In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
from patsy import dmatrices
import sklearn
import seaborn as sns
import datetime as dt
from datetime import datetime
from collections import Counter
from nltk.corpus import stopwords
stop = stopwords.words('english')
from matplotlib import rc
import plotly.express as px
```

In [2]:
```
dataframe = pd.read_csv(r"C:\Users\ctoqu\Desktop\Comcast_telecom_complaints_data.csv")
```

In [3]:
```
dataframe.head(5)
```

Out[3]:

	Ticket #	Customer Complaint	Date	Date_month_year	Time	Received Via	City	State	Zip code
0	250635	Comcast Cable Internet Speeds	22-04-15	22-Apr-15	3:53:50 PM	Customer Care Call	Abingdon	Maryland	21009
1	223441	Payment disappear - service got disconnected	04-08-15	04-Aug-15	10:22:56 AM	Internet	Acworth	Georgia	30102
2	242732	Speed and Service	18-04-15	18-Apr-15	9:55:47 AM	Internet	Acworth	Georgia	30101
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
4	307175	Comcast not working and no service to boot	26-05-15	26-May-15	1:25:26 PM	Internet	Acworth	Georgia	30101

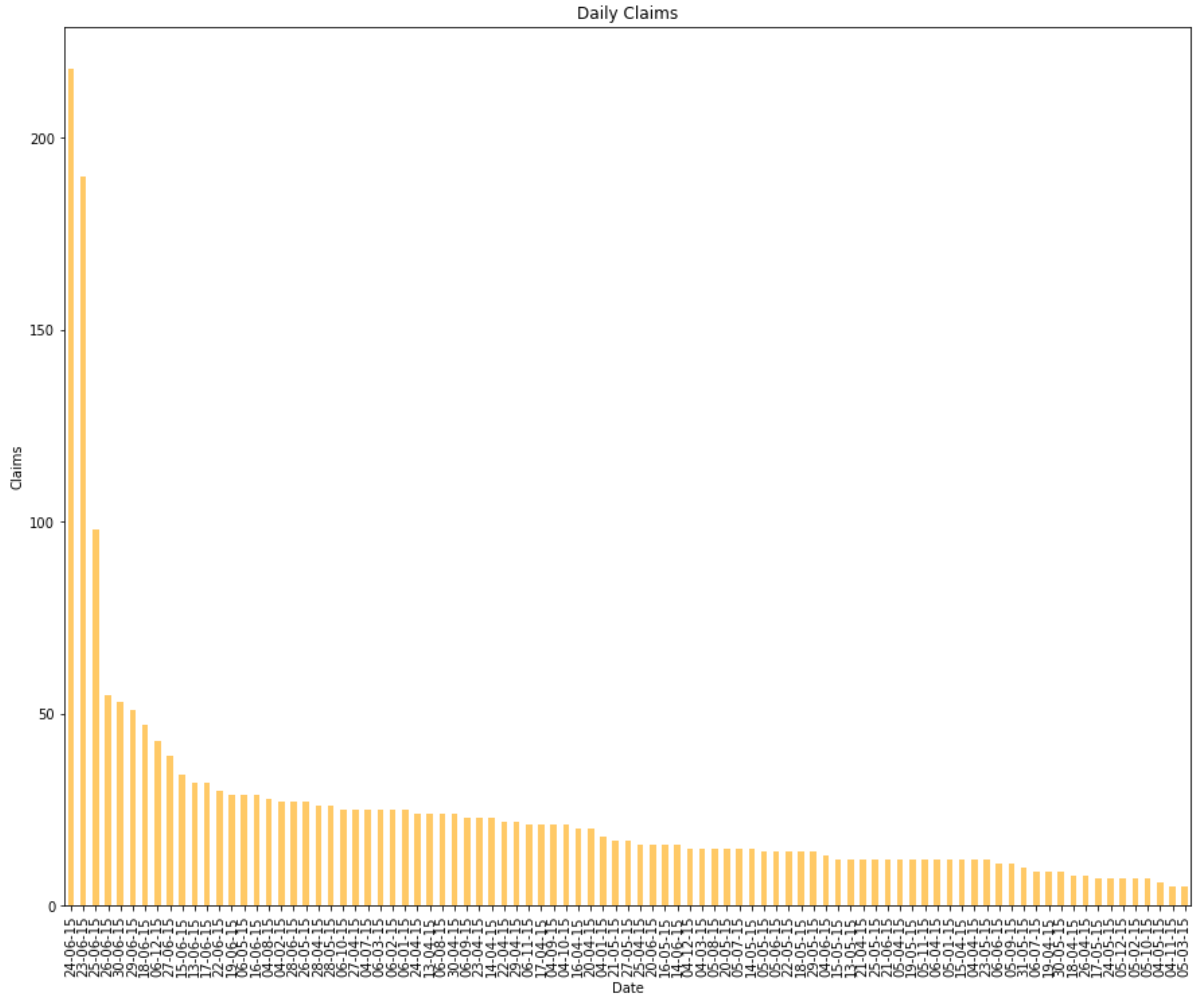
In [4]:
```
titles = dataframe.columns.values
print(titles)
```

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

In [5]:
```
dataframe['Date'].value_counts(dropna=False)
```

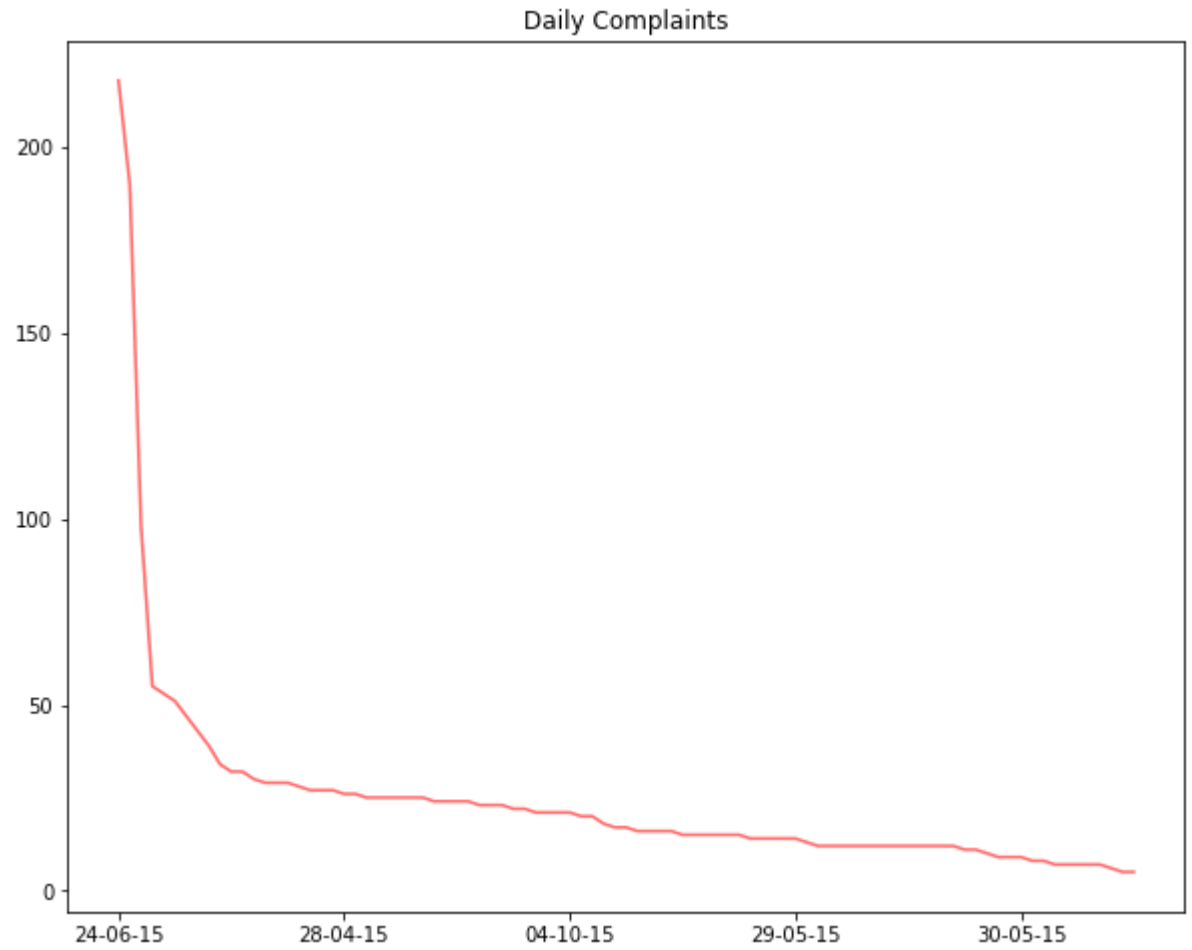
Out[5]:
```
24-06-15    218
23-06-15    190
25-06-15     98
26-06-15     55
30-06-15     53
            ...
05-02-15      7
05-10-15      7
04-05-15      6
04-11-15      5
05-03-15      5
Name: Date, Length: 91, dtype: int64
```

In [6]:
```
#Trend chart of daily complaints
plt.figure(figsize=(15,12))
dataframe['Date'].value_counts(dropna=False).plot(kind='bar', color = 'orange'
,alpha=.6)
plt.title('Daily Claims')
plt.xlabel('Date')
plt.ylabel('Claims')
plt.show()
```



The graphic above shows an important increment of complaints during June with a number of complaints of 218 on June 24, 190 on June 23 and 98 on June 25.

In [7]:
```
#Trend Chart(2) Monthly
plt.figure(figsize=(10,8))
dataframe.Date.value_counts().plot(kind='line',color = 'r',alpha=.6)
plt.title('Daily Complaints')
plt.show()
```



In [8]:
```
# Number of complaints by day
dataframe['Date_month_year'].value_counts(dropna=False)
```

Out[8]:
```
24-Jun-15    218
23-Jun-15    190
25-Jun-15     98
26-Jun-15     55
30-Jun-15     53
             ...
24-May-15      7
05-Oct-15      7
04-May-15      6
05-Mar-15      5
04-Nov-15      5
Name: Date_month_year, Length: 91, dtype: int64
```

In [9]:
```
dataframe.sort_values('Date_month_year')
```

Out[9]:

	Ticket #	Customer Complaint	Date	Date_month_year	Time	Received Via	City	State
1416	218108	Comcast Business Phone/Internet Contract Disag...	04-04-15	04-Apr-15	6:39:55 PM	Internet	Newnan	Georgia
1483	217985	bait and switch services for monetary gain	04-04-15	04-Apr-15	4:07:36 PM	Internet	Orcutt	California
584	217999	Misleading information given	04-04-15	04-Apr-15	4:21:46 PM	Internet	Des Moines	Washington
561	218043	comcast services	04-04-15	04-Apr-15	5:32:05 PM	Internet	Denver	Colorado
1892	218168	Multiple Unauthorized and Unwarranted Credit C...	04-04-15	04-Apr-15	8:10:35 PM	Customer Care Call	Shoreview	Minnesota
...	...	...	...	...	...	...	...	...
634	315836	Complaint against Comcast for incredibly bad s...	31-05-15	31-May-15	4:47:08 PM	Customer Care Call	Edgewood	Washington
668	315602	Comcast deceptive selling billing lack of serv...	31-05-15	31-May-15	10:40:45 AM	Internet	Evans	Georgia
329	315997	Hidden Product Installation Fee	31-05-15	31-May-15	8:27:53 PM	Customer Care Call	Camp Hill	Pennsylvania
874	315608	Comcast Service Failure	31-05-15	31-May-15	10:56:25 AM	Internet	Houston	Texas
464	315874	Comcast Termination Fee	31-05-15	31-May-15	5:54:06 PM	Customer Care Call	Colorado Springs	Colorado

2224 rows × 11 columns

In [10]:
```
print(type('Date_month_year'))
```

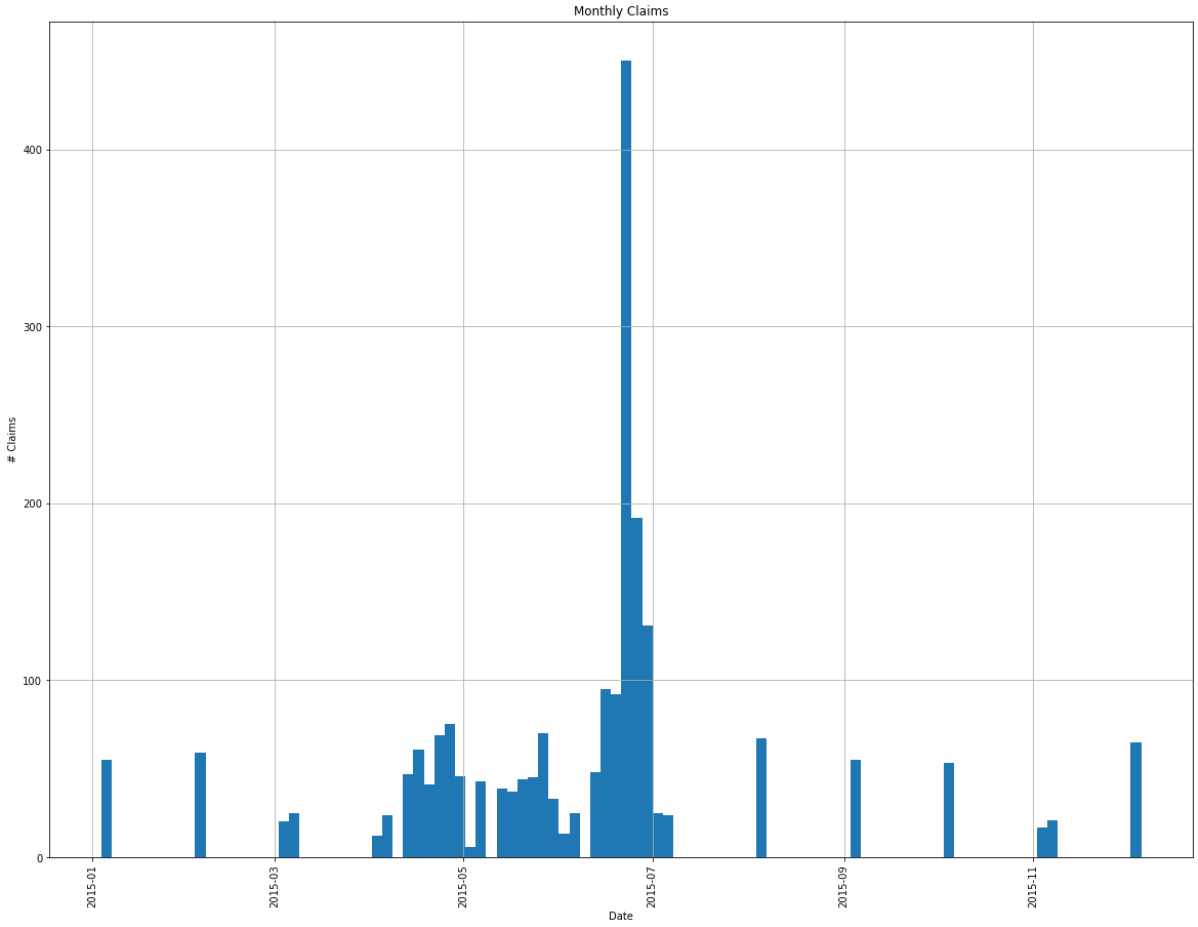
```
<class 'str'>
```

In [11]:
```
# Convert 'Date_month_year' column from string to datetime class
dataframe['Date_month_year']=pd.to_datetime(dataframe['Date_month_year'])
dataframe.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   datetime64[ns]
 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: datetime64[ns](1), int64(1), object(9)
memory usage: 191.2+ KB
```

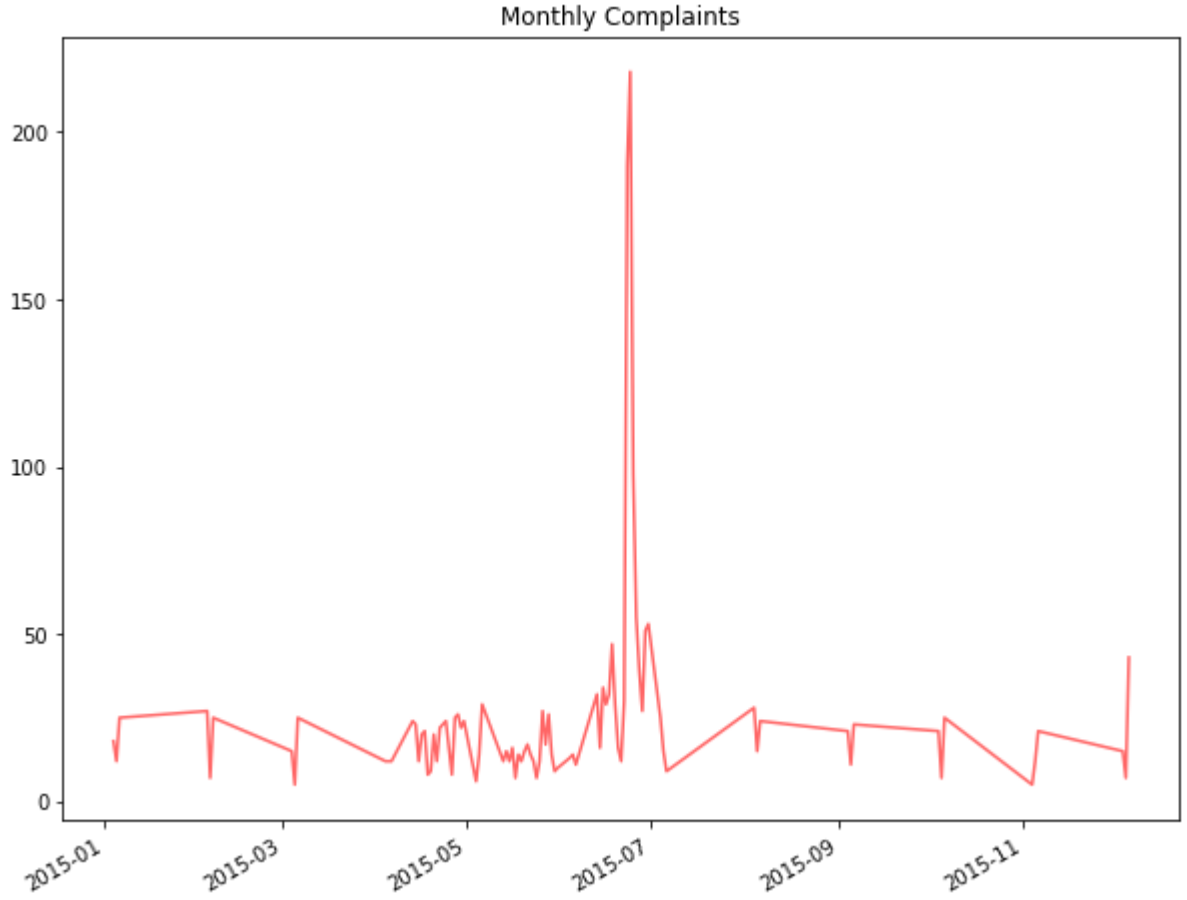
In [12]:
```
# #Trend chart of complaints by month

plt.figure(figsize=(20,15))
dataframe['Date_month_year'].hist(bins=100)
plt.xticks(rotation=90)
plt.title('Monthly Claims')
plt.xlabel('Date')
plt.ylabel('# Claims')
plt.show()
```



As seen on the previous graph, this monthly chart shows an overall increment of complaints during the months of May and June with a pic of in the month of June.

In [13]:
```
#Trend Chart Monthly
plt.figure(figsize=(10,8))
dataframe.Date_month_year.value_counts().plot(kind='line',color = 'r',alpha=.6
)
plt.title('Monthly Complaints')
plt.show()
```



In [14]:
```
#look compaints type in june
```

In [15]:
```
type("Customer Complaint")
```

Out[15]:
```
str
```

In [16]:
```
# Exclude stopwords with Python's list comprehension and pandas.DataFrame.appl
y.
dataframe['Customer Complaints_without_stopwords'] = dataframe['Customer Compl
aint'].apply(lambda x: ' '.join([word for word in x.split() if word not in (st
op)]))
print(dataframe)
```

```
     Ticket #                                Customer Complaint      Date  \
0      250635                     Comcast Cable Internet Speeds  22-04-15
1      223441       Payment disappear - service got disconnected  04-08-15
2      242732                                  Speed and Service  18-04-15
3      277946  Comcast Imposed a New Usage Cap of 300GB that ...  05-07-15
4      307175         Comcast not working and no service to boot  26-05-15
...       ...                                                ...       ...
2219   213550                           Service Availability  04-02-15
2220   318775         Comcast Monthly Billing for Returned Modem  06-02-15
2221   331188                           complaint about comcast  06-09-15
2222   360489             Extremely unsatisfied Comcast customer  23-06-15
2223   363614               Comcast, Ypsilanti MI Internet Speed  24-06-15

     Date_month_year          Time        Received Via        City     State
\
0         2015-04-22   3:53:50 PM  Customer Care Call    Abingdon  Maryland
1         2015-08-04  10:22:56 AM            Internet     Acworth   Georgia
2         2015-04-18   9:55:47 AM            Internet     Acworth   Georgia
3         2015-07-05  11:59:35 AM            Internet     Acworth   Georgia
4         2015-05-26   1:25:26 PM            Internet     Acworth   Georgia
...              ...          ...                 ...         ...       ...
2219      2015-02-04   9:13:18 AM  Customer Care Call  Youngstown   Florida
2220      2015-02-06   1:24:39 PM  Customer Care Call   Ypsilanti  Michigan
2221      2015-09-06   5:28:41 PM            Internet   Ypsilanti  Michigan
2222      2015-06-23  11:13:30 PM  Customer Care Call   Ypsilanti  Michigan
2223      2015-06-24  10:28:33 PM  Customer Care Call   Ypsilanti  Michigan

      Zip code  Status Filing on Behalf of Someone  \
0        21009  Closed                          No
1        30102  Closed                          No
2        30101  Closed                         Yes
3        30101    Open                         Yes
4        30101  Solved                          No
...        ...     ...                         ...
2219     32466  Closed                          No
2220     48197  Solved                          No
2221     48197  Solved                          No
2222     48197  Solved                          No
2223     48198    Open                         Yes

                Customer Complaints_without_stopwords
0                       Comcast Cable Internet Speeds
1         Payment disappear - service got disconnected
2                                      Speed Service
3     Comcast Imposed New Usage Cap 300GB punishes s...
4                         Comcast working service boot
...                                                ...
2219                           Service Availability
2220            Comcast Monthly Billing Returned Modem
2221                              complaint comcast
2222            Extremely unsatisfied Comcast customer
2223              Comcast, Ypsilanti MI Internet Speed

[2224 rows x 12 columns]
```

In [17]:
```
dataframe.head()
```

Out[17]:

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

In [18]:
```
#Frequency table showing the words that appear frequently in the complaints:

from collections import Counter
Counter(" ".join(dataframe["Customer Complaints_without_stopwords"].str.lower
()).split()).most_common(20)
```

Out[18]:
```
[('comcast', 1160),
 ('internet', 508),
 ('service', 411),
 ('billing', 273),
 ('data', 217),
 ('speed', 112),
 ('cap', 103),
 ('customer', 87),
 ('issues', 85),
 ('caps', 76),
 ('practices', 70),
 ('charges', 69),
 ('complaint', 68),
 ('speeds', 65),
 ('bill', 59),
 ('xfinity', 58),
 ('unfair', 58),
 ('-', 53),
 ('services', 52),
 ('throttling', 51)]
```

This Frequency table shows the words that appear more frequently in the complaints:

- Comcast is the word used more frequently but is often used in combination with Internet, Billing or customer service.
- The second most common word is Internet. We also see other terms related to this type of issue like speed, cap(s),data,throttling, xfinity and,service.
- Third is Billing, including other common words like charges and bill.
- Finally we see other words being used frequently in the compalints like Customer Service, practices and, unfair.

In [19]:
```
dataframe["Customer Complaint"].dtypes
```

Out[19]:
```
dtype('O')
```

In [20]:
```
#Create new categorical Value with Open and Closed
dataframe['Open_Closed'] = dataframe['Status']
dataframe['Open_Closed'] = dataframe['Status']
```

In [21]:
```
# Setting Open and closed categories
dataframe['Open_Closed'].replace('Pending','Open', inplace=True)
dataframe['Open_Closed'].replace('Solved','Closed', inplace=True)
```

In [22]:
```
#Open and Closed counts
dataframe['Open_Closed'].value_counts()
```

Out[22]:
```
Closed    1707
Open       517
Name: Open_Closed, dtype: int64
```

In [38]:
```
#Percentage of Open and Close counts
dataframe.Open_Closed.value_counts(normalize=True)
```

Out[38]:
```
Closed    0.767536
Open      0.232464
Name: Open_Closed, dtype: float64
```

In [23]:
```
dataframe.head(10)
```

Out[23]:

	Ticket #	Customer Complaint	Date	Date_month_year	Time	Received Via	City	State	Zip code
0	250635	Comcast Cable Internet Speeds	22-04-15	2015-04-22	3:53:50 PM	Customer Care Call	Abingdon	Maryland	21009
1	223441	Payment disappear - service got disconnected	04-08-15	2015-08-04	10:22:56 AM	Internet	Acworth	Georgia	30102
2	242732	Speed and Service	18-04-15	2015-04-18	9:55:47 AM	Internet	Acworth	Georgia	30101
3	277946	Comcast Imposed a New Usage Cap of 300GB that ...	05-07-15	2015-07-05	11:59:35 AM	Internet	Acworth	Georgia	30101
4	307175	Comcast not working and no service to boot	26-05-15	2015-05-26	1:25:26 PM	Internet	Acworth	Georgia	30101
5	338519	ISP Charging for arbitrary data limits with ov...	06-12-15	2015-12-06	9:59:40 PM	Internet	Acworth	Georgia	30101
6	361148	Throttling service and unreasonable data caps	24-06-15	2015-06-24	10:13:55 AM	Customer Care Call	Acworth	Georgia	30101
7	359792	Comcast refuses to help troubleshoot and corre...	23-06-15	2015-06-23	6:56:14 PM	Internet	Adrian	Michigan	49221
8	318072	Comcast extended outages	06-01-15	2015-01-06	11:46:30 PM	Customer Care Call	Alameda	California	94502
9	371214	Comcast Raising Prices and Not Being Available...	28-06-15	2015-06-28	6:46:31 PM	Customer Care Call	Alameda	California	94501

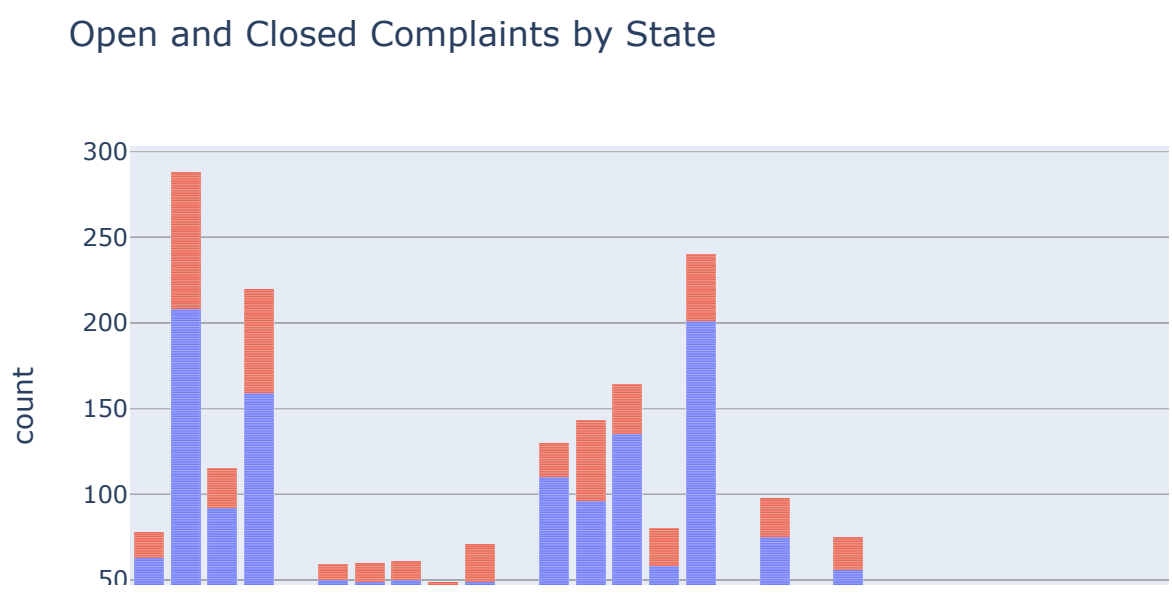
In [24]:
```
dataframe["State"].dtypes
```

Out[24]:
```
dtype('O')
```

In [25]:
```
#Stacked bar chart by state and Open/Close cases (w Ploty)

fig = px.bar(dataframe, x="State", color="Open_Closed", title="Open and Closed
Complaints by State", barmode = 'stack')

fig.show()
```



The state with more complaints overall in 2015 was Georgia, followed by Florida and California

In [26]:
```
# Setting data to analyse Q3
#setting Date_Month_Year as index
Q3_df = dataframe.set_index("Date_month_year")
```

In [27]:
```
Q3_df.head()
```

Out[27]:

Date_month_year	Ticket #	Customer Complaint	Date	Time	Received Via	City	State	Zip code	Sta
2015-04-22	250635	Comcast Cable Internet Speeds	22-04-15	3:53:50 PM	Customer Care Call	Abingdon	Maryland	21009	Clo
2015-08-04	223441	Payment disappear - service got disconnected	04-08-15	10:22:56 AM	Internet	Acworth	Georgia	30102	Clo
2015-04-18	242732	Speed and Service	18-04-15	9:55:47 AM	Internet	Acworth	Georgia	30101	Clo
2015-07-05	277946	Comcast Imposed a New Usage Cap of 300GB that ...	05-07-15	11:59:35 AM	Internet	Acworth	Georgia	30101	C
2015-05-26	307175	Comcast not working and no service to boot	26-05-15	1:25:26 PM	Internet	Acworth	Georgia	30101	So

In [28]:
```
#Filtering dates for Q3 with  pandas.DataFrame.loc

Q3 = Q3_df.loc['2015-07-01':'2015-09-30']
```

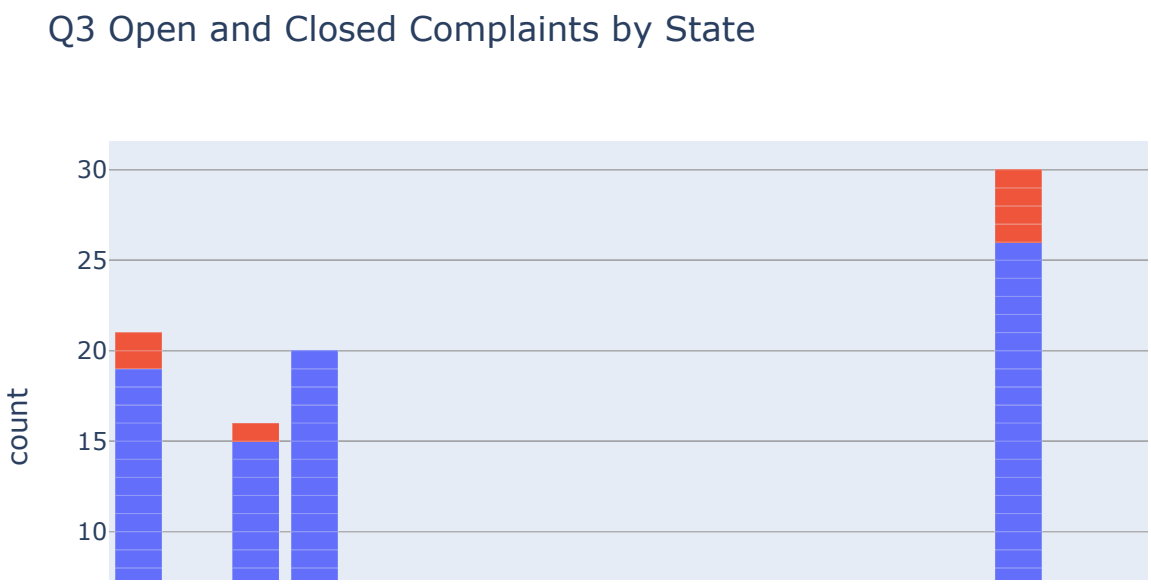
In [29]:
```
Q3.head()
```

Out[29]:

Date_month_year	Ticket #	Customer Complaint	Date	Time	Received Via	City	State	Zip code	Sta
2015-08-04	223441	Payment disappear - service got disconnected	04-08-15	10:22:56 AM	Internet	Acworth	Georgia	30102	Clo
2015-07-05	277946	Comcast Imposed a New Usage Cap of 300GB that ...	05-07-15	11:59:35 AM	Internet	Acworth	Georgia	30101	O
2015-07-06	327657	Internet out all the time but they have a mono...	06-07-15	8:55:43 PM	Customer Care Call	Alexandria	Virginia	22305	So
2015-08-06	328742	horrible cable service and customer service	06-08-15	3:18:58 PM	Internet	Alexandria	Virginia	22312	So
2015-08-06	328165	Speed	06-08-15	12:03:37 PM	Customer Care Call	Alexandria	Virginia	22304	So

In [30]:
```
#Bar chart by State and Open and Closed cases
fig = px.bar(Q3, x="State", color="Open_Closed", title="Q3 Open and Closed Com
plaints by State", barmode = 'stack')

fig.show()
```



The state with more complaints during Q3 is Florida

In [31]:
```
Q3['Open_Closed'].value_counts()
```

Out[31]:
```
Closed    156
Open       15
Name: Open_Closed, dtype: int64
```

In [32]:
```
Q3['Received Via'].value_counts()
```

Out[32]:
```
Customer Care Call    92
Internet              79
Name: Received Via, dtype: int64
```

In [41]:
```
Q3['Received Via'].value_counts(normalize=True)
```

Out[41]:
```
Customer Care Call    0.538012
Internet              0.461988
Name: Received Via, dtype: float64
```

In [33]:
```
Q3_Open = Q3.loc[Q3.Open_Closed == "Open"]
```

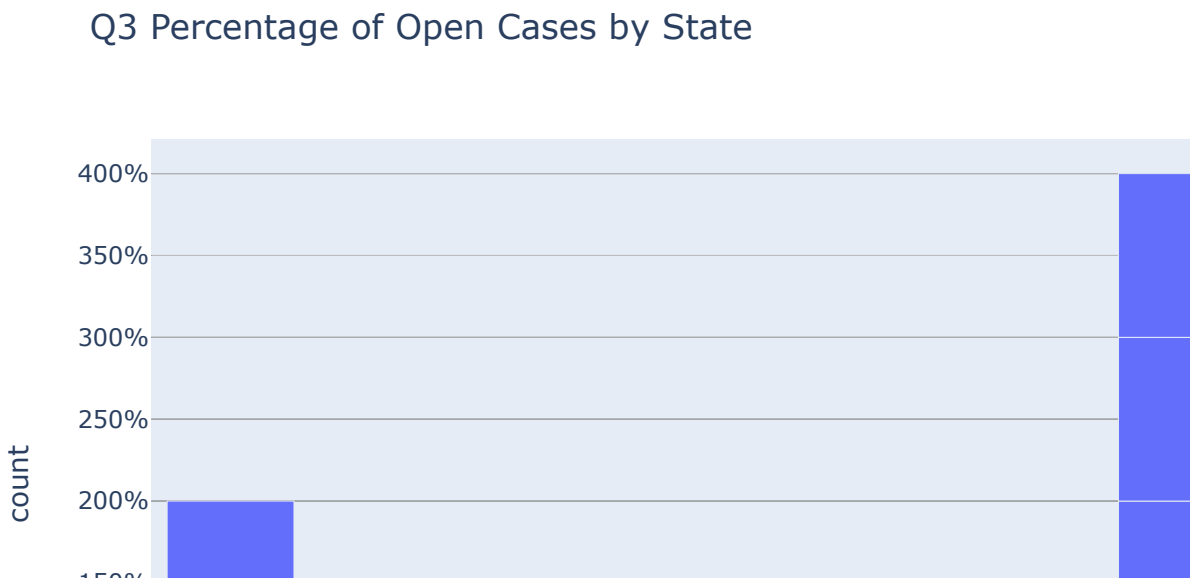
In [34]:
```
Q3_Open.head()
```

Out[34]:

Date_month_year	Ticket #	Customer Complaint	Date	Time	Received Via	City	State	Zip code
2015-07-05	277946	Comcast Imposed a New Usage Cap of 300GB that ...	05-07-15	11:59:35 AM	Internet	Acworth	Georgia	30101
2015-07-05	278509	Data caps	05-07-15	2:55:45 PM	Internet	Atlanta	Georgia	30363
2015-07-05	279378	Forced Bundling of Internet Service with CATV	05-07-15	9:16:13 PM	Customer Care Call	Ayer	Massachusetts	1432
2015-07-05	279279	Comcast Cable connection from street is underw...	05-07-15	7:52:46 PM	Internet	Castro Valley	California	94552
2015-07-06	327672	Comcast Internet	06-07-15	9:26:44 PM	Internet	Charleston	South Carolina	48230

In [35]:
```
#State with the highest percentage of unresolved complaints is Florida folowed
by Georgia and NJ
fig = px.bar(Q3_Open, x="State", color="Open_Closed", title="Q3 Percentage of
 Open Cases by State", barmode = 'stack')
fig.layout.yaxis.tickformat = '%'

fig.show()
```

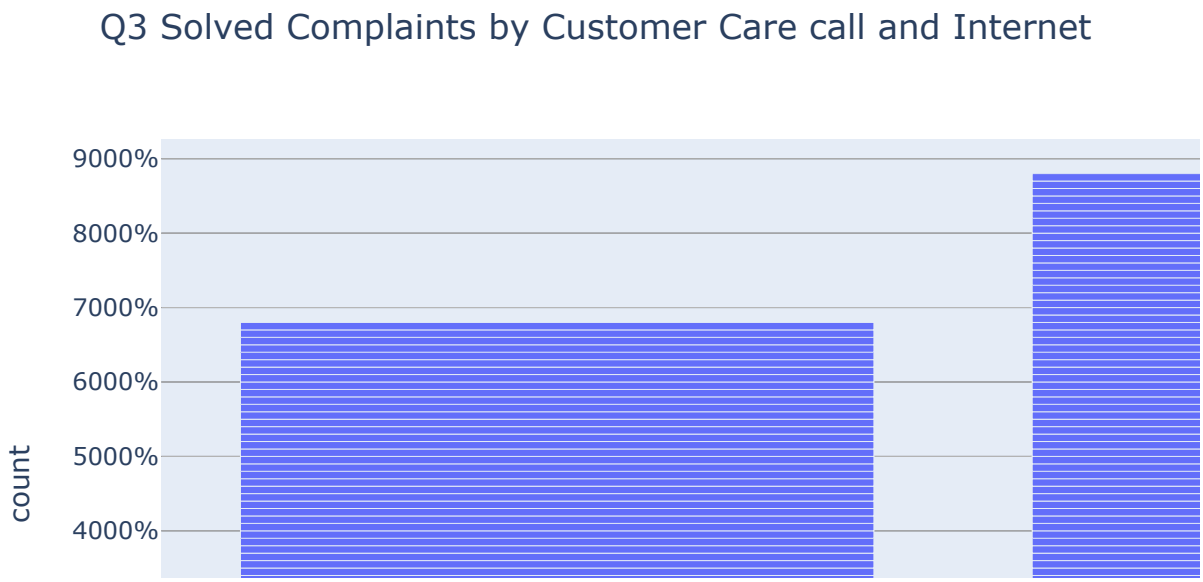


Florida is the state with a higher percentage of unresolved complaints during Q3

In [36]:
```
Q3_Closed = Q3.loc[Q3.Open_Closed == "Closed"]
```

In [37]:
```
#Percentage of Solved cases during Q3 Via Internet and Customer center Call

fig = px.bar(Q3_Closed, x="Received Via", color="Open_Closed", title="Q3 Solve
d Complaints by Customer Care call and Internet ", barmode = 'stack')
fig.layout.yaxis.tickformat = '%'
fig.show()
```



In [42]:
```
#Percentage of Closed complaints via Internet and Customer Care Call during Q3
Q3_Closed['Received Via'].value_counts(normalize=True)
```

Out[42]:
```
Customer Care Call    0.564103
Internet              0.435897
Name: Received Via, dtype: float64
```

During Q3 56% of the complaints were solved with Costumer care calls.

In [ ]:
```

```