

Analysis 311 is a non-emergency phone number that people can call to find information about services, make complaints, or report problems like graffiti or road damage

the Question for this data :

- What are the most closed sources ?
- What are the most closed month ?

In [1]:

```
#Load libraries:
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sb
```

In [2]:

```
#Load Data and display it:
data=pd.read_excel('cases_sf_clean.xlsx',encoding='utf-8-sig')
```

In [3]:

```
data.head()
```

Out[3]:

	CaseID	Opened	Updated	Closed	Days from Open to Update	Days from Open to Close	Days from Update to Close	Status	Responsible Agency	Cate
0	7298748	2017-06-25 14:08:43	2017-06-27 09:36:00	2017-06-28 06:51:27	2	3	1	Closed	DPT Abandoned Vehicles Work Queue	Aband Ve
1	7298766	2017-06-25 14:15:34	2017-06-26 00:00:00	2017-06-30 07:59:54	1	5	4	Closed	DPT Abandoned Vehicles Work Queue	Aband Ve
2	7446890	2017-08-01 09:31:08	2017-08-02 02:24:00	2017-08-03 09:10:05	1	2	1	Closed	DPT Meter_Bike Queue	Dam Pro
3	7446243	2017-08-01 07:56:48	2017-08-01 12:00:00	2017-08-02 08:44:54	0	1	1	Closed	DPT Abandoned Vehicles Work Queue	Aband Ve
4	9079212	2018-06-01 08:33:00	2018-06-04 12:00:00	2018-06-08 07:43:00	3	7	4	Closed	DPT Abandoned Vehicles Work Queue	Aband Ve

In [4]:

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 191334 entries, 0 to 191333  
Data columns (total 14 columns):  
CaseID                191334 non-null int64  
Opened                191334 non-null datetime64[ns]  
Updated               191334 non-null datetime64[ns]  
Closed                191334 non-null datetime64[ns]  
Days from Open to Update  191334 non-null int64  
Days from Open to Close  191334 non-null int64  
Days from Update to Close 191334 non-null int64  
Status                191334 non-null object  
Responsible Agency     191334 non-null object  
Category               191334 non-null object  
Request Type           191334 non-null object  
Request Details         191262 non-null object  
Neighborhood           191098 non-null object  
Source                  191334 non-null object  
dtypes: datetime64[ns](3), int64(4), object(7)  
memory usage: 20.4+ MB
```

In [5]:

```
data.shape
```

Out[5]:

```
(191334, 14)
```

In [6]:

```
# check for the sum missing data  
data.isnull().sum().sum()
```

Out[6]:

```
308
```

In [7]:

```
## Delete the missing value  
data.dropna(inplace=True)
```

In [8]:

```
data['Status'].value_counts()
```

Out[8]:

```
Closed    191026  
Name: Status, dtype: int64
```

In [9]:

```
data['Source'].value_counts()
```

Out[9]:

```
Phone          88874
Mobile/Open311 66229
Web            34564
Twitter         803
Integrated Agency 459
Other Department 96
Email           1
Name: Source, dtype: int64
```

In [10]:

```
## we change the rows datetime to Month
```

```
data['Status'] = data['Status'].map({'Closed': 1})
```

In [11]:

```
df=data.groupby('Source').sum()['Status']
```

In [12]:

```
df
```

Out[12]:

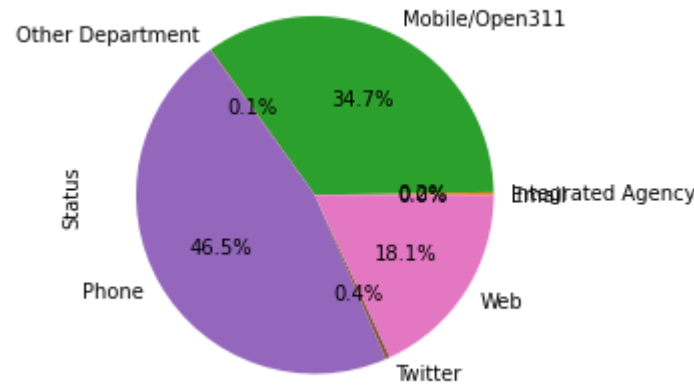
```
Source
Email          1
Integrated Agency 459
Mobile/Open311 66229
Other Department 96
Phone          88874
Twitter         803
Web            34564
Name: Status, dtype: int64
```

In [13]:

```
df.plot.pie(figsize=(4,15),autopct='%1.1f%%')
```

Out[13]:

<matplotlib.axes._subplots.AxesSubplot at 0x16e78b1abe0>



We note that the mobile phone is more closer

In [14]:

```
data.describe()
```

Out[14]:

	CaseID	Days from Open to Update	Days from Open to Close	Days from Update to Close	Status
count	1.910260e+05	191026.000000	191026.000000	191026.000000	191026.0
mean	8.261890e+06	6.325500	12.588417	6.262917	1.0
std	9.138244e+05	19.508197	33.042597	19.515070	0.0
min	6.679618e+06	0.000000	0.000000	0.000000	1.0
25%	7.337821e+06	0.000000	0.000000	0.000000	1.0
50%	8.422340e+06	1.000000	3.000000	1.000000	1.0
75%	9.065070e+06	4.000000	8.000000	4.000000	1.0
max	9.650676e+06	541.000000	605.000000	577.000000	1.0

Days from Open to Update :

- Mean : 6.33
- Mediam: 1
- Standard Deviation :19.51

Days from Open to Close :

- Mean : 12.6
- Median: 3
- Standrd Deviation:33.04

Days from Update to Close:

- Mean :6.3
- Median:1
- Standrd Deviation:19.52

In [15]:

```
data['Source'].value_counts()
```

Out[15]:

```
Phone          88874
Mobile/Open311 66229
Web            34564
Twitter         803
Integrated Agency 459
Other Department 96
Email           1
Name: Source, dtype: int64
```

In [18]:

```
data['Closed']=data['Opened'].dt.month_name()
```

In [19]:

```
# we change the rows datetime to Month
data['Opened']=data['Opened'].dt.month_name()
```

In [20]:

```
data.head(1)
```

Out[20]:

	CaseID	Opened	Updated	Closed	Days from Open to Update	Days from Open to Close	Days from Update to Close	Status	Responsible Agency	Categ
0	7298748	June	2017-06-27 09:36:00	June	2	3	1	1	DPT Abandoned Vehicles Work Queue	Abandon Veh

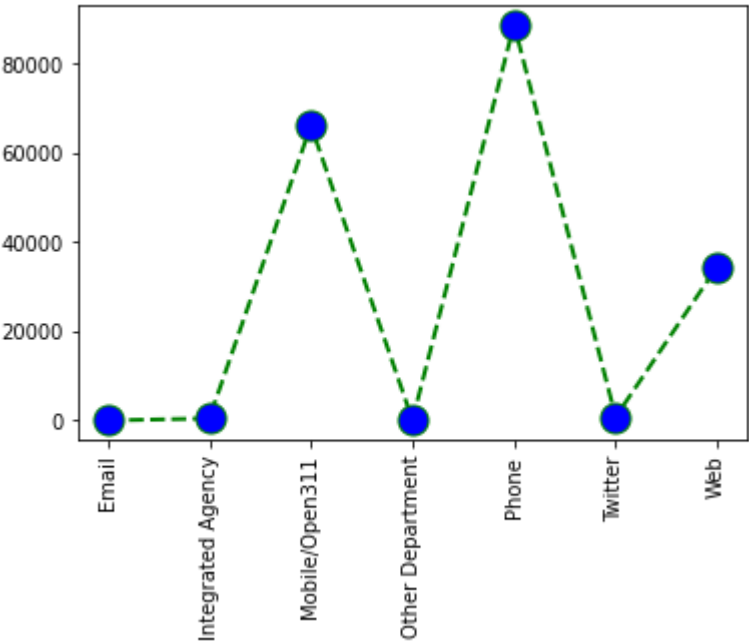
In [21]:

```
df2=data.groupby('Source').sum()['Status']

plt.plot(df2, color='green', linestyle='dashed', linewidth = 2,
         marker='o', markerfacecolor='blue', markersize=15)
plt.xticks(rotation=90)
```

Out[21]:

([0, 1, 2, 3, 4, 5, 6], <a list of 7 Text major ticklabel objects>)



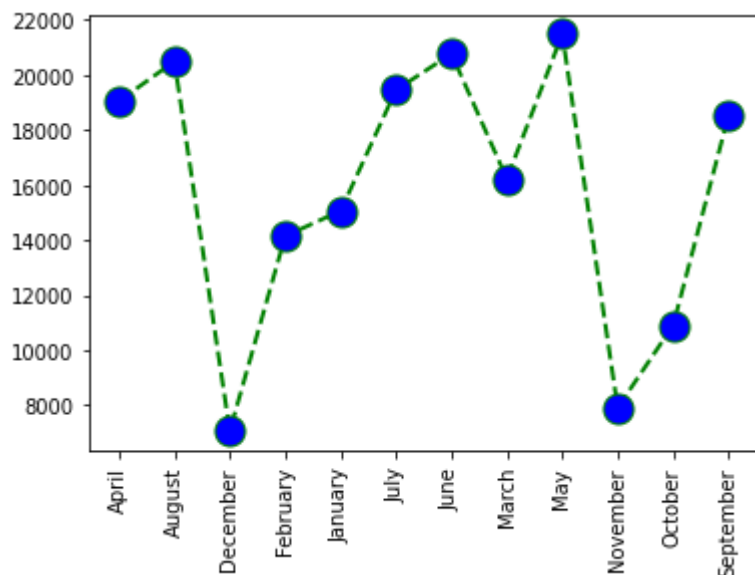
we note that the email is less in closed cases

In [22]:

```
df3=data.groupby('Closed').sum()['Status']  
plt.plot(df3, color='green', linestyle='dashed', linewidth = 2,  
         marker='o', markerfacecolor='blue', markersize=15)  
plt.xticks(rotation=90)
```

Out[22]:

```
([0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11],  
 <a list of 12 Text major ticklabel objects>)
```



we note that April is more in closed cases

In []: