```
import pandas as pd

import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np

dataset = pd.read_csv("C:/Users/VICTUS/Downloads/New
folder/Superstore.csv", encoding='latin1')

dataset.head(2)
```

```
   Row ID        Order ID  Order Date   Ship Date     Ship Mode
Customer ID  \
0       1  CA-2016-152156  08-11-2016  11-11-2016  Second Class    CG-
12520
1       2  CA-2016-152156  08-11-2016  11-11-2016  Second Class    CG-
12520

  Customer Name   Segment        Country       City  ... Postal Code
Region  \
0   Claire Gute  Consumer  United States  Henderson  ...       42420
South
1   Claire Gute  Consumer  United States  Henderson  ...       42420
South

        Product ID   Category Sub-Category  \
0  FUR-BO-10001798  Furniture    Bookcases
1  FUR-CH-10000454  Furniture       Chairs

                                          Product Name   Sales  Quantity
\
0                  Bush Somerset Collection Bookcase  261.96         2

1  Hon Deluxe Fabric Upholstered Stacking Chairs,...  731.94         3


   Discount    Profit
0       0.0   41.9136
1       0.0  219.5820

[2 rows x 21 columns]
```

```
dataset.shape
```

```
(9994, 21)
```

```
dataset.isnull().sum()
```

```
Row ID           0
Order ID         0
Order Date       0
Ship Date        0
```

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

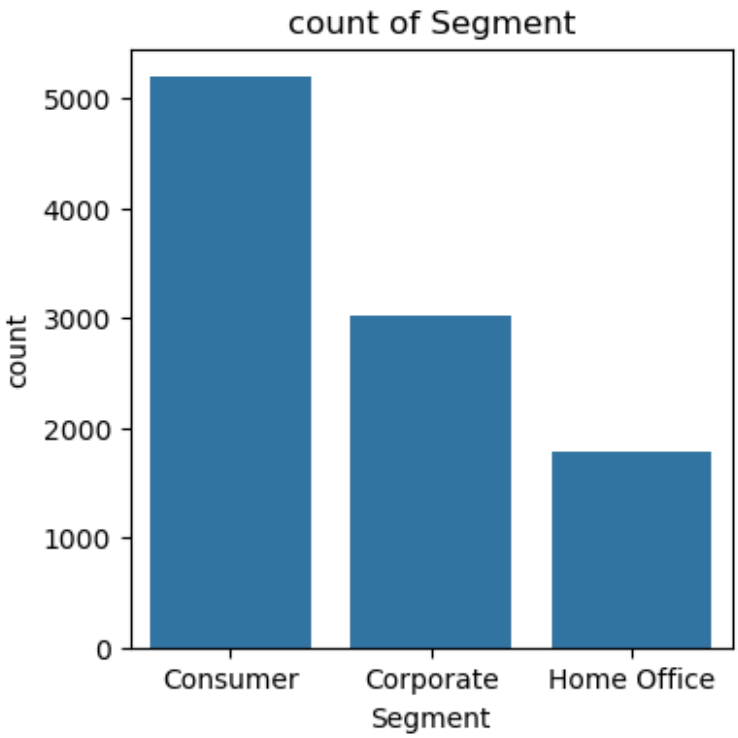
```
dataset['Segment'].value_counts()
```

```
Segment
Consumer       5191
Corporate      3020
Home Office    1783
Name: count, dtype: int64
```

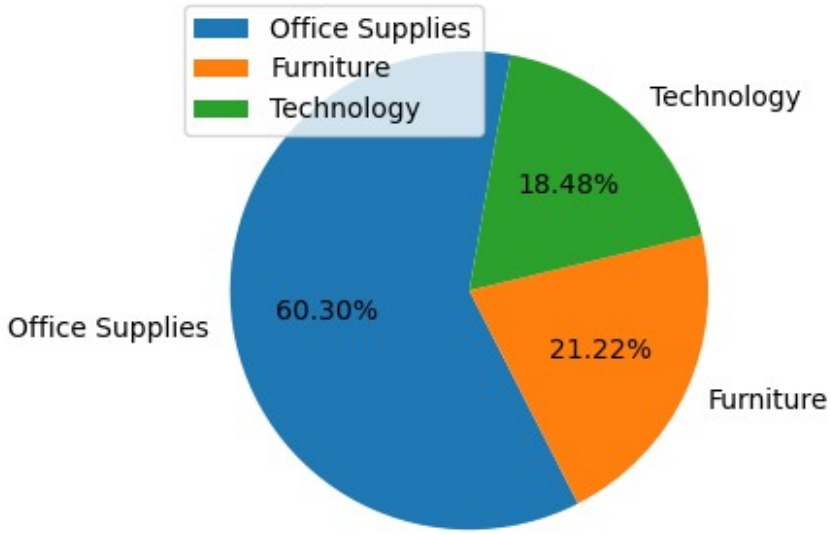
```
dataset['Segment'].unique()
```

```
array(['Consumer', 'Corporate', 'Home Office'], dtype=object)
```

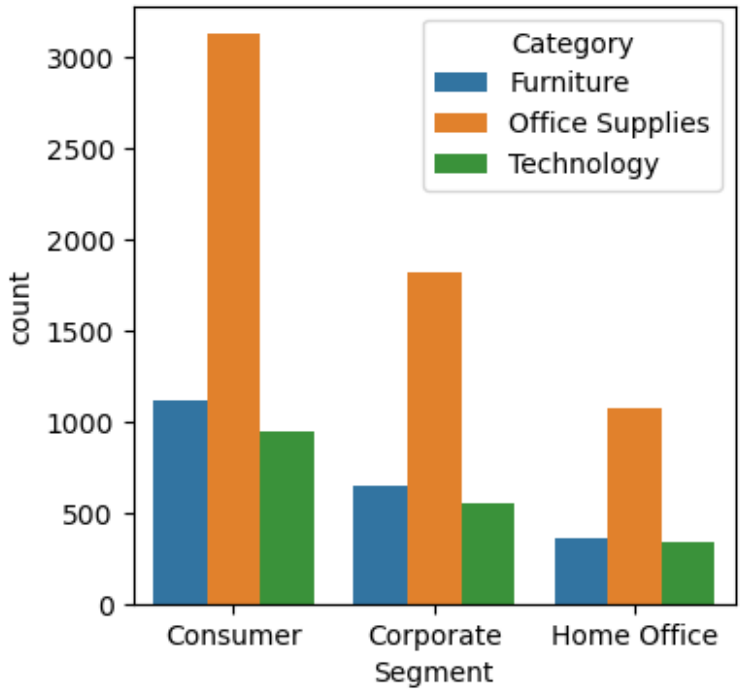
```
plt.figure(figsize=(4,4))
sns.countplot(x='Segment', data=dataset)
plt.title("count of Segment")
plt.savefig("count of Segment.pdf")
plt.show()
```



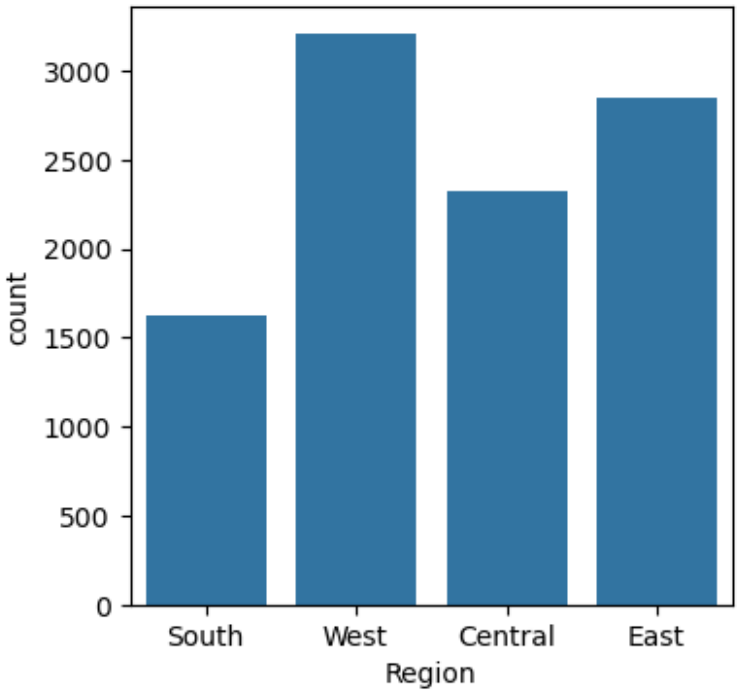

```
dataset['Category'].value_counts()

Category
Office Supplies    6026
Furniture          2121
Technology         1847
Name: count, dtype: int64

x= dataset['Category'].value_counts().index
y= dataset['Category'].value_counts().values

plt.figure(figsize=(4,4))
plt.pie(y,labels=x,startangle = 80,autopct= "%0.2f%%")
plt.legend(loc=0)
plt.show()
```



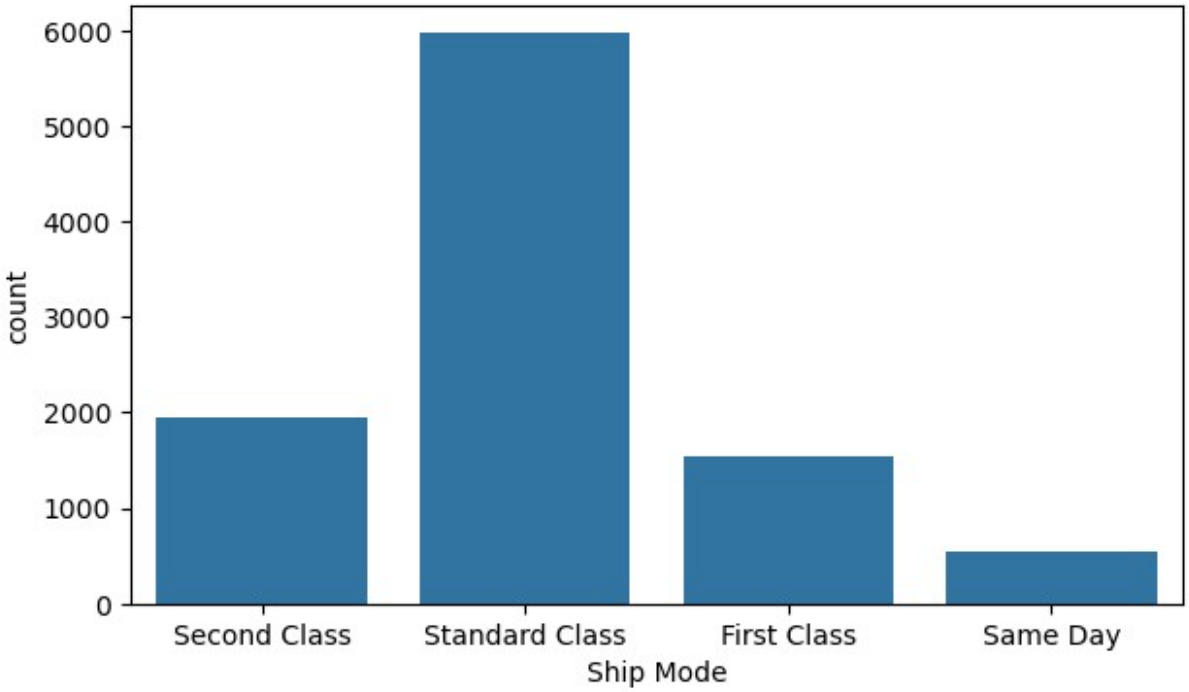

```
plt.figure(figsize=(4,4))
sns.countplot(x="Segment",data=dataset,hue="Category")
plt.show()
```




```
plt.figure(figsize=(4,4))
sns.countplot(x="Region",data=dataset)
plt.show()
```




```
plt.figure(figsize=(7,4))
sns.countplot(x="Ship Mode",data=dataset)
plt.show()
```




```
dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9994 entries, 0 to 9993
Data columns (total 21 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   Row ID         9994 non-null   int64
 1   Order ID       9994 non-null   object
 2   Order Date     9994 non-null   object
 3   Ship Date      9994 non-null   object
 4   Ship Mode      9994 non-null   object
 5   Customer ID    9994 non-null   object
 6   Customer Name  9994 non-null   object
 7   Segment        9994 non-null   object
 8   Country        9994 non-null   object
 9   City           9994 non-null   object
 10  State          9994 non-null   object
 11  Postal Code    9994 non-null   int64
 12  Region         9994 non-null   object
 13  Product ID     9994 non-null   object
 14  Category       9994 non-null   object
 15  Sub-Category   9994 non-null   object
 16  Product Name   9994 non-null   object
 17  Sales          9994 non-null   float64
 18  Quantity       9994 non-null   int64
 19  Discount       9994 non-null   float64
 20  Profit         9994 non-null   float64
dtypes: float64(3), int64(3), object(15)
memory usage: 1.6+ MB
```

```
dataset['Order Date'] = pd.to_datetime(dataset['Order Date'],
format='%d-%m-%Y')

dataset['Ship Date'] = pd.to_datetime(dataset['Ship Date'],
format='%d-%m-%Y')

dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9994 entries, 0 to 9993
Data columns (total 21 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   Row ID         9994 non-null   int64
 1   Order ID       9994 non-null   object
 2   Order Date     9994 non-null   datetime64[ns]
 3   Ship Date      9994 non-null   datetime64[ns]
 4   Ship Mode      9994 non-null   object
 5   Customer ID    9994 non-null   object
 6   Customer Name  9994 non-null   object
 7   Segment        9994 non-null   object
```

```
 8   Country        9994 non-null   object
 9   City           9994 non-null   object
 10  State          9994 non-null   object
 11  Postal Code    9994 non-null   int64
 12  Region         9994 non-null   object
 13  Product ID     9994 non-null   object
 14  Category       9994 non-null   object
 15  Sub-Category   9994 non-null   object
 16  Product Name   9994 non-null   object
 17  Sales          9994 non-null   float64
 18  Quantity       9994 non-null   int64
 19  Discount       9994 non-null   float64
 20  Profit         9994 non-null   float64
dtypes: datetime64[ns](2), float64(3), int64(3), object(13)
memory usage: 1.6+ MB
```

```
dataset["Order Year"]= dataset['Order Date'].dt.year
```

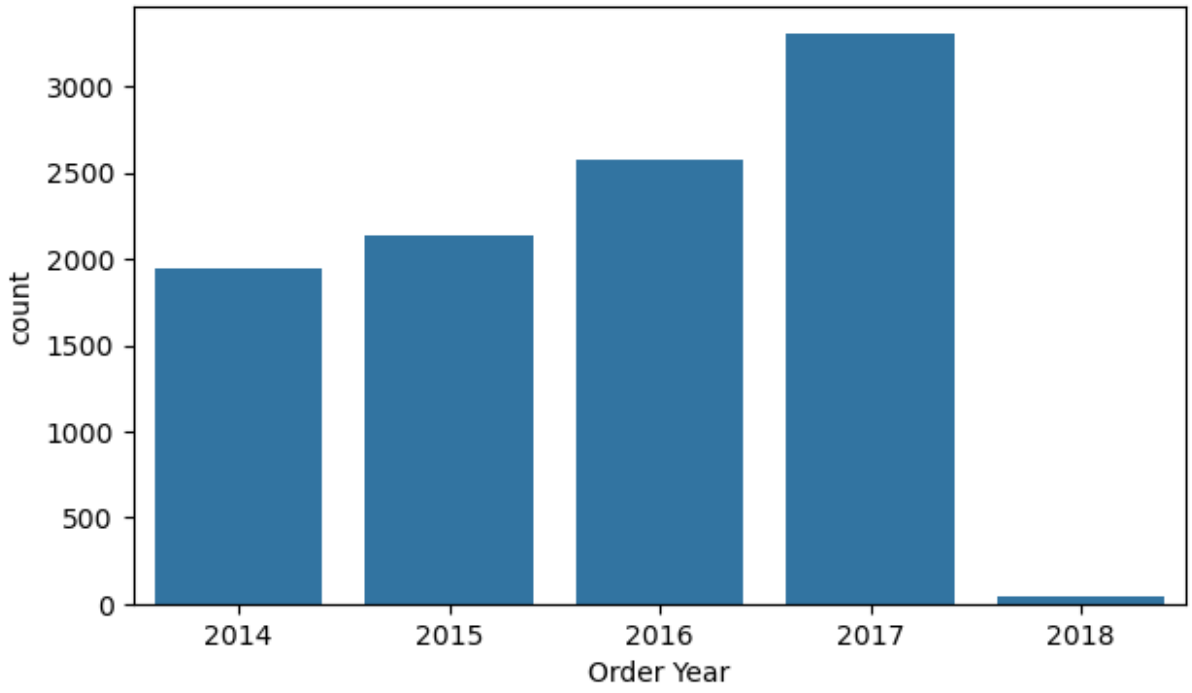
```
dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9994 entries, 0 to 9993
Data columns (total 22 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   Row ID         9994 non-null   int64
 1   Order ID       9994 non-null   object
 2   Order Date     9994 non-null   datetime64[ns]
 3   Ship Date      9994 non-null   datetime64[ns]
 4   Ship Mode      9994 non-null   object
 5   Customer ID    9994 non-null   object
 6   Customer Name  9994 non-null   object
 7   Segment        9994 non-null   object
 8   Country        9994 non-null   object
 9   City           9994 non-null   object
 10  State          9994 non-null   object
 11  Postal Code    9994 non-null   int64
 12  Region         9994 non-null   object
 13  Product ID     9994 non-null   object
 14  Category       9994 non-null   object
 15  Sub-Category   9994 non-null   object
 16  Product Name   9994 non-null   object
 17  Sales          9994 non-null   float64
 18  Quantity       9994 non-null   int64
 19  Discount       9994 non-null   float64
 20  Profit         9994 non-null   float64
 21  Order Year     9994 non-null   int32
dtypes: datetime64[ns](2), float64(3), int32(1), int64(3), object(13)
memory usage: 1.6+ MB
```

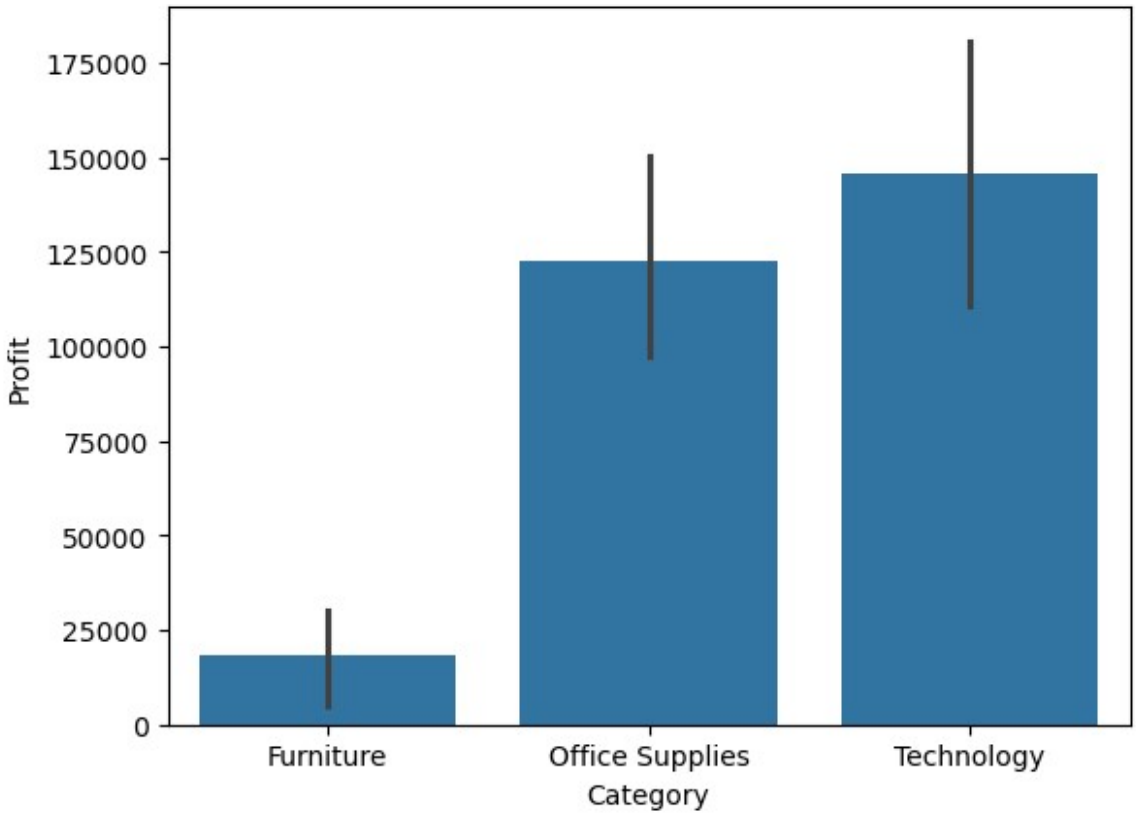
```
dataset['Order Year'].value_counts()
```

```
Order Year
2017    3303
2016    2578
2015    2131
2014    1940
2018      42
Name: count, dtype: int64
```

```
plt.figure(figsize=(7,4))
sns.countplot(x="Order Year",data=dataset)
plt.show()
```




```
sns.barplot(x="Category", y="Profit", data=dataset, estimator='sum')
plt.show()
```




```
dataset['State'].value_counts()[:5]
```

```
State
California      2001
New York        1128
Texas            985
Pennsylvania     587
Washington       506
Name: count, dtype: int64
```