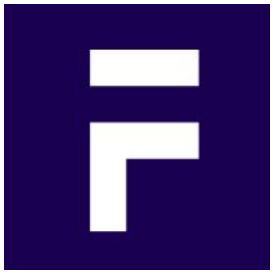

Thank you for taking the time to complete the **Figure Data Science Challenge**!

The purpose here is to assess your problem solving skills and data science modeling expertise. Please read the instructions below carefully to make your modeling process easier.

The data challenge is to build a loan default prediction model using publicly available loan data. Ideally we would like to have this completed over the next **4 days.**

Details about the dataset:

- The dataset that will be used for this challenge is Freddie Mac Single Family Loan-level Dataset. (When you create your account, you might need to check your spam folder if you don't see it immediately in your email). You can download the dataset at: http://www.freddiemac.com/research/datasets/sf_loanlevel_dataset.html
- There is a user-guide for this dataset, and it has 33 pages. To save your time, you can go directly to Page 7 - 17 and read the other pages when necessary. User-guide of the dataset can be found at: http://www.freddiemac.com/fmac-resources/research/pdf/user_guide.pdf
- The loan-level dataset has two downloadable sets: quarterly or annually. You can choose either one for next steps. To save your time, you can use the 2009 Q1 dataset for the model. The 2009 Q1 dataset includes data of loans originated in 2009 Q1.
- Within the dataset, you can find loan origination data and monthly performance data. You can join the origination data with the monthly data using 'loan sequence number'.

Questions:

1. Use the dataset and build a model to predict the probability of default at the time of origination. Here the definition of default is 90 days or more delinquent. The delinquency status is captured by 'Current Loan Delinquency Status' in the monthly performance data. In addition to delinquency status, if the 'zero balance code' in the performance data shows value '03', '06', or '09', the loan is also considered in default.

2. The output of the model you trained in the previous step is the probability of default. However, in some instances one needs both a prediction regarding the likelihood of default, as well as, the timing of the default. Propose and train an alternative model that also delivers an estimate of the time-to-default.

3. (Bonus question): Other than default, loan terminations can also be caused by prepayment. In other words, borrowers can pay off their loan before it matures. How would your solution to the previous question change in the presence of prepayment risk? Please describe.

Deliverables:

- Please send us your results, including your code, when you've completed the project
- Give a 45 minute presentation on the modeling process at the onsite interview. You can choose the format of the presentation (PPT or Jupyter Notebook, etc)
- If anything is unclear, please state the assumptions you are making and explain why they are justifiable.

We look forward to discussing your results!