

Name : Sufiyan Khan
Roll No. : A12
Class : TE - AI & DS

Batch : A1

EXPERIMENT 4

- Aim: a) Solving exercises in Data Exploration
b) Solving exercises in Data Preprocessing

Data Preprocessing:

Code & output:

```
dwmexp4.ipynb
File Edit View Run Kernel Tabs Settings Help

[8]: import pandas as pd
import os

[9]: df=pd.read_csv('house.csv')
df.head()

[9]:
```

	area_type	availability	location	size	society	total_sqft	bath	balcony	price
0	Super built-up Area	19-Dec	Electronic City Phase II	2 BHK	Coomee	1056	2.0	1.0	39.07
1	Plot Area	Ready To Move	Chikka Tirupathi	4 Bedroom	Theanmp	2600	5.0	3.0	120.00
2	Built-up Area	Ready To Move	Uttarahalli	3 BHK	NaN	1440	2.0	3.0	62.00
3	Super built-up Area	Ready To Move	Lingadheeranahalli	3 BHK	Soiewre	1521	3.0	1.0	95.00
4	Super built-up Area	Ready To Move	Kothanur	2 BHK	NaN	1200	2.0	1.0	51.00

```
[10]: df.shape
[10]: (13320, 9)

[11]: df.isnull()

[11]:
```

	area_type	availability	location	size	society	total_sqft	bath	balcony	price
0	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False
2	False	False	False	False	True	False	False	False	False
3	False	False	False	False	False	False	False	False	False
4	False	False	False	False	True	False	False	False	False

mple 0 1 Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 22 dwmexp4.ipynb 1

Name : Sufiyan Khan

Roll No. : A12

Class : TE - AI & DS

Batch : A1

dwmexp4.ipynb

File Edit View Run Kernel Tabs Settings Help

Python 3 (ipykernel)

```
[12]: df.isnull().sum()
```

```
[12]: area_type    0
availability    0
location        1
size           16
society        5502
total_sqft      0
bath           73
balcony        609
price           0
dtype: int64
```

```
[13]: df.isnull().sum().sum()
```

```
[13]: 6201
```

filling null values

```
[14]: df2=df.fillna(value=0)
df2
```

```
[14]:
```

	area_type	availability	location	size	society	total_sqft	bath	balcony	price
0	Super built-up Area	19-Dec	Electronic City Phase II	2 BHK	Coomee	1056	2.0	1.0	39.07
1	Plot Area	Ready To Move	Chikka Tirupathi	4 Bedroom	Theanmp	2600	5.0	3.0	120.00
2	Built-up Area	Ready To Move	Uttarahalli	3 BHK	0	1440	2.0	3.0	62.00
3	Super built-up Area	Ready To Move	Lingadheeranahalli	3 BHK	Soiewre	1521	3.0	1.0	95.00
4	Super built-up Area	Ready To Move	Kothanur	2 BHK	0	1200	2.0	1.0	51.00
...

Python 3 (ipykernel) Idle

Mode: Command | Ln 1, Col 12 | dwmexp4.ipynb | 1

dwmexp4.ipynb

File Edit View Run Kernel Tabs Settings Help

Python 3 (ipykernel)

```
13319 Super built-up Area Ready To Move Doddathoguru 1 BHK 0 550 1.0 1.0 17.00
```

13320 rows x 9 columns

```
[15]: df2.isnull().sum().sum()
```

```
[15]: 0
```

```
[16]: df3=df.fillna(value=5)
df3
```

```
[16]:
```

	area_type	availability	location	size	society	total_sqft	bath	balcony	price
0	Super built-up Area	19-Dec	Electronic City Phase II	2 BHK	Coomee	1056	2.0	1.0	39.07
1	Plot Area	Ready To Move	Chikka Tirupathi	4 Bedroom	Theanmp	2600	5.0	3.0	120.00
2	Built-up Area	Ready To Move	Uttarahalli	3 BHK	5	1440	2.0	3.0	62.00
3	Super built-up Area	Ready To Move	Lingadheeranahalli	3 BHK	Soiewre	1521	3.0	1.0	95.00
4	Super built-up Area	Ready To Move	Kothanur	2 BHK	5	1200	2.0	1.0	51.00
...
13315	Built-up Area	Ready To Move	Whitefield	5 Bedroom	ArsiaEx	3453	4.0	0.0	231.00
13316	Super built-up Area	Ready To Move	Richards Town	4 BHK	5	3600	5.0	5.0	400.00
13317	Built-up Area	Ready To Move	Raja Rajeshwari Nagar	2 BHK	Mahla T	1141	2.0	1.0	60.00
13318	Super built-up Area	18-Jun	Padmanabhanagar	4 BHK	SollyCI	4689	4.0	1.0	488.00
13319	Super built-up Area	Ready To Move	Doddathoguru	1 BHK	5	550	1.0	1.0	17.00

13320 rows x 9 columns

Python 3 (ipykernel) Idle

Mode: Command | Ln 1, Col 12 | dwmexp4.ipynb | 1

Name : Sufiyan Khan

Roll No. : A12

Class : TE - AI & DS

Batch : A1

dwmexp4.ipynb

File Edit View Run Kernel Tabs Settings Help

Notebook Python 3 (ipykernel)

```
[17]: df4=df.fillna(value=df['balcony'].mean())
df4
```

2	Built-up Area	Ready To Move	Uttarahalli	3 BHK	1.584376	1440	2.0	3.000000	62.00
3	Super built-up Area	Ready To Move	Lingadheeranahalli	3 BHK	Soiewre	1521	3.0	1.000000	95.00
4	Super built-up Area	Ready To Move	Kothanur	2 BHK	1.584376	1200	2.0	1.000000	51.00
...
13315	Built-up Area	Ready To Move	Whitefield	5 Bedroom	ArsiaEx	3453	4.0	0.000000	231.00
13316	Super built-up Area	Ready To Move	Richards Town	4 BHK	1.584376	3600	5.0	1.584376	400.00
13317	Built-up Area	Ready To Move	Raja Rajeshwari Nagar	2 BHK	Mahla T	1141	2.0	1.000000	60.00
13318	Super built-up Area	18-Jun	Padmanabhanagar	4 BHK	SollyCI	4689	4.0	1.000000	488.00
13319	Super built-up Area	Ready To Move	Doddathoguru	1 BHK	1.584376	550	1.0	1.000000	17.00

13320 rows x 9 columns

```
##dropping null values

[18]: df5=df.dropna(how='all')
df5
```

```
[18]:
```

	area_type	availability	location	size	society	total_sqft	bath	balcony	price
0	Super built-up Area	19-Dec	Electronic City Phase II	2 BHK	Coomee	1056	2.0	1.0	39.07
1	Plot Area	Ready To Move	Chikka Tirupathi	4 Bedroom	Theanmp	2600	5.0	3.0	120.00
2	Built-up Area	Ready To Move	Uttarahalli	3 BHK	NaN	1440	2.0	3.0	62.00
3	Super built-up Area	Ready To Move	Lingadheeranahalli	3 BHK	Soiewre	1521	3.0	1.0	95.00

simple 0 1 Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 12 dwmexp4.ipynb 1

dwmexp4.ipynb

File Edit View Run Kernel Tabs Settings Help

Notebook Python 3 (ipykernel)

```
##dropping null values

[18]: df5=df.dropna(how='all')
df5
```

```
[18]:
```

	area_type	availability	location	size	society	total_sqft	bath	balcony	price
0	Super built-up Area	19-Dec	Electronic City Phase II	2 BHK	Coomee	1056	2.0	1.0	39.07
1	Plot Area	Ready To Move	Chikka Tirupathi	4 Bedroom	Theanmp	2600	5.0	3.0	120.00
2	Built-up Area	Ready To Move	Uttarahalli	3 BHK	NaN	1440	2.0	3.0	62.00
3	Super built-up Area	Ready To Move	Lingadheeranahalli	3 BHK	Soiewre	1521	3.0	1.0	95.00
4	Super built-up Area	Ready To Move	Kothanur	2 BHK	NaN	1200	2.0	1.0	51.00
...
13315	Built-up Area	Ready To Move	Whitefield	5 Bedroom	ArsiaEx	3453	4.0	0.0	231.00
13316	Super built-up Area	Ready To Move	Richards Town	4 BHK	NaN	3600	5.0	NaN	400.00
13317	Built-up Area	Ready To Move	Raja Rajeshwari Nagar	2 BHK	Mahla T	1141	2.0	1.0	60.00
13318	Super built-up Area	18-Jun	Padmanabhanagar	4 BHK	SollyCI	4689	4.0	1.0	488.00
13319	Super built-up Area	Ready To Move	Doddathoguru	1 BHK	NaN	550	1.0	1.0	17.00

13320 rows x 9 columns

```
[19]: df5=df.dropna(how='any')
df5
```

```
[19]:
```

	area_type	availability	location	size	society	total_sqft	bath	balcony	price
0	Super built-up Area	19-Dec	Electronic City Phase II	2 BHK	Coomee	1056	2.0	1.0	39.07

Name : Sufiyan Khan

Roll No. : A12

Class : TE - AI & DS

Batch : A1

dwmxp4.ipynb

File Edit View Run Kernel Tabs Settings Help

13320 rows x 9 columns

Notebook Python 3 (ipykernel)

```
[19]: df5=df.dropna(how='any')
df5
```

```
[19]:
```

	area_type	availability	location	size	society	total_sqft	bath	balcony	price
0	Super built-up Area	19-Dec	Electronic City Phase II	2 BHK	Coomee	1056	2.0	1.0	39.07
1	Plot Area	Ready To Move	Chikka Tirupathi	4 Bedroom	Theanmp	2600	5.0	3.0	120.00
3	Super built-up Area	Ready To Move	Lingadheeranahalli	3 BHK	Soiewre	1521	3.0	1.0	95.00
5	Super built-up Area	Ready To Move	Whitefield	2 BHK	DuenaTa	1170	2.0	1.0	38.00
11	Plot Area	Ready To Move	Whitefield	4 Bedroom	Priry M	2785	5.0	3.0	295.00
...
13313	Super built-up Area	Ready To Move	Uttarahalli	3 BHK	Akila R	1345	2.0	1.0	57.00
13314	Super built-up Area	Ready To Move	Green Glen Layout	3 BHK	SoosePr	1715	3.0	3.0	112.00
13315	Built-up Area	Ready To Move	Whitefield	5 Bedroom	ArsiaEx	3453	4.0	0.0	231.00
13317	Built-up Area	Ready To Move	Raja Rajeshwari Nagar	2 BHK	Mahla T	1141	2.0	1.0	60.00
13318	Super built-up Area	18-Jun	Padmanabhanagar	4 BHK	SollyCl	4689	4.0	1.0	488.00

7496 rows x 9 columns

```
[20]: import numpy as np
```

```
[21]: df6=df.replace(to_replace=np.nan, value=2122)
df6
```

```
[21]:
```

	area_type	availability	location	size	society	total_sqft	bath	balcony	price
0	Super built-up Area	19-Dec	Electronic City Phase II	2 BHK	Coomee	1056	2.0	1.0	39.07
1	Plot Area	Ready To Move	Chikka Tirupathi	4 Bedroom	Theanmp	2600	5.0	3.0	120.00
2	Built-up Area	Ready To Move	Uttarahalli	3 BHK	2122	1440	2.0	3.0	62.00
3	Super built-up Area	Ready To Move	Lingadheeranahalli	3 BHK	Soiewre	1521	3.0	1.0	95.00
4	Super built-up Area	Ready To Move	Kothanur	2 BHK	2122	1200	2.0	1.0	51.00
...
13315	Built-up Area	Ready To Move	Whitefield	5 Bedroom	ArsiaEx	3453	4.0	0.0	231.00
13316	Super built-up Area	Ready To Move	Richards Town	4 BHK	2122	3600	5.0	2122.0	400.00
13317	Built-up Area	Ready To Move	Raja Rajeshwari Nagar	2 BHK	Mahla T	1141	2.0	1.0	60.00
13318	Super built-up Area	18-Jun	Padmanabhanagar	4 BHK	SollyCl	4689	4.0	1.0	488.00
13319	Super built-up Area	Ready To Move	Doddathoguru	1 BHK	2122	550	1.0	1.0	17.00

13320 rows x 9 columns

Data Exploration:

Code & output:

Name : Sufiyan Khan

Roll No. : A12

Class : TE - AI & DS

Batch : A1

dwmexp4.ipynb

File Edit View Run Kernel Tabs Settings Help

Notebook Python 3 (ipykernel)

	area_type	availability	location	bhk	bath	total_sqft	area	balcony	price
0	Super built-up Area	19-Dec	Electronic City Phase II	2 BHK	Coomee	1056	2.0	1.0	39.07
1	Plot Area	Ready To Move	Chikka Tirupathi	4 Bedroom	Theanmp	2600	5.0	3.0	120.00
2	Built-up Area	Ready To Move	Uttarahalli	3 BHK	2122	1440	2.0	3.0	62.00
3	Super built-up Area	Ready To Move	Lingadheeranahalli	3 BHK	Soiewre	1521	3.0	1.0	95.00
4	Super built-up Area	Ready To Move	Kothanur	2 BHK	2122	1200	2.0	1.0	51.00
...
13315	Built-up Area	Ready To Move	Whitefield	5 Bedroom	ArsiaEx	3453	4.0	0.0	231.00
13316	Super built-up Area	Ready To Move	Richards Town	4 BHK	2122	3600	5.0	2122.0	400.00
13317	Built-up Area	Ready To Move	Raja Rajeshwari Nagar	2 BHK	Mahla T	1141	2.0	1.0	60.00
13318	Super built-up Area	18-Jun	Padmanabhanagar	4 BHK	SollyCl	4689	4.0	1.0	488.00
13319	Super built-up Area	Ready To Move	Doddathoguru	1 BHK	2122	550	1.0	1.0	17.00

13320 rows x 9 columns

```
[36]: df6.bath.mean()
```

```
[36]: 14.307432432432432
```

```
[31]: df6.balcony.median()
```

```
[31]: 2.0
```

```
[32]: df6.total_sqft.mode()
```

```
[32]: 0    1200
      Name: total_sqft, dtype: object
```

dwmexp4.ipynb

File Edit View Run Kernel Tabs Settings Help

Notebook Python 3 (ipykernel)

2	Built-up Area	Ready To Move	Uttarahalli	3 BHK	2122	1440	2.0	3.0	62.00
3	Super built-up Area	Ready To Move	Lingadheeranahalli	3 BHK	Soiewre	1521	3.0	1.0	95.00
4	Super built-up Area	Ready To Move	Kothanur	2 BHK	2122	1200	2.0	1.0	51.00
...
13315	Built-up Area	Ready To Move	Whitefield	5 Bedroom	ArsiaEx	3453	4.0	0.0	231.00
13316	Super built-up Area	Ready To Move	Richards Town	4 BHK	2122	3600	5.0	2122.0	400.00
13317	Built-up Area	Ready To Move	Raja Rajeshwari Nagar	2 BHK	Mahla T	1141	2.0	1.0	60.00
13318	Super built-up Area	18-Jun	Padmanabhanagar	4 BHK	SollyCl	4689	4.0	1.0	488.00
13319	Super built-up Area	Ready To Move	Doddathoguru	1 BHK	2122	550	1.0	1.0	17.00

13320 rows x 9 columns

```
[36]: df6.bath.mean()
```

```
[36]: 14.307432432432432
```

```
[31]: df6.balcony.median()
```

```
[31]: 2.0
```

```
[32]: df6.total_sqft.mode()
```

```
[32]: 0    1200
      Name: total_sqft, dtype: object
```

```
[33]: df6.bath.std()
```

```
[33]: 156.47397009048743
```

Name : Sufiyan Khan

Roll No. : A12

Class : TE - AI & DS

Batch : A1

Conclusion: In conclusion, the DWM experiment on Data Exploration and Data Preprocessing highlighted the importance of understanding and preparing data effectively. Through data exploration, we uncovered patterns and anomalies, while data preprocessing ensured clean, reliable data for analysis. These steps are vital for informed decision-making and building accurate models, forming a strong basis for our data science skills going forward.