

```

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

df=pd.read_csv('Users/dilthyachigiri/Documents/Data Analysis/Python/Amazon_Sales_Analysis-main/Amazon Sale Report.csv',encoding='unicode_escape')

In [9]:
df.shape
Out[9]:
(128976, 21)

In [10]:
df.head()
Out[10]:
   index  Order ID  Date      Status  Fulfillment  Sales Channel  ship-service-level  Category  Size  Courier Status  ...  currency  Amount  ship-city  ship-state  ship-postal-code  ship-country  B2B  fulfilled-by  New  Pending$
0      0  405-8078761-5731545  04-30-22  Shipped  Cancelled  Merchant  Amazon.in  Standard  T-shirt  S  On the Way  ...  INR  647.62  MUMBAI  MAHARASHTRA  400001.0  IN  False  Easy Ship  NaN  NaN
1      1  171-9181851-1101146  04-30-22  Shipped  Delivered to Buyer  Merchant  Amazon.in  Standard  Shirt  3XL  Shipped  ...  INR  405.00  BENGALURU  KARNATAKA  560005.0  IN  False  Easy Ship  NaN  NaN
2      2  404-0697679-7271146  04-30-22  Shipped  Amazon  Amazon.in  Expedited  Shirt  3XL  Shipped  ...  INR  329.00  NAVI MUMBAI  MAHARASHTRA  410210.0  IN  True  NaN  NaN  NaN
3      3  404-0513277-8132951  04-30-22  Cancelled  Merchant  Amazon.in  Expedited  Blazer  L  On the Way  ...  INR  395.00  PULCHERRY  TAMILNADU  600008.0  IN  False  Easy Ship  NaN  NaN
4      4  407-1089797-7240320  04-30-22  Shipped  Amazon  Amazon.in  Expedited  Trouvers  3XL  Shipped  ...  INR  674.00  CHENNAI  TAMIL NADU  600072.0  IN  False  NaN  NaN  NaN
5 rows x 21 columns

In [11]:
df.tail(10)
Out[11]:
   index  Order ID  Date      Status  Fulfillment  Sales Channel  ship-service-level  Category  Size  Courier Status  ...  currency  Amount  ship-city  ship-state  ship-postal-code  ship-country  B2B  fulfilled-by  New  Pending$
0      0  405-8078761-5731545  04-30-22  Shipped  Cancelled  Merchant  Amazon.in  Standard  T-shirt  S  On the Way  ...  INR  647.62  MUMBAI  MAHARASHTRA  400001.0  IN  False  Easy Ship  NaN  NaN
1      1  171-9181851-1101146  04-30-22  Shipped  Delivered to Buyer  Merchant  Amazon.in  Standard  Shirt  3XL  Shipped  ...  INR  405.00  BENGALURU  KARNATAKA  560005.0  IN  False  Easy Ship  NaN  NaN
2      2  404-0697679-7271146  04-30-22  Shipped  Amazon  Amazon.in  Expedited  Shirt  3XL  Shipped  ...  INR  329.00  NAVI MUMBAI  MAHARASHTRA  410210.0  IN  True  NaN  NaN  NaN
3      3  404-0513277-8132951  04-30-22  Cancelled  Merchant  Amazon.in  Expedited  Blazer  L  On the Way  ...  INR  395.00  PULCHERRY  TAMILNADU  600008.0  IN  False  Easy Ship  NaN  NaN
4      4  407-1089797-7240320  04-30-22  Shipped  Amazon  Amazon.in  Expedited  Trouvers  3XL  Shipped  ...  INR  674.00  CHENNAI  TAMIL NADU  600072.0  IN  False  NaN  NaN  NaN
5      5  404-1609841-4307705  04-30-22  Shipped  Amazon  Amazon.in  Expedited  T-shirt  XL  Shipped  ...  INR  824.00  CHAZHADAR  UTTAR PRADESH  201322.0  IN  False  NaN  NaN  NaN
6      6  404-1609841-4307705  04-30-22  Shipped  Amazon  Amazon.in  Expedited  T-shirt  XL  Shipped  ...  INR  824.00  CHAZHADAR  UTTAR PRADESH  201322.0  IN  False  NaN  NaN  NaN
7      7  405-7607737-7092945  04-30-22  Shipped  Delivered to Buyer  Merchant  Amazon.in  Standard  Shirt  S  Shipped  ...  INR  399.00  HYDERABAD  TELANGANA  500032.0  IN  False  Easy Ship  NaN  NaN
8      8  405-8449345-5201656  04-30-22  Cancelled  Merchant  Amazon.in  Expedited  T-shirt  3XL  Cancelled  ...  NaN  NaN  HYDERABAD  TELANGANA  500032.0  IN  False  NaN  NaN  NaN
9      9  402-8937951-0911320  04-30-22  Shipped  Amazon  Amazon.in  Expedited  Shirt  3XL  Shipped  ...  INR  666.00  Chennai  TAMIL NADU  600041.0  IN  False  NaN  NaN  NaN
10 rows x 21 columns

In [12]:
df.tail(10)
Out[12]:
   index  Order ID  Date      Status  Fulfillment  Sales Channel  ship-service-level  Category  Size  Courier Status  ...  currency  Amount  ship-city  ship-state  ship-postal-code  ship-country  B2B  fulfilled-by  New  Pending$
128971  128971  404-6001340-8773107  05-31-22  Shipped  Amazon  Amazon.in  Expedited  Shirt  XL  Shipped  ...  INR  517.0  HYDERABAD  TELANGANA  500033.0  IN  False  NaN  NaN  NaN
128972  128972  407-9647469-3103350  05-31-22  Shipped  Amazon  Amazon.in  Expedited  T-shirt  M  Shipped  ...  INR  599.0  GURUGRAM  HARYANA  122004.0  IN  False  NaN  NaN  NaN
128973  128973  407-9647469-3103350  05-31-22  Shipped  Amazon  Amazon.in  Expedited  Blazer  XL  Shipped  ...  INR  753.33  HYDERABAD  TELANGANA  500049.0  IN  False  NaN  NaN  NaN
128974  128974  402-4384140-8773107  05-31-22  Shipped  Amazon  Amazon.in  Expedited  T-shirt  XS  Shipped  ...  INR  1129.0  HALKI  Gujarat  382950.0  IN  False  NaN  NaN  NaN
128975  128975  404-7436040-8738312  05-31-22  Shipped  Amazon  Amazon.in  Expedited  T-shirt  S  Shipped  ...  INR  666.0  Rajpur CHATTISGARH  490014.0  IN  False  NaN  NaN  NaN
5 rows x 21 columns

In [13]:
df.info()
Out[13]:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 128976 entries, 0 to 128975
Data columns (total 22 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   Fulfillment            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   Courier Status           128976 non-null  object
9   Currency                 128976 non-null  object
10  Amount                 121176 non-null  float64
11  ship-city              128941 non-null  object
12  ship-state             128941 non-null  object
13  ship-postal-code        128941 non-null  float64
14  ship-country            128941 non-null  object
15  B2B                     8 non-null        float64
16  fulfilled-by            39223 non-null   object
17  Pending$                8 non-null        float64
18  New$                    128976 non-null  float64
dtypes: float64(4), int64(2), object(14)
memory usage: 19.8+ MB

In [14]:
df.info()
Out[14]:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 128976 entries, 0 to 128975
Data columns (total 22 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   Fulfillment            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   Courier Status           128976 non-null  object
9   Currency                 128976 non-null  object
10  Amount                 121176 non-null  float64
11  ship-city              128941 non-null  object
12  ship-state             128941 non-null  object
13  ship-postal-code        128941 non-null  float64
14  ship-country            128941 non-null  object
15  B2B                     8 non-null        float64
16  fulfilled-by            39223 non-null   object
17  Pending$                8 non-null        float64
18  New$                    128976 non-null  float64
dtypes: float64(4), int64(2), object(14)
memory usage: 17.8+ MB

In [15]:
# checking null value
pd.isnull(df)
Out[15]:
   index  Order ID  Date      Status  Fulfillment  Sales Channel  ship-service-level  Category  Size  Courier Status  ...  currency  Amount  ship-city  ship-state  ship-postal-code  ship-country  B2B  fulfilled-by
0  False  False  False  False  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  False  False  False  False
2  False  False  False  False  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  False  False  False  False
4  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  True
...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...  ...
128971  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  True
128972  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  True
128973  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  True
128974  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  True
128975  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  True
128976 rows x 19 columns

In [16]:
#sum will give total values of null values
pd.isnull(df).sum()
Out[16]:
index              0
Order ID           0
Date               0
Status             0
Fulfillment        0
Sales Channel       0
ship-service-level  0
Category           0
Courier Status      0
Currency           7808
Amount            7808
ship-city          35
ship-state         35
ship-postal-code   35
ship-country       35
B2B                8
fulfilled-by       8
Pending$           8
New$              89713
dtype: int64

In [17]:
df.shape
Out[17]:
(128976, 19)

In [18]:
#drop null values
df.dropna(inplace=True)

In [19]:
df.shape
Out
```