Anthony Ivy

Doug MacIntosh

Anna Jacobs

Final Project Proposal

Variables that Affect Housing Sale Timeline in the Chicagoland Area

1.1 Project Idea

To create a predictive model to determine which factors have the largest impact on Chicagoland residential housing sales timeline. Some of the questions we plan to explore are as follows: (i) which variables affect how quickly property is sold and for what price, (ii) what is the magnitude of impact for these variables, (iii) does zipcode affect the probability of sale in a given time period, and (iv) which variables cause a property to sell above the median listing price in an area.

1.2 Data Set

Our data is compiled from collected 2020-2021 MLS data and the US Census Bureau. The MSL data is comprised of information from sold homes in the Chicago area and neighboring suburbs. Because the data set is from a single year, we intend on breaking it into training and testing data to help us create our models. Some variables in the MSL data set include sale price, number of beds and baths, neighborhood, whether there is a garage, and size of the property which we believe logically will affect sales. The US Census Bureau data includes demographic information by zip code which will also be useful to look into. A longer time horizon for the MLS data set would be more ideal, however with over 3000 homes in this set, and an average selling time of 83 days, there is enough overturn to gain the necessary insights.

1.3 Methods Used

We intend to use multiple linear regression, random forest, boosted, k-NN, and neural networks to develop our initial models. From there, we will tune each model and determine which variables provide the most impact before comparing the root mean squared error against the testing data to determine the best model.