

In [5]:

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

In [6]:

```
data=pd.read_csv("C:/Users/DELL/Documents/Simplilearn Python/Comcast_telecom_complaints_data/Comcast_telecom_complaints_d
```

In [7]:

```
data.head()
```

Out[7]:

	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

In [8]:

```
data['Date']=pd.to_datetime(data['Date'])
data['Month']=data['Date'].dt.month_name()
```

In [9]:

```
data.head()
```

Out[9]:

	Ticket #	Customer Complaint	Date	Date_month_year	Time	Received Via	City	State	Zip code	Status	Filing on Behalf of Someone	Month
0	250635	Comcast Cable Internet Speeds	2015-04-22	22-Apr-15	3:53:50 PM	Customer Care Call	Abingdon	Maryland	21009	Closed	No	April
1	223441	Payment disappear - service got disconnected	2015-04-08	04-Aug-15	10:22:56 AM	Internet	Acworth	Georgia	30102	Closed	No	April
2	242732	Speed and Service	2015-04-18	18-Apr-15	9:55:47 AM	Internet	Acworth	Georgia	30101	Closed	Yes	April
3	277946	Comcast Imposed a New Usage Cap of 300GB that ...	2015-05-07	05-Jul-15	11:59:35 AM	Internet	Acworth	Georgia	30101	Open	Yes	May
4	307175	Comcast not working and no service to boot	2015-05-26	26-May-15	1:25:26 PM	Internet	Acworth	Georgia	30101	Solved	No	May

In [10]:

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

Out[10]:

```
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
```

In [11]:

```
#Chart for date-wise complaints
```

In [12]:

```
datewise_group=data.groupby(['Date'])['Ticket #'].count()
datewise_df=pd.DataFrame(datewise_group).reset_index()
datewise_df.head()
```

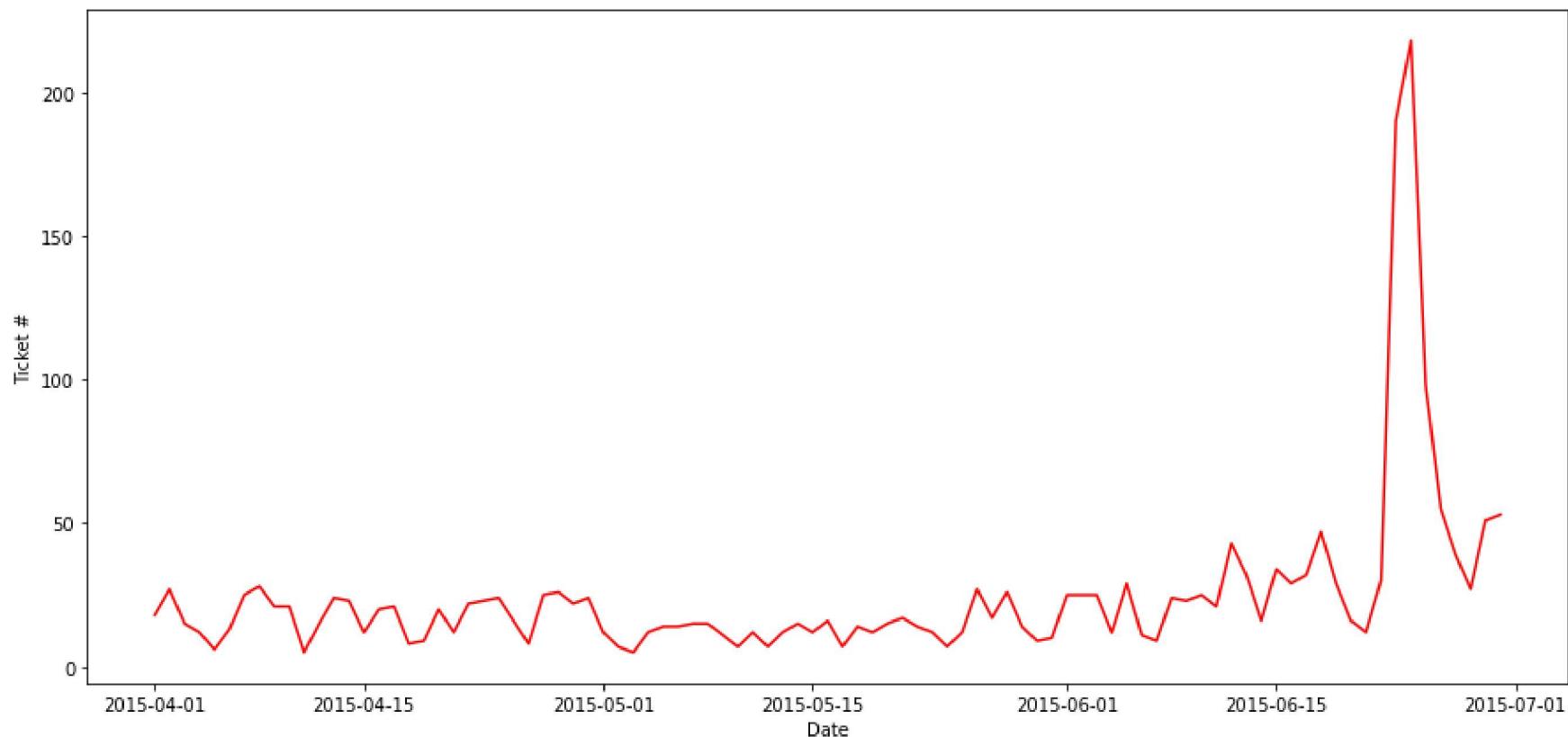
Out[12]:

	Date	Ticket #
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

In [13]:

```
plt.figure(figsize=(15,7))
sns.lineplot(data=datewise_df,x='Date', y='Ticket #', color='red')
```

Out[13]:



In [14]:

```
#Chart for month-wise complaints
monthwise_group=data.groupby(['Month'])['Ticket #'].count()
monthwise_df=pd.DataFrame(monthwise_group).reset_index()
monthwise_df.head()
```

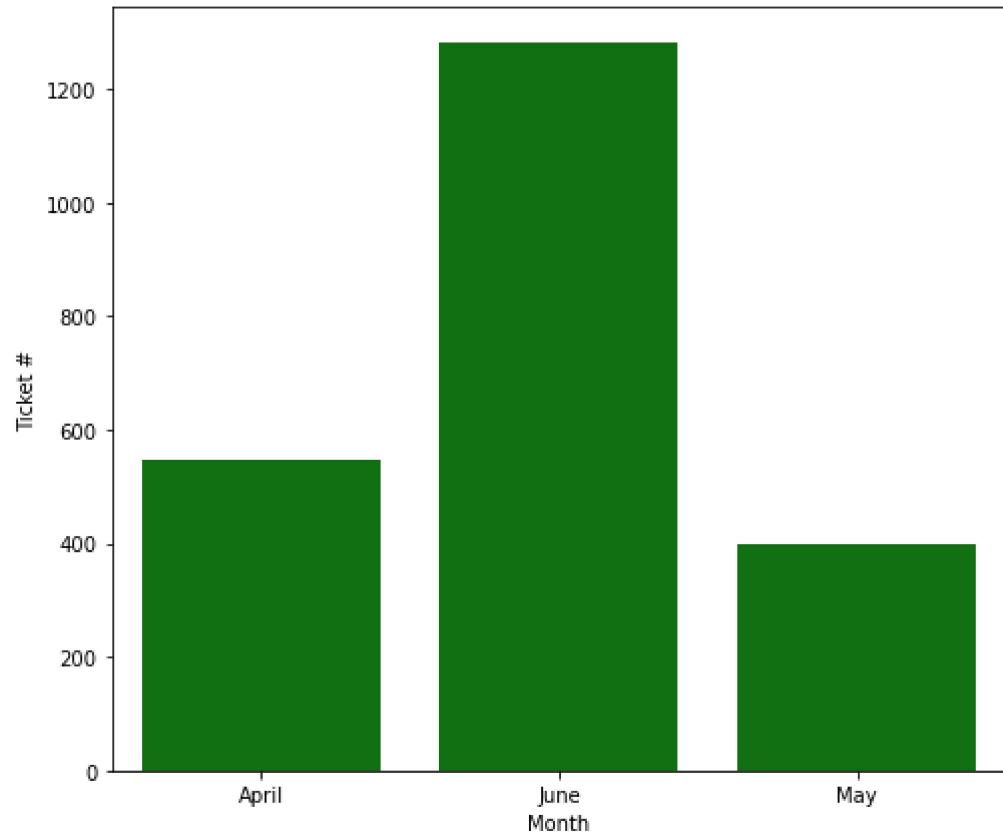
Out[14]:

	Month	Ticket #
0	April	545
1	June	1280
2	May	399

In [15]:

```
plt.figure(figsize=(8,7))
sns.barplot(data=monthwise_df,x='Month', y='Ticket #', color='green')
```

Out[15]:



In [16]:

```
#Frequency of complaint types  
data['Customer Complaint'].value_counts()[:14]
```

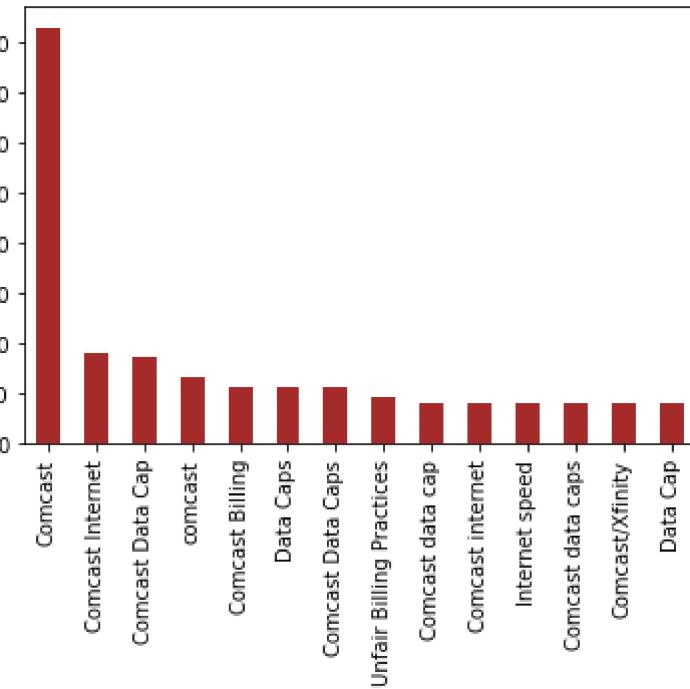
Out[16]:

Comcast	83
Comcast Internet	18
Comcast Data Cap	17
comcast	13
Comcast Billing	11
Data Caps	11
Comcast Data Caps	11
Unfair Billing Practices	9
Comcast data cap	8
Comcast internet	8
Internet speed	8
Comcast data caps	8
Comcast/Xfinity	8

Data Cap 8  
Name: Customer Complaint, dtype: int64

In [17]: `data['Customer Complaint'].value_counts()[:14].plot.bar(color="brown")`

Out[17]: <AxesSubplot:>



In [18]: `data['Newstatus']=[ 'Open' if status=='Open' or status=='Pending' else 'Closed' for status in data['Status']]`

In [19]: `data`

Out[19]:

	Ticket #	Customer Complaint	Date	Date_month_year	Time	Received Via	City	State	Zip code	Status	Filing on Behalf of Someone	Month	Newstatus
0	250635	Comcast Cable Internet Speeds	2015-04-22	22-Apr-15	3:53:50 PM	Customer Care Call	Abingdon	Maryland	21009	Closed	No	April	Closed

	Ticket #	Customer Complaint	Date	Date_month_year	Time	Received Via	City	State	Zip code	Status	Filing on Behalf of Someone	Month	Newstatus
1	223441	Payment disappear - service got disconnected	2015-04-08	04-Aug-15	10:22:56 AM	Internet	Acworth	Georgia	30102	Closed	No	April	Closed
2	242732	Speed and Service	2015-04-18	18-Apr-15	9:55:47 AM	Internet	Acworth	Georgia	30101	Closed	Yes	April	Closed
3	277946	Comcast Imposed a New Usage Cap of 300GB that ...	2015-05-07	05-Jul-15	11:59:35 AM	Internet	Acworth	Georgia	30101	Open	Yes	May	Open
4	307175	Comcast not working and no service to boot	2015-05-26	26-May-15	1:25:26 PM	Internet	Acworth	Georgia	30101	Solved	No	May	Closed
...	...	...	...	...	...	...	...	...	...	...	...	...	...
2219	213550	Service Availability	2015-04-02	04-Feb-15	9:13:18 AM	Customer Care Call	Youngstown	Florida	32466	Closed	No	April	Closed
2220	318775	Comcast Monthly Billing for Returned Modem	2015-06-02	06-Feb-15	1:24:39 PM	Customer Care Call	Ypsilanti	Michigan	48197	Solved	No	June	Closed
2221	331188	complaint about comcast	2015-06-09	06-Sep-15	5:28:41 PM	Internet	Ypsilanti	Michigan	48197	Solved	No	June	Closed
2222	360489	Extremely unsatisfied Comcast customer	2015-06-23	23-Jun-15	11:13:30 PM	Customer Care Call	Ypsilanti	Michigan	48197	Solved	No	June	Closed
2223	363614	Comcast, Ypsilanti MI Internet Speed	2015-06-24	24-Jun-15	10:28:33 PM	Customer Care Call	Ypsilanti	Michigan	48198	Open	Yes	June	Open

2224 rows × 13 columns

```
In [20]: statewise=data.groupby(['State','Newstatus']).size().unstack()
```

```
In [21]: statewise
```

```
Out[21]:
```

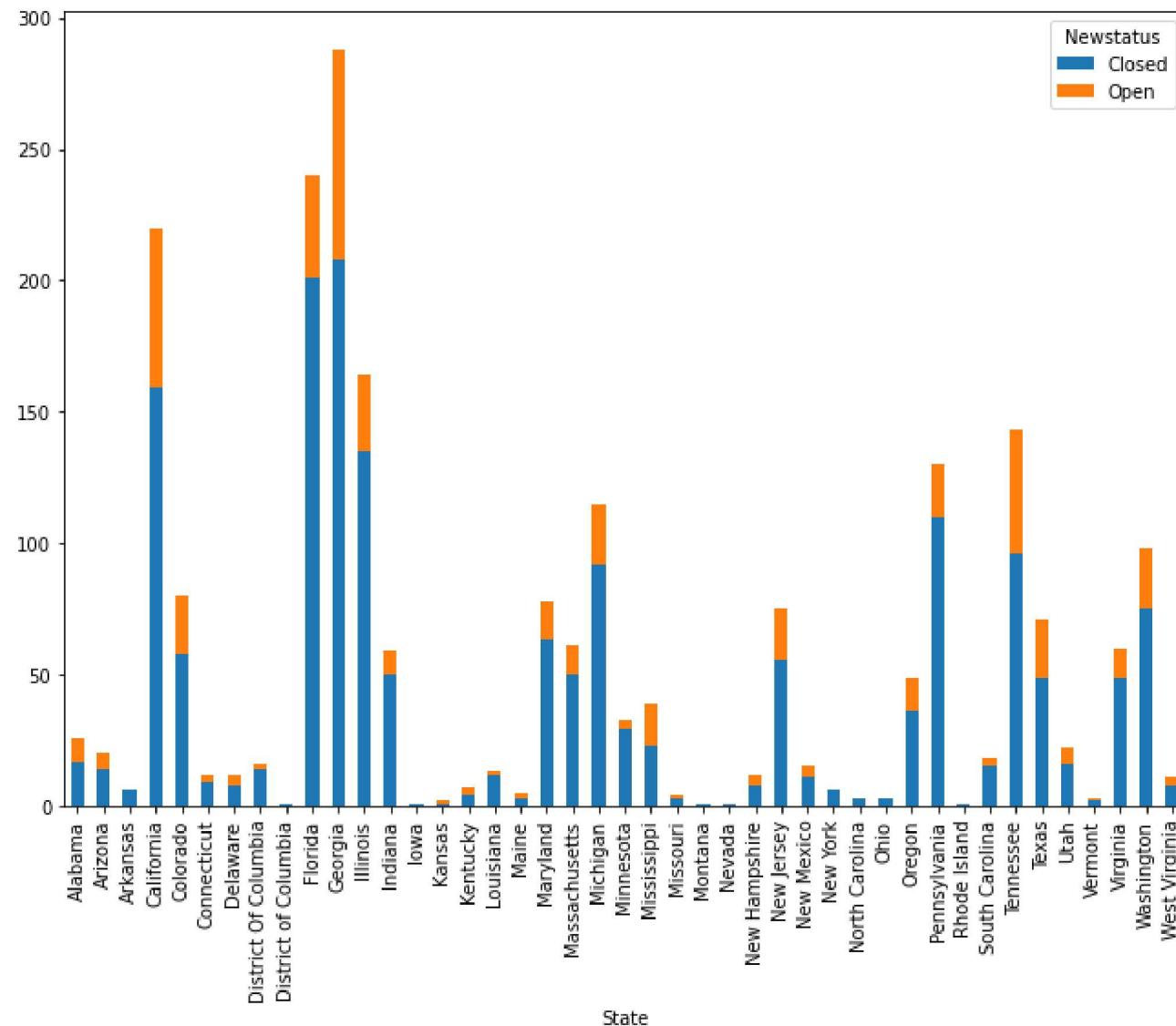
	Newstatus	Closed	Open
State			
<b>Alabama</b>	17.0	9.0	
<b>Arizona</b>	14.0	6.0	
<b>Arkansas</b>	6.0	NaN	
<b>California</b>	159.0	61.0	
<b>Colorado</b>	58.0	22.0	
<b>Connecticut</b>	9.0	3.0	
<b>Delaware</b>	8.0	4.0	
<b>District Of Columbia</b>	14.0	2.0	
<b>District of Columbia</b>	1.0	NaN	
<b>Florida</b>	201.0	39.0	
<b>Georgia</b>	208.0	80.0	
<b>Illinois</b>	135.0	29.0	
<b>Indiana</b>	50.0	9.0	
<b>Iowa</b>	1.0	NaN	
<b>Kansas</b>	1.0	1.0	
<b>Kentucky</b>	4.0	3.0	
<b>Louisiana</b>	12.0	1.0	
<b>Maine</b>	3.0	2.0	
<b>Maryland</b>	63.0	15.0	

Newstatus Closed Open

State		
<b>Massachusetts</b>	50.0	11.0
<b>Michigan</b>	92.0	23.0
<b>Minnesota</b>	29.0	4.0
<b>Mississippi</b>	23.0	16.0
<b>Missouri</b>	3.0	1.0
<b>Montana</b>	1.0	NaN
<b>Nevada</b>	1.0	NaN
<b>New Hampshire</b>	8.0	4.0
<b>New Jersey</b>	56.0	19.0
<b>New Mexico</b>	11.0	4.0
<b>New York</b>	6.0	NaN
<b>North Carolina</b>	3.0	NaN
<b>Ohio</b>	3.0	NaN
<b>Oregon</b>	36.0	13.0
<b>Pennsylvania</b>	110.0	20.0
<b>Rhode Island</b>	1.0	NaN
<b>South Carolina</b>	15.0	3.0
<b>Tennessee</b>	96.0	47.0
<b>Texas</b>	49.0	22.0
<b>Utah</b>	16.0	6.0
<b>Vermont</b>	2.0	1.0
<b>Virginia</b>	49.0	11.0
<b>Washington</b>	75.0	23.0
<b>West Virginia</b>	8.0	3.0

```
In [22]: statewise.plot.bar(stacked=True, figsize=(11,8))
```

```
Out[22]: <AxesSubplot:xlabel='State'>
```



```
In [23]: data.groupby('State').size().sort_values(ascending=False)
```

```
Out[23]: State
Georgia          288
Florida          240
California       220
Illinois         164
Tennessee        143
Pennsylvania     130
Michigan          115
Washington        98
Colorado          80
Maryland           78
New Jersey        75
Texas              71
Massachusetts     61
Virginia          60
Indiana            59
Oregon             49
Mississippi       39
Minnesota          33
Alabama            26
Utah                22
Arizona             20
South Carolina      18
District Of Columbia 16
New Mexico          15
Louisiana           13
Connecticut          12
New Hampshire        12
Delaware             12
West Virginia        11
Kentucky              7
Arkansas               6
New York               6
Maine                 5
Missouri               4
North Carolina        3
Vermont                 3
Ohio                  3
Kansas                 2
District of Columbia    1
Rhode Island            1
Iowa                  1
Nevada                 1
Montana                 1
dtype: int64
```

```
In [24]: # The above data shows that Georgia has the maximum number of complaints.
```

```
In [33]: statewise['percent']=statewise['Open']/(statewise['Open'].sum()+statewise['Closed'])* 100
```

```
In [34]: statewise.sort_values(by='percent',ascending=False)
```

```
Out[34]:
```

	Newstatus	Closed	Open	percent
	State			
<b>Georgia</b>	208.0	80.0	11.034483	
<b>California</b>	159.0	61.0	9.023669	
<b>Tennessee</b>	96.0	47.0	7.667210	
<b>Florida</b>	201.0	39.0	5.431755	
<b>Illinois</b>	135.0	29.0	4.447853	
<b>Texas</b>	49.0	22.0	3.886926	
<b>Washington</b>	75.0	23.0	3.885135	
<b>Colorado</b>	58.0	22.0	3.826087	
<b>Michigan</b>	92.0	23.0	3.776683	
<b>New Jersey</b>	56.0	19.0	3.315881	
<b>Pennsylvania</b>	110.0	20.0	3.189793	
<b>Mississippi</b>	23.0	16.0	2.962963	
<b>Maryland</b>	63.0	15.0	2.586207	
<b>Oregon</b>	36.0	13.0	2.350814	
<b>Virginia</b>	49.0	11.0	1.943463	
<b>Massachusetts</b>	50.0	11.0	1.940035	
<b>Alabama</b>	17.0	9.0	1.685393	
<b>Indiana</b>	50.0	9.0	1.587302	

Newstatus	Closed	Open	percent
State			
<b>Arizona</b>	14.0	6.0	1.129944
<b>Utah</b>	16.0	6.0	1.125704
<b>Delaware</b>	8.0	4.0	0.761905
<b>New Hampshire</b>	8.0	4.0	0.761905
<b>New Mexico</b>	11.0	4.0	0.757576
<b>Minnesota</b>	29.0	4.0	0.732601
<b>Kentucky</b>	4.0	3.0	0.575816
<b>West Virginia</b>	8.0	3.0	0.571429
<b>Connecticut</b>	9.0	3.0	0.570342
<b>South Carolina</b>	15.0	3.0	0.563910
<b>Maine</b>	3.0	2.0	0.384615
<b>District Of Columbia</b>	14.0	2.0	0.376648
<b>Kansas</b>	1.0	1.0	0.193050
<b>Vermont</b>	2.0	1.0	0.192678
<b>Missouri</b>	3.0	1.0	0.192308
<b>Louisiana</b>	12.0	1.0	0.189036
<b>Arkansas</b>	6.0	NaN	NaN
<b>District of Columbia</b>	1.0	NaN	NaN
<b>Iowa</b>	1.0	NaN	NaN
<b>Montana</b>	1.0	NaN	NaN
<b>Nevada</b>	1.0	NaN	NaN
<b>New York</b>	6.0	NaN	NaN
<b>North Carolina</b>	3.0	NaN	NaN
<b>Ohio</b>	3.0	NaN	NaN

```
Newstatus  Closed  Open  percent
```

```
State
```

State	Newstatus	Closed	Open	percent
Rhode Island	1.0	NaN	NaN	

```
In [ ]: #From the above table, Georgia has the highest percentage of unresolved complaints.
```

```
In [35]: #Considering both "closed" and "solved" cases
resolved_closed=data.groupby(['Received Via','Newstatus']).size().unstack()
resolved_closed['percent']=resolved_closed['Closed']/resolved_closed['Closed'].sum()*100
resolved_closed
```

```
Out[35]:      Newstatus  Closed  Open  percent
```

```
Received Via
```

Received Via	Newstatus	Closed	Open	percent
Customer Care Call	864	255	50.615114	
Internet	843	262	49.384886	

```
In [ ]: #Resolved cases via Customer care calls is 50.61% while through internet is 49.38%
```

```
In [28]: #Considering only "solved" cases
resolved_solved=data.groupby(['Received Via','Status']).size().unstack()
resolved_solved['percent']=resolved_solved['Solved']/resolved_solved['Solved'].sum()*100
resolved_solved['percent']
```

```
Out[28]: Received Via
Customer Care Call    49.023638
Internet              50.976362
Name: percent, dtype: float64
```