Practical Data Science

**Project Proposal – Analyzing Kiva Loan Data**

Team # 3

Suresh Rangarajan

Charlotte George

David Cohodes

November 12, 2012

**Data Source**

Kiva

URL: <http://build.kiva.org/>

Kiva provides a RESTful Web-service API for accessing data on lenders, loans and other related Kiva objects. Kiva API returns the response in XML and JSON formats depending on the URL accessed. We are planning to use the JSON format in our project. Specifically, we plan on utilizing Lenders, Teams and Loans objects with the KIVA API.

While there are several other apps already developed using this API, we plan to use the snapshot data provided by Kiva as our training data set to predict certain outcomes rather providing a customized interface to view the data.

**Questions we plan to ask and ideas planned for analysis and expected results**

1. Kiva runs promotion campaigns to incentivize lenders to make loans. We would like to identify patterns in the existing data to make these campaigns more effective.
   1. For example, we would try to map the locations of the lenders and the borrowers of loans and identify patterns to promote specific borrowers’ loans to lenders. Here, visualization may help a campaign sponsor to make subjective decisions.
   2. Map the demographics of the lenders and borrowers to target the promotion campaigns effectively.
2. Build a model that calculates certain probabilities:
   1. Probability of a loan (Kiva loan score?) being fully funded given the industry, location and other relevant parameters.
   2. Probability of a borrower receiving a loan given an industry, location and other relevant parameters from a specific lender
3. Where are the most loans made
   1. Aggregate data by country and potentially city
   2. Graphically plot where the most loans are made on a map
4. Analyze the demographics of borrowers to find out the following:
   1. Look at whether specific industries attract more loans in general.
   2. Are there specific triggers in an individual’s profile that make them more likely to get a loan (specific key words, location, industry, age)?
5. Analyze the demographics of lenders to find out the following:
   1. For a lender who has registered with Kiva but has never made a loan, what is the best approach to help him make a loan?
   2. For a lender who has not made a loan recently, what would be the best approach to help him make a loan again?
6. Analyze the data on teams and loans to determine the following:
   1. Which teams make the most loans?
      1. Analyze data set to sort loan volume by team and aggregate this information
      2. Graph teams and corresponding loans
   2. Is there anything in common between the teams that make the most loans?
      1. Use real expressions to see if there are any key words that tip off whether a specific group will make loans
   3. Analyze whether there are indicators as to whether a specific team will make multiple loans to the same individual
   4. Analyze whether the number of members of team has any impact on the volume of loans made.