PROJECT-4 F

October 21, 2022

```python
[1]: import numpy as np
     import pandas as pd
     import seaborn as sns
     import matplotlib.pyplot as plt
     import datetime
     %matplotlib inline
```

1 Import data into Python environment.

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

```python
[3]: df.head()
```

```
[3]:    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
```

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

```python
[4]:    # Converting to `datetime format`
        df ['Date']= pd.to_datetime (df['Date'])
        df ['Date_month_year']= pd.to_datetime (df['Date_month_year'])
        df ['Time']= pd.to_datetime (df['Time'])

        # Dropping the repeated `Date_month_year ` column.
        df= df.drop('Date_month_year',axis=1)

        ## Creating new table df1 with Date & Customer complaint
        df1=df[['Date','Customer Complaint']]

        # Adding new 'Month' column to 'df1'
        df1['Month']=df['Date'].dt.month
        df1
```

C:\Users\harsh\AppData\Local\Temp\ipykernel_12036\694767487.py:13:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df1['Month']=df['Date'].dt.month

| [4]:		Date	Customer Complaint	Month
	0	2015-04-22	Comcast Cable Internet Speeds	4
	1	2015-04-08	Payment disappear - service got disconnected	4
	2	2015-04-18	Speed and Service	4
	3	2015-05-07	Comcast Imposed a New Usage Cap of 300GB that …	5
	4	2015-05-26	Comcast not working and no service to boot	5
	…	…	…	…
	2219	2015-04-02	Service Availability	4
	2220	2015-06-02	Comcast Monthly Billing for Returned Modem	6
	2221	2015-06-09	complaint about comcast	6
	2222	2015-06-23	Extremely unsatisfied Comcast customer	6
	2223	2015-06-24	Comcast, Ypsilanti MI Internet Speed	6

[2224 rows x 3 columns]

2.0.1 Creating Chart

```python
[35]: from matplotlib import style
      style.use('ggplot')
      plt.figure(figsize=(20,15))

      plt.subplot(2,1,1)
      df1['Month'].value_counts(sort=False).rename(index={4:'April',5:'May',6:
      →'June'}).plot()
      plt.title('Trend chart for the number of complaints')
      plt.xlabel('Months')
      plt.ylabel('No. Of Complaints')


      plt.subplot(2,1,2)
      df["Date"].value_counts(sort=False).plot()
      plt.title('Trend chart for the number of complaints')
      plt.xlabel('Date')
      plt.ylabel('No. Of Complaints')
      plt.show()
```
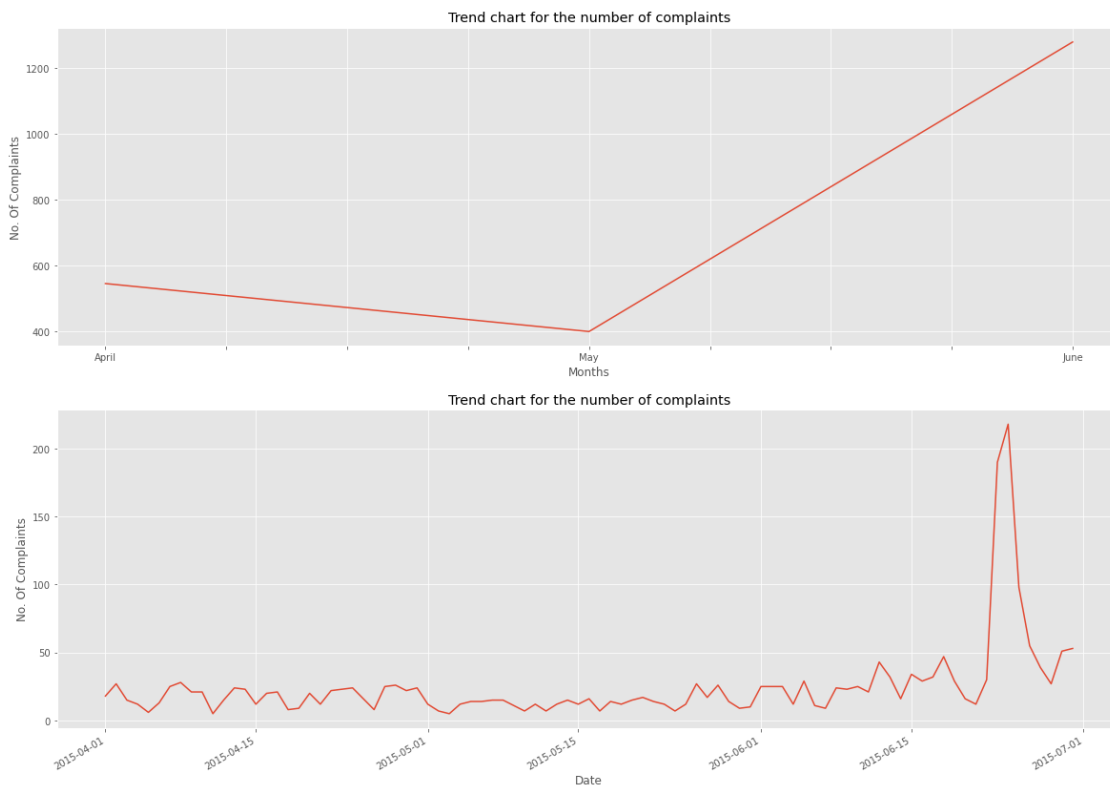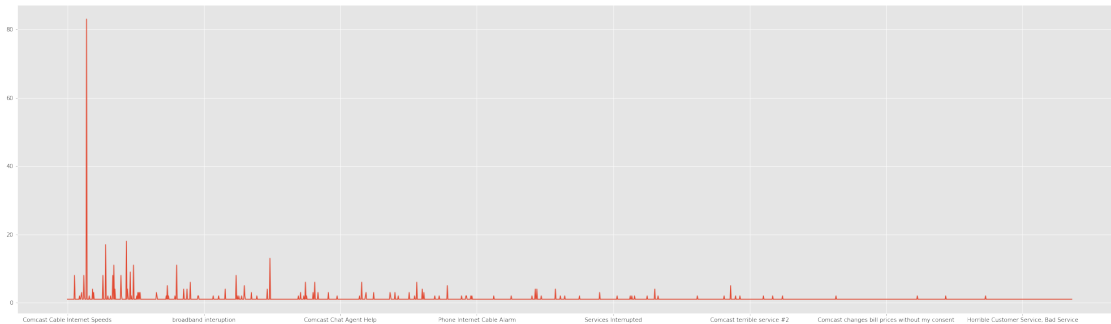
3 Provide a table with the frequency of complaint types.

```
[57]: plt.figure(figsize=(35,10))
      df['Customer Complaint'].value_counts(sort=False).plot(kind='line')
      plt.show()
```


3.1 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.

```
[46]: df["NewStatus"] = ["Open" if Status=="Open" or Status=="Pending" else "Closed"␣
      ↪for Status in df["Status"]]
      df
```

```
[46]:       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

                           Time         Received Via       City     State  Zip code  \
      0     2022-10-21 15:53:50  Customer Care Call    Abingdon  Maryland     21009
      1     2022-10-21 10:22:56            Internet     Acworth   Georgia     30102
      2     2022-10-21 09:55:47            Internet     Acworth   Georgia     30101
      3     2022-10-21 11:59:35            Internet     Acworth   Georgia     30101
      4     2022-10-21 13:25:26            Internet     Acworth   Georgia     30101
```

```
   …                  …                    …           …         …        …
2219 2022-10-21 09:13:18  Customer Care Call  Youngstown   Florida   32466
2220 2022-10-21 13:24:39  Customer Care Call   Ypsilanti  Michigan   48197
2221 2022-10-21 17:28:41             Internet   Ypsilanti  Michigan   48197
2222 2022-10-21 23:13:30  Customer Care Call   Ypsilanti  Michigan   48197
2223 2022-10-21 22:28:33  Customer Care Call   Ypsilanti  Michigan   48198


        Status Filing on Behalf of Someone  Closed  Open NewStatus
0       Closed                          No  Closed   NaN    Closed
1       Closed                          No  Closed   NaN    Closed
2       Closed                         Yes  Closed   NaN    Closed
3         Open                         Yes     NaN  Open      Open
4       Solved                          No  Solved   NaN    Closed
…          …                            …       …     …         …
2219    Closed                          No  Closed   NaN    Closed
2220    Solved                          No  Solved   NaN    Closed
2221    Solved                          No  Solved   NaN    Closed
2222    Solved                          No  Solved   NaN    Closed
2223      Open                         Yes     NaN  Open      Open


[2224 rows x 13 columns]
```

4 Provide state wise status of complaints in a stacked bar chart.

```
[49]: Status_complaints = df.groupby(["State","NewStatus"]).size().unstack().fillna(0)
      Status_complaints
```

```
[49]: NewStatus           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
```

[53]: `Status_complaints.plot(kind="bar", figsize=(30,10), stacked=True)`

[53]: `<AxesSubplot:xlabel='State'>`
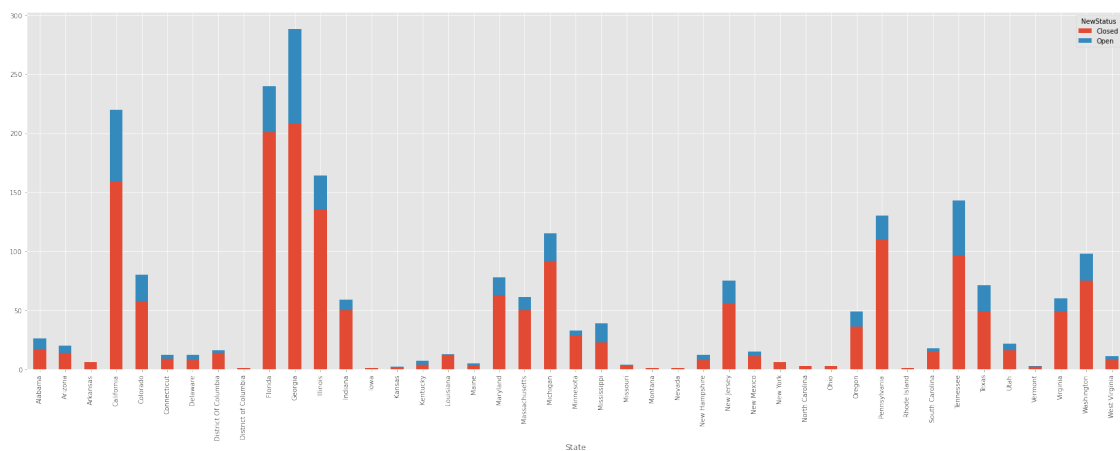
4.1 Q. Which state has the maximum complaints

4.1.1 As shown in the above Stacked bar graph maximum complaint is in "Georgia" State.

4.2 Q. Which state has the highest percentage of unresolved complaints

4.2.1 As shown in the above Stacked bar graph highest percentage of unresolved complaints is in "Georgia" State.

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

```
[66]: complaint_rece = df.groupby(['NewStatus','Received Via']).size().unstack().
      ↪fillna(0)
      complaint_rece
```

```
[66]: Received Via  Customer Care Call  Internet
      NewStatus
      Closed                       864       843
      Open                         255       262
```

```
[69]: Internet = (complaint_rece['Internet'][0] / (complaint_rece['Internet'][0] +␣
      ↪complaint_rece['Internet'][1]))*100
      CustomerCareCall_resolved = (complaint_rece['Customer Care Call'][0] /␣
      ↪(complaint_rece['Customer Care Call'][0] + complaint_rece['Customer Care␣
      ↪Call'][1]))*100
      print('Percentage of complaints resolved through Internet:',Internet,'%')
      print('Percentage of complaints resolved through Customer Care Call:
      ↪',CustomerCareCall_resolved,'%')
```

```
Percentage of complaints resolved through Internet: 76.289592760181 %
Percentage of complaints resolved through Customer Care Call: 77.21179624664879
%
```