PYTHON PROJECT

ON

Comcast Telecom Consumer Complaints

Submitted

By

UTPALA MOHAPATRA

PROJECT DESCRIPTION

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
- Zipcode: Customer zip
- Status: Status of complaint
- Filing on behalf of someone

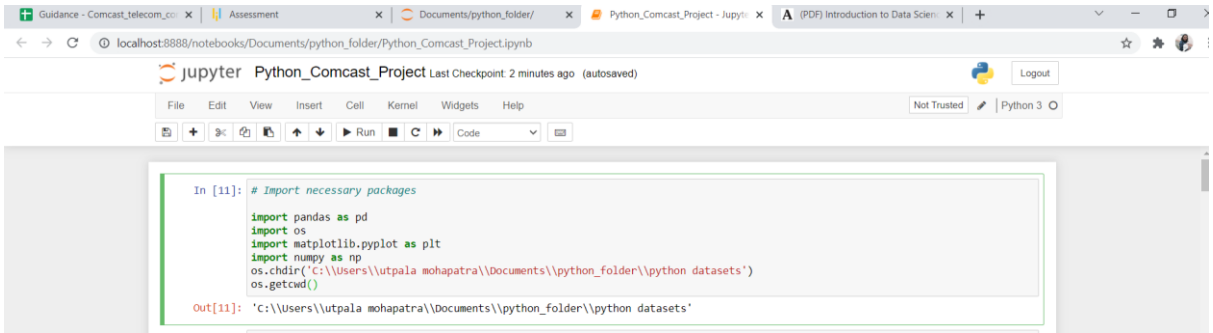
ANALYSIS OF TASK

Problem statement – 1>

Import data into Python environment.

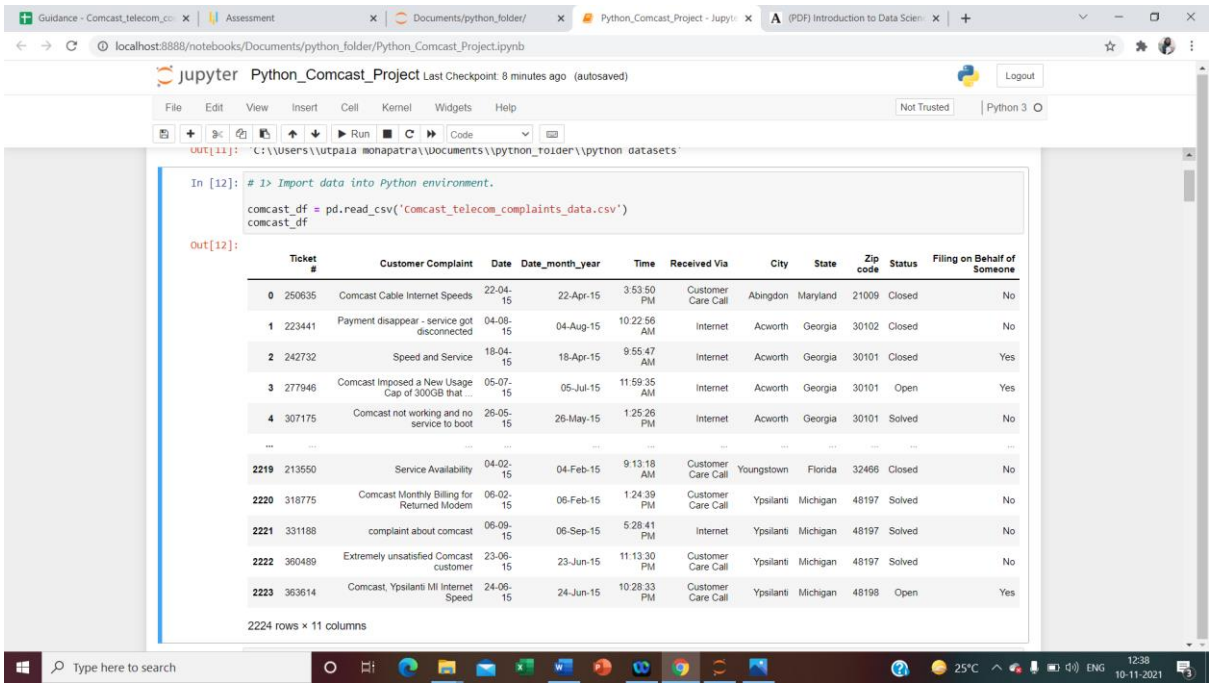
Step 1 – Import necessary Packages

Packages used are Pandas,Matplotlib,Numpy,Os



Step 2 – Import data to python environment

Use read.csv function to import the comcast file



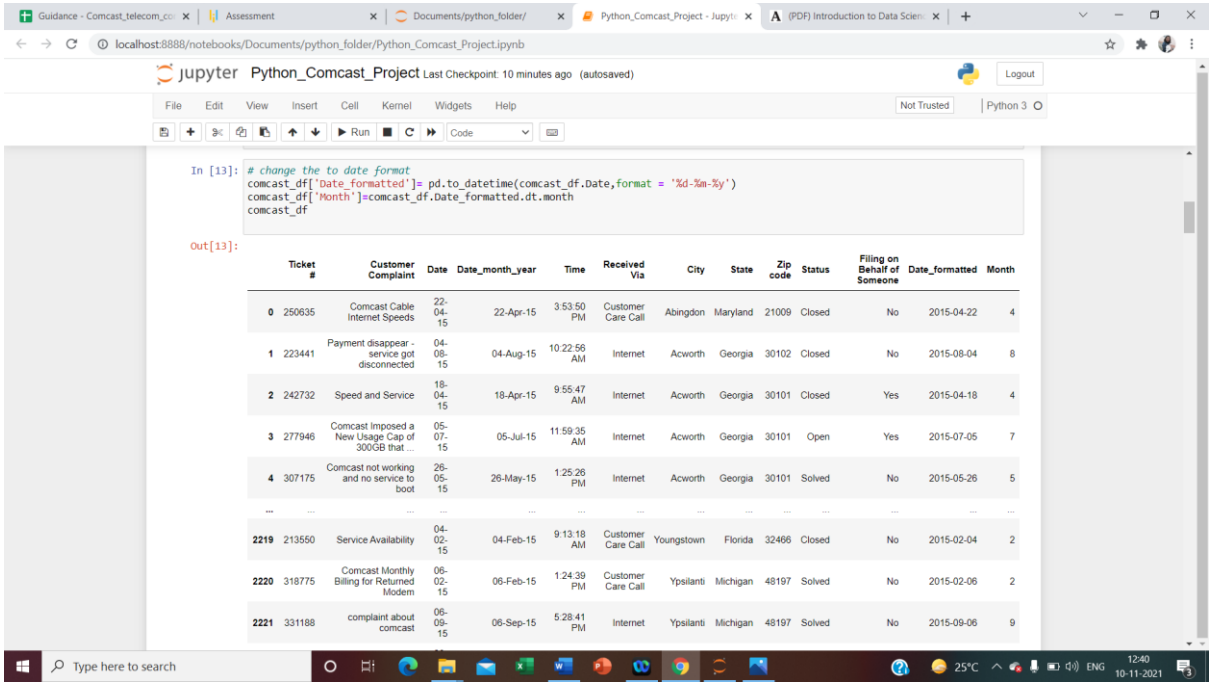
	Ticket #	Customer Complaint	Date	Date_month_year	Time	Received Via	City	State	Zip code	Status	Filing on Behalf of Someone
0	250635	Comcast Cable Internet Speeds	22-04-15	22-Apr-15	3:53:50 PM	Customer Care Call	Abingdon	Maryland	21009	Closed	No
1	223441	Payment disappear - service got disconnected	04-08-15	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

Problem statement -2>

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

Step 1 – Change Date to date format

By using pandas to_datetime function convert Date variable to date format. Then add two new Variables Date_formatted and Month to the comcast_df Dataframe.



	Ticket #	Customer Complaint	Date	Date_month_year	Time	Received Via	City	State	Zip code	Status	Filing on Behalf of Someone	Date_formatted	Month
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	4
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	8
2	242732	Speed and Service	18-04-15	18-Apr-15	9:55:47 AM	Internet	Acworth	Georgia	30101	Closed	Yes	2015-04-18	4
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	7
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	5
...	...	...	...	...	...	...	...	...	...	...	...	...	...
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	2
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	2
2221	331188	complaint about comcast	06-09-15	06-Sep-15	5:28:41 PM	Internet	Ypsilanti	Michigan	48197	Solved	No	2015-09-06	9

Step 2 – Find Number of complaints per Day

By using groupby().agg() with Date_formatted and Customer Complaint got the new dataframe comcast_count_day showing number of complaints per day.

```
In [14]: # Number of complaints/day
         comcast_count_day =pd.DataFrame(comcast_df.groupby('Date_formatted').agg({'Customer Complaint':'count'}))
         comcast_count_day
```

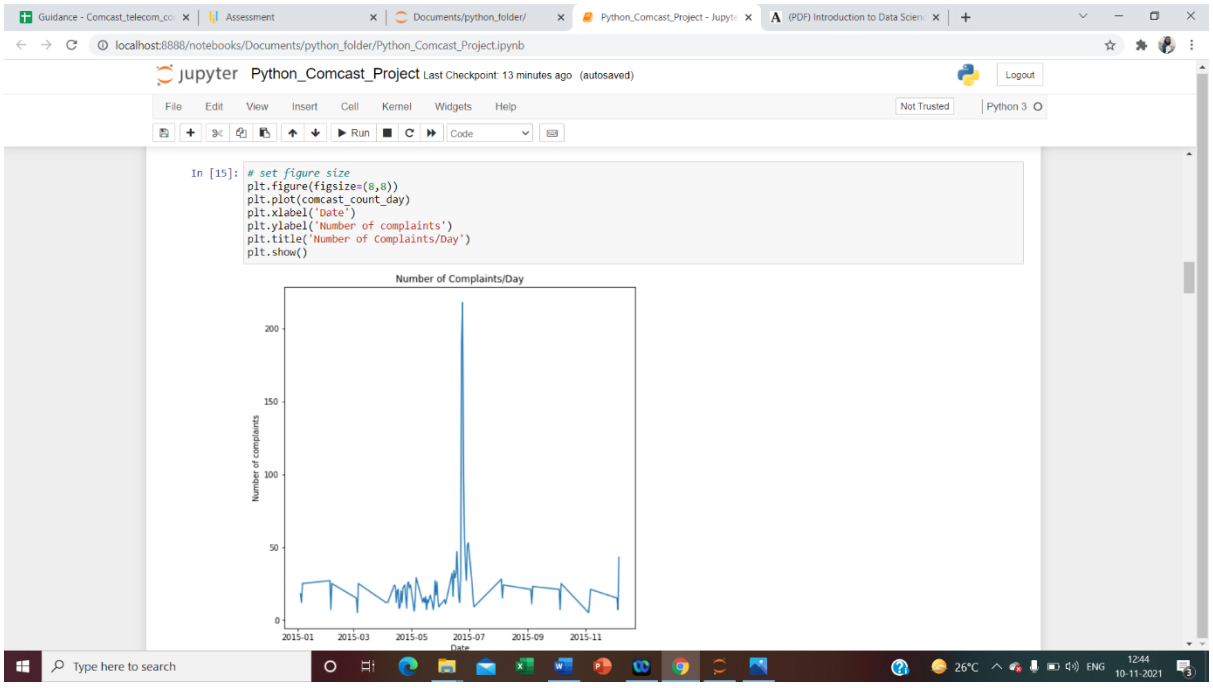
Out[14]:

Date_formatted	Customer Complaint
2015-01-04	18
2015-01-05	12
2015-01-06	25
2015-02-04	27
2015-02-05	7
...	...
2015-11-05	12
2015-11-06	21
2015-12-04	15
2015-12-05	7
2015-12-06	43

91 rows × 1 columns

Step 3 – Ploting Number of complaints per day

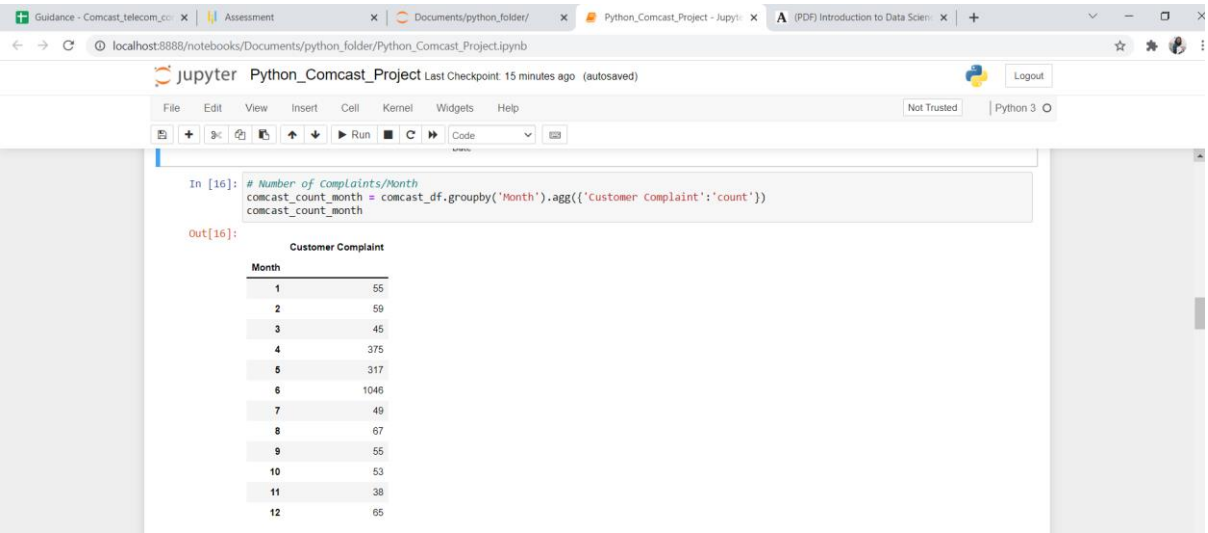
By using plot function of matplotlib package number of complaints per day can be visualized.



Insights - 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.

Step 4 – Find Number of complaints per Month

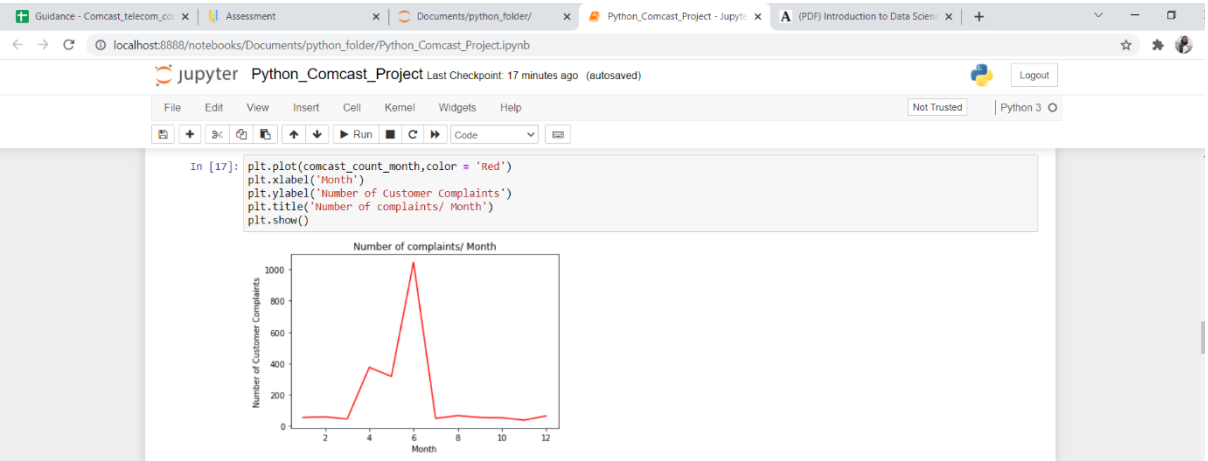
By using groupby().agg() with Month and Customer Complaint got the new dataframe comcast_count_month showing number of complaints per month.



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

Step 5 – Ploting Number of complaints per month

By using plot function of matplotlib package number of complaints per month can be visualized.



Insights- 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.

Problem statement -3>

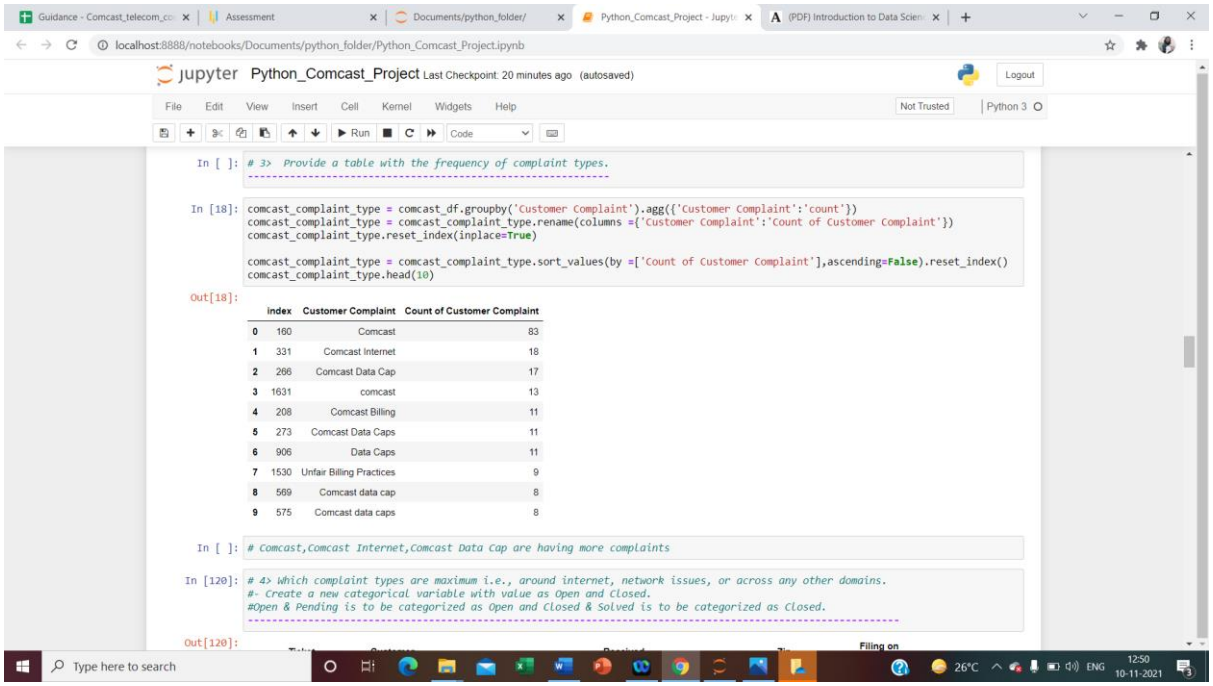
Provide a table with the frequency of complaint types. And find which complaint types are maximum?

Step 1 – Create a table showing complaint types and their frequency

By using groupby on ‘Customer Compalint’ variable and aggregate function on ‘count ‘ , the frequency of the complaint types can be shown. Rename() is used for renaming the frequency as ‘count of customer complaint’.

Step 2 –Sorting the frequency in descendind order to find the complaint types with maximum frequency

By using sort_values() on ‘count of customer complaint’ in descending order ,the complaint types with maximum frequency can be shown.



	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

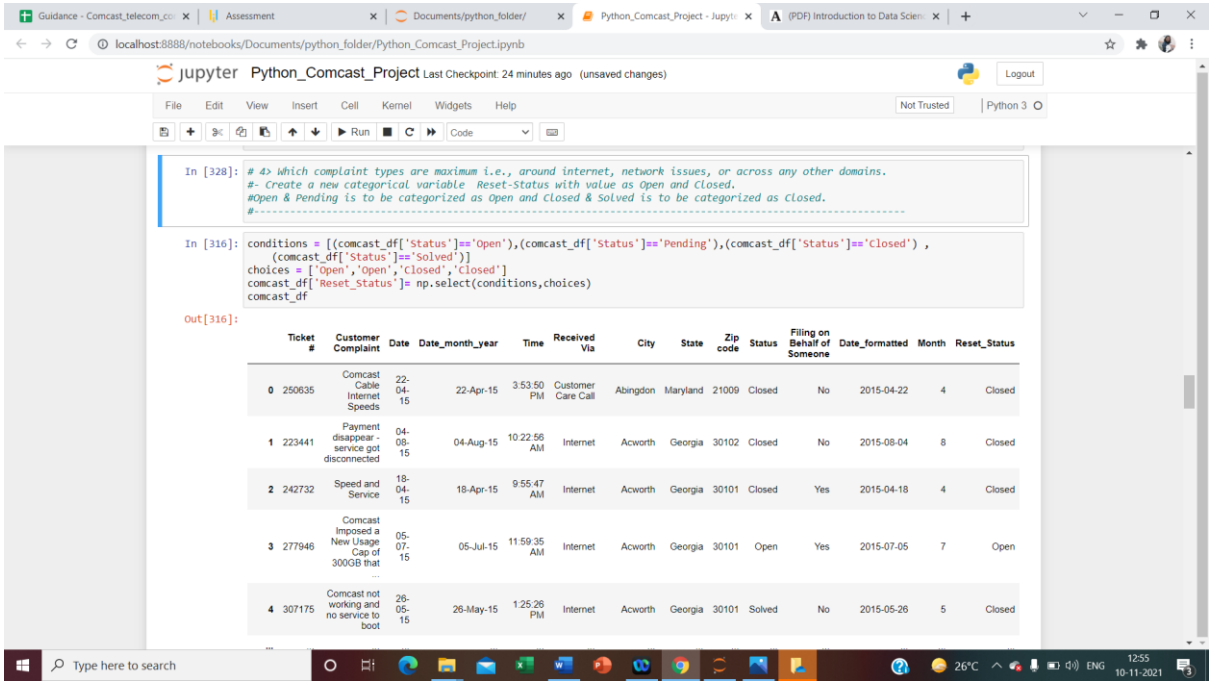
From the table above it can be seen that comcast, comcast internet and comcast data cap are having maximum complaints.

Problem statement -4>

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.

Step 1 – Create new variable Reset_status

Two new variables 'conditions' and 'choices' has been created and then numpy.select method is applied on these two variables to create the new variable 'Reset_Status which shows all the 'Open' and 'Pending 'as 'Open' and all the 'Closed' and 'Solved' as 'Closed'.



	Ticket #	Customer Complaint	Date	Date_month_year	Time	Received Via	City	State	Zip code	Status	Filing on Behalf of Someone	Date_formatted	Month	Reset_Status
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	4	Closed
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	8	Closed
2	242732	Speed and Service	18-04-15	18-Apr-15	9:55:47 AM	Internet	Acworth	Georgia	30101	Closed	Yes	2015-04-18	4	Closed
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	7	Open
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	5	Closed

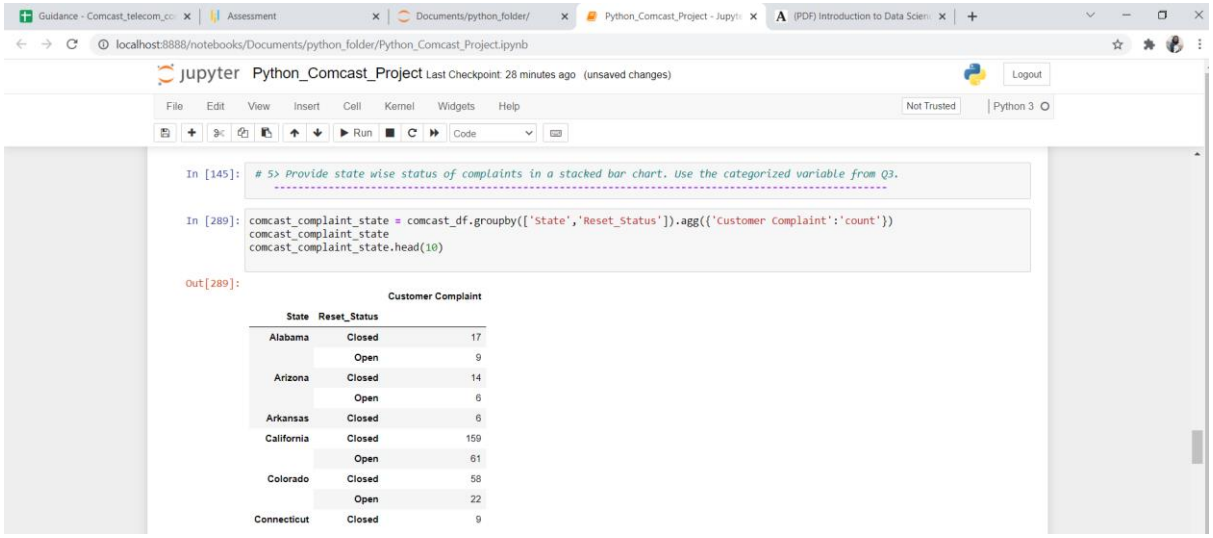
Problem statement -5>

Provide state wise status of complaints in a stacked bar chart. Use the categorized variable from Q4. Provide insights on:

- Which state has the maximum complaints
- Which state has the highest percentage of unresolved complaints

Step 1 –Find the frequency of variable Reset_Status for every State

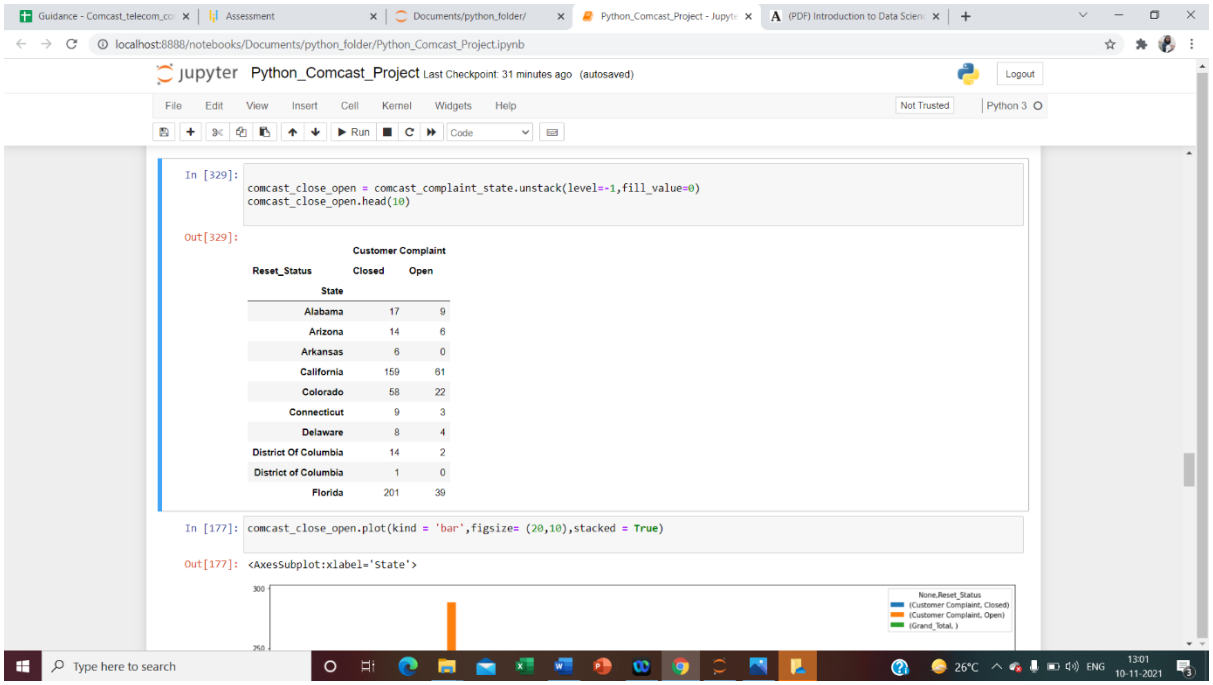
By applying groupby on ‘State’ and ‘Reset_Status’ variable and aggregate function of count on ‘Customer Complaint’ the new table ‘customer_complaint_state is created showing the frequency of Reset_Status variable for each State.



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

Step 2 – Create a new table showing the frequency of ‘Closed’ and ‘Open’ for each State.

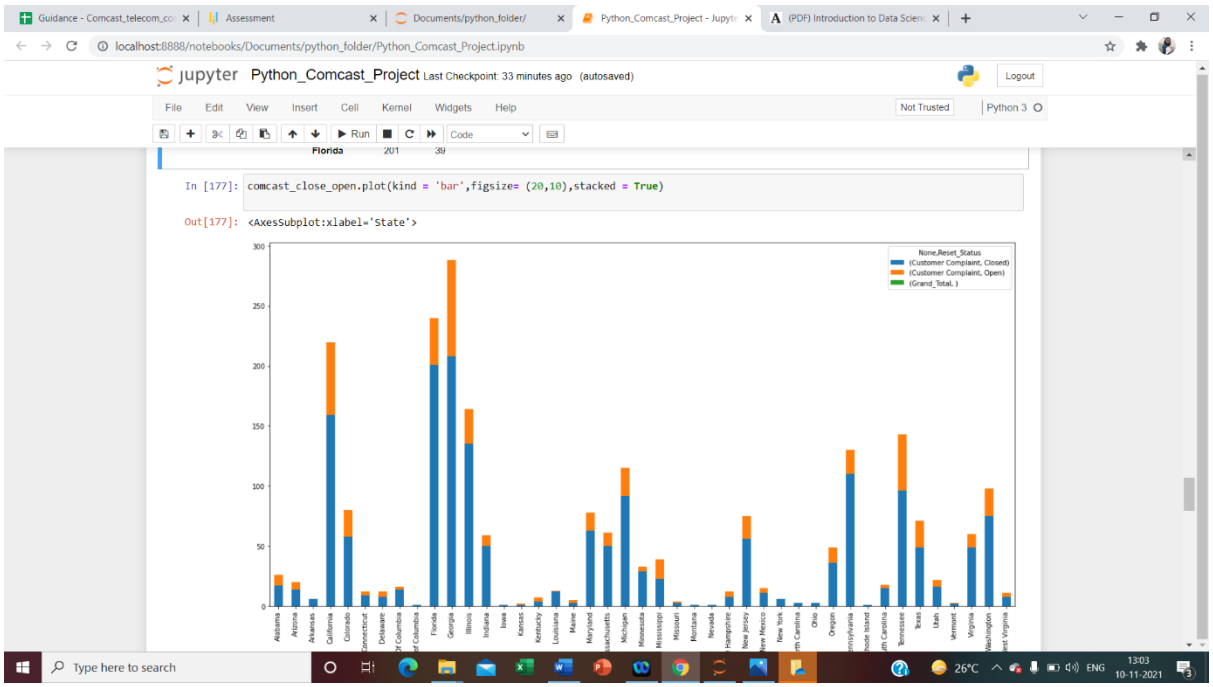
By applying unstack method for level (-1) frequency of ‘Open ‘and ‘Closed’ has been shown in the new table ‘comcast_close_open’ for each state.



Reset_Status	Customer Complaint	
State	Closed	Open
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

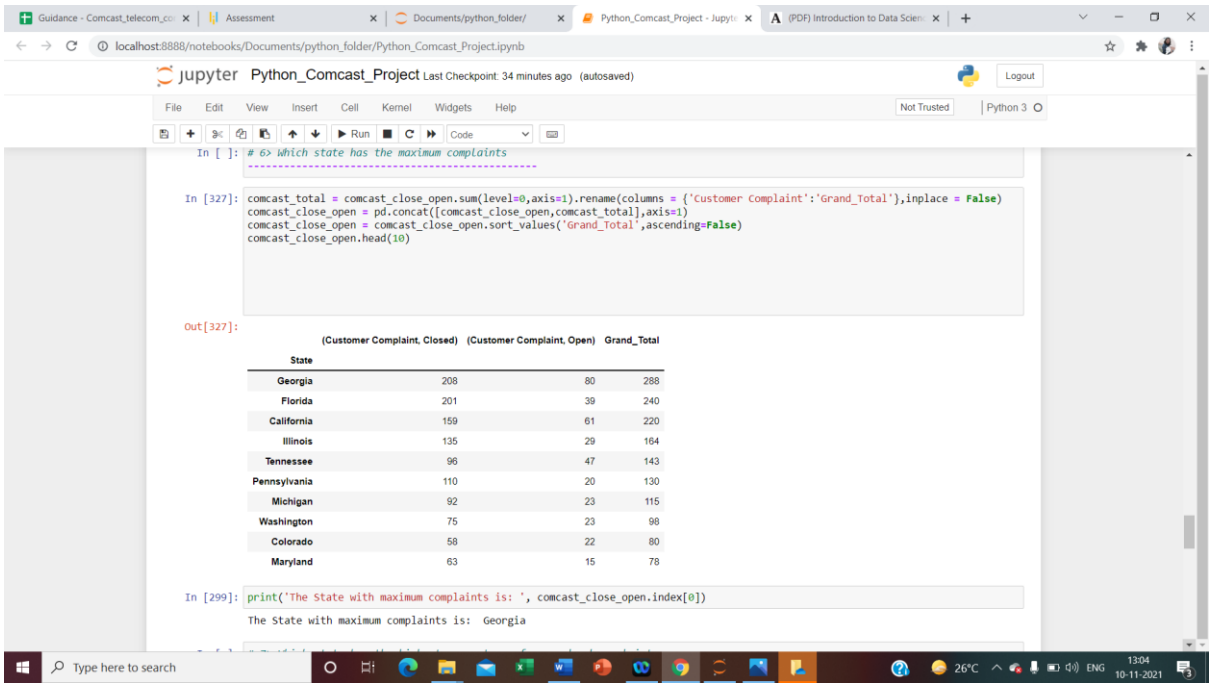
Step 2- Plot a Stacked Bar plot

By applying plot function with kind ‘bar’ and stacked True a stacked bar plot has been visualized for the ‘Open’ and ‘Close’ variable.



Step 3- which State has the maximum Complaints

By applying sum() on level 0 a new vaiable ‘Grand_Total’ is created which is showing the total of ‘Open’ and ‘Closed’ variable for each State then attached to the comcast_close_open dataframe by using pd.concat method. Next the new variable is sorted in descending order by applying sort_value method. Head(10) function is used to show only first 10 records.



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

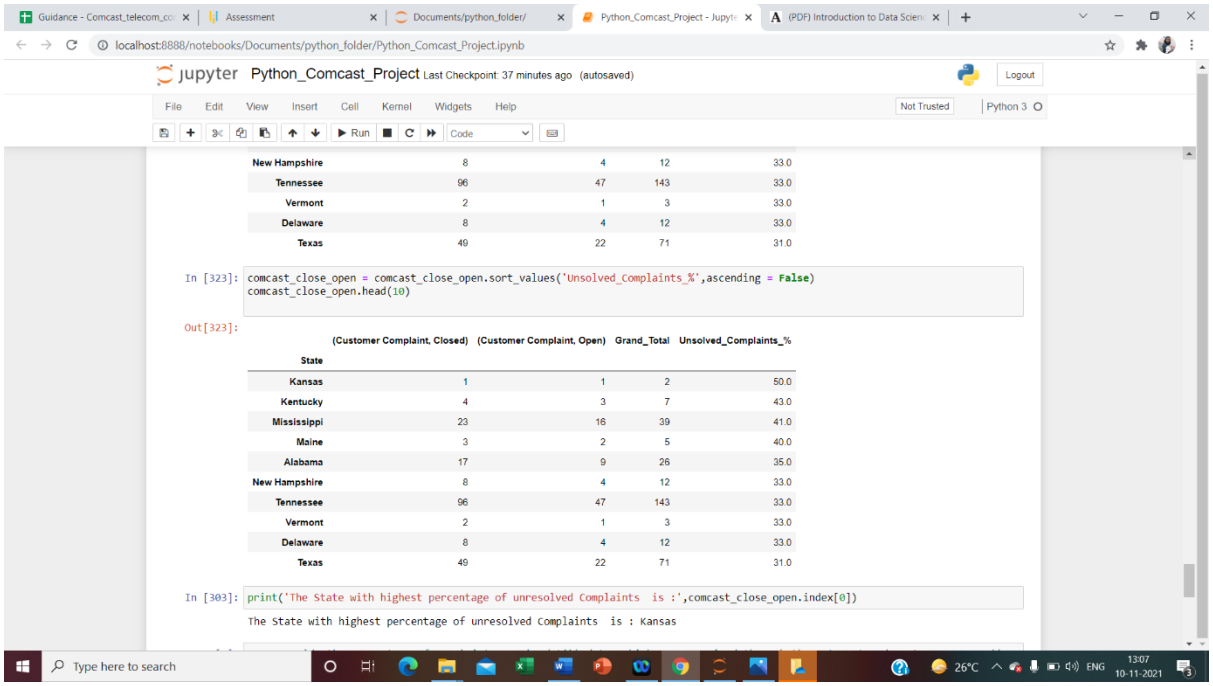
From the above table it is found that The State ‘Georgia’ has the maximum number of complaints i.e 288 complaints.

Step 4 – Which State has the highest percentage of unresolved Complaints

A new variable 'Unsolved_Complaints_%' is created by applying the following formula.

Unsolved_Complaints_% = ((Customer Complaint,Open) / Grand_Total) *100

The value is then rounded off by applying round().
Then the new variable is sorted in descending order by applying sort_value().



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 [323]:
```
comcast_close_open = comcast_close_open.sort_values('Unsolved_Complaints_%',ascending = False)
comcast_close_open.head(10)
```

Out[323]:

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 [303]:
```
print('The State with highest percentage of unresolved Complaints  is :',comcast_close_open.index[0])
```

```
The State with highest percentage of unresolved Complaints  is : Kansas
```

From the above table it is found that the state 'Kansas' has the Highest percentage of unresolved complaints i.e 50%.

Problem statement -6>

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

Step 1-Find the Grand Total of Open and Closed Complaints column wise and rowwise for Customer Care Call and Internet types.

By applying pandas crosstab function on 'Received Via' and 'Reset_Status' variable the total frequency of Open and Closed both row wise and column wise is shown in the new variable 'Grand_Total'.

Step 2- Find the percentage of Complaint resolved till date

By applying the following formula a new variable 'Resolved_till_date_%' is created showing percentage of complaints resolved till date.

Resolved_till_date_% =(Closed / Grand_Total)*100.

Then round() is applied to round off the value.