In [270]:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as cbrn
%matplotlib inline
```

In [271]:

```
#Import data into Python environment.
cmplnt_df = pd.read_csv(r"C:\Users\jms1cob\Downloads\1568699544_comcast_telecom_complaints_data\Comcast_telecom_complaints_data.csv")
```

In [272]:

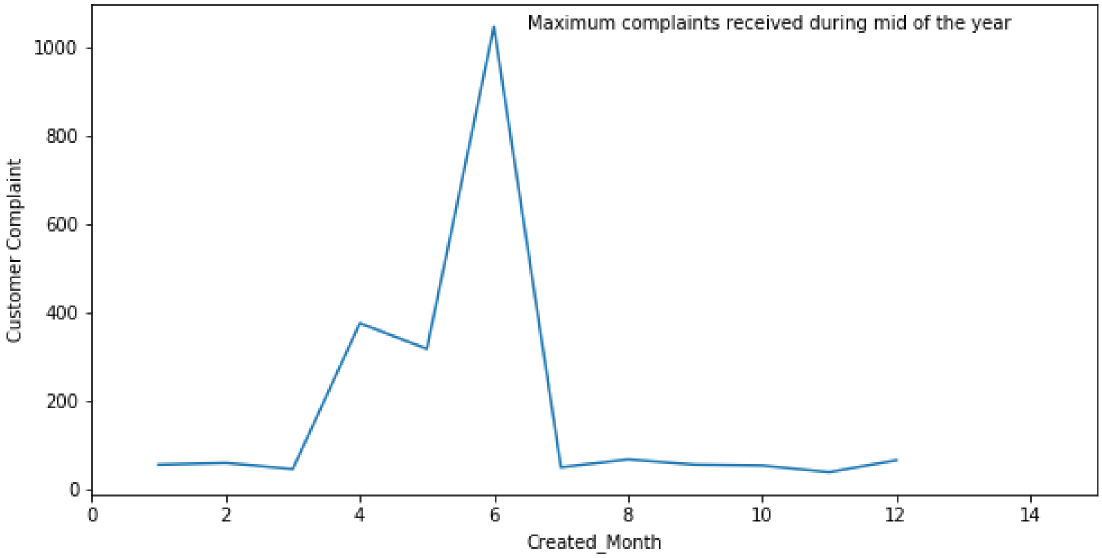
```
#Provide the trend chart for the number of complaints at monthly and daily granularity levels.
cmplnt_df['Date_month_year'] = pd.to_datetime(cmplnt_df['Date_month_year'])
cmplnt_df['Created_Day'] = cmplnt_df['Date_month_year'].apply(lambda dmy: dmy.day)
cmplnt_df['Created_Day_of_Week'] = cmplnt_df['Date_month_year'].apply(lambda dmy: dmy.dayofweek)
cmplnt_df['Created_Month'] =  cmplnt_df['Date_month_year'].apply(lambda dmy: dmy.month)
```

In [273]:

```
#Number of complaints monthly granularity level
plt.figure(figsize=(10,5))
cmplnt_df = cmplnt_df.sort_values(by='Created_Month')
bymonth = cmplnt_df.groupby('Created_Month').count().reset_index()
lp = cbrn.lineplot(x='Created_Month', y= 'Customer Complaint', data = bymonth, sort=False,markers = "o")
ax = lp.axes
ax.set_xlim(0,15)
ax.annotate('Maximum complaints received during mid of the year', color='black',xy=(6.5, 1040))
```

Out[273]:

```
Text(6.5, 1040, 'Maximum complaints received during mid of the year')
```



In [274]:

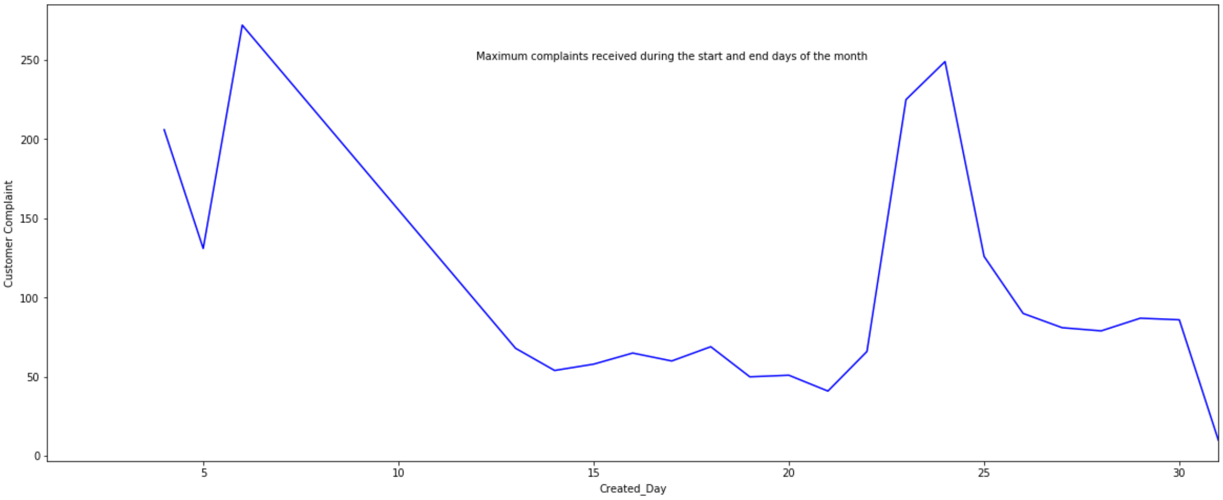
```
#Most number of complaints are received during mid of the year (June -> 6)
```

In [275]:

```
#Number of complaints daily granularity level
plt.figure(figsize=(20,8))
byday = cmplnt_df.groupby('Created_Day').count().reset_index()
lp = cbrn.lineplot(x='Created_Day', y= 'Customer Complaint', data = byday, sort=False, c
olor = 'blue',markers = "o", )
ax = lp.axes
ax.set_xlim(1,31)
ax.annotate('Maximum complaints received during the start and end days of the month', co
lor='black',xy=(12, 250))
```

Out[275]:

```
Text(12, 250, 'Maximum complaints received during the start and end days of the month')
```



In [276]:

```
#The most number of complaints are received during the start and end days of the month,
 with less number of complaints during mid of the month.
#Dates with maximum complaints = 5 to 10 AND 21 to 26
```

In [277]:

```
#Provide a table with the frequency of complaint types.
cmplnt_df['Customer Complaint'] = cmplnt_df['Customer Complaint'].str.title()
cmplnt_freq = cmplnt_df['Customer Complaint'].value_counts()
cmplnt_freq
```

```
Out[277]: Comcast                                                                          102
          Comcast Data Cap                                                                  30
          Comcast Internet                                                                  29
          Comcast Data Caps                                                                 21
          Comcast Billing                                                                   18
          Comcast Service                                                                   15
          Internet Speed                                                                    15
          Unfair Billing Practices                                                          13
          Data Caps                                                                         13
          Data Cap                                                                          12
          Comcast Complaint                                                                 11
          Comcast/Xfinity                                                                   11
          Comcast Internet Service                                                          10
          Billing                                                                            9
          Billing Issues                                                                     8
          Service Issues                                                                     5
          Complaint Against Comcast                                                          5
          Comcast Issues                                                                     5
          Slow Internet                                                                      5
          Comcast Billing Practices                                                          5
          Comcast Cable                                                                      5
          Comcast Billing Complaint                                                          5
          Internet Service                                                                   5
          Internet                                                                           5
          Slow Internet Speeds                                                               4
          Comcast Customer Service                                                           4
          Speed                                                                              4
          Comcast Internet Complaint                                                         4
          Internet Throttling                                                                4
          Internet Availability                                                              4
                                                                                          ...
          Internet Unreliable                                                                1
          Unwanted Email With No Opt Out Options                                             1
          Comcast - Unfair Billing Policies                                                  1
          Miss Represented And Over Charged Now                                              1
          Comcast Internet Issue                                                             1
          Charged For Equipment That I Didn'T Have                                           1
          Comcast Internet Service Provider                                                  1
          Internet Provider                                                                  1
          Double Billing From Comcast                                                        1
          Comcast Lied To Me                                                                 1
          Comcast Data Limits And Overage Charges                                            1
          Xfinity False Advertisement                                                        1
          Comcast Throttling Or Otherwise Under-Delivering Internet Service.                 1
          Fraudulent Charges And Poor Or No Service                                          1
          Data Caps Discourage Commerce                                                      1
          Comcast Unfair Product Bundling                                                    1
          Unnecessary Charge By Comcast For Technician Visit                                 1
          Re: Internet Too Expensive                                                         1
          2 Months And Comcast Has Not Fixed Problem                                         1
          Internet And Phone Service                                                         1
          Very Low Wireless Speed Not Working Well                                           1
          Comcast Forces Rep Visit Then Goes Back On Their Promise                           1
          Complaints About Comcast                                                           1
          Youtube Being Throttled?                                                           1
          Xfinity Throttling My Connection                                                   1
          Over Billing, No Reason Code For Excess Billing, Fraud - Imo                       1
          Xfinity Not Providing Service In My Portion Of The Neighborhood                    1
          Internet Speed/Service                                                             1
          Comcast Asking For Interstructure Fee                                              1
          Comcast Illegally Introduces Service Fees Without Notice Or Customer Consent       1
          Name: Customer Complaint, Length: 1740, dtype: int64
```

In [278]:
```
#Which complaint types are maximum i.e., around internet, network issues, or across any
 other domains.
from collections import Counter
cmplt_common = Counter(" ".join(cmplnt_df["Customer Complaint"]).split()).most_common(50
)
```

In [279]:
```
import nltk
from nltk.corpus import stopwords
```

In [280]:
```
stop_words = set(stopwords.words('english'))
new_stopwords = ['comcast','now','company','day','someone','thing','also','got','way','c
all']
stop_words = stop_words.union(new_stopwords)
cmplt_common_dict = dict(cmplt_common)
areas_flitered = [w for w in cmplt_common_dict.keys() if w.lower() not in stop_words]
print(areas_flitered[0],",",areas_flitered[1],"and",areas_flitered[2],"are the first thr
ee areas with maximum complaints, \
with",cmplt_common_dict[areas_flitered[0]],",",cmplt_common_dict[areas_flitered[1]],"an
d",cmplt_common_dict[areas_flitered[2]],\
     "complaints respectively.")
```

```
Internet , Service and Billing are the first three areas with maximum complaints, with 5
08 , 411 and 273 complaints respectively.
```

In [281]:
```
#Internet , Service and Billing are the first three areas with maximum complaints, with
 508 , 411 and 273 complaints respectively.
```

In [282]:
```
#Create a new categorical variable with value as Open and Closed. Open & Pending is to b
e categorized as Open and Closed & Solved is to be categorized as Closed.
cmplnt_df['Cmplnt_Status'] = ["Open" if Status=="Open" or Status=="Pending" else "Close
d" for Status in cmplnt_df["Status"]]
```

In [283]:
```
cmplnt_df['Cmplnt_Status'].unique()
```

Out[283]:
```
array(['Closed', 'Open'], dtype=object)
```

In [284]:
```
#Provide state wise status of complaints in a stacked bar chart. Use the categorized var
iable from Q3. Provide insights on:
```

In [285]:
```
#Provide state wise status of complaints in a stacked bar chart. Use the categorized var
iable 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 thro
ugh the Internet and customer care calls.

cmplnt_df['State'] = cmplnt_df['State'].str.title()
state_data = cmplnt_df.groupby(['State','Cmplnt_Status']).size().unstack().fillna(0)
```

In [286]:

```
state_data
```

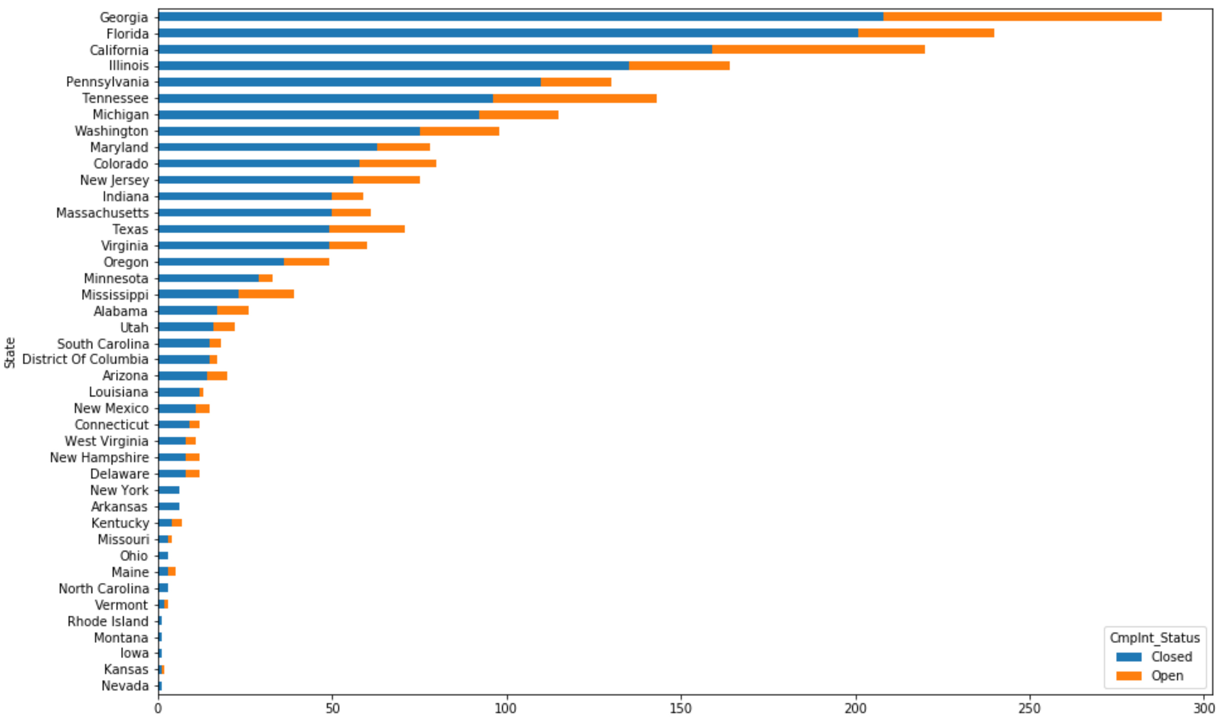
Out[286]:

Cmplnt_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
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

Cmplnt_Status	Closed	Open
State		
Virginia	49.0	11.0
Washington	75.0	23.0
West Virginia	8.0	3.0

In [287]:
```
state_data.sort_values('Closed',axis = 0,ascending=True).plot(kind="barh", figsize=(15,10), stacked=True)
```

Out[287]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1918593de48>
```



In [288]:
```
#Which state has the maximum complaints
```

In [289]:
```
cmplnt_df.groupby(["State"]).size().sort_values(ascending=False).to_frame().rename({0:"Complaint count"}, axis=1)[:1]
```

Out[289]:

	Complaint count
State	
Georgia	288

In [290]:
```
#Georgia state has the maximum complaints
```

In [291]:
```
#Which state has the highest percentage of unresolved complaints
```

In [292]:

```
cmplnt_unresolved = cmplnt_df.groupby(["State","Cmplnt_Status"]).size().unstack().fillna(0)
cmplnt_unresolved.sort_values('Closed',axis = 0,ascending=False)[:1]
```

Out[292]:

Cmplnt_Status	Closed	Open
State		
Georgia	208.0	80.0

In [293]:

```
#Highest percentage of unresolved complaints
cmplnt_unresolved['Resolved_cmp_prct'] = cmplnt_unresolved['Closed']/cmplnt_unresolved['Closed'].sum()*100
cmplnt_unresolved['Unresolved_cmp_prct'] = cmplnt_unresolved['Open']/cmplnt_unresolved['Open'].sum()*100
```

In [294]:

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

Out[294]:

Cmplnt_Status	Closed	Open	Resolved_cmp_prct	Unresolved_cmp_prct
State				
Georgia	208.0	80.0	12.18512	15.473888

In [295]:

```
#Georgia state has the highest percentage of unresolved complaints. (15.47%)
```

In [296]:

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

In [297]:

```
cmplnt_resolved = cmplnt_df.groupby(['Received Via','Cmplnt_Status']).size().unstack().fillna(0)
cmplnt_resolved['resolved'] = cmplnt_resolved['Closed']/cmplnt_resolved['Closed'].sum()*100
cmplnt_resolved['resolved']
```

Out[297]:

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

In [298]:

```
#Complaints which were received via internet & customer care calls were being resolved almost equally (~50%).
```