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

Choose Files House Price India.csv

• House Price India.csv(text/csv) - 1524561 bytes, last modified: 10/3/2023 - 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(uploaded['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	Postal Code	Latt:
0	6762810145	42491	5	2.50	3650	9050	2.0	0	4	5		1921	0	122003	52
1	6762810635	42491	4	2.50	2920	4000	1.5	0	0	5		1909	0	122004	52
2	6762810998	42491	5	2.75	2910	9480	1.5	0	0	3		1939	0	122004	52
3	6762812605	42491	4	2.50	3310	42998	2.0	0	0	3		2001	0	122005	52
4	6762812919	42491	3	2.00	2710	4500	1.5	0	0	4		1929	0	122006	52

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	Postal Code
1	4615	6762830250	42734	2	1.5	1556	20000	1.0	0	0	4	 1957	0	122066
1	4616	6762830339	42734	3	2.0	1680	7000	1.5	0	0	4	 1968	0	122072
1	4617	6762830618	42734	2	1.0	1070	6120	1.0	0	0	3	 1962	0	122056
1	4618	6762830709	42734	4	1.0	1030	6621	1.0	0	0	4	 1955	0	122042
1	4619	6762831463	42734	3	1.0	900	4770	1.0	0	0	3	 1969	2009	122018

5 rows × 23 columns

df

```
number
                                                                                         number
                                                                                                 condition
                                                                     number
                                          number of living
                                                                lot
                                                                             waterfront
                                                                                                                 Built Renovation Postal
                                                                                                    of the ...
                          Date
                                      of
                                          bathrooms
                                                        area
                                                               area
                                                                                present
                                                                                                                   Year
                                                                                                                               Year
                                                                                                                                       Code
                                                                     floors
                                                                                          views
                                bedrooms
                                                                                                     house
             6762810145 42491
                                                                                                                                  0 122003
                                                2.50
                                                               9050
                                                                                                                  1921
                                                                                                            ...
df.columns
     'Area of the house(excluding basement)', 'Area of the basement',
            'Built Year', 'Renovation Year', 'Postal Code', 'Lattitude', 'Longitude', 'living_area_renov', 'lot_area_renov', 'Number of schools nearby', 'Distance from the airport', 'Price'],
           dtvpe='object')
                                                _...
df.dtypes
     id
                                                 int64
     Date
                                                 int64
     number of bedrooms
                                                 int64
     number of bathrooms
                                               float64
     living area
                                                 int64
                                                 int64
     lot area
     number of floors
                                               float64
     waterfront present
                                                 int64
     number of views
                                                 int64
     condition of the house
                                                 int64
     grade of the house
                                                 int64
     Area of the house(excluding basement)
                                                 int64
     Area of the basement
                                                 int64
     Built Year
                                                 int64
     Renovation Year
                                                 int64
     Postal Code
                                                 int64
                                               float64
     Lattitude
                                               float64
     Longitude
                                                 int64
     living_area_renov
     lot_area_renov
                                                 int64
     Number of schools nearby
                                                 int64
     Distance from the airport
                                                 int64
     Price
                                                 int64
     dtype: object
df.info()
     <class 'pandas.core.frame.DataFrame'>
     RangeIndex: 14620 entries, 0 to 14619
     Data columns (total 23 columns):
      #
          Column
                                                  Non-Null Count Dtype
     ---
      a
          id
                                                  14620 non-null
                                                                  int64
      1
                                                  14620 non-null
          Date
                                                                   int64
          number of bedrooms
                                                  14620 non-null
          number of bathrooms
                                                  14620 non-null
                                                                  int64
          living area
                                                  14620 non-null
                                                  14620 non-null
                                                                   int64
          lot area
          number of floors
                                                  14620 non-null
      6
                                                                   float64
                                                  14620 non-null
          waterfront present
                                                                   int64
      8
          number of views
                                                  14620 non-null
                                                                   int64
          condition of the house
                                                  14620 non-null
                                                                   int64
      10
          grade of the house
                                                  14620 non-null
                                                                   int64
          Area of the house(excluding basement) 14620 non-null
                                                  14620 non-null
          Area of the basement
          Built Year
                                                  14620 non-null
                                                  14620 non-null
      14
          Renovation Year
         Postal Code
                                                  14620 non-null
      15
                                                                   int64
                                                  14620 non-null
      16
          Lattitude
                                                                   float64
      17
                                                  14620 non-null
                                                                   float64
          Longitude
                                                  14620 non-null
      18
          living_area_renov
                                                                   int64
      19
         lot area renov
                                                  14620 non-null
                                                                   int64
      20
         Number of schools nearby
                                                  14620 non-null
                                                                   int64
      21
         Distance from the airport
                                                  14620 non-null
                                                                   int64
      22 Price
                                                  14620 non-null
     dtypes: float64(4), int64(19)
     memory usage: 2.6 MB
df.shape
```

Univariate Analysis

(14620, 23)

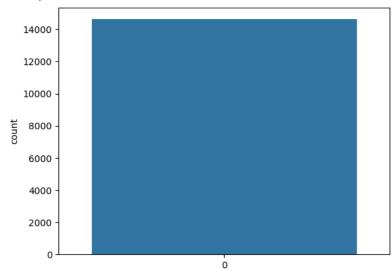
```
print(df.describe())
     std
            6.237575e+03
                              67.347991
                                                   0.938719
                                                                         0.769934
            6.762810e+09
                                                    1.000000
                                                                          0.500000
     min
                           42491.000000
     25%
            6.762815e+09
                          42546.000000
                                                   3.000000
                                                                         1.750000
                           42600.000000
                                                   3.000000
                                                                          2.250000
     50%
            6.762821e+09
     75%
            6.762826e+09
                           42662,000000
                                                   4.000000
                                                                          2,500000
                           42734,000000
                                                   33.000000
                                                                         8.000000
     max
            6.762832e+09
             living area
                               lot area number of floors waterfront present
     count
            14620.000000
                           1.462000e+04
                                             14620.000000
                                                                  14620.000000
     mean
             2098.262996
                           1.509328e+04
                                                 1.502360
                                                                      0.007661
              928.275721
                           3.791962e+04
                                                 0.540239
                                                                      0.087193
     std
              370.000000
                                                 1.000000
                                                                      0.000000
     min
                           5.200000e+02
                                                                      0.000000
     25%
             1440.000000
                           5.010750e+03
                                                 1.000000
     50%
             1930.000000
                           7.620000e+03
                                                 1.500000
                                                                      0.000000
             2570.000000
                          1.080000e+04
                                                                      0.000000
     75%
                                                  2.000000
            13540.000000
                          1.074218e+06
                                                  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
                                                       . . .
     25%
                   0.000000
                                            3.000000
                                                             1951.000000
                                                       . . .
     50%
                   0.000000
                                            3.000000
                                                             1975.000000
                                                       . . .
     75%
                   0.000000
                                            4.000000
                                                             1997.000000
                                                       . . .
                                            5.000000
                   4.000000
                                                             2015.000000
     max
            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
                   0.000000
                              122003.000000
                                                 52.385900
                                                             -114.709000
     min
     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
                2015.000000
                             122072.000000
                                                53.007600
                                                             -113.505000
     max
            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
                  1850.000000
                                   7620.000000
                                                                 2.000000
     50%
     75%
                  2380.000000
                                  10125.000000
                                                                 3.000000
                  6110.000000
                                 560617.000000
                                                                 3.000000
     max
            Distance from the airport
                                               Price
     count
                         14620.000000 1.462000e+04
                             64.950958
                                       5.389322e+05
     mean
     std
                              8,936008
                                        3.675324e+05
     min
                             50.000000
                                        7.800000e+04
     25%
                             57.000000
                                        3.200000e+05
                                        4.500000e+05
     50%
                             65.000000
     75%
                             73.000000
                                        6.450000e+05
                             80.000000
                                        7.700000e+06
     max
```

plt.hist(df['Price'])

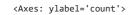
[8 rows x 23 columns]

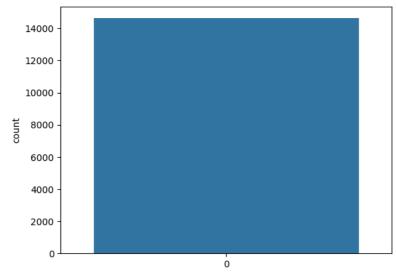
sns.countplot(df['grade of the house'])

<Axes: ylabel='count'>



sns.countplot(df['living_area_renov'])





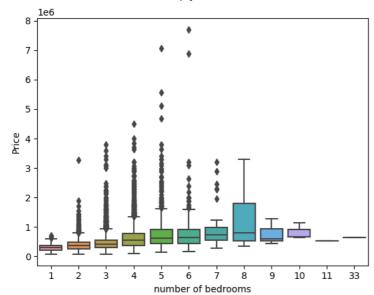
sns.boxplot(x=df['Price'])

<Axes: xlabel='Price'>

Bivariate Analysis

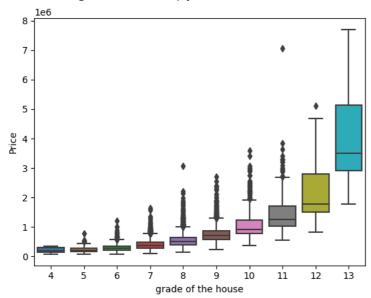
| sns.boxplot(x=df['number of bedrooms'],y=df['Price'])

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

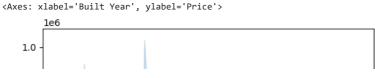


 $\verb|sns.boxplot(x=df['grade of the house'],y=df['Price'])|\\$

<Axes: xlabel='grade of the house', ylabel='Price'>

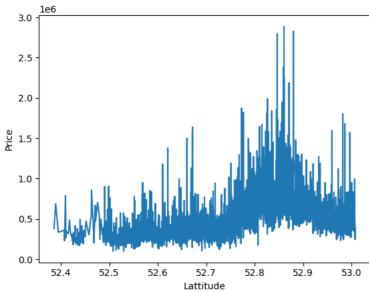


sns.lineplot(x=df['Built Year'],y=df['Price'])

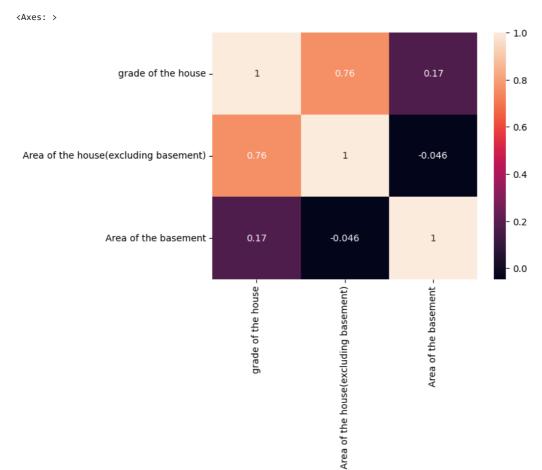


 $sns.lineplot(x=df.groupby('Lattitude').mean().index,y=df.groupby('Lattitude').mean()['Price'])\\ plt.show$

<function matplotlib.pyplot.show(close=None, block=None)>



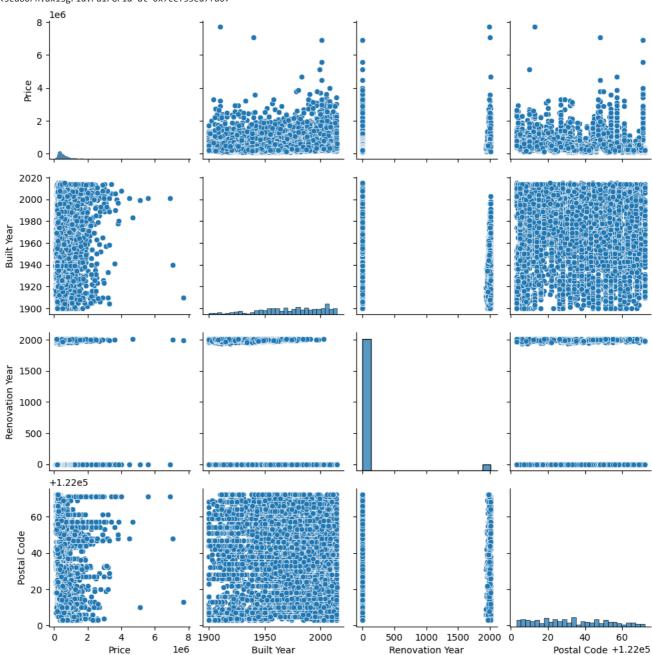
 $sns.heatmap(df[['grade of the house', \\ 'Area of the house(excluding basement)', 'Area of the basement']].corr(), annot=True)$



Multivariate Analysis

sns.pairplot(df[['Price','Built Year', 'Renovation Year', 'Postal Code']])

<seaborn.axisgrid.PairGrid at 0x7cef55ed7fa0>



df.duplicated().sum()

0

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

```
Waterfront present vs Price
       1e6
sns.scatterplot(df['number of views'])
    <Axes: ylabel='number of views'>
            3.5
       3.0
    number of views
       2.5
       2.0
       1.5
       1.0
       0.5
       0.0
            0
                 2000
                        4000
                              6000
                                     8000
                                           10000
                                                 12000
                                                        14000
```

```
plt.subplots(figsize=(17,17))
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	-0.016	-0.016	0.0059	-0.012	0.018	-0.023	-0.018	-0.032	-5e-05	0.00
	number of bedrooms	-0.016	1	0.51	0.57	0.034	0.18	0.0063	0.079	0.027	0.35	0.47	0.3	0.15	0.016	-0.044	-0.013	0.14	0.39	0.029	0.003
	number of bathers and	0.036	0.51	1	0.75	0.001	0.5	0.06	0.10	0.13	0.66	0.60	0.20	٥٢	0.05	0.11	0.001	0.22	0.57	0.070	0.00
	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	0.57	0.079	0.00
	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	0.76	0.18	0.00
	lot area	0.0044	0.034	0.081	0.17	1	0.0041	0.026	0.078	0.0085	0.11	0.18	0.02	0.052	0.0068	0.07	-0.091	0.22	0.15	0.71	-0.0
	number of floors -	-0.01	0.18	0.5	0.35	0.0041	1	0.016	0.02	-0.27	0.46	0.53	-0.24	0.48	0.0067	-0.13	0.051	0.13	0.29	-0.01	0.00
		0.010	0.005	0.05	0.11	0.005	0.016	,	0.4	0.010	0.00	0.070	0.005	0.024	0.005	0.030	0.022	0.010	0.005	0.022	0.00
	waterfront present -	0.012	0.0003	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	0.086	0.032	0.00
	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€	-0.08	0.28	0.072	0.0
nt(df.de	escribe())		0.007	0.10	0.063					•	0.15	0.17	0.10	0.00	0.050	0.045					ا
std	6.237575e+03 67.3479	991		e	.9387	719			0.769	934											
min	6.762810e+09 42491.0000		0.500																		
25% 50%	6.762815e+09 42546.0000 6.762821e+09 42600.0000				.0000				1.756 2.256												
75%	6.762826e+09 42662.000				.0000				2.500												
max	6.762832e+09 42734.0000				.0000				8.000												
	living area lot a	non n	umbar	of f	loons	- wat	onfno	nt nr	nacant	. \											
count	living area lot area number of floors waterfront present \ 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																				
min	370.000000 5.200000e+02 1.000000 0.000000																				
25%	1440.000000 5.010750e+03 1.000000 0.0000000																				
50% 75%	1930.000000 7.620000e+03 1.500000 0.0000000 2570.000000 1.0800000+04 2.000000 0.000000																				
7 370	2570.000000 1.080000e+04 2.000000 0.000000 13540.000000 1.074218e+06 3.500000 1.000000																				
max	13540.000000 1.074218e-	⊦06		3.5	00000	9		1.6													
max	13540.000000 1.074218e-		the				Built		90000												
count	number of views condit: 14620.000000	ion of	1620.0	house	·	. E	520.00	Year	90000												
count mean	number of views condit: 14620.000000 0.233105	ion of	1620.6 3.4	house 000000 130506		. E . 146	520.00 970.92	Year 00000 26402	90000												
count mean std	number of views condit: 14620.000000 0.233105 0.766259	ion of	1620.0 3.4 0.6	house 900000 130506		. E . 146 . 19	520.06 970.92 29.49	Year 90000 26402 93625	900000												
count mean std min	number of views condit: 14620.000000 0.233105 0.766259 0.000000	ion of	1620.0 3.4 0.6 1.0	house 100000 130506 164151		. E	520.06 970.92 29.49 900.06	Year 00000 26402 03625 00000	900000												
count mean std	number of views condit: 14620.000000 0.233105 0.766259 0.000000 0.000000	ion of	1620.6 3.4 0.6 1.6 3.6	house 100000 130506 164151 100000		. E. 146 . 19 . 19	520.06 970.92 29.49 900.06	Year 00000 26402 03625 00000	900000												
count mean std min 25%	number of views condit: 14620.000000 0.233105 0.766259 0.000000	ion of	1620.6 3.4 0.6 1.6 3.6 3.6	house 100000 130506 164151		. 146 . 19 . 19 . 19	520.06 970.92 29.49 900.06	Year 00000 26402 93625 00000 00000	900000												
count mean std min 25% 50%	number of views condit: 14620.000000 0.233105 0.766259 0.000000 0.000000	ion of	1620.6 3.4 0.6 1.6 3.6 3.6	house 100000 130506 164151 100000 100000		. 146 . 146 . 19 . 19 . 19	520.06 970.92 29.49 900.06 951.06	Year 90000 26402 93625 90000 90000 90000	900000												
count mean std min 25% 50% 75%	number of views condit: 14620.000000 0.233105 0.766259 0.000000 0.000000 0.0000000 4.0000000	ion of	1620.6 3.4 0.6 1.6 3.6 4.6 5.6	house 90000 130506 664151 900000 900000 900000		. E. 146 . 196 . 199 . 199 . 199 . 199	20.00 270.92 29.49 20.00 251.00 275.00 275.00 215.00	Year 90000 26402 93625 90000 90000 90000	\												
count mean std min 25% 50% 75% max	number of views 14620.000000 0.233105 0.766259 0.000000 0.000000 0.000000 0.000000 0.000000	ion of 14 al Cod	3.4 0.6 1.0 3.0 4.0 5.0	house 900000 30506 664151 900000 900000 900000 Latt	itude	. E . 146 . 19	520.06 970.92 29.49 900.06 951.06 975.06 997.06 15.06 Longi	Year 90000 26402 93625 90000 90000 90000 1tude	\												
count mean std min 25% 50% 75% max count	number of views 14620.000000	ion of 14 31 Cod .00000 .06224	1620.6 3.4 0.6 1.6 3.6 4.6 5.6	house 900000 130506 664151 900000 900000 900000 Latt	itude	. E. 146 . 19 . 19 . 19 . 19 . 19 . 20	520.06 970.92 29.49 900.06 951.06 975.06 997.06 115.06 Longi	Year 90000 26402 93625 90000 90000 90000 1tude 90000 94007	\												
count mean std min 25% 50% 75% max count mean std	number of views 14620.000000 0.233105 0.766259 0.000000 0.000000 0.000000 4.000000 4.000000 14620 90.924008 122033 416.216661 19	ion of 14 al Cod .00006 .06224 .08241	1620.6 3.4 0.6 1.6 3.6 4.6 5.6 16 10 14	house 900000 130506 664151 900000 900000 Latt 6620.6 52.7 0.1	itude 92848 37522	. E 146 . 19	520.06 970.92 29.45 900.06 951.06 975.06 915.06 Longi 520.06 114.46 0.14	Year 90000 26402 93625 90000 90000 90000 90000 1tude 90000 94007	\												
count mean std min 25% 50% 75% max count mean std min	number of views 14620.000000 0.233105 0.766259 0.000000 0.000000 0.000000 0.000000 4.000000 14620 90.924008 122033 416.216661 19 0.000000 122003	ion of 14 al Cod .00006 .06224 .08241	1620.6 3.4 0.6 1.0 3.0 4.0 5.0 14 14 18	house 900000 330506 664151 900000 900000 000000 Latt 6620.6 52.7 0.1	itude 92848 37522	. E . 146	520.06 970.92 29.49 900.06 951.06 975.06 915.06 Longi 520.06 114.76	Year 90000 26402 93625 90000 90000 90000 90000 1tude 90000 94007 11326	\												
count mean std min 25% 50% 75% max count mean std	number of views 14620.000000 0.233105 0.766259 0.000000 0.000000 0.000000 4.000000 4.000000 14620 90.924008 122033 416.216661 19	ion of 14 al Cod .00006 .06224 .08241 .00006	1620.6 3.4 0.6 1.6 3.6 4.6 5.6 le 100 14	house 900000 130506 64151 900000 900000 100000 Latt 1620.6 52.7 0.1 52.3	itude 92848 37522	. E. 146. 19	520.06 970.92 29.45 900.06 951.06 975.06 915.06 Longi 520.06 114.46 0.14	Year 00000 26402 33625 00000 00000 00000 00000 1tude 00000 14007 11326 09000	\												
count mean std min 25% 50% 75% max count mean std min 25%	number of views 14620.000000 0.233105 0.766259 0.000000 0.000000 0.000000 0.000000 4.000000 14620 14620 90.924008 122033 146.216661 19 0.000000 122017 0.000000 122017 0.000000 122018	al Cod .00006 .06224 .08241 .00006 .00006	620.6 3.4 0.6 1.6 3.6 4.6 5.6 14 14 14 18 90 90	house 900000 130506 64151 900000 900000 100000 Latt 620.6 52.7 0.1 52.3 52.7	itude 1000000 192848 37522 85900	. E. 146. 19	520.06 970.92 29.49 900.06 951.06 975.06 997.06 Longi 520.06 114.46 0.14 114.51	Year 00000 26402 33625 00000 00000 00000 00000 1tude 00000 14007 11326 09000 21000	\												
count mean std min 25% max count mean std min 25% 50% 75% max	number of views 14620.000000 0.233105 0.766259 0.000000 0.000000 0.000000 0.000000 4.000000 14620 90.924008 122033 416.216661 19 0.000000 122017 0.000000 122048 2015.000000 122072	al Cod .00006 .06224 .08241 .00006 .00006	620.6 3.4 0.6 1.6 3.6 4.6 5.6 14 14 14 18 90 90	house 900000 130506 664151 900000 900000 900000 Latt 6620.6 52.7 0.1 52.3 52.7 52.8	itude 100000 192848 37522 85900	. E. 146. 19	520.06 970.92 29.49 900.06 951.06 975.06 997.06 Longi 520.06 114.46 0.14 114.51	Year 90000 26402 93625 90000 90000 90000 1tude 90000 14007 11326 99000 21000	\												
count mean std min 25% 50% 75% max count mean std min 25% 50% 75% max	number of views condit: 14620.000000 0.233105 0.766259 0.0000000 0.000000 0.000000 4.000000 Renovation Year Posta 14620.000000 14620 90.924008 122033 416.216661 19 0.000000 122003 0.000000 122017 0.000000 122017 0.000000 122048 2015.000000 122072	al Cod .00006 .06224 .08241 .00006 .00006	3.4 0.6 1.6 3.6 3.6 4.6 5.6 de 90 14 4.4 8.8 90 90 90	house 130506	itude 000000 92848 37522 85906 007600	. E E 1446 149 149 149 149 149 149 149 149 149 149	520.06 970.92 29.49 900.06 951.06 975.06 997.06 114.46 0.14 114.51 114.51 114.31	Year 100000 16402 13625 1000000	\ \												
count mean std min 25% 75% max count mean std min 25% 50% 75%	number of views condit: 14620.000000 0.233105 0.766259 0.0000000 0.000000 0.000000 4.000000 Renovation Year Posta 14620.000000 14620 90.924008 122033 416.216661 19 0.000000 122003 0.000000 122017 0.000000 122032 0.000000 122048 2015.0000000 122072 living_area_renov lot_1	ion of 14 14 14 16 17 18 18 18 18 18 18 18 18 18	3.4 0.6 1.6 3.6 3.6 4.6 5.6 6 14 14 8.8 90 90 90 90 90 90	house 130506	itude 000000 92848 37522 85906 007600	. E E 1446 149 149 149 149 149 149 149 149 149 149	520.06 770.92 29.45 800.06 8051.06 8075.06 8075.06 8075.06 8075.06 8074 8074 8074 8074 8074 8074 8074 8074	Year 100000 16402 13625 1000000) 												
count mean std min 25% 50% 75% max count mean std min 25% 50% 75% max count count count count	number of views condit: 14620.000000 0.233105 0.766259 0.0000000 0.0000000 0.0000000 4.0000000 14620.0000000 1220033 416.216661 19 0.000000 122003 0.000000 122017 0.000000 122032 0.000000 122048 2015.0000000 1240 1996.702257 125	ion of 14 14 14 14 16 17 18 18 18 18 18 18 18 18 18	3.4 0.6.20.0 1.6 3.6 3.6 5.6 5.6 14 4.8 90 90 90 90 90 90 90 90 90 90 90 90 90	house 130506	itude 000000 92848 37522 85906 007600	. E E 1446 149 149 149 149 149 149 149 149 149 149	520.06 770.92 29.45 29.45 6000.06 600000000000000000000000000000	Year 199000 26402 23625 23625 23625 246000 200000 200000 200000 200000 24007 11326 299000 210000 250000 250000 250000 250000 250000 250000 250000 250000 250000 250000 250000 2500000000	\ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \												
count mean std min 25% 50% 75% max count mean std min 25% 50% 75% max count mean std min std min	number of views condit: 14620.000000 0.233105 0.766259 0.000000 0.000000 0.000000 0.000000 0.000000	al Cod .00000 .06224 .08241 .00000 .00000 .00000 .00000 .00000	3.4 0.6 1.6 3.6 3.6 4.6 4.6 4.6 4.8 8 90 90 90 90 90 90 90 90 90 90 90 90 90	house 130506	itude 000000 92848 37522 85906 007600	. E E 1446 149 149 149 149 149 149 149 149 149 149	520.06 770.92 29.44 29.49 600.06 600.07	Year 100000 0.6402 100000 0.00000 0.00000 0.00000 0.00000 0.00000 0.00000 0.00000 0.00000 0.00000 0.00000 0.00000 0.00000 0.000000	/ / \												
count mean std min 25% 50% 75% max count mean std min 25% 50% 75% max count mean std min 25%	number of views condit: 14620.000000 0.233105 0.766259 0.0000000 0.000000 0.0000000 0.0000000 0.000000	al Cod 00000 06224 08241 00000 00000 00000 00000 573.56 578.41	3.4.0.6.1.0.3.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0	house 130506	itude 000000 92848 37522 85906 007600	. E E 1446 149 149 149 149 149 149 149 149 149 149	520.06 770.92 29.45 8000.06 8000.06 8075.06 8075.06 8075.06 80114.46 8014 8014 8015.06 8016 8016 8016 8016 8016 8016 8016 80	Year 100000 16402) / / 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1												
count mean std min 25% 50% 75% max count mean std min 25% 50% 75% max count mean std min 25% 50%	number of views condit: 14620.000000 0.233105 0.766259 0.0000000 0.0000000 0.0000000 0.0000000	ion of 14 14 14 14 16 17 18 18 18 18 18 18 18 18 18	3.4 0.6.3 3.6 3.6 3.6 5.6 5.6 6 6 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9	house 130506	itude 000000 92848 37522 85906 007600	. E E 1446 149 149 149 149 149 149 149 149 149 149	520.06 770.92 29.44 7800.06 775.06 77	Year 10000 16402 1	\ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \												
count mean std min 25% 50% 75% max count mean std min 25% 50% 75% max count mean std min 25%	number of views condit: 14620.000000 0.233105 0.766259 0.0000000 0.0000000 0.0000000 0.0000000	al Cod 00000 06224 08241 00000 00000 00000 00000 573.56 578.41	3.4 0.6.3.0 3.6 3.6 5.6 5.6 14.4 14.4 18.8 190 190 190 190 190 190 190 190 190 190	house 130506	itude 000000 92848 37522 85906 007600	. E E 1446 149 149 149 149 149 149 149 149 149 149	520.06 770.92 29.45 29.45 29.45 29.45 3775.06	Year 100000 16402	\ / \ / \ / \ / \ / \ / \ / \ / \ / \ /												
count mean std min 25% 50% 75% max count mean std min 25% 50% 75% max count mean std min 25% 50% 75% 75% 75% 75%	number of views 14620.000000 0.233105 0.766259 0.000000 0.000000 0.000000 0.000000 0.000000	al Cod .00000 .06224 .08241 .000000	3.4 0.6.20.0 1.0 3.0 3.0 4.0 5.0 5.0 144 4.8 80 90 90 90 90 90 90 90 90 90 90 90 90 90	house 000000 130506 664151 000000 000000 000000 Latt 6620.6 52.7 6.1.1 52.3 52.8 52.8 53.6	itude 000000 92848 337522 85906 007606 007606	. E E 1446 149 149 149 149 149 149 149 149 149 149	520.06 770.92 29.45 29.45 29.45 29.45 3775.06	Year 100000 26402 2500000 26402 2500000 26402 250000 260000 260000 2600000 26000000 2600000000	\ / \ / \ / \ / \ / \ / \ / \ / \ / \ /												
count mean std min 25% 50% 75% max count mean std min 25% 50% 75% max count mean std min 25% 50% 75%	number of views condit: 14620.000000 0.233105 0.766259 0.0000000 0.0000000 0.0000000 0.0000000	al Cod .00000 .06224 .08241 .000000	3.4 0.6.20.0 1.0 3.0 3.0 4.0 5.0 5.0 144 4.8 80 90 90 90 90 90 90 90 90 90 90 90 90 90	house 1000000 130506 1604151 1000000 1000000 1000000 1000000 152.7 0.1.1 52.3 52.7 52.8 52.8 53.6 Numb	itude 00000 92848 37522 85906 00640 007606 007606	. E E 1446 149 149 149 149 149 149 149 149 149 149	520.06 770.92 29.45 29.45 29.45 29.45 3775.06	Year 100000 26402 2500000 26402 2500000 26402 250000 260000 260000 2600000 26000000 2600000000	\ / \ / \ / \ / \ / \ / \ / \ / \ / \ /												
count mean std min 25% 50% 75% max count mean std min 25% 50% 75% max count mean std min 25% 50% 75% max	number of views 14620.000000 0.233105 0.766259 0.000000 0.000000 0.000000 0.000000 4.000000 14620 90.924008 122033 416.216661 19 0.000000 122017 0.000000 122017 0.000000 122032 0.000000 122032 0.000000 122032 0.000000 122032 14620.000000 122032 0.000000 122032 0.000000 122032 0.000000 122032 0.000000 122032 0.000000 122032 0.000000 122032 0.000000 122032 0.000000 122032 0.000000 122032 0.000000 122032 0.000000 122032 0.0000000 122032 0.0000000 122032 0.0000000 122032 0.0000000 122032 0.0000000 140000000 5600000000000000000000000000000	ion of 14 all Cod 00000 06224 08241 00000 00000 00000 00000 000000 15551.00 125.00 1100000 110000000000000000000000000	3.4 0.6.3.6 3.6 3.6 5.6 5.6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6	house 000000 330506 664151 000000 000000 000000 52.7 0.1 52.3 52.8 52.8 53.0 Numb	itude 000000 92848 37522 85900 006400 007600	. E E 1446 149 149 149 149 149 149 149 149 149 149	520.06 770.92 29.45 29.45 29.45 29.45 3775.06	Year 100000 26402 2500000 26402 2500000 26402 250000 260000 260000 2600000 26000000 2600000000	\ / \ / \ / \ / \ / \ / \ / \ / \ / \ /												
count mean std min 25% 50% 75% max count	number of views condit: 14620.000000 0.233105 0.766259 0.0000000 0.000000 0.0000000 0.0000000 0.000000	ion of 14 all Coc 000000 000000 000000 000000 000000 0000	3.4.0.6.3.6.3.6.3.6.3.6.3.6.3.6.3.6.3.6.3.6	house 1000000 130506 130506 130506 1000000 1000000 1000000 1000000 152.3 52.7 6.1 52.8 52.8 52.8 53.6 Numb	itude 00000092848 37522 859007600 007600	. E E 1446 149 149 149 149 149 149 149 149 149 149	520.06 770.92 29.45 29.45 29.45 29.45 3775.06	Year 100000 26402 2500000 26402 2500000 26402 250000 260000 260000 2600000 26000000 2600000000	\ / \ / \ / \ / \ / \ / \ / \ / \ / \ /												
count mean std min 25% 50% 75% max count mean std min 25% 50% 75% max count mean std min 25% 50% 75% max count mean std min std min std min	number of views 14620.000000 0.233105 0.766259 0.0000000 0.0000000 0.0000000 0.0000000	ion of 14 14 14 16 17 17 17 18 18 18 18 18 18 18	3.4.0.6.1.6.3.6.3.6.3.6.3.6.3.6.3.6.3.6.3.6.3	house 1000000 130506 6641511 10000000 1000000 1000000 1000000 1000000	itude 00000092848 337522 85900 007600 007600	. E E 1446 149 149 149 149 149 149 149 149 149 149	520.06 770.92 29.45 29.45 29.45 29.45 3775.06	Year 100000 26402 2500000 26402 2500000 26402 250000 260000 260000 2600000 26000000 2600000000	\ / \ / \ / \ / \ / \ / \ / \ / \ / \ /												
count mean std min 25% 50% 75% max	number of views 14620.000000 0.233105 0.766259 0.0000000 0.0000000 0.0000000 0.0000000	al Cod .00000 .06224 .08241 .000000	3.4 0.6.6.3.4 3.6 3.6 5.6 5.6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6	house 000000 130506 130506 100000 100000 100000 100000 100000 100000 100000 100000 100000 100000 100000 100000 100000 100000 1000000	itude 000000 92848 337522 85906 007606 007606	. E E 1446 149 149 149 149 149 149 149 149 149 149	520.06 770.92 29.45 29.45 29.45 29.45 3775.06	Year 100000 26402 2500000 26402 2500000 26402 250000 260000 260000 2600000 26000000 2600000000	\ / \ / \ / \ / \ / \ / \ / \ / \ / \ /												
count mean std min 25% 50% 75% max count mean std min 25% 50% 75% max count mean std min 25% 50% 75% max count mean std min 25% 50% 75% 50%	number of views 14620.000000 0.233105 0.766259 0.0000000 0.0000000 0.0000000 0.0000000	al Cod .00000 .06224 .08241 .000000	3.4 0.6.6 3.4 3.6 3.6 5.6 5.6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6	house 1000000 130506 164151 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 10000000 1000000 1000000 1000000 1000000 1000000 1000000 10000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 10000000 1000000 1000000 1000000 1000000 1000000 1000000 10000000 1000000 1000000 1000000 1000000 1000000 1000000 10000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 10000000 1000000 1000000 1000000 1000000 1000000 1000000 10000000 1000000 1000000 1000000 1000000 1000000 1000000 10000000 1000000 1000000 1000000 1000000 1000000 1000000 10000000 1000000 1000000 100000 1000000 1000000 1000000 1000000 1000000 10000000 100000 1000000 1000000 1000000 1000000 1000000 100000000	itude 000000 92848 37522 85906 00640 007606 007606	. E E 1446 149 149 149 149 149 149 149 149 149 149	520.06 770.92 29.45 29.45 29.45 29.45 3775.06	Year 100000 26402 2500000 26402 2500000 26402 250000 260000 260000 2600000 26000000 2600000000	\ / \ / \ / \ / \ / \ / \ / \ / \ / \ /												
count mean std min 25% 50% 75% max	number of views 14620.000000 0.233105 0.766259 0.0000000 0.0000000 0.0000000 0.0000000	al Cod 000000 06224 08241 000000 000000 000000 000000 125.00 000000 125.00 000000 125.00 000000 125.00 000000 000000 000000 000000 000000 0000	3.4 0.6.6.3.4 3.6 3.6 5.6 5.6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6	house 1000000 130506 1600000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 10000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 10000000 1000000 1000000 1000000 1000000 1000000 1000000 10000000 1000000 1000000 1000000 1000000 1000000 1000000 10000000 1000000 1000000 1000000 1000000 1000000 1000000 10000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 1000000 10000000 1000000 1000000 1000000 1000000 1000000 1000000 10000000 1000000 1000000 1000000 1000000 1000000 1000000 10000000 1000000 1000000 1000000 1000000 1000000 10000000 10000000 100	itude 000000 92848 37522 85900 006400 007600 007600	. E E 1446 149 149 149 149 149 149 149 149 149 149	520.06 770.92 29.45 29.45 29.45 29.45 3775.06	Year 100000 26402 2500000 26402 2500000 26402 250000 260000 260000 2600000 26000000 2600000000	\ / \ / \ / \ / \ / \ / \ / \ / \ / \ /												

print(df.count())

```
14620
     Date
                                               14620
                                               14620
     number of bedrooms
     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
     grade of the house
                                               14620
     Area of the house(excluding basement)
                                               14620
     Area of the basement
                                               14620
     Built Year
                                               14620
     Renovation Year
                                               14620
     Postal Code
                                               14620
     Lattitude
                                               14620
     Longitude
                                               14620
                                               14620
     living_area_renov
     lot_area_renov
                                               14620
     Number of schools nearby
                                               14620
     Distance from the airport
                                               14620
                                               14620
     dtype: int64
print(df.corr())
     Longitude
                                                      0.341221
                                                                       0.258066
     living area renov
                                                      1.000000
                                                                       0.189225
                                                                      1.000000
                                                      0.189225
     lot area renov
                                                                      -0.025014
     Number of schools nearby
                                                     -0.001203
     Distance from the airport
                                                     -0.005673
                                                                      -0.014587
     Price
                                                      0.584924
                                                                       0.075535
                                             Number of schools nearby
     id
                                                            -0.004821
     Date
                                                             -0.004071
     number of bedrooms
                                                             0.003397
     number of bathrooms
                                                              0.002180
                                                             0.002370
     living area
                                                             -0.012671
     lot area
     number of floors
                                                             -0.007579
                                                             0.001563
     waterfront present
     number of views
                                                             0.008004
     condition of the house
                                                             -0.006939
     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
                                                             0.014949
     Lattitude
     Longitude
                                                             -0.010163
     living_area_renov
                                                             -0.001203
     lot_area_renov
                                                             -0.025014
     Number of schools nearby
                                                             1.000000
     Distance from the airport
                                                              0.004035
     Price
                                                             0.009890
                                             Distance from the airport
     id
                                                              -0.004542 -0.773114
                                                              0.011457 -0.027919
     Date
     number of bedrooms
                                                                        0.308460
                                                              -0.006157
     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
     number of views
                                                              -0.001657
     condition of the house
                                                              -0.002136
                                                                         0.041376
     grade of the house
                                                              0.004940
                                                                         0.671814
                                                              0.001222
     Area of the house(excluding basement)
                                                                         0.615220
     Area of the basement
                                                              0.002926
                                                                         0.330202
                                                              -0.003968
     Built Year
                                                                         0.050307
     Renovation Year
                                                              0.005342 0.133173
     Postal Code
                                                              0.011528 -0.115908
     Lattitude
                                                              0.007193
                                                                         0.297490
     Longitude
                                                              -0.003100
                                                                         0.024414
     living_area_renov
                                                              -0.005673
                                                                         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['grade of the house'].value_counts())

```
10/3/23, 3:28 PM
        8
               4137
        9
               1828
        6
               1324
        10
               804
                280
        11
               154
        5
        12
                55
        4
                17
        13
                10
        Name: grade of the house, dtype: int64
   print('Mean: ',df['Price'].mean())
   print('Median: ',df['Distance from the airport'].median())
   print('Mode: ',df['grade of the house'].mode())
        Mean: 538932.2183310534
        Median: 65.0
        Mode: 0
        Name: grade of the house, dtype: int64
   print(df.isnull().sum())
        id
                                                  9
        Date
                                                  0
        number of bedrooms
        number of bathrooms
                                                  0
        living area
                                                  0
        lot area
        number of floors
                                                  0
        waterfront present
                                                  0
        number of views
                                                  0
        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
        Lattitude
                                                  0
        Longitude
                                                  0
        living_area_renov
                                                  0
        lot area renov
                                                  0
        Number of schools nearby
                                                  0
        Distance from the airport
                                                  0
        Price
        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=df.drop(['Price','Date'],axis=1)
   x.set_index(['id'],inplace=True)
```

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

x.head()

	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	Renov
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()

```
\blacksquare
                  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 \ Gradient Boosting Regressor
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)
                       GradientBoostingRegressor
     GradientBoostingRegressor(max_depth=5, n_estimators=400)
y_pred = model.predict(x_test)
model.score(x_test,y_test)
     0.9122400099200828
r2_score(y_pred,y_test)
     0.90195538227743
y_pred
     array([497766.12740438, 244495.3776842 , 293819.40063242, ..., 698495.60350629, 297006.00386358, 245881.76921871])
y_pred_list = y['id'][-len(y_pred):].tolist()
y_pred_df=pd.DataFrame(y_pred_list,columns=['ID'])
y_pred_df['Predicted Price']= y_pred.round(2)
y_pred_df
\vdash
                     ID Predicted Price
                                             扁
        0
            6762811233
                                497766.13
            6762811403
                                244495.38
        1
            6762811775
                                293819.40
                                397555.35
        3
            6762811861
       4
            6762812009
                                474843.29
      1457 6762830250
                               1041014.57
            6762830339
                                317512.59
      1458
      1459
            6762830618
                                698495.60
      1460 6762830709
                                297006.00
      1461 6762831463
                                245881.77
     1462 rows × 2 columns
```