

Ex-08-Data-Visualization-

' AIM

To Perform Data Visualization on a complex dataset and save the data to a file.

' Explanation

Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.

' ALGORITHM

' STEP 1

Read the given Data

' STEP 2

Clean the Data Set using Data Cleaning Process

' STEP 3

Apply Feature generation and selection techniques to all the features of the data set

' STEP 4

Apply data visualization techniques to identify the patterns of the data.

' CODE

```
import pandas as pd
import numpy as np
df=pd.read_csv("Superstore.csv")

df.head()

#Data Visualization using Seaborn

import seaborn as sns
from matplotlib import pyplot as plt
```

```
#1.Line plot

plt.figure(figsize=(8,5))
sns.lineplot(x="Segment",y="Region",data=df,marker='o')
plt.xticks(rotation = 90)
sns.lineplot(x='Ship Mode',y='Category', hue ="Segment",data=df)
sns.lineplot(x="Category",y="Sales",data=df,marker='o')
```

#2.Scatterplot

```
sns.scatterplot(x='Category',y='Sub-Category',data=df)
sns.scatterplot(x='Category', y='Sub-Category', hue ="Segment",data=df)
plt.figure(figsize=(10,7))
sns.scatterplot(x="Region",y="Sales",data=df)
plt.xticks(rotation = 90)
```

#3.Boxplot

```
sns.boxplot(x="Sub-Category",y="Discount",data=df)
sns.boxplot( x="Profit", y="Category",data=df)
```

#4.Barplot

```
sns.barplot(x="Sub-Category",y="Sales",data=df)
plt.xticks(rotation = 90)
sns.barplot(x="Category",y="Sales",data=df)
plt.xticks(rotation = 90)
```

#5.Pointplot

```
sns.pointplot(x=df[ "Quantity"],y=df[ "Discount"])
```

#6.Count plot

```
sns.countplot(x="Category",data=df)
sns.countplot(x="Sub-Category",data=df)
```

#7.Histogram

```
sns.histplot(data=df,x ='Ship Mode',hue='Sub-Category')
```

#8.KDE Plot

```
sns.kdeplot(x="Profit", data = df,hue='Category')
```

#Data Visualization Using Matplotlib

#1.Plot

```
plt.plot(df['Category'], df['Sales'])
plt.show()
```

```
#2.Heatmap
```

```
df.corr()  
plt.subplots(figsize=(12,7))  
sns.heatmap(df.corr(),annot=True)
```

```
#3.Piechart
```

```
df1=df.groupby(by=["Ship Mode"]).sum()  
labels=[]  
for i in df1.index:  
    labels.append(i)  
colors=sns.color_palette("bright")  
plt.pie(df1["Sales"],labels=labels,autopct="%0.0f%%")  
plt.show()  
  
df3=df.groupby(by=["Category"]).sum()  
labels=[]  
for i in df3.index:  
    labels.append(i)  
plt.figure(figsize=(8,8))  
colors = sns.color_palette('pastel')  
plt.pie(df3["Profit"],colors = colors,labels=labels, autopct = '%0.0f%')  
plt.show()
```

```
#4.Histogram
```

```
plt.hist(df["Sub-Category"],facecolor="peru",edgecolor="blue",bins=10)  
plt.show()
```

```
#5.Bargraph
```

```
plt.bar(df.index,df['Category'])  
plt.show()
```

```
#6.Scatterplot
```

```
plt.scatter(df["Region"],df["Profit"], c ="blue")  
plt.show()
```

```
#7.Boxplot
```

```
plt.boxplot(x="Sales",data=df)  
plt.show()
```

' OUPUT

```
In [1]: import pandas as pd
import numpy as np
df=pd.read_csv("Superstore.csv")
```



```
In [2]: df.head()
```

Out[2]:

	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	...	Postal Code	Region	Product ID	Category	Sub-Category
0	1	CA-2016-152156	11/8/2016	11/11/2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	...	42420	South	FUR-BO-10001798	Furniture	Bookcases
1	2	CA-2016-152156	11/8/2016	11/11/2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	...	42420	South	FUR-CH-10000454	Furniture	Chairs
2	3	CA-2016-138688	6/12/2016	6/16/2016	Second Class	DV-13045	Darin Van Huff	Corporate	United States	Los Angeles	...	90036	West	OFF-LA-10000240	Office Supplies	Labels
3	4	US-2015-108966	10/11/2015	10/18/2015	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	...	33311	South	FUR-TA-10000577	Furniture	Tables
4	5	US-2015-108966	10/11/2015	10/18/2015	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	...	33311	South	OFF-ST-10000760	Office Supplies	Storage

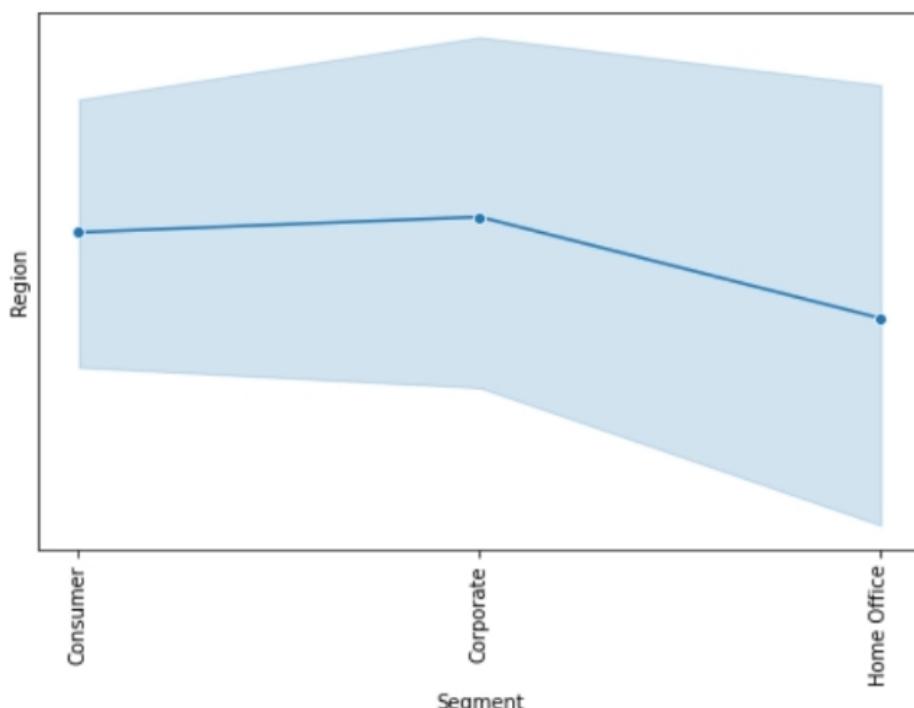
5 rows × 21 columns

```
In [7]: import seaborn as sns
from matplotlib import pyplot as plt
```

```
In [8]: #1. Line plot
```

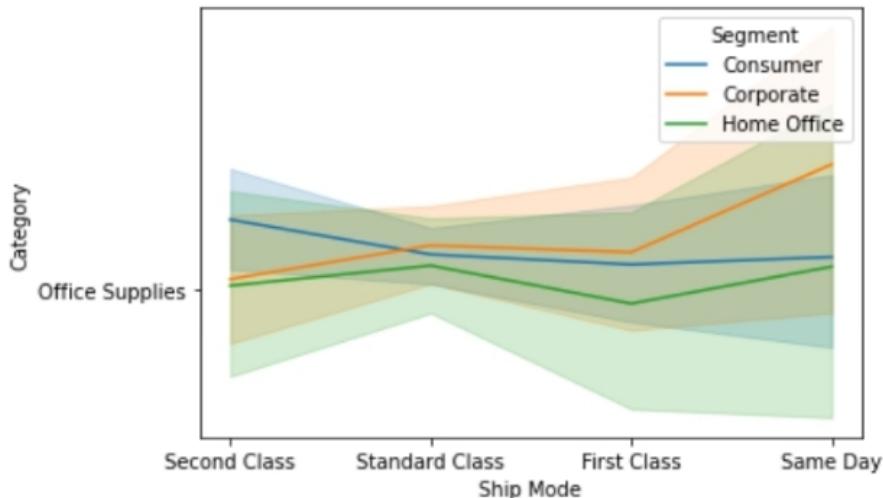
```
plt.figure(figsize=(8,5))
sns.lineplot(x="Segment",y="Region",data=df,marker='o')
plt.xticks(rotation = 90)
```

```
Out[8]: ([0, 1, 2], [Text(0, 0, ''), Text(0, 0, ''), Text(0, 0, '')])
```



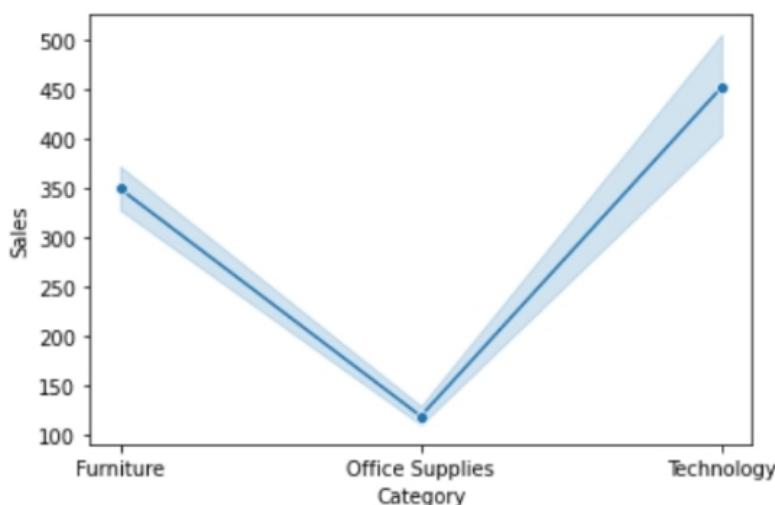
```
In [9]: sns.lineplot(x='Ship Mode',y='Category', hue ="Segment",data=df)
```

```
Out[9]: <AxesSubplot:xlabel='Ship Mode', ylabel='Category'>
```



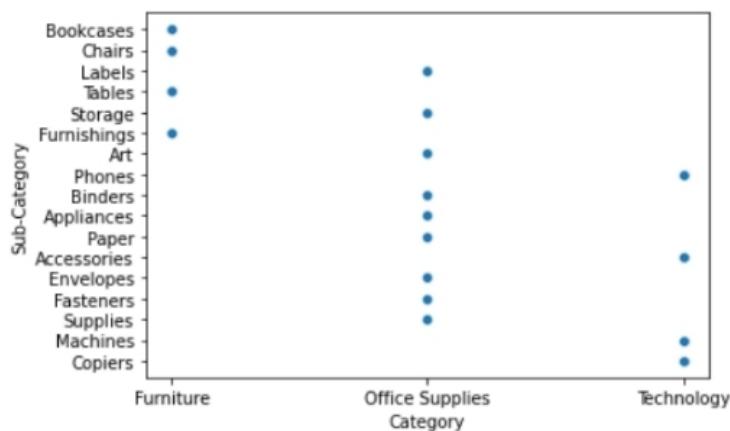
```
In [10]: sns.lineplot(x="Category",y="Sales",data=df,marker='o')
```

```
Out[10]: <AxesSubplot:xlabel='Category', ylabel='Sales'>
```



```
In [11]: #2.Scatterplot
sns.scatterplot(x='Category',y='Sub-Category',data=df)
```

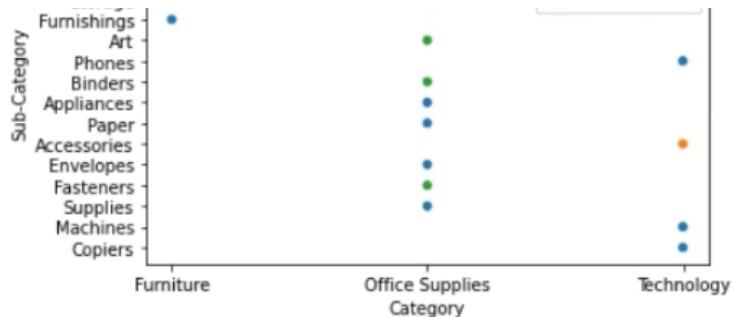
```
Out[11]: <AxesSubplot:xlabel='Category', ylabel='Sub-Category'>
```



```
In [12]: sns.scatterplot(x='Category', y='Sub-Category', hue = "Segment",data=df)
```

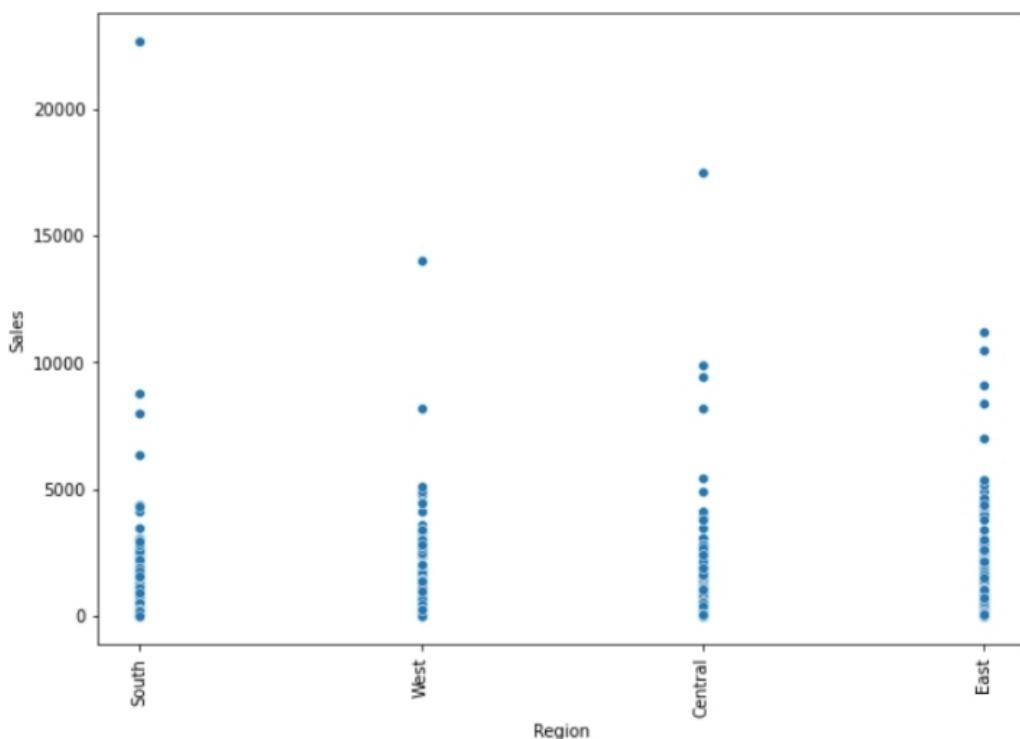
```
Out[12]: <AxesSubplot:xlabel='Category', ylabel='Sub-Category'>
```





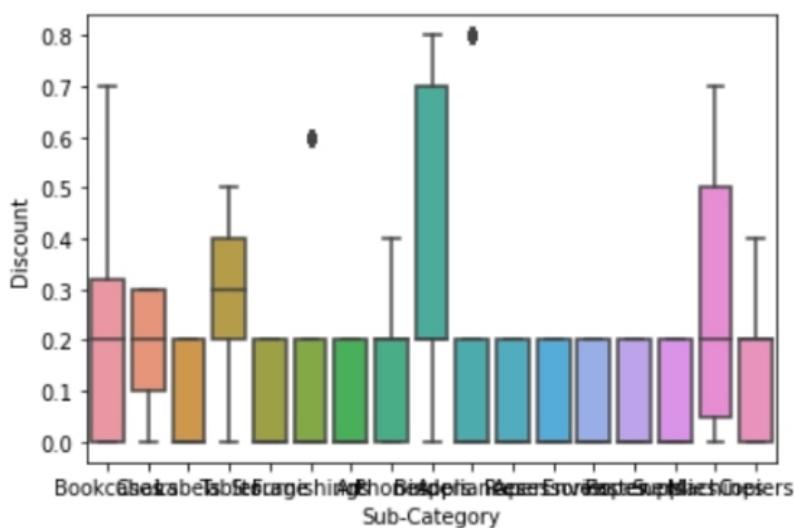
```
In [13]: plt.figure(figsize=(10,7))
sns.scatterplot(x="Region",y="Sales",data=df)
plt.xticks(rotation = 90)
```

```
Out[13]: ([0, 1, 2, 3], [Text(0, 0, ''), Text(0, 0, ''), Text(0, 0, ''), Text(0, 0, '')])
```



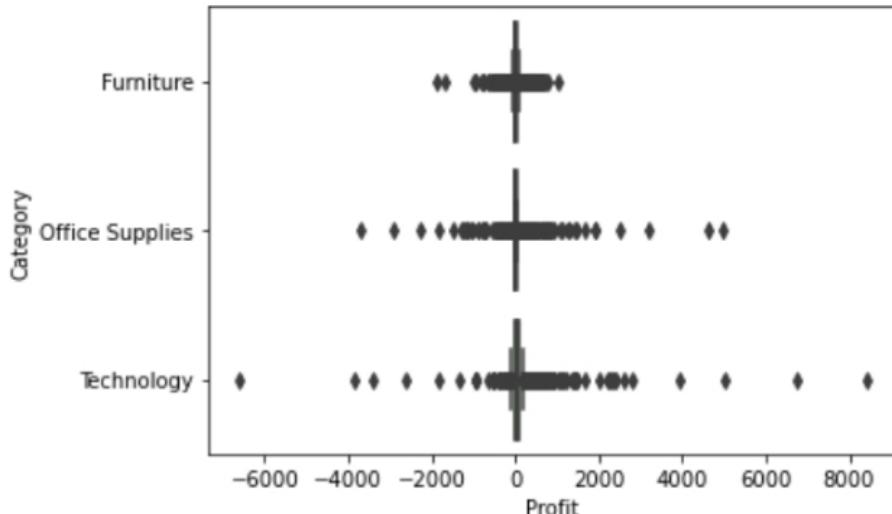
```
In [14]: #3.Boxplot
sns.boxplot(x="Sub-Category",y="Discount",data=df)
```

```
Out[14]: <AxesSubplot:xlabel='Sub-Category', ylabel='Discount'>
```



```
In [15]: sns.boxplot( x="Profit", y="Category",data=df)
```

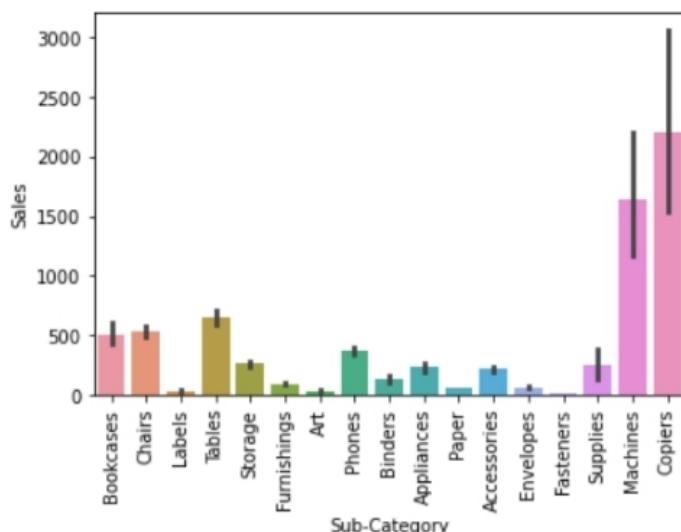
```
Out[15]: <AxesSubplot:xlabel='Profit', ylabel='Category'>
```



```
In [16]: #4.Barplot
```

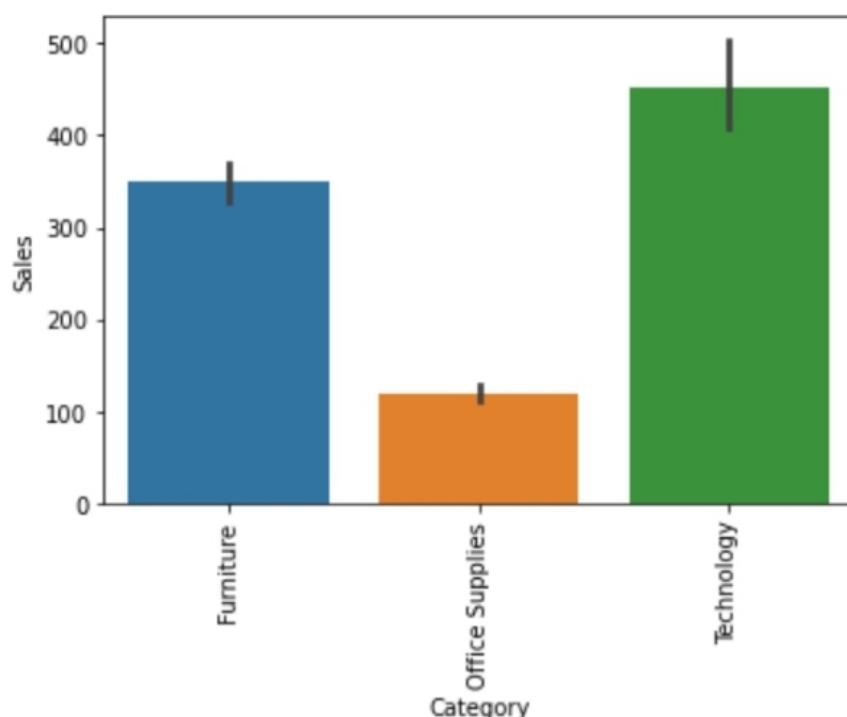
```
sns.barplot(x="Sub-Category",y="Sales",data=df)
plt.xticks(rotation = 90)
```

```
Out[16]: (array([ 0,  1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11, 12, 13, 14, 15, 16]),
[Text(0, 0, 'Bookcases'),
Text(1, 0, 'Chairs'),
Text(2, 0, 'Labels'),
Text(3, 0, 'Tables'),
Text(4, 0, 'Storage'),
Text(5, 0, 'Furnishings'),
Text(6, 0, 'Art'),
Text(7, 0, 'Phones'),
Text(8, 0, 'Binders'),
Text(9, 0, 'Appliances'),
Text(10, 0, 'Paper'),
Text(11, 0, 'Accessories'),
Text(12, 0, 'Envelopes'),
Text(13, 0, 'Fasteners'),
Text(14, 0, 'Supplies'),
Text(15, 0, 'Machines'),
Text(16, 0, 'Copiers')])
```



```
In [17]: sns.barplot(x="Category",y="Sales",data=df)
plt.xticks(rotation = 90)
```

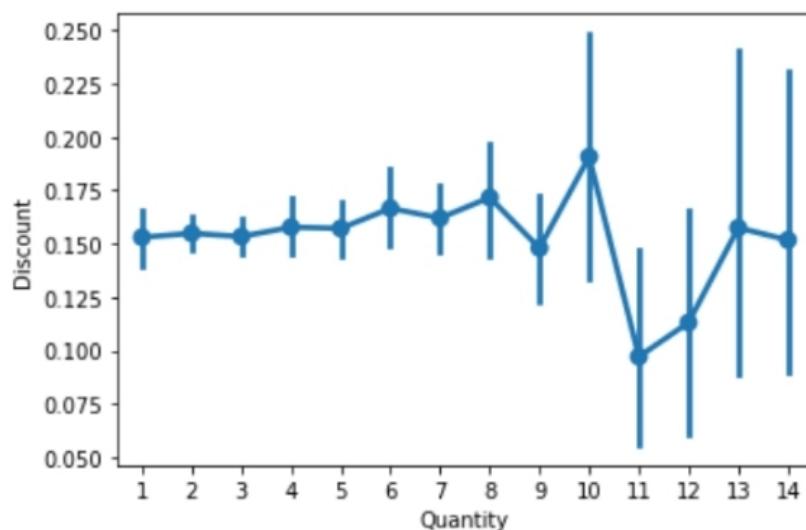
```
Out[17]: (array([0, 1, 2]),
 [Text(0, 0, 'Furniture'),
 Text(1, 0, 'Office Supplies'),
 Text(2, 0, 'Technology')])
```



```
In [18]: #5.Pointplot
```

```
sns.pointplot(x=df["Quantity"],y=df["Discount"])
```

```
Out[18]: <AxesSubplot:xlabel='Quantity', ylabel='Discount'>
```

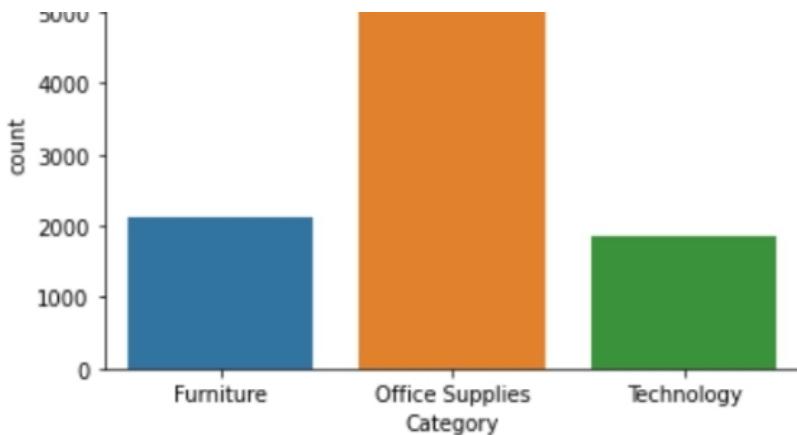


```
In [19]: #6.Count plot
```

```
sns.countplot(x="Category",data=df)
```

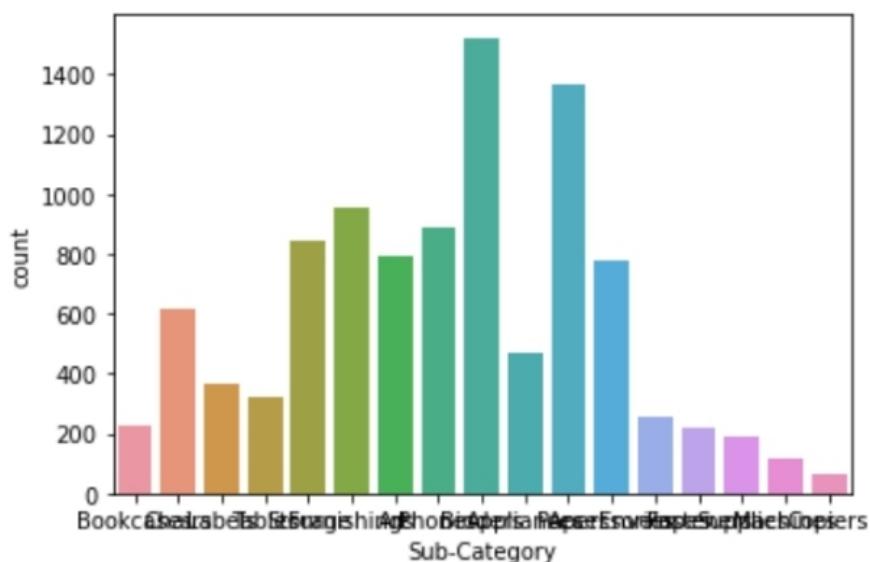
```
Out[19]: <AxesSubplot:xlabel='Category', ylabel='count'>
```





```
In [20]: sns.countplot(x="Sub-Category",data=df)
```

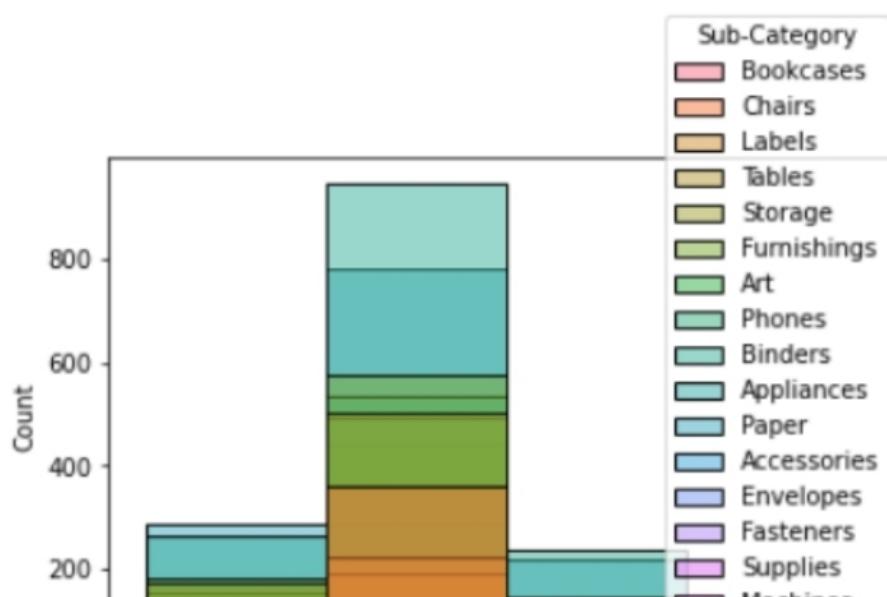
```
Out[20]: <AxesSubplot:xlabel='Sub-Category', ylabel='count'>
```



```
In [11]: #7.Histogram
```

```
sns.histplot(data=df,x ='Ship Mode',hue='Sub-Category')
```

```
Out[11]: <AxesSubplot:xlabel='Ship Mode', ylabel='Count'>
```

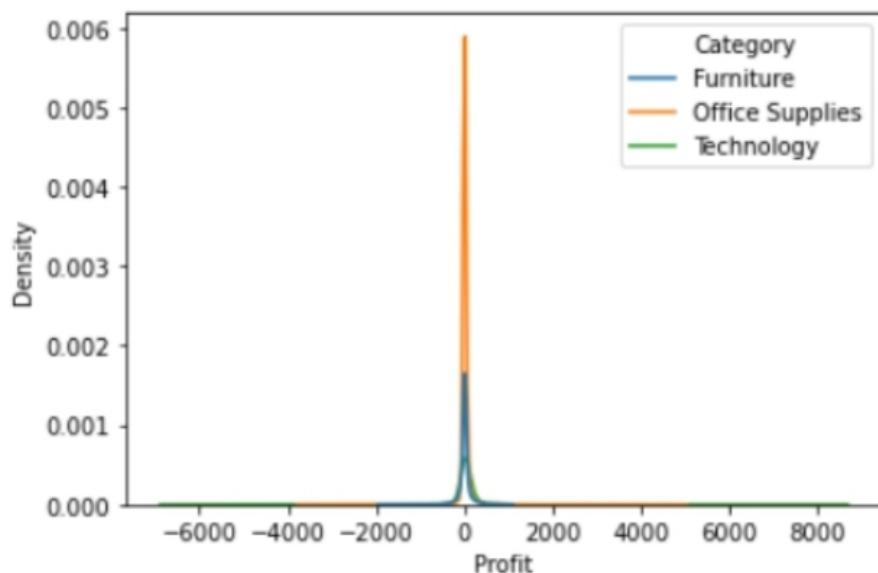




In [12]: #8.KDE Plot

```
sns.kdeplot(x="Profit", data = df,hue='Category')
```

Out[12]: <AxesSubplot:xlabel='Profit', ylabel='Density'>



Data Visualization Using Matplotlib:

In [13]: #1.Plot

```
plt.plot(df['Category'], df['Sales'])  
plt.show()
```





In [14]: #2.Heatmap

```
df.corr()
plt.subplots(figsize=(12,7))
sns.heatmap(df.corr(), annot=True)
```

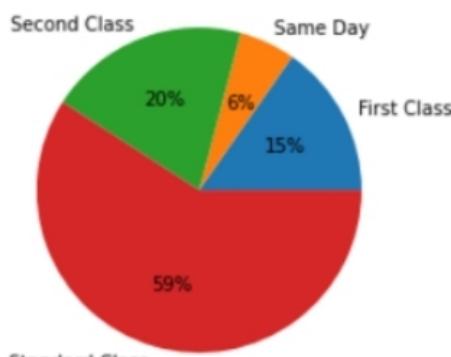
Out[14]: <AxesSubplot:>

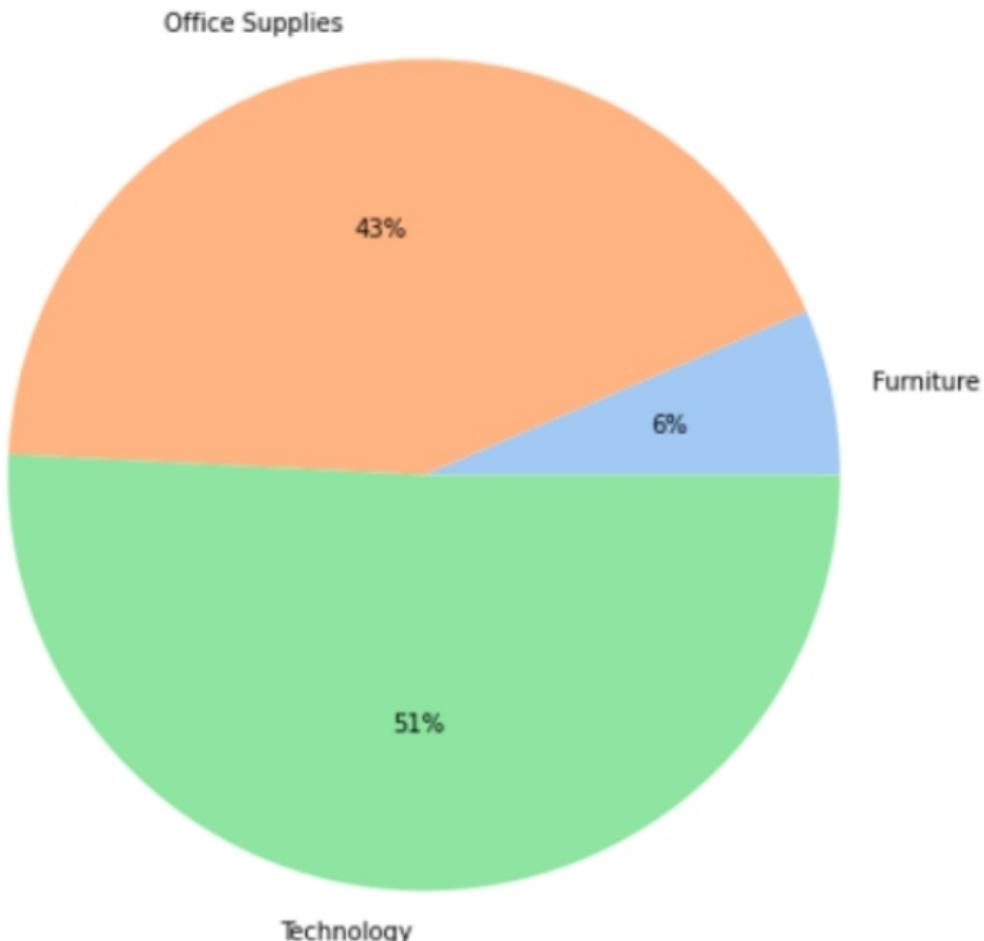


In [15]: #3.Piechart

```
df1=df.groupby(by=["Ship Mode"]).sum()
labels=[]
for i in df1.index:
    labels.append(i)
colors=sns.color_palette("bright")
plt.pie(df1["Sales"],labels=labels,autopct="%0.0f%%")
plt.show()

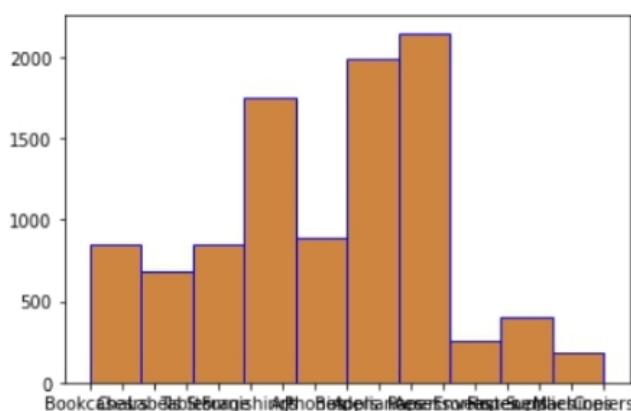
df3=df.groupby(by=["Category"]).sum()
labels=[]
for i in df3.index:
    labels.append(i)
plt.figure(figsize=(8,8))
colors = sns.color_palette('pastel')
plt.pie(df3["Profit"],colors = colors,labels=labels, autopct = '%0.0f%%')
plt.show()
```





```
In [16]: #4.Histogram
```

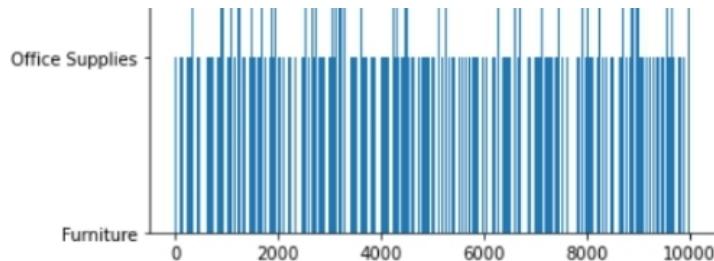
```
plt.hist(df["Sub-Category"], facecolor="peru", edgecolor="blue", bins=10)  
plt.show()
```



```
In [18]: #5.Bargraph
```

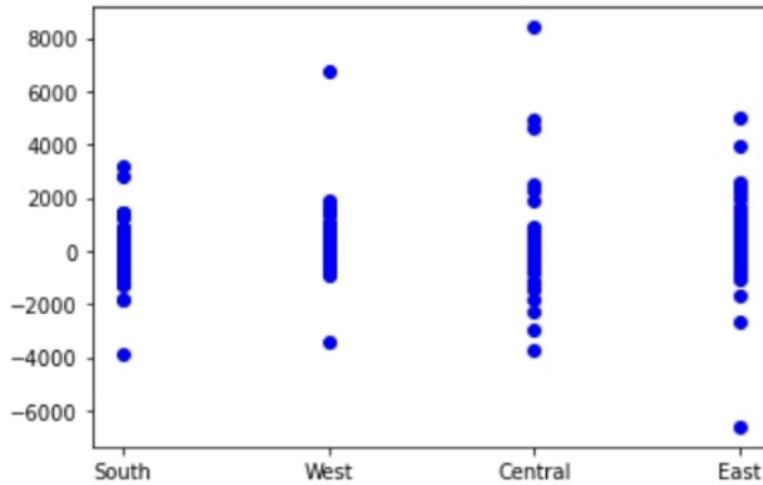
```
plt.bar(df.index, df['Category'])  
plt.show()
```





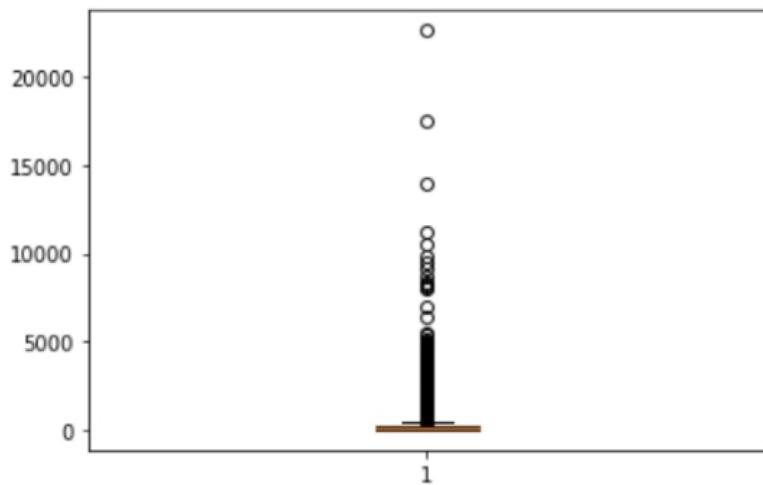
```
In [19]: #6.Scatterplot
```

```
plt.scatter(df["Region"],df["Profit"], c ="blue")
plt.show()
```



```
In [20]: #7.Boxplot
```

```
plt.boxplot(x="Sales",data=df)
plt.show()
```



RESULT

Hence, Data Visualization is applied on the complex dataset using libraries like Seaborn and Matplotlib successfully and the data is saved to file.