

This project uses the King County House Sales dataset, which can be found in kc house data.csv in the data folder in this repo. The description of the column names can be found in column names.md in the same folder.

## **Business Problem**

Abodes by Abhineet, aka Ab<sup>2</sup>, is a company that buys, renovates and resells homes. Ab<sup>2</sup> has seen an exponential take off over the past year and no longer has the staff to sort through all of the home listings in their area of operations - King's County in Seattle, Washington. With their extra capital they have decided to bring on a data scientist whose first responsibility is to create a tool that will identify sale listings worth investigating. Ab<sup>2</sup> has access to a large dataset of homes sold in the county in 2014 and 2015, and will provide it to the analyst.

# Results

In the end, we produced a Multiple Linear Regression model with an adjusted  $R^2 = .88$  and a Prob(F-stat) < .001. Thus, we can be fairly confident in the ability of our model to produce solid home pricing estimates. With confidence in our model, we built an estimator that takes in the information you may easily find in a home for sale listing, converts it to have all the attributes needed to be input into our model predictor, and returns an approximate price for any given home. Note that this estimator is very specific to King's County, WA. Based on our model, we can make some general statements. Keeping all other predictors the same, for a unit increase in:

- 1. bedrooms, the price of the home will decrease
- 2. bathrooms, the price of the home will increase
- 3. sqft\_lot, the price will increase slightly
- 4. sqft\_above, the price will increase strongly
- 5. sqft\_basement, the price will increase
- 6. yr\_built, for each year the home is newer the price decreases slightly
- 7. latitude, for each tenth of a degree northward the price will increase
- 8. longitude, for each tenth of a degree westward the price will increase
- 9. sqft\_living15, for each extra 100 sqft on average the 15 homes nearby have the price of your home will increase
- 10. sqft\_lot15, for each sqft of land your 15 nearest neighbors have on average, your home price decreases, but only very slightly

For the categorical columns, we can say:

- 1. There is less of a unit increase in price for more than one floor when compared to just having a floor at all (silly to say, but it's just important that the house exists).
- 2. The price of a home increases radically for a home that has waterfront property.
- 3. Relative to the most common grade of 7, houses graded less sell for less and houses graded for more sell for more in a way that is relative to their grade.

- 4. Houses with a view experience positive increases in unit increase in price per increase in view, scaling as the rating of the view increases.
- 5. Homes in some zipcodes sell for more or less relative to the zipcode with the most houses in it. We would have to do further investigation on the location of these zipcodes to make sense of this information.

# **Conclusions**

Based on our results, we can assert that the most important factors in determining our housing price are: that is has bedrooms, but only as many as necessary so that they don't take up additional living space; That it has as many bathrooms as is reasonable, such that people do not have to share; that the home is large and has a liveable basement, that the home is older and thus probably built closer to the city center; that your home is located as far west and as far north as possible, again because it places you closest to waterfront property and closest to the city center / industrial centers. The lot size does not matter so much, likely because expensive homes nearer the city have less lot space while lower priced homes further away have more lot space.

#### Releases

No releases published Create a new release

### **Packages**

No packages published Publish your first package

## Languages

Jupyter Notebook 100.0%