In [1]:
```python
# Import necessary packages

import pandas as pd
import os
import matplotlib.pyplot as plt
import numpy as np
os.chdir('C:\\Users\\utpala mohapatra\\Documents\\python_folder\\python datasets')
os.getcwd()
```

Out[1]: 'C:\\Users\\utpala mohapatra\\Documents\\python_folder\\python datasets'

In [2]:
```python
# 1> Import data into Python environment.

comcast_df = pd.read_csv('Comcast_telecom_complaints_data.csv')
comcast_df
```

Out[2]:

	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
...	...	...	...	...	...	...	...	...	...	...	...
2219	213550	Service Availability	04-02-15	04-Feb-15	9:13:18 AM	Customer Care Call	Youngstown	Florida	32466	Closed	No
2220	318775	Comcast Monthly Billing for Returned Modem	06-02-15	06-Feb-15	1:24:39 PM	Customer Care Call	Ypsilanti	Michigan	48197	Solved	No
2221	331188	complaint about comcast	06-09-15	06-Sep-15	5:28:41 PM	Internet	Ypsilanti	Michigan	48197	Solved	No
2222	360489	Extremely unsatisfied Comcast customer	23-06-15	23-Jun-15	11:13:30 PM	Customer Care Call	Ypsilanti	Michigan	48197	Solved	No
2223	363614	Comcast, Ypsilanti MI Internet Speed	24-06-15	24-Jun-15	10:28:33 PM	Customer Care Call	Ypsilanti	Michigan	48198	Open	Yes

2224 rows × 11 columns

In [3]:
```python
# 2> Provide the trend chart for the number of complaints at monthly and daily granularity levels.
--------------------------------------------------------------------------------------------
```

```
  File "<ipython-input-3-d24b6d7714ff>", line 2
    ------------------------------------------------------------------------------------------
                                                                                             ^
SyntaxError: invalid syntax
```

In [4]:
```python
# change the to date format and add new columns Date_formatted and Month
comcast_df['Date_formatted']= pd.to_datetime(comcast_df.Date,format = '%d-%m-%y')
comcast_df['Month']=comcast_df.Date_formatted.dt.month
comcast_df
```

Out[4]:

	Ticket #	Customer Complaint	Date	Date_month_year	Time	Received Via	City	State	Zip code	Status	Filing on Behalf of Someone	Date_formatted	Mon
0	250635	Comcast Cable Internet Speeds	22-04-15	22-Apr-15	3:53:50 PM	Customer Care Call	Abingdon	Maryland	21009	Closed	No	2015-04-22	
1	223441	Payment disappear - service got disconnected	04-08-15	04-Aug-15	10:22:56 AM	Internet	Acworth	Georgia	30102	Closed	No	2015-08-04	
2	242732	Speed and Service	18-04-15	18-Apr-15	9:55:47 AM	Internet	Acworth	Georgia	30101	Closed	Yes	2015-04-18	
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	2015-07-05	
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	2015-05-26	
...	...	...	...	...	...	...	...	...	...	...	...	...	
2219	213550	Service Availability	04-02-15	04-Feb-15	9:13:18 AM	Customer Care Call	Youngstown	Florida	32466	Closed	No	2015-02-04	
2220	318775	Comcast Monthly Billing for Returned Modem	06-02-15	06-Feb-15	1:24:39 PM	Customer Care Call	Ypsilanti	Michigan	48197	Solved	No	2015-02-06	
2221	331188	complaint about comcast	06-09-15	06-Sep-15	5:28:41 PM	Internet	Ypsilanti	Michigan	48197	Solved	No	2015-09-06	
2222	360489	Extremely unsatisfied Comcast customer	23-06-15	23-Jun-15	11:13:30 PM	Customer Care Call	Ypsilanti	Michigan	48197	Solved	No	2015-06-23	
2223	363614	Comcast, Ypsilanti MI Internet Speed	24-06-15	24-Jun-15	10:28:33 PM	Customer Care Call	Ypsilanti	Michigan	48198	Open	Yes	2015-06-24	

2224 rows × 13 columns

In [5]:
```python
# Number of complaints/day
comcast_count_day =pd.DataFrame(comcast_df.groupby('Date_formatted').agg({'Customer Complaint':'count'}))
comcast_count_day=comcast_count_day.reset_index()
comcast_count_day
```

Out[5]:

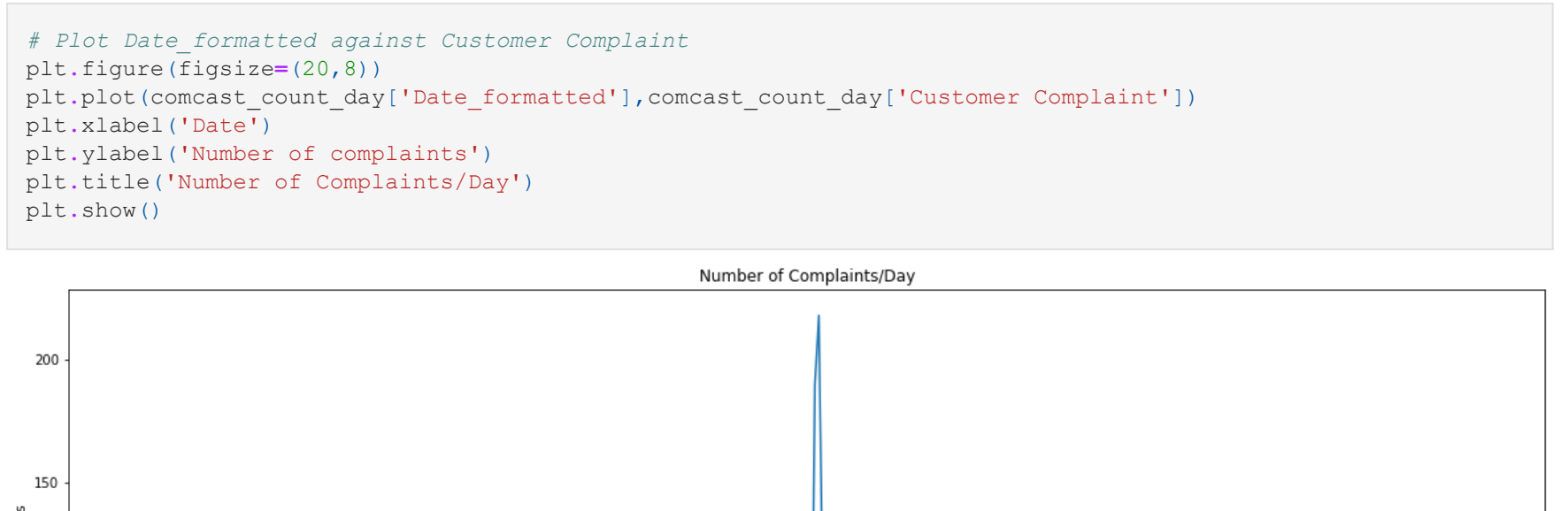
	Date_formatted	Customer Complaint
0	2015-01-04	18
1	2015-01-05	12
2	2015-01-06	25
3	2015-02-04	27
4	2015-02-05	7
...	...	...
86	2015-11-05	12
87	2015-11-06	21
88	2015-12-04	15
89	2015-12-05	7
90	2015-12-06	43

91 rows × 2 columns

In [6]:
```python
print('The date when maximum complaints were registered was: ',comcast_count_day.iloc[comcast_count_day['Custom
```

```
The date when maximum complaints were registered was:  2015-06-24 00:00:00
```

In [7]:
```python
# Plot Date_formatted against Customer Complaint
plt.figure(figsize=(20,8))
plt.plot(comcast_count_day['Date_formatted'],comcast_count_day['Customer Complaint'])
plt.xlabel('Date')
plt.ylabel('Number of complaints')
plt.title('Number of Complaints/Day')
plt.show()
```



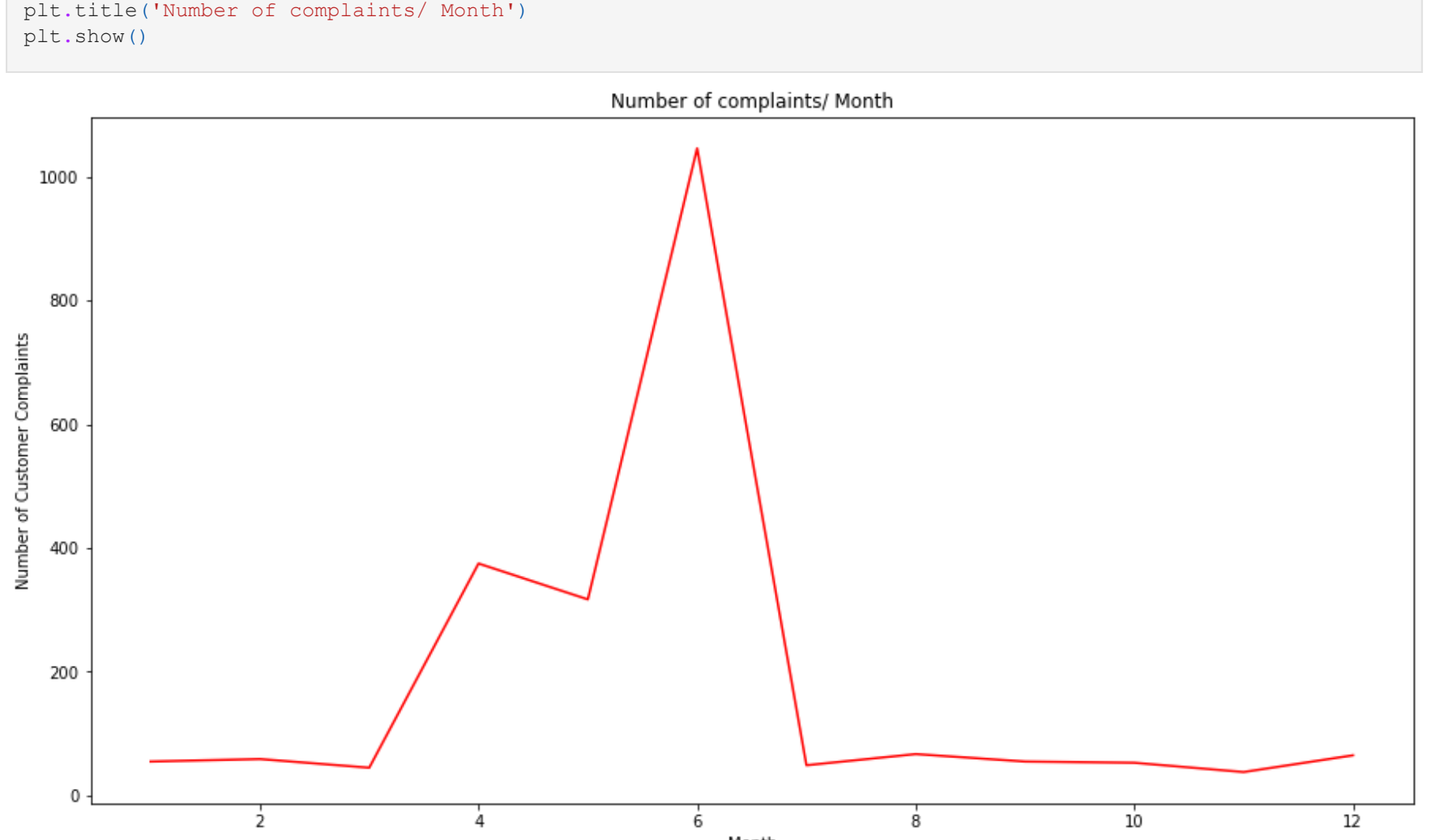
In [ ]:
```python
# From the plot we can see  there is a sudden hype in complaints in the end of July ,
# there might be some Technical fault around this time.
```

In [8]:
```python
# Number of Complaints/Month
comcast_count_month = comcast_df.groupby('Month').agg({'Customer Complaint':'count'})
comcast_count_month
```

Out[8]:

Month	Customer Complaint
1	55
2	59
3	45
4	375
5	317
6	1046
7	49
8	67
9	55
10	53
11	38
12	65

In [9]:
```python
plt.figure(figsize=(15,8))
plt.plot(comcast_count_month,color = 'Red',)
plt.xlabel('Month')
plt.ylabel('Number of Customer Complaints')
plt.title('Number of complaints/ Month')
plt.show()
```



In [ ]:
```python
# From the plot above it is found that from the month of march onwards the complaints are increasing
# till July and then there is a sudden decrease in complaints.
```

In [ ]:
```python
# 3>  Provide a table with the frequency of complaint types.
------------------------------------------------------------
```

In [10]:
```python
comcast_complaint_type = comcast_df.groupby('Customer Complaint').agg({'Customer Complaint':'count'})
comcast_complaint_type = comcast_complaint_type.rename(columns ={'Customer Complaint':'Count of Customer Compla
comcast_complaint_type.reset_index(inplace=True)


comcast_complaint_type = comcast_complaint_type.sort_values(by =['Count of Customer Complaint'],ascending=False
comcast_complaint_type.head(10)
```

Out[10]:

	index	Customer Complaint	Count of Customer Complaint
0	160	Comcast	83
1	331	Comcast Internet	18
2	266	Comcast Data Cap	17
3	1631	comcast	13
4	208	Comcast Billing	11
5	273	Comcast Data Caps	11
6	906	Data Caps	11
7	1530	Unfair Billing Practices	9
8	569	Comcast data cap	8
9	575	Comcast data caps	8

In [ ]:
```python
#Q> Which complaint types are maximum i.e., around internet, network issues, or across any other domains.
#ANS> Comcast,Comcast Internet,Comcast Data Cap are having more complaints
```

In [ ]:
```python
#-4> Create a new categorical variable  Reset-Status with value as Open and Closed.
#Open & Pending is to be categorized as Open and Closed & Solved is to be categorized as Closed.
#------------------------------------------------------------------------------------------------
```

In [11]:
```python
conditions = [(comcast_df['Status']=='Open'),(comcast_df['Status']=='Pending'),(comcast_df['Status']=='Closed')
    (comcast_df['Status']=='Solved')]
choices = ['Open','Open','Closed','Closed']
comcast_df['Reset_Status']= np.select(conditions,choices)
comcast_df
```

Out[11]:

	Ticket #	Customer Complaint	Date	Date_month_year	Time	Received Via	City	State	Zip code	Status	Filing on Behalf of Someone	Date_formatted	Mon
0	250635	Comcast Cable Internet Speeds	22-04-15	22-Apr-15	3:53:50 PM	Customer Care Call	Abingdon	Maryland	21009	Closed	No	2015-04-22	
1	223441	Payment disappear - service got disconnected	04-08-15	04-Aug-15	10:22:56 AM	Internet	Acworth	Georgia	30102	Closed	No	2015-08-04	
2	242732	Speed and Service	18-04-15	18-Apr-15	9:55:47 AM	Internet	Acworth	Georgia	30101	Closed	Yes	2015-04-18	
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	2015-07-05	
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	2015-05-26	
...	...	...	...	...	...	...	...	...	...	...	...	...	
2219	213550	Service Availability	04-02-15	04-Feb-15	9:13:18 AM	Customer Care Call	Youngstown	Florida	32466	Closed	No	2015-02-04	
2220	318775	Comcast Monthly Billing for Returned Modem	06-02-15	06-Feb-15	1:24:39 PM	Customer Care Call	Ypsilanti	Michigan	48197	Solved	No	2015-02-06	
2221	331188	complaint about comcast	06-09-15	06-Sep-15	5:28:41 PM	Internet	Ypsilanti	Michigan	48197	Solved	No	2015-09-06	
2222	360489	Extremely unsatisfied Comcast customer	23-06-15	23-Jun-15	11:13:30 PM	Customer Care Call	Ypsilanti	Michigan	48197	Solved	No	2015-06-23	
2223	363614	Comcast, Ypsilanti MI Internet Speed	24-06-15	24-Jun-15	10:28:33 PM	Customer Care Call	Ypsilanti	Michigan	48198	Open	Yes	2015-06-24	

2224 rows × 14 columns

In [ ]:
```python
 # 5> Provide state wise status of complaints in a stacked bar chart. Use the categorized variable from Q3.
   ----------------------------------------------------------------------------------------------
```

In [12]:
```python
comcast_complaint_state = comcast_df.groupby(['State','Reset_Status']).agg({'Customer Complaint':'count'})
comcast_complaint_state
comcast_complaint_state.head(10)
```

Out[12]:

State	Reset_Status	Customer Complaint
Alabama	Closed	17
	Open	9
Arizona	Closed	14
	Open	6
Arkansas	Closed	6
California	Closed	159
	Open	61
Colorado	Closed	58
	Open	22
Connecticut	Closed	9

In [13]:
```python
comcast_close_open = comcast_complaint_state.unstack(level=-1,fill_value=0)
comcast_close_open.head(10)
```

Out[13]:

	Customer Complaint	
Reset_Status	Closed	Open
State		
Alabama	17	9
Arizona	14	6
Arkansas	6	0
California	159	61
Colorado	58	22
Connecticut	9	3
Delaware	8	4
District Of Columbia	14	2
District of Columbia	1	0
Florida	201	39

In [ ]:
```python
comcast_close_open.plot(kind = 'bar',figsize= (20,10),stacked = True)
```

In [ ]:
```python
# 6> Which state has the maximum complaints
--------------------------------------------
```

In [14]:
```python
comcast_total = comcast_close_open.sum(level=0,axis=1).rename(columns = {'Customer Complaint':'Grand_Total'},in
comcast_close_open = pd.concat([comcast_close_open,comcast_total],axis=1)
comcast_close_open = comcast_close_open.sort_values('Grand_Total',ascending=False)
comcast_close_open.head(10)
```

Out[14]:

State	(Customer Complaint, Closed)	(Customer Complaint, Open)	Grand_Total
Georgia	208	80	288
Florida	201	39	240
California	159	61	220
Illinois	135	29	164
Tennessee	96	47	143
Pennsylvania	110	20	130
Michigan	92	23	115
Washington	75	23	98
Colorado	58	22	80
Maryland	63	15	78

In [ ]:
```python
print('The State with maximum complaints is: ', comcast_close_open.index[0])
```

In [ ]:
```python
# 7> Which state has the highest percentage of unresolved complaints
--------------------------------------------------------------------
```

In [15]:
```python
comcast_close_open['Unsolved_Complaints_%'] = round(comcast_close_open[('Customer Complaint', 'Open')]/comcast_
comcast_close_open.head(10)
```

Out[15]:

State	(Customer Complaint, Closed)	(Customer Complaint, Open)	Grand_Total	Unsolved_Complaints_%
Georgia	208	80	288	28.0
Florida	201	39	240	16.0
California	159	61	220	28.0
Illinois	135	29	164	18.0
Tennessee	96	47	143	33.0
Pennsylvania	110	20	130	15.0
Michigan	92	23	115	20.0
Washington	75	23	98	23.0
Colorado	58	22	80	28.0
Maryland	63	15	78	19.0

In [16]:
```python
comcast_close_open = comcast_close_open.sort_values('Unsolved_Complaints_%',ascending = False)
comcast_close_open.head(10)
```

Out[16]:

State	(Customer Complaint, Closed)	(Customer Complaint, Open)	Grand_Total	Unsolved_Complaints_%
Kansas	1	1	2	50.0
Kentucky	4	3	7	43.0
Mississippi	23	16	39	41.0
Maine	3	2	5	40.0
Alabama	17	9	26	35.0
New Hampshire	8	4	12	33.0
Tennessee	96	47	143	33.0
Vermont	2	1	3	33.0
Delaware	8	4	12	33.0
Texas	49	22	71	31.0

In [ ]:
```python
print('The State with highest percentage of unresolved Complaints  is :',comcast_close_open.index[0])
```

In [ ]:
```python
# 8> Provide the percentage of complaints resolved till date, which were received through the Internet and cust
------------------------------------------------------------------------------------------------------------
```

In [17]:
```python
comcast_resolved =pd.crosstab(comcast_df['Received Via'],comcast_df['Reset_Status'],margins=True,margins_name='
comcast_resolved
```

Out[17]:

Reset_Status	Closed	Open	Grand_Total
Received Via			
Customer Care Call	864	255	1119
Internet	843	262	1105
Grand_Total	1707	517	2224

In [ ]:
```python
comcast_resolved['Resolved_till_date_%']=round(comcast_resolved['Closed']/comcast_resolved['Grand_Total']*100)
comcast_resolved
```