



**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
Team Members:	1.YOGALAKSHMI .J 2.VENKATRAMAN.M 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 the Kaggle dataset page for 'House Price dataset of India' by Mohamed Afsal. The dataset is described as containing 23 columns and 14620 rows with no null values. It includes a 'Data Card', 'Code (16)', and 'Discussion (2)'. The 'About Dataset' section mentions it was useful for a project. The 'Usability' score is 8.82. The 'License' is 'Other (specified in description)'. The 'Expected update frequency' is 'Not specified'. The 'Tags' section is empty. The 'Download (492 kB)' button is visible.

LOAD THE DATASET INTO GOOGLE COLAB :

The screenshot shows a Google Colab notebook titled 'Assignment1.ipynb'. The code cell [7] contains the following Python code:

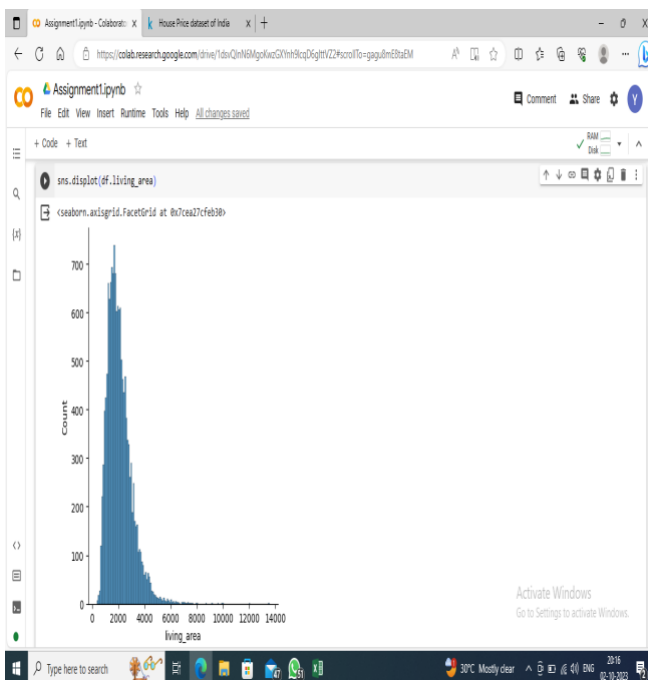
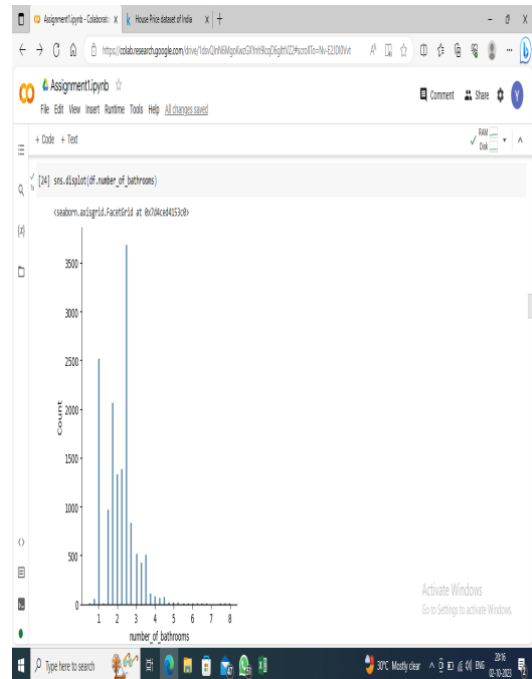
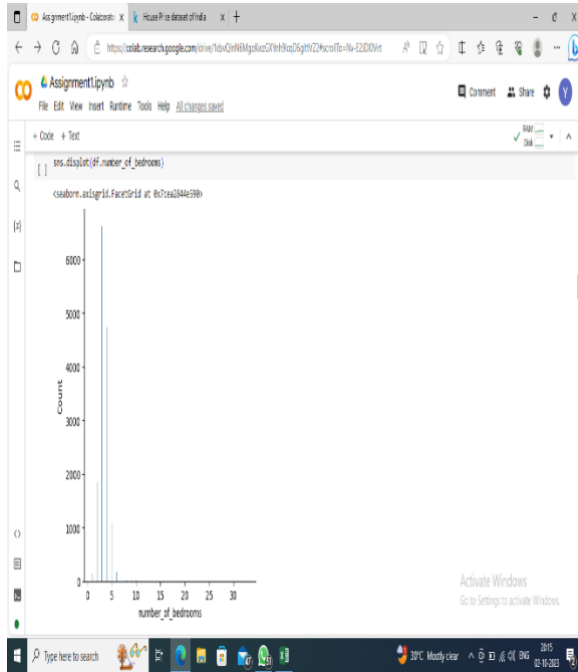
```
[7] import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
df = pd.read_csv('house price detail assignment.csv')
df
```

The output of the code is a pandas DataFrame with the following columns: id, Date, number_of_bedrooms, number_of_bathrooms, living_area, lot_area, number_of_floors, and waterfront_pre. The DataFrame contains 14618 rows of data.

	id	Date	number_of_bedrooms	number_of_bathrooms	living_area	lot_area	number_of_floors	waterfront_pre
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	

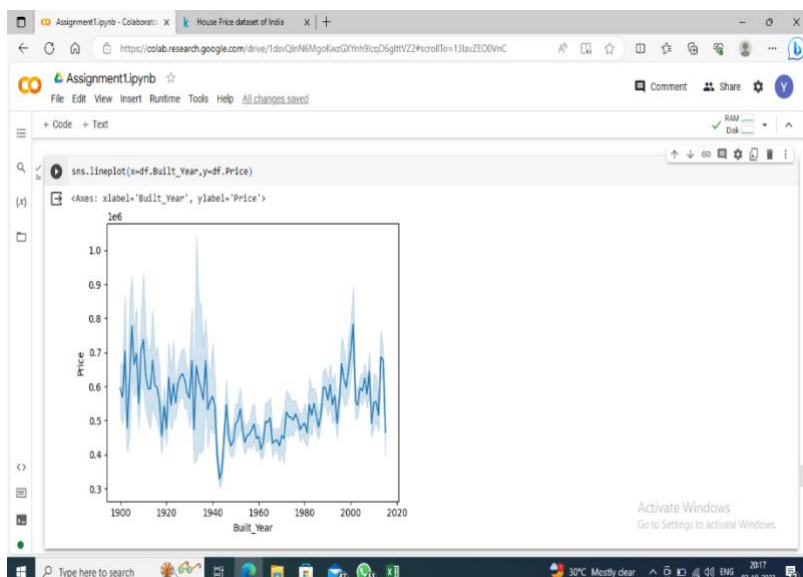
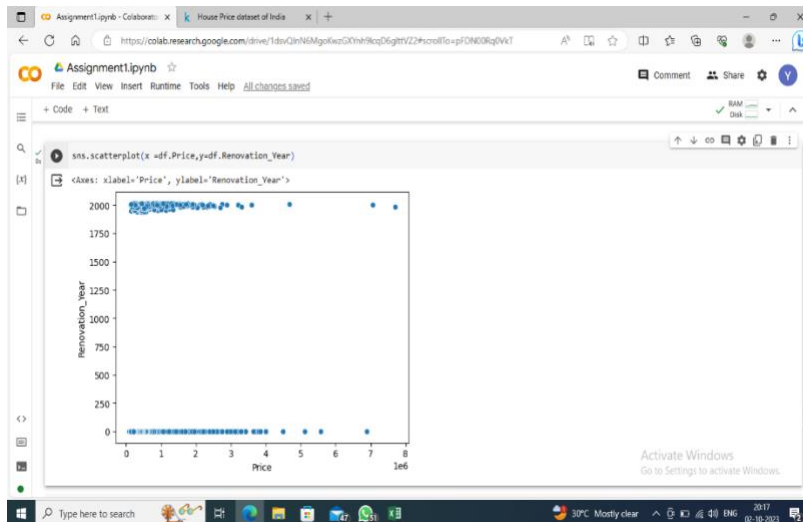
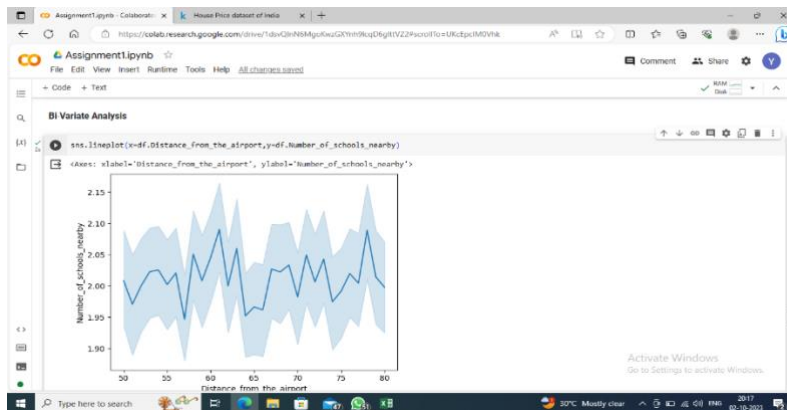


UNIVARIATE ANALYSIS :





BI-VARIATE ANALYSIS :





PERFORM DESCRIPTIVE STATISTICS ON THE DATASET :

Assignment1.ipynb - Collaborator x House Price dataset of India x +

https://colab.research.google.com/drive/1dsVQlnN6MgoKwzGXyh9lqD6glttVZ2#scrollTo=13lavZEO0VnC

Assignment1.ipynb

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

```
[ ] 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	14620.000000	1.462000e+04	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+03	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

Activate Windows
Go to Settings to activate Windows.

HANDLE THE MISSING VALUES :

Assignment1.ipynb - Collaborator x House Price dataset of India x +

https://colab.research.google.com/drive/1dsVQlnN6MgoKwzGXyh9lqD6glttVZ2#scrollTo=13lavZEO0VnC

Assignment1.ipynb

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

```
[ ] 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
Latitude	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()
```

Activate Windows
Go to Settings to activate Windows.



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/1dsvQInN6MgoKwzGXYnh9lcqD6glttVZ2?usp=sharing>