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 QUESION

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.

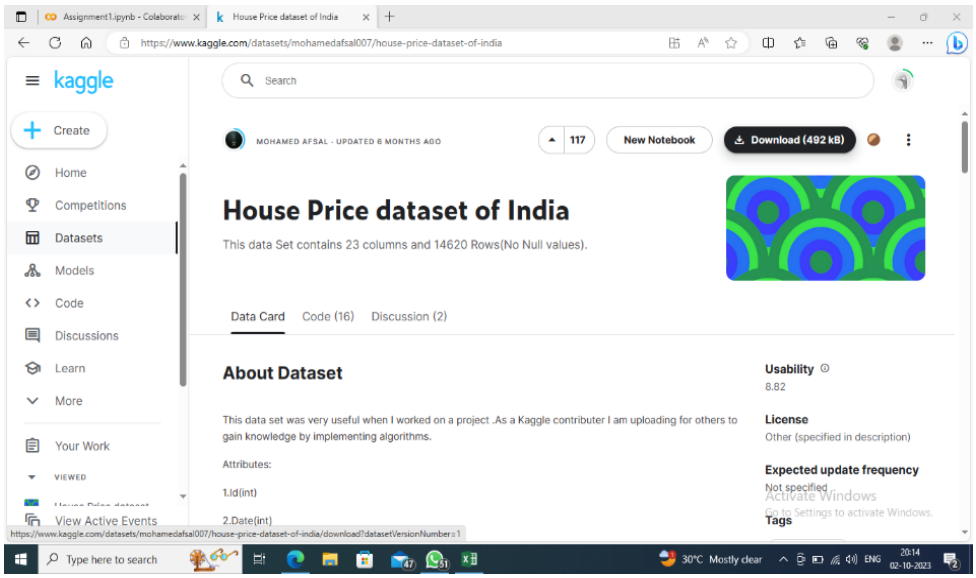
- Univariate Analysis
- Bi - Variate Analysis
- Multi-Variate Analysis

4. Perform descriptive statistics on the dataset.

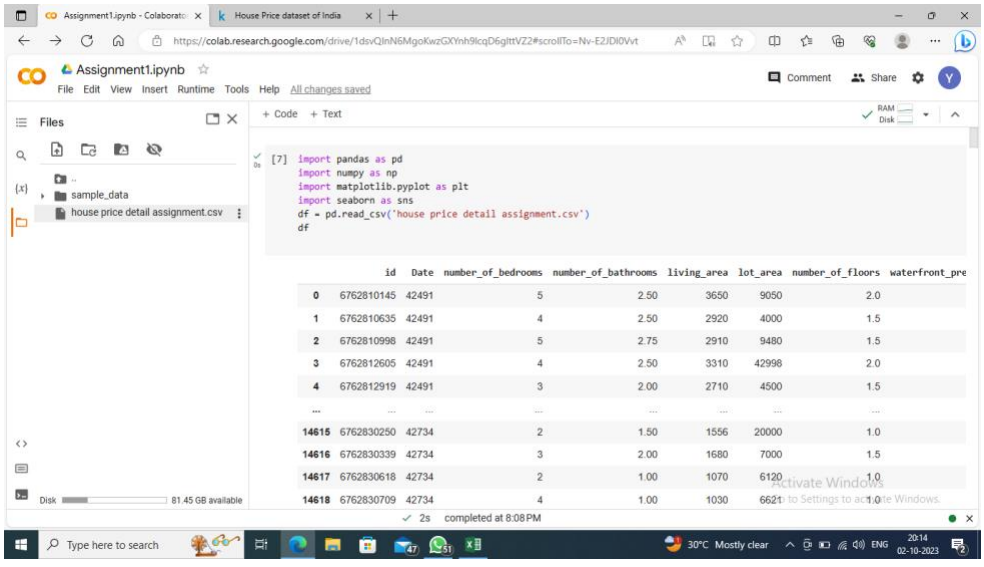
5. Handle the Missing values

ASSIGNMENT SOLUTION SCREENSHOTS

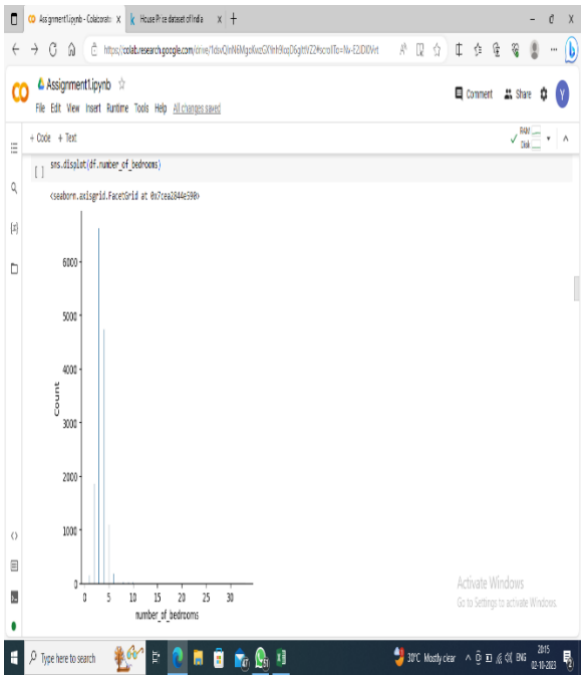
DOWNLOAD DATASET FROM THE LINK IN THE PDF :

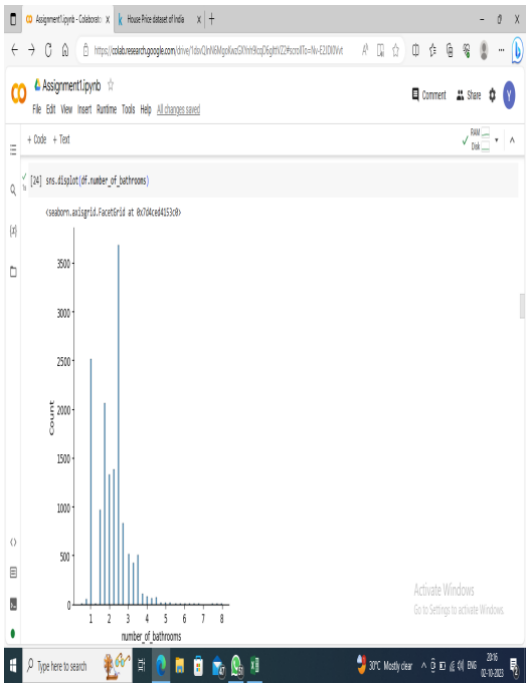


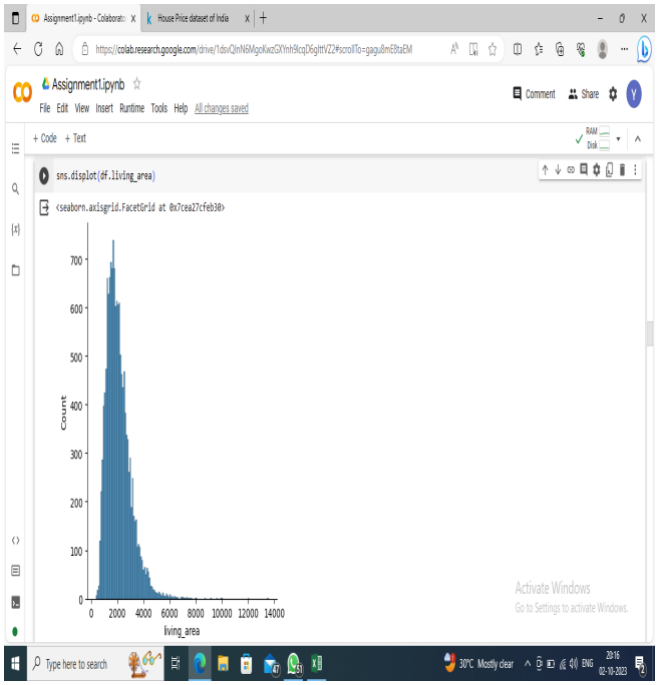
LOAD THE DATASET INTO GOOGLE COLAB :

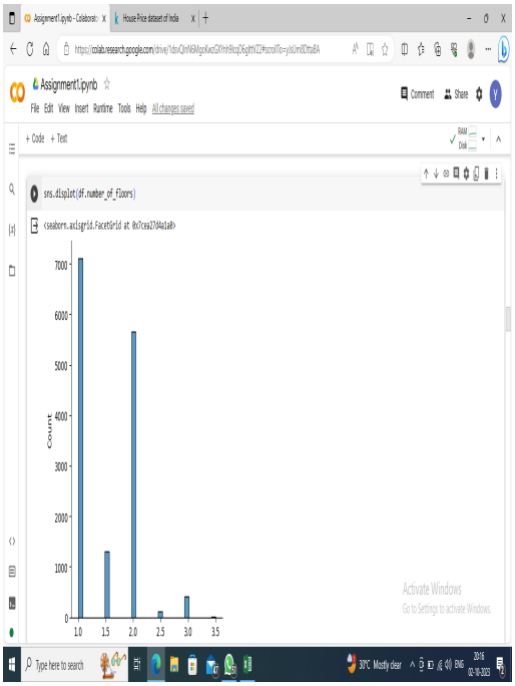


UNIVARIATE ANALYSIS :

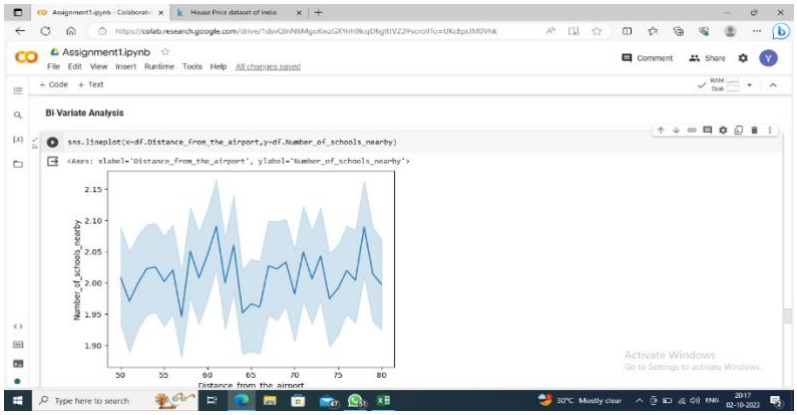


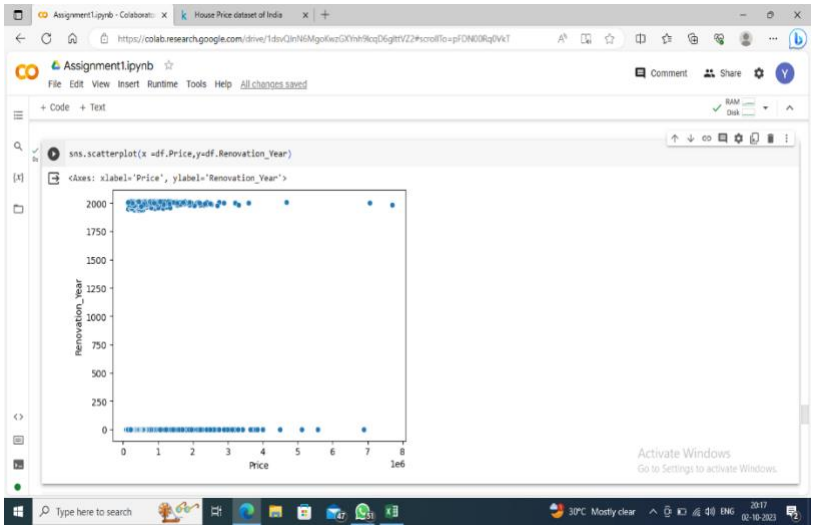


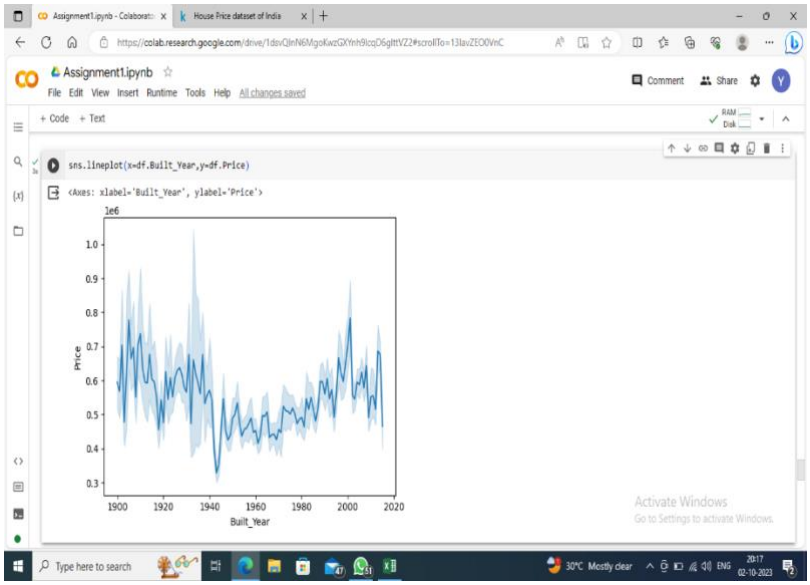




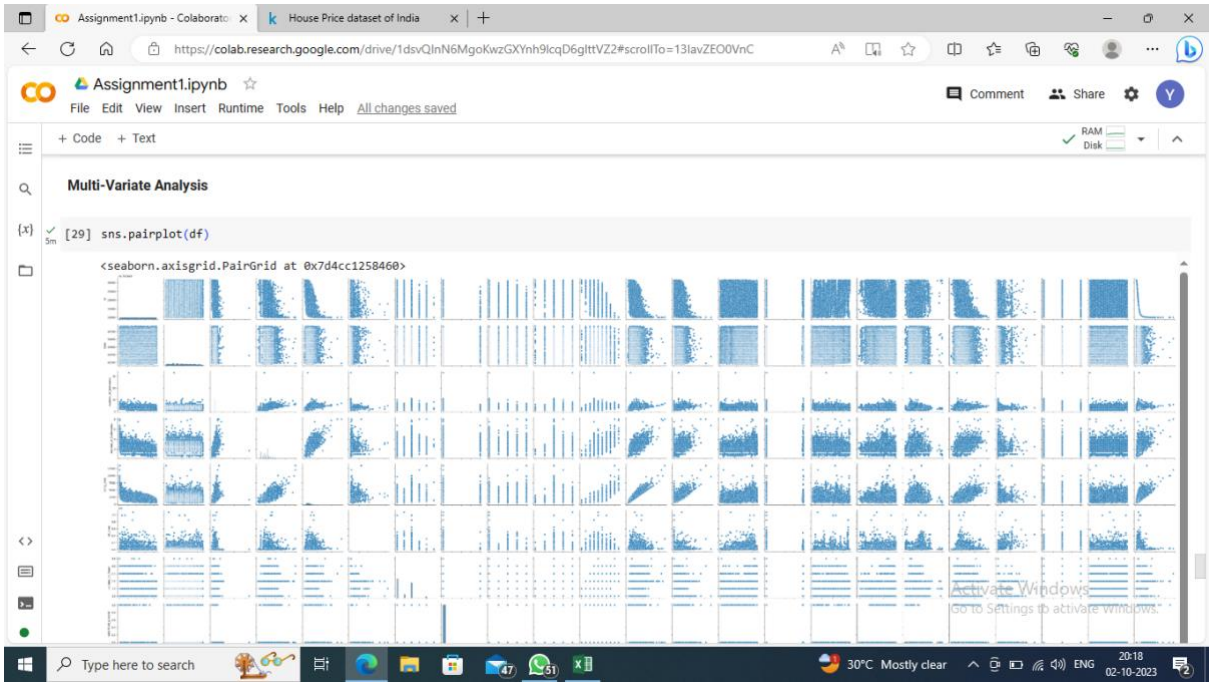
BI-VARIATE ANALYSIS :

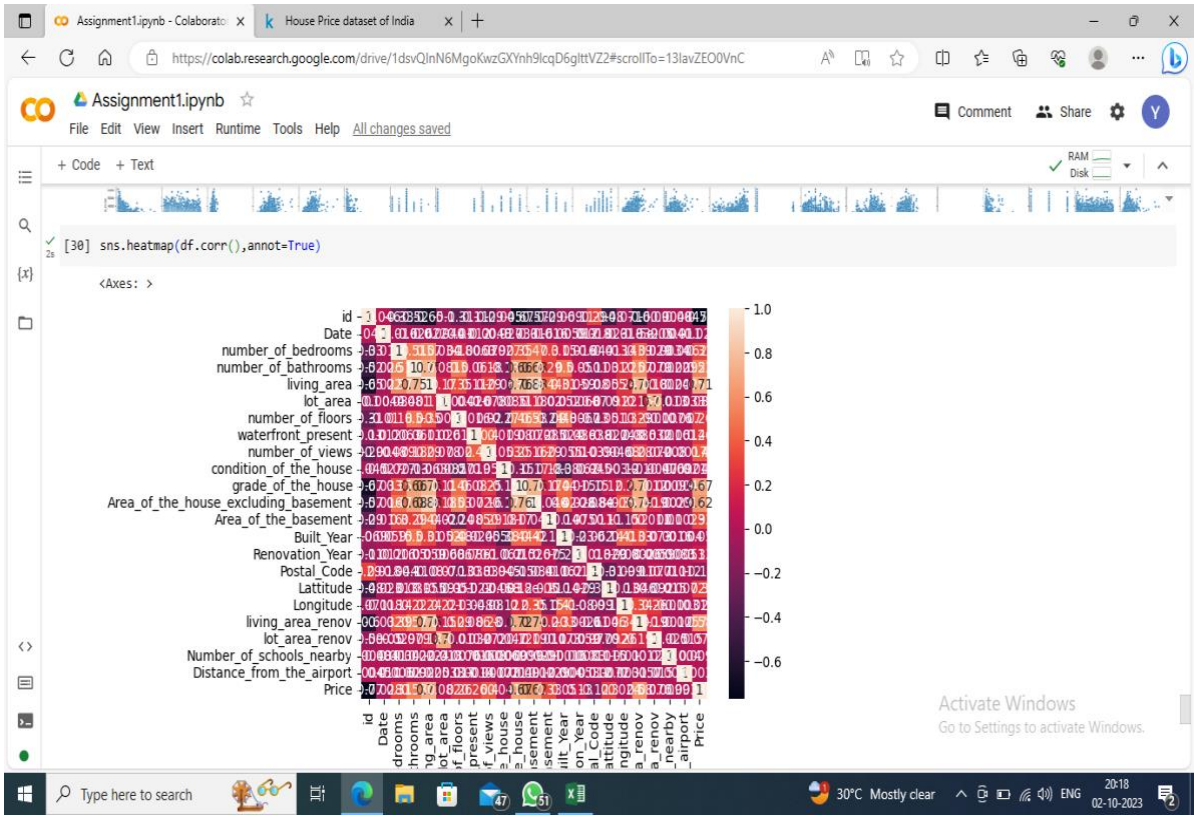




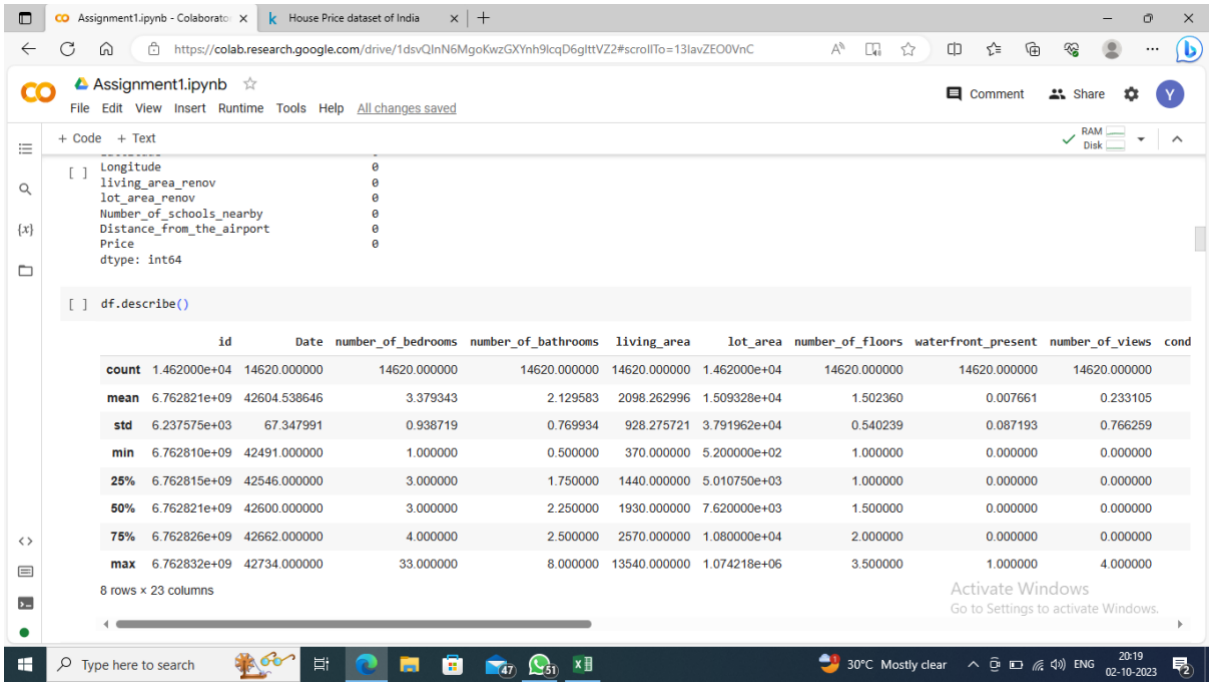


MULTI-VARIATE ANALYSIS:





PERFORM DESCRIPTIVE STATISTICS ON THE DATASET :



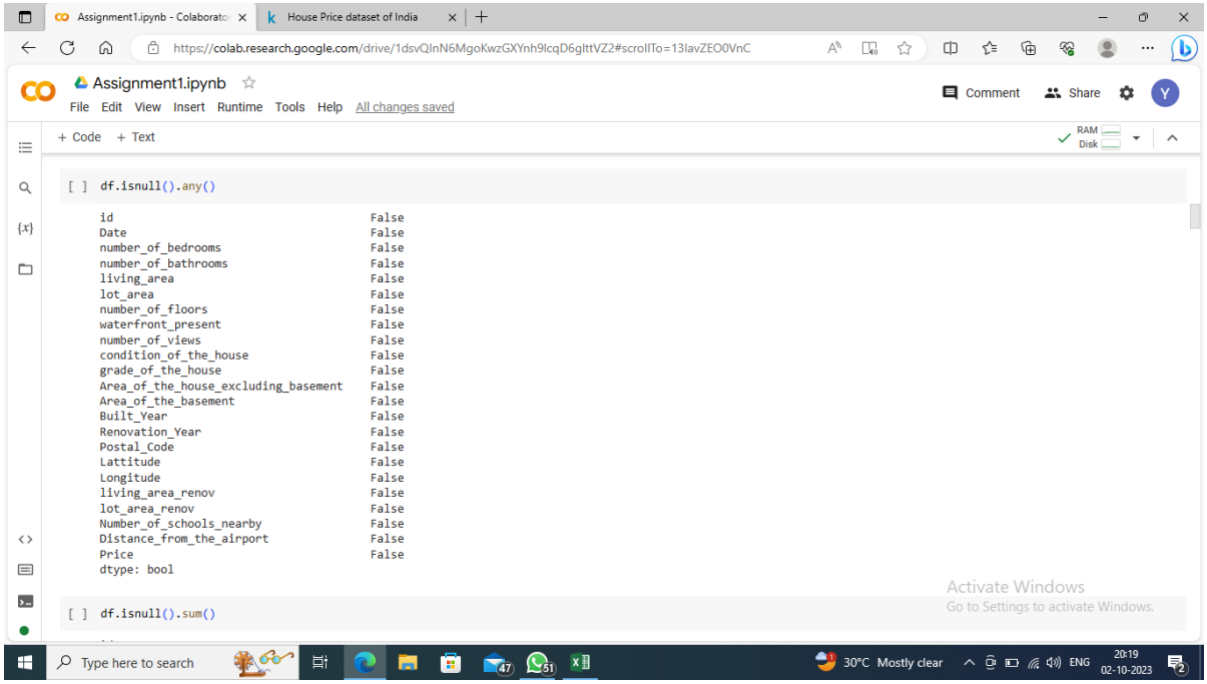
```
Longitude                      0
living_area_renov              0
lot_area_renov                 0
Number_of_schools_nearby       0
Distance_from_the_airport      0
Price                          0
dtype: int64
```

```
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+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 × 23 columns

HANDLE THE 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
```

```
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/1dsvQInN6MgoKwzGXYnh9lcqD6glttVZ2?usp=sharing