

Name - Angad Singh. Registration No. - 21BPS1420

Assignment 2:

```
# Task 1: Download the dataset
from google.colab import files
uploaded = files.upload()

Choose Files HousePriceIndia.csv
• HousePriceIndia.csv(text/csv) - 1524561 bytes, last modified: 9/6/2023 - 100% done
Saving HousePriceIndia.csv to HousePriceIndia (1).csv

# Task 2: Load the dataset
import pandas as pd
df = pd.read_csv ('HousePriceIndia.csv')
print (df.head ())

  id  Date  number of bedrooms  number of bathrooms  living area  \
0  6762810145  42491                  5             2.50      3650
1  6762810635  42491                  4             2.50      2920
2  6762810998  42491                  5             2.75      2910
3  6762812605  42491                  4             2.50      3310
4  6762812919  42491                  3             2.00      2710

  lot area  number of floors  waterfront present  number of views  \
0      9050            2.0          0                 0
1      4000            1.5          0                 0
2      9480            1.5          0                 0
3      42998           2.0          0                 0
4      4500            1.5          0                 0

  condition of the house  ...  Built Year  Renovation Year  Postal Code  \
0                  5  ...      1921          0      122003
1                  5  ...      1909          0      122004
2                  3  ...      1939          0      122004
3                  3  ...      2001          0      122005
4                  4  ...      1929          0      122006

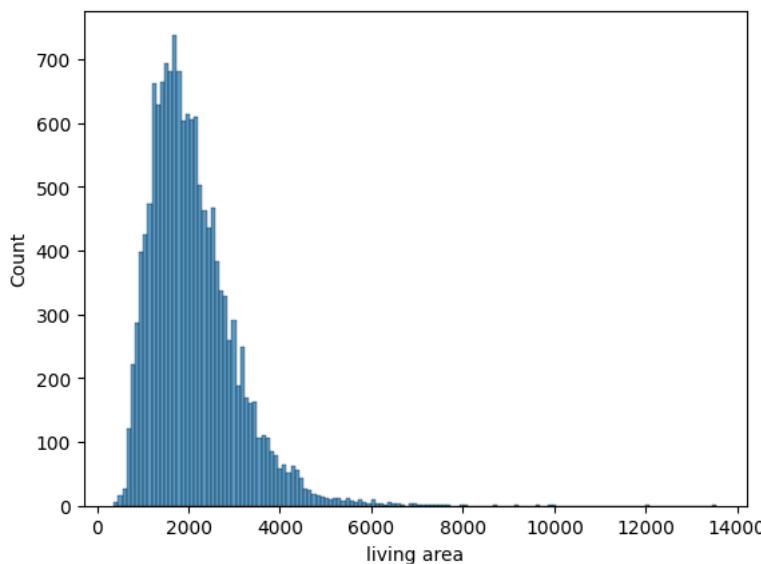
  Latitude  Longitude  living_area_renov  lot_area_renov  \
0    52.8645   -114.557          2880            5400
1    52.8878   -114.470          2470            4000
2    52.8852   -114.468          2940            6600
3    52.9532   -114.321          3350            42847
4    52.9047   -114.485          2060            4500

  Number of schools nearby  Distance from the airport  Price
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]

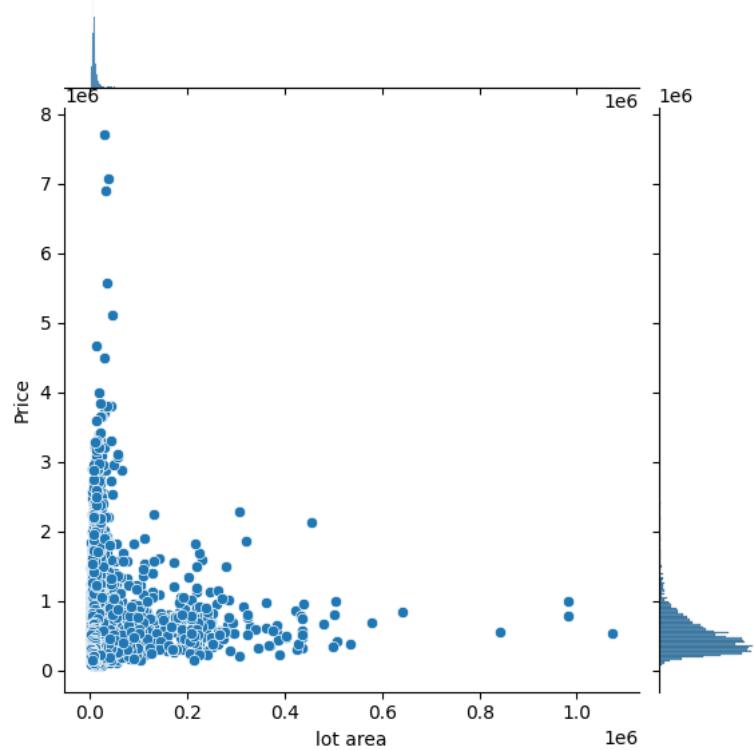
```
# Task 3 - 1: Univariate Analysis
import seaborn as sns
sns.histplot(df['living area'])
```

<Axes: xlabel='living area', ylabel='Count'>

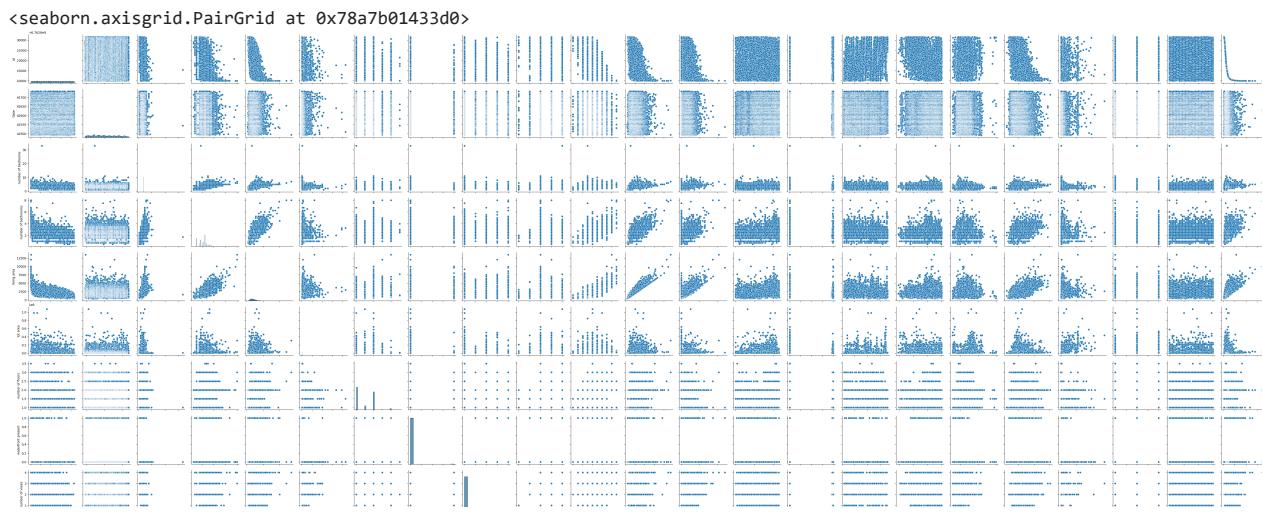


```
# Task 3 - 2: Bivariate Analysis  
sns.jointplot(x='lot area', y='Price', data=df)
```

```
<seaborn.axisgrid.JointGrid at 0x78a7b3a61600>
```



```
# Task 3 - 3: Multivariate Analysis  
sns.pairplot(df)
```



```
# Task 4: Check the descriptive statistics of 'df'
```

```
df.describe ()
```

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

8 rows × 23 columns



```
# Task 5: Handle the Missing values
```

```
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
Latitude                     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
```

✓ 0s completed at 11:44 PM

