

Springboard Capstone 1 Proposal - Beijing House Pricing Prediction

Nancy Mao

Problem

The Beijing housing market is highly competitive. Figuring out the best selling price is the number one decision to make when someone is going to sell a house. The purpose of the project is to build models to predict house pricing in Beijing China.

Potential Client

Real estate agencies can be the potential clients. A house pricing prediction model with good performance would be very valuable for a real estate agent who could make use of the information provided on a daily basis to decide the house listing price for his or her own clients.

Source of data

[Kaggle](#) Housing price of Beijing from 2011 to 2017, fetching from Lianjia.com

It includes URL, ID, Lng, Lat, CommunityID, TradeTime, DOM(days on market), Followers, Total price, Price, Square, Living Room, number of Drawing rooms, Kitchen and Bathroom, Building Type, Construction time. renovation condition, building structure, Ladder ratio(which is the proportion between number of residents on the same floor and number of elevators of ladder. It describes how many ladders a resident has on average), elevator, Property rights for five years (It's related to China restricted purchase of houses policy), Subway, District, Community average price.

Approach

1. Data Cleaning and Preparation
 - a. Missing data visualization and imputation
 - b. Variable conversion
2. Exploratory Data Analysis
 - a. Conduct univariate & multivariate analysis to understand how each features is distributed and correlated to the house price and how features correlated to each other
 - b. Inferential Statistics: conduct hypothesis testS to understand which features are significantly correlated to house pricing

- c. Time Series Analysis
- 3. Machine Learning Modeling
 - a. Apply 7 different models to get initial modeling results.
 - i. Linear Regression
 - ii. Ridge Regression
 - iii. Lasso Regression
 - iv. ElasticNet Regression
 - v. Random Forest
 - vi. K-Nearest Neighbors
 - vii. Gradient Boosting
 - b. Evaluate each model and decide models that might be more suitable.
 - c. Add more features iteratively to get better prediction

Deliverables

- 1. Code
 - data cleaning
 - data exploration analysis & data storytelling
 - machine learning model
- 2. Presentation Slide Deck