### 21BAI1131-SHUVAM JENA

import pandas as
pdimport numpy as
np
import matplotlib.pyplot as
pltimport seaborn as sns

TASK1----- Downloading the dataset

TASK2----- Load the dataset

	id	Date	number_of_bedrooms_number_of_bathrooms	\
0	6762810145	42491	5 2.50	
1	6762810635	42491	4 2.50	
2	6762810998	42491	5 2.75	
3	6762812605	42491	4 2.50	
4	6762812919	42491	3 2.00	
14615	6762830250	42734	2 1.50	
14616	6762830339	42734	3 2.00	
14617	6762830618	42734	2 1.00	
14618	6762830709	42734	4 1.00	
14619	6762831463	42734	3 1.00	

	living_area	lot_area	number_of_floors	waterfront_present
0	3650	9050	2.0	0
1	2920	4000	1.5	0
2	2910	9480	1.5	0
3	3310	42998	2.0	0
4	2710	4500	1.5	0
14615	1556	20000	1.0	0
14616	1680	7000	1.5	0
14617	1070	6120	1.0	0
14618	1030	6621	1.0	0
14619	900	4770	1.0	0

	number_of_views	condition_of_the_house	 Built_Year	\	
0	4	5	 1921		
1	0	5	 1909		
2	0	3	 1939		
3	0	3	 2001		

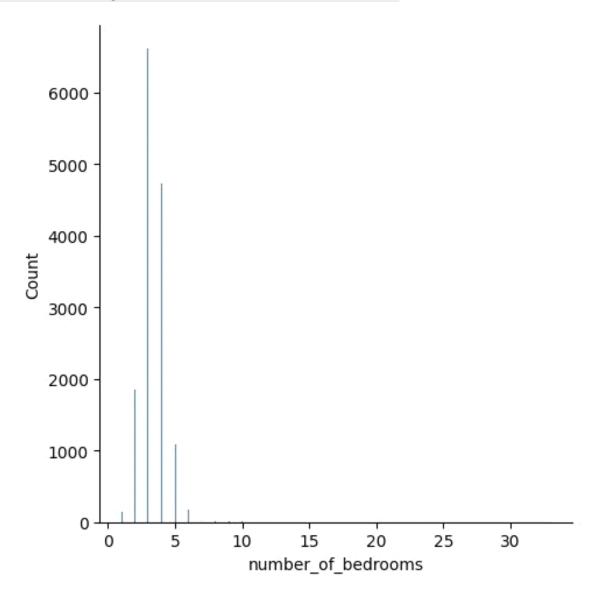
```
0
                                                                 1929
                                                     . . .
 14615
                        0
                                                  4
                                                                 1957
                                                     . . .
 14616
                        0
                                                  4 ...
                                                                 1968
 14617
                        0
                                                  3
                                                                 1962
 14618
                                                  4 ...
                        0
                                                                 1955
 14619
                        0
                                                  3 ...
                                                                 1969
         52.8852 -114.468
         122004
         52.8645 -114.557
         122003
         0
         Renovation_Year Postal_Code Lattitude Longitude
 living area renov \
0
122004
52.8878 -114.470
122005
52.9532 -114.321
. . .
. . .
. . .
. . .
 52.7289 -114.507
 122056
 52.9047 -114.485
 122006
 0
 14615
                        0
                        122066
          52.6191
                       -114.472
                        0
                        122072
          52.5075
                       -114.393
```

```
0
122042
52.7157 -114.411
                 2009
                          122018
                   52.5338 -114.552
 900
       lot_area_renov Number_of_schools_nearby
 Distance_from_the_airport \0 5400
 58
 1
                4000
 51
 2
                6600
 53
 3
               42847
 76
                4500
 4
 51
                . . .
 14615 17286
 76
            7480
 14616
 59
                                        2
                                        1
                                        3
                                        1
                                        3
```

## Univariate Analysis

sns.displot(df['number\_of\_bedrooms'])

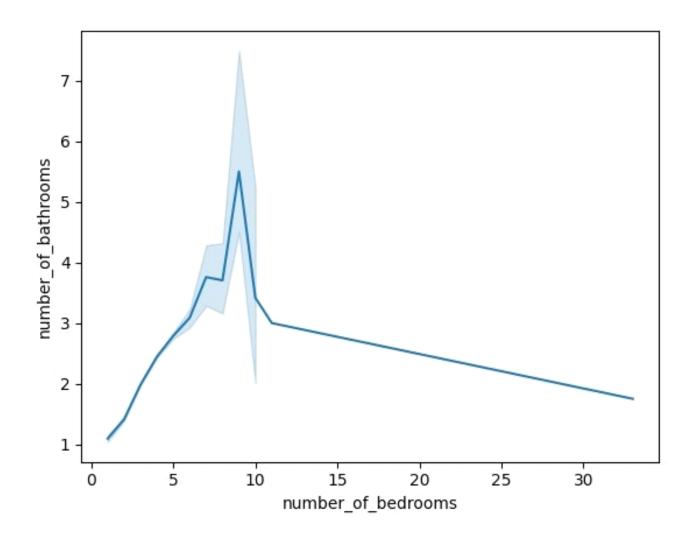
<seaborn.axisgrid.FacetGrid at 0x7815064595d0>



## **Bivariate Analysis**

sns.lineplot(x=df['number\_of\_bedrooms'],y=df['number\_of\_bathrooms'])

<Axes: xlabel='number\_of\_bedrooms', ylabel='number\_of\_bathrooms'>



# MULTIVARIATE ANALYSIS

sns.heatmap(df.corr(),annot=True)

<Axes: >

```
id - 1, 0.46338 502 65:0. 301 101:0 994 507 507:0 995 90129:0 80 701:500 900 4404 3
                                             Date -04 1 .00 602 6 20 804 0 .00 1020 4862 908 601 6 1080 5091 621 802 601 6532 0000 400 .D
                        number of bedrooms -0.630 11 51.570 840.6079 07354 0.8.0.590.60490.304890 290.84663
                       number of bathrooms -0.5202.5 10.0 081.5.00518.1 00500.29.5.005.0.D 8120.5.0 790.009
                                      living area -0.6500 20.751 ). 107 35 101-290 0, 7068 344 B 10-590 @ 552 0.70 1 800 060 71
                                         lot area -0...D04384811 1.004026780810 18020512166709 1021 1 1 0...D0338
                             number of floors -0.30 001 8.5-3.50 1.00.692, 274553, 244896.64.55108290 00.0602
                           waterfront present -0.0-01020606010261 1 0040 090807083020860882 04386 0300 0014
                             number of views $0290.4609182090780274 1 05335165290551-03390460828037403001
                      condition of the house -.04.002070.70.30.0030807.70.95 1).305.07108-0806924.500 31-00.100.0006924
                          grade_of_the_house -0.6703-0.6670.10.460826.1 10.70.107-040-05051 2.2.70100099.67
Area of the house(excluding basement) -1-6.701(0.6088) 1086 30/2/6.17761 .04.02028 849 03.760 90000
                      Area of the basement -0-2.90 D68.209449-02.20.405209 108-D70-4 1 ).0.407.500.10.1502.00 000 00029
                                       Built Year -- 06090596.6.810529802465384440.1 1 )-233-620441.8307301060.4
                              Renovation Year -)-0.1010206050509060608060.06021.60267052 11 001 6-2990 6000066600803 3
                                    Postal Code - 1090.894.401.08907.0.808808.9450.508601.00621 10:6199.90.107.001.0-D21
                                        Lattitude -0.-0802001088905.509905-D232040688.80+0451.01.4-D28 1 ).01.864.60902105 0 2
                                       Longitude -- . 07/00/20422224 22-0304808102 2.35. 1541-080991 1) 334260000.502
                             living_area_renov - 06603 20 0.70. 155 29 862 8.0. 727 0. 268 8 92 6. D 468 1 3.0 900 0 0
                                lot_area_renov -0.560032909.07.00.00507000472.0901.07.005397.09.2051 12.005000
                Distance_from_the_airport -00.405.000000290206.303000.0001400.000209.045031920.0015031050 1000
                                             Price 0:070081
                                                                                    the house
                                                                                  condition of the house
                                                           number of bedrooms
                                                              number_of_bathrooms
                                                                  living_area
                                                                     lot area
                                                                        number of floors
                                                                                           Area of the basement
                                                                                                 Renovation_Year
                                                                                                    Postal Code
                                                                                                       Lattitude
                                                                                                          Longitude
                                                                                                                       Distance_from_the_airport
                                                                           waterfront present
                                                                              number of views
                                                                                        Area_of_the_house(excluding basement)
                                                                                              Built Year
                                                                                                              iving_area_renov
                                                                                                                 lot area renov
                                                                                                                    Number of schools nearby
                                                                                    grade of
```

TASK4 --- To perform the descriptive statistics (It shows all the numerical data)

```
2.250000
75% 6.762826e+09 42662.000000 4.000000
2.500000
max 6.762832e+09 42734.000000 33.000000
8.000000
```

living\_area lot\_area number\_of\_floors waterfront\_present \ count 14620.000000 1.462000e+04 14620.000000 14620.000000

- 0.000000
- 0.000000
- 0.087193
- 0.007661

mean		2098.262996	1.509328e+04
1.502360	std	928.275721	3.791962e+04
0.540239	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
- 1.000000
- 0.000000

75%		2570.000000	1.080000e+04
2.000000	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	0.00000	1.000000	 1900.000000	
25%	0.00000	3.000000	 1951.000000	
50%	0.00000	3.000000	 1975.000000	
75%	0.00000	4.000000	 1997.000000	
max	4.000000	5.00000	 2015.000000	

	Renovation_Year	Postal_Code	Lattitude	Longitude	\
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	

25%	1490.000000	5097.750000	1.000000
50%	1850.000000	7620.000000	2.000000
75%	2380.000000	10125.000000	3.000000
max	6110.000000	560617.000000	3.000000

	Distance_from_the_airport	Price	
count	14620.000000	1.462000e+04	
mean	64.950958	5.389322e+05	
std	8.936008	3.675324e+05	
min	50.00000	7.800000e+04	
25%	57.00000	3.200000e+05	
50%	65.00000	4.500000e+05	
75%	73.00000	6.450000e+05	
max	80.00000	7.700000e+06	

TASK-05 ---- Check the missing values

There are no null values in the given dataset

df.isnull().sum()

