```
import pandas as pd
import numpy as np
import matplotlib as mpl
import matplotlib.pyplot as plt
import seaborn as sns
```

```
d=pd.read_csv("SampleSuperstore .csv")
df=pd.DataFrame(d)
print('Data has been succesfully imported')
df.head
```

```
Data has been succesfully imported
<bound method NDFrame.head of            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
...              ...        ...            ...              ...         ...
9989    Second Class   Consumer  United States            Miami     Florida
9990  Standard Class   Consumer  United States       Costa Mesa  California
9991  Standard Class   Consumer  United States       Costa Mesa  California
9992  Standard Class   Consumer  United States       Costa Mesa  California
9993    Second Class   Consumer  United States      Westminster  California

      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
...           ...    ...              ...          ...       ...       ...
9989        33180  South        Furniture  Furnishings   25.2480         3
9990        92627   West        Furniture  Furnishings   91.9600         2
9991        92627   West       Technology       Phones  258.5760         2
9992        92627   West  Office Supplies        Paper   29.6000         4
9993        92683   West  Office Supplies   Appliances  243.1600         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
...        ...       ...
9989      0.20    4.1028
9990      0.00   15.6332
9991      0.20   19.3932
9992      0.00   13.3200
9993      0.00   72.9480

[9994 rows x 13 columns]>
```

```
df.info
```

```
<bound method DataFrame.info of            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
...              ...        ...            ...              ...         ...
9989    Second Class   Consumer  United States            Miami     Florida
9990  Standard Class   Consumer  United States       Costa Mesa  California
9991  Standard Class   Consumer  United States       Costa Mesa  California
9992  Standard Class   Consumer  United States       Costa Mesa  California
9993    Second Class   Consumer  United States      Westminster  California

      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
...           ...    ...              ...          ...       ...       ...
9989        33180  South        Furniture  Furnishings   25.2480         3
9990        92627   West        Furniture  Furnishings   91.9600         2
9991        92627   West       Technology       Phones  258.5760         2
9992        92627   West  Office Supplies        Paper   29.6000         4
9993        92683   West  Office Supplies   Appliances  243.1600         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
...        ...       ...
9989      0.20    4.1028
9990      0.00   15.6332
9991      0.20   19.3932
9992      0.00   13.3200
9993      0.00   72.9480

[9994 rows x 13 columns]>
```

```
df.isnull().sum()
```

```
Ship Mode       0
Segment         0
Country         0
City            0
State           0
Postal Code     0
Region          0
Category        0
Sub-Category    0
Sales           0
Quantity        0
Discount        0
Profit          0
dtype: int64
```

```
df.describe
```

```
<bound method NDFrame.describe of            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
...              ...        ...            ...              ...         ...
9989    Second Class   Consumer  United States            Miami     Florida
9990  Standard Class   Consumer  United States       Costa Mesa  California
9991  Standard Class   Consumer  United States       Costa Mesa  California
9992  Standard Class   Consumer  United States       Costa Mesa  California
9993    Second Class   Consumer  United States      Westminster  California

      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
...           ...    ...              ...          ...       ...       ...
9989        33180  South        Furniture  Furnishings   25.2480         3
9990        92627   West        Furniture  Furnishings   91.9600         2
9991        92627   West       Technology       Phones  258.5760         2
9992        92627   West  Office Supplies        Paper   29.6000         4
9993        92683   West  Office Supplies   Appliances  243.1600         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
...        ...       ...
9989      0.20    4.1028
9990      0.00   15.6332
9991      0.20   19.3932
9992      0.00   13.3200
9993      0.00   72.9480

[9994 rows x 13 columns]>
```

```
df.drop(["Postal Code"],axis=1,inplace =True)
```

```
df
```

	Ship Mode	Segment	City	State	Category	Sub-Category	Sales	Qua
0	Second Class	Consumer	Henderson	Kentucky	Furniture	Bookcases	261.9600	
1	Second Class	Consumer	Henderson	Kentucky	Furniture	Chairs	731.9400	
2	Second Class	Corporate	Los Angeles	California	Office Supplies	Labels	14.6200	
3	Standard Class	Consumer	Fort Lauderdale	Florida	Furniture	Tables	957.5775	
4	Standard	Consumer	Fort	Florida	Office	Storage	22.3680	

```
#total sales
print('Total sales are {}'.format(df["Sales"].sum()))

#total profit
print("Total profit is {}".format(df["Profit"].sum()))
```

```
Total sales are 2297200.8603000003
Total profit is 286397.0217
```

```
df.columns
```

```
Index(['Ship Mode', 'Segment', 'City', 'State', 'Postal Code', 'Category',
       'Sub-Category', 'Sales', 'Quantity', 'Discount', 'Profit'],
      dtype='object')
```

```
df.shape
```

```
(9994, 10)
```

```
df.duplicated().sum()
```

```
50
```

```
df.drop_duplicates(inplace=True)
```

```
df
```

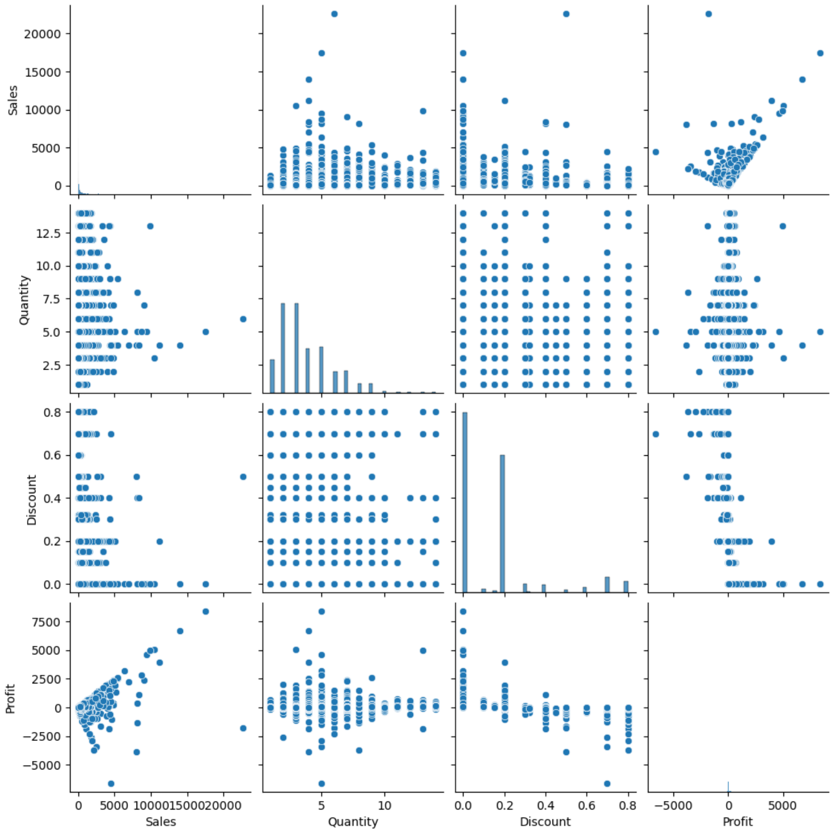
	Ship Mode	Segment	City	State	Category	Sub-Category	Sales	Qua
0	Second Class	Consumer	Henderson	Kentucky	Furniture	Bookcases	261.9600	
1	Second Class	Consumer	Henderson	Kentucky	Furniture	Chairs	731.9400	
2	Second Class	Corporate	Los Angeles	California	Office Supplies	Labels	14.6200	
3	Standard Class	Consumer	Fort Lauderdale	Florida	Furniture	Tables	957.5775	
4	Standard Class	Consumer	Fort Lauderdale	Florida	Office Supplies	Storage	22.3680	
...	...	...	...	...	...	...	...	
9989	Second Class	Consumer	Miami	Florida	Furniture	Furnishings	25.2480	
9990	Standard	Consumer	Costa Mesa	California	Furniture	Furnishings	91.9600	

```
df.duplicated().sum()
```

```
0
```

```
sns.pairplot(df)
```

```
<seaborn.axisgrid.PairGrid at 0x7f6b44242c50>
```



1..As discount increases profit decreases 2.As discount increases sales decreases 3.As sales increases profit increases

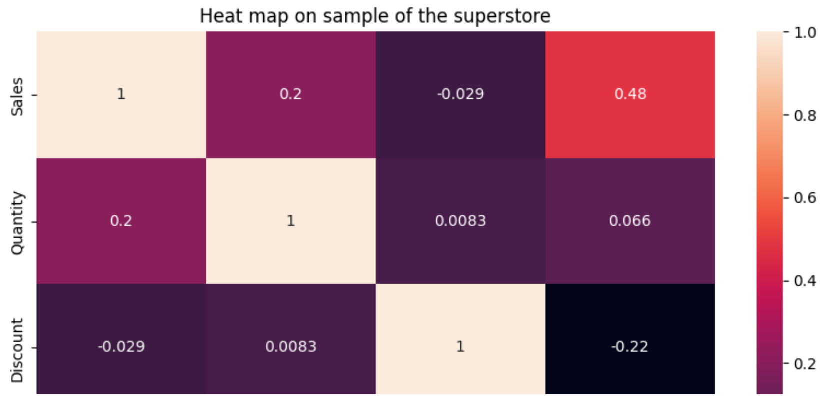
```
df.corr()
```

```
<ipython-input-31-2f6f6606aa2c>:1: FutureWarning: The default value of numeric_onl
  df.corr()
```

	Sales	Quantity	Discount	Profit
Sales	1.000000	0.200469	-0.028625	0.479078
Quantity	0.200469	1.000000	0.008307	0.066089
Discount	-0.028625	0.008307	1.000000	-0.219939
Profit	0.479078	0.066089	-0.219939	1.000000

```
plt.figure(figsize=(10,6))
sns.heatmap(df.corr(),annot=True)
plt.title("Heat map on sample of the superstore")
plt.show()
```

```
<ipython-input-32-436654308bd9>:2: FutureWarning: The default value of numeric_onl
  sns.heatmap(df.corr(),annot=True)
```

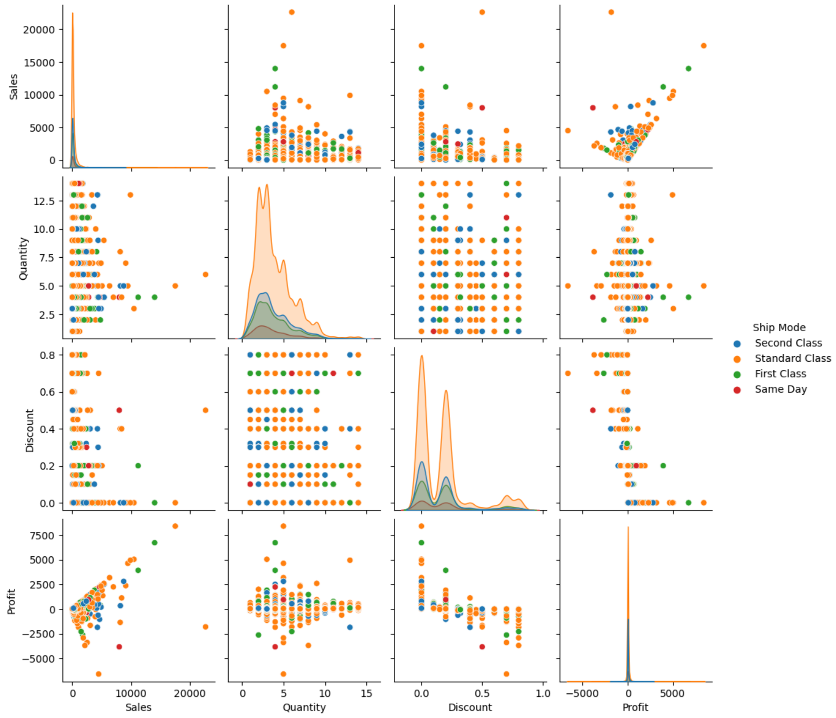


1.Most correlation:sales and profit

2.Least correlation:Discount and Quantity

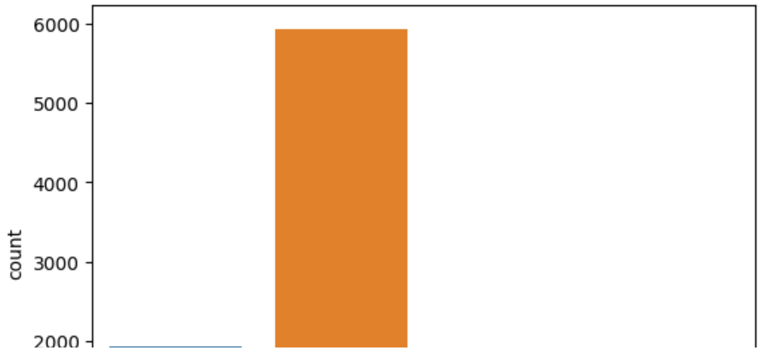
```
sns.pairplot(df,hue="Ship Mode")
```

```
<seaborn.axisgrid.PairGrid at 0x7f6b40b17940>
```



```
sns.countplot(x=df["Ship Mode"])
```

```
<Axes: xlabel='Ship Mode', ylabel='count'>
```



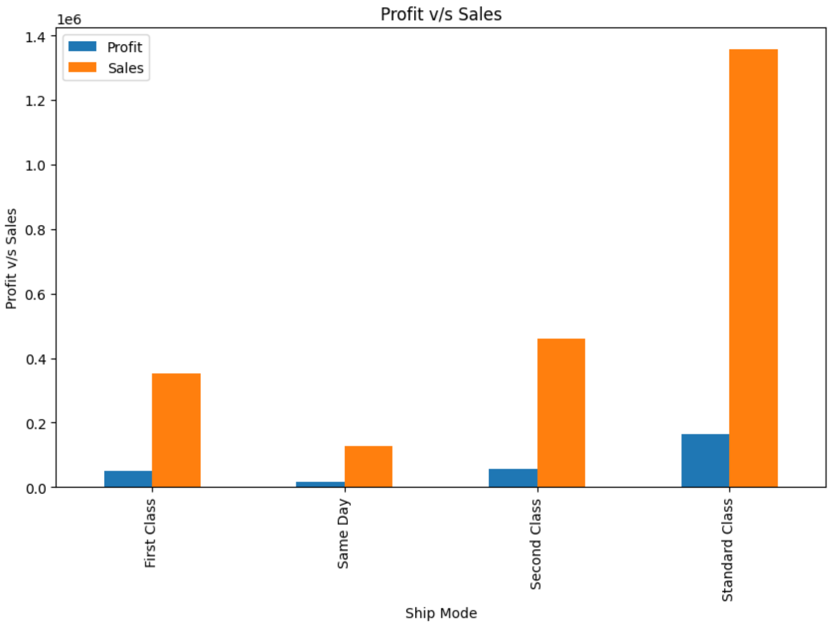
1.Most opted ship mode:Standard class

2.Least opted ship mode:same day

```
ps=df.groupby("Ship Mode")[["Profit","Sales"]].sum()
ps
```

	Profit	Sales
Ship Mode		
First Class	48910.4477	3.512746e+05
Same Day	15871.8869	1.283217e+05
Second Class	57425.5716	4.591240e+05
Standard Class	163889.6517	1.355879e+06

```
ps.plot(kind="bar",figsize=(10,6))
plt.title("Profit v/s Sales")
plt.ylabel("Profit v/s Sales")
plt.show()
```



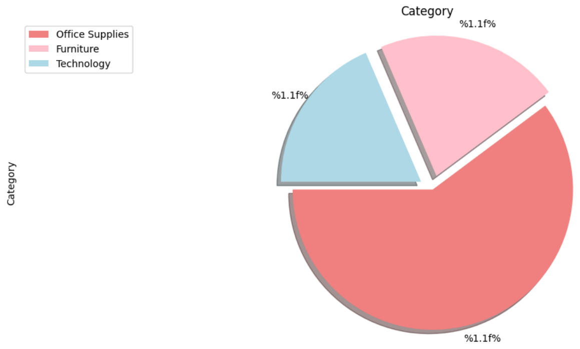
1. Maximum profit and sales:Standard class.
2. Minimum profit and sales:Same day.

```
cat=df["Category"].value_counts()
cat
```

```
Office Supplies    5986
Furniture          2114
Technology         1844
Name: Category, dtype: int64
```

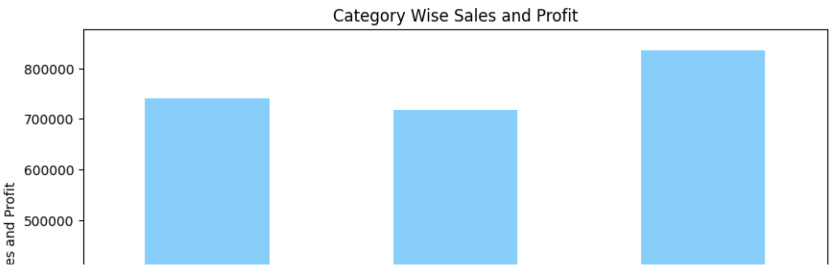
```
explode_list=[0,0.1,0.1]
colors_list=["lightcoral","pink","lightblue"]
cat.plot(kind="pie",figsize=(15,6),autopct="%%1.1f%%", shadow=True, startangle=180,explode=explode_list, labels=None, pctdistance=1.12, c
```

```
plt.title("Category")
plt.axis("equal")
plt.legend(labels=cat.index,loc="upper left")
plt.show()
```



```
#category wise sales
cs=df.groupby("Category").Sales.sum()
#category wise profit
cp=df.groupby("Category").Profit.sum()
```

```
#sales
ax0=cs.plot(kind="bar",figsize=(10,6),color='lightskyblue')

#profit
ax0=cp.plot(kind="bar",figsize=(10,6),color="lightcoral")

ax0.set_title("Category Wise Sales and Profit")
ax0.set_ylabel("Total Sales and Profit")

plt.show()
```



1.Maximum Profit and sales in Category -Technology 2.Minimum Profit and Sales in Cateogory- Office supplies

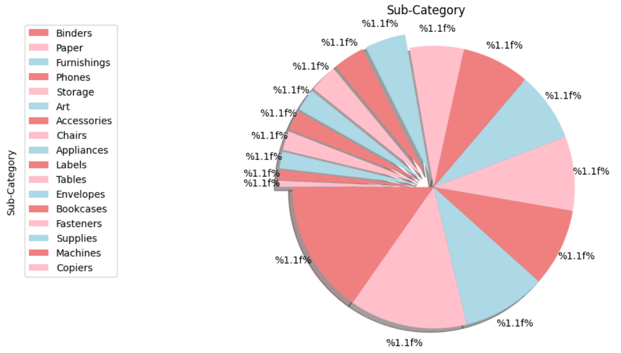
Sub-category

```
sub=df["Sub-Category"].value_counts()
sub
```

```
    Binders        1518
    Paper          1344
    Furnishings     954
    Phones          888
    Storage         845
    Art             793
    Accessories     773
    Chairs          615
    Appliances      466
    Labels          359
    Tables          319
    Envelopes       254
    Bookcases       226
    Fasteners       217
    Supplies        190
    Machines        115
    Copiers          68
    Name: Sub-Category, dtype: int64
```

```
from matplotlib import cm
cmap = cm.get_cmap("Spectral")
explode_list=[0,0,0,0,0,0,0,0,0,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1]
sub.plot(kind="pie",figsize=(15,6),autopct="%%1.1f%%", shadow=True, startangle=180,explode=explode_list, labels=None, pctdistance=1.12,


plt.title("Sub-Category")
plt.axis("equal")
plt.legend(labels=sub.index,loc="upper left")
plt.show()
```

```
    <ipython-input-62-366d5d746f55>:2: MatplotlibDeprecationWarning: The get_cmap func
      cmap = cm.get_cmap("Spectral")
```



1.Major Distribution of business in binders,papers and furnishing.

2.Least Distribution Of Business in Copiers,Machines and supplies.

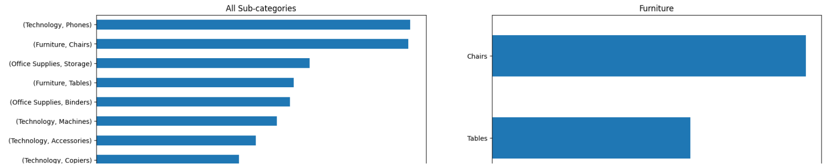
Category and Sub-Category

```
x=df.groupby(["Category","Sub-Category"]).Sales.sum()
y=df.groupby(["Category","Sub-Category"]).Profit.sum()
```

Sales

```
fig=plt.figure()
ax0=fig.add_subplot(2,2,1)
ax1=fig.add_subplot(2,2,2)
ax2=fig.add_subplot(2,2,3)
ax3=fig.add_subplot(2,2,4)

#Furniture
x["Furniture"].sort_values(ascending=True).plot(kind="barh",figsize=(20,20),ax=ax1)
ax1.set_title('Furniture')
ax1.set_xlabel('Sub-Categories')
ax1.set_ylabel('Sales')

#Technology
x["Technology"].sort_values(ascending=True).plot(kind="barh",figsize=(20,20),ax=ax2)
ax2.set_title('Technology')
ax2.set_xlabel('Sub-Categories')
ax2.set_ylabel('Sales')

#Office supplies
x["Office Supplies"].sort_values(ascending=True).plot(kind="barh",figsize=(20,20),ax=ax3)
ax3.set_title('Office Supplies')
ax3.set_xlabel('Sub-Categories')
ax3.set_ylabel('Sales')

#total
x.sort_values(ascending=True).plot(kind="barh",figsize=(20,20),ax=ax0)
ax0.set_title('All Sub-categories')
ax0.set_xlabel('Sub-Categories')
ax0.set_ylabel('Sales')
```

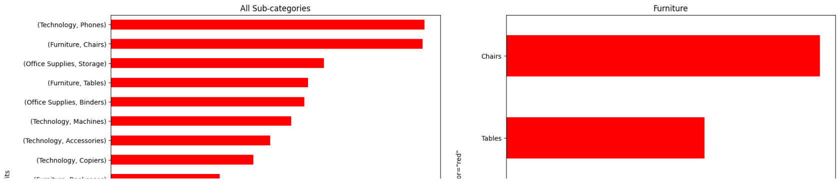
```
Text(0, 0.5, 'Sales')
```



Profit

```
fig=plt.figure()
ax0=fig.add_subplot(2,2,1)
ax1=fig.add_subplot(2,2,2)
ax2=fig.add_subplot(2,2,3)
ax3=fig.add_subplot(2,2,4)

#Furniture
x["Furniture"].sort_values(ascending=True).plot(kind="barh",figsize=(20,20),ax=ax1,color="red")
ax1.set_title('Furniture')
ax1.set_xlabel('Sub-Categories')
ax1.set_ylabel('Profits,color="red"')

#Technology
x["Technology"].sort_values(ascending=True).plot(kind="barh",figsize=(20,20),ax=ax2,color="red")
ax2.set_title('Technology')
ax2.set_xlabel('Sub-Categories')
ax2.set_ylabel('Profits')

#Office supplies
x["Office Supplies"].sort_values(ascending=True).plot(kind="barh",figsize=(20,20),ax=ax3,color="red")
ax3.set_title('Office Supplies')
ax3.set_xlabel('Sub-Categories')
ax3.set_ylabel('Profits')

#total
x.sort_values(ascending=True).plot(kind="barh",figsize=(20,20),ax=ax0,color="red")
ax0.set_title('All Sub-categories')
ax0.set_xlabel('Sub-Categories')
ax0.set_ylabel('Profits')
```

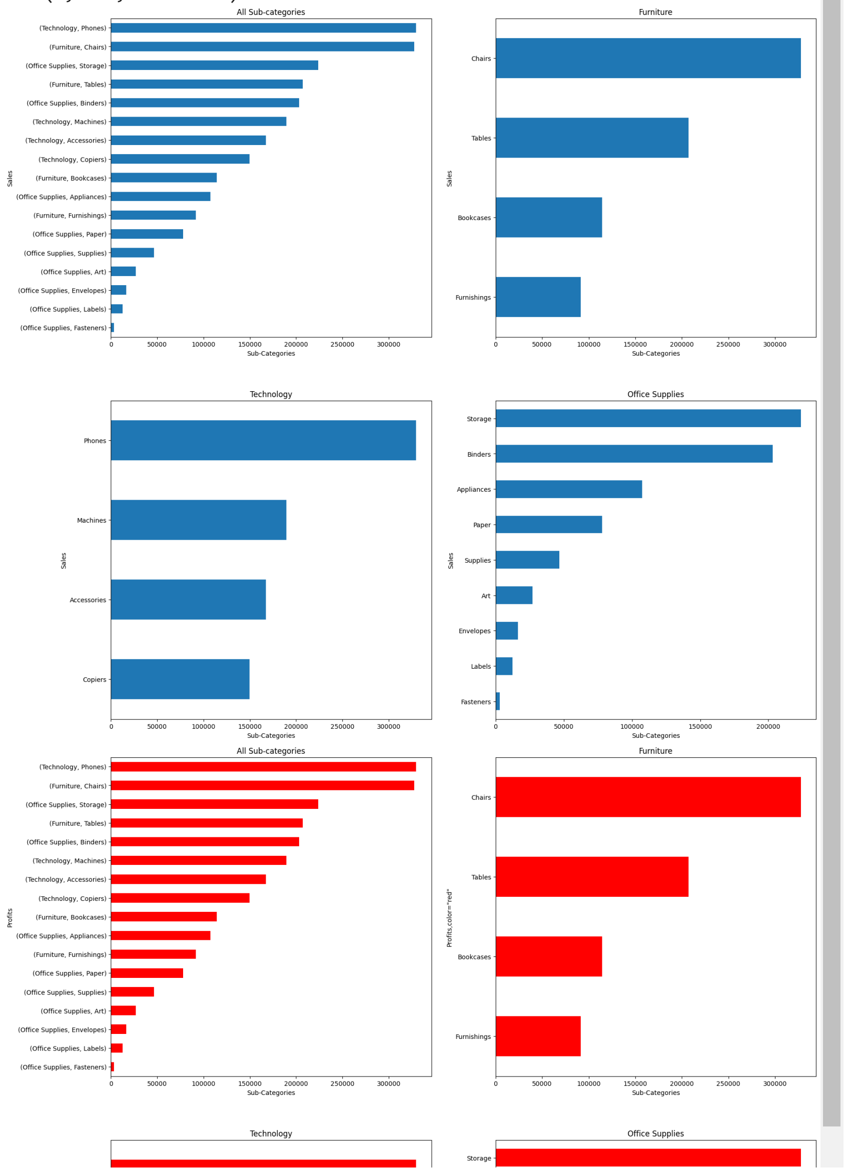
```
Text(0, 0.5, 'Profits')
```



Combined sales and profit

```
#sales
fig=plt.figure()
ax0=fig.add_subplot(2,2,1)
ax1=fig.add_subplot(2,2,2)
ax2=fig.add_subplot(2,2,3)
ax3=fig.add_subplot(2,2,4)

#Furniture
x["Furniture"].sort_values(ascending=True).plot(kind="barh",figsize=(20,20),ax=ax1)
ax1.set_title('Furniture')
ax1.set_xlabel('Sub-Categories')
ax1.set_ylabel('Sales')

#Technology
x["Technology"].sort_values(ascending=True).plot(kind="barh",figsize=(20,20),ax=ax2)
ax2.set_title('Technology')
ax2.set_xlabel('Sub-Categories')
ax2.set_ylabel('Sales')

#Office supplies
x["Office Supplies"].sort_values(ascending=True).plot(kind="barh",figsize=(20,20),ax=ax3)
ax3.set_title('Office Supplies')
ax3.set_xlabel('Sub-Categories')
ax3.set_ylabel('Sales')

#total
x.sort_values(ascending=True).plot(kind="barh",figsize=(20,20),ax=ax0)
ax0.set_title('All Sub-categories')
ax0.set_xlabel('Sub-Categories')
ax0.set_ylabel('Sales')

#profit
fig=plt.figure()
ax0=fig.add_subplot(2,2,1)
ax1=fig.add_subplot(2,2,2)
ax2=fig.add_subplot(2,2,3)
ax3=fig.add_subplot(2,2,4)

#Furniture
x["Furniture"].sort_values(ascending=True).plot(kind="barh",figsize=(20,20),ax=ax1,color="red")
ax1.set_title('Furniture')
ax1.set_xlabel('Sub-Categories')
ax1.set_ylabel('Profits,color="red"')

#Technology
x["Technology"].sort_values(ascending=True).plot(kind="barh",figsize=(20,20),ax=ax2,color="red")
ax2.set_title('Technology')
ax2.set_xlabel('Sub-Categories')
ax2.set_ylabel('Profits')

#Office supplies
x["Office Supplies"].sort_values(ascending=True).plot(kind="barh",figsize=(20,20),ax=ax3,color="red")
ax3.set_title('Office Supplies')
ax3.set_xlabel('Sub-Categories')
ax3.set_ylabel('Profits')

#total
x.sort_values(ascending=True).plot(kind="barh",figsize=(20,20),ax=ax0,color="red")
ax0.set_title('All Sub-categories')
ax0.set_xlabel('Sub-Categories')
ax0.set_ylabel('Profits')
```

```
Text(0, 0.5, 'Profits')
```



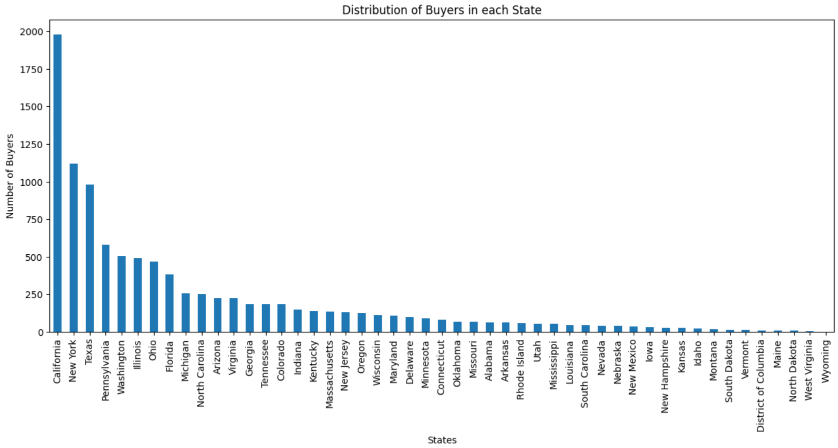
```
state=df["State"].value_counts()
state
```

```
    California              1980
    New York                1119
    Texas                    980
    Pennsylvania             581
    Washington               502
    Illinois                 490
    Ohio                     468
    Florida                  383
    Michigan                 254
    North Carolina           249
    Arizona                  224
    Virginia                 224
    Georgia                  184
    Tennessee                183
    Colorado                 182
    Indiana                  149
    Kentucky                 139
    Massachusetts            135
    New Jersey               130
    Oregon                   123
    Wisconsin                110
    Maryland                 105
    Delaware                  96
    Minnesota                 89
    Connecticut               82
    Oklahoma                  66
    Missouri                  66
    Alabama                   61
    Arkansas                  60
    Rhode Island              56
    Utah                      53
    Mississippi               53
    Louisiana                 42
    South Carolina            42
```

```
Nevada                    39
Nebraska                  38
New Mexico                37
Iowa                      30
New Hampshire             27
Kansas                    24
Idaho                     21
Montana                   15
South Dakota              12
Vermont                   11
District of Columbia      10
Maine                      8
North Dakota               7
West Virginia              4
Wyoming                    1
Name: State, dtype: int64
```

```
state.plot(kind='bar',figsize=(15,6))
plt.title("Distribution of Buyers in each State")
plt.xlabel("States")
plt.ylabel("Number of Buyers")
```

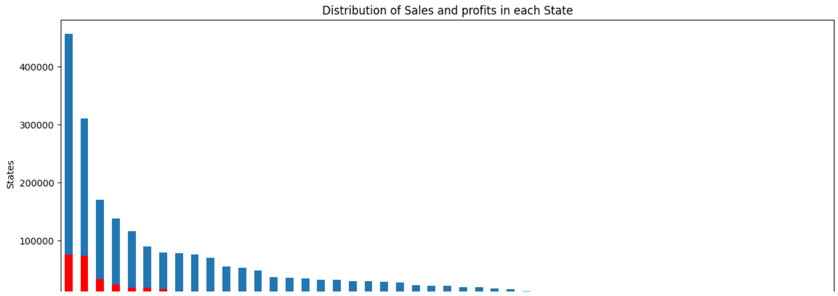
```
Text(0, 0.5, 'Number of Buyers')
```



```
ss=df.groupby(["State"]).Sales.sum()
sp=df.groupby(["State"]).Profit.sum()
```

```
#State wise distribution of sales
ss.sort_values(ascending=False).plot(kind="bar", figsize=(15, 6))
plt.xlabel("States")
plt.ylabel("States")

#State wise distribution of Profits
sp.sort_values(ascending=False).plot(kind="bar", figsize=(15,6), color="red")
plt.title("Distribution of Sales and profits in each State")
plt.xlabel("States")
plt.ylabel("States")
```

```
Text(0, 0.5, 'States')
```



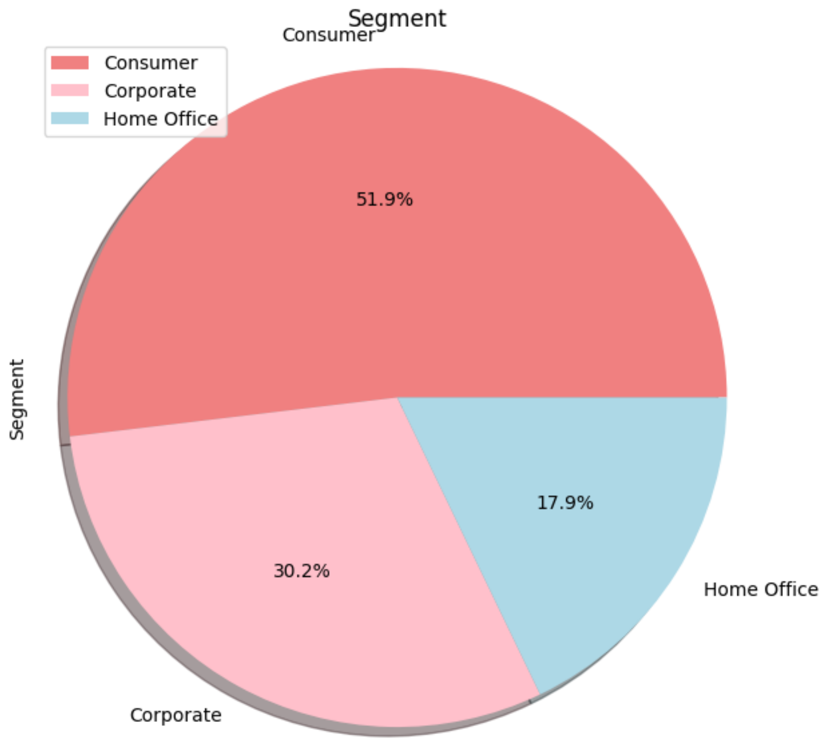
1.Maximum sales in profit comes from California and New york 2.Minimum sales and profit comes from Oregon,Florida,Arizona,Tennesse,Colorado,North Carolina,llionis

Segment

```
seg=df["Segment"].value_counts()
seg
```

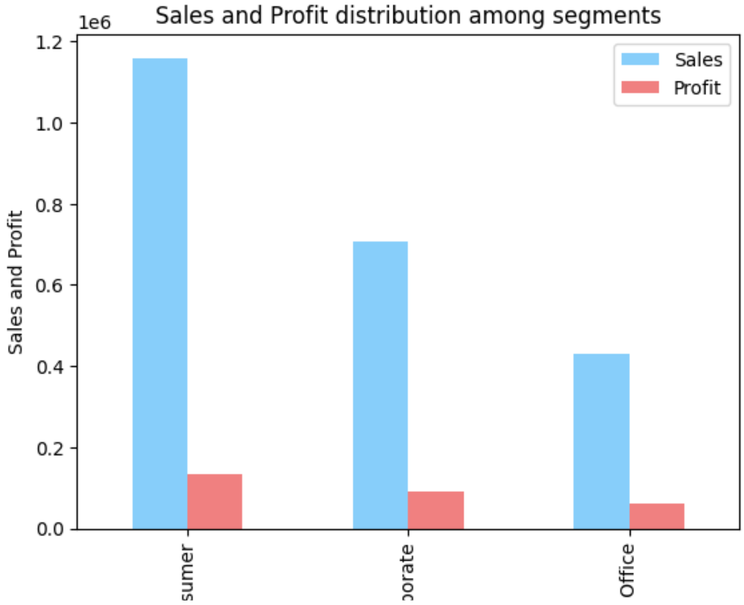
```
Consumer       5160
Corporate      3008
Home Office    1776
Name: Segment, dtype: int64
```

```
seg.plot(kind="pie",figsize=(7,7),autopct="%1.1f%%",shadow=True,label=None,colors=["lightcoral","pink","lightblue"])
plt.title("Segment")
plt.legend(labels=seg.index,loc="upper left")
plt.axis("equal")
plt.show()
```



Sales and profit in different segments

```
df.groupby("Segment")[["Sales", "Profit"]].sum().plot(kind="bar", color=["lightskyblue", "lightcoral"])
plt.ylabel("Sales and Profit")
plt.xlabel("Segment")
plt.title("Sales and Profit distribution among segments")
plt.show()
```



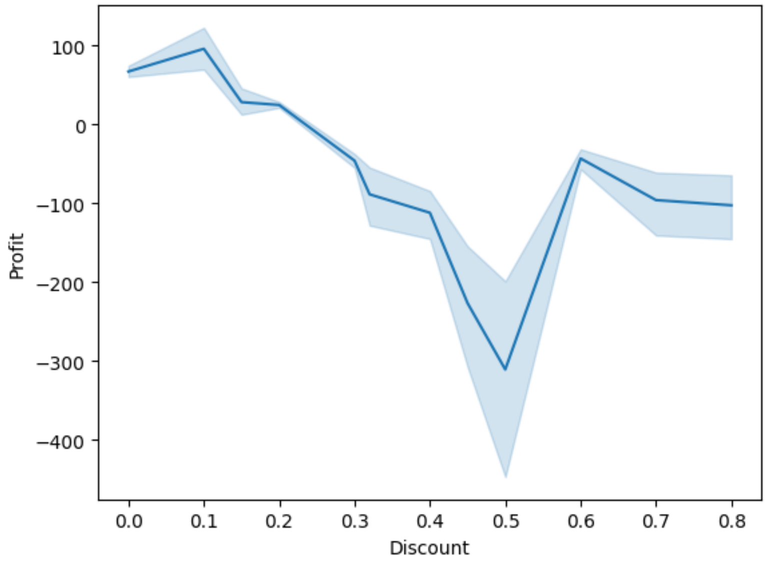
Most sales and profits are in consumer segment and least in Home office

Relation between Profit and Discount

```
df['Discount'].corr(df["Profit"])
```

```
-0.21993898249765037
```

```
sns.lineplot(x="Discount",y="Profit",data=df)
plt.show()
```

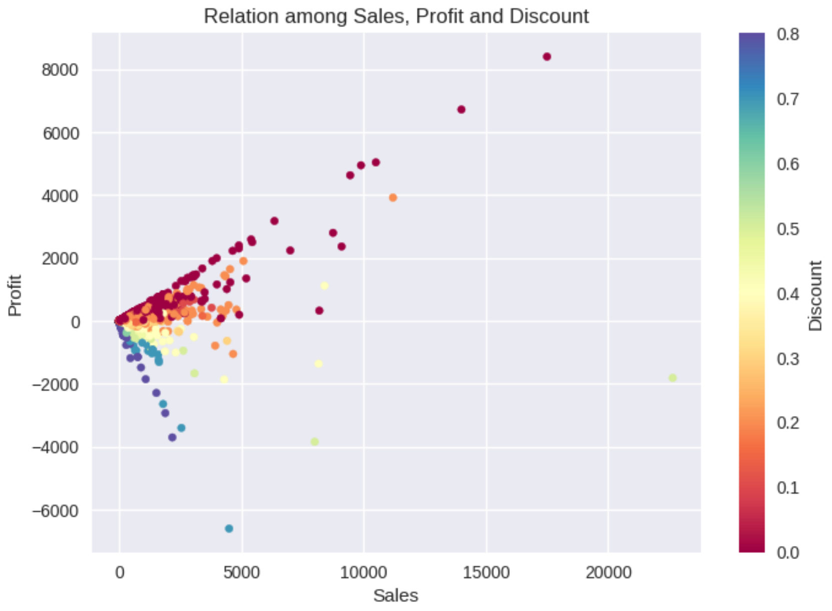


Profit and Discount have negative relation

Relation Among Sales,Profit and Discount

```
plt.style.use("seaborn")

df.plot(kind="scatter",x="Sales",y="Profit",c="Discount", colormap="Spectral")

plt.xlabel("Sales")

plt.ylabel("Profit")

plt.title("Relation among Sales, Profit and Discount")

plt.show()
```

```
<ipython-input-91-8055c9375965>:1: MatplotlibDeprecationWarning: The seaborn style
  plt.style.use("seaborn")
```


Relation among Sales, Profit and Discount
Profit
Sales
Discount