Import libraries

```
import numpy as np
import pandas as pd
import matplotlib as mpl
import matplotlib.pyplot as plt
```

Import data

```
df_amazon_sales = pd.read_csv("E:/Python_Amazon_Sales_Analysis/Amazon_Sales_Analysis/Amazon Sales Report.csv", encoding='unicode_escape')
```

```
In [6]: df_amazon_sales
Out[6]:
            Index  Order ID  Date  ...  Courier Status  \
0               0  405-8078784  04-30-22  ...       On the Way
1               1  171-9198151  04-30-22  ...  Shipped - Delivered to Buyer
2               2  404-8824857  04-30-22  ...       Shipped
3               3  408-7436540  04-30-22  ...       Shipped
4               4  408-9690857  04-30-22  ...       Shipped
...           ...          ...       ...  ...           ...
128971     128971  402-9867403  05-31-22  ...       Shipped
128972     128972  407-9689760  05-31-22  ...       Shipped
128973     128973  407-7823160  05-31-22  ...       Shipped
128974     128974  408-7436540  05-31-22  ...       Shipped
128975     128975  408-8884388  05-31-22  ...       Shipped

                 Unfulfilled  Sales Channel  ship-service-level  Style  SKU  \
0                  Merchant  Amazon.in  Standard  T-shirt  S  SH-RL-TS-S
1                  Merchant  Amazon.in  Expedited  Kurta  XL  JNE3797-KR-XL
2                  Merchant  Amazon.in  Standard  Kurta  XL  JNE3405-KR-XL
3                  Merchant  Amazon.in  Expedited  Blouse  L  SAR8431-BL-L
4                  Merchant  Amazon.in  Expedited  Kurta  XXL  JNE3797-KR-XXL
...                     ...        ...        ...    ...  ...          ...
128971             Merchant  Amazon.in  Expedited  Kurta  XL  JNE3405-KR-XL
128972             Merchant  Amazon.in  Expedited  Kurta  XL  JNE3405-KR-XL
128973             Merchant  Amazon.in  Expedited  Kurta  XL  JNE3405-KR-XL
128974             Merchant  Amazon.in  Standard  Kurta  XL  JNE3405-KR-XL
128975             Merchant  Amazon.in  Standard  Kurta  XL  JNE3405-KR-XL

                 Category  Size  Courier Status  Qty  Amount  ship-city  \
0              Shirt  S  On the Way  0  647.62  MUMBAI
1              Kurta  XL  Shipped  1  406.00  BENGALURU
2              Kurta  XL  Shipped  1  329.00  NAVI MUMBAI
3              Blouse  L  On the Way  1  753.33  PUDUCHERRY
4              Kurta  XXL  Shipped  1  574.00  CHENNAI
...                ...  ...           ...  ...     ...        ...
128971         Kurta  XL  Shipped  1  329.00  CHANDIGARH
128972         Kurta  XL  Shipped  1  329.00  HYDERABAD
128973         Kurta  XL  Shipped  1  329.00  TELANGANA
128974         Kurta  XL  Shipped  1  329.00  HYDERABAD
128975         Kurta  XL  Shipped  1  329.00  Nagar

                 ship-state  ship-postal-code  ship-country  \
0              MAHARASHTRA  400001.0  IN
1              KARNATAKA  560082.0  IN
2              MAHARASHTRA  410210.0  IN
3              PUDUCHERRY  605008.0  IN
4              TAMIL NADU  600073.0  IN
...                   ...       ...       ...
128971         CHANDIGARH  160009.0  IN
128972         TELANGANA  500032.0  IN
128973         TELANGANA  500032.0  IN
128974         TELANGANA  500032.0  IN
128975         Gujarat  396195.0  IN

                 Fulfilled by  Unnamed: 22
0              Easy Ship  NaN
1              Easy Ship  NaN
2              Easy Ship  NaN
3              Easy Ship  NaN
4              Easy Ship  NaN
...                  ...      ...
128971         Easy Ship  NaN
128972         Easy Ship  NaN
128973         Easy Ship  NaN
128974         Easy Ship  NaN
128975         Easy Ship  NaN

[128976 rows x 23 columns]
```

```
In [7]: df.head()
Out[7]:
   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  Fulfilled by  Unnamed: 22
0      0  405-8078784  04-30-22  Cancelled  Merchant  Amazon.in  Standard  T-shirt  S  On the Way  0  INR  647.62  MUMBAI  MAHARASHTRA  400001.0  IN  Easy Ship  NaN
1      1  171-9198151  04-30-22  Shipped - Delivered to Buyer  Merchant  Amazon.in  Expedited  Kurta  XL  Shipped  1  INR  406.00  BENGALURU  KARNATAKA  560082.0  IN  Easy Ship  NaN
2      2  404-8824857  04-30-22  Shipped  Merchant  Amazon.in  Standard  Kurta  XL  Shipped  1  INR  329.00  NAVI MUMBAI  MAHARASHTRA  410210.0  IN  Easy Ship  NaN
3      3  408-7436540  04-30-22  Shipped  Merchant  Amazon.in  Expedited  Blouse  L  On the Way  1  INR  753.33  PUDUCHERRY  PUDUCHERRY  605008.0  IN  Easy Ship  NaN
4      4  408-9690857  04-30-22  Shipped  Merchant  Amazon.in  Expedited  Kurta  XXL  Shipped  1  INR  574.00  CHENNAI  TAMIL NADU  600073.0  IN  Easy Ship  NaN
5      5  402-9867403  05-31-22  Shipped  Amazon.in  Expedited  Kurta  XL  Shipped  1  INR  329.00  CHANDIGARH  CHANDIGARH  160009.0  IN  Easy Ship  NaN
6      6  407-9689760  05-31-22  Shipped  Amazon.in  Expedited  Kurta  XL  Shipped  1  INR  329.00  HYDERABAD  TELANGANA  500032.0  IN  Easy Ship  NaN
7      7  407-7823160  05-31-22  Shipped - Delivered to Buyer  Amazon.in  Expedited  Shirt  S  Shipped  1  INR  599.00  HYDERABAD  TELANGANA  500032.0  IN  Easy Ship  NaN
8      8  408-7436540  05-31-22  Shipped  Amazon.in  Expedited  T-shirt  XL  Canceled  1  INR  Not a sale  HYDERABAD  TELANGANA  500032.0  IN  Easy Ship  NaN
9      9  408-8884388  05-31-22  Shipped  Amazon.in  Standard  Kurta  XL  Shipped  1  INR  329.00  Nagar  Gujarat  396195.0  IN  Easy Ship  NaN

10 rows x 23 columns
```

```
In [8]: df.tail()
Out[8]:
        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  Fulfilled by  Unnamed: 22
128971  128971  402-9867403  05-31-22  Shipped  Amazon.in  Expedited  Kurta  XL  Shipped  1  INR  329.0  CHANDIGARH  CHANDIGARH  160009.0  IN  Easy Ship  NaN
128972  128972  407-9689760  05-31-22  Shipped  Amazon.in  Expedited  Kurta  XL  Shipped  1  INR  329.0  HYDERABAD  TELANGANA  500032.0  IN  Easy Ship  NaN
128973  128973  407-7823160  05-31-22  Shipped  Amazon.in  Expedited  Kurta  XL  Shipped  1  INR  329.0  HYDERABAD  TELANGANA  500032.0  IN  Easy Ship  NaN
128974  128974  408-7436540  05-31-22  Shipped  Amazon.in  Standard  Kurta  XL  Shipped  1  INR  329.0  HYDERABAD  TELANGANA  500032.0  IN  Easy Ship  NaN
128975  128975  408-8884388  05-31-22  Shipped  Amazon.in  Standard  Kurta  XL  Shipped  1  INR  329.0  Nagar  Gujarat  396195.0  IN  Easy Ship  NaN
```

```
In [9]: df.shape
Out[9]: (128976, 23)
```

```
In [10]: df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 128976 entries, 0 to 128975
Data columns (total 23 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            128976 non-null  object
 12  Amount              128976 non-null  float64
 13  ship-city           128975 non-null  object
 14  ship-state          128975 non-null  object
 15  ship-postal-code    128975 non-null  float64
 16  ship-country        128975 non-null  object
 17  Fulfilled by        39732 non-null  object
 18  Unnamed: 22         128976 non-null  object
dtypes: float64(2), int64(2), object(19)
memory usage: 18.7 MB
```

```
In [11]: df.isnull().sum()
Out[11]:
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                 1
ship-state                1
ship-postal-code          1
ship-country              1
Fulfilled by         89244
Unnamed: 22               0
dtype: int64
```

```
In [12]: df.isnull().sum()/df.shape*100
Out[12]:
Index                0.000000
Order ID             0.000000
Date                 0.000000
Status               0.000000
Fulfilment           0.000000
Sales Channel        0.000000
ship-service-level   0.000000
Category             0.000000
Size                 0.000000
Courier Status       0.000000
Qty                  0.000000
Currency             0.000000
Amount               0.000000
ship-city            0.000775
ship-state           0.000775
ship-postal-code     0.000775
ship-country         0.000775
Fulfilled by        69.190055
Unnamed: 22          0.000000
dtype: int64
```

```
In [13]: df.describe()
Out[13]:
              Index            Qty         Amount  ship-postal-code
count  128976.000000  128976.000000  128976.000000  128975.000000
mean    64487.500000       0.904160     642.826309  463401.207025
std     37232.853992       0.313354     281.211687  191932.276167
min         0.000000       0.000000       0.000000    110001.0000
25%     32243.250000       1.000000     449.000000    370480.0000
50%     64487.500000       1.000000     605.000000    500033.0000
75%     96731.750000       1.000000     788.000000    600042.0000
max    128971.000000      15.000000    9584.000000    989828.0000
```

```
In [14]: df.describe(include='object')
Out[14]:
                 Date  Status  Fulfilment  Sales Channel  ship-service-level  Category  Size  Courier Status  Currency  ship-city  ship-state  ship-country  Fulfilled by
count          128976  128976      128976         128976              128976    128976  128976          128976    128976     128975     128975        128975         7732
unique            365      13           2              1                   2        39      38               5         1      14836         69           1            2
top         05-13-22  Shipped  Merchant  Amazon.in  Expedited  Set  S  On the Way  INR  MUMBAI  MAHARASHTRA  IN  Easy Ship
freq           1004   77804       89244         128976              86250     32232   9299           77804    128976      14706      22644        128975         5794
```

```
In [15]: df["Amount"].sum()
Out[15]: 82916166.72
```

```
In [16]: df["Amount"].value_counts()
Out[16]:
Amount
605.00    13
649.00    12
329.00    11
399.00    11
384.00    10
          ...
1099.00    1
1149.00    1
1199.00    1
1249.00    1
1499.00    1
Name: count, Length: 1413, dtype: int64
```

```
In [17]: df.columns
Out[17]: 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',
       'Fulfilled by', 'Unnamed: 22'],
      dtype='object')
```

```
In [18]: df["ship-postal-code"].dtype
Out[18]: dtype('float64')
```

```
In [19]: df["ship-postal-code"].astype(int)
```

```
In [20]: df["ship-postal-code"].dtype
Out[20]: dtype('int64')
```

```
In [21]: df["Qty"].value_counts()
Out[21]:
Qty
1    128971
2         3
0         1
15        1
Name: count, dtype: int64
```

```
In [22]: df.rename(columns={'Qty':'shipment qty'})
```

```
In [23]: df
Out[23]:
        Index  Order ID  Date  Status  Fulfilment  Sales Channel  ship-service-level  Category  Size  Courier Status  shipment qty  Currency  Amount  ship-city  ship-state  ship-postal-code  ship-country  Fulfilled by  Unnamed: 22
0           0  405-8078784  04-30-22  Cancelled  Merchant  Amazon.in  Standard  T-shirt  S  On the Way  0  INR  647.62  MUMBAI  MAHARASHTRA  400001.0  IN  Easy Ship  NaN
1           1  171-9198151  04-30-22  Shipped - Delivered to Buyer  Merchant  Amazon.in  Expedited  Kurta  XL  Shipped  1  INR  406.00  BENGALURU  KARNATAKA  560082.0  IN  Easy Ship  NaN
2           2  404-8824857  04-30-22  Shipped  Merchant  Amazon.in  Standard  Kurta  XL  Shipped  1  INR  329.00  NAVI MUMBAI  MAHARASHTRA  410210.0  IN  Easy Ship  NaN
3           3  408-7436540  04-30-22  Shipped  Merchant  Amazon.in  Expedited  Blouse  L  On the Way  1  INR  753.33  PUDUCHERRY  PUDUCHERRY  605008.0  IN  Easy Ship  NaN
4           4  408-9690857  04-30-22  Shipped  Merchant  Amazon.in  Expedited  Kurta  XXL  Shipped  1  INR  574.00  CHENNAI  TAMIL NADU  600073.0  IN  Easy Ship  NaN
5           5  402-9867403  05-31-22  Shipped  Amazon.in  Expedited  Kurta  XL  Shipped  1  INR  329.00  CHANDIGARH  CHANDIGARH  160009.0  IN  Easy Ship  NaN
6           6  407-9689760  05-31-22  Shipped  Amazon.in  Expedited  Kurta  XL  Shipped  1  INR  329.00  HYDERABAD  TELANGANA  500032.0  IN  Easy Ship  NaN
7           7  407-7823160  05-31-22  Shipped - Delivered to Buyer  Amazon.in  Expedited  Shirt  S  Shipped  1  INR  599.00  HYDERABAD  TELANGANA  500032.0  IN  Easy Ship  NaN
8           8  408-7436540  05-31-22  Shipped  Amazon.in  Expedited  T-shirt  XL  Canceled  1  INR  Not a sale  HYDERABAD  TELANGANA  500032.0  IN  Easy Ship  NaN
9           9  408-8884388  05-31-22  Shipped  Amazon.in  Standard  Kurta  XL  Shipped  1  INR  329.00  Nagar  Gujarat  396195.0  IN  Easy Ship  NaN
10         10  407-5943161  05-31-22  Shipped  Amazon.in  Expedited  Set  XXL  Shipped  1  INR  1199.0  Vapi  Gujarat  396191.0  IN  Easy Ship  NaN
11         11  408-5436091  05-31-22  Shipped  Amazon.in  Expedited  Kurta  S  Shipped  1  INR  399.0  Raipur  CHHATTISGARH  492014.0  IN  Easy Ship  NaN

12 rows x 23 columns
```

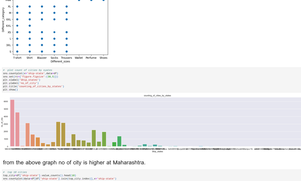
Exploratory Data Analysis

NO OF ITEMS WITH DIFFERENT SIZES

```
In [24]: sns.countplot(x='Size',data=df)
Out[24]: <Axes: xlabel='Size', count='count'>
In [25]: plt.title('No_of_items_with_different_sizes')
Out[25]: Text(0.5, 1.0, 'No_of_items_with_different_sizes')
In [26]: plt.show()
```

from the above graph M ,XLL sizes are sold highly in the market.

```
In [27]: df.groupby('Size')['ship_postal_code'].count().sort_values(ascending=False)
Out[27]: Size
M      32232
L      29304
XL     25305
XXL    15048
S      11559
3XL     5184
4XL     2000
5XL      165
6XL      179
Name: ship_postal_code, dtype: int64
```



from the above graph most no. of the sold item size is M and less no. of sold item size is 4XL

COURIER STATUS

```
In [28]: sns.countplot(x='Courier Status',data=df)
Out[28]: <Axes: xlabel='Courier Status', count='count'>
In [29]: plt.title('counting_of_different_courier_status')
Out[29]: Text(0.5, 1.0, 'counting_of_different_courier_status')
In [30]: plt.show()
```


from the above graph most of the product are on shipping condition.

```
In [31]: df[['Size']]
Out[31]: array(['S', 'XL', 'XXL', 'L', 'XL', 'XXL', 'XL', 'XL', 'XL', 'S',
       'XL', 'S', 'XL', 'S', 'XL', 'S', 'XL', 'S', 'XL', 'S
```