```
# Import the Module

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# Read the csv file

df=pd.read_csv(r'E:\Data Science with Python Simplilearn\\
Comcast_telecom_complaints_data.csv')
df.head()
```

```
  Ticket #                                  Customer Complaint
Date  \
0   250635                     Comcast Cable Internet Speeds  22-04-
15
1   223441       Payment disappear - service got disconnected  04-08-
15
2   242732                                   Speed and Service  18-04-
15
3   277946  Comcast Imposed a New Usage Cap of 300GB that ...  05-07-
15
4   307175         Comcast not working and no service to boot  26-05-
15

  Date_month_year         Time        Received Via      City     State
\
0       22-Apr-15   3:53:50 PM  Customer Care Call  Abingdon  Maryland

1       04-Aug-15  10:22:56 AM            Internet   Acworth   Georgia

2       18-Apr-15   9:55:47 AM            Internet   Acworth   Georgia

3       05-Jul-15  11:59:35 AM            Internet   Acworth   Georgia

4       26-May-15   1:25:26 PM            Internet   Acworth   Georgia


   Zip code  Status Filing on Behalf of Someone
0     21009  Closed                          No
1     30102  Closed                          No
2     30101  Closed                         Yes
3     30101    Open                         Yes
4     30101  Solved                          No
```

```
# Observing the data information in column wise

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
```

```
# Converting date column from object to datetime format by using
predefined function

df['Date']=pd.to_datetime(df['Date'])
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
```

```
# Adding New Column "month"

df['month']=df['Date'].dt.month_name()
df
```

```
     Ticket #                                  Customer Complaint
Date  \
```

```
0      250635                      Comcast Cable Internet Speeds 2015-
04-22
1      223441       Payment disappear - service got disconnected 2015-
04-08
2      242732                                  Speed and Service 2015-
04-18
3      277946  Comcast Imposed a New Usage Cap of 300GB that ... 2015-
05-07
4      307175         Comcast not working and no service to boot 2015-
05-26
...       ...                                                 ...
...
2219   213550                               Service Availability 2015-
04-02
2220   318775         Comcast Monthly Billing for Returned Modem 2015-
06-02
2221   331188                            complaint about comcast 2015-
06-09
2222   360489             Extremely unsatisfied Comcast customer 2015-
06-23
2223   363614               Comcast, Ypsilanti MI Internet Speed 2015-
06-24

     Date_month_year          Time        Received Via        City
State  \
0          22-Apr-15   3:53:50 PM  Customer Care Call    Abingdon
Maryland
1          04-Aug-15  10:22:56 AM            Internet     Acworth
Georgia
2          18-Apr-15   9:55:47 AM            Internet     Acworth
Georgia
3          05-Jul-15  11:59:35 AM            Internet     Acworth
Georgia
4          26-May-15   1:25:26 PM            Internet     Acworth
Georgia
...              ...          ...                 ...         ...
...
2219       04-Feb-15   9:13:18 AM  Customer Care Call  Youngstown
Florida
2220       06-Feb-15   1:24:39 PM  Customer Care Call   Ypsilanti
Michigan
2221       06-Sep-15   5:28:41 PM            Internet   Ypsilanti
Michigan
2222       23-Jun-15  11:13:30 PM  Customer Care Call   Ypsilanti
Michigan
2223       24-Jun-15  10:28:33 PM  Customer Care Call   Ypsilanti
Michigan

      Zip code  Status Filing on Behalf of Someone  month
0        21009  Closed                          No  April
```

```
1        30102  Closed                          No  April
2        30101  Closed                         Yes  April
3        30101    Open                         Yes    May
4        30101  Solved                          No    May
...        ...     ...                         ...    ...
2219     32466  Closed                          No  April
2220     48197  Solved                          No   June
2221     48197  Solved                          No   June
2222     48197  Solved                          No   June
2223     48198    Open                         Yes   June

[2224 rows x 12 columns]
```

```
# Finding no. of complaints per day

df['Date'].value_counts()
```

```
2015-06-24    218
2015-06-23    190
2015-06-25     98
2015-06-26     55
2015-06-30     53
              ...
2015-05-10      7
2015-05-24      7
2015-04-05      6
2015-04-11      5
2015-05-03      5
Name: Date, Length: 91, dtype: int64
```

```
# Grouping the date size to find the no of complaint in the record

dates=df.groupby('Date').size()
dates
```

```
Date
2015-04-01    18
2015-04-02    27
2015-04-03    15
2015-04-04    12
2015-04-05     6
              ..
2015-06-26    55
2015-06-27    39
2015-06-28    27
2015-06-29    51
2015-06-30    53
Length: 91, dtype: int64
```

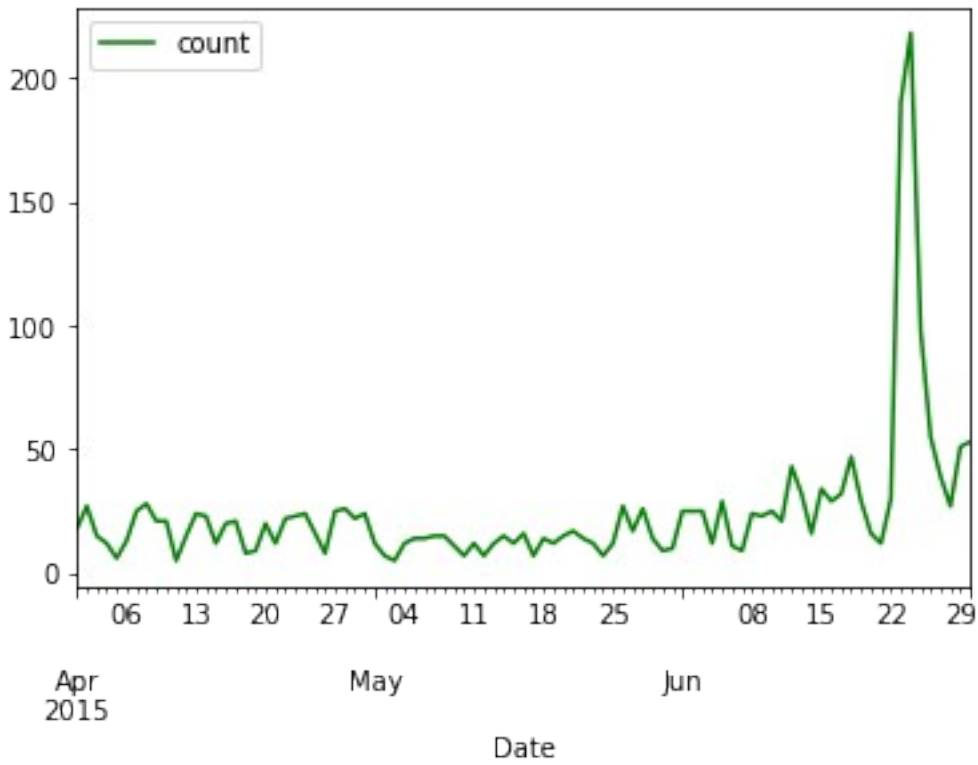
```
# Reset and Renaming the dates column name to find the count of
compalints
```

```
daily=pd.DataFrame(dates).reset_index()
daily=daily.rename(columns={0:'count'})
daily
```

```
         Date  count
0  2015-04-01     18
1  2015-04-02     27
2  2015-04-03     15
3  2015-04-04     12
4  2015-04-05      6
..        ...    ...
86 2015-06-26     55
87 2015-06-27     39
88 2015-06-28     27
89 2015-06-29     51
90 2015-06-30     53

[91 rows x 2 columns]
```

```
# creating plot graph to understand the month wise complaint

daily.plot(x='Date',y='count',kind='line',color='green')
```

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



```
# Grouping the maximum complaint record months of data

month=df.groupby('month').size()
month
```

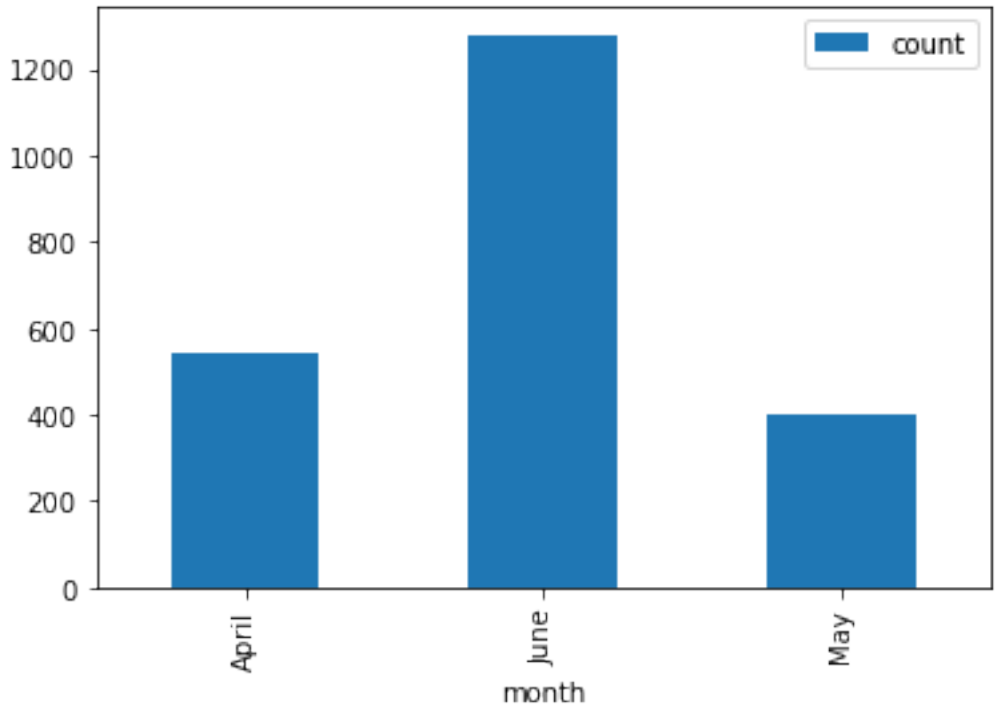
```
month
April     545
June     1280
May       399
dtype: int64
```

```
# Reset and Renaming the month column name to find the count of
compalints

month=pd.DataFrame(month).reset_index()
month=month.rename(columns={0:'count'})
month
```

	month	count
0	April	545
1	June	1280
2	May	399

```
# Using Bar graph for month column to identitfy the no. of complaint
records on the month

month.plot(x='month',y='count',kind='bar')
```

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



```
# Identify the complaint category by using value_counts function

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

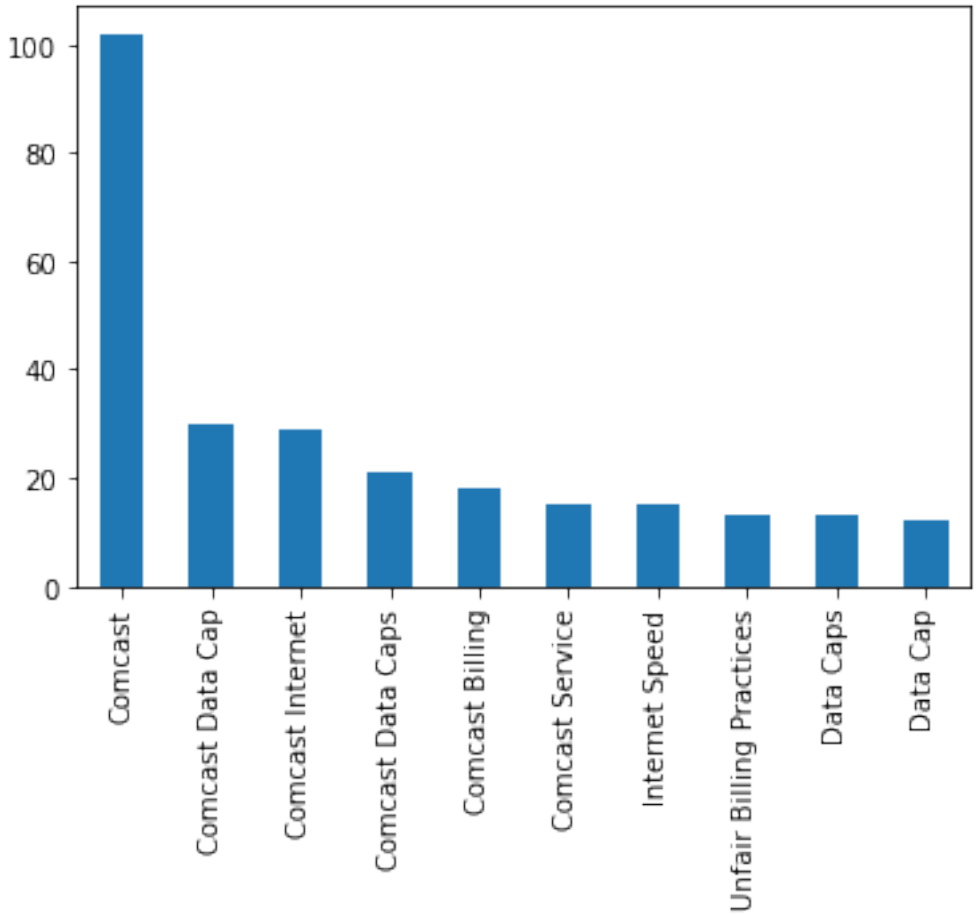
```
Comcast                                             83
Comcast Internet                                    18
Comcast Data Cap                                    17
comcast                                             13
Comcast Billing                                     11
                                                    ..
Improper Billing and non resolution of issues        1
Deceptive trade                                      1
intermittent internet                                1
Internet Speed on Wireless Connection                1
Comcast, Ypsilanti MI Internet Speed                 1
Name: Customer Complaint, Length: 1841, dtype: int64
```

```
# Using Lambda function to changing the string items upper case in
each words

df['Customer Complaint']=df['Customer Complaint'].apply(lambda x :
x.title())
df['Customer Complaint'].value_counts()
```

```
Comcast                                                           102
Comcast Data Cap                                                   30
Comcast Internet                                                   29
Comcast Data Caps                                                  21
Comcast Billing                                                    18
                                                                 ...
Monthly Data Caps                                                   1
Comcast/Xfinity Poor Service, Fraudulent Billing And Collection     1
Lost Emails/Billing                                                 1
Improper Billing And Non Resolution Of Issues                       1
Comcast, Ypsilanti Mi Internet Speed                                1
Name: Customer Complaint, Length: 1740, dtype: int64
```

```
# Finding the most complaints category of "Customer Complaint" by
using plot bar graph

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

```
<AxesSubplot:>
```




```
# Recall the dataset
```

```
df
```

```
     Ticket #                                  Customer Complaint
Date  \
0      250635                       Comcast Cable Internet Speeds 2015-
04-22
1      223441       Payment Disappear - Service Got Disconnected 2015-
04-08
2      242732                                    Speed And Service 2015-
04-18
3      277946  Comcast Imposed A New Usage Cap Of 300Gb That ... 2015-
05-07
4      307175         Comcast Not Working And No Service To Boot 2015-
05-26
...       ...                                                  ...
...
2219   213550                                Service Availability 2015-
04-02
2220   318775         Comcast Monthly Billing For Returned Modem 2015-
06-02
```

```
2221   331188                            Complaint About Comcast 2015-
06-09
2222   360489              Extremely Unsatisfied Comcast Customer 2015-
06-23
2223   363614                Comcast, Ypsilanti Mi Internet Speed 2015-
06-24

     Date_month_year         Time        Received Via        City
State  \
0          22-Apr-15   3:53:50 PM  Customer Care Call    Abingdon
Maryland
1          04-Aug-15  10:22:56 AM            Internet     Acworth
Georgia
2          18-Apr-15   9:55:47 AM            Internet     Acworth
Georgia
3          05-Jul-15  11:59:35 AM            Internet     Acworth
Georgia
4          26-May-15   1:25:26 PM            Internet     Acworth
Georgia
...              ...          ...                 ...         ...
...
2219       04-Feb-15   9:13:18 AM  Customer Care Call  Youngstown
Florida
2220       06-Feb-15   1:24:39 PM  Customer Care Call   Ypsilanti
Michigan
2221       06-Sep-15   5:28:41 PM            Internet   Ypsilanti
Michigan
2222       23-Jun-15  11:13:30 PM  Customer Care Call   Ypsilanti
Michigan
2223       24-Jun-15  10:28:33 PM  Customer Care Call   Ypsilanti
Michigan

      Zip code  Status Filing on Behalf of Someone  month
0        21009  Closed                          No  April
1        30102  Closed                          No  April
2        30101  Closed                         Yes  April
3        30101    Open                         Yes    May
4        30101  Solved                          No    May
...        ...     ...                         ...    ...
2219     32466  Closed                          No  April
2220     48197  Solved                          No   June
2221     48197  Solved                          No   June
2222     48197  Solved                          No   June
2223     48198    Open                         Yes   June

[2224 rows x 12 columns]
```

```
# Identify the complaints by using string contains by keywords

internet_issues1=df[df['Customer
```

```
Complaint'].str.contains('Network')].count()
internet_issues2=df[df['Customer
Complaint'].str.contains('Network')].count()
internet_issues3=df[df['Customer
Complaint'].str.contains('Internet')].count()
total_internet=internet_issues1+internet_issues2+internet_issues3
total_internet
```

```
Ticket #                        535
Customer Complaint              535
Date                            535
Date_month_year                 535
Time                            535
Received Via                    535
City                            535
State                           535
Zip code                        535
Status                          535
Filing on Behalf of Someone     535
month                           535
dtype: int64
```

```
# Identify the complaints by using string contains by keywords

billing_issue1=df[df['Customer
Complaint'].str.contains('Billing')].count()
billing_issue2=df[df['Customer
Complaint'].str.contains('Charges')].count()
total_billing=billing_issue1+billing_issue2
total_billing
```

```
Ticket #                        365
Customer Complaint              365
Date                            365
Date_month_year                 365
Time                            365
Received Via                    365
City                            365
State                           365
Zip code                        365
Status                          365
Filing on Behalf of Someone     365
month                           365
dtype: int64
```

```
# Identify the complaints by using string contains by keywords

service_issue1=df[df['Customer
Complaint'].str.contains('Service')].count()
service_issue2=df[df['Customer
Complaint'].str.contains('Customer')].count()
```

```
total_service=service_issue1+service_issue2
total_service
```

```
Ticket #                        584
Customer Complaint              584
Date                            584
Date_month_year                 584
Time                            584
Received Via                    584
City                            584
State                           584
Zip code                        584
Status                          584
Filing on Behalf of Someone     584
month                           584
dtype: int64
```

```
# Identify the complaints by using string contains by keywords.

other_issue=2224-(total_billing+total_internet+total_service)
other_issue
```

```
Ticket #                        740
Customer Complaint              740
Date                            740
Date_month_year                 740
Time                            740
Received Via                    740
City                            740
State                           740
Zip code                        740
Status                          740
Filing on Behalf of Someone     740
month                           740
dtype: int64
```

```
# Findind the status column unique String values

df['Status'].unique()
```

```
array(['Closed', 'Open', 'Solved', 'Pending'], dtype=object)
```

```
# Modifying the status column by using if else conditon

df['New_Status']=[ "Open" if x=="Open" or x=="Pending" else "Closed"
for x in df['Status']]

df
```

```
     Ticket #                                  Customer Complaint
Date  \
0      250635                    Comcast Cable Internet Speeds 2015-
```

```
04-22
1      223441       Payment Disappear - Service Got Disconnected 2015-
04-08
2      242732                                    Speed And Service 2015-
04-18
3      277946  Comcast Imposed A New Usage Cap Of 300Gb That ... 2015-
05-07
4      307175         Comcast Not Working And No Service To Boot 2015-
05-26
...       ...                                                  ...
...
2219   213550                                Service Availability 2015-
04-02
2220   318775         Comcast Monthly Billing For Returned Modem 2015-
06-02
2221   331188                             Complaint About Comcast 2015-
06-09
2222   360489             Extremely Unsatisfied Comcast Customer 2015-
06-23
2223   363614               Comcast, Ypsilanti Mi Internet Speed 2015-
06-24

     Date_month_year          Time        Received Via        City
State  \
0          22-Apr-15   3:53:50 PM  Customer Care Call    Abingdon
Maryland
1          04-Aug-15  10:22:56 AM            Internet     Acworth
Georgia
2          18-Apr-15   9:55:47 AM            Internet     Acworth
Georgia
3          05-Jul-15  11:59:35 AM            Internet     Acworth
Georgia
4          26-May-15   1:25:26 PM            Internet     Acworth
Georgia
...              ...          ...                 ...         ...
...
2219       04-Feb-15   9:13:18 AM  Customer Care Call  Youngstown
Florida
2220       06-Feb-15   1:24:39 PM  Customer Care Call   Ypsilanti
Michigan
2221       06-Sep-15   5:28:41 PM            Internet   Ypsilanti
Michigan
2222       23-Jun-15  11:13:30 PM  Customer Care Call   Ypsilanti
Michigan
2223       24-Jun-15  10:28:33 PM  Customer Care Call   Ypsilanti
Michigan

      Zip code  Status Filing on Behalf of Someone  month New_Status
0        21009  Closed                          No  April     Closed
1        30102  Closed                          No  April     Closed
```

```
2        30101  Closed                          Yes  April     Closed
3        30101    Open                          Yes    May       Open
4        30101  Solved                           No    May     Closed
...        ...     ...                          ...    ...        ...
2219     32466  Closed                           No  April     Closed
2220     48197  Solved                           No   June     Closed
2221     48197  Solved                           No   June     Closed
2222     48197  Solved                           No   June     Closed
2223     48198    Open                          Yes   June       Open

[2224 rows x 13 columns]
```

```
# Finding the staus of complaint by state wise and fill the NaN by
using fillna()

state_complain=df.groupby(['State','New_Status']).size().unstack().fil
lna(0)
state_complain
```

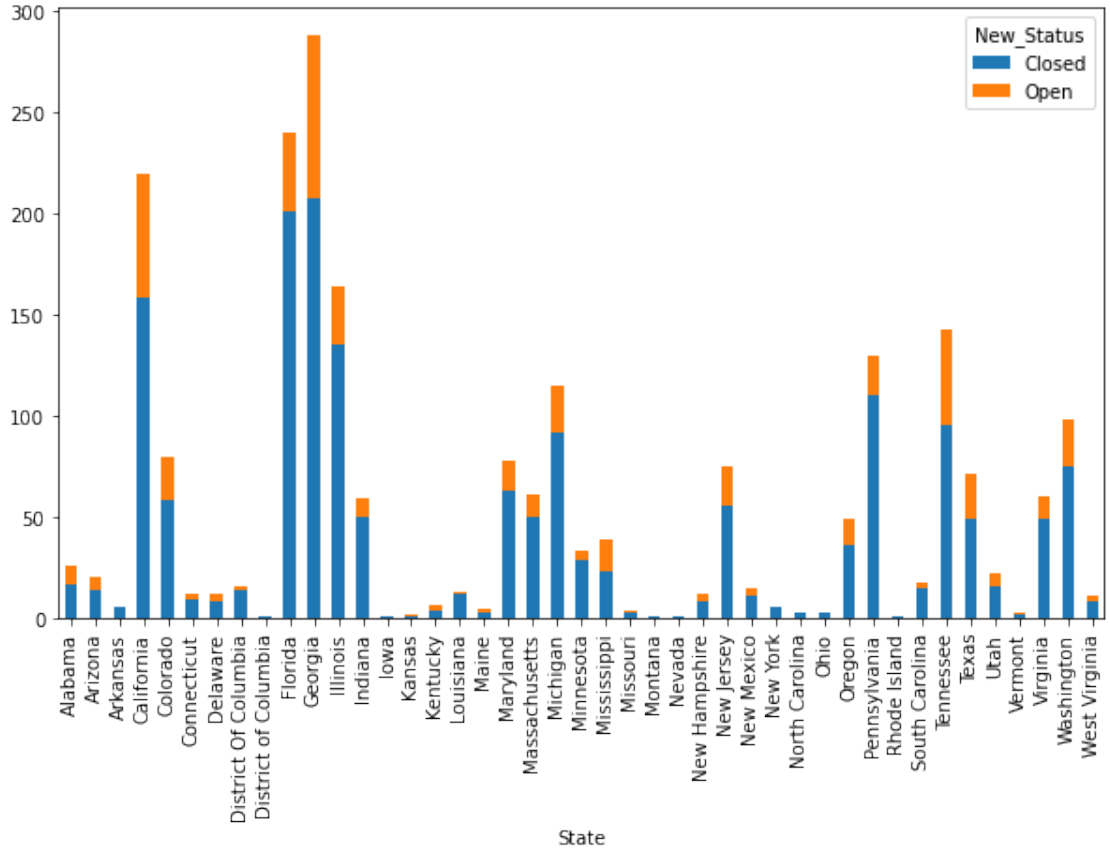
New_Status	Closed	Open
State		
Alabama	17.0	9.0
Arizona	14.0	6.0
Arkansas	6.0	0.0
California	159.0	61.0
Colorado	58.0	22.0
Connecticut	9.0	3.0
Delaware	8.0	4.0
District Of Columbia	14.0	2.0
District of Columbia	1.0	0.0
Florida	201.0	39.0
Georgia	208.0	80.0
Illinois	135.0	29.0
Indiana	50.0	9.0
Iowa	1.0	0.0
Kansas	1.0	1.0
Kentucky	4.0	3.0
Louisiana	12.0	1.0
Maine	3.0	2.0
Maryland	63.0	15.0
Massachusetts	50.0	11.0
Michigan	92.0	23.0
Minnesota	29.0	4.0
Mississippi	23.0	16.0
Missouri	3.0	1.0
Montana	1.0	0.0
Nevada	1.0	0.0
New Hampshire	8.0	4.0
New Jersey	56.0	19.0
New Mexico	11.0	4.0

```
New York                6.0   0.0
North Carolina          3.0   0.0
Ohio                    3.0   0.0
Oregon                 36.0  13.0
Pennsylvania          110.0  20.0
Rhode Island            1.0   0.0
South Carolina         15.0   3.0
Tennessee              96.0  47.0
Texas                  49.0  22.0
Utah                   16.0   6.0
Vermont                 2.0   1.0
Virginia               49.0  11.0
Washington             75.0  23.0
West Virginia           8.0   3.0
```

```
# creating stack bar graph for finding the state wise complaint
```

```
state_complain.plot.bar(stacked=True,figsize=(10,6))
```

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




```
# Findly the top complaint states by using value count function
```

```
df['State'].value_counts()[:5]
```

```
Georgia       288
Florida       240
California    220
Illinois      164
Tennessee     143
Name: State, dtype: int64
```

```
# Grouping the unresolved data by using groupby function.
# filling the missing values by using fillna function.
# sort the 'Open' values by using sort_values function
```

```
unresolved_data=df.groupby(['State','New_Status']).size().unstack().fi
llna(0).sort_values(by='Open')
unresolved_data
```

New_Status	Closed	Open
State		
New York	6.0	0.0
Nevada	1.0	0.0
Arkansas	6.0	0.0
Rhode Island	1.0	0.0
Ohio	3.0	0.0
Iowa	1.0	0.0
District of Columbia	1.0	0.0
North Carolina	3.0	0.0
Montana	1.0	0.0
Louisiana	12.0	1.0
Kansas	1.0	1.0
Missouri	3.0	1.0
Vermont	2.0	1.0
District Of Columbia	14.0	2.0
Maine	3.0	2.0
Connecticut	9.0	3.0
Kentucky	4.0	3.0
South Carolina	15.0	3.0
West Virginia	8.0	3.0
New Hampshire	8.0	4.0
New Mexico	11.0	4.0
Minnesota	29.0	4.0
Delaware	8.0	4.0
Arizona	14.0	6.0
Utah	16.0	6.0
Alabama	17.0	9.0
Indiana	50.0	9.0
Virginia	49.0	11.0
Massachusetts	50.0	11.0
Oregon	36.0	13.0
Maryland	63.0	15.0
Mississippi	23.0	16.0
New Jersey	56.0	19.0

```
Pennsylvania           110.0  20.0
Colorado                58.0  22.0
Texas                   49.0  22.0
Washington              75.0  23.0
Michigan                92.0  23.0
Illinois               135.0  29.0
Florida                201.0  39.0
Tennessee               96.0  47.0
California             159.0  61.0
Georgia                208.0  80.0
```

```
# Showing the unresolved data by using percentage method

unresolved_data['unresolved_cmp_perct']=unresolved_data['Open']/unreso
lved_data['Open'].sum()
unresolved_data
```

```
New_Status            Closed  Open  unresolved_cmp_perct
State
New York                 6.0   0.0              0.000000
Nevada                   1.0   0.0              0.000000
Arkansas                 6.0   0.0              0.000000
Rhode Island             1.0   0.0              0.000000
Ohio                     3.0   0.0              0.000000
Iowa                     1.0   0.0              0.000000
District of Columbia     1.0   0.0              0.000000
North Carolina           3.0   0.0              0.000000
Montana                  1.0   0.0              0.000000
Louisiana               12.0   1.0              0.001934
Kansas                   1.0   1.0              0.001934
Missouri                 3.0   1.0              0.001934
Vermont                  2.0   1.0              0.001934
District Of Columbia    14.0   2.0              0.003868
Maine                    3.0   2.0              0.003868
Connecticut              9.0   3.0              0.005803
Kentucky                 4.0   3.0              0.005803
South Carolina          15.0   3.0              0.005803
West Virginia            8.0   3.0              0.005803
New Hampshire            8.0   4.0              0.007737
New Mexico              11.0   4.0              0.007737
Minnesota               29.0   4.0              0.007737
Delaware                 8.0   4.0              0.007737
Arizona                 14.0   6.0              0.011605
Utah                    16.0   6.0              0.011605
Alabama                 17.0   9.0              0.017408
Indiana                 50.0   9.0              0.017408
Virginia                49.0  11.0              0.021277
Massachusetts           50.0  11.0              0.021277
Oregon                  36.0  13.0              0.025145
Maryland                63.0  15.0              0.029014
```

```
Mississippi             23.0  16.0             0.030948
New Jersey              56.0  19.0             0.036750
Pennsylvania           110.0  20.0             0.038685
Colorado                58.0  22.0             0.042553
Texas                   49.0  22.0             0.042553
Washington              75.0  23.0             0.044487
Michigan                92.0  23.0             0.044487
Illinois               135.0  29.0             0.056093
Florida                201.0  39.0             0.075435
Tennessee               96.0  47.0             0.090909
California             159.0  61.0             0.117988
Georgia                208.0  80.0             0.154739
```

```
# Grouping the complaint by using 'Received Via'and 'New_Status'
column

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

```
New_Status          Closed  Open
Received Via
Customer Care Call     864   255
Internet               843   262
```

```
# Finding the resolved data by using percentage

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

```
Received Via
Customer Care Call    50.615114
Internet              49.384886
Name: resolved, dtype: float64
```