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	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	7066950392	298 Drugs Driveway	N	NaN	True
3	1004	Dwight	Schrute	123-543-2345	980 Paper Avenue, Pennsylvania, 18503	Yes	Y	True
4	1005	Jon	Snow	876|678|3469	123 Dragons Road	Y	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	N	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	Yes	No	True
10	1011	Samwise	Gamgee	NaN	612 Shire Lane, Shire	Yes	No	True
11	1012	Harry	...Potter	7066950392	2394 Hogwarts Avenue	Y	NaN	True
12	1013	Don	Draper	123-543-2345	2039 Main Street	Yes	N	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
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	7066950392	298 Drugs Driveway	N	NaN	True
3	1004	Dwight	Schrute	123-543-2345	980 Paper Avenue, Pennsylvania, 18503	Yes	Y	True
4	1005	Jon	Snow	876|678|3469	123 Dragons Road	Y	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	N	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	Yes	No	True
10	1011	Samwise	Gamgee	NaN	612 Shire Lane, Shire	Yes	No	True
11	1012	Harry	...Potter	7066950392	2394 Hogwarts Avenue	Y	NaN	True
12	1013	Don	Draper	123-543-2345	2039 Main Street	Yes	N	False
13	1014	Leslie	Knope	876|678|3469	343 City Parkway	Yes	No	False
14	1015	Toby	Flenderson_	304-762-2467	214 HR Avenue	N	No	False
15	1016	Ron	Weasley	123-545-5421	2395 Hogwarts Avenue	No	N	False
16	1017	Michael	Scott	123/643/9775	121 Paper Avenue, Pennsylvania	Yes	No	False
17	1018	Clark	Kent	7066950392	3498 Super Lane	Y	NaN	True
18	1019	Creed	Braton	N/a	N/a	N/a	Yes	True
19	1020	Anakin	Skywalker	876|678|3469	910 Tatooine Road, Tatooine	Yes	N	True

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

```
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
          1      123/643/9775
          2        7066950392
          3      123-543-2345
          4      876|678|3469
          5      304-762-2467
          6               nan
          7      876|678|3469
          8               N/a
          9      123-545-5421
          10              nan
          11       7066950392
          12     123-543-2345
          13     876|678|3469
          14     304-762-2467
          15     123-545-5421
          16     123/643/9775
          17       7066950392
          18              N/a
          19     876|678|3469
          20     876|678|3469
          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
```

1	1002	Abed	Nadir	1236439775	93 West Main Street	No	Yes	False
2	1003	Walter	White	7066950392	298 Drugs Driveway	N	NaN	True
3	1004	Dwight	Schrute	1235432345	980 Paper Avenue, Pennsylvania, 18503	Yes	Y	True
4	1005	Jon	Snow	8766783469	123 Dragons Road	Y	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	N	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	Y	NaN	True
12	1013	Don	Draper	1235432345	2039 Main Street	Yes	N	False
13	1014	Leslie	Knope	8766783469	343 City Parkway	Yes	No	False
14	1015	Toby	Flenderson	3047622467	314 HR Avenue	N	No	False

```
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	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	Skywalker	8766783469	910 Tatooine Road, Tatooine	Yes	No	True
20	1020	Anakin	Skywalker	8766783469	910 Tatooine Road, Tatooine	Yes	No	True

In [210]:
```python
# 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	Holmes	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	Skywalker	8766783469	910 Tatooine Road, Tatooine	Yes	No	True
20	1020	Anakin	Skywalker	8766783469	910 Tatooine Road, Tatooine	Yes	No	True

In [212]:
```python
# Change the data type of of phone no. through lambda function

data['Phone_Number']=data['Phone_Number'].apply(lambda x:str(x))
```

In [213]:
```python
# 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	Gandalf	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]:
```python
# 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	Holmes	876-678-3469	98 Clue Drive	No	No	False
8	1009	Gandalf	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]:
```python
# Split the column into 3 other columns
data[['Street_Address','State','Postal_Code']]=data['Address'].str.split(',',expand=True)
```

In [218]: `data`

Out[218]:

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

In [219]:
```python
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
```

14	1015	Toby	Flenderson	304-762-2467	214 HR Avenue	No	No	False	214 HR Avenue	Nor
15	1016	Ron	Weasley	123-545-5421	2395 Hogwarts Avenue	No	No	False	2395 Hogwarts Avenue	Nor
16	1017	Michael	Scott	123-643-9775	121 Paper Avenue, Pennsylvania	Yes	No	False	121 Paper Avenue	Pennsylvan
17	1018	Clark	Kent	706-695-0392	3498 Super Lane	Yes	NaN	True	3498 Super Lane	Nor
18	1019	Creed	Braton				Yes	True		Nor
19	1020	Anakin	Skywalker	876-678-3469	910 Tatooine Road, Tatooine	Yes	No	True	910 Tatooine Road	Tatooi
20	1020	Anakin	Skywalker	876-678-3469	910 Tatooine Road, Tatooine	Yes	No	True	910 Tatooine Road	Tatooi

In [220]:
```python
# Now the fillna come to rescue by creating a blank in all those place where NaN placed.
data = data.fillna('')
data
```

1	1002	Abed	Nadir	123-643-9775	Street	No	Yes	False	Street	
2	1003	Walter	White	706-695-0392	298 Drugs Driveway	No		True	298 Drugs Driveway	
3	1004	Dwight	Schrute	123-543-2345	980 Paper Avenue, Pennsylvania, 18503	Yes	Yes	True	980 Paper Avenue	Pennsylvania
4	1005	Jon	Snow	876-678-3469	123 Dragons Road	Yes	No	True	123 Dragons Road	
5	1006	Ron	Swanson	304-762-2467	768 City Parkway	Yes	Yes	True	768 City Parkway	
6	1007	Jeff	Winger		1209 South Street	No	No	False	1209 South Street	
7	1008	Sherlock	Holmes	876-678-3469	98 Clue Drive	No	No	False	98 Clue Drive	
8	1009	Gandalf			123 Middle Earth	Yes		False	123 Middle Earth	

In [221]:
```python
#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]:
```python
# Drop a column that has no feasibility

data= data.drop('Not_Useful_Column',axis=1)
```

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

In [227]:
```python
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	Winger	000-000-0000	1209 South Street	No	No	1209 South		