King County Home Sales

Using data to understand what impacts home sale price

Flatiron School Data Science
Project 2

Zaid Shoorbajee

Overview

This analysis examines data about home sales in King County, Washington in 2014 and 2015.

The goal is to make inferences about **how much each attribute of a home affects its sale price.**

Outline

- Business problem
- Data and modeling
- Model results
- Conclusions and limitations

Business problem

A real estate agency needs to advise its clients on how to increase the estimated sale price of their homes.

What aspects of a home affect price the most?

- Construction quality? Square footage? A nice view?

What actions can a homeowner take to get their house to sell for more?

- Renovations?
- Emphasize certain attributes in a listing?

Data

Information about 21,597 homes sold in King County between May 2014 and May 2015.

Includes details such as:

- Square footage (living area, lot, basement)
- Number of floors, bedrooms, bathrooms
- Home condition and grade
- Construction and renovation dates
- Location
- Quality of view
- Waterfront?

Modeling

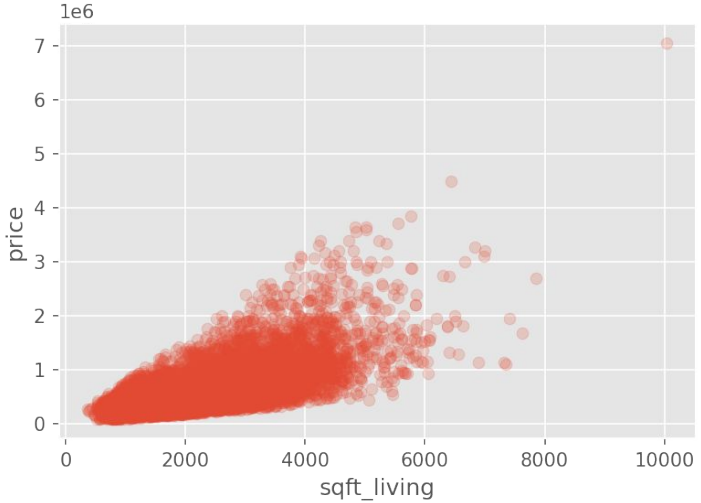
Used ordinary least squares regression.

- Tries to predict price based on a variable or multiple variables

Baseline model:

- Price as affected by square footage of living space.

Price as affected by square footage of living space



Modeling, cont.

Baseline model was poorly fit.

- R^2 of 0.441
- i.e. > half of the variation in price not explained by square footage

Improved the model by:

- augmenting some variables
- adding and removing other variables

Results: Final model

Predicts price using:

- Square footage of the living space
- Condition
- Grade
- Number of floors*
- View
- Waterfront
- Renovated in the last 10 years

This model has an improved fit: R^2 score is 0.59.

Generally, an improvement in any of these attributes tends to increase price.

* Floors:
Included because it improved the model, but the results were less intuitive.
E.g.: going from 1 floor to 1.5 floors would increase the price, but going to 2 floors, would decrease it.
Unable to make recommendations based on floors.

Recommendations

Renovate to improve:

- Square footage of the living space
- Condition
- Grade

Emphasize in listing and marketing:

- View
- Waterfront
- Recent renovations

Limitations

Unable to predict sale price in terms of dollars.

- Price and square footage were augmented in order to run a linear regression.
- Model does tells us whether something increases or decreases price.
- And we can compare the effect of one variable vs. another.
- But model is unable to convert the results into real dollar amounts.

Model fit is not stellar

- R^2 is 0.59.
- Can make inferences about effect on price.
- But much of the variation of the price is still unexplained (41%).

Thank you

Email: zshoorbajee@gmail.com

GitHub: @zshoorbajee

LinkedIn: linkedin.com/in/zshoorbajee/