In [239…

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

In [240…

```python
df=pd.read_csv('sales.csv')
```

In [241…

```python
#Check Null value
df.isnull().sum()
```

Out[241…

```
Order ID            80
Product             80
Quantity_Ordered    80
Price Each          80
Order Date          80
Purchase Address    80
dtype: int64
```

Data Cleaning

In [242…

```python
# Rename columns

df.rename(columns={'Order ID': 'Order_ID', 'Price Each': 'Price_Each'}, inplace=True)
```

In [243…

```python
# Make all columns  numeric

df['Order_ID']=pd.to_numeric(df['Order_ID'], errors='coerce')
df['Quantity_Ordered']=pd.to_numeric(df['Quantity_Ordered'], errors='coerce')
df['Price_Each']=pd.to_numeric(df['Price_Each'], errors='coerce')
df.head(5)
```

Out[243…

	Order_ID	Product	Quantity_Ordered	Price_Each	Order Date	Purchase Address
0	295665.0	Macbook Pro Laptop	1.0	1700.00	12/30/19 00:01	136 Church St, New York City, NY 10001
1	295666.0	LG Washing Machine	1.0	600.00	12/29/19 07:03	562 2nd St, New York City, NY 10001
2	295667.0	USB-C Charging Cable	1.0	11.95	12/12/19 18:21	277 Main St, New York City, NY 10001
3	295668.0	27in FHD Monitor	1.0	149.99	12/22/19 15:13	410 6th St, San Francisco, CA 94016
4	295669.0	USB-C Charging Cable	1.0	11.95	12/18/19 12:38	43 Hill St, Atlanta, GA 30301

In [244…

```python
# Solving Nan-Value Problems

df['Order_ID']=df['Order_ID'].interpolate(method='values')
df['Product']=df['Product'].fillna(method='ffill')
df['Quantity_Ordered']=df['Quantity_Ordered'].fillna(method='bfill')
df['Price_Each']=df['Price_Each'].interpolate(method='values')
df['Purchase Address']=df['Purchase Address'].fillna(method='bfill')
df['Order Date']=df['Order Date'].fillna(method='bfill')
df.isnull().sum()
```

Out[244…

```
Order_ID            0
Product             0
Quantity_Ordered    0
Price_Each          0
Order Date          0
Purchase Address    0
dtype: int64
```

In [245…

```python
# Make it integer

df['Order_ID']=df['Order_ID'].astype('int64')
df['Quantity_Ordered']=df['Quantity_Ordered'].astype('int64')
df['Price_Each']=df['Price_Each'].astype('int64')
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 25117 entries, 0 to 25116
Data columns (total 6 columns):
 #   Column            Non-Null Count  Dtype 
---  ------            --------------  ----- 
 0   Order_ID          25117 non-null  int64 
 1   Product           25117 non-null  object
 2   Quantity_Ordered  25117 non-null  int64 
 3   Price_Each        25117 non-null  int64 
 4   Order Date        25117 non-null  object
 5   Purchase Address  25117 non-null  object
dtypes: int64(3), object(3)
memory usage: 1.1+ MB
```

Visualisation

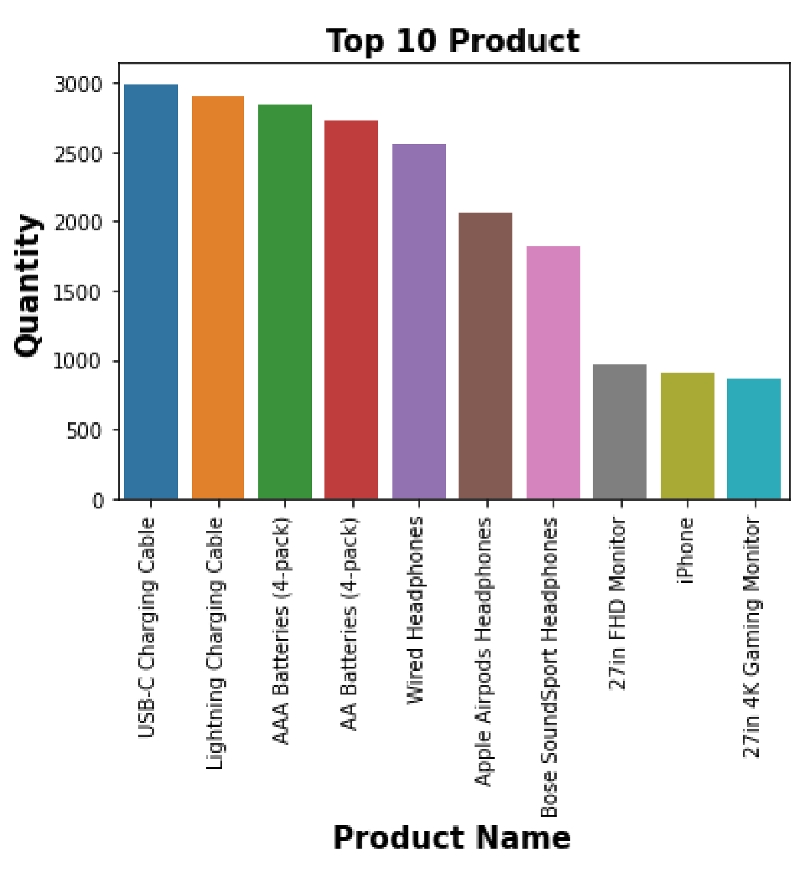
In [246…

```python
# Barplot of Top 10 Product

sns.barplot(x=df.Product.value_counts().head(10).keys(), y=df.Product.value_counts().head(10))
plt.xticks(rotation=90)
plt.title('Top 10 Product', fontsize=15, fontweight='bold')
plt.xlabel('Product Name', fontsize=15, color='black', fontweight='bold')
plt.ylabel('Quantity', fontsize=15, color='black', fontweight='bold')
```

Out[246…

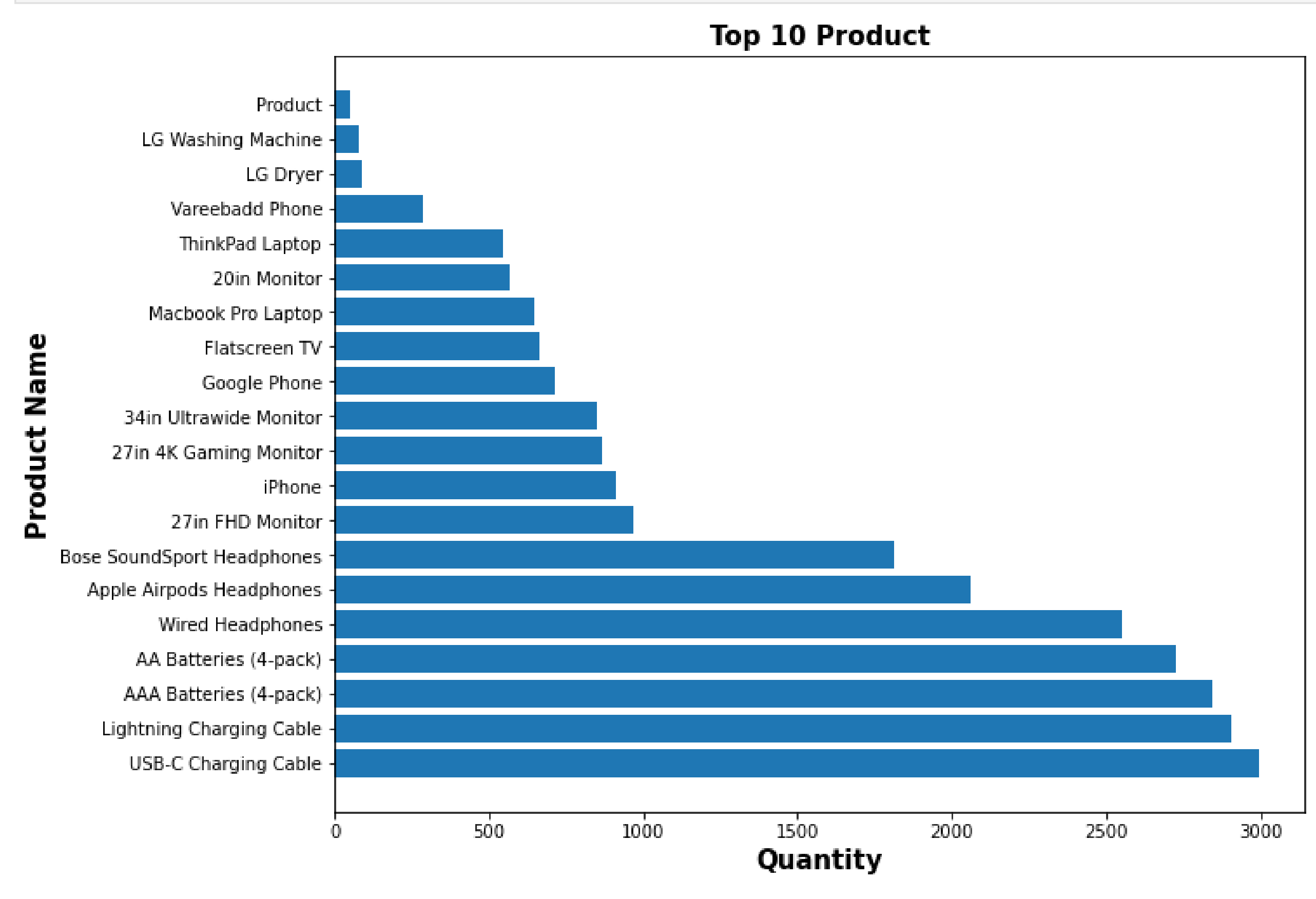
```
Text(0, 0.5, 'Quantity')
```



In [259…

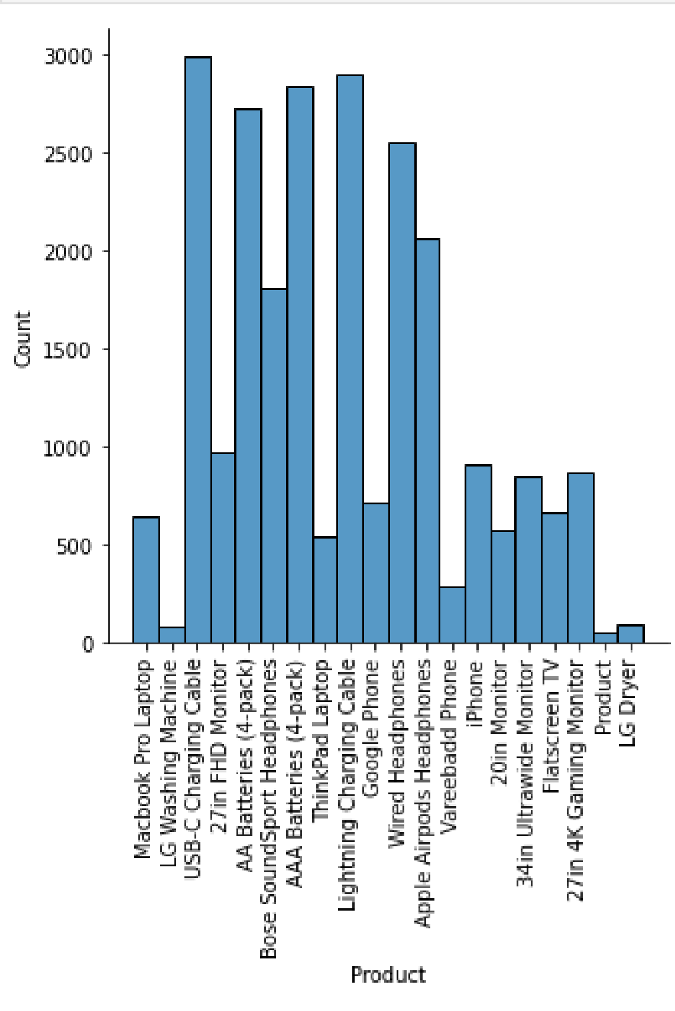
```python
# Top 10 Product in Horizontal way

plt.figure(figsize=(10,8))
plt.barh(df['Product'].value_counts().keys(), df['Product'].value_counts())
plt.title('Top 10 Product', fontsize=15, fontweight='bold')
plt.xlabel('Quantity', fontsize=15, color='black', fontweight='bold')
plt.ylabel('Product Name', fontsize=15, color='black', fontweight='bold')
plt.show()
```



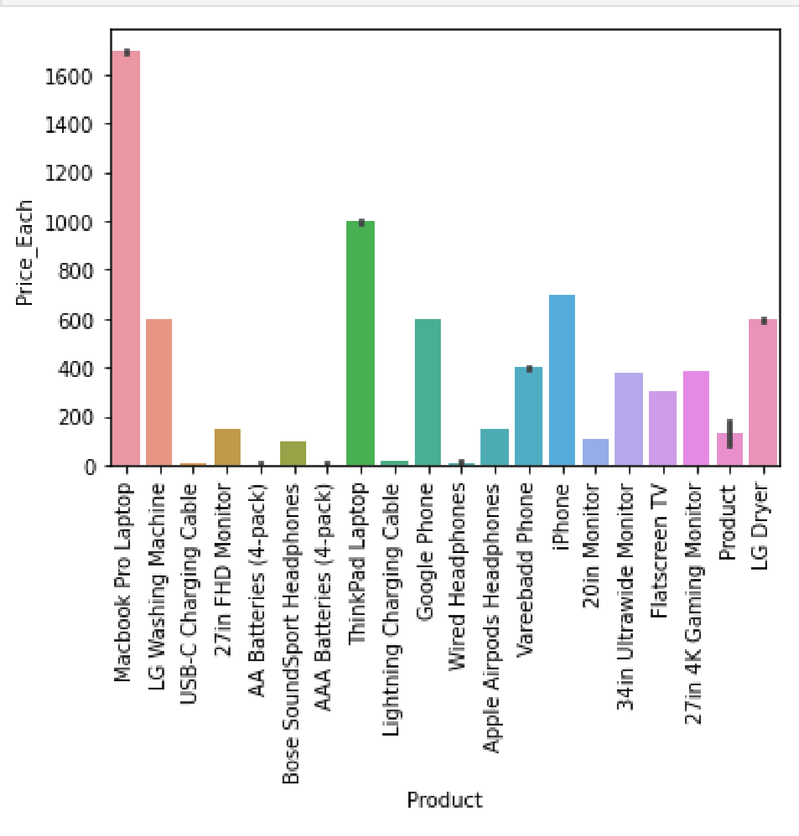
In [248…

```python
sns.displot(df['Product'])
plt.xticks(rotation=90)
plt.show()
```



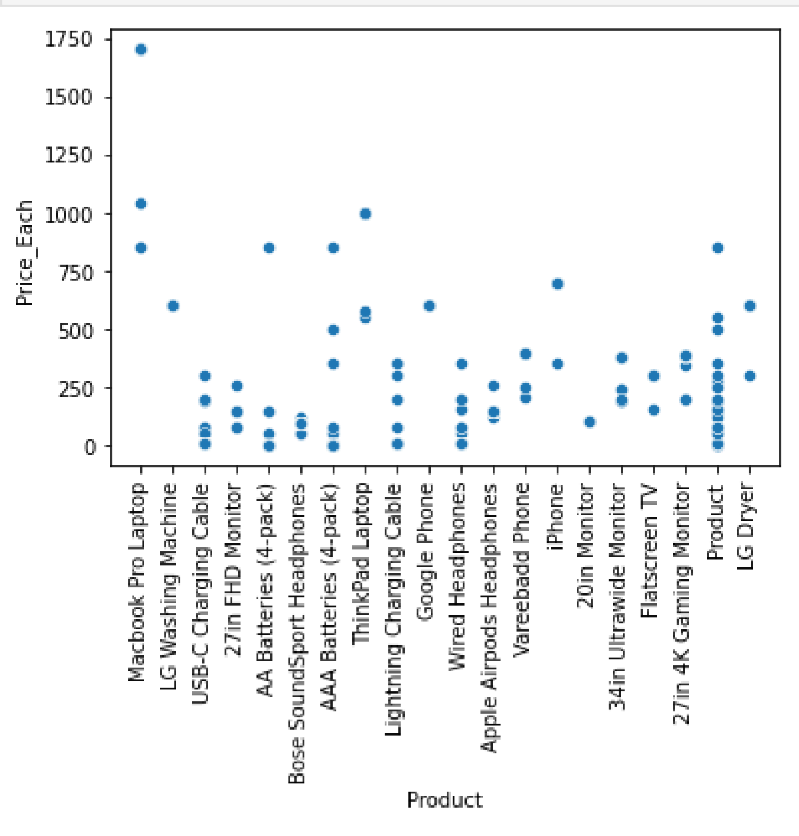
In [249…

```python
sns.barplot(x='Product', y='Price_Each', data=df)
plt.xticks(rotation=90)
plt.show()
```



In [250…

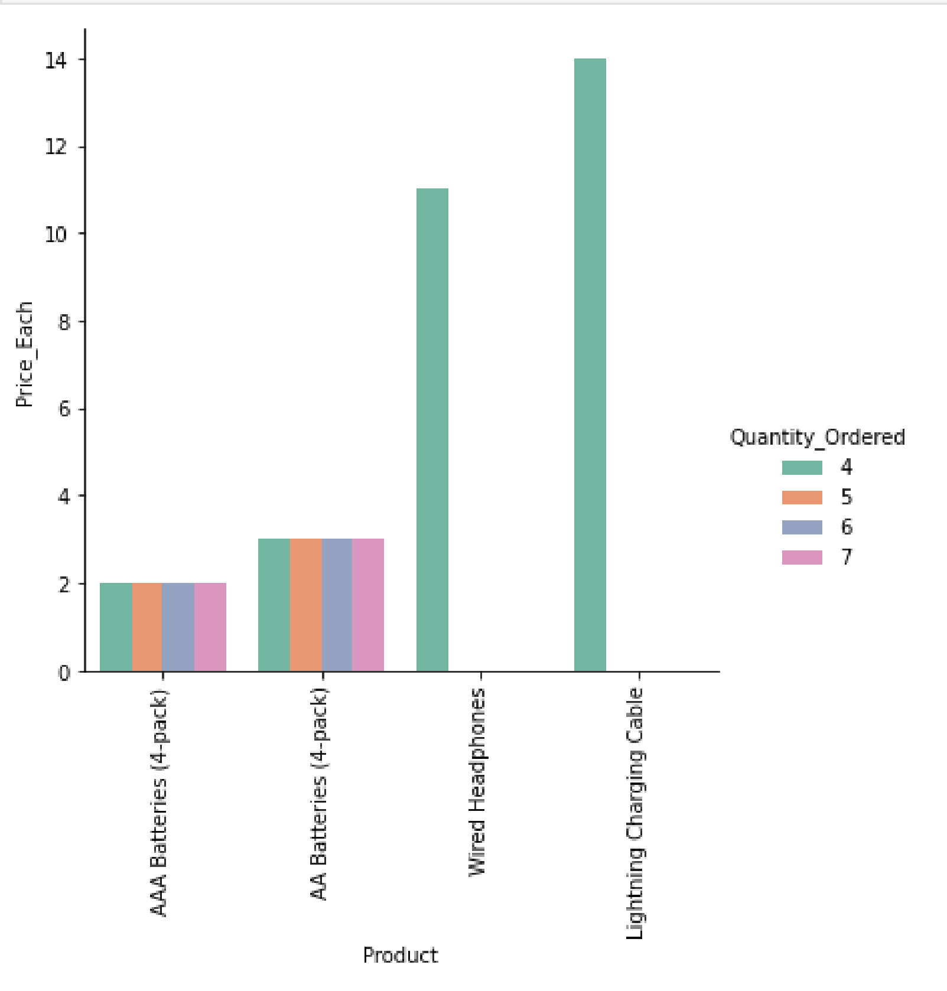
```python
sns.scatterplot(x='Product', y='Price_Each', data=df)
plt.xticks(rotation=90)
plt.show()
```



In [251…

```python
selected=df[df.Quantity_Ordered.isin([4,5,6,7])]

sns.catplot(x='Product', y='Price_Each', data=selected, kind='bar', hue='Quantity_Ordered', palette='Set2')
plt.xticks(rotation=90)
plt.show()
```



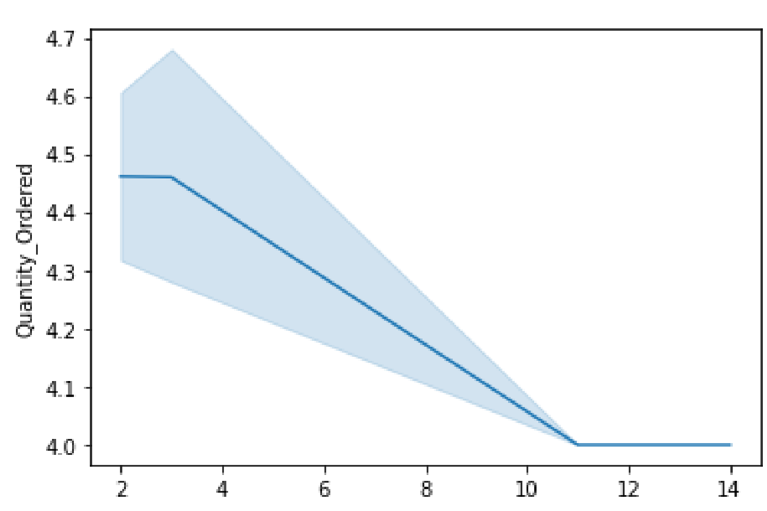
In [256…

```python
# We can see that with the increase in prices, the amount of purchases has also decreased.

dff=df[df.Quantity_Ordered.isin([4,5,6,7])]
sns.lineplot(x='Price_Each', y='Quantity_Ordered', data=dff,
        estimator='mean')
```

Out[256…

```
<AxesSubplot:xlabel='Price_Each', ylabel='Quantity_Ordered'>
```



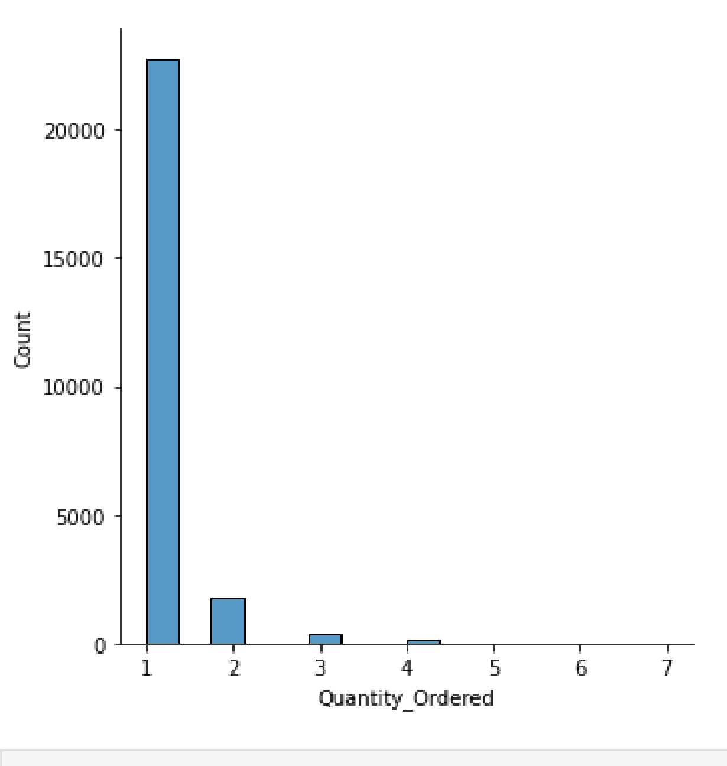
In [253…

```python
# Count of Quantity

sns.displot(df.Quantity_Ordered)
```

Out[253…

```
<seaborn.axisgrid.FacetGrid at 0x1da43cf6790>
```



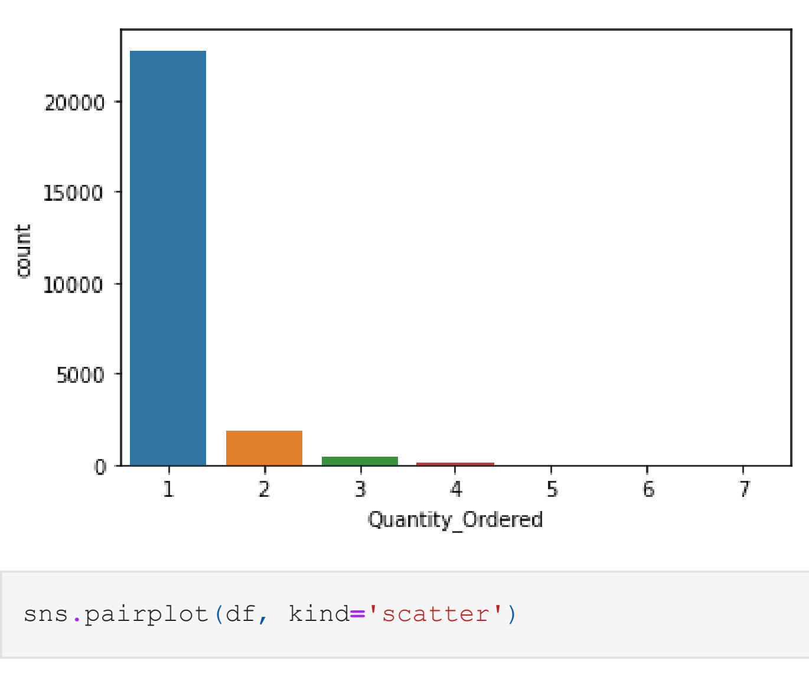
In [254…

```python
# Or,

sns.countplot(x='Quantity_Ordered', data=df,
            order=df.Quantity_Ordered.value_counts().index)
```

Out[254…

```
<AxesSubplot:xlabel='Quantity_Ordered', ylabel='count'>
```

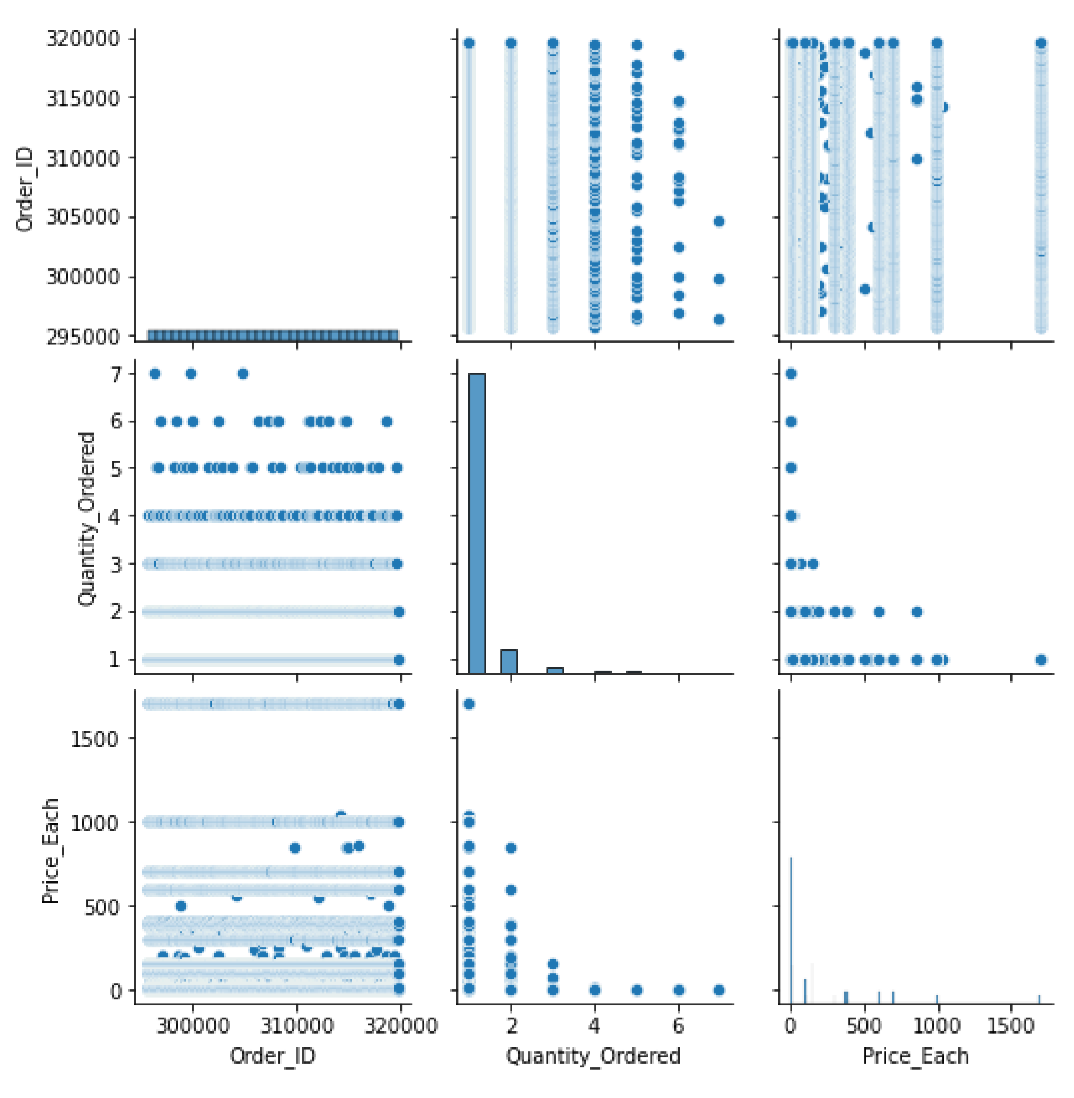


In [258…

```python
sns.pairplot(df, kind='scatter')
```

Out[258…

```
<seaborn.axisgrid.PairGrid at 0x1da445aeca0>
```



In [ ]:

In [ ]:

In [ ]: