import pandas as pd
import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

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.csv

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	со
0	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 rc	ows × 23 colum	ins								

df.tail()

		id	Date	number of bedrooms	number of bathrooms	living area	lot area	number of floors	waterfront present	number of views	
146	315	6762830250	42734	2	1.5	1556	20000	1.0	0	0	
146	616	6762830339	42734	3	2.0	1680	7000	1.5	0	0	
146	317	6762830618	42734	2	1.0	1070	6120	1.0	0	0	
146	618	6762830709	42734	4	1.0	1030	6621	1.0	0	0	
146	319	6762831463	42734	3	1.0	900	4770	1.0	0	0	
5 rov	5 rows × 23 columns										

df

```
number
                                                                     number
                                                                                         number
                                          number of living
                                                               lot
                                                                            waterfront
                                      of
                                                                        of
                                          bathrooms
                                                        area
                                                               area
                                                                                present
                                bedrooms
                                                                     floors
                                                                                          views
             6762810145 42491
                                       5
                                                2.50
                                                        3650
                                                               9050
                                                                        2.0
                                                                                      0
                                                                                              4
df.columns
     Index(['id', 'Date', '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', '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
                                                 int64
     living area
     lot area
                                                 int64
     number of floors
                                               float64
     waterfront present
                                                 int64
     number of views
                                                 int64
     condition of the house
                                                 int64
                                                 int64
     grade of the house
                                                 int64
     Area of the house(excluding basement)
     Area of the basement
                                                 int64
     Built Year
                                                 int64
     Renovation Year
                                                 int64
     Postal Code
                                                 int64
     Lattitude
                                               float64
     Longitude
                                               float64
                                                 int64
     living_area_renov
                                                 int64
     lot_area_renov
     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
          number of bedrooms
                                                  14620 non-null
                                                                  int64
      2
      3
          number of bathrooms
                                                  14620 non-null
                                                                  float64
          living area
                                                  14620 non-null
      4
          lot area
                                                  14620 non-null
                                                                   int64
          number of floors
                                                  14620 non-null
                                                                  float64
      6
          waterfront present
                                                  14620 non-null int64
                                                  14620 non-null
          number of views
                                                  14620 non-null
          condition of the house
                                                                  int64
      10 grade of the house
                                                  14620 non-null int64
          Area of the house(excluding basement) 14620 non-null
      12 Area of the basement
                                                  14620 non-null
                                                                  int64
          Built Year
                                                  14620 non-null
      13
                                                                  int64
      14
          Renovation Year
                                                  14620 non-null
                                                                  int64
      15 Postal Code
                                                  14620 non-null
                                                                  int64
                                                  14620 non-null
                                                                  float64
      16 Lattitude
      17 Longitude
                                                  14620 non-null
                                                                  float64
                                                  14620 non-null
      18 living_area_renov
      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
          Price
                                                  14620 non-null int64
     dtypes: float64(4), int64(19)
     memory usage: 2.6 MB
```

df.shape

print(df.describe())

```
Date number of bedrooms number of bathrooms
                 id
count 1.462000e+04 14620.000000
                                         14620.000000
                                                               14620.000000
       6.762821e+09 42604.538646
                                             3.379343
mean
                                                                   2.129583
      6.237575e+03
                        67.347991
                                             0.938719
                                                                  0.769934
std
       6.762810e+09 42491.000000
                                             1.000000
                                                                  0.500000
min
25%
       6.762815e+09 42546.000000
                                             3.000000
                                                                   1.750000
50%
       6.762821e+09
                    42600.000000
                                             3.000000
                                                                   2.250000
75%
       6.762826e+09 42662.000000
                                             4.999999
                                                                   2.500000
       6.762832e+09
                    42734.000000
                                            33.000000
                                                                   8.000000
max
                         lot area number of floors waterfront present \
        living area
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
                                           1.000000
                                                                9.999999
         370.000000
                    5.200000e+02
min
25%
        1440.000000
                     5.010750e+03
                                           1.000000
                                                                0.000000
50%
        1930.000000
                    7.620000e+03
                                           1.500000
                                                                0.000000
        2570.000000
                    1.080000e+04
                                           2.000000
                                                                0.000000
75%
max
       13540.000000 1.074218e+06
                                           3.500000
                                                                1.000000
       number of views condition of the house ...
                                                       Built Year
                                                     14620.000000
count
          14620.000000
                                  14620.000000
mean
              0.233105
                                      3.430506
                                                      1970.926402
              0.766259
                                      0.664151
                                                        29.493625
std
                                                . . .
              0.000000
                                      1.000000
                                                       1900.000000
min
25%
              0.000000
                                      3.000000
                                                       1951.000000
50%
              0.000000
                                      3.000000
                                                       1975.000000
                                                . . .
              0.000000
                                      4.000000
                                                      1997.000000
75%
                                                . . .
                                                      2015.000000
              4.000000
                                      5.000000
max
       Renovation Year
                          Postal Code
                                          Lattitude
                                                        Longitude \
          14620.000000
                        14620.000000 14620.000000
                                                     14620,000000
count
mean
             90.924008
                        122033.062244
                                          52.792848
                                                      -114.404007
                            19.082418
std
            416.216661
                                           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
max
           2015.000000 122072.000000
                                          53.007600
                                                      -113.505000
       living_area_renov lot_area_renov
                                          Number of schools nearby
                                                      14620.000000
count
            14620.000000
                            14620.000000
mean
             1996.702257
                            12753.500068
                                                           2.012244
std
              691.093366
                            26058.414467
                                                           0.817284
              460.000000
                              651.000000
                                                          1.000000
min
             1490.000000
25%
                             5097.750000
                                                          1.000000
50%
             1850.000000
                             7620.000000
                                                           2.000000
75%
             2380.000000
                            10125.000000
                                                           3.000000
max
             6110.000000
                           560617,000000
                                                           3.000000
       Distance from the airport
                    14620.000000 1.462000e+04
count
mean
                       64.950958 5.389322e+05
                        8.936008 3.675324e+05
std
min
                       50.000000
                                 7.800000e+04
                       57.000000 3.200000e+05
25%
50%
                       65.000000 4.500000e+05
75%
                       73.000000 6.450000e+05
```

plt.hist(df['lot area'])

10/17/23, 11:12 AM - Colaboratory

## Assignment-3

```
array([5.200000e+02, 1.078898e+05, 2.152596e+05, 3.226294e+05,
           4.299992e+05, 5.373690e+05, 6.447388e+05, 7.521086e+05, 8.594784e+05, 9.668482e+05, 1.074218e+06]),
     <BarContainer object of 10 artists>)
     14000
     12000
     10000
sns.countplot(df
               'number of bedrooms'])
        14000
        12000
        10000
         8000
         6000
         4000
         2000
            0
                                           Ó
```

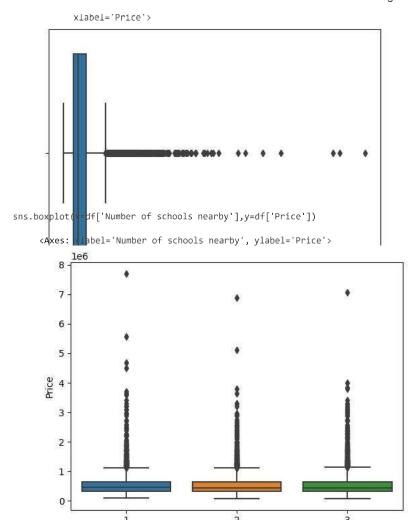
sns.countplot(df['number of floors'])

<Axes: ylabel='count'>

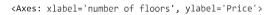
sns.boxplot(x=df['Price'])

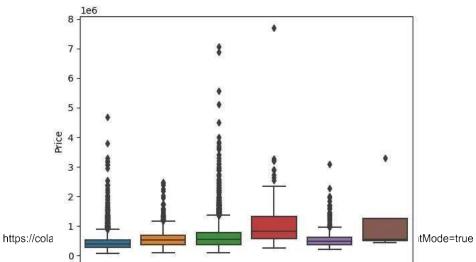
<Axes:

## Assignment-3



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

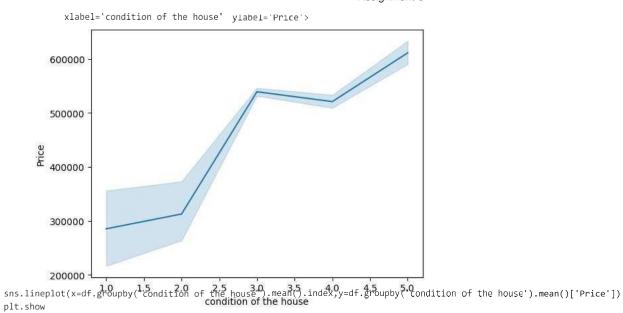




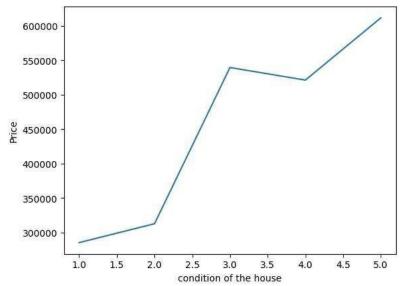
Number of schools nearby

<Axes:

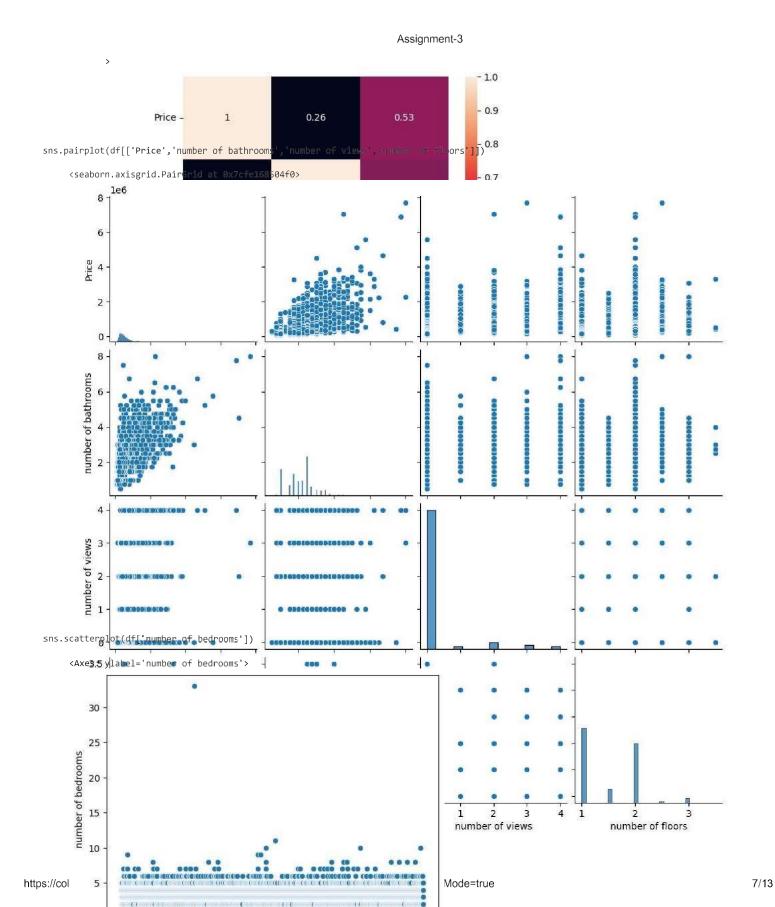


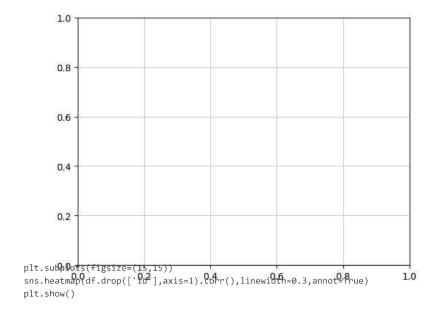


<function matplotlib.pyplot.show(close=None, block=None)>



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





print(df.describe())

	id	Date	number of bedrooms	s number of bathrooms	. \
count	1.462000e+04	14620.000000	14620.000000	14620.000000	
mean	6.762821e+09	42604.538646	3.379343	3 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	3.00000	1.750000	
50%	6.762821e+09	42600.000000	3.00000	2.250000	
75%	6.762826e+09	42662.000000	4.000000	2.500000	
max	6.762832e+09	42734.000000	33.000000	8.000000	
	living area	lot area	number of floors	waterfront present \	
count	14620.000000	1.462000e+04	14620.000000	14620.000000	
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	

1038, 000000 7, 000000 100000000000000000000	25%		5.010750e+03	1.0	000000	,		0.0	00000												_
Table 1.620. Beases 1.67218-e6 3				1	288888			0.0	88888						es vientines						
Tunber of Apphaer of bedress 1420, 600-603 1				- 1 <sup>2</sup> 3	.0150	.026	0.022	0.0049	-0.01	0.012	0.0048	0.027	0.033	-0.016	-0.016	0.005	0.012	0.018	-0.023	0.018	-(
Count   146/20,000000	max	13540.000000	1.0/4218e+06	3.	288888			1.0	pease												
Count   146/20,000000		numbon of vito	us sandition of t	ha haus			2011	Mana	1												
0.233185   0.7467764 of bathrooms									0.18-	0.006	0.079	0.027	0.35	0.47	0.3	0.15	0.016	-0.044	-0.013	0.14	
State   Stat				8.88888	0				,								_ YVV	8	3		
### 0				5,43050	6		3/0.9.	26402				1							1		
25% 0.088889   100									0.5	0.06	0.18	-0.13	0.66	0.68	0.29	0.5	0.05	-0.11	0.031	0.22	
Sex	min	0.0000	00	1.88888	0	19	900.00	90000										III N	1000		
Renovation vear routal accessors and accessors and accessors and accessors accessors and accessors accesso	25%	0.0000	00	1.8888	0	19	951.00	30000													
Renovation vear routal accessors and accessors and accessors and accessors accessors and accessors accesso	50%	0.0000	00 living area	0 0220	0 - 7 . 6	7519	975.00	00000	0.35	011	0.29	0.063	0.76	0.88	0.44	0.31	0.059	-0.08	0.055	0.24	5
Renovation Year   1628 area - No.4	75%	0.0000	00	4 2 2 2 2 2 2		19	997.00	00000	9.55	100000	3.43	0.000	17.5.5.T	0.00	5000000	B. 7.7.	0.000		0.055	A	
Renovation Year   1628 area - No.4	max	4.0000	99	5 00000		26			0. 0				9	- 20							
Court   14626, 088808   1462		1.0000			3333					200.0	2 222		COLUMN TO SERVICE STATE OF THE	12.11.2	2000		V 200	100	202022	20,000	
Court   14626, 088808   1462		Ponovation Vo	lot area	0.00440	.034 0.	.081	0.17	itudo	0.004	0.026	0.078	0.008	0.11	0.18	0.02	0.052	0.0068	0.07	-0.091	0.22	
mean   90.924088   min   10.8241   min   10.82408   min   10.80608   min	count			14000					, \											, ,	
10				14520.	200000	140	520.00	90000													
10	mean	90.9240	osnumber of floors	-0.01 (	J. 18 48	0.5	d 354°	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	
0.000 0.000	std	416.2166	61 19.082418		137522		0.14	11326													
### 12.83   ### 12	min	0.0000	00 122003.000000	52 F	385900		114.7	39000						-				- 4	-		
### 12.83   ### 12	25%	0.0000	00, 122017,000000	0.0125	00600	106	14.5	19000	0.016	7	0.4	0.010	0.00	0.072	0.005	0.024	0.006	0.020	0.022	0.040	6
## 1275%   0.0000000   127048.0000000   0.00000000000000000000000000	50%	0.0000	00 122032.000000	0.0120.	1.400	.00	114.4	21000	0.016	+	0.4	0.019	0.08	0.072	0.065	0.024	0.080	0.036	0.022	0.040	ř
Titing area penor lot_area_renor lot					-		114-3	15000										- 8	0		
Count 14662 600120 or of 150 0050 0.027 0.027 0.13 0.06 0.0088 0.27 0.019 0.053 1 0.15 0.17 0.18 0.38 0.0620.045 0.003 0.12 mean 1996.7022 12753.0506 0.027 0.027 0.018 0.058 0.27 0.019 0.053 1 0.15 0.17 0.18 0.38 0.0620.045 0.003 0.12 0.14 0.019 0.15 0.15 0.15 0.12 0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.				E 3	007500		112.50	85000	201000000	20.000				*****					I		
count 14620_89389 of 14620_89389 of 14620_89389 of 12620_89389 of 12620_89380 of	IIIdA	2013.0000	number of views	0.00400	.079 u	.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	6-0.08	
count 14620_89389 of 14620_89389 of 14620_89389 of 12620_89389 of 12620_89380 of			7 .	-0.000			-	Maria Maria	100	100						10000		W 1		77	
std 691,99366 z698,41444 d8. 269869 of the 51,99366 z038 z038 z038 z038 z038 z038 z038 z038					per et	Sch	ools (	earby										- 1			
std 691,99366 z698,41444 d8. 269869 of the 51,99366 z038 z038 z038 z038 z038 z038 z038 z038	count	14620.00	0000 14620.0000	0 0270	027 -0	13	0.063	0.008	6-0 27	0 019	0.053	1	-0.15	-0 17	018	-0.38	0.062	0.045	0 003	-0.12	
### ### ### ### ### ### ### ### ### ##	mean	1996.70	2257 12753.5000	3.3270	UL /	45	0.002	512244	0.21	0.013	0.033	-	0.10	0.1/	0.10	0.50	0.002	ر	0.003	0.12	
### ### ### ### ### ### ### ### ### ##	std	691.09	3366 26058.4144	67			0.	<del>317284</del>											100		
1856, 086000 7620,00000 7620,00000 7620,00000 0.47 0.68 0.89 0.18 0.53 0.072 0.16 0.17 0.76 1 0.046 0.42 0.026 0.084 8.60 0.35 0.35 0.35 0.35 0.35 0.35 0.35 0.3		160 00	9999 651 9999		and the same of		1.6	agagag	N - 2.00	188718660	O DESCRIPTION OF THE PARTY OF T	00 (SA) (SA)			VARABLES	10 200		//santra	800000000000000000000000000000000000000	114000007	
1856, 086000 7620,00000 7620,00000 7620,00000 0.47 0.68 0.89 0.18 0.53 0.072 0.16 0.17 0.76 1 0.046 0.42 0.026 0.084 8.60 0.35 0.35 0.35 0.35 0.35 0.35 0.35 0.3		1400.00	grade of the house	0.033	0.35	.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
Take of the house from the airport of bedrooms number of bedrooms Longitude -0.01c 0.14 0.22 0.24 0.22 0.13 0.04c 0.08 0.29 0.06 0.08 0.05 0.15 0.1 0.01 0.02 0.09 0.06 0.08 0.10 0.3 0.09 0.06 0.08 0.10 0.10 0.02 0.00 0.05 0.01 0.10 0.01 0.01 0.01 0.01							1.	2000000													
Distance from the airport ment of 0.01e 0.3 0.29 0.44 0.02 0.24 0.085 0.29 0.18 0.17 0.04e 1 0.14 0.075 0.011 0.11 0.15 mean 64.050958 5.38							2.	900000													
Distance from the airport ment of 0.01e 0.3 0.29 0.44 0.02 0.24 0.085 0.29 0.18 0.17 0.04e 1 0.14 0.075 0.011 0.11 0.15 mean 64.050958 5.38	75%rea	of the houselex	0000 10125.0000	0.016	3 47 6	68	0.83	200000	0.53	0.072	016	-0 17	0.76	1	0.046	0.42	0.026	0.084	8e-0	50.35	8
mean std 8.936988 5.38 5.00 50 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 min 50.000000 7.860 0.055 0.05 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 0.41 0.41 0.41 0.41 0.41 0.41	max	6110.00	0000 560617.0000	.00			0.00	300000	Market St.	0.072	0.10	9.47	4.70	-	0.0-0	100	0.020	0.00-	,.oc o		8
mean std 8.936988 5.38 5.00 50 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 min 50.000000 7.860 0.055 0.05 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 0.41 0.41 0.41 0.41 0.41 0.41																		- 8			
mean std 8.936988 5.38 5.00 50 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 min 50.000000 7.860 0.055 0.05 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 0.41 0.41 0.41 0.41 0.41 0.41		Distance from	the airport	acaRnic	e. s. s		22,152	7000000	22/12/02	Anna da Gran	-238/8/1	20232	2007-01-000	22012000	2	12/12/20	Mary Mary	Maria de Maria	Bridges)	8 E W C 12	
mean std 8.936988 5.38 5.00 50 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 min 50.000000 7.860 0.055 0.05 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 0.41 0.41 0.41 0.41 0.41 0.41	count	Are	1462 The pasement	-0.016	9.3	1.29	0.44	0.02	-0.24	0.085	0.29	0.18	0.17	-0.046	1	-0.14	0.075	-0.011	0.11	-0.15	
std min 50.000000 79.56.0 005 0.15 0.5 0.31 0.052 0.48 0.024 0.055 0.36 0.44 0.42 0.14 1 0.23 0.062 0.14 0.41 0.42 0.14 1 0.23 0.062 0.14 0.41 0.42 0.14 0.42 0.14 1 0.23 0.062 0.14 0.41 0.45 0.062 0.065 0																					
min 50, 008-001 50, 105 0.15 0.5 0.31 0.052 0.48 0.024 0.053 0.38 0.44 0.42 0.14 1 0.23 0.062 0.14 0.41 1 0.23 0.062 0.14 0.41 1 0.23 0.062 0.14 0.41 1 0.23 0.062 0.14 0.41 1 0.23 0.062 0.14 0.41 1 0.23 0.062 0.15 0.026 0.075 0.23 1 0.018 0.029 0.080 0.065 0.065 0.065 0.086 0.1 0.062 0.015 0.026 0.075 0.23 1 0.018 0.029 0.080 0.			0 036000 3 63		e							- 1						)	4		
25% 57.000000 3.20% 55.000000 4.5.5.00.012.0.016 0.05 0.059.0.065.006 0.086 0.1 0.062.0.015 0.026 0.075 -0.23 1 0.018 0.029 0.080 75% 65.0000000 4.5.5.00.012.0.016 0.086 0.1 0.059.0.065.0.006 0.086 0.1 0.062.0.015 0.026 0.075 -0.23 1 0.018 0.029 0.080			Built Year	0.00590	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	
## 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 0.018 0.029 0.080 print(df.count())  ### Destal 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 - 0.001			50.000000 7.80	angage+ di	Alternation		100		tention to the	and the same of	STATISTICS.		) <del>-1</del> 000	and the second	Monte		***************************************	one on the	- Accessing	satism de	
Postal Code - 0.018 0.044 - 0.11	25%		57.000000 3.20	9999e-4			- 5					ti i	2 2					- 80	7		
Postal Code -0.018 0.044 0.11 0.08 0.07 0.13 0.038 0.039 0.045 0.15 0.08 0.011 0.062 0.018 1 0.31 0.099 print(df.count())  id Lattitude -0.02 99.013 0.031 0.055 0.091 0.051 0.022 0.004 0.003 0.12 8.8e-050.11 -0.14 0.029 0.31 1 -0.13 0 0.008 0.009	50%		65.000000 4.50	0.0130	016 0	OF	0.050	0066	006	0.006	0.1	0.063	0015	0.036	0.075	0.22	4	0.010	0.000	0.00	5
id Lattitude -0.023 9.013 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 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	75%		75.000000 6.45	WY14V	010	.05 K	ม.นอย	J.0066	J.UU6 .	0.066	0.1	0.062	0.015	0.026	0.075	-0.23	7	0.016	0.029	-0.06	4
id Lattitude -0.023 9.013 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 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0									_					- 4	_		- 5		6		
id Lattitude -0.023 9.013 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 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0																					
id			Postal Code -	0.018-0	0.044 -0	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	
Date number of bedrooms number of bedrooms number of bathrooms Longitude -0.016 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 -0.095 -0.13 1  Iving area lot area number of floors   Iving 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 0.11 0.046 0.34	print(df.co	unt())					1000	W-	10.00	-	CONT.	100000	1000	000-00	0 000					100	
Date number of bedrooms number of bedrooms number of bathrooms Longitude -0.016 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 -0.095 -0.13 1  Iving area lot area number of floors   Iving 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 0.11 0.046 0.34							- (					1				7		- "			
Date number of bedrooms number of bedrooms number of bathrooms Longitude -0.016 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 -0.095 -0.13 1  Iving area lot area number of floors   Iving 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 0.11 0.046 0.34	id		Lattitude -	0.0230	0130	031	0.055	0 091	0.051	0 023	0046	0003	0.12-	8 8 e - 0	50 11	-0 14	0 029	-0.31	1	-0.13	0
number of bedrooms number of bathrooms Longitude -0.016 0.14 0.22 0.24 0.22 0.13 0.048 0.08 0.12 0.2 0.35 0.15 0.41 0.08 0.099 0.13 1  living area lot area number of floors waterfront present number of views condition of the house 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 0.005 0.077 0.092 0.26 0.1  Area of the batermater of schools nearby -0.0042.003 0.002 0.002 0.001 0.008 0.008 0.009 0.002 0.001 0.001 0.008 0.011 0.015 -0.01 0.008 0.011 0.015 -0.010 0.008 0.008 0.009 0.002 0.004 0.001 0.008 0.012 0.004 0.001 0.008 0.012 0.004 0.001 0.008 0.012 0.004 0.008 0.009 0.002 0.004 0.001 0.008 0.012 0.004 0.008 0.012 0.004 0.013 0.012 0.013 0.012 0.013 0.014 0.013 0.014 0.013 0.014			Lattitude	0.0233	V-1-1	031	0.055	0.031	0.031	0.022		0.005	0.12	۰.٥٥ ٥	JU. 11	0.11	0.023	0.51	-	0.13	Ĭ
Number of bathrooms   Longitude - 0.015 0.14   0.22   0.24   0.22   0.13   0.045 0.08   0.12   0.2   0.35   0.15   0.41   0.08   0.095   0.13   1		of hadrooms		1402		$\rightarrow$	- 6					ŭ						- %	(/-		
living area lot area number of floors 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 0.11 0.046 0.34 waterfront present number of views condition of the house lot_area_renov -5e-050.029 0.079 0.18 0.71 -0.01 0.032 0.072 0.004 0.12 0.19 0.011 0.073 0.00590.077 0.092 0.26 0.1 grade of the house (excluding basement) Area of the house(excluding basement) Area of the basember of schools nearby -0.0040.002 0.002 0.002 0.001 0.0080 0.0630009 0.002 0.01 0.0010 0.008 0.010 0.008 0.008 0.009 0.002 0.01 0.0010 0.008 0.010 0.008 0.008 0.010 0.015 0.			and constraints and the	1452	A cities W			WWW.SELWAY	No. of the last	68/7/2008	(1) 44 (1) 14 (1)	(24) (25) (20)	1,600,000	Constitute of	186-1172-1172-1172-1172-1172-1172-1172-117	W - 100	TORK VALUE OF	1440 1900 1900	200 100 000	200	
lot area number of floors   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 0.11   0.046   0.34   0.0016			Longitude -	-0.0184	3.14 C	).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	
number of floors waterfront present number of views condition of the house lot_area_renov -5e-059.029 0.079 0.18 0.71 -0.01 0.032 0.072 0.004 0.12 0.19 0.011 0.073 0.00590.077 -0.092 0.26 0.1	living	area		1462	0																
waterfront present number of views condition of the house lot_area_renov -5e-650.029 0.079 0.18 0.71 -0.01 0.032 0.0720.004 0.12 0.19 0.011 0.0730.00590.077 -0.092 0.26 0.1 grade of the house (excluding basement) Area of the house(excluding basement) Area of the basement of schools nearby -0.0040.0030.0020.0020.0024 0.0130.0070.00160.0080.006900099.0029 0.01-0.0016.0008B.011 0.015 -0.01-0.00 Built Year Renovation Year Postal Code Distance from the airport -0.0113.00600.0090.00250.00330.0170.001-0.0010.0020.0049.00120.0029-0.0040.00530.0120.00720.003-0.00 Lattitude Longitude living_area_renov Number of schools nearby Distance from the airport 9146200	lot ar	'ea		1462	0																
waterfront present number of views condition of the house lot_area_renov -5e-650.029 0.079 0.18 0.71 -0.01 0.032 0.0720.004 0.12 0.19 0.011 0.0730.00590.077 -0.092 0.26 0.1 grade of the house (excluding basement) Area of the house(excluding basement) Area of the basement of schools nearby -0.0040.0030.0020.0020.0024 0.0130.0070.00160.0080.006900099.0029 0.01-0.0016.0008B.011 0.015 -0.01-0.00 Built Year Renovation Year Postal Code Distance from the airport -0.0113.00600.0090.00250.00330.0170.001-0.0010.0020.0049.00120.0029-0.0040.00530.0120.00720.003-0.00 Lattitude Longitude living_area_renov Number of schools nearby Distance from the airport 9146200	number	of floors	living area renov -	-0.03252	039 0	57	0.76	0.15	0.29	0.086	0.28	-0.1	0.72	0.74	0.2	0.33-	0.002	6-0.11	0.046	0.34	
number of views condition of the house lot_area_renov -5e-050.029 0.079 0.18 0.71 -0.01 0.032 0.0720.00470.12 0.19 0.011 0.073 0.00590.077 0.092 0.26 0.1  grade of the house Area of the house(excluding basement) Area of the base(excluding basement) Area of the base(excluding basement) Built Year Renovation Year Postal Code Distance from the airport -0.0112.0062.0093 0.0025 0.003 0.017 0.0014 0.0012 0.002 0.0049 0.012 0.0029 0.004 0.0053 0.012 0.0072 0.003 0.0012 0.00	waterf	ront present		1462	0	Distance of the last		1. State of the	100000000	CONS.5788	N. Laborator O	(10,000			77.CT	Call Control				SOUTH	7:
condition of the house lot_area_renov -5e-030.029 0.079 0.18 0.71 -0.01 0.032 0.072 0.004 0.12 0.19 0.011 0.073 0.00590.077 -0.092 0.26 0.1 grade of the house Area of the house (excluding basement) Area of the batember of schools nearby 0.004 0.003 0.002 0.002 0.013 0.007 0.0016 0.008 0.0069 0.009 0.002 0.01 -0.0016 0.008 0.011 0.015 -0.01 0.008 0.011 0.015 -0.01 0.008 0.011 0.015 -0.01 0.008 0.009 0.009 0.009 0.002 0.01 -0.0016 0.008 0.009 0.009 0.009 0.002 0.01 -0.0016 0.008 0.011 0.015 -0.01 0.008 0.011 0.015 -0.01 0.008 0.012 0.012 0.013 0.012 0.008 0.008 0.00				1.46.2					) 2. 3	4								- 6	5		
grade of the house Area of the house(excluding basement) Area of the battember of schools nearby Built Year Renovation Year Postal Code Distance from the airport - 0.0113.0060.0090.00250.00330.0170.00140.00150.0020.0049.00120.00290.0049.00120.00290.0049.00120.00290.00490.00120.00290.00290.00290.00290.00490.00120.0029			- SWITE COMPANY	MERCANIA P	2000	Name of	1000		227202		NAME OF TAXABLE PARTY.		W. SEW	100220	SECTION SECTION	COLUMN A	C ROSE	ALC: NO.		20.20	0.8243
Area of the house(excluding basement) Area of the batember of schools nearby Built Year Renovation Year Postal Code Lattitude Longitude living_area_renov Number of schools nearby Distance from the airport Distance from the air	Condit	.c.+b- b	se lot_area_renov -	-se-usu	.029 0.	0/9	0.18	0.71	-0.01	0.032	0.072	0.004	0.12	0.19	0.011	0.073	ນ.ບບວາ	0.077	0.092	0.∠0	0.13
Area of the balanmeer of schools nearby 0.0042.90340.00220.00240.0130.0070.00160.00820.00840.00940.00290.0140.00088.011 0.015 -0.0140.00880.00840.00940.00140.0018				1462	0																
Built Year Renovation Year Postal Code		*	,	1452	0																
Built Year Renovation Year Postal Code			r of schools nearby	0.0042.	0034).0	0022	.0024	0.013	0.007	0.0016	0.008	0.0063	0009	9.002	0.01-	0.001	.0008	B.011	0.015	-0.01-	0.00
Postal Code Lattitude Longitude Longitude living_area_renov Number of schools nearby Distance from the airport 1446201 Stance from the airport 1446200 Stance	Built	Year		1462	8			1000			War and			4-14					NAME OF		
Postal Code Lattitude Longitude Longitude living_area_renov Number of schools nearby Distance from the airport 1446201 Stance from the airport 1446201 Stance from the airport 1446201 Stance from the airport Price dtype: int64  Distance from the airport -0.0112.006 0.009 0.002 0.003 0.017 0.001 0.001 0.001 0.002 0.004 0.002	Renova	tion Year		1482	0							n :		-			- 4	- 4	,		
Lattitude Longitude living_area_renov lot_area_renov Number of schools nearby Distance from the airport Price dtype: int64  Longitude  13:30  Price -0.026 6.31 0.53 0.71 0.082 0.26 0.26 0.4 0.041 0.67 0.62 0.33 0.05 0.13 0.12 0.3 0.024 0.5  146201  Swell title to the price of schools nearby Distance from the airport Price dtype: int64			ce from the airport	0.0112	0067	000	0025	003	0.017	001	0.001	0.003	0.046	0.001	າ ດດວດ	0.004	005	0.013	0077	0.03	0.00
Longitude living_area_renov lot_area_renov Number of schools nearby Distance from the airport Price dtype: int64  Price -0.028 0.31 0.53 0.71 0.082 0.26 0.26 0.4 0.041 0.67 0.62 0.33 0.05 0.13 -0.12 0.3 0.024 0.5  0.028 0.31 0.53 0.71 0.082 0.26 0.26 0.4 0.041 0.67 0.62 0.33 0.05 0.13 -0.12 0.3 0.024 0.5		Distail	ce nom the airport	0.0110.	0000	,052	.0021	0.0033	0.017	5.001*	3.001	J.002	5.004	5.0012	J.0025	0.004	3.003	0.012	J.0072	3.003	5.00
living_area_renov   Price - 0.028 0.31 0.53 0.71 0.082 0.26 0.26 0.4 0.041 0.67 0.62 0.33 0.05 0.13 -0.12 0.3 0.024 0.5				1-4-2	0											ě .		- X	()		
lot_area_renov  Number of schools nearby Distance from the airport Price dtype: int64  Distance from the airport Distance	_		1 marine 100 marine	404				Territoria de la constanta de													
Number of schools nearby  Distance from the airport  Price  Ot views a sement of floors area of the policy in the policy of the			Price -	-0.028	9.31 0	.53	0.71	0.082	0.26	0.26	0.4	0.041		0.62	0.33	0.05	0.13	-0.12	0.3	0.024	0.5
Distance from the airport  And the house seminaria as sement, he house of views area as a sement, stal Code dear a sement and state of floors.  The house of the seminaria area as a sement and the house of the seminaria as a sement and the house of the state of the seminaria as a sement and the seminaria area as a seminaria as a semina				1482	0			1000								W.					
Distance from the airport  And the house seminaria as sement, he house of views area as a sement, stal Code dear a sement and state of floors.  The house of the seminaria area as a sement and the house of the seminaria as a sement and the house of the state of the seminaria as a sement and the seminaria area as a seminaria as a semina	Number	of schools ne	arby	1462	01	ONLY TO	72.15	1	1	gdy -		100	10	-	Copt.	1 0	900		1	1	y I
				4462	.e≦	T S	ea	ea	Ors	i,	WS	Se	Se	Ŧ	T.	ā	ā	de	e	e	3
				462	.øē	00	ā	ā	<u>ŏ</u>	356	ē	00	no	je j	ne	9	9	8	ž	Ĕ	E C
		int64			다 다	אר	DG.	to	<del>+</del>	)re	<b>1</b>	ح	5	en	e	=	nc	=	E	ig.	Ξ,
	dtvne.				- March 1	100 Table 1	Bess	7	0	00000	0	<u>e</u>	Φ.	S	07	3		0.10	CTS.	_	G
CONTROL DESCRIPTION OF DESCRIPTION OF THE PROPERTY OF THE PROP	dtype:				Φ	75			3521 U.	And .				are.	10	~	-	+	19	0	m

print(df.corr())

```
Date number of bedrooms \
                                          id
                                                               -0.329034
id
                                    1.000000 0.045966
Date
                                    0.045966 1.000000
                                                                -0.015663
number of bedrooms
                                   -0.329034 -0.015663
                                                                1.000000
number of bathrooms
                                                                 0.509784
                                   -0.516909 -0.026485
living area
                                   -0.648127 -0.021958
                                                                 0.570526
```

3

```
lot area
                                           -0.100269 0.004392
                                                                           0.034416
     number of floors
                                           -0.312305 -0.010335
                                                                           0.177294
                                           -0.112937 0.012006
                                                                          -0.006257
     waterfront present
                                                                          0.078665
     number of views
                                           -0.293004 -0.004782
     condition of the house
                                           -0.045061 -0.027402
                                                                           0.026597
     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
                                           -0.068645 -0.005869
     Built Year
                                                                           0.152954
     Renovation Year
                                           -0.109155 -0.011636
                                                                          0.016132
     Postal Code
                                            0.294709 0.018243
                                                                          -0.044156
                                           -0.479334 -0.023327
     Lattitude
                                                                          -0.013163
                                           -0.070841 -0.018231
                                                                          0.135712
     Longitude
                                           -0.599900 -0.032495
     living_area_renov
                                                                           0.389855
                                           -0.089604 -0.000050
                                                                           0.029400
     lot_area_renov
     Number of schools nearby
                                           -0.004821 -0.004071
                                                                          0.003397
                                           -0.004542 0.011457
                                                                          -0.006157
     Distance from the airport
     Price
                                           -0.773114 -0.027919
                                                                          0.308460
                                            number of bathrooms living area \
     id
                                                      -0.516909
                                                                   -0.648127
                                                       -0.026485
                                                                    -0.021958
    Date
     number of bedrooms
                                                       0.509784
                                                                    0.570526
     number of bathrooms
                                                       1.000000
                                                                    0.753517
     living area
                                                       0.753517
                                                                     1.000000
     lot area
                                                       0.080806
                                                                     0.174420
     number of floors
                                                       0.502924
                                                                    0.354743
     waterfront present
                                                       0.060104
                                                                    0.105837
     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
                                                                    0.441491
                                                       0.287190
                                                                     0.309602
     Built Year
                                                       0.498127
     Renovation Year
                                                       0.049669
                                                                    0.059400
     Postal Code
                                                       -0.105546
                                                                    -0.080303
                                                                    0.054518
     Lattitude
                                                       0.031156
     Longitude
                                                       0.223904
                                                                    0.240208
                                                                     0.757571
     living_area_renov
                                                       0.570530
                                                                    0.180312
     lot area renov
                                                       0.078627
     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
                                            0.004392
                                                             -0.010335
    Date
     number of bedrooms
                                            0.034416
                                                              0.177294
     number of bathrooms
                                            0.080806
                                                              0.502924
                                            0.174420
                                                              0.354743
     living area
                                            1.000000
                                                              -0.004138
     lot area
     number of floors
                                           -0.004138
                                                              1.000000
print(df['Distance from the airport'].value_counts())
     70
           511
     64
           500
     50
           495
     79
           487
     60
           486
     67
           485
     56
           484
     53
           484
           482
     59
     69
           481
     58
           474
     55
           473
     74
           471
     72
           471
           470
     66
     62
           470
```

https://colab.resquech.google.com/drive/1DjaXDvNn2EE-xLl8uE\_0HmX2TWdl6OGf#printMode=true

75 466

61 443 51 442

Name: Distance from the airport, dtype: int64

print('Mean:',df['Area of the basement'].mean())
print('Median:',df['Distance from the airport'].median())
print('Mode:',df['grade of the house'].mode())

Mean: 296.47906976744184

Median: 65.0 Mode: 0 7

Name: grade of the house, dtype; int64

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

id a Date 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 Renovation Year Postal Code 0 Lattitude Longitude living\_area\_renov 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	g h
id									
6762810145	5	2.50	3650	9050	2.0	0	4	5	
6762810635	4	2.50	2920	4000	1.5	0	0	5	
6762810998	5	2.75	2910	9480	1.5	0	0	3	
6762812605	4	2.50	3310	42998	2.0	0	0	3	
6762812919	3	2.00	2710	4500	1.5	0	0	4	

y.head()

```
Price
     0 6762810145 2380000
      4 6762910635 1400000
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)
\verb|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)
     0.9130901919275989
r2_score(y_pred,y_test)
     0.9026890923961436
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 rc	ws × 2 column:	s