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

In [14]:
```python
df=pd.read_csv('Amazon Sale Report.csv')
```

In [10]:
```python
import os
```

In [11]:
```python
os.getcwd()
```

Out[11]: 'C:\\Users\\suran\\Downloads\\Amazon sales analysis'

In [12]:
```python
os.chdir('C:\\Users\\suran\\Downloads')
```

In [15]:
```python
df
```

Out[15]:

	index	Order ID	Date	Status	Fulfilment	Sales Channel	ship-service-level	Category	Size	Courier Status	...	currency	Amount	ship-city	ship-state	ship-postal-code	ship-country	B2B
0	0	405-8078784-5731545	04-30-22	Cancelled	Merchant	Amazon.in	Standard	T-shirt	S	On the Way	...	INR	647.62	MUMBAI	MAHARASHTRA	400081.0	IN	False
1	1	171-9198151-1101146	04-30-22	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	Shirt	3XL	Shipped	...	INR	406.00	BENGALURU	KARNATAKA	560085.0	IN	False
2	2	404-0687676-7273146	04-30-22	Shipped	Amazon	Amazon.in	Expedited	Shirt	XL	Shipped	...	INR	329.00	NAVI MUMBAI	MAHARASHTRA	410210.0	IN	True
3	3	403-9615377-8133951	04-30-22	Cancelled	Merchant	Amazon.in	Standard	Blazzer	L	On the Way	...	INR	753.33	PUDUCHERRY	PUDUCHERRY	605008.0	IN	False
4	4	407-1069790-7240320	04-30-22	Shipped	Amazon	Amazon.in	Expedited	Trousers	3XL	Shipped	...	INR	574.00	CHENNAI	TAMIL NADU	600073.0	IN	False
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
128971	128970	406-6001380-7673107	05-31-22	Shipped	Amazon	Amazon.in	Expedited	Shirt	XL	Shipped	...	INR	517.00	HYDERABAD	TELANGANA	500013.0	IN	False
128972	128971	402-9551604-7544318	05-31-22	Shipped	Amazon	Amazon.in	Expedited	T-shirt	M	Shipped	...	INR	999.00	GURUGRAM	HARYANA	122004.0	IN	False
128973	128972	407-9547469-3152358	05-31-22	Shipped	Amazon	Amazon.in	Expedited	Blazzer	XXL	Shipped	...	INR	690.00	HYDERABAD	TELANGANA	500049.0	IN	False
128974	128973	402-6184140-0545956	05-31-22	Shipped	Amazon	Amazon.in	Expedited	T-shirt	XS	Shipped	...	INR	1199.00	Halol	Gujarat	389350.0	IN	False
128975	128974	408-7436540-8728312	05-31-22	Shipped	Amazon	Amazon.in	Expedited	T-shirt	S	Shipped	...	INR	696.00	Raipur	CHHATTISGARH	492014.0	IN	False

128976 rows × 21 columns

In [16]:
```python
df.shape
```

Out[16]: (128976, 21)

In [17]:
```python
df.head(10)
```

Out[17]:

	index	Order ID	Date	Status	Fulfilment	Sales Channel	ship-service-level	Category	Size	Courier Status	...	currency	Amount	ship-city	ship-state	ship-postal-code	ship-country	B2B	fulfilled-by
0	0	405-8078784-5731545	04-30-22	Cancelled	Merchant	Amazon.in	Standard	T-shirt	S	On the Way	...	INR	647.62	MUMBAI	MAHARASHTRA	400081.0	IN	False	Easy Ship
1	1	171-9198151-1101146	04-30-22	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	Shirt	3XL	Shipped	...	INR	406.00	BENGALURU	KARNATAKA	560085.0	IN	False	Easy Ship
2	2	404-0687676-7273146	04-30-22	Shipped	Amazon	Amazon.in	Expedited	Shirt	XL	Shipped	...	INR	329.00	NAVI MUMBAI	MAHARASHTRA	410210.0	IN	True	NaN
3	3	403-9615377-8133951	04-30-22	Cancelled	Merchant	Amazon.in	Standard	Blazzer	L	On the Way	...	INR	753.33	PUDUCHERRY	PUDUCHERRY	605008.0	IN	False	Easy Ship
4	4	407-1069790-7240320	04-30-22	Shipped	Amazon	Amazon.in	Expedited	Trousers	3XL	Shipped	...	INR	574.00	CHENNAI	TAMIL NADU	600073.0	IN	False	NaN
5	5	404-1490984-4578765	04-30-22	Shipped	Amazon	Amazon.in	Expedited	T-shirt	XL	Shipped	...	INR	824.00	GHAZIABAD	UTTAR PRADESH	201102.0	IN	False	NaN
6	6	408-5748499-6859555	04-30-22	Shipped	Amazon	Amazon.in	Expedited	T-shirt	L	Shipped	...	INR	653.00	CHANDIGARH	CHANDIGARH	160036.0	IN	False	NaN
7	7	406-7807733-3785945	04-30-22	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	Shirt	S	Shipped	...	INR	399.00	HYDERABAD	TELANGANA	500032.0	IN	False	Easy Ship
8	8	407-5443024-5233168	04-30-22	Cancelled	Amazon	Amazon.in	Expedited	T-shirt	3XL	Cancelled	...	NaN	NaN	HYDERABAD	TELANGANA	500008.0	IN	False	NaN
9	9	402-4393761-0311520	04-30-22	Shipped	Amazon	Amazon.in	Expedited	Shirt	XXL	Shipped	...	INR	363.00	Chennai	TAMIL NADU	600041.0	IN	False	NaN

10 rows × 21 columns

In [18]:
```python
df.tail()
```

Out[18]:

	index	Order ID	Date	Status	Fulfilment	Sales Channel	ship-service-level	Category	Size	Courier Status	...	currency	Amount	ship-city	ship-state	ship-postal-code	ship-country	B2B	fulfilled-by
128971	128970	406-6001380-7673107	05-31-22	Shipped	Amazon	Amazon.in	Expedited	Shirt	XL	Shipped	...	INR	517.0	HYDERABAD	TELANGANA	500013.0	IN	False	NaN
128972	128971	402-9551604-7544318	05-31-22	Shipped	Amazon	Amazon.in	Expedited	T-shirt	M	Shipped	...	INR	999.0	GURUGRAM	HARYANA	122004.0	IN	False	NaN
128973	128972	407-9547469-3152358	05-31-22	Shipped	Amazon	Amazon.in	Expedited	Blazzer	XXL	Shipped	...	INR	690.0	HYDERABAD	TELANGANA	500049.0	IN	False	NaN
128974	128973	402-6184140-0545956	05-31-22	Shipped	Amazon	Amazon.in	Expedited	T-shirt	XS	Shipped	...	INR	1199.0	Halol	Gujarat	389350.0	IN	False	NaN
128975	128974	408-7436540-8728312	05-31-22	Shipped	Amazon	Amazon.in	Expedited	T-shirt	S	Shipped	...	INR	696.0	Raipur	CHHATTISGARH	492014.0	IN	False	NaN

5 rows × 21 columns

In [19]:
```python
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 [20]:
```python
#Drop unrelated/blank columns
df.drop(['New','PendingS'],axis=1,inplace=True)
```

In [21]:
```python
pd.isnull(df)
```

Out[21]:

	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
0	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	True
3	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	True
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
128971	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	True
128972	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	True
128973	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	True
128974	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	True
128975	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	True

128976 rows × 19 columns

In [22]:
```python
pd.isnull(df).sum()
```

Out[22]:

```
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 [23]:
```python
#drop null values
df.dropna(inplace=True)
```

In [24]:
```python
df.shape
```

Out[24]: (37514, 19)

In [25]:
```python
df.columns
```

Out[25]:

```
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 [26]:
```python
#change Data type
df['ship-postal-code']=df['ship-postal-code'].astype('int')
```

In [27]:
```python
#checking whether the data type change or not
df['ship-postal-code'].dtype
```

Out[27]: dtype('int32')

In [28]:
```python
df['Date']=pd.to_datetime(df['Date'])
```

In [29]:
```python
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 [34]:
```python
#Rename column
df.rename(columns={'Qty':'Quantity'},inplace=True)
df
```

Out[34]:

	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
0	0	405-8078784-5731545	2022-04-30	Cancelled	Merchant	Amazon.in	Standard	T-shirt	S	On the Way	0	INR	647.62	MUMBAI	MAHARASHTRA	400081	IN	False
1	1	171-9198151-1101146	2022-04-30	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	Shirt	3XL	Shipped	1	INR	406.00	BENGALURU	KARNATAKA	560085	IN	False
3	3	403-9615377-8133951	2022-04-30	Cancelled	Merchant	Amazon.in	Standard	Blazzer	L	On the Way	0	INR	753.33	PUDUCHERRY	PUDUCHERRY	605008	IN	False
7	7	406-7807733-3785945	2022-04-30	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	Shirt	S	Shipped	1	INR	399.00	HYDERABAD	TELANGANA	500032	IN	False
12	12	405-5513694-8146768	2022-04-30	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	Shirt	XS	Shipped	1	INR	399.00	Amravati.	MAHARASHTRA	444606	IN	False
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
128875	128874	405-4724097-1016369	2022-06-01	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	T-shirt	S	Shipped	1	INR	854.00	ALLUR	ANDHRA PRADESH	524315	IN	False
128876	128875	403-9524128-9243508	2022-06-01	Cancelled	Merchant	Amazon.in	Standard	Blazzer	XL	On the Way	0	INR	734.29	Barabanki	UTTAR PRADESH	225001	IN	False
128888	128887	405-6493630-8542756	2022-05-31	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	Trousers	M	Shipped	1	INR	518.00	NOIDA	UTTAR PRADESH	201301	IN	False
128891	128890	407-0116398-1810752	2022-05-31	Cancelled	Merchant	Amazon.in	Standard	Wallet	Free	On the Way	0	INR	398.10	MADURAI	TAMIL NADU	625007	IN	False
128892	128891	403-0317423-9322704	2022-05-31	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	Blazzer	M	Shipped	1	INR	721.00	UTTAR BAGDOGRA	WEST BENGAL	734014	IN	False

37514 rows × 19 columns

In [35]:
```python
#describe() method return description of the data in the dataframe(i.e count,mean,std,min.etc)
df.describe()
```

Out[35]:

	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 [36]:
```python
df.describe(include='object')
```

Out[36]:

	Order ID	Status	Fulfilment	Sales Channel	ship-service-level	Category	Size	Courier Status	currency	ship-city	ship-state	ship-country	fulfilled-by
count	37514	37514	37514	37514	37514	37514	37514	37514	37514	37514	37514	37514	37514
unique	34664	11	1	1	1	8	11	3	1	4698	58	1	1
top	171-5057375-2831560	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	T-shirt	M	Shipped	INR	BENGALURU	MAHARASHTRA	IN	Easy Ship
freq	12	28741	37514	37514	37514	14062	6806	31859	37514	2839	6236	37514	37514

In [37]:
```python
#Use describe() method for specific columns
df[['Quantity','Amount']].describe()
```

Out[37]:

	Quantity	Amount
count	37514.000000	37514.000000
mean	0.867383	646.553960
std	0.354160	279.952414
min	0.000000	0.000000
25%	1.000000	458.000000
50%	1.000000	629.000000
75%	1.000000	771.000000
max	5.000000	5495.000000

Exploratory Data Analysis

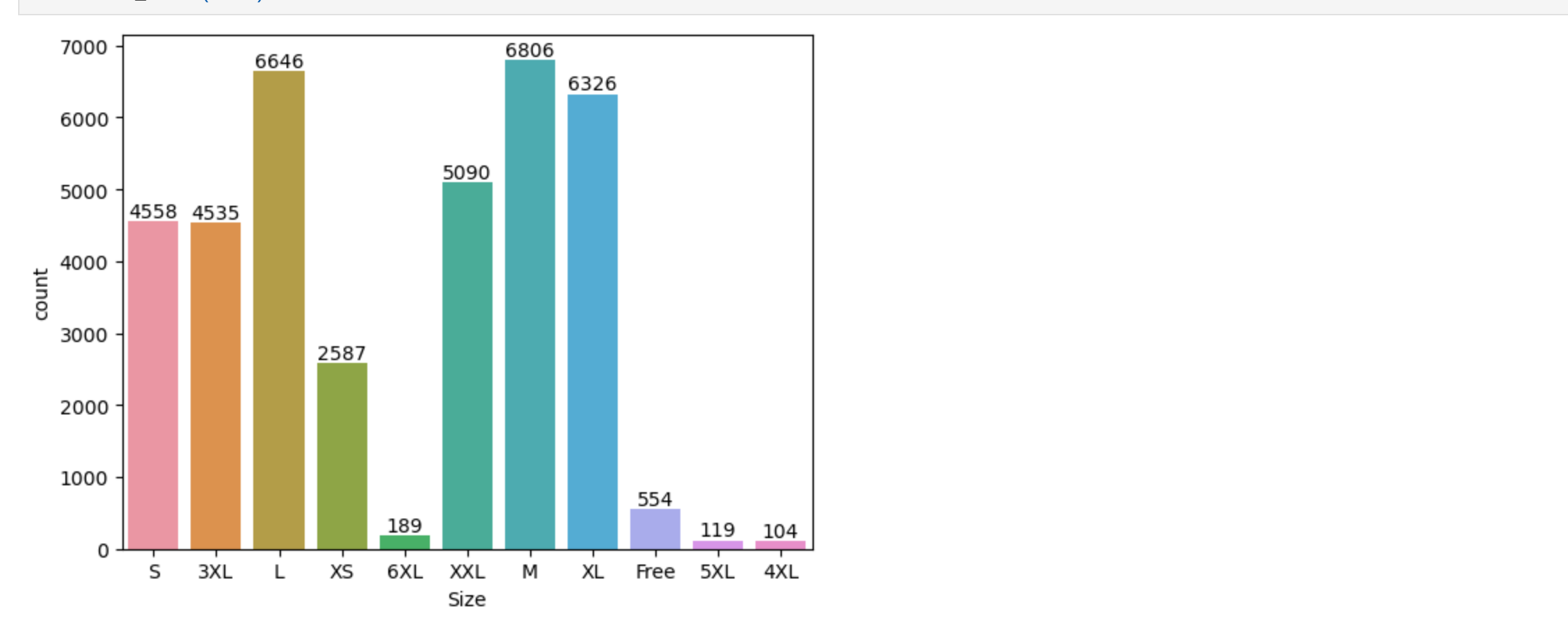
In [38]:
```python
df.columns
```

Out[38]:

```
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')
```

Size

In [42]:
```python
ax=sns.countplot(x='Size',data=df)

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



- From above graph shows that most of the people buy M-Size.

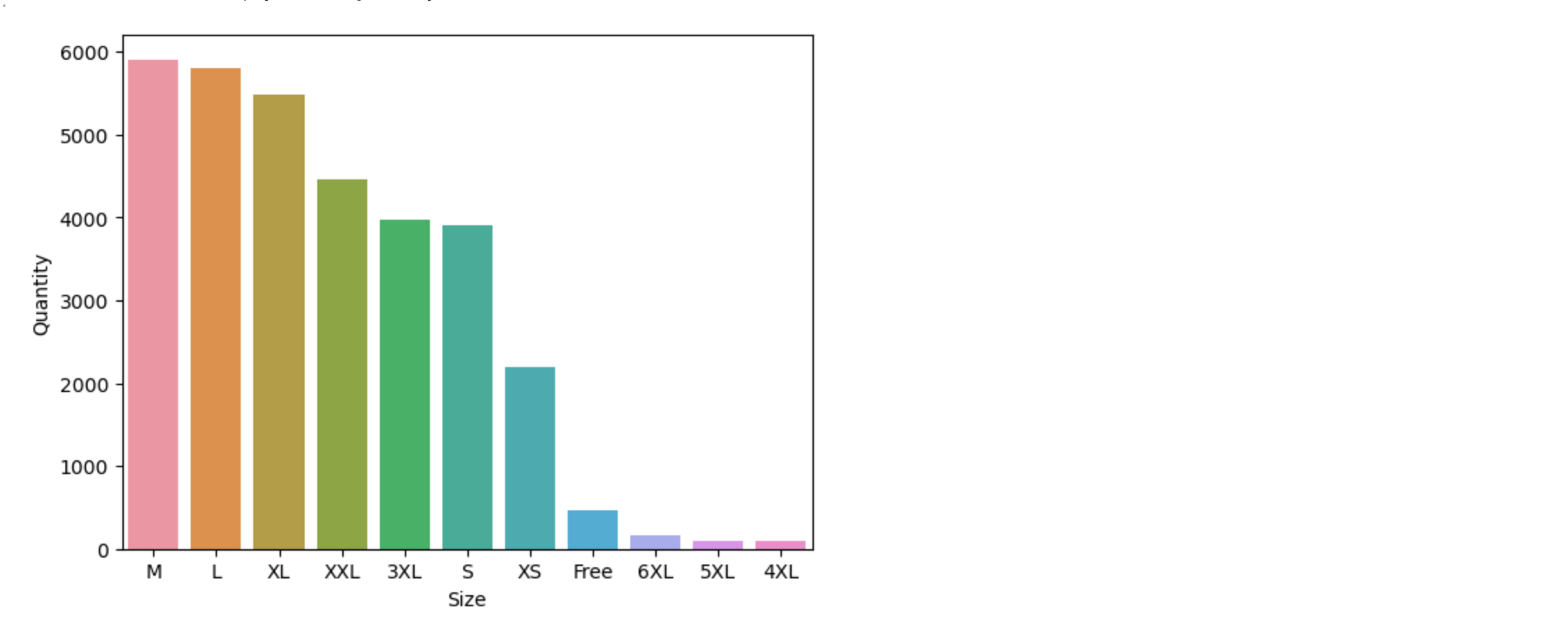
In [43]:
```python
#using group by
df.groupby(["Size"],as_index=False)['Quantity'].sum().sort_values(by="Quantity",ascending=False)
```

Out[43]:

	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 [46]:
```python
S_Qty=df.groupby(["Size"],as_index=False)['Quantity'].sum().sort_values(by="Quantity",ascending=False)

sns.barplot(x='Size',y='Quantity',data=S_Qty)
```

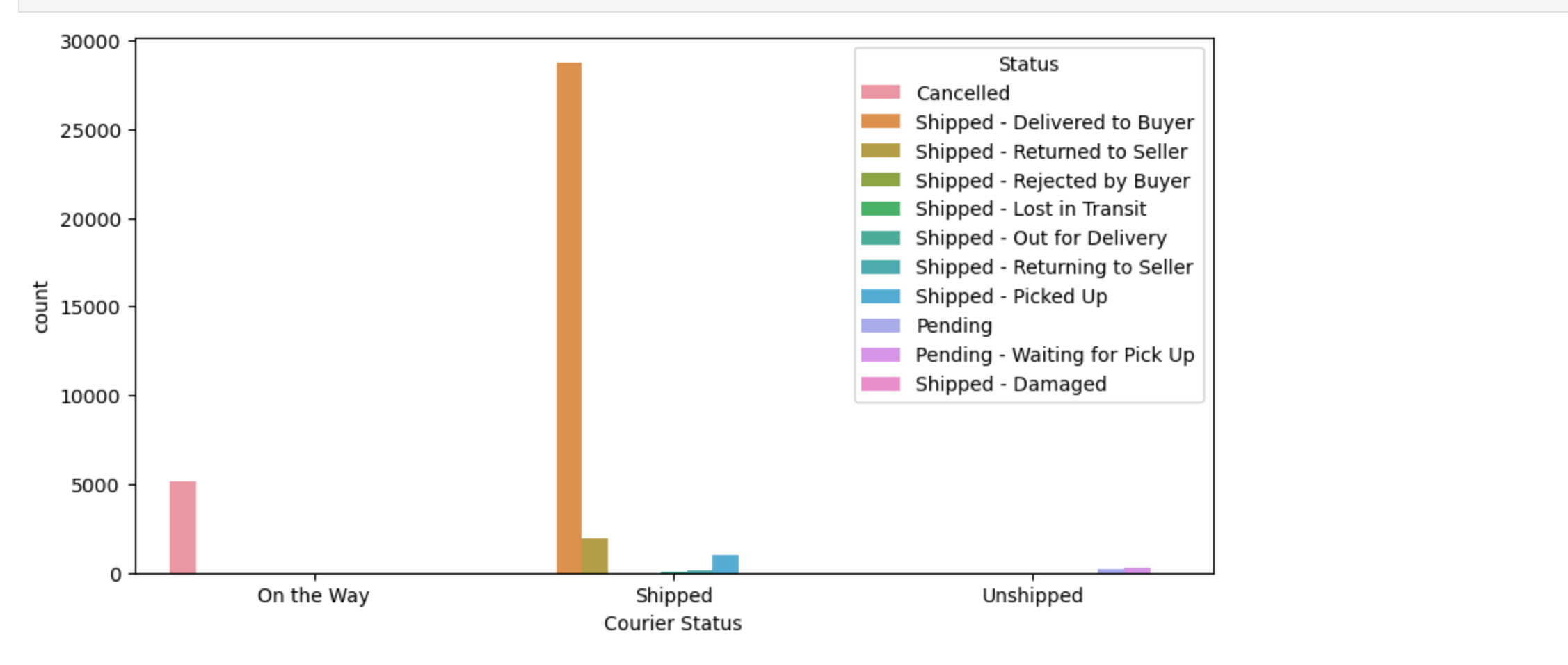
Out[46]: <Axes: xlabel='Size', ylabel='Quantity'>



- From above graph shows that most of the quantity buy M-size sales

Courier status

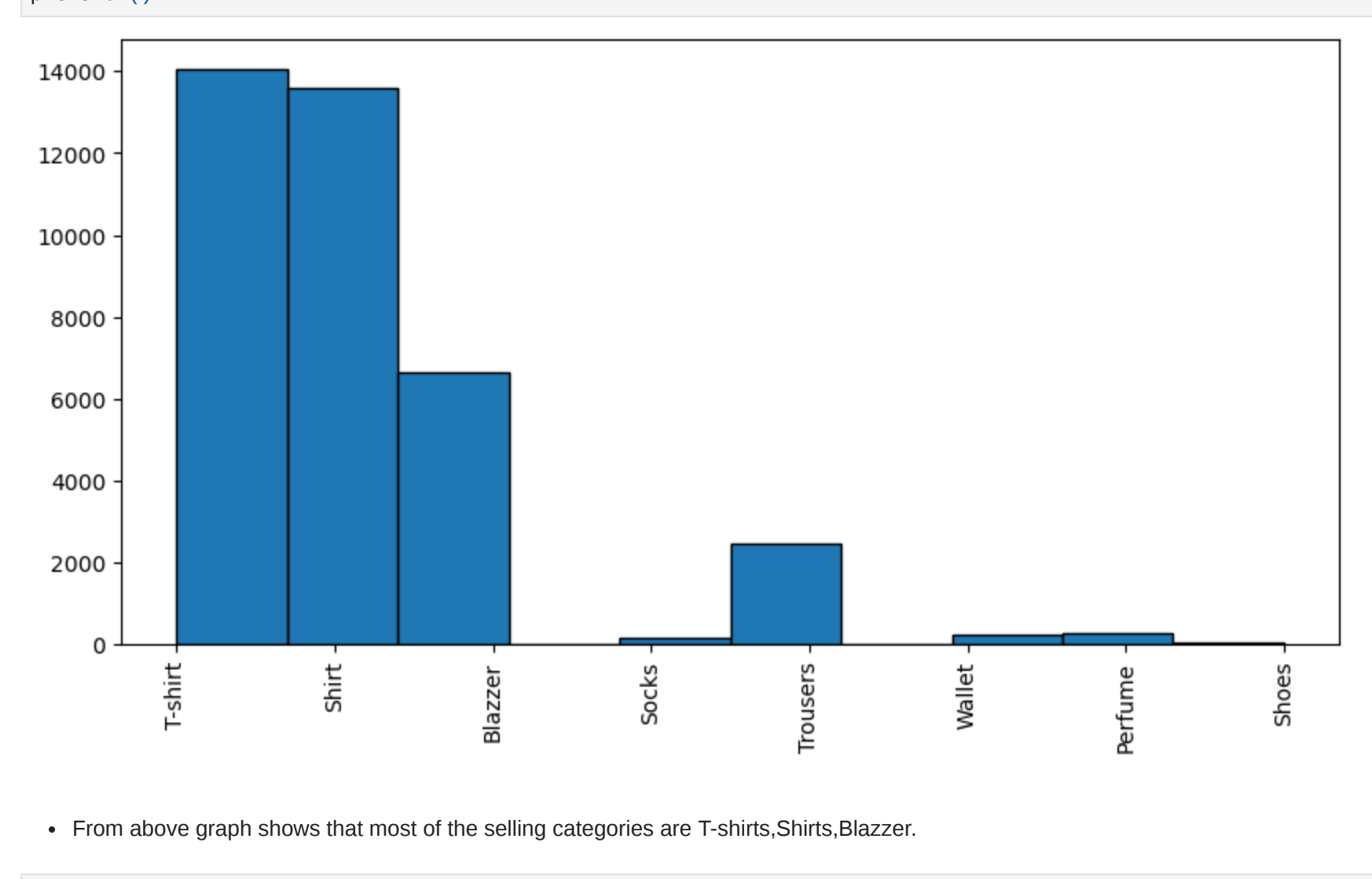
In [47]:
```python
plt.figure(figsize=(10,5))
sns.countplot(data=df,x='Courier Status',hue='Status')
plt.show()
```



- from the above graph shows that majority of the orders are shipped through courier.

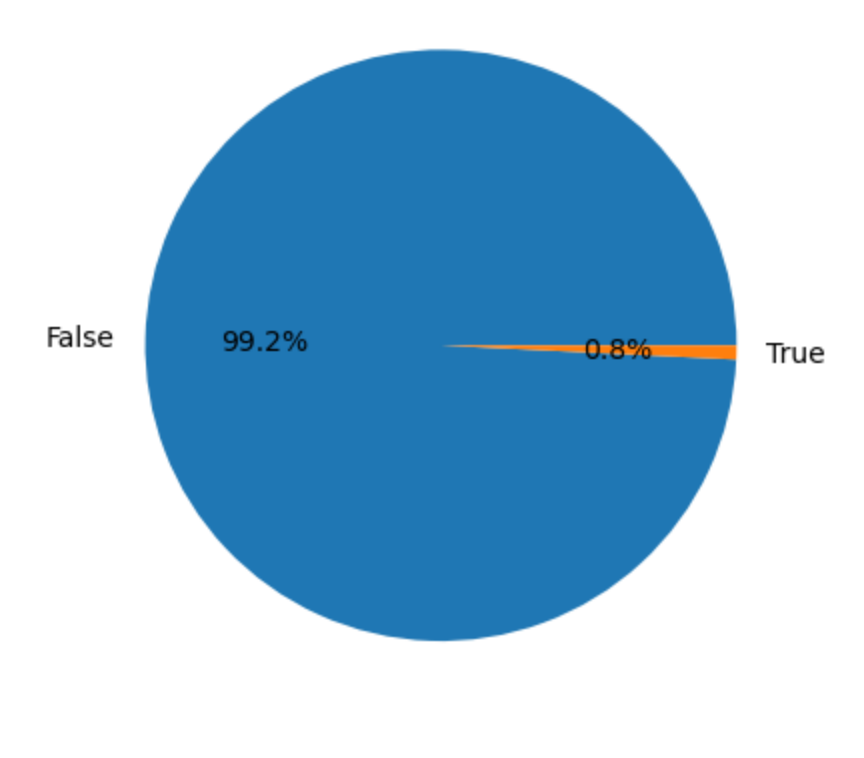
Category

In [53]:
```python
df['Category']=df['Category'].astype(str)
Column_data=df['Category']
plt.figure(figsize=(10,5))
plt.hist(Column_data, bins=10, edgecolor='Black')
plt.xticks(rotation=90)
plt.show()
```



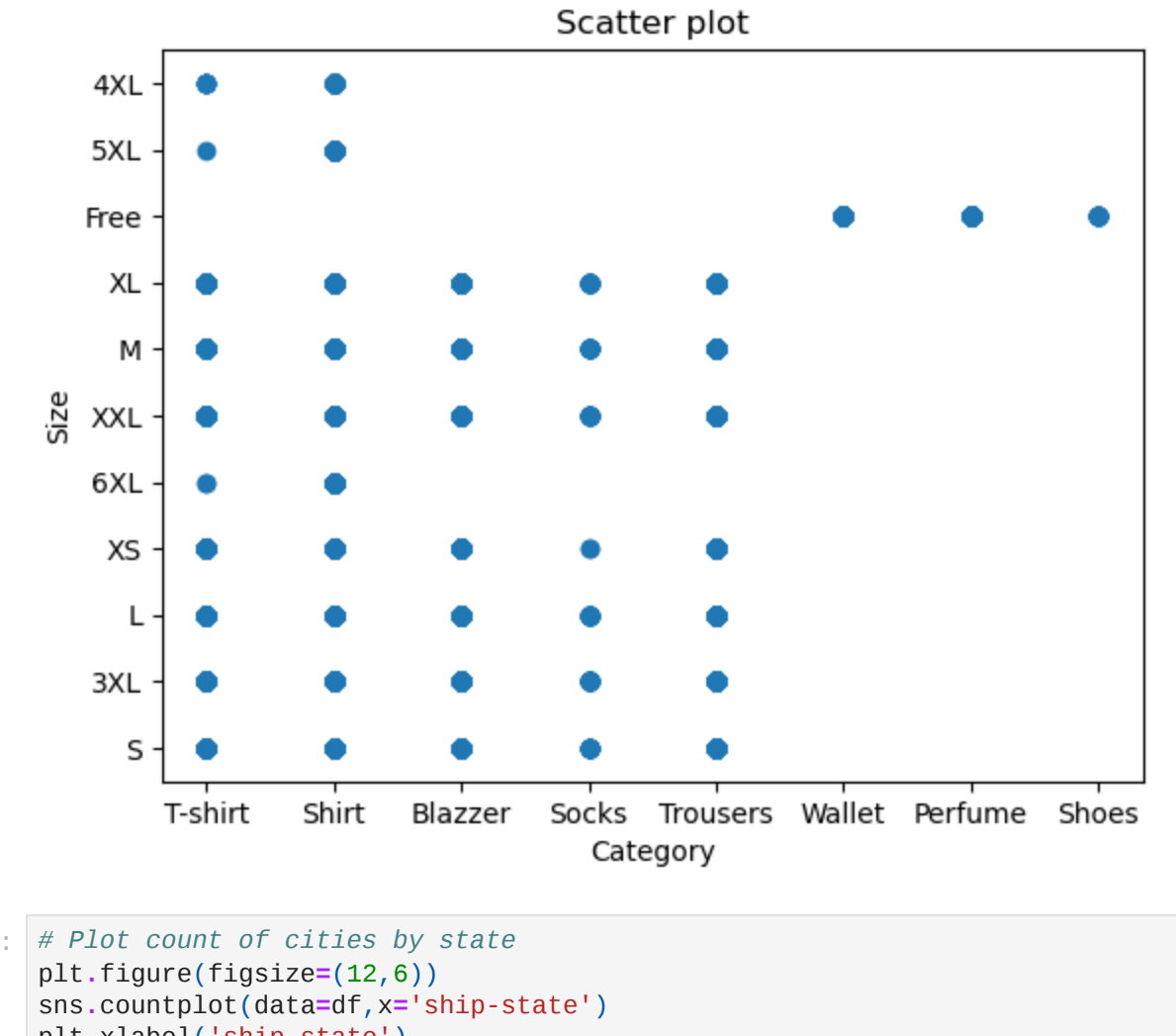
- From above graph shows that most of the selling categories are T-shirts,Shirts,Blazzer.

In [57]:
```python
# Checking B2B Data by using pie chart
B2B_check= df['B2B'].value_counts()
plt.pie(B2B_check, labels=B2B_check.index,autopct='%1.1f%%')
plt.show()
```

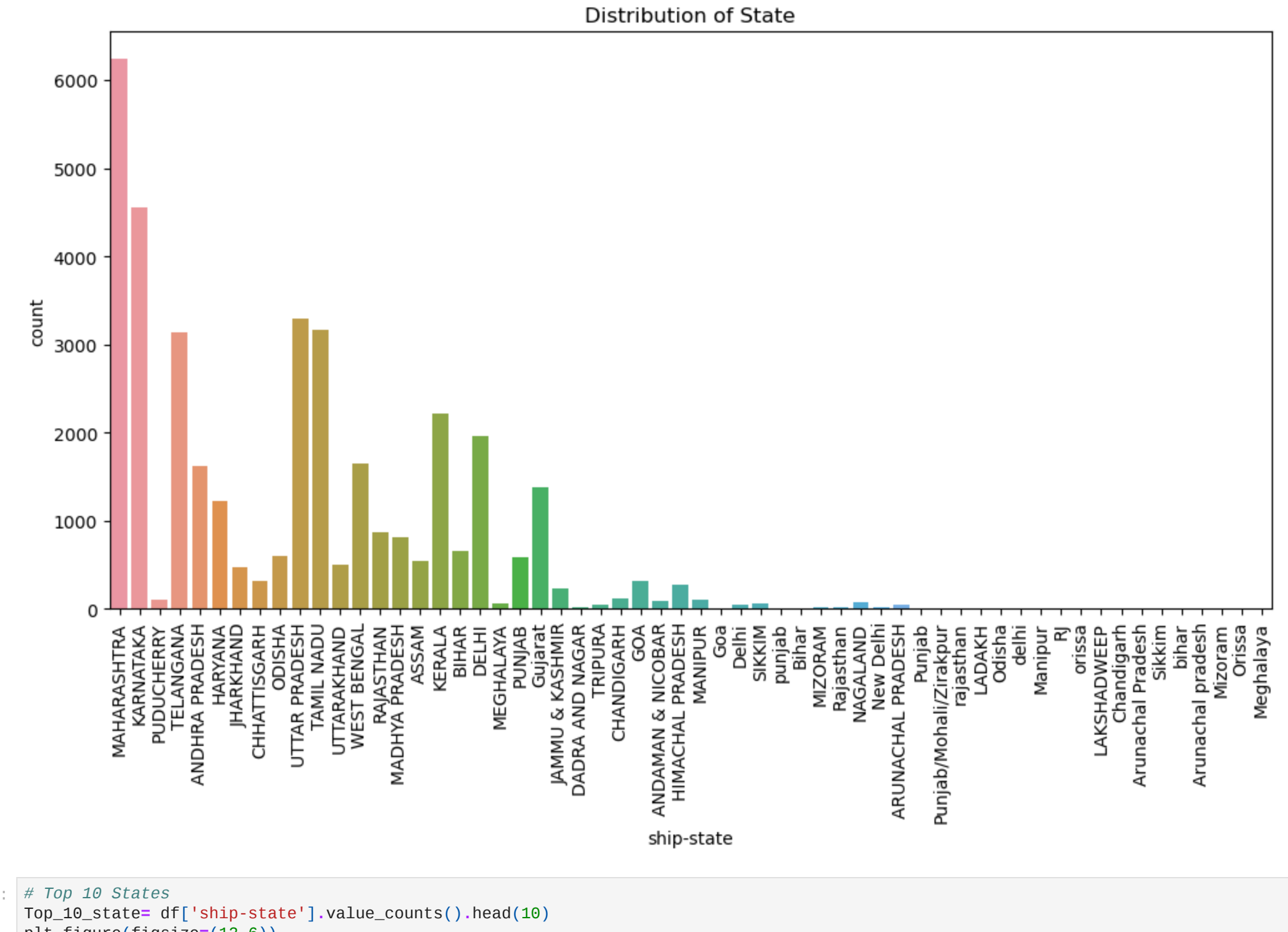


- from the above pie chart shows that maximum 99.2% of buyers are retailers and 0.8% are B2B buyers.

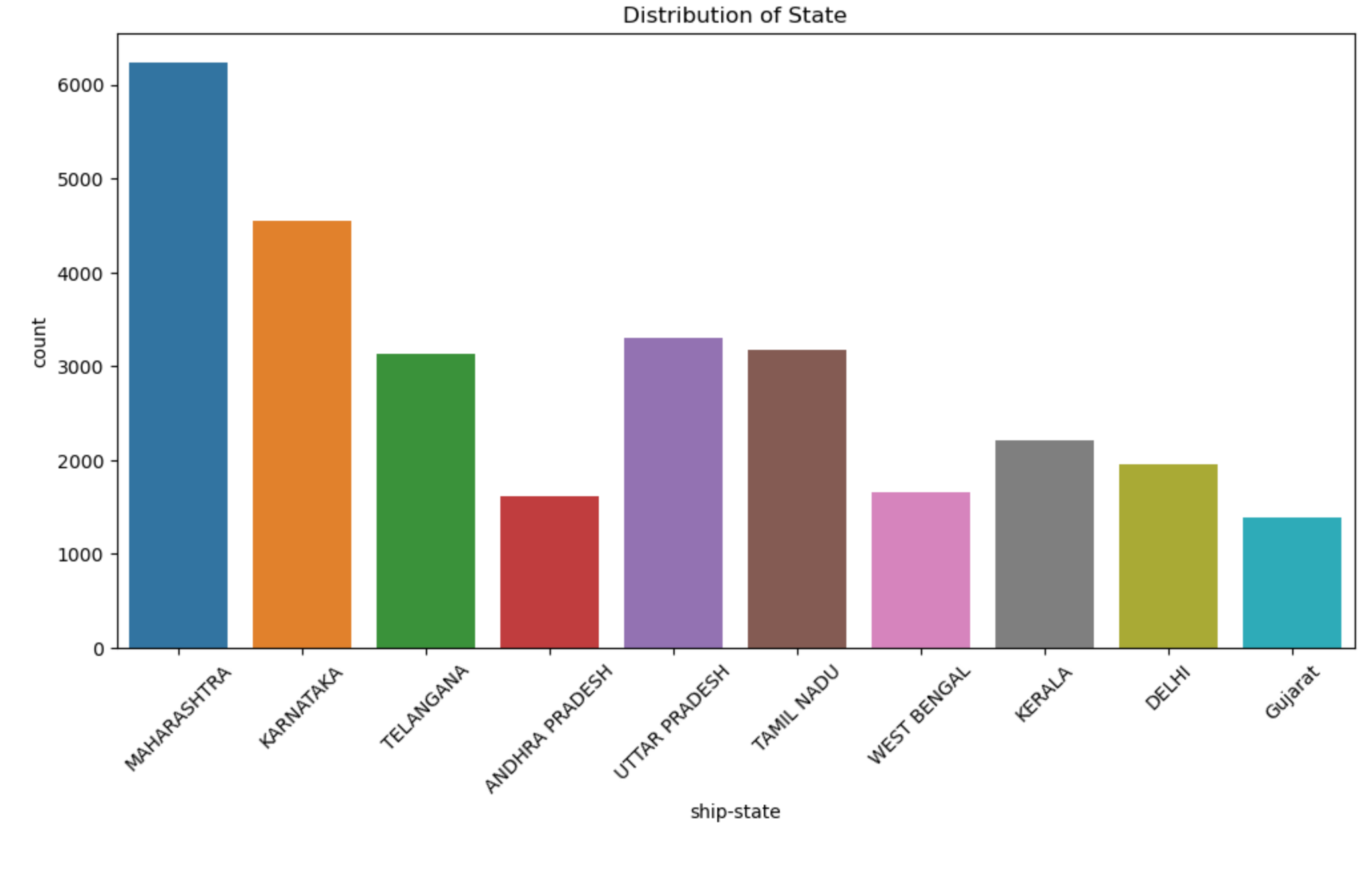
In [58]:
```python
# Use scatter plot
x_data=df['Category']
y_data=df['Size']

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



In [61]:
```python
# Plot count of cities by state
plt.figure(figsize=(12,6))
sns.countplot(data=df,x='ship-state')
plt.xlabel('ship-state')
plt.ylabel('count')
plt.title('Distribution of State')
plt.xticks(rotation=90)
plt.show()
```



In [66]:
```python
# Top 10 States
Top_10_state= df['ship-state'].value_counts().head(10)
plt.figure(figsize=(12,6))
sns.countplot(data=df[df['ship-state'].isin(Top_10_state.index)],x='ship-state')
plt.xlabel('ship-state')
plt.ylabel('count')
plt.title('Distribution of State')
plt.xticks(rotation=45)
plt.show()
```



- From above graph shows that most of the buyers are from Maharashtra,Karnataka.

Conclusion

The data reveals that the business has a significant base in Maharashtra and Karnataka state,mainly serves retailers,fulfill orders through Amazon courier and high demand on T-shirts, Shirts and Blazzer ;M-size as the preferred choice among buyers.

In [ ]: