house-price-prediction

July 2, 2024

1 Importing of Libraries

```
[1]: import numpy as np
     import pandas as pd
     import matplotlib.pyplot as plt
     import seaborn as sns
     from sklearn.model_selection import train_test_split
     from sklearn.preprocessing import LabelEncoder
     from sklearn.preprocessing import OneHotEncoder
     from sklearn.compose import ColumnTransformer
     from sklearn.preprocessing import StandardScaler
     from sklearn.linear_model import LinearRegression
     from sklearn.metrics import mean_squared_error, r2_score
     import warnings
     warnings.filterwarnings('ignore')
```

2 Data Loading

```
[2]: df = pd.read_csv(r"..\Datasets\House Price India.csv")
     df
```

```
[2]:              id   Date  number of bedrooms  number of bathrooms  \
     0      6762810635  42491                   4                 2.50
     1      6762810998  42491                   5                 2.75
     2      6762812605  42491                   4                 2.50
     3      6762812919  42491                   3                 2.00
     4      6762813105  42491                   3                 2.50
     ...           ...    ...                 ...                  ...
     14614  6762830250  42734                   2                 1.50
     14615  6762830339  42734                   3                 2.00
     14616  6762830618  42734                   2                 1.00
     14617  6762830709  42734                   4                 1.00
     14618  6762831463  42734                   3                 1.00

            living area  lot area  number of floors  waterfront present  \
     0              2920      4000               1.5                   0
```

1	2910	9480	1.5	0
2	3310	42998	2.0	0
3	2710	4500	1.5	0
4	2600	4750	1.0	0
…	…	…	…	…
14614	1556	20000	1.0	0
14615	1680	7000	1.5	0
14616	1070	6120	1.0	0
14617	1030	6621	1.0	0
14618	900	4770	1.0	0

	number of views	condition of the house	…	Built Year \
0	0	5	…	1909
1	0	3	…	1939
2	0	3	…	2001
3	0	4	…	1929
4	0	4	…	1951
…	…	… …	…	…
14614	0	4	…	1957
14615	0	4	…	1968
14616	0	3	…	1962
14617	0	4	…	1955
14618	0	3	…	1969

	Renovation Year	Postal Code	Lattitude	Longitude	living_area_renov \
0	0	122004	52.8878	-114.470	2470
1	0	122004	52.8852	-114.468	2940
2	0	122005	52.9532	-114.321	3350
3	0	122006	52.9047	-114.485	2060
4	0	122007	52.9133	-114.590	2380
…	…	…	…	…	…
14614	0	122066	52.6191	-114.472	2250
14615	0	122072	52.5075	-114.393	1540
14616	0	122056	52.7289	-114.507	1130
14617	0	122042	52.7157	-114.411	1420
14618	2009	122018	52.5338	-114.552	900

	lot_area_renov	Number of schools nearby	Distance from the airport \
0	4000	2	51
1	6600	1	53
2	42847	3	76
3	4500	1	51
4	4750	1	67
…	…	…	…
14614	17286	3	76
14615	7480	3	59
14616	6120	2	64

```
14617              6631                      3                54
14618              3480                      2                55

           Price
0          1400000
1          1200000
2           838000
3           805000
4           790000
...            ...
14614       221700
14615       219200
14616       209000
14617       205000
14618       146000

[14619 rows x 23 columns]
```

3 Data Cleaning

```
[3]: df.columns
```

```
[3]: 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')
```

```
[4]: df = df.drop(['id', 'Date', 'Built Year', 'Renovation Year', 'Postal Code',␣
     ↪'Lattitude', 'Longitude'], axis = 1)
     df.head()
```

```
[4]:    number of bedrooms  number of bathrooms  living area  lot area  \
     0                   4                 2.50         2920      4000
     1                   5                 2.75         2910      9480
     2                   4                 2.50         3310     42998
     3                   3                 2.00         2710      4500
     4                   3                 2.50         2600      4750

        number of floors  waterfront present  number of views  \
     0               1.5                   0                0
     1               1.5                   0                0
     2               2.0                   0                0
```

3

```
3                         1.5                          0                          0
4                         1.0                          0                          0

    condition of the house  grade of the house  \
0                         5                          8
1                         3                          8
2                         3                          9
3                         4                          8
4                         4                          9

    Area of the house(excluding basement)  Area of the basement  \
0                                     1910                    1010
1                                     2910                       0
2                                     3310                       0
3                                     1880                     830
4                                     1700                     900

    living_area_renov  lot_area_renov  Number of schools nearby  \
0                2470            4000                         2
1                2940            6600                         1
2                3350           42847                         3
3                2060            4500                         1
4                2380            4750                         1

    Distance from the airport     Price
0                           51  1400000
1                           53  1200000
2                           76   838000
3                           51   805000
4                           67   790000
```

[5]: `df.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 14619 entries, 0 to 14618
Data columns (total 16 columns):
 #   Column                                 Non-Null Count  Dtype
---  ------                                 --------------  -----
 0   number of bedrooms                     14619 non-null  int64
 1   number of bathrooms                    14619 non-null  float64
 2   living area                            14619 non-null  int64
 3   lot area                               14619 non-null  int64
 4   number of floors                       14619 non-null  float64
 5   waterfront present                     14619 non-null  int64
 6   number of views                        14619 non-null  int64
 7   condition of the house                 14619 non-null  int64
 8   grade of the house                     14619 non-null  int64
```

```
 9   Area of the house(excluding basement)  14619 non-null  int64
 10  Area of the basement                    14619 non-null  int64
 11  living_area_renov                       14619 non-null  int64
 12  lot_area_renov                          14619 non-null  int64
 13  Number of schools nearby                14619 non-null  int64
 14  Distance from the airport               14619 non-null  int64
 15  Price                                   14619 non-null  int64
dtypes: float64(2), int64(14)
memory usage: 1.8 MB
```

[6]: `df.isnull().sum()`

```
[6]: number of bedrooms                      0
     number of bathrooms                     0
     living area                             0
     lot area                                0
     number of floors                        0
     waterfront present                      0
     number of views                         0
     condition of the house                  0
     grade of the house                      0
     Area of the house(excluding basement)   0
     Area of the basement                    0
     living_area_renov                       0
     lot_area_renov                          0
     Number of schools nearby                0
     Distance from the airport               0
     Price                                   0
     dtype: int64
```

[7]: `for i in df.columns: print(i, ':\n', df[i].unique())`

```
number of bedrooms :
 [ 4  5  3  2  7  6  9  1  8 33 10 11]
number of bathrooms :
 [2.5  2.75 2.   3.25 1.75 2.25 1.   1.5  3.   8.   3.5  4.25 4.   5.
 3.75 4.5  5.75 1.25 6.5  4.75 0.75 5.25 5.5  6.25 6.75 7.5  0.5  6.
 7.75]
living area :
 [ 2920  2910  3310  2710  2600  3660  2240  2390  2200  2820  1820  1520
  1750  2730  2360  3240  2330  1940  2860  1600  2190   880  2210  1710
  2680  2430  1270  2160   800  1770  1970  1450  1920  2000  1570  1930
  1960  3630  1320  1560  3110  1090  1350  1580  1160  1340   850  2540
  1630  1850  2090  1800  1180  1480  1370  1200  1210   900  1650  1010
  2850  2280  2530 13540  2220  4010  4340  3190  1330  4420  4490  3690
  2170  3400  3180  2350  2120  2010  2450  2490  1680  2800  1390  2440
  1990  2300  1670  1880  1900  1140  2740  2550  1790  2070  3200  2340
  3040  1840  1590  1410  1408  1620  2110   840  2230  3090  1571  1120
```

2370	2310	1050	1550	1240	2050	810	1490	4510	3760	3490	3370
2690	3020	3740	3880	2290	3030	930	3380	3480	2060	2150	2260
3569	2130	3100	2970	2460	1430	1170	2020	1420	1660	3470	1030
1230	1100	2810	1440	1260	2890	1500	1310	1640	2331	1110	720
1890	1510	1300	1980	1690	780	6210	4270	3560	2960	4460	3900
4130	2410	2520	2650	1760	3680	2270	2030	3830	2510	4120	1860
3700	2720	1540	990	2656	3305	1610	2180	820	1150	1040	1606
1830	2140	3000	2420	1870	760	1250	1720	1400	1380	910	3510
3820	3410	2015	3160	2640	1910	3010	2480	2660	3290	2250	1360
1780	1060	1020	980	1460	3206	960	2320	860	1810	1730	950
1190	2100	520	7270	5190	3720	5860	4230	3750	3810	3780	3540
2830	2790	2950	2400	2840	2470	3060	1445	1352	1470	998	1220
2163	3650	3460	5420	3070	3280	3270	4530	2980	3300	2700	4700
4630	2620	2760	2670	2040	2560	2780	1290	1280	2380	1080	830
700	970	5300	4430	5030	3120	4860	3530	4400	3520	3150	3610
1740	1950	1530	3440	2876	3360	3001	650	790	3260	4910	4470
3320	2580	1700	5774	3430	1159	2257	3550	4210	3920	3950	2930
3930	3580	4560	2870	1232	6490	5520	2134	1726	1070	870	1000
5340	5010	4180	2500	770	2080	680	4380	2880	3500	2770	3230
3050	2303	1092	5305	4850	5330	3488	2750	2630	1396	5060	6040
3600	3860	4260	3340	3250	3220	590	3140	550	2052	1347	2570
940	740	4390	4740	4080	4220	3840	1814	4060	750	580	430
4580	5584	3670	4590	3080	3790	5940	3330	2610	3390	920	1296
1255	1982	1658	4440	3420	3910	2990	4140	2798	4250	3890	3753
2495	6980	4730	4960	5960	2900	2590	4410	710	5550	5700	4290
3990	4240	2675	3640	4370	4500	1495	730	1175	2961	3800	2481
1765	4720	560	2437	670	4670	3902	5180	5660	3590	4610	4160
2658	1463	2192	2068	1130	4280	3450	4150	4600	890	1405	1811
3730	5270	2940	3870	2732	1392	2009	1651	10040	8670	1295	3170
2456	2414	982	620	4030	5730	3960	2683	1769	420	4310	4200
3770	4070	3210	1384	3130	1212	9640	4100	4360	4170	1494	1413
1088	6430	6050	4680	2807	2795	3850	1689	1584	660	5990	3620
4750	2496	2238	5480	4640	3366	6630	6900	3940	380	5230	4570
2242	4550	3831	6390	2341	4690	4620	2034	690	5850	4660	4225
4000	3570	2025	1481	4480	4040	2085	1444	6640	2432	1715	4350
4830	5000	3350	1068	7050	3970	2375	2701	370	8020	3002	1954
2105	5070	4110	4930	3316	5430	3595	4050	2286	1601	1995	6500
4710	7320	3176	1785	1984	2538	630	5780	2783	3710	490	4800
5470	6070	5720	1714	1889	4575	4300	640	4650	5930	5320	2846
2448	1322	4920	1256	4475	833	3274	2311	1578	3045	2259	1798
3980	3691	2064	3281	2154	2029	4520	1278	2229	2038	1095	1852
2115	3786	3906	3555	2031	1239	5150	4065	3136	2993	1983	1834
2497	1489	1847	3217	5090	5760	5240	6400	1778	3172	2093	2217
902	4540	7350	1233	2475	4190	2979	4168	9200	5790	570	1833
1264	1867	1594	2075	1747	5310	4330	2198	2575	2623	6055	1313
3361	2692	2393	1076	1048	7000	5210	2815	1782	2007	2927	2267
6030	1509	2382	1646	1894	6550	5220	4870	1639	6880	3847	2641
2245	4320	2313	480	4810	1358	600	610	4285	4133	6085	6160

```
4090  2166  2398   530  5810  1553  5370  1458  6200  3931  1484  2425
1981  4790  5020  3135  3732  1728  3672  2738  5510  5400  4450  2095
6510  4770  3236   901  5440  3597  2717  2744  1808  6240  2517  9890
4083  3223  1914  1934  1986  1794  7100  3845  5110  2168  2598  1381
5350  2835  6930  3284  1427  2891  5290  5170  2734  1992   500  4020
5130  5450  2708  2584  5050  7480  2403  2005  3202  2329  6380  1845
1757   440  4890  5490 12050  2601  1516  3118  2632   540  7080  2074
2963  5570  2344  2434  1275  8010  7710  2441  2223   470  2849  2519
1465  2588  3004  2643  5100  2557  2678  1365  2452  1936  2885  2514
1861  2542  5710  2145  2655  1657  1397  2689  3545  2506  2716  3087
2208  2195  2423  1522   893  2251  1763  5770  4760   460  6110  2714
1613  4115  1078  6330  1679  2906  2789  3266  2864   809  5410  6563
1909  1788  7620  4980  2578  1912  3273  3216  5635  1752  6410  2672
2811  1615  5540  2206  5120  1805  6840  2531  7400  2905  1676  2056
5830  1899  5080  5610  5844  1422  2092  2406  1072  2301  1987  5620
1556]
```
lot area :
```
[ 4000  9480 42998 …  3770 10425  6621]
```
number of floors :
```
[1.5 2.   1.   2.5 3.   3.5]
```
waterfront present :
```
[0 1]
```
number of views :
```
[0 2 1 4 3]
```
condition of the house :
```
[5 3 4 2 1]
```
grade of the house :
```
[ 8  9 10  7  6 12 11  5  4 13]
```
Area of the house(excluding basement) :
```
[1910 2910 3310 1880 1700 3660 1550 1440 1300 2820 1640 1520 2710 1470
 1560 1360 2730 3240 1970 1140 2190 1130  880 1460 1710 2680 1570 1270
 1080  800 1770 1600 1450 1070 1000 1930 1650 2300 1040 3110 1090 1350
 1580 1160 1340  850 2540 1280 2200 1800 1180 1480 1370 1200 1210 2920
  900 1010 1990 2010 9410 2220 2850 4340 1690 1330 3410 3200 3690 1610
 3400 2780 3180 1810 1620 2450 2490 1680 1660 1050 1400 2440 1960 1670
 1320 1900 1940 1110  780 1790 3040 1540 1590 1410 1408 2110  840 2230
 1750 1571 1120 2370 2310 1240 1430  810 1490 4510 3930 2400 2280 2690
 2080 3740 3880 2290 3030  930 3380 1030 2150 3569 1170 2000 1850 2460
  740 2020 1420 2740 3470 1020 1100  960 1260 1310 1230  860 2331  720
  980 1630 1060 4760 3560 4460 1890 3900 3170 1780 3680 2130 1720 2030
 1860 3700 2720 2430  990 2656 2245 2180  820 1760 1150 1606 1830 1500
 1870 1220  760 1250 1920 1380  910 3510 2015 1840 3010 1390 2660 3290
 3160 2250 2550 1980 1510 2090 1290 3206 2320  650  790 1190  520 3000
 6420 3390 4910 4230 3750 2560 3780 3540 2790 2950 2260 2840 2470 2330
 2340 2100 3060 2600 1352  798 2163 2160 2360 3490 3890 3720 3300 3420
 2830 2350 3210 2890 2620 2670 2040  870  670 2380  830  770  700  970
 4570 4430 3250 3120 3820 3530 4400 2650 3150 3190 2210 1740 1530 3440
 3360 2140 3001 1950 2170 4470 3810 2510 3320 3100 2580 4490 3430 2700
```

```
1159 2240  950 1820 1363 2120 3550 4210 2900 3950 3800 2060 1232 2070
3940 3570 3280 2134  940 1726  890 2480 5010 2880 2640 3480 2500 3370
 680 4380 2390 2770 3230 3050 2800 1730 2303 1092 2860 3745 2760 4850
3488 2630  920  550 1396 5060 4100 3860 4260 3090 2930 2980  590 3140
2420 2052 1347 4740 4080 4220 3500 3840  944  750  580  430 4580 5584
3670 4060 4590 3790 4950 3330 2610 2810 2520 1296 1255 1982 1658 2530
2050 3130 2990 2798 3336 2495 5330 4270 4770 4140  710 3350 4180 4290
 730  630 2675 3610 3270 1405 4130 2870 2961 2481 1765  560 2437 4670
2782 4070 3590 2658 1463 2192 2068 4280 4150 3080 3070 1811 3260 2590
4010 3760 3020 1392 2009 1651 7680 6120 4500 1105 2456 2414  806  620
4030 4410 3960 2683 2570 1769 1341  420 3770 4120 3650 1144  610 2960
3450 1212 4820 4420 4170 3870 2940 2970  995 1494 1413 1088 6430 3850
6050 3640 2807 2795 2410 1689 1584  660 4050 3600 3990 3910 2496 2238
2750 2966 4930 4000 3730  690  380 3220 2242 3831 4560 3580 3920 2341
2270 4620 4250 4660 4225 2025 1481 3460 4040 1444 6640 2432 1715 4350
3710 4390 1068 4320 3340 5190 2375 2701  370 8020 3002 1954 2105 5070
4110 4370 3316 5430 3595 2286 1536 1995 5180 3830 7320 2726 1595 1564
2538 4440 2783  490 6070 1714 1889 3905 4300 3520  640 5320 1976 2448
1087 4240 1256 4475  833 3620 3274 2311 1578 4800 3045 1491 1798 3980
3691 2064 1175 3281 2154 2029 1002 2229 2038 1095 1852 2115 1934 4480
3906 3555 2031 1239 4065 3136 2233 1983 1834 2497 1489 1847 2587 5090
5240 1778 3172 2093 2217  902 4200 4750  963 2475 4360 2979 3222 6200
 570 4610 1094 5000 1833 1264 1867 2075 1747 2198 2575 2623 1313 3361
2692 2393 1076 1048 4940 2815 1782 2007 2927 2267  500 1509 2382 1646
5400 4190  600 1479 5990 2299 2641 2313  480 4810 1358 3485 4133 6085
4090 2166 4860 2398  530 1553 5370 4330 1484  765 2425 1981 4790 5020
3135 2932 1728 3006 2738 4830 1295 3236  901 5440 3597 2717 2744 1808
2517 8860 4083 3223 1914 1746 1794 2659 1427 3845 2168 2598 1381 2835
4310 3284 2891 4540 2734 1288 4020 5130 5450 2708 2584 5310 5050 5080
2403 1605 3202 2329 6380 3630 1845 1757  440 4890 5490 8570 2601  998
3118 2632  540 5760 2074 2963 2732 2434 1275 6090 2441 2223  470 2849
2519 2588 3004 2643 2557 2678 1315 2452 1936 2885 2024 2542 5710 2145
2655 1657 1397 5550 2689 2506 1785 2716 3087 2208 2423 1248  893 1763
5770  460 6110 2714 1613 5220 4115 1078 4900 1679 1165 2906 2789 3266
2864  809 5153 1766 1788 5980 2578 1912 2671 3216 4450 4285 5140 4600
1752 5610 2811 1615  866 1984 1805 1322 2531 1384 6290 2905 1676 2056
5830 1899 4160 5844 1422 2092 2406 1072 2301 1987 1556]
Area of the basement :
[1010    0  830  900  690  950  180  280 1170 1000  360  800  670  470
 750  860 1080  170  850  310 1330  350  760  730  520 1140  630 4130
1160 1500 1290 1300  560  140  540  500  340  780  720  840  250 1220
 680 1370  160  740  910 1600  300   80  440  270 1340  920  620 1360
1210 1400  590  940  990 1030  870  960  700  400 1200  430  790  210
1100  880  240 1580  490  650  420  480  530 1450  640 1130  600 1380
1230  570  260  370  330 1060  820  220  510 1090  890 1460 1070  100
  70  660 1050 1800 1760 1250  290  550  450  710  145  200  410  150
1120 1690 1530 1350 1280  980 1420   90 1260  930 1190  610 1180  320
1780 1040  810 1270  390 1540  580 1020  516 1700 1480 1284  770  120
```

8

```
 894   380 1740 1110 2550 1950   130 1510   970 1910 2330 1670   230    60
1560 1660   906   110 1940 1620 2030   190 2180 2070 2100   417   460 1650
  50 2060 2200 2850 2000 1320   435 1860 2110 2150 2300   265 1610   862
2360   176 1490 1150 1440 4820 1850 1590    20 2080 1630 1640 1830 1720
 875 2020 1310   475 2730 1430    65 1730 1840 1520 1870 1550 1390 1470
2010   235 2190   768 1680 1810 1410   276 1852 1570 2170 2590 1960 2600
 946    10 3000   266   506 2500 3500 1240 2090 1900 1820   374 1248 2810
1548 2240 2120   508 1281 2570 1790   666 3260 1816   243 2350 2620 2720
 704 2400 3480   518 2160 1710   784   415 2580   283 2040   935    40 1135
1890   274   861 2610   225 1920   143   602 1275 1990 1770   652 2490 1930]
living_area_renov :
[2470 2940 3350 2060 2380 3320 1570 2010 2320 2820 1910 2390 2410 1300
2730 1860 4050 2570 2200 2590 2860 1090 3000 1340 2780 2080 2260 2990
1560 1320 1850 1150 1770 2340 1680 1260 1450 2070 2290 1960 2830 1440
1790 1160 1480 1100 2280 1590 1410 2310 1750 2130 1400 1380 1580 3030
1280 1940 1390 2315 2240 2350 2140 4850 1870 2610 2720 3100 4420 4530
3430 2550 1670 3070 2020 3180 2970 1690 2750 2170 3715 1950 2580 1810
3010 1350 1720 1800 2840 2330 1060 2160 2030 1880 1520 2500 1290 1470
1890 1730 2220 1840 2670 1200 1408 1620 1430 1630 1310 1760 1820 1220
1980 1130 1170 1510 1240 2488 3510 2490 2540 2120 2040 3040 3240 3130
3770 2790 2800 2530 2450 2520 2770 2000 1780 2210 1420 1660 1970 1270
1460 1500 1930 1330 1740 1370 2090 1230 2441   840 2360 1650 1490   900
 820 1700 4100 2960 3470 3820 2430 4130 2190 1990 2250 3200 2850 2560
1640 2870 2510 1180 2600 1540 1250 1040 1360 1516 2230 2440 2011 1010
1140 1070   910 1326 3450 2930 2900 3260 2920 2950 3620 1900 1210 3140
2300 1190 2527 2150 2980 1920 1600 1357 1572 4460 3890 3660 3230 3500
3080 3880 2700 2690 2100 2270 1110 1439   998 1714 1610 1550 1020 3220
4760 2890 3530 2400 3600 2480 3170 3640 2370   980 1080 1120 1830   890
1710 3740 4040 4240 4440 3290 2180 3120   990 2650 3060 1364 2420 3480
4560 3210 3390 3360 2910   950   920 1030 1530 3860 4210 3700 2740 2810
2460 2660 1232   850 3490 3150 1445 2114 1404 3910 3160 3580 2760   930
3300 5170 4060 3920 2880 3610 2303 1862 1050 3850 3840 1000 2110 2680
2050 2620 3790 2415 3440 2640 3110 2052 2095 3630 2710 3270 5030 3680
 970 1571 1307 1658 3540 4290 2358 3370 1665 3494 2434   860   880 3930
3710 4140 1365 4020 3690 3750 3590 1346 3330 2630 1518 3190 1495 2305
3730 2037 2363 1765 3810 4090 3280 4390 2027   960 2437   770   700 4900
3960 3050 2578 1484 2583 1914 4280 2412 4070 3380 1405 1811 3250 3550
2518 3020 2106 2009 1188 4630 3800 4670 3950 1295 2478   740 3310 4180
2683 2955 4000 3400 3900 3670 3780 4400 3420   830   460 1256 1494 1098
3720 3560 2028 1459 1584 3340 2496 1934 2456 4470 4170 3980 1798 2376
2594 2214 1768 4550 4010 2554 4950 1277 1156   940 2667 5080 5790 3830
3639 1664 1481 4080 2502 4620 3410 3090 3618 2912 2238 1078 5070 3970
4490 3570 2516   780 1767 4160 3760 3520 2566 1678 4920 3650 4510 4030
3625 2165 2156 2641 3460 4340   800 4680 4300 2234   760 3990 4640 1746
1569 1696 2815 1309   870 2458 4750 3045 1894 2648 1802 2598 2154 2029
1616 2738 2634 2166 2673 1137 4270 4310 1979 1537 1847 4150 2996 1546
1813 2704 5380 3721 4190 2475   790 4362   806 4330 2597 1522 1466 1264
2616 1536 4042 4230 2198 2575 4890 3112 1745 1448 2574 2439 1076   810
```

```
 4913 2798 2189 1528 3940 2533 2622 5200 2056 1458 1509 2382 1975 4120
 4110 4590 4690 2451 1984 2323 1358 5600 2142 3191 1336 4320 4830 4225
 2474 3425 2316 2688 2112 3557 5110 1716 2725 2396 1981 4930 3008 1554
 1442 1463 4480 1638 3236 1138 2876 3193  750 2424 2901 4540 1303 1919
 2049 2077 1381  710 1282 2612 1941 2136 4370 2875 2555 2304 1443 3159
 2767 4940 4570 2425 1268 1399 1356 2221  720 4770 2665 3078 2344 2246
 1639 2724 2092 2389 2406 1566 1168  670 2419 2014 2879 2015 3543 2619
 1092 1608 1884 1691 2927 4800 2495 1845 1763 4410 2873 2258 1427  690
  620 2405 4200 1415 2547 3087 2091 4650 2822 2961 2647 3870 3726 4600
 2765 2242 2728 1056 1429 2604 6110 4220 5340 2255 4730 3413 1886 3515
 1321 1677 4250 1425 2697 1654 1162]
lot_area_renov :
 [ 4000  6600 42847 …  1608 17286  6631]
Number of schools nearby :
 [2 1 3]
Distance from the airport :
 [51 53 76 67 72 71 73 69 80 74 55 70 75 60 58 50 64 66 79 78 61 52 68 62
 63 77 56 65 57 59 54]
Price :
 [1400000 1200000  838000 …  221700  219200  146000]
```

[8]:
```python
for i in ['number of bathrooms', 'number of floors']: df[i] = df[i].
  ↪astype('int')
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 14619 entries, 0 to 14618
Data columns (total 16 columns):
 #   Column                               Non-Null Count  Dtype
---  ------                               --------------  -----
 0   number of bedrooms                   14619 non-null  int64
 1   number of bathrooms                  14619 non-null  int32
 2   living area                          14619 non-null  int64
 3   lot area                             14619 non-null  int64
 4   number of floors                     14619 non-null  int32
 5   waterfront present                   14619 non-null  int64
 6   number of views                      14619 non-null  int64
 7   condition of the house               14619 non-null  int64
 8   grade of the house                   14619 non-null  int64
 9   Area of the house(excluding basement) 14619 non-null  int64
 10  Area of the basement                 14619 non-null  int64
 11  living_area_renov                    14619 non-null  int64
 12  lot_area_renov                       14619 non-null  int64
 13  Number of schools nearby             14619 non-null  int64
 14  Distance from the airport            14619 non-null  int64
 15  Price                                14619 non-null  int64
dtypes: int32(2), int64(14)
memory usage: 1.7 MB
```
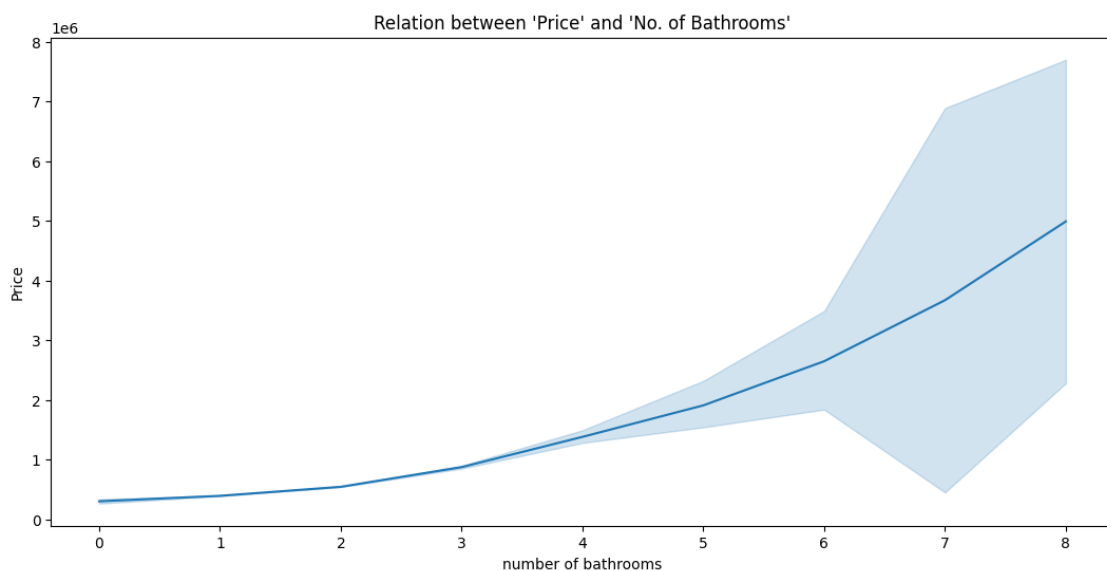
4 Data Analysis

```
[9]: plt.figure(figsize = (13,6))
     sns.barplot(data = df, x = 'number of bedrooms', y = 'Price')
     plt.title("Relation between 'Price' and 'No. of Bedrooms'")
     plt.show()
```
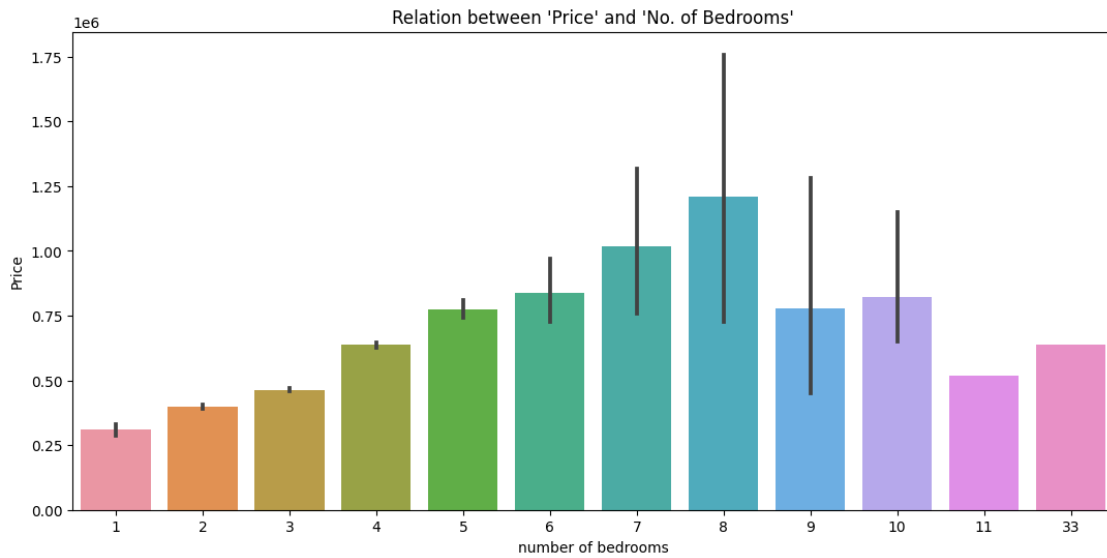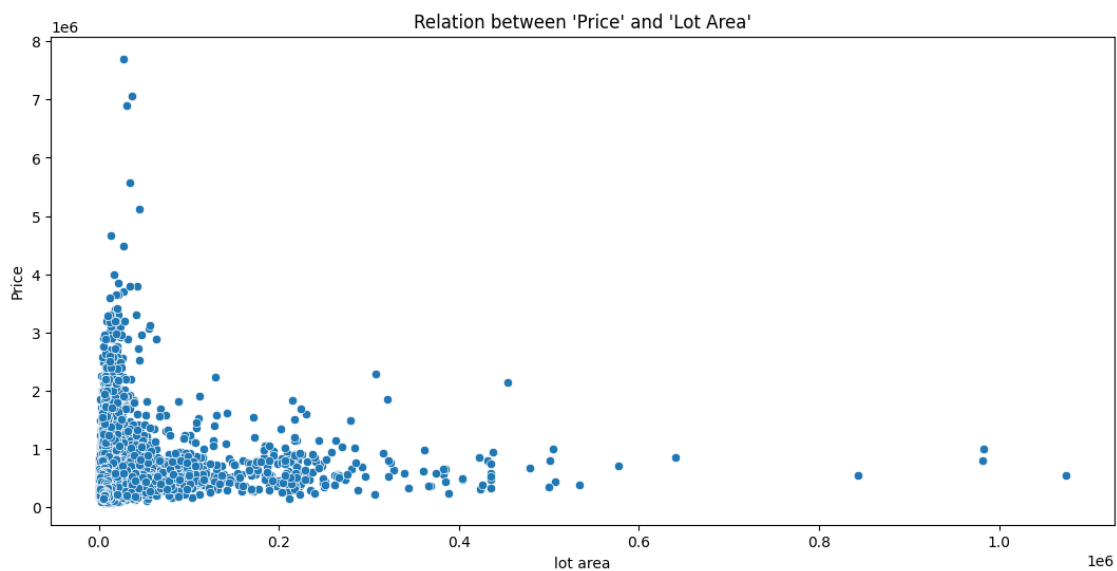


```
[10]: plt.figure(figsize = (13,6))
      sns.lineplot(data = df, x = 'number of bathrooms', y = 'Price')
      plt.title("Relation between 'Price' and 'No. of Bathrooms'")
      plt.show()
```

```
[11]: plt.figure(figsize = (13,6))
      sns.scatterplot(data = df, x = 'living area', y = 'Price')
      plt.title("Relation between 'Price' and 'Living Area'")
      plt.show()
```


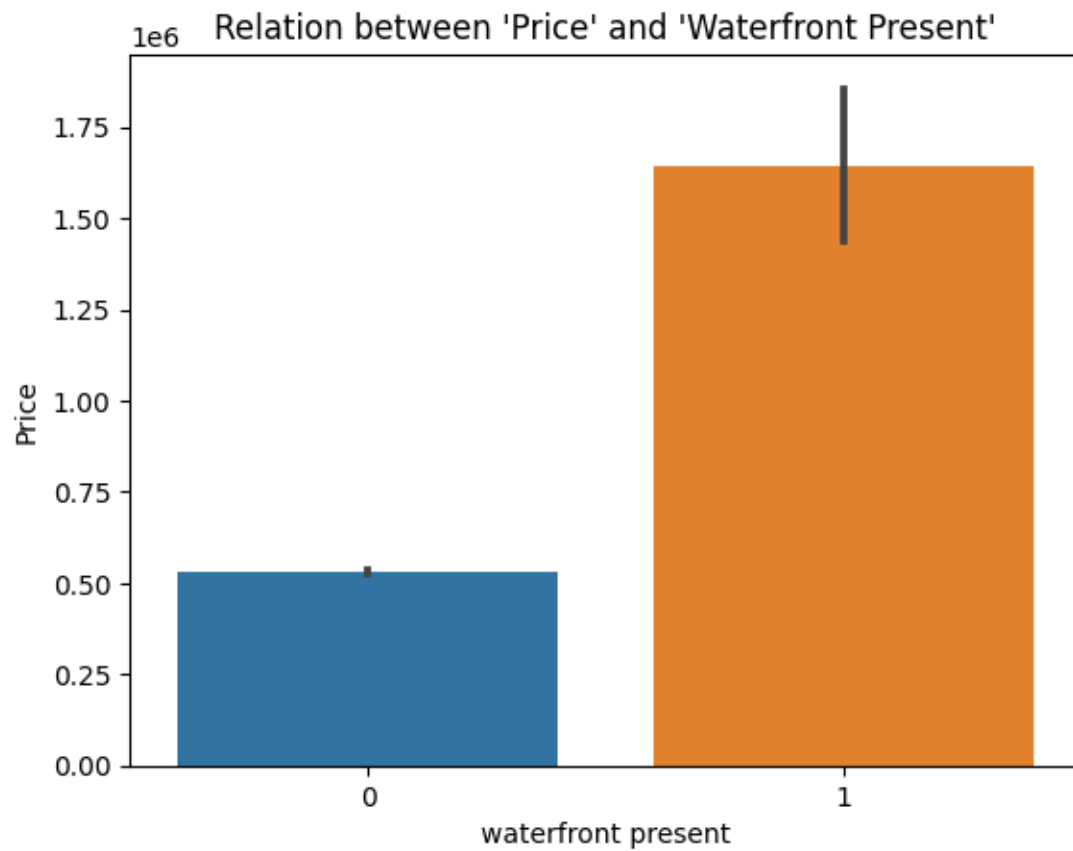Relation between 'Price' and 'Living Area'

```
[12]: plt.figure(figsize = (13,6))
      sns.scatterplot(data = df, x = 'lot area', y = 'Price')
      plt.title("Relation between 'Price' and 'Lot Area'")
      plt.show()
```
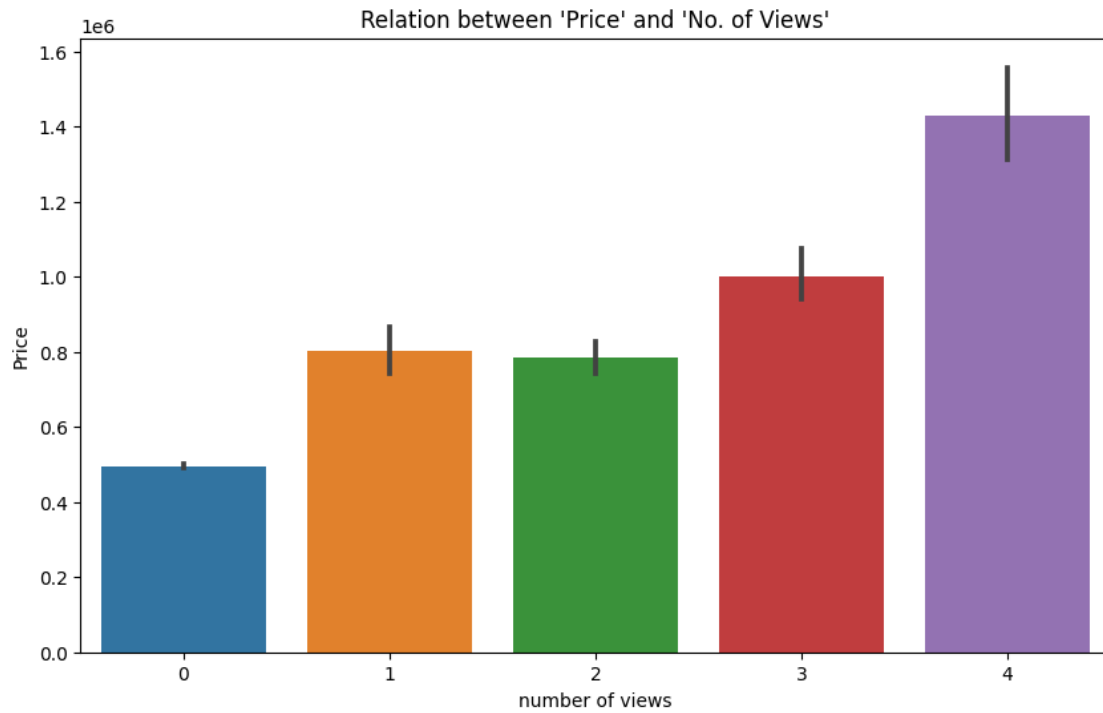

Relation between 'Price' and 'Lot Area'

```
[13]: sns.barplot(data = df, x = 'number of floors', y = 'Price')
      plt.title("Relation between 'Price' and 'No. of Floors'")
      plt.show()
```


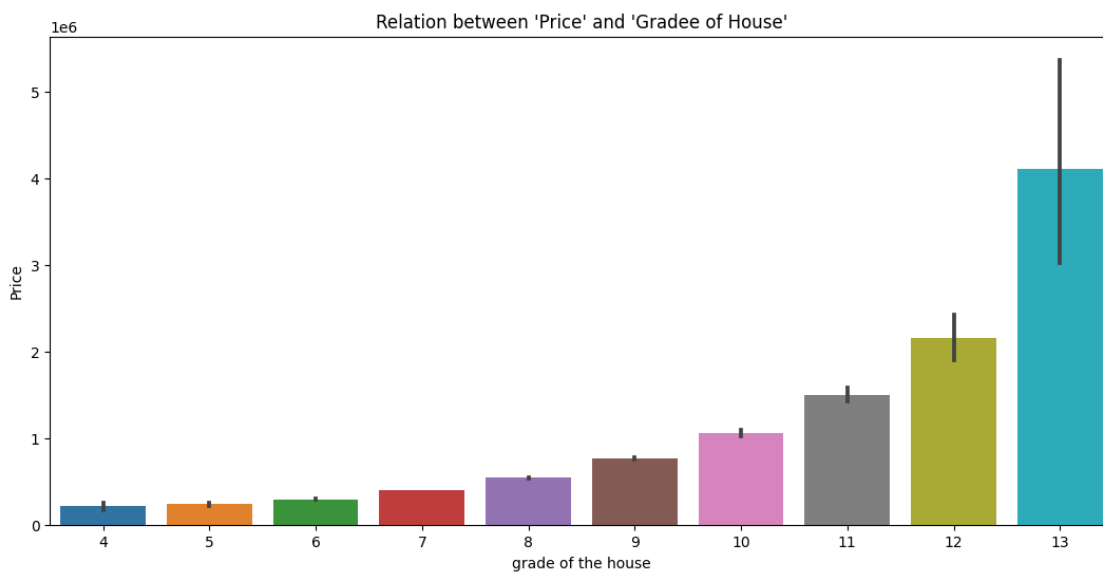Relation between 'Price' and 'No. of Floors'
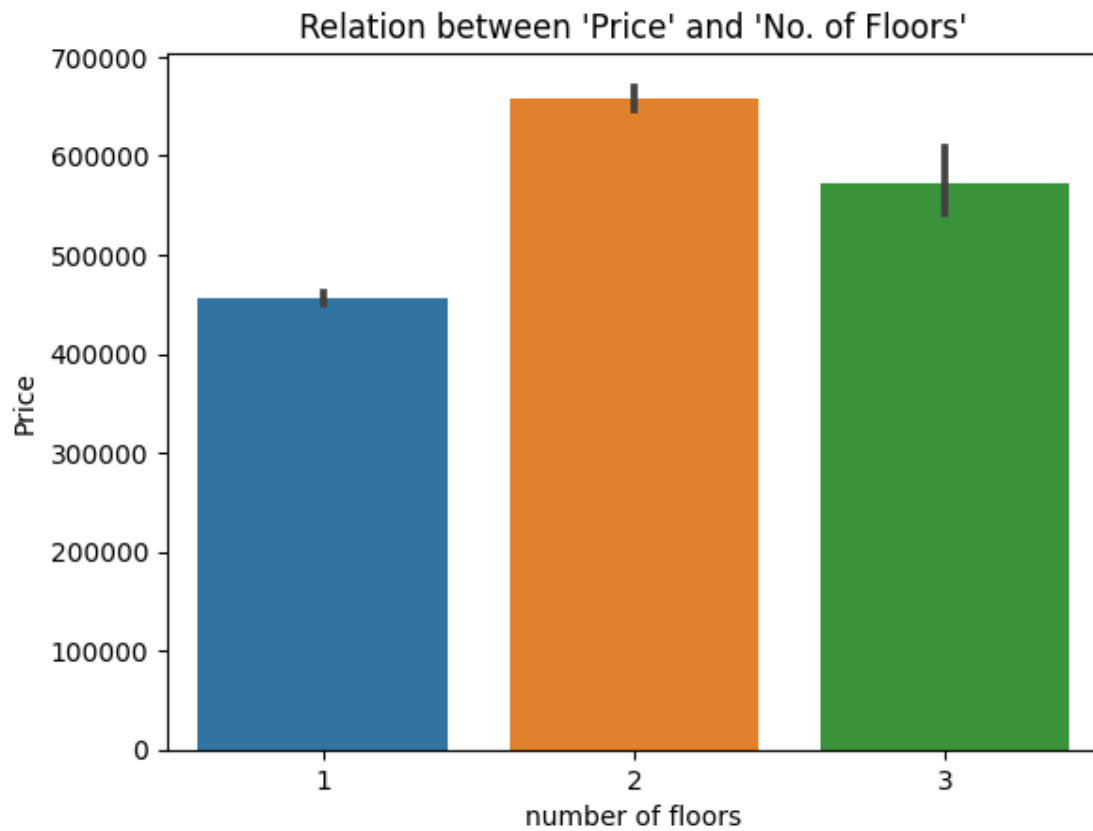
```
[14]: sns.barplot(data = df, x = 'waterfront present', y = 'Price')
      plt.title("Relation between 'Price' and 'Waterfront Present'")
      plt.show()
```

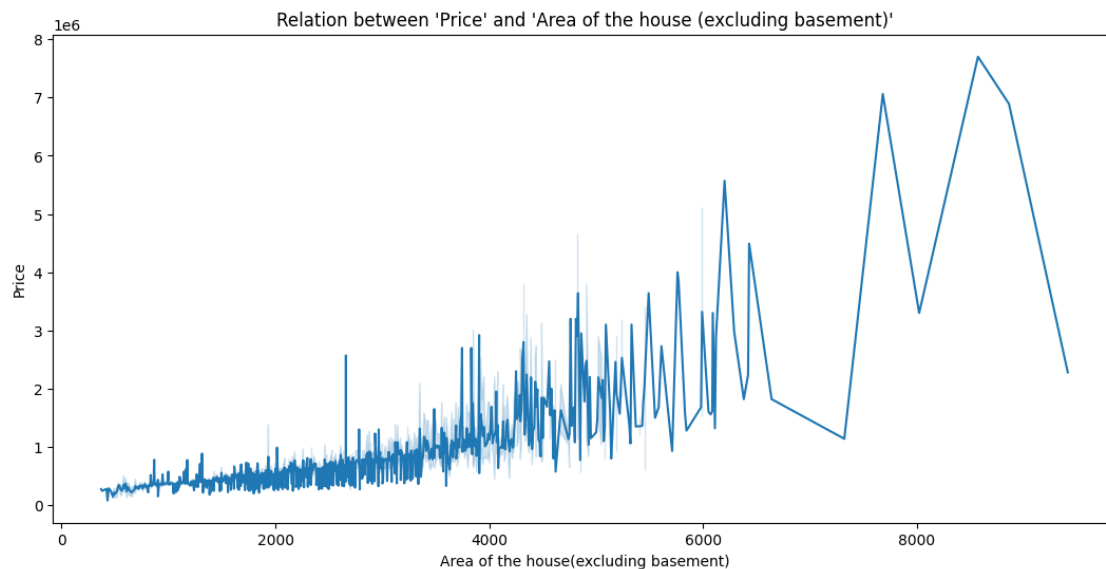Relation between 'Price' and 'Waterfront Present'

```
[15]: plt.figure(figsize = (10,6))
      sns.barplot(data = df, x = 'number of views', y = 'Price')
      plt.title("Relation between 'Price' and 'No. of Views'")
      plt.show()
```

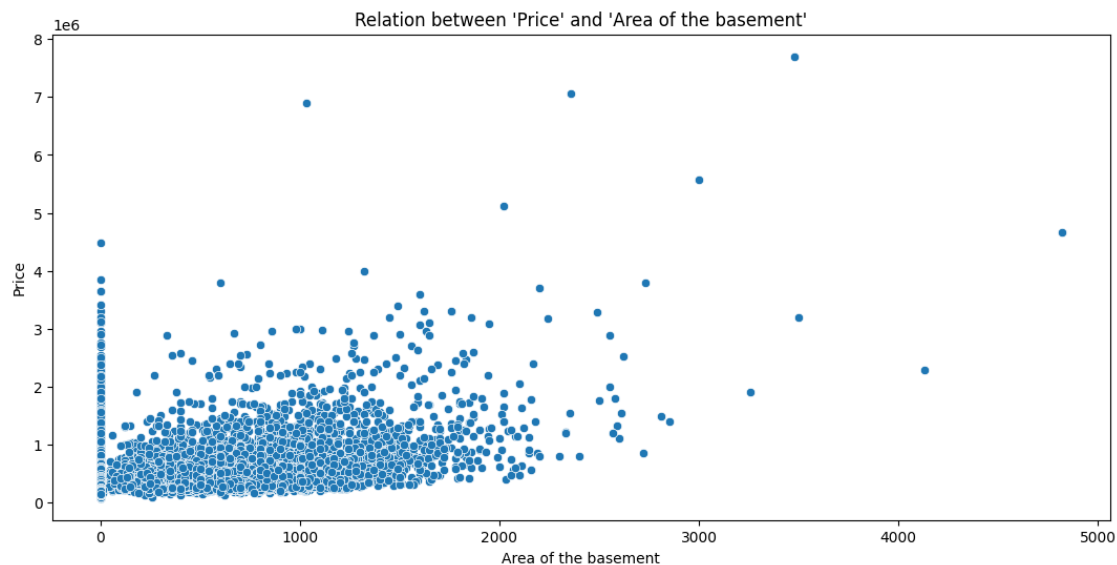Relation between 'Price' and 'No. of Views'

```
[16]: plt.figure(figsize = (13,6))
      sns.barplot(data = df, x = 'grade of the house', y = 'Price')
      plt.title("Relation between 'Price' and 'Gradee of House'")
      plt.show()
```
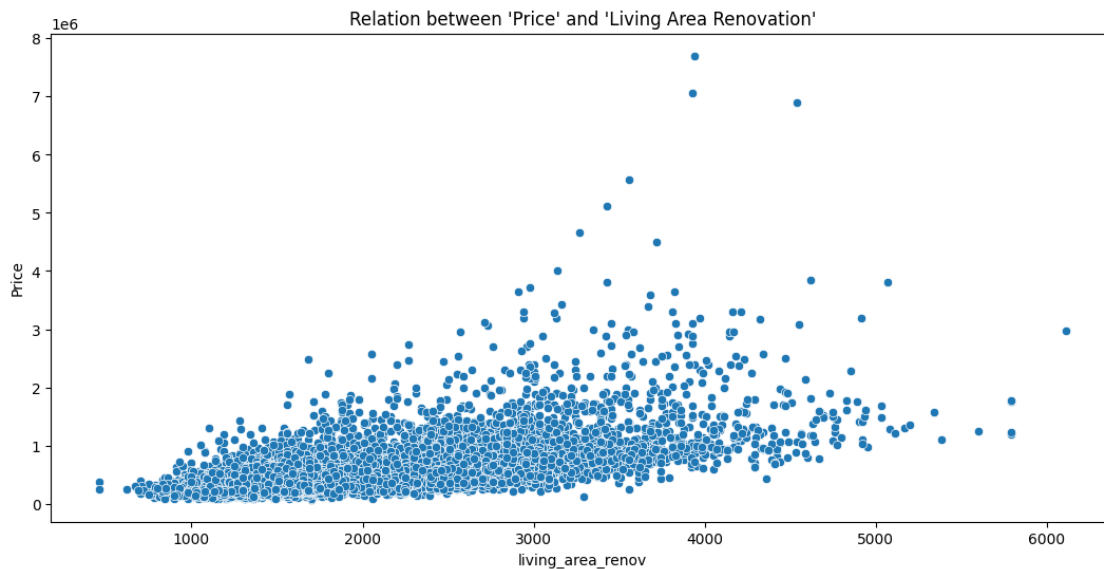

Relation between 'Price' and 'Gradee of House'

15

```
[17]: plt.figure(figsize = (13,6))
      sns.lineplot(data = df, x = 'Area of the house(excluding basement)', y =␣
       ↪'Price')
      plt.title("Relation between 'Price' and 'Area of the house (excluding␣
       ↪basement)'")
      plt.show()
```



```
[18]: plt.figure(figsize = (13,6))
      sns.scatterplot(data = df, x = 'Area of the basement', y = 'Price')
      plt.title("Relation between 'Price' and 'Area of the basement'")
      plt.show()
```

```
[19]: plt.figure(figsize = (13,6))
      sns.scatterplot(data = df, x = 'living_area_renov', y = 'Price')
      plt.title("Relation between 'Price' and 'Living Area Renovation'")
      plt.show()
```


Relation between 'Price' and 'Living Area Renovation'

```
[20]: plt.figure(figsize = (13,6))
      sns.scatterplot(data = df, x = 'lot_area_renov', y = 'Price')
      plt.title("Relation between 'Price' and 'Lot Area Renovation'")
      plt.show()
```


Relation between 'Price' and 'Lot Area Renovation'
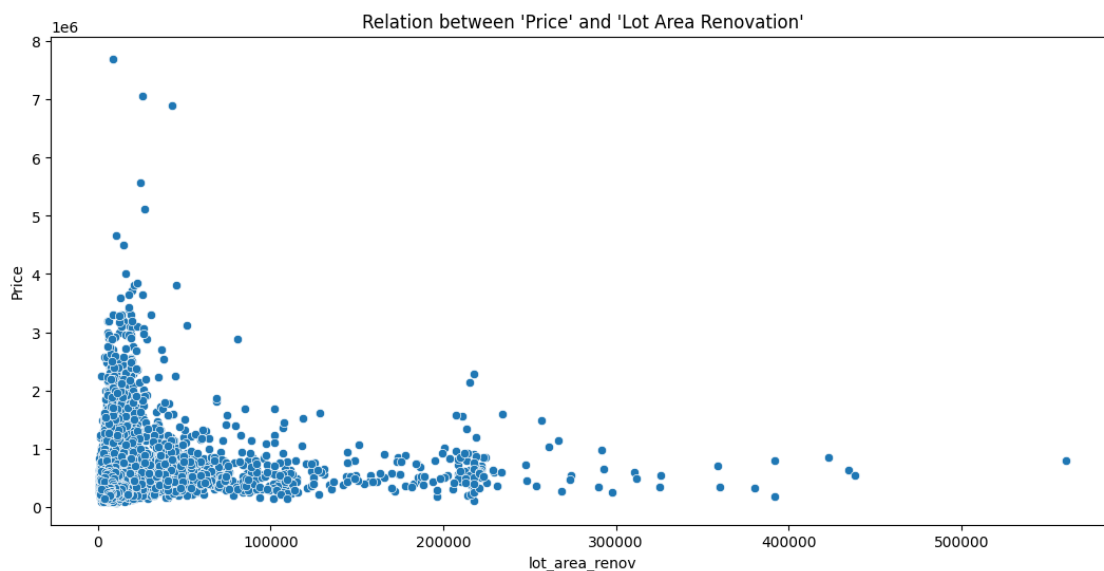
```
[21]: sns.barplot(data = df, x = 'Number of schools nearby', y = 'Price')
      plt.title("Relation between 'Price' and 'Number of schools nearby'")
      plt.show()
```

Relation between 'Price' and 'Number of schools nearby'
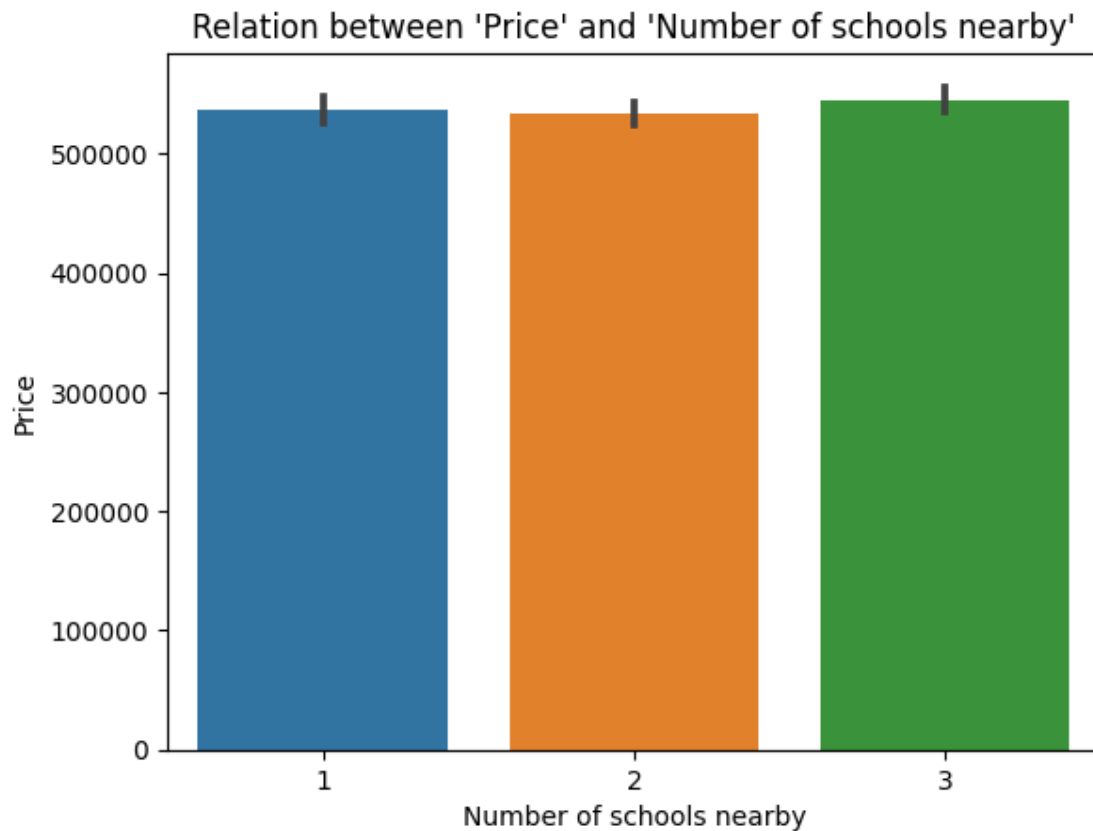

```
[22]: plt.figure(figsize = (15,6))
      sns.lineplot(data = df, x = 'Distance from the airport', y = 'Price')
      plt.title("Relation between 'Price' and 'Distance from the airport'")
      plt.show()
```
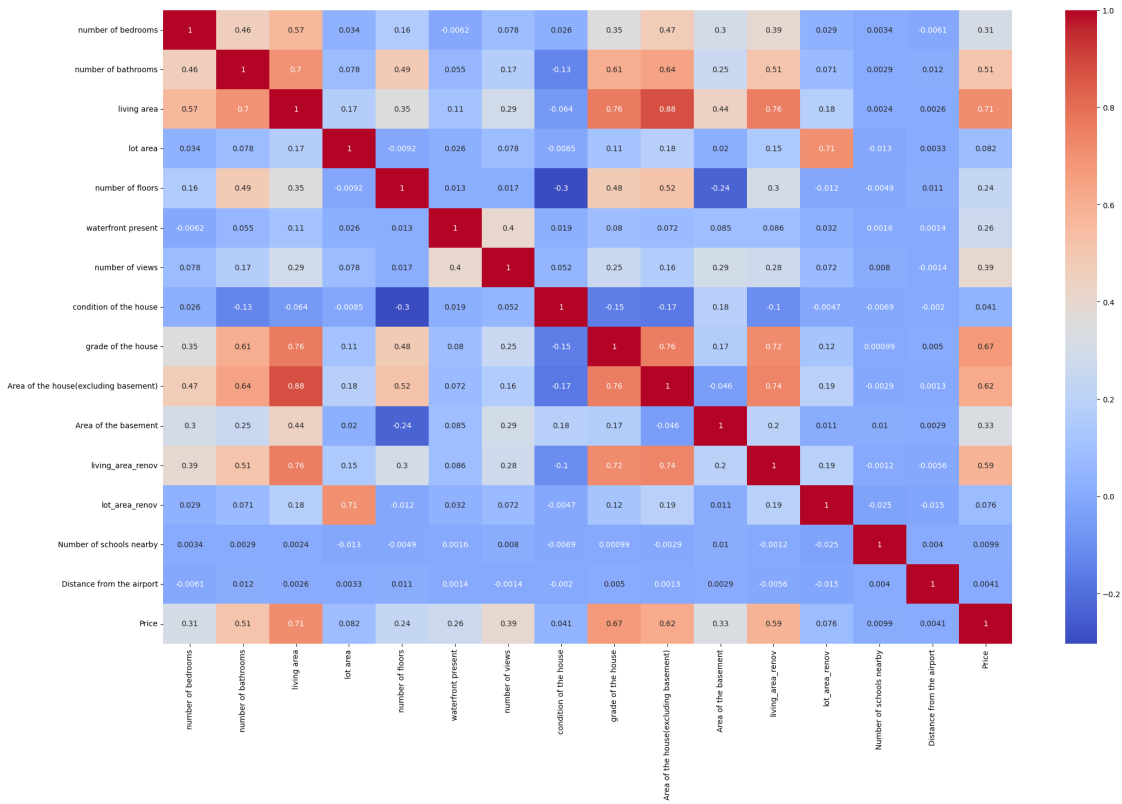
Relation between 'Price' and 'Distance from the airport'
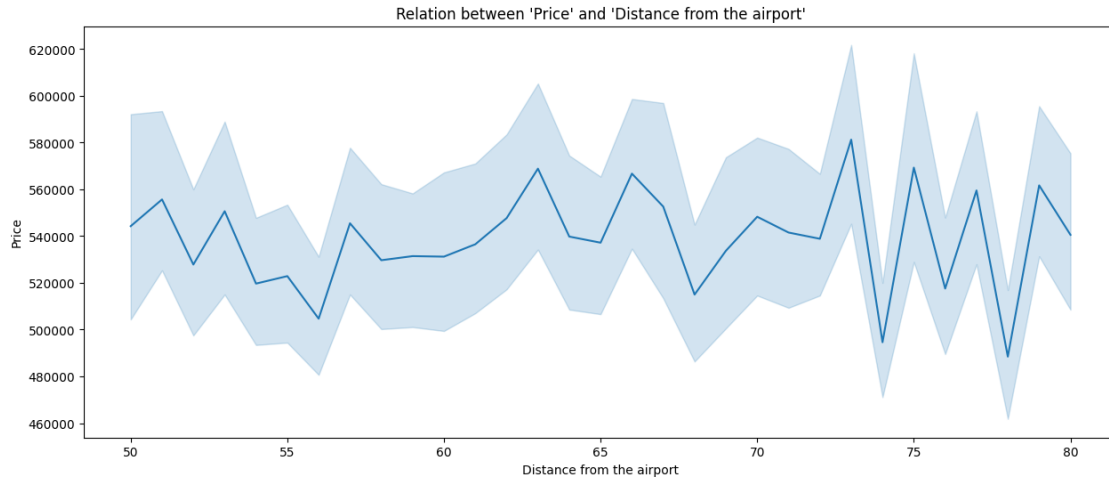
5 Co-Relation Heatmap

```
[23]: plt.figure(figsize = (25, 15))
      sns.heatmap(df.corr(), annot = True, cmap = 'coolwarm')
      plt.show()
```

6 Input & Output Creation

```
[24]: ip = df.drop(['Price'], axis = 1)
      ip
```

[24]:

	number of bedrooms	number of bathrooms	living area	lot area \
0	4	2	2920	4000
1	5	2	2910	9480
2	4	2	3310	42998
3	3	2	2710	4500
4	3	2	2600	4750
...	...	...	...	...
14614	2	1	1556	20000
14615	3	2	1680	7000
14616	2	1	1070	6120
14617	4	1	1030	6621
14618	3	1	900	4770

	number of floors	waterfront present	number of views \
0	1	0	0
1	1	0	0
2	2	0	0
3	1	0	0
4	1	0	0
...	...	...	...
14614	1	0	0
14615	1	0	0
14616	1	0	0
14617	1	0	0
14618	1	0	0

	condition of the house	grade of the house \
0	5	8
1	3	8
2	3	9
3	4	8
4	4	9
...	...	...
14614	4	7
14615	4	7
14616	3	6
14617	4	6
14618	3	6

	Area of the house(excluding basement)	Area of the basement \

```
0                              1910                  1010
1                              2910                     0
2                              3310                     0
3                              1880                   830
4                              1700                   900
...                             ...                   ...
14614                          1556                     0
14615                          1680                     0
14616                          1070                     0
14617                          1030                     0
14618                           900                     0

       living_area_renov  lot_area_renov  Number of schools nearby  \
0                   2470            4000                         2
1                   2940            6600                         1
2                   3350           42847                         3
3                   2060            4500                         1
4                   2380            4750                         1
...                  ...             ...                       ...
14614               2250           17286                         3
14615               1540            7480                         3
14616               1130            6120                         2
14617               1420            6631                         3
14618                900            3480                         2

       Distance from the airport
0                             51
1                             53
2                             76
3                             51
4                             67
...                          ...
14614                         76
14615                         59
14616                         64
14617                         54
14618                         55

[14619 rows x 15 columns]
```

```
[25]: op = df['Price']
      op
```

```
[25]: 0          1400000
      1          1200000
      2           838000
      3           805000
```

```
4        790000
         …
14614    221700
14615    219200
14616    209000
14617    205000
14618    146000
Name: Price, Length: 14619, dtype: int64
```

7   Train Test Split

```
[26]: x_train, x_test, y_train, y_test = train_test_split(ip, op, train_size = 0.7)
      df.shape, x_train.shape, x_test.shape
```

```
[26]: ((14619, 16), (10233, 15), (4386, 15))
```

8   Standard Scaler Transform

```
[27]: sc = StandardScaler()
      sc
```

```
[27]: StandardScaler()
```

```
[28]: x_train = sc.fit_transform(x_train)
      x_test = sc.fit_transform(x_test)
```

9   Implementation of ML Model

```
[29]: lr = LinearRegression()
      lr.fit(x_train, y_train)
```

```
[29]: LinearRegression()
```

10   Prediction

```
[30]: ycap = lr.predict(x_test)
      ycap
```

```
[30]: array([576391.6933568 , 573317.2142694 , 211286.27847616, …,
             485421.00881984, 355424.01888008, 117648.16965879])
```

```
[31]: pdf = pd.DataFrame({'y':list(y_test),
                          'prediction':ycap})
```

```
pdf
```

[31]:
```
              y    prediction
0        410000  5.763917e+05
1        600000  5.733172e+05
2        150000  2.112863e+05
3        325000  2.422692e+05
4        385000  4.000103e+05
...         ...           ...
4381     980000  1.005363e+06
4382     405000  3.023342e+05
4383     310950  4.854210e+05
4384     588000  3.554240e+05
4385     650000  1.176482e+05

[4386 rows x 2 columns]
```

11 Metrics

[32]:
```python
mse = mean_squared_error(y_test, ycap)
print("Mean Squared Error: ", mse)
```

```
Mean Squared Error:  52638013499.98443
```

[33]:
```python
r2 = r2_score(y_test, ycap)
print("R2 Score:", r2)
print("Accuracy:", r2*100)
```

```
R2 Score: 0.6265047911138439
Accuracy: 62.65047911138439
```

[35]:
```python
plt.figure(figsize = (13, 6))
sns.kdeplot(df['Price'], label = "Actual Value")
sns.kdeplot(y_test, label = "Testing Values")
sns.kdeplot(ycap, label = "Predicted Values")
plt.legend()
plt.show()
```