

AMAZON SALES ANALYSIS

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

```
In [2]: df=pd.read_csv(r"C:\Users\hp\Downloads\amazon_sales.csv - Sheet1.csv")
```

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

Out[3]:

	index	Order ID	Date	Status	Fulfilment	Sales Channel	ship-service-level	Category	Size	Courier Status	Qty	currency	Amount	
0	0	405-8078784-5731545	4/30/2022	Cancelled	Merchant	Amazon.in	Standard	T-shirt	S	On the Way	0	INR	647.62	
1	1	171-9198151-1101146	4/30/2022	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	Shirt	3XL	Shipped	1	INR	406.00	BEN
2	2	404-0687676-7273146	4/30/2022	Shipped	Amazon	Amazon.in	Expedited	Shirt	XL	Shipped	1	INR	329.00	NAVI
3	3	403-9615377-8133951	4/30/2022	Cancelled	Merchant	Amazon.in	Standard	Blazzer	L	On the Way	0	INR	753.33	PUDU
4	4	407-1069790-7240320	4/30/2022	Shipped	Amazon	Amazon.in	Expedited	Trousers	3XL	Shipped	1	INR	574.00	C

```
In [4]: 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 [5]: df.describe()
```

Out[5]:

	index	Qty	Amount	ship-postal-code
count	128976.000000	128976.000000	121176.000000	128941.000000
mean	64486.130427	0.904401	648.562176	463945.677744
std	37232.897832	0.313368	281.185041	191458.488954
min	0.000000	0.000000	0.000000	110001.000000
25%	32242.750000	1.000000	449.000000	382421.000000
50%	64486.500000	1.000000	605.000000	500033.000000
75%	96730.250000	1.000000	788.000000	600024.000000
max	128974.000000	15.000000	5584.000000	989898.000000

In [7]: df.shape

Out[7]: (128976, 19)

In [8]: # Check null values

In [9]: df.isnull()

Out[9]:

	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	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
...
128971	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
128972	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
128973	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
128974	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
128975	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False

128976 rows × 19 columns

In [10]: df.isnull().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.dropna(inplace=True)

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

```
In [19]: df['ship-postal-code'].dtype
```

```
Out[19]: dtype('int32')
```

```
In [20]: df['Date']=pd.to_datetime(df['Date'])
```

```
In [21]: df['Date'].dtype
```

```
Out[21]: dtype('<M8[ns]')
```

```
In [23]: #Describe the data with object data type
```

```
In [22]: df.describe(include='object')
```

Out[22]:

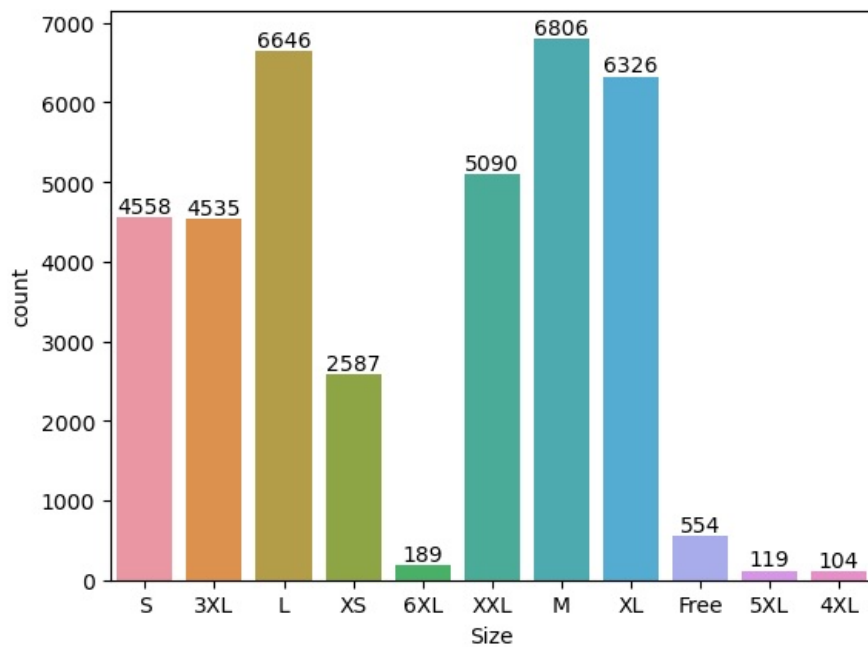
	Order ID	Status	Fulfilment	Sales Channel	ship-service-level	Category	Size	Courier Status	currency	ship-city	ship-state	count
count	37514	37514	37514	37514	37514	37514	37514	37514	37514	37514	37514	37514
unique	34664	11	1	1	1	8	11	3	1	4698	58	
top	171-5057375-2831560	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	T-shirt	M	Shipped	INR	BENGALURU	MAHARASHTRA	
freq	12	28741	37514	37514	37514	14062	6806	31859	37514	2839	6236	

```
In [27]: df[['Qty','Amount']].describe()
```

Out[27]:

	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

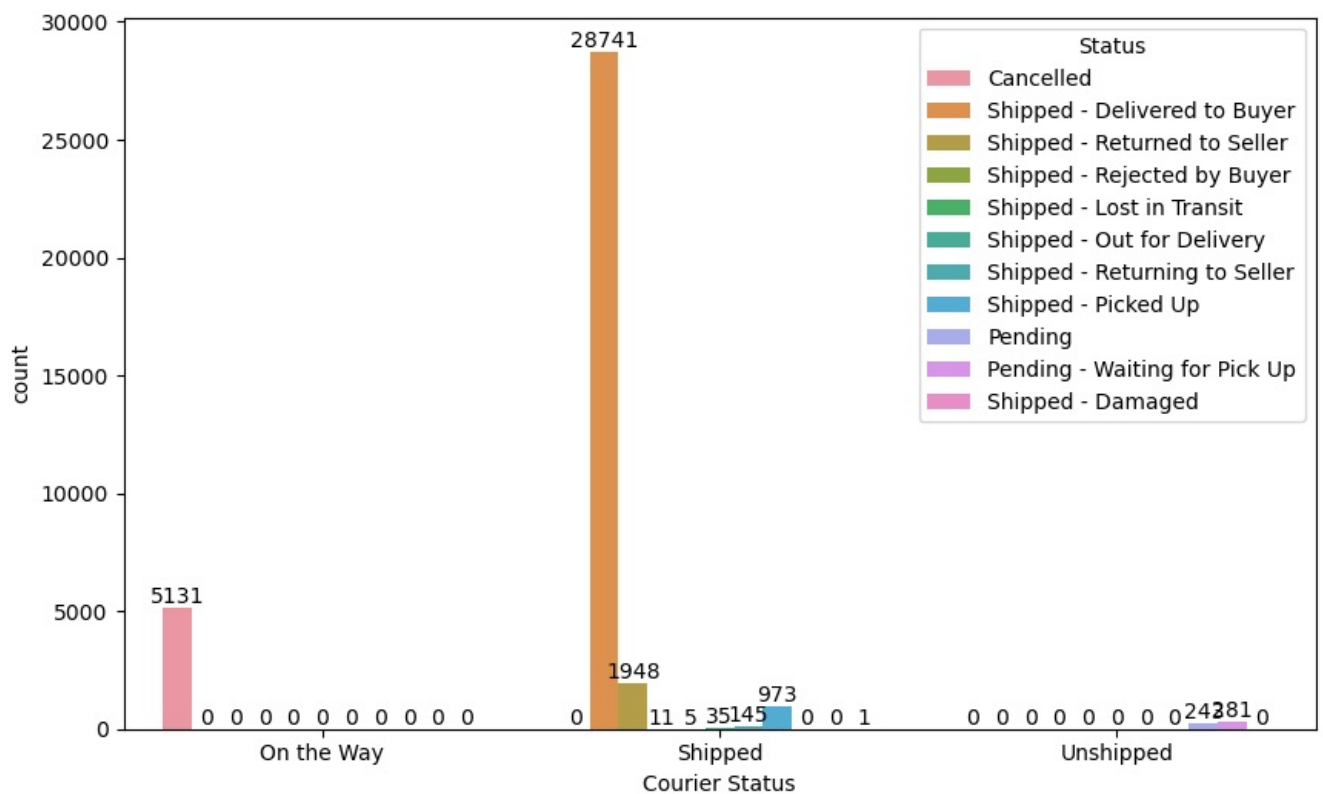
```
In [32]: ax=sns.countplot(x='Size', data=df)
         for bars in ax.containers:
             ax.bar_label(bars)
```



Most of the people buys M size

In [38]: #Courier status

```
In [48]: plt.figure(figsize=(10,6))
ax=sns.countplot(x='Courier Status', data=df, hue='Status')
for bars in ax.containers:
    ax.bar_label(bars)
plt.show()
```



Majority of the orders are the shipped by courier

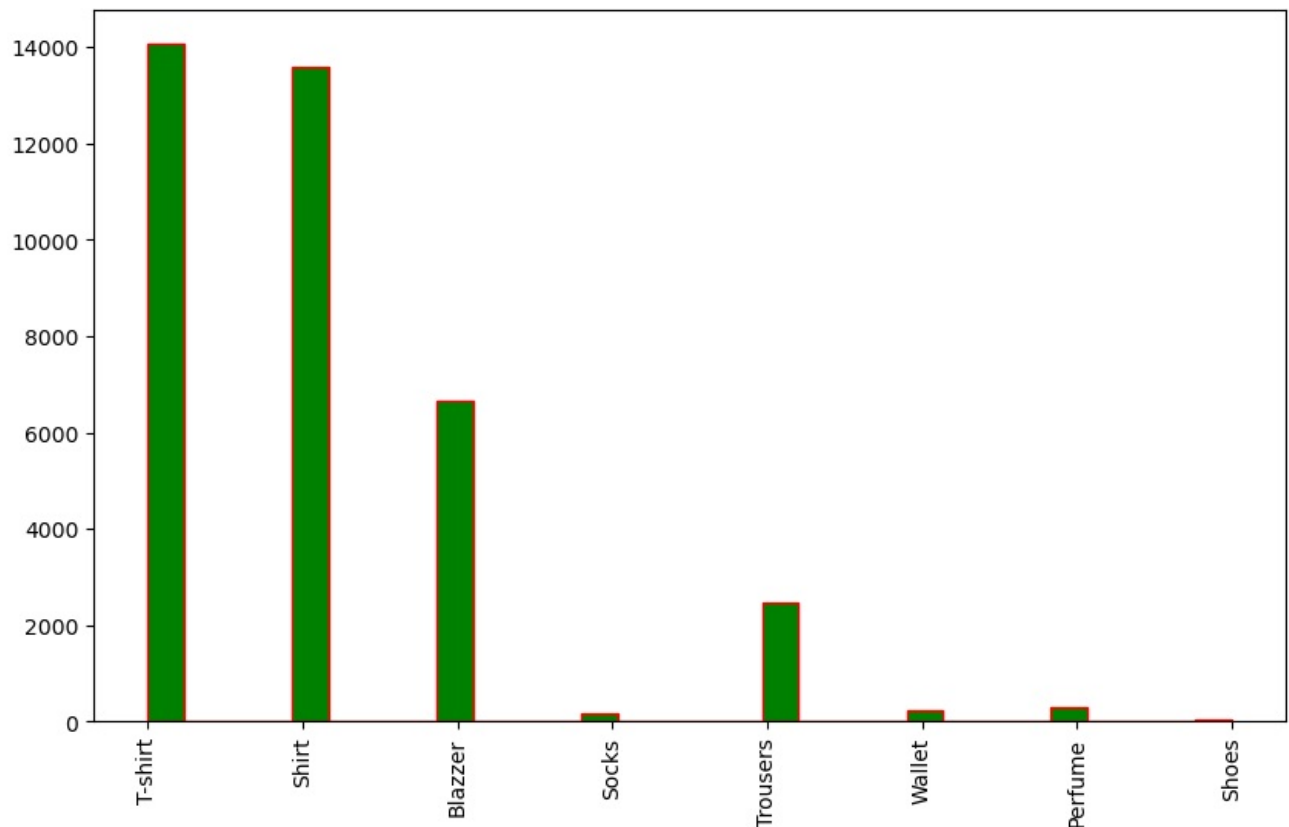
In [50]: df.head(2)

Out[50]:

	index	Order ID	Date	Status	Fulfilment	Sales Channel	ship-service-level	Category	Size	Courier Status	Qty	currency	Amount	ship-ci
0	0	405-8078784-5731545	2022-04-30	Cancelled	Merchant	Amazon.in	Standard	T-shirt	S	On the Way	0	INR	647.62	MUMB
1	1	171-9198151-1101146	2022-04-30	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	Shirt	3XL	Shipped	1	INR	406.00	BENGALUF

In [55]:

```
df['Category']=df['Category'].astype('str')
c_d=df['Category']
plt.figure(figsize=(10,6))
plt.hist(c_d,bins=30,edgecolor='Red',color='Green')
plt.xticks(rotation=90)
plt.show()
```



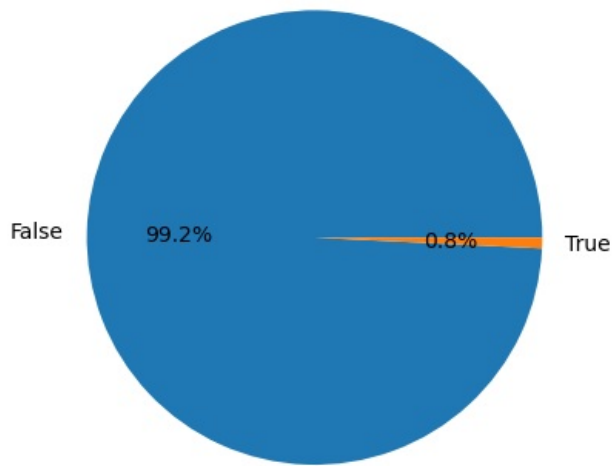
Most of the buying products are T-shirt

In [57]:

```
#Check B2B data
```

In [64]:

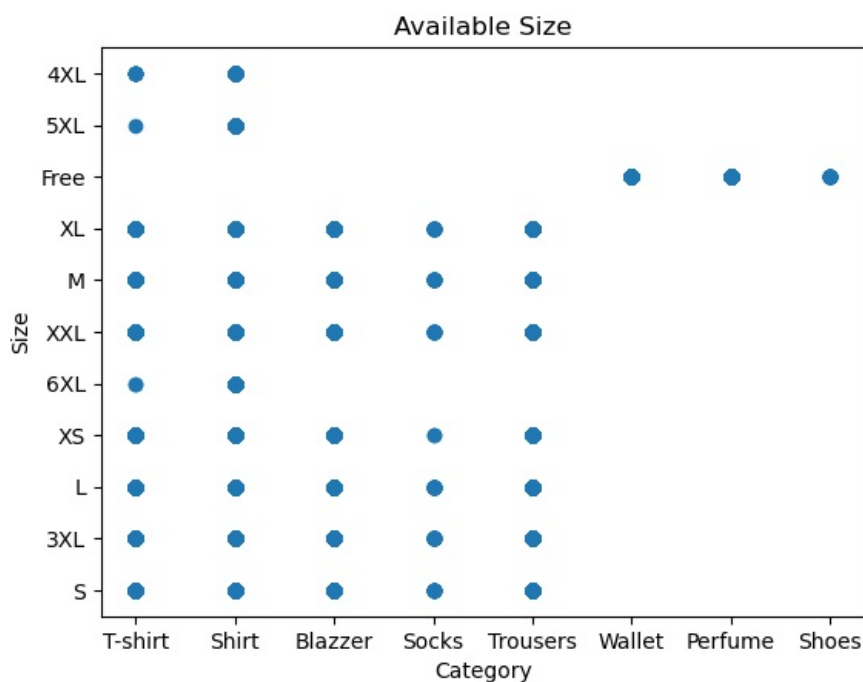
```
check_B2B=df['B2B'].value_counts()
#Create pie chart
plt.pie(check_B2B,labels=check_B2B.index, autopct='%1.1f%%')
plt.show()
```



Maximum 99.2% buyers are retailers and 0.85% are Wholesalers

In [66]: #Scatter plots

```
In [71]: X=df['Category']
Y=df['Size']
plt.scatter(X,Y)
plt.xlabel('Category')
plt.ylabel('Size')
plt.title('Available Size')
plt.show()
```



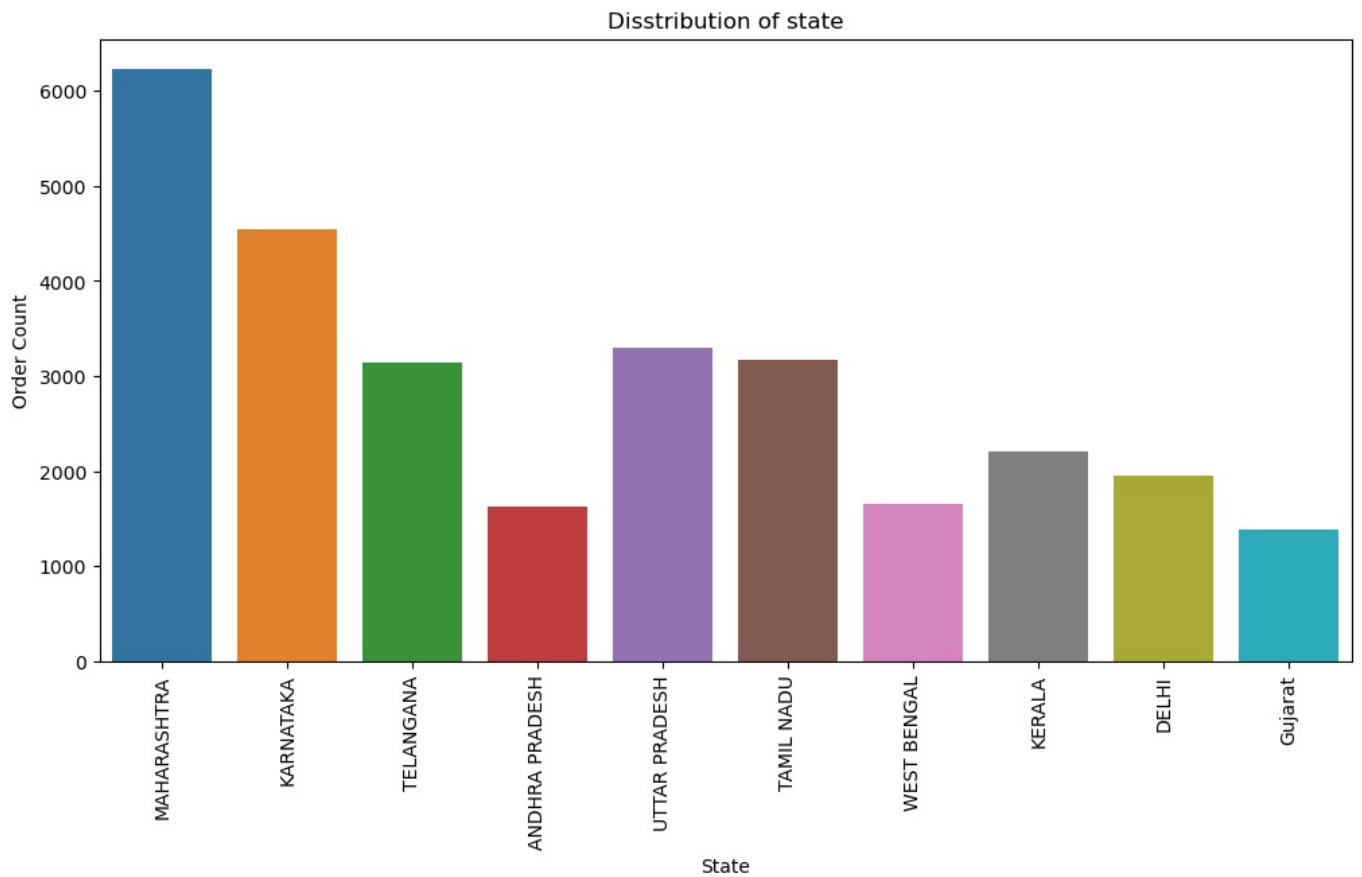
In [72]: df.head(2)

Out[72]:

	index	Order ID	Date	Status	Fulfilment	Sales Channel	ship-service-level	Category	Size	Courier Status	Qty	currency	Amount	ship-ci
0	0	405-8078784-5731545	2022-04-30	Cancelled	Merchant	Amazon.in	Standard	T-shirt	S	On the Way	0	INR	647.62	MUMB
1	1	171-9198151-1101146	2022-04-30	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	Shirt	3XL	Shipped	1	INR	406.00	BENGALUR

```
In [79]: top10_states=df['ship-state'].value_counts().head(10)
plt.figure(figsize=(12,6))
sns.countplot(data=df[df['ship-state'].isin(top10_states.index)], x='ship-state')
```

```
plt.xlabel('State')
plt.ylabel('Order Count')
plt.title('Disstribution of state')
plt.xticks(rotation=90)
plt.show()
```



Most of the buyers are from Maharastra

In []:

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js