

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

```
In [95]: # Clean data
# we requiried who do_not_contact only
```

```
In [2]: df=pd.read_excel("F:\\pandas data cleaning\\Customer Call List.xlsx")
```

```
In [ ]:
```

```
In [4]: df.head()
```

Out[4]:

	CustomerID	First_Name	Last_Name	Phone_Number	Address	Paying Customer	Do_Not_Co
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	



```
In [5]: df.columns
```

Out[5]: Index(['CustomerID', 'First_Name', 'Last_Name', 'Phone_Number', 'Address', 'Paying Customer', 'Do_Not_Contact', 'Not_Useful_Column'], dtype='object')

```
In [6]: df["CustomerID"]
```

```
Out[6]: 0      1001
        1      1002
        2      1003
        3      1004
        4      1005
        5      1006
        6      1007
        7      1008
        8      1009
        9      1010
       10      1011
       11      1012
       12      1013
       13      1014
       14      1015
       15      1016
       16      1017
       17      1018
       18      1019
       19      1020
       20      1020
        Name: CustomerID, dtype: int64
```

```
In [7]: df["First_Name"]
```

```
Out[7]: 0      Frodo
        1      Abed
        2      Walter
        3      Dwight
        4      Jon
        5      Ron
        6      Jeff
        7      Sherlock
        8      Gandalf
        9      Peter
       10      Samwise
       11      Harry
       12      Don
       13      Leslie
       14      Toby
       15      Ron
       16      Michael
       17      Clark
       18      Creed
       19      Anakin
       20      Anakin
        Name: First_Name, dtype: object
```

```
In [10]: df["Last_Name"]=df["Last_Name"].str.strip("./")
```

```
In [11]: df["Last_Name"]
```

```
Out[11]: 0      Baggins
1      Nadir
2      White
3      Schrute
4      Snow
5      Swanson
6      Winger
7      Holmes
8      NaN
9      Parker
10     Gamgee
11     Potter
12     Draper
13     Knope
14     Flenderson
15     Weasley
16     Scott
17     Kent
18     Braton
19     Skywalker
20     Skywalker
Name: Last_Name, dtype: object
```

```
In [12]: df.columns
```

```
Out[12]: Index(['CustomerID', 'First_Name', 'Last_Name', 'Phone_Number', 'Address',
               'Paying Customer', 'Do_Not_Contact', 'Not_Useful_Column'],
              dtype='object')
```

```
In [15]: df["Phone_Number"]=df["Phone_Number"].astype(str)
```

```
In [18]: df["Phone_Number"]=df["Phone_Number"].str.replace("[^a-zA-Z0-9]", "")
```

C:\Users\win10\AppData\Local\Temp\ipykernel_8612\3141684028.py:1: FutureWarning: The default value of regex will change from True to False in a future version.

```
df["Phone_Number"]=df["Phone_Number"].str.replace("[^a-zA-Z0-9]", "")
```

```
In [22]: df["Phone_Number"]=df["Phone_Number"].apply(lambda x:x[0:3]+"-"+x[3:6]+"-"+x[6:10])
```

```
In [23]: df["Phone_Number"]
```

```
Out[23]: 0      123-545-5421
1      123-643-9775
2      706-695-0392
3      123-543-2345
4      876-678-3469
5      304-762-2467
6           nan--
7      876-678-3469
8           Na--
9      123-545-5421
10          nan--
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
18           Na--
19     876-678-3469
20     876-678-3469
Name: Phone_Number, dtype: object
```

```
In [25]: df["Phone_Number"]=df["Phone_Number"].replace("nan--", "")
```

```
In [27]: df["Phone_Number"]=df["Phone_Number"].replace("Na--", "")
```

```
In [28]: df["Phone_Number"]
```

```
Out[28]: 0      123-545-5421
1      123-643-9775
2      706-695-0392
3      123-543-2345
4      876-678-3469
5      304-762-2467
6
7      876-678-3469
8
9      123-545-5421
10
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
18
19     876-678-3469
20     876-678-3469
Name: Phone_Number, dtype: object
```

```
In [29]: df.columns
```

```
Out[29]: Index(['CustomerID', 'First_Name', 'Last_Name', 'Phone_Number', 'Address',  
              'Paying Customer', 'Do_Not_Contact', 'Not_Useful_Column'],  
              dtype='object')
```

```
In [46]: df[["Street", "City", "ZipCode"]] = df["Address"].str.split(",", expand=True)
```

```
In [47]: df.columns
```

```
Out[47]: Index(['CustomerID', 'First_Name', 'Last_Name', 'Phone_Number', 'Address',  
              'Paying Customer', 'Do_Not_Contact', 'Not_Useful_Column', 'Street',  
              'City', 'ZipCode'],  
              dtype='object')
```

```
In [49]: df.drop(columns=["Address", "Not_Useful_Column"], inplace=True)
```

In [50]: df

Out[50]:

	CustomerID	First_Name	Last_Name	Phone_Number	Paying Customer	Do_Not_Contact	Stre
0	1001	Frodo	Baggins	123-545-5421	Yes	No	1: Sh La
1	1002	Abed	Nadir	123-643-9775	No	Yes	93 We Ma Stre
2	1003	Walter	White	706-695-0392	N	NaN	2: Dru Drivew
3	1004	Dwight	Schrute	123-543-2345	Yes	Y	9: Pap Aven
4	1005	Jon	Snow	876-678-3469	Y	No	1: Drago Ro
5	1006	Ron	Swanson	304-762-2467	Yes	Yes	768 C Parkw
6	1007	Jeff	Winger		No	No	12: Sou Stre
7	1008	Sherlock	Holmes	876-678-3469	N	No	98 Cl Dri
8	1009	Gandalf	NaN		Yes	NaN	1: Midc Eai
9	1010	Peter	Parker	123-545-5421	Yes	No	25 Ma Stre
10	1011	Samwise	Gamgee		Yes	No	6 Sh La
11	1012	Harry	Potter	706-695-0392	Y	NaN	23: Hogwa Aven
12	1013	Don	Draper	123-543-2345	Yes	N	20: Ma Stre
13	1014	Leslie	Knope	876-678-3469	Yes	No	343 C Parkw
14	1015	Toby	Flenderson	304-762-2467	N	No	214 H Aven
15	1016	Ron	Weasley	123-545-5421	No	N	23: Hogwa Aven
16	1017	Michael	Scott	123-643-9775	Yes	No	1: Pap Aven
17	1018	Clark	Kent	706-695-0392	Y	NaN	34: Sup La
18	1019	Creed	Braton		N/a	Yes	N

	CustomerID	First_Name	Last_Name	Phone_Number	Paying Customer	Do_Not_Contact	Stre
19	1020	Anakin	Skywalker	876-678-3469	Yes	N	Tatooi Ro. 9
20	1020	Anakin	Skywalker	876-678-3469	Yes	N	Tatooi Ro. 9

```
In [61]: df["Paying Customer"]=df["Paying Customer"].replace("Yes","y")
```

```
In [63]: df["Paying Customer"]=df["Paying Customer"].replace("No","N")
```

```
In [65]: df["Paying Customer"]
```

```
Out[65]: 0      y
1      N
2      N
3      y
4      Y
5      y
6      N
7      N
8      y
9      y
10     y
11     Y
12     y
13     y
14     N
15     N
16     y
17     Y
18    N/a
19     y
20     y
Name: Paying Customer, dtype: object
```

```
In [66]: df.columns
```

```
Out[66]: Index(['CustomerID', 'First_Name', 'Last_Name', 'Phone_Number',
               'Paying Customer', 'Do_Not_Contact', 'Street', 'City', 'ZipCode'],
              dtype='object')
```

```
In [69]: df["Do_Not_Contact"]=df["Do_Not_Contact"].replace("Yes","Y")
```

```
In [70]: df["Do_Not_Contact"]=df["Do_Not_Contact"].replace("No","N")
```



```
In [71]: df["Do_Not_Contact"]
```

```
Out[71]: 0      N
1      Y
2     NaN
3      Y
4      N
5      Y
6      N
7      N
8     NaN
9      N
10     N
11    NaN
12     N
13     N
14     N
15     N
16     N
17    NaN
18     Y
19     N
20     N
Name: Do_Not_Contact, dtype: object
```

In [72]: df

Out[72]:

	CustomerID	First_Name	Last_Name	Phone_Number	Paying Customer	Do_Not_Contact	Stre
0	1001	Frodo	Baggins	123-545-5421	y	N	1: Sh La
1	1002	Abed	Nadir	123-643-9775	N	Y	93 We Ma Stre
2	1003	Walter	White	706-695-0392	N	NaN	2: Dru Drivew
3	1004	Dwight	Schrute	123-543-2345	y	Y	9: Pap Aven
4	1005	Jon	Snow	876-678-3469	Y	N	1: Drago Ro.
5	1006	Ron	Swanson	304-762-2467	y	Y	768 C Parkw
6	1007	Jeff	Winger		N	N	12: Sou Stre
7	1008	Sherlock	Holmes	876-678-3469	N	N	98 Cl Dri
8	1009	Gandalf	NaN		y	NaN	1: Midc Eai
9	1010	Peter	Parker	123-545-5421	y	N	25 Ma Stre
10	1011	Samwise	Gamgee		y	N	6 Sh La
11	1012	Harry	Potter	706-695-0392	Y	NaN	23: Hogwa Aven
12	1013	Don	Draper	123-543-2345	y	N	20: Ma Stre
13	1014	Leslie	Knope	876-678-3469	y	N	343 C Parkw
14	1015	Toby	Flenderson	304-762-2467	N	N	214 H Aven
15	1016	Ron	Weasley	123-545-5421	N	N	23: Hogwa Aven
16	1017	Michael	Scott	123-643-9775	y	N	1: Pap Aven
17	1018	Clark	Kent	706-695-0392	Y	NaN	34: Sup La
18	1019	Creed	Braton		N/a	Y	N

	CustomerID	First_Name	Last_Name	Phone_Number	Paying Customer	Do_Not_Contact	Stre
19	1020	Anakin	Skywalker	876-678-3469	y	N	9 Tatooi Ro.
20	1020	Anakin	Skywalker	876-678-3469	y	N	9 Tatooi Ro.

In [74]:

Out[74]:

	CustomerID	First_Name	Last_Name	Phone_Number	Paying Customer	Do_Not_Contact	Street
3	1004	Dwight	Schrute	123-543-2345	y	Y	980 Paper Avenue

In [76]:

df.replace(np.nan, "", inplace=True)

In [77]:

df

Out[77]:

	CustomerID	First_Name	Last_Name	Phone_Number	Paying Customer	Do_Not_Contact	
0	1001	Frodo	Baggins	123-545-5421	y	N	
1	1002	Abed	Nadir	123-643-9775	N	Y	93
2	1003	Walter	White	706-695-0392	N		Dri
3	1004	Dwight	Schrute	123-543-2345	y	Y	A
4	1005	Jon	Snow	876-678-3469	Y	N	Dr

```
In [82]: for x in df.index:
         if df.loc[x,"Phone_Number"]=="":
             df.drop(x,inplace=True)
df
```

Out[82]:

	CustomerID	First_Name	Last_Name	Phone_Number	Paying Customer	Do_Not_Contact
0	1001	Frodo	Baggins	123-545-5421	y	N
1	1002	Abed	Nadir	123-643-9775	N	Y
2	1003	Walter	White	706-695-0392	N	
3	1004	Dwight	Schrute	123-543-2345	y	Y
4	1005	Jon	Snow	876-678-3469	Y	N

```
In [85]: for x in df.index:
        if df.loc[x,"Do_Not_Contact"]=="":
            df.drop(x,inplace=True)
df
```

Out[85]:

	CustomerID	First_Name	Last_Name	Phone_Number	Paying Customer	Do_Not_Contact	Stre
0	1001	Frodo	Baggins	123-545-5421	y	N	1: Sh La
1	1002	Abed	Nadir	123-643-9775	N	Y	93 We Ma Stre
3	1004	Dwight	Schrute	123-543-2345	y	Y	9: Pap Aven
4	1005	Jon	Snow	876-678-3469	Y	N	1: Drago Ro.
5	1006	Ron	Swanson	304-762-2467	y	Y	768 C Parkw
7	1008	Sherlock	Holmes	876-678-3469	N	N	98 Cl Dri
9	1010	Peter	Parker	123-545-5421	y	N	25 Ma Stre
12	1013	Don	Draper	123-543-2345	y	N	20 Ma Stre
13	1014	Leslie	Knope	876-678-3469	y	N	343 C Parkw
14	1015	Toby	Flenderson	304-762-2467	N	N	214 F Aven
15	1016	Ron	Weasley	123-545-5421	N	N	23 Hogwa Aven
16	1017	Michael	Scott	123-643-9775	y	N	1: Pap Aven
19	1020	Anakin	Skywalker	876-678-3469	y	N	9 Tatooi Ro.
20	1020	Anakin	Skywalker	876-678-3469	y	N	9 Tatooi Ro.



```
In [88]: df2=df[df.loc[:, "Do_Not_Contact"]=="N"]
```

```
In [91]: df2.reset_index(drop=True,inplace=True)
```

In [92]:

df

Out[92]:

	CustomerID	First_Name	Last_Name	Phone_Number	Paying Customer	Do_Not_Contact	Street
0	1001	Frodo	Baggins	123-545-5421	y	N	1. Sh La
1	1002	Abed	Nadir	123-643-9775	N	Y	93 We Ma Stre
3	1004	Dwight	Schrute	123-543-2345	y	Y	9. Pap Aven
4	1005	Jon	Snow	876-678-3469	Y	N	1. Drago Ro.
5	1006	Ron	Swanson	304-762-2467	y	Y	768 C Parkw
7	1008	Sherlock	Holmes	876-678-3469	N	N	98 Cl Dri
9	1010	Peter	Parker	123-545-5421	y	N	25 Ma Stre
12	1013	Don	Draper	123-543-2345	y	N	20 Ma Stre
13	1014	Leslie	Knope	876-678-3469	y	N	343 C Parkw
14	1015	Toby	Flenderson	304-762-2467	N	N	214 H Aven
15	1016	Ron	Weasley	123-545-5421	N	N	23 Hogwa Aven
16	1017	Michael	Scott	123-643-9775	y	N	1. Pap Aven
19	1020	Anakin	Skywalker	876-678-3469	y	N	9 Tatooi Ro.
20	1020	Anakin	Skywalker	876-678-3469	y	N	9 Tatooi Ro.

In [93]:

df2

Out[93]:

	CustomerID	First_Name	Last_Name	Phone_Number	Paying Customer	Do_Not_Contact	Street
0	1001	Frodo	Baggins	123-545-5421	y	N	1: Sh La
1	1005	Jon	Snow	876-678-3469	Y	N	1: Drago Ro
2	1008	Sherlock	Holmes	876-678-3469	N	N	98 Cl Dri
3	1010	Peter	Parker	123-545-5421	y	N	25 Mæ Stre
4	1013	Don	Draper	123-543-2345	y	N	20 Mæ Stre
5	1014	Leslie	Knope	876-678-3469	y	N	343 C Parkw
6	1015	Toby	Flenderson	304-762-2467	N	N	214 F Aven
7	1016	Ron	Weasley	123-545-5421	N	N	23 Hogwa Aven
8	1017	Michael	Scott	123-643-9775	y	N	1: Pap Aven
9	1020	Anakin	Skywalker	876-678-3469	y	N	9 Tatooi Ro
10	1020	Anakin	Skywalker	876-678-3469	y	N	9 Tatooi Ro

In []: