House Price Prediction King County, WA

Kathryn Davé

December 4, 2018

Purpose and Goals of Study:

- The purpose of this study was to devise a formula for predicting home prices with a reasonable accuracy given various facts about a property in King County, WA.

- The ultimate goal would be to maximize home sale price and profits for a given property.

Source Data:

- The data used for this study consists of house sale prices in King County, Washington (which includes Seattle) between the time period of May 2014 and May 2015.

- There are more than 21,000 home sales recorded in this dataset.

- Each home sale record consists of 20 pieces of information relevant to the property including:
  - Number of bedrooms and bathrooms
  - Square footage of the home, lot and basement (if present)
  - Total floors (levels) in house
  - An overall condition score (between 1-5)
  - An overall grade based on King County grading system (number between 4-13)
  - The location based on zip code, latitude and longitude
  - The year the home was built and/or if renovated
  - If the property is waterfront or has a water view
  - The number of times the home has been viewed

Methodology:

- The methodology used to determine a predictive house price model consisted of several analytic steps:
  - Exploratory analysis of the data to understand the amount and type of the source data
  - Data pre-processing to ensure the validity and consistency of the data
  - Data modelling and refining:
    - An iterative process that looked at the impact of adding and removing home features to determine the most relevant and predictive ones
    - Compared statistical values of multiple linear regression (R-squared & Mean Square Error) to provide an estimate of the strength of the relationship between the model and home price
  - Predictive modelling and obtaining results:
    - Validating the model after identifying most predictive features and obtaining the final relational equation

Key Findings:

- The results of the multiple linear regression in the King County dataset identified the most important predictors of house price as:

  - The <u>grade</u> assigned by the county system
  - The line of <u>latitude</u> of the property
  - The overall <u>square footage</u> of the living space
  - Number of <u>bedrooms</u> and <u>bathrooms</u> in the home

Results & Conclusions:

- The model will predict with about 68% accuracy the home sale price of a property in King County, given the number of bedrooms, bathrooms, square footage, line of latitude, and county grade

- This was verified and can be visualized by a roughly linear relationship between the model's predicted home price values and the actual observed home price values:

Scatter plot of final revised model