
```

In [3]:

```
df.describe()
```

Out[3]:

	User_ID	Age	Marital_Status	Orders	Amount
count	1.125100e+04	11251.000000	11251.000000	11251.000000	11239.000000
mean	1.003004e+06	35.421207	0.420318	2.489290	9453.610858
std	1.716125e+03	12.754122	0.493632	1.115047	5222.355869
min	1.000001e+06	12.000000	0.000000	1.000000	188.000000
25%	1.001492e+06	27.000000	0.000000	1.500000	5443.000000
50%	1.003065e+06	33.000000	0.000000	2.000000	8109.000000
75%	1.004430e+06	43.000000	1.000000	3.000000	12675.000000
max	1.006040e+06	92.000000	1.000000	4.000000	23952.000000

In [4]:

```
df.shape
```

Out[4]:

```
(11251, 13)
```

In [5]:

```
df.info() #amount variable has null values
```

```
<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]:

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

Out[6]:

```
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 [7]:

```
df.head(5)
```

Out[7]:

	User_ID	Cust_name	Product_ID	Gender	Age Group	Age	Marital_Status	State	
0	1002903	Sanskriti	P00125942	F	26-35	28	0	Maharashtra	W
1	1000732	Kartik	P00110942	F	26-35	35	1	Andhra Pradesh	Sc
2	1001990	Bindu	P00118542	F	26-35	35	1	Uttar Pradesh	
3	1001425	Sudevi	P00237842	M	0-17	16	0	Karnataka	Sc
4	1000588	Joni	P00057942	M	26-35	28	1	Gujarat	W

In [8]:

```
df.drop(['Cust_name'] , axis = 1 , inplace = True) #drop identifier column
```

In [9]:

```
df.columns
```

Out[9]:

```
Index(['User_ID', 'Product_ID', 'Gender', 'Age Group', 'Age', 'Marital_Sta
tus',
       'State', 'Zone', 'Occupation', 'Product_Category', 'Orders', 'Amoun
t'],
      dtype='object')
```

In [10]:

```
df.dropna()
```

Out[10]:

	User_ID	Product_ID	Gender	Age Group	Age	Marital_Status	State	Zone
0	1002903	P00125942	F	26-35	28	0	Maharashtra	Western
1	1000732	P00110942	F	26-35	35	1	Andhra Pradesh	Southern
2	1001990	P00118542	F	26-35	35	1	Uttar Pradesh	Central
3	1001425	P00237842	M	0-17	16	0	Karnataka	Southern
4	1000588	P00057942	M	26-35	28	1	Gujarat	Western
...	...	...	...	...	...	...	...	...
11246	1000695	P00296942	M	18-25	19	1	Maharashtra	Western
11247	1004089	P00171342	M	26-35	33	0	Haryana	Northern
11248	1001209	P00201342	F	36-45	40	0	Madhya Pradesh	Central
11249	1004023	P00059442	M	36-45	37	0	Karnataka	Southern
11250	1002744	P00281742	F	18-25	19	0	Maharashtra	Western

11239 rows × 12 columns

In [11]:

```
contvar=[]
for i in df.columns:
    print(i)
    if df[i].dtype!='o':
        print(i)
        contvar.append(i)
print(contvar)
```

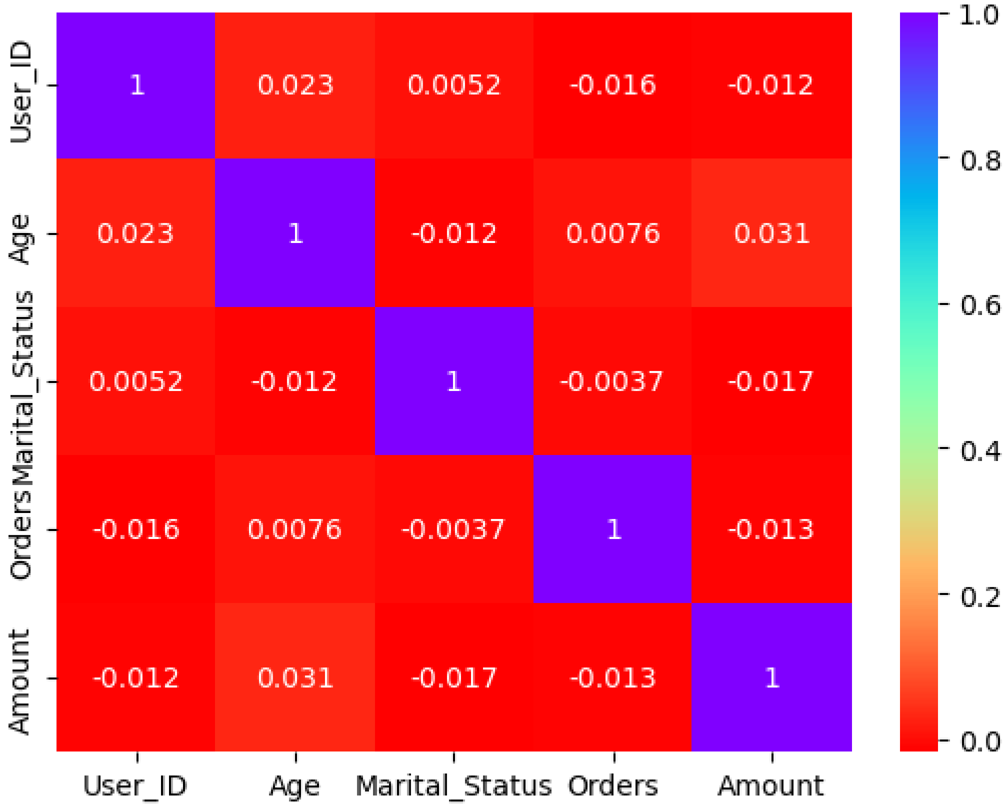
```
User_ID
User_ID
Product_ID
Product_ID
Gender
Gender
Age Group
Age Group
Age
Age
Marital_Status
Marital_Status
State
State
Zone
Zone
Occupation
Occupation
Product_Category
Product_Category
Orders
Orders
Amount
Amount
['User_ID', 'Product_ID', 'Gender', 'Age Group', 'Age', 'Marital_Status',
'State', 'Zone', 'Occupation', 'Product_Category', 'Orders', 'Amount']
```

In [12]:

```
corrdf=df.corr()
corrdf
sns.heatmap(corrdf,cmap ='rainbow_r', annot=True)
```

Out[12]:

```
<AxesSubplot:>
```



	User_ID	Age	Marital_Status	Orders	Amount
User_ID	1	0.023	0.0052	-0.016	-0.012
Age	0.023	1	-0.012	0.0076	0.031
Marital_Status	0.0052	-0.012	1	-0.0037	-0.017
Orders	-0.016	0.0076	-0.0037	1	-0.013
Amount	-0.012	0.031	-0.017	-0.013	1

In [13]:

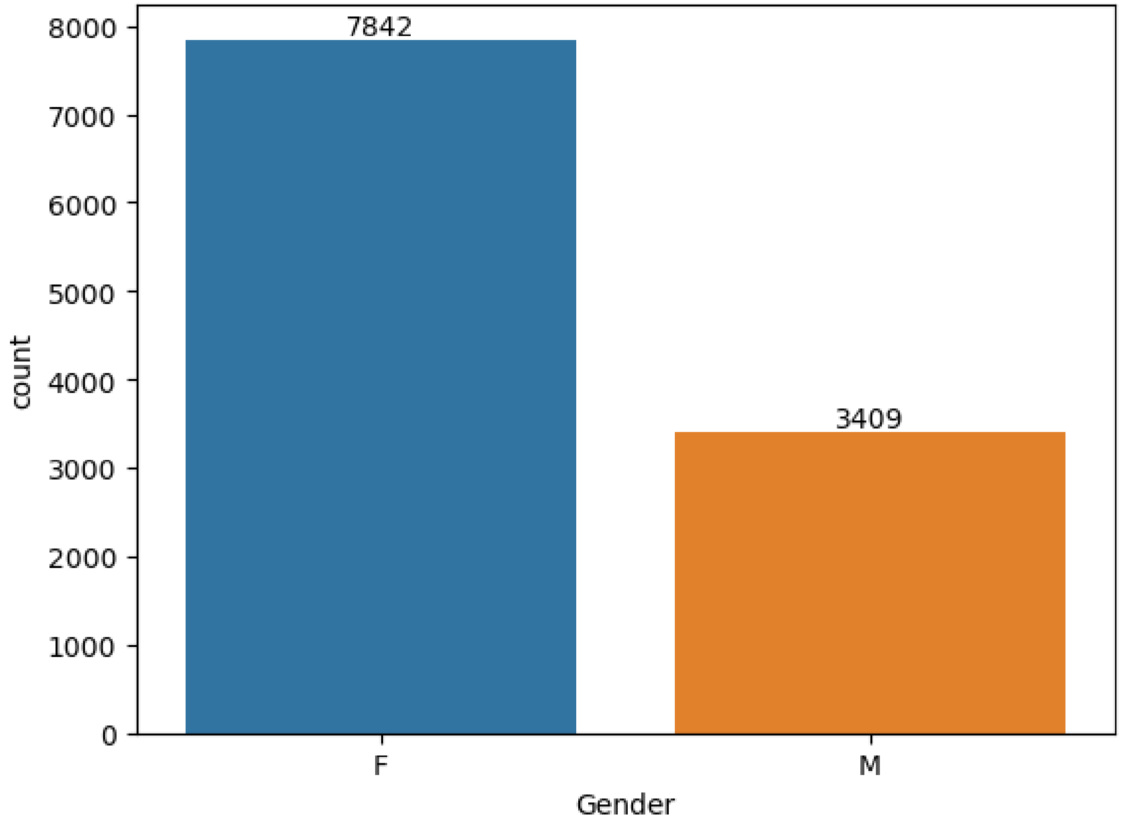
```
df.columns
```

Out[13]:

```
Index(['User_ID', 'Product_ID', 'Gender', 'Age Group', 'Age', 'Marital_Sta
tus',
       'State', 'Zone', 'Occupation', 'Product_Category', 'Orders', 'Amoun
t'],
      dtype='object')
```

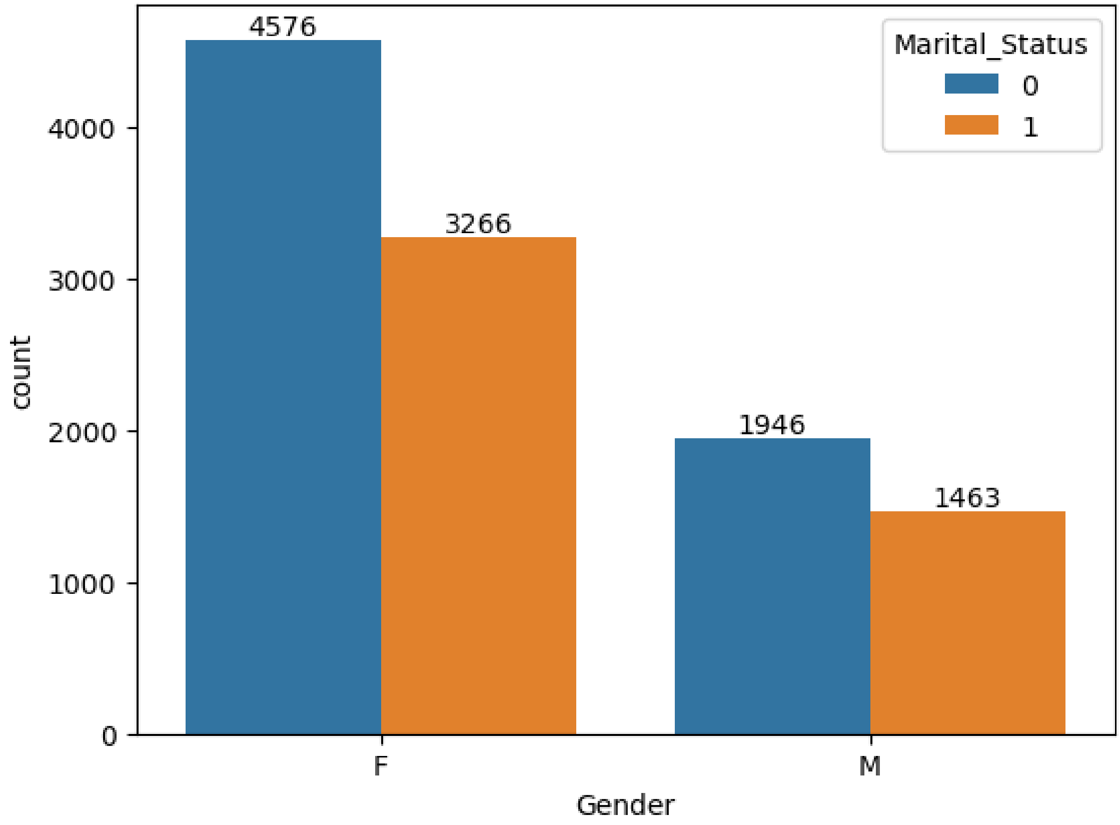
In [14]:

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



In [15]:

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



By the abover graph , we understand that married females purchase more.

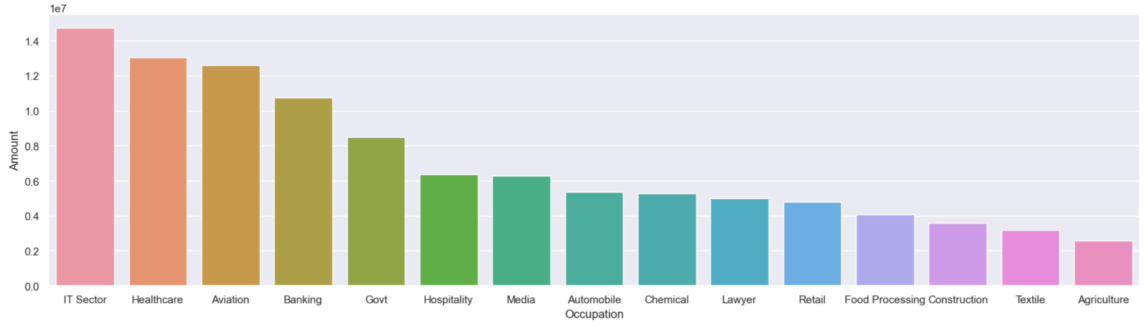
In [16]:

```
##  amount spent by Occuppation
Occupation = df.groupby(['Occupation'], as_index=False)['Amount'].sum().sort_values(by='

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

Out[16]:

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



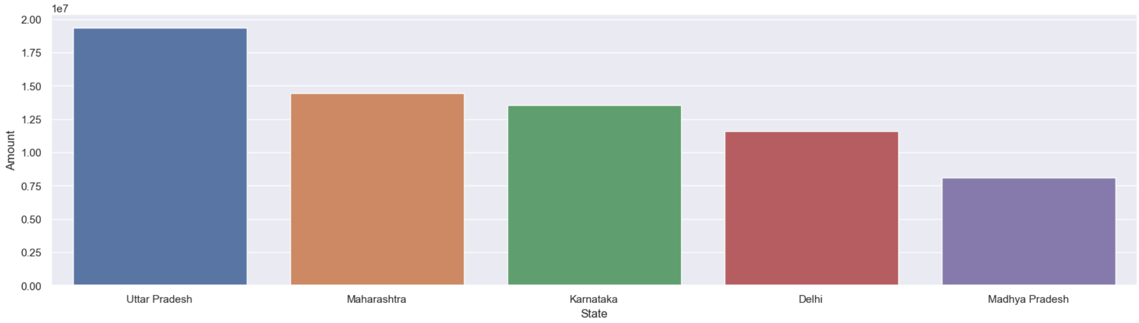
From above graph we can see that buyers who spend high amount working in IT SECTOER ,

In [17]:

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

Out[17]:

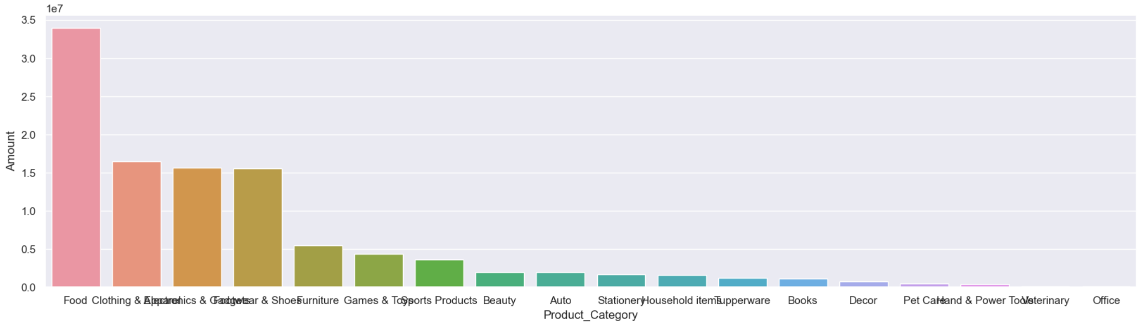
```
<AxesSubplot:xlabel='State', ylabel='Amount'>
```



From above graph we can see that higher amount spend by those who live in Uttar Pardesh , Maharashtra , Karanatka.

In [18]:

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



from above graph we can understand that people spend more on FOOD , CLOTHING and ELECTRONIC APPLIANCES

conclusion

Married women are our target audience. so marrird women who lives in Uttaer Pardesh , Maharashtra and Karnataka spend higher amount in food , clothing and electronic appliances , also people who worked in IT sector ,Healthcare and aviation are more likely to spend more.

In [ ]:

In [ ]: