



```
In [1]: #importing major libraries
import pandas as pd
import numpy as np

#additional libraries
import warnings
warnings.filterwarnings('ignore')
```

```
In [2]: df=pd.read_excel('Customer Call List.xlsx')
```

```
In [3]: df.head()
```

```
Out[3]:
```

	CustomerID	First_Name	Last_Name	Phone_Number	Address	Paying Customer
0	1001	Frodo	Baggins	123-545-5421	123 Shire Lane, Shire	Yes
1	1002	Abed	Nadir	123/643/9775	93 West Main Street	No
2	1003	Walter	/White	7066950392	298 Drugs Driveway	N
3	1004	Dwight	Schrute	123-543-2345	980 Paper Avenue, Pennsylvania, 18503	Yes
4	1005	Jon	Snow	876 678 3469	123 Dragons Road	Y

## data wrangling

- data collection: csv, xlsx, surveys, documents, apis, webscrap, databases, warehouse etc
- data assessing
- data cleaning

```
In [4]: #numpy and pandas
```

```
In [5]: df
```

Out[5]:

	CustomerID	First_Name	Last_Name	Phone_Number	Address	Paying_Customer
0	1001	Frodo	Baggins	123-545-5421	123 Shire Lane, Shire	Yes
1	1002	Abed	Nadir	123/643/9775	93 West Main Street	No
2	1003	Walter	/White	7066950392	298 Drugs Driveway	No
3	1004	Dwight	Schrute	123-543-2345	980 Paper Avenue, Pennsylvania, 18503	Yes
4	1005	Jon	Snow	876 678 3469	123 Dragons Road	No
5	1006	Ron	Swanson	304-762-2467	768 City Parkway	Yes
6	1007	Jeff	Winger	NaN	1209 South Street	No
7	1008	Sherlock	Holmes	876 678 3469	98 Clue Drive	No
8	1009	Gandalf	NaN	N/a	123 Middle Earth	Yes
9	1010	Peter	Parker	123-545-5421	25th Main Street, New York	Yes
10	1011	Samwise	Gamgee	NaN	612 Shire Lane, Shire	Yes
11	1012	Harry	...Potter	7066950392	2394 Hogwarts Avenue	No
12	1013	Don	Draper	123-543-2345	2039 Main Street	Yes
13	1014	Leslie	Knope	876 678 3469	343 City Parkway	Yes
14	1015	Toby	Flenderson_	304-762-2467	214 HR Avenue	No
15	1016	Ron	Weasley	123-545-5421	2395 Hogwarts Avenue	No
16	1017	Michael	Scott	123/643/9775	121 Paper Avenue, Pennsylvania	Yes
17	1018	Clark	Kent	7066950392	3498 Super Lane	No

	CustomerID	First_Name	Last_Name	Phone_Number	Address	Paying Customer
<b>18</b>	1019	Creed	Braton	N/a	N/a	N/a
<b>19</b>	1020	Anakin	Skywalker	876 678 3469	910 Tatooine Road, Tatooine	Yes
<b>20</b>	1020	Anakin	Skywalker	876 678 3469	910 Tatooine Road, Tatooine	Yes

## NOW CLEANING GARBAGE FROM LASTNAME

```
In [6]: #Column
df.Last_Name=df.Last_Name.str.strip('/. _')
df
```

Out[6]:

	CustomerID	First_Name	Last_Name	Phone_Number	Address	Paying_Customer
0	1001	Frodo	Baggins	123-545-5421	123 Shire Lane, Shire	Yes
1	1002	Abed	Nadir	123/643/9775	93 West Main Street	No
2	1003	Walter	White	7066950392	298 Drugs Driveway	No
3	1004	Dwight	Schrute	123-543-2345	980 Paper Avenue, Pennsylvania, 18503	Yes
4	1005	Jon	Snow	876 678 3469	123 Dragons Road	No
5	1006	Ron	Swanson	304-762-2467	768 City Parkway	Yes
6	1007	Jeff	Winger	NaN	1209 South Street	No
7	1008	Sherlock	Holmes	876 678 3469	98 Clue Drive	No
8	1009	Gandalf	NaN	N/a	123 Middle Earth	Yes
9	1010	Peter	Parker	123-545-5421	25th Main Street, New York	Yes
10	1011	Samwise	Gamgee	NaN	612 Shire Lane, Shire	Yes
11	1012	Harry	Potter	7066950392	2394 Hogwarts Avenue	No
12	1013	Don	Draper	123-543-2345	2039 Main Street	Yes
13	1014	Leslie	Knope	876 678 3469	343 City Parkway	Yes
14	1015	Toby	Flenderson	304-762-2467	214 HR Avenue	No
15	1016	Ron	Weasley	123-545-5421	2395 Hogwarts Avenue	No
16	1017	Michael	Scott	123/643/9775	121 Paper Avenue, Pennsylvania	Yes
17	1018	Clark	Kent	7066950392	3498 Super Lane	No

	CustomerID	First_Name	Last_Name	Phone_Number	Address	Payin! Customere
<b>18</b>	1019	Creed	Braton	N/a	N/a	N/a
<b>19</b>	1020	Anakin	Skywalker	876 678 3469	910 Tatooine Road, Tatooine	Ye
<b>20</b>	1020	Anakin	Skywalker	876 678 3469	910 Tatooine Road, Tatooine	Ye

```
In [7]: #duplicacy
df.duplicated().sum()
```

```
Out[7]: np.int64(1)
```

```
In [8]: df.drop_duplicates(inplace=True)
```

```
In [9]: df
```

Out[9]:

	CustomerID	First_Name	Last_Name	Phone_Number	Address	Paying_Customer
0	1001	Frodo	Baggins	123-545-5421	123 Shire Lane, Shire	Yes
1	1002	Abed	Nadir	123/643/9775	93 West Main Street	No
2	1003	Walter	White	7066950392	298 Drugs Driveway	No
3	1004	Dwight	Schrute	123-543-2345	980 Paper Avenue, Pennsylvania, 18503	Yes
4	1005	Jon	Snow	876 678 3469	123 Dragons Road	No
5	1006	Ron	Swanson	304-762-2467	768 City Parkway	Yes
6	1007	Jeff	Winger	NaN	1209 South Street	No
7	1008	Sherlock	Holmes	876 678 3469	98 Clue Drive	No
8	1009	Gandalf	NaN	N/a	123 Middle Earth	Yes
9	1010	Peter	Parker	123-545-5421	25th Main Street, New York	Yes
10	1011	Samwise	Gamgee	NaN	612 Shire Lane, Shire	Yes
11	1012	Harry	Potter	7066950392	2394 Hogwarts Avenue	No
12	1013	Don	Draper	123-543-2345	2039 Main Street	Yes
13	1014	Leslie	Knope	876 678 3469	343 City Parkway	Yes
14	1015	Toby	Flenderson	304-762-2467	214 HR Avenue	No
15	1016	Ron	Weasley	123-545-5421	2395 Hogwarts Avenue	No
16	1017	Michael	Scott	123/643/9775	121 Paper Avenue, Pennsylvania	Yes
17	1018	Clark	Kent	7066950392	3498 Super Lane	No

	CustomerID	First_Name	Last_Name	Phone_Number	Address	Paying Customer
<b>18</b>	1019	Creed	Braton	N/a	N/a	N/a
<b>19</b>	1020	Anakin	Skywalker	876 678 3469	910 Tatooine Road, Tatooine	Yes

```
In [10]: #dropping columns
df.shape
```

```
Out[10]: (20, 8)
```

```
In [11]: df.drop(columns='Not_Useful_Column',inplace=True)
```

```
In [12]: df
```

Out[12]:

	CustomerID	First_Name	Last_Name	Phone_Number	Address	Paying_Customer
0	1001	Frodo	Baggins	123-545-5421	123 Shire Lane, Shire	Yes
1	1002	Abed	Nadir	123/643/9775	93 West Main Street	No
2	1003	Walter	White	7066950392	298 Drugs Driveway	No
3	1004	Dwight	Schrute	123-543-2345	980 Paper Avenue, Pennsylvania, 18503	Yes
4	1005	Jon	Snow	876 678 3469	123 Dragons Road	No
5	1006	Ron	Swanson	304-762-2467	768 City Parkway	Yes
6	1007	Jeff	Winger	NaN	1209 South Street	No
7	1008	Sherlock	Holmes	876 678 3469	98 Clue Drive	No
8	1009	Gandalf	NaN	N/a	123 Middle Earth	Yes
9	1010	Peter	Parker	123-545-5421	25th Main Street, New York	Yes
10	1011	Samwise	Gamgee	NaN	612 Shire Lane, Shire	Yes
11	1012	Harry	Potter	7066950392	2394 Hogwarts Avenue	No
12	1013	Don	Draper	123-543-2345	2039 Main Street	Yes
13	1014	Leslie	Knope	876 678 3469	343 City Parkway	Yes
14	1015	Toby	Flenderson	304-762-2467	214 HR Avenue	No
15	1016	Ron	Weasley	123-545-5421	2395 Hogwarts Avenue	No
16	1017	Michael	Scott	123/643/9775	121 Paper Avenue, Pennsylvania	Yes
17	1018	Clark	Kent	7066950392	3498 Super Lane	No

	CustomerID	First_Name	Last_Name	Phone_Number	Address	Paying_Customer
<b>18</b>	1019	Creed	Braton	N/a	N/a	N/a
<b>19</b>	1020	Anakin	Skywalker	876 678 3469	910 Tatooine Road, Tatooine	Yes

```
In [13]: df['Paying_Customer']
df=df.rename({'Paying_Customer':'Paying_Customer'},axis=1)
```

```
In [14]: df.Paying_Customer=df.Paying_Customer.str.replace('Yes','Y')
df.Paying_Customer=df.Paying_Customer.str.replace('No','N')
#df.Paying_Customer
#np.where(condition,value,where)
df.Paying_Customer=np.where(df.Paying_Customer=='N/a','N',df.Paying_Customer)
```

```
In [15]: df.Do_Not_Contact=df.Do_Not_Contact.str.replace('Yes','Y')
df.Do_Not_Contact=df.Do_Not_Contact.str.replace('No','N')
#df.Paying_Customer
#np.where(condition,value,where)
df.Do_Not_Contact=np.where(df.Do_Not_Contact=='N/a','N',df.Do_Not_Contact)
```

```
In [16]: df
```

Out[16]:

	CustomerID	First_Name	Last_Name	Phone_Number	Address	Paying_C
<b>0</b>	1001	Frodo	Baggins	123-545-5421	123 Shire Lane, Shire	
<b>1</b>	1002	Abed	Nadir	123/643/9775	93 West Main Street	
<b>2</b>	1003	Walter	White	7066950392	298 Drugs Driveway	
<b>3</b>	1004	Dwight	Schrute	123-543-2345	980 Paper Avenue, Pennsylvania, 18503	
<b>4</b>	1005	Jon	Snow	876 678 3469	123 Dragons Road	
<b>5</b>	1006	Ron	Swanson	304-762-2467	768 City Parkway	
<b>6</b>	1007	Jeff	Winger	NaN	1209 South Street	
<b>7</b>	1008	Sherlock	Holmes	876 678 3469	98 Clue Drive	
<b>8</b>	1009	Gandalf	NaN	N/a	123 Middle Earth	
<b>9</b>	1010	Peter	Parker	123-545-5421	25th Main Street, New York	
<b>10</b>	1011	Samwise	Gamgee	NaN	612 Shire Lane, Shire	
<b>11</b>	1012	Harry	Potter	7066950392	2394 Hogwarts Avenue	
<b>12</b>	1013	Don	Draper	123-543-2345	2039 Main Street	
<b>13</b>	1014	Leslie	Knope	876 678 3469	343 City Parkway	
<b>14</b>	1015	Toby	Flenderson	304-762-2467	214 HR Avenue	
<b>15</b>	1016	Ron	Weasley	123-545-5421	2395 Hogwarts Avenue	
<b>16</b>	1017	Michael	Scott	123/643/9775	121 Paper Avenue, Pennsylvania	
<b>17</b>	1018	Clark	Kent	7066950392	3498 Super Lane	
<b>18</b>	1019	Creed	Braton	N/a	N/a	

	CustomerID	First_Name	Last_Name	Phone_Number	Address	Paying_C
19	1020	Anakin	Skywalker	876 678 3469	910 Tatooine Road, Tatooine	

```
In [17]: df.isnull().sum()
df.Do_Not_Contact.mode()
df.Do_Not_Contact.fillna('N',inplace=True)
```

```
In [18]: df
```

Out[18]:

	CustomerID	First_Name	Last_Name	Phone_Number	Address	Paying_C
<b>0</b>	1001	Frodo	Baggins	123-545-5421	123 Shire Lane, Shire	
<b>1</b>	1002	Abed	Nadir	123/643/9775	93 West Main Street	
<b>2</b>	1003	Walter	White	7066950392	298 Drugs Driveway	
<b>3</b>	1004	Dwight	Schrute	123-543-2345	980 Paper Avenue, Pennsylvania, 18503	
<b>4</b>	1005	Jon	Snow	876 678 3469	123 Dragons Road	
<b>5</b>	1006	Ron	Swanson	304-762-2467	768 City Parkway	
<b>6</b>	1007	Jeff	Winger	NaN	1209 South Street	
<b>7</b>	1008	Sherlock	Holmes	876 678 3469	98 Clue Drive	
<b>8</b>	1009	Gandalf	NaN	N/a	123 Middle Earth	
<b>9</b>	1010	Peter	Parker	123-545-5421	25th Main Street, New York	
<b>10</b>	1011	Samwise	Gamgee	NaN	612 Shire Lane, Shire	
<b>11</b>	1012	Harry	Potter	7066950392	2394 Hogwarts Avenue	
<b>12</b>	1013	Don	Draper	123-543-2345	2039 Main Street	
<b>13</b>	1014	Leslie	Knope	876 678 3469	343 City Parkway	
<b>14</b>	1015	Toby	Flenderson	304-762-2467	214 HR Avenue	
<b>15</b>	1016	Ron	Weasley	123-545-5421	2395 Hogwarts Avenue	
<b>16</b>	1017	Michael	Scott	123/643/9775	121 Paper Avenue, Pennsylvania	
<b>17</b>	1018	Clark	Kent	7066950392	3498 Super Lane	
<b>18</b>	1019	Creed	Braton	N/a	N/a	

	CustomerID	First_Name	Last_Name	Phone_Number	Address	Paying_C
19	1020	Anakin	Skywalker	876 678 3469	910 Tatooine Road, Tatooine	

```
In [19]: df.dropna(subset='Phone_Number',inplace=True)
df.Phone_Number = df.Phone_Number.astype(str).str.replace('[^0-9]','',regex=True)
```

```
In [23]: df.Phone_Number=np.where(df.Phone_Number=='',np.nan,df.Phone_Number) #n/a -->
df.Phone_Number
```

```
Out[23]: 0    1235455421
1    1236439775
2    7066950392
3    1235432345
4    8766783469
5    3047622467
7    8766783469
9    1235455421
11   7066950392
12   1235432345
13   8766783469
14   3047622467
15   1235455421
16   1236439775
17   7066950392
19   8766783469
Name: Phone_Number, dtype: object
```

```
In [24]: df.dropna(subset='Phone_Number',inplace=True)
df.shape
```

```
Out[24]: (16, 7)
```

```
In [25]: df.Phone_Number
```

```
Out[25]: 0    1235455421
         1    1236439775
         2    7066950392
         3    1235432345
         4    8766783469
         5    3047622467
         7    8766783469
         9    1235455421
        11    7066950392
        12    1235432345
        13    8766783469
        14    3047622467
        15    1235455421
        16    1236439775
        17    7066950392
        19    8766783469
        Name: Phone_Number, dtype: object
```

```
In [26]: print(df.Phone_Number.apply(lambda x: x[0:3]+'-'+x[3:6]+'-'+x[6:]))
         df.Phone_Number=df.Phone_Number.apply(lambda x: x[0:3]+'-'+x[3:6]+'-'+x[6:])

0    123-545-5421
1    123-643-9775
2    706-695-0392
3    123-543-2345
4    876-678-3469
5    304-762-2467
7    876-678-3469
9    123-545-5421
11   706-695-0392
12   123-543-2345
13   876-678-3469
14   304-762-2467
15   123-545-5421
16   123-643-9775
17   706-695-0392
19   876-678-3469
        Name: Phone_Number, dtype: object
```

```
In [32]: name='Anoushka'
         name[0:3]+'-'+name[3:6]+'-'+name[6:]
```

```
Out[32]: 'Ano-ush-ka'
```

```
In [33]: temp=df['Address'].str.split(',',expand=True)
         temp.columns=['Street','State','Zip']
         temp.head()
```

Out[33]:

	Street	State	Zip
0	123 Shire Lane	Shire	None
1	93 West Main Street	None	None
2	298 Drugs Driveway	None	None
3	980 Paper Avenue	Pennsylvania	18503
4	123 Dragons Road	None	None

In [34]:

```
pd.concat((df,temp),axis=1) #temporary
```

Out[34]:

	CustomerID	First_Name	Last_Name	Phone_Number	Address	Paying_C
0	1001	Frodo	Baggins	123-545-5421	123 Shire Lane, Shire	
1	1002	Abed	Nadir	123-643-9775	93 West Main Street	
2	1003	Walter	White	706-695-0392	298 Drugs Driveway	
3	1004	Dwight	Schrute	123-543-2345	980 Paper Avenue, Pennsylvania, 18503	
4	1005	Jon	Snow	876-678-3469	123 Dragons Road	
5	1006	Ron	Swanson	304-762-2467	768 City Parkway	
7	1008	Sherlock	Holmes	876-678-3469	98 Clue Drive	
9	1010	Peter	Parker	123-545-5421	25th Main Street, New York	
11	1012	Harry	Potter	706-695-0392	2394 Hogwarts Avenue	
12	1013	Don	Draper	123-543-2345	2039 Main Street	
13	1014	Leslie	Knope	876-678-3469	343 City Parkway	
14	1015	Toby	Flenderson	304-762-2467	214 HR Avenue	
15	1016	Ron	Weasley	123-545-5421	2395 Hogwarts Avenue	
16	1017	Michael	Scott	123-643-9775	121 Paper Avenue, Pennsylvania	
17	1018	Clark	Kent	706-695-0392	3498 Super Lane	
19	1020	Anakin	Skywalker	876-678-3469	910 Tatooine Road,	

CustomerID	First_Name	Last_Name	Phone_Number	Address	Paying_C
Tatooine					

```
In [35]: df[['Street','State','Zip']] = df['Address'].str.split(',',expand=True)
df.head(2)    #permanent basis
```

```
Out[35]:
```

	CustomerID	First_Name	Last_Name	Phone_Number	Address	Paying_Custor
<b>0</b>	1001	Frodo	Baggins	123-545-5421	123 Shire Lane, Shire	
<b>1</b>	1002	Abed	Nadir	123-643-9775	93 West Main Street	

```
In [40]: df.to_csv('Phone_number.csv',index=False)    #----- file of good manner after
```

```
In [41]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 16 entries, 0 to 19
Data columns (total 10 columns):
#   Column                Non-Null Count  Dtype
---  -
0   CustomerID            16 non-null    int64
1   First_Name            16 non-null    object
2   Last_Name             16 non-null    object
3   Phone_Number          16 non-null    object
4   Address               16 non-null    object
5   Paying_Customer       16 non-null    object
6   Do_Not_Contact        16 non-null    object
7   Street               16 non-null    object
8   State                 5 non-null     object
9   Zip                   1 non-null     object
dtypes: int64(1), object(9)
memory usage: 1.4+ KB
```

```
In [ ]:
```