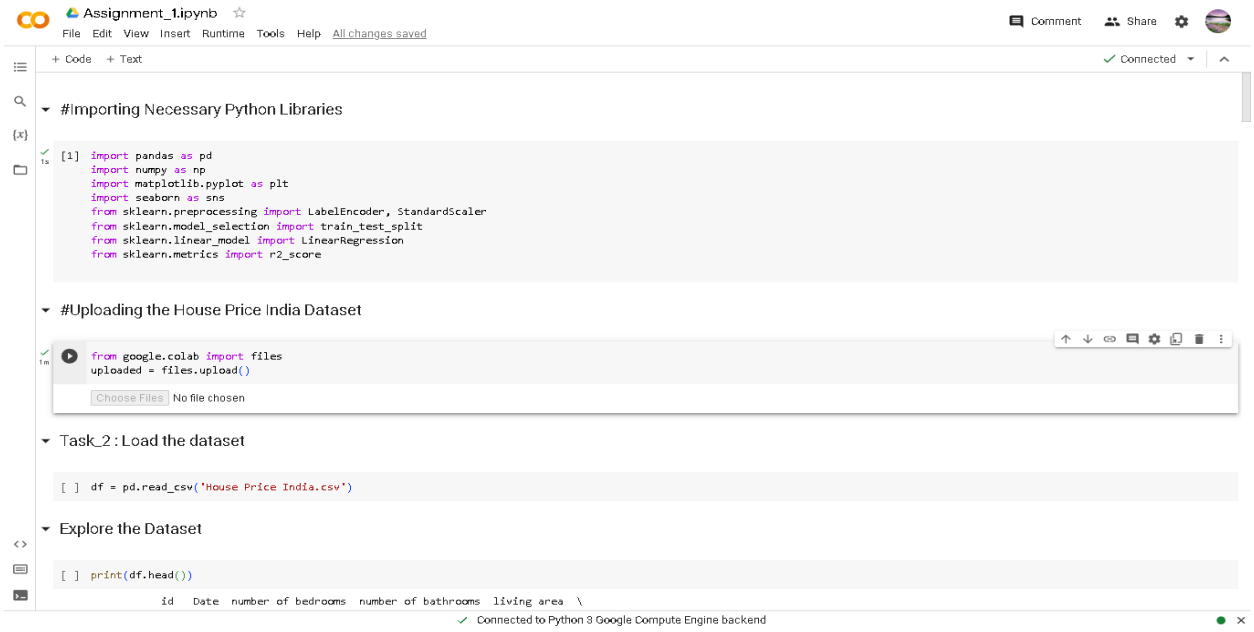
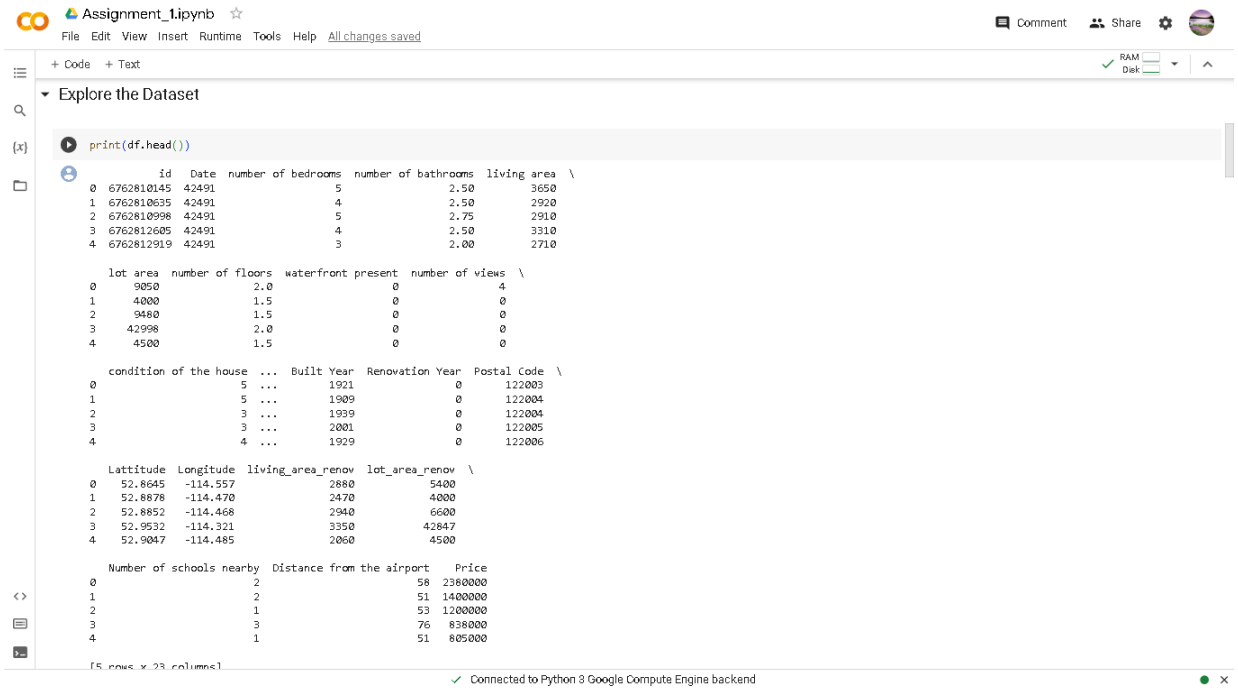
Assignment – 3

House Price India Analysis

Name: Sathiya Rubha M

NM Id: au611220104134





Assignment_1.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text

RAM Disk

```
[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  Lattitude                              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 visualizatins

Connected to Python 3 Google Compute Engine backend

Assignment_1.ipynb

File Edit View Insert Runtime Tools Help All changes saved

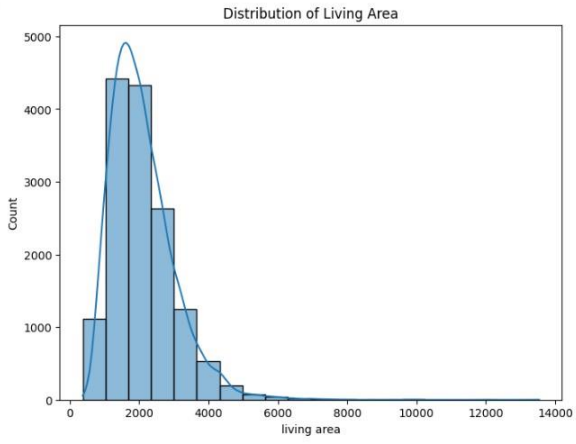
Comment Share

+ Code + Text

RAM Disk

1.Univariate Analysis

```
plt.figure(figsize=(8, 6))
sns.histplot(df['living area'], kde=True, bins=20)
plt.title('Distribution of Living Area')
plt.show()
```



Connected to Python 3 Google Compute Engine backend

Assignment_1.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text

RAM Disk

2.Bi-Variate Analysis

```
plt.figure(figsize=(8, 6))
sns.scatterplot(x='living area', y='Price', data=df)
plt.title('Living Area vs Price')
plt.show()
```

Connected to Python 3 Google Compute Engine backend

Assignment_1.ipynb

File Edit View Insert Runtime Tools Help All changes saved

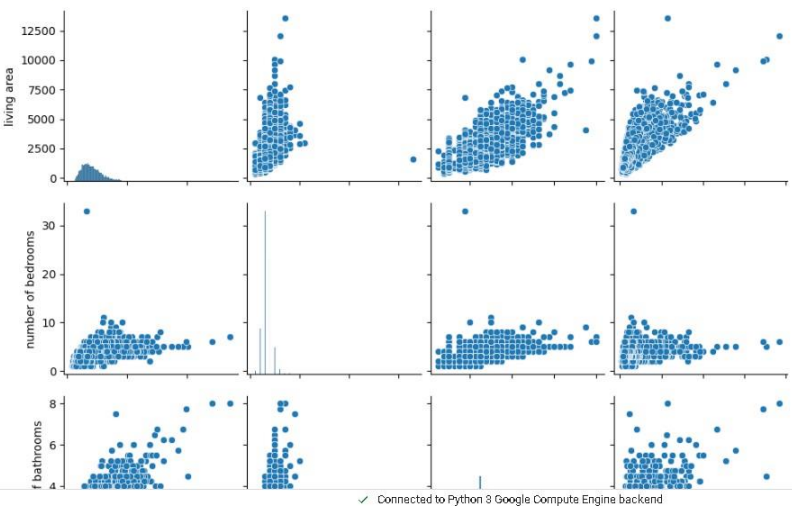
Comment Share

+ Code + Text

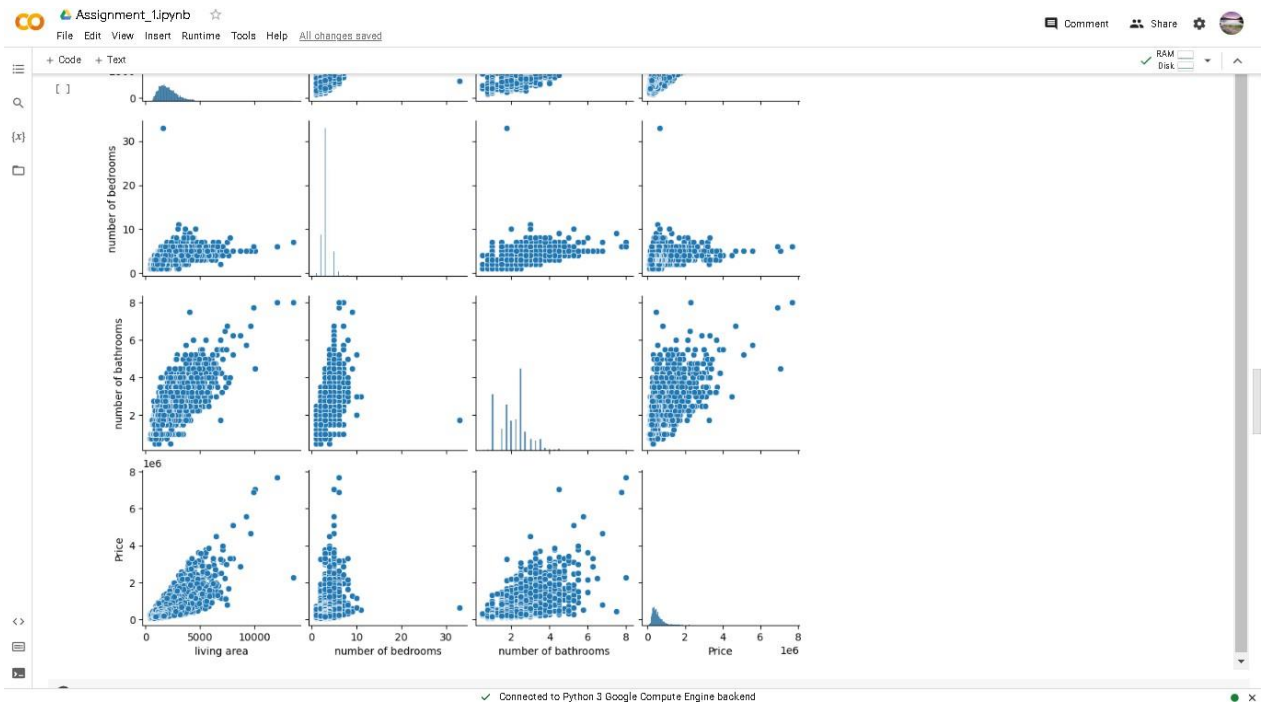
RAM Disk

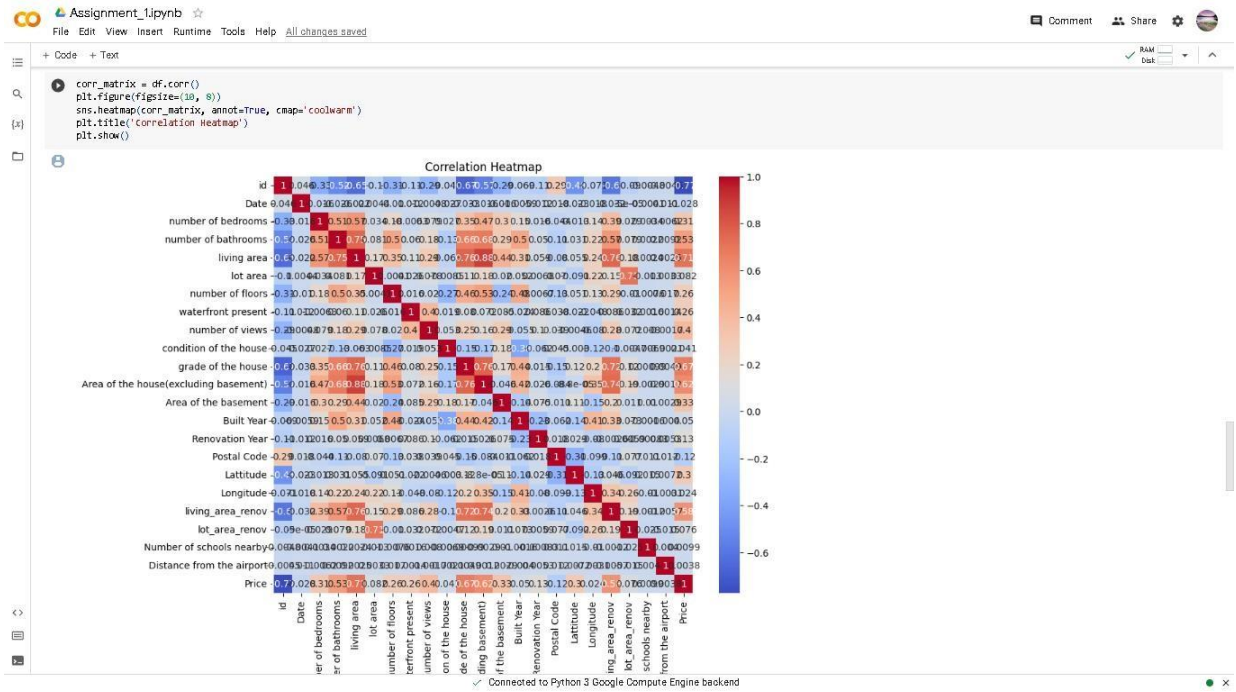
3.Multivariate Analysis

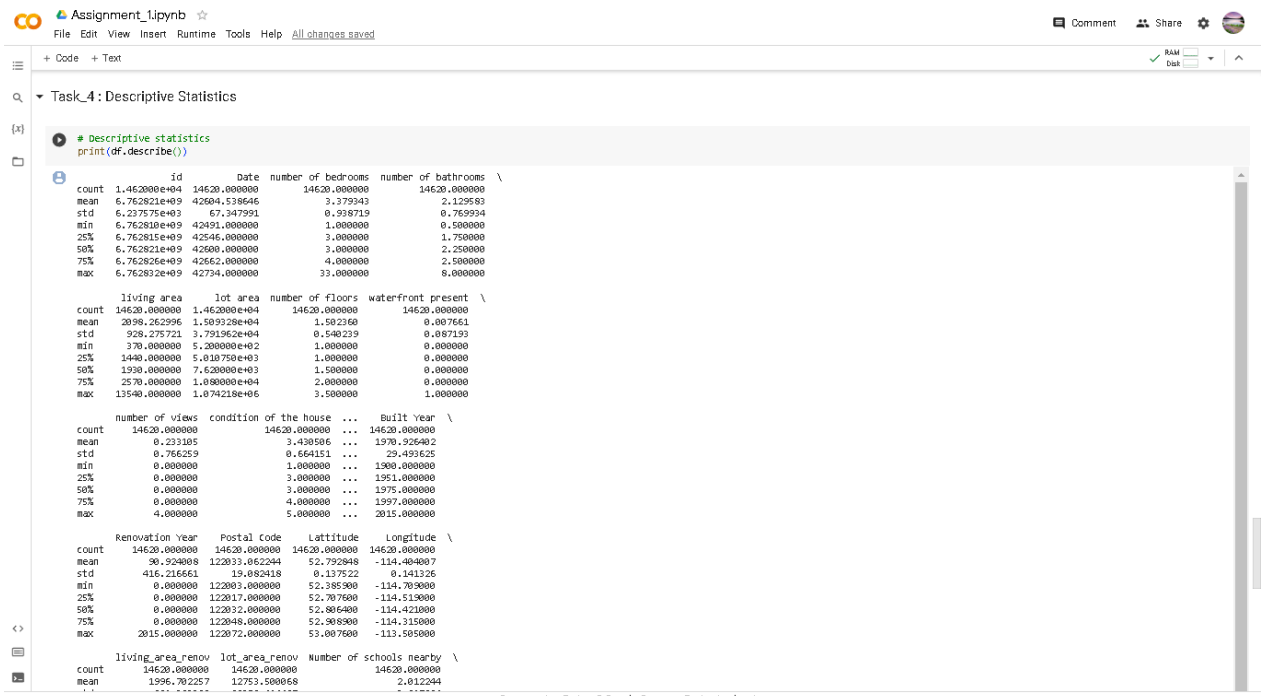
```
sns.pairplot(df[['living area', 'number of bedrooms', 'number of bathrooms', 'Price']])
plt.show()
```

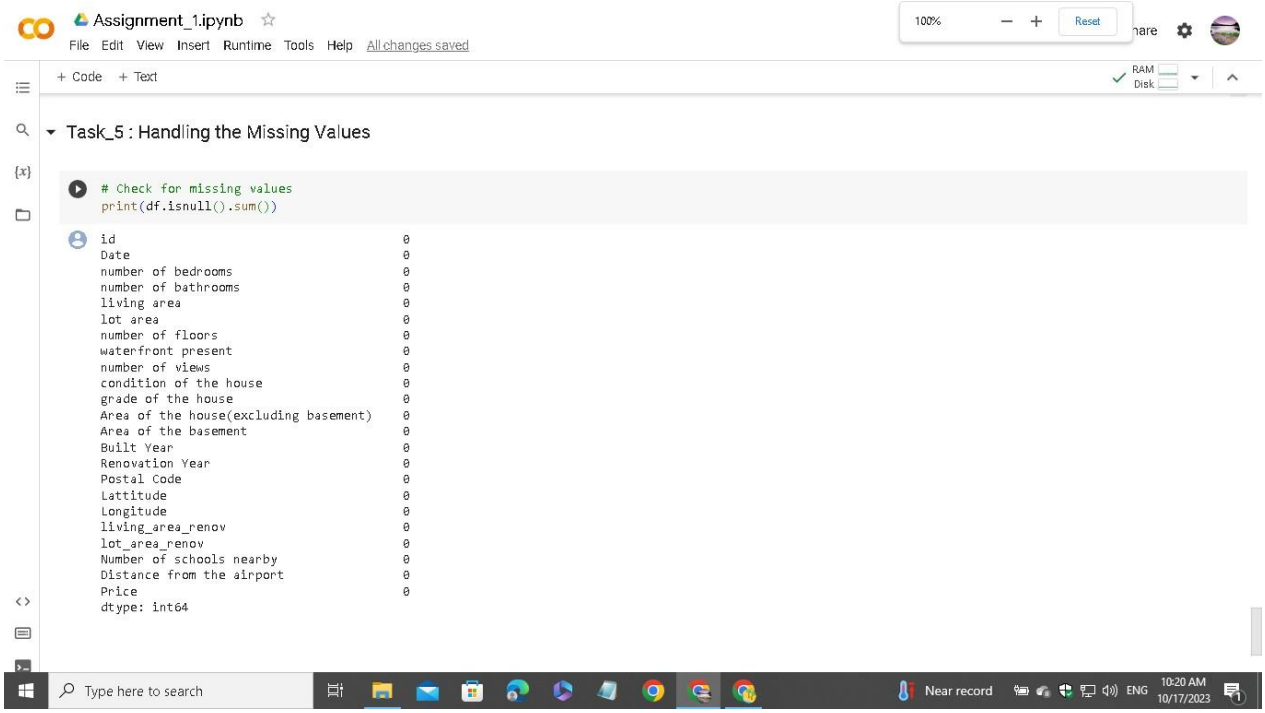


Connected to Python 3 Google Compute Engine backend


Assignment_1.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment Share
+ Code + Text
RAM Disk
number of bedrooms
number of bathrooms
Price
1e6
living area
number of bedrooms
number of bathrooms
Price
1e6
Connected to Python 3 Google Compute Engine backend


Assignment_1.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment Share
+ Code + Text
RAM Disk
corr_matrix = df.corr()
plt.figure(figsize=(10, 8))
sns.heatmap(corr_matrix, annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.show()
Correlation Heatmap
id
Date
number of bedrooms
number of bathrooms
living area
lot area
number of floors
waterfront present
number of views
condition of the house
grade of the house
Area of the house(excluding basement)
Area of the basement
Built Year
Renovation Year
Postal Code
Lattitude
Longitude
living_area_renov
lot_area_renov
Number of schools nearby
Distance from the airport
Price
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
Connected to Python 3 Google Compute Engine backend





Python file link:

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