

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

[In (2)]: df=pd.read_csv('Amazon Sale Report.csv',encoding='unicode_escape')

[In (3)]: df.shape
Out(3): (128976, 21)

[In (4)]: df.head()
Out(4):
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  PendingS
0      0      405-8078784-  2022-04-30      Cancelled  Merchant  Amazon.in  Standard  T-shirt  S  On the Way  ...  INR  647.82  MUMBAI  MAHARASHTRA  400031.0  IN  False  Easy Ship  NaN  NaN
1      1      171-9198151-  04-30-22      Shipped  Delivered to Buyer  Amazon.in  Standard  Shirt  3XL  Shipped  ...  INR  406.00  BENGALURU  KARNATAKA  560005  IN  False  NaN  NaN  NaN
2      2      404-0697076-  04-30-22      Shipped  Amazon  Amazon.in  Expedited  Shirt  XL  Shipped  ...  INR  329.00  NAVI MUMBAI  MAHARASHTRA  401200.0  IN  True  NaN  NaN  NaN
3      3      403-9615077-  04-30-22      Cancelled  Merchant  Amazon.in  Standard  Blazer  L  On the Way  ...  INR  753.33  PUDUCHERRY  PUDUCHERRY  605008.0  IN  False  Easy Ship  NaN  NaN
4      4      407-1069790-  04-30-22      Shipped  Amazon  Amazon.in  Expedited  Trousers  S  Shipped  ...  INR  574.00  CHENNAI  TAMIL NADU  600073.0  IN  False  NaN  NaN  NaN
5 rows * 21 columns

[In (5)]: df.tail()
Out(5):
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  PendingS
128971  128970      406-401380-  05-31-22      Shipped  Amazon  Amazon.in  Expedited  Shirt  XL  Shipped  ...  INR  517.0  HYDERABAD  TELANGANA  500013.0  IN  False  NaN  NaN  NaN
128972  128971      402-951906-  05-31-22      Shipped  Amazon  Amazon.in  Expedited  T-shirt  M  Shipped  ...  INR  999.0  GURUGRAM  HARYANA  121004.0  IN  False  NaN  NaN  NaN
128973  128972      407-364968-  05-31-22      Shipped  Amazon  Amazon.in  Expedited  Blazer  XXL  Shipped  ...  INR  1199.0  HYDERABAD  TELANGANA  500049.0  IN  False  NaN  NaN  NaN
128974  128973      402-918410-  05-31-22      Shipped  Amazon  Amazon.in  Expedited  T-shirt  XS  Shipped  ...  INR  1199.0  Faisal  Gujarat  385000.0  IN  False  NaN  NaN  NaN
128975  128974      406-7430540-  05-31-22      Shipped  Amazon  Amazon.in  Expedited  T-shirt  S  Shipped  ...  INR  696.0  Rapur  CHHATTISGARH  492014.0  IN  False  NaN  NaN  NaN
5 rows * 21 columns

[In (6)]: df.info()
Out(6):
<class 'pandas.core.frame.DataFrame'>
Int64Index: 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   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   Size               128976 non-null  object
 9   Courier Status      128976 non-null  object
10  Qty               128976 non-null  int64
11  currency           12176 non-null   float64
12  Amount            12176 non-null   float64
13  ship-city          12894 non-null   object
14  ship-state         12894 non-null   object
15  ship-postal-code   12894 non-null   float64
16  ship-country       12894 non-null   object
17  B2B                12896 non-null   bool
18  fulfilled-by       39243 non-null   object
19  PendingS           0 non-null      float64
20  NewPending         0 non-null      float64
dtypes: bool(1), float64(10), int64(12), object(14)
memory usage: 19.4+ MB

[In (7)]: df.drop(["New","Pending"],axis=1,inplace=True)

[In (8)]: df.info()
Out(8):
<class 'pandas.core.frame.DataFrame'>
Int64Index: 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   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   Size               128976 non-null  object
 9   Courier Status      128976 non-null  object
10  Qty               12176 non-null   float64
11  currency           12176 non-null   float64
12  Amount            12176 non-null   float64
13  ship-city          12894 non-null   object
14  ship-state         12894 non-null   object
15  ship-postal-code   12894 non-null   float64
16  ship-country       12894 non-null   object
17  B2B                12896 non-null   bool
18  fulfilled-by       39243 non-null   object
dtypes: bool(1), float64(10), int64(12), object(14)
memory usage: 17.4+ MB

[In (9)]: pd.isnull(df)
Out(9):
index      Order ID  Date      Status  Fulfillment  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  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
2  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
4  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False
...
128971  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False
128972  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False
128973  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False
128974  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False
128975  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False  False
128976 rows * 19 columns

[In (10)]: pd.isnull(df).sum()
Out(10):
index      0
Order ID    0
Date        0
Status      0
Fulfillment 0
Sales Channel 0
ship-service-level 0
Size         0
Courier Status 0
Qty          0
currency     1900
Amount       1900
ship-city    35
ship-state   35
ship-postal-code 35
ship-country 35
B2B          0
fulfilled-by 8973
dtype: int64

[In (11)]: df.shape
Out(11): (128976, 19)

[In (12)]: df.dropna(inplace=True)

[In (13)]: df.shape
Out(13): (127314, 19)

[In (14)]: df.columns
Out(14):
index      'index', 'Order ID', 'Date', 'Status', 'Fulfillment', '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")
Out(15):
df["ship-postal-code"].dtype
Out(16):
dtype('int64')

[In (18)]: df["Date"]>pd.to_datetime(df["Date"])

[In (19)]: df.columns
Out(19):
index      'index', 'Order ID', 'Date', 'Status', 'Fulfillment', '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 (20)]: df.rename(columns={"Qty":"Quantity"})

[In (20)]:
index      Order ID  Date      Status  Fulfillment  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-  2022-04-30      Cancelled  Merchant  Amazon.in  Standard  T-shirt  S  On the Way  0  INR  647.82  MUMBAI  MAHARASHTRA  400031  IN  False  Easy Ship
1      1      171-9198151-  2022-04-30      Shipped  Delivered to Buyer  Amazon.in  Standard  Shirt  3XL  Shipped  1  INR  406.00  BENGALURU  KARNATAKA  560005  IN  False  Easy Ship
3      3      403-9615077-  2022-04-30      Cancelled  Merchant  Amazon.in  Standard  Blazer  L  On the Way  0  INR  753.33  PUDUCHERRY  PUDUCHERRY  605008  IN  False  Easy Ship
7      7      406-7807733-  2022-04-30      Shipped  Delivered to Buyer  Amazon.in  Standard  Shirt  S  Shipped  1  INR  389.00  HYDERABAD  TELANGANA  500032  IN  False  Easy Ship
12     12     405-511264-  2022-04-30      Shipped  Delivered to Buyer  Amazon.in  Standard  Shirt  XS  Shipped  1  INR  599.00  Annamalai  MAHARASHTRA  444606  IN  False  Easy Ship
...
128971  128970      406-4724037-  2022-06-01      Shipped  Delivered to Buyer  Merchant  Amazon.in  Standard  T-shirt  S  Shipped  1  INR  854.00  ALLUR  ANDHRA PRADESH  524315  IN  False  Easy Ship
128976  128975      403-9624128-  2022-06-01      Cancelled  Merchant  Amazon.in  Standard  Blazer  XL  On the Way  0  INR  734.29  BARANAKI  UTTAR PRADESH  225001  IN  False  Easy Ship
128888  128987      404-07296-  2022-05-31      Shipped  Delivered to Buyer  Merchant  Amazon.in  Standard  Trousers  M  Shipped  1  INR  518.00  NOIDA  UTTAR PRADESH  201301  IN  False  Easy Ship
128891  128990      407-016396-  2022-05-31      Cancelled  Merchant  Amazon.in  Standard  Wallet  Free  On the Way  0  INR  398.10  MAQURAI  TAMIL NADU  625007  IN  False  Easy Ship
128982  128991      403-0317423-  2022-05-31      Shipped  Delivered to Buyer  Merchant  Amazon.in  Standard  Blazer  M  Shipped  1  INR  721.00  OTTARA BADOOGRA  WEST BENGAL  734014  IN  False  Easy Ship
37514 rows * 19 columns

[In (21)]: df.describe()
Out(21):
index      Order ID  Date      Status  Fulfillment  Sales Channel  ship-service-level  Category  Size  Courier Status  Qty  currency  Amount  ship-city  ship-state  ship-postal-code  ship-country  B2B  fulfilled-by
count      37514      37514      37514      37514      37514      37514      37514      37514      37514      37514      37514      37514      37514      37514      37514      37514      37514
min      0.000000      2022-05-11 07:56:47.30383840  0.867383  646.55360  463297.582754  0.000000  0.000000  110001.000000
25%      27235.250000      2022-04-20 09:00:00  1.000000  458.000000  370465.000000
50%      63470.500000      2022-05-09 09:00:00  1.000000  629.000000  600019.000000
75%      91790.750000      2022-06-01 09:00:00  1.000000  771.000000  800042.000000
max      38844.800000      2022-06-29 09:00:00  1.000000  5495.000000  988998.000000
std      36844.853039      NaN  0.564160  279.352414  194650.426537

[In (22)]: df.describe(include='object')
Out(22):
index      Order ID  Date      Status  Fulfillment  Sales Channel  ship-service-level  Category  Size  Courier Status  Qty  currency  Amount  ship-city  ship-state  ship-postal-code  ship-country  B2B  fulfilled-by
count      37514      37514      37514      37514      37514      37514      37514      37514      37514      37514      37514      37514      37514      37514      37514      37514      37514
unique      34644
top  171-5057375-283560  Shipped  Delivered to Buyer  Merchant  Amazon.in  Standard  T-shirt  M  Shipped  INR  BENGALURU  MAHARASHTRA  IN  Easy Ship
freq      12      26741  37514  37514  37514  14062  6806  31859  37514  2839  6236  37514  37514

[In (23)]: df[["Qty","Amount"]].describe()
Out(23):
Qty  Amount
count  37514.000000  37514.000000
mean    0.867383    646.553600
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

Exploratory Data Analysis

[In (24)]: df.columns
Out(24):
index      'index', 'Order ID', 'Date', 'Status', 'Fulfillment', '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

Size

[In (25)]: %sns.countplot(x="Size",data=df)
Out(25):


[In (26)]: %sns.boxplot(x="Size",data=df)
Out(26):


Note: From above Graph you can see that most of the people buys M-Size

[In (27)]: df.groupby(["Size"],as_index=False)["Qty"].sum().sort_values(by="Qty",ascending=False)
Out(27):
Size  Qty
5  M  5905
6  L  5795
8  XL  5481
10  XXL  4465
7  S  3972
9  XS  2191
4  Free  667
3  6XL  170
2  5XL  104
1  4XL  93

[In (28)]: %sns.jointplot(x="Size",y="Qty",data=df,kind="scatter")
Out(28):


Note: From above Graph you can see that most of the Qty buys M-Size in the sales

Courier Status

[In (29)]: sns.countplot(data=df,ax="Courier Status",hue="Status")
Out(29):


[In (30)]: plt.figure(figsize=(10,3))
sns.countplot(data=df,ax="Courier Status",hue="Status")
plt.show()
Out(30):


Note: From above Graph the majority of the orders are shipped through the courier.

[In (31)]: df["Size"].hist()
Out(31):


[In (32)]: <Axes: >
Out(32):


Note: From above Graph you can see that most of the buyers are T-shirt

[In (33)]: B2B_check = df["B2B"].value_counts()
plt.pie(B2B_check,labels=B2B_check.autoticks["%1.1f%%"],
plt.show()
Out(33):


Note: From above chart you can see that maximum i.e. 99.3% of buyers are retailers and 0.7% are B2B buyers

[In (34)]: B2B_Check = df["B2B"].value_counts()
plt.pie(B2B_Check,labels=B2B_Check.autoticks["%1.1f%%"],
plt.show()
Out(34):


Note: From above chart you can see that most of the Fulfillment are amazon

[In (35)]: %data = df[["Category"]]
%data = df["Size"]
plt.scatter(x=data,y=data)
plt.xlabel("Category")
plt.ylabel("Size")
plt.title("Scatter Plot")
plt.show()
Out(35):


[In (37)]: plt.figure(figsize=(12, 6))
sns.countplot(data=df,ax="ship-state")
plt.xlabel("ship-state")
plt.ylabel("count")
plt.title("Distribution of State")
plt.show()
Out(37):


[In (38)]: ship_state = df["ship-state"].value_counts().head(10)
plt.figure(figsize=(12, 6))
sns.countplot(data=df,ax="ship-state")
plt.xlabel("ship-state")
plt.ylabel("count")
plt.title("Distribution of State")
plt.show()
Out(38):


Note: From above Graph you can see that most of the buyers are Maharashtra state

Conclusion

The data analysis reveals that the business has a significant customer base in Maharashtra state, mainly serves retailers, fulfills orders through Amazon, experiences high demand for T-shirts, and sees M-Size as the preferred choice among buyers.
```