Comcast Telecom Consumer Complaints

In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt
```

In [2]:
```
#1-Import data into Python environment.
df = pd.read_csv('/Users/Abeer/Downloads/Comcast_telecom_complaints_data.csv')
df.head()
```

Out[2]:

	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 [3]:
```
#Check for Null
df.info()
```

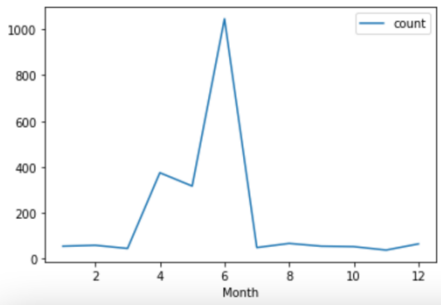
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2224 entries, 0 to 2223
Data columns (total 11 columns):
 #   Column                      Non-Null Count  Dtype 
---  ------                      --------------  ----- 
 0   Ticket #                    2224 non-null   object
 1   Customer Complaint          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 [4]:
```
#2-Provide the trend chart for the number of complaints at monthly and daily granularity levels.

#Convert col Date_month_year data type to date
df['Date_month_year'] = pd.to_datetime(df['Date_month_year'])
df['Month'] = df['Date_month_year'].dt.month
df['Day'] = df['Date_month_year'].dt.day_name()
```

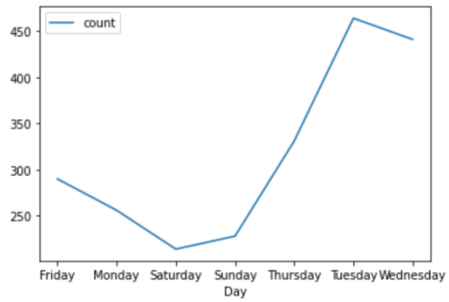
In [5]:
```
#Monthly
Month = df.groupby([df['Month']]).agg({'count'}).sort_values(by='Month')
Month['Ticket #'].plot(kind='line')
```

Out[5]: <AxesSubplot:xlabel='Month'>



In [6]:
```
#Daily
Day = df.groupby([df['Day']]).agg({'count'})
Day['Ticket #'].plot(kind='line')
```

Out[6]: <AxesSubplot:xlabel='Day'>



In [7]:
```
#3-Provide a table with the frequency of complaint types.
complaints = pd.DataFrame({'index':range(df.shape[0])})
df['Customer Complaint'] = df['Customer Complaint'].str.lower()
complaints['Internet'] = df['Customer Complaint'].str.extract("(internet)")
complaints['Network'] = df['Customer Complaint'].str.extract("(network)")
complaints['bills'] = df['Customer Complaint'].str.extract("(billing)")
complaints['charges'] = df['Customer Complaint'].str.extract("(charges)")
complaints['email'] = df['Customer Complaint'].str.extract("(email)")
complaints['data_cap'] = df['Customer Complaint'].str.extract("(data capacity)")
complaints_freq= complaints.notnull().sum().sort_values(ascending=False)
print(complaints_freq)
```

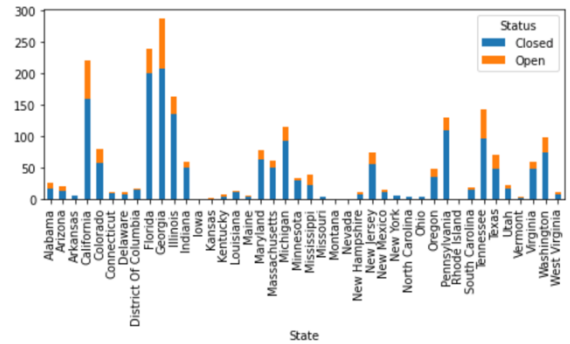
```
index       2224
Internet     532
bills        298
charges       77
email         16
Network        2
data_cap       0
dtype: int64
```

In [8]:
```
#4-Create a new categorical variable with value as Open and Closed.
df.loc[df.Status=='Solved','Status']='Closed'
df.loc[df.Status=='Pending','Status']='Open'
df['Status'].value_counts()
```

Out[8]:
```
Closed    1707
Open       517
Name: Status, dtype: int64
```

In [21]:
```
#5-Provide state wise status of complaints in a stacked bar chart.
comp_st=pd.crosstab(df.State,df.Status)
comp_st.plot(kind='bar',figsize=(8,3),stacked=True)
```

Out[21]: <AxesSubplot:xlabel='State'>



In [23]:
```
#6-Provide the percentage of complaints resolved till date, which were received through the Internet and custome
st = df.groupby(['Status']).size()
st_per = st/st.sum()*100
st_per
```

Out[23]:
```
Status
Closed    76.753597
Open      23.246403
dtype: float64
```

In [24]:
```
resolved_percent = df.groupby(['Received Via','Status']).size().unstack()
resolved_percent['resolved'] = resolved_percent['Closed']/resolved_percent['Closed'].sum()*100
resolved_percent
```

Out[24]:

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

Insights:

- June has the largest number of complaints.
- Number of complaints is significantly larger on Tuesdays and Wednesdays.
- Most of the complaints are related to Internet issues.
- Georgia has the maximum complaints. Also the highest percentage of unresolved complaints.
- 76.8% of the complaints are resolved.
- 50.6% of the resolved complaints were received through customer care call and 49.3 were received through the internet.