```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

data = pd.read_csv("sales.csv")

data.head()
```

```
   User_ID  Cust_name Product_ID Gender Age Group  Age  Marital_Status  \
0  1002903  Sanskriti  P00125942      F     26-35   28               0

1  1000732     Kartik  P00110942      F     26-35   35               1

2  1001990      Bindu  P00118542      F     26-35   35               1

3  1001425     Sudevi  P00237842      M      0-17   16               0

4  1000588       Joni  P00057942      M     26-35   28               1


            State      Zone       Occupation Product_Category  Orders  \
Amount
0      Maharashtra   Western       Healthcare             Auto       1
23952.0
1   Andhra?Pradesh  Southern             Govt             Auto       3
23934.0
2    Uttar Pradesh   Central       Automobile             Auto       3
23924.0
3        Karnataka  Southern     Construction             Auto       2
23912.0
4          Gujarat   Western  Food Processing             Auto       2
23877.0
```

```python
data.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
```

```
data.isnull().sum()
```

```
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
```

```
data.dropna(inplace=True)
data.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
```

```
data["Amount"].dtype
```

```
dtype('float64')
```

```python
data["Product_ID"].dtype
```

```
dtype('O')
```

```python
data["Orders"].dtype
```

```
dtype('int64')
```

```python
data["Order"] = data["Orders"].astype("int8")
```

```python
data["Amount"].describe()
```

```
count    11239.000000
mean      9453.610858
std       5222.355869
min        188.000000
25%       5443.000000
50%       8109.000000
75%      12675.000000
max      23952.000000
Name: Amount, dtype: float64
```

#EDA

```python
data["Gender"].value_counts()
```

```
F    7832
M    3407
Name: Gender, dtype: int64
```

```python
data["Gender"].value_counts().plot(kind="pie",autopct="%0.2f%%")
plt.show()
```
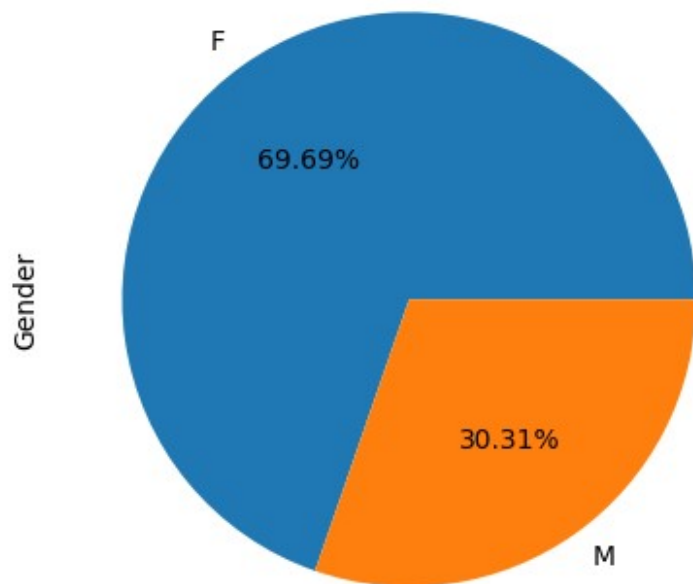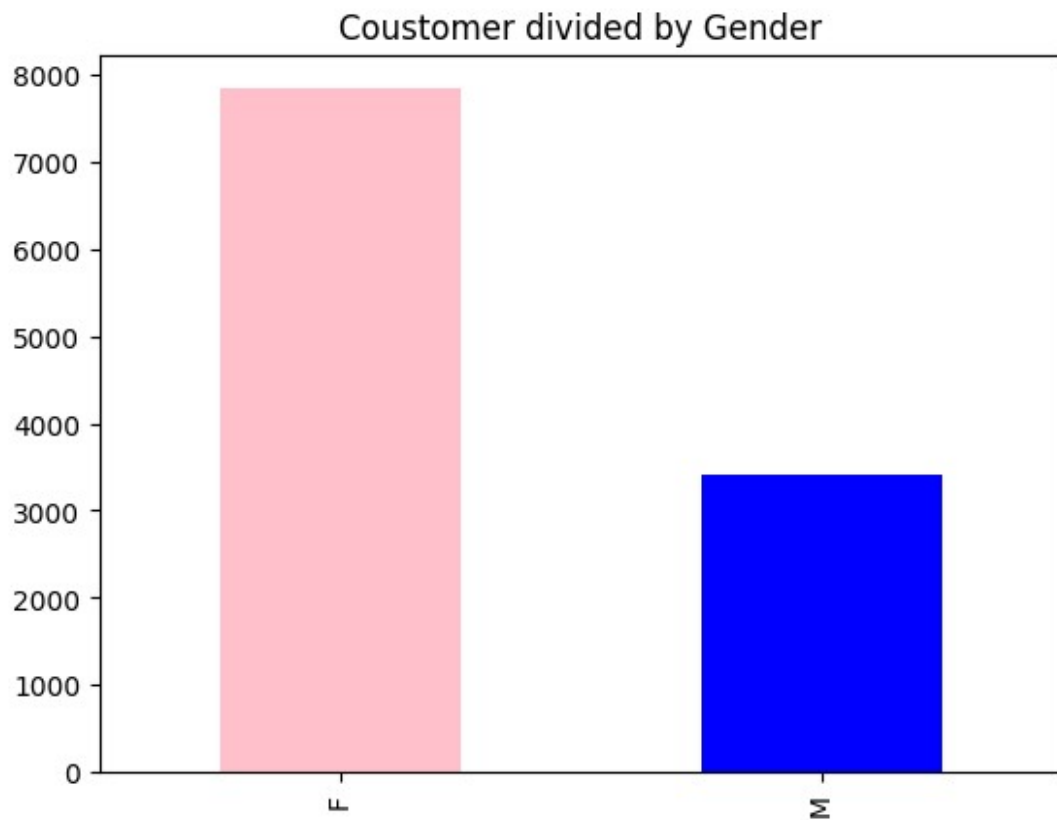
```
 data.groupby("Gender")["Amount"].sum()

Gender
F     74335856.43
M     31913276.00
Name: Amount, dtype: float64

data.groupby("Gender")["Amount"].sum().plot(kind="pie",autopct="%0.2f%
%")
plt.show()
```
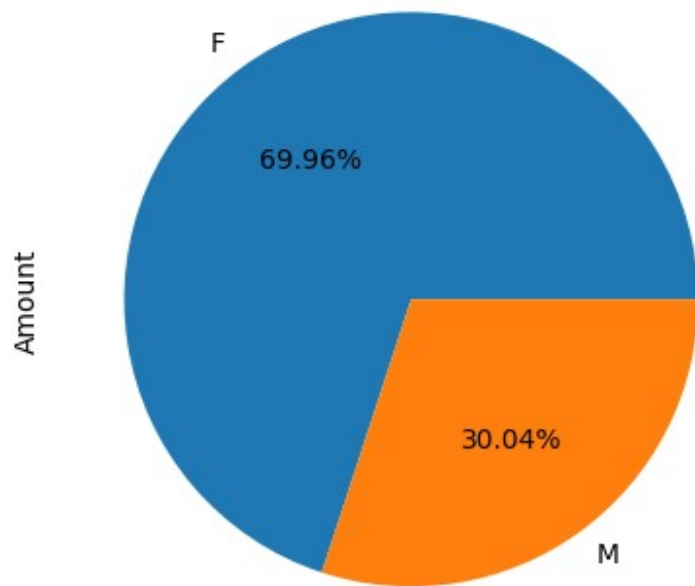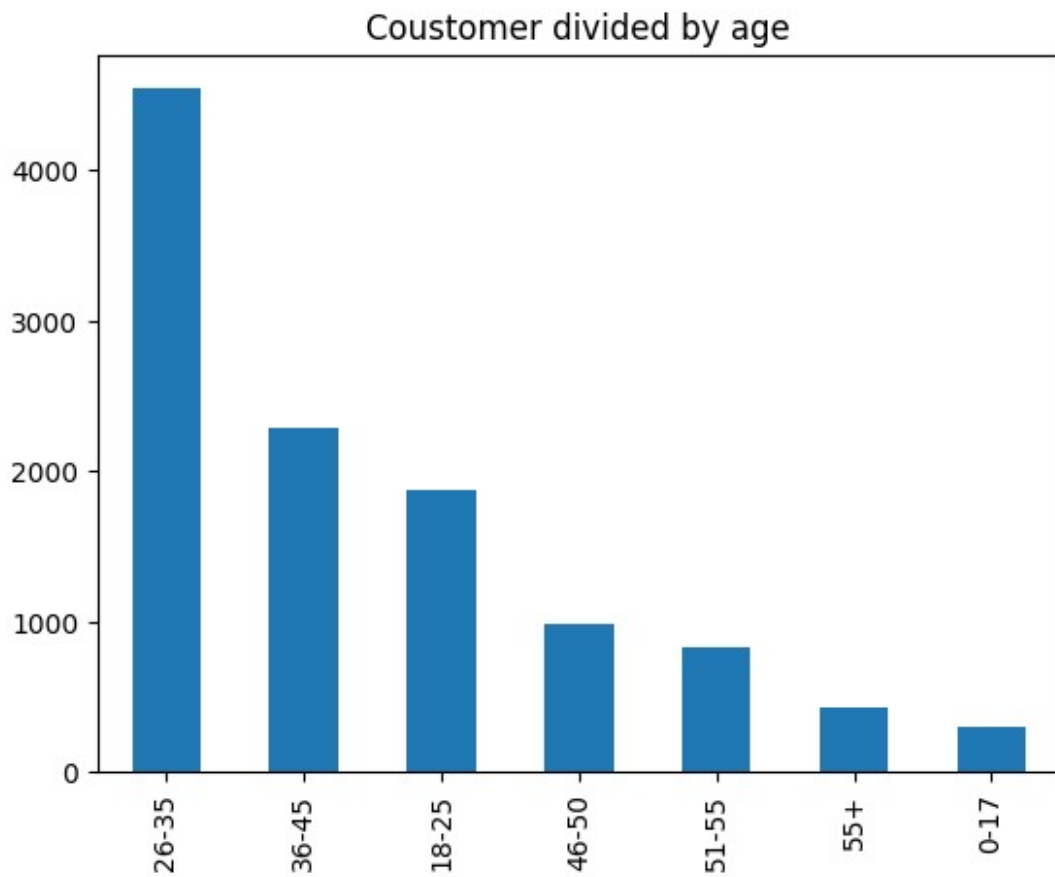
```
data["Gender"].value_counts().plot(kind="bar",color=("pink","blue"))
plt.title("Coustomer divided by Gender")
plt.show()
```

Coustomer divided by Gender

```
data["Age Group"].value_counts().plot(kind="bar")
plt.title("Coustomer divided by age")
plt.show()
```
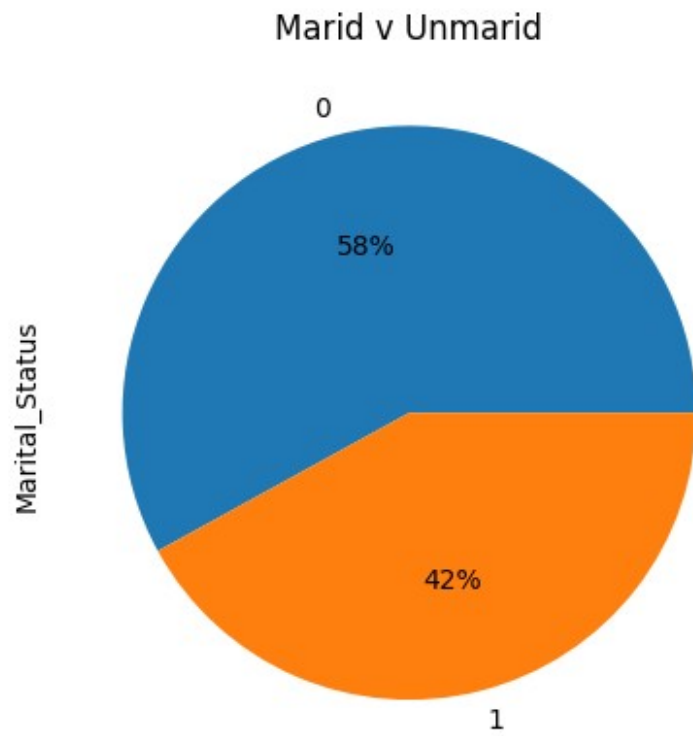
Coustomer divided by age


```
data.groupby("Age Group")["Amount"].sum().plot(kind="bar")
plt.title("Coustmer divided by spending & Age")
plt.show()
```

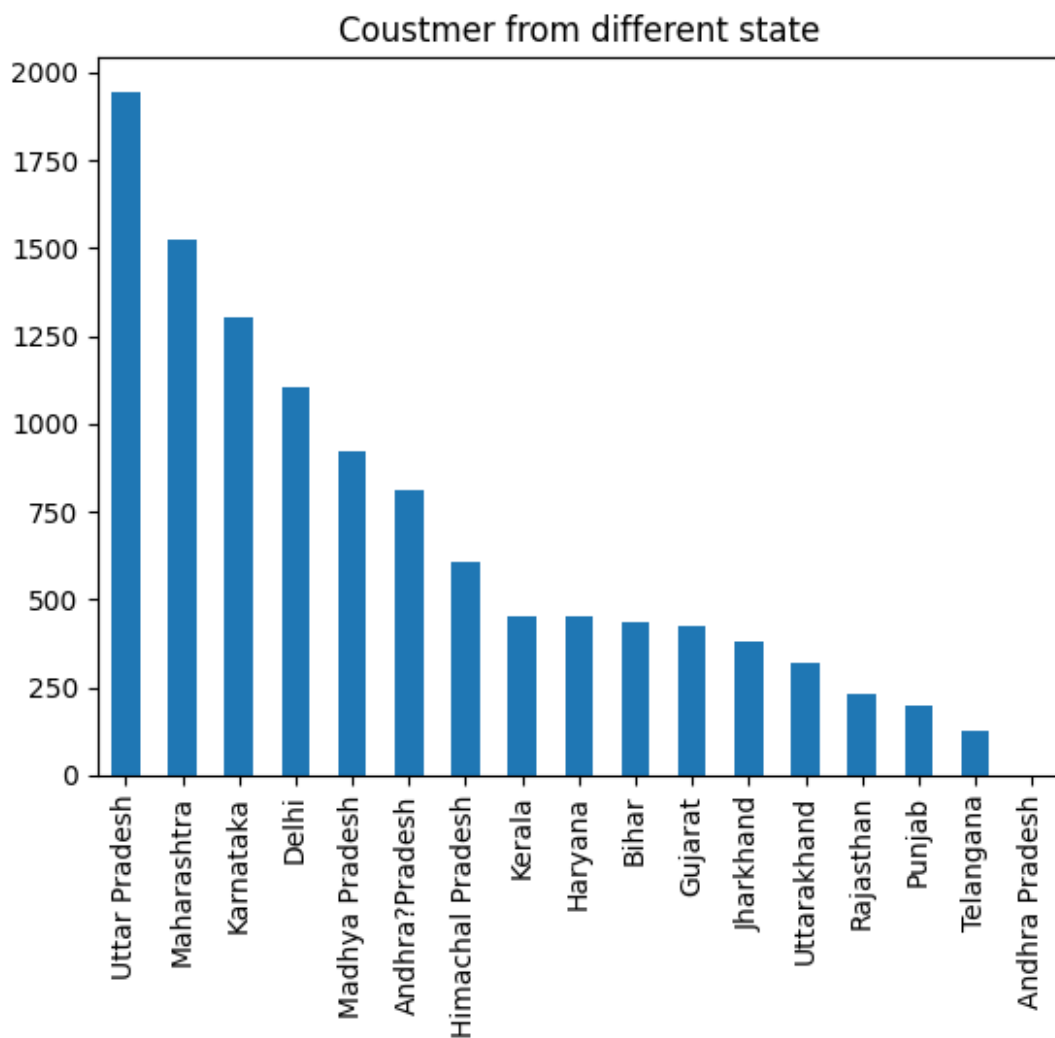Coustmer divided by spending & Age
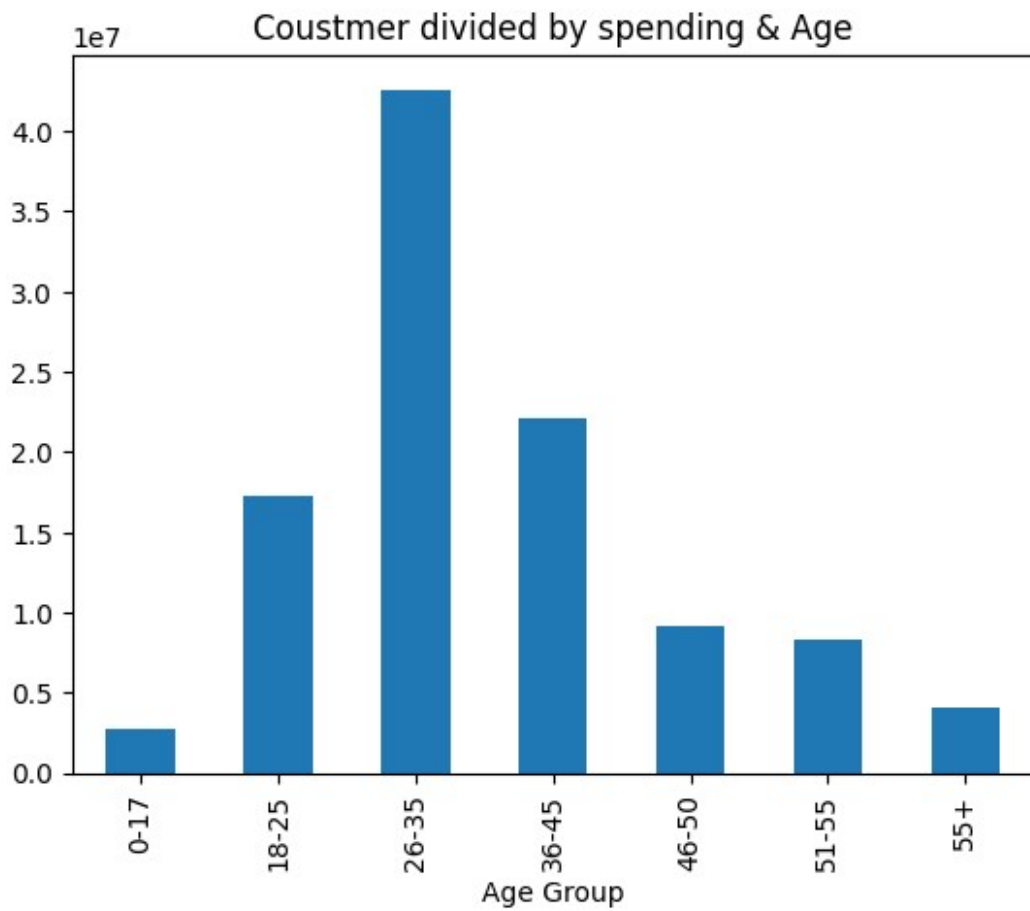
```
data["Marital_Status"].value_counts().plot(kind="pie",autopct="%0.0f%
%")
plt.title("Marid v Unmarid")
plt.show()
```

Marid v Unmarid
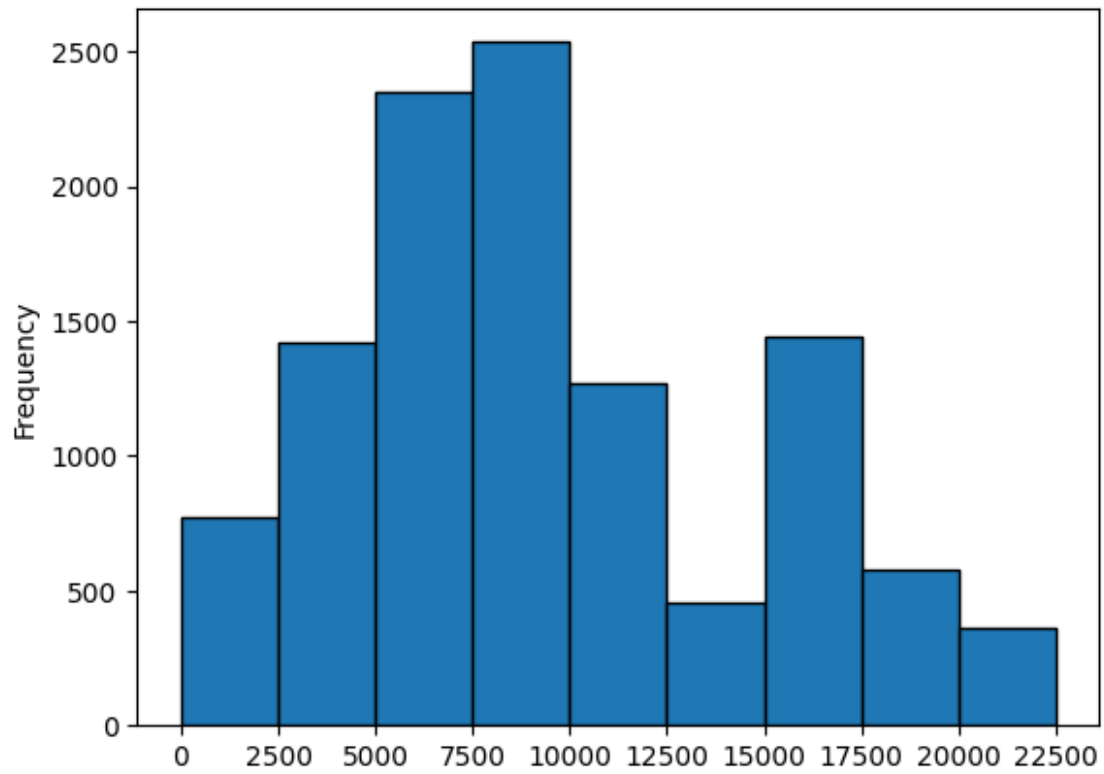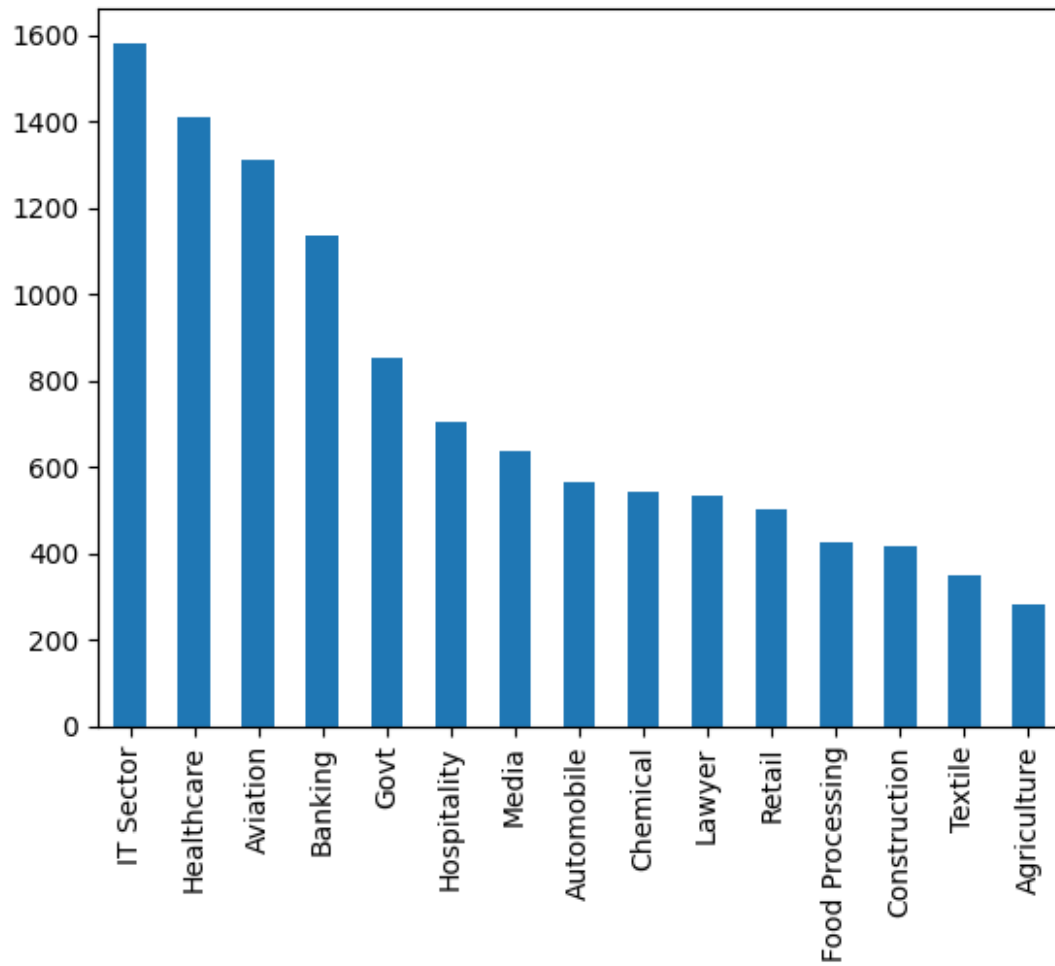


```
data["State"].value_counts().plot(kind="bar")
plt.title("Coustmer from different state")
plt.show()
```
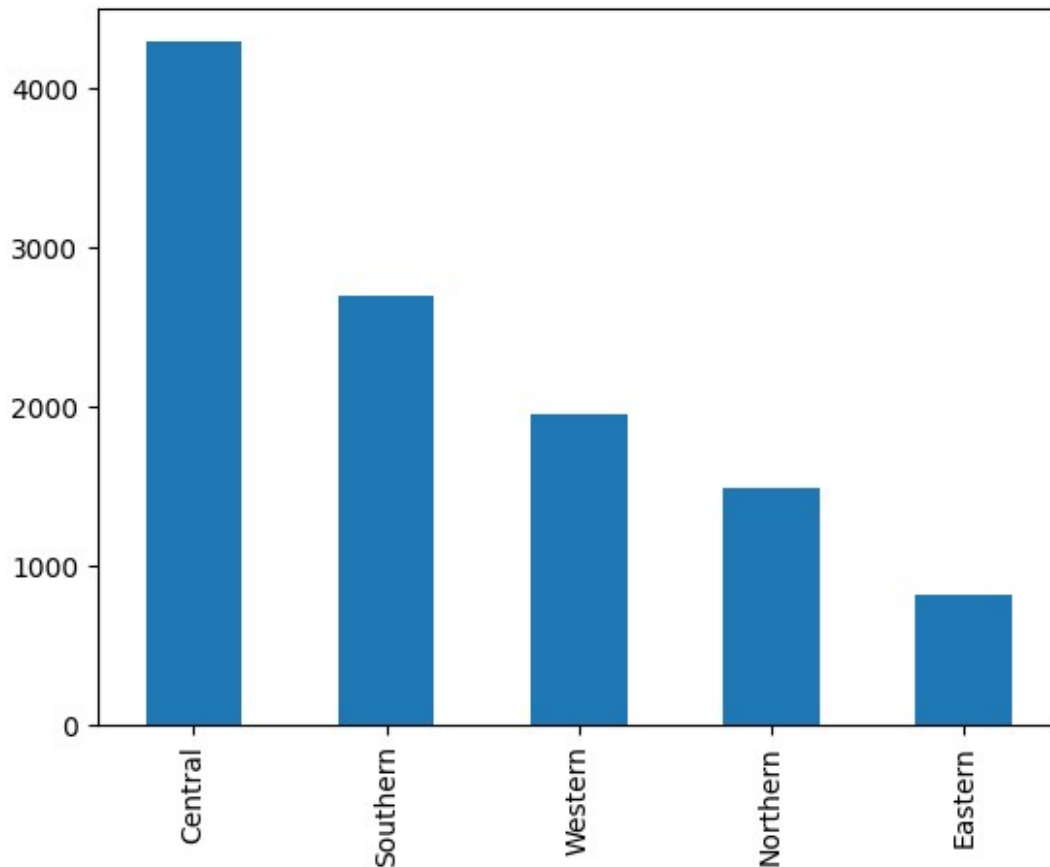
Coustmer from different state

```
data["Amount"].plot(kind="hist",bins=range(0,25000,2500),edgecolor="bl
ack")
plt.xticks(range(0,25000,2500))
plt.show()
```

```
data["Occupation"].value_counts().plot(kind="bar")
plt.show()
```

```
data["Zone"].value_counts().plot(kind="bar")
plt.show()
```

```
#Conclusion

# Married Woman b/w 26-35 Years From UP, Maharastra , karnataka
Working in IT
#Healthcare and more
#like to Buy Product Are Food , Clothing And Electronics Category .
```

```python
data.columns
```

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

```python
data["Product_Category"].value_counts()
```

```
Clothing & Apparel      2655
Food                    2490
Electronics & Gadgets   2087
Footwear & Shoes        1059
Household items          520
```

```
Beauty                        422
Games & Toys                  386
Sports Products               356
Furniture                     352
Pet Care                      212
Office                        113
Stationery                    112
Books                         103
Auto                           97
Decor                          96
Veterinary                     81
Tupperware                     72
Hand & Power Tools             26
Name: Product_Category, dtype: int64
```

data["Occupation"].value_counts()

```
IT Sector           1583
Healthcare          1408
Aviation            1310
Banking             1137
Govt                 854
Hospitality          703
Media                637
Automobile           565
Chemical             541
Lawyer               531
Retail               501
Food Processing      423
Construction         414
Textile              349
Agriculture          283
Name: Occupation, dtype: int64
```

data["Zone"].value_counts()

```
Central     4289
Southern    2693
Western     1952
Northern    1491
Eastern      814
Name: Zone, dtype: int64
```