In [31]:
```python
1  import numpy as np
2  import pandas as pd
3  import matplotlib.pyplot as plt
4  %matplotlib inline
5  import seaborn as sns
6
7  df = pd.read_csv("Diwali Sales Data.csv", encoding='unicode_escape')
```

In [32]:
```python
1  df.shape
```

Out[32]: (11251, 15)

In [33]:
```python
1  df.head()
```

Out[33]:

	User_ID	Cust_name	Product_ID	Gender	Age Group	Age	Marital_Status	State	Zone	Occupation	Produc
0	1002903	Sanskriti	P00125942	F	26-35	28	0	Maharashtra	Western	Healthcare	
1	1000732	Kartik	P00110942	F	26-35	35	1	Andhra Pradesh	Southern	Govt	
2	1001990	Bindu	P00118542	F	26-35	35	1	Uttar Pradesh	Central	Automobile	
3	1001425	Sudevi	P00237842	M	0-17	16	0	Karnataka	Southern	Construction	
4	1000588	Joni	P00057942	M	26-35	28	1	Gujarat	Western	Food Processing	

In [34]:
```python
1  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 [35]:
```python
1  # dropo unrelated/blank columns
2  df.drop(['Status','unnamed1'],axis=1,inplace=True)
```

In [36]:
```python
1  # check for null values
2  pd.isnull(df).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]:
```python
1  # drop null values
2  df.dropna(inplace=True)
```

In [38]:
```python
1  # Change datatype
2  df['Amount'] = df['Amount'].astype('int')
```

In [39]:
```python
1  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 [40]:
```python
1  # describe() method returns description of the data in the DataFrame (i.e.
   count,mean,std,etc)
2  df.describe()
```
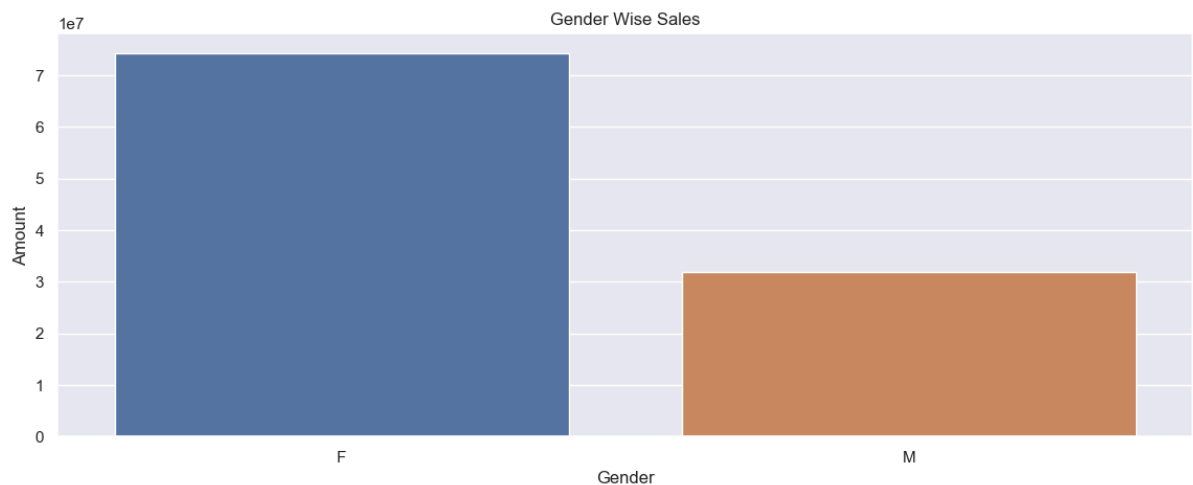
Out[40]:

	User_ID	Age	Marital_Status	Orders	Amount
count	1.123900e+04	11239.000000	11239.000000	11239.000000	11239.000000
mean	1.003004e+06	35.410357	0.420055	2.489634	9453.610553
std	1.716039e+03	12.753866	0.493589	1.114967	5222.355168
min	1.000001e+06	12.000000	0.000000	1.000000	188.000000
25%	1.001492e+06	27.000000	0.000000	2.000000	5443.000000
50%	1.003064e+06	33.000000	0.000000	2.000000	8109.000000
75%	1.004426e+06	43.000000	1.000000	3.000000	12675.000000
max	1.006040e+06	92.000000	1.000000	4.000000	23952.000000

Exploratory Data Analysis

Gender

In [41]:
```python
1  sales_gen = df.groupby(['Gender'], as_index=False)
   ['Amount'].sum().sort_values(by='Amount',ascending=False)
2
3  sns.barplot(x='Gender',y='Amount',data=sales_gen,hue='Gender', palette='deep')
4  plt.title("Gender Wise Sales")
5  plt.show()
```
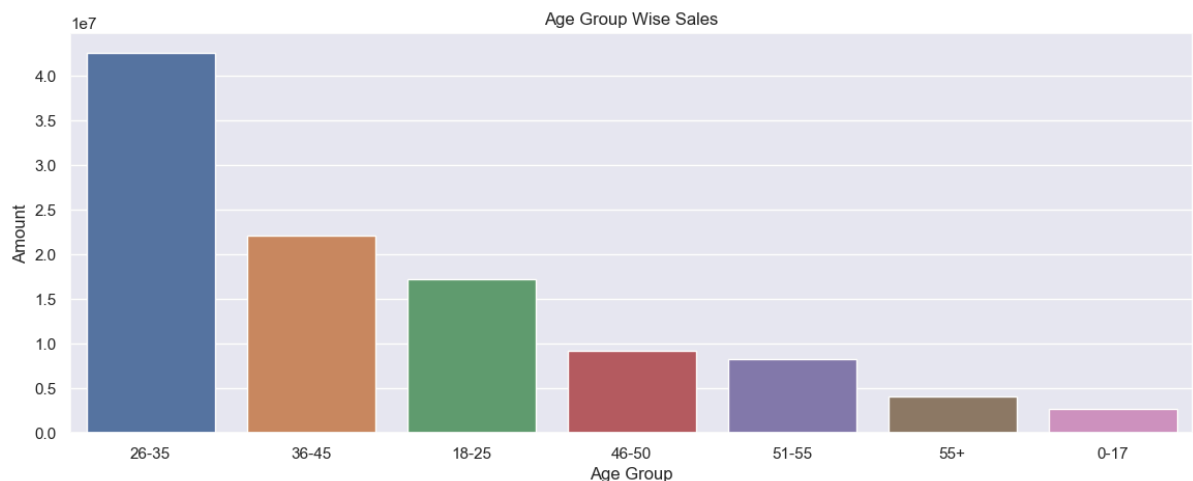

From above we can see that most of the buyers are Females and even the purchasing power of are greater than men

Age

In [42]:
```python
1  # Total Amount vs Age Group
2  sales_age = df.groupby(['Age Group'],as_index=False)
   ['Amount'].sum().sort_values(by='Amount',ascending=False)
3
4  sns.barplot(data=sales_age,x='Age Group',y='Amount', hue='Age Group',palette='deep')
5  plt.title("Age Group Wise Sales")
6  plt.show()
```


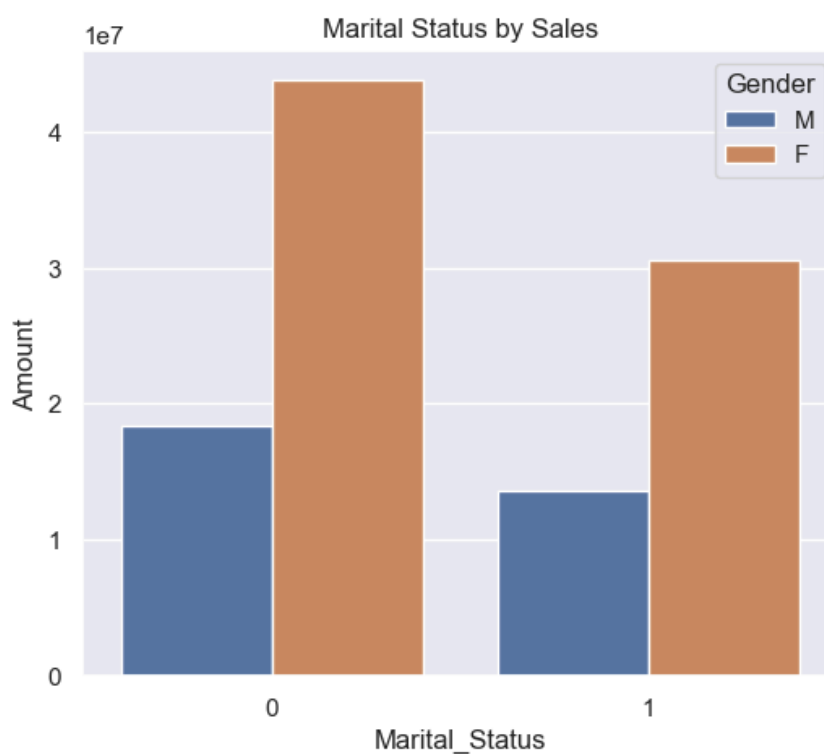**From above graph we can see that most of buyers are of age group between 26-35**

State

In [43]:
```python
# total number of orders from top 10 states

sales_state = df.groupby(['State'],as_index=False)
['Orders'].sum().sort_values(by='Orders',ascending=False).head(10)

sns.set(rc={'figure.figsize':(15,5)})
sns.barplot(data=sales_state, x="State", y ="Orders",hue='State',palette="deep")
plt.title('Top 10 State Selling Product')
plt.show()
```


From above graphs we can see that unexpectaly most of the orders are from Uter Pradesh, Maharashtra and Karnataka respectively but total sales/amount in from UP,Karnataka and then Maharashtra

Marital Status

In [44]:
```python
sales_state = df.groupby(['Marital_Status','Gender'],as_index=False)
['Amount'].sum().sort_values(by='Amount',ascending=True)

sns.set(rc={"figure.figsize":(6,5)})
sns.barplot(data=sales_state, x="Marital_Status",y="Amount", hue='Gender')
plt.title("Marital Status by Sales")
plt.show()
```


From above graphs we can see that buyers are married womans ther are high purchasing power

Occupation

In [45]:
```
1  occu_sales = df.groupby(['Occupation'],as_index=False)
   ['Amount'].sum().sort_values(by='Amount',ascending=False)
2  sns.set(rc={'figure.figsize':(20,5)})
3  sns.barplot(x="Occupation",y="Amount", data=occu_sales,hue='Occupation',palette='deep')
4  plt.title("Occupation By Sales")
5  plt.show()
```
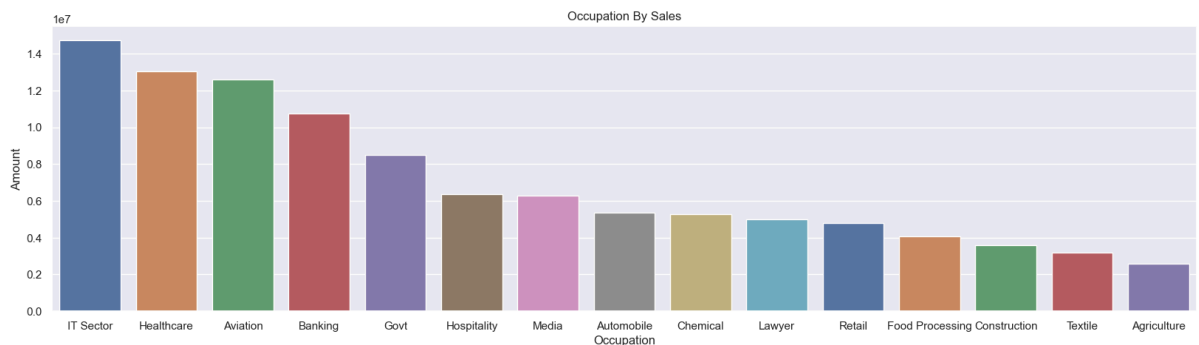

From above graphs we can see that most of the buyers are working IT Sector, Healthcare and Aviaton sector

Product Category

In [46]:
```
1  product_sales = df.groupby(['Product_Category'],as_index=False)
   ['Amount'].sum().sort_values(by='Amount',ascending=False).head(10)
2
3  sns.set(rc={'figure.figsize':(14,5)})
4  sns.barplot(data=product_sales, x='Product_Category',y='Amount',hue='Product_Category')
5  plt.title("Product Category By Sales")
6  plt.xticks(rotation=45)
7  plt.show()
```
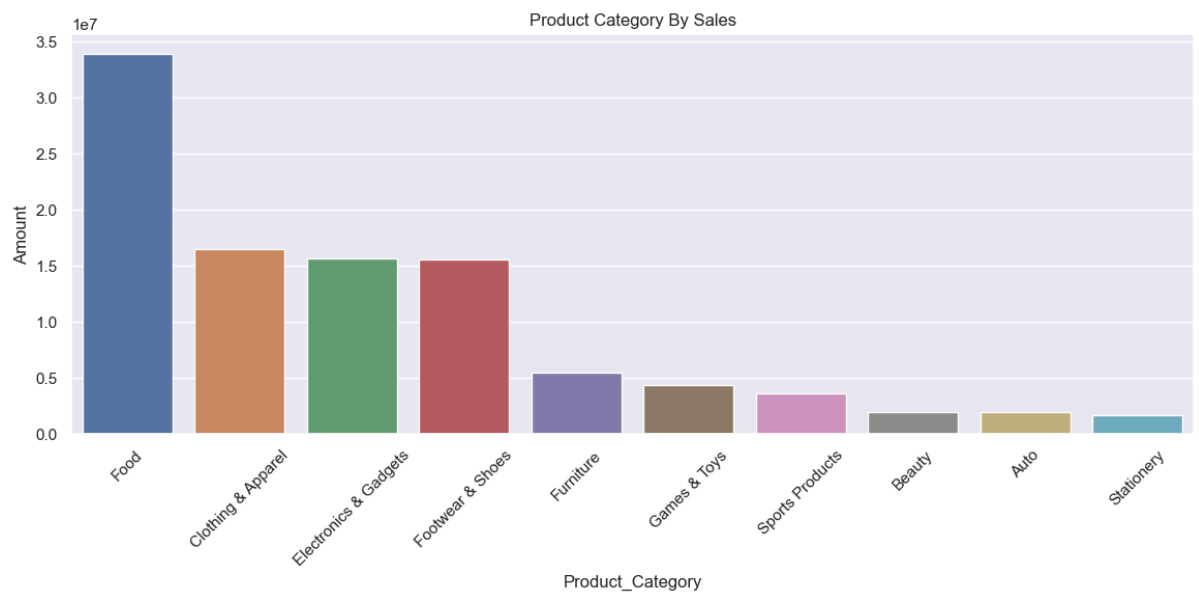

From above graphs we can see that most of the sold products are from Food, Footwear adn Electronics category