In [30]:
```
import pandas as pd
import numpy as mp
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
```

In [31]:
```
df = pd.read_csv(r'C:\Users\Hanzala\Downloads\Python_Diwali_Sales_Analysis-main\
```

In [32]:
```
df.head()
```

Out[32]:

	User_ID	Cust_name	Product_ID	Gender	Age Group	Age	Marital_Status	Stat
0	1002903	Sanskriti	P00125942	F	26-35	28	0	Maharashtı
1	1000732	Kartik	P00110942	F	26-35	35	1	Andhra Prades
2	1001990	Bindu	P00118542	F	26-35	35	1	Uttar Prades
3	1001425	Sudevi	P00237842	M	0-17	16	0	Karnatak
4	1000588	Joni	P00057942	M	26-35	28	1	Gujara

In [33]:
```
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 [34]:
```
df.drop(['Status','unnamed1'],axis=1, inplace=True)
```

In [35]:
```
df.head()
```

Out[35]:

	User_ID	Cust_name	Product_ID	Gender	Age Group	Age	Marital_Status	Stat
0	1002903	Sanskriti	P00125942	F	26-35	28	0	Maharashtı
1	1000732	Kartik	P00110942	F	26-35	35	1	Andhra Prades
2	1001990	Bindu	P00118542	F	26-35	35	1	Uttar Prades
3	1001425	Sudevi	P00237842	M	0-17	16	0	Karnatak
4	1000588	Joni	P00057942	M	26-35	28	1	Gujara

In [36]:
```
df.isnull().sum()
```

Out[36]:
```
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 [37]:
```
df.dropna(inplace=True)
```

In [38]:
```
df.isnull().sum()
```

Out[38]:
```
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 [39]:
```
df.columns
```

Out[39]:
```
Index(['User_ID', 'Cust_name', 'Product_ID', 'Gender', 'Age Group', 'Age',
       'Marital_Status', 'State', 'Zone', 'Occupation', 'Product_Category',
       'Orders', 'Amount'],
      dtype='object')
```

In [94]:
```
change= df.rename(columns={'State': 'New States'})
```

In [95]:
```
change.head()
```

Out[95]:

	User_ID	Cust_name	Product_ID	Gender	Age Group	Age	Marital_Status	New State
0	1002903	Sanskriti	P00125942	F	26-35	28	No	Maharashtı
1	1000732	Kartik	P00110942	F	26-35	35	No	Andhra Prades
2	1001990	Bindu	P00118542	F	26-35	35	No	Uttar Prades
3	1001425	Sudevi	P00237842	M	0-17	16	No	Karnatak
4	1000588	Joni	P00057942	M	26-35	28	No	Gujara

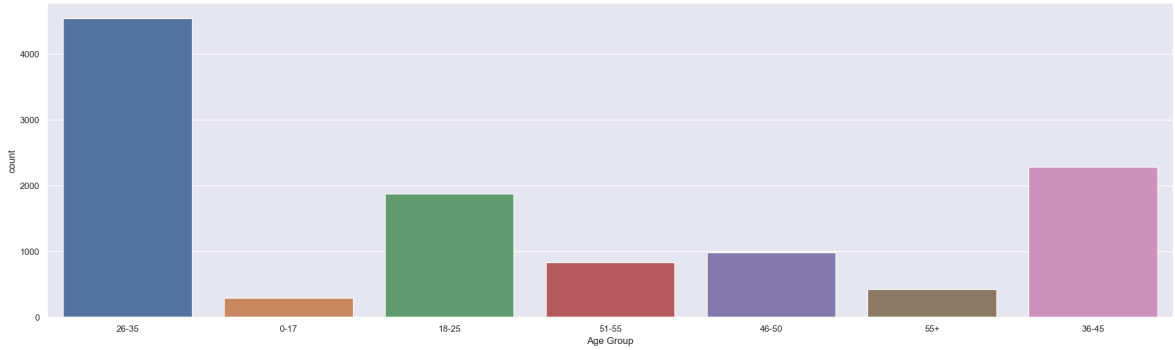
In [42]:
```
df.isnull().sum()
```

Out[42]:
```
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
```

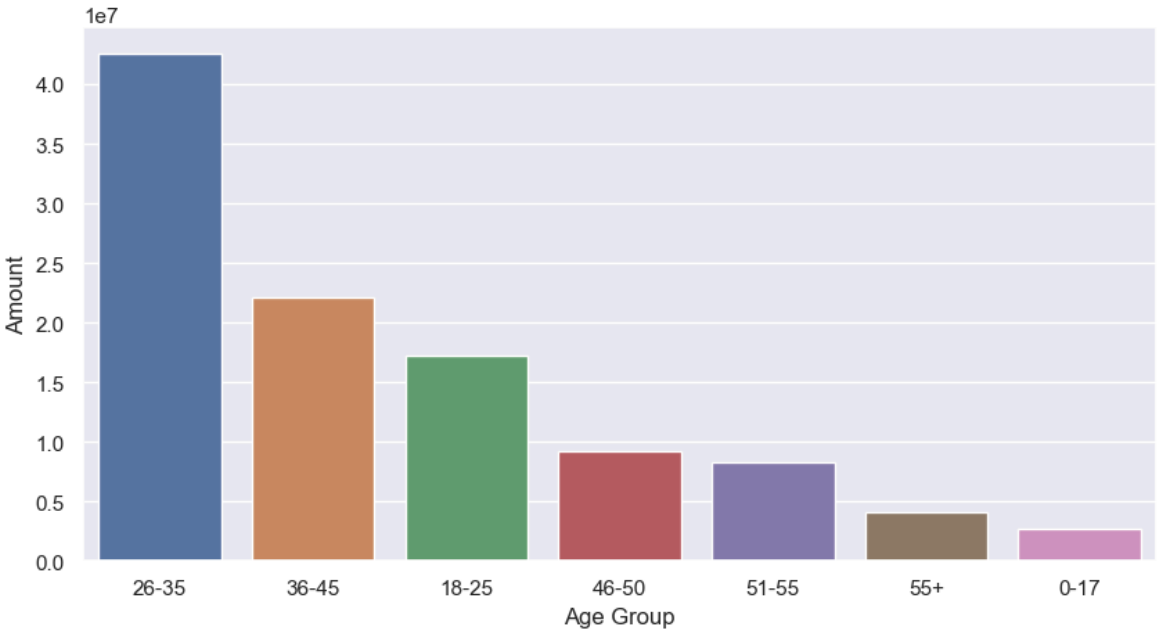
Exploratory Data Analysis

Age Group

In [43]:
```
# ORDERS BY AGE GROUP

graph = sns.countplot(x='Age Group', data=df, hue='Age Group')
sns.set(rc = {'figure.figsize':(10,5)})
plt.show()
```



In [44]:
```
ag_sales = df.groupby(['Age Group'],as_index = False) ['Amount'].sum().sort_valu
sns.set(rc = {'figure.figsize':(10,5)})
sns.barplot(data = ag_sales, x ='Age Group', y = 'Amount', hue ='Age Group')
plt.show()
```



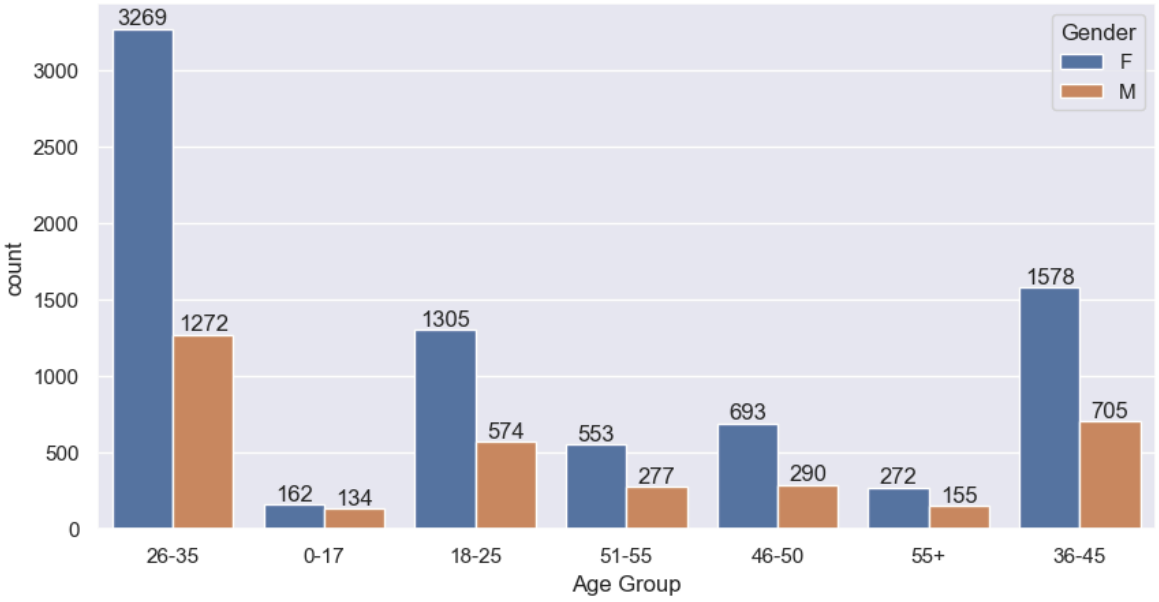
In [45]:
```
# From Above Graph Most Of The Buyers are Between 26-35 Age.
```

Order by Age-Group And Gender

In [46]:
```
df.columns
```

Out[46]:
```
Index(['User_ID', 'Cust_name', 'Product_ID', 'Gender', 'Age Group', 'Age',
       'Marital_Status', 'State', 'Zone', 'Occupation', 'Product_Category',
       'Orders', 'Amount'],
      dtype='object')
```

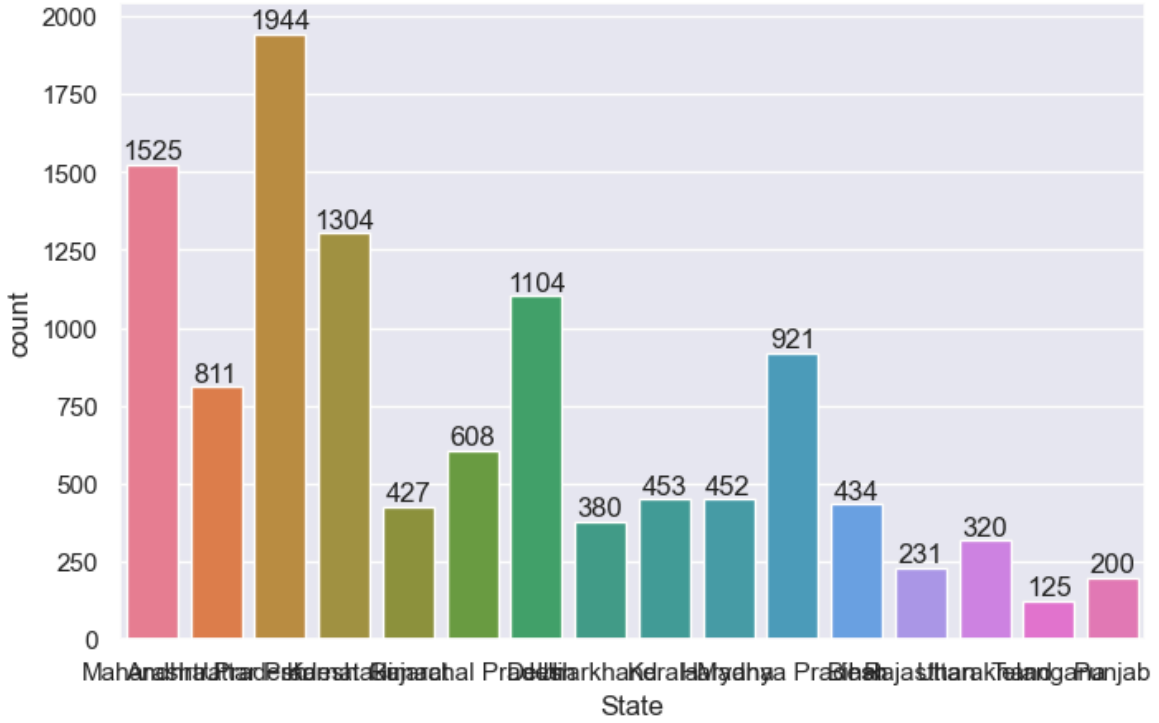
In [47]:
```
gen = sns.countplot(x= 'Age Group', data=df, hue='Gender')
sns.set(rc = {'figure.figsize':(8,5)})

for bars in gen.containers:
    gen.bar_label(bars)

plt.show()
```



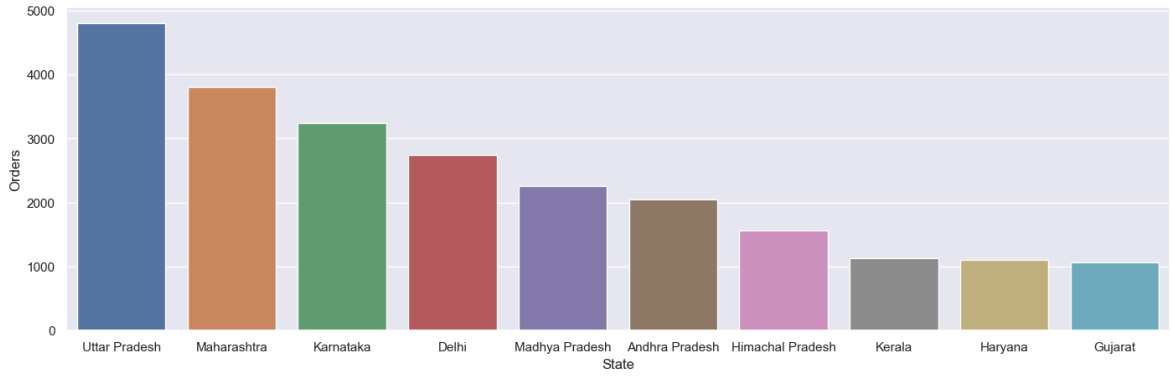
In [48]:
```
# From Above Graph Where Most Of Buyers Are Females And The Age Group is Between
```

State

In [49]:
```
sta = sns.countplot(x = 'State', data = df, hue='State')
sns.set(rc = {'figure.figsize':(18,5)})
for bars in sta.containers:
    sta.bar_label(bars)
plt.show()
```



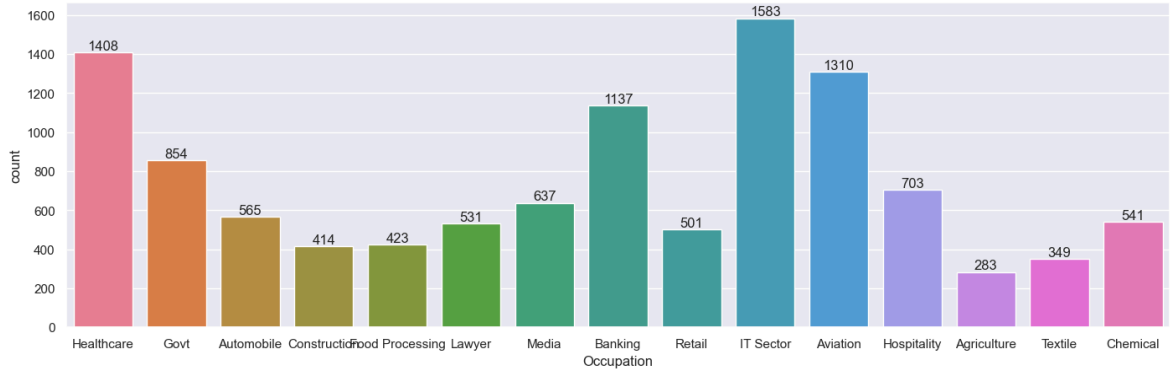
In [50]:
```
top_sta = df.groupby(['State'], as_index = False) ['Orders'].sum().sort_values(b
sns.set(rc = {'figure.figsize':(17,5)})
sns.barplot(data = top_sta, x = 'State', y = 'Orders', hue = 'State')
plt.show()
```



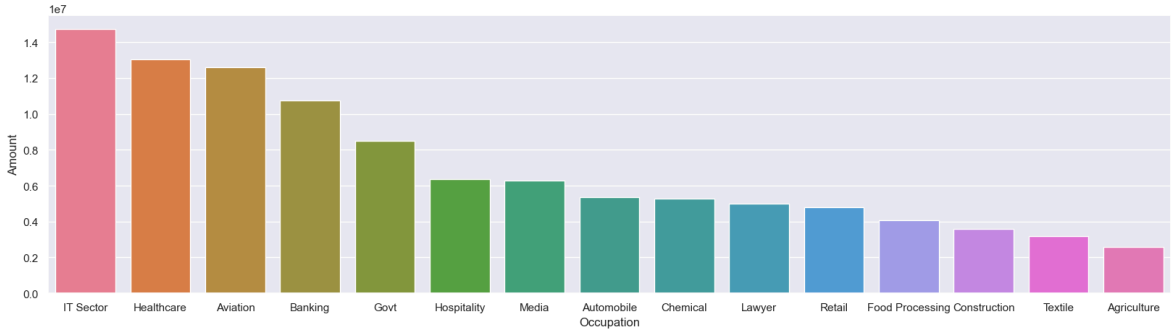
In [51]:
```
# "The first graph displays the number of orders per state."

# "The second graph displays the top 10 states by total Orders."
```

Occupation

In [52]:
```
occ = sns.countplot(data = df, x = 'Occupation', hue = 'Occupation')
for bars in occ.containers:
    occ.bar_label(bars)
plt.show()
```



In [53]:
```
occ_amt = df.groupby(['Occupation'],as_index = False) ['Amount'].sum().sort_valu
sns.set(rc = {'figure.figsize':(20,5)})
sns.barplot(data = occ_amt, x = 'Occupation', y = 'Amount', hue = 'Occupation')
plt.show()
```



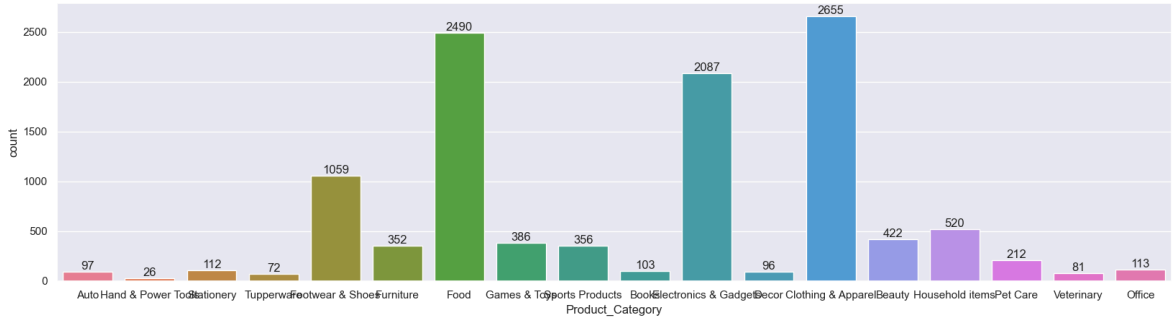
In [54]:
```
# 1 Graph of Occupation :- The first graph illustrates sales distribution by occ
# 2 Graph of Occupation :- The second graph show that the most of buyers belong
```

In [55]:
```
df.columns
```

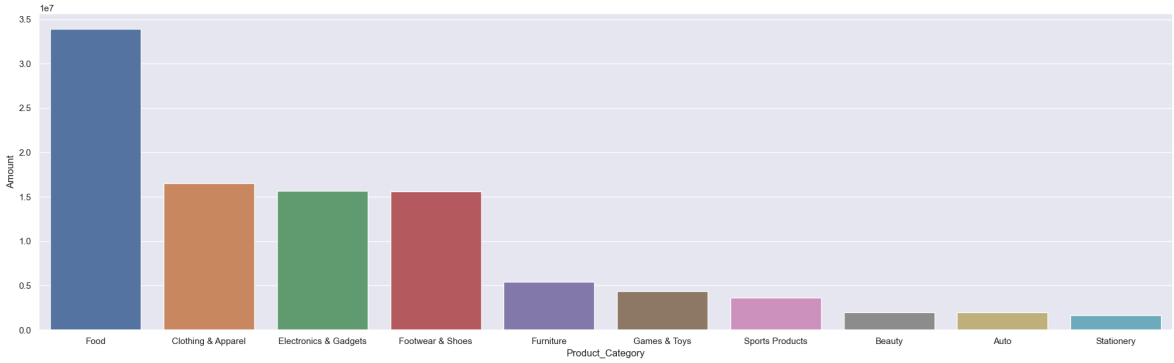
Out[55]:
```
Index(['User_ID', 'Cust_name', 'Product_ID', 'Gender', 'Age Group', 'Age',
       'Marital_Status', 'State', 'Zone', 'Occupation', 'Product_Category',
       'Orders', 'Amount'],
      dtype='object')
```

Product_Category

In [56]:
```
pd = sns.countplot(x = 'Product_Category', data = df, hue='Product_Category')
sns.set(rc={'figure.figsize':(25,7)})
for bars in pd.containers:
    pd.bar_label(bars)

plt.show()
```



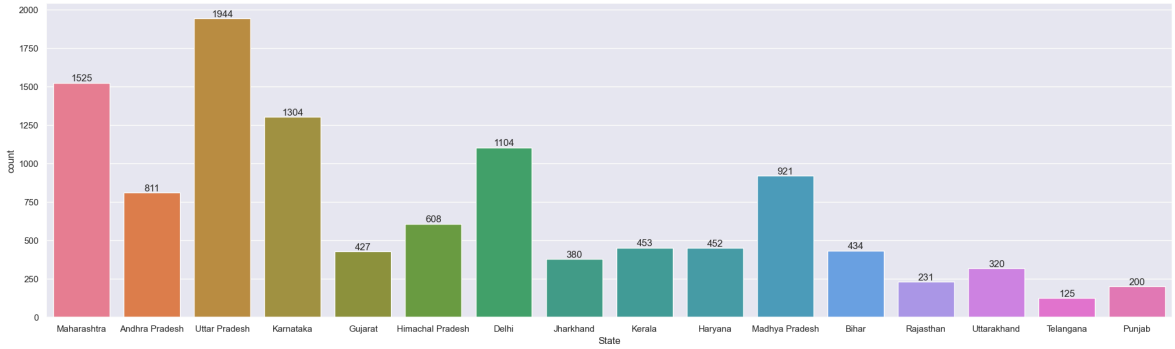
In [62]:
```
pd_sales = df.groupby(['Product_Category'],as_index = False) ['Amount'].sum().so
sns.barplot(data = pd_sales, x = 'Product_Category', y = 'Amount', hue = 'Produc
sns.set(rc={'figure.figsize':(25,5)})
plt.show()
```



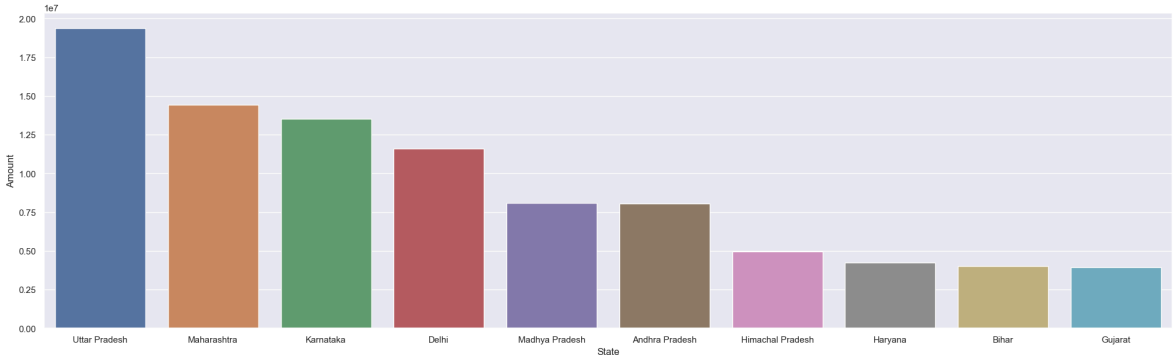
In [ ]:
```
# 1 Graph of Product Category :- The first graph shows the sales distribution ac
# 2 Graph of Product Category :- The second graph indicates that the highest sal
```

Sales By State

In [59]:
```
st = sns.countplot(data = df , x ='State', hue = 'State')
for bars in st.containers:
    st.bar_label(bars)
plt.show()
```



In [60]:
```
sta_amt = df.groupby(['State'],as_index = False) ['Amount'].sum().sort_values(by
sns.barplot(data = sta_amt, x = 'State', y = 'Amount', hue = 'State')
plt.show()
```



In [64]:
```
df.isnull().sum().sum()
```

Out[64]: 0