

In [1]:

```
import pandas as pd
```

In [2]:

```
customer_data = pd.read_excel("Customer Call List.xlsx")
customer_data
```

Out[2]:

	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
20	1020	Anakin	Skywalker	876 678 3469	910 Tatooine Road, Tatooine	Yes	N	True

In [3]:

```
customer_data.shape
```

Out[3]:

```
(21, 8)
```

In [4]:

```
customer_data.columns
```

Out[4]:

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

In [5]:

```
customer_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 21 entries, 0 to 20  
Data columns (total 8 columns):  
#   Column                Non-Null Count  Dtype  
---  -  
0   CustomerID            21 non-null    int64  
1   First_Name            21 non-null    object  
2   Last_Name             20 non-null    object  
3   Phone_Number          19 non-null    object  
4   Address               21 non-null    object  
5   Paying Customer       21 non-null    object  
6   Do_Not_Contact        17 non-null    object  
7   Not_Useful_Column     21 non-null    bool  
dtypes: bool(1), int64(1), object(6)  
memory usage: 1.3+ KB
```

In [6]:

```
customer_data.isnull().sum()
```

Out[6]:

```
CustomerID            0  
First_Name           0  
Last_Name            1  
Phone_Number         2  
Address              0  
Paying Customer      0  
Do_Not_Contact       4  
Not_Useful_Column    0  
dtype: int64
```

In [7]:

```
customer_data.isnull()
```

Out[7]:

	CustomerID	First_Name	Last_Name	Phone_Number	Address	Paying Customer	Do_Not_Contact	Not_Useful_Column
0	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	True	False
3	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False
5	False	False	False	False	False	False	False	False
6	False	False	False	True	False	False	False	False

7	CustomerID	First_Name	Last_Name	Phone_Number	Address	Paying_Customer	Do_Not_Contact	Not_Useful_Column
8	False	False	True	False	False	False	True	False
9	False	False	False	False	False	False	False	False
10	False	False	False	True	False	False	False	False
11	False	False	False	False	False	False	True	False
12	False	False	False	False	False	False	False	False
13	False	False	False	False	False	False	False	False
14	False	False	False	False	False	False	False	False
15	False	False	False	False	False	False	False	False
16	False	False	False	False	False	False	False	False
17	False	False	False	False	False	False	True	False
18	False	False	False	False	False	False	False	False
19	False	False	False	False	False	False	False	False
20	False	False	False	False	False	False	False	False

In [8]:

```
customer_data.dropna(inplace=True)
```

In [9]:

```
customer_data.shape
```

Out[9]:

(15, 8)

In [10]:

```
customer_data.Last_Name.unique()
```

Out[10]:

```
array(['Baggins', 'Nadir', 'Schrute', 'Snow', 'Swanson', 'Holmes',  
      'Parker', 'Draper', 'Knope', 'Flenderson_', 'Weasley', 'Scott',  
      'Braton', 'Skywalker'], dtype=object)
```

In [11]:

```
customer_data.Last_Name = customer_data.Last_Name.str.replace("_", "")
```

In [12]:

```
customer_data
```

Out[12]:

	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
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 Parkwav	Yes	Yes	True

7	CustomerID	First_Name	Last_Name	Phone_Number	Address	Paying_Customer	Do_Not_Contact	Not_Useful_Columns
9	1010	Peter	Parker	123-545-5421	25th Main Street, New York	Yes	No	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
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
20	1020	Anakin	Skywalker	876 678 3469	910 Tatooine Road, Tatooine	Yes	N	True

In [13]:

```
customer_data.drop(18,inplace=True)
```

In [14]:

```
customer_data.Phone_Number.unique()
```

Out[14]:

```
array(['123-545-5421', '123/643/9775', '123-543-2345', '876|678|3469',
      '304-762-2467'], dtype=object)
```

In [15]:

```
customer_data.Phone_Number = customer_data.Phone_Number.str.replace("/", "").str.replace(
"|", "").str.replace("-", "").str.replace("Na--", "N/a")
customer_data.Phone_Number.astype("string")
customer_data.Phone_Number = customer_data.Phone_Number.apply(lambda x: x[:3] + "-" + x[
3:6] + "-" + x[6:])
```

In [16]:

```
customer_data["Phone_Number"]
```

Out[16]:

```
0      123-545-5421
1      123-643-9775
3      123-543-2345
4      876-678-3469
5      304-762-2467
7      876-678-3469
9      123-545-5421
12     123-543-2345
13     876-678-3469
14     304-762-2467
15     123-545-5421
16     123-643-9775
19     876-678-3469
20     876-678-3469
Name: Phone_Number, dtype: object
```

In [17]:

```
customer_data.duplicated().sum()
```

Out[17]:

1

In [18]:

```
customer_data.drop_duplicates(inplace=True)
```

In [19]:

```
customer_data
```

Out[19]:

	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
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
7	1008	Sherlock	Holmes	876-678-3469	98 Clue Drive	N	No	False
9	1010	Peter	Parker	123-545-5421	25th Main Street, New York	Yes	No	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
19	1020	Anakin	Skywalker	876-678-3469	910 Tatooine Road, Tatooine	Yes	N	True

In [20]:

```
customer_data["Paying Customer"] = customer_data["Paying Customer"].str.replace("Yes", "Y")\
).str.replace("No", "N")
```

In [21]:

```
customer_data["Do_Not_Contact"] = customer_data["Do_Not_Contact"].str.replace("Yes", "Y")\
.str.replace("No", "N")
```

In [22]:

```
customer_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 13 entries, 0 to 19
Data columns (total 8 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   CustomerID      13 non-null    int64
1   First_Name      13 non-null    object
2   Last_Name       13 non-null    object
3   Phone_Number    13 non-null    object
4   Address         13 non-null    object
5   Paying Customer  13 non-null    object
6   Do_Not_Contact  13 non-null    object
7   Not_Useful_Column 13 non-null    object
```

```
3 Phone_Number      13 non-null    object
4 Address           13 non-null    object
5 Paying Customer   13 non-null    object
6 Do_Not_Contact    13 non-null    object
7 Not_Useful_Column 13 non-null    bool
dtypes: bool(1), int64(1), object(6)
memory usage: 845.0+ bytes
```

In [23]:

```
customer_data.columns
```

Out[23]:

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

In [26]:

```
customer_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 13 entries, 0 to 19
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype
---  -
0   CustomerID            13 non-null    int64
1   First_Name            13 non-null    object
2   Last_Name             13 non-null    object
3   Phone_Number          13 non-null    object
4   Address               13 non-null    object
5   Paying Customer       13 non-null    object
6   Do_Not_Contact        13 non-null    object
7   Not_Useful_Column     13 non-null    bool
dtypes: bool(1), int64(1), object(6)
memory usage: 845.0+ bytes
```

In [27]:

```
customer_data.isnull().sum()
```

Out[27]:

```
CustomerID      0
First_Name      0
Last_Name       0
Phone_Number    0
Address         0
Paying Customer 0
Do_Not_Contact  0
Not_Useful_Column 0
dtype: int64
```

In [28]:

```
customer_data[customer_data.duplicated()]
```

Out[28]:

CustomerID	First_Name	Last_Name	Phone_Number	Address	Paying Customer	Do_Not_Contact	Not_Useful_Column
------------	------------	-----------	--------------	---------	-----------------	----------------	-------------------

In [29]:

```
customer_data
```

Out[29]:

CustomerID	First_Name	Last_Name	Phone_Number	Address	Paying Customer	Do_Not_Contact	Not_Useful_Column	
0	1001	Freda	Bergine	123 545 5404	123 Shire Lane,	v	N	True

0	1001	Frodo	Baggins	123-543-5421	Shire Address 93 West Main Street	Y	N	True
CustomerID	First_Name	Last_Name	Phone_Number		Paying Customer	Do_Not_Contact	Not_Useful_Column	
1	1002	Abed	Nadir	123-643-9775		N	Y	False
3	1004	Dwight	Schrute	123-543-2345	980 Paper Avenue, Pennsylvania, 18503	Y	Y	True
4	1005	Jon	Snow	876-678-3469	123 Dragons Road	Y	N	True
5	1006	Ron	Swanson	304-762-2467	768 City Parkway	Y	Y	True
7	1008	Sherlock	Holmes	876-678-3469	98 Clue Drive	N	N	False
9	1010	Peter	Parker	123-545-5421	25th Main Street, New York	Y	N	True
12	1013	Don	Draper	123-543-2345	2039 Main Street	Y	N	False
13	1014	Leslie	Knope	876-678-3469	343 City Parkway	Y	N	False
14	1015	Toby	Flenderson	304-762-2467	214 HR Avenue	N	N	False
15	1016	Ron	Weasley	123-545-5421	2395 Hogwarts Avenue	N	N	False
16	1017	Michael	Scott	123-643-9775	121 Paper Avenue, Pennsylvania	Y	N	False
19	1020	Anakin	Skywalker	876-678-3469	910 Tatooine Road, Tatooine	Y	N	True

In [31]:

```
customer_data.shape
```

Out[31]:

(13, 8)

In [32]:

```
customer_data.to_excel("Cleaned_Call_List.xlsx")
```

In []: