In [111…

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

In [112…

```
df=pd.read_csv('1568699544_comcast_telecom_complaints_data (1).zip')
```

In [113…

```
df.head()
```

Out[113]:

	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 [114…

```
df.tail()
```

Out[114]:

	Ticket #	Customer Complaint	Date	Date_month_year	Time	Received Via	City	State	Zip code	Status	Filing on Behalf of Someone
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

In [115…

```
df.shape
```

Out[115]:

```
(2224, 11)
```

Task-1 Provide the trend chart for the number of complaints as monthly and daily granularity levels.

In [116…

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

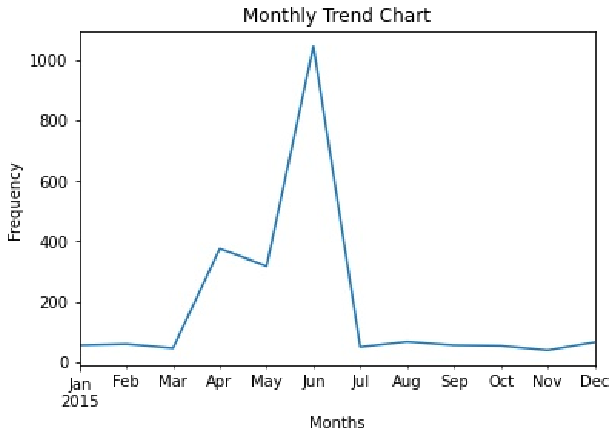
plotting monthly chart

In [123…

```
months=df.groupby(pd.Grouper(freq='M')).size().plot()
plt.xlabel('Months')
plt.ylabel('Frequency')
plt.title('Monthly Trend Chart')
```

Out[123]:

```
Text(0.5, 1.0, 'Monthly Trend Chart')
```



df['Date'].value_counts()[:8]

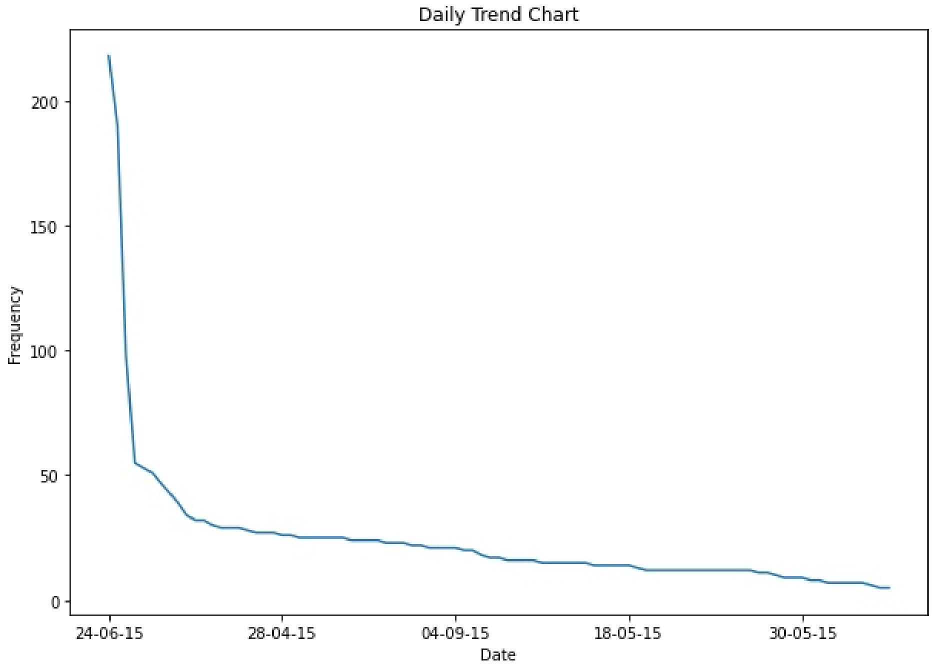
plotting daily chart

In [73]:

```
plt.figure(figsize=(10,7))
df['Date'].value_counts().plot()
plt.xlabel('Date')
plt.ylabel('Frequency')
plt.title('Daily Trend Chart')
```

Out[73]:

```
Text(0.5, 1.0, 'Daily Trend Chart')
```



In [ ]:

```

```

Task-2 Provide a table with the frequency of complaint types

In [76]:

```
df['Customer Complaint'].value_counts()[:9]
```

Out[76]:

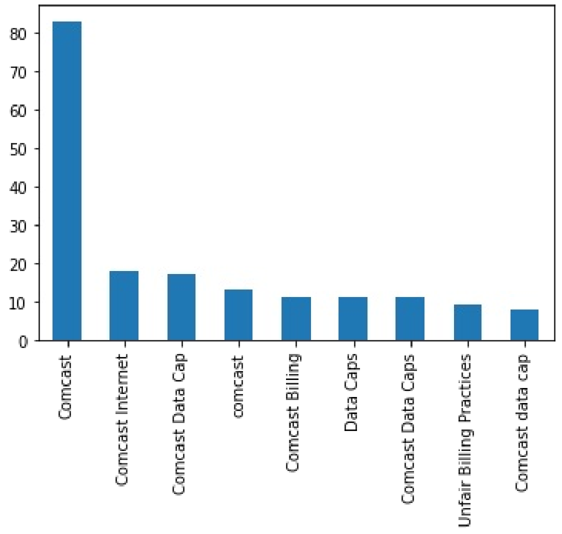
```
Comcast                      83
Comcast Internet             18
Comcast Data Cap             17
comcast                      13
Comcast Billing              11
Data Caps                    11
Comcast Data Caps            11
Unfair Billing Practices      9
Comcast data cap              8
Name: Customer Complaint, dtype: int64
```

In [77]:

```
df['Customer Complaint'].value_counts()[:9].plot.bar()
```

Out[77]:

```
<AxesSubplot:>
```



Task3-Which complaint types are maximum i.e., around internet, network issues, or across any other domains.

In [78]:

```
internet_issue1=df[df['Customer Complaint'].str.contains('networks')].count()
```

In [79]:

```
internet_issue2=df[df['Customer Complaint'].str.contains('speed')].count()
```

In [80]:

```
internet_issue3=df[df['Customer Complaint'].str.contains('data')].count()
```

In [81]:

```
internet_issue4=df[df['Customer Complaint'].str.contains('internet')].count()
```

In [82]:

```
billing_issue1=df[df['Customer Complaint'].str.contains('bill')].count()
```

In [83]:

```
billing_issue2=df[df['Customer Complaint'].str.contains('billing')].count()
```

In [84]:

```
billing_issue3=df[df['Customer Complaint'].str.contains('charges')].count()
```

In [85]:

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

In [86]:

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

In [87]:

```
total_internet_issues=internet_issue1+internet_issue2+internet_issue3+internet_issue4
print(total_internet_issues)
```

```
Ticket #                       373
Customer Complaint             373
Date                           373
Time                           373
Received Via                   373
City                           373
State                          373
Zip code                       373
Status                         373
Filing on Behalf of Someone    373
dtype: int64
```

In [88]:

```
total_billing_issues=billing_issue1+billing_issue2+billing_issue3
print(total_billing_issues)
```

```
Ticket #                       353
Customer Complaint             353
Date                           353
Time                           353
Received Via                   353
City                           353
State                          353
Zip code                       353
Status                         353
Filing on Behalf of Someone    353
dtype: int64
```

In [89]:

```
total_service_issues=service_issue1+service_issue2
print(total_service_issues)
```

```
Ticket #                       360
Customer Complaint             360
Date                           360
Time                           360
Received Via                   360
City                           360
State                          360
Zip code                       360
Status                         360
Filing on Behalf of Someone    360
dtype: int64
```

In [90]:

```
other_issues=2224-(total_internet_issues+total_billing_issues+total_service_issues)
print(other_issues)
```

```
Ticket #                       1138
Customer Complaint             1138
Date                           1138
Time                           1138
Received Via                   1138
City                           1138
State                          1138
Zip code                       1138
Status                         1138
Filing on Behalf of Someone    1138
dtype: int64
```

In [91]:

```
# We can conclude that other_isssues are maximum.
```

Task4-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.

In [92]:

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

Out[92]:

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

In [93]:

```
df['newstatus'] = ["Open" if Status=="Open" or Status=="Pending" else "Closed" for Status in df["Status"]]
df=df.drop(['Status'], axis=1)
df
```

Out[93]:

	Ticket #	Customer Complaint	Date	Time	Received Via	City	State	Zip code	Filing on Behalf of Someone	newstatus
Date_month_year										
2015-04-22	250635	Comcast Cable Internet Speeds	22-04-15	3:53:50 PM	Customer Care Call	Abingdon	Maryland	21009	No	Closed
2015-08-04	223441	Payment disappear - service got disconnected	04-08-15	10:22:56 AM	Internet	Acworth	Georgia	30102	No	Closed
2015-04-18	242732	Speed and Service	18-04-15	9:55:47 AM	Internet	Acworth	Georgia	30101	Yes	Closed
2015-07-05	277946	Comcast Imposed a New Usage Cap of 300GB that ...	05-07-15	11:59:35 AM	Internet	Acworth	Georgia	30101	Yes	Open
2015-05-26	307175	Comcast not working and no service to boot	26-05-15	1:25:26 PM	Internet	Acworth	Georgia	30101	No	Closed
...	...	...	...	...	...	...	...	...	...	...
2015-02-04	213550	Service Availability	04-02-15	9:13:18 AM	Customer Care Call	Youngstown	Florida	32466	No	Closed
2015-02-06	318775	Comcast Monthly Billing for Returned Modem	06-02-15	1:24:39 PM	Customer Care Call	Ypsilanti	Michigan	48197	No	Closed
2015-09-06	331188	complaint about comcast	06-09-15	5:28:41 PM	Internet	Ypsilanti	Michigan	48197	No	Closed
2015-06-23	360489	Extremely unsatisfied Comcast customer	23-06-15	11:13:30 PM	Customer Care Call	Ypsilanti	Michigan	48197	No	Closed
2015-06-24	363614	Comcast, Ypsilanti MI Internet Speed	24-06-15	10:28:33 PM	Customer Care Call	Ypsilanti	Michigan	48198	Yes	Open

2224 rows × 10 columns

Task5- Provide state wise status of complaints in a stacked bar chart

In [94]:

```
state_complaint=df.groupby(['State','newstatus']).size().unstack()
```

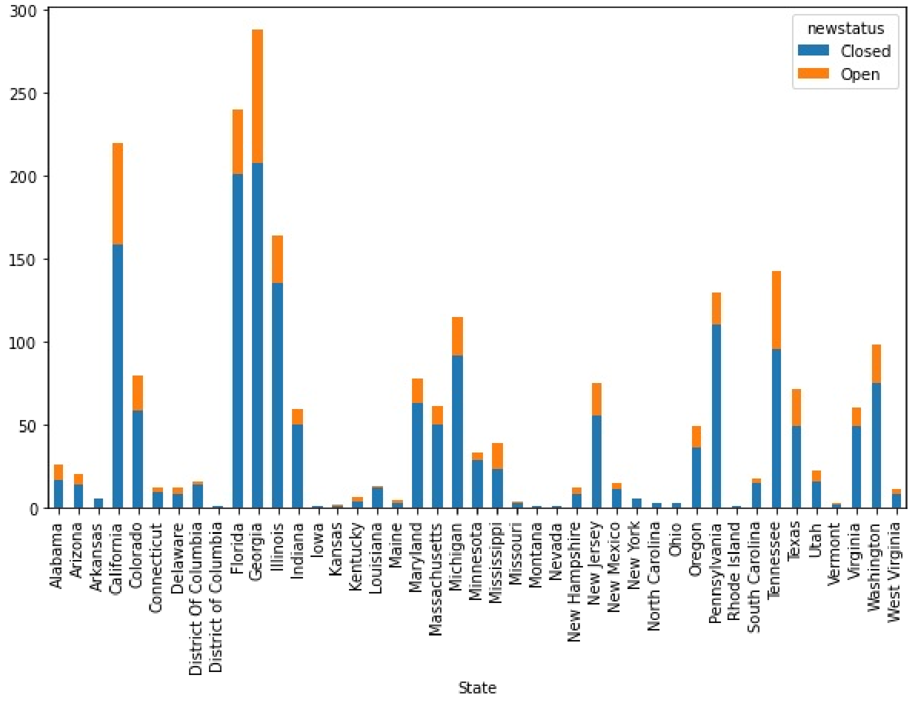
In [95]:

```
print(state_complaint)
```

```
newstatus             Closed  Open
State                             
Alabama                 17.0   9.0
Arizona                 14.0   6.0
Arkansas                 6.0   NaN
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   NaN
Florida                201.0  39.0
Georgia                208.0  80.0
Illinois               135.0  29.0
Indiana                 50.0   9.0
Iowa                     1.0   NaN
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   NaN
Nevada                   1.0   NaN
New Hampshire            8.0   4.0
New Jersey              56.0  19.0
New Mexico              11.0   4.0
New York                 6.0   NaN
North Carolina           3.0   NaN
Ohio                     3.0   NaN
Oregon                  36.0  13.0
Pennsylvania           110.0  20.0
Rhode Island             1.0   NaN
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 [96]:
```
state_complaint.plot.bar(figsize=(10,6),stacked=True)
```

Out[96]:
```
<AxesSubplot:xlabel='State'>
```



Task6- Which state has the maximum complaints

In [97]:
```
df.groupby(['State']).size().sort_values(ascending=False)
```

Out[97]:

```
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 [98]:

```
# We can conclude georgia state has maximum complaints
```

Task7- Which state has the highest percentage of unresolved complaints

In [99]:

```
print(df['newstatus'].value_counts())
```

```
Closed    1707
Open       517
Name: newstatus, dtype: int64
```

In [100…

```
unresolved_data=df.groupby(['State','newstatus']).size().unstack().sort_values(by='Open',ascending=False)
unresolved_data['unresolved_data_prct']=unresolved_data['Open']/unresolved_data['Open'].sum()*100
print(unresolved_data)
```

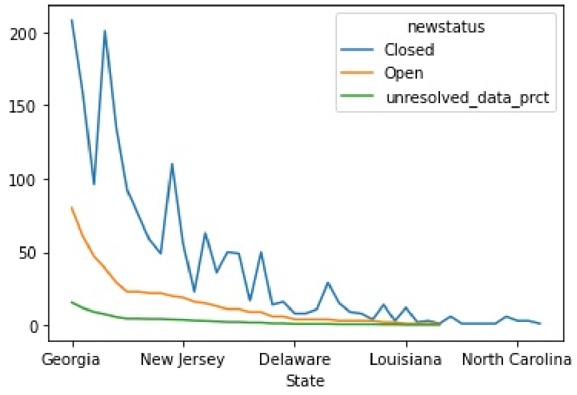
```
newstatus             Closed  Open  unresolved_data_prct
State                                                   
Georgia                208.0  80.0             15.473888
California             159.0  61.0             11.798839
Tennessee               96.0  47.0              9.090909
Florida                201.0  39.0              7.543520
Illinois               135.0  29.0              5.609284
Michigan                92.0  23.0              4.448743
Washington              75.0  23.0              4.448743
Colorado                58.0  22.0              4.255319
Texas                   49.0  22.0              4.255319
Pennsylvania           110.0  20.0              3.868472
New Jersey              56.0  19.0              3.675048
Mississippi             23.0  16.0              3.094778
Maryland                63.0  15.0              2.901354
Oregon                  36.0  13.0              2.514507
Massachusetts           50.0  11.0              2.127660
Virginia                49.0  11.0              2.127660
Alabama                 17.0   9.0              1.740812
Indiana                 50.0   9.0              1.740812
Arizona                 14.0   6.0              1.160542
Utah                    16.0   6.0              1.160542
Delaware                 8.0   4.0              0.773694
New Hampshire            8.0   4.0              0.773694
New Mexico              11.0   4.0              0.773694
Minnesota               29.0   4.0              0.773694
South Carolina          15.0   3.0              0.580271
Connecticut              9.0   3.0              0.580271
West Virginia            8.0   3.0              0.580271
Kentucky                 4.0   3.0              0.580271
District Of Columbia    14.0   2.0              0.386847
Maine                    3.0   2.0              0.386847
Louisiana               12.0   1.0              0.193424
Vermont                  2.0   1.0              0.193424
Missouri                 3.0   1.0              0.193424
Kansas                   1.0   1.0              0.193424
Arkansas                 6.0   NaN                   NaN
District of Columbia     1.0   NaN                   NaN
Iowa                     1.0   NaN                   NaN
Montana                  1.0   NaN                   NaN
Nevada                   1.0   NaN                   NaN
New York                 6.0   NaN                   NaN
North Carolina           3.0   NaN                   NaN
Ohio                     3.0   NaN                   NaN
Rhode Island             1.0   NaN                   NaN
```

In [101…

```
unresolved_data.plot()
```

Out[101]:

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



In [102…

```
#We can conclude that Georgia state has highest unresolved complaint
```

Task8- Provide the percentage of complaints resolved till date, which were received through the Internet and customer care calls.

In [103…

```
resolved_data=df.groupby(['Received Via','newstatus']).size().unstack().sort_values(by='Closed',ascending=False
resolved_data['resolved_data_prct']=resolved_data['Closed']/unresolved_data['Closed'].sum()*100
print(resolved_data)
```

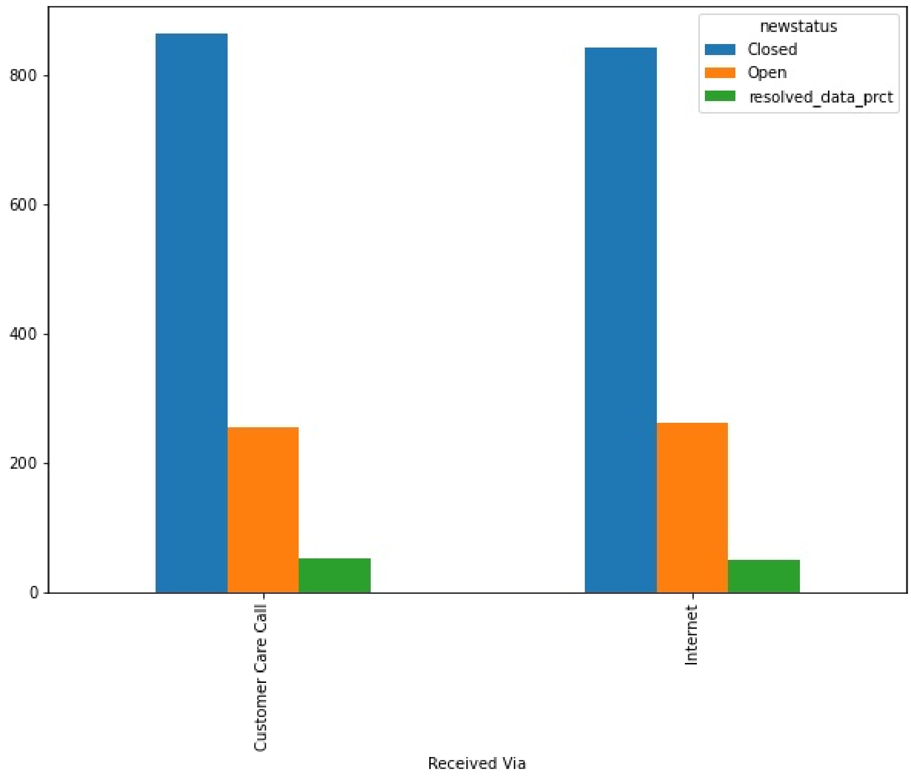
```
newstatus          Closed  Open  resolved_data_prct
Received Via                                       
Customer Care Call    864   255           50.615114
Internet              843   262           49.384886
```

In [104…

```
resolved_data.plot(kind='bar',figsize=(10,7))
```

Out[104]:

```
<AxesSubplot:xlabel='Received Via'>
```



We can conclude from this graph that there are 50.61%complaints resolved from customercarecall and 49.39% complaints resolved from internet

In [ ]:

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js