

Importing Necessary Libraries

In [144...]

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
plt.rcParams['figure.figsize'] = [15, 8]
import seaborn as sns
sns.set_style("whitegrid")
```

In [145...]

```
data=pd.read_csv("Comcast_telecom_complaints_data.csv")
```

In [146...]

```
data.head()
```

Out[146]:

	Ticket #	Customer Complaint	Date	Date_month_year	Time	Received Via	City	State	Zip code	Status	Filing on Behalf of Someone
0	250635	Comcast Cable Internet Speeds	22-04-15	22-Apr-15	3:53:50 PM	Customer Care Call	Abingdon	Maryland	21009	Closed	No
1	223441	Payment disappear - service got disconnected	04-08-15	04-Aug-15	10:22:56 AM	Internet	Acworth	Georgia	30102	Closed	No
2	242732	Speed and Service	18-04-15	18-Apr-15	9:55:47 AM	Internet	Acworth	Georgia	30101	Closed	Yes
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	Yes
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	No

In [149...]

```
data.dtypes
```

```
Out[149]:
```

Ticket #	object
Customer Complaint	object
Date	object
Date_month_year	object
Time	object
Received Via	object
City	object
State	object
Zip code	int64
Status	object
Filing on Behalf of Someone	object
dtype:	object

```
In [148...]: data.isna().sum()
```

```
Out[148]:
```

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 [150...]: data["Date_month_year"] = pd.to_datetime(data["Date_month_year"])
```

C:\Users\Asus\AppData\Local\Temp\ipykernel_17100\1230593614.py:1: UserWarning:

Could not infer format, so each element will be parsed individually, falling back to `dateutil`. To ensure parsing is consistent and as-expected, please specify a format.

```
In [ ]: data.dtypes
```

Extracting month and day from Date column to get trend chart

```
In [152...]: data["month"] = data["Date_month_year"].dt.month  
data["day"] = data["Date_month_year"].dt.day
```

In [153...]

`data.head()`

Out[153]:

	Ticket #	Customer Complaint	Date	Date_month_year	Time	Received Via	City	State	Zip code	Status	Filing on Behalf of Someone	month	day
0	250635	Comcast Cable Internet Speeds	22-04-15	2015-04-22	3:53:50 PM	Customer Care Call	Abingdon	Maryland	21009	Closed	No	4	22
1	223441	Payment disappear - service got disconnected	04-08-15	2015-08-04	10:22:56 AM	Internet	Acworth	Georgia	30102	Closed	No	8	4
2	242732	Speed and Service	18-04-15	2015-04-18	9:55:47 AM	Internet	Acworth	Georgia	30101	Closed	Yes	4	18
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	Open	Yes	7	5
4	307175	Comcast not working and no service to boot	26-05-15	2015-05-26	1:25:26 PM	Internet	Acworth	Georgia	30101	Solved	No	5	26

In [160...]

`month_group=data.groupby("month", as_index=False)["Customer Complaint"].count()`

In [161...]

`month_group`

Out[161]:

	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

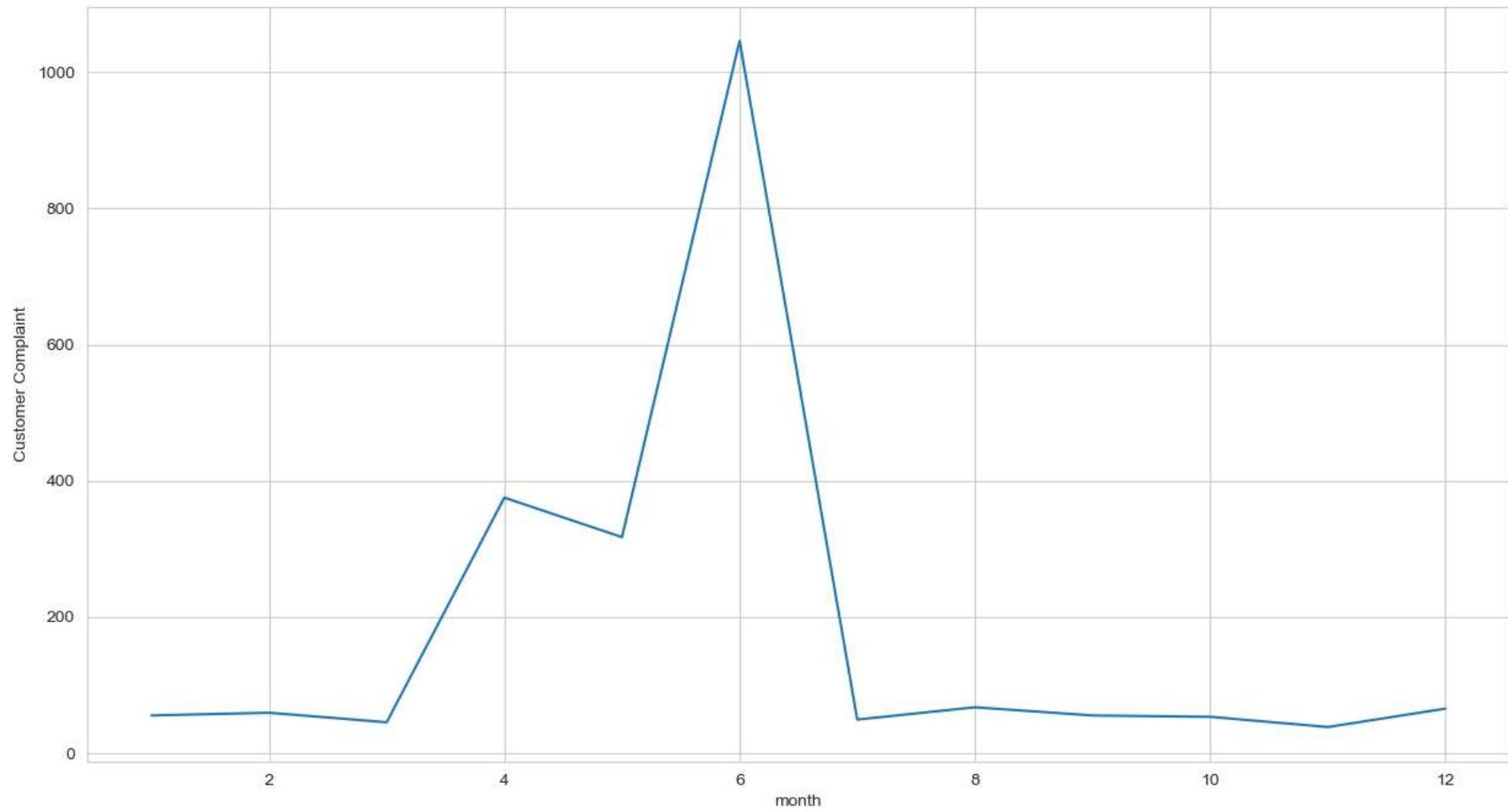
TREND CHART BASED ON MONTHLY COMPLAINTS

In [162...]

```
sns.lineplot(data=month_group, x="month", y="Customer Complaint")
```

Out[162]:

```
<AxesSubplot: xlabel='month', ylabel='Customer Complaint'>
```



```
In [164]: daily_group=data.groupby("day", as_index=False)[ "Customer Complaint" ].count()
```

```
In [165]: daily_group
```

Out[165]:

	day	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

TREND CHART BASED ON DAILY COMPLAINTS

```
In [166]: sns.lineplot(data=daily_group, x="day", y="Customer Complaint")
```

```
Out[166]: <AxesSubplot: xlabel='day', ylabel='Customer Complaint'>
```

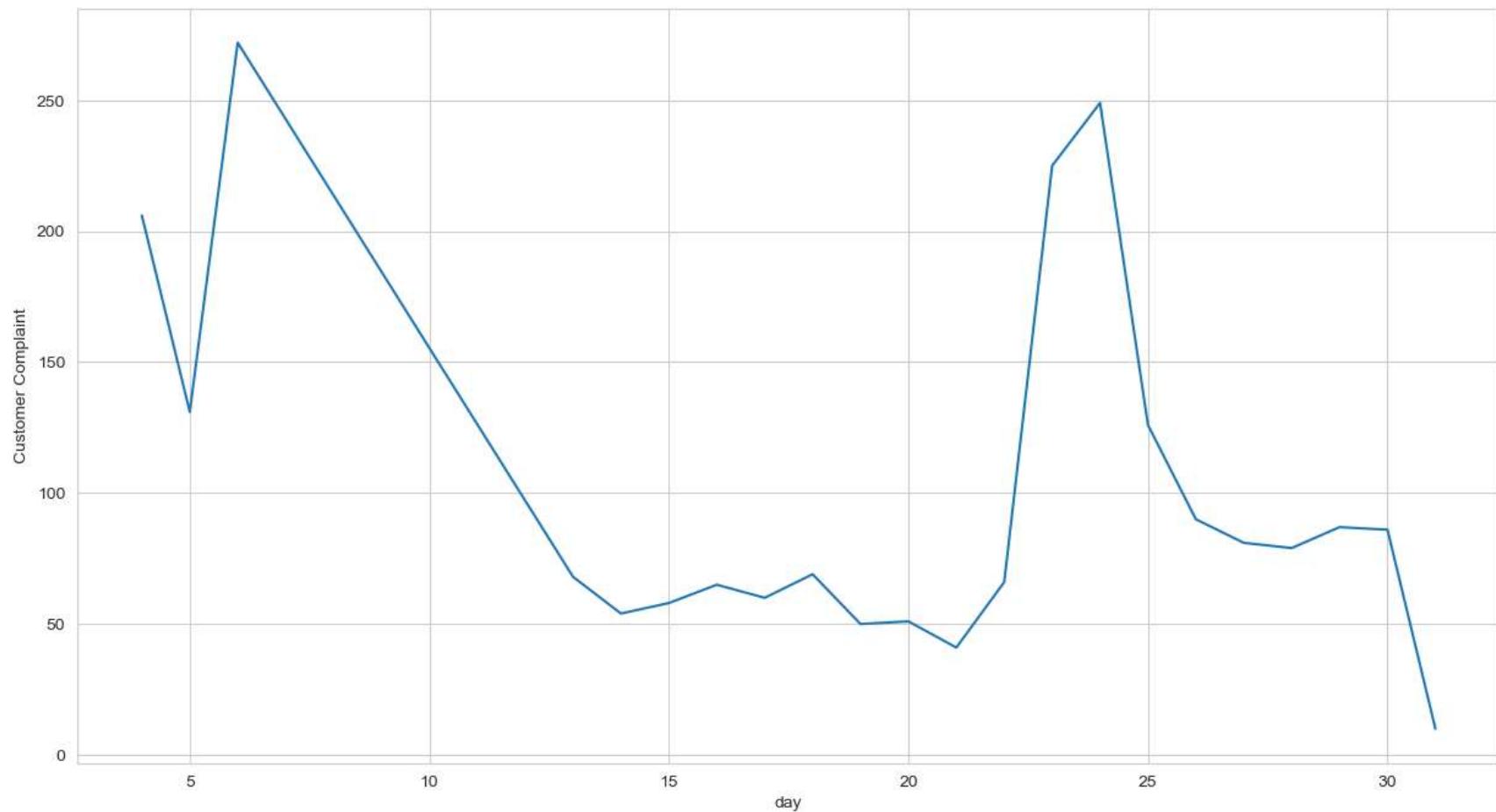


table with the frequency of complaint types

```
In [167]: complaints_frequency=data["Customer Complaint"].value_counts()
```

```
In [168]: complaints_frequency=complaints_frequency.rename_axis('Customer Complaints').reset_index(name='Number of Complaints')
```

```
In [169]: complaints_frequency
```

Out[169]:

	Customer Complaints	Number of Complaints
0	Comcast	83
1	Comcast Internet	18
2	Comcast Data Cap	17
3	comcast	13
4	Comcast Billing	11
...
1836	Improper Billing and non resolution of issues	1
1837	Deceptive trade	1
1838	intermittent internet	1
1839	Internet Speed on Wireless Connection	1
1840	Comcast, Ypsilanti MI Internet Speed	1

1841 rows × 2 columns

In [170...]: `data.rename(columns={"Received Via" : "Complaint types"}, inplace=True)`In []: `data.head()`In [172...]: `data["Complaint types"].value_counts()`Out[172]:
Complaint types
Customer Care Call 1119
Internet 1105
Name: count, dtype: int64In [173...]: `data["Status"].value_counts()`Out[173]:
Status
Solved 973
Closed 734
Open 363
Pending 154
Name: count, dtype: int64

Create a new categorical variable with value as Open and Closed. Open & Pending is to be categorized as Open and Closed & Solved is to be categorized as Closed.

In [174...]

```
conditions=[(data["Status"]=="Open") | (data["Status"]=="Pending"),
            (data["Status"]=="Closed") | (data["Status"]=="Solved"))
values=["Open", "Closed"]
data["New_Status"]=np.select(conditions, values)
```

In [58]:

```
#data["Open"]="Not Open"
#data.loc[(data["Status"]=="Open") | (data["Status"]=="Pending"), "Open"] = "Open"
```

In [59]:

```
#data["Closed"]="Not Closed"
#data.loc[(data["Status"]=="Closed") | (data["Status"]=="Solved"), "Closed"] = "Closed"
```

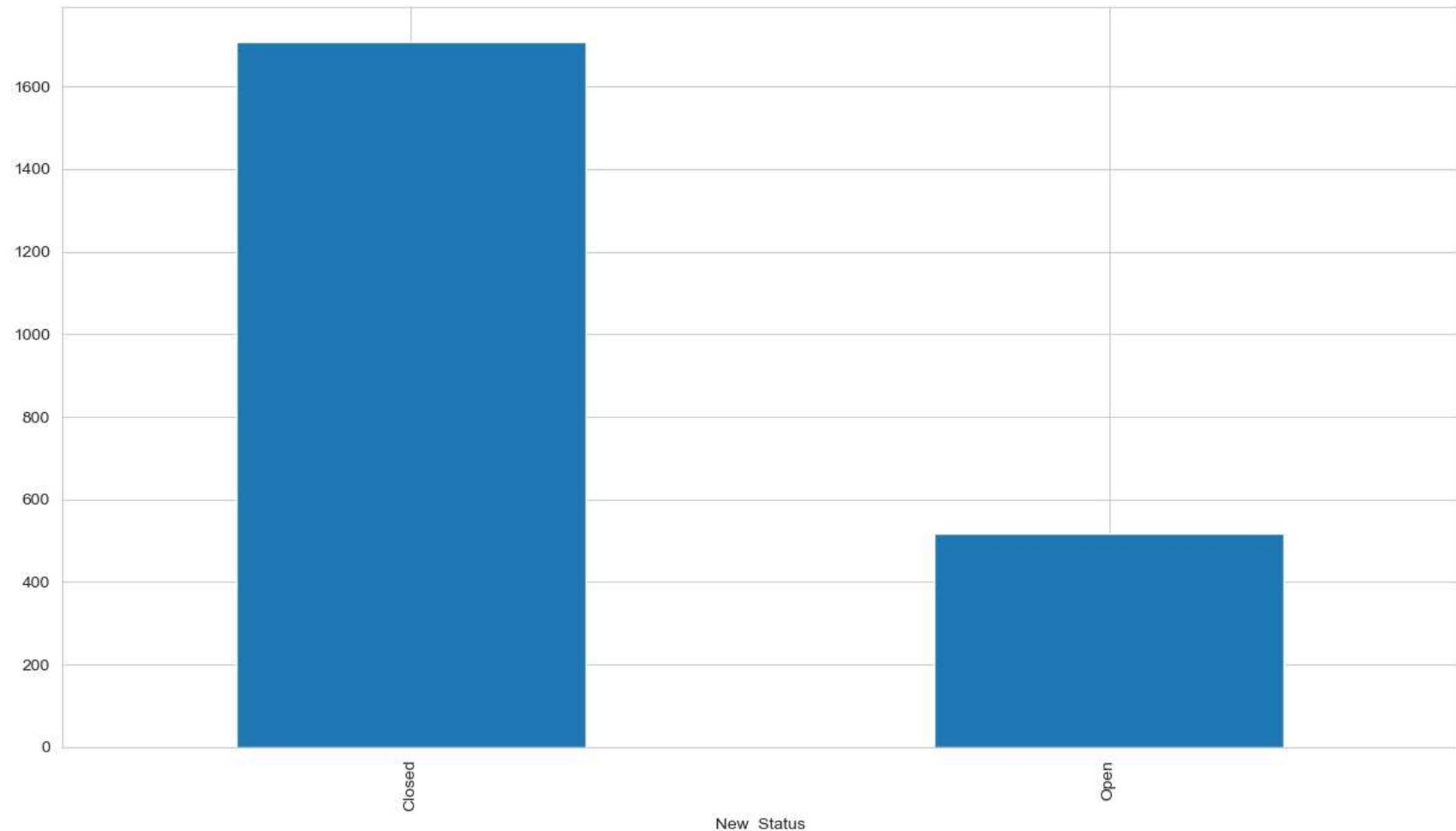
In []: data.head()

In [176...]

```
data["New_Status"].value_counts().plot(kind="bar")
```

Out[176]:

```
<AxesSubplot: xlabel='New_Status'>
```



state wise status of complaints in a stacked bar chart

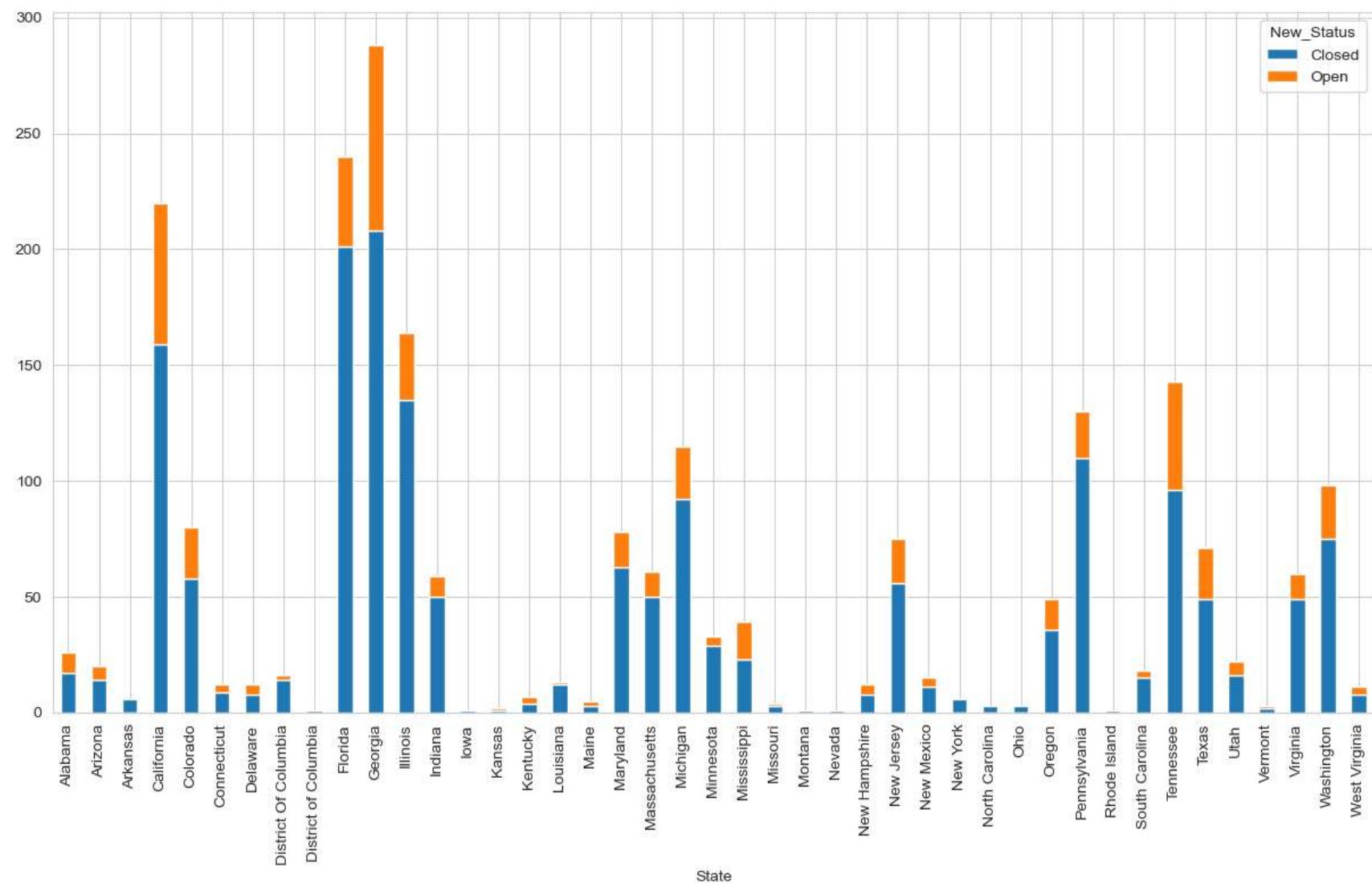
```
In [184]: #state_group=data.groupby(["State", "New_Status"], as_index=False).count()
```

```
In [188]: state_group=data.groupby(["State", "New_Status"]).size().reset_index().pivot(columns='New_Status', index='State', value
```

```
In [190]: state_group.plot(kind='bar', stacked=True)
```

```
Out[190]: <AxesSubplot: xlabel='State'>
```

Comcast_telecom_complaints_data_Analysis



```
In [194]: state_group.sort_values(by="Closed", ascending=False)
```

Out[194]:

	New_Status	Closed	Open
State			
Georgia	208.0	80.0	
Florida	201.0	39.0	
California	159.0	61.0	
Illinois	135.0	29.0	
Pennsylvania	110.0	20.0	
Tennessee	96.0	47.0	
Michigan	92.0	23.0	
Washington	75.0	23.0	
Maryland	63.0	15.0	
Colorado	58.0	22.0	
New Jersey	56.0	19.0	
Indiana	50.0	9.0	
Massachusetts	50.0	11.0	
Texas	49.0	22.0	
Virginia	49.0	11.0	
Oregon	36.0	13.0	
Minnesota	29.0	4.0	
Mississippi	23.0	16.0	
Alabama	17.0	9.0	
Utah	16.0	6.0	
South Carolina	15.0	3.0	
Arizona	14.0	6.0	
District Of Columbia	14.0	2.0	
Louisiana	12.0	1.0	

New_Status	Closed	Open
State		
New Mexico	11.0	4.0
Connecticut	9.0	3.0
West Virginia	8.0	3.0
New Hampshire	8.0	4.0
Delaware	8.0	4.0
Arkansas	6.0	NaN
New York	6.0	NaN
Kentucky	4.0	3.0
Maine	3.0	2.0
North Carolina	3.0	NaN
Missouri	3.0	1.0
Ohio	3.0	NaN
Vermont	2.0	1.0
District of Columbia	1.0	NaN
Rhode Island	1.0	NaN
Iowa	1.0	NaN
Kansas	1.0	1.0
Nevada	1.0	NaN
Montana	1.0	NaN

From the stacked bar and above we can say that *Georgia* state has the maximum complaints

```
In [ ]: #state_group=data.groupby(["State", "New_Status"], as_index=False).count()
```

```
In [ ]: #state_group.reset_index().sort_values(by=["City", "New_Status"], ascending=False)
```

```
In [203...]: counts=data["Complaint_types"].value_counts()  
percentage=data["Complaint_types"].value_counts(normalize=True).mul(100).round(2).astype(str)+"%"  
perc_of_complaints=pd.concat([counts, percentage], axis=1, keys=["counts", "percentage"])
```

Customer Care Call complaints are of more than 50%

```
In [204...]: perc_of_complaints
```

```
Out[204]:
```

Complaint types	counts	percentage
Customer Care Call	1119	50.31%
Internet	1105	49.69%

state has the highest percentage of unresolved complaints

```
In [205...]: state_group.sort_values(by="Open", ascending=False)
```

Out[205]:

	New_Status	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	
Michigan	92.0	23.0	
Washington	75.0	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	
Massachusetts	50.0	11.0	
Virginia	49.0	11.0	
Alabama	17.0	9.0	
Indiana	50.0	9.0	
Arizona	14.0	6.0	
Utah	16.0	6.0	
Delaware	8.0	4.0	
New Hampshire	8.0	4.0	
New Mexico	11.0	4.0	
Minnesota	29.0	4.0	

New_Status	Closed	Open
State		
South Carolina	15.0	3.0
Connecticut	9.0	3.0
West Virginia	8.0	3.0
Kentucky	4.0	3.0
District Of Columbia	14.0	2.0
Maine	3.0	2.0
Louisiana	12.0	1.0
Vermont	2.0	1.0
Missouri	3.0	1.0
Kansas	1.0	1.0
Arkansas	6.0	NaN
District of Columbia	1.0	NaN
Iowa	1.0	NaN
Montana	1.0	NaN
Nevada	1.0	NaN
New York	6.0	NaN
North Carolina	3.0	NaN
Ohio	3.0	NaN
Rhode Island	1.0	NaN

From above table georgia has highest complaints in open state i.e, unresolved complaints

In []: