* Project [Data Science With Python]: Comcast Telecom Consumer Complaints

1. Import Libraries

In [40]:
```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
```

2. Loading the Datasets

In [41]:
```python
df = pd.read_csv("C:/Users/lenovo/OneDrive/डेस्कटॉप/Comcast_telecom_complaints_data.csv")
```

In [42]:
```python
df.head()
```

Out[42]:

	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	22-04-2015	22-Apr-15	3:53:50 PM	Customer Care Call	Abingdon	Maryland	21009	Closed	No
1	223441	Payment disappear - service got disconnected	04-08-2015	04-Aug-15	10:22:56 AM	Internet	Acworth	Georgia	30102	Closed	No
2	242732	Speed and Service	18-04-2015	18-Apr-15	9:55:47 AM	Internet	Acworth	Georgia	30101	Closed	Yes
3	277946	Comcast Imposed a New Usage Cap of 300GB that ...	05-07-2015	05-Jul-15	11:59:35 AM	Internet	Acworth	Georgia	30101	Open	Yes
4	307175	Comcast not working and no service to boot	26-05-2015	26-May-15	1:25:26 PM	Internet	Acworth	Georgia	30101	Solved	No

In [43]:
```python
df.tail()
```

Out[43]:

	Ticket #	Customer Complaint	Date	Date_month_year	Time	Received Via	City	State	Zip code	Status	Filing on Behalf of Someone
2219	213550	Service Availability	04-02-2015	04-Feb-15	9:13:18 AM	Customer Care Call	Youngstown	Florida	32466	Closed	No
2220	318775	Comcast Monthly Billing for Returned Modem	06-02-2015	06-Feb-15	1:24:39 PM	Customer Care Call	Ypsilanti	Michigan	48197	Solved	No
2221	331188	complaint about comcast	06-09-2015	06-Sep-15	5:28:41 PM	Internet	Ypsilanti	Michigan	48197	Solved	No
2222	360489	Extremely unsatisfied Comcast customer	23-06-2015	23-Jun-15	11:13:30 PM	Customer Care Call	Ypsilanti	Michigan	48197	Solved	No
2223	363614	Comcast, Ypsilanti MI Internet Speed	24-06-2015	24-Jun-15	10:28:33 PM	Customer Care Call	Ypsilanti	Michigan	48198	Open	Yes

In [44]:
```python
df.shape
```

Out[44]:
```
(2224, 11)
```

In [45]:
```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2224 entries, 0 to 2223
Data columns (total 11 columns):
 #   Column                       Non-Null Count  Dtype
---  ------                       --------------  -----
 0   Ticket #                     2224 non-null   object
 1   Customer Complaint           2224 non-null   object
 2   Date                         2224 non-null   object
 3   Date_month_year              2224 non-null   object
 4   Time                         2224 non-null   object
 5   Received Via                 2224 non-null   object
 6   City                         2224 non-null   object
 7   State                        2224 non-null   object
 8   Zip code                     2224 non-null   int64
 9   Status                       2224 non-null   object
 10  Filing on Behalf of Someone  2224 non-null   object
dtypes: int64(1), object(10)
memory usage: 191.2+ KB
```

In [46]:
```python
df.describe()
```

Out[46]:

	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 [47]:
```python
df.isnull().sum()
```

Out[47]:
```
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
```

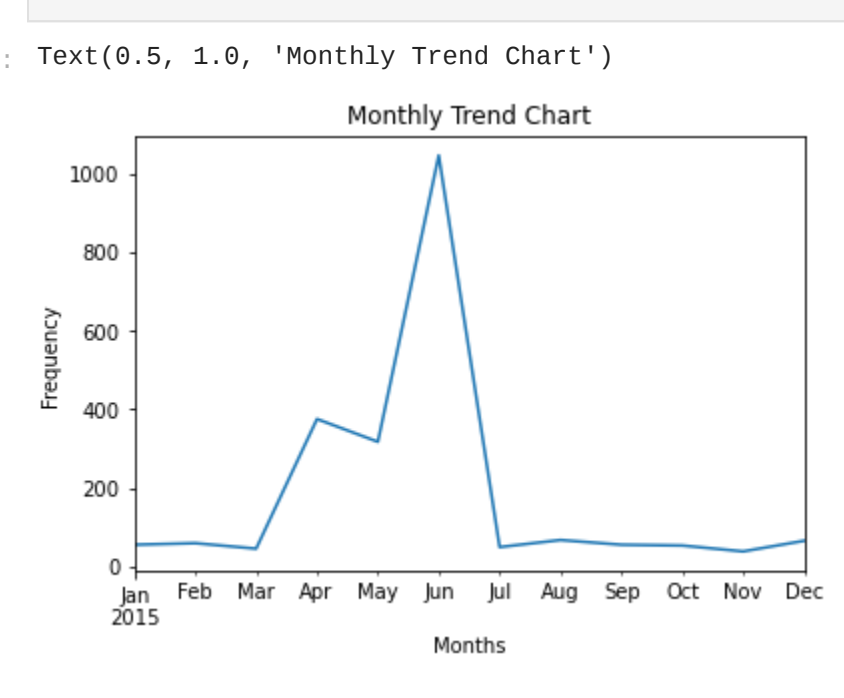
3. Provide the trend chart for the number of complaints at monthly and daily granularity levels.

Monthly Trend Chart

In [51]:
```python
df['Date_month_year']=df['Date_month_year'].apply(pd.to_datetime)
df=df.set_index('Date_month_year')
```

In [52]:
```python
months=df.groupby(pd.Grouper(freq="M")).size().plot()
plt.xlabel("Months")
plt.ylabel("Frequency")
plt.title("Monthly Trend Chart")
```

Out[52]:
```
Text(0.5, 1.0, 'Monthly Trend Chart')
```



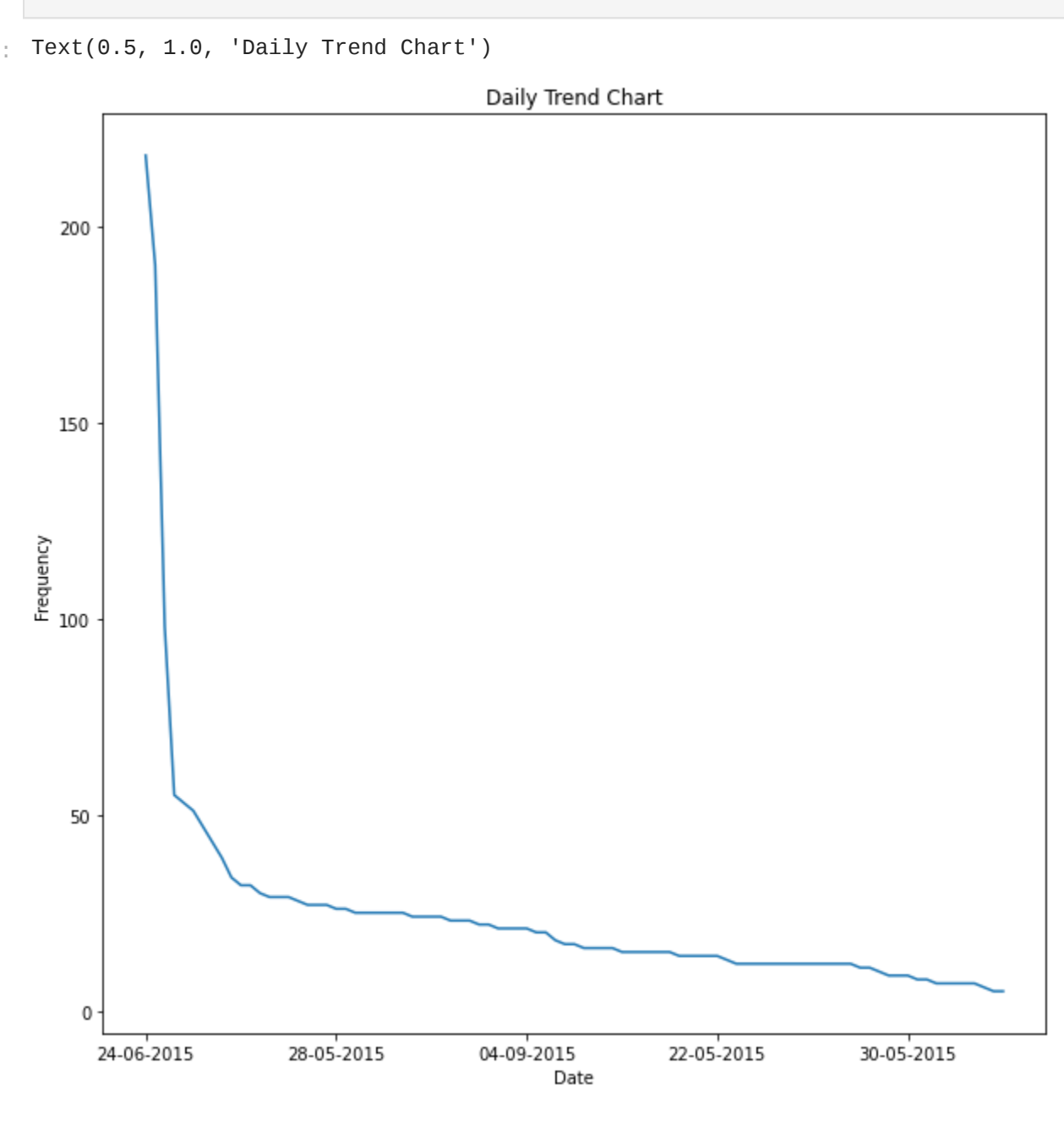
In [13]:
```python
df['Date'].value_counts()
```

Out[13]:
```
24-06-2015    218
23-06-2015    190
25-06-2015     98
26-06-2015     55
30-06-2015     53
             ...
05-12-2015      7
05-10-2015      7
04-05-2015      6
05-03-2015      5
04-11-2015      5
Name: Date, Length: 91, dtype: int64
```

Plotting daily trend chart

In [14]:
```python
df=df.sort_values(by='Date')
plt.figure(figsize=(10,10))
df['Date'].value_counts().plot()
plt.xlabel("Date")
plt.ylabel("Frequency")
plt.title("Daily Trend Chart")
```

Out[14]:
```
Text(0.5, 1.0, 'Daily Trend Chart')
```



4. Providing Tables with the frequency of complaint types

In [15]:
```python
df['Customer Complaint'].value_counts()
```

Out[15]:
```
Comcast                                                                              83
Comcast Internet                                                                     18
Comcast Data Cap                                                                     17
comcast                                                                              13
Comcast Billing                                                                      11
                                                                                     ..
Comcast started charging a lease fee for equipment I own after enabling automatic bill payment    1
Comcast billing error                                                                 1
Comcast Internet Services                                                             1
internet service from comcast                                                         1
Comcast charges                                                                       1
Name: Customer Complaint, Length: 1841, dtype: int64
```

5. Which complaint types are maximum i.e., around internet, network issues, or across any other domains

In [17]:
```python
internet_issue_1=df[df['Customer Complaint'].str.contains("network")].count()
```

In [18]:
```python
internet_issue_2=df[df['Customer Complaint'].str.contains("speed")].count()
```

In [19]:
```python
internet_issue_3=df[df['Customer Complaint'].str.contains("data")].count()
```

In [20]:
```python
internet_issue_4=df[df['Customer Complaint'].str.contains("internet")].count()
```

In [21]:
```python
billing_issue_1=df[df['Customer Complaint'].str.contains("bill")].count()
```

In [22]:
```python
billing_issue_2=df[df['Customer Complaint'].str.contains("billing")].count()
```

In [23]:
```python
billing_issue_3=df[df['Customer Complaint'].str.contains("charges")].count()
```

In [24]:
```python
service_issue_1=df[df['Customer Complaint'].str.contains("network")].count()
```

In [25]:
```python
service_issue_2=df[df['Customer Complaint'].str.contains("service")].count()
```

In [26]:
```python
Total_internet_issues = internet_issue_1 + internet_issue_2 + internet_issue_3 + internet_issue_4
```

In [27]:
```python
print(Total_internet_issues)
```

```
Ticket #                       374
Customer Complaint             374
Date                           374
Date_month_year                374
Time                           374
Received Via                   374
City                           374
State                          374
Zip code                       374
Status                         374
Filing on Behalf of Someone    374
dtype: int64
```

In [28]:
```python
total_billing_issues = billing_issue_1 + billing_issue_2 + billing_issue_3
```

In [32]:
```python
print(total_billing_issues)
```

```
Ticket #                       353
Customer Complaint             353
Date                           353
Date_month_year                353
Time                           353
Received Via                   353
City                           353
State                          353
Zip code                       353
Status                         353
Filing on Behalf of Someone    353
dtype: int64
```

In [34]:
```python
total_service_issues = service_issue_1 + service_issue_2
```

In [35]:
```python
print(total_service_issues)
```

```
Ticket #                       311
Customer Complaint             311
Date                           311
Date_month_year                311
Time                           311
Received Via                   311
City                           311
State                          311
Zip code                       311
Status                         311
Filing on Behalf of Someone    311
dtype: int64
```

In [55]:
```python
other_issues=2224-(Total_internet_issues + total_billing_issues + total_service_issues)
print(other_issues)
```

```
Ticket #                       1186
Customer Complaint             1186
Date                           1186
Date_month_year                1186
Time                           1186
Received Via                   1186
City                           1186
State                          1186
Zip code                       1186
Status                         1186
Filing on Behalf of Someone    1186
dtype: int64
```

6. Create a new categorical with value open & closed. Open & Pending is to be refer as Open '&' Closed & Solved is to be refer as closed

In [57]:
```python
df.Status.unique()
```

Out[57]:
```
array(['Closed', 'Open', 'Solved', 'Pending'], dtype=object)
```

In [62]:
```python
df["New_Status"]=["Open" if Status=="Open" or Status=="Pending" else "Closed" for Status in df["Status"]]
df=df.drop(['Status'],axis=1)
df
```

Out[62]:

	Ticket #	Customer Complaint	Date	Time	Received Via	City	State	Zip code	Filing on Behalf of Someone	New_Status
Date_month_year										
2015-04-22	250635	Comcast Cable Internet Speeds	22-04-2015	3:53:50 PM	Customer Care Call	Abingdon	Maryland	21009	No	Closed
2015-08-04	223441	Payment disappear - service got disconnected	04-08-2015	10:22:56 AM	Internet	Acworth	Georgia	30102	No	Closed
2015-04-18	242732	Speed and Service	18-04-2015	9:55:47 AM	Internet	Acworth	Georgia	30101	Yes	Closed
2015-07-05	277946	Comcast Imposed a New Usage Cap of 300GB that ...	05-07-2015	11:59:35 AM	Internet	Acworth	Georgia	30101	Yes	Open
2015-05-26	307175	Comcast not working and no service to boot	26-05-2015	1:25:26 PM	Internet	Acworth	Georgia	30101	No	Closed
...	...	...	...	...	...	...	...	...	...	...
2015-02-04	213550	Service Availability	04-02-2015	9:13:18 AM	Customer Care Call	Youngstown	Florida	32466	No	Closed
2015-02-06	318775	Comcast Monthly Billing for Returned Modem	06-02-2015	1:24:39 PM	Customer Care Call	Ypsilanti	Michigan	48197	No	Closed
2015-09-06	331188	complaint about comcast	06-09-2015	5:28:41 PM	Internet	Ypsilanti	Michigan	48197	No	Closed
2015-06-23	360489	Extremely unsatisfied Comcast customer	23-06-2015	11:13:30 PM	Customer Care Call	Ypsilanti	Michigan	48197	No	Closed
2015-06-24	363614	Comcast, Ypsilanti MI Internet Speed	24-06-2015	10:28:33 PM	Customer Care Call	Ypsilanti	Michigan	48198	Yes	Open

2224 rows × 10 columns

7. Which state has Maximum Complaints?

In [66]:
```python
df.groupby(['State']).size().sort_values(ascending=False)
```

Out[66]:
```
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
dtype: int64
```

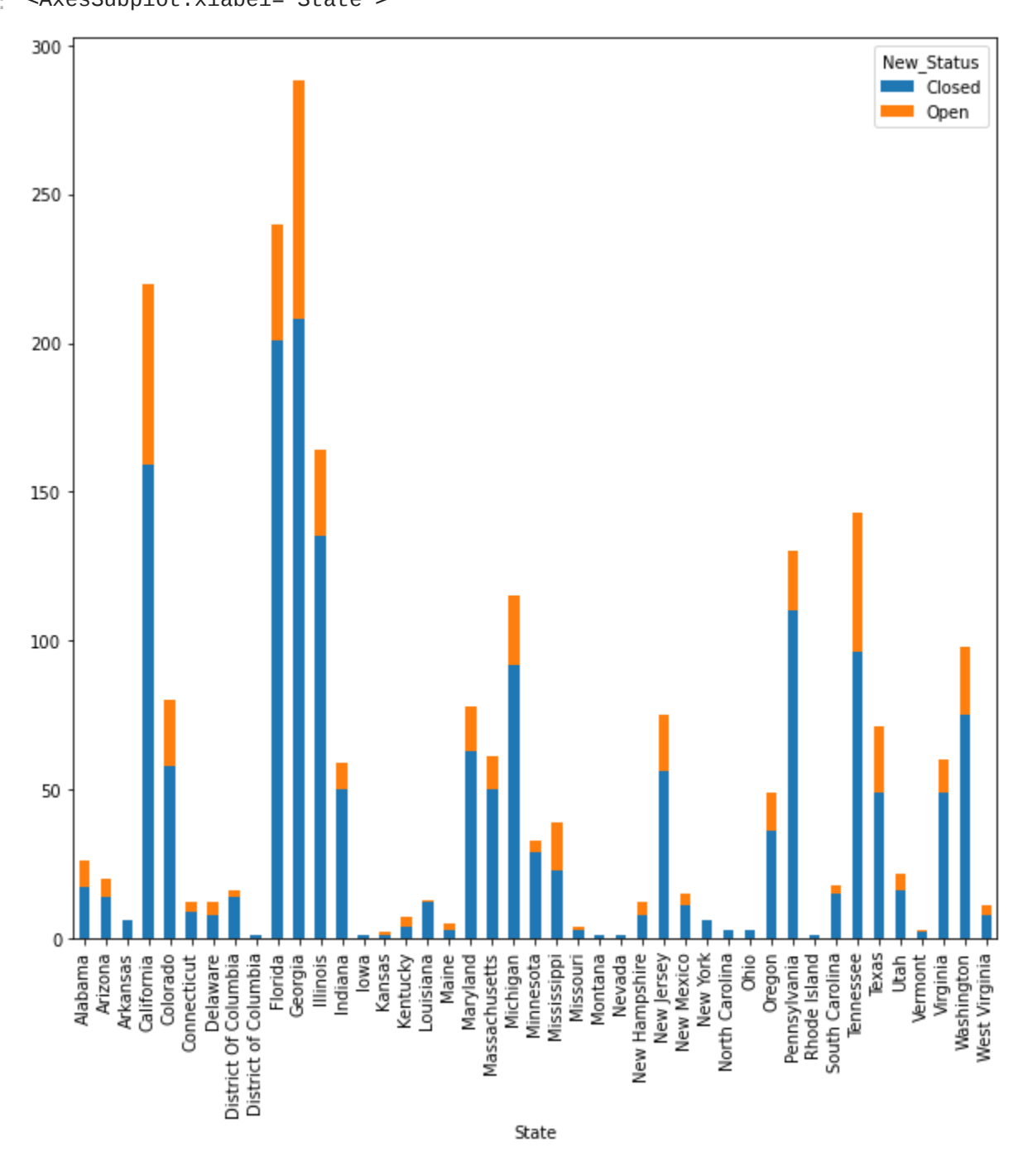
8. Provide state wise status of complaints in a stacked bar chart

In [80]:
```python
Status_Complaints=df.groupby(["State","New_Status"]).size().unstack()
print(Status_Complaints)
```

```
New_Status            Closed  Open
State
Alabama                 17.0   9.0
Arizona                 14.0   6.0
Arkansas                 6.0   NaN
California             159.0  61.0
Colorado                58.0  22.0
Connecticut              9.0   3.0
Delaware                 8.0   4.0
District Of Columbia    14.0   2.0
District of Columbia     1.0   NaN
Florida                201.0  39.0
Georgia                208.0  80.0
Illinois               135.0  29.0
Indiana                 50.0   9.0
Iowa                     1.0   NaN
Kansas                   1.0   1.0
Kentucky                 4.0   3.0
Louisiana               12.0   1.0
Maine                    3.0   2.0
Maryland                63.0  15.0
Massachusetts           50.0  11.0
Michigan                92.0  23.0
Minnesota               29.0   4.0
Mississippi             23.0  16.0
Missouri                 3.0   1.0
Montana                  1.0   NaN
Nevada                   1.0   NaN
New Hampshire            8.0   4.0
New Jersey              56.0  19.0
New Mexico              11.0   4.0
New York                 6.0   NaN
North Carolina           3.0   NaN
Ohio                     3.0   NaN
Oregon                  36.0  13.0
Pennsylvania           110.0  20.0
Rhode Island             1.0   NaN
South Carolina          15.0   3.0
Tennessee               96.0  47.0
Texas                   49.0  22.0
Utah                    16.0   6.0
Vermont                  2.0   1.0
Virginia                49.0  11.0
Washington              75.0  23.0
West Virginia            8.0   3.0
```

In [81]:
```python
Status_Complaints.plot.bar(figsize=(10,10),stacked=True)
```

Out[81]:
```
<AxesSubplot:xlabel='State'>
```



9. The State which has highest percentage of unresolved complaints

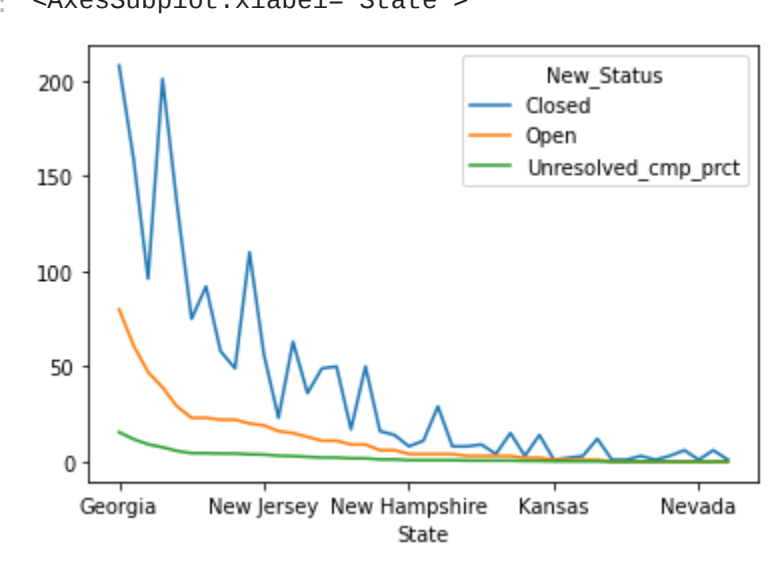
In [82]:
```python
print(df['New_Status'].value_counts())
```

```
Closed    1707
Open       517
Name: New_Status, dtype: int64
```

In [86]:
```python
unresolved_data=df.groupby(["State",'New_Status']).size().unstack().fillna(0).sort_values(by='Open',ascending=False)
unresolved_data['Unresolved_cmp_prct']=unresolved_data['Open']/unresolved_data['Open'].sum()*100
print(unresolved_data)
unresolved_data.plot()
```

```
New_Status            Closed  Open  Unresolved_cmp_prct
State
Georgia                208.0  80.0            15.473888
California             159.0  61.0            11.798839
Tennessee               96.0  47.0             9.090909
Florida                201.0  39.0             7.543520
Illinois               135.0  29.0             5.609284
Washington              75.0  23.0             4.448743
Michigan                92.0  23.0             4.448743
Colorado                58.0  22.0             4.255319
Texas                   49.0  22.0             4.255319
Pennsylvania           110.0  20.0             3.868472
New Jersey              56.0  19.0             3.675048
Mississippi             23.0  16.0             3.094778
Maryland                63.0  15.0             2.901354
Oregon                  36.0  13.0             2.514507
Virginia                49.0  11.0             2.127660
Massachusetts           50.0  11.0             2.127660
Alabama                 17.0   9.0             1.740812
Indiana                 50.0   9.0             1.740812
Utah                    16.0   6.0             1.160542
Arizona                 14.0   6.0             1.160542
New Hampshire            8.0   4.0             0.773694
New Mexico              11.0   4.0             0.773694
Minnesota               29.0   4.0             0.773694
Delaware                 8.0   4.0             0.773694
West Virginia            8.0   3.0             0.580271
Connecticut              9.0   3.0             0.580271
Kentucky                 4.0   3.0             0.580271
South Carolina          15.0   3.0             0.580271
Maine                    3.0   2.0             0.386847
District Of Columbia    14.0   2.0             0.386847
Kansas                   1.0   1.0             0.193424
Vermont                  2.0   1.0             0.193424
Missouri                 3.0   1.0             0.193424
Louisiana               12.0   1.0             0.193424
Montana                  1.0   0.0             0.000000
Rhode Island             1.0   0.0             0.000000
Ohio                     3.0   0.0             0.000000
District of Columbia     1.0   0.0             0.000000
North Carolina           3.0   0.0             0.000000
New York                 6.0   0.0             0.000000
Nevada                   1.0   0.0             0.000000
Arkansas                 6.0   0.0             0.000000
Iowa                     1.0   0.0             0.000000
```

Out[86]:
```
<AxesSubplot:xlabel='State'>
```



10. Provide the percentage of complaints resolved till date, which were received through the Internet and customer care calls.

In [91]:
```python
resolved_data=df.groupby(['Received Via','New_Status']).size().unstack().fillna(0)
resolved_data['resolved']= resolved_data['Closed']/resolved_data['Closed'].sum()*100
resolved_data['resolved']
```

Out[91]:
```
Received Via
Customer Care Call    50.615114
Internet              49.384886
Name: resolved, dtype: float64
```

In [96]:
```python
resolved_data.plot(kind="bar",figsize=(8,8))
```

Out[96]:
```
<AxesSubplot:xlabel='Received Via'>
```