SYED MOHAMMED RAZZUQ

21BLC1396

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.

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

```
path="/content/drive/MyDrive/dataset/House Price India.csv"
df=pd.read_csv(path)
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	Lattitude	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	5400
1	6762810635	42491	4	2.50	2920	4000	1.5	0	0	5	...	1909	0	122004	52.8878	-114.470	2470	4000
2	6762810998	42491	5	2.75	2910	9480	1.5	0	0	3	...	1939	0	122004	52.8852	-114.468	2940	6600
3	6762812605	42491	4	2.50	3310	42998	2.0	0	0	3	...	2001	0	122005	52.9532	-114.321	3350	42847
4	6762812919	42491	3	2.00	2710	4500	1.5	0	0	4	...	1929	0	122006	52.9047	-114.485	2060	4500

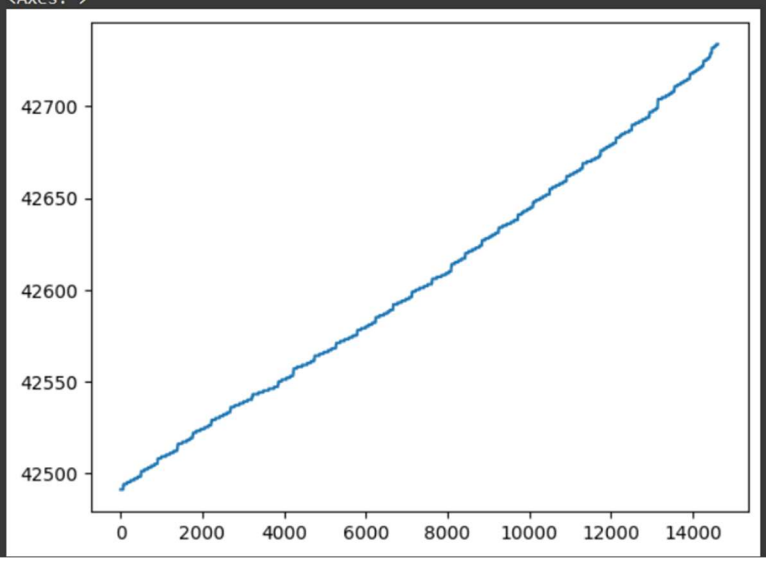
5 rows × 23 columns

```
[ ] df.shape
```

```
(14620, 23)
```

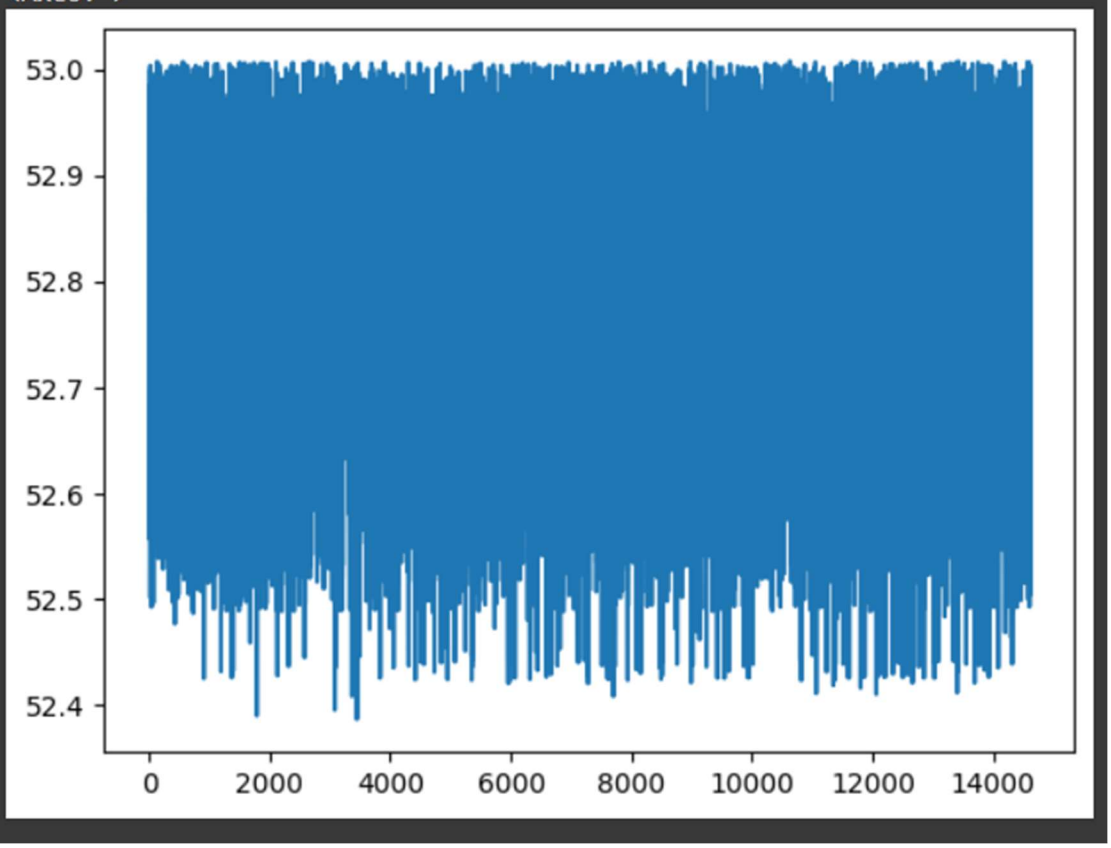
```
df.Date.plot()
```

```
<Axes: >
```



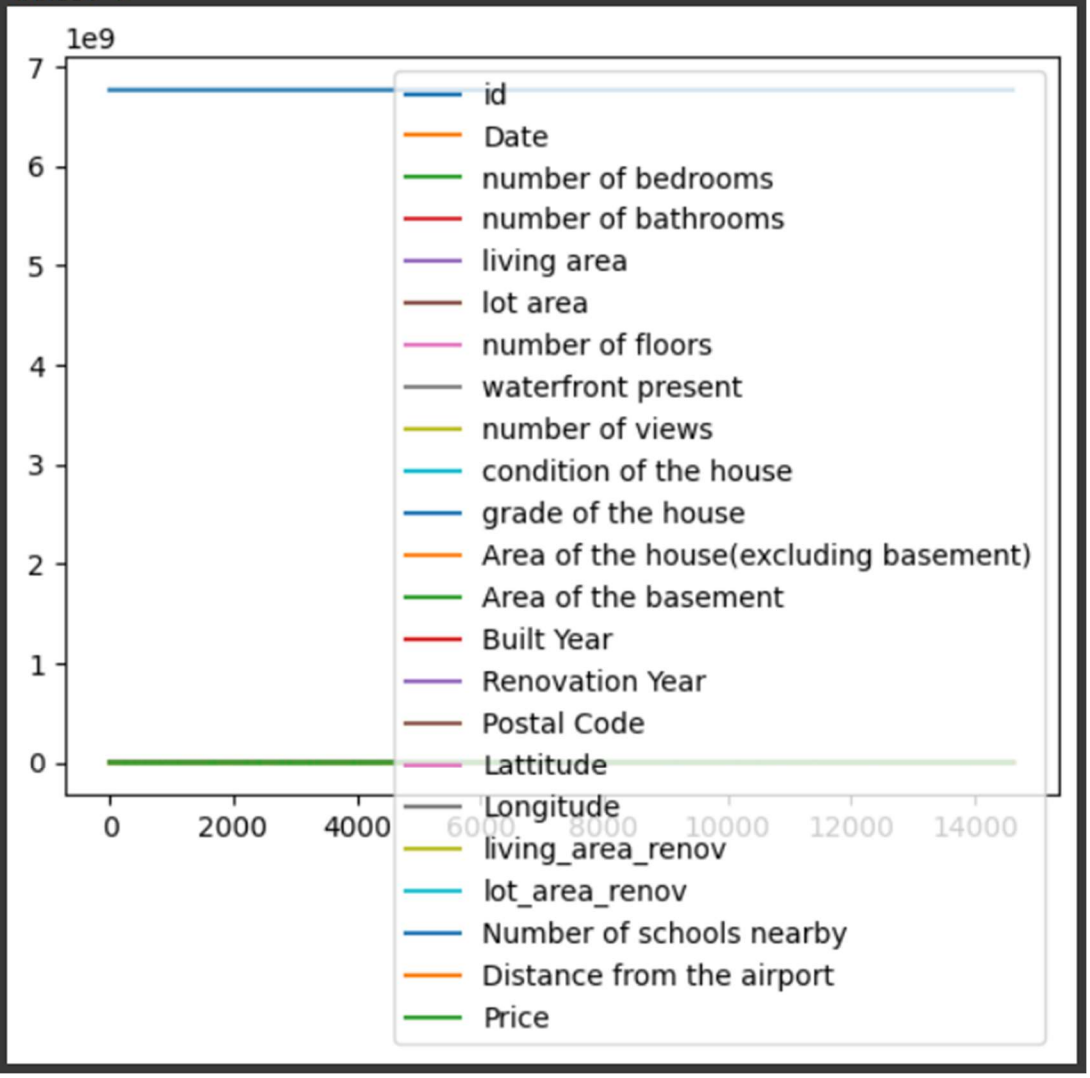

```
df.Lattitude.plot()
```

```
<Axes: >
```



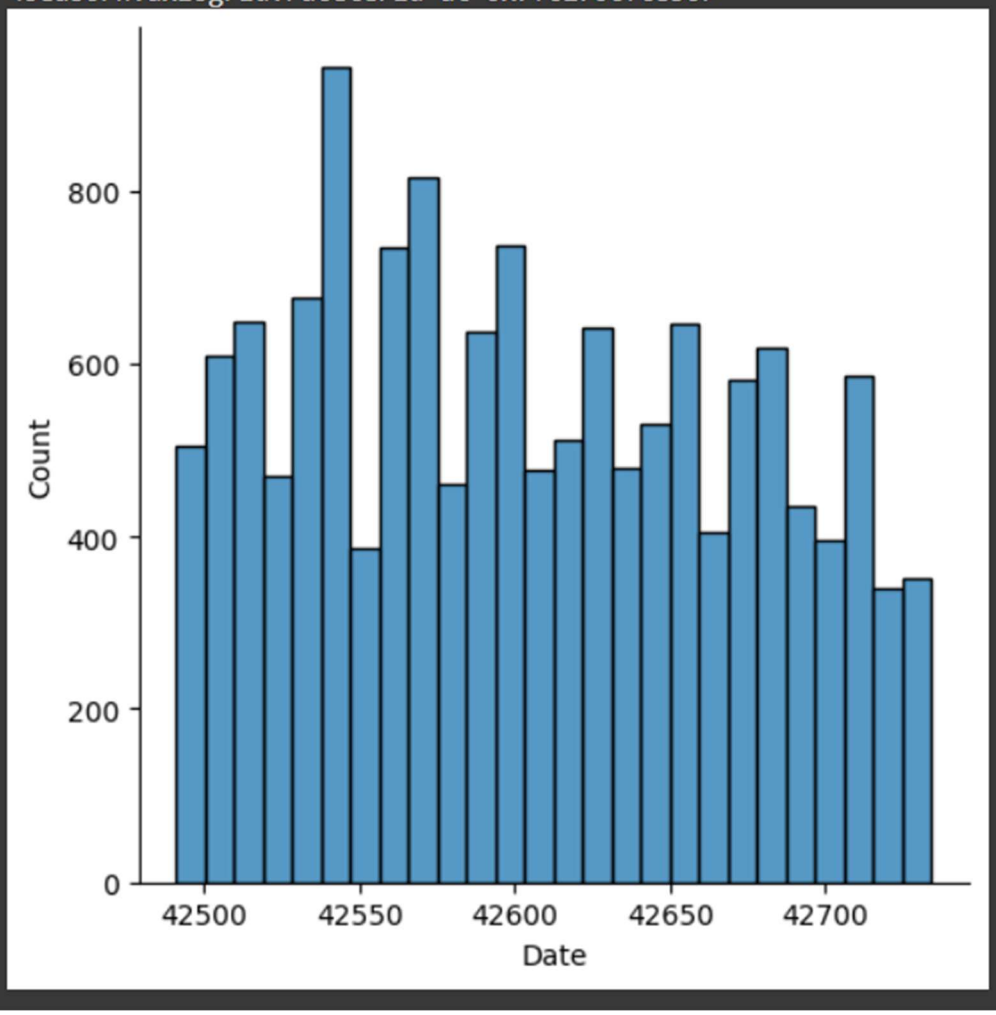
```
df.plot()
```

```
<Axes: >
```


1e9
7
6
5
4
3
2
1
0
0
2000
4000
6000
8000
10000
12000
14000
id
Date
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
Lattitude
Longitude
living_area_renov
lot_area_renov
Number of schools nearby
Distance from the airport
Price

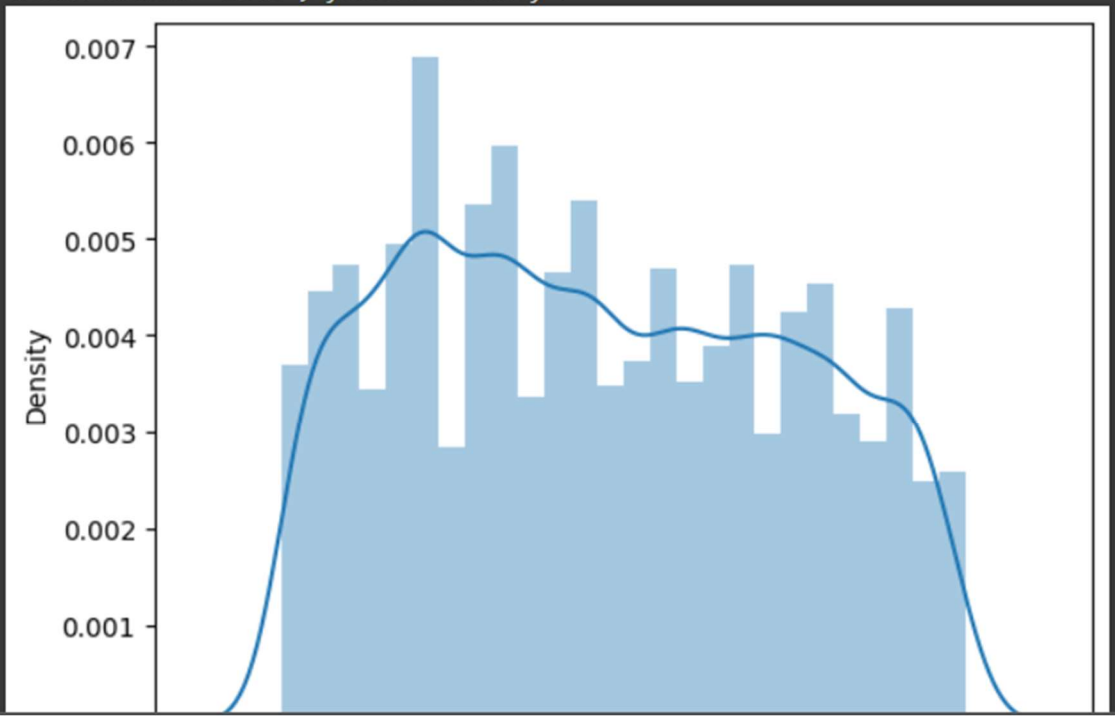
```
sns.displot(df.Date)
```

```
<seaborn.axisgrid.FacetGrid at 0x7fc170cfc850>
```



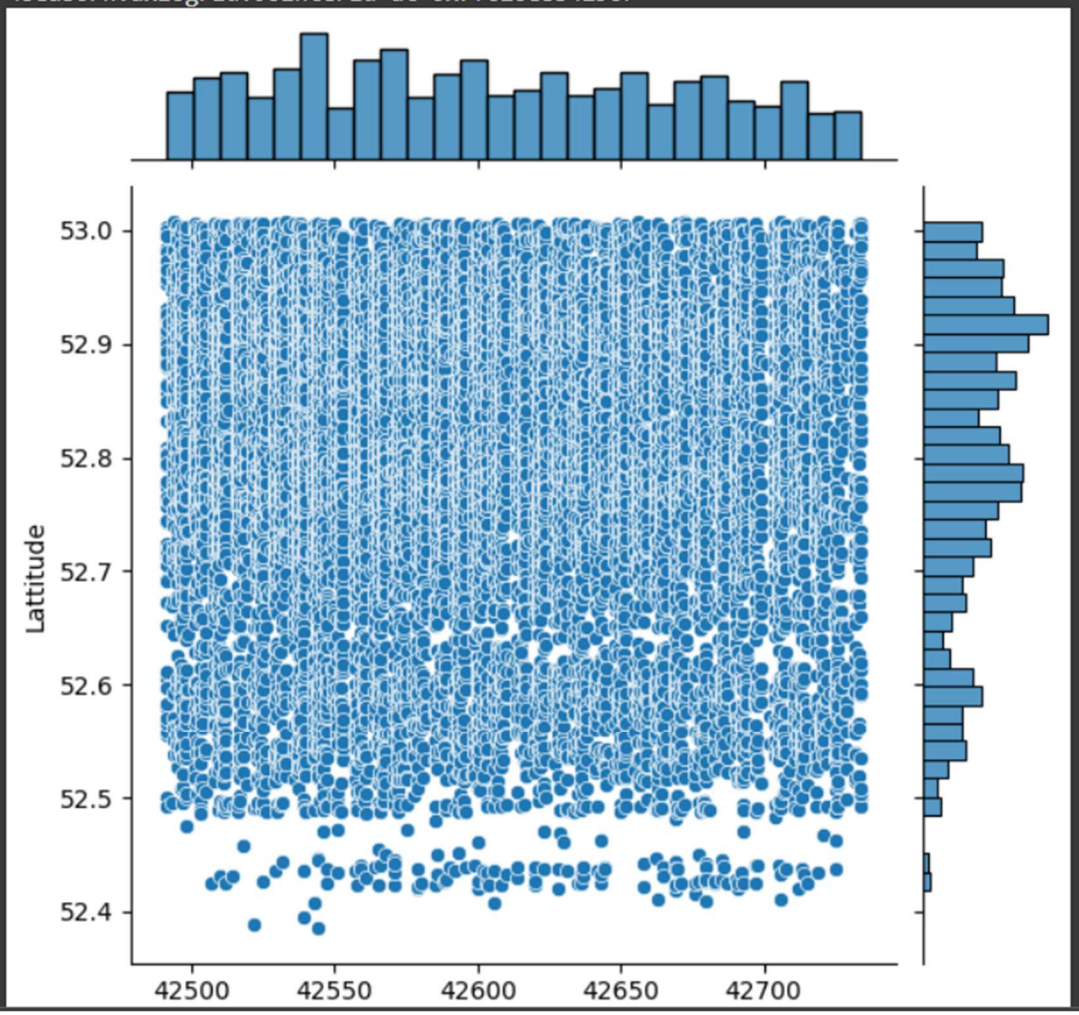
```
sns.distplot(df.Date)
```

```
<ipython-input-13-82d9cb3bf0f8>: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)
<Axes: xlabel='Date', ylabel='Density'>
```

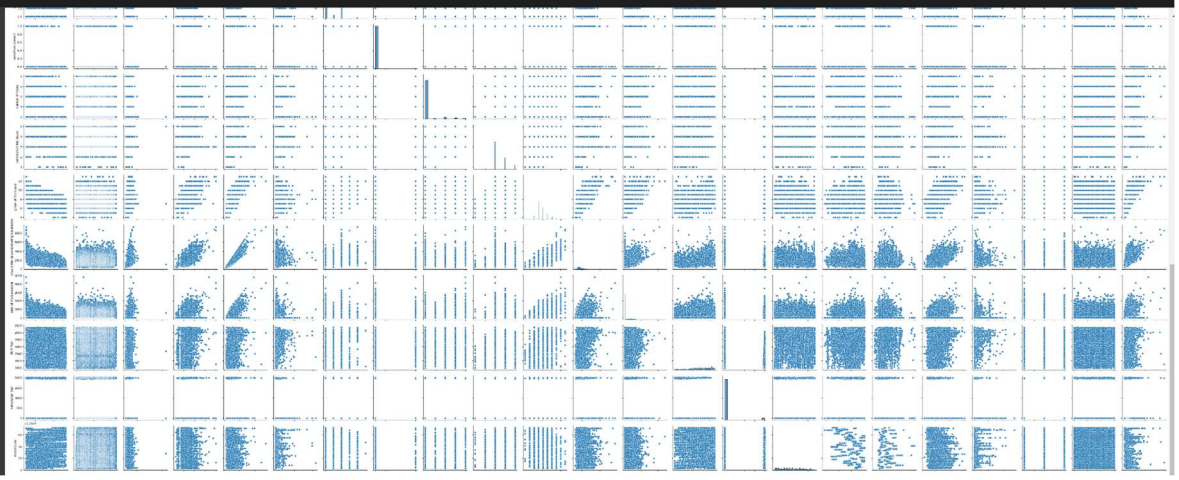


```
sns.jointplot(x= 'Date',y ='Lattitude',data=df)
```

```
<seaborn.axisgrid.JointGrid at 0x7fc16e8b4190>
```




```
#multivariate analysis
sns.pairplot(df)
```



descriptive statistics

```
df.describe()
```

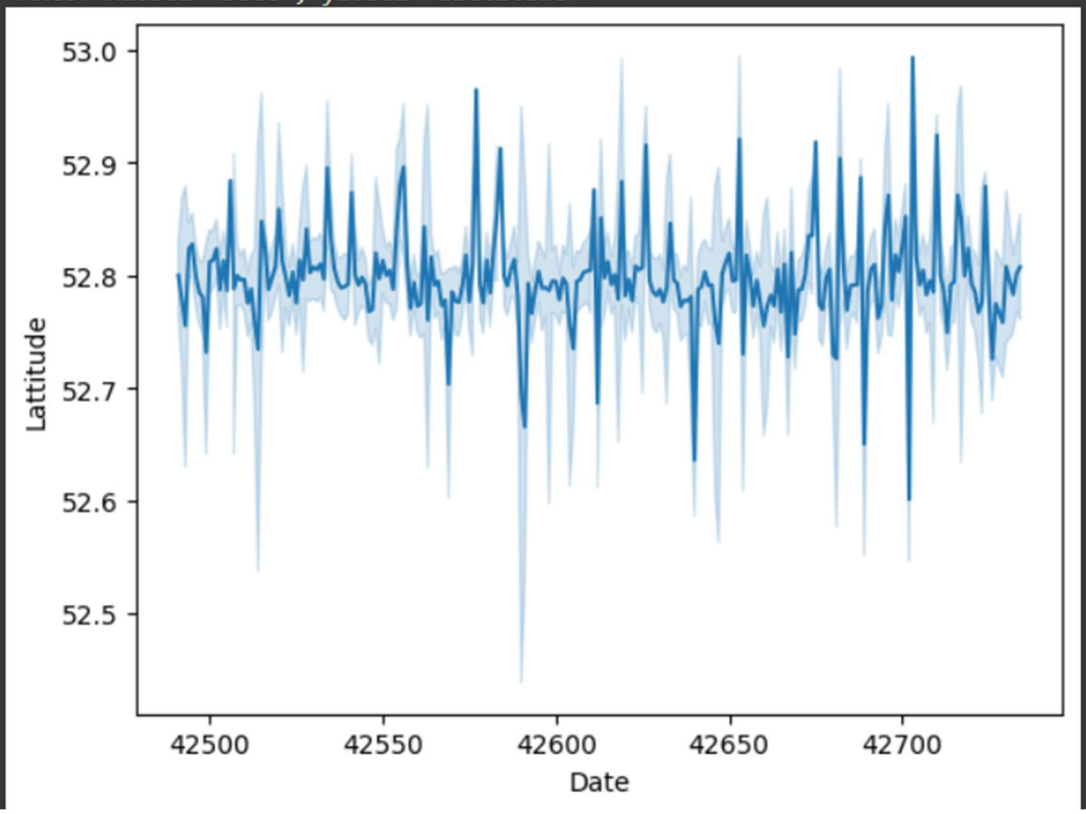
	id	Date	number of bedrooms	number of bathrooms	living area	lot area	number of floors	waterfront present	number of views	condition of the house	...	Built Year	Renovation Year	Postal C
count	1.462000e+04	14620.000000	14620.000000	14620.000000	14620.000000	1.462000e+04	14620.000000	14620.000000	14620.000000	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	...	29.493625	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 × 23 columns

Bi variate

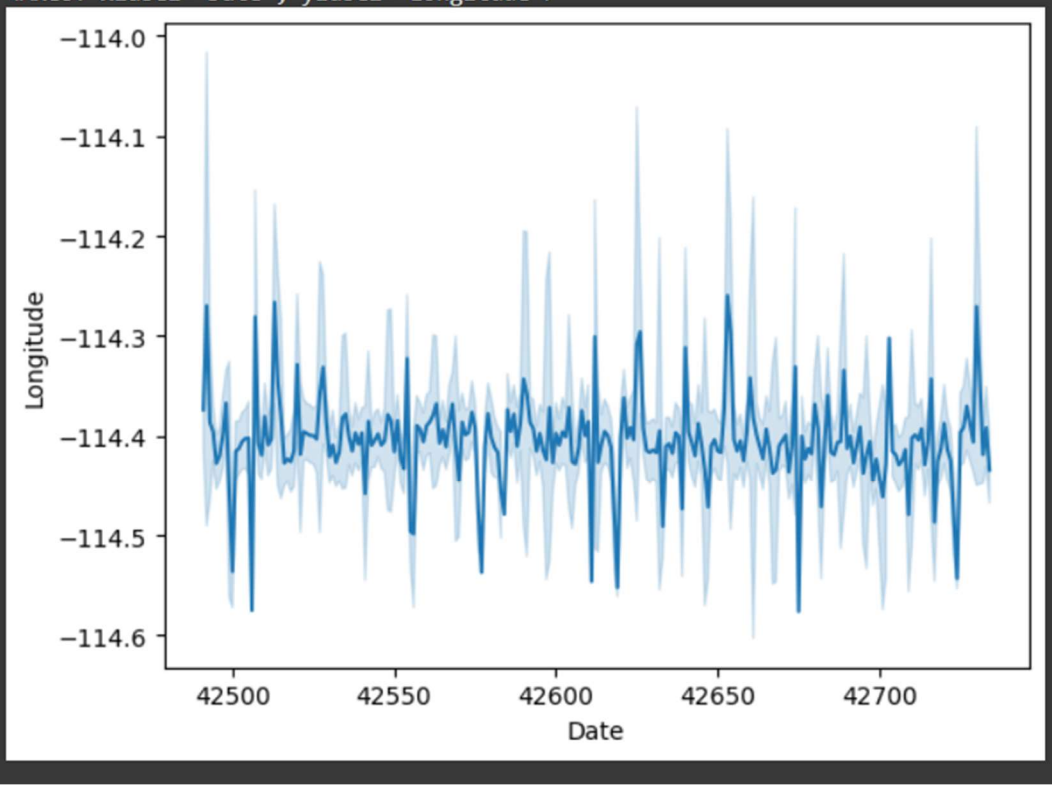
```
sns.lineplot(x = df.Date,y=df.Lattitude)
```

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



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

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



```
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.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
Lattitude                                 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
```

```
[ ] df.fillna(0)
```

	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	Lattitude	Longitude	living_area_renov	lot_area_rer
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	4(
2	6762810998	42491	5	2.75	2910	9480	1.5	0	0	3	...	1939	0	122004	52.8852	-114.468	2940	6E
3	6762812605	42491	4	2.50	3310	42998	2.0	0	0	3	...	2001	0	122005	52.9532	-114.321	3350	42E
4	6762812919	42491	3	2.00	2710	4500	1.5	0	0	4	...	1929	0	122006	52.9047	-114.485	2060	4E
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
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	6E
14619	6762831463	42734	3	1.00	900	4770	1.0	0	0	3	...	1969	2009	122018	52.5338	-114.552	900	34

14620 rows × 23 columns