

# ASSIGNMENT – 2

NEVATHA SRI.R

21BEC1803

VIT CHENNAI

## Assignment 2

Perform the Below Tasks to complete the assignment:-

Tasks:-

1. Download the dataset: [Dataset](#)
2. Load the dataset.
3. Perform the Below Visualizations.
  - Univariate Analysis
  - Bi - Variate Analysis
  - Multivariate Analysis
4. Perform descriptive statistics on the dataset.
5. Handle the Missing values.

### TASK – 1

CODE:

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

```
+ Code + Text All changes saved

ASSIGNMENT 2

NEVATHA SRI.R

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

## TASK – 2

```
df = pd.read_csv('/content/House Price India.csv')
df
```

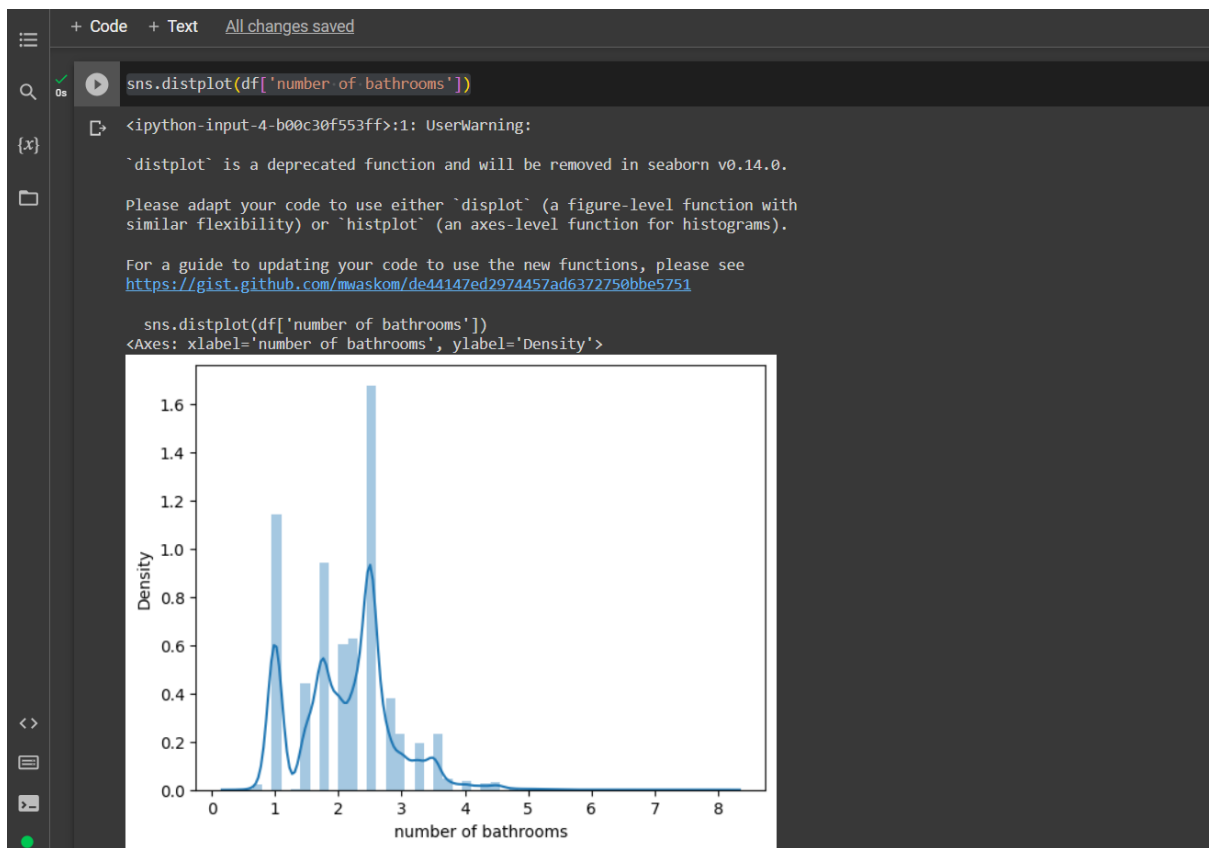
```
+ Code + Text All changes saved
[2] import numpy as np
import matplotlib.pyplot as plt
```

```
df = pd.read_csv('/content/House Price India.csv')
df
```

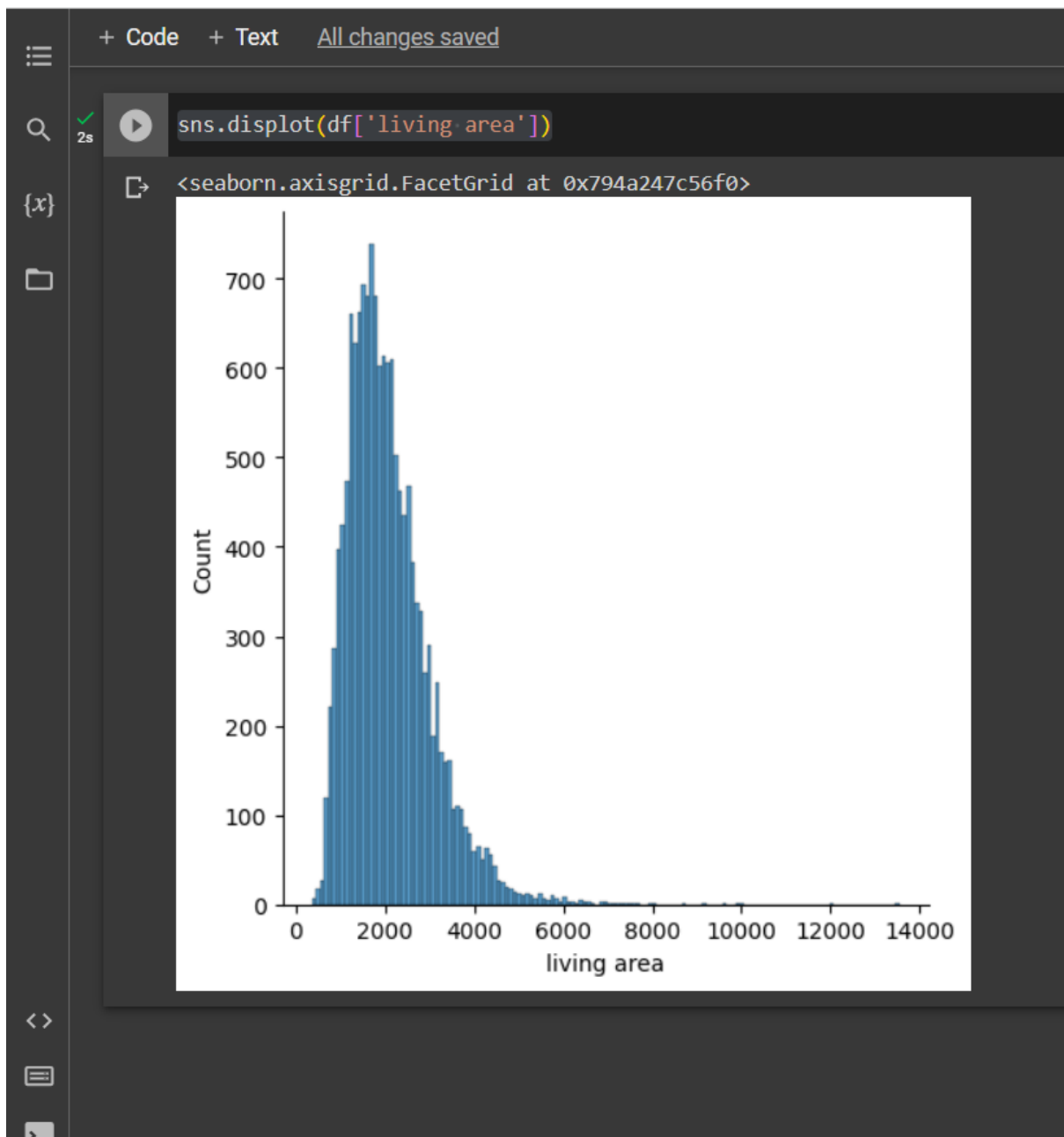
	id	Date	number of bedrooms	number of bathrooms	living area	lot area	number of floors	waterfront present	number of views	condition of the house	...	Built Year	Renovation Year	Postal Code	Latitude	Longitude	living_area_renov	lot_area_renov
0	6762810145	42491	5	2.50	3650	9050	2.0	0	4	5	...	1921	0	122003	52.8645	-114.557	2880	54
1	6762810635	42491	4	2.50	2920	4000	1.5	0	0	5	...	1909	0	122004	52.8878	-114.470	2470	40
2	6762810998	42491	5	2.75	2910	9480	1.5	0	0	3	...	1939	0	122004	52.8852	-114.468	2940	66
3	6762812605	42491	4	2.50	3310	42998	2.0	0	0	3	...	2001	0	122005	52.9532	-114.321	3350	428
4	6762812919	42491	3	2.00	2710	4500	1.5	0	0	4	...	1929	0	122006	52.9047	-114.485	2060	48
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
14615	6762830250	42734	2	1.50	1556	20000	1.0	0	0	4	...	1957	0	122066	52.6191	-114.472	2250	172
14616	6762830339	42734	3	2.00	1680	7000	1.5	0	0	4	...	1968	0	122072	52.5075	-114.393	1540	74
14617	6762830618	42734	2	1.00	1070	6120	1.0	0	0	3	...	1962	0	122056	52.7289	-114.507	1130	61
14618	6762830709	42734	4	1.00	1030	6621	1.0	0	0	4	...	1955	0	122042	52.7157	-114.411	1420	66
14619	6762831463	42734	3	1.00	900	4770	1.0	0	0	3	...	1969	2009	122018	52.5338	-114.552	900	34

14620 rows x 23 columns

```
sns.distplot(df['number of bathrooms'])
```



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



```
df['number of floors'].value_counts()
plt.pie(df['number of floors'].value_counts(), [0,0,0,0,0.2,0.4], labels =
['1.0', '2.0', '1.5', '3.0', '2.5', '3.5'], autopct = '%1.1f%%', shadow =
True)
plt.title('Number of floors')
plt.show()
```

✓  
0s



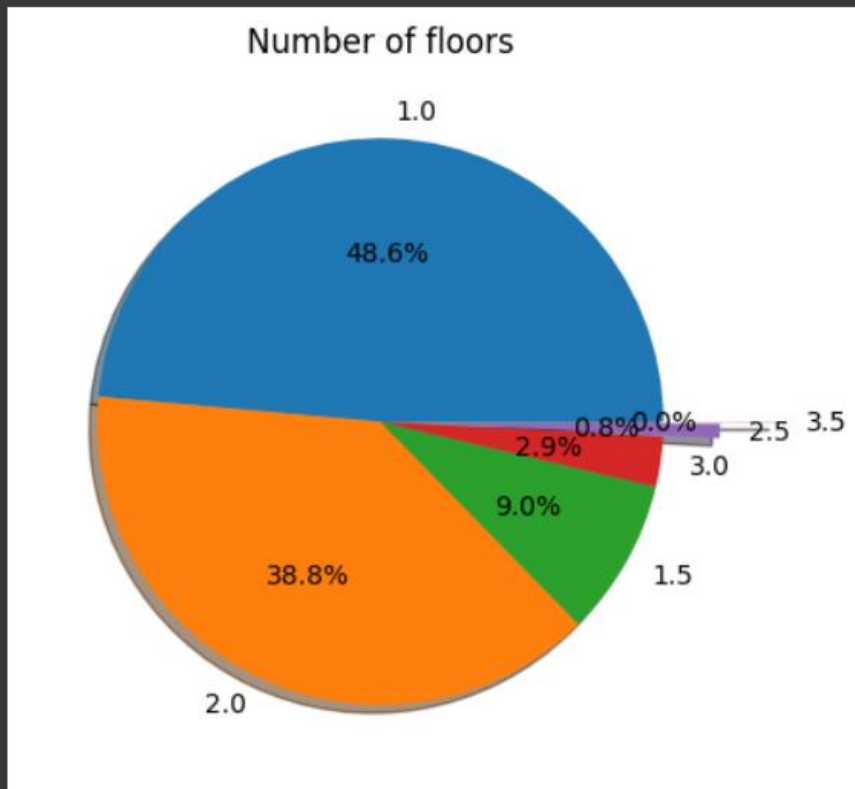
```
df['number of floors'].value_counts()
```

```
1.0    7103
2.0    5666
1.5    1311
3.0     418
2.5     118
3.5        4
Name: number of floors, dtype: int64
```

✓  
0s



```
plt.pie(df['number of floors'].value_counts(), [0,0,0,0,0.2,0.4], labels =
= ['1.0', '2.0', '1.5', '3.0', '2.5', '3.5'], autopct = '%1.1f%%', shadow =
True)
plt.title('Number of floors')
plt.show()
```



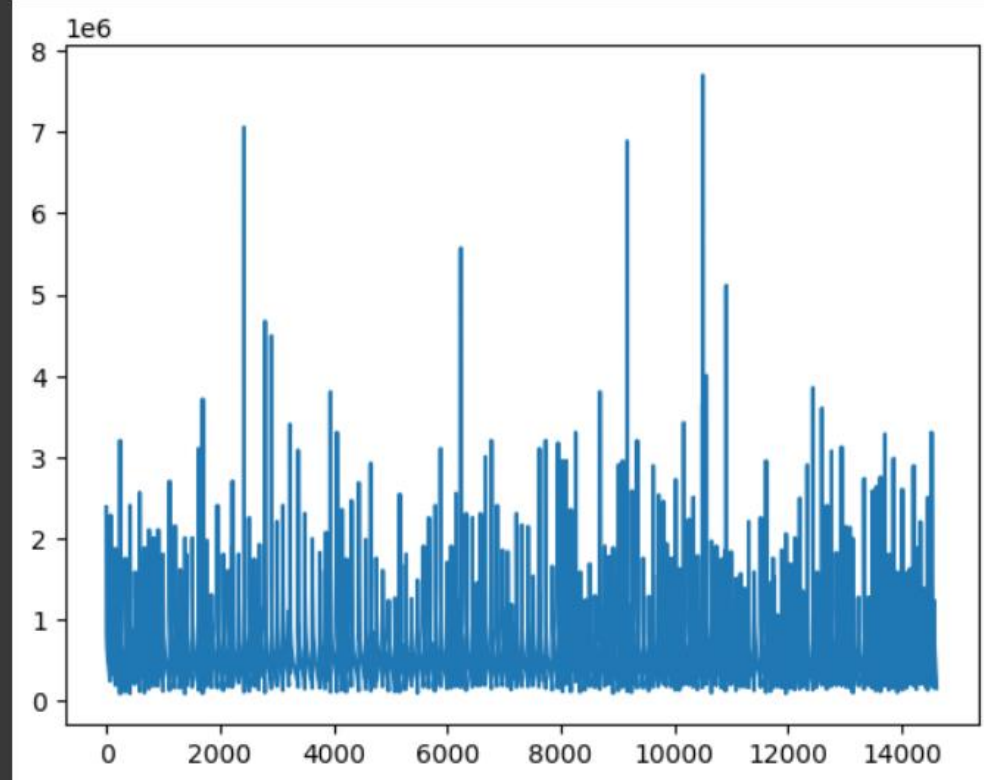
```
df['Price'].plot()
```

✓  
0s



```
df['Price'].plot()
```

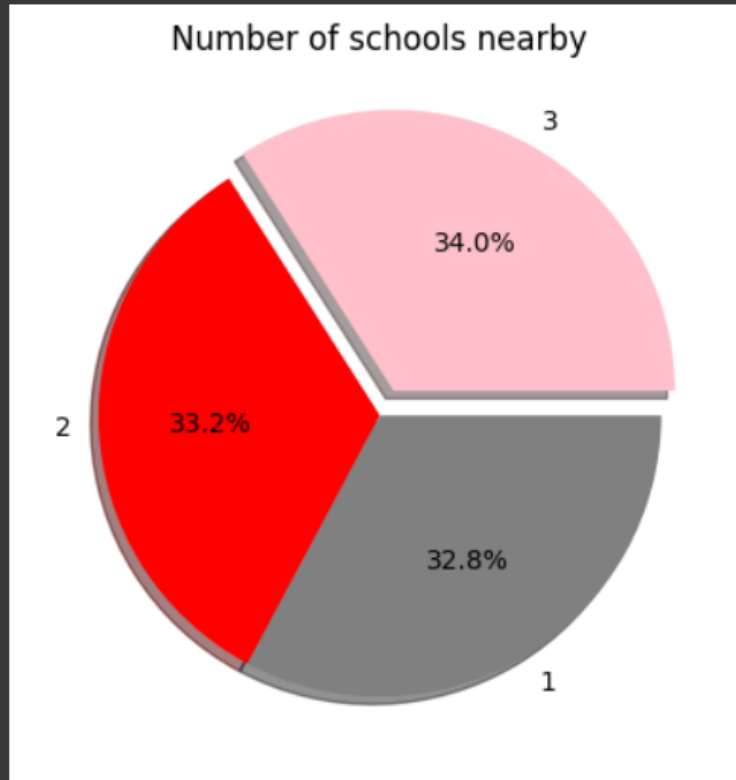
<Axes: >



0s



```
df['Number of schools nearby'].value_counts()
plt.pie(df['Number of schools nearby'].value_counts(),[0.1,0,0],labels
= ['3','2','1'],autopct = '%1.1f%%',shadow =
True,colors=['pink','red','grey'])
plt.title('Number of schools nearby')
plt.show()
```

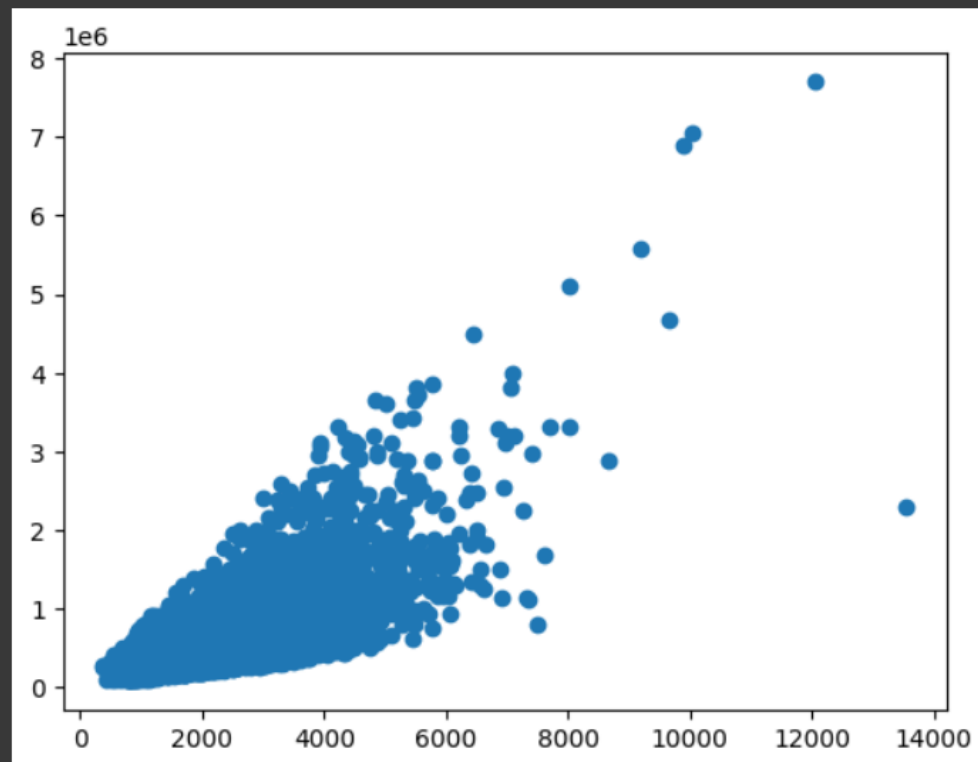


### TASK – 3 BIVARIATE ANALYSIS

✓  
0s



```
plt.scatter(df['living area'],df.Price)  
plt.show()
```





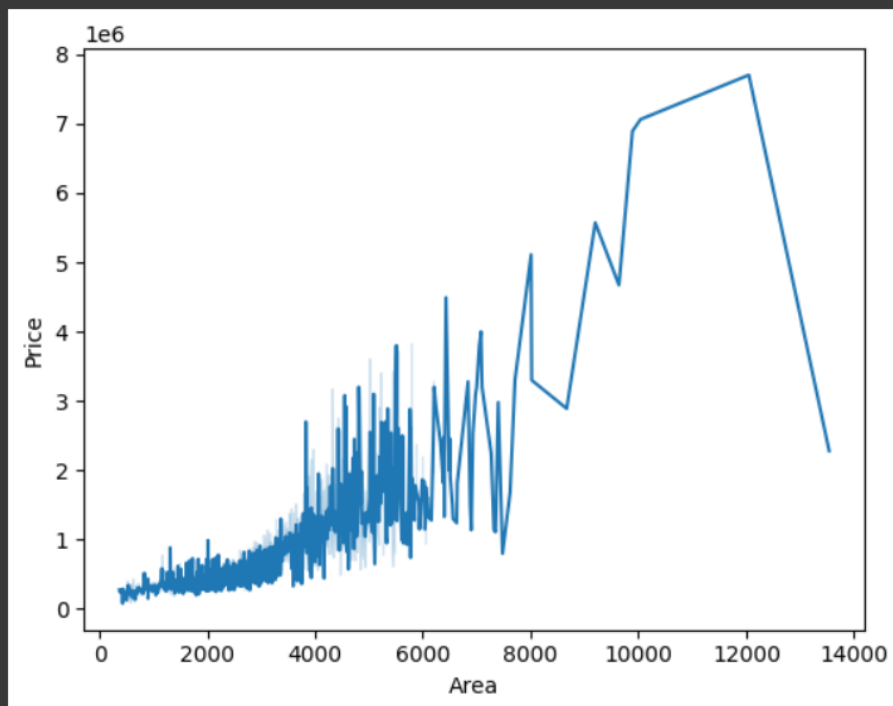


+ Code + Text



13s

```
sns.lineplot(x=df['living area'],y=df.Price)
plt.xlabel('Area')
plt.ylabel('Price')
plt.show()
```



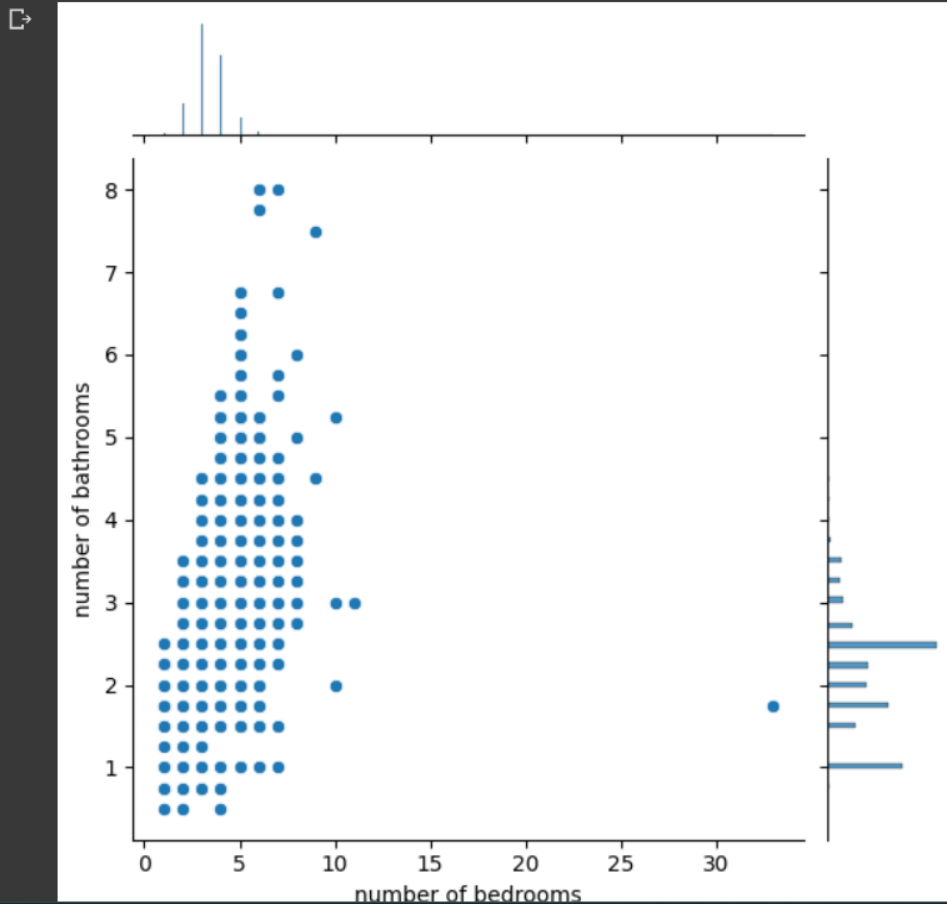


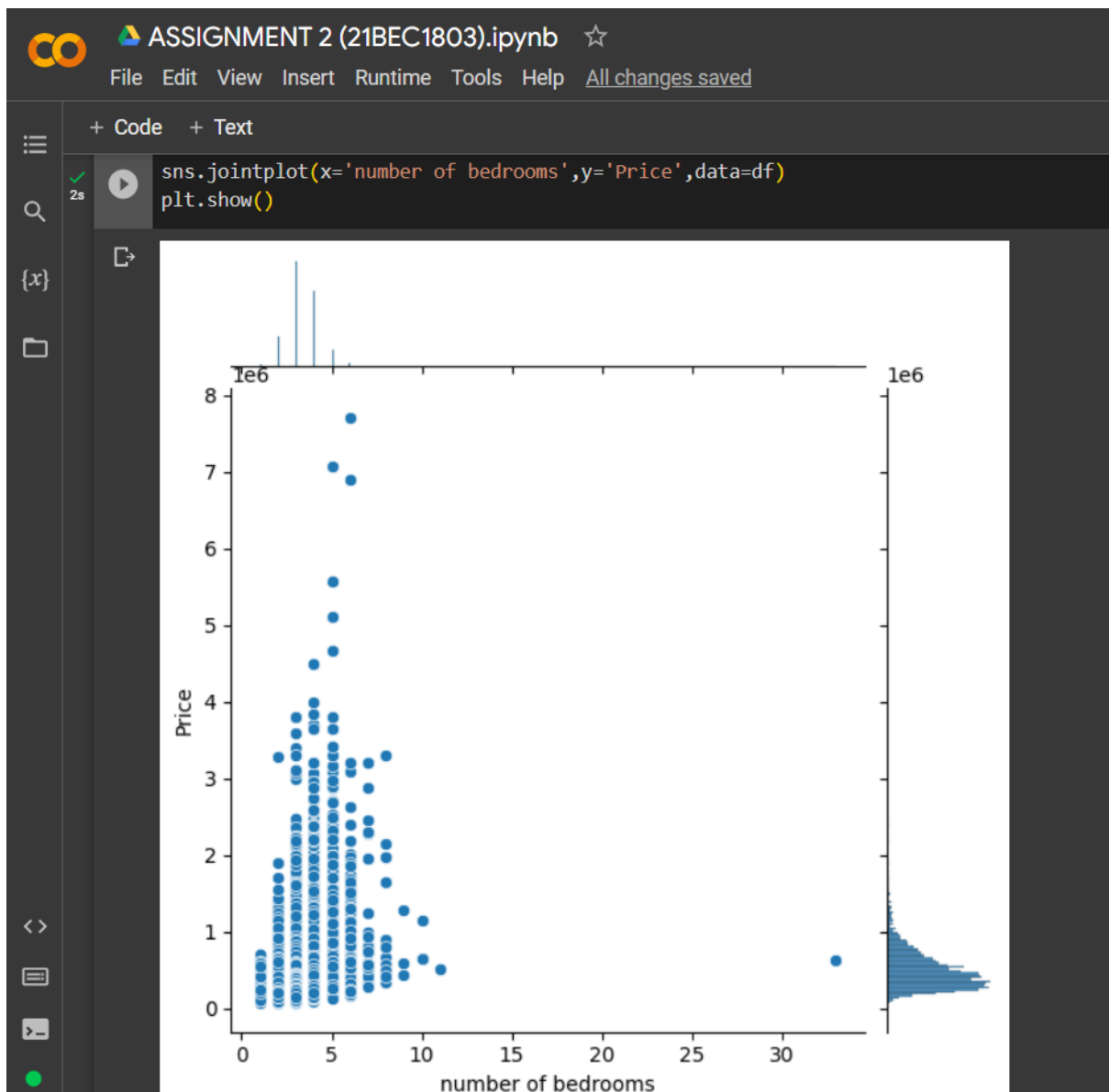
## ASSIGNMENT 2 (21BEC1803).ipynb ☆

File Edit View Insert Runtime Tools Help [All changes saved](#)

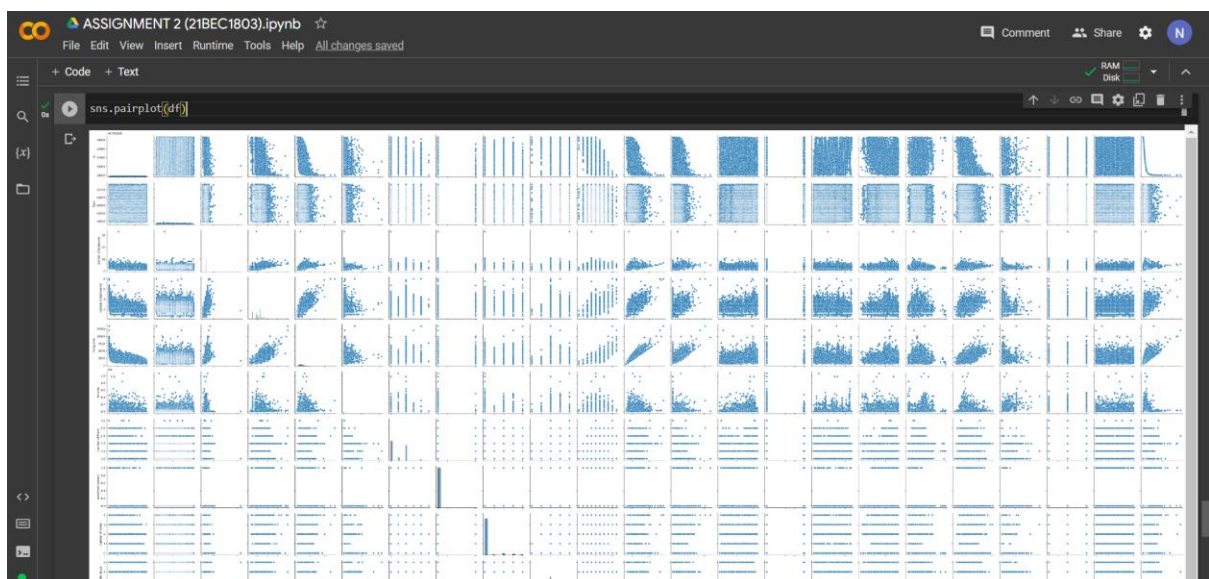
+ Code + Text

```
✓ 2s sns.jointplot(x='number of bedrooms',y='number of bathrooms',data=df)  
plt.show()
```





### TASK 3 – MULTIVARIATE ANALYSIS



## TASK – 4

```
df.describe()
```

ASSIGNMENT 2 (21BEC1803).ipynb

File Edit View Insert Runtime Tools Help

+ Code + Text

df.describe()

	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 C
count	1.462000e+04	14620.000000	14620.000000	14620.000000	14620.000000	1.462000e+04	14620.000000	14620.000000	14620.000000	14620.000000	...	14620.000000	14620.000000	14620.000
mean	6.762821e+09	42604.538646	3.379343	2.129583	2098.262996	1.509328e+04	1.502360	0.007661	0.233105	3.430506	...	1970.926402	90.924008	122033.062
std	6.237575e+03	67.347991	0.938719	0.769934	928.275721	3.791962e+04	0.540239	0.087193	0.766259	0.664151	...	28.493625	416.216661	19.082
min	6.762810e+09	42491.000000	1.000000	0.500000	370.000000	5.200000e+02	1.000000	0.000000	0.000000	1.000000	...	1900.000000	0.000000	122003.000
25%	6.762815e+09	42546.000000	3.000000	1.750000	1440.000000	5.010750e+03	1.000000	0.000000	0.000000	3.000000	...	1951.000000	0.000000	122017.000
50%	6.762821e+09	42600.000000	3.000000	2.250000	1930.000000	7.620000e+03	1.500000	0.000000	0.000000	3.000000	...	1975.000000	0.000000	122032.000
75%	6.762826e+09	42662.000000	4.000000	2.500000	2570.000000	1.080000e+04	2.000000	0.000000	0.000000	4.000000	...	1997.000000	0.000000	122048.000
max	6.762832e+09	42734.000000	33.000000	8.000000	13540.000000	1.074218e+06	3.500000	1.000000	4.000000	5.000000	...	2015.000000	2015.000000	122072.000

8 rows x 23 columns

## TASK – 5

ASSIGNMENT 2 (21BEC1803).ipynb

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

df.isnull().any()

#Since there are no null values, there is no need to handle the missing values!

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
Latitude	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	