In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [2]:
```
df = pd.read_excel('Superstore_USA 1.xlsx')
```

In [3]:
```
df.head(5)
```

Out[3]:

	Row ID	Order Priority	Discount	Unit Price	Shipping Cost	Customer ID	Customer Name	Ship Mode	Customer Segment	Product Category	...	Regic
0	18606	Not Specified	0.01	2.88	0.50	2	Janice Fletcher	Regular Air	Corporate	Office Supplies	...	Centr
1	20847	High	0.01	2.84	0.93	3	Bonnie Potter	Express Air	Corporate	Office Supplies	...	We
2	23086	Not Specified	0.03	6.68	6.15	3	Bonnie Potter	Express Air	Corporate	Office Supplies	...	We
3	23087	Not Specified	0.01	5.68	3.60	3	Bonnie Potter	Regular Air	Corporate	Office Supplies	...	We
4	23088	Not Specified	0.00	205.99	2.50	3	Bonnie Potter	Express Air	Corporate	Technology	...	We

5 rows × 24 columns

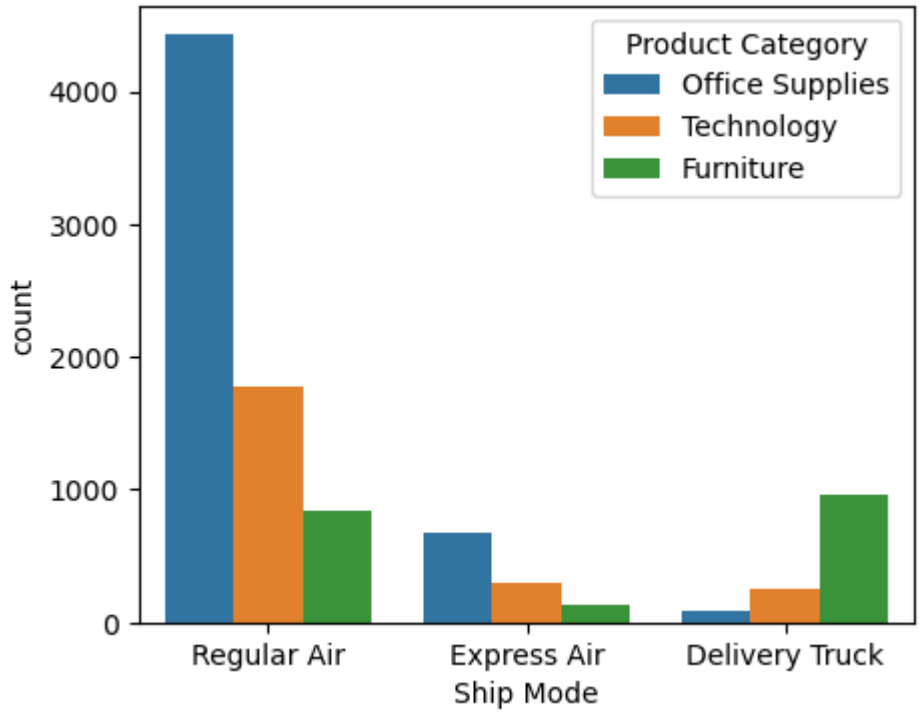
In [4]:
```
df.isnull().sum()
```

Out[4]:
```
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
```

In [5]:
```
df['Product Base Margin'].fillna(df['Product Base Margin'].mean(),inplace=True)
```

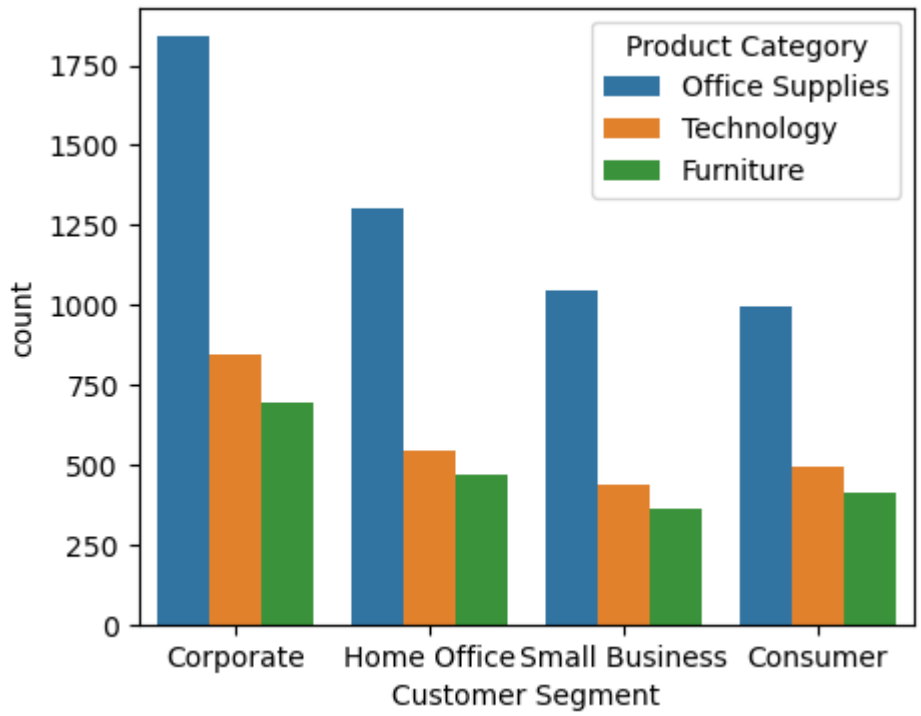
In [6]:
```
plt.figure(figsize=(5,4))
sns.countplot(x='Ship Mode',data=df,hue='Product Category')
plt.show
```

Out[6]:
```
<function matplotlib.pyplot.show(close=None, block=None)>
```



In [7]:
```
plt.figure(figsize=(5,4))
sns.countplot(x='Customer Segment',data=df,hue='Product Category')
plt.show
```

Out[7]:
```
<function matplotlib.pyplot.show(close=None, block=None)>
```



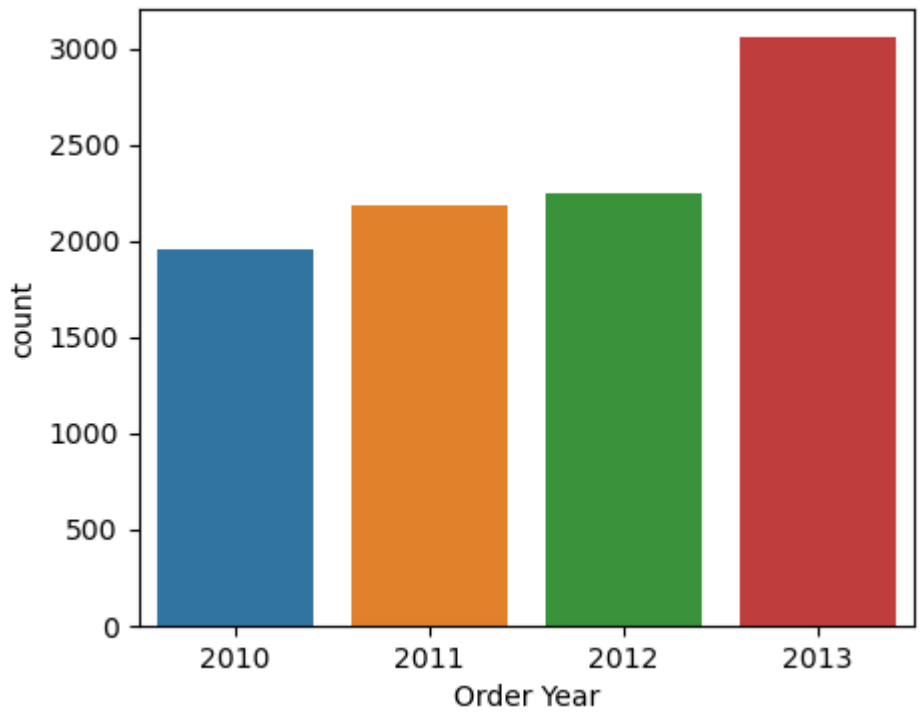
In [8]:
```
#sales over years
df['Order Year']=df['Order Date'].dt.year
```

In [9]:
```
df.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   9426 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
```

In [10]:
```
plt.figure(figsize=(5,4))
sns.countplot(x='Order Year',data=df)
plt.show
```

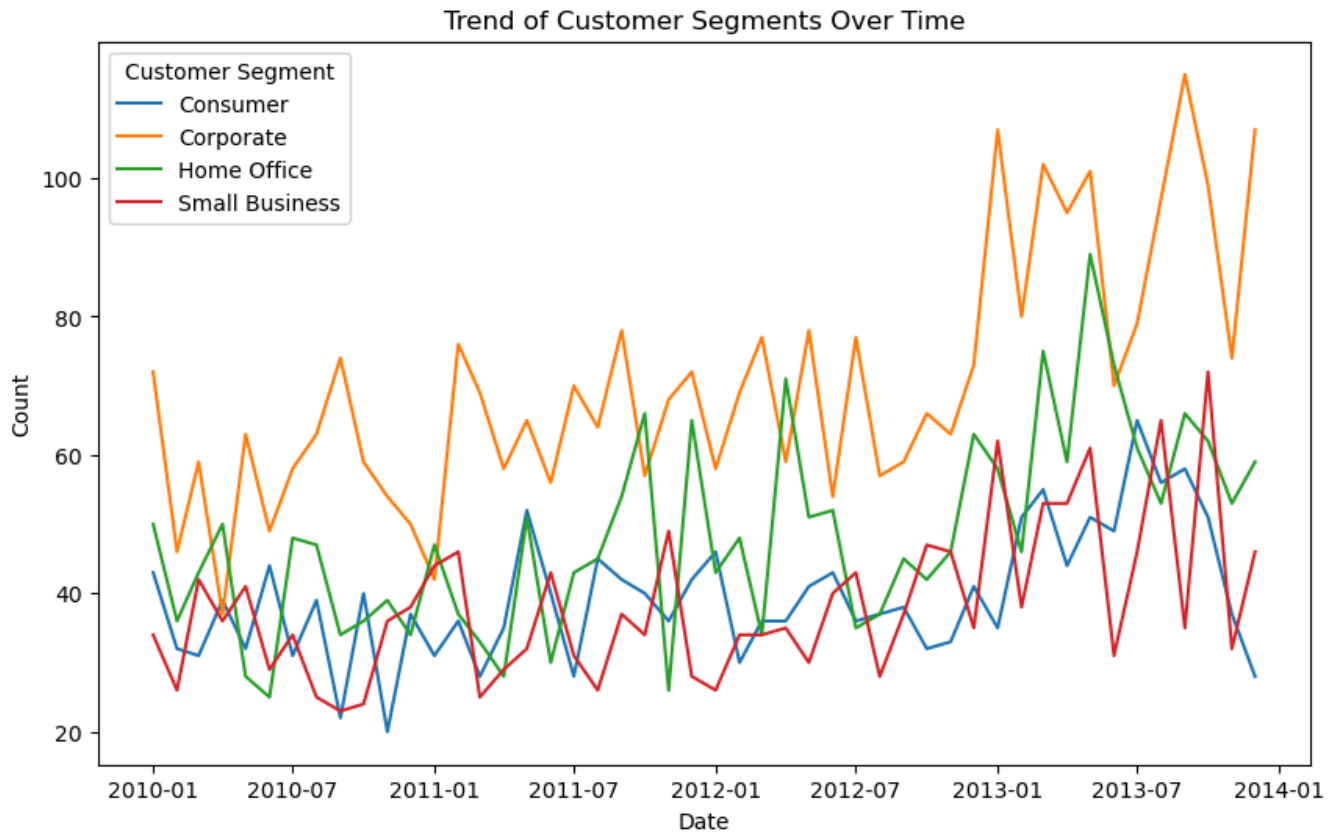
Out[10]:
```
<function matplotlib.pyplot.show(close=None, block=None)>
```



In [18]:
```
df['Order Date'] = pd.to_datetime(df['Order Date'])
trend= df.groupby([df['Order Date'].dt.to_period('M'), 'Customer Segment']).size().re
trend['Order Date'] =trend['Order Date'].dt.to_timestamp()
```

In [19]:
```
plt.figure(figsize=(10, 6))
sns.lineplot(data=trend, x='Order Date', y='Count', hue='Customer Segment')
```

```
plt.title('Trend of Customer Segments Over Time')
plt.xlabel('Date')
plt.ylabel('Count')
plt.show()
```



In [ ]: