This project will use a housing sales price dataset from Kaggle listed below, in order to predict a sales price for a given home and/or determine if a home is under or over priced for the market, according to a model using regression analysis and determine the features or group of features that most influence said price.

https://www.kaggle.com/competitions/house-prices-advanced-regression-techniques/

Recently there was news of Zillow attempting to transition into a homebuying and flipping business model which it ended shortly after it began. There has been a lot of speculation about this. Possible explanations for the aforementioned failure include costs associated with pricing and acquiring homes, outweighing any profit, volatility in the housing market, or an inability to purchase or sell at an appropriate volume, or a lack of capital, or any number of other possibilities. The closing costs associated with the purchase of a home and subsequent sale may be too high as a percentage of the value of the home, which would impact the profit from sale. The goal was likely to use a predictive model to find neighborhoods with quickly rising value trends where they coud purchase homes for less than they could reasonably sell them for a short time later, or identify features that justify a higher than average price, and find homes with said features that they could acquire at an average or below average cost, while avoiding homes otherwise. The hypothesis I'm putting forth is that the failure is not due to an inability to appropriately price homes with predictive models, which will be demonstrated by a reasonably accurate predictive model in this project.

There could be many other factors that led Zillow to halt their pursuit of this business model, such as poor marketing, or the expenses of marketing, or low response to marketing, such as in the case of mass mail marketing. Or… Perhaps this failure was due to the implementation of poor predictive models.

Success will be defined by the ability to apply this model to homes listed for sale in a test dataset to determine if they are over or underpriced, and identify those that are underpriced by at least 10% with a relatively high r-squared value and low mean absolute error, yet to be defined, within a given geographical area. With this model, a client such as Zillow could pursue a profitable house flipping business model, or make a solid determination not to pursue said business model.

The deliverables will include a written report, detailing the findings, the jupyter notebook version of the model and a slide deck to summarize the findings.