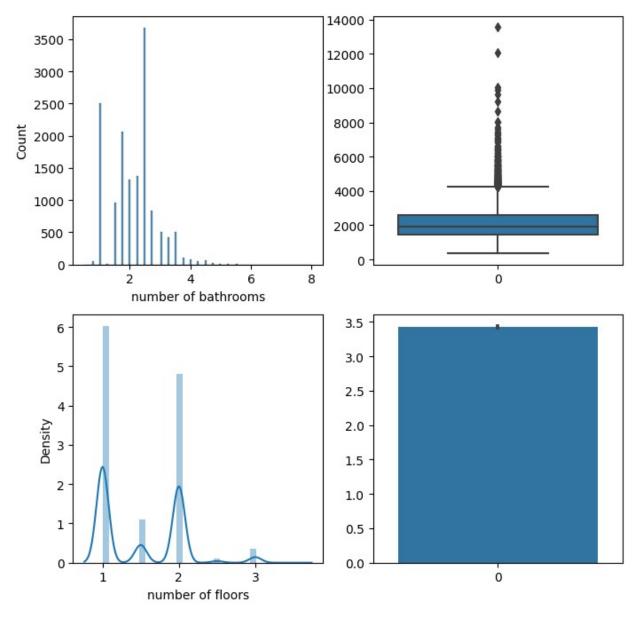
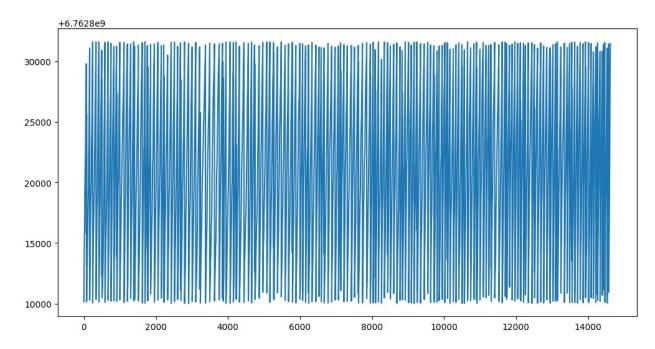
```
'''Name:Akshara
   Registration number: 21BAI1399
   Branch: CSE with specialization in AI and ML'''
{"type":"string"}
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
df=pd.read_csv('/content/House Price India.csv')
df
                id
                      Date
                            number of bedrooms
                                                  number of bathrooms \
0
       6762810145
                     42491
                                               5
                                                                   2.50
1
                                               4
       6762810635
                     42491
                                                                   2.50
2
                                               5
                     42491
                                                                   2.75
       6762810998
3
                                               4
       6762812605
                     42491
                                                                   2.50
4
                                               3
       6762812919
                     42491
                                                                   2.00
                                                                    . . .
       6762830250
                     42734
                                               2
14615
                                                                   1.50
                                               3
       6762830339
14616
                     42734
                                                                   2.00
                                               2
14617
       6762830618
                     42734
                                                                   1.00
                                               4
14618
       6762830709
                     42734
                                                                   1.00
                                               3
14619 6762831463
                    42734
                                                                   1.00
       living area lot area
                                 number of floors waterfront present
0
               3650
                          9050
                                               2.0
1
               2920
                          4000
                                               1.5
                                                                        0
2
                                               1.5
                                                                        0
               2910
                          9480
3
               3310
                         42998
                                               2.0
                                                                        0
4
                                               1.5
                                                                        0
               2710
                          4500
                . . .
                                                . . .
                                                                        0
               1556
14615
                         20000
                                               1.0
                                                                        0
                                               1.5
14616
               1680
                          7000
14617
               1070
                          6120
                                               1.0
                                                                        0
                                                                        0
14618
               1030
                          6621
                                               1.0
14619
                900
                          4770
                                               1.0
                                                                        0
       number of views
                          condition of the house
                                                          Built Year
0
                       4
                                                                 1921
                                                     . . .
                                                  5
1
                       0
                                                                 1909
                                                     . . .
2
                       0
                                                  3
                                                                 1939
3
                       0
                                                  3
                                                                 2001
4
                       0
                                                  4
                                                                 1929
. . .
                                                                  . . .
                      . .
                                                 . .
14615
                       0
                                                 4
                                                                 1957
                                                     . . .
14616
                       0
                                                 4
                                                                 1968
                       0
                                                  3
14617
                                                                 1962
                                                     . . .
                       0
14618
                                                                 1955
```

14619		0				3 .		1969
	Renovation `	Year	Postal C	Code L	_attitud	de L	ongitude	
living_ 0	_area_renov	\ 0	122	2003	52.864	15	-114.557	
2880		U	122	.003	J2.00 ²	+3	-114.557	
1		0	122	2004	52.887	78	-114.470	
2470 2		0	122	2004	52.885	52	-114.468	
2940								
3 3350		0	122	2005	52.953	32	-114.321	
4		0	122	2006	52.904	17	-114.485	
2060								
		• • •			• 1			
14615		0	122	2066	52.619	91	-114.472	
2250		0	122	0.72	E2 E0	76	114 202	
14616 1540		0	122	2072	52.507	/ 5	-114.393	
14617		0	122	2056	52.728	39	-114.507	
1130 14618		0	122	2042	52.715	57	-114.411	
1420		U	122	.042	JZ./15	,	-114.411	
14619	2	2009	122	2018	52.533	38	-114.552	
900								
o i so o st	lot_area_re	nov	Number of	schoo	ols near	rby	Distance	from the
airport		100				2		
58								
1 51	40	900				2		
2	66	500				1		
53	420	247				_		
3 76	428	34 /				3		
4	45	500				1		
51								
14615	172	286				3		
76 14616	7/	180				3		
59	72	100				5		
14617	61	120				2		
64 14618	66	531				3		
54	00	JJ 1				J		

```
14619
                 3480
                                               2
55
         Price
0
       2380000
1
       1400000
2
       1200000
3
        838000
4
        805000
        221700
14615
        219200
14616
14617
        209000
14618
        205000
14619
        146000
[14620 rows x 23 columns]
from matplotlib import rcParams
rcParams['figure.figsize']=8,8
fig,axes=plt.subplots(2,2)
sns.histplot(data=df['number of bathrooms'],ax=axes[0,0])
sns.boxplot(data=df['living area'],ax=axes[0,1])
sns.distplot(df['number of floors'],ax=axes[1,0])
sns.barplot(df['condition of the house'],ax=axes[1,1])
<ipython-input-65-175fa27109b6>:6: UserWarning:
`distplot` is a deprecated function and will be removed in seaborn
v0.14.0.
Please adapt your code to use either `displot` (a figure-level
function with
similar flexibility) or `histplot` (an axes-level function for
histograms).
For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751
  sns.distplot(df['number of floors'],ax=axes[1,0])
<Axes: >
```



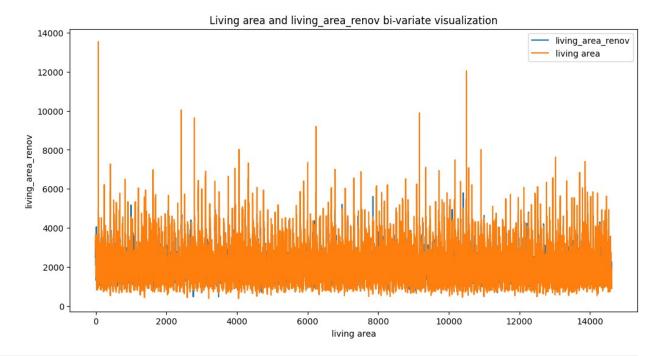
```
plt.figure(figsize=(12,6))
plt.plot(df['id'])
[<matplotlib.lines.Line2D at 0x7ef431291f30>]
```



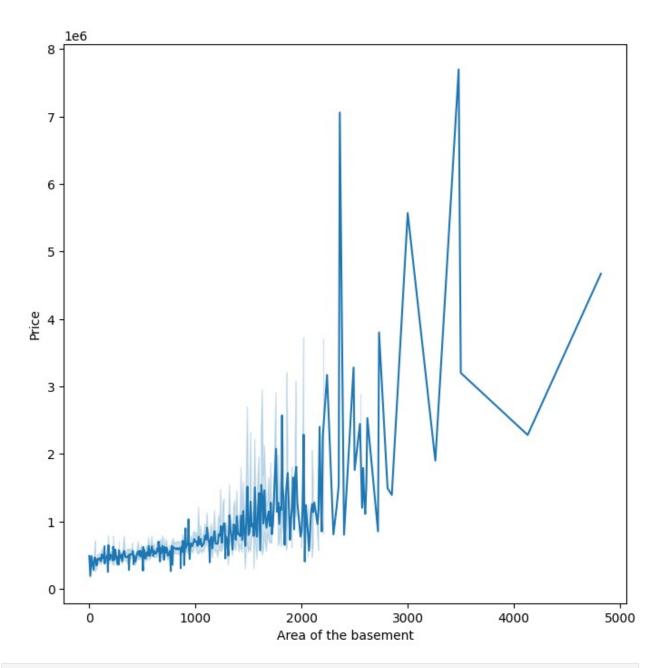
#Bi-Variate Analysis

```
plt.figure(figsize=(12,6))
plt.title("Living area and living_area_renov bi-variate
visualization")
plt.xlabel("living area")
plt.ylabel("living_area_renov")
plt.plot(df['living_area_renov'])
plt.plot(df['living area'])
plt.legend(['living_area_renov','living area'])

<matplotlib.legend.Legend at 0x7ef431244e80>
```

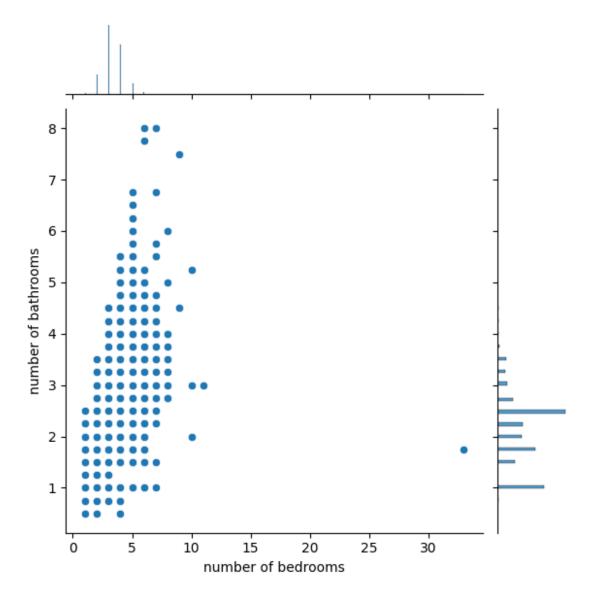


sns.lineplot(x='Area of the basement',y='Price',data=df)
<Axes: xlabel='Area of the basement', ylabel='Price'>



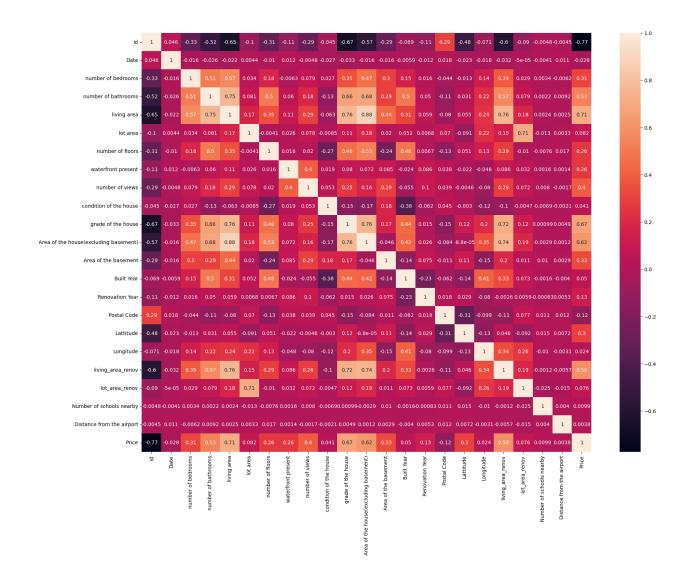
sns.jointplot(x= 'number of bedrooms',y ='number of
bathrooms',data=df)

<seaborn.axisgrid.JointGrid at 0x7ef4307ffb50>



#Multivariate analysis

```
plt.figure(figsize=(20,15))
sns.heatmap(df.corr(),annot=True)
<Axes: >
```



#4.Performing descriptive statististics on the dataset

df.des	cribe()			
	id	Date	number of bedrooms	number of
bathro	oms \			
count	1.462000e+04	14620.000000	14620.000000	
14620.0	000000			
mean	6.762821e+09	42604.538646	3.379343	
2.12958	83			
std	6.237575e+03	67.347991	0.938719	
0.76993	34			
min	6.762810e+09	42491.000000	1.000000	
0.50000	00			
25%	6.762815e+09	42546.000000	3.000000	
1.75000	00			
50%	6.762821e+09	42600.000000	3.000000	
2.25000	00			

75% 6.762826e+09 42662.000000 4.000000 2.500000 Living area					
max 6.762832e+09 42734.000000 33.000000 living area present \ count 14620.000000 1.462000e+04 14620.000000 waterfront present \ count 14620.000000 1.462000e+04 14620.000000 waterfront mean 2098.262996 1.509328e+04 1.502360 0.087193 min 370.00000 5.200000e+02 1.000000 25% 1440.00000 5.010750e+03 1.000000 0.000000 0.000000 0.000000 number of views condition of the house Built Year \ 14620.00000 count 14620.00000 1.000000 number of views condition of the house Built Year \ 14620.00000 count 14620.000000 1.04218e+06 3.500000 1.000000 1970.926402 std 0.766259 0.664151 29.493625 nin 0.000000 1.000000 1990.000000 25% 0.000000 3.000000 19975.000000 25% 0.000000 1.000000			42662.000000	4.0000	00
Description	max	6.762832e+09	42734.000000	33.0000	00
Count 14620.000000	nrocon		lot area	number of floors	waterfront
mean 2098.262996 1.509328e+04 1.502360 0.007661 928.275721 3.791962e+04 0.540239 0.087193 min 370.000000 5.200000e+02 1.000000 25% 1440.000000 5.010750e+03 1.000000 50% 1930.000000 7.620000e+03 1.500000 0.000000 25% 1500.00000 1.080000e+04 2.000000 0.000000 1.000000 1.000000 1.000000 max 13540.000000 1.000000 1.000000 number of views condition of the house Built Year \tag{20.000000} 1.000000 14620.000000 1.4620.000000 1.000000 and 0.766259 0.664151 29.493625 min 0.000000 3.000000 1.900000 1.900000 50% 0.000000 3.000000 1.975.000000 2.000000 75% 0.000000 4.000000 1.997.000000 1.000000 8 0.000000 3.000000 1.997.000000 1.000000 1.000000	count	14620.000000	1.462000e+04	14620.000000	
0.087193 min 370.000000 5.200000e+02 1.000000 25% 1440.000000 5.010750e+03 1.000000 50% 1930.000000 7.620000e+03 1.500000 0.000000 0.000000 1.080000e+04 2.000000 max 13540.000000 1.074218e+06 3.500000 1.000000 number of views condition of the house Built Year \	mean	2098.262996	1.509328e+04	1.502360	
min 370.000000 5.200000e+02 1.000000 25% 1440.00000 5.010750e+03 1.000000 50% 1930.000000 7.620000e+03 1.500000 0.000000 2570.000000 1.080000e+04 2.000000 0.000000 1.074218e+06 3.500000 1.000000 14620.000000 14620.000000 1.000000 14620.000000 14620.000000 1.000000 14620.000000 14620.000000 1.000000 14620.000000 14620.000000 1.000000 14620.000000 14620.000000 1.000000 14620.000000 1970.926402 1.000000 1.000000 1990.000000 1.000000 1.000000 1990.000000 25% 0.000000 3.000000 1997.000000 75% 0.000000 3.000000 1997.000000 1000000 14620.000000 14620.000000 14620.000000 12000.00000 12000.00000 1000000 12000.00000 12000.00000 14620.000000 10000			3.791962e+04	0.540239	
25% 1440.000000 5.010750e+03 1.000000 0.000000 0.000000	min	370.000000	5.200000e+02	1.000000	
1930.000000	25%	1440.000000	5.010750e+03	1.000000	
75% 2570.000000 1.080000e+04 2.000000 0.000000 max 13540.000000 1.074218e+06 3.500000 1.000000	50%	1930.000000	7.620000e+03	1.500000	
max 13540.000000 1.074218e+06 3.500000 number of views condition of the house Built Year \ count 14620.000000 14620.000000 14620.000000 mean 0.233105 3.430506 1970.926402 std 0.766259 0.664151 29.493625 min 0.000000 1.000000 1990.0000000 25% 0.000000 3.000000 1951.000000 50% 0.000000 3.000000 1951.000000 75% 0.000000 4.000000 1975.000000 max 4.000000 1997.000000 5.000000 1997.000000 max 4.000000 14620.000000 1997.000000 max 4.000000 12033.062244 52.792848 -114.404007 std 416.216661 19.082418 0.137522 0.141326 min 0.000000 122033.000000 52.385900 -114.709000 25% 0.000000 122033.000000 52.385900 -114.709000 25% 0.000000 122032.000000 52.385900 -114.709000 50% 0.000000 122032.000000 52.886400 -114.412000 75% 0.000000 122048.000000 52.908900 -114.315000 max 2015.000000 122048.000000 52.908900 -114.315000 max 2015.000000 122072.000000 53.007600 -113.505000 Number of schools nearby \ count 14620.000000 14620.000000 14620.000000 max 2015.000000 14620.000000 53.007600 -113.505000 Living_area_renov lot_area_renov Number of schools nearby \ count 14620.000000 14620.000000 53.007600 -113.505000 Living_area_renov lot_area_renov Number of schools nearby \ count 14620.000000 14620.000000 53.007600 -113.505000 Margarea_renov lot_area_renov Number of schools nearby \ count 14620.000000 14620.000000 651.000000 Living_area_renov lot_area_renov l			1.080000e+04	2.000000	
count 14620.000000 14620.000000 14620.000000 mean 0.233105 3.430506 1970.926402 std 0.766259 0.664151 29.493625 min 0.000000 1.000000 1900.000000 25% 0.000000 3.000000 1951.000000 50% 0.000000 3.000000 1975.000000 75% 0.000000 4.000000 1997.000000 max 4.000000 5.000000 1997.000000 mean 90.924008 122033.062244 52.792848 -114.404007 std 416.216661 19.082418 0.137522 0.141326 min 0.000000 122003.000000 52.385900 -114.709000 25% 0.000000 122017.000000 52.707600 -114.519000 50% 0.000000 122032.000000 52.806400 -114.421000 75% 0.000000 122048.000000 52.908900 -114.315000	max	13540.000000	1.074218e+06	3.500000	
count 14620.000000 14620.000000 14620.000000 14620.000000 mean 90.924008 122033.062244 52.792848 -114.404007 std 416.216661 19.082418 0.137522 0.141326 min 0.000000 122003.000000 52.385900 -114.709000 25% 0.000000 122017.000000 52.707600 -114.519000 50% 0.000000 122032.000000 52.806400 -114.421000 75% 0.000000 122048.000000 52.908900 -114.315000 max 2015.000000 122072.000000 53.007600 -113.505000 Living_area_renov lot_area_renov Number of schools nearby \ count 14620.000000 14620.000000 14620.000000 mean 1996.702257 12753.500068 2.012244 std 691.093366 26058.414467 0.817284 min 460.000000 651.000000 1.000000	mean std min 25% 50% 75%	14620.0000 0.2331 0.7662 0.0000 0.0000 0.0000	00 05 59 00 00 00	14620.000000 3.430506 0.664151 1.000000 3.000000 4.000000	14620.000000 1970.926402 29.493625 1900.000000 1951.000000 1975.000000
count $1\overline{4}620.\overline{0}00000$ $1\overline{4}620.\overline{0}00000$ 14620.000000 mean 1996.702257 12753.500068 2.012244 std 691.093366 26058.414467 0.817284 min 460.000000 651.000000 1.000000	mean std min 25% 50% 75%	14620.0000 90.9240 416.2166 0.0000 0.0000 0.0000	00 14620.000 08 122033.062 61 19.082 00 122003.000 00 122017.000 00 122032.000 00 122048.000	9000 14620.000000 2244 52.792848 2418 0.137522 9000 52.385900 9000 52.707600 9000 52.806400 9000 52.908900	14620.000000 -114.404007 0.141326 -114.709000 -114.519000 -114.421000 -114.315000
	mean std min	14620.00 1996.70 691.09 460.00	$ \begin{array}{rrr} 0000 & 1\overline{4}620 \\ 2257 & 12753 \\ 3366 & 26058 \\ 0000 & 651 \\ \end{array} $.000000 .500068 .414467 .000000	14620.000000 2.012244 0.817284 1.000000

```
50%
             1850.000000
                              7620.000000
                                                            2.000000
75%
             2380.000000
                            10125.000000
                                                            3.000000
             6110.000000
                            560617.000000
                                                            3.000000
max
       Distance from the airport
                                          Price
                    14620.000000
                                  1.462000e+04
count
                       64.950958
                                  5.389322e+05
mean
                        8.936008
std
                                  3.675324e+05
                                  7.800000e+04
                       50.000000
min
25%
                       57.000000
                                  3.200000e+05
50%
                       65.000000
                                  4.500000e+05
75%
                       73.000000
                                  6.450000e+05
max
                       80.000000 7.700000e+06
[8 rows x 23 columns]
```

#Handling the missing values

```
df.isnull().any()
df.isnull().sum()
id
                                           0
Date
                                           0
number of bedrooms
                                           0
number of bathrooms
                                           0
living area
                                           0
lot area
                                           0
number of floors
                                           0
                                           0
waterfront present
number of views
                                           0
condition of the house
                                           0
grade of the house
Area of the house(excluding basement)
                                           0
Area of the basement
                                           0
Built Year
                                           0
Renovation Year
                                           0
Postal Code
                                           0
Lattitude
                                           0
Longitude
                                           0
                                           0
living area renov
lot area renov
                                           0
Number of schools nearby
                                           0
Distance from the airport
                                           0
Price
                                           0
dtype: int64
```