Load the dataset.

```
import pandas as pd
import matplotlib.pyplot as plt
from matplotlib import rcParams
import seaborn as sns
```

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

living area	lot area	number of floors	waterfront present	number of views	condition of the house	...	Built Year	Renovation Year	Postal Code	Lattitude	Longit
3650	9050	2.0	0	4	5	...	1921	0	122003	52.8645	-114.
2920	4000	1.5	0	0	5	...	1909	0	122004	52.8878	-114.
2910	9480	1.5	0	0	3	...	1939	0	122004	52.8852	-114.
3310	42998	2.0	0	0	3	...	2001	0	122005	52.9532	-114.
2710	4500	1.5	0	0	4	...	1929	0	122006	52.9047	-114.
...	...	...	...	...	...	...	...	...	...	...	
1556	20000	1.0	0	0	4	...	1957	0	122066	52.6191	-114.
1680	7000	1.5	0	0	4	...	1968	0	122072	52.5075	-114.
1070	6120	1.0	0	0	3	...	1962	0	122056	52.7289	-114.
1030	6621	1.0	0	0	4	...	1955	0	122042	52.7157	-114.
900	4770	1.0	0	0	3	...	1969	2009	122018	52.5338	-114.

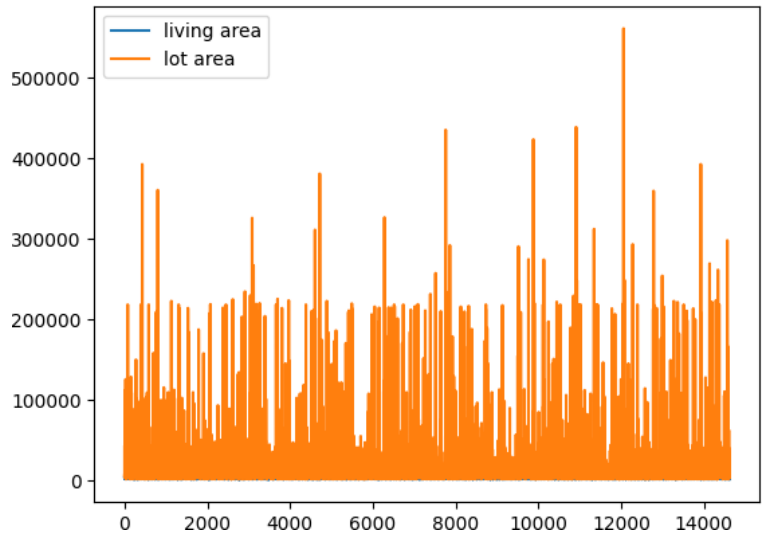
```
df.shape
```

```
(14620, 23)
```

Uni variate Analysis

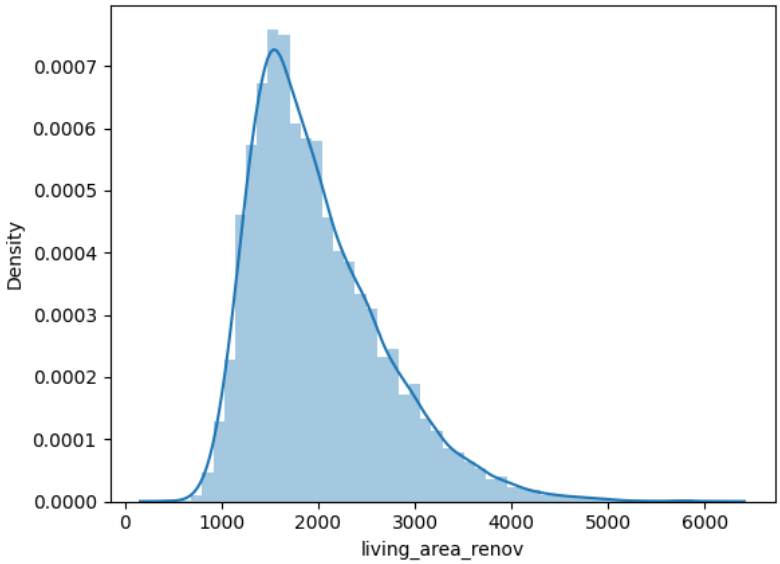
```
df.living_area_renov.plot()
df.lot_area_renov.plot()
plt.legend(['living area','lot area'])
```

```
<matplotlib.legend.Legend at 0x79a08c113d00>
```



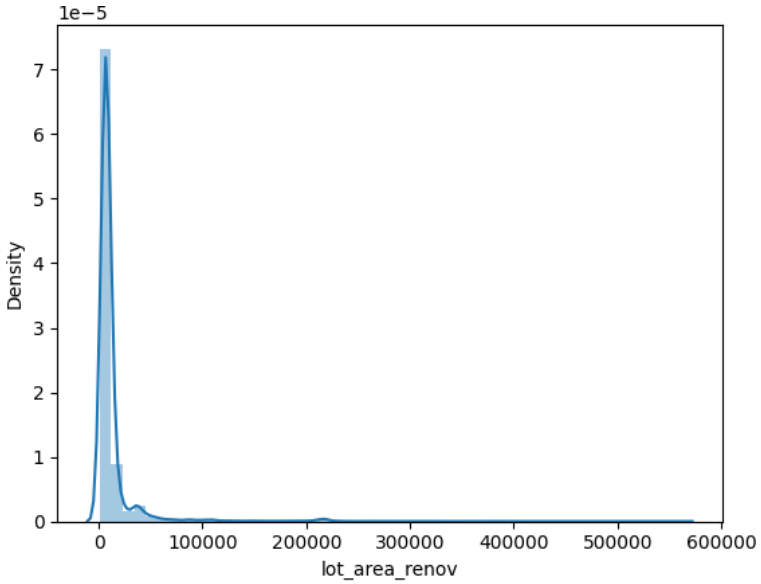

```
sns.distplot(df.living_area_renov)
```

```
<ipython-input-18-b944eacf0633>: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 with
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.living_area_renov)
<Axes: xlabel='living_area_renov', ylabel='Density'>
```




```
sns.distplot(df.lot_area_renov)
```

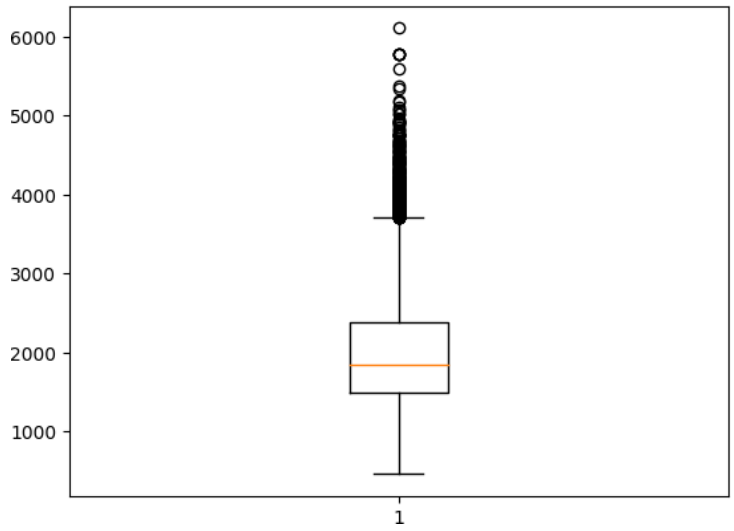
```
<ipython-input-9-fbd8f64c04a5>:1: UserWarning:

`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 with
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.lot_area_renov)
<Axes: xlabel='lot_area_renov', ylabel='Density'>
```



Bivariate Analysis

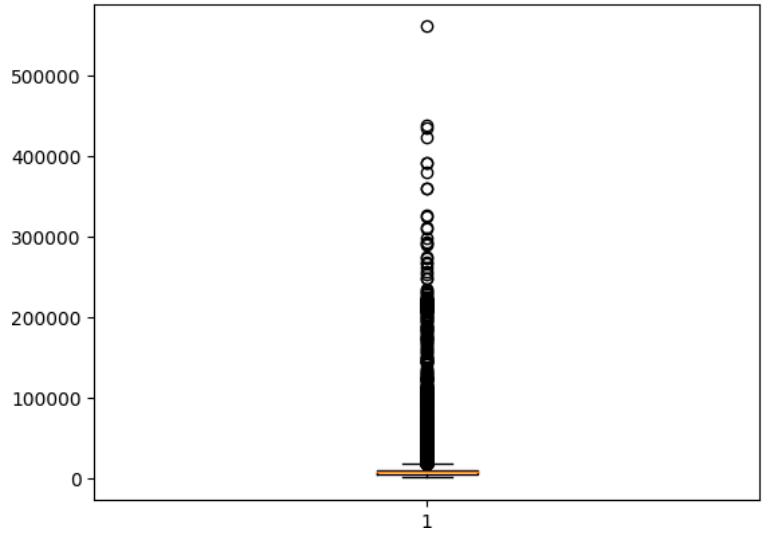
```
plt.boxplot(df['living_area_renov'])
```

```
{'whiskers': [<matplotlib.lines.Line2D at 0x7841f9734040>,
  <matplotlib.lines.Line2D at 0x7841f97341c0>],
 'caps': [<matplotlib.lines.Line2D at 0x7841f9734460>,
  <matplotlib.lines.Line2D at 0x7841f9734700>],
 'boxes': [<matplotlib.lines.Line2D at 0x7841f96f7d60>],
 'medians': [<matplotlib.lines.Line2D at 0x7841f97349a0>],
 'fliers': [<matplotlib.lines.Line2D at 0x7841f9734c40>],
 'means': []}
```




```
plt.boxplot(df['lot_area_renov'])
```

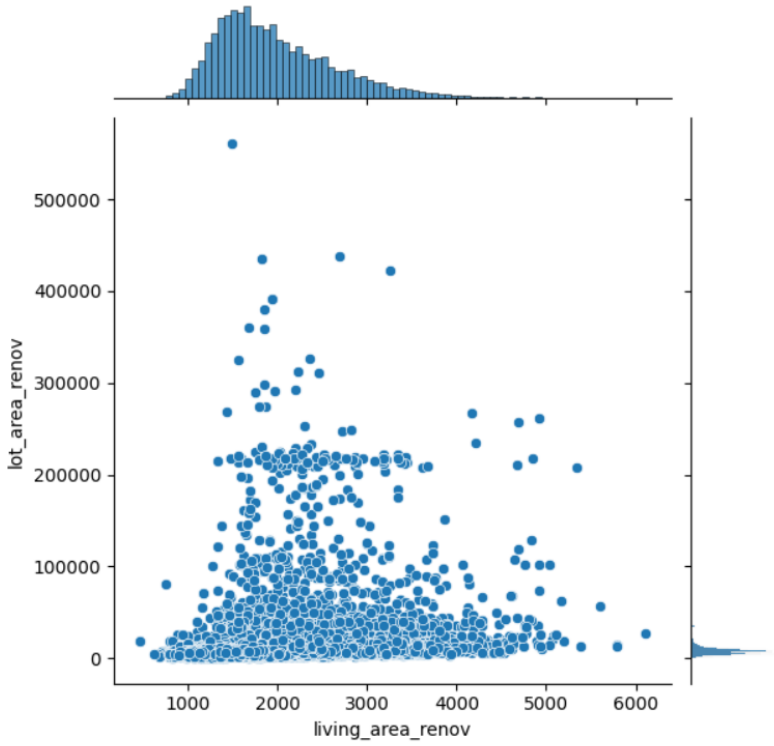
```
{'whiskers': [<matplotlib.lines.Line2D at 0x7841f95858d0>,
  <matplotlib.lines.Line2D at 0x7841f9585b70>],
 'caps': [<matplotlib.lines.Line2D at 0x7841f9585e10>,
  <matplotlib.lines.Line2D at 0x7841f95860b0>],
 'boxes': [<matplotlib.lines.Line2D at 0x7841f9585630>],
 'medians': [<matplotlib.lines.Line2D at 0x7841f9586350>],
 'fliers': [<matplotlib.lines.Line2D at 0x7841f95865f0>],
 'means': []}
```




```
sns.jointplot(x= 'living_area_renov',y ='lot_area_renov',data=df)
```

```
<seaborn.axisgrid.JointGrid at 0x7841f7d4f040>
```

```
<seaborn.axisgrid.JointGrid at 0x7841f7d4f040>
```



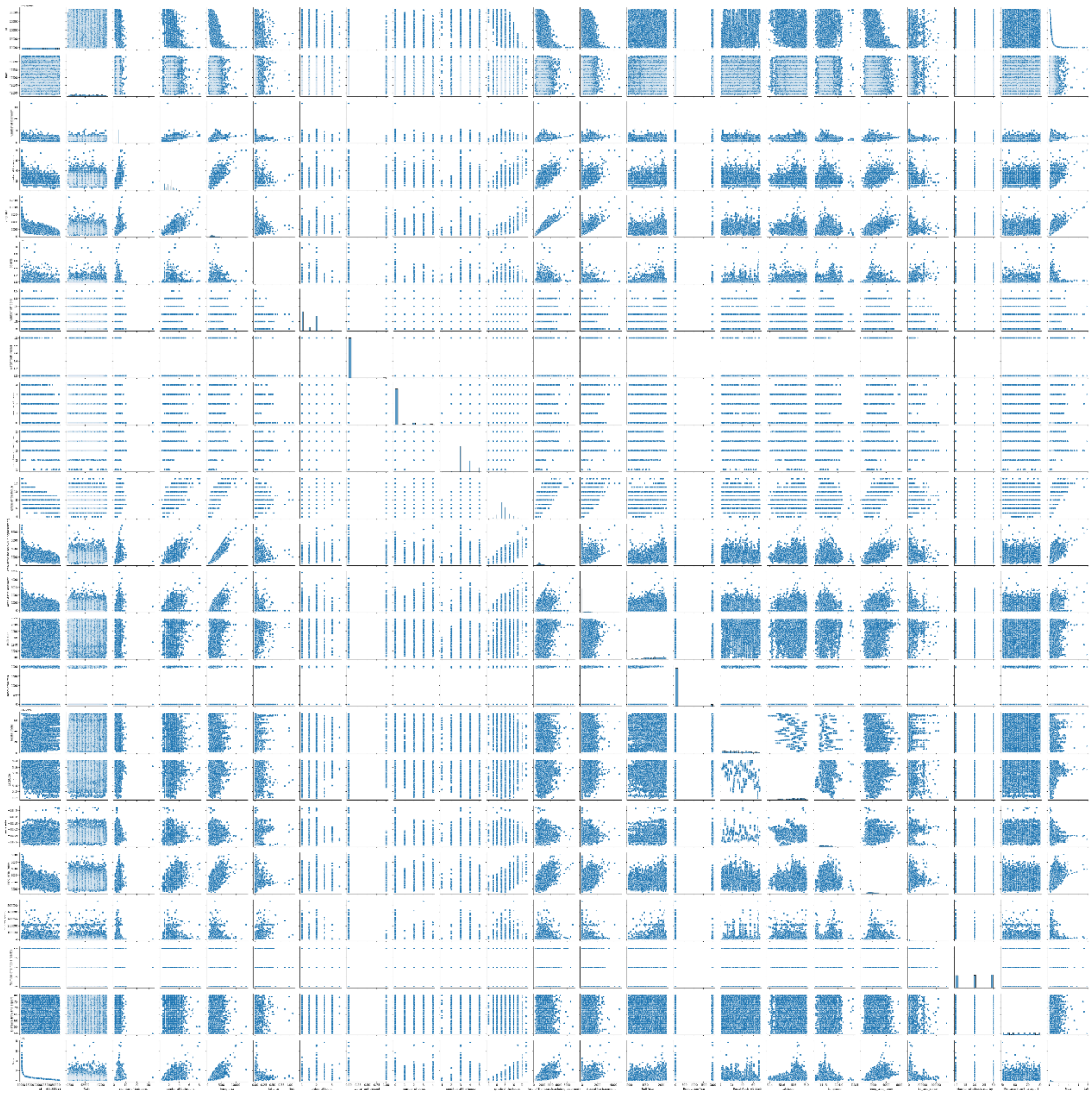
Multivariate analysis

```
import pandas as pd
import matplotlib.pyplot as plt
from matplotlib import rcParams
import seaborn as sns
```

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

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

```
<seaborn.axisgrid.PairGrid at 0x79a0e62aab60>
```



Perform descriptive statistics on the dataset.

```
desc_stats = df.describe()
print(desc_stats)
```

```
                  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
```

Handle the Missing values

```
missing_values = df.fillna(df.mean(), inplace=True)
print(missing_values)
```

```
None
```

Drive link: https://colab.research.google.com/drive/15Q8D1mcA_0FVvY79iJOw59qdI396vlF6?usp=drive_link