

Louvain School of Management

Analysis of the impact of lexical complexity on the outcome of crowdfunding campaigns

Author - BOLAND Camille
Supervisor - THEWISSEN James
Academic year - 2021-2022
Dissertation for the master of Management (GEST2M)
Corporate Finance
Daytime schedule

LOUVAIN SCHOOL OF MANAGEMENT

MASTER THESIS

CAMILLE BOLAND

CORPORATE FINANCE TRACK

Analysis of the impact of lexical
complexity on the outcome of
crowdfunding campaigns

ACADEMIC YEAR: 2021-2022

PROMOTOR: THEWISSEN JAMES



LOUVAIN
School of Management



Abstract

The existing literature on crowdfunding identifies funding success factors by focusing mainly on campaign-specific characteristics, such as the funding goal set, the duration of the project or the presence of a visual component. However, the content of project descriptions has been largely ignored. In this research, we investigate the impact of lexical complexity on funding success. Using a dataset collected on the crowdfunding platform Kickstarter, we find that this linguistic feature is significantly associated with the success of the campaign. We further examine this influence and find that lexical complexity can both encourage and discourage potential backers to contribute. Overall, this research indicates that project descriptions have significant predictive power on the outcome of a crowdfunding campaign and that its analysis can help project owners and backers to assess the likelihood of funding.

Keywords: Crowdfunding; Content Analysis; Lexical Complexity; Empirical Study

I would like to express my sincere gratitude to my thesis supervisor, James Thewissen, Professor of Finance at UCLouvain, for his availability, guidance and valuable advice throughout the writing of this thesis. I would also like to thank him for providing me with all the necessary tools to carry out this research. His experience has been a precious help not only for the literature review but also for the development of the statistical model.

Contents

1 Introduction	1
2 Literature Review	5
2.1 The lexical complexity	5
2.2 Determinants of successful crowdfunding campaigns	6
2.2.1 The particularities of crowdfunding campaigns	6
2.2.2 The message framing	7
2.2.3 The linguistic features	9
3 Research Question	10
4 Data Selecion	11
5 Methodology & Variables Description	12
5.1 Methodology	12
5.2 Variables Description	13
5.2.1 Dependent Variables	13
5.2.2 Independent Variables	14
5.2.3 Control Variables	15
6 Empirical Results	16
6.1 Descriptive Statistics	17
6.2 Regression Results	18
6.2.1 Analysis of Results	18
6.2.2 Discussion of Results	22
6.2.3 Research Question	24
7 Conclusion	25
8 Bibliography	27
9 Appendix	30

1 Introduction

In order to enter the market and grow, it is essential for early-stage ventures to raise funds. However, it appears that fundraising is one of the main difficulties involved in the creation process of a company. Indeed, there are many innovative and successful entrepreneurial projects that do not see the light of day due to a lack of financial resources. In the current corporate society, early-stage ventures struggle to have access to traditional financing solutions which are usually granted to mature companies that have already demonstrated a certain profitability. Traditional sources of funding were previously restricted to personal savings, funds raised from friends, family, and fools, bank loans and financial markets. However, in recent years, crowdfunding has become an essential financing alternative allowing young entrepreneurs to raise sufficient funds in order to create, develop, and ensure the sustainability of their business. As the name suggests, crowdfunding is the process of collecting small amounts of money from a pool of individuals through specialized platforms. A major part of a crowdfunding campaign is based on textual description (Rothman and al., 1997); it is therefore of utmost importance to determine the factors that would predict the success of a campaign. Textual description contains informational value that allows potential investors to learn more about the project owner, but also to understand the value of the project in which they are willing to invest (Cappa and al., 2021). Their decision to contribute to a particular project therefore mainly relies on the information disclosed in this description. While a great deal of literature has demonstrated the importance of content descriptions in the success of crowdfunding campaigns, the linguistic aspect has received relatively little attention from scholars. For this reason, this study focuses on the linguistic aspect of crowdfunding campaigns and attempts to identify the impact of a linguistic feature, namely the lexical complexity, on their success rate.

This report is motivated by the significant growth experienced by crowdfunding in recent years. In fact, the concept of crowdfunding is nothing new. Like the construction of the Statue of Liberty, many projects have been developed after being funded by the general public. But its emergence and popularity have recently been greatly accelerated for various reasons (Agrawal and al., 2011). First, in an economy where asymmetry of information between entrepreneurs and investors persists and where conditions for granting bank loans are tightening, access to external financing is relatively limited for start-ups. Consequently, crowdfunding is increasingly becoming part of the business model of companies as it represents a valuable source of funding that allows young entrepreneurs to achieve the first stage of financing. The emergence of this method of fundraising is also driven by the evolution of computer technologies. Indeed, this evolution has not only allowed the development of specialized platforms in the field of the collaborative economy, but it has also facilitated the connection

between entrepreneurs and potential contributors, allowing to reach out to a larger number of them. In addition, crowdfunding is an alternative that is known to enjoy a very broad lexical diversity and heterogeneity. Indeed, campaign descriptions are written by small entrepreneurs who have a lot of freedom in the way they convey information to potential investors and in the linguistic framework they use. By contrast with traditional financial documents, the crowdfunding campaigns are not or only barely standardized, which leads to considerable diversity. Finally, it is relevant to look at the lexical complexity in the context of crowdfunding since this linguistic feature tends to increase with distance. The more geographically distant an entrepreneur is from his potential contributors, the greater the lexical complexity. And as mentioned earlier, thanks to the internet, crowdfunding allows entrepreneurs to seek funds from a public all over the world.

Over the years, entrepreneurial fundraising has been studied from different approaches to raise the main issues related to this model, but also to discover the mechanisms that characterize this process. To date, some authors have focused on the factors that predict the outcome of a crowdfunding campaign (Mollick, 2014; Mitra and Gilbert, 2014). Other authors have looked at the motivations of entrepreneurs to turn to this financing alternative (Schwienbacher and Larralde, 2010), incentives of investors to contribute to fundraising (Belleflamme and al., 2013), or the impact of the geographical distance that persists between entrepreneurs and contributors (Agrawal and al., 2011). Recently, literature on the subject has extended and started to consider the language aspect of the fundraising. Numerous studies have demonstrated the importance of textual description within a crowdfunding campaign. Some authors have shown that the way a message is written impacts its effectiveness and therefore that the message framing influences fundraising (Moradi and Dass, 2019). One question that comes to mind is to determine the impact of the lexical complexity of a crowdfunding campaign description on its funding success. The aim of this paper is therefore to analyze the extent to which lexical complexity influences the investment decision of individuals and consequently the level of funding.

The present study supports the idea that the degree of lexical complexity of textual descriptions of crowdfunding campaigns effectively has an impact on the level of funding. First, we assume that a more lexically complex writing is positively related to the talent of the project owner. Talented entrepreneurs generally demonstrate more developed linguistic skills than less talented entrepreneurs. Consequently, we predict that crowdfunding campaigns with more lexically complex descriptions are perceived as more professional, which encourages the contribution of backers. However, we also assume that an overly complex textual description can reduce its readability and thus hinder the understanding of potential investors, leading them to distance themselves from the project. Therefore,

we expect the relationship between the degree of complexity of the campaign presentation and the investment decision of backers to be sometimes positive and sometimes negative. This suggests that there is an optimal level of lexical complexity that demonstrates the proficiency of the entrepreneur but does not interfere with the understanding of backers.

In order to support this research, 5000 campaigns launched on Kickstarter between 2012 and 2019 are analyzed. For each campaign, a wide range of information has been collected. Some of this information relating to the campaign itself such as the amount pledged, the number of backers, the duration of the campaign, etc. And some other information relates to the lexical aspect of the project descriptions. To answer the research question of this paper appropriately, an empirical analysis has been performed through various regressions. The latter revealed that the lexical complexity does influence the outcome of a crowdfunding campaign.

In an effort to study this problematic in the most appropriate way, this report will be structured in two distinct parts. In the first part, a review of the existing literature on the subject will be carried out to put some existing studies into perspective and to determine the impact of textual description in entrepreneurial narratives on fundraising. Following this, the research question will be developed. In the second part, which will be devoted to an empirical analysis, the data and methodology will be presented. Subsequently, the results of this study will be analyzed and compared to the previous results to detect the similarities as well as the discrepancies. Finally, this second part will provide a conclusion that will summarize the research carried out in this report, identify its limitations, and suggest potential recommendations to improve future research.

This report finds itself at the crossroads between financial literature and linguistics and contributes to the current literature through diverse dimensions.

First, this study contributes to the research on crowdfunding. Given its significant growth, crowdfunding is a financing alternative that is subject to a great deal of academic research. To date, majority of existing literature on crowdfunding focuses on the mechanisms that characterize this process, the campaigns' success factors and the motivations of funders and entrepreneurs to take part in this financing alternative. However, the influence of textual information on the level of funding has received little attention from researchers. To the best of our knowledge, our study is the first one to investigate the impact of lexical complexity on investors' decision making in the context of the collaborative economy. By analyzing the content of textual descriptions, this study measures the influence of lexical richness on funding success and thus improves the understanding of crowdfunding.

Second, this paper completes existing academic research that highlights the impact of qualitative information on the investment decision of the general public. It differs from previous studies that investigate the influence of content, length, and readability on contributors' decisions (Mitra and Gilbert, 2014; Moy and al., 2018), by examining a new linguistic feature, namely lexical complexity. However, this paper is not the first one to investigate the influence of lexical aspect on investor behavior. Joenväärä and al. (2019) have recently explored informational content of hedge funds' strategy descriptions by examining the effect of lexical diversity and sophistication on fund quality. Our study differs from the latter in various ways. First of all, the funding alternative considered in the two studies is fundamentally different. Indeed, a hedge fund represents a pooled investment fund, which means that the investors expect some financial return on the amount contributed, whereas crowdfunding is a way of raising funds without promising any financial return to backers. In contrast to our study, Joenväärä and al. (2019) focus exclusively on the influence of lexical diversity and sophistication on the investment decision, but they do not consider the dimension of lexical density. Our study broadens the scope of the literature by taking into account three interrelated dimensions which characterize the notion of lexical complexity. Therefore, our contribution lies in the fact that we will quantify the monetary impact of lexical complexity on investor decision making in a crowdfunding process.

Third, this work contributes significantly to project owners and potential contributors. The results of the present study allow entrepreneurs to have a better understanding of the investment behavior of funders in relation to the lexical complexity of textual descriptions. This allows entrepreneurs to assess the extent to which complexity is sought and tolerated by investors and consequently adapt the content of campaign descriptions to best meet the contributor's expectations. Moreover, this work provides potential contributors with information that allows them to assess the linguistic content of the descriptions written by project owners and thus anticipate the outcome (success or failure) of crowdfunding campaigns.

2 Literature Review

This section presents a review of the literature that aims to put results of previous work on the subject into perspective. This review will highlight the issues related to the collaborative economy, it will also help to better understand the underlying purpose of this report and to develop the research question.

First of all, the main topic of this paper, namely the lexical complexity, will be defined in the most appropriate way. Then, we will focus on the importance of linguistic aspect and on the factors that have been determined to be significant in predicting the outcome of a crowdfunding campaign.

2.1 The lexical complexity

The lexical complexity is a notion that has attracted the attention of many researchers over the years and that has been extensively studied. However, this concept remains relatively challenging to analyze.

The major challenge in studying the lexical complexity is that it is a notion that cannot be defined in absolute terms. Indeed, each individual will tend to perceive linguistic items differently depending on the context in which they are presented. Therefore, there are currently several definitions of the lexical complexity, but none of them is commonly accepted. In this paper, it has been decided to use to the definition given by Bulté and Housen (2012). According to them, the lexical complexity refers to the “degree of elaboration, the size, breadth, width, or richness of the learner’s lexical system and the way in which the learner uses and masters it”. This linguistic aspect is composed of three interrelated dimensions, namely the lexical sophistication, density, and diversity. Many authors, like Bulté and Housen (2012), support the idea that these three dimensions allow the quantification of the degree of complexity within a text.

Another challenge lies in the fact that there is no universal measure that allows researchers to quantify the degree of complexity in an essay. Although there are various methods for studying the three dimensions separately, there is no one unique method which considers all three aspects simultaneously. Moreover, in the operationalization process of this linguistic aspect, researchers encounter difficulties in comparing texts of different length or language. It is important to point out that the length of a writing is an element that should not be neglected because the longer the text, the higher the chances that it will be complex. It is also not recommended to compare the linguistic complexity of texts written in two different foreign languages.

2.2 Determinants of successful crowdfunding campaigns

This section identifies several signals that help to frame the assessment of the future performance of a crowdfunding campaign. As the main objective of this report is to examine the effect of the lexical complexity on the fundraising outcomes, it seems appropriate to mainly focus on quality signals directly related to entrepreneurial narratives. Indeed, the textual description is the most relevant component in order to analyze the effect of linguistic content (Gorbatai and Nelson, 2015).

2.2.1 The particularities of crowdfunding campaigns

Crowdfunding is a method of fundraising that differs from traditional funding channels through several dimensions.

The first particularity concerns the communication medium. Indeed, in crowdfunding, a major part of the campaign is based on the description of the project (Rothman and al., 1993). As the whole process takes place online, without face-to-face interaction with potential backers, it is obvious that the description plays a fundamental role in convincing the crowd of the quality of the project (Mitra and Gilbert, 2014). This description is an essential component for the entrepreneurs, as well as for the resource providers. On the one hand, it is the only communication tool available to entrepreneurs in order to promote their ideas, attract the attention of the crowd and convince them to commit funds in their projects. And, on the other hand, it represents a significant source of information for the investors who aim to invest money in a project and seek to assess its quality.

The second particularity is about the type of investors targeted. In traditional financing, ventures have to convince professionals who are experienced in the field of finance. But, in crowdfunding, ventures appeal directly to the generosity of the general public and therefore to individuals who are not necessarily specialized in business (Adamska-Mieruszevska and al., 2021). The difference in the type of investors will tend to modify the motivations that drive individuals to invest. Indeed, traditional investors are professionals who are willing to invest in a project they consider to be profitable and that will generate a good return on investment (Gorbatai and Nelson, 2015). In contrast to crowdfunding, where contributors are not looking for a potential financial return as they will not receive the amount invested back. They choose to commit funds to support a cause close to their hearts and in line with their personal values.

The third particularity relates to the process of persuasion. As a result of the difference in the investment motivations, the process and language used in order to convince resource providers will not be the same as in the traditional methods. While an effective communication to convince traditional investors will mainly rely on financial language that highlights the profitability and the expected return of the project, this type of language will often be perceived as unattractive and too complex by the general public. In order to get funded using crowdfunding, it is advisable to use an attractive, clear, and understandable language.

The fourth particularity is related to the geographical constraints. While in general, the success rate of traditionally funded entrepreneurial project is strongly impacted by the geographical constraint, in the context of crowdfunding, the impact of this constraint seems to be significantly reduced. Agrawal and al. (2011) have explored the potential relationship between the distance and the investment decision of individuals, and they found that the constraint of the geographical proximity tends to be relaxed in crowdfunding campaigns. Indeed, as the whole process of this fundraising model takes place online, it is possible to connect entrepreneurs and investors who are located in geographically distant areas. However, the geographical distance can have various effects on the willingness of the crowd to invest in entrepreneurial projects and even if crowdfunding has the potential to mitigate some of these effects, it cannot eliminate them entirely.

2.2.2 The message framing

Previous studies on health communication have shown that the message framing tends to interact with the individuals' behavior and therefore influence their decision making (Rothman and Salovey, 1997; Gallagher and Updegraff, 2012). In entrepreneurial fundraising, the message framing is an effective way to highlight the value of a project (Das and al., 2008). We can therefore expect that the way the message is framed will have an influence on the fundraising outcome. There is a growing body of the literature dealing with the storytelling and the linguistic style associated with it, which argues that the message has a real impact on the investment decision (Chen and al., 2020; Larrimore and al., 2011; Parhankangas and Renko, 2017; Zhao and Ryu, 2020). In the context of peer-to-peer lending, Chen and al. (2020) examined the extent to which investors consider information disclosed in order to make their investment decisions and found that information disclosure can significantly impact the probability of financing. Furthermore, Larrimore and al. (2011) investigated the existing relationship between the language used in peer-to-peer lending and the persuasion success. The results showed that the use of a long, concrete, and quantitative narrative is positively correlated with the funding success. In the context of crowdfunding, Parhankangas and Renko (2017) looked at the linguistic style of descriptions and the correlation between this component and the fundraising

outcome and they noted that a linguistic style which facilitates the understanding of the project tends to increase the chances of success of the campaign.

The literature identifies specific message framings that are hidden in descriptions of crowdfunding campaigns. It distinguishes between positive and negative framing or rational and emotional framing. First, prior studies have tested the effectiveness of positive and negative framing, but the results obtained differ from one study to another. Indeed, there are some empirical studies claiming that a positive frame in a description influences directly and positively the level of funding. For example, in their paper on the role of the language in online fundraising, Gorbatai and Nelson (2015) have shown that the use of a language that generates positive emotions is positively correlated with fundraising success. However, there are other studies suggesting that negative frame leads to more favorable fundraising outcomes. Moradi and Dass (2019) have carried out a research on the impact of the message framing on the outcome of a crowdfunding campaign. In this research, they defined the positive framing as the language associated with benefits and the negative framing as the language associated with costs and demonstrated that negative frame significantly influences the investment decision of the contributors, while positive frame does not have a direct positive impact on the fundraising level. Second, the rational and emotional message framings have been addressed by Chen and al. (2016) in their study about what leads to a successful promotional campaign on a crowdfunding platform. For utilitarian products, which are useful and practical products, the authors noted that emotional framing increases the level of donation, while rational framing has no impact on it. For hedonic products, which are pleasure products, both emotional and rational message frames influence the donation level. Consequently, Chen and al. (2016) suggest that the use of an emotional framing would be positively correlated with the level of funding in a crowdfunding campaign.

The content of a campaign presentation plays a major role in the outcome of the fundraising. Mitra and Gilbert (2014) performed a study that examined the predictive power of the description content by analyzing the words and phrases used by entrepreneurs to describe their project. Through this paper, they demonstrated that the way the entrepreneurs communicate with the crowd can significantly influence the success of the campaign. Indeed, words related to reciprocity, optimism or confidence of the entrepreneur are positively correlated to the success of a crowdfunding campaign, whereas pessimistic words and lack of confidence of the entrepreneur are more likely to lead to a non-funding campaign.

Furthermore, the way a message is framed is directly related to the socio-demographic profile of the entrepreneur. Indeed, Mollick (2014) found that crowdfunding projects tend to be closely correlated with the lifestyle and the culture of the geographical area in which the project creator lives. Therefore, individuals may be unconsciously influenced by the distance and invest in projects that are in line with their cultural expectations. Moreover, Gorbatai and Nelson (2015) argue that the language used by the entrepreneur is directly related to his socio-demographic, as well as geographical characteristics. It is obvious that individuals will have different oral and written communication styles depending on their geographical location and demographic group and this will be reflected in the descriptions of crowdfunding campaigns. Once again, investors may be sensitive to this aspect and favor the funding of projects that are close to their own communication style.

2.2.3 The linguistic features

In crowdfunding, the quality of communication between project owners and potential investors is a key component to ensure the success of the campaign. In order to gain the trust of the crowd and thus reach the fundraising goal, the entrepreneurs should definitely use a qualitative language in the project presentation. According to Mollick (2014), the underlying quality of the project is associated with the success of the campaign. In order to measure this quality, the author examined, among other things, the presence of spelling mistakes in the project description and noted that the chances of success of projects with spelling mistakes are relatively lower. Khelladi and al. (2018) further investigated the question by hypothesizing that the linguistic quality of description influences individuals' level of trust and therefore their investment decision. The results indicated that a description without spelling or grammatical errors leads to a higher level of confidence and thus to successful fundraising.

The readability of crowdfunding descriptions is an additional component that is essential in order to guarantee the success of the campaign. As potential supporters seek to have a good understanding of the project in which they are willing to invest, they tend to prefer projects with a clear and readable presentation. In their paper which aim to study the impact of the language on the success of reward-based crowdfunding campaigns, Adamska-Mieruszevska and al. (2021) prove that the readability of the text is positively related to the fundraising outcome. Another component that helps to make project descriptions relatively more comprehensible is the level of complexity of the language used. Although the use of a more complex language can be associated with the professionalism of the project owner, which builds confidence among individuals (Zhou and al., 2018), an overly complex description could discourage the community from contributing to the project.

In order to assess the reliability and professionalism of the entrepreneurs, the contributors can focus on lexical indices such as the lexical diversity and sophistication of project descriptions. It is therefore possible to separate good managers from bad ones by analyzing the linguistic content used in their campaign. By studying the information content of hedge funds strategy descriptions, Joenväärä and al. (2019) have examined the effect of the lexical diversity and sophistication on fund quality. Concerning the lexical diversity, the authors argued that there is a positive relationship between this textual measure and the fundraising outcome which means that descriptions with a high degree of lexical diversity perform better than those with less diversity. It is obvious that talented managers use a richer vocabulary to express themselves than less talented managers as they have more sophisticated linguistic skills. Therefore, lexical diversity allows investors to recognize the professionalism of the entrepreneur and to convince them to contribute. Concerning the lexical sophistication, the findings of Joenväärä and al. (2019) are mixed. On the one hand, they noticed that a more sophisticated description seems to predict the quality of hedge funds, but on the other hand they noted that the use of a more sophisticated language can be associated with deceptive behavior. Indeed, it appears that less skilled managers attempt to deceive their audience by writing with a high degree of syntactic complexity. Similarly, Moffitt and Burns (2009) agreed that fraudulent authors use a complex vocabulary. By making the text more difficult to read, bad manager chooses to deceive investors by obfuscating their understanding, but also by hiding bad news and the lack of professionalism.

3 Research Question

Based on the literature review presented in the previous section, we assume that lexical complexity has an influence on the outcome of a crowdfunding campaign for various reasons. First of all, the literature has shown that the project description is a major element of a crowdfunding campaign (Rothman and al., 1993). Since it represents the only source of information of the campaign, the textual description of the project is carefully considered by the resource providers when they are making their financial decision. Moreover, prior studies have shown that the message framing plays an important role in the investor decision making (Moradi and Dass, 2019). Considering these statements, we assume that the content, as well as the linguistic characteristics of crowdfunding project description, influence the backers' decision making and consequently the outcome of the fundraising. Furthermore, prior studies have shown that lexical indices can be used to gauge the general linguistic competence of an entrepreneur and thus give an indication of his level of performance (Kondal, 2015). This observation leads us to believe that the linguistic quality of textual descriptions is perceived by the crowd as a signal of the entrepreneur's professionalism. Therefore, we suggest that the degree of lexical complexity may affect the ability to attract funds. Indeed, this linguistic competence is

perceived by potential backers as a signal of quality, which fosters their confidence and encourages them to contribute. However, it is important to highlight that the lexical complexity should not hinder the readability of the campaign, which is also important for successful fundraising (Adamska-Mieruszevska and al., 2021). As individuals generally seek to understand the project before making their investment decision, we believe that an obscured understanding leads individuals to distance themselves from the project and thus discourage their participation in the campaign.

In this report, we support the idea that lexical complexity does have an impact on the success of crowdfunding projects, however it may also be that this linguistic feature does not elicit any particular reaction from potential investors. Indeed, lexical complexity is a concept that is relatively difficult to define and may therefore not be widely considered by the general public. Moreover, the linguistic features of textual descriptions may not be the primary focus of individuals in their decision-making process as other elements, such as risk disclosure, may have more impact on individuals' behavior. Finally, it may be that other determinants of the campaign, such as the nature of the project, the type of crowdfunding or the experience of the entrepreneur, are more significant in the decision making of potential contributors than the level of lexical complexity of textual descriptions.

This discussion leads us to address the following research question: *Does lexical complexity have an impact on the success rate of crowdfunding campaigns?*

In order to answer this question, we obviously consider the variable that indicates the success or failure of the campaign, but we extend our analysis by including three other variables that represent success in some way. First of all, we consider the total amount pledged to the campaign, which can be associated with the success of the campaign since the higher this amount, the more successful the campaign. Then, we look at the number of individuals who contribute to the campaign. This can also represent the success rate of the campaign as the greater the number of backers, the more positive the outcome of the campaign. Finally, we take into account the duration of the campaign that can give an indication of the outcome of the fundraising, considering that the faster the campaign is funded, the more successful it will be.

4 Data Selection

In order to conduct our exploratory research in the most appropriate way, we have gathered a detailed dataset of lexical complexity, as well as campaign attributes for a series of projects launched on Kickstarter.

Kickstarter is one of the most prominent reward-based crowdfunding platforms which supports a broad variety of projects that are divided into 15 distinct categories representing the different industries. Since its inception in April 2009, the platform has already successfully funded more than 220,000 projects by raising over \$6.5 billion from 21 million individuals . Kickstarter operates with an « all or nothing » strategy, which means that a project will only be funded if the minimum funding goal is reached. By using this strategy, the global platform offers some security to the backers. Kickstarter data is frequently used in crowdfunding literature as it is easily accessible for researchers and relatively exhaustive. These datasets have already been used several times in the scientific literature, for instance by Mollick (2014), Mitra and Gilbert (2014) and Madsen and McMullin (2018).

To answer our research question, we have collected a sample of data based on 5,000 projects pitched on Kickstarter between September 2012 and January 2019 and which contains both information related to the campaign and information allowing to quantify the level of lexical complexity.

5 Methodology & Variables Description

This section describes the methodology followed in order to answer our research question and the main variables used in order to perform our regressions.

5.1 Methodology

In order to analyze the impact of lexical complexity on the outcome of crowdfunding campaigns, an empirical analysis will be conducted through various regressions. To perform this analysis, two different databases are used. The first database contains quantitative and qualitative information about the characteristics of campaigns launched on Kickstarter which will be used as dependent and independent variables in the regression. The second database contains information related to the lexical complexity as it includes all the variables used to measure the level of lexical sophistication, diversity, and density of the project description. The data provided in the latter will be used as independent variables. Then, these two databases are merged to create a new one that gathers all the dependent and independent variables needed to run the regressions. It contains a total of 5,000 observations, with each observation representing a campaign launched on Kickstarter between 2012 and 2019.

Thereafter, we apply the Principal Component Analysis method to measure the variables of lexical sophistication and lexical diversity. This method is frequently used to extract and summarize the information contained in a large number of variables into a smaller number of synthetic variables that are more easily analyzed and interpreted. This step aims to create one single variable for the lexical sophistication based on the 5 variables related to this component available in our database and one single variable for the lexical diversity based on the 19 variables related to this component available in our database.

The empirical analysis of this paper is performed using four regressions, all four analyzing the impact of lexical complexity on salient factors of a successful crowdfunding campaign. Since our dependent variables consist of three continuous variables and one binary variable, there are three linear regressions and one logistic regression in our model. As a reminder, linear regression models are used to highlight the relationship between a continuous dependent variable and a set of independent explanatory variables, whereas logistic regression models are used to highlight the relationship between a binary dependent variable and a set of independent explanatory variables. The model is expressed as follows:

$$y = \alpha + \beta LexicalComplexity + \gamma Controls + \epsilon \quad (1)$$

where ϵ denotes the robust standard errors which are corrected for the heteroscedasticity.

5.2 Variables Description

In the following subsections, the meaning of the main variables, as well as their importance in this research are analyzed. These variables are divided into three distinct groups, namely the dependent variables, independent variables, and control variables.

5.2.1 Dependent Variables

The first dependent variable is *USD_Pledged* which is the total amount pledged to crowdfunding campaign in US dollars. This variable is usually used in the finance literature to measure the funding success of a campaign (Madsen and McMullin, 2018). Indeed, this funding success can be determined by comparing the total amount pledged in US dollars with the size of the goal in US dollar.

The second dependent variable used for analysis is related to the number of backers of the campaign. The *Backers_Count* variable represents the number of individuals pledging to support a project. Since the literature has already highlighted the importance of backers' support in achieving the goal of a crowdfunding campaign, it seems relevant to take this aspect into consideration in our analysis (Petruzzelli and al., 2019).

The third dependent variable is related to the duration of the campaign. The *Duration* variable corresponds to the number of days during which the project accepts contributions. Since the projects presented on Kickstarter cannot last longer than two months, the number of days will vary between 1 and 60 days. Since previous studies on crowdfunding have already included this variable in their analysis, we will be able to compare whether our results are consistent with those previously obtained (Mollick, 2014).

In order to take into account the skewed distribution of these three variables, we use a logarithmic transformation of these variables in our regressions.

The fourth dependent variable, *Success*, is a binary variable which indicates whether the campaign has been successfully funded or not. This binary variable takes the value 0 for campaigns that were not successfully funded and the value 1 for the campaigns that were successfully funded. Since Kickstarter applies an "All or Nothing" pledge mode, a campaign will only be considered as successful if it has reached its funding goal.

5.2.2 Independent Variables

The independent variables of the regressions correspond to the three main components of the lexical complexity, namely the lexical sophistication, the lexical diversity, and the lexical density. These variables will be used in order to measure the impact, and its significance, of the lexical complexity on the outcome of crowdfunding campaigns.

Lexical.Sophistication: The lexical sophistication refers to the proportion of advanced, difficult, or rare words that are used in a learner's text (O'Dell and al., 2000). As there is no universal definition of advanced or difficult words, the degree of sophistication of a text is primarily determined by the rarity of the words it contains (Bulté and Housen, 2012). This component of the lexical complexity provides a significant measure of the learner's overall writing competence. The lexical sophistication of a text is measured using 5 different variables that are detailed in *Figure 1*. As explained previously, we apply the PCA method for this variable, which provides a single variable that represents the notion of lexical sophistication.

Diversity.Pca: The lexical diversity is the range of a learner’s vocabulary as found in their language use (Kim, 2014). This component is directly related to the number of different words that are used in a text (Johansson, 2008). A paper that enjoys a large number of different words and few repetitions is considered to be significantly diverse and reflects the author’s good lexical skills. Over the years, various measures to quantify the degree of diversity in a text have been developed. Some of them are relatively intuitive but do not appear to be the most efficient because they are strongly impacted by the length of the text studied. Therefore, in order to best quantify this component, it is measured using 19 different variables that are detailed in *Figure 1*. As explained previously, we also apply the PCA method for this variable, which provides a single variable that represents the notion of lexical diversity.

Ld: The lexical density of a piece of writing is directly related to the proportion of lexical items it contains (Johansson, 2008). The elements considered as lexical items are nouns, verbs, adjectives, as well as adverbs. In order to quantify this component in the most appropriate way, authors such as Bulté and Housen (2012) usually use a ratio that divides the number of lexical words by the total number of words in a text. This last component is measured by using one variable (*LD*).

5.2.3 Control Variables

In addition to the independent variables related to the lexical complexity, 9 control variables which may affect the funding outcome are included in the regressions.

First, we control the impact of some textual attributes of the campaign. The variable *AboutWords* represents the number of words contained in the description of the project. This variable provides an indication of the length of the pitch provided by the entrepreneur in the “About this project” section on Kickstarter. The variable *RiskWords*, which is the number of words in the “Risks and Challenges” section, is also introduced in the model in order to measure the amount of risk information provided to the potential backers. The *Readability* variable measures the readability of the text based on the Flesch Reading Ease Score. This score evaluates the level of readability by using two separate variables, the average number of words per sentence and the average number of syllables per word. A high score means that the project description is relatively easy to read, and conversely a low score means that the project description is particularly difficult to read.

Then, we control the impact of certain characteristics specific to the campaign. We include some variables in order to control the impact of the location, *US_Project* and *factor(Location_Country)*. The variable *US_Project* reflects whether the project is located in the United States or in another country, while the variable *factor(Location_Country)* goes further by indicating the country of origin of the project. Since there are several countries and they cannot take the form of a single value, this variable is expressed as a factor.

Furthermore, we introduce dummy variables for each category of Kickstarter *factor(Category)* in order to consider the specificities related to the different industries. As a reminder, the platform classifies projects among 15 categories: Art, Comics, Dance, Design, Fashion, Film & Video, Food, Game, Music, Photography, Publishing, Technology, Theater. This variable is also expressed as a factor since there are several categories and they cannot take the form of a single value. The variable *Google_Trend* is integrated in order to assess the level of popularity of the platform Kickstarter. This variable corresponds to the monthly index of Google Trends searches for the word “Kickstarter”. We control this variable because the popularity of the platform may affect the success of its campaigns.

Finally, we control the experience and the sentiment of the entrepreneur by including the variables *Experience* and *Optimism*. The *Experience* variable indicates the entrepreneur’s prior experience in the field of crowdfunding. It is a valuable indicator to predict the outcome of a campaign as it provides information on the entrepreneurial skills of the project owner. The *Optimism* variable refers to the feeling which reflects the attitude and state of mind of the project owner. Previous studies have included this variable and have demonstrated that a project description which reflects a positive sentiment promotes the fundraising success (Wang and al, 2017).

All the variables are summarized and defined in *Figure 2*.

6 Empirical Results

This part consists of the presentation of the descriptive statistics of the variables used in this report, the analysis of the results obtained after having performed the linear and logistic regressions developed above and how they answer the research question.

6.1 Descriptive Statistics

Before proceeding to the analysis of the results, the descriptive statistics of all the variables of the regression model are examined. As it can be seen in *Figure 3*, these descriptive statistics consist of the number of observations, the mean, the standard deviation, the minimum, as well as the maximum.

Figure 3: Descriptive Statistics

	N	Mean	St. Dev.	Min	Max
Dependent Variables					
<i>USD Pledged</i>	5,000	10,106.470	63,246.440	0.000	2,650,311.000
<i>Backers Count</i>	5,000	112.956	596.938	0	21,975
<i>Duration</i>	5,000	32.843	11.434	1.919	60.042
<i>Success</i>	5,000	0.417	0.493	0	1
Independent Variables					
<i>Lexical Sophistication</i>	5,000	0.00	1.7207	-11.271	4.4612
<i>Lexical Diversity</i>	5,000	0.00	2.77403	-11.080	6.35961
<i>Lexical Density</i>	5,000	0.528	0.058	0.120	0.800
Control Variables					
<i>Risk Words</i>	5,000	114.633	96.251	0	1,260
<i>About Words</i>	5,000	574.835	559.591	4	5,742
<i>Readability</i>	5,000	58.751	17.963	-552.312	103.687
<i>Experience</i>	5,000	0.492	2.078	0	60
<i>Google Trend</i>	5,000	51.066	16.453	21	100
<i>Us Project</i>	5,000	0.719	0.450	0	1
<i>Optimism</i>	4,991	0.028	0.014	0.000	0.168

The final sample of this report includes 5,000 observations corresponding to crowdfunding campaigns pitched on Kickstarter. The number of observations for the control variable *Optimism* are slightly lower, which is simply due to the fact that these data were not available for all the campaigns.

By analyzing this table, we observe that 41,7% of the campaigns in the sample have been successfully funded, which represents approximately 2,000 projects. We also see that a standard campaign on Kickstarter is funded over a period of just over a month and is supported on average by 112 backers. By looking at the *USD Pledged* variable, we notice a substantial variability in the total amount pledged by the backers, ranging from zero dollar to over 2.5 million dollars. Furthermore, the statistics report that 71,9% of the campaigns are located in the United States and that many projects have been launched by entrepreneurs who already have some experience in crowdfunding.

Regarding the project description, we observe that the section “About this Project” contains on average 575 words (574,8 words). We see that there is a high variability in terms of project description length as the shortest pitch contains only 4 words, while the longest one contains over 5,700 words.

The variables *Location Country* and *Category* have been omitted from this table as they are not relevant due to the fact that these variables represent different country codes and categories.

6.2 Regression Results

The regressions performed in the context of this research can now be analyzed on the basis of the results deployed in *Table 1*. The first subsection of this part aims to present the results obtained and what it implies, and the second subsection will discuss and interpret the significant variables.

6.2.1 Analysis of Results

As previously mentioned, the main objective of these regressions is to analyze the impact of lexical complexity, using its three components, on the success rate of crowdfunding campaigns. To achieve this, we run four regressions based on our dependent variables, namely *USD Pledged*, *Backers Count*, *Duration* and *Success*.

First Regression - *USD Pledged*

As a reminder, this regression focuses on the potential relationship between lexical complexity and the total amount pledged to crowdfunding campaign which is expressed in USD. By looking at the results obtained, we notice that our regression model fits relatively well and has a significant predictive power on the total amount pledged to the campaign. Indeed, we observe that the F Statistic is 19.669 with a p-value lower than 0.01, which is highly significant. Furthermore, we find that 28.89% of the variation in the total amount pledged to the campaign can be explained by our model, as shown by the value of the multiple R-squared.

	<i>Dependent variable:</i>
	<i>USD Pledged</i>
Observations	4,991
R ²	0.289
Adjusted R ²	0.274
Residual Std. Error	2.758 (df = 4889)
F Statistic	19.669*** (df = 101; 4889)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Concerning the independent variables, we can observe that all three components of lexical complexity are significant at a 100% confidence interval. First, lexical sophistication and lexical diversity are negatively correlated with the total amount pledged to the campaign, meaning that if the level of lexical sophistication and diversity increases, the total amount pledged decreases. Conversely, we notice that the coefficient of the lexical density is positive, which implies that a higher level of density will result in a higher total amount pledged.

Regarding the control variables, we find that the number of words present in the “About this Project” section is significant at a 100% confidence interval. This variable is positively correlated with the dependent variable of this regression, suggesting that text descriptions with a high number of words tend to favor the pledged amount. We also see that the number of words in the “Risks Challenges” section is significant at a 100% confidence interval and is positively correlated with the total amount pledged. Furthermore, the relationship we observe can be explained by the entrepreneur’s prior experience in crowdfunding. Indeed, this variable is highly significant and positively correlated with the final promised contribution of the backers. In addition, the *Google Trend* variable also shows some significance in our model, at a 99.9% confidence interval, and its coefficient suggests that the less the platform name is searched on Google, the higher the pledged amount will be. Finally, we have a significant relationship if we include the different categories in our regression. Indeed, we can observe that there are 4 categories that are significant at the 100% confidence interval, 2 categories at 99% and 3 categories at 95%. Some of these categories encourage the contribution of backers and others have the opposite effect.

Second Regression - *Backers Count*

The main purpose of this second regression is to study the influence of our independent variables on the numbers of individuals pledging to support a campaign. The results obtained indicate that the F statistics of this model is 21.525 with a p-value lower than 0.01. This highly significant p-value allows us to support the fit of our model. Moreover, we find that 30.78% of the variation in the total number of backers in the campaign can be explained by our model, as indicated by the value of the multiple R-squared.

	<i>Dependent variable:</i>
	<i>Backers Count</i>
Observations	4,991
R ²	0.308
Adjusted R ²	0.294
Residual Std. Error	1.596 (df = 4889)
F Statistic	21.525*** (df = 101; 4889)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Expect for some slight differences, the results of this second model are relatively similar to those obtained in the previous one. Therefore, the variables that significantly explain the relationship between the lexical complexity and the total amount pledged to the campaign can also significantly explain the relationship studied in this second regression.

However, we observe that the readability of the project description has a positive predictive power on the number of backers of the campaign as it is significant at a 100% confidence interval. This means that the higher the Flesch Reading Ease Score, the more individuals support the campaign. Once again, it is relevant to include the variable *Category* in our model as some of them have the power to influence the results. Indeed, 6 categories are significant at a 100% confidence interval, and 2 categories at 95%.

Third Regression - *Duration*

In this third regression, we aim to study the impact of the lexical complexity on the duration of a campaign. The F statistics of this model is 2.748 and the p-value is lower than 0.01, which is highly significant. In addition, we find that 5.373% of the variation in the campaign duration can be explained by our model, as shown by the value of the multiple R-squared, which is quite low.

<i>Dependent variable:</i>	
<i>Duration</i>	
Observations	4,991
R ²	0.054
Adjusted R ²	0.034
Residual Std. Error	0.377 (df = 4889)
F Statistic	2.748*** (df = 101; 4889)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

The results obtained for this regression model are somewhat different from those obtained for the first two models. First of all, we observe that the variables related to the lexical complexity are not significant in our model. This implies that the lexical sophistication, diversity, and density of a project description will not have a significant effect on the duration of a campaign.

Then, we can see that the number of words included in the project description is significant at a 90% confidence interval and has a negative coefficient. We can therefore assume that the higher the number of words in the project description, the shorter the duration of the campaign. Moreover, the results indicate that the *Risks Words* variable is significant at a 95% confidence interval and is positively correlated with the campaign duration. This suggests that a campaign will take longer to be successfully funded when the “Risks and Challenges” section is longer. In addition, we notice that the readability of the project and the entrepreneur’s experience in crowdfunding have a highly significant impact on the duration (at a 100% confidence interval). These two variables have a negative coefficient suggesting that the more readable the pitch and the more experienced the entrepreneur, the faster the campaign will be funded. Finally, we find that projects launched in certain categories

are more likely to influence the duration of a campaign than others. Among the 15 categories, 2 of them are significant at the 100% confidence interval, 2 at 99.9% and 1 at 95%.

Fourth Regression - *Success*

The last regression of this research is probably the most explicit regression as the main focus is on the success of the campaign. Indeed, it studies the impact of lexical complexity on the *Success* variable which indicates whether the campaign has been successfully funded or not. We note that the variation in the success of a crowdfunding campaign explained by our model is 17.48%, as evidenced by the value of the Pseudo R-squared.

<i>Dependent variable:</i>	
<i>Success</i>	
Observations	4,991
Log Likelihood	-2,797.334
Akaike Inf. Crit.	5,798.667
Pseudo R ²	0.1748
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

The results of this regression indicate that the three components of lexical complexity can significantly explain the success of a project. Indeed, lexical sophistication and lexical density are both significant at 100% confidence interval and lexical diversity at 95%. Similarly to the first two regressions, we see that the coefficient of lexical sophistication is negative, suggesting that the more sophisticated a text, the less likely the project is to result in success. In contrast, lexical diversity and density appear to be positively correlated with the success of the campaign.

Furthermore, it emerges from the results that the *About Words* variable can also explain the relationship studied in this regression since it is positively correlated and significant at 100% confidence interval. Therefore, we can deduce that the higher the number of words contained in a project description, the more likely the project is to be funded. We can establish the same observation for the variable *Experience* which has a positive coefficient and is significant at 100% confidence interval. This means that a campaign submitted by an entrepreneur who has already launched a project on Kickstarter in the past is more likely to reach its funding goal. The *Optimism* variable is significant at 95% confidence interval and negatively correlated to the dependent variable of this regression, indicating that the more positive the entrepreneur's attitude tends to be, the less likely the campaign is to succeed. Finally, we observe that many project categories have a significant impact on the success of a campaign, whether positive or negative. Among the 15 categories, 7 are significant at 100% confidence interval, 2 at 99.9%, 1 at 95% and 1 at 90%.

6.2.2 Discussion of Results

We will now analyze the significant variables in our regression models and compare them to previous literature.

Independent Variables

The variable *Lexical Sophistication* appears in our regression models to be highly significant and negatively correlated with the funding success of a crowdfunding campaign. This means that the higher the proportion of rare words in a project description, the less likely the campaign is to reach its funding goal. This observation is in line with the existing literature. Adamska-Mieruszewska and al. (2021) provide an explanation for this observation. According to their paper, project descriptions that contain a high proportion of rare words are more difficult to read, which decreases the chances of successful funding. Joenväärä and al. (2019) agree with the results proposed by Adamska-Mieruszewska and al. (2021) but provide a different interpretation to this phenomenon. They suggest that the use of more sophisticated language would be associated with deceptive behavior of the entrepreneur. Indeed, it would seem that less qualified entrepreneurs try to express themselves in a more sophisticated manner to mask their lack of professionalism.

The *Lexical Diversity* variable represents the proportion of different words used by the entrepreneur to describe his project. The results of our study suggest that this variable is positively correlated with campaign success, which coincides with previous findings in the literature. Indeed, Gorbatai and Nelson (2015) demonstrated that lexical diversity has the power to influence investor decision-making and is positively correlated with online fundraising success. Joenväärä and al. (2019) confirm this positive relationship by stating that lexical diversity provides investors with an indication of the entrepreneur's professionalism, which promotes their contribution. However, the results of our study indicate that lexical diversity may also be negatively correlated with some variables representing the success rate of a campaign. This observation may be explained by the fact that too much diversity may hinder the readability of the project description. According to Zhou and al. (2018), crowdfunding projects are usually written in informal language that is not very diverse and therefore easy to understand. Based on this statement, we can assume that a high level of lexical diversity reduces readability and therefore decreases the likelihood of successful funding. In addition, it is important to note that crowdfunding is a financing alternative that allows to raise funds from a large public, not necessarily native English speakers, who may therefore struggle to understand a too diversified vocabulary.

The *Lexical Density* variable appears to be positively correlated with the outcome of a crowdfunding campaign. This finding can be explained by the fact that lexical density reflects the learner's language proficiency (Kondal, 2015). Indeed, in order to write a lexically dense pitch, the entrepreneur should have good linguistic skills, but above all should be able to master them. Therefore, we can deduce that a high level of lexical density is associated with a good linguistic performance of the entrepreneur, which is perceived by the backers as a signal of quality.

Control Variables

The *About Words* variable is positive, which is consistent with our expectations. Indeed, we assume that a project description with a large number of words is more detailed and provides more information. Therefore, it reduces the information asymmetry and increases potential investors' confidence and willingness to contribute. This is supported by Chen et al. (2020) who found that disclosure does indeed have a significant impact on the probability of funding. This finding also supports the study conducted by Zhou and al. (2018) which shows that the length of the text description positively impacts the success of a campaign.

The *Risk Words* variable also has a positive coefficient, which is in line with our expectations. Indeed, we expect that a longer « Risks Challenges » section will provide more information about the risks associated with the project and therefore reduce the uncertainty of potential investors. This observation confirms prior results which show that although backers are less supportive of high-risk projects since the introduction of this section, a longer risk description helps to minimize this effect (Madsen and McMullin, 2018).

The *Readability* variable is positively correlated with fundraising outcome. This corresponds to the findings in the previous literature. Indeed, Adamska-Mieruszevska and al. (2021) concluded that text readability has a significant influence on crowdfunding success. Our third regression suggests that better text readability decreases the duration of a campaign, which makes sense since a readable text increases the ease of understanding by the crowd and speeds up their decision making.

The *Experience* variable shows a positive relationship with the success rate of a crowdfunding campaign. This means that if project owner has successfully funded one or more campaigns on the Kickstarter platform in the past, chances of funding of the current project are higher. This finding is quite logical as the entrepreneur's past experience demonstrates his professionalism and reliability (Zhou and al., 2018). In addition, this observation could also indicate that the project owner may

have created a network of backers on Kickstarter in previous campaigns, which could accelerate the funding process and therefore reduce the length of the campaign. This is consistent with the results of our third regression.

The *Google Trend* variable is negatively correlated with the success of the campaign. This indicates that the less the platform's name is searched on Google, the more likely the project is to reach its funding goal. This contrasts with previous findings in the literature that show that campaigns actually benefit from the popularity of the platform on which they are launched (Liu and al. 2022). However, this result can be explained by the fact that a higher visibility of the page leads to more competition, which works against the campaign. Indeed, Chen (2021) has shown that the higher the competition, the less successful the project will be.

The *Optimism* variable has a negative effect on the success of a campaign, a relatively surprising finding that contrasts with the study conducted by Wang and al (2017). Indeed, it is expected that the feeling of optimism released by the entrepreneur demonstrates some confidence about the quality and seriousness of his project (Mitra and Gilbert, 2014). However, this finding is in line with the results obtained by Zhou and al. (2018) and Parhankangas and Ehrlich (2014) which suggest that an excessive feeling of optimism from the entrepreneur can reduce the credibility of his project and consequently negatively affect the outcome of the fundraising.

The *Category* variable impacts the variables representing the success of a campaign significantly, sometimes positively and sometimes negatively. This can be explained in various ways. First, it may be that some project categories require a lower minimum funding amount, which will make it easier to reach the funding goal and therefore ensure project success (Zhou and al., 2018). Second, some categories may be more popular than others among backers, such as those that generally deliver concrete products as a reward. As a result, projects in these categories are more likely to meet the campaign's funding goal.

6.2.3 Research Question

As a reminder, this paper is driven by the following research question: *Does lexical complexity have an impact on the success rate of crowdfunding campaigns?* The results obtained through this research and presented in the previous section demonstrate that the lexical complexity of project descriptions has a real impact on the outcome of crowdfunding campaigns, which allows us to confirm our initial reasoning.

Indeed, the regressions performed in this report show that the lexical sophistication, diversity, and density are three components that can significantly explain the final outcome of a project. Since these three components defined the lexical complexity as a whole, we extrapolate the results obtained and we draw the conclusion that lexical complexity influences the success rate of crowdfunding campaigns. Although we notice that the third regression does not support our observation, this does not contradict it. It may simply be that the duration of a campaign was not one of the most relevant variables to represent the level of success of a project.

7 Conclusion

The significant growth of the crowdfunding justifies the importance of the research on this subject. In recent years, many studies have been carried out in the literature with the aim of understanding the underlying mechanism of this financing alternative and identifying the factors associated with its success. However, although the existing literature has demonstrated its importance, it appears that the content and especially the linguistic characteristics of project descriptions have been somewhat ignored. The purpose of this study is therefore to fill this gap by highlighting the importance of the linguistic aspect of textual descriptions in the outcome of the campaigns and contributing to the understanding of crowdfunding.

Based on a dataset collected from crowdfunding platform Kickstarter, we study the impact of lexical complexity on the success rate of crowdfunding campaigns. In order to conduct this study, we rely on the three components that characterize lexical complexity, namely lexical sophistication, diversity, and density, and we identify four variables that represent the success of a project. We then analyze the potential relationship between these three linguistic components and the success variables using linear and logistic regression models. The results confirm our initial reasoning and show that the lexical complexity of project descriptions does indeed have an impact on the outcome of campaigns. Moreover, this report highlights the fact that lexical complexity, through its three components, can both encourage and discourage the contribution of potential funders. Indeed, on the one hand, lexical complexity can reflect the professionalism and proficiency of the project owner, which increases the willingness of backers to contribute, and on the other hand, lexical complexity can reduce the readability of the description and thus the understanding of backers, which decreases their willingness to contribute.

However, this research is subject to some limitations. Firstly, the data sample used to carry out this research has been collected on only one crowdfunding platform, Kickstarter. There are other well-known platforms, such as Indiegogo or KissKissBankBank, which operate differently or are based on other crowdfunding models. Therefore, the generalization of our findings is slightly limited. Secondly, our analysis focuses mainly on the project descriptions, that have been carefully prepared before the launch of the campaign. There may be other, more spontaneous sources of information, such as comments, which better reflect the linguistic skills of the project owner. Thirdly, this study extrapolates the results obtained for the lexical sophistication, diversity, and density to the notion of lexical complexity. It might be appropriate to summarize the information contained in these three components into a unique variable in order to obtain a result which directly represents the influence of the lexical complexity.

Future research can be performed in order to validate the conclusions drawn in this report. First of all, additional research can be conducted to address the limitations of our study. Then, a similar study could be conducted using a sample of data collected from multiple crowdfunding platforms. Indeed, it may be that backers who participate in other platforms, based on other crowdfunding models, have different motivations when making their investment decision. Finally, our results suggest that there is an optimal level of lexical complexity in project descriptions. Therefore, a future research could further investigate our results by hypothesizing that there is a positive but curvilinear relationship between the degree of lexical complexity and the level of funding of crowdfunding campaigns.

8 Bibliography

- Adamska-Mieruszewska, J., Mrzygłód, U., Suchanek, M., Fornalska-Skurczyńska, A. (2021). Keep it simple. The impact of language on crowdfunding success. *Economics Sociology*, 14(1), 130-144.
- Agrawal, A. K., Catalini, C., Goldfarb, A. (2011). The geography of crowdfunding (No. w16820). *National bureau of economic research*.
- Belleflamme, P., Lambert, T., Schwienbacher, A. (2013). Individual crowdfunding practices. *Venture Capital*, 15(4), 313-333.
- Bulté, B., Housen, A. (2012). Defining and operationalising L2 complexity. Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA, 23-46.
- Cappa, F., Pinelli, M., Maiolini, R., Leone, M. I. (2021). “Pledge” me your ears! The role of narratives and narrator experience in explaining crowdfunding success. *Small Business Economics*, 57(2), 953-973.
- Chen, L. (2021). Investigating the Impact of Competition and Incentive Design on Performance of Crowdfunding Projects: A Case of Independent Movies. *J. Theor. Appl. Electron. Commer. Res.*, 16, 791-810.
- Chen, S., Thomas, S., Kohli, C. (2016). What really makes a promotional campaign succeed on a crowdfunding platform?: Guilt, utilitarian products, emotional messaging, and fewer but meaningful rewards drive donations. *Journal of Advertising Research*, 56(1), 81-94.
- Chen, X., Huang, B., Shaban, M. (2020). Naïve or sophisticated? Information disclosure and investment decisions in peer-to-peer lending. *Journal of Corporate Finance*, 101805.
- Das, E., Kerkhof, P., Kuiper, J. (2008). Improving the effectiveness of fundraising messages: The impact of charity goal attainment, message framing, and evidence on persuasion. *Journal of Applied Communication Research*, 36(2), 161-175.
- Gallagher, K. M., Updegraff, J. A. (2012). Health message framing effects on attitudes, intentions, and behavior: a meta-analytic review. *Annals of behavioral medicine : a publication of the Society of Behavioral Medicine*, 43(1), 101–116.
- Gorbatai, A.D., Nelson, L. (2015). Gender and the Language of Crowdfunding.
- Joenväärä, J., Tiu, C., Karppinen, J. (2019). What do hedge funds say?
- Johansson, V. (2008). Lexical diversity and lexical density in speech and writing: A developmental perspective. *Working papers/Lund University, Department of Linguistics and Phonetics*, 53, 61-79.
- Khelladi, I., Castellano, S. Bertaud, A. (2018). Les déterminants de la confiance de l'internaute dans le contexte d'une collecte de crowdfunding. *Innovations*, 56, 117-140.

- Kim, J. (2014). Predicting L 2 Writing Proficiency Using Linguistic Complexity Measures : A Corpus-Based Study.
- Kondal, B. (2015). Effects of Lexical Density and Lexical Variety in Language Performance and Proficiency.
- Larrimore, L., Jiang, L., Larrimore, J., Markowitz, D.M., Gorski, S. (2011). Peer to Peer Lending: The Relationship Between Language Features, Trustworthiness, and Persuasion Success. *Journal of Applied Communication Research*, 39, 19 - 37.
- Liu, J., Yang, L., Wang, Z., Hahn, J. (2015). Winner Takes All? The "Blockbuster Effect" in Crowdfunding Platforms Research-in-Progress.
- Madsen, J.M., McMullin, J.L. (2018). Economic Consequences of Risk and Ability Disclosures: Evidence From Crowdfunding. *ERPNI: Entrepreneurs (Finance) (Topic)*.
- Mitra, T., Gilbert, E. (2014, February). The language that gets people to give: Phrases that predict success on kickstarter. *In Proceedings of the 17th ACM conference on Computer supported cooperative work social computing* (pp. 49-61).
- Moffitt, K.C., Burns, M.B. (2009). What Does That Mean? Investigating Obfuscation and Readability Cues as Indicators of Deception in Fraudulent Financial Reports. *AMCIS*.
- Mollick, E. (2014). The dynamics of crowdfunding: An exploratory study. *Journal of business venturing*, 29(1), 1-16.
- Moradi, M., Dass, M. (2019). An investigation into the effects of message framing on crowdfunding funding level. *Journal of Electronic Commerce Research*, 20(4), 238-254.
- O'Dell, F., Read, J., McCarthy, M. (2000). Assessing vocabulary. *Cambridge university press*.
- Parhankangas, A., Ehrlich, M. (2014). How entrepreneurs seduce business angels: : An impression management approach. *Journal of Business Venturing*, 29, 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.
- Petruzzelli, A.M., Natalicchio, A., Panniello, U., Roma, P. (2019). Understanding the crowdfunding phenomenon and its implications for sustainability. *Technological Forecasting and Social Change*.
- Rothman, A. J., Salovey, P. (1997). Shaping perceptions to motivate healthy behavior: the role of message framing. *Psychological bulletin*, 121(1), 3-19.
- Rothman, A. J., Salovey, P., Antone, C., Keough, K., Martin, C. D. (1993). The influence of message framing on intentions to perform health behaviors. *Journal of Experimental Social Psychology*, 29(5), 408-433.

Schwiebacher, A., Larralde, B. (2010). Crowdfunding of small entrepreneurial ventures. *Handbook of entrepreneurial finance*, Oxford University Press, Forthcoming.

Wang, W., Zhu, K., Wang, H., Wu, Y.J. (2017). The Impact of Sentiment Orientations on Successful Crowdfunding Campaigns through Text Analytics. *IET Softw.*, 11, 229-238.

Zhao, L., Ryu, S. (2020). Reward-based crowdfunding research and practice. *Advances in crowdfunding*, 119-143.

Zhou, M., Lu, B., Fan, W. et al. Project description and crowdfunding success: an exploratory study. *Inf Syst Front* 20, 259–274 (2018).

9 Appendix

Figure 1 : Variables of Lexical Complexity

Lexical Density	
<i>LD</i>	Lexical Density
Lexical Sophistication	
<i>LS1</i>	Lexical Sophistication – I
<i>LS2</i>	Lexical Sophistication – II
<i>VS1</i>	Verb Sophistication – I
<i>CVS1</i>	Corrected Verb Sophistication – I
<i>VS2</i>	Verb Sophistication - II
Lexical Diversity	
<i>NDW</i>	Number of Different Words
<i>NDWZ</i>	Number of Different Words in the first words of sample
<i>NDWERZ</i>	Number of Different Words Expected Random in the first words of sample
<i>NDWESZ</i>	Number of Different Words Expected Sequence in the first words of sample
<i>TTR</i>	Type-Token Ratio
<i>MSTTR</i>	Mean Segmental Type-Token Ratio
<i>CTTR</i>	Corrected Type-Token Ratio
<i>RTTR</i>	Root Type-Token Ratio
<i>LOGTTR</i>	Bilogarithmic Type-Token Ratio
<i>UBER</i>	Uber Index
<i>LV</i>	Lexical Word Variation
<i>VV1</i>	Verb Variation – I
<i>SVV1</i>	Squared Verb Variation – I
<i>CVV1</i>	Corrected Verb Variation – I
<i>VV2</i>	Verb Variation – II
<i>NV</i>	Noun Variation
<i>ADJV</i>	Adjective Variation
<i>ADVV</i>	Adverb Variation
<i>MODV</i>	Modifier Variation

Figure 2: Variables Description

Dependent Variables	
<i>USD Pledged*</i>	Amount pledged to the crowdfunding campaign in US dollar.
<i>Backers Count*</i>	Number of backers pledging to support the campaign.
<i>Duration*</i>	Number of days for which the project accepted contributions.
<i>Success</i>	Success of the campaign. Takes the value of 1 if the project succeeded, and 0 otherwise.
Independent Variables	
<i>Lexical Sophistication</i>	Lexical sophistication of the project description measured on the basis of 5 variables.
<i>Lexical Diversity</i>	Lexical diversity of the project description measured on the basis of 19 variables.
<i>Lexical Density</i>	Lexical density of the project description measured on the basis of 1 variables.
Control Variables	
<i>Risk Words</i>	Number of words in the “Risks & Challenges” section of the pitch provided by the entrepreneur.
<i>About Words</i>	Number of words in the “About this project” section of the pitch provided by the entrepreneur.
<i>Readability</i>	Readability of the pitch based on the Flesch Reading Ease Score.
<i>Location Country**</i>	Code of the country where the entrepreneur is located.
<i>Experience</i>	Number of Kickstarter campaigns launched by the entrepreneur previously.
<i>Google Trend</i>	Measure of popularity of Kickstarter based on monthly Google searches.
<i>Category**</i>	Dummy variables for each Kickstarter category.
<i>Us Project</i>	Takes the value of 1 if the campaign was launched in the US, 0 otherwise
<i>Optimism</i>	Sentiment reflecting the entrepreneur’s attitude and state of mind.

Notes: *Natural logarithm of the values are used in the regression analyses.

** Variables expressed as a factor in the regressions analyses.

Table 1: Regression Results

	<i>Dependent variables:</i>			
	Regression 1	Regression 2	Regression 3	Regression 4
	<i>USD Pledged</i>	<i>Backers Count</i>	<i>Duration</i>	<i>Success</i>
lexical.sophistication	−0.350*** (0.028)	−0.197*** (0.016)	−0.001 (0.004)	−0.262*** (0.024)
diversity.pca	−0.114*** (0.024)	−0.051*** (0.014)	−0.005 (0.003)	0.041** (0.021)
ld	4.603*** (0.764)	2.435*** (0.442)	−0.038 (0.105)	4.966*** (0.740)
RiskWords	0.004*** (0.0004)	0.002*** (0.0003)	0.0001** (0.0001)	0.0004 (0.0004)
AbtWords	0.002*** (0.0002)	0.001*** (0.0001)	−0.00005* (0.00003)	0.002*** (0.0002)
AbtSentence	−0.010*** (0.003)	−0.006*** (0.002)	0.001* (0.0004)	−0.018*** (0.003)
Readability	0.003 (0.002)	0.004*** (0.001)	−0.001*** (0.0003)	0.003 (0.002)
factor(location_country)AN	1.286 (3.948)	−0.510 (2.284)	0.549 (0.540)	−26.878 (2,058.243)
factor(location_country)AT	0.099 (3.006)	0.132 (1.739)	0.205 (0.411)	−12.116 (1,455.398)
factor(location_country)AU	−0.531 (2.834)	−0.181 (1.640)	0.099 (0.388)	−12.296 (1,455.398)
factor(location_country)BE	−0.494 (2.900)	−0.221 (1.678)	0.008 (0.397)	−13.104 (1,455.398)
factor(location_country)BG	−1.233 (3.424)	−1.178 (1.981)	0.075 (0.468)	−26.688 (1,780.988)
factor(location_country)BR	−1.334 (3.428)	−0.865 (1.984)	0.272 (0.469)	−27.871 (1,724.810)
factor(location_country)BS	−0.099 (3.431)	0.099 (1.985)	0.178 (0.469)	−11.094 (1,455.398)
factor(location_country)BY	0.722 (3.950)	−0.544 (2.286)	0.468 (0.540)	−27.513 (2,058.243)
factor(location_country)CA	0.365 (2.827)	0.125 (1.636)	0.113 (0.387)	−12.391 (1,455.398)
factor(location_country)CH	2.471 (3.009)	0.671 (1.741)	−0.044 (0.412)	−11.588 (1,455.398)
factor(location_country)CN	3.386 (3.044)	2.470 (1.761)	0.113 (0.416)	3.258 (1,563.203)
factor(location_country)CO	−0.266 (3.243)	−0.451 (1.876)	0.019 (0.444)	−26.717 (1,673.645)

	<i>Dependent variables:</i>			
	Regression 1	Regression 2	Regression 3	Regression 4
	<i>USD Pledged</i>	<i>Backers Count</i>	<i>Duration</i>	<i>Success</i>
factor(location_country)CR	−2.640 (3.942)	−1.082 (2.281)	0.004 (0.539)	−28.694 (2,058.243)
factor(location_country)CU	−0.398 (3.246)	−0.326 (1.878)	0.152 (0.444)	−11.958 (1,455.398)
factor(location_country)CZ	−0.314 (3.141)	−0.411 (1.817)	0.124 (0.430)	−27.412 (1,607.634)
factor(location_country)DE	−0.066 (2.841)	−0.164 (1.644)	0.095 (0.389)	−12.705 (1,455.398)
factor(location_country)DK	0.949 (2.933)	0.393 (1.697)	0.117 (0.401)	−11.744 (1,455.398)
factor(location_country)DO	−3.380 (3.445)	−2.014 (1.993)	0.072 (0.471)	−27.018 (1,753.608)
factor(location_country)EC	2.353 (3.942)	2.040 (2.281)	0.255 (0.539)	3.366 (2,058.243)
factor(location_country)EE	−1.323 (3.436)	−0.584 (1.988)	−0.296 (0.470)	−26.281 (1,774.392)
factor(location_country)ES	−0.554 (2.871)	−0.284 (1.661)	−0.004 (0.393)	−12.818 (1,455.398)
factor(location_country)FR	0.971 (2.859)	0.430 (1.654)	0.061 (0.391)	−11.644 (1,455.398)
factor(location_country)GB	0.273 (2.823)	0.134 (1.633)	0.088 (0.386)	−12.188 (1,455.398)
factor(location_country)GE	−3.852 (3.957)	−0.926 (2.289)	−0.018 (0.541)	−25.966 (2,058.243)
factor(location_country)GR	1.037 (3.244)	0.539 (1.877)	0.342 (0.444)	−12.323 (1,455.398)
factor(location_country)GT	−0.186 (3.954)	−0.444 (2.287)	0.529 (0.541)	−26.592 (2,058.243)
factor(location_country)HK	1.146 (2.905)	0.738 (1.681)	0.155 (0.397)	−10.671 (1,455.398)
factor(location_country)HN	−5.178 (3.961)	−2.479 (2.292)	0.803 (0.542)	−28.921 (2,058.243)
factor(location_country)HR	−3.022 (3.435)	−1.355 (1.988)	−0.314 (0.470)	−27.219 (1,781.916)
factor(location_country)HT	1.853 (3.434)	0.241 (1.987)	0.063 (0.470)	−11.402 (1,455.398)

	<i>Dependent variables:</i>			
	Regression 1	Regression 2	Regression 3	Regression 4
	<i>USD Pledged</i>	<i>Backers Count</i>	<i>Duration</i>	<i>Success</i>
factor(location_country)HU	1.299 (3.241)	0.224 (1.875)	-0.694 (0.443)	-10.315 (1,455.398)
factor(location_country>ID	-0.487 (3.433)	-0.340 (1.987)	0.441 (0.470)	-11.237 (1,455.398)
factor(location_country)IE	-0.073 (2.924)	0.081 (1.692)	0.187 (0.400)	-11.678 (1,455.398)
factor(location_country)IL	0.923 (3.086)	0.344 (1.785)	0.128 (0.422)	-27.051 (1,589.259)
factor(location_country)IN	1.971 (3.084)	0.875 (1.784)	0.379 (0.422)	-11.704 (1,455.398)
factor(location_country)IS	2.246 (3.954)	1.422 (2.288)	0.260 (0.541)	3.454 (2,058.243)
factor(location_country)IT	-0.396 (2.843)	-0.269 (1.645)	0.163 (0.389)	-12.573 (1,455.398)
factor(location_country)JP	1.287 (2.884)	0.260 (1.669)	0.165 (0.395)	-11.966 (1,455.398)
factor(location_country)KE	-5.033 (3.966)	-2.315 (2.294)	0.720 (0.543)	-26.113 (2,058.243)
factor(location_country)KH	2.655 (3.427)	1.262 (1.983)	-0.102 (0.469)	-11.145 (1,455.398)
factor(location_country)LR	0.284 (3.945)	-1.211 (2.283)	0.464 (0.540)	-27.008 (2,058.243)
factor(location_country)LT	1.136 (3.433)	0.965 (1.986)	0.412 (0.470)	3.656 (1,777.504)
factor(location_country)MC	0.929 (3.949)	-0.226 (2.285)	0.102 (0.540)	-27.581 (2,058.243)
factor(location_country)MG	3.080 (3.940)	0.358 (2.280)	-0.046 (0.539)	-26.414 (2,058.243)
factor(location_country)MM	-0.803 (3.430)	-0.320 (1.984)	0.486 (0.469)	-11.160 (1,455.398)
factor(location_country)MT	-1.204 (3.430)	-1.113 (1.985)	-0.260 (0.469)	-26.911 (1,781.318)
factor(location_country)MX	-1.532 (2.873)	-0.578 (1.662)	0.223 (0.393)	-12.501 (1,455.398)
factor(location_country)NE	1.964 (3.946)	2.574 (2.283)	0.083 (0.540)	1.772 (2,058.243)
factor(location_country)NG	2.676 (3.944)	-1.264 (2.282)	0.774 (0.540)	-27.300 (2,058.243)

	<i>Dependent variables:</i>			
	Regression 1	Regression 2	Regression 3	Regression 4
	<i>USD Pledged</i>	<i>Backers Count</i>	<i>Duration</i>	<i>Success</i>
factor(location_country)NI	-4.675 (3.954)	-1.981 (2.287)	0.037 (0.541)	-26.654 (2,058.243)
factor(location_country>NL	0.212 (2.857)	-0.147 (1.653)	0.159 (0.391)	-12.814 (1,455.398)
factor(location_country)NO	-0.289 (2.946)	-0.337 (1.704)	0.033 (0.403)	-13.780 (1,455.398)
factor(location_country)NP	3.778 (3.943)	2.880 (2.281)	0.095 (0.539)	3.705 (2,058.243)
factor(location_country)NZ	-0.184 (2.871)	-0.083 (1.661)	0.010 (0.393)	-12.524 (1,455.398)
factor(location_country)PE	3.112 (3.429)	1.486 (1.984)	-0.309 (0.469)	-11.764 (1,455.398)
factor(location_country)PH	-1.461 (3.243)	-1.043 (1.876)	-0.444 (0.444)	-12.993 (1,455.398)
factor(location_country)PL	-3.821 (3.435)	-1.346 (1.988)	-0.692 (0.470)	-26.111 (1,724.853)
factor(location_country)PR	-2.366 (3.141)	-1.232 (1.818)	0.286 (0.430)	-27.192 (1,591.168)
factor(location_country)PS	-2.224 (3.955)	-1.519 (2.289)	0.520 (0.541)	-27.321 (2,058.243)
factor(location_country)PT	-0.964 (3.244)	-0.371 (1.877)	-0.581 (0.444)	-12.390 (1,455.398)
factor(location_country)RO	2.947 (3.951)	1.764 (2.286)	0.163 (0.541)	2.768 (2,058.243)
factor(location_country)RU	-0.852 (3.079)	-0.270 (1.782)	0.159 (0.421)	-12.290 (1,455.398)
factor(location_country)RW	-7.323* (3.945)	-3.634 (2.282)	-0.067 (0.540)	-27.546 (2,058.243)
factor(location_country)SE	1.099 (2.888)	0.965 (1.671)	0.061 (0.395)	-11.901 (1,455.398)
factor(location_country)SG	1.855 (2.909)	0.978 (1.683)	-0.055 (0.398)	-10.803 (1,455.398)
factor(location_country)SI	2.656 (3.435)	1.607 (1.987)	0.266 (0.470)	-11.065 (1,455.398)
factor(location_country)SJ	2.236 (3.954)	1.478 (2.288)	0.095 (0.541)	3.195 (2,058.243)
factor(location_country)SN	2.580 (3.427)	1.555 (1.983)	-0.055 (0.469)	2.572 (1,746.945)
factor(location_country)TH	-0.762 (3.242)	-0.226 (1.876)	-0.267 (0.444)	-12.129 (1,455.398)

	<i>Dependent variables:</i>			
	Regression 1	Regression 2	Regression 3	Regression 4
	<i>USD Pledged</i>	<i>Backers Count</i>	<i>Duration</i>	<i>Success</i>
factor(location_country)TR	-3.584 (3.951)	-1.864 (2.286)	0.724 (0.541)	-30.488 (2,058.243)
factor(location_country)TW	3.624 (3.143)	2.722 (1.819)	0.427 (0.430)	-10.459 (1,455.398)
factor(location_country)UA	-1.561 (3.140)	-1.132 (1.817)	-0.005 (0.430)	-12.198 (1,455.398)
factor(location_country)UG	2.698 (3.955)	1.554 (2.288)	-0.714 (0.541)	4.059 (2,058.243)
factor(location_country)US	0.547 (2.822)	0.267 (1.633)	0.100 (0.386)	-12.215 (1,455.398)
factor(location_country)UY	1.276 (3.950)	1.297 (2.285)	0.781 (0.540)	-27.078 (2,058.243)
factor(location_country)VN	1.567 (3.944)	-0.353 (2.282)	-0.349 (0.540)	2.781 (2,058.243)
factor(location_country)ZA	3.256 (3.422)	1.095 (1.980)	0.065 (0.468)	-11.679 (1,455.398)
experience	0.101*** (0.019)	0.084*** (0.011)	-0.018*** (0.003)	0.440*** (0.038)
google_trend	-0.008*** (0.002)	-0.004*** (0.001)	0.001 (0.0003)	-0.003 (0.002)
factor(category)comics	0.788*** (0.255)	0.893*** (0.148)	0.044 (0.035)	0.444** (0.212)
factor(category)crafts	-1.042*** (0.281)	-0.580*** (0.163)	0.047 (0.038)	-1.064*** (0.251)
factor(category)dance	0.748* (0.389)	0.335 (0.225)	0.005 (0.053)	0.908*** (0.324)
factor(category)design	0.927*** (0.201)	0.691*** (0.116)	0.120*** (0.028)	-0.311* (0.161)
factor(category)fashion	-0.519** (0.205)	-0.259** (0.119)	0.031 (0.028)	-0.723*** (0.170)
factor(category)film -0.075	video (0.174)	0.262 (0.101)	0.101 (0.024)	0.035 (0.139)
factor(category)food	-0.155 (0.209)	-0.077 (0.121)	0.078*** (0.029)	-0.584*** (0.173)
factor(category)games	0.399** (0.195)	0.636*** (0.113)	0.014 (0.027)	-0.478*** (0.159)
factor(category)journalism	-1.165*** (0.379)	-0.450** (0.219)	0.081 (0.052)	-1.463*** (0.408)

	<i>Dependent variables:</i>			
	Regression 1 <i>USD Pledged</i>	Regression 2 <i>Backers Count</i>	Regression 3 <i>Duration</i>	Regression 4 <i>Success</i>
factor(category)music	1.056*** (0.183)	0.695*** (0.106)	0.072*** (0.025)	0.811*** (0.144)
factor(category)photography	0.021 (0.275)	0.091 (0.159)	0.015 (0.038)	-0.207 (0.220)
factor(category)publishing	-0.232 (0.185)	-0.030 (0.107)	0.052** (0.025)	-0.200 (0.148)
factor(category)technology	-0.467** (0.195)	-0.166 (0.113)	0.152*** (0.027)	-1.330*** (0.171)
factor(category)theater	1.264*** (0.281)	0.690*** (0.163)	-0.046 (0.039)	1.039*** (0.228)
US_Project				
Optimism	-1.366 (2.933)	0.112 (1.697)	0.248 (0.401)	-5.929** (2.511)
Constant	2.283 (2.869)	0.529 (1.660)	3.322*** (0.393)	8.931 (1,455.398)
Observations	4,991	4,991	4,991	4,991
R ²	0.289	0.308	0.054	
Pseudo R ²				0.1748
Adjusted R ²	0.274	0.294	0.034	
Log Likelihood				-2,797.334
Akaike Inf. Crit.				5,798.667
Residual Std. Error (df = 4889)	2.758	1.596	0.377	
F Statistic (df = 101; 4889)	19.669***	21.525***	2.748***	

Note:

*p<0.1; **p<0.05; ***p<0.01

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
Louvain School of Management

Place des Doyens, 1 bte L2.01.01, 1348 Louvain-la-Neuve
Boulevard Emile Devreux 6, 6000 Charleroi, Belgique
Chaussée de Binche 151, 7000 Mons, Belgique

www.uclouvain.be/lsm