Ca' Foscari
University
of Venice

Master's Degree Programme

In

Data Analytics for Business and Society

Final Thesis

The Social Impact of Crowdfunding and The Increasing Microlending Potential: The Case Study of Kiva

Supervisor
Ch. Prof. Francesco Rullani

Graduand
Caterina Tedeschi
Matriculation Number 883594

Academic Year
2021 / 2022

To My Family

Acknowledgements

To Professor Rullani, I extend my sincere thanks for the trust he placed in me, for his willingness to follow me during the development of the work, and for the valuable advice he shared with professionalism and deep sensitivity, motivating me to achieve more ambitious results.

I would also like to thank Professor Maiolini for dedicating his time and attention to me, supporting me in the correct selection and identification of the data and variables processed to realize an accurate and meticulous work.

Index

Introduction

Crowdfunding is considered as one of the most diffused Fintech solutions at the global level. The wide range of its structures has enabled it to satisfy the demands of clients from diverse sectors. For instance, it has been implemented by start-ups to promote their idea in the market or by companies seeking capital to expand their business.

In fact, the first chapter illustrates the crowdfunding ecosystem, examining the several actors involved in its mechanism, the main characteristics, advantages and disadvantages. Furthermore, some insights are provided on the actual reality where crowdfunding operates, the most popular platforms and comparisons with other financing solutions that can be implemented to start a business. Moreover, crowdfunding can be also applied in developing countries to ensure access to credit for underprivileged borrowers.

The second chapter analyses prosocial crowdfunding from different perspectives to reveal its potential. One specific typology of prosocial crowdfunding is represented by Kiva, a micro-lending platform and an international nonprofit organization that operates in underdeveloped countries to support local borrowers. Numerous studies have been conducted on Kiva and several aspects have been investigated. Therefore, this chapter outlines a comparison of diverse research questions based on Kiva’s performance.

The last chapter provides an experimental analysis of Kiva’s data and includes four sections: data cleaning, data visualization, topic modeling and data modeling. In particular, two variables, named *use* and *description,* will be examined with topic modeling. The first indicates the loan purpose, whereas the second delineates the borrower’s profile.

I Alternative Finance: Innovative Solutions in The Actual Reality

Crowdfunding is one of the principal manifestations that has emerged in the Fintech scenario due to the impact of innovation in the financial sector. It originated in the entertainment field and then expanded in other areas. So important is its contribution in our reality that the capital generated through crowdfunding reached a value of $34 billion, considering only the initial actors, who adopted this solution in their lives from 2011 to 2015. Actually, the highest transaction value is $451.30 million in and it was raised in the United States in 2023. It is also expected to present an annual growth rate of 2.46%, while the total amount is predicted to be $ 1.21 billion by 2027 (Statista, 2022).

1.1. The Crowdfunding Ecosystem

The term "crowdfunding" derives from "crowdsourcing", a collection of opinions, services, information and knowledge generated by a group of people sourced via internet. Crowdfunding offers the possibility to promote a project by raising money from the "crowd" through the implementation of an online platform (Jovanovic, 2019). It has diverse manifestations, from simple donations to high-risk investments, and two main phases can be distinguished: the pre-funding part and the post-funding one. The first comprehends the details of the strategy, marketing communication and the funding period. Whereas, the second describes what will happen when the crowdfunding campaign is over: all participants will know whether the initiative has achieved its goal, if profits have been distributed and whether the project can be concretely realized.

Multiple elements can implicitly influence the success of a crowdfunding campaign at the early stage. Firstly, the network of the project creator is a crucial aspect to consider because it represents the quality of the program and its trustworthiness. Moreover, the initial financial support is another positive sign that reveals higher likelihood of receiving future funds. Sustainability is an additional factor to evaluate, even though it does not necessarily denote the success of a campaign. In fact, according to the cooperative capitalism approach, investors are not only attracted by the future profits, but also by the values of the lender community (Bockel et al., 2021).

On the other hand, in the post-funding part, the degree of collaboration and inclusion of backers will consistently affect long-term success. While, in the case of sustainable projects, only a minority publish final data on the effectiveness of the campaign (Kaartemo, 2017).

Crowdfunding combines technological strategies with financial ones and, in fact, represents an alternative solution to existing intermediaries. An intriguing theory that should be applied to analyze this phenomenon from different perspectives is the so-called "stakeholder approach". The stakeholder figure indicates an individual or a group of people who can influence and be influenced by the realization of the company's purposes. The aim of stakeholder theory is the examination of roles within an organization that are completely distant from the traditional ones, but, at the same time, whose contribution is crucial to the organization's success (Laplume et al., 2008).

Website providers are the online structure that supports communication between the founder and backers. It offers multiple services such as constant email updates, news about the project, and a dedicated section for sharing ideas and thoughts. To promote the campaign on social media, a direct link is present to facilitate access to information and attract more investors. Furthermore, third party vendors are involved to enable secure payment processing and guarantee privacy. Whereas, the level of transparency depends on the typology of crowdfunding. (Ordanini et al., 2011).

The project founder designates the originator of the idea. This role can be covered by a start-up, an entrepreneur, an investor etc. Before engaging in a crowdfunding campaign, two aspects should be considered. The borrower's business knowledge and abilities are an indication of his background activities, market experience and, obviously, reliability. While, his product notions shows the degree of involvement in the realization of the item and an evidence of his marketing skills. A founder may present a level of professional training in one or both of the aforementioned areas. In general, his principal aim is to collect financial resources in order to examine a market idea, increase his network of relationships, or acquire validation for a future partnership with lenders (Gerber et al., 2012).

Backers perform numerous functions ranging from raising capital to supporting the project in actual and digital realities. In fact, if they are members of a certain company, they can spread the news and involve more individuals in the monetary phase. At the same time, they could use their personal profile on social media to draw more attention from the online community. According to several theories, different motivations drive backers to joining in a crowdfunding activity such as being recognized for their participation, being an official campaign contributor, aiding a cause. In reality, the genuine reason is a combination of the previous ones, considering also the relative extrinsic and intrinsic benefits derived from the typology of crowdfunding (Beaulieu, Sarker, 2015).

1.2. The Main Features of Crowdfunding

Since 2010, four types of crowdfunding have spread: investment-based crowdfunding, lending-based crowdfunding, reward-based crowdfunding and donation-based crowdfunding. All are characterized by the promotion of an idea on an online platform in which several investors can participate to contribute to the cause through the transfer of capital.

In particular, equity crowdfunding (or investment-based crowdfunding) is an emerging phenomenon in entrepreneurship research. In fact, the number of dedicated studies has incremented by 620% from 2012 to 2017 (Mochkabadi et al., 2020). In this model, an effective share of a company is purchased through an online investment. Firms choose a crowdfunding platform to advertise their offerings and the crowdfunding audience is invited to monetarily support the organization. One of the crucial aspects concerns the rewards for investors: they correspond to the set of ownership and administrative rights to participate in the organization. Once a share is bought, the single contributor has become an official equity stakeholder and his profits will be received in the form of dividends or capital gains. This also means that the investor will be designated as a residual claimant in the investee firm and, as a result, will assume the risk associated with the company's future performance to the extent of his share. Moreover, the crowdfunding platforms will receive a fixed fee from the borrowing firm for offering public relations activities and a flat rate, the amount of which depends on the financial resources raised. This framework is usually the most chosen by start-ups and early-stage companies due to the short period of time and the substantial amount of resources that can be collected. In general, it is also the preferred solution for business-oriented projects, for companies with institutional or commercial clients because it offers several possibilities. Primarily, they have access to bulk purchases rather than individual profits and, secondly, they would obtain favorable rates in procurement. On top of that, they earn the supplier company's share, which is equivalent to an indirect gain from its main services and products (Adhikary et al., 2018).

Peer-to-peer crowdfunding (or lending-based crowdfunding) is a form of alternative financing between individuals, organized through online platforms and aimed at developing entrepreneurial projects with a specific purpose. Originally, it was defined as a valid solution for individuals who did not provide any form of collateral to protect the lender from eventual default. Recently, it has diffused in the business sector as a further option for enterprises in the process of raising capital (Pierrakis, 2019). In fact, from the firm's perspective, it is a cost-effective solution because the platform, which represents the intermediary, facilitates the funding phase and attributes higher credit scores to the organization itself. From the lenders' point of view, they have the opportunity to purchase a portion of the loan at an interest rate whose value depends on market fluctuations (Bruton et al., 2015). This mechanism is based on uncollateralized loans and the concept of "all or nothing". The availability of more and more data on individuals and businesses encourages the definition of precise credit scores in order to facilitate the screening ability of lenders in identifying the correct interest rates (Iyer et al., 2009).

Reward-based crowdfunding and donation-based crowdfunding fall under the community-based crowdfunding model. In particular, the former is characterized by a return that does not present a financial nature and is proportioned to each backer's contribution. For example, an author might offer copies of his book before publication, or a massage therapist might reward donations with gift certificates for future services. Whereas, donation-based crowdfunding, also called donor - based crowdfunding, involves people raising money to support a cause without expecting anything in return. It is most prevalent among charities and nonprofit organizations that use this form of crowdfunding to solicit donations from supporters who believe in the cause they represent (Jovanovic, 2019).

An additional distinction between platforms regards fundraising strategies. "All or Nothing" (AON) refers to the crowdfunding campaign in which neither the funder nor the crowd (that contributed to it) will receive the financial resources if the predefined goal is not reached. It indicates a higher level of commitment and is utterly diffused in equity and lending crowdfunding. "Keep it All" (KIA) is the opposite approach: the company will keep the entire amount even if the aim is not met. It is the most preferred in the context of reward and donation crowdfunding where the importance of the cause prevails on the remaining factors (Shneor R et al., 2020).

Further categorization evaluates crowdfunding typologies according to three value propositions: hedonism, altruism and for-profit. Reward-based crowdfunding falls into the first group, where innovative projects are promoted in order to satisfy crowd interest and spread positive sentiments. Altruism is more related to donation-based crowdfunding, where the ultimate target is to support a social cause usually involving charitable organizations or foundations. For-profit refers to campaigns where the financial return is crucial and, in fact, concerns only the business environment (Shneor R et al., 2020; Haas et al., 2014).

To better understand the motivations behind the growth of crowdfunding, a fascinating aspect to investigate involves describing the highlights of the European legal framework in which this financial solution has been embedded. According to the European Commission 2017, the regulatory architecture of crowdfunding in Europe was rather heterogeneous. Indeed, some countries introduced specific legislation to regulate crowdfunding or improved the flexibility of the previous regulatory framework to adequately respond to the challenges raised by the sector. In general, the growth of the crowdfunding industry has been encouraged to enhance competition among different firms and agents willing to embrace this form of business, reduce the spread of information asymmetries and comply with transparency requirements for participants. Unfortunately, the weakness of the strategy of harmonization and regulatory cohesion advocated by the European Institutions has generated an evident fragmentation in the legislative scenario of the European Member States (Cicchiello, 2019; Hooghiemstra, 2020).

The European Crowdfunding Regulation (CFR) was issued on 7th October 2020 and its scope comprehends investment-based crowdfunding and lending-based crowdfunding. The principal aim of the regulation is to create a single European market to facilitate access to finance for entrepreneurs, start-ups and SMEs in general. This situation has augmented the growth of crowdfunding platforms, supported cross-border activities to surpass the internal barriers and obstacles, stimulated the interest and attraction of investors in this sector, guaranteed the same level of information for the participants and defined uniform rules valid in all EU countries. In addition, the CFR introduces an European passport for crowdfunding service providers (CSPs) (Macchiavello, 2021).

Some forms of crowdfunding are not yet included in the scope of the regulation. Nonetheless, the introduction of a "European passport" has elevated market efficiency, economies of scale and scope between European and National crowdfunding platforms by moving from a purely domestic legislation to a single and remarkable one.

Once the characteristics of crowdfunding have been delineated and the legal framework has explicated its boundless spread, a classification of the actual most recognized platforms is necessary to have a complete picture of the phenomenon. According to Investopedia (Kearl et al., 2022), the best crowdfunding platforms of 2023 are:

- Indiegogo;
- SeedInvest Technology;
- Mightycause;
- StartEngine;
- GoFundMe;
- Patreon.

Indiegogo was developed in 2008 by Danae Ringelmann, Eric Schell, and Slava Rubin and it is considered the best choice for entrepreneurs and investors to promote their ideas and initiatives. It offers the opportunity to match the project's campaign description with the founder's social network profiles, such as Twitter and Meta, in order to share more information about his personal experience and reach a wider audience. The borrower can also cover the role of lender by investing in other products available on Indiegogo and supporting other initiatives. It is present in 235 countries and has funded more than 800,000 ideas in just five years. The only drawback is the charge of 5% platform fee added to a third-party payment processing fee. However, the duration of the campaign is 60 days and, once the aim is achieved, the founder will receive the capital raised in 15 business days directly into his bank account.

SeedInvest Technology was created in New York in 2012 to support start-ups in their initial phase. To mention a few number of its successes: over 250 companies have been funded, more than 700,000 investors have been involved and more than $465 million in financial banking has been reached. In addition, if the project goal is not achieved in the predetermined time, costs will be added to the organization. Because of its popularity, selection among start-ups is a challenging process and includes a due diligence check. Once a certain project is chosen, the entrepreneur must take a few steps to officially launch his idea on the platform. These include creating a profile, closing the round and making some payments.[1] SeedInvest will charge a 5% equity fee and a 7.5% placement fee. Finally, the campaign can last 45 to 60 days, it depends on the business itself.

Mightycause was founded in 2006 as a platform for nonprofit organizations and individuals to promote a social cause, whose number exceeds 150,000. It has multiple solutions for different necessities: from lending-based crowdfunding campaigns, to special occasions, giving days or one-year duration. Moreover, it is compatible with customer relationship management (CRM) systems like Salesforce, offers also a free trial and the possibility to directly promote the idea through an integration on social network channels. The platform provides two subscriptions: the Essential and the Advanced plan. The former offers the basic components for a crowdfunding campaign and costs $59 per month. While, the second one also includes CRM services, statistical analysis and, of course, its price is higher than the previous one and is $99 per month. Beyond the choice of package, the average processing fee is 1.2% of the collected funds, added to 29 cents per transaction.

StartEngine was launched in 2014 by Ron Miller, Howard Marks, and Paul Kessler and now has ABC's "Shark Tank" host Kevin O'Leary as a strategic advisor. It adopts a lay-person-friendly investment approach allowing individuals to become a lender of a specific start-up or firm. The choice is facilitated by the company description with information on evaluation, price per share, number of contributors, etc. Only personal information and payment methods need to be entered into the application. Furthermore, StartEngine is also a platform where start-ups can be registered to seek capital. Actually, the total number of companies involved is 550 and the network has more than 1,000,000 users.

GoFundMe started in 2010 and is one of the most popular crowdfunding platforms online for numerous reasons, such as expert support available for any assistance, 0% platform fee, flexibility in raising funds for a charity, a friend or for one’s personal needs. In addition, setting up a project requires only a few steps: creating the campaign with images, videos

[1] Start-ups in their early-stages seek capital from outside sources, called rounds, in which investors participate with their own financial resources in exchange for equity or partial ownership of the firm itself.

or textual information related to the description of the applicant and the desirable amount that he would like to receive; sharing it via social networks, messages and emails; tracking monetary movements made by contributors. For the aforementioned aspects, GoFundMe has reached more than 200 million donors and raised over $17 billion. The only payment is the processing fee of 2.9% and $0.30 per transaction.

Patreon was founded in 2013 by Sam Yam and YouTube musician Jack Conte and is one of the leading platforms representing the crowdfunding industry dedicated to creative professionals. It has supported over 250,000 artists, writers, musicians and podcasters and has raised more than $3.5 billion. Monthly subscriptions are available, each distinguished by the services offered and the monthly fee. The first package is called “Lite”, takes 5% of the capital raised and it is aimed at creators at the beginning of their careers. The second option is “Pro” with an 8% fee and developed to promote artists’ business. The last plan is “Premium” and the associated payment is 12%. It is designed for clients who have a minimum revenue of $5,000 on the platform and have a social media community of at least 100,000 members. All solutions involve some software integrations, business tools and the use of Patreon mobile app.

1.3. The Main Aspects of a Successful Crowdfunding Idea

The perfect recipe that ensures the success of a crowdfunding project depends on the combination of diverse ingredients such as the pledging conditions (the final goal and duration), the features of the project (the campaign description, texts, videos, images) and the characteristics of the founder (previous experience, reliability and social network size) (Koch, Siering, 2019).

Firstly, the principal aim of the campaign is a crucial indication of the amount of money desired and needed to finance the project. It is considered a measure of the size of the plan and its importance. In general, it is appropriately defined in order to receive the expected reaction of investors. If the target is too ambitious, lenders would perceive a higher level of uncertainty associated with the project itself and this will produce a negative effect on their actual participation. Therefore, the definition of the final goal generates a consistent impact on the crowdfunding success, especially in the monetary phase.

Secondly, the funding period is another signal that determines the final outcome. A long campaign duration could represent an insufficient level of trust on the part of the founder and, consequently, investors may doubt the trustworthiness of the founder and the defined program. For this reason, short periods are preferred and more common.

Information disclosure is a delicate issue that should be handled scrupulously. It includes a textual description, images and videos. The first explains the characteristics of the project, its details and crucial information. So important is the correct use of words that it can generate a positive or negative impact on the success of the project. Indeed, if the explanation is clear, it increases the interpretability of the plan for interested investors and communicates a sense of reliability of the founder and the selected platform. However, it could also produce the opposite effect in the case of excessive information. It would increase investors' effort to understand the campaign and their decision making ability (Barbi, Bigelli, 2017; Mollick, 2014; Pitschner ,Pitschner-Finn, 2014).

Images play a more important role than textual information because they directly capture the user's attention and interest in the campaign (Glenberg, Langston, 1992). If a picture has the right colors, size and figures, it will certainly be highlighted in the description. As a result, the webpage views will increase, the project will be more scrutinized by investors and it will reach a wider audience. Numerous studies state that the longer the duration of the webpage, the greater the chances that the visitor will be attracted to the campaign and may become a future contributor (Unnava, Burnkrant, 1991). It is also true that too many images on the same page will drive away any interest in the project.

Videos are a key factor in the description of project characteristics because of the richness derived from the combination of motion and audio information. Dynamic visualization is always preferred over static images (Park, Hopkins, 1993). Project promotion is more effectively expressed through videos than simple textual description. One minute of video can transfer more content than two written pages of explanations of the idea, gaining more views and more interest from the crowd (Jiang, Benbasat, 2007). Therefore, the perfect amount of text, images and videos would definitely generate a positive effect on the success of the crowdfunding project.

Risk disclosure is one of the main feature of a crowdfunding campaign because it outlines any hazards, threats or exposures examined and identified through a risk management approach (Linsley & Shrives, 2006). A wide range of risks could be present, such as delaying the deadline, not reaching the set of predetermined goals, postponing loan repayment etc. Despite the negative consequences that might occur, the section dedicated to the project's risks should always be integrated into the crowdfunding platform to ensure transparent communication to web page visitors (Jahansoozi, 2006). It is also a smart insight of the founder for producing a risk analysis, a fundamental part to augment the probabilities of project success. On the other hand, an overly detailed report could also alarm investors and convey a high level of uncertainty about the outcome of the project.

The way a sentence is phrased, the words chosen and their juxtaposition can evoke positive or negative emotions in the reader. Numerous researches proved that the sentiments felt influence individual decision making (Bagozzi et al., 1999). In the case of investors, the use of specific terms affect their willingness to participate in the project and, consequently, the probabilities of financing the initiative.

Regarding the sphere of founder characteristics, previous project experience is considered as a positive element at the funding stage. Crowdfunding projects launched by a single entrepreneur are described in his personal page. Lenders can visit it to get an overview of his activities and the results of all his initiatives. A founder with previous works is considered more reliable than individuals at the beginning of their careers (Abdul-Rahman and Hailes, 2000). One theory that can confirm this general trend is called "Matthew Effect" and states that a success happened in the past has a high probability of occurring in the future (Merton, 1968). In fact, a founder with excellent reviews, ratings and successful projects will also receive financial resources to promote his future ideas.

The founder's network is an additional variable that, in combination with his experience, increases his perceived credibility with the audience. The number of friends and followers on social media is a measure of the user's popularity and represents a public screen through which he can share his ideas, values and communicate his identity and personality. As a result, the borrower's public profile and related contacts positively impact on the success of financing. For this reason, several entrepreneurs include a link to their public profile on social networks in the project presentation, to augment the amount of information that could contribute to the founder’s transparency, reliability and trustworthiness (Donath & Boyd, 2004; Yu et al., 2017).

After the interpretation of the various factors that influence the success of a crowdfunding campaign, a further clarification needs to be added to explicate what are the ultimate effects of combining founder characteristics with the details of the project. In fact, if the borrower signals are present, investors will attribute more importance to them than to project information (Gulati, Higgins, 2003). The positive feelings generated by the description of the founder’s experience, network and popularity are more decisive than the lack of textual details on the campaign. This happens because a reliable founder persuades more investors who will consider the aim of the project concretely achievable. It is also true that the opposite situation might occur: the founder’s business experience could present less specificities and, therefore, the crucial part becomes the data and particulars about the campaign, with the proper composition of textual information, images and videos.

1.4. Key Financing Solutions in The Early Stage of a Business

The principal requirement of start-ups in their initial phase is the search for capital. This process can take several months because it is one of the most challenging parts. In fact, once the business idea has been identified and analyzed from all perspectives to understand how to implement it, the financial part is another significant obstacle to overcome. In addition to crowdfunding and its various typologies, other options are available to select, based on the needs of the organization.

Angel financing refers to investors who support start-ups or newly opened businesses in exchange for equity or partial ownership of the firm itself. Recently, the size of the investment has increased from $25,000 to $100,000 for a single company and continues to grow due to the excellent results already obtained. Indeed, this strategy has been adopted by successful enterprises such as Facebook, WhatsApp and Uber. Some constant aspects of small businesses are always evaluated by this category of investors. Firstly, the values on which the company is based and what should be promoted also in its future development. Secondly, the positive consequences that will be certainly realized if the total amount is delivered. Thirdly, how the resources will be implemented in reality, based on the definition of the business plan. In particular, what technologies to include to be competitive in the market. From a financial point of view, another aspect to consider is the current level of risk associated with the business. In fact, the level of uncertainty affects investors' requests. It is also true that finding angel figures is a daunting challenge for small businesses. Fortunately, it has been facilitated by crowdfunding sites such as Kickstarter and Indiegogo, where numerous investors are signed up, or considering the network of entrepreneurs, venture capitalists, investment bankers, angel contributors and the AngelList website. Social media can be a useful tool at this stage of funding. In particular, LinkedIn offers the possibility to individualize some professionals interested in the same sector with which the start-up is affiliated. Obviously, the most effective solution is to create a network with experts and leading figures who might be attracted by the opportunity to fund a project. Public events are also another alternative to be noticed by experts. In fact, it is the perfect setting for building relationships with potential clients, employees, accountants, investment bankers and strategic partners.

Regarding venture capital companies, they are another opinion for start-ups and small business enterprises (SMEs) in their funding phase. Indeed, they offer professionals who can provide strategic assistance, financial resources and introduce their clients to a network of contacts. So efficient and effective is the contribution of venture capitalists in a business that the possibility of being selected among numerous organizations is really challenging. They are usually attracted to start-ups whose funders have previous experience or have received some public attention. This also means that they have to promote a project with high potential.

Numerous studies confirm that venture capitalists promote a borrower and his organization by following some criteria. Starting with the geographical analysis to understand whether that particular idea could be appreciated in a specific place. This aspect is closely related to the choice of the sector. It depends on the results of a market examination combined with what the venture capital firm can actually offer. The third part to consider is description of the company and current situation: what is the stage of development at this moment? How many resources have they already benefited? What is the number of this round? (Jeong et al., 2020; Harroch, Sullivan, 2019).

Some tactics that will certainly not be appreciated or considered are the request via emails to be received for an eventual interview or presentation of the organization, idea, business plan and financial situation. Whereas, the most popular approach is the introduction through a trusted professional who could be a current client of the VC, a colleague, a lawyer.

Once the start-up has secured a meeting with the venture capital firm, the following step is the creation of a “pitch deck”. This is a brief presentation of the company’s overview. It is also a commonly used practice for angel investors. It should include a graphic analysis of the current trend; insert images to communicate the values that will represent the fundamentals of the company; include all possible elements that help demonstrate that the project would have broad market reach and impact; justify all the decisions that have already been made in the organization to define an excellent business plan; send the presentation before the meeting and provide a demo of the products for the meeting. Some rules not to follow regarding the writing phase: too much textual or visual information is not appreciated; some details should be inserted in the pitch deck, others would be better explained verbally; competitors are always a challenge to be taken into account; in the data analysis of the actual situation, include data from the current month or a month before the presentation even if it has been prepared in the last eight months (Harroch, 2017).

Beyond the realization of the pitch deck, the venture process may require several months. In fact, just scheduling the first meeting could take weeks and will be followed by other meetings for the final selection. Also considering the participation of other start-ups and small and medium enterprises, this phase is the most crucial for the choice of venture capitalists. Once an organization is designated, the procedure continues with the negotiation part where legal documents are analyzed to decide on the investment amount. The borrower’s request could be higher than the value the VC would offer and, again, the negotiation terms need to be delineated. In this phase, the principal indicators to include are the estimated capital amount, the form of the investment (the most diffused are convertible preferred stock), the organization’s evaluation, the liquidation preference of the equity investment. This is followed by a selection of rights for the lender ranging from the Board observer rights, approval or “veto” rights, “preemptive rights” (rights to be present in future financings activities), rights to receive periodic reports, redemption right, drag-along rights (the company has the power to force all shareholders to vote for a sale of the organization only if the sale has been approved by a certain percentage of shareholders), anti-dilution protection (to protect the investment from eventual dilution in future rounds).

This preliminary analysis is essential to prevent unpleasant situations. The venture capital firm cannot predict what the beneficiary's future behavior will be. Indeed, the company's reaction after receiving funds is a crucial indicator of its learning capacity and internal organization. It will utterly impact its future performance and sustainable growth. On the other side, venture capitalists may choose to invest in a specific start-up to support it or for opportunistic intents. During initial meetings, the borrower has to disclose core information about his company and, sometimes, the lender will use the data received for his personal business. Starting from the idea, how it was developed, what technology to implement, what industry to choose etc. As a result, the reputation of venture capital will be negatively impacted and, at the same time, the future of the start-up will be inflated. For this reason, organizations in their early stages need to consider numerous factors before identifying the best option to receive external resources as the credibility and experience of the lender.

At the funding stage, another solution could be evaluated: small business loans. They are offered by a wide number of investors and present various characteristics in order to be adopted by businesses with diverse needs.

The small business line of credit is a significant support to manage the firms' resources and ongoing expenses. It fixes a limit on the maximum amount accessible and requires a monthly payment for service. In general, interest accrues when the line of credit is activated and the capital is effectively used. The requested amount is amortized over the year and an annual fee has to be paid to renew the subscription. If the agreement terms are not fulfilled, the company is forced to pay the entire amount at that moment.

The company's accounts receivable is another credit facility. The most attractive aspect is the variable interest rate and the guarantee of immediate funds based on the level of the firm's accounts receivable. The sum of resources disbursed through the AR line will increase again when the accounts receivable are repaid by the company's customers.

Another popular funding method is the working capital loan, a debt used directly by the company to finance common operations. In fact, its performance may be affected by market volatility, seasonal fluctuations, implementation of a single activity and similar situations. Some working capital loans are unsecured and this could cause an issue for a company in its first stage. To compensate for the lack of a valuable credit history, possible answers are pledge collateral or personal guarantees. The amount of capital commonly requested ranges from $5,000 to $100,000 and the preferred typology of loans has short duration (minimum one month, maximum one year).

Small businesses term loans are selected for capital expenses, business activities or expansion. They usually have a set dollar amount, a monthly interest rate and a repayment period ranging from six months to three years. The payment can be amortized over the life of the loan. Other interesting details are the forms of interest, variable or fixed, and term loans, secure or unsecured. It is an option chosen by small enterprises that need financial resources for a consistent expense.

Another category similar to small business term loans is small business loans. These are low interest rate loans offered by banks with more favorable interest rates and repayment terms than other loans for small and medium enterprises. This is made possible by the guarantee of the U.S that Small Business Administration (SBA), which affects the strict requirements of the selective process. The capital usually requested varies from $30,000 to $5 million.

When purchasing equipment, equipment loans are available as a favorable solution. They are characterized by a down payment of 20% of the purchase price of the asset and the collateral is the asset itself. The loan amount ranges from $5,000 to $500,000, the capital is amortized over two to four years, the interest can be fixed or variable and the rate is monthly. Equipment can also refer to vehicles and software.

Beyond small business loans, several backers can support SMEs, starting with online direct lenders. They comprehend reliable enterprises that offer working capital loans, cash advances and loans with a short duration in amounts from $5,000 to $500,000. Some websites act as lead generation services by providing access to several lenders; some cases are Fundera and LendingTree.

The most notorious lenders are commercial banks, characterized by rigorous loan underwriting criteria and long application process, but also by a high level of reliability. Some acknowledged examples are JP Morgan, Citibank and Wells Fargo. In the case of local organizations, numerous community banks can provide small business loans.

Peer-to-peer lending sites can act as intermediaries between small borrowers, institutional lenders and individuals. Some popular platforms are Funding Circle, Prosper and LendingClub.

Some bank lenders include the possibility of issuing loans backed by SBA, offering the aforementioned advantages to small and medium enterprises.

In this scenario of numerous and competitive financial options, the best strategy to understand which alternative to select is the analysis of key terms: the interest rate calculated on the loan, whether it is variable, fixed or occurs monthly, quarterly; if the loan payment presents a deadline or can be amortized over time; the calculation of administrations fess, underwriting fees, loan processing fees; any collateral or securities required; whether a lender can call for loan default and under what circumstances it is possible (Harroch, Sullivan, 2019).

1.5. The Importance of Microfinance in Developing Countries

The term Third World Countries refers to developing and underdeveloped nations. At this moment, more than 10% of the global population lives in deplorable conditions and the principal causes are political and economic inequalities rather than the lack of resources. Some general situations are always present in Africa, Latin American and Asia such as high birth and infant mortality rates, ignorant populations, low economic development, illness, starvation, agriculture, poor health care and death. Furthermore, one of the main issues is the exclusion of underdeveloped nations from financial institutions. This affects the possibility to start a business and, consequently, it adds to the causes of raising unemployment and poverty rate.

Access to official lenders is not guaranteed to 90% of the population in developing countries and this phenomenon can be represented through the “Vicious Poverty Round”. Investment capacity is limited and this utterly impacts on declining productivity, low domestic saving and revenues fall. If entrepreneurs have barriers in engaging new business ventures, economic development stalls. Although microfinance is a significant tool for seeking capital, it is consistently influenced by restrictions and hurdles. Once some funds are received from low-income families, they are invested to help with children's nutrition and ensure they can attend school (Robinson, 2002). These are only the first steps to build the fundamentals for economic growth. The support of financial services enhance the family’s self-confidence for increasing their assets and income. Therefore, the application and diffusion of microfinance allows the creation of an environment where political participation and democracy should be guaranteed.

To understand how microfinance works, it is better to start with an explanation of the concept of microfinance. It is a financial institution that guarantees access to capital to small organizations and low-income individuals. Some of the principal services are microsavings, microcredits, microinsurance and payment mechanisms (Kagan, Julia, 2018). It is a crucial option for situations where the borrower’s profile does not present the compulsory requirements to be accepted by banks. It is a typical condition prevalent in the poorest segments of the population, socially and geographically isolated from other members of society.

In the beginning, only two principal methods of microfinance were applied to offer financial services to clients. The former was the Relationship Banking for businesses and individual corporations. The latter referred to group requests for obtaining financial resources and, therefore, group-based arrangements were created to satisfy this particular request as well.

Actually, microfinance is the principal solution diffused in underdeveloped countries to enable low-income families to manage their earnings, promote economic development, job opportunities and the growth of small and micro enterprises. On the contrary, numerous criticisms have emerged to highlight the debt risk that could be caused by microfinance due to the low educational level of borrowers and challenges in starting a business. Related answers can be found in the description of strategies promoted through microfinance to reduce the poverty rate and offer information, supplies and resources.

Before microcredit, the bank was the only channel for seeking capital. Unfortunately, borrowers had no collateral or security to repay their debt. So high was their risk of default that underprivileged individuals were excluded from obtaining any loans. Numerous international agencies tried to reduce the poverty level through resource distribution which was significantly hindered. Therefore, only a portion of the population effectively benefited from the funds offered. Whereas, microfinance guarantees access to modest loans with reasonable interest rate and reimbursement procedures. So impressive and efficient is the program that this process has filled the gaps caused by other financial institutions and has also reached women in remote and rural areas.

Numerous studies have proven that preferential loans cannot be implemented as a valid solution. Due to the limited number of funds, resources are distributed only among companies with a higher reputation and, especially, strong banking contacts. The rest of the firms and small enterprises remain isolated and the poverty rate is not affected. In rare situations funds are obtained by small firms, but they are loans with a modest recovery rate. The general failure behind the financial institutions is utterly increased by the presence of government owned banks that are either too passive or have insufficient managerial capabilities, preventing the possibility of any improvement.

On the other hand, microfinance has achieved this goal through collaboration among organizations, thanks to which transaction costs and default rates have been reduced. Indeed, the combination of group loans, compulsory deposits and flexible payment schemes allow the distribution of funds to people who effectively need them. In the community, the relationship between loan officers and borrowers is crucial to identify the details of the investment and match them with the amount of capital to be requested. Collateral is substituted by group behavior that enhances the credibility and reliability of borrowers. The creation of a community is another fundamental phase to define a collaborative decision-making process and an accounting system in village organizations. To improve the capabilities of borrowers to repay the loan, some loan-granting institutions provide educational aid and training activities. The aforementioned aspects contribute to the success of microfinance in underdeveloped nations and promote its global implementation as a measure that effectively reduces poverty rate. In general, the difficulties of capital search in abject populations are the result of the lack of access and comprehension of multiple factors: primary resources, information on market prices of goods produced in those territories, health situation, services and structure of public institutions and, especially, information related to their legal rights (Almaamari et al., 2022).

Another aspect to be evaluated relates to the social condition of women in developing countries. Financial institutions consider the household poverty rate too high to be reduced. On the contrary, microfinance operates in this sector as well, providing job opportunities for women who previously had no possibilities to maintain themselves. It supports their empowerment by promoting several activities from entrepreneurial lessons, educational training to the theoretical and realistic business definition. As a result, women acquire more confidence and self-esteem, becoming more aware of their potential and expanding their knowledge. They usually invest the resources received for their children's education. In other cases, they use the loans for the development of their enterprise. This means that women become self-sufficient and can contribute to their family's expenses. Another risk may also occur: the loan is not invested efficiently and becomes a burden rather than a source. To avoid this situation, before granting a loan, an analysis of how it will be implemented and invested is essential.

One of numerous studies on women-led enterprises in developing nations was conducted in South Punjab, Pakistan, where Islamic microfinance institutions (IMFIs) operate. Data were collected through questionnaires from female borrowers who benefited from IMFIs. The results prove that 62% of them launched their own business through microfinance resources and the IMF contributed more to their entrepreneurial performance compared to non-IMFIs. In addition, the descriptive analysis reveals interesting information: women with five or more children and between the ages of 31 and 40 start their own business, which may vary according to their educational level; early age marriages are the main reason for girls to drop out of school, whereas the principal motivations for launching a women-led firm are gender inequality and responsibility for the welfare of their children; the preferred sectors for investment are boutique, tailoring, stitching, retail shops and beauty salons (Faridi et al., 2022).

Therefore, microfinance institutions represent the most efficient option for underprivileged populations to effectively improve their living conditions and create a better future for the next generation.

II Prosocial Crowdfunding & Kiva Experience

The crowdfunding phenomenon has completely transformed entrepreneurial finance in the past decade. Despite its global spread, researchers have just begun to analyze the complex dynamics of crowdfunding. In particular, lending-based prosocial crowdfunding is a typology of which our awareness and comprehension is still limited (Berns et al., 2020).

This chapter offers an overview of the potential of crowdfunding for projects promoting a social cause, followed by a description of a specific platform that operates effectively in this field, called Kiva, and what studies have emerged through the analysis of main variables in Kiva's database.

2.1. Prosocial Orientation in Crowdfunding

In developing countries, local microfinance institutions publish details of loans issued on prosocial crowdfunding platforms. They apply the loan-based crowdfunding model to create social value through their initiatives (Meyskens and Bird, 2015). The success of their mechanism is based on the combination of multiple factors, such as interest-free microloans, open-calls directly on online platforms and small contributions from numerous individuals without the intervention of traditional financial intermediaries (Mollick, 2014: 2). They represent a further innovation in the Fintech scenario and an evolution of peer-to-peer finance (Bruton et al., 2015).

2.1.1. The Moral Foundation Theory applied in the Crowdfunding Scenario

Numerous studies claim that investors are more attracted to non-financial returns. In fact, projects launched by borrowers who live in daunting conditions are more preferred by lenders because they induce particular feelings of compassion and comprehension (Belleflamme et al., 2014). Indeed, if campaign descriptions are written with human-related language or with words associated with altruism, they exhibit a higher likelihood to be chosen (Pietraszkiewicz et al. , 2017). When this 'warm-glow' sentiment is evoked and non-economic benefits are alluded to, the probabilities of fundraising increase considerably. This can happen in several situations, for example when borrowers highlight their organizational values or if psychological qualities such as resilience, certainty, hope and optimism are mentioned in the explanation of the initiative. The emotional aspect is absolutely underlined in prosocial crowdfunding to attract more funds, and this characteristic is also shared with angel investors, where entrepreneurs have the ability to emotionally engage investors and influence their financial choices (Huang and Pearce, 2015).

Another driver of crowdfunding participation is the willingness of being a member of a community that shares the same values promoted through the campaign (Gleasure and Feller, 2016). It has also been proved that the crowd tends to prefer charitable projects, where obtaining a loan means making a consistent difference in the lives of borrowers (Heller and Badding 2012). Furthermore, investors involved in prosocial lending follow the principles of microfinance. In this field, the impact of single donation is inferior than in equity-crowdfunding, because it enhances the benefits to borrowers and their community in the long run.

Information asymmetry is another aspect to consider in prosocial crowdfunding. In general, it is characterized by an open call in which borrowers can be selected from the total number of lenders. Information asymmetries emerge when borrowers, one part, have the information that lenders, who represent the other part, would like to receive in order to make a proper decision (Stiglitz, 2002). If the number of loans is significant, investors will prefer projects where the human side is highlighted (Allison et al., 2015). In this circumstances, moral foundations in the borrowers' profile could diminish the information asymmetry related to lending and enhance the social return of the campaign.

The aforementioned aspects are a consequence of the application of Moral Foundations Theory (MFT). According to the MFT, if the borrower's description presents some clues of morality, the lender will evaluate them as the borrower's capacities for creating social value. As a result, this typology of an entrepreneur's profile influences the backers' decision making process.

This theory was formulated to delineate a general framework of moral foundations. Numerous cultures of the actual society have built their psychological system on these principles. They can be perceived as an evolution of an individual's moral foundations. They are innate and inherited from the previous generation and differ from virtues because they can be modified through the sentiments felt by an individual during his lifetime. The Moral Foundation Theory classifies the moral system into five general foundations: fairness/reciprocity, harm/care, ingroup/loyalty, authority/respect, purity/sanctity. The first and the second are universal, all individuals have them; the rest are defined as primarily conservative.

Fairness/reciprocity comprehends trustworthiness, justice and integrity and encourages relationship building among individuals. Harm/Care refers to kindness, compassion and the human propensity to protect and support individuals from harm. Ingroup/loyalty strengthens values such as self-sacrifice for the group, devotion, patriotism and distrust of members of other groups. Authority/respect is based on voluntary defense, admiration and esteem for good leaders. Purity/Sanctity promotes sentiments of elevation, spiritual virtue and sacredness with intuitions about divinity (Haidt and Graham, 2007).

Borrower descriptions with some moral system-related cues have been shown to exhibit shorter funding periods in prosocial crowdfunding. In fact, it decreases the level of information asymmetry and amplifies the potential of borrowing for social value creation. Furthermore, universal moral foundations have a higher effect on lenders' response than predominantly conservative ones, especially if they regard authority/respect or purity/sanctity. However, to produce an immediate reaction of investors, primarily moral foundations should be highlighted in the descriptive part about the borrower's core business. The final result, combined with the universal moral foundations, will utterly attract a crowd of pro-social lenders and certainly reduce the funding period of the loan (Jancenelle et al., 2018).

2.1.2. The Effects of Prosocial Framing on Crowdfunding Performance

Another feature to evaluate in this typology of crowdfunding is the hybrid orientation. When social and economic values are both promoted in a campaign, the odds of success are higher than initiatives dedicated to only one inclination. So impressive is the backers' support that funds are collected faster and the possibilities of reaching the final goal are definitely multiplied (Moss et al., 2018).

This theory is also confirmed in reward-based crowdfunding, in particular in technology-oriented projects, where an altruistic approach integrated with sustainability values are more likely to attract the attention of the crowd and its monetary contribution (Calic and Mosakowski, 2016). However, Kim et al. (2016) suggested another perspective: in the actual reality, projects that promote a social cause have less chances to achieve their target. Nonetheless, if they succeed, the final amount of money earned is consistently more significant than the one raised in the rest of typologies of campaigns.

Further research asserts that pledges generated through crowdfunding can be divided into three categories: guarantees not corresponding to a benefit (for example nonreciprocal giving), promises equal to a concrete reward (like transactions) and deals that exceed the payoff (like reciprocal giving). If lenders' return is non-material, the likelihood of a campaign's success is drastically reduced. At the same time, projects with reciprocal benefits are positively related with the achievement of the final aim, while the chances diminish for initiatives based only on transactions (André et al., 2017). Moreover, as has already been explained, affective commitment and positive sentiments generated by the borrower's description consistently support the success of a project.

Two aspects have not yet been evaluated in the aforementioned analysis: the crowdfunding context and the implicit combination of social and economic objectives on the same platform. Regarding the first element, some platforms, like Kiva.org, clearly present the social mission and investors are aware of noneconomic compensation derived from their participation. In fact, scholars have confirmed that the social cause is an additional factor in determining the success of a campaign. On the contrary, the second category of platforms, like Kickstarter, promotes the self-interest of lenders and the benefits generated by the social mission of the initiative are only a consequence of the structure of the project. In fact, reward-based crowdfunding attracts investors because of the reward itself that will be received once the final target is reached. This general debate has more relevance in a scenario where the personal interest should be balanced with the social interest of investors. In fact, several studies analyze profit and economic companies separately, considering them as different phenomena, in order to detect the main investors' motivations to participate in crowdfunding. According to theories of entrepreneurship, the most popular concept is economic rationality: lenders are usually motivated by their own interest, which cannot be simultaneously combined with a social interest. As a result, initiatives characterized by an economic return cannot also pursue a social cause.

This thesis is confirmed in crowdfunding campaigns financed by institutional investors. On the other hand, the number of businesses that integrate and merge personal and social aims is continuously increasing and is achieving a wider audience and a more significant amount of resources (Van de Ven et al. 2007). Another interesting aspect regards the difficulties of promoting a pro-social message in the crowdfunding scenario, even though it has an innovative and digital background. Several analyses of communication have demonstrated that the level of information is so overblown that it is challenging the identification of right and useful news. Therefore, further examination of the framing of information is fundamental to properly choose the crowdfunding platform to invest.

The concept of framing relates to the organization, selection and packaging of information about a cause, a service, an issue, a product or a situation that allows the audience to interpret, understand and elaborate it (Giorgi and Weber 2015). It has two main functions: acting as a marker in the distinction between two similar objects; gaining the appreciation and support of the crowd. In the former case, when the term “green” is used, marketers highlight the environmental cause promoted by the project to attract investors interested in ecological themes. In effect, framing encourages audience selection from a set of options. While, if framing is implemented in social movements, financial reports or political meetings, it is considered as a powerful means to select a specific choice and motivate the audience in the achievement of a particular goal.

Numerous scholars have analyzed prosocial framing in the context of crowdfunding and argued that it effectively contributes to improving a company’s performance and promoting optimistic investment recommendations (Olsen et al. 2014). In fact, when the environmental issue is mentioned in a particular initiative, clients’ behavior improves significantly compared to the general attitude so far.

Nonetheless, how the frame is inserted into a project description is crucial to generate the expected reaction of the crowd. The frame can be included in written texts by following two general approaches.

Firstly, some entrepreneurs may use multiple terms related to a specific theme to emphasize a particular frame. At the same time, too many connections could produce the opposite effect. In general, one or at most two references are sufficient to capture the audience's attention. If this ground rule is not respected, framing references could be perceived as suspect (Levin et al., 1998). In addition, organizations that promote their activities with too much emphasis, have less probabilities of being funded (Parhankangas and Ehrlich, 2014).

Secondly, when a concept is highlighted in the textual description, the importance of other notions is seriously reduced (Burke, 1984). Therefore the resulting explanation may not present a complete and satisfying overview of the initiative, generating a sense of uncertainty. For instance, the accentuation of prosocial framing will hide the rest of values and ideas described in the campaign.

This can directly lead to another fundamental aspect to evaluate: self-interest always prevails over social interest. This means that individuals appreciate prosocial characteristics of products only if they are combined with their functionality (Crane 2001). Therefore, if the emphasis on prosocial framing is correct, the other benefits associated with a reward will also be highlighted and the consumer will make a proper decision. For example, a zero-emission campaign adopted by a firm and the natural ingredients mixed to create their facial products will be utterly appreciated by the client if the effects are also impressive on skin texture or final price.

The same situation is also evident in crowdfunding projects, where too much information only on customer relationship management will produce negative effects because the rest of crucial concepts fall in the background and the probabilities of success diminish. The solution is a balanced representation of economic and prosocial values. Furthermore, the importance of prosocial framing varies according to the typology of crowdfunding. For instance, in the reward-based form, items' characteristics should be underlined to draw the attention of the crowd. Moreover, the scheme of the information represents a fundamental aspect to consider, especially in digital scenarios, where it depends on the hypertext structure (Zhang and Salvendy, 2001). Interested investors can land on the campaign's page through a web-link or going directly to the platform page. In the latter case, every project displays a title, an image, a brief promotional phase and a hyperlink to its page. From a backer's point of view, he will see the title first, then the blurb and, only at a second phase, his focus will shift to the rest of the details. This means that the most important decision for borrowers regards the definition of the title and the blurb. When the text is too complex to be understood, the investor's attention will be dedicated only to the headline. If prosocial terms are inserted in these two sections, the opposite effect will result due to the low word count. Consequently, prosocial cues should be mentioned in the project description and strategically combined with the rest of the concepts to efficiently promote the initiative and reach the audience.

Crowdedness is another crucial aspect and represents the vast number of alternatives spread across digital markets that also enhances the competition among participants (Hansen and Haas, 2001). This phenomenon is extremely popular in the crowdfunding scenario where so high is the number of projects that it is really difficult to capture the attention of backers. The large set of initiatives requires a significant cognitive burden from individuals who have limited information processing capacities. When the amount of information increases, common human behavior applies cognitive strategies that make decisions less accurate and precise, but more feasible; they usually restrict attention to a limited number of features of a specific alternative. This means that individuals maximize the benefits and minimize the costs related to information processing.

Numerous studies prove that consumers only notice the most interesting characteristics of the products when the amount of information is consistent (Lurie, 2004). This means that the linguistic role is fundamental in the choice of a particular project. Consequently, prosocial terms should be added in crowdfunding initiatives to reduce cognitive burden. In fact, they can be considered as selection attributes and drive investors' attention to particular themes. Therefore, as crowdedness increases on crowdfunding platforms, prosocial framing should be considered as a strategic tool to attract more lenders and augment the likelihood of a campaign's success.

To sum up, prosocial framing is a fundamental aspect to include in the textual description of a crowdfunding project. If it is moderately emphasized, the possibilities of collecting more financial resources arise, whereas, in the case of overuse, the opposite effect is generated. In addition, the final result depends on the level of crowdedness of the online platform: as crowdedness augments, the negative outcome of excessive prosocial orientation is reduced, while the positive effect of limited prosocial orientation is reinforced (Defazio et al., 2021).

2.1.3. Sustainability-Oriented Campaigns and Their Implications in Crowdfunding

In the crowdfunding sector, the project creator should establish a relationship with the audience in order to have higher probabilities to succeed. Updates on projects development are significantly appreciated by backers and an appropriate linguistic style should be adopted to capture their attention (Cho and Kim, 2017). These aspects should be even more highlighted for sustainability-oriented projects, where investor can receive only intangible benefits (Parhankangas and Renko, 2017). This typology of initiative promotes long-term objectives, like the development of a social situation, and could be perceived as a lack of reward when compared to commercial projects. For this reason, the role of the creator is fundamental to involve investors in social initiatives rather than economic ones. The resulting uncertainty is a combination of the risky nature of the investment and the difficulty of measuring the impact of the project in our reality due to its long-term effects. To solve this issue, funders should be more focused on the realization of the campaign. A more detailed description of the project with images and videos should be associated with an explanation of the concrete realization of the project. According to Moral Foundation Theory, if the aforementioned factors are integrated with positive values and sentiments, the fundraising performance of the project can utterly increase and reduce the perceived unpredictability (Baucus and Mitteness, 2016). Therefore, creators of sustainability-oriented campaigns should strengthen their communication skills and support backers in their decision-making process to let them understand the opportunities and the importance promoted by social projects.

Numerous studies state that some characteristics of backers influence their funding decisions such as risk attitude, geographical distance and culture (Zhao et al., 2017). Moreover, the success of the campaign depends on the combination of multiple elements related to lenders, starting from their level of participation and integration, promotion of the project via social media, attraction of more potential backers who might be interested in the initiative etc. In general, the online word of mouth mechanism generates social benefits for the project because it increases its popularity. Lenders of sustainability-oriented projects are expected to leverage more than funders of commercial initiatives the aforementioned factors, for the higher level of uncertainty of reaching the final goal and the non-tangible reward that will be received from this typology of project (Hörisch, 2015). To overcome this challenge, creators should individualize certain figures among the crowd of investors who believe most in the social cause. They can be designated as official ambassadors of the campaign to spread news, credibility and positive feelings on their social network. Consequently, trustworthiness and reliability of the funder are enhanced, whereas perceived risks and negative sentiments are drastically reduced. Therefore, the role covered by backers in sustainable and social projects affects more campaign success than commercial initiatives (Petruzzelli et al., 2019).

Some dissimilarities have emerged in the confrontation of multiple theories on sustainability-oriented campaigns. Several studies affirm that the likelihood of success is higher for this type of initiatives (Calic and Mosakowski, 2016). On the other hand, other scholars believe that social and environmental projects are less profitable than commercial ones (Hörisch, 2015). It is also true that the probabilities of success are significantly influenced by the growth of the market. Nevertheless, a common theory should be defined to understand which factors are determinant for the promotion of social and sustainable campaigns or which elements should be included to reach a higher audience.

In general, to lower the level of uncertainty and mitigate the intangible nature of the goals being pursued, more tangible goods can be offered. However, increasing the number of pledges could encourage a further distinction among lenders, to understand if they really believe in project cause or are only interested in economic returns (Belleflamme et al., 2014). Based on the aforementioned theories, social and sustainable projects that provide more tangible rewards are preferred over the ones offering non-material goods. It is also true that a segmentation realized through the detection of the principal investors' motivations has a positive impact on the success probabilities (Petruzzelli et al., 2019).

A further aspect to evaluate regards local institutional support to launch a crowdfunding campaign. This aspect acquires more importance for sustainability-oriented projects, where institutional contribution is considered crucial to spread consensus and awareness among the population. Themes such as circular economy, climate change and pollution fall under this typology of crowdfunding and, when the cause is also supported by institutions, more lenders will be interested in the project and the chances of success will increase significantly.

In addition, cross-networking is another variable influencing a crowdfunding campaign. On one side, project creators choose the platforms with more lenders because popularity is one of the multiple factors influencing a crowdfunding initiative. On the other side, backers prefer platforms where high-impact projects are promoted, the financial return is impressive and the pledges received satisfy their expectations. This means that the structure of the platform is fundamental to facilitate access, guarantee supply-demand matching and advertise campaigns (Galuszka and Brzozowska, 2017).

Therefore, the crowdfunding platform covers a crucial role in the success of sustainability-oriented projects. In fact, if eco-friendly initiatives are launched on crowdfunding platforms, they have more probabilities of being funded. The marketing phase is also facilitated: copying the direct hyperlink to the project on social media augments the visualizations and reaches a wider audience (Petruzzelli et al., 2019).

The phenomenon of crowdfunding has only recently been analyzed by several scholars, despite impressive achievements realized over the past decade. Some studies prove that good performance and professional investors raise the amount of financial resources collected and this positive impact will be also reflected in the final market (Viotto da Cruz, 2018). Numerous findings affirm that crowdfunding has broadened access to the financial world and this effect has enhanced the growth of small and medium enterprises in economic and social sectors. While, only a relatively small number of scholars have examined what happens once a crowdfunding campaign has reached the desired outcome. The theory regarding the positive relationship between good performance and campaign success is confirmed not only for commercial crowdfunding, but is even more valid in social projects for diverse reasons. Firstly, sustainable-oriented initiatives find numerous obstacles in seeking financial resources through the traditional intermediaries (Choi and Gray, 2008). Consequently, crowdfunding has become the most popular and reliable alternative in the Fintech scenario. Secondly, prosocial crowdfunding draws public attention to social or environmental themes and promotes sustainable attitudes to be adopted in the actual reality. As a result, by sensitizing public awareness, crowdfunding has accelerated the process of the realization of a more sustainable society (Petruzzelli et al., 2019).

2.2 The case of Kiva

Kiva is an acknowledged crowdfunding platform founded in 2005 in San Francisco and operates across underdeveloped countries, where more than 1.7 billion people have difficulty accessing financial services. In fact, it targets borrowers who would like to receive support to improve their living conditions, launch their business or develop the quality of services offered by their small enterprises.

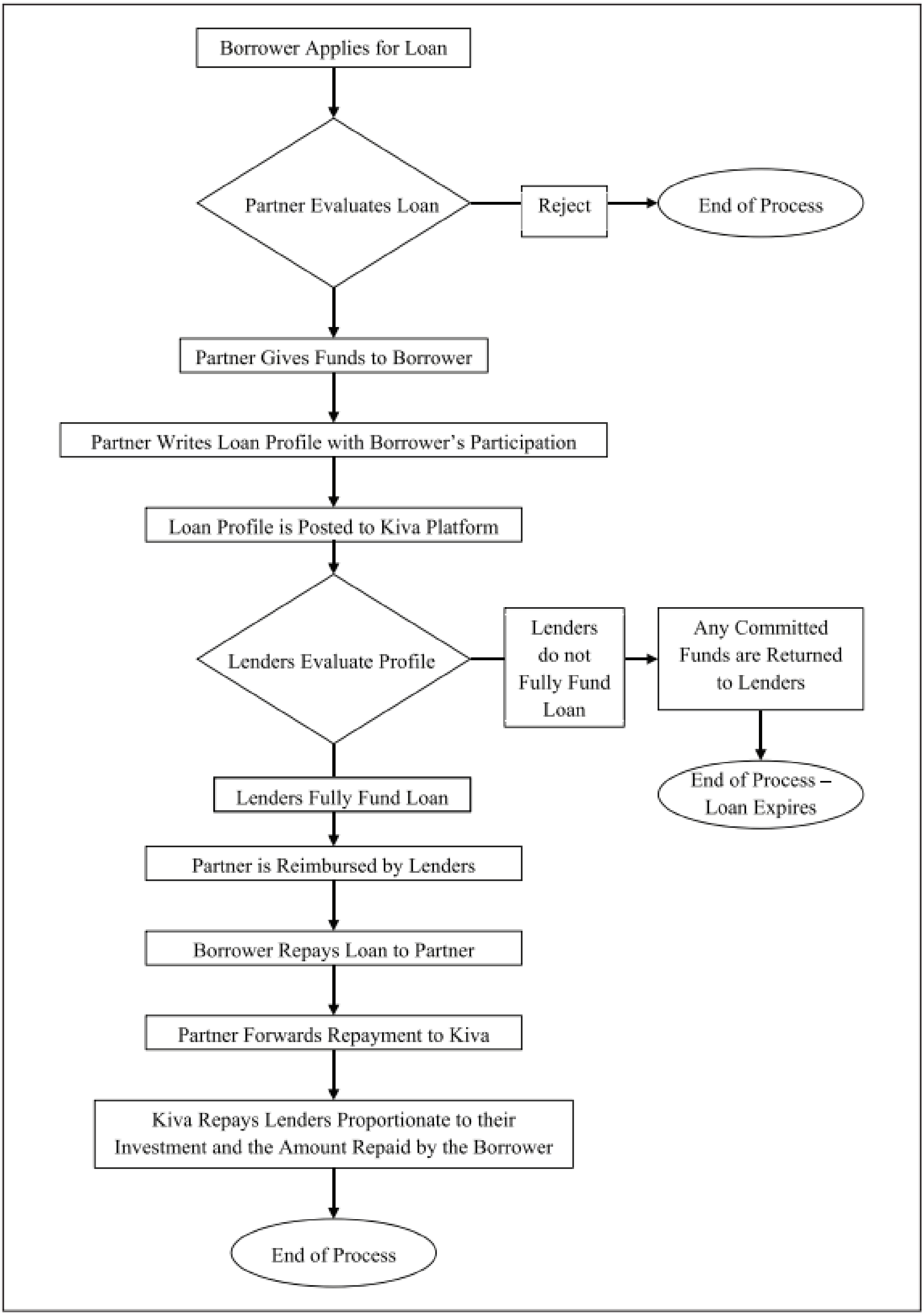
Actually, Kiva presents 2.2 million of lenders and 4.6 million of borrowers and operates in more than 70 countries on 5 continents. These significant results have been achieved through the collaboration with 450 volunteers and 4.351 Field Partners and Trustees, who support the connection with borrowers. Field Partners are usually microfinance institutions, but can also include schools, social enterprises and NGOs. Whereas, Trustees endorse loans to borrowers who prefer direct loans, which are not administered by Field Partners. They share a desire of effectively improving people's life conditions guaranteeing access to credit. Indeed, Kiva provides the possibility to invest in different projects simultaneously, with a minimum offer of $25. Lenders can also choose to make a donation in addition to the loans. In fact, operating costs are covered by Field Partner service fees and the contributions of backers, foundations and Kiva supporters. This fundraising model not only enhances sustainability, but also stimulates the crowd to aid the organization concretely. To have a general idea of the magnitude of this organization, just consider that 1.9 million of loans have been financed and an average of $2.5 million in loans is funded each week.

Kiva is constantly innovating to satisfy the needs of lenders with more flexible terms, promote community-wide projects and reduce costs to borrowers. Expanding a business generates benefits for the entrepreneur, his family and the community where he lives. It allows market growth creating more job opportunities and improving living conditions. In fact, one of the Kiva's most iconic slogans is "Lifting one, to lift many".

Its popularity and effectiveness are also reinforced by the application of rigorous and rigid due diligence requirements. In fact, Kiva encourages all lenders to inquire about the risks of investing in micro-lending organizations. The loan repayment rate can vary and it depends on economic and demographic variables. How the loan is administered is another fundamental aspect to evaluate in defining the related level of due-diligence. In fact, the majority of loans are managed by local partners with a presence of more than 80 countries. Kiva selects Field Partners before allowing them to post loan requests on the platform. It is also true that South America borrowers can only apply for direct loans due to their unprivileged conditions that significantly influence their credibility and reliability. Therefore, this typology of loan has a higher default rate and the repayment terms could be not met. Nonetheless, this demonstrates that Kiva reaches and supports populations that even microlenders cannot serve. In fact, it has received several recognitions for its commitment, such as the Wall Street Journal Financial Inclusion Challenge in 2015 for operational effectiveness or the Google's Global Impact Award for expanding financial access and sharing results (kiva.org).

Several studies have been conducted on Kiva to examine it from different perspectives and detect the principal determinants of its impressive global success. One interesting aspect to analyse is its general structure. It encompasses four actors: borrowers, lenders, partners and the platform. The cycle begins with borrowers seeking a loan through one of Kiva's partners (Figure 2.2.1). The latter plays the role of financial intermediaries between the borrowers and the lenders and evaluates the application by considering various aspects, such as the borrowers' profile, his needs, his family situation, his motivations for asking for a loan and how the loan will be effectively invested. Further financial analysis is provided to identify the borrowers' repayment capacity, borrowers' reliability, creditworthiness, estimated duration of the loan, relative default rate and other technical requirements. Once the due diligence is completed, the partner posts an explicit and coherent description of the loan details and the borrower profile on the platform. On the other hand, lenders examine the initiatives promoted on Kiva and decide which one to invest in. The following step comprehends the collection of funds, after which Kiva transfers the amount raised to the partner, who directly finances the borrower while also taking into account the interest rate calculation. The evaluation of this parameter depends on economic aspects, like the

Figure 2.2.1. The Kiva structure



Source: Moleskis M, Alegre I, Canela MA (2019), *Crowdfunding Entrepreneurial or Humanitarian Needs? The Influence of Signals and Biases on Decisions*, Nonprofit and Voluntary Sector Quarterly, 48 (3), 552 - 571 risk characteristics associated with a borrower, and the partners' ongoing operations. In the final stage, Kiva remunerates lenders proportionally to their investment (zero interest rate loans) and the amount repaid by the borrower, with an average repayment rate of 96.4% (Moleskis et al, 2019).

2.3. The Comparison of Research Questions based on The Kiva Dataset

In the past decade, numerous scholars have conducted diverse analysis through the implementation of the Kiva dataset. Various themes have emerged such as the impact of gender differences in crowdfunding or racial discrimination in microfinance etc. (Gafni et al., 2021; Luo, Ge, 2018). To detect which aspects of prosocial crowdfunding have not yet been discussed in the actual literature, the next paragraphs will illustrate a comparison of the topics emerged through the analysis of Kiva variables.

2.3.1. The Impact of Loan Purpose on The Success of a Prosocial Crowdfunding Campaign

The first study examines how much the success of a project is influenced by the characteristics of loans. In particular, two main categories can be distinguished: essential loans and business loans. The former is usually requested to support health, housing or educational needs. Whereas, the latter is a useful tool to promote business causes such as starting a new activity, expanding small enterprises and improving the quality of goods or services offered.

The data analysis reaches three different results. Firstly, in prosocial crowdfunding platforms, loans aimed to meet urgent needs are preferred over consumption loans due to the ethical behaviors of the backers. Secondly, as the size of investment increases, the default rate associated with a loan rises. Therefore, the level of uncertainty is higher, the risk of over-indebtedness becomes more feasible and backers will only finance loans of inferior size. Thirdly, loans promoted by female borrowers are more popular. Indeed, women receive expected financial resources more rapidly if they cover essential needs than ones launched by male borrowers. Unfortunately, the gender difference may cause an adverse effect: men would feel inferior to women and force them to dedicate only to household activities and family necessities (Gafni et al., 2021).

2.3.2. Racial discrimination in Microlending Platforms

In the traditional market, ethnicity is an indicator of creditworthiness and significantly influences investors in their decision making process. On the contrary, this information is not provided in crowdfunding. Consequently, other variables are perceived as signals of borrowers' reliability, such as the funding process of the campaign and the general attitude of lenders (Zhang and Liu 2012). In addition, White borrowers are considered more trustworthy than minorities.

Another prevalent attitude in crowdfunding concerns funders, who tend to behave more prosocial when their profile is easily identifiable on the online platform and, for this reason, support the minority with more enthusiasm (Chen and Li, 2009). Moreover, investors' contributions are relatively small in the case of microfinance. Therefore, they focus more on information related to social welfare of the campaign and reduce the attention on riskier factors like race or ethnicity.

The empirical analysis proves that ethnic/racial discrimination is not a determinant variable for the success of a crowdfunding campaign. Nonetheless, some differences can be pointed out. Investors support African Americans' loans with an inferior average amount, while American lenders are less represented in loans to Africa Americas than to Whites. Even though the ethnic or racial discrimination might also prevail in crowdfunding. At the moment the nationality of the borrower does not influence the probability of being funded because lenders come from diverse countries and are located all over the word (Luo, Ge, 2018).

2.3.3. The Determinant of Credit Default in P2P Microfinancing Platforms

Risk uncertainty has been extensively researched in credit markets and information asymmetry is a primary factor that increases repayment risks in peer-to-peer microfinancing platforms. In fact, microfinance institutions play the role of intermediaries between lenders and borrowers in order to facilitate their communication. In addition, the majority of backers do not have a financial background and cannot evaluate the actual default rate associated with the diverse loans (Herzenstein et al., 2011). On the contrary, MFIs have the possibility of monitoring the borrower’s behavior, obtaining more information on his profile and his credibility. Furthermore, they can apply some tactics to obtain an appropriate interest rate and fair contract terms considering also the comparison with other transactions in related markets (Hoff & Stiglitz, 1990; Stiglitz & Weiss, 1983). In fact, the analysis conducted on Kiva data confirms that the role of MFI is crucial for the prevention of repayment issues by individual entrepreneurs and groups of borrowers (Dorfleitner, Oswald, 2016).

Group lending with joint liability is an innovative aspect in lending crowdfunding. It substantially reduces the default risk rate because more borrowers can guarantee for each other’s reliability. It also facilitates monitoring by MFIs due to a previous screening is applied by each borrower to create a homogeneous group. Another advantage regards the dilution of borrowing costs: even though every borrower has the same loan contract, the related costs are reduced if the level of trustworthiness is higher. Consequently, more small entrepreneurs can join market initiatives and the group's average probability of loss of repayment decreases (Morduch, 1999). In addition, empirical study states that the default risk rate and group lending are not negatively related (Dorfleitner, Oswald, 2016).

Progressive lending is considered a significant incentive for loan repayment because it supports the borrower’s access to future loans based on the evaluation of his actual and past repayment history. Indeed, customers who have fulfilled the terms of the contract, are more likely to participate in future crowdfunding initiatives. In addition, associated default rates will consistently decrease and, consequently, their official profile will certainly improve (Morduch, 1999).

It is also true that some loans features should be examined in relation to the progressive lending concept. In fact, long loan terms could cause serious issues for borrowers living in poor conditions. They are forced to save enough money for repayment and, at the same time, the maturity of their activities could have already come to an end. On the other side, short term loans may negatively influence borrowers’ conditions because they would have a short period to regenerate the funds received. Kiva’s data confirm the theoretical assumptions: as the size and duration of the loan grows, the probability of default increases (Dorfleitner, Oswald, 2016).

Traditional microloan contracts provide regular installments one or two weeks after loan disbursement. Microfinance institutions also adopt this practice, despite the high transactional expenses, because they receive immediate information about their clients' repayment discipline and ability to repay. At the same time, this represents an education method to support borrowers comply with their obligations on a regular basis. In fact, MFIs can evaluate the borrower’s risk profile and income level. Repayment obligations should usually be satisfied after the loan is disbursed. This means that the borrower’s financial resources should have been collected before the loan application. In this situation, MFIs can intervene to support the borrower and lend part of the total investment amount, ensuring repayment even if unsecured (Armendariz de Aghion & Morduch, 2000). Few studies have approached this question and their results have not reached the same conclusions. In this case, two theoretical hypotheses can be delineated: the probability of default diminishes if loans have highly regular repayment obligations; the default rate increases with loans characterized by a grace period. Empirical analysis, based on Kiva data, states that grace period and loan term positively influence the probability of default (Dorfleitner, Oswald, 2016).

In peer-to-peer crowdfunding, female borrowers exhibit better repayment behavior in comparison to men. Indeed, microfinance institutions show that the women’s borrower profile is associated with a low default risk rate. According to global research, women are not only better at managing credit risks, but they also prevent credit default (D'Espallier et al., 2011). Beyond financial analysis, women-led businesses outperform male entrepreneurs (Kalnins & Willians, 2014). The achieved outcomes can also be evaluated through consideration of theories of moral hazard and risk-aversion.

Moral hazard is manifested when asymmetric information is present between two parties. In particular, after the agreement between the two parties is defined, one of them changes his behavior. Whereas, asymmetric information regards situations in which one party has more material knowledge than the other one. The related concept is adverse selection: it refers to situations where retailers have more information than purchasers. This phenomenon is usually characterized by asymmetric information. For example, when a seller knows more details about product quality than the buyer (Nickolas et al., 2022).

Therefore, women have a low moral hazard risk derived from a combination of multiple factors, such as observance of social sanctions, low level of mobility and limited choice of financial alternatives (Armendariz de Aghion & Morduch, 2010). The latter motivation expresses the social background of women in underdeveloped countries. They are usually excluded from the labor market of developing nations because of the numerous access barriers (Emran et al., 2011). The only solution for them to start their own business and improve their living conditions is microfinance. For this reason, they always comply with the terms of the loan contract and can be considered as better customers for MFIs. Nevertheless, Kiva's data show a weak relationship between the probability of default and women's role as professional micro borrowers. These results may evidence the consequences of the gender effect (Dorfleitner, Oswald, 2016).

III The Characteristics of Successfully Funded Projects:

Empirical Analysis of Microlending on Kiva

The empirical analysis was conducted to investigate the microlending situation on Kiva from 2005 to 2021. Considering the theoretical background illustrated in the previous chapters, some aspects of Kiva have not yet been examined in the actual literature. In fact, the borrower's profile has only been described with some variables such as gender, age, individual or group activities, nationalities etc. In addition, numerous loans features have been evaluated: funding time, default rate, size, terms, status (repaid or expired) etc. (Dorfleitner, Oswald, 2016; Luo, Ge, 2018). Nonetheless, the actual application of the loan associated with the borrower's personal situations have not yet been identified.

Therefore, two research questions can be individualized:

1. What are the borrower's backgrounds usually related to specific loan purposes?
2. What are the characteristics of the project that positively impact the borrower's possibility of being funded?

In the Kiva dataset, the general framework where the borrower's lives and his personal conditions are expressed by the *description* variable, whereas the future implementation of the funds in the borrower's life is represented by the *use* variable.

3.1. Dataset Description

The Kiva dataset comprehends different sections[2]:

- A single csv-format file of 219.337 observations and 26 variables covering the years 2005 to 2010 and four xlsx-format files with only the variables *use* and *description*;
- Regarding data related to later years, the followed structure comprehends a single csv-file per year, which variables are 26 and the number of observations ranging from 100.000 to 230.000, and two or at most four xlsx-format files with the variables *use* and *description*. For instance, the dataset of 2011 presents 109.334

[2] The Kiva dataset is available at http://kivatools.com/downloads

observations; while, the same number of observations regarding the variables *use* and *description* are divided into two xlsx-format files;

- Five xlsx-format files related to the Partners, Trustees, Activities, Sectors and Countries. In particular, the file of Partners has 563 observations and 18 variables; the "Trustees" file presents 1511 observations and 2 variables; the "Activities" file has 163 observations and 2 columns; the "Sector" file displays 15 observations and 2 variables; the "Countries" file shows 98 observations and 2 variables.

The overall study was performed on Jupyter Notebook, a web-based interactive computing platform available on Anaconda software, implementing Python Programming Language. The analysis is divided into diverse parts: starting with data cleaning, followed by data visualization, topic modeling and data modeling.

3.2. Data Cleaning

Data cleaning, or data cleansing, aims to improve the quality of data through the identification of inconsistencies and errors. When several data sources are integrated, this phase acquires more relevance because the probabilities of incongruences significantly augments (Rahm, Do, 2000) (see Appendix A).

After loading libraries and reading the aforementioned files, some common steps are applied in all datasets, starting with the calculation of null values across the different variables. The column *TrusteeId* presents the ID of the Trustees and it has only null values. For this reason, it is deleted in all datasets (this also implies that the eventual join with the dataset of Trustee is not applicable). In addition, some rows have all the information of the diverse variables in just one column. The best choice is to remove them due to the few number of cases. For example, in the dataset 2005-2010 they are equal to 33. Other null values are also in *city* and *loanName* variables and are substituted with the value 'Not specified' to avoid dropping the values of the remaining columns. Moreover, several datatypes are not correct as in the case of *fundraisingDate*, transformed into date format, or *ActivityID*, changed into integer. The last operation regards the application of "info" and "describe" methods in order to obtain a complete overview of the dataset and the main indicators such as mean, standard deviator, maximum, minimum, quantiles of each column[3].

Further operations are applied in specific cases, depending on data inconsistencies. In fact, in the feature *plannedExpirationDate* of the dataset of 2018, not only the date is displayed, but also the time. The last one is eliminated through the “slice” method that extracts only a section of a string.

The next step regards joining the diverse datasets per year. For example, the dataset that covers years from 2005 to 2010, is joined with the different datasets of the same year that comprehends *use*, *description*, *ActivityName*, *SectorName*, *CountryName*, *PName* and other features related to the description of partners. Two typologies of join are applied: the inner and the outer. The first selects records from two datasets if there are matching values in a common variables to both datasets (in this case is the *id* of the transaction). Whereas, the second returns unmatched values and matched values from both datasets (Brumm B, 2019). The latter is implemented for the creation of a unique dataframe with *use* and *description* variables, distinct per year because they are divided into two to four files to reduce the file size and facilitate the upload.

The results of the several joins are the creation of 12 datasets (one per year, except for the first that includes years from 2005 to 2010) with 46 variables and a number of observations ranging from 100.000 to 230.000. Due to the limitations of the RAM of Jupyter Notebook, a unique dataset is defined: 3000 observations are randomly taken from the 12 datasets and the resulted ones are outer joined. The final dataset involves all years and displays 36.000 observation and 46 features.

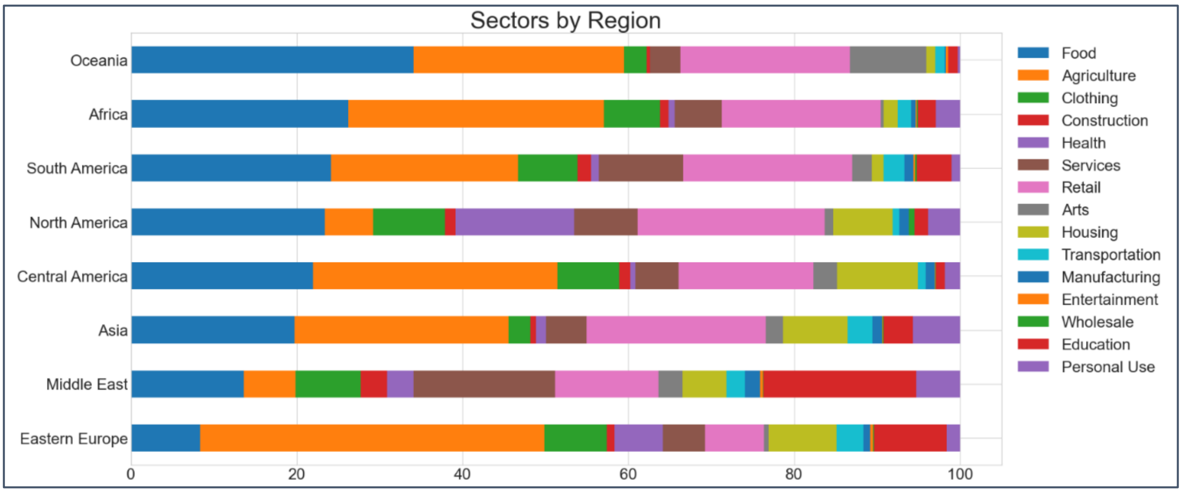
3.3. Data Visualization

Data Visualization is the graphical representation of data. It is considered a crucial part to individualize principal trends, calculate statistical summaries and apply data transformation in order to facilitate the interpretation and obtain a clarification of the phenomenon (Unwin A, 2020).

[3] See pandas documentation, version 1.5.3

To comprehend the current situation in developing countries, *Region* and *SectorName* variables are analyzed. The first refers to the geographical areas of the borrowers, while the second distinguishes the several sectors diffused in the nations described in the dataset. The Figure 3.3.1 shows that food, agriculture and retail industries are prevalent in Oceania, Africa, Asia, Eastern Europe and North, South and Central America. It is also noticeable that agriculture equals to 40% in Eastern Europe, food is around 35% in Oceania and retail presents a value of 20% in all areas, except that Middle and Easter Europe, where education and services have a significant diffusion.

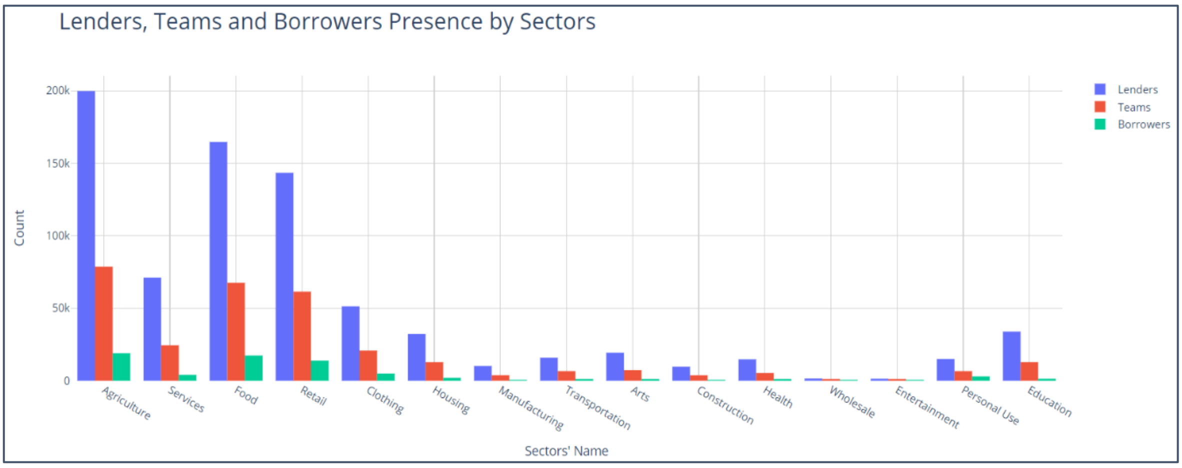
Figure 3.3.1 – Sectors by Region



The sectors' distribution can be also examined from a different perspective, considering the impact generated on Kiva performance by distinctive subjects. Lenders, teams and borrowers are fundamental factors to highlight in order to have a complete overview of the Kiva's structure. In fact, the presence of lenders is definitely more consistent than the ones of borrowers and teams. In particular, agriculture, food and retail are again the main categories where lenders prefer to invest in because the majority of borrowers' and teams' requests are concentrated in the three fields. The total number of lenders equals to 199.747 k in agriculture, while is around 150 k in food and retail. Whereas, borrowers are more homogenously distributed in the three sectors with a total amount of around 68 k.

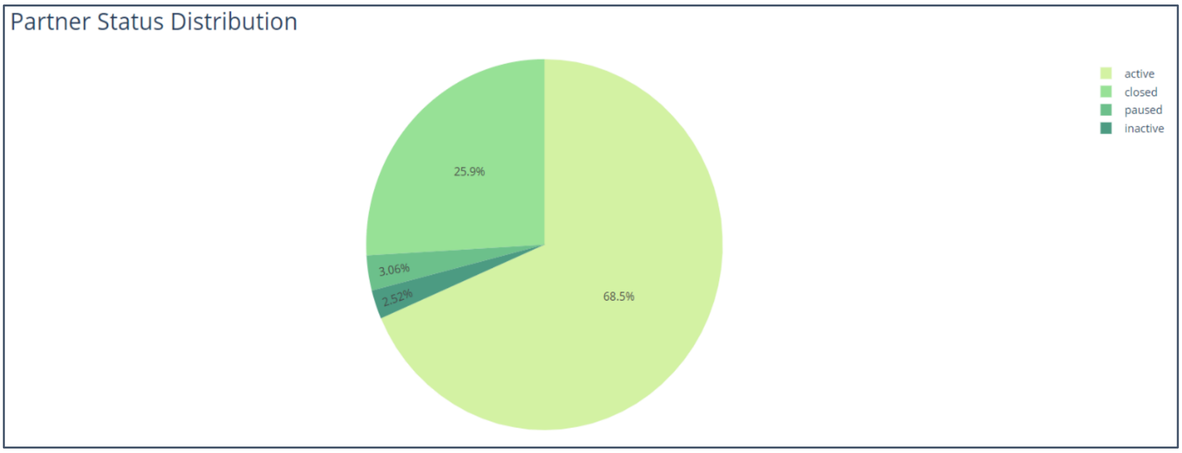
On the other hand, the number of teams is utterly inferior than the previous two classes and their average distribution is around 17k (Figure 3.3.2).

Figure 3.3.2 – Lenders, Teams and Borrower Presence by Sectors



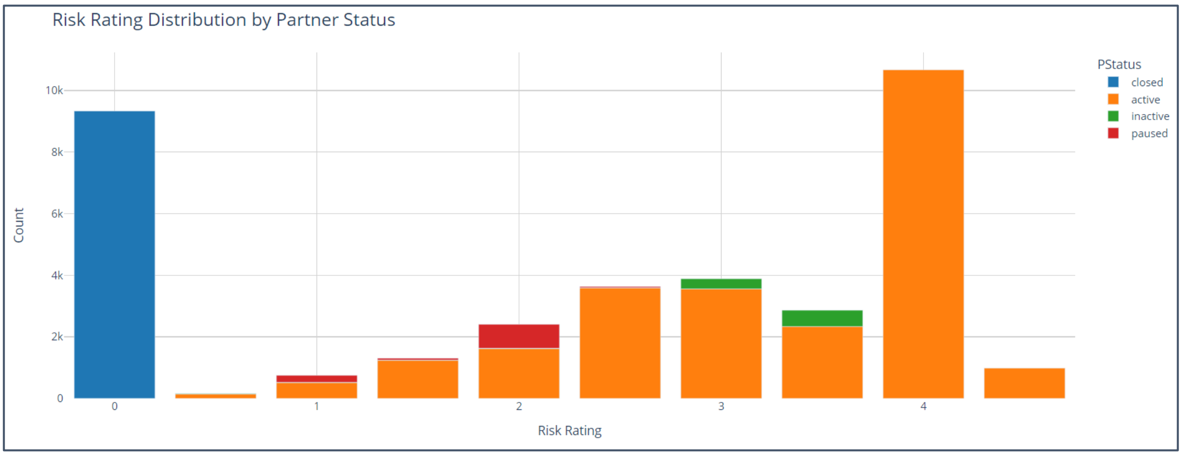
Another major role is played by partners, the Kiva intermediaries between borrowers or teams, and lenders. They can distinguished into four groups based on their current status: 24k are active and represents the 68.5% of the total. Whereas, 25.9% are closed, followed by 3.06% as paused and 2.52% as inactive (Figure 3.3.3). To better investigate the general motivations beyond their status, it should be evaluated the relative default rate categories.

Figure 3.3.3 – Partner Status Distribution



The Field Partner risk rating reveals the risk associated to institutional default of every Kiva's partners. It can be expressed with a minimum value of 0 until a maximum value of 5, but the interpretation is opposite. The Field Partner represented by rating of 5 has the lower risk of institutional default. On the contrary, the partners associated with rate equals to 0 have higher possibilities of default (kiva.org). As can be notice in the Figure 3.3.4, the role of intermediaries is crucial to increase the probabilities of repayment and, in fact, active partners have a default risk of 5, while the ones with a closed status are correlated to the maximum level of uncertainty.

Figure 3.3.4 – Risk Rating Distribution by Partner Status



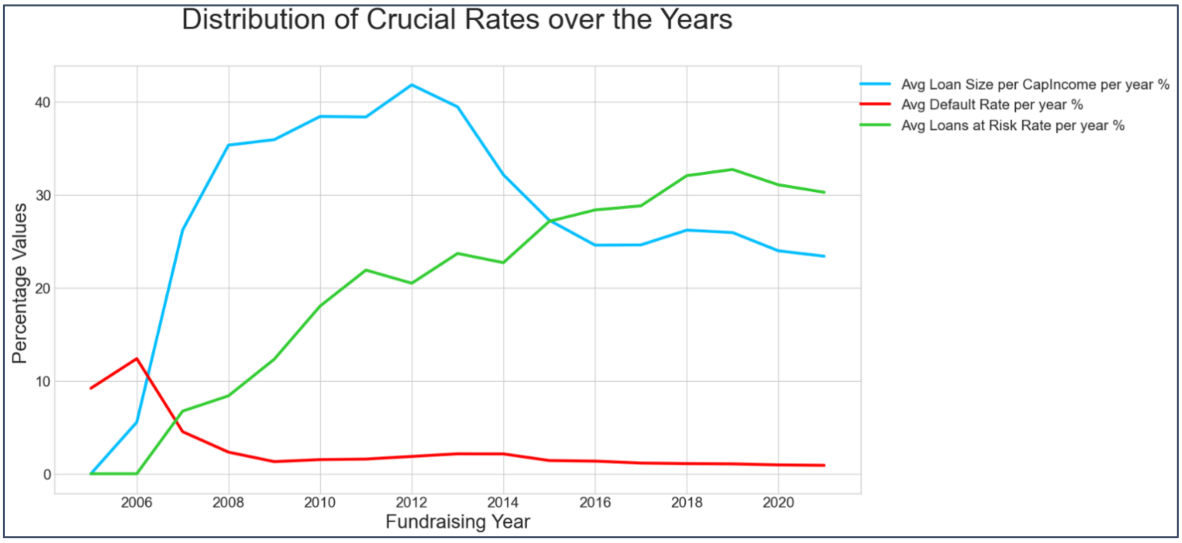
To better examine strengths and stability of Kiva's microfunding, other indicators can be underlined. A Field Partner's average loan size corresponds to the percentage of the a nation's gross annual income per capita. This means that loans with a low percentage of gross national income per capita are more diffused in underdeveloped countries. At the same time, they are more expensive for the Field Partner to generate, disburse and raise.

The default rate associated with the numerous partners equals to the ratio of amount of ended loans defaulted and amount of ended loans. In general, a situation of default occurs in two cases: when the raise of funds from a partner or a borrower is doubtful; when the total amount repaid as a quarterly reconciliation is inferior than the amount that should have collected as of 360 days prior and no repayments are confirmed during this time.

Loans at risk rate regards the percentage of arrears. In fact, it is equal to the percentage of loans being paid back by a Field Partners with a past due in repayment by at least one day. It does not indicate specifically a delinquency of a borrower or a Field Partner. However, it is a crucial index to consider in the repayment phase (kiva.org).

The Figure 3.3.5 illustrates that the average default rate reached its peak in 2006, followed by a declining trend. Whereas, the average loan size per capital income increased until 2012. In the next two years it started to decrease, while, since 2016, it shows a constant value. The only measure with an exponential trend is the average loan at risk rate. As a result, the graph confirms the theoretical assumptions of the previous chapters: even though the probabilities of not being repaid is growing, lenders choose to invest in unprivileged borrowers' project because they are more interested in supporting the social cause than obtaining the financial return.

Figure 3.3.5 – Distribution of Crucial Rates over the Years

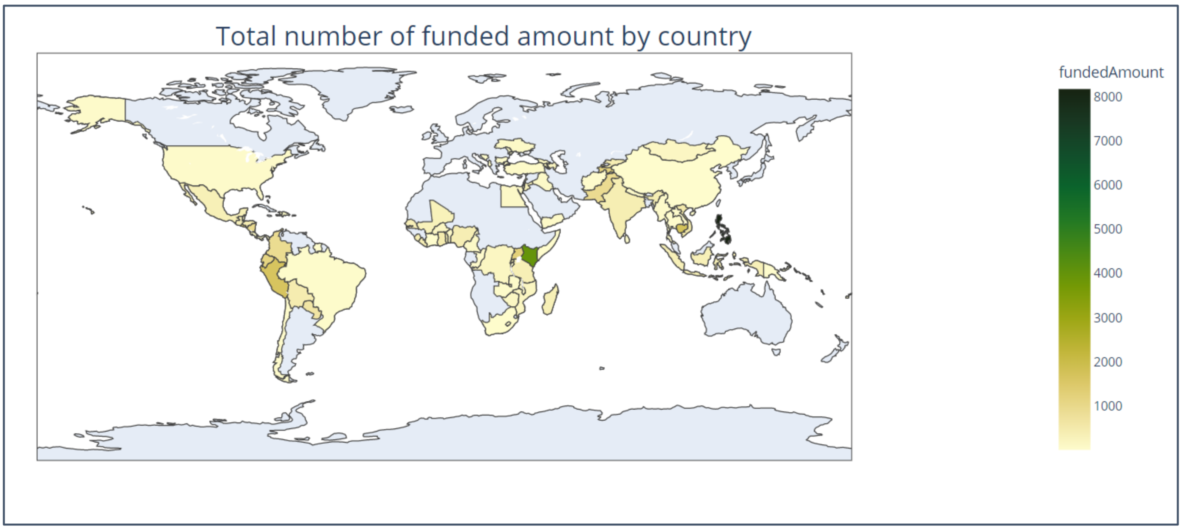


To understand the global situation impacted by the Kiva mechanism, the map offers the possibility to immediately individualize the funded amount associated with each country. The first place is held by Philippines, followed by Kenya and Peru (Figure 3.3.6)[4].

---

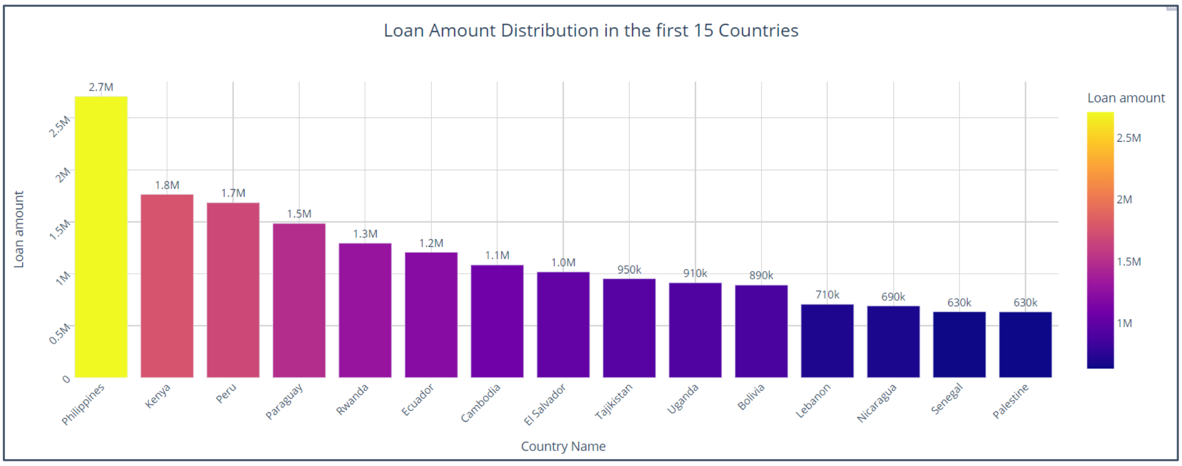
[4] To realize the graph shown in the Figure 3.6, another dataset is considered. It comprehends the list of ISO country codes provided by Wikipedia.

Figure 3.3.6 – Total Number of Funded Amount by Country



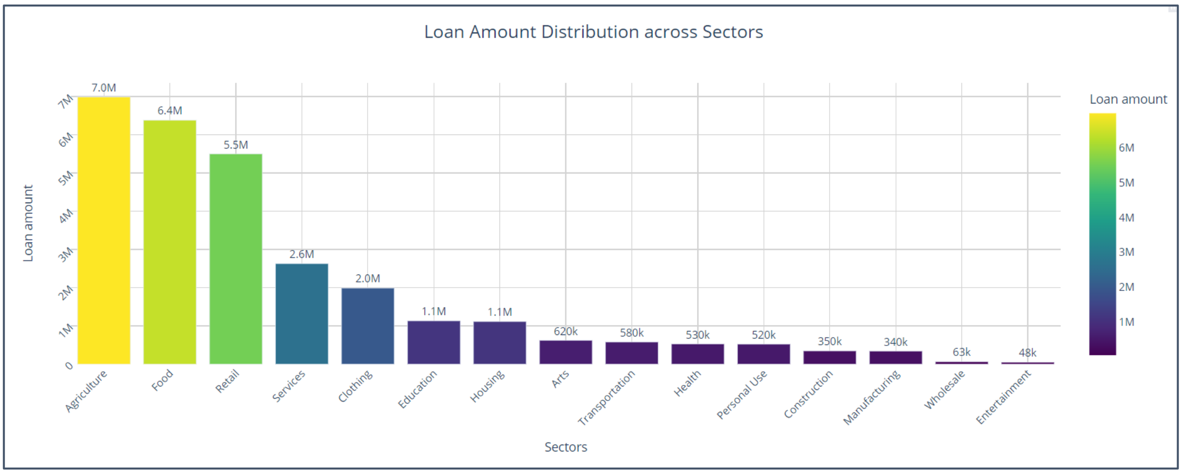
In fact, the situation is confirmed in the Figure 3.3.7 in which is highlighted the loan amount distribution across countries. Philippines presents the highest number of funds, equals to $2.7million, followed by Kenya with $1.8 m and Peru with $1.7 m. The last position is covered by Palestine, where the total is $630 k and still represents an impressive result.

Figure 3.3.7 - Loan Amount Distribution in The First 15 Countries



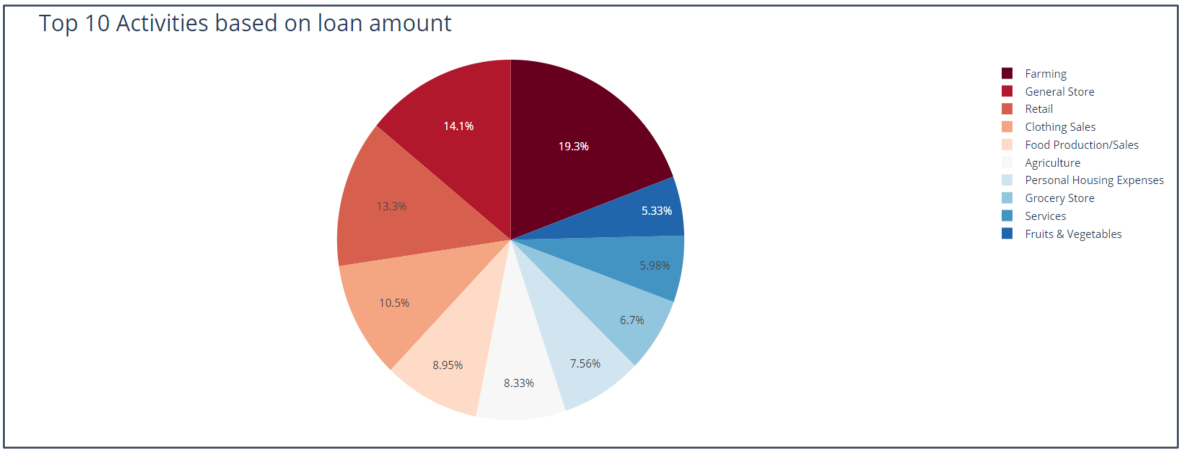
To have a more detailed insight, another interesting aspect to note is the distribution of loan amount across sectors. As can be deduced from the Figure 3.3.1, agriculture, food and retail are not only the most dominant industries, but also the branches with the most consistent loan amount. A value of 7.0 m is achieved in agriculture, followed by 6.4 m in food and 5.5m in retail (Figure 3.3.8).

Figure 3.3.8 – Loan Amount Distribution Across Sectors



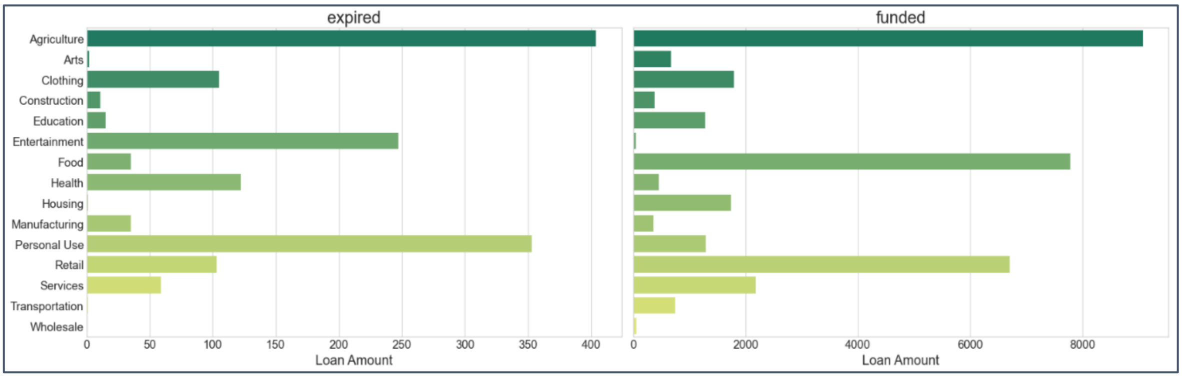
On the other hand, the activities with the more significant loan amount are farming with a percentage of 19.3%, general store with 14.1% and retail with 13.3%. The rest of them are showed in the Figure 3.3.9, where it is evident that the last place is covered by fruits and vegetables, if only the primary ten activities are estimated.

Figure 3.3.9 – Top Activities based on Loan Amount



Another point of view it is offered by the relationship between the *LoanAmount* variable and *status*. The latter indicates whether the loan is funded or past due. The two outcomes can be analyzed with a further variable: *SectorName*. As a result, the Figure 3.3.10 illustrates that the highest values of expired and funded loans are both present in the agriculture field. It is also true that the majority of loans are repaid, while only a small segment are expired. The left bar plot evidences that personal use and entertainment are the additional branches with the most significant amount of expired loans. Whereas, the right graph confirms that the maximum values are of funded loans are reached in food and retail.

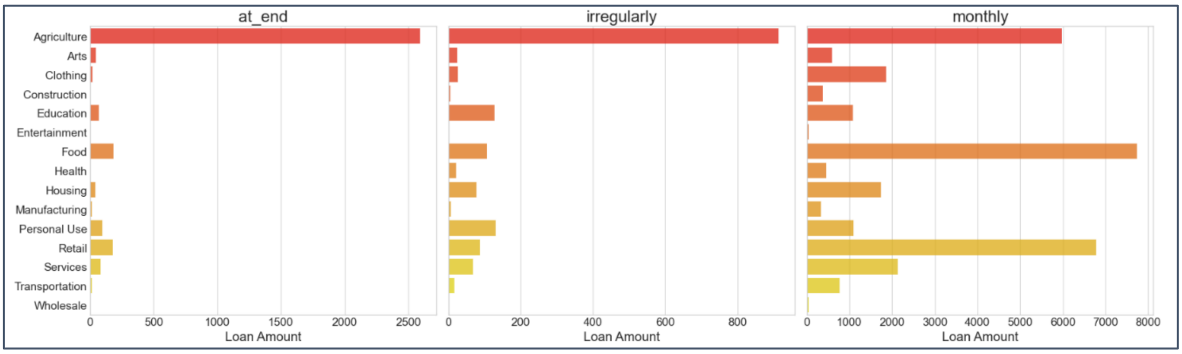
Figure 3.3.10 – The Status of Loans Categorized by Sectors



The last bar plot evaluates the relationship between *the loanAmount, repaymentInterval* and *SectorName*. As it can be noticed in the Figure 3.3.11, three forms of repayment intervals are provided: monthly, at the end of the loan duration and irregularly. The most frequent is the first one, followed by the second one and the third one. In addition, the "irregular" and "at the end" typologies are more applied in the agriculture sector.[5]

[5] To examine how the illustrated graphs are realized, see the Appendix B where some examples are included.

Figure 3.3.11 – The Repayment Interval Divided by Sectors



3.4 Topic Modeling

Topic modeling is a clustering method for textual data in order to rapidly identify the main themes discussed in a document (Blei, 2012). It has been implemented in numerous fields such as social sciences and economics. Recently, it has been also extended to social networks to detect the main concepts diffused on the posts and individualize eventual issues (Kim, Ju, 2019). In addition, it also represents a strategic tool in environmental sector. In fact, it has been applied to study the public perception on questions regarding climate change, pollution effects, economic growth (Tvinnereim, Fløttum, et al., 2017). Consequently, it has become one the most valid means in dealing with textual data, considering the substantial utilization over the last decade (Savin et. Al, 2023). In this analysis, topic modeling has been fundamental for the examination of *use* and *description*, two alphanumeric variables. Before this step, some crucial operations should be previously involved and regards the so called "Preprocessing" phase. It comprehends the removal of stopwords, followed by the tokenization and the lemmatization (see Appendix C).

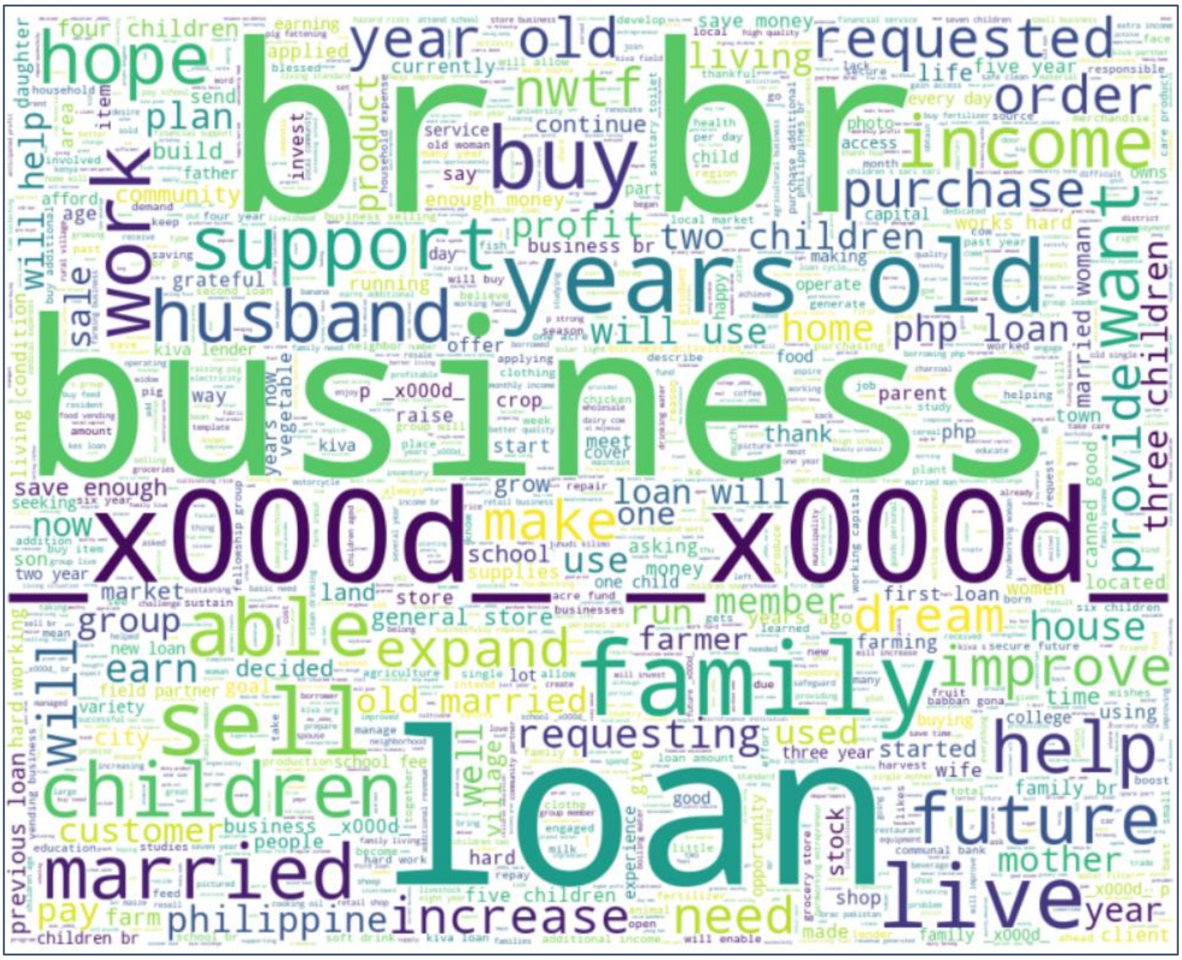
Stopwords are the most common words such has pronouns, prepositions, articles etc. and do not insert interesting information in the text. A list of stopwords is included in the WordCloud library and facilitates their elimination.

Once deleted the stopwords, the word cloud provides the visualization of the most frequent words in a certain column. The *use* variable presents terms such as buy, purchase, sell, business, fertilizer, pay, family, canned, good, as it is shown in the Figure 3.4.1. Therefore the principal loan purpose is buying a certain good, selling it, doing something for the family, purchasing something in the agriculture or retail sector etc.

Figure 3.4.1 – Word Cloud of the *use* variable

On the other hand, the word cloud of the *description* variable is not really explicable. In fact, syntax formatting elements influence on the identification of the most recurrent terms. Nonetheless, some emerging topics can be individualized such as business, loan, family, purchase, married, help, live, future etc. (Figure 3.4.2)

Figure 3.4.2 – Word Cloud of the *description* variable



Punctuation removal, tokenization, stemmatization or lemmatization techniques are applied in the next phase to define a document-term matrix. It lists the frequency calculated on each words present in a document or a text. Tokenization allows to divide a text in a list of words and can be applied by splitting the different terms on the white spaces or punctuation. On the contrary, stemmatization is a method to delete the ending of words and leave the remaining, most important part. For instance, “purchases” become “purchase” or “happiness” is transformed into “happy”. Lemmatization is similar to the previous technique and reduces all terms to their “base form”. For example, the words “gone” or “went” are modified into (to) “be”.

The further step is called “feature selection” and it is fundamental to guarantee the effectiveness of the topic model. Indeed, the *use* and *description* variables can present numerous unique terms, characteristic by a minimum frequency, and words with a frequency too much higher in comparison with the others. The both situations will be negatively impact on the modeling phase because they do not contribute to the identification of meaningful topics. Two methods can be applied in this case: the term frequency-inverse document frequency or the frequency cut off. The first associates a low score to words too frequent or too rare. The second establishes cut-off under which the words are too rare or over which the words are too frequent. In this analysis, the second technique has been applied with some modifications due to the data characteristics (Jacobi et al., 2015).

The *use* variable does not present critical aspects; only “buy”, “purchase”, “sell” words have a frequency of more than 5000. This value affects consistently the results of the topic model because the remaining terms have all a frequency lower than 5000. In addition, “etc”, “will” and “additional” are also removed for their effluent meaning. The final list of words are used for the creation of the dictionary and the corpus, fundamental elements for the topic model. The dictionary defines the mapping from word to ids and it is used to define the vocabulary size. Whereas the corpus is a vector of the documents that will be used for training. In particular, it is a matrix of shape (number of documents, numbers of terms).[6]

For the *description* variable, the analysis is more complex for numerous reasons: the amount of textual data for each observations is utterly elevate; the personal name and surname of borrowers are present in each observations; city names are diffused; the name and surname of the author of the description are included in each observation; adverbs are repeated multiple times for the same observations and their meaning is not useful to detect the main topics; the same occurs also for some adjectives or verbs too much general. As a result, a cut-off cannot be inserted to reduce the number of too frequent or too rare words. Therefore another analysis has been conducted. To eliminate the author’s information in each observation, the position of the word “translated” is crucial. In fact, after this term, only the writer’s description is present. The next step is creating two columns of the *description* variable: one with the content before the word “translated”, one after that. The second one is deleted to remove all the author’s characteristics.

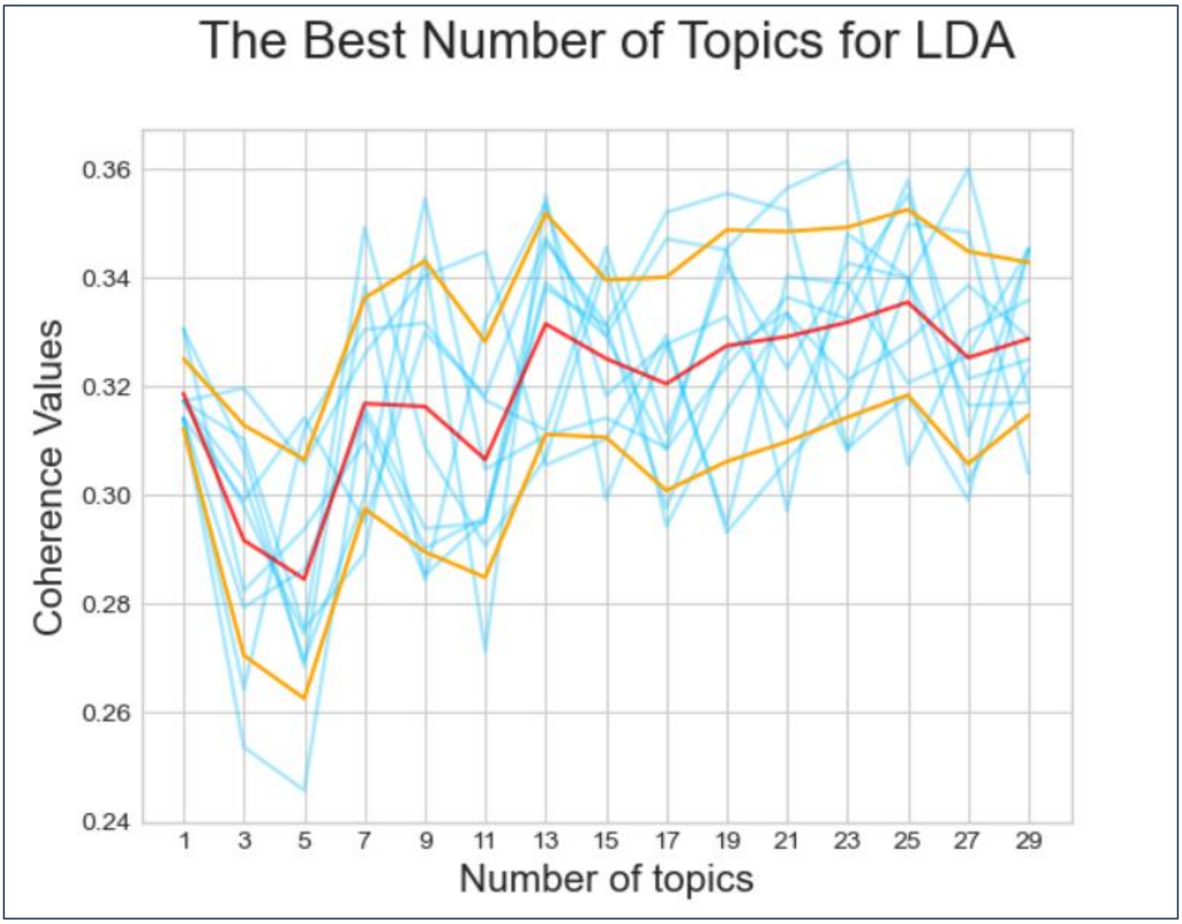
[6] See Gensim documentation, version 4.3.0

The following step comprehends the elimination of the borrower's surname and name in the *description* variable. In the dataset, the *loanName* feature displays these information of the borrower. Consequently, a for loop is defined to delete all the value present in l*oanName* from *description's* list of terms. Another modification regards the removal of the words with a length less than 3 combined with the elimination of the most common time adverbs or adverbs of manner.

Once all the previous steps have been concluded, the next phase regards the definition of the right number of themes to be inserted in topic modeling. In fact, the main target is to define k number of topics. Granularity expresses the level of detail of the model and depends on the k value. If the granularity is too small, it would be more challenging the topic identification. If it is too high, topics would be similar one to the other and the topic modeling would lose its effectiveness Jacobi et al., 2015).

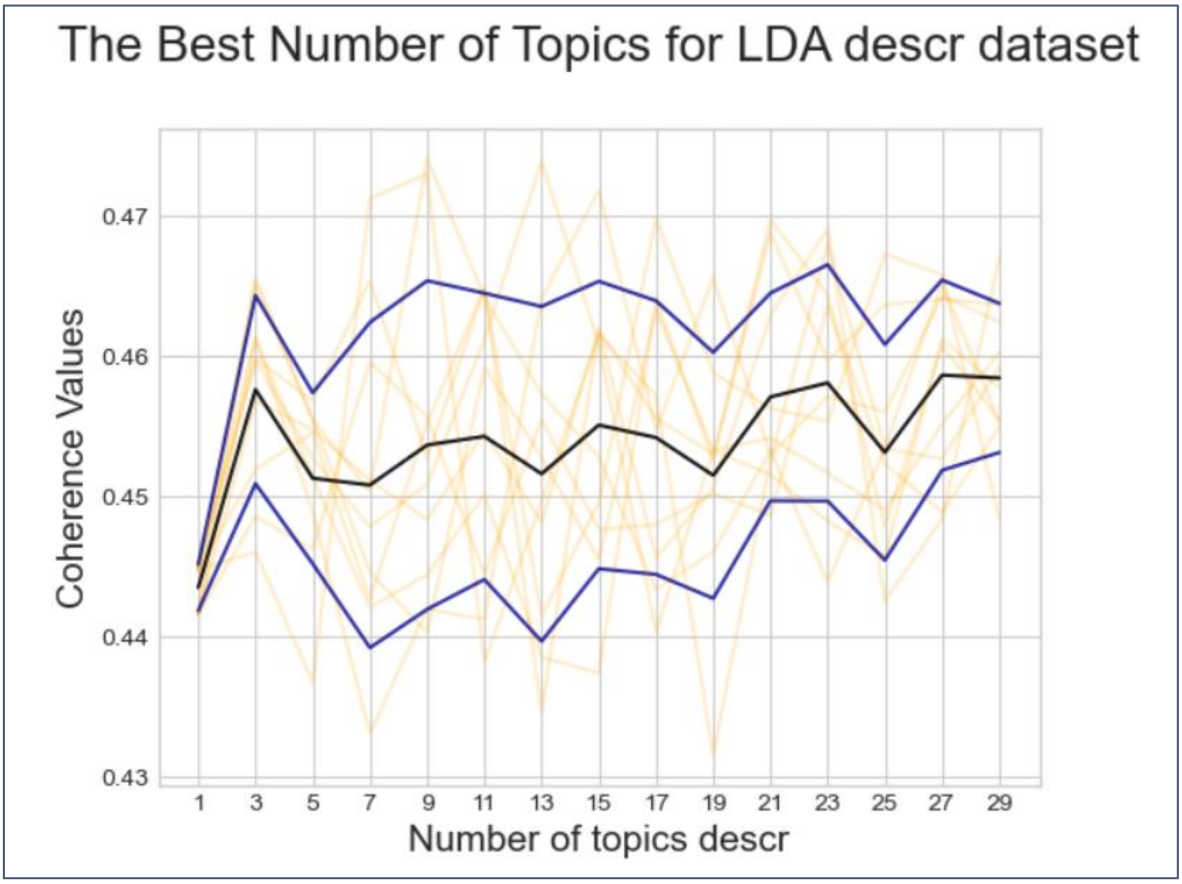
Numerous method are identified and combined to determine the optimal value of k. Firstly, one of the most popular metrics is called "coherence" and evaluates the coherence of semantic information in the model. It is expressed through the conditional probabilities defined for the association of words for which the document has the first n words of a specific topic simultaneously. This means that higher is the probability that the words with the highest topic ranking occur in a text together, higher is the classification effect of the model (Gan, Qi, 2021). In this analysis, the coherence is estimated in 10 trials and the results are saved in a dataframe and saved with a csv file. A line plot is also provided to show the curve derived from the calculation of the mean of coherence values for each number of topic. In addition, the other two lines represented in the Figure 3.4.3 represent the confidence interval resulting from the mean of coherence values minus the standard deviation (the lowest line) and the mean of coherence values plus the standard deviation. To choose the optimal coherence score, a graphical method is applied: the elbow method. When the coherence curve creates an angle at the elbow, the corresponding value in that point corresponds to the best number of topics to apply in the topic model (Maithya et al., 2022; Gurdiel et al., 2021). The Figure 3.4.3 illustrates that the optimal number of topics for the *use* variable can be: 7, 9, 13 and 25.

Figure 3.4.3 – The Coherence Plot and The Elbow Method for The *use* Variable



On the contrary, the optimal number of topics for the description variable can be: 3, 11, 15, 17, 23.

Figure 3.4.4 - The Coherence Plot and The Elbow Method for The *description* Variable



To identity the optimal number of topics between the ones selected through the coherence score and the elbow method, a sensitivity analysis is applied. Sensitivity analysis highlights how the uncertainty in the result of a model can be caused by sources of uncertainty in the model parameters. It is also combined with the uncertainty analysis regarding uncertainty in model prediction without individualizing which hypothesis have impacted more the final outcome (Saltelli et al. 2019). In this study, the sensitivity analysis is implemented to compare the outcomes of the different topic modelling. All parameters of the topic modelling remain the same, except the number of topics that are identified through the coherence and the elbow method.

The model selected for the following analysis is the Latent Dirichlet Allocation (LDAvis), an unsupervised technique that identifies the diverse themes emerged from a document based on the patterns of (co)-occurrence of words in the text (Jacobi et al., 2015). It is based on the concept of consistency. Every topic presents an internal consistency. It refers to the terms that occur together in the specific document and/or are related to the same topic (Chang et al., 2009). Another fundamental parameter of the LDA model is lambda and indicates relevance of a word in a specific topic. It facilitates the ranking of terms based on the level of topic interpretation. Sievert and Shirley have conducted a study to determine the optimal value of lambda. It has been discovered that if lambda equals to 0.6, it generates an estimated 70% probability of correct words related to a topic. On the contrary, if the values are closer to 0 or 1, the estimated proportions of correct identifications are closer to 53% and 63% respectively (Sievert, Shirley, 2014).

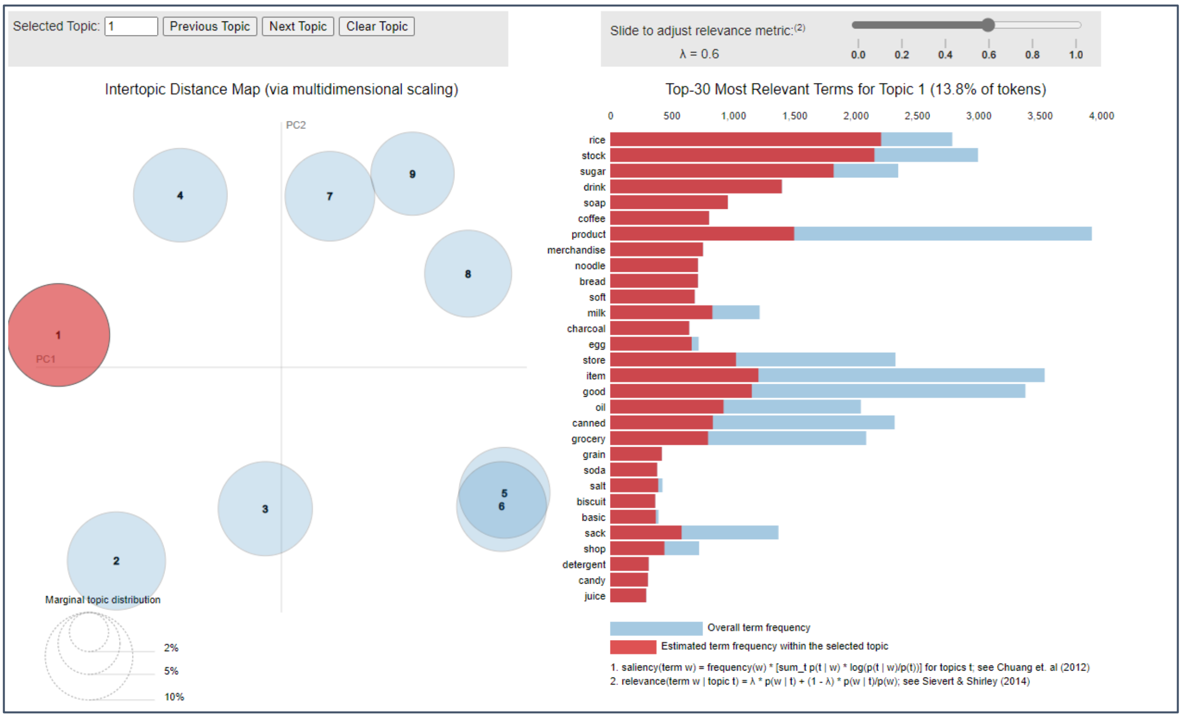
Nonetheless, the mathematical goodness of fit measures applied for the definition of the optimal number of topics can achieve a result not clearly interpretable from the human perspective. It happens more frequently with an high number of topics because the algorithm could identify some terms not semantically relevant for human (Chang et al. 2009).

In addition, another issue regards the randomness emerged in the LDA model, even though the random_state parameter was inserted. For example, some words related to the third topic of the *use* variable were exchanged with the terms identified in the fourth topic. Therefore, at the second run, the words remains the same, but the third topic has become the fourth, while the fourth has become the third. To guarantee the reproducibility of results, once all the parameters are set, the different LDA outputs are saved with the joblib library.

From the sensitivity analysis in combination with the results of the coherence and the elbow method, the optimal number of topics for the *use* variable is 9, as can be noticed in the Figure 3.4.5.[7] After having analyzed the words present in each bubble, a title to each topic has been assigned: topic n.1 – purchase, topic n.2 – farming, topic n.3 – investment, topic n.4 – trading, topic n.5 – family_expenses, topic n.6 – family_needs, topic n.7 – business, topic n.8 - improvement, topic n.9 – personal_activity.

[7] To see the terms identified in the other topics, see Appendix C.

Figure 3.4.5 – The Optimal Number of Topics of the *use* Variable



On the other hand, the definition of topics for the *description* variable presents some difficulties for the interpretability. Probably, it is caused by the similar frequency of the words and the amount of textual data of the *description* column. For example, the original data of the *use* and *description* features related to the first observation of the dataset are:

use: "Purchasing sheep to expand his business"

description: "For the past 2 years, Bobomurod has worked in agriculture, raising animals. He raises sheep which he cares for greatly since thanks to them, he can cover his household expenses. Bobomurod showed us his animals with great happiness as his sheep were about to be vaccinated against the various sickness that affect that type of livestock. <P>_x000D_"

Therefore, the solution could be comparing the results obtained through the application of a specific model that will crucial also in the last part of the analyses. Logistic regression model, also called logit model, is a typology of regression analysis used to find the probability of a certain event occurring. It is used to model dichotomous outcome variables and, in particular, the log odds of the dependent variable is modeled as a linear combination of the predictor variables.

Considering the aforementioned research questions, the dependent variable is *status* (it indicates if the loan has been funded or expired), while the independent variables are the topics of the *use* and *description* variables. The number of topics of *use* is fixed and equals to 9. Whereas, to define the number of topics of *description*, the outcomes obtained from the coherence and elbow methos are selected (k = 3, 11, 15, 17, 23).

Therefore, a dataframe is created with only a subset of variables: *status*, whose categorical values are transformed into integers for the logit model (1, if status = funded, 0, if status = expired) and the topic contribution based on the number of topics of *use* and *description*. This means that the topic contribution has been calculated five different times as the number of topics for *description* variable. Randomness influences also the results of the different models for the aforementioned motivations and, to guarantee the reproducibility of results, the diverse outcomes are saved through the joblib library. Even though the p-values changed in the different runs, the evaluation of the model has achieved the same outcome: the optimal number of topics for the *description* variable is 15. In this phase, two elements have been considered: the R squared and the p-values. The first "(R^2 or the coefficient of determination) is a statistical measure in a regression model that determines the proportion of variance in the dependent variable that can be explained by the independent variable" (CFI Team, 2022). While, "the p-value (probability value) is a probability measure of finding the observed, or more extreme, results, when the null hypothesis of a given statistical test is true. The p-value is a primary value used to quantify the statistical significance of the results of a hypothesis test. The main interpretation of the p-value is whether there's enough evidence to reject the null hypothesis. If the p-value is reasonably low (less than the level of significance), we can state that there is enough evidence to reject the null hypothesis. Otherwise, we should not reject the null hypothesis. The conclusions about the hypothesis test are drawn when the p-value of a test is compared against the level of significance, which plays the role of a benchmark. The most typical levels of significance are 0.10, 0.05, and 0.01. The level of significance of 0.05 is considered conventional and the most commonly used." (CFI Team, 2022). In this analysis both indicators are considered because the R-squared do not consistently vary in the different models.

As it is evident in the Figure 3.4.6, the R- squared is 4%, but four topic variables of *description* have a p-value around 0.05 and five of them shows a p-value less than 0.05.

Figure 3.4.6 – Logistic Regression Model with *status*, 9 topics of *use*, 15 topic of *description*

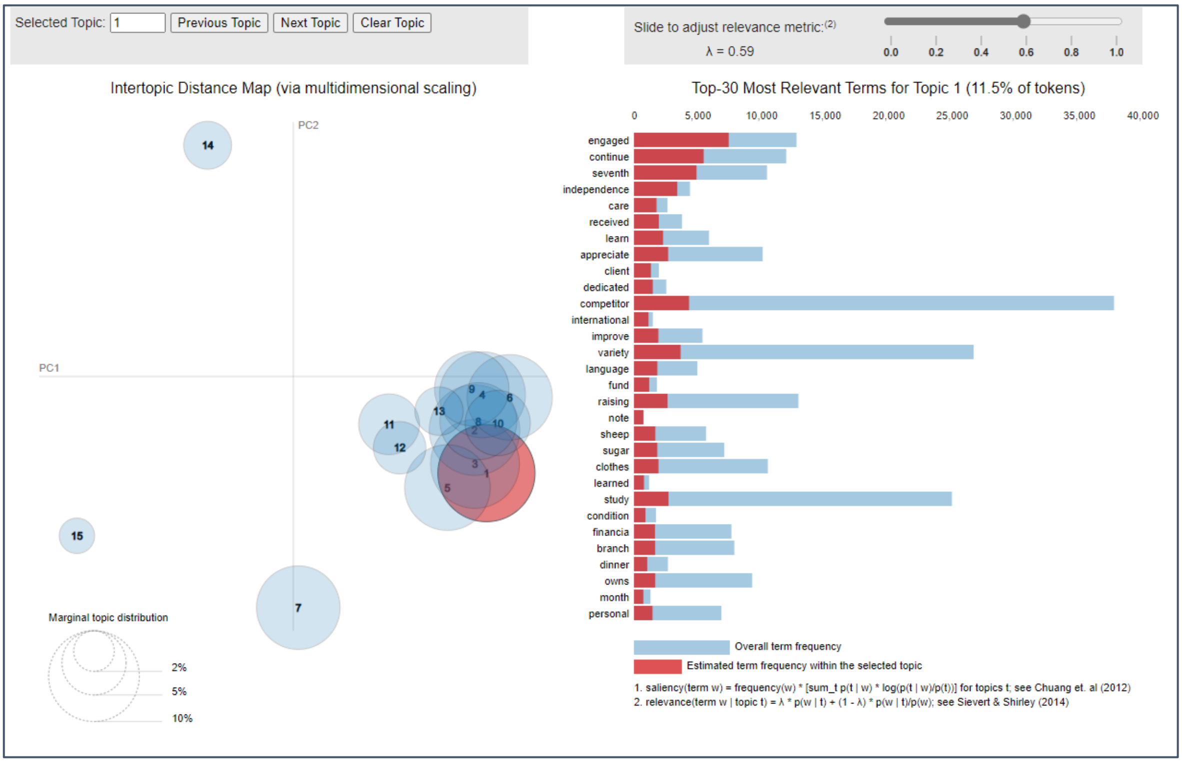
Logit Regression Results

Dep. Variable:	status	**No. Observations:**	36000
Model:	Logit	**Df Residuals:**	35975
Method:	MLE	**Df Model:**	24
Date:	Sun, 22 Jan 2023	**Pseudo R-squ.:**	0.04083
Time:	20:13:51	**Log-Likelihood:**	-5959.7
converged:	True	**LL-Null:**	-6213.4
Covariance Type:	nonrobust	**LLR p-value:**	4.828e-92

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
const	0.8385	0.910	0.921	0.357	-0.945	2.622
topic_contribution_use_0	1.3953	0.868	1.608	0.108	-0.306	3.097
topic_contribution_use_1	1.7014	0.873	1.949	0.051	-0.010	3.413
topic_contribution_use_2	1.7890	0.879	2.036	0.042	0.067	3.511
topic_contribution_use_3	2.1727	0.877	2.479	0.013	0.455	3.891
topic_contribution_use_4	1.7888	0.882	2.028	0.043	0.060	3.517
topic_contribution_use_5	2.1264	0.878	2.422	0.015	0.406	3.847
topic_contribution_use_6	2.1315	0.874	2.438	0.015	0.418	3.845
topic_contribution_use_7	1.4489	0.873	1.659	0.097	-0.263	3.161
topic_contribution_use_8	1.0368	0.876	1.183	0.237	-0.681	2.754
topic_contribution_descr_0	0.6062	0.330	1.837	0.066	-0.041	1.253
topic_contribution_descr_1	2.4942	0.629	3.966	0.000	1.262	3.727
topic_contribution_descr_2	0.5913	0.322	1.833	0.067	-0.041	1.223
topic_contribution_descr_3	2.2876	0.345	6.637	0.000	1.612	2.963
topic_contribution_descr_4	1.2521	0.400	3.129	0.002	0.468	2.036
topic_contribution_descr_5	0.3183	0.335	0.951	0.342	-0.338	0.974
topic_contribution_descr_6	1.0946	0.343	3.189	0.001	0.422	1.767
topic_contribution_descr_7	5.1938	0.837	6.206	0.000	3.553	6.834
topic_contribution_descr_8	-0.0323	0.320	-0.101	0.920	-0.659	0.595
topic_contribution_descr_9	-0.1268	0.344	-0.369	0.712	-0.801	0.547
topic_contribution_descr_10	0.6859	0.440	1.558	0.119	-0.177	1.549
topic_contribution_descr_11	0.6029	0.332	1.818	0.069	-0.047	1.253
topic_contribution_descr_12	-0.2597	0.323	-0.804	0.421	-0.893	0.374
topic_contribution_descr_13	0.1311	0.333	0.393	0.694	-0.522	0.784
topic_contribution_descr_14	0.7101	0.408	1.742	0.082	-0.089	1.509

Therefore, 15 number of topics for the *description* variable are selected for the LDA model. To improve the interpretability and better understand the borrower's background, the words related to the same bubble are also manually individualized in the original textual data. The observations with the highest topic contribution are considered more important than the remaining one. Consequently, the titles assigned to each topic of the *description* variables are: topic n.1 – face_difficulties, topic n.2 – entrepreneurship, topic n. 3 – causes, topic n. 4 – asking_for_loan, topic n.5 – retail, topic n. 6 – progress, topic n. 7 – borrowers_situation, topic n. 8 – women_reasons, topic n. 9 – financial_resources, topic n. 10 – job, topic n. 11 – procurement, topic n. 12 – growth, topic n. 13 – local_activity, topic n. 14 – poverty, topic n. 15 – loyal_borrowers (Figure 3.4.7).

Figure 3.4.7 – The Optimal Number of Topics of the *description* Variable

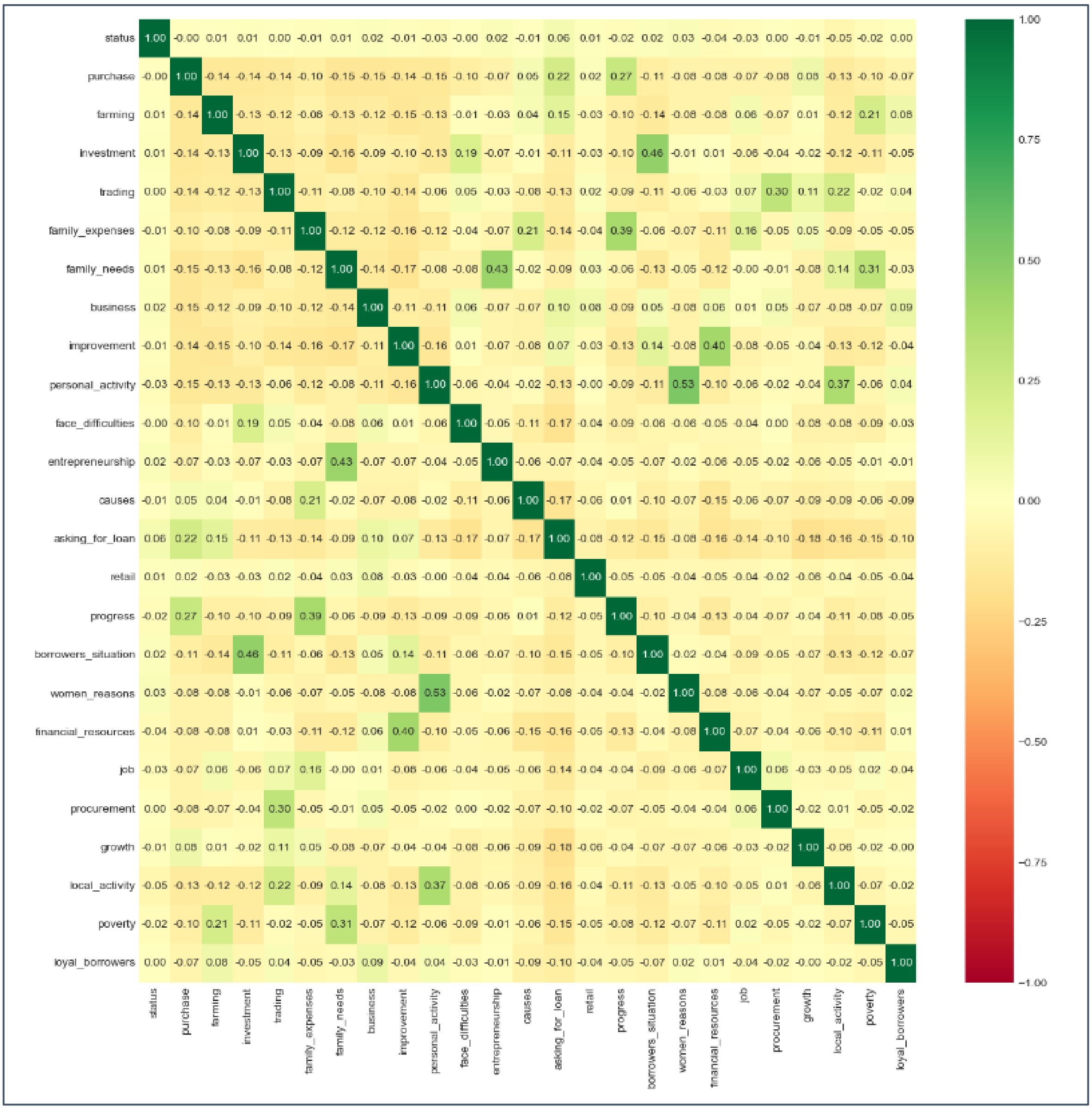


The last step of this phase comprehends the creation of the dataset with *id* of the loan transaction, the *status* variable, the 9 topics of the *use* variable and the 15 topics of the *description* variable. To answer to the first research question, the correlation matrix should be defined. It indicates the correlation between all pairs of variables of the dataframe and presents a set of ones in the diagonal because each variable is perfectly correlated with itself. It is also a symmetric matrix because the resulted values in the upper right triangle are the same in the lower left triangle.

In the Figure 3.4.8 it is evident that some variables are highly correlated with values equal or more than 0.20. In particular, not only the relationships between specific topics emerged in the literature analysis are confirmed in the correlation matrix, but also other aspects arise:

- women usually request a loan to launch their personal activity (*women_reasons – personal_activity*);
- borrowers, who want to invest in small businesses, present an accurate description of their situation to attract the investors' attention (*borrowers_*situation - *investment*);
- if the need of financial resources is utterly highlighted in the loan application, they will be generally implemented to improve the quality of the goods or services provided by the borrowers' local organizations (*financial_resources – improvement; face-difficulties – investment*);
- starting a business is associated to collect new capital to satisfy family's needs (*entrepreneurship – family; local_activity – personal_activity*);
- family expenses are usually finalized to the achievement of better living conditions, for example building a sanitary toilet in the house, paying tuition for university or school, etc. (*family _expenses – progress*; *causes – family_expenses; job – family_expenses*);
- several goods should be purchased to enhance the borrowers' activity in order to expand the offer and resell them (*progress – purchase; procurement – trading; local_activity – trading; asking for loan – purchase*);
- numerous family needs are caused by the consistent poverty rate of the developing nations (*poverty – family_needs*);
- the majority of population who lives in poor conditions is usually dedicated to the farming sector (poverty – farming).

Figure 3.4.8 – Correlation Matrix



3.5. Data Modeling

To respond to the second research question, the last part of the analysis is divided in two segments: the first comprehends the entire dataset with 36000 observations, whereas the second is based on a smaller one with 18000 rows.

The outcome variable of the following models, *status,* comprehends the category of funded loans with 34.507 observations and the one of expired loans with 1.493 rows. This means that the two classes are unbalanced. For this reason, the sensitivity analyses is applied to evaluate the effects on the final models, when the number of observation in the outcome variable are modified. In addition, the dataset created in the topic modeling part is merged with the original dataset to involved and examine all available data. Considering that the RAM of Jupyter Notebook is limited, the variables are distinguished into three groups: *SectorName* and *Region*, partner variables and the residual. The variables on the number of topics of *use* and *description*, plus *status*, are always comprehended.

Numerous operations have been applied in all sections. Firstly, the substitution of the *status* values with 1 and 0 values, 1 is the loan is funded, 0 in the opposite case because the logistic regression accepts only discrete and categorical variables as outcome. The "get dummies" function is implemented for all categorical variables to convert them into dummy or indicator variables. This means that if a categorical variable has three classes, three variables will be created, one for each class, and they will present the value 0 or 1. The function iterates over the object passed and checks whether the element at a certain index matches the column heading. If it does, the value of the variable will be 1, otherwise 0.[8]

On the other hand, numerical variables are normalized by calculating the ratio between the difference of the actual observations minus the minimum value of the column, as numerator, and the difference of the maximum value of the feature minus its minimum, as denominator.

The third step regards the multicollinearity problem and the calculation of the variance inflation factor. In fact, one of the assumptions in binary logistic regression to achieve a valid result is the absence of multicollinearity. This means that explanatory variables should not be highly correlated with each other. To identify which variables cause this issue, the variance inflation factor (VIF) can be performed. It illustrates how much variance of the coefficient estimate is influenced by multicollinearity. If the VIF crosses the value of 10, it indicates multicollinearity. In logistic regression, also values more than 2.5 could cause a problem. One of the most common method to reduce multicollinearity is removing the variable with the highest VIF (Midi et al., 2013; Senaviratna et al. 2019).

[8] See pandas documentation, version 1.5.3.

Therefore, the following approach has been adopted: calculate the VIF and consider a distinction between the variables created through topic modeling and the ones generate with the get dummies function. In the first group, delete the variables associated with the highest topic contribution, one for *use* and one for *description*. In the second group, analyse the observations in each class of the original categorical variable and select the ones with the highest and lowest number of observations. Consequently, remove the variables created with the get dummies that corresponds to the identified classes. Run the logistic regression with the remaining variables and identify if other variables determine the multicollinearity. If it is true, delete the one with the highest VIF and run again the model until the VIF values are small enough to not affect the model outcome (see Appendix D).

The above steps are applied in three datasets with 36.000 observations and also in the three datasets with 18.000 observations (the observations in the smaller dataset are randomly selected to reduce the size of the category with the highest number of data).

In the dataset related to sectors and regions, the same variables have been deleted for the VIF calculation: *Sector_Agriculture, Region_Asia, retail* and *investment*.

Considering the full dataset and the p-values, the Figure 5.1 shows that the projects launched on Kiva's platform have more probabilities to be funded if the loan description comprehends an explanation of the borrower's situation or the principal reasons of a woman for which the loan is requested. In addition, borrowers who comply with the repayment terms and have an higher reputability are more preferred by lenders. In particular, if the borrowers applies for the loan to start his own business, he has more likelihood of receiving the financial resources. The sectors more fascinating for bakers are arts, education and manufacturing (*entrepreneurship, borrowers_situation, women_reasons, loyal_borrowers, asking_for_loan, SectorName_Arts, SectorName_Education, SectorName_Manufactoring*).

On the other hand, the variables that negatively influence the success of a projects are *purchase, personal_activity, causes, progress, job, poverty, local_activity, SectorName_Clothing, SectorName_Housing, SectorName_Personal Use, SectorName_Retail, SectorName_Transportation, Region_Africa, Region_Central America,* *Region_Eastern Europe, Region_Middle Est, Region_North America, Region_Oceania, Region_South America* (Figure 3.5.1 and 3.5.2).

Figure 3.5.1 – First part of Logit Model of Sector, Region and Topic Variables (full dataset)

Logit Regression Results

Dep. Variable:	status	**No. Observations:**	36000
Model:	Logit	**Df Residuals:**	35958
Method:	MLE	**Df Model:**	41
Date:	Sat, 28 Jan 2023	**Pseudo R-squ.:**	0.07153
Time:	19:15:38	**Log-Likelihood:**	-5769.0
converged:	True	**LL-Null:**	-6213.4
Covariance Type:	nonrobust	**LLR p-value:**	7.962e-160

	coef	std err	z	P>|z|	[0.025	0.975]
const	4.1356	0.271	15.271	0.000	3.605	4.666
purchase	-0.4866	0.218	-2.234	0.025	-0.913	-0.060
farming	0.0289	0.231	0.125	0.901	-0.424	0.482
trading	0.1676	0.232	0.723	0.470	-0.287	0.622
family_expenses	-0.1550	0.231	-0.672	0.502	-0.607	0.297
family_needs	-0.2152	0.250	-0.860	0.390	-0.706	0.275
business	0.2039	0.217	0.938	0.348	-0.222	0.630
improvement	-0.2545	0.196	-1.301	0.193	-0.638	0.129
personal_activity	-0.7775	0.241	-3.232	0.001	-1.249	-0.306
face_difficulties	0.3391	0.270	1.257	0.209	-0.190	0.868
entrepreneurship	2.5118	0.666	3.774	0.000	1.207	3.816
causes	-0.6243	0.245	-2.546	0.011	-1.105	-0.144
progress	-0.7444	0.273	-2.725	0.006	-1.280	-0.209
borrowers_situation	0.4578	0.273	1.678	0.093	-0.077	0.993
women_reasons	4.3928	0.861	5.102	0.000	2.705	6.080
financial_resources	-0.1037	0.244	-0.425	0.671	-0.581	0.374
job	-0.9739	0.278	-3.508	0.000	-1.518	-0.430
procurement	-0.5172	0.394	-1.313	0.189	-1.289	0.255
growth	0.4029	0.275	1.465	0.143	-0.136	0.942
poverty	-1.0952	0.272	-4.022	0.000	-1.629	-0.562
loyal_borrowers	0.8532	0.384	2.219	0.026	0.100	1.607

This means that lenders consider also the level of uncertainty related to each loan. It has been increased in the last year as it is shown in the Figure 3.3.5. In fact, the variables that negatively affect the project's success are associated with borrowers who live in

Figure 3.5.2 – Second part of Logit Model of Sector, Region and Topic Variables (full dataset)

asking_for_loan	1.1891	0.267	4.461	0.000	0.667	1.711
local_activity	-0.4643	0.257	-1.810	0.070	-0.967	0.039
SectorName_Arts	2.3189	0.718	3.232	0.001	0.913	3.725
SectorName_Clothing	-0.7954	0.147	-5.400	0.000	-1.084	-0.507
SectorName_Construction	0.4676	0.331	1.411	0.158	-0.182	1.117
SectorName_Education	1.8509	0.311	5.944	0.000	1.241	2.461
SectorName_Food	-0.1915	0.121	-1.588	0.112	-0.428	0.045
SectorName_Health	-0.2884	0.224	-1.288	0.198	-0.727	0.150
SectorName_Housing	-0.6369	0.171	-3.723	0.000	-0.972	-0.302
SectorName_Manufacturing	2.6209	1.009	2.598	0.009	0.644	4.598
SectorName_Personal Use	-0.3587	0.208	-1.726	0.084	-0.766	0.049
SectorName_Retail	-0.7549	0.115	-6.572	0.000	-0.980	-0.530
SectorName_Services	-0.2125	0.151	-1.410	0.158	-0.508	0.083
SectorName_Transportation	-0.4315	0.199	-2.167	0.030	-0.822	-0.041
Region_Africa	-0.4080	0.099	-4.131	0.000	-0.602	-0.214
Region_Central America	-1.0981	0.107	-10.276	0.000	-1.308	-0.889
Region_Eastern Europe	-1.1152	0.146	-7.614	0.000	-1.402	-0.828
Region_Middle East	-1.1324	0.125	-9.063	0.000	-1.377	-0.888
Region_North America	-0.6311	0.202	-3.127	0.002	-1.027	-0.236
Region_Oceania	-1.0388	0.180	-5.784	0.000	-1.391	-0.687
Region_South America	-0.8512	0.110	-7.756	0.000	-1.066	-0.636

unprivileged conditions. The same result is also confirmed in the logit model applied on the under sample (Figure 3.5.3, 3.5.4). The result of the full and small dataset are also influenced by the removal of the four variables; this means that *Sector_Agriculture, Region_Asia, retail* and *investment* generate a consistent effect on the initiatives published on the Kiva platform.

Figure 3.5.3 – First part of Logit Model of Sector, Region and Topic Variables (small dataset)

Logit Regression Results

Dep. Variable:	status	**No. Observations:**	18000
Model:	Logit	**Df Residuals:**	17958
Method:	MLE	**Df Model:**	41
Date:	Sat, 28 Jan 2023	**Pseudo R-squ.:**	0.08352
Time:	19:38:41	**Log-Likelihood:**	-4716.4
converged:	True	**LL-Null:**	-5146.2
Covariance Type:	nonrobust	**LLR p-value:**	9.438e-154

	coef	std err	z	P>|z|	[0.025	0.975]
const	3.4170	0.281	12.145	0.000	2.866	3.968
purchase	-0.4654	0.225	-2.068	0.039	-0.906	-0.024
farming	0.0394	0.239	0.165	0.869	-0.430	0.509
trading	0.1109	0.240	0.463	0.643	-0.359	0.581
family_expenses	-0.1489	0.237	-0.628	0.530	-0.614	0.316
family_needs	-0.2594	0.259	-1.002	0.316	-0.767	0.248
business	0.1866	0.223	0.837	0.403	-0.251	0.624
improvement	-0.2370	0.201	-1.182	0.237	-0.630	0.156
personal_activity	-0.7050	0.251	-2.812	0.005	-1.196	-0.214
face_difficulties	0.3137	0.279	1.123	0.261	-0.234	0.861
entrepreneurship	2.6464	0.676	3.915	0.000	1.322	3.971
causes	-0.6198	0.254	-2.437	0.015	-1.118	-0.121
asking_for_loan	1.1812	0.274	4.306	0.000	0.644	1.719
progress	-0.7968	0.285	-2.800	0.005	-1.354	-0.239
borrowers_situation	0.4373	0.282	1.549	0.121	-0.116	0.991
women_reasons	4.4595	0.882	5.054	0.000	2.730	6.189
financial_resources	-0.0656	0.255	-0.258	0.797	-0.565	0.434
job	-1.0251	0.289	-3.549	0.000	-1.591	-0.459
procurement	-0.4710	0.402	-1.171	0.242	-1.260	0.318
growth	0.4322	0.287	1.507	0.132	-0.130	0.994
local_activity	-0.5036	0.268	-1.876	0.061	-1.030	0.022
poverty	-1.0964	0.286	-3.832	0.000	-1.657	-0.536

Figure 3.5.4 – Second part of Logit Model of Sector, Region and Topic Variables (small dataset)

loyal_borrowers	0.9930	0.397	2.503	0.012	0.215	1.771
SectorName_Arts	2.2408	0.720	3.113	0.002	0.830	3.651
SectorName_Clothing	-0.8067	0.152	-5.296	0.000	-1.105	-0.508
SectorName_Construction	0.4134	0.339	1.219	0.223	-0.252	1.078
SectorName_Education	1.8631	0.318	5.867	0.000	1.241	2.485
SectorName_Food	-0.1961	0.124	-1.576	0.115	-0.440	0.048
SectorName_Health	-0.3292	0.235	-1.403	0.161	-0.789	0.131
SectorName_Housing	-0.6688	0.179	-3.733	0.000	-1.020	-0.318
SectorName_Manufacturing	2.6528	1.011	2.624	0.009	0.671	4.634
SectorName_Personal Use	-0.4207	0.215	-1.959	0.050	-0.842	0.000
SectorName_Retail	-0.7892	0.119	-6.655	0.000	-1.022	-0.557
SectorName_Services	-0.2126	0.156	-1.360	0.174	-0.519	0.094
SectorName_Transportation	-0.4523	0.208	-2.174	0.030	-0.860	-0.045
Region_Africa	-0.4044	0.101	-3.996	0.000	-0.603	-0.206
Region_Central America	-1.1355	0.111	-10.236	0.000	-1.353	-0.918
Region_Eastern Europe	-1.0990	0.154	-7.125	0.000	-1.401	-0.797
Region_Middle East	-1.1210	0.131	-8.584	0.000	-1.377	-0.865
Region_North America	-0.5954	0.209	-2.850	0.004	-1.005	-0.186
Region_Oceania	-1.0942	0.187	-5.840	0.000	-1.461	-0.727
Region_South America	-0.8809	0.113	-7.773	0.000	-1.103	-0.659

Regarding the datasets on partners, the variables deleted for high VFI are *investment, retail, PStatus_active, PStatementType_net_billing, PAvgBorrowerCostType_PY, PloansPosted, PloansAtRiskRate, PRiskRating.*

To augment the probabilities of the success of a project, the model applied on both datasets displays that the campaign should definitively include the following variables: *trading, family_needs, asking_for_loan, borrowers_situation, women_reasons, PtotalAmountRaised, PavgBorrowerCost, PStatus_closed, PStatementType_dual_statement, PavgBorrowerCostType_APR.* As a result, the borrower profile is still fundamental for the selection of a project. When the motivation to ask for a loan are explained in combination with the family conditions, lenders will feel more involved and will finance the initiative. Moreover, as in the sector and region dataset, the attention dedicated to women prevails in comparison to the other factors (highest coefficient). In addition, trading activity is most popular to buy and sell goods or services. At the same time, the role played by partners is fundamental in the project selection and the principal aspects that lenders consider are the amount raised (to evaluate the size of the investment), the average borrower cost and the dual statement type. On the contrary, the outcome of the dataset regarding sector and region is confirmed also in this case. Backers prefer to invest in situations associated with a low level of precautionsness. This relates not only the borrowers living conditions, but also the specific characteristics of the partners such as the default rate, the loss derived from the currency exchange, the status etc. These conditions are represented by the following variables: *improvement,*

Figure 3.5.5 - First part of Logit Model of Topic and Partners' Variables (full dataset)

Logit Regression Results

Dep. Variable:	status	**No. Observations:**	33875
Model:	Logit	**Df Residuals:**	33839
Method:	MLE	**Df Model:**	35
Date:	Sat, 28 Jan 2023	**Pseudo R-squ.:**	0.05669
Time:	21:06:17	**Log-Likelihood:**	-5610.0
converged:	True	**LL-Null:**	-5947.2
Covariance Type:	nonrobust	**LLR p-value:**	2.374e-119

	coef	std err	z	P>|z|	[0.025	0.975]
const	3.4773	0.633	5.497	0.000	2.238	4.717
purchase	-0.3359	0.211	-1.591	0.112	-0.750	0.078
farming	0.0496	0.223	0.222	0.824	-0.388	0.487
trading	0.4967	0.227	2.192	0.028	0.053	0.941
family_expenses	0.1951	0.229	0.853	0.393	-0.253	0.643
family_needs	0.6396	0.236	2.708	0.007	0.177	1.102
business	0.2906	0.222	1.309	0.191	-0.145	0.726
improvement	-0.3580	0.192	-1.863	0.063	-0.735	0.019
personal_activity	-0.6946	0.227	-3.066	0.002	-1.139	-0.251
face_difficulties	0.0820	0.256	0.320	0.749	-0.420	0.585
entrepreneurship	0.8263	0.651	1.269	0.204	-0.449	2.102
causes	-0.0604	0.243	-0.248	0.804	-0.538	0.417
asking_for_loan	0.6182	0.320	1.929	0.054	-0.010	1.246
progress	-0.6764	0.280	-2.417	0.016	-1.225	-0.128
borrowers_situation	0.6156	0.275	2.236	0.025	0.076	1.155
women_reasons	4.3071	0.836	5.150	0.000	2.668	5.946
financial_resources	-0.4916	0.244	-2.014	0.044	-0.970	-0.013
job	-0.2638	0.281	-0.938	0.348	-0.815	0.287

personal_activity, progress, financial_resources, local_activity, PavgLoanSizePercentPerCapitaIncome, PdefaultRate, PcurrencyExchangeLossRate, PchargesFeesInterest, PStatus_inactive (Figure 3.5.5, 3.5.6, 3.5.7, 3.5.8).

In addition, the results are also influenced by the deleted variables for the multicollinearity issues. Therefore, they should be also considered as fundamental for the impressive impact on the project's outcome.

Figure 3.5.6 - Second part of Logit Model of Topic and Partners' Variables (full dataset)

procurement	-0.1816	0.406	-0.448	0.654	-0.977	0.613
growth	0.1362	0.264	0.516	0.606	-0.381	0.653
local_activity	-0.6378	0.247	-2.577	0.010	-1.123	-0.153
poverty	-0.3200	0.263	-1.217	0.223	-0.835	0.195
loyal_borrowers	0.0714	0.365	0.195	0.845	-0.645	0.787
PtotalAmountRaised	1.0534	0.212	4.963	0.000	0.637	1.469
PavgBorrowerCost	0.8973	0.203	4.423	0.000	0.500	1.295
PavgProfitability	0.6400	0.441	1.450	0.147	-0.225	1.505
PavgLoanSizePercentPerCapitaIncome	-0.3031	0.157	-1.926	0.054	-0.611	0.005
PdefaultRate	-2.3481	0.518	-4.533	0.000	-3.363	-1.333
PcurrencyExchangeLossRate	-3.1541	0.678	-4.650	0.000	-4.484	-1.825
PchargesFeesInterest	-1.2628	0.481	-2.627	0.009	-2.205	-0.321
ParrearsRate	-0.1665	0.119	-1.394	0.163	-0.401	0.068
PStatus_closed	0.3514	0.077	4.551	0.000	0.200	0.503
PStatus_inactive	-0.7696	0.135	-5.699	0.000	-1.034	-0.505
PStatus_paused	0.1030	0.169	0.610	0.542	-0.228	0.434
PStatementType_dual_statement	1.6681	0.602	2.772	0.006	0.489	2.848
PavgBorrowerCostType_APR	0.6712	0.150	4.481	0.000	0.378	0.965

Figure 3.5.7 - First part of Logit Model of Topic and Partners' Variables (small dataset)

Logit Regression Results

Dep. Variable:	status	**No. Observations:**	18000
Model:	Logit	**Df Residuals:**	17964
Method:	MLE	**Df Model:**	35
Date:	Sat, 28 Jan 2023	**Pseudo R-squ.:**	0.06369
Time:	21:16:31	**Log-Likelihood:**	-4691.4
converged:	True	**LL-Null:**	-5010.5
Covariance Type:	nonrobust	**LLR p-value:**	6.366e-112

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
const	2.9107	0.646	4.505	0.000	1.644	4.177
purchase	-0.3391	0.217	-1.561	0.119	-0.765	0.087
farming	0.0043	0.229	0.019	0.985	-0.445	0.453
trading	0.4218	0.233	1.811	0.070	-0.035	0.878
family_expenses	0.1480	0.235	0.630	0.529	-0.312	0.608
family_needs	0.5738	0.241	2.379	0.017	0.101	1.047
business	0.2478	0.227	1.090	0.276	-0.198	0.694
improvement	-0.3617	0.198	-1.832	0.067	-0.749	0.025
personal_activity	-0.7786	0.235	-3.308	0.001	-1.240	-0.317
face_difficulties	0.0291	0.267	0.109	0.913	-0.494	0.552
entrepreneurship	0.7885	0.647	1.219	0.223	-0.479	2.056
causes	-0.1077	0.253	-0.426	0.670	-0.603	0.387
asking_for_loan	0.5847	0.330	1.773	0.076	-0.062	1.231
progress	-0.6763	0.289	-2.338	0.019	-1.243	-0.109
borrowers_situation	0.5390	0.285	1.892	0.059	-0.019	1.097
women_reasons	4.4268	0.856	5.171	0.000	2.749	6.105
financial_resources	-0.5297	0.254	-2.083	0.037	-1.028	-0.031

Figure 3.5.8 - Second part of Logit Model of Topic and Partners' Variables (small dataset)

job	-0.3362	0.292	-1.150	0.250	-0.909	0.237
procurement	-0.2115	0.416	-0.508	0.611	-1.027	0.604
growth	0.0623	0.274	0.228	0.820	-0.474	0.598
local_activity	-0.6082	0.257	-2.364	0.018	-1.112	-0.104
poverty	-0.2677	0.273	-0.982	0.326	-0.802	0.267
loyal_borrowers	0.1099	0.380	0.289	0.772	-0.635	0.855
PtotalAmountRaised	1.0117	0.217	4.668	0.000	0.587	1.437
PavgBorrowerCost	0.8658	0.210	4.126	0.000	0.454	1.277
PavgProfitability	0.5846	0.456	1.282	0.200	-0.309	1.479
PavgLoanSizePercentPerCapitaIncome	-0.3272	0.164	-2.000	0.045	-0.648	-0.007
PdefaultRate	-2.3372	0.553	-4.229	0.000	-3.420	-1.254
PcurrencyExchangeLossRate	-3.2229	0.752	-4.286	0.000	-4.697	-1.749
PchargesFeesInterest	-1.2429	0.487	-2.552	0.011	-2.197	-0.288
ParrearsRate	-0.1522	0.124	-1.231	0.218	-0.394	0.090
PStatus_closed	0.3461	0.079	4.371	0.000	0.191	0.501
PStatus_inactive	-0.7824	0.143	-5.486	0.000	-1.062	-0.503
PStatus_paused	0.0550	0.175	0.315	0.753	-0.287	0.397
PStatementType_dual_statement	1.6896	0.613	2.754	0.006	0.487	2.892
PavgBorrowerCostType_APR	0.6663	0.154	4.337	0.000	0.365	0.967

An interesting variable have been created in the dataset with residual variables. It is called funding_days and indicates how many days are necessary to fund a loan. In fact, it is obtained from the difference between the fundraising date and the disbursal date.

The variables removed for the multicollinearity problem in the datasets with residual variables are: *investment, retail, fundedAmount, repaymentInterval_monthly, lossliabilityCurrenctExchange_shared, lendersTotalCount.*

In particular, both datasets consider that the following variables produce a positive impact on the project's success: *trading, family_needs, women_reasons, flexibleFundraisingEnabled, distributionModel_fieldPartner, repaymentInterval_irregularly, lossLiabilityCurrencyExchange_lender, lossLiabilityCurrencyExchange_none, lossLiabilityCurrencyExchange_partner.*

Figure 3.5.8 - First part of Logit Model of Topic and Residual Variables (full dataset)

Logit Regression Results

Dep. Variable:	status	**No. Observations:**	36000
Model:	Logit	**Df Residuals:**	35969
Method:	MLE	**Df Model:**	30
Date:	Sat, 28 Jan 2023	**Pseudo R-squ.:**	0.1158
Time:	21:46:29	**Log-Likelihood:**	-5493.8
converged:	True	**LL-Null:**	-6213.4
Covariance Type:	nonrobust	**LLR p-value:**	3.469e-284

	coef	std err	z	P>|z|	[0.025	0.975]
purchase	-0.4561	0.210	-2.171	0.030	-0.868	-0.044
farming	-0.2528	0.219	-1.157	0.247	-0.681	0.175
trading	0.4890	0.225	2.174	0.030	0.048	0.930
family_expenses	0.1002	0.230	0.435	0.664	-0.351	0.552
family_needs	0.8800	0.238	3.698	0.000	0.414	1.346
business	0.2418	0.220	1.098	0.272	-0.190	0.674
improvement	-0.2662	0.190	-1.400	0.161	-0.639	0.106
personal_activity	-0.5523	0.225	-2.458	0.014	-0.993	-0.112
face_difficulties	-1.2011	0.194	-6.203	0.000	-1.581	-0.822
entrepreneurship	0.5560	0.637	0.873	0.383	-0.692	1.804
causes	-0.9898	0.175	-5.656	0.000	-1.333	-0.647
progress	-1.4339	0.221	-6.483	0.000	-1.867	-1.000
women_reasons	5.2379	0.893	5.862	0.000	3.487	6.989
financial_resources	-1.7022	0.172	-9.871	0.000	-2.040	-1.364
job	-1.5484	0.220	-7.027	0.000	-1.980	-1.116
procurement	-0.2818	0.356	-0.791	0.429	-0.980	0.416
growth	-1.0536	0.199	-5.297	0.000	-1.443	-0.664
local_activity	-1.8248	0.191	-9.558	0.000	-2.199	-1.451
poverty	-0.9316	0.208	-4.476	0.000	-1.339	-0.524
loyal_borrowers	-1.3379	0.309	-4.336	0.000	-1.943	-0.733
borrowers_situation	-0.5523	0.213	-2.596	0.009	-0.969	-0.135
borrowerCount	0.1820	0.414	0.440	0.660	-0.629	0.993

Whereas the opposite effect is generated if the following variables are present: *purchase, personal_activity, face_difficulties, causes, progress, financial_resources, job, growth, local_activity, poverty, loyal_borrowers, borrowers_situation, lenderRepaymentTerm, funding_days.*

The results obtained from the partners, sectors and regions analysis are also confirmed in this case. In fact, women' motivations to receive a loan are the most important for lenders' evaluation of the project. Instead, borrowers' descriptions negatively impact on the campaign's success if they are associated with the technical specifics of loans (Figure 3.5.8, 3.5.9, 3.5.10, 3.5.11).

Figure 3.5.9 - Second part of Logit Model of Topic and Residuals Variables (full dataset)

lenderRepaymentTerm	-7.4605	0.406	-18.373	0.000	-8.256	-6.665
flexibleFundraisingEnabled	1.0126	0.248	4.082	0.000	0.526	1.499
distributionModel_fieldPartner	4.8717	0.286	17.060	0.000	4.312	5.431
repaymentInterval_at_end	0.1003	0.105	0.960	0.337	-0.105	0.305
repaymentInterval_irregularly	0.5288	0.138	3.838	0.000	0.259	0.799
lossLiabilityCurrencyExchange_lender	1.6464	0.384	4.290	0.000	0.894	2.399
lossLiabilityCurrencyExchange_none	0.9147	0.088	10.388	0.000	0.742	1.087
lossLiabilityCurrencyExchange_partner	2.7591	0.280	9.851	0.000	2.210	3.308
funding_days	-2.3520	0.921	-2.554	0.011	-4.157	-0.547

Figure 3.5.10 - First part of Logit Model of Topic and Residuals Variables (small dataset)

Logit Regression Results

Dep. Variable:	status	**No. Observations:**	18000
Model:	Logit	**Df Residuals:**	17969
Method:	MLE	**Df Model:**	30
Date:	Sat, 28 Jan 2023	**Pseudo R-squ.:**	0.1383
Time:	20:43:39	**Log-Likelihood:**	-4434.5
converged:	True	**LL-Null:**	-5146.2
Covariance Type:	nonrobust	**LLR p-value:**	8.140e-281

	coef	std err	z	P>|z|	[0.025	0.975]
purchase	-0.4706	0.220	-2.136	0.033	-0.902	-0.039
farming	-0.2471	0.229	-1.080	0.280	-0.696	0.201
trading	0.5103	0.234	2.177	0.029	0.051	0.970
family_expenses	0.0844	0.239	0.353	0.724	-0.384	0.553
family_needs	0.8122	0.246	3.300	0.001	0.330	1.294
business	0.2356	0.227	1.037	0.300	-0.210	0.681
improvement	-0.2639	0.198	-1.334	0.182	-0.652	0.124
personal_activity	-0.4553	0.238	-1.917	0.055	-0.921	0.010
face_difficulties	-1.1982	0.202	-5.945	0.000	-1.593	-0.803
entrepreneurship	0.6867	0.653	1.051	0.293	-0.594	1.967
causes	-0.8947	0.181	-4.930	0.000	-1.250	-0.539
progress	-1.4111	0.231	-6.098	0.000	-1.865	-0.958
borrowers_situation	-0.5547	0.220	-2.516	0.012	-0.987	-0.123
women_reasons	5.7243	0.958	5.974	0.000	3.846	7.603
financial_resources	-1.6225	0.181	-8.946	0.000	-1.978	-1.267
job	-1.4474	0.234	-6.197	0.000	-1.905	-0.990
procurement	-0.2546	0.364	-0.699	0.484	-0.969	0.459
growth	-1.0304	0.207	-4.975	0.000	-1.436	-0.624
local_activity	1.8062	0.202	8.959	0.000	2.201	1.411

Figure 3.5.11 - Second part of Logit Model of Topic and Residuals Variables (small dataset)

poverty	-0.8400	0.217	-3.865	0.000	-1.266	-0.414
loyal_borrowers	-1.3058	0.323	-4.048	0.000	-1.938	-0.674
borrowerCount	0.0299	0.427	0.070	0.944	-0.807	0.867
lenderRepaymentTerm	-8.0130	0.437	-18.333	0.000	-8.870	-7.156
flexibleFundraisingEnabled	0.9770	0.251	3.891	0.000	0.485	1.469
distributionModel_fieldPartner	4.1780	0.300	13.944	0.000	3.591	4.765
repaymentInterval_at_end	0.0785	0.110	0.715	0.475	-0.137	0.294
repaymentInterval_irregularly	0.4942	0.142	3.492	0.000	0.217	0.772
lossLiabilityCurrencyExchange_lender	1.7215	0.388	4.437	0.000	0.961	2.482
lossLiabilityCurrencyExchange_none	0.9537	0.091	10.466	0.000	0.775	1.132
lossLiabilityCurrencyExchange_partner	2.7360	0.281	9.729	0.000	2.185	3.287
funding_days	-1.8070	0.741	-2.438	0.015	-3.260	-0.354

Conclusion

In conclusion, microfinance is the most effective and efficient financial service in developing countries. It significantly offers the possibility to start a business, sustain small activities in providing high-quality goods and services and improve the living conditions of underprivileged borrowers. In particular, lending crowdfunding platforms are utterly diffused for promoting social causes. Kiva is one of the most acknowledge organizations that consistently support local borrowers, especially women, whose main desire is to guarantee the welfare of their children, promote their ideas and realize them concretely to reduce gender inequality. Both literature and the empirical analysis have highlighted the importance of the borrower's description in the loan application and the level of default risk associated with a specific project.

Regarding the data analysis, some open points that could be explored further are:

- Randomness regarding the topic modeling;
- The comparison between different methods to define the optimal number of topics;
- Adding the examination of other dependent variables like the *funding_days*;
- Applying further techniques to detect the frequent presence of p-values with a value of 0.000 in the logit models.

Appendix

A – Data Cleaning

Figure A.1 – Import libraries and read datasets in csv-format files

```
import pandas as pd
import numpy as np
from pathlib import Path

import warnings
warnings.filterwarnings("ignore")

import matplotlib.pyplot as plt
import plotly.express as px
import plotly.graph_objects as go
%matplotlib inline
import seaborn as sns
plt.style.use('seaborn-whitegrid')
palette = plt.get_cmap("tab10")
```

```
#read all datasets

original_loan_2005_2010 = pd.read_csv("C:\\Users\\catet\\Desktop\\dataset_modificati\\2005-2010_loan.csv", low_memory=False)
original_loan_2011 = pd.read_csv("C:\\Users\\catet\\Desktop\\dataset_modificati\\2011_loan.csv", low_memory=False)
original_loan_2012 = pd.read_csv("C:\\Users\\catet\\Desktop\\dataset_modificati\\2012_loan.csv", low_memory=False)
original_loan_2013 = pd.read_csv("C:\\Users\\catet\\Desktop\\dataset_modificati\\2013_loan.csv", low_memory=False)
original_loan_2014 = pd.read_csv("C:\\Users\\catet\\Desktop\\dataset_modificati\\2014_loan.csv", low_memory=False)
original_loan_2015 = pd.read_csv("C:\\Users\\catet\\Desktop\\dataset_modificati\\2015_loan.csv", low_memory=False)
original_loan_2016 = pd.read_csv("C:\\Users\\catet\\Desktop\\dataset_modificati\\2016_loan.csv", low_memory=False)
original_loan_2017 = pd.read_csv("C:\\Users\\catet\\Desktop\\dataset_modificati\\2017_loan.csv", low_memory=False)
original_loan_2018 = pd.read_csv("C:\\Users\\catet\\Desktop\\dataset_modificati\\2018_loan.csv", low_memory=False)
original_loan_2019 = pd.read_csv("C:\\Users\\catet\\Desktop\\dataset_modificati\\2019_loan.csv", low_memory=False)
original_loan_2020 = pd.read_csv("C:\\Users\\catet\\Desktop\\dataset_modificati\\2020_loan.csv", low_memory=False)
original_loan_2021 = pd.read_csv("C:\\Users\\catet\\Desktop\\dataset_modificati\\2021_loan.csv", low_memory=False)
```

Figure A.2 – Check null values in all variables and delete two columns

```
#check null values in the dataset loan_2005_2010
#the columns TrusteeId and plannedExpirationDate have only null values

for col in original_loan_2005_2010:
    print(col, sum(pd.isnull(original_loan_2005_2010[col])))
```

```
id 0
ActivityId 33
SectorId 33
CountryIsoCode 33
Region 33
PartnerId 33
TrusteeId 219337
fundedAmount 33
city 12010
lendersTotalCount 33
teamsTotalCount 33
flexibleFundraisingEnabled 33
lossLiabilityCurrencyExchange 33
borrowerCount 33
distributionModel 33
fundraisingDate 33
hasCurrencyExchangeLossLenders 33
repaymentInterval 33
plannedExpirationDate 219337
raisedDate 183
status 33
loanName 49
loanAmount 33
lenderRepaymentTerm 33
disbursalDate 33
fundraisingYEAR; 33
```

```
# delete columns TrusteeId and plannedExpirationDate because they have all Nans

original_loan_2005_2010.pop('TrusteeId')
original_loan_2005_2010.pop('plannedExpirationDate')
```

```
0        NaN
1        NaN
2        NaN
3        NaN
4        NaN
          ..
219332   NaN
219333   NaN
219334   NaN
219335   NaN
219336   NaN
Name: plannedExpirationDate, Length: 219337, dtype: float64
```

Figure A.3 – Remove every observations where *id* is not an integer datatype

```
#remove every row of the dataset where the "id" column is not an integer
#create a new dataset for data cleaning

print("Length of dataset before filtering on id: ", len(original_loan_2005_2010))
loan_2005_2010 = original_loan_2005_2010[original_loan_2005_2010['id'].apply(lambda x: str(x).isdigit())]

print("Length of dataset after filtering on id: ", len(loan_2005_2010))
```

```
Length of dataset before filtering on id:  219337
Length of dataset after filtering on id:  219304
```

Figure A.4 – Substitute null values

```
#substitute null values in the columns city and loanName

loan_2005_2010["city"].fillna("Not specified", inplace = True)
loan_2005_2010["loanName"].fillna("Not specified", inplace = True)
```

Figure A.5 – Transform into date type

```
#transform into date type

loan_2005_2010['fundraisingDate']= pd.to_datetime(loan_2005_2010['fundraisingDate'])
loan_2005_2010['raisedDate']= pd.to_datetime(loan_2005_2010['raisedDate'])
loan_2005_2010['disbursalDate']= pd.to_datetime(loan_2005_2010['disbursalDate'])
```

Figure A.6 - Transform into integers

```
#transform into integers

loan_2005_2010['id'] = loan_2005_2010['id'].astype(int)
loan_2005_2010['ActivityId'] = loan_2005_2010['ActivityId'].astype(int)
loan_2005_2010['SectorId'] = loan_2005_2010['SectorId'].astype(int)
loan_2005_2010['PartnerId'] = loan_2005_2010['PartnerId'].astype(int)
loan_2005_2010['fundedAmount'] = loan_2005_2010['fundedAmount'].astype(int)
loan_2005_2010['lendersTotalCount'] = loan_2005_2010['lendersTotalCount'].astype(int)
loan_2005_2010['teamsTotalCount'] = loan_2005_2010['teamsTotalCount'].astype(int)
loan_2005_2010['flexibleFundraisingEnabled'] = loan_2005_2010['flexibleFundraisingEnabled'].astype(int)
loan_2005_2010['borrowerCount'] = loan_2005_2010['borrowerCount'].astype(int)
loan_2005_2010['hasCurrencyExchangeLossLenders'] = loan_2005_2010['hasCurrencyExchangeLossLenders'].astype(int)
loan_2005_2010['loanAmount'] = loan_2005_2010['loanAmount'].astype(int)
loan_2005_2010['lenderRepaymentTerm'] = loan_2005_2010['lenderRepaymentTerm'].astype(int)
loan_2005_2010["fundraisingYEAR"] = loan_2005_2010["fundraisingYEAR"].astype(int)
```

Figure A.7 – Number of columns, rows and their data type

```
loan_2005_2010.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 219304 entries, 0 to 219336
Data columns (total 24 columns):
 #   Column                             Non-Null Count   Dtype
---  ------                             --------------   -----
 0   id                                 219304 non-null  int32
 1   ActivityId                         219304 non-null  int32
 2   SectorId                           219304 non-null  int32
 3   CountryIsoCode                     219304 non-null  object
 4   Region                             219304 non-null  object
 5   PartnerId                          219304 non-null  int32
 6   fundedAmount                       219304 non-null  int32
 7   city                               219304 non-null  object
 8   lendersTotalCount                  219304 non-null  int32
 9   teamsTotalCount                    219304 non-null  int32
 10  flexibleFundraisingEnabled         219304 non-null  int32
 11  lossLiabilityCurrencyExchange      219304 non-null  object
 12  borrowerCount                      219304 non-null  int32
 13  distributionModel                  219304 non-null  object
 14  fundraisingDate                    219304 non-null  datetime64[ns]
 15  hasCurrencyExchangeLossLenders     219304 non-null  int32
 16  repaymentInterval                  219304 non-null  object
 17  raisedDate                         219154 non-null  datetime64[ns]
 18  status                             219304 non-null  object
 19  loanName                           219304 non-null  object
 20  loanAmount                         219304 non-null  int32
 21  lenderRepaymentTerm                219304 non-null  int32
 22  disbursalDate                      219304 non-null  datetime64[ns]
 23  fundraisingYEAR                    219304 non-null  int32
dtypes: datetime64[ns](3), int32(13), object(8)
memory usage: 31.0+ MB
```

Figure A.8 – Principal indicators of the dataset

```
loan_2005_2010.describe()
```

	id	ActivityId	SectorId	PartnerId	fundedAmount	lendersTotalCount	teamsTotalCount
count	219304.000000	219304.000000	219304.000000	219304.000000	219304.000000	219304.000000	219304.000000
mean	138742.505107	76.417471	7.126751	85.558540	733.745281	21.908702	6.163002
std	72908.585083	48.161224	4.208898	41.150186	691.143431	20.854195	7.487861
min	84.000000	9.000000	1.000000	1.000000	25.000000	1.000000	0.000000
25%	81035.750000	35.000000	4.000000	60.000000	325.000000	9.000000	1.000000
50%	140368.500000	67.000000	7.000000	91.000000	550.000000	17.000000	4.000000
75%	201750.250000	104.000000	12.000000	119.000000	950.000000	28.000000	8.000000
max	262603.000000	224.000000	17.000000	180.000000	10000.000000	330.000000	101.000000

Figure A.9 – Delete time values in *plannedExpirationDate*

```
loan_2018['plannedExpirationDate'] = loan_2018['plannedExpirationDate'].str.slice(0, 10)
loan_2018['plannedExpirationDate']= pd.to_datetime(loan_2018['plannedExpirationDate'])
```

Figure A.10 – Outer join between the datasets with *use* and *description* variables

```
#load and read other datasets; create a unique one in the join3_2005_2010

A_2005_2010_excel = pd.read_excel('C:\\Users\\catet\\Desktop\\dataset_modificati\\2005-2010_use_and_descriptions_A.xlsx')
A_2005_2010 = pd.DataFrame(A_2005_2010_excel, columns=['id', 'use', 'description'])

B_2005_2010_excel = pd.read_excel('C:\\Users\\catet\\Desktop\\dataset_modificati\\2005-2010_use_and_descriptions_B.xlsx')
B_2005_2010 = pd.DataFrame(B_2005_2010_excel, columns=['id', 'use', 'description'])

C_2005_2010_excel = pd.read_excel('C:\\Users\\catet\\Desktop\\dataset_modificati\\2005-2010_use_and_descriptions_C.xlsx')
C_2005_2010 = pd.DataFrame(C_2005_2010_excel, columns=['id', 'use', 'description'])

D_2005_2010_excel = pd.read_excel('C:\\Users\\catet\\Desktop\\dataset_modificati\\2005-2010_use_and_descriptions_D.xlsx')
D_2005_2010 = pd.DataFrame(D_2005_2010_excel, columns=['id', 'use', 'description'])

join1_2005_2010 = pd.merge(A_2005_2010, B_2005_2010, how='outer')
join2_2005_2010 = pd.merge(join1_2005_2010, C_2005_2010, how='outer')
join3_2005_2010 = pd.merge(join2_2005_2010, D_2005_2010, how='outer')
```

Figure A.11 – Inner join between the dataset with loan information and the ones of activity, sector, country and partner

```
u_loan_2005_2010 = pd.merge(loan_2005_2010, join3_2005_2010, on = 'id')

#dataset with the activity name
loan_2005_2010_ac= pd.merge(u_loan_2005_2010, Activity_data, on = 'ActivityId')

#dataset with the sector name
loan_2005_2010_sec = pd.merge(loan_2005_2010_ac, sector_data, on = 'SectorId')

#dataset with the country name
loan_2005_2010_cou = pd.merge(loan_2005_2010_sec, country_data, on = 'CountryIsoCode')

#join with the partner dataset
cleaned_final_2005_2010 = pd.merge(loan_2005_2010_cou, partner_data, on = 'PartnerId')
```

Figure A.12 – Creation of a single dataset called *small_data*

```
#take randomly 4000 observations from each dataset to reduce the dimensionality
import random
random_state = 123

small_2005_2010 = cleaned_final_2005_2010.sample(n = 3000, replace = False, random_state = random_state)
small_2011 = cleaned_final_2011.sample(n = 3000, replace = False, random_state = random_state)
small_2012 = cleaned_final_2012.sample(n = 3000, replace = False, random_state = random_state)
small_2013 = cleaned_final_2013.sample(n = 3000, replace = False, random_state = random_state)
small_2014 = cleaned_final_2014.sample(n = 3000, replace = False, random_state = random_state)
small_2015 = cleaned_final_2015.sample(n = 3000, replace = False, random_state = random_state)
small_2016 = cleaned_final_2016.sample(n = 3000, replace = False, random_state = random_state)
small_2017 = cleaned_final_2017.sample(n = 3000, replace = False, random_state = random_state)
small_2018 = cleaned_final_2018.sample(n = 3000, replace = False, random_state = random_state)
small_2019 = cleaned_final_2019.sample(n = 3000, replace = False, random_state = random_state)
small_2020 = cleaned_final_2020.sample(n = 3000, replace = False, random_state = random_state)
small_2021 = cleaned_final_2021.sample(n = 3000, replace = False, random_state = random_state)
```

```
#create a unique dataset

data_10_11 = pd.merge(small_2005_2010, small_2011, how='outer')
data_11_12 = pd.merge(data_10_11, small_2012, how='outer')
data_12_13 = pd.merge(data_11_12, small_2013, how='outer')
data_13_14 = pd.merge(data_12_13, small_2014, how='outer')
data_14_15 = pd.merge(data_13_14, small_2015, how='outer')
data_15_16 = pd.merge(data_14_15, small_2016, how='outer')
data_16_17 = pd.merge(data_15_16, small_2017, how='outer')
data_17_18 = pd.merge(data_16_17, small_2018, how='outer')
data_18_19 = pd.merge(data_17_18, small_2019, how='outer')
data_19_20 = pd.merge(data_18_19, small_2020, how='outer')
small_data = pd.merge(data_19_20, small_2021, how='outer')
```

B – Some Examples of Data Transformation

Figure B.1 – Realization of the Line Plot for the "Distribution of Crucial Rates over the Years"

```
df1 = small_data.sort_values(['fundraisingYEAR'],ascending=False).groupby('fundraisingYEAR')['PavgLoanSizePercentPerCapitaIncome'].mean().reset_index()
df2 = small_data.sort_values(['fundraisingYEAR'],ascending=False).groupby('fundraisingYEAR')['PdefaultRate'].mean().reset_index()
df3 = small_data.sort_values(['fundraisingYEAR'],ascending=False).groupby('fundraisingYEAR')['PloansAtRiskRate'].mean().reset_index()

fig, ax = plt.subplots(figsize=(15, 8))

plt.plot('fundraisingYEAR', 'PavgLoanSizePercentPerCapitaIncome', data=df1, color='deepskyblue',
         label = 'Avg Loan Size per CapIncome per year %', linewidth = 3)
plt.plot( 'fundraisingYEAR',  'PdefaultRate', data=df2, color='r',
         label = 'Avg Default Rate per year %', linewidth = 3)
plt.plot('fundraisingYEAR','PloansAtRiskRate', data=df3,
         label = 'Avg Loans at Risk Rate per year %', color='limegreen', linewidth = 3)

plt.title("Distribution of Crucial Rates over the Years\n", fontsize=30)
plt.xlabel("Fundraising Year", fontsize=20)
plt.ylabel("Percentage Values ", fontsize=20)
plt.xticks(fontsize=15)
plt.yticks(fontsize=15)

plt.legend(loc=1, bbox_to_anchor=(1.4, 1), fontsize=15)
plt.show()
```

Figure B.2 – Realization of the Map for the “Total Number of Funded Amount by Country”

```
#import this dataset downloaded from kaggle website to have extra columns for the creation of the following map

dataset_country_new = pd.read_csv("C:\\Users\\catet\\Desktop\\Tesi_Magistrale\\wikipedia-iso-country-codes.csv", low_memory=False)
dataset_country_new = dataset_country_new.rename(columns={"Alpha-2 code": "CountryIsoCode"})
data_for_map = pd.merge(small_data, dataset_country_new, on = 'CountryIsoCode')
```

```
#the column Alpha-3 code is crucial for the following graph
df_for_map = data_for_map.groupby(['CountryName', 'Alpha-3 code'], as_index = False)['fundedAmount'].count()

df_for_map.sort_values(by=['fundedAmount'], ascending=False, inplace=True)
```

```
fig = px.choropleth(df_for_map, locations="Alpha-3 code",
                    color="fundedAmount",
                   hover_name="CountryName",
                   color_continuous_scale=px.colors.sequential.speed)

fig.update_layout(
    title={
        'text': "Total number of funded amount by country",
        'y':0.9,
        'x':0.5,
        'xanchor': 'center',
        'yanchor': 'bottom'},title_font_size=25 )
```

C – Topic Modeling

Figure C.1 - Divide *description* into two columns

```
#split each row of the "description" column when "translated" word is present
#because it is followed by the name and surname of the description author, a meaningless information for the topic analysis
new_column = smallest_data["description"].str.split("translated", expand = True)

# create two columns and delete the one with the author's details
smallest_data["description_final"]= new_column[0]

smallest_data["description_previous"]= new_column[1]

smallest_data.drop(columns = ["description"], inplace = True)
smallest_data.drop(columns = ["description_previous"], inplace = True)
```

Figure C.2 - Tokenization, Lemmatization, Stopwords of the *description* variable

```python
# add meaningless words for the topic analysis in d_add_stopwords (such as common verbs, first names, last names etc)
d_add_stopwords = ['will', 'able', 'requested', 'want', 'new','future', 'nwtf','day',
                  '_x000D_', 'additional', 'almost', 'enough', 'etc', 'mrs', 'abject', 'ibn', 'betty', 'laura',
                 'necessary', 'got', 'philippine', 'ricardo', 'cham', 'besides', 'soft', 'along',
                  'greenstein', 'resulted', 'later', 'complex', 'mavlavi', 'translation', 'weak', 'bantama', 'elmina',  'nkwa',
                  'naivasha', 'firdousi', 'permettront', 'carolina', 'asasah', 'purok', 'taken', 'wednesday', 'hluvuku', 'grameen',
                  'mindanao', 'quecha', "lu'u", 'located', 'ilse', 'ackerman', 'marty', 'lis, a', 'molle', 'del', 'thi',
                 'high', 'greatly', 'togo', 'even', 'bothered', 'without', 'with', 'samira', 'annette', 'marie', 'valentina', 'sto', 'different',
                  'difference', 'addition', 'low', 'meet', 'tell', 'say', 'end', 'known', 'many', 'asked', 'kes', 'prior', 'ulises',
                 'especially', 'primarily', 'friday', 'cluster', 'tengo', 'viernes', 'next', 'rudaki', 'avicenna', 'ferdowsi',
                  'omar', 'khayyam', 'jomi', 'rumi', 'didn', 'anyone', 'whil', 'varied', 'multi', 'strong', 'multi', 'varied', 'post',
                  'puno', 'nearly', 'nao', 'specialty', 'serf', 'ketut', 'doles', 'suwiti', 'mahartini', 'saka', 'year', 'impro',
                  'donteav', 'hasn', 'rod', 'past', 'mauriceo', 'tac', 'adalathon', 'adalathan', 'till', 'former', 'basically', 'approximately',
                  'per', 'fortich', 'sits', 'perhaps', 'epic', 'secondary', 'whenever', 'href']

d_common_adverbs = ['always', 'usually', 'normally', 'generally', 'often', 'frequently', 'sometimes', 'occasionally', 'hardly', 'ever', 'rarely',
                    'never', 'daily','weekly', 'monthly', 'yearly', 'every', 'hours', 'week', 'gladly', 'gently', 'quietly', 'safely', 'truthfully',
                    'warmly', 'wildly', 'carefully', 'wisely', 'hard', 'fast', 'straight', 'well', 'angrily', 'boldly', 'before', 'later', 'since', 'soon',
                    'still', 'yet', 'early', 'earlier', 'eventually', 'recently', 'previously', 'finally', 'above', 'behind', 'below', 'down', 'far',
                    'outside', 'towards', 'under', 'upstairs', 'back', 'over', 'away', 'off', 'therefore', 'thus', 'accordingly', 'indeed', 'however',
                    'thereby', 'consequently', 'subsequently', 'indeed', 'now', 'nevertheless', 'nonetheless' , 'around', 'previous', 'forward',
                    'tomorrow', 'yesterday' , 'today', 'tonight', 'then', 'morning', 'evening', 'afternoon', 'currently', 'moment', 'momentarily', 'case']

d_adverbs_manner = ['accidentally', 'angrily','anxiously','awkwardly','badly','beautifully','blindly','boldly','bravely','brightly','busily','calmly',
                    'carefully','carelessly','cautiously','cheerfully','clearly','closely','correctly','courageously','cruelly','daringly','deliberately',
                    'doubtfully','eagerly','easily','elegantly','enormously','enthusiastically','equally','eventually','exactly','faithfully',
                    'fast','fatally','fiercely','fondly','foolishly','fortunately','frankly','frantically', 'generously','gently', 'gladly','gracefully',
                    'greedily','happily','hard','hastily','healthily','honestly','hungrily','hurriedly','inadequately','ingeniously','innocently',
                    'inquisitively','irritably','joyously','justly','kindly','lazily','loosely','loudly','madly','mortally','mysteriously',
                    'neatly','nervously','noisily','obediently','openly','painfully','patiently','perfectly','politely','poorly','powerfully',
                    'promptly','punctually','quickly','quietly','rapidly','rarely','really','recklessly','regularly','reluctantly','repeatedly',
                    'rightfully','roughly','rudely','sadly','safely','selfishly','sensibly','seriously','sharply','shyly','silently','sleepily',
                    'slowly','smoothly','softly','solemnly','speedily','stealthily','sternly','straight','stupidly','successfully','suddenly',
                    'suspiciously','swiftly','tenderly','tensely','thoughtfully','tightly','truthfully','unexpectedly','victoriously',
                    'violently','vivaciously','warmly','weakly','wearily','well','wildly','wisely' ]



texts_descr_final = []

# Loop through document list
for j in descr_topic_modeling.iteritems():
    # clean and tokenize document string
    d_raw = str(j[1]).lower()
    d_tokens = tokenizer.tokenize(d_raw)

    # remove stop words from tokens
    d_stopped_tokens = [d_raw for d_raw in d_tokens if not d_raw in en_stop]

     #remove more common adverbs
    d_without_adverbs = [d_raw for d_raw in d_stopped_tokens if not d_raw in d_common_adverbs]

    #remove more common adverbs of manner
    d_without_adv_manner = [d_raw for d_raw in d_without_adverbs if not d_raw in d_adverbs_manner]

    #remove borrowers' name and surname
    d_without_name = [d_raw for d_raw in d_without_adv_manner if not d_raw in loan_names]

    # remove stop words from tokens
    d_stopped_tokens_new = [d_raw for d_raw in d_without_name if not d_raw in d_add_stopwords]

    # Lemmatize tokens
    d_lemma_tokens = [lemmatizer.lemmatize(d_tokens) for d_tokens in d_stopped_tokens_new]
```

Figure C.3 - The function to the calculation of coherence

```python
# create a function to compute the coherence value for different number of topics

def compute_coherence_values(dictionary, corpus, texts, limit, start=1, step=1, trials = 10, save_to_csv = False, name = None):

    # calculate coherence values in 10 trials and save them in a dataframe
    # create a numpy ndarray with trials rows and limit - start / step columns

    coherence_values_all_trials = np.zeros((trials, int((limit - start) / step+1)))

    for i in range(trials):
        coherence_values = []
        model_list = []
        for num_topics in range(start, limit, step):

            ldamodel = gensim.models.LdaMulticore(corpus=corpus, id2word=dictionary, num_topics=num_topics, workers=4)
            model_list.append(ldamodel)
            coherencemodel = CoherenceModel(model=ldamodel, texts=texts, dictionary=dictionary,  coherence='c_v')
            coherence_values.append(coherencemodel.get_coherence())
        coherence_values = np.array(coherence_values)
        coherence_values_all_trials[i,:] = coherence_values

    if save_to_csv:
        # save the coherence values in a csv file
        df = pd.DataFrame(coherence_values_all_trials)
        if name is None:
            name = 'coherence_values_all_trials'
        df.to_csv(name+".csv", index=False)
    return coherence_values_all_trials
```

Figure C.4 – Seven number of topics for the *use* variable

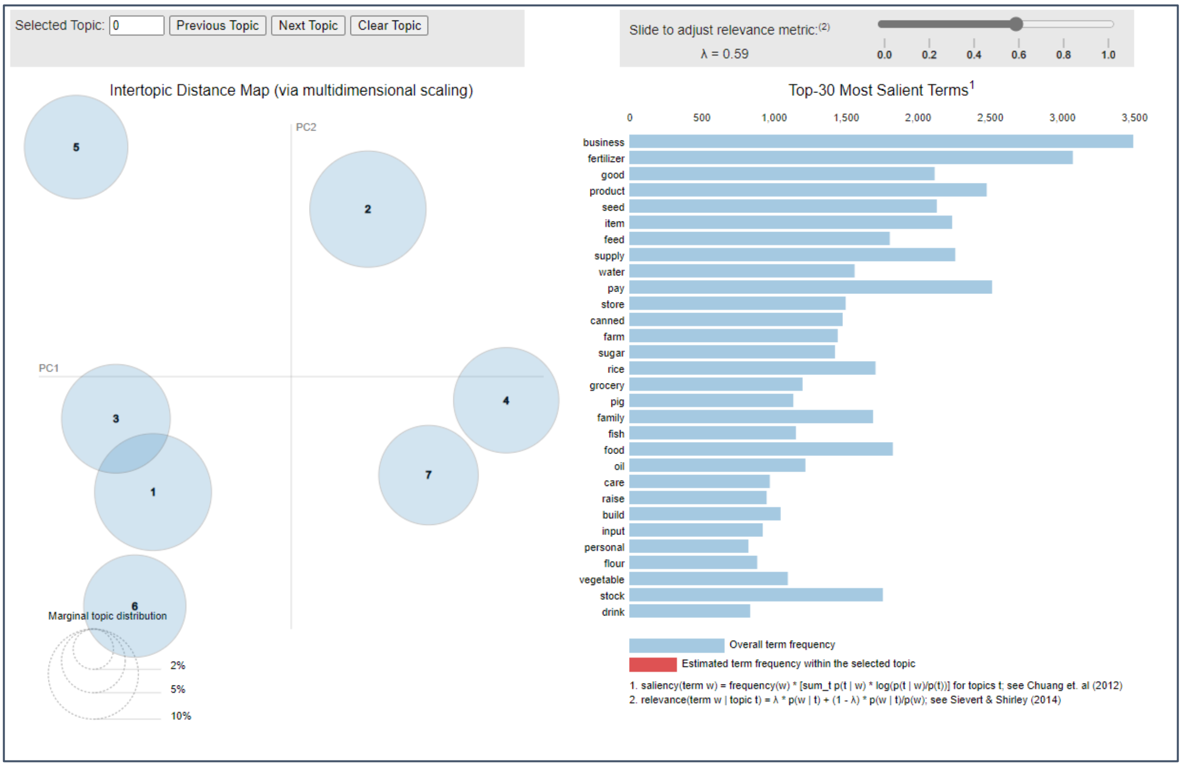


Figure C.5 – The second out of nine topics of the *use* variable

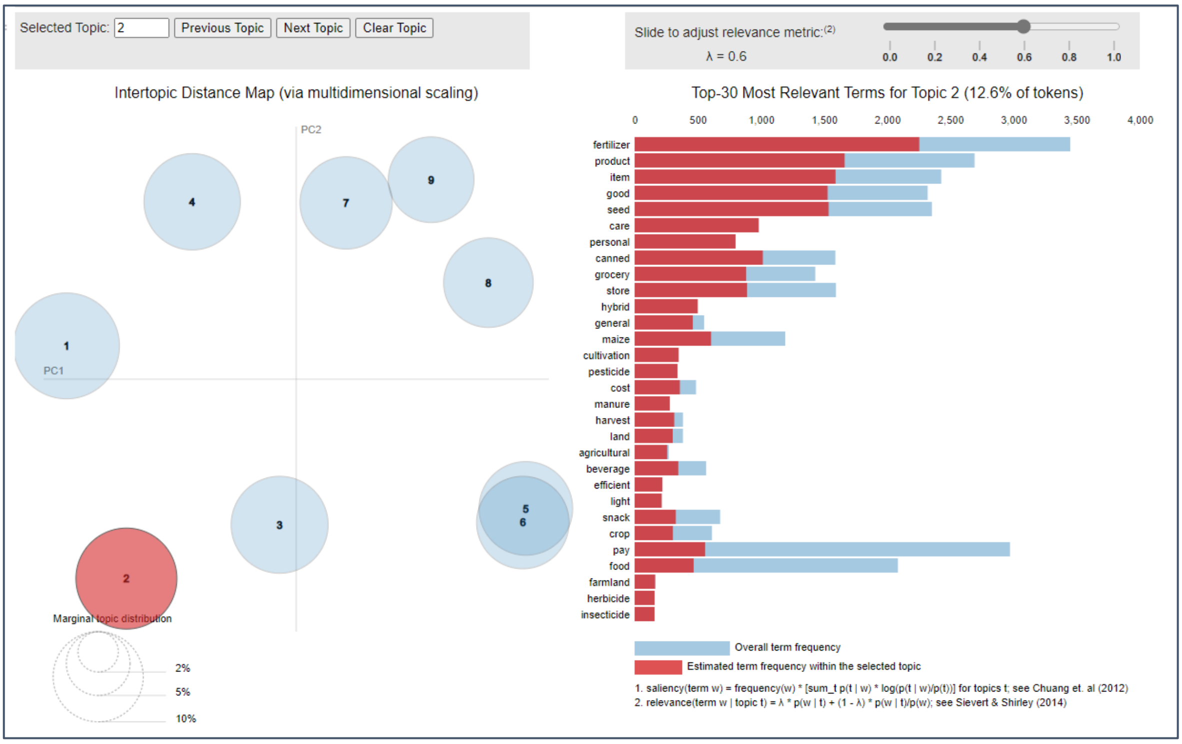


Figure C.6 – The third out of nine topics of the *use* variable

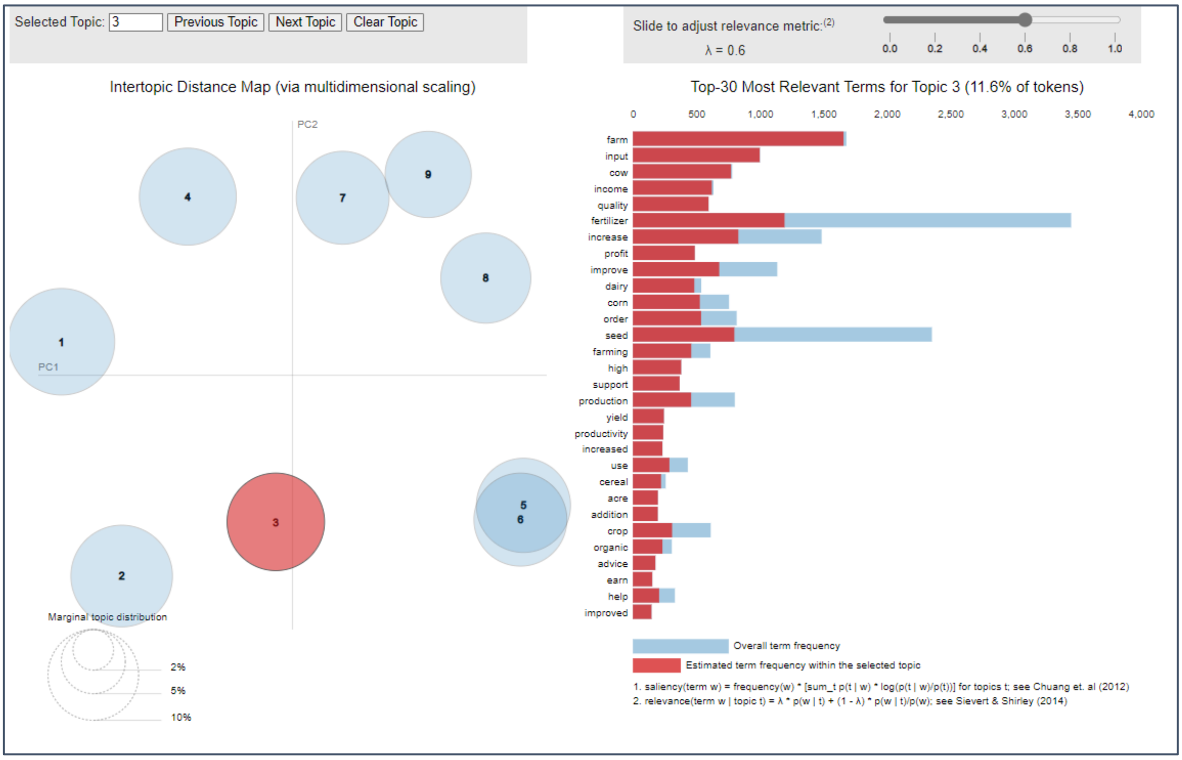


Figure C.7 – The fourth out of nine topics of the *use* variable

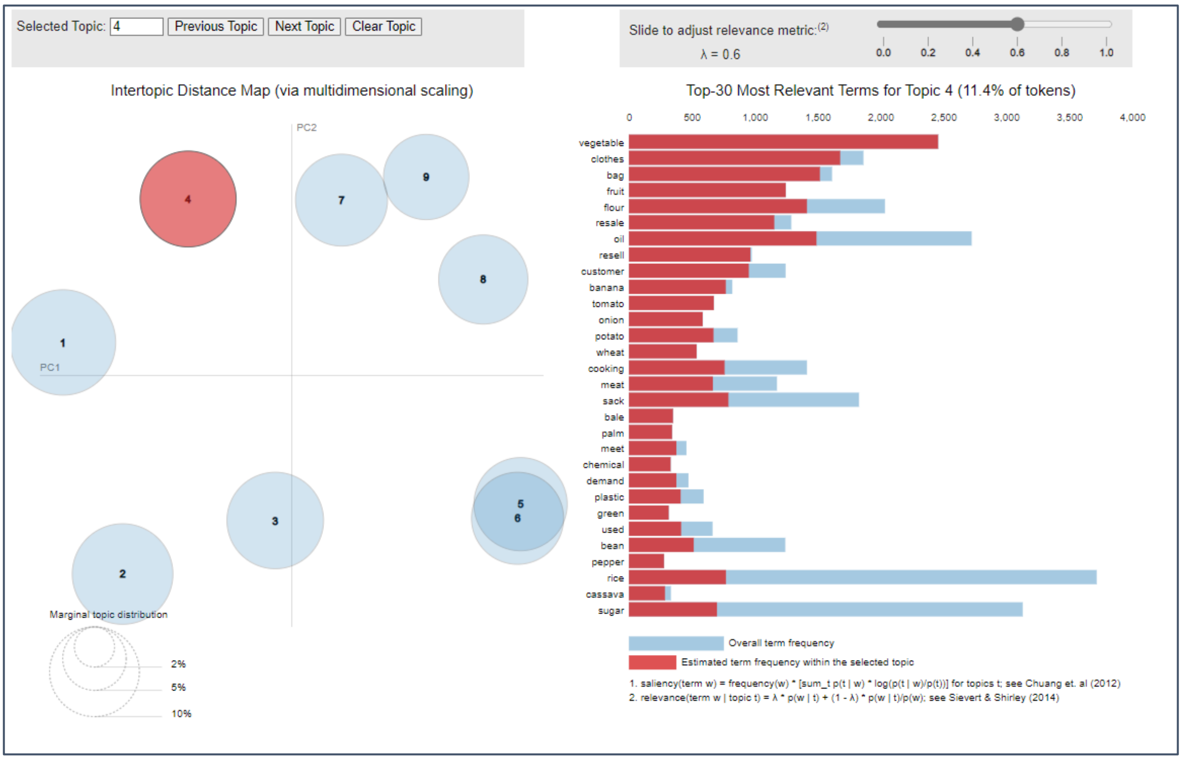


Figure C.8 – The fifth out of nine topics of the *use* variable

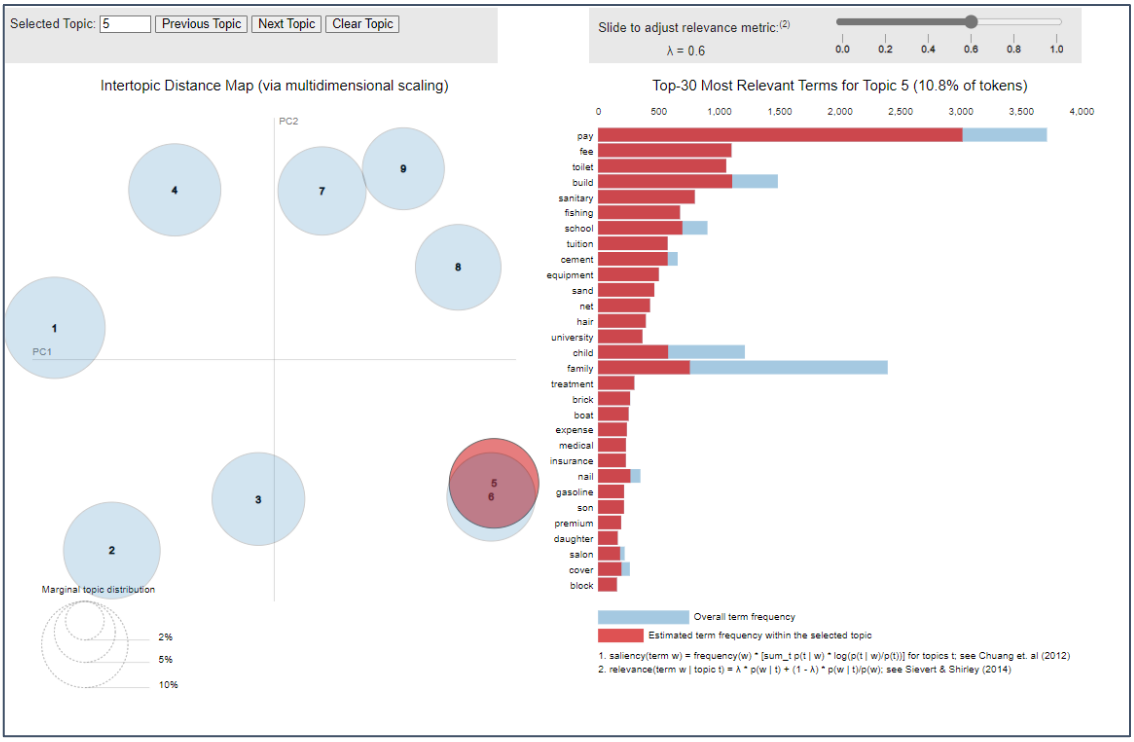


Figure C.9 – The sixth out of nine topics of the *use* variable

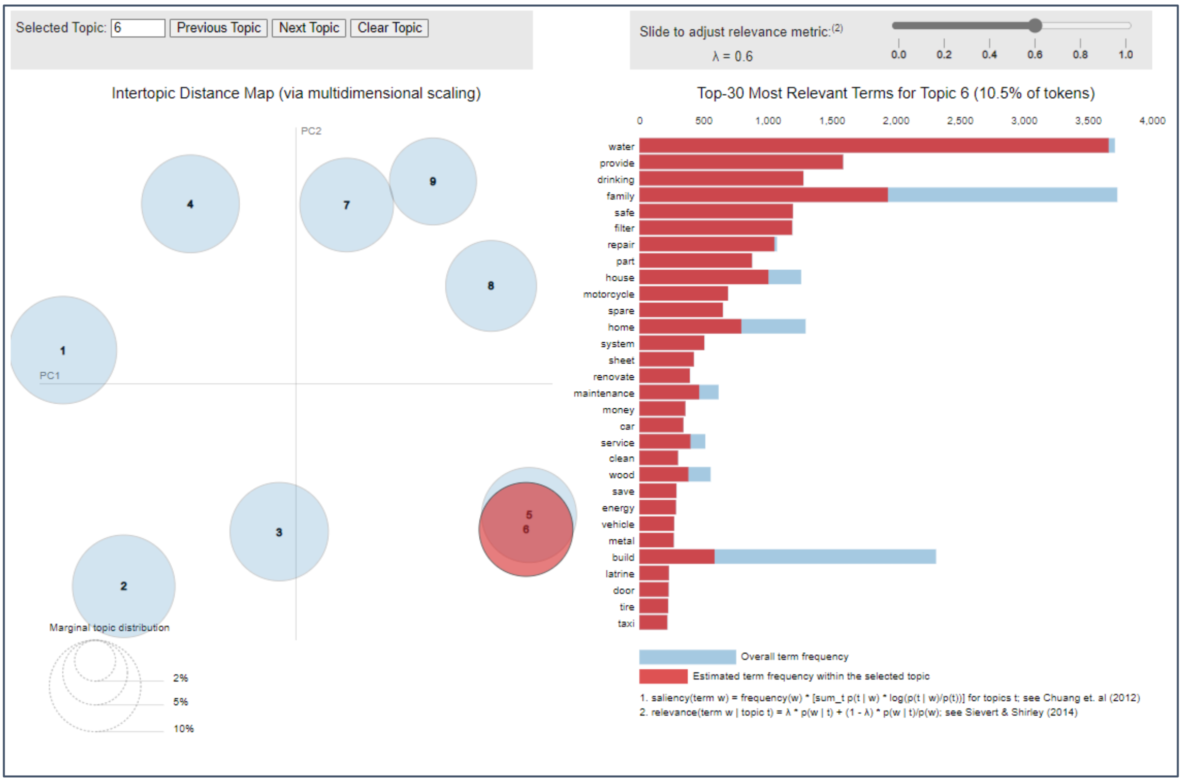


Figure C.10 – The seventh out of nine topics of the *use* variable

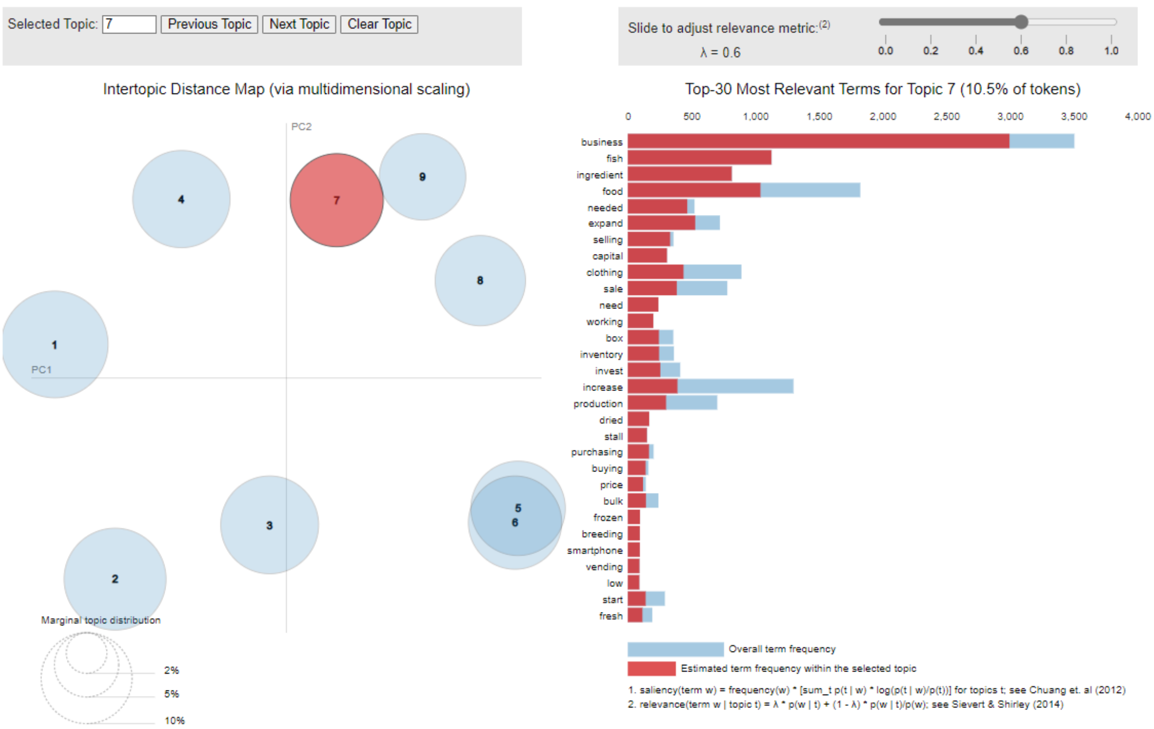


Figure C.11 – The eight out of nine topics of the *use* variable

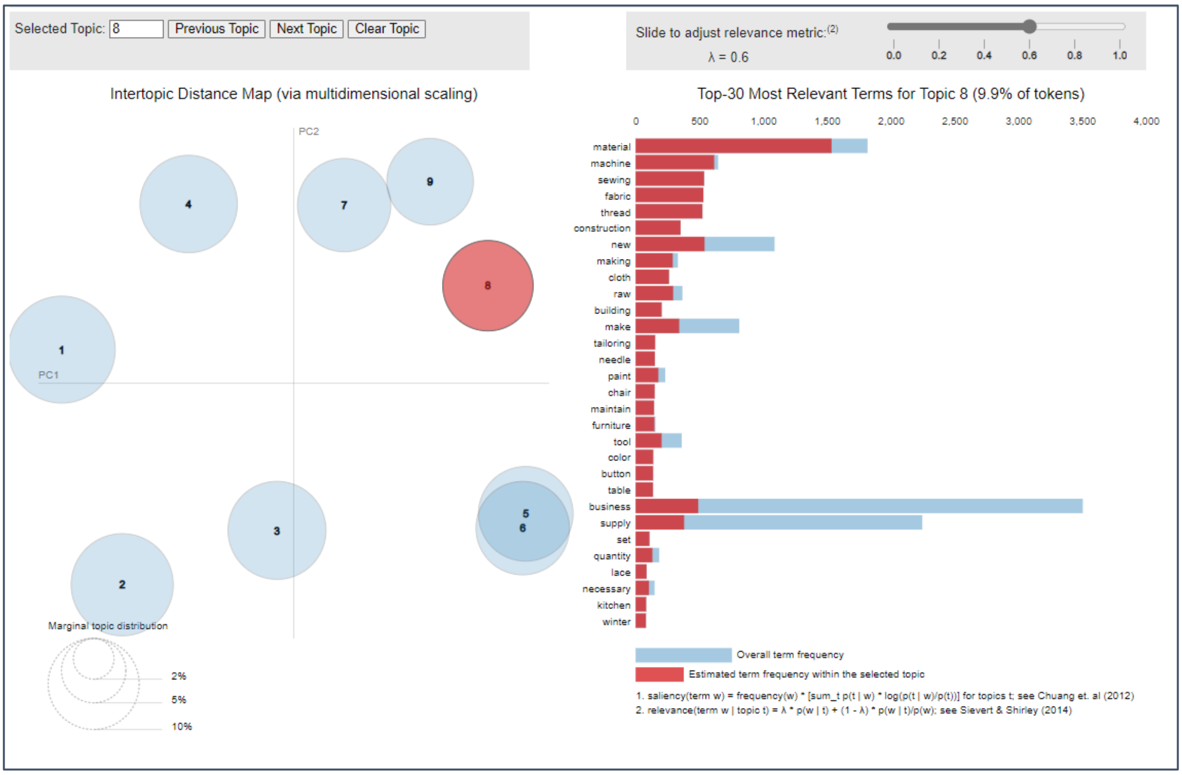


Figure C.12 – The ninth out of nine topics of the *use* variable

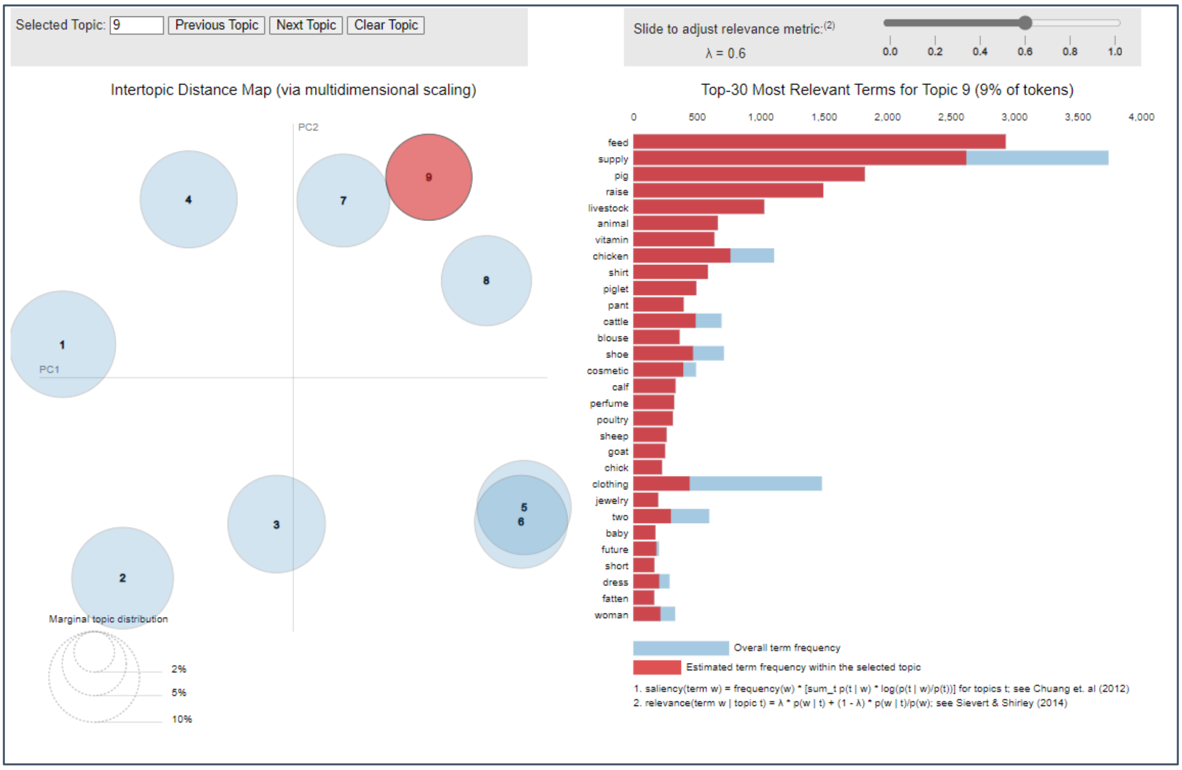


Figure C.13 – Three number of topics for the *description* variable

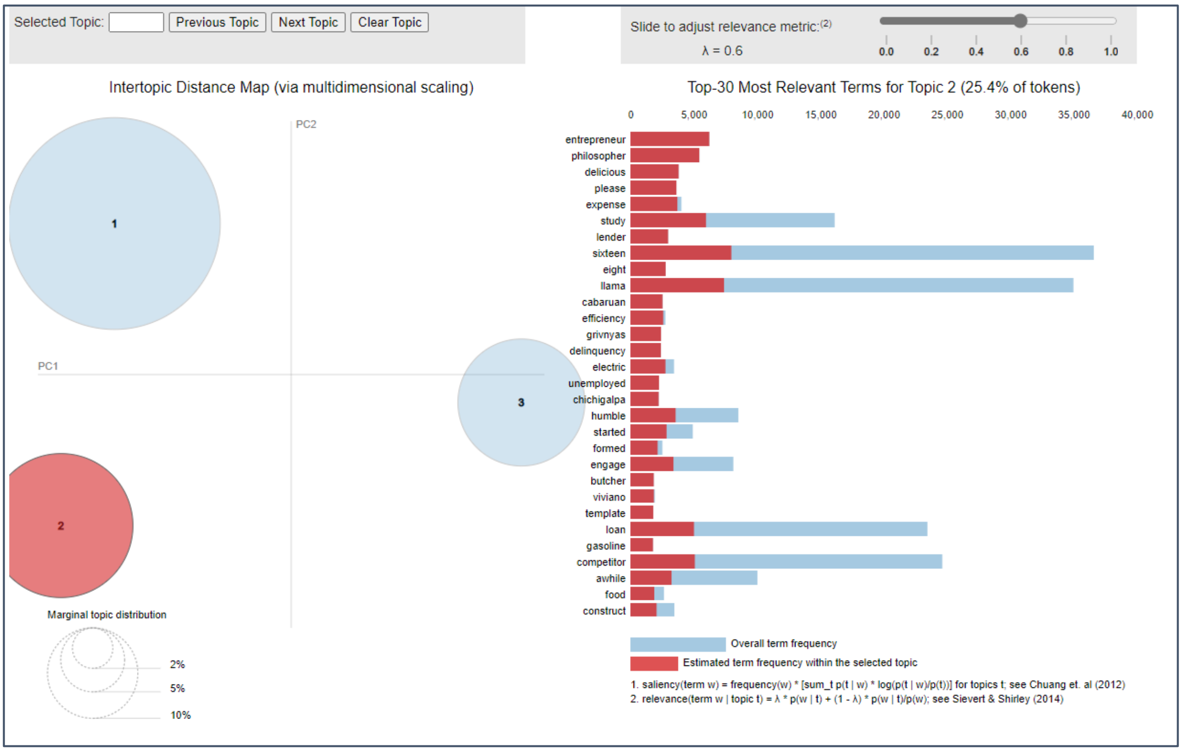


Figure C.14 – The second out of fifteen topics of the *description* variable

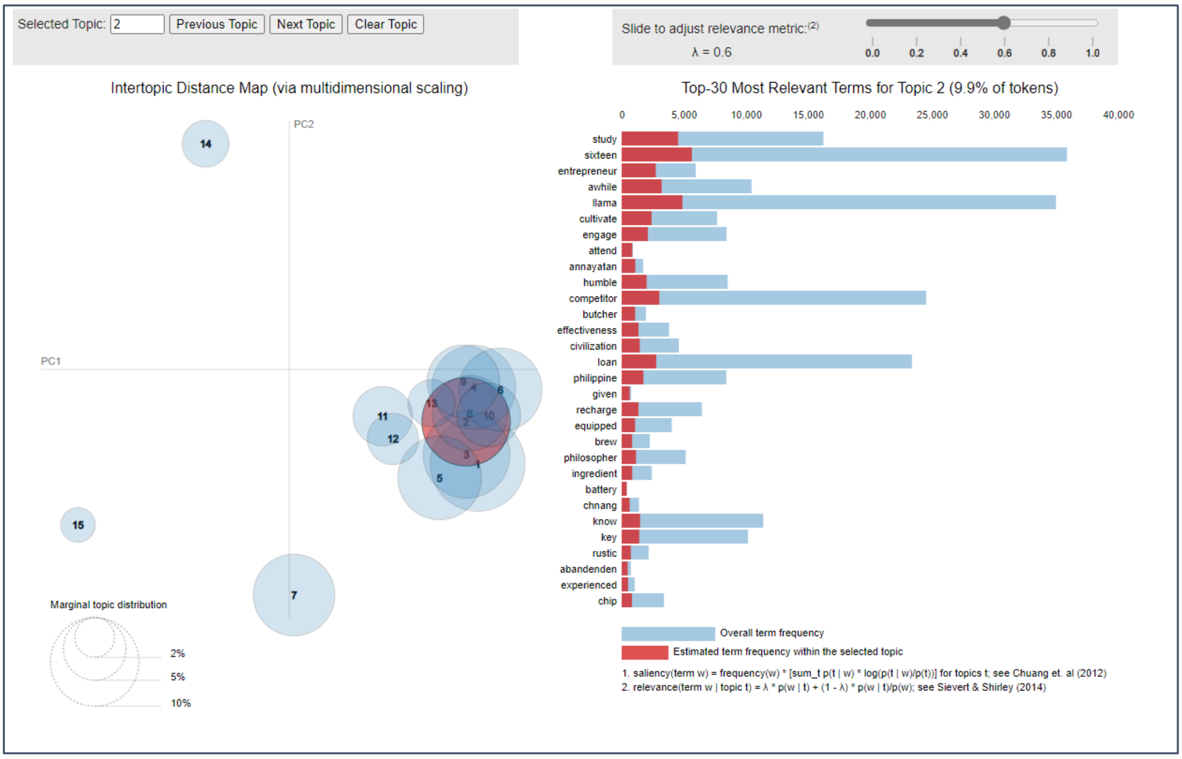


Figure C.15 – The third out of fifteen topics of the *description* variable

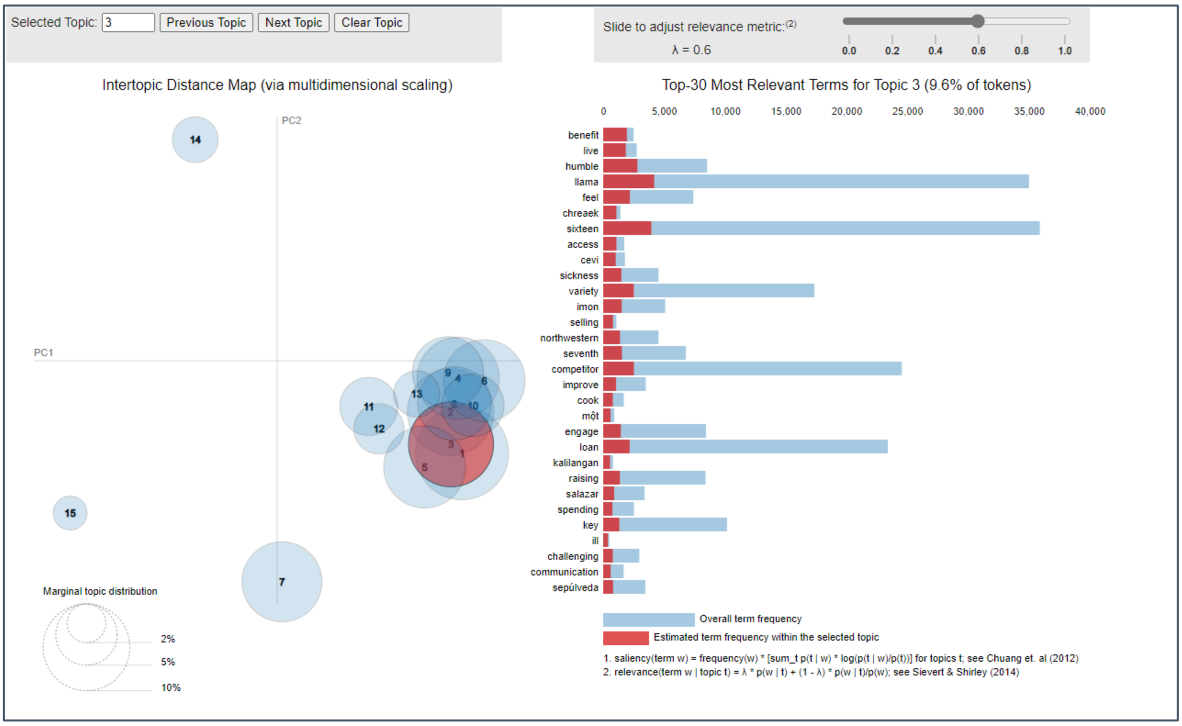


Figure C.16 – The forth out of fifteen topics of the *description* variable

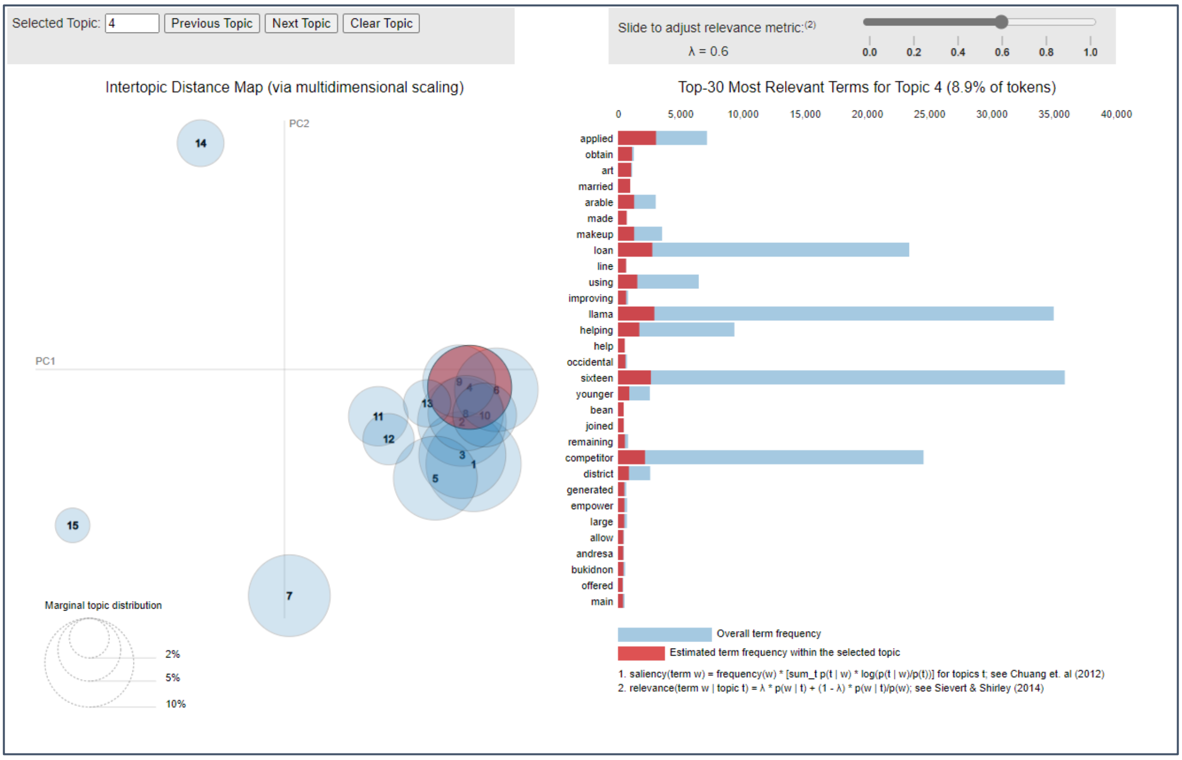


Figure C.17 – The fifth out of fifteen topics of the *description* variable

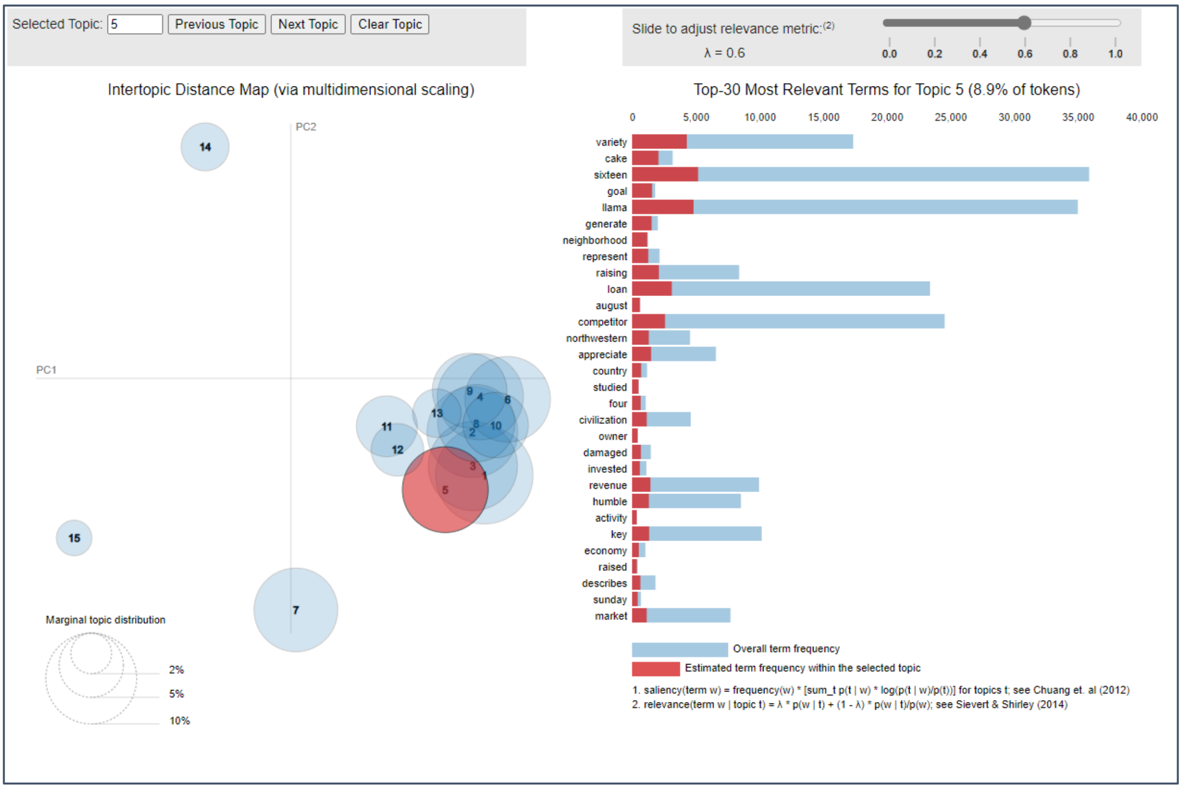


Figure C.18 – The sixth out of fifteen topics of the *description* variable

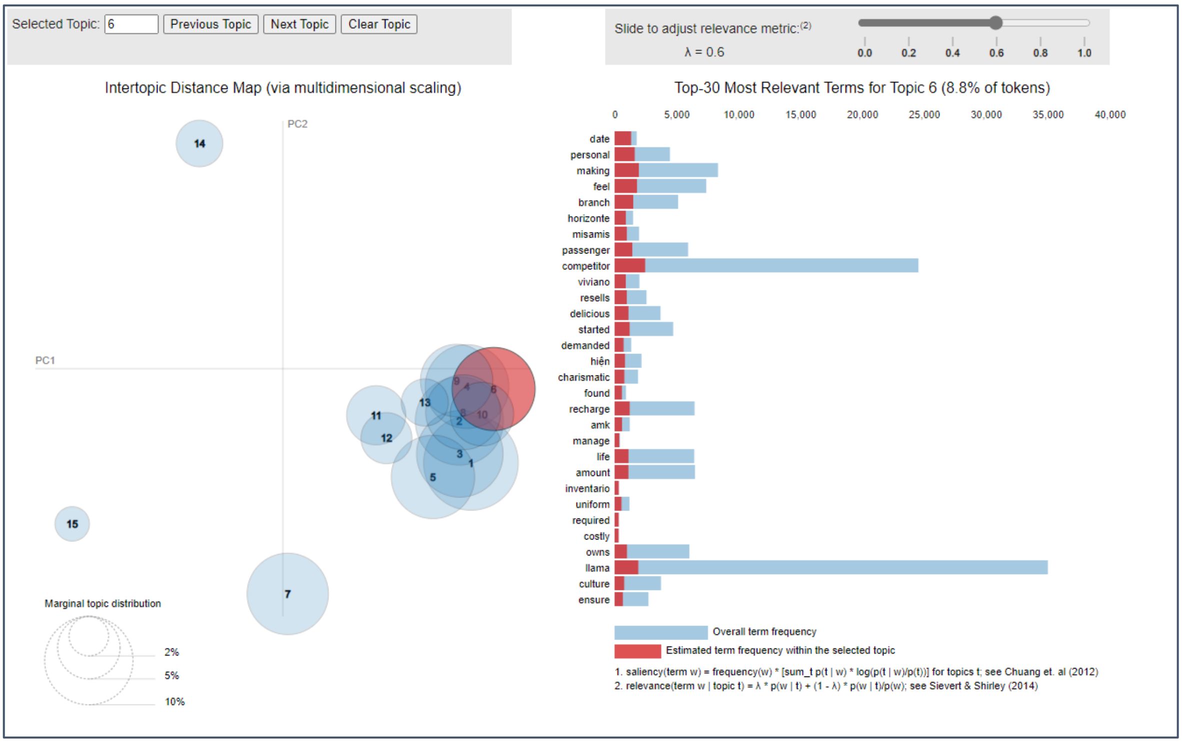


Figure C.19 – The seventh out of fifteen topics of the *description* variable

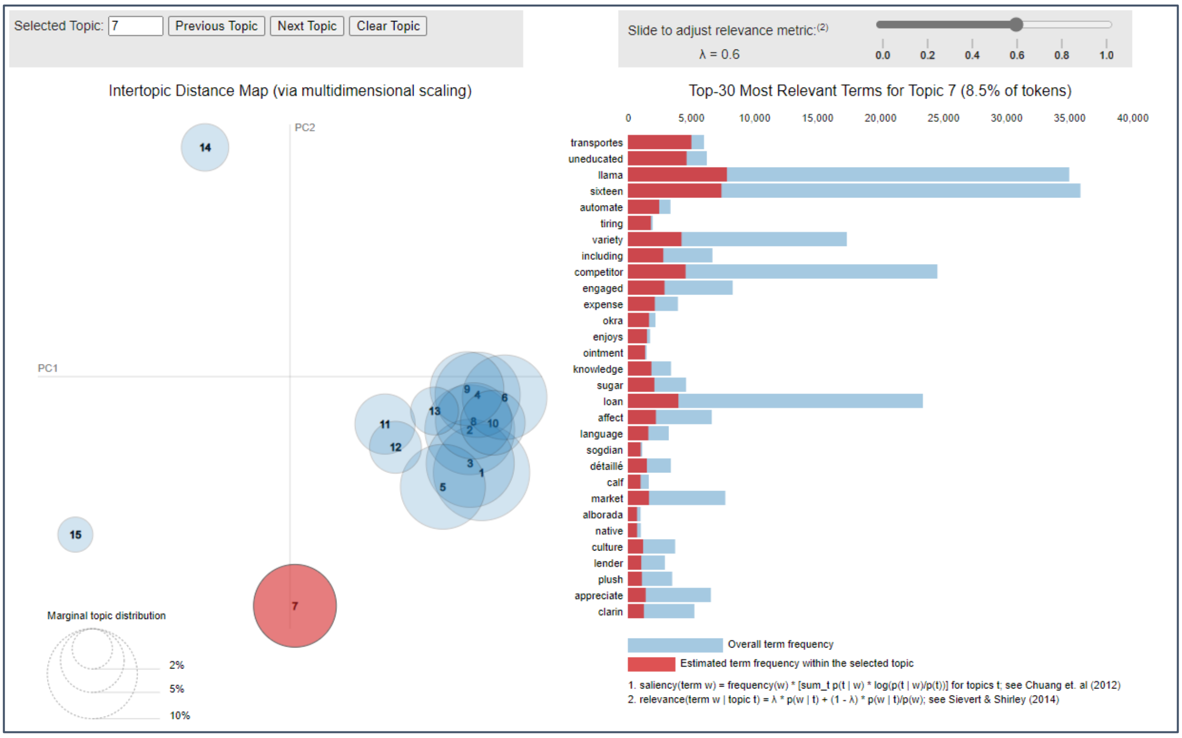


Figure C.20 – The eight out of fifteen topics of the *description* variable

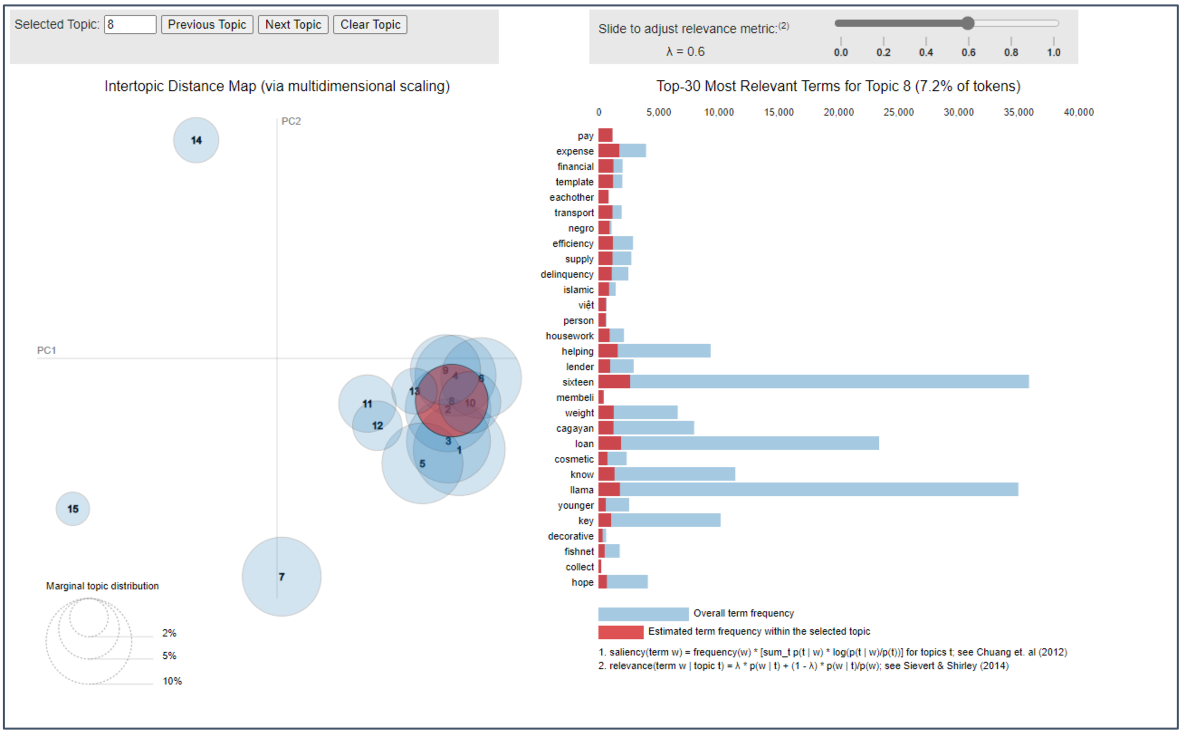


Figure C.21 – The ninth out of fifteen topics of the *description* variable

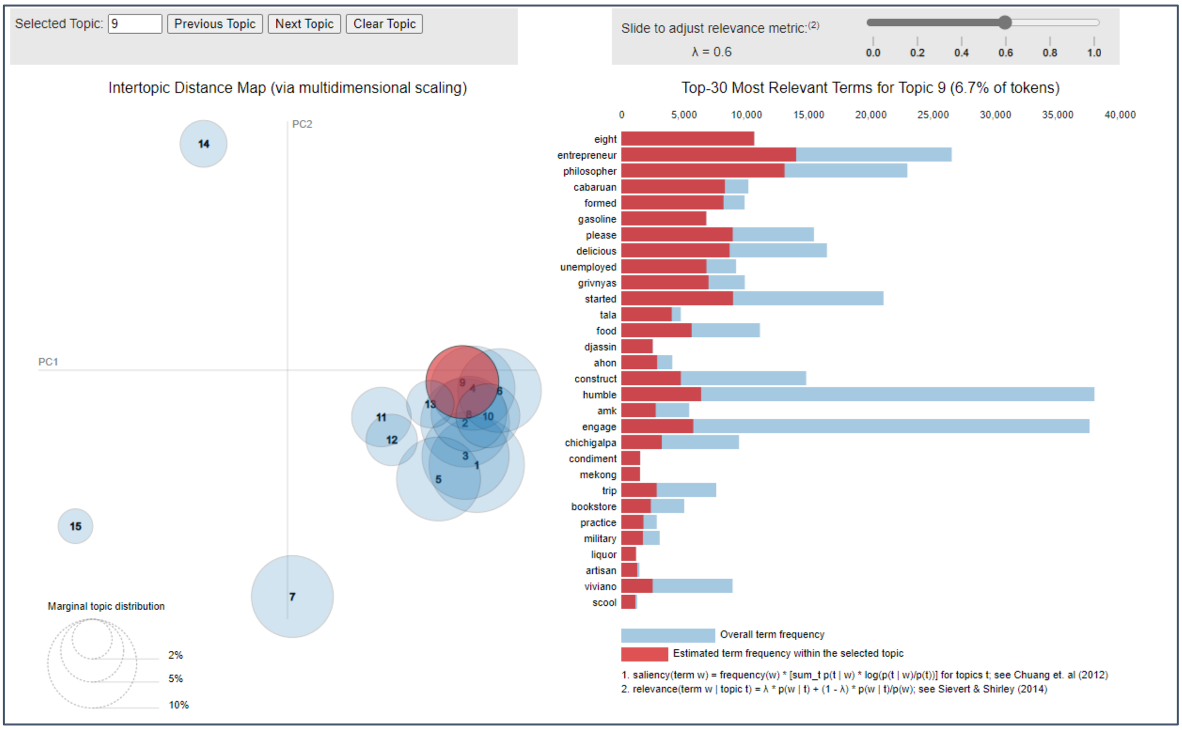


Figure C.22 – The tenth out of fifteen topics of the *description* variable

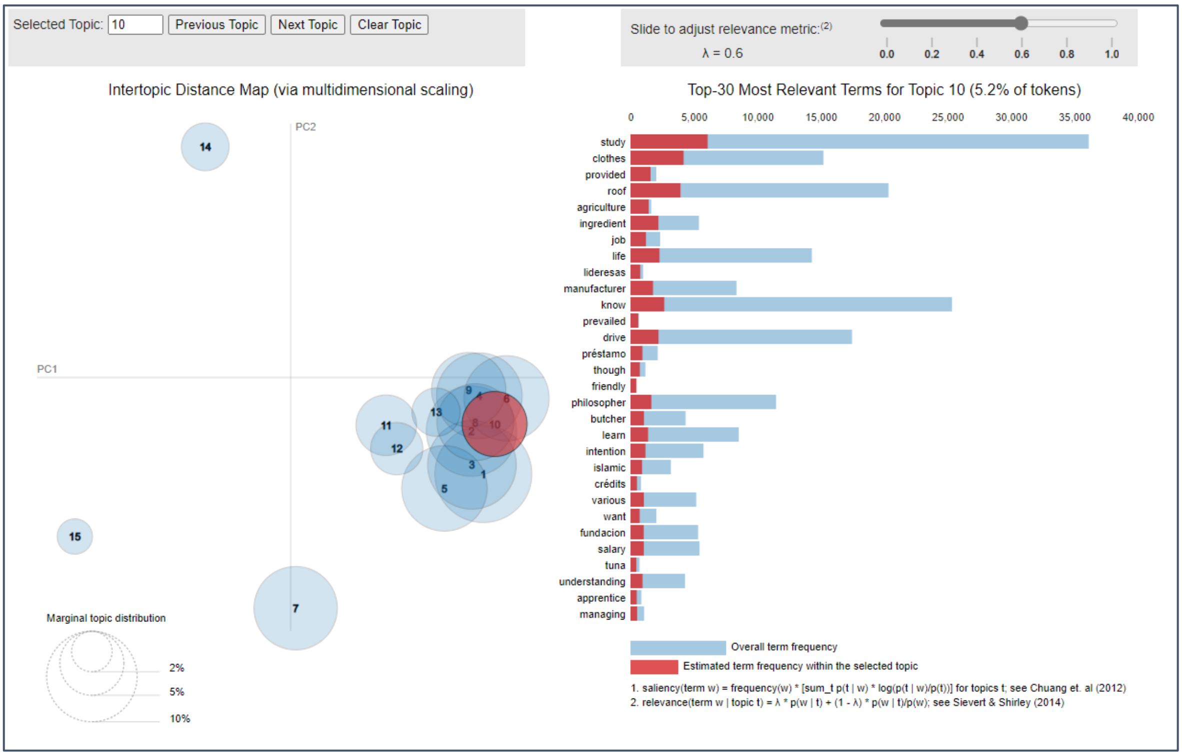


Figure C.23 – The eleventh out of fifteen topics of the *description* variable

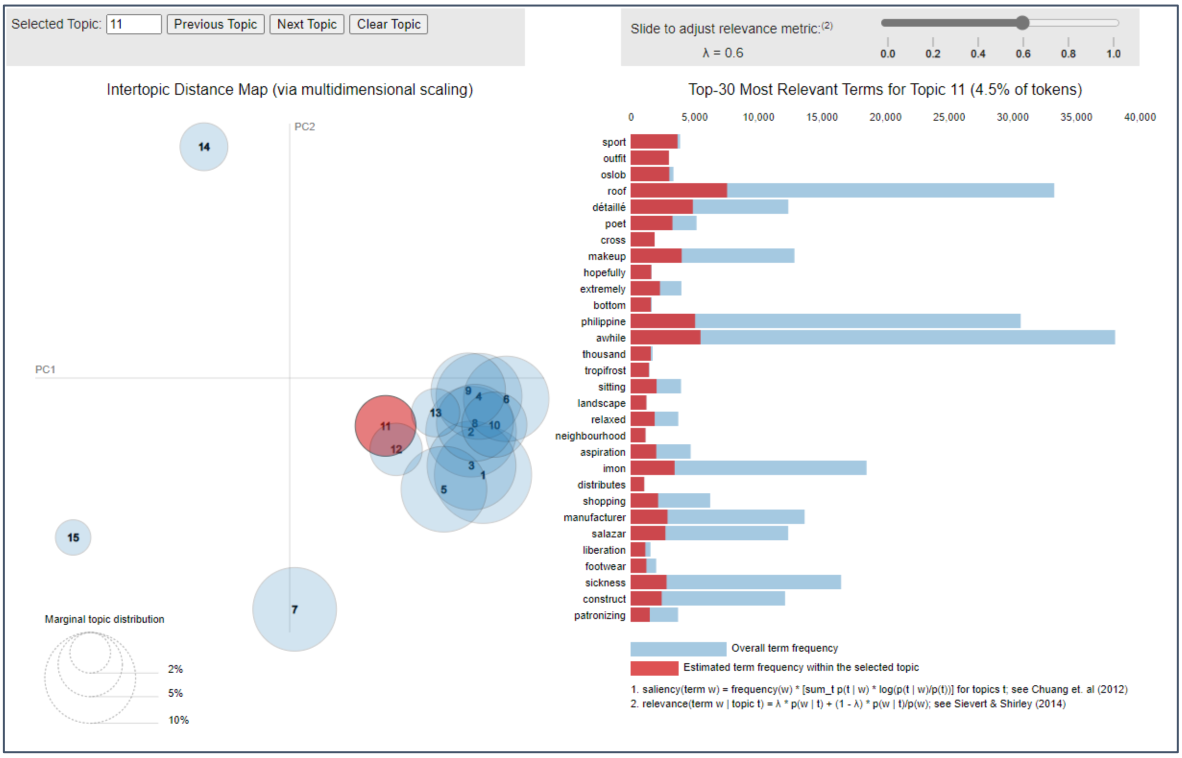


Figure C.24 – The twelfth out of fifteen topics of the *description* variable

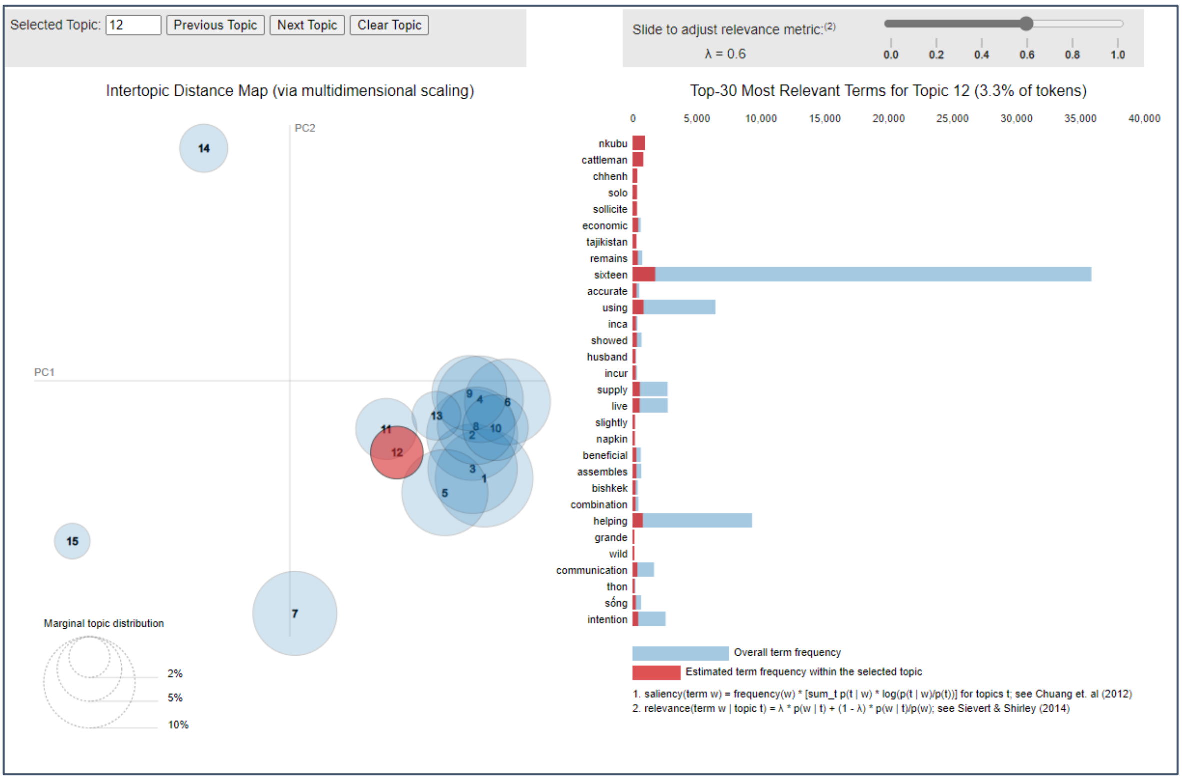


Figure C.25 – The thirteenth out of fifteen topics of the *description* variable

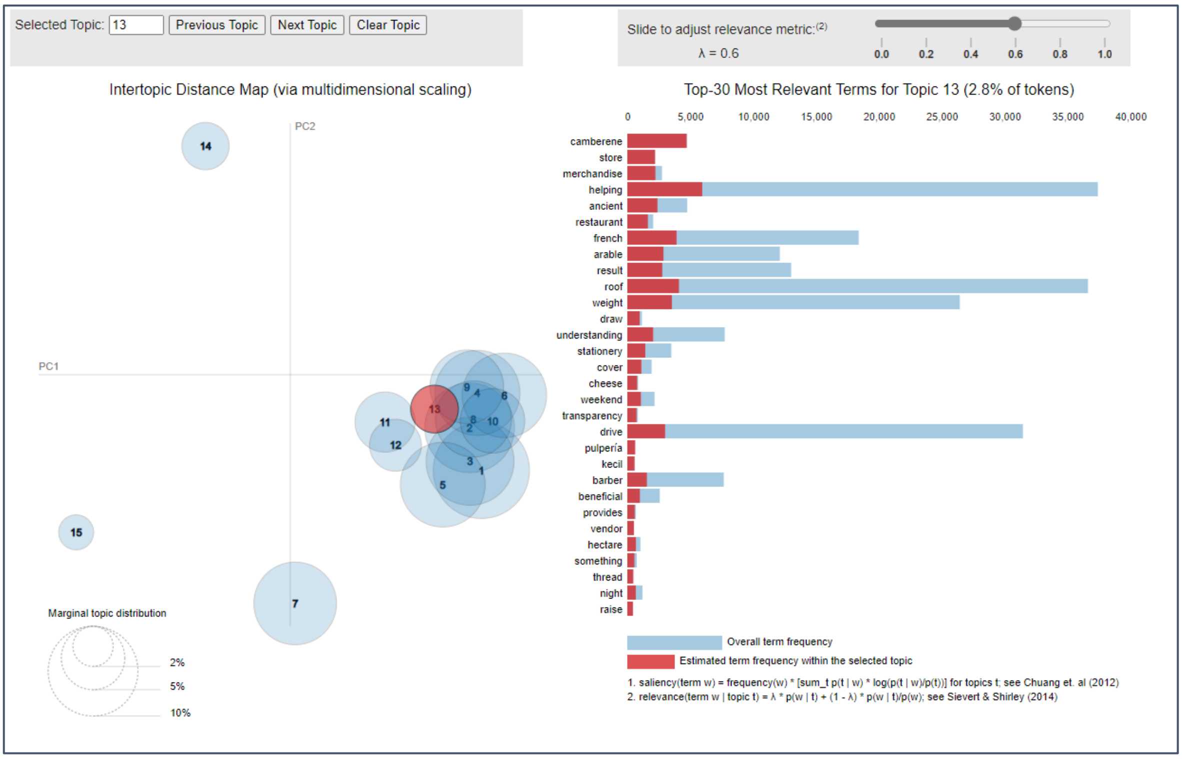


Figure C.26 – The fourteenth out of fifteen topics of the *description* variable

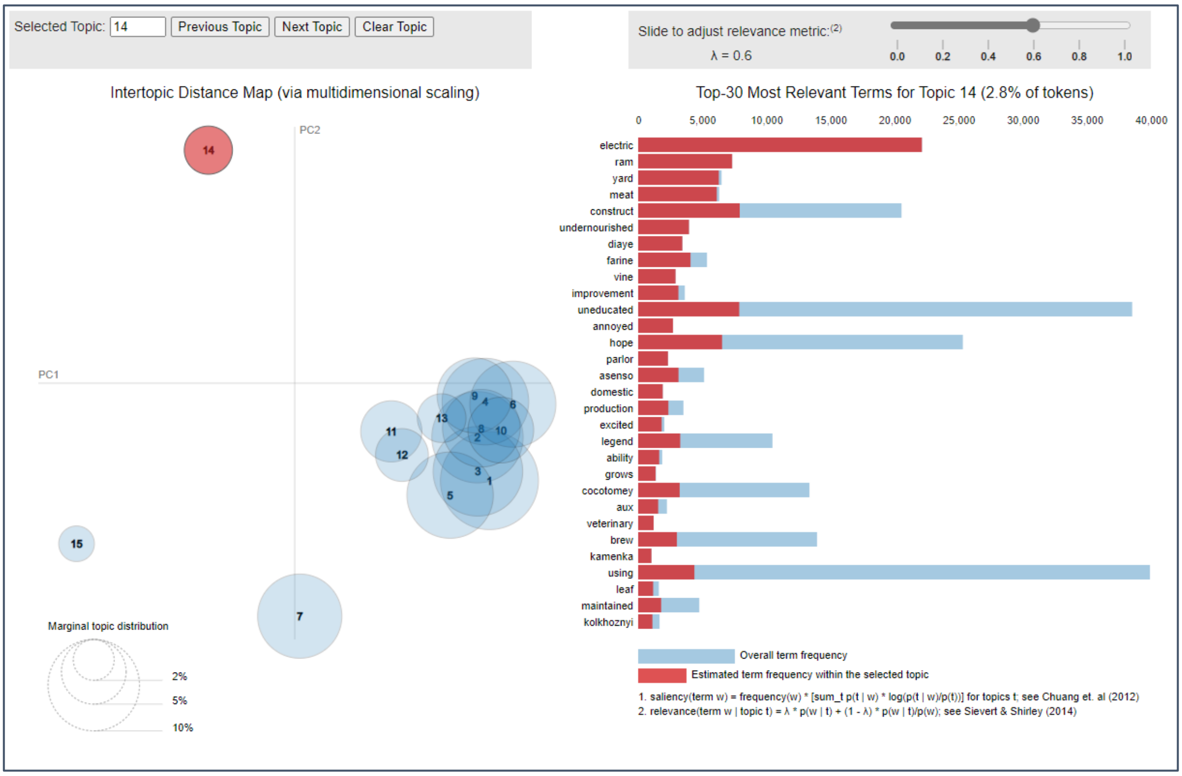


Figure C.27 – The fifteenth out of fifteen topics of the *description* variable

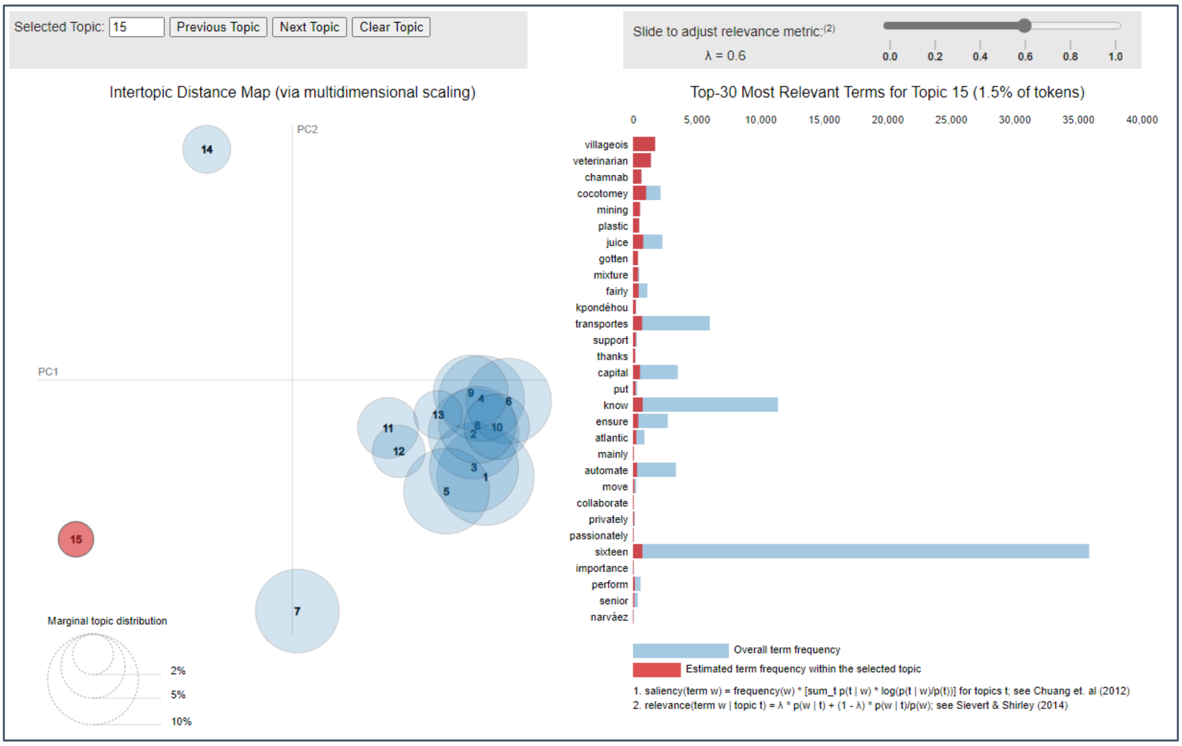


Figure C.28 – Function for the calculation of the topic contribution

```
# create a function that will calculate the topic contribution of "use" and "description" columns for each row of the smallest_data.csv dataset

#This vector will be used as a feature for the classification task in later steps
def get_topic_contribution(ldamodel, corpus, num_topics):
    vector = ldamodel[corpus]  # get topic probability distribution for a document
    topic_contribution = np.zeros((len(vector), num_topics))
    for i in range(len(vector)):
        for j in range(len(vector[i])):
            try:
                topic_contribution[i,vector[i][j][0]] = vector[i][j][1]
            except:
                print('error list topic out of range, row and topic', i, j)
                # print(vector)
    return topic_contribution
```

Figure C.29 – Logit Model with *status*, 9 topics of *use*, 3 topics of *description*

Logit Regression Results

Dep. Variable:	status	**No. Observations:**	36000
Model:	Logit	**Df Residuals:**	35987
Method:	MLE	**Df Model:**	12
Date:	Sun, 22 Jan 2023	**Pseudo R-squ.:**	0.02353
Time:	19:22:15	**Log-Likelihood:**	-6067.2
converged:	True	**LL-Null:**	-6213.4
Covariance Type:	nonrobust	**LLR p-value:**	1.794e-55

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
const	1.8940	2.426	0.781	0.435	-2.860	6.648
topic_contribution_use_0	1.2535	0.846	1.482	0.138	-0.405	2.912
topic_contribution_use_1	1.8391	0.845	2.175	0.030	0.182	3.496
topic_contribution_use_2	2.1194	0.851	2.491	0.013	0.452	3.787
topic_contribution_use_3	2.3037	0.852	2.703	0.007	0.633	3.974
topic_contribution_use_4	1.5448	0.859	1.799	0.072	-0.138	3.228
topic_contribution_use_5	2.1557	0.849	2.539	0.011	0.491	3.820
topic_contribution_use_6	2.2344	0.850	2.628	0.009	0.568	3.901
topic_contribution_use_7	1.0786	0.849	1.271	0.204	-0.584	2.742
topic_contribution_use_8	1.4354	0.848	1.694	0.090	-0.226	3.097
topic_contribution_descr_0	-0.9746	2.301	-0.423	0.672	-5.485	3.536
topic_contribution_descr_1	0.7978	2.289	0.349	0.727	-3.689	5.284
topic_contribution_descr_2	-0.3601	2.297	-0.157	0.875	-4.863	4.143

Figure C.30 – Logit Model with *status*, 9 topics of *use*, 11 topics of *description*

Logit Regression Results

Dep. Variable:	status	**No. Observations:**	36000
Model:	Logit	**Df Residuals:**	35979
Method:	MLE	**Df Model:**	20
Date:	Fri, 27 Jan 2023	**Pseudo R-squ.:**	0.02767
Time:	11:47:29	**Log-Likelihood:**	-6041.5
converged:	True	**LL-Null:**	-6213.4
Covariance Type:	nonrobust	**LLR p-value:**	8.447e-61

	coef	std err	z	P>|z|	[0.025	0.975]
const	1.7942	0.885	2.029	0.043	0.061	3.528
topic_contribution_use_0	1.3551	0.783	1.731	0.083	-0.179	2.889
topic_contribution_use_1	1.4251	0.784	1.817	0.069	-0.112	2.962
topic_contribution_use_2	1.9170	0.791	2.422	0.015	0.366	3.468
topic_contribution_use_3	1.7160	0.790	2.173	0.030	0.168	3.264
topic_contribution_use_4	1.5202	0.795	1.912	0.056	-0.038	3.079
topic_contribution_use_5	1.7819	0.790	2.255	0.024	0.233	3.330
topic_contribution_use_6	1.8209	0.789	2.307	0.021	0.274	3.368
topic_contribution_use_7	1.3005	0.786	1.655	0.098	-0.240	2.841
topic_contribution_use_8	0.7557	0.788	0.959	0.337	-0.788	2.300
topic_contribution_descr_0	-0.8349	0.453	-1.844	0.065	-1.722	0.053
topic_contribution_descr_1	1.6687	0.656	2.542	0.011	0.382	2.955
topic_contribution_descr_2	2.5064	0.557	4.500	0.000	1.415	3.598
topic_contribution_descr_3	0.0325	0.466	0.070	0.944	-0.881	0.946
topic_contribution_descr_4	0.2004	0.485	0.413	0.679	-0.750	1.151
topic_contribution_descr_5	-0.2117	0.522	-0.406	0.685	-1.234	0.811
topic_contribution_descr_6	-0.7476	0.446	-1.675	0.094	-1.622	0.127
topic_contribution_descr_7	-0.3779	0.474	-0.797	0.426	-1.308	0.552
topic_contribution_descr_8	-0.4082	0.457	-0.892	0.372	-1.305	0.488
topic_contribution_descr_9	4.3215	0.829	5.211	0.000	2.696	5.947
topic_contribution_descr_10	-0.5538	0.475	-1.166	0.244	-1.485	0.377

Figure C.31 – Logit Model with *status*, 9 topics of *use*, 17 topics of *description* (first part)

Logit Regression Results

Dep. Variable:	status	**No. Observations:**	36000
Model:	Logit	**Df Residuals:**	35973
Method:	MLE	**Df Model:**	26
Date:	Sun, 22 Jan 2023	**Pseudo R-squ.:**	0.04805
Time:	20:23:41	**Log-Likelihood:**	-5914.8
converged:	True	**LL-Null:**	-6213.4
Covariance Type:	nonrobust	**LLR p-value:**	2.343e-109

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
const	1.4023	0.791	1.773	0.076	-0.148	2.953
topic_contribution_use_0	1.1837	0.752	1.574	0.115	-0.290	2.658
topic_contribution_use_1	1.6325	0.755	2.163	0.031	0.153	3.112
topic_contribution_use_2	1.5945	0.762	2.094	0.036	0.102	3.087
topic_contribution_use_3	1.8931	0.757	2.501	0.012	0.409	3.377
topic_contribution_use_4	1.5665	0.764	2.050	0.040	0.069	3.064
topic_contribution_use_5	1.6346	0.760	2.152	0.031	0.146	3.123
topic_contribution_use_6	1.8734	0.757	2.476	0.013	0.391	3.356
topic_contribution_use_7	1.1274	0.756	1.490	0.136	-0.355	2.610
topic_contribution_use_8	0.7007	0.756	0.927	0.354	-0.780	2.182
topic_contribution_descr_0	-0.0569	0.438	-0.130	0.897	-0.915	0.801
topic_contribution_descr_1	-0.2503	0.311	-0.804	0.421	-0.861	0.360
topic_contribution_descr_2	0.4002	0.331	1.207	0.227	-0.249	1.050
topic_contribution_descr_3	1.9685	0.334	5.891	0.000	1.314	2.623
topic_contribution_descr_4	2.8498	0.548	5.200	0.000	1.776	3.924

Figure C.32 – Logit Model with *status*, 9 topics of *use*, 17 topics of *description* (second part)

topic_contribution_descr_5	-0.1497	0.327	-0.457	0.647	-0.791	0.492
topic_contribution_descr_6	0.9246	0.327	2.823	0.005	0.283	1.566
topic_contribution_descr_7	6.7742	1.154	5.871	0.000	4.513	9.036
topic_contribution_descr_8	-0.0238	0.316	-0.075	0.940	-0.643	0.595
topic_contribution_descr_9	-0.6131	0.336	-1.827	0.068	-1.271	0.045
topic_contribution_descr_10	0.6537	0.412	1.585	0.113	-0.155	1.462
topic_contribution_descr_11	-0.0504	0.320	-0.158	0.875	-0.677	0.576
topic_contribution_descr_12	-0.3756	0.317	-1.184	0.236	-0.997	0.246
topic_contribution_descr_13	0.1038	0.320	0.324	0.746	-0.524	0.732
topic_contribution_descr_14	3.6655	0.982	3.734	0.000	1.742	5.589
topic_contribution_descr_15	-0.1092	0.302	-0.361	0.718	-0.702	0.484
topic_contribution_descr_16	-0.7656	0.369	-2.073	0.038	-1.489	-0.042

Figure C.33 – Logit Model with *status*, 9 topics of *use*, 23 topics of *description* (first part)

Logit Regression Results

Dep. Variable:	status	**No. Observations:**	36000
Model:	Logit	**Df Residuals:**	35967
Method:	MLE	**Df Model:**	32
Date:	Sun, 22 Jan 2023	**Pseudo R-squ.:**	0.04481
Time:	19:59:07	**Log-Likelihood:**	-5935.0
converged:	True	**LL-Null:**	-6213.4
Covariance Type:	nonrobust	**LLR p-value:**	4.681e-97

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
const	1.9594	0.845	2.318	0.020	0.303	3.616
topic_contribution_use_0	0.7801	0.814	0.959	0.338	-0.814	2.375
topic_contribution_use_1	1.1892	0.816	1.457	0.145	-0.411	2.789
topic_contribution_use_2	1.3935	0.819	1.702	0.089	-0.212	2.999
topic_contribution_use_3	1.6372	0.820	1.997	0.046	0.031	3.244
topic_contribution_use_4	1.2574	0.824	1.527	0.127	-0.357	2.871
topic_contribution_use_5	1.4945	0.819	1.824	0.068	-0.111	3.100
topic_contribution_use_6	1.6133	0.818	1.971	0.049	0.009	3.217
topic_contribution_use_7	0.6810	0.816	0.834	0.404	-0.919	2.281
topic_contribution_use_8	0.3729	0.816	0.457	0.648	-1.227	1.973
topic_contribution_descr_0	0.6884	0.372	1.851	0.064	-0.040	1.417
topic_contribution_descr_1	-0.2252	0.400	-0.563	0.573	-1.009	0.559
topic_contribution_descr_2	-0.1774	0.295	-0.602	0.547	-0.755	0.400
topic_contribution_descr_3	-0.6106	0.315	-1.937	0.053	-1.228	0.007
topic_contribution_descr_4	2.0332	0.522	3.898	0.000	1.011	3.056
topic_contribution_descr_5	-0.4261	0.321	-1.327	0.184	-1.055	0.203
topic_contribution_descr_6	0.6902	0.384	1.796	0.073	-0.063	1.443
topic_contribution_descr_7	4.5069	0.894	5.039	0.000	2.754	6.260
topic_contribution_descr_8	-0.6261	0.360	-1.740	0.082	-1.331	0.079
topic_contribution_descr_9	-1.0460	0.318	-3.290	0.001	-1.669	-0.423

Figure C.34 – Logit Model with *status*, 9 topics of *use*, 23 topics of *description* (second part)

topic_contribution_descr_10	0.3353	1.242	0.270	0.787	-2.099	2.770
topic_contribution_descr_11	0.9570	0.427	2.242	0.025	0.120	1.794
topic_contribution_descr_12	-0.7882	0.307	-2.571	0.010	-1.389	-0.187
topic_contribution_descr_13	-0.3154	0.314	-1.004	0.316	-0.931	0.301
topic_contribution_descr_14	0.0494	0.465	0.106	0.915	-0.862	0.961
topic_contribution_descr_15	-0.2212	0.398	-0.556	0.578	-1.001	0.558
topic_contribution_descr_16	1.1538	0.502	2.299	0.021	0.170	2.137
topic_contribution_descr_17	-0.7783	0.340	-2.289	0.022	-1.445	-0.112
topic_contribution_descr_18	-0.1030	0.325	-0.317	0.751	-0.739	0.533
topic_contribution_descr_19	0.6145	0.363	1.691	0.091	-0.098	1.327
topic_contribution_descr_20	-0.0828	0.309	-0.268	0.789	-0.688	0.522
topic_contribution_descr_21	1.6646	0.502	3.314	0.001	0.680	2.649
topic_contribution_descr_22	1.8313	0.325	5.629	0.000	1.194	2.469

D – Data Modeling

Figure D.1 – Example of get dummies for *SectorName* and *Region*

```
#apply get dummies to categorical variables

df_sector_region = pd.concat([df_sector_region.drop("SectorName", axis = 1),
                pd.get_dummies(df_sector_region["SectorName"], prefix = "SectorName", drop_first=False)], axis=1, sort=False)

df_sector_region = pd.concat([df_sector_region.drop("Region", axis = 1),
                pd.get_dummies(df_sector_region["Region"], prefix = "Region", drop_first=False)], axis=1, sort=False)
```

Figure D.2 – Example of data normalization

```
#normalize all quantitative variables
max_valuef = df_residual_var_norm['fundedAmount'].max()
min_valuef = df_residual_var_norm['fundedAmount'].min()
df_residual_var_norm['fundedAmount'] = (df_residual_var_norm['fundedAmount'] - min_valuef) / (max_valuef - min_valuef)

max_valuel = df_residual_var_norm['loanAmount'].max()
min_valuel = df_residual_var_norm['loanAmount'].min()
df_residual_var_norm['loanAmount'] = (df_residual_var_norm['loanAmount'] - min_valuel) / (max_valuel - min_valuel)

max_valuele = df_residual_var_norm['lendersTotalCount'].max()
min_valuele = df_residual_var_norm['lendersTotalCount'].min()
df_residual_var_norm['lendersTotalCount'] = (df_residual_var_norm['lendersTotalCount'] - min_valuele) / (max_valuele - min_valuele)

max_valuet = df_residual_var_norm['teamsTotalCount'].max()
min_valuet = df_residual_var_norm['teamsTotalCount'].min()
df_residual_var_norm['teamsTotalCount'] = (df_residual_var_norm['teamsTotalCount'] - min_valuet ) / (max_valuet - min_valuet)

max_valueb = df_residual_var_norm['borrowerCount'].max()
min_valueb = df_residual_var_norm['borrowerCount'].min()
df_residual_var_norm['borrowerCount'] = (df_residual_var_norm['borrowerCount'] - min_valueb ) / (max_valueb - min_valueb)

max_valuelr = df_residual_var_norm['lenderRepaymentTerm'].max()
min_valuelr = df_residual_var_norm['lenderRepaymentTerm'].min()
df_residual_var_norm['lenderRepaymentTerm'] = (df_residual_var_norm['lenderRepaymentTerm'] - min_valuelr ) / (max_valuelr - min_valuelr)

max_valuefd = df_residual_var_norm['funding_days'].max()
min_valuefd = df_residual_var_norm['funding_days'].min()
df_residual_var_norm['funding_days'] = (df_residual_var_norm['funding_days'] - min_valuefd ) / (max_valuefd - min_valuefd)
```

Figure D.3 – Example of the application of the variance inflation factor

```
df_vif_18000_rvn
```

	Variable	VIF
0	purchase	65.378332
1	farming	54.131091
2	investment	51.420204
3	trading	42.694952
4	family_expenses	44.696754
5	family_needs	56.200968
6	business	48.281006
7	improvement	64.855128
8	personal_activity	50.194848
9	face_difficulties	6.699242
10	entrepreneurship	3.510154
11	causes	8.130395
12	asking_for_loan	15.838976
13	retail	3.245870
14	progress	6.531245
15	borrowers_situation	7.583758
16	women_reasons	4.086683
17	financial_resources	7.869348
18	job	3.764001
19	procurement	2.394767
20	growth	5.481506
21	local_activity	7.356121
22	poverty	6.248133
23	loyal_borrowers	2.704423
24	fundedAmount	209.517652
25	loanAmount	143.174726

List of Figures

23. First part of Logit Model of Sector, Region and Topic Variables (small dataset)

24. Second part of Logit Model of Sector, Region and Topic Variables (small dataset)

25. First part of Logit Model of Topic and Partners' Variables (full dataset)

26. Second part of Logit Model of Topic and Partners' Variables (small dataset)

27. First part of Logit Model of Topic and Partners' Variables (full dataset)

28. Second part of Logit Model of Topic and Partners' Variables (small dataset)

29. First part of Logit Model of Topic and Residual Variables (full dataset)

30. Second part of Logit Model of Topic and Residual Variables (full dataset)

31. First part of Logit Model of Topic and Residual Variables (small dataset)

32. Second part of Logit Model of Topic and Residual Variables (small dataset)

33. Import libraries and read datasets in csv-format files

34. Check null values in all variables and delete two columns

35. Remove every observations where id is not an integer datatype

36. Substitute null values

37. Transform into date type

38. Transform into integers

39. Number of columns, rows and their data type

40. Principal indicators of the dataset

41. Delete time values in plannedExpirationDate

42. Outer join between the datasets with use and description variables

43. Inner join between the dataset with loan information and the ones of activity, sector, country and partner

44. Creation of a single dataset called small_data

45. Realization of the Line Plot for the "Distribution of Crucial Rates over the Years"

Bibliography

Abdul-Rahman A Hailes S(2000), *Supporting Trust in Virtual Communities,* In Proceedings of the Hawai'i International Conference on System Sciences (HICSS), Maui (HI), US

Adhikary BK, Kutsuna K, Hoda T (2018), *Crowdfunding-types and models*. *In: Crowdfunding*, SpringerBriefs in Economics(), Springer, Singapore, 9 – 20

Allison TH, Davis BC, Short JC, Webb JW (2015), *Crowdfunding in a prosocial microlending environment: Examining the role of intrinsic versus extrinsic cues*, Entrepreneurship Theory and Practice 39 (1), 53–73

Almaamari OA, Vedava P (2022*), The importance of microfinance institutions in the development of developing countries*, International Journal of Economics, 10 (4) 28 – 35

André K, Bureau S, Gautier A, Rubel O (2017), *Beyond the opposition between altruism and self-interest: Reciprocal giving in reward-based crowdfunding*, Journal of Business Ethics, 146(2), 313–332

Armendáriz de Aghion B, Morduch J (2000), *Microfinance beyond group lendin*g, Economics of Transition, 8, 401–420

Armendáriz de Aghion B, Morduch J (2010*), The economics of microfinance (2nd ed.)*, Cambridge, MA: MIT Press

Bagozzi RP, Gopinath M, Nyer PU (1999), *The Role of Emotions in Marketing,* Journal of the Academy of Marketing Science, 27(2), 184–206

Barbi M, Bigelli M (2017), *Crowdfunding practices in and outside the US,* Research in International Business and Finance, 42, 208– 223.

Baucus MS, Mitteness CR (2016*), Crowdfrauding: avoiding Ponzi entrepreneurs when investing in new venture*s, Elsevier, 59 (1), 37–50

Beaulieu T, Sarker S (2015), *A conceptual framework for understanding crowdfunding*, Communications of the Association for Information Systems, 37, 1 – 31

Belleflamme P, Lambert T and Schwienbacher A (2014), *Crowdfunding: Tapping the right crowd*, Journal of Business Venturing 29 (5), 585–609

Berns JP, Figueroa-Armijos M, da Motta Veiga SP, Dunne TC (2020), *Dynamics of Lendign-Based Prosocial Crowdfunidng: Using a Social Responsibility Lens*, Journal of Business Ethics, Springer, 161, 169 – 185

Blei D (2012), *Probabilistic Topic Models*, Communications of the ACM, 55, 77–84

Bockel A, Horisch J, Tenner I (2021), *A systematic literature review of crowdfunding and sustainability: highlighting what really matters*, Management review quarterly, Springer, 71, 433 – 453

Brumm B (2019), Outer Join, *Beginning Oracle SQL for Oracle Database 18c,* Apress, Berkeley, CA., Springer, 263 - 273

Bruton G, Khavul S, Siegel D, Wright M (2015), *New financial alternatives in seeding entrepreneurship: Microfinance, crowdfunding, and peer-to-peer innovations*, Entrepreneurship Theory and Practice 39 (1), 9–26.

Burke K (1984), *Attitudes toward history*, California: University of California Press

Calic G, Mosakowski E (2016*), Kicking of social entrepreneurship: How a sustainability orientation influences crowdfunding succes*s, Journal of Management Studies, 53(5), 738–767

CFI Teams (2022), R-Squared, Corporate Financial Institute, https://corporatefinanceinstitute.com/resources/data-science/r-squared/

CFI Teams (2022), P-Value, Corporate Financial Institute, https://corporatefinanceinstitute.com/resources/data-science/p-value/

Chang J, Sean G, Chong W, Jordan BG, and David MB (2009), *Reading Tea Leaves: How Humans Interpret Topic Models,* In Advances in Neural Information Processing Systems, 288–296.

Chen Y, Li SX (2009), *Group identity and social preferences*, American Economic Review, 99 (1), 431 – 457

Cho M, Kim G (2017), *A cross-cultural comparative analysis of crowdfunding projects in the United States and South Korea*, Elsevier, 72, 312–320

Choi, DY, Gray ER (2008), *The venture development processes of "sustainable" entrepreneurs*, Management Research News, 31 (8), 558–569

Cicchiello A. F. (2019), *Harmonizing the crowdfunding regulation in Europe: need, challenges, and risks*, Journal of Small Business & Entrepreneurship, 32 (4), 1 – 22

Crane A (2001), *Unpacking the ethical product*, Journal of Business Ethics, 30(4), 361–373

Defazio D, Franzoni C, Rossi-Lamastra C (2021*), How Pro-social Framing Affects the Success of Crowdfunding Projects: The Role of Emphasis and Information Crowdedness*, Journal of Business Ethics, Springer, 171, 357 – 378

D'Espallier B, Guérin I, Mersland R (2011), *Women and repayment in microfinance: A global analysis*, World Development, 39, 758–772.

Donath JS, Boyd DM (2004), *Public Displays of Connection*, Technology Journal, 22(4)

Dorfleitner G, Oswald EM (2016), *Repayment behavior in peer-to-peer microfinancing: Empirical evidence from Kiva,* Review of Financial Economics, Elsevier, 30, 45 – 59

Emran MS, Morshed AM, Stiglitz J (2011*), Microfinance and missing markets,* MPRA working paper Nr. 41451, Munich: Munich Personal RePEc Archive

Faridi MZ, Nawaz S, Bibi S (2022), *Islamic microfinance and women entrepreneurial success: mediating role of human capital*, Journal of Accounting and Finance in Emerging Economies 8, (1), 1 – 12

Gafni H, Hudon M, Pèrilleux A (2021), *Business or basic needs? The impact of loan purpose on social crowdfunding platforms,* Journal of Business Ethics, Springer, 173, 777 – 793

Galuszka P, Brzozowska B (2017), *Crowdfunding: towards a redefinition of the artist's role – the case of MegaTotal*, International Journal of Cultural Studies, 20 (1), 83–99

Gan J, Qi Y (2021), *Selection for the optimal number of topics for LDA topic model – taking patent policy analysis as an example*, Entropy, 23 (10)

Gerber EM, Hui JS, Kuo PY (2012), *Crowdfunding: Why people are motivated to post and fund projects on crowdfunding platforms*, In Proceedings of the International Workshop on Design, Influence, and Social Technologies: Techniques, Impacts and Ethics.

Giorgi S, Weber K (2015), *Marks of distinction*, Administrative Science Quarterly, 60(2), 333–367

Gleasure R and Feller J (2016a), *A rift in the ground: Theorizing the evolution of anchor values in crowdfunding communities through the oculus rift case study*, Journal of the Association for Information Systems 17 (10), 708–736

Glenberg AM, Langston WE (1992), *Comprehension of Illustrated Text: Pictures Help to Build Mental Models,* Journal of Memory and Language, 31(2), 129–151.

Gulati R, Higgins MC (2003), *Which Ties Matter When? The Contingent Effects of Interorganizational Partnerships on IPO Success,* Strategic Management Journal, 24(2), 127–144.

Gurdiel Pinto L, Morales Mediano J, Cifuentes Quintero JA (2021), *A comparison study between coherence and perplexity for determining the number of topics in practitioners interview analysis,* IV Congreso Iberoamericano AJICEDE

Haas P, Blohm I, Leimeister JM (2014), *An Empirical Taxonomy of Crowdfunding Intermediaries,* International Conferences of Information Systems 2014, Auckland, New Zealand

Haidt J and Graham J (2007), *When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize*, Social Justice Research 20 (1), 98–116.

Hansen MT, Haas M (2001), *Competing for attention in knowledge markets: Electronic document dissemination in a management consulting company*, Administrative Science Quarterly, 46(1), 1–28

Harroch R (2017), *How To Create A Great Investor Pitch Deck For Startups Seeking Financing*, Forbes

Harroch R D, Sullivan M (2019), *Startup Financing: 5 Key Funding Options For Your Company*, Forbes

Heller LR and Badding KD (2012), *For compassion or money? The factors influencing the funding of micro loans,* The Journal of Socio-Economics 41 (6), 831–835

Herzenstein M, Dholakia U, Andrews R (2011), *Strategic herding behavior in peerto-peer loan auctions*, Journal of Interactive Marketing, 25, 27–36.

Hoff K, Stiglitz J (1990), *Introduction: Imperfect information and rural credit markets – puzzles and policy perspectives*, World Bank Economic Review, 4, 235–250

Hooghiemstra S. N. (2020), *The European Crowdfunding Regulation – Towards harmonization of (Equity- and lending-based) Crowdfunding in Europe?,* SSRN

Hörisch J (2015), *Crowdfunding for environmental ventures: an empirical analysis of the influence of environmental orientation on the success of crowdfunding initiatives*, Journal of Cleaner Production, 107, 636–645

Huang L and Pearce JL (2015), *Managing the unknowable: The effectiveness of early-stage investor gut feel in entrepreneurial investment decisions*, Administrative Science Quarterly 60 (4), 634–670

Iyer R, Khwaja AI, Luttmer EFP, Shue K (2009), *Screening in new credit markets: Can individual lenders infer borrower creditworthiness in peer-to-peer lending?*, NBER Working Paper Series

Jacobi C, Van Atteveldt, Welbers K (2018), *Quantitative analysis of large amounts of journalistic texts using topic modelling*, Rethinking Research Methods in an Age of Digital Journalism

Jahansoozi J. (2006), *Organization-Stakeholder Relationships: Exploring Trust and Transparency,* Journal of Management Development, 25(10), 942–955

Jancenelle VE, Javalgi RRG (2018), *The effect of moral foundations in prosocial crowdfunding*, International Small Business Journal: Researching Entrepreneurship, 36 (8), 932 – 951

Jeong J, Kim J, Son H, Nam D (2020), *The role of venture capital investment in startup' sustainable growth and performance: Focusing on absorptive capacity and venture capitalists' reputation*, Sustainability, 12, 8

Jiang Z, Benbasat I (2007), *The Effects of Presentation Formats and Task Complexity on Online Consumers' Product UnderstandinG,* MIS Quarterly, 31(3), 475–500.

Jovanovic T. (2019), *Crowdfunding: What do we know so far?,* International Journal of Innovation and Technology Management, World Scientific, 16 (1)

Kaartemo V (2017), *The elements of a successful crowdfunding campaign: A systematic literature review of crowdfunding performance*, International Review of Entrepreneurship, 15 (3)

Kagan, Julia (2018), *Microfinance*, Investopedia

Kalnins A, Willians M (2014), *When do female-owned businesses out-survive maleowned businesses? A disaggregated approach by industry and geography*, Journal of Business Venturing, 29, 822–835

Kearl M, Klammer M M (2022), *Best Crowdfunding platforms*, Investopedia

Kim L, Ju J (2019), *Can media forecast technological progress? A text-mining approach to the on-line newspaper and blog's representation of prospective industrial technologies*, Information Processing & Management, 56(4), 1506–1525

Kim PH, Bufart M, Croidieu G (2016), *TMI: Signaling credible claims in crowdfunding campaign narratives,* Group and Organization Management, 41(6), 717–750

Kiva, https://www.kiva.org/

Koch JA, Siering M (2019), *The recipe of successful crowdfunding campaigns: an analysis of crowdfunding success factors and their interrelations*, Electronic Markets, Springer, 29, 661 – 679

Laplume AO, Sonpar K, Litz RA (2008), *Stakeholder theory: Reviewing a theory that moves us,* Journal of Management, 34(6), 1152-1189

Levin IP, Schneider SL, Gaeth GJ (1998), *All frames are not created equal: A typology and critical analysis of framing effect*s, Organizational Behavior and Human Decision Processes, 76(2), 149–188

Linsley PM, Shrives PJ (2006), *Risk reporting: A study of risk disclosures in the annual reports of UK companies,* The British Accounting Review, 38(4), 387–404

Luo X, Ge L (2018), *Racial /ethnic discrimination in crowdfunding: Evidence form Kiva*, PACIS 2019 Proceedings (Pacific Asia Conference on Information Systems), 312

Lurie NH (2004), *Decision making in information-rich environments: The role of information structure*, Journal of Consumer Research, 30(4), 473–486

Macchiavello E. (2021), *The European Crowdfunding Service Providers Regulation: The Future of Marketplace Lending and Investing in Europe and the 'Crowdfunding Nature' Dilemma,* European Business Law Review, 32 (3), 557 – 608

Maithya EM, Nderu L, Njagi D (2022), *A multilevel clustering model for coherent topic discovery in short texts*, IST Africa Conference

Merton RK (1968), *The Matthew effect in Science,* Science, 159(3810), 56–63

Meyskens M and Bird L (2015), *Crowdfunding and value creation*, Entrepreneurship Research Journal, 5 (2), 155–166

Midi H, Sarkar SK, Rana S (2015), *Collinearity diagnosis of binary logistic regression model*, Journal of Interdisciplinary Mathematics, 13 (3), 253 - 267

Mochkabadi K, Volkmann CK (2020), *Equity crowdfunding: a systematic review of the literature*, Small Business Economics, Springer, 54, 75 – 118

Moleskis M, Alegre I, Canela MA (2019), *Crowdfunding Entrepreneurial of Humanitarian Needs? The Influence of Signals and Biases on Decisions*, Nonprofit and Voluntary Sector Quarterly, 48 (3), 552 - 571

Mollick E (2014*), The dynamics of crowdfunding: An exploratory study,* Journal of Business Venturing, 29 (1), 1–16

Morduch J. (1999), *The microfinance promise*, Journal of Economic Literature, 37, 1569–1614

Moss TW, Renko M, Block E, Meyskens M (2018), *Funding the story of hybrid ventures: Crowdfunder lending preferences and linguistic hybridity*, Journal of Business Venturing, 33(5), 643–659

Nickolas S, Kelly R, Jiwon MA (2022*), Understanding the Difference Between Moral Hazard and Adverse Selection,* Investopedia

Olsen MC, Slotegraaf RJ, Chandukala SR (2014), *Green claims and message frames: How green new products change brand attitude*, Journal of Marketing, 78(5), 119–137

Ordanini A, Miceli L, Pizzetti M, Parasuraman A (2011), *Crowd-funding: Transforming customers into investors through innovative service platforms,* Journal of Service Management, 22(4), 443- 470

Parhankangas A, Ehrlich M (2014), *How entrepreneurs seduce business angels: An impression management approach,* Journal of Business Venturing, 29(4), 543–564

Parhankangas A, Renko M (2017), *Linguistic style and crowdfunding success among social and commercial entrepreneurs*, Journal of Business Venturing, 32 (2), 215–236.

Park OC, Hopkins R (1993), *Instructional conditions for using dynamic visual displays: a review,* Instructional Science, 21, 427– 449

Petruzzelli AM, Natalicchio A, Panniello U, Roma P (2019), *Understanding the crowdfunding phenomenon and its implications for sustainability*, Technological Forecasting & Social Change, Elsevier, 141, 138 – 148

Pierrakis Y (2019), *Peer-to-peer lending to businesses: Investors' characteristics, investment criteria and motivation*, The International Journal of Entrepreneurship, 20, No. 04, 239 – 250

Pietraszkiewicz A, Soppe B and Formanowicz M (2017*), Go pro bono: Prosocial language as a success factor in crowdfunding*, Social Psychology, 48 (5), 265–278

Pitschner S, Pitschner-Finn S (2014), *Non-profit differentials in crowd-based financing: Evidence from 50,000 campaigns,* Economics Letters, 123(3), 391–394

Rahm E, Do HH (2000), *Data cleaning: Problems and current approaches*, Bulletin of the Technical Committee on Data Engineering, IEEE Computer Society, 23 (4)

Robinson MS (2001), *The Microfinance Revolution: Sustainable Finance for the Poor*, World Bank Publications

Saltelli A, Aleksankina K et al. (2019), *Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices*, Environmental Modelling & Software, Elsevier

Savin I, Chukavina K, Pushkarev A (2022), Topic-based classification and identification of global trends for startup companies, Small Business Economics, Springer, 60, 659 - 689

Senaviratna NAMR, Cooray TMJA (2019), *Diagnosing multicollinearity of logistic regression model*, Asian Journal of Probability and Statistics, 5(2), 1-9

Shneor R, Zhao L, Flaten BT (2020), *Advances in crowdfunding: Research and practice*, Springer Nature

Sievert C, Shirley K (2014), *LDAvis: A method for visualizing and interpreting topics, Proceedings of the workshop on interactive language leaning, visualization and interfaces,* Association for Computational Linguistics, 63 - 70

Stiglitz J, Weiss A (1983), *Incentive effects of terminations: Applications to the credit and labor markets*, American Economic Review, 73, 912–927

Stiglitz JE (2002), *Information and the change in the paradigm in economics*, American Economic Review, 92 (3), 460–501

Tvinnereim E, Liu X, Jamelske EM (2017), *Public perceptions of air pollution and climate change: Diferent manifestations, similar causes, and concerns*, Climatic Change, 140, 399–412

Unnava HR, Burnkrant RE (1991), *An Imagery-Processing View of the Role of Pictures in Print Advertisements,* Journal of Marketing Research, 28(2), 226–231

Unwin A (2020), *Why is data visualization important? What is important in data visualization?*, Harvard Data Science Review 2.1

Van de Ven AH, Sapienza HJ, Villanueva J (2007*), Entrepreneurial pursuits of self- and collective interests*, Strategic Entrepreneurship Journal, 1(3–4), 353–370

Viotto da Cruz J (2018), *Beyond financing: crowdfunding as an informational mechanism*, Journal of Business Venturing, 33 (3), 371–393

Yu Z, Zhang D, Wang Z, Guo B, Roussaki I, Doolin K, Claffey E. (2017), *Toward Context-Aware Mobile Social Networks*, IEEE Communications Magazine, 55(10), 168–175

Zhang H, Salvendy G (2001), *The implications of visualization ability and structure preview design for web information search tasks*, International Journal of Human-Computer Interaction, 13(1), 75–95

Zhang J, Liu P (2012), *Rational Herding in Microloan Markets,* Management Science, 58 (5), 892- 912

Zhao Q, Chen CD, Wang JL, Chen PC (2017), *Determinants of backers' funding intention in crowdfunding: social exchange theory and regulatory focus*, Telematics and Informatics, 34 (1), 370–384

(2022), *Crowdfunding – Worldwide*, Statista Market Forecast