python-project

September 15, 2023

[3]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

[4]:
```
# load the dataset
df= pd.read_csv("SampleSuperstore.csv")
df.head()
```

[4]:
```
        Ship Mode    Segment        Country             City       State  \
0    Second Class   Consumer  United States        Henderson    Kentucky
1    Second Class   Consumer  United States        Henderson    Kentucky
2    Second Class  Corporate  United States      Los Angeles  California
3  Standard Class   Consumer  United States  Fort Lauderdale     Florida
4  Standard Class   Consumer  United States  Fort Lauderdale     Florida

   Postal Code Region         Category Sub-Category     Sales  Quantity  \
0        42420  South        Furniture    Bookcases  261.9600         2
1        42420  South        Furniture       Chairs  731.9400         3
2        90036   West  Office Supplies       Labels   14.6200         2
3        33311  South        Furniture       Tables  957.5775         5
4        33311  South  Office Supplies      Storage   22.3680         2

   Discount    Profit
0      0.00   41.9136
1      0.00  219.5820
2      0.00    6.8714
3      0.45 -383.0310
4      0.20    2.5164
```

[8]:
```
df.head()
```

[8]:
```
        Ship Mode    Segment        Country             City       State  \
0    Second Class   Consumer  United States        Henderson    Kentucky
1    Second Class   Consumer  United States        Henderson    Kentucky
2    Second Class  Corporate  United States      Los Angeles  California
3  Standard Class   Consumer  United States  Fort Lauderdale     Florida
4  Standard Class   Consumer  United States  Fort Lauderdale     Florida
```

```
  Region         Category Sub-Category     Sales  Quantity  Discount    Profit
0  South        Furniture    Bookcases  261.9600         2      0.00   41.9136
1  South        Furniture       Chairs  731.9400         3      0.00  219.5820
2   West  Office Supplies       Labels   14.6200         2      0.00    6.8714
3  South        Furniture       Tables  957.5775         5      0.45 -383.0310
4  South  Office Supplies      Storage   22.3680         2      0.20    2.5164
```

[10]:

```python
print(df['Ship Mode'].unique())
print(df['Segment'].unique())
print(df['Country'].unique())
print(df['Category'].unique())
print(df['Sub-Category'].unique())
```

```
['Second Class' 'Standard Class' 'First Class' 'Same Day']
['Consumer' 'Corporate' 'Home Office']
['United States']
['Furniture' 'Office Supplies' 'Technology']
['Bookcases' 'Chairs' 'Labels' 'Tables' 'Storage' 'Furnishings' 'Art'
 'Phones' 'Binders' 'Appliances' 'Paper' 'Accessories' 'Envelopes'
 'Fasteners' 'Supplies' 'Machines' 'Copiers']
```

[11]:

```python
#statistical data
df.describe()
```

[11]:

```
              Sales     Quantity     Discount       Profit
count   9994.000000  9994.000000  9994.000000  9994.000000
mean     229.858001     3.789574     0.156203    28.656896
std      623.245101     2.225110     0.206452   234.260108
min        0.444000     1.000000     0.000000 -6599.978000
25%       17.280000     2.000000     0.000000     1.728750
50%       54.490000     3.000000     0.200000     8.666500
75%      209.940000     5.000000     0.200000    29.364000
max    22638.480000    14.000000     0.800000  8399.976000
```

[12]:

```python
df.info()
```

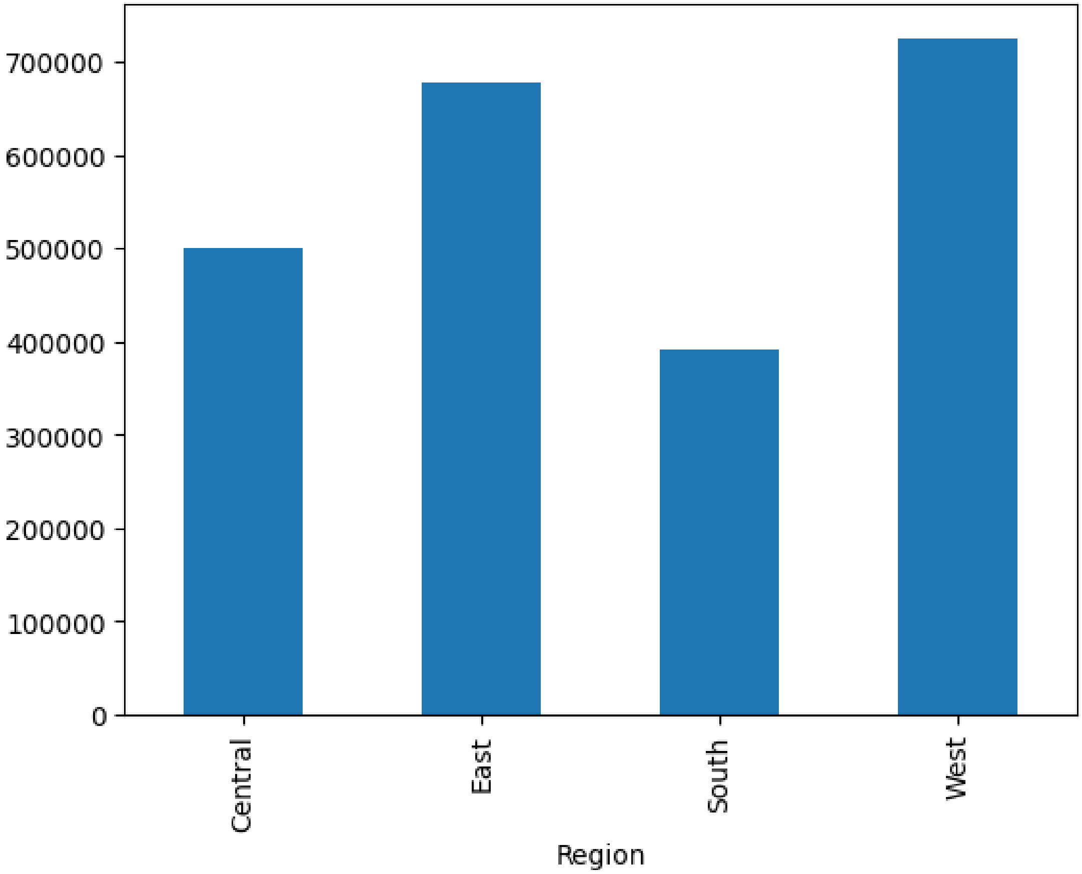
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9994 entries, 0 to 9993
Data columns (total 12 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   Ship Mode     9994 non-null   object
 1   Segment       9994 non-null   object
 2   Country       9994 non-null   object
 3   City          9994 non-null   object
 4   State         9994 non-null   object
 5   Region        9994 non-null   object
 6   Category      9994 non-null   object
 7   Sub-Category  9994 non-null   object
 8   Sales         9994 non-null   float64
 9   Quantity      9994 non-null   int64
 10  Discount      9994 non-null   float64
 11  Profit        9994 non-null   float64
dtypes: float64(3), int64(1), object(8)
memory usage: 937.1+ KB
```

[14]:
```
df.isnull().sum()
```

[14]:
```
Ship Mode       0
Segment         0
Country         0
City            0
State           0
Region          0
Category        0
Sub-Category    0
Sales           0
Quantity        0
Discount        0
Profit          0
dtype: int64
```

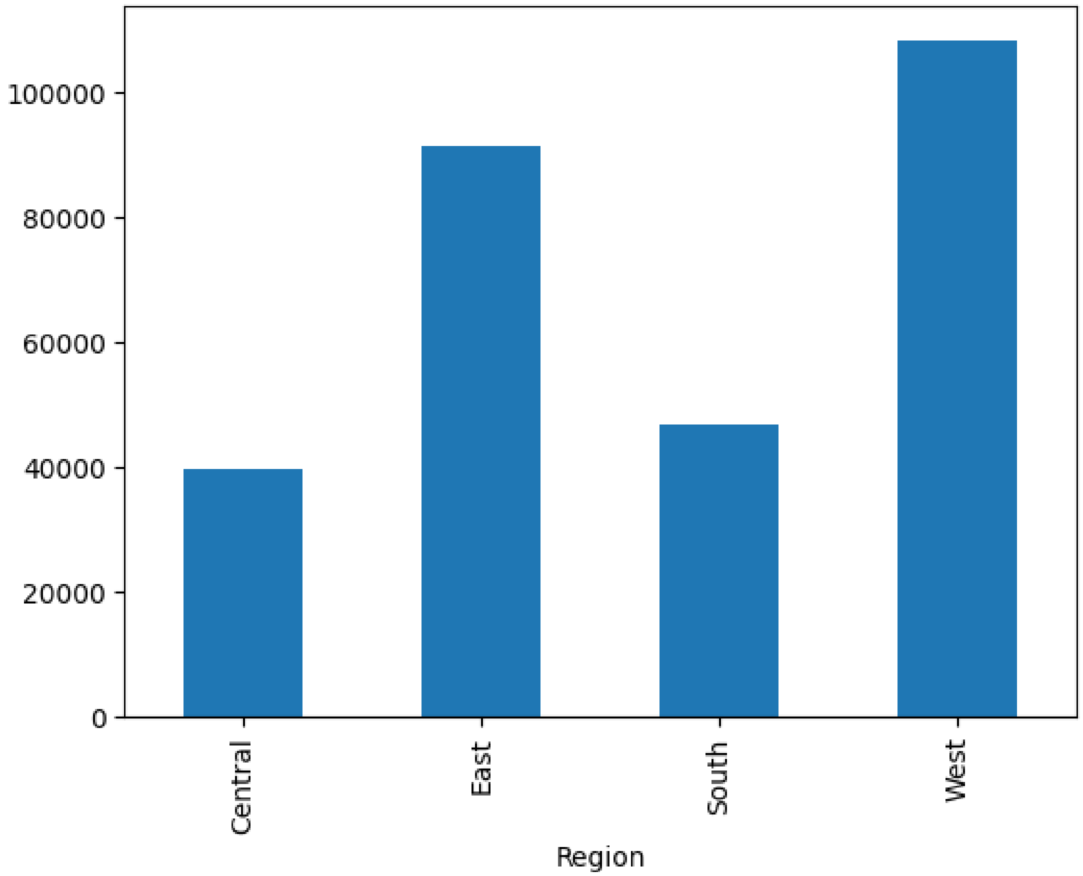
[15]:
```
#analyse the data
#profit in diff region  and sales analysis based on region
df.groupby("Region")["Sales"].sum().plot.bar()
```

[15]:
```
<Axes: xlabel='Region'>
```



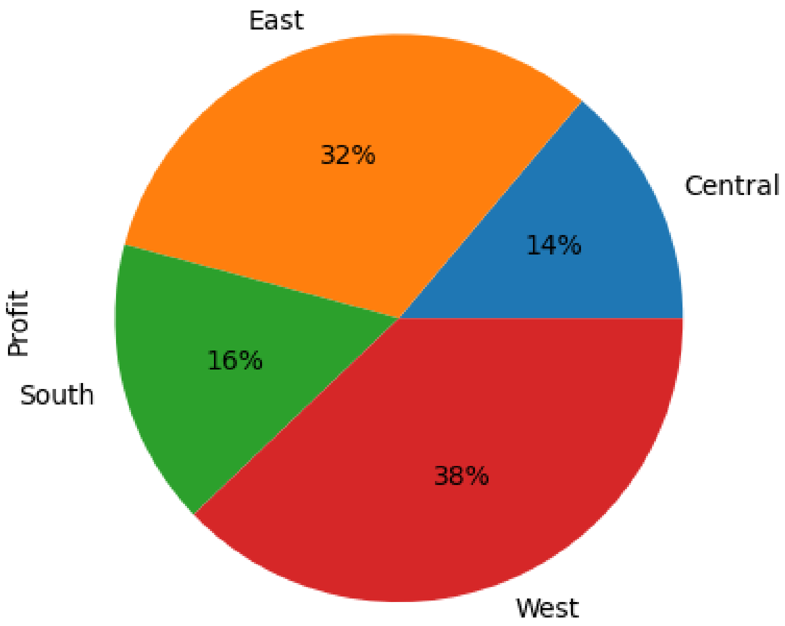
[16]:
```
df.groupby("Region")["Profit"].sum().plot.bar()
#profit by region
```

[16]:
```
<Axes: xlabel='Region'>
```



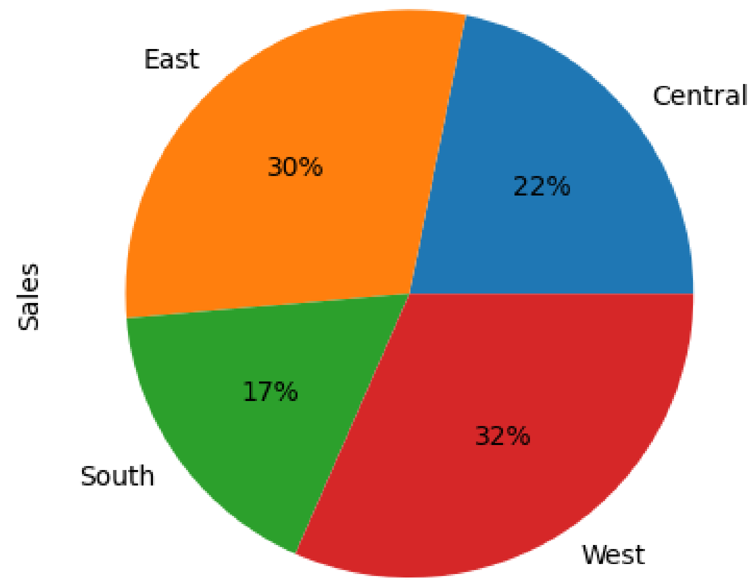
```
[17]: df.groupby("Region")["Profit"].sum().plot.pie(autopct="%1.0f%%")
```

```
[17]: <Axes: ylabel='Profit'>
```



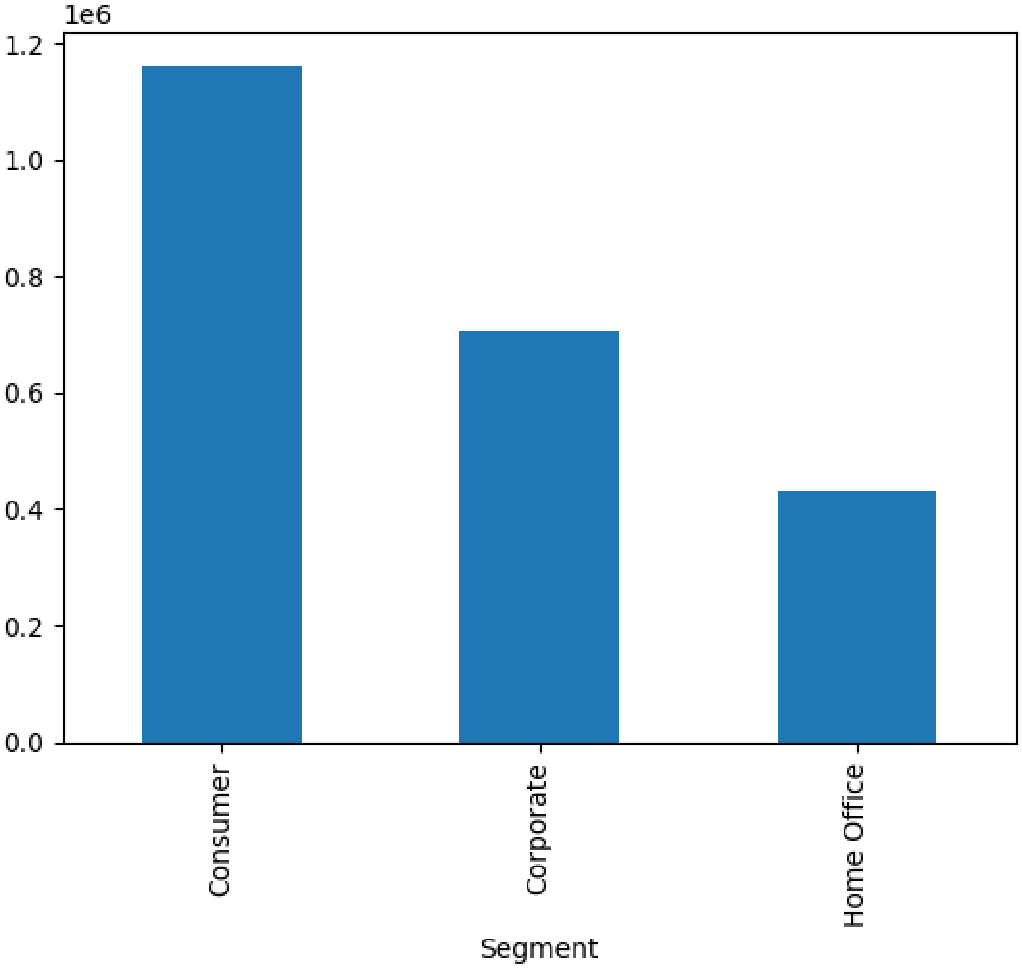
```
[18]: df.groupby("Region")["Sales"].sum().plot.pie(autopct="%1.0f%%")
```

```
[18]: <Axes: ylabel='Sales'>
```



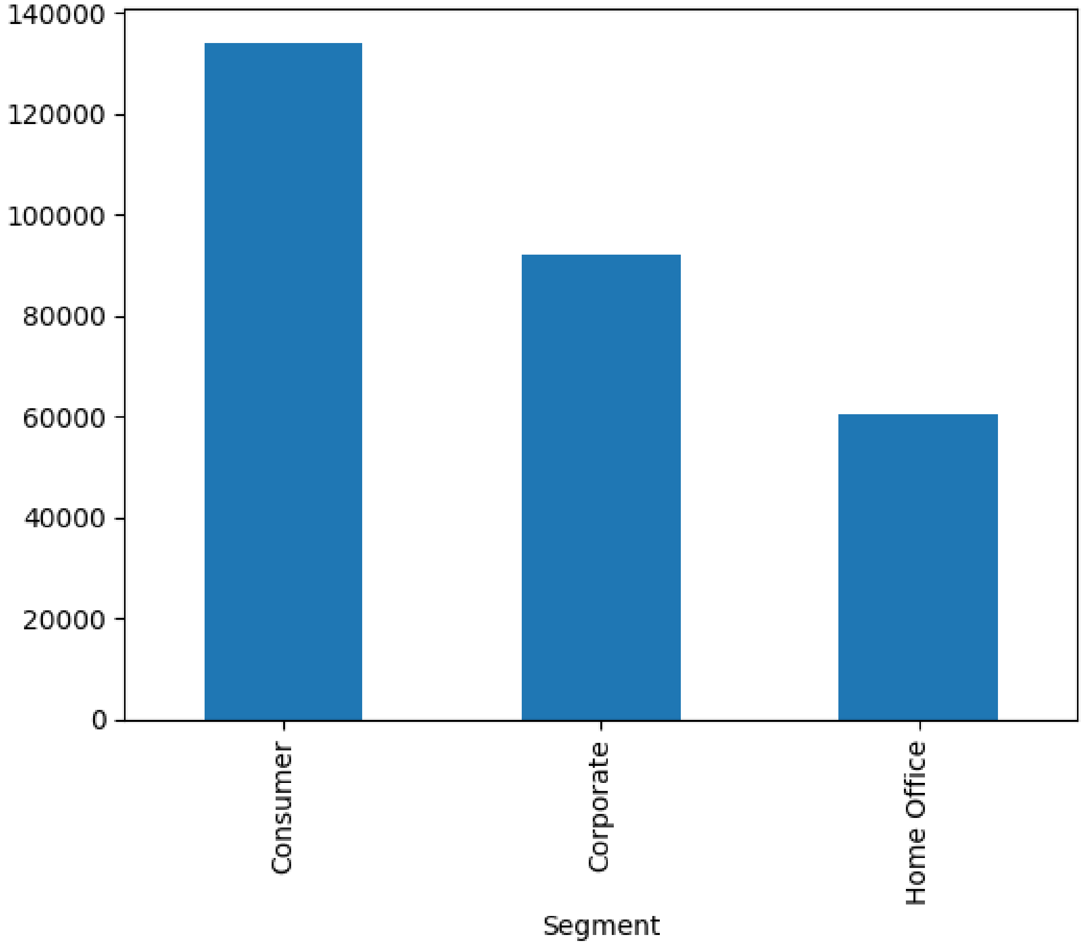
```
[19]: df.groupby("Segment")["Sales"].sum().plot.bar()
```

```
[19]: <Axes: xlabel='Segment'>
```



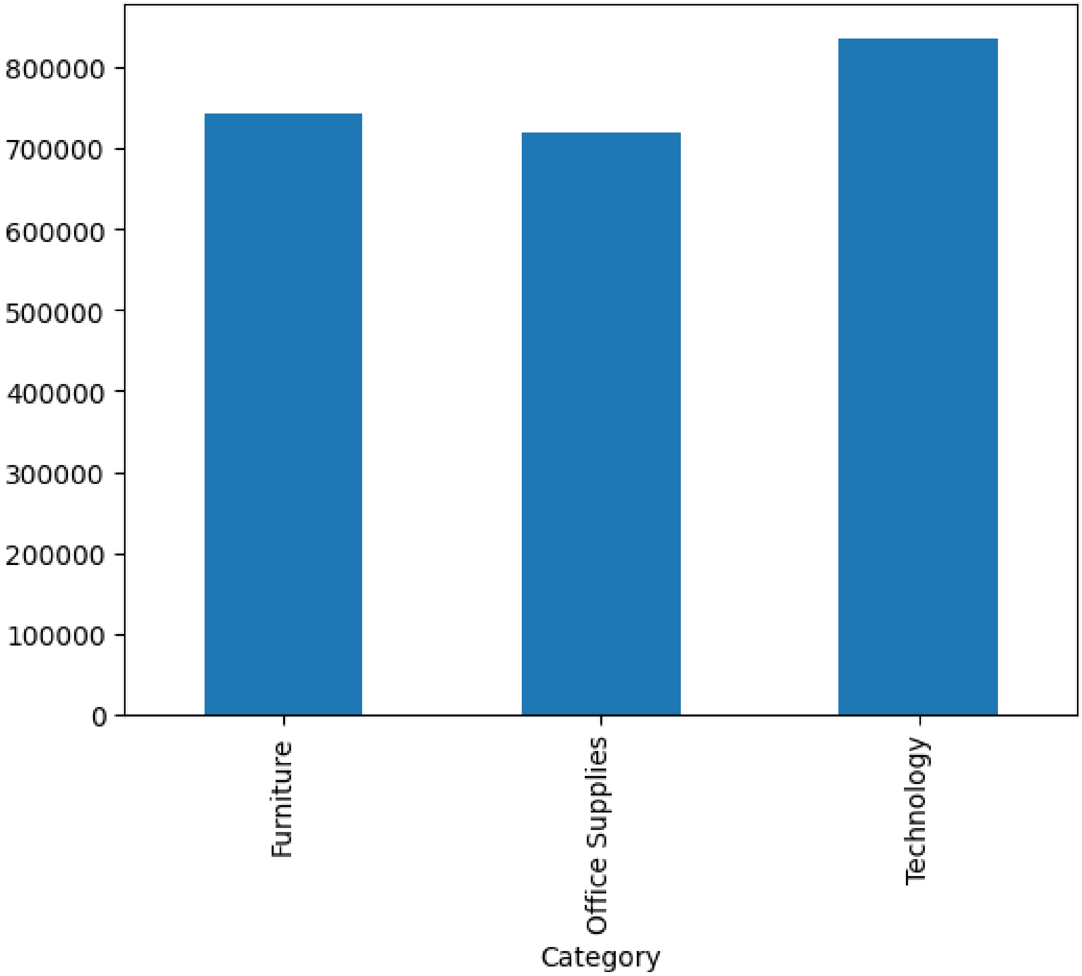
```
[20]: df.groupby("Segment")["Profit"].sum().plot.bar()
```

```
[20]: <Axes: xlabel='Segment'>
```



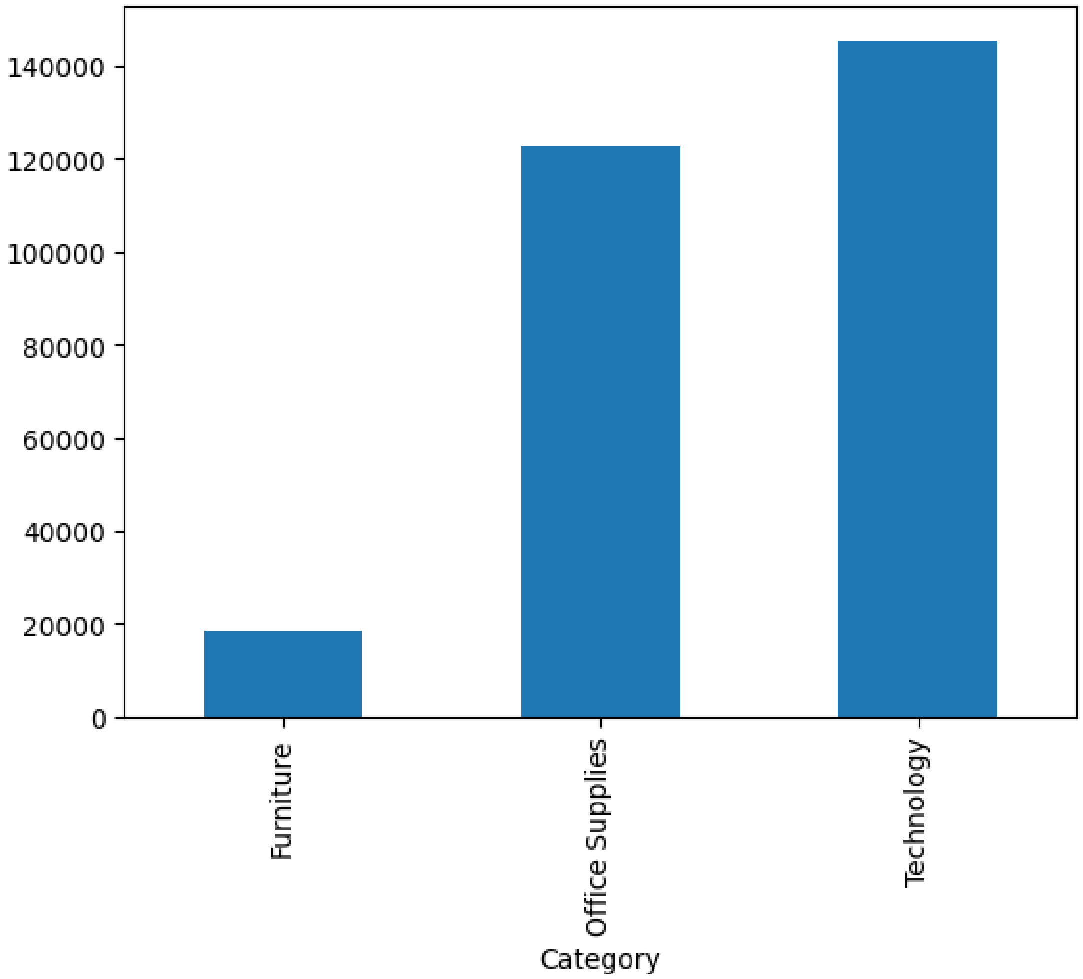
```
[21]: df.groupby("Category")["Sales"].sum().plot.bar()
```

```
[21]: <Axes: xlabel='Category'>
```



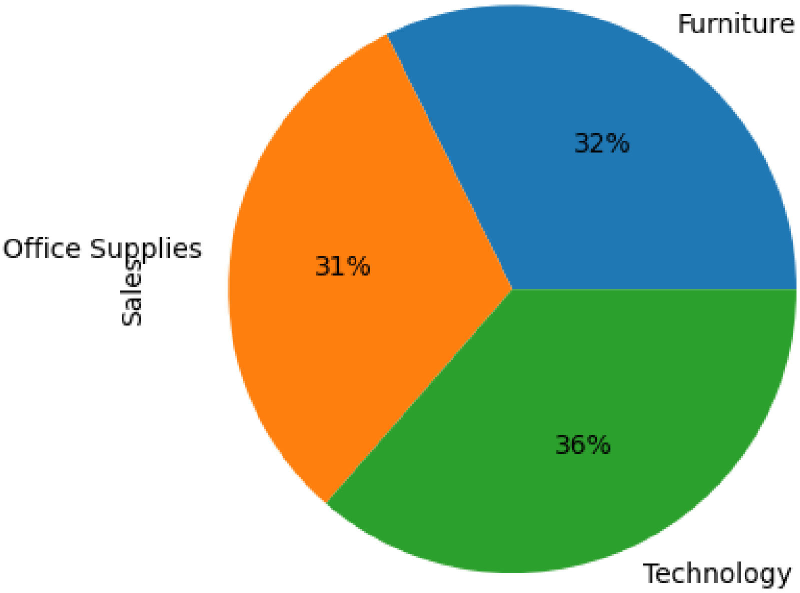
```
[22]: df.groupby("Category")["Profit"].sum().plot.bar()
```

```
[22]: <Axes: xlabel='Category'>
```



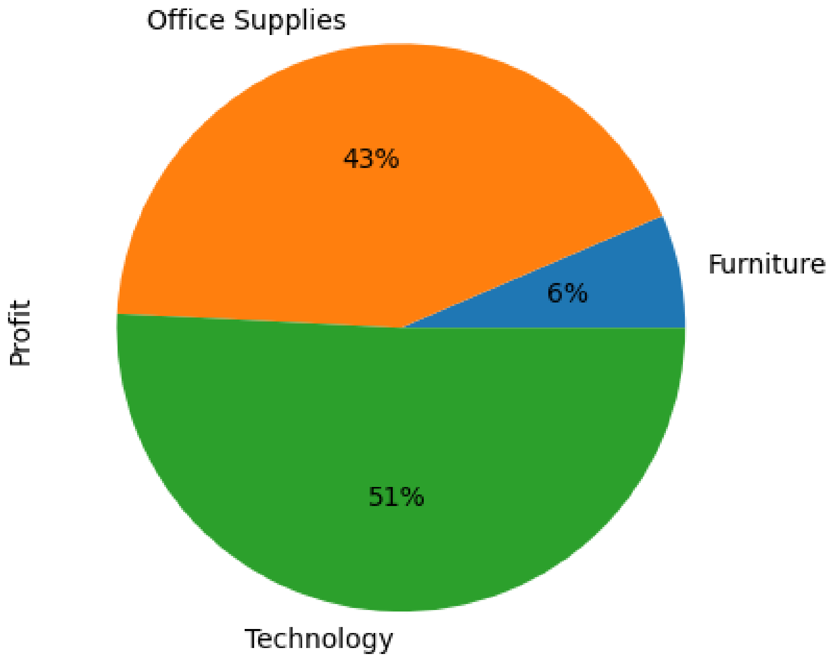
```
[23]: df.groupby("Category")["Sales"].sum().plot.pie(autopct="%1.0f%%")
```

```
[23]: <Axes: ylabel='Sales'>
```



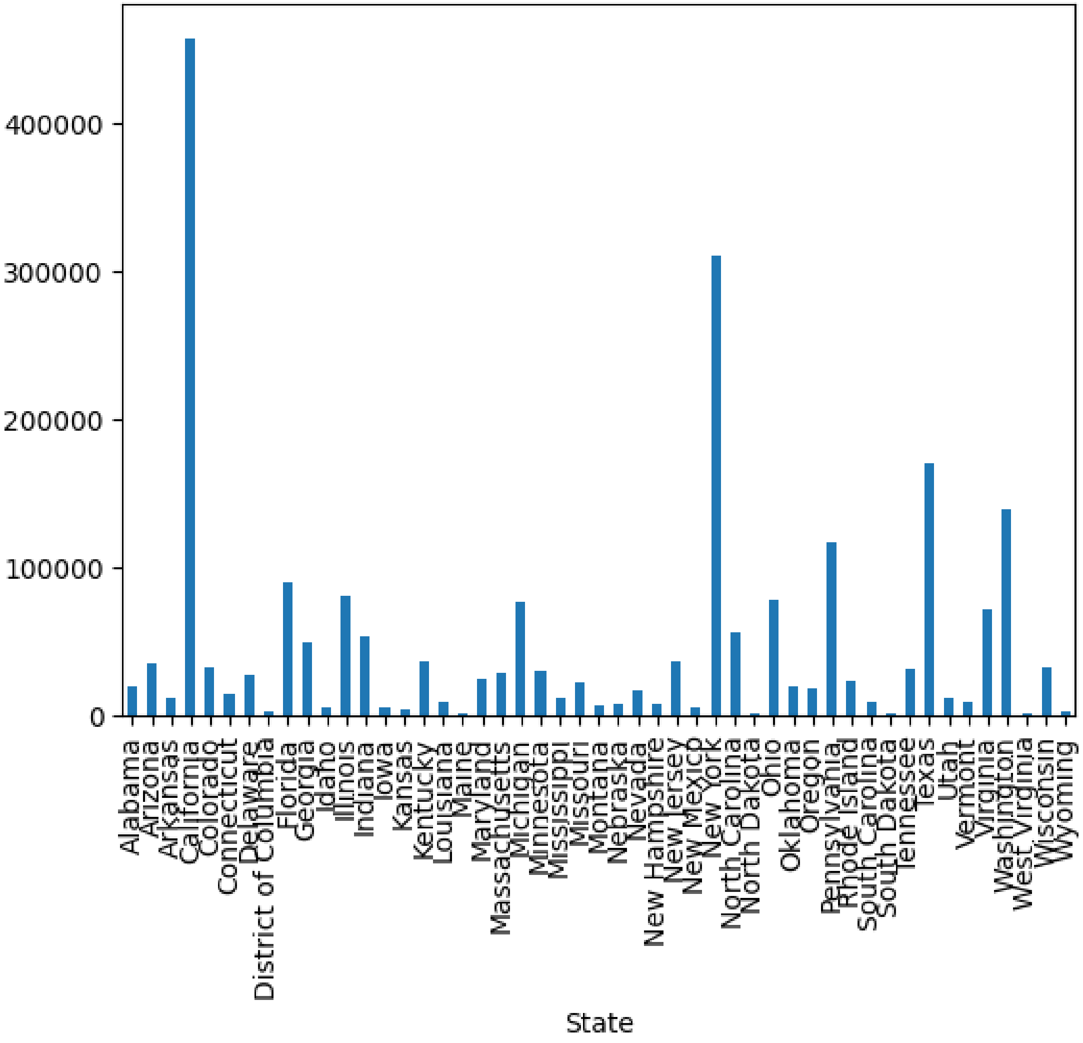
```
[24]: df.groupby("Category")["Profit"].sum().plot.pie(autopct="%1.0f%%")
```

```
[24]: <Axes: ylabel='Profit'>
```



```
[26]: df.groupby("State")["Sales"].sum().plot.bar()
```

```
[26]: <Axes: xlabel='State'>
```

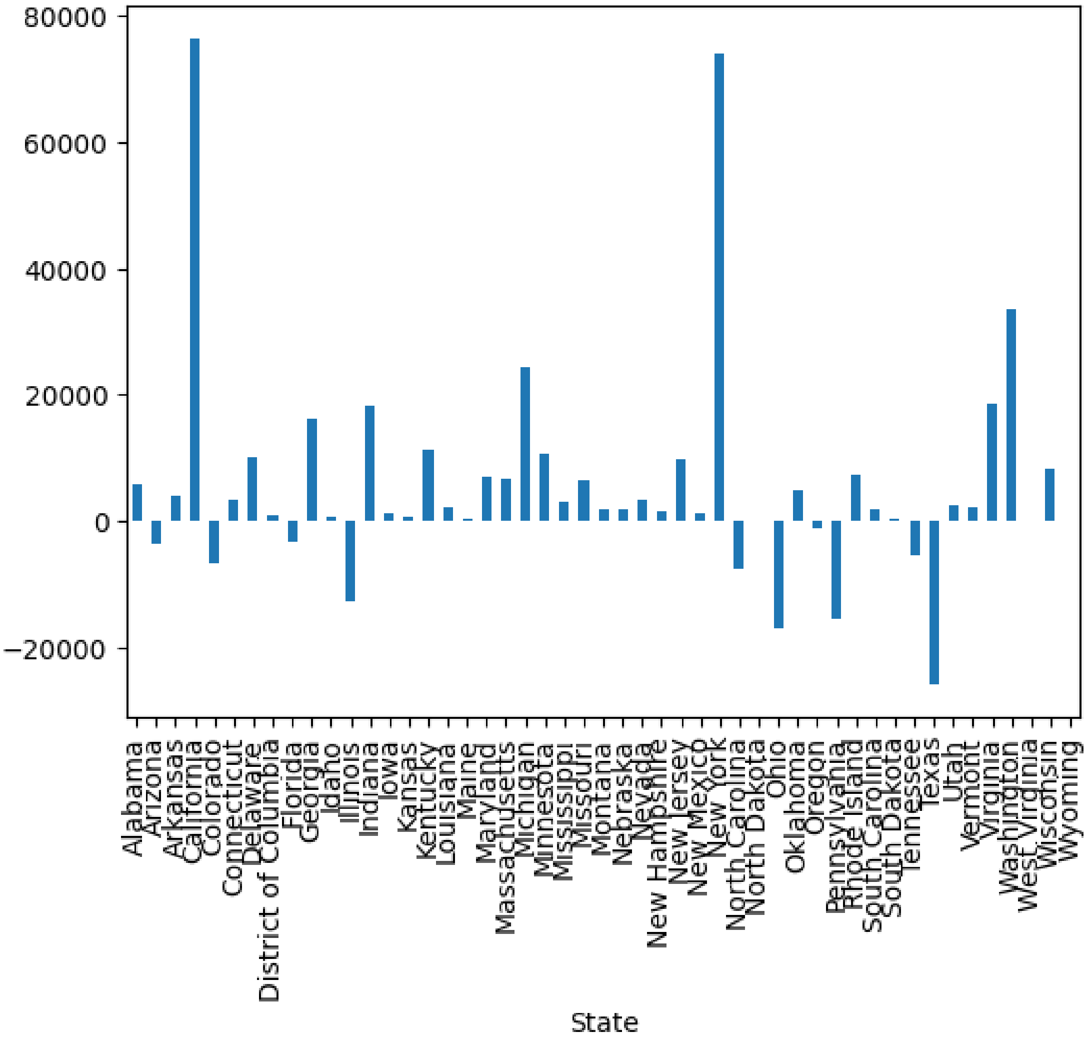


[27]:

```
df.groupby("State")["Profit"].sum().plot.bar()
```

[27]:

```
<Axes: xlabel='State'>
```


80000
60000
40000
20000
0
−20000
Alabama
Arizona
Arkansas
California
Colorado
Connecticut
Delaware
District of Columbia
Florida
Georgia
Idaho
Illinois
Indiana
Iowa
Kansas
Kentucky
Louisiana
Maine
Maryland
Massachusetts
Michigan
Minnesota
Mississippi
Missouri
Montana
Nebraska
Nevada
New Hampshire
New Jersey
New Mexico
New York
North Carolina
North Dakota
Ohio
Oklahoma
Oregon
Pennsylvania
Rhode Island
South Carolina
South Dakota
Tennessee
Texas
Utah
Vermont
Virginia
Washington
West Virginia
Wisconsin
Wyoming
State