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

Choose Files No file chosen Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.

Saving House Price India csv to House Price India (1) csv

import pandas as pd
import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

import io

df = pd.read_csv("/content/House Price India.csv")

df.head()

 \Box

	id	Date	number of bedrooms	number of bathrooms	living area	lot area	number of floors	waterfront present	number of views
(6762810145	42491	5	2.50	3650	9050	2.0	0	4
1	6762810635	42491	4	2.50	2920	4000	1.5	0	0
2	6762810998	42491	5	2.75	2910	9480	1.5	0	0
3	6762812605	42491	4	2.50	3310	42998	2.0	0	0
4	6762812919	42491	3	2.00	2710	4500	1.5	0	0

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	14615	6762830250	42734	2	1.5	1556	20000	1.0	0	0	4		1957	0	122066	
1	14616	6762830339	42734	3	2.0	1680	7000	1.5	0	0	4		1968	0	122072	
1	14617	6762830618	42734	2	1.0	1070	6120	1.0	0	0	3		1962	0	122056	
1	14618	6762830709	42734	4	1.0	1030	6621	1.0	0	0	4		1955	0	122042	
1	14619	6762831463	42734	3	1.0	900	4770	1.0	0	0	3		1969	2009	122018	
1	14618	6762830618 6762830709	42734	2	1.0	1070 1030	6120 6621	1.0	0	0	4		1962 1955	0	1220 1220)56)42

5 rows × 23 columns

df

	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
0	6762810145	42491	5	2.50	3650	9050	2.0	0	4	5		1921	0	122003
1	6762810635	42491	4	2.50	2920	4000	1.5	0	0	5		1909	0	122004
2	6762810998	42491	5	2.75	2910	9480	1.5	0	0	3		1939	0	122004
3	6762812605	42491	4	2.50	3310	42998	2.0	0	0	3		2001	0	122005
4	6762812919	42491	3	2.00	2710	4500	1.5	0	0	4		1929	0	122006
14615	6762830250	42734	2	1.50	1556	20000	1.0	0	0	4		1957	0	122066
14616	6762830339	42734	3	2.00	1680	7000	1.5	0	0	4		1968	0	122072
14617	6762830618	42734	2	1.00	1070	6120	1.0	0	0	3		1962	0	122056
14618	6762830709	42734	4	1.00	1030	6621	1.0	0	0	4		1955	0	122042
14619	6762831463	42734	3	1.00	900	4770	1.0	0	0	3		1969	2009	122018
14620 rd	ows × 23 colum	ns												

```
df.columns
    'number of views', 'condition of the house', 'grade of the house',
            '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'],
           dtype='object')
df.dtypes
     id
                                               int64
    Date
                                               int64
     number of bedrooms
                                               int64
     number of bathrooms
                                             float64
     living area
                                               int64
     lot area
                                               int64
     number of floors
                                             float64
     waterfront present
                                               int64
     number of views
                                               int64
     condition of the house
                                               int64
     grade of the house
     Area of the house(excluding basement)
                                               int64
     Area of the basement
                                               int64
     Built Year
                                               int64
     Renovation Year
                                               int64
     Postal Code
                                               int64
    Lattitude
                                             float64
     Longitude
                                             float64
     living_area_renov
                                               int64
     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
     0 id
                                                14620 non-null int64
     1
         Date
                                                14620 non-null int64
     2
         number of bedrooms
                                                14620 non-null int64
     3
         number of bathrooms
                                                14620 non-null float64
     4
         living area
                                                14620 non-null int64
                                                14620 non-null int64
         lot area
         number of floors
                                                14620 non-null
                                                                float64
                                                14620 non-null int64
         waterfront present
     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 int64
     11
     12 Area of the basement
                                                14620 non-null int64
     13 Built Year
                                                14620 non-null int64
     14 Renovation Year
                                                14620 non-null int64
     15
         Postal Code
                                                14620 non-null int64
     16 Lattitude
                                                14620 non-null
                                                                float64
         Longitude
                                                14620 non-null
                                                                float64
                                                14620 non-null int64
     18 living area renov
                                                14620 non-null int64
     19 lot area renov
     20 Number of schools nearby
                                                14620 non-null int64
     21
         Distance from the airport
                                                14620 non-null
                                                                int64
     22 Price
                                                14620 non-null int64
     dtypes: float64(4), int64(19)
     memory usage: 2.6 MB
df.shape
     (14620, 23)
```

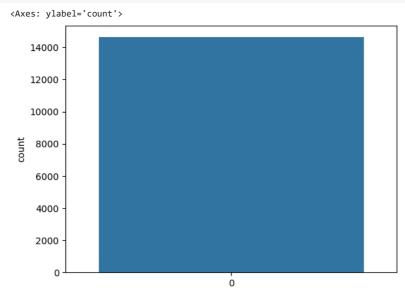
Double-click (or enter) to edit

UNIVARIATE ANAYSIS

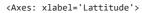
print(df.describe())

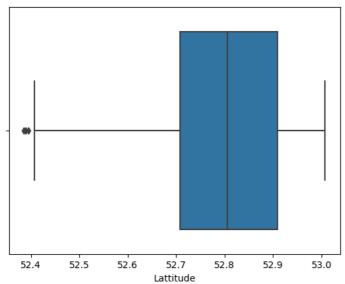
```
Date number of bedrooms number of bathrooms
                 id
count 1.462000e+04
                     14620,000000
                                                                14620.000000
                                          14620.000000
mean
       6.762821e+09
                     42604.538646
                                               3,379343
                                                                    2.129583
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
50%
       6.762821e+09
                      42600.000000
                                               3.000000
                                                                     2.250000
75%
       6.762826e+09
                     42662.000000
                                               4.000000
                                                                     2.500000
       6.762832e+09
                     42734.000000
                                              33.000000
                                                                     8.000000
max
                          lot area number of floors waterfront present
        living area
                                                             14620.000000
       14620,000000
count
                     1,462000e+04
                                        14620.000000
        2098.262996
                     1.509328e+04
                                            1.502360
                                                                 0.007661
mean
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
                                                                 0.000000
50%
        1930.000000
                     7.620000e+03
                                             1.500000
                                                                 0.000000
75%
        2570.000000
                     1.080000e+04
                                             2.000000
                                                                 0.000000
       13540.000000 1.074218e+06
                                                                 1.000000
max
                                             3.500000
       number of views condition of the house
                                                         Built Year
          14620.000000
                                   14620.000000
                                                       14620,000000
count
              0.233105
                                       3.430506
                                                        1970.926402
mean
                                       0.664151
std
              0.766259
                                                          29.493625
min
              0.000000
                                       1.000000
                                                        1900.000000
25%
              0.000000
                                       3.000000
                                                        1951.000000
                                                 . . .
50%
              0.000000
                                       3.000000
                                                        1975.000000
                                                  . . .
75%
               0.000000
                                       4.000000
                                                        1997.000000
                                                  . . .
max
              4.000000
                                        5.000000
                                                        2015.000000
       Renovation Year
                           Postal Code
                                           Lattitude
                                                          Longitude
count
          14620.000000
                          14620.000000
                                        14620.000000
                                                       14620.000000
             90.924008
                        122033.062244
                                           52.792848
                                                        -114.404007
mean
std
            416.216661
                             19.082418
                                            0.137522
                                                           0.141326
min
              0.000000
                        122003.000000
                                            52.385900
                                                        -114,709000
25%
              0.000000
                         122017.000000
                                            52.707600
                                                        -114.519000
              0.000000
50%
                         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
            14620.000000
                             14620.000000
                                                        14620.000000
count
             1996.702257
                             12753.500068
                                                            2.012244
mean
              691.093366
                             26058.414467
                                                            0.817284
std
              460,000000
                               651.000000
                                                            1,000000
min
25%
             1490.000000
                              5097.750000
                                                            1.000000
50%
             1850.000000
                              7620.000000
                                                            2.000000
75%
             2380.000000
                             10125.000000
                                                            3.000000
             6110.000000
                            560617.000000
                                                            3.000000
max
       Distance from the airport
                    14620.000000
                                   1.462000e+04
count
                        64.950958
                                   5.389322e+05
mean
                        8.936008
                                   3.675324e+05
std
                        50.000000
                                   7.800000e+04
min
25%
                        57.000000
                                   3.200000e+05
50%
                        65,000000
                                   4.500000e+05
75%
                        73.000000
                                   6.450000e+05
```

plt.hist(df['number of floors'])



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





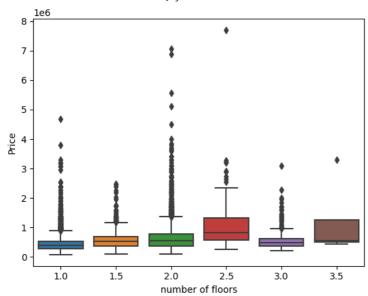
BIVARIATE ANAYSIS

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

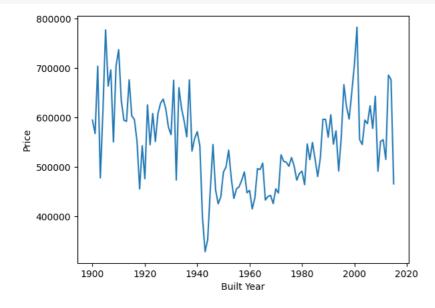
```
<Axes: xlabel='Built Year', ylabel='Price'>

1e6
8
sns.boxplot(x=df['number of floors'],y=df['Price'])
```

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



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



sns.heatmap(df[['Price','living area','lot area']].corr(),annot=True)

<Axes: >

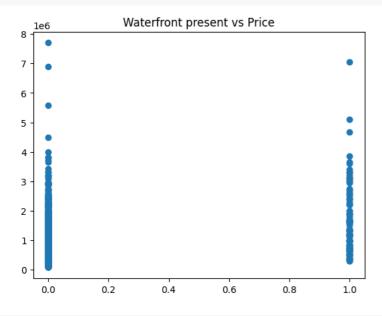
- 1.0

 $(\texttt{df}[['Price', 'number\ of\ views', 'grade\ of\ the\ house', 'condition\ of\ the\ house']])$

	Price	number of views	grade of the house	condition of the house
0	2380000	4	10	5
1	1400000	0	8	5
2	1200000	0	8	3
3	838000	0	9	3
4	805000	0	8	4
14615	221700	0	7	4
14616	219200	0	7	4
14617	209000	0	6	3
14618	205000	0	6	4
14619	146000	0	6	3

14620 rows × 4 columns

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	0.004	0.027	-0.033	-0.01€	-0.016	0.0059	0.012	0.018	-0.023	-0.018	-0.032	-5e-05	0.00410
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	0.39	0.029	0.003 4 0.
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	0.00220.0
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	0.0024).(
lot area -	0.0044	0.034	0.081	0.17	1	0.004	ທ.026	0.078	0.008	0.11	0.18	0.02	0.052	0.0068	0.07	-0.091	0.22	0.15	0.71	0.0130.0
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	0.29	-0.01-	0.00760.
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	0.086	0.032	0.00160.
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	0.0040	5-0.08	0.28	0.072	0.0080.
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	-0.1	0.004	D.006 9 0.
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	0.72	0.120	.00099.
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	0.74	0.19-	0.0029.
Area of the basement -	-0.016	0.3	0.29	0.44	0.02	-0.24	0.085	0.29	0.18	0.17	-0.04€	1	-0.14	0.075	-0.011	0.11	-0.15	0.2	0.011	0.01 0.0
Built Year -	0.005	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	0.33	0.073	0.00160
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-	0.0020	0.005€	.00083.
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	-0.11	0.077	0.011 0.
Lattitude -	-0.023	-0.013	0.031	0.055	-0.091	0.051	-0.022	0.004	0.003	0.12-	8.8e-0	50.11	-0.14	0.029	-0.31	1	-0.13	0.046	-0.092	0.0150.0
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	0.34	0.26	-0.01-0.
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	1	0.19-	0.00120.
lot_area_renov -	-5e-05	0.029	0.079	0.18	0.71	-0.01	0.032	0.072	0.004 ⁻	0.12	0.19	0.011	0.073	0.0059	0.077	-0.092	0.26	0.19	1	-0.025-0
Number of schools nearby -	0.004	D.0034	9.0022	0.0024	0.013	0.0070	3 .001€	0.008	0.00&	0009	9.0029	0.01-	0.001	8.0008	₿.011	0.015	-0.01-	0.001	0.025	1 0.
Distance from the airport -	0.011	0.0062	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	D.005	0.015	0.004

print(df.describe())

```
Date number of bedrooms number of bathrooms
count 1.462000e+04
                        14620.000000
                                               14620.000000
                                                                        14620.000000
mean
       6.762821e+09
                        42604.538646
                                                    3.379343
                                                                            2.129583
       6.237575e+03
                           67.347991
                                                    0.938719
                                                                            0.769934
std
       6.762810e+09
                        42491.000000
                                                    1.000000
                                                                            0.500000
min
                                                                            1.750000
25%
       6.762815e+09
                        42546.000000
                                                    3.000000
                                                                            2.250000
50%
       6.762821e+09 42600.000000
                                                    3,000000
75%
                                                    4.000000
                                                                            2.500000
        6.762826e+09
                        42662.000000
                                                   33.000000
                        42734.000000
                                                                            8.000000
max
        6.762832e+09
         living area
                             lot area % \left( 1\right) =\left( 1\right) \left( 1\right) =\left( 1\right) \left( 1\right) \left( 1\right)  waterfront present \left( 1\right) \left( 1\right) \left( 1\right) \left( 1\right) \left( 1\right) 
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
                                                                         0.000000
          370.000000
                        5.200000e+02
                                                  1.000000
min
25%
         1440.000000
                        5.010750e+03
                                                  1.000000
                                                                         0.000000
50%
         1930.000000
                        7.620000e+03
                                                  1.500000
                                                                         0.000000
75%
                                                                         0.000000
         2570.000000
                        1.080000e+04
                                                  2.000000
                                                  3.500000
        13540.000000 1.074218e+06
                                                                         1.000000
max
        number of views
                           condition of the house \\
                                                               Built Year
count
           14620.000000
                                       14620.000000
                                                             14620.000000
```

```
1970.926402
                   0.233105
                                            3,430506
    mean
                   0.766259
                                                               29,493625
     std
                                            0.664151
                                                             1900.000000
     min
                   0.000000
                                            1.000000
     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
               14620.000000
                               14620.000000
                                             14620.000000
                                                            14620.000000
     count
                  90.924008
                             122033,062244
                                                52,792848
                                                             -114,404007
     mean
                 416.216661
                                  19.082418
                                                                0.141326
     std
                                                 0.137522
                   0.000000
                             122003.000000
                                                             -114,709000
     min
                                                 52.385900
     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
                 14620.000000
                                  14620.000000
                                                             14620.000000
     count
                  1996.702257
                                  12753.500068
                                                                 2.012244
    mean
     std
                   691.093366
                                  26058.414467
                                                                 0.817284
                   460,000000
                                    651.000000
                                                                 1,000000
     min
                  1490.000000
                                                                 1.000000
     25%
                                   5097.750000
     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
                             8.936008
                                        3.675324e+05
    std
     min
                             50.000000
                                        7.800000e+04
     25%
                             57.000000
                                        3.200000e+05
                             65.000000
                                        4.5000000+05
     50%
     75%
                             73.000000
                                        6.450000e+05
print(df.count())
     id
                                               14620
     Date
                                                14620
     number of bedrooms
                                               14620
                                                14620
     number of bathrooms
     living area
                                                14620
                                                14620
     lot area
     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
                                               14620
     Renovation Year
     Postal Code
                                                14620
                                               14620
     Lattitude
     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())
                                                    id
                                                            Date number of bedrooms
                                             1,000000
                                                        0.045966
                                                                            -0.329034
     id
                                             0.045966
                                                        1.000000
                                                                            -0.015663
    Date
                                                                             1.000000
     number of bedrooms
                                             -0.329034 -0.015663
     number of bathrooms
                                             -0.516909 -0.026485
                                                                             0.509784
     living area
                                             -0.648127 -0.021958
                                                                             0.570526
     lot area
                                             -0.100269 0.004392
                                                                             0.034416
     number of floors
                                                                             0.177294
                                             -0.312305 -0.010335
     waterfront present
                                             -0.112937
                                                        0.012006
                                                                            -0.006257
     number of views
                                             -0.293004 -0.004782
                                                                             0.078665
                                             -0.045061 -0.027402
                                                                             0.026597
     condition of the house
     grade of the house
                                             -0.673448 -0.033097
                                                                             0.352945
     Area of the house(excluding basement) -0.565116 -0.015994
                                                                             0.473599
     Area of the basement
                                             -0.290806 -0.015711
                                                                             0.300332
     Built Year
                                             -0.068645 -0.005869
                                                                             0.152954
     Renovation Year
                                             -0.109155 -0.011636
                                                                             0.016132
     Postal Code
                                             0.294709 0.018243
                                                                            -0.044156
     Lattitude
                                             -0.479334 -0.023327
                                                                            -0.013163
     Longitude
                                             -0.070841 -0.018231
                                                                             0.135712
     living_area_renov
                                             -0.599900 -0.032495
                                                                             0.389855
     lot_area_renov
                                             -0.089604 -0.000050
```

```
-0.004821 -0.004071
                                                                           0.003397
    Number of schools nearby
     Distance from the airport
                                            -0.004542 0.011457
                                                                           -0.006157
                                            -0.773114 -0.027919
                                                                           0.308460
    Price
                                             number of bathrooms living area
     id
                                                       -0.516909
                                                                    -0.648127
    Date
                                                       -0.026485
                                                                    -0.021958
     number of bedrooms
                                                        0.509784
                                                                     0.570526
     number of bathrooms
                                                        1.000000
                                                                     0.753517
     living area
                                                        0.753517
                                                                     1.000000
                                                        0.080806
                                                                     0.174420
     lot area
     number of floors
                                                        0.502924
                                                                     0.354743
                                                        0.060104
                                                                     0.105837
     waterfront present
     number of views
                                                        0.183789
                                                                     0.287728
     condition of the house
                                                       -0.128232
                                                                     -0.063358
     grade of the house
                                                        0.663054
                                                                     0.761835
     Area of the house(excluding basement)
                                                        0.684391
                                                                     0.875793
     Area of the basement
                                                        0.287190
                                                                     0.441491
     Built Year
                                                                     0.309602
                                                        0.498127
     Renovation Year
                                                        0.049669
                                                                     0.059400
     Postal Code
                                                       -0.105546
                                                                     -0.080303
     Lattitude
                                                                     0.054518
                                                        0.031156
     Longitude
                                                        0.223904
                                                                     0.240208
     living_area_renov
                                                        0.570530
                                                                     0.757571
     lot_area_renov
                                                        0.078627
                                                                     0.180312
     Number of schools nearby
                                                        0.002180
                                                                     0.002370
    Distance from the airport
                                                        0.009206
                                                                     0.002511
     Price
                                                        0.531735
                                                                     0.712169
                                            lot area number of floors \
     id
                                            -0.100269
                                                              -0.312305
    Date
                                             0.004392
                                                              -0.010335
     number of bedrooms
                                             0.034416
                                                               0.177294
     number of bathrooms
                                             0.080806
                                                               0.502924
                                             0.174420
     living area
                                                               0.354743
     lot area
                                             1.000000
                                                              -0.004138
print(df['number of floors'].value_counts())
```

```
1.0
        7103
        5666
2.0
        1311
1.5
3.0
         418
2.5
         118
3.5
           4
```

Name: number of floors, 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

HANDLE THE MISSED VALUES

```
print(df.isnull().sum())
```

```
Date
                                          0
number of bedrooms
                                          0
number of bathrooms
                                          0
living area
                                          0
lot area
                                          0
number of floors
                                          0
waterfront present
                                          a
number of views
                                          0
condition of the house
                                          0
grade of the house
                                          0
Area of the house(excluding basement)
Area of the basement
Built Year
                                          0
Renovation Year
                                          0
Postal Code
                                          0
Lattitude
                                          0
Longitude
                                          a
living_area_renov
                                          0
lot_area_renov
                                          0
Number of schools nearby
                                          0
Distance from the airport
                                          0
                                          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=df.drop(['Price','Date'],axis=1)
x.set_index(['id'],inplace=True)
y=df[['id','Price']]
x.head()
```

ic	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
6762810145	5 5	2.50	3650	9050	2.0	0	4	5	10	3370	280	1921	
6762810635	5 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	5 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
```

805000

4 6762812919

```
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)
```

```
▼ GradientBoostingRegressor

GradientBoostingRegressor(max_depth=5, n_estimators=400)
```

```
y_pred = model.predict(x_test)
model.score(x_test,y_test)
```

0.9127724700080724

```
r2_score(y_pred,y_test)
```

0.9025621094930966

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
```

	ID	Predicted Price
0	6762811233	497766.13
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 rows × 2 columns