**Predicting the Probability of a Loan being Crowdfunded on Kiva**

**What is Kiva?**

Kiva Microfunds is a non-profit organization that allows people to lend money via the internet to low-income entrepreneurs and students throughout the world. Kiva's mission is "to connect people through lending to alleviate poverty." Since 2005, Kiva has crowd-funded more than a million loans, totaling more than $1 billion. The Kiva platform has attracted a community of well over a million lenders from around the world.

**How does the business model work?**

Kiva is a crowdfunding website where people all over the world can make interest free loans to entrepreneurs worldwide. These Funds are posted and disbursed by Kiva’s field partner organizations worldwide.

**What is the problem we are trying to solve?**

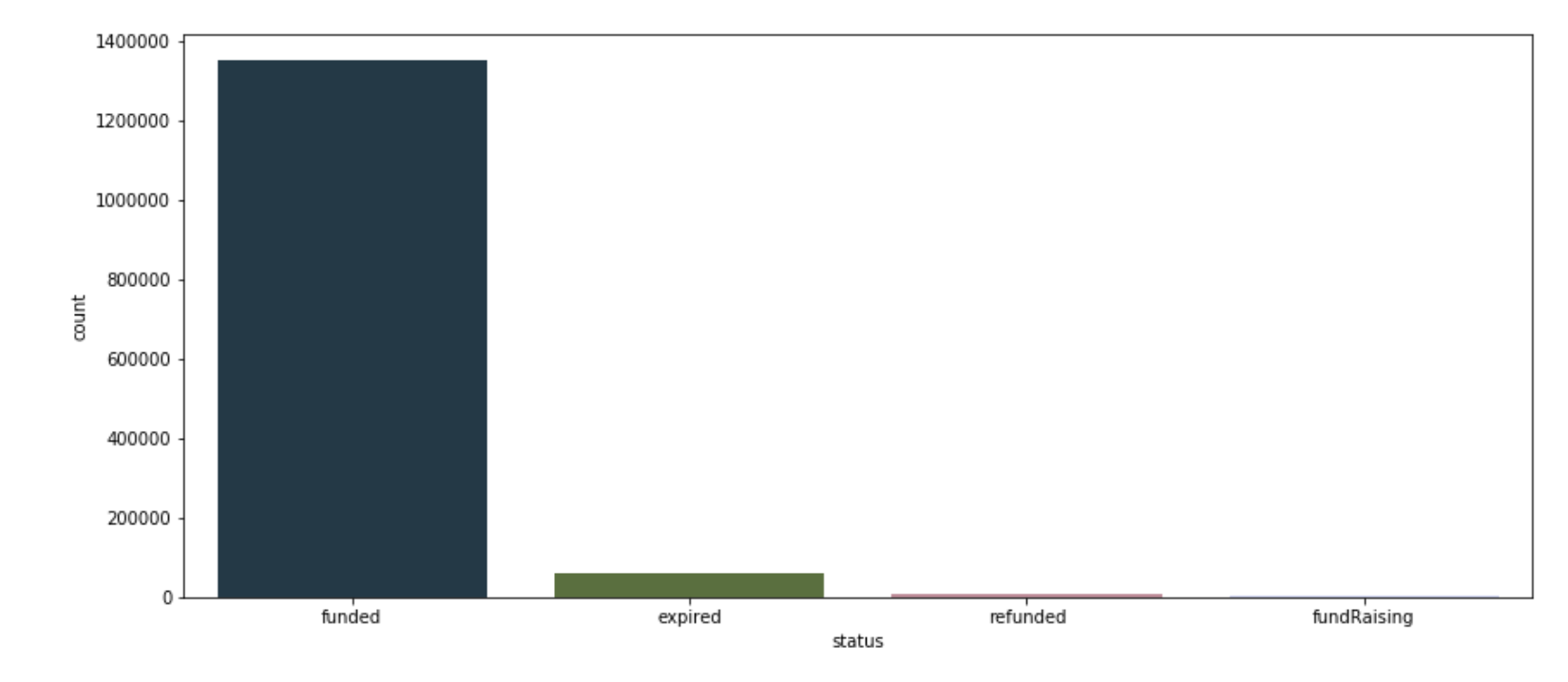
At any point in time there are 4000 + entrepreneurs who need funding. However, if loans are not fully funded within 30 days of posting, **they expire, and no one gets the money.** In this scenario, the field partner has to make the loan themselves, increasing a risk of loss exposure on their books

**How will this analysis help?**

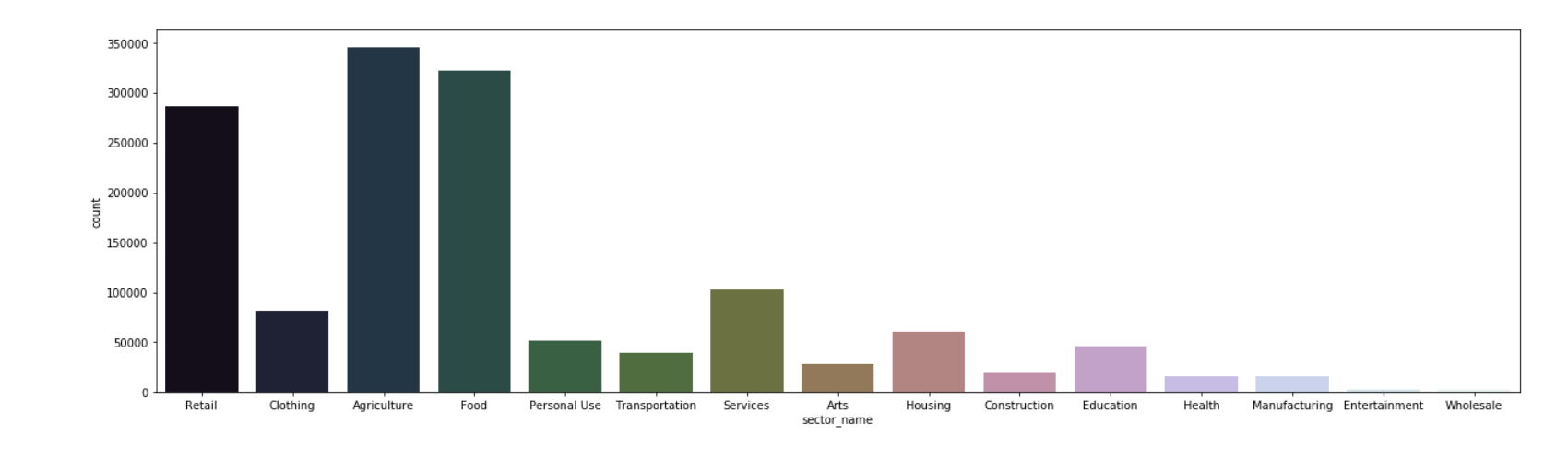
This analysis will help them post the loans most likely to get funded on Kiva and reduce exposures of losses on their books.

By learning about which loans are mostly likely to get funded, they can focus their energy on promoting those aspects of the loan on the webpage.

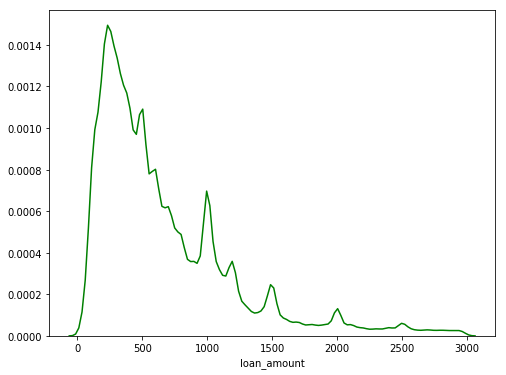
**Exploratory Analysis:**

****

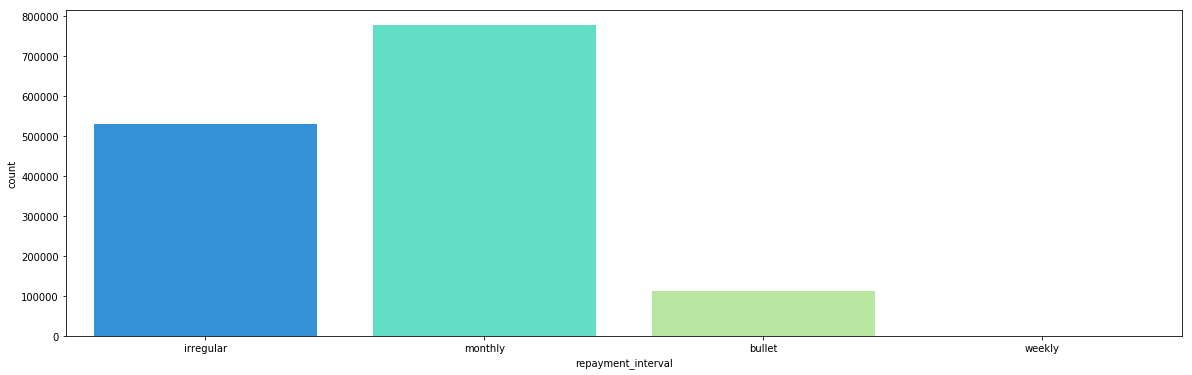
Most loans in this sample are funded, skewing our data considerably.



Majority of borrowers are in the Agricultural sector which intuitively makes sense



The mean loan amount is $800 and the distribution is right skewed.



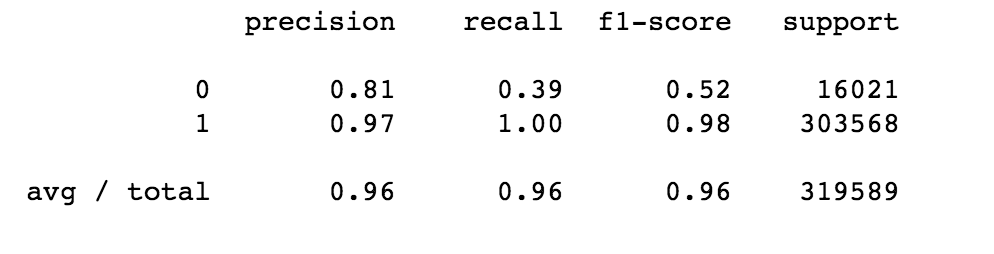
Repayment intervals are generally monthly, but sometimes they can be irregular due to the seasonality and variation in farming products

**Exploratory Analysis:**

**The Models:**

1. Logistic Regression

**Logistic regression** is the regression analysis to conduct when the dependent variable is binary. In this case, loan\_funded =1 and expired=0.

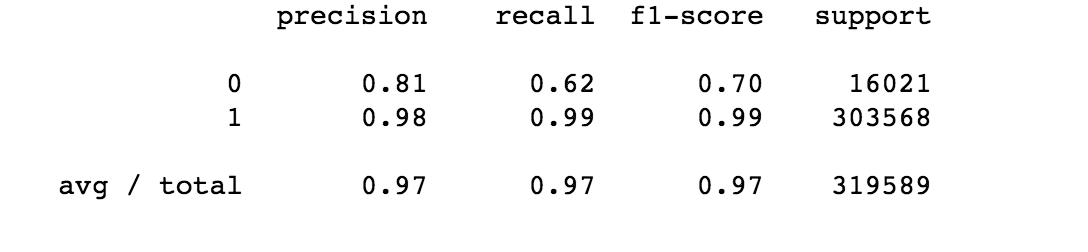


This model is good at determining if a loan will be funded, but not good at determining which loans will not fund. this is very likely to the class imbalance issue mentioned earlier.

The model correctly identifies 40% of all expired loans as "not funded"

2. Random Forest

**Random Forest** is a flexible, easy to use machine learning algorithm that produces, even without hyper-parameter tuning, a great result most of the time. It is also one of the most used algorithms, because it’s simplicity and the fact that it can be used for both classification and regression tasks.



This model is better than the previous one at determining which loans won’t be funded. It correctly identifies 62% of expired loans as “not funded”

**Recommendations:**

* Loan Amount is a significant indictor of whether a loan will get funded or not. In general, smaller loan amounts have a higher probability to get funded
* Shorter loan terms are more likely to be funded
* Journal entries are significant! The partner should make an effort to write these as lenders are more likely to feel connected to the borrowers this way
* Females are more likely to be funded than males. Partners should prioritize posting those loans.