In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [2]:
```
df = pd.read_csv('C:\\Users\\sobran\\Documents\\Python DS project\\comcast telecom c
```

In [3]:
```
df.describe()
```

Out[3]:

	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 [4]:
```
df.describe(include = 'all')
```

Out[4]:

	Ticket #	Customer Complaint	Date	Date_month_year	Time	Received Via	City	State	Zip coc
count	2224	2224	2224	2224	2224	2224	2224	2224	2224.0000(
unique	2224	1841	91	91	2190	2	928	43	Na
top	376227	Comcast	24-06-2015	24-Jun-15	1:29:58 PM	Customer Care Call	Atlanta	Georgia	Na
freq	1	83	218	218	2	1119	63	288	Na
mean	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	47994.39343
std	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	28885.27942
min	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1075.0000(
25%	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	30056.5000(
50%	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	37211.0000(
75%	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	77058.7500(
max	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	99223.0000(

In [5]:
```
df.head()
```

Out[5]:

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

In [6]:
```
df.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 [7]:
```
df.isnull().sum()
```

Out[7]:
```
Ticket #                     0
Customer Complaint           0
Date                         0
Date_month_year              0
Time                         0
Received Via                 0
City                         0
State                        0
Zip code                     0
Status                       0
```

```
Filing on Behalf of Someone    0
dtype: int64
```

In [8]:
```
df = df.drop(['Ticket #','Time'],axis=1)
```

In [9]:
```
df.head()
```

Out[9]:

	Customer Complaint	Date	Date_month_year	Received Via	City	State	Zip code	Status	Filing on Behalf of Someone
0	Comcast Cable Internet Speeds	22-04-2015	22-Apr-15	Customer Care Call	Abingdon	Maryland	21009	Closed	No
1	Payment disappear - service got disconnected	04-08-2015	04-Aug-15	Internet	Acworth	Georgia	30102	Closed	No
2	Speed and Service	18-04-2015	18-Apr-15	Internet	Acworth	Georgia	30101	Closed	Yes
3	Comcast Imposed a New Usage Cap of 300GB that ...	05-07-2015	05-Jul-15	Internet	Acworth	Georgia	30101	Open	Yes
4	Comcast not working and no service to boot	26-05-2015	26-May-15	Internet	Acworth	Georgia	30101	Solved	No

In [11]:
```
df['Date_month_year'] = df['Date_month_year'].apply(pd.to_datetime)
```

In [12]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2224 entries, 0 to 2223
Data columns (total 9 columns):
 #   Column                      Non-Null Count  Dtype         
---  ------                      --------------  -----         
 0   Customer Complaint          2224 non-null   object        
 1   Date                        2224 non-null   object        
 2   Date_month_year             2224 non-null   datetime64[ns]
 3   Received Via                2224 non-null   object        
 4   City                        2224 non-null   object        
 5   State                       2224 non-null   object        
 6   Zip code                    2224 non-null   int64         
 7   Status                      2224 non-null   object        
 8   Filing on Behalf of Someone  2224 non-null   object        
dtypes: datetime64[ns](1), int64(1), object(7)
memory usage: 156.5+ KB
```

In [13]:
```
df = df.set_index('Date_month_year')
```

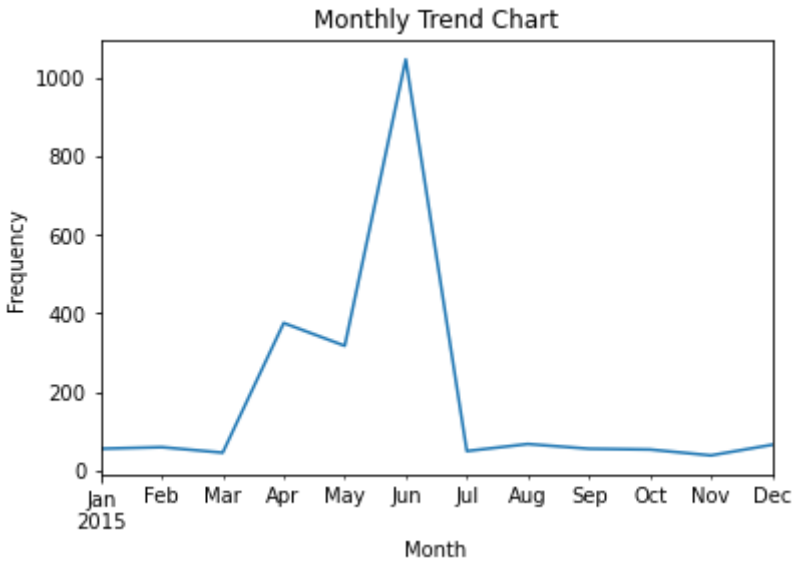
In [14]:
```
df.head()
```

Out[14]:

Date_month_year	Customer Complaint	Date	Received Via	City	State	Zip code	Status	Filing on Behalf of Someone
2015-04-22	Comcast Cable Internet Speeds	22-04-2015	Customer Care Call	Abingdon	Maryland	21009	Closed	No
2015-08-04	Payment disappear - service got disconnected	04-08-2015	Internet	Acworth	Georgia	30102	Closed	No
2015-04-18	Speed and Service	18-04-2015	Internet	Acworth	Georgia	30101	Closed	Yes
2015-07-05	Comcast Imposed a New Usage Cap of 300GB that ...	05-07-2015	Internet	Acworth	Georgia	30101	Open	Yes
2015-05-26	Comcast not working and no service to boot	26-05-2015	Internet	Acworth	Georgia	30101	Solved	No

In [15]:
```
df.groupby(pd.Grouper(freq='M')).size().plot()
plt.xlabel('Month')
plt.ylabel('Frequency')
plt.title('Monthly Trend Chart')
```

Out[15]:
```
Text(0.5, 1.0, 'Monthly Trend Chart')
```



In [17]:
```
df['Date'].value_counts()[:15]
```

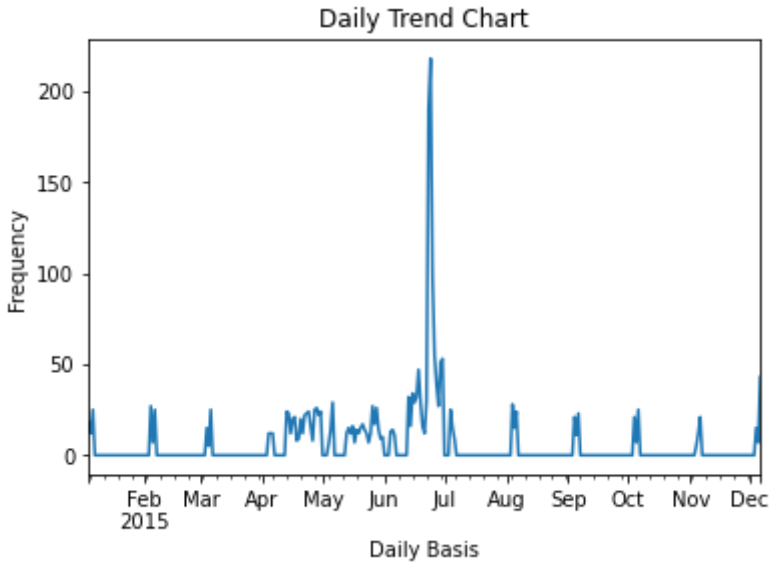
Out[17]:
```
24-06-2015    218
23-06-2015    190
25-06-2015     98
26-06-2015     55
```

```
30-06-2015     53
29-06-2015     51
18-06-2015     47
06-12-2015     43
27-06-2015     39
15-06-2015     34
17-06-2015     32
13-06-2015     32
22-06-2015     30
16-06-2015     29
19-06-2015     29
Name: Date, dtype: int64
```

In [20]:

```
df.groupby(pd.Grouper(freq='D')).size().plot()
plt.xlabel('Daily Basis')
plt.ylabel('Frequency')
plt.title('Daily Trend Chart')
```

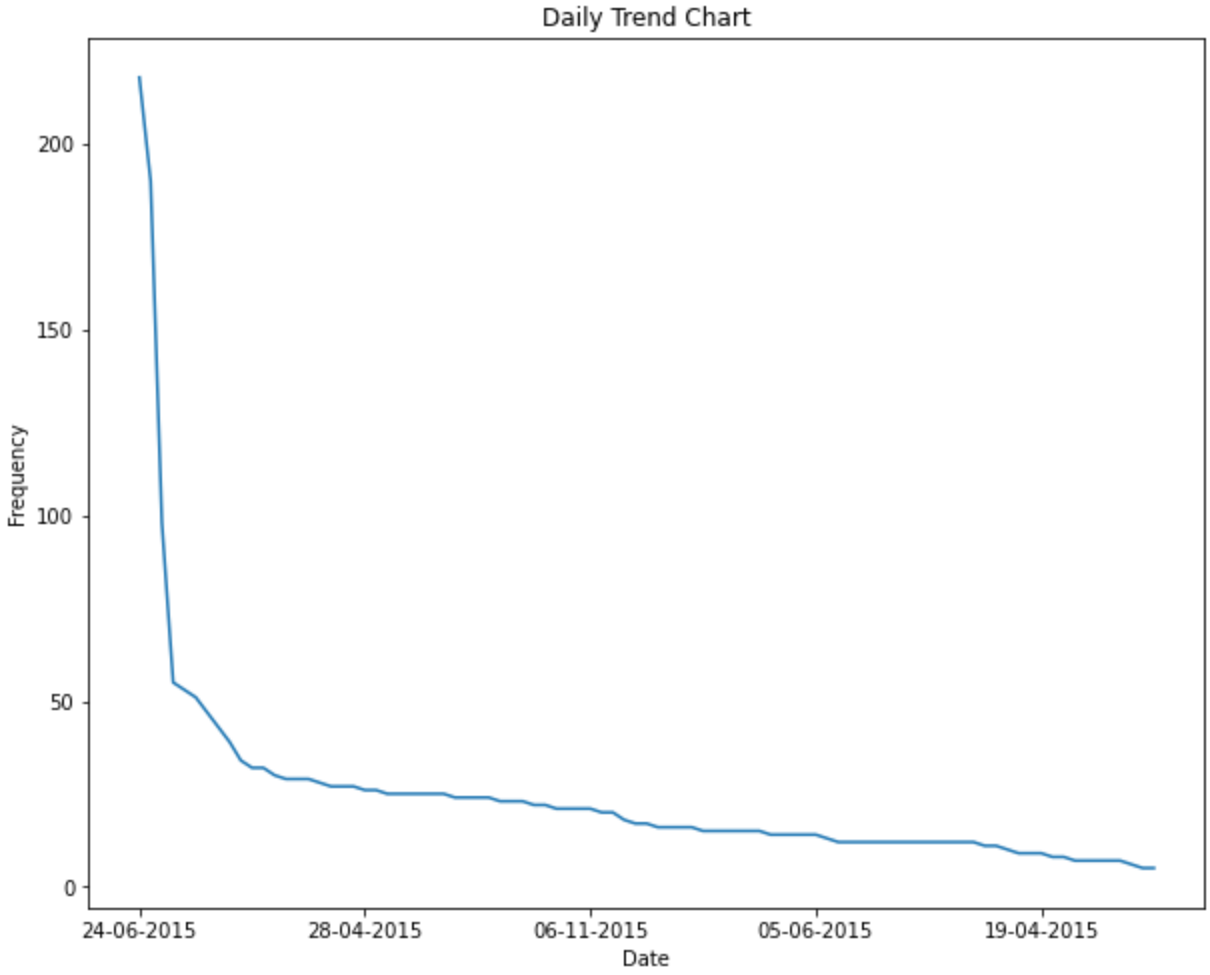
Out[20]: Text(0.5, 1.0, 'Daily Trend Chart')



In [19]:

```
df = df.sort_values(by='Date')
plt.figure(figsize=(10,8))
df['Date'].value_counts().plot()
plt.xlabel('Date')
plt.ylabel('Frequency')
plt.title('Daily Trend Chart')
```

Out[19]: Text(0.5, 1.0, 'Daily Trend Chart')



In [21]:
```
df['Customer Complaint'].value_counts()
```

Out[21]:
```
Comcast                                                          83
Comcast Internet                                                 18
Comcast Data Cap                                                 17
comcast                                                          13
Data Caps                                                        11
                                                                 ..
Comcast Internet, cable, and phone outtages                       1
Unable to renew IP address                                        1
Lack of availability                                              1
Comcast blocking DirecTv signals                                  1
Comcast Billing for Late Payment/Disconnect due to their error    1
Name: Customer Complaint, Length: 1841, dtype: int64
```

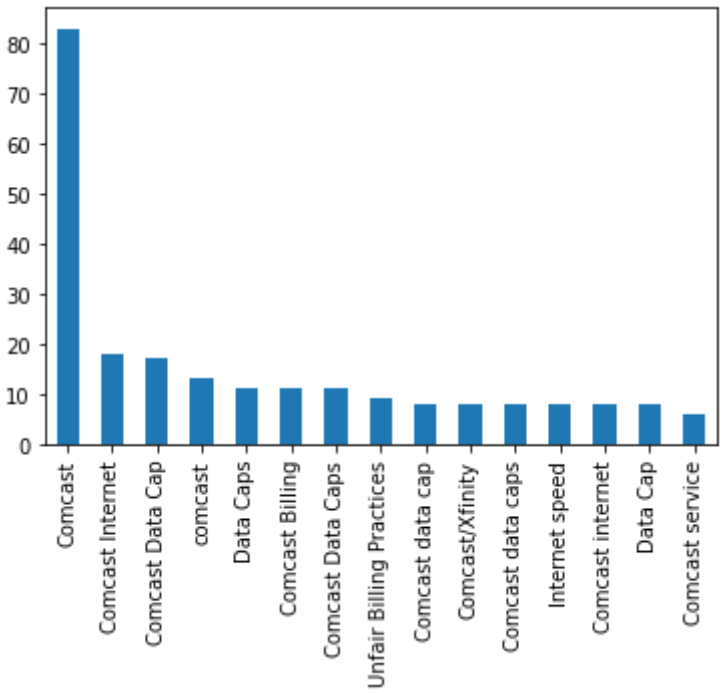
In [22]:
```
df['Customer Complaint'].value_counts()[:15]
```

Out[22]:
```
Comcast                    83
Comcast Internet           18
Comcast Data Cap           17
comcast                    13
Data Caps                  11
Comcast Billing            11
Comcast Data Caps          11
Unfair Billing Practices    9
Comcast data cap            8
Comcast/Xfinity             8
Comcast data caps           8
Internet speed              8
Comcast internet            8
Data Cap                    8
```

```
Comcast service              6
Name: Customer Complaint, dtype: int64
```

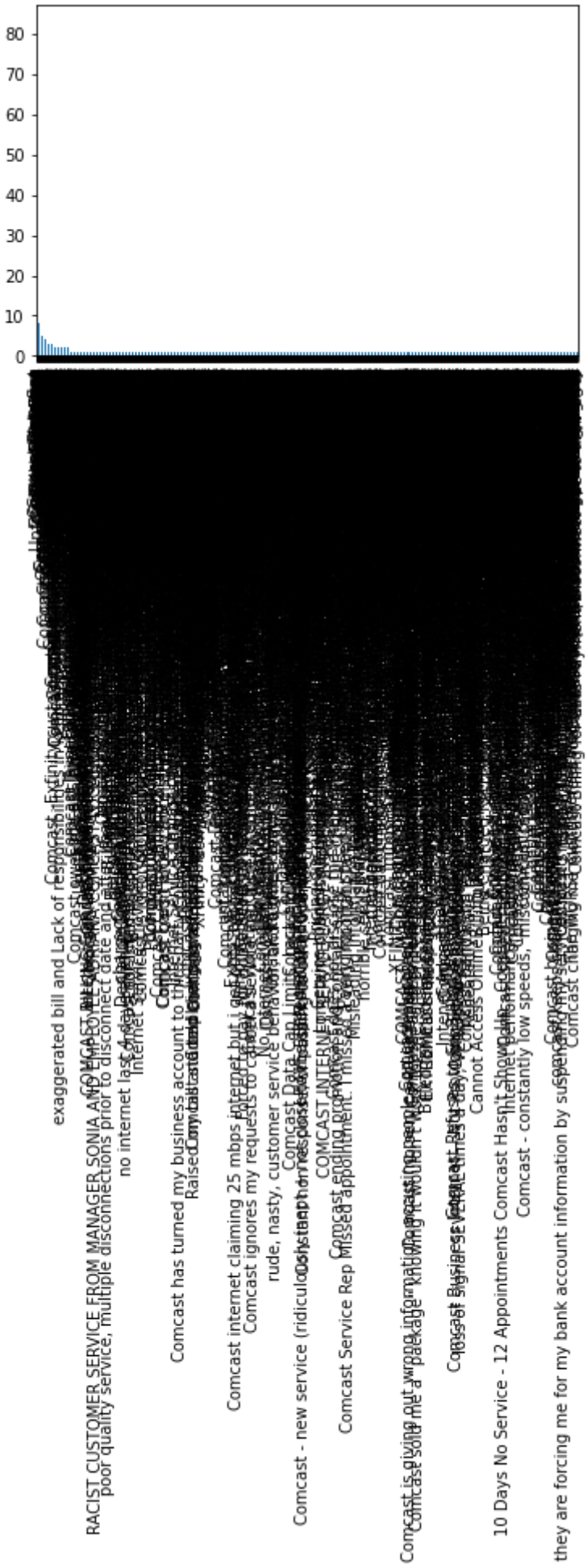
In [26]:
```
df['Customer Complaint'].value_counts()[:15].plot.bar()
```

Out[26]: <AxesSubplot:>



In [27]:
```
df['Customer Complaint'].value_counts().plot.bar()
```

Out[27]: <AxesSubplot:>



In [28]:
```
df['Customer Complaint'].unique()
```

Out[28]:
```
array(['Fraudulent claims reported to collections agency',
       'Comcast refusal of service', 'Comcast Cable', ...,
       'Comcast of East Windsor NJ Complaint',
       'Complaint against Comcast for incredibly bad service',
       'Questionable internet slowdown'], dtype=object)
```

In [31]:
```
internet_issue1 = df[df['Customer Complaint'].str.contains('network')].count()
#print(internet_issue1)
```

In [32]:
```
internet_issue2 = df[df['Customer Complaint'].str.contains('speed')].count()
```

In [33]:
```
internet_issue3 = df[df['Customer Complaint'].str.contains('data')].count()
```

In [34]:
```
internet_issue4 = df[df['Customer Complaint'].str.contains('internet')].count()
```

In [35]:
```
billing_issue1 = df[df['Customer Complaint'].str.contains('billing')].count()
```

In [36]:
```
billing_issue2 = df[df['Customer Complaint'].str.contains('charges')].count()
```

In [37]:
```
billing_issue3 = df[df['Customer Complaint'].str.contains('bill')].count()
```

In [38]:
```
service_issue1 = df[df['Customer Complaint'].str.contains('service')].count()
```

In [39]:
```
service_issue2 = df[df['Customer Complaint'].str.contains('customer')].count()
```

In [40]:
```
total_issue_internet = internet_issue1 + internet_issue2 + internet_issue3 + interne
toprint(total_issue_internet)
```

```
Customer Complaint            374
Date                          374
Received Via                  374
City                          374
State                         374
Zip code                      374
Status                        374
Filing on Behalf of Someone   374
dtype: int64
```

In [41]:
```
total_billing_issues = billing_issue1 + billing_issue2 + billing_issue3
print(total_billing_issues)
```

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

In [42]:
```
total_service_issues = service_issue1 + service_issue2
```

```
print(total_service_issues)
```

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

In [44]:
```
df.shape
```

Out[44]: (2224, 8)

In [45]:
```
other_issues = 2224 - (total_billing_issues + total_service_issues + total_issue_int
```

In [46]:
```
print(other_issues)
```

```
Customer Complaint            1137
Date                          1137
Received Via                  1137
City                          1137
State                         1137
Zip code                      1137
Status                        1137
Filing on Behalf of Someone   1137
dtype: int64
```

In [50]:
```
df['newStatus']= ['Open' if Status=='Open' or Status=='Pending' else 'Closed' for St
```

In [51]:
```
df.head(15)
```

Out[51]:

Date_month_year	Customer Complaint	Date	Received Via	City	State	Zip code	Status	Filing on Behalf of Someone
2015-01-04	Fraudulent claims reported to collections agency	04-01-2015	Customer Care Call	Atlanta	Georgia	30312	Closed	No
2015-01-04	Comcast refusal of service	04-01-2015	Customer Care Call	Wayne	Pennsylvania	19087	Closed	No
2015-01-04	Comcast Cable	04-01-2015	Internet	Franklin	Tennessee	37067	Closed	No
2015-01-04	Data Overages	04-01-2015	Internet	Savannah	Georgia	31406	Closed	No
2015-01-04	Comcast	04-01-2015	Internet	North Huntingdon	Pennsylvania	15642	Closed	No

Date_month_year	Customer Complaint	Date	Received Via	City	State	Zip code	Status	Filing on Behalf of Someone
2015-01-04	Comcast harassment	04-01-2015	Customer Care Call	Schaumburg	Illinois	60193	Closed	No
2015-01-04	Comcast is ignoring me now. They are making NO...	04-01-2015	Internet	Golden	Colorado	80401	Closed	No
2015-01-04	Unable to get in touch with anyone that has th...	04-01-2015	Customer Care Call	Huntsville	Alabama	35801	Closed	No
2015-01-04	Comcast Lied About Pricing And Installation	04-01-2015	Customer Care Call	Newark	California	94560	Closed	No
2015-01-04	overcharged	04-01-2015	Internet	Newman	California	95360	Closed	No
2015-01-04	No Internet service	04-01-2015	Internet	Philadelphia	Pennsylvania	19128	Closed	No
2015-01-04	Comcast Customer Service; Theft; Inconsistency	04-01-2015	Customer Care Call	Philadelphia	Pennsylvania	19121	Closed	No
2015-01-04	Incorrect Billing	04-01-2015	Customer Care Call	Boynton Beach	Florida	33426	Closed	No
2015-01-04	comcast cable	04-01-2015	Customer Care Call	Lockport	Illinois	60441	Closed	No
2015-01-04	Comcast speeds as low as 12 MB/s, paying for 1...	04-01-2015	Customer Care Call	Washington	Pennsylvania	15301	Closed	No

In [52]:
```
df.sample(15)
```

Out[52]:

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

Date_month_year	Customer Complaint	Date	Received Via	City	State	Zip code	Status	Filing on Behalf of Someone
2015-05-06	Rates	06-05-2015	Internet	Bala Cynwyd	Pennsylvania	19004	Closed	No
2015-06-26	Incorrect Billing and Service from Comcast	26-06-2015	Internet	Bloomington	Indiana	47401	Solved	No
2015-08-06	Speed	06-08-2015	Customer Care Call	Alexandria	Virginia	22304	Solved	No
2015-06-27	Comcast monopoly bundling practices	27-06-2015	Internet	Alexandria	Virginia	22304	Open	No
2015-04-25	Comcast blocking HBO Go on Playstation systems	25-04-2015	Customer Care Call	Redmond	Washington	98052	Open	No
2015-06-13	Data Cap	13-06-2015	Customer Care Call	Cumming	Georgia	30041	Solved	No
2015-08-04	new fcc internet rules slowing down system	04-08-2015	Internet	Middletown	Connecticut	6457	Closed	No
2015-01-05	Comcast is ripping me off, and I can't be quie...	05-01-2015	Internet	Chicago	Illinois	60610	Solved	No
2015-04-18	Comcast customer service and billing complaint	18-04-2015	Customer Care Call	Enola	Pennsylvania	17025	Closed	No
2015-01-06	Comcast Internet Service Bad Quality	06-01-2015	Internet	Cupertino	California	95014	Closed	No
2015-06-29	Comcast service	29-06-2015	Customer Care Call	Detroit	Michigan	48202	Open	No

Date_month_year	Customer Complaint	Date	Received Via	City	State	Zip code	Status	Filing on Behalf of Someone
2015-06-24	Re: Unfair Billing Practices Complaints agains...	24-06-2015	Internet	Vernon Hills	Illinois	60061	Solved	No
2015-06-16	Comcast does not disclose the Internet speed I...	16-06-2015	Customer Care Call	Baltimore	Maryland	21218	Open	No
2015-04-24	comcast unwilling to resolve data usage issue	24-04-2015	Customer Care Call	Cartersville	Georgia	30120	Closed	No
2015-12-04	Comcast Agreement	04-12-2015	Customer Care Call	Boulder	Colorado	80303	Closed	No

In [55]:
```
df.groupby(['State']).size().sort_values(ascending=False)[:10]
```

Out[55]:
```
State
Georgia         288
Florida         240
California      220
Illinois        164
Tennessee       143
Pennsylvania    130
Michigan        115
Washington       98
Colorado         80
Maryland         78
dtype: int64
```

In [56]:
```
state_complain = df.groupby(['State','newStatus']).size().unstack()
```

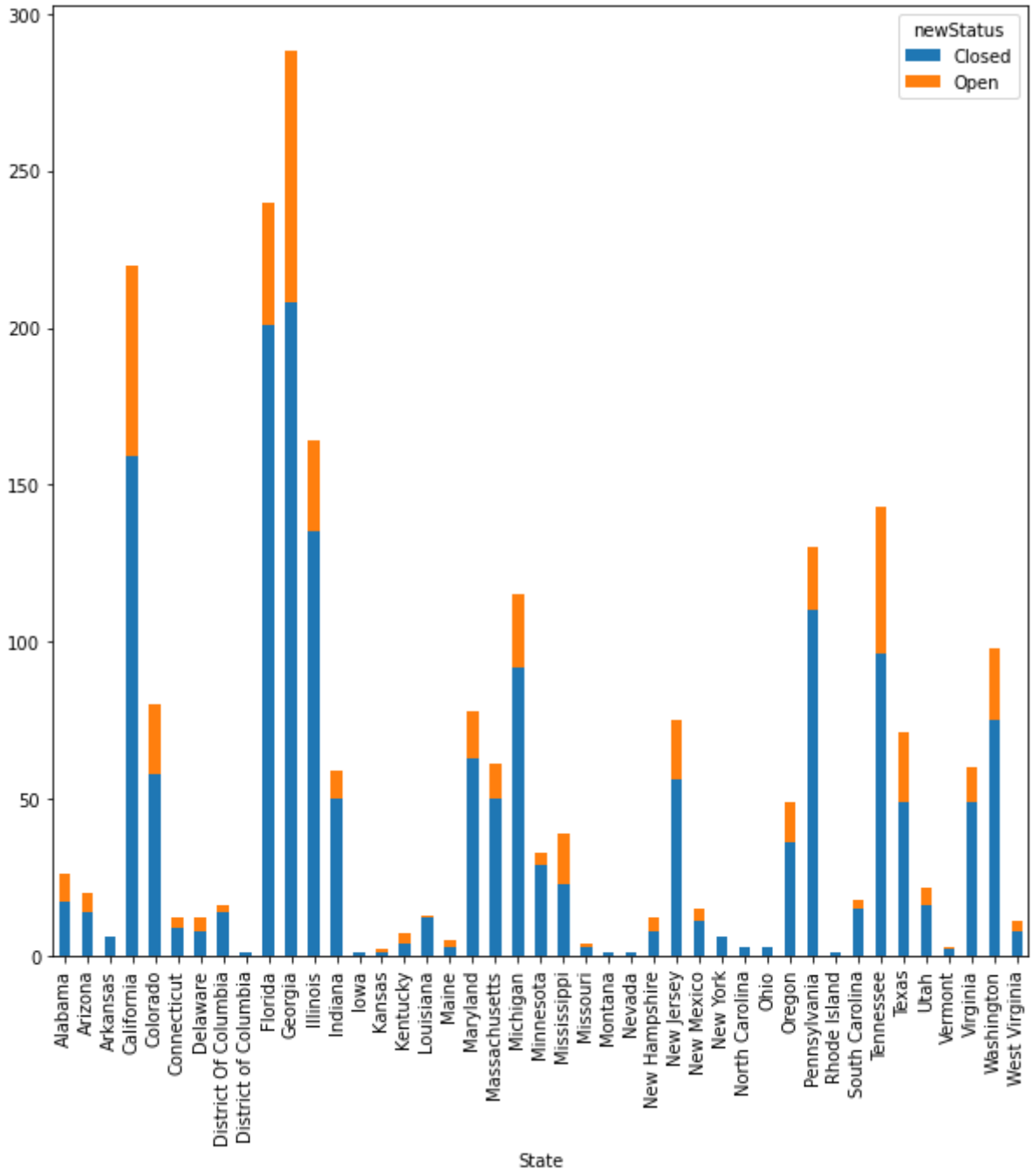
In [57]:
```
print(state_complain)
```

```
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 [58]:
```
state_complain.plot.bar(figsize=(10,10),stacked=True)
```

Out[58]:
```
<AxesSubplot:xlabel='State'>
```



In [59]:
```
df.newStatus.value_counts()
```

Out[59]:
```
Closed    1707
Open       517
Name: newStatus, dtype: int64
```

In [65]:
```
unresolved_data = df.groupby(['State','newStatus']).size().unstack().fillna(0).sort_
print(unresolved_data)
```

```
newStatus             Closed  Open
State
Georgia                208.0  80.0
California             159.0  61.0
Tennessee               96.0  47.0
Florida                201.0  39.0
Illinois               135.0  29.0
Washington              75.0  23.0
Michigan                92.0  23.0
Colorado                58.0  22.0
Texas                   49.0  22.0
```

```
Pennsylvania           110.0  20.0
New Jersey              56.0  19.0
Mississippi             23.0  16.0
Maryland                63.0  15.0
Oregon                  36.0  13.0
Virginia                49.0  11.0
Massachusetts           50.0  11.0
Alabama                 17.0   9.0
Indiana                 50.0   9.0
Utah                    16.0   6.0
Arizona                 14.0   6.0
New Hampshire            8.0   4.0
New Mexico              11.0   4.0
Minnesota               29.0   4.0
Delaware                 8.0   4.0
West Virginia            8.0   3.0
Connecticut              9.0   3.0
Kentucky                 4.0   3.0
South Carolina          15.0   3.0
Maine                    3.0   2.0
District Of Columbia    14.0   2.0
Kansas                   1.0   1.0
Vermont                  2.0   1.0
Missouri                 3.0   1.0
Louisiana               12.0   1.0
Montana                  1.0   0.0
Rhode Island             1.0   0.0
Ohio                     3.0   0.0
District of Columbia     1.0   0.0
North Carolina           3.0   0.0
New York                 6.0   0.0
Nevada                   1.0   0.0
Arkansas                 6.0   0.0
Iowa                     1.0   0.0
```

In [66]:
```
unresolved_data['unresolved_cmp_prct']=unresolved_data['Open']/unresolved_data['Open
```

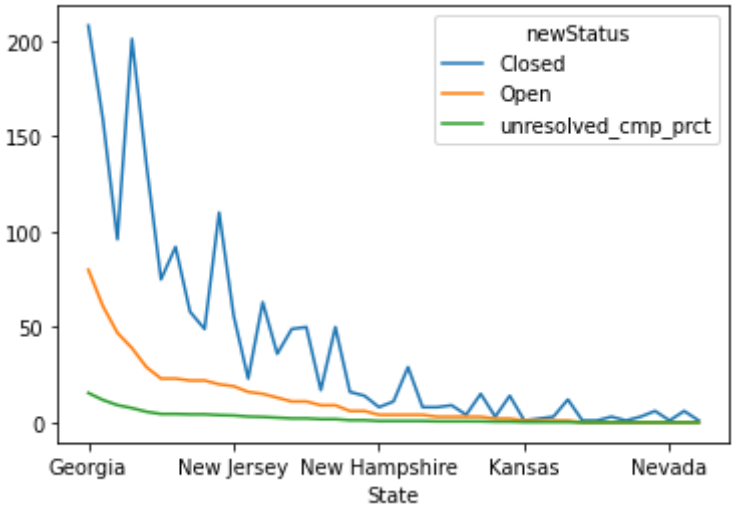
In [67]:
```
print(unresolved_data)
```

```
newStatus             Closed  Open  unresolved_cmp_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
Washington              75.0  23.0             4.448743
Michigan                92.0  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
Virginia                49.0  11.0             2.127660
Massachusetts           50.0  11.0             2.127660
Alabama                 17.0   9.0             1.740812
Indiana                 50.0   9.0             1.740812
Utah                    16.0   6.0             1.160542
Arizona                 14.0   6.0             1.160542
New Hampshire            8.0   4.0             0.773694
New Mexico              11.0   4.0             0.773694
Minnesota               29.0   4.0             0.773694
Delaware                 8.0   4.0             0.773694
West Virginia            8.0   3.0             0.580271
Connecticut              9.0   3.0             0.580271
```

```
Kentucky                4.0   3.0             0.580271
South Carolina         15.0   3.0             0.580271
Maine                   3.0   2.0             0.386847
District Of Columbia   14.0   2.0             0.386847
Kansas                  1.0   1.0             0.193424
Vermont                 2.0   1.0             0.193424
Missouri                3.0   1.0             0.193424
Louisiana              12.0   1.0             0.193424
Montana                 1.0   0.0             0.000000
Rhode Island            1.0   0.0             0.000000
Ohio                    3.0   0.0             0.000000
District of Columbia    1.0   0.0             0.000000
North Carolina          3.0   0.0             0.000000
New York                6.0   0.0             0.000000
Nevada                  1.0   0.0             0.000000
Arkansas                6.0   0.0             0.000000
Iowa                    1.0   0.0             0.000000
```

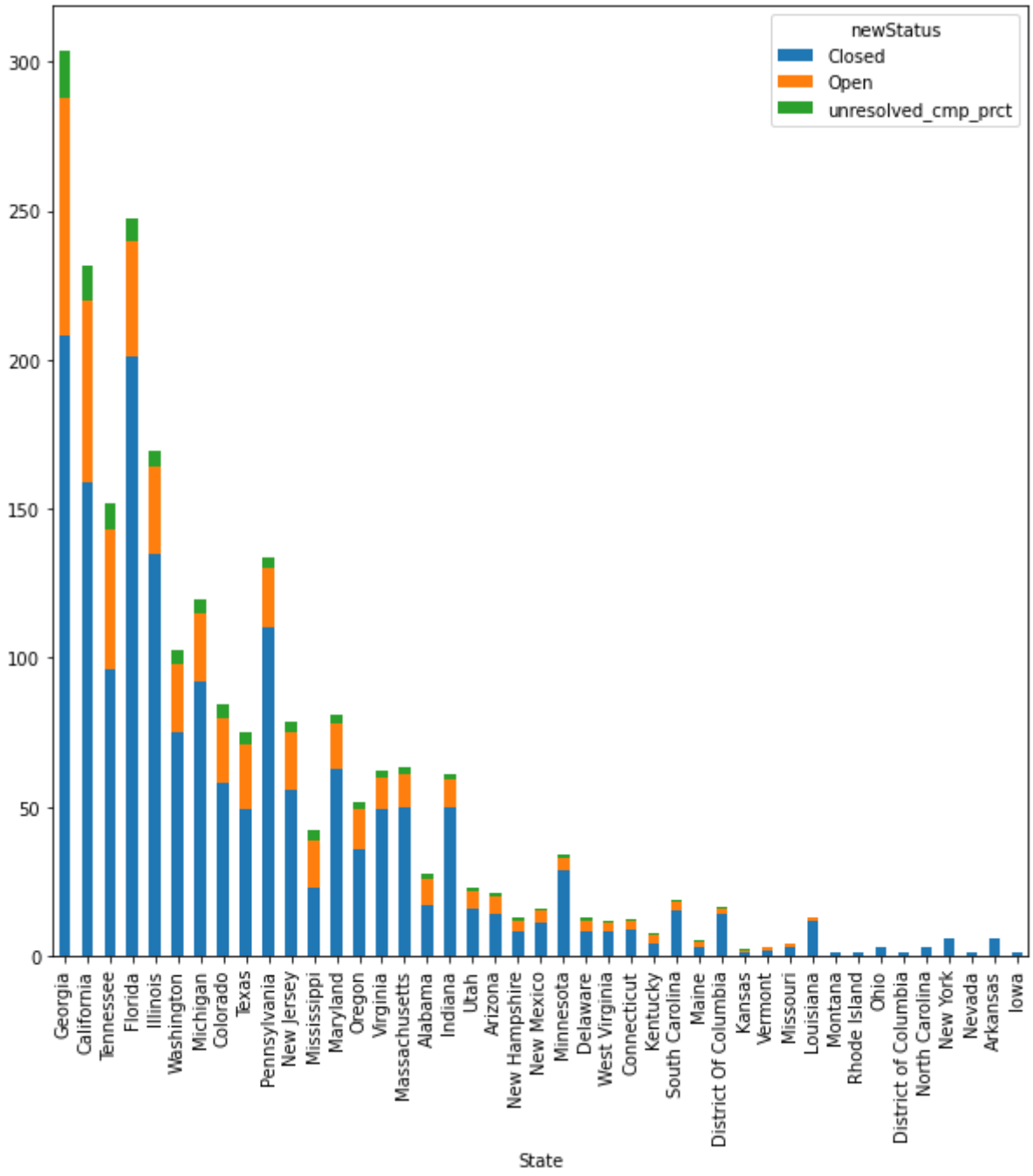
In [68]:
```
unresolved_data.plot()
```

Out[68]: <AxesSubplot:xlabel='State'>



In [69]:
```
unresolved_data.plot.bar(figsize=(10,10),stacked=True)
```

Out[69]: <AxesSubplot:xlabel='State'>



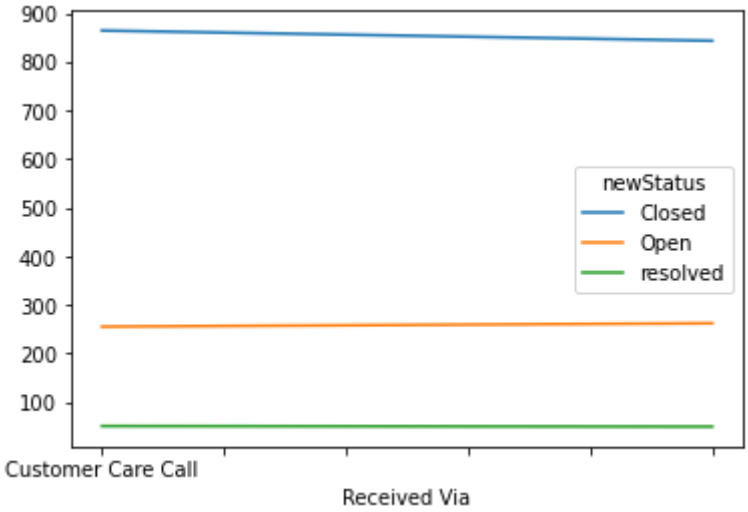
In [70]:
```
resolved_data = df.groupby(['Received Via','newStatus']).size().unstack().fillna(0)
```

In [71]:
```
resolved_data['resolved'] = resolved_data['Closed']/resolved_data['Closed'].sum()*10
print(resolved_data)
```

```
newStatus           Closed  Open   resolved
Received Via
Customer Care Call     864   255  50.615114
Internet               843   262  49.384886
```

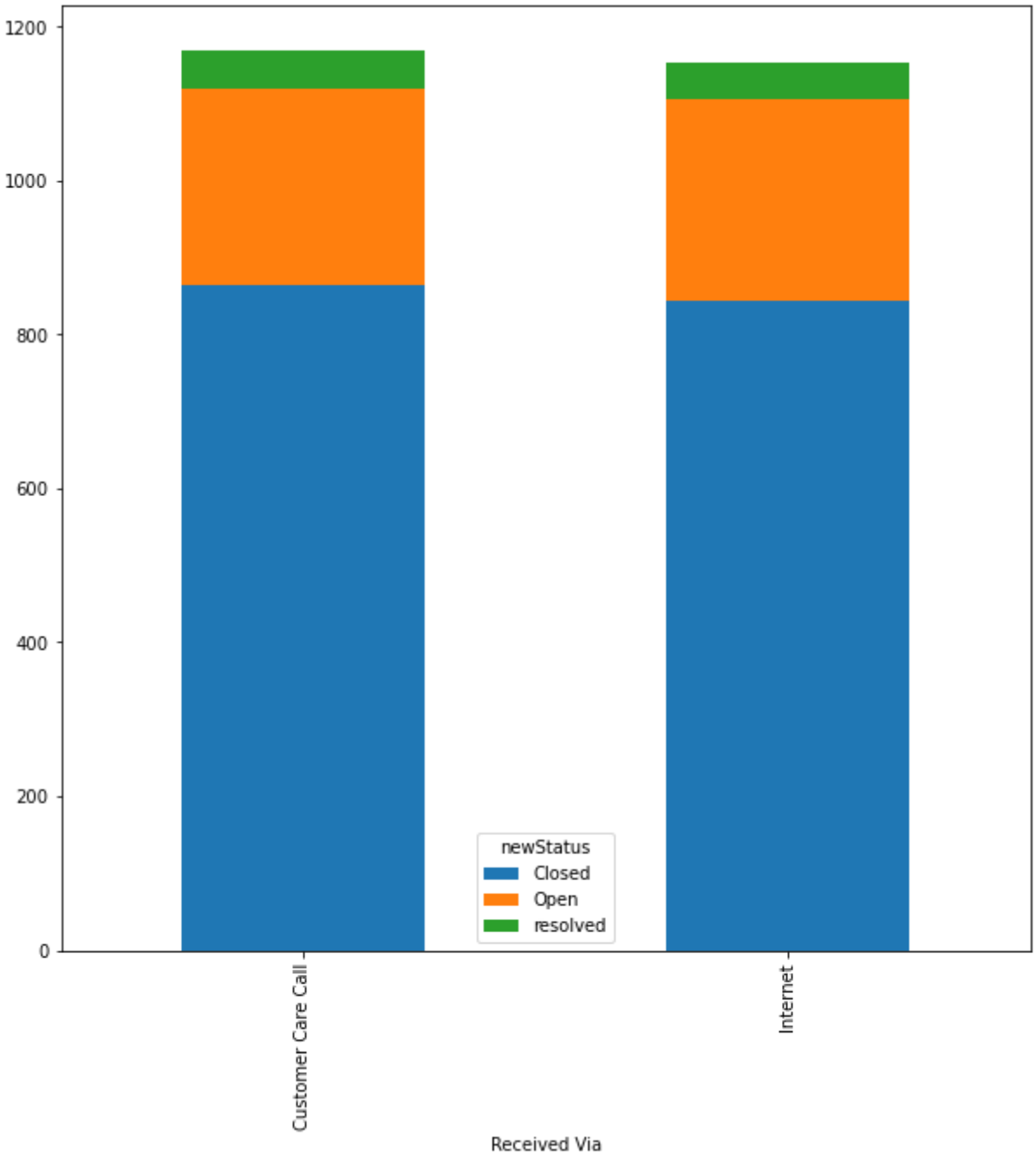
In [72]:
```
resolved_data.plot()
```

Out[72]:
```
<AxesSubplot:xlabel='Received Via'>
```



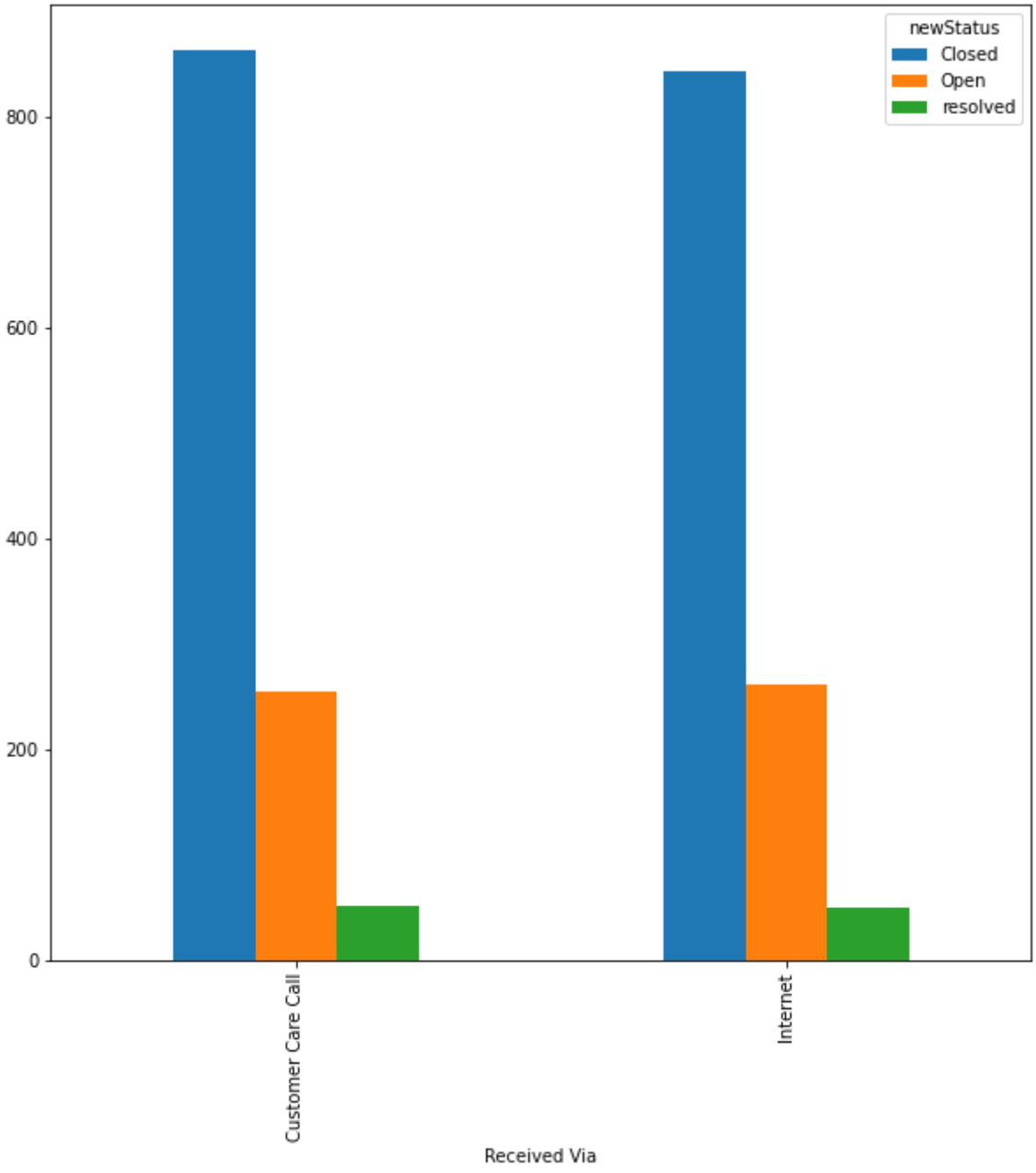
In [73]:
```
resolved_data.plot.bar(figsize=(10,10),stacked=True)
```

Out[73]: <AxesSubplot:xlabel='Received Via'>



In [74]:
```
resolved_data.plot(kind='bar',figsize=(10,10))
```

Out[74]: <AxesSubplot:xlabel='Received Via'>



In [ ]:
```

```