Files

+ Code + Text

```
[9] import pandas as pd
    df = pd.read_csv('House Price India.csv')
    df.head()
```

	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	Renovation Year	Postal Code	Lattitude	Longi
	6762810145	42491	5	2.50	3650	9050	2.0	0	4	5	...	1921	0	122003	52.8645	-11.
	6762810635	42491	4	2.50	2920	4000	1.5	0	0	5	...	1909	0	122004	52.8878	-11.
	6762810998	42491	5	2.75	2910	9480	1.5	0	0	3	...	1939	0	122004	52.8852	-11.
	6762812605	42491	4	2.50	3310	42998	2.0	0	0	3	...	2001	0	122005	52.9532	-11.
	6762812919	42491	3	2.00	2710	4500	1.5	0	0	4	...	1929	0	122006	52.9047	-11.

ows × 23 columns

```
[10] df.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
```

Disk ▬▬▬▬ 81.32 GB available

Files

Connecting to a runtime to enable file browsing.

+ Code + Text

```
df.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
```
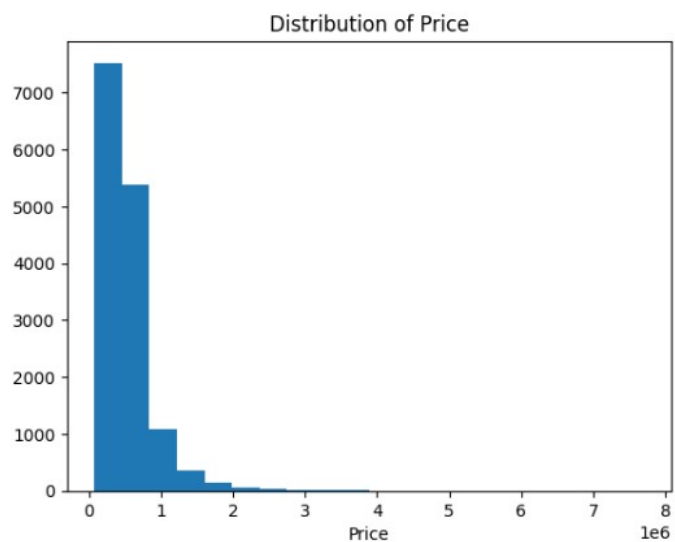
+ Code   + Text

```python
import matplotlib.pyplot as plt
plt.hist(df['Price'],bins=20)
plt.xlabel('Price')
plt.title('Distribution of Price')
plt.show()
```


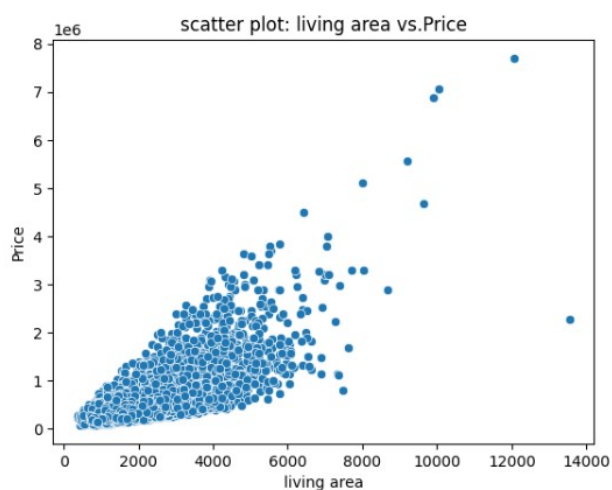Distribution of Price

+ Code   + Text

```python
import seaborn as sns
sns.scatterplot(x='living area',y='Price', data=df)
plt.xlabel('living area')
plt.title('scatter plot: living area vs.Price')
plt.show()
```


scatter plot: living area vs.Price
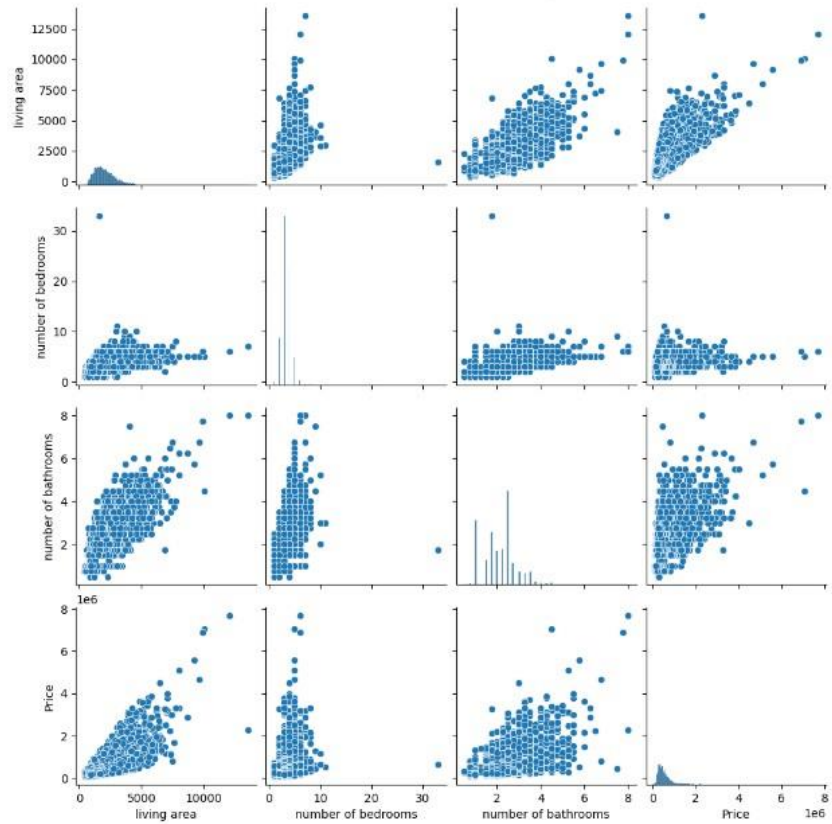
≡ Files ◰ ✕

🔍 Connecting to a runtime to enable file browsing.

{x}

📁

+ Code + Text

```
[ ]  sns.pairplot(df[['living area','number of bedrooms','number of bathrooms','Price']])
     plt.show()
```

+ Code + Text

Files

Connecting to a runtime to enable file browsing.

```python
print(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                3.438506  ...   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         98.924008  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     5897.750000                  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
```

+ Code + Text

```
df.dropna(inplace=True)
df['living area'].fillna(df['living area'].mean(),inplace=True)
print(df)
```

```
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_reno
0                    0       122003    52.8645   -114.557               288
1                    0       122004    52.8878   -114.470               247
2                    0       122004    52.8852   -114.468               294
3                    0       122005    52.9532   -114.321               335
4                    0       122006    52.9047   -114.485               206
...                ...          ...        ...        ...               ...
14615                0       122066    52.6191   -114.472               225
14616                0       122072    52.5875   -114.393               154
14617                0       122056    52.7289   -114.507               113
14618                0       122042    52.7157   -114.411               142
14619             2009       122018    52.5338   -114.552                90
```