In [ ]:
```
#Loading data
import pandas as pd
obj=pd.read_csv("C:\\Users\\satya\\Documents\\Amazon Sale Report.csv")
```

In [39]:
```
# using info function to see important information about dataset
print(obj.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 128975 entries, 0 to 128974
Data columns (total 24 columns):
 #   Column              Non-Null Count   Dtype  
---  ------              --------------   -----  
 0   index               128975 non-null  int64  
 1   Order ID            128975 non-null  object 
 2   Date                128975 non-null  object 
 3   Status              128975 non-null  object 
 4   Fulfilment          128975 non-null  object 
 5   Sales Channel       128975 non-null  object 
 6   ship-service-level  128975 non-null  object 
 7   Style               128975 non-null  object 
 8   SKU                 128975 non-null  object 
 9   Category            128975 non-null  object 
 10  Size                128975 non-null  object 
 11  ASIN                128975 non-null  object 
 12  Courier Status      122103 non-null  object 
 13  Qty                 128975 non-null  int64  
 14  currency            121180 non-null  object 
 15  Amount              121180 non-null  float64
 16  ship-city           128942 non-null  object 
 17  ship-state          128942 non-null  object 
 18  ship-postal-code    128942 non-null  float64
 19  ship-country        128942 non-null  object 
 20  promotion-ids       79822 non-null   object 
 21  B2B                 128975 non-null  bool   
 22  fulfilled-by        39277 non-null   object 
 23  Unnamed: 22         79925 non-null   object 
dtypes: bool(1), float64(2), int64(2), object(19)
memory usage: 22.8+ MB
None
```

In [13]:
```
# to remove columns that contain more than 30 perecent null values
threshold=0.3
limit=len(obj) * threshold
objnew=obj.dropna(axis=1,thresh=len(obj) - limit)
print(objnew.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 128975 entries, 0 to 128974
Data columns (total 21 columns):
 #   Column              Non-Null Count   Dtype  
---  ------              --------------   -----  
 0   index               128975 non-null  int64  
 1   Order ID            128975 non-null  object 
 2   Date                128975 non-null  object 
 3   Status              128975 non-null  object 
 4   Fulfilment          128975 non-null  object 
 5   Sales Channel       128975 non-null  object 
 6   ship-service-level  128975 non-null  object 
 7   Style               128975 non-null  object 
 8   SKU                 128975 non-null  object 
 9   Category            128975 non-null  object 
 10  Size                128975 non-null  object 
 11  ASIN                128975 non-null  object 
 12  Courier Status      122103 non-null  object 
 13  Qty                 128975 non-null  int64  
 14  currency            121180 non-null  object 
 15  Amount              121180 non-null  float64
 16  ship-city           128942 non-null  object 
 17  ship-state          128942 non-null  object 
 18  ship-postal-code    128942 non-null  float64
 19  ship-country        128942 non-null  object 
 20  B2B                 128975 non-null  bool   
dtypes: bool(1), float64(2), int64(2), object(16)
memory usage: 19.8+ MB
None
```

In [14]:
```
# counting how many null values present in each column
print(objnew.isnull().sum())
```

```
index                   0
Order ID                0
Date                    0
Status                  0
Fulfilment              0
Sales Channel           0
ship-service-level      0
Style                   0
SKU                     0
Category                0
Size                    0
ASIN                    0
Courier Status       6872
Qty                     0
currency             7795
Amount               7795
ship-city              33
ship-state             33
ship-postal-code       33
ship-country           33
B2B                     0
dtype: int64
```

In [38]:
```
#filling Courier status missing  values
mode_value = objnew['Courier Status'].mode()[0]
objnew['Courier Status'] = objnew['Courier Status'].fillna(mode_value)
print(objnew['Courier Status'].isna().sum())
```

```
0
```

In [37]:
```
#filling currency status missing value
mode_values = objnew['currency'].mode()[0]
objnew['currency'] = objnew['currency'].fillna(mode_values)
print(objnew['currency'].isna().sum())
```

```
0
```

In [36]:
```
#filling Amount missing values
mean_amount = objnew['Amount'].mean()
objnew['Amount']=objnew['Amount'].fillna(objnew['Amount'].mean)
print(objnew['Amount'].isna().sum())
```

```
0
```

In [35]:
```
#filling ship-city missing values
mode_values = objnew['ship-city'].mode()[0]
objnew['ship-city'] = objnew['ship-city'].fillna(mode_values)
print(objnew['ship-city'].isna().sum())
```

```
0
```

In [34]:
```
#filling ship-state missing values
mode_values = objnew['ship-state'].mode()[0]
objnew['ship-state'] = objnew['ship-state'].fillna(mode_values)
print(objnew['ship-state'].isna().sum())
```

```
0
```

In [33]:
```
#filling postal-code missing values
mode_values = objnew['ship-postal-code'].mode()[0]
objnew['ship-postal-code'] = objnew['ship-postal-code'].fillna(mode_values)
print(objnew['ship-postal-code'].isna().sum())
```

```
0
```

In [32]:
```
#filling ship-country missing values
mode_values = objnew['ship-country'].mode()[0]
objnew['ship-country'] = objnew['ship-country'].fillna(mode_values)
print(objnew['ship-country'].isna().sum())
```

```
0
```

In [ ]:
```
```