Predicting King County Housing Prices

Key findings from OLS Regression Analysis

Serena Quiroga, July 2019

Introduction & Approach

- **Objective:** To accurately predict housing prices in King County based on the available features by finding the most important variables that influence price.

- **Approach:** I cleaned, explored, and modeled the King County Housing dataset and used multivariate linear regression to determine which features have the **best predictive power for housing prices.**

- The dataset has information on *prices*, *number of bedrooms*, *bathrooms*, *floors*, *size*, *year built and renovated*, and more. However, only a handful of these features really count...

Key Findings

- The variables you might think are important at first glance, actually don't matter all that much…*waterfront properties, basements, year built, renovated*

- The size of the living space is what counts!

- Grades do matter – houses with higher grades have a higher, positive correlation with price.

- Location makes a difference in prices.

- How many times a house has been viewed can make a notable difference in price.

- More floors and more bedrooms can mean a lower price.

What doesn't seem to matter

- **Waterfront properties** – whether a house is on the waterfront or not does not show a strong influence on price.

- **Does renovation make a big difference?** Not quite. We would need more data on houses that have been renovated in order to determine this, because only 739 houses in our data have been renovated, with most of them undergoing renovations in the past 2-3 decades.

- **Year Built...**

Year Renovated for Renovated Houses ONLY
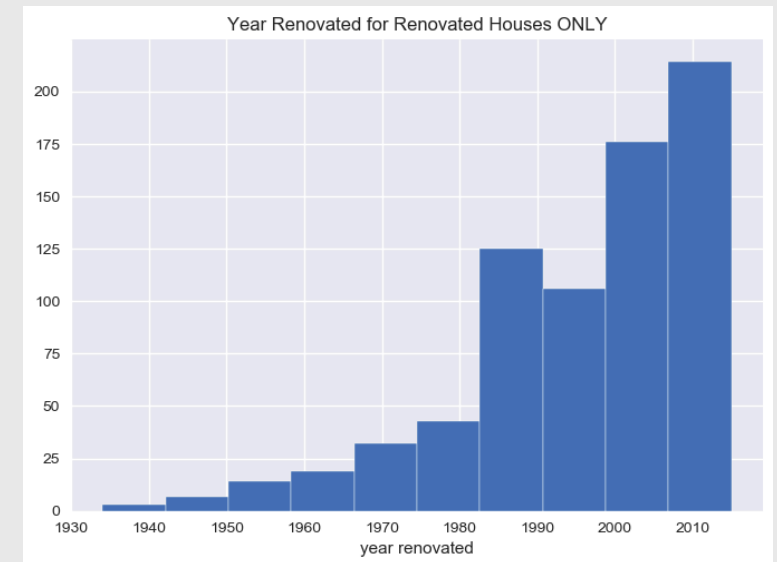
Year Built

In our dataset, we have houses built in 1900 through 2015. We can see plenty of variability over the years, with some notable spikes that correspond to fateful financial moments in U.S. history.

We smoothed the data out by looking at housing prices per decade. However, in doing so, one could be losing important nuances tied to specific years. Regardless, year built decades did not show strong explanatory power for price (3.5%), so we removed it.

Year Built vs. Mean Annual Price ($)
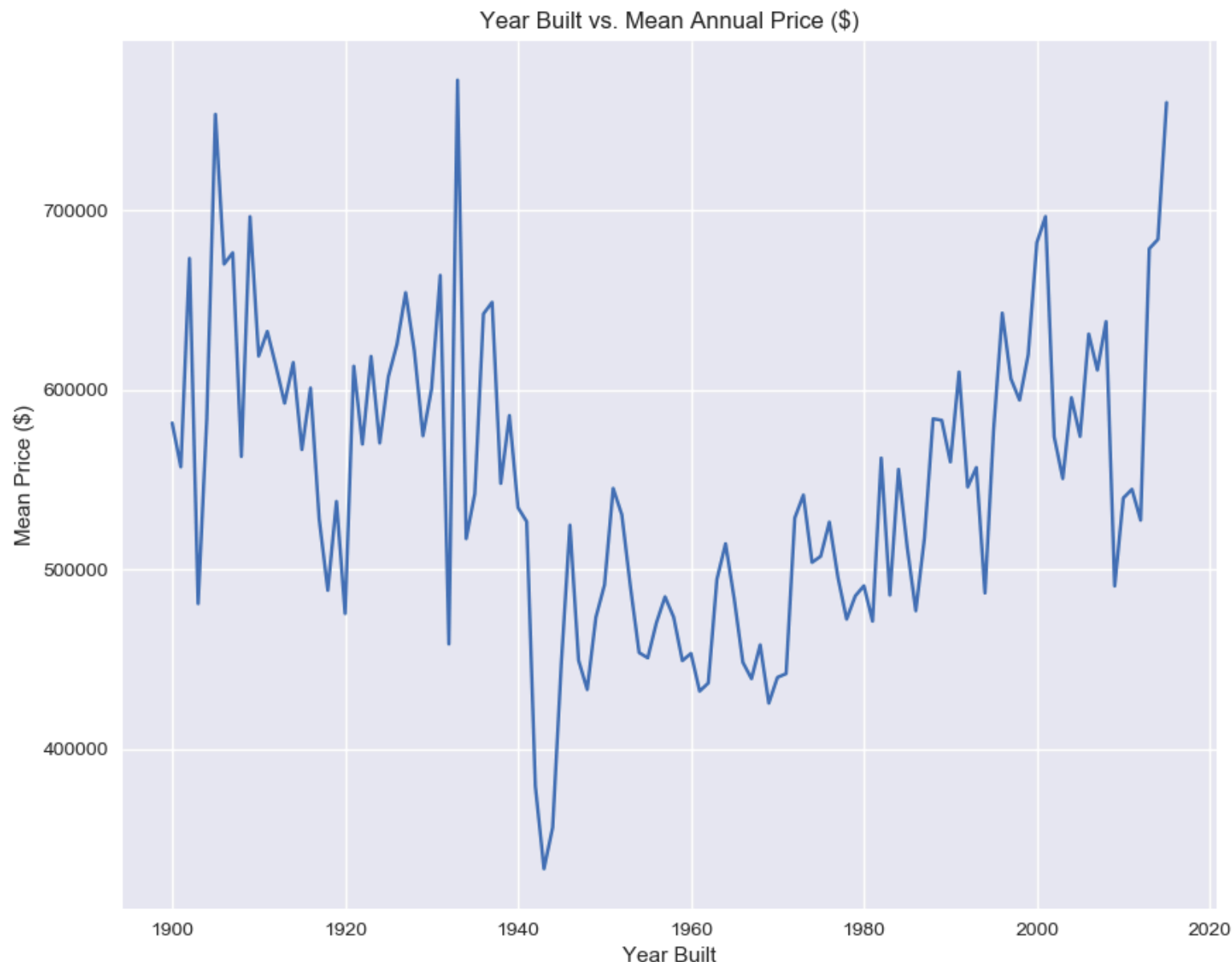
Bigger is better, and grades *do* matter

- The size (in square footage) of the living space has the highest predictive power for understanding housing prices.
  - A 1% increase in square footage is associated with a $2,712.91 increase in price, holding all else constant.

- Only the best grades (11-13) seem to matter in terms of having a statistically significant and positive impact on price.

(Log) Sqft Living & Price
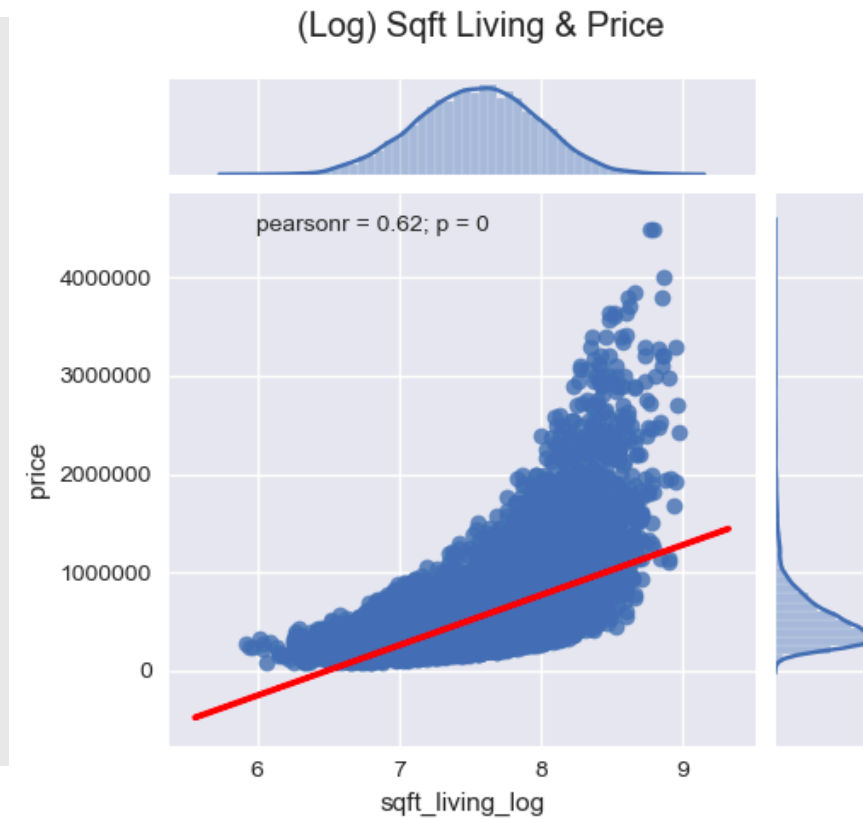
Location, Location, Location.

We used a tool to cluster the data by location (10 locations), and we can see some general patterns and concentrated pockets where price varies according to location.

Location 2 offers the most expensive housing prices, while location 4 has the biggest drop in prices, holding all else constant.

Price According to Geo-Coordinates
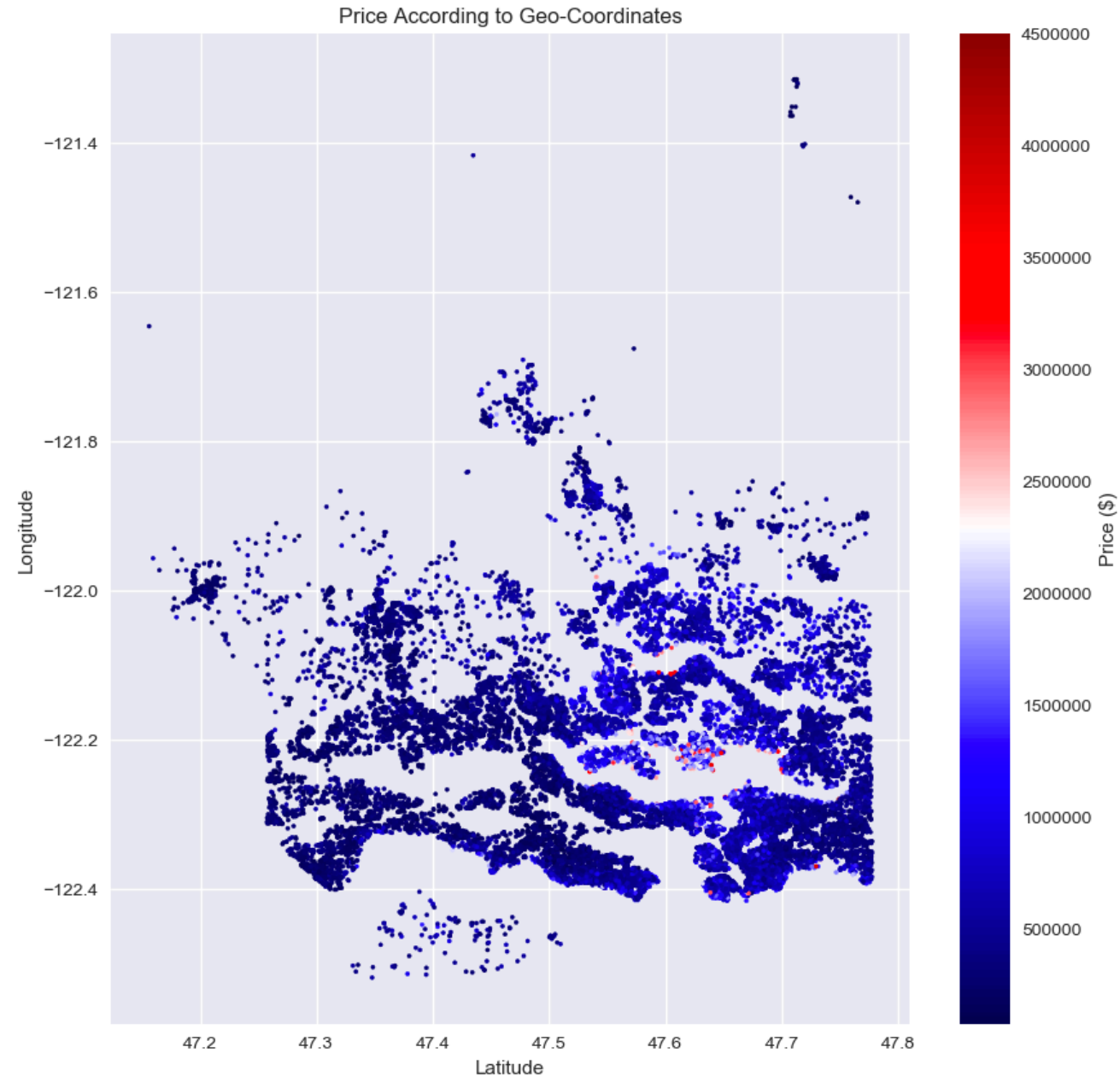
Summary

- Using data on square footage of living space, number of times viewed, grades, location, number of floors and bedrooms we are able to explain 71% of the changes in housing prices.

- Square footage has the biggest (positive) impact on price, followed by high grades (of 12 or 13) and a high view count (4).

- More exploration is needed to understand the negative relationship between bedrooms, floors and price.