from google.colab import files
upload = files.upload()

Choose Files House Price India.csv

House Price India.csv(text/csv) - 1524561 bytes, last modified: 9/30/2023 - 100% done Saving House Price India.csv to House Price India.csv

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

import io

df = pd.read_csv(io.BytesIO(upload['House Price India.csv']))

df.head()

	id	Date	number of bedrooms	number of bathrooms	living area	lot area	number of floors	waterfront present	number of views	condition of the house	 Built Year	Renovation Year	Ро
0	6762810145	42491	5	2.50	3650	9050	2.0	0	4	5	 1921	0	12
1	6762810635	42491	4	2.50	2920	4000	1.5	0	0	5	 1909	0	12
2	6762810998	42491	5	2.75	2910	9480	1.5	0	0	3	 1939	0	12
3	6762812605	42491	4	2.50	3310	42998	2.0	0	0	3	 2001	0	12
4	6762812919	42491	3	2.00	2710	4500	1.5	0	0	4	 1929	0	12

5 rows × 23 columns

df.tail()

	id	Date	number of bedrooms	number of bathrooms	living area	lot area	number of floors	waterfront present	number of views	condition of the house	 Built Year	Renovation Year	
14615	6762830250	42734	2	1.5	1556	20000	1.0	0	0	4	 1957	0	1
14616	6762830339	42734	3	2.0	1680	7000	1.5	0	0	4	 1968	0	1
14617	6762830618	42734	2	1.0	1070	6120	1.0	0	0	3	 1962	0	1
14618	6762830709	42734	4	1.0	1030	6621	1.0	0	0	4	 1955	0	1
14619	6762831463	42734	3	1.0	900	4770	1.0	0	0	3	 1969	2009	1

5 rows x 23 columns

df

	id	Date	O†	number of bathrooms	living area	lot area	number of floors	waterfr pres		number of views	condition of the hous	ie	•••	Built Year	Renovatio Yea
0	6762810145	42491	5	2.50	3650	9050	2.0		0	4		5		1921	
olumns															
Index	(['id', 'Date'														
	'living area 'number of v 'Area of the 'Built Year'	views', 'cor house(excl , 'Renovati	ndition o luding bas ion Year'	f the hous sement)', , 'Postal	e', 'grad 'Area of Code', 'L	de of the the bas attitud	ne house' sement',								
	'Longitude', 'Number of s	chools near					'Price']	,							
14616	dtype='object 6 6762830339		3	2.00	1680	7000	1.5		0	0		4		1968	
dtypes															
id				int64	ļ										
Date				int64											
	r of bedrooms			int64											
	r of bathrooms g area	•		float64 int64											
lot a				int64											
	r of floors			float64											
	front present			int64											
number	r of views			int64	ļ										
	tion of the ho	ouse		int64											
0	of the house			int64											
	of the house(_	asement)	int64											
	of the basemer	it		int64											
Built	ation Year			int64 int64											
	1 Code			int64											
Latti				float64											
Longi	tude			float64	ļ										
living	g_area_renov			int64	ļ										
lot_a	rea_renov			int64											
_															
Number	r of schools r	-		int64											
Number Distar	nce from the a	-		int64	ļ										
Number Distar Price	nce from the a	-			ļ										
Number Distar Price dtype	nce from the a	-		int64	ļ										
Number Distar Price dtype info() <class rangel<="" td=""><td>nce from the a : object s 'pandas.core Index: 14620 e</td><td>eirport e.frame.Data entries, 0 t</td><td>to 14619</td><td>int64</td><td>ļ</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></class>	nce from the a : object s 'pandas.core Index: 14620 e	eirport e.frame.Data entries, 0 t	to 14619	int64	ļ										
Number Distar Price dtype info() <class (<="" data="" rangel="" td=""><td>nce from the a : object s 'pandas.core</td><td>eirport e.frame.Data entries, 0 t</td><td>to 14619</td><td>int64 int64 Non-</td><td>Null Cour</td><td></td><td>pe</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></class>	nce from the a : object s 'pandas.core	eirport e.frame.Data entries, 0 t	to 14619	int64 int64 Non-	Null Cour		pe								
Number Distar Price dtype info() <class (<="" data="" rangel="" td=""><td>nce from the a : object s 'pandas.core Index: 14620 e columns (total Column</td><td>eirport e.frame.Data entries, 0 t</td><td>to 14619</td><td>int64 int64 Non</td><td>Null Cour</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></class>	nce from the a : object s 'pandas.core Index: 14620 e columns (total Column	eirport e.frame.Data entries, 0 t	to 14619	int64 int64 Non	Null Cour										
Number Distar Price dtype info() <class (<="" data="" rangel="" td=""><td>nce from the a : object s 'pandas.core Index: 14620 e columns (total</td><td>eirport e.frame.Data entries, 0 t</td><td>to 14619</td><td>int64 int64 Non 1462</td><td>Null Cour</td><td>ll int</td><td>64</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></class>	nce from the a : object s 'pandas.core Index: 14620 e columns (total	eirport e.frame.Data entries, 0 t	to 14619	int64 int64 Non 1462	Null Cour	ll int	64								
Number Distar Price dtype info() <class (="" 0="" 1="" :="" [<="" data="" rangel="" td=""><td>nce from the a : object s 'pandas.core Index: 14620 e columns (total Column</td><td>e.frame.Data entries, 0 t 23 columns</td><td>to 14619</td><td>int64 int64 Non 1462 1462</td><td>Null Cour</td><td>ll inte</td><td>64 64</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></class>	nce from the a : object s 'pandas.core Index: 14620 e columns (total Column	e.frame.Data entries, 0 t 23 columns	to 14619	int64 int64 Non 1462 1462	Null Cour	ll inte	64 64								
Number Distar Price dtype info() <class (<="" data="" rangel="" td=""><td>nce from the a : object s 'pandas.core Index: 14620 e columns (total Column id Date</td><td>e.frame.Data entries, 0 t 23 columns</td><td>to 14619</td><td>int64 int64 Non 1462 1462</td><td>.Null Cour 20 non-nul 20 non-nul</td><td>ll inte</td><td>64 64 64</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></class>	nce from the a : object s 'pandas.core Index: 14620 e columns (total Column id Date	e.frame.Data entries, 0 t 23 columns	to 14619	int64 int64 Non 1462 1462	.Null Cour 20 non-nul 20 non-nul	ll inte	64 64 64								
Number Distar Price dtype info() <class (<="" data="" rangel="" td=""><td>nce from the a : object s 'pandas.core Index: 14620 e columns (total Column id Date number of bedr</td><td>e.frame.Data entries, 0 t 23 columns</td><td>to 14619</td><td>Non- 1462 1462 1462 1462</td><td>Null Cour </td><td>ll intell intell intell intelli</td><td>64 64 64 64 at64</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></class>	nce from the a : object s 'pandas.core Index: 14620 e columns (total Column id Date number of bedr	e.frame.Data entries, 0 t 23 columns	to 14619	Non- 1462 1462 1462 1462	Null Cour 	ll intell intell intell intelli	64 64 64 64 at64								
Number Distar Price dtype info() <class (="" 0="" 1="" 2="" 3="" 4="" 5="" 5<="" :="" [="" data="" rangel="" td="" =""><td>s 'pandas.core Index: 14620 e columns (total Column id Date number of bedr number of bath living area lot area</td><td>e.frame.Data entries, 0 t 23 columns</td><td>to 14619</td><td>Non- 1462 1462 1462 1462</td><td>Null Cour </td><td>11 into</td><td>64 64 64 at 64 at 64</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></class>	s 'pandas.core Index: 14620 e columns (total Column id Date number of bedr number of bath living area lot area	e.frame.Data entries, 0 t 23 columns	to 14619	Non- 1462 1462 1462 1462	Null Cour 	11 into	64 64 64 at 64 at 64								
Number Distar Price dtype info() <class (<="" data="" rangel="" td=""><td>s 'pandas.core Index: 14620 e columns (total Column id Date number of bedr number of bath living area lot area number of floc</td><td>e.frame.Data entries, 0 t . 23 columns rooms nrooms</td><td>to 14619</td><td>Non</td><td>Null Cour </td><td>ll into</td><td>64 64 64 64 64 64 64 at64</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></class>	s 'pandas.core Index: 14620 e columns (total Column id Date number of bedr number of bath living area lot area number of floc	e.frame.Data entries, 0 t . 23 columns rooms nrooms	to 14619	Non	Null Cour 	ll into	64 64 64 64 64 64 64 at64								
Number Distar Price dtype info() <class (<="" data="" rangel="" td=""><td>s 'pandas.core Index: 14620 e columns (total Column id Date number of bedr number of bear living area lot area number of floo waterfront pre</td><td>e.frame.Data entries, 0 t 23 columns rooms prooms</td><td>to 14619</td><td>Non</td><td>Null Cour </td><td>ll into</td><td>64 64 64 at64 64 64 64 at64</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></class>	s 'pandas.core Index: 14620 e columns (total Column id Date number of bedr number of bear living area lot area number of floo waterfront pre	e.frame.Data entries, 0 t 23 columns rooms prooms	to 14619	Non	Null Cour 	ll into	64 64 64 at64 64 64 64 at64								
Number Distar Price dtype info() cclass Rangel Data (s 'pandas.core Index: 14620 e columns (total Column id Date number of bedr number of bath living area number of floo waterfront pre number of view	e.frame.Data entries, 0 t 23 columns rooms prooms prooms esent	to 14619	Non	Null Cour 	11 into	64 64 64 at64 64 64 at64 64								
Number Distar Price dtype info() <class (="" 0="" 0<="" 1="" 2="" 3="" 4="" 5="" 6="" 7="" 8="" 9="" :="" [="" data="" rangel="" td="" =""><td>s 'pandas.core Index: 14620 e columns (total Column id Date number of bedr number of bath living area lot area number of floc waterfront pre number of view condition of t</td><td>e.frame.Data entries, 0 t . 23 columns rooms prooms prooms essent us the house</td><td>to 14619</td><td>Non</td><td>Null Cour </td><td>11 into</td><td>64 64 64 at64 64 64 at64 64 64</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></class>	s 'pandas.core Index: 14620 e columns (total Column id Date number of bedr number of bath living area lot area number of floc waterfront pre number of view condition of t	e.frame.Data entries, 0 t . 23 columns rooms prooms prooms essent us the house	to 14619	Non	Null Cour 	11 into	64 64 64 at64 64 64 at64 64 64								
Number Distar Price dtype info() <class #="" (="" ()="" 0="" 1="" 10="" 2="" 3="" 4="" 5="" 6="" 7="" 8="" 8<="" 9="" :="" [="" data="" rangel="" td="" =""><td>s 'pandas.core Index: 14620 e columns (total Column id Date number of bedr number of bath living area number of floo waterfront pre number of view</td><td>e.frame.Data entries, 0 t 23 columns rooms prooms prooms prooms prooms prooms prooms prooms prooms prooms prooms prooms prooms</td><td>to 14619 s):</td><td>Non</td><td>Null Cour </td><td>11 into 11 int</td><td>64 64 64 64 64 64 64 64 64 64</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></class>	s 'pandas.core Index: 14620 e columns (total Column id Date number of bedr number of bath living area number of floo waterfront pre number of view	e.frame.Data entries, 0 t 23 columns rooms prooms prooms prooms prooms prooms prooms prooms prooms prooms prooms prooms prooms	to 14619 s):	Non	Null Cour 	11 into 11 int	64 64 64 64 64 64 64 64 64 64								
Number Distar Price dtype info() <class !="" (="" 0="" 1="" 10="" 11="" 2="" 3="" 4="" 5="" 6="" 7="" 8="" 9="" :="" <="" [="" data="" rangel="" td="" §=""><td>s 'pandas.core Index: 14620 e columns (total Column id Date number of bedr number of bath living area lot area number of floo waterfront pre number of view condition of t grade of the h</td><td>e.frame.Data entries, 0 t 23 columns crooms crooms</td><td>to 14619 s):</td><td>Non</td><td>Null Cour </td><td>ll intill intill</td><td>64 64 64 at64 64 64 at64 64 64</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></class>	s 'pandas.core Index: 14620 e columns (total Column id Date number of bedr number of bath living area lot area number of floo waterfront pre number of view condition of t grade of the h	e.frame.Data entries, 0 t 23 columns crooms	to 14619 s):	Non	Null Cour 	ll intill	64 64 64 at64 64 64 at64 64 64								
Number Distar Price dtype info() <class (<="" data="" rangel="" td=""><td>s 'pandas.core Index: 14620 e columns (total Column id Date number of bedr number of bath living area lot area number of floo waterfront pre number of view condition of t grade of the ba Hea of the ba Built Year</td><td>e.frame.Data e.frame.Data entries, 0 t 23 columns rooms rooms esent vs che house house ouse(excludi</td><td>to 14619 s):</td><td>Non</td><td>Null Cour </td><td>ll intill intill</td><td>64 64 64 64 64 64 64 64 64 64 64 64 64</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></class>	s 'pandas.core Index: 14620 e columns (total Column id Date number of bedr number of bath living area lot area number of floo waterfront pre number of view condition of t grade of the ba Hea of the ba Built Year	e.frame.Data e.frame.Data entries, 0 t 23 columns rooms rooms esent vs che house house ouse(excludi	to 14619 s):	Non	Null Cour 	ll intill	64 64 64 64 64 64 64 64 64 64 64 64 64								
Number Distar Price dtype info() cclass Rangel Data (s 'pandas.core Index: 14620 e columns (total Column id Date number of bedr number of batr living area living area number of floo waterfront pre number of view condition of t grade of the h Area of the ba Built Year Renovation Year	e.frame.Data e.frame.Data entries, 0 t 23 columns rooms rooms esent vs che house house ouse(excludi	to 14619 s):	Non	Null Cour 	ll intill	64 64 64 64 64 64 64 64 64 64 64 64 64								
Number Distar Price dtype dtype info()	s 'pandas.core Index: 14620 e columns (total Column id Date number of bedr number of batr living area number of floc waterfront pre number of view condition of t grade of the h Area of the ba Built Year Renovation Year	e.frame.Data e.frame.Data entries, 0 t 23 columns rooms rooms esent vs che house house ouse(excludi	to 14619 s):	Non	Null Cour 	ll intill	64 64 64 64 64 64 64 64 64 64 64 64 64 6								
Number Distar Price dtype info()	s 'pandas.core Index: 14620 e columns (total Column id Date number of bedr number of bath living area lot area number of view condition of t grade of the h Area of the h Area of the h Area of the h Area of the M Renovation Yea Postal Code Lattitude	e.frame.Data e.frame.Data entries, 0 t 23 columns rooms rooms esent vs che house house ouse(excludi	to 14619 s):	int64 int64 int64 Non	Null Cour 	ll intill	64 64 64 64 64 64 64 64 64 64 64 64 64 6								
Number Distar Price dtype info()	s 'pandas.core Index: 14620 e columns (total Column id Date number of bedr number of bath living area lot area number of floo waterfront pre number of the he Area of the he Area of the ba Built Year Renovation Yea Postal Code Lattitude Longitude	e.frame.Data e.frame.Data entries, 0 t 23 columns crooms p	to 14619 s):	int64 int64 int64 Non	Null Cour 	ll intill	64 64 64 64 64 64 64 64 64 64 64 64 64 6								
Number Distar Price dtype info()	s 'pandas.core Index: 14620 e columns (total Column id Date number of bedr number of bath living area lot area number of vie condition of t grade of the h Area of the ba Built Year Renovation Yea Postal Code Lattitude Longitude living_area_re	e.frame.Data e.frame.Data entries, 0 t . 23 columns rooms rooms ers esent us the house house house puse(excluding esement	to 14619 s):	int64 int64 int64 Non	Null Cour 	ll intill	64 64 64 64 64 64 64 64 66 64 66 64 64 6								
Number Distar Price dtype info()	s 'pandas.core Index: 14620 e columns (total Column id Date number of bedr number of bath living area lot area number of floo waterfront pre number of the he Area of the he Area of the ba Built Year Renovation Yea Postal Code Lattitude Longitude	e.frame.Data e.frame.Data entries, 0 t 23 columns rooms rooms esent vs the house house house eouse(excluding	to 14619 s): ing basem	int64 int64 int64 Non	Null Cour 	ll intill	64 64 64 64 64 64 64 64 64 64 64 64 64 6								
Number Distar Price dtype info()	s 'pandas.core Index: 14620 e columns (total Column id Date number of bedr number of batr living area lot area number of total condition of te Area of the ba Built Year Renovation Yea Postal Code Lattitude Longitude living_area_re lot_area_renov	e.frame.Data e.frame.Data entries, 0 to 23 columns rooms rooms esent ws che house house puse(excluding aremov pools nearby	to 14619 s): ing basem	int64 int64 int64 Non	Null Cour 	ll intill	64 64 64 64 64 64 64 64 64 64 64 64 64 6								
Number Distar Price dtype info()	s 'pandas.core Index: 14620 e columns (total Column id Date number of bedr number of batr living area number of view condition of t grade of the h Area of the ba Built Year Renovation Year Postal Code Lattitude Longitude living_area_re lot_area_renow Number of sche Distance from Price	e.frame.Data e.frame.Data entries, 0 to . 23 columns croom	to 14619 s): ing basem	int64 int64 int64 Non 1462 1462 1462 1462 1462 1462 1462 1462	Null Cour 	ll intill	64 64 64 64 64 64 64 64 64 66 64 66 64 64								
Number Distar Price dtype dtype info()	s 'pandas.core Index: 14620 e columns (total Column id Date number of bedr number of batr living area number of floc waterfront pre number of the h Area of the h Area of the ba Built Year Renovation Year Postal Code Lattitude Longitude living_area_re lot_area_renow Number of scho Distance from	e.frame.Data e.frame.Data entries, 0 t 23 columns rooms rooms rooms esent use buse esent use esement ar	to 14619 s): ing basem	int64 int64 int64 Non 1462 1462 1462 1462 1462 1462 1462 1462	Null Cour 	ll intill	64 64 64 64 64 64 64 64 64 66 64 66 64 64								

(14620, 23)

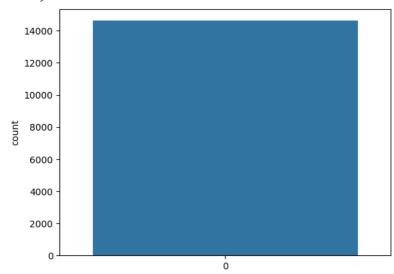
Univariate Analysis

```
print(df.describe())
     std
            6.237575e+03
                              67.347991
                                                    0.938719
                                                                          0.769934
            6.762810e+09
                           42491.000000
                                                    1.000000
                                                                          0.500000
     min
     25%
            6.762815e+09
                           42546.000000
                                                    3.000000
                                                                          1.750000
     50%
            6.762821e+09
                           42600.000000
                                                    3.000000
                                                                          2.250000
            6.762826e+09
                           42662.000000
                                                    4.000000
                                                                          2.500000
     75%
            6.762832e+09
                           42734.000000
                                                   33.000000
                                                                          8.000000
     max
                                         number of floors waterfront present
             living area
                               lot area
     count
            14620.000000
                           1.462000e+04
                                              14620.000000
                                                                  14620.000000
     mean
             2098.262996
                           1.509328e+04
                                                  1.502360
                                                                      0.007661
     std
              928.275721
                           3.791962e+04
                                                  0.540239
                                                                       0.087193
              370.000000
                           5.200000e+02
                                                  1.000000
                                                                       0.000000
     min
                                                  1.000000
                                                                       0.000000
     25%
             1440.000000
                           5.010750e+03
     50%
             1930.000000
                           7.620000e+03
                                                  1.500000
                                                                       0.000000
     75%
             2570.000000
                           1.080000e+04
                                                  2.000000
                                                                       0.000000
                          1.074218e+06
            13540.000000
                                                  3.500000
                                                                       1.000000
     max
            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
                                                       . . .
                   0.000000
                                             1.000000
                                                             1900.000000
     min
                                                       . . .
                   0.000000
                                             3.000000
                                                             1951.000000
     25%
                                                       . . .
     50%
                   0.000000
                                             3.000000
                                                             1975.000000
                                                       . . .
     75%
                   0.000000
                                             4.000000
                                                             1997.000000
                                                       . . .
                   4.000000
                                             5.000000
                                                             2015.000000
     max
            Renovation Year
                                Postal Code
                                                Lattitude
                                                               Longitude
               14620.000000
                               14620.000000 14620.000000 14620.000000
     count
     mean
                  90.924008 122033.062244
                                                52.792848
                                                             -114.404007
     std
                 416.216661
                                  19.082418
                                                 0.137522
                                                                0.141326
                   0.000000
                              122003.000000
                                                 52.385900
                                                             -114.709000
     min
     25%
                   0.000000
                              122017.000000
                                                 52.707600
                                                             -114.519000
                                                 52.806400
     50%
                   0.000000
                              122032.000000
                                                             -114.421000
                                                 52.908900
     75%
                   0.000000
                              122048.000000
                                                             -114.315000
     max
                 2015.000000
                             122072.000000
                                                 53.007600
                                                             -113.505000
                                                Number of schools nearby
            living area renov lot area renov
     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
                          14620.000000
                                        1.462000e+04
     count
                             64.950958
                                        5.389322e+05
     mean
     std
                              8.936008
                                        3.675324e+05
     min
                             50.000000
                                        7.8000000+04
     25%
                             57.000000
                                        3.200000e+05
     50%
                             65.000000
                                        4.500000e+05
     75%
                             73.000000
                                        6.450000e+05
                             80.000000
                                        7.700000e+06
     [8 rows x 23 columns]
```

plt.hist(df['Postal Code'])

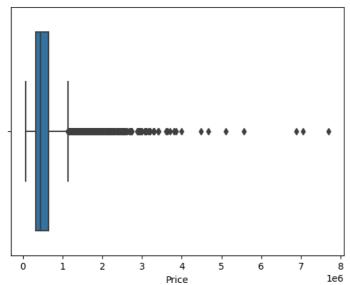
sns.countplot(df['waterfront present'])

<Axes: ylabel='count'>



sns.boxplot(x=df['Price'])

<Axes: xlabel='Price'>



Bivariate Analysis

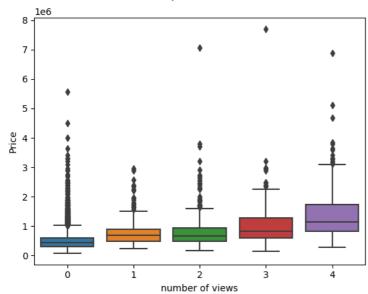
 $\verb|sns.boxplot(x=df['number of bathrooms'],y=df['Price'])|\\$





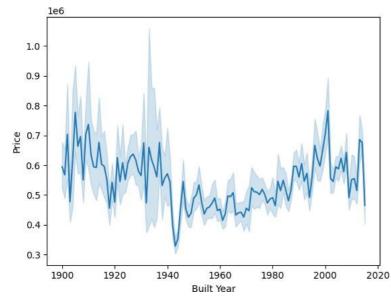
sns.boxplot(x=df['number of views'],y=df['Price'])

<Axes: xlabel='number of views', ylabel='Price'>



sns.lineplot(x=df['Built Year'],y=df['Price'])

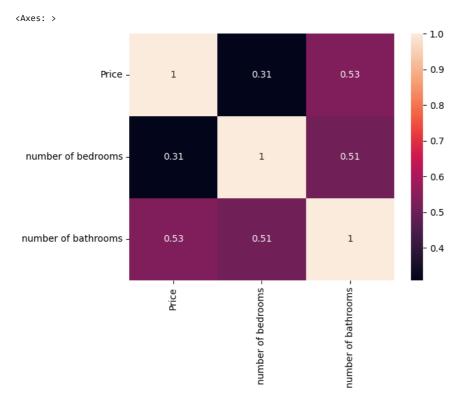
<Axes: xlabel='Built Year', ylabel='Price'>



sns.lineplot(x=df.groupby('Built Year').mean().index,y=df.groupby('Built Year').mean()['Price'])
plt.show()



sns.heatmap(df[['Price','number of bedrooms','number of bathrooms']].corr(),annot=True)

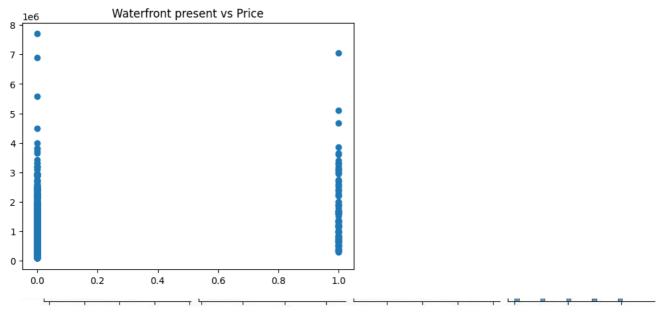


Multivariate Analysis

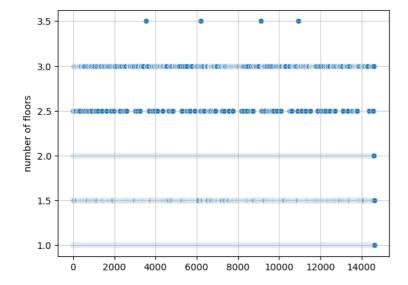
 $\verb|sns.pairplot(df[['Price', 'number of bedrooms', 'number of bathrooms', 'number of floors']]|)| \\$

0

plt.scatter(df['waterfront present'],df['Price'])
plt.title("Waterfront present vs Price")
plt.grid(linestyle='-',linewidth=0.)



sns.scatterplot(df['number of floors'])
plt.grid(linestyle='-',linewidth=0.5)



plt.subplots(figsize=(15,15))
sns.heatmap(df.drop(['id'],axis=1).corr(),linewidth=0.3,annot=True)
plt.show()

Date -	- 1	-0.016	-0.026	-0.022	0.0044	-0.01	0.012	0.0048	0.027	-0.033	-0.01€	-0.016	0.0059	0.012	0.018	-0.023	0.018
number of bedrooms	-0.016	1	0.51	0.57	0.034	0.18-	0.006	0.079	0.027	0.35	0.47	0.3	0.15	0.016	-0.044	-0.013	0.14
number of bathrooms	-0.026	0.51	1	0.75	0.081	0.5	0.06	0.18	-0.13	0.66	0.68	0.29	0.5	0.05	-0.11	0.031	0.22
living area	-0.022	0.57	0.75	1	0.17	0.35	0.11	0.29	-0.063	0.76	0.88	0.44	0.31	0.059	-0.08	0.055	0.24
lot area ·	0.0044	0.034	0.081	0.17	1	0.004	ს .026	0.078	0.008	50.11	0.18	0.02	0.052	0.0068	0.07	-0.091	0.22
number of floors	-0.01	0.18	0.5	0.35-	0.004:	1	0.016	0.02	-0.27	0.46	0.53	-0.24	0.48	0.0067	-0.13	0.051	0.13
waterfront present	0.012	0.006	0.06	0.11	0.026	0.016	1	0.4	0.019	0.08	0.072	0.085	-0.024	0.086	0.038	-0.022	0.048
number of views	0.0048	0.079	0.18	0.29	0.078	0.02	0.4	1	0.053	0.25	0.16	0.29	-0.055	0.1	0.039	0.004	5-0.08
condition of the house	-0.027	0.027	-0.13	-0.063	0.008	5-0.27	0.019	0.053	1	-0.15	-0.17	0.18	-0.38	-0.062	0.045	-0.003	-0.12
grade of the house	-0.033	0.35	0.66	0.76	0.11	0.46	0.08	0.25	-0.15	1	0.76	0.17	0.44	0.015	-0.15	0.12	0.2
Area of the house(excluding basement)	-0.016	0.47	0.68	0.88	0.18	0.53	0.072	0.16	-0.17	0.76	1	-0.046	0.42	0.026	-0.084	3.8e-0	50.35
Area of the basement	-0.016	0.3	0.29	0.44	0.02	-0.24	0.085	0.29	0.18	0.17	-0.046	1	-0.14	0.075	-0.011	0.11	-0.15
Built Year	0.0059	0.15	0.5	0.31	0.052	0.48	-0.024	-0.055	-0.38	0.44	0.42	-0.14	1	-0.23	-0.062	-0.14	0.41
Renovation Year	-0.012	0.016	0.05	0.059	0.0068	0.0067	0.086	0.1	-0.062	0.015	0.026	0.075	-0.23	1	0.018	0.029	-0.08-
Postal Code	-0.018	-0.044	-0.11	-0.08	0.07	-0.13	0.038	0.039	0.045	-0.15	-0.084	-0.011	-0.062	0.018	1	-0.31	0.099
Lattitude -	-0.023	-0.013	0.031	0.055	-0.091	0.051	-0.022	0.0046	0.003	0.12-	3.8e-0	50.11	-0.14	0.029	-0.31	1	-0.13
Longitude ·	-0.018	0.14	0.22	0.24	0.22	0.13	-0.048	-0.08	-0.12	0.2	0.35	-0.15	0.41	-0.08	-0.099	-0.13	1
living_area_renov	-0.032	0.39	0.57	0.76	0.15	0.29	0.086	0.28	-0.1	0.72	0.74	0.2	0.33-	0.0026	5-0.11	0.046	0.34
lot_area_renov ·	-5e-05	0.029	0.079	0.18	0.71	-0.01	0.032	0.072	0.004 ⁻	70.12	0.19	0.011	0.073	0.0059	0.077	-0.092	0.26
Number of schools nearby	0.004	D.0034	9.0022	0.0024	-0.013	0.0070	3 .001€	0.008	ე.00&	10009	9 .0029	0.01-	0.001	5.0008	B .011	0.015	-0.01-
Distance from the airport	0.011	0.006:	D.0092	0.0025	0.0033	0.017	0.0014	0.001 -	0.002	0.0049	0.0012	0.0029	-0.004	0.0053	0.012	0.0072	0.003
Price ·	-0.028	0.31	0.53	0.71	0.082	0.26	0.26	0.4	0.041	0.67	0.62	0.33	0.05	0.13	-0.12	0.3	0.024
	- Date	edrooms -	athrooms -	/ing area -	lot area -	of floors -	t present -	of views -	he house -	he house -	sement) -	asement -	3uilt Year -	tion Year -	stal Code -	Lattitude -	ongitude -

print(df.describe())

```
std
            6.237575e+03
                              67 347991
                                                    0.938719
                                                                          0.769934
     min
            6.762810e+09
                           42491.000000
                                                    1.000000
                                                                          0.500000
     25%
            6.762815e+09
                           42546.000000
                                                    3.000000
                                                                          1.750000
                                                    3.000000
            6.762821e+09
                           42600.000000
                                                                          2.250000
     75%
            6.762826e+09
                           42662.000000
                                                    4.000000
                                                                          2.500000
            6.762832e+09
                                                   33.000000
                                                                          8.000000
                           42734.000000
                               lot area number of floors waterfront present
             living area
            14620.000000
     count
                           1.462000e+04
                                              14620.000000
                                                                  14620.000000
     mean
             2098.262996
                           1.509328e+04
                                                  1.502360
                                                                       0.007661
     std
              928.275721
                           3.791962e+04
                                                  0.540239
                                                                       0.087193
                                                                       0.000000
     min
              370,000000
                           5.200000e+02
                                                  1,000000
     25%
              1440.000000
                           5.010750e+03
                                                  1.000000
                                                                       0.000000
     50%
              1930.000000
                           7.620000e+03
                                                  1.500000
                                                                       0.000000
             2570.000000
                           1.080000e+04
     75%
                                                  2.000000
                                                                       0.000000
     max
            13540.000000
                          1.074218e+06
                                                  3.500000
                                                                       1.000000
            number of views condition of the house
                                                              Built Year
                                                       . . .
     count
               14620.000000
                                        14620.000000
                                                            14620.000000
                                                      . . .
                   0.233105
                                             3,430506
                                                             1970,926402
     mean
                                            0.664151
                                                               29.493625
     std
                   0.766259
                                                       ...
                                            1,000000
     min
                   0.000000
                                                             1900,000000
     25%
                   0.000000
                                             3.000000
                                                             1951.000000
     50%
                   0.000000
                                             3.000000
                                                             1975.000000
     75%
                   0.000000
                                             4.000000
                                                              1997.000000
                                                       . . .
                   4.000000
                                             5.000000
                                                             2015.000000
     max
            Renovation Year
                                Postal Code
                                                 Lattitude
                                                               Longitude
     count
               14620.000000
                               14620.000000
                                             14620.000000
                                                            14620.000000
                  90.924008
                             122033.062244
                                                 52.792848
                                                             -114.404007
     mean
                                  19.082418
                                                  0.137522
                                                                0.141326
     std
                 416.216661
     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
                  1996.702257
                                  12753.500068
                                                                  2.012244
     mean
                                                                  0.817284
     std
                   691.093366
                                  26058.414467
                                                                  1.000000
     min
                   460.000000
                                    651.000000
     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
     count
                          14620.000000
                                        1.462000e+04
                             64.950958
                                        5.389322e+05
     mean
                              8.936008
                                        3.675324e+05
     std
                             50.000000
                                        7.800000e+04
     min
     25%
                             57,000000
                                        3,200000e+05
                             65.000000
     50%
                                        4.500000e+05
     75%
                             73 000000
                                        6.450000e+05
     max
                             80.000000
                                        7.700000e+06
     [8 rows x 23 columns]
print(df.count())
     id
                                                14620
     Date
                                                14620
     number of bedrooms
                                                14620
     number of bathrooms
                                                14620
     living area
                                                14620
     lot area
                                                14620
     number of floors
                                                14620
     waterfront present
                                                14620
     number of views
                                                14620
     condition of the house
                                                14620
                                                14620
     grade of the house
     Area of the house(excluding basement)
                                                14620
     Area of the basement
                                                14620
     Built Year
                                                14620
     Renovation Year
                                                14620
     Postal Code
                                                14629
     Lattitude
                                                14620
     Longitude
                                                14620
     living_area_renov
                                                14620
                                                14620
     lot_area_renov
     Number of schools nearby
                                                14620
     Distance from the airport
                                                14620
                                                14620
     Price
     dtype: int64
```

print(df.corr())

```
0.258066
     Longitude
                                                     0.341221
     living_area_renov
                                                     1.000000
                                                                     0.189225
                                                     0.189225
                                                                     1.000000
     lot_area_renov
     Number of schools nearby
                                                    -0.001203
                                                                    -0.025014
                                                                     -0.014587
     Distance from the airport
                                                    -0.005673
                                                     0.584924
                                            id
                                                           -0.004821
     Date
                                                           -0.004071
     number of bedrooms
                                                            0.003397
     number of bathrooms
                                                            0.002180
     living area
                                                            0.002370
     lot area
                                                            -0.012671
     number of floors
     waterfront present
                                                            0.001563
     number of views
                                                            0.008004
                                                           -0.006939
     condition of the house
     grade of the house
                                                            0.000986
     Area of the house(excluding basement)
                                                           -0.002894
     Area of the basement
                                                            0.010284
     Built Year
                                                           -0.001631
     Renovation Year
                                                           -0.000826
     Postal Code
                                                            0.010605
     Lattitude
                                                            0.014949
     Longitude
                                                           -0.010163
     living_area_renov
                                                           -0.001203
                                                           -0.025014
     lot area renov
     Number of schools nearby
                                                            1.000000
     Distance from the airport
                                                            0.004035
     Price
                                                            0.009890
                                            Distance from the airport
                                                                          Price
     id
                                                            -0.004542 -0.773114
     Date
                                                             0.011457 -0.027919
     number of bedrooms
                                                             -0.006157
                                                                       0.308460
     number of bathrooms
                                                             0.009206
                                                                       0.531735
     living area
                                                             0.002511
                                                                       0.712169
     lot area
                                                             0.003291
                                                                       0.081992
     number of floors
                                                             0.016567
                                                                       0.262732
     waterfront present
                                                             0.001448
                                                                       0.263687
                                                            -0.001657
     number of views
                                                                       0.395973
     condition of the house
                                                            -0.002136
                                                                       0.041376
     grade of the house
                                                             0.004940
                                                                       0.671814
     Area of the house(excluding basement)
                                                             0.001222
     Area of the basement
                                                             0.002926
     Built Year
                                                            -0.003968
     Renovation Year
                                                             0.005342 0.133173
     Postal Code
                                                             0.011528 -0.115908
     Lattitude
                                                             0.007193 0.297490
                                                            -0.003100
     Longitude
                                                                       0.024414
                                                            -0.005673
     living_area_renov
                                                                       0.584924
     lot_area_renov
                                                            -0.014587
                                                                       0.075535
     Number of schools nearby
                                                             0.004035
                                                                       0.009890
     Distance from the airport
                                                             1.000000
                                                                       0.003804
                                                             0.003804 1.000000
     [23 rows x 23 columns]
print(df['Number of schools nearby'].value_counts())
          4973
          4853
          4794
     Name: Number of schools nearby, dtype: int64
print('Mean:',df['Distance from the airport'].mean())
print('Median:',df['Area of the basement'].median())
print('Mode:',df['grade of the house'].mode())
     Mean: 64.95095759233926
     Median: 0.0
     Mode: 0
     Name: grade of the house, dtype: int64
Handling the Missing Values
print(df.isnull().sum())
     id
     Date
                                              0
     number of bedrooms
                                              0
     number of bathrooms
                                              0
     living area
                                              0
     lot area
                                              0
     number of floors
```

```
10/3/23, 10:19 AM
```

```
waterfront present
                                                a
                                                0
     number of views
     condition of the house
                                                0
     grade of the house
                                                0
     Area of the house(excluding basement)
                                                0
     Area of the basement
                                                0
     Built Year
     Renovation Year
                                                0
     Postal Code
                                                0
                                                0
     Lattitude
                                                a
     Longitude
                                                0
     living_area_renov
     lot_area_renov
                                                0
     \hbox{Number of schools nearby}
                                                0
     Distance from the airport
                                                0
     Price
                                                0
     dtype: int64
df.dropna(inplace=True)
df.fillna(0,inplace=True)
df.interpolate(inplace=True)
from sklearn.preprocessing import StandardScaler
from sklearn.preprocessing import MinMaxScaler
```

x.head()

y=df[['id','Price']]

x=df.drop(['Price','Date'],axis=1)
x.set_index(['id'],inplace=True)

	number of bedrooms	number of bathrooms	living area	lot area	number of floors	waterfront present	number of views	condition of the house	grade of the house	Area of the house(excluding basement)	Area of the basement	Built Year	Re
id													
6762810145	5	2.50	3650	9050	2.0	0	4	5	10	3370	280	1921	
6762810635	4	2.50	2920	4000	1.5	0	0	5	8	1910	1010	1909	
6762810998	5	2.75	2910	9480	1.5	0	0	3	8	2910	0	1939	
6762812605	4	2.50	3310	42998	2.0	0	0	3	9	3310	0	2001	
6762812919	3	2.00	2710	4500	1.5	0	0	4	8	1880	830	1929	

y.head()

```
id Price

0 6762810145 2380000

1 6762810635 1400000

2 6762810998 1200000

3 6762812605 838000

4 6762812919 805000
```

```
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.ensemble import GradientBoostingRegressor
from sklearn.metrics import r2_score

x_train,x_test,y_train,y_test = train_test_split(x,y['Price'],test_size =0.1,random_state=2)
model = GradientBoostingRegressor(n_estimators=400,max_depth=5,min_samples_split=2,learning_rate=0.1)
model.fit(x_train,y_train)

r GradientBoostingRegressor
```

GradientBoostingRegressor(max_depth=5, n_estimators=400)

y_pred_df

	ID	Predicted Price										
0	6762811233	497766.13	th									
1	6762811403	244495.38										
2	6762811775	293819.40										
3	6762811861	397555.35										
4	6762812009	474843.29										
1457	6762830250	1041014.57										
1458	6762830339	317512.59										
1459	6762830618	698495.60										
1460	6762830709	297006.00										
1461	6762831463	245881.77										
1462 ro	1462 rows × 2 columns											