In [1]:
```
#Importing Libraries needed in the project
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LinearRegression
from sklearn import metrics
```

In [2]:
```
# Import data into Python environment.

df_comcast_comp = pd.read_csv('Comcast_telecom_complaints_data.csv')
df_comcast_comp.head()
```

Out[2]:

	Ticket #	Customer Complaint	Date	Date_month_year	Time	Received Via	City	State	Zip code	Status
0	250635	Comcast Cable Internet Speeds	22-04-15	22-Apr-15	3:53:50 PM	Customer Care Call	Abingdon	Maryland	21009	Closed
1	223441	Payment disappear - service got disconnected	04-08-15	04-Aug-15	10:22:56 AM	Internet	Acworth	Georgia	30102	Closed
2	242732	Speed and Service	18-04-15	18-Apr-15	9:55:47 AM	Internet	Acworth	Georgia	30101	Closed
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
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

In [3]:
```
df_comcast_comp.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
```

In [4]:
```
df_comcast_comp.shape
```

Out[4]:
```
(2224, 11)
```

In [5]:
```
df_comcast_comp["date_index"] = df_comcast_comp["Date_month_year"] + " " + df_comcast_c
df_comcast_comp["date_index"] = pd.to_datetime(df_comcast_comp["date_index"])
df_comcast_comp["Date_month_year"] = pd.to_datetime(df_comcast_comp["Date_month_year"])
```

In [6]:
```
df_comcast_comp.dtypes
```

Out[6]:
```
Ticket #                             object
Customer Complaint                   object
Date                                 object
Date_month_year              datetime64[ns]
Time                                 object
Received Via                         object
City                                 object
State                                object
Zip code                              int64
Status                               object
Filing on Behalf of Someone          object
date_index                   datetime64[ns]
dtype: object
```

In [7]:
```
df_comcast_comp = df_comcast_comp.set_index(df_comcast_comp["date_index"])
```

In [15]:
```
#MOnthly Level
df_comcast_comp.head()
```

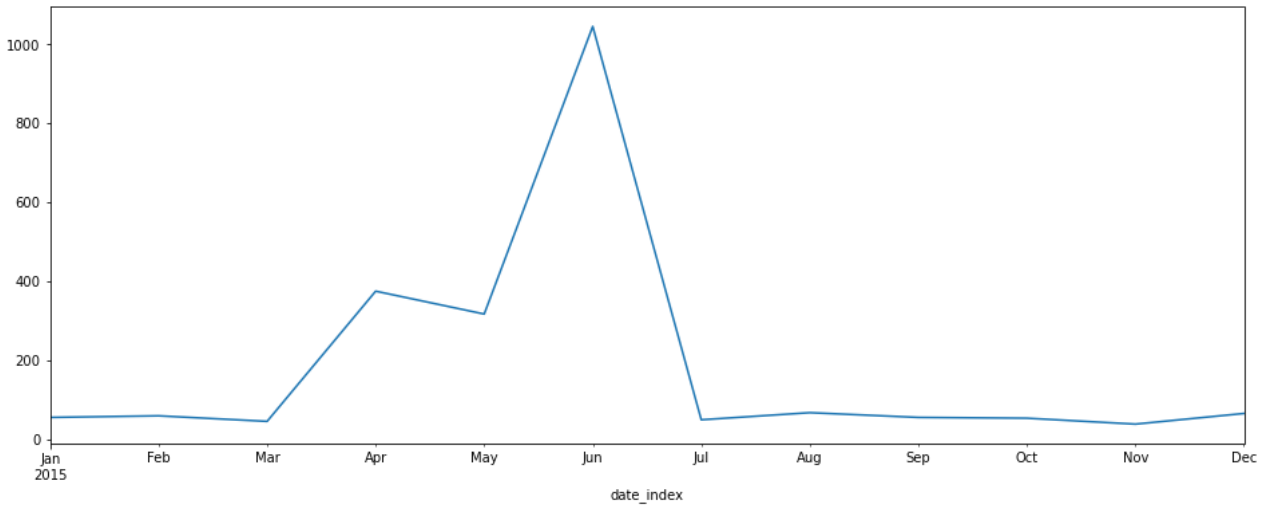
Out[15]:

	Ticket #	Customer Complaint	Date	Date_month_year	Time	Received Via	City	State	Zi coc
date_index									
2015-04-22 15:53:50	250635	Comcast Cable Internet Speeds	22-04-15	2015-04-22	3:53:50 PM	Customer Care Call	Abingdon	Maryland	210(
2015-08-04 10:22:56	223441	Payment disappear - service got disconnected	04-08-15	2015-08-04	10:22:56 AM	Internet	Acworth	Georgia	301(
2015-04-18 09:55:47	242732	Speed and Service	18-04-15	2015-04-18	9:55:47 AM	Internet	Acworth	Georgia	301(
2015-07-05 11:59:35	277946	Comcast Imposed a New Usage Cap of 300GB that ...	05-07-15	2015-07-05	11:59:35 AM	Internet	Acworth	Georgia	301(

date_index	Ticket #	Customer Complaint	Date	Date_month_year	Time	Received Via	City	State	Zi coc
2015-05-26 13:25:26	307175	Comcast not working and no service to boot	26-05-15	2015-05-26	1:25:26 PM	Internet	Acworth	Georgia	301(

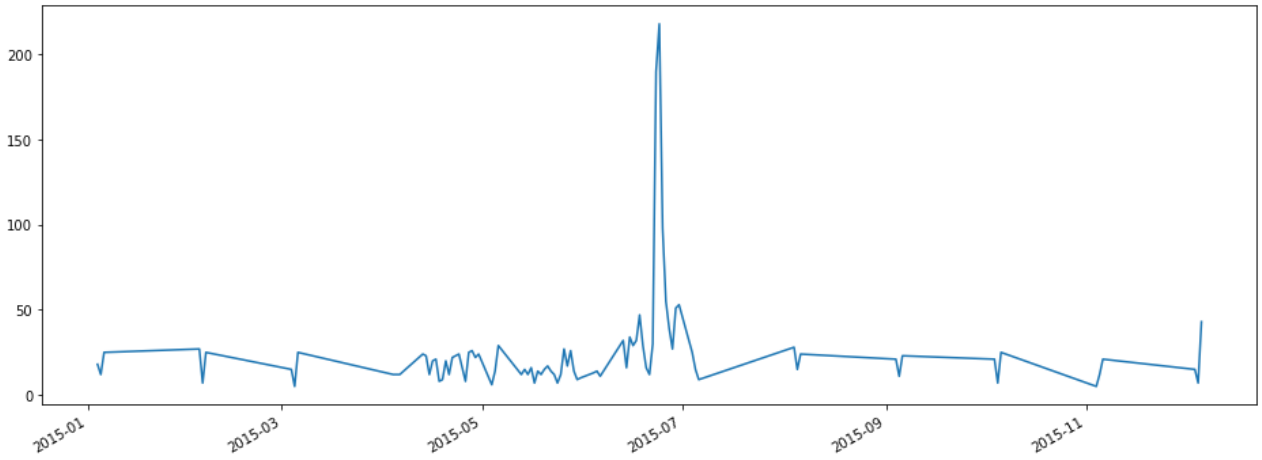
In [28]:
```
#Trend chart for the number of complaints at monthly and daily granularity levels.
df_comcast_comp["Date_month_year"].groupby(pd.Grouper(freq='M')).size().plot(figsize=(1
```

Out[28]: <AxesSubplot:xlabel='date_index'>



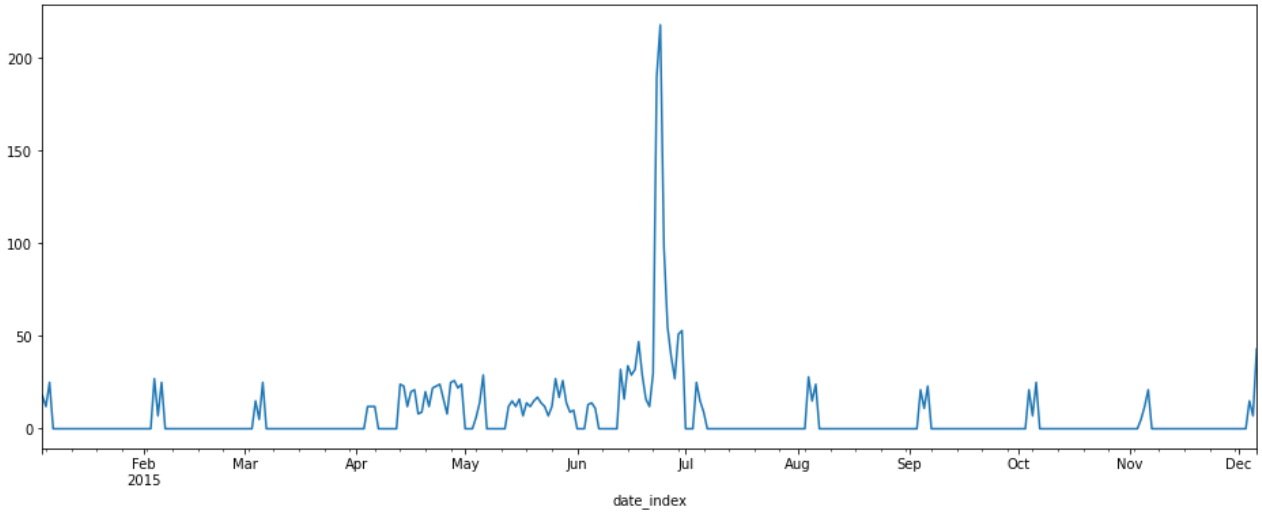
In [30]:
```
df_comcast_comp["Date_month_year"].value_counts().plot(figsize=(16, 6))
```

Out[30]: <AxesSubplot:>



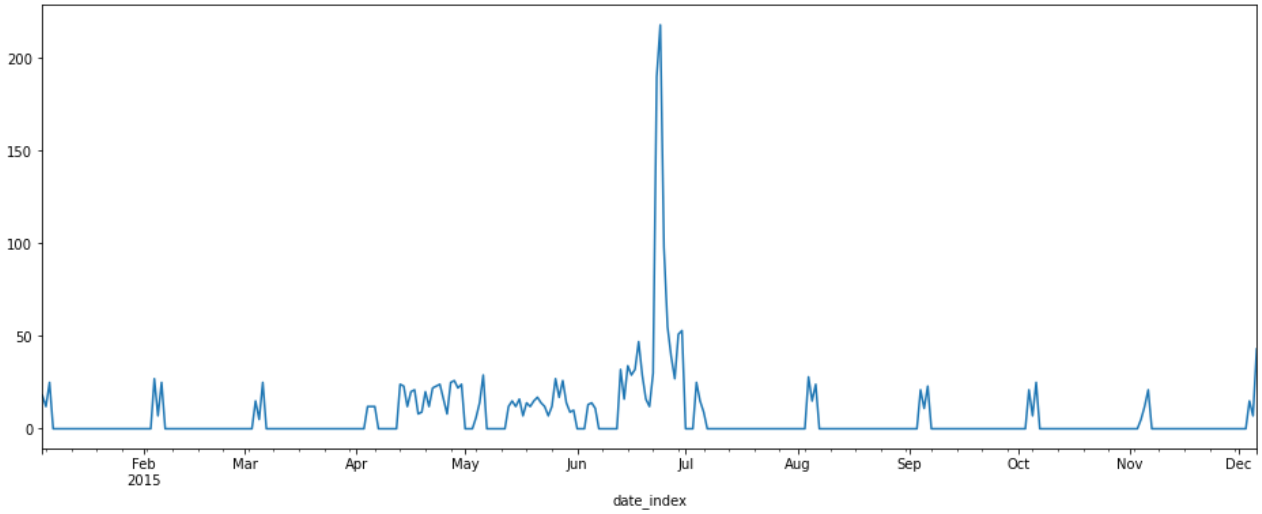
In [27]:
```
df_comcast_comp["Date_month_year"].groupby(pd.Grouper(freq='D')).size().plot(figsize=(
```

Out[27]: <AxesSubplot:xlabel='date_index'>



In [34]:
```
df_comcast_comp["Date_month_year"].groupby(pd.Grouper(freq='D')).size().plot(figsize=(
```

Out[34]: <AxesSubplot:xlabel='date_index'>



In [38]:
```
df_bystate = df_comcast_comp[['Ticket #', 'State']].groupby('State').size()
df_bystate
```

Out[38]:
```
State
Alabama                 26
Arizona                 20
Arkansas                 6
California             220
Colorado                80
Connecticut             12
Delaware                12
District Of Columbia    16
District of Columbia     1
Florida                240
Georgia                288
Illinois               164
Indiana                 59
Iowa                     1
Kansas                   2
Kentucky                 7
Louisiana               13
Maine                    5
Maryland                78
```

```
Massachusetts           61
Michigan               115
Minnesota               33
Mississippi             39
Missouri                 4
Montana                  1
Nevada                   1
New Hampshire           12
New Jersey              75
New Mexico              15
New York                 6
North Carolina           3
Ohio                     3
Oregon                  49
Pennsylvania           130
Rhode Island             1
South Carolina          18
Tennessee              143
Texas                   71
Utah                    22
Vermont                  3
Virginia                60
Washington              98
West Virginia           11
dtype: int64
```

In [41]:
```
df_bystate.nlargest(5)
```

Out[41]:
```
State
Georgia       288
Florida       240
California    220
Illinois      164
Tennessee     143
dtype: int64
```

In [42]:
```
#state has the maximum complaints
print("State with Maximum Complaints", df_bystate.nlargest(1))
```

```
State with Maximum Complaints State
Georgia    288
dtype: int64
```

In [45]:
```
df_comcast_comp[['Ticket #', 'Status']].groupby("Status").size()
```

Out[45]:
```
Status
Closed     734
Open       363
Pending    154
Solved     973
dtype: int64
```

In [47]:
```
#New Categorical variable with value as Open and Closed.
df_comcast_comp["StatusNew"] = ['Open' if x =='Open' or x =='Pending' else 'Closed' for
print("Open complaints ",df_comcast_comp['StatusNew'].loc[df_comcast_comp["StatusNew"]
```

```
Open complaints  517 Closed Complinats 1707
```

In [56]:
```
df_grouped = df_comcast_comp.groupby(['State', 'StatusNew'])['State'].size()
df_grouped.head()
```

Out[56]:
```
State     StatusNew
Alabama   Closed       17
          Open          9
```

```
Arizona   Closed       14
          Open          6
Arkansas  Closed        6
Name: State, dtype: int64
```

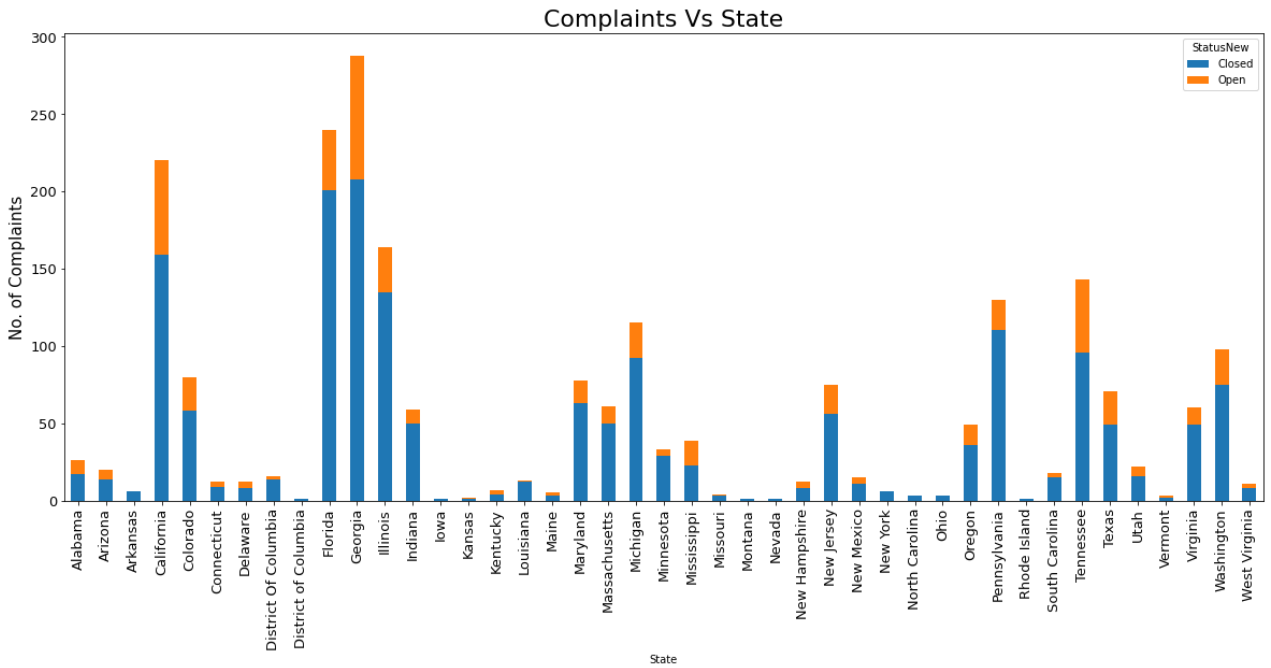
In [55]:
```
#state wise status of complaints in a stacked bar chart
df_StatusComplaints = df_comcast_comp.groupby(['State', 'StatusNew']).size().unstack().
df_StatusComplaints
```

Out[55]:

StatusNew	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

StatusNew	Closed	Open
State		
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 [57]:
```
ax = df_grouped.unstack().plot(kind='bar', figsize=(20,8), fontsize=13, stacked=True)
ax.set_alpha(0.8)
ax.set_title("Complaints Vs State", fontsize=22)
ax.set_ylabel("No. of Complaints", fontsize=15);
plt.show()
```



In [ ]:
```
fig = plt.figure(figuresize(10,6))
```

In [67]:
```
df_StatusComplaints['Resolved Complaints'] = (df_StatusComplaints['Closed'])/((df_Statu
print( " percentage of complaints resolved till date ", df_StatusComplaints['Resolved C
```

```
percentage of complaints resolved till date  79.51045168976003
```

In [73]:
```
#state has the highest percentage of unresolved complaints
print("State with more unresolved complaints till date ", df_StatusComplaints['Resolved
```

```
State with more unresolved complaints till date  State
Kansas    0.5
Name: Resolved Complaints, dtype: float64
```

In [ ]: