**Predictive Insights in Real Estate: Advanced Modeling Strategies for Accurate**

# Price Forecasting

**Overview of Dataset**

This dataset has been sourced from [magicbricks.com](http://magicbricks.com/). It offers a comprehensive compilation of real estate details, encompassing numerous variables related to residential properties. It provides insights into property characteristics like (area, latitude, longitude, Bedrooms, Bathrooms, Balcony, Status, neworold parking, Furnished\_status, Lift, Landmarks, type\_of\_building, desc, and Price\_sqft). With a wealth of information, the [dataset](https://www.kaggle.com/datasets/goelyash/housing-price-dataset-of-delhiindia) facilitates an in-depth exploration of the relationships between these variables and the ultimate sale price (Price) of the properties. This dataset can be utilised as a valuable tool for constructing a predictive model, and enhancing decision-making for various stakeholders in the real estate market, including homebuyers, sellers, and industry professionals.

# Dataset Description

1. **Price:** The target variable representing the property's price.
2. **Address:** The location of the property.
3. **Area:** The size of the property in square feet.
4. **Latitude:** The geographic coordinate of the property.
5. **Longitude:** The geographic coordinate of the property.
6. **Bedrooms:** The number of bedrooms in the property.
7. **Bathrooms:** The number of bathrooms in the property.
8. **Balcony:** Indicates the number of balconies the property has.
9. **Status:** Indicates whether the property is under construction or ready.
10. **neworold:** Indicates whether the property is new or old.
11. **Parking:** Indicates the parking availability.
12. **Furnished Status:** Indicates whether the property is fully furnished, semi-furnished or unfurnished.
13. **Lift:** Indicates the total no of lifts.
14. **Landmarks:** Significant nearby locations or features.
15. **Type of Building:** The architectural type of the property.
16. **Description:** Additional information about the property.
17. **Price per Sqft:** The price of the property per square foot.

The primary goal of using the real estate dataset is to construct a robust predictive model for housing prices (Price), considering a variety of features. The challenge encompasses unravelling intricate relationships among predictors like property characteristics. The task involves crafting and optimising a machine-learning model that comprehensively captures the complexities of the real estate market and an interactive dashboard for insightful visualisation. This model aims to serve as a valuable tool for stakeholders, including homebuyers, sellers, and real estate professionals, enabling them to make well-informed decisions based on accurate predictions of property values.

**Note to the Students:** This project is an open-ended one. The students need to bring in more variety and ideas to the project to make it more interesting. They may incorporate creative ideas to make it more sophisticated or think about how they may be able to scale the project.

# Steps to Approach the Problem

* **Data Exploration and Cleaning:** Review the dataset structure and column types, and apply data cleaning and appropriate data manipulation techniques.
* **Create Visualisations:** Create visualisations to understand feature distributions and

generate correlation matrices to identify relationships between variables. Additionally, create new features to capture temporal dynamics.

* **Model Training with Regularisation:** Train the forecasting model incorporating basic as well as Ridge and Lasso regression techniques.
* **Hyperparameter Tuning for Regularisation:** Fine-tune the regularisation parameters for

Ridge and Lasso regression. Explore a range of alpha values to find the optimal level of regularisation. Understand the impact of regularisation on model coefficients and its role in preventing model complexity.

* **Comparison with Non-Regularized Models:** Compare the performance of Ridge and Lasso regression with basic regression. Assess improvements in terms of generalisation and reduction of overfitting.
* **Bias-Variance Trade-off Analysis:** Find the optimal model complexity that minimises both bias and variance to achieve a well-balanced predictive performance.
* **Model Testing and Evaluation:** Test and evaluate the models to ensure their effectiveness. Evaluate how regularisation affects the model’s performance.
* **Geometric representation:** Gain a visual insight into how Ridge and Lasso regularisation

shapes the coefficient space, aiding in the comprehension of their impact on model parameters.

* **Dashboard Development:** Create an interactive dashboard for visualising the analysis of geometric representation.
* **Documentation and Communication:** Document the inclusion of Ridge and Lasso regression in the modelling process. Clearly communicate the impact and benefits of regularisation on the predictive model. Also, state why the usage of linear regression wasn’t optimal for this project.