

Project Proposal

May 2021

I have chosen to do a project using the Ames, Iowa housing dataset which we have used throughout the assignments in the course. This dataset covers housing and location information concerning homes in the Ames, Iowa area. The data has both categorical and non-categorical variables and has missing or null values. The data set is of 1500 residential homes in Ames Iowa consisting of data from 2006 to 2012. It consists of a total of 82 variables around this housing and location data.

My research question is, can I using the Ames housing data set, accurately predict sales prices of the homes, using regression techniques and multiple models to create the most accurate predictor. The aim will be to create, using the techniques learned throughout this class, to create a predictor model using regression techniques that accurately predicts the aforementioned housing prices. I will perform data exploration and the other requirements of the project in R studio and share my results.

I plan to use a regression technique, most likely a Lasso regression model in R studio to predict the housing prices. I also plan to add clustering, and a k-means regression model to further view the data. I will conduct the EDA and data transformation and cleanup within R studio as well, and finally conclude with the model predictor and the corresponding results.