

```
In [51]: import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import matplotlib.ticker as ticker
from sklearn.linear_model import LinearRegression
import math
```

```
In [52]: data=pd.read_csv("C:/Users/Athichan/Downloads/House Price India.csv")
```

```
In [53]: print(data.shape)
data.head()
```

Out [53]:

	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	Number of schools nearby	Distance from the airport	Price
0	6762810145	42491	5	2.50	3650	9050	2.0	0	4	5 ...	1921	0	122003	52.8645	-114.557	2880	5400	2	58	2380000
1	6762810635	42491	4	2.50	2920	4000	1.5	0	0	5 ...	1909	0	122004	52.8878	-114.470	2470	4000	2	51	1400000
2	6762810998	42491	5	2.75	2910	9480	1.5	0	0	3 ...	1939	0	122004	52.8852	-114.468	2940	6600	1	53	1200000
3	6762812605	42491	4	2.50	3310	42998	2.0	0	0	3 ...	2001	0	122005	52.9532	-114.321	3350	42847	3	76	838000
4	6762812919	42491	3	2.00	2710	4500	1.5	0	0	4 ...	1929	0	122006	52.9047	-114.485	2060	4500	1	51	805000

5 rows × 23 columns

```
In [54]: data.isnull().sum()
```

Out [54]:

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

```
In [55]: data.describe().T
```

Out [55]:

	count	mean	std	min	25%	50%	75%	max
id	14620.0	6.762821e+09	6237.574799	6.762815e+09	6.762821e+09	6.762821e+09	6.762821e+09	6.762832e+09
Date	14620.0	4.260454e+04	67.347991	4.249100e+04	4.254600e+04	4.260000e+04	4.266200e+04	4.273400e+04
number of bedrooms	14620.0	3.379343e+00	0.938719	1.000000e+00	3.000000e+00	3.000000e+00	4.000000e+00	3.000000e+01
number of bathrooms	14620.0	2.129583e+00	0.769934	5.000000e-01	1.750000e+00	2.250000e+00	2.500000e+00	8.000000e+00
living area	14620.0	2.086263e+03	928.275721	3.700000e+02	1.440000e+03	1.930000e+03	2.570000e+03	1.354000e+04
lot area	14620.0	1.509328e+04	37919.621304	5.200000e+02	5.010750e+03	7.620000e+03	1.080000e+04	1.074218e+06
number of floors	14620.0	1.502360e+00	0.540239	1.000000e+00	1.000000e+00	1.500000e+00	2.000000e+00	3.500000e+00
waterfront present	14620.0	7.660739e-03	0.087193	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	1.000000e+00
number of views	14620.0	2.331053e-01	0.766259	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	4.000000e+00
condition of the house	14620.0	3.430506e+00	0.664151	1.000000e+00	3.000000e+00	3.000000e+00	4.000000e+00	5.000000e+00
grade of the house	14620.0	7.682421e+00	1.175033	4.000000e+00	7.000000e+00	7.000000e+00	8.000000e+00	1.300000e+01
Area of the house(excluding basement)	14620.0	1.801784e+03	833.809963	3.700000e+02	1.200000e+03	1.580000e+03	2.240000e+03	9.410000e+03
Area of the basement	14620.0	2.964791e+02	448.551409	0.000000e+00	0.000000e+00	0.000000e+00	5.800000e+02	4.820000e+03
Built Year	14620.0	1.970926e+03	29.493625	1.900000e+03	1.951000e+03	1.975000e+03	1.997000e+03	2.015000e+03
Renovation Year	14620.0	9.082401e+01	416.216661	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	2.015000e+03
Postal Code	14620.0	1.220331e+05	19.082418	1.220030e+05	1.220170e+05	1.220320e+05	1.220480e+05	1.220720e+05
Latitude	14620.0	5.279285e+01	0.137522	5.238590e+01	5.270760e+01	5.280640e+01	5.290890e+01	5.300760e+01
Longitude	14620.0	-1.144040e+02	0.141326	-1.147090e+02	-1.145190e+02	-1.144210e+02	-1.143150e+02	-1.135050e+02
living_area_renov	14620.0	1.996702e+03	691.093366	4.600000e+02	1.490000e+03	1.850000e+03	2.380000e+03	6.110000e+03
lot_area_renov	14620.0	1.275350e+04	26058.414467	6.510000e+02	5.097750e+03	7.620000e+03	1.012500e+04	5.606170e+05
Number of schools nearby	14620.0	2.012244e+00	0.817284	1.000000e+00	1.000000e+00	2.000000e+00	3.000000e+00	3.000000e+00
Distance from the airport	14620.0	6.495096e+01	8.936008	5.000000e+01	5.700000e+01	6.500000e+01	7.300000e+01	8.000000e+01
Price	14620.0	5.389322e+05	367532.380804	7.800000e+04	3.200000e+05	4.500000e+05	6.450000e+05	7.700000e+06

```
In [56]: data.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
```

```
In [57]: data.describe().T
```

Out [57]:

	count	mean	std	min	25%	50%	75%	max
id	14620.0	6.762821e+09	6237.574799	6.762815e+09	6.762821e+09	6.762821e+09	6.762826e+09	6.762832e+09
Date	14620.0	4.260454e+04	67.347991	4.249100e+04	4.254600e+04	4.260000e+04	4.266200e+04	4.273400e+04
number of bedrooms	14620.0	3.379343e+00	0.938719	1.000000e+00	3.000000e+00	3.000000e+00	4.000000e+00	3.000000e+01
number of bathrooms	14620.0	2.129583e+00	0.769934	5.000000e-01	1.750000e+00	2.250000e+00	2.500000e+00	8.000000e+00
living area	14620.0	2.086263e+03	928.275721	3.700000e+02	1.440000e+03	1.930000e+03	2.570000e+03	1.354000e+04
lot area	14620.0	1.509328e+04	37919.621304	5.200000e+02	5.010750e+03	7.620000e+03	1.080000e+04	1.074218e+06
number of floors	14620.0	1.502360e+00	0.540239	1.000000e+00	1.000000e+00	1.500000e+00	2.000000e+00	3.500000e+00
waterfront present	14620.0	7.660739e-03	0.087193	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	1.000000e+00
number of views	14620.0	2.331053e-01	0.766259	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	4.000000e+00
condition of the house	14620.0	3.430506e+00	0.664151	1.000000e+00	3.000000e+00	3.000000e+00	4.000000e+00	5.000000e+00
grade of the house	14620.0	7.682421e+00	1.175033	4.000000e+00	7.000000e+00	7.000000e+00	8.000000e+00	1.300000e+01
Area of the house(excluding basement)	14620.0	1.801784e+03	833.809963	3.700000e+02	1.200000e+03	1.580000e+03	2.240000e+03	9.410000e+03
Area of the basement	14620.0	2.964791e+02	448.551409	0.000000e+00	0.000000e+00	0.000000e+00	5.800000e+02	4.820000e+03
Built Year	14620.0	1.970926e+03	29.493625	1.900000e+03	1.951000e+03	1.975000e+03	1.997000e+03	2.015000e+03
Renovation Year	14620.0	9.082401e+01	416.216661	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	2.015000e+03
Postal Code	14620.0	1.220331e+05	19.082418	1.220030e+05	1.220170e+05	1.220320e+05	1.220480e+05	1.220720e+05
Latitude	14620.0	5.279285e+01	0.137522	5.238590e+01	5.270760e+01	5.280640e+01	5.290890e+01	5.300760e+01
Longitude	14620.0	-1.144040e+02	0.141326	-1.147090e+02	-1.145190e+02	-1.144210e+02	-1.143150e+02	-1.135050e+02
living_area_renov	14620.0	1.996702e+03	691.093366	4.600000e+02	1.490000e+03	1.850000e+03	2.380000e+03	6.110000e+03
lot_area_renov	14620.0	1.275350e+04	26058.414467	6.510000e+02	5.097750e+03	7.620000e+03	1.012500e+04	5.606170e+05
Number of schools nearby	14620.0	2.012244e+00	0.817284	1.000000e+00	1.000000e+00	2.000000e+00	3.000000e+00	3.000000e+00
Distance from the airport	14620.0	6.495096e+01	8.936008	5.000000e+01	5.700000e+01	6.500000e+01	7.300000e+01	8.000000e+01
Price	14620.0	5.389322e+05	367532.380804	7.800000e+04	3.200000e+05	4.500000e+05	6.450000e+05	7.700000e+06

```
In [58]: data = data.drop(columns=['Id','Date'])
data
```

Out [58]:

	number of bedrooms	number of bathrooms	living area	lot area	number of floors	waterfront present	number of views	condition of the house	grade of the house	Area of the house(excluding basement)	Built Year	Renovation Year	Postal Code	Latitude	Longitude	living_area_renov	lot_area_renov	Number of schools nearby	Distance from the airport	Price
0	5	2.50	3650	9050	2.0	0	4	5	10	3370 ...	1921	0	122003	52.8645	-114.557	2880	5400	2	58	2380000
1	4	2.50	2920	4000	1.5	0	0	5	8	1910 ...	1909	0	122004	52.8878	-114.470	2470	4000	2	51	1400000
2	5	2.75	2910	9480	1.5	0	0	3	8	2910 ...	1939	0	122004	52.8852	-114.468	2940	6600	1	53	1200000
3	4	2.50	3310	42998	2.0	0	0	3	9	3310 ...	2001	0	122005	52.9532	-114.321	3350	42847	3	76	838000
4	3	2.00	2710	4500	1.5	0	0	4	8	1880 ...	1929	0	122006	52.9047	-114.485	2060	4500	1	51	805000
...
14615	2	1.50	1556	20000	1.0	0	0	4	7	1556 ...	1957	0	122066	52.6191	-114.472	2250	17286	3	76	221700
14616	3	2.00	1680	7000	1.5	0	0	4	7	1680 ...	1968	0	122072	52.5075	-114.393	1540	7480	3	59	219200
14617	2	1.00	1070	6120	1.0	0	0	3	6	1070 ...	1962	0	122056	52.7289	-114.507	1130	6120	2	64	209000
14618	4	1.00	1030	6621	1.0	0	0	4	6	1030 ...	1955	0	122042	52.7157	-114.411	1420	6631	3	54	205000
14619	3	1.00	900	4770	1.0	0	0	3	6	900 ...	1969	2009	122018	52.5338	-114.552	900	3480	2	55	146000

14620 rows × 21 columns

```
In [59]: def rename_columns(columns):
return columns.lower().replace(' ','_')
```

```
In [60]: data.rename(columns=rename_columns, inplace=True)
```

```
In [61]: data.columns
```

```
Out [61]: Index(['number_of_bedrooms', 'number_of_bathrooms', 'living_area', 'lot_area',
'number_of_floors', 'waterfront_present', 'number_of_views',
'condition_of_the_house', 'grade_of_the_house',
'area_of_the_house(excluding_basement)', 'area_of_the_basement',
'built_year', 'renovation_year', 'postal_code', 'latitude',
'longitude', 'living_area_renov', 'lot_area_renov',
'number_of_schools_nearby', 'distance_from_the_airport', 'price'],
      dtype='object')
```

Univariate

```
In [62]: sns.countplot(data['number_of_bathrooms'])
```

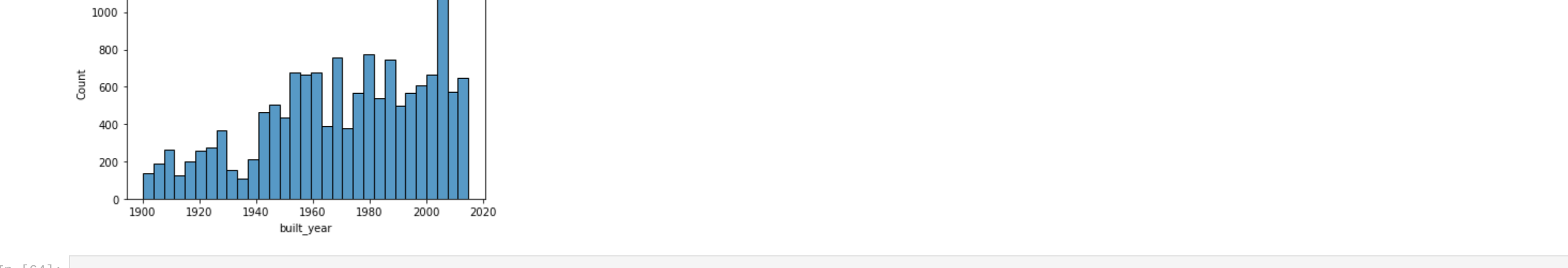
```
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be 'data', and passing other arguments without an explicit keyword will result in an error or misinterpretation.
warnings.warn(msg, FutureWarning)
```

```
Out [62]: <AxesSubplot: xlabel='number_of_bathrooms', ylabel='count'>
```



```
In [63]: sns.histplot(data['built_year'])
```

```
Out [63]: <AxesSubplot: xlabel='built_year', ylabel='Count'>
```



```
In [64]: matrix = np.triu(data.corr())
sns.heatmap(data.corr(), mask=matrix, cmap='vlag',
            linewidths=0.5, linecolor='white', annot=True, annot_kws={'size':5})
```