Jeremy Mwenda

CS591 Project 4/19/2015

For my project, I will be analyzing data from Kiva.org, a non-profit organization that enables people to lend money to enterprenuers in underserved parts of the world. Kiva works with field partners in various countries to administer the loans. The field partners are organizations such as microfinance institutions, schools, social businesses, and non-profit organizations trusted by Kiva. Although, both Kiva and its lenders do not charge interest on the loans, the field partners are allowed to charge an interest on the loans to cover their operating cost. A borrower applies for a loan through a field partner, the field partner then evaluates the borrower and the reason for the loan to decide whether to approve the loan. If approved, the field parner posts the loan on Kiva's website where lenders can see the loan and contribute funds toward the loan. When enough funding has been provided by lenders, the amount of the loan is then released to the field partner who then issues it to the borrower. The borrower pays back the loan on installment terms pre-agreed upon.

The dataset contains a list of loan details downloaded from Kiva.org.

The following are some of the important attributes of a loan:

Loan ID	Town/City & Country
Borrower Name	Posted date
Status – fundraising, funded, disbursed, paid, defaulted, expired, refunded	Funded date
Loan amount	Disbursal date
Activity	Amount funded
Sector e.g. agriculture, retail, clothing	Borrower gender
Use – short description on how loan will be used	Number of borrowers (for group loans)
Deliquent	Repayment interval
Field partner ID	Translator – if borrower doesn’t speak english
Forex loss – if lender made loss due to currency exchange rate	Lender count – if loan was funded by multiple people

Basic Statistics on the data

Number of loans: 89,492

Number of loans by Gender of the borrower:

Female	67,144
Male	22,348

Average loan amount (in USD) for each Gender:

Female	730.86
Male	858.92

Loan amount aggregates in US dollars:

Mean	762.84
Median	550
Min	25
Max	49800
Stdev	761.78

Repayment rate: number of loans paid / (loans defaulted + loans paid):

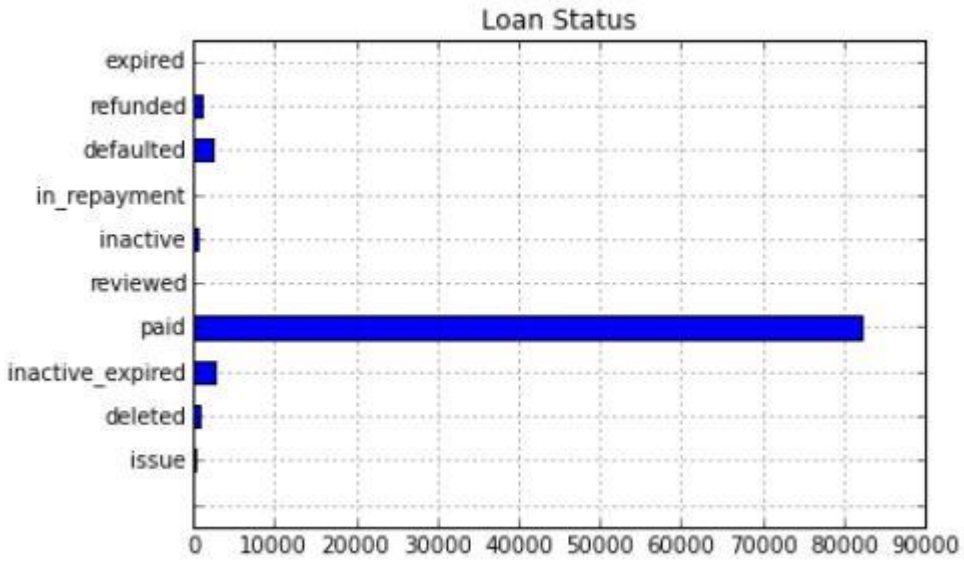
97.19%

Reported Kiva repayment rate for all loans to date (from Kiva.org website):

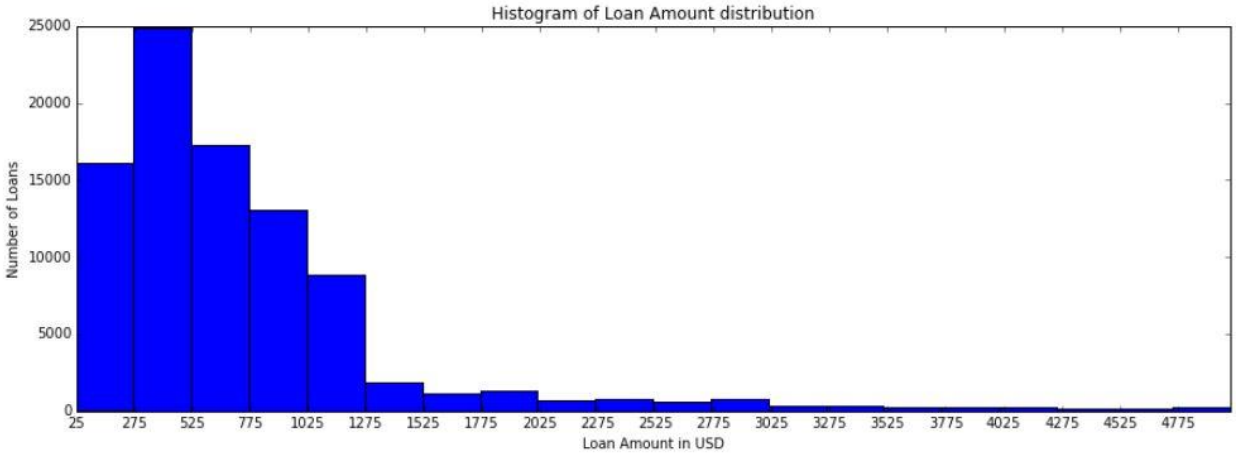
99.01%

...next page

Below is a graph of loans statuses. A majority of the loans are in 'paid' status.

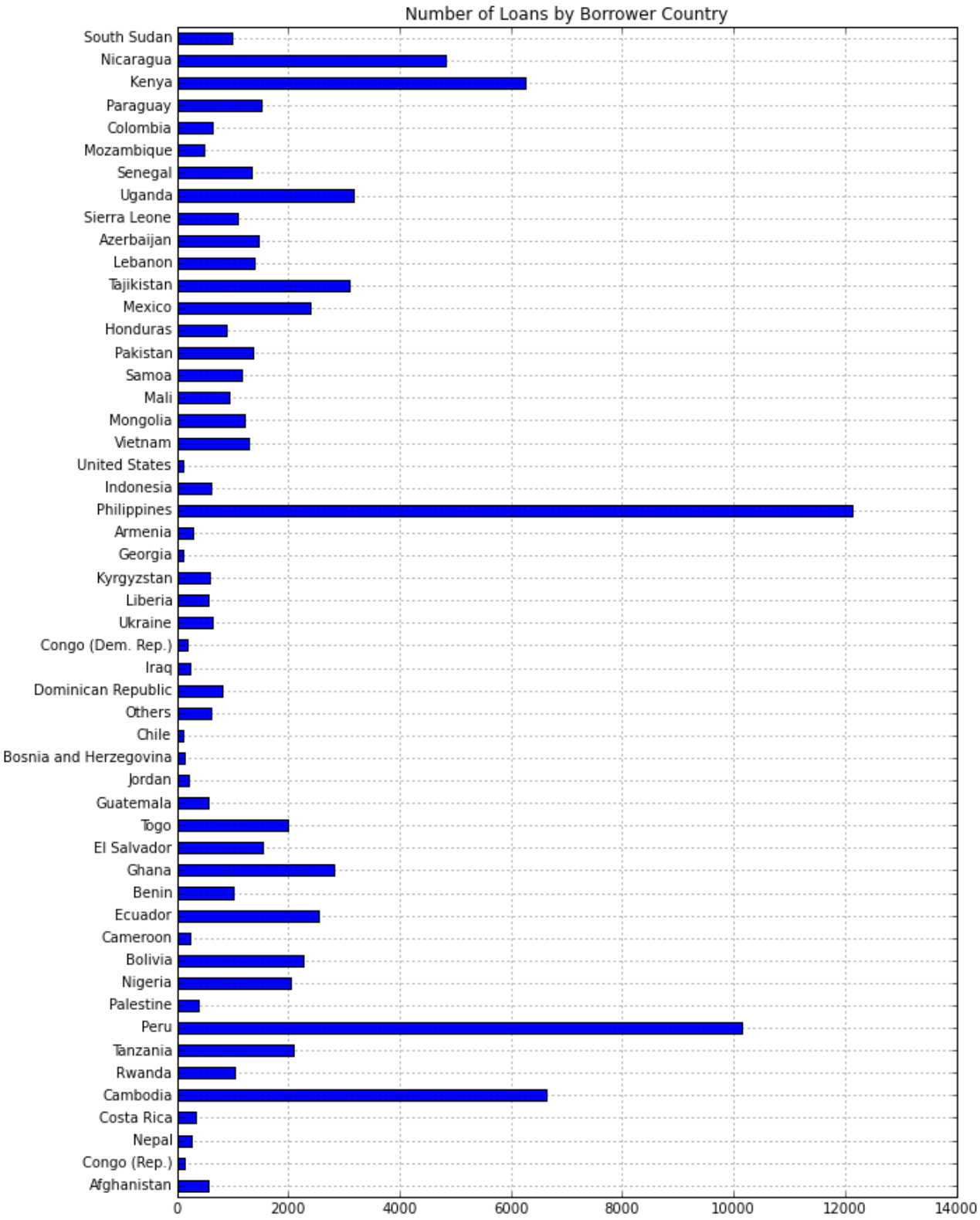


Most of the loans are below USD 1000. Below is a histogram of loan amount distribution:

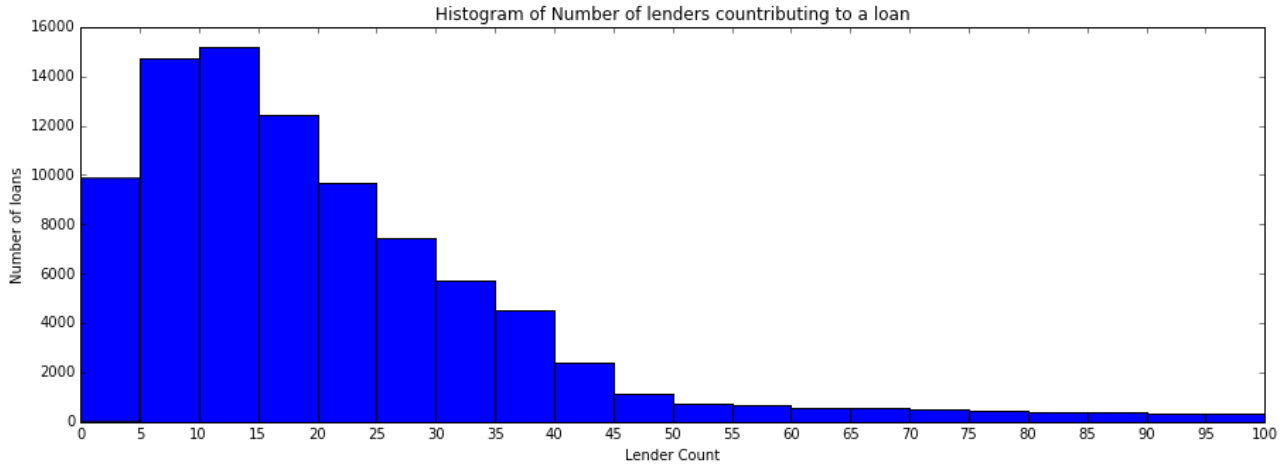


...next page

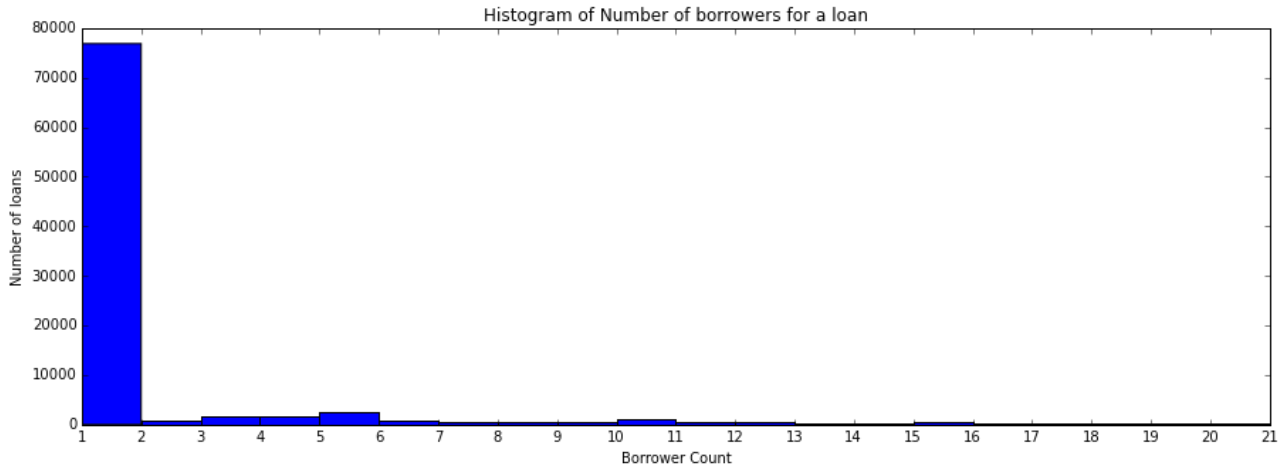
Most of the loans went to developing countries, with Philippines, Peru, Cambodia, and Kenya receiving the most loans:



Most loans had multiple lenders. Kiva allows a minimun contribution of $25 by a lender toward funding a loan. The averange lender count per loan is 21. Below is a histogram of the lender count distribution:

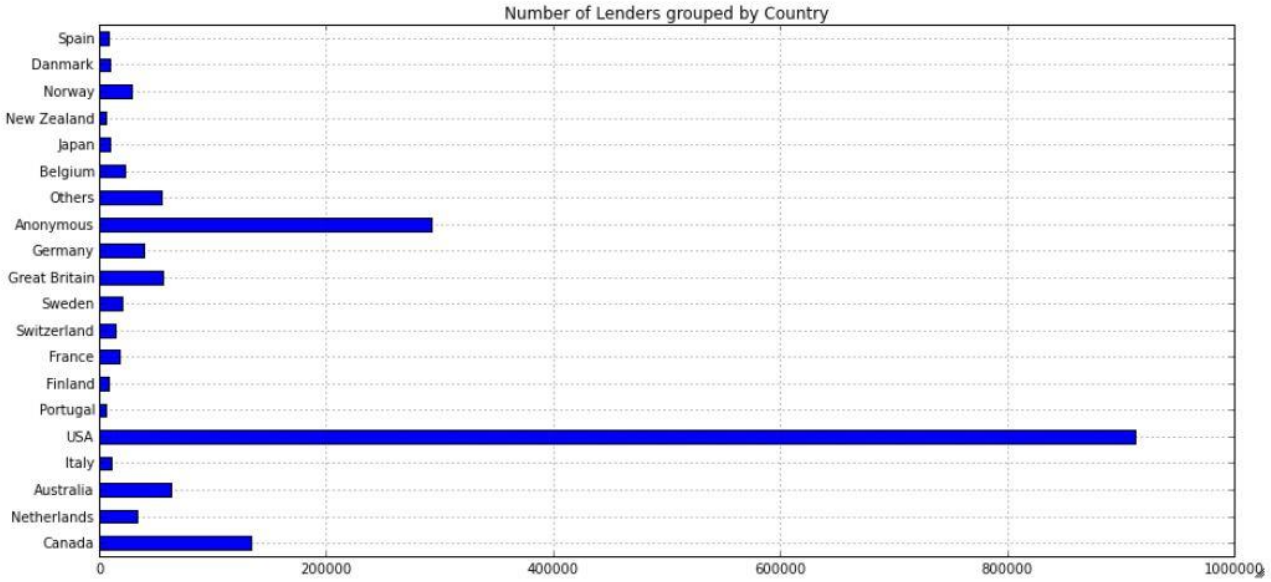


Kiva allows both individuals and groups to borrow a loan. Most loans were issued to individuals. Below is the borrower count distribution:



...next page

Most of the lenders came from the United States, which is where Kiva.org started.



Hypotheses to test

- The probability of defaulting on a loan is positively correlated to loan amount. I will use a logistic regression model to test this.
- Loan applications with longer descriptions are more likely to be funded than those with short descriptions. The loan description contains among other details, the activity that the borrower needs the loan money, and explains how the borrower will be able to repay the loan. I will test this using a logistic regression model.
- Loan applications with a picture of the borrower attached are more likely to be funded that those without.
- I will also look at various sectors that are funded, to determine which sectors are more likely to be funded. Sectors funded include retail, agriculture, food, clothing, etc.
- I will also analyze the texts of the description to try and find if the probability of a loan getting funded is correlated to the 'tone' of the description. Using some words in the description such as words that portray the borrower as hard-working or words that elicit empathy, may increase the likelihood of getting a loan funded.

Data collection

I downloaded a snapshot of loan data from Kiva.org. This data is in JSON format.

I retrieved lender data for the loans from Kiva lenders API. This data is also in JSON format.

I used python to analyze and calculate the basic statistics on the data.

References

http://www.kiva.org/

http://en.wikipedia.org/wiki/Kiva_(organization)