|  |
| --- |
| Exciting news! Kiva's next chapter | Kiva |

We envision a world where all people hold the power to create opportunity for themselves and others.

from the Kiva Website

**Microfinance Funding**

A Datascience Capstone Project

# **Background:**

Kiva is a non-profit that provides an online platform that connects individual lenders to low-income entrepreneurs around the world to finance microloans. Each borrower has a profile page that shares their story with a picture, information about how the loan will be used, the amount requested and details about the loan (length, repayment schedule, funding model, etc.). The loans are posted for approximately 30 days and during that time lenders can contribute in $25 increments until the loan is fully funded.

Currently, the median loan is for approximately $500, funded by 14 individual lenders and 95% of the loans are fully funded during the funding window. Each loan has characteristics that make it potentially appealing to different lenders. Distinguishing features include where the borrower is located, borrower demographics, how the loan will be used, and loan characteristics. Kiva would like to better understand what makes a loan more attractive to lenders and predict what loans are at risk of not being funded.



# **Problem Statement:**

How can Kiva anticipate which microloans are at risk of not being funded allowing for strategic marketing and better feedback to borrowers?

Is it possible to predict which microfinance loans will be funded and how quickly, considering borrower location, borrower demographics, loan purpose, and loan structure? Creating this model will allow Kiva to proactively help borrowers structure their loans as well as provide strategic placement (marketing) of loans on their platform.

# **Data & Problem Approach:**

The data for this project comes from Kiva via their developer tools website which produces a nightly snapshot of detailed information about all loans since 2006. [<https://www.kiva.org/build/data-snapshots>] The entire dataset is over 2M records but will be divided into smaller subsets for model development. Two separate approaches will be considered: utilizing the most recent 6 months of loans as well as taking 100,000 random loans from the entire dataset.

The approach for this project will be start with reducing the dataset to a manageable size and cleaning the data. It is anticipated that additional features will be created to summarize some of the text heavy fields. During exploratory data analysis the data will be visualized and high-level feature correlation will be performed to identify the most promising predictive characterstics about the loan. Finally, a variety of models will be developed, tuned, validated and evaluated.