```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib as mplt
import warnings

d = pd.read_csv("C:\\Users\\Admin\\Desktop\\dataset\\house price
dataset.csv")

df = pd.DataFrame(d)

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]
```

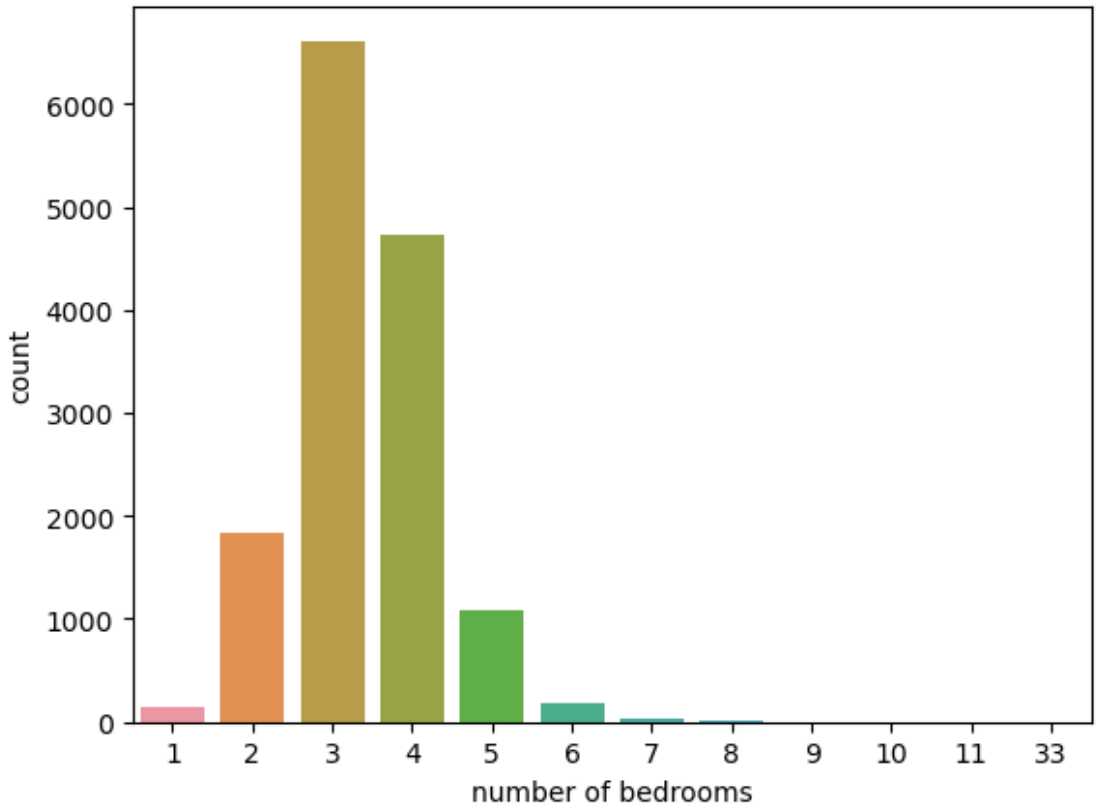
```
df.columns
```

```
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')
```

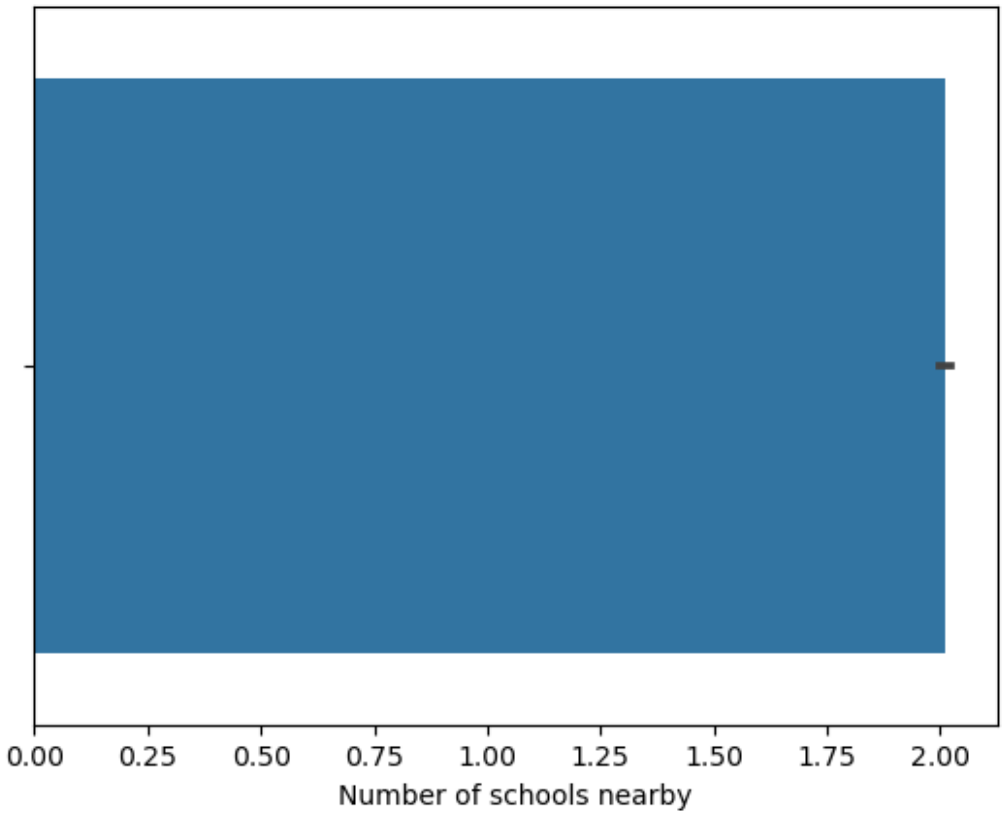
```
sns.countplot(data=df,x='number of bedrooms')
```

```
<Axes: xlabel='number of bedrooms', ylabel='count'>
```



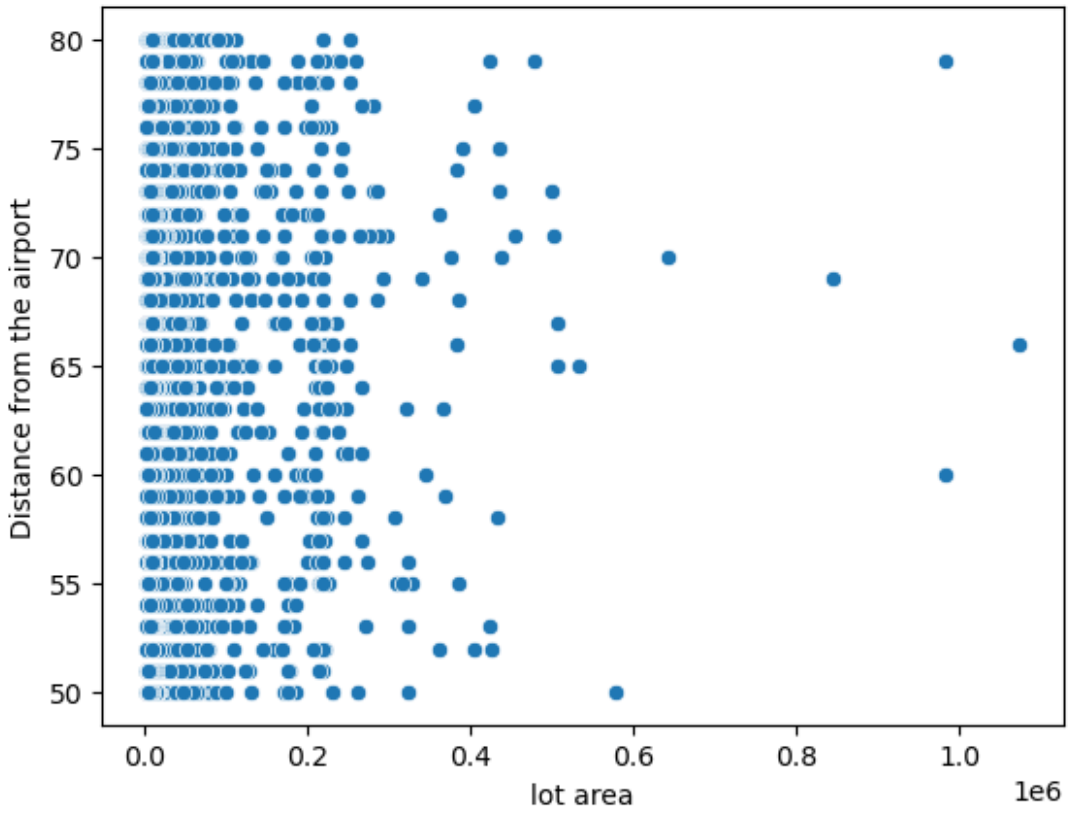

```
sns.barplot(data=df,x='Number of schools nearby')
```

```
<Axes: xlabel='Number of schools nearby'>
```



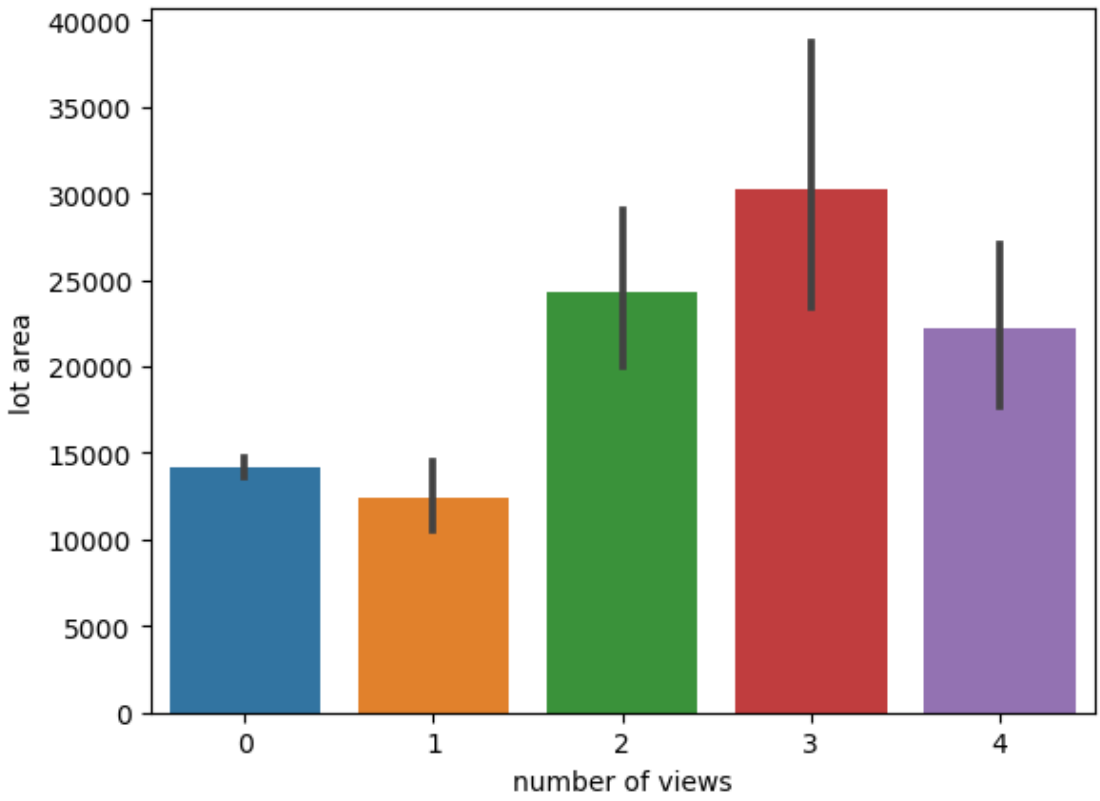

```
sns.scatterplot(data=df,x='lot area',y='Distance from the airport')
```

```
<Axes: xlabel='lot area', ylabel='Distance from the airport'>
```



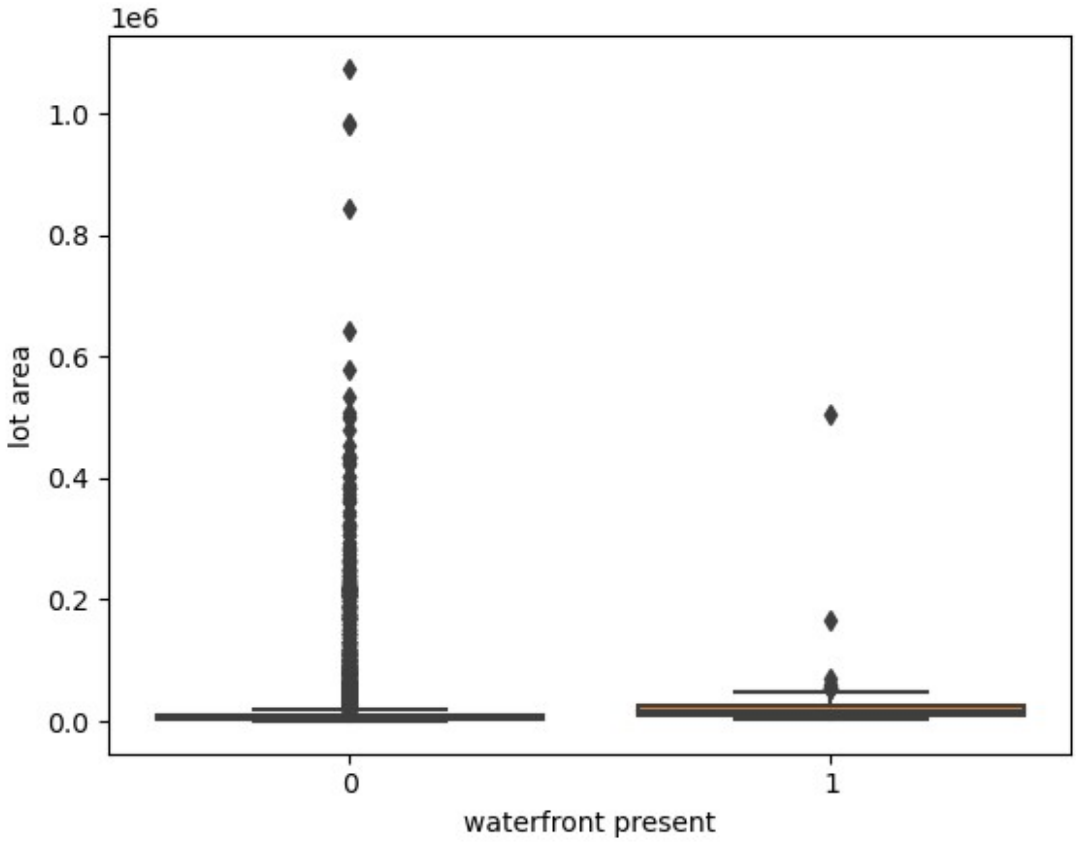

```
sns.barplot(data=df,x='number of views',y='lot area')
```

```
<Axes: xlabel='number of views', ylabel='lot area'>
```




```
sns.boxplot(data=df,x='waterfront present',y='lot area')
```

```
<Axes: xlabel='waterfront present', ylabel='lot area'>
```

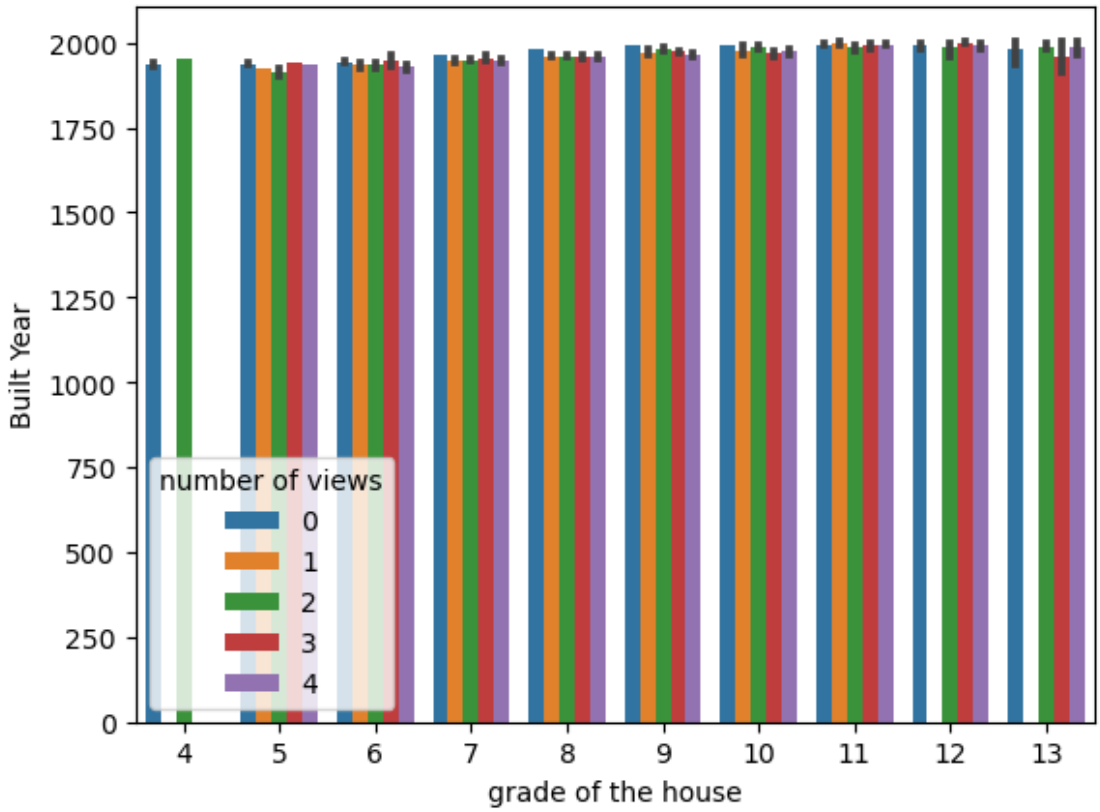



```
sns.heatmap(df.corr())
```

```
<Axes: >
```

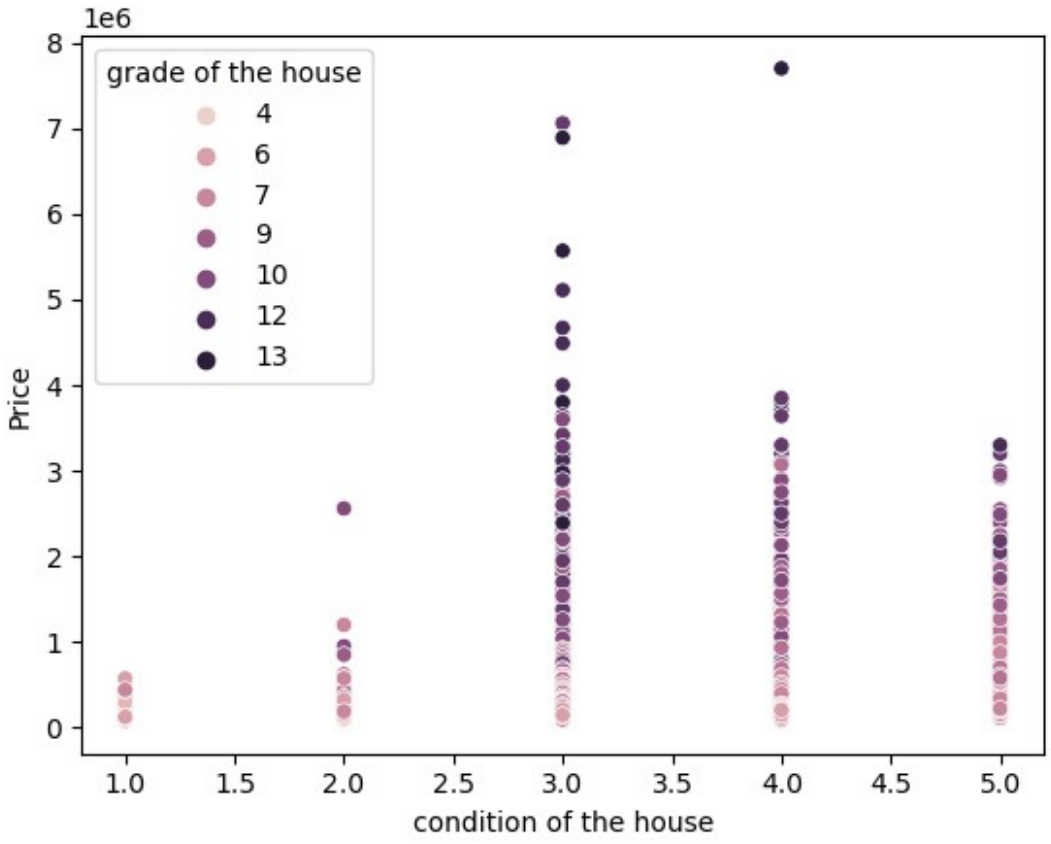

```
sns.barplot(data=df,x='grade of the house',y='Built Year',hue='number of views')
```

```
<Axes: xlabel='grade of the house', ylabel='Built Year'>
```




```
sns.scatterplot(data=df,x='condition of the
house',y='Price',hue='grade of the house')
```

```
<Axes: xlabel='condition of the house', ylabel='Price'>
```




```
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]
```

```
df['Postal Code'].value_counts()
```

```
122028    432
122005    416
122006    397
122007    396
122033    383
         ...
122013     77
122062     62
122036     49
122060     39
122071     37
Name: Postal Code, Length: 70, dtype: int64
```

```
df['Postal Code'].value_counts().to_frame()
```

```
        Postal Code
122028          432
122005          416
122006          397
122007          396
122033          383
...             ...
122013           77
122062           62
122036           49
122060           39
122071           37

[70 rows x 1 columns]
```

```
df['Built Year'].mean()
```

```
1970.9264021887825
```

```
df=df['Built Year'].fillna(1971)
```

```
print(df)
```

```
0        1921
1        1909
2        1939
3        2001
```

```
4        1929
         ...
14615    1957
14616    1968
14617    1962
14618    1955
14619    1969
Name: Built Year, Length: 14620, dtype: int64
```

```
df.isnull()
```

```
          id   Date  number of bedrooms  number of bathrooms  living
area  \
0      False  False               False                False
False
1      False  False               False                False
False
2      False  False               False                False
False
3      False  False               False                False
False
4      False  False               False                False
False
...      ...    ...                 ...                  ...
...
14615  False  False               False                False
False
14616  False  False               False                False
False
14617  False  False               False                False
False
14618  False  False               False                False
False
14619  False  False               False                False
False

       lot area  number of floors  waterfront present  number of views
\
0         False             False               False            False

1         False             False               False            False

2         False             False               False            False

3         False             False               False            False

4         False             False               False            False

...         ...               ...                 ...              ...
```

```
14615     False             False               False            False

14616     False             False               False            False

14617     False             False               False            False

14618     False             False               False            False

14619     False             False               False            False


       condition of the house  ...  Built Year  Renovation Year
Postal Code  \
0                         False  ...       False            False
False
1                         False  ...       False            False
False
2                         False  ...       False            False
False
3                         False  ...       False            False
False
4                         False  ...       False            False
False
...                         ...  ...         ...              ...
...
14615                     False  ...       False            False
False
14616                     False  ...       False            False
False
14617                     False  ...       False            False
False
14618                     False  ...       False            False
False
14619                     False  ...       False            False
False

       Lattitude  Longitude  living_area_renov  lot_area_renov  \
0          False      False              False           False
1          False      False              False           False
2          False      False              False           False
3          False      False              False           False
4          False      False              False           False
...          ...        ...                ...             ...
14615      False      False              False           False
14616      False      False              False           False
14617      False      False              False           False
14618      False      False              False           False
14619      False      False              False           False
```

```
       Number of schools nearby  Distance from the airport  Price
0                         False                      False  False
1                         False                      False  False
2                         False                      False  False
3                         False                      False  False
4                         False                      False  False
...                         ...                        ...    ...
14615                     False                      False  False
14616                     False                      False  False
14617                     False                      False  False
14618                     False                      False  False
14619                     False                      False  False

[14620 rows x 23 columns]
```

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

```
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
```