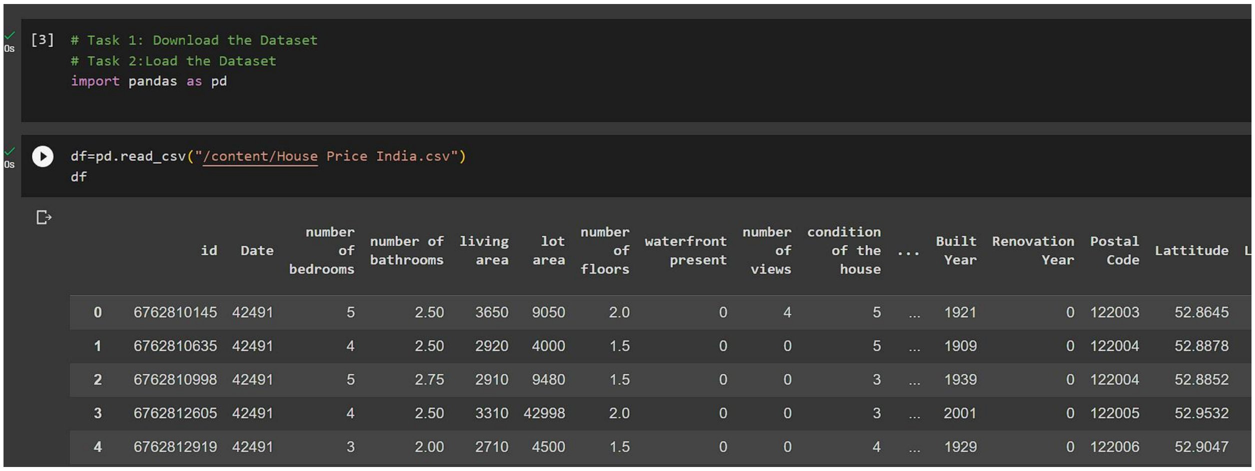
Name: Kelvin J Anil
Reg No : 21BCE0002

Assignment 2, AI ML, Evening Batch

Colab:
https://colab.research.google.com/drive/1YPGiT3ZzoLYuZGVQvHxFKI0AWMyblBTT?usp=sharing

1. Download the dataset: Dataset
2. Load the dataset.

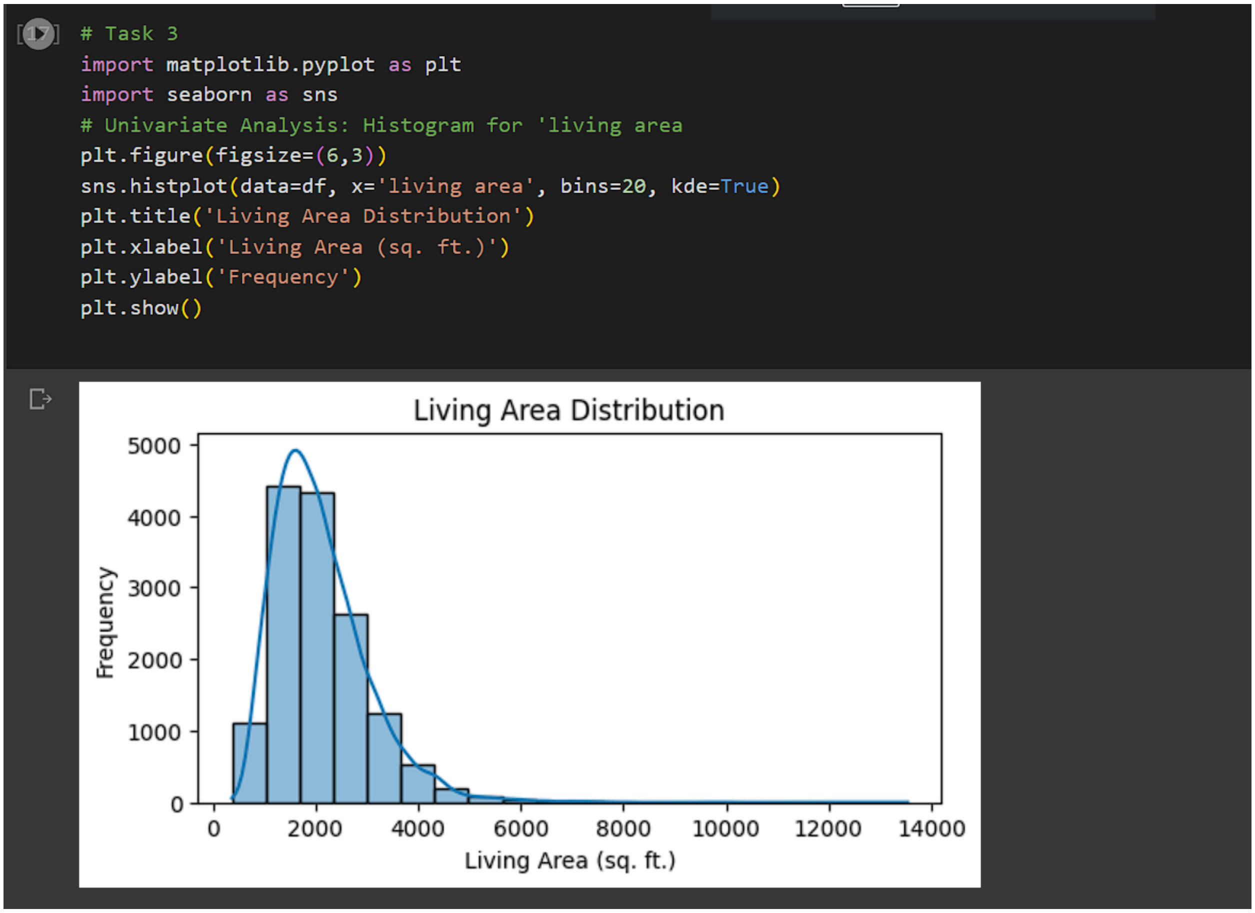


```
[3] # Task 1: Download the Dataset
    # Task 2:Load the Dataset
    import pandas as pd
```

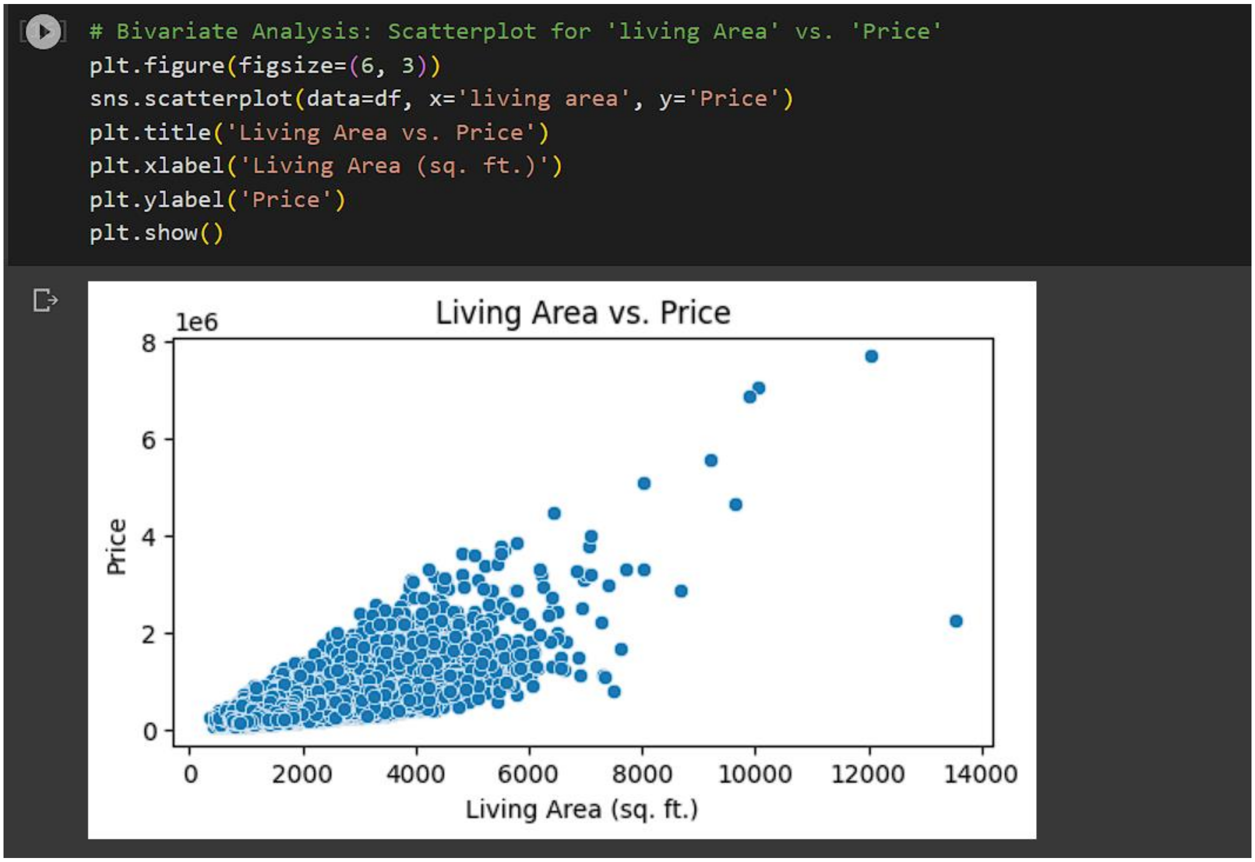
```
df=pd.read_csv("/content/House Price India.csv")
df
```

	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	L
0	6762810145	42491	5	2.50	3650	9050	2.0	0	4	5	...	1921	0	122003	52.8645	
1	6762810635	42491	4	2.50	2920	4000	1.5	0	0	5	...	1909	0	122004	52.8878	
2	6762810998	42491	5	2.75	2910	9480	1.5	0	0	3	...	1939	0	122004	52.8852	
3	6762812605	42491	4	2.50	3310	42998	2.0	0	0	3	...	2001	0	122005	52.9532	
4	6762812919	42491	3	2.00	2710	4500	1.5	0	0	4	...	1929	0	122006	52.9047	

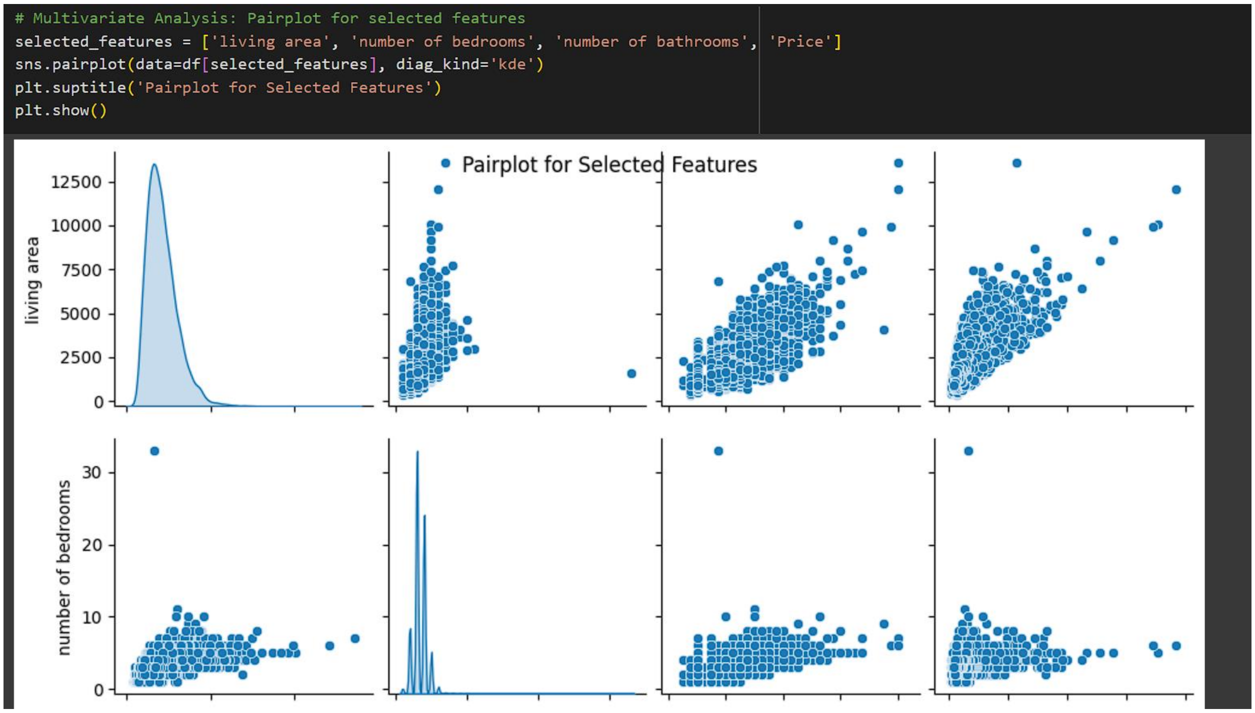
3. Perform the Below Visualizations.
   - Univariate Analysis
   - Bi - Variate Analysis
   - Multivariate Analysis



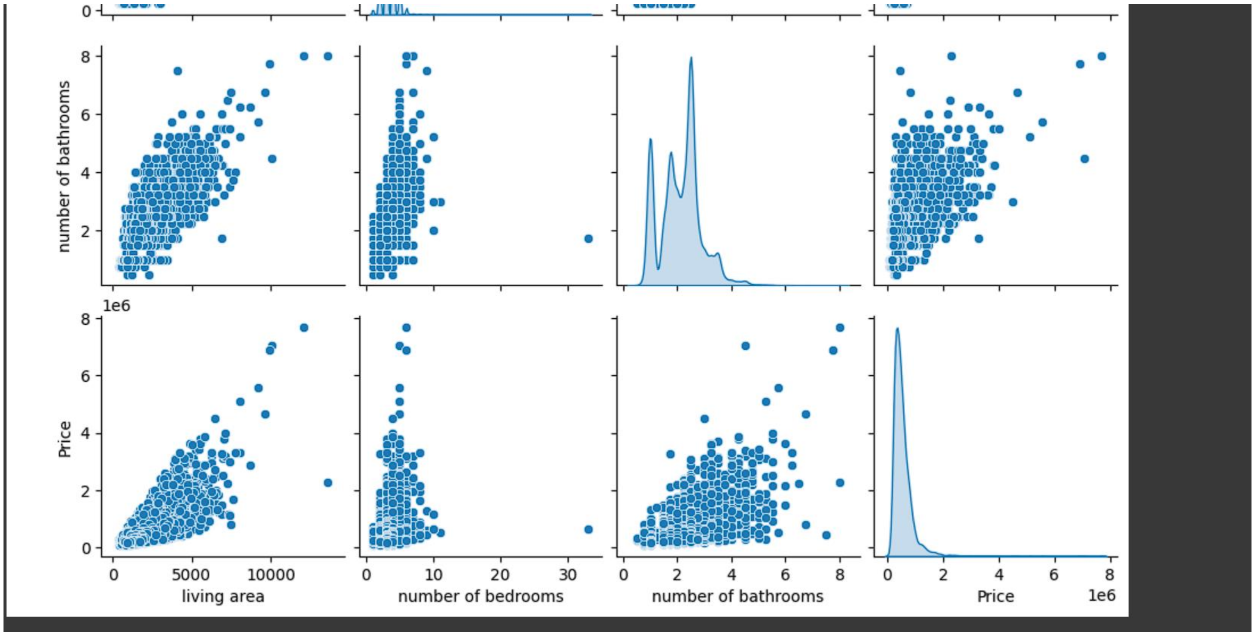
```python
# Task 3
import matplotlib.pyplot as plt
import seaborn as sns
# Univariate Analysis: Histogram for 'living area
plt.figure(figsize=(6,3))
sns.histplot(data=df, x='living area', bins=20, kde=True)
plt.title('Living Area Distribution')
plt.xlabel('Living Area (sq. ft.)')
plt.ylabel('Frequency')
plt.show()
```



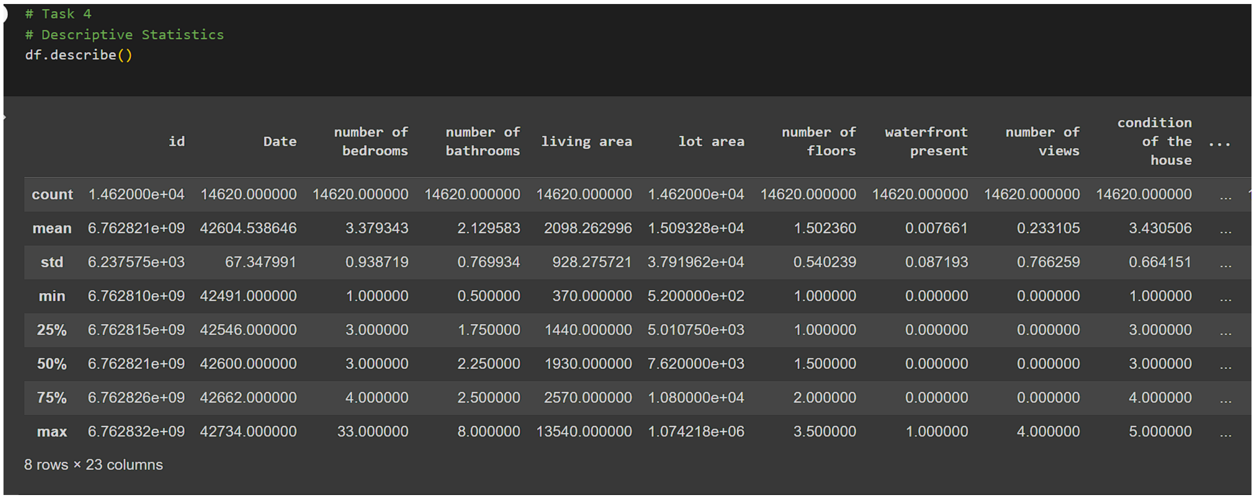
```
# Bivariate Analysis: Scatterplot for 'living Area' vs. 'Price'
plt.figure(figsize=(6, 3))
sns.scatterplot(data=df, x='living area', y='Price')
plt.title('Living Area vs. Price')
plt.xlabel('Living Area (sq. ft.)')
plt.ylabel('Price')
plt.show()
```



```
# Multivariate Analysis: Pairplot for selected features
selected_features = ['living area', 'number of bedrooms', 'number of bathrooms', 'Price']
sns.pairplot(data=df[selected_features], diag_kind='kde')
plt.suptitle('Pairplot for Selected Features')
plt.show()
```



4. Perform descriptive statistics on the dataset.

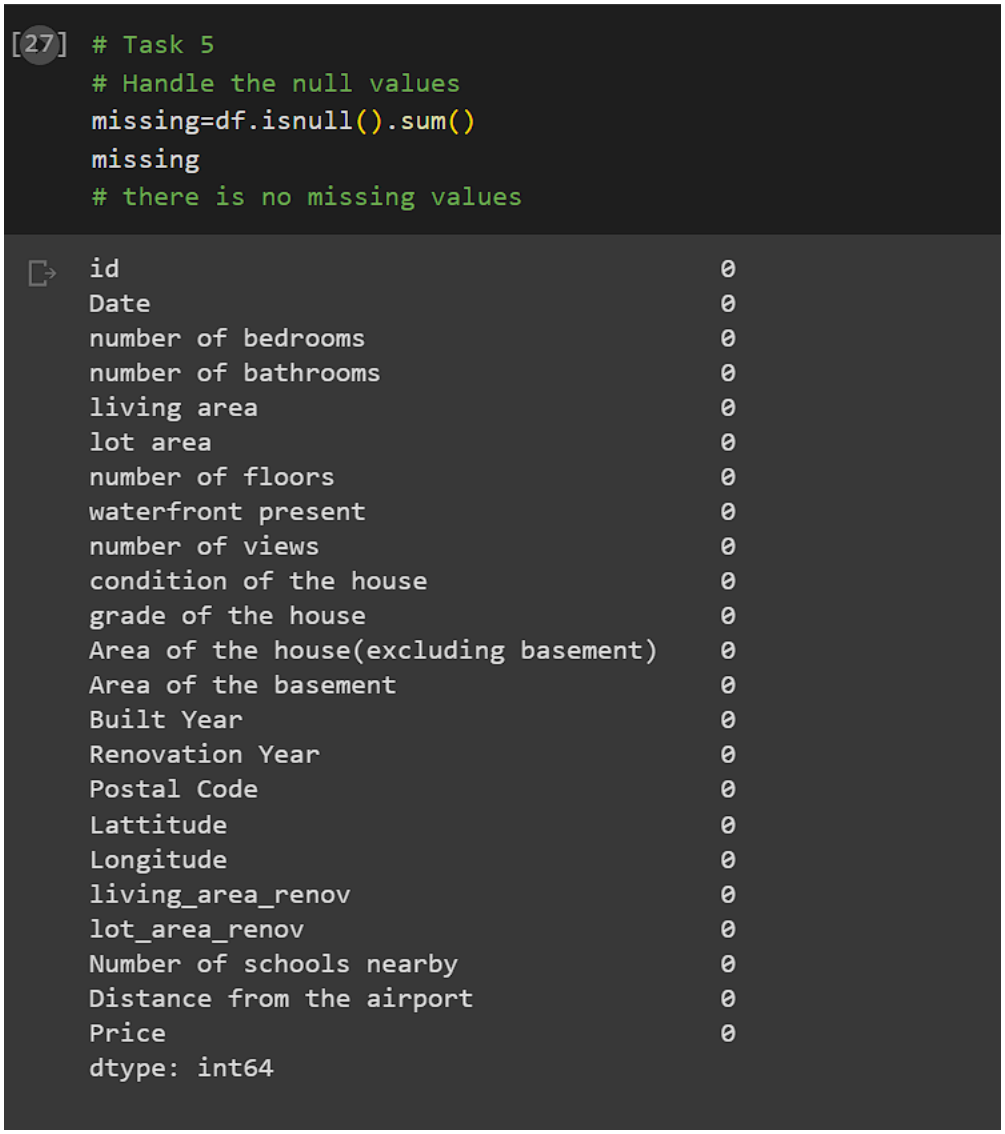


```
# Task 4
# Descriptive Statistics
df.describe()
```

	id	Date	number of bedrooms	number of bathrooms	living area	lot area	number of floors	waterfront present	number of views	condition of the house	...
count	1.462000e+04	14620.000000	14620.000000	14620.000000	14620.000000	1.462000e+04	14620.000000	14620.000000	14620.000000	14620.000000	...
mean	6.762821e+09	42604.538646	3.379343	2.129583	2098.262996	1.509328e+04	1.502360	0.007661	0.233105	3.430506	...
std	6.237575e+03	67.347991	0.938719	0.769934	928.275721	3.791962e+04	0.540239	0.087193	0.766259	0.664151	...
min	6.762810e+09	42491.000000	1.000000	0.500000	370.000000	5.200000e+02	1.000000	0.000000	0.000000	1.000000	...
25%	6.762815e+09	42546.000000	3.000000	1.750000	1440.000000	5.010750e+03	1.000000	0.000000	0.000000	3.000000	...
50%	6.762821e+09	42600.000000	3.000000	2.250000	1930.000000	7.620000e+03	1.500000	0.000000	0.000000	3.000000	...
75%	6.762826e+09	42662.000000	4.000000	2.500000	2570.000000	1.080000e+04	2.000000	0.000000	0.000000	4.000000	...
max	6.762832e+09	42734.000000	33.000000	8.000000	13540.000000	1.074218e+06	3.500000	1.000000	4.000000	5.000000	...

8 rows × 23 columns

5. Handle the Missing values.



```
# Task 5
# Handle the null values
missing=df.isnull().sum()
missing
# there is no missing values
```

```
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
```