NAAN MUDHALVAN MANDATORY COURSE
AUGUST – DECEMBER 2023 (ODD SEMESTER)
DATA ANALYTICS WITH TABLEAU

Course Name:	Data Analytics With Tableau
Team Number:	**23**
Assignment Submitted To:	Anna University – Naan Mudhalvan
Year:	IV
Department:	Artificial Intelligence and Data Science
Semester	VII
Team Members:	1. Manikandan S 2. Manjusri T G 3. Rajesh R 4. Surya S
Guided By:	Ms.M.Noorunnisha – AP / CSE (Mamce - 8126)
SPOC Name:	Dr.K.Manikandan

ASSIGNMENT QUESTION

Perform the Below Tasks to complete the assignment:-

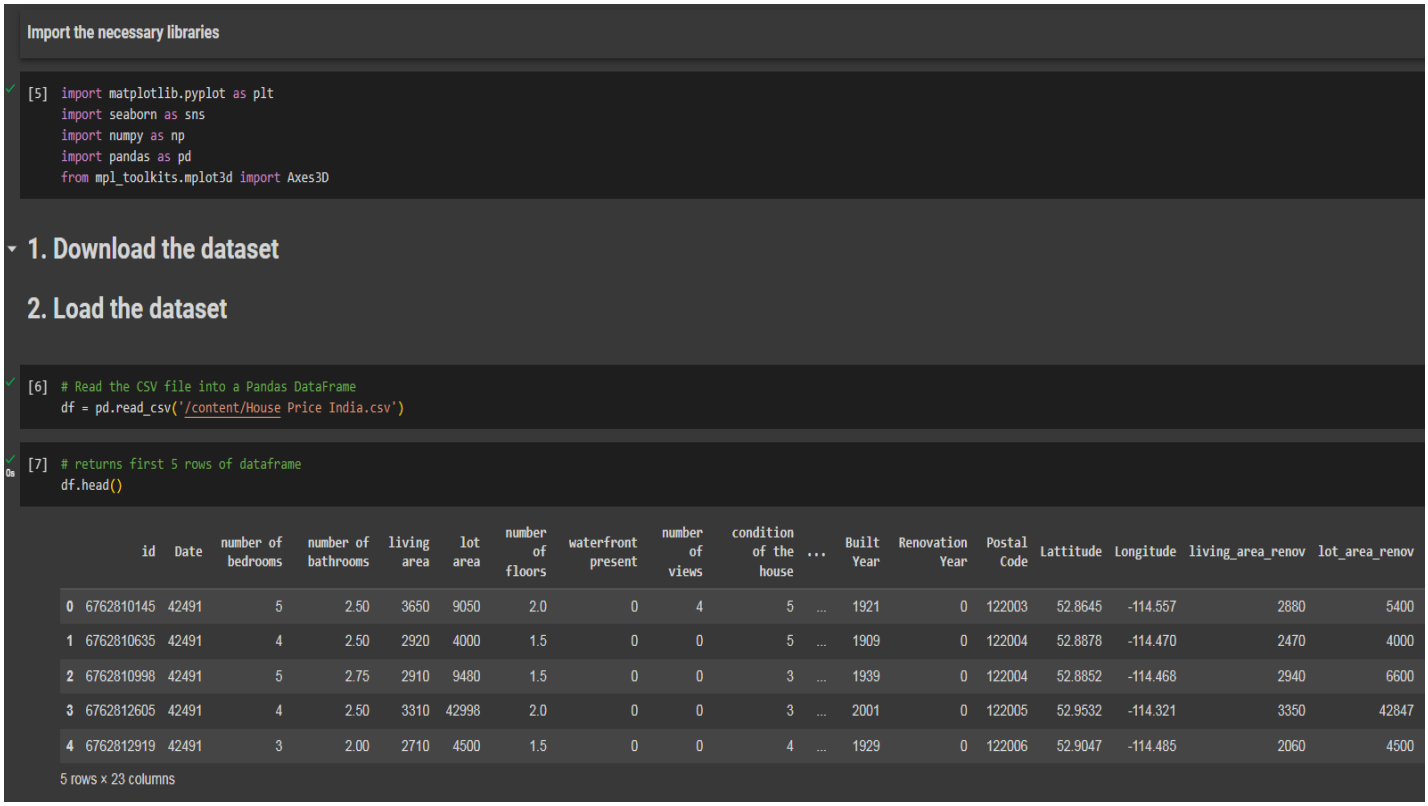
Tasks:-

1. Download the dataset: Dataset
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

Google Colab Link -

https://colab.research.google.com/drive/1xOI5LGPFBWqG6SIcJ1p5qwuUfxSD_8iF?usp=sharing



Import the necessary libraries

```
[5] import matplotlib.pyplot as plt
    import seaborn as sns
    import numpy as np
    import pandas as pd
    from mpl_toolkits.mplot3d import Axes3D
```

1. Download the dataset

2. Load the dataset

```
[6] # Read the CSV file into a Pandas DataFrame
    df = pd.read_csv('/content/House Price India.csv')
```

```
[7] # returns first 5 rows of dataframe
    df.head()
```

	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	Longitude	living_area_renov	lot_area_renov
0	6762810145	42491	5	2.50	3650	9050	2.0	0	4	5	...	1921	0	122003	52.8645	-114.557	2880	5400
1	6762810635	42491	4	2.50	2920	4000	1.5	0	0	5	...	1909	0	122004	52.8878	-114.470	2470	4000
2	6762810998	42491	5	2.75	2910	9480	1.5	0	0	3	...	1939	0	122004	52.8852	-114.468	2940	6600
3	6762812605	42491	4	2.50	3310	42998	2.0	0	0	3	...	2001	0	122005	52.9532	-114.321	3350	42847
4	6762812919	42491	3	2.00	2710	4500	1.5	0	0	4	...	1929	0	122006	52.9047	-114.485	2060	4500

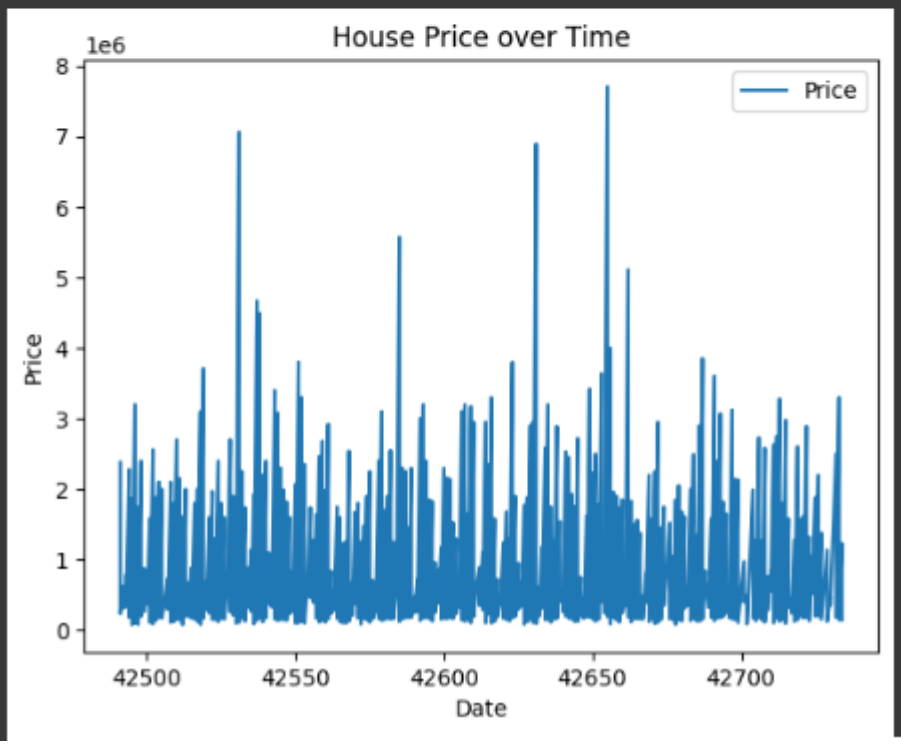
5 rows × 23 columns

3. Perform the Below Visualizations

1. Univariate Analysis

Line Plots / Bar Charts

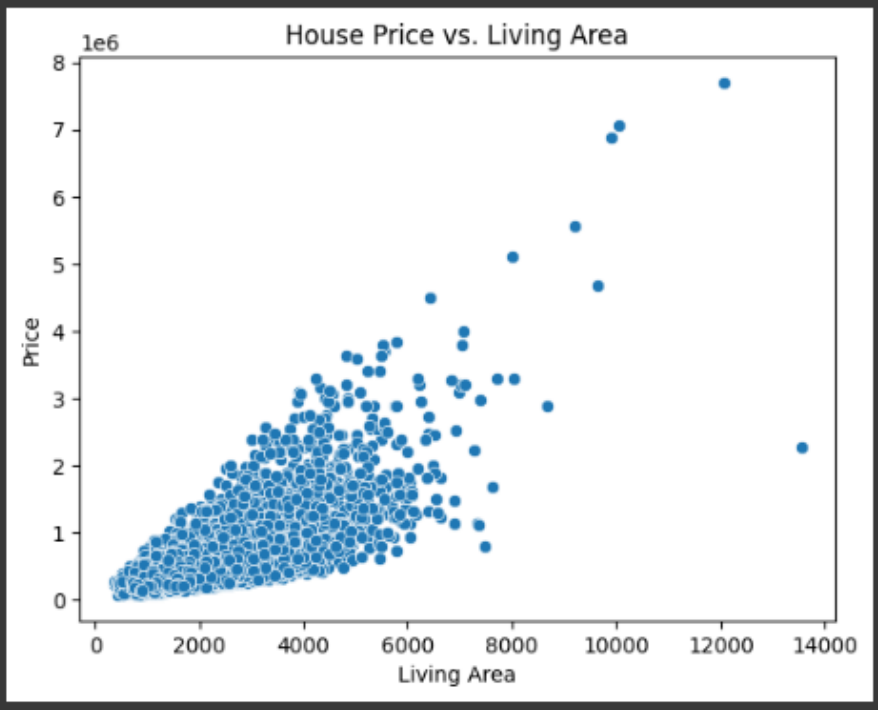
```
[10] # Create a line plot of the house price over time
     df.plot(x='Date', y='Price', kind='line')
     plt.title('House Price over Time')
     plt.xlabel('Date')
     plt.ylabel('Price')
     plt.show()

     # Create a bar chart of the number of bedrooms in each house
     df['number of bedrooms'].value_counts().plot(kind='bar')
     plt.title('Number of Bedrooms')
     plt.xlabel('Number of Bedrooms')
     plt.ylabel('Count')
     plt.show()
```



2. Bi-Variate Analysis

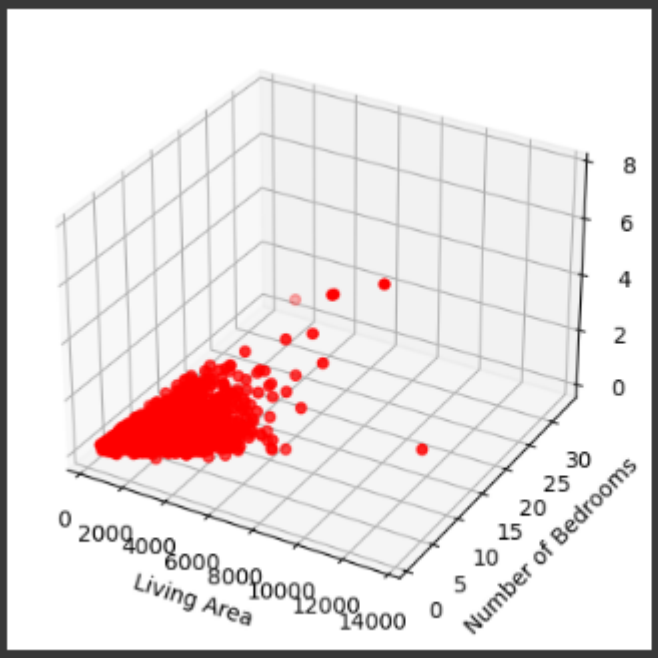
Scatterplot

```
# Create a scatterplot of the house price and living area
sns.scatterplot(x='living area', y='Price', data=df)
plt.title('House Price vs. Living Area')
plt.xlabel('Living Area')
plt.ylabel('Price')
plt.show()
```



3. Multi-Variate Analysis

```
# Create a 3D scatter plot of the features in the dataset
fig = plt.figure()
ax = fig.add_subplot(111, projection='3d')

ax.scatter(df['living area'], df['number of bedrooms'], df['Price'], c='red')

# Set the axis labels
ax.set_xlabel('Living Area')
ax.set_ylabel('Number of Bedrooms')
ax.set_zlabel('Price')

# Show the plot
plt.show()
```



4. Perform descriptive statistics on the dataset.

```
[20] # Print the descriptive statistics of the dataset
     df.describe()
```

	id	Date	number of bedrooms	number of bathrooms	living area	lot area	number of floors	waterfront present	
count	1.462000e+04	14620.000000	14620.000000	14620.000000	14620.000000	1.462000e+04	14620.000000	14620.000000	1
mean	6.762821e+09	42604.538646	3.379343	2.129583	2098.262996	1.509328e+04	1.502360	0.007661	
std	6.237575e+03	67.347991	0.938719	0.769934	928.275721	3.791962e+04	0.540239	0.087193	
min	6.762810e+09	42491.000000	1.000000	0.500000	370.000000	5.200000e+02	1.000000	0.000000	
25%	6.762815e+09	42546.000000	3.000000	1.750000	1440.000000	5.010750e+03	1.000000	0.000000	
50%	6.762821e+09	42600.000000	3.000000	2.250000	1930.000000	7.620000e+03	1.500000	0.000000	
75%	6.762826e+09	42662.000000	4.000000	2.500000	2570.000000	1.080000e+04	2.000000	0.000000	
max	6.762832e+09	42734.000000	33.000000	8.000000	13540.000000	1.074218e+06	3.500000	1.000000	

8 rows × 23 columns

5. Handle the Missing values

```
# Check for missing values
df.isnull().sum()
```

```
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
```

Conclusion

- There are no missing values in the dataset.
- Mean house price in India: ₹57.21 lakhs.
- Median house price in India: ₹48 lakhs.
- Factors affecting house price: Living area, number of bedrooms, and house grade.
- Larger living areas, more bedrooms, and higher grades lead to higher prices.