Introduction

The United States housing bubble is one of the sectors that was severely affected by the 2007-2009 recession in the US. The extremely rise in subprime mortgage delinquencies and foreclosures, and the resulting decline of securities leads to a financial crisis over the United States economic (North Carolina Department of Statistics 2016). Over half of the US states was affected by the declination of housing price and reached its largest drop in 2008(cite). Housing price, hence, is an important sector in predicting and presenting the economic growth of a country, especially the United States. According to Mark Zandi, chief economist of Moody's Economy.com, housing prices declines of 10-15 percent are enough to get significant credit problem and to eliminate the homeowner's equity (Bernasek). Ultimately, having a model to define which is the group of significant factors in predicting housing price is a big contribution to avoid future recession and a potential financial crisis.

The purpose of this project is to build a regression model to predict the housing sale price. So, we will analyze the relationship between the housing market and a large number of categorical variables in home value assessments from a data science and academic purpose instead of from an economic perspective like the majority of the literature. The Ames Housing dataset conducted by Dean De Cock for data science education from 2006 to 2010 was utilized in the analysis. With an easily understood predictor variable of home sale prices, this dataset provides an opportunity to explore and conduct multiples regression models from simple inference oriented models to the more complex prediction oriented models. I hypothesized that factors related to house area and house quality are significant in my future model.

Materials & Methods