```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from matplotlib import rcParams
import seaborn as sns

df = pd.read_csv('/content/House Price India.csv')
dfh = df.head()
dfl = df.tail()
dfh
```

```
           id   Date  number of bedrooms  number of bathrooms  living
area  \
0  6762810145  42491                   5                 2.50
3650
1  6762810635  42491                   4                 2.50
2920
2  6762810998  42491                   5                 2.75
2910
3  6762812605  42491                   4                 2.50
3310
4  6762812919  42491                   3                 2.00
2710

   lot area  number of floors  waterfront present  number of views  \
0      9050               2.0                   0                4
1      4000               1.5                   0                0
2      9480               1.5                   0                0
3     42998               2.0                   0                0
4      4500               1.5                   0                0

   condition of the house  ...  Built Year  Renovation Year  Postal
Code  \
0                       5  ...        1921                0
122003
1                       5  ...        1909                0
122004
2                       3  ...        1939                0
122004
3                       3  ...        2001                0
122005
4                       4  ...        1929                0
122006

   Lattitude  Longitude  living_area_renov  lot_area_renov  \
0    52.8645   -114.557               2880            5400
1    52.8878   -114.470               2470            4000
2    52.8852   -114.468               2940            6600
3    52.9532   -114.321               3350           42847
4    52.9047   -114.485               2060            4500
```

```
   Number of schools nearby  Distance from the airport    Price
0                         2                         58  2380000
1                         2                         51  1400000
2                         1                         53  1200000
3                         3                         76   838000
4                         1                         51   805000

[5 rows x 23 columns]
```

df

```
                id   Date  number of bedrooms  number of bathrooms  \
0      6762810145  42491                   5                 2.50
1      6762810635  42491                   4                 2.50
2      6762810998  42491                   5                 2.75
3      6762812605  42491                   4                 2.50
4      6762812919  42491                   3                 2.00
...           ...    ...                 ...                  ...
14615  6762830250  42734                   2                 1.50
14616  6762830339  42734                   3                 2.00
14617  6762830618  42734                   2                 1.00
14618  6762830709  42734                   4                 1.00
14619  6762831463  42734                   3                 1.00

       living area  lot area  number of floors  waterfront present  \
0             3650      9050               2.0                   0
1             2920      4000               1.5                   0
2             2910      9480               1.5                   0
3             3310     42998               2.0                   0
4             2710      4500               1.5                   0
...            ...       ...               ...                 ...
14615         1556     20000               1.0                   0
14616         1680      7000               1.5                   0
14617         1070      6120               1.0                   0
14618         1030      6621               1.0                   0
14619          900      4770               1.0                   0

       number of views  condition of the house  ...  Built Year  \
0                    4                       5  ...        1921
1                    0                       5  ...        1909
2                    0                       3  ...        1939
3                    0                       3  ...        2001
4                    0                       4  ...        1929
...                ...                     ...  ...         ...
14615                0                       4  ...        1957
14616                0                       4  ...        1968
14617                0                       3  ...        1962
14618                0                       4  ...        1955
14619                0                       3  ...        1969
```

```
       Renovation Year  Postal Code  Lattitude  Longitude
living_area_renov  \
0                    0       122003    52.8645   -114.557
2880
1                    0       122004    52.8878   -114.470
2470
2                    0       122004    52.8852   -114.468
2940
3                    0       122005    52.9532   -114.321
3350
4                    0       122006    52.9047   -114.485
2060
...                ...          ...        ...        ...
...
14615                0       122066    52.6191   -114.472
2250
14616                0       122072    52.5075   -114.393
1540
14617                0       122056    52.7289   -114.507
1130
14618                0       122042    52.7157   -114.411
1420
14619             2009       122018    52.5338   -114.552
900

       lot_area_renov  Number of schools nearby  Distance from the
airport  \
0                5400                         2
58
1                4000                         2
51
2                6600                         1
53
3               42847                         3
76
4                4500                         1
51
...               ...                       ...
...
14615           17286                         3
76
14616            7480                         3
59
14617            6120                         2
64
14618            6631                         3
54
14619            3480                         2
```

```
55
```

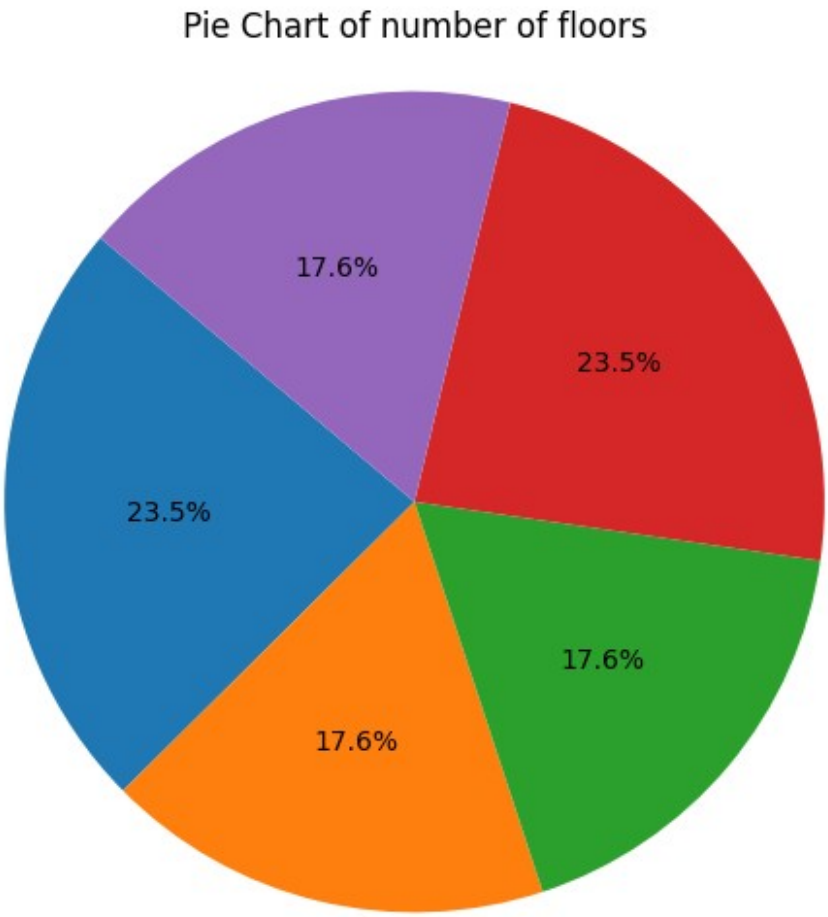
```
         Price
0      2380000
1      1400000
2      1200000
3       838000
4       805000
...        ...
14615   221700
14616   219200
14617   209000
14618   205000
14619   146000

[14620 rows x 23 columns]
```

```
df.shape
```

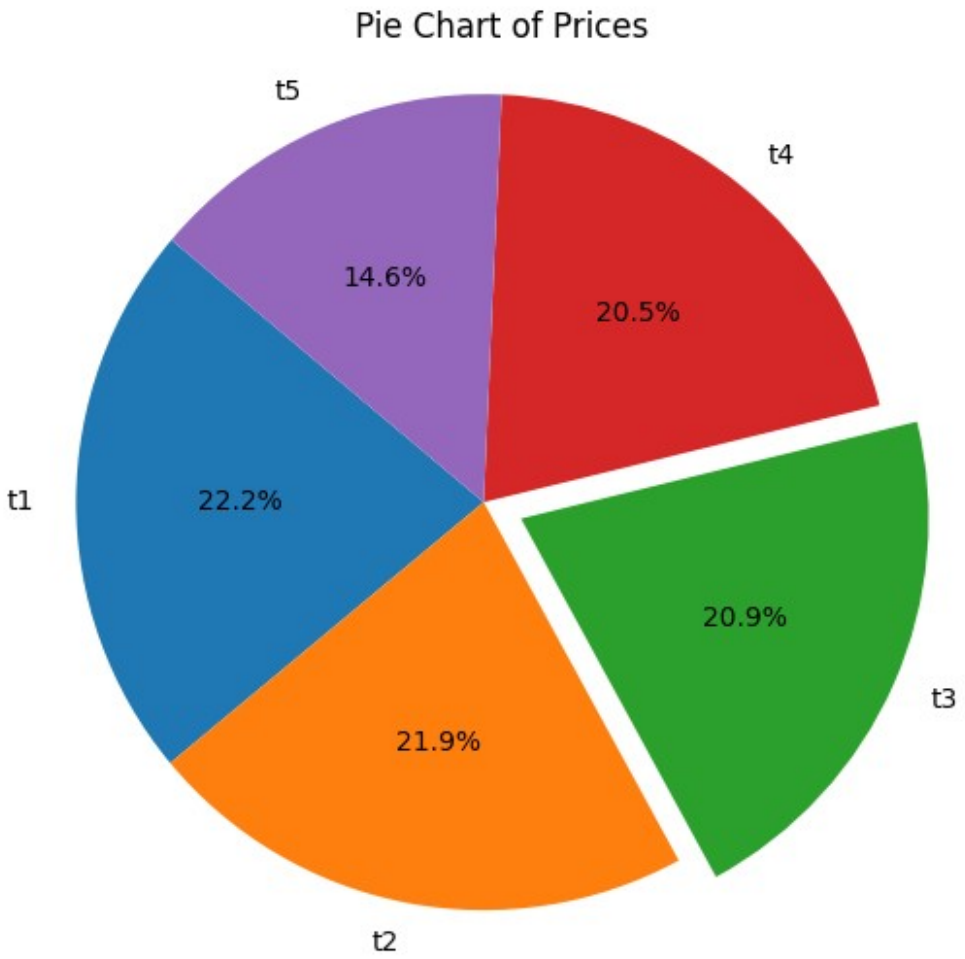
```
(14620, 23)
```

Univariate Analysis

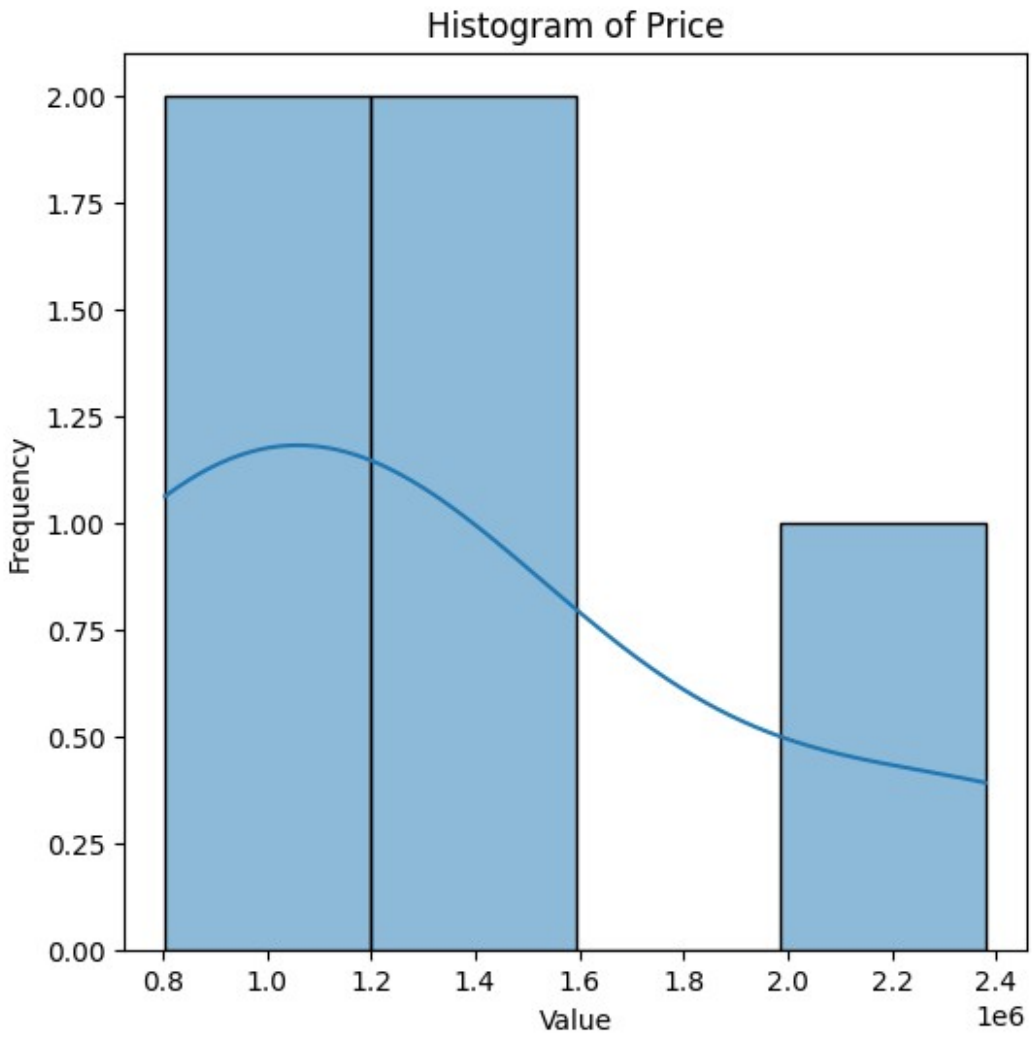
```
#Pie plot of the df(Head)
plt.figure(figsize=(6, 6))  # Set the figure size
plt.pie(dfh['number of floors'],autopct='%1.1f%%', startangle=140)
plt.title('Pie Chart of number of floors')
plt.axis('equal')  # Equal aspect ratio ensures that pie is drawn as a
circle.
plt.show()
```




```
#pie plot of df(tail)
plt.figure(figsize=(6, 6))  # Set the figure size
plt.pie(dfl['Price'],
[0,0,0.1,0,0],labels=['t1','t2','t3','t4','t5'],autopct='%1.1f%%',
startangle=140)
plt.title('Pie Chart of Prices')
plt.axis('equal')  # Equal aspect ratio ensures that pie is drawn as a
circle.
plt.show()
```

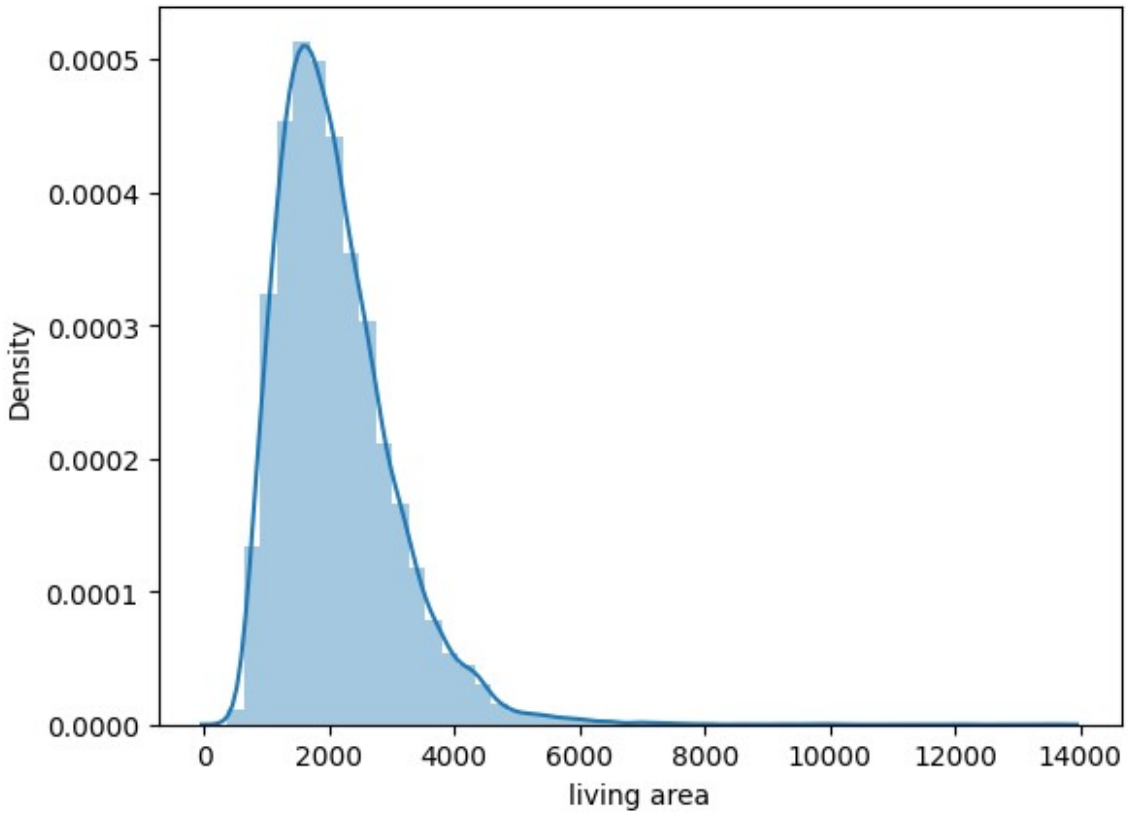



```
#histogram plot of df(head)
plt.figure(figsize=(6, 6))
sns.histplot(dfh['Price'], kde=True)
plt.title('Histogram of Price')
plt.xlabel('Value')
plt.ylabel('Frequency')
plt.show()
```



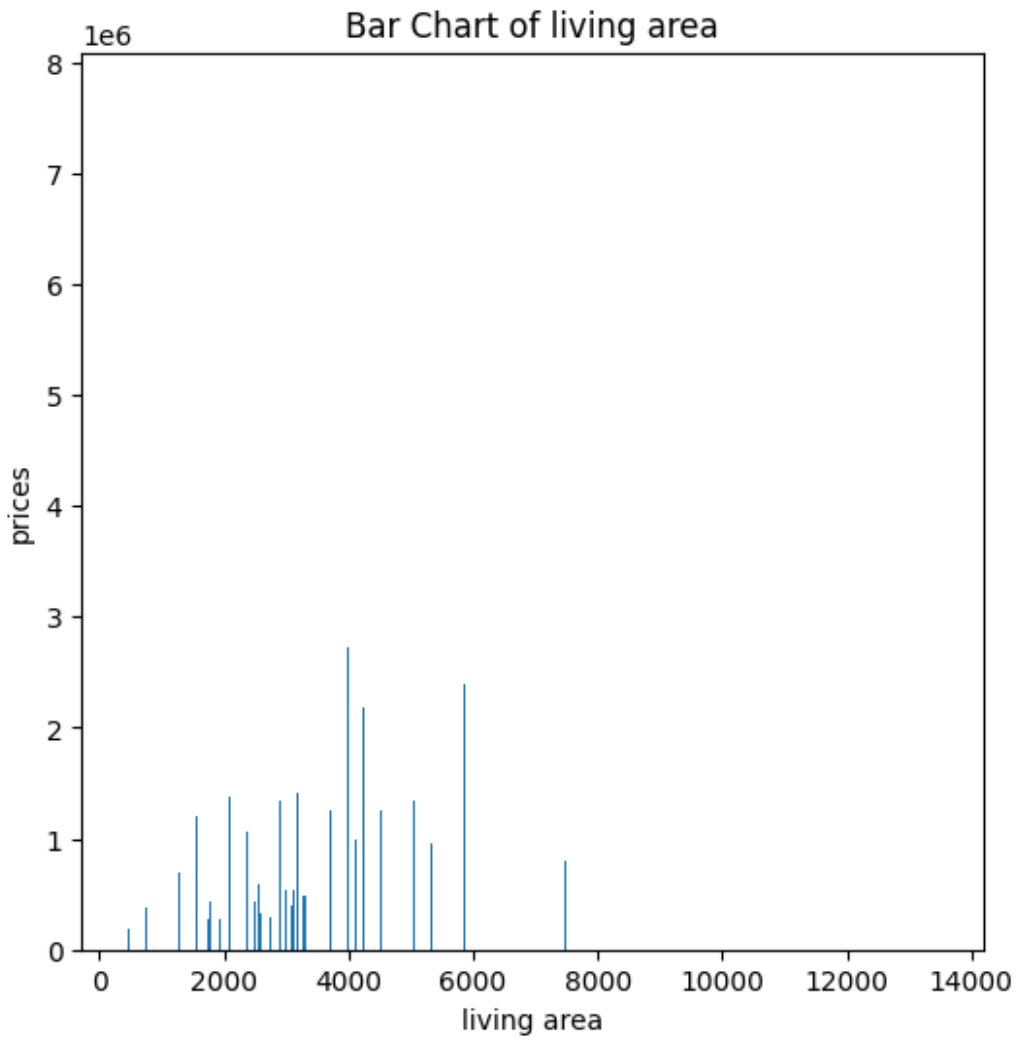

```
sns.distplot(df['living area'])
```

```
<ipython-input-8-f6cb9bd3998b>:1: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn
v0.14.0.

Please adapt your code to use either `displot` (a figure-level
function with
similar flexibility) or `histplot` (an axes-level function for
histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df['living area'])

<Axes: xlabel='living area', ylabel='Density'>
```

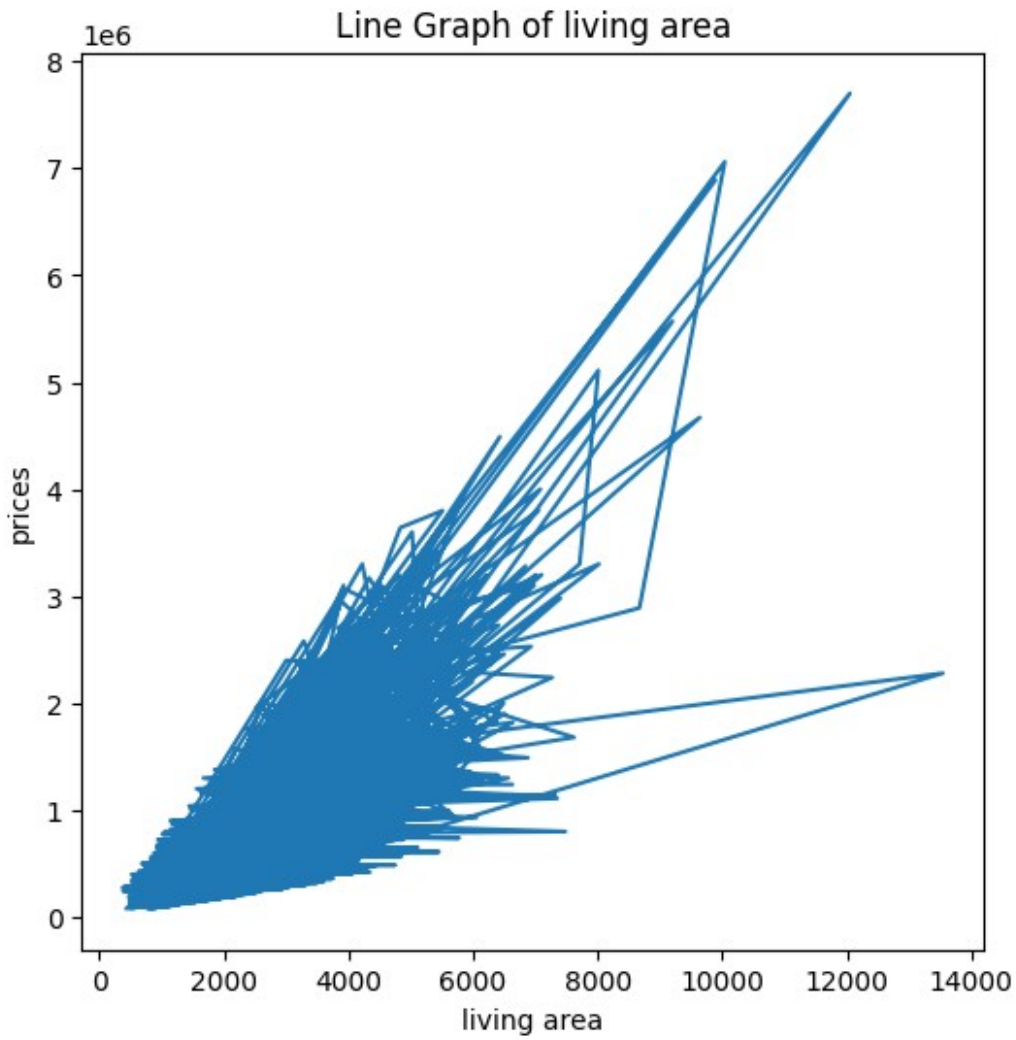


Bivariate Analysis

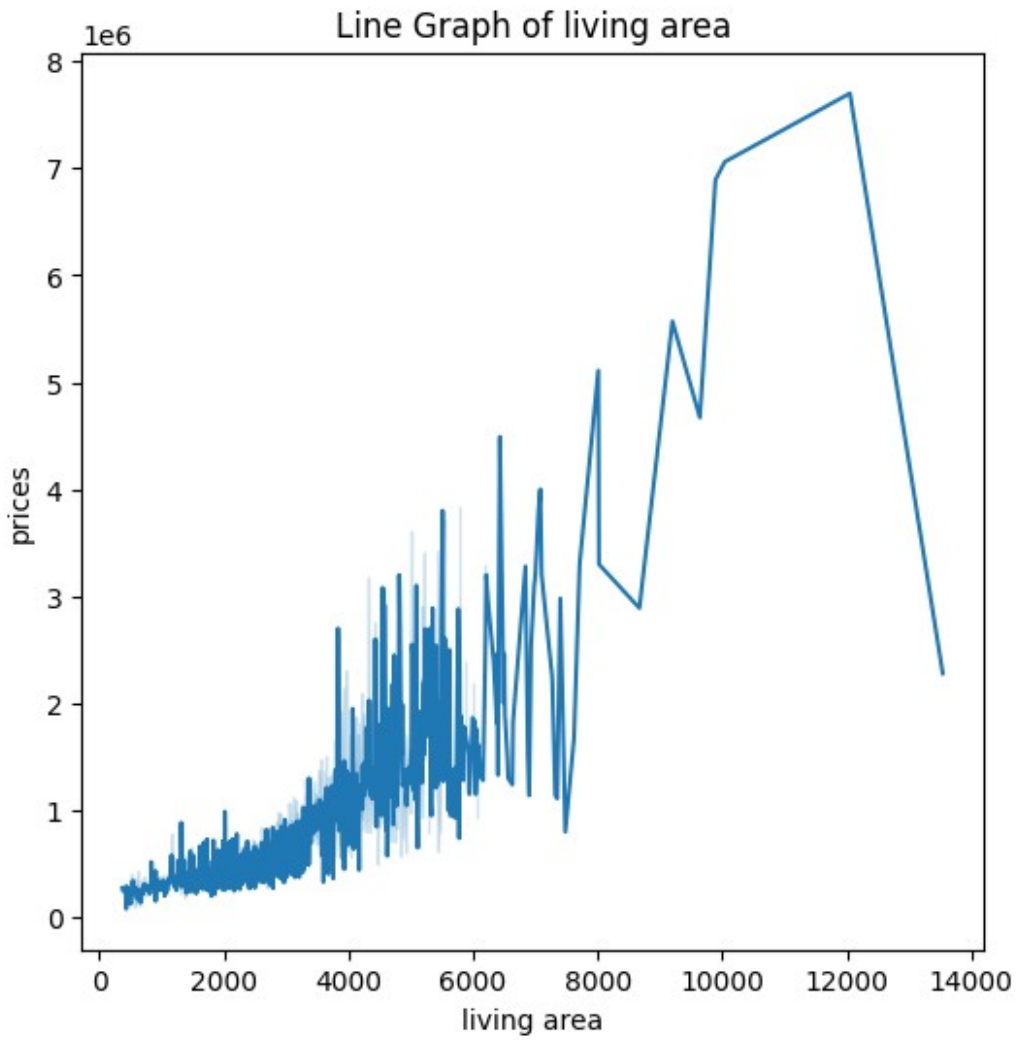
```
#Barplot of df [Comparison between 'living area' feature and 'Price']
plt.figure(figsize=(6, 6))
plt.bar(df['living area'], df['Price'])
plt.title('Bar Chart of living area')
plt.xlabel('living area')
plt.ylabel('prices')
plt.show()
```




```
#linegraph of df [Comparison between 'living area' feature and
'Price']
plt.figure(figsize=(6, 6))
plt.plot(df['living area'], df['Price'])
plt.title('Line Graph of living area')
plt.xlabel('living area')
plt.ylabel('prices')
plt.show()
```

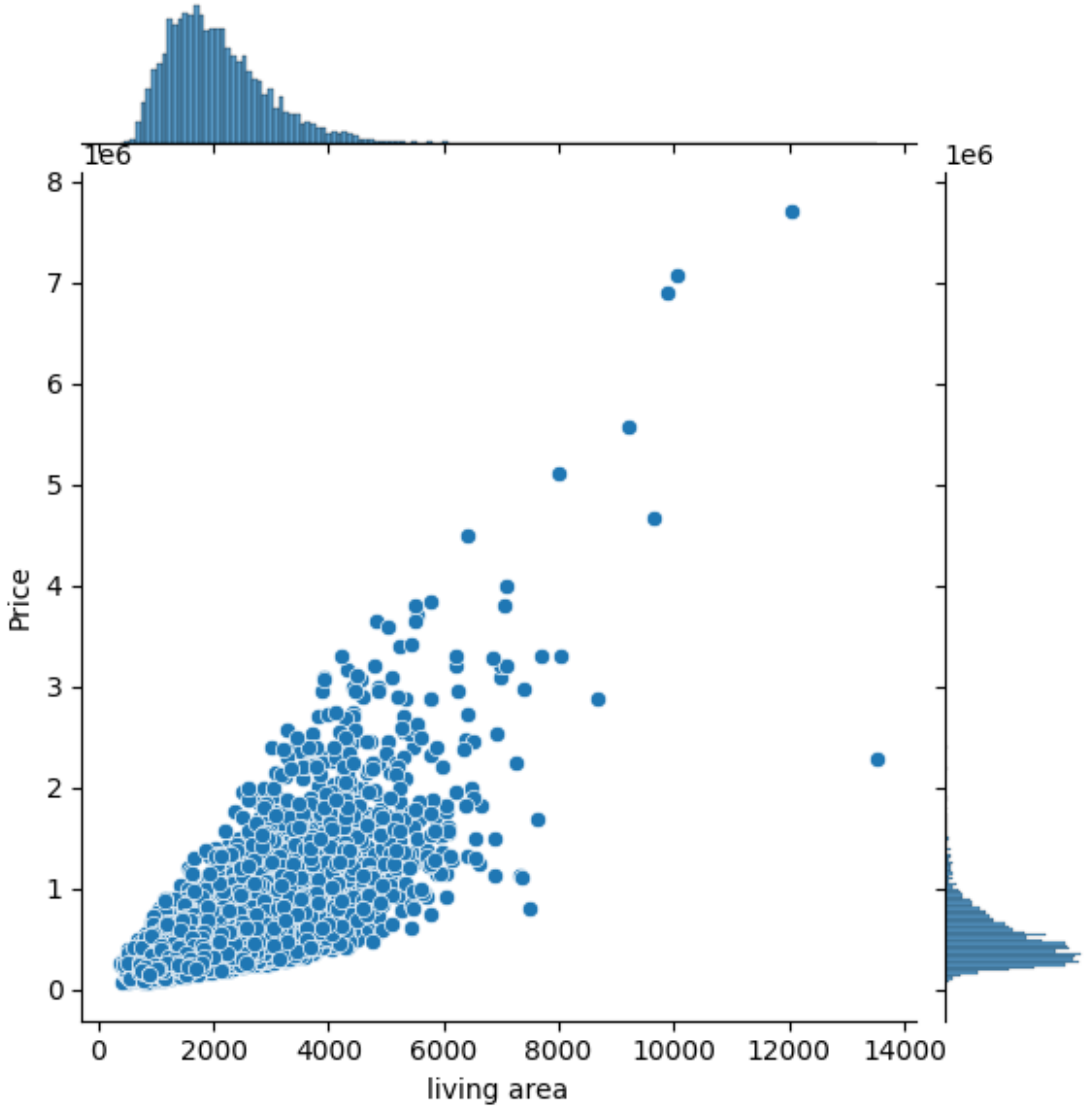



```
#lineplot of df [Comparison between 'living area' feature and 'Price']
plt.figure(figsize=(6, 6))
sns.lineplot(x= df['living area'], y= df['Price'])
plt.title('Line Graph of living area')
plt.xlabel('living area')
plt.ylabel('prices')
plt.show()
```




```
#Satterplot of df [Comparison between 'living area' feature and
'Price']
plt.figure(figsize=(6, 6))
plt.scatter(df['living area'], df['Price'])
plt.title('Scatter Plot of living area')
plt.xlabel('living area')
plt.ylabel('prices')
plt.show()
```

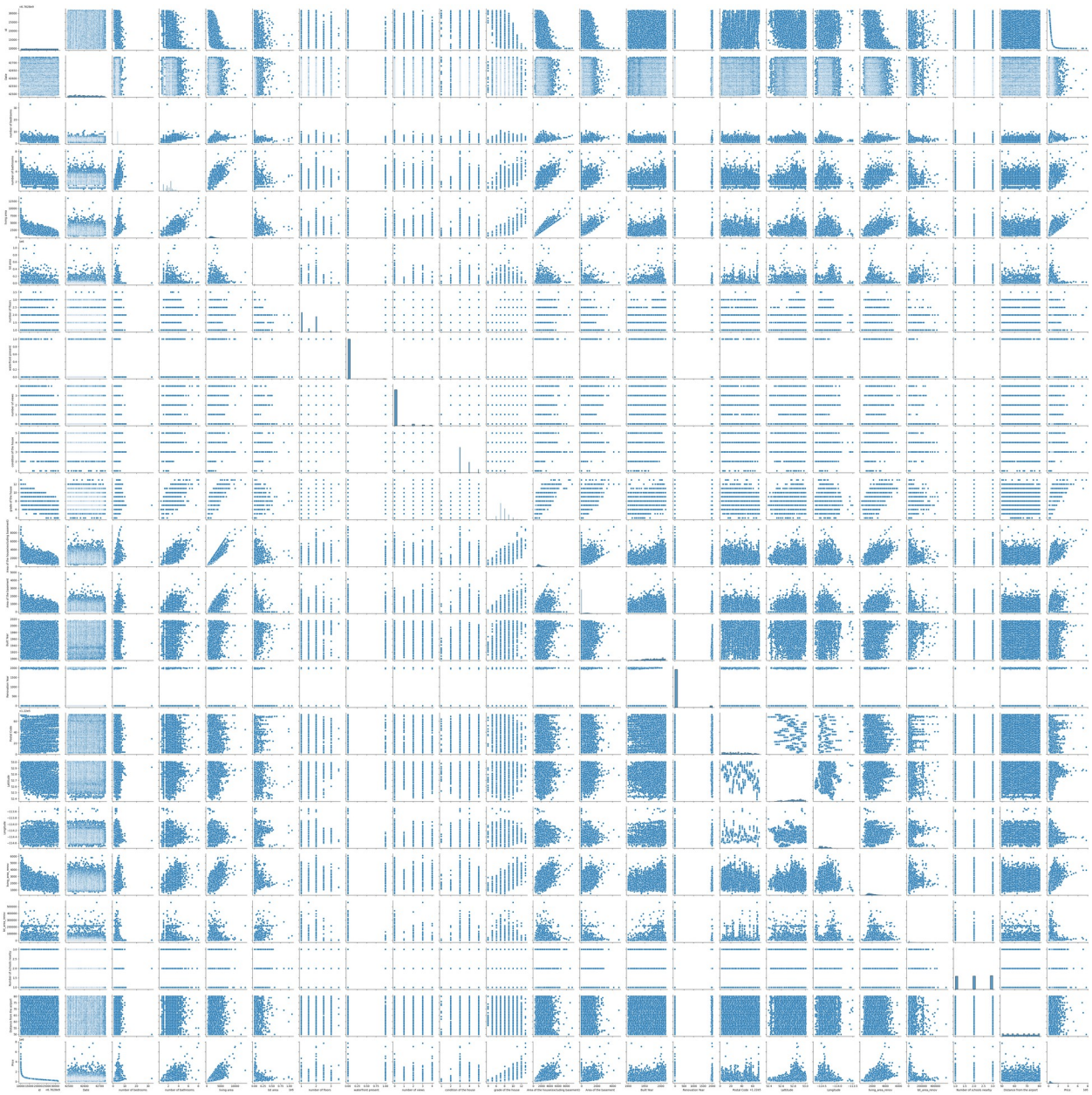

```
# JointPlot [Comparison between 'living area' feature and 'Price']
sns.jointplot(x='living area',y= 'Price',data =df)
<seaborn.axisgrid.JointGrid at 0x7e171bd75d80>
```



Multivariate Analysis

```
#Pair Plot
sns.pairplot(df)
```

```
<seaborn.axisgrid.PairGrid at 0x7e170055edd0>
```



```
df.shape
```

```
(14620, 23)
```

```
df.corr()
```

```
                                                id      Date  number of
bedrooms  \
id                                        1.000000  0.045966           -
0.329034
Date                                      0.045966  1.000000           -
0.015663
number of bedrooms                       -0.329034 -0.015663
```

```
1.000000
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                         -0.312305 -0.010335
0.177294
waterfront present                       -0.112937  0.012006           -
0.006257
number of views                          -0.293004 -0.004782
0.078665
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
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
0.029400
Number of schools nearby                 -0.004821 -0.004071
0.003397
Distance from the airport                -0.004542  0.011457           -
0.006157
Price                                    -0.773114 -0.027919
0.308460

                                          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
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                               0.287190
0.441491
Built Year                                         0.498127
0.309602
Renovation Year                                    0.049669
0.059400
Postal Code                                       -0.105546    -
0.080303
Lattitude                                          0.031156
0.054518
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
living area                             0.174420          0.354743
lot area                                1.000000         -0.004138
```

```
number of floors                          -0.004138          1.000000
waterfront present                         0.026282          0.016316
number of views                            0.078308          0.020153
condition of the house                    -0.008548         -0.269928
grade of the house                         0.110546          0.463082
Area of the house(excluding basement)      0.183553          0.525643
Area of the basement                       0.019755         -0.242976
Built Year                                 0.051615          0.481565
Renovation Year                            0.006848          0.006705
Postal Code                                0.070131         -0.129788
Lattitude                                 -0.090983          0.050731
Longitude                                  0.221432          0.127550
living_area_renov                          0.149744          0.285093
lot_area_renov                             0.706812         -0.010120
Number of schools nearby                  -0.012671         -0.007579
Distance from the airport                  0.003291          0.016567
Price                                      0.081992          0.262732

                                          waterfront present  number of
views  \
id                                                -0.112937        -
0.293004
Date                                               0.012006        -
0.004782
number of bedrooms                                -0.006257
0.078665
number of bathrooms                                0.060104
0.183789
living area                                        0.105837
0.287728
lot area                                           0.026282
0.078308
number of floors                                   0.016316
0.020153
waterfront present                                 1.000000
0.400206
number of views                                    0.400206
1.000000
condition of the house                             0.018644
0.052533
grade of the house                                 0.079831
0.254532
Area of the house(excluding basement)              0.071865
0.162672
Area of the basement                               0.085441
0.293062
Built Year                                        -0.024226        -
0.055357
Renovation Year                                    0.085865
```

```
0.102944
Postal Code                                     0.038318
0.039268
Lattitude                                      -0.021795        -
0.004555
Longitude                                      -0.047791        -
0.079706
living_area_renov                               0.085743
0.281452
lot_area_renov                                  0.032055
0.072300
Number of schools nearby                        0.001563
0.008004
Distance from the airport                       0.001448        -
0.001657
Price                                           0.263687
0.395973

                                    condition of the house  ...  \
id                                              -0.045061  ...
Date                                            -0.027402  ...
number of bedrooms                               0.026597  ...
number of bathrooms                             -0.128232  ...
living area                                     -0.063358  ...
lot area                                        -0.008548  ...
number of floors                                -0.269928  ...
waterfront present                               0.018644  ...
number of views                                  0.052533  ...
condition of the house                           1.000000  ...
grade of the house                              -0.152530  ...
Area of the house(excluding basement)           -0.167695  ...
Area of the basement                             0.180609  ...
Built Year                                      -0.381718  ...
Renovation Year                                 -0.062126  ...
Postal Code                                      0.045334  ...
Lattitude                                       -0.002998  ...
Longitude                                       -0.121189  ...
living_area_renov                               -0.099743  ...
lot_area_renov                                  -0.004748  ...
Number of schools nearby                        -0.006939  ...
Distance from the airport                       -0.002136  ...
Price                                            0.041376  ...

                                    Built Year  Renovation Year  \
id                                   -0.068645        -0.109155
Date                                 -0.005869        -0.011636
number of bedrooms                    0.152954         0.016132
number of bathrooms                   0.498127         0.049669
living area                           0.309602         0.059400
```

```
lot area                                 0.051615         0.006848
number of floors                         0.481565         0.006705
waterfront present                      -0.024226         0.085865
number of views                         -0.055357         0.102944
condition of the house                  -0.381718        -0.062126
grade of the house                       0.440358         0.014501
Area of the house(excluding basement)    0.419369         0.025727
Area of the basement                    -0.138843         0.075104
Built Year                               1.000000        -0.233683
Renovation Year                         -0.233683         1.000000
Postal Code                             -0.062349         0.018006
Lattitude                               -0.143153         0.028908
Longitude                                0.414591        -0.080050
living_area_renov                        0.328625        -0.002601
lot_area_renov                           0.072874         0.005869
Number of schools nearby                -0.001631        -0.000826
Distance from the airport               -0.003968         0.005342
Price                                    0.050307         0.133173

                                       Postal Code  Lattitude
Longitude  \
id                                       0.294709  -0.479334  -
0.070841
Date                                     0.018243  -0.023327  -
0.018231
number of bedrooms                      -0.044156  -0.013163
0.135712
number of bathrooms                     -0.105546   0.031156
0.223904
living area                             -0.080303   0.054518
0.240208
lot area                                 0.070131  -0.090983
0.221432
number of floors                        -0.129788   0.050731
0.127550
waterfront present                       0.038318  -0.021795  -
0.047791
number of views                          0.039268  -0.004555  -
0.079706
condition of the house                   0.045334  -0.002998  -
0.121189
grade of the house                      -0.146342   0.115256
0.203754
Area of the house(excluding basement)   -0.083730  -0.000088
0.345899
Area of the basement                    -0.010542   0.112989  -
0.145879
Built Year                              -0.062349  -0.143153
0.414591
```

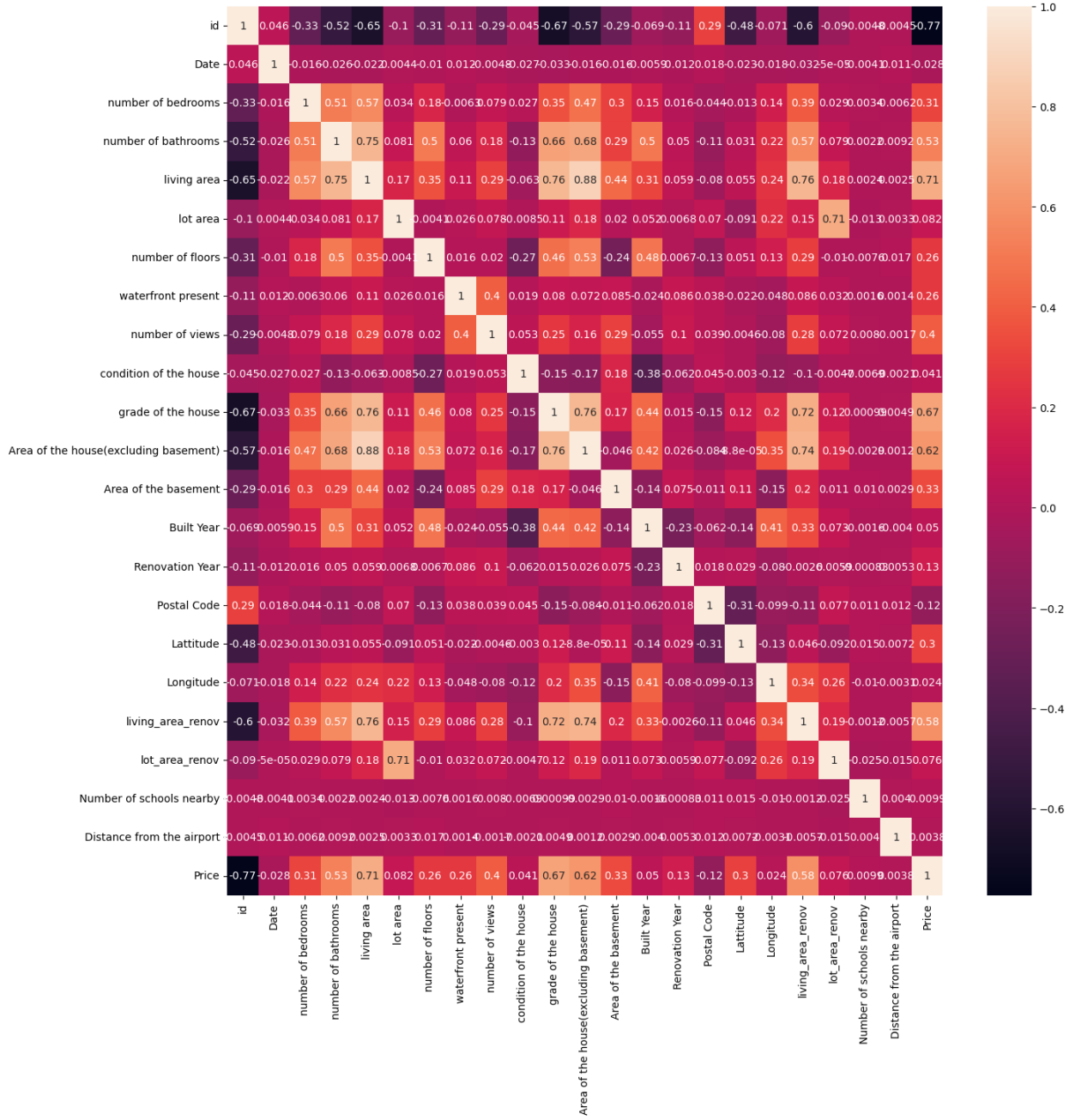
```
Renovation Year                              0.018006   0.028908  -
0.080050
Postal Code                                  1.000000  -0.310172  -
0.099003
Lattitude                                   -0.310172   1.000000  -
0.131472
Longitude                                   -0.099003  -0.131472
1.000000
living_area_renov                           -0.108454   0.046148
0.341221
lot_area_renov                               0.077483  -0.091622
0.258066
Number of schools nearby                     0.010605   0.014949  -
0.010163
Distance from the airport                    0.011528   0.007193  -
0.003100
Price                                       -0.115908   0.297490
0.024414

                                        living_area_renov
lot_area_renov  \
id                                                -0.599900       -
0.089604
Date                                              -0.032495       -
0.000050
number of bedrooms                                 0.389855
0.029400
number of bathrooms                                0.570530
0.078627
living area                                        0.757571
0.180312
lot area                                           0.149744
0.706812
number of floors                                   0.285093       -
0.010120
waterfront present                                 0.085743
0.032055
number of views                                    0.281452
0.072300
condition of the house                            -0.099743       -
0.004748
grade of the house                                 0.720019
0.116725
Area of the house(excluding basement)              0.737744
0.194670
Area of the basement                               0.196403
0.011283
Built Year                                         0.328625
0.072874
```

```
Renovation Year                                -0.002601
0.005869
Postal Code                                    -0.108454
0.077483
Lattitude                                       0.046148       -
0.091622
Longitude                                       0.341221
0.258066
living_area_renov                               1.000000
0.189225
lot_area_renov                                  0.189225
1.000000
Number of schools nearby                       -0.001203       -
0.025014
Distance from the airport                      -0.005673       -
0.014587
Price                                           0.584924
0.075535

                                       Number of schools nearby  \
id                                                    -0.004821
Date                                                  -0.004071
number of bedrooms                                     0.003397
number of bathrooms                                    0.002180
living area                                            0.002370
lot area                                              -0.012671
number of floors                                      -0.007579
waterfront present                                     0.001563
number of views                                        0.008004
condition of the house                                -0.006939
grade of the house                                     0.000986
Area of the house(excluding basement)                 -0.002894
Area of the basement                                   0.010284
Built Year                                            -0.001631
Renovation Year                                       -0.000826
Postal Code                                            0.010605
Lattitude                                              0.014949
Longitude                                             -0.010163
living_area_renov                                     -0.001203
lot_area_renov                                        -0.025014
Number of schools nearby                               1.000000
Distance from the airport                              0.004035
Price                                                  0.009890

                                       Distance from the airport
Price
id                                                    -0.004542 -
0.773114
Date                                                   0.011457 -
```

```
0.027919
number of bedrooms                                -0.006157
0.308460
number of bathrooms                                0.009206
0.531735
living area                                        0.002511
0.712169
lot area                                           0.003291
0.081992
number of floors                                   0.016567
0.262732
waterfront present                                 0.001448
0.263687
number of views                                   -0.001657
0.395973
condition of the house                            -0.002136
0.041376
grade of the house                                 0.004940
0.671814
Area of the house(excluding basement)              0.001222
0.615220
Area of the basement                               0.002926
0.330202
Built Year                                        -0.003968
0.050307
Renovation Year                                    0.005342
0.133173
Postal Code                                        0.011528 -
0.115908
Lattitude                                          0.007193
0.297490
Longitude                                         -0.003100
0.024414
living_area_renov                                 -0.005673
0.584924
lot_area_renov                                    -0.014587
0.075535
Number of schools nearby                           0.004035
0.009890
Distance from the airport                          1.000000
0.003804
Price                                              0.003804
1.000000

[23 rows x 23 columns]
```

```
plt.figure(figsize=(15, 15))
sns.heatmap(df.corr(),annot=True)
```

```
<Axes: >
```



Description

```
df.describe()
```

```
                 id          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
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
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
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
max    13540.000000  1.074218e+06          3.500000
1.000000

       number of views  condition of the house  ...    Built Year  \
count     14620.000000            14620.000000  ...  14620.000000
mean          0.233105                3.430506  ...   1970.926402
std           0.766259                0.664151  ...     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
mean         90.924008  122033.062244     52.792848   -114.404007
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
50%           0.000000  122032.000000     52.806400   -114.421000
75%           0.000000  122048.000000     52.908900   -114.315000
max        2015.000000  122072.000000     53.007600   -113.505000
```

```
       living_area_renov  lot_area_renov  Number of schools nearby  \
count       14620.000000    14620.000000              14620.000000
mean         1996.702257    12753.500068                  2.012244
std           691.093366    26058.414467                  0.817284
min           460.000000      651.000000                  1.000000
25%          1490.000000     5097.750000                  1.000000
50%          1850.000000     7620.000000                  2.000000
75%          2380.000000    10125.000000                  3.000000
max          6110.000000   560617.000000                  3.000000

       Distance from the airport         Price
count               14620.000000  1.462000e+04
mean                   64.950958  5.389322e+05
std                     8.936008  3.675324e+05
min                    50.000000  7.800000e+04
25%                    57.000000  3.200000e+05
50%                    65.000000  4.500000e+05
75%                    73.000000  6.450000e+05
max                    80.000000  7.700000e+06

[8 rows x 23 columns]
```

Handling the missing values

```
null_values = df.isnull().sum()
null_values
```

```
id                                          0
Date                                        0
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
Built Year                                  0
Renovation Year                             0
Postal Code                                 0
Lattitude                                   0
Longitude                                   0
living_area_renov                           0
lot_area_renov                              0
Number of schools nearby                    0
```

```
Distance from the airport                0
Price                                    0
dtype: int64
```

```
# Identify columns with missing values
missing_values = df.columns[df.isnull().any()]
missing_values
```

```
Index([], dtype='object')
```

```
# Calculate the median for each column with missing values and store
it in a dictionary
median_values = {}
for column in missing_values:
    median = df[column].median()
    median_values[column] = median

median_values
df.fillna(median_values, inplace=True)
```