# EDA PROJECT

**Name-Dhanraj Rajeev Singh Section-K20RU**

# Roll No.- RK20RUB53

**Registration No.- 12002739 Submitted to- Abhijeet Dutta Sir**

House Rent Dataset

# Dataset Description:

1. Dataset contains a file named House\_Rent\_Dataset.csv which is imported in jupyter notebook for analysis.
2. This file contains 4746 rows and 12 columns.
3. Python libraries/packages will come handy for data analysis. Libraries such as numpy, matplotlib, pandas, seaborn and more are used.

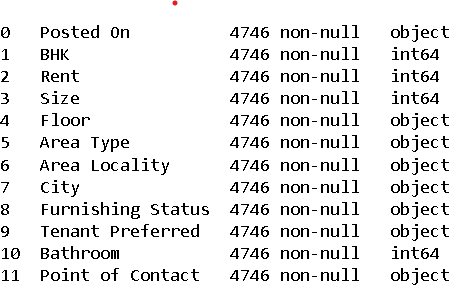
# Why this dataset?

Growing up I was used to shifting from one state to another constantly changing houses. So, I was always aware about factors of the house such as size, rent, BHK’s. Whether the room is fully furnished or semi-furnished, how many bathrooms are there? Questions like these are important to ask when shifting to another house. So, I figured since this is personal to me, I can understand more about houses and how their prices can vary by performing multiple visualizations in jupyter notebook.

# Table contents of House\_Rent\_Dataset:

Tables:-



Datatype: -

# Approach:-

1. Understanding the dataset using basic commands.
2. Analyzing all columns and rows.
3. Cleaning the dataset if required.
4. Performing Univariate, Bivariate, and multivariate analysis.
5. Create insights.
6. Conclusion of the EDA.

# Insights of the dataset: -

1. What factors contribute to the price of a house? What is the reason for the same?
2. Is the price of the house directly correlated with the size of the house?
3. Does location of the house play an important role in determining the price?
4. Is furnishing status a vital factor?
5. Bathroom and BHK columns have a high correlation with the price.
6. Most of the house rentals are of low-price value.
7. Bachelor/Family is a high tenant preference.
8. Furnishing status doesn’t play a major role in terms of preference of the tenant.
9. Locality is the major factor determining the price of rent. Kolkata having low rent rates meanwhile Mumbai having high rates.

# How this can help a franchise: -

This can help real estate franchises understand what commodities a tenant is looking for when purchasing a rental property. Franchises can also benefit by knowing the prices of similar type of houses, such that they are not under valuing their property by selling it cheaper. The locality is a major factor, if franchises are unaware of the prices of the locality, they can often downplay themselves, hence benefitting the tenants.

Graphical user interface

Description automatically generated with medium confidence

Table

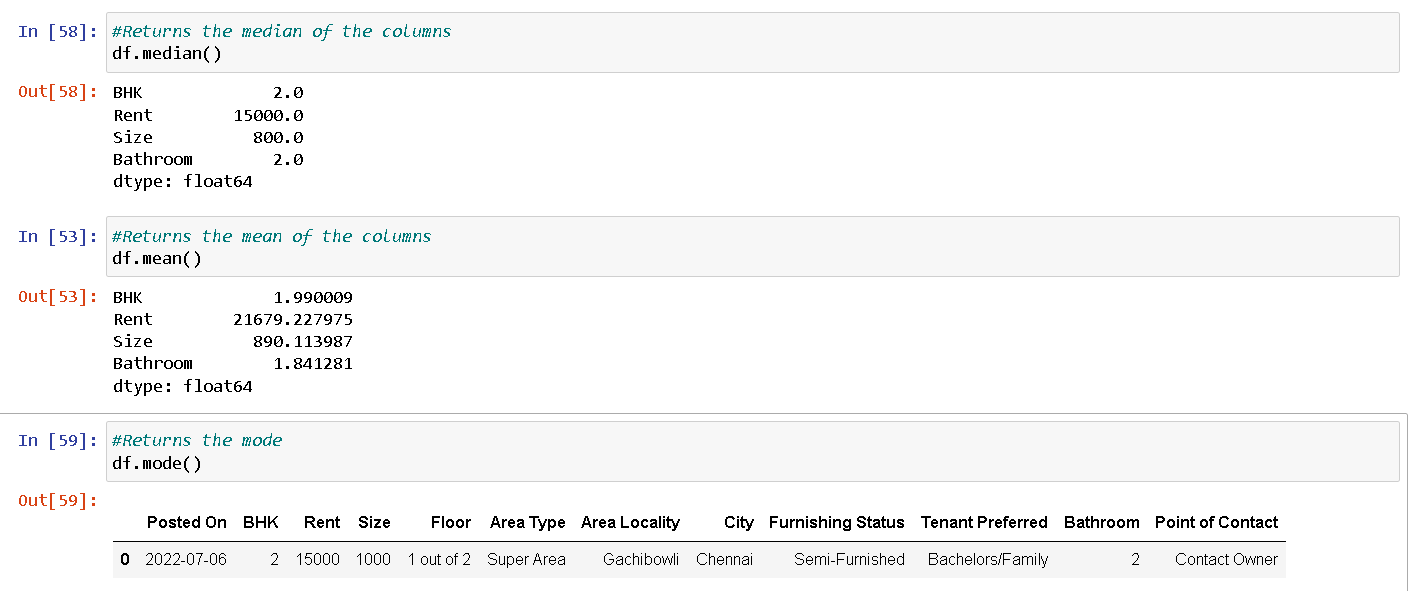
Description automatically generated

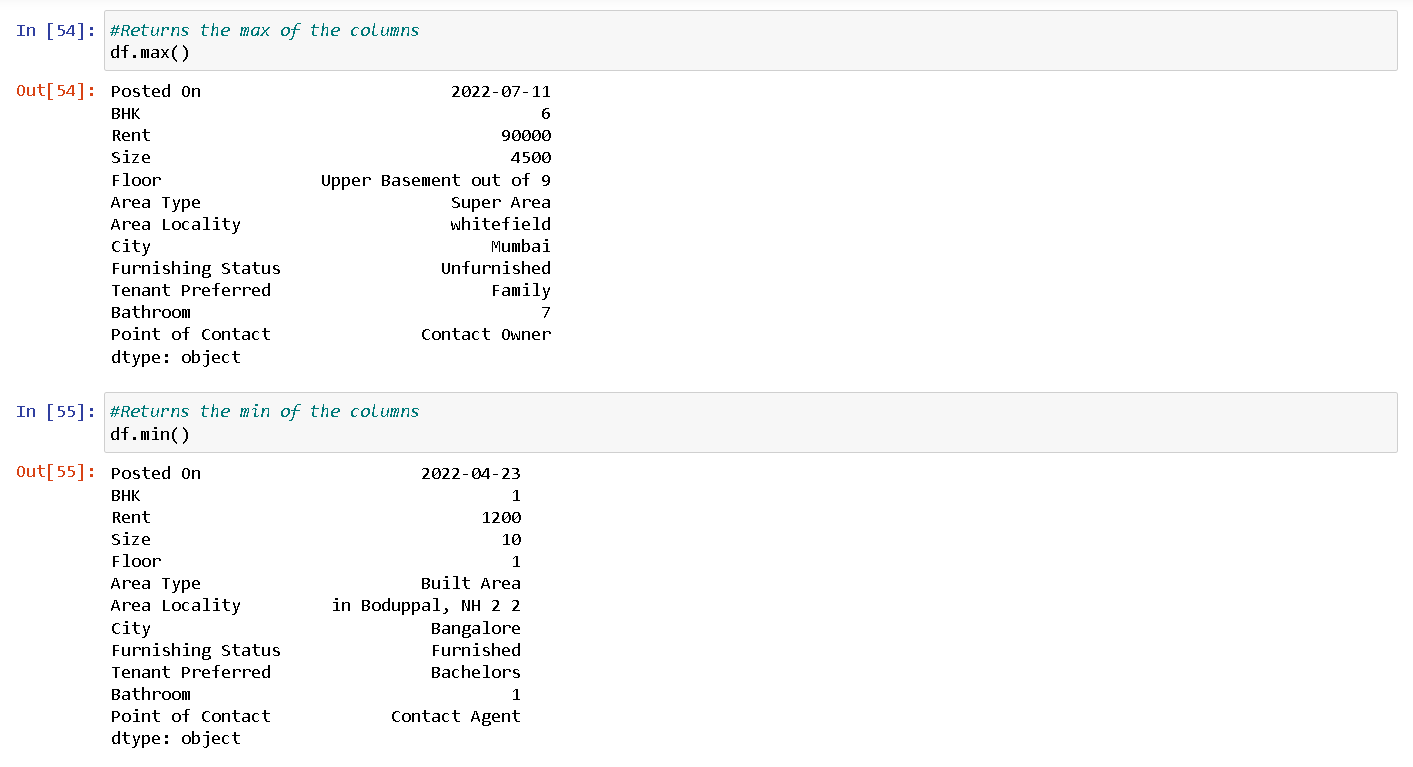
Graphical user interface, text

Description automatically generated

Graphical user interface, text, application

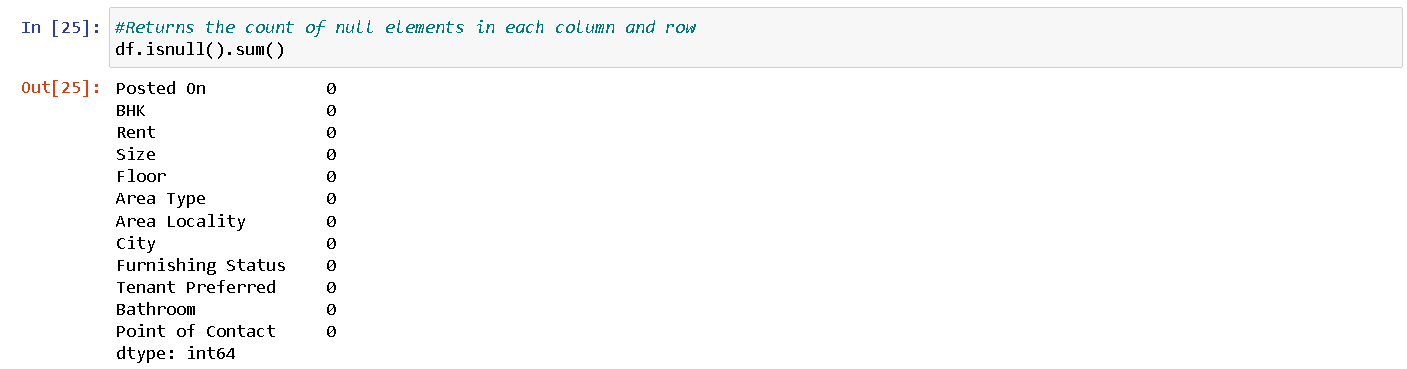
Description automatically generated

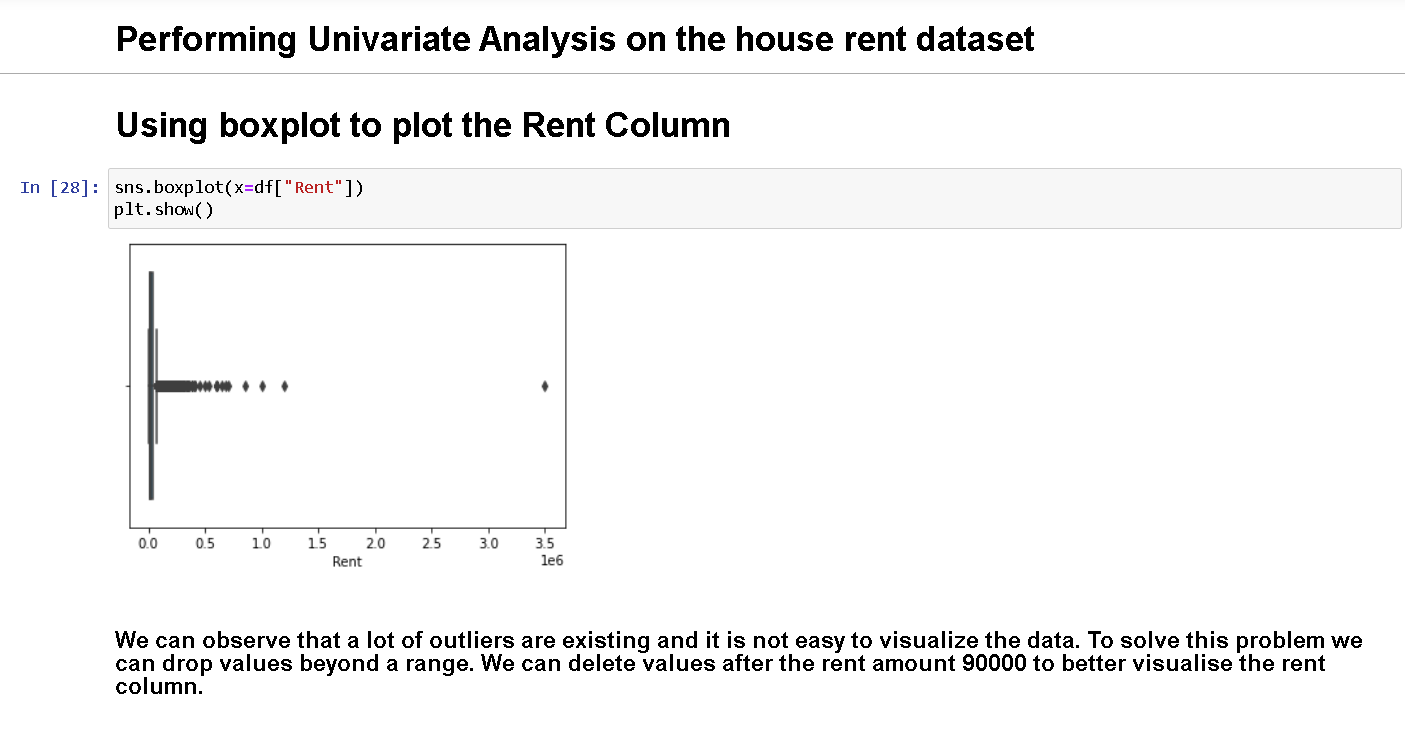


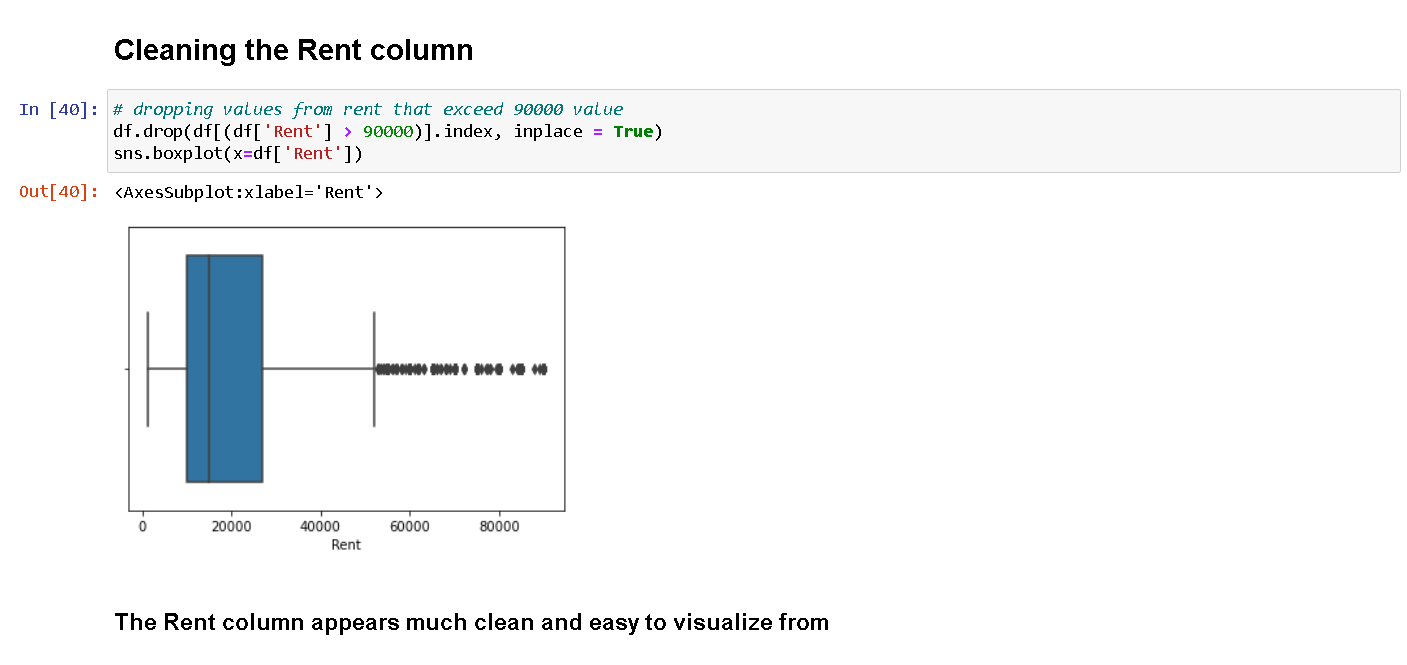


Graphical user interface, text, application, email

Description automatically generated







Graphical user interface

Description automatically generated with medium confidence

Graphical user interface

Description automatically generated with low confidence

Graphical user interface, application

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, application

Description automatically generated  
  
Schematic

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

A hypothesis test is a formal statistical test we use to reject or fail to reject some statistical

hypothesis. A one sample t-test is used to test whether the mean of a population is equal to

some value.

Here, we want to know whether the mean Rent is equal to 21600 units

or not.

The ttest\_1samp() is used from the scipy.stats library to perform a one sample t-test.

The two hypotheses for this one sample t-test are as follows:

a) H0: µ = 21600 (the mean rent is equal to 21600 Rs)

b) HA: µ ≠21600 (the mean rent is not equal to 21600 Rs)

The t test statistic is 0.7693038489296401 and the corresponding two-sided p-value

is 0.2933036995166342.

Because the p-value of our test (0.2933036995166342) is greater than alpha = 0.05, we fail

to reject the null hypothesis of the test.

We do not have sufficient evidence to say that the mean rent is equal to 21600 Rs

Conclusion of this analysis: -

1. Furnishing status and preferred tenant type doesn’t contribute to the increase in the rent of the house.
2. Rent prices of houses are generally expensive when purchasing through an agent in comparison to purchasing directly from the owner.
3. Size of these houses are mostly concentrated at 1000. Outliers to this fact exist.
4. Mean rent is priced at Rs 21679.227975 while median rent is 15000.
5. Lot of outliers exist in our analysis when analyzing the Rent column. In our analysis the range of the Rent price varies from 15000 to approximately 28000.
6. Bathroom and size, BHK and size, BHK and bathroom are highly correlated.