In [1]:
```python
##importing all libaries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
import scipy as sm
from bs4 import BeautifulSoup
from datetime import datetime
```

In [2]:
```python
##Import data into Python environment.
data=pd.read_csv("Comcast_telecom_complaints_data.csv")
```

In [3]:
```python
data.head(5)
```

Out[3]:

	Ticket #	Customer Complaint	Date	Date_month_year	Time	Received Via	City	State	Zip code	Status	Filing on Behalf of Someone
0	250635	Comcast Cable Internet Speeds	22-04-15	22-Apr-15	3:53:50 PM	Customer Care Call	Abingdon	Maryland	21009	Closed	No
1	223441	Payment disappear - service got disconnected	04-08-15	04-Aug-15	10:22:56 AM	Internet	Acworth	Georgia	30102	Closed	No
2	242732	Speed and Service	18-04-15	18-Apr-15	9:55:47 AM	Internet	Acworth	Georgia	30101	Closed	Yes
3	277946	Comcast Imposed a New Usage Cap of 300GB that ...	05-07-15	05-Jul-15	11:59:35 AM	Internet	Acworth	Georgia	30101	Open	Yes
4	307175	Comcast not working and no service to boot	26-05-15	26-May-15	1:25:26 PM	Internet	Acworth	Georgia	30101	Solved	No

In [4]:
```python
data.shape
```

Out[4]:
```
(2224, 11)
```

In [5]:
```python
data.info
```

Out[5]:
```
<bound method DataFrame.info of      Ticket #                                 Customer Complaint      Date  \
0      250635                      Comcast Cable Internet Speeds  22-04-15
1      223441       Payment disappear - service got disconnected  04-08-15
2      242732                                  Speed and Service  18-04-15
3      277946  Comcast Imposed a New Usage Cap of 300GB that ...  05-07-15
4      307175         Comcast not working and no service to boot  26-05-15
...       ...                                                ...       ...
2219   213550                               Service Availability  04-02-15
2220   318775         Comcast Monthly Billing for Returned Modem  06-02-15
2221   331188                           complaint about comcast  06-09-15
2222   360489              Extremely unsatisfied Comcast customer  23-06-15
2223   363614           Comcast, Ypsilanti MI Internet Speed  24-06-15

     Date_month_year         Time        Received Via        City     State  \
0          22-Apr-15   3:53:50 PM  Customer Care Call    Abingdon  Maryland
1          04-Aug-15  10:22:56 AM            Internet     Acworth   Georgia
2          18-Apr-15   9:55:47 AM            Internet     Acworth   Georgia
3          05-Jul-15  11:59:35 AM            Internet     Acworth   Georgia
4          26-May-15   1:25:26 PM            Internet     Acworth   Georgia
...              ...          ...                 ...         ...       ...
2219       04-Feb-15   9:13:18 AM  Customer Care Call  Youngstown   Florida
2220       06-Feb-15   1:24:39 PM  Customer Care Call   Ypsilanti  Michigan
2221       06-Sep-15   5:28:41 PM            Internet   Ypsilanti  Michigan
2222       23-Jun-15  11:13:30 PM  Customer Care Call   Ypsilanti  Michigan
2223       24-Jun-15  10:28:33 PM  Customer Care Call   Ypsilanti  Michigan

      Zip code  Status Filing on Behalf of Someone
0        21009  Closed                          No
1        30102  Closed                          No
2        30101  Closed                         Yes
3        30101    Open                         Yes
4        30101  Solved                          No
...        ...     ...                         ...
2219     32466  Closed                          No
2220     48197  Solved                          No
2221     48197  Solved                          No
2222     48197  Solved                          No
2223     48198    Open                         Yes

[2224 rows x 11 columns]>
```

In [6]:
```python
type(data)
```

Out[6]:
```
pandas.core.frame.DataFrame
```

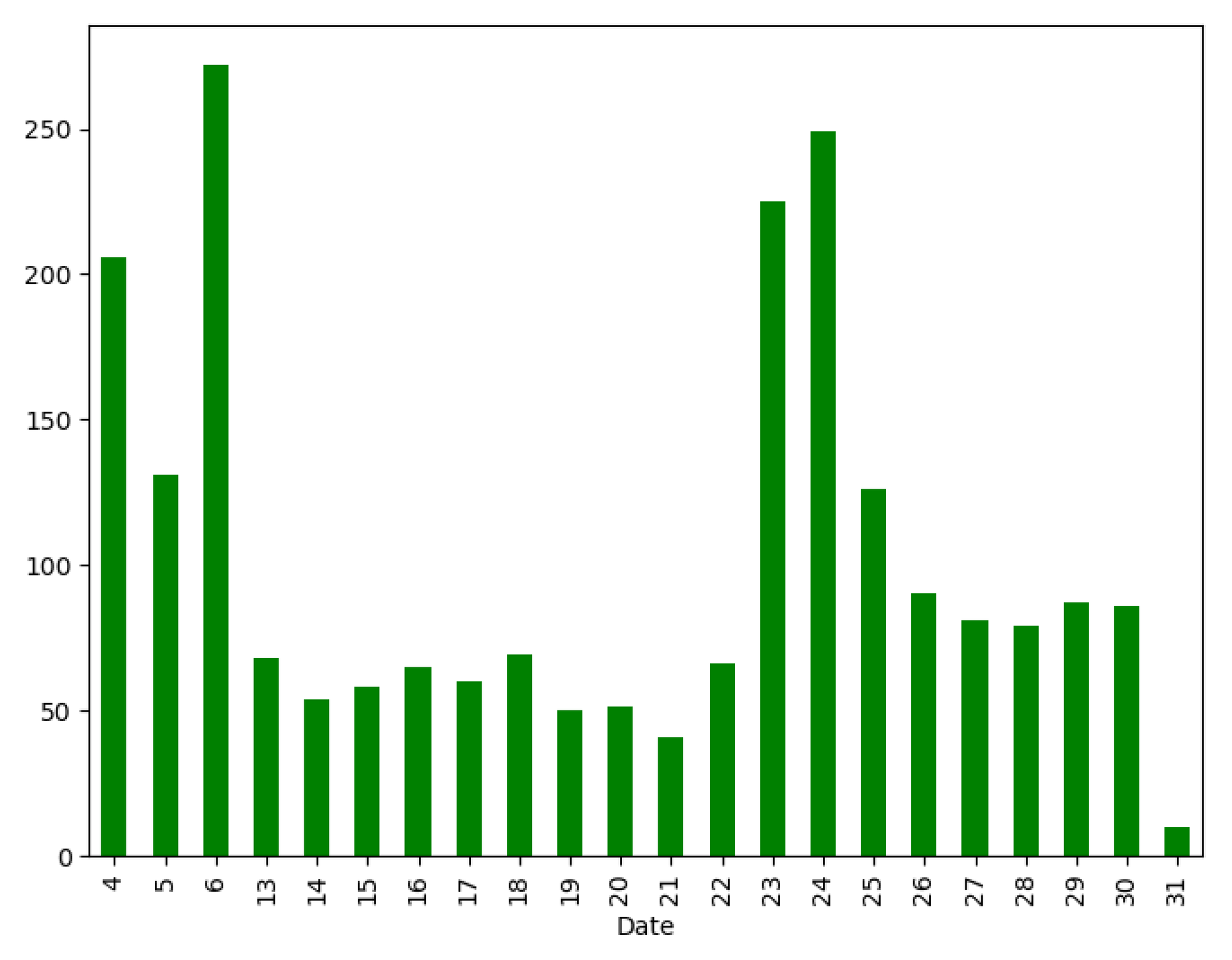
In [7]:
```python
print(data.describe())
```

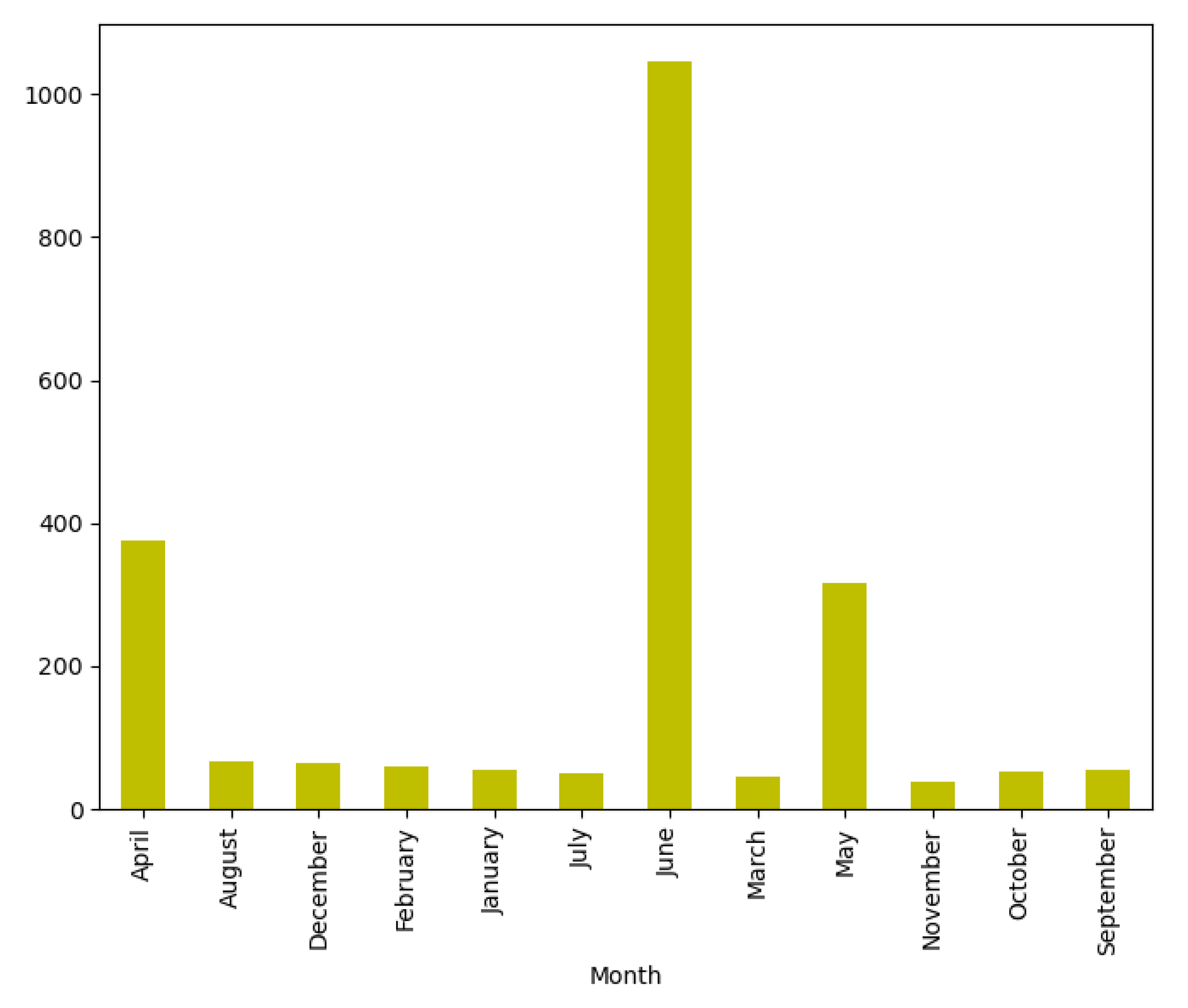
```
           Zip code
count   2224.000000
mean   47994.393435
std    28885.279427
min     1075.000000
25%    30056.500000
50%    37211.000000
75%    77058.750000
max    99223.000000
```

In [8]:
```python
data.isnull().sum(axis=0)
```

Out[8]:
```
Ticket #                       0
Customer Complaint             0
Date                           0
Date_month_year                0
Time                           0
Received Via                   0
City                           0
State                          0
Zip code                       0
Status                         0
Filing on Behalf of Someone    0
dtype: int64
```

In [9]:
```python
## trend chart for the number of complaints at monthly and daily granularity levels.
data['Month']=pd.to_datetime(data['Date_month_year']).dt.month_name()
data['Date']=pd.to_datetime(data['Date_month_year']).dt.day
##Graph for date wise
data.groupby(['Date'])['Customer Complaint'].count().plot(kind='bar',color='g',figsize=(8,6))
plt.show()
##Graph for date wise
data.groupby(['Month'])['Customer Complaint'].count().plot(kind='bar',color='y',figsize=(8,6))
plt.show()
```





In [34]:
```python
#  A table with the frequency of complaint types.
data['Customer Complaint'].value_counts().to_frame().reset_index()
```

Out[34]:

	index	Customer Complaint
0	Comcast	83
1	Comcast Internet	18
2	Comcast Data Cap	17
3	comcast	13
4	Comcast Billing	11
...	...	...
1836	Improper Billing and non resolution of issues	1
1837	Deceptive trade	1
1838	intermittent internet	1
1839	Internet Speed on Wireless Connection	1
1840	Comcast, Ypsilanti MI Internet Speed	1

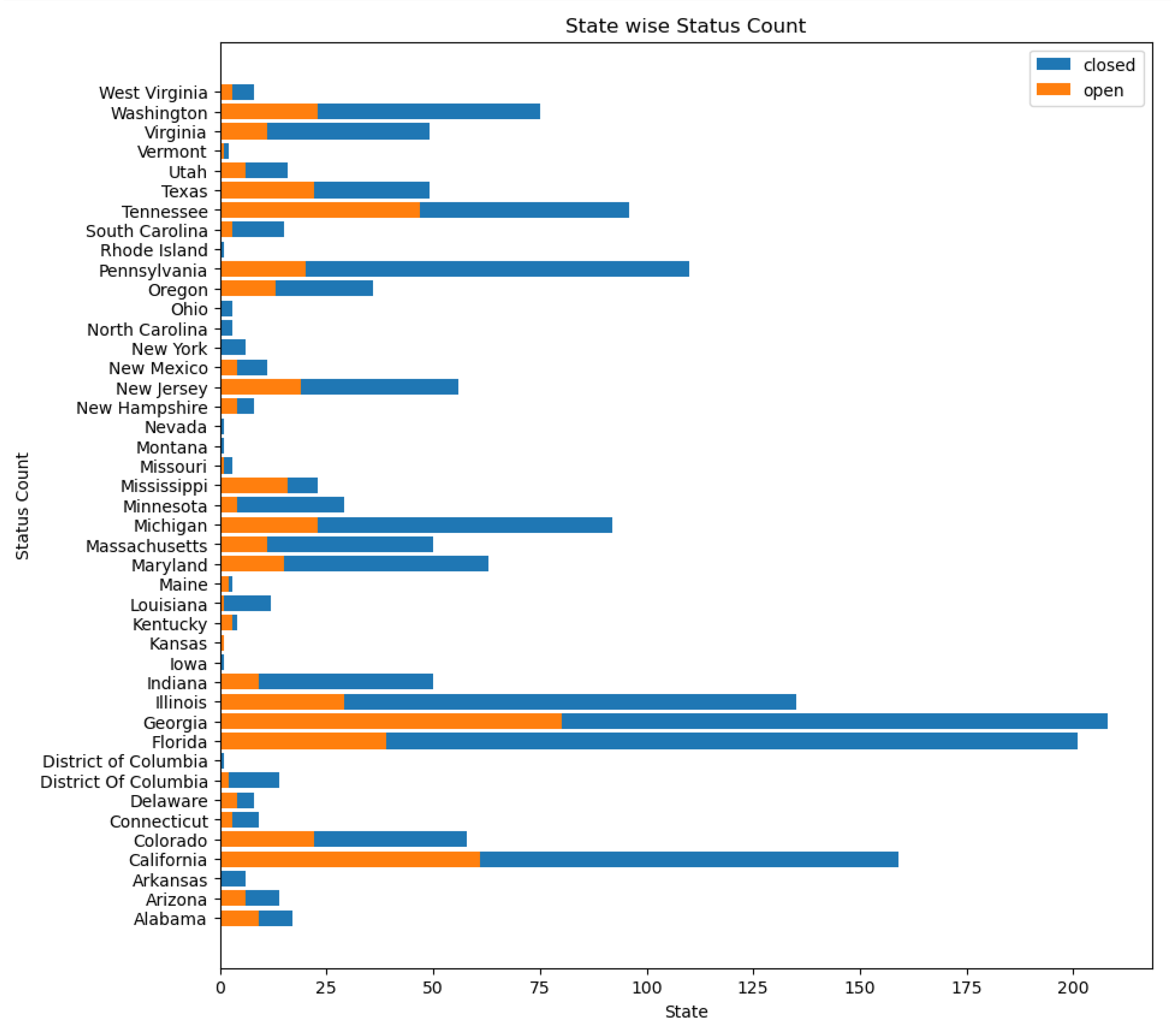
1841 rows × 2 columns

In [35]:
```python
## Which complaint types are maximum i.e., around internet, network issues, or across any other domains.
data['Customer Complaint'].value_counts()
```

Out[35]:
```
Comcast                                          83
Comcast Internet                                 18
Comcast Data Cap                                 17
comcast                                          13
Comcast Billing                                  11
                                                 ..
Improper Billing and non resolution of issues     1
Deceptive trade                                   1
intermittent internet                             1
Internet Speed on Wireless Connection             1
Comcast, Ypsilanti MI Internet Speed              1
Name: Customer Complaint, Length: 1841, dtype: int64
```

In [12]:
```python
## Create a new categorical variable with value as Open and Closed. Open & Pending is to be categorized as Open and Closed & Solved is to be categorized as Closed.
data['Status']=data['Status'].apply(lambda x:'Open'
                                    if ((x=='Open')|(x=='Pending'))
                                     else'Closed')
```

In [25]:
```python
## Provide state wise status of complaints in a stacked bar chart.
op=data[data['Status']=='Open'].groupby(['State'])['Status'].count().to_frame().reset_index()
close=data[data['Status']=='Closed'].groupby(['State'])['Status'].count().to_frame().reset_index()
fig=plt.figure(figsize=(10,10))
plt.barh(close.State,close.Status)
plt.barh(op.State,op.Status)
plt.xlabel("State")
plt.ylabel("Status Count")
plt.legend(["closed","open"])
plt.title("State wise Status Count")
plt.show()
```



In [29]:
```python
## maximun complaints
data.groupby('State')['Customer Complaint'].agg("count").sort_values(ascending=False)
```

Out[29]:
```
State
Georgia                 288
Florida                 240
California              220
Illinois                164
Tennessee               143
Pennsylvania            130
Michigan                115
Washington               98
Colorado                 80
Maryland                 78
New Jersey               75
Texas                    71
Massachusetts            61
Virginia                 60
Indiana                  59
Oregon                   49
Mississippi              39
Minnesota                33
Alabama                  26
Utah                     22
Arizona                  20
South Carolina           18
District Of Columbia     16
New Mexico               15
Louisiana                13
Connecticut              12
New Hampshire            12
Delaware                 12
West Virginia            11
Kentucky                  7
Arkansas                  6
New York                  6
Maine                     5
Missouri                  4
North Carolina            3
Vermont                   3
Ohio                      3
Kansas                    2
District of Columbia      1
Rhode Island              1
Iowa                      1
Nevada                    1
Montana                   1
Name: Customer Complaint, dtype: int64
```

In [28]:
```python
## state in which the highest percentage of unresolved complaints.
state_unsolved=data.loc[data['Status']=='Open',['State']].value_counts()
state_unsolved/state_unsolved.sum()*100
```

Out[28]:
```
State
Georgia                 15.473888
California              11.798839
Tennessee                9.090909
Florida                  7.543520
Illinois                 5.609284
Michigan                 4.448743
Washington               4.448743
Colorado                 4.255319
Texas                    4.255319
Pennsylvania             3.868472
New Jersey               3.675048
Mississippi              3.094778
Maryland                 2.901354
Oregon                   2.514507
Massachusetts            2.127660
Virginia                 2.127660
Alabama                  1.740812
Indiana                  1.740812
Arizona                  1.160542
Utah                     1.160542
Delaware                 0.773694
New Hampshire            0.773694
New Mexico               0.773694
Minnesota                0.773694
South Carolina           0.580271
Connecticut              0.580271
West Virginia            0.580271
Kentucky                 0.580271
District Of Columbia     0.386847
Maine                    0.386847
Louisiana                0.193424
Vermont                  0.193424
Missouri                 0.193424
Kansas                   0.193424
dtype: float64
```

In [ ]:
```python
## The percentage of complaints resolved till date, which were received through the Internet and customer care calls.
data[data['Status']=='Closed'].groupby('Status')['Received Via'].value_counts(normalize=True)*100
```

Out[ ]:
```
Status  Received Via
Closed  Customer Care Call    50.615114
        Internet              49.384886
Name: Received Via, dtype: float64
```

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]: