THE AMAZON DATA ANALYSIS

PRESENTED BY SONIA

THIS AMAZON SALES REPORT IS A GREAT EXAMPLE OF HOW DATA ANALYSIS CAN BE USED TO OPTIMIZE E-COMMERCE BUSINESSES. BY ANALYZING SALES DATA BY COUNTRY, STATE, MARKET, CATEGORY, AND SHIP MODE, BUSINESSES CAN IDENTIFY TRENDS AND MAKE DATA-DRIVEN DECISIONS TO IMPROVE SALES AND PROFITABILITY.

```python
In [4]: import numpy as np
        import pandas as pd
        import matplotlib.pyplot as plt
        import seaborn as sns
```

```python
In [2]: dataset = pd.read_csv("D:\sonia\skill lync\Amazon Sales data.csv")
```

```python
In [3]: dataset
```

Out[3]:

	Region	Country	Item Type	Sales Channel	Order Priority	Order Date	Order ID	Ship Date	Units Sold	Unit Price	Unit Cost	Total Revenue	Total Cost	P
0	Australia and Oceania	Tuvalu	Baby Food	Offline	H	5/28/2010	669165933	6/27/2010	9925	255.28	159.42	2533654.00	1582243.50	95141
1	Central America and the Caribbean	Grenada	Cereal	Online	C	8/22/2012	963881480	9/15/2012	2804	205.70	117.11	576782.80	328376.44	24840
2	Europe	Russia	Office Supplies	Offline	L	05-02-2014	341417157	05-08-2014	1779	651.21	524.96	1158502.59	933903.84	22459
3	Sub-Saharan Africa	Sao Tome and Principe	Fruits	Online	C	6/20/2014	514321792	07-05-2014	8102	9.33	6.92	75591.66	56065.84	1952
4	Sub-Saharan	Rwanda	Office	Offline	L	02-01-	115456712	02-06-	5062	651.21	524.96	3296425.02	2657347.52	63907

```python
In [4]: dataset.head(2)
```

Out[4]:

	Region	Country	Item Type	Sales Channel	Order Priority	Order Date	Order ID	Ship Date	Units Sold	Unit Price	Unit Cost	Total Revenue	Total Cost	Total Profit	Order Date2	Ship Date2
0	Australia and Oceania	Tuvalu	Baby Food	Offline	H	5/28/2010	669165933	6/27/2010	9925	255.28	159.42	2533654.0	1582243.50	951410.50	28-05-2010	27-06-2010
1	Central America and the Caribbean	Grenada	Cereal	Online	C	8/22/2012	963881480	9/15/2012	2804	205.70	117.11	576782.8	328376.44	248406.36	22-08-2012	15-09-2012

```python
In [5]: dataset.shape
```

Out[5]: (100, 16)

```python
In [6]: dataset.isnull().sum()
```

Out[6]:
```
Region          0
Country         0
Item Type       0
Sales Channel   0
```

```
8]:  dataset['Sales Channel'].value_counts()

8]:  Sales Channel
     Offline    50
     Online     50
     Name: count, dtype: int64


9]:  dataset['Order Priority'].unique()

9]:  array(['H', 'C', 'L', 'M'], dtype=object)


0]:  dataset['Order Priority'].value_counts()

0]:  Order Priority
     H    30
     L    27
     C    22
     M    21
     Name: count, dtype: int64
```
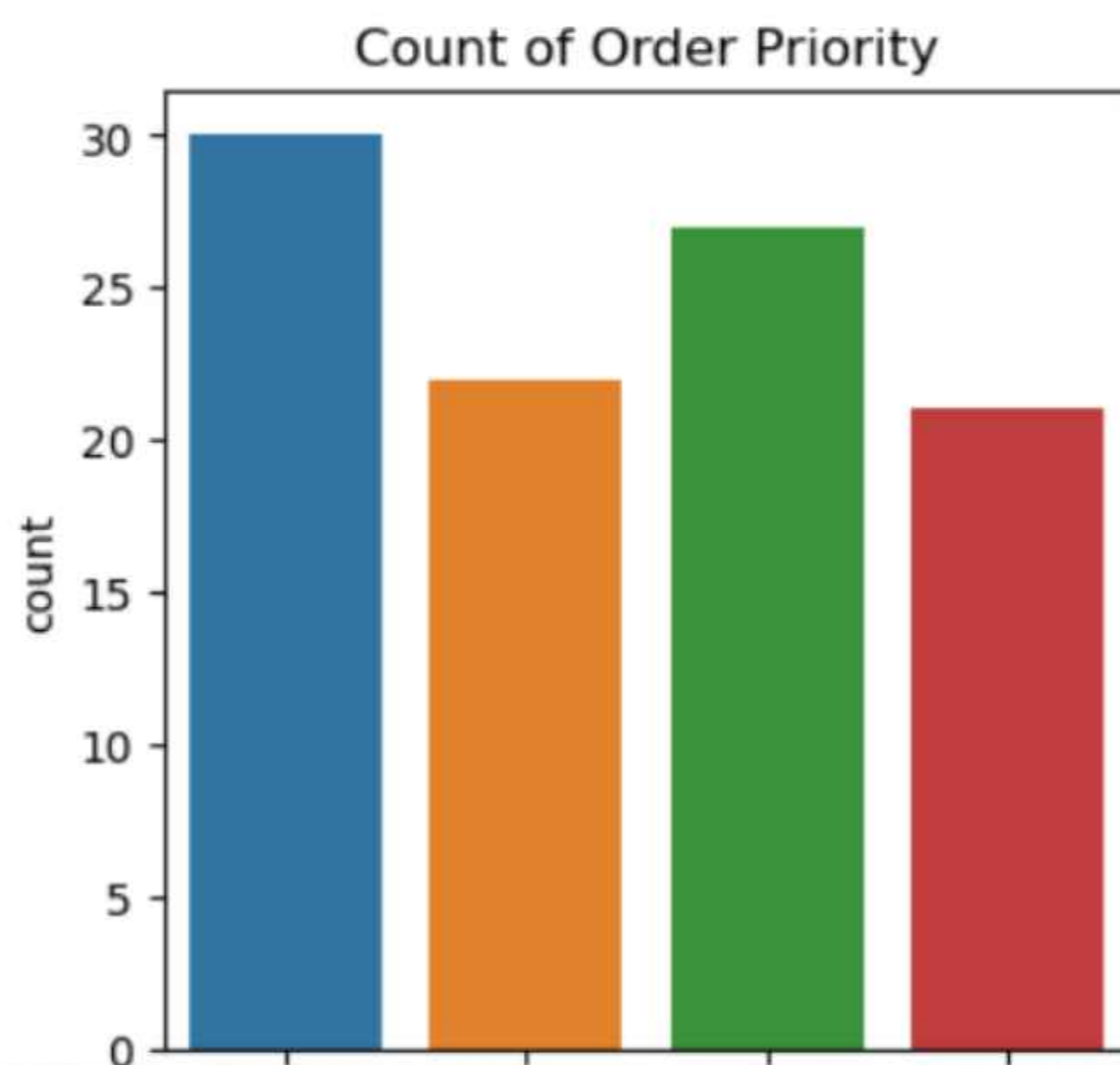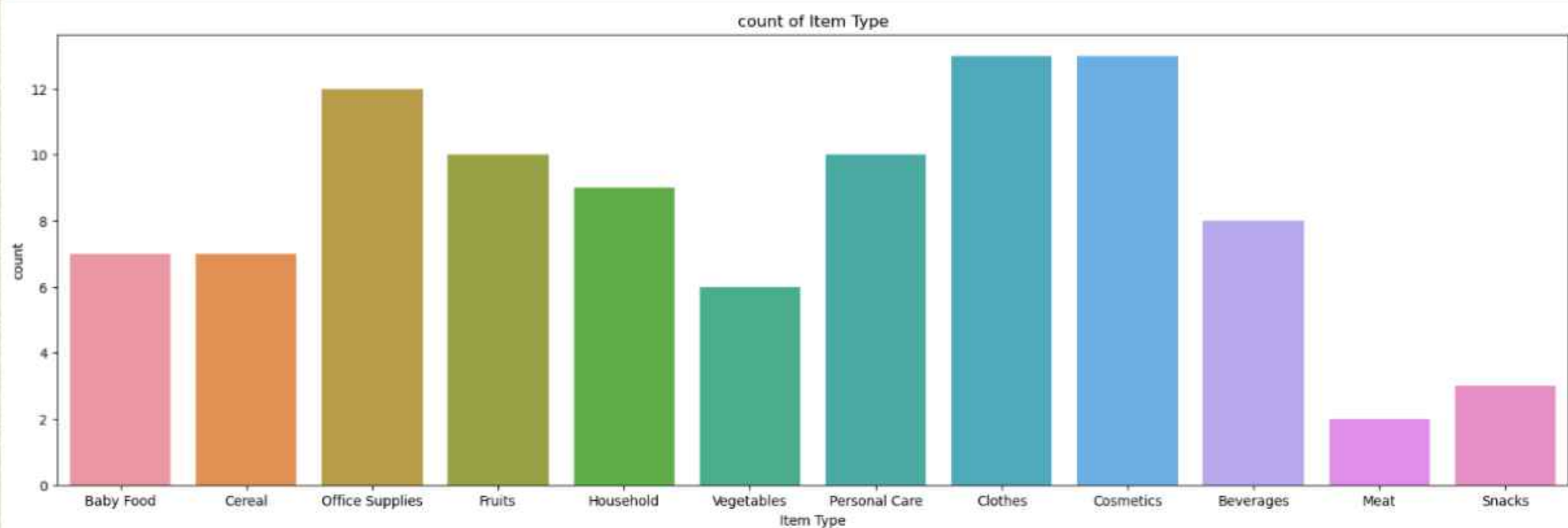
```
In [12]:  # FROM THE BELOW CHART WE CAN SEE THAT MOST OF ORDER PRIORITY
          plt.figure(figsize=(4,4))
          sns.countplot(x="Order Priority",data =dataset)
          plt.title("Count of Order Priority")
          plt.savefig("count of Order Priority.jpg")
          plt.show()
```
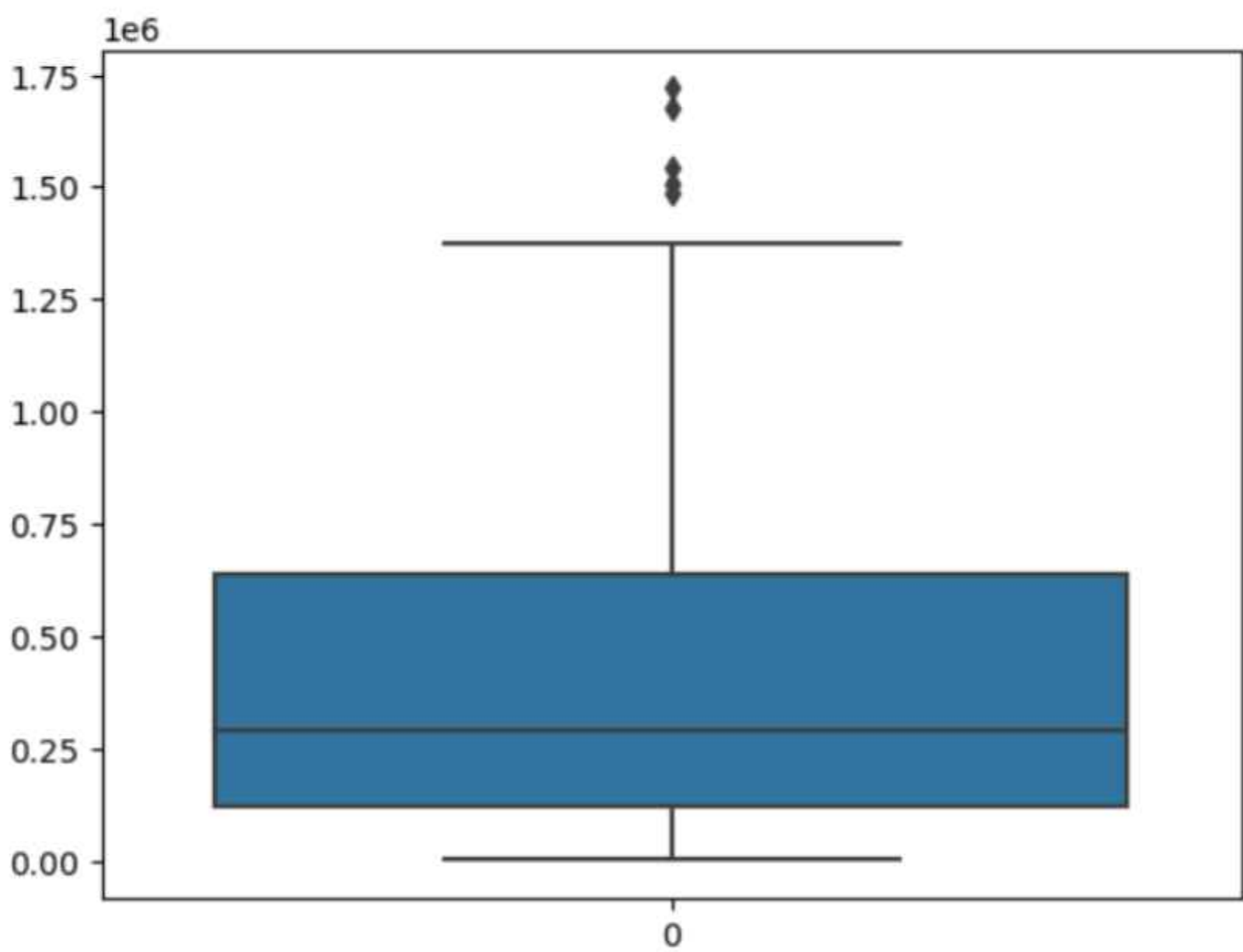

Count of Order Priority

```
#FROM THE  ELOW CHART WE CAN SEE THAT HOW MUCH ITEM TYPE CONTRIBUTE TO REVENUE GENERATE
plt.figure(figsize=(20,6))
sns.countplot(x="Item Type",data =dataset)
plt.title("count of Item Type")
plt.savefig("count of Item Type.jpg")
plt.show()
```
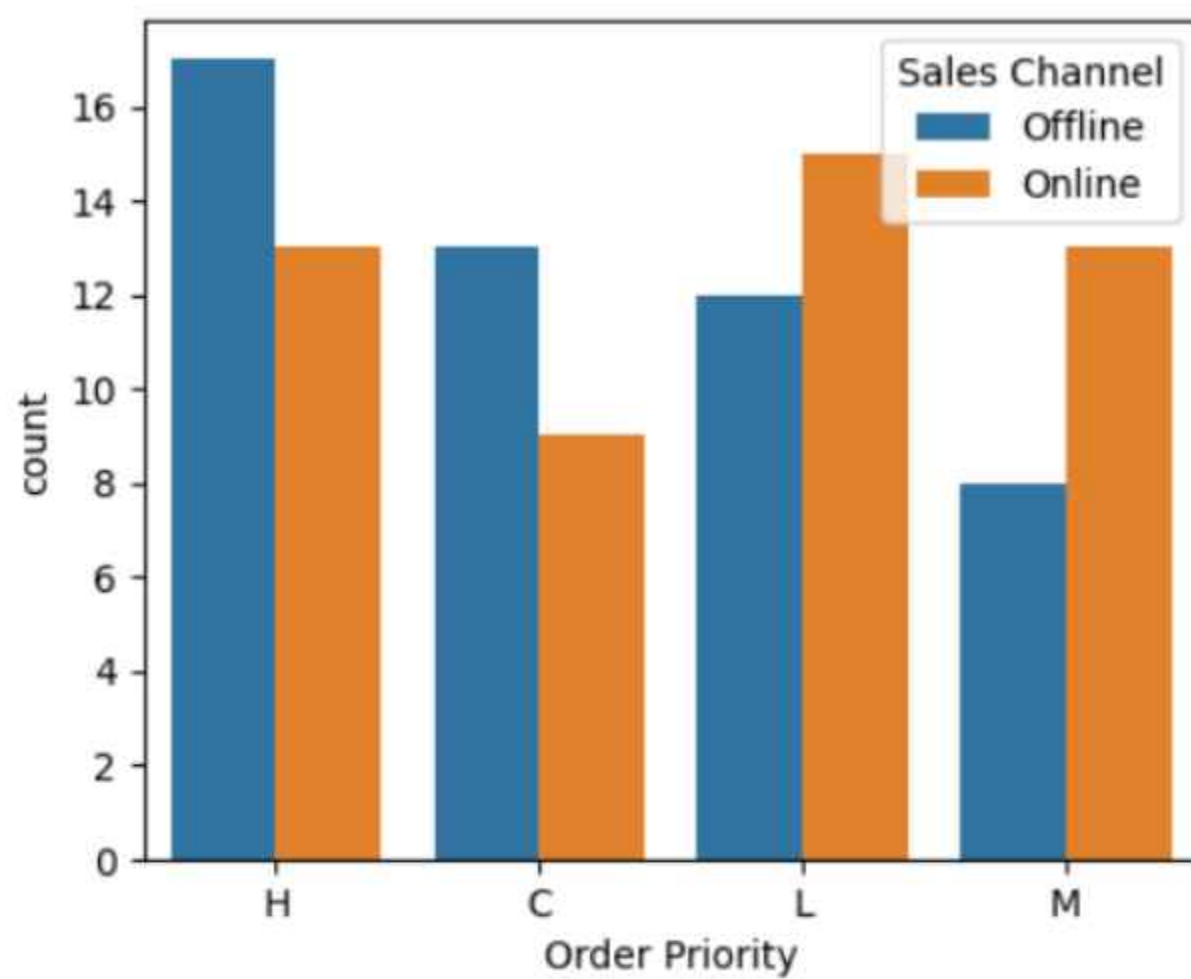

count of Item Type

```
In [63]:   # FROM THE BELOW CHART WE CAN ABOARSED ABOUT THE TOTAL PROFIT BY BOXPLOT
           sns.boxplot(dataset['Total Profit'])

Out[63]:   <Axes: >
```

```
: # FROM THE BELOW CHART WE CAN SEE THAT H CONTRIBUTE MORE OFFLINE BUT L CONTRI MORE IN ONLINE
plt.figure(figsize=(5,4))
sns.countplot(x='Order Priority',data= dataset,hue="Sales Channel")
plt.show()
```



```
In [65]: # Top 5 most sold Item type

         dataset['Item Type'].value_counts()[:5]

Out[65]: Item Type
         Clothes             13
         Cosmetics           13
         Office Supplies      12
         Fruits              10
         Personal Care        10
         Name: count, dtype: int64
```

```
: # FROM THE BELOW CHART WE CAN SEE THAT H CONTRIBUTE MORE OFFLINE BUT L CONTRI MORE IN ONLINE
plt.figure(figsize=(5,4))
sns.countplot(x='Order Priority',data= dataset,hue="Sales Channel")
plt.show()
```

HERE ARE SOME SPECIFIC INSIGHTS THAT CAN BE GLEANED FROM THIS REPORT

1. THERE ARE ARE 100 ROW AND 16 COLUMNS.

2 THE UNITED STATES IS AMAZON'S LARGEST MARKET, ACCOUNTING FOR 42.46 OF TOTAL SALES.

3. THE TECHNOLOGY CATEGORY IS AMAZON'S BEST-SELLING CATEGORY, ACCOUNTING FOR 4.7 MILLION IN SALES.

4. STANDARD CLASS IS THE MOST POPULAR SHIPPING MODE, ACCOUNTING FOR 7.58 MILLION IN SALES.

5. THERE ARE OFFLINE AND ONLINE CHANNEL

6. THERE ARE 4 UNIQUE ORDER PRIORITY H ,C L,M.

7. THERE IS H IS HIGHEST ORDER PRIORTY AND OFFLINE CHANNEL CONTRIBUTION.

8. CLOTHES AND COSMETICS BOTH ARE HIGHEST IN CONTRIBUTE OF REVENUE.

9. SUB-SHARAN AFRICA IS THE HIGHEST REGION OF OFFLINE CHANNEL AND TOTAL PROFIT AND ALSO REVENUE.

10. THA GAMBIA IS THE TOP SELLING COUNTRY.

11. THE HONDURAS COUNTRY GENERATE MAXIMUM REVENUE.

BY ANALYZING THIS DATA, AMAZON CAN

FOCUS MARKETING EFFORTS ON COUNTRIES AND STATES WITH HIGH GROWTH POTENTIAL.

DEVELOP NEW PRODUCTS IN CATEGORIES THAT ARE IN HIGH DEMAND.

OFFER FASTER OR CHEAPER SHIPPING OPTIONS TO COMPETE MORE EFFECTIVELY.

OVER ALL CONCLUSION WE HAVE TO INCREASE OUR SALES IN NORTH AMERICA, LESS TOTAL PROFIT MARGIN IS FROM FRUIT ,

WE HAVE INCREASE AND TOP MOST PROFIT MARGIN IS FROM COSMETIC AND FROM HOUSEHOLD

WE OBSERVE THAT ORDER PRIORITY BY M IS THE OFFICE SUPPLER IS HIGHER THAN OTHER

IN SUB SAHARAN AFRICA REGION OFFILE SALE IS GREATER THAN ONLINE

ORDER PRIOINERITY BY H OFFLINE IS GREATER THAN ONLINE

THE HONDURAS COUNTRY GENERATE MAXIMUM REVENUE

THE SUB SAHARAN AFRICA GENERATE MAXIMUM REVENUE AND TOTAL PROFIT

THE GAMBIA IS THE TOP SELLING COUNTRY

THANK YOU