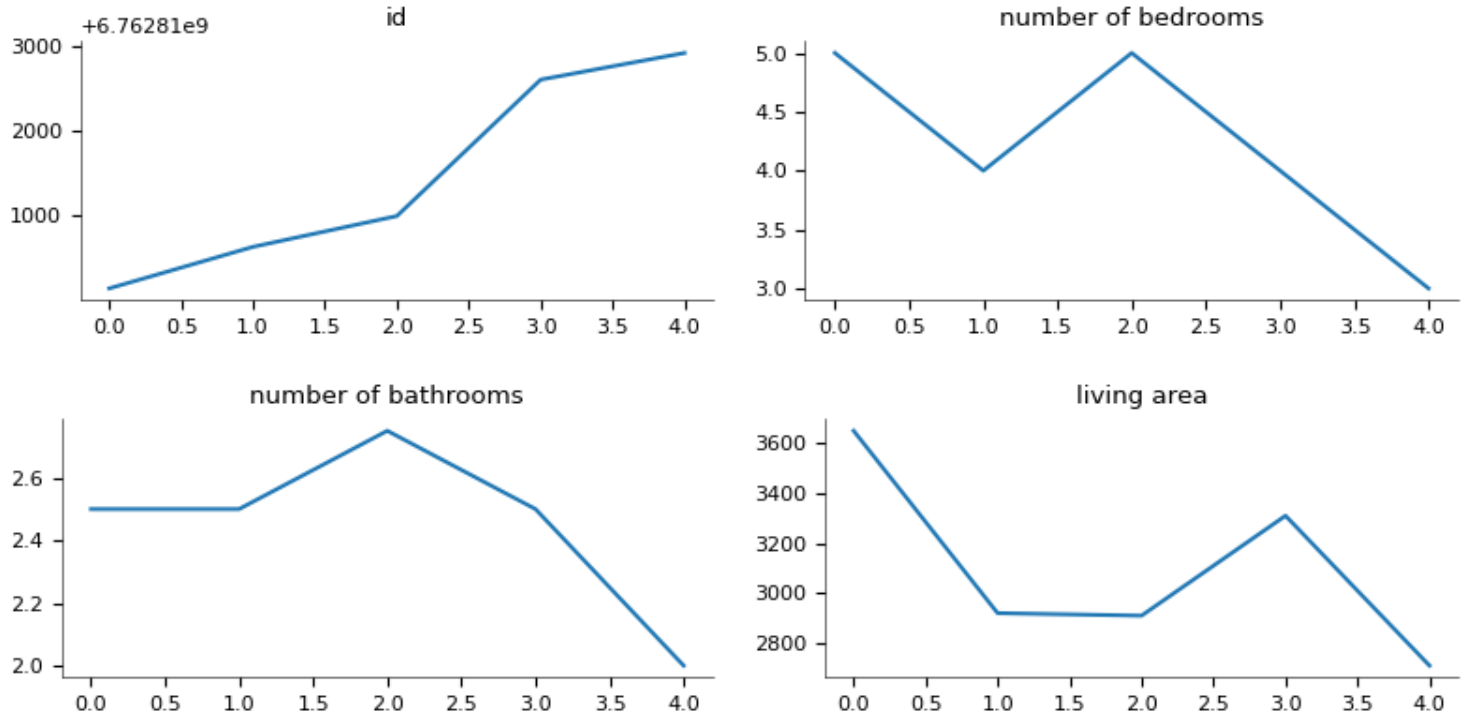
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt


df=pd.read_excel("/content/House prediction.xlsx")
df.head()
```

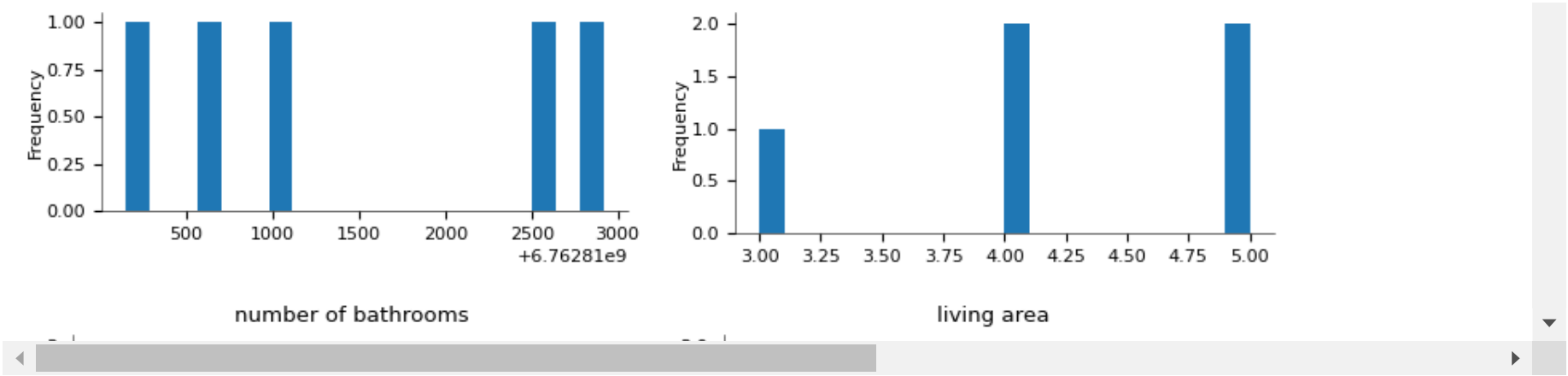
	id	Date	number of bedrooms	number of bathrooms	living area	lot area	number of floors	waterfront present	number of views	condition of the house	...	Built Year	Renovation Year	Post Co
0	6762810145	42491	5	2.50	3650	9050	2.0	0	4	5	...	1921	0	1220
1	6762810635	42491	4	2.50	2920	4000	1.5	0	0	5	...	1909	0	1220
2	6762810998	42491	5	2.75	2910	9480	1.5	0	0	3	...	1939	0	1220
3	6762812605	42491	4	2.50	3310	42998	2.0	0	0	3	...	2001	0	1220
4	6762812919	42491	3	2.00	2710	4500	1.5	0	0	4	...	1929	0	1220

5 rows × 23 columns

Values



Distributions


Frequency
1.00
0.75
0.50
0.25
0.00
500
1000
1500
2000
2500
3000
+6.76281e9
number of bathrooms
Frequency
2.0
1.5
1.0
0.5
0.0
3.00
3.25
3.50
3.75
4.00
4.25
4.50
4.75
5.00
living area

```
df.shape
```

```
(14620, 23)
```

```
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  Lattitude                              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
```

```
df.isnull().any()
```

```
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
```

```
df.describe
```

```
<bound method NDFrame.describe of                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
```

```
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
0	6762810145	42491	5	2.50	3650	9050	2.0	0	4	5	...	1921	0	122003
1	6762810635	42491	4	2.50	2920	4000	1.5	0	0	5	...	1909	0	122004
2	6762810998	42491	5	2.75	2910	9480	1.5	0	0	3	...	1939	0	122004
3	6762812605	42491	4	2.50	3310	42998	2.0	0	0	3	...	2001	0	122005
4	6762812919	42491	3	2.00	2710	4500	1.5	0	0	4	...	1929	0	122006

5 rows × 23 columns

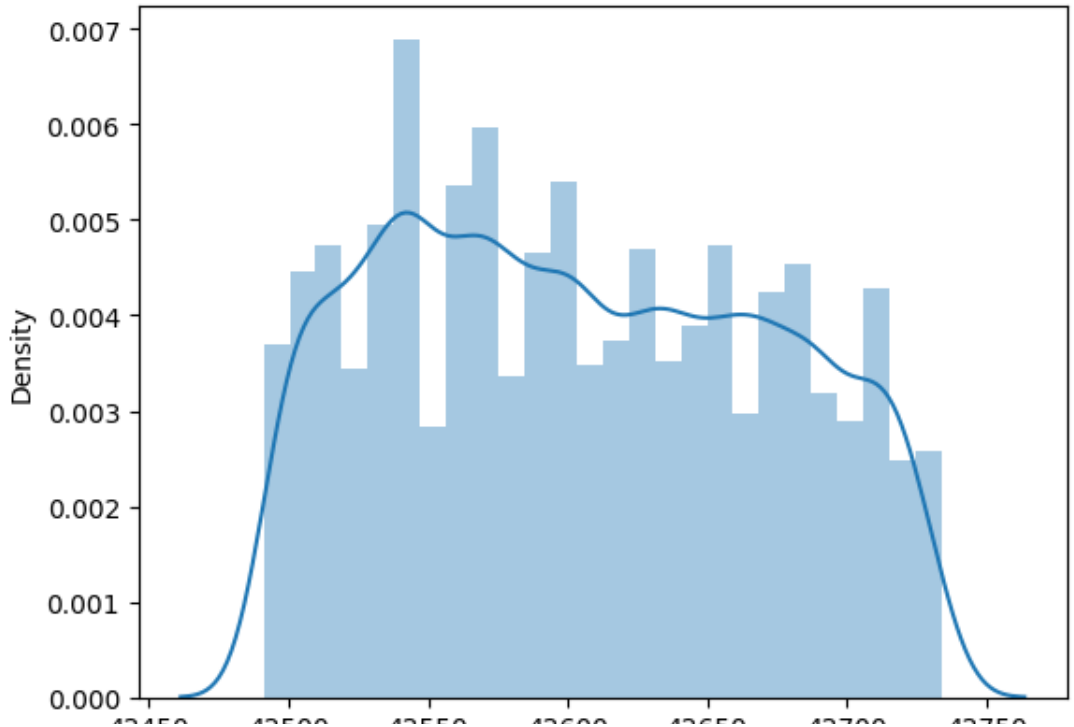
```
df.corr()
```

	id	Date	number of bedrooms	number of bathrooms	living area	lot area	number of floors	waterfront present	number of views	condition of the house	..
id	1.000000	0.045966	-0.329034	-0.516909	-0.648127	-0.100269	-0.312305	-0.112937	-0.293004	-0.045061	.
Date	0.045966	1.000000	-0.015663	-0.026485	-0.021958	0.004392	-0.010335	0.012006	-0.004782	-0.027402	.
number of bedrooms	-0.329034	-0.015663	1.000000	0.509784	0.570526	0.034416	0.177294	-0.006257	0.078665	0.026597	.
number of bathrooms	-0.516909	-0.026485	0.509784	1.000000	0.753517	0.080806	0.502924	0.060104	0.183789	-0.128232	.
living area	-0.648127	-0.021958	0.570526	0.753517	1.000000	0.174420	0.354743	0.105837	0.287728	-0.063358	.
lot area	-0.100269	0.004392	0.034416	0.080806	0.174420	1.000000	-0.004138	0.026282	0.078308	-0.008548	.
number of floors	-0.312305	-0.010335	0.177294	0.502924	0.354743	-0.004138	1.000000	0.016316	0.020153	-0.269928	.
waterfront present	-0.112937	0.012006	-0.006257	0.060104	0.105837	0.026282	0.016316	1.000000	0.400206	0.018644	.
number of views	-0.293004	-0.004782	0.078665	0.183789	0.287728	0.078308	0.020153	0.400206	1.000000	0.052533	.
condition of the house	-0.045061	-0.027402	0.026597	-0.128232	-0.063358	-0.008548	-0.269928	0.018644	0.052533	1.000000	.
grade of the house	-0.673448	-0.033097	0.352945	0.663054	0.761835	0.110546	0.463082	0.079831	0.254532	-0.152530	.
Area of the house(excluding basement)	-0.565116	-0.015994	0.473599	0.684391	0.875793	0.183553	0.525643	0.071865	0.162672	-0.167695	.
Area of the											
Built Year											.

UNIVARIATE

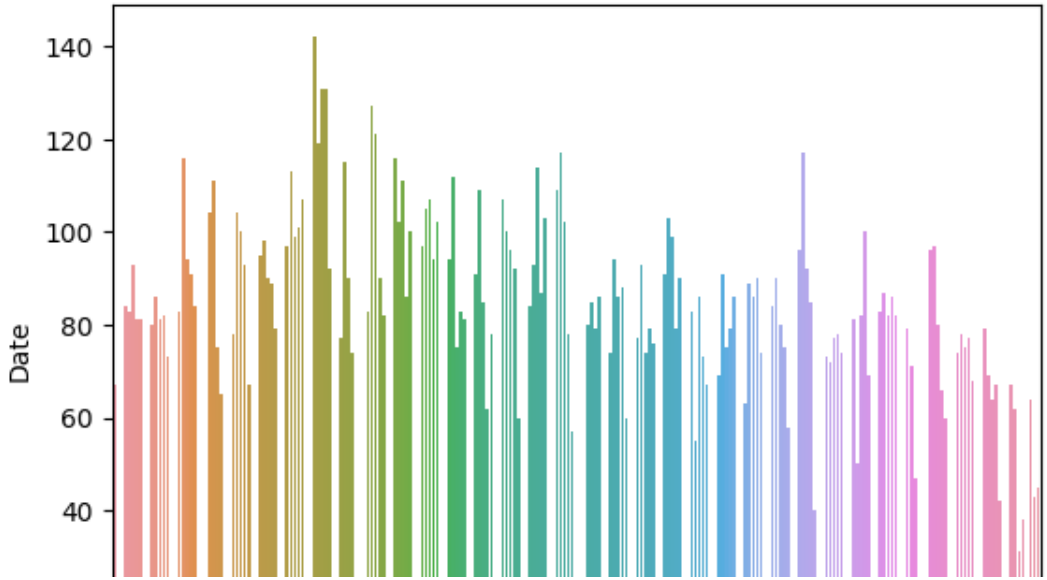
```
sns.distplot(df.Date) #for numerical column
```

```
<ipython-input-12-4e261051fa23>:1: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df.Date) #for numerical column
<Axes: xlabel='Date', ylabel='Density'>
```



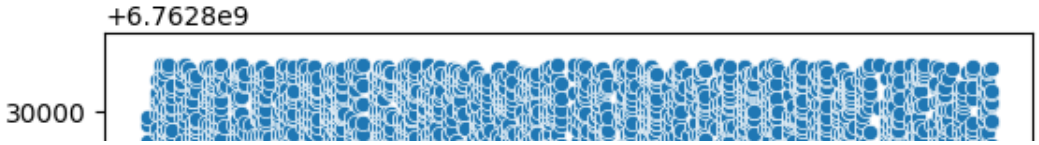

```
sns.barplot(x =df.Date.value_counts().index,y =df.Date.value_counts() )
```

```
<Axes: ylabel='Date'>
```




```
sns.scatterplot(x = df.Date,y=df.id)
```

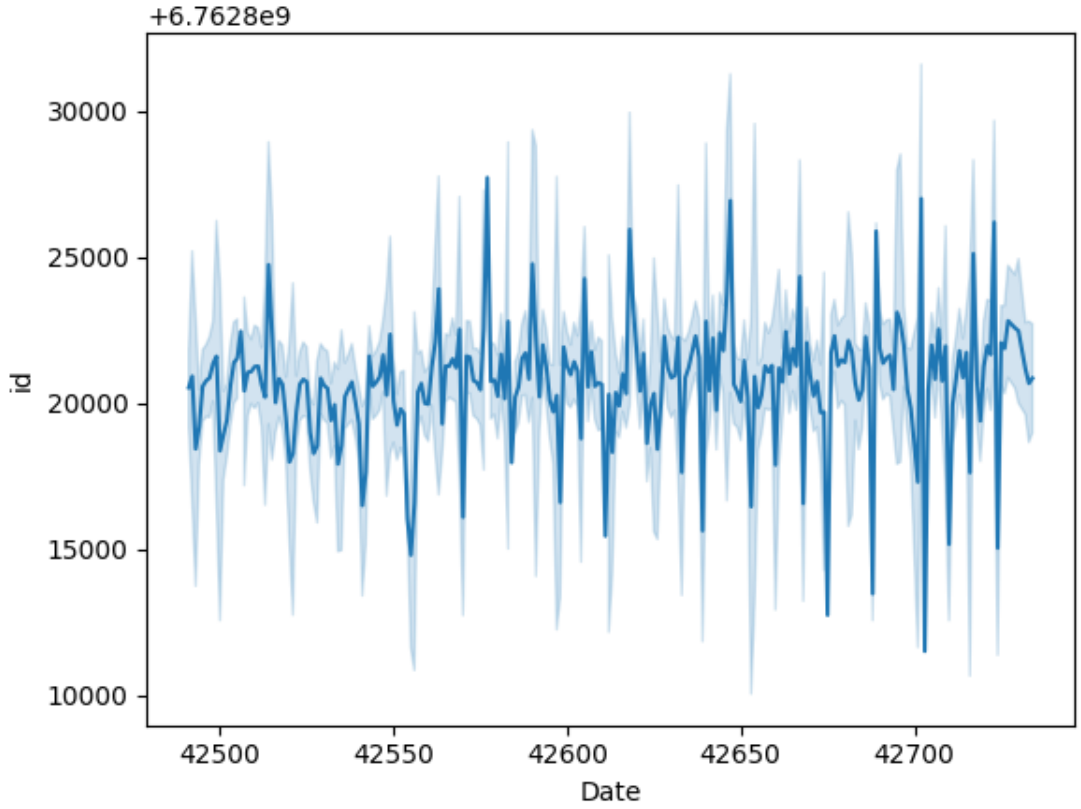
```
<Axes: xlabel='Date', ylabel='id'>
```



BIVARIATE ANALYSIS

```
sns.lineplot(x=df.Date,y=df.id)
```

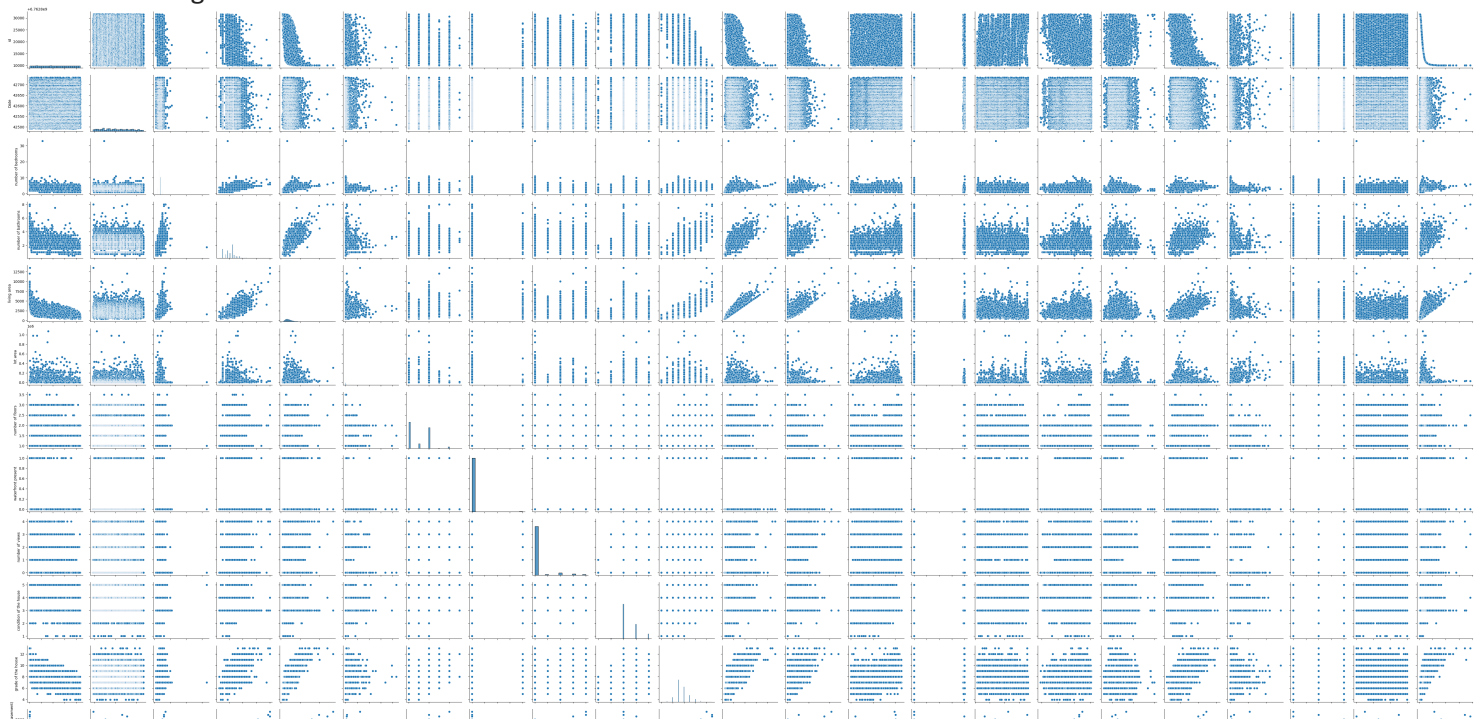
```
<Axes: xlabel='Date', ylabel='id'>
```



MULTIVARIATE ANALYSIS

```
sns.pairplot(df)
```

```
<seaborn.axisgrid.PairGrid at 0x7d7bd8be4f10>
```



```
sns.heatmap(df.corr(),annot=True)
```

```
<Axes: >
```

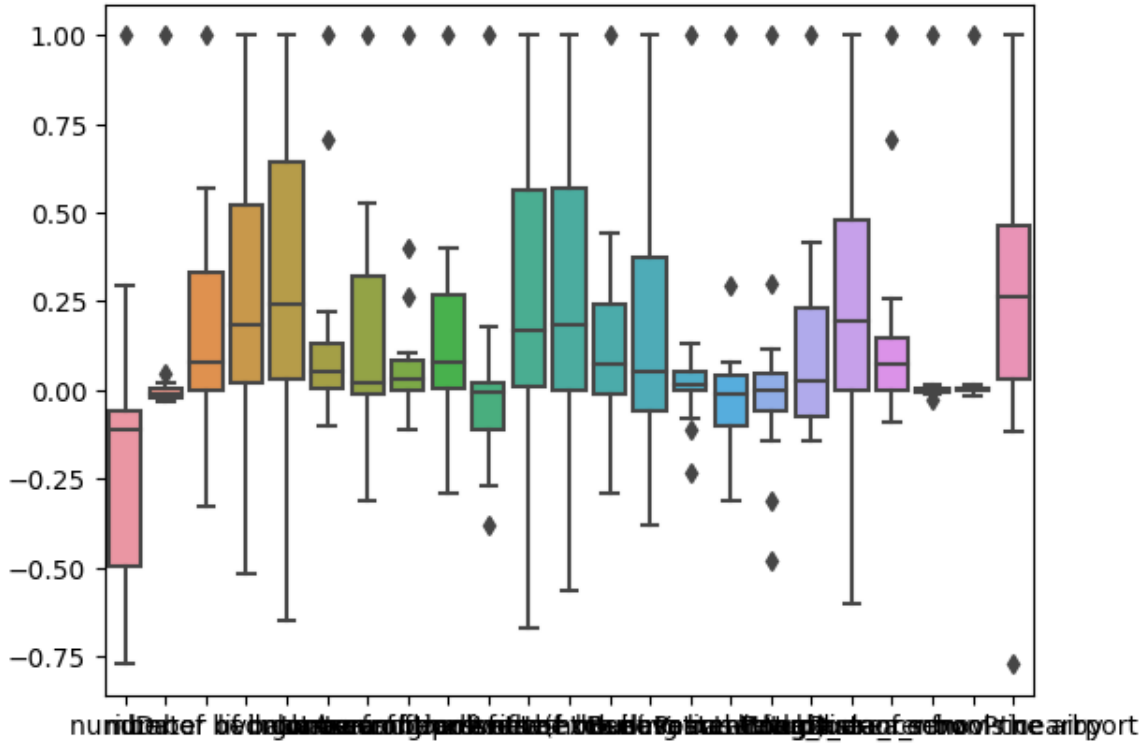

```
df.head()
```

	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
0	6762810145	42491	5	2.50	3650	9050	2.0	0	4	5	...	1921	0	122003
1	6762810635	42491	4	2.50	2920	4000	1.5	0	0	5	...	1909	0	122004
2	6762810998	42491	5	2.75	2910	9480	1.5	0	0	3	...	1939	0	122004
3	6762812605	42491	4	2.50	3310	42998	2.0	0	0	3	...	2001	0	122005
4	6762812919	42491	3	2.00	2710	4500	1.5	0	0	4	...	1929	0	122006

5 rows × 23 columns

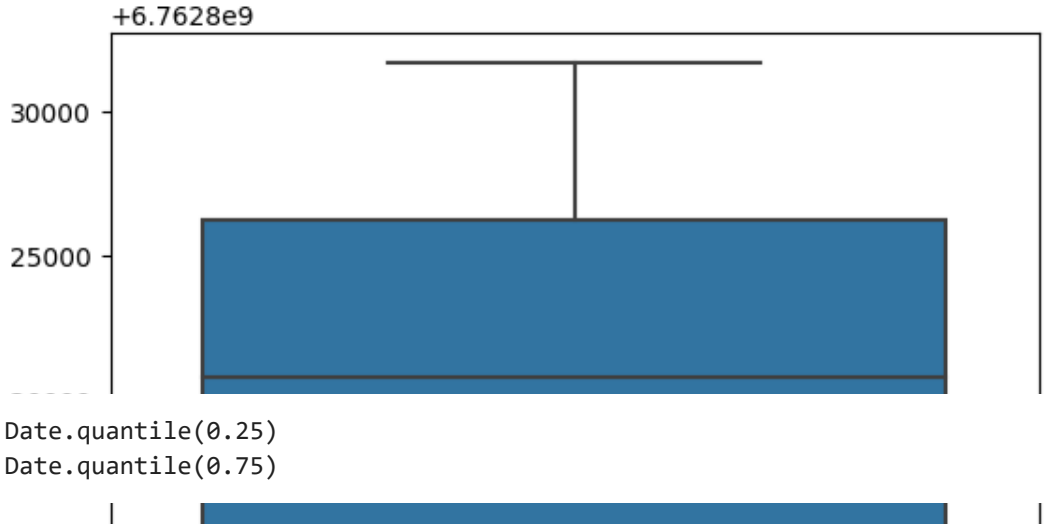
```
sns.boxplot(df.corr())
```

```
<Axes: >
```




```
sns.boxplot(df.id)
```

```
<Axes: >
```




```
q1=df.Date.quantile(0.25)
q3=df.Date.quantile(0.75)
```

```
print(q1)
print(q3)
```

```
42546.0
42662.0
```

```
from scipy import stats
```

```
Date_zscore=stats.zscore(df.Date)
Date_zscore
```

```
0       -1.685908
1       -1.685908
2       -1.685908
3       -1.685908
4       -1.685908
           ...
14615    1.922341
14616    1.922341
14617    1.922341
14618    1.922341
14619    1.922341
Name: Date, Length: 14620, dtype: float64
```

```
df_z=df[np.abs(Date_zscore)<=3]
df_z
```

	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	Pos C
0	6762810145	42491	5	2.50	3650	9050	2.0	0	4	5	...	1921	0	122
1	6762810635	42491	4	2.50	2920	4000	1.5	0	0	5	...	1909	0	122
2	6762810998	42491	5	2.75	2910	9480	1.5	0	0	3	...	1939	0	122
3	6762812605	42491	4	2.50	3310	42998	2.0	0	0	3	...	2001	0	122
4	6762812919	42491	3	2.00	2710	4500	1.5	0	0	4	...	1929	0	122
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
14615	6762830250	42734	2	1.50	1556	20000	1.0	0	0	4	...	1957	0	122
14616	6762830339	42734	3	2.00	1680	7000	1.5	0	0	4	...	1968	0	122
14617	6762830618	42734	2	1.00	1070	6120	1.0	0	0	3	...	1962	0	122
14618	6762830709	42734	4	1.00	1030	6621	1.0	0	0	4	...	1955	0	122
14619	6762831463	42734	3	1.00	900	4770	1.0	0	0	3	...	1969	2009	122

14620 rows × 23 columns

```
sns.boxplot(df.Date)
```