

Loan Repayment Challenge

1 Introduction

At MoneyLion we are constantly working to assess the risk of our applicants more accurately. Being successful in this objective enables us to better price customers and mitigate losses on our portfolio of loans. The following challenge asks you to work with a data set of loan repayment. It is intentionally meant to be open-ended. The point is not to arrive at a predetermined answer or search for the lowest possible standard error. Rather, the hope is that it will force you to ask relevant questions about the data, do some preliminary exploration, perform the necessary manipulations or aggregations, generate visualizations, and reach conclusions or insights. The most important thing to remember is that we are evaluating your thought process and ideas! The more you explain your thinking, in a clear and succinct manner, the better. If you get stuck, describe what additional information or data you might look to collect, and trying a different idea is highly encouraged.

2 Data

You are provided with 3 files: `loan.csv`, `payment.csv`, and `clarity_underwriting_variables.csv`. The files are comma-separated, with the column names in the first row. The detailed description of each column is available in the MoneyLion Data Scientist Assessment Data Dictionary document. This can be found in the `dictionaries.zip` file.

3 Rules

You may use any language, packages, or external libraries for the challenge, though Python or R are preferred. An IPython notebook might be the best way to show your code and write your comments/thoughts to follow along. You may refer online for technical reasons but **DO NOT** directly use any help from other people, sources, online forums, etc., your submission should be solely your ideas and work. There is no hard limit on the amount of time that can be spent on this challenge. Please disclose how much time you spent on the challenge when you've completed it.

4 Deliverables

A zip file with contents grouped into the following sub-directories (you may omit empty directories):

- Proposal / Plan on how you would approach the problem
- data (only include new data files if you have any, **DO NOT** send back to us the assessment data!) [Optional]
- notebooks containing the data cleaning and a bit of EDA that justify your proposal or plan (we appreciate if you include an HTML format of your notebook as well as your notebook) [Optional]
- code (if you have some of your own helper functions you use in your assessment) [Optional]

5 Guidelines

In this assessment, we do not expect a model built (feel free to attempt if you want to), but we want to see a proposal or plan around how you would approach the problem of predicting loan risk or quality (loan repayment) on a given loan application. Before diving into developing a plan, start with some exploratory data analysis (EDA). Take a look at some distributions of some of the variates. Think about how would you define a loan as risky. How would you select your sample if you were to build

this model? Back your proposal with some EDA and insights you discovered from the data.

6 Tips

Here are some of the skills we are looking for:

- What insights can you get from the data, did you try to understand it?
- Data preparation, exploration, and visualization.
- Clearly communicate your thought process, you do not want your reviewer to scratch their head figuring out what you are doing.
- Back your assumptions with evidence.