Modeling Assigment #2

Brandon Moretz

### Introduction

To accurately forecast the value of a home, we must find a relevant dataset that contains accurate information of comparable inventory so that we can explore the significant variables of a home which ultimately determine the sale price of the residence. Once we have explored the data set and selected an appropriate sample from the population, our task will be to create both single and multivariate regression models that leverages these key indicators in the data to predict the value of a home given based upon its features. Once we have constructed the models, we will form hypothesis tests at our stated confidence intervals and conduct statistical significance tests upon these models.

In this report, we will use the Ames dataset which is an alternative to the famous Boston housing data to perform exploratory data analysis through variable derivation, validation, selection and visualization to measure the relevance of these indicators as they pertain to the value of the home in terms of a dollar estimate.

### Sample definition

This data is from the Ames Iowa Assessor’s Office and contains characteristics regarding residential properties sold in Ames from 2006 to 2010.

The Ames housing dataset contains approximately three-thousand observations of eighty-two variables collected from the Ames Assessor’s Office specifically for assessing value of individual residential properties sold in Ames, Iowa from 2006 to 2010. Given that this data was collected for specifically this purpose, it should be an ideal source of information for our observational study and resulting regression modeling.

For the sample of homes, we are looking for a “normal” set of homes to build our regression models. For normal in this example, we will choose to only look at single-family style homes (including residential zoning) and non-abnormal sale conditions. We also restrict our analysis to homes with a sale price less than $700,000 as the vast majority of our homes meet this criteria. We can see the waterfall of our sample size with each of the preceding chart:



### Sample definition

### Research

### Conclusion