ANALYZING THE INFLUENCE OF PROPERTY FEATURES ON HOUSING PRICES IN URBAN REAL ESTATE MARKETS USING MULTIPLE LINEAR REGRESSION IN MINITAB

1. Background & Objectives

Urban housing markets are characterized by fluctuating prices influenced by multiple property attributes. For both investors and policymakers, understanding these factors is key. This project uses Minitab's multiple linear regression to quantify how variables such as square footage, location, number of bedrooms, and proximity to amenities affect property prices.

2. Methodology

Data Source:

Synthetic dataset of 500 urban housing transactions from a Tier-1 city, including:

- **Dependent Variable:** Sale Price (USD)
- **Independent Variables:**
    - Square Footage
    - Number of Bedrooms
    - Number of Bathrooms
    - Distance to City Center (km)
    - Proximity to Metro Station (Yes/No)
    - Year Built

Software Used:

Minitab 21

Steps Followed in Minitab:

1. **Data Cleaning & Import:**
    - Imported data as Excel file.
    - Checked for missing values and removed entries with incomplete data.

2. **Exploratory Statistics:**
    - *Stat > Basic Statistics > Display Descriptive Statistics*
    - Scatterplots created for each independent variable against price.
3. **Variable Encoding:**
    - Dummy variable created for proximity to metro (1 = Yes, 0 = No).
4. **Model Setup:**
    - *Stat > Regression > Regression > Fit Regression Model*
    - Sale Price as response; others as predictors.
5. **Diagnostics:**
    - Checked **VIF** for multicollinearity.
    - Used **Residual Plots** to test for homoscedasticity.
    - Assessed **R-squared** and **Adjusted R-squared** values.

3. Results & Interpretation

- **Significant Predictors ($p < 0.05$):**
    - Square Footage
    - Distance to City Center
    - Proximity to Metro
    - Year Built
- **Model Statistics:**
    - $R^2 = 0.82$, Adjusted $R^2 = 0.80$
    - VIF values all $< 5 \rightarrow$ no multicollinearity issues
    - Residuals showed no major deviation from normality
- **Key Insight:**
    - Each additional 100 sq. ft. increases price by \$7,500
    - Homes within 1 km of metro stations were valued 12% higher on average

4. Visual Outputs (Created in Minitab)

- Scatterplot: Square Footage vs. Sale Price
- Regression Coefficient Table
- Residual vs. Fitted Value Plot
- Histogram of Residuals

5. Recommendations

- Developers should focus on building within 3 km of metro stations to command premium pricing.
- Investment strategies should prioritize newer, mid-sized homes near central business districts.
- Realtors can use these metrics for more accurate property appraisals.

6. Future Work

- Introduce interaction terms (e.g., age × size) to explore compounding effects.
- Run separate regressions for different city zones to capture spatial heterogeneity.
- Compare Minitab results with machine learning models in R or Python for robustness.

7. Stakeholder Relevance

Academic Use:

- Serves as an ideal case for teaching multiple regression with dummy variables.
- Applicable in real estate economics, urban planning, or applied statistics courses.

Corporate Use (Real Estate/Analytics Firms):

- Equips teams with evidence-based pricing strategies.
- Supports dynamic dashboards and valuation models with regression-backed logic.