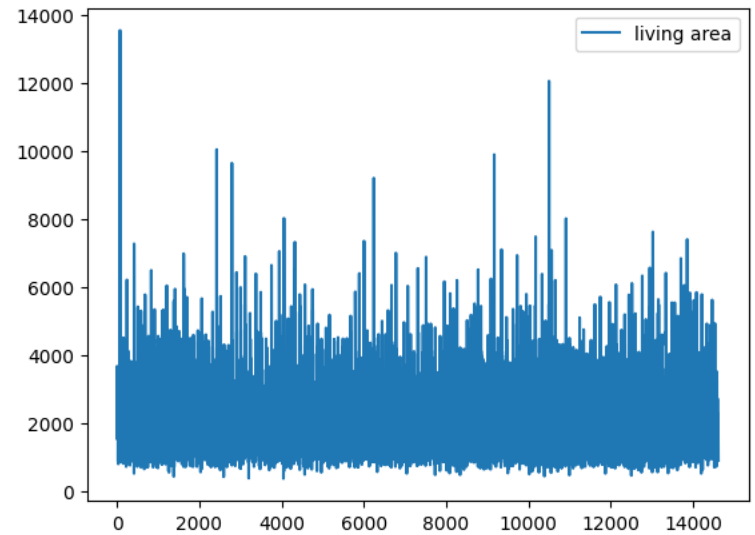
```
# Reg No: 21BEC2297
# Name: Keshav Goyal
import pandas as pd
import matplotlib.pyplot as plt
from matplotlib import rcParams
import seaborn as sns


df = pd.read_csv('/content/House_Price_India.csv') # Importing the dataset


df
```

	id	Date	number of bedrooms	number of bathrooms	living area	lot area	number of floors	waterfront present	number of views
0	6762810145	42491	5	2.50	3650	9050	2.0	0	4
1	6762810635	42491	4	2.50	2920	4000	1.5	0	0
2	6762810998	42491	5	2.75	2910	9480	1.5	0	0
3	6762812605	42491	4	2.50	3310	42998	2.0	0	0
4	6762812919	42491	3	2.00	2710	4500	1.5	0	0
...	...	...	...	...	...	...	...	...	...
14615	6762830250	42734	2	1.50	1556	20000	1.0	0	0
14616	6762830339	42734	3	2.00	1680	7000	1.5	0	0
14617	6762830618	42734	2	1.00	1070	6120	1.0	0	0
14618	6762830709	42734	4	1.00	1030	6621	1.0	0	0
14619	6762831463	42734	3	1.00	900	4770	1.0	0	0

14620 rows × 23 columns

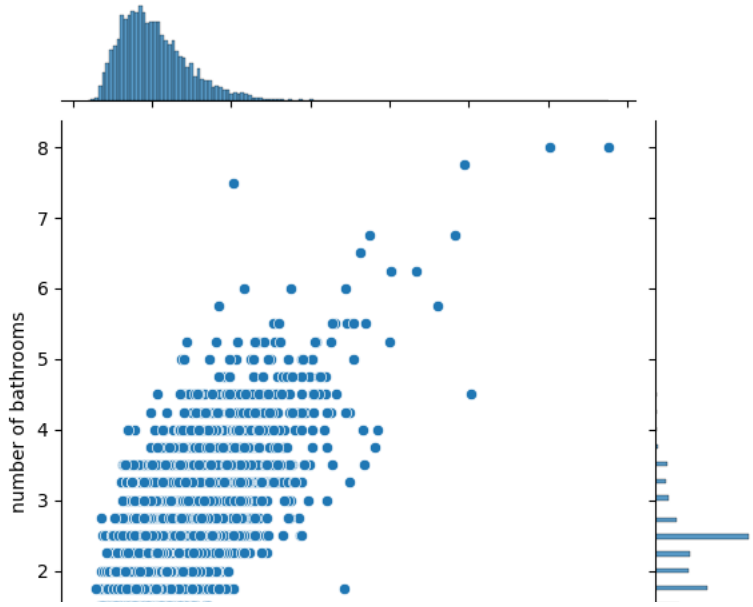
```
df.loc[:, ['living area']].plot()
```

```
<Axes: >
```



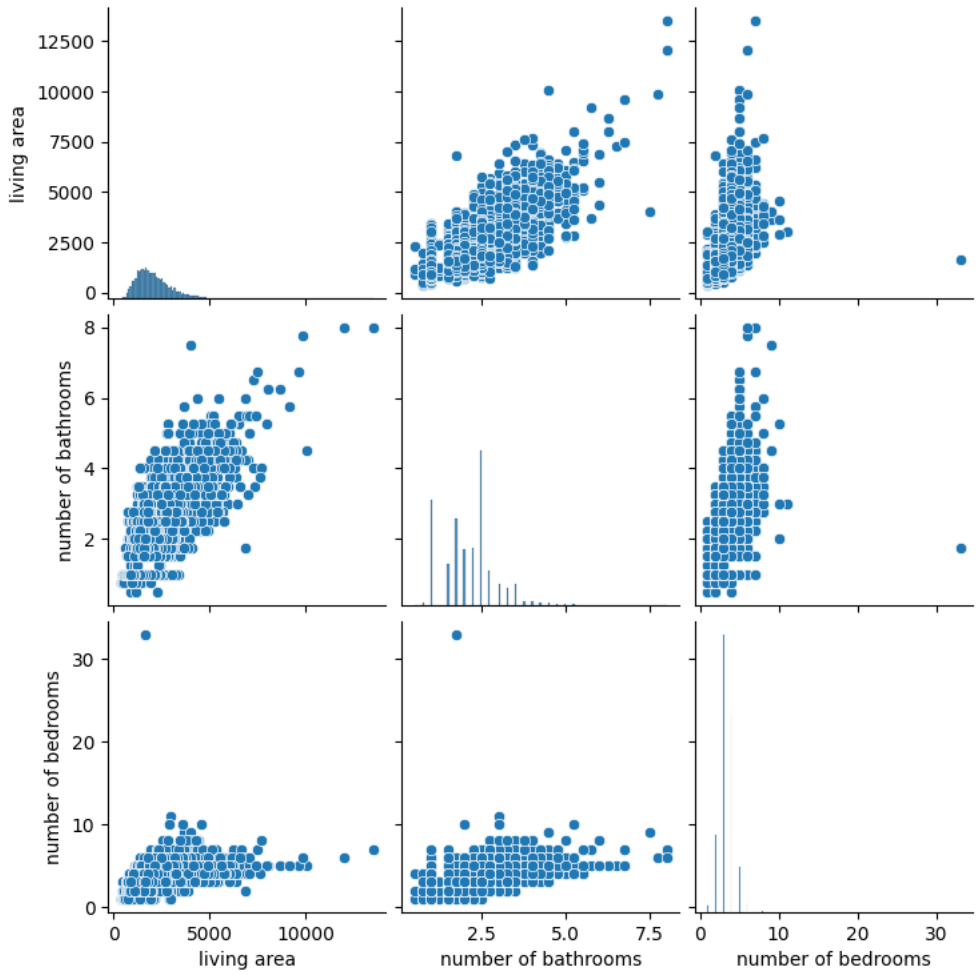

```
# Example of a joint plot
sns.jointplot(x='living area', y='number of bathrooms',data=df)
```

```
<seaborn.axisgrid.JointGrid at 0x7f9163fc3040>
```




```
#Pair Plot
sns.pairplot(df.loc[:,['living area','number of bathrooms','number of bedrooms']])
```

```
<seaborn.axisgrid.PairGrid at 0x78dcb9da7850>
```




```
df.describe() # Descriptive statistics
```

	id	Date	number of bedrooms	number of bathrooms	living area	lot area	number of floors
count	1.462000e+04	14620.000000	14620.000000	14620.000000	14620.000000	1.462000e+04	14620.000000
mean	6.762821e+09	42604.538646	3.379343	2.129583	2098.262996	1.509328e+04	1.502360
std	6.237575e+03	67.347991	0.938719	0.769934	928.275721	3.791962e+04	0.540239
min	6.762810e+09	42491.000000	1.000000	0.500000	370.000000	5.200000e+02	1.000000
25%	6.762815e+09	42546.000000	3.000000	1.750000	1440.000000	5.010750e+03	1.000000
50%	6.762821e+09	42600.000000	3.000000	2.250000	1930.000000	7.620000e+03	1.500000
75%	6.762826e+09	42662.000000	4.000000	2.500000	2570.000000	1.080000e+04	2.000000

```
df.isnull().any() # checking is there any null values in our dataset
# Since no NULL value hence no need to do anything
```

```
id                                    False
Date                                  False
number of bedrooms                    False
number of bathrooms                   False
living area                           False
lot area                              False
number of floors                      False
waterfront present                    False
number of views                       False
condition of the house                False
grade of the house                    False
Area of the house(excluding basement)    False
Area of the basement                  False
Built Year                            False
Renovation Year                       False
Postal Code                           False
Lattitude                             False
Longitude                             False
living_area_renov                     False
lot_area_renov                        False
Number of schools nearby              False
Distance from the airport             False
Price                                 False
dtype: bool
```