

ARJUN COLLEGE OF TECHNOLOGY

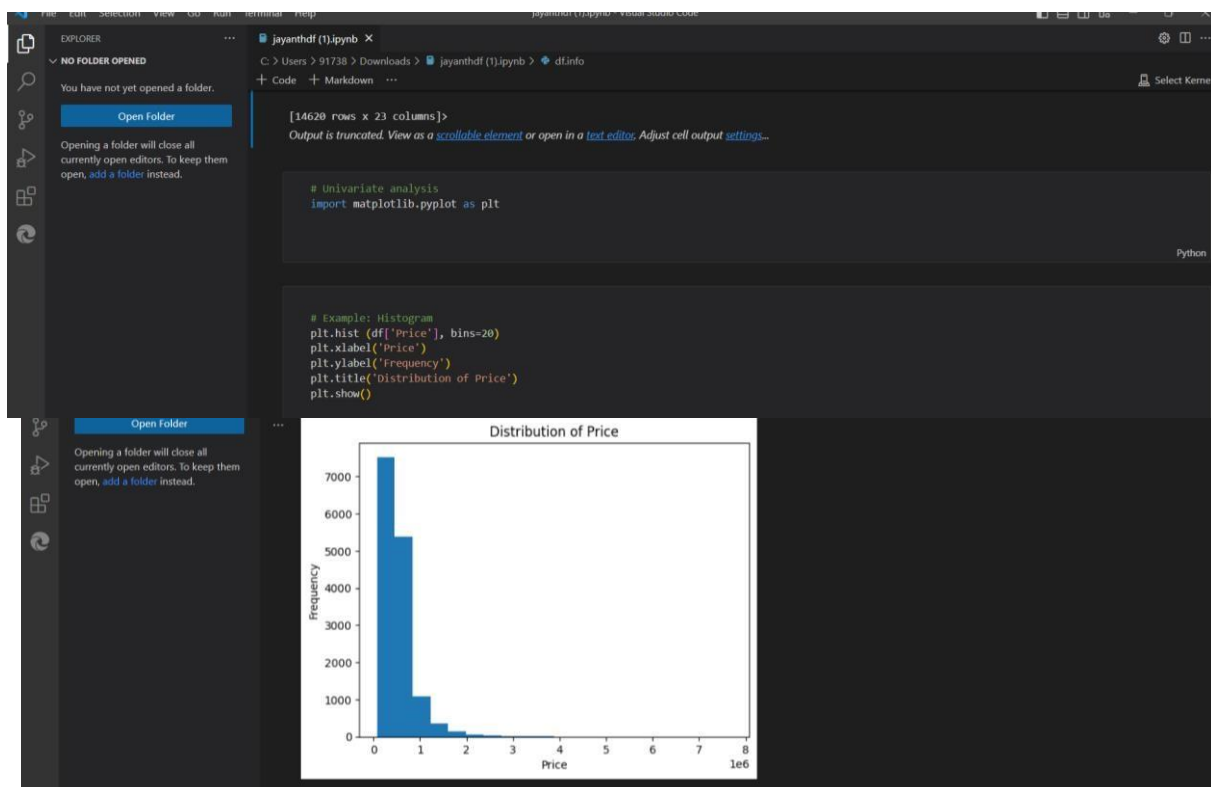
ASSIGNMENT – 3

NAAN MUDHALVAN

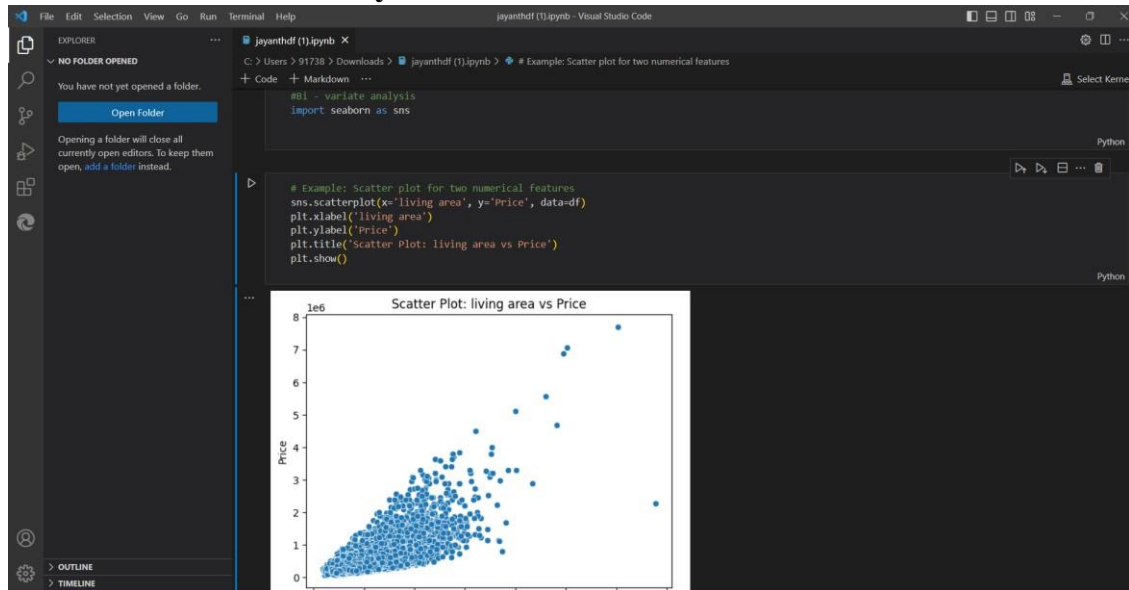
NAME : TOTAMSETTY SAICHARAN

REG.NO : 723920104049

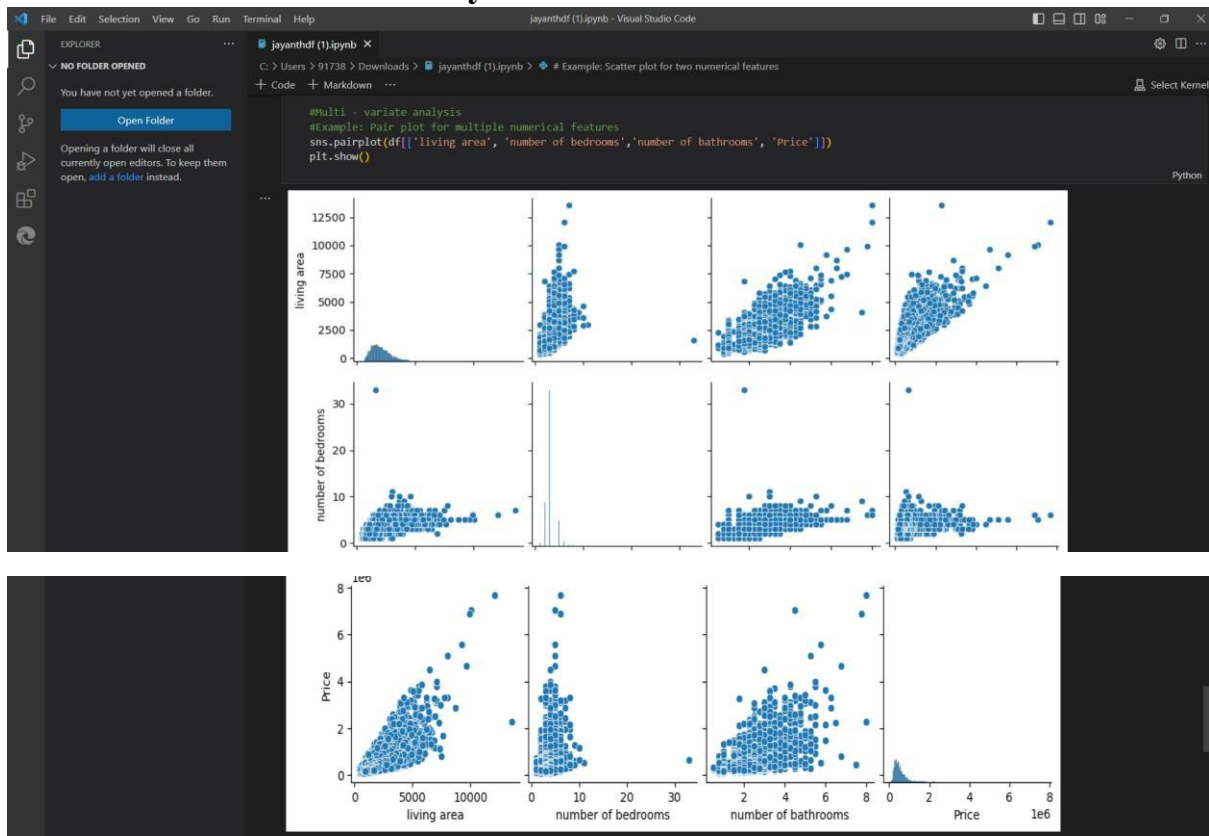
STEP-1: Univariate Analysis



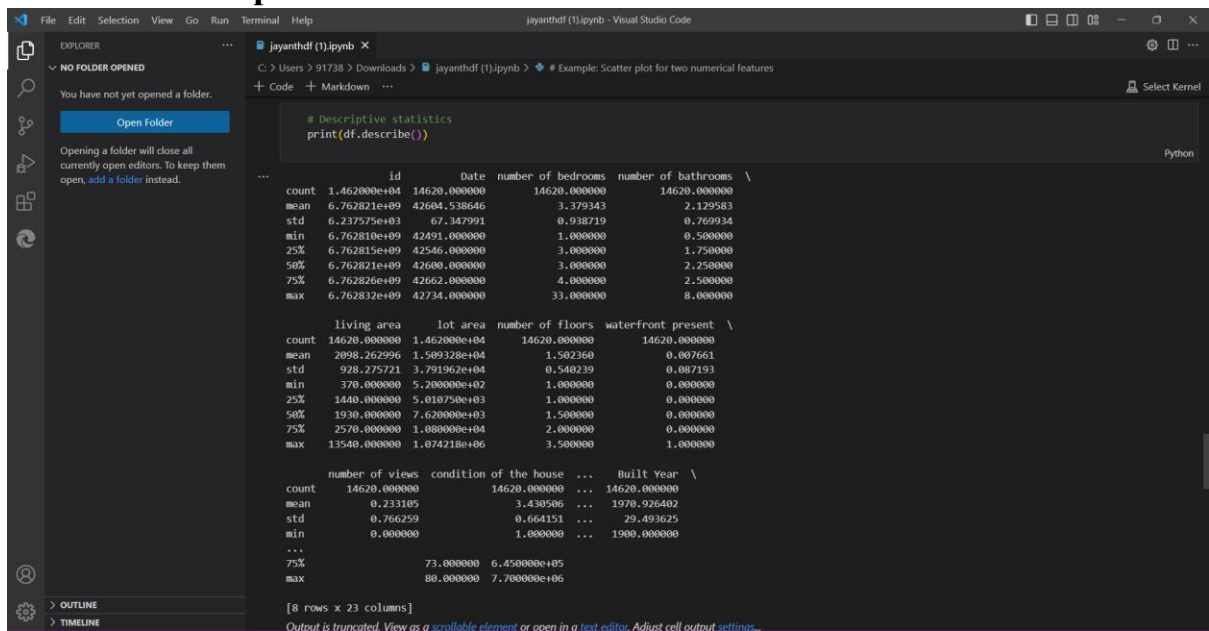
STEP-2: Bi - variate Analysis



STEP-3: Multi - variate Analysis



STEP-4: Descriptive statistics



The screenshot shows a Jupyter Notebook in Visual Studio Code. The code cell contains the following Python code:

```
# Descriptive statistics
print(df.describe())
```

The output of the code is a summary of the data, showing statistics for various features. The output is truncated, showing only the first 8 rows of the 23 columns.

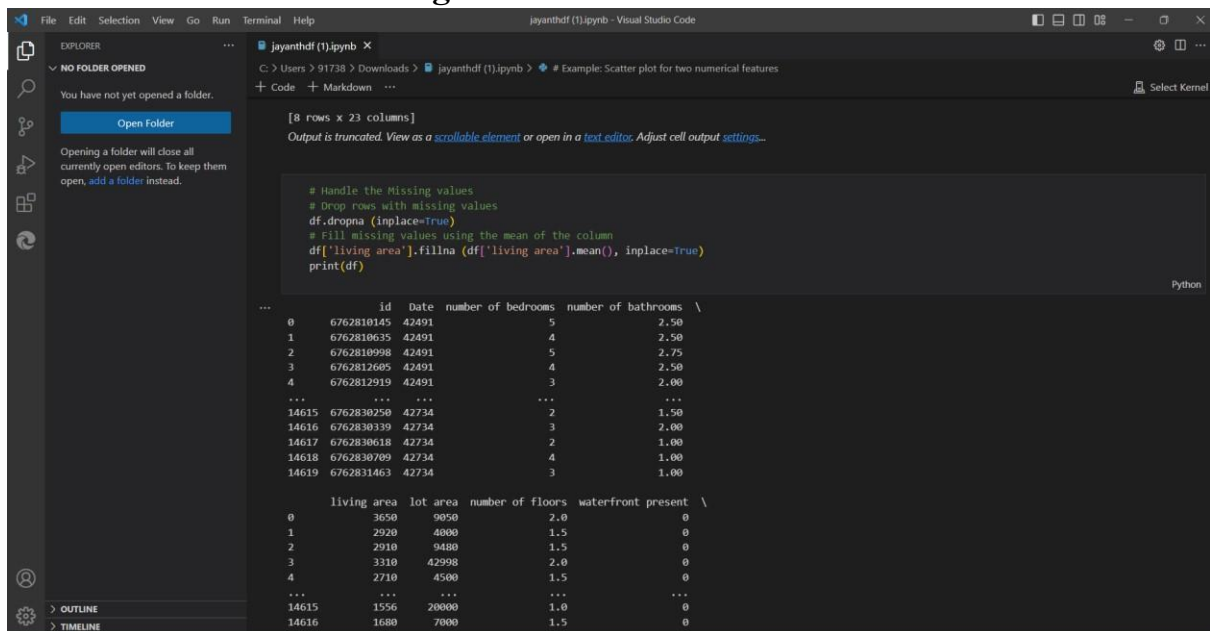
	id	Date	number of bedrooms	number of bathrooms	\
count	1.462000e+04	14620.000000	14620.000000	14620.000000	
mean	6.762821e+09	42684.538546	3.379343	2.129583	
std	6.237575e+03	67.347991	0.938719	0.769934	
min	6.762810e+09	42491.000000	1.000000	0.500000	
25%	6.762815e+09	42546.000000	3.000000	1.750000	
50%	6.762821e+09	42680.000000	3.000000	2.250000	
75%	6.762826e+09	42682.000000	4.000000	2.500000	
max	6.762832e+09	42734.000000	33.000000	8.000000	

	living area	lot area	number of floors	waterfront present	\
count	14620.000000	1.462000e+04	14620.000000	14620.000000	
mean	2098.262996	1.509328e+04	1.502360	0.007661	
std	928.275721	3.791962e+04	0.540239	0.087193	
min	370.000000	5.200000e+02	1.000000	0.000000	
25%	1440.000000	5.010750e+03	1.000000	0.000000	
50%	1930.000000	7.620000e+03	1.500000	0.000000	
75%	2570.000000	1.080000e+04	2.000000	0.000000	
max	13540.000000	1.074218e+06	3.500000	1.000000	

	number of views	condition of the house	... Built Year	\
count	14620.000000	14620.000000	... 14620.000000	
mean	0.233105	3.438586	... 1970.926402	
std	0.766259	0.664151	... 29.493625	
min	0.000000	1.000000	... 1900.000000	
...				
75%	73.000000	6.450000e+05		
max	80.000000	7.700000e+06		

[8 rows x 23 columns]
Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...

STEP-4: Handle the Missing values



The screenshot shows a Jupyter Notebook in Visual Studio Code. The code cell contains the following Python code:

```
# Handle the Missing values
# Drop rows with missing values
df.dropna(inplace=True)
# Fill missing values using the mean of the column
df['living area'].fillna(df['living area'].mean(), inplace=True)
print(df)
```

The output of the code is a summary of the data, showing statistics for various features. The output is truncated, showing only the first 8 rows of the 23 columns.

	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	9400	1.5	0	
3	3310	42900	2.0	0	
4	2710	4500	1.5	0	
...	
14615	1556	20000	1.0	0	
14616	1680	7000	1.5	0	
...	