Comcast_Telecom_consumer_complaint

April 12, 2022

0.0.1 Importing the libraries

[1]:
```
import pandas as pd
```

[2]:
```
data1 = pd.read_csv('Comcast_telecom_complaints_data.csv')
data1
```

[2]:

	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			
4	307175	Comcast not working and no service to boot	26-05-15	26-May-15	1:25:26 PM	Internet	Acworth	Georgia			
…	…	…	…	…	…	…	…	…			
2219	213550	Service Availability	04-02-15	04-Feb-15	9:13:18 AM	Customer Care Call	Youngstown	Florida			
2220	318775	Comcast Monthly Billing for Returned Modem	06-02-15	06-Feb-15	1:24:39 PM	Customer Care Call	Ypsilanti	Michigan			
2221	331188	complaint about comcast	06-09-15	06-Sep-15	5:28:41 PM	Internet	Ypsilanti	Michigan			
2222	360489	Extremely unsatisfied Comcast customer	23-06-15	23-Jun-15	11:13:30 PM	Customer Care Call	Ypsilanti	Michigan			
2223	363614	Comcast, Ypsilanti MI Internet Speed	24-06-15	24-Jun-15	10:28:33 PM	Customer Care Call	Ypsilanti	Michigan			

```
3        30101    Open                          Yes
4        30101  Solved                           No
…          …     …                             …
2219     32466  Closed                           No
2220     48197  Solved                           No
2221     48197  Solved                           No
2222     48197  Solved                           No
2223     48198    Open                          Yes

[2224 rows x 11 columns]
```

[3]:
```
data1.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
```

[4]:
```
data1['Date'] = pd.to_datetime(data1['Date'])
```

[5]:
```
data1.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
```

[6]:
```
data1['Month'] = data1['Date'].dt.month_name()
```

[7]:
```
data1.head()
```

[7]:

	Ticket #	Customer Complaint	Date	Date_month_year	Time	Received Via	City	State	Zip code	Status	Filing on Behalf of Someone	Month
0	250635	Comcast Cable Internet Speeds	2015-04-22	22-Apr-15	3:53:50 PM	Customer Care Call	Abingdon	Maryland	21009	Closed	No	April
1	223441	Payment disappear - service got disconnected	2015-04-08	04-Aug-15	10:22:56 AM	Internet	Acworth	Georgia	30102	Closed	No	April
2	242732	Speed and Service	2015-04-18	18-Apr-15	9:55:47 AM	Internet	Acworth	Georgia	30101	Closed	Yes	April
3	277946	Comcast Imposed a New Usage Cap of 300GB that …	2015-05-07	05-Jul-15	11:59:35 AM	Internet	Acworth	Georgia	30101	Open	Yes	May
4	307175	Comcast not working and no service to boot	2015-05-26	26-May-15	1:25:26 PM	Internet	Acworth	Georgia	30101	Solved	No	May

[8]:
```
data1['Date'].value_counts()
```

[8]:
```
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
```

[9]:
```
dates=data1.groupby('Date').count()['Ticket #']
```

[10]:
```
daily = pd.DataFrame(dates).reset_index()
```

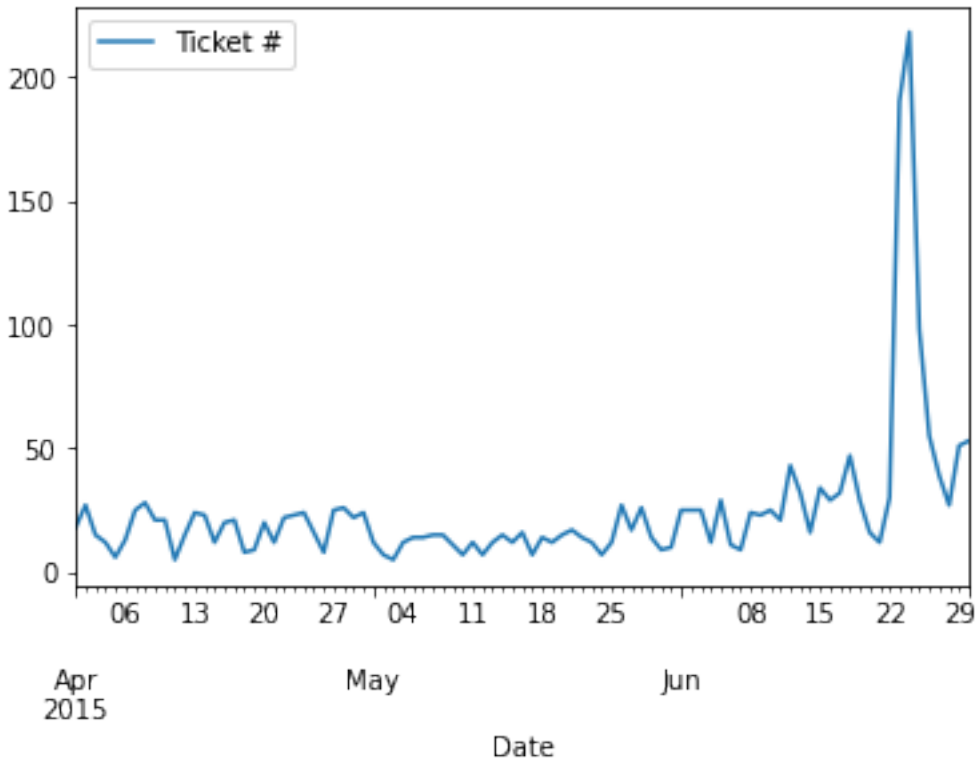
[11]:
```
daily.head()
```

[11]:

	Date	Ticket #
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

[12]:
```
daily.plot(x="Date",y="Ticket #",kind="line")
```

[12]:
```
<AxesSubplot:xlabel='Date'>
```



[13]:
```
months = data1.groupby('Month').count()['Ticket #']
```

[14]:
```
months
```

[14]:
```
Month
April     545
June     1280
May       399
Name: Ticket #, dtype: int64
```

[15]:
```
months_df = pd.DataFrame(months).reset_index()
```

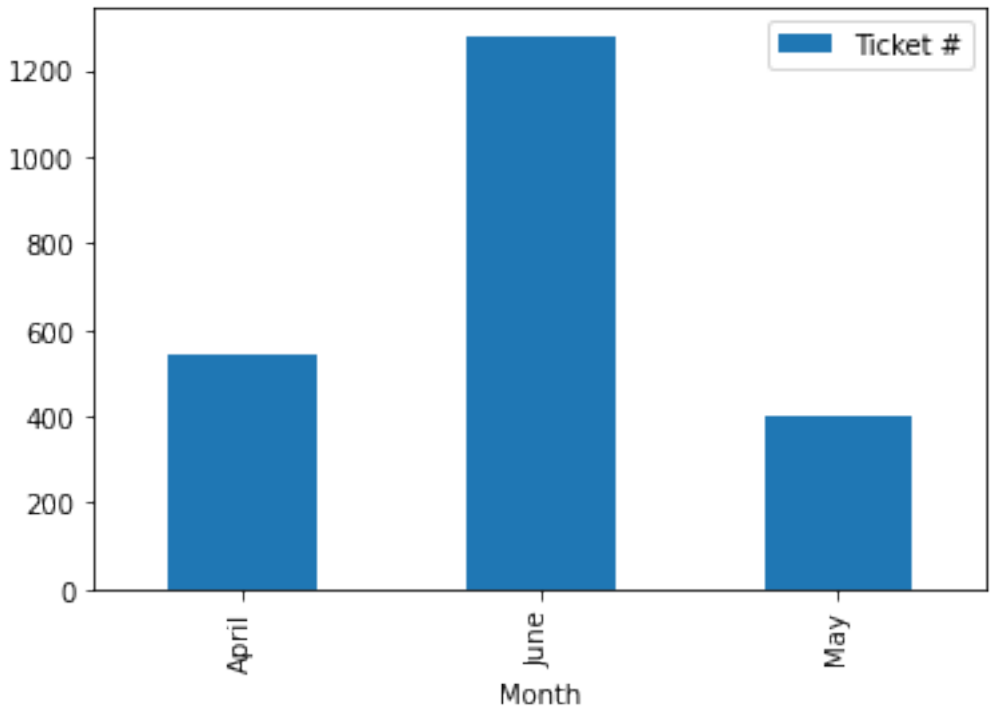
[16]:
```
months_df.head()
```

[16]:

	Month	Ticket #
0	April	545
1	June	1280
2	May	399

[17]:
```
import matplotlib.pyplot as plt
```

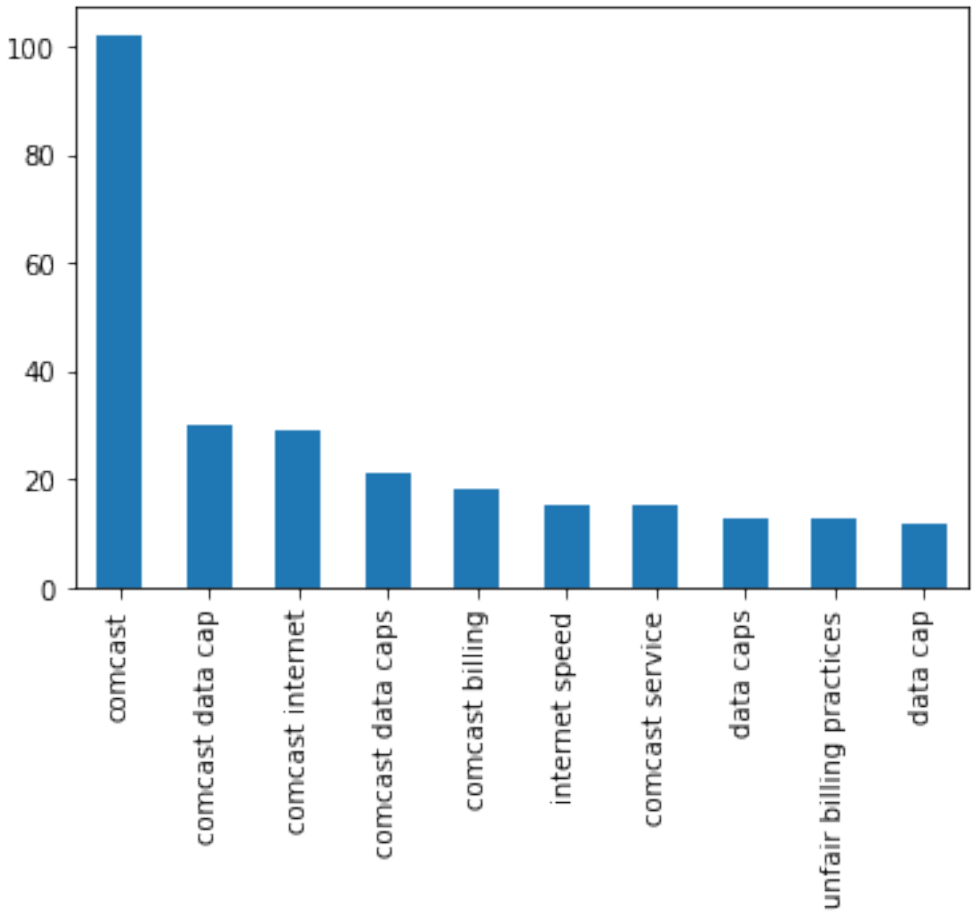
[18]:
```
months_df.plot(x="Month",y="Ticket #",kind="bar")
plt.show()
```



[19]:
```
data1['Customer Complaint'] = data1["Customer Complaint"].apply(lambda x :
↪str(x).lower())
```

[20]:
```
data1['Customer Complaint'].value_counts()[:10].plot.bar()
```

[20]: <AxesSubplot:>



[21]:
```
data1['Status'].unique()
```

[21]:
```
array(['Closed', 'Open', 'Solved', 'Pending'], dtype=object)
```

[22]:
```
data1["New_status"] =  ['Open' if st == "Open" or st=="Pending" else "Closed"␣
↪for st in data1['Status']]
```

[23]:
```
data1["New_status"]
```

[23]:
```
0        Closed
1        Closed
2        Closed
3          Open
4        Closed
          ...
2219     Closed
```

```
2220    Closed
2221    Closed
2222    Closed
2223      Open
Name: New_status, Length: 2224, dtype: object
```

[24]:
```
data1.head()
```

[24]:
```
  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

  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  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
```

[25]:
```
data1["New_status"].unique()
```

[25]:
```
array(['Closed', 'Open'], dtype=object)
```

[26]:
```
state_complaint = data1.groupby(['State', 'New_status']).size().unstack().
↪fillna(0)
```

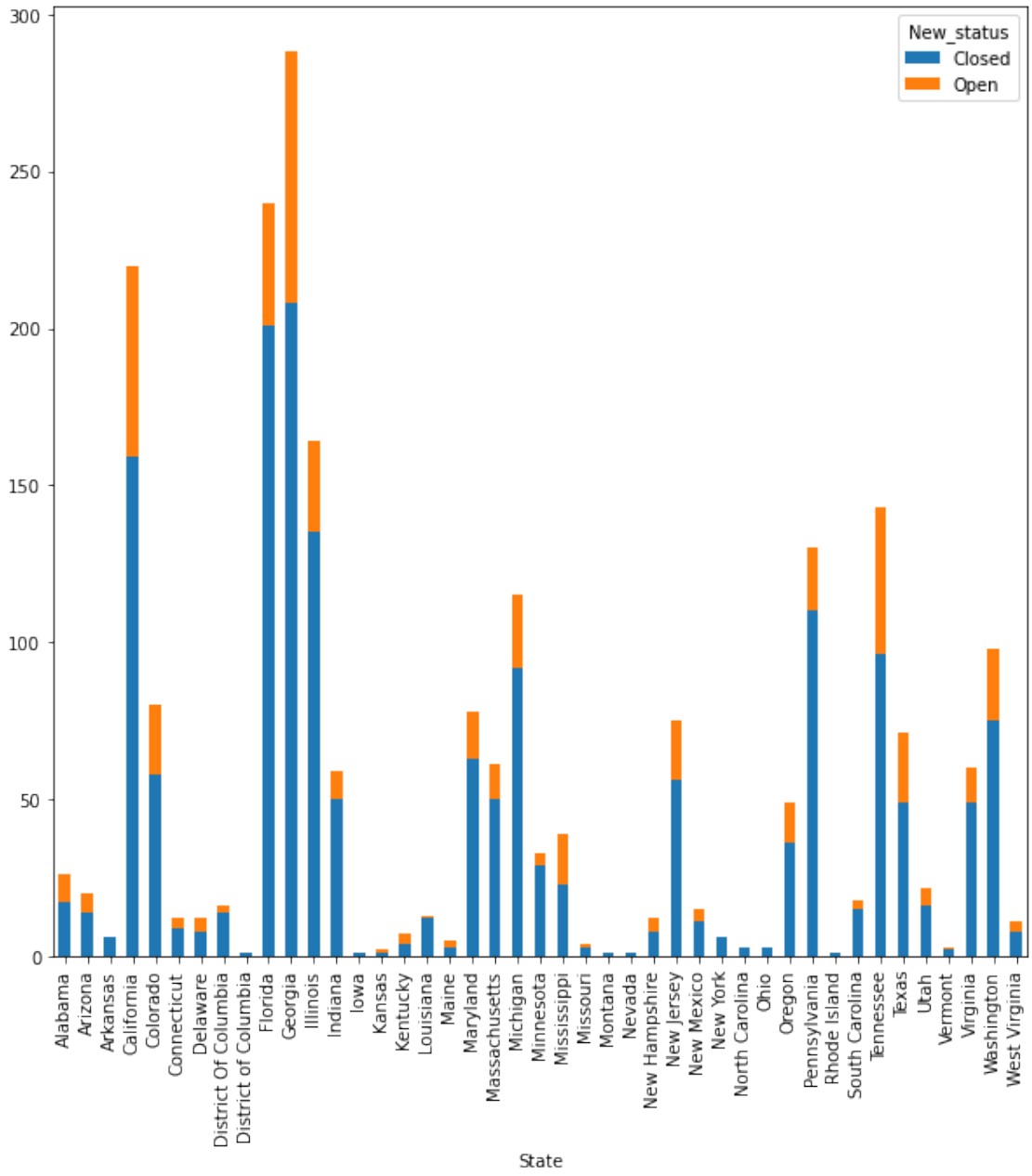
[27]:
```
state_complaint
```

[27]:
```
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
```

[28]:
```
state_complaint.plot.bar(stacked = 'True', figsize= (10,10))
```

[28]:
```
<AxesSubplot:xlabel='State'>
```



0.0.2 State having maximum complaints

[29]:
```
data1.groupby(["State"]).size().sort_values(ascending = False)[:5]
```

[29]:
```
State
Georgia       288
Florida       240
```

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

0.0.3 Finding the state having the highest percentage of unresolved complaints

[30]:
```
unresolved_complaints = data1.groupby(['State', "New_status"]).size().unstack().
↪sort_values(by = "Open", ascending = False)
```

[31]:
```
unresolved_complaints['Unresolved_cmp_predict'] = unresolved_complaints["Open"]/
↪unresolved_complaints["Open"].sum()*100
```

[32]:
```
unresolved_complaints
```

[32]:

New_status	Closed	Open	Unresolved_cmp_predict
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
Michigan	92.0	23.0	4.448743
Washington	75.0	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
Massachusetts	50.0	11.0	2.127660
Virginia	49.0	11.0	2.127660
Alabama	17.0	9.0	1.740812
Indiana	50.0	9.0	1.740812
Arizona	14.0	6.0	1.160542
Utah	16.0	6.0	1.160542
Delaware	8.0	4.0	0.773694
New Hampshire	8.0	4.0	0.773694
New Mexico	11.0	4.0	0.773694
Minnesota	29.0	4.0	0.773694
South Carolina	15.0	3.0	0.580271
Connecticut	9.0	3.0	0.580271
West Virginia	8.0	3.0	0.580271
Kentucky	4.0	3.0	0.580271
District Of Columbia	14.0	2.0	0.386847

```
Maine                    3.0   2.0              0.386847
Louisiana               12.0   1.0              0.193424
Vermont                  2.0   1.0              0.193424
Missouri                 3.0   1.0              0.193424
Kansas                   1.0   1.0              0.193424
Arkansas                 6.0   NaN                   NaN
District of Columbia     1.0   NaN                   NaN
Iowa                     1.0   NaN                   NaN
Montana                  1.0   NaN                   NaN
Nevada                   1.0   NaN                   NaN
New York                 6.0   NaN                   NaN
North Carolina           3.0   NaN                   NaN
Ohio                     3.0   NaN                   NaN
Rhode Island             1.0   NaN                   NaN
```

0.0.4 Percentage of complaints resolved till date, which were received through the internet and customer care calls

[33]:
```
resolved_complaints = data1.groupby(["Received Via", "New_status"]).size().
↪unstack()
```

[34]:
```
resolved_complaints["Resolved"] = resolved_complaints["Closed"]/
↪resolved_complaints["Closed"].sum()*100
```

[35]:
```
resolved_complaints["Resolved"]
```

[35]:
```
Received Via
Customer Care Call    50.615114
Internet              49.384886
Name: Resolved, dtype: float64
```

0.0.5 Finding the maximum complaint types

[36]:
```
complaint_types = data1["Customer Complaint"].unique()
complaint_types
```

[36]:
```
array(['comcast cable internet speeds',
       'payment disappear - service got disconnected',
       'speed and service', …,
       'comcast monthly billing for returned modem',
       'extremely unsatisfied comcast customer',
       'comcast, ypsilanti mi internet speed'], dtype=object)
```

[37]:
```
complaint_type = data1['Customer Complaint'].value_counts()
```

[38]:
```
complaint_type.head()
```

[38]:
```
comcast              102
comcast data cap      30
comcast internet      29
comcast data caps     21
comcast billing       18
Name: Customer Complaint, dtype: int64
```

[39]:
```
internet_issues1 = data1[data1["Customer Complaint"].str.contains(("network"))].
↪count()
```

[41]:
```
internet_issues2 = data1[data1["Customer Complaint"].str.contains("data")].
↪count()
```

[42]:
```
internet_issues3 = data1[data1["Customer Complaint"].str.contains("speed")].
↪count()
```

[43]:
```
internet_issues4 = data1[data1["Customer Complaint"].str.contains("internet")].
↪count()
```

[44]:
```
billing_issues1 = data1[data1["Customer Complaint"].str.contains("billing")].
↪count()
```

[46]:
```
billing_issues2 = data1[data1["Customer Complaint"].str.contains("charges")].
↪count()
```

[47]:
```
billing_issues3 = data1[data1["Customer Complaint"].str.contains("bills")].
↪count()
```

[48]:
```
service_issues1 = data1[data1["Customer Complaint"].str.contains("service")].
↪count()
```

[49]:
```
service_issues2 = data1[data1["Customer Complaint"].str.contains("customer")].
↪count()
```

[50]:
```
total_internet_issues = internet_issues1 + internet_issues2 + internet_issues3␣
↪+ internet_issues4
```

[51]:
```
total_internet_issues
```

[51]:
```
Ticket #                 945
Customer Complaint       945
Date                     945
Date_month_year          945
Time                     945
Received Via             945
```

```
City                            945
State                           945
Zip code                        945
Status                          945
Filing on Behalf of Someone     945
Month                           945
New_status                      945
dtype: int64
```

[52]:
```
total_billing_issues = billing_issues1 + billing_issues2 + billing_issues3
```

[53]:
```
total_billing_issues
```

[53]:
```
Ticket #                        378
Customer Complaint              378
Date                            378
Date_month_year                 378
Time                            378
Received Via                    378
City                            378
State                           378
Zip code                        378
Status                          378
Filing on Behalf of Someone     378
Month                           378
New_status                      378
dtype: int64
```

[54]:
```
total_service_issue = service_issues1 + service_issues2
total_service_issue
```

[54]:
```
Ticket #                        584
Customer Complaint              584
Date                            584
Date_month_year                 584
Time                            584
Received Via                    584
City                            584
State                           584
Zip code                        584
Status                          584
Filing on Behalf of Someone     584
Month                           584
New_status                      584
dtype: int64
```

[56]:
```
data1["Customer Complaint"].shape
```

[56]: (2224,)

[57]:
```
other_issues = 2224 - (total_internet_issues + total_service_issue +␣
↪total_billing_issues)
other_issues
```

[57]:
```
Ticket #                        317
Customer Complaint              317
Date                            317
Date_month_year                 317
Time                            317
Received Via                    317
City                            317
State                           317
Zip code                        317
Status                          317
Filing on Behalf of Someone     317
Month                           317
New_status                      317
dtype: int64
```

0.0.6 From this we can conclude that internet issues are more