AIML Assignment-3

Rollno: 1233

Name: Kirti Vichare

Mumbai house rent dataset (practice problem)

Dataset used:**Mumbai house rent**

click the link to get the datafile

https://drive.google.com/file/d/1JEmTt5uLGN4cwVYCgA2fzzH2Y3TYoVTz/view?usp=sharing

Dataset fields

1. Locality: Locality in Mumbai

2. Type : Type of Flat

3. Rent/Month : Rent per month in Rupees.

4. Build_up_area(sq.ft) : Build up area in sq.ft.

5. Furnishing : Type of Furnishing

6. Bathroom: Total Number of Bathrooms **7.**

   Balcony : Total Number of Balcony available.

8. Parking : Total Number of Parknig available. **9.**

Carpet_area(sq.ft) : Carpet area in sq.ft.

```
# import required libraries (numpy,pandas, matplotlib,
seaborn) import numpy as np import pandas as pd import
matplotlib.pyplot as plt import seaborn as sns
```

```
# upload the data file
Mumbai_House_Rent.csv from google.colab
import files files.upload()
              2390Mumbai…_Rent.csv
    2390Mumbai_House_Rent.csv(text/csv) - 1185989 bytes, last modified: 5/29/2022 - 100% done
    Saving 2390Mumbai_House_Rent.csv to 2390Mumbai_House_Rent.csv
    {'2390Mumbai_House_Rent.csv':
    b'Locality,Type,Rent/Month,Build_up_area(sq.ft),Furn # import the dataset
    and store in dataframe named
'df' file_name
='/content/2390Mumbai_House_Rent.csv'
```

```
df = pd.read_csv(file_name)  #
Display dataframe contents df
```

	Locality	Type	Rent/Month	Build_up_area(sq.ft)	Furnishing	Bathroom
0	Andheri	1 RK	20000	350 sq.ft	Semi Furnished	
1	Andheri	Apartment 3 BHK				
2	Andheri	Apartment	100000	1850 sq.ft	Semi Furnished	
3	Andheri	2 BHK				
4	Andheri	Apartment	45000	1240 sq.ft	Semi Furnished	
		1 BHK Apartment	31500	650 sq.ft	Unfurnished	
		2 BHK Apartment	44000	1250 sq.ft	Unfurnished	
...	...	...	...	...	...	.
15381		1 RK Apartment	25000	300 sq.ft	Fully Furnished	
15382	Worli	3 BHK	215000	3132 sq.ft	Semi Furnished	
	Worli	Apartment				

```
#display all columns in the dataframe/dataset
df.columns
```

```
Index(['Locality', 'Type', 'Rent/Month', 'Build_up_area(sq.ft)',
       'Furnishing', 'Bathrooms', 'Balcony', 'Parking',
       'Carpet_area(sq.ft)'], dtype='object')
```

```
#dispaly first 10 records
df.head(10)
```

		1 RK			Semi
0	Andheri	20000 350 sq.ft	1		
		Apartment			Furnished
		3 BHK			Semi
1	Andheri	100000	1850 sq.ft	3	
		Apartment			Furnished
2	BHK	Semi			
2	Andheri	45000 1240 sq.ft	2		
		Apartment			Furnished
1 BHK					
3	Andheri	31500 650 sq.ft	Unfurnished 2		

Apartment

2 BHK

4   Andheri 44000 1250 sq.ft Unfurnished   2

Apartment

3 BHK Fully

5   Andheri 65000 1050 sq.ft   3

                Apartment   Furnished

```
Number of rows in dataset 15386 Number of columns in
dataset 9
```

```
#Display localitywies total houses for rent # e.g. Goregaon 1085
#       Powai 1001
df.Locality.value_counts()
```

```
Goregaon       1085
Powai          1001
Chembur         985
Andheri         969
Bandra          958
Malad           800
Worli           738
Kurla           727
Jogeshwari      670
Bhandup         665
Vile Parle      655
Khar            572
Ghatkopar       571
Sion            550
Vikhroli        543
Dadar           521
Mulund          497
Parel           457
Prabhadevi      443
Juhu            438
Mahim           201
Govandi         169
Wadala          134
Lokhandwala     128
Tardeo          126
Byculla         101
Lalbaug         101
Malabar Hill     74
Girgaon          73
Marine Drive     66
Mahalakshmi      65
Santacruz        62
Grant Road       47
Vidyavihar       38
Matunga          35
Nariman Point    29
```

```
Masjid          27
Colaba          23
Fort            22
Dharavi         20
Name:        dtype int6
Locality,      :      4
```

```python
# Display top 5 entries of localitywies total houses for rent #1 #a =
df.Locality.unique()
#for i in
range(len(a)):
# print(df.sort_values(by=['Locality']))
# print(pd.DataFrame(df, 'Locality', index=[i]))
```

```
Masjid          27
Colaba          23
Fort            22
Dharavi         20
```

```
#print(df.groupby('Locality').head(5))
# print(df['Locality'][df['Locality']==a[i]].head(5))
df.Locality.value_counts().head(5)
```

```
Goregaon 1085
Powai    1001
Chembur   985
Andheri   969
Bandra    958
Name: Locality, dtype: int64
```

```
#display Juhu area info/records and store them new dataframe with name
'new_df' new_df=df[(df.Locality=="Juhu")]
new_df
```

	Locality	Type	Rent/Month	Build_up_area(sq.ft)	Furnishing	Bathrooms
		3 BHK Apartment	125000	1300 sq.ft	Fully Furnished	3
		2 BHK Apartment	80000	910 sq.ft	Fully Furnished	2
		3 BHK Apartment	190000	1750 sq.ft	Unfurnished	3
		3 BHK Apartment	90000	1020 sq.ft	Semi Furnished	3
6883	Juhu	2 BHK Apartment	80000	920 sq.ft	Fully Furnished	2
...	...	...	...	...	...	...
7312	Juhu	2 BHK Apartment	85000	952 sq.ft	Fully Furnished	2
7313	Juhu	2 BHK Apartment	85000	985 sq.ft	Fully Furnished	2

```
#display all records/flats having more than rs. 300000 rent per month in juhu
area.
Store new_df=df
new_df[(new_df['Locality']=="Juhu") & (new_df['Rent/Month']>30000)]
```

	Locality	Type	Rent/Month	Build_up_area(sq.ft)	Furnishing	Bathrooms
6882	Juhu	3 BHK Apartment	90000	1020 sq.ft	Semi Furnished	3
6883	Juhu	2 BHK Apartment	80000	920 sq.ft	Fully Furnished	2

	...	...	...	...		...	...	...
7312	Juhu	85000	952 sq.ft	2 BHK Apartment	2			Fully Furnished
7313	Juhu	85000	985 sq.ft	2 BHK Apartment	2			Fully Furnished

```
Build_up_area(sq.ft) 0
Furnishing             0
```

```
Build_up_area(sq.ft) 0
Furnishing             0
```

```
       Bathrooms              0
       Balcony                0
       Parking               35
       Carpet_area(sq.ft)     0
       dtype: int64
#display various areas (sq. ft) of flats available
df['Carpet_area(sq.ft)'].unique()
```

```
          '249 sq.ft', '327 sq.ft', '328 sq.ft', '1530 sq.ft', '638 sq.ft',
          '821 sq.ft', '511 sq.ft', '935 sq.ft', '863 sq.ft', '335 sq.ft',
          '776 sq.ft', '897 sq.ft', '446 sq.ft', '474 sq.ft', '684 sq.ft',
          '1325 sq.ft', '718 sq.ft', '279 sq.ft', '891 sq.ft', '265 sq.ft',
          '402 sq.ft', '426 sq.ft', '345 sq.ft', '2500 sq.ft', '1480 sq.ft',
          '2400 sq.ft', '1920 sq.ft', '1395 sq.ft', '2100 sq.ft',
          '1375 sq.ft', '1900 sq.ft', '1075 sq.ft', '1475 sq.ft',
          '1383 sq.ft', '981 sq.ft', '1301 sq.ft', '1865 sq.ft', '845 sq.ft',
          '2300 sq.ft', '1860 sq.ft', '1115 sq.ft', '1672 sq.ft',
          '1290 sq.ft', '1016 sq.ft', '724 sq.ft', '1320 sq.ft',
          '1109 sq.ft', '774 sq.ft', '1365 sq.ft', '553 sq.ft', '1156 sq.ft',
          '1520 sq.ft', '757 sq.ft', '619 sq.ft', '926 sq.ft', '512 sq.ft',
          '427 sq.ft', '362 sq.ft', '1006 sq.ft', '384 sq.ft', '568 sq.ft',
          '972 sq.ft', '1086 sq.ft', '487 sq.ft', '657 sq.ft', '727 sq.ft',
          '472 sq.ft', '418 sq.ft', '605 sq.ft', '695 sq.ft', '617 sq.ft',
          '677 sq.ft', '1440 sq.ft', '383 sq.ft', '210 sq.ft', '832 sq.ft',
          '834 sq.ft', '805 sq.ft', '1418 sq.ft', '954 sq.ft', '1336 sq.ft',
          '1208 sq.ft', '1750 sq.ft', '2675 sq.ft', '1132 sq.ft',
          '1295 sq.ft', '267 sq.ft', '767 sq.ft', '628 sq.ft', '2250 sq.ft',
          '1061 sq.ft', '1525 sq.ft', '618 sq.ft', '833 sq.ft', '808 sq.ft',
          '469 sq.ft', '719 sq.ft', '582 sq.ft', '594 sq.ft', '417 sq.ft',
          '674 sq.ft', '1245 sq.ft', '1413 sq.ft', '1540 sq.ft', '666 sq.ft',
          '1193 sq.ft', '1154 sq.ft', '496 sq.ft', '706 sq.ft', '1233 sq.ft',
          '569 sq.ft', '363 sq.ft', '306 sq.ft', '457 sq.ft', '1030 sq.ft',
          '45 sq.ft', '973 sq.ft', '145 sq.ft', '1024 sq.ft', '851 sq.ft',
          '1153 sq.ft', '963 sq.ft', '647 sq.ft', '662 sq.ft', '663 sq.ft',
          '781 sq.ft', '654 sq.ft', '1056 sq.ft', '964 sq.ft', '683 sq.ft',
          '671 sq.ft', '743 sq.ft', '778 sq.ft', '793 sq.ft', '656 sq.ft',
          '1013 sq.ft', '1253 sq.ft', '801 sq.ft', '634 sq.ft', '1035 sq.ft',
          '636 sq.ft', '1096 sq.ft', '399 sq.ft', '1376 sq.ft', '1795 sq.ft',
          '1460 sq.ft', '857 sq.ft', '1206 sq.ft', '2090 sq.ft',
          '1850 sq.ft', '1380 sq.ft', '195 sq.ft', '977 sq.ft', '538 sq.ft',
          '1875 sq.ft', '637 sq.ft', '477 sq.ft', '749 sq.ft', '679 sq.ft',
          '1403 sq.ft', '952 sq.ft', '1145 sq.ft', '644 sq.ft', '1756 sq.ft',
          '1410 sq.ft', '479 sq.ft', '738 sq.ft', '492 sq.ft', '439 sq.ft',
          '1239 sq.ft', '1590 sq.ft', '1430 sq.ft', '379 sq.ft', '888 sq.ft',
          '1321 sq.ft', '1770 sq.ft', '1563 sq.ft', '1680 sq.ft',
          '1840 sq.ft', '1675 sq.ft', '371 sq.ft', '358 sq.ft', '181 sq.ft',
          '956 sq.ft', '868 sq.ft', '936 sq.ft', '571 sq.ft', '1089 sq.ft',
          '938 sq.ft', '1331 sq.ft', '957 sq.ft', '934 sq.ft', '459 sq.ft',
          '483 sq.ft', '969 sq.ft', '1311 sq.ft', '1513 sq.ft', '576 sq.ft',
          '1493 sq.ft', '1510 sq.ft', '1303 sq.ft', '1456 sq.ft',
          '611 sq.ft', '324 sq.ft', '498 sq.ft', '1370 sq.ft', '1788 sq.ft',
          '301 sq.ft', '794 sq.ft', '646 sq.ft', '527 sq.ft', '549 sq.ft',
          '1092 sq.ft', '473 sq.ft', '493 sq.ft', '859 sq.ft', '1046 sq.ft',
          '1276 sq.ft', '661 sq.ft', '764 sq.ft', '883 sq.ft', '649 sq.ft',
          '282 sq.ft', '256 sq.ft', '528 sq.ft', '659 sq.ft', '843 sq.ft',
          '229 sq.ft', '1397 sq.ft', '316 sq.ft', '1065 sq.ft', '261 sq.ft',
```

```
'632 sq.ft', '1263 sq.ft', '1001 sq.ft', '708 sq.ft', '1201 sq.ft',
'877 sq.ft', '424 sq.ft', '534 sq.ft', '1655 sq.ft', '1476 sq.ft',
'1611 sq.ft', '1227 sq.ft', '1645 sq.ft', '652 sq.ft', '509 sq.ft',
'2070 sq.ft', '1960 sq.ft', '1348 sq.ft', '1347 sq.ft',
'1538 sq.ft', '1304 sq.ft', '4000 sq.ft', '2170 sq.ft',
'1663 sq.ft', '1262 sq.ft', '1029 sq.ft', '1373 sq.ft',
'1312 sq.ft', '839 sq.ft', '974 sq.ft', '2080 sq.ft', '1298 sq.ft',
'1910 sq.ft', '245 sq.ft', '2180 sq.ft', '2980 sq.ft',
'1385 sq.ft', '1415 sq.ft', '286 sq.ft', '313 sq.ft', '321 sq.ft',
'2800 sq.ft'], dtype=object)
```

```
#Display number of bathrooms in the dataset
df.Bathrooms.unique()
array(['1', '3', '2', 'Missing', '4', '5', '6', '7'], dtype=object)
```

```
#display dataframe new_df data
new_df=df
new_df
```

	Locality	Type	Rent/Month	Build_up_area(sq.ft)	Furnishing	Bathrooms
0	Andheri	1 RK Apartment	20000	350 sq.ft	Semi Furnished	1
1	Andheri	3 BHK Apartment	100000	1850 sq.ft Furnished	Semi	3
2	Andheri	2 BHK Apartment	45000	1240 sq.ft	Semi Furnished	2
3	Andheri	1 BHK Apartment	31500	650 sq.ft	Unfurnished	2
4	Andheri	2 BHK Apartment	44000	1250 sq.ft	Unfurnished	2
...	...	...	...	...	...	...
15381	Worli	1 RK Apartment	25000	300 sq.ft	Fully Furnished	1
15382	Worli	3 BHK Apartment	215000	3132 sq.ft	Semi Furnished	3

```
#delete rows from the dataframe 'new_df' dataframe which contains
empty cells # Deleting the rows which have Empty cells
#3 new_df =
df.dropna() new_df
```

...	...	...	...	...	...	...
15381	Worli	1 RK Apartment	25000	300 sq.ft	Fully Furnished	1
		3 BHK			Semi	

	...	...	...
15381	1 RK Apartment	25000
15383	2 BHK Apartment	100000
15384	3 BHK Apartment	240000
15385	3 BHK Apartment	250000

15351 rows × 2 columns

```
print(test)
```

```
print(test)
```

```
          Localit              Type Rent/Month Build_up_area(sq.f \
0              y 1 RK Apartmen 20000        t)
        Andheri                 t                  350 sq.ft
1         Andheri 3 BHK Apartmen   100000      1850 sq.ft t
2         Andheri 2 BHK Apartmen    45000      1240 sq.ft t
3         Andheri 1 BHK Apartmen    31500       650 sq.ft
                                 t
4         Andheri 2 BHK Apartmen 44000 1250 sq.ft ... ... t ... ...
                           ...
15379 Worli 1 BHK Apartmen 22000      400 sq.ft t
15382   Worli 3 BHK Apartmen    215000      3132 sq.ft t
15383   Worli 2 BHK Apartmen    100000      1180 sq.ft t
15384   Worli 3 BHK Apartmen    240000      3400 sq.ft t
15385   Worli 3 BHK Apartmen    250000      3200 sq.ft
                                 t


           Furnishing Bathrooms Balcony Parking
                             Carpet_area(sq.ft)
0      Semi Furnished     1 Missing  0.0   250 sq.ft
1      Semi Furnished     3    1     2.0   1400 sq.ft
2      Semi Furnished     2    4     1.0   850 sq.ft
3      Unfurnished 2 Missing 1.0 425 sq.ft 4 Unfurnished 2 3 1.0 800 sq.ft
        ... ... ... ... ... ...
   15379 Semi Furnished      1 Missing     1.0       300 sq.ft
15382  Semi Furnished     3    2     3.0   765 sq.ft
15383  Fully Furnished    2    3     1.0   920 sq.ft
15384  Semi Furnished     4    2     2.0   2100 sq.ft
15385  Fully Furnished    4    2     2.0   1900 sq.ft


   [1220 rows x columns]
4      ------------------------- 9
     Locality           Type Rent/Month               \
   8  3 BHK Build_up_area(sq.ft) Apartment 100000
       Andher               2200 sq.ft
           i

   9 Andheri 3 BHK Apartment 55000     1350 sq.f       t     11
        Andheri 3 BHK Apartment    72000      987 sq.f   t

      14    Andheri 2 BHK Apartment    42000      1000 sq.f      t
  21    Andher 3 BHK Apartmen    75000          1482 sq.ft

 ...        i             t       ...                ...
   15381 Worli 1 RK Apartment    25000         300 sq.ft

           Furnishing Bathrooms Balcony Parking
                           Carpet_area(sq.ft)
8        Fully Furnished  3    4    2.0   1600 sq.ft
9        Unfurnished      2    2    1.0   1064 sq.ft
  11    Semi Furnished    2    2    2.0        880 sq.ft
 14    Fully Furnished 2 Missing  1.0   Missing 21 Fully Furnished 3    4
                         1.0   1100 sq.ft
 ...        ...      ...     ...    ...              ...
15353 Semi Furnished 3 2 2.0 1200 sq.ft 15364 Semi Furnished 1 Missing 1.0
400 sq.ft
15377 Fully Furnished     1 Missing     1.0          Missin
                                                     g
```

```
15380 Fully Furnished        3        1        0.0        1200
                                                          sq.ft
15381 Fully Furnished        1 Missing        0.0         Missin
                                                          g

[3182 rows 9 columns]  x
```
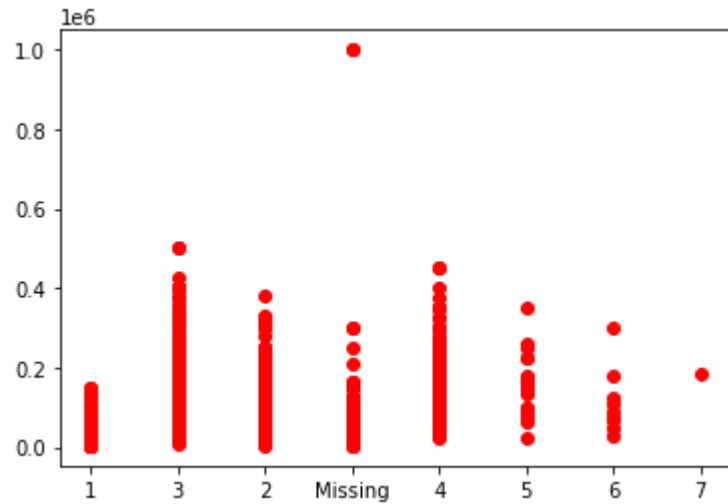
```python
#draw scatter plot of Bathrooms and Rent/Month (use training data)
plt.scatter(df["Bathrooms"],df["Rent/Month"],color='red')
```

```
plt.show
```

    <function matplotlib.pyplot.show>


```
# describe statistical analysis of the dataframe 'df' data #8 df.describe()
```

	Rent/Month	Parking
count	15386.000000	15351.000000
mean	54306.799363	0.971272
std	43373.356708	0.672316
min	3500.000000	0.000000
25%	30000.000000	1.000000
50%	43000.000000	1.000000
75%	65000.000000	1.000000
max	1000000.000000	4.000000

```
# display scatter plot of 'Rent/Month' and 'Parking'
plt.scatter(df['Rent/Month'],df['Parking'],color='blue')
```

```
# display Parking column null values index/row number entries #10
df.isnull().sum()
```

```
Locality                0
Type                    0
Rent/Month              0
Build_up_area(sq.ft)    0
Furnishing              0
Bathrooms               0
Balcony                 0
Parking                 35
Carpet_area(sq.ft)      0
dtype: int64
```

```
# find the mean value of Parking column and store in variable pVal #replace the
null values of Parking column with pVal. #11 pVal =
df['Parking'].mean() df = df.fillna(pVal)
pVal
```

```
0.9712722298221614
```

```
# Verify, is there any column has null
values? #12
df.isnull().sum()
```

```
Locality                0
Type                    0
Rent/Month              0
Build_up_area(sq.ft) 0
Furnishing              0
Bathrooms               0
Balcony                 0
Parking                 0
Carpet_area(sq.ft)    0 dtype: int64
```

```
# display crosstab report of Type and Furnishing
pd.crosstab(df.Type,df.Furnishing)
```

Furnishing Type	Fully Furnished	Semi Furnished	Unfurnished
1 BHK Apartment	1144	1851	1759
1 RK Apartment	278	417	563
2 BHK Apartment	1804	2833	1614
3 BHK Apartment	996	1563	564

```
# display pairplot of data sns.pairplot(df)
plt.show()
```
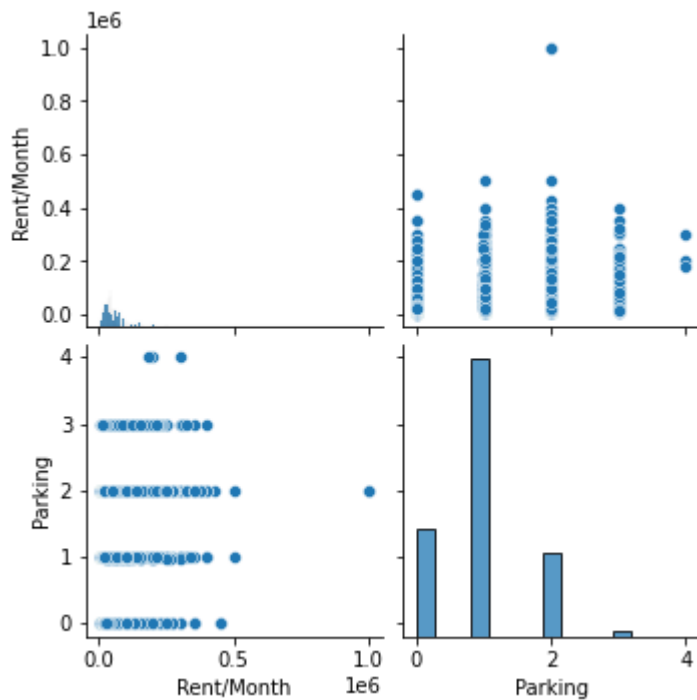


```
# display distribution of Rent/Month column data #13
df.hist(column="Rent/Month")
array([[<matplotlib.axes._subplots.AxesSubplot object at 0x7f5e27d3a990>]],
      dtype=object)
```
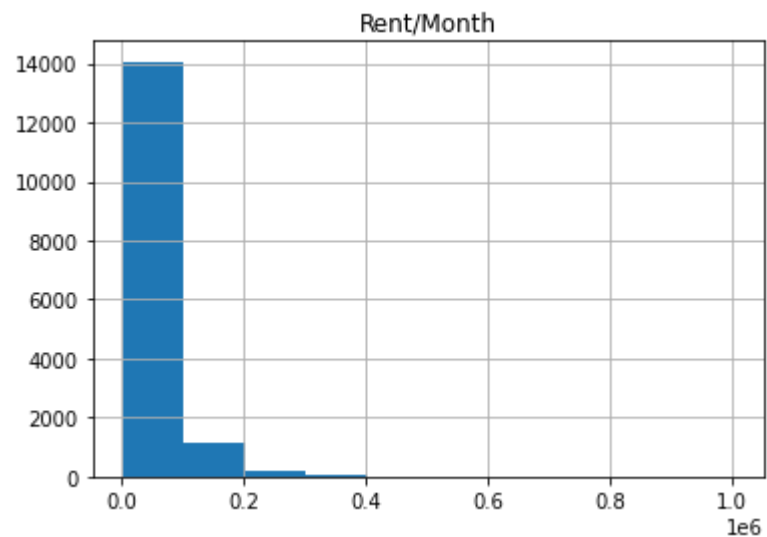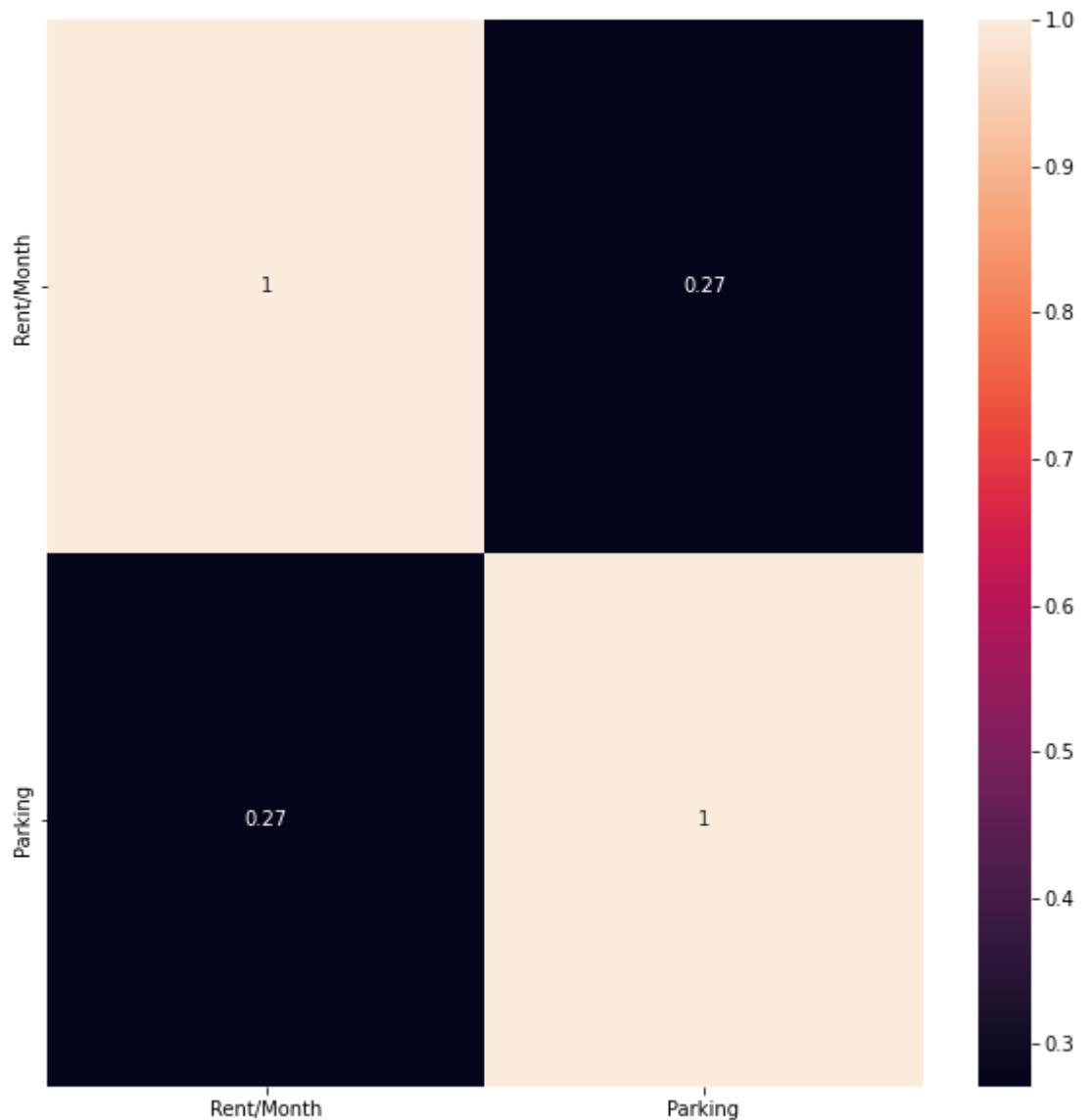


```
#display correlation of data df.corr()
```

	Rent/Month	Parking
Rent/Month	1.000000	0.271151
Parking	0.271151	1.000000

```
# display heatmap of dataframe df
plt.figure(figsize =(10,10))
sns.heatmap(df.corr(),annot=True)
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f5e27cb9210>


```
df.columns
```

Index(['Locality', 'Type', 'Rent/Month', 'Build_up_area(sq.ft)',
       'Furnishing', 'Bathrooms', 'Balcony', 'Parking',
       'Carpet_area(sq.ft)'], dtype='object')

```
#display boxplot of Parking column #14
df.boxplot(column="Parking")
```

✓