

```
pip install openpyxl
```

```
Collecting openpyxl
```

```
  Downloading openpyxl-3.1.5-py2.py3-none-any.whl.metadata (2.5 kB)
```

```
Collecting et_xmlfile (from openpyxl)
```

```
  Downloading et_xmlfile-2.0.0-py3-none-any.whl.metadata (2.7 kB)
```

```
Downloading openpyxl-3.1.5-py2.py3-none-any.whl (250 kB)
```

```
Downloading et_xmlfile-2.0.0-py3-none-any.whl (18 kB)
```

```
Installing collected packages: et_xmlfile, openpyxl
```

```
----- 1/2 [openpyxl]
----- 1/2 [openpyxl]
----- 1/2 [openpyxl]
----- 1/2 [openpyxl]
----- 1/2 [openpyxl]
----- 1/2 [openpyxl]
----- 2/2 [openpyxl]
```

```
Successfully installed et_xmlfile-2.0.0 openpyxl-3.1.5
```

```
Note: you may need to restart the kernel to use updated packages.
```

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
df = pd.read_excel("C:\\Users\\stupa\\Desktop\\E\\
Superstore_USA.xlsx")
```

```
df.head(10)
```

	Row	ID	Order	Priority	Discount	Unit	Price	Shipping	Cost
Customer	ID	\							
0	18606	Not Specified		0.01	2.88		0.50		
2									
1	20847	High		0.01	2.84		0.93		
3									
2	23086	Not Specified		0.03	6.68		6.15		
3									
3	23087	Not Specified		0.01	5.68		3.60		
3									
4	23088	Not Specified		0.00	205.99		2.50		
3									
5	23597	Medium		0.09	55.48		14.30		
3									
6	25549	Low		0.08	120.97		26.30		
3									
7	20228	Not Specified		0.02	500.98		26.00		
5									
8	19483	Low		0.08	6.48		6.81		
5									
9	24782	High		0.01	90.24		0.99		

6

	Customer Name	Ship Mode	Customer Segment	Product
Category ... \				
0	Janice Fletcher	Regular Air	Corporate	Office Supplies ...
1	Bonnie Potter	Express Air	Corporate	Office Supplies ...
2	Bonnie Potter	Express Air	Corporate	Office Supplies ...
3	Bonnie Potter	Regular Air	Corporate	Office Supplies ...
4	Bonnie Potter	Express Air	Corporate	Technology ...
5	Bonnie Potter	Express Air	Corporate	Office Supplies ...
6	Bonnie Potter	Delivery Truck	Corporate	Technology ...
7	Ronnie Proctor	Delivery Truck	Home Office	Furniture ...
8	Ronnie Proctor	Regular Air	Home Office	Office Supplies ...
9	Dwight Hwang	Regular Air	Home Office	Office Supplies ...

	Region	State or Province	City	Postal Code	Order Date
Ship Date \					
0	Central	Illinois	Addison	60101	2012-05-28
2012-05-30					
1	West	Washington	Anacortes	98221	2010-07-07
2010-07-08					
2	West	Washington	Anacortes	98221	2011-07-27
2011-07-28					
3	West	Washington	Anacortes	98221	2011-07-27
2011-07-28					
4	West	Washington	Anacortes	98221	2011-07-27
2011-07-27					
5	West	Washington	Anacortes	98221	2011-11-09
2011-11-11					
6	West	Washington	Anacortes	98221	2013-07-01
2013-07-08					
7	West	California	San Gabriel	91776	2010-12-13
2010-12-15					
8	West	California	San Gabriel	91776	2012-05-12
2012-05-21					
9	West	California	San Jose	95123	2011-05-26
2011-05-26					

	Profit	Quantity ordered	new	Sales	Order ID
0	1.3200		2	5.90	88525

1	4.5600	4	13.01	88522
2	-47.6400	7	49.92	88523
3	-30.5100	7	41.64	88523
4	998.2023	8	1446.67	88523
5	1388.0523	37	2011.67	88524
6	1001.4453	12	1451.37	88526
7	4390.3665	12	6362.85	90193
8	-141.2600	18	113.25	90197
9	1045.4673	16	1515.17	90194

[10 rows x 24 columns]

df.shape

(9426, 24)

df.isnull().sum()

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	72
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

```
df["Product Base Margin"] = df["Product Base Margin"].fillna(df["Product Base Margin"].mean())
```

```
dataset["Order Priority"].value_counts()
```

Order Priority	
High	1970

```

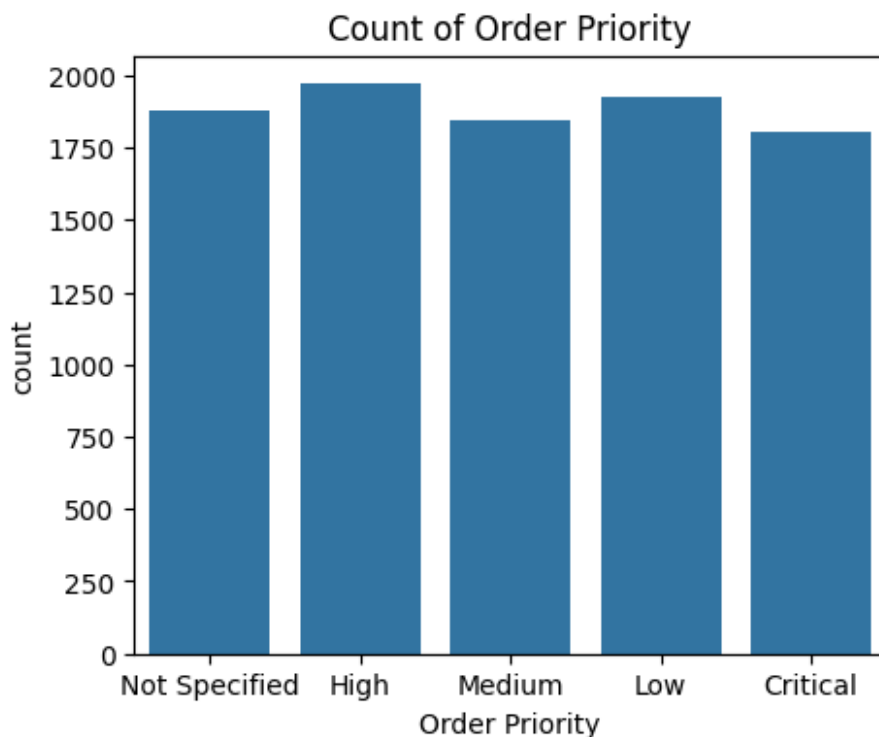
Low          1926
Not Specified 1881
Medium       1844
Critical     1804
Critical      1
Name: count, dtype: int64

df["Order Priority"] = df["Order Priority"].str.strip().str.title()
print(df["Order Priority"].value_counts())

Order Priority
High          1970
Low           1926
Not Specified 1881
Medium        1844
Critical      1805
Name: count, dtype: int64

plt.figure(figsize=(5, 4))
sns.countplot(x="Order Priority", data=df)
plt.title("Count of Order Priority")
plt.savefig("Count of Order Priority.jpg")
plt.show()

```



```
df["Ship Mode"].value_counts()
```

```

Ship Mode
Regular Air      7036
Delivery Truck   1283
Express Air      1107
Name: count, dtype: int64

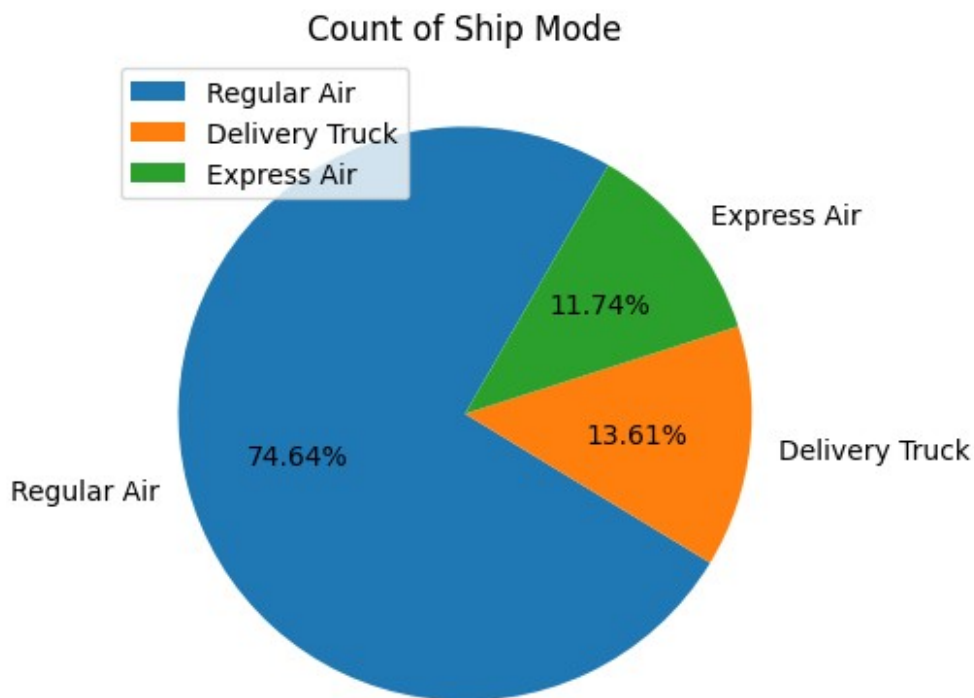
```

```

x=df["Ship Mode"].value_counts().index
y=df["Ship Mode"].value_counts().values

plt.pie(y, labels=x, startangle = 60, autopct="%0.2f%%")
plt.title("Count of Ship Mode")
plt.savefig("Count of Ship Mode.jpg")
plt.legend(loc=2)
plt.show()

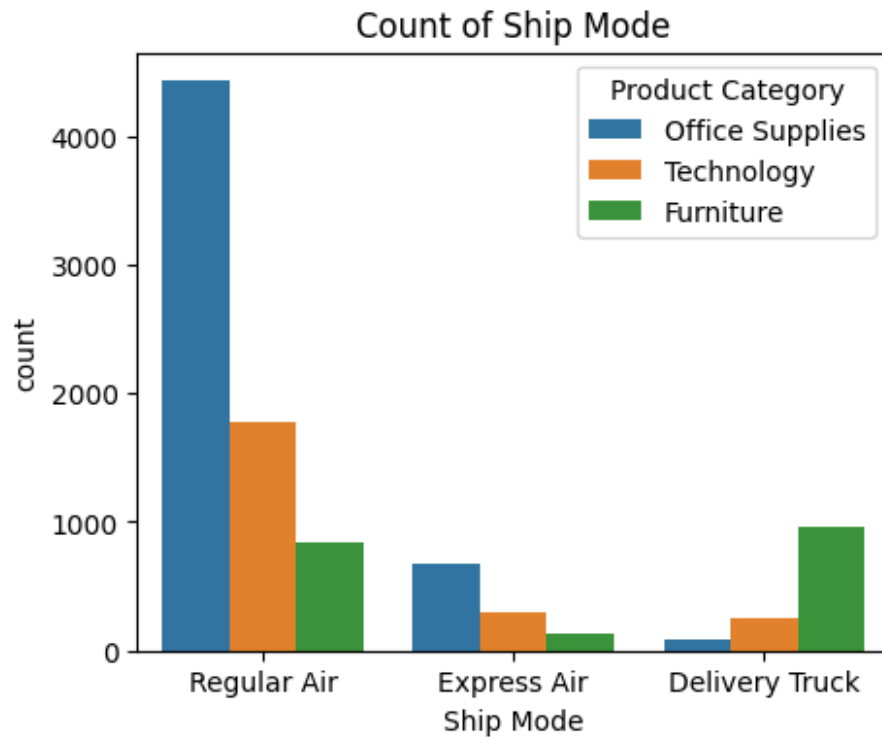
```



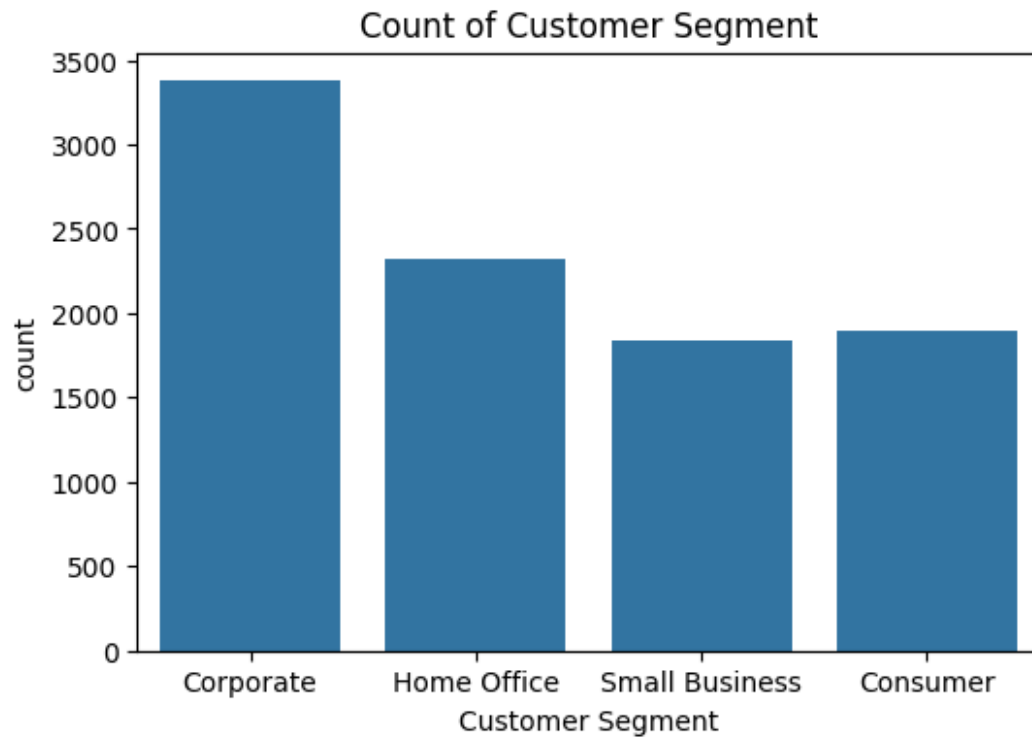
```

plt.figure(figsize=(5, 4))
sns.countplot(x="Ship Mode", data=df, hue="Product Category")
plt.title("Count of Ship Mode")
plt.savefig("Count of Ship Mode Bar.jpg")
plt.show()

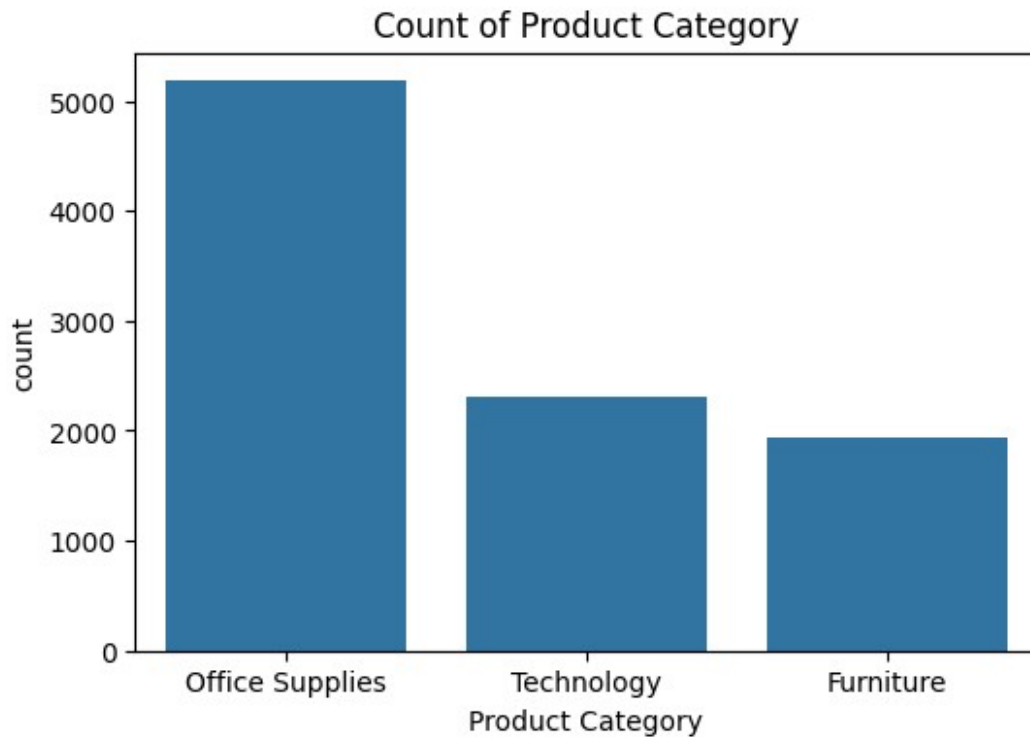
```



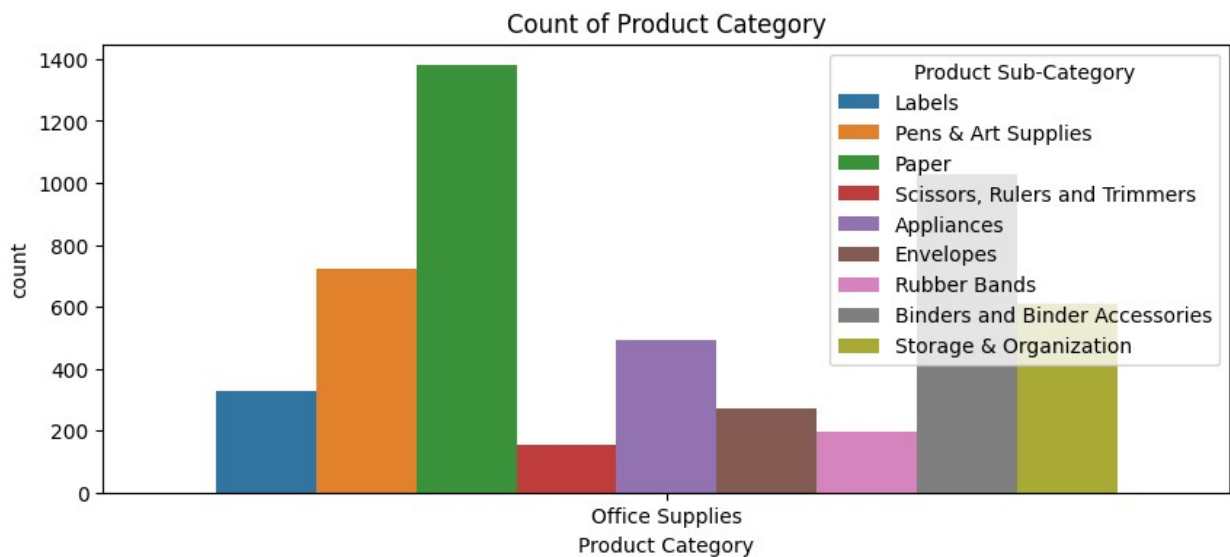
```
plt.figure(figsize=(6,4))
sns.countplot(x="Customer Segment", data=df)
plt.title("Count of Customer Segment")
plt.savefig("Customer Segment.jpg")
plt.show()
```



```
plt.figure(figsize=(6,4))
sns.countplot(x="Product Category", data=df)
plt.title("Count of Product Category")
plt.savefig("Product Category.jpg")
plt.show()
```



```
plt.figure(figsize=(10,4))
sns.countplot(x="Product Category", data=df[df["Product Category"]=="Office Supplies"], hue="Product Sub-Category")
plt.title("Count of Product Category")
plt.savefig("Product Category.jpg")
plt.show()
```



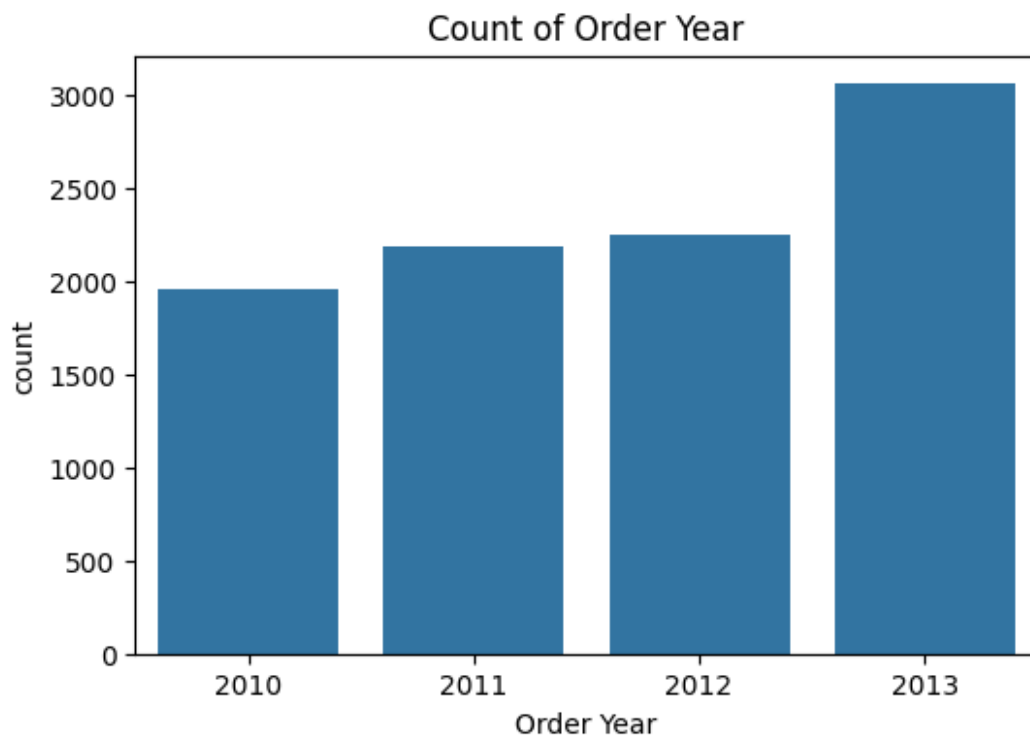
```
df["Order Year"]=df["Order Date"].dt.year
```



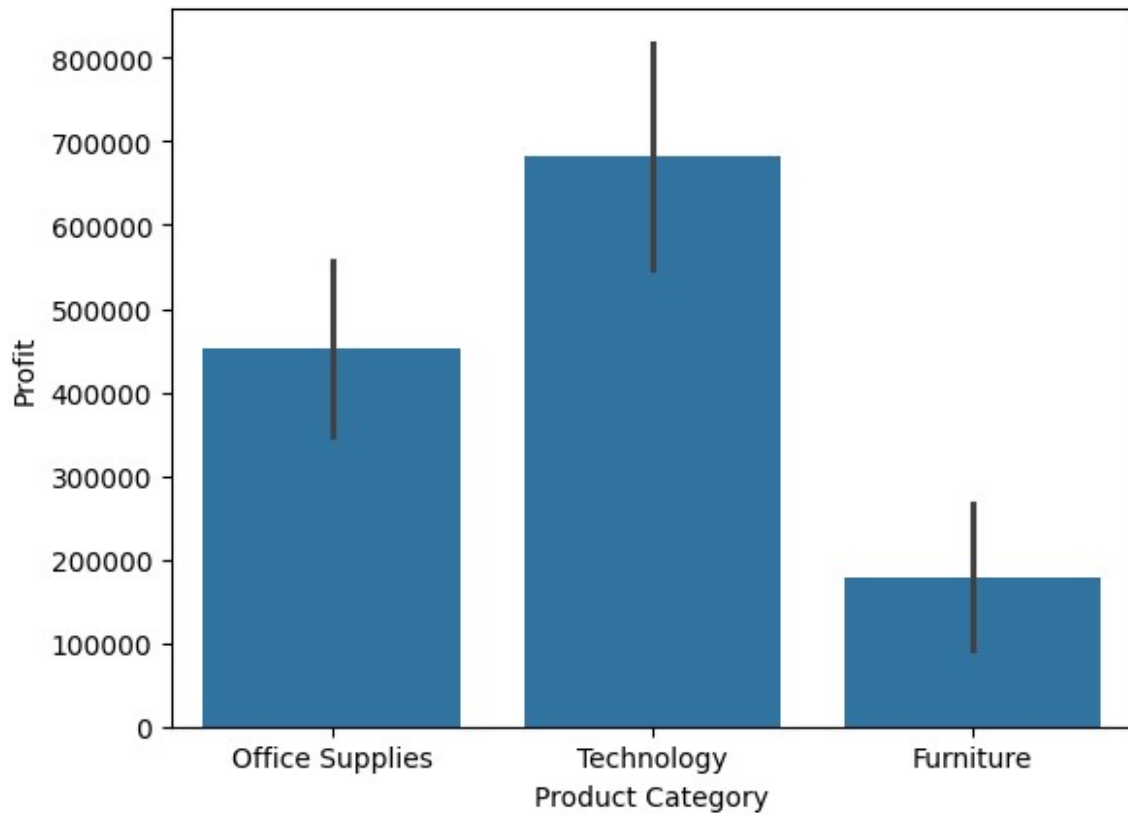
```
df["Order Year"].value_counts()

Order Year
2013    3054
2012    2241
2011    2179
2010    1952
Name: count, dtype: int64

plt.figure(figsize=(6,4))
sns.countplot(x="Order Year", data=df)
plt.title("Count of Order Year")
plt.savefig("Order Year.jpg")
plt.show()
```



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



```
df["State or Province"].value_counts()[:5]
```

```
State or Province
California      1021
Texas           646
Illinois        584
New York        574
Florida         522
Name: count, dtype: int64
```