

Team 23:

Kiva.org / Moran Koren

Eyal Ben Baruch - 318651494

Alon Fridental - 209370774

Gilad Schwarz – 207334194

Github repo: https://github.com/Eyalbenba/Kiva_R_Project/tree/main/final_report

The Impact of Lender Count on Loan Repayment Rates/Kiva platform:

Introduction

Kiva is a non-profit organization that operates a revolutionary online lending platform. It connects lenders who are looking to provide microloans to entrepreneurs and individuals in need around the world. Kiva's mission is to alleviate poverty by providing access to affordable capital for those who may not have access to traditional banking systems.

Through Kiva's platform, individuals can lend small amounts of money to borrowers in various countries and sectors. These borrowers then use the funds to start or expand their businesses, improve their education, or meet other financial needs.

It's worth noting that while lenders on Kiva do not earn interest on their loans, the repayment of funds allows them to reinvest in other borrowers or withdraw their money from the platform. Kiva has gained popularity as a way for people to make a positive social impact through small-scale lending.

Our Research Question Is: Can you predict if a loan will be returned and How does the number of lenders affect the loan repayment rates on the Kiva platform?

Our research focuses on predicting loan return rates by examining the loan characteristics and other relevant variables, the main goal is to determine if these factors can serve as reliable predictors for estimating the likelihood of loan repayment.

The following natural question we aim to investigate is the relationship between lender count and loan repayment rates, shedding light on the potential impact of collective lending efforts on borrower repayment behaviors.

Our expectation prior to our work on the data was that an increased number of lenders may positively influence repayment rates. The logic behind this expectation is that when a loan has a higher number of lenders might hint about their belief in that loan cause and its possibility to be returned.

Prior research has explored factors influencing borrower behaviors and provided insights about loan features, notably a study titled "The Role of Rating and Loan Characteristics in Online Microfunding Behaviors" conducted by Paruthi, Gaurav et al. One of the research questions expands about the role that various loan features have in lending behavior. They're results show that lenders appear to favor loans in sectors that are often times aligned with sectors that official institutes like to contribute to, which is pretty similar to the "Wisdom of The Crowd" idea. Our study differs from theirs and aims to examine loan return rates and the influence of lender count on loan repayment rates by harnessing their insights. We're also aware that assessing risks in loan in the banking world exists, but our study is about understanding loan outcomes by investigating the specific impact of lenders count on loan repayment opposed to a standard loan.

By conducting this research, we seek to contribute valuable insights into the dynamics between lender count and loan repayment rates on the Kiva platform. The findings will provide a deeper understanding of the effectiveness of collective lending efforts.

We hope our model can contribute to understanding the factors that influence loan repayment and provide insights into the effectiveness of using regression models for predicting loan outcomes and we intend to back up our findings with analyzing and measurement tools.

Data overview

The data contains two primary entities: borrowers and lenders. Borrowers are described as loan projects representing the individual borrowing activities on the Kiva platform, while lenders are the individuals or groups providing financial support to these projects.

We merged 2 datasets (BigML and loans CSV's) in order to form a complete set of information about the entities stated above, meaning each row is an observation about a loan given.

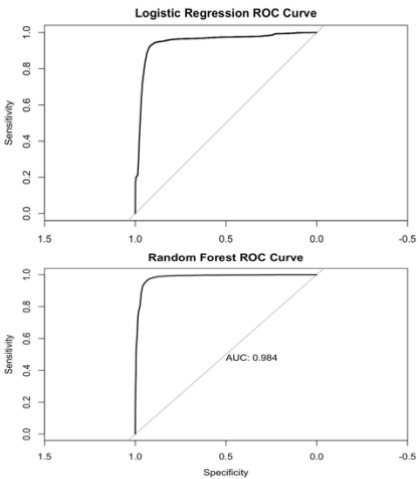
Broad explanation about the Data Variables:

Each observation consists of a loan made on Kiva which is represented by a row in the dataframe. The variables that make up that row divide into:

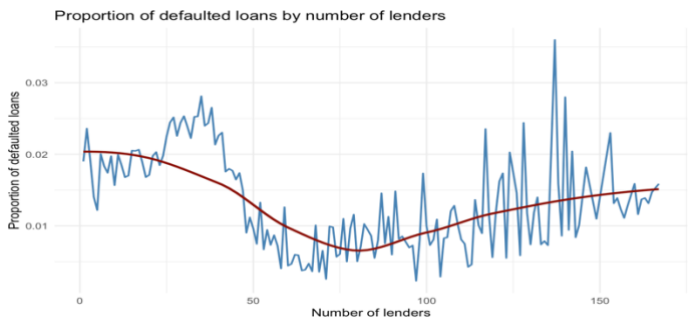
- Ethnicity Variables such as Native Language, Country etc..
- Loan Characteristics such as Lender Term , Number of total Lenders, Loan use etc..
- Business Variables such as Sector , Activity Name etc. . .

Methods and Results

Our first research question aimed to predict loan defaults, and for this purpose, we employed two models: logistic regression and random forest. Both models performed well, but the random forest model demonstrated a slight advantage over the logistic regression model in terms of predictive accuracy. Moreover, the random forest model provided valuable insights into the key attributes influencing loan repayment. By identifying the most crucial factors, we gained a better understanding of the drivers behind successful loan repayment. Additionally, the random forest model assisted us in exploring and explaining the observed increase in default rates when the number of lenders exceeded 80. Its ability to capture complex interactions among variables aided in formulating potential explanations for this phenomenon, thereby enhancing our overall understanding of the data.



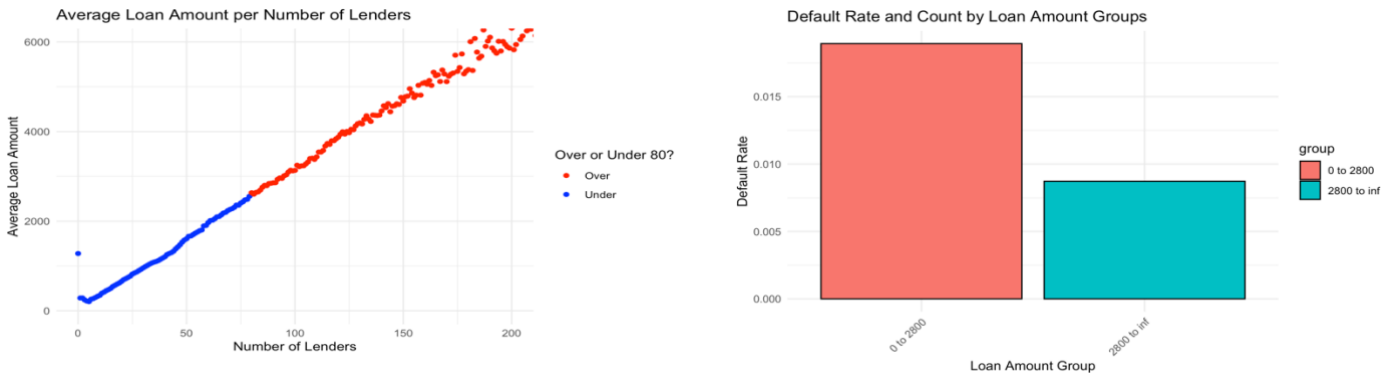
Moving on to our second research question, we sought to understand the impact of the number of lenders on loan repayment rates within the Kiva platform. We began by visualizing the relationship between the number of lenders (Num_Lenders) and loan default rates. A sudden increase in default rates was observed after the number of lenders surpassed 80.



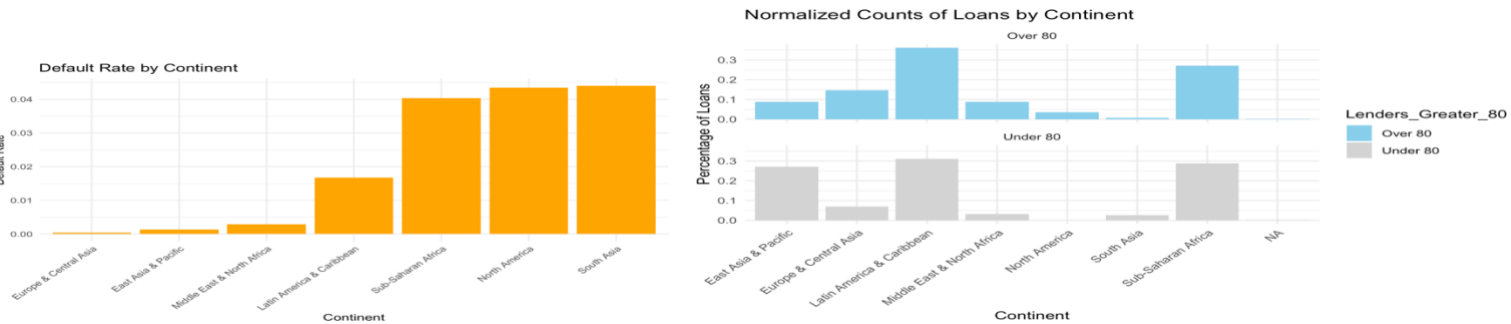
To investigate this increase in default rates, we formulated several hypotheses. Hypothesis 1 stated that the lack of sufficient data for loans with over 80 lenders might be the reason in increase in default rates. Hypothesis 2 focused on assessing whether different sectors in loans influenced the default rates. Hypothesis 3 explored the possibility that an increase in loan amounts in these loans could be a factor contributing to the rise in default rates. Finally, Hypothesis 4 examined the potential influence of different continents on loan defaults

Among these hypotheses, Hypotheses 3 and 4 proved to be the most interesting:

Regarding Hypothesis 3, we initially hypothesized that an increase in the number of lenders would lead to higher loan amounts and, consequently, higher default rates due to the potential challenges in repayment. However, upon analyzing the data, we were surprised to find that loans with higher amounts, associated with a larger number of lenders, actually exhibited lower default rates compared to loans with lower amounts. This unexpected result challenges our initial hypothesis and suggests the presence of other factors influencing the relationship between loan amount, number of lenders, and default rates.



Regarding Hypothesis 4, we examined the distribution of continents for loans with over 80 lenders compared to loans with fewer lenders. We discovered a significant decrease in loans from East Asia and an increase in loans from North America and Africa within the group of loans with over 80 lenders. Further analysis of default rates revealed that East Asia exhibited remarkably low default rates, while North America and Africa had higher default rates. This finding suggests that the higher default rate observed in loans with over 80 lenders, as depicted in our initial graph, can be attributed to these regional variations.



Overall, this analysis of the Kiva dataset provided valuable insights into loan default prediction and the relationship between the number of lenders and loan repayment rates. The choice of employing both logistic regression and random forest models allowed us to compare their performance and select the most effective model for our predictive task. Additionally, the investigation of various hypotheses related to the number of lenders beyond 80 lenders shed light on potential factors influencing loan defaults, offering valuable knowledge and implications for future research in this area.

Limitations and Future Work:

Limitations:

Despite the valuable insights gained from our analysis, there are certain limitations to acknowledge. First, our study focused solely on the Kiva dataset, which may not fully represent the diversity of loan repayment patterns across different regions or platforms. Additionally, our analysis primarily relied on numerical and categorical variables available in the dataset, limiting the exploration of potential unobserved factors that could influence loan defaults. Furthermore, the data used for our analysis had a specific time range, and loan repayment dynamics may vary over time, suggesting the need for periodic updates and continuous monitoring.

Future Work:

To further enhance our understanding of loan defaults and the impact of the number of lenders, future research could explore additional data sources to validate and expand upon our findings. Regarding the analysis of the data - Inside our data set there was a column that loaners wrote their loan description, it could be interesting to incorporate this text data using LLMs (e.g. GPT-4), by doing so the findings could offer insights into qualitative factors influencing loan defaults and funding (e.g. longer description ==? more funding). Furthermore, conducting longitudinal studies could shed light on the temporal aspects of loan repayment behavior and capture any changes or trends over time. Future work could also involve investigating the perspective of the lenders themselves and understanding the factors that contribute to successful loan repayment from their point of view. Conducting surveys or interviews with lenders who have successfully repaid their loans could provide valuable insights into their strategies, motivations, and experiences throughout the lending process. By gaining a deeper understanding of the lender's perspective, we can uncover additional factors that may influence loan repayment and potentially identify best practices or interventions that can help improve repayment rates. This approach would provide a more holistic view of the lending ecosystem and contribute to the development of targeted interventions and policies to support successful loan repayment.