# Amazon Sales Report

In [7]:
```
## Import libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
```

In [13]:
```
df=pd.read_csv("Amazon Sale Report.csv")
```

In [15]:
```
df.head(3)
```

Out[15]:

| | index | Order ID | Date | Status | Fulfilment | Sales Channel | ship-service-level | Category | Size | Cou St |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 0 | 405-8078784-5731545 | 04-30-22 | Cancelled | Merchant | Amazon.in | Standard | T-shirt | S | Or |
| **1** | 1 | 171-9198151-1101146 | 04-30-22 | Shipped - Delivered to Buyer | Merchant | Amazon.in | Standard | Shirt | 3XL | Ship |
| **2** | 2 | 404-0687676-7273146 | 04-30-22 | Shipped | Amazon | Amazon.in | Expedited | Shirt | XL | Ship |

3 rows × 21 columns

In [19]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 128976 entries, 0 to 128975
Data columns (total 21 columns):
 #   Column              Non-Null Count   Dtype  
---  ------              --------------   -----  
 0   index               128976 non-null  int64  
 1   Order ID            128976 non-null  object 
 2   Date                128976 non-null  object 
 3   Status              128976 non-null  object 
 4   Fulfilment          128976 non-null  object 
 5   Sales Channel       128976 non-null  object 
 6   ship-service-level  128976 non-null  object 
 7   Category            128976 non-null  object 
 8   Size                128976 non-null  object 
 9   Courier Status      128976 non-null  object 
 10  Qty                 128976 non-null  int64  
 11  currency            121176 non-null  object 
 12  Amount              121176 non-null  float64
 13  ship-city           128941 non-null  object 
 14  ship-state          128941 non-null  object 
 15  ship-postal-code    128941 non-null  float64
 16  ship-country        128941 non-null  object 
 17  B2B                 128976 non-null  bool   
 18  fulfilled-by        39263 non-null   object 
 19  New                 0 non-null       float64
 20  PendingS            0 non-null       float64
dtypes: bool(1), float64(4), int64(2), object(14)
memory usage: 19.8+ MB
```

In [21]:
```
df.shape
```

Out[21]:
```
(128976, 21)
```

In [25]:
```
df.drop(["New","PendingS"],axis=1,inplace=True)
```

In [27]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 128976 entries, 0 to 128975
Data columns (total 19 columns):
 #   Column              Non-Null Count   Dtype  
---  ------              --------------   -----  
 0   index               128976 non-null  int64  
 1   Order ID            128976 non-null  object 
 2   Date                128976 non-null  object 
 3   Status              128976 non-null  object 
 4   Fulfilment          128976 non-null  object 
 5   Sales Channel       128976 non-null  object 
 6   ship-service-level  128976 non-null  object 
 7   Category            128976 non-null  object 
 8   Size                128976 non-null  object 
 9   Courier Status      128976 non-null  object 
 10  Qty                 128976 non-null  int64  
 11  currency            121176 non-null  object 
 12  Amount              121176 non-null  float64
 13  ship-city           128941 non-null  object 
 14  ship-state          128941 non-null  object 
 15  ship-postal-code    128941 non-null  float64
 16  ship-country        128941 non-null  object 
 17  B2B                 128976 non-null  bool   
 18  fulfilled-by        39263 non-null   object 
dtypes: bool(1), float64(2), int64(2), object(14)
memory usage: 17.8+ MB
```

In [31]:
```
## checking the null values
df.isnull().sum()
```

Out[31]:
```
index                     0
Order ID                  0
Date                      0
Status                    0
Fulfilment                0
Sales Channel             0
ship-service-level        0
Category                  0
Size                      0
Courier Status            0
Qty                       0
currency               7800
Amount                 7800
ship-city                35
ship-state               35
ship-postal-code         35
ship-country             35
B2B                       0
fulfilled-by          89713
dtype: int64
```

In [33]:
```
## drop the null values
df.dropna(inplace=True)
```

In [35]:
```
## checking the null values
df.isnull().sum()
```

Out[35]:
```
index                0
Order ID             0
Date                 0
Status               0
Fulfilment           0
Sales Channel        0
ship-service-level   0
Category             0
Size                 0
Courier Status       0
Qty                  0
currency             0
Amount               0
ship-city            0
ship-state           0
ship-postal-code     0
ship-country         0
B2B                  0
fulfilled-by         0
dtype: int64
```

In [37]:
```
df.columns
```

Out[37]:
```
Index(['index', 'Order ID', 'Date', 'Status', 'Fulfilment', 'Sales Channel',
       'ship-service-level', 'Category', 'Size', 'Courier Status', 'Qty',
       'currency', 'Amount', 'ship-city', 'ship-state', 'ship-postal-code',
       'ship-country', 'B2B', 'fulfilled-by'],
      dtype='object')
```

In [39]:
```
## change the datatype
df['ship-postal-code']=df['ship-postal-code'].astype('int')
```

In [41]:
```
## chceking the whether the data type chnage or not
df["ship-postal-code"].dtype
```

Out[41]:
```
dtype('int32')
```

In [43]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 37514 entries, 0 to 128892
Data columns (total 19 columns):
 #   Column              Non-Null Count  Dtype  
---  ------              --------------  -----  
 0   index               37514 non-null  int64  
 1   Order ID            37514 non-null  object 
 2   Date                37514 non-null  object 
 3   Status              37514 non-null  object 
 4   Fulfilment          37514 non-null  object 
 5   Sales Channel       37514 non-null  object 
 6   ship-service-level  37514 non-null  object 
 7   Category            37514 non-null  object 
 8   Size                37514 non-null  object 
 9   Courier Status      37514 non-null  object 
 10  Qty                 37514 non-null  int64  
 11  currency            37514 non-null  object 
 12  Amount              37514 non-null  float64
 13  ship-city           37514 non-null  object 
 14  ship-state          37514 non-null  object 
 15  ship-postal-code    37514 non-null  int32  
 16  ship-country        37514 non-null  object 
 17  B2B                 37514 non-null  bool   
 18  fulfilled-by        37514 non-null  object 
dtypes: bool(1), float64(1), int32(1), int64(2), object(14)
memory usage: 5.3+ MB
```

In [45]:
```
df["Date"]=pd.to_datetime(df["Date"])
```

In [47]:
```
df["Date"].dtype
```

Out[47]:
```
dtype('<M8[ns]')
```

In [49]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 37514 entries, 0 to 128892
Data columns (total 19 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   index               37514 non-null  int64
 1   Order ID            37514 non-null  object
 2   Date                37514 non-null  datetime64[ns]
 3   Status              37514 non-null  object
 4   Fulfilment          37514 non-null  object
 5   Sales Channel       37514 non-null  object
 6   ship-service-level  37514 non-null  object
 7   Category            37514 non-null  object
 8   Size                37514 non-null  object
 9   Courier Status      37514 non-null  object
 10  Qty                 37514 non-null  int64
 11  currency            37514 non-null  object
 12  Amount              37514 non-null  float64
 13  ship-city           37514 non-null  object
 14  ship-state          37514 non-null  object
 15  ship-postal-code    37514 non-null  int32
 16  ship-country        37514 non-null  object
 17  B2B                 37514 non-null  bool
 18  fulfilled-by        37514 non-null  object
dtypes: bool(1), datetime64[ns](1), float64(1), int32(1), int64(2), object(13)
memory usage: 5.3+ MB
```

In [55]:
```
## rename Columns
df=df.rename(columns={'Qty':'Quantity'})
```

In [57]:
```
df.head(3)
```

Out[57]:

| | index | Order ID | Date | Status | Fulfilment | Sales Channel | ship-service-level | Category | Size | Cou St |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 0 | 405-8078784-5731545 | 2022-04-30 | Cancelled | Merchant | Amazon.in | Standard | T-shirt | S | Or |
| **1** | 1 | 171-9198151-1101146 | 2022-04-30 | Shipped - Delivered to Buyer | Merchant | Amazon.in | Standard | Shirt | 3XL | Ship |
| **3** | 3 | 403-9615377-8133951 | 2022-04-30 | Cancelled | Merchant | Amazon.in | Standard | Blazzer | L | Or |

In [61]:
```
## describe return dseecriptive Data like mean median mode-- only for numericals
df.describe()
```

Out[61]:

| | index | Quantity | Amount | ship-postal-code |
|---|---|---|---|---|
| **count** | 37514.000000 | 37514.000000 | 37514.000000 | 37514.000000 |
| **mean** | 60953.809858 | 0.867383 | 646.553960 | 463291.552754 |
| **std** | 36844.853039 | 0.354160 | 279.952414 | 194550.425637 |
| **min** | 0.000000 | 0.000000 | 0.000000 | 110001.000000 |
| **25%** | 27235.250000 | 1.000000 | 458.000000 | 370465.000000 |
| **50%** | 63470.500000 | 1.000000 | 629.000000 | 500019.000000 |
| **75%** | 91790.750000 | 1.000000 | 771.000000 | 600042.000000 |
| **max** | 128891.000000 | 5.000000 | 5495.000000 | 989898.000000 |

In [63]:
```
df.describe(include='object')
```

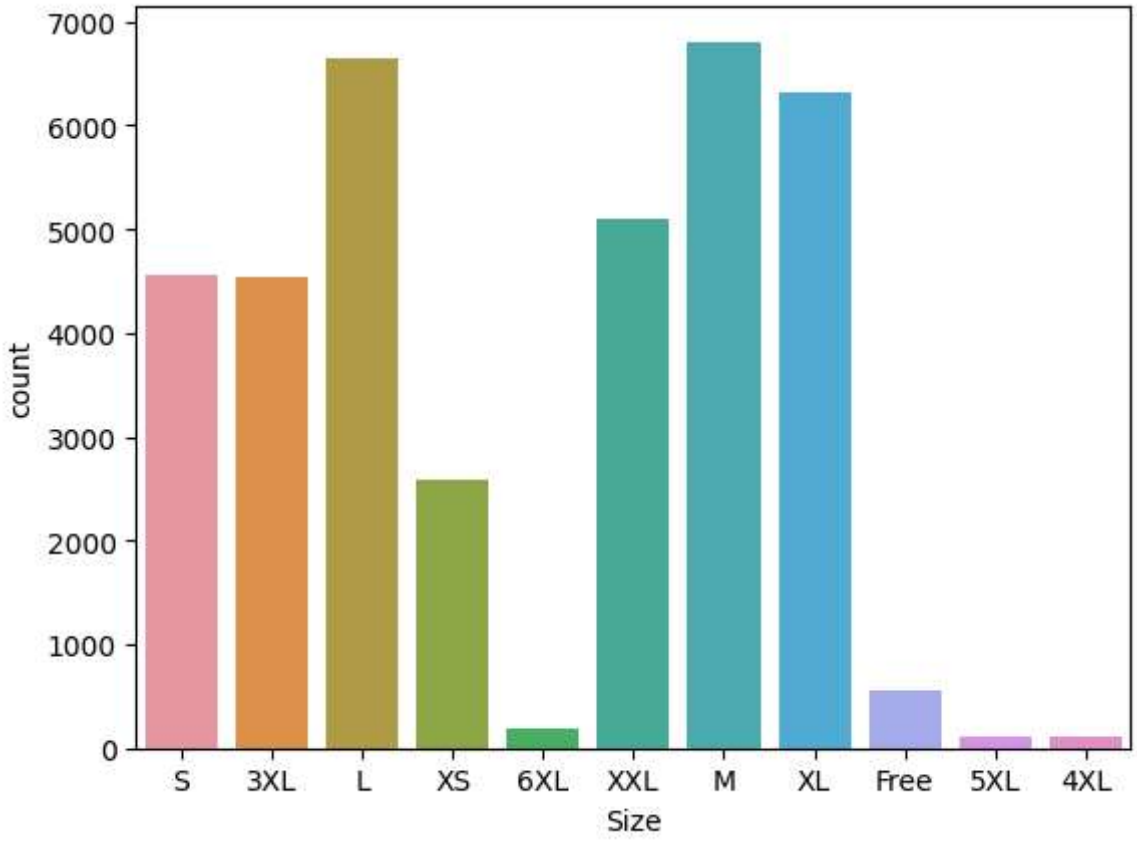
Out[63]:

| | Order ID | Status | Fulfilment | Sales Channel | ship-service-level | Category | Size | Courier Status | cu |
|---|---|---|---|---|---|---|---|---|---|
| **count** | 37514 | 37514 | 37514 | 37514 | 37514 | 37514 | 37514 | 37514 | |
| **unique** | 34664 | 11 | 1 | 1 | 1 | 8 | 11 | 3 | |
| **top** | 171-5057375-2831560 | Shipped - Delivered to Buyer | Merchant | Amazon.in | Standard | T-shirt | M | Shipped | |
| **freq** | 12 | 28741 | 37514 | 37514 | 37514 | 14062 | 6806 | 31859 | |

In [69]:
```
### Exploratory Data Anlaysis
df.columns
```

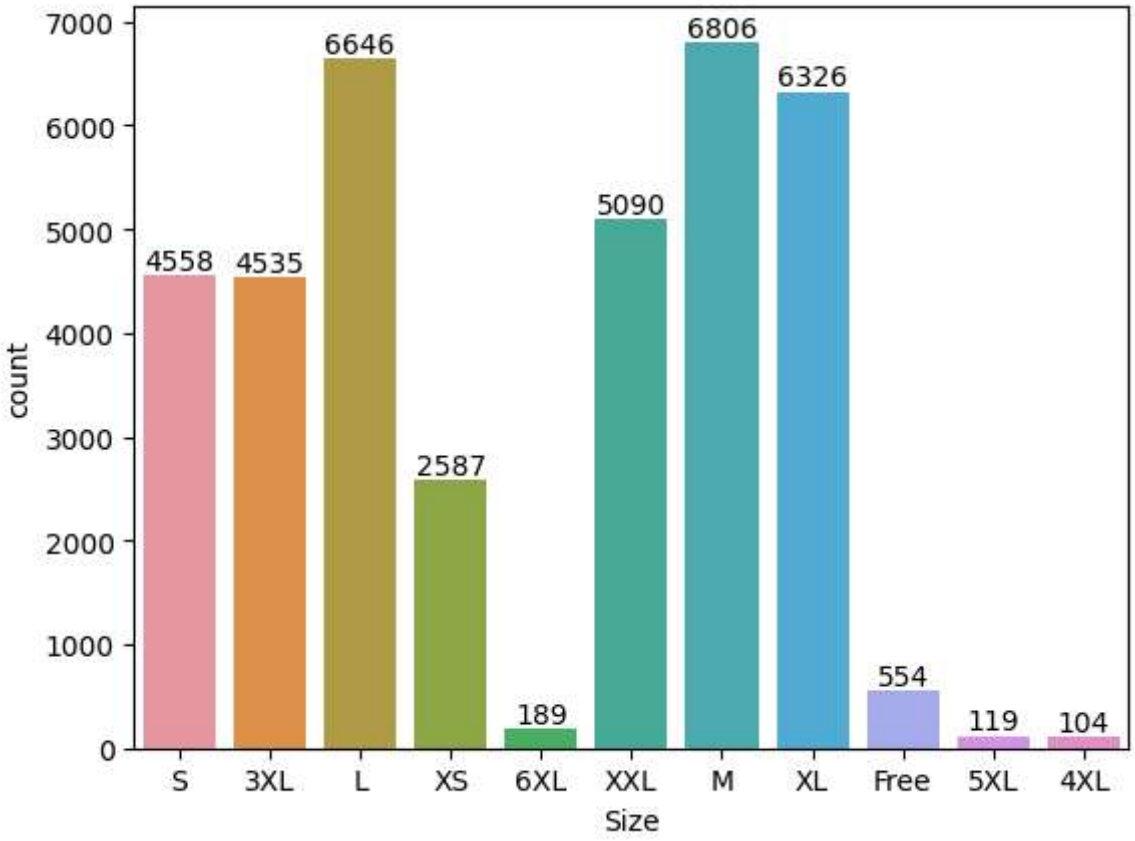
Out[69]:
```
Index(['index', 'Order ID', 'Date', 'Status', 'Fulfilment', 'Sales Channel',
       'ship-service-level', 'Category', 'Size', 'Courier Status', 'Quantity',
       'currency', 'Amount', 'ship-city', 'ship-state', 'ship-postal-code',
       'ship-country', 'B2B', 'fulfilled-by'],
      dtype='object')
```

In [73]:
```
ax=sns.countplot(x="Size",data=df)
```




In [77]:
```
ax=sns.countplot(x="Size",data=df)

for bars in ax.containers:
    ax.bar_label(bars)
```




Note= Most Of the People Buys- M- Size

## Groupby By

- This function used to group data based on one or more columns in the DataFrame

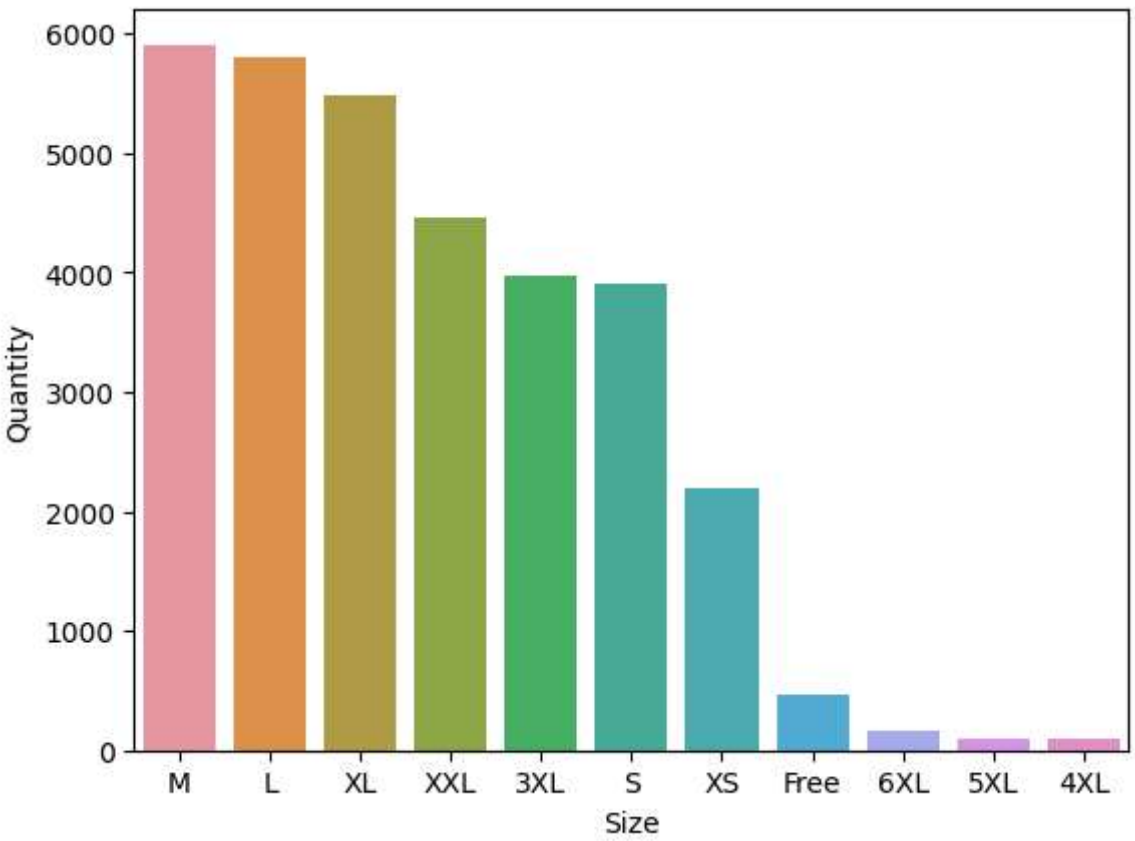
In [84]:
```
df.groupby(["Size"],as_index=False)["Quantity"].sum().sort_values(by="Quantity", as
```

Out[84]:

| | Size | Quantity |
|---|---|---|
| **6** | M | 5905 |
| **5** | L | 5795 |
| **8** | XL | 5481 |
| **10** | XXL | 4465 |
| **0** | 3XL | 3972 |
| **7** | S | 3896 |
| **9** | XS | 2191 |
| **4** | Free | 467 |
| **3** | 6XL | 170 |
| **2** | 5XL | 104 |
| **1** | 4XL | 93 |

In [86]:
```
S_Qty=df.groupby(["Size"],as_index=False)["Quantity"].sum().sort_values(by="Quantit
sns.barplot(x='Size', y='Quantity',data=S_Qty)
```

Out[86]: <AxesSubplot: xlabel='Size', ylabel='Quantity'>

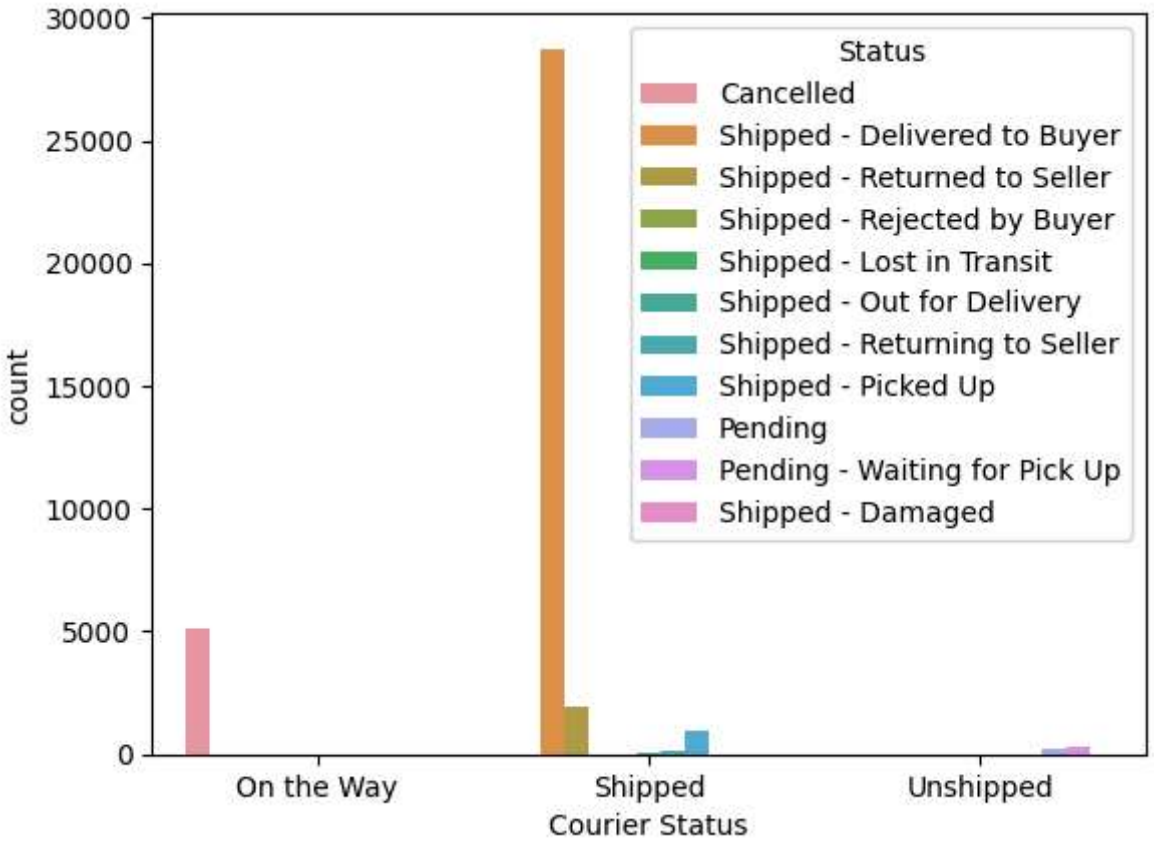



Note= Most Of the People Buys- M- Size

## Courier Status

In [96]:
```
sns.countplot(data=df, x="Courier Status", hue="Status")
```

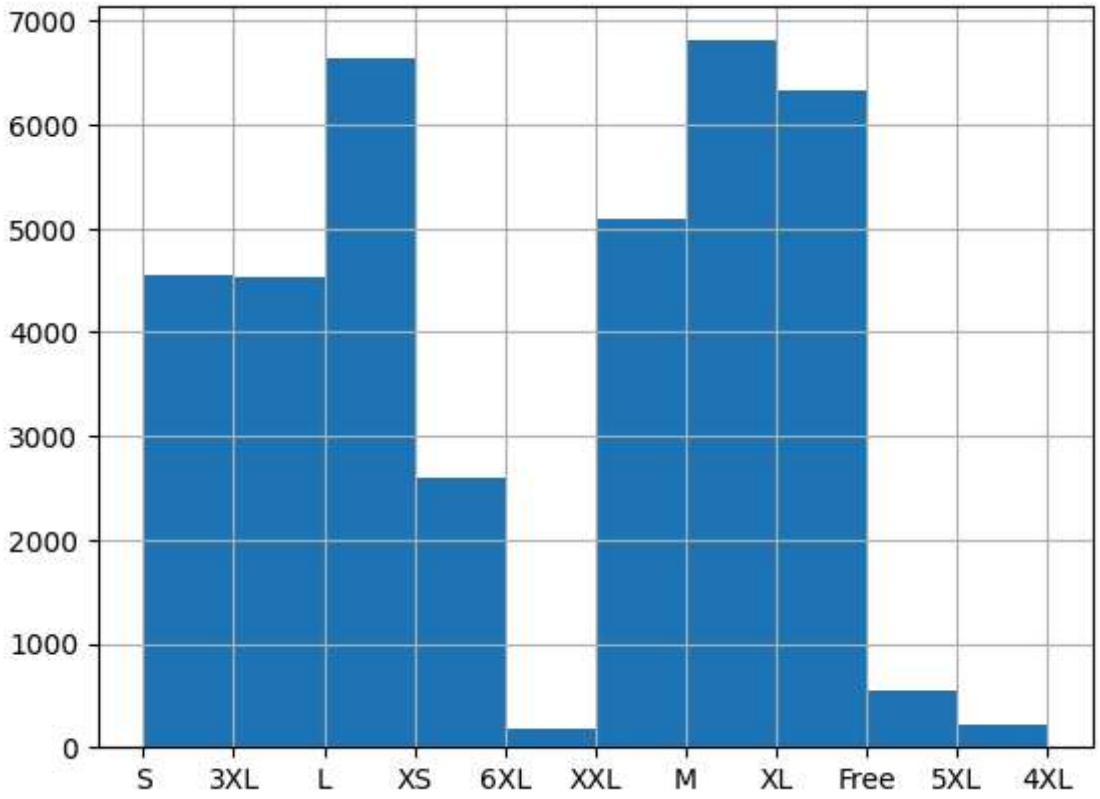
Out[96]:
```
<AxesSubplot: xlabel='Courier Status', ylabel='count'>
```




## From above Graph the Majority of the orders are shipped throgh the courier

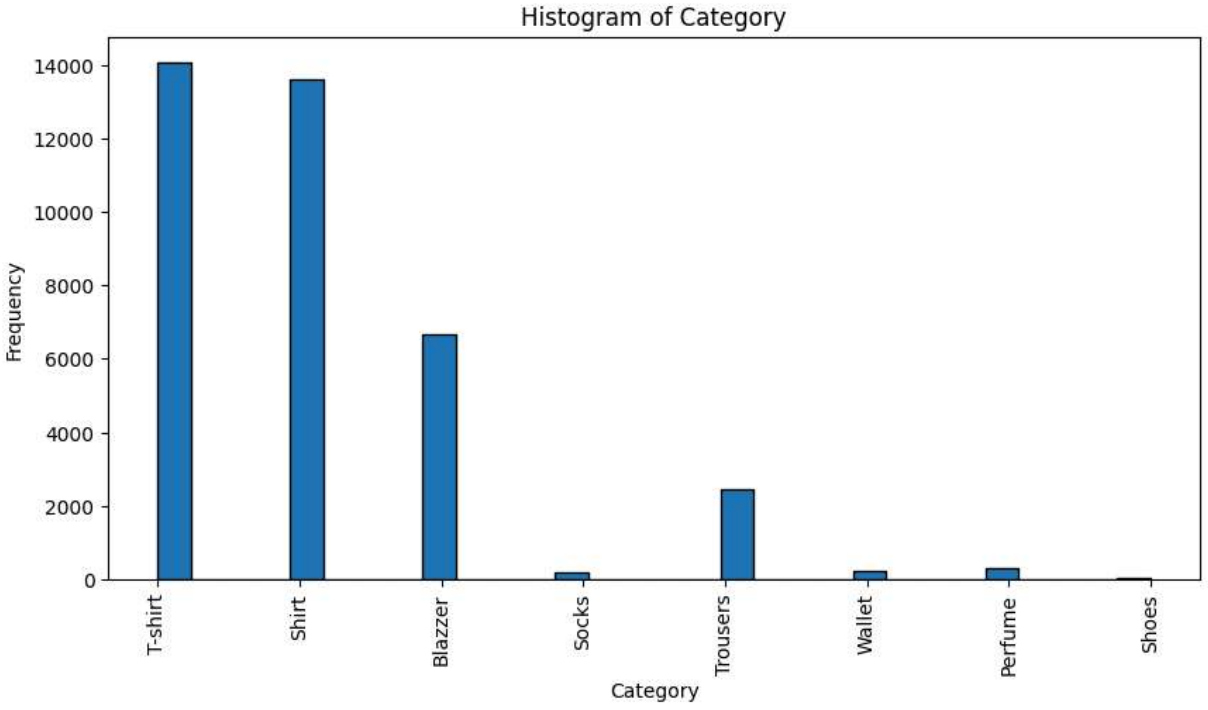
In [99]:
```
#### Histogram
df["Size"].hist()
```

Out[99]:
```
<AxesSubplot: >
```




In [103…

```
plt.figure(figsize=(10, 5))
plt.hist(column_data, bins=30, edgecolor='black')
plt.xticks(rotation=90)
plt.title("Histogram of Category")
plt.xlabel("Category")
plt.ylabel("Frequency")
plt.show()
```

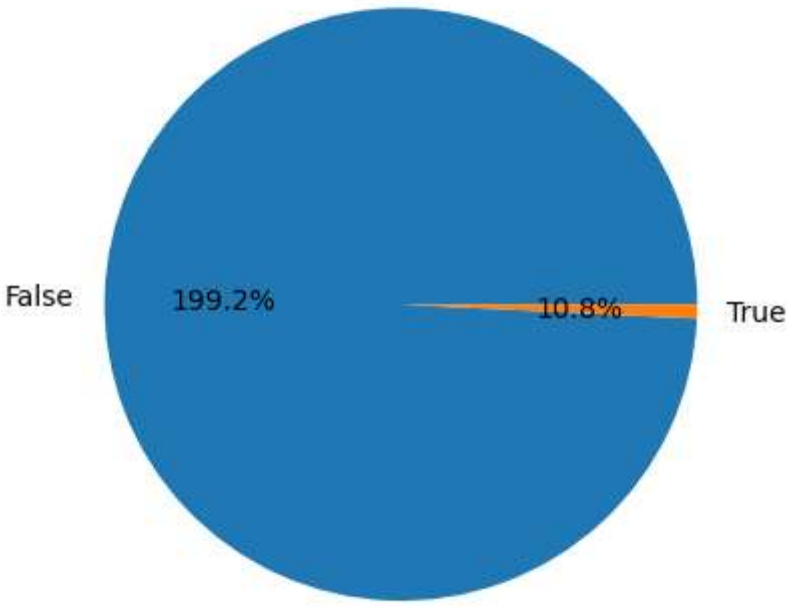


## Note: From above graph you can see that most of the buyers are T-shirt

In [112…

```
#### Checking B2B data  by using pie chart
B2B_check=df["B2B"].value_counts()

#plot tha pie chart
plt.pie(B2B_check,labels=B2B_check.index, autopct='1%.1f%%')
plt.show()
```

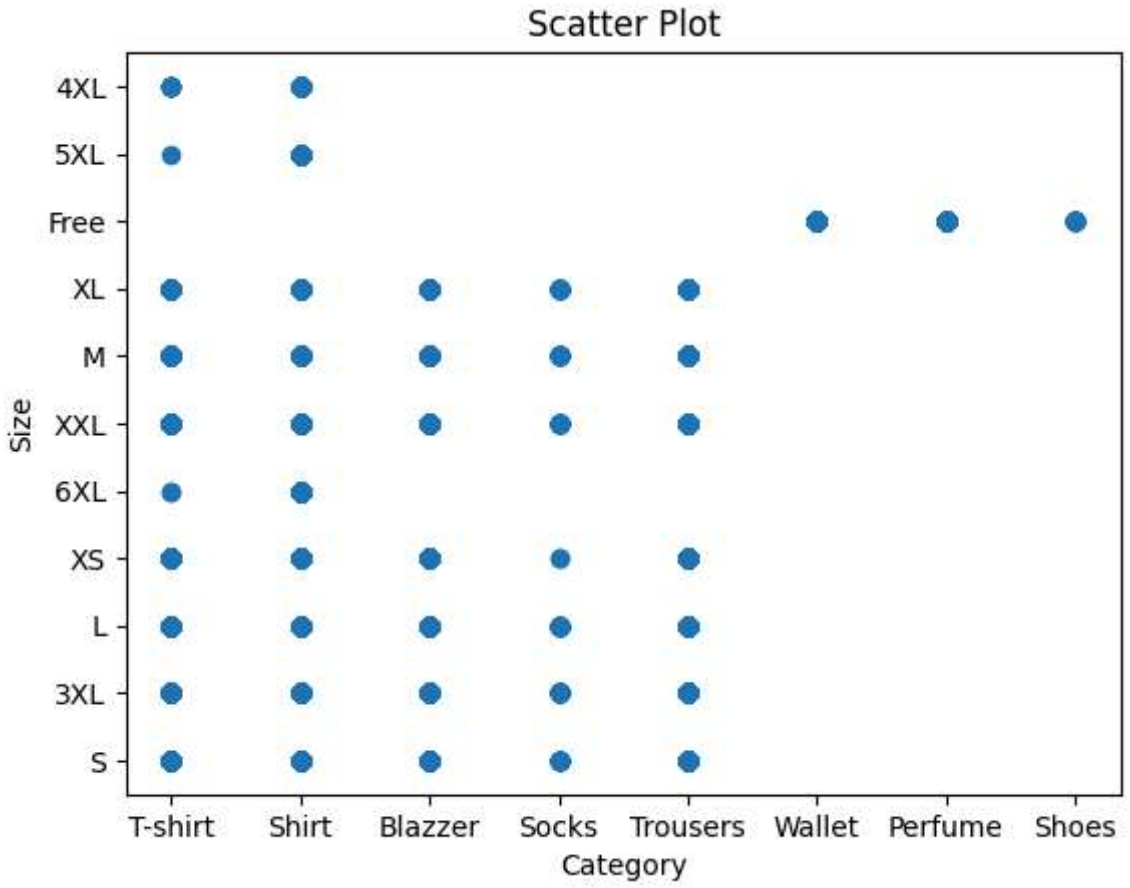



## From aobe chart we can see that maximum i.e. 99.3% of buyers are retailers and 0.8% are B2B Buyers

In [118…

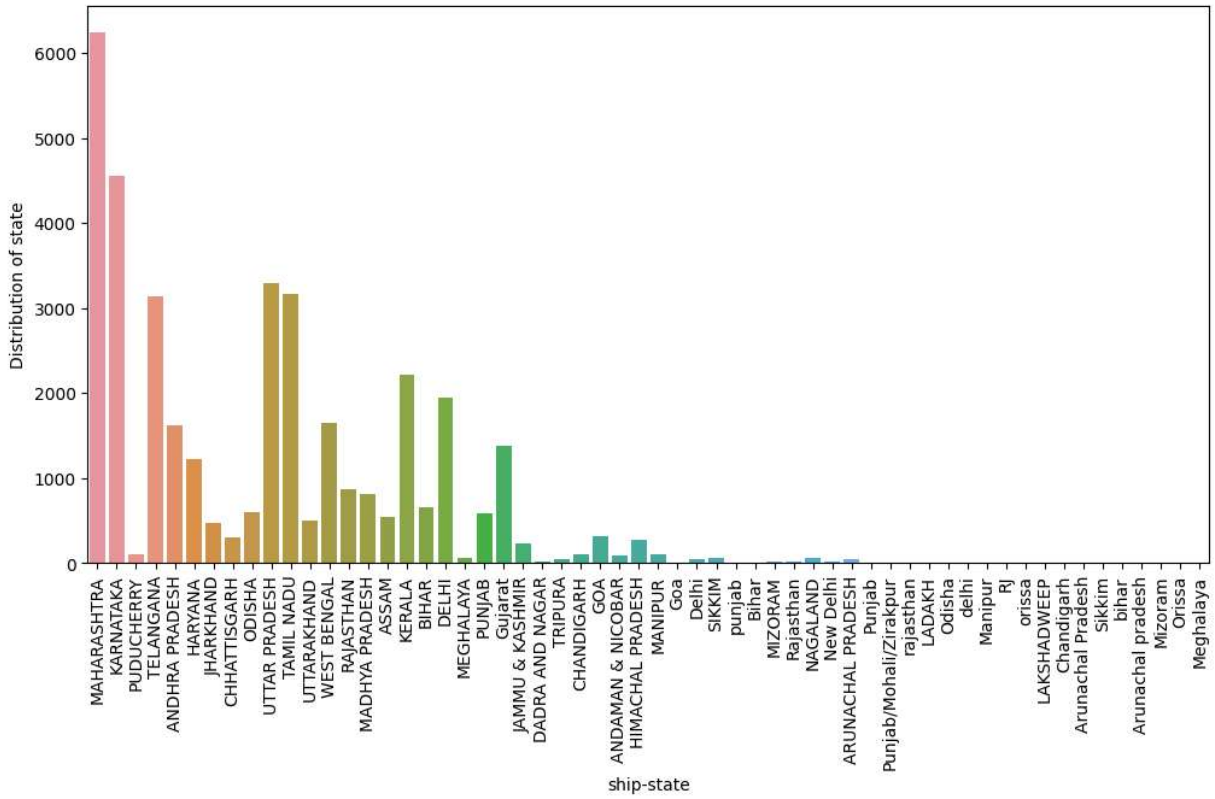
```
### Prepare data for scatter plot
x_data=df["Category"]
y_data=df["Size"]

plt.scatter(x_data,y_data)
plt.xlabel("Category")
plt.ylabel("Size")
plt.title("Scatter Plot")
plt.show()
```



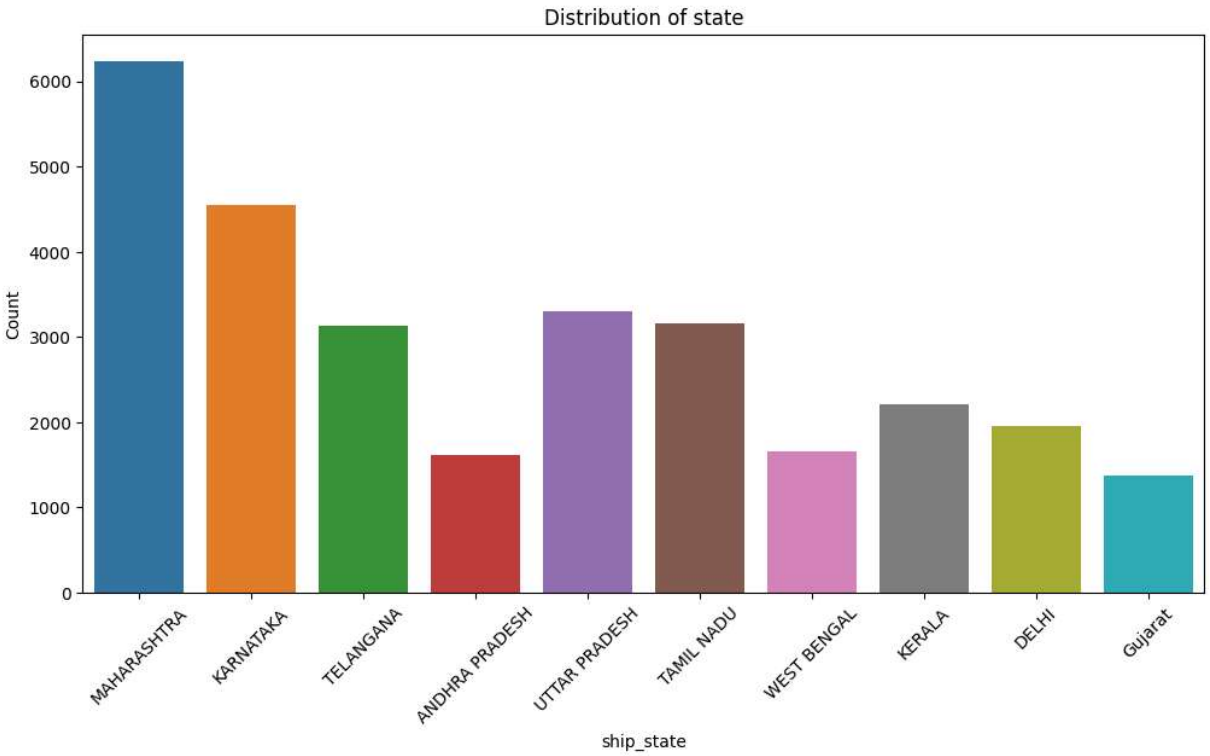

In [122…

```
## PLot count of cities by state
plt.figure(figsize=(12,6))
sns.countplot(data=df,x='ship-state')
plt.xlabel('ship-state')
plt.ylabel('Distribution of state')
plt.xticks(rotation=90)
plt.show()
```




In [128…

```
### Top 10 states
top_10_states=df["ship-state"].value_counts().head(10)
#plot count of cities by state
plt.figure(figsize=(12,6))
sns.countplot(data=df[df["ship-state"].isin(top_10_states.index)],x='ship-state')
plt.xlabel("ship_state")
plt.ylabel("Count")
plt.title("Distribution of state")
plt.xticks(rotation=45)
plt.show()
```




**Note: From above graph you can see that most of the buyers from Maharashtra**

## Conclusion

- The data analysis reveals that the business has a significant customer base in Maharashtra state, mainly serves reatailers, fullfills order through Amazon, experiances high demand of T-shirts and see M-Size as the preffered choice among buyers

In [ ]: