```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline

df=pd.read_csv('Amazon Sale Report.csv')

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
```

```
df.head()
```

```
   index             Order ID      Date                            Status
\
0      0  405-8078784-5731545  04-30-22                         Cancelled

1      1  171-9198151-1101146  04-30-22  Shipped - Delivered to Buyer

2      2  404-0687676-7273146  04-30-22                           Shipped

3      3  403-9615377-8133951  04-30-22                         Cancelled
```

```
4      4  407-1069790-7240320  04-30-22                          Shipped


  Fulfilment Sales Channel ship-service-level  Category Size Courier
Status  \
0   Merchant     Amazon.in           Standard   T-shirt    S     On
the Way
1   Merchant     Amazon.in           Standard     Shirt  3XL
Shipped
2     Amazon     Amazon.in          Expedited     Shirt   XL
Shipped
3   Merchant     Amazon.in           Standard   Blazzer    L     On
the Way
4     Amazon     Amazon.in          Expedited  Trousers  3XL
Shipped

   ...  currency  Amount    ship-city   ship-state ship-postal-code  \
0  ...       INR  647.62       MUMBAI  MAHARASHTRA         400081.0
1  ...       INR  406.00    BENGALURU    KARNATAKA         560085.0
2  ...       INR  329.00  NAVI MUMBAI  MAHARASHTRA         410210.0
3  ...       INR  753.33   PUDUCHERRY   PUDUCHERRY         605008.0
4  ...       INR  574.00      CHENNAI   TAMIL NADU         600073.0

   ship-country    B2B  fulfilled-by New  PendingS
0            IN  False     Easy Ship NaN       NaN
1            IN  False     Easy Ship NaN       NaN
2            IN   True           NaN NaN       NaN
3            IN  False     Easy Ship NaN       NaN
4            IN  False           NaN NaN       NaN

[5 rows x 21 columns]
```

```
df.drop(['New','PendingS'], axis=1, inplace=True) #dropping
unnecessary columns

df.isnull() #checking null values
```

```
        index  Order ID   Date  Status  Fulfilment  Sales Channel  \
0       False     False  False   False       False          False
1       False     False  False   False       False          False
2       False     False  False   False       False          False
3       False     False  False   False       False          False
4       False     False  False   False       False          False
...       ...       ...    ...     ...         ...            ...
128971  False     False  False   False       False          False
128972  False     False  False   False       False          False
128973  False     False  False   False       False          False
128974  False     False  False   False       False          False
128975  False     False  False   False       False          False

        ship-service-level  Category   Size  Courier Status    Qty
```

```
currency  \
0                     False     False  False           False  False
False
1                     False     False  False           False  False
False
2                     False     False  False           False  False
False
3                     False     False  False           False  False
False
4                     False     False  False           False  False
False
...                     ...       ...    ...             ...    ...
...
128971                False     False  False           False  False
False
128972                False     False  False           False  False
False
128973                False     False  False           False  False
False
128974                False     False  False           False  False
False
128975                False     False  False           False  False
False

        Amount  ship-city  ship-state  ship-postal-code  ship-country
B2B  \
0        False      False       False             False         False
False
1        False      False       False             False         False
False
2        False      False       False             False         False
False
3        False      False       False             False         False
False
4        False      False       False             False         False
False
...        ...        ...         ...               ...           ...
...
128971   False      False       False             False         False
False
128972   False      False       False             False         False
False
128973   False      False       False             False         False
False
128974   False      False       False             False         False
False
128975   False      False       False             False         False
False
```

```
        fulfilled-by
0              False
1              False
2               True
3              False
4               True
...              ...
128971          True
128972          True
128973          True
128974          True
128975          True

[128976 rows x 19 columns]
```

```
df.isnull().sum() #calculating the total number of null values
according to column wise
```

```
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
```

```
df.dropna(inplace = True) #dropping null values
```

```
df.shape
```

```
(37514, 19)
```

```
df.isnull().sum() #no missing values present after dropping na values
```

```
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
```

```
# change data type
df['ship-postal-code']=df['ship-postal-code'].astype('int')

#checking whether the  data type change or not
df['ship-postal-code'].dtype

dtype('int32')

df['Date']=pd.to_datetime (df['Date']) #change data type

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
```

```
df.describe() #summary statistics
```

```
                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
```

```
df.describe(include='object')
```

```
                    Order ID                         Status
Fulfilment  \
count                 37514                          37514      37514

unique                34664                             11          1

top     171-5057375-2831560  Shipped - Delivered to Buyer   Merchant

freq                     12                          28741      37514


       Sales Channel ship-service-level Category   Size Courier Status
\
count          37514               37514    37514  37514          37514

unique             1                   1        8     11              3

top        Amazon.in            Standard  T-shirt      M        Shipped

freq           37514               37514    14062   6806          31859


       currency  ship-city   ship-state ship-country fulfilled-by
count     37514      37514        37514        37514        37514
unique        1       4698           58            1            1
top         INR  BENGALURU  MAHARASHTRA           IN    Easy Ship
freq      37514       2839         6236        37514        37514
```

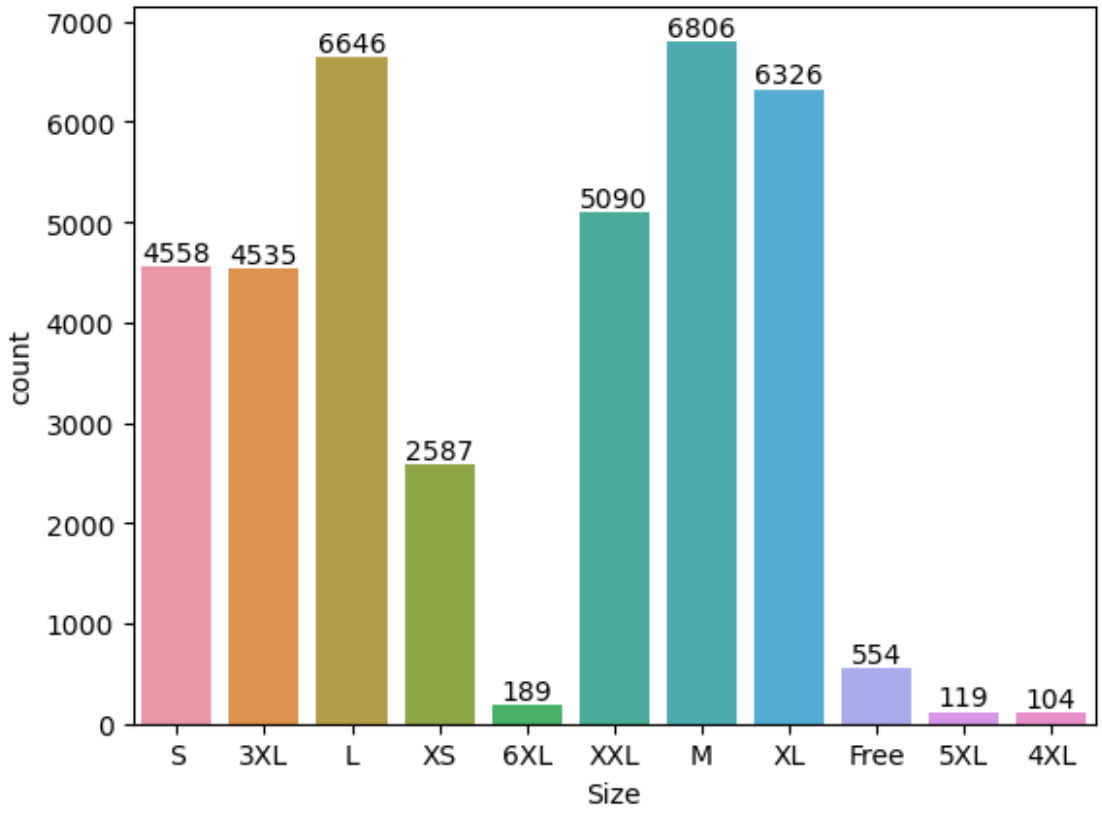
```
df.columns
```

```
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')
```

Exploratory Data Analysis

```
ax=sns.countplot(x='Size' ,data=df)
for bars in ax.containers:
    ax.bar_label(bars) ## Counting and displaying the total number of
counts according to size
```

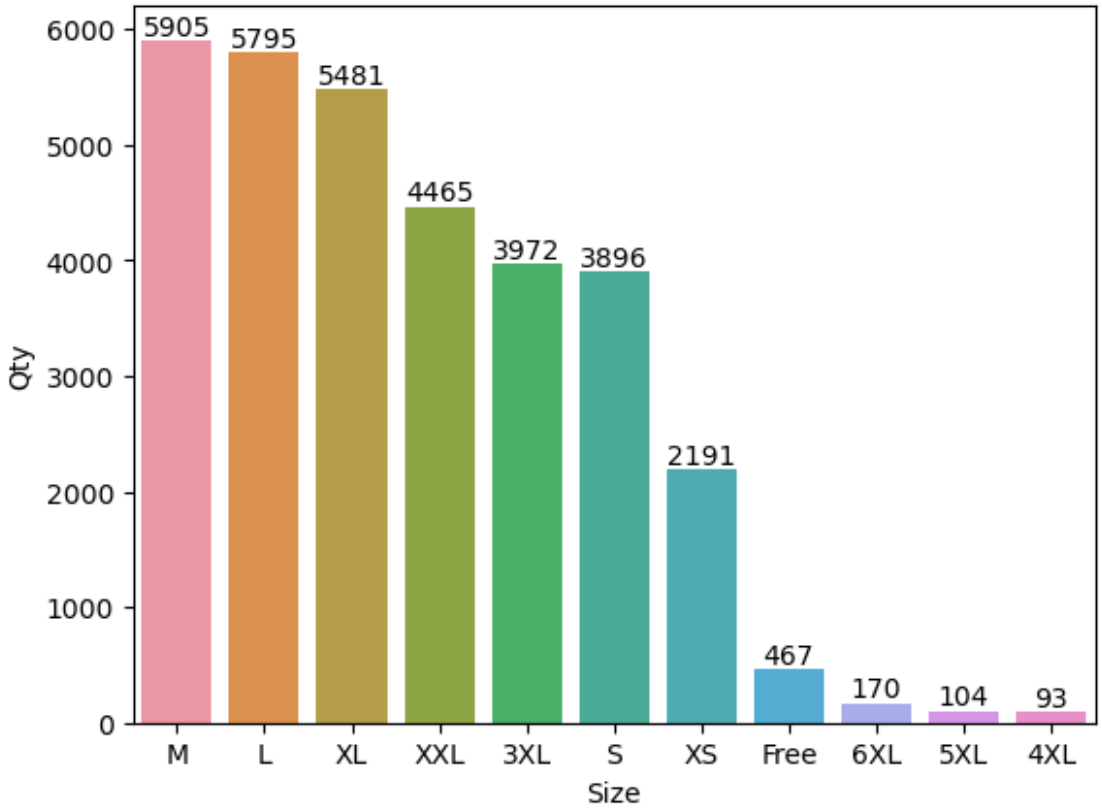



```
df.groupby('Size')['Qty'].sum().sort_values(ascending =False)
```

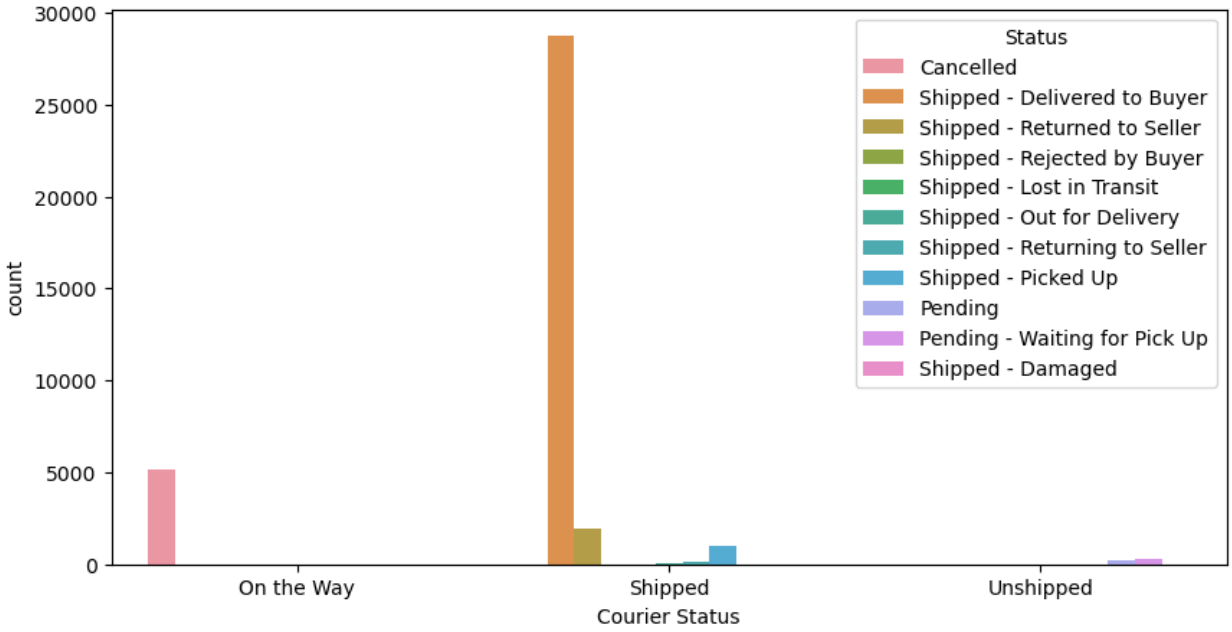
```
Size
M       5905
L       5795
XL      5481
XXL     4465
3XL     3972
S       3896
```

```
XS      2191
Free     467
6XL      170
5XL      104
4XL       93
Name: Qty, dtype: int64
```

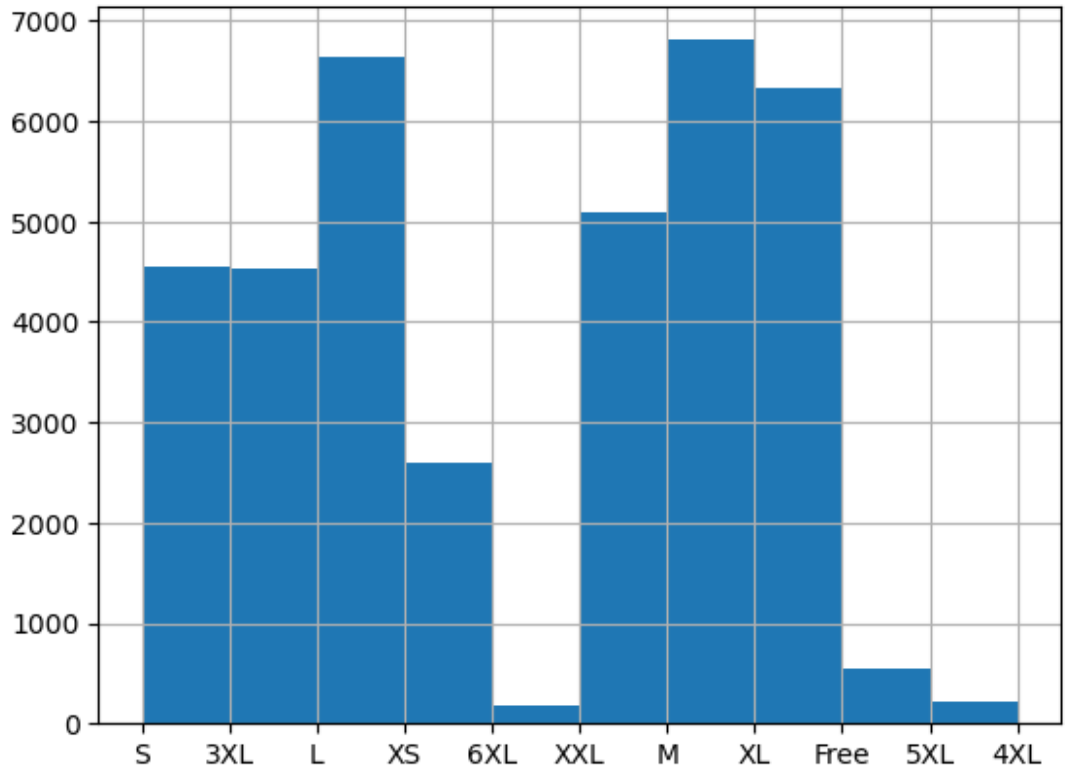
```
S_Qty=df.groupby(['Size'], as_index=False)
['Qty'].sum().sort_values(by='Qty',ascending=False)

ax=sns.barplot(x='Size',y='Qty', data=S_Qty)
for bars in ax.containers:
    ax.bar_label(bars) ## Counting and displaying the total number of
counts according to size
plt.show()
```



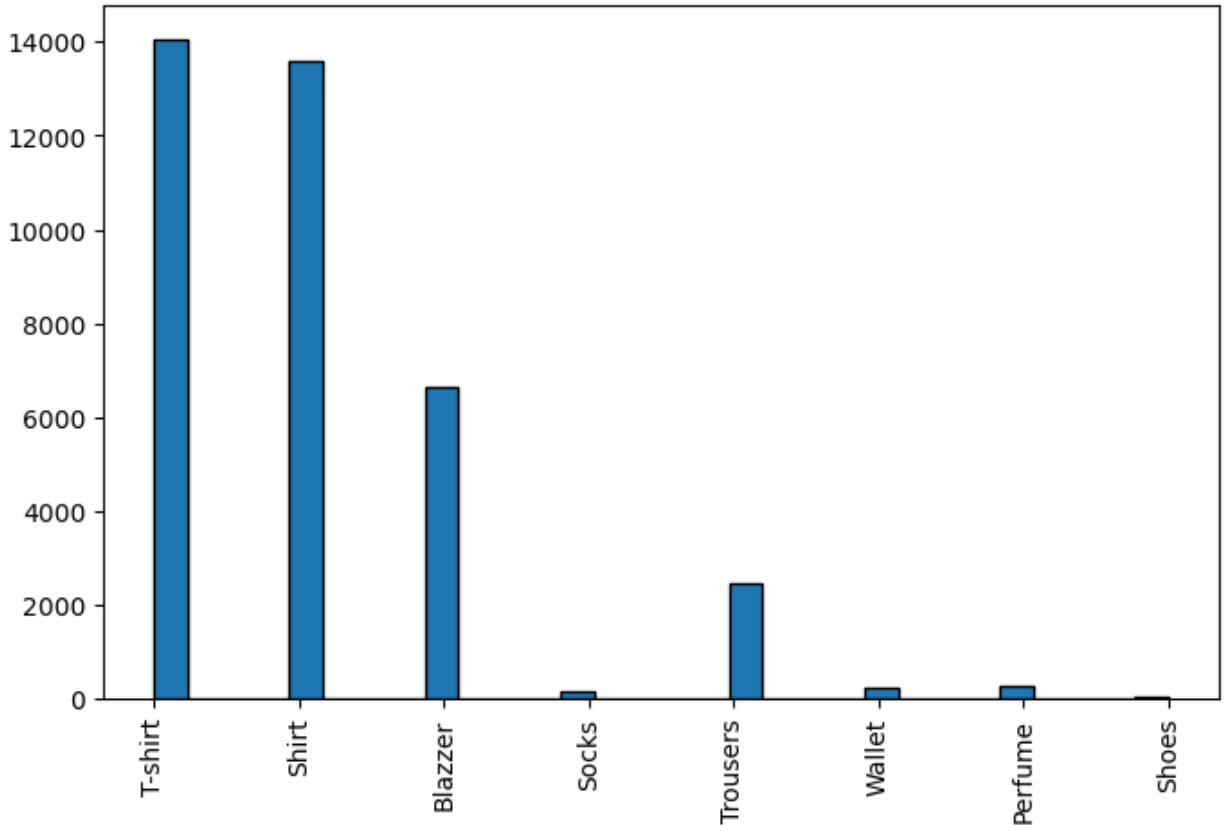

```
#Courier Status
plt.figure(figsize=(10,5))
ax=sns.countplot(data=df, x='Courier Status',hue= 'Status')
plt.show()
```



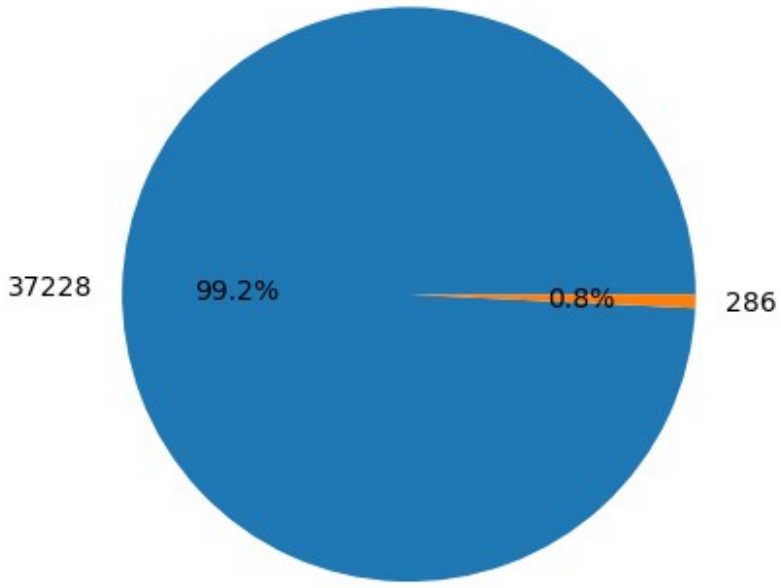

```
#histogram
df['Size'].hist()
plt.show()
```



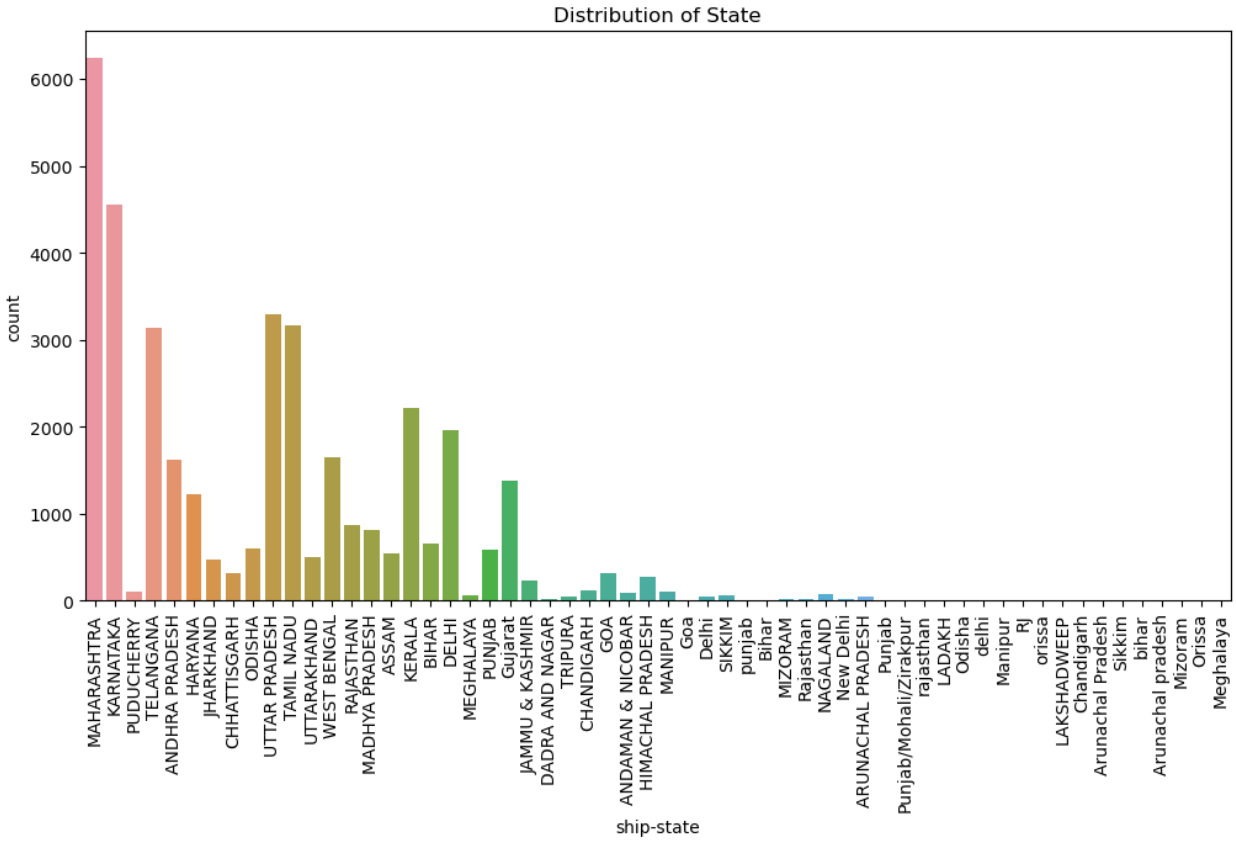
```
df['Category'] = df['Category'].astype(str)
column_data = df['Category']
plt.figure(figsize=(8, 5))
plt.hist(column_data, bins=30, edgecolor='Black')
plt.xticks(rotation=90)
plt.show()
```



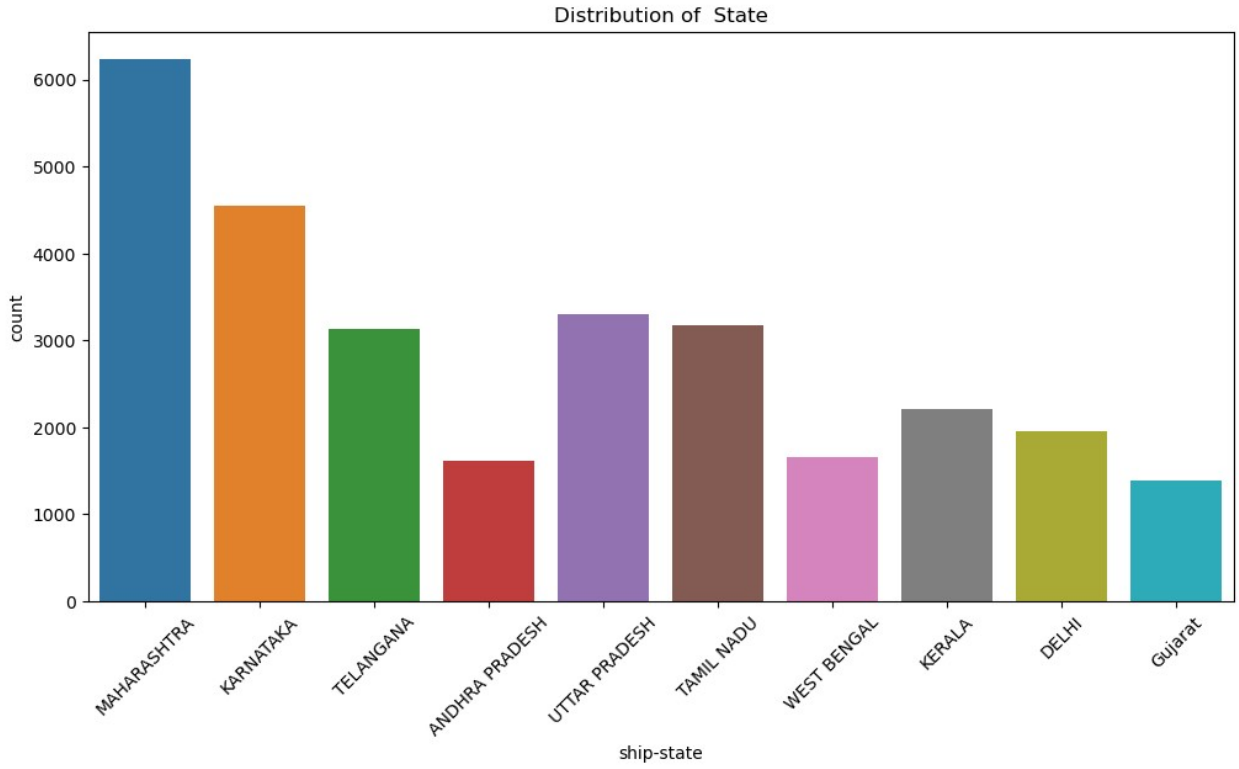

```
## Note: From above Graph you can see that most of the buyers are T-shirt

# Checking B2B Data  by using pie chart
B2B_Check = df['B2B'].value_counts()
#  Plot the pie chart
plt.pie(B2B_Check, labels=B2B_Check, autopct='%1.1f%%')
#plt.axis('equal')
plt.show()
```




```
##Note : From above chart we can see that maximum i.e. 99.3% of buyers
are retailers and 0.7% are B2B buyers

# 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()
```




```
##Note : From above bar graph we can see that maximum i.e.Maharastra
has highest disrtibution by state.

# top_10_States
top_10_state = 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_state.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 bar graph we can see top 10 states by distribution and maximum i.e.Maharastra has highest disrtibution by state.