Real Estate Final Report

Problem:

Find a way to accurately predict sales prices of homes without needing a large amount of data

Approach:

I was given a spreadsheet with data on homes and other commercial properties. I cleaned the data and filtered it down to what we needed for this project. After that I looked for correlations between the variables we had and what the sales price ended up being. From there multiple models were built and evaluated to determine the best approach to predicting the sales price. A random forest model ended up being the best choice.

Findings:

The random forest model was able to predict the sale value with an r-squared of over 0.8. This could likely be improved with extra information such as number of bedrooms and bathrooms and square footage.

Recommendations:

This model can be used with the assessed value of the home, the month you're looking to sell in, the city the home is in, and whether the home is a single family residence or multi family residence. The model could also be improved if we had access to more data. This could change the model type or it could remain the same.