```
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
```

id	Date	number_of_bedrooms	number_of_bathrooms	living_area	lot_area	number_of_floors	waterfront_present	numb
6762810145	42491	5	2.50	3650	9050	2.0	0	
6762810635	42491	4	2.50	2920	4000	1.5	0	
6762810998	42491	5	2.75	2910	9480	1.5	0	
6762812605	42491	4	2.50	3310	42998	2.0	0	
6762812919	42491	3	2.00	2710	4500	1.5	0	
...	...	...	...	...	...	...	...	
6762830250	42734	2	1.50	1556	20000	1.0	0	
6762830339	42734	3	2.00	1680	7000	1.5	0	
6762830618	42734	2	1.00	1070	6120	1.0	0	
6762830709	42734	4	1.00	1030	6621	1.0	0	
6762831463	42734	3	1.00	900	4770	1.0	0	

ws × 23 columns

```
df.head()
```

	id	Date	number_of_bedrooms	number_of_bathrooms	living_area	lot_area	number_of_floors	waterfront_present
0	6762810145	42491	5	2.50	3650	9050	2.0	0
1	6762810635	42491	4	2.50	2920	4000	1.5	0
2	6762810998	42491	5	2.75	2910	9480	1.5	0
3	6762812605	42491	4	2.50	3310	42998	2.0	0
4	6762812919	42491	3	2.00	2710	4500	1.5	0

5 rows × 23 columns

```
df.shape
```

```
(14620, 23)
```

```
df.describe()
```

	id	Date	number_of_bedrooms	number_of_bathrooms	living_area	lot_area	number_of_floors	wat
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	
max	6.762832e+09	42734.000000	33.000000	8.000000	13540.000000	1.074218e+06	3.500000	

8 rows × 23 columns

```
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
```

```
df.isnull().any()
```

```
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
```

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

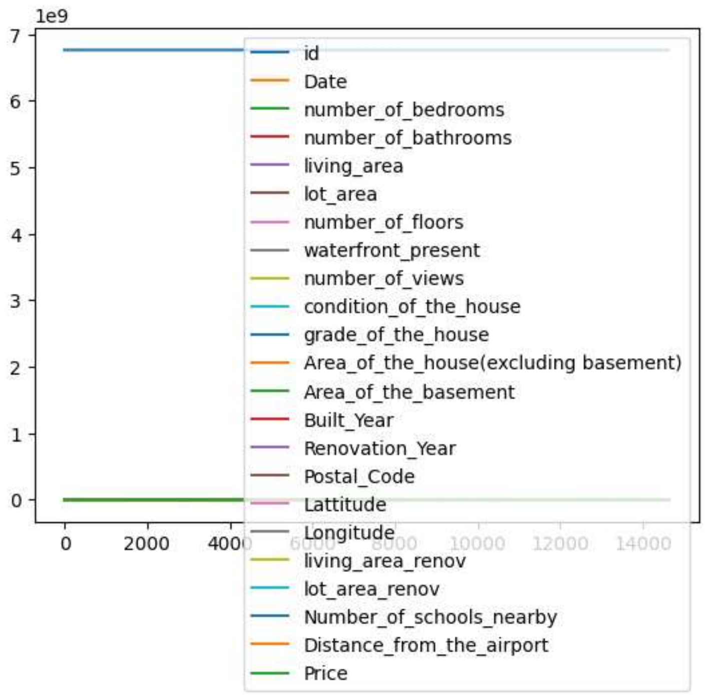
```
id                                       0
Date                                     0
number_of_bedrooms                       0
number_of_bathrooms                      0
living_area                              0
lot_area                                 0
number_of_floors                         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
```

Double-click (or enter) to edit

Univariate Analysis

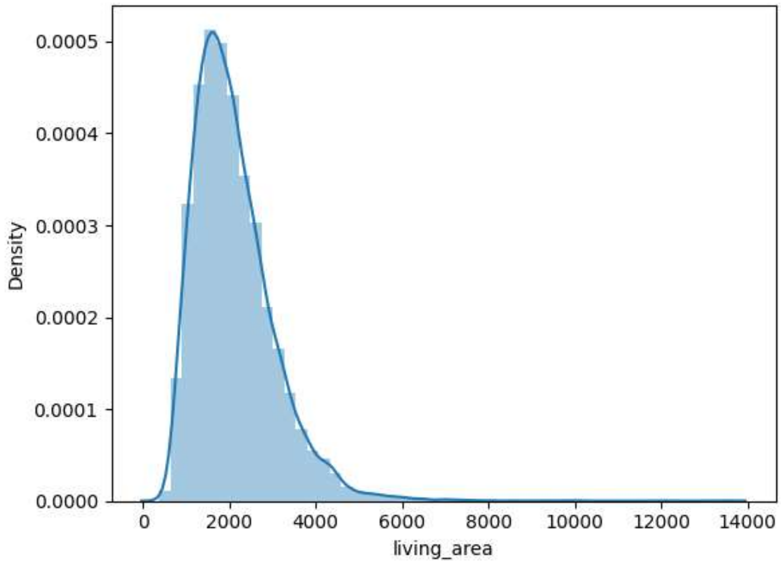
```
df.plot()
```

```
<Axes: >
```



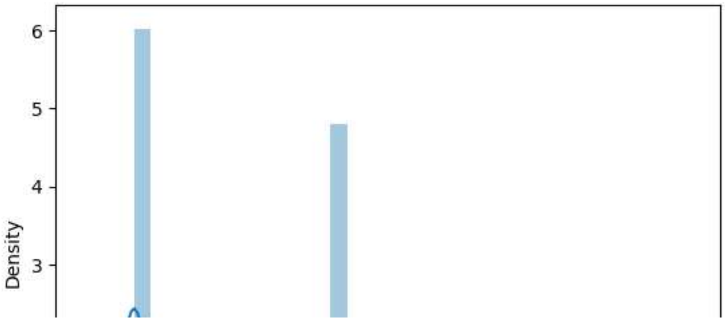
```
sns.distplot(df.living_area)
```

```
<ipython-input-13-2fe1fc3439c6>: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)
<Axes: xlabel='living_area', ylabel='Density'>
```



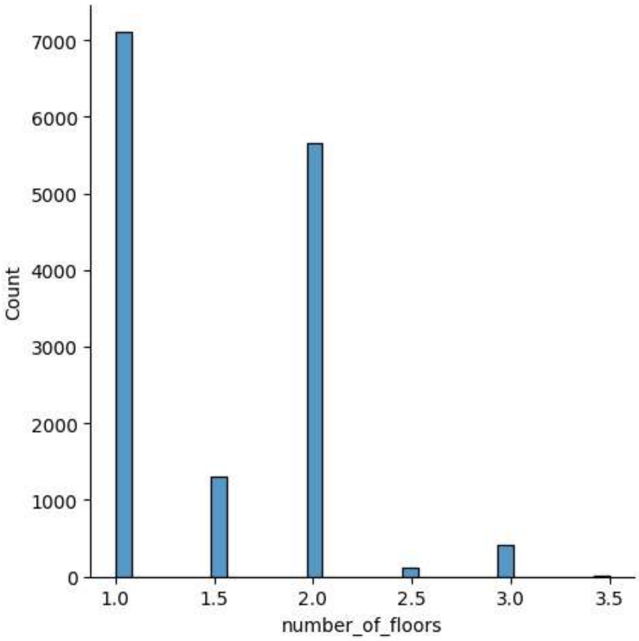
```
sns.distplot(df['number_of_floors'])
```

```
<ipython-input-15-6342cd93d98e>: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['number_of_floors'])
<Axes: xlabel='number_of_floors', ylabel='Density'>
```



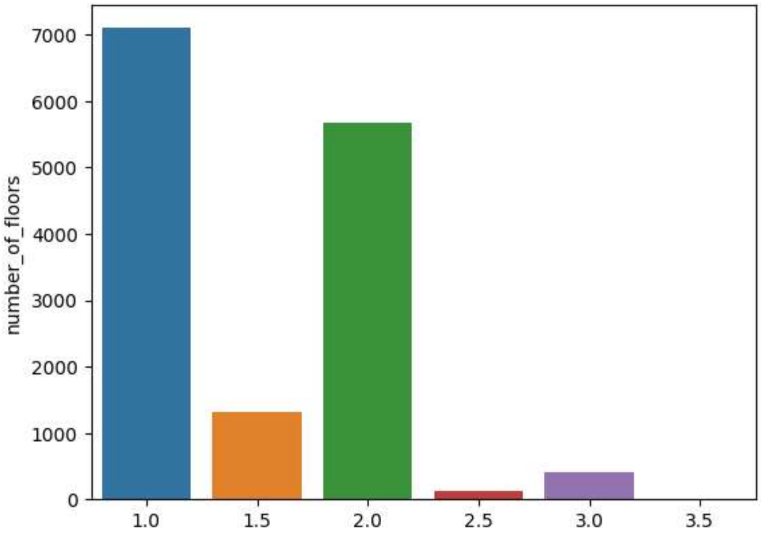
```
sns.displot(df['number_of_floors'])
```

```
<seaborn.axisgrid.FacetGrid at 0x7e7e042c5b70>
```



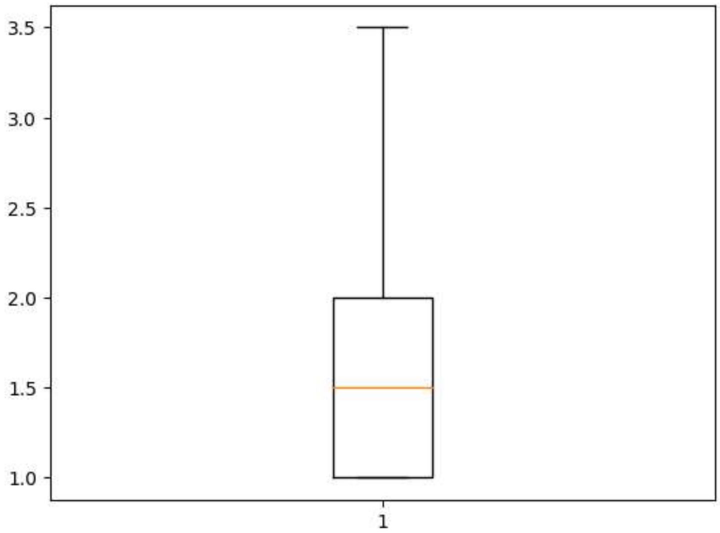
```
sns.barplot(x = df['number_of_floors'].value_counts().index, y = df['number_of_floors'].value_counts())
```

```
<Axes: ylabel='number_of_floors'>
```



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

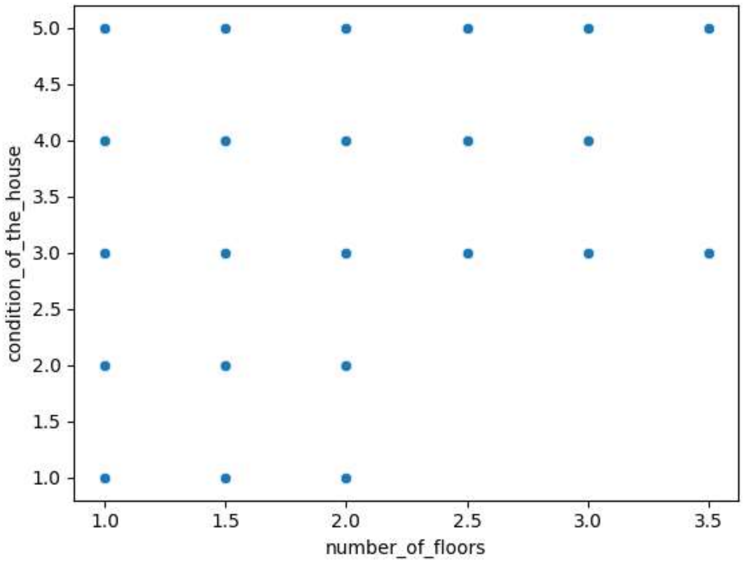
```
{'whiskers': [<matplotlib.lines.Line2D at 0x7e7e000983d0>,
  <matplotlib.lines.Line2D at 0x7e7e000991b0>],
 'caps': [<matplotlib.lines.Line2D at 0x7e7dfdd0b7c0>,
  <matplotlib.lines.Line2D at 0x7e7dfdb80820>],
 'boxes': [<matplotlib.lines.Line2D at 0x7e7e0009a680>],
 'medians': [<matplotlib.lines.Line2D at 0x7e7dfdc83130>],
 'fliers': [<matplotlib.lines.Line2D at 0x7e7dfdc82bf0>],
 'means': []}
```



Bivariate Analysis

```
sns.scatterplot(x = df['number_of_floors'], y = df['condition_of_the_house'])
```

```
<Axes: xlabel='number_of_floors', ylabel='condition_of_the_house'>
```



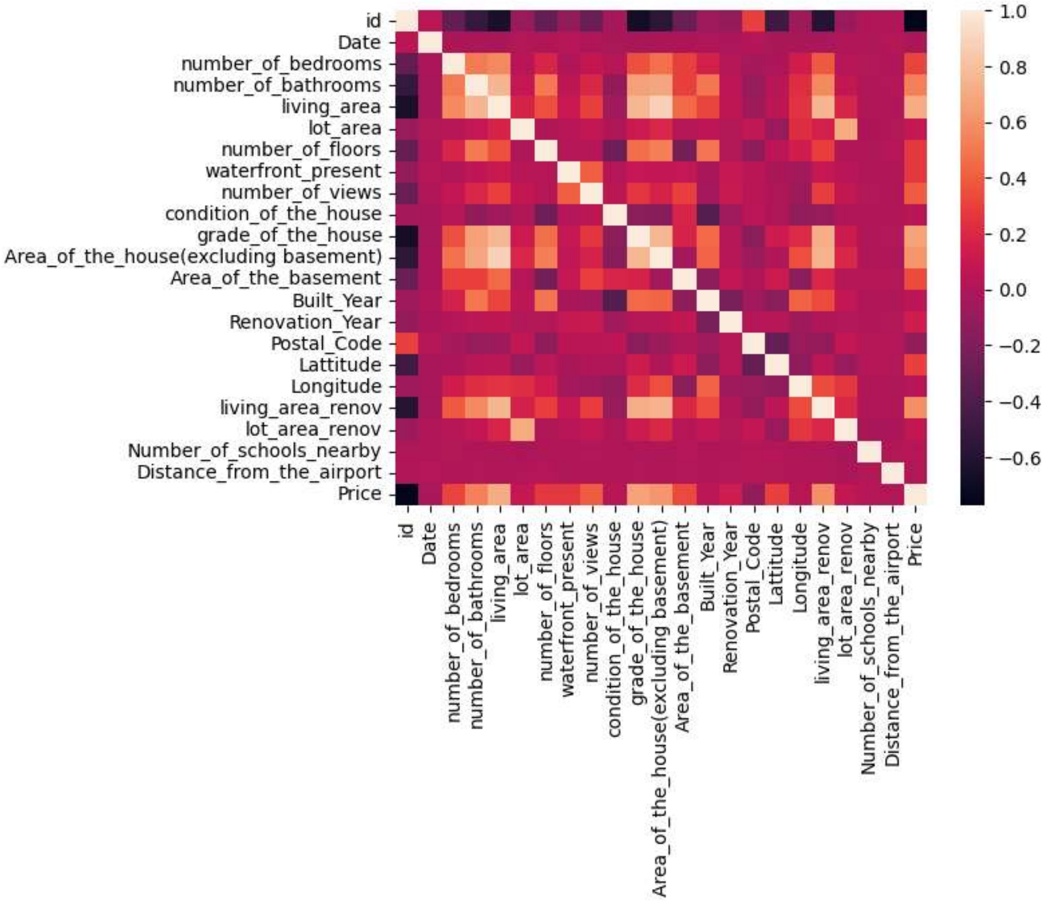
```
sns.lineplot(x = df['number_of_floors'], y = df['condition_of_the_house'])
```

```
<Axes: xlabel='number_of_floors', ylabel='condition_of_the_house'>
```

Multivariate Analysis

```
sns.heatmap(df.corr())
```

```
<Axes: >
```



```
df.describe()
```

```
                  id           Date  number_of_bedrooms  number_of_bathrooms   living_area      lot_area  number_of_floors  waterfront_
count   1.462000e+04   14620.000000        14620.000000         14620.000000  14620.000000  1.462000e+04      14620.000000         1462(
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             (
max     6.762832e+09   42734.000000           33.000000             8.000000  13540.000000  1.074218e+06          3.500000             '
8 rows × 23 columns
```

there are no missing values

✓ 0s completed at 5:23 PM