

NAME: EDE RENUKA MADHAV

REG NO: 21BCT0244

VIT VELLORE

SLOT: 6:00 PM TO 8:00 PM

1. Download the dataset
2. Load the dataset

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

```
df=pd.read_csv("/content/House Price India.csv")
df.head()
```

	id	Date	number of bedrooms	number of bathrooms	living area \
0	6762810145	42491	5	2.50	3650
1	6762810635	42491	4	2.50	2920
2	6762810998	42491	5	2.75	2910
3	6762812605	42491	4	2.50	3310
4	6762812919	42491	3	2.00	2710

	lot area	number of floors	waterfront present	number of views \
0	9050	2.0	0	4
1	4000	1.5	0	0
2	9480	1.5	0	0
3	42998	2.0	0	0
4	4500	1.5	0	0

	condition of the house	...	Built Year	Renovation Year	Postal Code \
0	5	...	1921	0	122003
1	5	...	1909	0	122004
2	3	...	1939	0	122004
3	3	...	2001	0	122005
4	4	...	1929	0	122006

	Latitude	Longitude	living_area_renov	lot_area_renov	\
0	52.8645	-114.557	2880	5400	
1	52.8878	-114.470	2470	4000	
2	52.8852	-114.468	2940	6600	
3	52.9532	-114.321	3350	42847	
4	52.9047	-114.485	2060	4500	

	Number of schools nearby	Distance from the airport	Price
0	2	58	2380000
1	2	51	1400000
2	1	53	1200000
3	3	76	838000
4	1	51	805000

[5 rows x 23 columns]

1. Perform the Visualizations

UNIVARATE ANALYSIS

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

```
<ipython-input-15-261e020c24e7>:1: UserWarning:
```

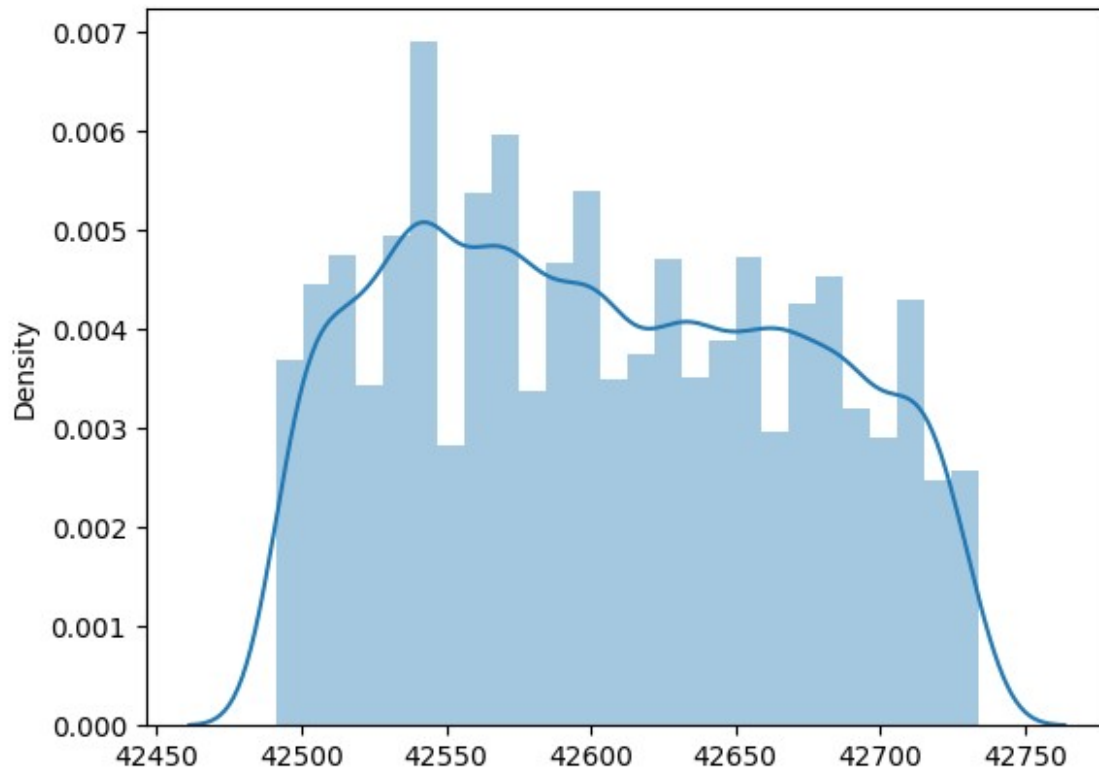
```
`distplot` is a deprecated function and will be removed in seaborn v0.14.0.
```

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>

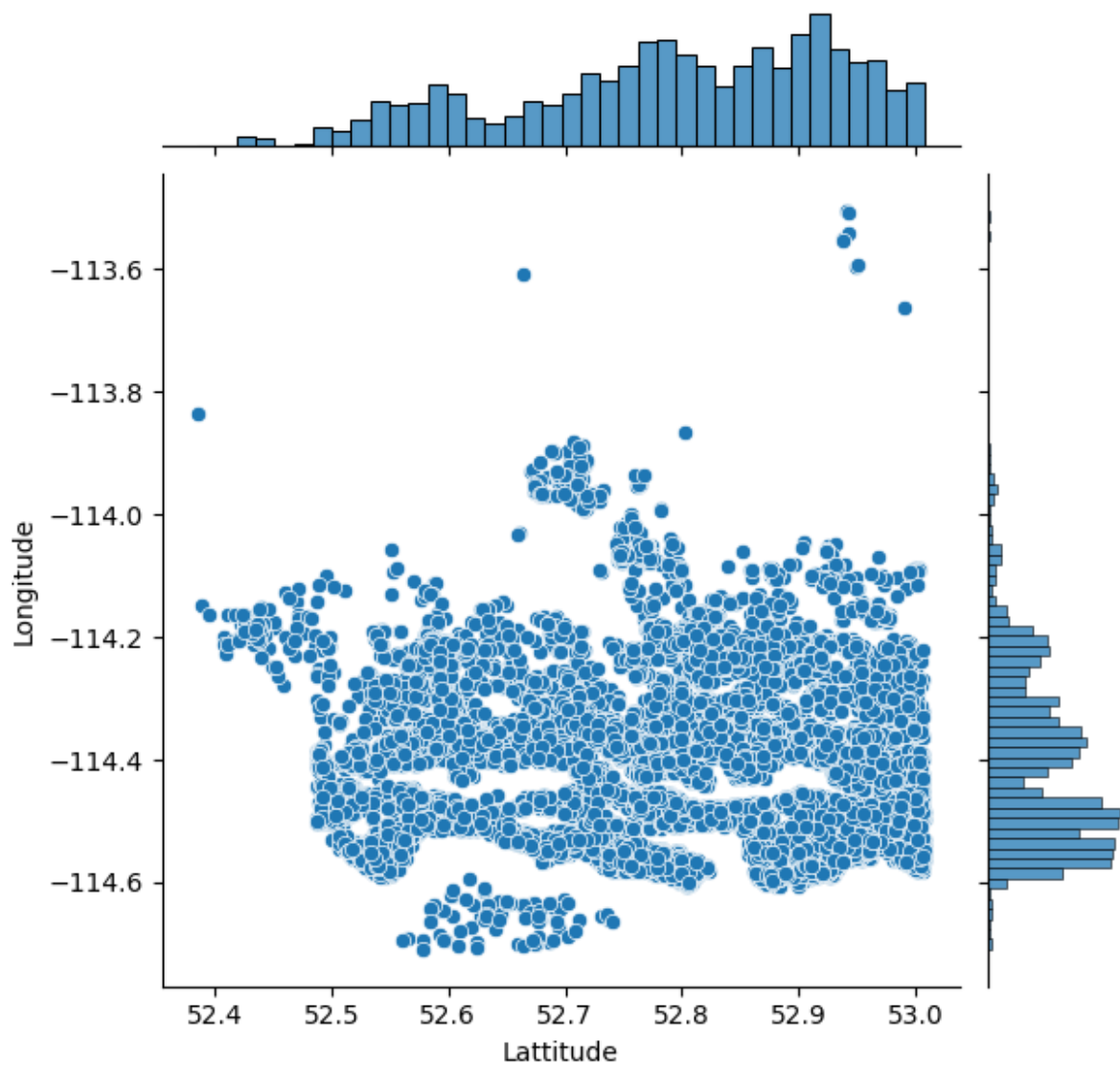
```
sns.distplot([df.Date])
```

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



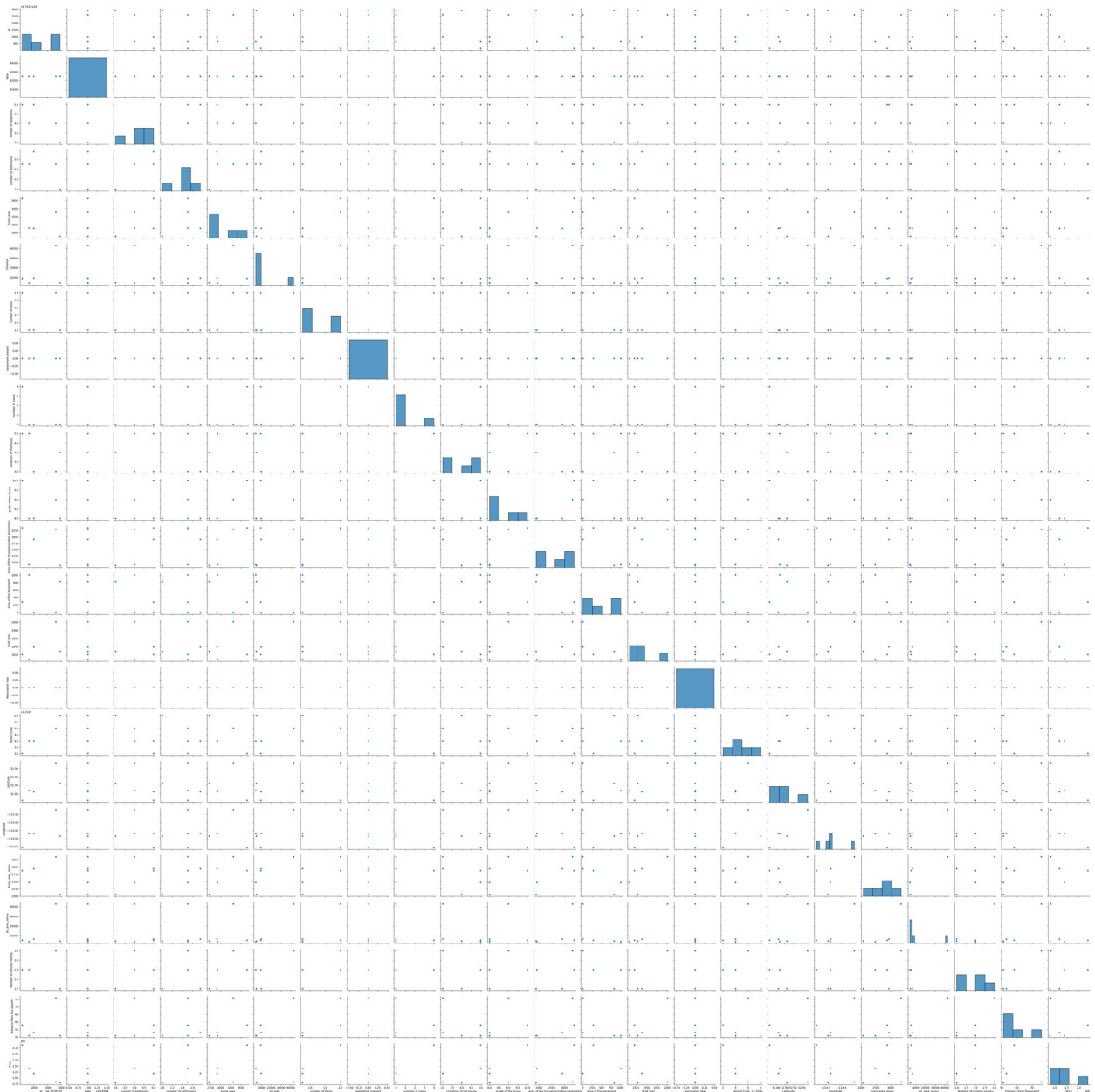
BIVARIATE ANALYSIS

```
sns.jointplot(x='Latitude', y='Longitude', data=df)  
<seaborn.axisgrid.JointGrid at 0x79c9a0166ef0>
```



MULTIVARIATE ANALYSIS

```
sns.pairplot(df)  
<seaborn.axisgrid.PairGrid at 0x79c93b093c70>
```



1. DESCRIPTIVE STATISTICS

```
df.describe()
```

	id	Date	number of bedrooms	number of bathrooms
count	5.000000e+00	5.0	5.00000	5.000000
mean	6.762811e+09	42491.0	4.20000	2.450000
std	1.231158e+03	0.0	0.83666	0.273861
min	6.762810e+09	42491.0	3.00000	2.000000

25%	6.762811e+09	42491.0	4.00000	2.500000
50%	6.762811e+09	42491.0	4.00000	2.500000
75%	6.762813e+09	42491.0	5.00000	2.500000
max	6.762813e+09	42491.0	5.00000	2.750000
	living area	lot area	number of floors	waterfront present
\				
count	5.000000	5.000000	5.000000	5.0
mean	3100.000000	14005.600000	1.700000	0.0
std	376.563408	16401.728165	0.273861	0.0
min	2710.000000	4000.000000	1.500000	0.0
25%	2910.000000	4500.000000	1.500000	0.0
50%	2920.000000	9050.000000	1.500000	0.0
75%	3310.000000	9480.000000	2.000000	0.0
max	3650.000000	42998.000000	2.000000	0.0
	number of views	condition of the house	...	Built Year
count	5.000000	5.0	...	5.000000
mean	0.800000	4.0	...	1939.800000
std	1.788854	1.0	...	35.933271
min	0.000000	3.0	...	1909.000000
25%	0.000000	3.0	...	1921.000000
50%	0.000000	4.0	...	1929.000000
75%	0.000000	5.0	...	1939.000000
max	4.000000	5.0	...	2001.000000
	Renovation Year	Postal Code	Lattitude	Longitude
count	5.0	5.000000	5.000000	5.000000
mean	0.0	122004.400000	52.899080	-114.460200
std	0.0	1.140175	0.033452	0.085905
min	0.0	122003.000000	52.864500	-114.557000
25%	0.0	122004.000000	52.885200	-114.485000
50%	0.0	122004.000000	52.887800	-114.470000
75%	0.0	122005.000000	52.904700	-114.468000
max	0.0	122006.000000	52.953200	-114.321000
	living_area_renov	lot_area_renov	Number of schools nearby	\
count	5.000000	5.000000	5.00000	

mean	2740.000000	12669.400000	1.800000
std	491.680791	16898.719827	0.83666
min	2060.000000	4000.000000	1.00000
25%	2470.000000	4500.000000	1.00000
50%	2880.000000	5400.000000	2.00000
75%	2940.000000	6600.000000	2.00000
max	3350.000000	42847.000000	3.00000

	Distance from the airport	Price
count	5.000000	5.000000e+00
mean	57.800000	1.324600e+06
std	10.568822	6.406721e+05
min	51.000000	8.050000e+05
25%	51.000000	8.380000e+05
50%	53.000000	1.200000e+06
75%	58.000000	1.400000e+06
max	76.000000	2.380000e+06

[8 rows x 23 columns]

1. HANDLING MISSING VALUES

```
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	

There are No Null Values for the given dataset