Regression Analysis for Housing Prices in Ames, Iowa

By:

Eugene Khoo, Florian Combelles, Joanne Chong,
Kenneth Goh, Tan Ming Jie

Agenda

- Background
- Problem Statement
- Existing Prediction Model
- Exploratory Data Analysis
- Recommended Prediction Model
- Conclusion and Recommendations

1.
Background

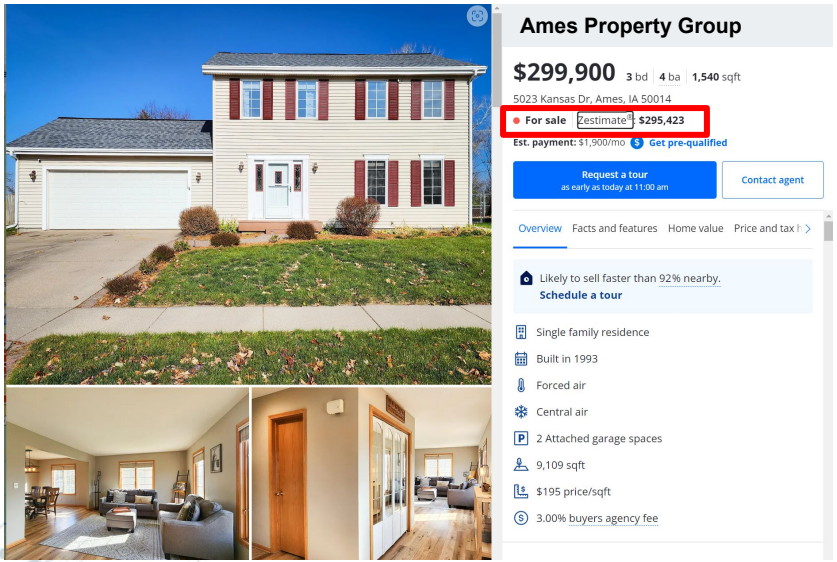
Who we are and what we do

Who we are

- Reputable real estate agency established in 1990
- Top agency in Ames, Iowa for achieving high sales volume in the last 20 years
- Well-known for our proprietary ***Prop-Price Estimator ©***
  - *What it is:* A prediction model that provides customers with best estimated price for their property

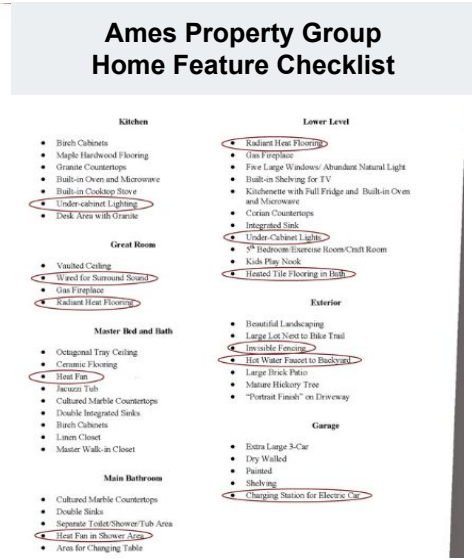
How does Prop-Price Estimator© work?

Paid buyer's access



Data collection

(For price estimation)



Ames Property Group
Home Feature Checklist

Kitchen
- Birch Cabinets
- Maple Hardwood Flooring
- Granite Countertops
- Built-in Oven and Microwave
- Built-in Cooktop Stove
- Under-cabinet Lighting
- Desk Area with Granite

Great Room
- Vaulted Ceiling
- Wired for Surround Sound
- Gas Fireplace
- Radiant Heat Flooring

Master Bed and Bath
- Octagonal Tray Ceiling
- Ceramic Flooring
- Heat Fan
- Jacuzzi Tub
- Cultured Marble Countertops
- Double Integrated Sinks
- Birch Cabinets
- Linen Closet
- Master Walk-in Closet

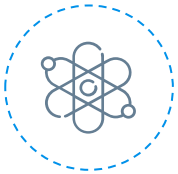
Main Bathroom
- Cultured Marble Countertops
- Double Sinks
- Separate Toilet/Shower/Tub Area
- Heat Fan in Shower Area
- Area for Changing Table

Lower Level
- Radiant Heat Flooring
- Gas Fireplace
- Five Large Windows/ Abundant Natural Light
- Built-in Shelving for TV
- Kitchenette with Full Fridge and Built-in Oven and Microwave
- Corian Countertops
- Integrated Sink
- Under-Cabinet Lights
- 5th Bedroom/Exercise Room/Craft Room
- Kids Play Nook
- Heated Tile Flooring in Bath

Exterior
- Beautiful Landscaping
- Large Lot Next to Bike Trail
- Invisible Fencing
- Hot Water Faucet to Backyard
- Large Brick Patio
- Mature Hickory Tree
- "Portrait Finish" on Driveway

Garage
- Extra Large 3-Car
- Dry Walled
- Painted
- Shelving
- Charging Station for Electric Car

Importance of accurate price prediction



Strategic Pricing

- Price estimation based on most sought-after property features
- Maximise seller profit based on market demand

Listing Duration

- Accurate price lead to increased interest and faster sales
- Reduce listing management time
- Decrease buyers' bargaining power

Increase Engagement

- Remain go-to agency for clients
- Generate trust in our data-driven technology

2.
Problem Statement

The challenge at hand

Problem statement & objective

- Increasingly competitive real estate market
  - Competitors offer higher-accuracy predictions
- Shrinking market share
  - More customers looking for better predictions

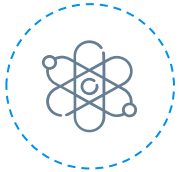
- Data scientists tasked to improve prediction accuracy
  - Enhance existing model or develop a new one
- Model must beat top competitors

3.

Existing Prediction Model

The science behind the Prop-Price Estimator©

What we are currently using



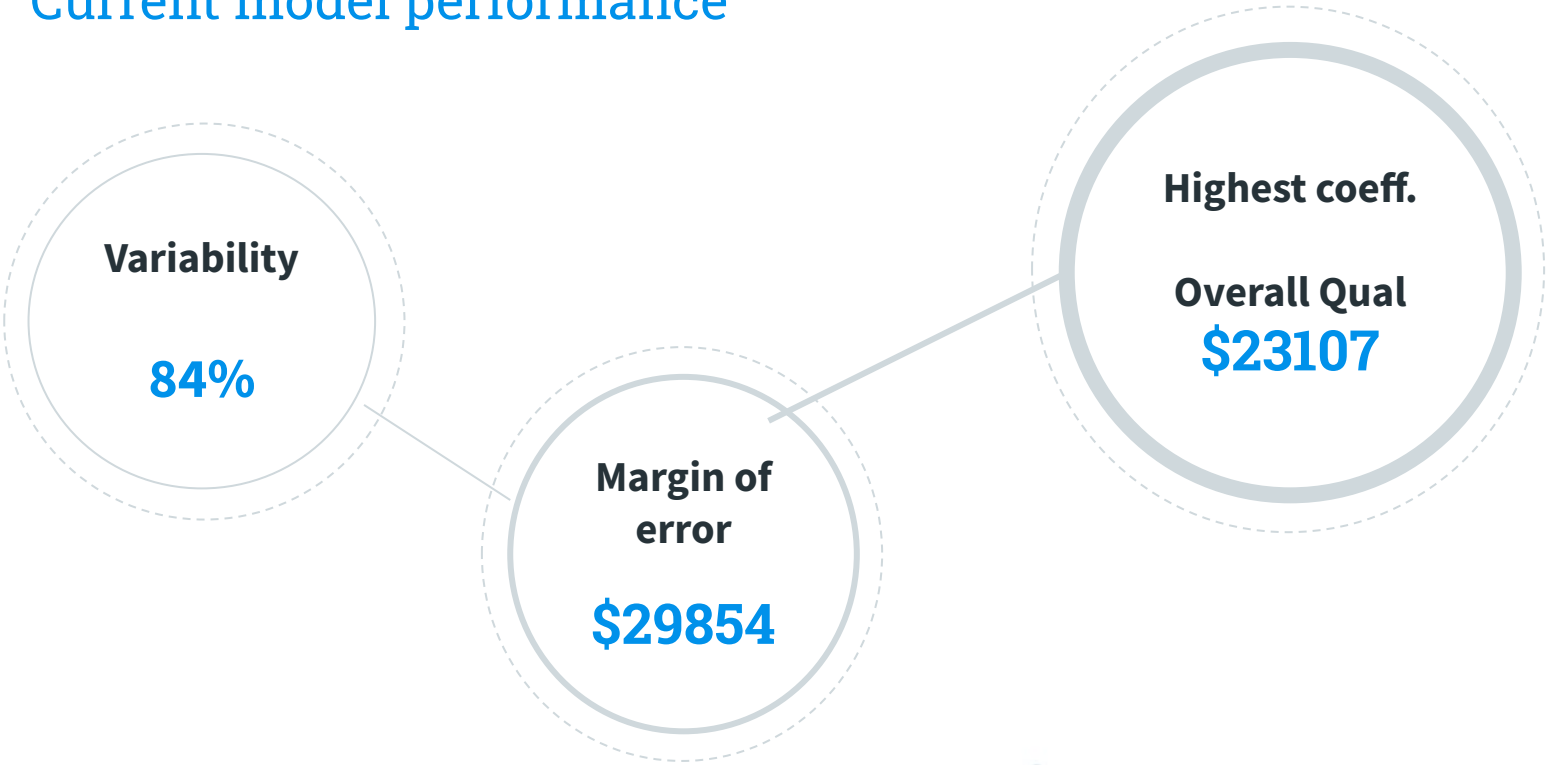
Ridge Regression Model

- Designed by external vendor
- Deployed 2 years ago
- Retrains with new data every two months

Key Variables

- Overall Material Quality
- Living Area size
- Neighborhood
- 1st Floor area
- Garage Area
- Basement Quality

Current model performance



Why we need to upgrade our model

External Factors

Market competition

- Rising number of competitors
- Shift of focus toward prediction accuracy
- Competitors error margin is 27.000$

Increasing customer demands

- For more accurate selling price
- Fast and simple prediction
- Better recommendations to optimize property value

Why we need to upgrade our model

Internal Factors

Brand

- Reputation as number 1 predictor in the market
- Defend market leader position in the housing market
- Client trust for over two generations

Cost management

- Run time of model
- Memory usage
- Better recommendation support for our agents

4.

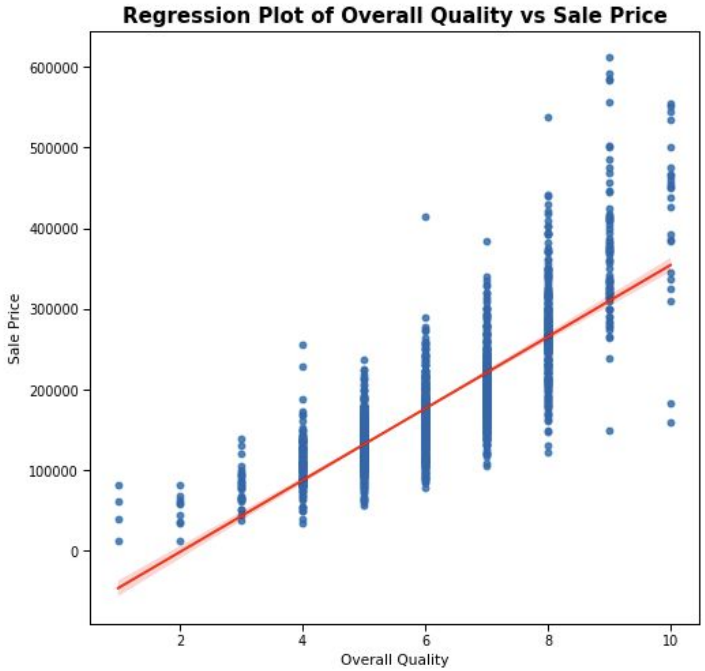
Exploratory Data Analysis

Deriving insights through new data set

Ground Living Area positively correlated with Sale Price^(⅓)

- The ground living area is 1,499 square feet
- Most ground living area range from 1,129 to 1,728 square feet

Overall Quality positively correlated with Sale Price



- The average overall quality is 6.11
- Average sale price for top, mid and low quality:
  - Rated 10: $417,397
  - Rated 6: $134,964
  - Rated 1: $28,725

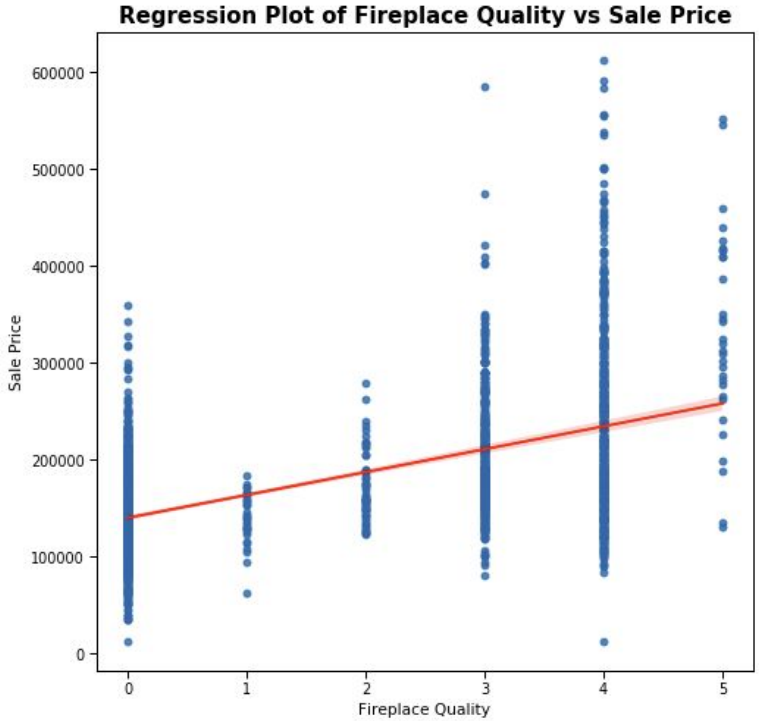
Kitchen Quality positively correlated with Sale Price

Ordinal values were scaled from 0-4

- The average housing prices based on kitchen quality:
  - Excellent - $336,424
  - Good - $211,639
  - Typical/Average - $139,502
  - Fair - $101,335
- Better-rated kitchen will result in higher price

Fireplace Quality positively correlated with Sale Price

Ordinal values were scaled from 0-5

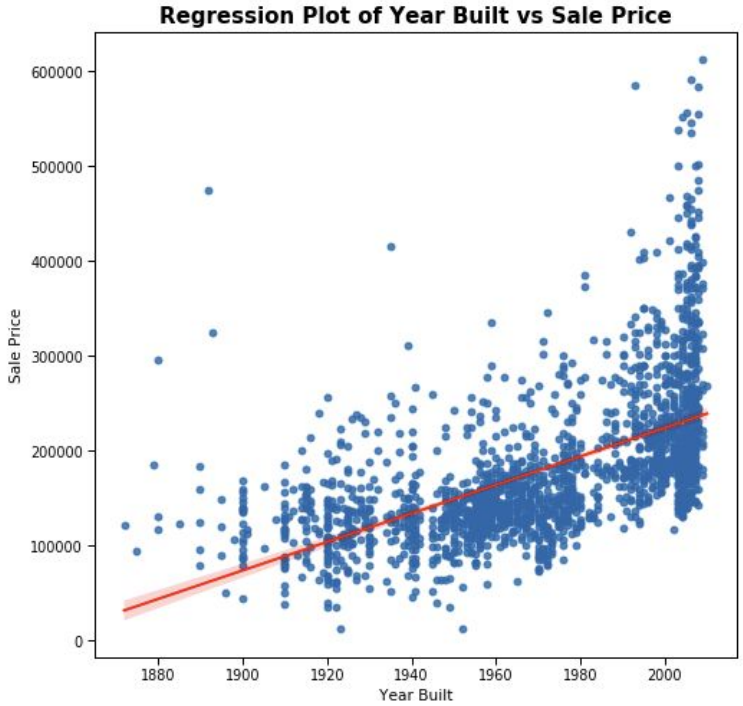


- The average housing prices based on kitchen quality:
  - Excellent - $331,481
  - Good - $233,578
  - Typical/Average - $204,894
  - Fair - $170,287
  - Poor - $139,264
- Better-rated fireplaces will result in higher price

Garage Cars positively correlated with Sale Price

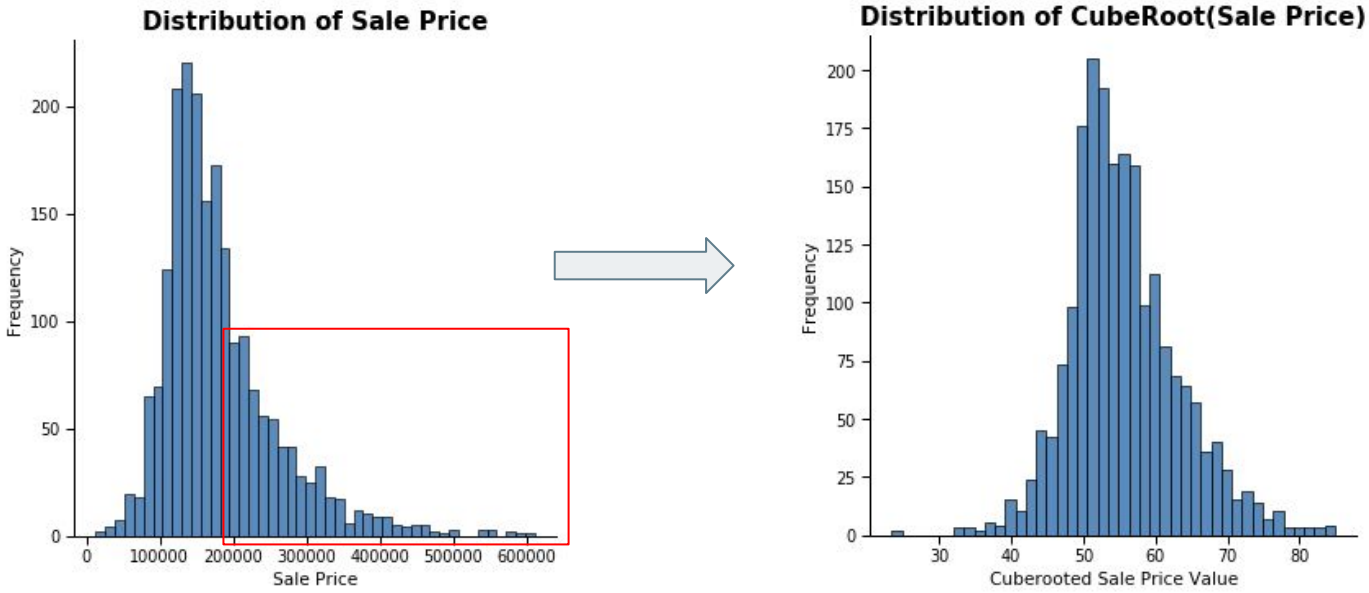
- The average housing prices based on number of cars in a garage:
  - 5 cars - $126,500
  - 4 cars - $229,653
  - 3 cars - $309,383
  - 2 cars - $184,037
  - 1 car - $126,916
  - No garage - $105,669
- Being able to fit more cars in a garage will result in higher price

Newer built homes have higher Sales Prices



- The average sale price of oldest homes:
  - 1872 - $122,000
  - 1875 - $94,000
  - 1879 - $185,000
- The average sale price of newer homes:
  - 2010 - $267,916
  - 2009 - $294,460
  - 2008 - $326,057

Applying cube root resolves right skewness of sale price distribution



The average sale price of a house in Ames, Iowa is $181,469

The smallest transaction was $79,259, while the largest price sold was $611,657

5.

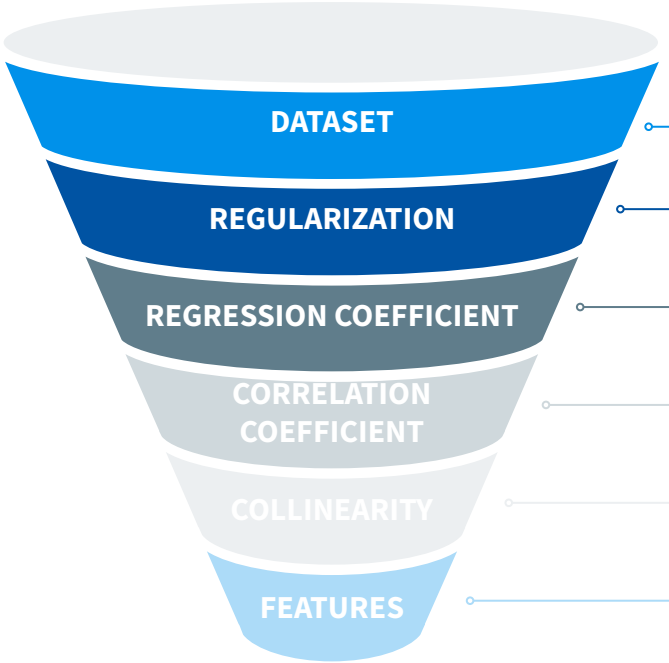
Recommended Pricing Model

How we can regain our edge

LINEAR REGRESSION
LIGHTYEAR
LINEAR REGRESSION EVERYWHERE
imgflip.com

Selected Features

Which features we want to use



183 columns

Lasso Regularisation

Highest regression coefficient (E.g. Overall Quality & Overall Condition)

Highest correlation coefficient (Selected features have average correlation coefficient of 0.5)

Removing features with collinearity above 0.7

Final selection of 13 Features

Selected housing features

5
Ordinal

- Overall Quality
- Overall Condition
- Kitchen Quality
- Fireplace Quality
- Basement Exposure

7
Numerical

- Year Built
- Misc Value
- Total Basement Area
- No. of cars the garage can fit
- Type 1 finished sf
- Abv ground living area
- Lot Area

1
Categorical

- Neighborhood (NorthRidge Heights)

Evaluation of Models

How we'll be measuring performance

High R2 Score
0 - 100%

+

Low Root Mean Squared Error
($) Value

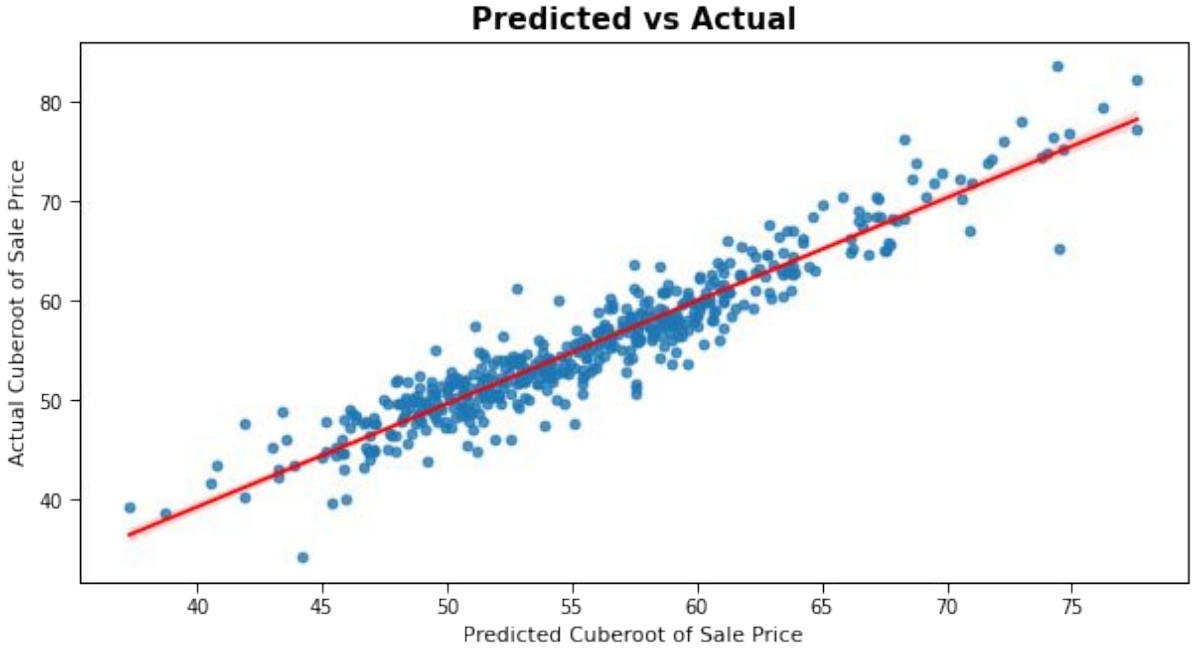
+

Optimal No.of Features

How our new model is faring

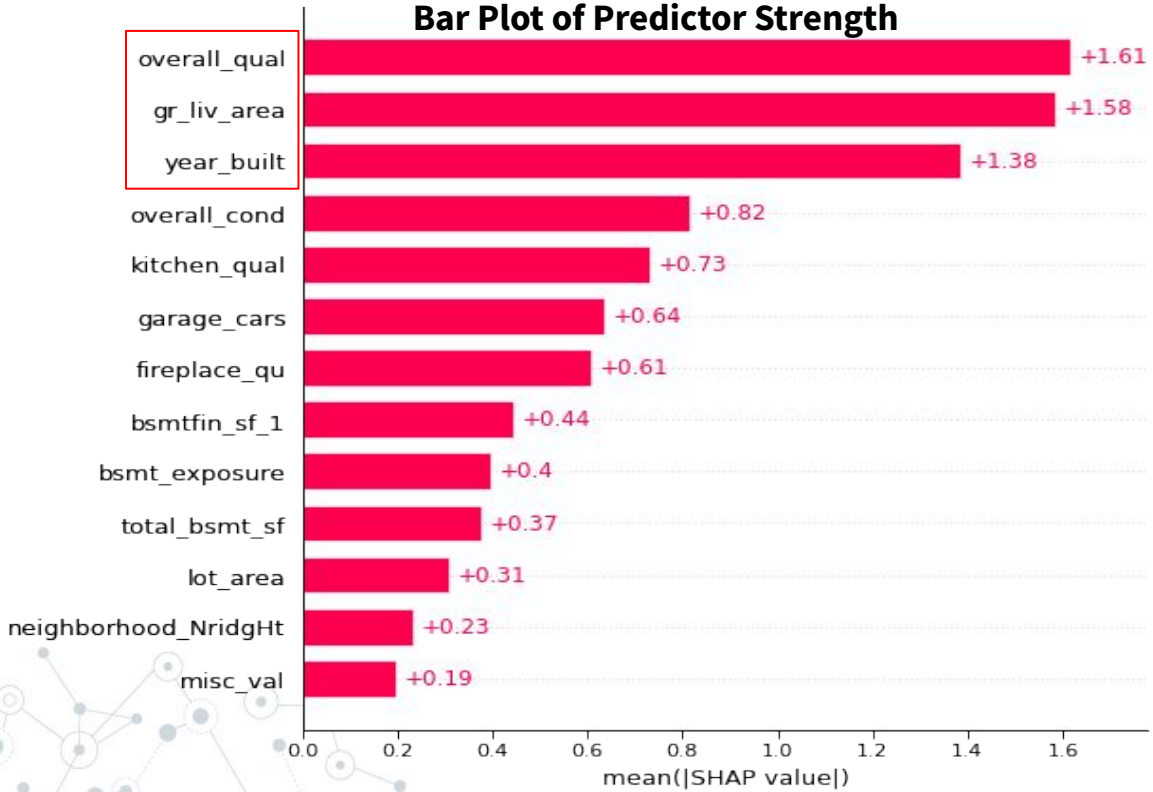
	TRAINING SET R^2 SCORE (%)	TEST SET R^2 SCORE (%)	MARGIN OF ERROR ($)	# OF HOUSING FEATURES
Current Model	83.22	83.75	29,854	6
Market Best	84.12	85.57	27,000	17
New Model	86.96	89.55	24,488	13

Predictive capabilities and reliability



Predicted values are within close proximity of actual values

Influence of each feature on pricing



Top 3 Predictors

1. Overall Quality
2. Above ground living area (Sq. ft.)
3. Year Built

6.

Conclusion & Recommendations

Summing up our findings

In a nutshell

Current Price Predictors

- ❑ Overall Material Quality
- ❑ Living Area size
- ❑ Neighborhood
- ❑ 1st Floor area
- ❑ Garage Area
- ❑ Basement Quality

With our new prediction model

New Set of Variables

- ❑ Overall quality
- ❑ Living area size
- ❑ Year built
- ❑ Overall condition
- ❑ Kitchen quality
- ❑ Capacity of garage
- ❑ Fireplace quality
- ❑ Basement size
- ❑ Basement exposure / access
- ❑ Lot area
- ❑ Neighborhood

Black: Existing variables
Blue: New additions

Prediction model limitations

Sensitivity

Model sensitivity to outliers whenever there's huge difference in features and/or price

Representation

Unrepresentative data from each neighbourhood

Overlaps

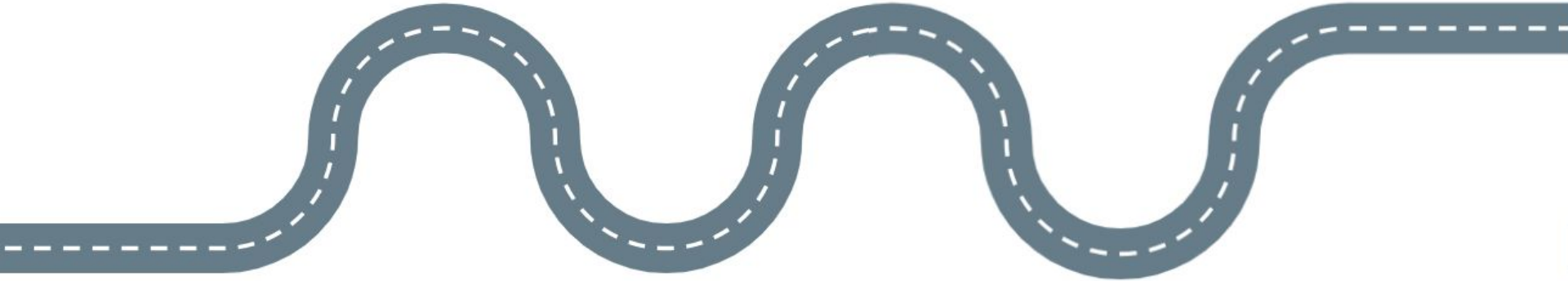
Overlapping characters and input for categorical variables, e.g. quality ratings "Average" and "Fair" sound similar

External Factors

Economical push that influence house prices such as the 2008 housing crisis in the US

Building more accurate prediction model

1. Counsel homeowners to improve property's market value through the top predictors
2. Ensure all home listings feature the top predictors
3. Further investigate features and ensure representation in data collection
4. Better define characteristics for categorical variables



Thanks!

Any questions?

You can find us at:

intelligence@ameshousegrp.com

7.
Appendix

Lot Area positively correlated with Sale Price

- The average lot area is 10,065 square feet
- Most lot area range from 7,500 to 11,514 square feet
- As lot size increase, the price increases as well

Total Basement Area increases with Sale Price

- The average basement area is 1,057 square feet
- Most basement area range from 793 to 1,318 square feet