Name:

School Name: BAY ATLANTIC UNIVERSITY

Course Number and Name:

Professor: Micheline Al Harrack

Final Exam: Project Topic

**Research question**

How can data collected from people residing in different areas help us to predict the rent of housing**.** In this case, data is provided from github with the link[**https://www.kaggle.com/datasets/iamsouravbanerjee/house-rent-prediction-dataset?select=House\_Rent\_Dataset.csv**](https://www.kaggle.com/datasets/iamsouravbanerjee/house-rent-prediction-dataset?select=House_Rent_Dataset.csv)**.** It is data from India country.

Housing in India varies from the places of else while maharajas to modern apartment buildings in big cities to tiny huts in far-flung villages. There has been tremendous growth in India's housing sector as incomes have risen. The Human Rights Measurement Initiative finds that India is doing 60.9% of what should be possible at its level of income for the right to housing.

Renting, also known as hiring or letting, is an agreement where a payment is made for the temporary use of a good, service, or property owned by another. A gross lease is when the tenant pays a flat rental amount, and the landlord pays for all property charges regularly incurred by the ownership. Renting can be an example of the sharing economy. The data has 4746 rows and 12 columns.

**Dataset Columns:**

* BHK: Number of Bedrooms, Hall, Kitchen.
* Rent: Price of the Houses/Apartments/Flats.
* Size: Size of the Houses/Apartments/Flats in Square Feet.
* Floor: Houses/Apartments/Flats situated on which Floor and Total Number of Floors (Example: Ground out of 2, 3 out of 5, etc.)
* Area Type: Size of the Houses/Apartments/Flats calculated on either Super Area or Carpet Area or Build Area.
* Area Locality: Locality of the Houses/Apartments/Flats.
* City: City where the Houses/Apartments/Flats are Located.
* Furnishing Status: Furnishing Status of the Houses/Apartments/Flats, either it is Furnished or Semi-Furnished or Unfurnished.
* Tenant Preferred: The type of Tenant Preferred by the Owner or Agent.
* Bathroom: Number of Bathrooms.
* Point of Contact: Whom should you contact for more information regarding the Houses/Apartments/Flats

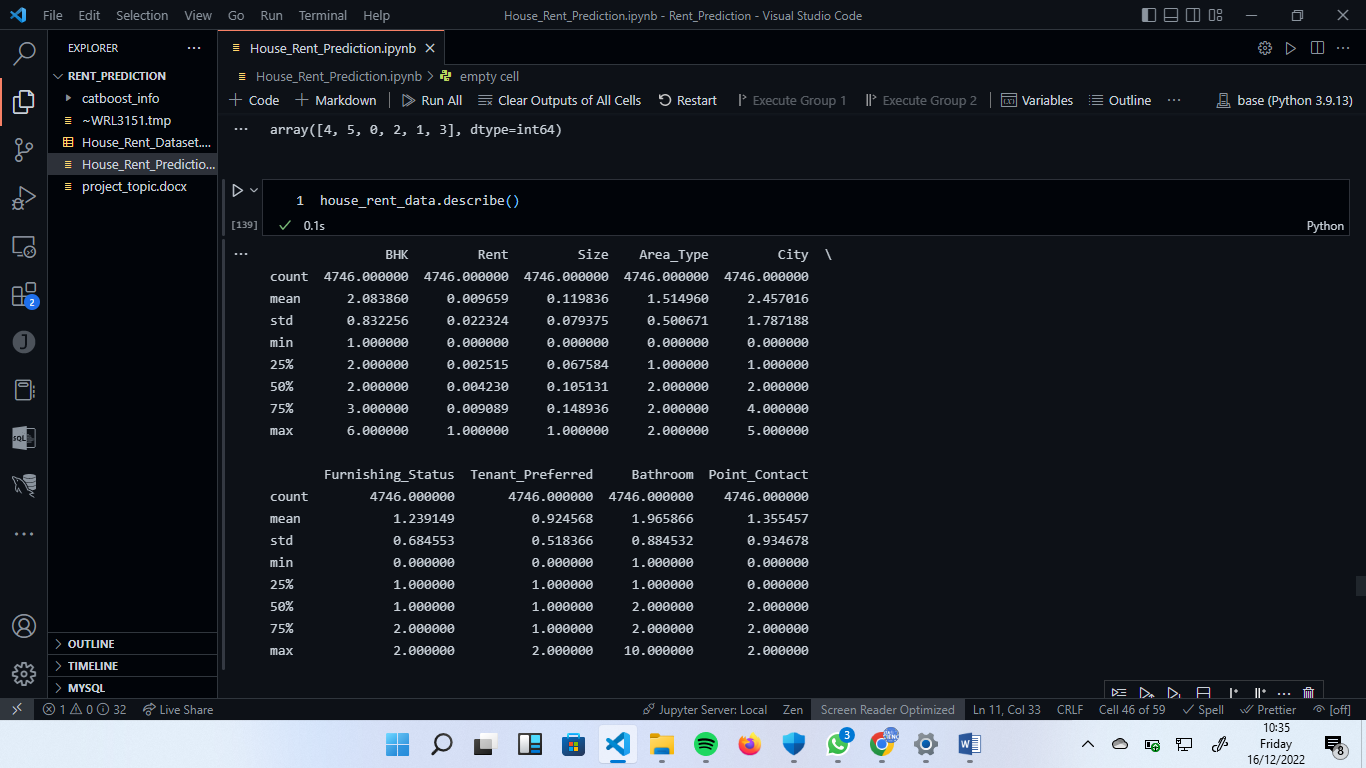
The first thing to is to load the dataset into Jupyter notebook.

**Data cleaning**

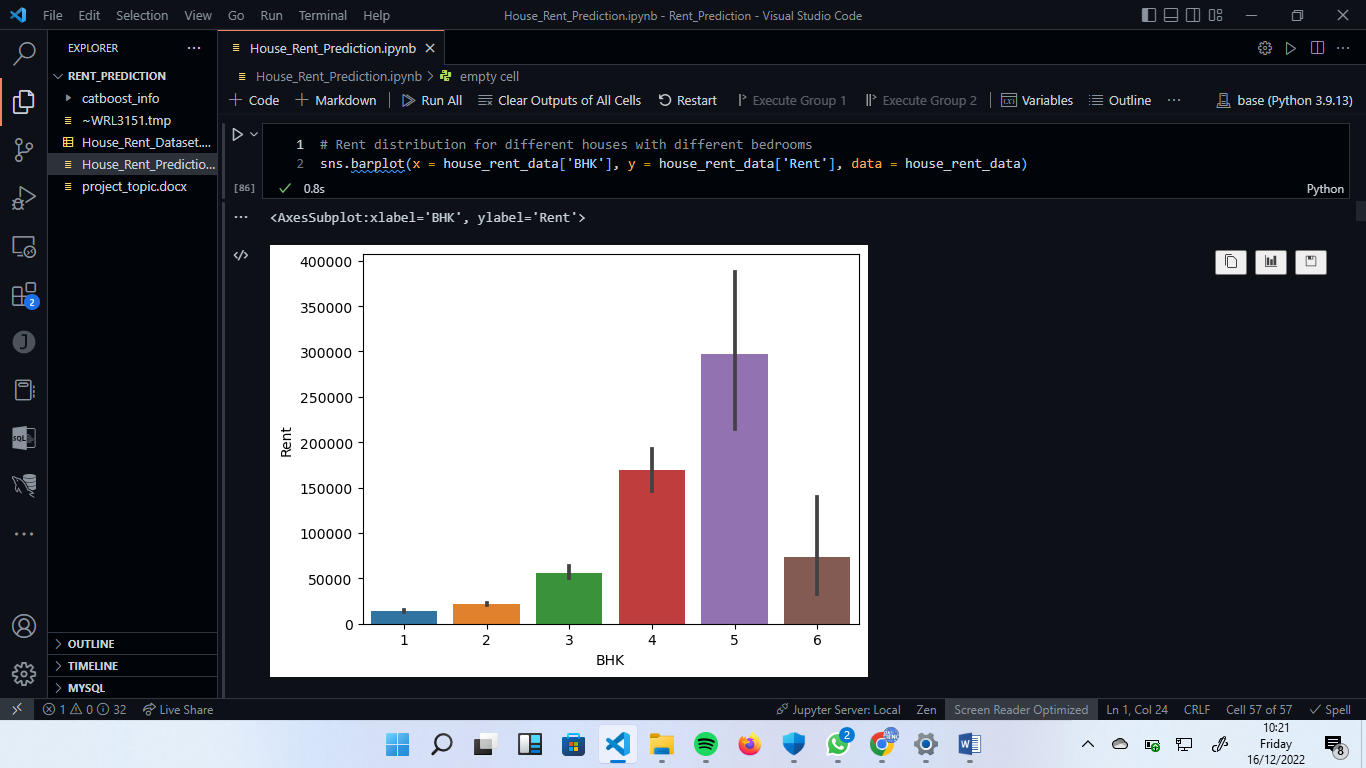
Can be observed that the data has zero missing values, there are no duplicates. The columns are named incorrectly and therefore there is need to rename some of the columns. The data columns Floor has values consist of numbers and strings, and therefore, can be split into the floor number and the total floors. And then, all categorical variables can be label encoded for prediction. The last thing to do is remove unused columns in the creating the model.

**Data analysis and Visualization**

Let’s discover some trends within the data. The summary statistics is given below

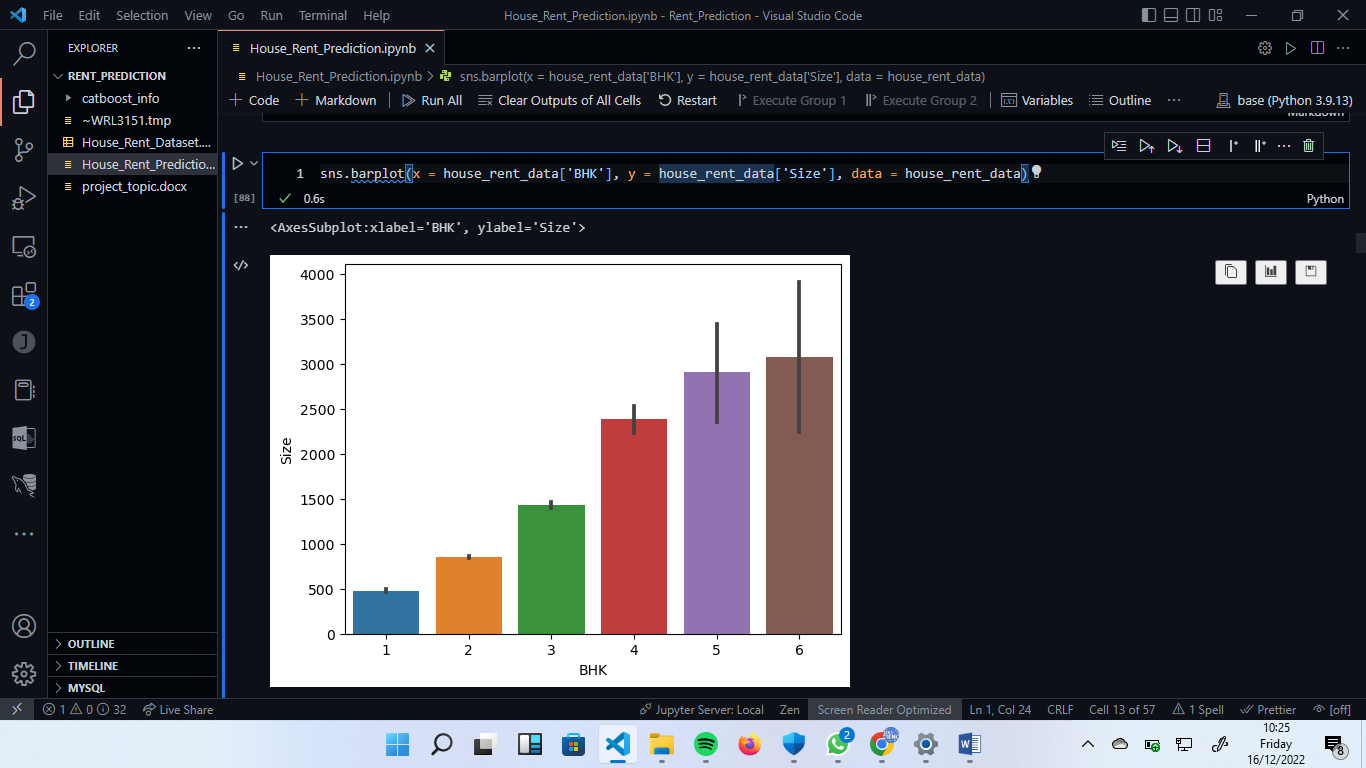


The rent distribution for all houses of different number of bedrooms, kitchens and etc.



It looks like houses that have 4 or 5 rooms are higher rent prices than 6.

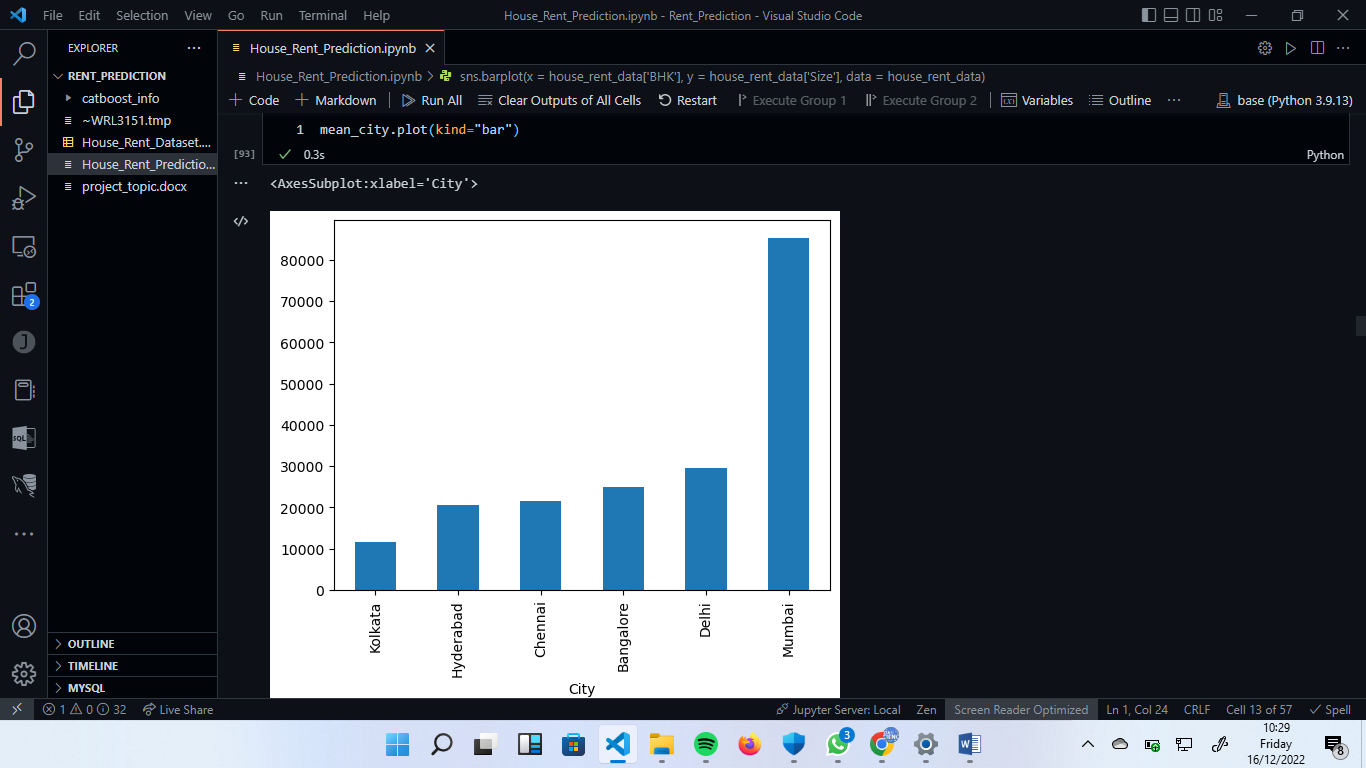
6 bedrooms houses can be out of the City or not big as 4 and 5 bedrooms houses.



Now, can be confirmed that, houses are out of the city. Examining the Rent column, It is clear that the data is skewed to the right



Mumbai city is very expensive to rent that others cities.



The house built on carpet are very expensive compared to the others

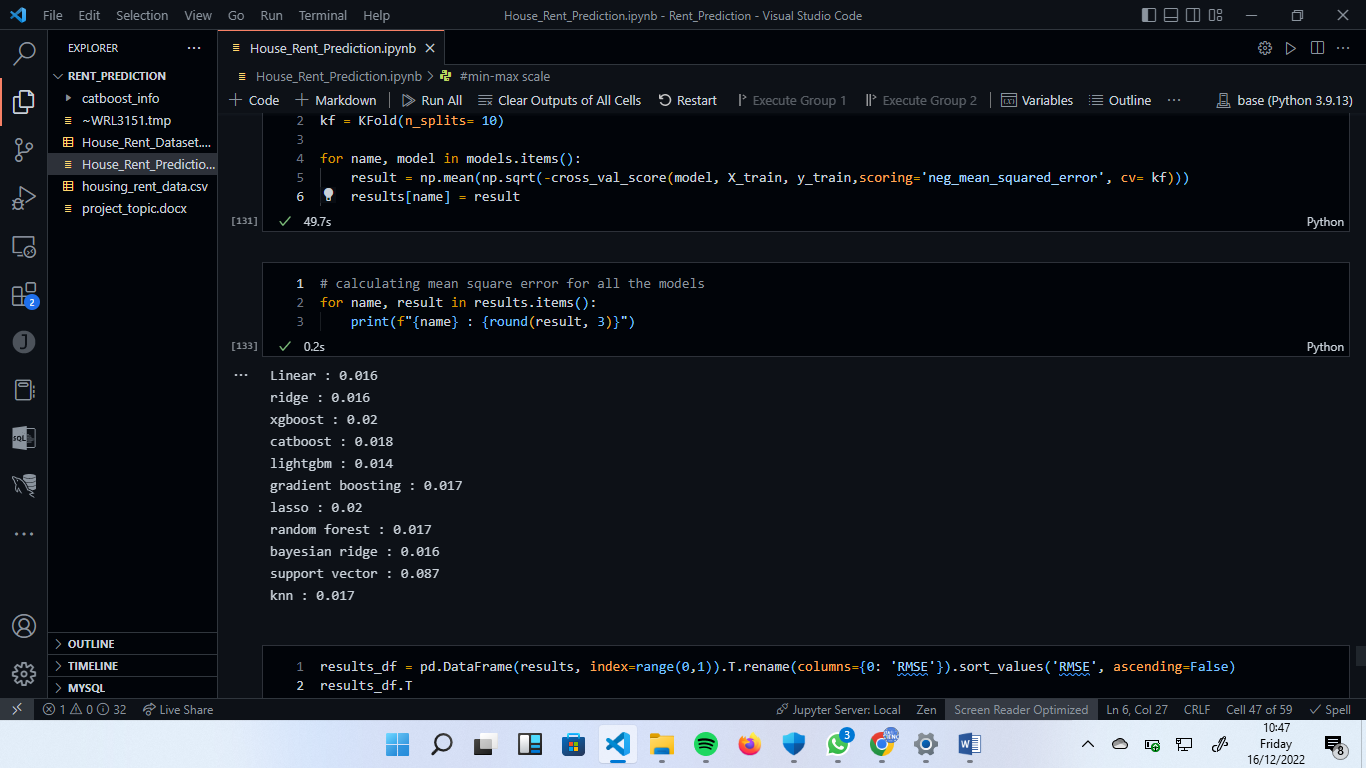


**Machine learning**

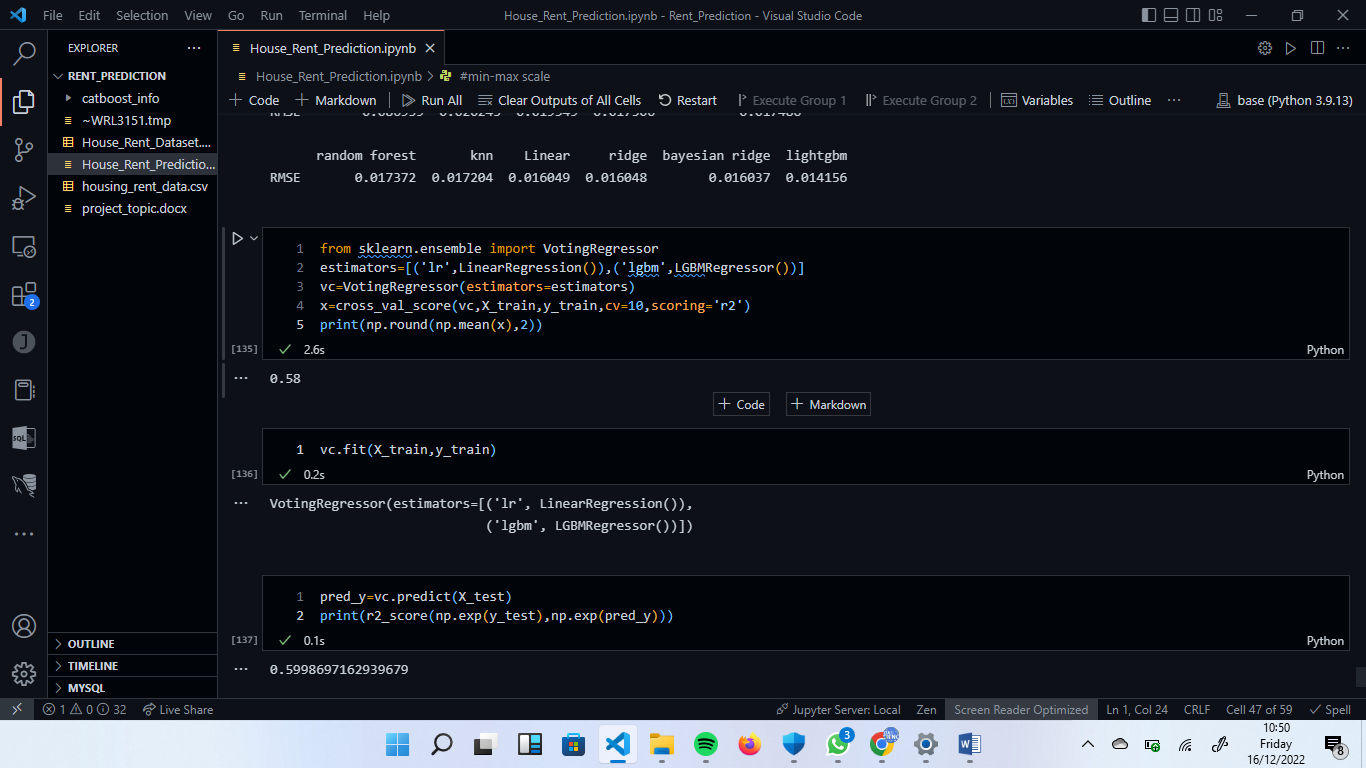
The dependent variable is Rent and others variables are independent variables. The correlations values for some variables are shown below



The first thing to do is divide data into training and testing sets. And since values of the Rent and Size are high, they can be normalized. Let’s train regressor and compare their performaces.



Let’s choose the linear regressor and Ligthgbm regressor and combine them**.**



The R2 score is 0.6 and therefore that is the best combination.

**Source:**

Obtained this data set from Kaggle’s “**House rent prediction dataset**”

**Reference:**

https://www.kaggle.com/datasets/iamsouravbanerjee/house-rent-prediction-dataset