

ASSIGNMENT 3

House Price India

-Sanjay S

```
Assignment - 3

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

[3] df=pd.read_csv('/content/House Price India.csv')

[4] df.head()
```

	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_
0	6762810145	42491	5	2.50	3650	9050	2.0	0	4	5	...	1921	0	122003	52.8645	-114.557	
1	6762810635	42491	4	2.50	2920	4000	1.5	0	0	5	...	1909	0	122004	52.8878	-114.470	
2	6762810998	42491	5	2.75	2910	9480	1.5	0	0	3	...	1939	0	122004	52.8852	-114.468	
3	6762812605	42491	4	2.50	3310	42998	2.0	0	0	3	...	2001	0	122005	52.9532	-114.321	
4	6762812919	42491	3	2.00	2710	4500	1.5	0	0	4	...	1929	0	122006	52.9047	-114.485	

5 rows x 23 columns

```
df.shape

(14620, 23)

Univariate Analysis

[7] df_price = df.loc[df['Price']>=3000000]
df_year = df.loc[df['Built Year']>1990]
df_ryear = df.loc[df['Renovation Year']>2000]

[8] df.info()
```

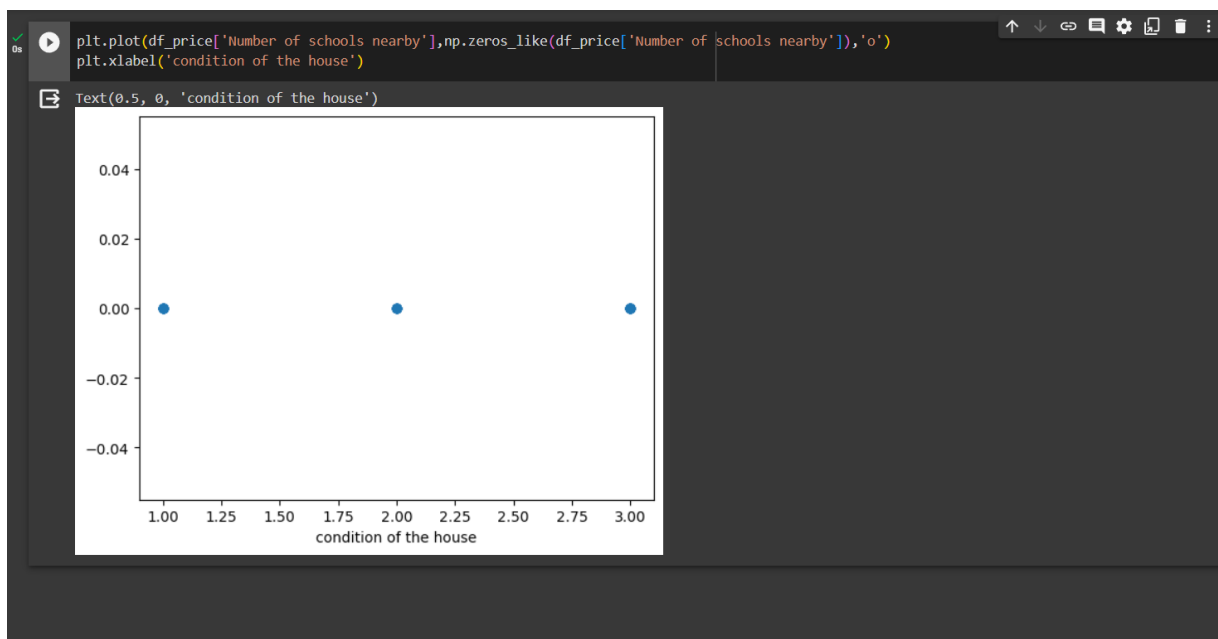
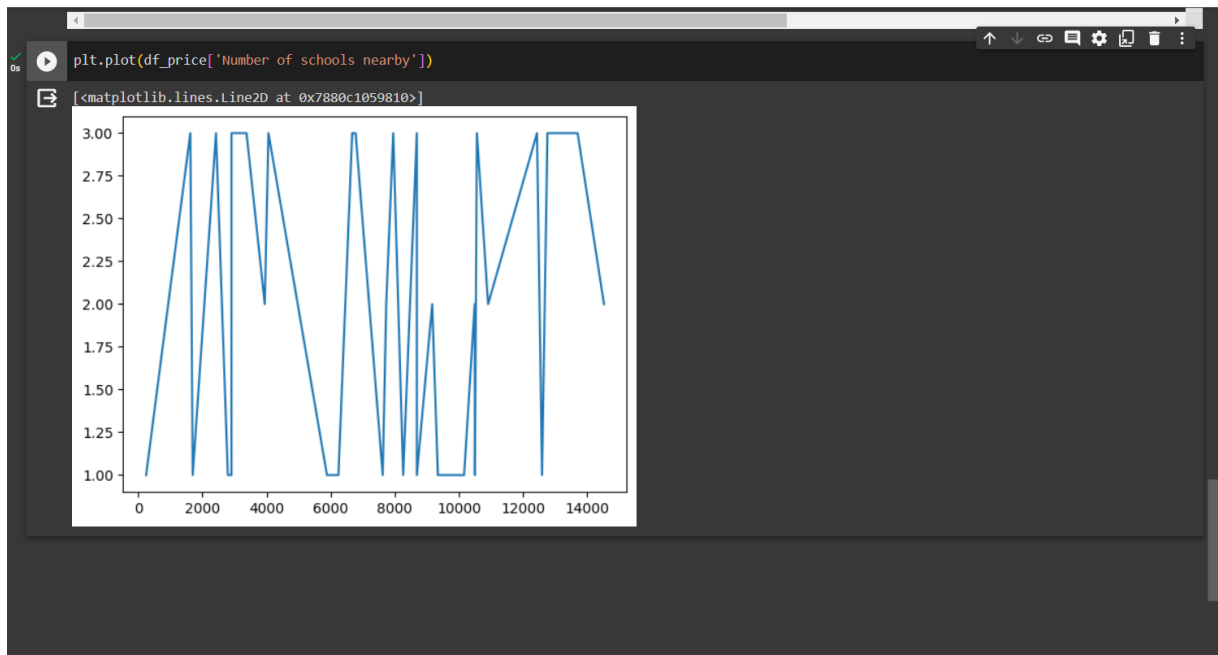
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 14620 entries, 0 to 14619
Data columns (total 23 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   id                                    14620 non-null  int64
1   Date                                14620 non-null  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
```

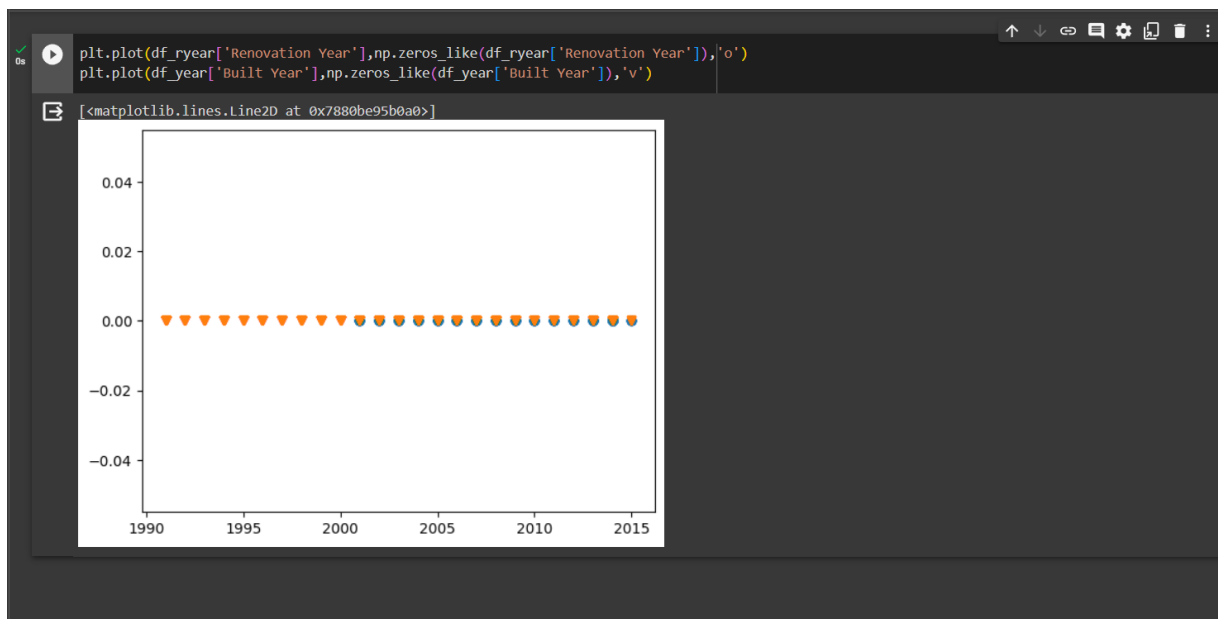
```
[8] df.info()

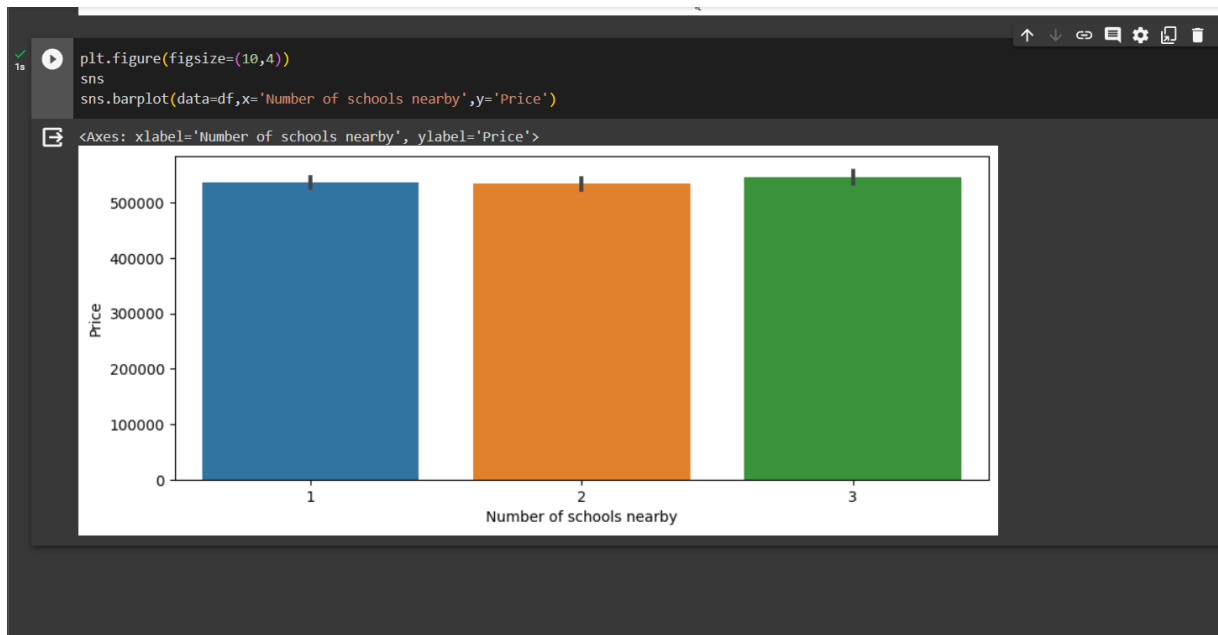
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 14620 entries, 0 to 14619
Data columns (total 23 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   id                                         14620 non-null  int64
1   Date                                       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  Latitude                                 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
```

```
[9] df_price
```

	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	li
243	6762810052	42496	7	4.50	6210	8856	2.5	0	2	5	...	1910	0	122061	52.8607	-114.544	
1622	6762810059	42518	6	4.25	6980	15682	3.0	0	4	4	...	1999	0	122057	52.7852	-114.421	
1697	6762810035	42519	4	3.50	5550	28078	2.0	0	2	4	...	2000	0	122071	52.8695	-114.424	
2424	6762810021	42531	5	4.50	10040	37325	2.0	1	2	3	...	1940	2001	122048	52.8800	-114.404	
2794	6762810027	42537	5	6.75	9640	13068	1.0	1	4	3	...	1983	2009	122057	52.7870	-114.400	
2907	6762810029	42538	4	3.00	6430	27517	2.0	0	0	3	...	2001	0	122048	52.8508	-114.409	
2908	6762810065	42538	4	4.25	4850	12445	2.0	1	4	5	...	1989	0	122033	52.9311	-114.434	
3234	6762810043	42543	3	4.50	5230	17826	2.0	1	4	3	...	2005	0	122057	52.7648	-114.433	
3376	6762810062	42544	4	5.00	4550	18641	1.0	1	4	3	...	2002	0	122019	52.8353	-114.267	
3946	6762810033	42551	5	5.50	7050	42840	1.0	0	2	4	...	1978	0	122048	52.8529	-114.410	
4061	6762810047	42552	5	6.25	8020	21738	2.0	0	0	3	...	2001	0	122027	52.7975	-114.379	
5887	6762810060	42579	5	5.25	5090	23669	2.0	0	0	3	...	2006	0	122048	52.8597	-114.406	
6244	6762810023	42585	5	5.75	9200	35069	2.0	0	0	3	...	2001	0	122071	52.8589	-114.423	
6674	6762810066	42592	3	3.50	4410	10756	2.0	1	4	3	...	2014	0	122053	52.7583	-114.395	
6781	6762810053	42593	4	3.25	7000	28206	1.0	1	4	4	...	1991	0	122020	52.8228	-114.276	







```
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 Y
count	1.462000e+04	14620.000000	14620.000000	14620.000000	14620.000000	1.462000e+04	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	...	1970.926
std	6.237575e+03	67.347991	0.938719	0.769934	928.275721	3.791962e+04	0.540239	0.087193	0.766259	0.664151	...	29.493
min	6.762810e+09	42491.000000	1.000000	0.500000	370.000000	5.200000e+02	1.000000	0.000000	0.000000	1.000000	...	1900.000
25%	6.762815e+09	42546.000000	3.000000	1.750000	1440.000000	5.010750e+03	1.000000	0.000000	0.000000	3.000000	...	1951.000
50%	6.762821e+09	42600.000000	3.000000	2.250000	1930.000000	7.620000e+03	1.500000	0.000000	0.000000	3.000000	...	1975.000
75%	6.762826e+09	42662.000000	4.000000	2.500000	2570.000000	1.080000e+04	2.000000	0.000000	0.000000	4.000000	...	1997.000
max	6.762832e+09	42734.000000	33.000000	8.000000	13540.000000	1.074218e+06	3.500000	1.000000	4.000000	5.000000	...	2015.000

8 rows x 23 columns

```
df['Area of the basement'].value_counts()
```

0	8842
800	157
600	148
700	147
500	146
...	...
276	1
768	1
2190	1
1840	1
1930	1

Name: Area of the basement, Length: 280, dtype: int64

```
df['Distance from the airport'].value_counts().to_frame()
```

Distance from the airport	
54	514
70	511
64	500
50	495
79	487
60	486
67	485
56	484
53	484
59	482
69	481
58	474
55	473
74	471
72	471
66	470
62	470
57	467

```
new_count = df['grade of the house'].value_counts().to_frame()  
new_count.rename(columns={'number of floors': 'new count'}, inplace=True)  
new_count
```

grade of the house	
7	6011
8	4137
9	1828
6	1324
10	804
11	280
5	154
12	55
4	17
13	10

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	