



# PROJECT REPORT

COMCAST TELECOM CONSUMER COMPLAINTS

BY: WALEED ALAMSLOKHI

## Consumer Complaints:

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.

## Data Dictionary:

- Ticket #: Ticket number assigned to each complaint
- Customer Complaint: Description of complaint
- Date: Date of complaint
- Time: Time of complaint
- Received Via: Mode of communication of the complaint
- City: Customer city
- State: Customer state
- Zip code: Customer zip
- Status: Status of complaint
- Filing on behalf of someone

## Analysis to be done:

To perform these tasks, you can use any of the different Python libraries such as NumPy, SciPy, Pandas, scikit-learn, matplotlib, and Beautiful Soup.

- Import data into Python environment.
- Provide the trend chart for the number of complaints at monthly and daily granularity levels.
- Provide a table with the frequency of complaint types.
- Which complaint types are maximum i.e., around internet, network issues, or across any other domains.
- 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.
- Provide state wise status of complaints in a stacked bar chart. Use the categorized variable from Q3. Provide insights on:
  - Which state has the maximum complaints
  - Which state has the highest percentage of unresolved complaints
  - Provide the percentage of complaints resolved till date, which were received through the Internet and customer care calls.

## SOURCE CODE :

```
import pandas as pd
import numpy as np
from pandas import datetime as dt

df_faa_dataset = pd.read_csv('/Users/a8/Desktop/Data Science//Comcast_telecom_complaints_data.csv')

complaint_df=pd.read_csv('/Users/a8/Desktop/Data Science//Comcast_telecom_complaints_data.csv')
complaint_df.head() #lookout first five rows
complaint_df.info()

complaint_df.Date=pd.to_datetime(complaint_df.Date)
complaint_df['Zip code']=complaint_df['Zip code'].astype(str)
complaint_df.info()
complaint_df.isnull().sum()
complaint_df['month']=complaint_df.Date.dt.month
complaint_df['month']=complaint_df['month'].replace([1,2,3,4,5,6,7,8,9,10,11,12],['Jan','Feb','March','April','May','June','July','Aug','Sep','Oct','Nov','Dec'])
complaint_df['month'].head()
import seaborn as sns
import matplotlib
import matplotlib.pyplot as plt
%matplotlib inline

sns.set_style('darkgrid')
matplotlib.rcParams['font.size'] = 14
matplotlib.rcParams['figure.figsize'] = (9, 5)
matplotlib.rcParams['figure.facecolor'] = '#00000000'
complaint_df['Customer Complaint'].groupby(complaint_df.Date).count().plot(kind='bar',figsize=(22,6))
complaint_df['month'].value_counts().plot(kind='bar',color='grey')
sns.countplot(x='month', hue='Status', data=complaint_df)
sns.countplot(x='month', hue='Status', data=complaint_df)

from collections import Counter
c_count = Counter(complaint_df['Customer Complaint'])
d=c_count.most_common(10)
d=pd.DataFrame(d,columns=['type','count'])
d

statewise_df=complaint_df[['Ticket #', 'State']].groupby('State').count().sort_values(by='Ticket #',ascending=False)
statewise_df.head()

statewise_df.plot(kind='bar')

complaint_df.Status.unique()
```

```
complaint_df['open/closed']=complaint_df['Status'].apply(lambda x: 'Closed' if (x == 'Solved' or x == 'Closed') else 'Open')
```

```
Statewise_solving_rate=complaint_df.groupby(['State','open/closed'])['open/closed'].count().unstack().fillna(0)
Statewise_solving_rate=pd.DataFrame(Statewise_solving_rate)
Statewise_solving_rate.head(2)
```

```
Statewise_solving_rate['unresolved complaint percent']=(Statewise_solving_rate['Open']/Statewise_solving_rate['Closed']*100)
```

```
Statewise_solving_rate.plot(kind='bar',stacked=True,figsize=(15,4))
```

```
compl=complaint_df.groupby(['Received Via','open/closed'])['Received Via'].count().unstack()
compl
```

```
compl=complaint_df['open/closed'].value_counts()
compl
```

```
totcompl=len(complaint_df)
totcompl
```

```
for x in compl.values:
    l=(x/totcompl*100)
    print(l)
```

## data

	Ticket #	Customer Complaint	Date	Time	Received Via	City	State	Zip code	Status	Filing on Behalf of Someone	Description
0	250635	Comcast Cable Internet Speeds	4/22/2015	3:53:50 PM	Internet	Abingdon	Maryland	21009	Closed	No	I have been contacting Comcast Internet Techni...
1	223441	Payment disappear - service got disconnected	4/8/2015	10:22:56 AM	Internet	Acworth	Georgia	30102	Closed	No	Back in January 2015 I made 2 payments: One fo...
2	242732	Speed and Service	4/18/2015	9:55:47 AM	Internet	Acworth	Georgia	30101	Closed	Yes	Our home is located at in Acworth Georgia 3010...
3	277946	Comcast Imposed a New Usage Cap of 300GB that ...	5/7/2015	11:59:35 AM	Internet	Acworth	Georgia	30101	Open	Yes	Comcast in the Atlanta area has just put into ...
4	307175	Comcast not working and no service to boot	5/26/2015	1:25:26 PM	Internet	Acworth	Georgia	30101	Solved	No	I have been a customer of Comcast of some sort...



# FUNCTIONS IMPLEMENTED

```
import pandas
```

pandas is a software library written for the Python programming language for data manipulation and analysis

```
import numpy
```

numpy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays

```
import seaborn
```

Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.

```
import matplotlib
```

Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python.

# Analysis Task

- Import data into Python environment.

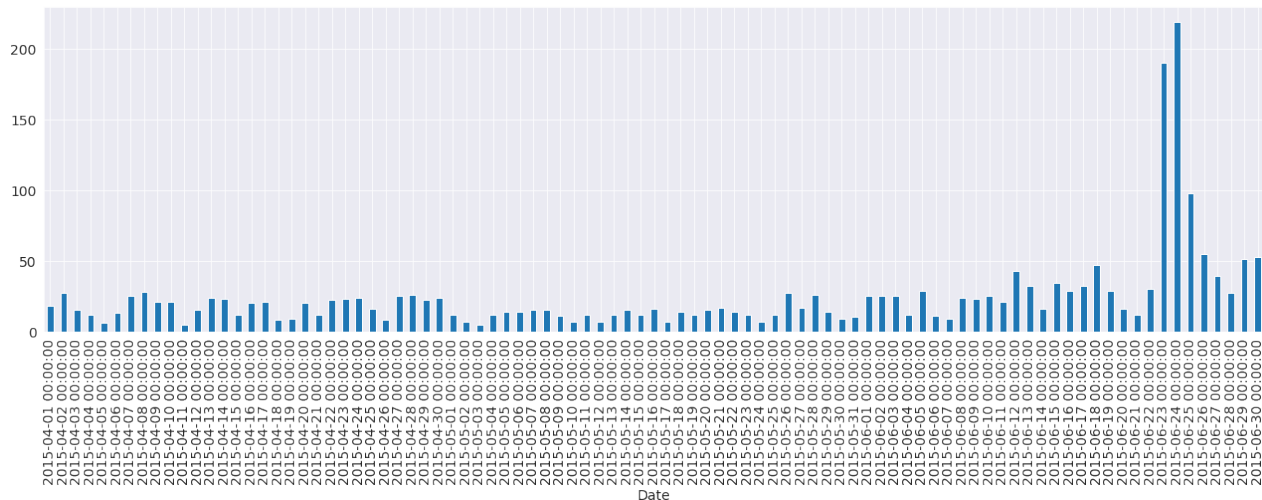
```
: import pandas as pd
import numpy as np
from pandas import datetime as dt

import seaborn as sns
import matplotlib
import matplotlib.pyplot as plt
%matplotlib inline

sns.set_style('darkgrid')
matplotlib.rcParams['font.size'] = 14
matplotlib.rcParams['figure.figsize'] = (9, 5)
matplotlib.rcParams['figure.facecolor'] = '#00000000'

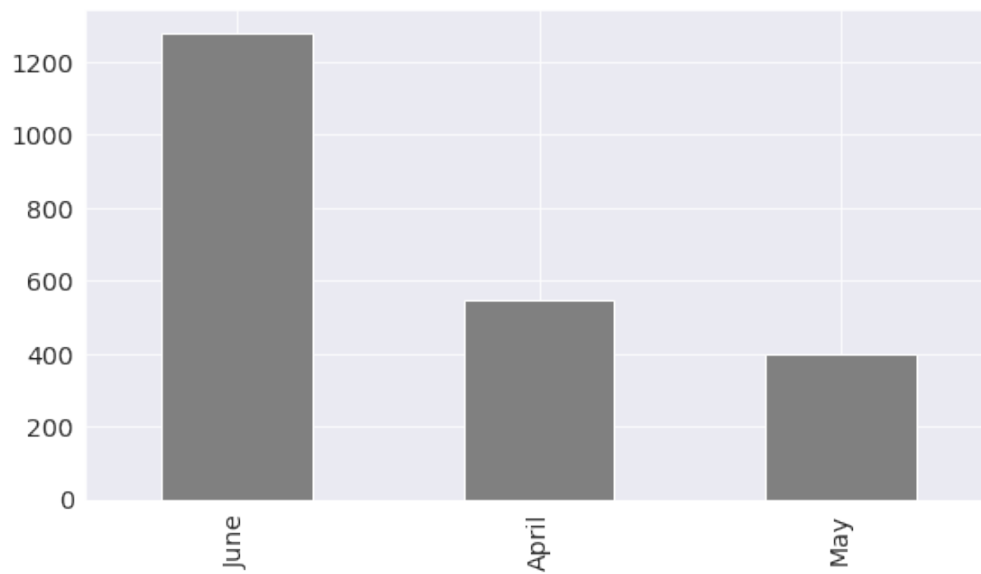
from collections import Counter
c_count = Counter(complaint_df['Customer Complaint'])
d=c_count.most_common(10)
d=pd.DataFrame(d,columns=['type','count'])
d
```

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



Highest complaints (more than 200) received on Date 24.06.2015

complaints over the month



We got highest number of complaints in the month of June according to dataset

- Provide a table with the frequency of complaint types.

	type	count
0	Comcast	83
1	Comcast Internet	18
2	Comcast Data Cap	17
3	comcast	13
4	Data Caps	11
5	Comcast Data Caps	11
6	Comcast Billing	11
7	Unfair Billing Practices	9
8	Internet speed	8
9	Comcast data caps	8

maximum complaints are coming for internet issue

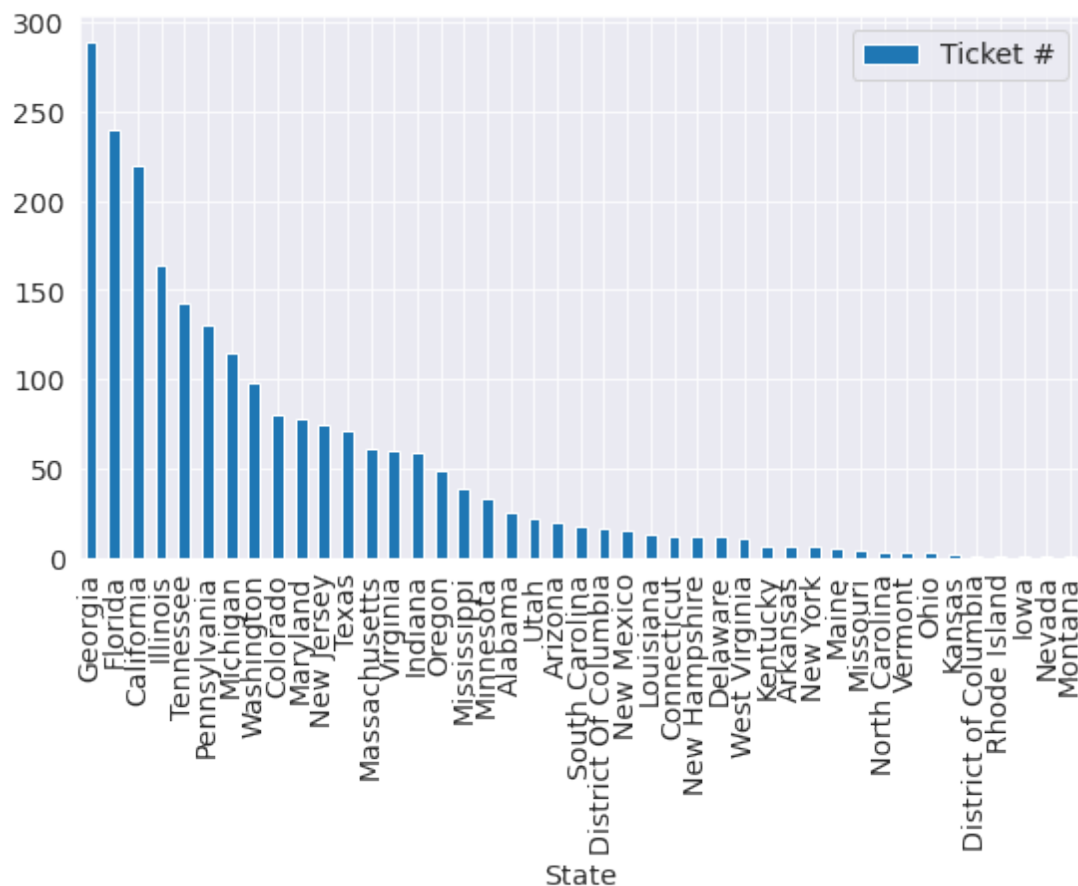
Which complaint types are maximum i.e., around internet, network issues, or across any other domains.

### Georgia state has received maximum

Georgia state has received maximum customer complaints we have to give special attention to that state Georgia, Florida and California are three states from where maximum complaints were received

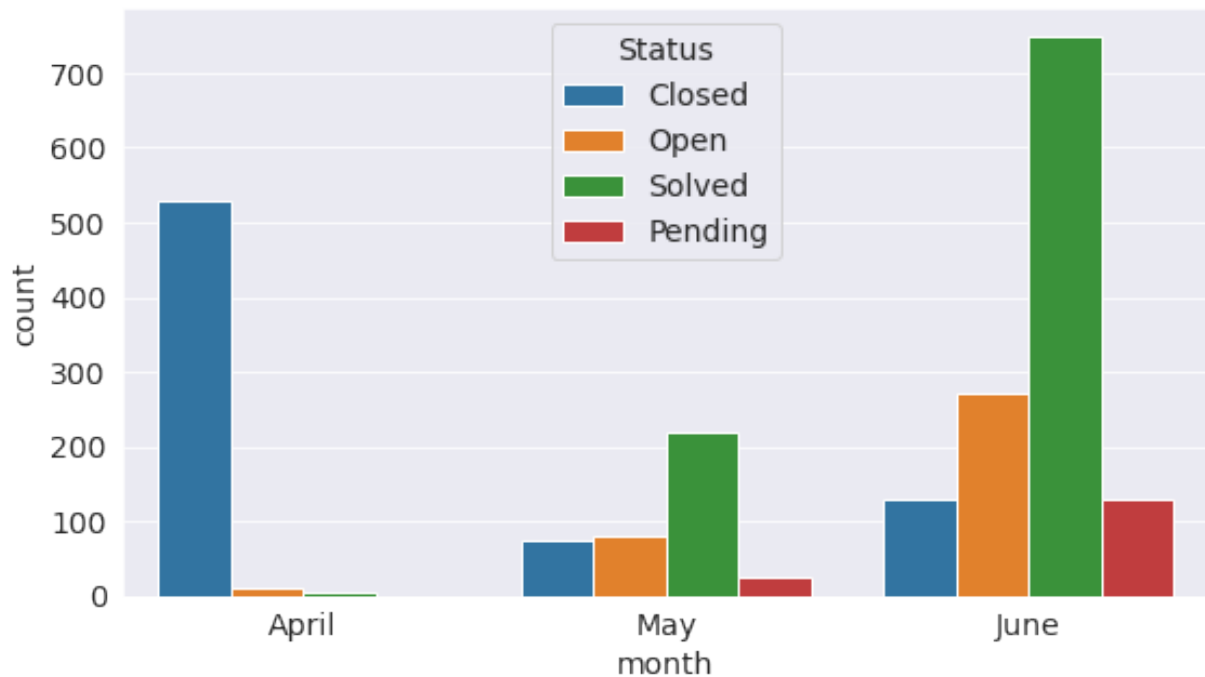
### Georgia state has received maximum

Ticket #	
State	
Georgia	289
Florida	240
California	220
Illinois	164
Tennessee	143





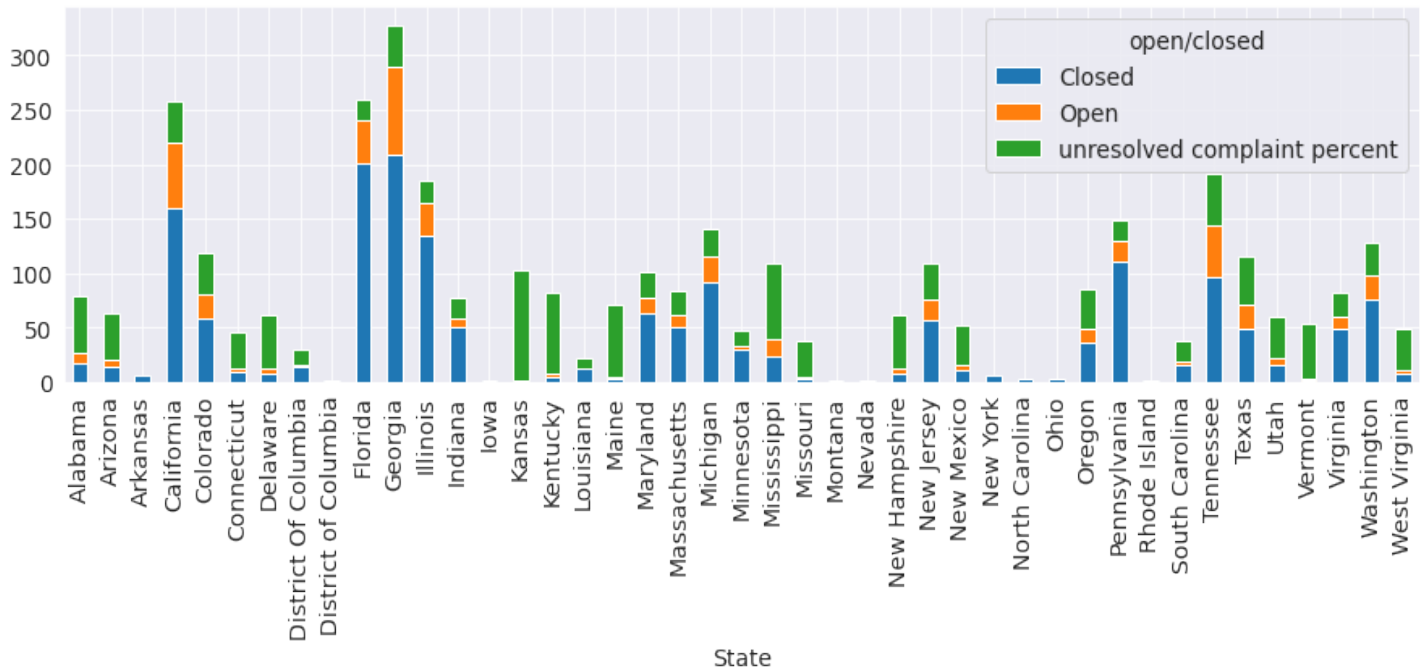
- Create a new categorical variable with value as **Open** and **Closed**. Open & Pending is to be categorized as Open and Closed & Solved is to be categorized as Closed.



open/closed	Closed	Open	unresolved complaint percent
State			
Alabama	17.0	9.0	52.941176
Arizona	14.0	6.0	42.857143
Arkansas	6.0	0.0	0.000000
California	159.0	61.0	38.364780
Colorado	58.0	22.0	37.931034

Closed	1707
Open	518

-Provide state wise status of complaints in a stacked bar chart. Use the categorized variable from Q3. Provide insights on:



from above stacked chart it is clear that max complaint states having around 70% complaint resolving rate  
and state Kansas has 1.0 unresolved complaint out of 1.0 received complaint thus unresolved percentage complaint is 100%