Kings County House Price Prediction - Regression Model

By David Boyd

What will we cover

- Executive Summary (1 pager to send out via email)
- Business Problem
- Methodology and data sources used
- Key findings & results
- Recommendations
- Future analysis

Executive Summary

- Performed a multivariate regression model to identify which features impact house prices the most
- The 5 most correlated features with house price in the data were: grade of property, SQ footage of living space above basement, # of bathrooms, # of bedrooms and SQ footage of 15 closest homes in area
- In order to improve the price of your home, focus on improving these qualities first:
    - Grade of the property
    - Condition of the property

Business Problem

The stakeholder is someone who currently owns a home and is looking to understand which **factors** have the **biggest impact** on **increasing or decreasing** the value of a home.

The goal is to help homeowners **assist in their decision making process** when considering different renovations.

Methodology & Data Used

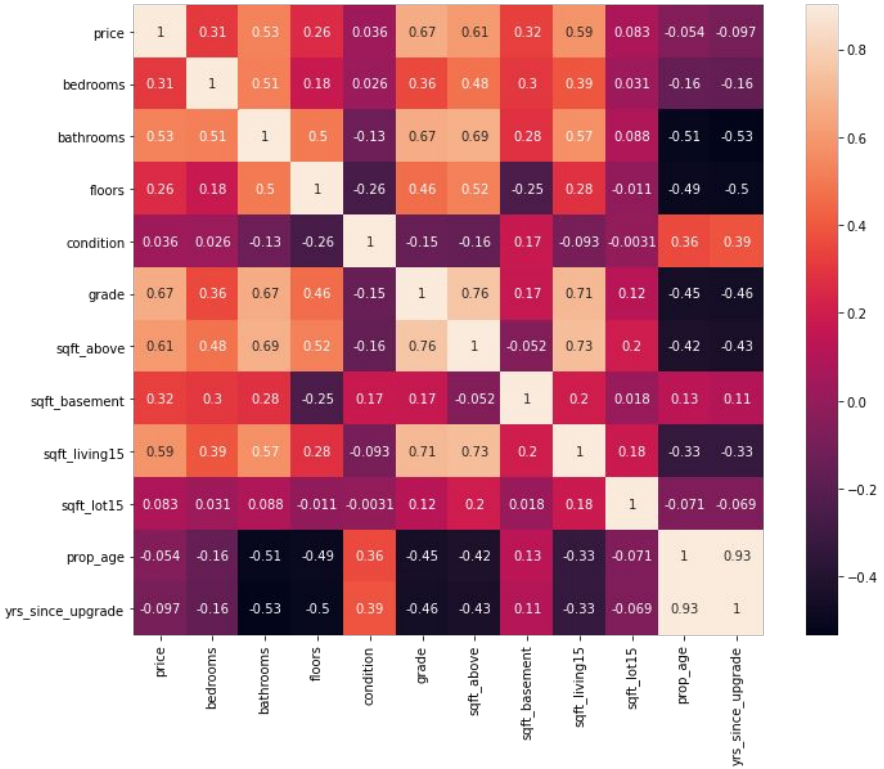
To create this model we used data from King's country data in a CSV which contained **18 different features**, with ~ **21,500** records.

- The data was scrubbed for **outliers**
- Data formats were improved
- Performed a **correlation analysis** to understand which features to choose
- Built a **multivariate regression** model to predict which features impact overall price

Key Findings

Looking at the correlation chart on the left, out of the data used, the features that are **most correlated** with price are:

- **Grade** of the property
- Sq footage of **above property** (living space above the basement)
- **# of bathrooms** in the property
- SQ footage of the **15 closest homes**
- **# of bedrooms** in the property



Recommendations

- Looking just at a baseline model, when only considering the SQ Footage of above living space, this accounted for **26%** of explainable variation in the price
- The final model accounted for **49%** of the explainable variation
- The features which have the most impact on house prices are the:
  - Grade of the property
  - Condition of the property
  - Overall above living space
  - If the property has been renovated since it was built

Areas for future improvement

In order to better improve the accuracy of the model and better understand other impacting features, I suggest looking into the following data:

- Stats around the **local neighborhood** (school quality, crime rate, etc)
- Does the property have a garage and if so, how many cars can it fit inside
- Proximity to local amenities
- **Socioeconomic data** around local area (Employment rate, etc)

Any Questions?

Email: dboyd580@gmai.com

Github: @db495

LinkedIn:
https://www.linkedin.com/in/david-boyd-16245ba1/