Final Report

PA-VA Realty is interested in understanding the important factors that influence house prices in Pittsburgh, Pennsylvania and Richmond, Virginia. Being provided with a historical data set of housing prices and detailed property information, I sought to identify these important factors, as well as build a model that would help the team of realtors easily pinpoint over and underpriced homes.

The most important property details to take note of for gauging housing prices are state, number of bathrooms, square footage of the house, roof material, and average household income of the zip code. To offer some intuition, housing price tends to be positively correlated with number of bathrooms, sqft, and the average income of the zip code area. Additionally, Virginia houses tend to be more expensive. However, the different factors do interact, so these are not strict rules. State, number of bathrooms, square footage of the house, roof material, and average household income of the zip code appear to be significant in at least 9 of the 13 models that were analyzed. All these predictors, except for average income, were the most important variables of the best model, which was the model that produced the smallest prediction error.

It is to be noted that many of the predictors are not only closely related to price but also closely related to one another. For example, number of total rooms, number of bedrooms, and number of bathrooms are naturally all positively correlated with each other. The similar information that they provide in predicting house price may be a reason why number of bathrooms was considered very important and number of total rooms was not.

While some factors are important in predicting housing prices, others are less so. In this study, number of fireplaces was removed from consideration because every house from Virginia was missing this information. The number of fireplaces may be important in determining house

prices, but a more complete data set would be necessary to find out. It was also found that housing description is rather insignificant in predicting housing prices. This is mostly due to 90.8% of the houses in this data set being classified as single-family houses, with categories like mobile homes constituting as little as 0.07% of the data. However, your mileage may vary, whether PA-VA Realty decides to expand its business to areas with more diverse housing styles (e.g. condominiums, row houses, multi-family houses).

In conclusion, state, number of bathrooms, square footage of the house, roof material, and average household income of the zip code appear to be the most influential factors of housing prices in Pittsburgh and Virginia. The generated predictive model considers the relationships between these factors and more to forecast a house's price, all in order for PA-VA Realty to seamlessly identify mispriced homes.