assignment-2

September 3, 2023

[4]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

[5]:
```
df = pd.read_csv('./dataset.csv')
```

[6]:
```
df.head(5)
```

[6]:

	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	Longitude	living_area_renov	lot_area_renov	Number of schools nearby	Distance from the airport	Price
0	6762810145	42491	5	2.50	3650	9050	2.0	0	4	5	…	1921	0	122003	52.8645	-114.557	2880	5400			
1	6762810635	42491	4	2.50	2920	4000	1.5	0	0	5	…	1909	0	122004	52.8878	-114.470	2470	4000			
2	6762810998	42491	5	2.75	2910	9480	1.5	0	0	3	…	1939	0	122004	52.8852	-114.468	2940	6600			
3	6762812605	42491	4	2.50	3310	42998	2.0	0	0	3	…	2001	0	122005	52.9532	-114.321	3350	42847			
4	6762812919	42491	3	2.00	2710	4500	1.5	0	0	4	…	1929	0	122006	52.9047	-114.485	2060	4500			

```
0                          2                          58   2380000
1                          2                          51   1400000
2                          1                          53   1200000
3                          3                          76    838000
4                          1                          51    805000

[5 rows x 23 columns]
```

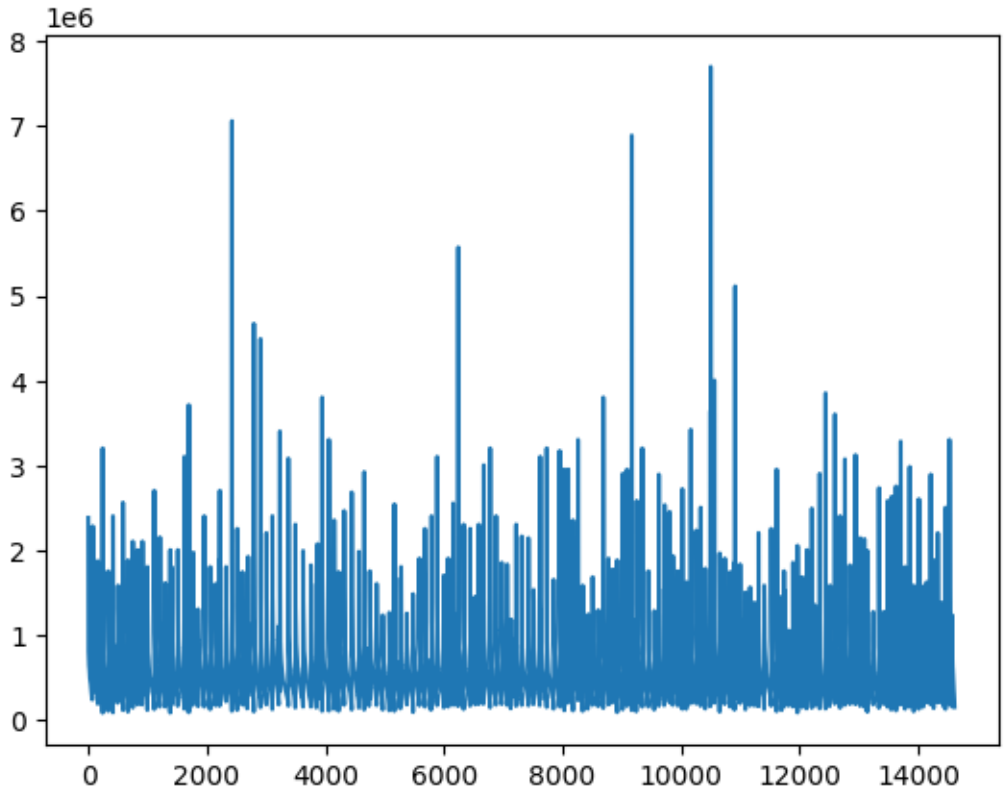
[21]:
```
df.shape
```

[21]:
```
(14620, 23)
```

##Univariate Analysis

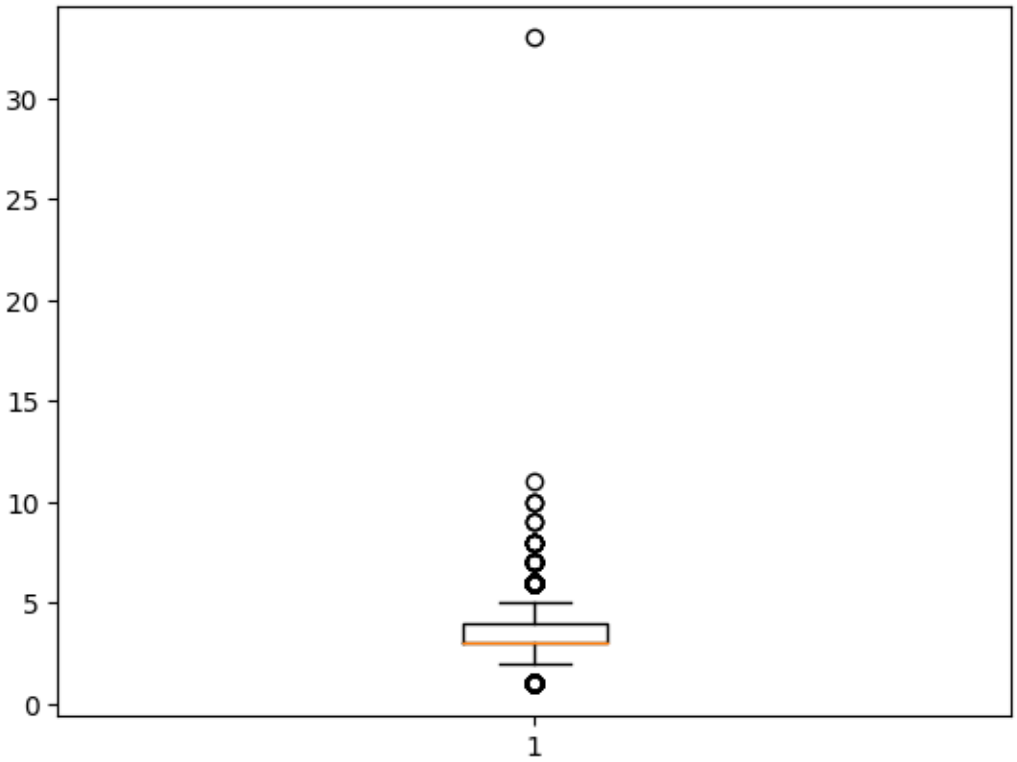
[9]:
```
df.Price.plot()
```

[9]:
```
<Axes: >
```



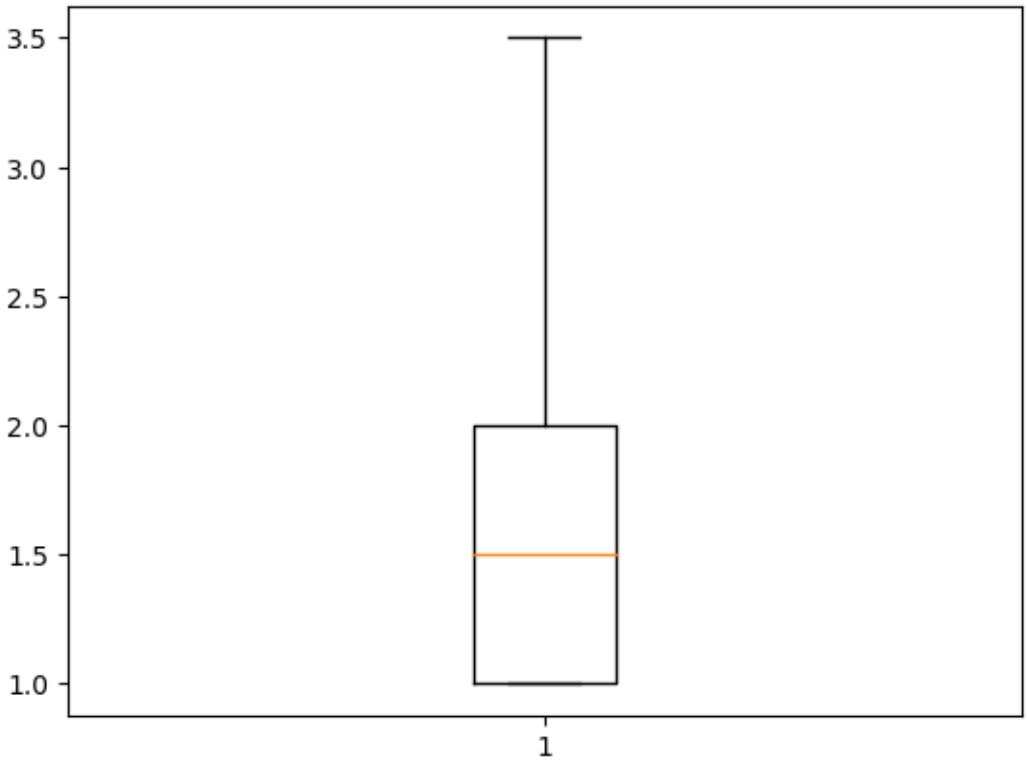
[47]:
```
plt.boxplot(df['number of bedrooms'])
```

[47]:
```
{'whiskers': [<matplotlib.lines.Line2D at 0x7ab8cb6150c0>,
  <matplotlib.lines.Line2D at 0x7ab8cb615360>],
 'caps': [<matplotlib.lines.Line2D at 0x7ab8cb6154e0>,
  <matplotlib.lines.Line2D at 0x7ab8cb615780>],
 'boxes': [<matplotlib.lines.Line2D at 0x7ab8cb614e20>],
 'medians': [<matplotlib.lines.Line2D at 0x7ab8cb615a20>],
 'fliers': [<matplotlib.lines.Line2D at 0x7ab8cb615cc0>],
 'means': []}
```



[53]:
```
plt.boxplot(df['number of floors'])
```

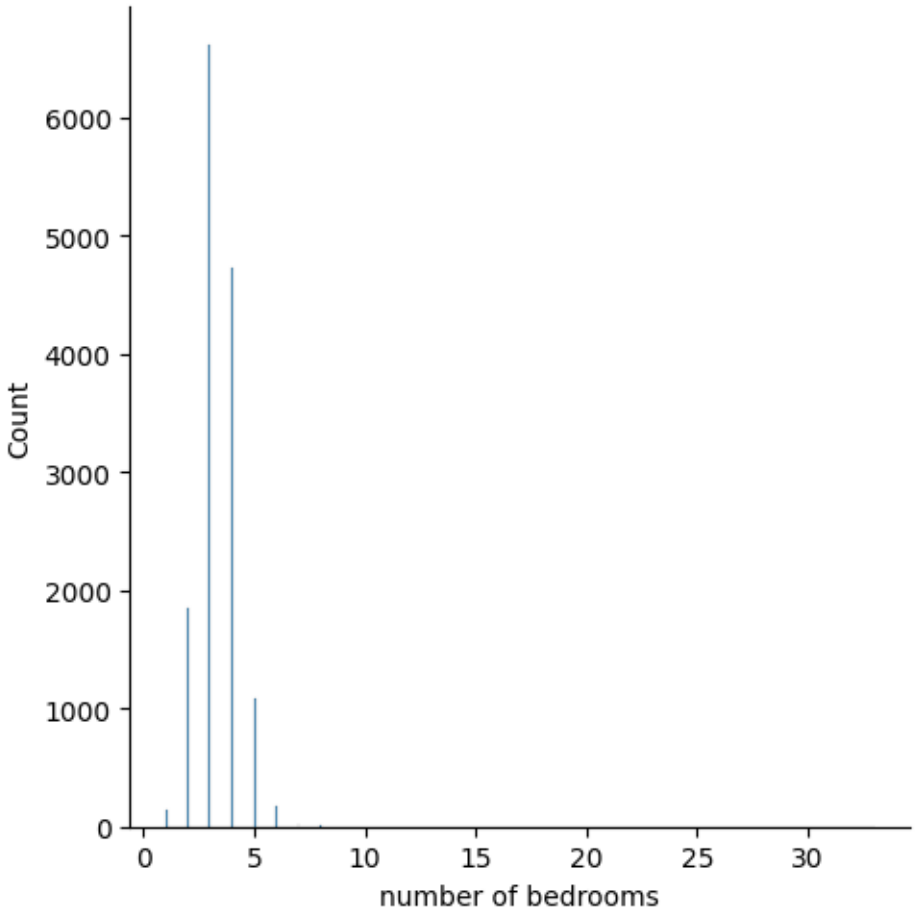
[53]:
```
{'whiskers': [<matplotlib.lines.Line2D at 0x7ab8cb2ca2c0>,
  <matplotlib.lines.Line2D at 0x7ab8cb2ca560>],
 'caps': [<matplotlib.lines.Line2D at 0x7ab8cb2ca800>,
  <matplotlib.lines.Line2D at 0x7ab8cb2caaa0>],
 'boxes': [<matplotlib.lines.Line2D at 0x7ab8cb2ca020>],
 'medians': [<matplotlib.lines.Line2D at 0x7ab8cb2cad40>],
 'fliers': [<matplotlib.lines.Line2D at 0x7ab8cb2cafe0>],
 'means': []}
```



0.1 Bi-variate Analysis

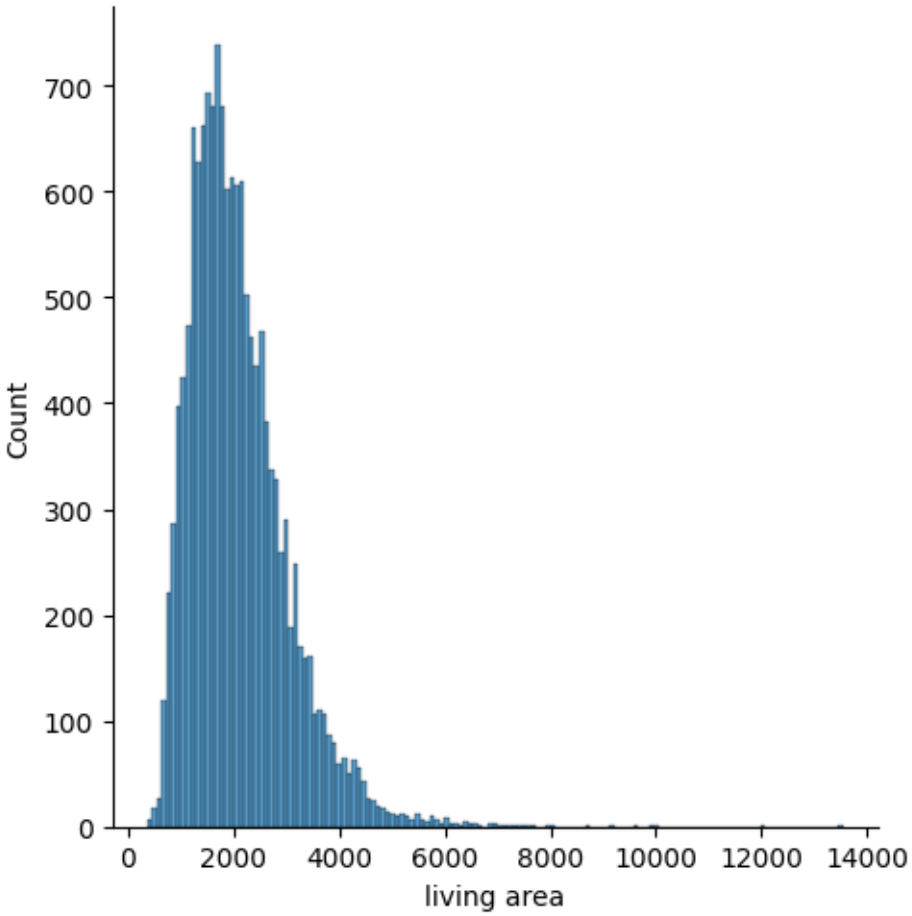
[40]:
```
sns.displot(df['number of bedrooms'])
```

[40]:
```
<seaborn.axisgrid.FacetGrid at 0x7ab8e49a1900>
```



[41]:
```
sns.displot(df['living area'])
```

[41]: <seaborn.axisgrid.FacetGrid at 0x7ab8d8025930>



0.2 Multi-variate Analysis

[56]:
```
sns.heatmap(df.corr())
```

[56]: <Axes: >


id
Date
number of bedrooms
number of bathrooms
living area
lot area
number of floors
waterfront present
number of views
condition of the house
grade of the house
Area of the house(excluding basement)
Area of the basement
Built Year
Renovation Year
Postal Code
Lattitude
Longitude
living_area_renov
lot_area_renov
Number of schools nearby
Distance from the airport
Price
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6