Data Science with Python

Certification Project

Comcast Telecom Consumer Complaints Analysis

Problem Statement:

Comcast is an American global telecommunication company. The firm has been providing terrible customer service. They continue to fall short despite repeated promises to improve. Only last month (October 2016) the authority fined them a $2.3 million, after receiving over 1000 consumer complaints.
The existing database will serve as a repository of public customer complaints filed against Comcast.
It will help to pin down what is wrong with Comcast's customer service.

Solution

✓ Import data into Python environment.

df = pd.read_csv('C:\\Users\\sobran\\Documents\\Python DS project\\comcast telecom consumer complaints\\Comcast_telecom_complaints_data.csv')

```
df.head()
```

	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	22-04-2015	22-Apr-15	3:53:50 PM	Customer Care Call	Abingdon	Maryland	21009	Closed	No
1	223441	Payment disappear - service got disconnected	04-08-2015	04-Aug-15	10:22:56 AM	Internet	Acworth	Georgia	30102	Closed	No
2	242732	Speed and Service	18-04-2015	18-Apr-15	9:55:47 AM	Internet	Acworth	Georgia	30101	Closed	Yes
3	277946	Comcast Imposed a New Usage Cap of 300GB that ...	05-07-2015	05-Jul-15	11:59:35 AM	Internet	Acworth	Georgia	30101	Open	Yes
4	307175	Comcast not working and no service to boot	26-05-2015	26-May-15	1:25:26 PM	Internet	Acworth	Georgia	30101	Solved	No

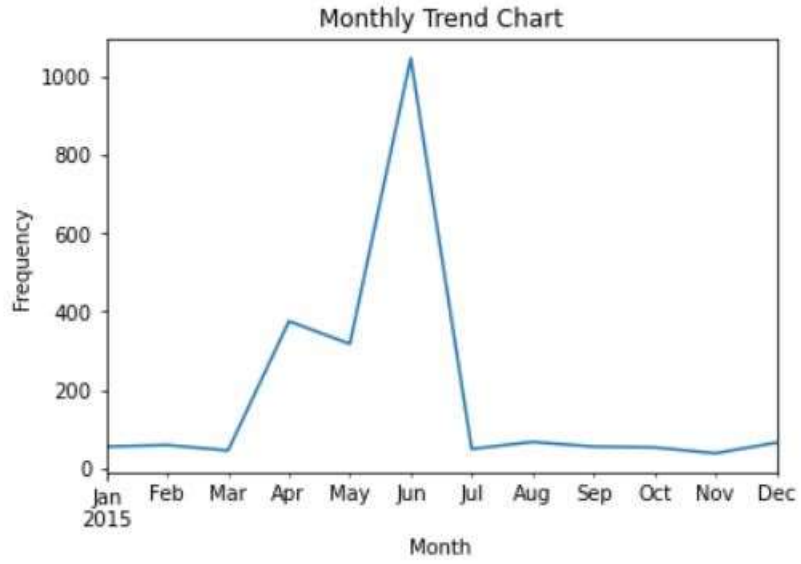
```
df.info()
```

```
<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
```

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

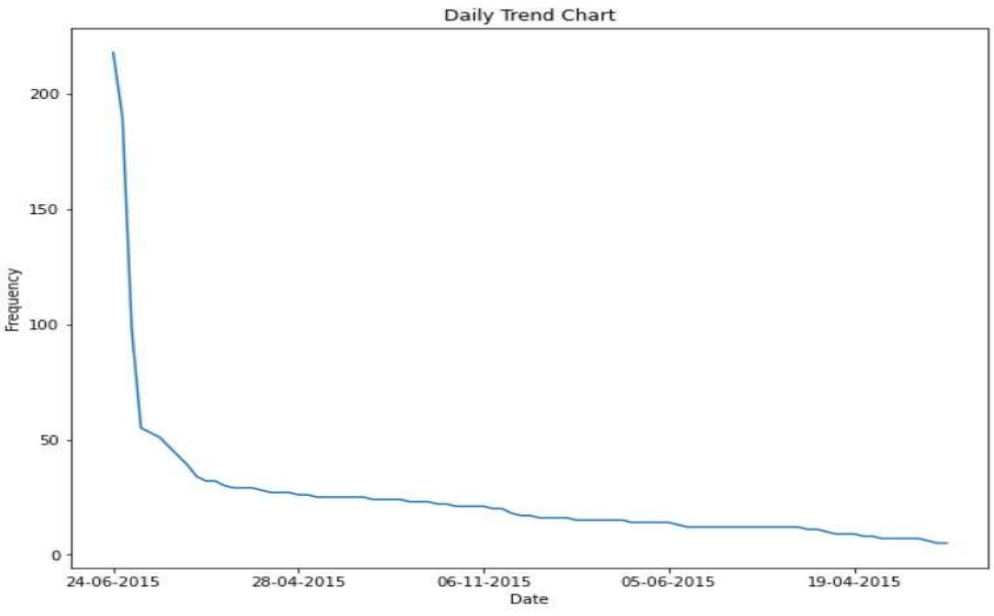
```
df.groupby(pd.Grouper(freq='M')).size().plot()
plt.xlabel('Month')
plt.ylabel('Frequency')
plt.title('Monthly Trend Chart')
```

```
Text(0.5, 1.0, 'Monthly Trend Chart')
```




```
df = df.sort_values(by='Date')
plt.figure(figsize=(10,8))
df['Date'].value_counts().plot()
plt.xlabel('Date')
plt.ylabel('Frequency')
plt.title('Daily Trend Chart')
```

```
Text(0.5, 1.0, 'Daily Trend Chart')
```



✓ Provide a table with the frequency of complaint types.

```
df['Customer Complaint'].value_counts()
```

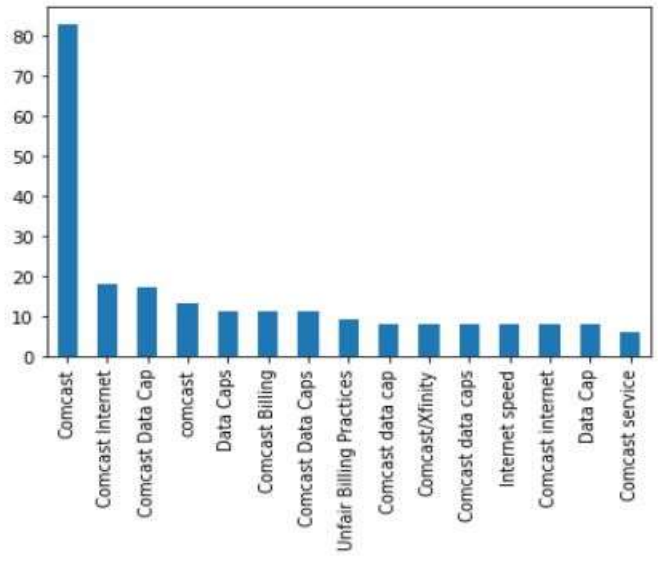
```
Comcast                                                          83
Comcast Internet                                                 18
Comcast Data Cap                                                 17
comcast                                                          13
Data Caps                                                        11
                                                                 ..
Comcast Internet, cable, and phone outtages                       1
Unable to renew IP address                                        1
Lack of availability                                              1
Comcast blocking DirecTv signals                                  1
Comcast Billing for Late Payment/Disconnect due to their error    1
Name: Customer Complaint, Length: 1841, dtype: int64
```

```
df['Customer Complaint'].value_counts()[:15]
```

```
Comcast                    83
Comcast Internet           18
Comcast Data Cap           17
comcast                    13
Data Caps                  11
Comcast Billing            11
Comcast Data Caps          11
Unfair Billing Practices    9
Comcast data cap            8
Comcast/Xfinity             8
Comcast data caps           8
Internet speed              8
Comcast internet            8
Data Cap                    8
Comcast service             6
Name: Customer Complaint, dtype: int64
```

```
df['Customer Complaint'].value_counts()[:15].plot.bar()
```

```
<AxesSubplot:>
```



✓ Which complaint types are maximum i.e., around internet, network issues, or across any other domains.

```
total_issue_internet = internet_issue1 + internet_issue2 + internet_issue3 + internet_issue4
toprint(total_issue_internet)
```

```
Customer Complaint            374
Date                          374
Received Via                  374
City                          374
State                         374
Zip code                      374
Status                        374
Filing on Behalf of Someone   374
dtype: int64
```

```
total_billing_issues = billing_issue1 + billing_issue2 + billing_issue3
print(total_billing_issues)
```

```
Customer Complaint            353
Date                          353
Received Via                  353
City                          353
State                         353
Zip code                      353
Status                        353
Filing on Behalf of Someone   353
dtype: int64
```

```
total_service_issues = service_issue1 + service_issue2
print(total_service_issues)
```

```
Customer Complaint            360
Date                          360
Received Via                  360
City                          360
State                         360
Zip code                      360
Status                        360
Filing on Behalf of Someone   360
dtype: int64
```

```
df.shape
```

```
(2224, 8)
```

```
other_issues = 2224 - (total_billing_issues + total_service_issues + total_issue_internet)
```

```
print(other_issues)
```

```
Customer Complaint            1137
Date                          1137
Received Via                  1137
City                          1137
State                         1137
Zip code                      1137
Status                        1137
Filing on Behalf of Someone   1137
dtype: int64
```

✓ 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.

```
df['newStatus']= ['Open' if Status=='Open' or Status=='Pending' else 'Closed' for Status in df['Status']]
```

```
df.sample(10)
```

Date_month_year	Customer Complaint	Date	Received Via	City	State	Zip code	Status	Filing on Behalf of Someone	newStatus
2015-06-13	Advertised Internet price not honored	13-06-2015	Internet	Peachtree City	Georgia	30269	Solved	No	Closed
2015-06-27	Comcast Throttling My Internet	27-06-2015	Customer Care Call	Fort Collins	Colorado	80524	Pending	Yes	Open
2015-08-05	Comcast Added Service After I Declined It	05-08-2015	Internet	Atlanta	Georgia	30345	Solved	No	Closed
2015-05-05	Comcast/earthlink	05-05-2015	Internet	Hoirn Lake	Mississippi	38637	Solved	No	Closed
2015-07-04	Transfer services	04-07-2015	Customer Care Call	Corryton	Tennessee	37721	Closed	No	Closed
2015-06-23	Comcast data cap at 300GB	23-06-2015	Customer Care Call	Charleston	South Carolina	29405	Pending	No	Open
2015-05-31	Comcast deceptive selling billing lack of serv...	31-05-2015	Internet	Evans	Georgia	30809	Closed	No	Closed
2015-02-06	Comcast internet	06-02-2015	Internet	Chicago	Illinois	60614	Closed	No	Closed
2015-06-20	terrible internet quality	20-06-2015	Customer Care Call	Philadelphia	Pennsylvania	19116	Solved	No	Closed
2015-05-15	Unannouced service outage not caused by weather	15-05-2015	Customer Care Call	Morrisville	Pennsylvania	19067	Solved	No	Closed

✓ Which state has the maximum complaints.

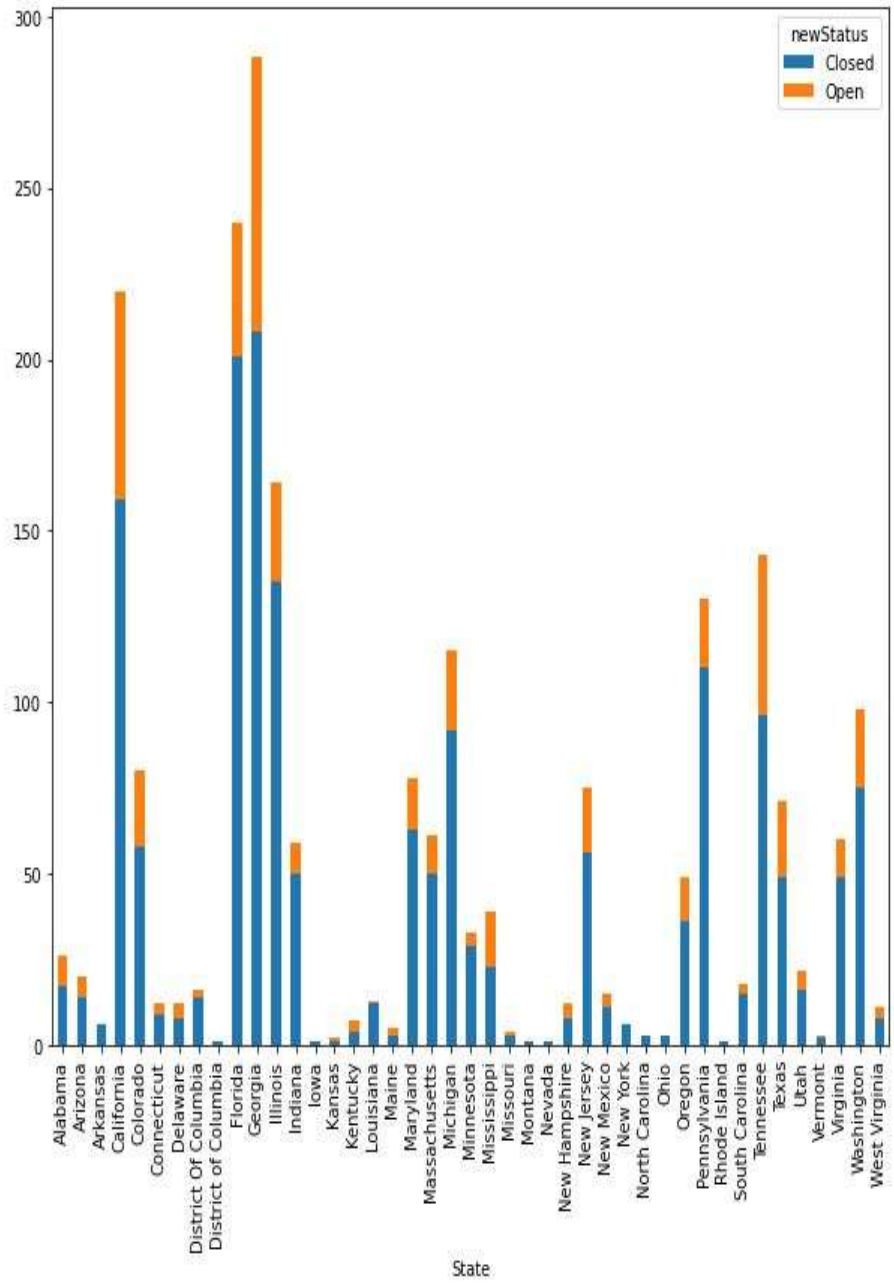
```
df.groupby(['State']).size().sort_values(ascending=False)[:10]
```

```
State
Georgia         288
Florida         240
California      220
Illinois        164
Tennessee       143
Pennsylvania    130
Michigan        115
Washington       98
Colorado         80
Maryland         78
dtype: int64
```

✓ Provide state wise status of complaints in a stacked bar chart. Use the categorized variable from Q3.

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

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



✓ Which state has the highest percentage of unresolved complaints.

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

```
print(unresolved_data)
```

newStatus	Closed	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

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

```
resolved_data = df.groupby(['Received Via','newStatus']).size().unstack().fillna(0)
```

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

```
newStatus           Closed  Open   resolved
Received Via
Customer Care Call     864   255  50.615114
Internet               843   262  49.384886
```

```
resolved_data.plot(kind='bar',figsize=(10,10))
```

```
<AxesSubplot:xlabel='Received Via'>
```