# Innobyte services Intership task

Amazon Sales report analysis

Name : Rubin Jain

# Introduction

In this report, we are analysing the amazon sales report data which containes the sales data of different amazon product to from 1 april to 1 july to provide useful business insights for expanding the business and understanding sales patterns.

# Libraries used

numpy

pandas

seaborn

matplotlib

In [1]:
```
#Importing Libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
C:\Users\Dell\anaconda3\lib\site-packages\scipy\__init__.py:155: UserWarning:
A NumPy version >=1.18.5 and <1.25.0 is required for this version of SciPy (d
etected version 1.26.4
  warnings.warn(f"A NumPy version >={np_minversion} and <{np_maxversion}"
```

In [3]:
```
#Importing Dataset
data=pd.read_csv("C:/Users/Dell/Downloads/Amazon Sale Report.csv",encodin
```

In [4]:
```
1 data
```

Out[4]:

| | index | Order ID | Date | Status | Fulfilment | Sales Channel | ship-service-level | Category | Size | Co St |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 0 | 405-8078784-5731545 | 04-30-22 | Cancelled | Merchant | Amazon.in | Standard | T-shirt | S | Or |
| **1** | 1 | 171-9198151-1101146 | 04-30-22 | Shipped - Delivered to Buyer | Merchant | Amazon.in | Standard | Shirt | 3XL | Shi |
| **2** | 2 | 404-0687676-7273146 | 04-30-22 | Shipped | Amazon | Amazon.in | Expedited | Shirt | XL | Shi |
| **3** | 3 | 403-9615377-8133951 | 04-30-22 | Cancelled | Merchant | Amazon.in | Standard | Blazzer | L | Or |
| **4** | 4 | 407-1069790-7240320 | 04-30-22 | Shipped | Amazon | Amazon.in | Expedited | Trousers | 3XL | Shi |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| **128971** | 128970 | 406-6001380-7673107 | 05-31-22 | Shipped | Amazon | Amazon.in | Expedited | Shirt | XL | Shi |
| **128972** | 128971 | 402-9551604-7544318 | 05-31-22 | Shipped | Amazon | Amazon.in | Expedited | T-shirt | M | Shi |
| **128973** | 128972 | 407-9547469-3152358 | 05-31-22 | Shipped | Amazon | Amazon.in | Expedited | Blazzer | XXL | Shi |
| **128974** | 128973 | 402-6184140-0545956 | 05-31-22 | Shipped | Amazon | Amazon.in | Expedited | T-shirt | XS | Shi |
| **128975** | 128974 | 408-7436540-8728312 | 05-31-22 | Shipped | Amazon | Amazon.in | Expedited | T-shirt | S | Shi |

128976 rows × 21 columns

# Performing EDA

In [5]:
```
1 data.head()
```

Out[5]:

| | index | Order ID | Date | Status | Fulfilment | Sales Channel | ship-service-level | Category | Size | Courier Status | .. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 0 | 405-8078784-5731545 | 04-30-22 | Cancelled | Merchant | Amazon.in | Standard | T-shirt | S | On the Way | .. |
| **1** | 1 | 171-9198151-1101146 | 04-30-22 | Shipped - Delivered to Buyer | Merchant | Amazon.in | Standard | Shirt | 3XL | Shipped | .. |
| **2** | 2 | 404-0687676-7273146 | 04-30-22 | Shipped | Amazon | Amazon.in | Expedited | Shirt | XL | Shipped | .. |
| **3** | 3 | 403-9615377-8133951 | 04-30-22 | Cancelled | Merchant | Amazon.in | Standard | Blazzer | L | On the Way | .. |
| **4** | 4 | 407-1069790-7240320 | 04-30-22 | Shipped | Amazon | Amazon.in | Expedited | Trousers | 3XL | Shipped | .. |

5 rows × 21 columns

In [6]:
```
1 len(data.columns)
```

Out[6]:
```
21
```

In [7]:
```
1 data.columns
```

Out[7]:
```
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', 'New', 'PendingS'],
      dtype='object')
```

In [8]:
```
1 data.tail()
```

Out[8]:

| | index | Order ID | Date | Status | Fulfilment | Sales Channel | ship-service-level | Category | Size | Cour Stat |
|---|---|---|---|---|---|---|---|---|---|---|
| **128971** | 128970 | 406-6001380-7673107 | 05-31-22 | Shipped | Amazon | Amazon.in | Expedited | Shirt | XL | Shipp |
| **128972** | 128971 | 402-9551604-7544318 | 05-31-22 | Shipped | Amazon | Amazon.in | Expedited | T-shirt | M | Shipp |
| **128973** | 128972 | 407-9547469-3152358 | 05-31-22 | Shipped | Amazon | Amazon.in | Expedited | Blazzer | XXL | Shipp |
| **128974** | 128973 | 402-6184140-0545956 | 05-31-22 | Shipped | Amazon | Amazon.in | Expedited | T-shirt | XS | Shipp |
| **128975** | 128974 | 408-7436540-8728312 | 05-31-22 | Shipped | Amazon | Amazon.in | Expedited | T-shirt | S | Shipp |

5 rows × 21 columns

In [9]:
```
1 data.describe()
```

Out[9]:

| | index | Qty | Amount | ship-postal-code | New | PendingS |
|---|---|---|---|---|---|---|
| **count** | 128976.000000 | 128976.000000 | 121176.000000 | 128941.000000 | 0.0 | 0.0 |
| **mean** | 64486.130427 | 0.904401 | 648.562176 | 463945.677744 | NaN | NaN |
| **std** | 37232.897832 | 0.313368 | 281.185041 | 191458.488954 | NaN | NaN |
| **min** | 0.000000 | 0.000000 | 0.000000 | 110001.000000 | NaN | NaN |
| **25%** | 32242.750000 | 1.000000 | 449.000000 | 382421.000000 | NaN | NaN |
| **50%** | 64486.500000 | 1.000000 | 605.000000 | 500033.000000 | NaN | NaN |
| **75%** | 96730.250000 | 1.000000 | 788.000000 | 600024.000000 | NaN | NaN |
| **max** | 128974.000000 | 15.000000 | 5584.000000 | 989898.000000 | NaN | NaN |

In [10]:
```
1 data.info()
```

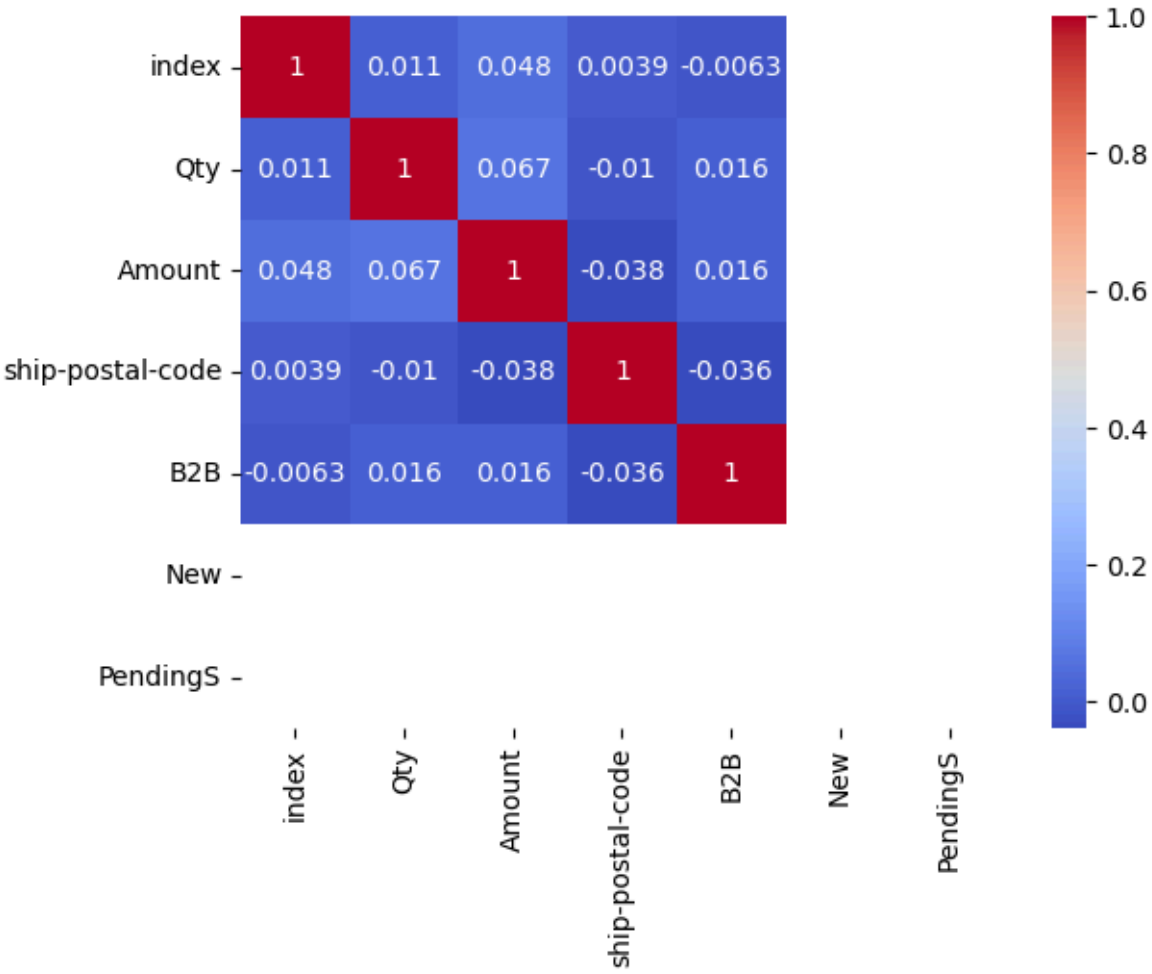
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 128976 entries, 0 to 128975
Data columns (total 21 columns):
 #   Column              Non-Null Count   Dtype  
---  ------              --------------   -----  
 0   index               128976 non-null  int64  
 1   Order ID            128976 non-null  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 [11]:
```
1 data.isna().sum()
```

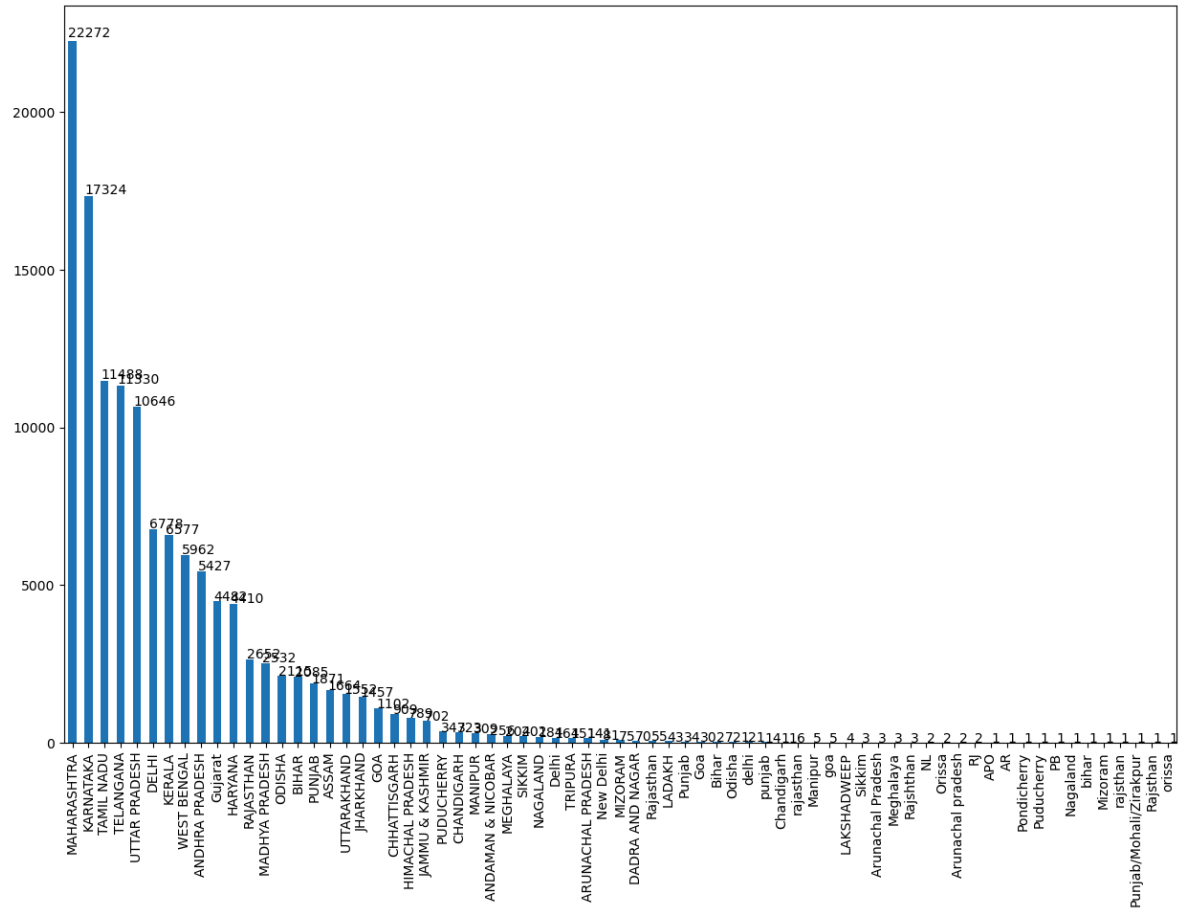
Out[11]:
```
index                     0
Order ID                  0
Date                      0
Status                    0
Fulfilment                0
Sales Channel             0
ship-service-level        0
Category                  0
Size                      0
Courier Status            0
Qty                       0
currency               7800
Amount                 7800
ship-city                35
ship-state               35
ship-postal-code         35
ship-country             35
B2B                       0
fulfilled-by          89713
New                  128976
PendingS             128976
dtype: int64
```

In [12]:
```
#correlation among the numerical columns
sns.heatmap(data.corr(),annot=True,cmap='coolwarm')
```

Out[12]: <AxesSubplot:>



In [13]:
```
#Number of orders shipped in each state of the country
plt.figure(figsize=(15, 10))
ax=data['ship-state'].value_counts().plot(kind='bar')
for p in ax.patches:
        ax.annotate(str(p.get_height()), (p.get_x() * 1.005, p.get_height
plt.show()
```




In [14]:
```
state_wise_orders=dict(data['ship-state'].value_counts().sort_values(asce
```

In [44]:
```
1 state_wise_orders
```

Out[44]:

```
{'MAHARASHTRA': 22272,
 'KARNATAKA': 17324,
 'TAMIL NADU': 11488,
 'TELANGANA': 11330,
 'UTTAR PRADESH': 10646,
 'DELHI': 6778,
 'KERALA': 6577,
 'WEST BENGAL': 5962,
 'ANDHRA PRADESH': 5427,
 'Gujarat': 4482,
 'HARYANA': 4410,
 'RAJASTHAN': 2652,
 'MADHYA PRADESH': 2532,
 'ODISHA': 2115,
 'BIHAR': 2085,
 'PUNJAB': 1871,
 'ASSAM': 1664,
 'UTTARAKHAND': 1552,
 'JHARKHAND': 1457,
 'GOA': 1102,
 'CHHATTISGARH': 909,
 'HIMACHAL PRADESH': 789,
 'JAMMU & KASHMIR': 702,
 'PUDUCHERRY': 347,
 'CHANDIGARH': 323,
 'MANIPUR': 309,
 'ANDAMAN & NICOBAR': 256,
 'MEGHALAYA': 204,
 'SIKKIM': 202,
 'NAGALAND': 184,
 'Delhi': 164,
 'TRIPURA': 151,
 'ARUNACHAL PRADESH': 141,
 'New Delhi': 81,
 'MIZORAM': 75,
 'DADRA AND NAGAR': 70,
 'Rajasthan': 55,
 'LADAKH': 43,
 'Punjab': 34,
 'Goa': 30,
 'Bihar': 27,
 'Odisha': 21,
 'delhi': 21,
 'punjab': 14,
 'Chandigarh': 11,
 'rajasthan': 6,
 'Manipur': 5,
 'goa': 5,
 'LAKSHADWEEP': 4,
 'Meghalaya': 3,
 'Rajshthan': 3,
 'Arunachal Pradesh': 3,
 'Sikkim': 3,
 'NL': 2,
 'Orissa': 2,
 'Arunachal pradesh': 2,
 'RJ': 2,
```

```
'bihar': 1,
'Rajsthan': 1,
'Punjab/Mohali/Zirakpur': 1,
'rajsthan': 1,
'Mizoram': 1,
'Puducherry': 1,
'Nagaland': 1,
'PB': 1,
'Pondicherry': 1,
'AR': 1,
'APO': 1,
'orissa': 1}
```

As we can see that the name of state is not in same pattern and because of that same state with capital and small starting letter is counted as different so we convert all of them to capitalize format and reobserve the orders in particular states

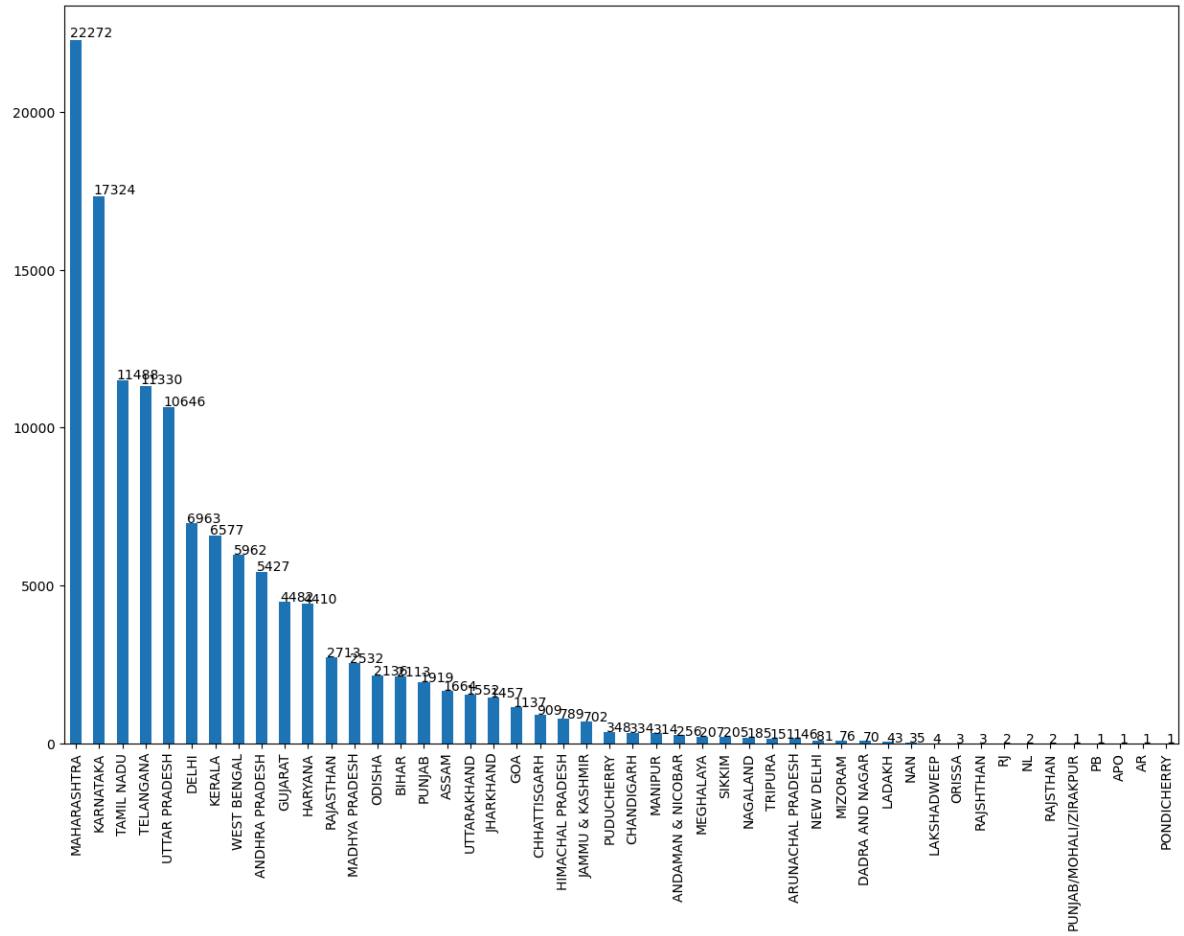
In [16]:
```
data['ship-state']=data['ship-state'].apply(lambda x:str(x))
```

In [18]:
```
data['ship-state']=data['ship-state'].apply(lambda x:x.upper())
```

In [22]:
```
state_wise_orders=dict(data['ship-state'].value_counts().sort_values(asce
state_wise_orders
```

Out[22]:
```
{'MAHARASHTRA': 22272,
 'KARNATAKA': 17324,
 'TAMIL NADU': 11488,
 'TELANGANA': 11330,
 'UTTAR PRADESH': 10646,
 'DELHI': 6963,
 'KERALA': 6577,
 'WEST BENGAL': 5962,
 'ANDHRA PRADESH': 5427,
 'GUJARAT': 4482,
 'HARYANA': 4410,
 'RAJASTHAN': 2713,
 'MADHYA PRADESH': 2532,
 'ODISHA': 2136,
 'BIHAR': 2113,
 'PUNJAB': 1919,
 'ASSAM': 1664,
 'UTTARAKHAND': 1552,
 'JHARKHAND': 1457,
 'GOA': 1137,
 'CHHATTISGARH': 909,
 'HIMACHAL PRADESH': 789,
 'JAMMU & KASHMIR': 702,
 'PUDUCHERRY': 348,
 'CHANDIGARH': 334,
 'MANIPUR': 314,
 'ANDAMAN & NICOBAR': 256,
 'MEGHALAYA': 207,
 'SIKKIM': 205,
 'NAGALAND': 185,
 'TRIPURA': 151,
 'ARUNACHAL PRADESH': 146,
 'NEW DELHI': 81,
 'MIZORAM': 76,
 'DADRA AND NAGAR': 70,
 'LADAKH': 43,
 'NAN': 35,
 'LAKSHADWEEP': 4,
 'ORISSA': 3,
 'RAJSHTHAN': 3,
 'RJ': 2,
 'NL': 2,
 'RAJSTHAN': 2,
 'PUNJAB/MOHALI/ZIRAKPUR': 1,
 'PB': 1,
 'APO': 1,
 'AR': 1,
 'PONDICHERRY': 1}
```

In [20]:
```
plt.figure(figsize=(15, 10))
ax=data['ship-state'].value_counts().plot(kind='bar')
for p in ax.patches:
        ax.annotate(str(p.get_height()), (p.get_x() * 1.005, p.get_height
plt.show()
```




In [24]:
```
data.duplicated().sum()
```

Out[24]: 168

In [25]:
```
data.drop_duplicates(inplace=True)
```

In [27]:
```
data.duplicated().sum()
```

Out[27]: 0

In [30]:
```
1 data.isna().sum()
```

Out[30]:
```
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             7789
Amount               7789
ship-city              33
ship-state              0
ship-postal-code       33
ship-country           33
B2B                     0
fulfilled-by        89595
New                128808
PendingS           128808
dtype: int64
```

As we can see that the column new and pending contains only null values so we can drop these columns also we have no need for index column because we have dataframe indexing

In [32]:
```
data.drop(columns=['New','PendingS','index'],inplace=True)
data
```

Out[32]:

| | Order ID | Date | Status | Fulfilment | Sales Channel | ship-service-level | Category | Size | Courier Status | Qty |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 405-8078784-5731545 | 04-30-22 | Cancelled | Merchant | Amazon.in | Standard | T-shirt | S | On the Way | |
| **1** | 171-9198151-1101146 | 04-30-22 | Shipped - Delivered to Buyer | Merchant | Amazon.in | Standard | Shirt | 3XL | Shipped | |
| **2** | 404-0687676-7273146 | 04-30-22 | Shipped | Amazon | Amazon.in | Expedited | Shirt | XL | Shipped | |
| **3** | 403-9615377-8133951 | 04-30-22 | Cancelled | Merchant | Amazon.in | Standard | Blazzer | L | On the Way | |
| **4** | 407-1069790-7240320 | 04-30-22 | Shipped | Amazon | Amazon.in | Expedited | Trousers | 3XL | Shipped | |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **128971** | 406-6001380-7673107 | 05-31-22 | Shipped | Amazon | Amazon.in | Expedited | Shirt | XL | Shipped | |
| **128972** | 402-9551604-7544318 | 05-31-22 | Shipped | Amazon | Amazon.in | Expedited | T-shirt | M | Shipped | |
| **128973** | 407-9547469-3152358 | 05-31-22 | Shipped | Amazon | Amazon.in | Expedited | Blazzer | XXL | Shipped | |
| **128974** | 402-6184140-0545956 | 05-31-22 | Shipped | Amazon | Amazon.in | Expedited | T-shirt | XS | Shipped | |
| **128975** | 408-7436540-8728312 | 05-31-22 | Shipped | Amazon | Amazon.in | Expedited | T-shirt | S | Shipped | |

128808 rows × 18 columns

In [33]:
```
1 data.isna().sum()
```

Out[33]:
```
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             7789
Amount               7789
ship-city              33
ship-state              0
ship-postal-code       33
ship-country           33
B2B                     0
fulfilled-by        89595
dtype: int64
```

In [36]:
```
1 data[data['Qty']==0]
```

Out[36]:

| | Order ID | Date | Status | Fulfilment | Sales Channel | ship-service-level | Category | Size | Courier Status | Qty | cur |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 405-8078784-5731545 | 04-30-22 | Cancelled | Merchant | Amazon.in | Standard | T-shirt | S | On the Way | 0 | |
| **3** | 403-9615377-8133951 | 04-30-22 | Cancelled | Merchant | Amazon.in | Standard | Blazzer | L | On the Way | 0 | |
| **8** | 407-5443024-5233168 | 04-30-22 | Cancelled | Amazon | Amazon.in | Expedited | T-shirt | 3XL | Cancelled | 0 | |
| **3** | 404-6019946-2909948 | 04-30-22 | Cancelled | Merchant | Amazon.in | Standard | T-shirt | M | On the Way | 0 | |
| **9** | 404-5933402-8801952 | 04-30-22 | Cancelled | Merchant | Amazon.in | Standard | Shirt | 3XL | On the Way | 0 | |
| **..** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| **4** | 406-3923120-4345139 | 05-31-22 | Cancelled | Amazon | Amazon.in | Expedited | Shirt | M | Cancelled | 0 | |
| **8** | 403-6757403-6097100 | 05-31-22 | Cancelled | Amazon | Amazon.in | Expedited | T-shirt | 3XL | Cancelled | 0 | |
| **9** | 403-6757403-6097100 | 05-31-22 | Cancelled | Amazon | Amazon.in | Expedited | T-shirt | 3XL | Cancelled | 0 | |
| **9** | 408-9513596-4393945 | 05-31-22 | Cancelled | Amazon | Amazon.in | Expedited | T-shirt | L | Cancelled | 0 | |
| **8** | 404-5182288-1653947 | 05-31-22 | Cancelled | Amazon | Amazon.in | Expedited | Shirt | XS | Cancelled | 0 | |

rows × 18 columns

In [37]:
```
1 data['Qty'].value_counts()
```

Out[37]:
```
1     115629
0      12792
2        340
3         32
4          9
5          2
15         1
9          1
13         1
8          1
Name: Qty, dtype: int64
```

In [39]:
```
#We know that quantity of order should be non zero so we discard entries w
data=data[data['Qty']>0]
data
```

Out[39]:

| | Order ID | Date | Status | Fulfilment | Sales Channel | ship-service-level | Category | Size | Courier Status | Qty |
|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 171-9198151-1101146 | 04-30-22 | Shipped - Delivered to Buyer | Merchant | Amazon.in | Standard | Shirt | 3XL | Shipped | 1 |
| **2** | 404-0687676-7273146 | 04-30-22 | Shipped | Amazon | Amazon.in | Expedited | Shirt | XL | Shipped | 1 |
| **4** | 407-1069790-7240320 | 04-30-22 | Shipped | Amazon | Amazon.in | Expedited | Trousers | 3XL | Shipped | 1 |
| **5** | 404-1490984-4578765 | 04-30-22 | Shipped | Amazon | Amazon.in | Expedited | T-shirt | XL | Shipped | 1 |
| **6** | 408-5748499-6859555 | 04-30-22 | Shipped | Amazon | Amazon.in | Expedited | T-shirt | L | Shipped | 1 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | .. |
| **128971** | 406-6001380-7673107 | 05-31-22 | Shipped | Amazon | Amazon.in | Expedited | Shirt | XL | Shipped | 1 |
| **128972** | 402-9551604-7544318 | 05-31-22 | Shipped | Amazon | Amazon.in | Expedited | T-shirt | M | Shipped | 1 |
| **128973** | 407-9547469-3152358 | 05-31-22 | Shipped | Amazon | Amazon.in | Expedited | Blazzer | XXL | Shipped | 1 |
| **128974** | 402-6184140-0545956 | 05-31-22 | Shipped | Amazon | Amazon.in | Expedited | T-shirt | XS | Shipped | 1 |
| **128975** | 408-7436540-8728312 | 05-31-22 | Shipped | Amazon | Amazon.in | Expedited | T-shirt | S | Shipped | 1 |

116016 rows × 18 columns

In [40]:
```
data.dropna(subset = ['Amount', 'currency'], inplace = True)
data.dropna(subset = ['ship-city', 'ship-state', 'ship-postal-code', 'shi
data
```

```
C:\Users\Dell\AppData\Local\Temp\ipykernel_14516\1884413495.py:1: SettingWith
CopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/s
table/user_guide/indexing.html#returning-a-view-versus-a-copy (https://panda
s.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-ver
sus-a-copy)
  data.dropna(subset = ['Amount', 'currency'], inplace = True)
C:\Users\Dell\AppData\Local\Temp\ipykernel_14516\1884413495.py:2: SettingWith
CopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/s
table/user_guide/indexing.html#returning-a-view-versus-a-copy (https://panda
s.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-ver
sus-a-copy)
  data.dropna(subset = ['ship-city', 'ship-state', 'ship-postal-code', 'ship-
country'], inplace = True)
```

Out[40]:

| Status | Fulfilment | Sales Channel | ship-service-level | Category | Size | Courier Status | Qty | currency | Amount | |
|---|---|---|---|---|---|---|---|---|---|---|
| Shipped - Delivered to Buyer | Merchant | Amazon.in | Standard | Shirt | 3XL | Shipped | 1 | INR | 406.0 | BE |
| Shipped | Amazon | Amazon.in | Expedited | Shirt | XL | Shipped | 1 | INR | 329.0 | NA |
| Shipped | Amazon | Amazon.in | Expedited | Trousers | 3XL | Shipped | 1 | INR | 574.0 | |
| Shipped | Amazon | Amazon.in | Expedited | T-shirt | XL | Shipped | 1 | INR | 824.0 | G |
| Shipped | Amazon | Amazon.in | Expedited | T-shirt | L | Shipped | 1 | INR | 653.0 | CH |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| Shipped | Amazon | Amazon.in | Expedited | Shirt | XL | Shipped | 1 | INR | 517.0 | HY |
| Shipped | Amazon | Amazon.in | Expedited | T-shirt | M | Shipped | 1 | INR | 999.0 | G |
| Shipped | Amazon | Amazon.in | Expedited | Blazzer | XXL | Shipped | 1 | INR | 690.0 | HY |
| Shipped | Amazon | Amazon.in | Expedited | T-shirt | XS | Shipped | 1 | INR | 1199.0 | |
| Shipped | Amazon | Amazon.in | Expedited | T-shirt | S | Shipped | 1 | INR | 696.0 | |

ıns

In [41]:
```
1 data.isna().sum()
```

Out[41]:
```
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        83524
dtype: int64
```

In [44]:
```
1 total_rows=len(data)
2 null_in_fulfilledby=data['fulfilled-by'].isna().sum()
```

In [46]:
```
1 null_percentage=(null_in_fulfilledby/total_rows)*100
2 null_percentage
```

Out[46]:
```
72.08796520058
```

Over 70% of data entry in fulfilled-by column is null values so we should drop that column

In [47]:
```
data.drop(columns = ['fulfilled-by'], inplace = True)
data
```

```
C:\Users\Dell\AppData\Local\Temp\ipykernel_14516\440946415.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)
  data.drop(columns = ['fulfilled-by'], inplace = True)
```

Out[47]:

| | Order ID | Date | Status | Fulfilment | Sales Channel | ship-service-level | Category | Size | Courier Status | Qty |
|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 171-9198151-1101146 | 04-30-22 | Shipped - Delivered to Buyer | Merchant | Amazon.in | Standard | Shirt | 3XL | Shipped | 1 |
| **2** | 404-0687676-7273146 | 04-30-22 | Shipped | Amazon | Amazon.in | Expedited | Shirt | XL | Shipped | 1 |
| **4** | 407-1069790-7240320 | 04-30-22 | Shipped | Amazon | Amazon.in | Expedited | Trousers | 3XL | Shipped | 1 |
| **5** | 404-1490984-4578765 | 04-30-22 | Shipped | Amazon | Amazon.in | Expedited | T-shirt | XL | Shipped | 1 |
| **6** | 408-5748499-6859555 | 04-30-22 | Shipped | Amazon | Amazon.in | Expedited | T-shirt | L | Shipped | 1 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | .. |
| **128971** | 406-6001380-7673107 | 05-31-22 | Shipped | Amazon | Amazon.in | Expedited | Shirt | XL | Shipped | 1 |
| **128972** | 402-9551604-7544318 | 05-31-22 | Shipped | Amazon | Amazon.in | Expedited | T-shirt | M | Shipped | 1 |
| **128973** | 407-9547469-3152358 | 05-31-22 | Shipped | Amazon | Amazon.in | Expedited | Blazzer | XXL | Shipped | 1 |
| **128974** | 402-6184140-0545956 | 05-31-22 | Shipped | Amazon | Amazon.in | Expedited | T-shirt | XS | Shipped | 1 |
| **128975** | 408-7436540-8728312 | 05-31-22 | Shipped | Amazon | Amazon.in | Expedited | T-shirt | S | Shipped | 1 |

115864 rows × 17 columns

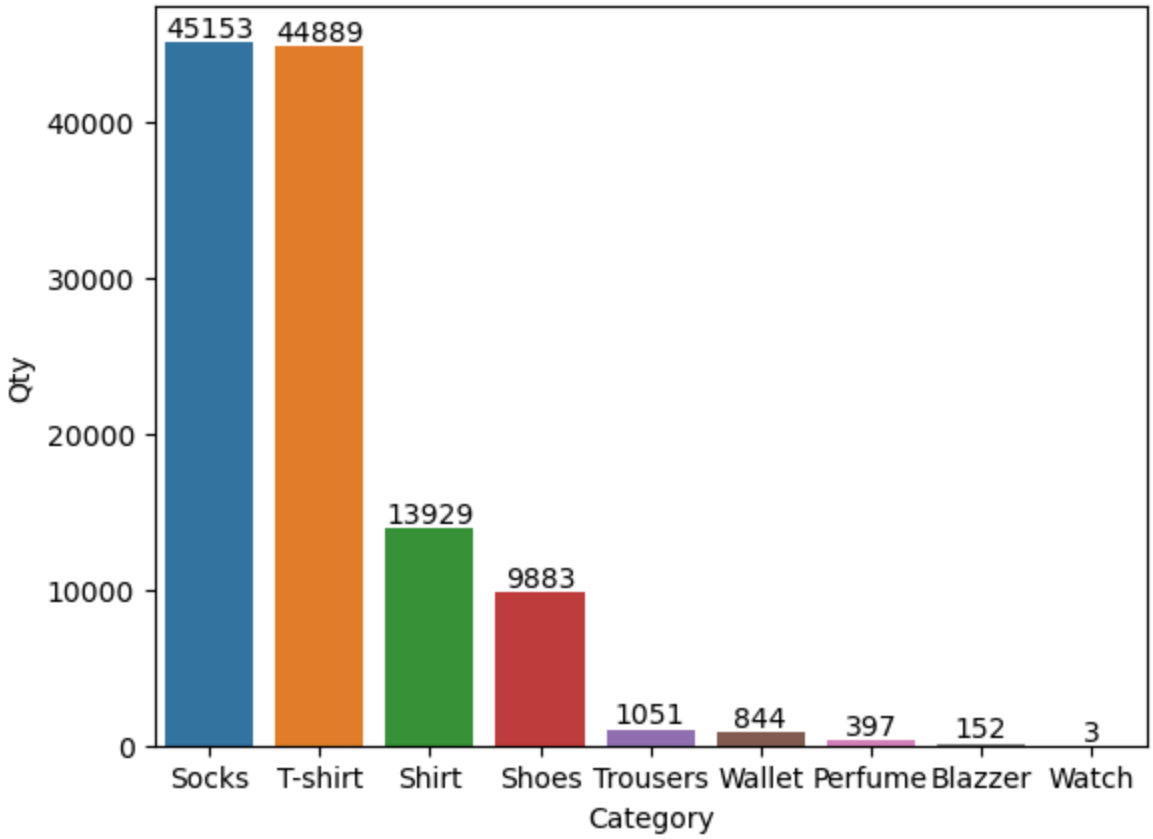
In [48]:
```
1 data.isna().sum()
```

Out[48]:
```
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
dtype: int64
```

In [49]:
```
1 data.columns
```

Out[49]:
```
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'],
      dtype='object')
```

In [57]:
```
Qty = data.groupby('Category')['Qty'].sum().tolist()
category = data.Category.unique().tolist()
df = pd.DataFrame({"Category" : category, "Qty" : Qty})
df.sort_values(by = 'Qty', ascending=False, inplace = True)
ax = sns.barplot(x = df['Category'], y = df['Qty'])
for i in ax.containers:
    ax.bar_label(i)
```

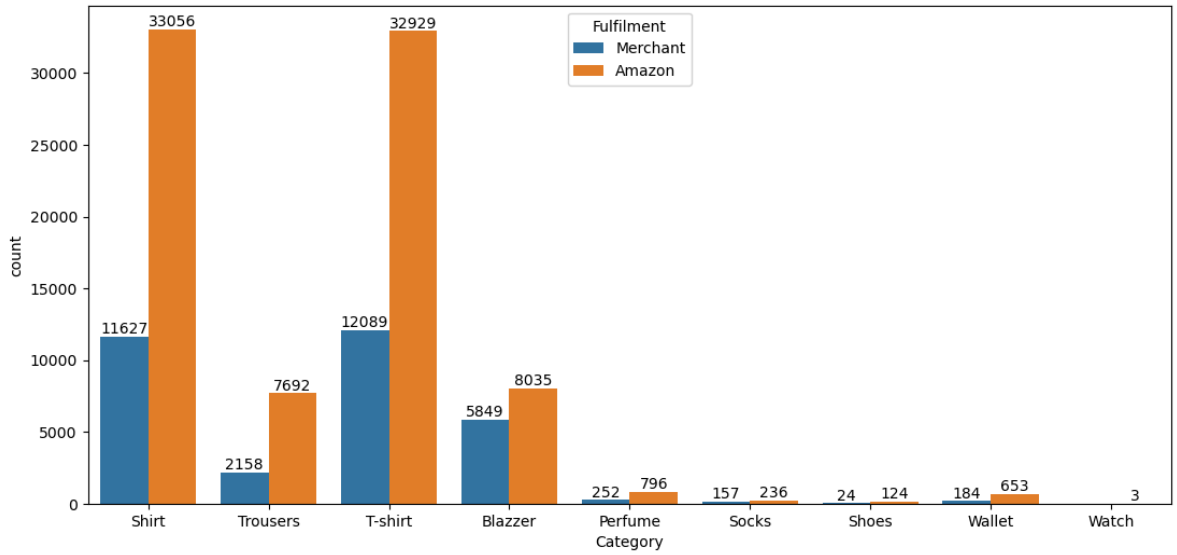



In the sales analysis, we can sayy that most selled item is socks and least sold item is watch

In [58]:
```
1 data.groupby('Category')['Fulfilment'].value_counts()
```

Out[58]:
```
Category  Fulfilment
Blazzer   Amazon         8035
          Merchant       5849
Perfume   Amazon          796
          Merchant        252
Shirt     Amazon        33056
          Merchant      11627
Shoes     Amazon          124
          Merchant         24
Socks     Amazon          236
          Merchant        157
T-shirt   Amazon        32929
          Merchant      12089
Trousers  Amazon         7692
          Merchant       2158
Wallet    Amazon          653
          Merchant        184
Watch     Amazon            3
Name: Fulfilment, dtype: int64
```

In [60]:
```
1 plt.figure(figsize = (13, 6))
2 ax = sns.countplot(data = data, x = 'Category', hue = 'Fulfilment')
3 for i in ax.containers:
4     ax.bar_label(i)
```



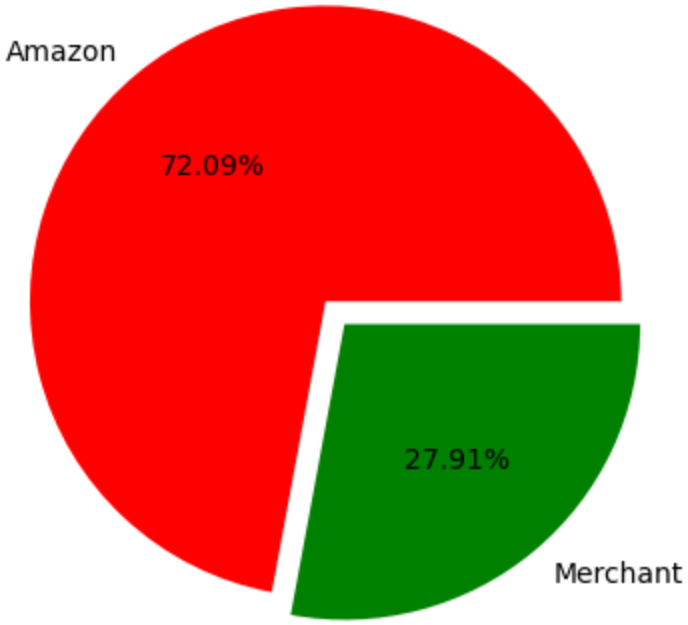

Amazon itself holds most of the stocks of these products as in every product number of orders fulfilled by amazon is greater than that of merchants

In [61]:
```
1 count = data1.Fulfilment.value_counts().tolist()
```

In [62]:
```
efficiency_amazon = count[0] / np.sum(count) * 100
print(f"efficiency of amazon : {efficiency_amazon}")
print(f"efficiency of merchant : {100 - efficiency_amazon}")
```

```
efficiency of amazon : 72.08796520058
efficiency of merchant : 27.91203479942
```

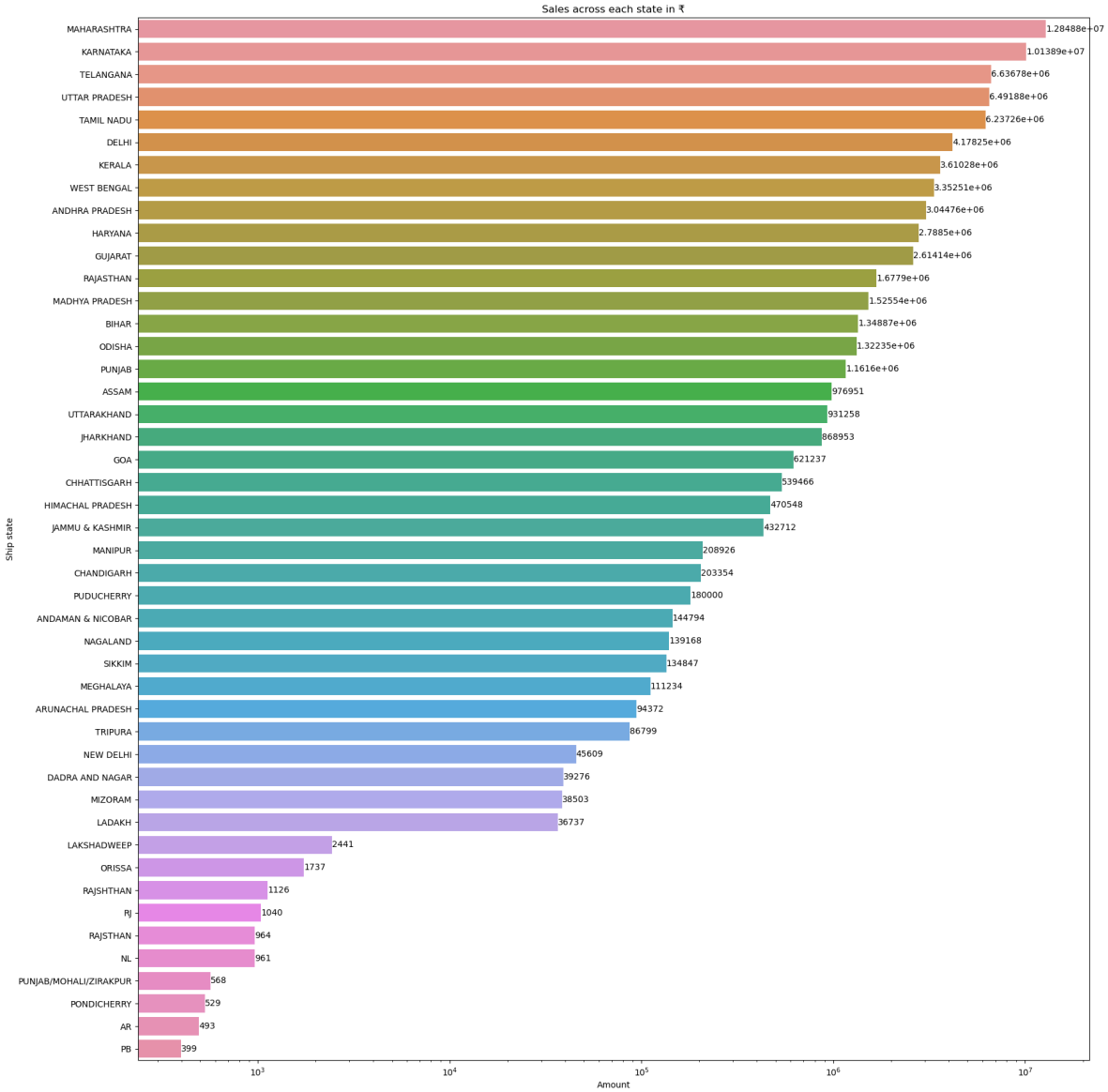
In [65]:
```
plt.pie(
    x = [efficiency_amazon, 100 - efficiency_amazon],
    labels = ['Amazon', 'Merchant'],
    autopct = "%1.2f%%",
    explode = [0, 0.1],
    colors = ['red', 'green']
)
plt.show()
```



In [66]:
```
#Turnover collected from each state
plt.figure(figsize = (20, 22))
df2 = pd.DataFrame(data.groupby('ship-state')['Amount'].sum())
df2 = df2.sort_values(by = 'Amount', ascending = False)
axs = sns.barplot(y = df2.index.tolist(), x = 'Amount', data = df2, log =
for i in axs.containers:
    axs.bar_label(i)

plt.ylabel('Ship state')
plt.title('Sales across each state in \u20B9')
```

Out[66]: Text(0.5, 1.0, 'Sales across each state in ₹')




Largest number of orders are placed in Maharashtra and so is the largest sum of amount.

In [68]:

```
state = data['ship-state'].unique().tolist()[1]
df = pd.DataFrame(data[data['ship-state'] == state].groupby('Size')['Cate
df
```

Out[68]:

| | | Category |
|---|---|---|
| **Size** | **Category** | |
| **3XL** | **Shirt** | 959 |
| | **T-shirt** | 937 |
| | **Blazzer** | 429 |
| | **Trousers** | 280 |
| | **Socks** | 12 |
| **4XL** | **Shirt** | 73 |
| | **T-shirt** | 15 |
| **5XL** | **Shirt** | 85 |
| | **T-shirt** | 17 |
| **6XL** | **Shirt** | 109 |
| | **T-shirt** | 9 |
| **Free** | **Perfume** | 223 |
| | **Wallet** | 109 |
| | **Shoes** | 17 |
| **L** | **Shirt** | 1283 |
| | **T-shirt** | 1257 |
| | **Blazzer** | 473 |
| | **Trousers** | 350 |
| | **Socks** | 9 |
| **M** | **T-shirt** | 1280 |
| | **Shirt** | 1259 |
| | **Blazzer** | 357 |
| | **Trousers** | 352 |
| | **Socks** | 4 |
| **S** | **T-shirt** | 1040 |
| | **Shirt** | 816 |
| | **Blazzer** | 271 |
| | **Trousers** | 266 |
| | **Socks** | 8 |
| **XL** | **Shirt** | 1308 |
| | **T-shirt** | 1203 |
| | **Blazzer** | 428 |
| | **Trousers** | 403 |
| | **Socks** | 13 |

| | | Category |
|---|---|---|
| **Size** | **Category** | |
| **XS** | **T-shirt** | 745 |
| | **Shirt** | 464 |
| | **Trousers** | 208 |
| | **Blazzer** | 185 |
| | **Socks** | 6 |
| **XXL** | **Shirt** | 1242 |
| | **T-shirt** | 928 |
| | **Trousers** | 412 |
| | **Blazzer** | 399 |
| | **Socks** | 6 |