Data Scientist Take Home

The Data

The dataset, and column definitions, can be found here: https://archive-beta.ics.uci.edu/ml/datasets/default+of+credit+card+clients.

Note that the dataset is quite small and can easily be stored in memory in its entirety.

The Task

Build a model to predict the likelihood of default, a model which can generalize to other data points. The column to be predicted is `default payment next month`.

Put your code in a Jupyter notebook. Be sure to generously intersperse Markdown cells in your notebook, in order to explain the steps you are taking and any assumptions you are making *(in words)*. Breaking down the larger task into smaller, more concrete steps is just as important to us as your final trained model and performance metrics, as it reveals your thought process!

For any functions that you write, please include clear and fleshed out docstrings. Also, please attempt to adhere to PEP-8 code formatting standards throughout your notebook.

There is no need to make slides to accompany your notebook. If you're invited to a virtual onsite, we'll ask you to walk us through your modeling exercise, and you can present your work *directly from your notebook*. (In this case -- but not in others -- your audience will be sufficiently code-savvy to respond well to this form of presentation.)

If you have any questions about this exercise, feel free to ask. But we encourage you to try answering them yourself. How you choose to interpret this task gives us additional insight into your thinking. Also, you'll notice that for a few of the columns in the dataset, some commonly occurring values lack definitions in the documentation; we are interested in seeing how you deal with this incompleteness of information.

Finally, once you have trained your model, please think of ways its predictions can be used in Credit Sesame's product. Be prepared to have a conversation about your envisioned use cases.

To submit your notebook, please provide us with a link to your notebook in GitHub. It's suggested that you spend no more than 2 to 4 hours on this exercise, but of course the amount of time spent will vary from person to person. We pledge to get back to you within *one week* of your submission and let you know if you've advanced to the "onsite" panel interview. Thank you so much for your time and effort working on this!