# Project Title

*"Analyzing the Impact of Housing Features on Sale Prices: A Regression-Based Study on the Ames Housing Dataset"*

# Introduction

Real estate valuation is a crucial aspect of the housing market, influencing buyers, sellers, and investors. Traditional home price assessments often rely on simple factors like square footage and the number of bedrooms. However, many location-based, structural, and quality-related features contribute significantly to house prices. This study aims to explore which features impact housing prices the most and develop an accurate predictive model using machine learning regression techniques.

# Goal of the Project

### Research Question/Topic

This research investigates which housing features significantly influence the final sale price, with a particular focus on The Impact of Location-Based Features on Housing Prices. The study will also compare regression techniques to assess their effectiveness in predicting house prices.

### Machine Learning Methods/Models

Ridge Regression (L2 Regularization) – Handles multicollinearity by reducing large coefficient magnitudes.
Lasso Regression (L1 Regularization) – Performs feature selection by eliminating insignificant predictors.
Polynomial Regression – Captures non-linear relationships between features and sale prices.
Elastic Net Regression – A combination of Ridge and Lasso, balancing feature selection and regularization.
Gradient Boosting (XGBoost)

# Dataset for Analysis/Demonstration

The study will utilize the Ames Housing Dataset, which contains detailed information about residential properties in Ames, Iowa. The dataset includes 79 explanatory variables that describe various characteristics of homes, such as lot size, neighborhood, house quality, basement area, garage type, and more. The target variable is Sale Price, representing the final transaction price of each house.
Dataset: House Prices- Advanced Regression Techniques

### Description of Dataset

The Ames Housing Dataset was compiled by Dean De Cock for use in real estate data science research. It provides a modern alternative to the Boston Housing dataset, offering a broader and more detailed view of property values. The dataset consists of training and test sets, where the goal is to predict SalePrice for the test set.
Feature Categories in the Dataset:

- Structural Features: House square footage, number of bedrooms/bathrooms, basement size, and house style.
- Location-Based Features: Neighborhood, street type, and proximity to amenities.
- Quality & Condition Features: Overall condition, material quality, and exterior features.
- Market & Financial Features: Sale type, house age, and zoning classification.
- Amenities & Extras: Garage type, porch area, presence of a pool, fence, etc.

# Metrics for Evaluation

- Root Mean Squared Error (RMSE) – Measures the average difference between predicted and actual house prices.
- Feature Importance Scores – Using Lasso, Ridge, and XGBoost, the study will analyze which features impact house prices the most.
- Cross-Validation Scores – To ensure model generalizability and avoid overfitting.