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

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

In [3]:
```
df.shape
```

Out[3]:
```
(128976, 21)
```

In [4]:
```
df.head()
```

Out[4]:

	index	Order ID	Date	Status	Fulfilment	Sales Channel	shi servic lev
0	0	405-8078784-5731545	04-30-22	Cancelled	Merchant	Amazon.in	Standa
1	1	171-9198151-1101146	04-30-22	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standa
2	2	404-0687676-7273146	04-30-22	Shipped	Amazon	Amazon.in	Expedite
3	3	403-9615377-8133951	04-30-22	Cancelled	Merchant	Amazon.in	Standa
4	4	407-1069790-7240320	04-30-22	Shipped	Amazon	Amazon.in	Expedite

5 rows × 21 columns

In [5]:
```
df.tail()
```

Out[5]:

	index	Order ID	Date	Status	Fulfilment	Sales Channel	
128971	128970	406-6001380-7673107	05-31-22	Shipped	Amazon	Amazon.in	E
128972	128971	402-9551604-7544318	05-31-22	Shipped	Amazon	Amazon.in	E
128973	128972	407-9547469-3152358	05-31-22	Shipped	Amazon	Amazon.in	E
128974	128973	402-6184140-0545956	05-31-22	Shipped	Amazon	Amazon.in	E
128975	128974	408-7436540-8728312	05-31-22	Shipped	Amazon	Amazon.in	E

5 rows × 21 columns

In [6]:
```
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 [8]:
```
df.drop(["New","PendingS"], axis=1, inplace=True)
```

In [9]:
```
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 [ ]:
```
pd.isnull(df)
```

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]:
```
pd.isnull(df).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 [14]:
```
df.columns
```

Out[14]:
```
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')
```

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

In [16]:
```
df['ship-postal-code'].dtype
```

Out[16]:
```
dtype('int32')
```

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

In [ ]:
```
df.rename(columns={'Qty':'Quantity'})
```

EXPLORATORY DATA ANALYSIS

In [ ]:
```
X=sns.countplot(x='Size', data=df)
```

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

from the above grph we seen that most of the people buy M-size followed by L

In [ ]:
```
gp=df.groupby(['Size'], as_index=False)['Qty'].sum().sort_values(by='Qty', asc
sns.barplot(x='Size',y='Qty',data=gp)
plt.show()
```

by above graph we seen that most sold was M-size

courier status

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

- most of the orders are in shipped status

In [ ]:
```
#to increse fig size
plt.figure(figsize=(10,5))
sns.countplot(data=df, x='Courier Status' , hue='Status')
plt.show()
```

In [ ]:
```
plt.hist(data= df, bins= 25, edgecolor='black', x='Category')
plt.xticks(rotation=360)
plt.show()
```

Most of the buyers buys T-shirts

In [ ]:
```
pie=df['B2B'].value_counts()
plt.pie(pie, labels=pie.index, autopct='%1.1f%%')
plt.show()
```

from the above chart we can say that 99.2 percent of buyers are retailers and .8 % are B2B

In [ ]:
```
plt.scatter(x='Category', y='Size', data=df)
plt.xlabel('category')
plt.ylabel('size')
plt.title ('scatter')
```

we can easily findout what are the sizes availabel for the respective categories

In [ ]:
```
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()
```

In [ ]:
```
#tp=df['ship-state'].value_counts().head(10)
```

```
#sns.countplot(data=df[df['ship-state'].isin(tp.index)], x='ship-state')
#plt.xticks(rotation=90)
#plt.show()
```

from the above graph we can see that most of the sales are from Maharastra

--CONCLUSION-- The Data Analysis reveals that the business has significant customer base in Maharastra ,mainly retailers , fullfills orders through Amazon. Experiences high demand for T-shirts and sees M-size as the preferred choice among buyers..