Handling Null values

we remove null values because it will create problem in machine learning model

In [4]:
```
import pandas as pd
import seaborn as sns
```

In [3]:
```
df=pd.read_csv('Bengaluru_House_Data.csv')
```

In [6]:
```
df.head(2)
```

Out[6]:

	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

first we will check where nulls values are there and how much

In [7]:
```
sns.heatmap(df.isnull())
```

Out[7]:

In [10]:
```
df.isnull()

#it will give bulean value
```

Out[10]:

	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 [11]:
```
df.isnull().sum()

#we are checking column wise
```

Out[11]:
```
area_type        0
availability     0
location         1
size            16
society       5502
total_sqft       0
bath            73
balcony        609
price            0
dtype: int64
```

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

#we are checking total
```

Out[12]:
```
6201
```

In [ ]:
```
#we can drop/replace/fill this null value
```

filing null values

we are creating new dataset df2 so that orginal is not effected

we can put any value , here we putting 0

In [13]:
```
df2=df.fillna(value=0)
```

In [14]:
```
sns.heatmap(df2.isnull())
```

Out[14]:

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

Out[15]:
```
0
```

In [18]:
```
#filling null vlaue with privious value row wise by using method =pad

df4=df.fillna(method='pad')
df4.head()
```

Out[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	Theanmp	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	Soiewre	1200	2.0	1.0	51.00

In [19]:
```
#filling null vlaue with next value row wise by using method =bfill

df5=df.fillna(method='bfill')
df5.head()
```

Out[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
2	Built-up Area	Ready To Move	Uttarahalli	3 BHK	Soiewre	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	DuenaTa	1200	2.0	1.0	51.00

In [20]:
```
#filling null vlaue with next value column wise by using method

df6=df.fillna(method='bfill' , axis=1)
df6.head()
```

Out[20]:

	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	1440	1440	2.0	3.0	62.0
3	Super built-up Area	Ready To Move	Lingadheeranahalli	3 BHK	Soiewre	1521	3.0	1.0	95.0
4	Super built-up Area	Ready To Move	Kothanur	2 BHK	1200	1200	2.0	1.0	51.0

In [21]:
```
#filling null vlaue with next value column wise by using method

df7=df.fillna(method='pad' , axis=1)
df7.head()
```

Out[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	3 BHK	1440	2.0	3.0	62.0
3	Super built-up Area	Ready To Move	Lingadheeranahalli	3 BHK	Soiewre	1521	3.0	1.0	95.0
4	Super built-up Area	Ready To Move	Kothanur	2 BHK	2 BHK	1200	2.0	1.0	51.0

In [23]:
```
#filling diff values in NULl in diff column , by using dict {}

df8=df.fillna({'society': 'abc' , 'balcony': 'xyz'})
df8.head()
```

Out[23]:

	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	abc	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	abc	1200	2.0	1.0	51.00

In [25]:
```
#filling null vlaue with the 'mean' of a column

df9=df.fillna(value=df['balcony'].mean())
df9.head()
```

Out[25]:

	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	1.584376	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	1.584376	1200	2.0	1.0	51.00

In [27]:
```
#filling null vlaue with the 'max' value of a column
#we can also use min value

df0=df.fillna(value=df['balcony'].max())
df0.head()
```

Out[27]:

	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	3.0	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	3.0	1200	2.0	1.0	51.00

drop

In [29]:
```
#this will remove all row which consist of even single null value

df10=df.dropna()
df10.head()
```

Out[29]:

	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	Prrry M	2785	5.0	3.0	295.00

In [30]:
```
#'how' is inbuild parameter which has 2(any , all) prameter
#any- it will remove row if t has even 1 null value
#all- all will remove row if it has all null values


df11=df.dropna(how='any')
df11.head()
```

Out[30]:

	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	Prrry M	2785	5.0	3.0	295.00

In [31]:
```
df12=df.dropna(how='all')
df12.head()
```

Out[31]:

	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

replace

it is like fillna but it has some more functions, it can be used to replace other value also

In [35]:
```
df13=df.replace(to_replace =3.0 ,value= 5.0)
df13.head()

#here we have replace 3.0 by 5.0
```

Out[35]:

	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	5.0	120.00
2	Built-up Area	Ready To Move	Uttarahalli	3 BHK	NaN	1440	2.0	5.0	62.00
3	Super built-up Area	Ready To Move	Lingadheeranahalli	3 BHK	Soiewre	1521	5.0	1.0	95.00
4	Super built-up Area	Ready To Move	Kothanur	2 BHK	NaN	1200	2.0	1.0	51.00

interpolate()

it will interpret value accourding to the column

In [39]:
```
df['balcony']=df['balcony'].interpolate(method='linear')
df
```

Out[39]:

	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	0.5	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	NaN	550	1.0	1.0	17.00

13320 rows × 9 columns

In [ ]:
```

```