



+	<table><tr><td>Student Name</td><td>Aarya Trifale</td></tr><tr><td>SRN No</td><td>31241458</td></tr><tr><td>Roll No</td><td>41</td></tr><tr><td>Program</td><td>Computer Engineering</td></tr><tr><td>Year</td><td>Second Year</td></tr><tr><td>Division</td><td>E</td></tr><tr><td>Subject</td><td>Fundamental Of Data science</td></tr><tr><td>Assignment No</td><td>2</td></tr></table>	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
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		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	4	Super built-up	Area		Ready To Move	
4	5	Super built-up	Area		Ready To Move	
...	
13315	13316	Built-up	Area		Ready To Move	
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	size	society	total_sqft	bath	balcony	\
0		Electronic City Phase II	2 BHK	Coomee	1056	2.0	1.0	
1		Chikka Tirupathi	4 Bedroom	Theanmp	2600	5.0	3.0	
2		Uttarahalli	3 BHK	NaN	1440	2.0	3.0	
3		Lingadheeranahalli	3 BHK	Soiewre	1521	3.0	1.0	

4	Kothanur	2 BHK	NaN	1200	2.0	1.0
...
13315	Whitefield	5 Bedroom	ArsiaEx	3453	4.0	0.0
13316	Richards Town	4 BHK	NaN	3600	5.0	NaN
13317	Raja Rajeshwari Nagar	2 BHK	Mahla T	1141	2.0	1.0
13318	Padmanabhanagar	4 BHK	SollyCl	4689	4.0	1.0
13319	Doddathoguru	1 BHK	NaN	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]

[30]: `file.head()` *#.head() returns the first few rows (the "head" of the dataframe i. e. excel file)*

[30]:	ID	area_type	availability	location \
0	1	Super built-up Area	2024-12-19 00:00:00	Electronic City Phase II
1	2	Plot Area	Ready To Move	Chikka Tirupathi
2	3	Built-up Area	Ready To Move	Uttarahalli
3	4	Super built-up Area	Ready To Move	Lingadheeranahalli
4	5	Super built-up Area	Ready To Move	Kothanur

	size	society	total_sqft	bath	balcony	price
0	2 BHK	Coomee	1056	2.0	1.0	39.07
1	4 Bedroom	Theanmp	2600	5.0	3.0	120.00
2	3 BHK	NaN	1440	2.0	3.0	62.00
3	3 BHK	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]:	ID	area_type	availability \
13315	13316	Built-up Area	Ready To Move
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      size society total_sqft  bath  balcony \
13315      Whitefield  5 Bedroom  ArsiaEx      3453   4.0     0.0
13316      Richards Town    4 BHK      NaN      3600   5.0     NaN
13317 Raja Rajeshwari Nagar    2 BHK  Mahla T      1141   2.0     1.0
13318      Padmanabhanagar    4 BHK  SollyCl      4689   4.0     1.0
13319      Doddathoguru     1 BHK      NaN       550   1.0     1.0

      price
13315  231.0
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]:
count  13320.000000  13247.000000  12711.000000  13320.000000
mean    6660.500000    2.692610    1.584376    112.565627
std     3845.297128    1.341458    0.817263    148.971674
min       1.000000    1.000000    0.000000     8.000000
25%     3330.750000    2.000000    1.000000    50.000000
50%     6660.500000    2.000000    2.000000    72.000000
75%     9990.250000    3.000000    2.000000   120.000000
max    13320.000000   40.000000    3.000000   3600.000000

```

```
[33]: file.isnull().sum() #this will return the number of missing values in dataset
```

```

[33]: ID          0
      area_type    0
      availability  0
      location     1
      size        16
      society     5502
      total_sqft   0
      bath        73
      balcony     609
      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']

```

```
[34]: 0      Coomee
      1      Theanmp
      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 BHK
      3      3 BHK
      4      2 BHK
      ...
      13315     5 Bedroom
      13316     4 BHK
      13317     2 BHK
      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
      1      5.0
      2      2.0
      3      3.0
      4      2.0
      ...
      13315     4.0
      13316     5.0
      13317     2.0
      13318     4.0
      13319     1.0
      Name: bath, Length: 13320, dtype: object
```

```
[37]: file['balcony'] = file['balcony'].fillna('1')
      file['balcony']
```

```
[37]: 0      1.0
      1      3.0
      2      3.0
      3      1.0
      4      1.0
      ...
      13315  0.0
      13316   1
      13317  1.0
      13318  1.0
      13319  1.0
      Name: balcony, Length: 13320, dtype: object
```

```
[38]: file['location'] = file['location'].fillna('pune')
      file['location']
```

```
[38]: 0      Electronic City Phase II
      1      Chikka Tirupathi
      2      Uttarahalli
      3      Lingadheeranahalli
      4      Kothanur
      ...
      13315      Whitefield
      13316      Richards Town
      13317      Raja Rajeshwari Nagar
      13318      Padmanabhanagar
      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      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      4  Super built-up  Area      Ready To Move
4      5  Super built-up  Area      Ready To Move
...    ...      ...
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	3 BHK	Null Values	1440	2.0	
3	Lingadheeranahalli	3 BHK	Soiewre	1521	3.0	
4	Kothanur	2 BHK	Null Values	1200	2.0	
...
6655	Sarjapur Road	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	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]

[57]: df3

[57]:

	ID	area_type	availability	\
6660	6661	Super built-up Area	Ready To Move	
6661	6662	Super built-up Area	Ready To Move	
6662	6663	Plot Area	Ready To Move	
6663	6664	Plot Area	Ready To Move	
6664	6665	Super built-up Area	Ready To Move	
...
13315	13316	Built-up Area	Ready To Move	
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	size	society	total_sqft	bath	balcony	\
----------	------	---------	------------	------	---------	---

6660	Sarjapur Road	3 BHK	MJinele	1186	2.0	2.0
6661	Thanisandra	2 BHK	SLityee	1039	2.0	2.0
6662	Mahalakshmi Puram	4 Bedroom	Null Values	1500	5.0	3.0
6663	Gottigere	5 Bedroom	Null Values	1500	4.0	1.0
6664	Malleshwaram	3 BHK	Maionra	2520	3.0	2.0
...
13315	Whitefield	5 Bedroom	ArsiaEx	3453	4.0	0.0
13316	Richards Town	4 BHK	Null Values	3600	5.0	1
13317	Raja Rajeshwari Nagar	2 BHK	Mahla T	1141	2.0	1.0
13318	Padmanabhanagar	4 BHK	SollyCl	4689	4.0	1.0
13319	Doddathoguru	1 BHK	Null Values	550	1.0	1.0

price

6660	40.0
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	4	Super built-up Area	Ready To Move
4	5	Super built-up Area	Ready To Move
...

13315	13316	Built-up	Area	Ready To Move
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	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	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	
13316	Richards Town	4 BHK	Null Values	3600	5.0	
13317	Raja Rajeshwari Nagar	2 BHK	Mahla T	1141	2.0	
13318	Padmanabhanagar	4 BHK	SollyCl	4689	4.0	
13319	Doddathoguru	1 BHK	Null Values	550	1.0	

	balcony	price
0	1.0	39.07
1	3.0	120.00
2	3.0	62.00
3	1.0	95.00
4	1.0	51.00
...
13315	0.0	231.00
13316	1	400.00
13317	1.0	60.00
13318	1.0	488.00
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
      ↪ df3, missing values are NaN
```

```
[45]:      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	4	Super built-up	Area	Ready To Move
4	5	Super built-up	Area	Ready To Move
...
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	3 BHK	Null Values	1440	2.0
3	Lingadheeranahalli	3 BHK	Soiewre	1521	3.0
4	Kothanur	2 BHK	Null Values	1200	2.0
...
6655	Sarjapur Road	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	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	availability	location \
0	6661	Super built-up Area	Ready To Move	Sarjapur Road
1	6662	Super built-up Area	Ready To Move	Thanisandra
2	6663	Plot Area	Ready To Move	Mahalakshmi Puram
3	6664	Plot Area	Ready To Move	Gottigere

4	6665	Super built-up	Area	Ready To Move	Malleshwaram
...
6655	13316	Built-up	Area	Ready To Move	Whitefield
6656	13317	Super built-up	Area	Ready To Move	Richards Town
6657	13318	Built-up	Area	Ready To Move	Raja Rajeshwari Nagar
6658	13319	Super built-up	Area	2024-06-18 00:00:00	Padmanabhanagar
6659	13320	Super built-up	Area	Ready To Move	Doddathoguru

	size	society	total_sqft	bath	balcony	price
0	3 BHK	MJinele	1186	2.0	2.0	40.0
1	2 BHK	SLityee	1039	2.0	2.0	39.5
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 BHK	Maionra	2520	3.0	2.0	150.0
...
6655	5 Bedroom	ArsiaEx	3453	4.0	0.0	231.0
6656	4 BHK	Null Values	3600	5.0	1	400.0
6657	2 BHK	Mahla T	1141	2.0	1.0	60.0
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]: df6 = df2.set_index('ID').join(df3.set_index('ID'), lsuffix= '_Left', rsuffix = '_Right', how = 'outer')
df6
```

```
[47]:
```

	area_type_Left	availability_Left	location_Left \
ID			
1	Super built-up Area	2024-12-19 00:00:00	Electronic City Phase II
2	Plot Area	Ready To Move	Chikka Tirupathi
3	Built-up Area	Ready To Move	Uttarahalli
4	Super built-up Area	Ready To Move	Lingadheeranahalli
5	Super built-up Area	Ready To Move	Kothanur
...
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	4 Bedroom	Theanmp	2600	5.0	3.0	
3	3 BHK	Null Values	1440	2.0	3.0	
4	3 BHK	Soiewre	1521	3.0	1.0	
5	2 BHK	Null Values	1200	2.0	1.0	
...	
13316	NaN	NaN	NaN	NaN	NaN	NaN
13317	NaN	NaN	NaN	NaN	NaN	NaN
13318	NaN	NaN	NaN	NaN	NaN	NaN
13319	NaN	NaN	NaN	NaN	NaN	NaN
13320	NaN	NaN	NaN	NaN	NaN	NaN

	price_Left	area_typeRight	availabilityRight	\
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	Built-up Area	Ready To Move	
13317	NaN	Super built-up Area	Ready To Move	
13318	NaN	Built-up Area	Ready To Move	
13319	NaN	Super built-up Area	2024-06-18 00:00:00	
13320	NaN	Super built-up Area	Ready To Move	

	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	
...	
13316	Whitefield	5 Bedroom	ArsiaEx	3453	
13317	Richards Town	4 BHK	Null Values	3600	
13318	Raja Rajeshwari Nagar	2 BHK	Mahla T	1141	
13319	Padmanabhanagar	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 18 columns]

```
[48]: df6 = df2.set_index('ID').join(df3.set_index('ID'), lsuffix= '_Left', rsuffix =
      ↪'_Right', how = 'inner')
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_Left	availability_Left	location_Left	\
ID				
1	Super built-up Area	2024-12-19 00:00:00	Electronic City Phase II	
2	Plot Area	Ready To Move	Chikka Tirupathi	
3	Built-up Area	Ready To Move	Uttarahalli	
4	Super built-up Area	Ready To Move	Lingadheeranahalli	
5	Super built-up Area	Ready To Move	Kothanur	
...	
6656	Super built-up Area	Ready To Move	Sarjapur Road	
6657	Built-up Area	Ready To Move	Sarakki Nagar	
6658	Built-up Area	Ready To Move	Attibele	
6659	Super built-up Area	Ready To Move	Thubarahalli	
6660	Plot Area	Ready To Move	Kundalahalli	

	size_Left	society_Left	total_sqft_Left	bath_Left	balcony_Left	\
ID						
1	2 BHK	Coomee	1056	2.0	1.0	
2	4 Bedroom	Theanmp	2600	5.0	3.0	
3	3 BHK	Null Values	1440	2.0	3.0	
4	3 BHK	Soiewre	1521	3.0	1.0	
5	2 BHK	Null Values	1200	2.0	1.0	

...
6656	1 BHK	Ahriaup	539	1.0	1.0	
6657	4 Bedroom	Null Values	1200	4.0	1.0	
6658	3 Bedroom	Null Values	2400	3.0	3.0	
6659	2 BHK	ShdhiSa	1200	2.0	2.0	
6660	4 Bedroom	Null Values	2500	5.0	2.0	

ID	price_Left	area_typeRight	availabilityRight	locationRight	sizeRight	\
1	39.07	NaN	NaN	NaN	NaN	
2	120.00	NaN	NaN	NaN	NaN	
3	62.00	NaN	NaN	NaN	NaN	
4	95.00	NaN	NaN	NaN	NaN	
5	51.00	NaN	NaN	NaN	NaN	
...	
6656	45.00	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	

ID	societyRight	total_sqftRight	bathRight	balconyRight	priceRight
1	NaN	NaN	NaN	NaN	NaN
2	NaN	NaN	NaN	NaN	NaN
3	NaN	NaN	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN	NaN
5	NaN	NaN	NaN	NaN	NaN
...
6656	NaN	NaN	NaN	NaN	NaN
6657	NaN	NaN	NaN	NaN	NaN
6658	NaN	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 =
    ↳ 'Right', how = 'right')
df6
```

```
[50]: area_type_Left availability_Left location_Left size_Left society_Left \
ID
6661      NaN      NaN      NaN      NaN      NaN
6662      NaN      NaN      NaN      NaN      NaN
6663      NaN      NaN      NaN      NaN      NaN
6664      NaN      NaN      NaN      NaN      NaN
```

6665	NaN	NaN	NaN	NaN	NaN
...
13316	NaN	NaN	NaN	NaN	NaN
13317	NaN	NaN	NaN	NaN	NaN
13318	NaN	NaN	NaN	NaN	NaN
13319	NaN	NaN	NaN	NaN	NaN
13320	NaN	NaN	NaN	NaN	NaN

	total_sqft_Left	bath_Left	balcony_Left	price_Left	\
ID					
6661	NaN	NaN	NaN	NaN	
6662	NaN	NaN	NaN	NaN	
6663	NaN	NaN	NaN	NaN	
6664	NaN	NaN	NaN	NaN	
6665	NaN	NaN	NaN	NaN	
...	
13316	NaN	NaN	NaN	NaN	
13317	NaN	NaN	NaN	NaN	
13318	NaN	NaN	NaN	NaN	
13319	NaN	NaN	NaN	NaN	
13320	NaN	NaN	NaN	NaN	

	area_typeRight	availabilityRight	locationRight	\
ID				
6661	Super built-up Area	Ready To Move	Sarjapur Road	
6662	Super built-up Area	Ready To Move	Thanisandra	
6663	Plot Area	Ready To Move	Mahalakshmi Puram	
6664	Plot Area	Ready To Move	Gottigere	
6665	Super built-up Area	Ready To Move	Malleswaram	
...	
13316	Built-up Area	Ready To Move	Whitefield	
13317	Super built-up Area	Ready To Move	Richards Town	
13318	Built-up Area	Ready To Move	Raja Rajeshwari Nagar	
13319	Super built-up Area	2024-06-18 00:00:00	Padmanabhanagar	
13320	Super built-up Area	Ready To Move	Doddathoguru	

	sizeRight	societyRight	total_sqftRight	bathRight	balconyRight	\
ID						
6661	3 BHK	MJinele	1186	2.0	2.0	
6662	2 BHK	SLityee	1039	2.0	2.0	
6663	4 Bedroom	Null Values	1500	5.0	3.0	
6664	5 Bedroom	Null Values	1500	4.0	1.0	
6665	3 BHK	Maionra	2520	3.0	2.0	
...	
13316	5 Bedroom	ArsiaEx	3453	4.0	0.0	
13317	4 BHK	Null Values	3600	5.0	1	
13318	2 BHK	Mahla T	1141	2.0	1.0	

13319	4 BHK	SollyCl	4689	4.0	1.0
13320	1 BHK	Null Values	550	1.0	1.0

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
      ↪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	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	4	Super built-up Area	Ready To Move
4	5	Super built-up Area	Ready To Move
...
13315	13316	Built-up Area	Ready To Move
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	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	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

13316	Richards Town	4 BHK	Null Values	3600	5.0
13317	Raja Rajeshwari Nagar	2 BHK	Mahla T	1141	2.0
13318	Padmanabhanagar	4 BHK	SollyCl	4689	4.0
13319	Doddathoguru	1 BHK	Null Values	550	1.0

	balcony	price
0	1.0	39.07
1	3.0	120.00
2	3.0	62.00
3	1.0	95.00
4	1.0	51.00
...
13315	0.0	231.00
13316	1	400.00
13317	1.0	60.00
13318	1.0	488.00
13319	1.0	17.00

[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

```
[55]: median = numerical_data.median() #using .median() to find median of
      ↪numerical_data
      print(median)
```

72.0

PART - 2

```
[61]: df1.loc[df1.price >= 2000]
```

```
[61]:
```

	ID		area_type		availability	\
	408	409	Super built-up	Area	2024-01-19 00:00:00	
	605	606	Super built-up	Area	2024-01-19 00:00:00	
	2623	2624	Plot	Area	2024-07-18 00:00:00	
	3180	3181	Super built-up	Area	Ready To Move	
	4162	4163	Built-up	Area	Ready To Move	
	6421	6422	Plot	Area	2024-09-18 00:00:00	
	7315	7316	Plot	Area	2024-04-18 00:00:00	
	8546	8547	Plot	Area	Ready To Move	
	10304	10305	Plot	Area	Ready To Move	
	11080	11081	Super built-up	Area	2024-01-18 00:00:00	
	11763	11764	Plot	Area	Ready To Move	
	12443	12444	Plot	Area	Ready To Move	
	13067	13068	Plot	Area	Ready To Move	
	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	4 Bedroom	Null Values	30000	4.0	
	3180	Shanthala Nagar	5 BHK	Kierser	8321	5.0	
	4162	Yemlur	4 Bedroom	Epplan	7000	5.0	
	6421	Bommenahalli	4 Bedroom	Prood G	2940	3.0	
	7315	Yemlur	4 Bedroom	Epplan	11000	5.0	
	8546	Dollars Colony	3 Bedroom	Null Values	7800	3.0	
	10304	5th Block Jayanagar	4 Bedroom	Null Values	10624	4.0	
	11080	Ashok Nagar	4 BHK	Null Values	8321	5.0	
	11763	Sadashiva Nagar	5 Bedroom	Null Values	9600	7.0	
	12443	Dollars Colony	4 Bedroom	Null Values	4350	8.0	
	13067	Defence Colony	10 Bedroom	Null Values	7150	13.0	
	13197	Ramakrishnappa Layout	4 Bedroom	Null Values	9200	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
12443	1	2600.0
13067	1	3600.0
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
1	2		Plot	Area	Ready To Move
2	3		Built-up	Area	Ready To Move
3	4	Super	built-up	Area	Ready To Move
4	5	Super	built-up	Area	Ready To Move
...
13315	13316		Built-up	Area	Ready To Move
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	size	society	total_sqft	bath	balcony \
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
4		Kothanur	2 BHK	Null Values	1200	2.0	1.0
...	
13315		Whitefield	5 BHK	ArsiaEx	3453	4.0	0.0
13316		Richards Town	4 BHK	Null Values	3600	5.0	1
13317		Raja Rajeshwari Nagar	2 BHK	Mahla T	1141	2.0	1.0

13318	Padmanabhanagar	4 BHK	SollyCl	4689	4.0	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()
```

