

Data Visualization project

KIVA CROWDFUNDING ANALYSIS



**SUBMITTED BY:**

HIMANSHU CHAWALE

UMA DALWANI

GAURI MUTHA

ROOP KISHEN PERSHED SETH

APURVA SHERKE

1. **INTRODUCTION**

Kiva is an online crowdfunding organization which extends financial services to poor and backward people in various countries around the world. Over $1 Billion USD has been lent in the form of loans to over 2 million people in the past. Making decisions to lend loans to the borrowers is based on various critical factors such as investment priorities, level of poverty, and understand the target communities of the lenders and informing them has been a challenging task. Kiva is headquartered in San Francisco, California.

1. **DATA SET REFERENCE**

The data files have been taken from Kaggle. The location of the data files used is –

<https://www.kaggle.com/cesarjaitmanlabaton/data-science-for-good-kiva-crowdfunding/data>

The following datasets have been used in the visualization:

1. **Kiva\_loans.csv** –This file provides details about the loans lent by Kiva between 2014 till 2017. This is the primary file of our analysis. The data used in this dataset are:

|  |  |
| --- | --- |
| **Column** | **Description** |
| funded\_amount | Dollar value of loan funded on Kiva.org |
| loan\_amount | Total dollar amount of loan applied |
| activity | Loan activity type |
| sector | Sector of loan activity |
| country | Country of the borrower |
| posted\_time | Date and time when loan was posted |
| funded\_time | Date and time at which loan was funded |
| term\_in\_months | Time taken for the loan to be funded |
| lender\_count | Number of lenders contributing to loan |
| borrower\_genders | Gender of borrower(s) |
| repayment\_interval | Frequency at which installments are paid |

1. **Kiva\_mpi\_region\_location.csv** – This file provides the Country Name, MPI, Latitudes and Longitudes of the countries that borrowed loans from Kiva. The data used in this dataset are:

|  |  |
| --- | --- |
| **Column** | **Description** |
| country | Country name |
| region | Name of location within country |
| MPI | Multi-dimensional poverty index for this region |
| lat | Latitude |
| lon | Longitude |

1. **Loan\_theme\_ids.csv** – This file provides the theme type of each of the loans.

|  |  |
| --- | --- |
| **Column** | **Description** |
| id | Unique ID for loan (Loan ID) |
| Loan Theme ID | Unique ID for loan theme |
| Loan Theme Type | General description of the loan theme category |
| Partner ID | Unique ID for field partners (Partner ID) |

1. **Loan\_themes\_by\_region.csv** ­– This file provides the themes of loans in various regions along with the partners who have lent the loans, amount and locations.

|  |  |
| --- | --- |
| **Column** | **Description** |
| Field Partner Name | Name of Field Partner |
| Sector | Sector of loan activity |
| Loan Theme ID | Unique ID for loan theme |
| Loan Theme Type | General description of the loan theme category |
| Country | Country name |
| Number | Number of loans funded in this location |
| mpi\_region | MPI Region where we think this loan theme is located |
| lat | Latitude |
| lon | Longitude |

1. **MPI\_national.csv -** This file contains data about the MPI of particular country along with the factors which defines MPI like deprivation intensity, headcount ratio etc.

|  |  |
| --- | --- |
| **Column** | **Description** |
| ISO | Unique ID for country |
| Country | Country name |
| MPI Urban | Multi-dimensional poverty index for urban areas within the country |
| Headcount Ratio Urban | Poverty headcount ratio(% of population listed as poor) within urban areas |
| Intensity of Deprivation Urban | Average distance below the poverty line of those listed as poor in urban areas |
| MPI Rural | Multi-dimensional poverty index for rural areas within the country |
| Headcount Ratio Rural | Poverty headcount ratio (% of population listed as poor) within rural areas within the country |
| Intensity of Deprivation Rural | Average distance below the poverty line of those listed as poor in rural areas |

1. **african\_conflicts.csv -** This file contains dates and locations of conflict events; Specific types of events including battles, civilian killings, riots, protests and recruitment activities;This dataset codes the dates and locations of all reported political violence and protest events in dozens of developing countries in Africa

|  |  |
| --- | --- |
| **Column** | **Description** |
| ACTOR1 / ACTOR2 | Name of first actor |
| ACTOR1\_ID/ ACTOR2\_ID | Id of the actor |
| ACTOR\_DYAD\_ID | Join of first 2 actors |
| ADMIN1/ADMIN2/ADMIN3 | Administrative region where event took place |
| ALLY\_ACTOR\_1/ALLY\_ACTOR\_2 | Allied party to actors |
| COUNTRY | Country of conflict |
| EVENT\_DATE | Date of conflict |
| EVENT\_ID\_CNTY | An individual identifier by number and country acronym |
| EVENT\_TYPE | Event occurrence, string eg “Riots against police” |
| FATALITIES | Integer value of fatalities that occurred, as reported by source |
| GEO\_PRECISION | A numeric code indicating the level of certainty of the geocode for the event |
| INTER1 / INTER2 | A numeric code indicating the type of Actor |
| INTERACTION | A numeric code indicating the interaction between types |
| LATITUDE | Latitude |
| LOCATION | Location where event occured |
| LONGITUDE | Longitude |
| NOTES | Additional notes |
| SOURCE | Source of conflict information |
| TIME\_PRECISION | A numeric code indicating the level of certainty of the date coded for the event |
| YEAR | Year when the event occured |

1. **countries and continents.csv**

World countries with their corresponding continents , official english names, official french names, Dial,ITU,Languages and so on

|  |  |
| --- | --- |
| **Column** | **Description** |
| Name | Name of country |
| official\_name\_en | Official Name |
| Capital | Capital of the country |
| Continent | Continent of the country |
| Language | Language code |

1. **DATA MANIPULATION**

The missing geo-code values using ‘ArcGIS Pro’ tool. However, even after this, certain values were still missing which were fixed manually, to locate the regions on the maps.

‘Borrower genders’ field in Kiva\_loans.csv dataset was also manipulated which had values in the form of comma separated values. 2 new columns were added, i.e., Female and Male. Each cell was populated after computing the count of respective gender for record.

1. **OVERALL INSIGHTS**

More than **80 countries** all over the world are receiving loan funding by Kiva organization.

**Sub-Saharan Africa** is the most popular region receiving loan funding by Kiva organization

Maximum number of loans are repaid in the period of **14** months

The more the number of lenders for a loan, less is the total amount of funding for that loan

The maximum number of field partners who are providing loan are present in **Kenya**

**Sub-Saharan** has the maximum poverty rate compared to other regions across the world

Compared to any developed country the developing/poor countries receive major funding from KIVA e.g., USA: Average time for any loan to be approved in USA is approximately **34 days** as compared to other countries like Kenya (**6 days**) or Philippines (**7 days**)

1. **VISUALIZATIONS**
2. Global Viz

This visualization shows the countries across the world, which have been provided loan funding.

**Encoding:** Geocoding, Color.

Since it is a location data, we are geocoding the longitude and latitude values using a map. Due to the structure of the map, we wanted a uniform distribution of funded amount and hence Color encoding was used.

**Insight**: Kiva has raised funds to many countries across the globe; Philippines and Kenya are the top funded countries

1. Lender Investment

This visualization shows the proportion of Loan amount to Funded amount provided by lenders, with lender count starting from 0 to 100

**Encoding**: X-Y plane, Value

Since we wanted data against 2 axes, hence we used x-y plane. Maximum amount of loans is depicted using high intensity value and minimum amount of loans is depicted using low intensity values

**Insight:** Few loans have gone unfunded. Loans which have been provided maximum funding have fewer number of lenders as compared to the ones who have higher number of lenders.

1. World Regions Most Funded

This visualization shows the regions which have been funded most across the world

**Encoding**: X-Y plane, Color

Since we wanted data against 2 axes, hence we used x-y plane. Color is used to highlight different regions across the world, since we wanted the most funded region to be highlighted in comparison with others

**Insight:** Maximum number of loans have been provided to Sub-Saharan Africa region

Europe and Central Asia have the minimum number of loans funded by Kiva.

1. MPI Distribution World Regions

This visualization shows average poverty index for different regions and countries across the world

**Encoding**: Size and Color

Size is used to highlight the poorest regions based on average MPI: with poorest region being the biggest square and comparatively less poor region with a smaller square.

Color is used to avoid the monochrome effect and make it more appealing and differentiable

**Insight**: Among the Sub-Saharan African region, Chad and South Sudan are the poorest countries based on the multidimensional poverty index

1. World MPI Map

This visualization shows average poverty rate for different regions spread geographically across the world

**Encoding:** Geocoding, Color, Size

Color is used to define MPI, where red is used to define poor regions and green is used to define regions of lower MPI. Size is used to define the measure of MPI, MPI of the poorest country will have biggest red circle and MPI of the less poor countries will be smallest green decreasing as per the scale

**Insight:** African Sub-Saharan region has the maximum poverty index for most of its countries

1. Amount Disbursed Region Map

This visualization shows the average amount of Loan provided to different regions across the world

**Encoding:** Geo-coding, Color

Color is used to show the average disbursement amount which has been provided by Kiva, maximum amount being depicted by green and lowest amount depicted by red

**Insight**: Kiva has fairly funded the countries which have higher MPI in comparison with World MPI Map

1. MPI Urban Rural Comp

This visualization shows the correlation between average MPI Rural and average MPI Urban for different countries

**Encoding:** X-Y plane, Color.

Color intensity defines correlation between average values of MPI Urban v/s MPI Rural

Darker color means rural MPI. Lighter color means urban MPI. Color was used to signify the variation.

**Insights:** MPI ratings of both are pretty much correlated to an extent.

The countries from **Europe, South America, North America** falls in the starting point of the plot, which shows the lower **MPI Ratings**.

The countries from **Asia** and mainly **Africa** in the later part of the plot, reflecting the higher **MPI Ratings**.

Interestingly, countries like **Somalia, Niger, Ethiopia, Burkina Faso** have much lower Urban MPI Ratings (0.3) and have higher Rural MPI Ratings (0.7).**South Sudan** has both Rural and Urban Ratings high.

1. Deprivation Intensity Correlation

This visualization shows the correlation between average intensity deprivation for rural and urban parts of the countries

**Insights**: the initial stages where the intensity rates are low for both they have a linear graph. But there are some outliers like that of Chad, Somalia, South Sudan which are having a different opinion

**Encoding:** X-Y plane, Color

Color intensity defines correlation between average intensity of deprivation values of MPI Urban v/s MPI Rural. Darker color means rural MPI. Lighter color means urban MPI. Color was used to signify the variation

1. Field Partners Count by Country

This visualization shows the count of field partners for different countries across the world

**Insights:** Kenya has the maximum number of field partners (**30**)

**Encoding:** X-Y plane, Color

Color is used to show better visual effect and prominently highlight the country with most number of partners

1. Loan Durations

This visualization shows the period for which loans have been requested for.

**Insights**: Most common term for taking a loan is 14 months and 8 months

**Encoding**: X-Y plane, Color

We have used color hue to show maximum loans for a loan period

1. Sub Saharan Loan

This visualization shows the distribution of Number of loans across the countries of sub-Saharan region

**Insights**: Kiva is funding majorly in sub-Saharan region and Kenya tops the chart. Kenya gets maximum help from Kenya.

**Encoding**: x-y plane, color

This visualization is using binary color coding to differentiate top 5 countries based on count of loans from with a single color and rest with another

1. Conflict Affected Africa

This visualization is based on an external dataset which shows the count of fatalities across various locations for a range of time and due to different events

**Insights**: Somalia sees most number of fatalities due to conflicts. Other regions affected by conflicts are: Congo, Nigeria, Sudan

**Encoding**: Color, size

Size is used to show which regions have most number of fatalities. Color is used to highlight its appearance

1. Repayment Interval

This viz shows the form of repayment done the most, by each country and the number of payments made using that mode.

**Insights:** Philippines has the most number of irregular payments, while Kenya has seen more of bullet payments

**Encoding**: Size, Color

1. Terms for each sector

This visualization shows the total number of months taken to repay the loan in each sector by borrowers from Kenya.

**Insights:** We see that the maximum number of loans have been taken in the Education sector, and then followed by the Agriculture sector. It takes an average of 31 months to repay the loans borrowed for education and 24.5 months for the Agriculture sector.

**Encoding**: Color.

Color differentiates the sector of the funding. Each sector will have its unique color

1. Gender Ratio

This visualization shows the gender distribution of the loan borrowers in all the countries.

**Insights:** We see that approximately the same number of males and females have borrowed loans, however, the number of females is slightly lower than the males.

**Encoding**: Color

We used color to differentiate percentage of males and females

1. Days for approval

This visualization shows the average number of days taken to approve loans in countries like Kenya, Philippines, and United States for various sectors.

**Insights:** The loans approved in Philippines takes an average of 5-10 days in all the sector while it takes approximately 25 days on an average in United States for any loan to get approved. Kiva prioritizes funding only the poorer countries rather than the rich countries, therefore, loan processing time in USA is much slower than in the other countries

**Encoding**: Color

Higher the number of days darker is the color and lower the number of days required for loan approval lighter is the color. The color hue makes it easier for the reader to see the range of the values easily and makes it differentiable.

1. Dashboard

This visualization shows us the Number of Loans provided by Kiva across different countries, and within different sectors, for different activities and which genders have requested loan at a go.

1. **CALCULATED FIELDS**

**Female Percentage** - The Female Percentage is calculated by dividing the total sum of all women by sum of the total number of men and women.

**Male Percentage** - The Male Percentage is calculated by dividing the total sum of all men by sum of the total number of men and women.

**Loan Approval Time** - This is calculated by subtracting the Funded Time from the Posted Time.

1. **RHETORICS USED**

**Logical** - It is logical to fund in the region with higher MPI

**Example where it is used:** Loan approval time in developed countries takes much longer than that of the other countries. (**Visualization Reference: Days for Approval**)

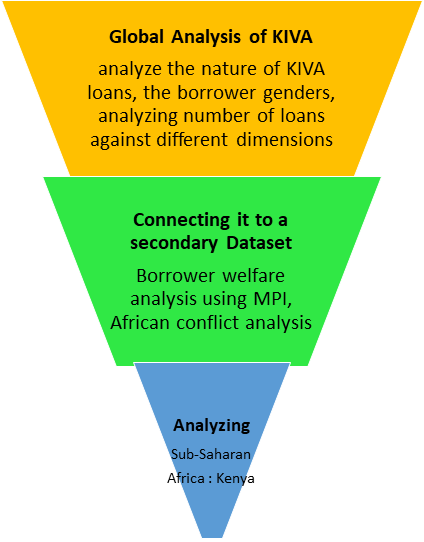
**Ethical** - It is not moral to fund the regions which is doing fairy good in terms of MPI, instead it is moral to fund the countries which are poorer that the ones which have lower MPI.

**Example used:** Philippines, even when it has comparatively lower MPI, it is the most funded country than other countries which have high MPI like South Sudan, Chad (**Visualization Reference: Global Viz**)

1. **STORYLINE AND APPROACH**

We divided our storyline in 3 categories:

1. A global analysis of Kiva funding
2. Insights from first category: Country specific analysis (Philippines)
3. World region specific analysis: Sub-Saharan Africa: Conflict analysis of African countries, Analysis of Kenya



1. **Global Analysis of KIVA**

KIVA’s mission is to alleviate the poverty across the globe. It was decided to do an exploratory analysis of KIVA dataset. The initial objective which was considered was to analyze:

1. Countries which were funded by KIVA
2. Top sectors which were given loans by the organization
3. What repayment interval types are available
4. Lenders who have provided the loan
5. Activities which are funded the most by KIVA
6. Gender Analysis
7. **Connecting to a Secondary Dataset**

On further research, it was seen that loan funding works on various parameters than just defined income-level. It interested the team that by using some additional datasets, the borrower’s welfare can be studied. The **MPI** combines two key pieces of information to measure acute poverty: **the incidence of poverty**, or the proportion of people (within a given population) who experience multiple deprivations, and **the intensity of their deprivation** - the average proportion of (weighted) deprivations they experience. We analyzed these factors and found out the correlation to lay down our insights. MPI visualization of poverty struck countries and KIVA’s help received to those countries was studied further.

1. **Analyzing Sub-Saharan (Kenya, Africa)**

Considering that poorer regions will be funded the most we further analyzed the countries with higher MPI (multidimensional poverty index) and how has KIVA’s helped them. Sub-Saharan being the most affected regions of conflicts as well as poverty, countries from those regions were considered for further analysis and for finding insights. As we were progressing with the analysis for Sub-Saharan Africa region, we found out interesting fact that even though there are other poorer countries in Africa, KIVA has not funded loans for them and instead Kenya with comparatively lower MPI was the country to receive the maximum funding amount. Since Kiva administers loans through a network of field partners we analyzed the field partners influence on funding of the loan. A final comparison about the poor country with the developed country made more sense.

**TERMS FOR REFERENCE**

|  |  |
| --- | --- |
| **Term Used** | **Description** |
| Crowdfunding | Practice of funding projects by raising many small amount of money |
| Multidimensional Poverty Index (MPI) | an international measure of acute poverty |
| Deprivation Intensity | the proportion of people (within a given population) who experience multiple deprivations |
| Incidence of Poverty | the average proportion of (weighted) deprivations people experience |
| Repayment Modes | The modes used by the borrowers to repay the loan : Irregular, Monthly, Weekly, Bullet (One time payment of whole sum) |