Project 3 - Comcast Telecom Consumer Complaints

In [1]:
```
# import python libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

In [2]:
```
### Import Data into Python Environment
```

In [3]:
```
df_complaints = pd.read_csv("Comcast_telecom_complaints_data.csv")
df_complaints.head()
```

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]:
```
# New column created date_index = timestamp
df_complaints["date_index"] = df_complaints["Date_month_year"] + " " + df_complaints["Time"]

df_complaints["date_index"]
# Typecasting to datetime
df_complaints["date_index"]= pd.to_datetime(df_complaints["date_index"])
df_complaints["Date_month_year"] = pd.to_datetime(df_complaints["Date_month_year"])
df_complaints= df_complaints.set_index(df_complaints["date_index"])
df_complaints.head()
```

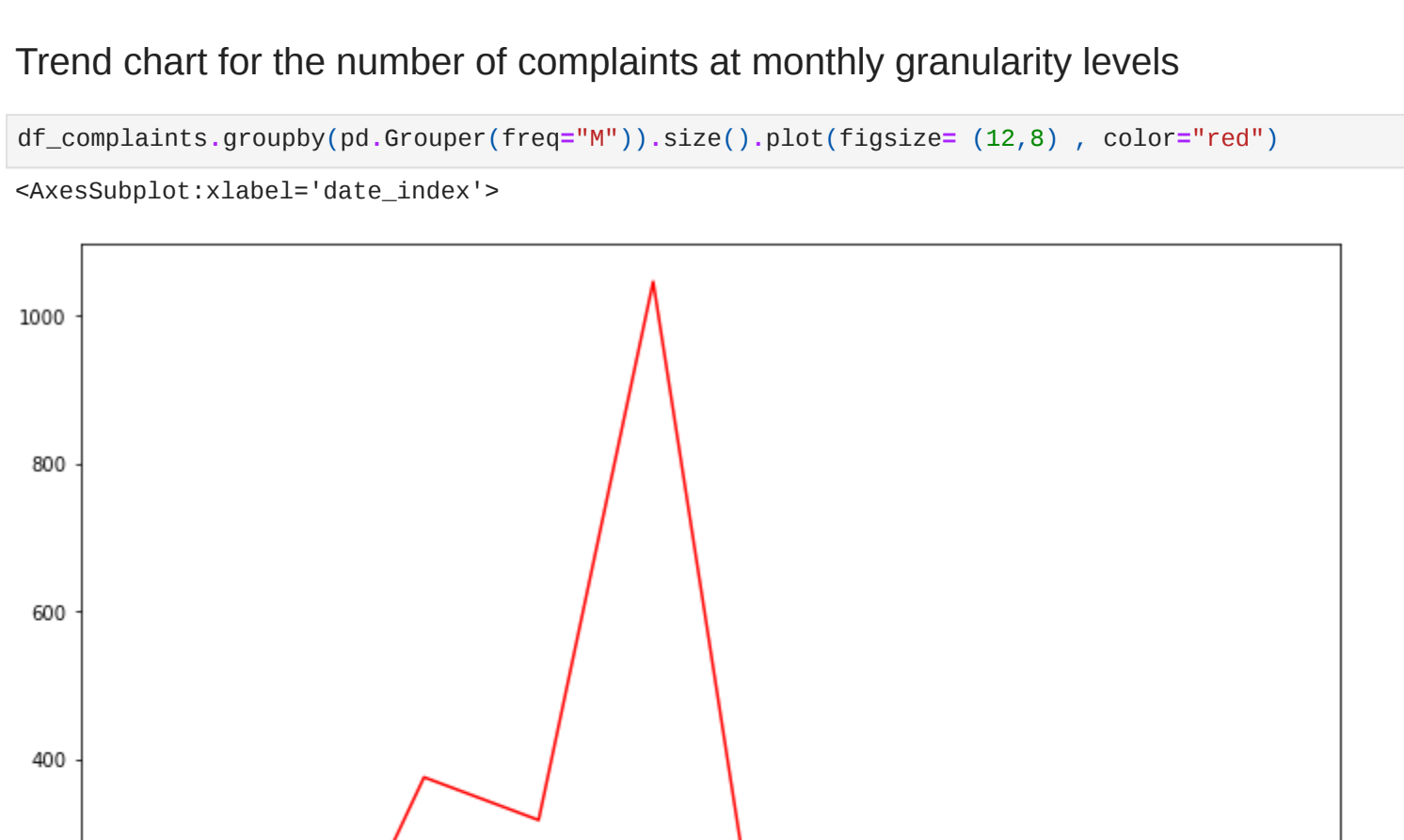
Out[4]:

date_index	Ticket #	Customer Complaint	Date	Date_month_year	Time	Received Via	City	State	Zip code	Status	Filing on Behalf of Someone	date_index
2015-04-22 15:53:50	250635	Comcast Cable Internet Speeds	22-04-15	2015-04-22	3:53:50 PM	Customer Care Call	Abingdon	Maryland	21009	Closed	No	2015-04-22 15:53:50
2015-08-04 10:22:56	223441	Payment disappear - service got disconnected	04-08-15	2015-08-04	10:22:56 AM	Internet	Acworth	Georgia	30102	Closed	No	2015-08-04 10:22:56
2015-04-18 09:55:47	242732	Speed and Service	18-04-15	2015-04-18	9:55:47 AM	Internet	Acworth	Georgia	30101	Closed	Yes	2015-04-18 09:55:47
2015-07-05 11:59:35	277946	Comcast Imposed a New Usage Cap of 300GB that ...	05-07-15	2015-07-05	11:59:35 AM	Internet	Acworth	Georgia	30101	Open	Yes	2015-07-05 11:59:35
2015-05-26 13:25:26	307175	Comcast not working and no service to boot	26-05-15	2015-05-26	1:25:26 PM	Internet	Acworth	Georgia	30101	Solved	No	2015-05-26 13:25:26

Trend chart for the number of complaints at monthly granularity levels

In [5]:
```
df_complaints.groupby(pd.Grouper(freq="M")).size().plot(figsize= (12,8) , color="red")
```

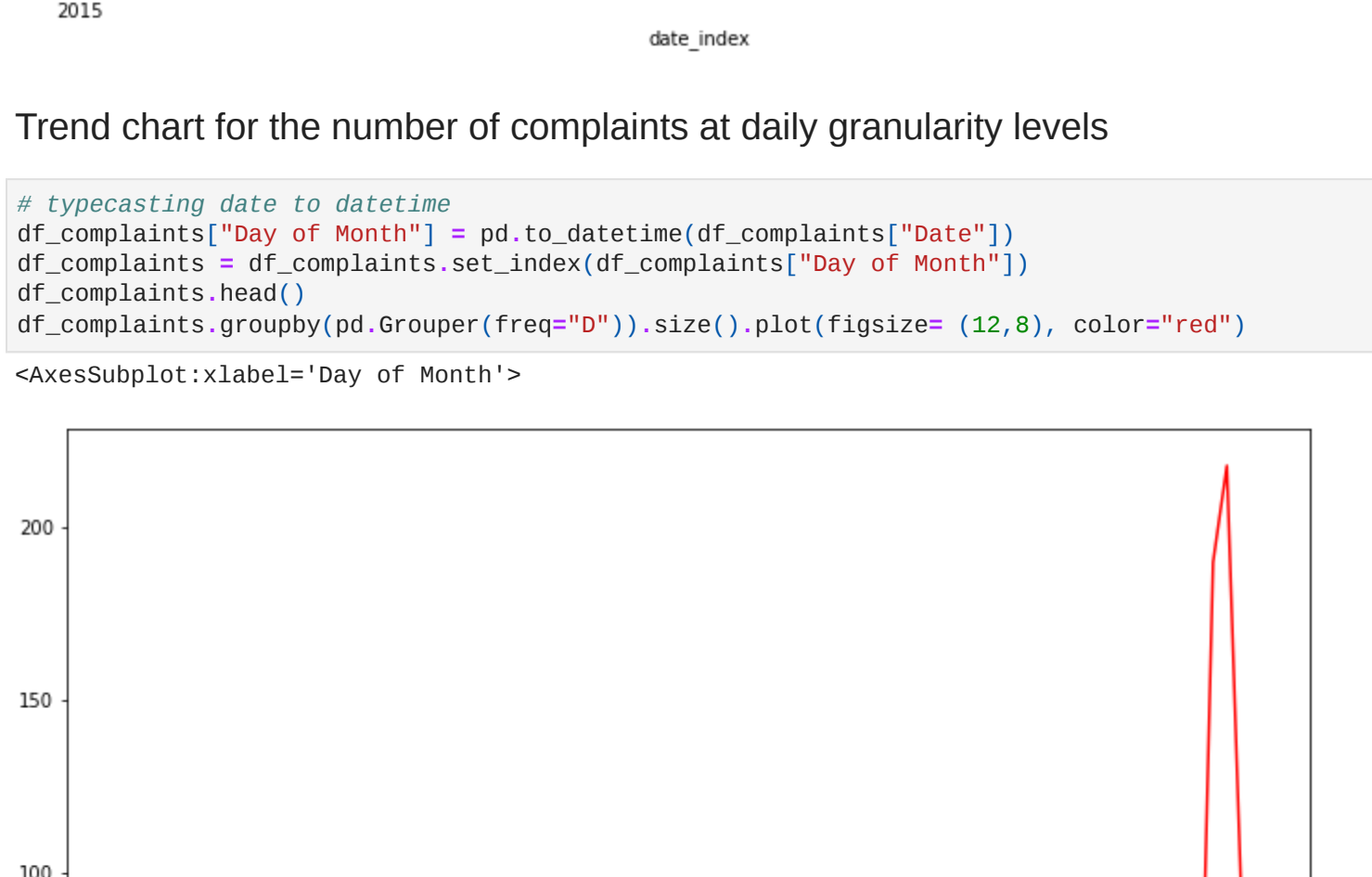
Out[5]: <AxesSubplot:xlabel='date_index'>



Trend chart for the number of complaints at daily granularity levels

In [6]:
```
# typecasting date to datetime
df_complaints["Day of Month"] = pd.to_datetime(df_complaints["Date"])
df_complaints = df_complaints.set_index(df_complaints["Day of Month"])
df_complaints.head()
df_complaints.groupby(pd.Grouper(freq="D")).size().plot(figsize= (12,8), color="red")
```

Out[6]: <AxesSubplot:xlabel='Day of Month'>



Provide a table with the frequency of complaint types.

In [7]:
```
df_type = df_complaints["Customer Complaint"].str.lower().value_counts()
df_type.head(25)
```

Out[7]:
```
comcast                             102
comcast data cap                     30
comcast internet                     29
comcast data caps                    21
comcast billing                      18
comcast service                      15
internet speed                       15
unfair billing practices             13
data caps                            13
data cap                             12
comcast complaint                    11
comcast/xfinity                      11
comcast internet service             10
billing                               9
billing issues                        8
comcast cable                         5
internet                              5
comcast billing complaint             5
comcast issues                        5
comcast billing practices             5
service issues                        5
slow internet                         5
internet service                      5
complaint against comcast             5
comcast unfair billing practices      4
Name: Customer Complaint, dtype: int64
```

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

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

Open & Pending is to be categorized as Open and Closed & Solved is to be categorized as Closed

In [8]:
```
df_complaints["newstatus"] =["Open" if Status=="Open" else "Closed" for Status in df_complaints["Status"]]
```

In [9]:
```
df_status= df_complaints.groupby('State').newstatus.value_counts().unstack()
df_status.head(25)
```

Out[9]:

newstatus	Closed	Open
State		
Alabama	22.0	4.0
Arizona	18.0	2.0
Arkansas	6.0	NaN
California	173.0	47.0
Colorado	68.0	12.0
Connecticut	10.0	2.0
Delaware	9.0	3.0
District Of Columbia	15.0	1.0
District of Columbia	1.0	NaN
Florida	205.0	35.0
Georgia	253.0	35.0
Illinois	141.0	23.0
Indiana	51.0	8.0
Iowa	1.0	NaN
Kansas	2.0	NaN
Kentucky	7.0	NaN
Louisiana	13.0	NaN
Maine	3.0	2.0
Maryland	65.0	13.0
Massachusetts	51.0	10.0
Michigan	97.0	18.0
Minnesota	31.0	2.0
Mississippi	30.0	9.0
Missouri	3.0	1.0
Montana	1.0	NaN

Provide state wise status of complaints in a stacked bar chart. Use the categorized variable from Q3

In [10]:
```
plt.figure(figsize=(15,7))
df_status.plot(kind = 'bar' ,stacked =True)
plt.rcParams['figure.dpi'] = 100
```

<Figure size 1080x504 with 0 Axes>

Georgia has maximum complaints

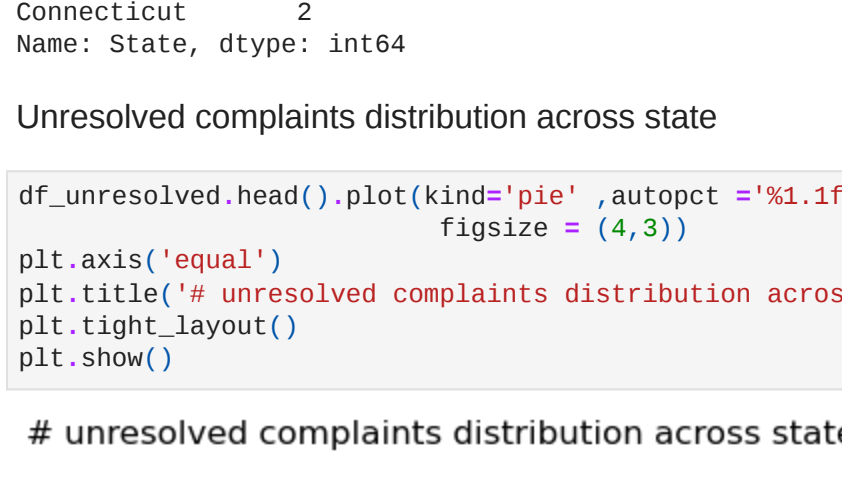
Which state has the maximum complaints

In [11]:
```
df_unresolved = df_complaints[df_complaints['newstatus']=='Open']
df_unresolved
colors = ['#639ace' , '#ca6b39' , '#7f67ca' ,'#5ba85f', '#c360aa' ,'#a7993f' ,'#cc566a']
df_unresolved = df_unresolved['State'].value_counts()
df_unresolved.head(25)
```

Out[11]:
```
California       47
Georgia          35
Florida          35
Tennessee        33
Illinois         23
Michigan         18
Washington       18
Texas            18
New Jersey       17
Pennsylvania     14
Maryland         13
Colorado         12
Oregon           10
Massachusetts    10
Mississippi       9
Virginia          8
Indiana           8
Utah              6
Alabama           4
New Mexico        3
New Hampshire     3
West Virginia     3
Delaware          3
Arizona           2
Connecticut       2
Name: State, dtype: int64
```

Unresolved complaints distribution across state

In [12]:
```
df_unresolved.head().plot(kind='pie' ,autopct ='%1.1f%%',
                          figsize = (4,3))
plt.axis('equal')
plt.title('# unresolved complaints distribution across state\n')
plt.tight_layout()
plt.show()
```



In [13]:
```
df_unresolved/len(df_unresolved)
```

Out[13]:
```
California              1.516129
Georgia                 1.129032
Florida                 1.129032
Tennessee               1.064516
Illinois                0.741935
Michigan                0.580645
Washington              0.580645
Texas                   0.580645
New Jersey              0.548387
Pennsylvania            0.451613
Maryland                0.419355
Colorado                0.387097
Oregon                  0.322581
Massachusetts           0.322581
Mississippi             0.290323
Virginia                0.258065
Indiana                 0.258065
Utah                    0.193548
Alabama                 0.129032
New Mexico              0.096774
New Hampshire           0.096774
West Virginia           0.096774
Delaware                0.096774
Arizona                 0.064516
Connecticut             0.064516
Maine                   0.064516
Minnesota               0.064516
South Carolina          0.064516
Vermont                 0.032258
Missouri                0.032258
District Of Columbia    0.032258
Name: State, dtype: float64
```

In [14]:
```
df_unresolved/df_complaints.shape[0]
```

Out[14]:
```
California              0.021133
Georgia                 0.015737
Florida                 0.015737
Tennessee               0.014838
Illinois                0.010342
Michigan                0.008094
Washington              0.008094
Texas                   0.008094
New Jersey              0.007644
Pennsylvania            0.006295
Maryland                0.005845
Colorado                0.005396
Oregon                  0.004496
Massachusetts           0.004496
Mississippi             0.004047
Virginia                0.003597
Indiana                 0.003597
Utah                    0.002698
Alabama                 0.001799
New Mexico              0.001349
New Hampshire           0.001349
West Virginia           0.001349
Delaware                0.001349
Arizona                 0.000899
Connecticut             0.000899
Maine                   0.000899
Minnesota               0.000899
South Carolina          0.000899
Vermont                 0.000450
Missouri                0.000450
District Of Columbia    0.000450
Name: State, dtype: float64
```

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

In [16]:
```
df_received = df_complaints[df_complaints['Received Via'].isin(['Internet','Customer Care Call'])]
df_received.head(25)
```

Out[16]:

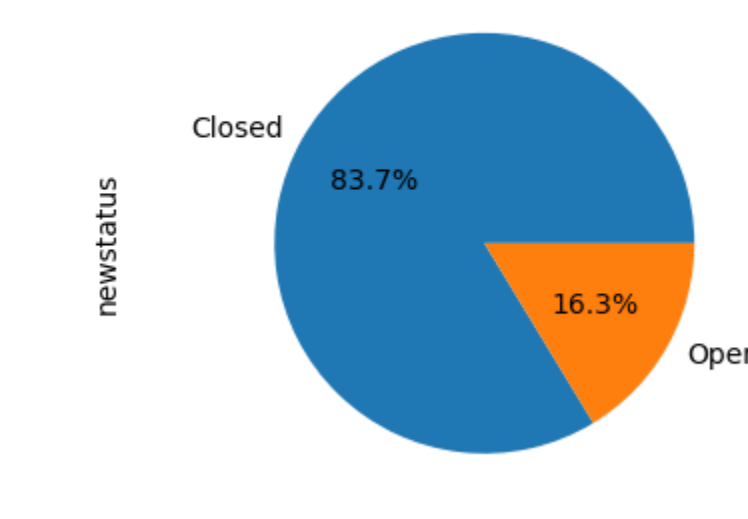
Day of Month	Ticket #	Customer Complaint	Date	Date_month_year	Time	Received Via	City	State	Zip code	Status	Filing on Behalf of Someone	date_index	Day of Month	newstatus
2015-04-22	250635	Comcast Cable Internet Speeds	22-04-15	2015-04-22	3:53:50 PM	Customer Care Call	Abingdon	Maryland	21009	Closed	No	2015-04-22 15:53:50	2015-04-22	Closed
2015-04-08	223441	Payment disappear - service got disconnected	04-08-15	2015-08-04	10:22:56 AM	Internet	Acworth	Georgia	30102	Closed	No	2015-08-04 10:22:56	2015-04-08	Closed
2015-04-18	242732	Speed and Service	18-04-15	2015-04-18	9:55:47 AM	Internet	Acworth	Georgia	30101	Closed	Yes	2015-04-18 09:55:47	2015-04-18	Closed
2015-05-07	277946	Comcast Imposed a New Usage Cap of 300GB that ...	05-07-15	2015-07-05	11:59:35 AM	Internet	Acworth	Georgia	30101	Open	Yes	2015-07-05 11:59:35	2015-05-07	Open
2015-05-26	307175	Comcast not working and no service to boot	26-05-15	2015-05-26	1:25:26 PM	Internet	Acworth	Georgia	30101	Solved	No	2015-05-26 13:25:26	2015-05-26	Closed
2015-06-12	338519	ISP Charging for arbitrary data limits with ov...	06-12-15	2015-12-06	9:59:40 PM	Internet	Acworth	Georgia	30101	Solved	No	2015-12-06 21:59:40	2015-06-12	Closed
2015-06-24	361148	Throttling service and unreasonable data caps	24-06-15	2015-06-24	10:13:55 AM	Customer Care Call	Acworth	Georgia	30101	Pending	No	2015-06-24 10:13:55	2015-06-24	Closed
2015-06-23	359792	Comcast refuses to help troubleshoot and corre...	23-06-15	2015-06-23	6:56:14 PM	Internet	Adrian	Michigan	49221	Solved	No	2015-06-23 18:56:14	2015-06-23	Closed
2015-06-01	318072	Comcast extended outages	06-01-15	2015-01-06	11:46:30 PM	Customer Care Call	Alameda	California	94502	Closed	No	2015-01-06 23:46:30	2015-06-01	Closed
2015-06-28	371214	Comcast Raising Prices and Not Being Available...	28-06-15	2015-06-28	6:46:31 PM	Customer Care Call	Alameda	California	94501	Open	Yes	2015-06-28 18:46:31	2015-06-28	Open
2015-04-24	255938	Billing after service was asked to be disconne...	24-04-15	2015-04-24	4:40:36 PM	Internet	Albuquerque	New Mexico	87106	Closed	No	2015-04-24 16:40:36	2015-04-24	Closed
2015-05-06	276409	YAHOO FAILURE TO RESTORE EMAIL SEARCH FEATURE	05-06-15	2015-06-05	3:09:49 PM	Customer Care Call	Albuquerque	New Mexico	87109	Closed	No	2015-06-05 15:09:49	2015-05-06	Closed
2015-06-13	339282	Comcast Violating Open Internet Rules by Block...	13-06-15	2015-06-13	4:03:18 PM	Internet	Albuquerque	New Mexico	87105	Open	Yes	2015-06-13 16:03:18	2015-06-13	Open
2015-06-23	360178	Internet speed	23-06-15	2015-06-23	9:23:23 PM	Internet	Albuquerque	New Mexico	87113	Solved	No	2015-06-23 21:23:23	2015-06-23	Closed
2015-06-30	376268	Internet Disconnects Every Night	30-06-15	2015-06-30	10:30:02 PM	Customer Care Call	Albuquerque	New Mexico	87116	Solved	No	2015-06-30 22:30:02	2015-06-30	Closed
2015-06-27	370137	Internet complaint	27-06-15	2015-06-27	3:25:03 PM	Customer Care Call	Albuquerqur	New Mexico	87102	Pending	No	2015-06-27 15:25:03	2015-06-27	Closed
2015-06-24	363695	Internet Availability and Speed	24-06-15	2015-06-24	11:47:33 PM	Customer Care Call	Alexandria	Indiana	46001	Solved	No	2015-06-24 23:47:33	2015-06-24	Closed
2015-04-16	238694	Comcast owes me $65 and claims I need to retur...	16-04-15	2015-04-16	10:04:57 AM	Internet	Alexandria	Virginia	22304	Closed	No	2015-04-16 10:04:57	2015-04-16	Closed
2015-04-11	230876	Horrible Internet Service	04-11-15	2015-11-04	7:48:05 PM	Customer Care Call	Alexandria	Virginia	22305	Closed	No	2015-11-04 19:48:05	2015-04-11	Closed
2015-06-02	318725	Failure to provide services that I am billed for.	06-02-15	2015-02-06	1:03:52 PM	Customer Care Call	Alexandria	Virginia	22314	Closed	No	2015-02-06 13:03:52	2015-06-02	Closed
2015-06-07	327657	Internet out all the time but they have a mono...	06-07-15	2015-07-06	8:55:43 PM	Customer Care Call	Alexandria	Virginia	22305	Solved	No	2015-07-06 20:55:43	2015-06-07	Closed
2015-06-08	328742	horrible cable service and customer service	06-08-15	2015-08-06	3:18:58 PM	Internet	Alexandria	Virginia	22312	Solved	No	2015-08-06 15:18:58	2015-06-08	Closed
2015-06-08	328165	Speed	06-08-15	2015-08-06	12:03:37 PM	Customer Care Call	Alexandria	Virginia	22304	Solved	No	2015-08-06 12:03:37	2015-06-08	Closed
2015-06-27	370538	Comcast monopoly bundling practices	27-06-15	2015-06-27	9:04:34 PM	Internet	Alexandria	Virginia	22304	Open	No	2015-06-27 21:04:34	2015-06-27	Open
2015-06-27	370363	COMCAST!	27-06-15	2015-06-27	6:37:29 PM	Internet	Alexandria	Virginia	22305	Solved	No	2015-06-27 18:37:29	2015-06-27	Closed

In [17]:
```
df_received.newstatus.value_counts()
```

Out[17]:
```
Closed    1861
Open       363
Name: newstatus, dtype: int64
```

complaints status through Internet & Customer Care

In [18]:
```
df_received.newstatus.value_counts().plot(kind='pie' ,autopct ='%1.1f%%',
                          figsize = (4,3))
plt.axis('equal')
plt.title('# complaints status through Internet & Customer Care\n')
plt.tight_layout()
plt.show()
```



In [19]:
```
df_received_closed = df_received[df_received['newstatus']=='closed']
```

In [20]:
```
df_received_closed.newstatus.value_counts()
```

Out[20]:
```
Series([], Name: newstatus, dtype: int64)
```