In [3]:
```
df=pd.read_csv(r'C:\Users\yuvra\Desktop\Amazon Sale Report.csv')
```

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

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

In [5]:
```
df.head()
```

Out[5]:

	index	Order ID	Date	Status	Fulfilment	Sales Channel	ship-service-level	Category	Size	Courier Status	...
0	0	405-8078784-5731545	04-30-22	Cancelled	Merchant	Amazon.in	Standard	T-shirt	S	On the Way	...
1	1	171-9198151-1101146	04-30-22	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	Shirt	3XL	Shipped	...
2	2	404-0687676-7273146	04-30-22	Shipped	Amazon	Amazon.in	Expedited	Shirt	XL	Shipped	...
3	3	403-9615377-8133951	04-30-22	Cancelled	Merchant	Amazon.in	Standard	Blazzer	L	On the Way	...
4	4	407-1069790-7240320	04-30-22	Shipped	Amazon	Amazon.in	Expedited	Trousers	3XL	Shipped	...

5 rows × 21 columns

In [6]:
```
df.tail()
```

128972	128971	9551604-7544318	31-22	Shipped	Amazon	Amazon.in	Expedited	T-shirt	M	Shipp
128973	128972	407-9547469-3152358	05-31-22	Shipped	Amazon	Amazon.in	Expedited	Blazzer	XXL	Shipp
128974	128973	402-6184140-0545956	05-31-22	Shipped	Amazon	Amazon.in	Expedited	T-shirt	XS	Shipp
128975	128974	408-7436540-8728312	05-31-22	Shipped	Amazon	Amazon.in	Expedited	T-shirt	S	Shipp

5 rows × 21 columns

In [7]:
```
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 [8]:
```
df.drop(['New','PendingS'] , axis=1 , inplace=True)
```

In [9]:
```
df.info()
```

```
 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 [10]:
```
pd.isnull(df).sum()
```

Out[10]:
```
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 [11]:
```
df.shape
```

Out[11]:
```
(128976, 19)
```

In [12]:
```
df.dropna(inplace=True)
```

In [13]:
```
df.shape
```

Out[13]:
```
(37514, 19)
```

In [14]:
```
df.columns
```

In [16]:
```
df['ship-postal-code'].dtype
```

Out[16]:
```
dtype('int32')
```

In [17]:
```
df['Date']=pd.to_datetime (df['Date'])
```

In [18]:
```
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 [19]:
```
df.rename(columns={'Qty':'Quantity'})
```

1	1	171-9198151-1101146	2022-04-30	-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	On
7	7	406-7807733-3785945	2022-04-30	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	Shirt	S	Ship
12	12	405-5513694-8146768	2022-04-30	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	Shirt	XS	Ship
...	...	...	...	...	...	...	...	...	...	
128875	128874	405-4724097-1016369	2022-06-01	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	T-shirt	S	Ship
128876	128875	403-9524128-9243508	2022-06-01	Cancelled	Merchant	Amazon.in	Standard	Blazzer	XL	On
128888	128887	405-6493630-8542756	2022-05-31	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	Trousers	M	Ship
128891	128890	407-0116398-1810752	2022-05-31	Cancelled	Merchant	Amazon.in	Standard	Wallet	Free	On
128892	128891	403-0317423-9322704	2022-05-31	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	Blazzer	M	Ship

37514 rows × 19 columns

In [20]:
```
df.describe()
```

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 [21]:
```
df.describe(include='object')
```

Out[21]:

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

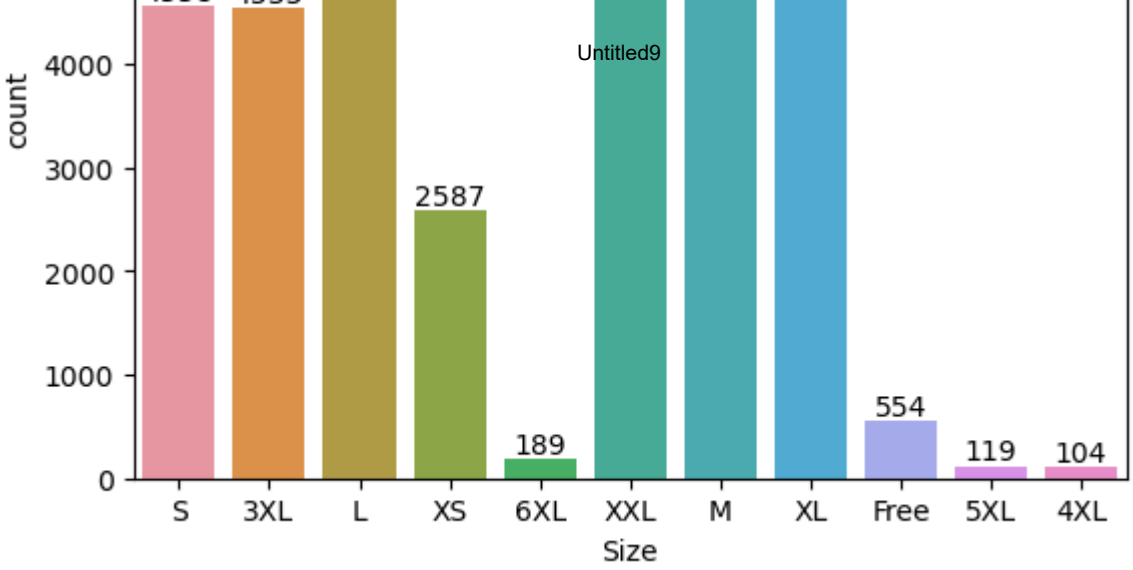
In [22]:
```
df[['Qty','Amount']].describe()
```

Out[22]:

	Qty	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

In [23]:
```
ax=sns.countplot(x='Size',data=df)

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



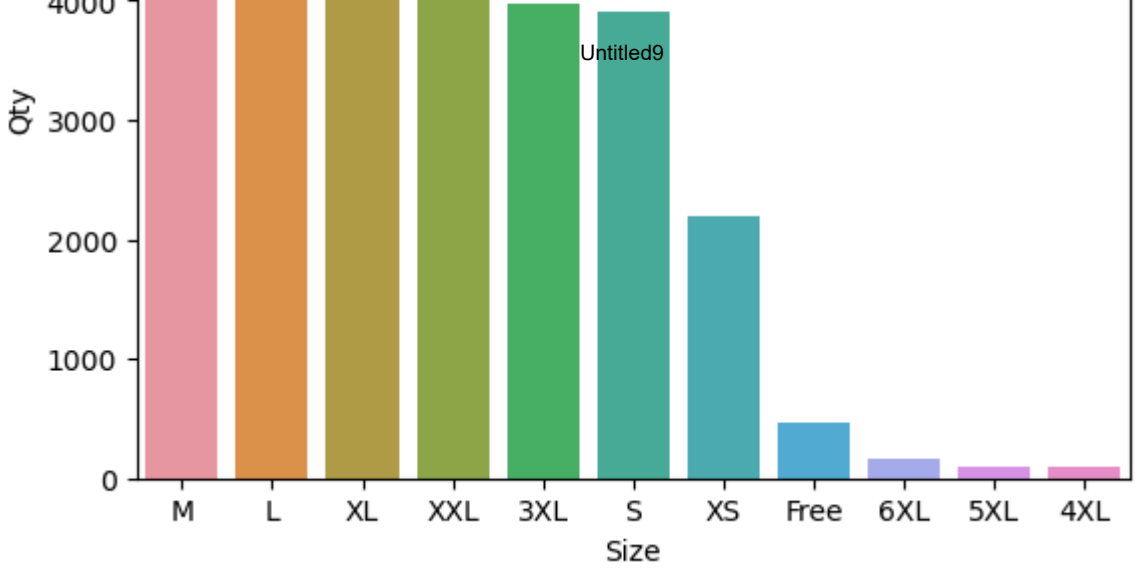
In [37]:
```
df.groupby(['Size'], as_index=False)['Qty'].sum().sort_values(by='Qty', ascending=F
```

Out[37]:

	Size	Qty
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

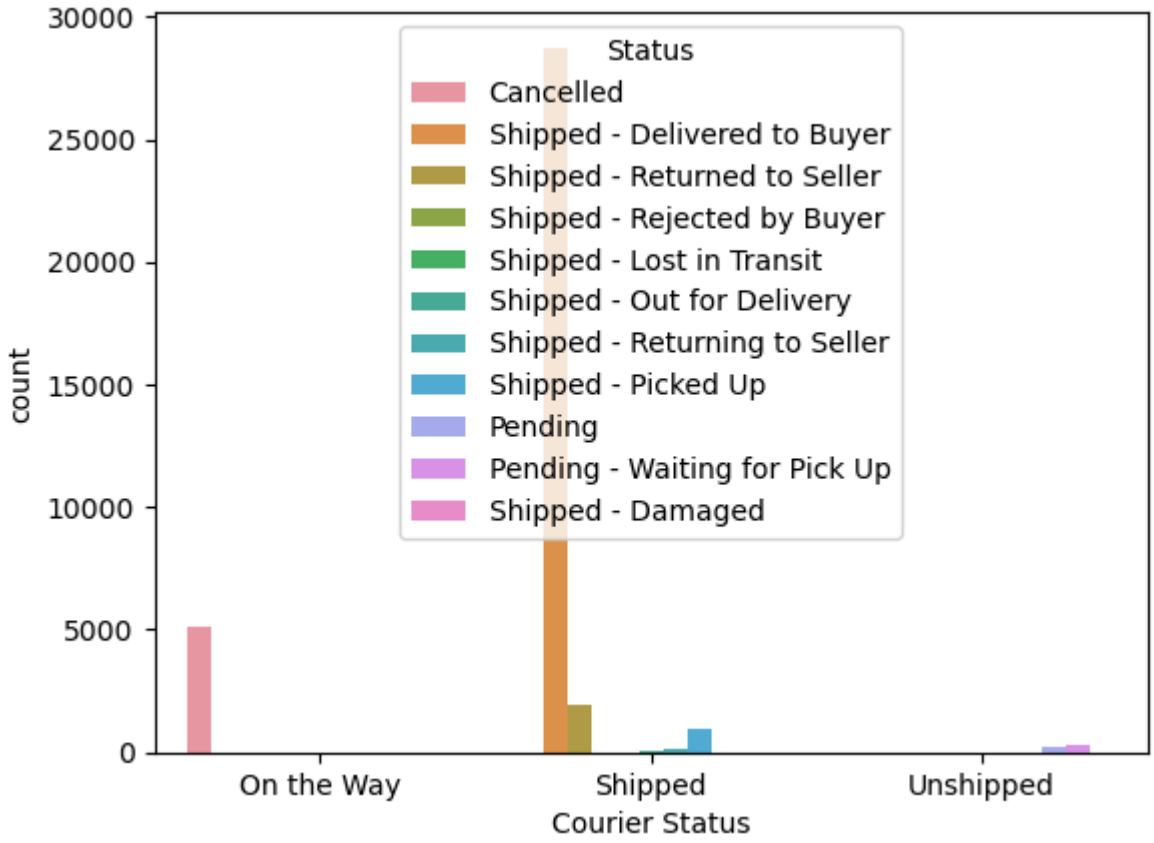
In [24]:
```
S_QTY=df.groupby(['Size'], as_index=False)['Qty'].sum().sort_values(by='Qty', ascer
sns.barplot(x='Size',y='Qty',data=S_QTY)
```

Out[24]:
```
<AxesSubplot:xlabel='Size', ylabel='Qty'>
```

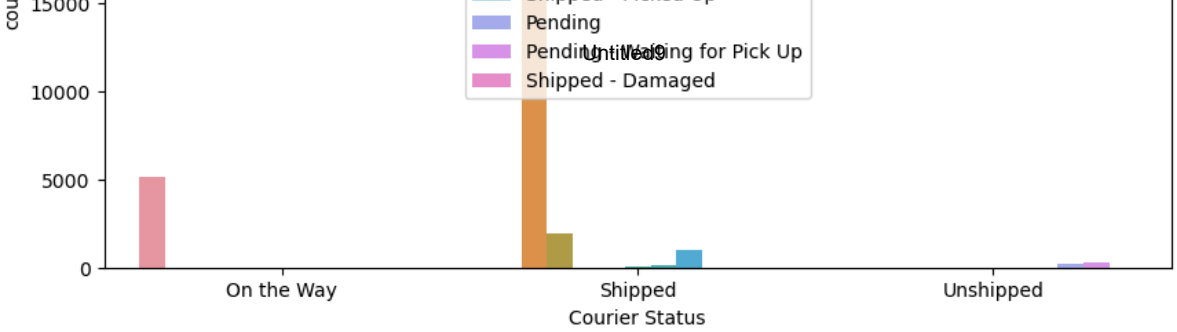


In [26]:
```
sns.countplot(data=df,x='Courier Status',hue='Status')
```

Out[26]:
```
<AxesSubplot:xlabel='Courier Status', ylabel='count'>
```

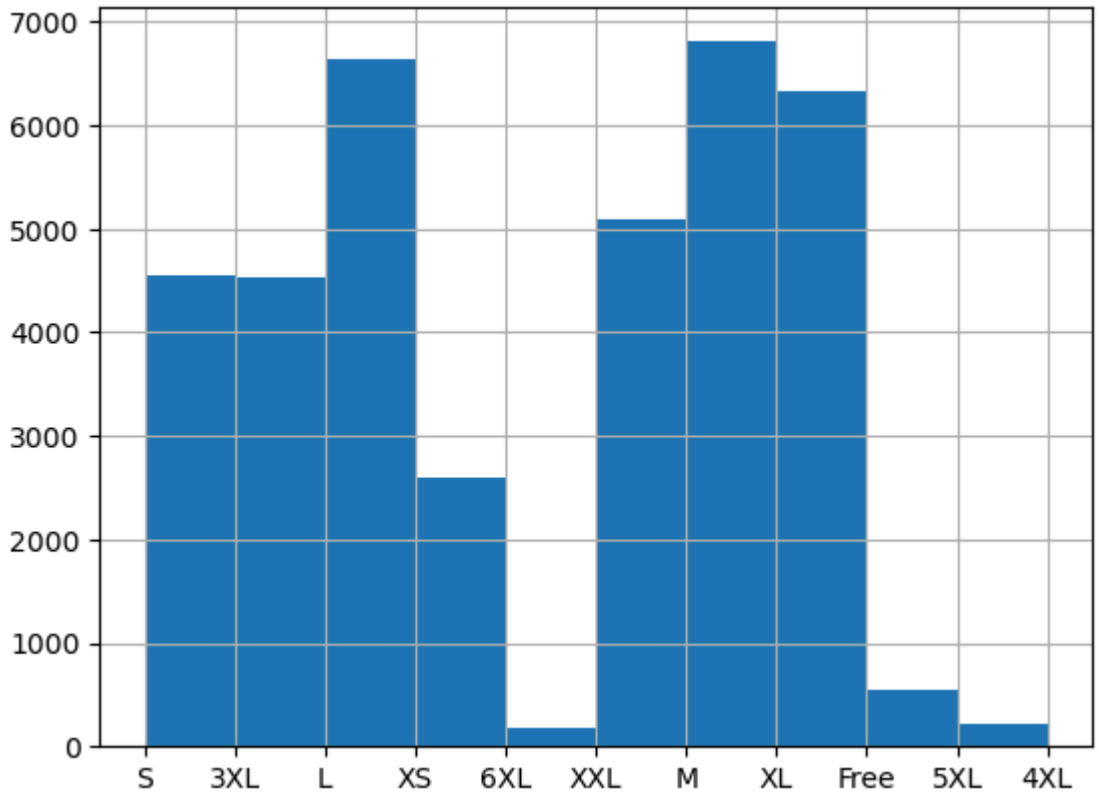


In [28]:
```
plt.figure(figsize=(10,5))
ax=sns.countplot(data=df,x='Courier Status',hue='Status')
plt.show()
```

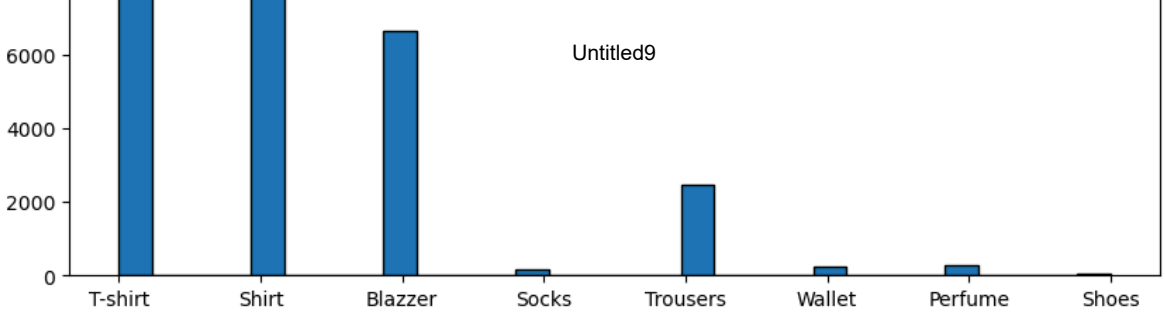


In [30]:
```
df['Size'].hist()
```

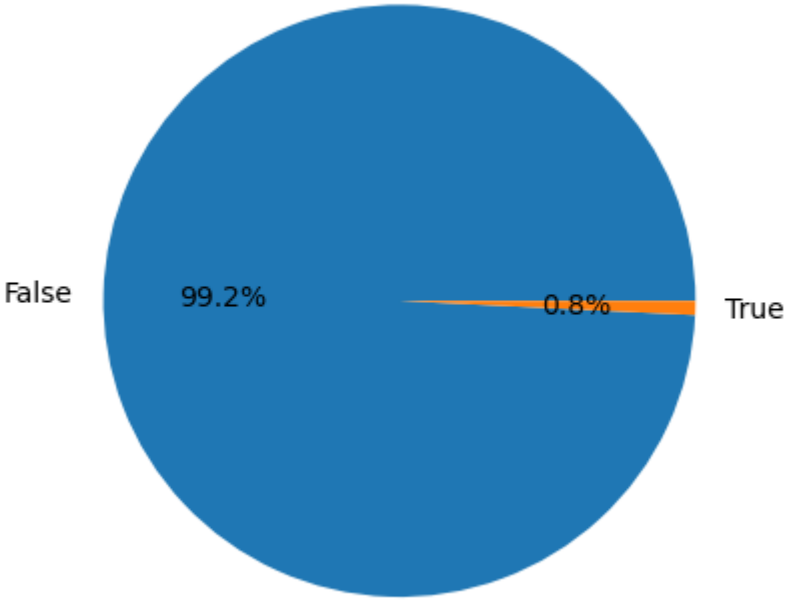
Out[30]: <AxesSubplot:>



In [35]:
```
df['Category']=df['Category'].astype(str)
column_data=df['Category']
plt.figure(figsize=(10, 5))
plt.hist(column_data,bins=30,edgecolor='black')
plt.show()
```

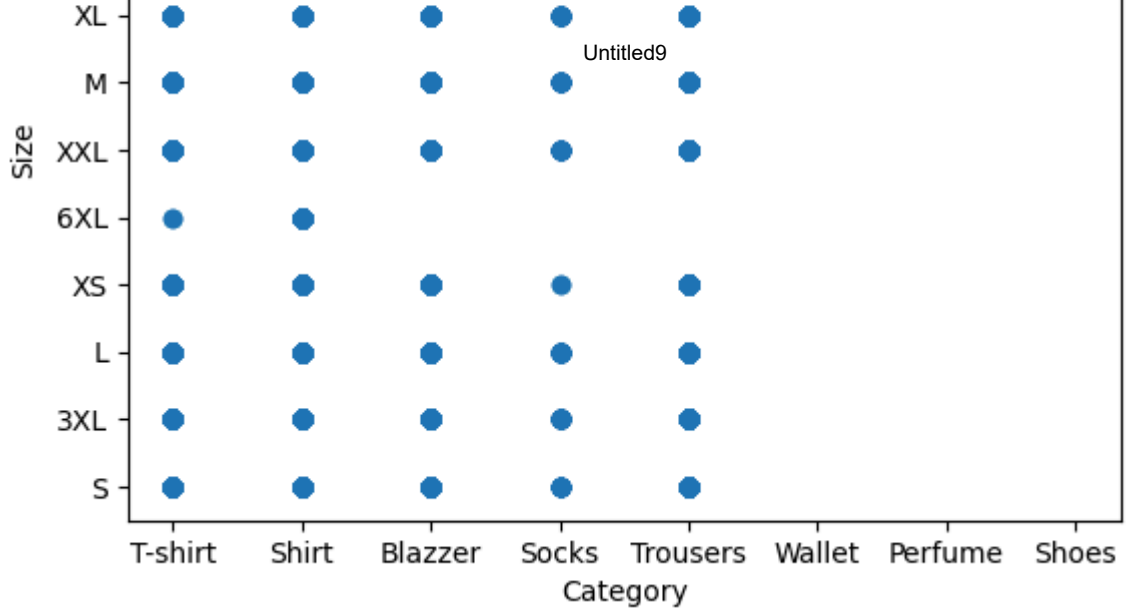


In [37]:
```
B2B_Check=df['B2B'].value_counts()
plt.pie(B2B_Check,labels=B2B_Check.index,autopct='%1.1f%%')
plt.show()
```

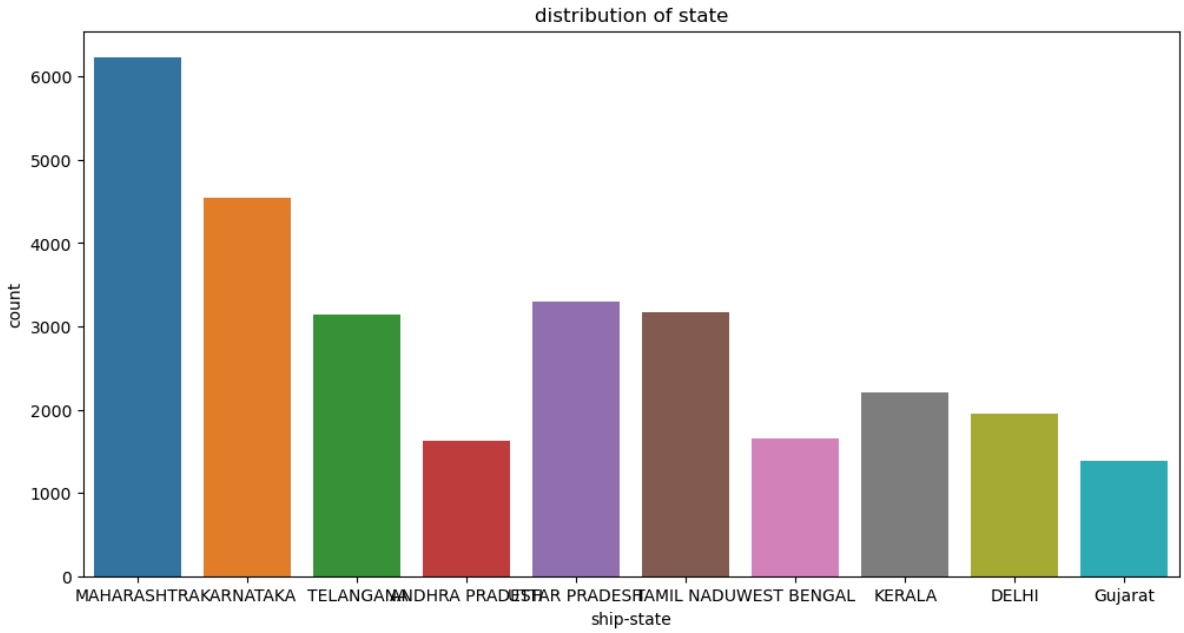


In [38]:
```
x_data=df['Category']
y_data=df['Size']

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



In [43]:
```
top10state=df['ship-state'].value_counts().head(10)
filtered_df = df[df['ship-state'].isin(top10state.index)]
plt.figure(figsize=(12,6))
sns.countplot(data=filtered_df, x='ship-state')
plt.xlabel('ship-state')
plt.ylabel('count')
plt.title('distribution of state')
plt.show()
```



In [ ]: