Home Remodel Profitability Analysis

Freddy Abrahamson

December 29th, 2021

Outline

- Business Problem
- Data Understanding
- Linear Regression explained
- Home Improvement recommendations

Business Problem

The goal is to be able to provide advice to homeowners, in the King County area, about how home renovations can increase the value of their homes, and by what amount.

To that end, we will use a linear regression model to gain these insights.

Data Understanding

- The data comes from the King County House Sales dataset, in form of a ‘csv’ file.
- King County is located in Washington State, and has a size of approximately 2300 sq. miles (U.S Census Bureau).
- The dataset has over 20,000 rows, and 20 columns.
- Each row contains information about a different home.
- Each column represents a different characteristic of that home, such as the zip code, square footage, number of bedrooms, number of bathrooms, building grade, and the number of floors.

Linear Regression Explained

- In order for us to be able to calculate by how much a particular home remodel, changes the home's price, we must first be able to account for all the different home prices.
- In other words: What causes home A to cost this much, and home B to cost that much ?
- This range of varying home prices is called the variance.
- A linear regression model would take all the factors that contribute to a home's price, such as square footage, the number of bathrooms, and the number of floors, and find the optimal coefficient to apply to each, so that as much variance as possible can be accounted for. The following would be an example of the resulting formula:

$$b_0 + (b_1 * sq\_ft) + (b_2 * bthrms) + (b_3 * flrs) = \text{home price}$$

**where b_0 is the constant, and b_1- b_3 are the coefficients

Linear Regression Explained: Model Results

- The metric that quantifies how much of the variance is accounted for is called the coefficient of determination or R^2.
- This metric is a fraction where the numerator is the amount of the variance that is accounted for, and the denominator is all of the variance.
- In the case of our model we have an R^2 of .653, which means that our model accounts for 65.3% of all the variance in price.

Home Improvement Recommendation: Explained

- For our model a natural log transformation was applied to the price.
- Transformations can be used to improve the performance of a linear regression model.
- A transformation is simply the replacement of a variable, by a function of that variable, which in this case is the natural log.
- In our case, our model will return the price as the power to which Euler's Number (e) should be raised .
- For example, if the model returns the price as 10, we know that the price is = e^10.
- Since only the price was log transformed and not the features of the home, we have what is called a log – level regression.
- In this case, the way to interpret the change in price, based on a change in a feature is:

$$\%(\Delta y) = 100 * (e^{b1*x} - 1)$$

**where e is approximately 2.718, Δ = change, y= price, x = feature of the house

Home Improvement Recommendations

1. A change from a .5 bathroom home to a three bathroom home, will result in an increase in price of 7.36%.
2. An expansion of the house from 1 floor to 2.5 floors, will result in a price increase of 21.86%.
3. An increase in living space from 1881 square feet, to 2500 square feet, will result in an increase in price of 17.73%.

Home Improvement Recommendation: Conclusion

- The change in price for a given unit of change in the features, is not necessarily the same.
- For example, the change in price from a 3 bedroom home, to a 4 bedroom home, is not necessarily the same as the change in price from a 1 bedroom to a 2 bedroom.
- The correct way to calculate the change is to calculate the price of your home as it is currently, and then make the modification you are interested in, and see how the price has changed.

Does Anyone have questions?

Thank You