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

dataset = pd.read_excel("C:\\Users\\HP\\Downloads\\file\\
Superstore_USA.xlsx")

dataset
```

```
      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
...      ...            ...       ...         ...            ...
...
9421   20275       Critical      0.06       35.89          14.72
3402
9422   20276       Critical      0.00        3.34           7.49
3402
9423   24491  Not Specified      0.08      550.98          45.70
3402
9424   25914           High      0.10      105.98          13.99
3403
9425   24492  Not Specified      0.09        7.78           2.50
3403

        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  ...
...               ...             ...              ...              ..
.  ...
9421   Frederick Cole     Regular Air         Consumer  Office
```

```
Supplies  ...
9422   Frederick Cole     Regular Air          Consumer  Office
Supplies  ...
9423   Frederick Cole  Delivery Truck          Consumer
Furniture  ...
9424    Tammy Buckley     Express Air          Consumer
Furniture  ...
9425    Tammy Buckley     Express Air          Consumer  Office
Supplies  ...

       Region State or Province        City  Postal Code Order Date  \
0     Central          Illinois     Addison        60101 2012-05-28
1        West        Washington   Anacortes        98221 2010-07-07
2        West        Washington   Anacortes        98221 2011-07-27
3        West        Washington   Anacortes        98221 2011-07-27
4        West        Washington   Anacortes        98221 2011-07-27
...       ...               ...         ...          ...         ...
9421     East     West Virginia  Charleston        25314 2013-05-14
9422     East     West Virginia  Charleston        25314 2013-05-14
9423     East     West Virginia  Charleston        25314 2013-09-12
9424     West           Wyoming    Cheyenne        82001 2010-02-08
9425     West           Wyoming    Cheyenne        82001 2013-09-12

      Ship Date       Profit  Quantity ordered new    Sales Order ID
0    2012-05-30     1.320000                     2     5.90    88525
1    2010-07-08     4.560000                     4    13.01    88522
2    2011-07-28   -47.640000                     7    49.92    88523
3    2011-07-28   -30.510000                     7    41.64    88523
4    2011-07-27   998.202300                     8  1446.67    88523
...         ...          ...                   ...     ...      ...
9421 2013-05-15   137.860000                    13   447.87    87532
9422 2013-05-14   -39.070000                     3    13.23    87532
9423 2013-09-14 -1225.029097                     4  2215.93    87533
9424 2010-02-11   349.485000                     5   506.50    87530
9425 2013-09-14    78.062400                    23   172.48    87533

[9426 rows x 24 columns]

dataset.head(5)

   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

     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  ...


    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

     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 rows x 24 columns]
```

```
dataset.shape
```

```
(9426, 24)
```

```
dataset.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      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
```

```
dataset["Product Base Margin"].fillna(dataset["Product Base
Margin"].mean(), inplace = True)
```

Order Priority

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

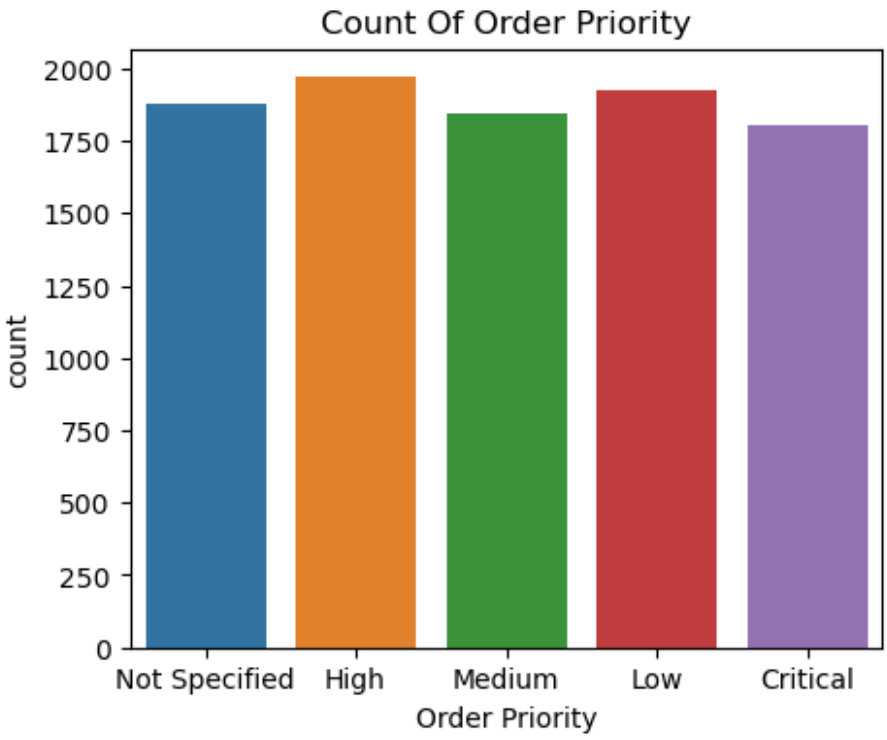
```
Order Priority
High             1970
Low              1926
Not Specified    1881
Medium           1844
Critical         1805
Name: count, dtype: int64
```

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

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

```
dataset["Order Priority"] = dataset["Order
Priority"].replace("Critical ", "Critical")

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



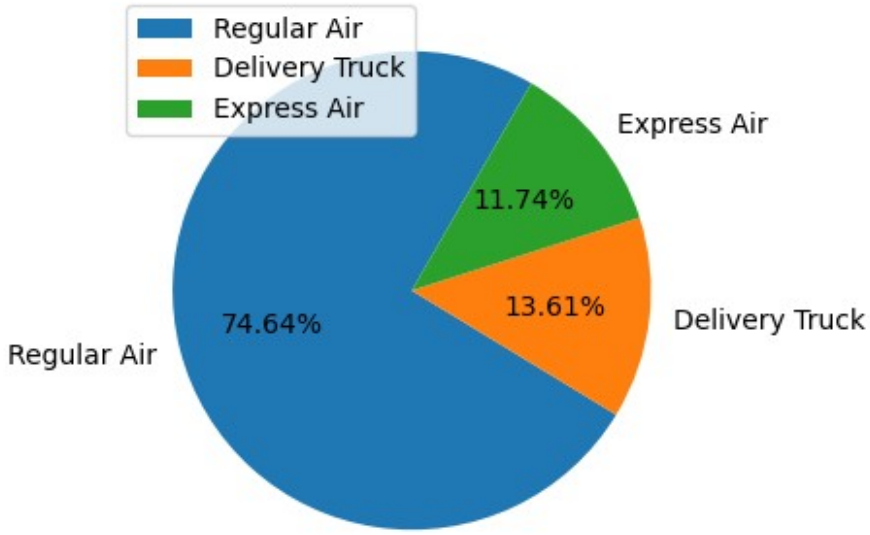
Ship mode

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

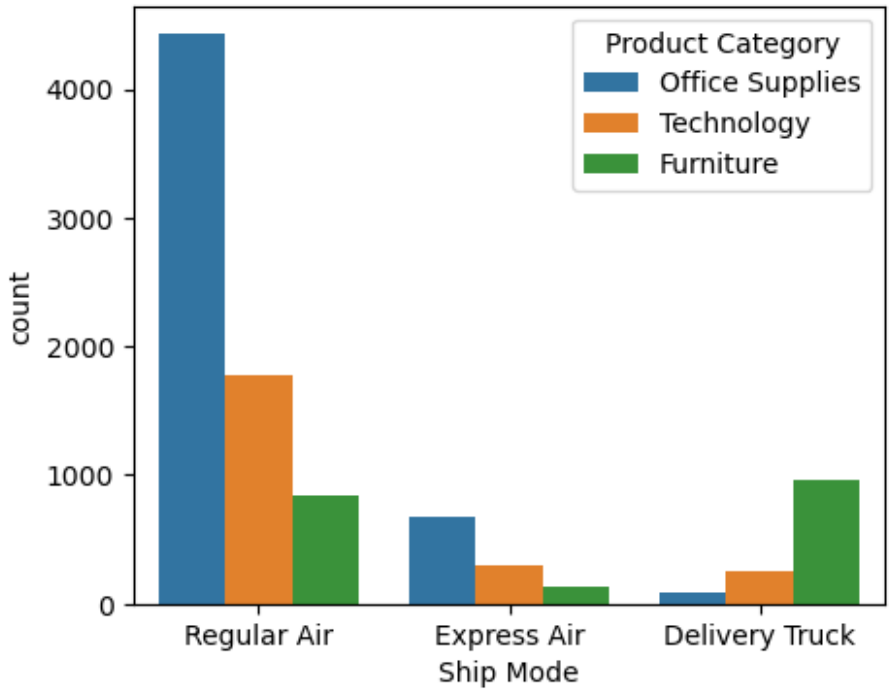
```
Ship Mode
Regular Air       7036
Delivery Truck    1283
Express Air       1107
Name: count, dtype: int64
```

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

plt.figure(figsize = (5,4))
plt.pie(y, labels = x, startangle = 60, autopct = "%0.2f%%")
plt.legend(loc = 2)
plt.show()
```

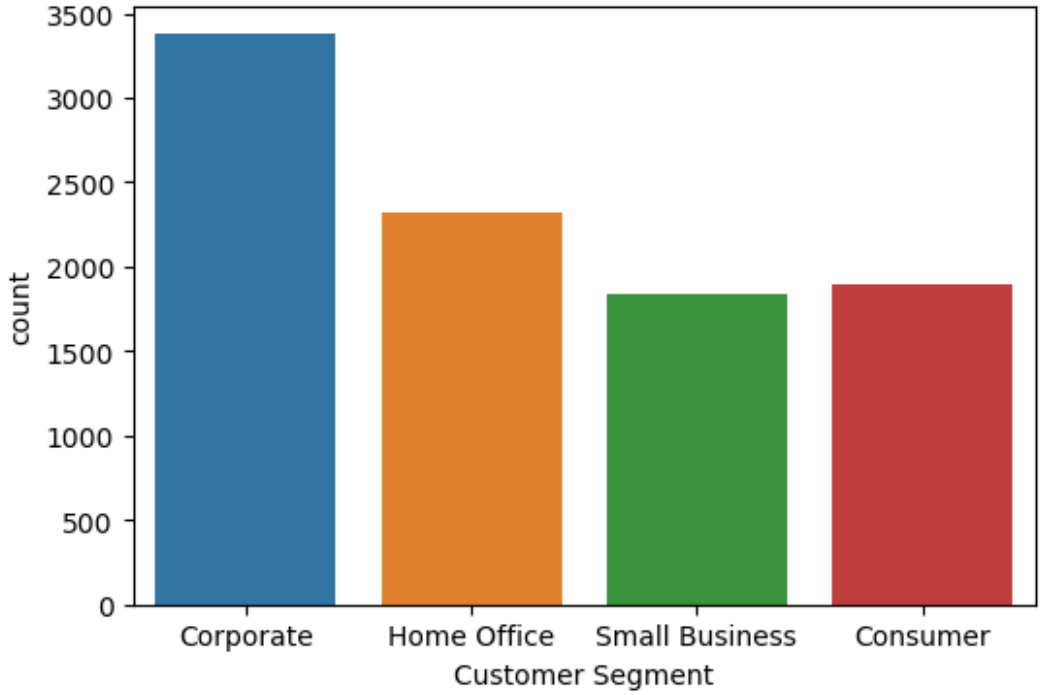



```
plt.figure(figsize = (5,4))
sns.countplot(x = "Ship Mode", data = dataset, hue = "Product
Category")
plt.show()
```



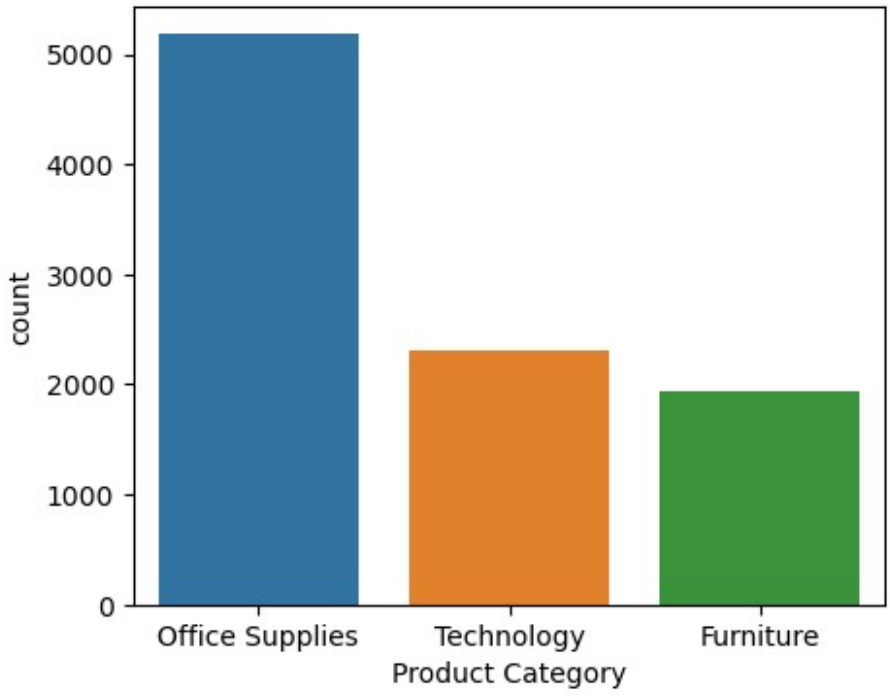
#Customer Segment

```
plt.figure(figsize = (6,4))
sns.countplot(x = "Customer Segment", data = dataset)
plt.show()
```

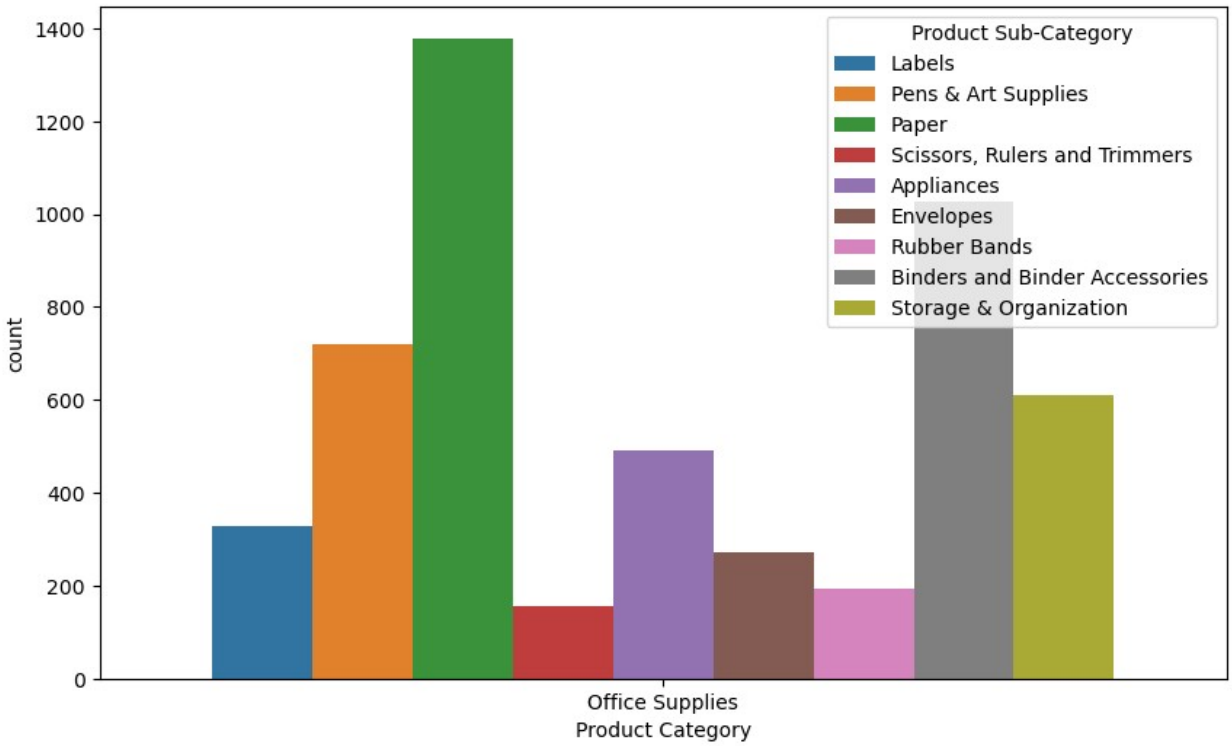


#Product Category

```
plt.figure(figsize = (5,4))
sns.countplot(x = "Product Category", data = dataset)
plt.show()
```




```
plt.figure(figsize = (10,6))
sns.countplot(x = "Product Category", data = dataset[dataset["Product
Category"] == "Office Supplies"], hue = "Product Sub-Category")
plt.show()
```



```
dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9426 entries, 0 to 9425
Data columns (total 24 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   Row ID                9426 non-null   int64
 1   Order Priority        9426 non-null   object
 2   Discount              9426 non-null   float64
 3   Unit Price            9426 non-null   float64
 4   Shipping Cost         9426 non-null   float64
 5   Customer ID           9426 non-null   int64
 6   Customer Name         9426 non-null   object
 7   Ship Mode             9426 non-null   object
 8   Customer Segment      9426 non-null   object
 9   Product Category      9426 non-null   object
 10  Product Sub-Category  9426 non-null   object
 11  Product Container     9426 non-null   object
 12  Product Name          9426 non-null   object
 13  Product Base Margin   9354 non-null   float64
 14  Region                9426 non-null   object
 15  State or Province     9426 non-null   object
 16  City                  9426 non-null   object
 17  Postal Code           9426 non-null   int64
 18  Order Date            9426 non-null   datetime64[ns]
 19  Ship Date             9426 non-null   datetime64[ns]
 20  Profit                9426 non-null   float64
 21  Quantity ordered new  9426 non-null   int64
 22  Sales                 9426 non-null   float64
 23  Order ID              9426 non-null   int64
dtypes: datetime64[ns](2), float64(6), int64(5), object(11)
memory usage: 1.7+ MB
```

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

```
dataset.info()
```

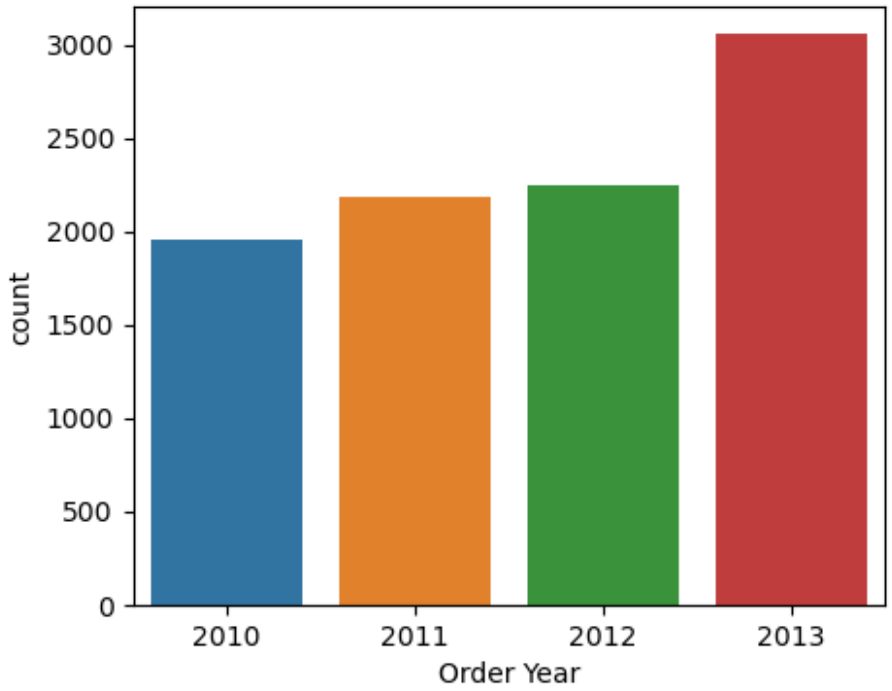
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9426 entries, 0 to 9425
Data columns (total 25 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   Row ID                9426 non-null   int64
 1   Order Priority        9426 non-null   object
 2   Discount              9426 non-null   float64
 3   Unit Price            9426 non-null   float64
 4   Shipping Cost         9426 non-null   float64
 5   Customer ID           9426 non-null   int64
 6   Customer Name         9426 non-null   object
```

```
 7   Ship Mode             9426 non-null   object
 8   Customer Segment      9426 non-null   object
 9   Product Category      9426 non-null   object
 10  Product Sub-Category  9426 non-null   object
 11  Product Container     9426 non-null   object
 12  Product Name          9426 non-null   object
 13  Product Base Margin   9354 non-null   float64
 14  Region                9426 non-null   object
 15  State or Province     9426 non-null   object
 16  City                  9426 non-null   object
 17  Postal Code           9426 non-null   int64
 18  Order Date            9426 non-null   datetime64[ns]
 19  Ship Date             9426 non-null   datetime64[ns]
 20  Profit                9426 non-null   float64
 21  Quantity ordered new  9426 non-null   int64
 22  Sales                 9426 non-null   float64
 23  Order ID              9426 non-null   int64
 24  Order Year            9426 non-null   int32
dtypes: datetime64[ns](2), float64(6), int32(1), int64(5), object(11)
memory usage: 1.8+ MB
```

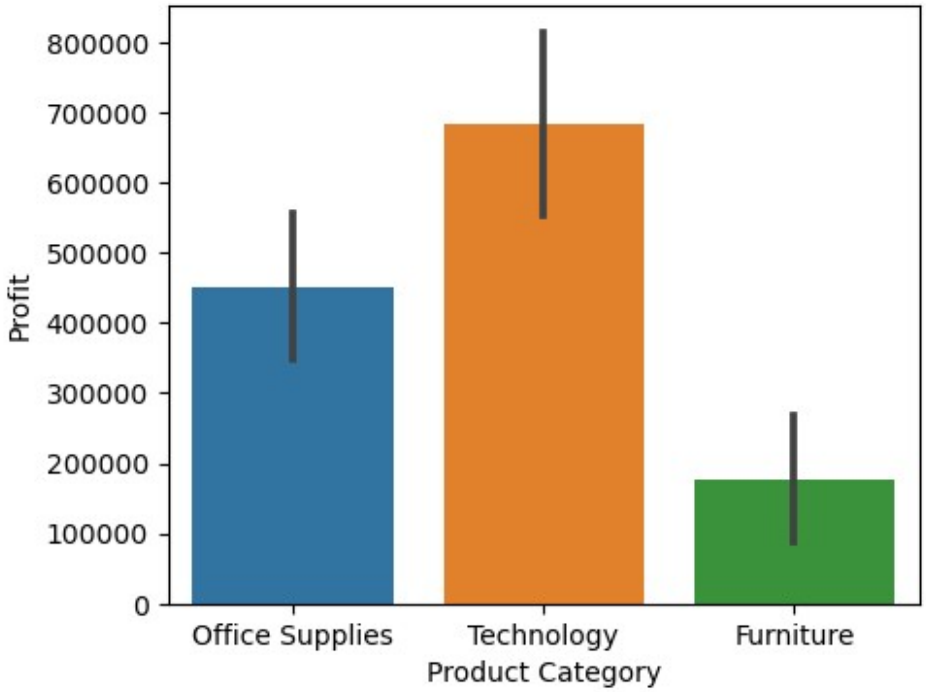
```
dataset["Order Year"].value_counts()
```

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

```
plt.figure(figsize = (5,4))
sns.countplot(x = "Order Year", data = dataset)
plt.show()
```




```
plt.figure(figsize = (5,4))
sns.barplot(x = "Product Category", y = "Profit", data = dataset,
estimator = "sum")
plt.show()
```



```
dataset["State or Province"].value_counts()
```

```
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
```

```
plt.figure(figsize = (5,4))
sns.barplot(x = "Product Category", y = "Product Base Margin", data =
dataset, estimator = "sum")
plt.show()
```