BASIC OUTPUTS OF THE PROJECT

```
In [42]: df_complaints = pd.read_csv("Comcast_telecom_complaints_data.csv")
```

```
In [43]: df_complaints.head()
```

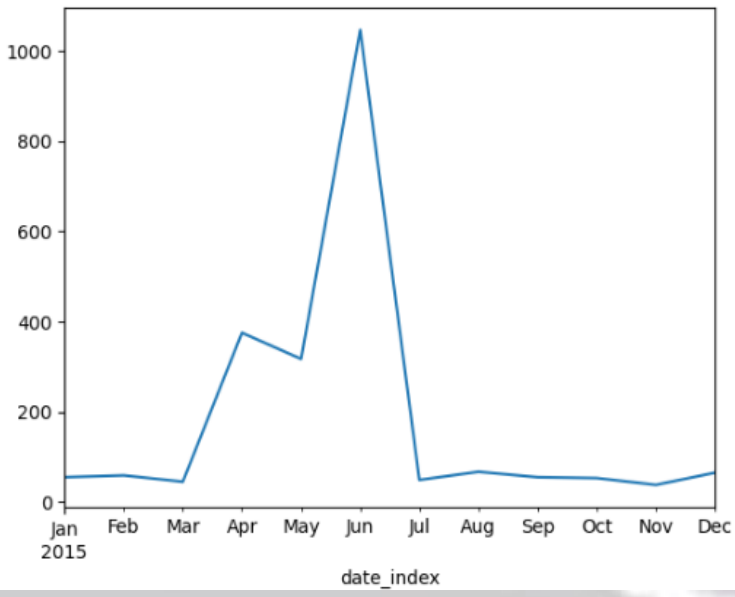
Out[43]:

	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

Trend chart for the number of complaints at monthly granularity levels

```
[45]: df_complaints.groupby(pd.Grouper(freq="M")).size().plot()
```

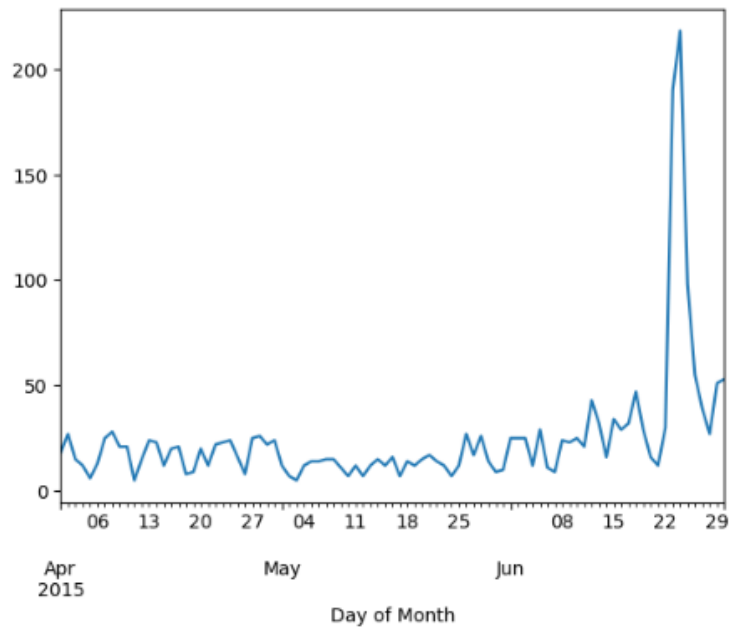
```
[45]: <AxesSubplot:xlabel='date_index'>
```



Trend chart for the number of complaints at daily granularity levels

[46]:
```
df_complaints['Day of Month'] = pd.to_datetime(df_complaints["Date"])
df_complaints = df_complaints.set_index(df_complaints["Day of Month"])
df_complaints.groupby(pd.Grouper(freq="D")).size().plot()
```

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



Provide a table with the frequency of complaint types

In [47]:
```
df_type = df_complaints["Customer Complaint"].value_counts()
```

In [48]:
```
df_type.head(25)
```

Out[48]:
```
Comcast                      83
Comcast Internet             18
Comcast Data Cap             17
comcast                      13
Comcast Billing              11
Data Caps                    11
Comcast Data Caps            11
Unfair Billing Practices      9
Comcast data cap              8
Comcast internet              8
Internet speed                8
Comcast data caps             8
Comcast/Xfinity               8
Data Cap                      8
Billing                       6
Comcast Service               6
Comcast billing               6
COMCAST                       6
Comcast service               6
Comcast Internet Service      5
Comcast complaint             5
Comcast Complaint             5
Internet Speed                5
availabilty                   4
Comcast Issues                4
Name: Customer Complaint, dtype: int64
```

```
In [49]: df_type = df_complaints['Customer Complaint'].str.upper().value_counts()

In [50]: df_type.head(25)
```

```
Out[50]: 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 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.

```
[62]: df_complaints["newStatus"] = ["Open" if Status=="Open" or Status=="Pending" else "Closed" for Status in df_complaints["Stat...
```

```
[63]: df_status = df_complaints.groupby('State').newStatus.value_counts().unstack()
      #df_status = df_complaints.groupby(['State','newStatus'])['CDUNT'].sum().unstack()
```

```
[64]: df_status.head(25)
```

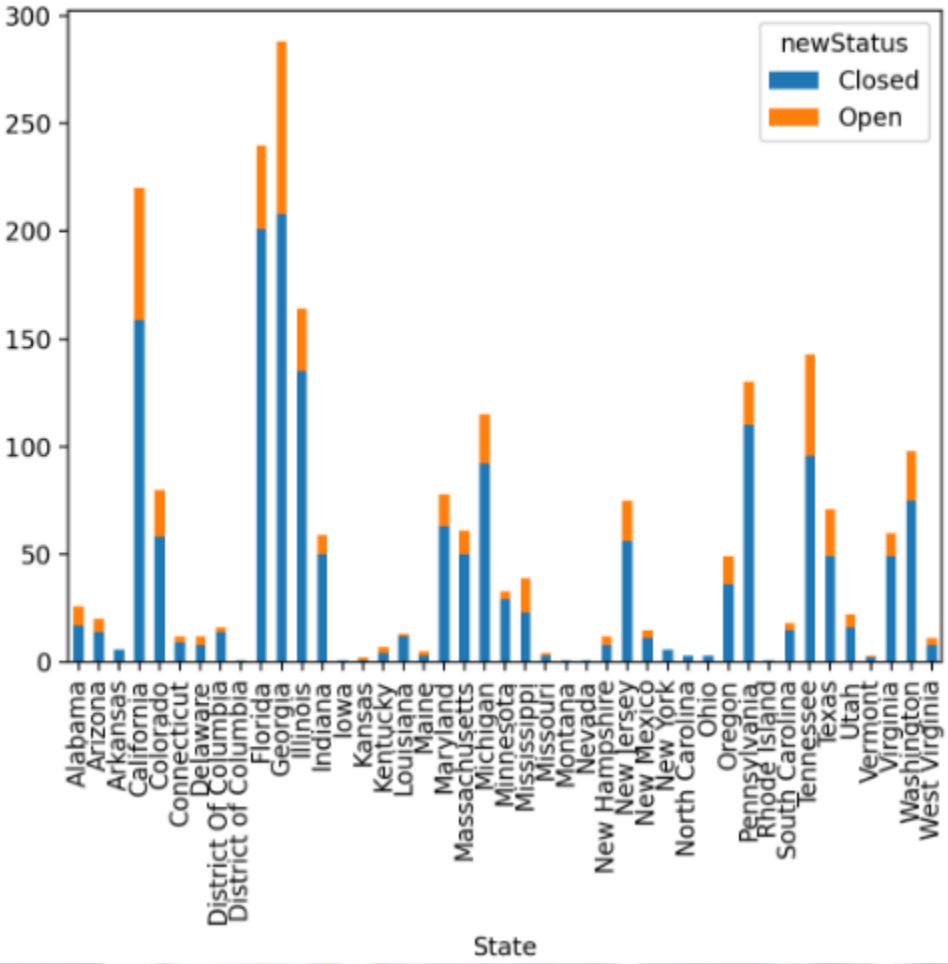
[64]:

newStatus	Closed	Open
State		
Alabama	17.0	3.0
Arizona	14.0	8.0
Arkansas	8.0	NaN
California	159.0	81.0
Colorado	58.0	22.0
Connecticut	9.0	9.0
Delaware	3.0	4.0
District Of Columbia	14.0	2.0
District of Columbia	1.0	NaN
Florida	201.0	99.0
Georgia	203.0	80.0
Illinois	125.0	29.0
Indiana	50.0	3.0
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	89.0	15.0
Massachusetts	50.0	11.0
Michigan	92.0	29.0
Minnesota	29.0	4.0
Mississippi	29.0	18.0
Missouri	3.0	1.0
Montana	1.0	NaN

```
[33]: plt.figure(figsize=(200,100))
      plt.rcParams['figure.dpi'] = 200
      df_states.plot(kind='bar', stacked=True)

[33]: <AxesSubplot:xlabel='State'>

      <Figure size 20000x10000 with 0 Axes>
```

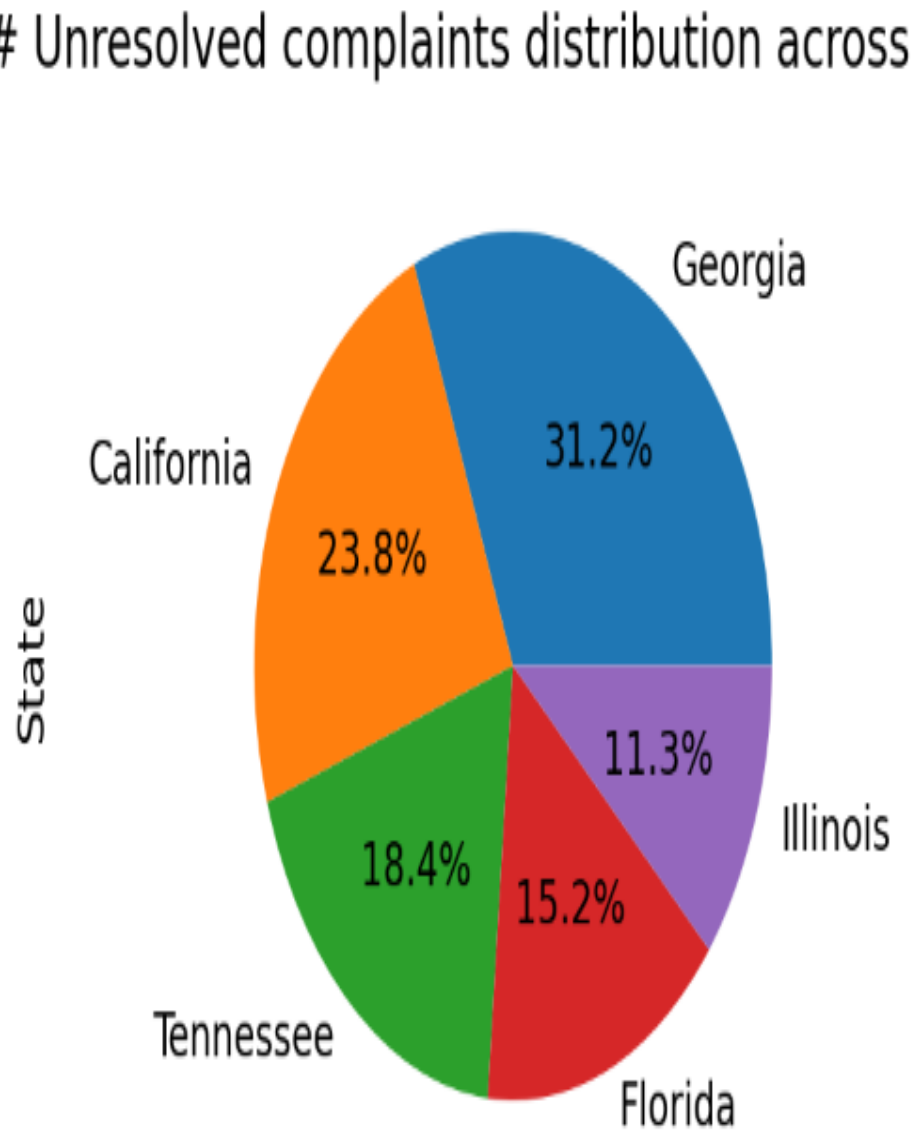


Georgia has maximum number of complaints

[56]:
```
# Unresolved complaints distribution across State
df_unresolved = df_complaints[df_complaints['newStatus']=='Open']
colors = ['#639ace','#ca6b39','#7f67ca','#5ba85f','#c360aa','#a7993f','#cc566a']
df_unresolved = df_unresolved['State'].value_counts()
df_unresolved.head(25)
```

[56]:
```
Georgia         80
California      61
Tennessee       47
Florida         39
Illinois        29
Michigan        23
Washington      23
Texas           22
Colorado        22
Pennsylvania    20
New Jersey      19
Mississippi     16
Maryland        15
Oregon          13
Massachusetts   11
Virginia        11
Indiana          9
Alabama          9
Arizona          6
Utah             6
New Hampshire    4
Delaware         4
Minnesota        4
New Mexico       4
Connecticut      3
Name: State, dtype: int64
```

In [57]:
```
df_unresolved.head().plot(kind='pie',autopct='%1.1f%%',
                          #explode = (0.15, 0, 0, 0, 0), startangle=45, shadow=False, colors = colors,
                          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 [58]:
```
df_received = df_complaints[df_complaints['Received Via'].isin(['Internet','Customer Care Call'])]
```

In [59]:
```
df_received.head()
```

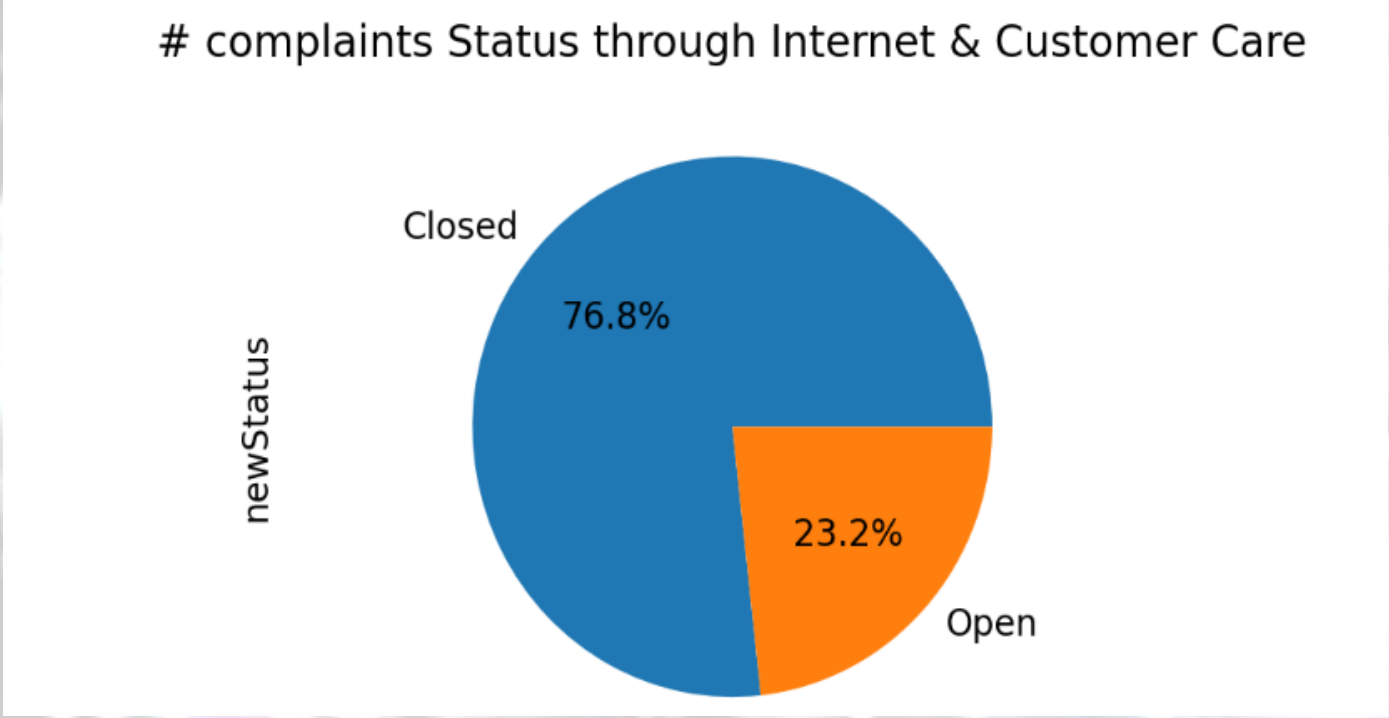
Out[59]:

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

In [60]:
```
df_received.newStatus.value_counts()
```

Out[60]:
```
Closed    1707
Open       517
Name: newStatus, dtype: int64
```

In [61]:
```
df_received.newStatus.value_counts().plot(kind='pie',autopct='%1.1f%%',
                        #explode = (0.15, 0, 0, 0, 0), startangle=45, shadow=False, colors = colors,
                        figsize = (4,3))
plt.axis('equal')
plt.title('# complaints Status through Internet & Customer Care\n')
plt.tight_layout()
plt.show()
```



In [62]:
```
df_received_closed = df_received[df_received['newStatus']=='Closed']
```

In [63]:
```
df_received_closed.newStatus.value_counts()
```

Out[63]:
```
Closed    1707
Name: newStatus, dtype: int64
```

Complaint types based on different location with average after grouping them 'Request_Closing_Time'

In [21]:
```
brGrpComplaintType = NYC.groupby('Complaint Type')
```

In [22]:
```
brGrpData = brGrpComplaintType.get_group('Blocked Driveway')
brGrpData.shape
```

Out[22]: (198831, 54)

In [23]:
```
NYC.isnull().sum()
```

Out[23]:
```
Unique Key                          0
Created Date                        0
Closed Date                      2151
Agency                              0
Agency Name                         0
Complaint Type                      0
Descriptor                       5487
Location Type                     711
Incident Zip                     2992
Incident Address                51649
Street Name                     51649
Cross Street 1                  57733
Cross Street 2                  57388
Intersection Street 1          131413
Intersection Street 2          131484
Address Type                     1257
City                             2992
Landmark                       134731
Facility Type                    2151
Status                              0
Due Date                            1
Resolution Description              0
Resolution Action Updated Date   2480
Community Board                     0
Borough                             0
X Coordinate (State Plane)       4818
Y Coordinate (State Plane)       4818
Park Facility Name                  0
Park Borough                        0
School Name                         0
School Number                       0
School Region                       1
School Code                         1
School Phone Number                 0
School Address                      0
School City                         0
School State                        0
School Zip                          1
School Not Found                    0
School or Citywide Complaint      134733
Vehicle Type                      134733
Taxi Company Borough              134733
Taxi Pick Up Location             134733
Bridge Highway Name               134760
Bridge Highway Direction          134760
Road Ramp                         134776
Bridge Highway Segment            134776
Garage Lot Name                   134733
Ferry Direction                   134777
Ferry Terminal Name               134778
Latitude                            4818
Longitude                           4818
Location                            4818
Request_Closing_Time                2151
dtype: int64
```

In [24]:
```
NYC['City'].dropna(inplace=True)
```

In [25]:
```
NYC['City'].shape
```

Out[25]: (364558,)

In [26]:
```
GroupData['City'].isnull().sum()
```

Out[26]: 333

In [27]:
```
GroupData['City'].fillna('Unknown City', inplace =True)
```

```
C:\Users\91805\AppData\Local\Temp\ipykernel_3488\2773330166.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  GroupData['City'].fillna('Unknown City', inplace =True)
```

Scatter plot displaying all the cities that raised complaint of type 'Blocked Driveway'

In [49]:
```
plt.figure(figsize=(10, 13))
plt.scatter(GroupData['Complaint Type'],GroupData['City'])
plt.title('Plot showing list of cities that raised complaint of type Blocked Driveway')
plt.show()
```