

```
In [2]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [3]: #df=pd.read_csv('Amazon Sale Report.csv',encoding='unicode_escape')
df=pd.read_csv('C:\Users\pc\Downloads\project py\Python_Amazon_Sales_Analysis-main\Amazon Sale Rep
df
```

	index	Order ID	Date	Status	Fulfilment	Sales Channel	ship-service-level	Category	Size	Courier Status	...	currency	Amou
	0	0	405-8078784-5731545	04-30-22	Cancelled	Merchant	Amazon.in	Standard	T-shirt	S	On the Way	...	INR 647.0
	1	1	171-9198151-1101146	04-30-22	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	Shirt	3XL	Shipped	...	INR 406.0
	2	2	404-0687676-7273146	04-30-22	Shipped	Amazon	Amazon.in	Expedited	Shirt	XL	Shipped	...	INR 329.0
	3	3	403-9615377-8133951	04-30-22	Cancelled	Merchant	Amazon.in	Standard	Blazzer	L	On the Way	...	INR 753.0
	4	4	407-1069790-7240320	04-30-22	Shipped	Amazon	Amazon.in	Expedited	Trousers	3XL	Shipped	...	INR 574.0
...
	128971	128970	406-8078784-7673107	05-31-22	Shipped	Amazon	Amazon.in	Expedited	Shirt	XL	Shipped	...	INR 517.0
	128972	128971	402-9551604-7544318	05-31-22	Shipped	Amazon	Amazon.in	Expedited	T-shirt	M	Shipped	...	INR 999.0
	128973	128972	407-9547469-3152358	05-31-22	Shipped	Amazon	Amazon.in	Expedited	Blazzer	XXL	Shipped	...	INR 690.0
	128974	128973	402-6184140-0545956	05-31-22	Shipped	Amazon	Amazon.in	Expedited	T-shirt	XS	Shipped	...	INR 1199.0
	128975	128974	408-7436540-8728312	05-31-22	Shipped	Amazon	Amazon.in	Expedited	T-shirt	S	Shipped	...	INR 696.0

128976 rows × 21 columns

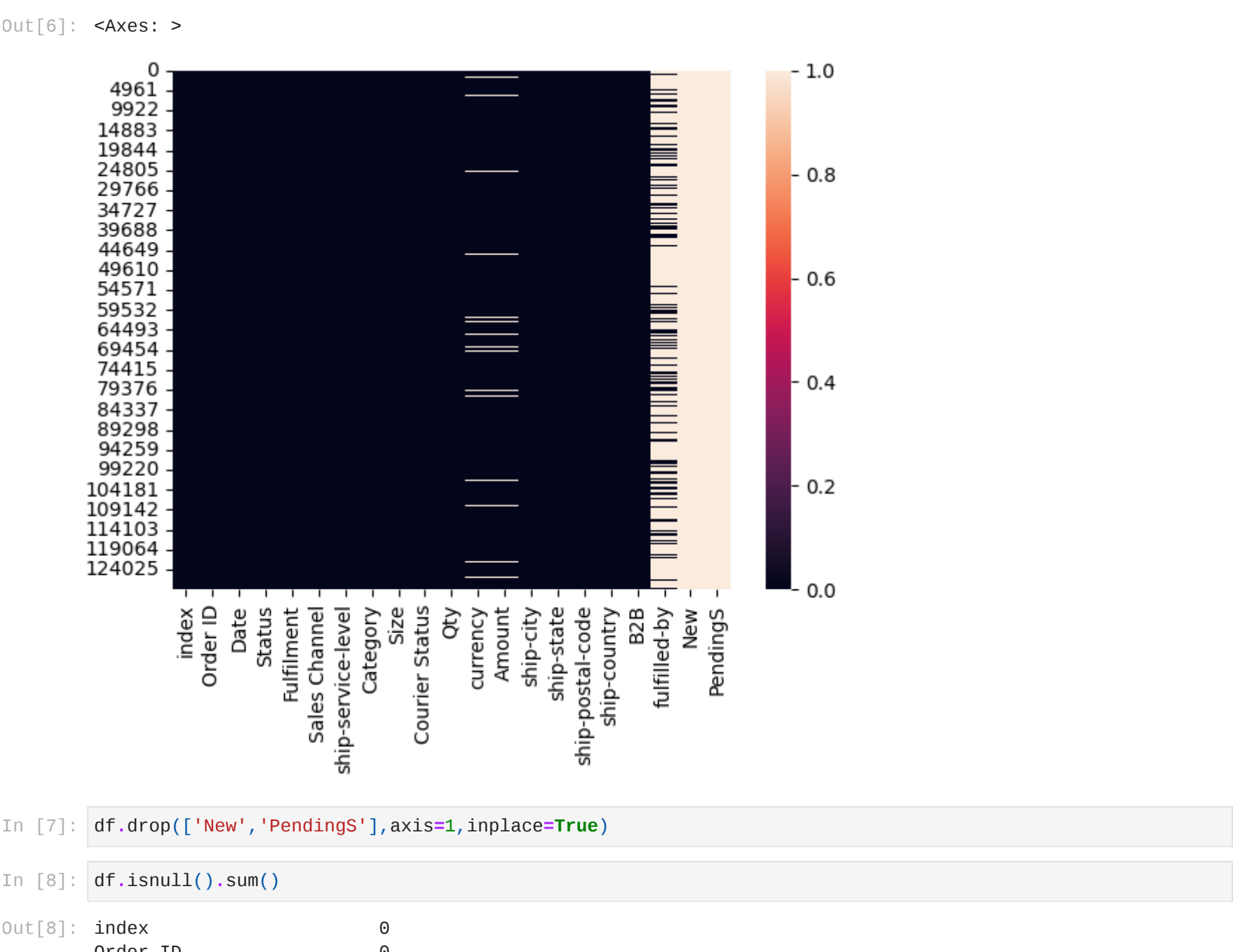
```
In [4]: df.shape
```

```
Out[4]: (128976, 21)
```

```
In [5]: 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 [6]: sns.heatmap(df.isnull())
```



```
In [7]: df.drop(['New','PendingS'],axis=1,inplace=True)
```

```
In [8]: df.isnull().sum()
```

```
Out[8]: 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              7880
Amount                7880
ship-city             35
ship-state            35
ship-postal-code      35
ship-country          35
B2B                   0
fulfilled-by         89713
dtype: int64
```

```
In [9]: #drop null values
df.dropna(inplace=True)
```

```
In [10]: df['ship-postal-code']=df['ship-postal-code'].astype('int')
```

```
In [11]: df['Date']=pd.to_datetime(df['Date'])
```

```
In [12]: df.shape
```

```
Out[12]: (37514, 19)
```

```
In [13]: df.isnull().sum()
```

```
Out[13]: 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 [14]: 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 [15]: df.describe()
```

	index	Qty	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 [16]: df.describe(include='object')
```

```
Out[16]:
```

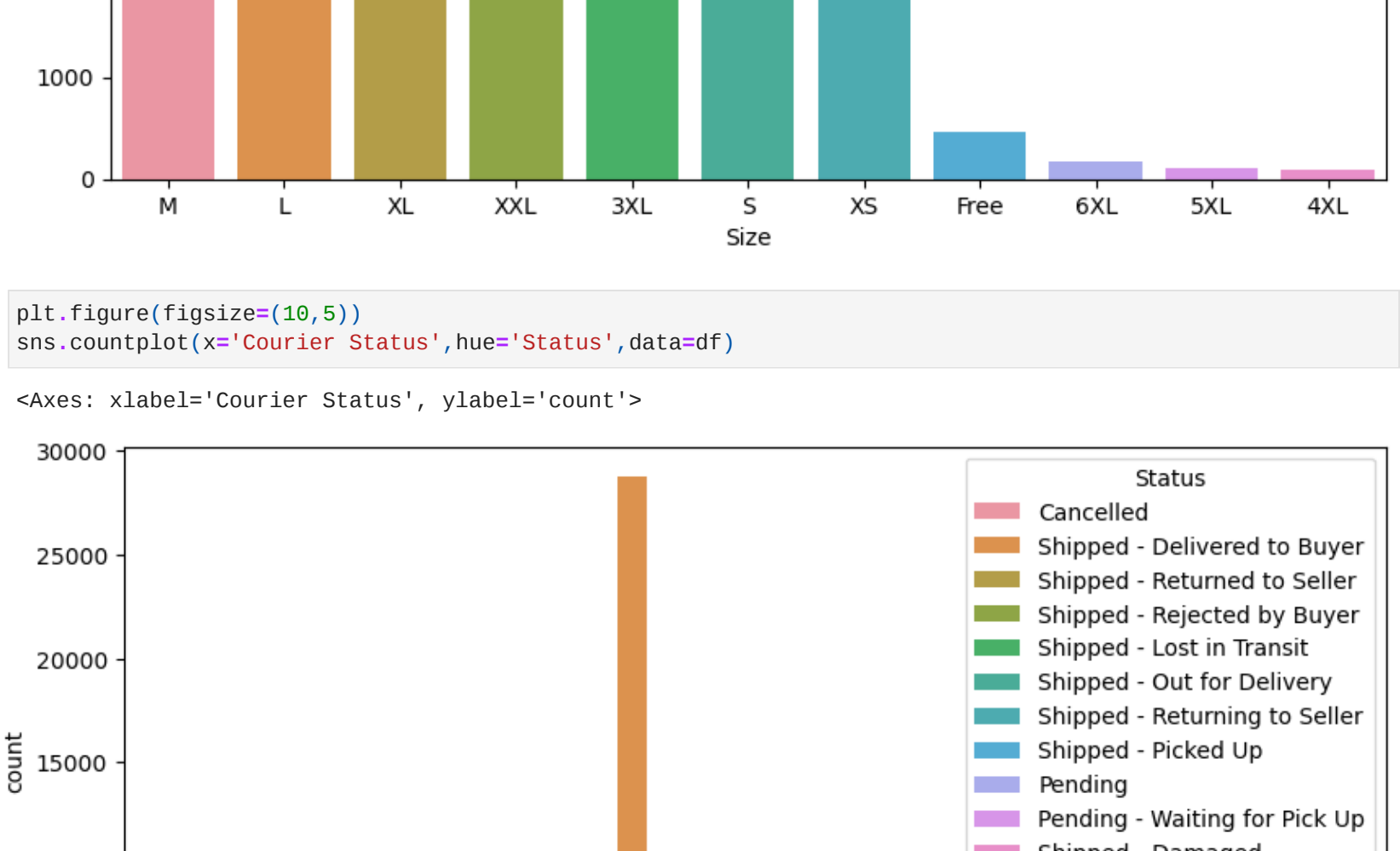
	Order ID	Status	Fulfilment	Sales Channel	ship-service-level	Category	Size	Courier Status	currency	ship-city	ship-country
count	37514	37514	37514	37514	37514	37514	37514	37514	37514	37514	37514
unique	34664	11	1	1	1	8	11	3	1	4698	37514
top	5057375-2831560	Shipped - Delivered to Buyer	Merchant	Amazon.in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

Data Analysis And Visualization

```
In [17]: df.columns
```

```
Out[17]: 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)
```

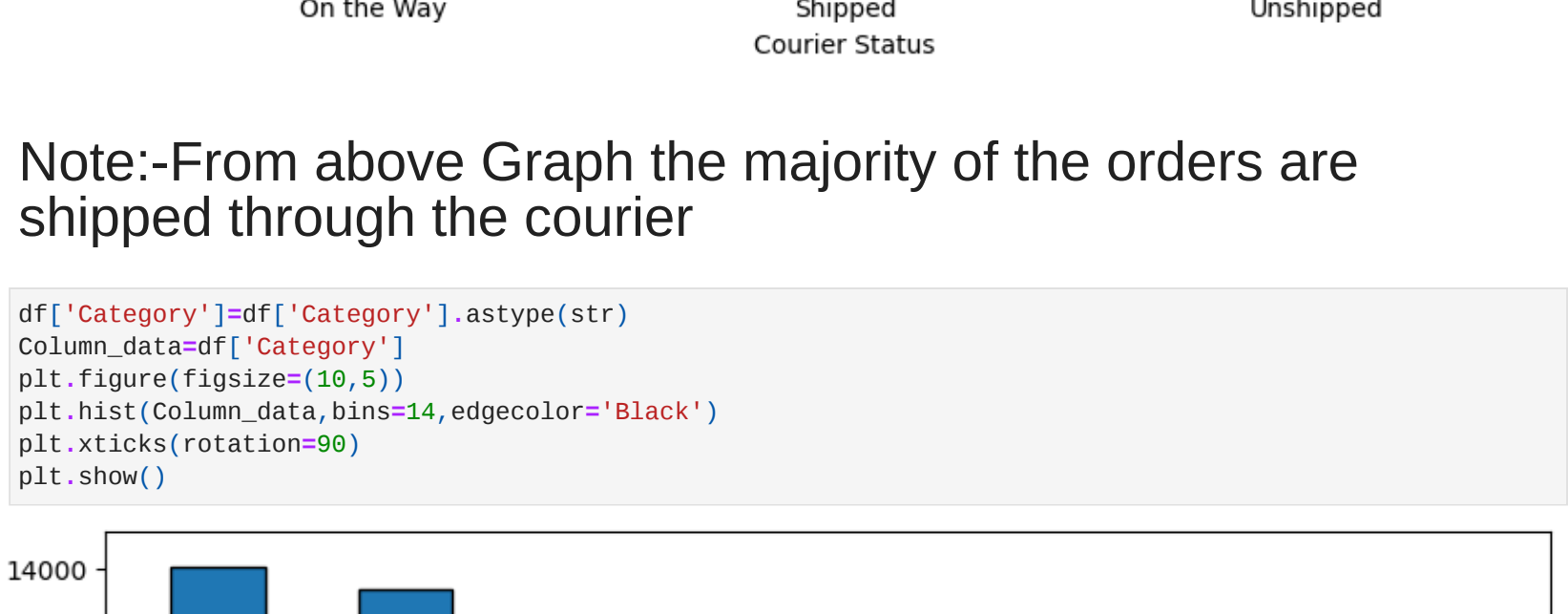
```
In [21]: plt.figure(figsize=(10,5))
s=sns.countplot(x='Size',data=df)
for bars in s.containers:
    s.bar_label(bars)
```



Note:-From above Graph can see that most of the people buys M-Size

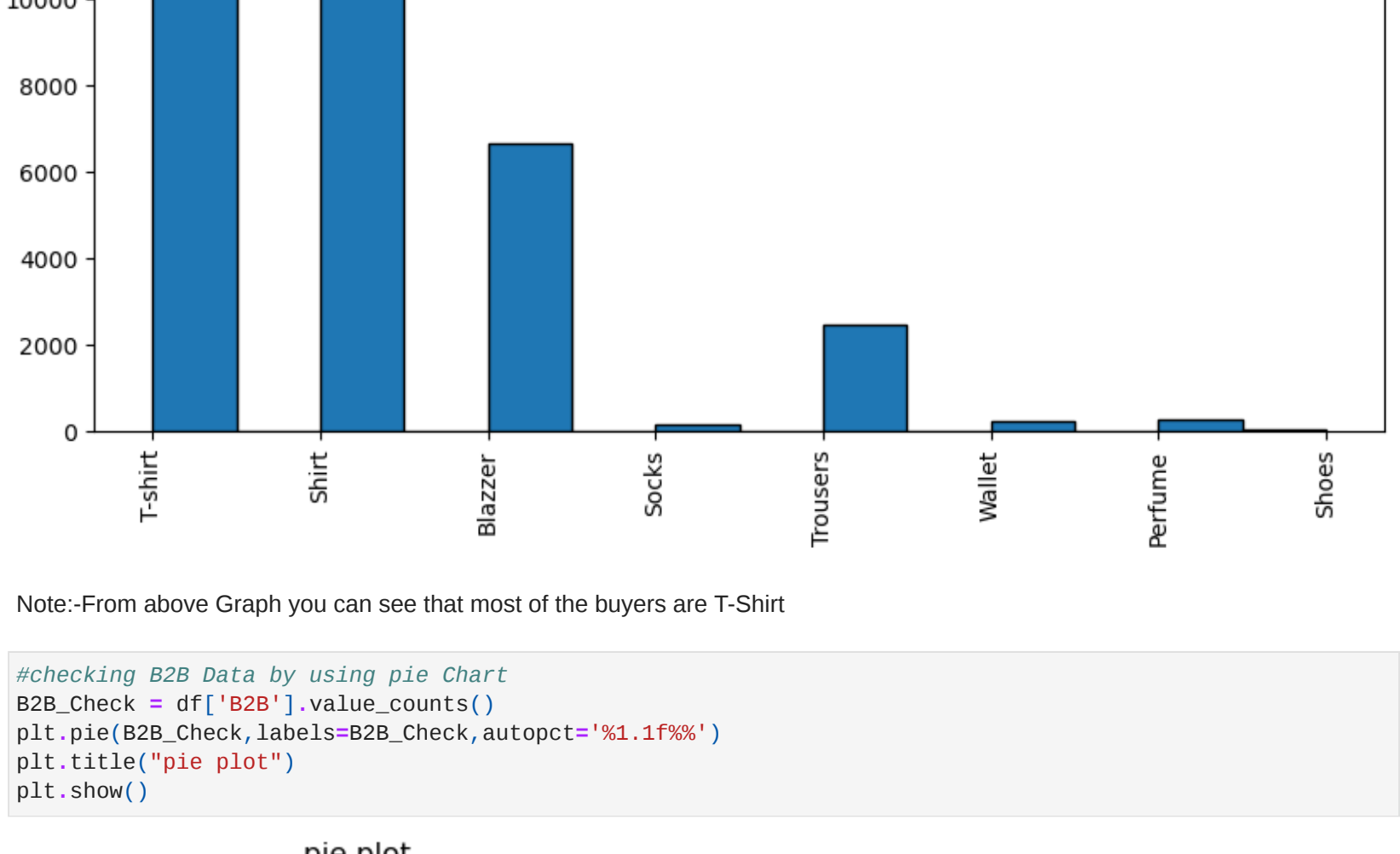
```
In [29]: plt.figure(figsize=(10,5))
S_Qty=df.groupby(['Size'],as_index=False)['Qty'].sum().sort_values(by='Qty',ascending=False)
sns.barplot(x='Size',y='Qty',data=S_Qty)
```

```
Out[29]: <Axes: xlabel='Size', ylabel='Qty'>
```



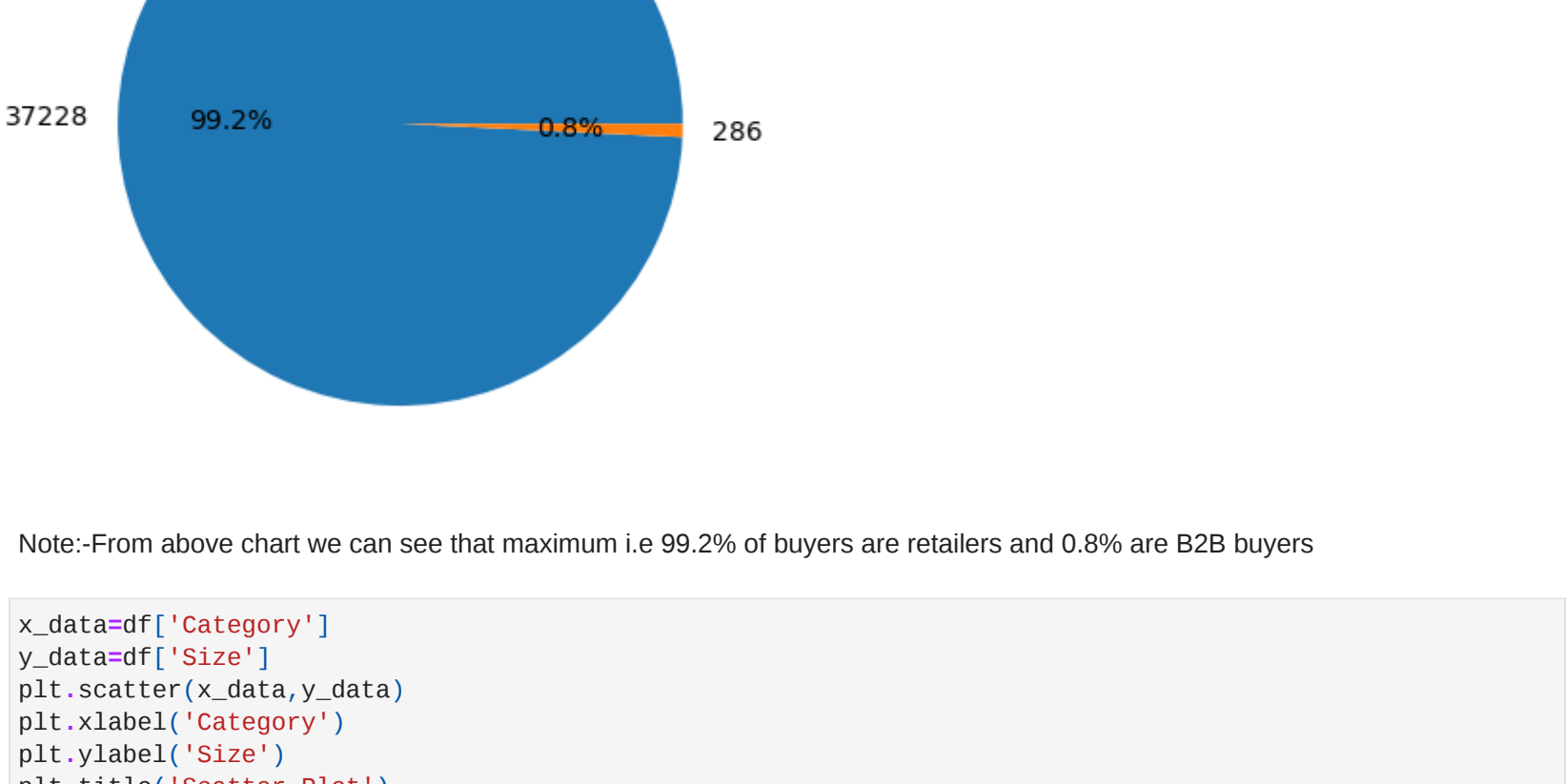
```
In [32]: plt.figure(figsize=(10,5))
sns.countplot(x='Courier Status',hue='Status',data=df)
```

```
Out[32]: <Axes: xlabel='Courier Status', ylabel='count'>
```



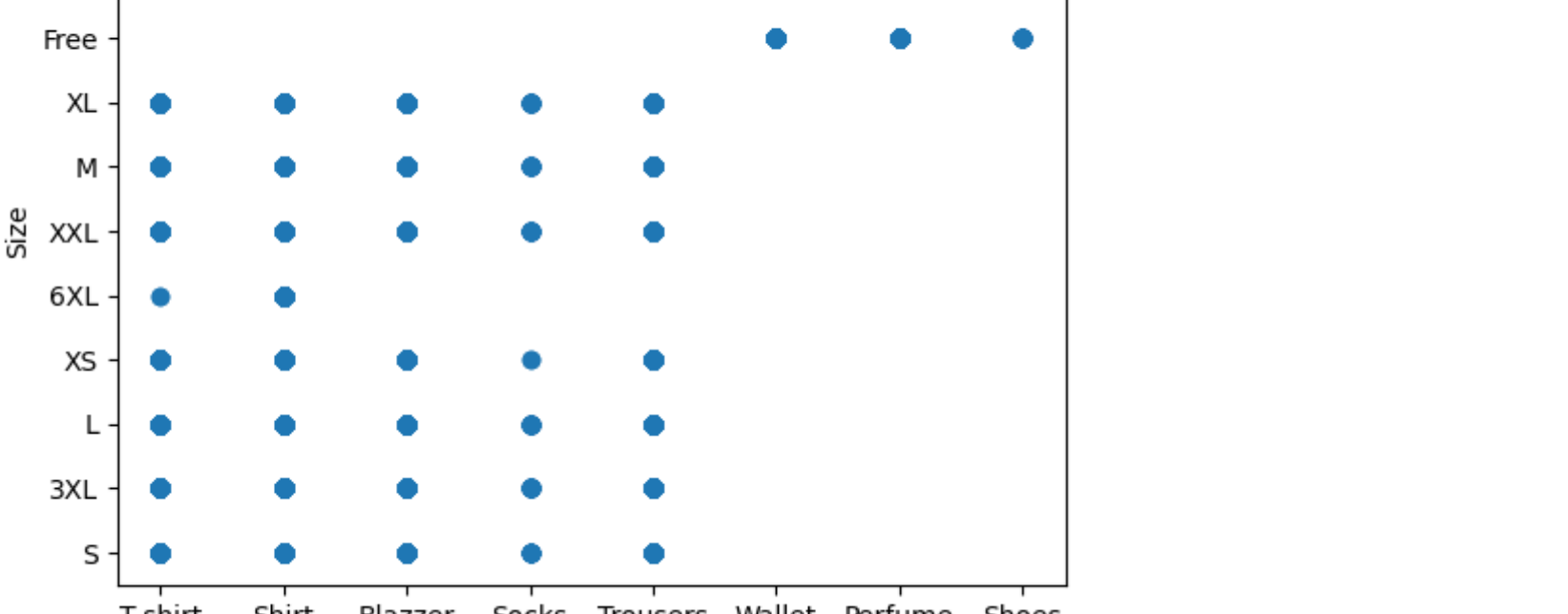
Note:-From above Graph the majority of the orders are shipped through the courier

```
In [42]: df['Category']=df['Category'].astype(str)
Column_data=df['Category']
plt.figure(figsize=(10,5))
plt.hist(Column_data,bins=14,edgecolor='Black')
plt.xticks(rotation=90)
plt.show()
```



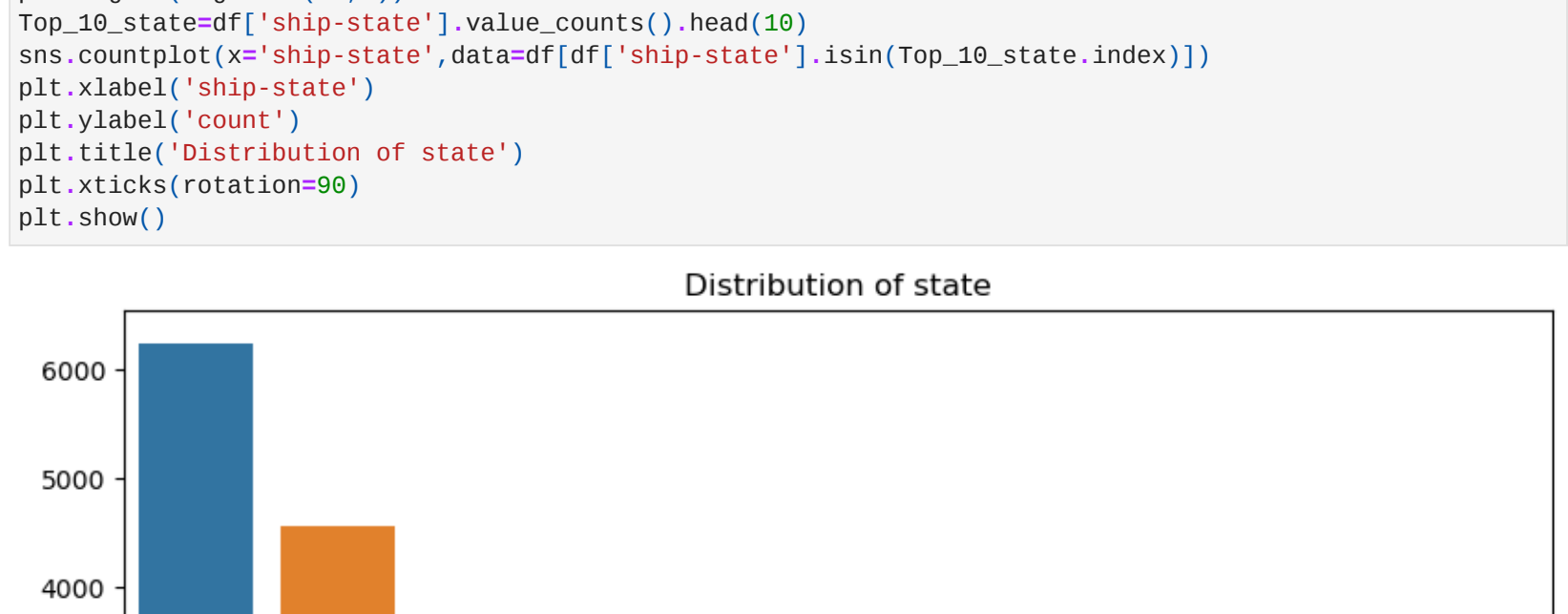
Note:-From above Graph you can see that most of the buyers are T-Shirt

```
In [47]: #checking B2B Data by using pie Chart
B2B_Check = df['B2B'].value_counts()
plt.pie(B2B_Check,labels=B2B_Check,autopct='%1.1f%%')
plt.title("pie plot")
plt.show()
```

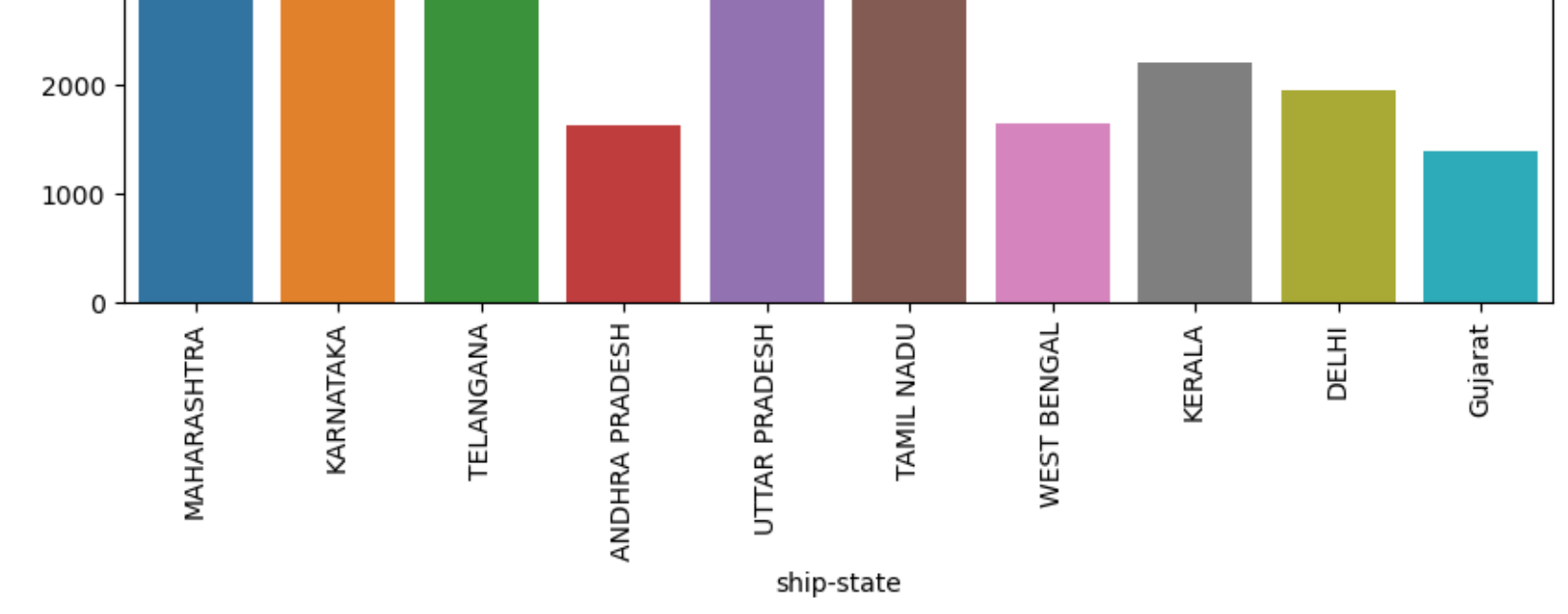


Note:-From above chart we can see that maximum i.e 99.2% of buyers are retailers and 0.8% are B2B buyers

```
In [48]: 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 [52]: plt.figure(figsize=(10,5))
Top_10_state=df['ship-state'].value_counts().head(10)
sns.countplot(x='ship-state',data=df[df['ship-state'].isin(Top_10_state.index)])
plt.xlabel('ship-state')
plt.ylabel('count')
plt.title('Distribution of state')
plt.xticks(rotation=90)
plt.show()
```



Note:-From above Graph you can See that most of buyers are Maharashtra

Conclusion:- The data analysis reveals that business has significant base in maharashtra state,mainly server retailers,fulfills orders through Amazone,experiences demand for T-Shirt,and Size M-Size as the preferred choice among buyers.

```
In [ ] :
```