

## COMCAST TELECOM CONSUMER COMPLAINTS

Import the library

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

```
In [2]: import numpy as np
```

```
In [3]: import matplotlib.pyplot as plt
```

```
In [4]: import seaborn as sns
```

Import the data set

```
In [10]: df = pd.read_csv('Comcast.csv')
```

```
In [11]: df.head(10)
```

```
Out[11]:
```

	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
5	338519	ISP Charging for arbitrary data limits with ov...	06-12-15	06-Dec-15	9:59:40 PM	Internet	Acworth	Georgia	30101	Solved	No
6	361148	Throttling service and unreasonable data caps	24-06-15	24-Jun-15	10:13:55 AM	Customer Care Call	Acworth	Georgia	30101	Pending	No
7	359792	Comcast refuses to help troubleshoot and corre...	23-06-15	23-Jun-15	6:56:14 PM	Internet	Adrian	Michigan	49221	Solved	No
8	318072	Comcast extended outages	06-01-15	06-Jan-15	11:46:30 PM	Customer Care Call	Alameda	California	94502	Closed	No
9	371214	Comcast Raising Prices and Not Being Available...	28-06-15	28-Jun-15	6:46:31 PM	Customer Care Call	Alameda	California	94501	Open	Yes

```
In [12]: #convert date to datetime
```

```
In [14]: df['Date']=pd.to.datetime(df['Date'])
```

```
In [16]: 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
```

```
In [ ]: #create month from date
```

```
In [18]: df['Month']=df['Date'].dt.month_name()
```

```
In [19]: df.head()
```

```
Out[19]:
```

	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

```
In [21]: daily = df['Date'].value_counts()
```

```
In [22]: daily
```

```
Out[22]:
```

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

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

```
In [23]: # create a DataFrame to plot the graph
```

```
In [25]: daily = pd.DataFrame(daily).reset_index()
```

```
In [26]: daily
```

```
Out[26]:
```

	index	Date
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	2015-05-24	7
87	2015-05-02	7
88	2015-04-05	6
89	2015-04-11	5
90	2015-05-03	5

91 rows x 2 columns

```
In [ ]: # rename the index and date
```

```
In [31]: daily.rename(columns ={'index':'Date', 'Date':'count'}, inplace=True)
```

```
In [32]: daily
```

```
Out[32]:
```

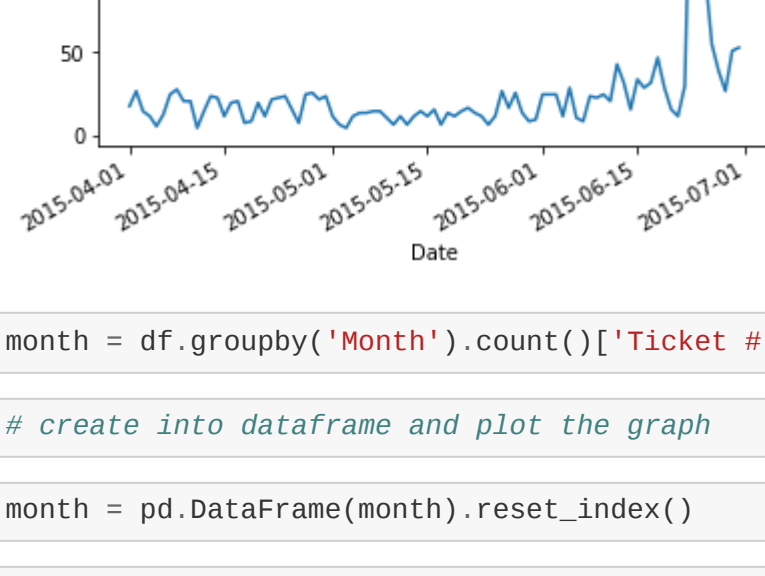
	Date	count
0	2015-06-24	218
1	2015-06-23	190
2	2015-06-25	98
3	2015-06-26	55
4	2015-06-30	53
...	...	...
86	2015-05-24	7
87	2015-05-02	7
88	2015-04-05	6
89	2015-04-11	5
90	2015-05-03	5

91 rows x 2 columns

```
In [ ]: # Now plot the graph date Vs Count
```

```
In [35]: daily.plot(x='Date', y='count', kind='line')
```

```
plt.show()
```



```
In [41]: month = df.groupby('Month').count()['Ticket #']
```

```
In [42]: # create into dataframe and plot the graph
```

```
In [43]: month = pd.DataFrame(month).reset_index()
```

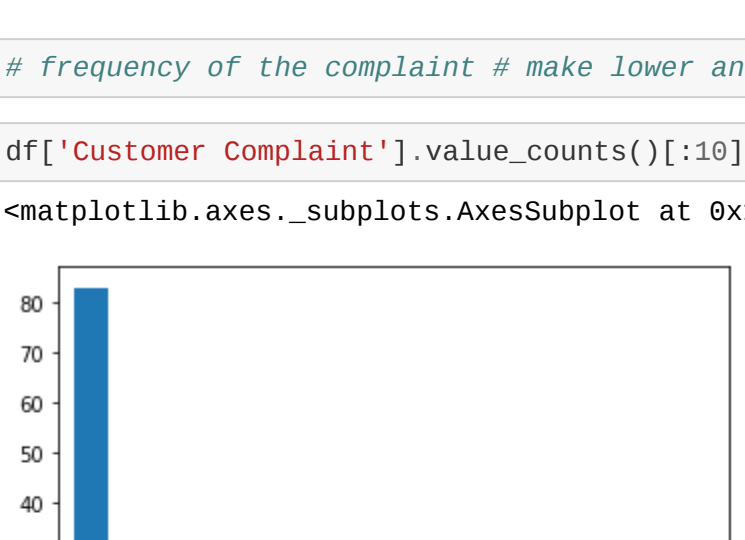
```
In [44]: month
```

```
Out[44]:
```

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

```
In [47]: month.plot(x='Month', y='Ticket #', kind='bar')
```

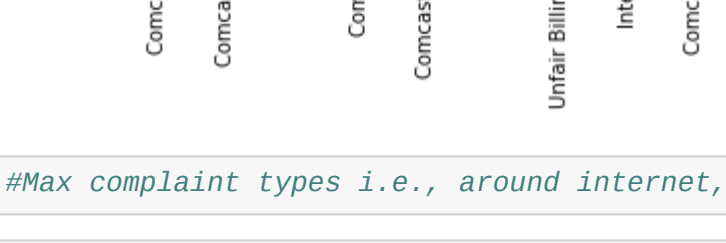
```
Out[47]: <matplotlib.axes._subplots.AxesSubplot at 0x1f3bd665d30>
```



```
In [ ]: # frequency of the complaint # make lower and upper case
```

```
In [50]: df['Customer Complaint'].value_counts()[1:10].plot.bar()
```

```
Out[50]: <matplotlib.axes._subplots.AxesSubplot at 0x1f3bd665d30>
```



```
In [ ]: #Max complaint types i.e., around internet, network issues, or across any other domains
```

```
In [55]: internet_issue1 = df[df['Customer Complaint'].str.contains('speed')].count()['Ticket #']
```

```
In [57]: internet_issue2 = df[df['Customer Complaint'].str.contains('data')].count()['Ticket #']
```

```
In [59]: internet_issue3 = df[df['Customer Complaint'].str.contains('network')].count()['Ticket #']
```

```
In [60]: total_internet_issue = internet_issue1 + internet_issue2 + internet_issue3
```

```
In [61]: total_internet_issue
```

```
Out[61]: 116
```

```
In [86]: billing_issue1 = df[df['Customer Complaint'].str.contains('billing')].count()['Ticket #']
```

```
In [87]: billing_issue1
```

```
Out[87]: 126
```

```
In [88]: billing_issue2 = df[df['Customer Complaint'].str.contains('charges')].count()['Ticket #']
```

```
In [89]: billing_issue3 = df[df['Customer Complaint'].str.contains('pricing')].count()['Ticket #']
```

```
In [97]: total_billing_issue = billing_issue1 + billing_issue2 + billing_issue3
```

```
In [98]: total_billing_issue
```

```
Out[98]: 193
```

```
In [92]: service_issue1 = df[df['Customer Complaint'].str.contains('service')].count()['Ticket #']
```

```
In [93]: service_issue2 = df[df['Customer Complaint'].str.contains('customer')].count()['Ticket #']
```

```
In [99]: total_service_issue = service_issue1 + service_issue2
```

```
In [100]: total_service_issue
```

```
Out[100]: 360
```

```
In [101]: total_internet_issue, total_billing_issue, total_service_issue
```

```
Out[101]: (116, 193, 360)
```

```
In [103]: df.shape
```

```
Out[103]: (2224, 12)
```

```
In [104]: # so we have to find other complain
```

```
In [105]: other_issue = 2224-(116+193+360)
```

```
In [106]: other_issue
```

```
Out[106]: 1555
```

```
In [107]: # As per my data_issue calculation, I found other issue has higher than total_internet_issue
#total_billing_issue, total_service_issue
```

```
In [ ]: #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 [108]: df['Newstatus']=[ 'Open' if st=='Open' or st == 'Pending' else 'Close' for st in df['Status']]
```

```
In [109]: df.head()
```

```
Out[109]:
```

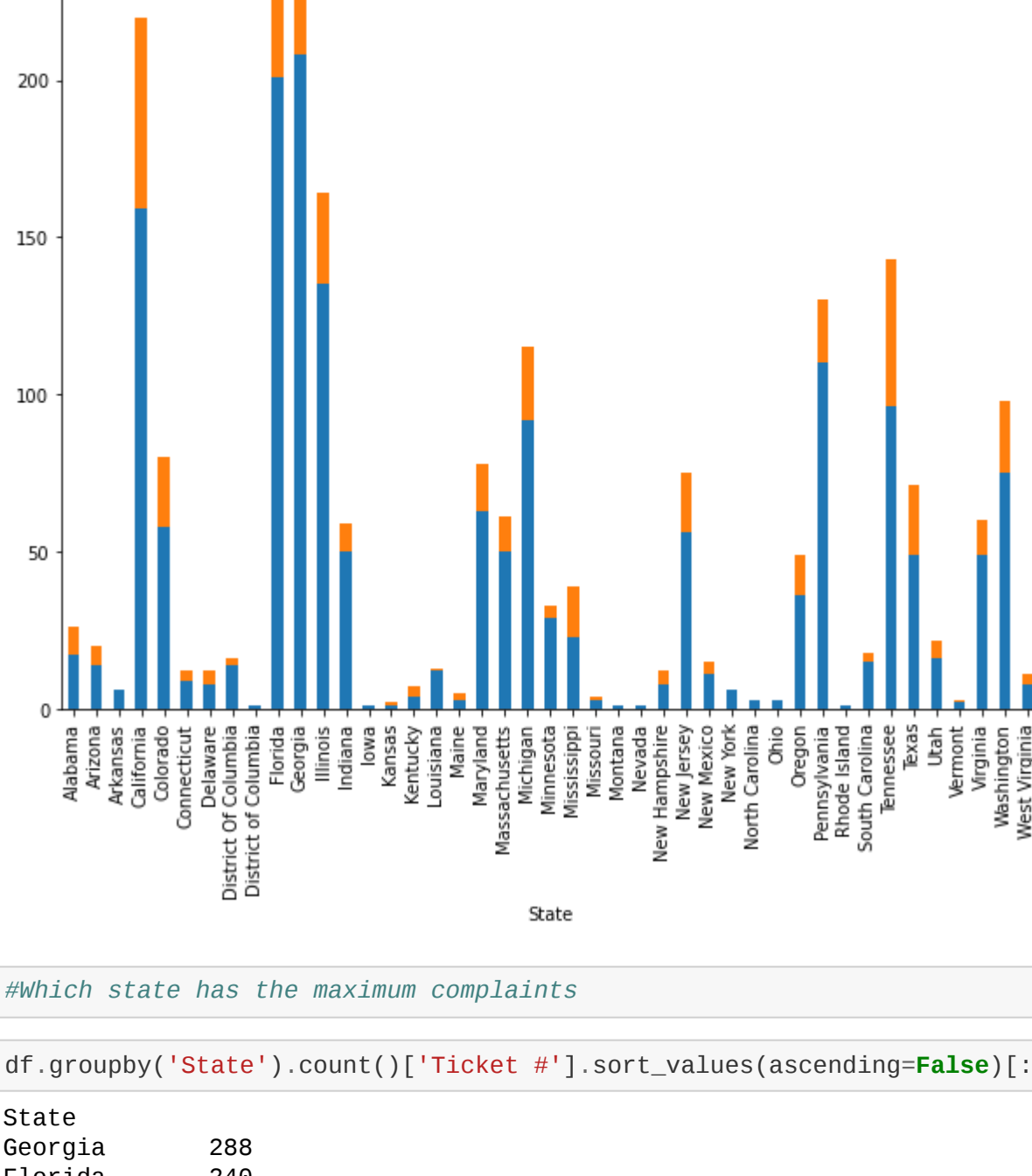
	Ticket #	Customer Complaint	Date	Date_month_year	Time	Received Via	City	State	Zip code	Status	Filing on Behalf of Someone	Month
0	250635	Comcast Cable Internet Speeds	2015-04-22	22-Apr-15	3:53:50 PM	Customer Care Call	Abingdon	Maryland	21009	Closed	No	April
1	223441	Payment disappear - service got disconnected	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

```
In [ ]: #now we have to Provide state wise status of complaints in a stacked bar chart
```

```
In [125]: state_complaint=df.groupby(['State', 'Newstatus']).size().unstack()
```

```
In [126]: state_complaint.plot(stacked=True, figsize=(10,10))
```

```
Out[126]: <matplotlib.axes._subplots.AxesSubplot at 0x1f3bf7dfb50>
```



```
In [127]: #Which state has the maximum complaints
```

```
In [128]: df.groupby('State').count()['Ticket #'].sort_values(ascending=False)[1:5]
```

```
Out[128]:
```

State	288
Georgia	288
Florida	240
California	220
Illinois	164
Tennessee	143

Name: Ticket #, dtype: int64

```
In [129]: #Which state has the maximum complaints
```

```
In [130]: df.groupby('State').count()['Ticket #'].sort_values(ascending=False)[1:5]
```

```
Out[130]:
```

State	288
Georgia	288
Florida	240
California	220
Illinois	164
Tennessee	143

Name: Ticket #, dtype: int64

```
In [131]: #Which state has the highest percentage of unresolved complaints
```

```
In [133]: unresolved_data=df.groupby(['State', 'Newstatus']).size().unstack().fillna(0).sort_values('Open', ascending=False)
```

```
In [134]: unresolved_data['unresolved_cmp_prcnt']=unresolved_data['Open']/unresolved_data['Open'].sum()*100
```

```
In [136]: unresolved_data.head()
```

```
Out[136]:
```

	Newstatus	Close	Open	unresolved_cmp_prcnt
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	

```
In [140]: resolved_data=df.groupby(['Received_Via', 'Newstatus']).size().unstack()
```

```
In [141]: resolved_data['resolved']=resolved_data['Close']/resolved_data['Close'].sum()*100
```

```
In [143]: resolved_data['resolved']
```

```
Out[143]:
```

Received_Via	50.615114
Customer Care Call	49.384886

Name: resolved, dtype: float64

```
In [ ]:
```