Project 4 Comcast Telecom Consumer Complaints

In [2]:
```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

Import data into Python environment

In [3]:
```python
df_comp = pd.read_csv("Comcast_telecom_complaints_data.csv")
```

In [4]:
```python
df_comp.head()
```

Out[4]:

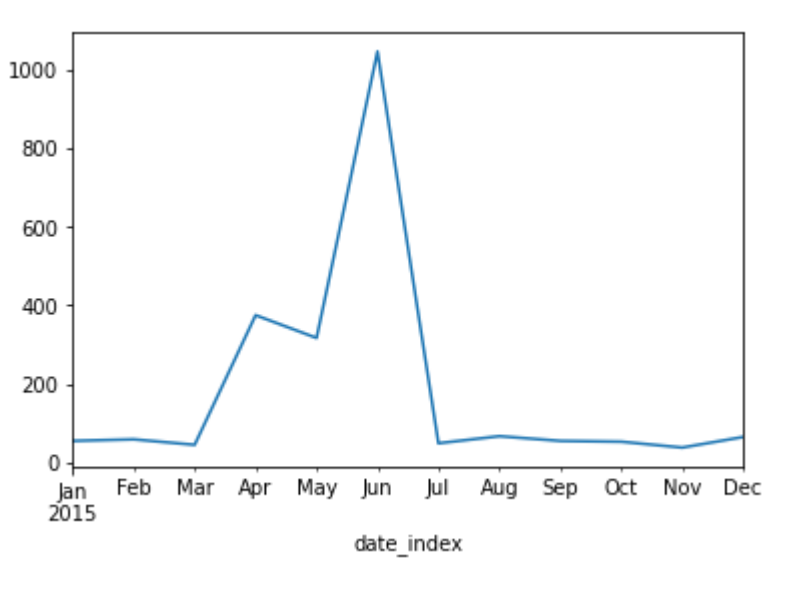
	Ticket #	Customer Complaint	Date	Date_month_year	Time	Received Via	City	State	Zip code	Status	Filing on Behalf of Someone
0	250635	Comcast Cable Internet Speeds	22-04-15	22-Apr-15	3:53:50 PM	Customer Care Call	Abingdon	Maryland	21009	Closed	No
1	223441	Payment disappear - service got disconnected	04-08-15	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 [11]:
```python
df_comp["date_index"] = df_comp["Date_month_year"].astype(str) + " " + df_comp["Time"]
df_comp["date_index"] = pd.to_datetime(df_comp["date_index"])
df_comp["Date_month_year"] = pd.to_datetime(df_comp["Date_month_year"])
df_comp = df_comp.set_index(df_comp["date_index"])
```

Trend chart for the number of complaints at monthly granularity levels

In [13]:
```python
df_comp.groupby(pd.Grouper(freq="M")).size().plot()
```

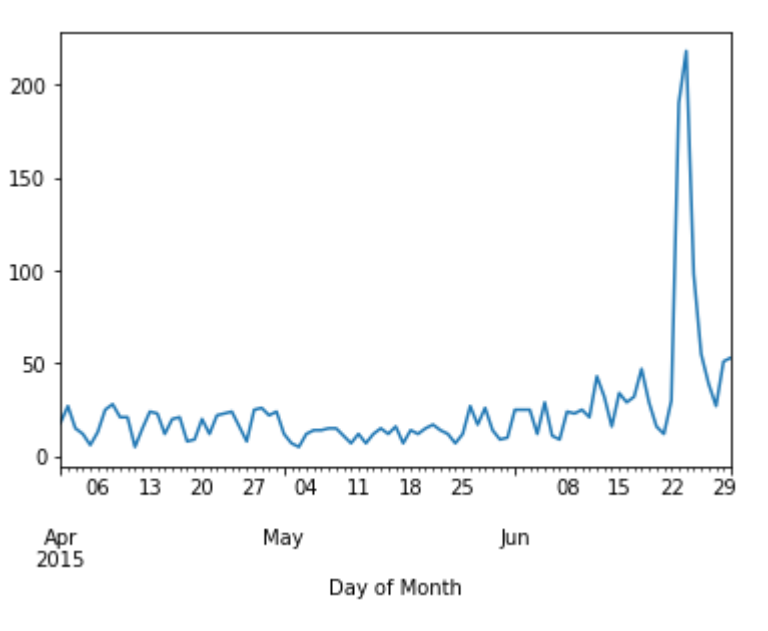
Out[13]: <AxesSubplot:xlabel='date_index'>



Trend chart for the number of complaints at daily granularity levels

In [14]:
```python
df_comp['Day of Month'] = pd.to_datetime(df_comp["Date"])
df_comp = df_comp.set_index(df_comp["Day of Month"])
df_comp.groupby(pd.Grouper(freq="D")).size().plot()
```

Out[14]: <AxesSubplot:xlabel='Day of Month'>



Provide a table with the frequency of complaint types

In [15]:
```python
df_type = df_comp["Customer Complaint"].value_counts()
```

In [32]:
```python
df_type.head(30)
```

Out[32]:
```
Comcast                             83
Comcast Internet                    18
Comcast Data Cap                    17
comcast                             13
Data Caps                           11
Comcast Billing                     11
Comcast Data Caps                   11
Unfair Billing Practices             9
Comcast data cap                     8
Comcast data caps                    8
Comcast internet                     8
Internet speed                       8
Data Cap                             8
Comcast/Xfinity                      8
COMCAST                              6
Comcast service                      6
Comcast Service                      6
Comcast billing                      6
Billing                              6
Comcast complaint                    5
Internet Speed                       5
Comcast Internet Service             5
Comcast Complaint                    5
Data cap                             4
Comcast Billing and Service Issues   4
availabilty                          4
Comcast Unfair Billing Practices     4
Billing Dispute                      4
Comcast Issues                       4
Slow Internet                        4
Name: Customer Complaint, dtype: int64
```

Python is case-sensitive it is treating Comcast and comcast as two different complaints.If all complaints are changed to upper case then it will give a correct count

In [25]:
```python
df_type = df_complaints['Customer Complaint'].str.upper().value_counts()
```

In [33]:
```python
df_type.head(30)
```

Out[33]:
```
Comcast                             83
Comcast Internet                    18
Comcast Data Cap                    17
comcast                             13
Data Caps                           11
Comcast Billing                     11
Comcast Data Caps                   11
Unfair Billing Practices             9
Comcast data cap                     8
Comcast data caps                    8
Comcast internet                     8
Internet speed                       8
Data Cap                             8
Comcast/Xfinity                      8
COMCAST                              6
Comcast service                      6
Comcast Service                      6
Comcast billing                      6
Billing                              6
Comcast complaint                    5
Internet Speed                       5
Comcast Internet Service             5
Comcast Complaint                    5
Data cap                             4
Comcast Billing and Service Issues   4
availabilty                          4
Comcast Unfair Billing Practices     4
Billing Dispute                      4
Comcast Issues                       4
Slow Internet                        4
Name: Customer Complaint, dtype: int64
```

Complaint types are maximum around Comcast , Comcast data Cap , Comcast Internet , Comcast data Cap , Comcast Billing

Create a new categorical variable with value as Open and Closed.

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

In [20]:
```python
df_comp["newStatus"] = ["Open" if Status=="Open" or Status=="Pending"
                        else "Closed" for Status in df_comp["Status"]]
```

In [29]:
```python
df_status = df_comp.groupby('State').newStatus.value_counts().unstack()
```

In [31]:
```python
df_status.head(30)
```

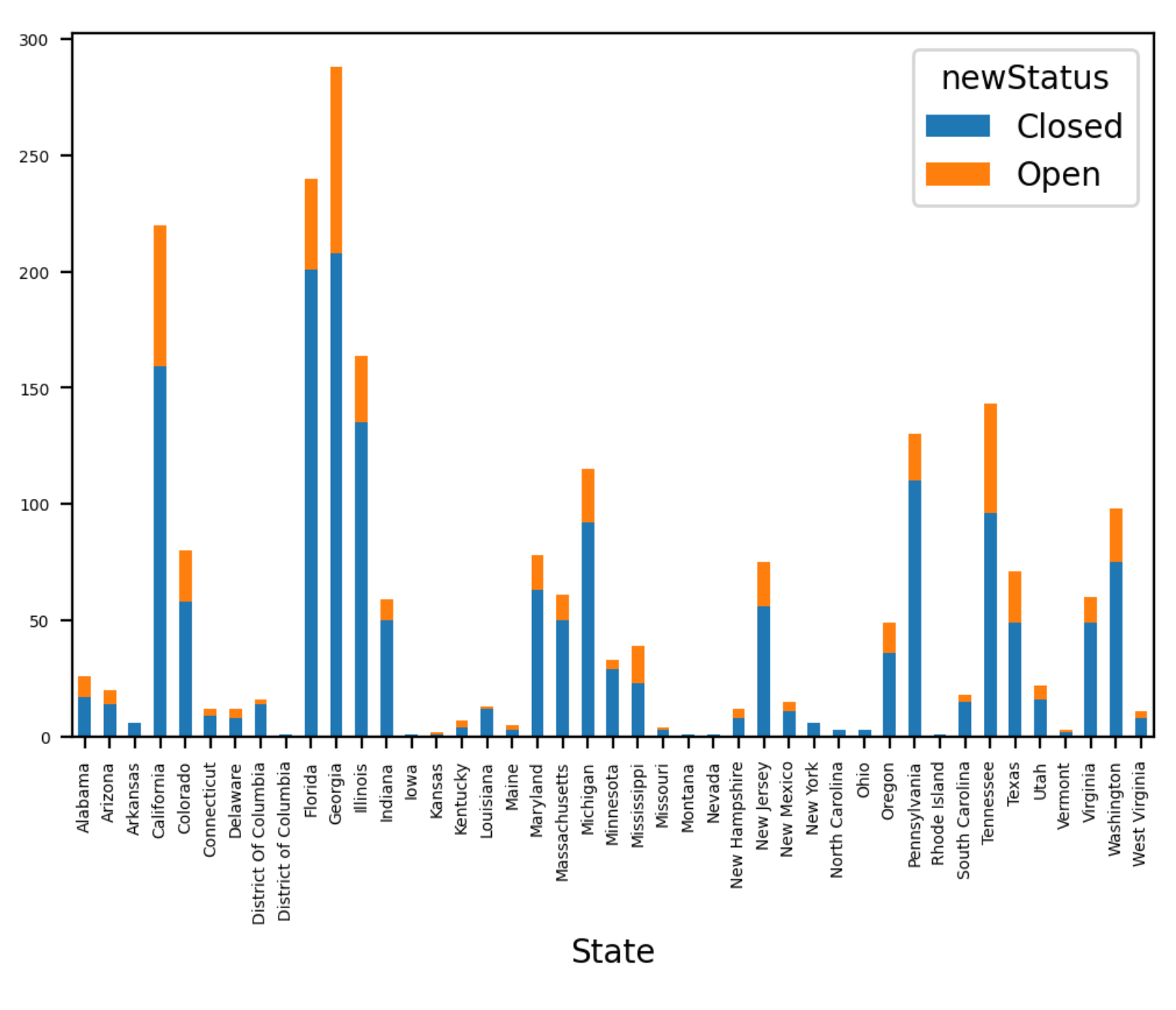
Out[31]:

newStatus	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

In [73]:
```python
plt.figure(figsize=(200,100))
plt.rcParams['figure.dpi'] = 200
df_status.plot(kind='bar', rot = 90, fontsize = 5, stacked=True)
```

Out[73]: <AxesSubplot:xlabel='State'>

<Figure size 40000x20000 with 0 Axes>



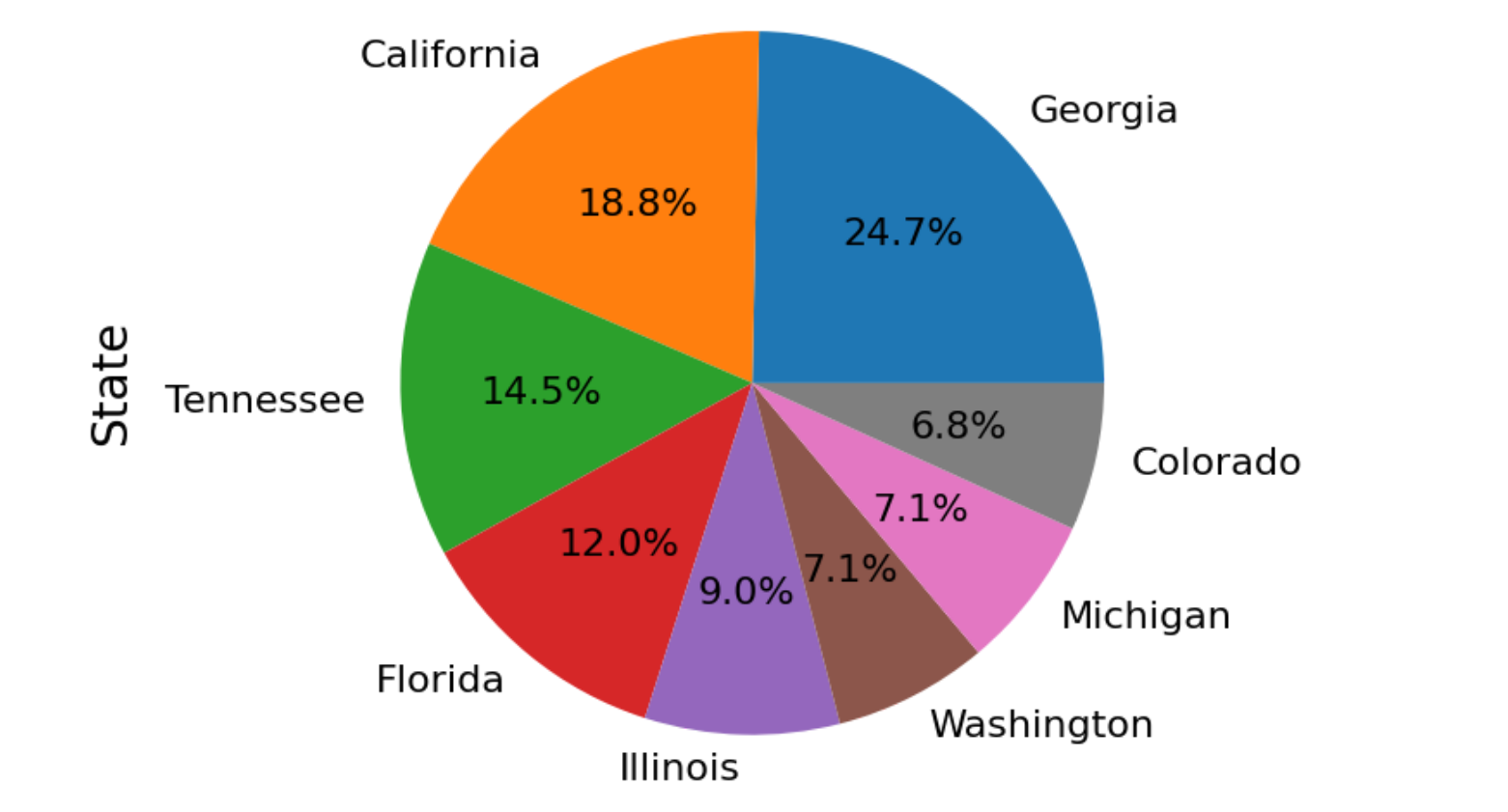
newStatus
Closed
Open
300
250
200
150
100
50
0
Alabama
Arizona
Arkansas
California
Colorado
Connecticut
Delaware
District Of Columbia
District of Columbia
Florida
Georgia
Illinois
Indiana
Iowa
Kansas
Kentucky
Louisiana
Maine
Maryland
Massachusetts
Michigan
Minnesota
Mississippi
Missouri
Montana
Nevada
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

Georgia has maximum number of complaints

In [62]:
```python
# Unresolved complaints distribution across State
df_unresolved = df_comp[df_comp['newStatus']=='Open']
colors = ['#639ace','#ca6b39','#7f67ca','#5ba85f','#c360aa','#a7993f','#cc566a']
df_unresolved = df_unresolved['State'].value_counts()
df_unresolved.head(30)
```

Out[62]:
```
Georgia                 80
California              61
Tennessee               47
Florida                 39
Illinois                29
Washington              23
Michigan                23
Colorado                22
Texas                   22
Pennsylvania            20
New Jersey              19
Mississippi             16
Maryland                15
Oregon                  13
Massachusetts           11
Virginia                11
Alabama                  9
Indiana                  9
Utah                     6
Arizona                  6
Delaware                 4
New Mexico               4
Minnesota                4
New Hampshire            4
West Virginia            3
Connecticut              3
Kentucky                 3
South Carolina           3
District Of Columbia     2
Maine                    2
Name: State, dtype: int64
```

In [86]:
```python
df_unresolved.head(8).plot(kind='pie',autopct='%1.1f%%',
                           fontsize = 8, figsize = (4,3))
plt.axis('equal')
plt.title('# Unresolved complaints distribution across State\n')
plt.tight_layout()
plt.show()
```



Georgia has maximum percentage of unresolved complaints

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

In [81]:
```python
df_received = df_comp[df_comp['Received Via'].isin(['Internet','Customer Care Call'])]
```

In [82]:
```python
df_received.head()
```

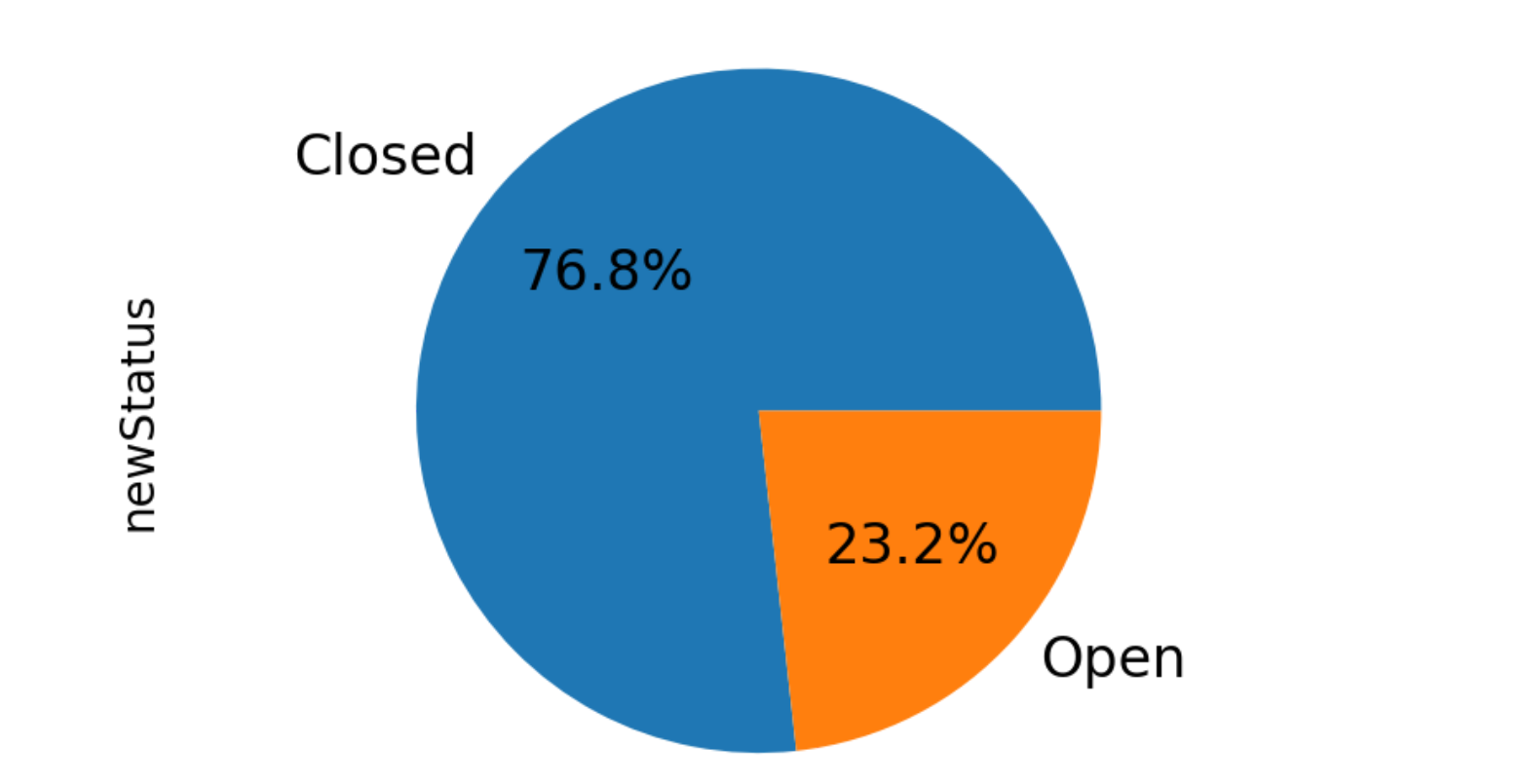
Out[82]:

	Ticket #	Customer Complaint	Date	Date_month_year	Time	Received Via	City	State	Zip code	Status	Filing on Behalf of Someone	date_in
Day of Month												
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	2015-0 15:5
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	2015-0 10:2
2015-04-18	242732	Speed and Service	18-04-15	2015-04-18	9:55:47 AM	Internet	Acworth	Georgia	30101	Closed	Yes	2015-0 09:5
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	2015-0 11:5
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	2015-0 13:2

In [83]:
```python
df_received.newStatus.value_counts()
```

Out[83]:
```
Closed    1707
Open       517
Name: newStatus, dtype: int64
```

In [91]:
```python
df_received.newStatus.value_counts().plot(kind='pie',autopct='%1.1f%%',
                           fontsize = 12, figsize = (4,3))
plt.axis('equal')
plt.title('Complaints Status through Internet & Customer Care\n')
plt.tight_layout()
plt.show()
```



In [92]:
```python
df_received_closed = df_received[df_received['newStatus']=='Closed']
```

In [93]:
```python
df_received_closed.newStatus.value_counts()
```

Out[93]:
```
Closed    1707
Name: newStatus, dtype: int64
```