In [3]:

```python
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [4]:

```python
df=pd.read_csv(r"C:\Users\abha mohan\Desktop\Comcast_telecom_complaints_data.csv")
```

In [5]:

```python
df.head()
```

Out[5]:

	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 [6]:

```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
```

NOW WE SHALL CONVERT DATE INTO DATETIME OBJECT

In [8]:

```python
df['Date']=pd.to_datetime(df['Date'])
```

In [9]:

```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   datetime64[ns]
 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: datetime64[ns](1), int64(1), object(9)
memory usage: 191.2+ KB
```

In [10]:

```python
df['Month']=df['Date'].dt.month_name()
```

In [11]:

```python
df.head()
```

Out[11]:

	Ticket #	Customer Complaint	Date	Date_month_year	Time	Received Via	City	State	Zip code	Status	Filing on Behalf of Someone	Month
0	250635	Comcast Cable Internet Speeds	2015-04-22	22-Apr-15	3:53:50 PM	Customer Care Call	Abingdon	Maryland	21009	Closed	No	April
1	223441	Payment disappear - service got disconnected	2015-04-08	04-Aug-15	10:22:56 AM	Internet	Acworth	Georgia	30102	Closed	No	April
2	242732	Speed and Service	2015-04-18	18-Apr-15	9:55:47 AM	Internet	Acworth	Georgia	30101	Closed	Yes	April
3	277946	Comcast Imposed a New Usage Cap of 300GB that ...	2015-05-07	05-Jul-15	11:59:35 AM	Internet	Acworth	Georgia	30101	Open	Yes	May
4	307175	Comcast not working and no service to boot	2015-05-26	26-May-15	1:25:26 PM	Internet	Acworth	Georgia	30101	Solved	No	May

TO UNDERSTAND COMPLAINTS ON DAILY BASIS

In [13]:

```python
daily=df['Date'].value_counts()
```

In [14]:

```python
daily=pd.DataFrame(daily).reset_index()
```

In [15]:

```python
daily.rename(columns={'index':'Date',
                      'Date':'Count'},inplace=True)
```

In [16]:

```python
daily
```

Out[16]:

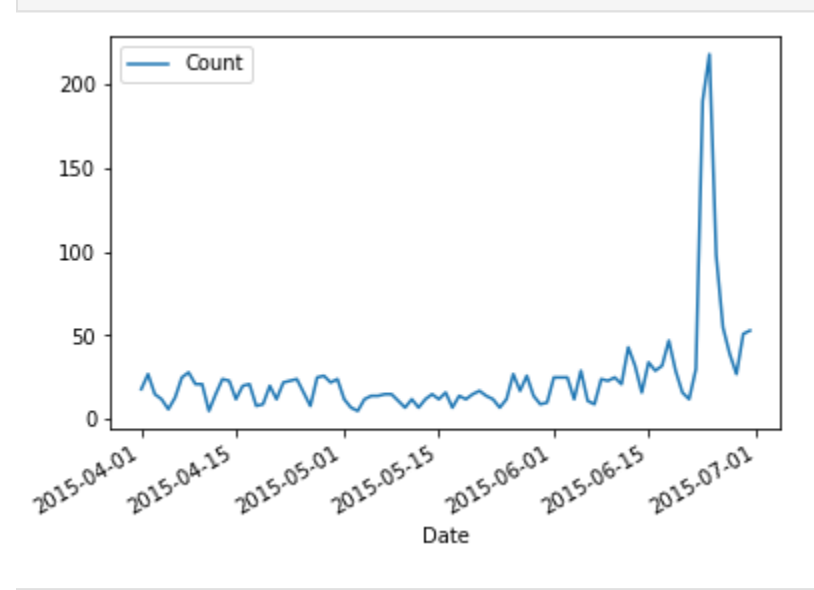
	Date	Count
0	2015-06-24	218
1	2015-06-23	190
2	2015-06-25	98
3	2015-06-26	55
4	2015-06-30	53
...	...	...
86	2015-05-10	7
87	2015-05-24	7
88	2015-04-05	6
89	2015-04-11	5
90	2015-05-03	5

91 rows × 2 columns

Graphically above data :

In [17]:

```python
daily.plot(x='Date',y='Count',kind='line')
plt.show()
```



In [19]:

```python
df.groupby('Month').count()['Ticket #']
```

Out[19]:

```
Month
April     545
June     1280
May       399
Name: Ticket #, dtype: int64
```

In [22]:

```python
mnth=pd.DataFrame(df.groupby('Month').count()['Ticket #']).reset_index()
```

In [23]:

```python
mnth
```

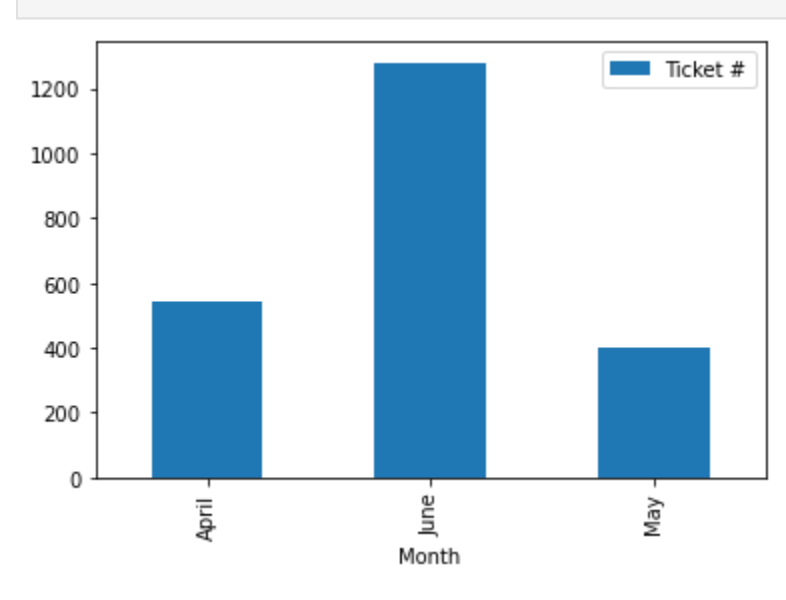
Out[23]:

	Month	Ticket #
0	April	545
1	June	1280
2	May	399

Graphically above data :

In [25]:

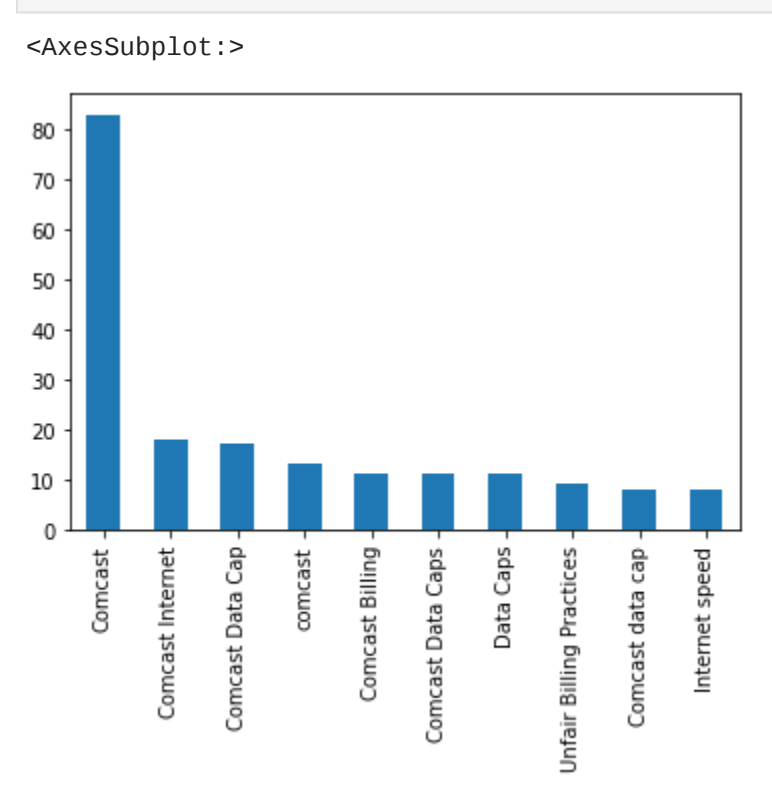
```python
mnth.plot(x='Month',y='Ticket #',kind='bar')
plt.show()
```



In [28]:

```python
df['Customer Complaint'].value_counts()[:10].plot.bar()
```

Out[28]: <AxesSubplot:>



In [29]:

```python
internet_issue1=df[df['Customer Complaint'].str.contains('speed')].count()['Ticket #']
internet_issue2=df[df['Customer Complaint'].str.contains('network')].count()['Ticket #']
internet_issue3=df[df['Customer Complaint'].str.contains('data')].count()['Ticket #']

total_internet_issue=internet_issue1+internet_issue2+internet_issue3
```

In [30]:

```python
total_internet_issue
```

Out[30]: 179

In [33]:

```python
billing_issue1=df[df['Customer Complaint'].str.contains('billing')].count()['Ticket #']
billing_issue2=df[df['Customer Complaint'].str.contains('charges')].count()['Ticket #']


total_billing_issue=billing_issue1+billing_issue2
```

In [34]:

```python
total_billing_issue
```

Out[34]: 169

In [35]:

```python
service_issue1=df[df['Customer Complaint'].str.contains('service')].count()['Ticket #']
service_issue2=df[df['Customer Complaint'].str.contains('customer')].count()['Ticket #']

total_service_issue=service_issue1+service_issue2
```

In [36]:

```python
total_service_issue
```

Out[36]: 360

In [38]:

```python
total_internet_issue,total_billing_issue,total_service_issue
```

Out[38]: (179, 169, 360)

In [39]:

```python
other=2224-(169+179+360)
```

In [40]:

```python
other
```

Out[40]: 1516

In [41]:

```python
df['Status'].unique()
```

Out[41]: array(['Closed', 'Open', 'Solved', 'Pending'], dtype=object)

In [42]:

```python
df['newstatus']=['Open' if st=='Open' or st=='Pending' else 'Close' for st in df['Status'] ]
```

In [43]:

```python
df
```

Out[43]:

	Ticket #	Customer Complaint	Date	Date_month_year	Time	Received Via	City	State	Zip code	Status	Filing on Behalf of Someone	Month	newstatus
0	250635	Comcast Cable Internet Speeds	2015-04-22	22-Apr-15	3:53:50 PM	Customer Care Call	Abingdon	Maryland	21009	Closed	No	April	Close
1	223441	Payment disappear - service got disconnected	2015-04-08	04-Aug-15	10:22:56 AM	Internet	Acworth	Georgia	30102	Closed	No	April	Close
2	242732	Speed and Service	2015-04-18	18-Apr-15	9:55:47 AM	Internet	Acworth	Georgia	30101	Closed	Yes	April	Close
3	277946	Comcast Imposed a New Usage Cap of 300GB that ...	2015-05-07	05-Jul-15	11:59:35 AM	Internet	Acworth	Georgia	30101	Open	Yes	May	Open
4	307175	Comcast not working and no service to boot	2015-05-26	26-May-15	1:25:26 PM	Internet	Acworth	Georgia	30101	Solved	No	May	Close
...	...	...	...	...	...	...	...	...	...	...	...	...	...
2219	213550	Service Availability	2015-04-02	04-Feb-15	9:13:18 AM	Customer Care Call	Youngstown	Florida	32466	Closed	No	April	Close
2220	318775	Comcast Monthly Billing for Returned Modem	2015-06-02	06-Feb-15	1:24:39 PM	Customer Care Call	Ypsilanti	Michigan	48197	Solved	No	June	Close
2221	331188	complaint about comcast	2015-06-09	06-Sep-15	5:28:41 PM	Internet	Ypsilanti	Michigan	48197	Solved	No	June	Close
2222	360489	Extremely unsatisfied Comcast customer	2015-06-23	23-Jun-15	11:13:30 PM	Customer Care Call	Ypsilanti	Michigan	48197	Solved	No	June	Close
2223	363614	Comcast, Ypsilanti MI Internet Speed	2015-06-24	24-Jun-15	10:28:33 PM	Customer Care Call	Ypsilanti	Michigan	48198	Open	Yes	June	Open

2224 rows × 13 columns

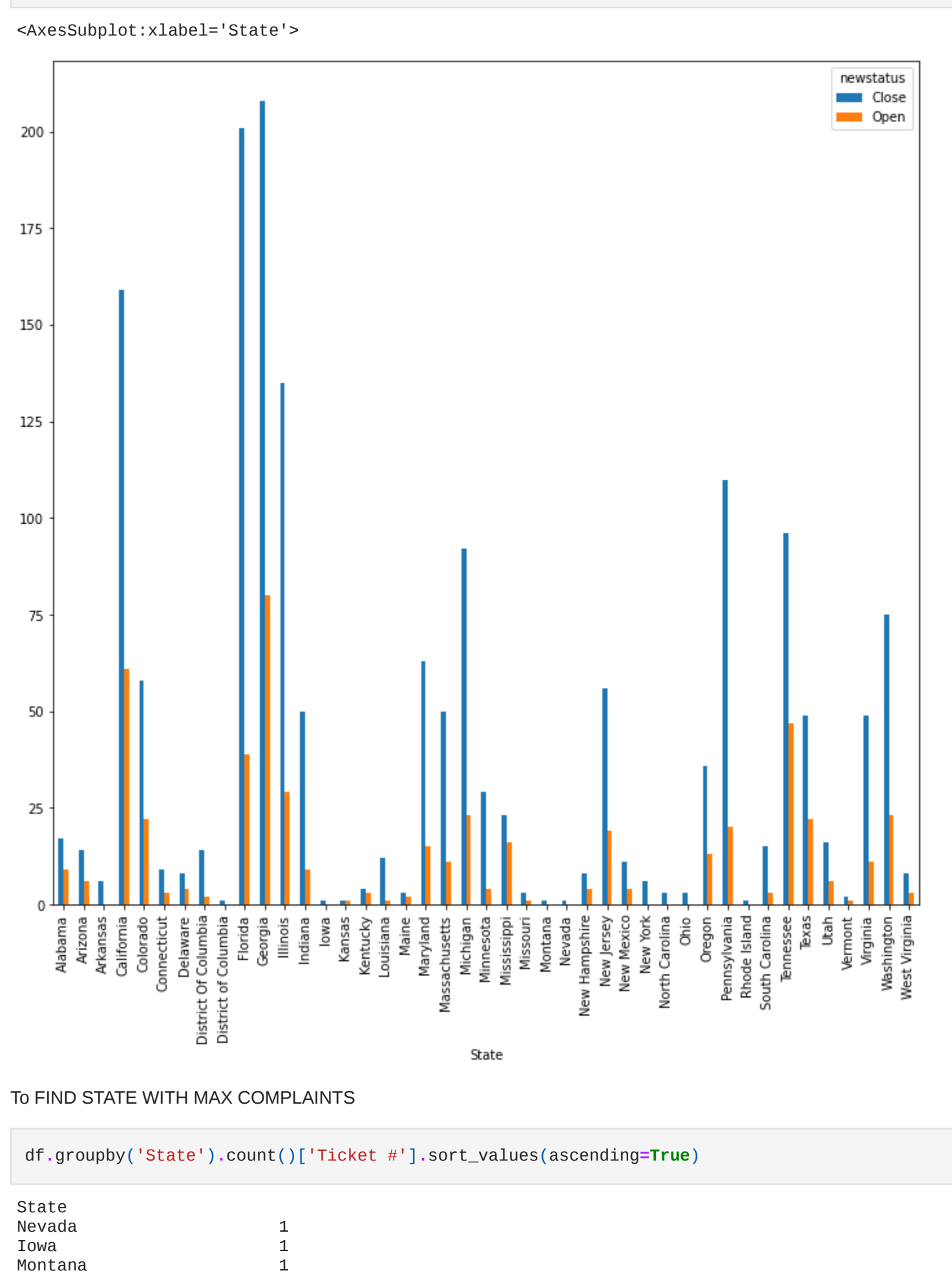
In [47]:

```python
state_cmp=df.groupby(['State','newstatus']).size().unstack()
```

In [51]:

```python
state_cmp.plot.bar(figsize=(12,12))
```

Out[51]: <AxesSubplot:xlabel='State'>



newstatus
Close
Open
200
175
150
125
100
75
50
25
0
Alabama
Arizona
Arkansas
California
Colorado
Connecticut
Delaware
District Of Columbia
District of Columbia
Florida
Georgia
Illinois
Indiana
Iowa
Kansas
Kentucky
Louisiana
Maine
Maryland
Massachusetts
Michigan
Minnesota
Mississippi
Missouri
Montana
Nevada
New Hampshire
New Jersey
New Mexico
New York
North Carolina
Ohio
Oregon
Pennsylvania
Rhode Island
South Carolina
Tennessee
Texas
Utah
Vermont
Virginia
Washington
West Virginia
State

To FIND STATE WITH MAX COMPLAINTS

In [54]:

```python
df.groupby('State').count()['Ticket #'].sort_values(ascending=True)
```

Out[54]:

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

TO FIND STATE WITH HIGHEST UNRESOLVED COMPLAINTS

In [55]:

```python
unresolved_data=df.groupby(['State','newstatus']).size().unstack().fillna(0).sort_values('Open',ascending=False)
```

In [56]:

```python
unresolved_data['unresolved_cmp_prct']=unresolved_data['Open']/unresolved_data['Open'].sum()*100
```

In [57]:

```python
unresolved_data
```

Out[57]:

newstatus	Close	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

COMPLAINTES RESOLVED TILL DATE RECEIVED VIA INTERNET AND CUSTOMER CARE

In [58]:

```python
resolved_data=df.groupby(['Received Via','newstatus']).size().unstack()
```

In [59]:

```python
resolved_data['resolved']=resolved_data['Close']/resolved_data['Close'].sum()*100
```

In [60]:

```python
resolved_data['resolved']
```

Out[60]:

```
Received Via
Customer Care Call    50.615114
Internet              49.384886
Name: resolved, dtype: float64
```

THANK YOU

In [ ]: