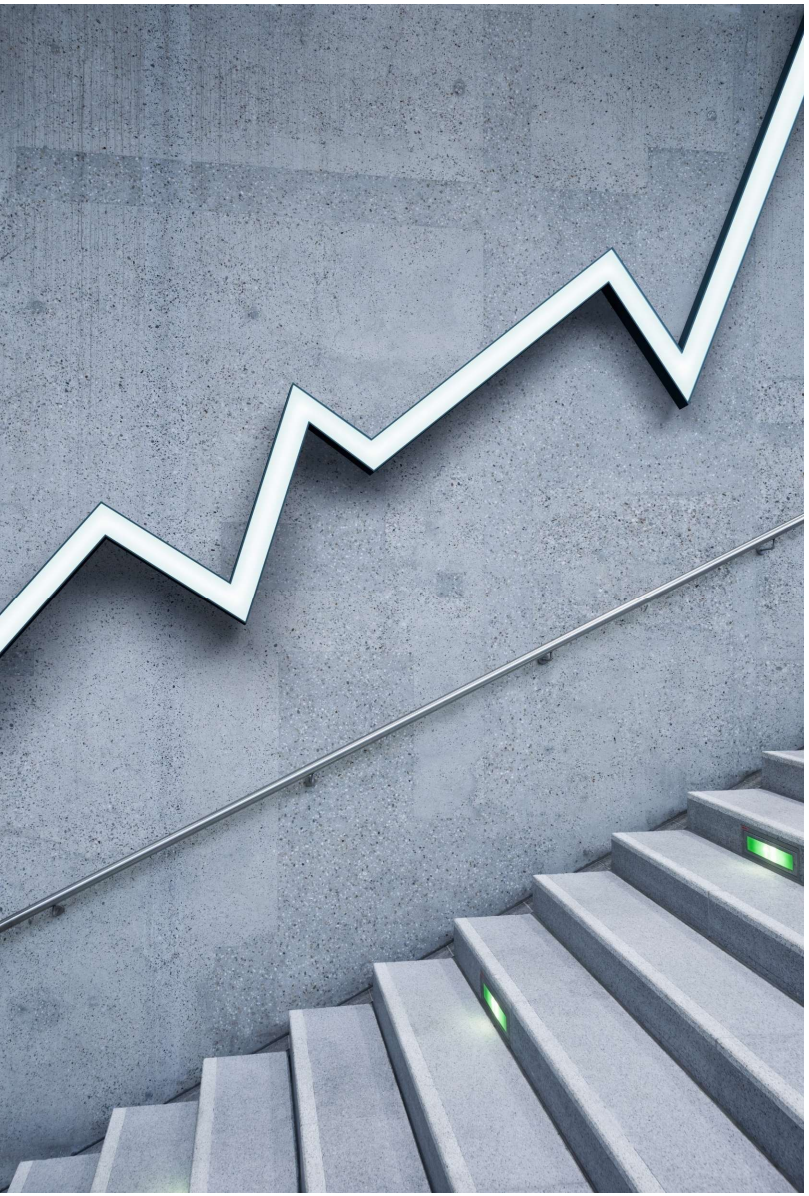


Launch a House Price Prediction Tool

RUTHY YAO

Summary

- A house valuation tool based on linear regression model for HomeConnect's property website.
- 9 input variables covering the basic building parameters and location features
- Goodness of fit of 65% – accommodating the requirement for both reliability and practicality.
- Three key insights from the price correlation with the housing features:
  - ✓ Waterfront houses can substantially drive the house price up.
  - ✓ Renovation can increase the house value.
  - ✓ House price is more correlated with the size of the living area than the land size.

Outline

- Business Problem
- The data
- Methodology
- Results and Insights
- Next Steps

Business Problem

- HomeConnect is planning to launch a house valuation tool for its residential property website aiming to enhance the website user experience and increase traffic.
- The SLT also want to gain some insight from the modelling on what matter the most for the house price so they can provide some value-add advices to their end customers.
- This is an initial trial. If it works, it will be launched officially on the website.

The Data

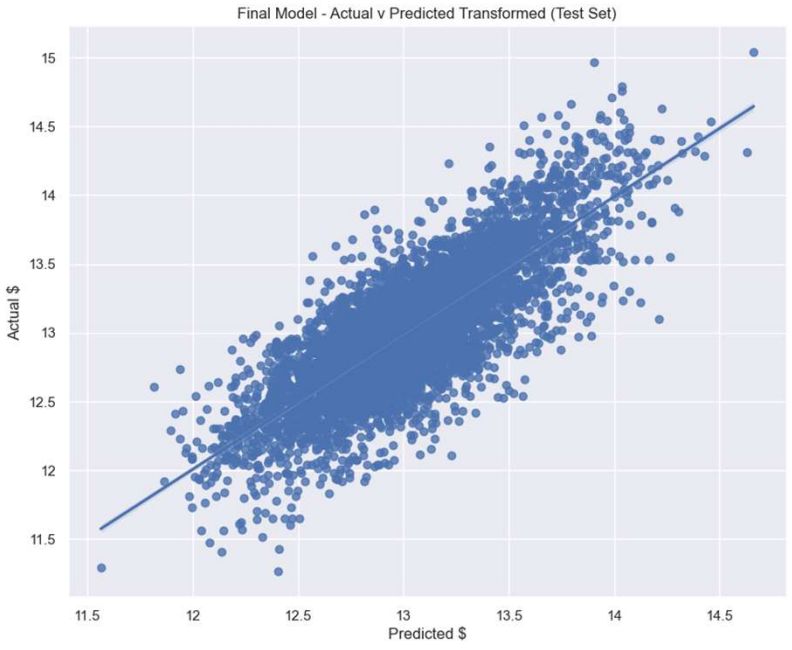
The dataset provides the house sold price and a wide variety of housing characteristics of the houses sold between 2014 and 2015 in King County, a northwest county in Washington State.

- Over 21,500 houses data
- Basic building parameters such as property size, floors, number of bedrooms, etc.
- Location features such as latitude and longitude.
- Housing condition and grade on construction quality.

Analytical Methodology

- Linear Regression modelling to identify the relationship between house price and the various independent variables
- Iteration approach to work out the best model.
- Recursive feature elimination to find the optimal number of features
- Train-test split to validate the model
- Evaluate the model fit:
  - Adjusted R-squared
  - Root mean square error (RMSE)

Results and Insights



Adjusted R-squared	65%
RMSE	$185k
No. of Features	9

Results and Insights

	input_variables	coefficient
0	sqft_living	268.174929
1	sqft_lot	1.215496
2	distance	-11542.310731
3	bedrooms	-43687.523173
4	bathrooms	22549.665787
5	floors	5327.534080
6	waterfront	689432.706624
7	is_renovated	90465.117861

- Houses at waterfront will see price up by $689k.
- Renovating the house can boost the house price by $90k.
- House price has higher correlation with the size of the living area compared with the size of the land.

Next Steps

- **Categorize the houses into metro, regional and rural** - Houses in different regions display distinctive features. Adding a region parameter will help refine the model and reduce the prediction error.
- **Bring in the national geospatial data to the model** - Incorporate a geospatial dataset of the whole country which is a stepping stone for the tool to be officially published on the website.
- **Incorporate condos or apartment housing data** - Adding different housing types will enrich the data and enhance the usability of the tool.

Thank you !

Email: zejia.yao@gmail.com

GitHub: https://github.com/RuthyYao