

Cribs Pricing

Team mates: Jun Tae Son, Omer Saif Cheema, Yusheng Zhu

Dataset: King County Home Sales (<https://www.kaggle.com/harlfoxem/housesalesprediction>)

The dataset comes from the official public records of home sales in King County, WA. There are 21613 instances and each represents a home sold from May 2014 to May 2015. The followings are the descriptions of each variable included in this dataset:

Dependent variable	Description
Price	Price of each home sold

Independent variable	Description
ID	Unique ID for each home sold
Data	Date of the home sale
Bedrooms	Number of bedrooms
Bathrooms	Number of bathrooms, where 0.5 counts for a room with a toilet but no shower
Sqft_living	Square footage of the apartments interior living space
Sqft_lot	Square footage of the land space
Floors	Number of floors
Waterfront	A dummy variable for whether the apartment was overlooking the waterfront or not
View	An index from 0 to 4 of how good the view of the property was
Condition	An index from 1 to 5 on the condition of the apartment
Grade	An index from 1 to 13, where 1-3 falls short of building construction and design, 7 has an average level of construction and design, and 11-13 have a high-quality level of construction and design
Sqft_above	The square footage of the interior housing space that is above ground level
Sqft_basement	The square footage of the interior housing space that is below ground level
yr_built	The year the house was initially built
yr_renovation	The year of the house's last renovation
zipcode	What zipcode are the house is in
lat	Lattitude
long	Longitude
Sqft_living15	The squarefootage of interior housing living space for the nearest 15 neighbors
Sqft_lot15	The square footage of the land lots of the nearest 15 neighbors

Problem description: The goal of the analysis is to generate the best fitted regression model for predicting home sale price in King county.

Proposed methodology: In this project, we will take the following 6 steps to analyze the dataset.

STEP1: Recode qualitative covariates and create dummy variables.

STEP2: Examine distribution of statistics to check normality assumption.

STEP3: Fit a full model and check for significant of coefficient estimates, multicollinearity, and regression assumptions.

STEP4: Apply variable selection procedures to find an optimal subset of predictors.

STEP5: Check for outliers and influential points.

STEP6: Write the final regression equation and interpret the coefficient estimates.

References:

Chica Olmo, J. (2007). Prediction of Housing Location Price By a Multivariate Spatial Method:

Cokriging. *Journal of Real Estate Research*, 29(1), pp.7-12. Available at:

http://pages.jh.edu/jrer/papers/pdf/past/vol29n01/05.91_114.pdf [Accessed February 27, 2017]

Glasserman, P. (2001). . Linear Regression Managerial Statistics. *Columbia Business School class materials*.

Available at: <https://www0.gsb.columbia.edu/faculty/pglasserman/B6014/Regression.pdf> [Accessed February 27, 2017]

Hu, G., Wang, J. and Feng, W. (2013). Multivariate Regression Modeling for Home Value Estimates with Evaluation Using Maximum Information Coefficient. *Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing 2012*, 443, pp.69-81. Available at:

<http://www.acisinternational.org/Springer/SamplePaper.pdf> [Accessed February 27, 2017]

Ng, A. (2015). Machine learning for a London housing price prediction mobile application. *Imperial College London*. Available at: http://www.doc.ic.ac.uk/~mpd37/theses/2015_beng_aaron-ng.pdf [Accessed February 27, 2017]

Pardoe, I. (2007). Modeling home prices using realtor data. *Journal of Statistics Education*, 16(2), pp.6-9.

Available at: <http://www2.amstat.org/publications/jse/v16n2/datasets.pardoe.pdf> [Accessed February 27, 2017]

Smith, M. (2014). Case 12: Housing Prices Multiple Regression – Multicollinearity and Model

Building. *University of Colorado Denver Business School*., pp.3-9. Available at:

<https://www.jmp.com/content/dam/jmp/documents/en/academic/case-study-library/case-study-library-12/business-case-studies/12-housingprices.pdf> [Accessed February 27, 2017]