# Ames Predicting Property Price

Prepared for : Management of BestEstate Co.

# Table of contents

# 01

# Introduction & Problem Statement

# Introduction

BestEstate Co
Web/App usage: 5.27 session by day. 158.1 sessions by month
Web/App success: 1.27 success session by month
Web/App price prediction accuracy : Around 81%

Goal: To develop a regression model that will make accurate predictions of house prices in the city of Ames in Iowa.

1. Homeowners can more accurately determine the asking price at which to list their property. They won't undervalue their home in this manner and miss out on possible profit.
2. Vice versa, homebuyers can avoid unintentionally making overpriced purchases by knowing what is a fair price to pay for a specific house.

# Problem Statement
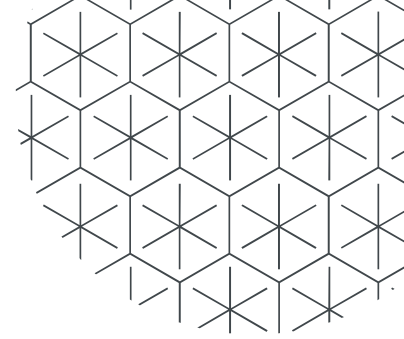
1. To determine the features that can be improved to drive up prices.

2. To pinpoint the features that have a negative impact on prices.
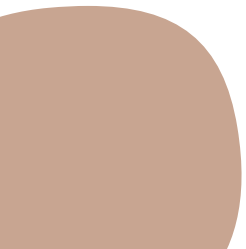
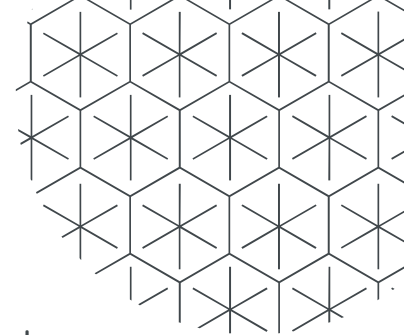# 02

# EDA

# Ames Housing Dataset

- 80 columns and over 2000 rows of different features relating to houses

- Used by the Ames Assessor's Office to compute assessed values for individual residential properties sold in Ames, IA from 2006 to 2010.
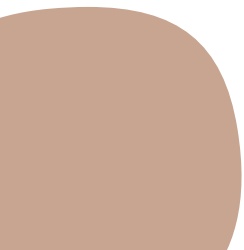
# Data Cleaning for columns with >5% missing values

| column name | no of null rows | % of missing values |
| --- | --- | --- |
| pool_qc | 2042 | 99.56 |
| misc_feature | 1986 | 96.83 |
| alley | 1911 | 93.17 |
| fence | 1651 | 80.50 |
| fireplace_qu | 1000 | 48.76 |
| lot_frontage | 330 | 16.09 |
| garage_yr_blt | 114 | 5.56 |
| garage_cond | 114 | 5.56 |
| garage_qual | 114 | 5.56 |
| garage_finish | 114 | 5.56 |
| garage_type | 113 | 5.51 |

# Sale Prices are mostly in the 150-200k range
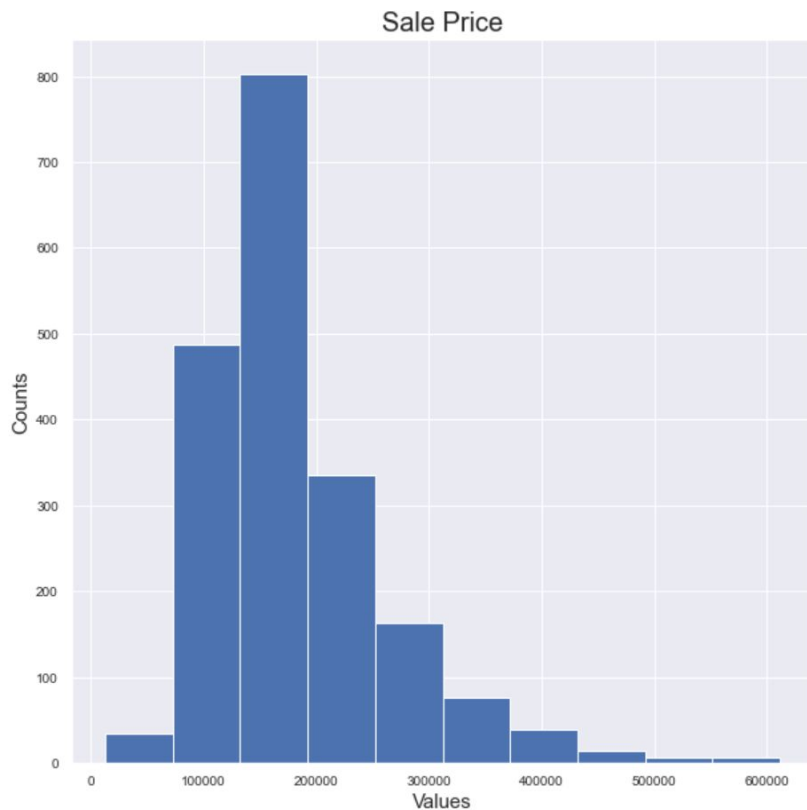


Sale Price

Top 10 Features correlating with Sale Price
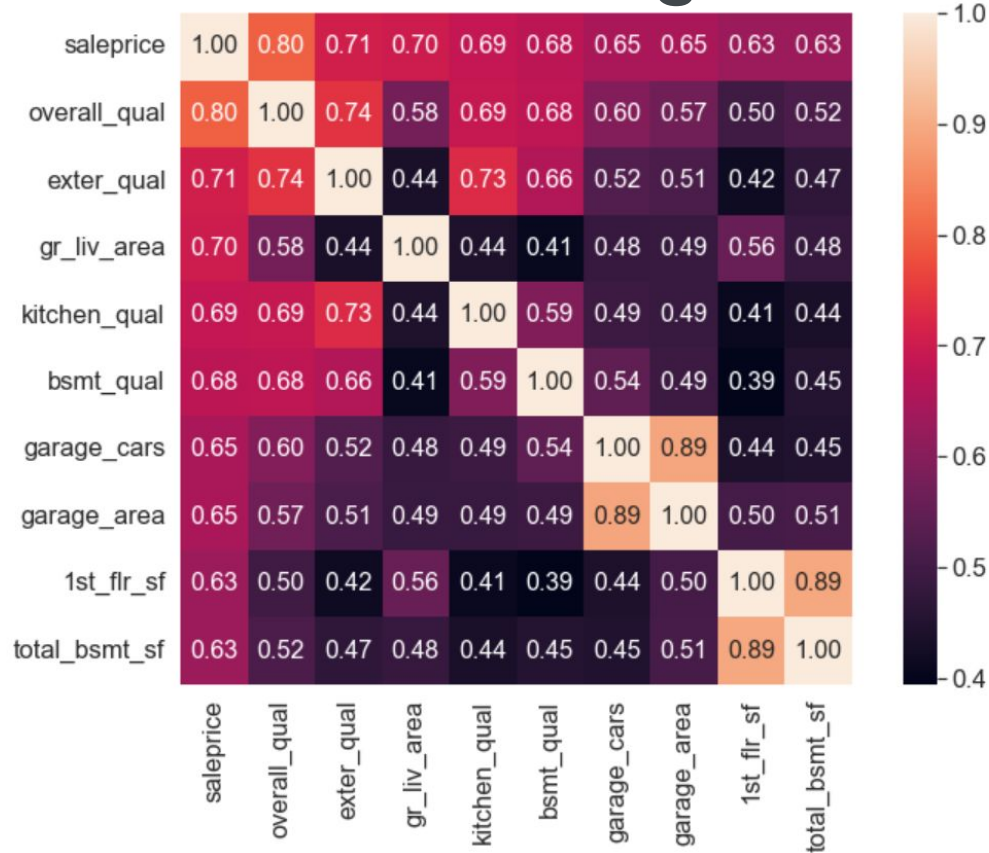
# Most of Ames houses have mid-range Overall Quality score
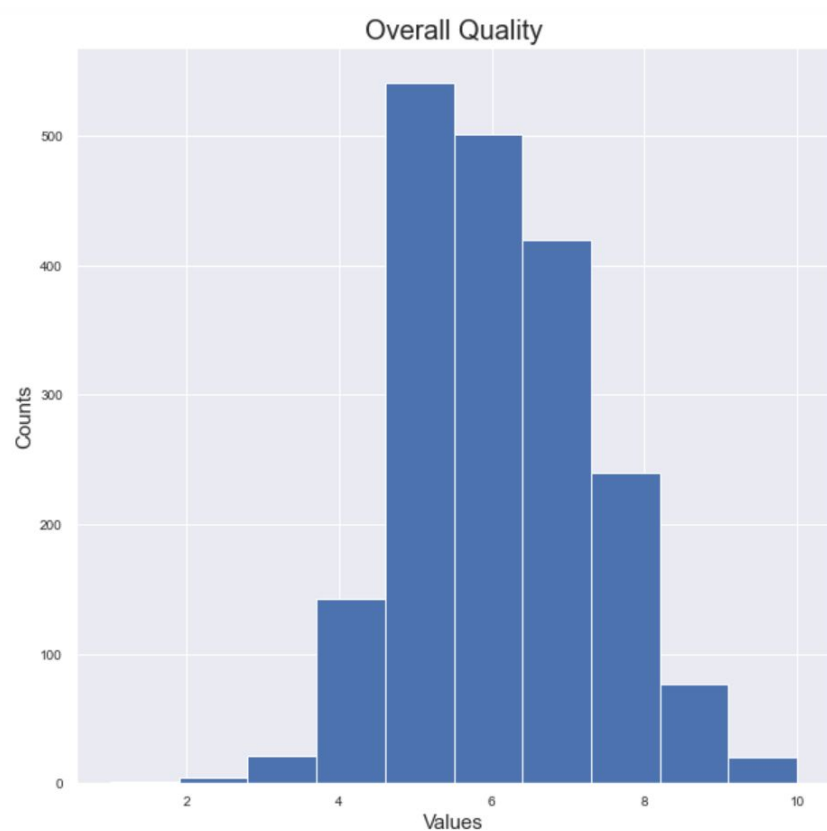
# Sale Price increases with Overall Quality
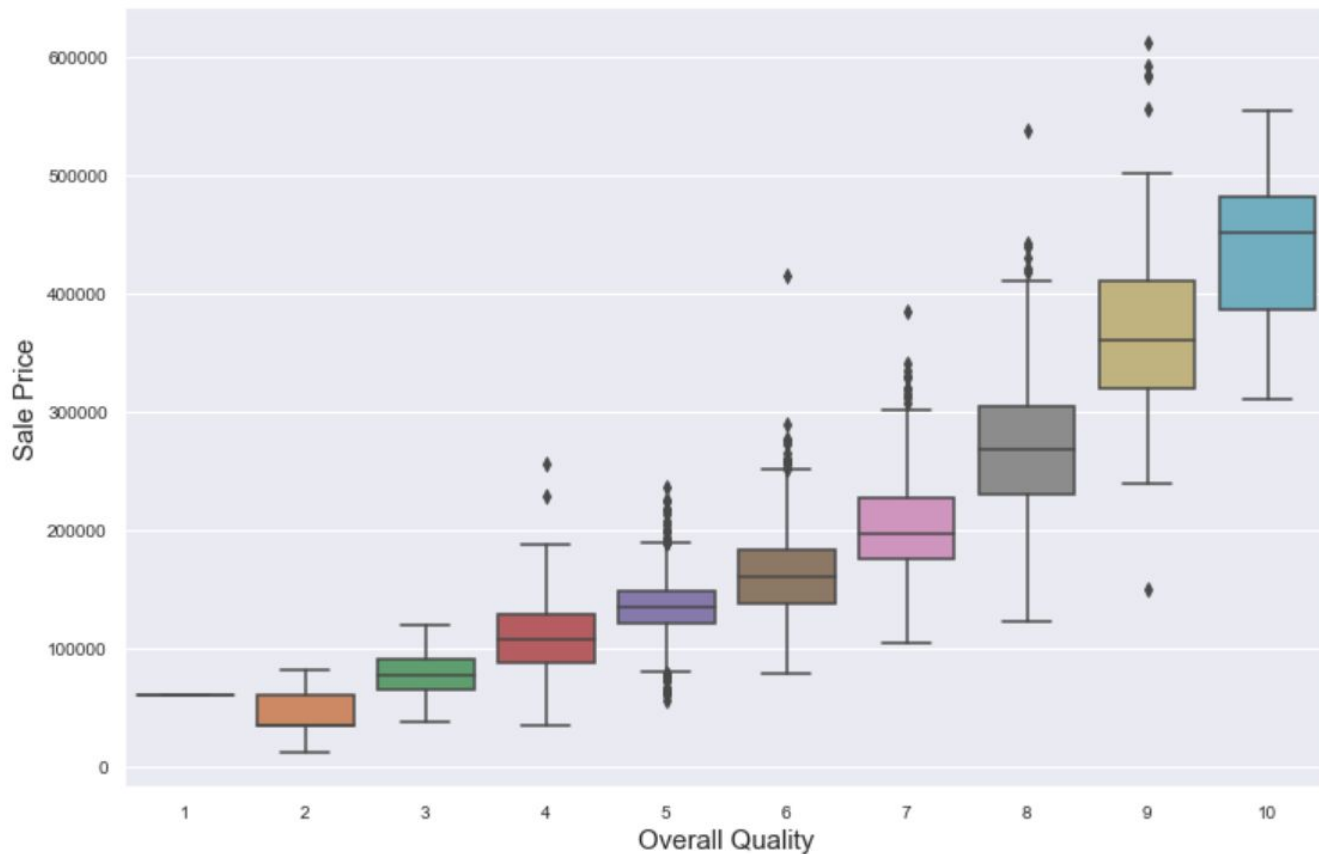
# In general, Sale Price increases as Above Grade Living Area increases



Above grade living area square feet



Above grade living area square feet vs Sale Price
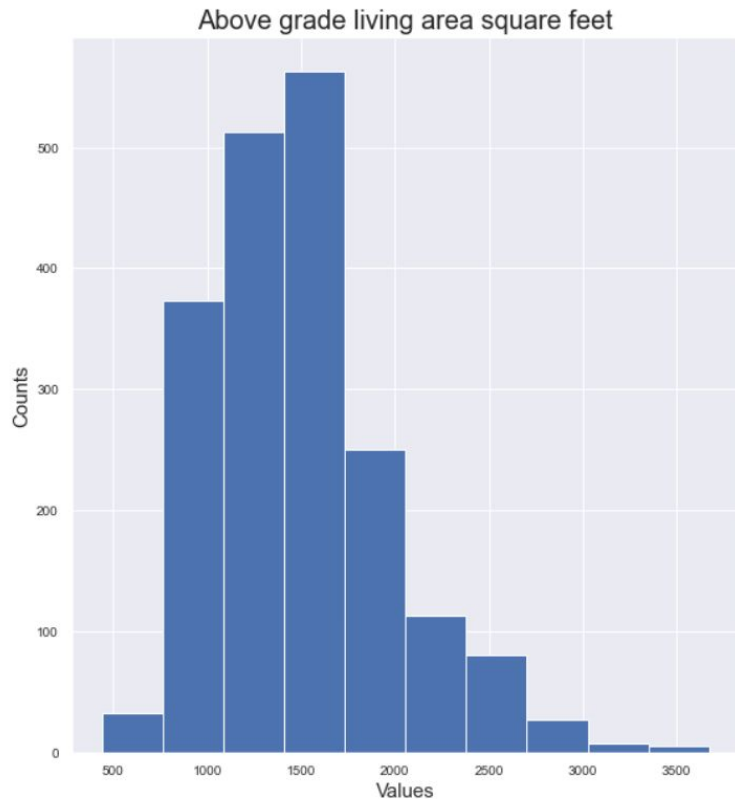
# Most houses have space for 2 cars but we see higher range of sale prices for houses with 3 cars garage space.
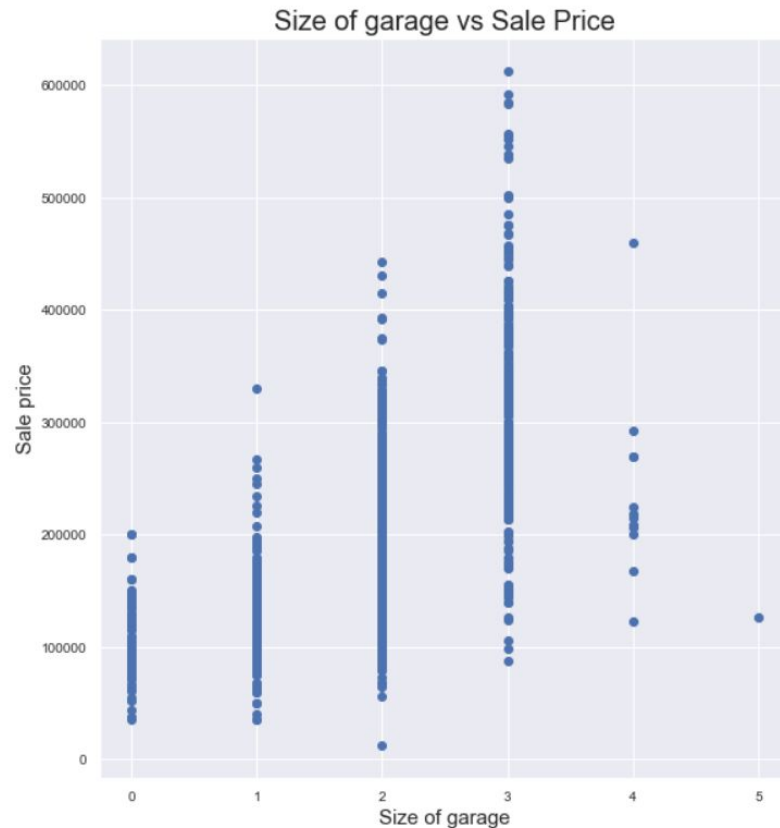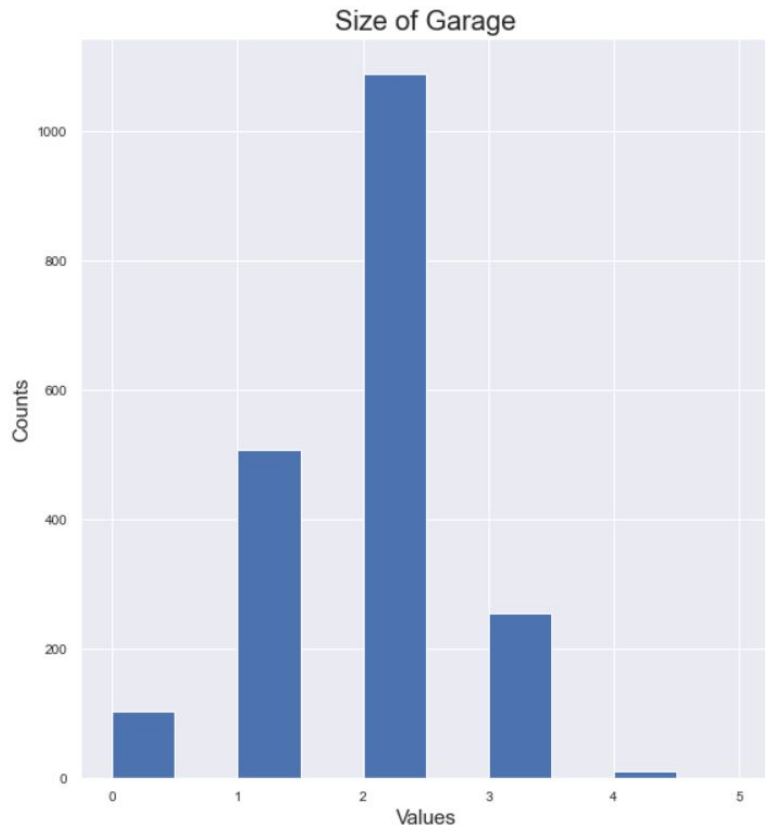
# 03
# Model Selection

# Model Test Results

| Model | Test Results |
|---|---|
| Baseline Model | −0.00115 |
| Linear Regression | 0.89199 |
| Ridge Regression | 0.89355 |
| Lasso Regression | 0.89233 |

- **Ridge Regression** is the best performing model!
- This is in line with our understanding of machine learning models.

# Interpretation of Model Coefficients

- Positive coefficient means increase in a variable will cause the sale price to increase.

- Negative coefficient means increase in a variable will then cause the sale price to decrease.

- Value signifies the impact a change in a variable will have on the sale price.

# Model Coefficients



| | |
|---|---|
| Abv Grd Liv Area (Log) | 0.18 |
| Overall Quality | 0.15 |
| Overall Condition | 0.07 |
| Bsmt Finished Area - 1st | 0.06 |
| 1st floor in sqft (Log) | 0.05 |
| Lot Area | 0.05 |
| Foundation - Concrete | 0.05 |
| Garage Cars | 0.05 |
| Bsmt Quality | 0.05 |
| Remodel Date | 0.04 |
| Home Functionality Rating | 0.04 |
| Paved Driveway | 0.03 |
| Type of Sale - New | 0.03 |
| Screen Porch Area in sqft | 0.02 |
| Exterior - Brick Face | 0.02 |
| Bsmt Exposure | 0.02 |
| NBH - Northridge Heights | 0.02 |
| Bsmt Full Bathrooms | 0.02 |
| Heating - Gravity Furnace | -0.02 |
| NBH - Edwards | -0.03 |
| NBH - OldTown | -0.03 |
| Zone - Resid Med Density | -0.04 |
| Zone - Commercial | -0.04 |

# % Increase for Every 1 Unit Increase in Variables



Bar chart titled "% Increase for Every 1 Unit Increase in Variables". The y-axis is labeled "% Increase" ranging from -5.00% to 20.00%. The x-axis is labeled "Variables".

| Variable | % Increase |
|---|---|
| Abv Grd Liv Area | 17.80% |
| Overall Quality | 11.27% |
| Overall Condition | 4.90% |
| Bsmt Finished Area - 1st | 4.54% |
| 1st floor in sqft | 3.81% |
| Lot Area | 3.67% |
| Foundation - Concrete | 3.53% |
| Garage Cars | 3.24% |
| Bsmt Quality | 3.17% |
| Remodel Date | 2.81% |
| Home Functionality | 2.60% |
| Paved Driveway | 2.24% |
| Type of Sale - New | 1.89% |
| Screen Porch Area in sqft | 1.75% |
| Exterior - Brick Face | 1.61% |
| Bsmt Exposure | 1.40% |
| NBH - Northridge Heights | 1.33% |
| Bsmt Full Bathrooms | 1.12% |
| Heating - Gravity Furnace | -1.58% |
| NBH - Edwards | -1.79% |
| NBH - OldTown | -1.79% |
| Zone - Resid Med Density | -2.53% |
| Zone - Commercial | -2.53% |

# Interpretation of Model Coefficients

- Average Sale Price = $183,000

| Variable | Coefficient values | % ↑ for every 1 Unit ↑ in Variable | % ↑ in Monetary Value |
|---|---|---|---|
| Abv Grd Liv Area (Sqft) | 0.178 | 17.80% | $33,000 |
| Overall Quality | 0.154 | 11.27% | $21,000 |
| Overall Condition | 0.069 | 4.90% | $9,000 |
| Zone - Resid Med Density | -0.037 | -2.53% | -$5,000 |
| Zone - Commercial | -0.037 | -2.53% | -$5,000 |

# 04
# Limitations & Improvements

# Limitations

Some inputs have unclear ratings:

- Condition, quality
    - Excellent, Good, Average, Fair

Multicollinearity:

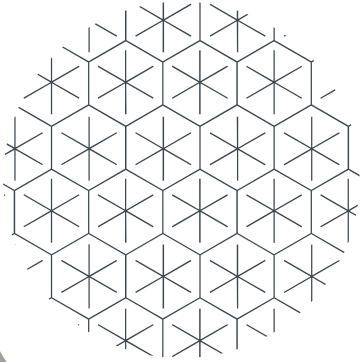- Corr threshold > 0.7

External Factors:

- eg. black swan events

# Future Improvements

- Uncover new valuable inputs

- Test different categorical groupings

- Hyperparameter Tuning

# Conclusion

Based on test results,

- Our model can predict 89.36% of the changes in sale prices accurately with the changes in inputs

Characteristics most impactful on sale price:

- 'Above grade living area square feet'
- Overall material & finish quality

Model accuracy ≈ Quality of Data

# 06

## Recommendations

# Recommendations

## Value-add to current product/service offering

- Buyers & Sellers should make informed decision
  - Utility
  - Investment

## Valuation is only a benchmark

- 3 valuation methods
  - Sales Comparison
  - Cost
  - Income Capitalisation
- "10 different valuers, 10 different valuations"