

Non-Technical Report

The goal of the project is to predict the price of houses in Pittsburgh, PA and Richmond, VA as well as understand what factors drive or explain those prices. To start, I have all the information about 1200 houses including their prices, and information about 600 more houses the price of which I need to predict. I analyzed the data I had, and then I trained a lot of different models to find the one that is the best at predicting.

The most important factor in driving the house value is the square footage of the house. The larger the square footage, the more expensive the house is. The second important factor is the number of bathrooms in the house, with houses with more bathrooms having a larger price. The total number of rooms and the number of bedrooms are also good predictors of a house price. The more rooms there are in a house, the more expensive it is.

The best models are good at predicting the prices of single-family houses, houses with shingle roof type, houses with brick exterior finish, houses with total number of rooms ranging from 5 to 10, with the number of bathrooms from 0 to 3, with 1 to 4 bedrooms, and houses built after 1900 year. We can predict the prices of such houses with a greater accuracy because we had a lot of data about them to start with.

There are some houses I would recommend to not trust the predictions on. The reason is that we did not have enough data on them to start with; thus, we are unsure about the possible price of those houses. They include mobile homes, rowhouses, houses with log or concrete exterior finish, and roll roof type, houses higher than 2 stories, old houses built before 1900, and houses with more than 5 bedrooms and more than 12 total rooms.