

# Assignment – 3

## House Price India Analysis

Name: Sathiya Rubha M

NM Id: au611220104134

```
Assignment_1.ipynb
File Edit View Insert Runtime Tools Help All changes saved
+ Code + Text
#Importing Necessary Python Libraries
[1] import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import LabelEncoder, StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score

#Uploading the House Price India Dataset
from google.colab import files
uploaded = files.upload()

Task_2: Load the dataset
[ ] df = pd.read_csv('House Price India.csv')

Explore the Dataset
[ ] print(df.head())
```

id Date number of bedrooms number of bathrooms living area \

Connected to Python 3 Google Compute Engine backend

```
Assignment_1.ipynb
File Edit View Insert Runtime Tools Help All changes saved
+ Code + Text
Explore the Dataset
print(df.head())
```

	id	Date	number of bedrooms	number of bathrooms	living area \
0	6762810145	42491	5	2.50	3650
1	6762810635	42491	4	2.50	2920
2	6762810998	42491	5	2.75	2910
3	6762812605	42491	4	2.50	3310
4	6762812919	42491	3	2.00	2710

	lot area	number of floors	waterfront present	number of views \
0	9050	2.0	0	4
1	4000	1.5	0	0
2	9480	1.5	0	0
3	42998	2.0	0	0
4	4500	1.5	0	0

	condition of the house ...	Built Year	Renovation Year	Postal Code \
0	5 ...	1921	0	122003
1	5 ...	1909	0	122004
2	3 ...	1939	0	122004
3	3 ...	2001	0	122005
4	4 ...	1929	0	122006

	Latitude	Longitude	living_area_renov	lot_area_renov \
0	52.8645	-114.557	2880	5400
1	52.8878	-114.470	4000	
2	52.8852	-114.468	2940	6600
3	52.9532	-114.321	3350	42847
4	52.9047	-114.485	2060	4500

	Number of schools nearby	Distance from the airport	Price
0	2	58	2360000
1	2	51	1400000
2	1	53	1200000
3	3	76	838000
4	1	51	805000

[5 rows x 23 columns]

Connected to Python 3 Google Compute Engine backend

Assignment\_1.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share Settings

+ Code + Text

RAM Disk

+ Code + Text

```

[5 rows x 23 columns]

print(df.info())

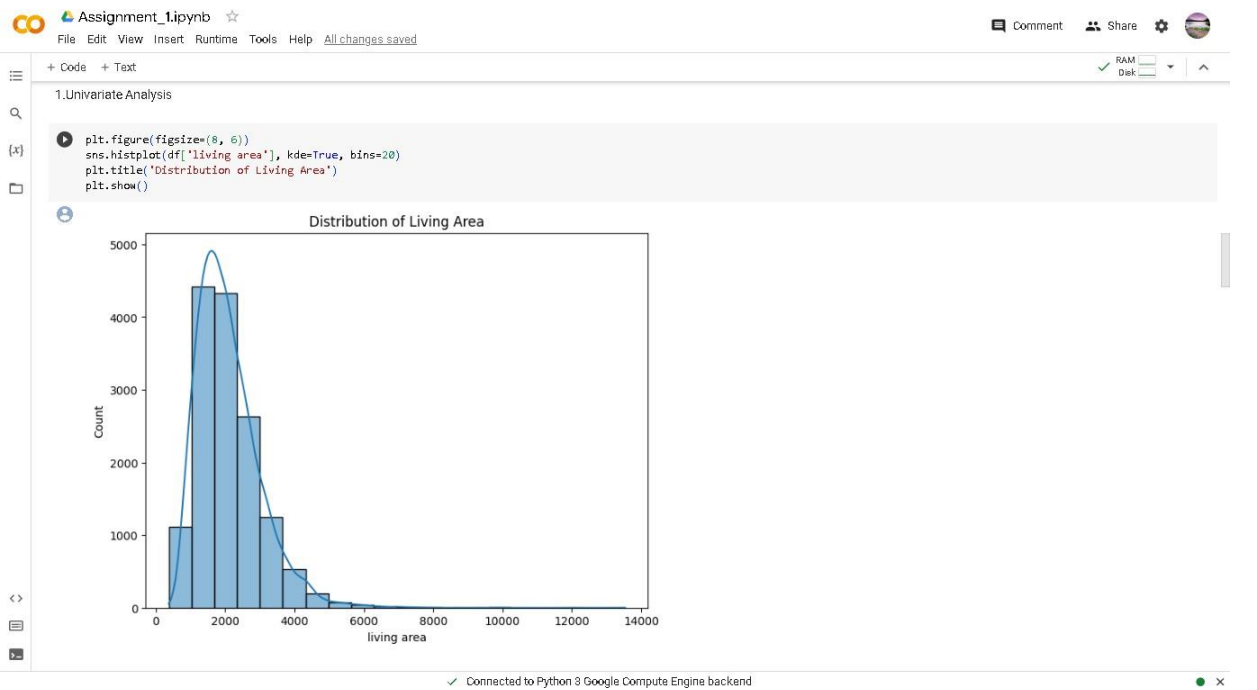
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 14620 entries, 0 to 14619
Data columns (total 23 columns):
#   Column              Non-Null Count  Dtype
---  -
0    id                  14620 non-null  int64
1    Date                14620 non-null  int64
2    number of bedrooms  14620 non-null  int64
3    number of bathrooms 14620 non-null  float64
4    living area         14620 non-null  int64
5    lot area            14620 non-null  int64
6    number of floors     14620 non-null  float64
7    waterfront present  14620 non-null  int64
8    number of views      14620 non-null  int64
9    condition of the house 14620 non-null  int64
10   grade of the house   14620 non-null  int64
11   Area of the house(excluding basement) 14620 non-null  int64
12   Area of the basement 14620 non-null  int64
13   Built Year           14620 non-null  int64
14   Renovation Year      14620 non-null  int64
15   Postal Code          14620 non-null  int64
16   Latitude             14620 non-null  float64
17   Longitude            14620 non-null  float64
18   living_area_renov    14620 non-null  int64
19   lot_area_renov       14620 non-null  int64
20   Number of schools nearby 14620 non-null  int64
21   Distance from the airport 14620 non-null  int64
22   Price                14620 non-null  int64
dtypes: float64(4), int64(19)
memory usage: 2.6 MB
None

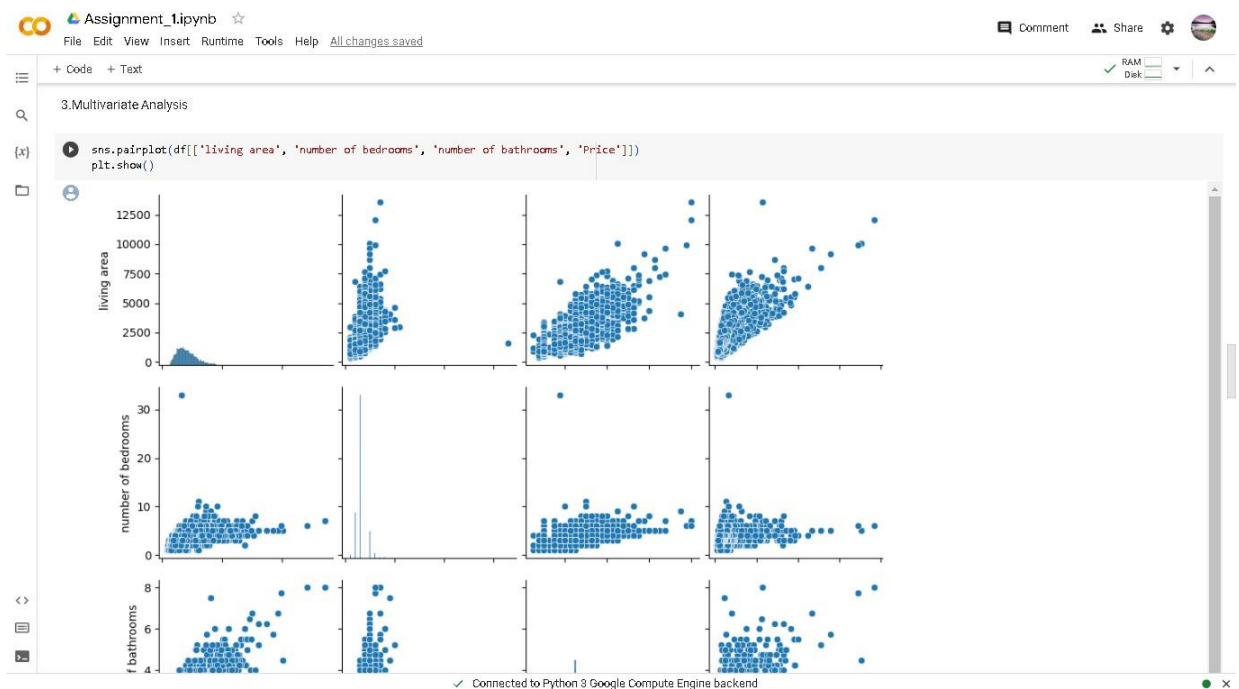
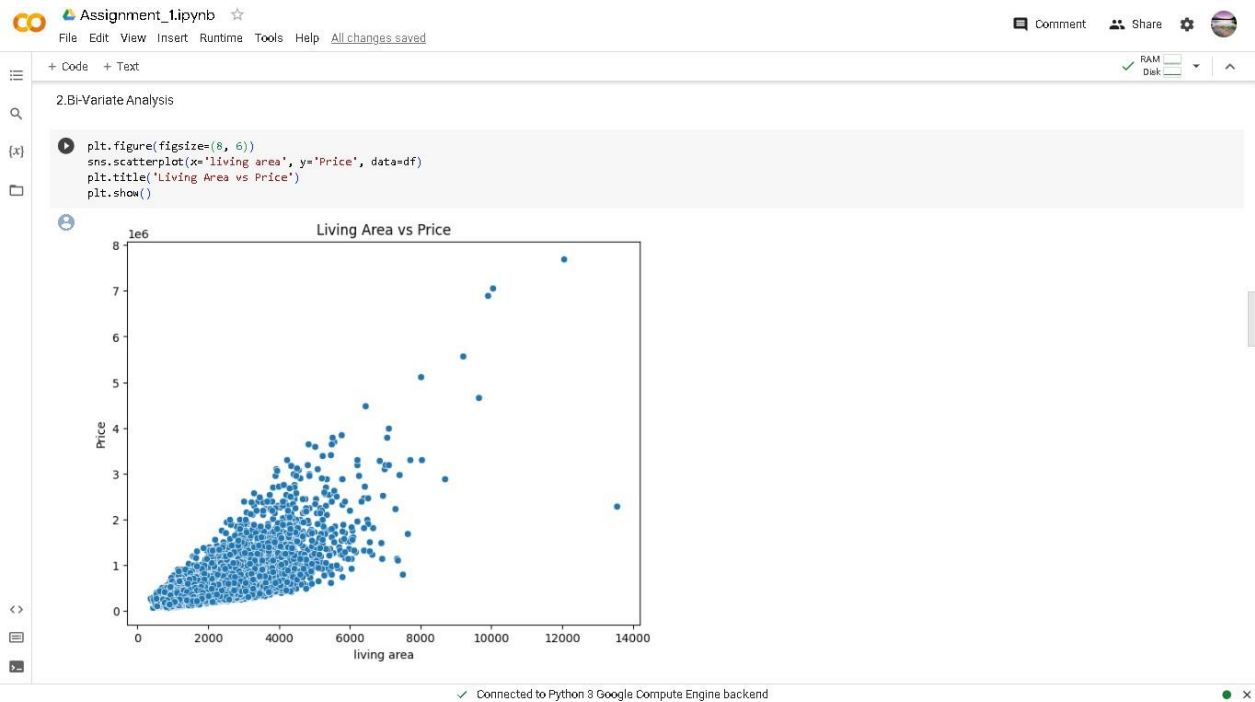
```

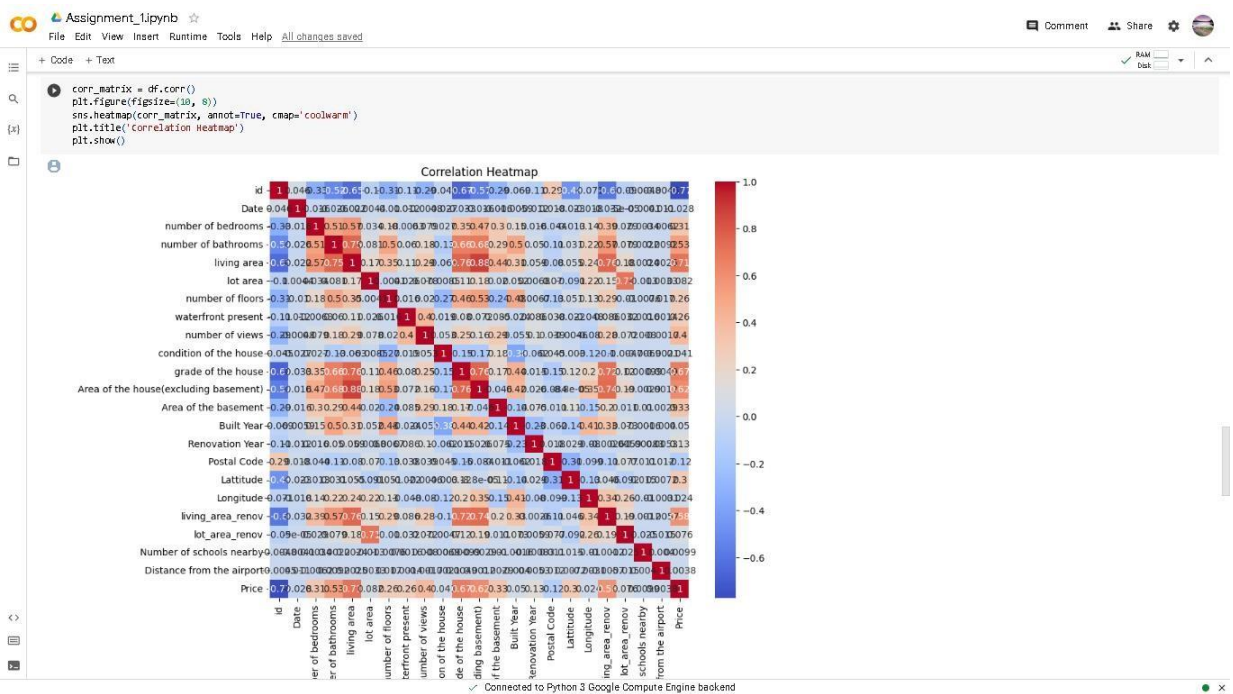
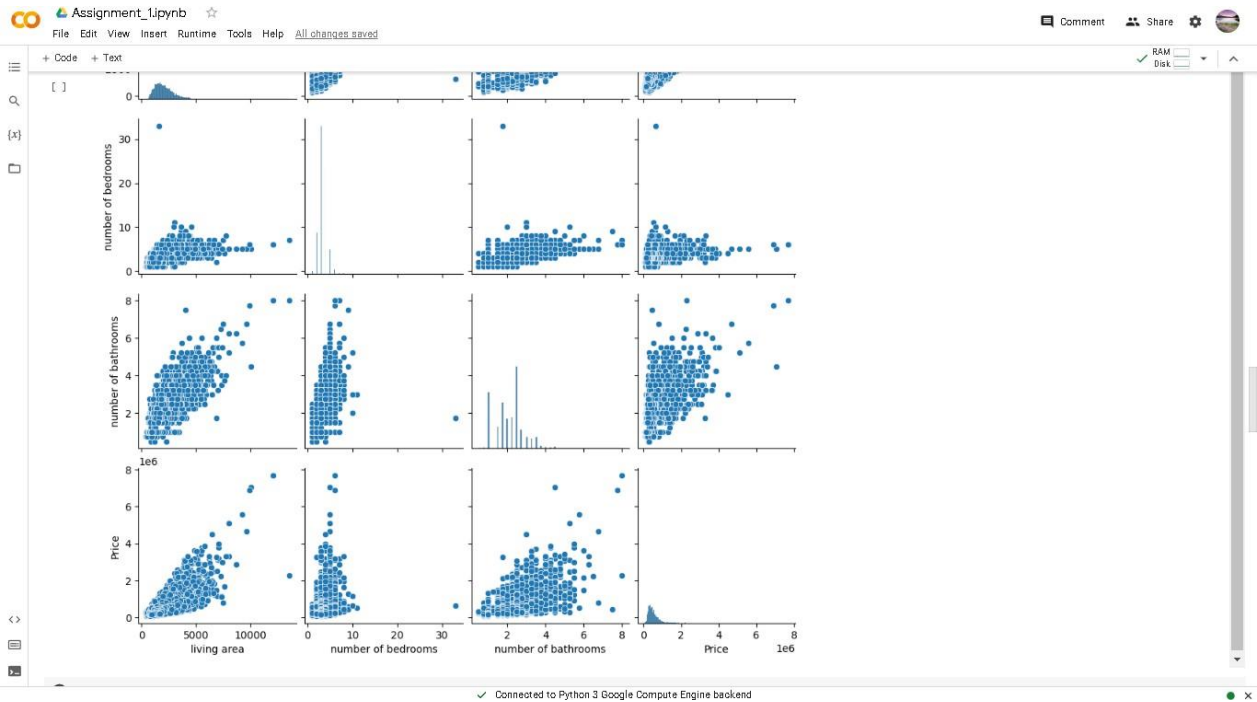
+ Code + Text

Task\_3 : Performing the below visualizatin

Connected to Python 3 Google Compute Engine backend







```
Assignment_1.ipynb
File Edit View Insert Runtime Tools Help All changes saved
+ Code + Text
Task_4 : Descriptive Statistics

# Descriptive statistics
print(df.describe())

count    1.462000e+04    14620.000000    14620.000000    14620.000000 \
mean     6.762821e+09    42604.538646     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    42600.000000     3.000000     2.250000
75%      6.762826e+09    42662.000000     4.000000     2.500000
max      6.762832e+09    42734.000000    33.000000     6.000000

count    14620.000000    1.462000e+04    14620.000000    14620.000000 \
mean     2098.262956    1.589320e+04     1.582360     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%     1530.000000     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

count    14620.000000    condition of the house ... Built Year \
mean     0.233105     3.430506 ... 1970.926402
std      0.766259     0.664151 ... 29.493625
min      0.000000     1.000000 ... 1900.000000
25%      0.000000     3.000000 ... 1951.000000
50%      0.000000     3.000000 ... 1975.000000
75%      0.000000     4.000000 ... 1997.000000
max      4.000000     5.000000 ... 2015.000000

Renovation Year    Postal Code    Latitude    Longitude \
count    14620.000000    14620.000000    14620.000000    14620.000000
mean      90.924008    122033.052244    52.752948    -114.404007
std      416.216661    19.082418     0.137522     0.141326
min        0.000000    122003.000000    52.385900    -114.709000
25%        0.000000    122017.000000    52.707600    -114.519000
50%        0.000000    122032.000000    52.905400    -114.421000
75%        0.000000    122048.000000    52.908900    -114.315000
max      2015.000000    122072.000000    53.007600    -113.505000

living_area_renov    lot_area_renov    Number of schools nearby \
count    14620.000000    14620.000000    14620.000000
mean     1996.702257    12753.500068     2.012244
dtype: object
```

```
Assignment_1.ipynb
File Edit View Insert Runtime Tools Help All changes saved
+ Code + Text
Task_5 : Handling the Missing Values

# Check for missing values
print(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
Latitude    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
```

Python file link:

<https://colab.research.google.com/drive/1hedicSny0goQesBOqmRQgxSARByorW2F?usp=sharing>