

Contents

I. Problem Description.....	2
II. The datasets are provided as cited below.....	2
III. Tasks:	3
Main task :	3
Visualization Tasks & others:.....	3
IV. Evaluation Metric:	3
IV. For Information:	3

I. Problem Description

Predict the Flat Resale Prices in Singapore

Investing in Flats is the biggest investment for most households in cities like Singapore which is very small but heavily urbanised. Therefore, being able to accurately value the flat 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.

In this competition, you are expected to create an analytical and modelling framework to predict the flat resale prices based on the quantitative and qualitative features provided in the dataset while answering other questions too cited below.

II. The datasets are provided as cited below:

To build the model and tune the model & visualization

- i. train1.csv
- ii. train2.csv
- iii. train3.csv
- iv. building_to_mrt_distance.csv (distance in meters) - Can be used for both train and test

For the test prediction & tool submissions

- v. test.csv
- vi. samplesubmission.csv (just for reference on tool submission format for the test predictions)

Target attribute : "resale_price"

III. Tasks

Main Task:

In this competition, you are expected to create an analytical and modelling framework to predict the flat resale prices based on the quantitative and qualitative features provided in the dataset while answering other questions too cited below. You may derive new features from the existing features and from the domain knowledge, which may help in improving the model efficiency.

Visualization Tasks & others:

1. Exploratory Data Analysis using visualizations in R Notebook or Jupiter notebook format

(Train data should be used for this task)

- List down the insights observed from the visualizations
- Explain the impact of most important attributes on target attribute observed from the visualizations.

2. Based on domain and data understanding, which attributes do you intuitively think would impact the flat resale price? Using the ML methods, which attributes do you find as important?

3. Is there any Bias or variance problem? If yes, how did you address it?

IV. Evaluation Metric:

- Consider the 'RMSE' as the error metric for regression task to tune the model.

V. For Information:

- Singapore Flats have 99-year lease.