Concepts of Data Cleaning on Dataset Named " Customer Call List"

In [199]: # import the pandas library import pandas as pd In [200]: # call/ import the data table data = pd.read_excel('Customer Call List.xlsx') In [201]: data 0 1001 123 Shire Lane, Shire Frodo 123-545-5421 Yes No **Baggins** True 1 1002 Abed Nadir 123/643/9775 93 West Main Street No Yes False 2 1003 Walter /White 7066950392 298 Drugs Driveway Ν NaN True 980 Paper Avenue, Pennsylvania, 18503 3 1004 Dwight Schrute 123-543-2345 Yes Υ True 1005 876|678|3469 123 Dragons Road Υ Jon Snow No True 304-762-2467 5 1006 Ron Swanson 768 City Parkway Yes Yes True Winger 1007 Jeff NaN 1209 South Street False 6 No No 1008 876|678|3469 98 Clue Drive Sherlock Ν No False 7 Holmes 1009 8 Gandalf NaN N/a 123 Middle Earth NaN False Yes 1010 123-545-5421 25th Main Street, New York 9 Peter Parker Yes No True 10 1011 Samwise Gamgee NaN 612 Shire Lane, Shire Yes No True 11 1012 Harry ...Potter 7066950392 2394 Hogwarts Avenue Υ NaN True 12 1013 Draper 123-543-2345 2039 Main Street Yes Ν False In [202]: # Drop the duplicates data.drop_duplicates() Out[202]: CustomerID First_Name Last_Name Phone_Number Address Paying Customer Do_Not_Contact Not_Useful_Column 123 Shire Lane, Shire 0 1001 Frodo 123-545-5421 Yes No True Baggins 1002 123/643/9775 1 Abed Nadir 93 West Main Street No Yes False /White 7066950392 2 1003 Walter 298 Drugs Driveway Ν NaN True 123-543-2345 980 Paper Avenue, Pennsylvania, 18503 1004 Υ 3 Dwight Schrute Yes True 1005 Jon Snow 876 678 3469 123 Dragons Road Υ No True 1006 Ron Swanson 304-762-2467 768 City Parkway Yes Yes True 1007 Winger NaN 1209 South Street No False Jeff No 1008 876|678|3469 98 Clue Drive Ν 7 Sherlock Holmes No False 8 1009 Gandalf NaN N/a 123 Middle Earth Yes NaN False 9 1010 Peter Parker 123-545-5421 25th Main Street, New York No True Yes 10 1011 Samwise Gamgee NaN 612 Shire Lane, Shire Yes No True Υ 2394 Hogwarts Avenue True 11 1012 Harry ...Potter 7066950392 NaN 1013 12 Don Draper 123-543-2345 2039 Main Street Yes Ν False 13 1014 Leslie Knope 876|678|3469 343 City Parkway Yes Νo False 304-762-2467 14 1015 Flenderson_ 214 HR Avenue Ν No False Toby 123-545-5421 2395 Hogwarts Avenue 1016 Weasley False 15 Ron No Ν Michael 16 1017 Scott 123/643/9775 121 Paper Avenue, Pennsylvania Yes No False 17 1018 7066950392 3498 Super Lane Υ NaN Clark Kent True 18 1019 N/a N/a N/a Creed **Braton** Yes True

910 Tatooine Road, Tatooine

In [203]: # Remove the unnecessary special character from Last name
data['Last_Name']=data['Last_Name'].str.strip('123/..._')

Skywalker

876|678|3469

Anakin

19

1020

Ν

Yes

True

```
In [204]: # change the datatype of column to string
           data['Phone_Number'] = data['Phone_Number'].astype(str)
           # clear the format of phone no. but it has not worked
           data['Phone_Number'] = data['Phone_Number'].replace('1','')
           # or data['Phone_Number'] = data['Phone_Number'].astype(str).str.replace('[a-zA-Z0-9]','')
           data['Phone_Number']
Out[204]: 0
                  123-545-5421
                  123/643/9775
                    7066950392
                  123-543-2345
           3
           4
                  876 678 3469
           5
                  304-762-2467
           6
           7
                  876 678 3469
           8
                            N/a
           9
                  123-545-5421
           10
           11
                    7066950392
           12
                  123-543-2345
                  876 678 3469
           13
                  304-762-2467
           14
                  123-545-5421
           15
                  123/643/9775
           16
           17
                    7066950392
           18
                            N/a
           19
                  876 678 3469
                  876 678 3469
           20
           Name: Phone_Number, dtype: object
In [205]: | data['Phone_Number'].astype(str)
           # Now I have changed the method of doing it, like i utilized multiple replace with str and it worked.
           data['Phone_Number'] =data['Phone_Number'].str.replace('-','').str.replace('/','').str.replace('|','')
In [206]: data
                      IUUZ
                                           Nauii
                                                     1230439773
                                                                               93 VVESt IVIAIIT Street
                                                                                                                           162
                                                                                                                                            raise
                                Apeu
                                                                               298 Drugs Driveway
             2
                      1003
                                           White
                                                    7066950392
                                                                                                             Ν
                                                                                                                          NaN
                                Walter
                                                                                                                                            True
                                                     1235432345 980 Paper Avenue, Pennsylvania, 18503
                                                                                                                            Υ
                      1004
                               Dwight
                                          Schrute
                                                                                                            Yes
                                                                                                                                            True
                      1005
                                           Snow
                                                    8766783469
                                                                                123 Dragons Road
                                                                                                                           No
                                                                                                                                            True
                                  Jon
                      1006
                                        Swanson
                                                    3047622467
                                                                                 768 City Parkway
                                 Ron
                                                                                                            Yes
                                                                                                                           Yes
                                                                                                                                            True
                      1007
                                  Jeff
                                          Winger
                                                                                 1209 South Street
                                                                                                            Νo
                                                                                                                           No
                                                                                                                                            False
                                                           nan
             7
                      1008
                              Sherlock
                                                    8766783469
                                                                                    98 Clue Drive
                                                                                                             Ν
                                                                                                                                            False
                                          Holmes
                                                                                                                           No
                                                                                  123 Middle Earth
             8
                      1009
                              Gandalf
                                            NaN
                                                            Na
                                                                                                            Yes
                                                                                                                          NaN
                                                                                                                                            False
                                                                          25th Main Street, New York
             9
                      1010
                                Peter
                                          Parker
                                                     1235455421
                                                                                                            Yes
                                                                                                                           Νo
                                                                                                                                            True
            10
                      1011
                                                                              612 Shire Lane, Shire
                                                                                                            Yes
                                                                                                                           No
                                                                                                                                            True
                              Samwise
                                         Gamgee
            11
                      1012
                                                    7066950392
                                Harry
                                           Potter
                                                                             2394 Hogwarts Avenue
                                                                                                                          NaN
                                                                                                                                            True
            12
                      1013
                                                     1235432345
                                 Don
                                                                                  2039 Main Street
                                                                                                            Yes
                                                                                                                             Ν
                                                                                                                                            False
                                          Draper
            13
                      1014
                                Leslie
                                                    8766783469
                                                                                 343 City Parkway
                                                                                                                            No
                                                                                                                                            False
                                          Knope
                                                                                                            Yes
                                                    2047020407
                                                                                   044 110 4......
In [207]: # Change the values through replace function
           data[data['Paying Customer']=='Y']=data[data['Paying Customer']=='Y'].replace('Y','Yes')
           data[data['Paying Customer']=='N']=data[data['Paying Customer']=='N'].replace('N','No')
In [208]: | # Change the values through replace function
           data[data['Do_Not_Contact']=='Y']=data[data['Do_Not_Contact']=='Y'].replace('Y','Yes')
           data[data['Do_Not_Contact']=='N']=data[data['Do_Not_Contact']=='N'].replace('N','No')
```

In [209]: data

Out[209]:

	CustomerID	First_Name	Last_Name	Phone_Number	Address	Paying Customer	Do_Not_Contact	Not_Useful_Column
0	1001	Frodo	Baggins	1235455421	123 Shire Lane, Shire	Yes	No	True
1	1002	Abed	Nadir	1236439775	93 West Main Street	No	Yes	False
2	1003	Walter	White	7066950392	298 Drugs Driveway	No	NaN	True
3	1004	Dwight	Schrute	1235432345	980 Paper Avenue, Pennsylvania, 18503	Yes	Yes	True
4	1005	Jon	Snow	8766783469	123 Dragons Road	Yes	No	True
5	1006	Ron	Swanson	3047622467	768 City Parkway	Yes	Yes	True
6	1007	Jeff	Winger	nan	1209 South Street	No	No	False
7	1008	Sherlock	Holmes	8766783469	98 Clue Drive	No	No	False
8	1009	Ganda l f	NaN	Na	123 Middle Earth	Yes	NaN	False
9	1010	Peter	Parker	1235455421	25th Main Street, New York	Yes	No	True
10	1011	Samwise	Gamgee	nan	612 Shire Lane, Shire	Yes	No	True
11	1012	Harry	Potter	7066950392	2394 Hogwarts Avenue	Yes	NaN	True
12	1013	Don	Draper	1235432345	2039 Main Street	Yes	No	False
13	1014	Leslie	Knope	8766783469	343 City Parkway	Yes	No	False
14	1015	Toby	Flenderson	3047622467	214 HR Avenue	No	No	False
15	1016	Ron	Weasley	1235455421	2395 Hogwarts Avenue	No	No	False
16	1017	Michael	Scott	1236439775	121 Paper Avenue, Pennsylvania	Yes	No	False
17	1018	Clark	Kent	7066950392	3498 Super Lane	Yes	NaN	True
18	1019	Creed	Braton	Na	N/a	N/a	Yes	True
19	1020	Anakin	Skywa l ker	8766783469	910 Tatooine Road, Tatooine	Yes	No	True
20	1020	Anakin	Skywa l ker	8766783469	910 Tatooine Road, Tatooine	Yes	No	True

In [210]: # Rename the column, we can do with more than one at the same time by adding inside the curly brackets data=data.rename(columns={'Paying Customer':'Paying_Customer'})

In [211]: data

Out[211]:

	CustomerID	First_Name	Last_Name	Phone_Number	Address	Paying_Customer	Do_Not_Contact	Not_Useful_Column
0	1001	Frodo	Baggins	1235455421	123 Shire Lane, Shire	Yes	No	True
1	1002	Abed	Nadir	1236439775	93 West Main Street	No	Yes	False
2	1003	Walter	White	7066950392	298 Drugs Driveway	No	NaN	True
3	1004	Dwight	Schrute	1235432345	980 Paper Avenue, Pennsylvania, 18503	Yes	Yes	True
4	1005	Jon	Snow	8766783469	123 Dragons Road	Yes	No	True
5	1006	Ron	Swanson	3047622467	768 City Parkway	Yes	Yes	True
6	1007	Jeff	Winger	nan	1209 South Street	No	No	False
7	1008	Sherlock	Ho l mes	8766783469	98 Clue Drive	No	No	False
8	1009	Gandalf	NaN	Na	123 Middle Earth	Yes	NaN	False
9	1010	Peter	Parker	1235455421	25th Main Street, New York	Yes	No	True
10	1011	Samwise	Gamgee	nan	612 Shire Lane, Shire	Yes	No	True
11	1012	Harry	Potter	7066950392	2394 Hogwarts Avenue	Yes	NaN	True
12	1013	Don	Draper	1235432345	2039 Main Street	Yes	No	False
13	1014	Leslie	Knope	8766783469	343 City Parkway	Yes	No	False
14	1015	Toby	Flenderson	3047622467	214 HR Avenue	No	No	False
15	1016	Ron	Weasley	1235455421	2395 Hogwarts Avenue	No	No	False
16	1017	Michael	Scott	1236439775	121 Paper Avenue, Pennsylvania	Yes	No	False
17	1018	Clark	Kent	7066950392	3498 Super Lane	Yes	NaN	True
18	1019	Creed	Braton	Na	N/a	N/a	Yes	True
19	1020	Anakin	Skywa l ker	8766783469	910 Tatooine Road, Tatooine	Yes	No	True
20	1020	Anakin	Skywa l ker	8766783469	910 Tatooine Road, Tatooine	Yes	No	True

```
In [212]: # Change the data type of of phone no. through Lambda function
data['Phone_Number']=data['Phone_Number'].apply(lambda x:str(x))
```

In [213]: # Change the format of phone no. through lambda function with apply function
data['Phone_Number']=data['Phone_Number'].apply(lambda x:x[0:3] + '-'+ x[3:6]+'-'+x[6:10])

In [214]: data

Out[214]:

	CustomerID	First_Name	Last_Name	Phone_Number	Address	Paying_Customer	Do_Not_Contact	Not_Useful_Column
0	1001	Frodo	Baggins	123-545-5421	123 Shire Lane, Shire	Yes	No	True
1	1002	Abed	Nadir	123-643-9775	93 West Main Street	No	Yes	False
2	1003	Walter	White	706-695-0392	298 Drugs Driveway	No	NaN	True
3	1004	Dwight	Schrute	123-543-2345	980 Paper Avenue, Pennsylvania, 18503	Yes	Yes	True
4	1005	Jon	Snow	876-678-3469	123 Dragons Road	Yes	No	True
5	1006	Ron	Swanson	304-762-2467	768 City Parkway	Yes	Yes	True
6	1007	Jeff	Winger	nan	1209 South Street	No	No	False
7	1008	Sherlock	Holmes	876-678-3469	98 Clue Drive	No	No	False
8	1009	Ganda l f	NaN	Na	123 Middle Earth	Yes	NaN	False
9	1010	Peter	Parker	123-545-5421	25th Main Street, New York	Yes	No	True
10	1011	Samwise	Gamgee	nan	612 Shire Lane, Shire	Yes	No	True
11	1012	Harry	Potter	706-695-0392	2394 Hogwarts Avenue	Yes	NaN	True
12	1013	Don	Draper	123-543-2345	2039 Main Street	Yes	No	False
13	1014	Leslie	Knope	876-678-3469	343 City Parkway	Yes	No	False
14	1015	Toby	Flenderson	304-762-2467	214 HR Avenue	No	No	False
15	1016	Ron	Weasley	123-545-5421	2395 Hogwarts Avenue	No	No	False
16	1017	Michael	Scott	123-643-9775	121 Paper Avenue, Pennsylvania	Yes	No	False
17	1018	Clark	Kent	706-695-0392	3498 Super Lane	Yes	NaN	True
18	1019	Creed	Braton	Na	N/a	N/a	Yes	True
19	1020	Anakin	Skywalker	876-678-3469	910 Tatooine Road, Tatooine	Yes	No	True
20	1020	Anakin	Skywalker	876-678-3469	910 Tatooine Road, Tatooine	Yes	No	True

```
In [215]: # Change the string value into blank.
data['Phone_Number'] = data['Phone_Number'].str.replace('nan--','')
data['Phone_Number'] = data['Phone_Number'].str.replace('Na--','')
```

In [216]: data

Out[216]:

	CustomerID	First_Name	Last_Name	Phone_Number	Address	Paying_Customer	Do_Not_Contact	Not_Useful_Column
0	1001	Frodo	Baggins	123-545-5421	123 Shire Lane, Shire	Yes	No	True
1	1002	Abed	Nadir	123-643-9775	93 West Main Street	No	Yes	False
2	1003	Walter	White	706-695-0392	298 Drugs Driveway	No	NaN	True
3	1004	Dwight	Schrute	123-543-2345	980 Paper Avenue, Pennsylvania, 18503	Yes	Yes	True
4	1005	Jon	Snow	876-678-3469	123 Dragons Road	Yes	No	True
5	1006	Ron	Swanson	304-762-2467	768 City Parkway	Yes	Yes	True
6	1007	Jeff	Winger		1209 South Street	No	No	False
7	1008	Sherlock	Ho l mes	876-678-3469	98 Clue Drive	No	No	False
8	1009	Ganda l f	NaN		123 Middle Earth	Yes	NaN	False
9	1010	Peter	Parker	123-545-5421	25th Main Street, New York	Yes	No	True
10	1011	Samwise	Gamgee		612 Shire Lane, Shire	Yes	No	True
11	1012	Harry	Potter	706-695-0392	2394 Hogwarts Avenue	Yes	NaN	True
12	1013	Don	Draper	123-543-2345	2039 Main Street	Yes	No	False
13	1014	Leslie	Knope	876-678-3469	343 City Parkway	Yes	No	False
14	1015	Toby	Flenderson	304-762-2467	214 HR Avenue	No	No	False
15	1016	Ron	Weasley	123-545-5421	2395 Hogwarts Avenue	No	No	False
16	1017	Michael	Scott	123-643-9775	121 Paper Avenue, Pennsylvania	Yes	No	False
17	1018	Clark	Kent	706-695-0392	3498 Super Lane	Yes	NaN	True
18	1019	Creed	Braton		N/a	N/a	Yes	True
19	1020	Anakin	Skywalker	876-678-3469	910 Tatooine Road, Tatooine	Yes	No	True
20	1020	Anakin	Skywalker	876-678-3469	910 Tatooine Road, Tatooine	Yes	No	True

```
In [217]: # Split the column into 3 other columns
data[['Street_Address','State','Postal_Code']]=data['Address'].str.split(',',expand=True)
```

In [218]: data

Out[218]:

State	Street_Address	Not_Useful_Column	Do_Not_Contact	Paying_Customer	Address	Phone_Number	Last_Name	First_Name	CustomerID	
Shire	123 Shire Lane	True	No	Yes	123 Shire Lane, Shire	123-545-5421	Baggins	Frodo	1001	0
None	93 West Main Street	False	Yes	No	93 West Main Street	123-643-9775	Nadir	Abed	1002	1
None	298 Drugs Driveway	True	NaN	No	298 Drugs Driveway	706-695-0392	White	Walter	1003	2
Pennsylvania	980 Paper Avenue	True	Yes	Yes	980 Paper Avenue, Pennsylvania, 18503	123-543-2345	Schrute	Dwight	1004	3
None	123 Dragons Road	True	No	Yes	123 Dragons Road	876-678-3469	Snow	Jon	1005	4
None	768 City Parkway	True	Yes	Yes	768 City Parkway	304-762-2467	Swanson	Ron	1006	5
None	1209 South Street	Fa l se	No	No	1209 South Street		Winger	Jeff	1007	6
None	98 Clue Drive	False	No	No	98 Clue Drive	876-678-3469	Holmes	Sherlock	1008	7
None	123 Midd l e Earth	False	NaN	Yes	123 Middle Earth		NaN	Gandalf	1009	8
New York	25th Main Street	True	No	Yes	25th Main Street, New York	123-545-5421	Parker	Peter	1010	9
Shire	612 Shire Lane	True	No	Yes	612 Shire Lane, Shire		Gamgee	Samwise	1011	10
None	2394 Hogwarts Avenue	True	NaN	Yes	2394 Hogwarts Avenue	706-695-0392	Potter	Harry	1012	11
None	2039 Main Street	False	No	Yes	2039 Main Street	123-543-2345	Draper	Don	1013	12
None	343 City Parkway	False	No	Yes	343 City Parkway	876-678-3469	Knope	Leslie	1014	13
None	214 HR Avenue	False	No	No	214 HR Avenue	304-762-2467	Flenderson	Toby	1015	14
None	2395 Hogwarts Avenue	False	No	No	2395 Hogwarts Avenue	123-545-5421	Weasley	Ron	1016	15
Pennsylvania	121 Paper Avenue	False	No	Yes	121 Paper Avenue, Pennsylvania	123-643-9775	Scott	Michael	1017	16
None	3498 Super Lane	True	NaN	Yes	3498 Super Lane	706-695-0392	Kent	Clark	1018	17
None	N/a	True	Yes	N/a	N/a		Braton	Creed	1019	18
Tatooine	910 Tatooine Road	True	No	Yes	910 Tatooine Road, Tatooine	876-678-3469	Skywa l ker	Anakin	1020	19
Tatooine	910 Tatooine Road	True	No	Yes	910 Tatooine Road, Tatooine	876-678-3469	Skywa l ker	Anakin	1020	20
•										

In [219]: da

data = data.replace('N/a','') # It is working because Na means Not Available, that means it is a string which can be replace wi data.replace('NaN','') # It is not working because NaN means there has a emply place or (Not a Number), so obviously it will n

4										>
14	1015	Toby	Flenderson	304-762-2467	214 HR Avenue	No	No	False	214 HR Avenue	Noi 🌥
15	1016	Ron	Weasley	123-545-5421	2395 Hogwarts Avenue	No	No	False	2395 Hogwarts Avenue	Noi
16	1017	Michae l	Scott	123-643-9775	121 Paper Avenue, Pennsylvania	Yes	No	False	121 Paper Avenue	Pennsylvar
17	1018	Clark	Kent	706-695-0392	3498 Super Lane	Yes	NaN	True	3498 Super Lane	Noi
18	1019	Creed	Braton				Yes	True		Noi
19	1020	Anakin	Skywalker	876-678-3469	910 Tatooine Road, Tatooine	Yes	No	True	910 Tatooine Road	Tatooii
20	1020	Anakin	Skywalker	876-678-3469	910 Tatooine Road, Tatooine	Yes	No	True	910 Tatooine Road	Tatooiı

```
In [220]: # Now the fillna come to rescue by creating a blank in all those place where NaN placed.
           data = data.fillna('')
           data
                 IUUZ
                           Apeu
                                      ıvauıı
                                               123-043-3113
                                                                                                                      гаізе
                                                                  Street
                                                                                                                                    Street
                                                               298 Drugs
                                                                                                                                 298 Drugs
       2
                1003
                          Walter
                                      White
                                               706-695-0392
                                                                                     No
                                                                                                                      True
                                                               Driveway
                                                                                                                                  Driveway
                                                               980 Paper
                                                                Avenue,
                                                                                                                                 980 Paper
       3
                                     Schrute
                1004
                          Dwight
                                               123-543-2345
                                                                                     Yes
                                                                                                     Yes
                                                                                                                       True
                                                                                                                                           Pennsylvania
                                                           Pennsylvania,
                                                                                                                                   Avenue
                                                                  18503
                                                             123 Dragons
                                                                                                                               123 Dragons
                1005
                                               876-678-3469
                            Jon
                                      Snow
                                                                                     Yes
                                                                                                     No
                                                                                                                      True
                                                                                                                                     Road
                                                                  Road
                                                                768 City
                                                                                                                                   768 City
                1006
                            Ron
                                   Swanson
                                               304-762-2467
                                                                                     Yes
                                                                                                     Yes
                                                                                                                      True
                                                                Parkway
                                                                                                                                  Parkway
                                                              1209 South
                                                                                                                                1209 South
                                     Winger
                1007
                             Jeff
                                                                                     No
                                                                                                                      False
                                                                                                     No
                                                                  Street
                                                                                                                                    Street
                                                                                                                               98 Clue Drive
                1008
                        Sherlock
                                               876-678-3469 98 Clue Drive
                                                                                                     No
                                                                                                                      False
                                     Holmes
                                                                                     No
                                                                                                                                123 Middle
                                                              123 Middle
                1009
                         Gandalf
                                                                                     Yes
                                                                                                                      False
                                                                  Earth
                                                                                                                                     Earth
In [221]: #Now we want to place Yes Where we have blanks in Do_Not_Contact, because we assume that those who did not mention will pick th
            for x in data.index:
                if data.loc[x, 'Do_Not_Contact'] == '':
                     data.at[x, 'Do_Not_Contact'] = 'No Feedback'
```

```
In [222]: # Drop a column that has no feasibility
          data= data.drop('Not_Useful_Column',axis=1)
```

```
In [226]: for x in data.index:
              if data.loc[x, 'Phone_Number'] == 'Not mentioned':
                  data.at[x, 'Phone_Number'] = '000-000-0000'
```

In [227]: data

Out[227]:

	CustomerID	First_Name	Last_Name	Phone_Number	Address	Paying_Customer	Do_Not_Contact	Street_Address	State	Postal_	_Code
0	1001	Frodo	Baggins	123-545-5421	123 Shire Lane, Shire	Yes	No	123 Shire Lane	Shire		
1	1002	Abed	Nadir	123-643-9775	93 West Main Street	No	Yes	93 West Main Street			
2	1003	Walter	White	706-695-0392	298 Drugs Driveway	No	No Feedback	298 Drugs Driveway			
3	1004	Dwight	Schrute	123-543-2345	980 Paper Avenue, Pennsylvania, 18503	Yes	Yes	980 Paper Avenue	Pennsylvania		18503
4	1005	Jon	Snow	876-678-3469	123 Dragons Road	Yes	No	123 Dragons Road			
5	1006	Ron	Swanson	304-762-2467	768 City Parkway	Yes	Yes	768 City Parkway			
6	1007	Jeff	Winaer	000-000-0000	1209 South	No	No	1209 South			