

COMCAST

December 12, 2022

```
[3]: import pandas as pd
import numpy as np
import datetime as dt
import seaborn as sns
import matplotlib.pyplot as plt
from matplotlib import style
from matplotlib import style
%matplotlib inline
```

```
[4]: #UPLOADING THE DATA SET
df = pd.read_csv("Comcast_telecom_complaints_data.csv", index_col=0,
↳ parse_dates=[2,3])
```

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

```
[5]:
```

	Customer Complaint	Date \
Ticket #		
250635	Comcast Cable Internet Speeds	2015-04-22
223441	Payment disappear - service got disconnected	2015-04-08
242732	Speed and Service	2015-04-18
277946	Comcast Imposed a New Usage Cap of 300GB that ...	2015-05-07
307175	Comcast not working and no service to boot	2015-05-26

	Date_month_year	Time	Received Via	City	State \
Ticket #					
250635	2015-04-22	3:53:50 PM	Customer Care Call	Abingdon	Maryland
223441	2015-08-04	10:22:56 AM	Internet	Acworth	Georgia
242732	2015-04-18	9:55:47 AM	Internet	Acworth	Georgia
277946	2015-07-05	11:59:35 AM	Internet	Acworth	Georgia
307175	2015-05-26	1:25:26 PM	Internet	Acworth	Georgia

	Zip code	Status	Filing on Behalf of Someone
Ticket #			
250635	21009	Closed	No
223441	30102	Closed	No
242732	30101	Closed	Yes
277946	30101	Open	Yes

307175	30101	Solved	No
--------	-------	--------	----

```
[6]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 2224 entries, 250635 to 363614
Data columns (total 10 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Customer Complaint                    2224 non-null   object
1   Date                                  2224 non-null   datetime64[ns]
2   Date_month_year                      2224 non-null   datetime64[ns]
3   Time                                  2224 non-null   object
4   Received Via                         2224 non-null   object
5   City                                  2224 non-null   object
6   State                                2224 non-null   object
7   Zip code                             2224 non-null   int64
8   Status                               2224 non-null   object
9   Filing on Behalf of Someone          2224 non-null   object
dtypes: datetime64[ns](2), int64(1), object(7)
memory usage: 191.1+ KB
```

```
[7]: df.isna().sum()
```

```
[7]: Customer Complaint
      Date
      Date_month_year
      Time
      Received Via
      City
      State
      Zip code
      Status
      Filing on Behalf of Someone
      dtype: int64
```

```
[8]: df.head()
```

```
[8]:
```

	Customer Complaint	Date \
Ticket #		
250635	Comcast Cable Internet Speeds	2015-04-22
223441	Payment disappear - service got disconnected	2015-04-08
242732	Speed and Service	2015-04-18
277946	Comcast Imposed a New Usage Cap of 300GB that ...	2015-05-07
307175	Comcast not working and no service to boot	2015-05-26

Date_month_year	Time	Received Via	City	State \
-----------------	------	--------------	------	---------

Ticket #	Date	Time	Customer Complaint	City	State
250635	2015-04-22	3:53:50 PM	Customer Care Call	Abingdon	Maryland
223441	2015-08-04	10:22:56 AM	Internet	Acworth	Georgia
242732	2015-04-18	9:55:47 AM	Internet	Acworth	Georgia
277946	2015-07-05	11:59:35 AM	Internet	Acworth	Georgia
307175	2015-05-26	1:25:26 PM	Internet	Acworth	Georgia

Ticket #	Zip code	Status	Filing on Behalf of Someone
250635	21009	Closed	No
223441	30102	Closed	No
242732	30101	Closed	Yes
277946	30101	Open	Yes
307175	30101	Solved	No

```
[9]: #trend chart for complaint at monthly levels
df['month'] = pd.DatetimeIndex(df['Date_month_year']).month
m = df['Customer Complaint'].groupby(df['month']).count().reset_index()
m
```

```
[9]:
```

	month	Customer Complaint
0	1	55
1	2	59
2	3	45
3	4	375
4	5	317
5	6	1046
6	7	49
7	8	67
8	9	55
9	10	53
10	11	38
11	12	65

```
[10]: df.columns
```

```
[10]: Index(['Customer Complaint', 'Date', 'Date_month_year', 'Time', 'Received Via',
        'City', 'State', 'Zip code', 'Status', 'Filing on Behalf of Someone',
        'month'],
        dtype='object')
```

```
[11]: df.head()
```

```
[11]:
```

Ticket #	Customer Complaint	Date
250635	Comcast Cable Internet Speeds	2015-04-22
223441	Payment disappear - service got disconnected	2015-04-08

242732	Speed and Service	2015-04-18
277946	Comcast Imposed a New Usage Cap of 300GB that ...	2015-05-07
307175	Comcast not working and no service to boot	2015-05-26

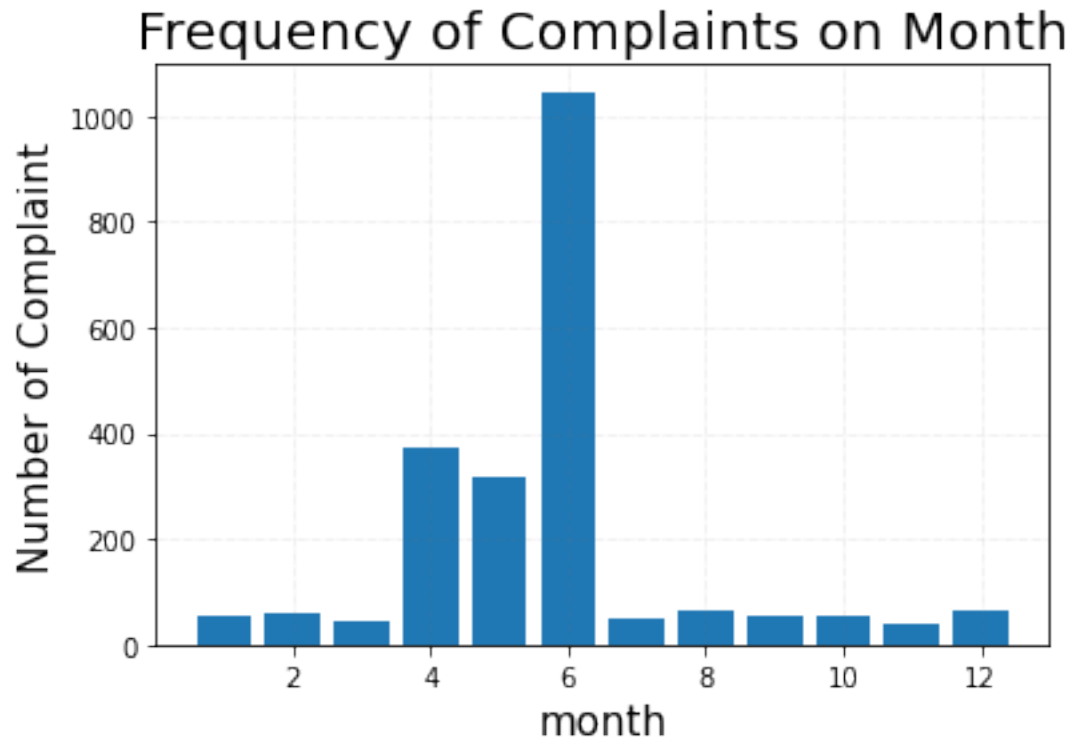
	Date_month_year	Time	Received Via	City	State \
Ticket #					
250635	2015-04-22	3:53:50 PM	Customer Care Call	Abingdon	Maryland
223441	2015-08-04	10:22:56 AM	Internet	Acworth	Georgia
242732	2015-04-18	9:55:47 AM	Internet	Acworth	Georgia
277946	2015-07-05	11:59:35 AM	Internet	Acworth	Georgia
307175	2015-05-26	1:25:26 PM	Internet	Acworth	Georgia

	Zip code	Status	Filing on Behalf of Someone	month
Ticket #				
250635	21009	Closed	No	4
223441	30102	Closed	No	8
242732	30101	Closed	Yes	4
277946	30101	Open	Yes	7
307175	30101	Solved	No	5

```
[12]: x=m['month']
y=m['Customer Complaint']
plt.bar(x,y)
plt.xlabel("month",fontsize=15)
plt.ylabel("Number of Complaint",fontsize=15)
plt.grid(b = True, color = 'grey',
        linestyle = '-.', linewidth = 0.5,
        alpha = 0.2)
plt.title("Frequency of Complaints on Month",fontsize=20)
```

```
/usr/local/lib/python3.7/site-packages/ipykernel_launcher.py:8:
MatplotlibDeprecationWarning: The 'b' parameter of grid() has been renamed
'visible' since Matplotlib 3.5; support for the old name will be dropped two
minor releases later.
```

```
[12]: Text(0.5, 1.0, 'Frequency of Complaints on Month')
```



```
[13]: #trend chart for companies at daily levels
import datetime
from dateutil.parser import parse
import pytz

df['Date of the Month'] = df['Date_month_year'].apply(lambda d: d.day)
d = df['Customer Complaint'].groupby(df['Date of the Month']).count().
    ↪reset_index()
d
```

```
[13]:
```

	Date of the Month	Customer Complaint
0	4	206
1	5	131
2	6	272
3	13	68
4	14	54
5	15	58
6	16	65
7	17	60
8	18	69
9	19	50
10	20	51
11	21	41

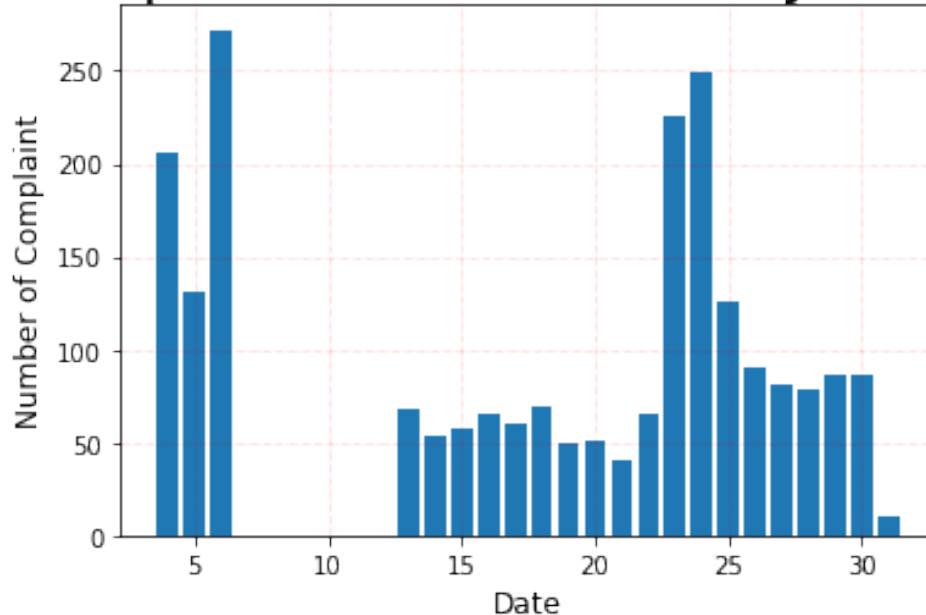
12	22	66
13	23	225
14	24	249
15	25	126
16	26	90
17	27	81
18	28	79
19	29	87
20	30	86
21	31	10

```
[14]: x=d['Date of the Month']
      y=d['Customer Complaint']
      plt.bar(x,y)
      plt.xlabel("Date",fontsize=12)
      plt.ylabel("Number of Complaint",fontsize=12)
      plt.grid(b = True, color = 'red',
                linestyle = '-.', linewidth = 0.5,
                alpha = 0.2)
      plt.title(" Complaints on Date of any month ",fontsize=25)
```

```
/usr/local/lib/python3.7/site-packages/ipykernel_launcher.py:8:
MatplotlibDeprecationWarning: The 'b' parameter of grid() has been renamed
'visible' since Matplotlib 3.5; support for the old name will be dropped two
minor releases later.
```

```
[14]: Text(0.5, 1.0, ' Complaints on Date of any month ')
```

Complaints on Date of any month



```
[15]: #frequency of complaint types
```

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

```
[16]: array(['Comcast Cable Internet Speeds',
          'Payment disappear - service got disconnected',
          'Speed and Service', ..., 'complaint about comcast',
          'Extremely unsatisfied Comcast customer',
          'Comcast, Ypsilanti MI Internet Speed'], dtype=object)
```

```
[17]: df['Customer Complaint'].nunique()
```

```
[17]: 1841
```

```
[18]: x=df['Customer Complaint'].value_counts()
```

```
[19]: d1=pd.DataFrame(x)
      d1
```

```
[19]:
```

	Customer Complaint
Comcast	83
Comcast Internet	18
Comcast Data Cap	17
comcast	13

Comcast Billing	11
...	...
Terrible Service	1
Comcast blocking HBOGO on Playstation 4 Consoles	1
Suspected Throttling	1
Internet connection both upload and download sp...	1
Unethical Charges	1

[1841 rows x 1 columns]

```
[20]: plt.rcParams["figure.figsize"] = [15, 5]
plt.rcParams["figure.autolayout"] = True
fig, ax = plt.subplots()
df['Customer Complaint'].value_counts().plot(ax=ax, kind='bar',
↳ xlabel='Complaint Type', ylabel='frequency')
```

```
[20]: <AxesSubplot:xlabel='Complaint Type', ylabel='frequency'>
```

```
/usr/local/lib/python3.7/site-packages/IPython/core/pylabtools.py:132:
UserWarning: Tight layout not applied. The bottom and top margins cannot be made
large enough to accommodate all axes decorations.
fig.canvas.print_figure(bytes_io, **kw)
```



```

307175      [Comcast not working and no service to boot]
...
213550      [Service Availability]
318775      [Comcast Monthly Billing for Returned Modem]
331188      [complaint about comcast]
360489      [Extremely unsatisfied Comcast customer]
363614      [Comcast, Ypsilanti MI Internet Speed]
Name: Customer Complaint, Length: 2224, dtype: object

```

```

[23]: df["Complaint Category"] = np.where(df["Customer Complaint"].str.
    ↳contains('Internet|Speed|Cable|connection|DSL|modem|ip|ISP',case=False,regex=True),
    ↳'Internet',
        np.where(df["Customer Complaint"].str.
    ↳contains('Payment|Disappear|Disconnected',case=False,regex=True), 'Paym',
        np.where(df["Customer Complaint"].str.
    ↳contains('Comcast|Imposed|New|Usage|Cap|of',case=False,regex=True), 'Usage',
        np.where(df["Customer Complaint"].str.
    ↳contains('complaint|breach|bully|non_
    ↳response|False|invalid',case=False,regex=True), 'Greviances',
        np.where(df["Customer Complaint"].str.
    ↳contains('email|mail',case=False,regex=True), 'Email',
        np.where(df["Customer Complaint"].str.
    ↳contains('policies|competition|Forgery|Fraud|Deceptive|deception|scam|business',case=False,
    ↳'Marketing',
        'Other')))))))

```

```

[24]: df["Complaint Category"]

```

```

[24]: Ticket #
250635      Internet
223441          Paym
242732      Internet
277946          Usage
307175          Usage
...
213550          Other
318775      Internet
331188          Usage
360489          Usage
363614      Internet
Name: Complaint Category, Length: 2224, dtype: object

```

```

[25]: d3= pd.DataFrame(df["Customer Complaint"].groupby(df["Complaint Category"]).
    ↳count().sort_values(ascending=True)).reset_index()
d3      #usage complaint types are maximum

```

```
[25]: Complaint Category Customer Complaint
0      Email                12
1      Paym                 14
2      Greviances           27
3      Marketing            32
4      Other                407
5      Internet             721
6      Usage               1011
```

```
[26]: #new categorcal variable
```

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

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

```
[28]: df['Stat']=np.where(df['Status'].str.
    ↳contains('Open|Pending',case=False,regex=True), 'Open',
        np.where(df['Status'].str.
    ↳contains('Solved|Closed',case=False,regex=True), 'Closed' ,
        'Other'))
```

```
[29]: df['Stat']
```

```
[29]: Ticket #
250635    Closed
223441    Closed
242732    Closed
277946     Open
307175    Closed
...
213550    Closed
318775    Closed
331188    Closed
360489    Closed
363614     Open
Name: Stat, Length: 2224, dtype: object
```

```
[30]: df.head()
```

```
[30]: Customer Complaint      Date \
Ticket #
250635      Comcast Cable Internet Speeds 2015-04-22
223441      Payment disappear - service got disconnected 2015-04-08
242732      Speed and Service 2015-04-18
277946      Comcast Imposed a New Usage Cap of 300GB that ... 2015-05-07
307175      Comcast not working and no service to boot 2015-05-26
```

	Date_month_year	Time	Received Via	City	State \
Ticket #					
250635	2015-04-22	3:53:50 PM	Customer Care Call	Abingdon	Maryland
223441	2015-08-04	10:22:56 AM	Internet	Acworth	Georgia
242732	2015-04-18	9:55:47 AM	Internet	Acworth	Georgia
277946	2015-07-05	11:59:35 AM	Internet	Acworth	Georgia
307175	2015-05-26	1:25:26 PM	Internet	Acworth	Georgia

	Zip code	Status	Filing on Behalf of Someone	month \
Ticket #				
250635	21009	Closed	No	4
223441	30102	Closed	No	8
242732	30101	Closed	Yes	4
277946	30101	Open	Yes	7
307175	30101	Solved	No	5

	Date of the Month	Complaint Category	Stat
Ticket #			
250635	22	Internet	Closed
223441	4	Paym	Closed
242732	18	Internet	Closed
277946	5	Usage	Open
307175	26	Usage	Closed

```
[31]: #state wise status
```

```
[32]: from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
df["Stat_le"]=le.fit_transform(df["Stat"])
```

```
[33]: df.head()
```

```
[33]:
```

	Customer Complaint	Date \
Ticket #		
250635	Comcast Cable Internet Speeds	2015-04-22
223441	Payment disappear - service got disconnected	2015-04-08
242732	Speed and Service	2015-04-18
277946	Comcast Imposed a New Usage Cap of 300GB that ...	2015-05-07
307175	Comcast not working and no service to boot	2015-05-26

	Date_month_year	Time	Received Via	City	State \
Ticket #					
250635	2015-04-22	3:53:50 PM	Customer Care Call	Abingdon	Maryland
223441	2015-08-04	10:22:56 AM	Internet	Acworth	Georgia
242732	2015-04-18	9:55:47 AM	Internet	Acworth	Georgia
277946	2015-07-05	11:59:35 AM	Internet	Acworth	Georgia
307175	2015-05-26	1:25:26 PM	Internet	Acworth	Georgia

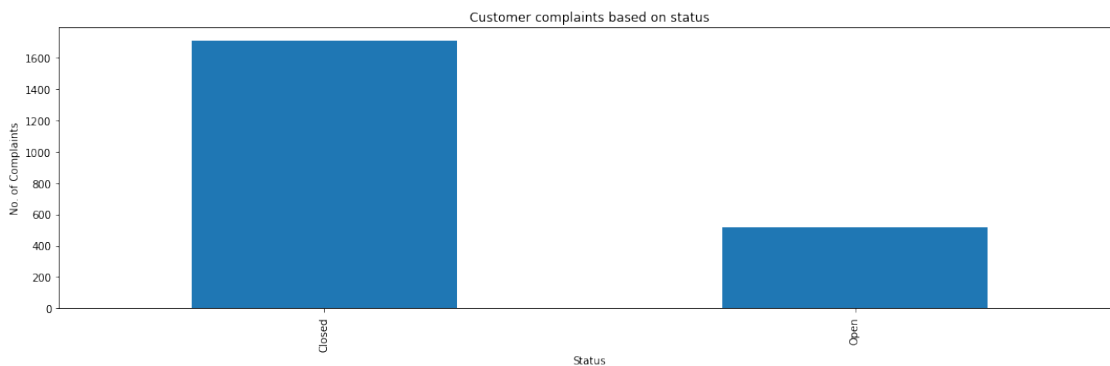
Ticket #	Zip code	Status	Filing on Behalf of Someone	month \
250635	21009	Closed	No	4
223441	30102	Closed	No	8
242732	30101	Closed	Yes	4
277946	30101	Open	Yes	7
307175	30101	Solved	No	5

Ticket #	Date of the Month	Complaint Category	Stat	Stat_le
250635	22	Internet	Closed	0
223441	4	Paym	Closed	0
242732	18	Internet	Closed	0
277946	5	Usage	Open	1
307175	26	Usage	Closed	0

```
[34]: n=df['Stat'].value_counts()
      d4=pd.DataFrame(n)
      d4
```

```
[34]:      Stat
      Closed  1707
      Open    517
```

```
[35]: n.plot(kind='bar')
      plt.xlabel("Status")
      plt.ylabel("No. of Complaints")
      plt.title("Customer complaints based on status")
      plt.show()
```



```
[36]: df.columns
```

```
[36]: Index(['Customer Complaint', 'Date', 'Date_month_year', 'Time', 'Received Via',  
          'City', 'State', 'Zip code', 'Status', 'Filing on Behalf of Someone',  
          'month', 'Date of the Month', 'Complaint Category', 'Stat', 'Stat_le'],  
          dtype='object')
```

```
[37]: df.columns
```

```
[37]: Index(['Customer Complaint', 'Date', 'Date_month_year', 'Time', 'Received Via',  
          'City', 'State', 'Zip code', 'Status', 'Filing on Behalf of Someone',  
          'month', 'Date of the Month', 'Complaint Category', 'Stat', 'Stat_le'],  
          dtype='object')
```

```
[38]: y=df['Stat'].groupby(df['State']).count().sort_values(ascending=False)
```

```
[39]: Dst=pd.DataFrame(y)  
Dst      #Georgia has maximum complaints
```

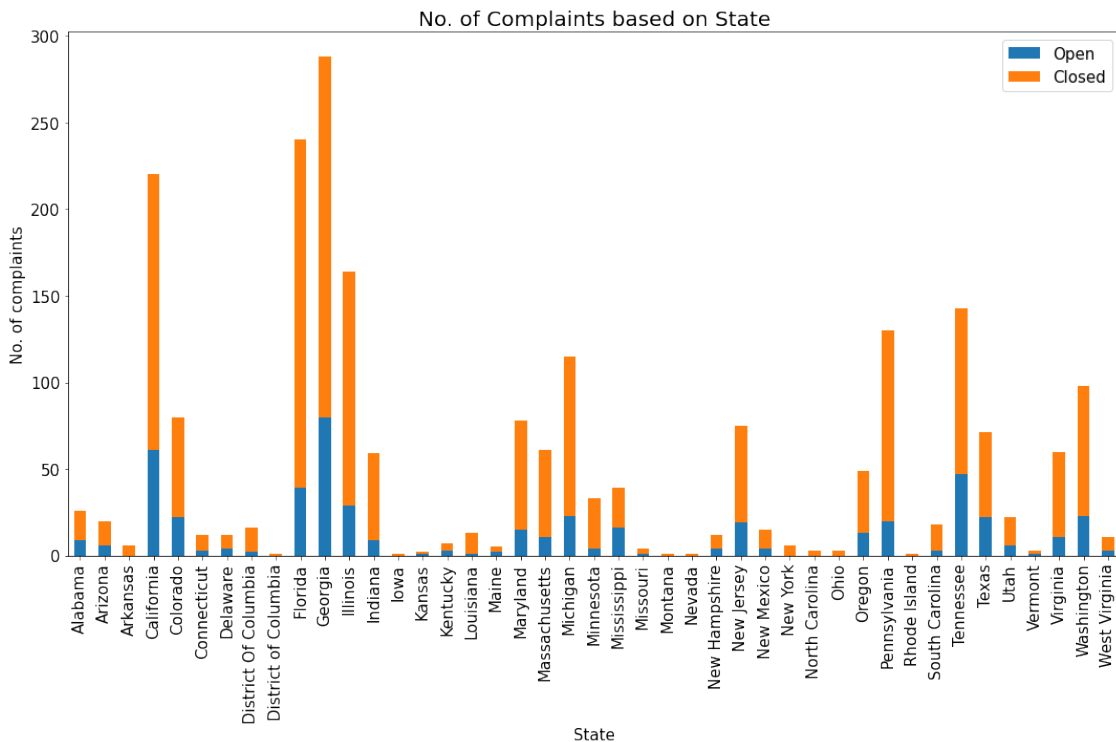
```
[39]:
```

State	Stat
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
New Hampshire	12
Connecticut	12
Delaware	12

West Virginia	11
Kentucky	7
New York	6
Arkansas	6
Maine	5
Missouri	4
North Carolina	3
Vermont	3
Ohio	3
Kansas	2
District of Columbia	1
Rhode Island	1
Montana	1
Iowa	1
Nevada	1

```
[40]: ch = df.groupby(['State', 'Stat'])['State'].count().unstack('Stat')
z = ch[['Open', 'Closed']].plot(kind='bar',
    ↪stacked=True, rot=90, figsize=(15,10), fontsize=15)
z.set_xlabel("State", fontsize=15)
z.set_ylabel("No. of complaints", fontsize=15)
z.set_title("No. of Complaints based on State", fontsize=20)
z.legend(labels=['Open', 'Closed'], fontsize=15)
```

[40]: <matplotlib.legend.Legend at 0x7fe2c0137190>



```
[41]: #From the graph we can infer that Georgia has the maximum complaints
```

```
[ ]:
```

```
[42]: #unresolved complaints
```

```
[43]: df.head()
```

```
[43]:
```

	Customer Complaint	Date \
Ticket #		
250635	Comcast Cable Internet Speeds	2015-04-22
223441	Payment disappear - service got disconnected	2015-04-08
242732	Speed and Service	2015-04-18
277946	Comcast Imposed a New Usage Cap of 300GB that ...	2015-05-07
307175	Comcast not working and no service to boot	2015-05-26

	Date_month_year	Time	Received Via	City	State \
Ticket #					
250635	2015-04-22	3:53:50 PM	Customer Care Call	Abingdon	Maryland
223441	2015-08-04	10:22:56 AM	Internet	Acworth	Georgia
242732	2015-04-18	9:55:47 AM	Internet	Acworth	Georgia
277946	2015-07-05	11:59:35 AM	Internet	Acworth	Georgia
307175	2015-05-26	1:25:26 PM	Internet	Acworth	Georgia

	Zip code	Status	Filing on Behalf of Someone	month \
Ticket #				
250635	21009	Closed	No	4
223441	30102	Closed	No	8
242732	30101	Closed	Yes	4
277946	30101	Open	Yes	7
307175	30101	Solved	No	5

	Date of the Month	Complaint Category	Stat	Stat_le
Ticket #				
250635	22	Internet	Closed	0
223441	4	Paym	Closed	0
242732	18	Internet	Closed	0
277946	5	Usage	Open	1
307175	26	Usage	Closed	0

```
[44]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 2224 entries, 250635 to 363614
Data columns (total 15 columns):
```


#	Column	Non-Null Count	Dtype
0	Customer Complaint	2224 non-null	object
1	Date	2224 non-null	datetime64[ns]
2	Date_month_year	2224 non-null	datetime64[ns]
3	Time	2224 non-null	object
4	Received Via	2224 non-null	object
5	City	2224 non-null	object
6	State	2224 non-null	object
7	Zip code	2224 non-null	int64
8	Status	2224 non-null	object
9	Filing on Behalf of Someone	2224 non-null	object
10	month	2224 non-null	int64
11	Date of the Month	2224 non-null	int64
12	Complaint Category	2224 non-null	object
13	Stat	2224 non-null	object
14	Stat_le	2224 non-null	int64

dtypes: datetime64[ns](2), int64(4), object(9)
memory usage: 358.0+ KB

```
[45]: df.head()
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 2224 entries, 250635 to 363614
Data columns (total 15 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Customer Complaint                    2224 non-null   object
1   Date                                 2224 non-null   datetime64[ns]
2   Date_month_year                       2224 non-null   datetime64[ns]
3   Time                                  2224 non-null   object
4   Received Via                          2224 non-null   object
5   City                                  2224 non-null   object
6   State                                 2224 non-null   object
7   Zip code                              2224 non-null   int64
8   Status                                2224 non-null   object
9   Filing on Behalf of Someone           2224 non-null   object
10  month                                 2224 non-null   int64
11  Date of the Month                     2224 non-null   int64
12  Complaint Category                    2224 non-null   object
13  Stat                                  2224 non-null   object
14  Stat_le                              2224 non-null   int64
dtypes: datetime64[ns](2), int64(4), object(9)
memory usage: 358.0+ KB
```

```
[46]: u = df[df['Stat']=='Open'].groupby(['State'])['Stat'].count().fillna(0)
u=pd.DataFrame(u)
u['UNRESOLVED COMPLAINTS']=u['Stat']
u=u.drop(['Stat'],axis=1)
u
```

```
[46]:
```

	UNRESOLVED COMPLAINTS
State	
Alabama	9
Arizona	6
California	61
Colorado	22
Connecticut	3
Delaware	4
District Of Columbia	2
Florida	39
Georgia	80
Illinois	29
Indiana	9
Kansas	1
Kentucky	3
Louisiana	1
Maine	2
Maryland	15
Massachusetts	11
Michigan	23
Minnesota	4
Mississippi	16
Missouri	1
New Hampshire	4
New Jersey	19
New Mexico	4
Oregon	13
Pennsylvania	20
South Carolina	3
Tennessee	47
Texas	22
Utah	6
Vermont	1
Virginia	11
Washington	23
West Virginia	3

```
[48]: df.head()
```

```
[48]:
```

	Customer Complaint	Date \
Ticket #		

250635	Comcast Cable Internet Speeds	2015-04-22
223441	Payment disappear - service got disconnected	2015-04-08
242732	Speed and Service	2015-04-18
277946	Comcast Imposed a New Usage Cap of 300GB that ...	2015-05-07
307175	Comcast not working and no service to boot	2015-05-26

	Date_month_year	Time	Received Via	City	State \
Ticket #					
250635	2015-04-22	3:53:50 PM	Customer Care Call	Abingdon	Maryland
223441	2015-08-04	10:22:56 AM	Internet	Acworth	Georgia
242732	2015-04-18	9:55:47 AM	Internet	Acworth	Georgia
277946	2015-07-05	11:59:35 AM	Internet	Acworth	Georgia
307175	2015-05-26	1:25:26 PM	Internet	Acworth	Georgia

	Zip code	Status	Filing on Behalf of Someone	month \
Ticket #				
250635	21009	Closed	No	4
223441	30102	Closed	No	8
242732	30101	Closed	Yes	4
277946	30101	Open	Yes	7
307175	30101	Solved	No	5

	Date of the Month	Complaint Category	Stat	Stat_le
Ticket #				
250635	22	Internet	Closed	0
223441	4	Paym	Closed	0
242732	18	Internet	Closed	0
277946	5	Usage	Open	1
307175	26	Usage	Closed	0

```
[49]: d = df[df['Stat']=='Closed'].groupby(['State'])['Stat'].count().fillna(0)
d=pd.DataFrame(d)
d['RESOLVED COMPLAINTS']=d['Stat']
d=d.drop(['Stat'],axis=1)
d
```

```
[49]: RESOLVED COMPLAINTS
```

State	
Alabama	17
Arizona	14
Arkansas	6
California	159
Colorado	58
Connecticut	9
Delaware	8
District Of Columbia	14
District of Columbia	1

Florida	201
Georgia	208
Illinois	135
Indiana	50
Iowa	1
Kansas	1
Kentucky	4
Louisiana	12
Maine	3
Maryland	63
Massachusetts	50
Michigan	92
Minnesota	29
Mississippi	23
Missouri	3
Montana	1
Nevada	1
New Hampshire	8
New Jersey	56
New Mexico	11
New York	6
North Carolina	3
Ohio	3
Oregon	36
Pennsylvania	110
Rhode Island	1
South Carolina	15
Tennessee	96
Texas	49
Utah	16
Vermont	2
Virginia	49
Washington	75
West Virginia	8

```
[50]: t = df.groupby(['State'])['Stat'].count()
      t=pd.DataFrame(t)
```

```
[52]: u
```

```
[52]: UNRESOLVED COMPLAINTS
      State
Alabama          9
Arizona          6
California       61
Colorado        22
Connecticut       3
```

Delaware	4
District Of Columbia	2
Florida	39
Georgia	80
Illinois	29
Indiana	9
Kansas	1
Kentucky	3
Louisiana	1
Maine	2
Maryland	15
Massachusetts	11
Michigan	23
Minnesota	4
Mississippi	16
Missouri	1
New Hampshire	4
New Jersey	19
New Mexico	4
Oregon	13
Pennsylvania	20
South Carolina	3
Tennessee	47
Texas	22
Utah	6
Vermont	1
Virginia	11
Washington	23
West Virginia	3

```
[53]: df.head()
```

```
[53]:
```

	Customer Complaint	Date \
Ticket #		
250635	Comcast Cable Internet Speeds	2015-04-22
223441	Payment disappear - service got disconnected	2015-04-08
242732	Speed and Service	2015-04-18
277946	Comcast Imposed a New Usage Cap of 300GB that ...	2015-05-07
307175	Comcast not working and no service to boot	2015-05-26

	Date_month_year	Time	Received Via	City	State \
Ticket #					
250635	2015-04-22	3:53:50 PM	Customer Care Call	Abingdon	Maryland
223441	2015-08-04	10:22:56 AM	Internet	Acworth	Georgia
242732	2015-04-18	9:55:47 AM	Internet	Acworth	Georgia
277946	2015-07-05	11:59:35 AM	Internet	Acworth	Georgia
307175	2015-05-26	1:25:26 PM	Internet	Acworth	Georgia

	Zip code	Status	Filing on Behalf of Someone	month	\
Ticket #					
250635	21009	Closed	No	4	
223441	30102	Closed	No	8	
242732	30101	Closed	Yes	4	
277946	30101	Open	Yes	7	
307175	30101	Solved	No	5	

	Date of the Month	Complaint Category	Stat	Stat_le
Ticket #				
250635	22	Internet	Closed	0
223441	4	Paym	Closed	0
242732	18	Internet	Closed	0
277946	5	Usage	Open	1
307175	26	Usage	Closed	0

```
[54]: St=df['State'].unique()
```

```
[55]: St=pd.DataFrame(St)
St.columns = ['State']
St
```

```
[55]:
      State
0      Maryland
1      Georgia
2      Michigan
3      California
4      New Mexico
5      Indiana
6      Virginia
7      Illinois
8      Pennsylvania
9      Massachusetts
10     Oregon
11     Texas
12     New Hampshire
13     Minnesota
14     Tennessee
15     Colorado
16     Florida
17     Alabama
18     Washington
19     New York
20     New Jersey
21     Maine
22     Missouri
```

```

23         West Virginia
24             Montana
25         Mississippi
26         Connecticut
27             Vermont
28             Kentucky
29         South Carolina
30             Ohio
31             Utah
32             Delaware
33             Arkansas
34             Nevada
35             Louisiana
36             Kansas
37             Arizona
38         North Carolina
39             Rhode Island
40 District Of Columbia
41 District of Columbia
42             Iowa

```

```

[56]: m=pd.merge(St,u,on=["State"],how="outer").fillna(0)
      m

```

```

[56]:
      State  UNRESOLVED COMPLAINTS
0      Maryland                15.0
1       Georgia                80.0
2      Michigan                23.0
3      California               61.0
4      New Mexico                 4.0
5       Indiana                 9.0
6       Virginia                11.0
7      Illinois                29.0
8      Pennsylvania             20.0
9      Massachusetts            11.0
10      Oregon                 13.0
11      Texas                 22.0
12  New Hampshire                 4.0
13      Minnesota                 4.0
14      Tennessee               47.0
15      Colorado                22.0
16      Florida                 39.0
17      Alabama                 9.0
18      Washington              23.0
19      New York                 0.0
20      New Jersey              19.0
21      Maine                   2.0

```

22	Missouri	1.0
23	West Virginia	3.0
24	Montana	0.0
25	Mississippi	16.0
26	Connecticut	3.0
27	Vermont	1.0
28	Kentucky	3.0
29	South Carolina	3.0
30	Ohio	0.0
31	Utah	6.0
32	Delaware	4.0
33	Arkansas	0.0
34	Nevada	0.0
35	Louisiana	1.0
36	Kansas	1.0
37	Arizona	6.0
38	North Carolina	0.0
39	Rhode Island	0.0
40	District Of Columbia	2.0
41	District of Columbia	0.0
42	Iowa	0.0

[57]: d

[57]: RESOLVED COMPLAINTS

State	
Alabama	17
Arizona	14
Arkansas	6
California	159
Colorado	58
Connecticut	9
Delaware	8
District Of Columbia	14
District of Columbia	1
Florida	201
Georgia	208
Illinois	135
Indiana	50
Iowa	1
Kansas	1
Kentucky	4
Louisiana	12
Maine	3
Maryland	63
Massachusetts	50
Michigan	92

Minnesota	29
Mississippi	23
Missouri	3
Montana	1
Nevada	1
New Hampshire	8
New Jersey	56
New Mexico	11
New York	6
North Carolina	3
Ohio	3
Oregon	36
Pennsylvania	110
Rhode Island	1
South Carolina	15
Tennessee	96
Texas	49
Utah	16
Vermont	2
Virginia	49
Washington	75
West Virginia	8

```
[58]: m2=pd.merge(d,m,on=["State"],how="outer")
```

```
[59]: m2
```

```
[59]:
```

	State	RESOLVED COMPLAINTS	UNRESOLVED COMPLAINTS
0	Alabama	17	9.0
1	Arizona	14	6.0
2	Arkansas	6	0.0
3	California	159	61.0
4	Colorado	58	22.0
5	Connecticut	9	3.0
6	Delaware	8	4.0
7	District Of Columbia	14	2.0
8	District of Columbia	1	0.0
9	Florida	201	39.0
10	Georgia	208	80.0
11	Illinois	135	29.0
12	Indiana	50	9.0
13	Iowa	1	0.0
14	Kansas	1	1.0
15	Kentucky	4	3.0
16	Louisiana	12	1.0
17	Maine	3	2.0
18	Maryland	63	15.0

19	Massachusetts	50	11.0
20	Michigan	92	23.0
21	Minnesota	29	4.0
22	Mississippi	23	16.0
23	Missouri	3	1.0
24	Montana	1	0.0
25	Nevada	1	0.0
26	New Hampshire	8	4.0
27	New Jersey	56	19.0
28	New Mexico	11	4.0
29	New York	6	0.0
30	North Carolina	3	0.0
31	Ohio	3	0.0
32	Oregon	36	13.0
33	Pennsylvania	110	20.0
34	Rhode Island	1	0.0
35	South Carolina	15	3.0
36	Tennessee	96	47.0
37	Texas	49	22.0
38	Utah	16	6.0
39	Vermont	2	1.0
40	Virginia	49	11.0
41	Washington	75	23.0
42	West Virginia	8	3.0

```
[60]: m3=pd.merge(St,Dst,on=["State"],how="outer")
      m3
```

```
[60]:
```

	State	Stat
0	Maryland	78
1	Georgia	288
2	Michigan	115
3	California	220
4	New Mexico	15
5	Indiana	59
6	Virginia	60
7	Illinois	164
8	Pennsylvania	130
9	Massachusetts	61
10	Oregon	49
11	Texas	71
12	New Hampshire	12
13	Minnesota	33
14	Tennessee	143
15	Colorado	80
16	Florida	240
17	Alabama	26

18	Washington	98
19	New York	6
20	New Jersey	75
21	Maine	5
22	Missouri	4
23	West Virginia	11
24	Montana	1
25	Mississippi	39
26	Connecticut	12
27	Vermont	3
28	Kentucky	7
29	South Carolina	18
30	Ohio	3
31	Utah	22
32	Delaware	12
33	Arkansas	6
34	Nevada	1
35	Louisiana	13
36	Kansas	2
37	Arizona	20
38	North Carolina	3
39	Rhode Island	1
40	District Of Columbia	16
41	District of Columbia	1
42	Iowa	1

```
[102]: p= m['UNRESOLVED COMPLAINTS'].div(m3['Stat']).mul(100)
p=pd.DataFrame(p)
p['State']=St['State']
p.columns=['% of unresolved complaints','State']

p
```

```
[102]:
```

	% of unresolved complaints	State
0	19.230769	Maryland
1	27.777778	Georgia
2	20.000000	Michigan
3	27.727273	California
4	26.666667	New Mexico
5	15.254237	Indiana
6	18.333333	Virginia
7	17.682927	Illinois
8	15.384615	Pennsylvania
9	18.032787	Massachusetts
10	26.530612	Oregon
11	30.985915	Texas
12	33.333333	New Hampshire

13	12.121212	Minnesota
14	32.867133	Tennessee
15	27.500000	Colorado
16	16.250000	Florida
17	34.615385	Alabama
18	23.469388	Washington
19	0.000000	New York
20	25.333333	New Jersey
21	40.000000	Maine
22	25.000000	Missouri
23	27.272727	West Virginia
24	0.000000	Montana
25	41.025641	Mississippi
26	25.000000	Connecticut
27	33.333333	Vermont
28	42.857143	Kentucky
29	16.666667	South Carolina
30	0.000000	Ohio
31	27.272727	Utah
32	33.333333	Delaware
33	0.000000	Arkansas
34	0.000000	Nevada
35	7.692308	Louisiana
36	50.000000	Kansas
37	30.000000	Arizona
38	0.000000	North Carolina
39	0.000000	Rhode Island
40	12.500000	District Of Columbia
41	0.000000	District of Columbia
42	0.000000	Iowa

```
[122]: p=p.loc[:,['State','% of unresolved complaints']]
p.sort_values(by=['% of unresolved complaints'],ascending=False)
#Kansas has the highest % of unresolved complaints
```

```
[122]:
```

	State	% of unresolved complaints
36	Kansas	50.000000
28	Kentucky	42.857143
25	Mississippi	41.025641
21	Maine	40.000000
17	Alabama	34.615385
27	Vermont	33.333333
32	Delaware	33.333333
12	New Hampshire	33.333333
14	Tennessee	32.867133
11	Texas	30.985915
37	Arizona	30.000000

1	Georgia	27.777778
3	California	27.727273
15	Colorado	27.500000
23	West Virginia	27.272727
31	Utah	27.272727
4	New Mexico	26.666667
10	Oregon	26.530612
20	New Jersey	25.333333
26	Connecticut	25.000000
22	Missouri	25.000000
18	Washington	23.469388
2	Michigan	20.000000
0	Maryland	19.230769
6	Virginia	18.333333
9	Massachusetts	18.032787
7	Illinois	17.682927
29	South Carolina	16.666667
16	Florida	16.250000
8	Pennsylvania	15.384615
5	Indiana	15.254237
40	District Of Columbia	12.500000
13	Minnesota	12.121212
35	Louisiana	7.692308
24	Montana	0.000000
38	North Carolina	0.000000
41	District of Columbia	0.000000
39	Rhode Island	0.000000
33	Arkansas	0.000000
34	Nevada	0.000000
19	New York	0.000000
30	Ohio	0.000000
42	Iowa	0.000000

```
[63]: #percentage of complaints resolved till date, which were received through the
      ↪Internet and customer care calls.
```

```
[114]: z=df[df['Stat']=='Closed']
      z1=z['Stat'].count()
      type(z1)
```

```
[114]: numpy.int64
```

```
[89]: t=df[(df['Stat']=='Closed') & (df['Received Via']=='Internet')]
      t
```

```
[89]:
```

	Customer Complaint	Date \
Ticket #		

223441	Payment disappear - service got disconnected	2015-04-08
242732	Speed and Service	2015-04-18
307175	Comcast not working and no service to boot	2015-05-26
338519	ISP Charging for arbitrary data limits with ov...	2015-06-12
359792	Comcast refuses to help troubleshoot and corre...	2015-06-23
...
326963	Internet Services	2015-06-06
256002	i had sent out a check payment comcast	2015-04-24
310847	Comcast Internet Service quality	2015-05-28
254488	problems with internet service	2015-04-24
331188	complaint about comcast	2015-06-09

Ticket #	Date_month_year	Time Received	Via	City	State \
223441	2015-08-04	10:22:56 AM	Internet	Acworth	Georgia
242732	2015-04-18	9:55:47 AM	Internet	Acworth	Georgia
307175	2015-05-26	1:25:26 PM	Internet	Acworth	Georgia
338519	2015-12-06	9:59:40 PM	Internet	Acworth	Georgia
359792	2015-06-23	6:56:14 PM	Internet	Adrian	Michigan
...
326963	2015-06-06	4:01:16 PM	Internet	Wyoming	Michigan
256002	2015-04-24	5:05:36 PM	Internet	York	Pennsylvania
310847	2015-05-28	8:56:14 AM	Internet	York	Pennsylvania
254488	2015-04-24	5:05:25 AM	Internet	York Haven	Pennsylvania
331188	2015-09-06	5:28:41 PM	Internet	Ypsilanti	Michigan

Ticket #	Zip code	Status	Filing on Behalf of Someone	month \
223441	30102	Closed	No	8
242732	30101	Closed	Yes	4
307175	30101	Solved	No	5
338519	30101	Solved	No	12
359792	49221	Solved	No	6
...
326963	49509	Closed	No	6
256002	17401	Closed	No	4
310847	17403	Solved	No	5
254488	17370	Closed	No	4
331188	48197	Solved	No	9

Ticket #	Date of the Month	Complaint Category	Stat	Stat_le
223441	4	Paym	Closed	0
242732	18	Internet	Closed	0
307175	26	Usage	Closed	0
338519	6	Internet	Closed	0
359792	23	Usage	Closed	0

...
326963	6	Internet	Closed	0
256002	24	Paym	Closed	0
310847	28	Internet	Closed	0
254488	24	Internet	Closed	0
331188	6	Usage	Closed	0

[843 rows x 15 columns]

```
[109]: t1=t['Stat'].count()
```

```
[111]: s=(t1/z1)*100  ##percentage of complaints resolved till date, which were
      ↪received through the Internet
s
```

```
[111]: 49.38488576449912
```

```
[115]: df['Received Via'].unique()
```

```
[115]: array(['Customer Care Call', 'Internet'], dtype=object)
```

```
[118]: cu=df[(df['Stat']=='Closed') & (df['Received Via']=='Customer Care Call')]
      cu=cu['Stat'].count()
```

```
[119]: s1=(cu/z1)*100  ##percentage of complaints resolved till date, which were
      ↪received through the Customer Care Call
s1
```

```
[119]: 50.61511423550088
```

```
[ ]: #END
```