



Learn Git and GitHub without any code!

Using the Hello World guide, you'll start a branch, write comments, and open a pull request.

[Read the guide](#)

 **dl-gd / flatiron-phase-2-project**

☆ 0 stars 🍴 0 forks

☆ Star

👁 Unwatch ▼

<> Code

! Issues

🔗 Pull requests

▶ Actions

📁 Projects

📖 Wiki

🛡 Security

🔗 master ▼



dl-gd final README ...

35 seconds ago

🕒 3

[View code](#)

README.md



Building A Pricing Model For First Time Home Buyers

Author: Dominic Garcia



King_County_WA

Overview

In this analysis, I inspect the King County House Sales dataset and iteratively develop a multiple regression model to analyze house prices.

Business Problem

There are lots of residents in King County, Washington who are considering buying their first home. These prospective buyers could benefit immensely from being able to accurately forecast the price of their first home based on a set of given parameters.

As an analyst for DLG Real Estate Agency, I am tasked with developing a regression model to help my fellow employees determine which homes are best for their clients.

Data

The dataset used contains house sale prices for King County, which were sold between May 2014 and May 2015. This dataset can be found in the project repository's Data folder.

Final Model

My final regression model, built to predict the price of a house in King County, can be evaluated using the following metrics.

- **R-squared = 0.808** : This value indicates that my model explains 80.8% of the variation in home prices from their mean value.
- **Root Mean Square Error (Using Test Data) = 114522** : On average, my model's predicted price is +/- \$114,522 from the home's actual value.
- **Q-Q Plot (see below)** : From the plot, it seems that the residuals produced by my model have a relatively Normal distribution within 2 standard deviations of the mean. However, there is some skew, especially towards the right. In other words, my model is generally best at predicting the prices of homes that cost up to about \$1,009,014.



Conclusions

My analysis leads to the following advice for any prospective first time home buyer in King County, WA:

- **Clients looking to save on their home purchase could start by looking in these zip codes: 98198, 98188, 98031, 98038, 98178, 98168 & 98058.** These area codes had the lowest relative impact on price in my final model.
- **Conversely, clients looking to make a bigger investment could start by looking in these zip codes: 98039, 98004, 98119, 98112, 98109, 98102 & 98040.** These area codes had the highest relative impact on price in my final model.
- **Clients looking to save money should consider the home's condition grade, as it seems to have a sizable impact on price.** Even going from an average grade to a high grade can increase a home's price significantly.
- **As expected, there seems to be significant correlation between a home's square footage & its price.** See scatterplot below.



Future Work

Given more time, I would look at the features I had to initially omit. This would involve using the *lat* & *long* columns to further inspect the impact of specific locations (beyond zip codes) on price. Additionally, I would look at the *sqft_living15* & *sqft_lot15* columns to get insight on what neighborhoods are most expensive to live in and the overall impact of comparative size of neighbors' homes.

For More Information

You can find more in-depth analysis in the [Jupyter Notebook](#), or get a cursory review through this [presentation](#).

For additional info, contact Dominic Garcia at dlgarcia.017@gmail.com

Releases

No releases published

[Create a new release](#)

Packages

No packages published

[Publish your first package](#)

Languages

● **Jupyter Notebook** 100.0%