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

In [2]:
```
df = pd.read_csv(r'C:\Users\BISHWJEET KUMAR\Downloads\Data Analysis\Diwali-Sales-Analysis-main\Diwali-Sales-Ana
```

In [148…
```
df.head(10)
```

Out[148…

	User_ID	Cust_name	Product_ID	Gender	Age Group	Age	Marital_Status	State	Zone	Occupation	Product_Category
0	1002903	Sanskriti	P00125942	F	26-35	28	0	Maharashtra	Western	Healthcare	Auto
1	1000732	Kartik	P00110942	F	26-35	35	1	Andhra Pradesh	Southern	Govt	Auto
2	1001990	Bindu	P00118542	F	26-35	35	1	Uttar Pradesh	Central	Automobile	Auto
3	1001425	Sudevi	P00237842	M	0-17	16	0	Karnataka	Southern	Construction	Auto
4	1000588	Joni	P00057942	M	26-35	28	1	Gujarat	Western	Food Processing	Auto
5	1000588	Joni	P00057942	M	26-35	28	1	Himachal Pradesh	Northern	Food Processing	Auto
6	1001132	Balk	P00018042	F	18-25	25	1	Uttar Pradesh	Central	Lawyer	Auto
7	1002092	Shivangi	P00273442	F	55+	61	0	Maharashtra	Western	IT Sector	Auto
8	1003224	Kushal	P00205642	M	26-35	35	0	Uttar Pradesh	Central	Govt	Auto
9	1003650	Ginny	P00031142	F	26-35	26	1	Andhra Pradesh	Southern	Media	Auto

In [3]:
```
df.drop(['Status','unnamed1'],axis=1,inplace=True)
```

In [4]:
```
df.loc[(df['Marital_Status']==0),'Relationship_status']= 'Single'
df.loc[(df['Marital_Status'] >= 1),'Relationship_status']= 'Married'
```

In [5]:
```
df['User_ID'].duplicated().sum()
```

Out[5]:
```
7496
```

In [ ]:
```
#ALL DUPLICATES USER_ID HAVE DIFFT CUST_NAME
```

In [6]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11251 entries, 0 to 11250
Data columns (total 14 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  Relationship_status  11251 non-null  object 
dtypes: float64(1), int64(4), object(9)
memory usage: 1.2+ MB
```

In [7]:
```
df.isnull().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
Relationship_status    0
dtype: int64
```

In [152…
```
print(df['Amount'].mean())
```

```
9453.610857727557
```

In [170…
```
df['Amount']=df['Amount'].replace(np.NaN,9453.61)
```

In [154…
```
df.head(30)
```

Out[154…

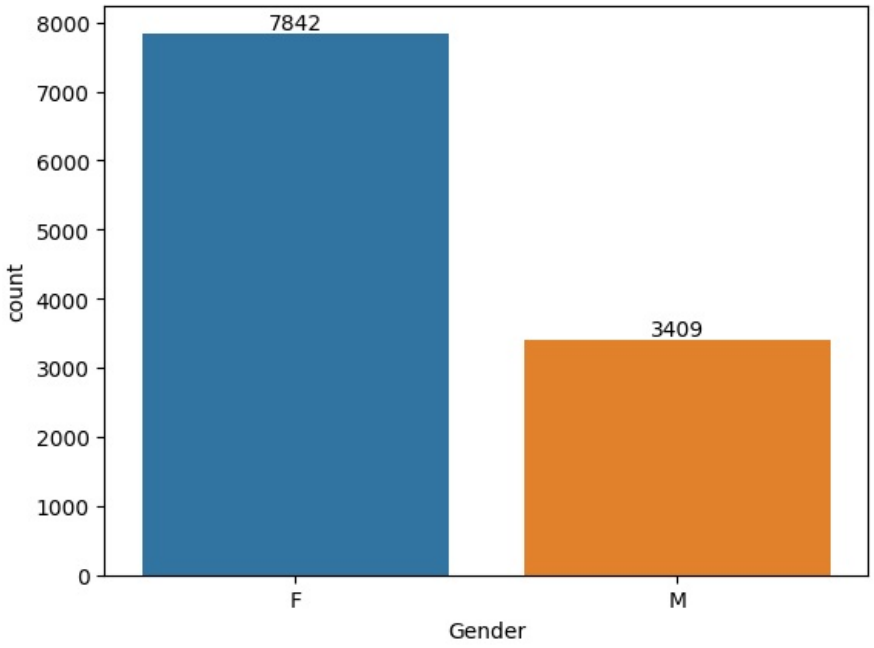
	User_ID	Cust_name	Product_ID	Gender	Age Group	Age	Marital_Status	State	Zone	Occupation	Product_Categor
0	1002903	Sanskriti	P00125942	F	26-35	28	0	Maharashtra	Western	Healthcare	Au
1	1000732	Kartik	P00110942	F	26-35	35	1	Andhra Pradesh	Southern	Govt	Au
2	1001990	Bindu	P00118542	F	26-35	35	1	Uttar Pradesh	Central	Automobile	Au
3	1001425	Sudevi	P00237842	M	0-17	16	0	Karnataka	Southern	Construction	Au
4	1000588	Joni	P00057942	M	26-35	28	1	Gujarat	Western	Food Processing	Au
5	1000588	Joni	P00057942	M	26-35	28	1	Himachal Pradesh	Northern	Food Processing	Au
6	1001132	Balk	P00018042	F	18-25	25	1	Uttar Pradesh	Central	Lawyer	Au
7	1002092	Shivangi	P00273442	F	55+	61	0	Maharashtra	Western	IT Sector	Au
8	1003224	Kushal	P00205642	M	26-35	35	0	Uttar Pradesh	Central	Govt	Au
9	1003650	Ginny	P00031142	F	26-35	26	1	Andhra Pradesh	Southern	Media	Au
10	1003829	Harshita	P00200842	M	26-35	34	0	Delhi	Central	Banking	Au
11	1000214	Kargatis	P00119142	F	18-25	20	0	Andhra Pradesh	Southern	Retail	Au
12	1004035	Elijah	P00080342	F	18-25	20	1	Andhra Pradesh	Southern	IT Sector	Au
13	1001680	Vasudev	P00324942	M	26-35	26	1	Andhra Pradesh	Southern	Automobile	Au
14	1003858	Cano	P00293742	M	46-50	46	1	Madhya Pradesh	Central	Hospitality	Au
15	1000813	Lauren	P00289942	F	18-25	24	0	Andhra Pradesh	Southern	Govt	Au
16	1005447	Amy	P00275642	F	46-50	48	1	Andhra Pradesh	Southern	IT Sector	Au
17	1001193	Mick	P00004842	F	26-35	29	0	Andhra Pradesh	Southern	Aviation	Au
18	1001883	Praneet	P00029842	M	51-55	54	1	Uttar Pradesh	Central	Hospitality	Au
19	1001883	Praneet	P00029842	M	51-55	54	1	Uttar Pradesh	Central	Hospitality	Au
20	1000113	Ellis	P00180642	F	18-25	19	1	Andhra Pradesh	Southern	Govt	Au
21	1000416	Hrisheekesh	P00181842	F	46-50	46	1	Uttar Pradesh	Central	Banking	Au
22	1005256	Grant	P00101742	F	26-35	30	0	Andhra Pradesh	Southern	IT Sector	Au
23	1001505	Gilcrest	P00271842	F	51-55	53	0	Uttar Pradesh	Central	Automobile	Au
24	1000900	Skaria	P00317842	M	55+	83	0	Karnataka	Southern	Automobile	Au
25	1005908	Eric	P00282642	F	26-35	33	0	Andhra Pradesh	Southern	IT Sector	Au
26	1001101	Gibson	P00234742	F	36-45	40	0	Uttar Pradesh	Central	Banking	Au
27	1004736	Mahima	P00058042	F	18-25	25	1	Andhra Pradesh	Southern	Banking	Au
28	1004037	Etezadi	P00190542	M	51-55	54	1	Andhra Pradesh	Southern	Govt	Hand & Pow Too
29	1002340	James	P00119642	F	36-45	39	1	Andhra Pradesh	Southern	Aviation	Au

In [155…
```
df['Amount']=df['Amount'].astype('int')
```

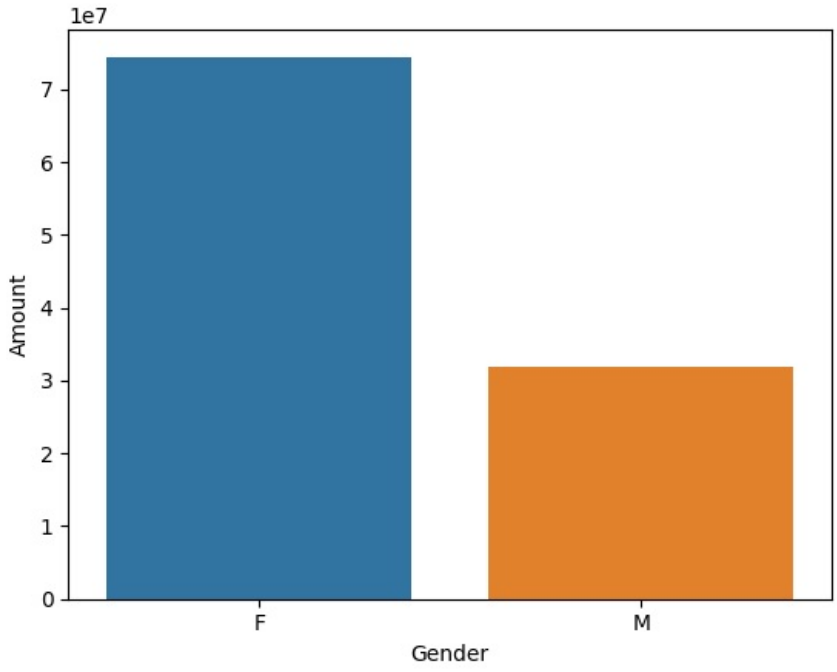
In [12]:
```
df.columns
```

Out[12]:
```
Index(['User_ID', 'Cust_name', 'Product_ID', 'Gender', 'Age Group', 'Age',
       'Marital_Status', 'State', 'Zone', 'Occupation', 'Product_Category',
       'Orders', 'Amount', 'Status', 'unnamed1', 'Relationship_status'],
      dtype='object')
```

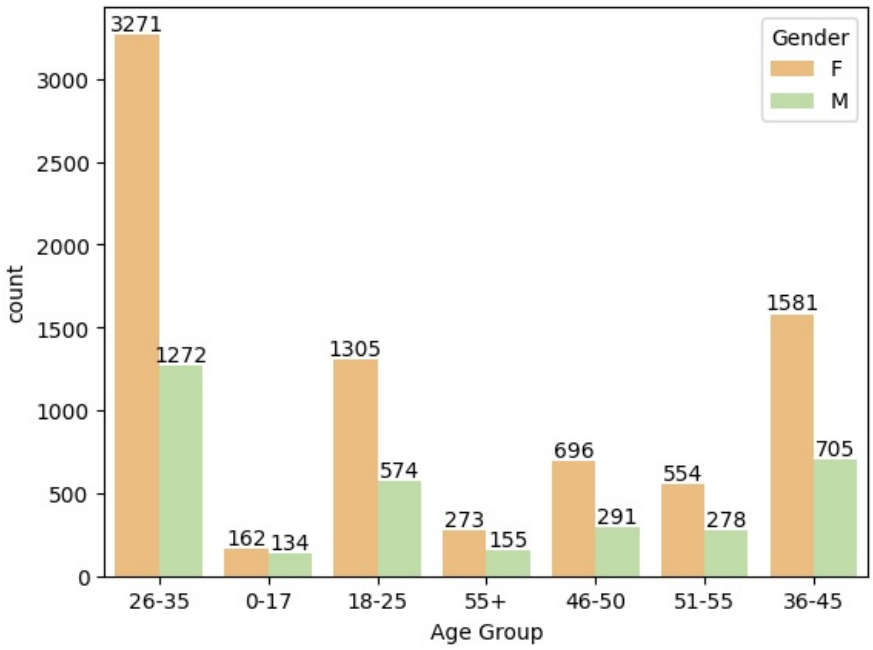
In [20]:
```
s= sns.countplot(x='Gender',data=df,hue='Gender')
for bars in s.containers:
    s.bar_label(bars)
```



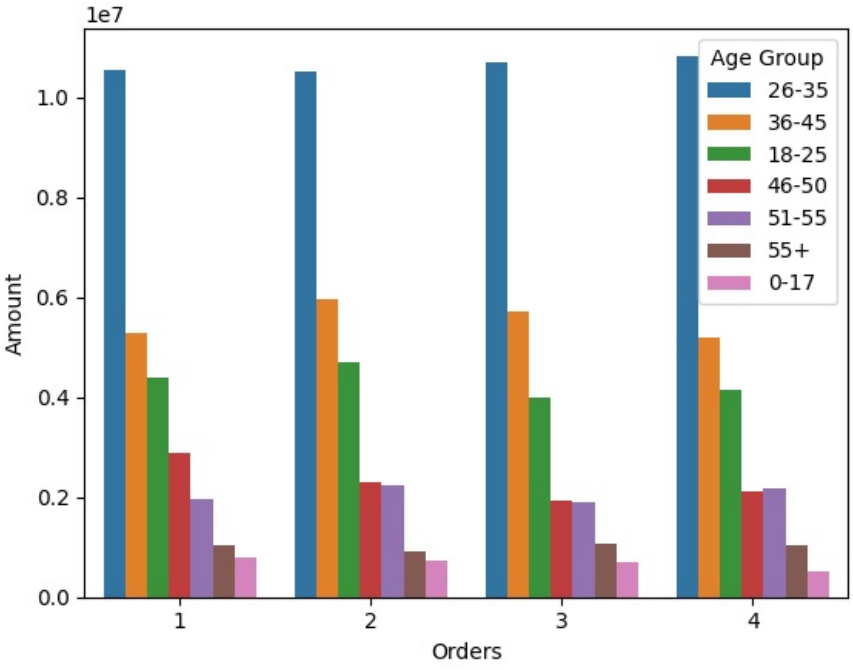
In [16]:
```
Sales_by_Gen=df.groupby(['Gender'],as_index=False)['Amount'].sum().sort_values(by='Amount',ascending=False)
sns.barplot(x= 'Gender',y= 'Amount',data=Sales_by_Gen,hue='Gender')
plt.show()
```



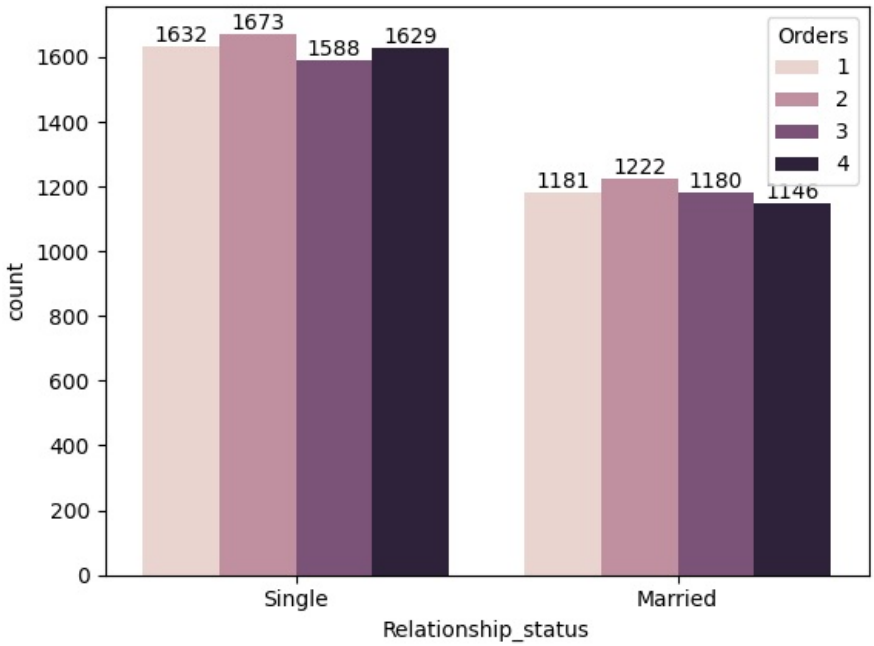
In [17]:
```
s= sns.countplot(data= df,x='Age Group',hue='Gender',palette='Spectral')
for bars in s.containers:
    s.bar_label(bars)
```



In [102…
```
Sales_by_Ord= df.groupby(['Age Group','Orders'],as_index=False)['Amount'].sum().sort_values(by='Amount',ascendi
sns.barplot(x='Orders',y='Amount',data=Sales_by_Ord,hue='Age Group')
plt.show()
```

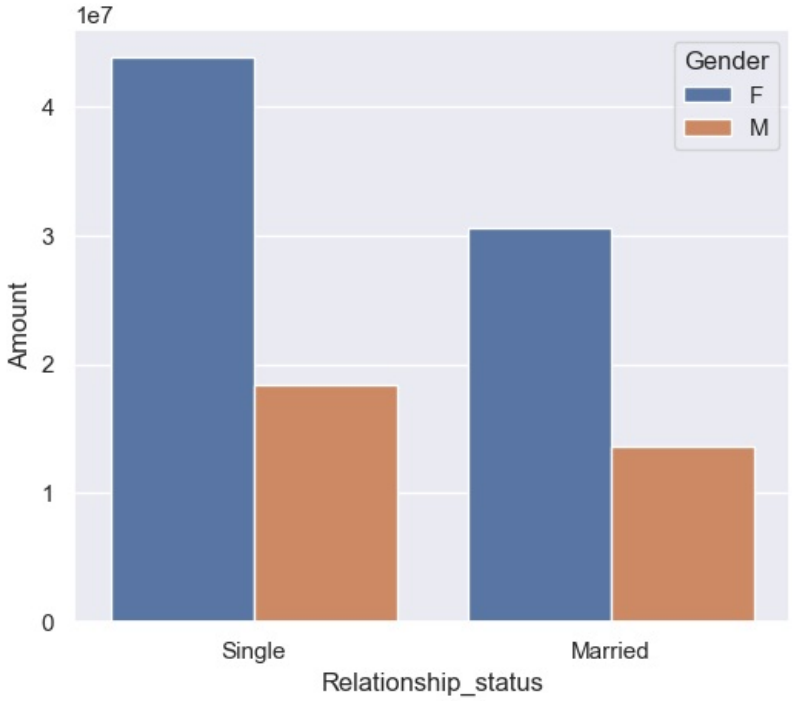


In [18]:
```
s= sns.countplot(x='Relationship_status',data=df,hue='Orders')
for bars in s.containers:
    s.bar_label(bars)
```

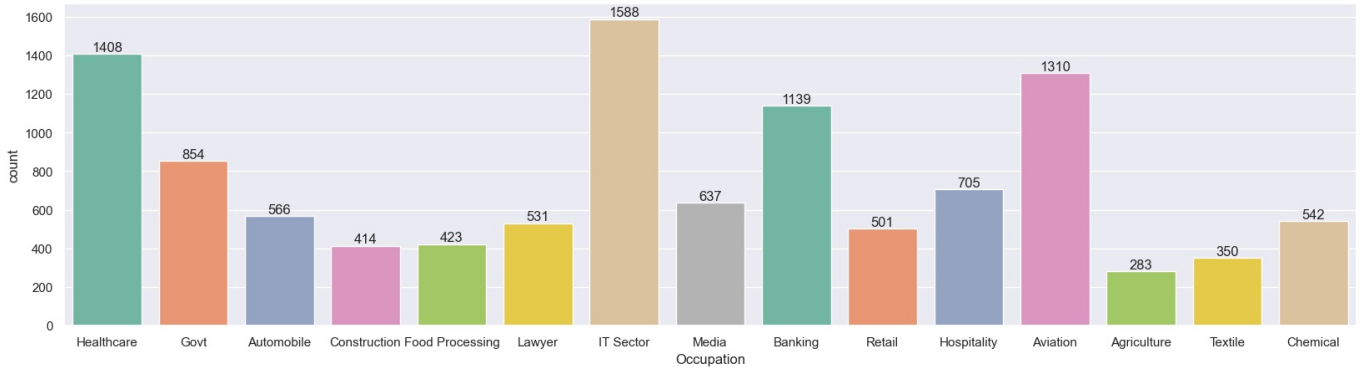


In [19]:
```
Reln_sales= df.groupby(['Relationship_status','Gender'],as_index=False)['Amount'].sum().sort_values(by='Amount'
sns.set(rc={'figure.figsize':(6,5)})
sns.barplot(data=Reln_sales,x='Relationship_status',y='Amount',hue= 'Gender')
```

Out[19]: <Axes: xlabel='Relationship_status', ylabel='Amount'>

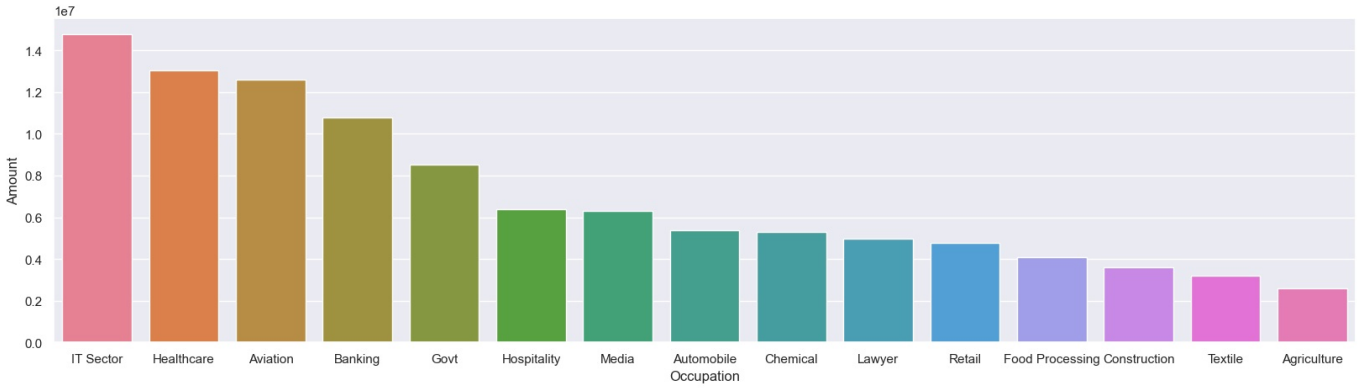


In [50]:
```
sns.set_palette(["red", "green", "blue", "yellow", "purple"])
s=sns.countplot(data=df, x= 'Occupation',palette='Set2',hue='Occupation')
sns.set(rc={'figure.figsize':(20,5)})
for bars in s.containers:
    s.bar_label(bars)
```



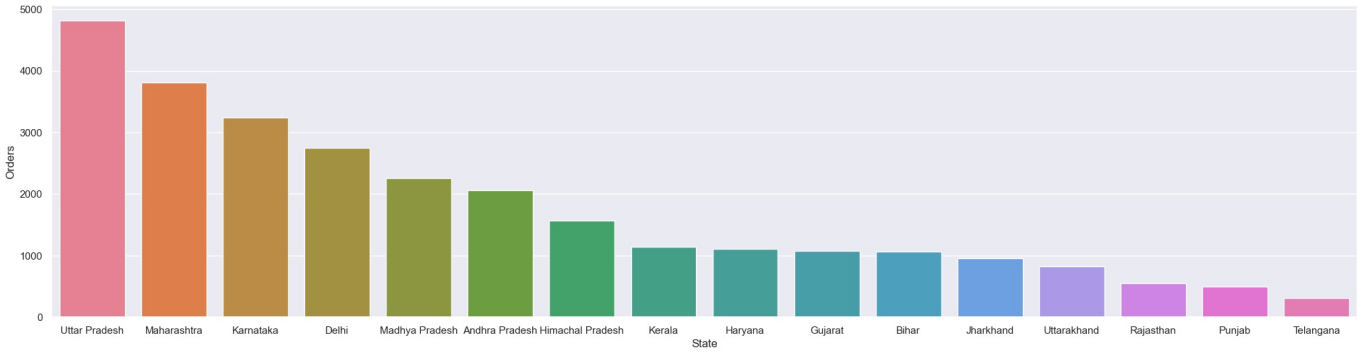
In [53]:
```
Sales_by_Occup= df.groupby(['Occupation'],as_index=False)['Amount'].sum().sort_values(by='Amount',ascending=Fals
sns.barplot(data=Sales_by_Occup,x='Occupation',y='Amount',hue='Occupation')
```

```
sns.set(rc={'figure.figsize':(20,5)})
```



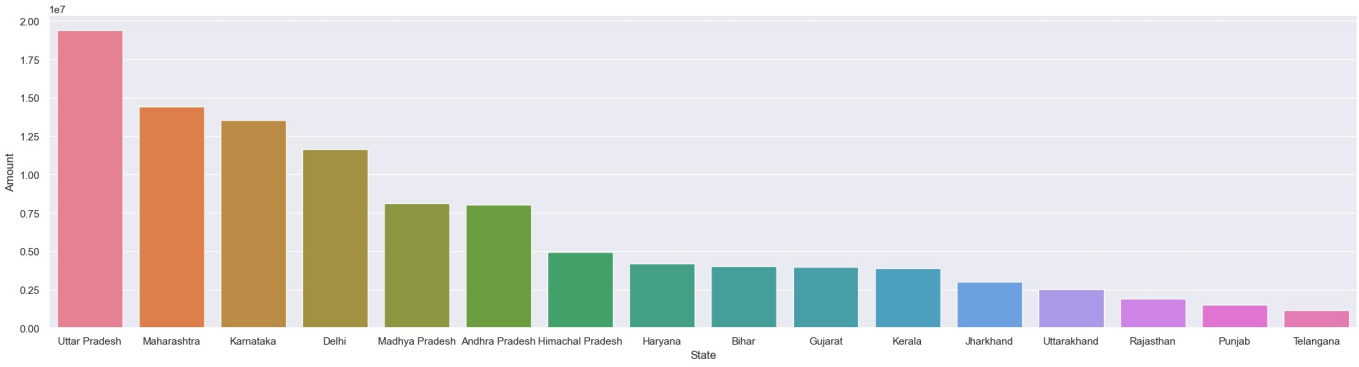
In [86]:
```
Order_by_State =df.groupby(['State'],as_index=False)['Orders'].sum().sort_values(by='Orders',ascending=False)
sns.set(rc={'figure.figsize':(25,6)})
sns.barplot(data=Order_by_State,x='State',y='Orders',hue='State')
```

Out[86]: <Axes: xlabel='State', ylabel='Orders'>

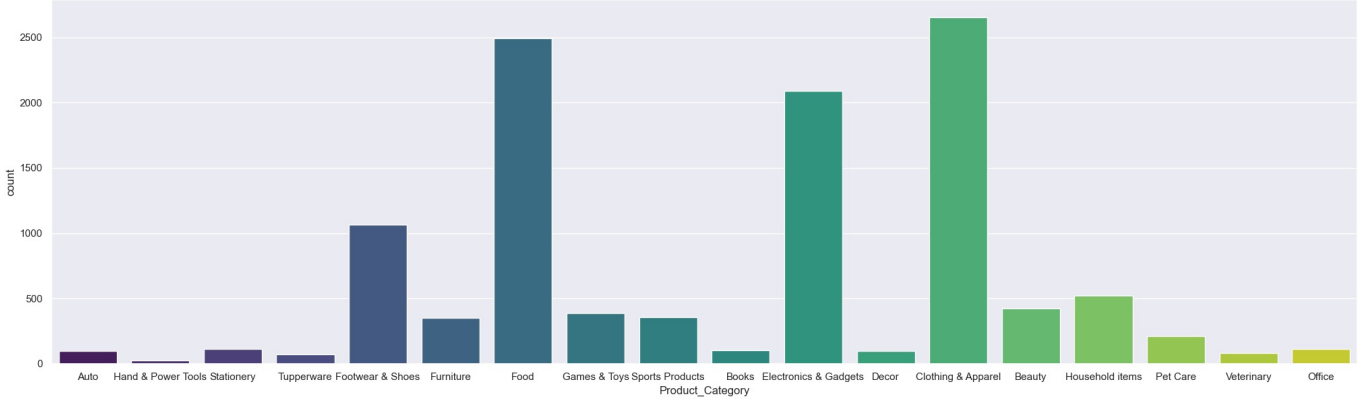


In [85]:
```
Sales_by_State =df.groupby(['State'],as_index=False)['Amount'].sum().sort_values(by='Amount',ascending=False)
sns.set(rc={'figure.figsize':(25,6)})
sns.barplot(data=Sales_by_State,x='State',y='Amount',hue='State')
```

Out[85]: <Axes: xlabel='State', ylabel='Amount'>

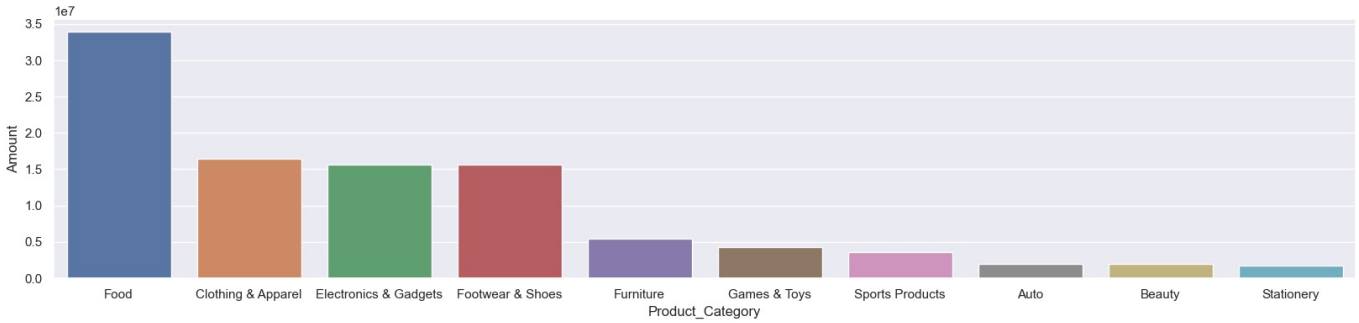


In [100…
```
sns.set(rc={'figure.figsize':(25,7)})
s= sns.countplot(data=df,x='Product_Category',hue='Product_Category',palette='viridis')
```



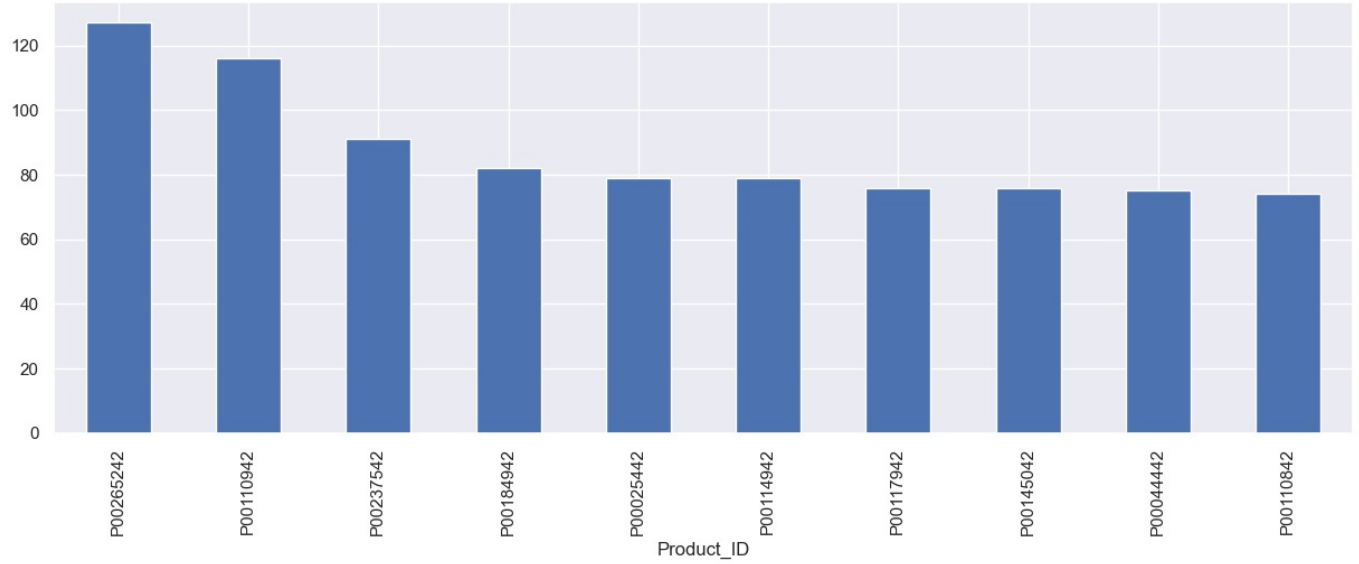
In [118…
```
Sales_by_Product =df.groupby(['Product_Category'],as_index=False)['Amount'].sum().sort_values(by='Amount',ascen
sns.set(rc={'figure.figsize':(20,4)})
sns.barplot(data= Sales_by_Product,x='Product_Category',y='Amount',hue='Product_Category')
```

Out[118… <Axes: xlabel='Product_Category', ylabel='Amount'>



In [122…
```
fig1,s= plt.subplots(figsize=(15,5))
df.groupby('Product_ID')['Orders'].sum().nlargest(10).sort_values(ascending=False).plot(kind='bar')
```

Out[122… <Axes: xlabel='Product_ID'>



In [ ]:

In [ ]: