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 df=pd.read\_csv('Bengaluru\_House\_Data.csv') In [3]: df.head(2)In [6]: availability location society total\_sqft bath balcony price Out[6]: area\_type size 0 Super built-up Area 19-Dec Electronic City Phase II 2 BHK Coomee 1056 2.0 1.0 39.07 Chikka Tirupathi 4 Bedroom Theanmp 3.0 120.00 Plot Area Ready To Move 2600 5.0 first we will check where nulls values are there and how much sns.heatmap(df.isnull()) In [7]: <AxesSubplot:> Out[7]: - 1.0 635 1270 1905 2540 3175 3810 4445 5080 5715 6350 6985 7620 8255 8890 9525 10160 10795 11430 12065 12700 - 0.8 - 0.6 - 0.4 - 0.2 wailability df.isnull() In [10]: #it will give bulean value total\_sqft bath Out[10]: area\_type availability location size society balcony price 0 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 False False True False True 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 df.isnull().sum() #we are checking column wise area\_type Out[11]: availability 0 location 1 size 16 society 5502 total\_sqft 0 73 bath balcony 609 price dtype: int64 df.isnull().sum().sum() In [12]: #we are checking total 6201 Out[12]: #we can drop/replace/fill this null value In [ ] filing null values we are creating new dataset df2 so that orginal is not effected we can put any value, here we putting 0 df2=df.fillna(value=0) In [13]: sns.heatmap(df2.isnull()) <AxesSubplot:> Out[14]: 0 - 635 - 1270 - 1905 - 2540 - 3175 - 3810 - 4445 - 5080 - 5715 - 6350 - 6985 - 7620 - 8255 - 8890 - 9525 - 10160 - 10795 - 11430 - 12065 - 12700 -0.100 -0.075 -0.050 -0.025 -0.000 -0.025-0.050-0.075 -0.100availability bath society df2.isnull().sum().sum() In [15]: Out[15]: #filling null vlaue with privious value row wise by using method =pad df4=df.fillna(method='pad') df4.head() availability location society total\_sqft bath balcony price Out[18]: area\_type size 0 Super built-up Area Electronic City Phase II 2 BHK Plot Area Ready To Move Chikka Tirupathi 4 Bedroom Theanmp 5.0 3.0 120.00 2600 2 Uttarahalli 1440 2.0 62.00 Built-up Area Ready To Move 3 BHK Theanmp 3.0 Lingadheeranahalli 95.00 3 Super built-up Area Ready To Move 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 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 Chikka Tirupathi 4 Bedroom Theanmp Plot Area Ready To Move 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 #filling null vlaue with next value column wise by using method In [20]: df6=df.fillna(method='bfill' , axis=1) df6.head() area\_type Out[20]: availability location size society total\_sqft bath balcony 1.0 39.07 19-Dec Electronic City Phase II 1056 2.0 Super built-up Area 2 BHK Coomee Plot Area Ready To Move Chikka Tirupathi 4 Bedroom Theanmp 2600 5.0 3.0 120.0 2 Uttarahalli **3 BHK** 2.0 62.0 Built-up Area Ready To Move 1440 1440 3.0 3 Super built-up Area Ready To Move Lingadheeranahalli 3 BHK 1521 3.0 95.0 Soiewre 1.0 4 Super built-up Area Ready To Move 2 BHK 1200 Kothanur 1200 2.0 1.0 51.0 #filling null vlaue with next value column wise by using method In [21]: df7=df.fillna(method='pad' , axis=1) df7.head() availability location society total\_sqft bath balcony price Out[21]: area\_type size 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 2600 5.0 3.0 120.0 Theanmp 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 1521 3.0 95.0 Soiewre 1.0 2.0 4 Super built-up Area Ready To Move Kothanur 2 BHK 2 BHK 1200 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 society total\_sqft bath balcony price size Coomee 0 Super built-up Area 19-Dec Electronic City Phase II 2 BHK 1056 2.0 39.07 1.0 1 Chikka Tirupathi 4 Bedroom Plot Area Ready To Move Theanmp 2600 5.0 3.0 120.00 Built-up Area Ready To Move Uttarahalli 3 BHK 1440 2.0 62.00 abc 3.0 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 #filling null vlaue with the 'mean' of a column df9=df.fillna(value=df['balcony'].mean()) df9.head() society total\_sqft bath balcony Out[25]: area\_type availability location price Coomee 19-Dec Electronic City Phase II 1056 2.0 39.07 Super built-up Area 2 BHK 1.0 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 Soiewre 1521 3.0 95.00 3 BHK 1.0 2 BHK 1.584376 4 Super built-up Area Ready To Move 2.0 51.00 Kothanur 1200 1.0 #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 society total\_sqft bath balcony price size 0 Super built-up Area 19-Dec Electronic City Phase II 2 BHK Coomee 1056 2.0 1.0 39.07 5.0 Plot Area Ready To Move Chikka Tirupathi 4 Bedroom Theanmp 2600 3.0 120.00 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 95.00 Soiewre 1521 3.0 1.0 4 Super built-up Area Ready To Move Kothanur 2 BHK 3.0 1200 2.0 1.0 51.00 drop #this will remove all row which consist of even single null value df10=df.dropna() df10.head() availability location society total\_sqft bath balcony price Out[29]: area\_type size Super built-up Area 19-Dec Electronic City Phase II 2 BHK Coomee 1056 2.0 1.0 39.07 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 95.00 3 BHK Soiewre 1521 3.0 1.0 5 Super built-up Area Ready To Move Whitefield 2 BHK DuenaTa 1170 2.0 38.00 Plot Area Ready To Move Whitefield 4 Bedroom 5.0 11 3.0 295.00 Prrry M 2785 #'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() availability location society total\_sqft bath balcony price Out[30]: area\_type size 0 Super built-up Area 19-Dec Electronic City Phase II 2 BHK Coomee 1056 2.0 39.07 Chikka Tirupathi 4 Bedroom Theanmp 3.0 120.00 Plot Area Ready To Move 2600 5.0 3 Super built-up Area Ready To Move Lingadheeranahalli 3 BHK Soiewre 1521 3.0 95.00 Whitefield 38.00 5 Super built-up Area Ready To Move 2 BHK DuenaTa 1170 2.0 1.0 Plot Area Ready To Move 11 Whitefield 4 Bedroom Prrry M 2785 3.0 295.00 df12=df.dropna(how='all') df12.head() Out[31]: area\_type availability location society total\_sqft bath balcony price size Super built-up Area 19-Dec Electronic City Phase II 2 BHK Coomee 1056 2.0 1.0 39.07 5.0 Plot Area Ready To Move Chikka Tirupathi 4 Bedroom 2600 3.0 120.00 2 Built-up Area Ready To Move Uttarahalli 3 BHK 1440 2.0 62.00 NaN 3.0 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 df13=df.replace(to\_replace =3.0 ,value= 5.0) In [35]: 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 Electronic City Phase II 2.0 39.07 19-Dec 2 BHK 1056 1.0 Coomee Plot Area Ready To Move 2600 5.0 5.0 120.00 Chikka Tirupathi 4 Bedroom Theanmp 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 1521 5.0 95.00 Soiewre 1.0 4 Super built-up Area Ready To Move 2.0 Kothanur 2 BHK 1200 1.0 51.00 NaN interpolate() it will interpret value accourding to the column df['balcony']=df['balcony'].interpolate(method='linear') In [39]: df area\_type availability location society total\_sqft bath balcony price Out[39]: size 19-Dec Electronic City Phase II 39.07 O Super built-up Area 2 BHK Coomee 1056 2.0 1.0 4 Bedroom Chikka Tirupathi 1 Plot Area Ready To Move 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 1521 3.0 95.00 Soiewre 1.0 2.0 4 Super built-up Area Ready To Move Kothanur 2 BHK NaN 1200 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 Ready To Move 13317 Built-up Area Raja Rajeshwari Nagar 2 BHK Mahla T 1141 2.0 1.0 60.00 4.0 1.0 488.00 13318 Super built-up Area 18-Jun 4 BHK SollyCl 4689 Padmanabhanagar **13319** Super built-up Area Doddathoguru 1 BHK NaN 550 1.0 1.0 17.00 13320 rows × 9 columns