In [7]:
```python
#here importing the pnadas and numpy library
import pandas as pd
import numpy as np
```

In [29]:
```python
#reading the csv file
dataframe = pd.read_csv("D:\TEIT_17_DSPL\house rent prediction data main\House
```

In [9]:
```python
#printing the number of samples and attributes of dataset
print(dataframe.shape)
```

```
(4746, 12)
```

In [11]:
```python
#printing the coiumns of data set
columns = list(dataframe.columns)
print(columns)
```

```
['Posted On', 'BHK', 'Rent', 'Size', 'Floor', 'Area Type', 'Area Locality',
'City', 'Furnishing Status', 'Tenant Preferred', 'Bathroom', 'Point of Contac
t']
```

In [12]:
```python
#print the first 5 samples
print(dataframe.head())
```

```
    Posted On  BHK   Rent  Size            Floor    Area Type  \
0  2022-05-18    2  10000  1100  Ground out of 2   Super Area
1  2022-05-13    2  20000   800      1 out of 3   Super Area
2  2022-05-16    2  17000  1000      1 out of 3   Super Area
3  2022-07-04    2  10000   800      1 out of 2   Super Area
4  2022-05-09    2   7500   850      1 out of 2  Carpet Area

             Area Locality     City Furnishing Status Tenant Preferred  \
0                   Bandel  Kolkata       Unfurnished  Bachelors/Family
1  Phool Bagan, Kankurgachi  Kolkata     Semi-Furnished  Bachelors/Family
2   Salt Lake City Sector 2  Kolkata     Semi-Furnished  Bachelors/Family
3              Dumdum Park  Kolkata       Unfurnished  Bachelors/Family
4             South Dum Dum  Kolkata       Unfurnished         Bachelors

   Bathroom Point of Contact
0         2    Contact Owner
1         1    Contact Owner
2         1    Contact Owner
3         1    Contact Owner
4         1    Contact Owner
```

In [18]:
```python
array = dataframe.values
print(array[1])
```

```
['2022-05-13' 2 20000 800 '1 out of 3' 'Super Area'
 'Phool Bagan, Kankurgachi' 'Kolkata' 'Semi-Furnished' 'Bachelors/Family'
 1 'Contact Owner']
```

In [30]:
```python
#created a new dataframe
new_dataframe = dataframe
isnull = new_dataframe.isnull()
# print(isnull) #checking null value or not
sum = new_dataframe.isnull().sum()
# print(sum)
mean = new_dataframe.isnull().mean()
# print(mean)
median = new_dataframe.isnull().median()
# print(median)
print(new_dataframe.shape)
```

```
(4746, 12)
```

In [37]:
```python
#finding duplicate in dataset
duplicate = new_dataframe.duplicated()
# print(duplicate)
# finding dupicate oin particular column
rent = new_dataframe.Rent.duplicated()
print(rent)
print(new_dataframe.Size.duplicated())
# finding any duplicate value present in dataset --> it will return false is t
print(new_dataframe.duplicated().any())
```

```
0       False
1       False
2       False
3        True
4       False
        ...
4741     True
4742     True
4743     True
4744     True
4745     True
Name: Rent, Length: 4746, dtype: bool
0       False
1       False
2       False
3        True
4       False
        ...
4741     True
4742     True
4743     True
4744     True
4745     True
Name: Size, Length: 4746, dtype: bool
False
```

In [38]:
```python
# printing datatype of column
dttypes = new_dataframe.dtypes
print(dttypes);
```

```
Posted On          object
BHK                 int64
Rent                int64
Size                int64
Floor              object
Area Type          object
Area Locality      object
City               object
Furnishing Status  object
Tenant Preferred   object
Bathroom            int64
Point of Contact   object
dtype: object
```

In [41]:
```python
# changing the attributes of dataset for training prupose
new_dataframe.replace({"Super Area" : "1" , "Carpet Area": "2"},inplace = True
print(new_dataframe.shape)
print(new_dataframe.head())
```

```
(4746, 12)
    Posted On  BHK   Rent  Size         Floor Area Type  \
0  2022-05-18    2  10000  1100  Ground out of 2         1
1  2022-05-13    2  20000   800       1 out of 3         1
2  2022-05-16    2  17000  1000       1 out of 3         1
3  2022-07-04    2  10000   800       1 out of 2         1
4  2022-05-09    2   7500   850       1 out of 2         2

                Area Locality     City Furnishing Status  Tenant Preferred  \
0                      Bandel  Kolkata       Unfunished  Bachelors/Family
1    Phool Bagan, Kankurgachi  Kolkata    Semi-Furnished  Bachelors/Family
2    Salt Lake City Sector 2  Kolkata    Semi-Furnished  Bachelors/Family
3                Dumdum Park  Kolkata       Unfunished  Bachelors/Family
4               South Dum Dum  Kolkata       Unfunished          Bachelors

   Bathroom Point of Contact
0         2    Contact Owner
1         1    Contact Owner
2         1    Contact Owner
3         1    Contact Owner
4         1    Contact Owner
```

In [44]:
```python
# replacing the Furninshing status
new_dataframe.replace({"Unfurnished" : "1" , "Furnished": "2" , "Semi-Furnishe
print(new_dataframe.shape)
print(new_dataframe.head())
```

```
(4746, 12)
    Posted On  BHK   Rent  Size            Floor  Area Type  \
0  2022-05-18    2  10000  1100  Ground out of 2          1
1  2022-05-13    2  20000   800       1 out of 3          1
2  2022-05-16    2  17000  1000       1 out of 3          1
3  2022-07-04    2  10000   800       1 out of 2          1
4  2022-05-09    2   7500   850       1 out of 2          2

               Area Locality     City  Furnishing Status  Tenant Preferred  \
0                     Bandel  Kolkata                  1  Bachelors/Family
1   Phool Bagan, Kankurgachi  Kolkata                  2  Bachelors/Family
2    Salt Lake City Sector 2  Kolkata                  2  Bachelors/Family
3               Dumdum Park  Kolkata                  1  Bachelors/Family
4              South Dum Dum  Kolkata                  1         Bachelors

   Bathroom Point of Contact
0         2    Contact Owner
1         1    Contact Owner
2         1    Contact Owner
3         1    Contact Owner
4         1    Contact Owner
```

In [ ]:

In [ ]: