

# Amazon Sales Report

## Importing libraries

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

## Importing dataset

```
df = pd.read_csv("/content/drive/MyDrive/Amazon Sale Report1.csv")
df.head()
```

## Output:

	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
0	0	405-8078784-5731545	04-30-22	Cancelled	Merchant	Amazon.in	Standard	T-shirt	S	On the Way	0	INR	647.62	MUMBAI	MAHARASHTRA	400081.0	IN	False	Easy Ship
1	1	171-9198151-1101146	04-30-22	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	Shirt	3XL	Shipped	1	INR	406.00	BENGALURU	KARNATAKA	560085.0	IN	False	Easy Ship
2	2	404-0687676-7273146	04-30-22	Shipped	Amazon	Amazon.in	Expedited	Shirt	XL	Shipped	1	INR	329.00	NAVI MUMBAI	MAHARASHTRA	410210.0	IN	True	NaN
3	3	403-9615377-8133951	04-30-22	Cancelled	Merchant	Amazon.in	Standard	Blazzer	L	On the Way	0	INR	753.33	PUDUCHERRY	PUDUCHERRY	605008.0	IN	False	Easy Ship
4	4	407-1069790-7240320	04-30-22	Shipped	Amazon	Amazon.in	Expedited	Trousers	3XL	Shipped	1	INR	574.00	CHENNAI	TAMIL NADU	600073.0	IN	False	NaN

```
df.info()
```

## Output:

```
<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
```

```
df.tail()
```

Output:

	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
128971	128970	406-6001380-7673107	05-31-22	Shipped	Amazon	Amazon.in	Expedited	Shirt	XL	Shipped	1	INR	517.0	HYDERABAD	TELANGANA	500013.0	IN	False	NaN
128972	128971	402-9551604-7544318	05-31-22	Shipped	Amazon	Amazon.in	Expedited	T-shirt	M	Shipped	1	INR	999.0	GURUGRAM	HARYANA	122004.0	IN	False	NaN
128973	128972	407-9547469-3152358	05-31-22	Shipped	Amazon	Amazon.in	Expedited	Biazzar	XXL	Shipped	1	INR	690.0	HYDERABAD	TELANGANA	500049.0	IN	False	NaN
128974	128973	402-6184140-0545956	05-31-22	Shipped	Amazon	Amazon.in	Expedited	T-shirt	XS	Shipped	1	INR	1199.0	Halol	Gujarat	389350.0	IN	False	NaN
128975	128974	408-7436540-8728312	05-31-22	Shipped	Amazon	Amazon.in	Expedited	T-shirt	S	Shipped	1	INR	696.0	Raipur	CHHATTISGARH	492014.0	IN	False	NaN

```
pd.isnull(df).sum()
```

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

Output: dtype: int64

```
df.shape
```

Output:

```
(128976, 19)
```

```
df.dropna(inplace=True)
```

```
df.shape
```

Output:

```
(37514, 19)
```

```
df.info()
```

Output:

```
<class 'pandas.core.frame.DataFrame'>
Index: 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  object  
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  float64 
16  ship-country            37514 non-null  object  
17  B2B                     37514 non-null  bool    
18  fulfilled-by            37514 non-null  object  
dtypes: bool(1), float64(2), int64(2), object(14)
memory usage: 5.5+ MB
```

```
df.head()
```

Output:

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
0	0	405-8078784-5731545 04-30-22	Cancelled	Merchant	Amazon.in	Standard	T-shirt	S	On the Way	0	INR	647.62	MUMBAI	MAHARASHTRA	400081.0	IN	False	Easy Ship
1	1	171-9198151-1101146 04-30-22	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	Shirt	3XL	Shipped	1	INR	406.00	BENGALURU	KARNATAKA	560085.0	IN	False	Easy Ship
3	3	403-9615377-8133951 04-30-22	Cancelled	Merchant	Amazon.in	Standard	Blazzer	L	On the Way	0	INR	753.33	PUDUCHERRY	PUDUCHERRY	605008.0	IN	False	Easy Ship
7	7	406-7807733-3785945 04-30-22	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	Shirt	S	Shipped	1	INR	399.00	HYDERABAD	TELANGANA	500032.0	IN	False	Easy Ship
12	12	405-5513694-8146768 04-30-22	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	Shirt	XS	Shipped	1	INR	399.00	Amravati	MAHARASHTRA	444606.0	IN	False	Easy Ship

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

```
df['ship-postal-code'].dtype
```

Output:

dtype('int64')

```
df['Date'] = pd.to_datetime(df['Date'])
df['Date'].dtype
```

Output:

dtype('<M8[ns]')

```
df.describe()
```

Output:

	index	Date	Qty	Amount	ship-postal-code
count	37514.000000	37514	37514.000000	37514.000000	37514.000000
mean	60953.809858	2022-05-11 07:56:47.303939840	0.867383	646.553960	463291.552754
min	0.000000	2022-03-31 00:00:00	0.000000	0.000000	110001.000000
25%	27235.250000	2022-04-20 00:00:00	1.000000	458.000000	370465.000000
50%	63470.500000	2022-05-09 00:00:00	1.000000	629.000000	500019.000000
75%	91790.750000	2022-06-01 00:00:00	1.000000	771.000000	600042.000000
max	128891.000000	2022-06-29 00:00:00	5.000000	5495.000000	989898.000000
std	36844.853039	NaN	0.354160	279.952414	194550.425637

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

Output:

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

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

Output:

	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

```
df.head()
```

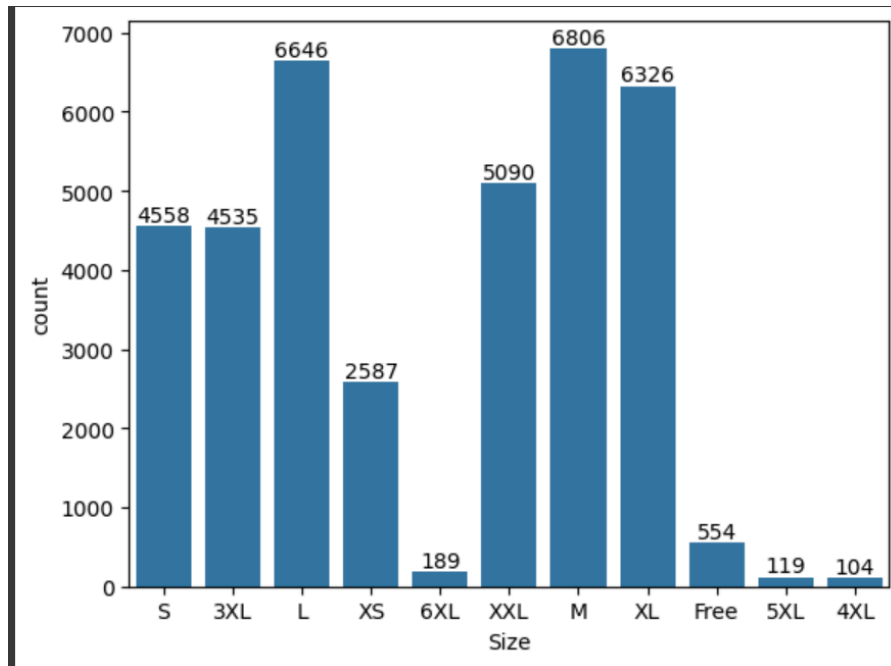
Output:

index	Order ID	Date	Status	Fulfilment	Sales Channel	ship-service-level	Category	Size	Courier	Qty	currency	Amount	ship-city	ship-state	ship-postal-code	ship-country	B2B	fulfilled-by
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	Easy Ship
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	Easy Ship
3	403-9815377-8133951	2022-04-30	Cancelled	Merchant	Amazon.in	Standard	Blazzer	L	On the Way	0	INR	753.33	PUDUCHERRY	PUDUCHERRY	605008	IN	False	Easy Ship
7	406-7807733-3785945	2022-04-30	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	Shirt	S	Shipped	1	INR	399.00	HYDERABAD	TELANGANA	500032	IN	False	Easy Ship
12	405-5513894-8146768	2022-04-30	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	Shirt	XS	Shipped	1	INR	399.00	Amravati.	MAHARASHTRA	444606	IN	False	Easy Ship

```
ax=sns.countplot(x='Size',data=df)
for bars in ax.containers:
```

```
ax.bar_label(bars)
```

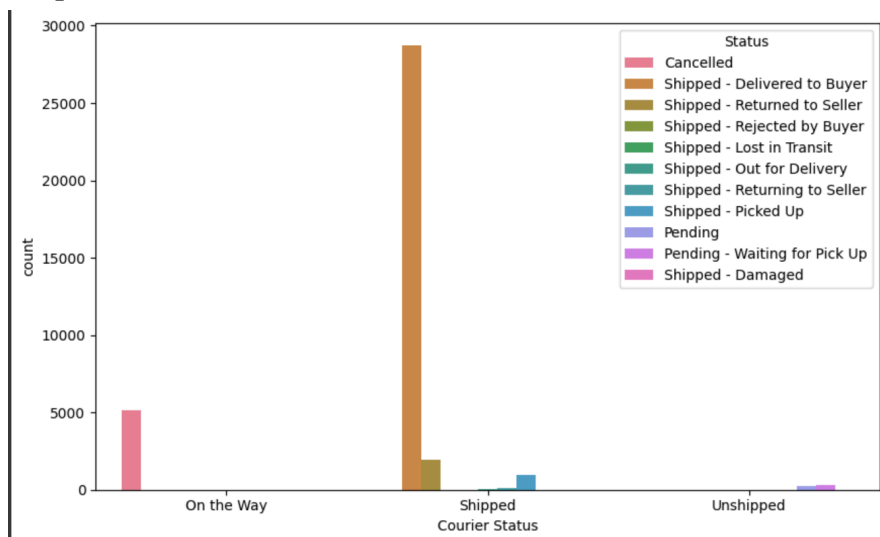
Output:



most of the people buy M-Size

```
plt.figure(figsize=(10,6))
sns.countplot(data=df,x='Courier Status',hue='Status')
plt.show()
```

Output:



Majority to the orders are shipped through the courier

```
df.head()
```

Output:

	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
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	Easy Ship
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	Easy Ship
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	Easy Ship
7	7	406-7807733-3785945	2022-04-30	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	Shirt	S	Shipped	1	INR	399.00	HYDERABAD	TELANGANA	500032	IN	False	Easy Ship
12	12	405-5513694-8146768	2022-04-30	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	Shirt	XS	Shipped	1	INR	399.00	Amravati	MAHARASHTRA	444606	IN	False	Easy Ship

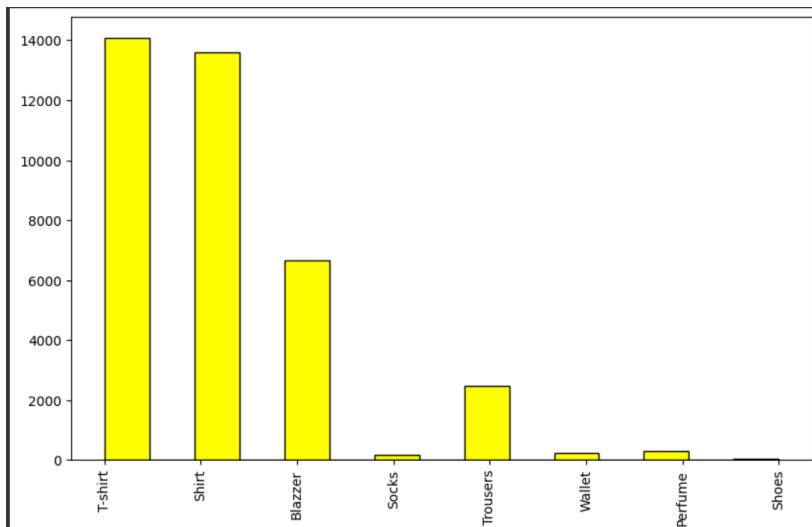
```
df.info()
```

Output:

```
<class 'pandas.core.frame.DataFrame'>
Index: 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  int64  
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), int64(3), object(13)
memory usage: 6.5+ MB
```

```
df['Category']=df['Category'].astype(str)
c_d=df['Category']
plt.figure(figsize=(10,6))
plt.hist(c_d,bins=15,edgecolor="Black",color="Yellow")
plt.xticks(rotation=90)
plt.show()
```

Output:



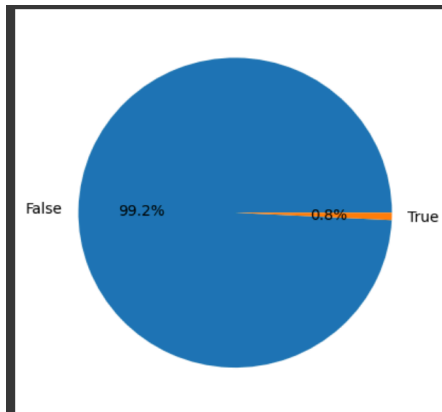
Most of the buyers are t-shirt

```
check_B2B=df['B2B'].value_counts()

plt.pie(check_B2B,labels=check_B2B.index, autopct='%1.1f%%')
plt.show()
```

Output:





Maximum 99.2% of buyers are retailers and 0.8% are wholesalers

```
x_data=df['Category']
y_data=df['Size']
plt.scatter(x_data,y_data)
plt.xlabel('Category')
plt.ylabel('Size')
plt.title('Available size')
plt.show()
```

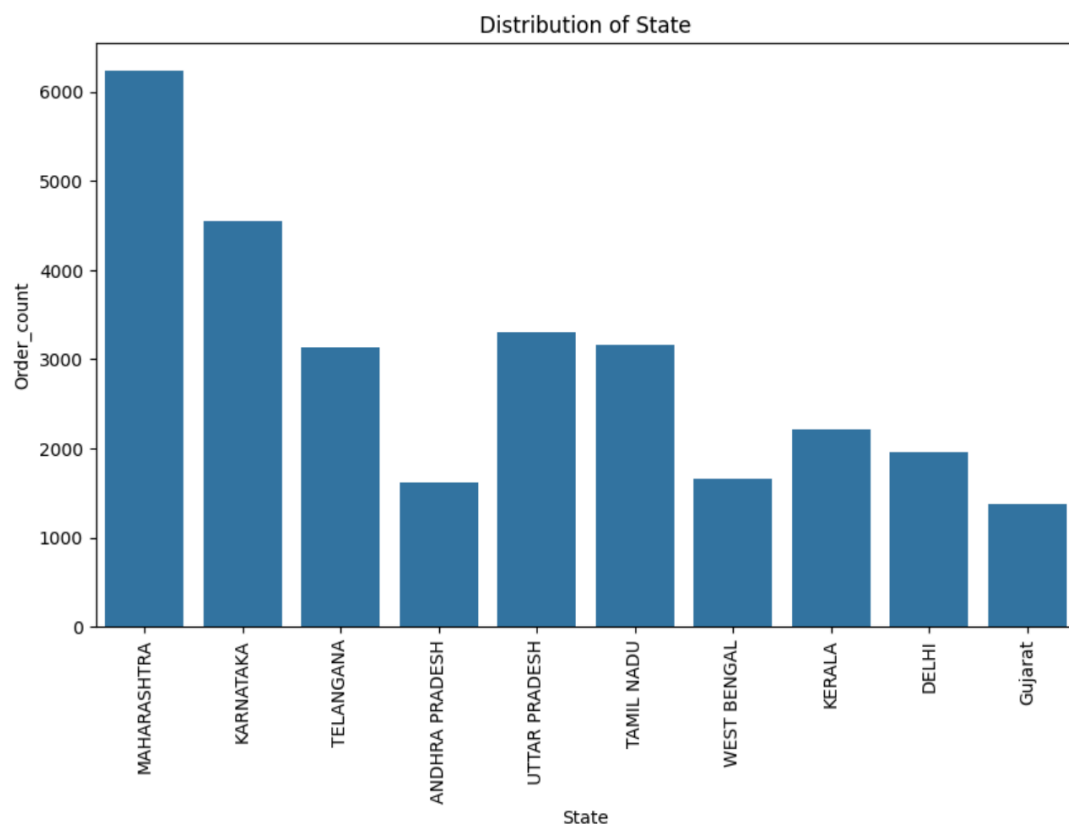
Output:



```
top10_state=df['ship-state'].value_counts().head(10) #for top 10 states
plt.figure(figsize=(10,6))
```

```
sns.countplot(data=df[df['ship-state'].isin(top10_state.index)],x="ship-state")
plt.xlabel('State')
plt.ylabel('Order_count')
plt.title('Distribution of State')
plt.xticks(rotation=90)
plt.show()
```

Output:



Most of the buyers are from maharashtra state

