

Archana Ramkumar
Dr. Lucas Mentch
Statistical Learning & Data Science
April 18, 2023

Final Report

The overall goal of this project is to gain insights into the property information that drives housing prices in Pittsburgh for PA Reality. To do this, I built a model that predicts house prices using the most relevant set of housing information among all the attributes given. The model I have built and tested will hopefully allow a team of realtors to identify overpriced or underpriced homes by comparing the list price to the predicted house price. In the set of historical data provided, there are 16 aspects of various property information including but not limited to: description of home, the exterior finish of the home, the number of bedrooms, bathrooms, fireplaces, the average household income in the zip code the house is in, distance from downtown Pittsburgh, etc. After considering a variety of models, I ended up choosing one that is overall more complex in terms of its algorithm because there is a lot of data points and property information to consider. I sought to find a model that would best be able to handle this data without misinterpreting any underlying relationships between any of the property characteristics and housing price.

From my analysis and model that I choose, the four most important attributes which resulted in the best predicted house prices in Pittsburgh is square footage of the home, followed by average household income in the zip code the house is in, lot area (in square footage), and the number of bathrooms in the house. Square footage of the home had the most significant association in determining price, meaning that higher priced homes had larger square footage. Similarly, bathrooms had the second most significant association in determining price. Based on these findings, PA reality can consider focusing on this subset of housing information when

looking at various properties and their list prices. One thing to note is that I cannot say for *certain* that these set of factors are the *only* aspects that matter in driving house prices in Pittsburgh as this is just based off my workflow and intuitive methods. In the future, I could better analyze the relationships *between* various aspects of property information more thoroughly to gain key insights into general patterns or related attributes that influence housing prices.

In general, the model and set of property information that I have deemed to be important can be used to identify under/overpriced homes. This information can be leveraged by both buyers and sellers to make more informed decisions when it comes to purchasing or selling a property at a reasonable price. Moreover, the model can be used to identify trends or patterns in the housing market. These trends could inform economic policy related to housing affordability and availability in the city of Pittsburgh. An example could be if properties now are going up, this model can be used to determine if the pricing is kept in check.