GA-DSI-123

Project 2

Ames Housing Data & Kaggle Challenge

Ayako Homma

February 17, 2023

Problem statement

The goal of this project is to conduct a comprehensive analysis of the factors that exert the greatest influence on the market value of homes in Ames, IA. Additionally, we aim to develop a robust model that can accurately predict home prices, achieving an R-squared score of at least 0.8.

By doing so, this project will enable local real estate agents to offer more informed guidance to their clients, helping them to make better decisions when buying, renting, or selling properties in the area.

Key steps in the research process

01
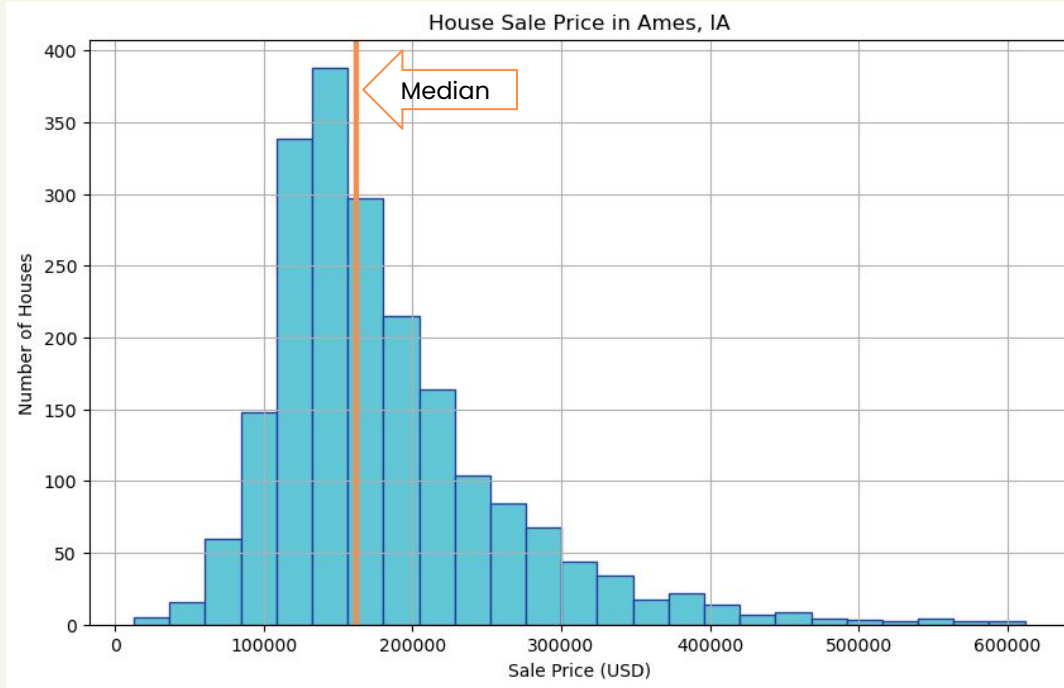
Data Import & Cleaning

02

Exploratory Data Analysis

03

Data Modeling & Evaluation

04

Conclusion & Recommendations

Median sale price is around $162K in Ames, IA

House Sale Price in Ames, IA

Median

What are some factors impacting the house value?

8 critical factors that influence a home's value

Determining the right list price is crucial. We'll explain the most important home resale value factors.
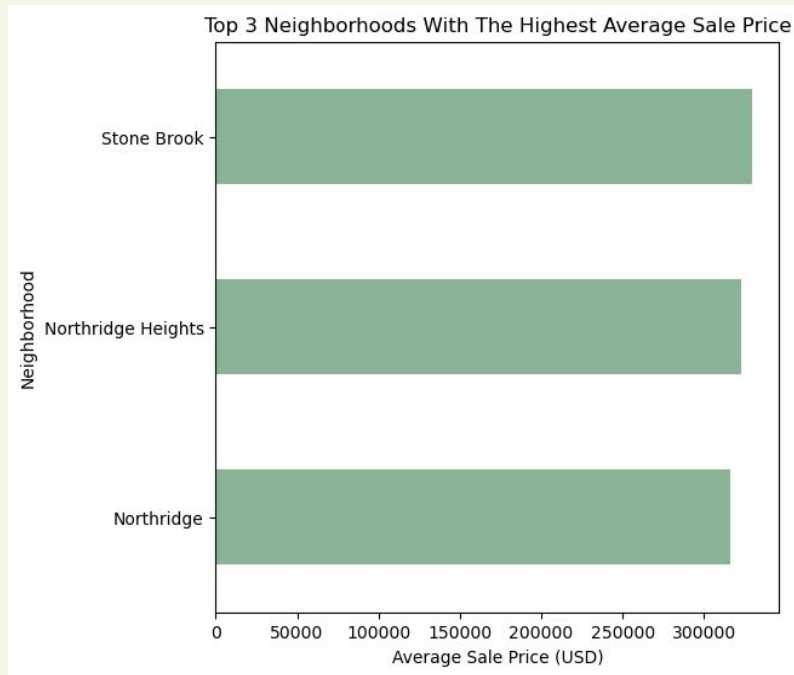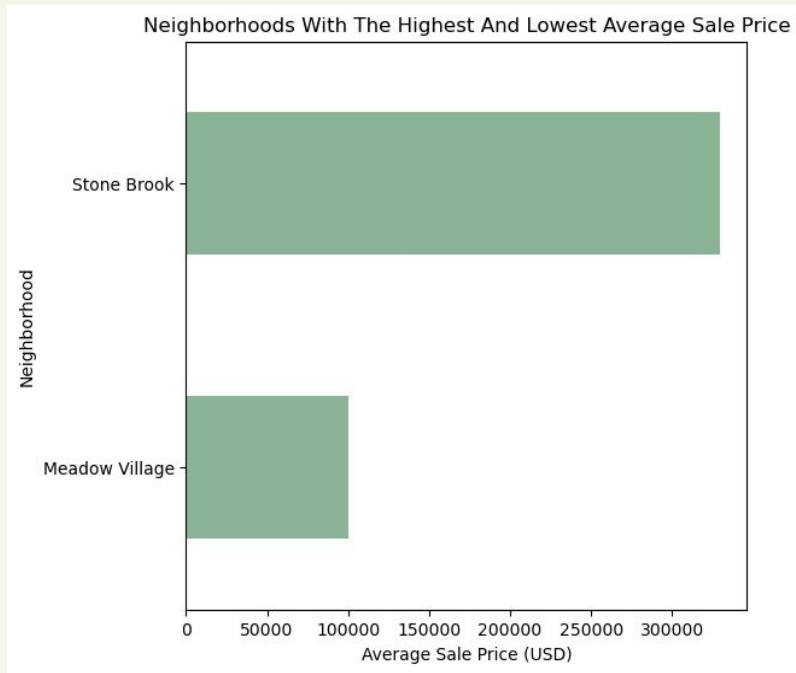
JG **Joe Gomez**
6/3/2022 · 8 min read

Neighborhood	Location	Home Size	Age & Condition
Upgrades & Updates	The Local Market	Economic Indicators	Interest Rates

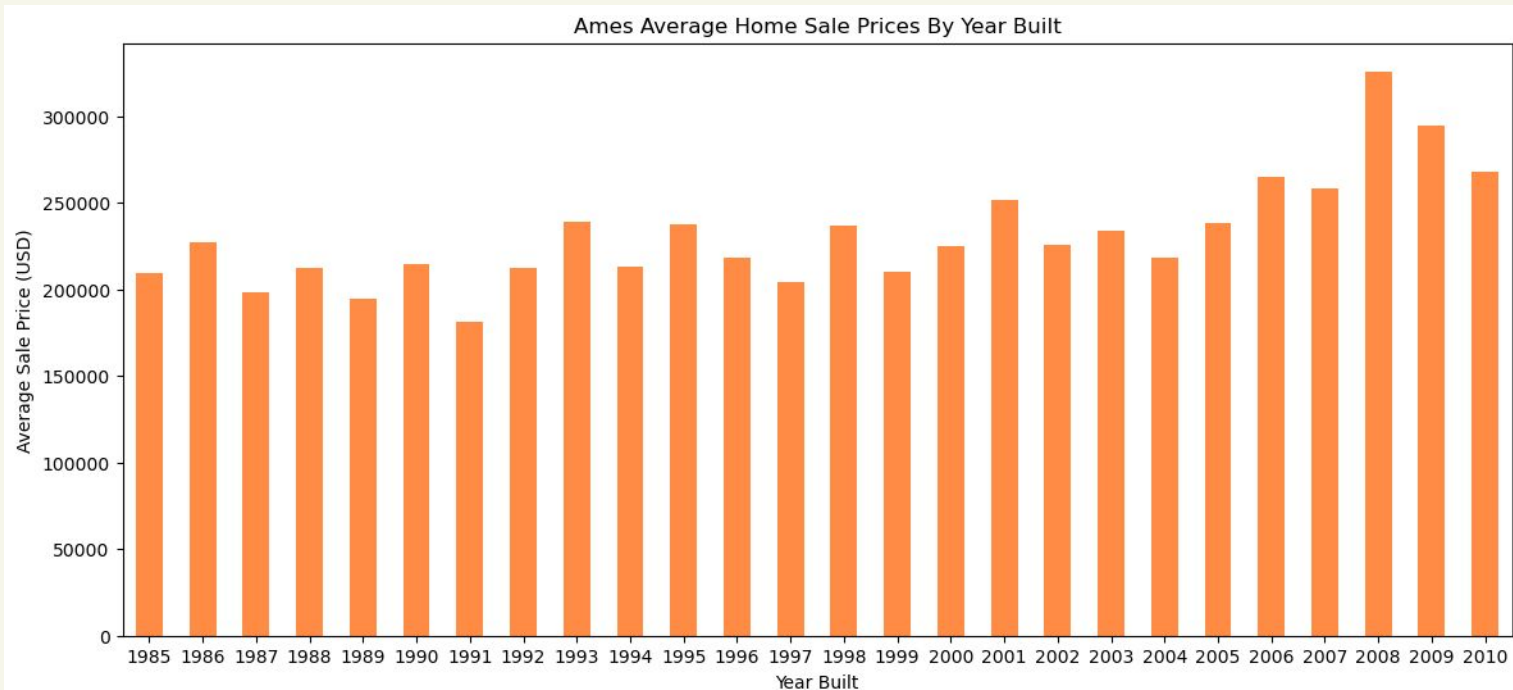# Neighborhood is one of a key factor influencing the sale price

Neighborhoods With The Highest And Lowest Average Sale Price

Top 3 Neighborhoods With The Highest Average Sale Price

As home size gets larger, sale price also goes up 📊

Home Size vs Home Sale Price
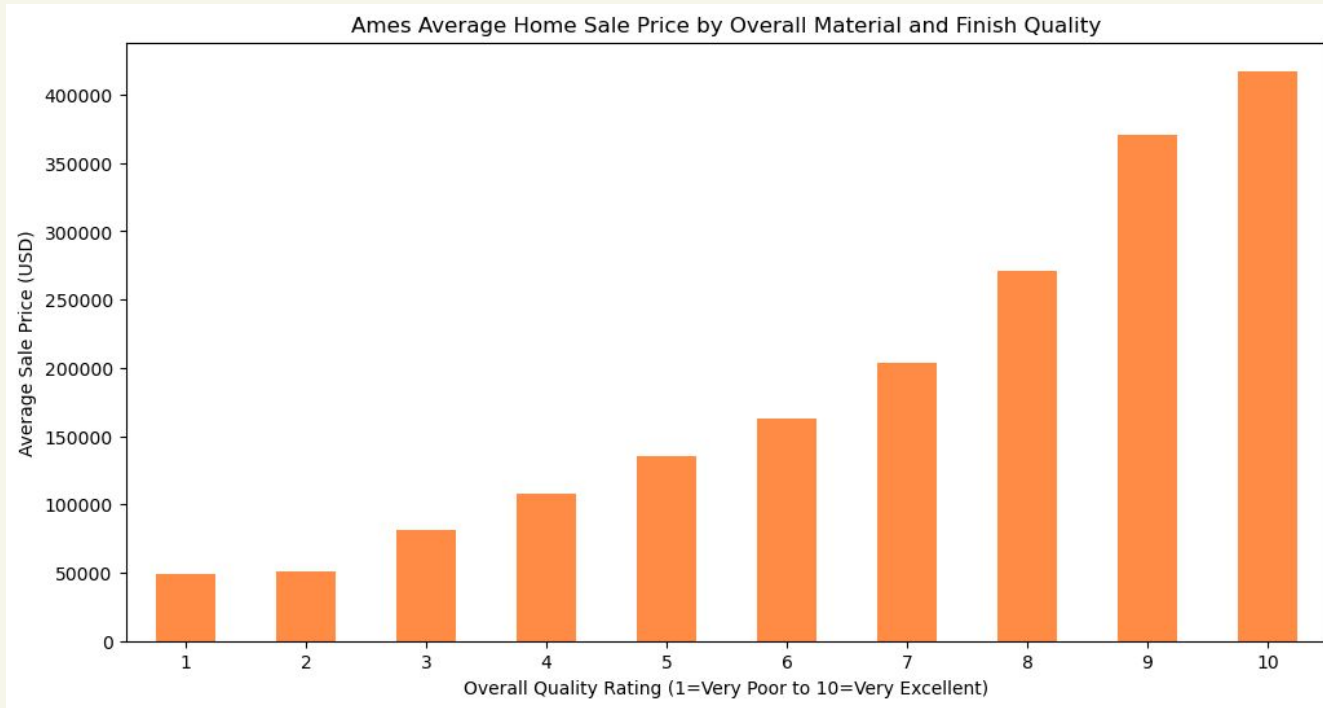
The sizes of outdoor spaces and bathrooms also impact sale price

Newer homes are sold at higher prices

Ames Average Home Sale Prices By Year Built
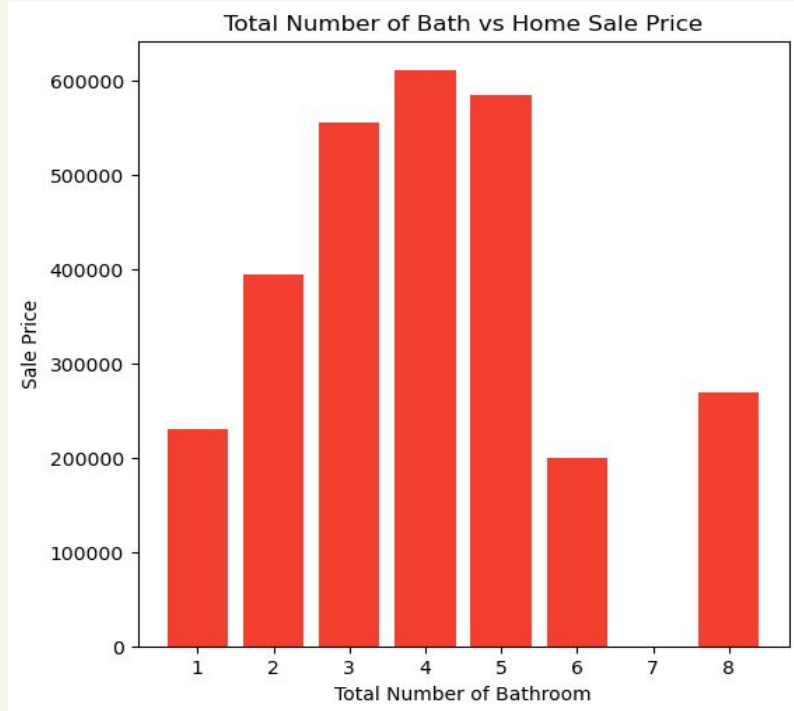
A move-in-ready home are sold at higher prices

Ames Average Home Sale Price by Overall Material and Finish Quality

Updates & Upgrades can add value to a home

Ames House Average Sale Price by Remodel Year

Home Value Comparison by Quality of Basement
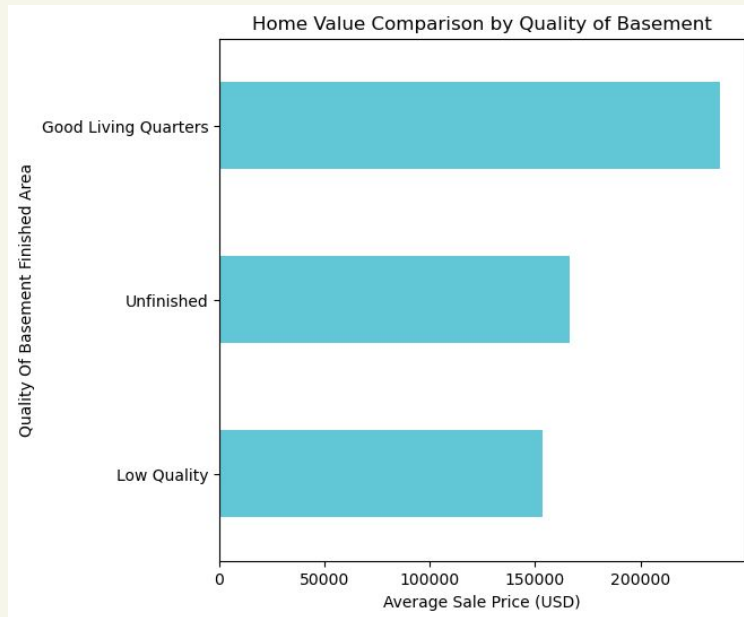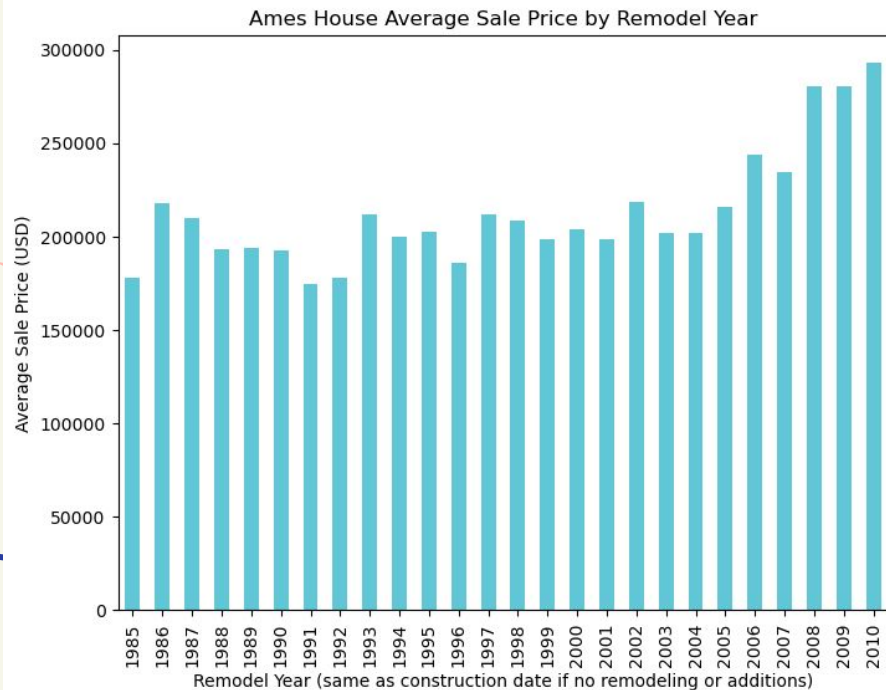
EDA takeaways

★ **Neighborhood & Location**:
  ○ There is a significant difference in sale price by neighborhood.
  ○ Northridge Heights, Northridge, and Stone Brook saw the highest average sales price of a home in the city of Ames, IA.

★ **Home Size & Usable Space**:
  ○ Home size positively impacts a home sale price.

★ **Age and condition**:
  ○ In general, newer homes are sold at higher prices.
  ○ Also, great quality of the house (like a move-in-ready home) has a higher sale price.

★ **Upgrades & updates**:
  ○ Recent modeling and addition also bring additional value to a home.
  ○ A house with a good living quality of basement could add more than 40% higher sale price than unfinished or low quality basement.

Models using different feature matrix were built and evaluated

Matrix features: a correlation coefficient with 'saleprice'	Training R-Squared Score	Test R-Squared Score	RMSE	Cross validation Score (cv=5)
equal to or greater than 0.5/-0.5	0.821	0.895	24,241.895	0.839, 0.879, 0.741, 0.866, 0.782
equal to or greater than 0.25/-0.25	0.868	0.904	23,088.501	0.887, 0.891, 0.777, 0.893, 0.825
equal to or greater than 0.15/-0.15	0.88	0.914	21,917.108	0.890, 0.900, 0.800, 0.886, 0.839

Best performing LR model

Feature Matrix (X):
Column features that have a correlation coefficient equal to or greater than 0.15/-0.15 with target variable (y), "sale price"

Training R-Squared Score
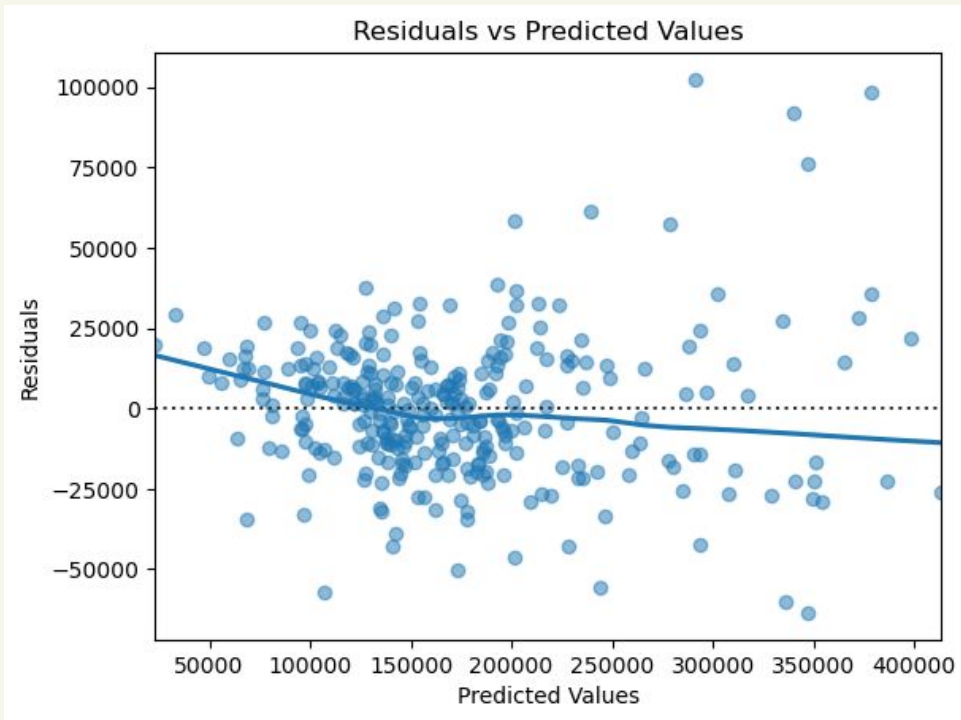0.88

Test R-Squared Score
0.914

RMSE
$21,917.108

Cross Validation Score (cv=5)
0.890, 0.900, 0.800, 0.886, 0.839

Residuals vs Predicted Values

Conclusion

★ Based on our analysis, we have developed a model that can accurately predict home prices in Ames, IA, with an R-squared score of 0.914.

★ Our analysis identified several key factors that have a significant influence on the market value of homes in the area, including the overall quality, the garage area, and the year the house was built.

★ By taking these factors into account, our model can provide local real estate agents with more accurate pricing information, enabling them to offer more informed guidance to their clients.

Recommendations

Here are some suggestion to improve our model:
★ **Feature engineering**: There are some features related to garage. A new interaction column could be created by combining the features related to garage.
★ **Regularization**: Lasso and Ridge regularization can be tested. Also StandardScaler could be used.
★ **Model selection**: Other models such as Random Forest, Gradient Boosting and Neural Networks could be tested.

Thank you!

Any questions?