AMAZON 2024 DATA ANALYSIS PROJECT

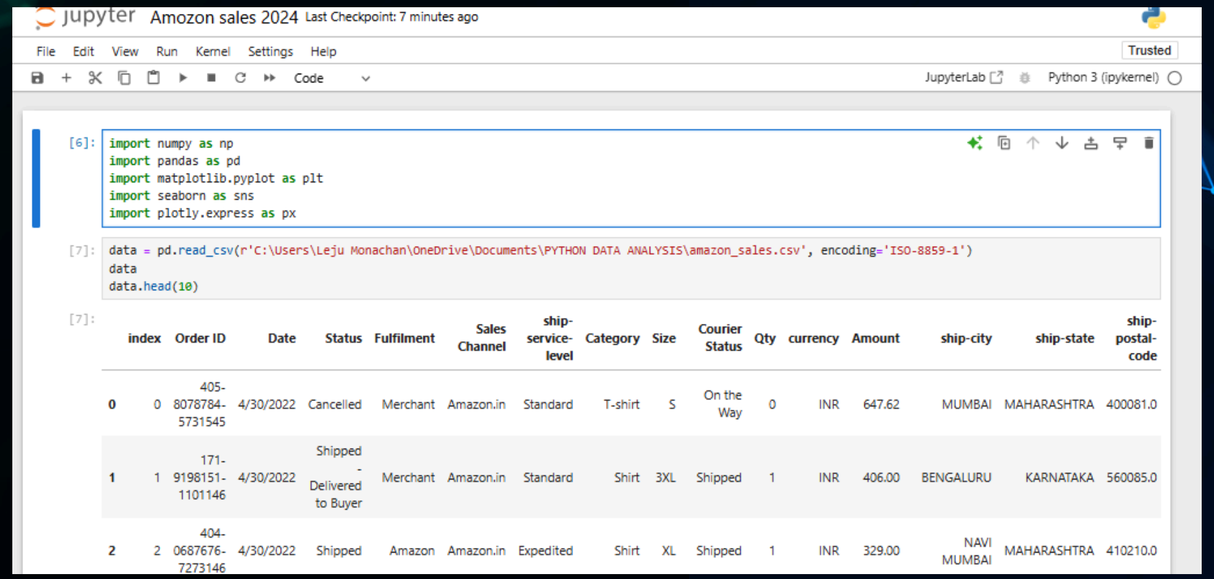
PYTHON PROJECT

https://github.com/lejumonachan/PYTHON-DATA-ANALYSIS-PROJECT

PYTHON PROJECT

In this project, the aim is to perform an in-depth analysis of Amazon's sales data for the year 2024, utilizing various Python libraries to gain valuable insights into sales performance, customer behavior, and geographic distribution

Importing cvs file into Jupyter Notebook using numpy and pandas



jupyter Amazon sales 2024 Last Checkpoint: 7 minutes ago

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

[6]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
```

[7]:
```
data = pd.read_csv(r'C:\Users\Leju Monachan\OneDrive\Documents\PYTHON DATA ANALYSIS\amazon_sales.csv', encoding='ISO-8859-1')
data
data.head(10)
```

[7]:

	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
0	0	405-8078784-5731545	4/30/2022	Cancelled	Merchant	Amazon.in	Standard	T-shirt	S	On the Way	0	INR	647.62	MUMBAI	MAHARASHTRA	400081.0
1	1	171-9198151-1101146	4/30/2022	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	Shirt	3XL	Shipped	1	INR	406.00	BENGALURU	KARNATAKA	560085.0
2	2	404-0687676-7273146	4/30/2022	Shipped	Amazon	Amazon.in	Expedited	Shirt	XL	Shipped	1	INR	329.00	NAVI MUMBAI	MAHARASHTRA	410210.0

Some Amazon data set deatils below

```
[8]: data.isnull
```

```
[8]: <bound method DataFrame.isnull of          index             Order ID       Date                       Status  \
0            0  405-8078784-5731545  4/30/2022                  Cancelled
1            1  171-9198151-1101146  4/30/2022  Shipped - Delivered to Buyer
2            2  404-0687676-7273146  4/30/2022                    Shipped
3            3  403-9615377-8133951  4/30/2022                  Cancelled
4            4  407-1069790-7240320  4/30/2022                    Shipped
...        ...                  ...        ...                        ...
128971  128970  406-6001380-7673107  5/31/2022                    Shipped
128972  128971  402-9551604-7544318  5/31/2022                    Shipped
128973  128972  407-9547469-3152358  5/31/2022                    Shipped
128974  128973  402-6184140-0545956  5/31/2022                    Shipped
128975  128974  408-7436540-8728312  5/31/2022                    Shipped

       Fulfilment Sales Channel ship-service-level  Category Size  \
0        Merchant     Amazon.in           Standard   T-shirt    S
1        Merchant     Amazon.in           Standard     Shirt  3XL
2          Amazon     Amazon.in          Expedited     Shirt   XL
3        Merchant     Amazon.in           Standard   Blazzer    L
4          Amazon     Amazon.in          Expedited  Trousers  3XL
...           ...           ...                ...       ...  ...
128971     Amazon     Amazon.in          Expedited     Shirt   XL
128972     Amazon     Amazon.in          Expedited   T-shirt    M
128973     Amazon     Amazon.in          Expedited   Blazzer  XXL
128974     Amazon     Amazon.in          Expedited   T-shirt   XS
128975     Amazon     Amazon.in          Expedited   T-shirt    S

       Courier Status  Qty currency   Amount    ship-city    ship-state  \
0          On the Way    0      INR   647.62       MUMBAI   MAHARASHTRA
1             Shipped    1      INR   406.00    BENGALURU     KARNATAKA
2             Shipped    1      INR   329.00  NAVI MUMBAI   MAHARASHTRA
3          On the Way    0      INR   753.33   PUDUCHERRY    PUDUCHERRY
4             Shipped    1      INR   574.00      CHENNAI    TAMIL NADU
...               ...  ...      ...      ...          ...           ...
128971        Shipped    1      INR   517.00    HYDERABAD     TELANGANA
128972        Shipped    1      INR   999.00     GURUGRAM       HARYANA
128973        Shipped    1      INR   690.00    HYDERABAD     TELANGANA
128974        Shipped    1      INR  1199.00        Halol       Gujarat
128975        Shipped    1      INR   696.00       Raipur  CHHATTISGARH

       ship-postal-code ship-country    B2B fulfilled-by
0              400081.0           IN  False    Easy Ship
1              560085.0           IN  False    Easy Ship
```

Some Amazon data set deatils below : Info()

```
[29]: data.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
```

Some Amazon data set deatils below : tail()

```
data.tail()
```

	index	Order ID	Date	Status	Fulfilment	Sales Channel	ship-service-level	Category	Size	Courier Status	Qty	currency	Amount	ship-city	ship-state	shi post co
128971	128970	406-6001380-7673107	5/31/2022	Shipped	Amazon	Amazon.in	Expedited	Shirt	XL	Shipped	1	INR	517.0	HYDERABAD	TELANGANA	50001
128972	128971	402-9551604-7544318	5/31/2022	Shipped	Amazon	Amazon.in	Expedited	T-shirt	M	Shipped	1	INR	999.0	GURUGRAM	HARYANA	12200
128973	128972	407-9547469-3152358	5/31/2022	Shipped	Amazon	Amazon.in	Expedited	Blazzer	XXL	Shipped	1	INR	690.0	HYDERABAD	TELANGANA	50004
128974	128973	402-6184140-0545956	5/31/2022	Shipped	Amazon	Amazon.in	Expedited	T-shirt	XS	Shipped	1	INR	1199.0	Halol	Gu	
128975	128974	408-7436540-8728312	5/31/2022	Shipped	Amazon	Amazon.in	Expedited	T-shirt	S	Shipped	1	INR	696.0	Raipur	CHHATTISG	

Some Amazon data set deatils below : describe ()

```
memory usage: 5.3+ MB
```

[15]:
```
data.describe(include = "object")
```

[15]:

	Order ID	Date	Status	Fulfilment	Sales Channel	ship-service-level	Category	Size	Courier Status	currency	ship-city	ship-state	ship-country	fulfilled-by
count	128976	128976	128976	128976	128976	128976	128976	128976	128976	121176	128941	128941	128941	39263
unique	120229	91	13	2	2	2	9	11	4	1	8948	69	1	1
top	403-4984515-8861958	5/3/2022	Shipped	Amazon	Amazon.in	Expedited	T-shirt	M	Shipped	INR	BENGALURU	MAHARASHTRA	IN	Easy Ship
freq	12	2085	77815	89713	128852	88630	50292	22373	109486	121176	11208	22272	128941	39263

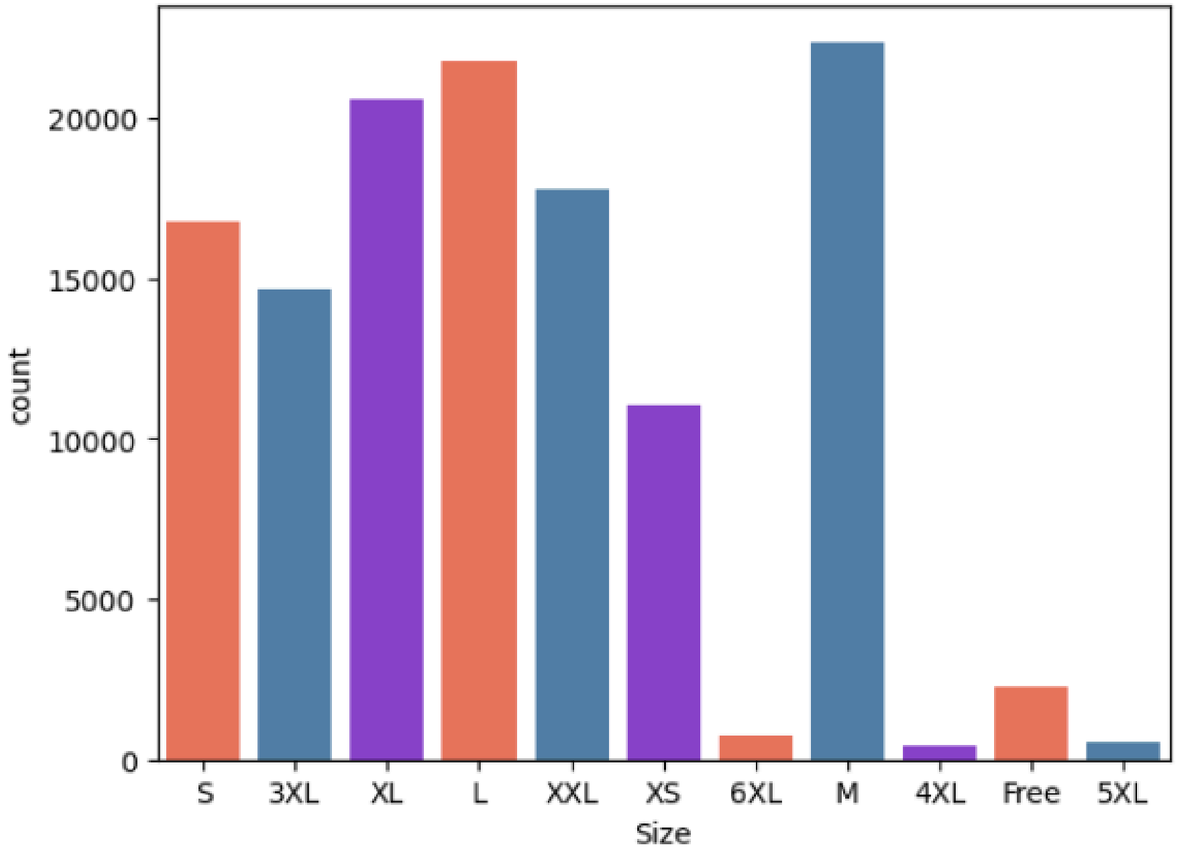
[101]:
```
data[['Qty','Amount']].describe()
```

[101]:

	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

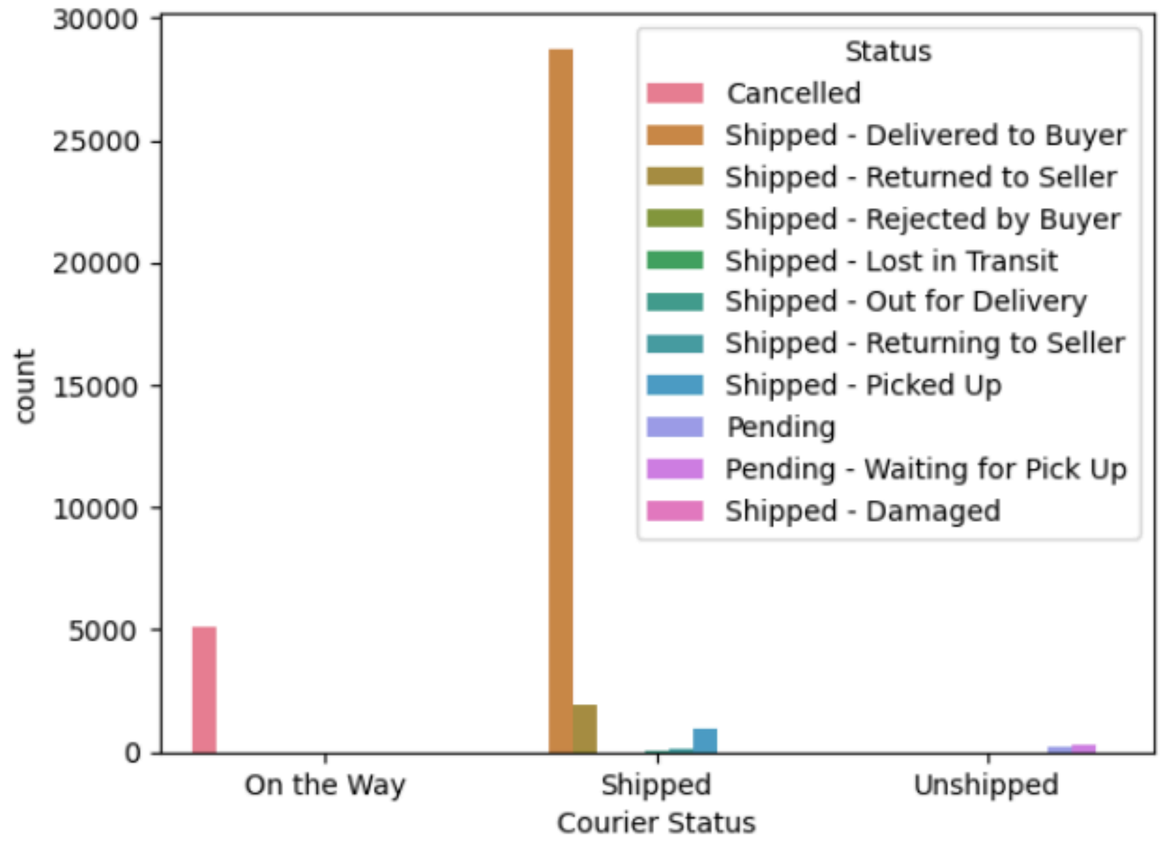
Medium size has the highest distribution

```
[19]: ax = sns.countplot(x='Size',data = data ,palette=['#FF6347', '#4682B4', '#8A2BE2'] )
      for bars in ax.coutainers:
          ax.bar_label(bars)
```



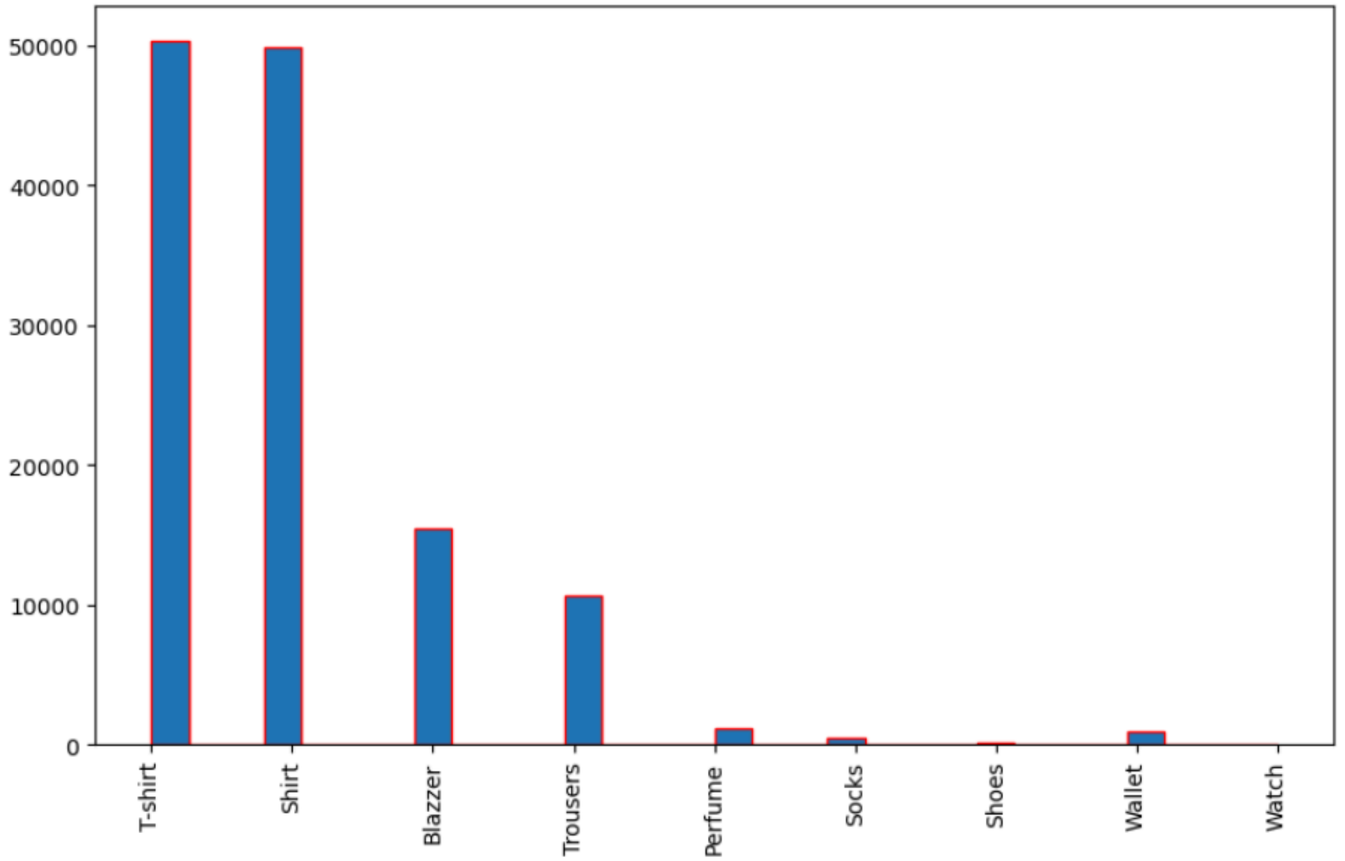
Courier Status Tracking list

```
[139]: sns.countplot(x = 'Courier Status',hue = 'Status',data = data)
       plt.show()
```



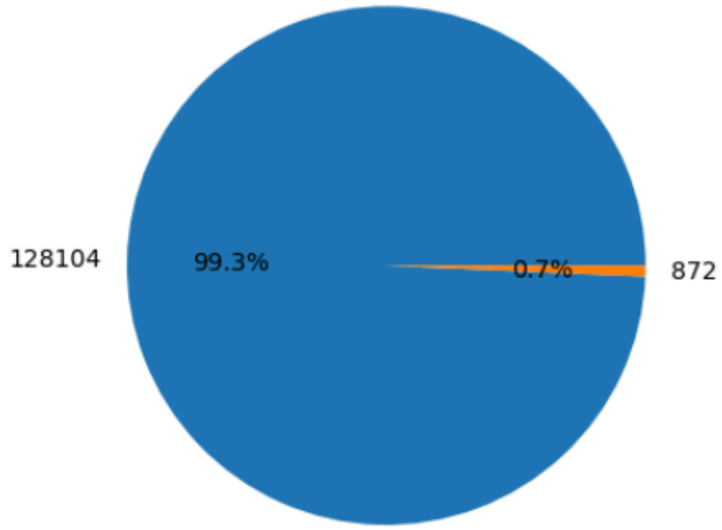
T -shirt and shirt has the highest count ()

[27]:
```
data['Category'] = data['Category'].astype('string')
cd = data['Category']
plt.figure(figsize = (10,6))
plt.hist(cd,bins=30,edgecolor="Red")
plt.xticks(rotation=90)
plt.show()
```



99.2% of buyers are Retailers and 0.8 has only Wholesaler

```
[43]: check_B2B = data['B2B'].value_counts()
      plt.pie(check_B2B , labels = check_B2B,autopct ='%1.1f%%')
      plt.show()
```



maximum 99.2% of buyers are retailes and 0.8% are wholesaler

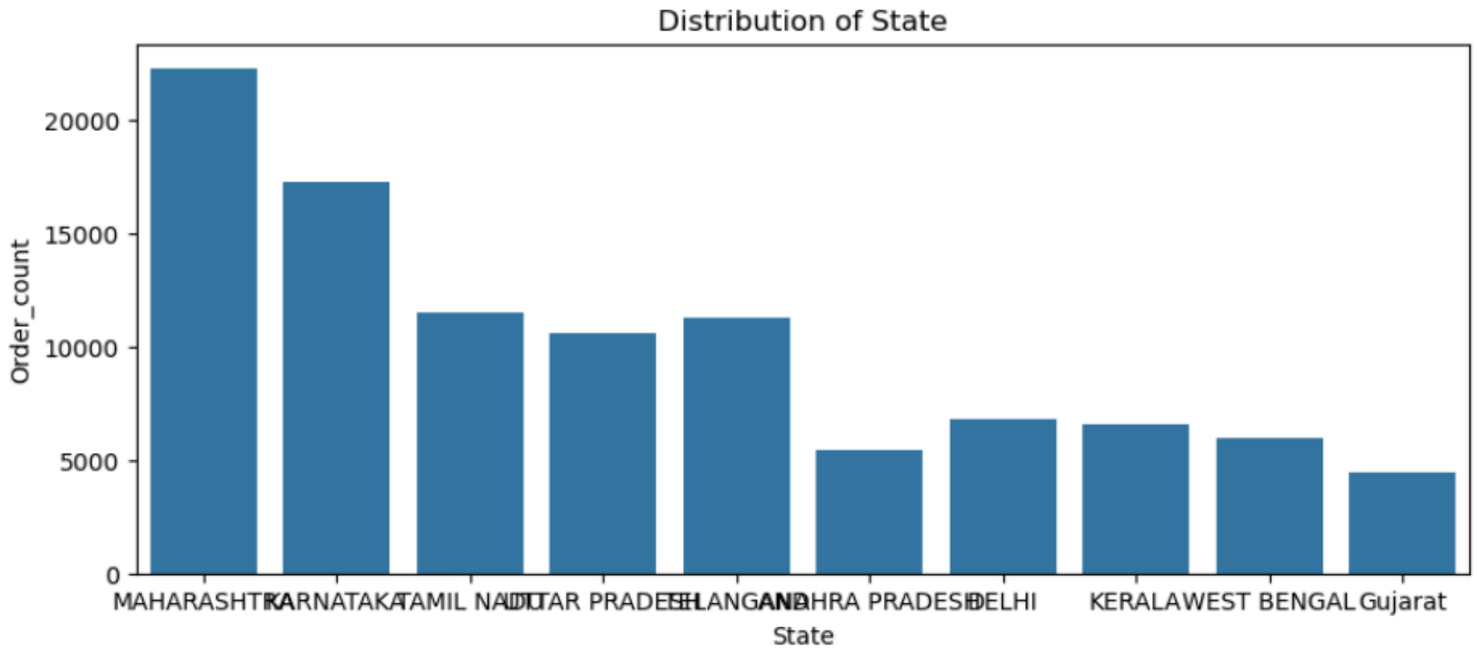
scatter plot by category bases

[67]:
```
data_x = data["Category"]
data_y = data["Size"]

plt.scatter(data_x,data_y)
plt.xlabel("Category")
plt.ylabel("Size")
plt.title("Available size")
plt.show()
```

Highest amount of Distribution by Statewise

```
[81]: top10_state =data["ship-state"].value_counts().head(10)
      plt.figure (figsize =(10,4))
      sns.countplot(data=data[data["ship-state"].isin(top10_state.index)],x="ship-state")
      plt.xlabel("State")
      plt.ylabel("Order_count")
      plt.title("Distribution of State")
      plt.show()
```



THANK YOU