In [23]:
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

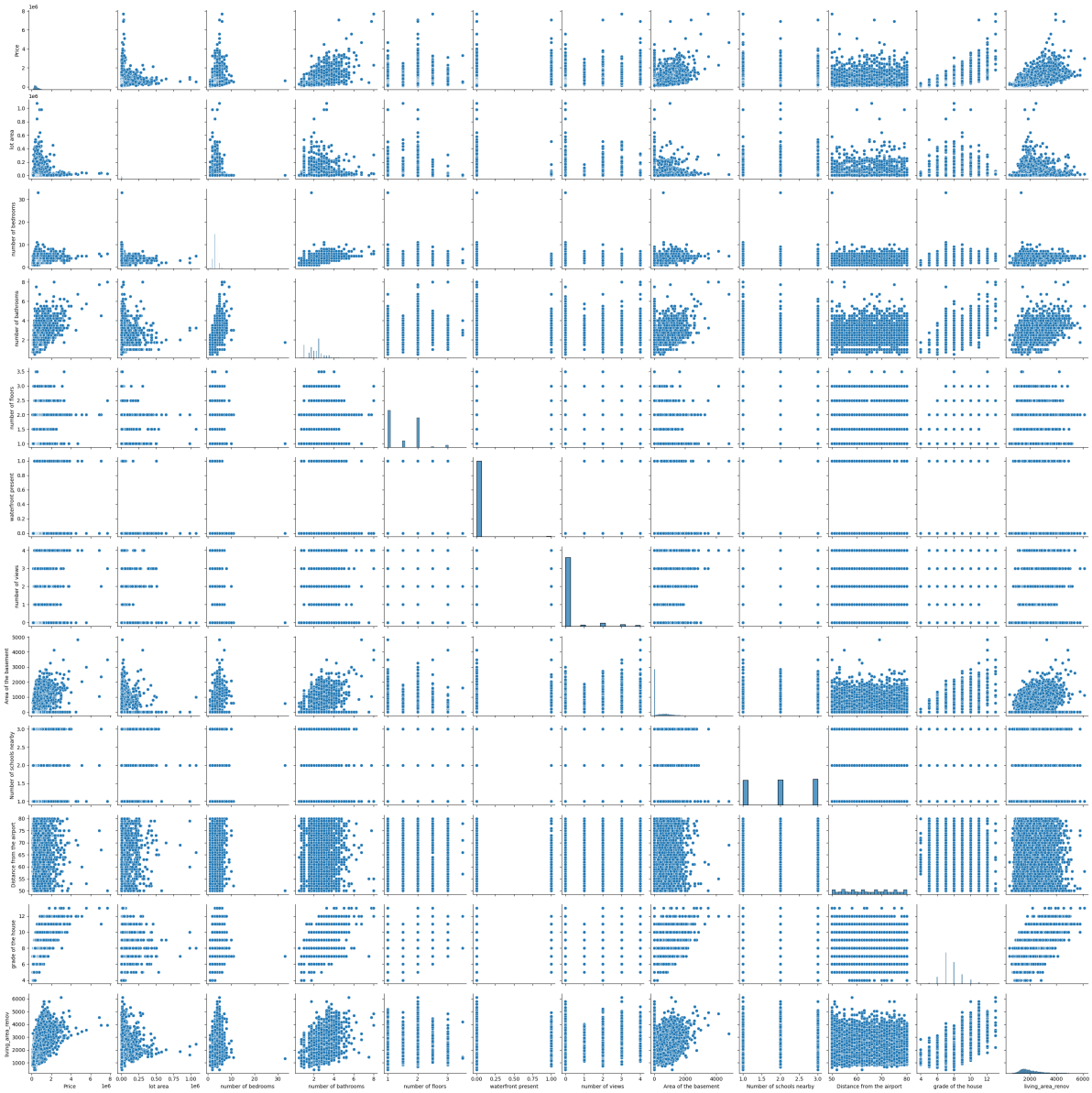
In [24]:
```
data = pd.read_csv('House Price India.csv')
```

In [25]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 14620 entries, 0 to 14619
Data columns (total 23 columns):
 #   Column                                 Non-Null Count  Dtype  
---  ------                                 --------------  -----  
 0   id                                     14620 non-null  int64  
 1   Date                                   14620 non-null  int64  
 2   number of bedrooms                     14620 non-null  int64  
 3   number of bathrooms                    14620 non-null  float64
 4   living area                            14620 non-null  int64  
 5   lot area                               14620 non-null  int64  
 6   number of floors                       14620 non-null  float64
 7   waterfront present                     14620 non-null  int64  
 8   number of views                        14620 non-null  int64  
 9   condition of the house                 14620 non-null  int64  
 10  grade of the house                     14620 non-null  int64  
 11  Area of the house(excluding basement)  14620 non-null  int64  
 12  Area of the basement                   14620 non-null  int64  
 13  Built Year                             14620 non-null  int64  
 14  Renovation Year                        14620 non-null  int64  
 15  Postal Code                            14620 non-null  int64  
 16  Lattitude                              14620 non-null  float64
 17  Longitude                              14620 non-null  float64
 18  living_area_renov                      14620 non-null  int64  
 19  lot_area_renov                         14620 non-null  int64  
 20  Number of schools nearby               14620 non-null  int64  
 21  Distance from the airport              14620 non-null  int64  
 22  Price                                  14620 non-null  int64  
dtypes: float64(4), int64(19)
memory usage: 2.6 MB
```

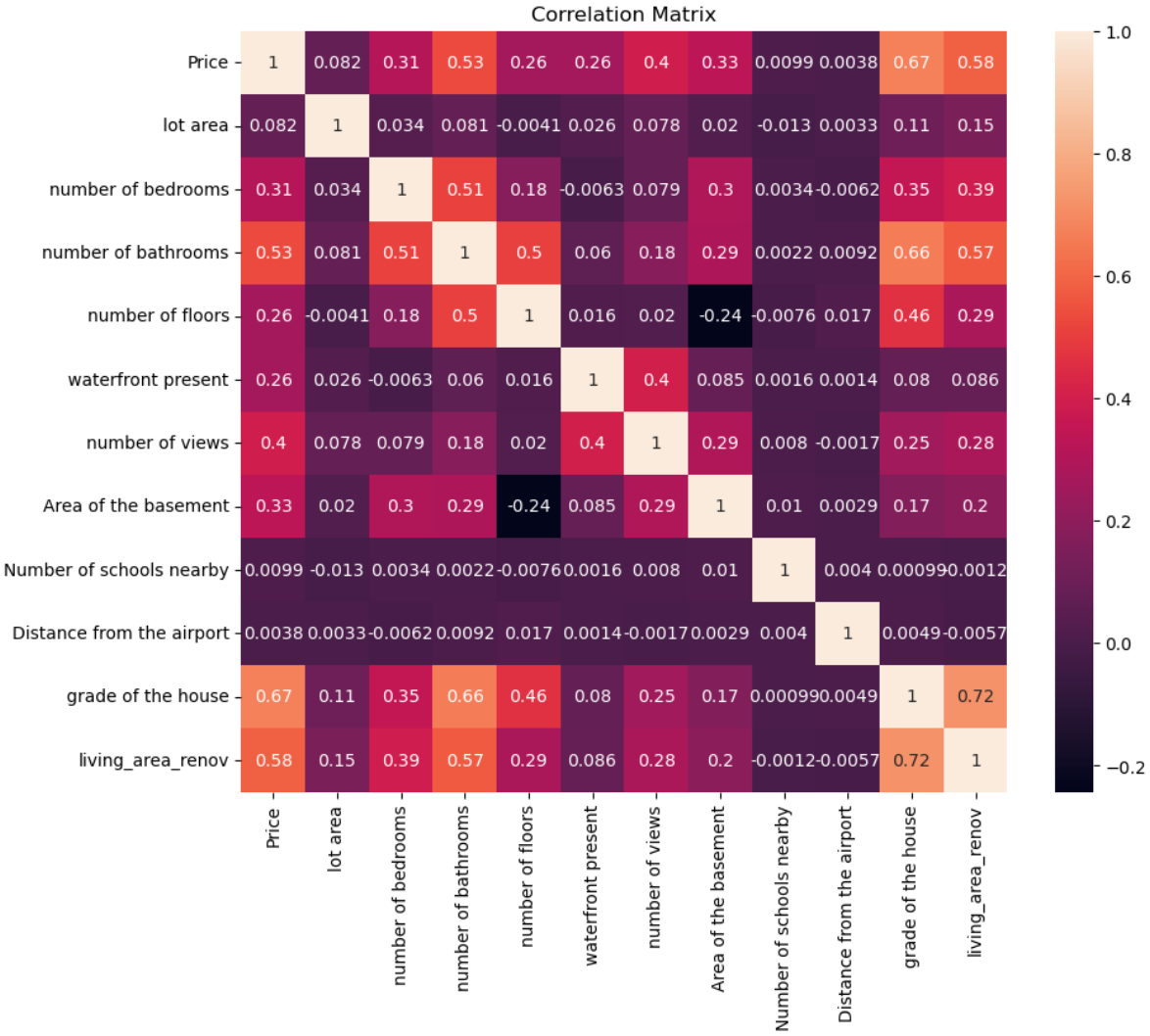
In [26]:
```
variables = ['Price', 'lot area', 'number of bedrooms', 'number of bathrooms', 'nu
```

In [27]:
```
sns.pairplot(data[variables])
plt.show()
```



In [28]:
```
correlation_matrix = data[variables].corr()
```

In [30]:
```
plt.figure(figsize=(10,8))
sns.heatmap(correlation_matrix, annot=True)
plt.title('Correlation Matrix')
plt.show()
```


Correlation Matrix
Price 1 0.082 0.31 0.53 0.26 0.26 0.4 0.33 0.0099 0.0038 0.67 0.58
lot area 0.082 1 0.034 0.081 -0.0041 0.026 0.078 0.02 -0.013 0.0033 0.11 0.15
number of bedrooms 0.31 0.034 1 0.51 0.18 -0.0063 0.079 0.3 0.0034 -0.0062 0.35 0.39
number of bathrooms 0.53 0.081 0.51 1 0.5 0.06 0.18 0.29 0.0022 0.0092 0.66 0.57
number of floors 0.26 -0.0041 0.18 0.5 1 0.016 0.02 -0.24 -0.0076 0.017 0.46 0.29
waterfront present 0.26 0.026 -0.0063 0.06 0.016 1 0.4 0.085 0.0016 0.0014 0.08 0.086
number of views 0.4 0.078 0.079 0.18 0.02 0.4 1 0.29 0.008 -0.0017 0.25 0.28
Area of the basement 0.33 0.02 0.3 0.29 -0.24 0.085 0.29 1 0.01 0.0029 0.17 0.2
Number of schools nearby 0.0099 -0.013 0.0034 0.0022 -0.0076 0.0016 0.008 0.01 1 0.004 0.00099 -0.0012
Distance from the airport 0.0038 0.0033 -0.0062 0.0092 0.017 0.0014 -0.0017 0.0029 0.004 1 0.0049 -0.0057
grade of the house 0.67 0.11 0.35 0.66 0.46 0.08 0.25 0.17 0.00099 0.0049 1 0.72
living_area_renov 0.58 0.15 0.39 0.57 0.29 0.086 0.28 0.2 -0.0012 -0.0057 0.72 1
Price
lot area
number of bedrooms
number of bathrooms
number of floors
waterfront present
number of views
Area of the basement
Number of schools nearby
Distance from the airport
grade of the house
living_area_renov
1.0
0.8
0.6
0.4
0.2
0.0
−0.2

In [ ]: