COMCAST TELECOM CONSUMER COMPLAINTS

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

df=pd.read_csv("G:/AMRIT/PROJECT1/New
folder/Comcast_telecom_complaints_data.csv")

df.head()
```

```
  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
```

```
df.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
```

```
#Convert the date into datetime obj
df['Date']=pd.to_datetime(df['Date'])

df.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
```

```
df['month']=df['Date'].dt.month_name()

df.head()
```

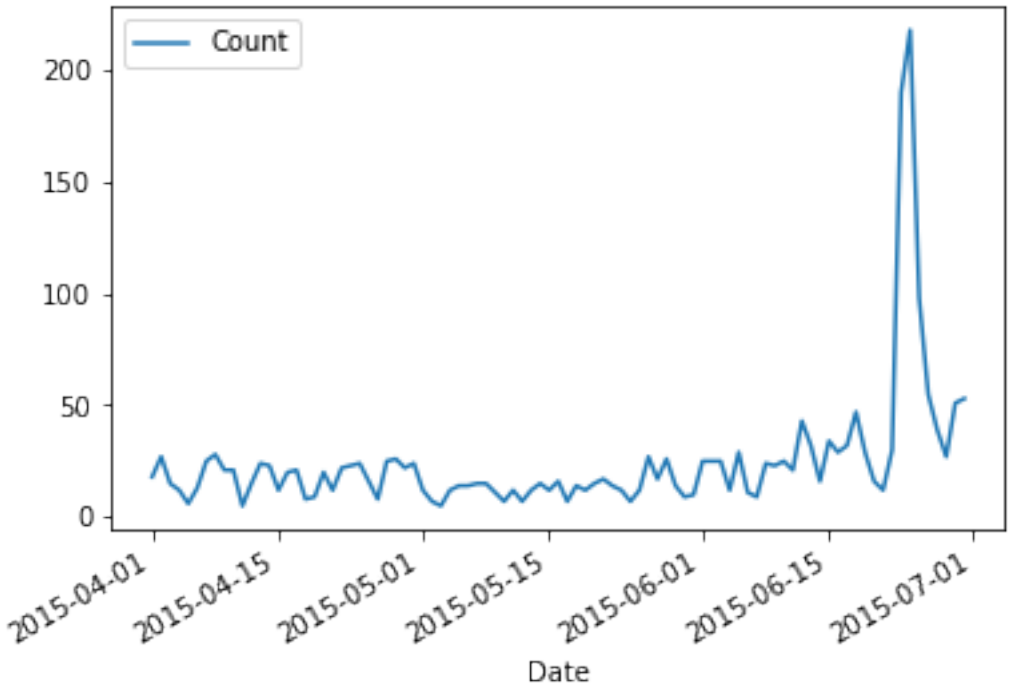
```
  Ticket #                                  Customer Complaint
Date  \
0   250635                       Comcast Cable Internet Speeds 2015-04-
22
1   223441       Payment disappear - service got disconnected 2015-04-
08
2   242732                                   Speed and Service 2015-04-
```

```
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
0     21009  Closed                          No  April
1     30102  Closed                          No  April
2     30101  Closed                         Yes  April
3     30101    Open                         Yes    May
4     30101  Solved                          No    May
```

```
#daily basis

daily=df['Date'].value_counts()

daily=pd.DataFrame(daily).reset_index()

daily.rename(columns={'index':'Date',
                      'Date':'Count'},inplace=True)

daily
```

```
         Date  Count
0  2015-06-24    218
1  2015-06-23    190
2  2015-06-25     98
3  2015-06-26     55
4  2015-06-30     53
..        ...    ...
86 2015-05-10      7
87 2015-05-24      7
88 2015-04-05      6
89 2015-04-11      5
90 2015-05-03      5

[91 rows x 2 columns]
```

```
daily.plot(x='Date',y='Count',kind='line')
plt.show()
```




```
df.groupby('month').count()['Ticket #']
```

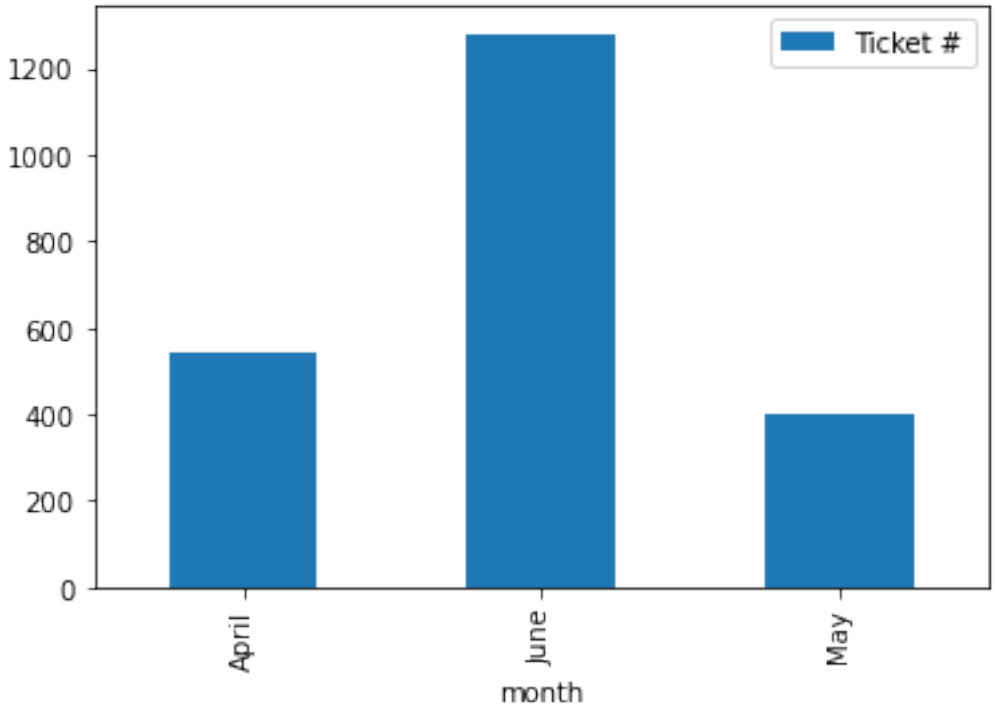
```
month
April     545
June     1280
May       399
Name: Ticket #, dtype: int64
```

```
mnth=pd.DataFrame(df.groupby('month').count()['Ticket
#']).reset_index()
```

```
mnth
```

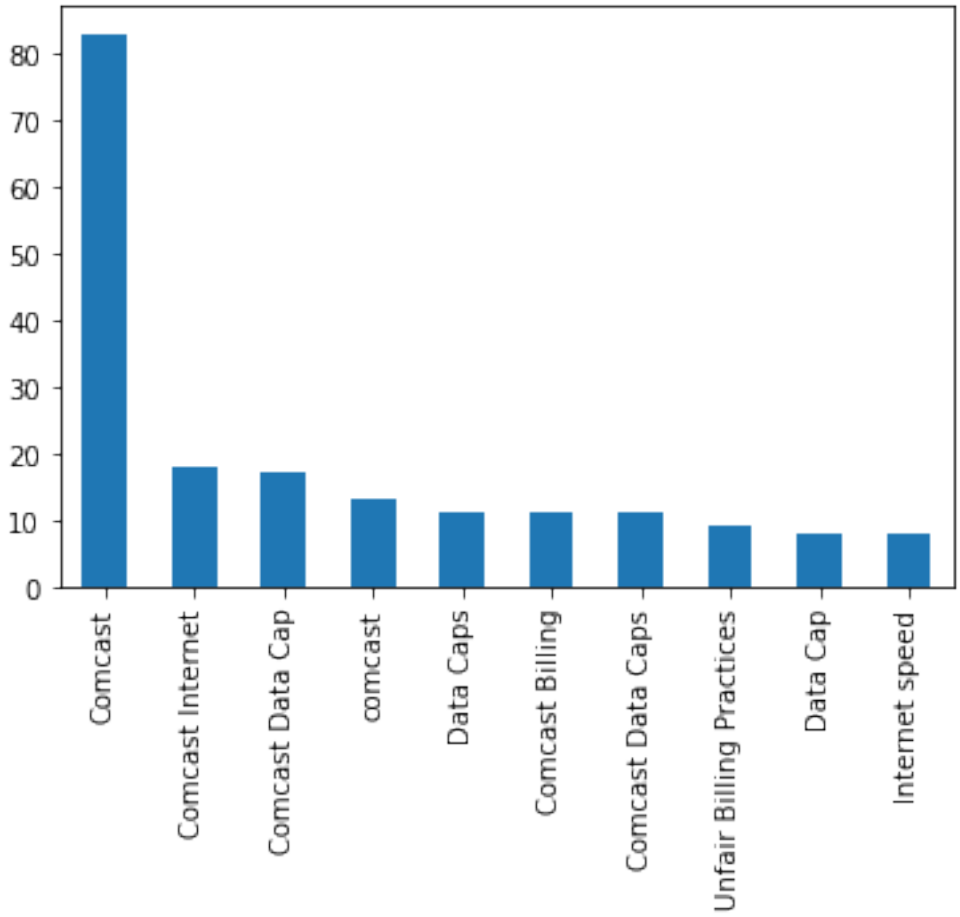
```
   month  Ticket #
0  April       545
1   June      1280
2    May       399
```

```
mnth.plot(x='month',y='Ticket #',kind='bar')
plt.show()
```




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

```
<AxesSubplot:>
```




```
internet_issue1=df[df['Customer
Complaint'].str.contains('speed')].count()['Ticket #']

internet_issue2=df[df['Customer
Complaint'].str.contains('network')].count()['Ticket #']

internet_issue3=df[df['Customer
Complaint'].str.contains('data')].count()['Ticket #']

total_internet_issue=internet_issue1+internet_issue2+internet_issue3

total_internet_issue

179

billing_issue1=df[df['Customer
Complaint'].str.contains('billing')].count()['Ticket #']

billing_issue2=df[df['Customer
Complaint'].str.contains('charges')].count()['Ticket #']

total_billing_issue=billing_issue1+billing_issue2
```

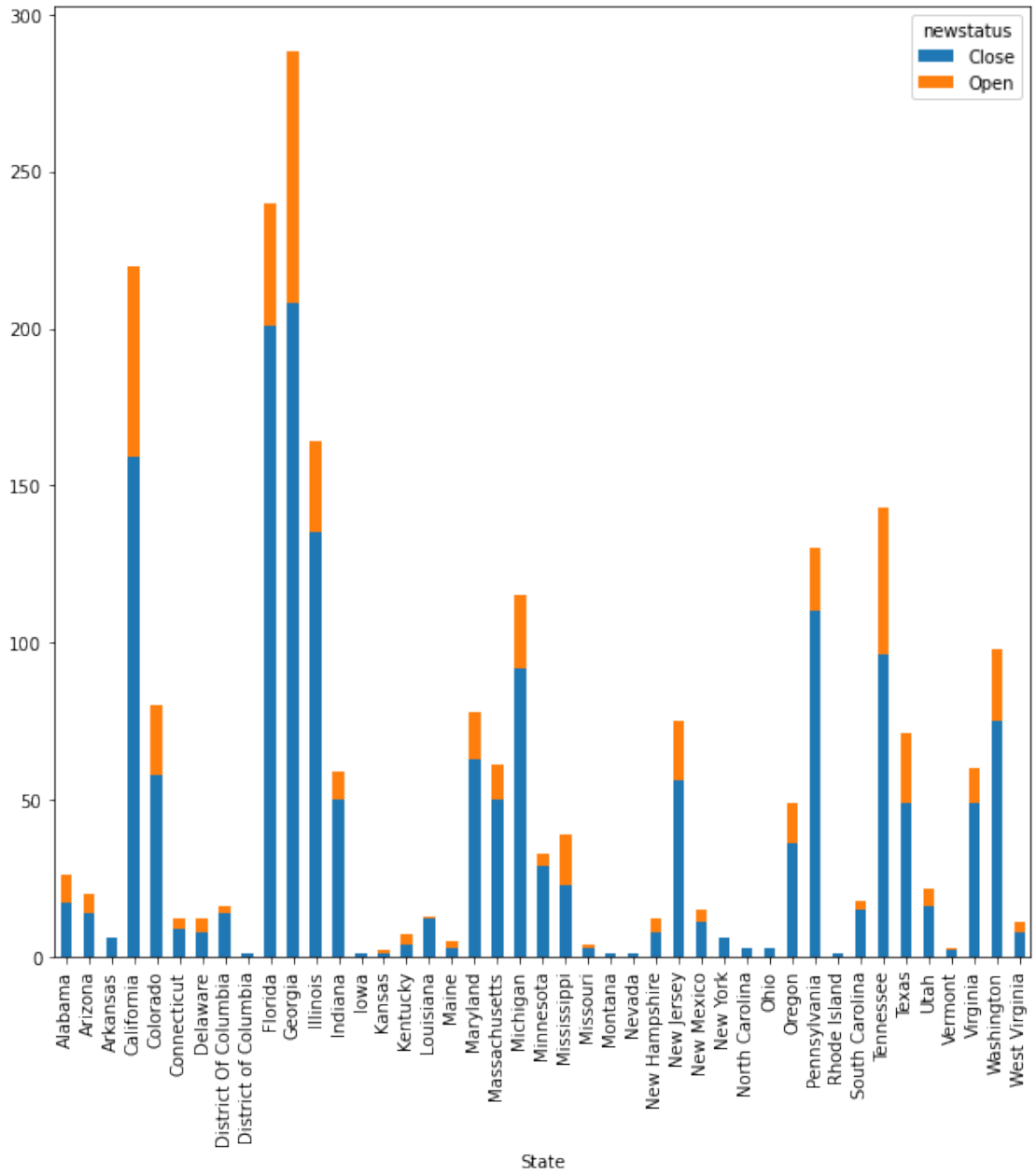
```
service_issue1=df[df['Customer
Complaint'].str.contains('service')].count()['Ticket #']

service_issue2=df[df['Customer
Complaint'].str.contains('customer')].count()['Ticket #']

total_service_issue=service_issue1+service_issue2

total_billing_issue,total_internet_issue,total_service_issue

(169, 179, 360)

other=2224-(169+179+360)

other

1516

df['Status'].unique()

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

df['newstatus']=['Open' if st=='Open' or st=='Pending' else 'Close'
for st in df['Status'] ]

df.head()

  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 newstatus
0     21009  Closed                         No  April     Close
1     30102  Closed                         No  April     Close
2     30101  Closed                        Yes  April     Close
3     30101    Open                        Yes    May      Open
4     30101  Solved                         No    May     Close
```

```
state_cmp=df.groupby(['State','newstatus']).size().unstack()
```

```
state_cmp.plot.bar(stacked=True,figsize=(10,10))
```

```
<AxesSubplot:xlabel='State'>
```



```
#Which state has the maximum complaints

df.groupby('State').count()['Ticket #'].sort_values(ascending=False)
[:5]
```

```
State
Georgia       288
Florida       240
California    220
Illinois      164
Tennessee     143
Name: Ticket #, dtype: int64
```

```
#Which state has the highest percentage of unresolved complaints

unresolved_data=df.groupby(['State','newstatus']).size().unstack().fil
lna(0).sort_values('Open',ascending=False)

unresolved_data['unresolved_cmp_prct']=unresolved_data['Open']/
unresolved_data['Open'].sum()*100

unresolved_data
```

newstatus	Close	Open	unresolved_cmp_prct
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
Washington	75.0	23.0	4.448743
Michigan	92.0	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
Virginia	49.0	11.0	2.127660
Massachusetts	50.0	11.0	2.127660
Alabama	17.0	9.0	1.740812
Indiana	50.0	9.0	1.740812
Utah	16.0	6.0	1.160542
Arizona	14.0	6.0	1.160542
New Hampshire	8.0	4.0	0.773694
New Mexico	11.0	4.0	0.773694
Minnesota	29.0	4.0	0.773694
Delaware	8.0	4.0	0.773694
West Virginia	8.0	3.0	0.580271
Connecticut	9.0	3.0	0.580271
Kentucky	4.0	3.0	0.580271

```
South Carolina         15.0   3.0            0.580271
Maine                   3.0   2.0            0.386847
District Of Columbia   14.0   2.0            0.386847
Kansas                  1.0   1.0            0.193424
Vermont                 2.0   1.0            0.193424
Missouri                3.0   1.0            0.193424
Louisiana              12.0   1.0            0.193424
Montana                 1.0   0.0            0.000000
Rhode Island            1.0   0.0            0.000000
Ohio                    3.0   0.0            0.000000
District of Columbia    1.0   0.0            0.000000
North Carolina          3.0   0.0            0.000000
New York                6.0   0.0            0.000000
Nevada                  1.0   0.0            0.000000
Arkansas                6.0   0.0            0.000000
Iowa                    1.0   0.0            0.000000
```

```
resolved_data=df.groupby(['Received
Via','newstatus']).size().unstack()
```

```
resolved_data['resolved']=resolved_data['Close']/
resolved_data['Close'].sum()*100
```

```
resolved_data['resolved']
```

```
Received Via
Customer Care Call    50.615114
Internet              49.384886
Name: resolved, dtype: float64
```