PRACTICAL NO 7

Aim: Develop an application to pre-process data imported from external sources .

Step 1: Load the csv file

Code:

```
import pandas as pd
df = pd.read_csv("airbnbData - airbnbData.csv")
```

Step 2: Show the first few rows

Code:

```
import pandas as pd


#1.display first 10 rows.
data=pd.read_csv("airbnbData.csv",encoding='latin-1')
print(data.head().to_string())
#print(data.to_string())
```

Output:

```
   ListingID  ...                                          ShortDesc
0     281552  ...  \n    Entire home/apt Мвеб 24 reviews Мвеб Har...
1     182613  ...  \n    Entire home/apt Мвеб 17 reviews Мвеб Cha...
2    1587540  ...  \n    Entire home/apt Мвеб 5 reviews Мвеб Char...
3     469506  ...  \n    Entire home/apt Мвеб 60 reviews Мвеб Bro...
4    3937268  ...  \n    Private room Мвеб 11 reviews Мвеб Brookl...

[5 rows x 65 columns]
```

Step 3: Show the values as NaN where the values are empty under hostname

Code:

```
print('---------------------- null values ----------------------')
empty=pd.isnull(data['HostName'])
print(empty)
```

```
---------- null values ----------
0       False
1       False
2       False
3       False
4       False
       ...
2017    False
2018    False
2019    False
2020    False
2021    False
Name: HostName, Length: 2022, dtype: bool
```

Output:

Step 4: Show the data types of each column

Code:

```
print("---------- sum of the null values ----------")
abt_lst=data['AboutListing'].isnull().sum()
print(abt_lst)
print(data.dtypes)
```

Output :

```
---------- sum of the null values ----------
3
ListingID      int64
Title         object
UserID         int64
baseurl       object
Price          int64
              ...
S_CheckIn     object
S_Checkout    object
S_NumBeds     object
S_PropType    object
ShortDesc     object
Length: 65, dtype: object
```

Step 5: Set index to id

Code :

```
air_df = df.set_index("HostName", append=False)
print(air_df.head())
```

Output:

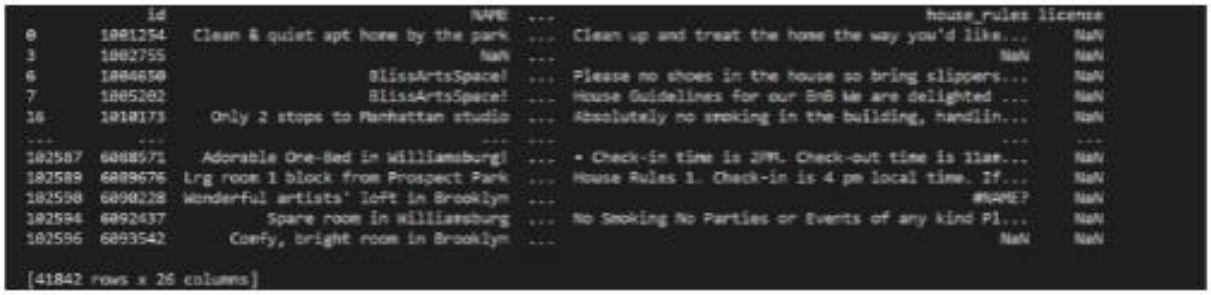
```
                ListingID  ...                                          ShortDesc
HostName                   ...
Mary Catherine     281552  ...  \n    Entire home/apt Мвеб 24 reviews Мвеб Har...
Max                182613  ...  \n    Entire home/apt Мвеб 17 reviews Мвеб Cha...
Finola            1587540  ...  \n    Entire home/apt Мвеб 5 reviews Мвеб Char...
Rupal              469506  ...  \n    Entire home/apt Мвеб 60 reviews Мвеб Bro...
Natasha           3937268  ...  \n    Private room Мвеб 11 reviews Мвеб Brookl...

[5 rows x 64 columns]
```

Step 6: Find the location of Brooklyn under neighborhood group

Code:

```
brooklyn_location = df.loc[df["neighbourhood group"] == "Brooklyn"]
print(brooklyn_location)
```

Output:



```
          id                                NAME  ...                                house_rules license
0    1001254  Clean & quiet apt home by the park  ...  Clean up and treat the home the way you'd like...     NaN
3    1002755                                 NaN  ...                                               NaN     NaN
6    1004650                    BlissArtsSpace!  ...  Please no shoes in the house so bring slippers...     NaN
7    1005202                    BlissArtsSpace!  ...  House Guidelines for our BnB We are delighted ...     NaN
16   1010173     Only 2 stops to Manhattan studio  ...  Absolutely no smoking in the building, handlin...     NaN
...      ...                                 ...  ...                                               ...     ...
102587  6088571   Adorable One-Bed in Williamsburg!  ...  • Check-in time is 2PM. Check-out time is 11am...     NaN
102589  6089676  Lrg room 1 block from Prospect Park  ...  House Rules 1. Check-in is 4 pm local time. If...     NaN
102598  6090228  Wonderful artists' loft in Brooklyn  ...                                            #NAME?     NaN
102594  6092437          Spare room in Williamsburg  ...  No Smoking No Parties or Events of any kind Pl...     NaN
102596  6093542       Comfy, bright room in Brooklyn  ...                                               NaN     NaN

[41842 rows x 26 columns]
```

Step 7: Find out how many null values are there under host_identifity_verified

Code:

```
null_count = df["host_identity_verified"].isnull().sum()
print("Total null values in host_identity_verified are column: "+ null_count)
```

Output :

```
Total null values in host_identity_verified are column :  289
```

Step 8: How many hotels are instant_bookable

Code:

```
df = pd.read_csv("airbnbData - airbnbData.csv")
instant_bookable_hotels = df[df["BookInstantly"] == "Yes"]
count_instant_bookable = len(instant_bookable_hotels)
print("Number of instant bookable hotels", count_instant_bookable)
```

Out

```
Number of instant bookable hotels 142
```