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

```python
In [2]: comcast=pd.read_csv('Comcast_telecom_complaints_data.csv')
```

```python
In [3]: comcast.head(5)
```

Out[3]:

| | 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 |

```python
In [4]: comcast.shape
```

Out[4]: (2224, 11)

```python
In [5]: comcast.dtypes
```

Out[5]:
```
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
```

```python
In [6]: comcast.describe()
```

Out[6]:

| | 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 |

```python
In [7]: comcast.columns
```

Out[7]:
```
Index(['Ticket #', 'Customer Complaint', 'Date', 'Date_month_year', 'Time',
       'Received Via', 'City', 'State', 'Zip code', 'Status',
       'Filing on Behalf of Someone'],
      dtype='object')
```

```python
In [8]: comcast.isnull().sum()
```

Out[8]:
```
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
```
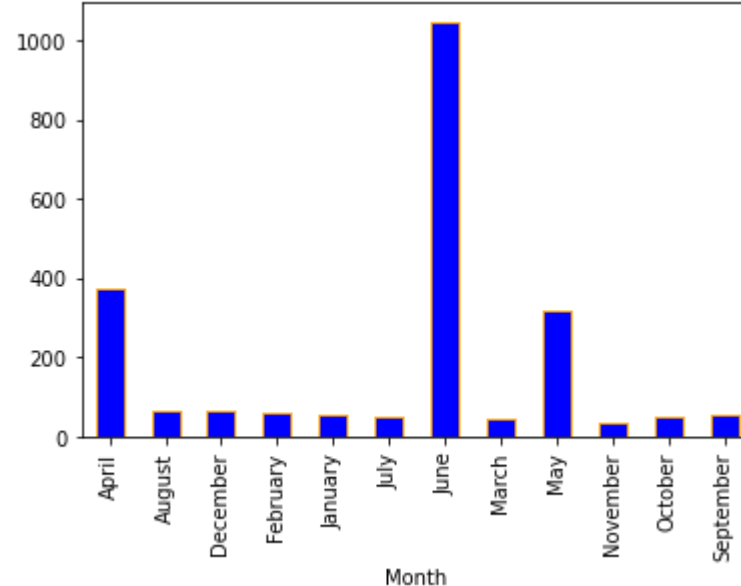
```python
In [9]: comcast['Month']=pd.to_datetime(comcast['Date_month_year']).dt.month_name()
        comcast['Date']=pd.to_datetime(comcast['Date_month_year']).dt.day
```

```python
In [10]: comcast.groupby(['Date'])['Customer Complaint'].count().plot(kind='bar',color="red",edgecolor="darkgreen",)
         plt.show()
```
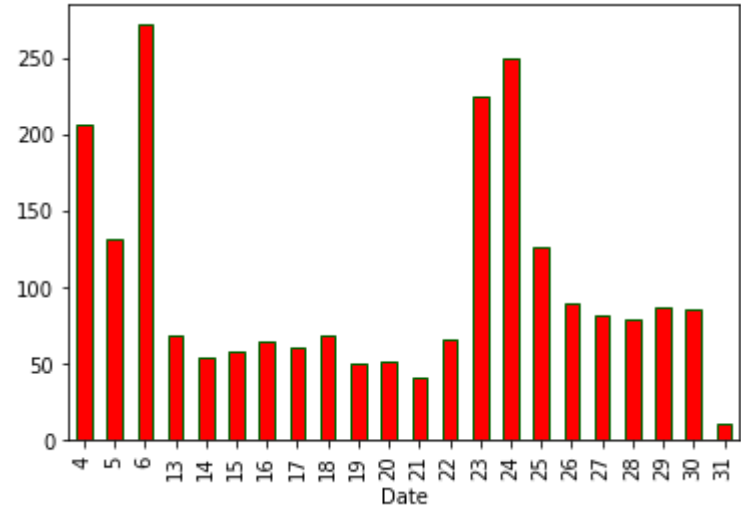


```python
In [11]: comcast.groupby(['Month'])['Customer Complaint'].count().plot(kind='bar',color="blue",edgecolor="orange")
         plt.show()
```



```python
In [12]: comcast['Customer Complaint'].value_counts().to_frame().reset_index()
```

Out[12]:

| | index | Customer Complaint |
|---|---|---|
| 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

```python
In [13]: comcast['Customer Complaint'].value_counts().head(5)
```

Out[13]:
```
Comcast              83
Comcast Internet     18
Comcast Data Cap     17
comcast              13
Comcast Billing      11
Name: Customer Complaint, dtype: int64
```

```python
In [14]: comcast['Status']=comcast['Status'].apply(lambda x: 'Open' if ((x=='Open') | (x=='Pending')) else 'Closed')
```

```python
In [15]: opn=comcast[comcast['Status']=='Open'].groupby(['State'])['Status'].count().to_frame().reset_index()
         clos=comcast[comcast['Status']=='Closed'].groupby(['State'])['Status'].count().to_frame().reset_index()
```

```python
In [16]: fig=plt.figure(figsize=(7,7))
         plt.barh(clos.State, clos.Status)
         plt.barh(opn.State, opn.Status)
         plt.ylabel("State")
         plt.xlabel("Status Count")
         plt.legend(["closed", "open"])
         plt.title("State wise Status Count")

         plt.show()
```
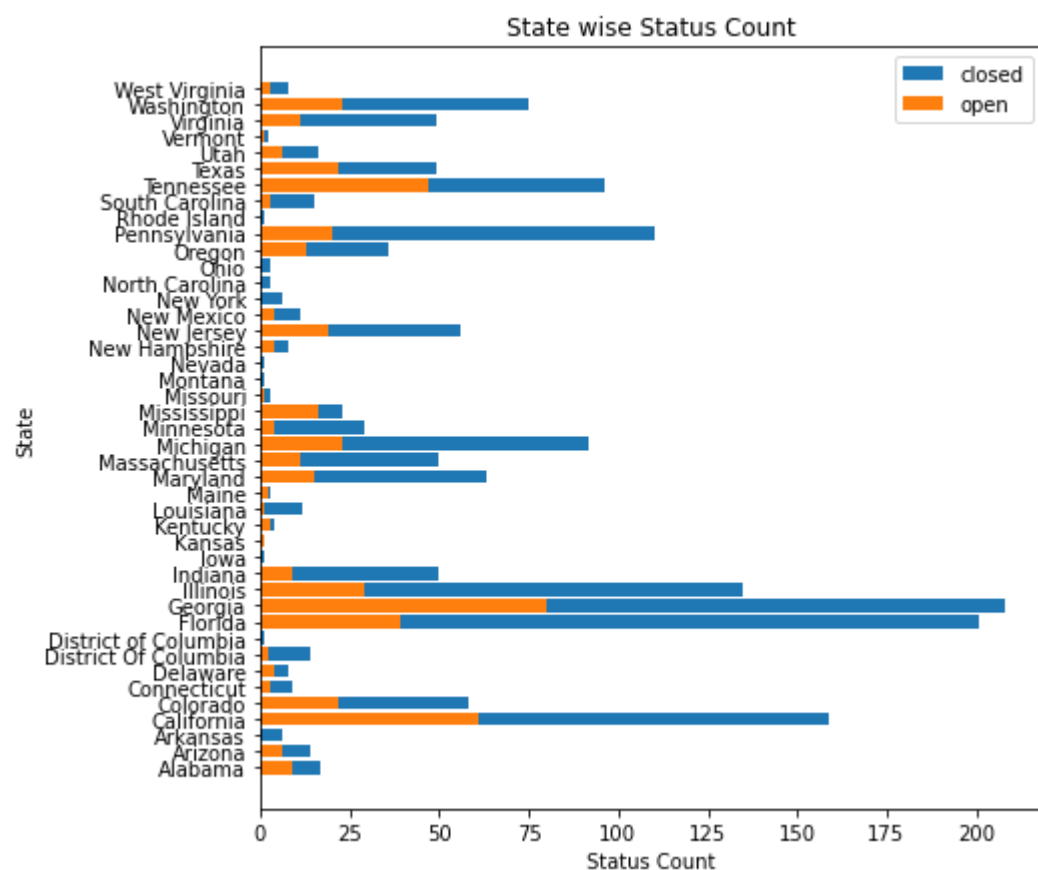


```python
In [17]: comcast.groupby("State")['Customer Complaint'].agg("count").sort_values(ascending=False).head(5)
```

Out[17]:
```
State
Georgia       288
Florida       240
California     220
Illinois      164
Tennessee     143
Name: Customer Complaint, dtype: int64
```

```python
In [18]: State_Unsolved=comcast.loc[comcast['Status']=='Open',['State']].value_counts()
         State_Unsolved.head(1)/State_Unsolved.sum()*100
```

Out[18]:
```
State
Georgia    15.473684
dtype: float64
```

```python
In [19]: comcast[comcast['Status']=='Closed'].groupby('Status')['Received Via'].value_counts(normalize=True)*100
```

Out[19]:
```
Status  Received Via
Closed  Customer Care Call    50.615114
        Internet              49.384886
Name: Received Via, dtype: float64
```

```python
In [ ]:
```