

```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [4]: df=pd.read_csv("SampleSuperstore.csv")
df.head()
```

Out[4]:

	Ship Mode	Segment	Country	City	State	Postal Code	Region	Category	Sub-Category	
0	Second Class	Consumer	United States	Henderson	Kentucky	42420	South	Furniture	Bookcases	261
1	Second Class	Consumer	United States	Henderson	Kentucky	42420	South	Furniture	Chairs	731
2	Second Class	Corporate	United States	Los Angeles	California	90036	West	Office Supplies	Labels	14
3	Standard Class	Consumer	United States	Fort Lauderdale	Florida	33311	South	Furniture	Tables	957
4	Standard Class	Consumer	United States	Fort Lauderdale	Florida	33311	South	Office Supplies	Storage	22

```
In [5]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9994 entries, 0 to 9993
Data columns (total 13 columns):
Ship Mode      9994 non-null object
Segment        9994 non-null object
Country        9994 non-null object
City           9994 non-null object
State          9994 non-null object
Postal Code    9994 non-null int64
Region         9994 non-null object
Category       9994 non-null object
Sub-Category   9994 non-null object
Sales          9994 non-null float64
Quantity       9994 non-null int64
Discount       9994 non-null float64
Profit         9994 non-null float64
dtypes: float64(3), int64(2), object(8)
memory usage: 1015.1+ KB
```

```
In [6]: df.isnull().sum()
```

```
Out[6]: 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
```

```
In [7]: df.columns
```

```
Out[7]: Index(['Ship Mode', 'Segment', 'Country', 'City', 'State', 'Postal Code',
              'Region', 'Category', 'Sub-Category', 'Sales', 'Quantity', 'Discount',
              'Profit'],
              dtype='object')
```

```
In [8]: df.shape
```

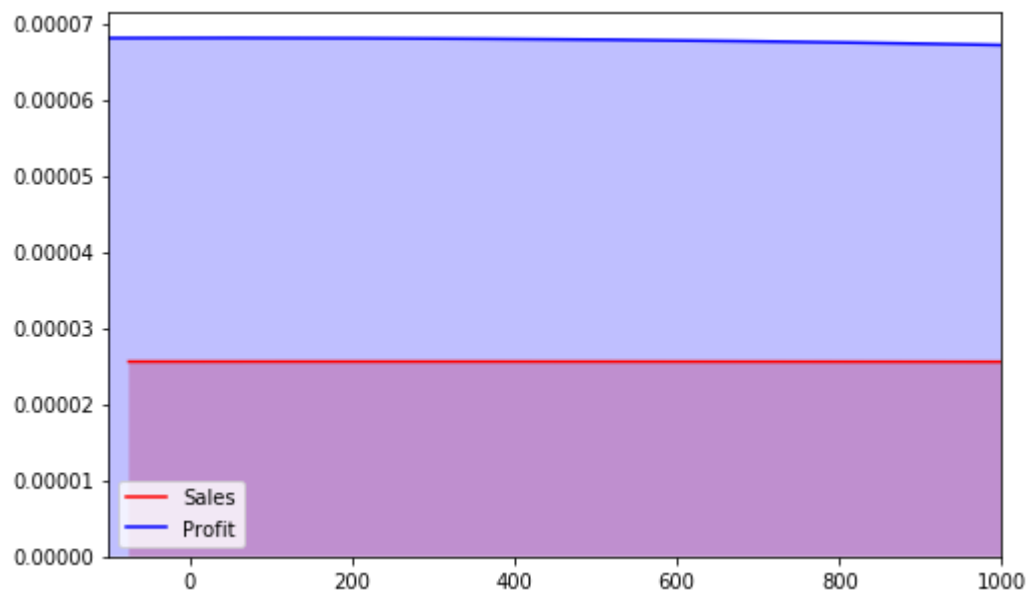
```
Out[8]: (9994, 13)
```

```
In [9]: df.nunique()
```

```
Out[9]: Ship Mode      4
        Segment      3
        Country      1
        City        531
        State       49
        Postal Code  631
        Region      4
        Category     3
        Sub-Category 17
        Sales      5825
        Quantity    14
        Discount    12
        Profit     7287
        dtype: int64
```

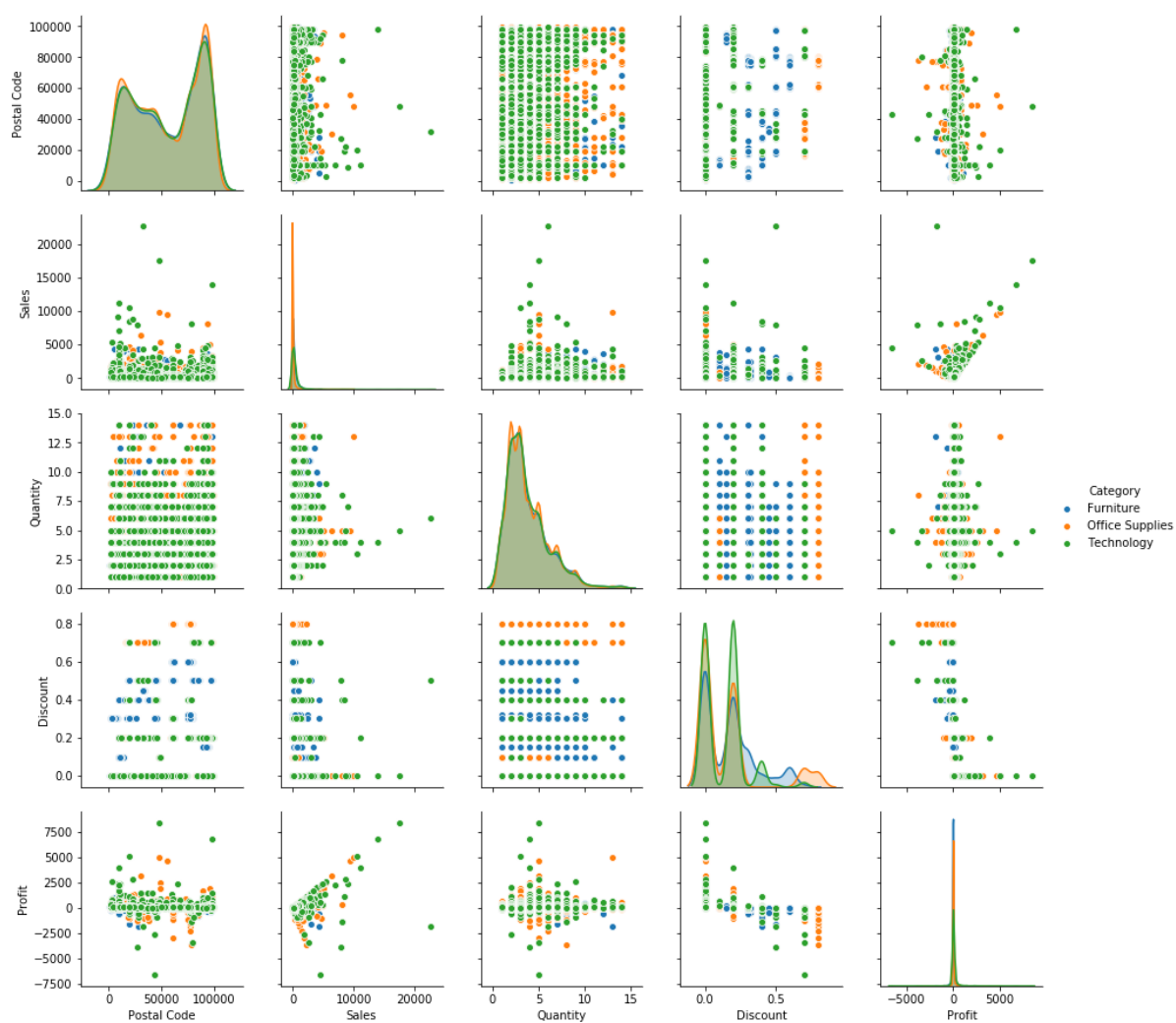
```
In [10]: plt.figure(figsize=(8,5))
sns.kdeplot(df['Sales'],color='red',label='Sales',shade=True,bw=25)
sns.kdeplot(df['Profit'],color='Blue',label='Profit',shade=True,bw=25)
plt.xlim([-100,1000])
plt.legend()
```

Out[10]: <matplotlib.legend.Legend at 0x1369bda5be0>



```
In [11]: sns.pairplot(df,hue='Category')
```

```
Out[11]: <seaborn.axisgrid.PairGrid at 0x1369be7c710>
```



```
In [12]: sns.pairplot(df,hue='Region')
```

```
Out[12]: <seaborn.axisgrid.PairGrid at 0x1369f1434a8>
```



```
In [13]: sns.pairplot(df,hue='Segment')
```

```
Out[13]: <seaborn.axisgrid.PairGrid at 0x136a1528b70>
```



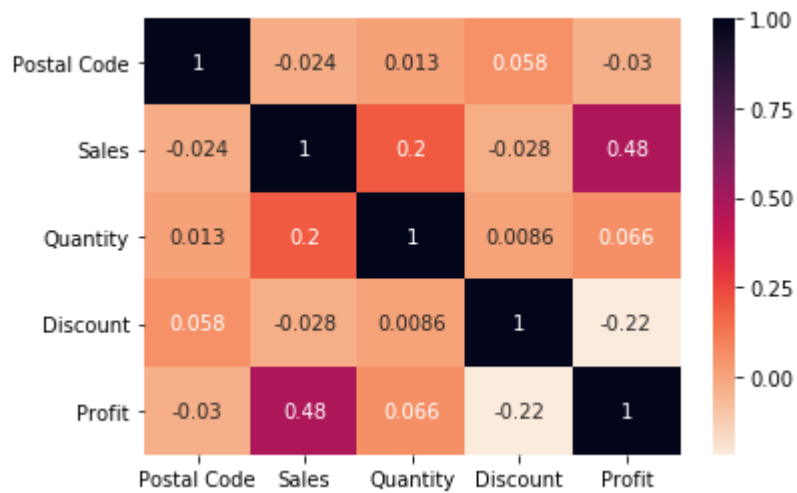
```
In [14]: df.corr()
```

```
Out[14]:
```

	Postal Code	Sales	Quantity	Discount	Profit
Postal Code	1.000000	-0.023854	0.012761	0.058443	-0.029961
Sales	-0.023854	1.000000	0.200795	-0.028190	0.479064
Quantity	0.012761	0.200795	1.000000	0.008623	0.066253
Discount	0.058443	-0.028190	0.008623	1.000000	-0.219487
Profit	-0.029961	0.479064	0.066253	-0.219487	1.000000

```
In [15]: sns.heatmap(df.corr(),cmap='rocket_r',annot=True)
```

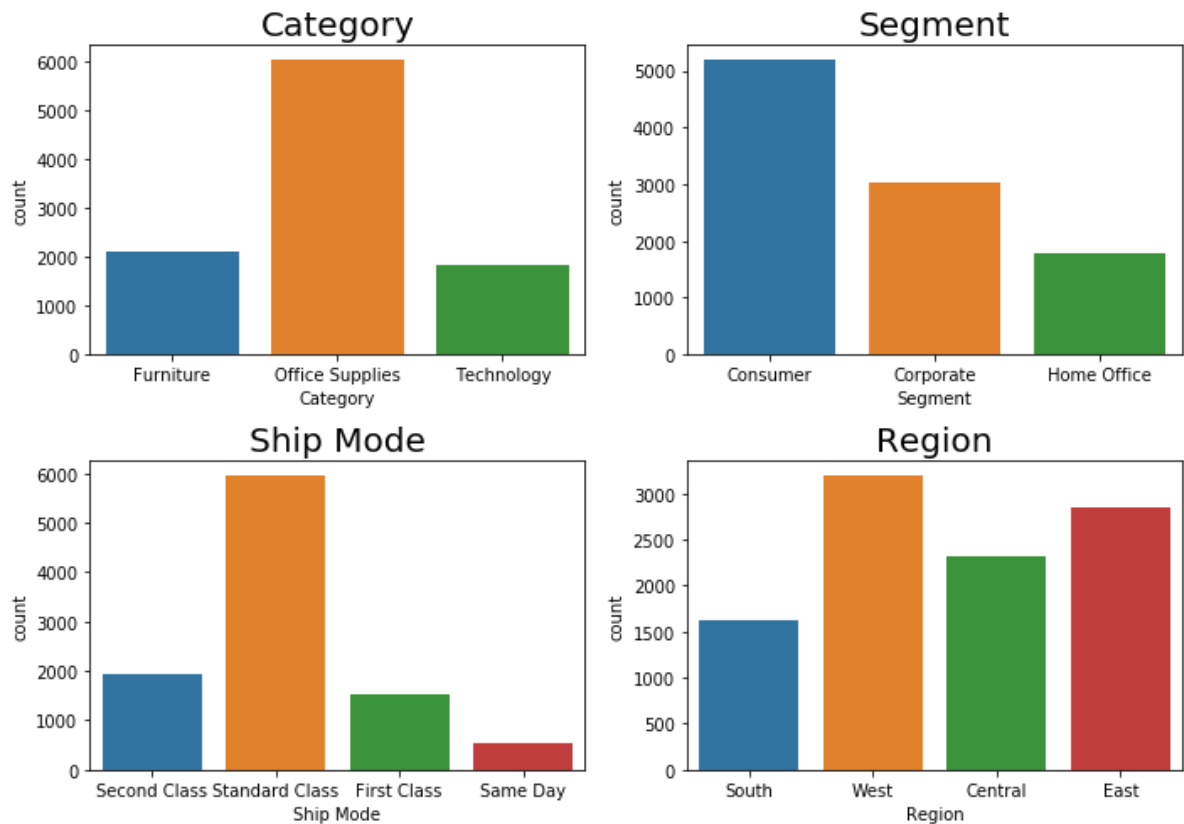
```
Out[15]: <matplotlib.axes._subplots.AxesSubplot at 0x136a38bda20>
```



```
In [16]: fig,axs=plt.subplots(nrows=2,ncols=2,figsize=(10,7));

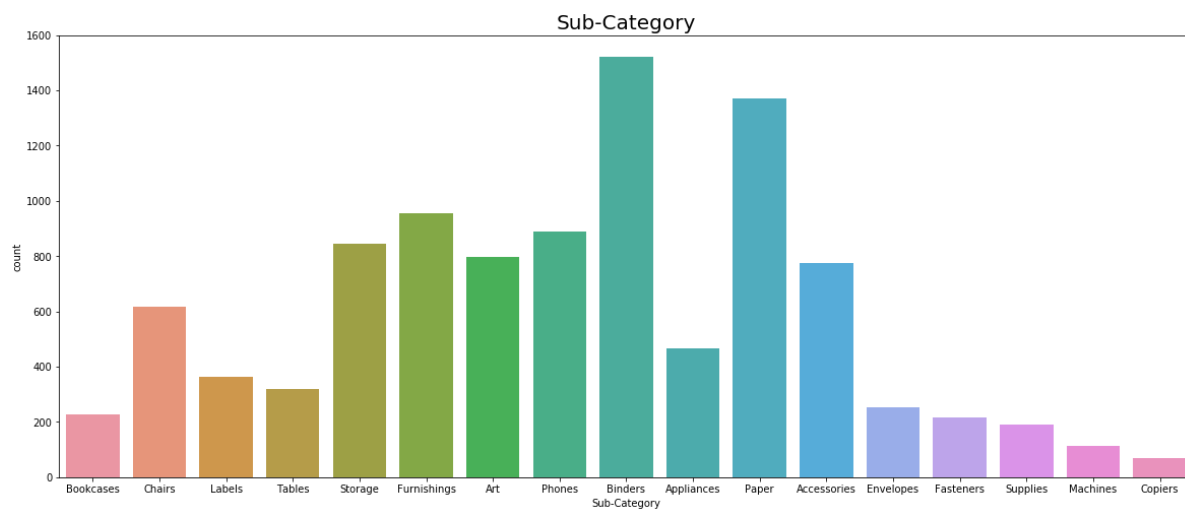
sns.countplot(df['Category'],ax=axs[0][0])
sns.countplot(df['Segment'],ax=axs[0][1])
sns.countplot(df['Ship Mode'],ax=axs[1][0])
sns.countplot(df['Region'],ax=axs[1][1])
axs[0][0].set_title('Category',fontsize=20)
axs[0][1].set_title('Segment',fontsize=20)
axs[1][0].set_title('Ship Mode',fontsize=20)
axs[1][1].set_title('Region',fontsize=20)

plt.tight_layout()
```



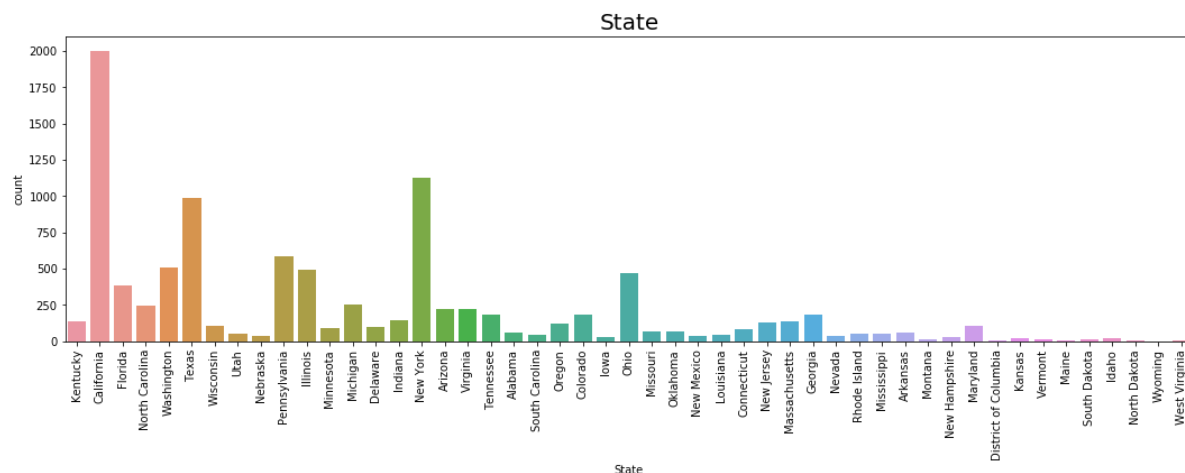

```
In [17]: plt.figure(figsize=(20,8))
sns.countplot(df['Sub-Category'])
plt.title('Sub-Category',fontsize=20)
```

Out[17]: Text(0.5, 1.0, 'Sub-Category')



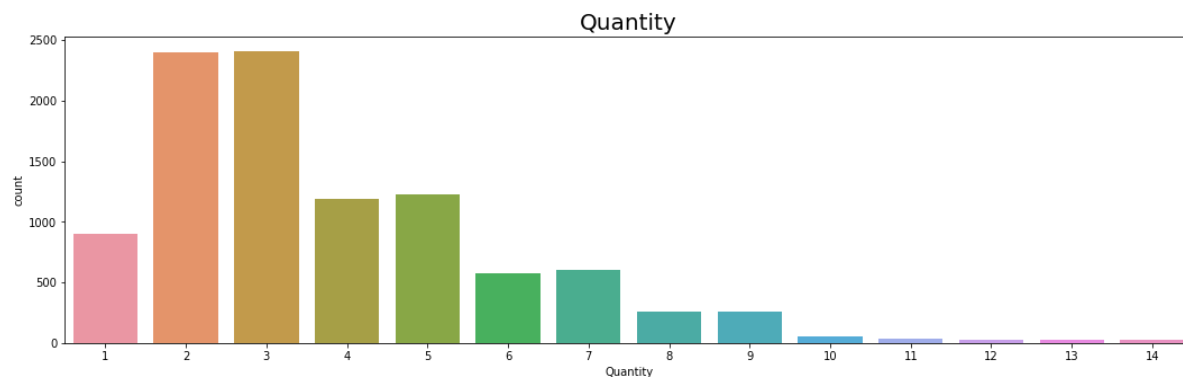
```
In [18]: plt.figure(figsize=(18,5))
sns.countplot(df['State'])
plt.xticks(rotation=90)
plt.title('State',fontsize=20)
```

Out[18]: Text(0.5, 1.0, 'State')



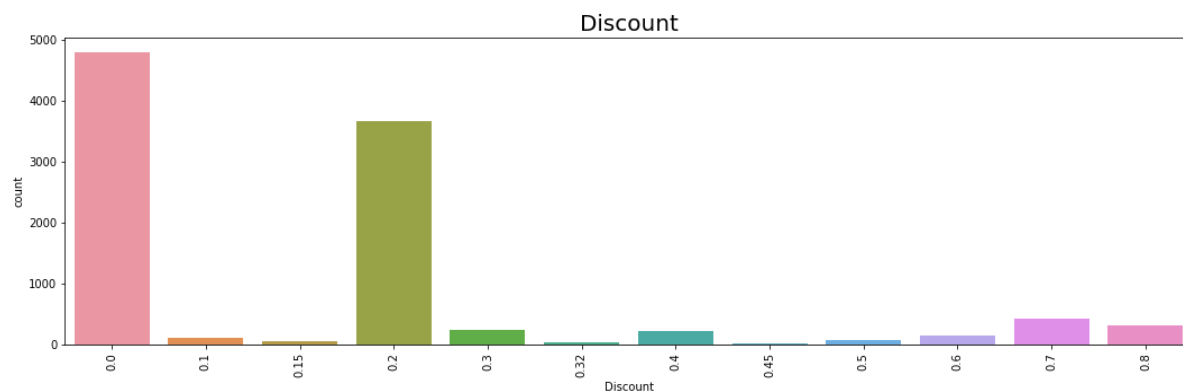
```
In [19]: plt.figure(figsize=(18,5))
sns.countplot(df['Quantity'])
plt.title('Quantity',fontsize=20)
```

Out[19]: Text(0.5, 1.0, 'Quantity')



```
In [20]: plt.figure(figsize=(18,5))
sns.countplot(df['Discount'])
plt.xticks(rotation=90)
plt.title('Discount',fontsize=20)
```

Out[20]: Text(0.5, 1.0, 'Discount')



```
In [21]: fig, axs = plt.subplots(ncols=2, nrows = 2, figsize = (10,10))
sns.distplot(df['Sales'], color = 'red', ax = axs[0][0])
sns.distplot(df['Profit'], color = 'green', ax = axs[0][1])
sns.distplot(df['Quantity'], color = 'orange', ax = axs[1][0])
sns.distplot(df['Discount'], color = 'blue', ax = axs[1][1])
axs[0][0].set_title('Sales Distribution', fontsize = 20)
axs[0][1].set_title('Profit Distribution', fontsize = 20)
axs[1][0].set_title('Quantity distribution', fontsize = 20)
axs[1][1].set_title('Discount Distribution', fontsize = 20)
plt.show()
```



```
In [22]: df['Country'].value_counts()
```

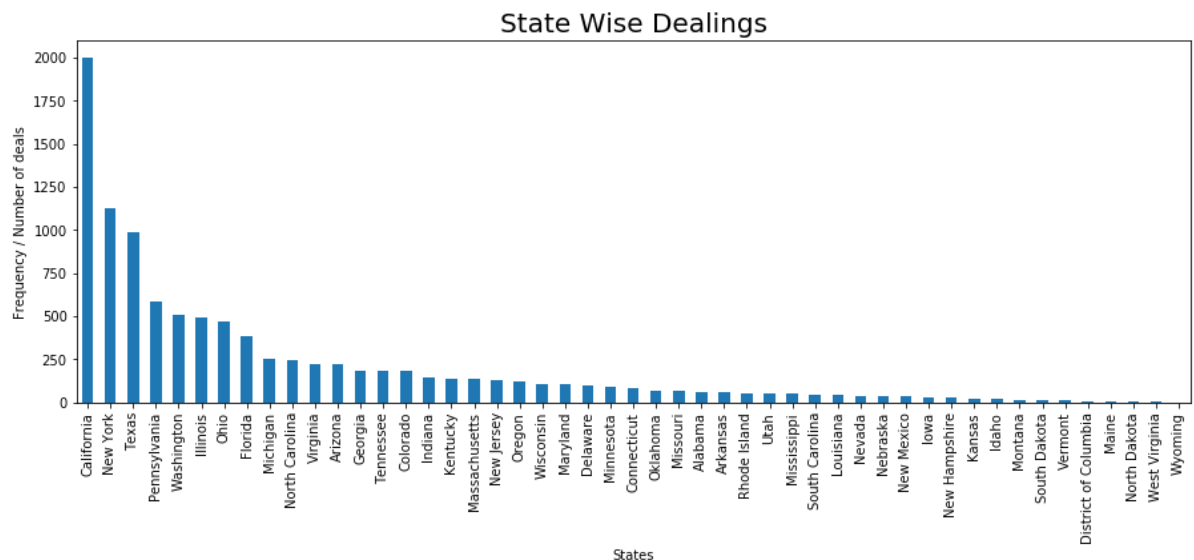
```
Out[22]: United States    9994
         Name: Country, dtype: int64
```

```
In [23]: df1 = df['State'].value_counts()
df1.head(10)
```

```
Out[23]: California      2001
New York      1128
Texas         985
Pennsylvania   587
Washington     506
Illinois       492
Ohio           469
Florida        383
Michigan       255
North Carolina 249
Name: State, dtype: int64
```

```
In [24]: df1.plot(kind='bar',figsize=(15,5))
plt.ylabel('Frequency / Number of deals')
plt.xlabel('States')

plt.title('State Wise Dealings', fontsize = 20)
plt.show()
```



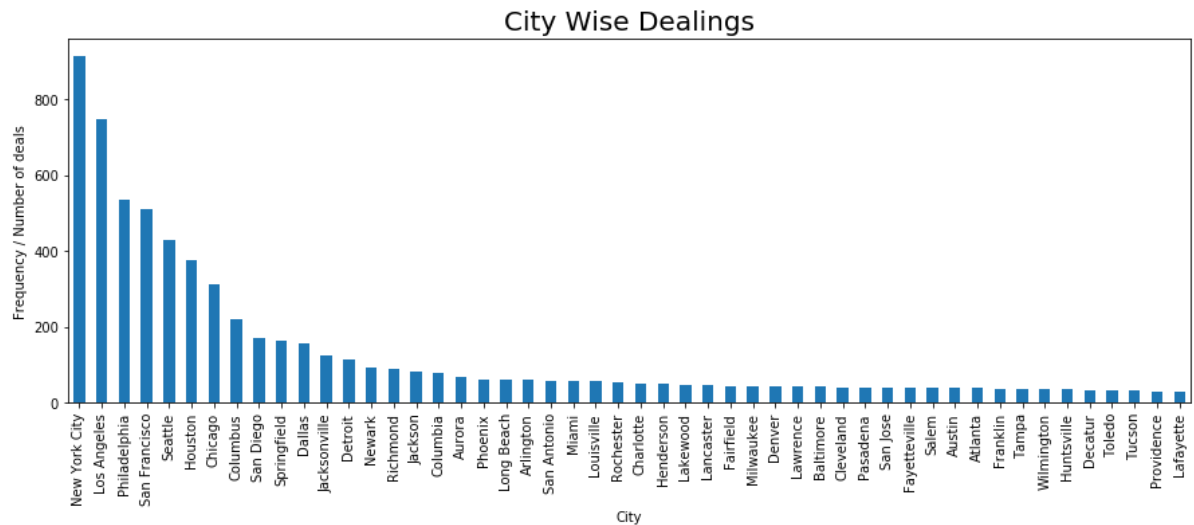
```
In [25]: df['State'].value_counts().mean()
```

```
Out[25]: 203.9591836734694
```

```
In [26]: df2 = df['City'].value_counts()
df2=df2.head(50)
```

```
In [27]: df2.plot(kind='bar',figsize=(15,5))
plt.ylabel('Frequency / Number of deals')
plt.xlabel('City')

plt.title('City Wise Dealings', fontsize = 20)
plt.show()
```



```
In [28]: df['City'].value_counts().mean()
```

```
Out[28]: 18.821092278719398
```

```
In [29]: df['Segment'].value_counts()
```

```
Out[29]: Consumer      5191
Corporate      3020
Home Office    1783
Name: Segment, dtype: int64
```

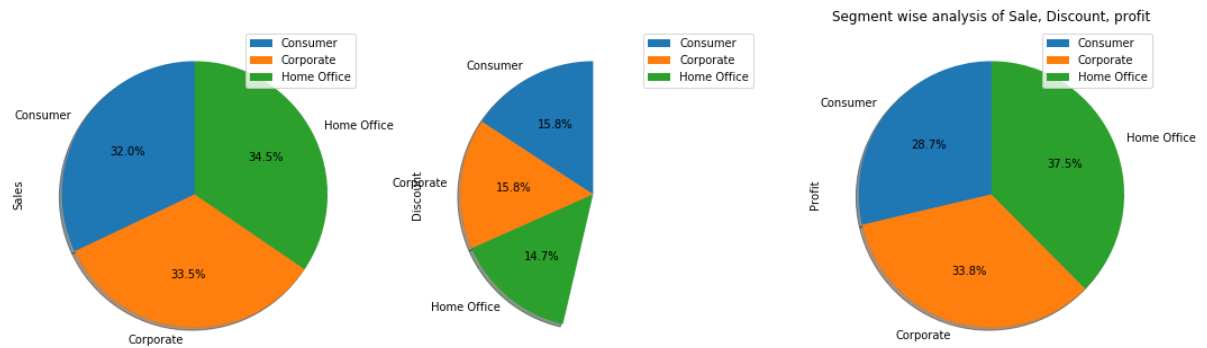
```
In [30]: df_segment= df.groupby(['Segment'])[['Sales', 'Discount', 'Profit']].mean()
df_segment
```

```
Out[30]:
```

	Sales	Discount	Profit
Segment			
Consumer	223.733644	0.158141	25.836873
Corporate	233.823300	0.158228	30.456667
Home Office	240.972041	0.147128	33.818664

```
In [31]: #1. sales 2. Discount 3. Profit
df_segment.plot.pie(subplots=True,
                    autopct='%1.1f%%',
                    figsize=(18, 20),
                    startangle=90,      # start angle 90° (Africa)
                    shadow=True,
                    labels = df_segment.index)
plt.title('Segment wise analysis of Sale, Discount, profit')
```

Out[31]: Text(0.5, 1.0, 'Segment wise analysis of Sale, Discount, profit')



```
In [32]: df['State'].value_counts().head(10)
```

Out[32]:

California	2001
New York	1128
Texas	985
Pennsylvania	587
Washington	506
Illinois	492
Ohio	469
Florida	383
Michigan	255
North Carolina	249

Name: State, dtype: int64

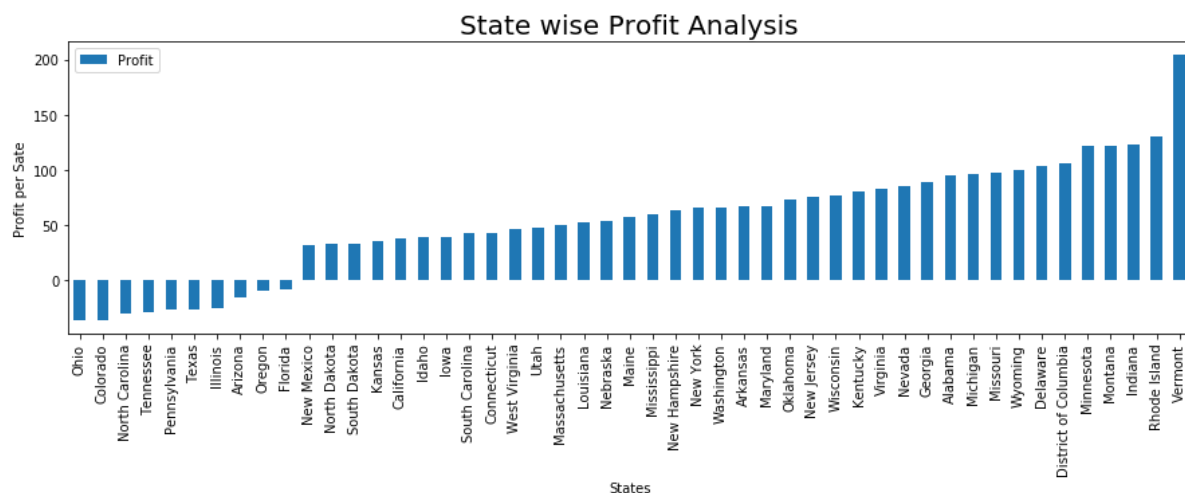
```
In [33]: df_state= df.groupby(['State'])[['Sales', 'Discount', 'Profit']].mean()
df_state.head(10)
```

Out[33]:

	Sales	Discount	Profit
State			
Alabama	319.846557	0.000000	94.865989
Arizona	157.508933	0.303571	-15.303235
Arkansas	194.635500	0.000000	66.811452
California	228.729451	0.072764	38.171608
Colorado	176.418231	0.316484	-35.867351
Connecticut	163.223866	0.007317	42.823071
Delaware	285.948635	0.006250	103.930988
District of Columbia	286.502000	0.000000	105.958930
Florida	233.612815	0.299347	-8.875461
Georgia	266.825217	0.000000	88.315453

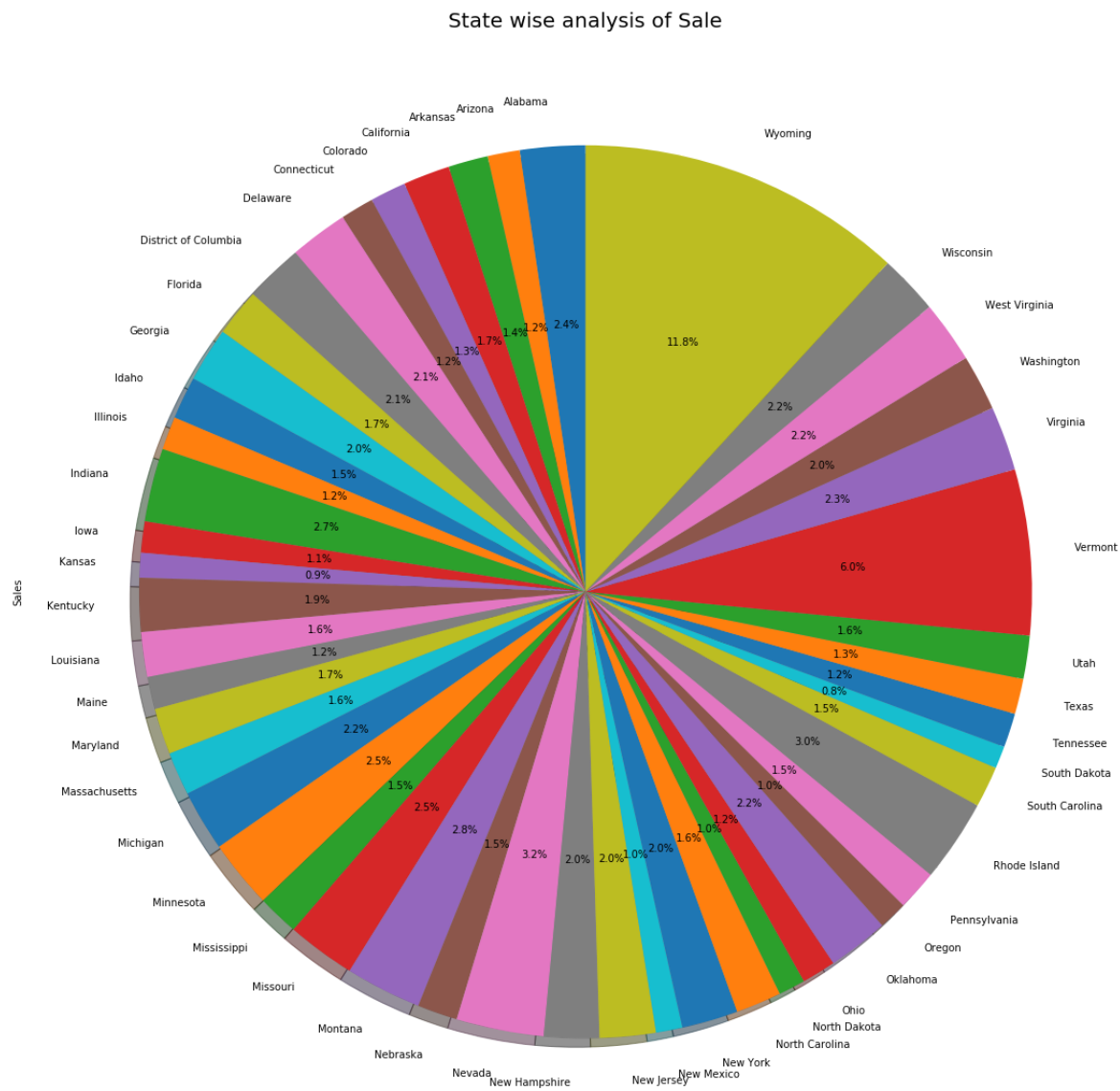
```
In [34]: df_state1=df_state.sort_values('Profit')

df_state1[['Profit']].plot(kind = 'bar', figsize = (15,4))
plt.title('State wise Profit Analysis', fontsize = 20)
plt.ylabel('Profit per Sate')
plt.xlabel('States')
plt.show()
```



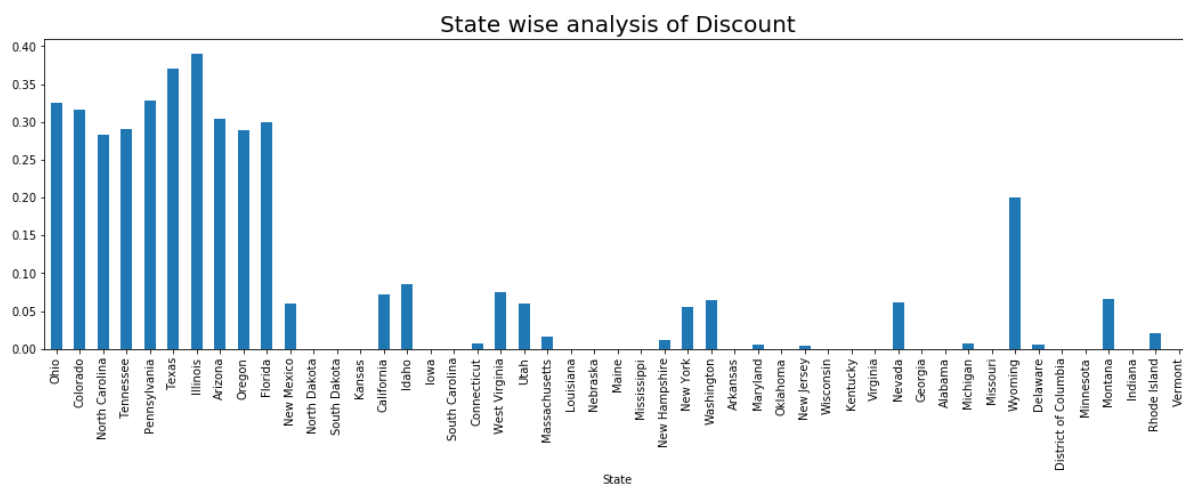
```
In [35]: df_state['Sales'].plot(kind='pie',
    figsize = (20,20),
    autopct='%1.1f%%',
    startangle=90,      # start angle 90° (Africa)
    shadow=True)
plt.title('State wise analysis of Sale',fontsize=20)
```

```
Out[35]: Text(0.5, 1.0, 'State wise analysis of Sale')
```




```
In [36]: df_state1['Discount'].plot(kind='bar',figsize=(18,5))
plt.title('State wise analysis of Discount', fontsize=20)
```

```
Out[36]: Text(0.5, 1.0, 'State wise analysis of Discount')
```



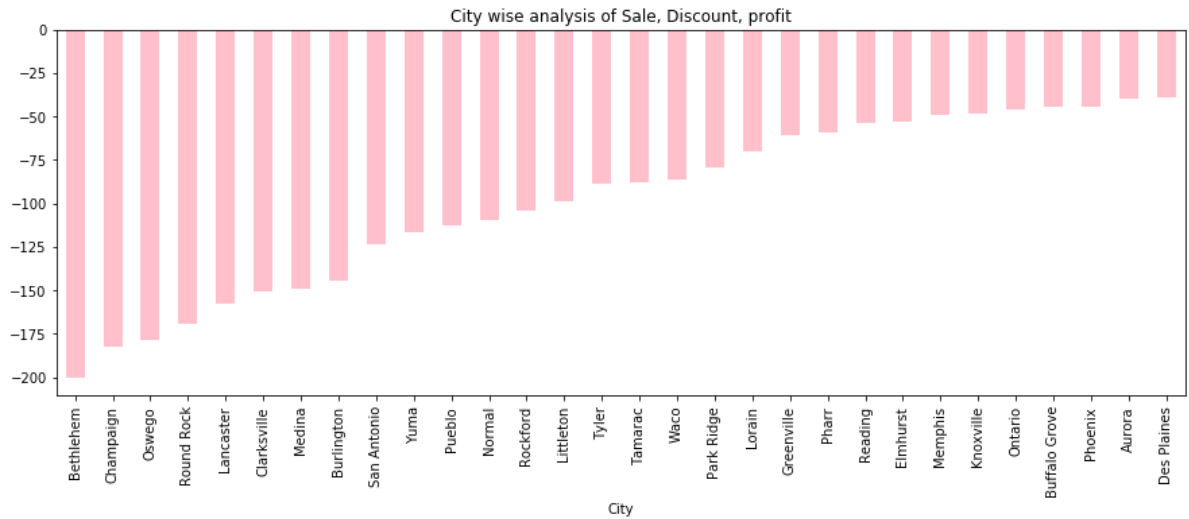
```
In [37]: df_city= df.groupby(['City'])[['Sales', 'Discount', 'Profit']].mean()
df_city = df_city.sort_values('Profit')
df_city.head()
```

```
Out[37]:
```

	Sales	Discount	Profit
City			
Bethlehem	337.926800	0.380000	-200.619160
Champaign	151.960000	0.600000	-182.352000
Oswego	107.326000	0.600000	-178.709200
Round Rock	693.436114	0.274286	-169.061614
Lancaster	215.031826	0.315217	-157.371052

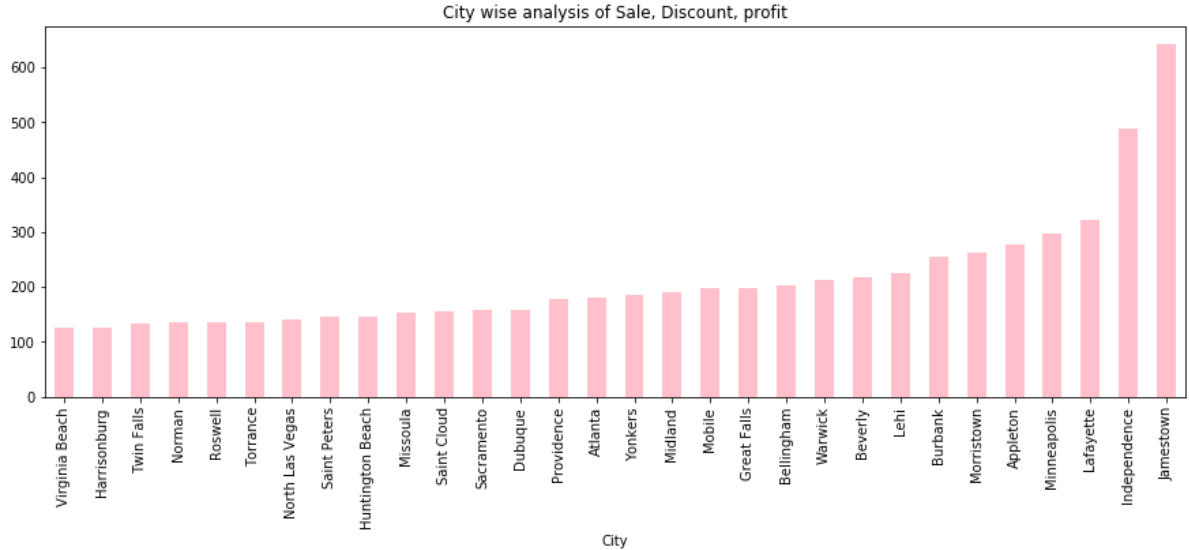
In [38]: *#1. Low Profit*
`df_city['Profit'].head(30).plot(kind='bar',figsize=(15,5),color = 'Pink')`
`plt.title('City wise analysis of Sale, Discount, profit')`

Out[38]: Text(0.5, 1.0, 'City wise analysis of Sale, Discount, profit')



In [39]: *#2. High Profit*
`df_city['Profit'].tail(30).plot(kind='bar',figsize=(15,5),color = 'Pink')`
`plt.title('City wise analysis of Sale, Discount, profit')`

Out[39]: Text(0.5, 1.0, 'City wise analysis of Sale, Discount, profit')



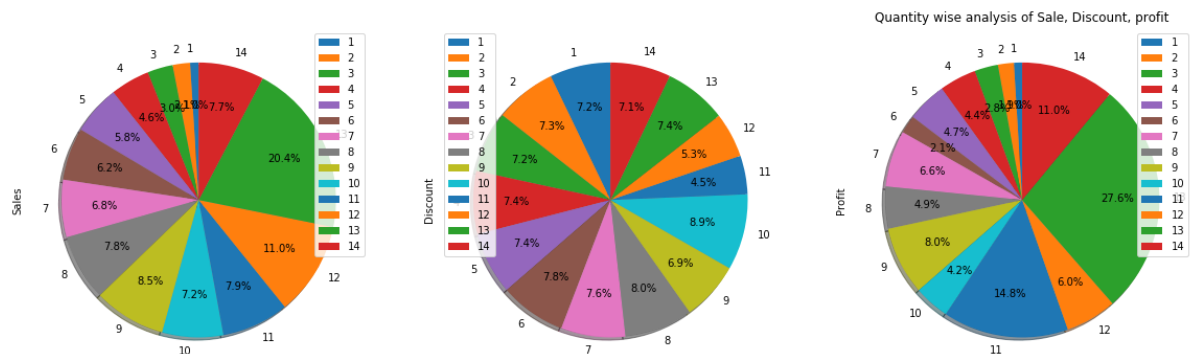
```
In [40]: df_quantity = df.groupby(['Quantity'])[['Sales', 'Discount', 'Profit']].mean()
df_quantity.head(10)
```

Out[40]:

	Sales	Discount	Profit
Quantity			
1	59.234632	0.152959	8.276396
2	120.354488	0.154858	16.006831
3	175.201578	0.153329	23.667715
4	271.764059	0.157708	37.131310
5	337.936339	0.157146	40.257394
6	362.101960	0.166556	18.051517
7	395.888393	0.161980	56.579163
8	458.210802	0.171595	42.244342
9	498.083683	0.147946	68.557716
10	422.046737	0.190702	35.862404

```
In [41]: #1. sales 2. Discount 3. Profit
df_quantity.plot.pie(subplots=True,
                    autopct='%1.1f%%',
                    figsize=(20, 20),
                    pctdistance=0.69,
                    startangle=90,      # start angle 90° (Africa)
                    shadow=True,
                    labels = df_quantity.index)
plt.title('Quantity wise analysis of Sale, Discount, profit')
```

Out[41]: Text(0.5, 1.0, 'Quantity wise analysis of Sale, Discount, profit')



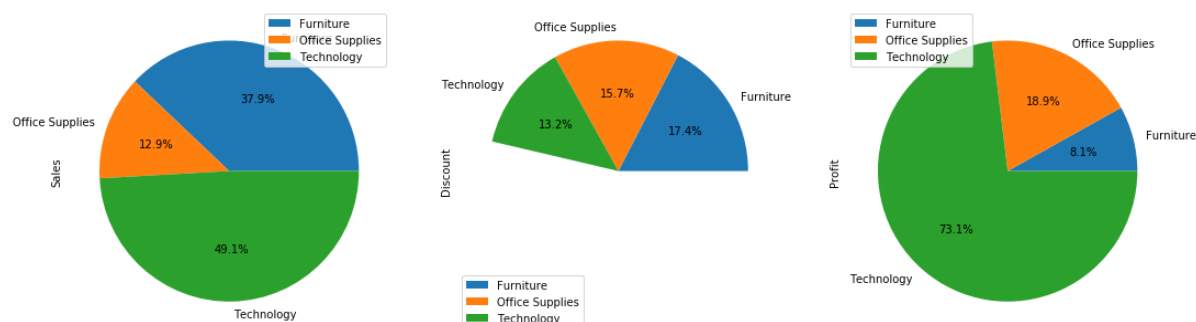
```
In [42]: df_category = df.groupby(['Category'])[['Sales', 'Discount', 'Profit']].mean()
df_category
```

Out[42]:

	Sales	Discount	Profit
Category			
Furniture	349.834887	0.173923	8.699327
Office Supplies	119.324101	0.157285	20.327050
Technology	452.709276	0.132323	78.752002

```
In [43]: df_category.plot.pie(subplots=True,
                             figsize=(18, 20),
                             autopct='%1.1f%%',
                             labels = df_category.index)
```

Out[43]: array([<matplotlib.axes._subplots.AxesSubplot object at 0x00000136A4E4F940>,
<matplotlib.axes._subplots.AxesSubplot object at 0x00000136A5746A58>,
<matplotlib.axes._subplots.AxesSubplot object at 0x00000136A574BD30>],
dtype=object)

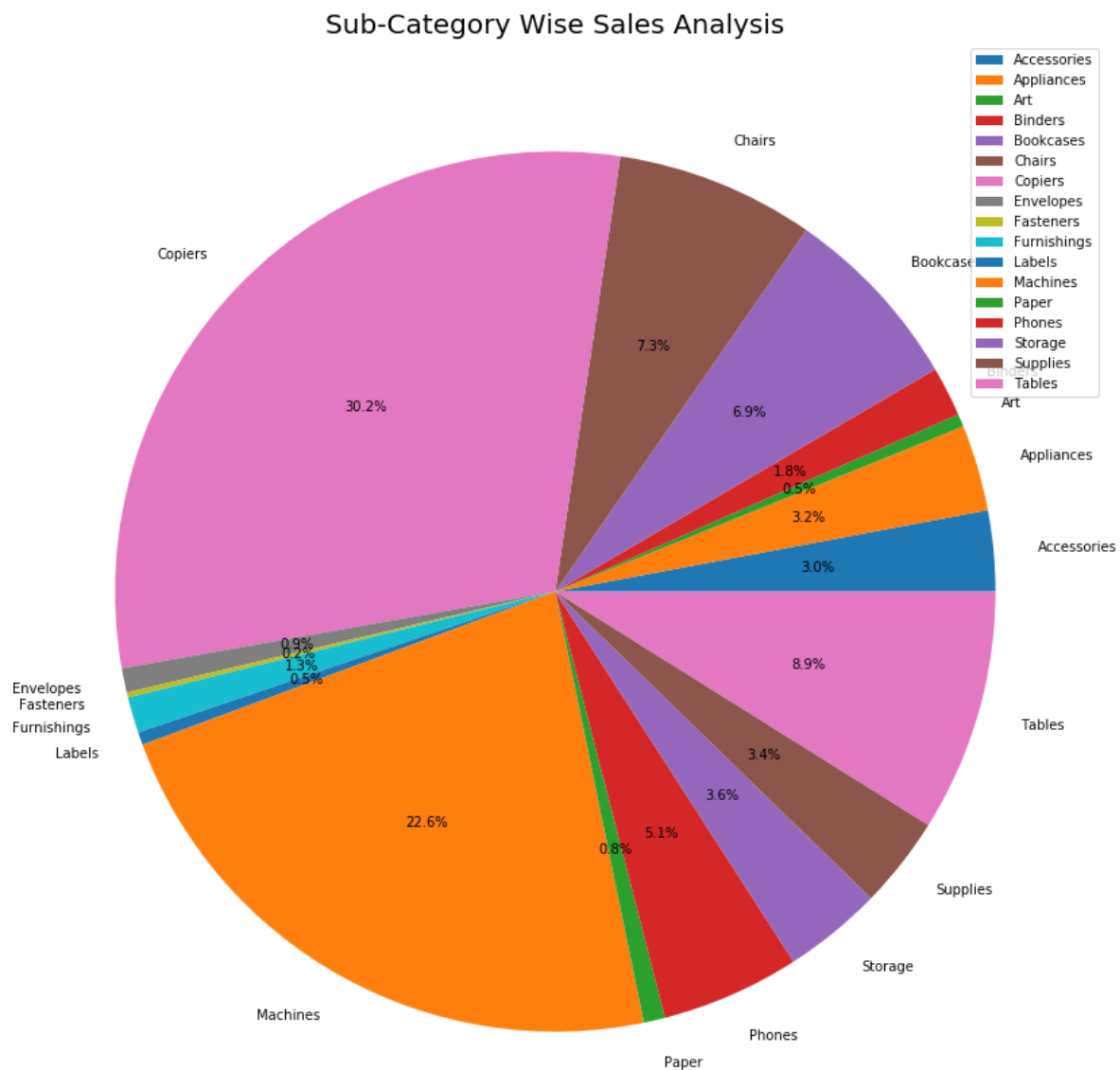


```
In [44]: df_sub_category = df.groupby(['Sub-Category'])[['Sales', 'Discount', 'Profit']]
         .mean()
         df_sub_category.head(10)
```

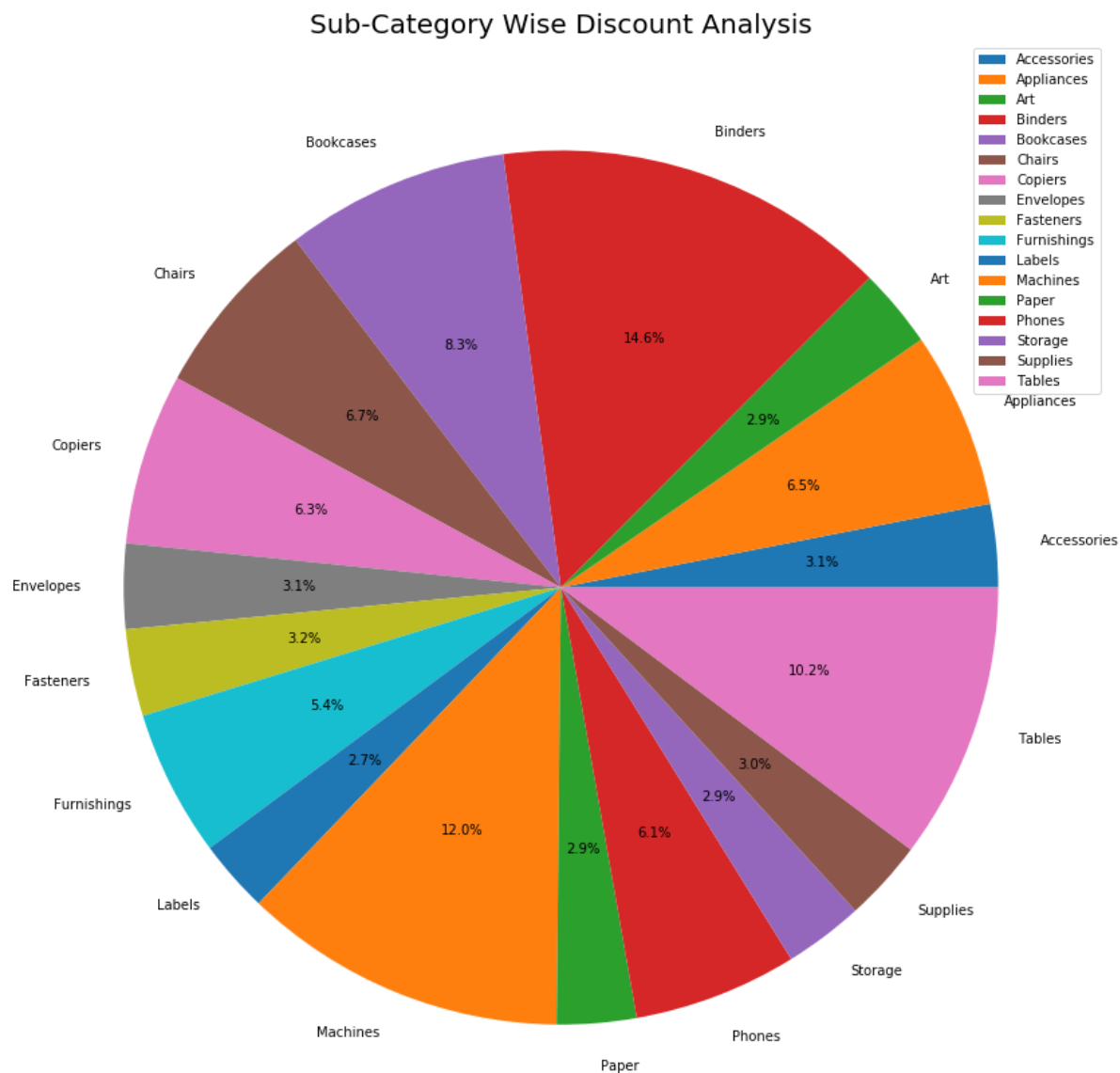
Out[44]:

	Sales	Discount	Profit
Sub-Category			
Accessories	215.974604	0.078452	54.111788
Appliances	230.755710	0.166524	38.922758
Art	34.068834	0.074874	8.200737
Binders	133.560560	0.372292	19.843574
Bookcases	503.859633	0.211140	-15.230509
Chairs	532.332420	0.170178	43.095894
Copiers	2198.941618	0.161765	817.909190
Envelopes	64.867724	0.080315	27.418019
Fasteners	13.936774	0.082028	4.375660
Furnishings	95.825668	0.138349	13.645918

```
In [45]: plt.figure(figsize = (15,15))
plt.pie(df_sub_category['Sales'], labels = df_sub_category.index, autopct = '%1.1f%%')
plt.title('Sub-Category Wise Sales Analysis', fontsize = 20)
plt.legend()
plt.xticks(rotation = 90)
plt.show()
```

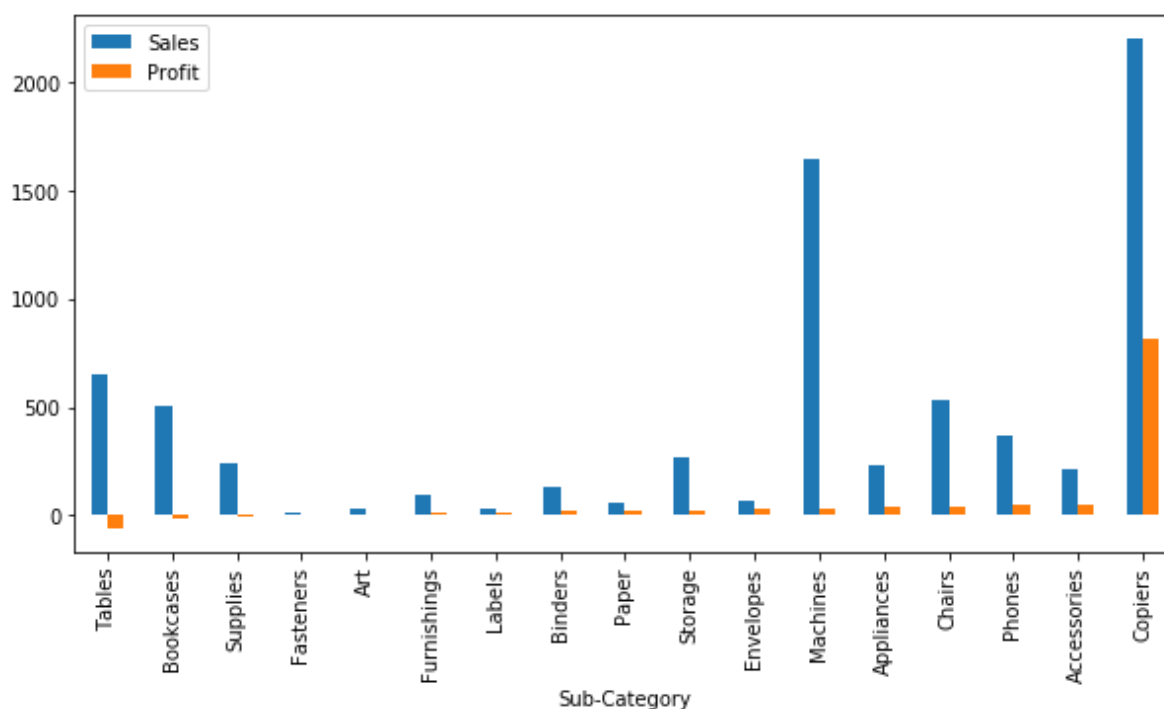


```
In [47]: plt.figure(figsize = (15,15))
plt.pie(df_sub_category['Discount'], labels = df_sub_category.index, autopct =
'%1.1f%%')
plt.title('Sub-Category Wise Discount Analysis', fontsize = 20)
plt.legend()
plt.xticks(rotation = 90)
plt.show()
```



```
In [48]: df_sub_category.sort_values('Profit')[['Sales', 'Profit']].plot(kind='bar',
                                             figsize= (10,5),
                                             label=[ 'Avg Sale
s Price($)', 'Profit($)'])
```

Out[48]: <matplotlib.axes._subplots.AxesSubplot at 0x136a5a20dd8>



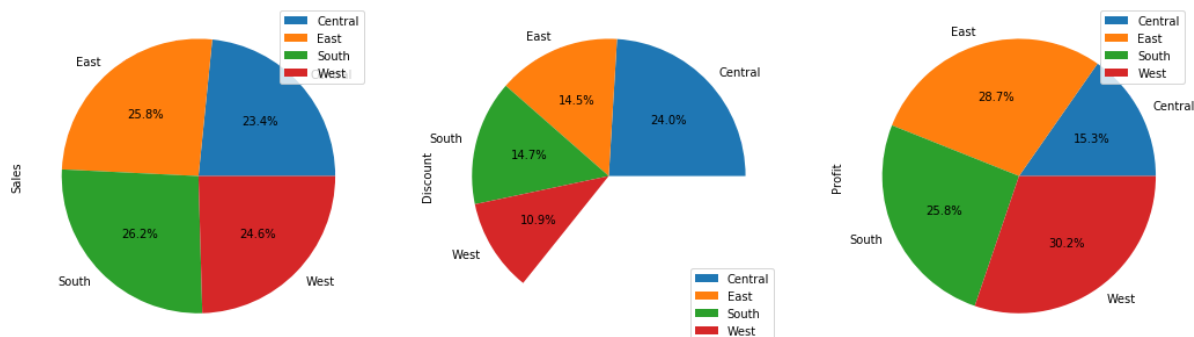
```
In [49]: df_region = df.groupby(['Region'])[['Sales', 'Discount', 'Profit']].mean()
df_region
```

Out[49]:

	Sales	Discount	Profit
Region			
Central	215.772661	0.240353	17.092709
East	238.336110	0.145365	32.135808
South	241.803645	0.147253	28.857673
West	226.493233	0.109335	33.849032


```
In [50]: df_region.plot.pie(subplots=True,
                             figsize=(18, 20),
                             autopct='%1.1f%%',
                             labels = df_region.index)
```

```
Out[50]: array([<matplotlib.axes._subplots.AxesSubplot object at 0x00000136A5CBD978>,
                  <matplotlib.axes._subplots.AxesSubplot object at 0x00000136A5CF42E8>,
                  <matplotlib.axes._subplots.AxesSubplot object at 0x00000136A5D1B668>],
              dtype=object)
```



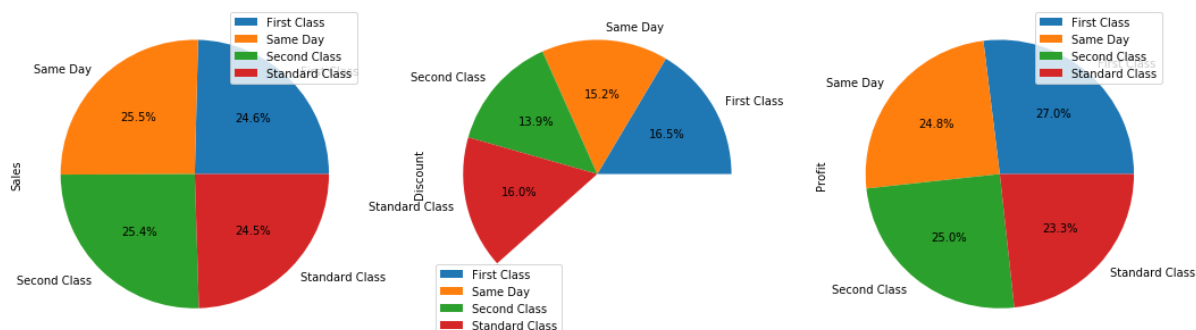
```
In [51]: df['Ship Mode'].value_counts()
```

```
Out[51]: Standard Class    5968
          Second Class     1945
          First Class      1538
          Same Day         543
          Name: Ship Mode, dtype: int64
```

```
In [52]: df_shipmode = df.groupby(['Ship Mode'])[['Sales', 'Discount', 'Profit']].mean()
```

```
In [53]: df_shipmode.plot.pie(subplots=True,
                               figsize=(18, 20),
                               autopct='%1.1f%%',
                               labels = df_shipmode.index)
```

```
Out[53]: array([<matplotlib.axes._subplots.AxesSubplot object at 0x00000136A5FCACC0>,
                  <matplotlib.axes._subplots.AxesSubplot object at 0x00000136A6006128>,
                  <matplotlib.axes._subplots.AxesSubplot object at 0x00000136A602B588>],
              dtype=object)
```



```
In [ ]:
```