# Assignment

In [12]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

***Loading The dataset***

In [3]:
```
dt=pd.read_csv("C:\Users\VINITH\Downloads\House Price India.csv")
dt
```

Out[3]:

| | id | Date | number of bedrooms | number of bathrooms | living area | lot area | number of floors | waterfront present | number of views | condition of the house | ... |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 6762810145 | 42491 | 5 | 2.50 | 3650 | 9050 | 2.0 | 0 | 4 | 5 | ... |
| **1** | 6762810635 | 42491 | 4 | 2.50 | 2920 | 4000 | 1.5 | 0 | 0 | 5 | ... |
| **2** | 6762810998 | 42491 | 5 | 2.75 | 2910 | 9480 | 1.5 | 0 | 0 | 3 | ... |
| **3** | 6762812605 | 42491 | 4 | 2.50 | 3310 | 42998 | 2.0 | 0 | 0 | 3 | ... |
| **4** | 6762812919 | 42491 | 3 | 2.00 | 2710 | 4500 | 1.5 | 0 | 0 | 4 | ... |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **14615** | 6762830250 | 42734 | 2 | 1.50 | 1556 | 20000 | 1.0 | 0 | 0 | 4 | ... |
| **14616** | 6762830339 | 42734 | 3 | 2.00 | 1680 | 7000 | 1.5 | 0 | 0 | 4 | ... |
| **14617** | 6762830618 | 42734 | 2 | 1.00 | 1070 | 6120 | 1.0 | 0 | 0 | 3 | ... |
| **14618** | 6762830709 | 42734 | 4 | 1.00 | 1030 | 6621 | 1.0 | 0 | 0 | 4 | ... |
| **14619** | 6762831463 | 42734 | 3 | 1.00 | 900 | 4770 | 1.0 | 0 | 0 | 3 | ... |

14620 rows × 23 columns

**Univariate Analysis**

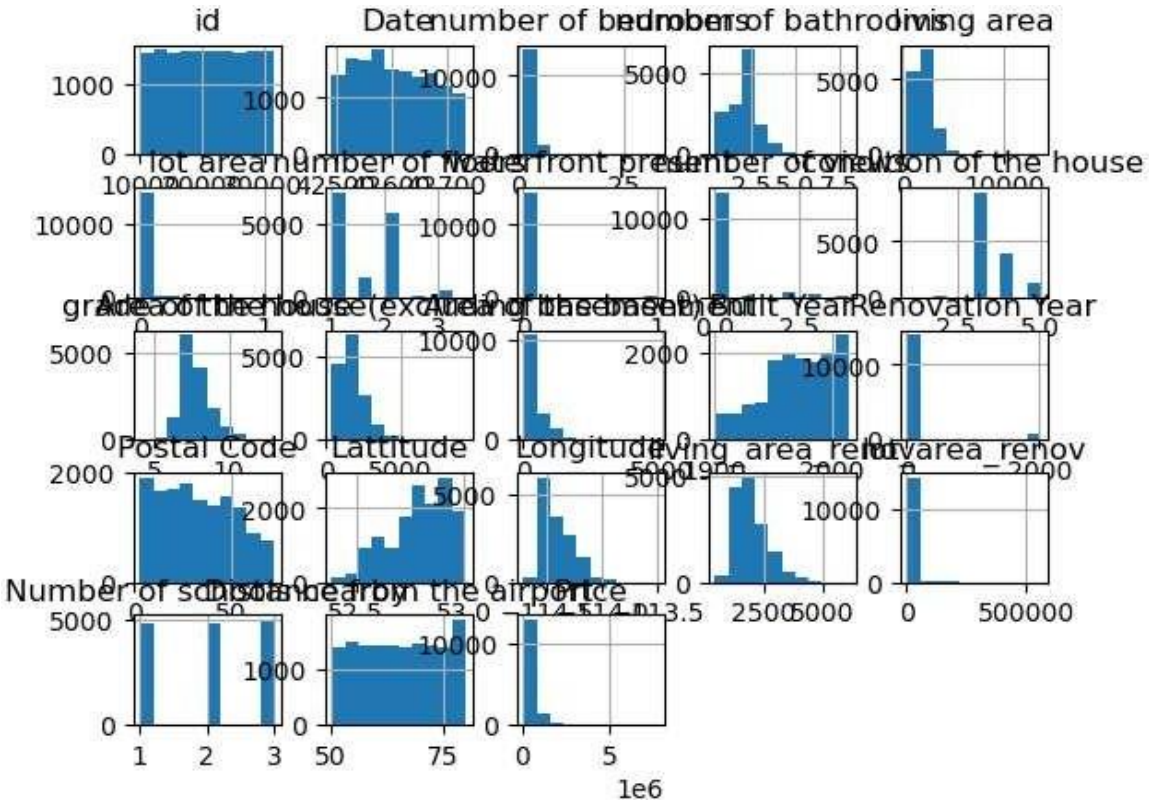
In [5]:
```
dt.describe()
```

Out[5]:

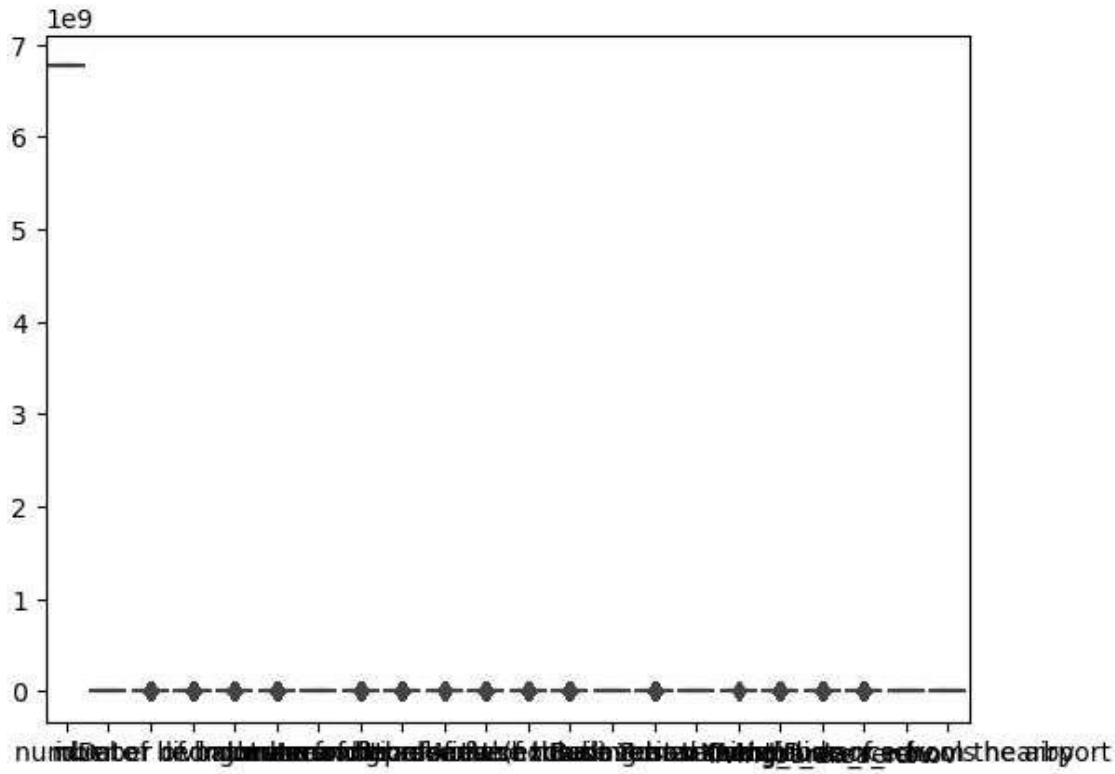
| | id | Date | number of bedrooms | number of bathrooms | living area | lot area | number of floors | |
|---|---|---|---|---|---|---|---|---|
| **count** | 1.462000e+04 | 14620.000000 | 14620.000000 | 14620.000000 | 14620.000000 | 1.462000e+04 | 14620.000000 | 1 |
| **mean** | 6.762821e+09 | 42604.538646 | 3.379343 | 2.129583 | 2098.262996 | 1.509328e+04 | 1.502360 | |
| **std** | 6.237575e+03 | 67.347991 | 0.938719 | 0.769934 | 928.275721 | 3.791962e+04 | 0.540239 | |
| **min** | 6.762810e+09 | 42491.000000 | 1.000000 | 0.500000 | 370.000000 | 5.200000e+02 | 1.000000 | |
| **25%** | 6.762815e+09 | 42546.000000 | 3.000000 | 1.750000 | 1440.000000 | 5.010750e+03 | 1.000000 | |
| **50%** | 6.762821e+09 | 42600.000000 | 3.000000 | 2.250000 | 1930.000000 | 7.620000e+03 | 1.500000 | |
| **75%** | 6.762826e+09 | 42662.000000 | 4.000000 | 2.500000 | 2570.000000 | 1.080000e+04 | 2.000000 | |
| **max** | 6.762832e+09 | 42734.000000 | 33.000000 | 8.000000 | 13540.000000 | 1.074218e+06 | 3.500000 | |

8 rows × 23 columns

In [8]:
```
dt.hist()
plt.show()
```



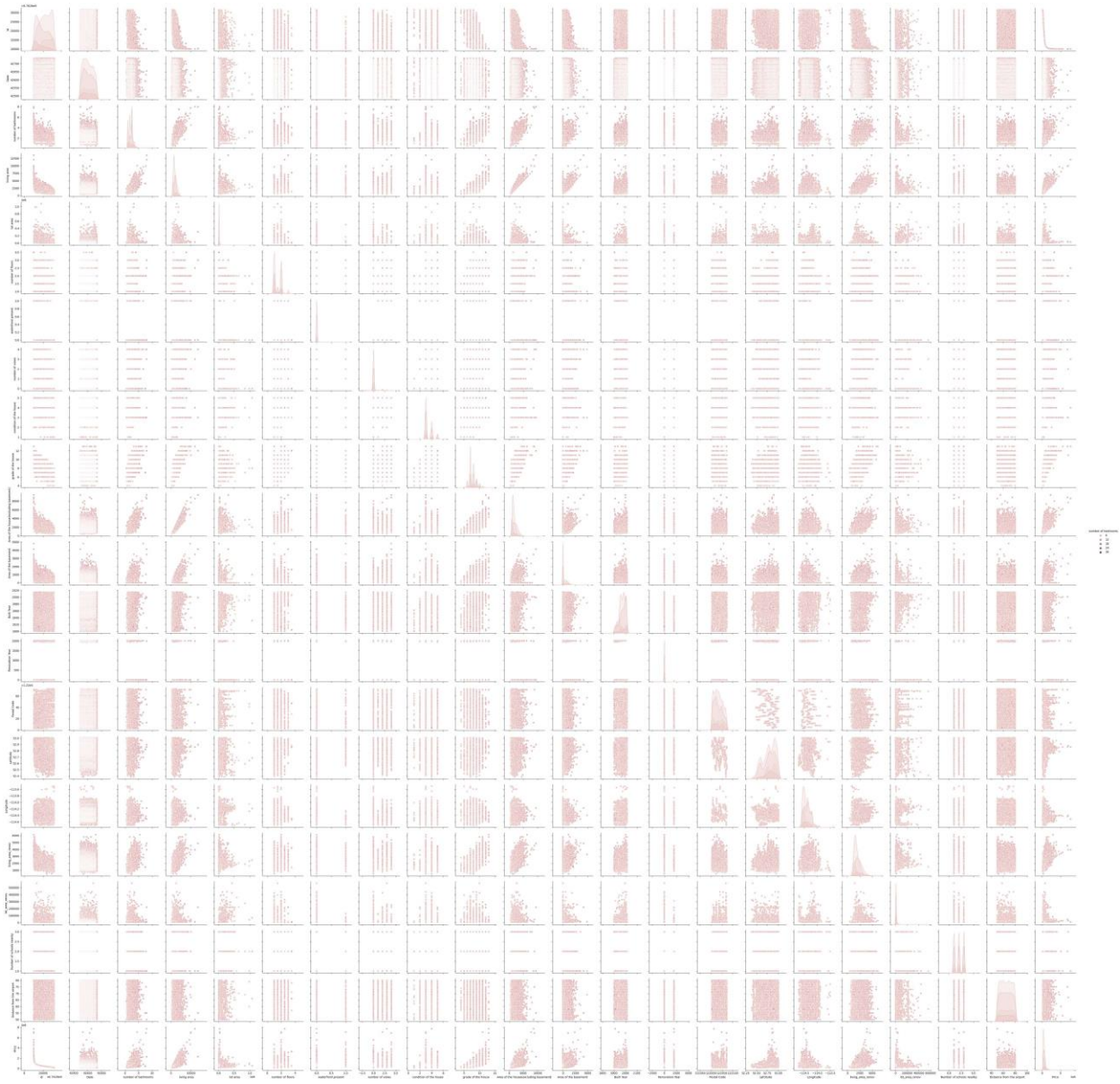
In [9]:
```
sns.boxplot(data=dt.iloc[:, :-1])
plt.show()
```



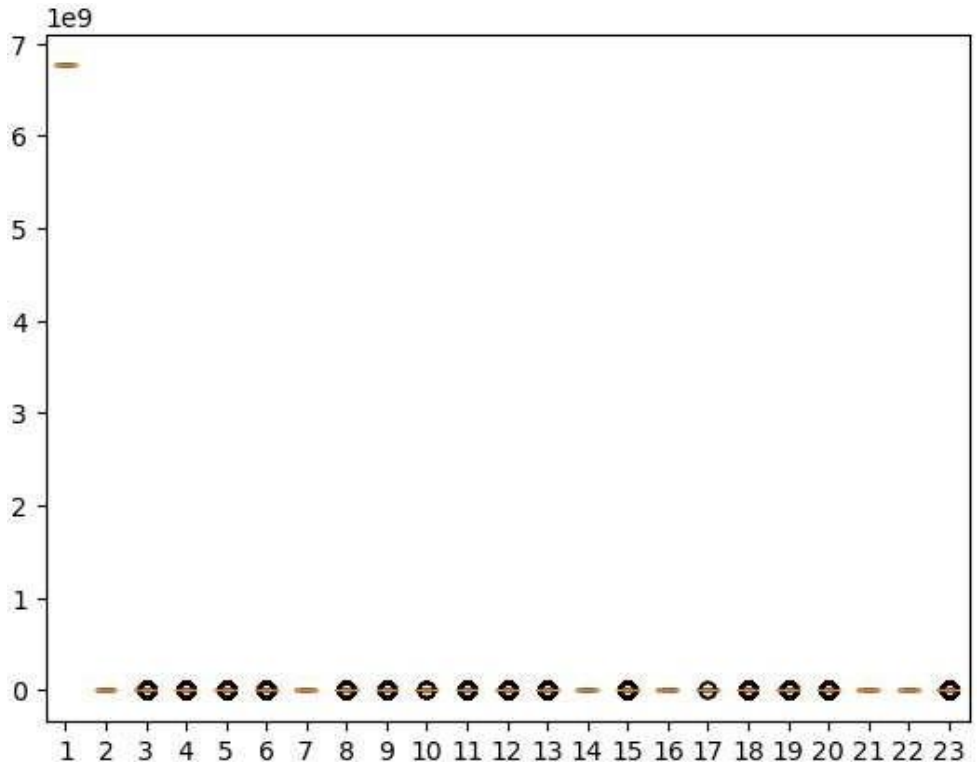

**Bivariate Analysis**

In [14]:
```
visual=sns.pairplot(dt,hue="number of bedrooms")
print(visual)
```

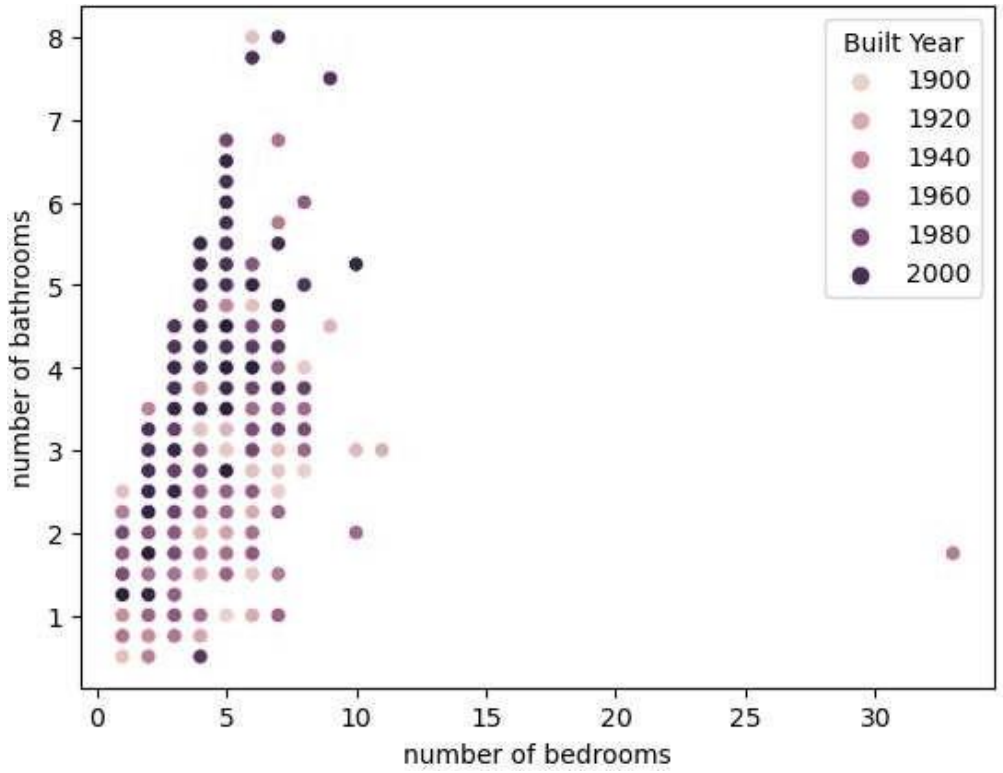
```
<seaborn.axisgrid.PairGrid object at 0x000001C3D46F8B80>
```



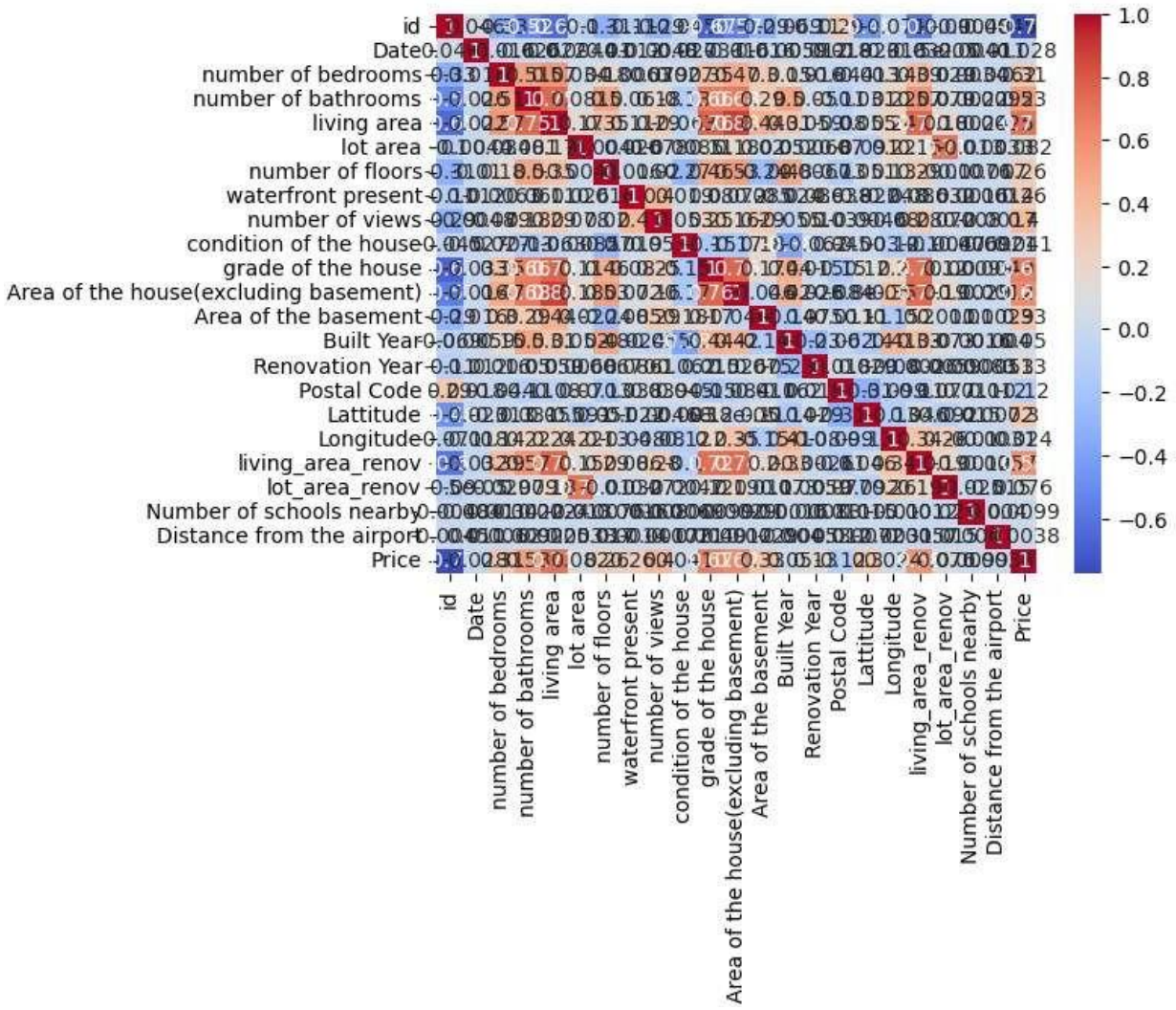
In [16]:
```
plt.boxplot(dt)
plt.show()
```



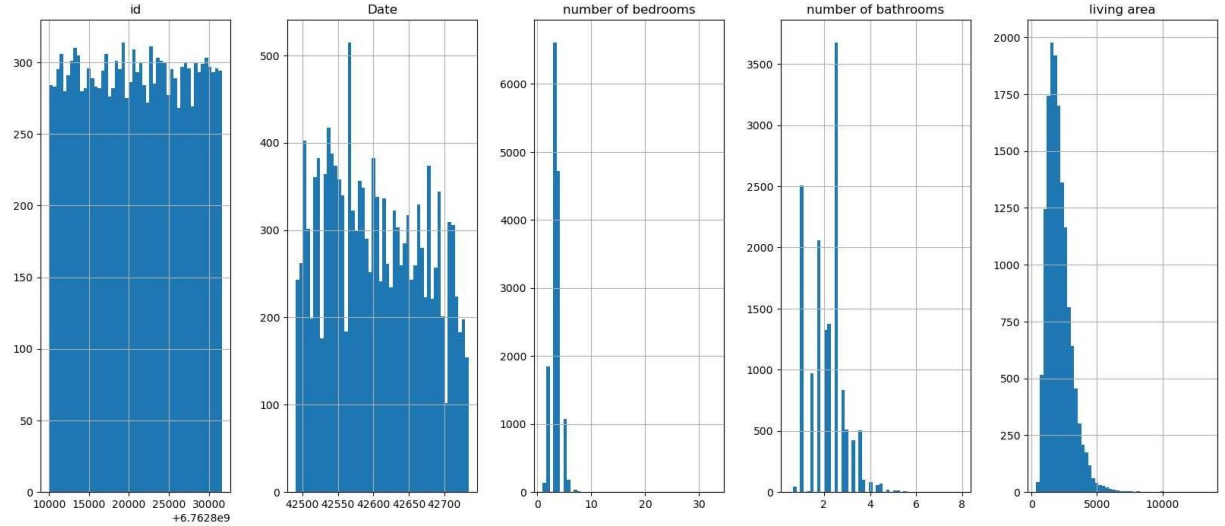

In [17]:
```
sns.scatterplot(x='number of bedrooms', y='number of bathrooms', data=dt, hue='Built Yea
plt.show()
```

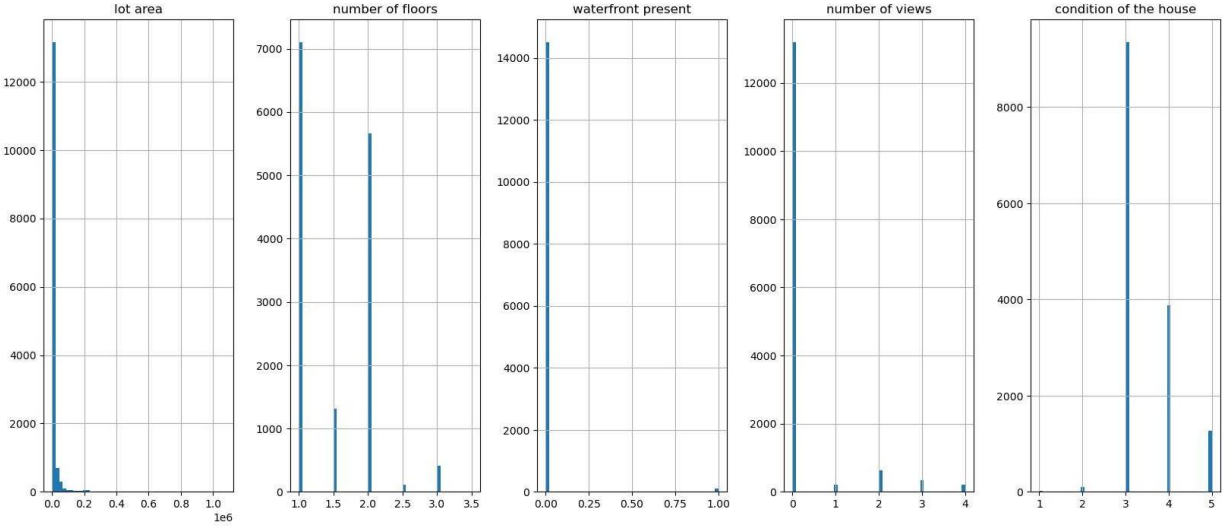


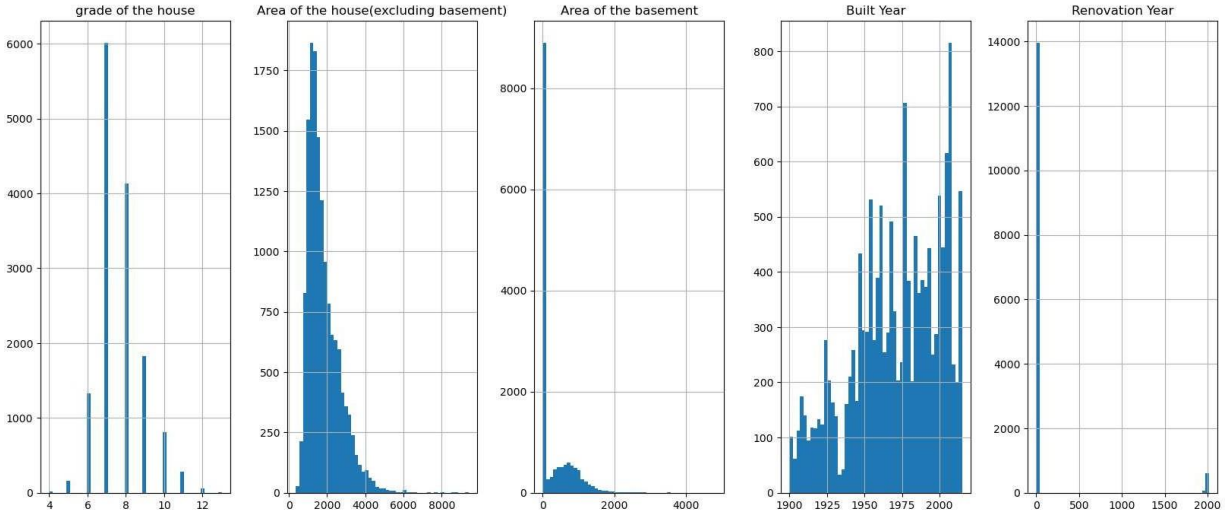
In [18]:
```
dt_corr = dt.corr()
sns.heatmap(dt_corr, annot=True, cmap='coolwarm')
plt.show()
```

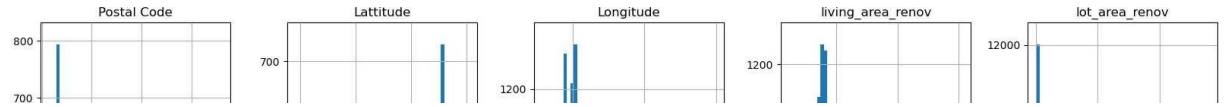


In [22]:
```
dt.hist(bins=50,figsize=(20,50));
```


id
Date
number of bedrooms
number of bathrooms
living area

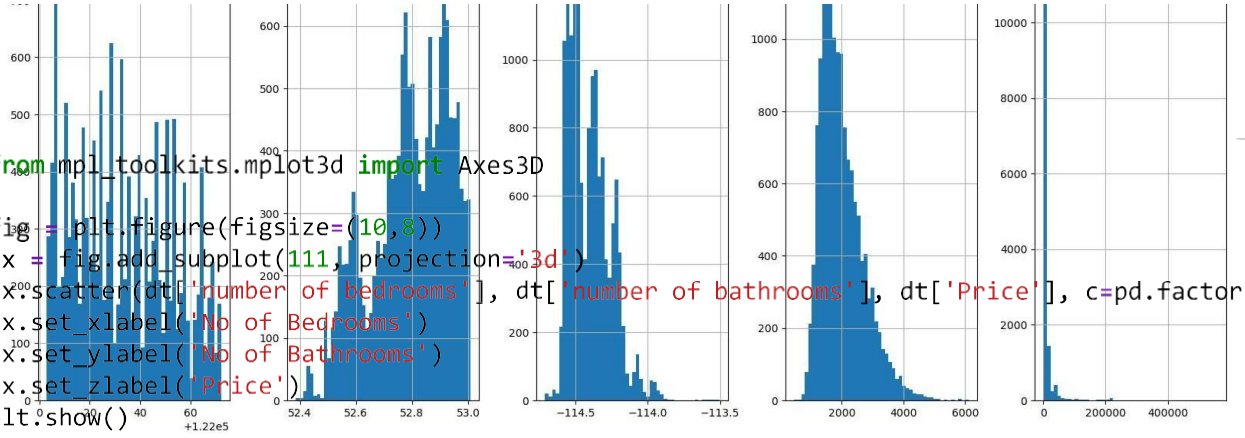

lot area
number of floors
waterfront present
number of views
condition of the house

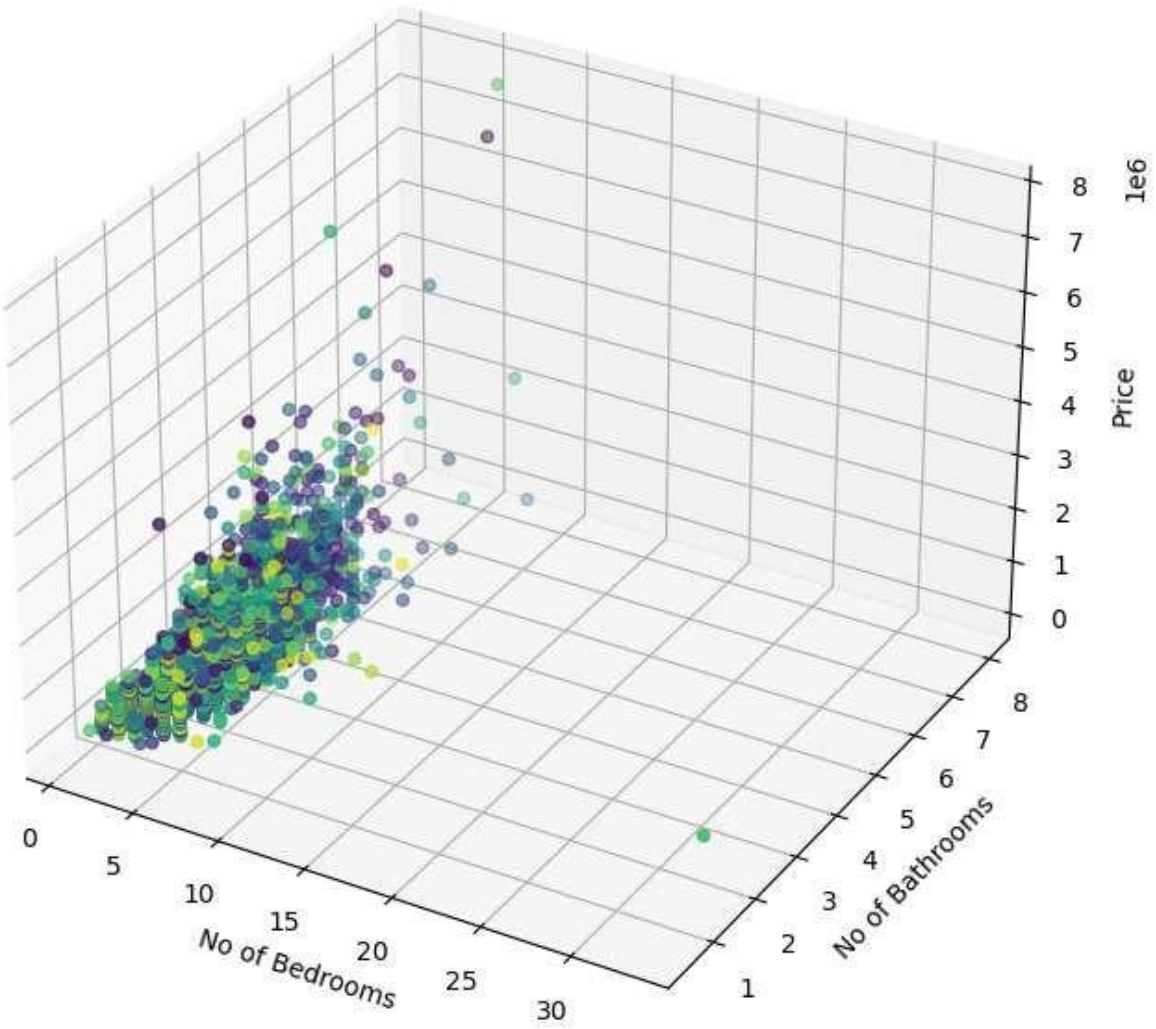

grade of the house
Area of the house(excluding basement)
Area of the basement
Built Year
Renovation Year


Postal Code
Lattitude
Longitude
living_area_renov
lot_area_renov

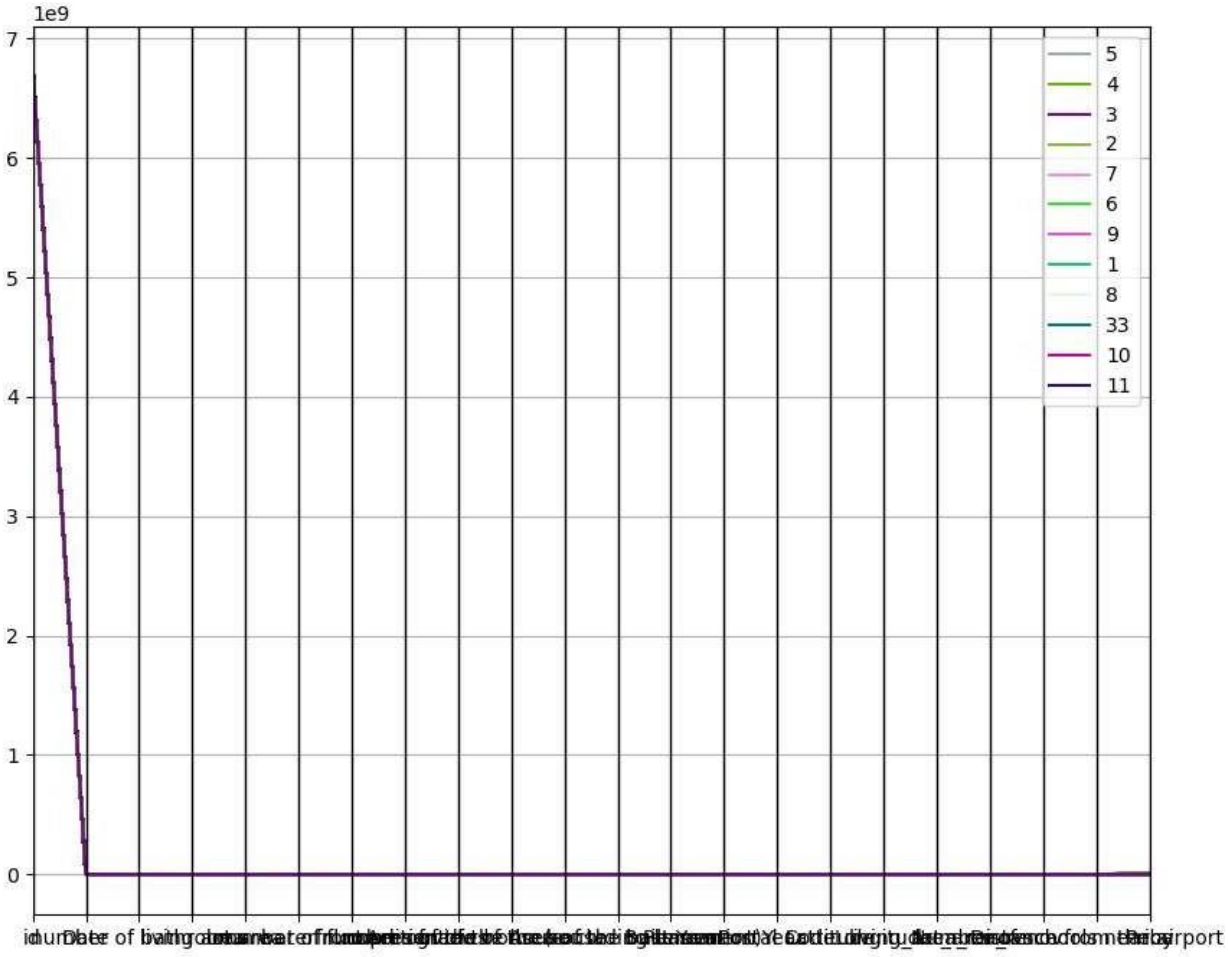
In [23]:
```
from mpl_toolkits.mplot3d import Axes3D

fig = plt.figure(figsize=(10,8))
ax = fig.add_subplot(111, projection='3d')
ax.scatter(dt['number of bedrooms'], dt['number of bathrooms'], dt['Price'], c=pd.factor
ax.set_xlabel('No of Bedrooms')
ax.set_ylabel('No of Bathrooms')
ax.set_zlabel('Price')
plt.show()
```

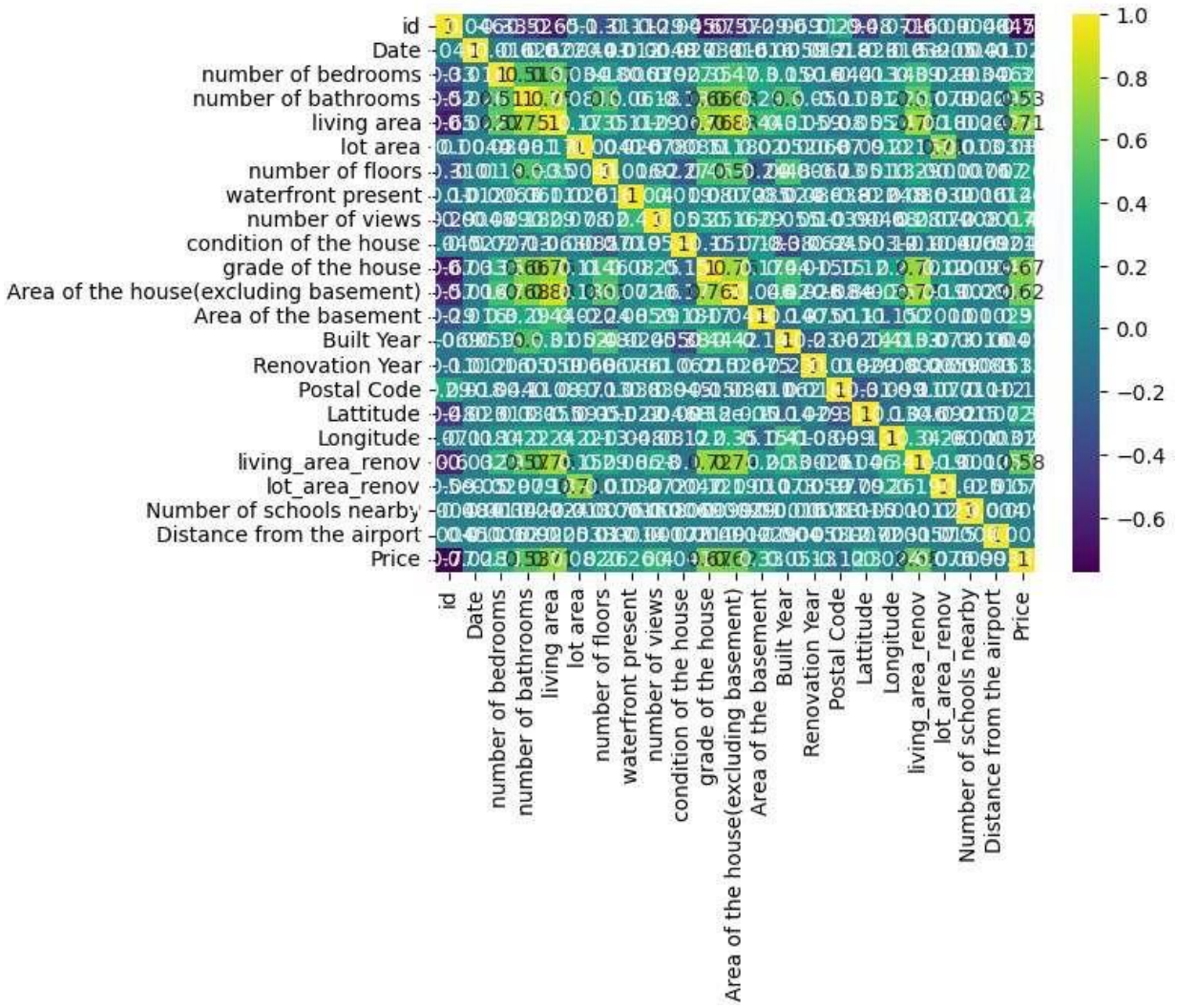




In [27]:
```
from pandas.plotting import parallel_coordinates
plt.figure(figsize=(10,8))
parallel_coordinates(dt, 'number of bedrooms')
plt.show()
```



In [28]:
```
dt_corr = dt.corr()
sns.heatmap(dt_corr, annot=True, cmap='viridis')
plt.show()
```




**Descriptive Analysis**

In [30]:
```
dt.describe()
```

Out[30]:

| | id | Date | number of bedrooms | number of bathrooms | living area | lot area | number of floors | |
|---|---|---|---|---|---|---|---|---|
| **count** | 1.462000e+04 | 14620.000000 | 14620.000000 | 14620.000000 | 14620.000000 | 1.462000e+04 | 14620.000000 | 1 |
| **mean** | 6.762821e+09 | 42604.538646 | 3.379343 | 2.129583 | 2098.262996 | 1.509328e+04 | 1.502360 | |
| **std** | 6.237575e+03 | 67.347991 | 0.938719 | 0.769934 | 928.275721 | 3.791962e+04 | 0.540239 | |
| **min** | 6.762810e+09 | 42491.000000 | 1.000000 | 0.500000 | 370.000000 | 5.200000e+02 | 1.000000 | |
| **25%** | 6.762815e+09 | 42546.000000 | 3.000000 | 1.750000 | 1440.000000 | 5.010750e+03 | 1.000000 | |
| **50%** | 6.762821e+09 | 42600.000000 | 3.000000 | 2.250000 | 1930.000000 | 7.620000e+03 | 1.500000 | |
| **75%** | 6.762826e+09 | 42662.000000 | 4.000000 | 2.500000 | 2570.000000 | 1.080000e+04 | 2.000000 | |
| **max** | 6.762832e+09 | 42734.000000 | 33.000000 | 8.000000 | 13540.000000 | 1.074218e+06 | 3.500000 | |

8 rows × 23 columns

In [31]:
```
dt.skew()
```

Out[31]:
```
id                                    -0.000802
Date                                   0.143747
number of bedrooms                     2.663257
number of bathrooms                    0.556663

living area                            1.538337
lot area                              10.155206
number of floors                       0.586158
waterfront present                    11.294672
number of views                        3.409219
condition of the house                 1.018018
grade of the house                     0.777584
Area of the house(excluding basement)  1.436446
Area of the basement                   1.609744
Built Year                            -0.472049
Renovation Year                        4.359764
Postal Code                            0.227735
Lattitude                             -0.523831
Longitude                              0.873803
living_area_renov                      1.081959
lot_area_renov                         7.774206
Number of schools nearby              -0.022519
Distance from the airport              0.006114
Price                                  4.269298
dtype: float64
```

In [32]:
```
dt.kurtosis()
```

Out[32]:
```
id                                      -1.201221
Date                                    -1.130823
number of bedrooms                      69.240310
number of bathrooms                      1.588195
living area                              6.073617
lot area                               164.757273
number of floors                        -0.523576
waterfront present                     125.586791
number of views                         10.968839
condition of the house                   0.351359
grade of the house                       1.048022
Area of the house(excluding basement)    3.402258
Area of the basement                     3.139635
Built Year                              -0.673474
Renovation Year                         17.011306
Postal Code                             -1.058364
Lattitude                               -0.619219
Longitude                                0.950315
living_area_renov                        1.428944
lot_area_renov                          79.360403
Number of schools nearby                -1.502552
Distance from the airport               -1.203048
Price                                   40.321918
dtype: float64
```

In [36]:
```
dt.groupby("number of bedrooms").max()
```

Out[36]:

| number of bedrooms | id | Date | number of bathrooms | living area | lot area | number of floors | waterfront present | number of views | condition of the house | grade of the house | .. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 6762831615 | 42733 | 2.50 | 3000 | 533610 | 3.0 | 1 | 4 | 5 | 9 | .. |
| **2** | 6762831616 | 42734 | 3.50 | 6840 | 982278 | 3.5 | 1 | 4 | 5 | 12 | .. |
| **3** | 6762831613 | 42734 | 4.50 | 6400 | 843309 | 3.5 | 1 | 4 | 5 | 13 | .. |
| **4** | 6762831588 | 42734 | 5.50 | 7620 | 982998 | 3.0 | 1 | 4 | 5 | 13 | .. |
| **5** | 6762831510 | 42734 | 6.75 | 10040 | 1074218 | 3.0 | 1 | 4 | 5 | 13 | .. |
| **6** | 6762831191 | 42734 | 8.00 | 12050 | 248600 | 3.0 | 1 | 4 | 5 | 13 | .. |
| **7** | 6762827935 | 42685 | 8.00 | 13540 | 307752 | 3.0 | 0 | 4 | 5 | 12 | .. |
| **8** | 6762825321 | 42722 | 6.00 | 7710 | 20666 | 3.5 | 0 | 3 | 5 | 12 | .. |
| **9** | 6762820817 | 42592 | 7.50 | 4050 | 6988 | 2.5 | 0 | 0 | 3 | 8 | .. |
| **10** | 6762815290 | 42732 | 5.25 | 4590 | 11914 | 2.0 | 0 | 2 | 4 | 9 | .. |
| **11** | 6762818607 | 42602 | 3.00 | 3000 | 4960 | 2.0 | 0 | 0 | 3 | 7 | .. |
| **33** | 6762815473 | 42545 | 1.75 | 1620 | 6000 | 1.0 | 0 | 0 | 5 | 7 | .. |

12 rows × 22 columns

In [37]:
```
dt.groupby("Built Year").max()
```

Out[37]:

| Built Year | id | Date | number of bedrooms | number of bathrooms | living area | lot area | number of floors | waterfront present | number of views | condition of the house | ... |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **1900** | 6762831268 | 42726 | 6 | 4.00 | 4380 | 262231 | 2.5 | 0 | 3 | 5 | ... |
| **1901** | 6762827764 | 42721 | 8 | 2.75 | 3440 | 7200 | 2.5 | 0 | 2 | 5 | ... |
| **1902** | 6762828814 | 42686 | 6 | 3.00 | 4480 | 6000 | 2.5 | 0 | 0 | 5 | ... |
| **1903** | 6762831286 | 42714 | 6 | 3.50 | 2800 | 46173 | 2.5 | 1 | 4 | 5 | ... |
| **1904** | 6762830724 | 42733 | 8 | 4.00 | 7710 | 47044 | 3.5 | 0 | 1 | 5 | ... |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **2011** | 6762829355 | 42729 | 5 | 4.00 | 5635 | 77832 | 3.0 | 0 | 3 | 3 | ... |
| **2012** | 6762831335 | 42726 | 6 | 4.50 | 4920 | 95950 | 3.0 | 0 | 3 | 3 | ... |
| **2013** | 6762831396 | 42726 | 7 | 5.00 | 5310 | 64441 | 3.0 | 0 | 3 | 3 | ... |
| **2014** | 6762831181 | 42734 | 6 | 5.00 | 5790 | 108865 | 3.0 | 1 | 4 | 3 | ... |
| **2015** | 6762829970 | 42734 | 5 | 4.00 | 4460 | 9240 | 3.0 | 0 | 2 | 3 | ... |

116 rows × 22 columns

In [38]:
```
dt.groupby("Built Year").mean()
```

Out[38]:

| Built Year | id | Date | number of bedrooms | number of bathrooms | living area | lot area | number of floors | waterfront present | o |
|---|---|---|---|---|---|---|---|---|---|
| **1900** | 6.762819e+09 | 42598.868852 | 3.311475 | 1.758197 | 1784.918033 | 12321.196721 | 1.483607 | 0.000000 | 0. |
| **1901** | 6.762819e+09 | 42592.761905 | 3.571429 | 1.535714 | 1825.476190 | 4377.238095 | 1.619048 | 0.000000 | 0. |
| **1902** | 6.762817e+09 | 42574.650000 | 3.700000 | 2.087500 | 2075.500000 | 4342.900000 | 1.825000 | 0.000000 | 0. |
| **1903** | 6.762821e+09 | 42609.363636 | 3.212121 | 1.613636 | 1596.848485 | 6999.424242 | 1.484848 | 0.030303 | 0. |
| **1904** | 6.762820e+09 | 42597.250000 | 3.000000 | 1.571429 | 1740.464286 | 6149.535714 | 1.357143 | 0.000000 | 0. |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | |
| **2011** | 6.762820e+09 | 42604.408163 | 3.469388 | 2.660714 | 2342.326531 | 6385.693878 | 2.000000 | 0.000000 | 0. |
| **2012** | 6.762821e+09 | 42601.970874 | 3.543689 | 2.645631 | 2395.242718 | 6329.757282 | 1.980583 | 0.000000 | 0. |
| **2013** | 6.762818e+09 | 42592.592308 | 3.923077 | 2.869231 | 2691.600000 | 7792.353846 | 2.000000 | 0.000000 | 0. |
| **2014** | 6.762818e+09 | 42617.628713 | 3.745050 | 2.729579 | 2634.391089 | 5566.292079 | 2.123762 | 0.004950 | 0. |
| **2015** | 6.762821e+09 | 42598.000000 | 3.250000 | 2.416667 | 2195.833333 | 3899.333333 | 2.291667 | 0.000000 | 0. |

116 rows × 22 columns

In [39]:
```
dt["Built Year"].value_counts()
```

Out[39]:
```
2014    404
2005    319
2006    300
2004    296
2003    295
       ...
1902     20
1935     18
1933     17
1934     15
2015     12
Name: Built Year, Length: 116, dtype: int64
```

In [42]:
```
dt.groupby("number of bedrooms").agg({'Built Year':'max'})
```

Out[42]:

| number of bedrooms | Built Year |
|---|---|
| 1 | 2015 |
| 2 | 2015 |
| 3 | 2015 |
| 4 | 2015 |
| 5 | 2015 |
| 6 | 2014 |
| 7 | 2013 |
| 8 | 1997 |
| 9 | 1996 |
| 10 | 2008 |
| 11 | 1918 |
| 33 | 1947 |

In [44]:
```
dt.groupby("number of bedrooms").agg({'Built Year':'min'})
```

Out[44]:

| number of bedrooms | Built Year |
|---|---|
| 1 | 1900 |
| 2 | 1900 |
| 3 | 1900 |
| 4 | 1900 |
| 5 | 1900 |
| 6 | 1900 |
| 7 | 1901 |
| 8 | 1901 |
| 9 | 1915 |
| 10 | 1913 |
| 11 | 1918 |
| 33 | 1947 |

### Handling Missing Values

In [45]:
```
dt.duplicated()
```

Out[45]:
```
0        False
1        False
2        False
3        False
4        False
         ...
14615    False
14616    False
14617    False
14618    False
Length: 14620, dtype: bool
```

In [46]:
```
dt.isna()
```

Out[46]:

| | id | Date | number of bedrooms | number of bathrooms | living area | lot area | number of floors | waterfront present | number of views | condition of the house | ... | Built Year | F |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | False | False | False | False | False | False | False | False | False | False | ... | False | |
| **1** | False | False | False | False | False | False | False | False | False | False | ... | False | |
| **2** | False | False | False | False | False | False | False | False | False | False | ... | False | |
| **3** | False | False | False | False | False | False | False | False | False | False | ... | False | |
| **4** | False | False | False | False | False | False | False | False | False | False | ... | False | |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| **14615** | False | False | False | False | False | False | False | False | False | False | ... | False | |
| **14616** | False | False | False | False | False | False | False | False | False | False | ... | False | |
| **14617** | False | False | False | False | False | False | False | False | False | False | ... | False | |
| **14618** | False | False | False | False | False | False | False | False | False | False | ... | False | |
| **14619** | False | False | False | False | False | False | False | False | False | False | ... | False | |

14620 rows × 23 columns

In [47]:
```
dt.dropna()
```

Out[47]:

| | id | Date | number of bedrooms | number of bathrooms | living area | lot area | number of floors | waterfront present | number of views | condition of the house | ... |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 6762810145 | 42491 | 5 | 2.50 | 3650 | 9050 | 2.0 | 0 | 4 | 5 | ... |
| **1** | 6762810635 | 42491 | 4 | 2.50 | 2920 | 4000 | 1.5 | 0 | 0 | 5 | ... |
| **2** | 6762810998 | 42491 | 5 | 2.75 | 2910 | 9480 | 1.5 | 0 | 0 | 3 | ... |
| **3** | 6762812605 | 42491 | 4 | 2.50 | 3310 | 42998 | 2.0 | 0 | 0 | 3 | ... |
| **4** | 6762812919 | 42491 | 3 | 2.00 | 2710 | 4500 | 1.5 | 0 | 0 | 4 | ... |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **14615** | 6762830250 | 42734 | 2 | 1.50 | 1556 | 20000 | 1.0 | 0 | 0 | 4 | ... |
| **14616** | 6762830339 | 42734 | 3 | 2.00 | 1680 | 7000 | 1.5 | 0 | 0 | 4 | ... |
| **14617** | 6762830618 | 42734 | 2 | 1.00 | 1070 | 6120 | 1.0 | 0 | 0 | 3 | ... |
| **14618** | 6762830709 | 42734 | 4 | 1.00 | 1030 | 6621 | 1.0 | 0 | 0 | 4 | ... |
| **14619** | 6762831463 | 42734 | 3 | 1.00 | 900 | 4770 | 1.0 | 0 | 0 | 3 | ... |

14620 rows × 23 columns