Advanced Real Estate Price Prediction

Team: The Outliers

Institution: BITS Pilani Digital

Course: Advanced Apex Project 1 (2025-26)

Objective: Robust ML pipeline for Ames housing price prediction

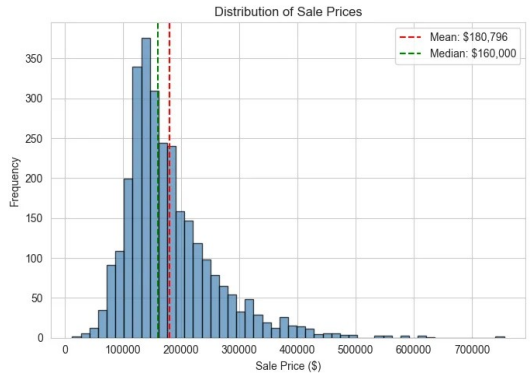
Problem Statement & Business Goal

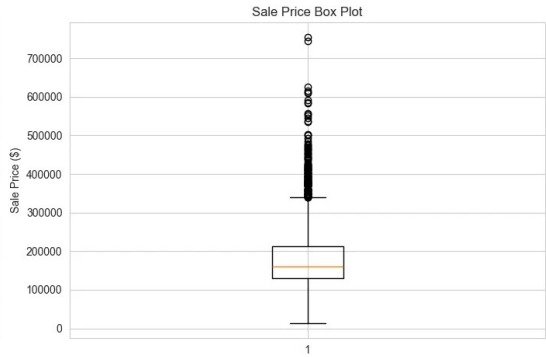
The Challenge:

- Real estate valuation is subjective and relies on heuristics
- Accurate prediction is critical for homeowners, buyers, and investors

Business Goal:

- Develop a data-driven model to minimize prediction error (RMSE)
- Provide interpretable insights into key value drivers





Methodology & Workflow

1. Data Acquisition:

Extracted Ames Housing dataset (2930 records, 82 features)

2. Preprocessing:

Imputed missing values (LotFrontage median) and removed outliers (IQR method)

3. Feature Engineering:

Created 'Total_SF', 'Total_Bath', 'House_Age' and log-transformed target variable

4. Modeling:

Trained Linear, Ridge, Lasso, ElasticNet. Selected Ridge Regression (Alpha=1.0) for stability

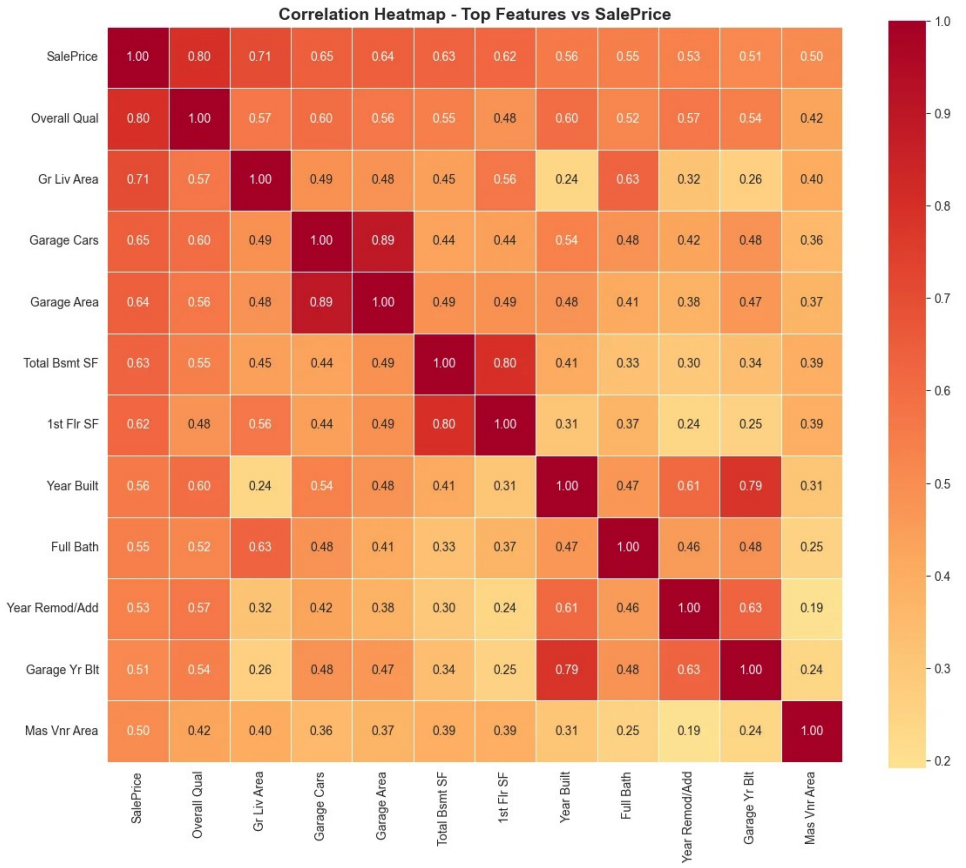
Key Insights & Results

Model Performance:

- Best Model: Ridge Regression
- R-squared: ~0.85 (85% of variance)
- RMSE: ~$34,746

Top Drivers of Price:

1. Overall Quality (Corr: 0.80)
2. Total Sq Ft (Corr: 0.78)
3. Neighborhood premium



Conclusion & Future Work

Conclusion:

- Model estimates prices within ~12% margin of error
- Quality and Condition are the ultimate differentiators

Recommendations:

- Sellers: Renovate for 'Overall Quality' ROI
- Buyers: Seek older homes in developing areas

Future Work:

- Incorporate external economic data (interest rates, inflation)
- Experiment with XGBoost/Gradient Boosting

Project Summary Dashboard

```
AMES HOUSING PRICE PREDICTION - KEY METRICS

Dataset Overview
  - Properties Analyzed: 2,930
  - Features Used: 73
  - Target: Sale Price (12,789 – 755,000)

Model Performance (Ridge Regression)
  - R-squared: 84.94% of price variance explained
  - RMSE: $34,746 average error
  - MAE: $21,557 average absolute error

Top Predictors
  1. Overall Quality (r = 0.80)
  2. Total SF - Engineered (r = 0.79)
  3. Living Area (r = 0.71)
  4. Garage Cars (r = 0.65)
  5. Total Basement SF (r = 0.63)

Key Finding: Quality and size are the primary drivers of home value.
```