E-commerse store Analysis

In [2]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as mp
import seaborn as sb
```

In [3]:
```
pd.set_option("display.max_columns",24)
```

In [6]:
```
dataset = pd.read_excel("Superstore_USA.xlsx")
```

In [7]:
```
dataset.head(2)
```

Out[7]:

	Row ID	Order Priority	Discount	Unit Price	Shipping Cost	Customer ID	Customer Name	Ship Mode	Customer Segment	Product Category	Product Sub-Category	Product Container	Product Name	Product Base Margin	Region	State or Province	City	Postal Code	Order Date	Ship Date	Profit	Quantity ordered new	Sales	Order ID
0	18606	Not Specified	0.01	2.88	0.50	2	Janice Fletcher	Regular Air	Corporate	Office Supplies	Labels	Small Box	Avery 49	0.36	Central	Illinois	Addison	60101	2012-05-28	2012-05-30	1.32	2	5.90	88525
1	20847	High	0.01	2.84	0.93	3	Bonnie Potter	Express Air	Corporate	Office Supplies	Pens & Art Supplies	Wrap Bag	SANFORD Liquid Accent™ Tank-Style Highlighters	0.54	West	Washington	Anacortes	98221	2010-07-07	2010-07-08	4.56	4	13.01	88522

Order Priority Analysis

In [14]:
```
#get the missing values
dataset.isnull().sum()
```

Out[14]:
```
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     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
```

In [13]:
```
#fill the missing values
dataset['Product Base Margin'].fillna(dataset['Product Base Margin'].mean(),inplace=True)
```

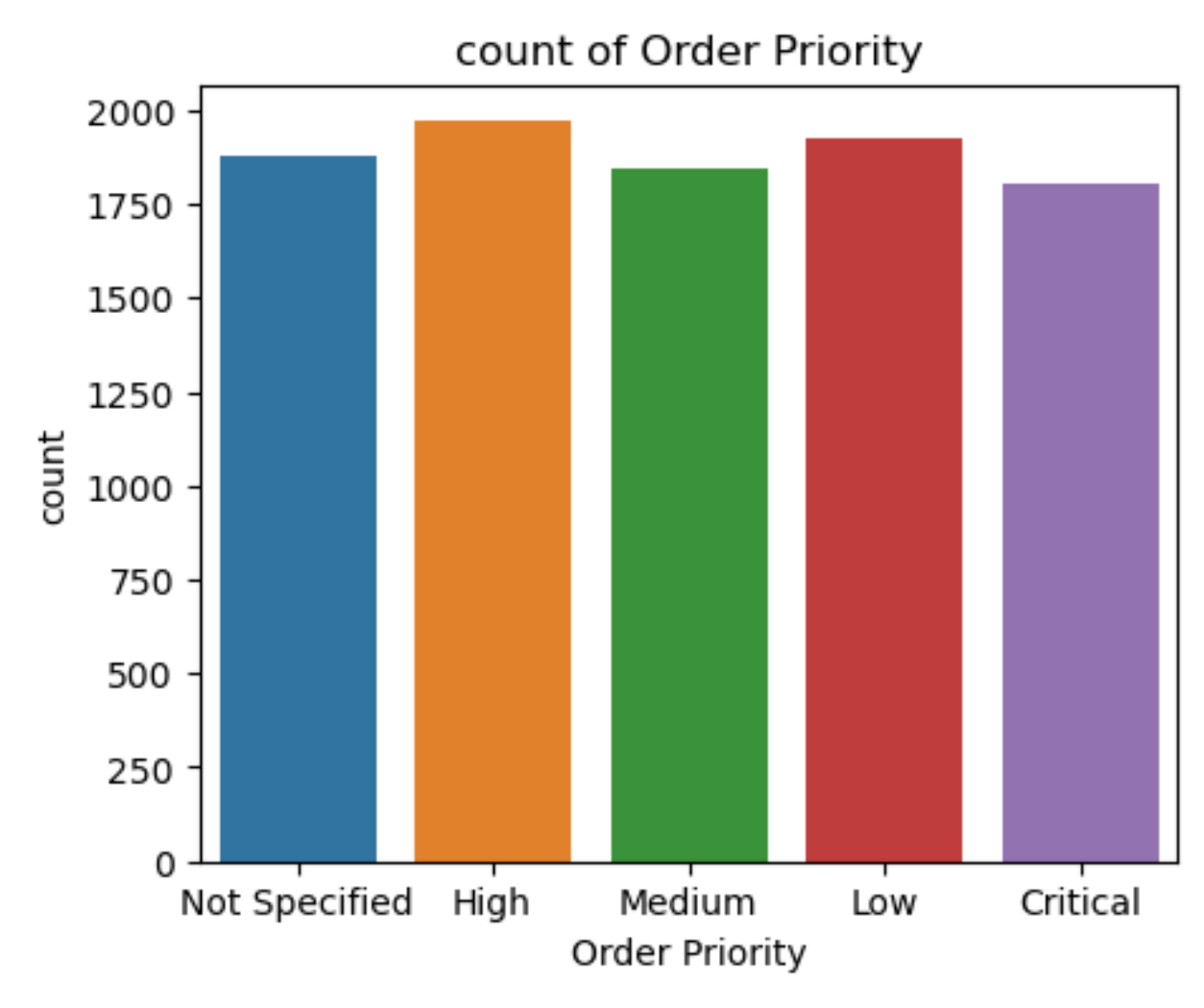
In [15]:
```
# total orders
dataset.shape
```

Out[15]:
```
(9426, 24)
```

In [29]:
```
#count the order
dataset['Order Priority'].value_counts()
```

Out[29]:
```
Order Priority
High             1970
Low              1926
Not Specified    1881
Medium           1844
Critical         1805
Name: count, dtype: int64
```

In [68]:
```
mp.figure(figsize=(5,4))
sb.countplot(x="Order Priority",data=dataset)
mp.title("count of Order Priority")
mp.savefig("count of Order Priority.jpg")
mp.show()
```



In [26]:
```
#to get the unique values
#dataset['Order Priority'].unique()
```

Out[26]:
```
array(['Not Specified', 'High', 'Medium', 'Low', 'Critical', 'Critical '],
      dtype=object)
```

In [28]:
```
#remove dupliacate value fields
#dataset['Order Priority']=dataset["Order Priority"].replace("Critical ","Critical")
```

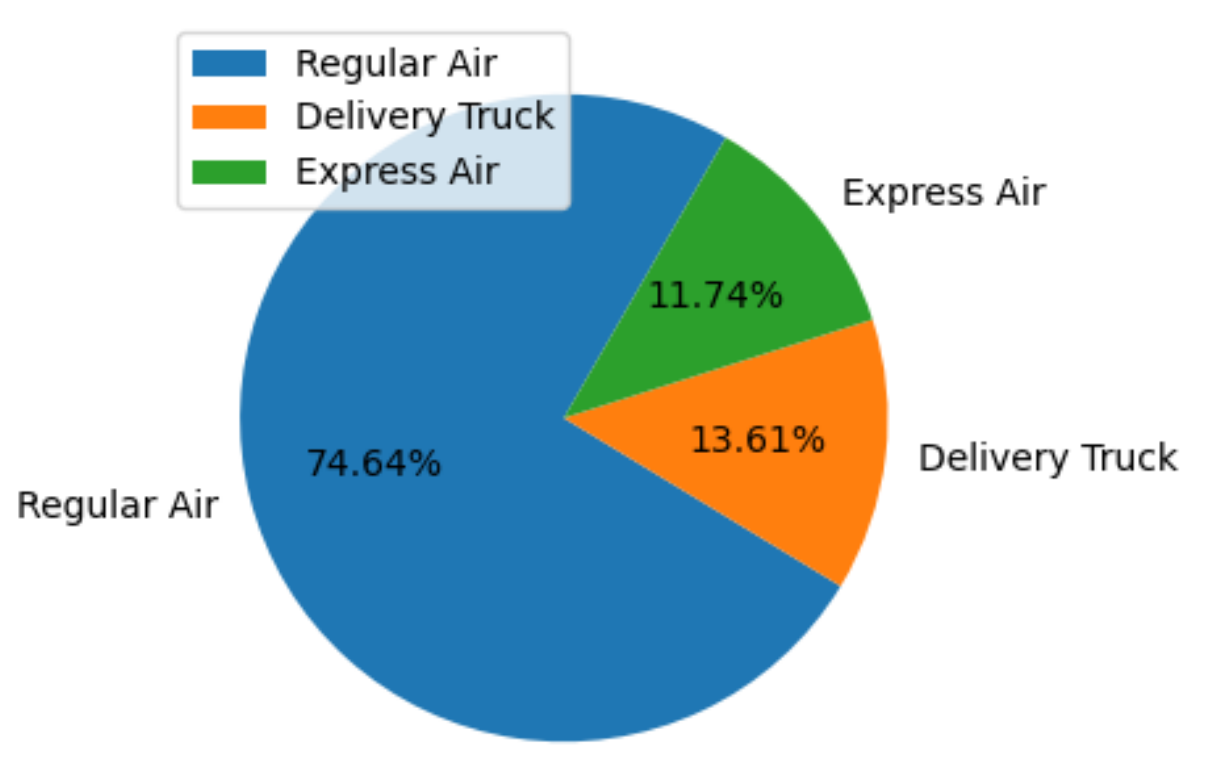
shipping mode analysis

In [38]:
```
#count the order
dataset['Ship Mode'].value_counts()
```

Out[38]:
```
Ship Mode
Regular Air       7036
Delivery Truck    1283
Express Air       1107
Name: count, dtype: int64
```

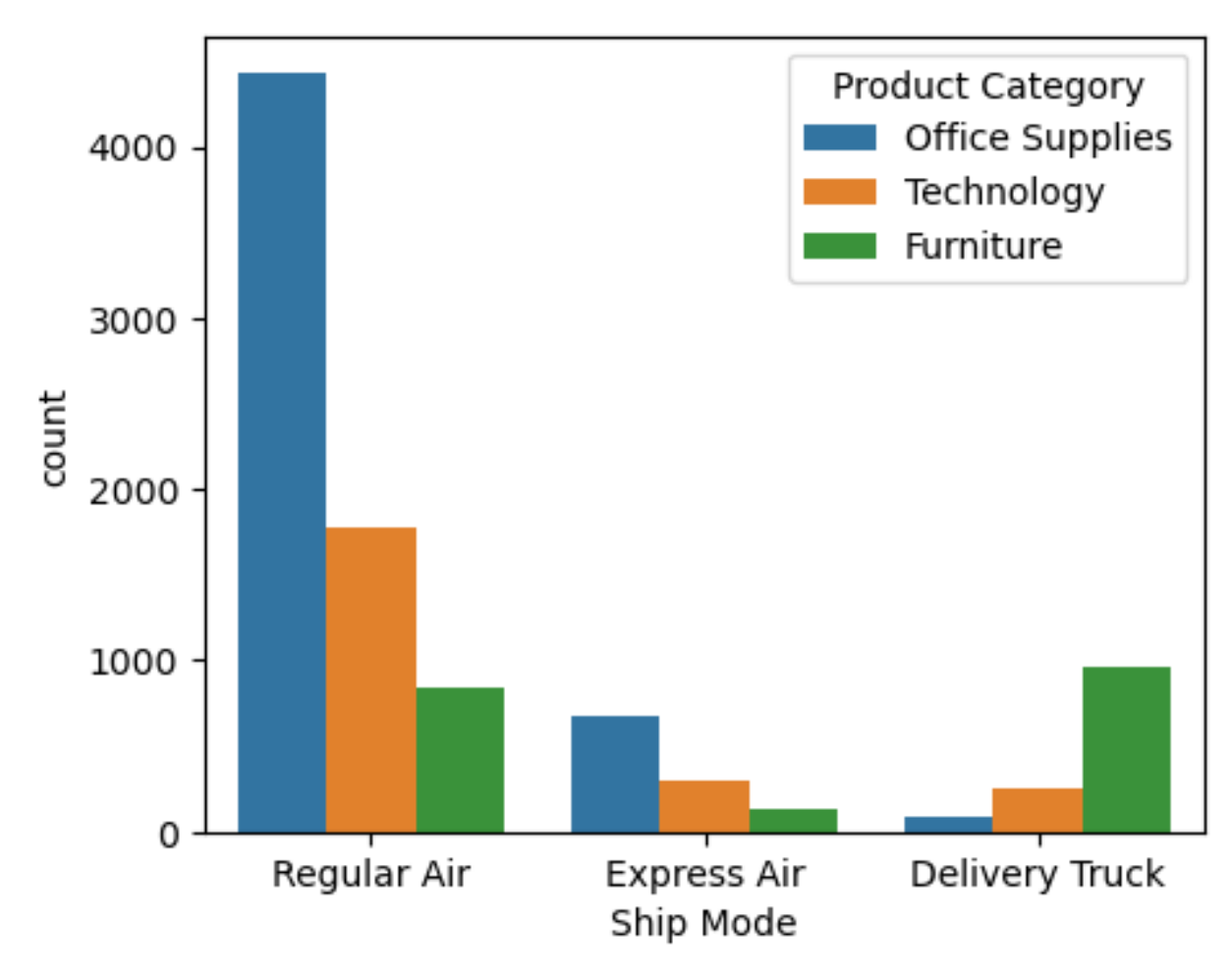
In [47]:
```
x=dataset['Ship Mode'].value_counts().index
y=dataset['Ship Mode'].value_counts().values
```

In [61]:
```
mp.figure(figsize=(5,4))
mp.pie(y,labels=x,startangle=60,autopct="%0.2f%%")
mp.legend(loc=2)
mp.show()
```



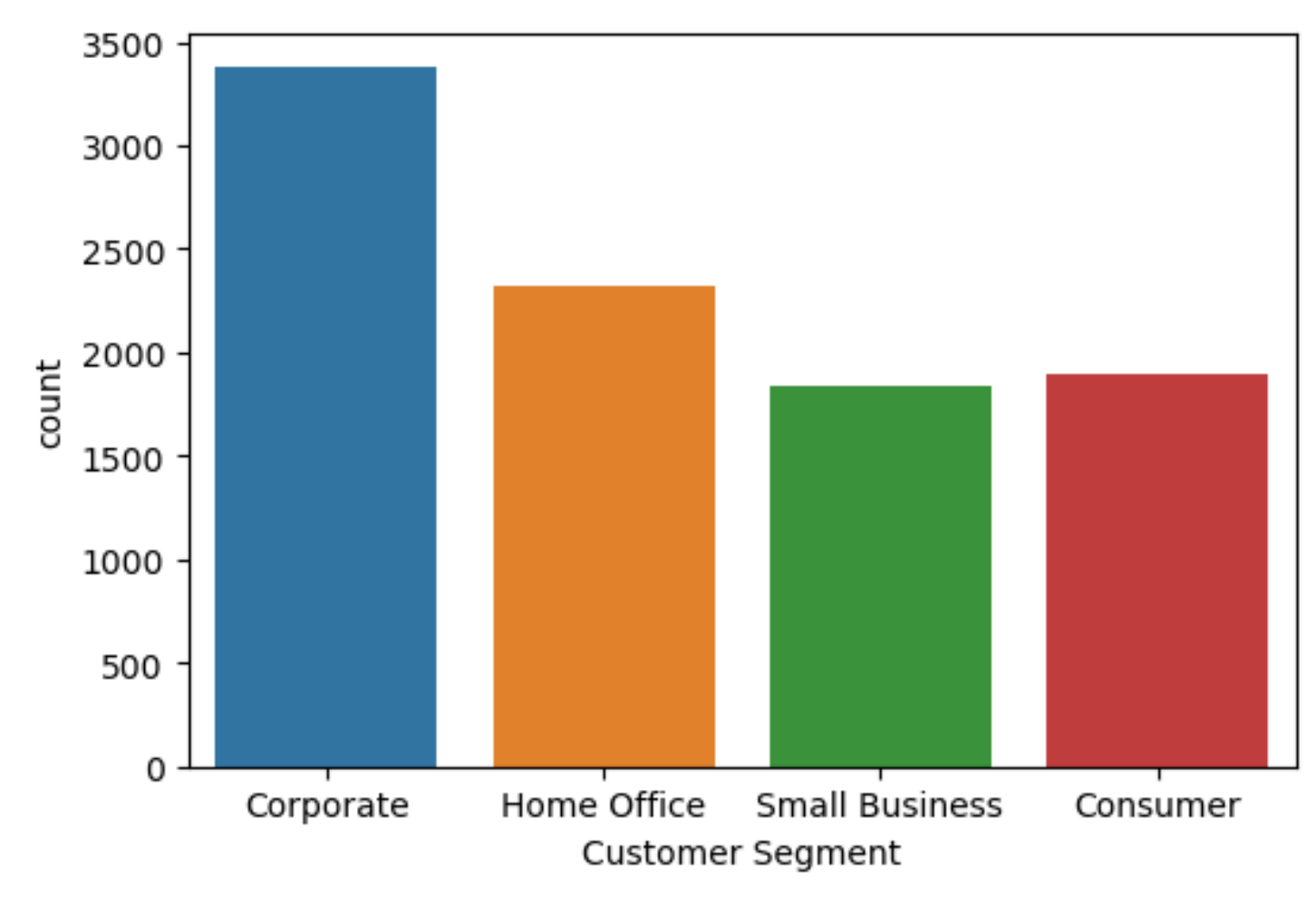
In [ ]:
```
#bivariate analysis
```

In [62]:
```
mp.figure(figsize=(5,4))
sb.countplot(x="Ship Mode",data=dataset,hue="Product Category")
mp.show()
```



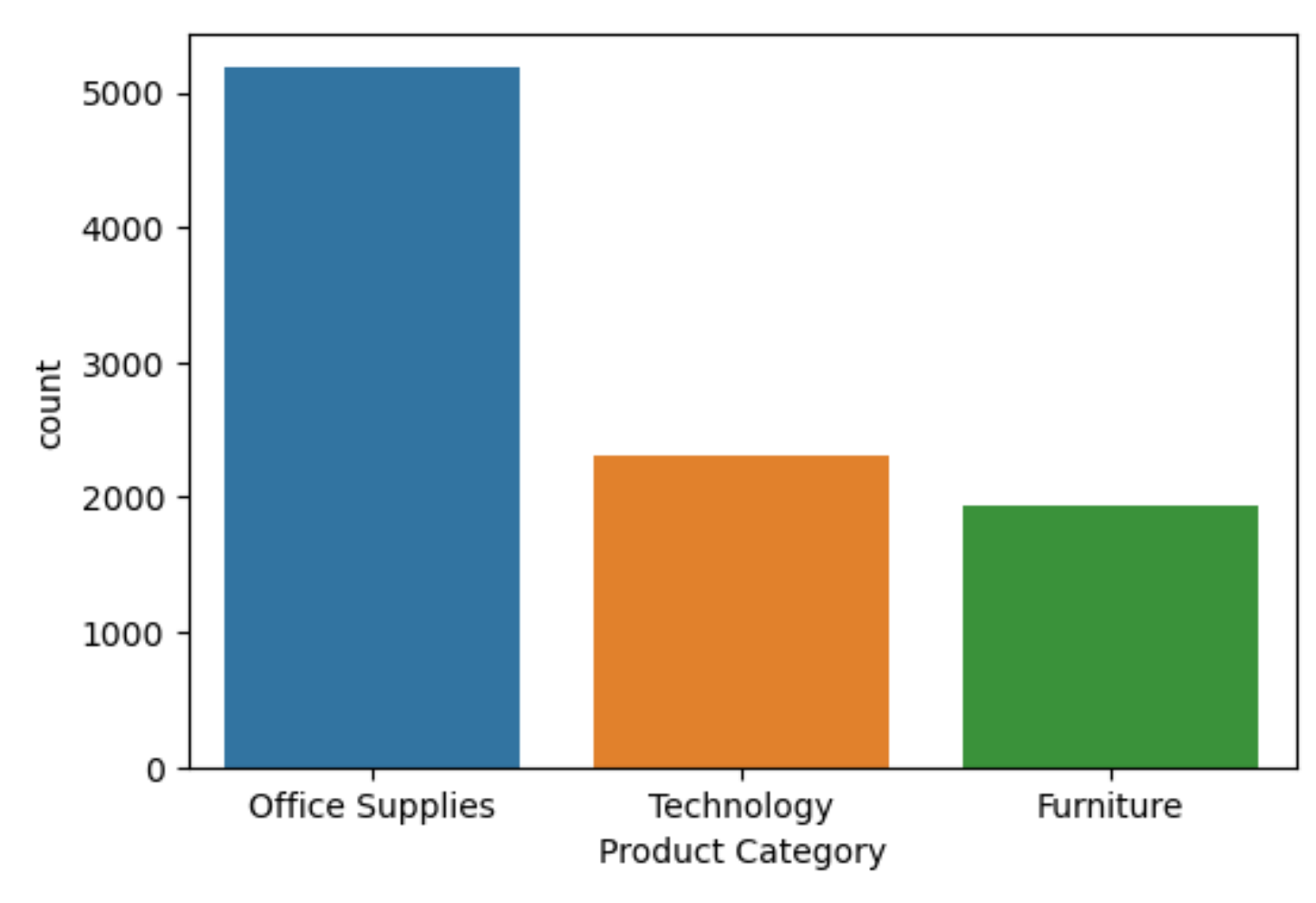
Customer Segment

In [64]:
```
mp.figure(figsize=(6,4))
sb.countplot(x="Customer Segment",data=dataset)
mp.show()
```

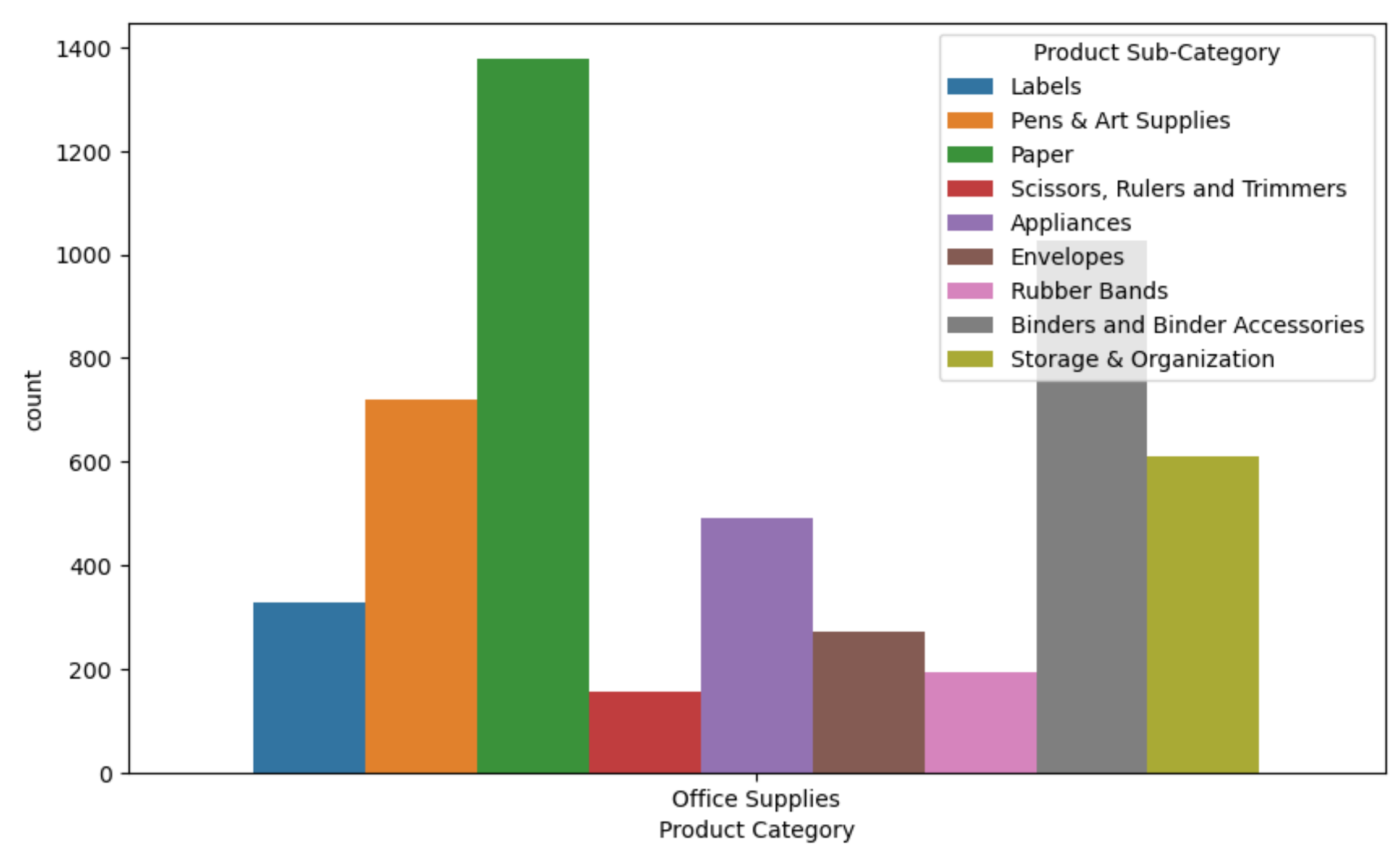


Product Category

In [65]:
```
mp.figure(figsize=(6,4))
sb.countplot(x="Product Category",data=dataset)
mp.show()
```



In [73]:
```
mp.figure(figsize=(10,6))
sb.countplot(x="Product Category",data=dataset[dataset["Product Category"]=="Office Supplies"],hue="Product Sub-Category")
mp.show()
```

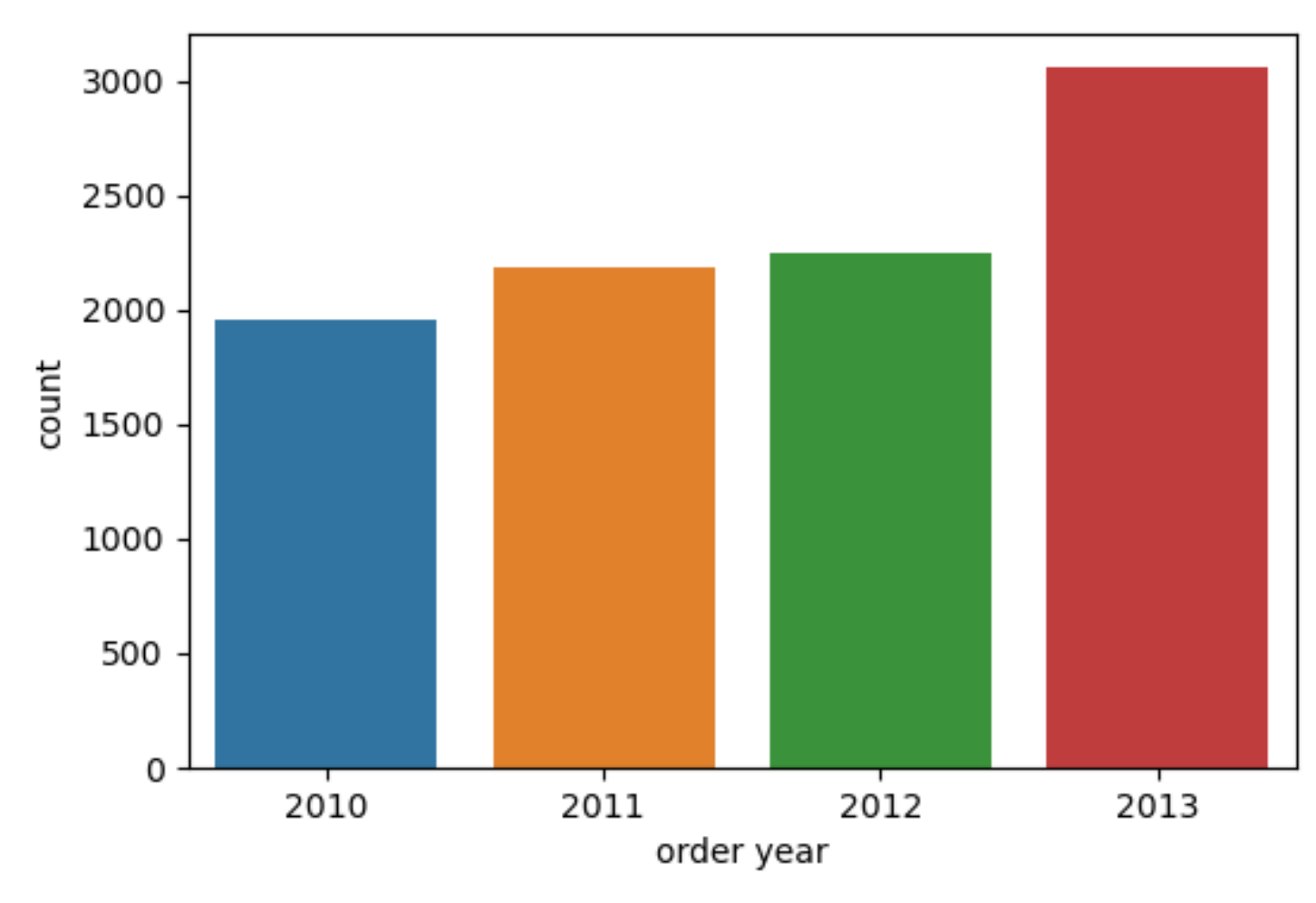


In [75]:
```
dataset['order year']= dataset['Order Date'].dt.year
```

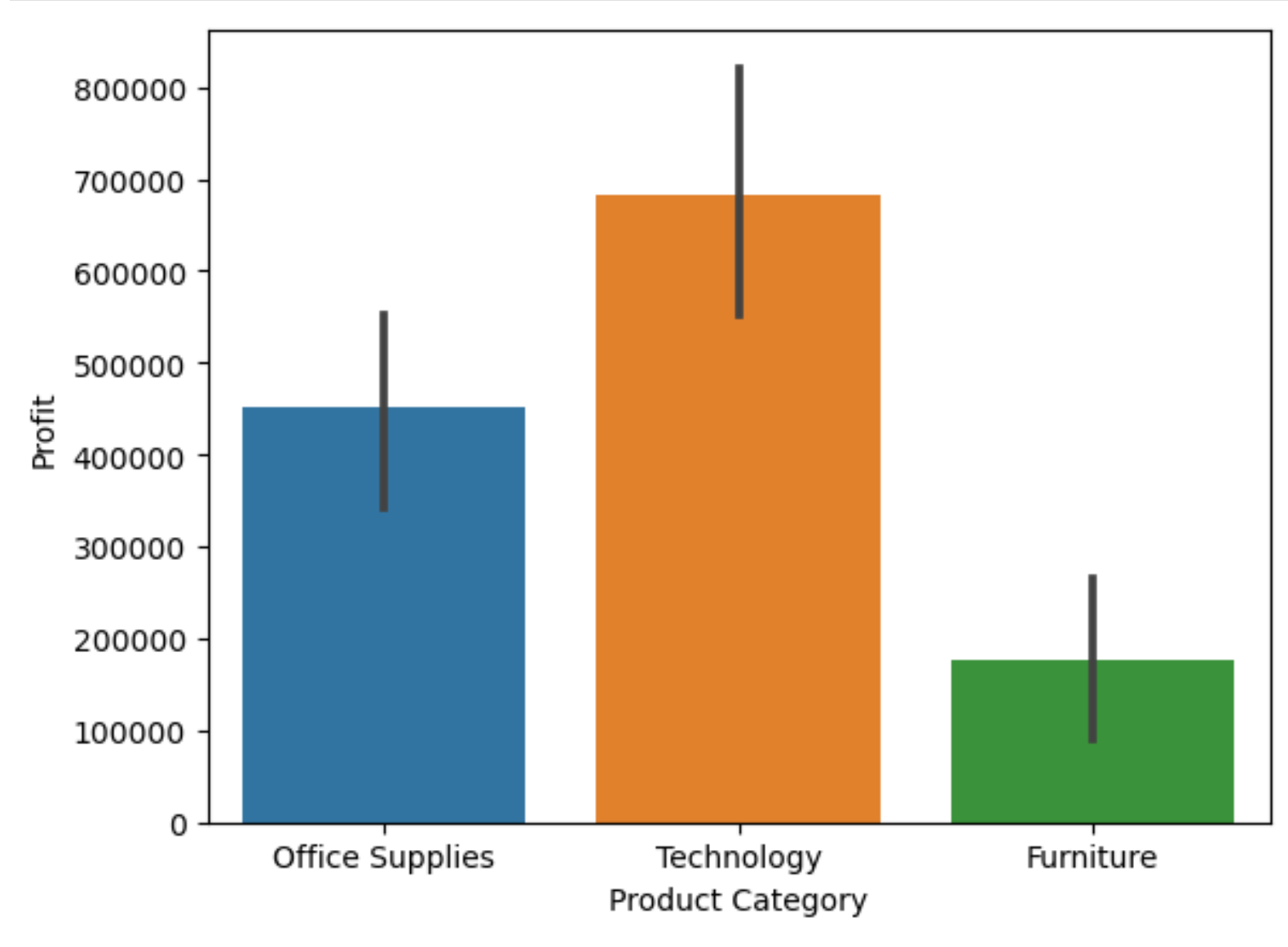
In [76]:
```
#find out which year has maximum sales
dataset['order year'].value_counts()
```

Out[76]:
```
order year
2013    3054
2012    2241
2011    2179
2010    1952
Name: count, dtype: int64
```

In [78]:
```
mp.figure(figsize=(6,4))
sb.countplot(x="order year",data=dataset)
mp.show()
```



In [81]:
```
# which category has maximum profit
sb.barplot(x="Product Category",y="Profit",data=dataset,estimator='sum')
mp.show()
```



In [82]:
```
# which state has what sales
dataset['State or Province'].value_counts()
```

Out[82]:
```
State or Province
California              1021
Texas                    646
Illinois                 584
New York                 574
Florida                  522
Ohio                     396
Washington               327
Michigan                 327
Pennsylvania             271
North Carolina           251
Indiana                  241
Minnesota                239
Massachusetts            222
Georgia                  214
Virginia                 198
Maryland                 178
Colorado                 177
New Jersey               177
Wisconsin                169
Oregon                   168
Tennessee                166
Missouri                 161
Iowa                     156
Utah                     146
Arizona                  134
Kansas                   133
Maine                    128
Alabama                  125
Arkansas                 123
Idaho                    114
South Carolina           105
Oklahoma                 104
Louisiana                 89
New Mexico                84
Kentucky                  83
Connecticut               82
Mississippi               78
Nebraska                  77
District of Columbia      68
Vermont                   61
New Hampshire             54
Montana                   49
West Virginia             43
Nevada                    43
North Dakota              34
South Dakota              28
Wyoming                   21
Rhode Island              20
Delaware                  15
Name: count, dtype: int64
```

In [ ]:

In [ ]:

In [ ]: