

# REAL ESTATE PRICE PREDICTION USING MULTIPLE REGRESSION IN EXCEL

## 1. Background and Problem Statement:

A real estate consulting firm aimed to build a **pricing estimation model** for mid-range residential properties in Pune, India. Although they had data on property size, number of bedrooms, and location, the sales team was unable to explain pricing patterns consistently. They requested a simple, Excel-based regression model that could help predict price based on key quantitative features and support their sales presentations.

## 2. Objectives:

- Build a multiple linear regression model in Excel to predict house prices
- Quantify the effect of area (sq. ft.), number of bedrooms, and distance from city center on sale price
- Validate model reliability using  $R^2$ , standard error, and residual plots
- Generate an Excel interface for price prediction by inputting property features

## 3. Methodology:

### 3.1 Dataset Overview

- **Observations:** 120 residential properties sold between Jan–Dec 2023
- **Variables:**
  - Price (dependent variable): in INR Lakhs
  - Area\_sqft: property built-up area in square feet
  - Bedrooms: number of bedrooms
  - Distance\_km: distance from the city center (Shivaji Nagar)

### 3.2 Data Preparation in Excel

- Used **named ranges** for independent variables
- Checked for missing values (none found)
- Used Excel formulas to calculate correlations before running regression

### 3.3 Regression Execution

- Used **Excel Data Analysis Toolpak** → **Regression**
- Dependent variable: Price
- Independent variables: Area\_sqft, Bedrooms, Distance\_km
- Verified regression assumptions:
  - Checked residual plots for normality
  - Confirmed low pairwise correlation between predictors (no multicollinearity)

## 4. Results and Interpretation:

### 4.1 Regression Output

Predictor	Coefficient	p-Value	Interpretation
Intercept	12.15	0.002	Base price for 0 sqft and 0 distance
Area_sqft	0.054	<0.001	Each additional sqft increases price by ₹5,400
Bedrooms	4.72	0.005	Each extra bedroom adds ₹4.72 lakhs
Distance_km	-1.26	<0.001	Price drops ₹1.26 lakhs per km from center

- **$R^2 = 0.78$**  → 78% of variation in price explained by the model
- **Standard Error = 3.45 lakhs**
- All predictors statistically significant at 5% level

### 4.2 Residual Diagnostics

- Residual plot showed **no pattern** (randomly scattered) → assumption of linearity met
- Histogram of residuals indicated **approximate normality**
- Variance Inflation Factor (VIF) approximated manually using  $R^2$  of each predictor → all < 2

## 5. Deliverables in Excel:

- **Interactive Price Estimator Tool:**
  - Users input area, bedrooms, and distance
  - Predicted price auto-calculated using regression formula

- Output sheet with:
  - Regression statistics
  - Model interpretation summary
  - Visualizations: scatter plots, residual plot, fitted line vs actual

## 6. Recommendations:

- Sales team should prioritize **larger properties within 5 km of center**, as price sensitivity to location is steep
- Develop custom pricing bands by zone based on regression equation
- Periodically retrain the model every 6 months using new property data

## 7. Stakeholder Relevance:

### Academic:

- A textbook example of multiple linear regression with only quantitative variables
- Demonstrates how Excel can be used to run, interpret, and visualize regression results

### Corporate:

- Provides real estate professionals with a lightweight pricing tool
- Enhances transparency and consistency in price justification to customers
- Builds foundational capabilities for advanced pricing models in future CRM integrations