**Data**

**Data Collection**

For this study, we initially collected a comprehensive dataset through Kiva API on May 15th, which included 6493 observations and 10 variables. These observations encompass a variety of attributes related to the projects, borrowers, and their engagement metrics. The numeric variables of interest collected are detailed below:

| **Variable** | **Min** | **Median** | **Mean** | **Max** |
| --- | --- | --- | --- | --- |
| sentiment | 0 | 6 | 7.28 | 33 |
| lenderRepaymentTerm | 5 | 14 | 15.1 | 122 |
| loanAmount | 100 | 825 | 1428 | 300000 |
| daysSinceFundraising | 0 | 14 | 14.5 | 42 |
| fundedAmount | 0 | 0.1000 | 0.1668 | 1.0000 |
| researchScore | 0 | 5.500 | 6.926 | 37.500 |

Note that *researchScore* here refers to an internal score given by Kiva internally to determine alignment with social impact or effectiveness in achieving positive social outcomes. The top 6 frequent *country* and *sector* distribution tables are as follows:

| **Country** | **Count** | **Sector** | **Count** |
| --- | --- | --- | --- |
| Philippines | 693 | Agriculture | 1910 |
| El Salvador | 610 | Food | 1687 |
| Kenya | 551 | Retail | 1247 |
| Ecuador | 421 | Services | 474 |
| Colombia | 377 | Clothing | 377 |
| Tajikistan | 361 | Housing | 206 |
| (Other) | 3480 | (Other) | 592 |

The rest two variables are *loan\_id* and *gender* (female: 4909, male: 1584).

However, given the limitations of our dataset, which only included a small number of completed projects, it was impractical to construct predictive models. As a result, we shifted our approach to cluster analysis to identify patterns and relationships within the data. This method allows us to gain insights into the dynamics and key factors influencing Kiva's fundraising process.