Pratyush Tyagi 21BCE2747 Vit Vellore

TASK 1 File Downloded House price India

```
import pandas as pd
import matplotlib.pyplot as plt
from matplotlib import rcParams
import seaborn as sns
```

TasK 2

```
df = pd.read_csv('/content/House Price India.csv')
df.head()
```
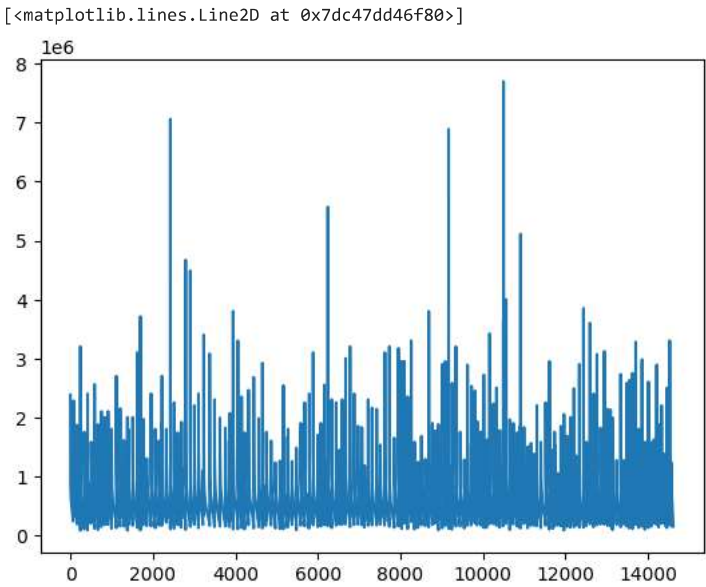
	id	Date	number of bedrooms	number of bathrooms	living area	lot area	number of floors	waterfront present	number of views	cc
0	6762810145	42491	5	2.50	3650	9050	2.0	0	4	
1	6762810635	42491	4	2.50	2920	4000	1.5	0	0	
2	6762810998	42491	5	2.75	2910	9480	1.5	0	0	
3	6762812605	42491	4	2.50	3310	42998	2.0	0	0	
4	6762812919	42491	3	2.00	2710	4500	1.5	0	0	

5 rows × 23 columns

Task 3

Univariavble

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

```
[<matplotlib.lines.Line2D at 0x7dc47dd46f80>]
```



```
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
```
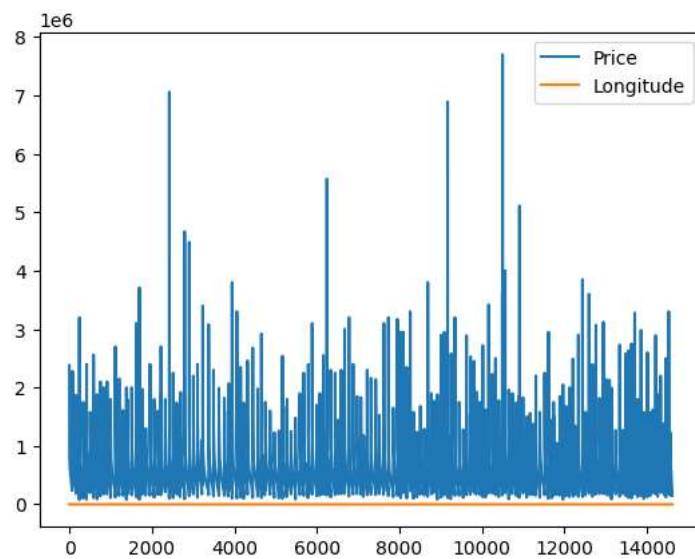
Bi variate

```
df.Price.plot()
df.Longitude.plot()
plt.legend(['Price','Longitude'])
```
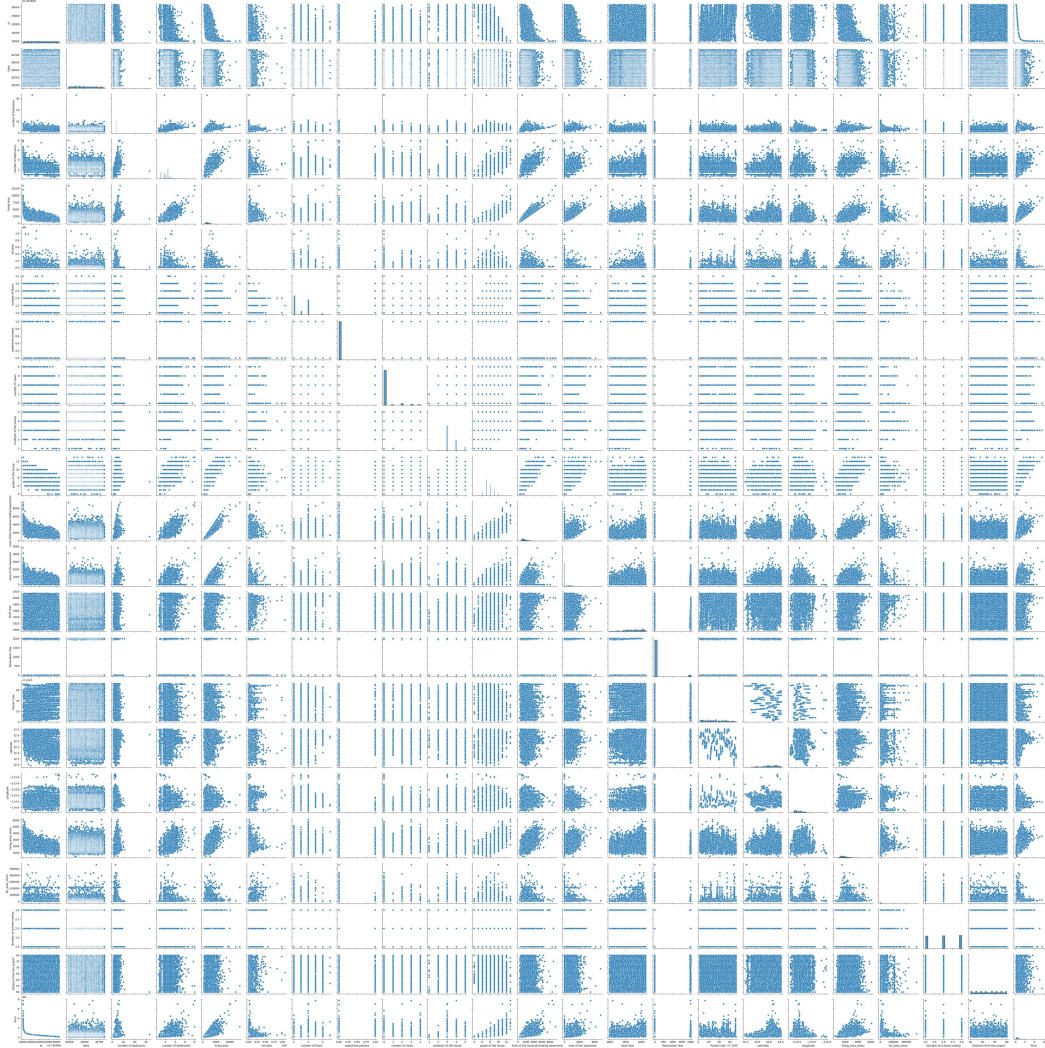
```
<matplotlib.legend.Legend at 0x7dc47dea03a0>
```


Mulrivariate

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

<seaborn.axisgrid.PairGrid at 0x7dc47dea05b0>


```
sns.heatmap(df.corr())
plt.figure(figsize=(60,60))
```

```
<Figure size 6000x6000 with 0 Axes>
```


Task 4 Descriptive staststics

```
df.describe()
```

	id	Date	number of bedrooms	number of bathrooms	living area	lot area	number of floors	waterfront present	number of views	cond...
count	1.462000e+04	14620.000000	14620.000000	14620.000000	14620.000000	1.462000e+04	14620.000000	14620.000000	14620.000000	14620.0
mean	6.762821e+09	42604.538646	3.379343	2.129583	2098.262996	1.509328e+04	1.502360	0.007661	0.233105	3.4
std	6.237575e+03	67.347991	0.938719	0.769934	928.275721	3.791962e+04	0.540239	0.087193	0.766259	0.6
min	6.762810e+09	42491.000000	1.000000	0.500000	370.000000	5.200000e+02	1.000000	0.000000	0.000000	1.0
25%	6.762815e+09	42546.000000	3.000000	1.750000	1440.000000	5.010750e+03	1.000000	0.000000	0.000000	3.0
50%	6.762821e+09	42600.000000	3.000000	2.250000	1930.000000	7.620000e+03	1.500000	0.000000	0.000000	3.0
75%	6.762826e+09	42662.000000	4.000000	2.500000	2570.000000	1.080000e+04	2.000000	0.000000	0.000000	4.0
max	6.762832e+09	42734.000000	33.000000	8.000000	13540.000000	1.074218e+06	3.500000	1.000000	4.000000	5.0

8 rows × 23 columns

Task 5 Handle the missing value

```
<Figure size 6000x6000 with 0 Axes>
```

```
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
```

✓ 0s   completed at 12:23 AM   ● ✕