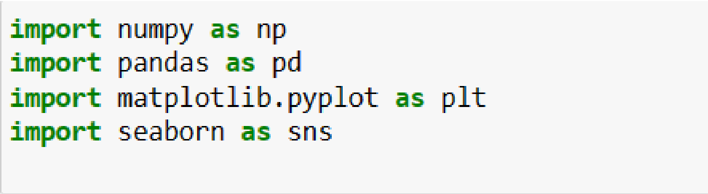
Name: - Shresthi Singh

Reg No.: - 21BCE2682

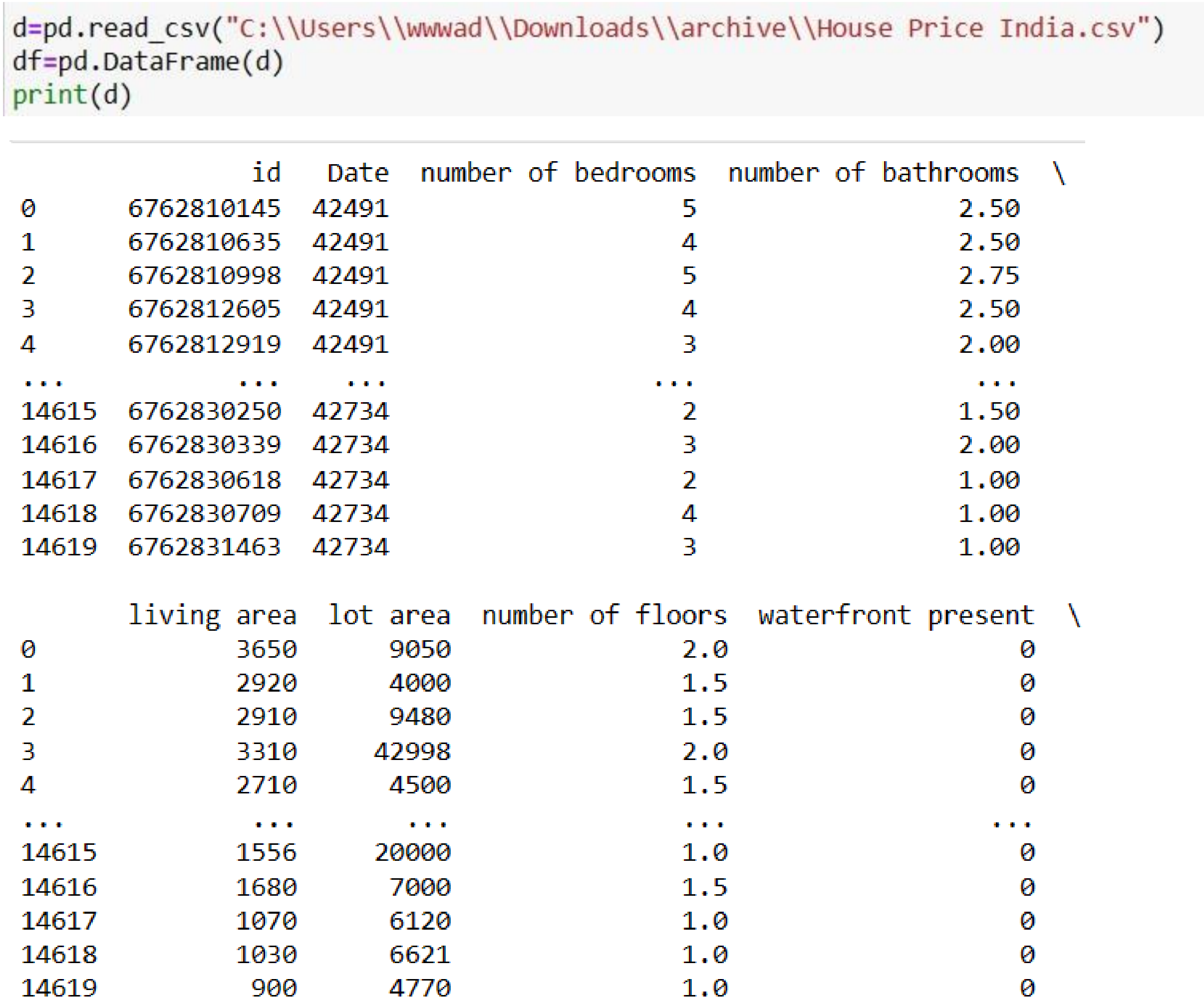
Assignment 2

1. *Download the dataset: Dataset*



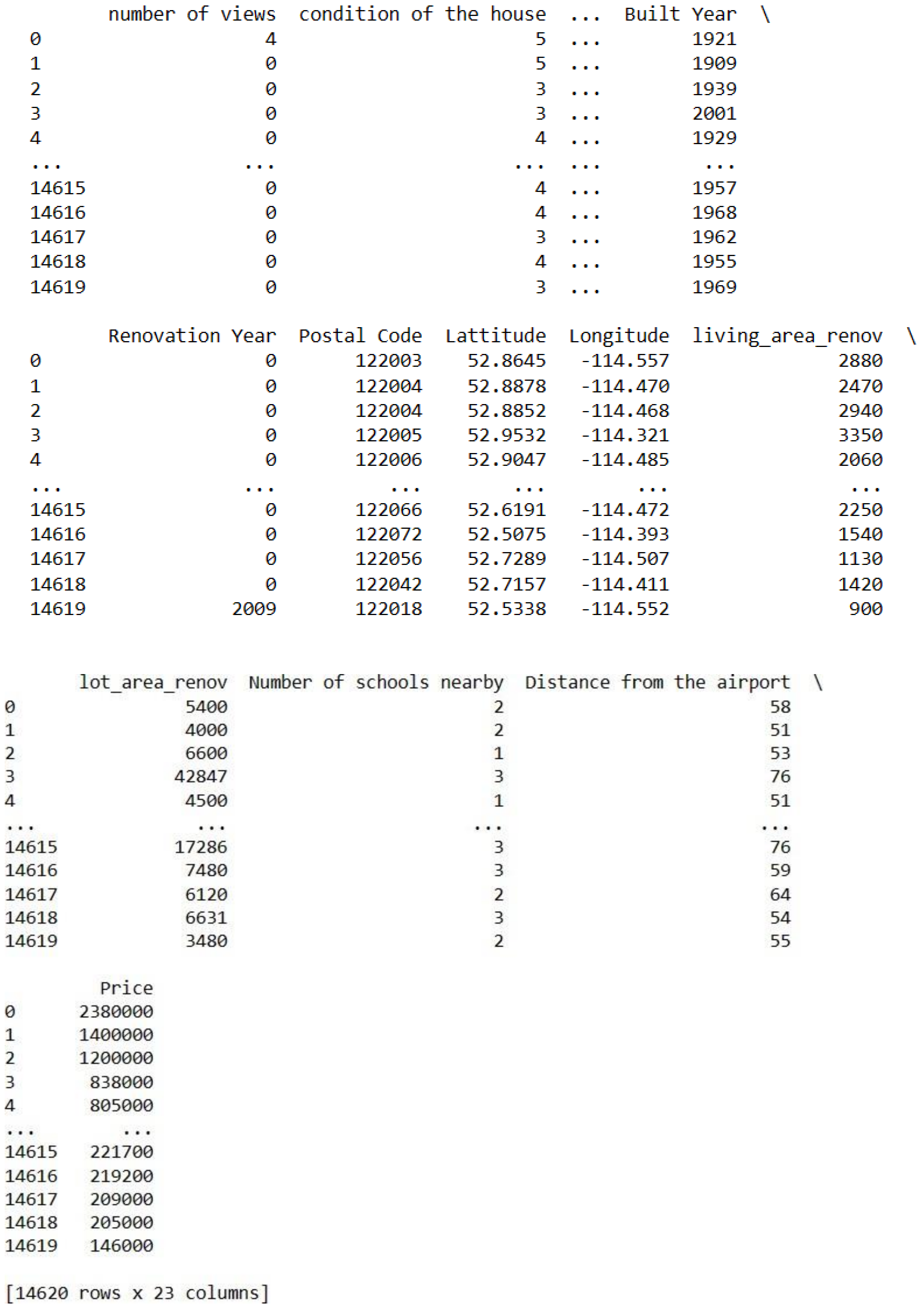
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

2. *Load the dataset.*



```
d=pd.read_csv("C:\\Users\\wwwad\\Downloads\\archive\\House Price India.csv")
df=pd.DataFrame(d)
print(d)
```

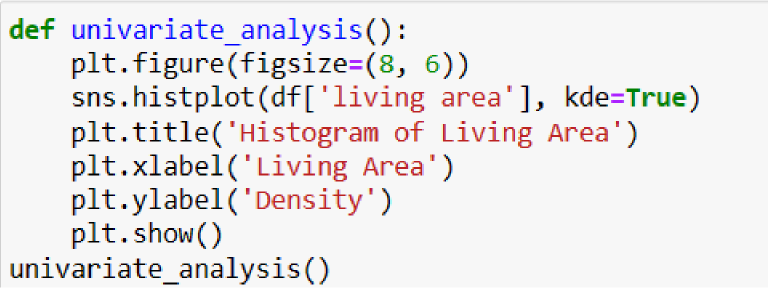
```
               id   Date  number of bedrooms  number of bathrooms  \
0      6762810145  42491                   5                 2.50   
1      6762810635  42491                   4                 2.50   
2      6762810998  42491                   5                 2.75   
3      6762812605  42491                   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   
```



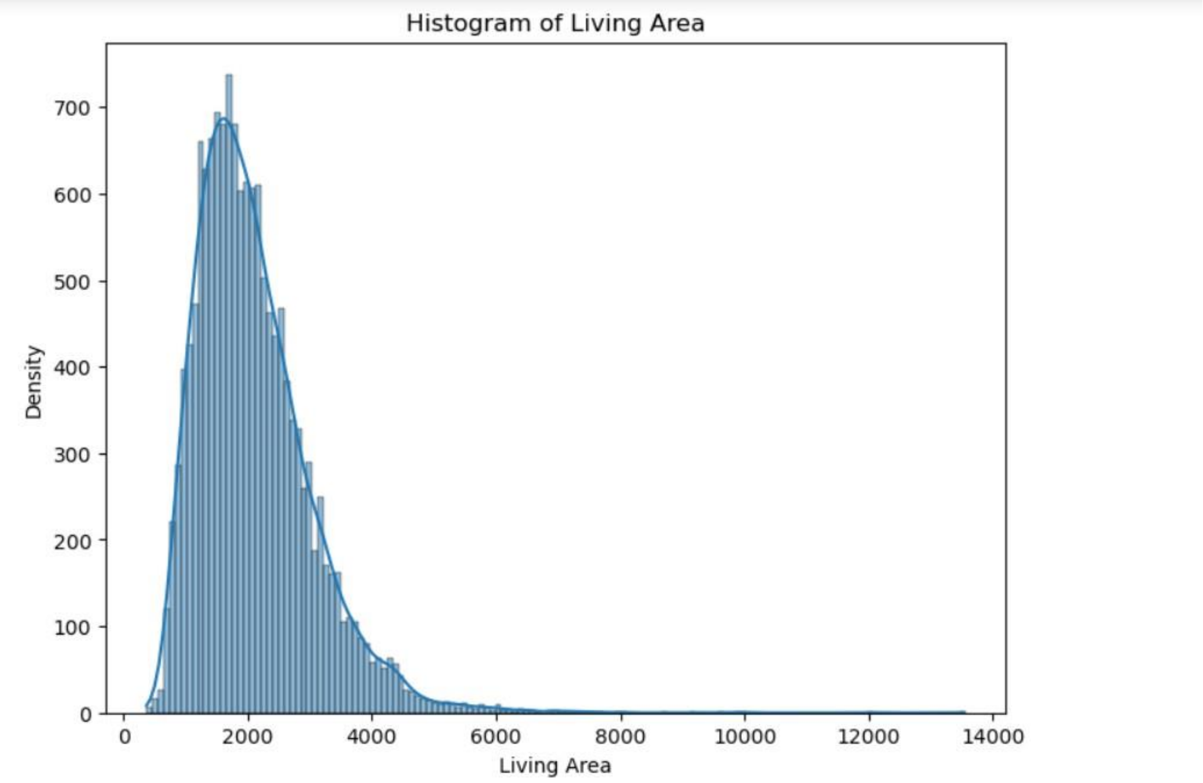
```
       number of views  condition of the house  ...  Built Year  \
0                    4                       5  ...        1921
1                    0                       5  ...        1909
2                    0                       3  ...        1939
3                    0                       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               1130
14618                0       122042    52.7157   -114.411               1420
14619             2009       122018    52.5338   -114.552                900


       lot_area_renov  Number of schools nearby  Distance from the airport  \
0                5400                         2                         58
1                4000                         2                         51
2                6600                         1                         53
3               42847                         3                         76
4                4500                         1                         51
...               ...                       ...                        ...
14615           17286                         3                         76
14616            7480                         3                         59
14617            6120                         2                         64
14618            6631                         3                         54
14619            3480                         2                         55

         Price
0      2380000
1      1400000
2      1200000
3       838000
4       805000
...        ...
14615   221700
14616   219200
14617   209000
14618   205000
14619   146000

[14620 rows x 23 columns]
```

3. *Perform the Below Visualizations.*

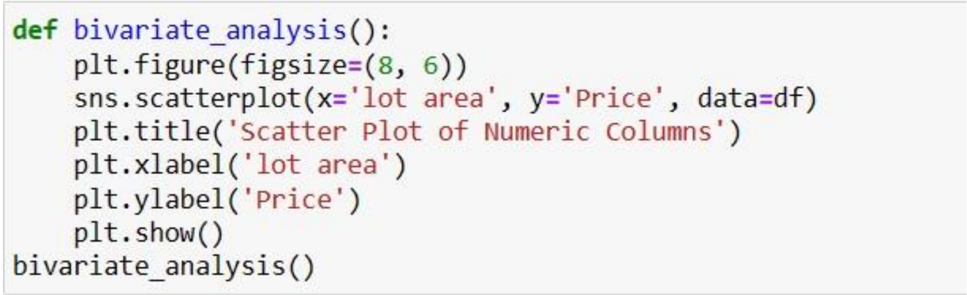
- ***Univariate Analysis***



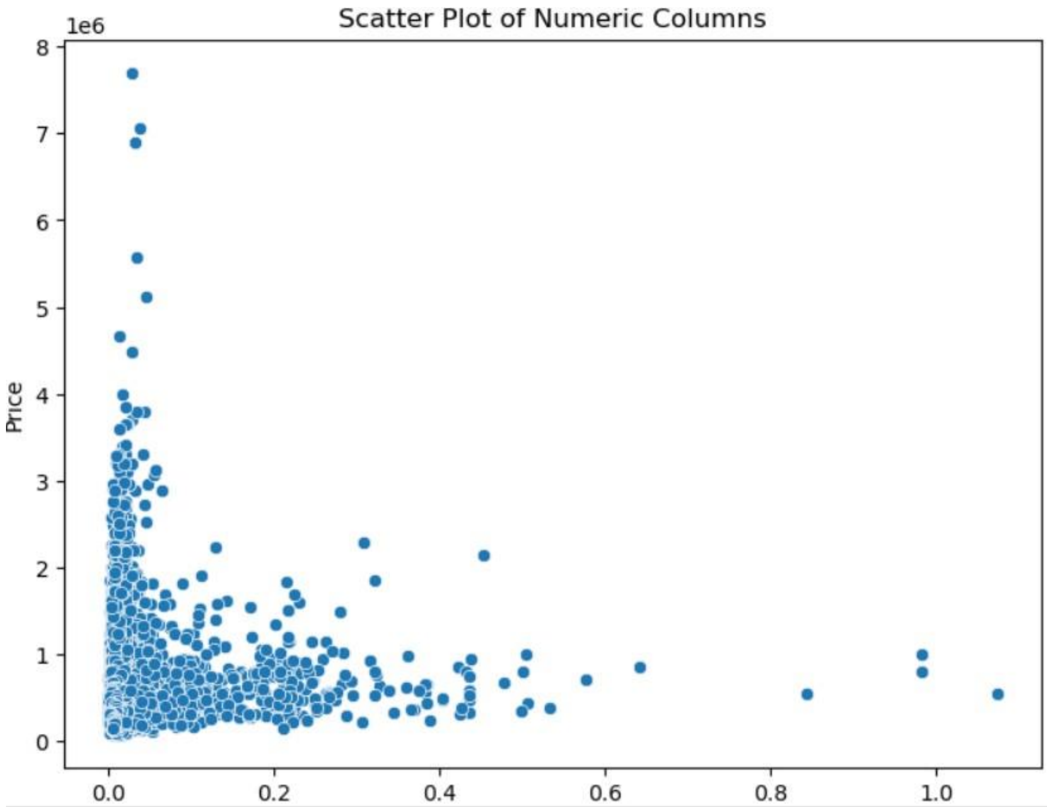
```
def univariate_analysis():
    plt.figure(figsize=(8, 6))
    sns.histplot(df['living area'], kde=True)
    plt.title('Histogram of Living Area')
    plt.xlabel('Living Area')
    plt.ylabel('Density')
    plt.show()
univariate_analysis()
```



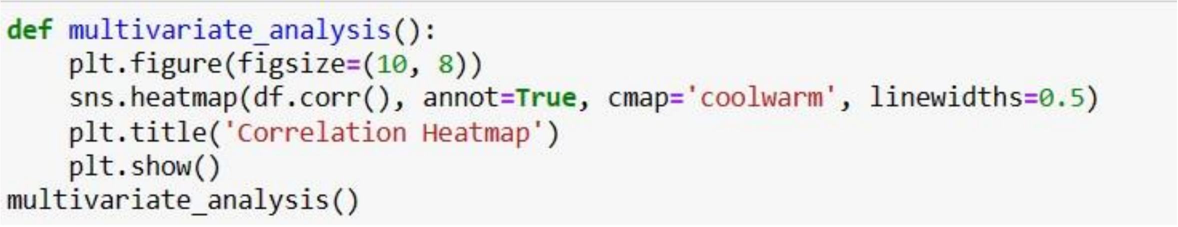
- ***Bi - Variate Analysis***



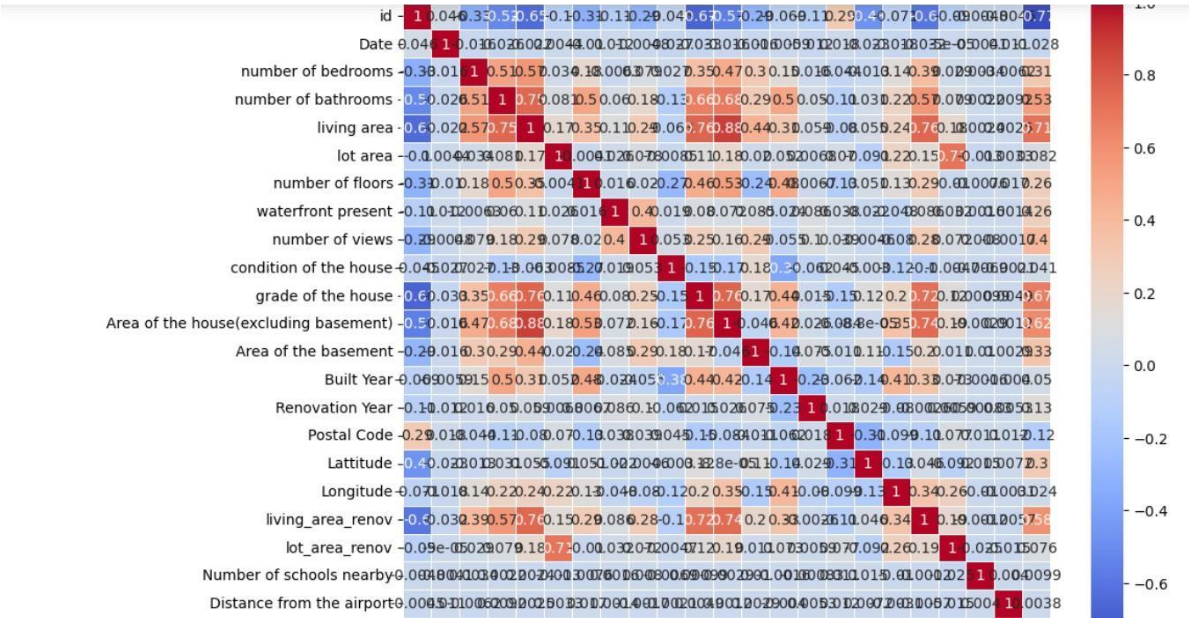
```
def bivariate_analysis():
    plt.figure(figsize=(8, 6))
    sns.scatterplot(x='lot area', y='Price', data=df)
    plt.title('Scatter Plot of Numeric Columns')
    plt.xlabel('lot area')
    plt.ylabel('Price')
    plt.show()
bivariate_analysis()
```



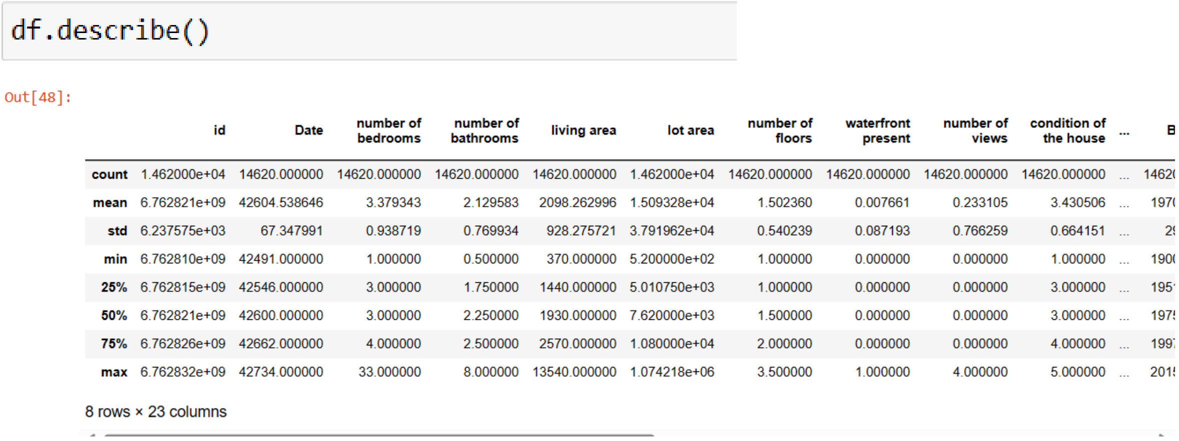
- ***Multivariate Analysis***



```
def multivariate_analysis():
    plt.figure(figsize=(10, 8))
    sns.heatmap(df.corr(), annot=True, cmap='coolwarm', linewidths=0.5)
    plt.title('Correlation Heatmap')
    plt.show()
multivariate_analysis()
```



4.Perform descriptive statistics on the dataset.



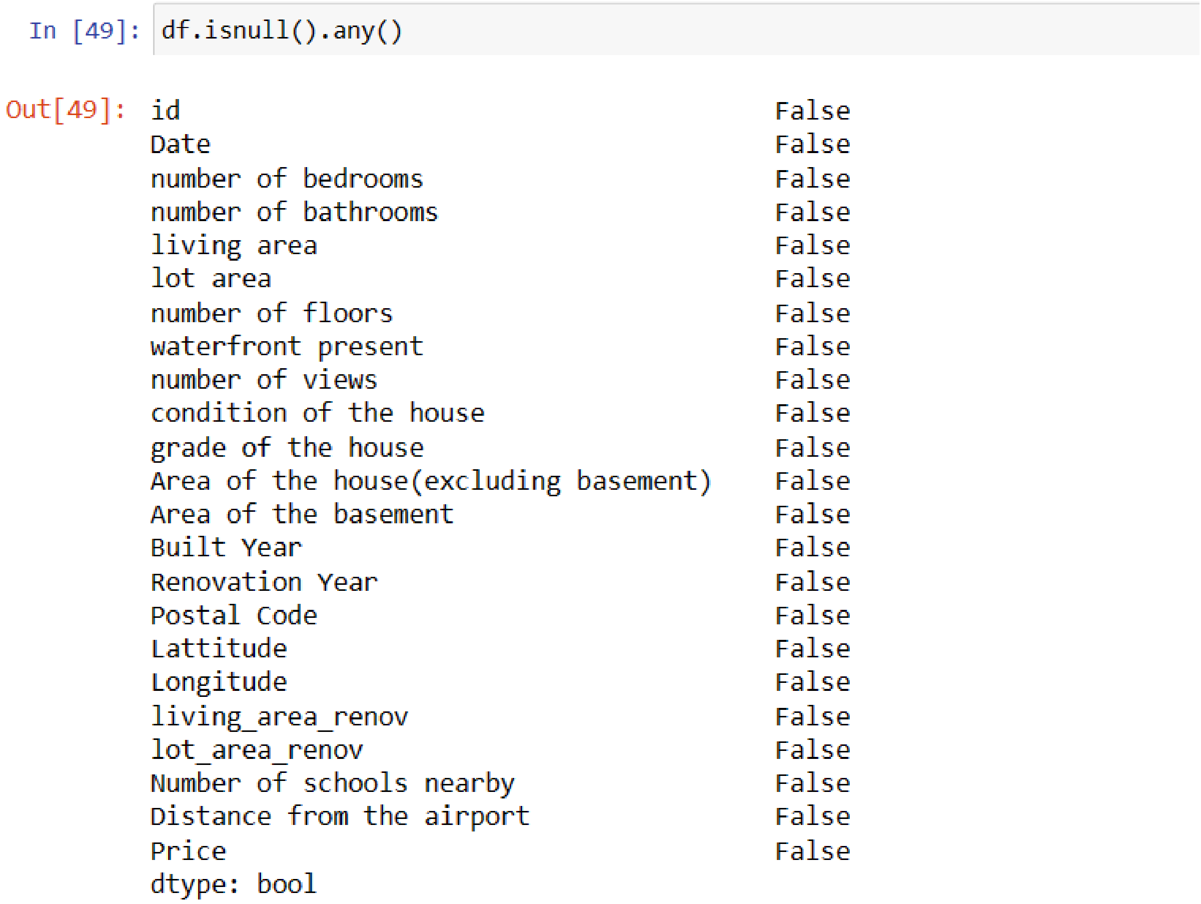
```
df.describe()
```

Out[48]:

	id	Date	number of bedrooms	number of bathrooms	living area	lot area	number of floors	waterfront present	number of views	condition of the house	...	B
count	1.462000e+04	14620.000000	14620.000000	14620.000000	14620.000000	1.462000e+04	14620.000000	14620.000000	14620.000000	14620.000000	...	1462
mean	6.762821e+09	42604.538646	3.379343	2.129583	2098.262996	1.509328e+04	1.502360	0.007661	0.233105	3.430506	...	197
std	6.237575e+03	67.347991	0.938719	0.769934	928.275721	3.791962e+04	0.540239	0.087193	0.766259	0.664151	...	2
min	6.762810e+09	42491.000000	1.000000	0.500000	370.000000	5.200000e+02	1.000000	0.000000	0.000000	1.000000	...	190
25%	6.762815e+09	42546.000000	3.000000	1.750000	1440.000000	5.010750e+03	1.000000	0.000000	0.000000	3.000000	...	195
50%	6.762821e+09	42600.000000	3.000000	2.250000	1930.000000	7.620000e+03	1.500000	0.000000	0.000000	3.000000	...	197
75%	6.762826e+09	42662.000000	4.000000	2.500000	2570.000000	1.080000e+04	2.000000	0.000000	0.000000	4.000000	...	199
max	6.762832e+09	42734.000000	33.000000	8.000000	13540.000000	1.074218e+06	3.500000	1.000000	4.000000	5.000000	...	201

8 rows × 23 columns

5. Handle the Missing values



```
In [49]: df.isnull().any()
```

```
Out[49]: 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
```