

Student Name	Aarya Trifale
SRN No	31241458
Roll No	41
Program	Computer Engineering
Year	Second Year
Division	E
Subject	Fundamental Of Data science
Assignment No	2

Assignment Number - 2

Assignment Number - 2

Name: Aarya Trifale, Roll no: 41

Problem Statement:

In the given Dataset:

A)Perform data wrangling for formats of merging, grouping and concatenation.

B)Explore the dataset using descriptive statistics to gain insights into the data distribution and identify patterns.

PART - A

```
[28]: #Importing pandas library and numpy library as pd and np resp so that we can use their methods to process our excel data.

import pandas as pd
import numpy as np
file = pd.read_excel("Bengaluru_House_Data_with_ID.xlsx") #opened excel file i.

e. dataframe using read_excel method and saved it in a variable file.
file
```

```
[28]:
                ID
                                                  availability
                               area_type
      0
                 1
                    Super built-up Area
                                          2024-12-19 00:00:00
                 2
      1
                              Plot
                                                 Ready To Move
                                   Area
      2
                 3
                          Built-up Area
                                                 Ready To Move
      3
                 4
                    Super built-up
                                                 Ready To Move
                                    Area
                    Super built-up
                                    Area
                                                 Ready To Move
                                                 Ready To Move
      13315 13316
                          Built-up
                                    Area
      13316 13317
                    Super built-up
                                    Area
                                                 Ready To Move
      13317
             13318
                          Built-up
                                    Area
                                                 Ready To Move
      13318
             13319
                    Super built-up Area
                                          2024-06-18 00:00:00
      13319
             13320
                    Super built-up Area
                                                 Ready To Move
                             location
                                                   society total_sqft
                                                                             balcony
                                             size
                                                                       bath
      0
             Electronic City Phase II
                                            2 BHK
                                                   Coomee
                                                                 1056
                                                                        2.0
                                                                                  1.0
      1
                     Chikka Tirupathi 4 Bedroom
                                                                 2600
                                                                        5.0
                                                                                  3.0
                                                   Theanmp
      2
                                                                                  3.0
                          Uttarahalli
                                            3 BHK
                                                       NaN
                                                                 1440
                                                                        2.0
      3
                   Lingadheeranahalli
                                            3 BHK Soiewre
                                                                 1521
                                                                        3.0
                                                                                  1.0
```

```
4.0
      13315
                            Whitefield
                                       5 Bedroom ArsiaEx
                                                                  3453
                                                                                   0.0
                        Richards Town
                                            4 BHK
                                                                          5.0
      13316
                                                        NaN
                                                                  3600
                                                                                   NaN
      13317
                Raja Rajeshwari Nagar
                                            2 BHK
                                                   Mahla T
                                                                  1141
                                                                          2.0
                                                                                   1.0
                      Padmanabhanagar
                                                    SollyCl
      13318
                                            4 BHK
                                                                  4689
                                                                          4.0
                                                                                   1.0
      13319
                          Doddathoguru
                                            1 BHK
                                                        NaN
                                                                   550
                                                                          1.0
                                                                                   1.0
              price
      0
              39.07
             120.00
      1
      2
              62.00
              95.00
      4
              51.00
      13315
             231.00
             400.00
      13316
      13317
              60.00
      13318
             488.00
      13319
              17.00
      [13320 rows x 10 columns]
[30]: file.head() #.head() returns the first few rows (the "head" of the dataframe i.
       \rightarrowe. excel file)
[30]:
         ID
                        area_type
                                           availability
                                                                           location \
      0
             Super built-up Area
                                    2024-12-19 00:00:00 Electronic City Phase II
      1
                       Plot Area
                                          Ready To Move
                                                                  Chikka Tirupathi
      2
                   Built-up Area
                                          Ready To Move
                                                                       Uttarahalli
          3
      3
             Super built-up Area
                                          Ready To Move
                                                                Lingadheeranahalli
      4
             Super built-up Area
                                          Ready To Move
                                                                          Kothanur
              size
                    society total_sqft
                                         bath balcony
                                                          price
             2 BHK
                    Coomee
                                   1056
                                          2.0
                                                    1.0
                                                          39.07
      0
                    Theanmp
      1
         4 Bedroom
                                   2600
                                          5.0
                                                    3.0 120.00
      2
             3 ВНК
                        NaN
                                   1440
                                          2.0
                                                    3.0
                                                          62.00
             3 BHK
      3
                    Soiewre
                                   1521
                                          3.0
                                                    1.0
                                                          95.00
      4
             2 BHK
                        NaN
                                   1200
                                          2.0
                                                    1.0
                                                          51.00
[31]: file.tail() #.tail() method returns a specified number of last row
[31]:
                                                   availability \
                ID
                                area_type
                           Built-up Area
                                                 Ready To Move
      13315 13316
                    Super built-up
                                                 Ready To Move
      13316 13317
                                    Area
      13317
             13318
                           Built-up Area
                                                 Ready To Move
      13318 13319
                    Super built-up Area
                                           2024-06-18 00:00:00
```

Kothanur

2 BHK

NaN

1200

2.0

1.0

4

```
society total_sqft
                           location
                                          size
                                                                      bath
                                                                            balcony \
      13315
                                                                                0.0
                         Whitefield
                                     5 Bedroom
                                                 ArsiaEx
                                                               3453
                                                                       4.0
      13316
                     Richards Town
                                         4 BHK
                                                     NaN
                                                               3600
                                                                       5.0
                                                                                NaN
                                         2 BHK
                                                Mahla T
      13317
             Raja Rajeshwari Nagar
                                                               1141
                                                                       2.0
                                                                                1.0
                   Padmanabhanagar
                                         4 BHK
                                                 SollyCl
                                                               4689
                                                                       4.0
                                                                                1.0
      13318
      13319
                      Doddathoguru
                                         1 BHK
                                                     NaN
                                                                 550
                                                                       1.0
                                                                                1.0
             price
             231.0
      13315
      13316
             400.0
      13317
              60.0
      13318 488.0
      13319
              17.0
[32]: file.describe() #.describe() calculates a few summary statistics for each column
[32]:
                                                balcony
                        ID
                                    bath
                                                                price
             13320.000000
                            13247.000000
                                          12711.000000
                                                         13320.000000
      count
      mean
              6660.500000
                                2.692610
                                               1.584376
                                                           112.565627
      std
              3845.297128
                                               0.817263
                                1.341458
                                                           148.971674
      min
                 1.000000
                                1.000000
                                               0.000000
                                                             8.000000
      25%
              3330.750000
                                2.000000
                                               1.000000
                                                            50.000000
                                2.000000
      50%
              6660.500000
                                               2.000000
                                                            72.000000
      75%
              9990.250000
                                3.000000
                                               2.000000
                                                           120.000000
             13320.000000
                               40.000000
                                               3.000000
                                                          3600.000000
      max
[33]: file.isnull().sum() #this will return the number of missing values in dataset
[33]: ID
                          0
                          0
      area_type
      availability
                          0
      location
                          1
                         16
      size
      society
                       5502
      total_sqft
                          0
      bath
                         73
                        609
      balcony
      price
                          0
      dtype: int64
[34]: file['society'] = file['society'].fillna('Null Values') #this fillna() replaces_
       missing values in the `society` column with `'Null Values'` and then shows
       ⇔the updated column.
      file['society']
```

Ready To Move

13319 13320 Super built-up Area

```
[34]: 0
                   Coomee
                   Theanmp
      1
      2
               Null Values
      3
                   Soiewre
      4
               Null Values
      13315
                   ArsiaEx
      13316
               Null Values
      13317
                   Mahla T
      13318
                   SollyCl
      13319
               Null Values
      Name: society, Length: 13320, dtype: object
[35]: file['size'] = file['size'].fillna('10bhk')
      file['size']
[35]: 0
                   2 BHK
      1
               4 Bedroom
      2
                   3 ВНК
      3
                   3 ВНК
      4
                   2 BHK
      13315
               5 Bedroom
      13316
                   4 BHK
                   2 BHK
      13317
      13318
                   4 BHK
      13319
                   1 BHK
      Name: size, Length: 13320, dtype: object
[36]: file['bath'] = file['bath'].fillna('19')
      file['bath']
[36]: 0
               2.0
               5.0
      1
      2
               2.0
      3
               3.0
      4
               2.0
      13315
               4.0
               5.0
      13316
      13317
               2.0
      13318
               4.0
               1.0
      13319
      Name: bath, Length: 13320, dtype: object
[37]: file['balcony'] = file['balcony'].fillna('1')
      file['balcony']
```

```
[37]: 0
               1.0
               3.0
      1
      2
               3.0
      3
               1.0
      4
               1.0
      13315
               0.0
      13316
                 1
      13317
               1.0
      13318
               1.0
               1.0
      13319
      Name: balcony, Length: 13320, dtype: object
[38]: file['location'] = file['location'].fillna('pune')
      file['location']
[38]: 0
               Electronic City Phase II
                       Chikka Tirupathi
      1
      2
                            Uttarahalli
      3
                     Lingadheeranahalli
      4
                               Kothanur
      13315
                             Whitefield
                          Richards Town
      13316
      13317
                  Raja Rajeshwari Nagar
                        Padmanabhanagar
      13318
      13319
                           Doddathoguru
      Name: location, Length: 13320, dtype: object
[39]: #dividing the dataset into two parts to perform various operations like merge,
       ⇔join, etc
      midpoint = len(file) //2
      df2 = file.iloc[:midpoint]
      df3 = file.iloc[midpoint:]
[56]: df2
[56]:
              ID
                                                availability \
                             area_type
                  Super built-up Area
                                         2024-12-19 00:00:00
      0
      1
               2
                            Plot Area
                                               Ready To Move
      2
               3
                                               Ready To Move
                        Built-up Area
                  Super built-up
                                               Ready To Move
      3
      4
                  Super built-up
                                               Ready To Move
                                  Area
      6655
            6656
                  Super built-up
                                  Area
                                               Ready To Move
      6656
            6657
                        Built-up Area
                                               Ready To Move
      6657
            6658
                        Built-up Area
                                               Ready To Move
```

```
6658
            6659
                   Super built-up Area
                                                Ready To Move
      6659
            6660
                             Plot
                                   Area
                                                Ready To Move
                             location
                                             size
                                                        society total_sqft bath
      0
            Electronic City Phase II
                                            2 BHK
                                                        Coomee
                                                                       1056
                                                                             2.0
      1
                     Chikka Tirupathi
                                       4 Bedroom
                                                        Theanmp
                                                                       2600
                                                                             5.0
      2
                          Uttarahalli
                                                   Null Values
                                                                       1440
                                                                             2.0
                                            3 BHK
      3
                   Lingadheeranahalli
                                            3 ВНК
                                                        Soiewre
                                                                       1521
                                                                             3.0
      4
                                            2 BHK
                             Kothanur
                                                   Null Values
                                                                       1200
                                                                             2.0
                       Sarjapur Road
      6655
                                            1 BHK
                                                        Ahriaup
                                                                        539
                                                                             1.0
      6656
                        Sarakki Nagar
                                       4 Bedroom
                                                   Null Values
                                                                       1200
                                                                             4.0
      6657
                             Attibele 3 Bedroom
                                                   Null Values
                                                                       2400
                                                                             3.0
      6658
                         Thubarahalli
                                            2 BHK
                                                        ShdhiSa
                                                                       1200
                                                                             2.0
      6659
                         Kundalahalli 4 Bedroom Null Values
                                                                       2500
                                                                            5.0
           balcony
                      price
      0
               1.0
                      39.07
      1
               3.0
                    120.00
      2
               3.0
                      62.00
      3
                      95.00
               1.0
      4
               1.0
                      51.00
               1.0
      6655
                      45.00
      6656
               1.0
                      85.00
      6657
               3.0
                     120.00
      6658
               2.0
                      79.55
      6659
               2.0
                    350.00
      [6660 rows x 10 columns]
[57]: df3
                ID
                                                    availability \
                                area_type
                                                   Ready To Move
      6660
              6661
                     Super built-up Area
      6661
              6662
                     Super built-up
                                      Area
                                                   Ready To Move
      6662
              6663
                               Plot
                                                   Ready To Move
                                      Area
      6663
              6664
                               Plot
                                      Area
                                                   Ready To Move
      6664
              6665
                     Super built-up
                                      Area
                                                   Ready To Move
             13316
                                                   Ready To Move
      13315
                           Built-up
                                      Area
      13316
             13317
                     Super built-up
                                      Area
                                                   Ready To Move
      13317
             13318
                           Built-up
                                      Area
                                                   Ready To Move
      13318
             13319
                     Super built-up
                                      Area
                                            2024-06-18 00:00:00
                     Super built-up
      13319
             13320
                                      Area
                                                   Ready To Move
                                                      society total_sqft bath balcony \
                           location
                                           size
```

[57]:

```
Sarjapur Road
                                                     MJinele
      6661
                                         2 BHK
                                                                         2.0
                                                                                  2.0
                        Thanisandra
                                                     SLityee
                                                                   1039
      6662
                 Mahalakshmi Puram
                                     4 Bedroom
                                                 Null Values
                                                                   1500
                                                                         5.0
                                                                                  3.0
      6663
                                     5 Bedroom
                          Gottigere
                                                 Null Values
                                                                   1500
                                                                         4.0
                                                                                  1.0
      6664
                      Malleshwaram
                                         3 BHK
                                                     Maionra
                                                                   2520
                                                                         3.0
                                                                                  2.0
                                                                   3453 4.0
                                                                                  0.0
      13315
                        Whitefield 5 Bedroom
                                                     ArsiaEx
                                                                   3600
      13316
                     Richards Town
                                         4 BHK
                                                Null Values
                                                                         5.0
                                                                                    1
             Raja Rajeshwari Nagar
                                         2 BHK
                                                     Mahla T
                                                                         2.0
                                                                                  1.0
      13317
                                                                   1141
      13318
                   Padmanabhanagar
                                         4 BHK
                                                     SollyCl
                                                                   4689
                                                                         4.0
                                                                                  1.0
      13319
                      Doddathoguru
                                         1 BHK
                                                                    550 1.0
                                                Null Values
                                                                                  1.0
             price
              40.0
      6660
      6661
              39.5
      6662
             233.0
      6663
             100.0
      6664
             150.0
      13315
             231.0
      13316
             400.0
      13317
              60.0
      13318
             488.0
      13319
              17.0
      [6660 rows x 10 columns]
     Using .merge()
[42]: df2.merge(df3) # Merge df2 and df3 based on the 'ID' column (default is inner_
       ⇒join)
[42]: Empty DataFrame
      Columns: [ID, area type, availability, location, size, society, total sqft,
      bath, balcony, price]
      Index: []
[43]: df2.merge(df3,how='outer') #Outer Merge: Keeps all rows from both dataframes,
       ⇔fills missing values with NaN
[43]:
                ID
                                area_type
                                                   availability \
      0
                 1
                    Super built-up Area
                                           2024-12-19 00:00:00
      1
                 2
                               Plot
                                     Area
                                                  Ready To Move
      2
                 3
                           Built-up
                                    Area
                                                  Ready To Move
      3
                    Super built-up
                                     Area
                                                  Ready To Move
      4
                    Super built-up
                                     Area
                                                 Ready To Move
```

3 BHK

6660

2.0

1186

2.0

```
13315 13316
                          Built-up
                                                  Ready To Move
                                     Area
      13316
            13317
                    Super built-up
                                                 Ready To Move
                                     Area
      13317
             13318
                           Built-up
                                     Area
                                                  Ready To Move
      13318
             13319
                    Super built-up
                                     Area
                                           2024-06-18 00:00:00
      13319
             13320
                    Super built-up Area
                                                  Ready To Move
                                                        society total_sqft bath \
                              location
                                             size
      0
                                                                      1056
             Electronic City Phase II
                                            2 BHK
                                                        Coomee
                                                                            2.0
      1
                     Chikka Tirupathi 4 Bedroom
                                                                      2600 5.0
                                                        Theanmp
      2
                          Uttarahalli
                                            3 BHK
                                                   Null Values
                                                                      1440 2.0
      3
                   Lingadheeranahalli
                                            3 BHK
                                                        Soiewre
                                                                      1521
                                                                            3.0
      4
                              Kothanur
                                            2 BHK
                                                   Null Values
                                                                      1200 2.0
      13315
                            Whitefield 5 Bedroom
                                                        ArsiaEx
                                                                      3453 4.0
                                            4 BHK
                                                                      3600 5.0
      13316
                        Richards Town
                                                   Null Values
      13317
                Raja Rajeshwari Nagar
                                            2 BHK
                                                       Mahla T
                                                                      1141 2.0
                                                                      4689 4.0
      13318
                      Padmanabhanagar
                                            4 BHK
                                                        SollyCl
      13319
                          Doddathoguru
                                                   Null Values
                                                                       550 1.0
                                            1 BHK
            balcony
                      price
                      39.07
      0
                1.0
      1
                3.0
                     120.00
      2
                3.0
                      62.00
                1.0
                      95.00
      3
      4
                1.0
                      51.00
      13315
                0.0 231.00
                  1 400.00
      13316
      13317
                1.0
                      60.00
                1.0 488.00
      13318
      13319
                1.0
                      17.00
      [13320 rows x 10 columns]
[44]: df2.merge(df3,how='inner') #Inner Merge: Keeps only the rows with matching 'ID'
       ⇒in both dataframes
[44]: Empty DataFrame
      Columns: [ID, area_type, availability, location, size, society, total_sqft,
      bath, balcony, price]
      Index: []
[45]: df2.merge(df3,how='left') #Left Merge: Keeps all rows from df2 and matches from
       \hookrightarrow df3, missing values are NaN
[45]:
              ID
                              area_type
                                                availability \
      0
                  Super built-up Area 2024-12-19 00:00:00
```

```
1
         2
                      Plot
                            Area
                                         Ready To Move
2
         3
                  Built-up
                            Area
                                         Ready To Move
3
            Super built-up
                             Area
                                         Ready To Move
4
            Super built-up
                             Area
                                         Ready To Move
6655
     6656
            Super built-up
                                         Ready To Move
                            Area
                  Built-up
                                         Ready To Move
6656
      6657
                            Area
6657
      6658
                  Built-up
                            Area
                                         Ready To Move
6658
                                         Ready To Move
      6659
            Super built-up
                            Area
6659
      6660
                      Plot
                            Area
                                         Ready To Move
                      location
                                      size
                                                 society total_sqft bath \
      Electronic City Phase II
                                                 Coomee
0
                                     2 BHK
                                                               1056
                                                                     2.0
1
              Chikka Tirupathi 4 Bedroom
                                                 Theanmp
                                                               2600
                                                                     5.0
2
                   Uttarahalli
                                     3 ВНК
                                            Null Values
                                                               1440
                                                                     2.0
3
            Lingadheeranahalli
                                     3 BHK
                                                 Soiewre
                                                               1521
                                                                     3.0
4
                                     2 BHK
                      Kothanur
                                            Null Values
                                                               1200
                                                                     2.0
                                     1 BHK
6655
                Sarjapur Road
                                                 Ahriaup
                                                                539
                                                                     1.0
6656
                 Sarakki Nagar 4 Bedroom
                                            Null Values
                                                               1200
                                                                     4.0
6657
                                            Null Values
                                                               2400
                      Attibele
                                 3 Bedroom
                                                                     3.0
6658
                  Thubarahalli
                                     2 BHK
                                                 ShdhiSa
                                                               1200
                                                                     2.0
6659
                  Kundalahalli 4 Bedroom Null Values
                                                               2500
                                                                     5.0
     balcony
               price
0
         1.0
               39.07
         3.0 120.00
1
2
         3.0
               62.00
3
         1.0
               95.00
4
         1.0
               51.00
6655
         1.0
               45.00
6656
         1.0
               85.00
6657
         3.0 120.00
6658
         2.0
               79.55
6659
         2.0 350.00
```

[6660 rows x 10 columns]

[46]: df2.merge(df3,how='right') #Right Merge: Keeps all rows from df3 and matches_ from df2, missing values are NaN

[46]:		ID	area	_type	availab	ility	loc	ation	\
(0 66	61 Super	built-up	Area	Ready To	Move	Sarjapur	Road	
	1 66	62 Super	built-up	Area	Ready To	Move	Thanis	andra	
:	2 66	63	Plot	Area	Ready To	Move	Mahalakshmi	Puram	
	3 66	64	Plot	Area	Ready To	Move	Gott	igere	

```
4
       6665
             Super built-up
                                            Ready To Move
                                                                      Malleshwaram
                               Area
                                            Ready To Move
6655
      13316
                    Built-up
                               Area
                                                                        Whitefield
                                            Ready To Move
6656
      13317
             Super built-up
                                                                     Richards Town
                               Area
                    Built-up
                                            Ready To Move
                                                            Raja Rajeshwari Nagar
6657
      13318
                               Area
6658
      13319
             Super built-up
                                     2024-06-18 00:00:00
                                                                   Padmanabhanagar
                               Area
             Super built-up
                                            Ready To Move
                                                                      Doddathoguru
6659
      13320
                               Area
                      society total_sqft bath balcony
           size
                                                          price
0
          3 ВНК
                      MJinele
                                      1186
                                            2.0
                                                     2.0
                                                           40.0
1
          2 BHK
                      SLityee
                                            2.0
                                                     2.0
                                                           39.5
                                      1039
2
      4 Bedroom
                  Null Values
                                     1500
                                            5.0
                                                     3.0
                                                          233.0
3
      5 Bedroom
                  Null Values
                                     1500
                                            4.0
                                                     1.0
                                                          100.0
4
          3 ВНК
                      Maionra
                                      2520
                                            3.0
                                                     2.0
                                                          150.0
6655
      5 Bedroom
                      ArsiaEx
                                     3453
                                            4.0
                                                     0.0
                                                          231.0
                                            5.0
                                                          400.0
6656
          4 BHK
                  Null Values
                                     3600
                                                       1
6657
          2 BHK
                      Mahla T
                                            2.0
                                                     1.0
                                                           60.0
                                     1141
6658
          4 BHK
                      SollyCl
                                     4689
                                            4.0
                                                     1.0
                                                          488.0
6659
          1 BHK
                  Null Values
                                       550
                                            1.0
                                                     1.0
                                                           17.0
```

[6660 rows x 10 columns]

using .Join()

Using join to combine two dataframes based on their index

Setting 'ID' as the index for both dataframes and performing an outer join>

'lsuffix' and 'rsuffix' are used to handle overlapping column names.>

```
[47]:
                                       availability_Left
                                                                        location_Left \
                    area_type_Left
      ID
      1
              Super built-up
                               Area
                                     2024-12-19 00:00:00
                                                            Electronic City Phase II
      2
                        Plot
                                            Ready To Move
                                                                     Chikka Tirupathi
                               Area
      3
                    Built-up
                                            Ready To Move
                                                                          Uttarahalli
                               Area
      4
              Super built-up
                               Area
                                            Ready To Move
                                                                  Lingadheeranahalli
              Super built-up
                                            Ready To Move
                                                                             Kothanur
                               Area
      13316
                                NaN
                                                      NaN
                                                                                  NaN
      13317
                                NaN
                                                      NaN
                                                                                  NaN
      13318
                                NaN
                                                      NaN
                                                                                  NaN
      13319
                                NaN
                                                      NaN
                                                                                  NaN
      13320
                                NaN
                                                      NaN
                                                                                  NaN
```

```
size_Left society_Left total_sqft_Left bath_Left balcony_Left \
ID
1
            2 BHK
                        Coomee
                                             1056
                                                         2.0
                                                                       1.0
2
                                                                       3.0
       4 Bedroom
                        Theanmp
                                             2600
                                                         5.0
3
            3 BHK
                   Null Values
                                             1440
                                                         2.0
                                                                       3.0
            3 ВНК
                        Soiewre
                                                         3.0
4
                                             1521
                                                                       1.0
5
            2 BHK
                   Null Values
                                             1200
                                                         2.0
                                                                       1.0
13316
              NaN
                                                                       NaN
                            NaN
                                              NaN
                                                         NaN
13317
              NaN
                            NaN
                                              NaN
                                                         NaN
                                                                       NaN
                            NaN
13318
              NaN
                                              NaN
                                                         NaN
                                                                       NaN
13319
              NaN
                            NaN
                                              NaN
                                                         NaN
                                                                       NaN
              NaN
13320
                            NaN
                                              NaN
                                                         NaN
                                                                       NaN
                                               availabilityRight
       price_Left
                           area_typeRight
ID
1
             39.07
                                       NaN
                                                               NaN
2
            120.00
                                       NaN
                                                               NaN
3
             62.00
                                       NaN
                                                               NaN
4
             95.00
                                       NaN
                                                              NaN
5
             51.00
                                       NaN
                                                              NaN
13316
               NaN
                                                    Ready To Move
                           Built-up
                                      Area
                                                    Ready To Move
13317
                     Super built-up
               NaN
                                      Area
13318
                           Built-up
                                      Area
                                                    Ready To Move
               NaN
13319
               NaN
                     Super built-up
                                      Area
                                             2024-06-18 00:00:00
13320
                    Super built-up
                                                    Ready To Move
               NaN
                                      Area
                locationRight
                               sizeRight societyRight total_sqftRight
ID
1
                           NaN
                                       NaN
                                                      NaN
                                                                       NaN
2
                           NaN
                                       NaN
                                                      NaN
                                                                       NaN
3
                           NaN
                                       NaN
                                                      NaN
                                                                       NaN
4
                                                      NaN
                           NaN
                                       NaN
                                                                       NaN
5
                           NaN
                                       NaN
                                                      NaN
                                                                       NaN
                                                 ArsiaEx
13316
                   Whitefield
                                5 Bedroom
                                                                      3453
13317
                Richards Town
                                     4 BHK
                                             Null Values
                                                                      3600
13318
       Raja Rajeshwari Nagar
                                     2 BHK
                                                 Mahla T
                                                                      1141
              Padmanabhanagar
13319
                                     4 BHK
                                                 SollyCl
                                                                      4689
13320
                 Doddathoguru
                                     1 BHK
                                             Null Values
                                                                       550
      bathRight balconyRight priceRight
ID
1
             NaN
                           NaN
                                        NaN
2
             NaN
                           NaN
                                        NaN
3
                                        NaN
             NaN
                           NaN
```

4	NaN	NaN	NaN
5	NaN	NaN	NaN
•••	•••	•••	•••
13316	4.0	0.0	231.0
13317	5.0	1	400.0
13318	2.0	1.0	60.0
13319	4.0	1.0	488.0
13320	1.0	1.0	17.0
[13320	rows x 1	8 columns]	
100	160	1 (1751)	/ 100

[48]: df6 = df2.set_index('ID').join(df3.set_index('ID'), lsuffix= '_Left', rsuffix =__ df6

[48]: Empty DataFrame

Columns: [area_type_Left, availability_Left, location_Left, size_Left, society_Left, total_sqft_Left, bath_Left, balcony_Left, price_Left, area_typeRight, availabilityRight, locationRight, sizeRight, societyRight, total_sqftRight, bathRight, balconyRight, priceRight] Index: []

[49]: df6 = df2.set_index('ID').join(df3.set_index('ID'), lsuffix= '_Left', rsuffix =__ ¬'Right', how = 'left') df6

[49]:		area_type_L	eft availabil	Lity_Left	location	_Left \
ID)					
1	Super	built-up A	rea 2024-12-19	00:00:00 E	lectronic City Pha	se II
2		Plot A	rea Ready	7 To Move	Chikka Tiru	ıpathi
3		Built-up A	rea Ready	7 To Move	Uttara	halli
4	Super	built-up A	rea Ready	7 To Move	Lingadheerana	halli
5	Super	built-up A	rea Ready	7 To Move	Kot	hanur
•••		•••		•••	•••	
66	556 Super	built-up A	rea Ready	7 To Move	Sarjapur	Road
66	557	Built-up A	rea Ready	7 To Move	Sarakki	Nagar
66	58	Built-up A	rea Ready	7 To Move	Att	ibele
66	59 Super	built-up A	rea Ready	7 To Move	Thubara	halli
66	60	Plot A	rea Ready	7 To Move	Kundala	halli
			T ()			
		Left society	_Left total_sqff	L_Left bath_	Left balcony_Left	\
ID						
1			omee	1056	2.0 1.0	
2	4 Bed	room Th	eanmp	2600	5.0 3.0	
3	3	BHK Null V	alues	1440	2.0 3.0	
4	3	BHK So	iewre	1521	3.0 1.0	
5	2	BHK Null V	alues	1200	2.0 1.0	

```
6656
                                                   539
                                                                            1.0
                 1 BHK
                             Ahriaup
                                                              1.0
             4 Bedroom
                                                                            1.0
      6657
                         Null Values
                                                  1200
                                                              4.0
                                                              3.0
                                                                            3.0
      6658
             3 Bedroom
                         Null Values
                                                  2400
      6659
                 2 BHK
                             ShdhiSa
                                                  1200
                                                              2.0
                                                                            2.0
      6660
            4 Bedroom
                        Null Values
                                                  2500
                                                              5.0
                                                                            2.0
            price_Left area_typeRight availabilityRight locationRight sizeRight \
      ID
      1
                  39.07
                                     NaN
                                                        NaN
                                                                        NaN
                                                                                   NaN
      2
                 120.00
                                     NaN
                                                        NaN
                                                                        NaN
                                                                                   NaN
      3
                  62.00
                                     NaN
                                                        NaN
                                                                        NaN
                                                                                   NaN
                                                                        NaN
      4
                  95.00
                                     NaN
                                                        NaN
                                                                                   NaN
      5
                  51.00
                                     NaN
                                                        NaN
                                                                        NaN
                                                                                   NaN
                  45.00
      6656
                                     NaN
                                                        NaN
                                                                        NaN
                                                                                   NaN
      6657
                  85.00
                                     NaN
                                                        NaN
                                                                                   NaN
                                                                        NaN
      6658
                 120.00
                                     NaN
                                                        NaN
                                                                        NaN
                                                                                   NaN
      6659
                  79.55
                                     NaN
                                                                        NaN
                                                                                   NaN
                                                        NaN
      6660
                 350.00
                                     NaN
                                                        NaN
                                                                        NaN
                                                                                   NaN
            societyRight total_sqftRight bathRight balconyRight priceRight
      ID
      1
                     NaN
                                       NaN
                                                  NaN
                                                                NaN
                                                                             NaN
      2
                     NaN
                                       NaN
                                                  NaN
                                                                NaN
                                                                             NaN
                                                  NaN
      3
                     NaN
                                       NaN
                                                                NaN
                                                                             NaN
                                       NaN
                                                  NaN
                                                                NaN
      4
                     NaN
                                                                             NaN
      5
                     NaN
                                       NaN
                                                  NaN
                                                                NaN
                                                                             NaN
      6656
                     NaN
                                       {\tt NaN}
                                                  NaN
                                                                             NaN
                                                                NaN
      6657
                     NaN
                                       NaN
                                                  NaN
                                                                NaN
                                                                             NaN
      6658
                     NaN
                                       {\tt NaN}
                                                  NaN
                                                                NaN
                                                                             NaN
      6659
                     NaN
                                       NaN
                                                  NaN
                                                                NaN
                                                                             NaN
      6660
                     NaN
                                       NaN
                                                  NaN
                                                                NaN
                                                                             NaN
      [6660 rows x 18 columns]
[50]: df6 = df2.set_index('ID').join(df3.set_index('ID'), lsuffix= '_Left', rsuffix =__
       df6
[50]:
             area_type_Left availability_Left location_Left size_Left society_Left \
      ID
      6661
                         NaN
                                            NaN
                                                            NaN
                                                                       NaN
                                                                                     {\tt NaN}
      6662
                         NaN
                                                            NaN
                                                                       NaN
                                            NaN
                                                                                     NaN
      6663
                                                                       NaN
                         NaN
                                            NaN
                                                            NaN
                                                                                     NaN
      6664
                         NaN
                                            NaN
                                                            NaN
                                                                       NaN
                                                                                     NaN
```

6665	NaN	1	NaN	NaN	NaN	NaN	
	 No.N		 No.N	 N o N	···	MaM	
13316	NaN Na		NaN	NaN NaN	NaN	NaN	
13317	NaN Na		NaN	NaN	NaN	NaN	
13318	NaN NaN		NaN	NaN NaN	NaN	NaN	
13319	NaN		NaN N-N	NaN N-N	NaN N-N	NaN N-N	
13320	NaN	l	NaN	NaN	NaN	NaN	
	total_sqft_Lef	ft bath_Left	balcony_Left	price_Left	\		
ID							
6661	Na			NaN			
6662	Na		NaN	NaN			
6663	Na			NaN			
6664	Na		NaN	NaN			
6665	Na	aN NaN	NaN	NaN			
•••	•••	•••	•••	•			
13316	Na			NaN			
13317	Na			NaN			
13318	Na		NaN	NaN			
13319	Na	aN NaN	NaN	NaN			
13320	Na	aN NaN	NaN	NaN			
	area_ty	peRight a	availabilityRig	ght	locationRight	5 \	
ID	a 1 13.		D 1		a		
6661	Super built-u	-	Ready To Mo		Sarjapur Road		
6662	Super built-u	-	Ready To Mo		Thanisandra		
6663	Plo		Ready To Mo		nalakshmi Puran		
6664	Plo		Ready To Mo		Gottigere		
6665	Super built-u	ıp Area	Ready To Mo	ove	Malleshwaran	n	
	5	•••					
13316		ıp Area	Ready To Mo		Whitefield		
13317	Super built-u	-	Ready To Mo		Richards Town		
13318		-	•	-	ajeshwari Nagan		
	-	-	24-06-18 00:00:		Padmanabhanagan		
13320	Super built-u	ıp Area	Ready To Mo	ove	Doddathogurı	1	
	sizeRight soc	rietvRight to	otal_sqftRight	hathRight h	nalconvRight \		
ID	Sizenight Boo	recyntight c	orar_sqr citigite	bacintigne i	diconymight	`	
6661	3 ВНК	MJinele	1186	2.0	2.0		
6662	2 BHK	SLityee	1039	2.0	2.0		
6663		ill Values	1500	5.0	3.0		
6664		ill Values	1500	4.0	1.0		
6665	3 BHK	Maionra	2520	3.0	2.0		
0000	O DIII	naroma			2.0		
 13316	 5 Bedroom	… ArsiaEx	3453	4.0	0.0		
13317		ill Values	3600	5.0	1		
13317	2 BHK	Mahla T	1141	2.0	1.0		
10010		riania 1	1141	2.0	1.0		

```
4 BHK
                                                                            SollyCl
                                                                                                                                   550
                                                                                                                                                               1.0
                                                                                                                                                                                                  1.0
                13320
                                              1 BHK Null Values
                                   priceRight
                ID
                6661
                                                   40.0
                6662
                                                   39.5
                6663
                                                 233.0
                6664
                                                 100.0
                6665
                                                 150.0
                                                 •••
                13316
                                                 231.0
                13317
                                                 400.0
                13318
                                                   60.0
                13319
                                                 488.0
                13320
                                                   17.0
                [6660 rows x 18 columns]
              Using .concat()
[51]: conc2 = pd.concat([df2,df3], axis = 0) #This adds rows of df3 below the rows of the concatangle of the
                  ⇔df2. Columns that don't match will have NaN values
                conc2
                                                                                                     #If axis=1 is used, the dataframes are_
                    ⇔concatenated side-by-side (column-wise)
[51]:
                                            ID
                                                                                                                                      availability \
                                                                                    area_type
                                              1
                                                      Super built-up Area
                                                                                                                  2024-12-19 00:00:00
                                              2
                1
                                                                                  Plot
                                                                                                  Area
                                                                                                                                   Ready To Move
                2
                                              3
                                                                       Built-up
                                                                                                                                   Ready To Move
                                                                                                  Area
                3
                                              4
                                                      Super built-up
                                                                                                                                   Ready To Move
                                                                                                Area
                                              5
                                                      Super built-up
                                                                                                                                   Ready To Move
                                                                                                 Area
                13315 13316
                                                                       Built-up
                                                                                                                                   Ready To Move
                                                                                                Area
                13316 13317
                                                     Super built-up Area
                                                                                                                                   Ready To Move
                                                                       Built-up
                                                                                                Area
                                                                                                                                   Ready To Move
                13317
                                  13318
                                                      Super built-up Area
                                                                                                                  2024-06-18 00:00:00
                13318
                                 13319
                13319
                                                      Super built-up Area
                                                                                                                                   Ready To Move
                                  13320
                                                                                                                                                    society total_sqft bath \
                                                                               location
                                                                                                                        size
                                                                                                                                                                                          1056 2.0
                0
                                   Electronic City Phase II
                                                                                                                                                    Coomee
                                                                                                                      2 BHK
                                                                                                                                                                                          2600 5.0
                1
                                                         Chikka Tirupathi
                                                                                                        4 Bedroom
                                                                                                                                                    Theanmp
                2
                                                                       Uttarahalli
                                                                                                                                        Null Values
                                                                                                                                                                                          1440 2.0
                                                                                                                      3 BHK
                3
                                                   Lingadheeranahalli
                                                                                                                      3 ВНК
                                                                                                                                                    Soiewre
                                                                                                                                                                                          1521 3.0
                4
                                                                               Kothanur
                                                                                                                      2 BHK
                                                                                                                                        Null Values
                                                                                                                                                                                          1200 2.0
                13315
                                                                         Whitefield 5 Bedroom
                                                                                                                                                   ArsiaEx
                                                                                                                                                                                          3453 4.0
```

4689

4.0

1.0

13319

```
3600 5.0
      13316
                        Richards Town
                                           4 BHK Null Values
      13317
                Raja Rajeshwari Nagar
                                            2 BHK
                                                       Mahla T
                                                                     1141 2.0
      13318
                      Padmanabhanagar
                                           4 BHK
                                                       SollyCl
                                                                     4689 4.0
      13319
                         Doddathoguru
                                                  Null Values
                                                                      550 1.0
                                           1 BHK
            balcony
                      price
      0
                1.0
                      39.07
      1
                3.0 120.00
      2
                      62.00
                3.0
      3
                1.0
                      95.00
      4
                1.0
                      51.00
      13315
                0.0 231.00
                  1 400.00
      13316
      13317
                1.0
                      60.00
      13318
                1.0 488.00
      13319
                      17.00
                1.0
      [13320 rows x 10 columns]
[52]: numerical_data = df1['price'] #adding a new numerical column in dataset df1
      print(numerical_data)
     0
               39.07
     1
              120.00
     2
               62.00
     3
               95.00
     4
               51.00
     13315
              231.00
     13316
              400.00
     13317
               60.00
     13318
              488.00
     13319
               17.00
     Name: price, Length: 13320, dtype: float64
[53]: mean = numerical_data.mean() #using .mean() to find mean of numerical_data
      print(mean)
     112.5656265015015
[54]: mode = numerical_data.mode() #using .mode() to find mode of numerical_data
      print(mode)
     0
          75.0
     Name: price, dtype: float64
```

```
[61]:
                 ID
                                                    availability \
                                 area_type
      408
                409
                     Super built-up
                                      Area
                                            2024-01-19 00:00:00
      605
                606
                     Super built-up
                                             2024-01-19 00:00:00
                                      Area
      2623
               2624
                                Plot
                                      Area
                                             2024-07-18 00:00:00
      3180
              3181
                     Super built-up
                                                   Ready To Move
                                      Area
      4162
              4163
                           Built-up
                                      Area
                                                   Ready To Move
      6421
              6422
                                Plot
                                      Area
                                            2024-09-18 00:00:00
      7315
                                      Area
                                             2024-04-18 00:00:00
              7316
                                Plot
      8546
              8547
                                Plot
                                      Area
                                                   Ready To Move
      10304
             10305
                                Plot
                                      Area
                                                   Ready To Move
      11080
             11081
                     Super built-up
                                             2024-01-18 00:00:00
                                      Area
      11763
             11764
                                Plot
                                      Area
                                                   Ready To Move
      12443
             12444
                                Plot
                                                   Ready To Move
                                      Area
      13067
              13068
                                Plot
                                                   Ready To Move
                                      Area
      13197
              13198
                                Plot
                                      Area
                                                   Ready To Move
      13200
             13201
                                Plot
                                      Area
                                                   Ready To Move
                           location
                                            size
                                                       society total_sqft
                                                                            bath
      408
                       Rajaji Nagar
                                           7 BHK
                                                   Null Values
                                                                     12000
                                                                              6.0
      605
                       Malleshwaram
                                           7 BHK
                                                   Null Values
                                                                     12000
                                                                              7.0
      2623
                   Dodsworth Layout
                                                   Null Values
                                                                     30000
                                                                              4.0
                                       4 Bedroom
                                                       Kierser
      3180
                    Shanthala Nagar
                                           5 BHK
                                                                      8321
                                                                              5.0
                             Yemlur
                                                                              5.0
      4162
                                       4 Bedroom
                                                       Epllan
                                                                      7000
      6421
                       Bommenahalli
                                       4 Bedroom
                                                       Prood G
                                                                      2940
                                                                              3.0
      7315
                              Yemlur
                                       4 Bedroom
                                                       Epllan
                                                                     11000
                                                                              5.0
      8546
                     Dollars Colony
                                       3 Bedroom
                                                   Null Values
                                                                      7800
                                                                              3.0
                5th Block Jayanagar
                                       4 Bedroom
                                                   Null Values
                                                                              4.0
      10304
                                                                     10624
      11080
                        Ashok Nagar
                                           4 BHK
                                                   Null Values
                                                                      8321
                                                                              5.0
      11763
                    Sadashiva Nagar
                                       5 Bedroom
                                                   Null Values
                                                                      9600
                                                                              7.0
                     Dollars Colony
                                       4 Bedroom
                                                   Null Values
                                                                      4350
                                                                              8.0
      12443
      13067
                     Defence Colony
                                      10 Bedroom
                                                   Null Values
                                                                      7150
                                                                             13.0
             Ramakrishnappa Layout
                                       4 Bedroom
                                                   Null Values
                                                                      9200
      13197
                                                                              4.0
      13200
                     Defence Colony
                                       6 Bedroom Null Values
                                                                      8000
                                                                              6.0
```

balcony price 408 3.0 2200.0 605 3.0 2200.0 2623 1 2100.0

```
3180
                3.0 2700.0
      4162
                  1 2050.0
      6421
                2.0
                     2250.0
      7315
                3.0
                     2000.0
      8546
                2.0 2000.0
      10304
                2.0
                     2340.0
      11080
                2.0 2912.0
      11763
                2.0 2736.0
                  1 2600.0
      12443
      13067
                     3600.0
                  1
      13197
                  1
                     2600.0
      13200
                3.0 2800.0
[64]: size_mapping = {
       '2 BHK': '2 BHK',
       '3 BHK': '3 BHK',
       '4 Bedroom': '4 BHK',
       '4 BHK': '4 BHK',
       '5 Bedroom': '5 BHK',
       '6 BHK': '6 BHK'
      }
      df1['size'] = df1['size'].map(size_mapping)
      df1
[64]:
                ID
                               area_type
                                                  availability \
      0
                 1
                    Super built-up Area
                                          2024-12-19 00:00:00
                 2
                              Plot
      1
                                    Area
                                                Ready To Move
      2
                 3
                          Built-up
                                    Area
                                                Ready To Move
      3
                 4
                    Super built-up
                                                Ready To Move
                                    Area
                 5
                    Super built-up
                                                Ready To Move
                                    Area
                                                Ready To Move
      13315 13316
                          Built-up
                                    Area
      13316 13317
                    Super built-up Area
                                                Ready To Move
      13317
            13318
                          Built-up Area
                                                Ready To Move
                    Super built-up Area 2024-06-18 00:00:00
      13318
            13319
      13319 13320
                    Super built-up Area
                                                Ready To Move
                                                  society total_sqft bath balcony \
                             location
                                        size
      0
             Electronic City Phase II
                                       2 BHK
                                                  Coomee
                                                                 1056
                                                                       2.0
                                                                               1.0
      1
                     Chikka Tirupathi 4 BHK
                                                  Theanmp
                                                                 2600
                                                                       5.0
                                                                               3.0
      2
                          Uttarahalli 3 BHK
                                              Null Values
                                                                 1440
                                                                       2.0
                                                                               3.0
      3
                   Lingadheeranahalli 3 BHK
                                                  Soiewre
                                                                 1521
                                                                       3.0
                                                                               1.0
                                                                       2.0
      4
                             Kothanur 2 BHK
                                              Null Values
                                                                 1200
                                                                               1.0
      13315
                           Whitefield 5 BHK
                                                                 3453
                                                                       4.0
                                                                               0.0
                                                  ArsiaEx
                                                                 3600
                                                                       5.0
      13316
                        Richards Town 4 BHK
                                              Null Values
                                                                                 1
                Raja Rajeshwari Nagar 2 BHK
                                                  Mahla T
                                                                 1141
                                                                       2.0
                                                                               1.0
      13317
```

```
13318
                                                  SollyCl
                                                                      4.0
                      Padmanabhanagar 4 BHK
                                                                4689
                                                                              1.0
      13319
                         Doddathoguru
                                         NaN Null Values
                                                                 550 1.0
                                                                              1.0
              price
      0
              39.07
      1
             120.00
      2
              62.00
      3
              95.00
      4
              51.00
      13315 231.00
      13316 400.00
      13317
              60.00
      13318 488.00
      13319
             17.00
      [13320 rows x 10 columns]
[65]: import matplotlib.pyplot as plt
      grouped = df1.groupby('size')['price'].mean().reset_index()
      plt.bar(grouped['size'], grouped['price'], color='green')
      plt.xlabel('Property Size (BHK/Bedroom)')
      plt.ylabel('Average Price (in Lakhs)')
      plt.title('Average Price by Property Size')
      plt.show()
```

