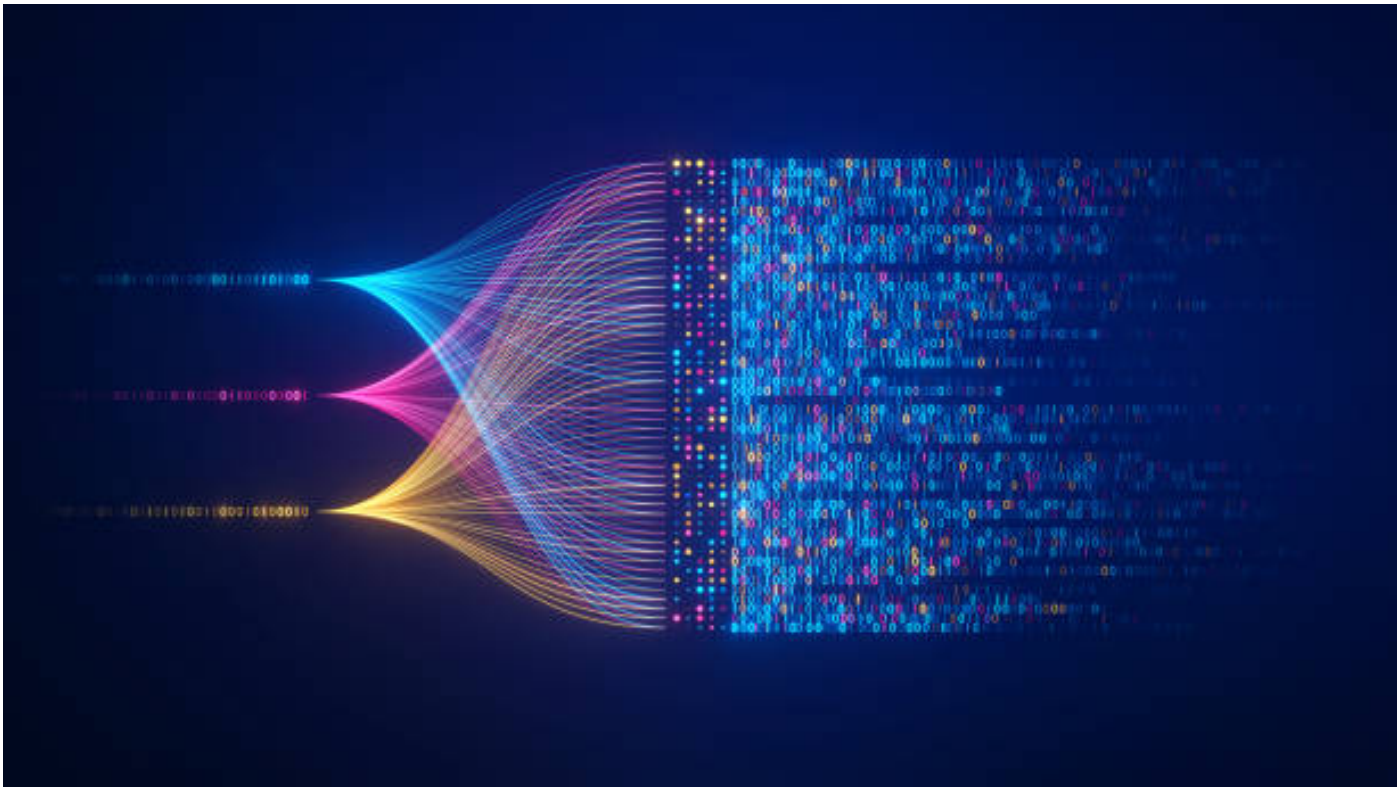
Triple Bam Corp.

ML Price Prediction Report



Made Possible By:

Lead: Luke

Doc Lead: Sam

Model Lead: Jarom

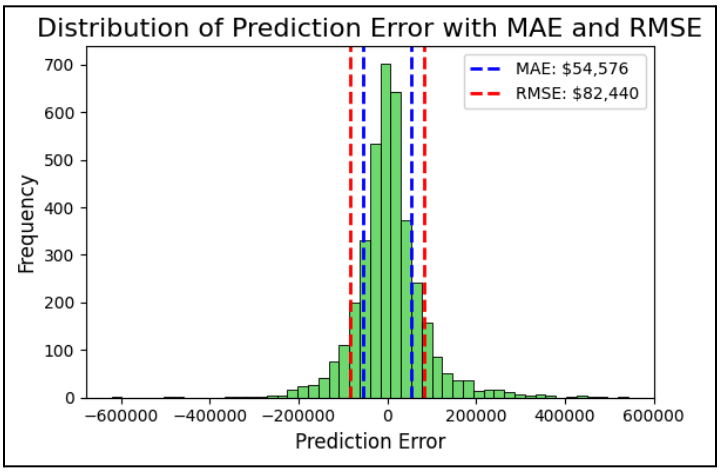
Chart Lead: Tyler

Export Lead: Scott

Background

Reddic Housing LLC is a company that provides housing price estimates to development firms. The team has been tasked with creating a machine learning model that will accurately predict housing prices based on set predictors. Our team has been tasked with heading the construction of this model and making sure it is as accurate as possible.

Multiple iterations of the model were tested. It was expected that the team would have to heavily doctor the data in order to have it accurately give predictions, but it was found that simply min-max scaling many of the data points fixed most of the issues the team would have run into.



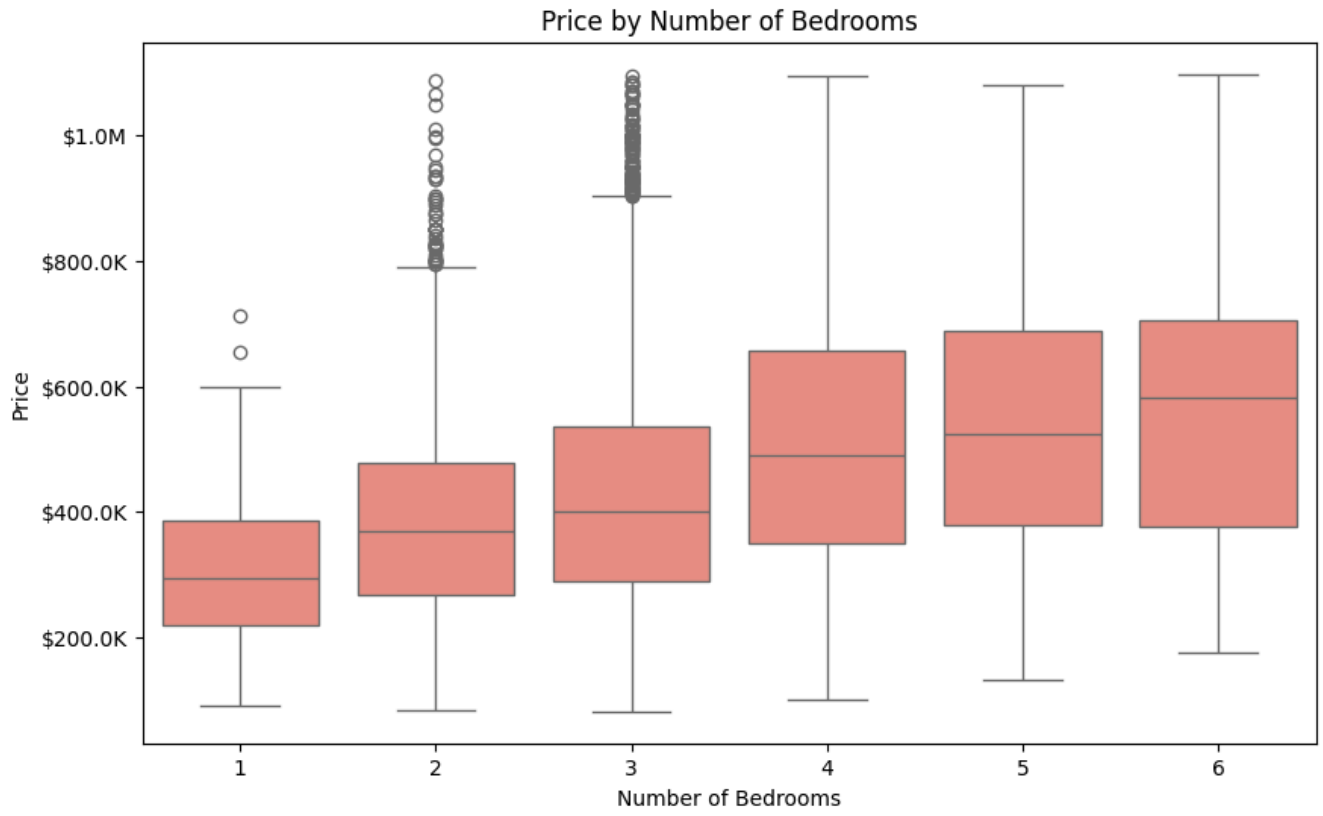
The current iteration of the model has a Mean Absolute Error (MAE) of $54,576. The mean of the data sits around 540,000, which means the MAE is only 10% off of the mean of the data. This is relatively reasonable, especially in the housing market where price variations can be significant.

The R^2 is sitting at .88, which is very impressive. This shows that the model is explaining about 88% of the variance in housing prices. Our Root Mean Squared Error (RMSE) is sitting around $82,000 which is almost 15% off of the mean, which shows that there are some larger errors in the form of outliers in the data. This makes sense as the housing market can be volatile at times and some of the data in the set seems to be incorrect, forming artificial outliers.

Findings

As our team analyzed the data, we identified several strong predictors of a home's price. One of the most influential factors was the number of bedrooms, which showed a steady correlation with higher prices. This makes sense, as more bedrooms often indicate a larger home with increased square footage—another key driver of home value. Similarly, the number of bathrooms and overall living space contributed significantly, as homes with more functional space tend to be more desirable and command higher prices.

Location also played a crucial role in price variation. By grouping zip codes, we found that homes in more desirable areas consistently had higher values. This suggests that regional demand, amenities, and neighborhood quality significantly impact pricing. The combination of structural features like square footage and location-based factors provided a strong foundation for our model, reinforcing common real estate trends while improving predictive accuracy.



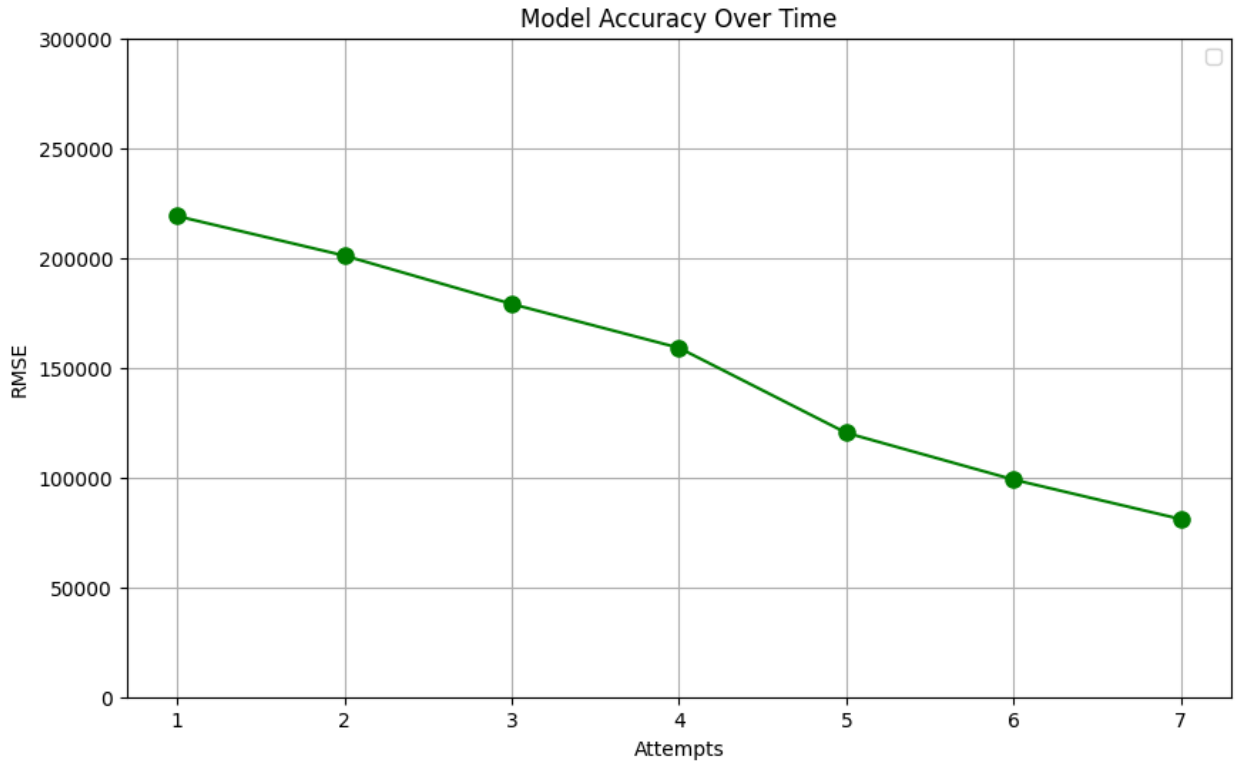
Results and Action Items

The team recommends using the model that has been created and discussed above. There should continue to be a strong focus on finding features that are strong indicators of price in the housing market. What these features are can be simply broken down into location and space. Space of the house being sold and of the 15 closest neighbors combines into a strong predictor of price. Location becomes a strong feature when we look at the scales we were given in the dataset (view, waterfront).

Housing prices can fluctuate wildly depending on the time of year and based purely on the economy and how many houses in the area are up for sale. The error in the model is relatively small compared to these fluctuations, so should be able to accurately give users information within around $50,000. If given additional features, this number could easily be lowered, especially if the given data has a strong correlation to price. Users should not have any issues estimating a price using our new model and there should be very little to no frustration with its estimations.

To get our model as accurate as possible we took advantage of all the connections in the data we found. One of the more fun ways we did this was by grouping similar zip codes. Because one of the strongest correlations we found in the data was price compared to the price of the neighborhood. Some zip codes have very limited representation and to increase accuracy we

grouped nearby zip codes together. Adding this change alone was able to reduce our RMSE by close to 10,000. As shown in the chart below, we started with an RMSE over 200k and in just a week of development we were able to get it down to 81K and it will only get better.



Python Notebook

https://colab.research.google.com/drive/1275gTP1HQHZvZ0fzf9TpwfHz1XlZ2O-r