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

In [2]:
```
df=pd.read_csv(r"C:\Users\Pratik patil\Downloads\Diwali Sales 1.csv",encoding='unic
```

In [29]:
```
df.head(10)
```

Out[29]:

	User_ID	Cust_name	Product_ID	Gender	Age Group	Age	Marital_Status	State	Zor
0	1002903	Sanskriti	P00125942	F	26-35	28	0	Maharashtra	Weste
1	1000732	Kartik	P00110942	F	26-35	35	1	Andhra Pradesh	Southe
2	1001990	Bindu	P00118542	F	26-35	35	1	Uttar Pradesh	Centr
3	1001425	Sudevi	P00237842	M	0-17	16	0	Karnataka	Southe
4	1000588	Joni	P00057942	M	26-35	28	1	Gujarat	Weste
5	1000588	Joni	P00057942	M	26-35	28	1	Himachal Pradesh	Northe
6	1001132	Balk	P00018042	F	18-25	25	1	Uttar Pradesh	Centr
8	1003224	Kushal	P00205642	M	26-35	35	0	Uttar Pradesh	Centr
9	1003650	Ginny	P00031142	F	26-35	26	1	Andhra Pradesh	Southe
10	1003829	Harshita	P00200842	M	26-35	34	0	Delhi	Centr

In [5]:
```
df.shape
```

Out[5]:
```
(11251, 15)
```

In [6]:
```
df.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
```

In [7]:
```
df.drop(["Status","unnamed1"],axis=1,inplace=True)
```

In [8]:
```
df.info()
```

```
<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 [10]:
```
df.isnull().sum()
```

Out[10]:
```
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 [14]:
```
df.dropna(inplace=True)
```

In [15]:
```
df.isnull().sum()
```

Out[15]:
```
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
dtype: int64
```

In [16]:
```
df.info()
```

```
<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 [23]:
```
df["Amount"] = pd.to_numeric(df["Amount"], errors='coerce').fillna(0).astype('int64
```

In [24]:
```
df["Amount"].dtype
```

Out[24]:
```
dtype('int64')
```

In [25]:
```
df.columns
```

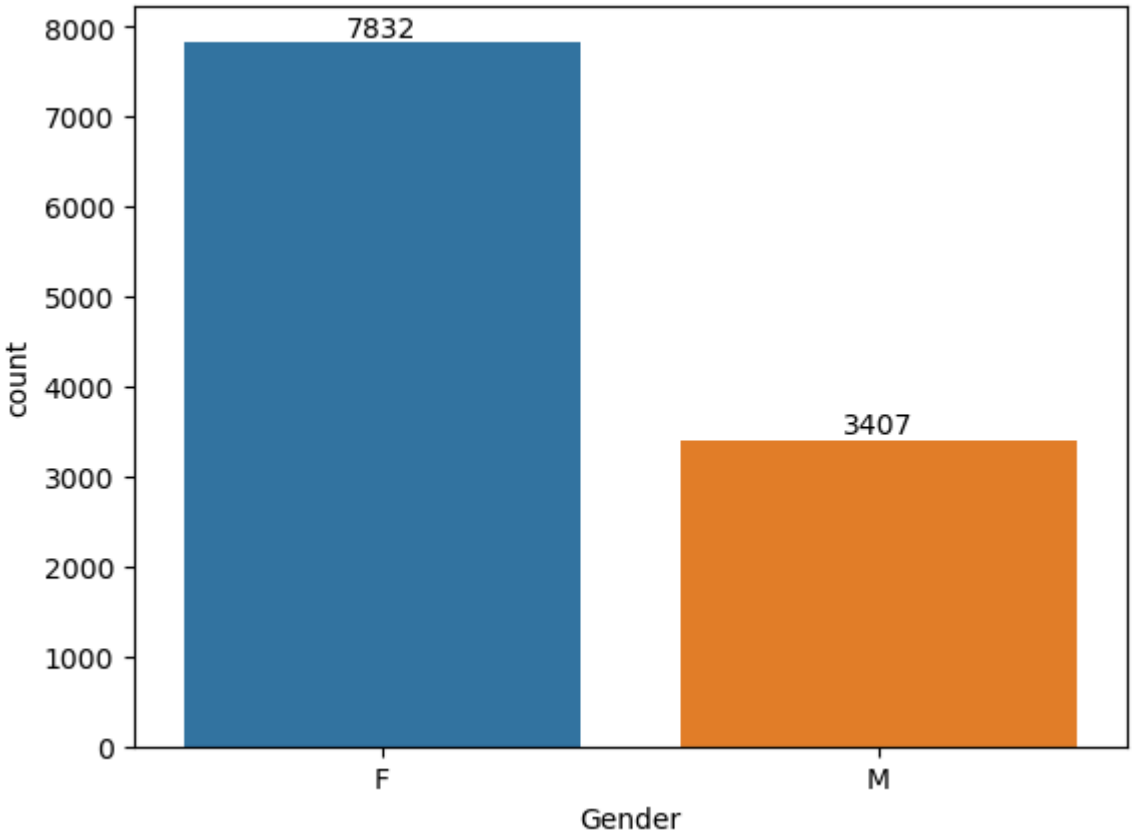
Out[25]:
```
Index(['User_ID', 'Cust_name', 'Product_ID', 'Gender', 'Age Group', 'Age',
       'Marital_Status', 'State', 'Zone', 'Occupation', 'Product_Category',
       'Orders', 'Amount'],
      dtype='object')
```

In [32]:
```
df[["Age","Orders"]].describe()
```

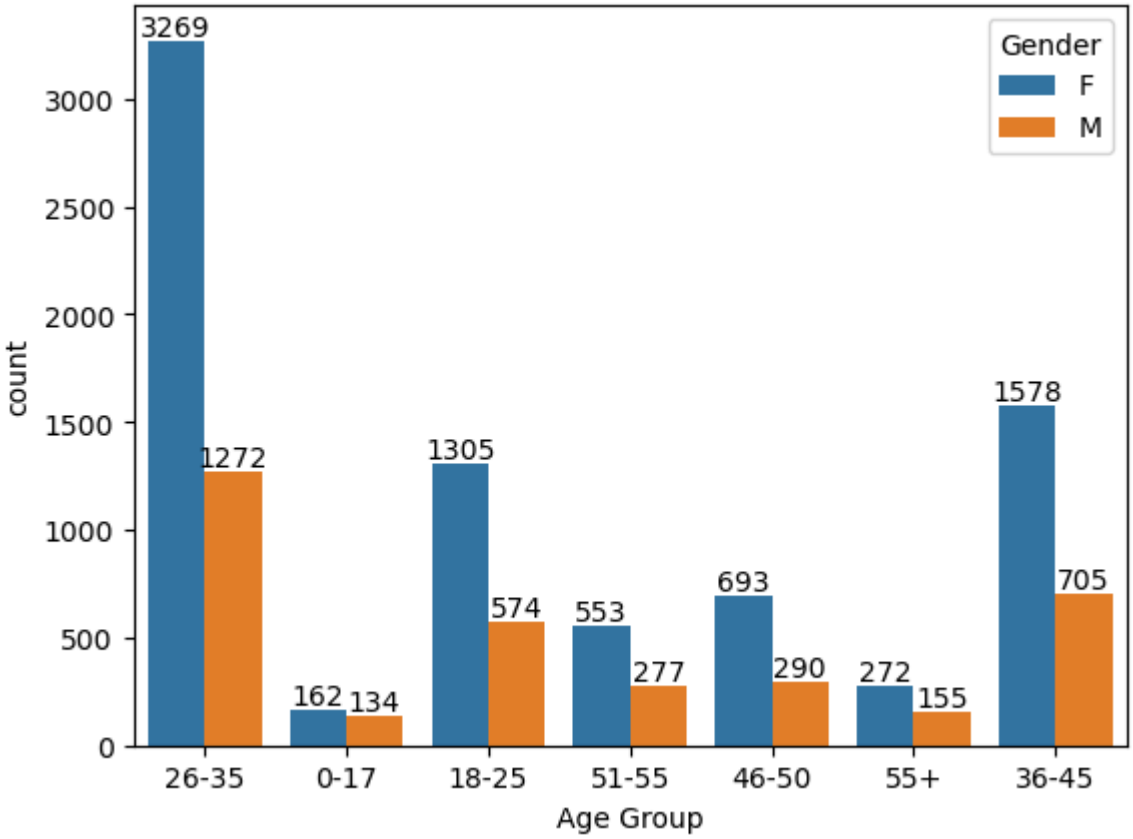
Out[32]:

	Age	Orders
count	11239.000000	11239.000000
mean	35.410357	2.489634
std	12.753866	1.114967
min	12.000000	1.000000
25%	27.000000	2.000000
50%	33.000000	2.000000
75%	43.000000	3.000000
max	92.000000	4.000000

In [33]:
```
pop=sb.countplot(x='Gender',data=df)
for bars in pop.containers:
    pop.bar_label(bars)
```

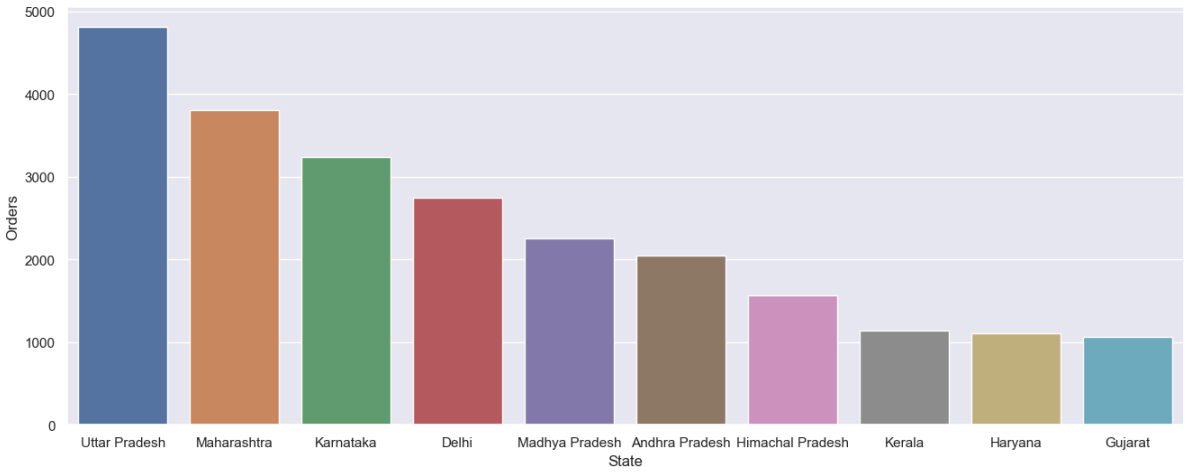


In [36]:
```
#age group and gender count
agm=sb.countplot(x='Age Group',data=df,hue='Gender')
for bars in agm.containers:
    agm.bar_label(bars)
```

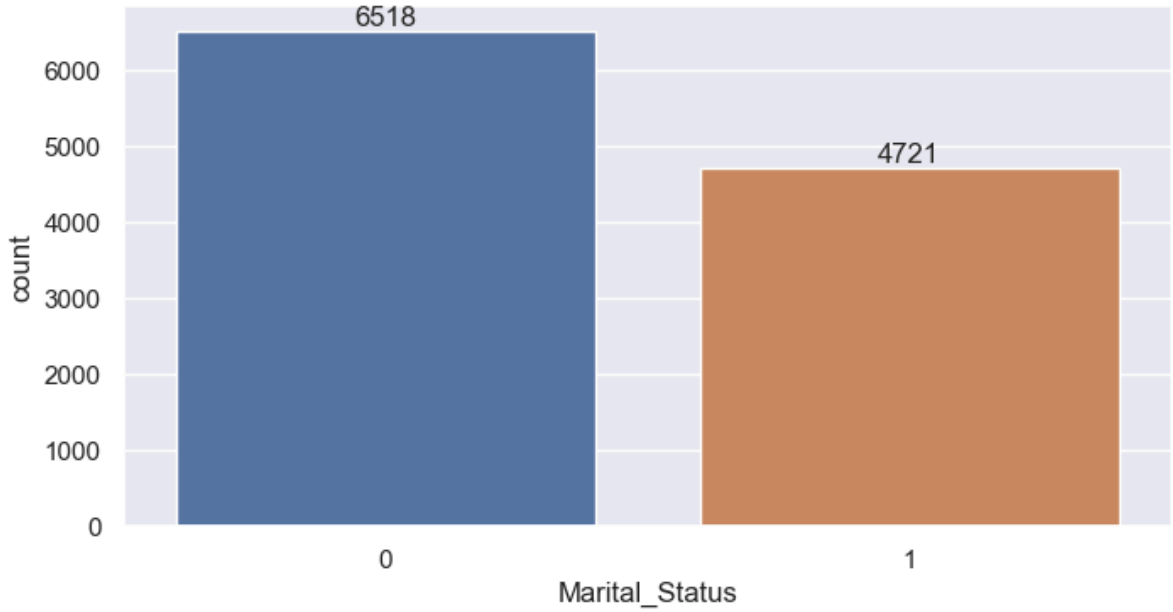


In [41]:
```
#statewise Sales
sales_state=df.groupby(['State'],as_index=False)['Orders'].sum().sort_values(by='Or
sb.set(rc={'figure.figsize':(16,6)})
sb.barplot(x='State',y='Orders',data=sales_state)
```

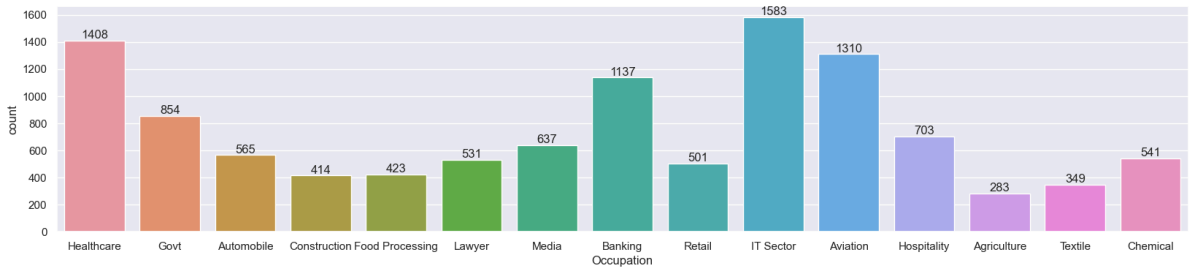
Out[41]:
```
<Axes: xlabel='State', ylabel='Orders'>
```



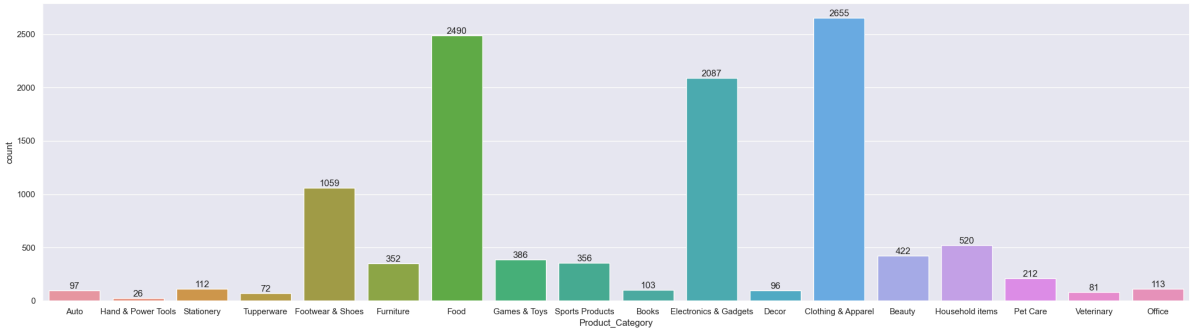
In [60]:
```
MS=sb.countplot(x='Marital_Status',data=df)
for bars in MS.containers:
    MS.bar_label(bars)
sb.set(rc={"figure.figsize":(4,8)})
```



In [63]:
```
occ=sb.countplot(x='Occupation',data=df)
for bars in occ.containers:
    occ.bar_label(bars)
sb.set(rc={'figure.figsize':(20,6)})
```



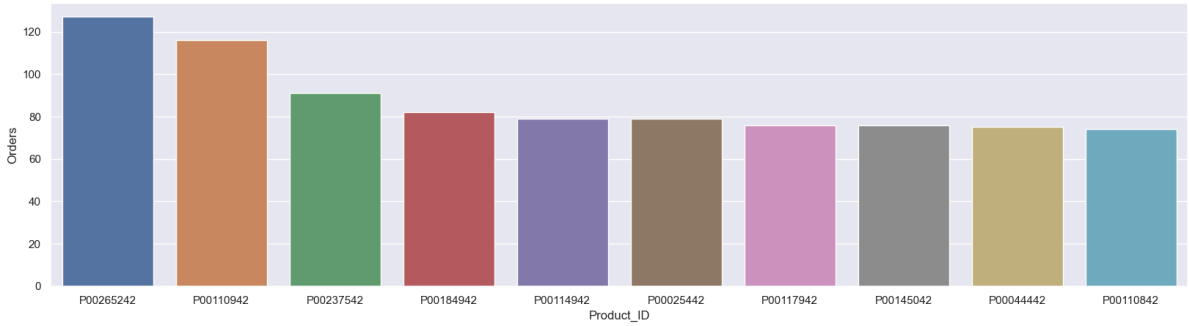
In [76]:
```
pc=sb.countplot(x='Product_Category',data=df)
for bars in pc.containers:
    pc.bar_label(bars)
sb.set(rc={'figure.figsize':(27,7)})
```



In [77]:
```
sales_prodid=df.groupby(["Product_ID"],as_index=False)["Orders"].sum().sort_values(
```

In [78]:
```
sb.set(rc={'figure.figsize':(20,5)})
sb.barplot(data=sales_prodid,x="Product_ID",y="Orders")
```

Out[78]:
```
<Axes: xlabel='Product_ID', ylabel='Orders'>
```



Coclusion

Married Women age group 26-35 yrs from UP,Maharashtra and Karnataka Working in IT.

In [ ]: