Data Cleaning in Python

```
import pandas as pd
import numpy as np

df = pd.read_excel(r"C:\Users\User\Downloads\Customer Call List.xlsx")
df
```

```
    CustomerID First_Name    Last_Name  Phone_Number  \
0         1001      Frodo      Baggins  123-545-5421
1         1002       Abed        Nadir  123/643/9775
2         1003     Walter       /White    7066950392
3         1004     Dwight      Schrute  123-543-2345
4         1005        Jon         Snow  876|678|3469
5         1006        Ron      Swanson  304-762-2467
6         1007       Jeff       Winger           NaN
7         1008   Sherlock       Holmes  876|678|3469
8         1009    Gandalf          NaN           N/a
9         1010      Peter       Parker  123-545-5421
10        1011    Samwise       Gamgee           NaN
11        1012      Harry    ...Potter    7066950392
12        1013        Don       Draper  123-543-2345
13        1014     Leslie        Knope  876|678|3469
14        1015       Toby  Flenderson_  304-762-2467
15        1016        Ron      Weasley  123-545-5421
16        1017   Michael         Scott  123/643/9775
17        1018      Clark         Kent    7066950392
18        1019      Creed       Braton           N/a
19        1020     Anakin    Skywalker  876|678|3469
20        1020     Anakin    Skywalker  876|678|3469

                                  Address Paying Customer
Do_Not_Contact  \
0                    123 Shire Lane, Shire             Yes
No
1                      93 West Main Street              No
Yes
2                       298 Drugs Driveway               N
NaN
3   980 Paper Avenue, Pennsylvania, 18503             Yes
Y
4                         123 Dragons Road               Y
No
5                         768 City Parkway             Yes
Yes
6                        1209 South Street              No
No
7                            98 Clue Drive               N
```

```
No
8                          123 Middle Earth             Yes
NaN
9             25th Main Street, New York             Yes
No
10                   612 Shire Lane, Shire             Yes
No
11                    2394 Hogwarts Avenue               Y
NaN
12                        2039 Main Street             Yes
N
13                        343 City Parkway             Yes
No
14                           214 HR Avenue               N
No
15                    2395 Hogwarts Avenue              No
N
16        121 Paper Avenue, Pennsylvania             Yes
No
17                         3498 Super Lane               Y
NaN
18                                     N/a             N/a
Yes
19          910 Tatooine Road, Tatooine             Yes
N
20          910 Tatooine Road, Tatooine             Yes
N

    Not_Useful_Column
0                True
1               False
2                True
3                True
4                True
5                True
6               False
7               False
8               False
9                True
10               True
11               True
12              False
13              False
14              False
15              False
16              False
17               True
18               True
```

```
19              True
20              True

#Remove the duplicates
df=df.drop_duplicates()
df

    CustomerID First_Name    Last_Name  Phone_Number  \
0         1001      Frodo      Baggins  123-545-5421
1         1002       Abed        Nadir  123/643/9775
2         1003     Walter       /White    7066950392
3         1004     Dwight      Schrute  123-543-2345
4         1005        Jon         Snow  876|678|3469
5         1006        Ron      Swanson  304-762-2467
6         1007       Jeff       Winger           NaN
7         1008   Sherlock       Holmes  876|678|3469
8         1009    Gandalf          NaN           N/a
9         1010      Peter       Parker  123-545-5421
10        1011    Samwise       Gamgee           NaN
11        1012      Harry    ...Potter    7066950392
12        1013        Don       Draper  123-543-2345
13        1014     Leslie        Knope  876|678|3469
14        1015       Toby  Flenderson_  304-762-2467
15        1016        Ron      Weasley  123-545-5421
16        1017   Michael         Scott  123/643/9775
17        1018      Clark         Kent    7066950392
18        1019      Creed       Braton           N/a
19        1020     Anakin    Skywalker  876|678|3469

                                  Address Paying Customer
Do_Not_Contact  \
0                     123 Shire Lane, Shire             Yes
No
1                      93 West Main Street              No
Yes
2                       298 Drugs Driveway               N
NaN
3   980 Paper Avenue, Pennsylvania, 18503             Yes
Y
4                         123 Dragons Road               Y
No
5                         768 City Parkway             Yes
Yes
6                        1209 South Street              No
No
7                            98 Clue Drive               N
No
8                        123 Middle Earth             Yes
NaN
9                25th Main Street, New York             Yes
```

```
No
10                    612 Shire Lane, Shire                Yes
No
11                     2394 Hogwarts Avenue                  Y
NaN
12                         2039 Main Street                Yes
N
13                         343 City Parkway                Yes
No
14                            214 HR Avenue                  N
No
15                     2395 Hogwarts Avenue                 No
N
16          121 Paper Avenue, Pennsylvania                 Yes
No
17                          3498 Super Lane                  Y
NaN
18                                      N/a                N/a
Yes
19             910 Tatooine Road, Tatooine                 Yes
N

    Not_Useful_Column
0                True
1               False
2                True
3                True
4                True
5                True
6               False
7               False
8               False
9                True
10               True
11               True
12              False
13              False
14              False
15              False
16              False
17               True
18               True
19               True

#Remove the columns which does not need
df=df.drop(columns="Not_Useful_Column")
df

    CustomerID First_Name    Last_Name  Phone_Number  \
0         1001      Frodo      Baggins  123-545-5421
```

```
1         1002       Abed        Nadir  123/643/9775
2         1003     Walter       /White    7066950392
3         1004     Dwight      Schrute  123-543-2345
4         1005        Jon         Snow  876|678|3469
5         1006        Ron      Swanson  304-762-2467
6         1007       Jeff       Winger           NaN
7         1008   Sherlock       Holmes  876|678|3469
8         1009    Gandalf          NaN           N/a
9         1010      Peter       Parker  123-545-5421
10        1011    Samwise       Gamgee           NaN
11        1012      Harry    ...Potter    7066950392
12        1013        Don       Draper  123-543-2345
13        1014     Leslie        Knope  876|678|3469
14        1015       Toby  Flenderson_  304-762-2467
15        1016        Ron      Weasley  123-545-5421
16        1017   Michael         Scott  123/643/9775
17        1018      Clark         Kent    7066950392
18        1019      Creed       Braton           N/a
19        1020     Anakin    Skywalker  876|678|3469

                                  Address Paying Customer
Do_Not_Contact
0                     123 Shire Lane, Shire             Yes
No
1                      93 West Main Street              No
Yes
2                        298 Drugs Driveway               N
NaN
3   980 Paper Avenue, Pennsylvania, 18503             Yes
Y
4                         123 Dragons Road               Y
No
5                         768 City Parkway             Yes
Yes
6                        1209 South Street              No
No
7                            98 Clue Drive               N
No
8                         123 Middle Earth             Yes
NaN
9                 25th Main Street, New York             Yes
No
10                    612 Shire Lane, Shire             Yes
No
11                     2394 Hogwarts Avenue               Y
NaN
12                        2039 Main Street             Yes
N
13                        343 City Parkway             Yes
```

```
No
14                              214 HR Avenue                N
No
15                        2395 Hogwarts Avenue              No
N
16            121 Paper Avenue, Pennsylvania               Yes
No
17                             3498 Super Lane               Y
NaN
18                                         N/a             N/a
Yes
19               910 Tatooine Road, Tatooine               Yes
N

#df["Last_Name"]=df["Last_Name"].str.lstrip("...")
#df["Last_Name"]=df["Last_Name"].str.lstrip("/")
#df["Last_Name"]=df["Last_Name"].str.rstrip("_")
df["Last_Name"]=df["Last_Name"].str.strip("123._/")
df

    CustomerID First_Name   Last_Name  Phone_Number  \
0         1001      Frodo     Baggins  123-545-5421
1         1002       Abed       Nadir  123/643/9775
2         1003     Walter       White    7066950392
3         1004     Dwight     Schrute  123-543-2345
4         1005        Jon        Snow  876|678|3469
5         1006        Ron     Swanson  304-762-2467
6         1007       Jeff      Winger           NaN
7         1008   Sherlock      Holmes  876|678|3469
8         1009    Gandalf         NaN           N/a
9         1010      Peter      Parker  123-545-5421
10        1011    Samwise      Gamgee           NaN
11        1012      Harry      Potter    7066950392
12        1013        Don      Draper  123-543-2345
13        1014     Leslie       Knope  876|678|3469
14        1015       Toby  Flenderson  304-762-2467
15        1016        Ron     Weasley  123-545-5421
16        1017   Michael        Scott  123/643/9775
17        1018      Clark        Kent    7066950392
18        1019      Creed      Braton           N/a
19        1020     Anakin   Skywalker  876|678|3469

                                  Address Paying Customer
Do_Not_Contact
0                     123 Shire Lane, Shire             Yes
No
1                      93 West Main Street               No
Yes
2                        298 Drugs Driveway               N
NaN
```

```
3   980 Paper Avenue, Pennsylvania, 18503               Yes
Y
4                             123 Dragons Road               Y
No
5                             768 City Parkway             Yes
Yes
6                            1209 South Street              No
No
7                                98 Clue Drive               N
No
8                             123 Middle Earth             Yes
NaN
9                 25th Main Street, New York             Yes
No
10                       612 Shire Lane, Shire             Yes
No
11                        2394 Hogwarts Avenue               Y
NaN
12                            2039 Main Street             Yes
N
13                            343 City Parkway             Yes
No
14                               214 HR Avenue               N
No
15                        2395 Hogwarts Avenue              No
N
16           121 Paper Avenue, Pennsylvania             Yes
No
17                             3498 Super Lane               Y
NaN
18                                         N/a             N/a
Yes
19              910 Tatooine Road, Tatooine             Yes
N
```

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

```
    CustomerID First_Name   Last_Name  Phone_Number  \
0         1001      Frodo     Baggins  123-545-5421
1         1002       Abed       Nadir  123/643/9775
2         1003     Walter       White           NaN
3         1004     Dwight     Schrute  123-543-2345
4         1005        Jon        Snow  876|678|3469
5         1006        Ron     Swanson  304-762-2467
6         1007       Jeff      Winger           NaN
7         1008   Sherlock      Holmes  876|678|3469
8         1009    Gandalf         NaN           N/a
9         1010      Peter      Parker  123-545-5421
10        1011    Samwise      Gamgee           NaN
```

```
11        1012      Harry      Potter           NaN
12        1013        Don      Draper  123-543-2345
13        1014     Leslie       Knope  876|678|3469
14        1015       Toby  Flenderson  304-762-2467
15        1016        Ron     Weasley  123-545-5421
16        1017   Michael        Scott  123/643/9775
17        1018      Clark        Kent           NaN
18        1019      Creed      Braton           N/a
19        1020     Anakin   Skywalker  876|678|3469

                                  Address Paying Customer
Do_Not_Contact
0                     123 Shire Lane, Shire             Yes
No
1                       93 West Main Street              No
Yes
2                        298 Drugs Driveway               N
NaN
3   980 Paper Avenue, Pennsylvania, 18503             Yes
Y
4                          123 Dragons Road               Y
No
5                          768 City Parkway             Yes
Yes
6                         1209 South Street              No
No
7                             98 Clue Drive               N
No
8                          123 Middle Earth             Yes
NaN
9                25th Main Street, New York             Yes
No
10                    612 Shire Lane, Shire             Yes
No
11                      2394 Hogwarts Avenue               Y
NaN
12                         2039 Main Street             Yes
N
13                         343 City Parkway             Yes
No
14                            214 HR Avenue               N
No
15                      2395 Hogwarts Avenue              No
N
16           121 Paper Avenue, Pennsylvania             Yes
No
17                          3498 Super Lane               Y
NaN
18                                      N/a             N/a
```

```
Yes
19             910 Tatooine Road, Tatooine             Yes
N

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

----------------------------------------------------------------------
-----
TypeError                                 Traceback (most recent call
last)
Cell In[69], line 1
----> 1 df["Phone_Number"]=df["Phone_Number"].apply(lambda
x:x[0:3]+'-'+x[3:6]+'-'+x[6:10])
      2 df

File ~\anaconda3\Lib\site-packages\pandas\core\series.py:4924, in
Series.apply(self, func, convert_dtype, args, by_row, **kwargs)
   4789 def apply(
   4790     self,
   4791     func: AggFuncType,
   (...)
   4796     **kwargs,
   4797 ) -> DataFrame | Series:
   4798     """
   4799     Invoke function on values of Series.
   4800 
   (...)
   4915     dtype: float64
   4916     """
   4917     return SeriesApply(
   4918         self,
   4919         func,
   4920         convert_dtype=convert_dtype,
   4921         by_row=by_row,
   4922         args=args,
   4923         kwargs=kwargs,
-> 4924     ).apply()

File ~\anaconda3\Lib\site-packages\pandas\core\apply.py:1427, in
SeriesApply.apply(self)
   1424     return self.apply_compat()
   1426 # self.func is Callable
-> 1427 return self.apply_standard()

File ~\anaconda3\Lib\site-packages\pandas\core\apply.py:1507, in
SeriesApply.apply_standard(self)
   1501 # row-wise access
   1502 # apply doesn't have a `na_action` keyword and for backward
```

```
compat reasons
   1503 # we need to give `na_action="ignore"` for categorical data.
   1504 # TODO: remove the `na_action="ignore"` when that default has
been changed in
   1505 #  Categorical (GH51645).
   1506 action = "ignore" if isinstance(obj.dtype, CategoricalDtype)
else None
-> 1507 mapped = obj._map_values(
   1508     mapper=curried, na_action=action,
convert=self.convert_dtype
   1509 )
   1511 if len(mapped) and isinstance(mapped[0], ABCSeries):
   1512     # GH#43986 Need to do list(mapped) in order to get treated
as nested
   1513     #  See also GH#25959 regarding EA support
   1514     return obj._constructor_expanddim(list(mapped),
index=obj.index)

File ~\anaconda3\Lib\site-packages\pandas\core\base.py:921, in
IndexOpsMixin._map_values(self, mapper, na_action, convert)
    918 if isinstance(arr, ExtensionArray):
    919     return arr.map(mapper, na_action=na_action)
--> 921 return algorithms.map_array(arr, mapper, na_action=na_action,
convert=convert)

File ~\anaconda3\Lib\site-packages\pandas\core\algorithms.py:1743, in
map_array(arr, mapper, na_action, convert)
   1741 values = arr.astype(object, copy=False)
   1742 if na_action is None:
-> 1743     return lib.map_infer(values, mapper, convert=convert)
   1744 else:
   1745     return lib.map_infer_mask(
   1746         values, mapper, mask=isna(values).view(np.uint8),
convert=convert
   1747     )

File lib.pyx:2972, in pandas._libs.lib.map_infer()

Cell In[69], line 1, in <lambda>(x)
----> 1 df["Phone_Number"]=df["Phone_Number"].apply(lambda
x:x[0:3]+'-'+x[3:6]+'-'+x[6:10])
      2 df

TypeError: 'float' object is not subscriptable

df["Phone_Number"]=df["Phone_Number"].str.replace('NaN','')
df

    CustomerID First_Name   Last_Name  Phone_Number  \
0         1001      Frodo     Baggins  123-545-5421
```

```
1          1002       Abed       Nadir  123/643/9775
2          1003     Walter       White           NaN
3          1004     Dwight     Schrute  123-543-2345
4          1005        Jon        Snow  876|678|3469
5          1006        Ron     Swanson  304-762-2467
6          1007       Jeff      Winger           NaN
7          1008   Sherlock      Holmes  876|678|3469
8          1009    Gandalf         NaN           N/a
9          1010      Peter      Parker  123-545-5421
10         1011    Samwise      Gamgee           NaN
11         1012      Harry      Potter           NaN
12         1013        Don      Draper  123-543-2345
13         1014     Leslie       Knope  876|678|3469
14         1015       Toby  Flenderson  304-762-2467
15         1016        Ron     Weasley  123-545-5421
16         1017   Michael        Scott  123/643/9775
17         1018      Clark        Kent           NaN
18         1019      Creed      Braton           N/a
19         1020     Anakin   Skywalker  876|678|3469

                                  Address Paying Customer
Do_Not_Contact
0                    123 Shire Lane, Shire             Yes
No
1                      93 West Main Street              No
Yes
2                        298 Drugs Driveway              N
NaN
3   980 Paper Avenue, Pennsylvania, 18503             Yes
Y
4                          123 Dragons Road              Y
No
5                         768 City Parkway            Yes
Yes
6                        1209 South Street             No
No
7                            98 Clue Drive              N
No
8                         123 Middle Earth            Yes
NaN
9                  25th Main Street, New York            Yes
No
10                   612 Shire Lane, Shire            Yes
No
11                     2394 Hogwarts Avenue              Y
NaN
12                        2039 Main Street            Yes
N
13                         343 City Parkway            Yes
```

```
No
14                             214 HR Avenue                N
No
15                      2395 Hogwarts Avenue               No
N
16         121 Paper Avenue, Pennsylvania                 Yes
No
17                            3498 Super Lane               Y
NaN
18                                        N/a             N/a
Yes
19            910 Tatooine Road, Tatooine                 Yes
N
```

```
df[["Street
Adress","State","ZipCode"]]=df["Address"].str.split(',',2,expands=True
)
df
```

```
---------------------------------------------------------------------------
-----
TypeError                                 Traceback (most recent call
last)
Cell In[77], line 1
----> 1 df[["Street
Adress","State","ZipCode"]]=df["Address"].str.split(',',2,expands=True
)
      2 df

File ~\anaconda3\Lib\site-packages\pandas\core\strings\
accessor.py:137, in
forbid_nonstring_types.<locals>._forbid_nonstring_types.<locals>.wrapp
er(self, *args, **kwargs)
    132     msg = (
    133         f"Cannot use .str.{func_name} with values of "
    134         f"inferred dtype '{self._inferred_dtype}'."
    135     )
    136     raise TypeError(msg)
--> 137 return func(self, *args, **kwargs)

TypeError: StringMethods.split() got an unexpected keyword argument
'expands'
```

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

```
    CustomerID First_Name   Last_Name  Phone_Number  \
0         1001      Frodo     Baggins  123-545-5421
1         1002       Abed       Nadir  123/643/9775
```

```
2          1003     Walter       White           NaN
3          1004     Dwight     Schrute  123-543-2345
4          1005        Jon        Snow  876|678|3469
5          1006        Ron     Swanson  304-762-2467
6          1007       Jeff      Winger           NaN
7          1008   Sherlock      Holmes  876|678|3469
8          1009    Gandalf         NaN           N/a
9          1010      Peter      Parker  123-545-5421
10         1011    Samwise      Gamgee           NaN
11         1012      Harry      Potter           NaN
12         1013        Don      Draper  123-543-2345
13         1014     Leslie       Knope  876|678|3469
14         1015       Toby  Flenderson  304-762-2467
15         1016        Ron     Weasley  123-545-5421
16         1017   Michael        Scott  123/643/9775
17         1018      Clark        Kent           NaN
18         1019      Creed      Braton           N/a
19         1020     Anakin   Skywalker  876|678|3469

                                             Address Paying Customer
Do_Not_Contact
0                        123 Shire Lane, Shire               Y
No
1                          93 West Main Street               N
Yes
2                           298 Drugs Driveway               N
NaN
3   980 Paper Avenue, Pennsylvania, 18503               Y
Y
4                             123 Dragons Road               Y
No
5                             768 City Parkway               Y
Yes
6                            1209 South Street               N
No
7                                98 Clue Drive               N
No
8                             123 Middle Earth               Y
NaN
9                25th Main Street, New York               Y
No
10                       612 Shire Lane, Shire               Y
No
11                        2394 Hogwarts Avenue               Y
NaN
12                            2039 Main Street               Y
N
13                            343 City Parkway               Y
No
```

```
14                               214 HR Avenue               N
No
15                        2395 Hogwarts Avenue               N
N
16          121 Paper Avenue, Pennsylvania               Y
No
17                             3498 Super Lane               Y
NaN
18                                         N/a             N/a
Yes
19             910 Tatooine Road, Tatooine               Y
N
```

```
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
```

```
    CustomerID First_Name   Last_Name  Phone_Number  \
0         1001      Frodo     Baggins  123-545-5421
1         1002       Abed       Nadir  123/643/9775
2         1003     Walter       White           NaN
3         1004     Dwight     Schrute  123-543-2345
4         1005        Jon        Snow  876|678|3469
5         1006        Ron     Swanson  304-762-2467
6         1007       Jeff      Winger           NaN
7         1008   Sherlock      Holmes  876|678|3469
8         1009    Gandalf         NaN           N/a
9         1010      Peter      Parker  123-545-5421
10        1011    Samwise      Gamgee           NaN
11        1012      Harry      Potter           NaN
12        1013        Don      Draper  123-543-2345
13        1014     Leslie       Knope  876|678|3469
14        1015       Toby  Flenderson  304-762-2467
15        1016        Ron     Weasley  123-545-5421
16        1017   Michael        Scott  123/643/9775
17        1018      Clark        Kent           NaN
18        1019      Creed      Braton           N/a
19        1020     Anakin   Skywalker  876|678|3469

                                  Address Paying Customer
Do_Not_Contact
0                    123 Shire Lane, Shire               Y
N
1                      93 West Main Street               N
Y
2                       298 Drugs Driveway               N
NaN
3   980 Paper Avenue, Pennsylvania, 18503               Y
Y
4                         123 Dragons Road               Y
```

```
N
5                              768 City Parkway               Y
Y
6                             1209 South Street               N
N
7                                98 Clue Drive                N
N
8                              123 Middle Earth               Y
NaN
9                    25th Main Street, New York               Y
N
10                        612 Shire Lane, Shire               Y
N
11                          2394 Hogwarts Avenue              Y
NaN
12                             2039 Main Street               Y
N
13                             343 City Parkway               Y
N
14                                214 HR Avenue               N
N
15                          2395 Hogwarts Avenue              N
N
16             121 Paper Avenue, Pennsylvania                 Y
N
17                              3498 Super Lane               Y
NaN
18                                          N/a             N/a
Y
19                 910 Tatooine Road, Tatooine                Y
N

#df.replace('N/a','')
#df.replace('NaN','')
df=df.filled('')
df

----------------------------------------------------------------------
-----
AttributeError                            Traceback (most recent call
last)
~\AppData\Local\Temp\ipykernel_10188\1390790910.py in ?()
      1 #df.replace('N/a','')
      2 #df.replace('NaN','')
----> 3 df=df.filled('')
      4 df

~\anaconda3\Lib\site-packages\pandas\core\generic.py in ?(self, name)
   6295             and name not in self._accessors
   6296             and
```

```
self._info_axis._can_hold_identifiers_and_holds_name(name)
   6297         ):
   6298             return self[name]
-> 6299         return object.__getattribute__(self, name)

AttributeError: 'DataFrame' object has no attribute 'filled'
```