Bengalore House Price prediction

Import Dataset

In [1]:
```
import pandas as pd
```

In [2]:
```
df=pd.read_csv('bengaluru_data.csv')
df.head()
```

Out[2]:

	area_type	availability	location	size	society	total_sqft	bath	balcony	pr
0	Super built-up Area	19-Dec	Electronic City Phase II	2 BHK	Coomee	1056	2.0	1.0	39
1	Plot Area	Ready To Move	Chikka Tirupathi	4 Bedroom	Theanmp	2600	5.0	3.0	120
2	Built-up Area	Ready To Move	Uttarahalli	3 BHK	NaN	1440	2.0	3.0	62
3	Super built-up Area	Ready To Move	Lingadheeranahalli	3 BHK	Soiewre	1521	3.0	1.0	95
4	Super built-up Area	Ready To Move	Kothanur	2 BHK	NaN	1200	2.0	1.0	51

In [3]:
```
df.shape
```

Out[3]:
```
(13320, 9)
```

In [4]:
```
df.isnull()
```

Out[4]:

	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
...	...	...	...	...	...	...	...	...	...
13315	False	False	False	False	False	False	False	False	False
13316	False	False	False	False	True	False	False	True	False
13317	False	False	False	False	False	False	False	False	False
13318	False	False	False	False	False	False	False	False	False
13319	False	False	False	False	True	False	False	False	False

13320 rows × 9 columns

In [5]:
```
df.isnull().sum()
```

Out[5]:
```
area_type          0
availability       0
location           1
size              16
society         5502
total_sqft         0
bath              73
balcony          609
price              0
dtype: int64
```

In [6]:
```
df.isnull().sum().sum()
```

Out[6]:
```
6201
```

Filling null values

In [7]:
```
df2=df.fillna(value=0)
df2
```

Out[7]:

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

13320 rows × 9 columns

In [8]:
```
df.isnull().sum().sum()
```

Out[8]:
```
6201
```

In [9]:
```
df3=df.fillna(value=5)
df3
```

Out[9]:

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

13320 rows × 9 columns

In [10]:
```
#Filling null value with a previous value
df4=df.fillna(method='pad')
df4
```

Out[10]:

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

13320 rows × 9 columns

In [11]:
```
df4.isnull().sum()
```

Out[11]:
```
area_type       0
availability    0
location        0
size            0
society         0
total_sqft      0
bath            0
balcony         0
price           0
dtype: int64
```

In [12]:
```
#Filling null value with the next value
df5=df.fillna(method='bfill')
df5
```

Out[12]:

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

13320 rows × 9 columns

In [13]:
```
df6=df.fillna(method='pad',axis=1)
df6
```

Out[13]:

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

13320 rows × 9 columns

In [14]:
```
df
```

Out[14]:

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

13320 rows × 9 columns

In [15]:
```
df7=df.fillna(method='bfill',axis=1)
df7
```

Out[15]:

	area_type	availability	location	size	society	total_sqft	bath	balcony
0	Super built-up Area	19-Dec	Electronic City Phase II	2 BHK	Coomee	1056	2.0	1.0
1	Plot Area	Ready To Move	Chikka Tirupathi	4 Bedroom	Theanmp	2600	5.0	3.0
2	Built-up Area	Ready To Move	Uttarahalli	3 BHK	1440	1440	2.0	3.0
3	Super built-up Area	Ready To Move	Lingadheeranahalli	3 BHK	Soiewre	1521	3.0	1.0
4	Super built-up Area	Ready To Move	Kothanur	2 BHK	1200	1200	2.0	1.0
...	...	...	...	...	...	...	...	...
13315	Built-up Area	Ready To Move	Whitefield	5 Bedroom	ArsiaEx	3453	4.0	0.0
13316	Super built-up Area	Ready To Move	Richards Town	4 BHK	3600	3600	5.0	400.0
13317	Built-up Area	Ready To Move	Raja Rajeshwari Nagar	2 BHK	Mahla T	1141	2.0	1.0
13318	Super built-up Area	18-Jun	Padmanabhanagar	4 BHK	SollyCl	4689	4.0	1.0
13319	Super built-up Area	Ready To Move	Doddathoguru	1 BHK	550	550	1.0	1.0

13320 rows × 9 columns

In [16]:
```
df
```

Out[16]:

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

13320 rows × 9 columns

In [17]:
```
# Filling different values in null in different columns
df8=df.fillna({'society':'abcd',
               'balcony':'efgh'})
df8
```

Out[17]:

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

13320 rows × 9 columns

In [18]:
```
#Filling null value with the mean of a column
df9=df.fillna(value=df['balcony'].mean())
df9
```

Out[18]:

	area_type	availability	location	size	society	total_sqft	bath	balcony
0	Super built-up Area	19-Dec	Electronic City Phase II	2 BHK	Coomee	1056	2.0	1.000000
1	Plot Area	Ready To Move	Chikka Tirupathi	4 Bedroom	Theanmp	2600	5.0	3.000000
2	Built-up Area	Ready To Move	Uttarahalli	3 BHK	1.584376	1440	2.0	3.000000
3	Super built-up Area	Ready To Move	Lingadheeranahalli	3 BHK	Soiewre	1521	3.0	1.000000
4	Super built-up Area	Ready To Move	Kothanur	2 BHK	1.584376	1200	2.0	1.000000
...	...	...	...	...	...	...	...	...
13315	Built-up Area	Ready To Move	Whitefield	5 Bedroom	ArsiaEx	3453	4.0	0.000000
13316	Super built-up Area	Ready To Move	Richards Town	4 BHK	1.584376	3600	5.0	1.584376
13317	Built-up Area	Ready To Move	Raja Rajeshwari Nagar	2 BHK	Mahla T	1141	2.0	1.000000
13318	Super built-up Area	18-Jun	Padmanabhanagar	4 BHK	SollyCl	4689	4.0	1.000000
13319	Super built-up Area	Ready To Move	Doddathoguru	1 BHK	1.584376	550	1.0	1.000000

13320 rows × 9 columns

Dropna()

In [19]:
```
df
```

Out[19]:

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

13320 rows × 9 columns

In [20]:
```
df10=df.dropna()
df10
```

Out[20]:

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

7496 rows × 9 columns

In [21]:
```
df11=df.dropna(how='all')
df11
```

Out[21]:

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

13320 rows × 9 columns

replace()

In [22]:
```
import numpy as np
df12=df.replace(to_replace=np.nan,value=4568932)
df12
```

Out[22]:

	area_type	availability	location	size	society	total_sqft	bath	balcon
0	Super built-up Area	19-Dec	Electronic City Phase II	2 BHK	Coomee	1056	2.0	1.
1	Plot Area	Ready To Move	Chikka Tirupathi	4 Bedroom	Theanmp	2600	5.0	3.
2	Built-up Area	Ready To Move	Uttarahalli	3 BHK	4568932	1440	2.0	3.
3	Super built-up Area	Ready To Move	Lingadheeranahalli	3 BHK	Soiewre	1521	3.0	1.
4	Super built-up Area	Ready To Move	Kothanur	2 BHK	4568932	1200	2.0	1.
...	...	...	...	...	...	...	...	
13315	Built-up Area	Ready To Move	Whitefield	5 Bedroom	ArsiaEx	3453	4.0	0.
13316	Super built-up Area	Ready To Move	Richards Town	4 BHK	4568932	3600	5.0	4568932.
13317	Built-up Area	Ready To Move	Raja Rajeshwari Nagar	2 BHK	Mahla T	1141	2.0	1.
13318	Super built-up Area	18-Jun	Padmanabhanagar	4 BHK	SollyCl	4689	4.0	1.
13319	Super built-up Area	Ready To Move	Doddathoguru	1 BHK	4568932	550	1.0	1.

13320 rows × 9 columns

In [23]:
```
df13=df.replace(to_replace=3.0,value=5.0)
df13
```

Out[23]:

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

13320 rows × 9 columns

interpolate()

In [24]:
```
df[' balcony']=df['balcony'].interpolate(method='linear')
df
```

Out[24]:

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

13320 rows × 10 columns

In [25]:
```
df[' balcony']=df['balcony'].interpolate(method='linear',limit_direction='fo
df
```

Out[25]:

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

13320 rows × 10 columns

Outliers

In [26]:
```
# Importing
import sklearn
from sklearn.datasets import load_diabetes
import pandas as pd
import seaborn as sns
import warnings
warnings.filterwarnings("ignore")
```

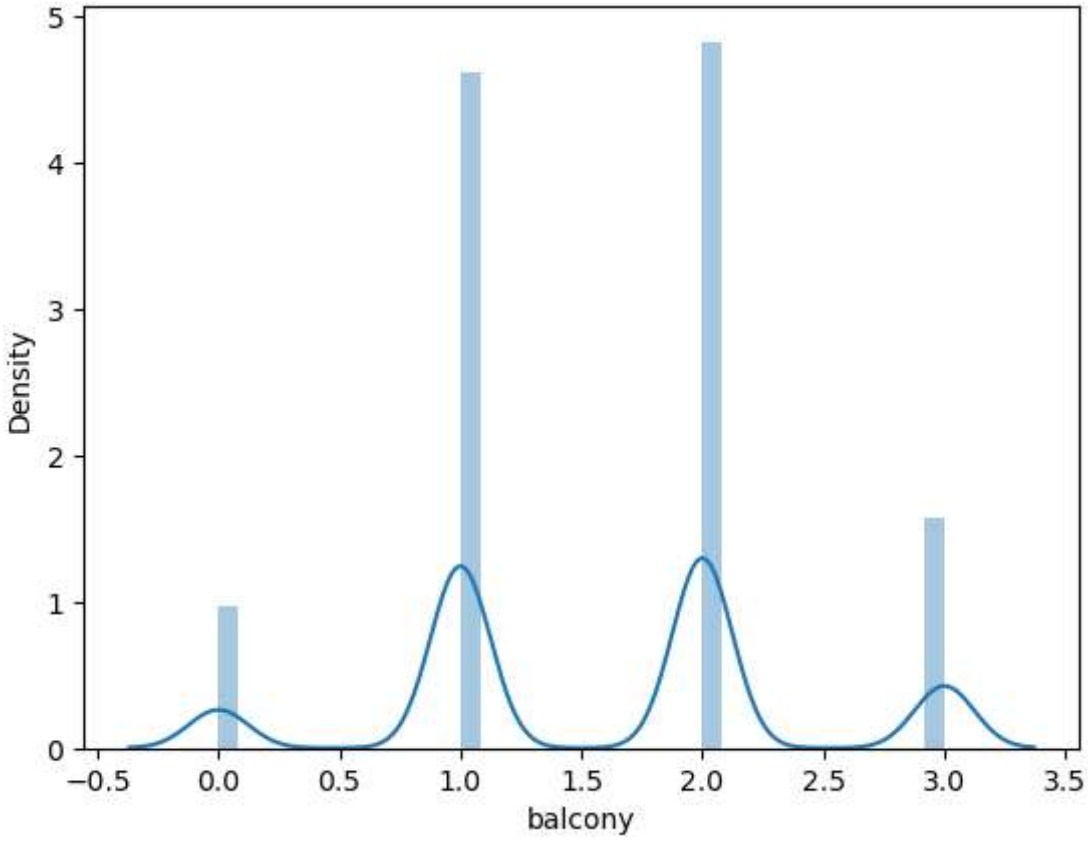
In [27]:
```
df.describe()
```

Out[27]:

	bath	balcony	price	balcony
count	13247.000000	12711.000000	13320.000000	13320.000000
mean	2.692610	1.584376	112.565627	1.584872
std	1.341458	0.817263	148.971674	0.808889
min	1.000000	0.000000	8.000000	0.000000
25%	2.000000	1.000000	50.000000	1.000000
50%	2.000000	2.000000	72.000000	2.000000
75%	3.000000	2.000000	120.000000	2.000000
max	40.000000	3.000000	3600.000000	3.000000

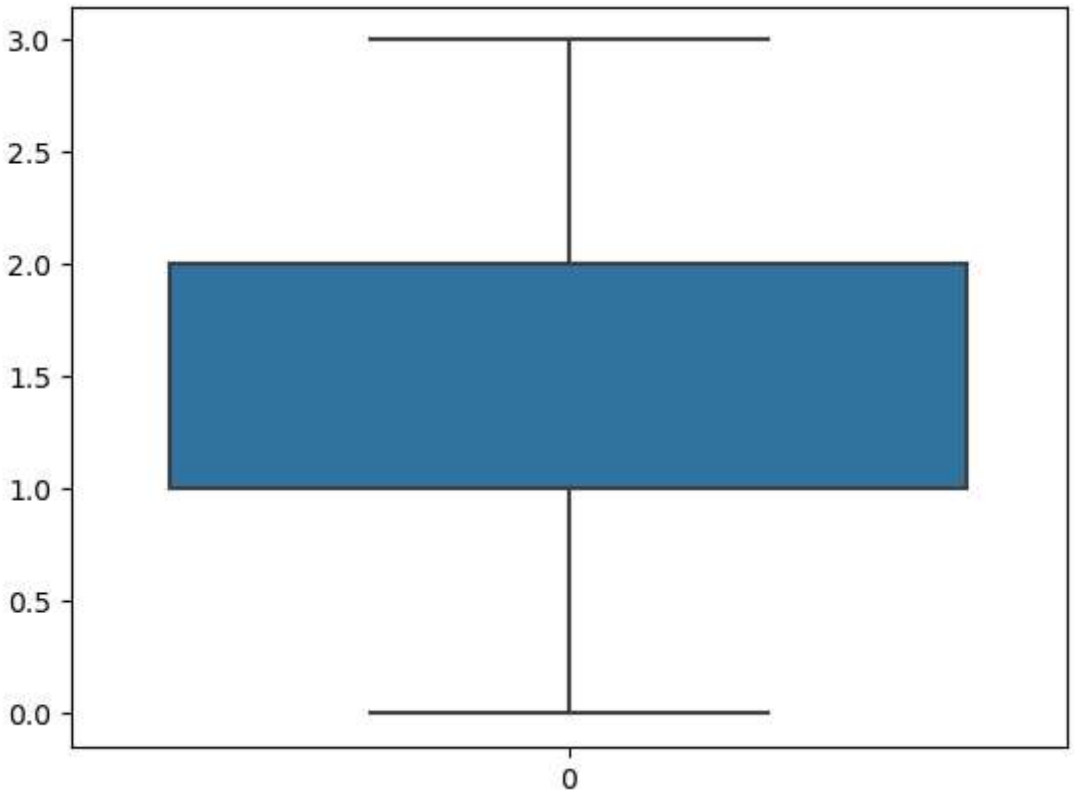
In [28]:
```
sns.distplot(df['balcony'])
```

Out[28]:
```
<Axes: xlabel='balcony', ylabel='Density'>
```



In [29]:
```
#to see outliers clearly
sns.boxplot(df['balcony'])
```

Out[29]: <Axes: >



IQR method

In [37]:
```
q1=df['balcony'].quantile(0.25)
q3=df['balcony'].quantile(0.75)
iqr=q3-q1
q1,q3,iqr
```

Out[37]: (1.0, 2.0, 1.0)

In [38]:
```

upper_limit=q3+(1.5*iqr)
lower_limit=q1-(1.5*iqr)
lower_limit,upper_limit
```

Out[38]: (-0.5, 3.5)

In [39]:
```
sns.boxplot(df['balcony'])
```

Out[39]:
```
<Axes: >
```