Assignment 2

21BPS1407
VIJAY SANKAR REDDY
EVENING BATCH

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

```
df = pd.read_csv("/content/House Price India.csv")
df.head
```

```
<bound method NDFrame.head of                 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
...                ...                     ...  ...         ...
```

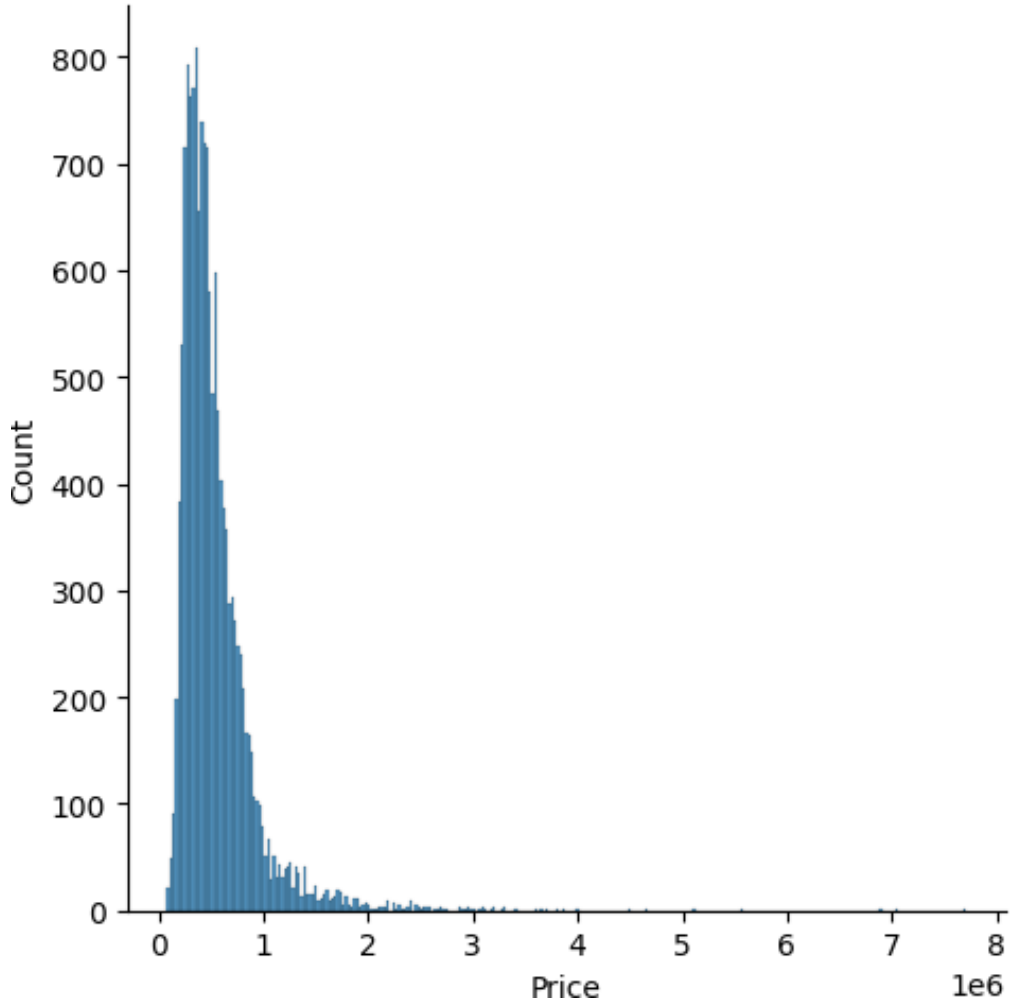
```
14615                0                       4  ...         1957
14616                0                       4  ...         1968
14617                0                       3  ...         1962
14618                0                       4  ...         1955
14619                0                       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
```

```
df.shape
```

```
(14620, 23)
```

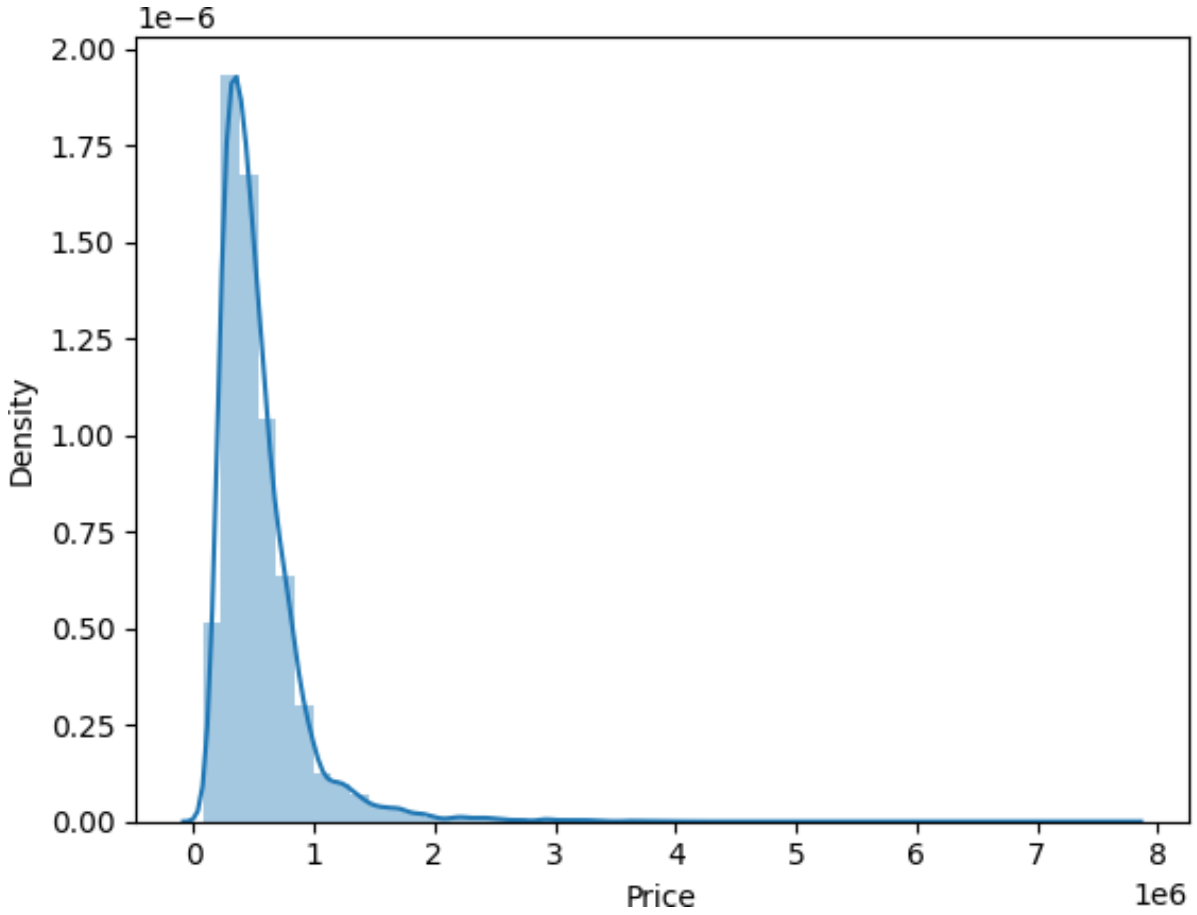
```
# UNIVARIATE ANALYSIS
import seaborn as sns
sns.displot(df.Price)
```

```
<seaborn.axisgrid.FacetGrid at 0x7f3e49cb2440>
```



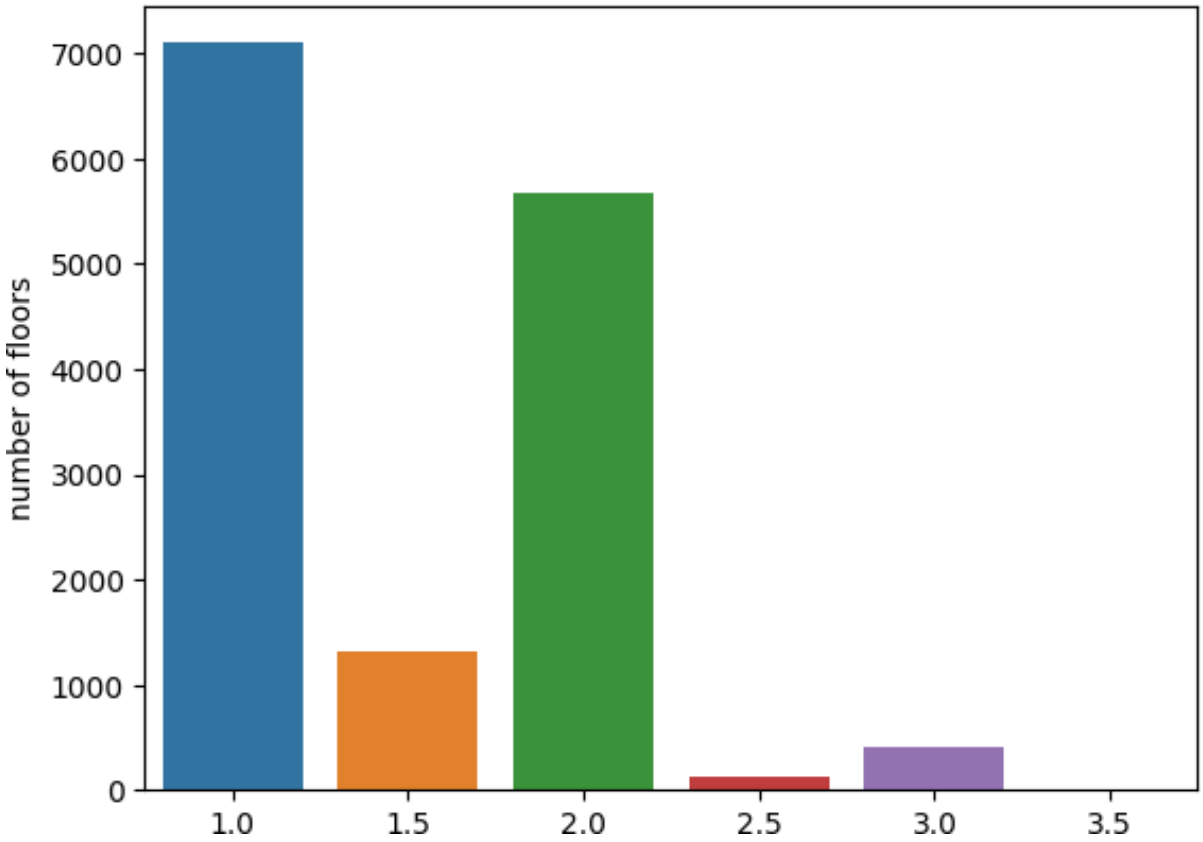
```
sns.distplot(df.Price)
```

```
<ipython-input-7-5e080168c38c>: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 wit
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.Price)
<Axes: xlabel='Price', ylabel='Density'>
```



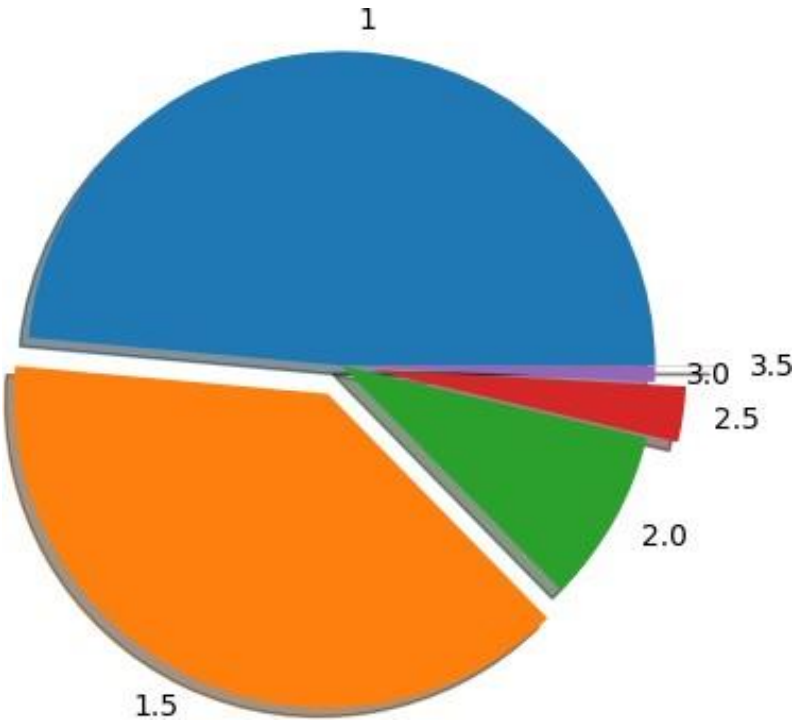
```
# BAR PLOT
sns.barplot(x = df['number of floors'].value_counts().index,y = df['number of fl
```

```
<Axes: ylabel='number of floors'>
```



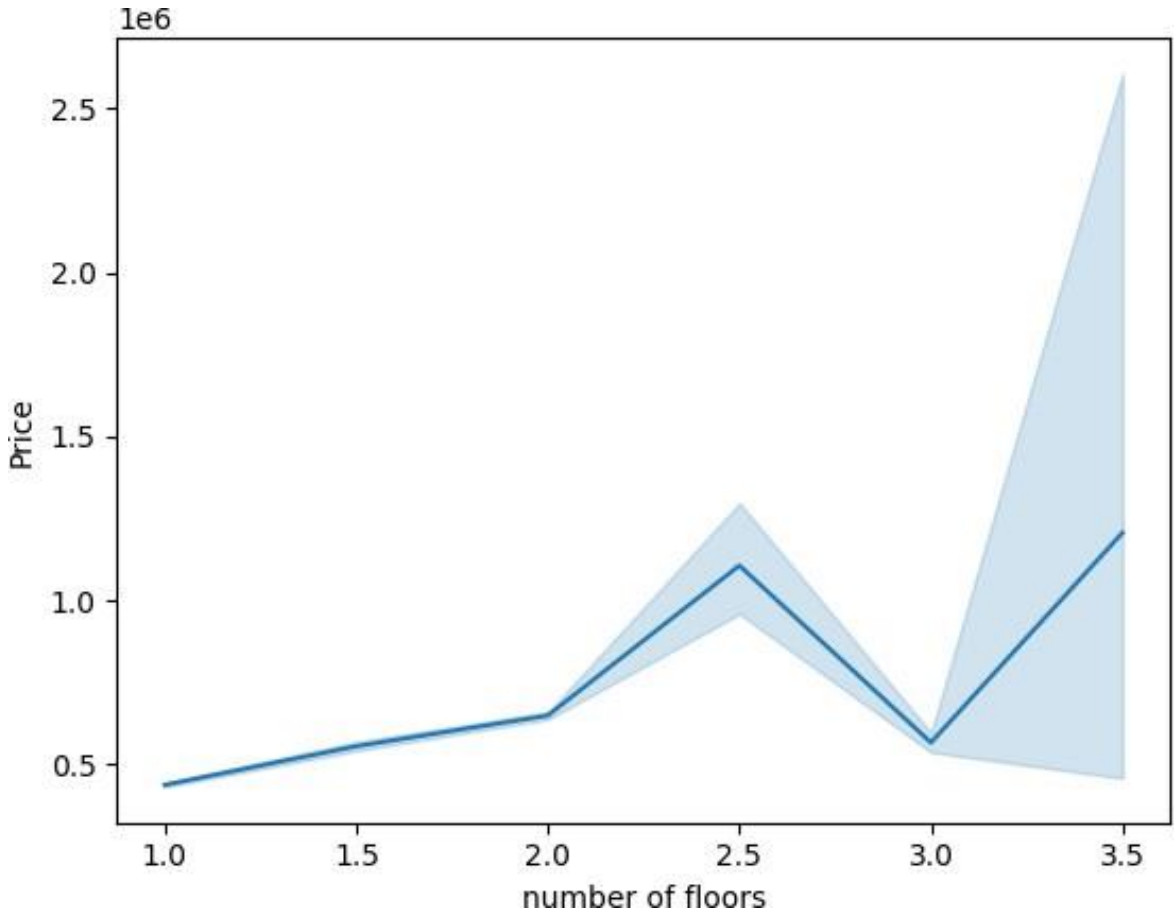
```
plt.pie(df['number of floors'].value_counts(),[0,0.1,0,0.1,0,0.2],labels = ['1',
```

```
([<matplotlib.patches.Wedge at 0x7f3e3c5058d0>,
  <matplotlib.patches.Wedge at 0x7f3e3c505450>,
  <matplotlib.patches.Wedge at 0x7f3e3c5047f0>,
  <matplotlib.patches.Wedge at 0x7f3e3c504100>,
  <matplotlib.patches.Wedge at 0x7f3e3c530550>,
  <matplotlib.patches.Wedge at 0x7f3e3c530c10>],
 [Text(0.048912807550267855, 1.098911978848875, '1'),
  Text(-0.5135468709023937, -1.0845596393865855, '1.5'),
  Text(0.9579791438698172, -0.5406255265065203, '2.0'),
  Text(1.187878966741221, -0.17012807050516252, '2.5'),
  Text(1.099596836099304, -0.029779154460806083, '3.0'),
  Text(1.2999995198044283, -0.0011173666612348623, '3.5')])
```



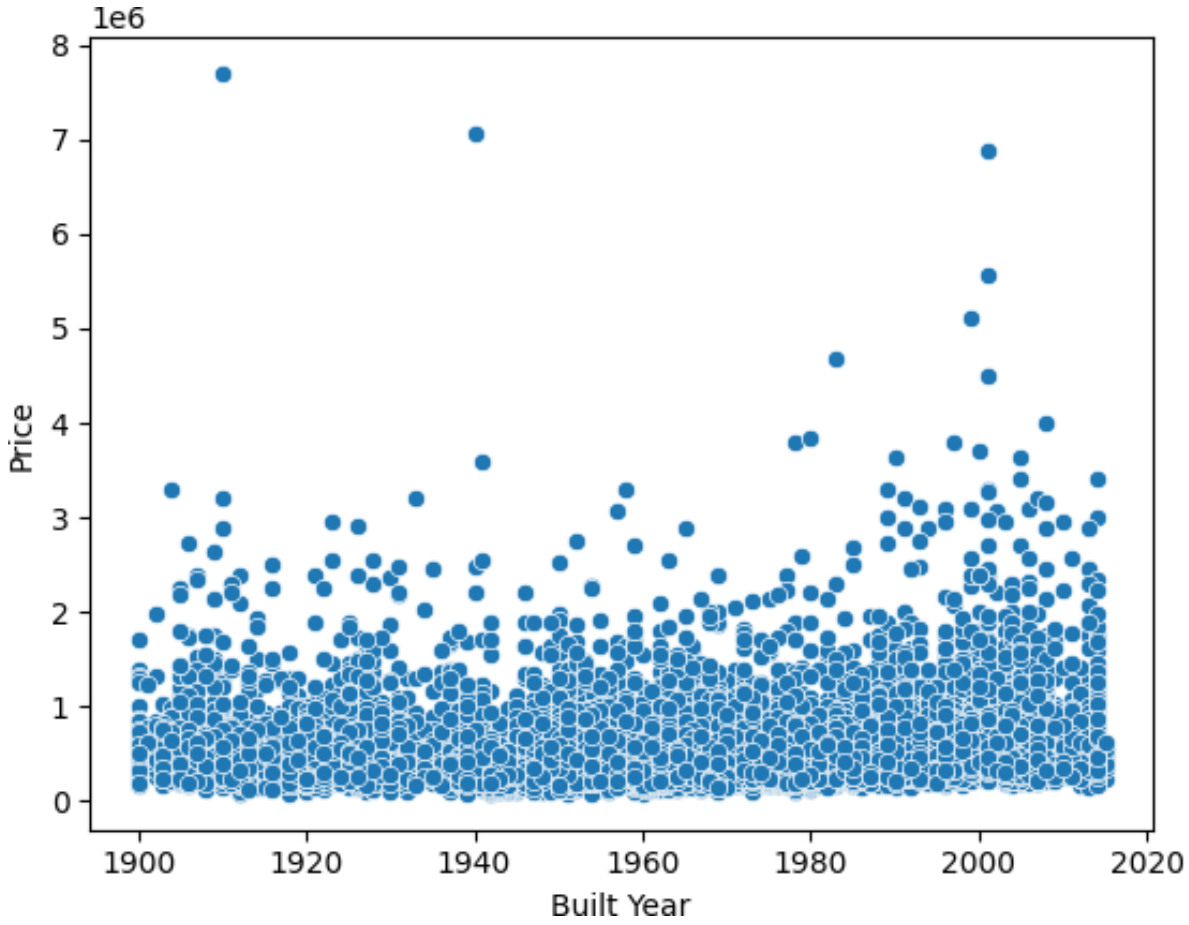
```
# BIVARIATE ANALYSIS
sns.lineplot(x = df['number of floors'],y = df['Price'])
```

```
<Axes: xlabel='number of floors', ylabel='Price'>
```



```
sns.scatterplot(x = df['Built Year'],y = df['Price'])
```

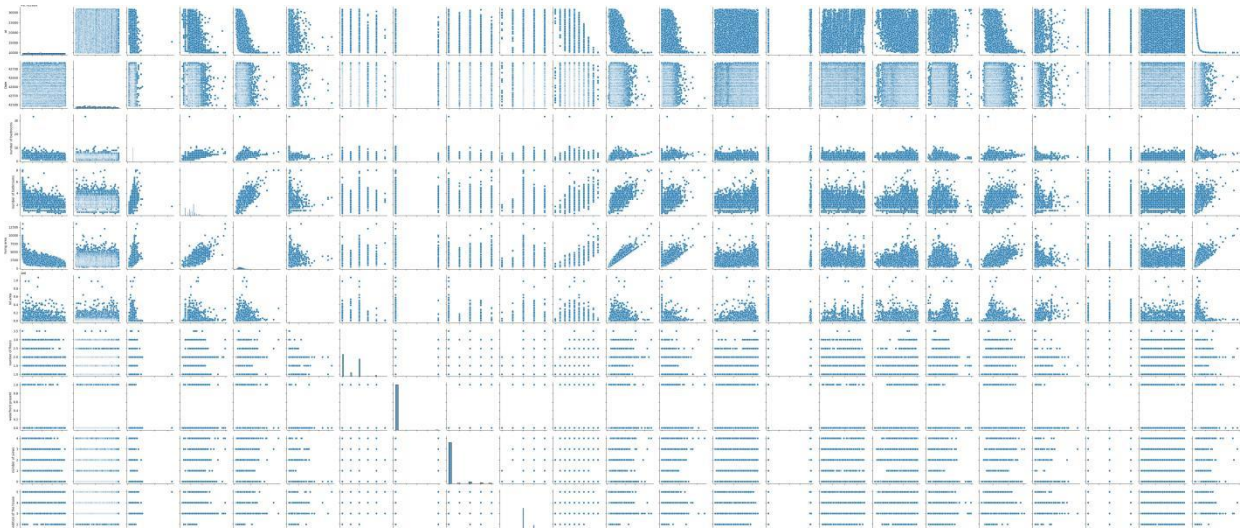
```
<Axes: xlabel='Built Year', ylabel='Price'>
```

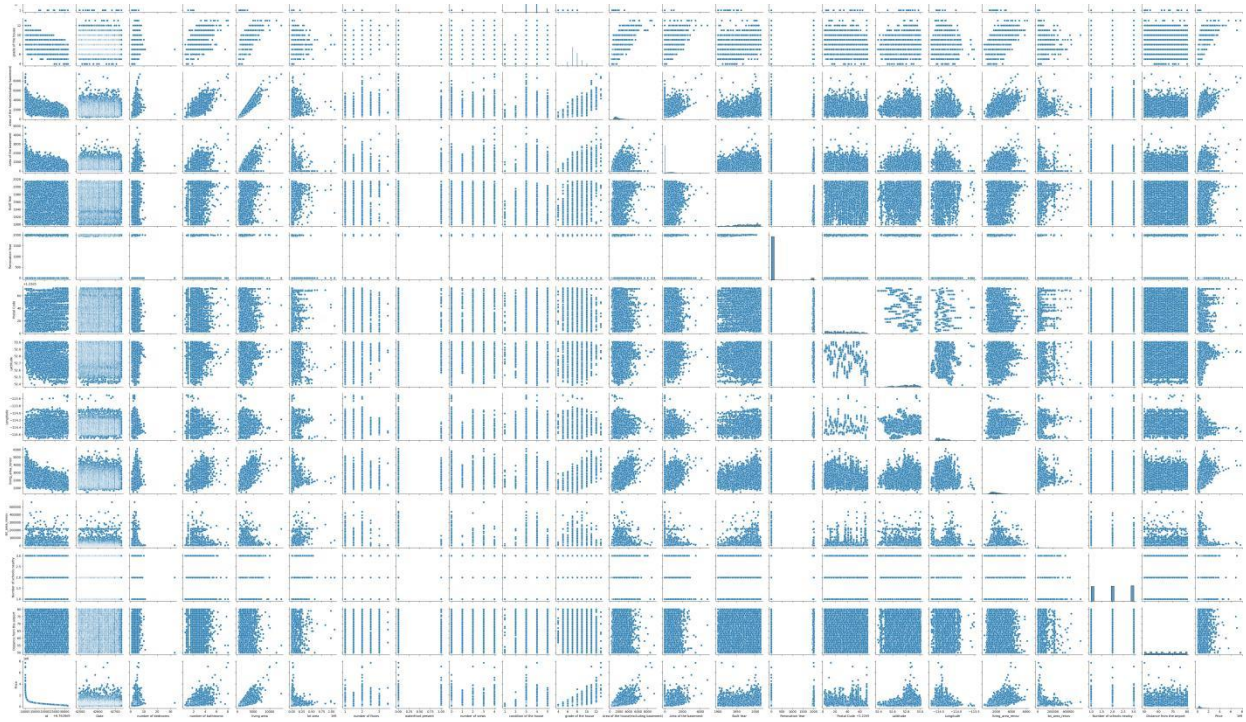



```
# MULTIVARIATE ANALYSIS
```

```
sns.pairplot(df)
```

```
<seaborn.axisgrid.PairGrid at 0x7f3e3c283700>
```





```
df.describe()
```

	id	Date	number of bedrooms	number of bathrooms	living area	lot
count	1.462000e+04	14620.000000	14620.000000	14620.000000	14620.000000	1.46200
mean	6.762821e+09	42604.538646	3.379343	2.129583	2098.262996	1.50932
std	6.237575e+03	67.347991	0.938719	0.769934	928.275721	3.79196
min	6.762810e+09	42491.000000	1.000000	0.500000	370.000000	5.20000
25%	6.762815e+09	42546.000000	3.000000	1.750000	1440.000000	5.01075
50%	6.762821e+09	42600.000000	3.000000	2.250000	1930.000000	7.62000
75%	6.762826e+09	42662.000000	4.000000	2.500000	2570.000000	1.08000
max	6.762832e+09	42734.000000	33.000000	8.000000	13540.000000	1.07421

8 rows × 23 columns

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

# hence there are no null values to handle in the dataset
```

```
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
```

Colab paid products - Cancel contracts here