Ames, Iowa

By: Alec Edgecliffe-Johnson

TABLE OF CONTENTS

01 **PURPOSE**
Problem statement - what and why

02 **DATA**
Introduction to the data

03 **METHODOLOGY**
Our approach to analyzing the data

04 **FINDINGS**
Key findings from our analysis

05 **RECOMMENDATIONS**
Recommendations and further steps based on our findings

OUR PURPOSE

To help local property developers in Ames, Iowa make better choices about how and where they construct houses in order to keep development local.

INTRODUCTION TO THE DATA

1. HOUSE CHARACTERISTICS

Features innate to the house like square footage, kitchen size and quality, fireplaces, storeys etc.

2. ADD-ONS TO THE HOUSE

Features not directly related to the main living area like decks, basement size, garage space etc.

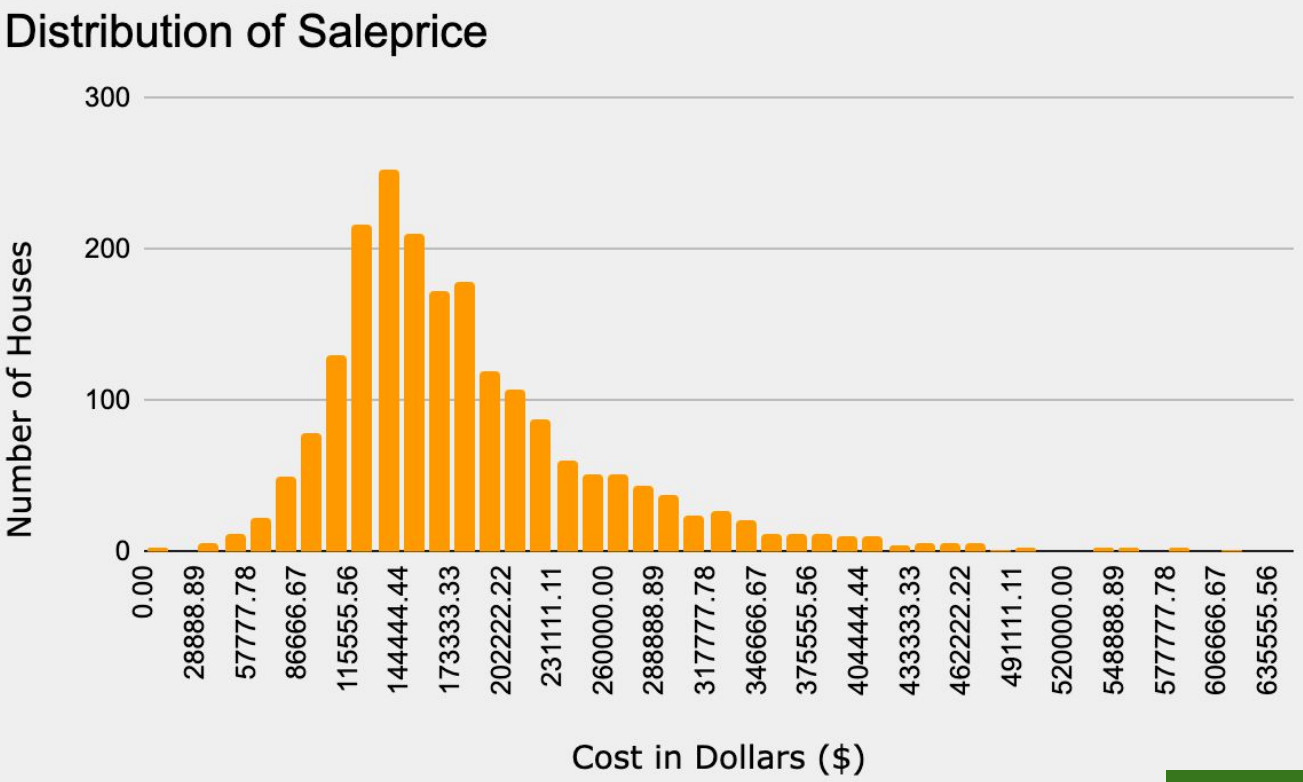
3. PLOT CHARACTERISTICS

Features related to the land on which the house sits like land size, frontage onto the street and gradient of land.

4. LOCATION

Where the house sits within the greater city and area - specifically neighborhoods of interest.

SALEPRICE OF HOMES IN AMES, IOWA



METHODOLOGY

EDA

Extracting and analysing correlations from the dataset for modeling purposes

MODELLING

Developing models and numerous iterations of features to provide the highest quality results.

CLEANING

Cleaning the data set including condensing features and converting to ordinal scales

PREPROCESSING

Feature selection and final engineering, scaling and regularization of data

EVALUATION

Evaluating our model and model outputs to make informed recommendations

CLEANING PROCESS

50%

Missing values and feature creation

- Reducing columns of up to 2,000 missing values. Putting in 0's and median for lot frontage
- Combination columns like bathrooms, deck and porch, age of house and remodel

35%

Fixing outliers and dummifying

- Dropping outliers outside of 3std from the mean
- Dropping columns likely to be collineated

15%

Conversion to ordinal data

- Converting qualitative to nominal categorical data to measure changes in, for example, quality

MODEL AND METRICS

01	**Baseline**	• Based on the mean of about $180,000 • No ability to explain variance in the data • Baseline RMSE of **~78,000**
02	**Multiple Linear Regression**	• Utilizes 14 selected features • **Train R2: 87%** • **RMSE of 30,050**
03	**Polynomial Regression**	• Utilizes 14 original selected features • **Train R2: 91%.** Slightly overfit. • **RMSE of 26,506**
04	**Ridge Model**	• Utilizes 32 selected features • **Train R2: 89%** • **RMSE of 26,638**
05	**Lasso Model**	• Utilizes 32 selected features • **Train R2: 89%** • **RMSE of 26,573**

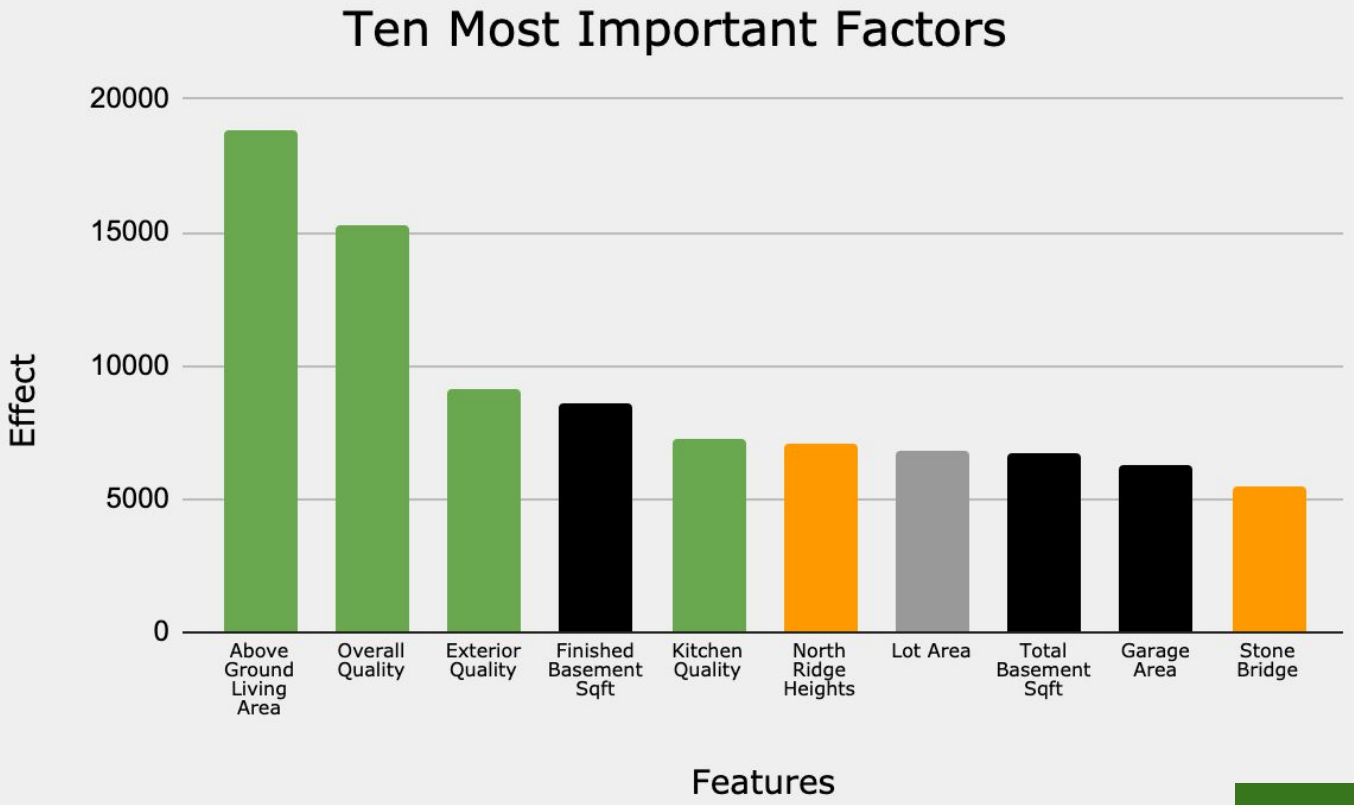
PRICE DETERMINANTS

HOUSE CHARACTERISTICS	40%
HOUSE ADD-ONS	30%
LOCATION	20%
PLOT CHARACTERISTICS	10%

SALIENT FEATURES

Above Ground Living Area	HOUSE CHARACTERISTICS
Finished Basement Sqft	HOUSE ADD-ONS
North Ridge Heights	LOCATION
Lot Area	PLOT CHARACTERISTICS

MOST IMPORTANT FACTORS



RECOMMENDATIONS

HOUSE

Total square footage above ground level as well as **internal and external quality** are key areas of focus. Furthermore, aim to design a **high quality kitchen**.

ADD-ONS

Include a **large finished basement,** plenty of **garage area**, and a **large basement** in general to attract a higher sale price.

LOT

Total lot area or the size of the plot of land as a whole is of importance to consumers, even more than the size of the garage and total basement area.

LOCATION

The **North Ridge Heights** and **Stone Bridge** neighborhoods tend to correlate with higher saleprices. Developing an understanding of why this is the case is important.

FURTHER STEPS

LOCATION ANALYSIS

With additional data we'd like to explore why some areas in Ames have a larger effect on saleprice than others (correlation or causation?) and help bring this insight to understand up and coming areas in the Greater Ames area.

DETAILS

With the current dataset we'd like to better understand how different aspects of house characteristics and types of features interact and model these interactions to better understand customer choices.

CREDITS

Thank you! Any questions?

Presentation template by Slidesgo