



NAAN MUDHALVAN MANDATORY COURSE

AUGUST – DECEMBER 2023 (ODD SEMESTER)

DATA ANALYTICS WITH TABLEAU

Course Name:

Data Analytics With Tableau

Team Number:

15

Assignment Submitted To:

Anna University – Naan Mudhalvan

Year:

IV

Department:

B.Tech Information technology

Semester

VII

1.YOGALAKSHMI .J

2.VENKATRAMAN.M

Team Members:

3.DHINESHKUMAR.N

4.GOVINDARAJ.A

Guided By:

Ms.M.Noorunnisha – AP / CSE
(Mamce - 8126)

SPOC Name:

Dr.K.Manikandan



ASSIGNMENT QUESTION

Perform the Below Tasks to complete the assignment:-

Tasks:-

1. Download the dataset: Dataset(<https://www.kaggle.com/datasets/mohamedafsal007/house-price-dataset-of-india/download?datasetVersionNumber=1>)
2. Load the dataset.
3. Perform the Below Visualizations.
 - Univariate Analysis
 - Bi - Variate Analysis
 - Multi-Variate Analysis
4. Perform descriptive statistics on the dataset.
5. Handle the Missing values



ASSIGNMENT SOLUTION SCREENSHOTS

DOWNLOAD DATASET FROM THE LINK IN THE PDF :

The screenshot shows a web browser window with the URL <https://www.kaggle.com/datasets/mohamedafsal007/house-price-dataset-of-india>. The page displays the dataset details, including its title, description, and various metrics. A prominent 'Download (492 kB)' button is visible. The left sidebar of the Kaggle interface is also visible.

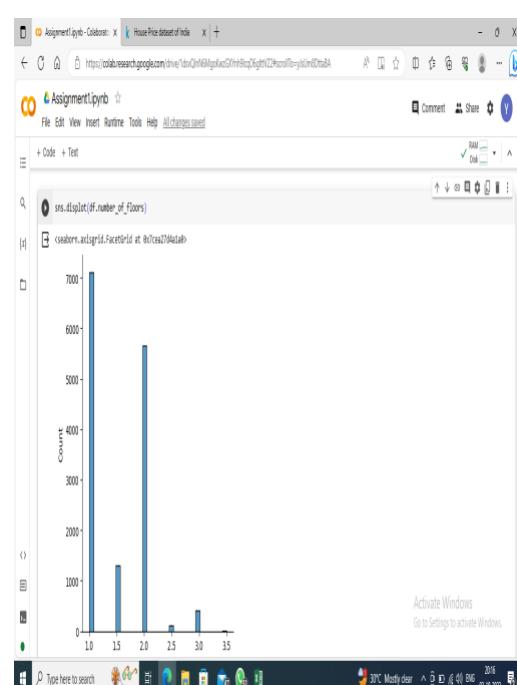
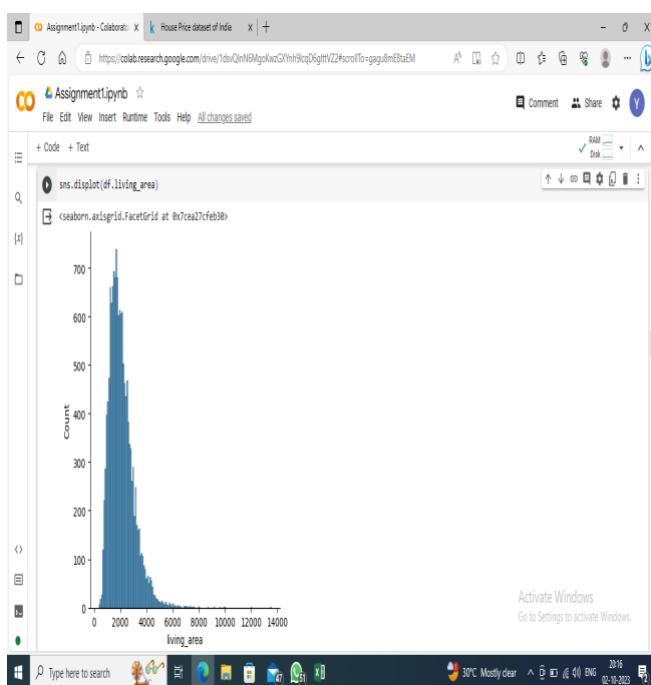
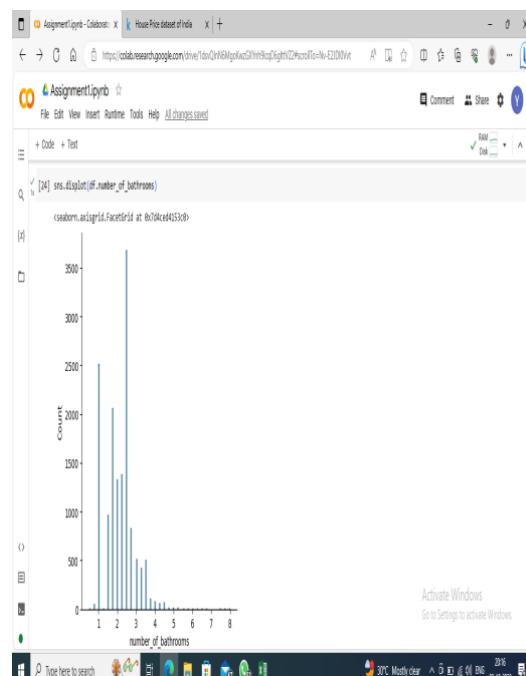
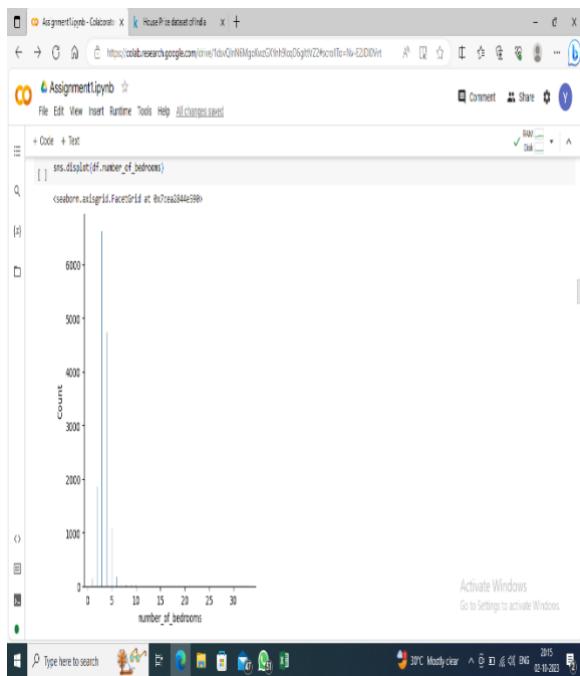
LOAD THE DATASET INTO GOOGLE COLAB :

The screenshot shows a Google Colab notebook titled 'Assignment1.ipynb'. A code cell at the top imports pandas, numpy, matplotlib.pyplot, and seaborn, and reads a CSV file named 'house price detail assignment.csv'. Below the code cell, a table displays the first 14 rows of the dataset, including columns like id, Date, number_of_bedrooms, number_of_bathrooms, living_area, lot_area, number_of_floors, and waterfront_prc.

	id	Date	number_of_bedrooms	number_of_bathrooms	living_area	lot_area	number_of_floors	waterfront_prc
0	6762810145	42491	5	2.50	3650	9050	2.0	
1	6762810635	42491	4	2.50	2920	4000	1.5	
2	6762810598	42491	5	2.75	2910	9480	1.5	
3	6762812605	42491	4	2.50	3310	42998	2.0	
4	6762812919	42491	3	2.00	2710	4500	1.5	
...	
14615	6762830250	42734	2	1.50	1556	20000	1.0	
14616	6762830339	42734	3	2.00	1680	7000	1.5	
14617	6762830618	42734	2	1.00	1070	6120	1.0	
14618	6762830709	42734	4	1.00	1030	6621	1.0	

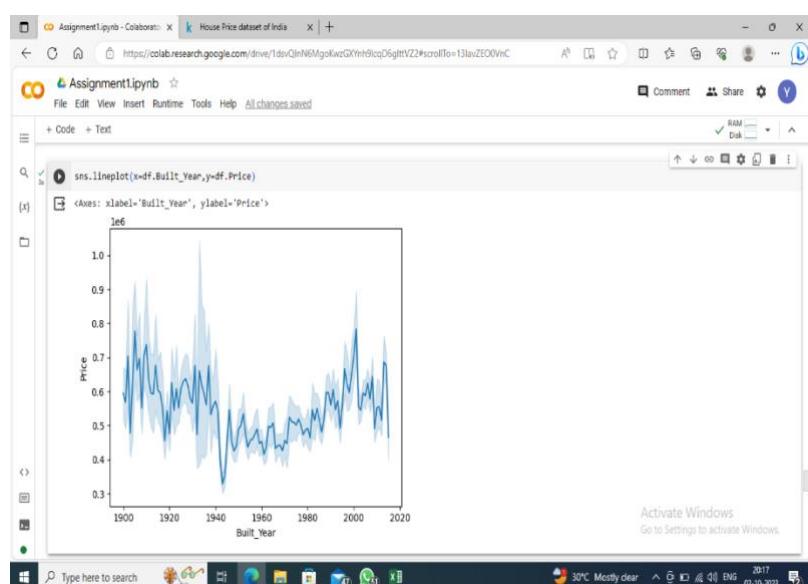
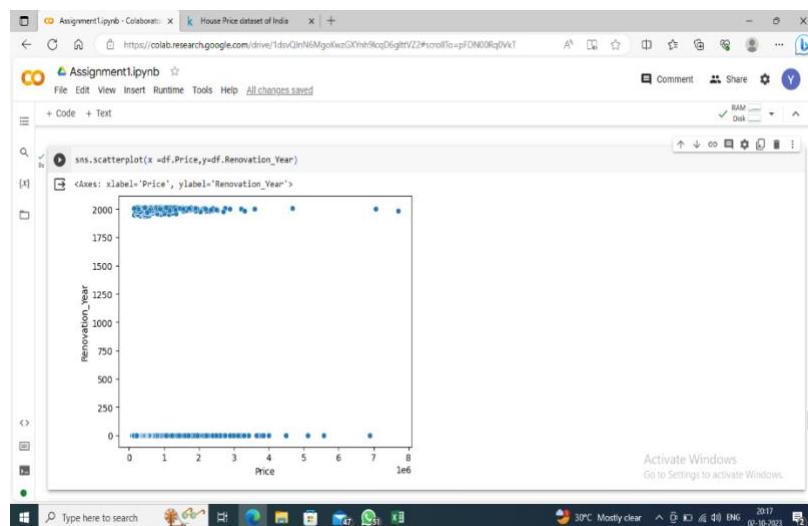
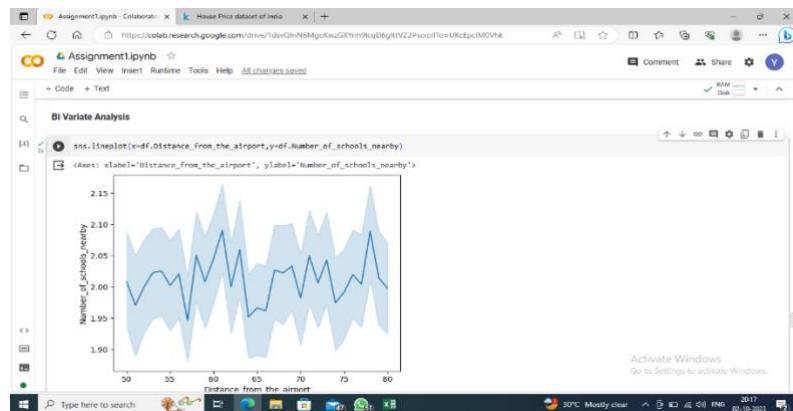


UNIVARIATE ANALYSIS :



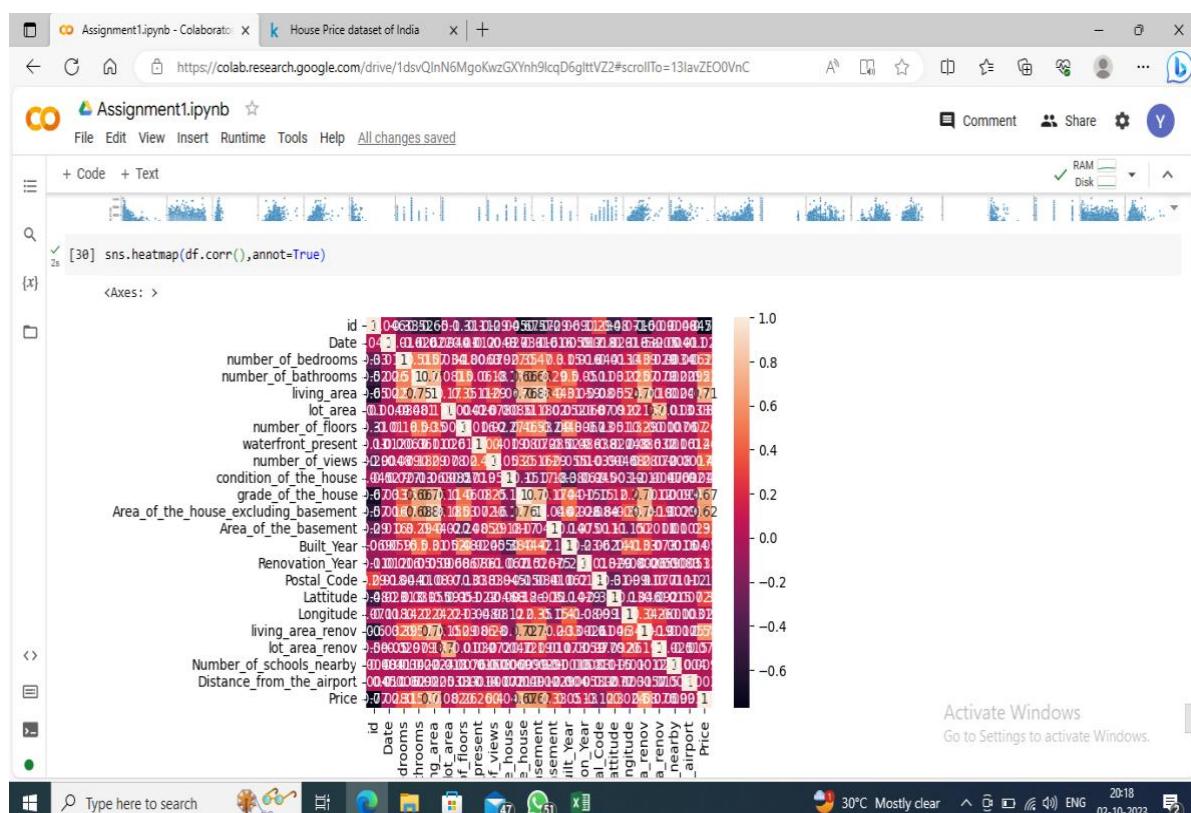
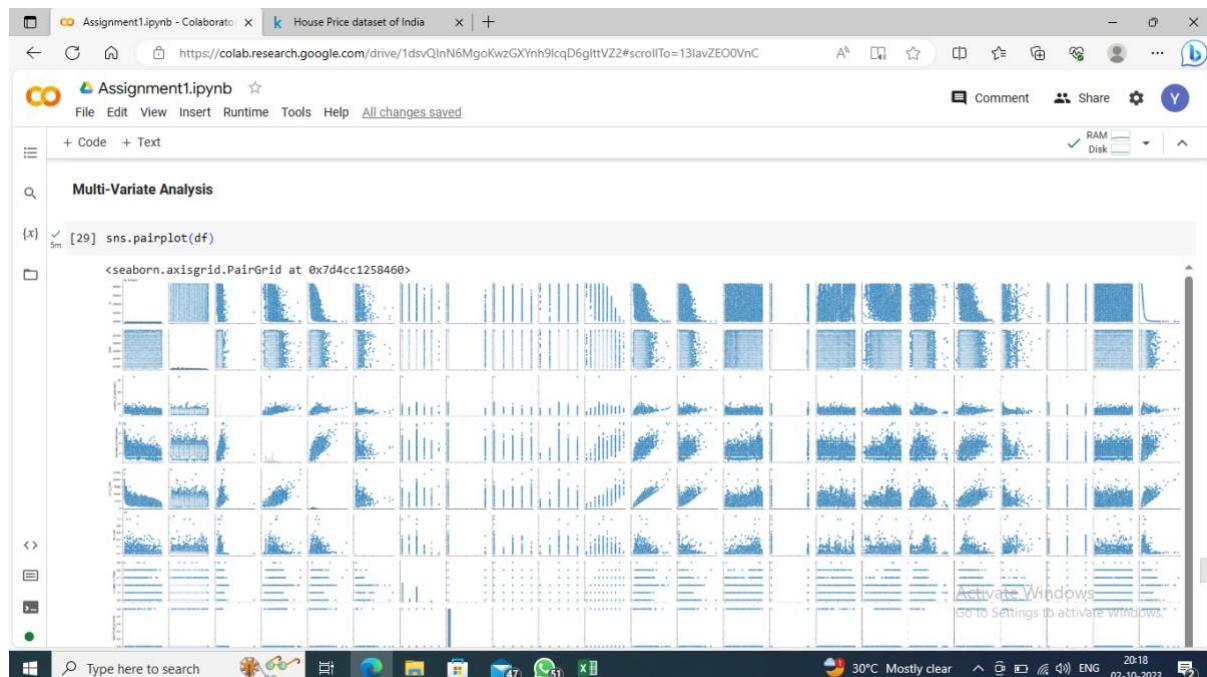


BI-VARIATE ANALYSIS :





MULTI-VARIATE ANALYSIS:





PERFORM DESCRIPTIVE STATISTICS ON THE DATASET :

The screenshot shows a Jupyter Notebook interface with the following code and output:

```
[ ] 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	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	
std	6.237575e+03	67.347991	0.938719	0.769934	928.275721	3.791962e+04	0.540239	0.087193	0.766259	
min	6.762810e+09	42491.000000	1.000000	0.500000	370.000000	5.200000e+02	1.000000	0.000000	0.000000	
25%	6.762815e+09	42546.000000	3.000000	1.750000	1440.000000	5.010750e+03	1.000000	0.000000	0.000000	
50%	6.762821e+09	42600.000000	3.000000	2.250000	1930.000000	7.620000e+03	1.500000	0.000000	0.000000	
75%	6.762826e+09	42662.000000	4.000000	2.500000	2570.000000	1.080000e+04	2.000000	0.000000	0.000000	
max	6.762832e+09	42734.000000	33.000000	8.000000	13540.000000	1.074218e+06	3.500000	1.000000	4.000000	

8 rows x 23 columns

HANDLE THE MISSING VALUES :

The screenshot shows a Jupyter Notebook interface with the following code and output:

```
[ ] df.isnull().any()
```

	id	Date	number_of_bedrooms	number_of_bathrooms	living_area	lot_area	number_of_floors	waterfront_present	number_of_views	cond
{x}	id	False	False	False	False	False	False	False	False	False
	Date	False	False	False	False	False	False	False	False	False
	number_of_bedrooms	False	False	False	False	False	False	False	False	False
	number_of_bathrooms	False	False	False	False	False	False	False	False	False
	living_area	False	False	False	False	False	False	False	False	False
	lot_area	False	False	False	False	False	False	False	False	False
	number_of_floors	False	False	False	False	False	False	False	False	False
	waterfront_present	False	False	False	False	False	False	False	False	False
	number_of_views	False	False	False	False	False	False	False	False	False
	condition_of_the_house	False	False	False	False	False	False	False	False	False
	grade_of_the_house	False	False	False	False	False	False	False	False	False
	Area_of_the_house_excluding_basement	False	False	False	False	False	False	False	False	False
	Area_of_the_basement	False	False	False	False	False	False	False	False	False
	Built_Year	False	False	False	False	False	False	False	False	False
	Renovation_Year	False	False	False	False	False	False	False	False	False
	Postal_Code	False	False	False	False	False	False	False	False	False
	Latitude	False	False	False	False	False	False	False	False	False
	Longitude	False	False	False	False	False	False	False	False	False
	living_area_renov	False	False	False	False	False	False	False	False	False
	lot_area_renov	False	False	False	False	False	False	False	False	False
	Number_of_schools_nearby	False	False	False	False	False	False	False	False	False
	Distance_from_the_airport	False	False	False	False	False	False	False	False	False
	Price	False	False	False	False	False	False	False	False	False
	dtype: bool									

```
[ ] df.isnull().sum()
```



In my dataset there is no null values.

So I used the data set as it as.

MY ASSIGNMENT GOOGLE COLAB LINK FOR REFERENCE :

LINK :

<https://colab.research.google.com/drive/1dsVQInN6MgoKwzGXNh9lcqD6gltvZ2?usp=sharing>