```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
```

importing data in Python Environment

```
df=pd.read_csv("Comcast_telecom_complaints_data.csv")
```

```
df.shape
```

```
(2224, 11)
```

```
df.head()
```

```
  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

  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


   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
```

```
df.isnull().sum()
```

```
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
```

```
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
```

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

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

```
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
```

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

```
df
```

```
     Ticket #                                    Customer Complaint
Date  \
0      250635                       Comcast Cable Internet Speeds 2015-
04-22
1      223441       Payment disappear - service got disconnected 2015-
04-08
2      242732                                    Speed and Service 2015-
04-18
3      277946  Comcast Imposed a New Usage Cap of 300GB that ... 2015-
05-07
4      307175         Comcast not working and no service to boot 2015-
05-26
...       ...                                                    ...
...
2219   213550                                Service Availability 2015-
04-02
2220   318775         Comcast Monthly Billing for Returned Modem 2015-
06-02
2221   331188                              complaint about comcast 2015-
06-09
2222   360489             Extremely unsatisfied Comcast customer 2015-
06-23
2223   363614               Comcast, Ypsilanti MI Internet Speed 2015-
06-24

     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  month
0        21009  Closed                          No  April
1        30102  Closed                          No  April
2        30101  Closed                         Yes  April
3        30101    Open                         Yes    May
4        30101  Solved                          No    May
...        ...     ...                         ...    ...
2219     32466  Closed                          No  April
2220     48197  Solved                          No   June
2221     48197  Solved                          No   June
2222     48197  Solved                          No   June
2223     48198    Open                         Yes   June

[2224 rows x 12 columns]
```

```
df['Date'].value_counts()
```

```
2015-06-24    218
2015-06-23    190
2015-06-25     98
2015-06-26     55
2015-06-30     53
              ...
2015-05-24      7
2015-05-02      7
2015-04-05      6
2015-04-11      5
2015-05-03      5
Name: Date, Length: 91, dtype: int64
```

```
dates=df.groupby("Date").size()
```

```
dates
```

```
Date
2015-04-01    18
2015-04-02    27
2015-04-03    15
2015-04-04    12
2015-04-05     6
```

```
              ..
2015-06-26    55
2015-06-27    39
2015-06-28    27
2015-06-29    51
2015-06-30    53
Length: 91, dtype: int64
```

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

```
daily=daily.rename(columns={0:'count'})
```

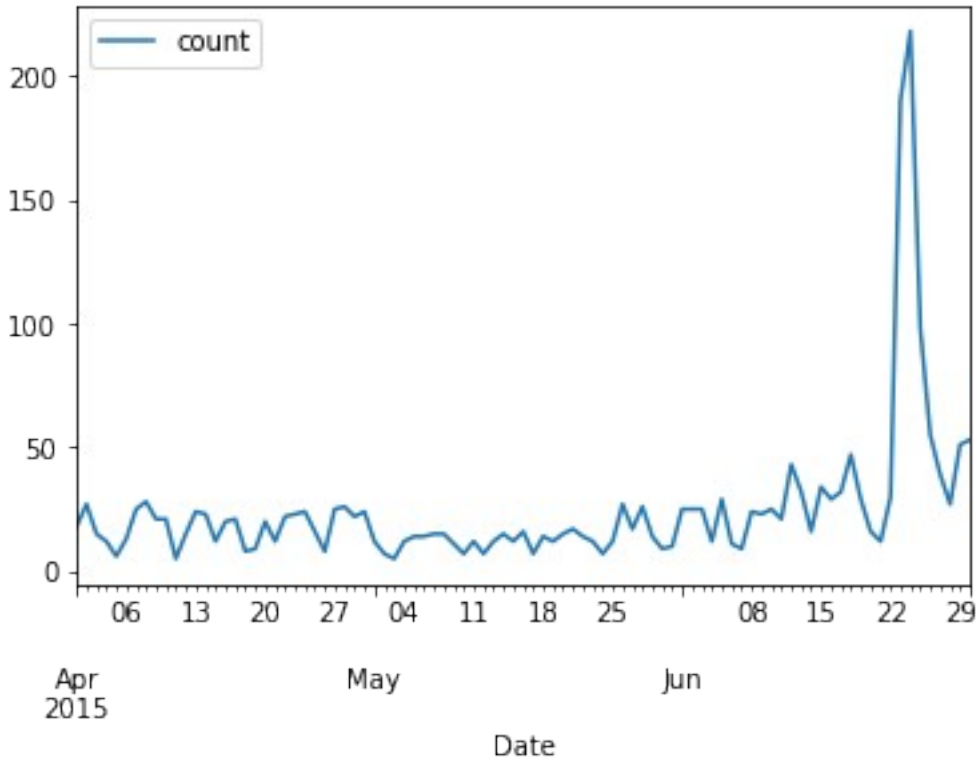
```
daily
```

```
         Date  count
0  2015-04-01     18
1  2015-04-02     27
2  2015-04-03     15
3  2015-04-04     12
4  2015-04-05      6
..        ...    ...
86 2015-06-26     55
87 2015-06-27     39
88 2015-06-28     27
89 2015-06-29     51
90 2015-06-30     53

[91 rows x 2 columns]
```

```
daily.plot(x='Date',y='count',kind='line')
```

```
<AxesSubplot:xlabel='Date'>
```




```
mnth=df.groupby('month').size()
```

```
mnth
```

```
month
April     545
June     1280
May       399
dtype: int64
```

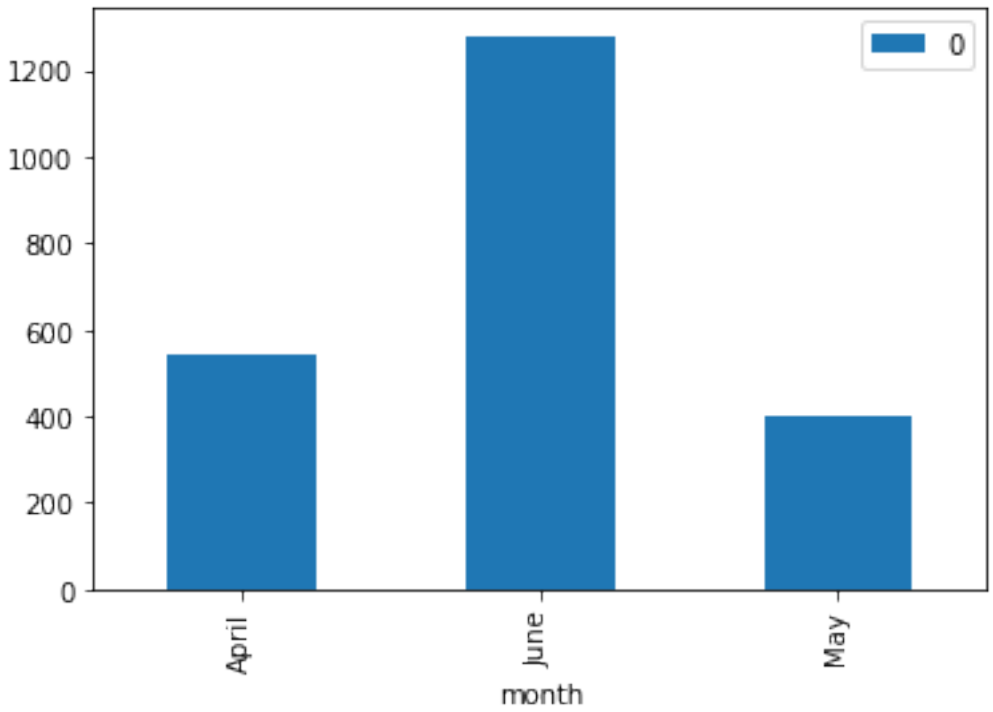
```
mnth_df=pd.DataFrame(mnth).reset_index()
```

```
mnth_df
```

```
   month     0
0  April   545
1   June  1280
2    May   399
```

```
mnth_df.plot(x='month',y=0,kind='bar')
```

```
<AxesSubplot:xlabel='month'>
```



Provide a table with the frequency of complaint types.

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

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

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

```
df['new_status']=['Open'   if st=="Open" or st=="Pending"   else
"Closed"   for st in df['Status']]
```

```
df
```

```
     Ticket #                                  Customer Complaint
Date  \
0      250635                     Comcast Cable Internet Speeds 2015-
04-22
1      223441       Payment disappear - service got disconnected 2015-
04-08
2      242732                                   Speed and Service 2015-
04-18
3      277946  Comcast Imposed a New Usage Cap of 300GB that ... 2015-
05-07
4      307175         Comcast not working and no service to boot 2015-
05-26
...       ...                                                  ...
...
2219   213550                               Service Availability 2015-
04-02
```

```
2220   318775         Comcast Monthly Billing for Returned Modem 2015-
06-02
2221   331188                             complaint about comcast 2015-
06-09
2222   360489             Extremely unsatisfied Comcast customer 2015-
06-23
2223   363614               Comcast, Ypsilanti MI Internet Speed 2015-
06-24

     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  month new_status
0        21009  Closed                          No  April     Closed
1        30102  Closed                          No  April     Closed
2        30101  Closed                         Yes  April     Closed
3        30101    Open                         Yes    May       Open
4        30101  Solved                          No    May     Closed
...        ...     ...                         ...    ...        ...
2219     32466  Closed                          No  April     Closed
2220     48197  Solved                          No   June     Closed
2221     48197  Solved                          No   June     Closed
2222     48197  Solved                          No   June     Closed
2223     48198    Open                         Yes   June       Open

[2224 rows x 13 columns]
```

Provide state wise status of complaints in a stacked bar chart

```
state_complain=df.groupby(['State','new_status']).size().unstack().fil
lna(0)
```

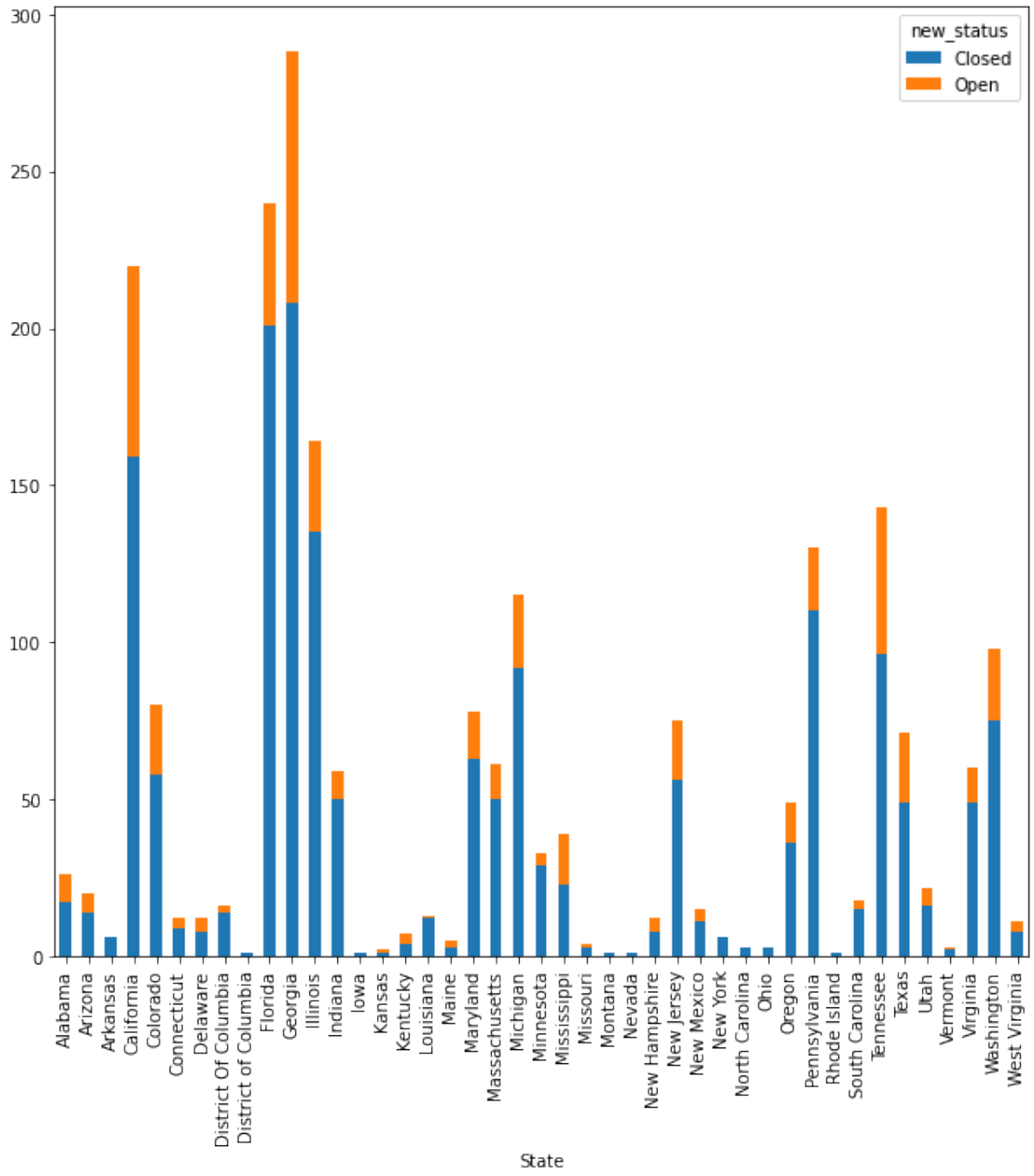
```
state_complain
```

new_status	Closed	Open
State		
Alabama	17.0	9.0
Arizona	14.0	6.0
Arkansas	6.0	0.0
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	0.0
Florida	201.0	39.0
Georgia	208.0	80.0
Illinois	135.0	29.0
Indiana	50.0	9.0
Iowa	1.0	0.0
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	0.0
Nevada	1.0	0.0
New Hampshire	8.0	4.0
New Jersey	56.0	19.0
New Mexico	11.0	4.0
New York	6.0	0.0
North Carolina	3.0	0.0
Ohio	3.0	0.0
Oregon	36.0	13.0
Pennsylvania	110.0	20.0
Rhode Island	1.0	0.0
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
```

```
state_complain.plot.bar(stacked=True,figsize=(10,10))
```

```
<AxesSubplot:xlabel='State'>
```



State which has the maximum complaints

```
df.groupby('State').size().sort_values(ascending=False)[:5]
```

```
State
Georgia       288
Florida       240
```

```
California    220
Illinois      164
Tennessee     143
dtype: int64
```

So we can conclude that State of Gorgia has the maximum complaints registered.

which state has the highest percentage of unresolved complaints

```
unresolved_data=df.groupby(['State','new_status']).size().unstack().fi
llna(0).sort_values(by='Open',ascending=False)

unresolved_data
```

new_status	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
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

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

```
unresolved_data
```

new_status	Closed	Open	unresolved_cmp_perct
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
```

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

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

resolved_data
```

```
new_status          Closed  Open
Received Via
Customer Care Call     864   255
Internet               843   262
```

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

resolved_data['resolved']
```

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