
Machine Learning

Sberbank Russian Housing Market

TAs
ntu.mlta@gmail.com

Task description

- In this challenge, complex interactions between housing features such as number of bedrooms and location are used to make pricing predictions .



<https://www.kaggle.com/c/sberbank-russian-housing-market>

 Featured Prediction Competition

Sberbank Russian Housing Market

Can you predict realty price fluctuations in Russia's volatile economy?

\$25,000
Prize Money

 Sberbank · 685 teams · 2 months to go (2 months to go until merger deadline)

[Overview](#) [Data](#) [Kernels](#) [Discussion](#) [Leaderboard](#) [More](#) [My Submissions](#) [Submit Predictions](#)

Overview

Description

Evaluation

Prizes

Housing costs demand a significant investment from both consumers and developers. And when it comes to planning a budget—whether personal or corporate—the last thing anyone needs is uncertainty about one of their biggets expenses. [Sberbank](#), Russia's oldest and largest bank, helps their customers by making predictions about realty prices so renters, developers, and lenders are more confident when they sign a lease or purchase a building.

<https://www.kaggle.com/c/sberbank-russian-housing-market>



The image is a screenshot of the Kaggle competition page for 'Sberbank Russian Housing Market'. The header features a trophy icon and the text 'Featured Prediction Competition'. The main title is 'Sberbank Russian Housing Market' with the subtitle 'Can you predict realty price fluctuations in Russia's volatile economy?'. Below this, it says 'Sberbank · 685 teams · 2 months to go (2 months to go until merger deadline)'. On the right, a red box highlights the prize money '\$25,000 Prize Money', with three red arrows pointing to it from the right. The navigation bar includes 'Overview', 'Data', 'Kernels', 'Discussion', 'Leaderboard', 'More', 'My Submissions', and 'Submit Predictions'. The 'Overview' section is active, showing a 'Description' tab selected. The description text reads: 'Housing costs demand a significant investment from both consumers and developers. And when it comes to planning a budget—whether personal or corporate—the last thing anyone needs is uncertainty about one of their biggets expenses. Sberbank, Russia's oldest and largest bank, helps their customers by making predictions about realty prices so renters, developers, and lenders are more confident when they sign a lease or purchase a building.'

Featured Prediction Competition

Sberbank Russian Housing Market

Can you predict realty price fluctuations in Russia's volatile economy?

Sberbank · 685 teams · 2 months to go (2 months to go until merger deadline)

\$25,000
Prize Money

Overview Data Kernels Discussion Leaderboard More My Submissions **Submit Predictions**

Overview

Description

Evaluation

Prizes

Housing costs demand a significant investment from both consumers and developers. And when it comes to planning a budget—whether personal or corporate—the last thing anyone needs is uncertainty about one of their biggets expenses. Sberbank, Russia's oldest and largest bank, helps their customers by making predictions about realty prices so renters, developers, and lenders are more confident when they sign a lease or purchase a building.

Data Files

- **train.csv, test.csv** : information about individual transactions.
- **macro.csv** : data on Russia's macroeconomy and financial sector
- **data_dictionary.txt** : explanations of the fields available in the other data files

Taining & Testing data

- 292 attributes per transcation
- continuous & discrete attributes
- including information of a house and local area.
- you can see explanation in data_dictionary.txt
- macro.csv could be joined by timestamp

Evaluation - RMSLE

$$\epsilon = \sqrt{\frac{1}{n} \sum_{i=1}^n (\log(p_i + 1) - \log(a_i + 1))^2}$$

n is the total number of observations in the (public/private) data set,

p_i is your prediction,

a_i is the actual response for i .

$\log(x)$ is the natural logarithm of x

Evaluation - RMSLE

$$\epsilon = \sqrt{\frac{1}{n} \sum_{i=1}^n (\log(p_i + 1) - \log(a_i + 1))^2}$$

n is the total number of observations in the (public/private) data set,

p_i is your prediction,

a_i is the actual response for i .

$\log(x)$ is the natural logarithm of x

- RMSLE penalizes an under-predicted estimate greater than an over-predicted estimate.

Characteristics of this task

- More attributes and instances of data
- Economic prior knowledge
- Prize money!!!!!!!!!!!!!!

Reminder

- Some missing labels
- Seperate discrete and continous attributes first.