King County Real Estate Analysis

Author: Andre Layton

Summary

This project analyzes the housing data from King County, WA for Keller Williams Realty to gain insight into features of a home impact the sales price of the property.

Initially, the main focus will be to select which feature other than heat sources impacts the price of the home. Although, once selected, the features will be modeled to determine the impact the features have on the property's value and by how much.

According to the models:

➔ The size/square footage of the living room and the heat sources will need to be renovated and should be the main priorities.
➔ Switch to gas-powered heat - when compared to electric-powered homes, gas-powered homes increased the sales price by approximately $97,000.
➔ Increase the size of the living room, keeping in mind that each square foot increase will yield an associated increase of $302 in the property's price.

Outline

- Business Problem
- Data
- Methods
- Results
- Conclusions

Business Problem

➔ Explore the relationship between the heat source of the home and its sales price, quantifying the impact, if possible.

➔ Determine another feature that can serve as a reliable predictor of a home's sales price.

➔ Select what renovations need to be made based on model results.

Data

➔ The initial dataset pulled from the King County government website contains many features, such as year the home was built, the number of bedrooms and bathrooms, etc.

➔ The final dataset contains the size of the living room (in square feet), the sales price, and three columns each representing a different heat source.
  ● The last three columns are the result of one-hot encoding, or "dummying", the heat source categorical column in order to make it numerical for modeling.

➔ We refine and focus our analysis to model two predictors, both individually and in a multiple regression test.

Modeling

➔ The modeling begins with a simple linear regression model between the sales price and the size/square footage of the living room.

➔ The size of the living room is the feature with the highest correlation to the sales price, in this dataset.

◆ In other words, it is the best predictor for sales price.

Square Footage of the Living Room vs. the Sales Price

Modeling (cont.)
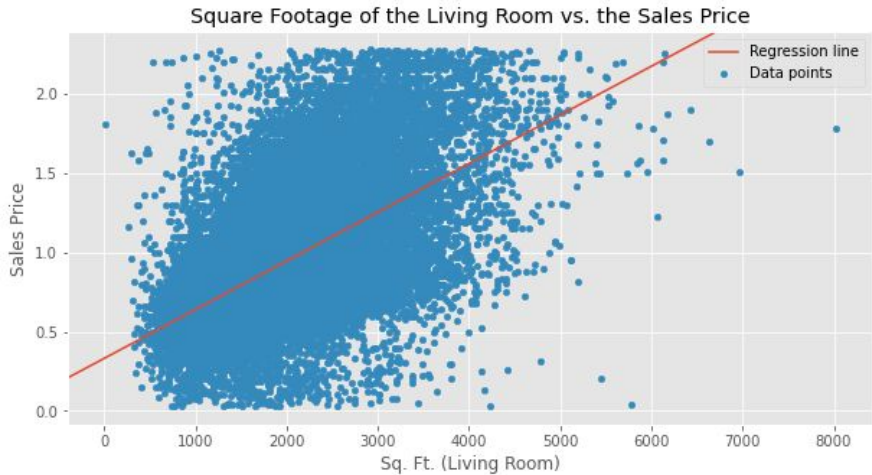
Sales Price vs. Heat Sources
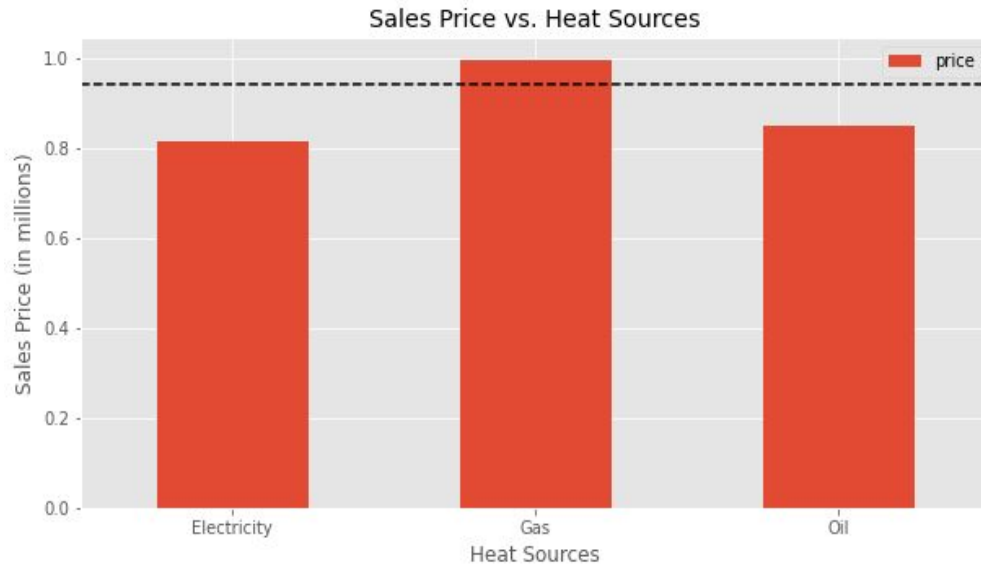
➔ The business problem also calls for research into how the heat sources impact and affect the sales price of the home.

➔ Gas homes are valued the most, on average, while electricity-powered properties are the cheapest/lowest priced.

➔ We used electricity power as a reference category to help describe the results more clearly.

Regression Results

➔ The best predictor of sales price is the size of the living room.
  - As the square footage of the living room increases by one, the sales price increases, accordingly, by $302.

➔ Gas-powered homes are valued at a higher average, and when compared to electric-powered homes, increases the sales price by $97,000.

➔ The mean absolute error is $279,238, meaning any given prediction will be off by this amount.

Conclusions

➔ Consider renovating the size of the living room and the heating sources towards gas power in order to drive the sales price higher and sell the property for higher, as well.

➔ Further work could include:
- Research into location and its effect on sales price.
- Expanding the dataset to include more features.
- A method to include more of the outliers into the modeling.

Thank You!

Email: alaygt6@gmail.com
GitHub: @therookiescientist-andre
LinkedIn: linkedin.com/in/ak-layton/