In [1]:
```
!pip install numpy
```

```
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: numpy in c:\programdata\anaconda3\lib\site-packages (1.24.3)
```

In [3]:
```
!pip install pandas
```

```
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: pandas in c:\programdata\anaconda3\lib\site-packages (2.0.3)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\programdata\anaconda3\lib\site-packages (from pandas) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in c:\programdata\anaconda3\lib\site-packages (from pandas) (2023.3.post1)
Requirement already satisfied: tzdata>=2022.1 in c:\programdata\anaconda3\lib\site-packages (from pandas) (2023.3)
Requirement already satisfied: numpy>=1.21.0 in c:\programdata\anaconda3\lib\site-packages (from pandas) (1.24.3)
Requirement already satisfied: six>=1.5 in c:\programdata\anaconda3\lib\site-packages (from python-dateutil>=2.8.2->pandas) (1.16.0)
```

In [7]:
```
import pandas as pd
```

In [8]:
```
dataset = pd.read_excel(r'C:\Users\lenovo\Desktop\P1-SuperStoreUS-2015.xlsx')
```

In [10]:
```
dataset.head(2)
```

Out[10]:

	Row ID	Order Priority	Discount	Unit Price	Shipping Cost	Customer ID	Customer Name	Ship Mode	Customer Segment	Product Category	...	Region	State or Province	City	Postal Code	Order Date	Ship Date	Profit	Quantity ordered new	Sales	Order ID
0	20847	High	0.01	2.84	0.93	3	Bonnie Potter	Express Air	Corporate	Office Supplies	...	West	Washington	Anacortes	98221	2015-01-07	2015-01-08	4.5600	4	13.01	88522
1	20228	Not Specified	0.02	500.98	26.00	5	Ronnie Proctor	Delivery Truck	Home Office	Furniture	...	West	California	San Gabriel	91776	2015-06-13	2015-06-15	4390.3665	12	6362.85	90193

2 rows × 25 columns

In [9]:
```
dataset.shape
```

Out[9]:
```
(1952, 25)
```

In [10]:
```
dataset.isnull().sum()
```

Out[10]:
```
Row ID                   0
Order Priority           0
Discount                 0
Unit Price               0
Shipping Cost            0
Customer ID              0
Customer Name            0
Ship Mode                0
Customer Segment         0
Product Category         0
Product Sub-Category     0
Product Container        0
Product Name             0
Product Base Margin     16
Country                  0
Region                   0
State or Province        0
City                     0
Postal Code              0
Order Date               0
Ship Date                0
Profit                   0
Quantity ordered new     0
Sales                    0
Order ID                 0
dtype: int64
```

Order Priority

In [17]:
```
dataset['Product Base Margin'].fillna(dataset['Product Base Margin'].mean(),inplace=True)
```

In [23]:
```
dataset['Order Priority'].value_counts()
```

Out[23]:
```
Order Priority
Low              398
Not Specified    396
High             391
Critical         391
Medium           376
Name: count, dtype: int64
```

In [20]:
```
dataset['Order Priority'].unique()
```

Out[20]:
```
array(['High', 'Not Specified', 'Critical', 'Medium', 'Low', 'Critical '],
      dtype=object)
```

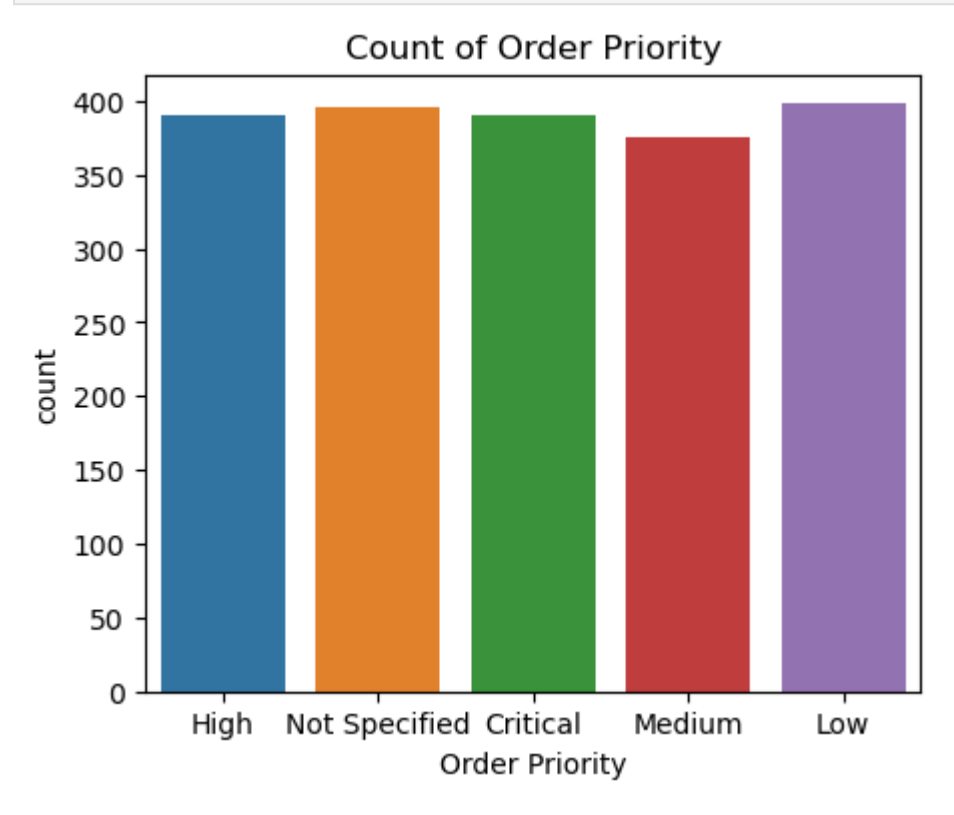
In [22]:
```
dataset['Order Priority'] = dataset['Order Priority'].replace('Critical ', 'Critical')
```

In [27]:
```
!pip install seaborn
!pip install matplotlib
```

```
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: seaborn in c:\programdata\anaconda3\lib\site-packages (0.12.2)
Requirement already satisfied: numpy!=1.24.0,>=1.17 in c:\programdata\anaconda3\lib\site-packages (from seaborn) (1.24.3)
Requirement already satisfied: pandas>=0.25 in c:\programdata\anaconda3\lib\site-packages (from seaborn) (2.0.3)
Requirement already satisfied: matplotlib!=3.6.1,>=3.1 in c:\programdata\anaconda3\lib\site-packages (from seaborn) (3.7.2)
Requirement already satisfied: contourpy>=1.0.1 in c:\programdata\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (1.0.5)
Requirement already satisfied: cycler>=0.10 in c:\programdata\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in c:\programdata\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (4.25.0)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\programdata\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (1.4.4)
Requirement already satisfied: packaging>=20.0 in c:\programdata\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (23.1)
Requirement already satisfied: pillow>=6.2.0 in c:\programdata\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (10.2.0)
Requirement already satisfied: pyparsing<3.1,>=2.3.1 in c:\programdata\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (3.0.9)
Requirement already satisfied: python-dateutil>=2.7 in c:\programdata\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in c:\programdata\anaconda3\lib\site-packages (from pandas>=0.25->seaborn) (2023.3.post1)
Requirement already satisfied: tzdata>=2022.1 in c:\programdata\anaconda3\lib\site-packages (from pandas>=0.25->seaborn) (2023.3)
Requirement already satisfied: six>=1.5 in c:\programdata\anaconda3\lib\site-packages (from python-dateutil>=2.7->matplotlib!=3.6.1,>=3.1->seaborn) (1.16.0)
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: matplotlib in c:\programdata\anaconda3\lib\site-packages (3.7.2)
Requirement already satisfied: contourpy>=1.0.1 in c:\programdata\anaconda3\lib\site-packages (from matplotlib) (1.0.5)
Requirement already satisfied: cycler>=0.10 in c:\programdata\anaconda3\lib\site-packages (from matplotlib) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in c:\programdata\anaconda3\lib\site-packages (from matplotlib) (4.25.0)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\programdata\anaconda3\lib\site-packages (from matplotlib) (1.4.4)
Requirement already satisfied: numpy>=1.20 in c:\programdata\anaconda3\lib\site-packages (from matplotlib) (1.24.3)
Requirement already satisfied: packaging>=20.0 in c:\programdata\anaconda3\lib\site-packages (from matplotlib) (23.1)
Requirement already satisfied: pillow>=6.2.0 in c:\programdata\anaconda3\lib\site-packages (from matplotlib) (10.2.0)
Requirement already satisfied: pyparsing<3.1,>=2.3.1 in c:\programdata\anaconda3\lib\site-packages (from matplotlib) (3.0.9)
Requirement already satisfied: python-dateutil>=2.7 in c:\programdata\anaconda3\lib\site-packages (from matplotlib) (2.8.2)
Requirement already satisfied: six>=1.5 in c:\programdata\anaconda3\lib\site-packages (from python-dateutil>=2.7->matplotlib) (1.16.0)
```

In [2]:
```
import seaborn as sns
import matplotlib.pyplot as plt
```

In [43]:
```
plt.figure(figsize=(5, 4))
sns.countplot(x='Order Priority', data=dataset)
plt.title("Count of Order Priority")
plt.savefig(r"C:\Users\lenovo\Desktop\Count of Order Priority.jpg")
plt.show()
```



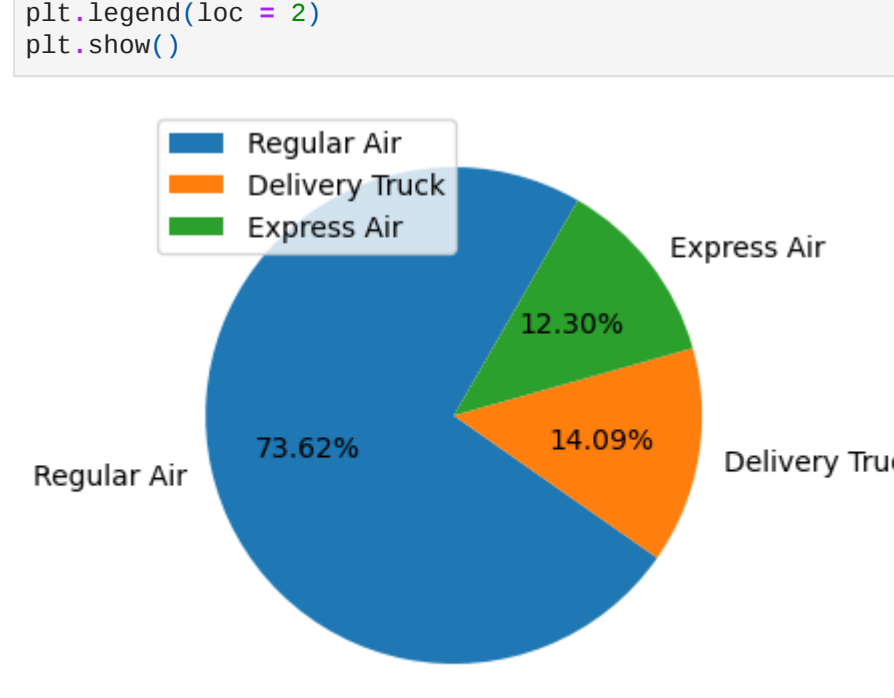
Ship Mode

In [43]:
```
dataset['Ship Mode'].value_counts()
```

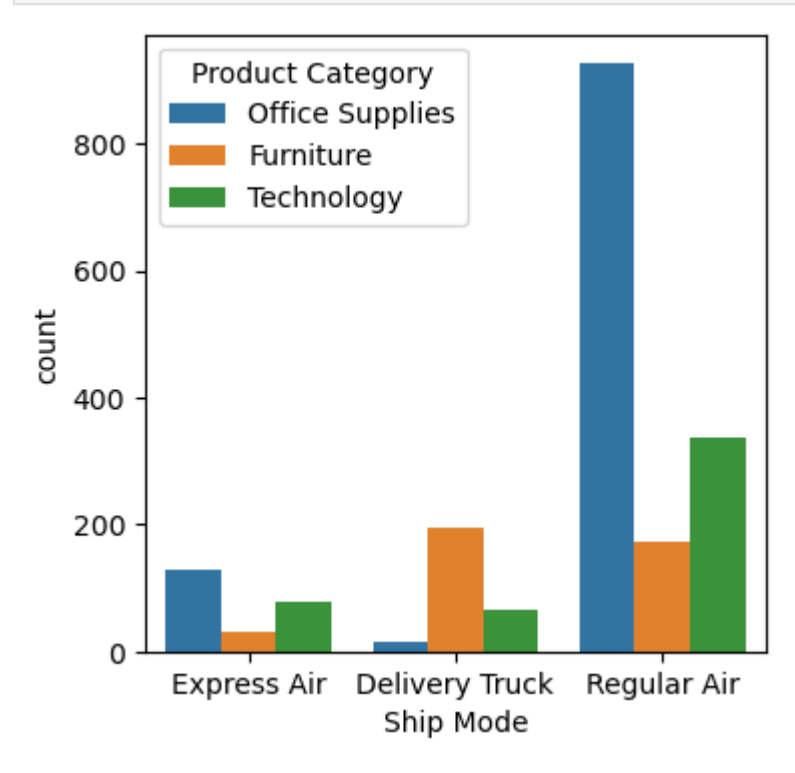
Out[43]:
```
Ship Mode
Regular Air       1437
Delivery Truck     275
Express Air        240
Name: count, dtype: int64
```

In [45]:
```
x = dataset['Ship Mode'].value_counts().index
y = dataset['Ship Mode'].value_counts().values
```

In [46]:
```
plt.figure(figsize = (4, 5))
plt.pie(y, labels=x, startangle = 60, autopct='%0.2f%%')
plt.savefig(r"C:\Users\lenovo\Desktop\Ship Mode.jpg")
plt.legend(loc = 2)
plt.show()
```



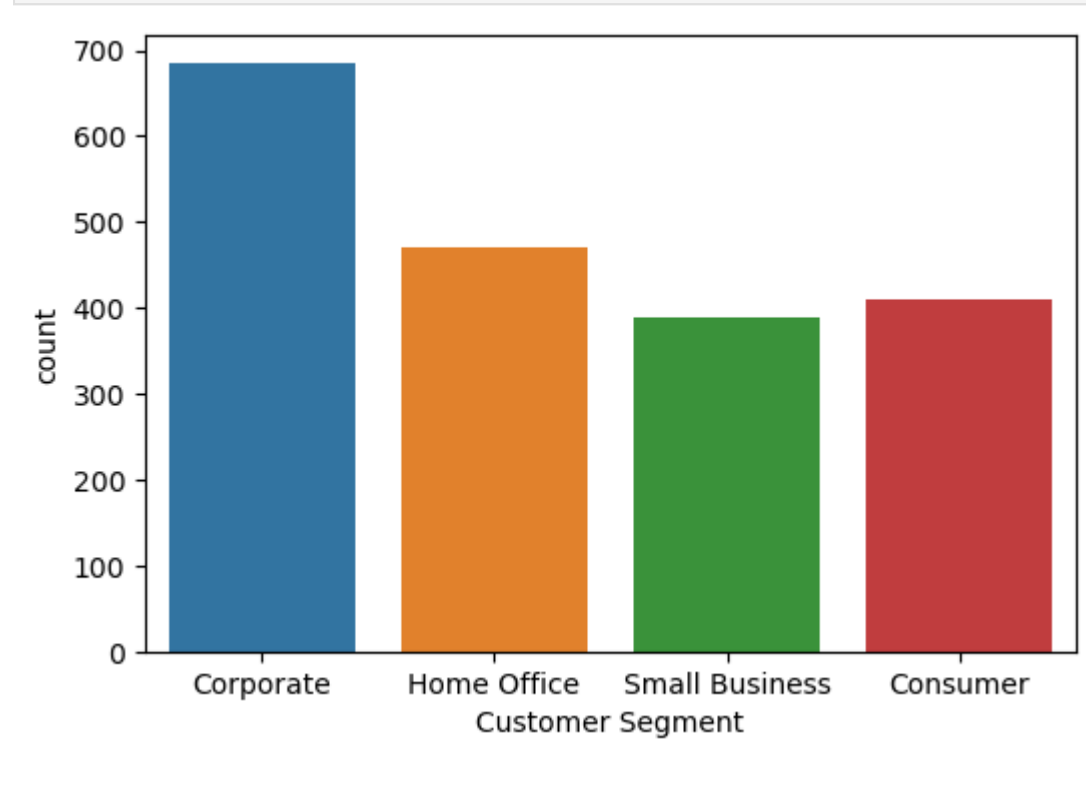
In [14]:
```
plt.figure(figsize = (4, 4))
sns.countplot(x="Ship Mode", data = dataset, hue="Product Category")
plt.show()
```



Customer Segment

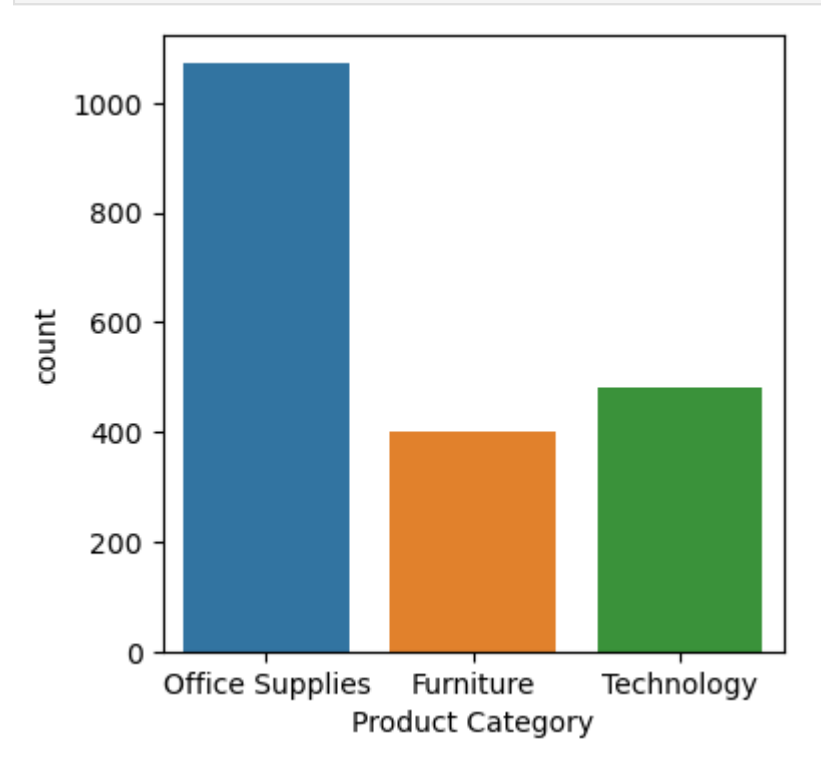
Customer Segment

In [47]:
```
plt.figure(figsize = (6,4))
sns.countplot(x="Customer Segment", data = dataset)
plt.savefig(r"C:\Users\lenovo\Desktop\Customer Segment.jpg")
plt.show()
```

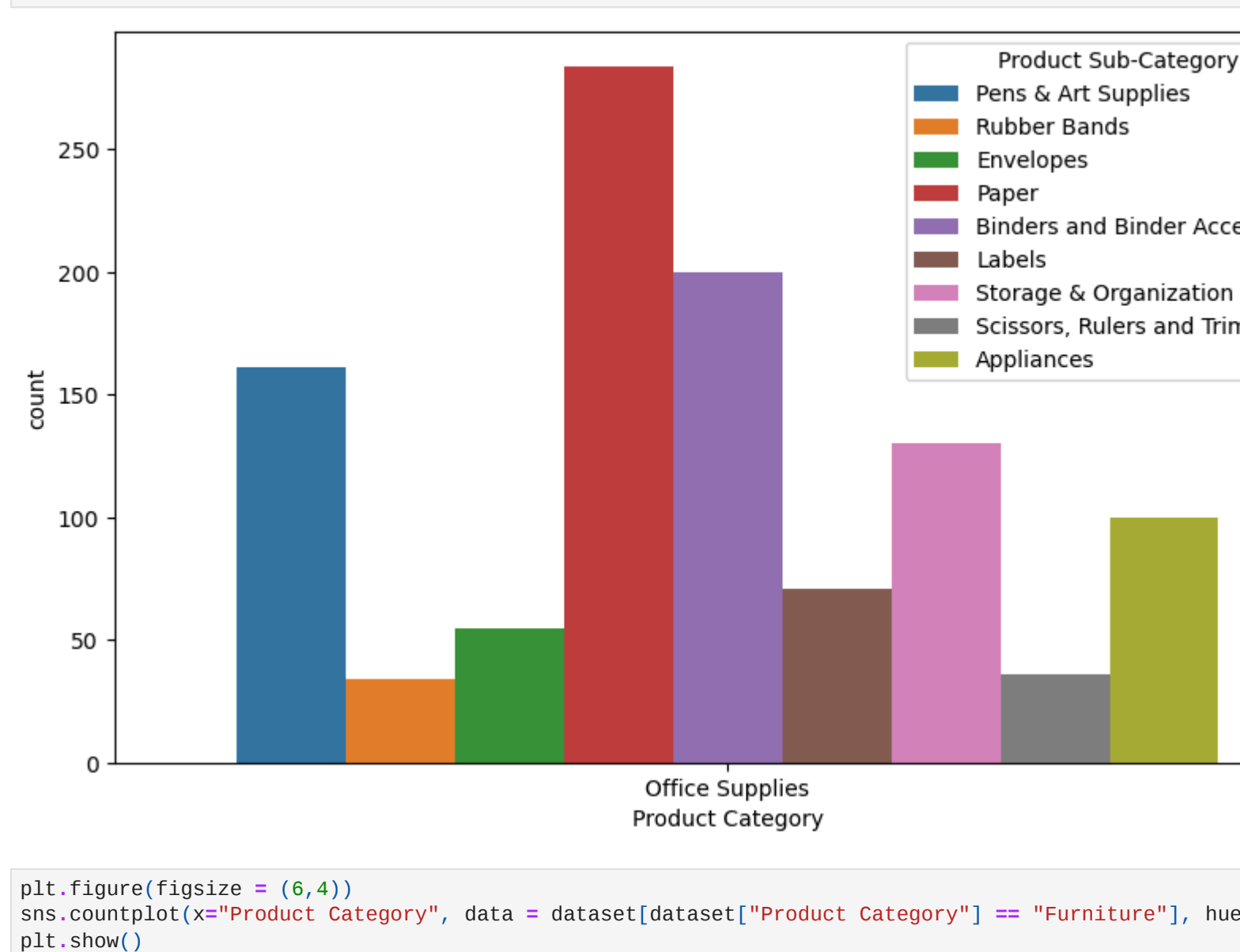


Product Category

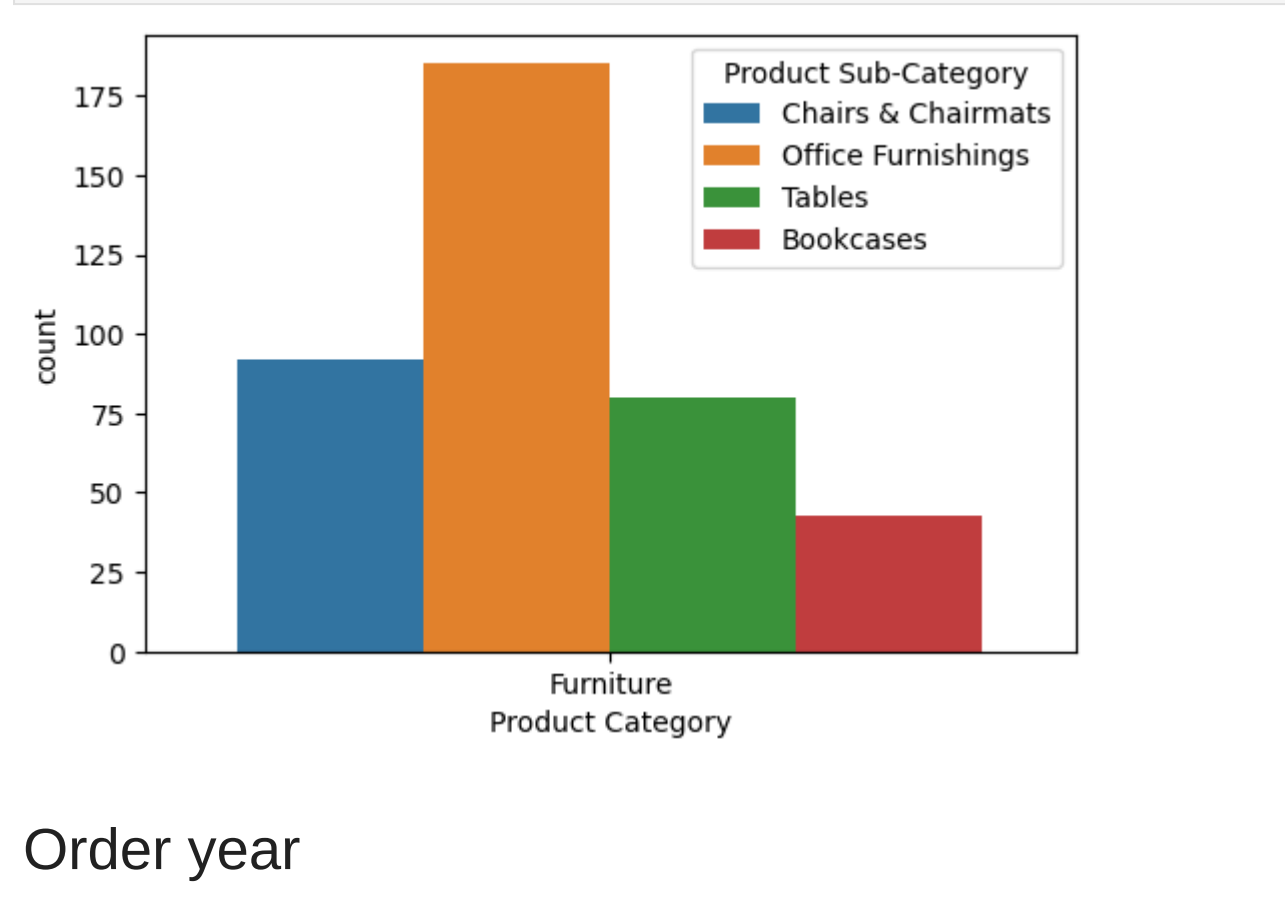
In [48]:
```
plt.figure(figsize = (4,4))
sns.countplot(x="Product Category", data = dataset)
plt.savefig(r"C:\Users\lenovo\Desktop\Product Category.jpg")
plt.show()
```



In [49]:
```
plt.figure(figsize = (10,6))
sns.countplot(x="Product Category", data = dataset[dataset["Product Category"] == "Office Supplies"], hue = "Product Sub-Category")
plt.savefig(r"C:\Users\lenovo\Desktop\Office Supplies.jpg")
plt.show()
```



In [23]:
```
plt.figure(figsize = (6,4))
sns.countplot(x="Product Category", data = dataset[dataset["Product Category"] == "Furniture"], hue = "Product Sub-Category")
plt.show()
```



Order year

In [26]:
```
dataset["Order year"] = dataset['Order Date'].dt.year
```

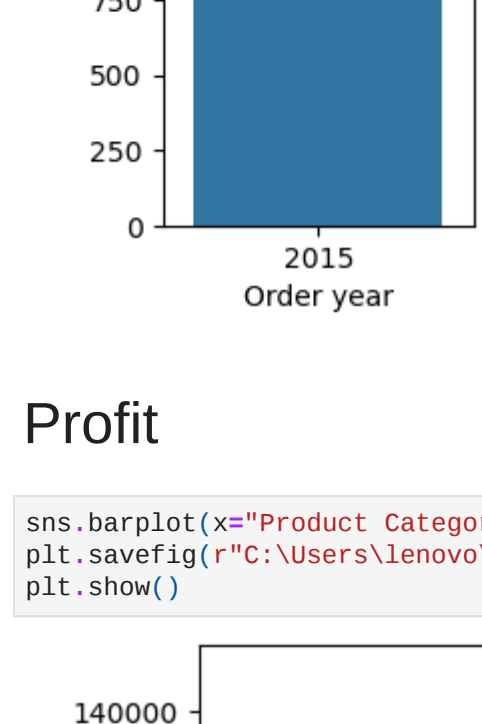
In [27]:
```
dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1952 entries, 0 to 1951
Data columns (total 26 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   Row ID                1952 non-null   int64
 1   Order Priority        1952 non-null   object
 2   Discount              1952 non-null   float64
 3   Unit Price            1952 non-null   float64
 4   Shipping Cost         1952 non-null   float64
 5   Customer ID           1952 non-null   int64
 6   Customer Name         1952 non-null   object
 7   Ship Mode             1952 non-null   object
 8   Customer Segment      1952 non-null   object
 9   Product Category      1952 non-null   object
 10  Product Sub-Category  1952 non-null   object
 11  Product Container     1952 non-null   object
 12  Product Name          1952 non-null   object
 13  Product Base Margin   1936 non-null   float64
 14  Country               1952 non-null   object
 15  Region                1952 non-null   object
 16  State or Province     1952 non-null   object
 17  City                  1952 non-null   object
 18  Postal Code           1952 non-null   int64
 19  Order Date            1952 non-null   datetime64[ns]
 20  Ship Date             1952 non-null   datetime64[ns]
 21  Profit                1952 non-null   float64
 22  Quantity ordered new  1952 non-null   int64
 23  Sales                 1952 non-null   float64
 24  Order ID              1952 non-null   int64
 25  Order year            1952 non-null   int32
dtypes: datetime64[ns](2), float64(6), int32(1), int64(5), object(12)
memory usage: 389.0+ KB
```

In [28]:
```
dataset['Order year'].value_counts()
```

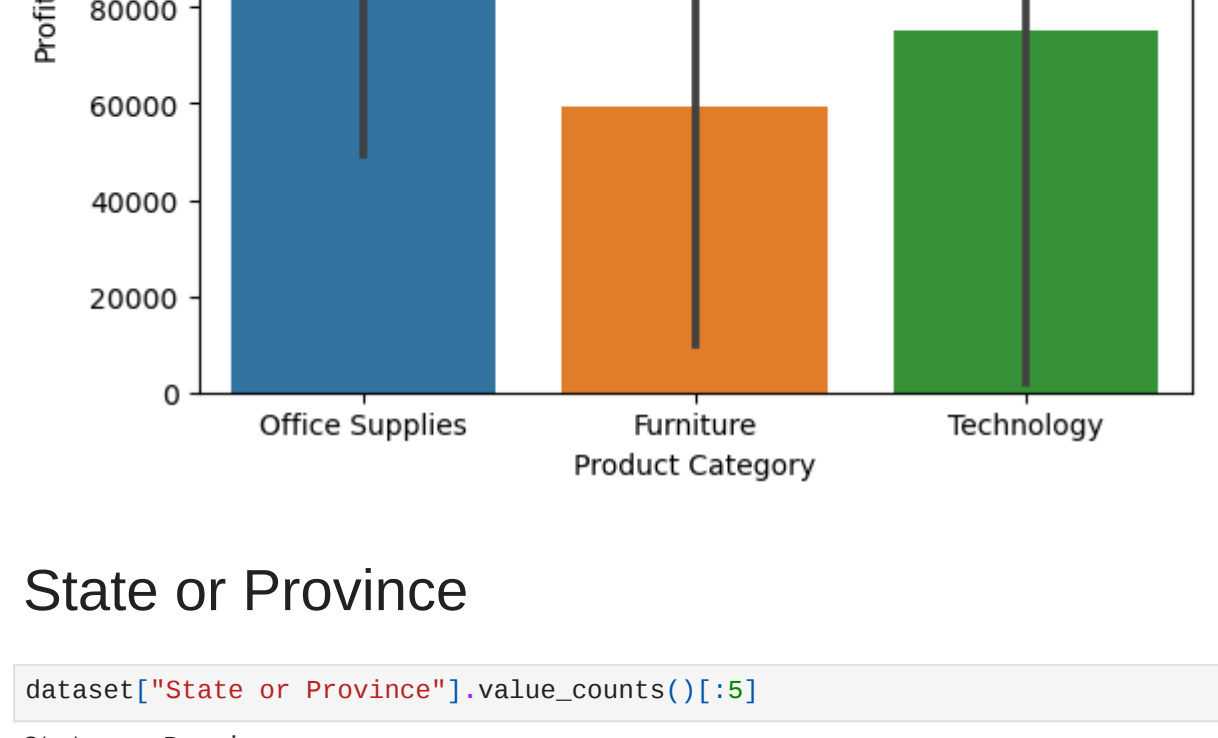
Out[28]:
```
Order year
2015    1952
Name: count, dtype: int64
```

In [31]:
```
plt.figure(figsize = (2,4))
sns.countplot(x="Order year", data = dataset)
plt.show()
```



Profit

In [50]:
```
sns.barplot(x="Product Category", y="Profit", data=dataset,estimator="sum")
plt.savefig(r"C:\Users\lenovo\Desktop\Profit.jpg")
plt.show()
```

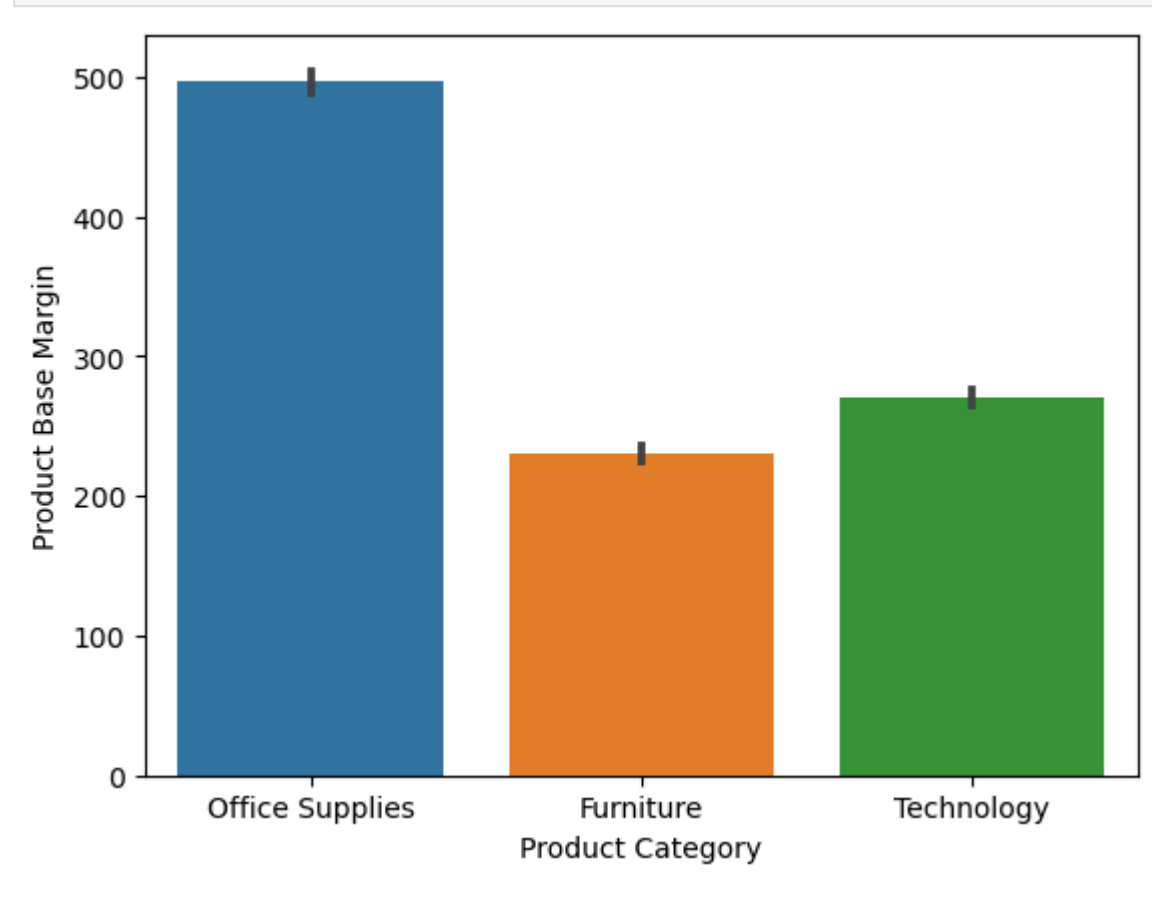


State or Province

In [38]:
```
dataset["State or Province"].value_counts()[:5]
```

Out[38]:
```
State or Province
California    214
New York      129
Texas         124
Florida       123
Illinois       88
Name: count, dtype: int64
```

In [51]:
```
sns.barplot(x="Product Category", y="Product Base Margin", data=dataset,estimator="sum")
plt.savefig(r"C:\Users\lenovo\Desktop\State or Province.jpg")
plt.show()
```



In [ ]: