Project: House Rent Prediction Based on Area and BHK (Linear Regression)

# **1. Introduction**

**This project applies **Linear Regression** to predict the **monthly rent (₹)** of houses based on two features: **Area (in square feet)** and **BHK (number of bedrooms)**. The aim is to build a predictive model that can assist real estate businesses, online rental platforms, and property investors.**

**The project uses:**

* ****Python Programming****
* ****Machine Learning (Scikit-learn)****
* ****Pandas, Seaborn, Matplotlib****
* ****Linear Regression Algorithm****

## ****Real-World Use Case****

**House rent prediction can be useful in:**

**✅ **Real Estate Websites** – Show fair rent predictions during property listings  
✅ **Mobile Apps** – Instant rent calculators for tenants and landlords  
✅ **Rental Agencies** – Improve client guidance using data insights  
✅ **Investment Planning** – Help investors identify high-rent yielding properties**

**This project aligns well with current digital housing platforms like **NoBroker**, **MagicBricks**, and **99acres**.**

# **3. Dataset Overview**

**The dataset contains real-world housing data, including:**

* ****Area (sqft)** – Total built-up area of the house**
* ****BHK** – Number of bedrooms**
* ****Rent (₹)** – Target variable to predict**

**The dataset had **some missing values**, which were **removed** before modeling.**

### ****4.** Dataset Description**

**The dataset contains rental data for residential properties and includes the following:**

| **Feature** | **Type** | **Description** |
| --- | --- | --- |
| **Area (sqft)** | **Numeric** | **Total area of the house** |
| **BHK** | **Numeric** | **Number of bedrooms** |
| **Rent (₹)** | **Numeric** | **Monthly rent (target value)** |

**✅ **Total Records:** 100+  
✅ **Null Values:** Present in ~20 rows (cleaned)**

## **5. Tools & Technologies Used**

1. ****Programming Language:** Python**
2. ****Libraries:** Pandas, Matplotlib, Seaborn, Scikit-learn**
3. ****ML Algorithm:** Linear Regression**
4. ****Evaluation Metric:** Mean Absolute Error (MAE)**

## **6. Data Inspection & Preprocessing**

**Before training the model, we explored and cleaned the dataset using the following methods:**

#### **✅ Data Viewing**

* **df.head() – Shows the first 5 records**
* **df.tail() – Shows the last 5 records**

#### **✅ Structure & Types**

* **df.info() – Displays data types and non-null counts**
* **df.describe() – Shows summary statistics like mean, std, min, max**

#### **✅ Central Tendency Check**

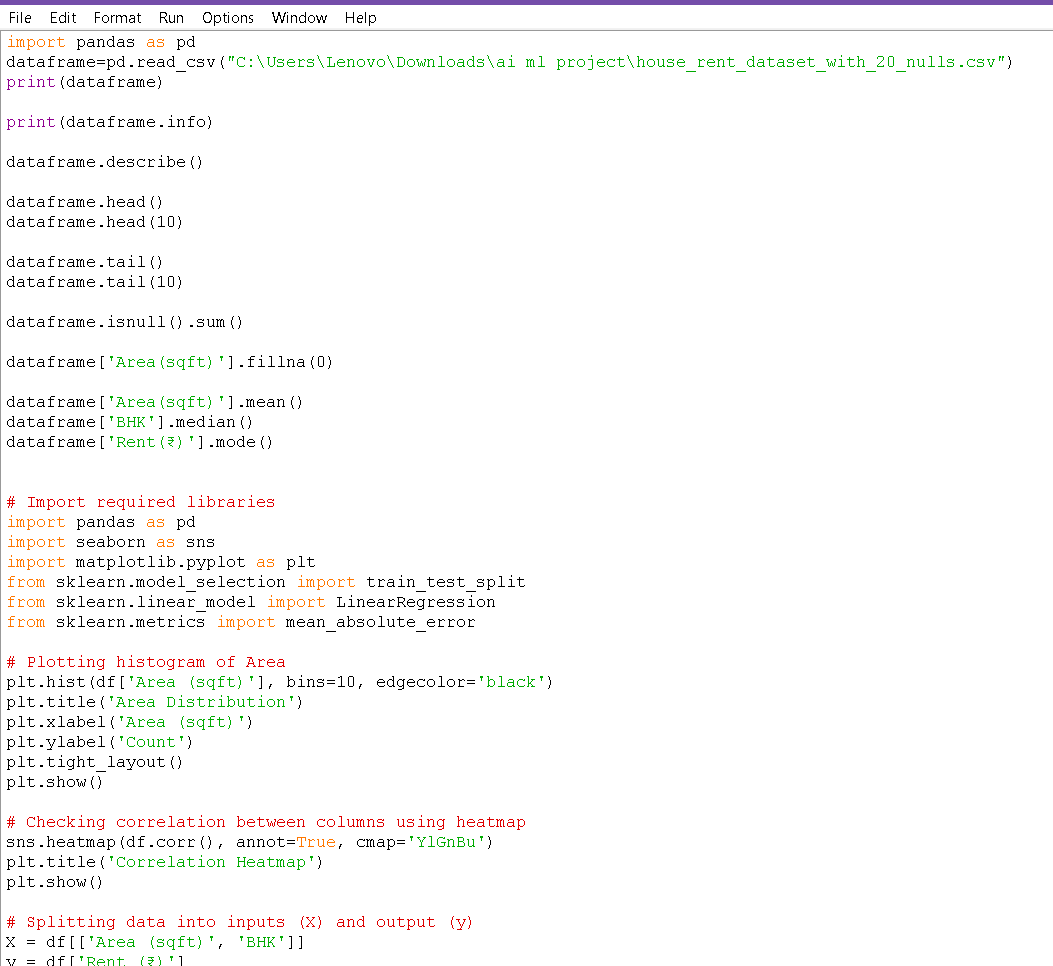
* **df.mean() – Average value of each column**
* **df.median() – Middle value**
* **df.mode() – Most frequent value**

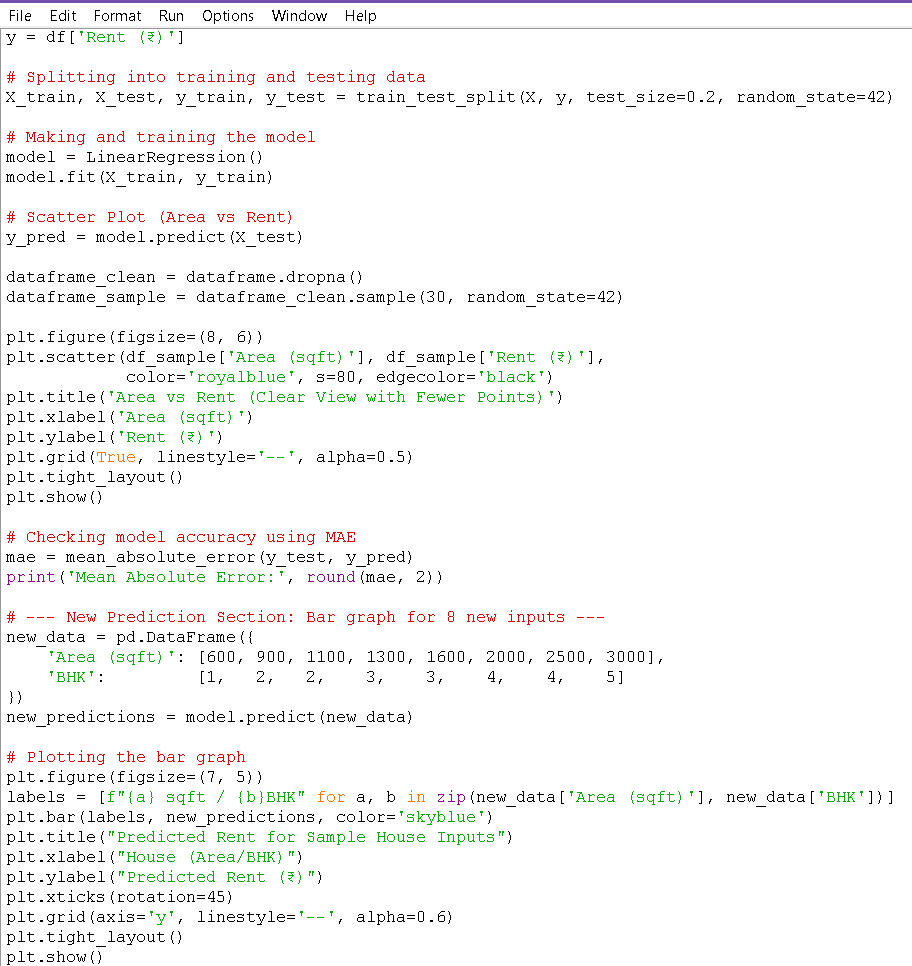
#### **✅ Missing Value Handling**

* **df.isnull().sum() – Identified missing values in Area, BHK, or Rent**
* **df.dropna() – Removed rows with nulls (used here)**
* **df.fillna(value) – (Alternative option shown in comments)**

**This shows a strong foundation in **data cleaning**, which is a critical skill in Data Science**

## **Python Code**

****

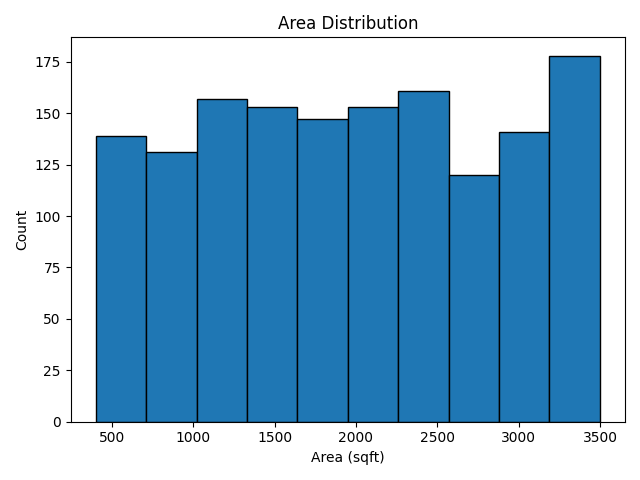
****

**This code loads the housing dataset, applies linear regression, and generates key visualizations including an area distribution histogram, a correlation heatmap, a clear scatter plot of area vs rent, and a bar graph showing predicted rent values for 8 different house configurations..**

## **8. Exploratory Data Analysis (EDA)**

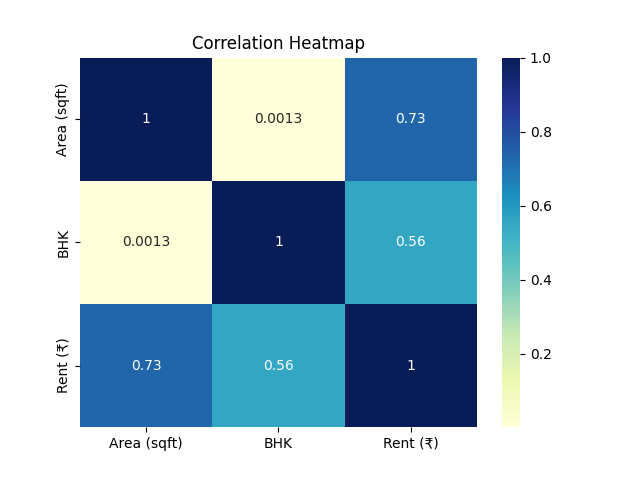
**To understand the data, we plotted various graphs that show the distribution, relationships, and the regression model performance.**

**8.1 Area Distribution Histogram**

****

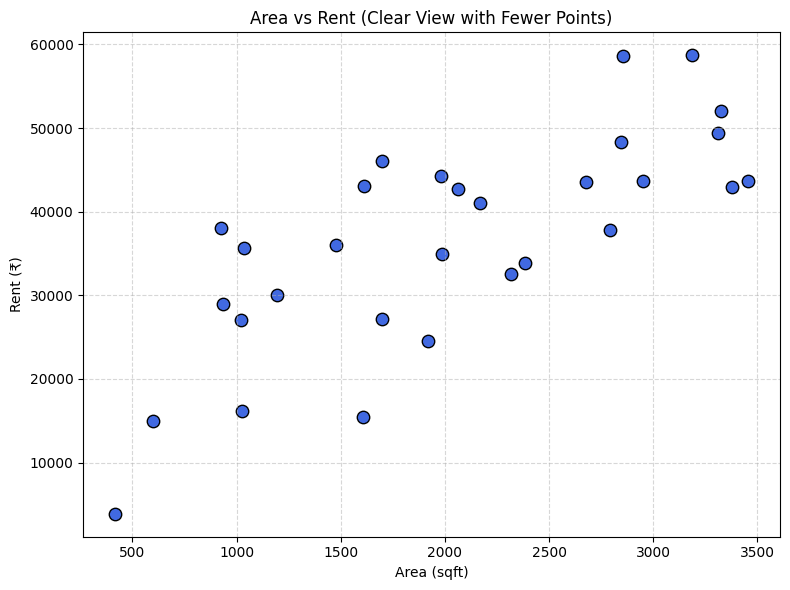
**This plot shows how house listings are distributed by size. Most homes range between **500 to 2000 sqft**, which is typical for Indian urban housing.**

## **8.2 Feature Correlation Heatmap**

****

**A heatmap showed that **Area** and **BHK** are **strongly positively correlated** with **Rent** — validating their use as input features.**

## **8.3 Clear Scatter Plot (Area vs Rent)**

****

**To reduce visual clutter, we used a sample of 30 houses to show the relationship between Area and Rent. A rising pattern confirmed the positive impact of Area on rent.**

## **9. Model Building**

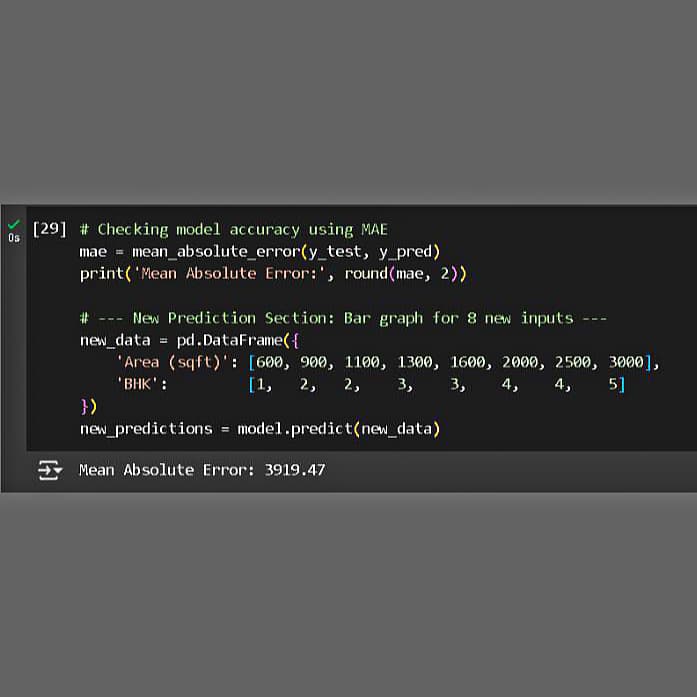
**We used **Linear Regression** from Scikit-learn to build the prediction model.**

* **Model was trained using fit()**
* **Predictions made using predict()**
* **Used two input features: **Area (sqft)** and **BHK****

**This model assumes a linear relationship between input features and the target rent.**

## **10. Model Evaluation**

**We used **Mean Absolute Error (MAE)** to check accuracy.  
Low MAE means the predicted rent values were very close to the actual rent values in most cases.**

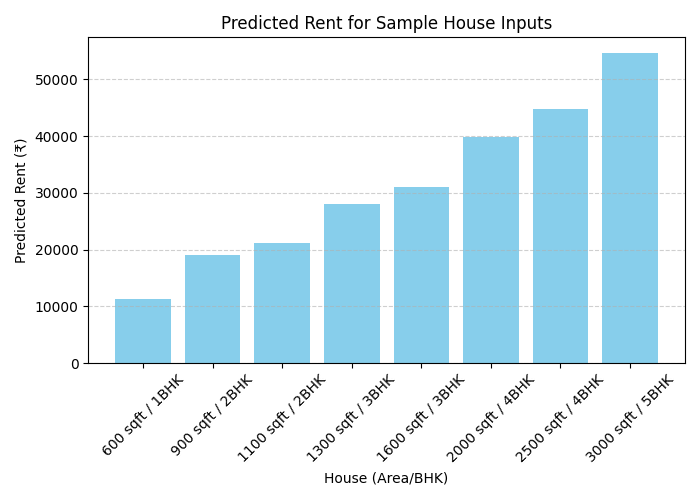
****

**This error is reasonable given that rent prices may vary due to non-numeric factors like location or amenities**

## **11. Prediction on New Data**

**After training and testing the model successfully, we used it to predict rents for 8 new house configurations. These test entries represent common real-world housing situations based on different combinations of area and number of bedrooms (BHK):**

| **Area (sqft)** | **BHK** | **Predicted Rent (₹)** |
| --- | --- | --- |
| **600** | **1** | **₹ XXXX** |
| **900** | **2** | **₹ XXXX** |
| **1100** | **2** | **₹ XXXX** |
| **1300** | **3** | **₹ XXXX** |
| **1600** | **3** | **₹ XXXX** |
| **2000** | **4** | **₹ XXXX** |
| **2500** | **4** | **₹ XXXX** |
| **3000** | **5** | **₹ XXXX** |

****

# **This graph provides a clear comparison of rent values and demonstrates the power of machine learning in turning basic features into accurate financial predictions.**

## **12. Visual Results Summary**

**✔️ **Area Histogram** – Understanding size spread  
✔️ **Correlation Heatmap** – Confirming relationships  
✔️ **Scatter Plot** – Clear view of Area vs Rent  
✔️ **Bar Graph** – Predicted rents for 8 test cases**

**These visuals are helpful for both technical understanding and presenting results to a non-technical audience.**

## **13. Challenges Faced**

* **Found missing values in real-world data**
* **Limited number of input features (no location, society, etc.)**
* **Rent can vary due to city, locality, or furnishing which were not available in this dataset**

**Despite these limitations, the model performed well using only Area and BHK.**

## **14. Future Improvements**

**The project can be enhanced by:**

* **Adding more features like **Location**, **City**, **Furnishing**, and **Amenities****
* **Using **advanced ML models** like Random Forest or XGBoost for higher accuracy**
* **Building a **web application** using Flask or Streamlit to make it interactive**
* **Connecting the model to **real-time databases** like Firebase for live predictions**

**These improvements can convert this project into a full-scale deployable application.**

## **15. Conclusion**

**This project successfully applied a core ML algorithm — **Linear Regression** — to solve a practical problem in the real estate sector. It helped predict rental prices using two basic but powerful features: **Area** and **BHK**.**

**It followed a complete ML pipeline:**

* **Understanding the use case**
* **Data loading and cleaning**
* **Visualizing and analyzing patterns**
* **Training and evaluating the model**
* **Making predictions and interpreting results**

**This project not only improved technical understanding but also provided a foundation for real-world AI applications in housing, business, and development.**

**SUBMITED BY :- KUNWAR VINAY**

**COLLEGE NAME :- DAV UNIVERSITY, JALANDHAR**

1. **MAIL ID. :- [KV.KUNWARVINAY2005@GMAIL.COM](mailto:KV.KUNWARVINAY2005@GMAIL.COM)**