

Mortgage PD scorecard

Files:

[Mortgage_sample.csv](#)

Modelling dataset with data about 50000 US mortgages

[Mortgage_metadata.xlsx](#)

Data dictionary for the mortgage dataset

Sample information:

The sample population comes from data on US residential mortgages and is provided by International Finance Research¹. Please note that it may be used for academic purposes only and any commercial use is strictly prohibited.

For most of the observations, you can see the target (default), but we have kept a “private” sample. This corresponds to real-life, where the validation function would keep an out-of-sample/out-of-time dataset. We leave it to you to consider if any analysis should be done on this sample 😊

Goal:

As indicated in the Credit Risk presentation, your task is to build a PD scorecard using the provided data. The goal is to create a model that will predict a probability of default for each mortgage.

The presentation contains an overview of a proposed modelling process and some considerations to consider when developing and assessing the model.

You will be assessed on the “good modelling practice” you employ. Remember, the best model is not necessarily the one with the highest performance metric. Your goal should be to build a scorecard with enough discriminatory power, but the steps taken during the modelling process are most important.

If you have any questions, please contact jan.nusko@cz.ey.com.

¹ The periods have been anonymized and the data has been processed by the Credit Risk Analytics team (<http://www.creditriskanalytics.net>)