Analysis Task

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

- 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.

The analysis results to be provided with insights wherever applicable.

Importing Libraries and Data into python environment

In [1]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
date_fields = [['Date', 'Time']]
comcast_df =pd.read_csv(r"D:\Baldev\Data Science\Python\Project\Comcast_telecom_complai
```

In [2]:
```
comcast_df
```

Out[2]:

	Date_Time	Ticket #	Customer Complaint	Date	Date_month_year	Time	Received Via	City	Sta
0	2015-04-22 15:53:50	250635	Comcast Cable Internet Speeds	22-04-15	22-Apr-15	3:53:50 PM	Customer Care Call	Abingdon	Maryla
1	2015-08-04 10:22:56	223441	Payment disappear - service got disconnected	04-08-15	04-Aug-15	10:22:56 AM	Internet	Acworth	Geor

	Date_Time	Ticket #	Customer Complaint	Date	Date_month_year	Time	Received Via	City	Sta
2	2015-04-18 09:55:47	242732	Speed and Service	18-04-15	18-Apr-15	9:55:47 AM	Internet	Acworth	Geor
3	2015-07-05 11:59:35	277946	Comcast Imposed a New Usage Cap of 300GB that ...	05-07-15	05-Jul-15	11:59:35 AM	Internet	Acworth	Geor
4	2015-05-26 13:25:26	307175	Comcast not working and no service to boot	26-05-15	26-May-15	1:25:26 PM	Internet	Acworth	Geor
...	...	...	...	...	...	...	...	...	
2219	2015-02-04 09:13:18	213550	Service Availability	04-02-15	04-Feb-15	9:13:18 AM	Customer Care Call	Youngstown	Flori
2220	2015-02-06 13:24:39	318775	Comcast Monthly Billing for Returned Modem	06-02-15	06-Feb-15	1:24:39 PM	Customer Care Call	Ypsilanti	Michig
2221	2015-09-06 17:28:41	331188	complaint about comcast	06-09-15	06-Sep-15	5:28:41 PM	Internet	Ypsilanti	Michig
2222	2015-06-23 23:13:30	360489	Extremely unsatisfied Comcast customer	23-06-15	23-Jun-15	11:13:30 PM	Customer Care Call	Ypsilanti	Michig
2223	2015-06-24 22:28:33	363614	Comcast, Ypsilanti MI Internet Speed	24-06-15	24-Jun-15	10:28:33 PM	Customer Care Call	Ypsilanti	Michig

2224 rows × 12 columns

In [3]:
```
comcast_df.head()
```

Out[3]:

	Date_Time	Ticket #	Customer Complaint	Date	Date_month_year	Time	Received Via	City	State	
0	2015-04-22 15:53:50	250635	Comcast Cable Internet Speeds	22-04-15	22-Apr-15	3:53:50 PM	Customer Care Call	Abingdon	Maryland	2

	Date_Time	Ticket #	Customer Complaint	Date	Date_month_year	Time	Received Via	City	State	
1	2015-08-04 10:22:56	223441	Payment disappear - service got disconnected	04-08-15	04-Aug-15	10:22:56 AM	Internet	Acworth	Georgia	3
2	2015-04-18 09:55:47	242732	Speed and Service	18-04-15	18-Apr-15	9:55:47 AM	Internet	Acworth	Georgia	3
3	2015-07-05 11:59:35	277946	Comcast Imposed a New Usage Cap of 300GB that ...	05-07-15	05-Jul-15	11:59:35 AM	Internet	Acworth	Georgia	3
4	2015-05-26 13:25:26	307175	Comcast not working and no service to boot	26-05-15	26-May-15	1:25:26 PM	Internet	Acworth	Georgia	3

In [4]:
```
comcast_df.tail()
```

Out[4]:

	Date_Time	Ticket #	Customer Complaint	Date	Date_month_year	Time	Received Via	City	State
2219	2015-02-04 09:13:18	213550	Service Availability	04-02-15	04-Feb-15	9:13:18 AM	Customer Care Call	Youngstown	Florida
2220	2015-02-06 13:24:39	318775	Comcast Monthly Billing for Returned Modem	06-02-15	06-Feb-15	1:24:39 PM	Customer Care Call	Ypsilanti	Michigan
2221	2015-09-06 17:28:41	331188	complaint about comcast	06-09-15	06-Sep-15	5:28:41 PM	Internet	Ypsilanti	Michigan
2222	2015-06-23 23:13:30	360489	Extremely unsatisfied Comcast customer	23-06-15	23-Jun-15	11:13:30 PM	Customer Care Call	Ypsilanti	Michigan
2223	2015-06-24 22:28:33	363614	Comcast, Ypsilanti MI Internet Speed	24-06-15	24-Jun-15	10:28:33 PM	Customer Care Call	Ypsilanti	Michigan

In [5]:
```
comcast_df.describe()
```

Out[5]:

	Zip code
count	2224.000000
mean	47994.393435
std	28885.279427
min	1075.000000
25%	30056.500000
50%	37211.000000
75%	77058.750000
max	99223.000000

In [6]:
```
comcast_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2224 entries, 0 to 2223
Data columns (total 12 columns):
 #   Column                       Non-Null Count  Dtype         
---  ------                       --------------  -----         
 0   Date_Time                    2224 non-null   datetime64[ns]
 1   Ticket #                     2224 non-null   object        
 2   Customer Complaint           2224 non-null   object        
 3   Date                         2224 non-null   object        
 4   Date_month_year              2224 non-null   object        
 5   Time                         2224 non-null   object        
 6   Received Via                 2224 non-null   object        
 7   City                         2224 non-null   object        
 8   State                        2224 non-null   object        
 9   Zip code                     2224 non-null   int64         
 10  Status                       2224 non-null   object        
 11  Filing on Behalf of Someone  2224 non-null   object        
dtypes: datetime64[ns](1), int64(1), object(10)
memory usage: 208.6+ KB
```

In [7]:
```
comcast_df.columns
```

Out[7]:
```
Index(['Date_Time', 'Ticket #', 'Customer Complaint', 'Date',
       'Date_month_year', 'Time', 'Received Via', 'City', 'State', 'Zip code',
       'Status', 'Filing on Behalf of Someone'],
      dtype='object')
```

Tickets in Daily Granularity

In [8]:
```
comcast_df.insert(loc=3,column='Day',value=comcast_df['Date_Time'].dt.day)
```

In [9]:
```
comcast_df
```

Out[9]:

	Date_Time	Ticket #	Customer Complaint	Day	Date	Date_month_year	Time	Received Via	City

	Date_Time	Ticket #	Customer Complaint	Day	Date	Date_month_year	Time	Received Via	City
0	2015-04-22 15:53:50	250635	Comcast Cable Internet Speeds	22	22-04-15	22-Apr-15	3:53:50 PM	Customer Care Call	Abingdon
1	2015-08-04 10:22:56	223441	Payment disappear - service got disconnected	4	04-08-15	04-Aug-15	10:22:56 AM	Internet	Acworth
2	2015-04-18 09:55:47	242732	Speed and Service	18	18-04-15	18-Apr-15	9:55:47 AM	Internet	Acworth
3	2015-07-05 11:59:35	277946	Comcast Imposed a New Usage Cap of 300GB that ...	5	05-07-15	05-Jul-15	11:59:35 AM	Internet	Acworth
4	2015-05-26 13:25:26	307175	Comcast not working and no service to boot	26	26-05-15	26-May-15	1:25:26 PM	Internet	Acworth
...	...	...	...	...	...	...	...	...	...
2219	2015-02-04 09:13:18	213550	Service Availability	4	04-02-15	04-Feb-15	9:13:18 AM	Customer Care Call	Youngstown
2220	2015-02-06 13:24:39	318775	Comcast Monthly Billing for Returned Modem	6	06-02-15	06-Feb-15	1:24:39 PM	Customer Care Call	Ypsilanti
2221	2015-09-06 17:28:41	331188	complaint about comcast	6	06-09-15	06-Sep-15	5:28:41 PM	Internet	Ypsilanti
2222	2015-06-23 23:13:30	360489	Extremely unsatisfied Comcast customer	23	23-06-15	23-Jun-15	11:13:30 PM	Customer Care Call	Ypsilanti
2223	2015-06-24 22:28:33	363614	Comcast, Ypsilanti MI Internet Speed	24	24-06-15	24-Jun-15	10:28:33 PM	Customer Care Call	Ypsilanti

2224 rows × 13 columns

In [10]:
```
comcast_df.insert(loc=4,column='Month',value=comcast_df['Date_Time'].dt.month)
```

In [11]:
```
comcast_df
```

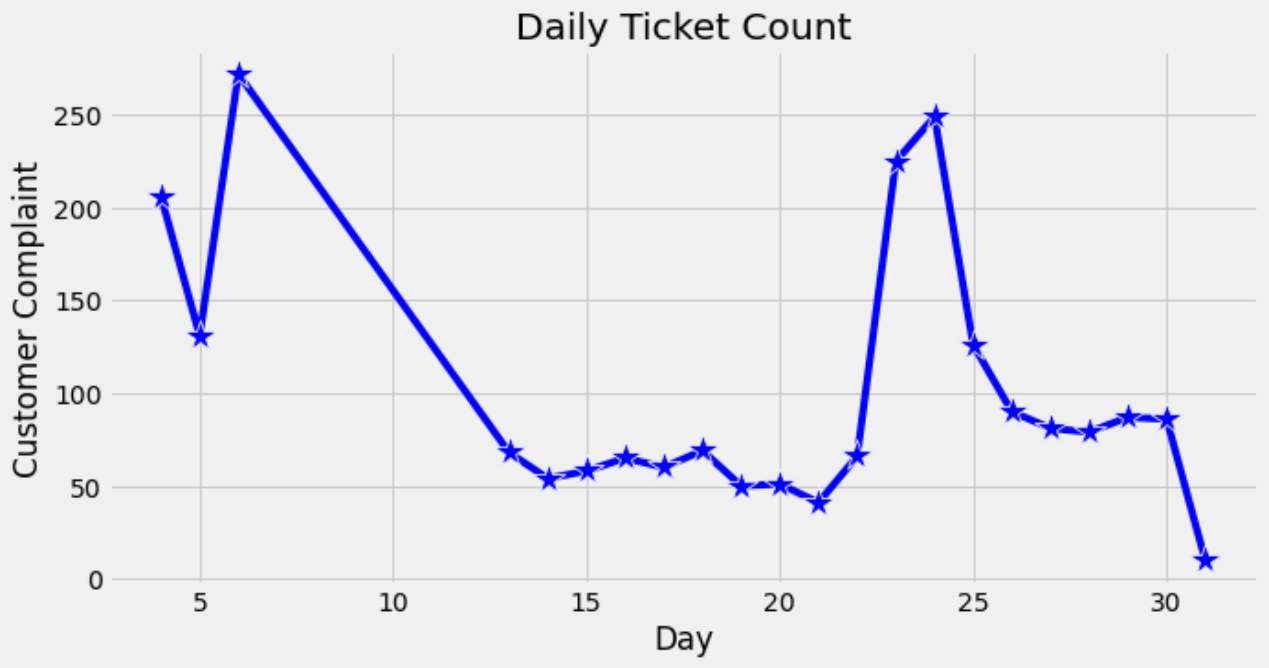
Out[11]:

	Date_Time	Ticket #	Customer Complaint	Day	Month	Date	Date_month_year	Time	Received Via	
0	2015-04-22 15:53:50	250635	Comcast Cable Internet Speeds	22	4	22-04-15	22-Apr-15	3:53:50 PM	Customer Care Call	Ab
1	2015-08-04 10:22:56	223441	Payment disappear - service got disconnected	4	8	04-08-15	04-Aug-15	10:22:56 AM	Internet	A
2	2015-04-18 09:55:47	242732	Speed and Service	18	4	18-04-15	18-Apr-15	9:55:47 AM	Internet	A
3	2015-07-05 11:59:35	277946	Comcast Imposed a New Usage Cap of 300GB that ...	5	7	05-07-15	05-Jul-15	11:59:35 AM	Internet	A
4	2015-05-26 13:25:26	307175	Comcast not working and no service to boot	26	5	26-05-15	26-May-15	1:25:26 PM	Internet	A
...	...	...	...	...	...	...	...	...	...	
2219	2015-02-04 09:13:18	213550	Service Availability	4	2	04-02-15	04-Feb-15	9:13:18 AM	Customer Care Call	Youn
2220	2015-02-06 13:24:39	318775	Comcast Monthly Billing for Returned Modem	6	2	06-02-15	06-Feb-15	1:24:39 PM	Customer Care Call	Y
2221	2015-09-06 17:28:41	331188	complaint about comcast	6	9	06-09-15	06-Sep-15	5:28:41 PM	Internet	Y
2222	2015-06-23 23:13:30	360489	Extremely unsatisfied Comcast customer	23	6	23-06-15	23-Jun-15	11:13:30 PM	Customer Care Call	Y
2223	2015-06-24 22:28:33	363614	Comcast, Ypsilanti MI Internet Speed	24	6	24-06-15	24-Jun-15	10:28:33 PM	Customer Care Call	Y

2224 rows × 14 columns

In [12]:
```
plt.style.use('fivethirtyeight')
```

```
plt.figure(figsize=(10,5))
byday = comcast_df.groupby('Day').count().reset_index()
lp = sns.lineplot(x= 'Day', y = 'Customer Complaint', data = byday,color="Blue", sort =
plt.title('Daily Ticket Count')
ax = lp.axes
plt.show()
```



Tickets in monthly Granularity

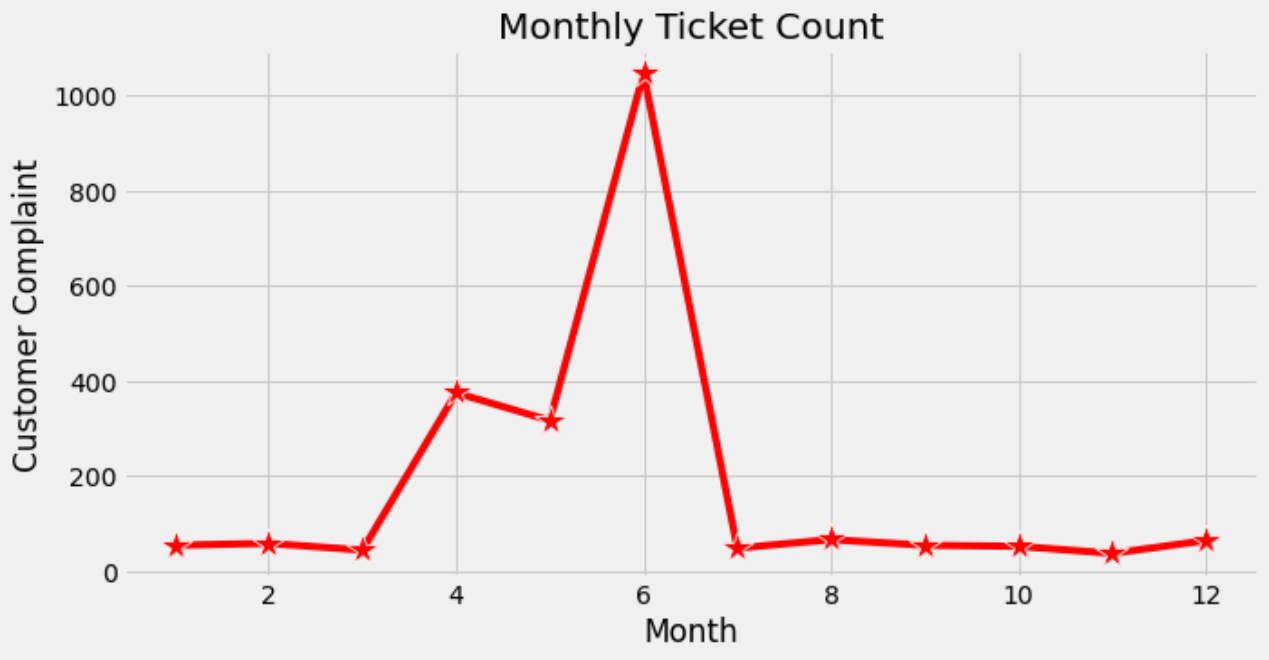
In [13]:

```
bymonth=comcast_df.groupby('Month').count().reset_index()
bymonth
```

Out[13]:

	Month	Date_Time	Ticket #	Customer Complaint	Day	Date	Date_month_year	Time	Received Via	City	State
0	1	55	55	55	55	55	55	55	55	55	55
1	2	59	59	59	59	59	59	59	59	59	59
2	3	45	45	45	45	45	45	45	45	45	45
3	4	375	375	375	375	375	375	375	375	375	375
4	5	317	317	317	317	317	317	317	317	317	317
5	6	1046	1046	1046	1046	1046	1046	1046	1046	1046	1046
6	7	49	49	49	49	49	49	49	49	49	49
7	8	67	67	67	67	67	67	67	67	67	67
8	9	55	55	55	55	55	55	55	55	55	55
9	10	53	53	53	53	53	53	53	53	53	53
10	11	38	38	38	38	38	38	38	38	38	38
11	12	65	65	65	65	65	65	65	65	65	65

In [14]:
```
plt.figure(figsize=(10,5))
lp = sns.lineplot(x= 'Month', y = 'Customer Complaint', data = bymonth,color="Red", sor
plt.title('Monthly Ticket Count')
ax = lp.axes
plt.show()
```



Frequency of complaint types

In [15]:
```
comcast_df['Customer Complaint']= comcast_df['Customer Complaint'].str.lower()
comcast_df
```

Out[15]:

	Date_Time	Ticket #	Customer Complaint	Day	Month	Date	Date_month_year	Time	Received Via	
0	2015-04-22 15:53:50	250635	comcast cable internet speeds	22	4	22-04-15	22-Apr-15	3:53:50 PM	Customer Care Call	Ab
1	2015-08-04 10:22:56	223441	payment disappear - service got disconnected	4	8	04-08-15	04-Aug-15	10:22:56 AM	Internet	A
2	2015-04-18 09:55:47	242732	speed and service	18	4	18-04-15	18-Apr-15	9:55:47 AM	Internet	A
3	2015-07-05 11:59:35	277946	comcast imposed a new usage cap of 300gb that ...	5	7	05-07-15	05-Jul-15	11:59:35 AM	Internet	A

	Date_Time	Ticket #	Customer Complaint	Day	Month	Date	Date_month_year	Time	Received Via	
4	2015-05-26 13:25:26	307175	comcast not working and no service to boot	26	5	26-05-15	26-May-15	1:25:26 PM	Internet	A
...	...	...	...	...	...	...	...	...	...	
2219	2015-02-04 09:13:18	213550	service availability	4	2	04-02-15	04-Feb-15	9:13:18 AM	Customer Care Call	Young
2220	2015-02-06 13:24:39	318775	comcast monthly billing for returned modem	6	2	06-02-15	06-Feb-15	1:24:39 PM	Customer Care Call	Y
2221	2015-09-06 17:28:41	331188	complaint about comcast	6	9	06-09-15	06-Sep-15	5:28:41 PM	Internet	Y
2222	2015-06-23 23:13:30	360489	extremely unsatisfied comcast customer	23	6	23-06-15	23-Jun-15	11:13:30 PM	Customer Care Call	Y
2223	2015-06-24 22:28:33	363614	comcast, ypsilanti mi internet speed	24	6	24-06-15	24-Jun-15	10:28:33 PM	Customer Care Call	Y

2224 rows × 14 columns

In [16]:
```
comcast_df.groupby(['Customer Complaint']).size().sort_values(ascending=False).to_frame
```

Out[16]:

	Customer Complaint	Count
0	comcast	102
1	comcast data cap	30
2	comcast internet	29
3	comcast data caps	21
4	comcast billing	18
...	...	...
1735	comcast not working and no service to boot	1
1736	comcast not refunding my credit	1
1737	comcast not refunding deposit for cancelled in...	1
1738	comcast not honoring agreement	1

	Customer Complaint	Count
1739	youtube being throttled?	1

1740 rows × 2 columns

Creating a new categorical variable as New_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 [64]:
```
comcast_df.Status.unique()
```

Out[64]:
```
array(['Closed', 'Open', 'Solved', 'Pending'], dtype=object)
```

In [18]:
```
comcast_df['Status']
```

Out[18]:
```
0       Closed
1       Closed
2       Closed
3         Open
4       Solved
         ...
2219    Closed
2220    Solved
2221    Solved
2222    Solved
2223      Open
Name: Status, Length: 2224, dtype: object
```

In [19]:
```
comcast_df['New_Status']= ["Open"if Status == "Pending" or Status == "Open" else 'Close
```

In [20]:
```
comcast_df
```

Out[20]:

	Date_Time	Ticket #	Customer Complaint	Day	Month	Date	Date_month_year	Time	Received Via	
0	2015-04-22 15:53:50	250635	comcast cable internet speeds	22	4	22-04-15	22-Apr-15	3:53:50 PM	Customer Care Call	Ab
1	2015-08-04 10:22:56	223441	payment disappear - service got disconnected	4	8	04-08-15	04-Aug-15	10:22:56 AM	Internet	A
2	2015-04-18 09:55:47	242732	speed and service	18	4	18-04-15	18-Apr-15	9:55:47 AM	Internet	A

	Date_Time	Ticket #	Customer Complaint	Day	Month	Date	Date_month_year	Time	Received Via	
3	2015-07-05 11:59:35	277946	comcast imposed a new usage cap of 300gb that ...	5	7	05-07-15	05-Jul-15	11:59:35 AM	Internet	A
4	2015-05-26 13:25:26	307175	comcast not working and no service to boot	26	5	26-05-15	26-May-15	1:25:26 PM	Internet	A
...	...	...	...	...	...	...	...	...	...	
2219	2015-02-04 09:13:18	213550	service availability	4	2	04-02-15	04-Feb-15	9:13:18 AM	Customer Care Call	Young
2220	2015-02-06 13:24:39	318775	comcast monthly billing for returned modem	6	2	06-02-15	06-Feb-15	1:24:39 PM	Customer Care Call	Y
2221	2015-09-06 17:28:41	331188	complaint about comcast	6	9	06-09-15	06-Sep-15	5:28:41 PM	Internet	Y
2222	2015-06-23 23:13:30	360489	extremely unsatisfied comcast customer	23	6	23-06-15	23-Jun-15	11:13:30 PM	Customer Care Call	Y
2223	2015-06-24 22:28:33	363614	comcast, ypsilanti mi internet speed	24	6	24-06-15	24-Jun-15	10:28:33 PM	Customer Care Call	Y

2224 rows × 15 columns

In [21]:
```
comcast_df.New_Status.unique()
```

Out[21]:
```
array(['Closed', 'Open'], dtype=object)
```

Provide state wise status of complaints in a stacked bar chart. Use the categorized variable

In [22]:
```
comcast_df['State']= comcast_df['State'].str.upper()
```

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

Out[23]:

	State	Count
0	GEORGIA	288
1	FLORIDA	240
2	CALIFORNIA	220
3	ILLINOIS	164
4	TENNESSEE	143
5	PENNSYLVANIA	130
6	MICHIGAN	115
7	WASHINGTON	98
8	COLORADO	80
9	MARYLAND	78
10	NEW JERSEY	75
11	TEXAS	71
12	MASSACHUSETTS	61
13	VIRGINIA	60
14	INDIANA	59
15	OREGON	49
16	MISSISSIPPI	39
17	MINNESOTA	33
18	ALABAMA	26
19	UTAH	22
20	ARIZONA	20
21	SOUTH CAROLINA	18
22	DISTRICT OF COLUMBIA	17
23	NEW MEXICO	15
24	LOUISIANA	13
25	CONNECTICUT	12
26	NEW HAMPSHIRE	12
27	DELAWARE	12
28	WEST VIRGINIA	11
29	KENTUCKY	7
30	ARKANSAS	6
31	NEW YORK	6
32	MAINE	5

	State	Count
33	MISSOURI	4
34	NORTH CAROLINA	3
35	VERMONT	3
36	OHIO	3
37	KANSAS	2
38	RHODE ISLAND	1
39	IOWA	1
40	NEVADA	1
41	MONTANA	1

In [24]:
```
state_complaints = pd.crosstab(index = comcast_df['State'],columns=comcast_df["New_Stat
```

In [25]:
```
state_complaints.fillna(0,inplace=True)
```

In [26]:
```
state_complaints
```

Out[26]:

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	15.0	2.0
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

New_Status	Closed	Open
State		
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

In [27]:
```
state_complaints['Total']= state_complaints['Closed']+state_complaints['Open']
```

State wise status of complaints in a stacked bar chart, Where the no. of open,closed and total tickets are shown

In [28]:
```
state_complaints
```

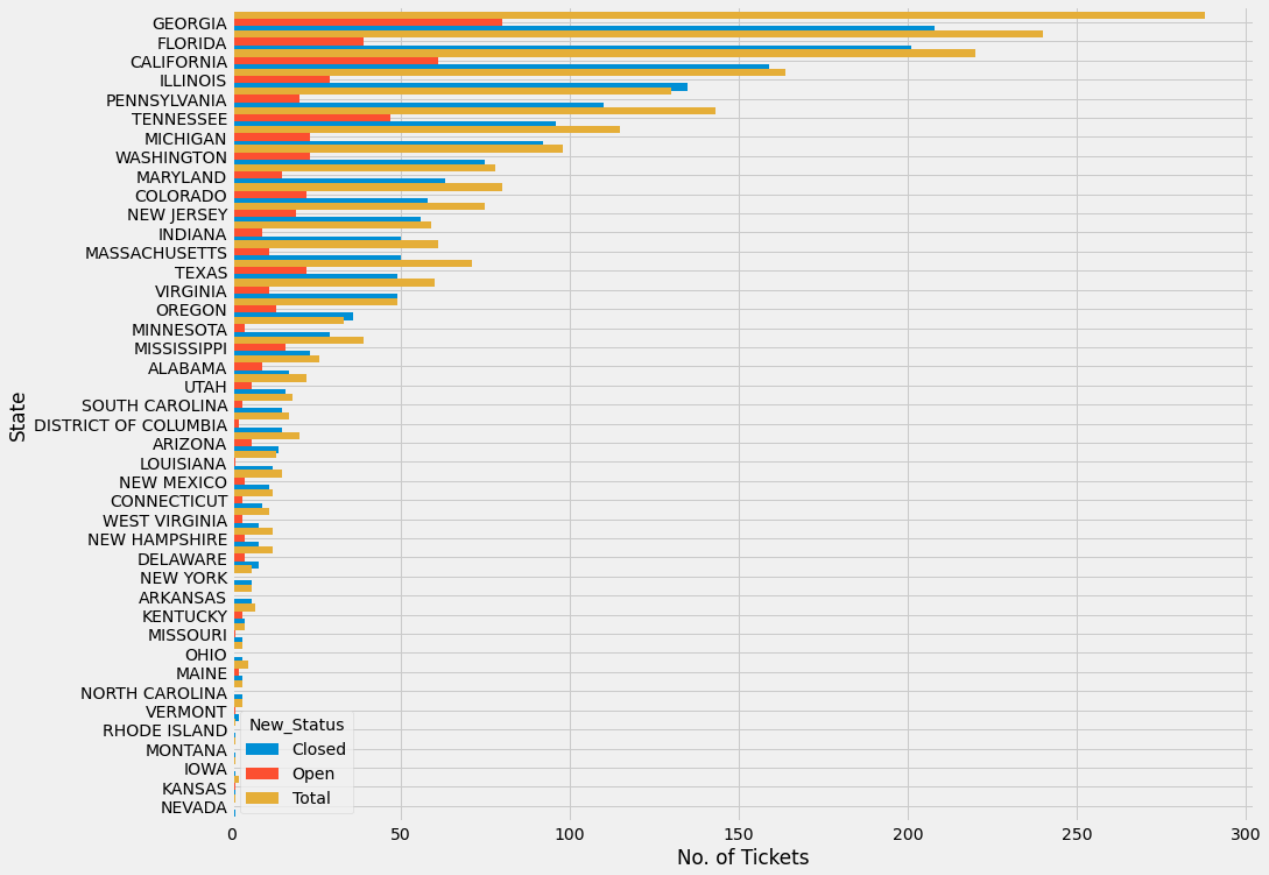
Out[28]:

New_Status	Closed	Open	Total
State			
ALABAMA	17.0	9.0	26.0
ARIZONA	14.0	6.0	20.0
ARKANSAS	6.0	0.0	6.0
CALIFORNIA	159.0	61.0	220.0
COLORADO	58.0	22.0	80.0
CONNECTICUT	9.0	3.0	12.0
DELAWARE	8.0	4.0	12.0
DISTRICT OF COLUMBIA	15.0	2.0	17.0
FLORIDA	201.0	39.0	240.0
GEORGIA	208.0	80.0	288.0
ILLINOIS	135.0	29.0	164.0
INDIANA	50.0	9.0	59.0
IOWA	1.0	0.0	1.0
KANSAS	1.0	1.0	2.0
KENTUCKY	4.0	3.0	7.0
LOUISIANA	12.0	1.0	13.0
MAINE	3.0	2.0	5.0
MARYLAND	63.0	15.0	78.0
MASSACHUSETTS	50.0	11.0	61.0
MICHIGAN	92.0	23.0	115.0
MINNESOTA	29.0	4.0	33.0
MISSISSIPPI	23.0	16.0	39.0
MISSOURI	3.0	1.0	4.0
MONTANA	1.0	0.0	1.0
NEVADA	1.0	0.0	1.0
NEW HAMPSHIRE	8.0	4.0	12.0
NEW JERSEY	56.0	19.0	75.0
NEW MEXICO	11.0	4.0	15.0
NEW YORK	6.0	0.0	6.0
NORTH CAROLINA	3.0	0.0	3.0
OHIO	3.0	0.0	3.0
OREGON	36.0	13.0	49.0

New_Status	Closed	Open	Total
State			
PENNSYLVANIA	110.0	20.0	130.0
RHODE ISLAND	1.0	0.0	1.0
SOUTH CAROLINA	15.0	3.0	18.0
TENNESSEE	96.0	47.0	143.0
TEXAS	49.0	22.0	71.0
UTAH	16.0	6.0	22.0
VERMONT	2.0	1.0	3.0
VIRGINIA	49.0	11.0	60.0
WASHINGTON	75.0	23.0	98.0
WEST VIRGINIA	8.0	3.0	11.0

In [29]:
```
n= state_complaints.sort_values('Closed',axis=0,ascending=True).plot(kind='barh',figsiz
plt.xlabel("No. of Tickets")
```

Out[29]: Text(0.5, 0, 'No. of Tickets')



In [32]:
```
state_complaints['Unresoleved %age']= state_complaints['Open']/state_complaints['Total'
```

In [33]:
```
state_complaints
```

Out[33]:

New_Status State	Closed	Open	Total	Unresoleved %age
ALABAMA	17.0	9.0	26.0	34.615385
ARIZONA	14.0	6.0	20.0	30.000000
ARKANSAS	6.0	0.0	6.0	0.000000
CALIFORNIA	159.0	61.0	220.0	27.727273
COLORADO	58.0	22.0	80.0	27.500000
CONNECTICUT	9.0	3.0	12.0	25.000000
DELAWARE	8.0	4.0	12.0	33.333333
DISTRICT OF COLUMBIA	15.0	2.0	17.0	11.764706
FLORIDA	201.0	39.0	240.0	16.250000
GEORGIA	208.0	80.0	288.0	27.777778
ILLINOIS	135.0	29.0	164.0	17.682927
INDIANA	50.0	9.0	59.0	15.254237
IOWA	1.0	0.0	1.0	0.000000
KANSAS	1.0	1.0	2.0	50.000000
KENTUCKY	4.0	3.0	7.0	42.857143
LOUISIANA	12.0	1.0	13.0	7.692308
MAINE	3.0	2.0	5.0	40.000000
MARYLAND	63.0	15.0	78.0	19.230769
MASSACHUSETTS	50.0	11.0	61.0	18.032787
MICHIGAN	92.0	23.0	115.0	20.000000
MINNESOTA	29.0	4.0	33.0	12.121212
MISSISSIPPI	23.0	16.0	39.0	41.025641
MISSOURI	3.0	1.0	4.0	25.000000
MONTANA	1.0	0.0	1.0	0.000000
NEVADA	1.0	0.0	1.0	0.000000
NEW HAMPSHIRE	8.0	4.0	12.0	33.333333
NEW JERSEY	56.0	19.0	75.0	25.333333
NEW MEXICO	11.0	4.0	15.0	26.666667
NEW YORK	6.0	0.0	6.0	0.000000
NORTH CAROLINA	3.0	0.0	3.0	0.000000
OHIO	3.0	0.0	3.0	0.000000

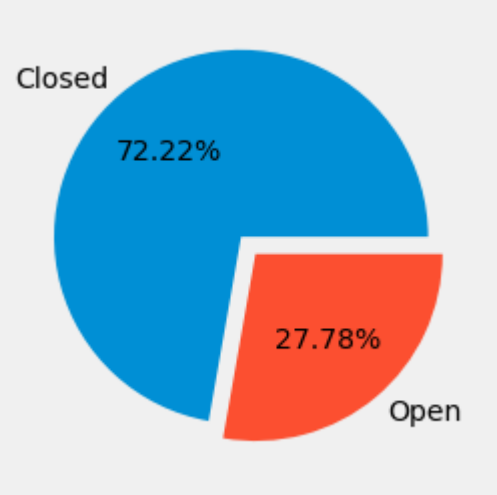
New_Status / State	Closed	Open	Total	Unresoleved %age
OREGON	36.0	13.0	49.0	26.530612
PENNSYLVANIA	110.0	20.0	130.0	15.384615
RHODE ISLAND	1.0	0.0	1.0	0.000000
SOUTH CAROLINA	15.0	3.0	18.0	16.666667
TENNESSEE	96.0	47.0	143.0	32.867133
TEXAS	49.0	22.0	71.0	30.985915
UTAH	16.0	6.0	22.0	27.272727
VERMONT	2.0	1.0	3.0	33.333333
VIRGINIA	49.0	11.0	60.0	18.333333
WASHINGTON	75.0	23.0	98.0	23.469388
WEST VIRGINIA	8.0	3.0	11.0	27.272727

Finding the state which has the highest percentage of unresolved complaints

In [48]:
```
Ticket_Status=comcast_df.groupby(["State","New_Status"]).size().unstack().fillna(0).max
Ticket_Status
```

Out[48]:
```
New_Status
Closed    208.0
Open       80.0
dtype: float64
```

In [49]:
```
# Total CLosed and open Tickets
labels = "Closed","Open"
explode = (0.07,0.05)
plt.pie(a,labels=labels,explode=explode,autopct='%1.2f%%')
plt.show()
```



In [52]:
```
Closed_Ticket = comcast_df.groupby(["State","New_Status"]).size().unstack().fillna(0)
```

```
Closed_Ticket.sort_values('Closed',axis = 0,ascending=False)[:1]
```

Out[52]:

State	New_Status Closed	Open
GEORGIA	208.0	80.0

In [55]:

```
Closed_Ticket['Resolved_cmp_prct'] = Closed_Ticket['Closed']/Closed_Ticket['Closed'].su
Closed_Ticket['Unresolved_cmp_prct'] = Closed_Ticket['Open']/Closed_Ticket['Open'].sum(
```

In [57]:

```
# Georgia state has highest Unresolved complaints when compared to other states
Closed_Ticket.sort_values('Unresolved_cmp_prct',axis = 0,ascending=False)[:1]
```

Out[57]:

State	New_Status Closed	Open	Resolved_cmp_prct	Unresolved_cmp_prct
GEORGIA	208.0	80.0	12.18512	15.473888

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

In [62]:

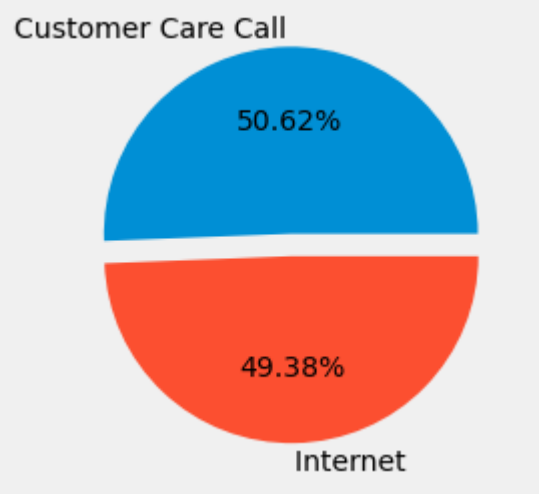
```
Complaints_Resolved = comcast_df.groupby(['Received Via','New_Status']).size().unstack(
Complaints_Resolved['resolved'] = Complaints_Resolved['Closed']/Complaints_Resolved['Cl
P=Complaints_Resolved['resolved']
P
```

Out[62]:

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

In [63]:

```
labels = "Customer Care Call","Internet "
explode = (0.07,0.05)
plt.pie(P,labels=labels,explode=explode,autopct='%1.2f%%')
plt.show()
```



Project Ends

In [ ]: