
Singapore Housing Prices

The goal of this Kaggle InClass challenge is to predict housing prices in Singapore.

A house represents the biggest investment for most households. Whether you're a home owner or a tenant, your mortgage payments or monthly rent will likely consume a significant portion of your income. Meanwhile, with some of the highest property prices in the world, real estate as an asset class poses an out-sized impact on the macroeconomic environment in Singapore.

Therefore, being able to accurately value current housing prices will not only facilitate market transactions by providing valuable guidance for all market participants (be it home owners, home buyers, landlords, tenants or banks that underwrite mortgages), but also provide useful insights for policy makers and government authorities in understanding the current state of the economy.

This is, however, an incredibly difficult endeavor. Real estate is notoriously known for its illiquid nature, owing to the fact that every single home is unique; there's literally only that ONE home at that exact location! Your goal in this challenge is to get as close to actual transaction prices as possible.

The competition is worth 20% of the course grade. You should participate either individually or in a team of two in the competition.

1 How To Participate

1.1 Accessing Competition Page

The competition has been hosted on Kaggle and can be accessed through following url : <https://www.kaggle.com/t/5e4e16a28d704f6591f1043769a6817a>. The training and test data for the competition can be accessed through the "Data" tab.

1.2 Software

You may use any software to do the learning and predictions. Scikit Learn has many algorithms. For deep learning, you can consider packages such as Keras or Apache Singa, Tensorflow, Theano, etc. You may also write your own learning and prediction programs.

1.3 Making a Submission

After you have trained your models using training data, you can make a submission on Kaggle through "Make a Submission" tab. You can find details about submission file format in the "Evaluation" section under "Submission Format". A sample submission file has also been provided in the data section for your help.

Before you make a submission make sure that your team name has following format (Your team name)-(Student id1)-(Student id2). The student ids are required for correct grading. Team name can be set by going to the "My Team" tab.

1.4 Leaderboard

Public leaderboard will show the score on part of the test set. You can see the rank of your trained model through it. The rankings on Private Leaderboard might be slightly different as it ranks the performance of your predictions on a different part of the test set.

2 Grading

The competition closes at midnight Wednesday 11 April UCT time. You are also required to submit a 5 page report describing your method. The report should be uploaded to the IVLE workbin by midnight Thursday 12 April Singapore time. You will be evaluated based on the scores on the Private Leaderboard and the report.

In addition, you are required to submit your source code to the IVLE workbin. Your assignment is considered incomplete if the source code is not submitted.

3 Prizes

This prediction competition is held in collaboration with UrbanZoom (<https://urbanzoom.com/>). The data for the competition is provided by UrbanZoom.

In addition, UrbanZoom is also providing the following prizes:

- First prize: \$400
- Second prize: \$250
- Third prize: \$100
- Technical award: \$250

The first three prizes are based on the private leaderboard score. The technical award based on interestingness and quality of technical work as described in the report. The winner of the first three prizes are also be eligible for the technical award.

4 How to do Well

Understand the data. Like most Kaggle competitions, understanding the problem and doing feature engineering will most likely to get you a good score.

That said, a better model may potentially perform better if there is enough data available – that is one reason we have a technical prize. What are some of the interesting properties of the data for building more interesting models? You are given data from the past and are trying to predict data from the future – the data is not i.i.d. but comes from a time series. Predicting the next value, given

a window of data from the recent past or using a recurrent neural network comes to mind. But there is more than that. There is spatial information associated with each property. Convolutional neural networks has worked well on many spatial (and temporal) problems, so you may want to think about using that. Unfortunately, the data is not on a regular grid: a regular CNN may not work and you may have to go further. Furthermore, most of the data is actually missing, prices are only observed when properties are sold! There is also data from HDBs as well as private housing. These have similarities as well as differences. Can you exploit them appropriately?