In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
import datetime
from matplotlib import style
style.use('fivethirtyeight')
from matplotlib.ticker import AutoMinorLocator
```

In [2]:
```
### Importing data into Python environment.
```

In [3]:
```
df1=pd.read_csv(r"F:\data sciene course\simplilearn\Data Science and Python\data Sci
```

In [4]:
```
##Importing to Time-Date formate through timestamp in python
```

In [5]:
```
Date_Time=[['Date','Time']]
Comcast_telecom_complaints_data=r"F:\data sciene course\simplilearn\Data Science and
ComeCasteMasterData=pd.read_csv(Comcast_telecom_complaints_data, parse_dates = Date_
ComeCasteMasterData.drop(['Date', 'Date_month_year','Time'], axis=1, inplace=True)
ComeCasteMasterData.head(5)
```

Out[5]:

	Date_Time	Ticket #	Customer Complaint	Received Via	City	State	Zip code	Status	Filing on Behalf of Someone
0	2015-04-22 15:53:50	250635	Comcast Cable Internet Speeds	Customer Care Call	Abingdon	Maryland	21009	Closed	No
1	2015-08-04 10:22:56	223441	Payment disappear - service got disconnected	Internet	Acworth	Georgia	30102	Closed	No
2	2015-04-18 09:55:47	242732	Speed and Service	Internet	Acworth	Georgia	30101	Closed	Yes
3	2015-07-05 11:59:35	277946	Comcast Imposed a New Usage Cap of 300GB that ...	Internet	Acworth	Georgia	30101	Open	Yes
4	2015-05-26 13:25:26	307175	Comcast not working and no service to boot	Internet	Acworth	Georgia	30101	Solved	No

In [6]:
```
#Providing the trend chart for the number of complaints at monthly and daily granula
```

In [7]:
```
#Extracting date and Month
ComeCasteMasterData['date_value'] = ComeCasteMasterData['Date_Time'].dt.date
ComeCasteMasterData["month_of_the_year"]=ComeCasteMasterData['Date_Time'].dt.strftim
ComeCasteMasterData.head()
```

Out[7]:

	Date_Time	Ticket #	Customer Complaint	Received Via	City	State	Zip code	Status	Filing on Behalf of Someone	date_
0	2015-04-22 15:53:50	250635	Comcast Cable Internet Speeds	Customer Care Call	Abingdon	Maryland	21009	Closed	No	201
1	2015-08-04 10:22:56	223441	Payment disappear - service got disconnected	Internet	Acworth	Georgia	30102	Closed	No	201
2	2015-04-18 09:55:47	242732	Speed and Service	Internet	Acworth	Georgia	30101	Closed	Yes	201
3	2015-07-05 11:59:35	277946	Comcast Imposed a New Usage Cap of 300GB that ...	Internet	Acworth	Georgia	30101	Open	Yes	201
4	2015-05-26 13:25:26	307175	Comcast not working and no service to boot	Internet	Acworth	Georgia	30101	Solved	No	201

In [8]:
```
#Calculating Nos of complians per day
df_perday=ComeCasteMasterData.groupby('date_value').size().reset_index().rename(colu
df_perday
```

Out[8]:

	date_value	Complains_Perday
0	2015-01-04	18
1	2015-01-05	12
2	2015-01-06	25
3	2015-02-04	27
4	2015-02-05	7
...	...	...
86	2015-11-05	12
87	2015-11-06	21
88	2015-12-04	15
89	2015-12-05	7
90	2015-12-06	43

91 rows × 2 columns

In [9]:
```
df_perday.max()
```

Out[9]:
```
date_value          2015-12-06
Complains_Perday           218
```

```
dtype: object
```

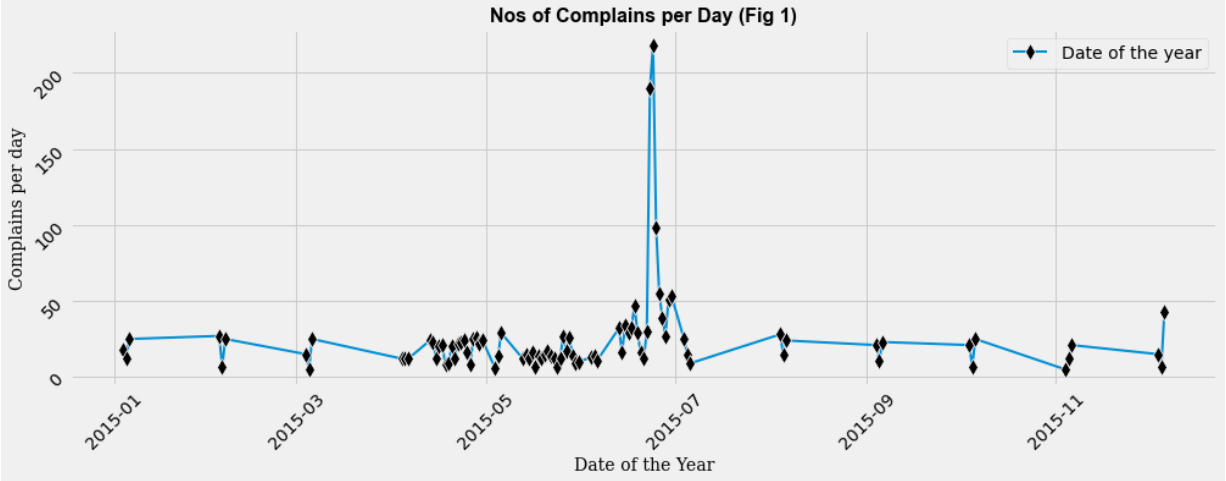
In [27]:
```
#Extracting the month from the timestamp
df=ComeCasteMasterData
df_perMonth=ComeCasteMasterData.groupby('month_of_the_year').size().reset_index().re
#df_perMonth=df_perMonth1.sort_values("Complains_Per_Month", ascending=False)
df_perMonth
```

Out[27]:

	month_of_the_year	Complains_Per_Month
0	Apr-15	375
1	Aug-15	67
2	Dec-15	65
3	Feb-15	59
4	Jan-15	55
5	Jul-15	49
6	Jun-15	1046
7	Mar-15	45
8	May-15	317
9	Nov-15	38
10	Oct-15	53
11	Sep-15	55

Showing trend chart per day basis

In [11]:
```
plt.figure(figsize = (15,5))
sns.lineplot(data=df_perday, x='date_value', y='Complains_Perday',marker = 'd', mark
plt.xlabel('Date of the Year', fontsize = 14, fontfamily = 'Serif')
plt.ylabel('Complains per day', fontsize = 14, fontfamily = 'Serif')
plt.title('Nos of Complains per Day (Fig 1)', fontsize = 16, fontfamily = 'Arial', f
#plt.annotate('Max',xy=(2015-06, 150),  xytext = (5,75), fontsize = 18, fontfamily =
plt.tick_params(labelrotation=45)
plt.legend()
plt.show()
```

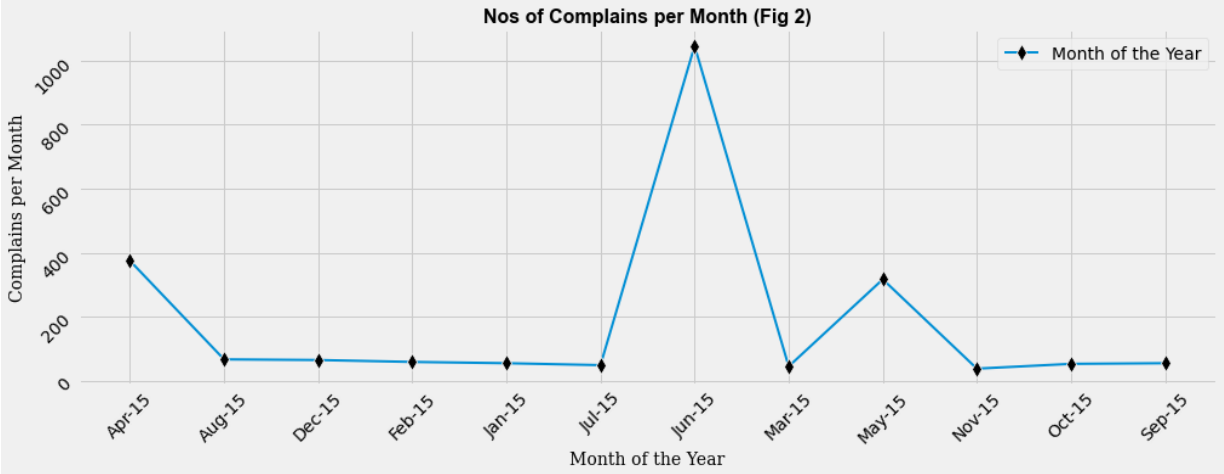


In [12]:
```
print(f"The complain is highest in the day \n{df_perday.max()}")
```

```
The complain is highest in the day
date_value          2015-12-06
Complains_Perday           218
dtype: object
```

Showing trend chart per Month basis

In [13]:
```
plt.figure(figsize = (15,5))
sns.lineplot(data=df_perMonth, x="month_of_the_year", y='Complains_Per_Month',marker
plt.xlabel('Month of the Year', fontsize = 14, fontfamily = 'Serif')
plt.ylabel('Complains per Month', fontsize = 14, fontfamily = 'Serif')
plt.title('Nos of Complains per Month (Fig 2)', fontsize = 16, fontfamily = 'Arial',
#plt.annotate('Max',xy=(2015-06, 150),  xytext = (5,75), fontsize = 18, fontfamily =
plt.tick_params(labelrotation=45)
plt.legend()
plt.show()
```



In [26]:
```
print(f"The complain is highest in the \n{df_perMonth.head(1)}")
```

```
The complain is highest in the
  month_of_the_year  Complains_Per_Month
6            Jun-15                 1046
```

From Fig1 and Fig 2 it can be observed that the complain was highest in the day 2015-12-06 and month wise it was highest in the Month Jun-15

#Table with the frequency of complaint types.

In [15]:
```
df2=ComeCasteMasterData
df2['Filtered Complaint'] = df2['Customer Complaint'].apply(lambda x : x.upper().rep
df2.head(3)
```

Out[15]:

	Date_Time	Ticket #	Customer Complaint	Received Via	City	State	Zip code	Status	Filing on Behalf of Someone	date_
0	2015-04-22 15:53:50	250635	Comcast Cable Internet Speeds	Customer Care Call	Abingdon	Maryland	21009	Closed	No	201

	Date_Time	Ticket #	Customer Complaint	Received Via	City	State	Zip code	Status	Filing on Behalf of Someone	date_
1	2015-08-04 10:22:56	223441	Payment disappear - service got disconnected	Internet	Acworth	Georgia	30102	Closed	No	201
2	2015-04-18 09:55:47	242732	Speed and Service	Internet	Acworth	Georgia	30101	Closed	Yes	201

In [16]:

```
df3=df2
#Defining the class of complain types
strings = ['SPEEDS',"SPEED", 'PAYMENT', 'SERVICE',"BILL","INTERNET", "XFINITY", "DAT
#Fitering based on this class
for x in strings:
    df3.loc[df3['Filtered Complaint'].str.contains(x,na=True),'Complain_type']=x
df3.loc[df3['Filtered Complaint'].str.contains("CHARGE",na=True),'Complain_type']="P
df3.loc[df3['Filtered Complaint'].str.contains("PAYMENT",na=True),'Complain_type']="
df3.loc[df3['Filtered Complaint'].str.contains("COMPLAIN",na=True),'Complain_type']=
df3.loc[df3['Filtered Complaint'].str.contains("CAP",na=True),'Complain_type']="DATA
df3.loc[df3['Filtered Complaint'].str.contains("DECEPTIVE",na=True),'Complain_type']
df3.loc[df3['Filtered Complaint'].str.contains("SPEED",na=True),'Complain_type']="IN
df3.loc[df3['Filtered Complaint'].str.contains("REFUSE",na=True),'Complain_type']="S
df3['Complain_type'] = df3['Complain_type'].fillna("OTHERS")
df3.head()
```

Out[16]:

	Date_Time	Ticket #	Customer Complaint	Received Via	City	State	Zip code	Status	Filing on Behalf of Someone	date_
0	2015-04-22 15:53:50	250635	Comcast Cable Internet Speeds	Customer Care Call	Abingdon	Maryland	21009	Closed	No	201
1	2015-08-04 10:22:56	223441	Payment disappear - service got disconnected	Internet	Acworth	Georgia	30102	Closed	No	201
2	2015-04-18 09:55:47	242732	Speed and Service	Internet	Acworth	Georgia	30101	Closed	Yes	201
3	2015-07-05 11:59:35	277946	Comcast Imposed a New Usage Cap of 300GB that ...	Internet	Acworth	Georgia	30101	Open	Yes	201
4	2015-05-26 13:25:26	307175	Comcast not working and no service to boot	Internet	Acworth	Georgia	30101	Solved	No	201

In [17]:

```
#Calculating Complain Frequency and Complain type percentage for each complain class
df_ComplainType=df3.groupby('Complain_type').size().reset_index().rename(columns={0:
```

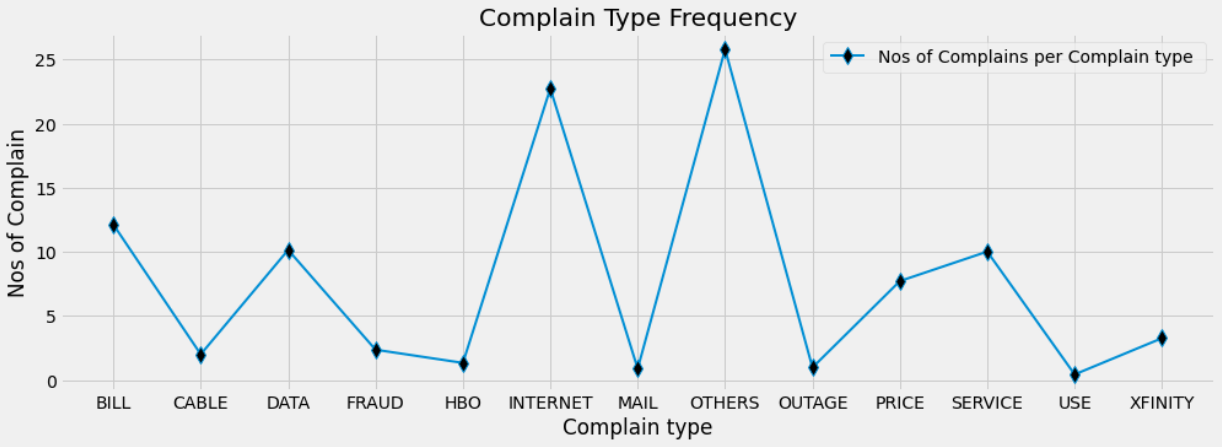
```
df_ComplainType["COMPLAIN_TYPE_PERCENT"]=df_ComplainType["Complain_Type_Frequency"]/
df_ComplainType["COMPLAIN_TYPE_PERCENT"] = np.round(df_ComplainType["COMPLAIN_TYPE_P
df_ComplainType
```

Out[17]:

	Complain_type	Complain_Type_Frequency	COMPLAIN_TYPE_PERCENT
0	BILL	270	12.14
1	CABLE	45	2.02
2	DATA	226	10.16
3	FRAUD	53	2.38
4	HBO	30	1.35
5	INTERNET	506	22.75
6	MAIL	20	0.90
7	OTHERS	574	25.81
8	OUTAGE	22	0.99
9	PRICE	172	7.73
10	SERVICE	223	10.03
11	USE	10	0.45
12	XFINITY	73	3.28

In [18]:

```
fig, ax = plt.subplots(figsize=(15,5))
x = df_ComplainType['Complain_type']
y = df_ComplainType['COMPLAIN_TYPE_PERCENT']
ax.plot(x, y,marker = 'd', markersize = 10, markerfacecolor = 'Black',  lw = 2, labe
ax.set_xlabel('Complain type')
ax.set_ylabel('Nos of Complain')
ax.set_title('Complain Type Frequency')
plt.legend()
plt.show()
```



From the graph it can be observed that the complains are mostly in Internet domain. Though in the graph and data the No of complains are most in OTHERS category, it is not to be considered as it is unclassified data which does not belong to any particular domain and even if it is classified, no one class under OTHERS category will be able to cross the Nos of complain in internet domain.

Creating 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 [19]:
```
df_Open_Close=ComeCasteMasterData
df_Open_Close.loc[df_Open_Close['Status'].str.contains('Closed|Solved',na=True),'Com
df_Open_Close.loc[df_Open_Close['Status'].str.contains('Open|Pending',na=True),'Comp
df_Open_Close.head()
```

Out[19]:

	Date_Time	Ticket #	Customer Complaint	Received Via	City	State	Zip code	Status	Filing on Behalf of Someone	date_
0	2015-04-22 15:53:50	250635	Comcast Cable Internet Speeds	Customer Care Call	Abingdon	Maryland	21009	Closed	No	201
1	2015-08-04 10:22:56	223441	Payment disappear - service got disconnected	Internet	Acworth	Georgia	30102	Closed	No	201
2	2015-04-18 09:55:47	242732	Speed and Service	Internet	Acworth	Georgia	30101	Closed	Yes	201
3	2015-07-05 11:59:35	277946	Comcast Imposed a New Usage Cap of 300GB that ...	Internet	Acworth	Georgia	30101	Open	Yes	201
4	2015-05-26 13:25:26	307175	Comcast not working and no service to boot	Internet	Acworth	Georgia	30101	Solved	No	201

In [20]:
```
#df_statwise_complain_grouped=df_Open_Close[["State",'Complain Closed/Open']].groupb
df_statwise_complain_grouped=df_Open_Close[["State",'Complain Closed/Open']].groupby
#df_statwise_complain_grouped
df_statwise_complain_grouped
```

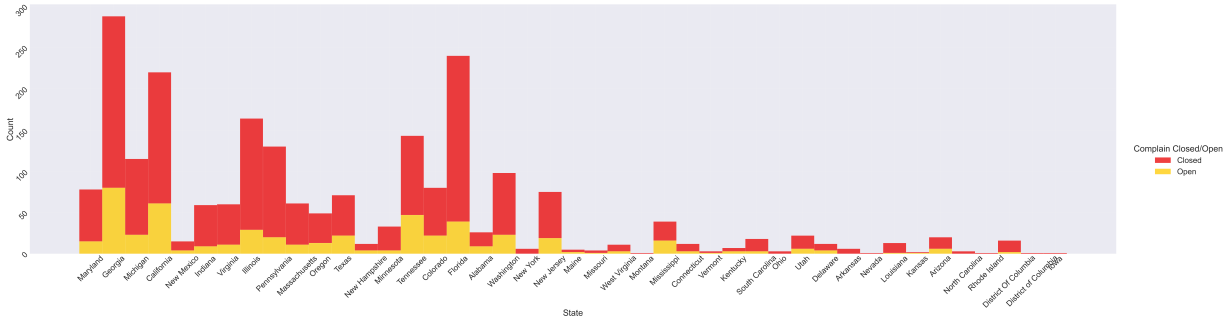
Out[20]:

	State	Complain Closed/Open	Complain_Type_Frequency
0	Alabama	Closed	17
1	Alabama	Open	9
2	Arizona	Closed	14
3	Arizona	Open	6
4	Arkansas	Closed	6
...	...	...	...
72	Virginia	Open	11
73	Washington	Closed	75
74	Washington	Open	23
75	West Virginia	Closed	8

	State	Complain Closed/Open	Complain_Type_Frequency
76	West Virginia	Open	3

77 rows × 3 columns

In [21]:
```
sns.set(font_scale=10)
sns.displot(data=df_Open_Close, x="State", hue='Complain Closed/Open',multiple="stac
plt.tick_params(labelrotation=45)
plt.show()
```



In [22]:
```
#Calculating the state with highest no of complains (Method2)
df_state_sorted=df_Open_Close.sort_values(by=["State"],ascending=False)
dfWithHigestComplain=df_state_sorted["State"].value_counts().head(1)
print(f"State with Highest No of clomplain is\n{dfWithHigestComplain}")
```

```
State with Highest No of clomplain is
Georgia    288
Name: State, dtype: int64
```

Which state has the highest percentage of unresolved complaints

In [23]:
```
StateWithHighestUnresolvedComplains=df_statwise_complain_grouped
StateWithHighestUnresolvedComplainsOpen=df_statwise_complain_grouped.copy()
#Filtering and copying the Open or unresolved complains
StateWithHighestUnresolvedComplainsOpen=StateWithHighestUnresolvedComplains[(StateWi
#Calculating the percentage of Unresolved complain
StateWithHighestUnresolvedComplainsOpen["Unresolved_Complain_Percentage"]=(StateWith

print(f"State with highest percentage of unresolved complaints \n {StateWithHighestU
```

```
State with highest percentage of unresolved complaints
 State                              West Virginia
Complain Closed/Open                        Open
Complain_Type_Frequency                       80
Unresolved_Complain_Percentage         15.473888
dtype: object
```

The percentage of complaints resolved till date, which were received through the Internet and customer care calls.

In [24]:
```
ComplainInternetCustomercare=df_Open_Close
ComplainInternetCustomercar=ComplainInternetCustomercare[(ComplainInternetCustomerca
ComplainInternetCustomercar=ComplainInternetCustomercare[(ComplainInternetCustomerca
PercentageOfComplainResovled=ComplainInternetCustomercar["Complain Closed/Open"].cou
print(f"The percentage of complaints resolved till date that were received through t
```

```
The percentage of complaints resolved till date that were received through the inter
```

```
net and customercare calls 76.75
```

In [ ]: