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

In [2]:
```python
df=pd.read_csv("Amazon.csv")
```

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

Out[3]:

	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	New	PendingS
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	NaN	NaN
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	NaN	NaN
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	NaN	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	NaN	NaN
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	NaN	NaN

5 rows × 21 columns

In [4]:
```python
df.tail()
```

Out[4]:

	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	New	PendingS
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	NaN	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	NaN	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	NaN	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	NaN	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	NaN	NaN

5 rows × 21 columns

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

Out[5]:
```
(128976, 21)
```

In [6]:
```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 [7]:
```python
#drop blanks columns
df.drop(['New','PendingS'],axis=1,inplace=True)
```

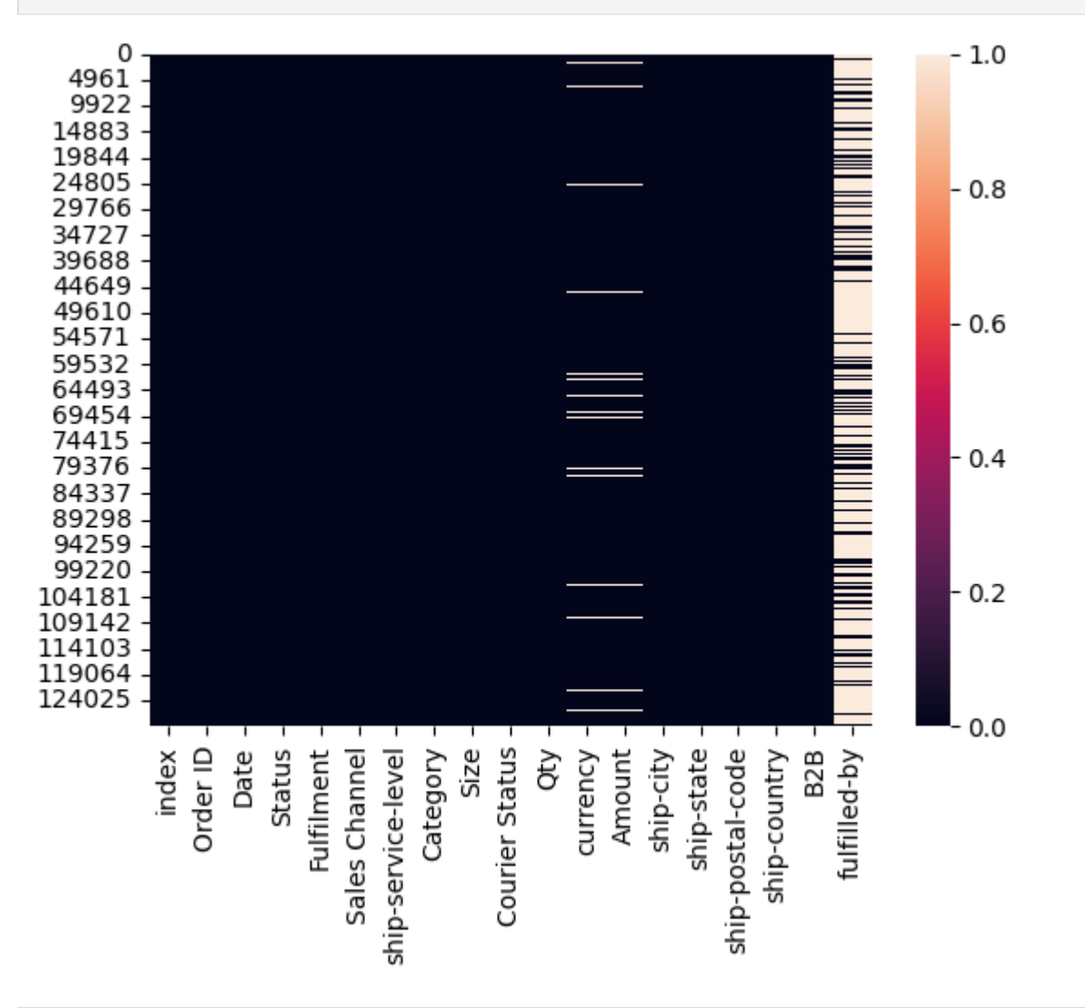
In [8]:
```python
df.info()
```

```
<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 [9]:
```python
df.isnull().sum()
```

Out[9]:
```
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 [10]:
```python
sns.heatmap(df.isnull())
plt.show()
```



In [11]:
```python
df.drop(['fulfilled-by'],axis=1,inplace=True)
df.dropna(inplace=True)
```

In [12]:
```python
df.isnull().sum()
```

Out[12]:
```
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
dtype: int64
```

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

Out[13]:
```
(121143, 18)
```

In [14]:
```python
df['ship-postal-code']=df['ship-postal-code'].astype(int)
```

In [15]:
```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 121143 entries, 0 to 128975
Data columns (total 18 columns):
 #   Column              Non-Null Count   Dtype  
---  ------              --------------   -----  
 0   index               121143 non-null  int64  
 1   Order ID            121143 non-null  object 
 2   Date                121143 non-null  object 
 3   Status              121143 non-null  object 
 4   Fulfilment          121143 non-null  object 
 5   Sales Channel       121143 non-null  object 
 6   ship-service-level  121143 non-null  object 
 7   Category            121143 non-null  object 
 8   Size                121143 non-null  object 
 9   Courier Status      121143 non-null  object 
 10  Qty                 121143 non-null  int64  
 11  currency            121143 non-null  object 
 12  Amount              121143 non-null  float64
 13  ship-city           121143 non-null  object 
 14  ship-state          121143 non-null  object 
 15  ship-postal-code    121143 non-null  int32  
 16  ship-country        121143 non-null  object 
 17  B2B                 121143 non-null  bool   
dtypes: bool(1), float64(1), int32(1), int64(2), object(13)
memory usage: 16.3+ MB
```

In [16]:
```python
df['Date']=pd.to_datetime(df['Date'])
```

In [17]:
```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 121143 entries, 0 to 128975
Data columns (total 18 columns):
 #   Column              Non-Null Count   Dtype         
---  ------              --------------   -----         
 0   index               121143 non-null  int64         
 1   Order ID            121143 non-null  object        
 2   Date                121143 non-null  datetime64[ns]
 3   Status              121143 non-null  object        
 4   Fulfilment          121143 non-null  object        
 5   Sales Channel       121143 non-null  object        
 6   ship-service-level  121143 non-null  object        
 7   Category            121143 non-null  object        
 8   Size                121143 non-null  object        
 9   Courier Status      121143 non-null  object        
 10  Qty                 121143 non-null  int64         
 11  currency            121143 non-null  object        
 12  Amount              121143 non-null  float64       
 13  ship-city           121143 non-null  object        
 14  ship-state          121143 non-null  object        
 15  ship-postal-code    121143 non-null  int32         
 16  ship-country        121143 non-null  object        
 17  B2B                 121143 non-null  bool          
dtypes: bool(1), datetime64[ns](1), float64(1), int32(1), int64(2), object(12)
memory usage: 16.3+ MB
```

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

Out[18]:

	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
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
2	2	404-0687676-7273146	2022-04-30	Shipped	Amazon	Amazon.in	Expedited	Shirt	XL	Shipped	1	INR	329.00	NAVI MUMBAI	MAHARASHTRA	410210	IN	True
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
4	4	407-1069790-7240320	2022-04-30	Shipped	Amazon	Amazon.in	Expedited	Trousers	3XL	Shipped	1	INR	574.00	CHENNAI	TAMIL NADU	600073	IN	False

In [36]:
```python
df.rename(columns={'Qty':'Quantity'},inplace=True)
```

In [37]:
```python
df.describe()
```

Out[37]:

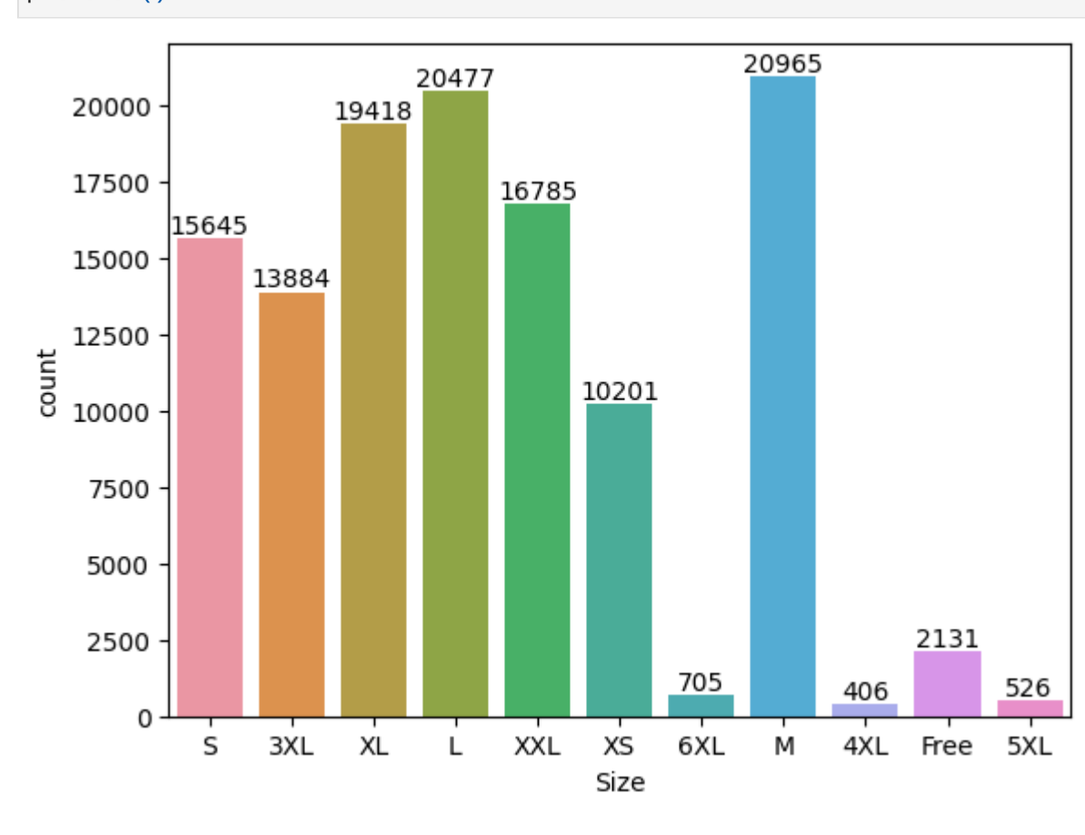
	index	Quantity	Amount	ship-postal-code
count	121143.000000	121143.000000	121143.000000	121143.000000
mean	64486.312655	0.961252	648.576874	463623.724507
std	37220.415404	0.214276	281.196896	191301.588170
min	0.000000	0.000000	0.000000	110001.000000
25%	32294.500000	1.000000	449.000000	382421.000000
50%	64477.000000	1.000000	605.000000	500032.000000
75%	96682.500000	1.000000	788.000000	600020.000000
max	128974.000000	8.000000	5584.000000	989898.000000

In [20]:
```python
df.describe(include=object)
```

Out[20]:

	Order ID	Status	Fulfilment	Sales Channel	ship-service-level	Category	Size	Courier Status	currency	ship-city	ship-state	ship-country
count	121143	121143	121143	121143	121143	121143	121143	121143	121143	121143	121143	121143
unique	112861	12	2	1	2	9	11	3	1	8697	68	1
top	171-5057375-2831560	Shipped	Amazon	Amazon.in	Expedited	T-shirt	M	Shipped	INR	BENGALURU	MAHARASHTRA	IN
freq	12	77589	83629	121143	82713	47038	20965	109458	121143	10675	21084	121143

In [39]:
```python
ax=sns.countplot(x='Size',data=df)
ax.bar_label(ax.containers[0])
plt.show()
```



Note:From above graph you can see that most people buys M-size

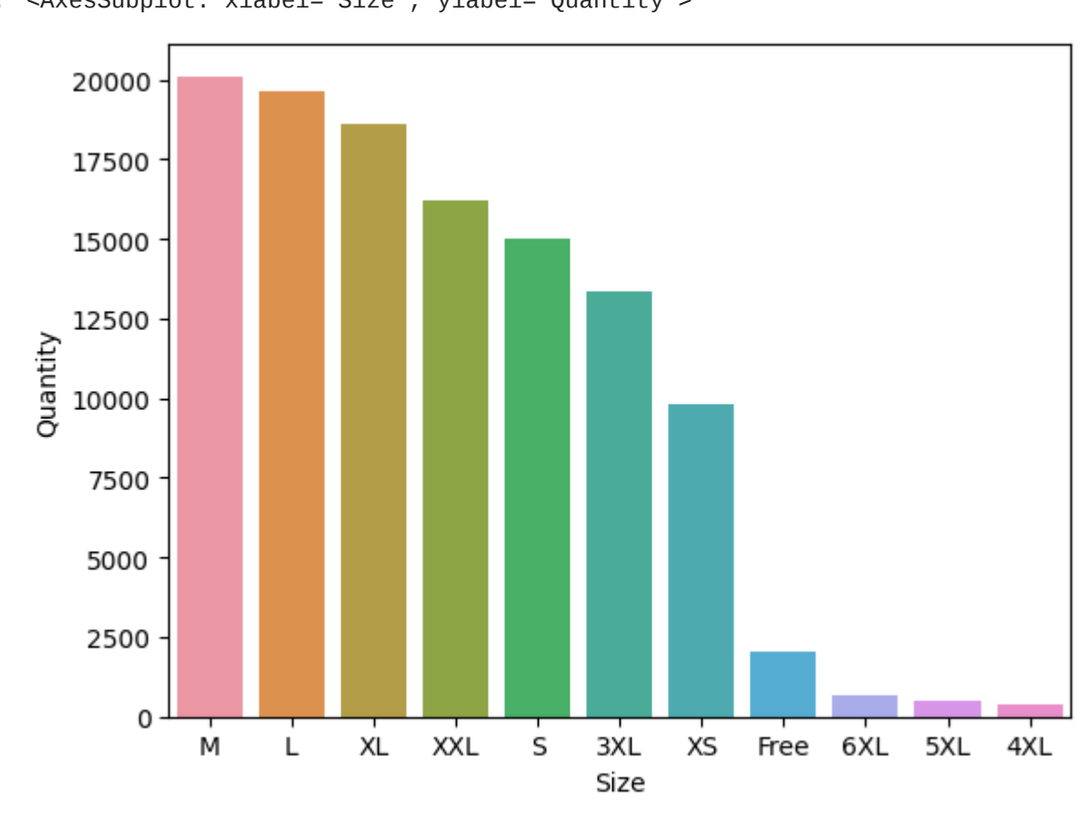
In [42]:
```python
df1=df.groupby(["Size"],as_index=False)['Quantity'].sum().sort_values(by='Quantity',ascending=False)
df1
```

Out[42]:

	Size	Quantity
6	M	20113
5	L	19667
8	XL	18611
10	XXL	16217
7	S	15019
0	3XL	13341
9	XS	9835
4	Free	2051
3	6XL	687
2	5XL	512
1	4XL	396

In [44]:
```python
sns.barplot(x='Size',y='Quantity',data=df1)
```

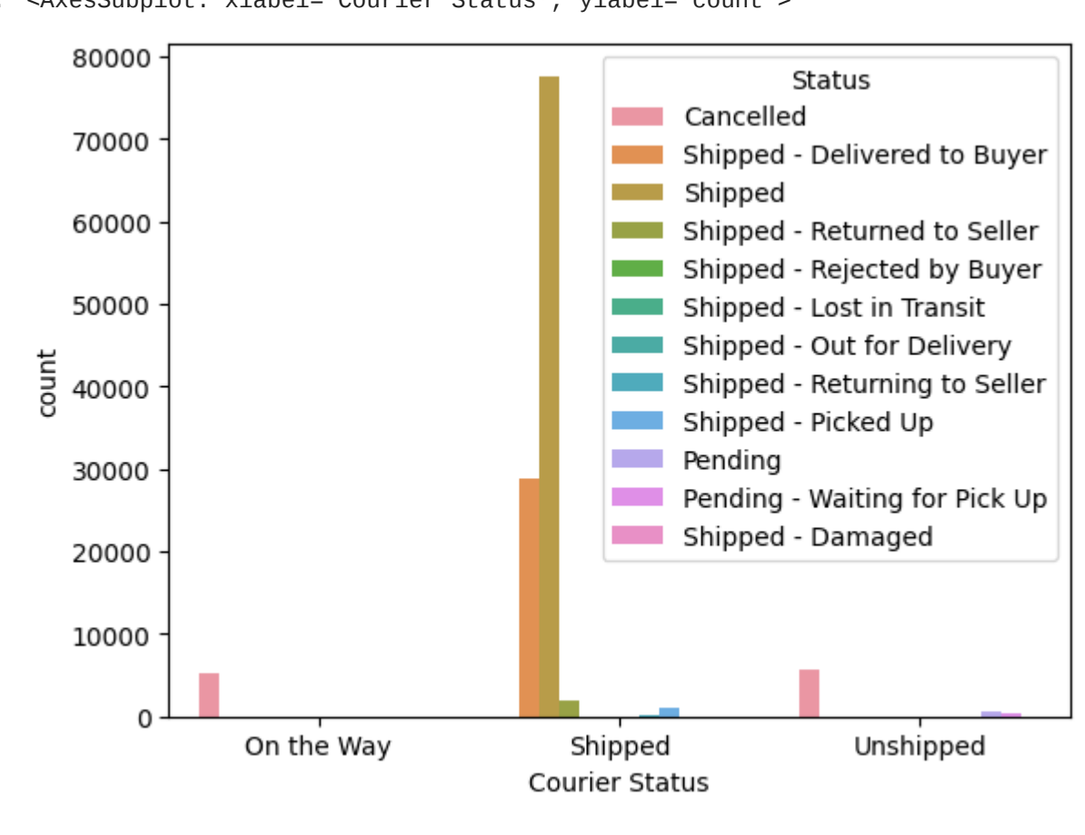
Out[44]:
```
<AxesSubplot: xlabel='Size', ylabel='Quantity'>
```



most of quentity buys m-size in the sales

In [23]:
```python
sns.countplot(x='Courier Status',data=df,hue='Status')
```

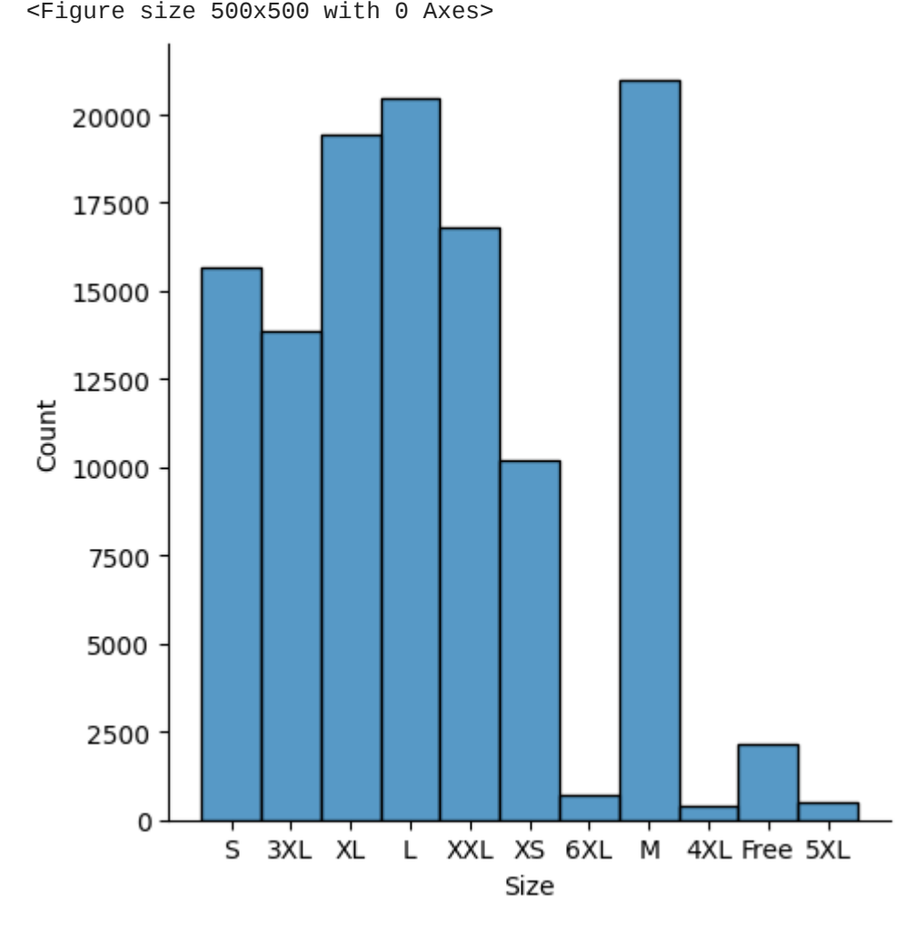
Out[23]:
```
<AxesSubplot: xlabel='Courier Status', ylabel='count'>
```



majorly of the orders are shipped by the courier.

In [24]:
```python
plt.figure(figsize=(5,5))
sns.displot(df['Size'])
plt.show()
```

```
<Figure size 500x500 with 0 Axes>
```

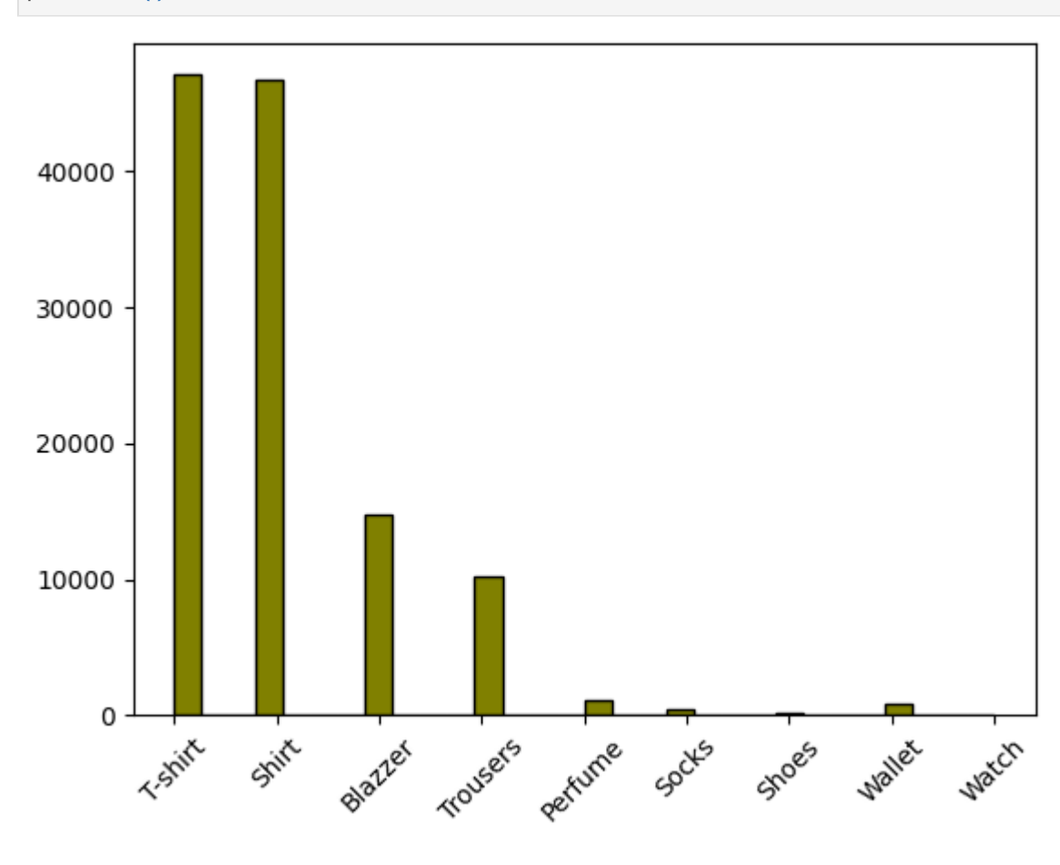


In [25]:
```python
df['Category']=df['Category'].astype(str)
df['Category']
```

In [26]:
```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 121143 entries, 0 to 128975
Data columns (total 18 columns):
 #   Column              Non-Null Count   Dtype         
---  ------              --------------   -----         
 0   index               121143 non-null  int64         
 1   Order ID            121143 non-null  object        
 2   Date                121143 non-null  datetime64[ns]
 3   Status              121143 non-null  object        
 4   Fulfilment          121143 non-null  object        
 5   Sales Channel       121143 non-null  object        
 6   ship-service-level  121143 non-null  object        
 7   Category            121143 non-null  object        
 8   Size                121143 non-null  object        
 9   Courier Status      121143 non-null  object        
 10  Qty                 121143 non-null  int64         
 11  currency            121143 non-null  object        
 12  Amount              121143 non-null  float64       
 13  ship-city           121143 non-null  object        
 14  ship-state          121143 non-null  object        
 15  ship-postal-code    121143 non-null  int32         
 16  ship-country        121143 non-null  object        
 17  B2B                 121143 non-null  bool          
dtypes: bool(1), datetime64[ns](1), float64(1), int32(1), int64(2), object(12)
memory usage: 16.3+ MB
```

In [47]:
```python
plt.hist(df['Category'],bins=30,color='olive',edgecolor='black')
plt.xticks(rotation=45)
plt.show()
```

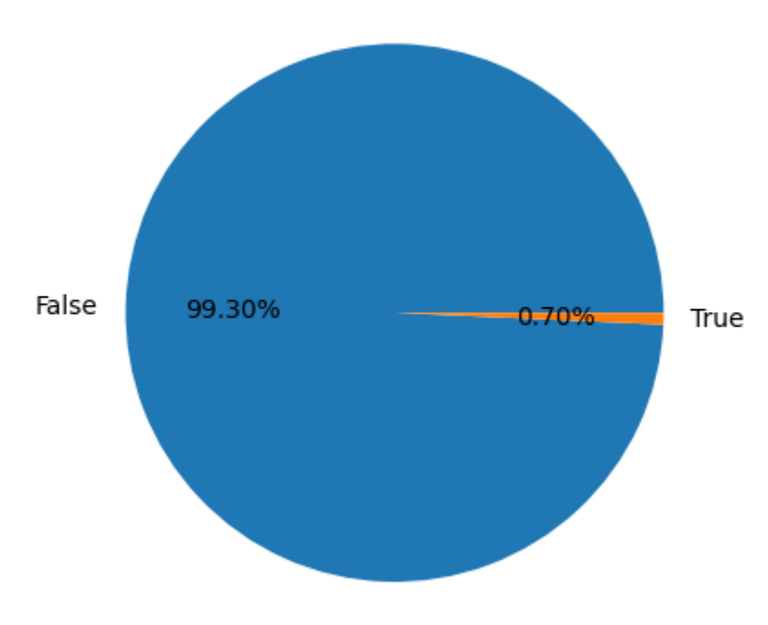


most of the buyers are T-shart and shart

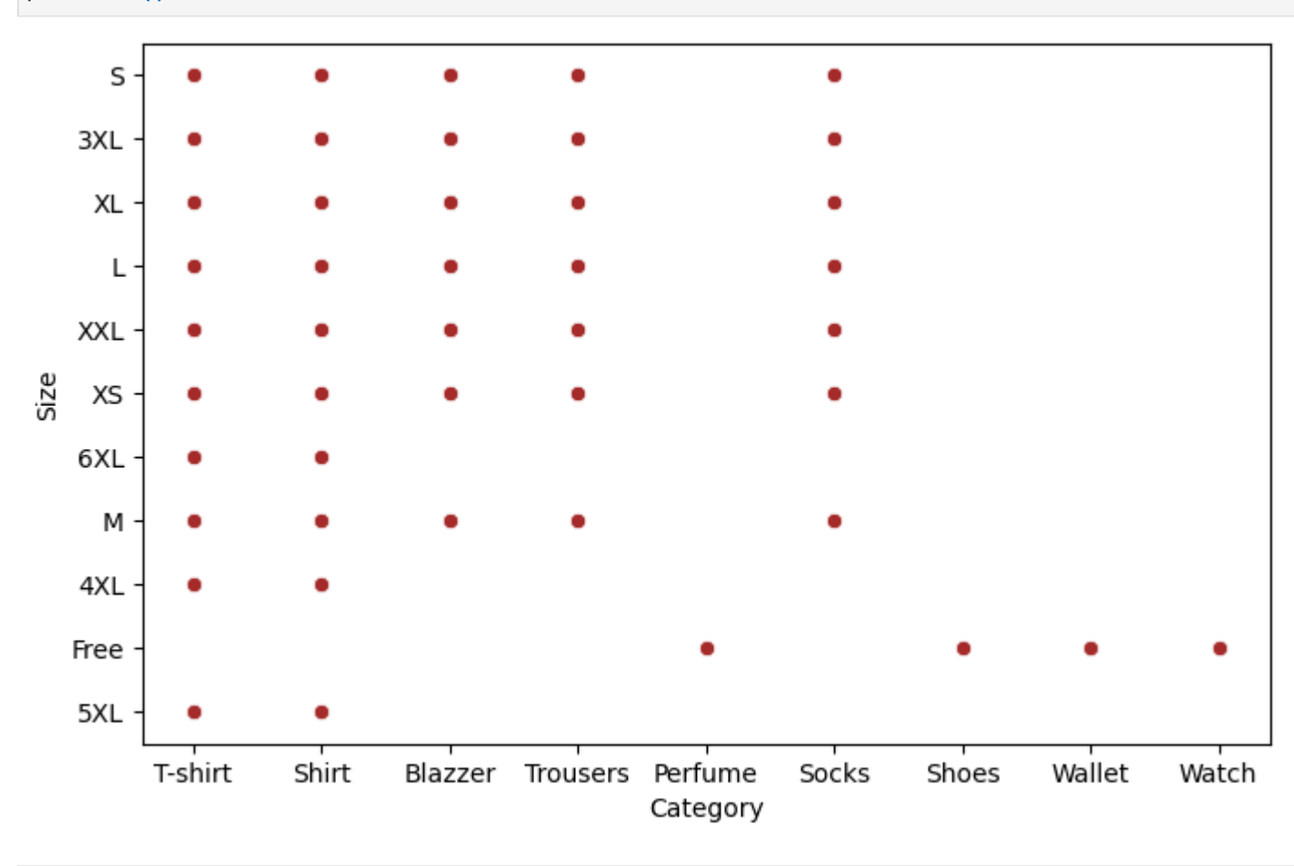
In [28]:
```python
p=df['B2B'].value_counts(normalize=True)
p
```

Out[28]:
```
False    0.993033
True     0.006967
Name: B2B, dtype: float64
```

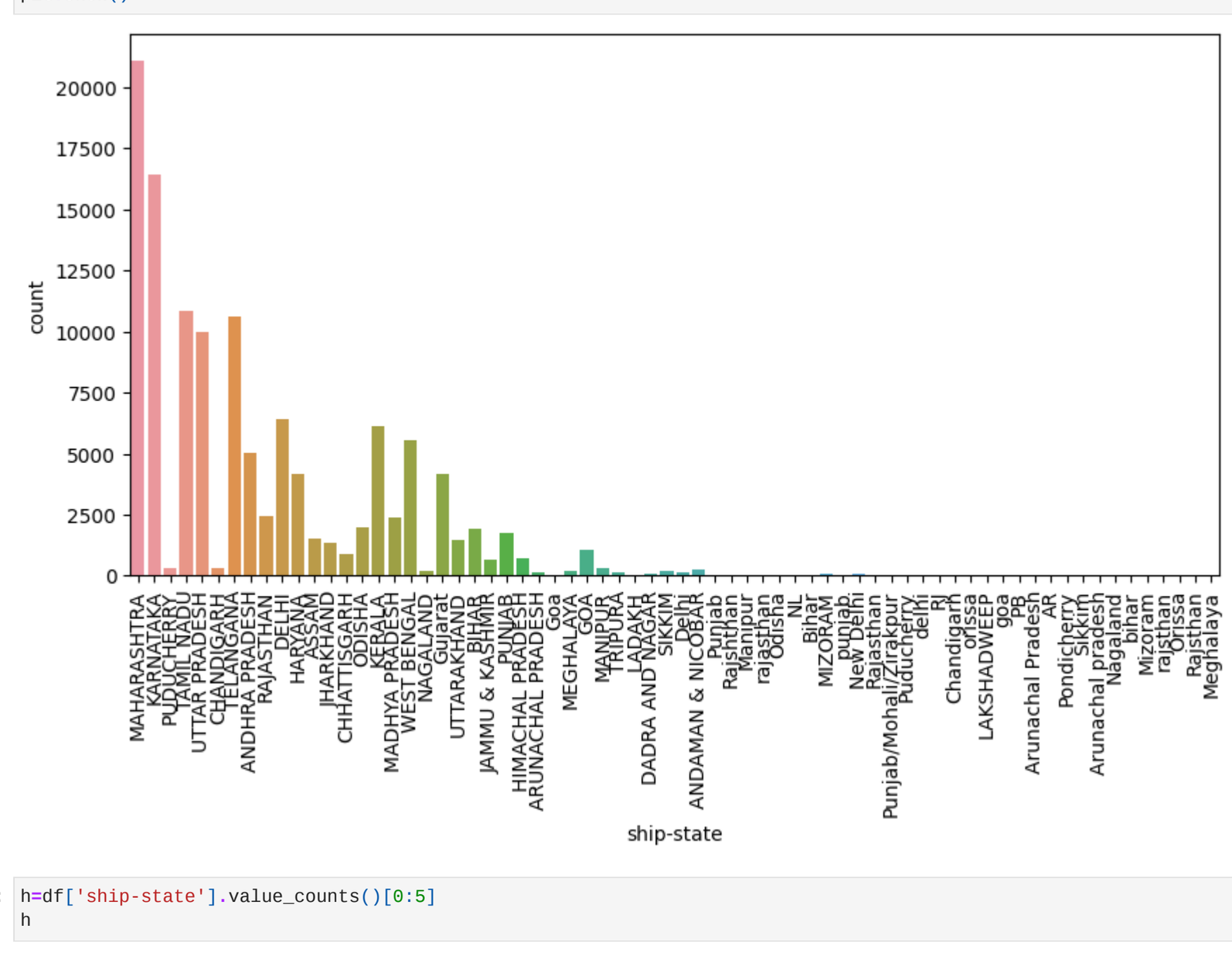
In [30]:
```python
plt.pie(p,labels=p.index,autopct='%0.2f%%')
plt.show()
```



In [53]:
```python
plt.figure(figsize=(8,5))
sns.scatterplot(x='Category',y='Size',data=df,color='brown')
plt.show()
```



In [32]:
```python
plt.figure(figsize=(10,5))
sns.countplot(x='ship-state',data=df)
plt.xticks(rotation=90)
plt.show()
```



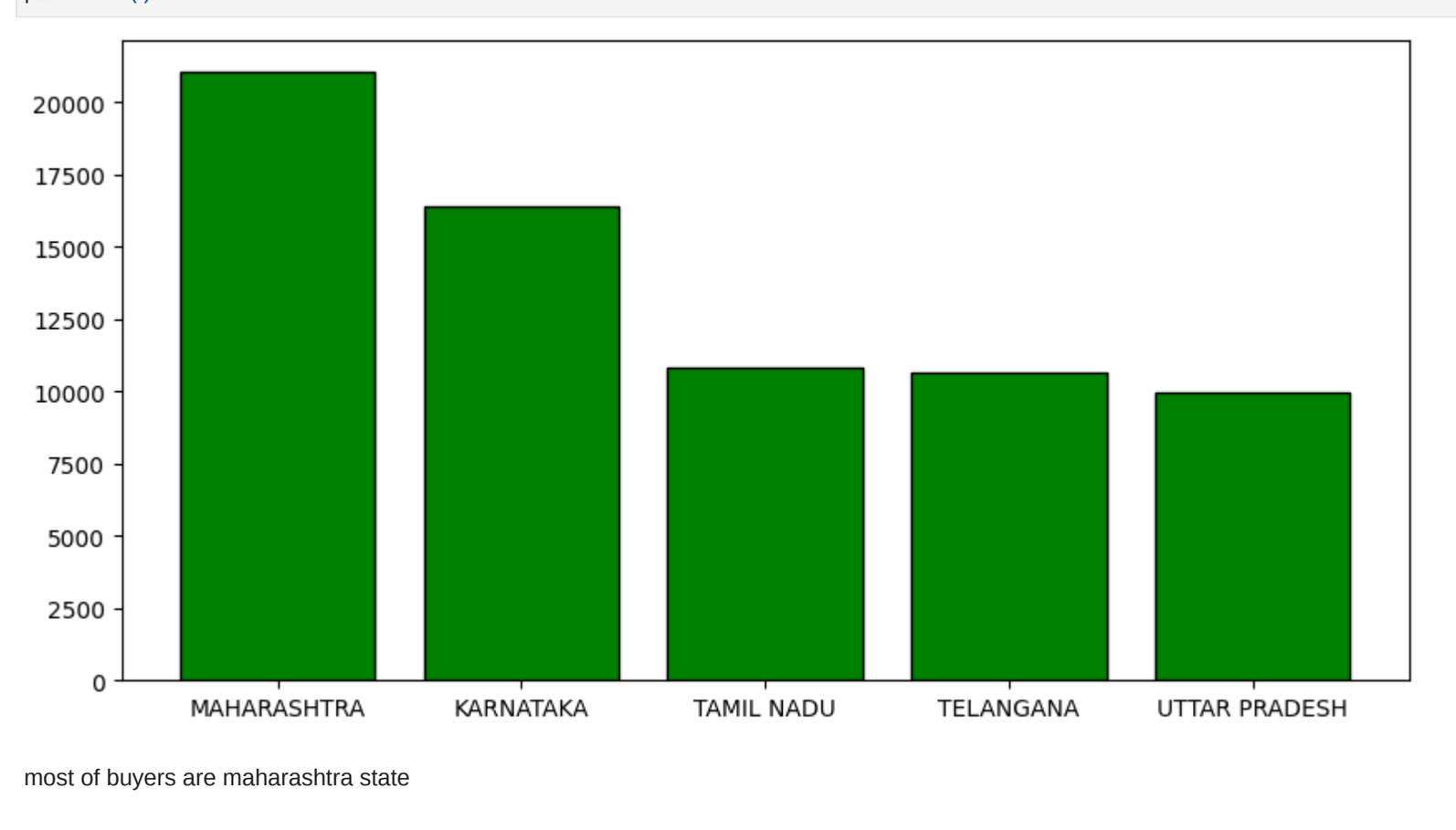
In [33]:
```python
h=df['ship-state'].value_counts()[0:5]
h
```

Out[33]:
```
MAHARASHTRA      21084
KARNATAKA        16390
TAMIL NADU       10813
TELANGANA        10636
UTTAR PRADESH     9956
Name: ship-state, dtype: int64
```

In [34]:
```python
v=list(df['ship-state'].value_counts()[0:5].keys())
v
```

Out[34]:
```
['MAHARASHTRA', 'KARNATAKA', 'TAMIL NADU', 'TELANGANA', 'UTTAR PRADESH']
```

In [35]:
```python
plt.figure(figsize=(10,5))
plt.bar(v,h,color='g',edgecolor='black')
plt.show()
```



most of buyers are maharashtra state

conclusion

The data analysis reveals that the business has a significant customer base in maharastra state ,mainly serves retailers, fulfills orders through amazon,experiences high demand for T-sharts,and see M-size as the preferrs most of buyers.

In [ ]: