In [54]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [55]:
```
df=pd.read_csv("E:\\simpliearn\\python projects\\Comcast_telecom_complaints_data.csv
```

In [56]:
```
df
```

Out[56]:

| | Ticket # | Customer Complaint | Date | Date_month_year | Time | Received Via | City | State | Z coc |
|---|---|---|---|---|---|---|---|---|---|
| **0** | 250635 | Comcast Cable Internet Speeds | 22-04-15 | 22-Apr-15 | 3:53:50 PM | Customer Care Call | Abingdon | Maryland | 210( |
| **1** | 223441 | Payment disappear - service got disconnected | 04-08-15 | 04-Aug-15 | 10:22:56 AM | Internet | Acworth | Georgia | 301( |
| **2** | 242732 | Speed and Service | 18-04-15 | 18-Apr-15 | 9:55:47 AM | Internet | Acworth | Georgia | 301( |
| **3** | 277946 | Comcast Imposed a New Usage Cap of 300GB that ... | 05-07-15 | 05-Jul-15 | 11:59:35 AM | Internet | Acworth | Georgia | 301( |
| **4** | 307175 | Comcast not working and no service to boot | 26-05-15 | 26-May-15 | 1:25:26 PM | Internet | Acworth | Georgia | 301( |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | |
| **2219** | 213550 | Service Availability | 04-02-15 | 04-Feb-15 | 9:13:18 AM | Customer Care Call | Youngstown | Florida | 324( |
| **2220** | 318775 | Comcast Monthly Billing for Returned Modem | 06-02-15 | 06-Feb-15 | 1:24:39 PM | Customer Care Call | Ypsilanti | Michigan | 481( |
| **2221** | 331188 | complaint about comcast | 06-09-15 | 06-Sep-15 | 5:28:41 PM | Internet | Ypsilanti | Michigan | 481( |
| **2222** | 360489 | Extremely unsatisfied Comcast customer | 23-06-15 | 23-Jun-15 | 11:13:30 PM | Customer Care Call | Ypsilanti | Michigan | 481( |
| **2223** | 363614 | Comcast, Ypsilanti MI Internet Speed | 24-06-15 | 24-Jun-15 | 10:28:33 PM | Customer Care Call | Ypsilanti | Michigan | 481( |

2224 rows × 11 columns

In [46]:
```
df.head()
```

Out[46]:

| | Ticket # | Customer Complaint | Date | Date_month_year | Time | Received Via | City | State | Zip code | St |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 250635 | Comcast Cable Internet Speeds | 22-04-15 | 22-Apr-15 | 3:53:50 PM | Customer Care Call | Abingdon | Maryland | 21009 | Cl |
| **1** | 223441 | Payment disappear - service got disconnected | 04-08-15 | 04-Aug-15 | 10:22:56 AM | Internet | Acworth | Georgia | 30102 | Cl |
| **2** | 242732 | Speed and Service | 18-04-15 | 18-Apr-15 | 9:55:47 AM | Internet | Acworth | Georgia | 30101 | Cl |
| **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 | ( |
| **4** | 307175 | Comcast not working and no service to boot | 26-05-15 | 26-May-15 | 1:25:26 PM | Internet | Acworth | Georgia | 30101 | Sc |

In [57]:
```
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 [48]:
```
df.isnull().sum()
```

Out[48]:
```
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 [ ]:

In [49]:
```
df=df.drop(['Ticket #','Time'],axis=1)
```

In [50]:
```
df
```

Out[50]:

| | Customer Complaint | Date | Date_month_year | Received Via | City | State | Zip code | Status | Filing Behalf Someo |
|---|---|---|---|---|---|---|---|---|---|
| **0** | Comcast Cable Internet Speeds | 22-04-15 | 22-Apr-15 | Customer Care Call | Abingdon | Maryland | 21009 | Closed | |
| **1** | Payment disappear - service got disconnected | 04-08-15 | 04-Aug-15 | Internet | Acworth | Georgia | 30102 | Closed | |
| **2** | Speed and Service | 18-04-15 | 18-Apr-15 | Internet | Acworth | Georgia | 30101 | Closed | |
| **3** | Comcast Imposed a New Usage Cap of 300GB that ... | 05-07-15 | 05-Jul-15 | Internet | Acworth | Georgia | 30101 | Open | |
| **4** | Comcast not working and no service to boot | 26-05-15 | 26-May-15 | Internet | Acworth | Georgia | 30101 | Solved | |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **2219** | Service Availability | 04-02-15 | 04-Feb-15 | Customer Care Call | Youngstown | Florida | 32466 | Closed | |
| **2220** | Comcast Monthly Billing for Returned Modem | 06-02-15 | 06-Feb-15 | Customer Care Call | Ypsilanti | Michigan | 48197 | Solved | |
| **2221** | complaint about comcast | 06-09-15 | 06-Sep-15 | Internet | Ypsilanti | Michigan | 48197 | Solved | |
| **2222** | Extremely unsatisfied Comcast customer | 23-06-15 | 23-Jun-15 | Customer Care Call | Ypsilanti | Michigan | 48197 | Solved | |
| **2223** | Comcast, Ypsilanti MI Internet Speed | 24-06-15 | 24-Jun-15 | Customer Care Call | Ypsilanti | Michigan | 48198 | Open | |

2224 rows × 9 columns

In [51]:
```
# Provide the trend chart for the number of complaints at monthly and daily granular
```

In [52]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2224 entries, 0 to 2223
Data columns (total 9 columns):
 #   Column                       Non-Null Count  Dtype 
---  ------                       --------------  ----- 
 0   Customer Complaint           2224 non-null   object
 1   Date                         2224 non-null   object
 2   Date_month_year              2224 non-null   object
 3   Received Via                 2224 non-null   object
 4   City                         2224 non-null   object
 5   State                        2224 non-null   object
 6   Zip code                     2224 non-null   int64 
 7   Status                       2224 non-null   object
 8   Filing on Behalf of Someone  2224 non-null   object
dtypes: int64(1), object(8)
memory usage: 156.5+ KB
```

In [58]:
```
df['Date_month_year']=df['Date_month_year'].apply(pd.to_datetime)
```

In [59]:
```
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   datetime64[ns]
 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 [ ]:
```
#setting date as index
```

In [60]:
```
df=df.set_index('Date_month_year')
```

In [61]:
```
df.head()
```

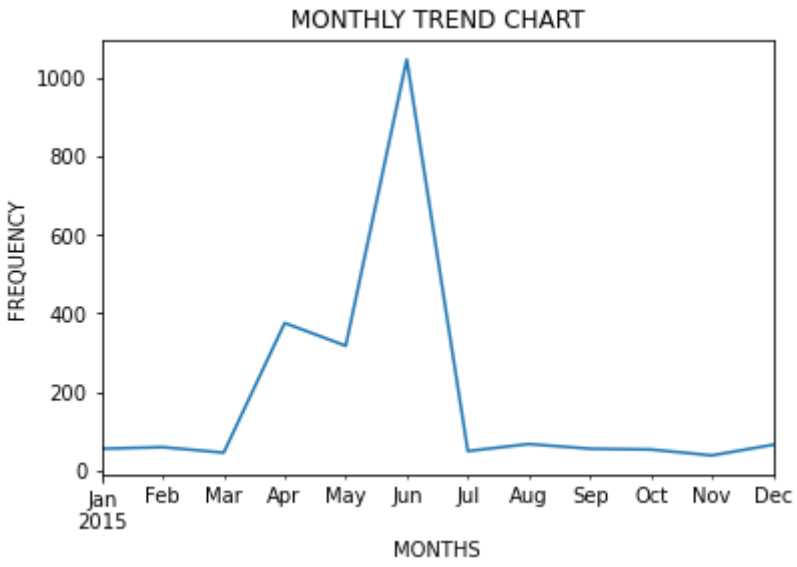
Out[61]:

| Date_month_year | Ticket # | Customer Complaint | Date | Time | Received Via | City | State | Zip code | Statu |
|---|---|---|---|---|---|---|---|---|---|
| **2015-04-22** | 250635 | Comcast Cable Internet Speeds | 22-04-15 | 3:53:50 PM | Customer Care Call | Abingdon | Maryland | 21009 | Close |

| Date_month_year | Ticket # | Customer Complaint | Date | Time | Received Via | City | State | Zip code | Statu |
|---|---|---|---|---|---|---|---|---|---|
| **2015-08-04** | 223441 | Payment disappear - service got disconnected | 04-08-15 | 10:22:56 AM | Internet | Acworth | Georgia | 30102 | Close |
| **2015-04-18** | 242732 | Speed and Service | 18-04-15 | 9:55:47 AM | Internet | Acworth | Georgia | 30101 | Close |
| **2015-07-05** | 277946 | Comcast Imposed a New Usage Cap of 300GB that ... | 05-07-15 | 11:59:35 AM | Internet | Acworth | Georgia | 30101 | Ope |
| **2015-05-26** | 307175 | Comcast not working and no service to boot | 26-05-15 | 1:25:26 PM | Internet | Acworth | Georgia | 30101 | Solve |

In [62]:
```
month=df.groupby(pd.Grouper(freq='M')).size().plot()
plt.xlabel('MONTHS')
plt.ylabel('FREQUENCY')
plt.title('MONTHLY TREND CHART')
```

Out[62]:
```
Text(0.5, 1.0, 'MONTHLY TREND CHART')
```




In [63]:
```
df['Date'].value_counts()
```

Out[63]:
```
24-06-15    218
23-06-15    190
25-06-15     98
26-06-15     55
30-06-15     53
           ...
05-02-15      7
05-12-15      7
04-05-15      6
05-03-15      5
```

```
04-11-15      5
Name: Date, Length: 91, dtype: int64
```

In [64]:
```
df=df.sort_values(by='Date')
```

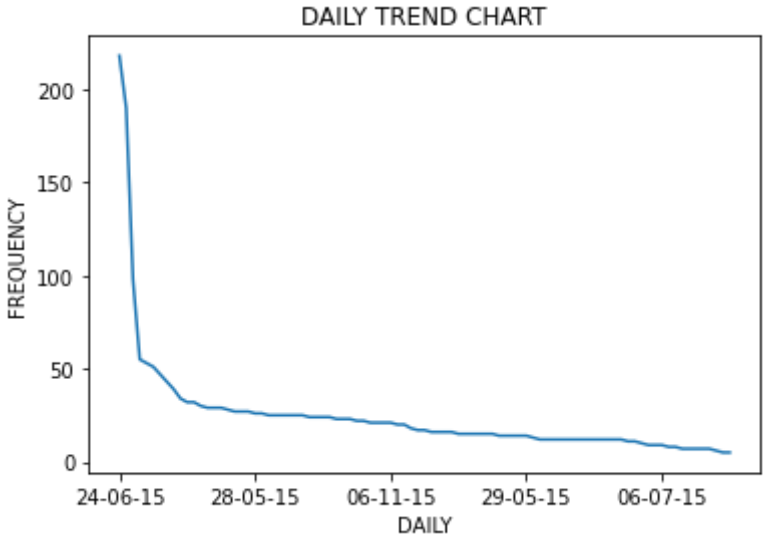
In [65]:
```
df.head()
```

Out[65]:

| Date_month_year | Ticket # | Customer Complaint | Date | Time | Received Via | City | State | Zip code | St |
|---|---|---|---|---|---|---|---|---|---|
| **2015-01-04** | 211976 | Fraudulent claims reported to collections agency | 04-01-15 | 1:26:53 PM | Customer Care Call | Atlanta | Georgia | 30312 | Cl |
| **2015-01-04** | 211677 | Comcast refusal of service | 04-01-15 | 12:01:06 PM | Customer Care Call | Wayne | Pennsylvania | 19087 | Cl |
| **2015-01-04** | 212507 | Comcast Cable | 04-01-15 | 3:54:43 PM | Internet | Franklin | Tennessee | 37067 | Cl |
| **2015-01-04** | 213120 | Data Overages | 04-01-15 | 8:05:57 PM | Internet | Savannah | Georgia | 31406 | Cl |
| **2015-01-04** | 211478 | Comcast | 04-01-15 | 10:47:35 AM | Internet | North Huntingdon | Pennsylvania | 15642 | Cl |

In [66]:
```
df['Date'].value_counts().plot()
plt.xlabel('DAILY')
plt.ylabel('FREQUENCY')
plt.title('DAILY TREND CHART')
```

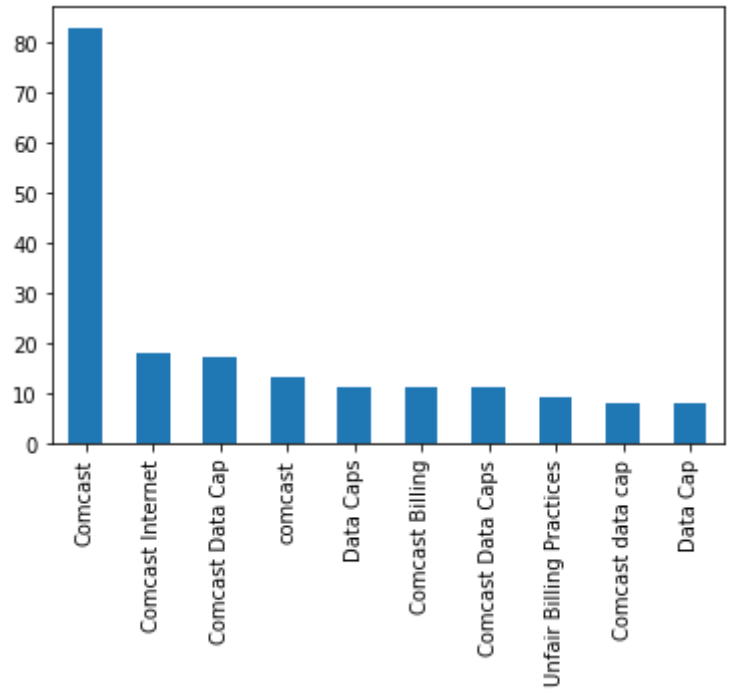
Out[66]:
```
Text(0.5, 1.0, 'DAILY TREND CHART')
```




In [ ]:
```
# Provide a table with the frequency of complaint types.
```

In [67]:
```
df['Customer Complaint'].value_counts()[:10].plot.bar()
```

Out[67]: <AxesSubplot:>




In [ ]:
```
# Which complaint types are maximum i.e., around internet, network issues, or across
```

In [68]:
```
internet_issues1=df[df['Customer Complaint'].str.contains('network')].count()
```

In [ ]:
```
internet_issues1
```

In [69]:
```
internet_issues2=df[df['Customer Complaint'].str.contains('speed')].count()
```

In [ ]:
```
internet_issues2
```

In [70]:
```
internet_issues3=df[df['Customer Complaint'].str.contains('data')].count()
```

In [ ]:
```
internet_issues3
```

In [71]:
```
internet_issues4=df[df['Customer Complaint'].str.contains('internet')].count()
```

In [ ]:
```
internet_issues4
```

In [72]:
```
billing_issues1=df[df['Customer Complaint'].str.contains('billing')].count()
```

In [ ]:
```
billing_issues1
```

In [73]:
```
billing_issues2=df[df['Customer Complaint'].str.contains('bill')].count()
```

In [ ]:
```
billing_issues2
```

In [74]:
```
billing_issues3=df[df['Customer Complaint'].str.contains('charges')].count()
```

In [77]:
```
service_issues1=df[df['Customer Complaint'].str.contains('service')].count()
```

In [78]:
```
service_issues2=df[df['Customer Complaint'].str.contains('customer')].count()
```

In [79]:
```
total_number_issues=internet_issues1+internet_issues2+internet_issues3+internet_issu
```

In [80]:
```
total_number_internet_issues=internet_issues1+internet_issues2+internet_issues3+inte
```

In [81]:
```
total_number_internet_issues
```

Out[81]:
```
Ticket #                      374
Customer Complaint            374
Date                          374
Time                          374
Received Via                  374
City                          374
State                         374
Zip code                      374
Status                        374
Filing on Behalf of Someone   374
dtype: int64
```

In [82]:
```
total_number_billing_issues=billing_issues1+billing_issues2+billing_issues3
```

In [83]:
```
total_number_billing_issues
```

Out[83]:
```
Ticket #                      353
Customer Complaint            353
Date                          353
Time                          353
Received Via                  353
City                          353
State                         353
Zip code                      353
Status                        353
Filing on Behalf of Someone   353
dtype: int64
```

In [84]:
```
total_number_service_issues=service_issues1+service_issues2
```

In [85]:
```
total_number_service_issues
```

Out[85]:
```
Ticket #                      360
Customer Complaint            360
Date                          360
Time                          360
Received Via                  360
City                          360
State                         360
Zip code                      360
Status                        360
Filing on Behalf of Someone   360
dtype: int64
```

In [86]:
```
total_number_issues
```

Out[86]:
```
Ticket #                      1087
Customer Complaint            1087
Date                          1087
Time                          1087
Received Via                  1087
City                          1087
State                         1087
Zip code                      1087
Status                        1087
Filing on Behalf of Someone   1087
dtype: int64
```

In [89]:
```
df.shape
```

Out[89]: (2224, 10)

In [90]:
```
other_issue=2224-total_number_issues
```

In [91]:
```
other_issue
```

Out[91]:
```
Ticket #                      1137
Customer Complaint            1137
Date                          1137
Time                          1137
Received Via                  1137
City                          1137
State                         1137
Zip code                      1137
Status                        1137
Filing on Behalf of Someone   1137
dtype: int64
```

In [92]:
```
#Create a new categorical variable with value as Open and Closed. Open & Pending is
```

In [94]:
```
df['Status'].unique()
```

Out[94]: array(['Closed', 'Open', 'Solved', 'Pending'], dtype=object)

In [98]:
```
df['newstatus']=['Open' if status=='Open' or status=='Pending' else 'Closed' for sta
```

In [99]:
```
df['newstatus']
```

Out[99]:
```
Date_month_year
2015-01-04    Closed
2015-01-04    Closed
2015-01-04    Closed
2015-01-04    Closed
2015-01-04    Closed
               ...
2015-05-31      Open
2015-05-31      Open
2015-05-31      Open
2015-05-31      Open
2015-05-31    Closed
Name: newstatus, Length: 2224, dtype: object
```

In [100…
```
df.sample(10)
```

Out[100…

| | Ticket # | Customer Complaint | Date | Time | Received Via | City | State | Zip code |
|---|---|---|---|---|---|---|---|---|
| **Date_month_year** | | | | | | | | |
| **2015-06-18** | 347736 | Data Cap | 18-06-15 | 12:35:23 AM | Internet | Tucson | Arizona | 85741 |
| **2015-07-04** | 222366 | intermittant phone and internet | 04-07-15 | 4:42:53 PM | Customer Care Call | Salem | New Hampshire | 3079 |
| **2015-06-24** | 360673 | Comcast high prices and throttling speeds | 24-06-15 | 12:57:49 AM | Customer Care Call | Houston | Texas | 77039 |

| Date_month_year | Ticket # | Customer Complaint | Date | Time | Received Via | City | State | Zip code |
|---|---|---|---|---|---|---|---|---|
| **2015-06-13** | 338674 | Comcast outage - Bay Area | 13-06-15 | 2:13:06 AM | Internet | East Palo Alto | California | 94303 |
| **2015-05-06** | 324795 | Rates | 06-05-15 | 9:59:39 AM | Internet | Bala Cynwyd | Pennsylvania | 19004 |
| **2015-03-06** | 322409 | FCC Complaint against comcast/xfinity on provi... | 06-03-15 | 10:20:53 PM | Customer Care Call | Canonsburg | Pennsylvania | 15317 |
| **2015-06-30** | 375249 | Comcast cable | 30-06-15 | 3:47:29 PM | Internet | Beach Haven | New Jersey | 8008 |
| **2015-06-16** | 342665 | cannot access my email or service at all | 16-06-15 | 9:51:17 AM | Customer Care Call | Gathersburg | Maryland | 20877 |
| **2015-05-26** | 306805 | Failure to deliver service | 26-05-15 | 11:29:54 AM | Customer Care Call | Bainbridge Island | Washington | 98110 |
| **2015-06-28** | 370846 | Comcast deceptive advertising, overage charges | 28-06-15 | 12:08:05 PM | Internet | Atlanta | Georgia | 30329 |

In [101…
```
#Provide state wise status of complaints in a stacked bar chart. Use the categorized
```

In [107…
```
state_complaints=df.groupby(['State','newstatus']).size().unstack()
```

In [108…
```
state_complaints
```

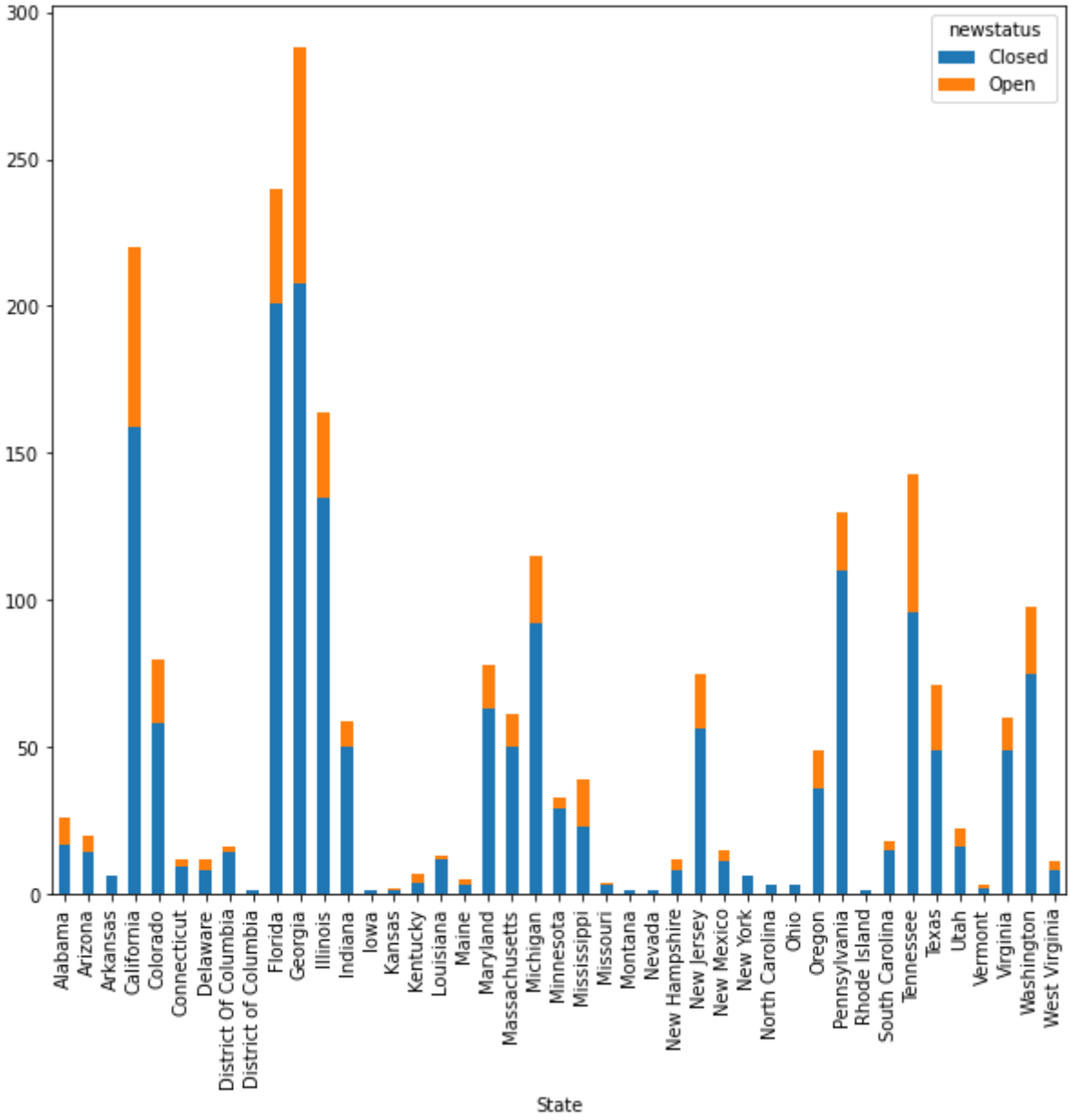
Out[108…

| State | newstatus Closed | Open |
|---|---|---|
| **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 |

| newstatus<br>State | Closed | Open |
|---|---|---|
| **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 [110…
```
state_complaints.plot.bar(figsize=(10,9),stacked=True)
#here i just analyse that georgia has highest number of open and closed issues
```

Out[110… <AxesSubplot:xlabel='State'>




In [111…
```
#Which state has the maximum complaints
```

In [112…
```
df.groupby(['State']).size().sort_values(ascending=False)
#georgia has highest number of complaints
```

Out[112…
```
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
New Hampshire            12
Connecticut              12
Delaware                 12
West Virginia            11
Kentucky                  7
New York                  6
Arkansas                  6
Maine                     5
Missouri                  4
North Carolina            3
Vermont                   3
Ohio                      3
Kansas                    2
District of Columbia      1
Rhode Island              1
Montana                   1
Iowa                      1
Nevada                    1
dtype: int64
```

In [113…
```
#Which state has the highest percentage of unresolved complaints
```

In [114…
```
df['newstatus'].value_counts()
```

Out[114…
```
Closed    1707
Open       517
Name: newstatus, dtype: int64
```

In [115…
```
unresolved_data=df.groupby(['State','newstatus']).size().unstack().fillna(0).sort_va
```

In [116…
```
unresolved_data
```

Out[116…

| newstatus | Closed | Open |
|---|---|---|
| State | | |
| **Georgia** | 208.0 | 80.0 |
| **California** | 159.0 | 61.0 |
| **Tennessee** | 96.0 | 47.0 |
| **Florida** | 201.0 | 39.0 |
| **Illinois** | 135.0 | 29.0 |
| **Washington** | 75.0 | 23.0 |
| **Michigan** | 92.0 | 23.0 |
| **Colorado** | 58.0 | 22.0 |
| **Texas** | 49.0 | 22.0 |
| **Pennsylvania** | 110.0 | 20.0 |
| **New Jersey** | 56.0 | 19.0 |
| **Mississippi** | 23.0 | 16.0 |

| newstatus<br>State | Closed | Open |
|---|---|---|
| **Maryland** | 63.0 | 15.0 |
| **Oregon** | 36.0 | 13.0 |
| **Virginia** | 49.0 | 11.0 |
| **Massachusetts** | 50.0 | 11.0 |
| **Alabama** | 17.0 | 9.0 |
| **Indiana** | 50.0 | 9.0 |
| **Utah** | 16.0 | 6.0 |
| **Arizona** | 14.0 | 6.0 |
| **New Hampshire** | 8.0 | 4.0 |
| **New Mexico** | 11.0 | 4.0 |
| **Minnesota** | 29.0 | 4.0 |
| **Delaware** | 8.0 | 4.0 |
| **West Virginia** | 8.0 | 3.0 |
| **Connecticut** | 9.0 | 3.0 |
| **Kentucky** | 4.0 | 3.0 |
| **South Carolina** | 15.0 | 3.0 |
| **Maine** | 3.0 | 2.0 |
| **District Of Columbia** | 14.0 | 2.0 |
| **Kansas** | 1.0 | 1.0 |
| **Vermont** | 2.0 | 1.0 |
| **Missouri** | 3.0 | 1.0 |
| **Louisiana** | 12.0 | 1.0 |
| **Montana** | 1.0 | 0.0 |
| **Rhode Island** | 1.0 | 0.0 |
| **Ohio** | 3.0 | 0.0 |
| **District of Columbia** | 1.0 | 0.0 |
| **North Carolina** | 3.0 | 0.0 |
| **New York** | 6.0 | 0.0 |
| **Nevada** | 1.0 | 0.0 |
| **Arkansas** | 6.0 | 0.0 |
| **Iowa** | 1.0 | 0.0 |

In [117…
```
unresolved_data['unresolved_cmp_prec']=unresolved_data['Open']/unresolved_data['Open
```

In [118…
```
unresolved_data
```

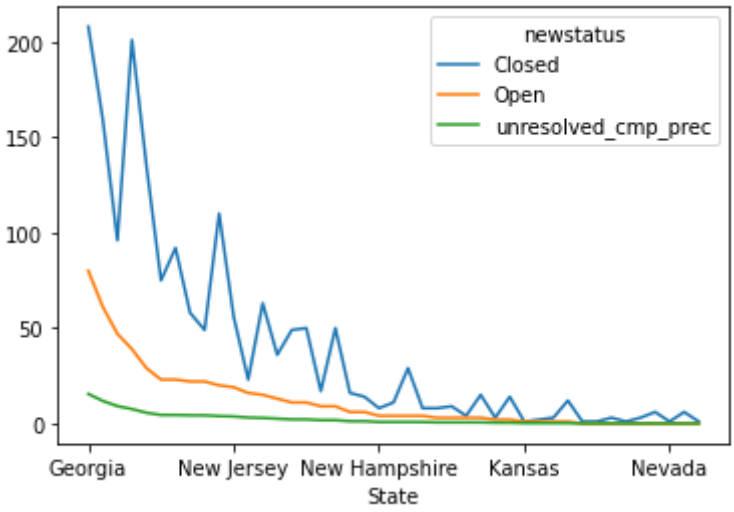
Out[118…

| newstatus<br>State | Closed | Open | unresolved_cmp_prec |
|---|---|---|---|
| **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 |

| newstatus | Closed | Open | unresolved_cmp_prec |
|---|---|---|---|
| State | | | |
| **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 |

In [119…
```
unresolved_data.plot()
```

Out[119… <AxesSubplot:xlabel='State'>




In [120…
```
# Provide the percentage of complaints resolved till date, which were received throu
```

In [122…
```
resolved_data=df.groupby(['Received Via','newstatus']).size().unstack().fillna(0)
```

In [123…
```
resolved_data
```

Out[123…

| newstatus | Closed | Open |
|---|---|---|
| Received Via | | |
| **Customer Care Call** | 864 | 255 |
| **Internet** | 843 | 262 |

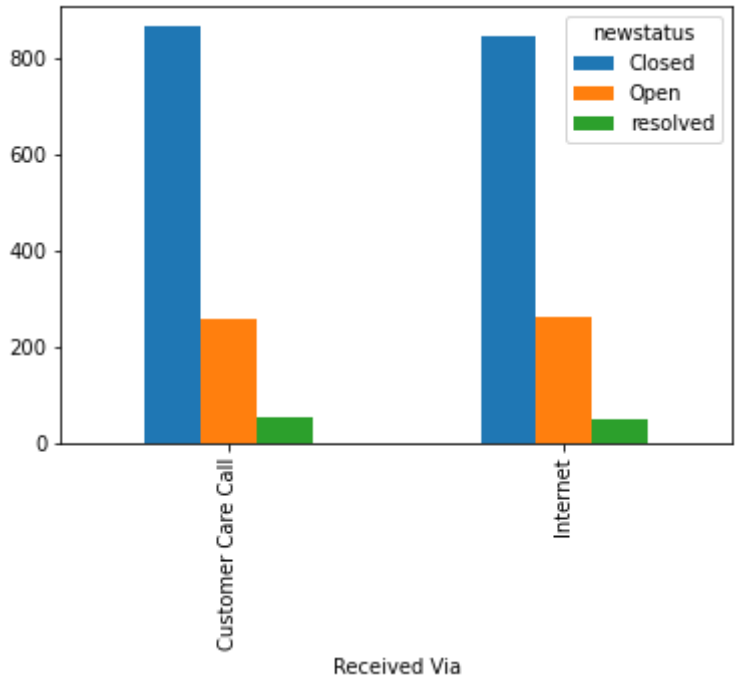
In [126…
```
resolved_data['resolved']=resolved_data['Closed']/resolved_data['Closed'].sum()*100
```

In [127…
```
resolved_data['resolved']
```

Out[127…
```
Received Via
Customer Care Call    50.615114
Internet              49.384886
Name: resolved, dtype: float64
```

In [130…
```
resolved_data.plot.bar()
```

Out[130… <AxesSubplot:xlabel='Received Via'>




In [ ]: