Kaggle Project

Century 21 Ames

Home Sell Price Prediction Models

Submission:

December 5, 2019

Contributors:

Jaclyn Coate

Josh Eysenbach

Andrew Leppla

Branum Stephan

Table of Contents

[1 Introduction 2](#_Toc25143574)

[2 Data Description 2](#_Toc25143578)

[3 Analysis Question 1: 3](#_Toc25143583)

[3.1 Problem Statement 3](#_Toc25143584)

[3.2 Build and Fit the Model 3](#_Toc25143586)

[3.3 Checking Assumptions 3](#_Toc25143588)

[3.3.1 Residuals Plots 3](#_Toc25143589)

[3.3.2 Influential point analysis (Cook’s D and Leverage) 3](#_Toc25143590)

[3.3.3 Make sure to address each assumption 3](#_Toc25143591)

[3.4 Comparing Competing Models 3](#_Toc25143592)

[3.4.1 Adj R2 3](#_Toc25143593)

[3.4.2 Internal CV Press 3](#_Toc25143594)

[3.5 Parameters 3](#_Toc25143595)

[3.5.1 Estimates 3](#_Toc25143596)

[3.5.2 Interpretation 3](#_Toc25143597)

[3.5.3 Confidence Intervals 3](#_Toc25143598)

[3.6 Conclusion 3](#_Toc25143599)

[3.6.1 A short summary of the analysis 3](#_Toc25143600)

[4 Analysis Question 2 3](#_Toc25143601)

[4.1 Problem Statement 3](#_Toc25143602)

[4.2 Model Selection 3](#_Toc25143603)

[4.2.1 Type of Selection 3](#_Toc25143604)

[4.3 Checking Assumptions 3](#_Toc25143605)

[4.3.1 Residual Plots 3](#_Toc25143606)

[4.3.2 Influential point analysis (Cook’s D and Leverage) 3](#_Toc25143607)

[4.3.3 Make sure to address each assumption 3](#_Toc25143608)

[4.4 Comparing Competing Models 3](#_Toc25143609)

[4.4.1 Adj R2 3](#_Toc25143610)

[4.4.2 Internal CV 3](#_Toc25143611)

[4.4.3 Kaggle Score 3](#_Toc25143612)

[4.5 Conclusion 4](#_Toc25143613)

[4.5.1 A short summary of the analysis 4](#_Toc25143614)

[5 Appendix 4](#_Toc25143615)

[5.1 Well comment SAS Code for Analysis 1 and 2 4](#_Toc25143616)

# Introduction

## When someone starts the processes of buying a home it is usually filled with daydreams of must haves and a few needs. This investigation is going to examine [through multilinear regression] the different components that contribute to the pricing of house buying and what Century 21 Ames can do to predict the pricing for home buyers in Ames, Iowa.

## With conservative methods our first analysis will provide a house sell price prediction model based solely on square footage and sell prices from only active Century 21 Ames neighborhoods [NAmes, Edwards, BrkSide]. With this initial model Century 21 Ames will be armed with a model to help them predict the sell prices of homes in the neighborhoods in which they are actively working.

## Using four different model selections we will be completing a second analysis that will build a predictor model for the selling price of house across all of Ames, Iowa. With this model analysis, Century 21 Ames, will have a strong predictor of all variables that highly affect sell prices across all the neighborhoods in Ames, Iowa; effectively empowering them to expand from their three active neighborhoods to much more.

# Data Description

## The data in this evaluation contains over 1400 rows and 80 different variables that could contribute to the selling price of a home in Ames, Iowa. The data was collected across 25 different neighborhoods from houses that have been built between 1872 – 2010.

## In our first analysis we will be using the following variables to predict sell price: *GrLivArea* [living area square footage], *SalesPrice* [sales prices of homes in neighborhoods], and *Neighborhoods* [NAmes, Edwards, BrkSide].

## In our second analysis we will be building a predictive model for sales prices of all the homes in Ames, Iowa. This will include the follow variables: XXX [] XXX [] XXX [] XXX [] XXX [] XXX [] XXX []XXX []XXX [].

## To find out more about this data and the definitions of all individual variables you can visit the Kaggle competition website [here](https://www.kaggle.com/c/house-prices-advanced-regression-techniques/data) [https://www.kaggle.com/c/house-price-advanced-regression-techniques/data].

# Analysis Question 1:

## Problem Statement

### Century 21 Ames wants the estimate of how the sale price of a house is related to the square footage of the living area of the home in their three active neighborhoods: NAmes, Edwards, and BrkSide.

## Build and Fit the Model

### 

## Checking Assumptions

### Residuals Plots

### Influential point analysis (Cook’s D and Leverage)

### Make sure to address each assumption

## Comparing Competing Models

### Adj R2

### Internal CV Press

## Parameters

### Estimates

### Interpretation

### Confidence Intervals

## Conclusion

### A short summary of the analysis

# Analysis Question 2

## Problem Statement

## Model Selection

### Type of Selection

#### Stepwise

#### Forward

#### Backward

#### Custom

## Checking Assumptions

### Residual Plots

### Influential point analysis (Cook’s D and Leverage)

### Make sure to address each assumption

## Comparing Competing Models

### Adj R2

### Internal CV

### Kaggle Score

## Conclusion

### A short summary of the analysis

# Appendix

## Well comment SAS Code for Analysis 1 and 2