[page de garde]

HEC MONTRÉAL

The Early Bird Gets the Worm: Underlying Measures of Success in Crowdfunding

par

Marc-André Poliquin

Jean-François Plante

HEC Montréal

Directeur de recherche

Sciences de la gestion

Spécialisation Science des données et Analytique d'affaires

Mémoire présenté en vue de l'obtention

Du grade de maîtrise ès sciences en gestion

(M. Sc.)

Septembre 2022

Résumé

La montée en popularité des réseaux de financement participatif qui permettent aux individus de rechercher des fonds auprès de la communauté en ligne souligne l'importance que les méthodes de financement alternatives commencent à avoir dans la société actuelle. En tant que tel, la détermination des facteurs qui définissent le succès est particulièrement pertinente. Plus précisément, trois facteurs ressortent comme des questions de recherche pertinentes à explorer sur la base de la littérature existante : (1) la fixation d'objectifs appropriés et l'importance d'un financement précoce, (2) la détermination des caractéristiques clés des campagnes qui suivent des trajectoires distinctes de montants promis et de bailleurs de fonds dans le temps, et (3) la présence et l'ampleur de l'impact du fardeau de l'étranger dans les réseaux de financement participatif. Les résultats de notre analyse indiquent que la fixation d'objectifs réalistes et l'atteinte de ces objectifs lors des premiers jours de financement sont des facteurs importants dans le résultat d'une campagne de financement participatif. Les chances de succès augmentent de manière significative plus tôt une campagne atteint 20% de son objectif de financement, la première semaine étant une étape clé et la deuxième étant un point de non-retour pour la plupart si le 20% n’est pas accumulé à ce point dans le temps. La classification fonctionnelle est présentée comme une méthode permettant de classer les courbes chronologiques de divers indicateurs clés de performance, tels que le financement quotidien, le nombre de soutiens, le pourcentage de l'objectif atteint et le ratio de dollar par jour. Nous constatons que les profils de financement des campagnes suivent généralement des courbes similaires avec un élan initial qui s'estompe rapidement, mais le regroupement fonctionnel a permis d'identifier différents sous-groupes de campagnes ayant des compositions et niveaux de réussite variables. Les campagnes réussies conservent leur élan initial un peu plus longtemps et bénéficient d'une impulsion plus forte fin de phase, liée à une meilleure aptitude sociale. Cela permet à ces campagnes d'atteindre leurs objectifs plus rapidement et plus efficacement. Nous avons ensuite analysé si les étrangers dans le réseau de financement participatif sont désavantagés en mesurant le biais induit par les variables économiques, géographiques et culturelles. Plus précisément, nous avons mesuré l'impact de la devise, du continent d'origine et de la langue maternelle d'une campagne sur ses chances de succès. Nous avons constaté que l'utilisation de devises étrangères à la plateforme agissait comme un handicap dans le réseau de financement participatif. En outre, l'analyse de la variable d'interaction entre la monnaie et le continent a montré que l'utilisation de l'USD sur d'autres continents était associée à une augmentation des chances de succès.

Mots-clés:

Fixation d’objectifs, Financement précoce, Développement d’étapes-clés, Classification de données temporelles, Fardeau de l'étranger

Méthodes de recherche:

Régression logistique, Modélisation de séries temporelles, Inférence, Classification fonctionnelle

Abstract

The rise in popularity of crowdfunding networks that allow individuals to seek funding from the online community highlights the importance that alternative funding methods are starting to have in present society. As such, figuring out the factors that define success is of particular relevance. Specifically, three factors stand out as relevant research questions to explore based on existing literature: (1) proper goal setting and the importance of early funding, (2) key characteristics of campaigns following distinct pledge and backer trajectories over time, and (3) the presence and extent of the impact of liability of foreignness in crowdfunding networks. The results of our analysis indicate that setting realistic goals and reaching early funding milestones are important factors in a crowdfunding campaign's outcome. Odds of success increase significantly the earlier a campaign reaches 20% of its funding goal, with the first week being a key milestone and the second being a deal-breaker for most. Functional clustering is introduced as a method to classify the time-series curves of various Key Performance Indicators (KPI) such as daily funding, backer count, percentage of goal reached and dollar per day. We find that campaign funding profiles typically follow similar curves with initial momentum that quickly fades, however functional clustering identified different subgroups of campaigns of varying levels of success. Successful campaigns maintain their initial momentum slightly longer and benefit from stronger end phase boosts that are linked to better social fitness. This allows these campaigns to reach their goals more quickly and efficiently. We then analyzed if foreigners in the crowdfunding network are at a disadvantage by measuring the bias induced by economic, geographic, and cultural variables. Namely, we measured the impact that a campaign's currency, continent of origin and first language have on its odds of success. We found that using foreign currencies foreign to the platform acted as a liability in the crowdfunding network.

Keywords:

Goal Setting, Early Funding, Milestones Benchmarking, Time-Series Clustering, Liability of Foreignness

Research Methods:

Logistic Regression, Time-Series Modelling, Inference, Functional Clustering

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AIC: Akaike Information Criterion

BIC: Bayesian Information Criterion

ICL: Integrated Completed Likelihood criterion

FDA: Functional Data Analysis

KPI: Key Performance Indicator

SVM: Support Vector Machine

KNN: K-Nearest Neighbor

US: United States

UK: United Kingdom

USD: United States Dollar

CAD: Canadian Dollar

GBP: Great Britain Pound

NZD: New Zealand Dollar

Acknowledgements

As I reflect on this research and my entire Master's studies, I realize that this has been a much bigger journey than I had initially conceived. Despite its challenges, I found it to be the most intellectually thrilling and crucial learning experience I have had the opportunity to pursue. Of course, it was all only made possible thanks to the support of my close friends, family and research director. I would like to extend many thanks to them as a token of my gratitude. To my girlfriend Michelle, for keeping me motivated every day and for leading by example so well. To my close friends Francis, Marc-Antoine, Jean-Simon and Samuel for their genuine interest, honest questions and critical opinions that enabled me to polish this work. Finally, to my family, for their relentless support and dedication, and to my research director Jean-François for his thoughtful advice, patience, leadership and continued help explaining tricky topics that were the key drivers in completing this master's thesis. He is truly passionate about the success of his students and a master of his craft, guiding me through the entire research process in a way that allowed me to learn and grow from this experience as I polished the upcoming research to what it is now. I decided to study Data Science as I was eager to find a new challenge and thought that my passion for technology, my unquenchable thirst for knowledge and my desire to understand human beings and the world we live in better would be a perfect fit for this field. I believe this research to be the culmination of years of fascinating work that helped me understand and develop some of the most important concepts and toolbox of the next century, as well as further my critical thinking in ways that I previously did not think imaginable. Data literacy runs much deeper than I thought, and I feel privileged to come out of this with a deep understanding of the perspectives and concepts that will shape the world of tomorrow.

Introduction

Crowdfunding first appeared as a concept in the early 2000s. With the rise of the Internet and the tumultuous economic times that soon followed, it is no surprise that entrepreneurs and innovators soon started looking for alternative sources of funding online. Rather than relying on banks to fund their projects, people could now seek funding from the online community through crowdfunding platforms. Launched in 2009, Kickstarter is one of the leading platforms in the crowdfunding space that people rely on for their creative ventures and business projects. In recent years, hundreds of new crowdfunding platforms have emerged, largely due to the fact that crowdfunding continues to grow in importance today (Fernandez-Blanco et al., 2020). In fact, some specialized platforms have even started to appear for particular types of ventures, such as real estate crowdfunding, which can be seen as potential alternatives to combat rising housing costs and widen access to real estate investments. As technology advances, the crowdfunding landscape continues to evolve. One example is how some newer platforms also integrate blockchain technology to process transactions (Hartmann et al., 2019). Another recent use of crowdfunding that is becoming more and more prevalent is social activism (Diaz & Cacheda, 2016). This shows how crowdfunding platforms can also be vectors for shifts in social paradigms as well as technological innovation. Further, our increased reliance and immersion to social media only adds fuel to the ever-growing fire that drives the growth of these crowdfunding platforms.

With that in mind, the importance of analyzing trends that define success, investor behavior as well as biases in the crowdfunding space becomes increasingly important. As such, the purpose of this study will be to dig deeper into the underlying structures of crowdfunding data to shed light on certain measures of success. This study will focus primarily on three interconnected topics. The first will attempt to develop early funding milestones that campaigners could use as guidelines to estimate their chances of succeeding at key moments of the campaign. The second will classify project profiles through functional clustering and determine key elements related to their success. Finally, the study will attempt to determine if geographical, economic or cultural biases such as language play a role in a project's odds of success.

Literature Review

Crowdfunding platforms have become a popular way for entrepreneurs, creative minds and social activists to raise capital for their various ventures (Fernandez-Blanco et al., 2020). The increasing popularity of crowdfunding platforms in recent years shows its usefulness in raising capital for projects that often simply can't be self-funded. These types of platforms are becoming increasingly popular various types of projects, including social activism (Hartmann et al., 2019), as well as social activism and political ventures for advocatory movements (Diaz & Cacheda, 2016). Kickstarter, one of the prominent platforms online, uses an "all-or-nothing" policy where the pledges received from backers are only collected if the campaign's goal is achieved (Crosetto & Regner, 2014). In other words, the commitment from the backers is always conditional on the project's goal being achieved. Achieving a campaign's goal automatically turns it into a success, while other projects that don't completely fulfill their goal are classified as failed campaigns. With this dynamic in mind, proper goal setting becomes a key element that needs to be considered to increase a campaign's odds of success, especially considering the fact that increasing goal size was negatively associated with success (Mollick, 2014).

Additionally, predicting the success of crowdfunding campaigns on Kickstarter through machine learning has been the subject of previous research. Research on crowdfunding campaign outcome prediction by Kaur et al. (2022) found that logistic regression models and Support Vector Machines (SVM) were among the most accurate to use for accurate predictions of Kickstarter data. The work also analyzes the funding behavior of campaigns and finds that social media plays a determinant role in increasing a campaign's chances of success (Kaur et al., 2022). While this research achieves high accuracy with these methods and displays the importance of social media activity, it does so by including total backer count and social media comments as predictor variables, which are metrics that would only be available once a campaign is over. Other research also supports the importance of social media features, such as "Launch hard or go home" by Etter et al. (2013), in which predictive models were developed and subsequently significantly improved thanks to social media network variables such as Facebook friend count and Twitter tweets. Promotional activities on social media networks have been shown to be an important predictor of success (Li et al., 2016). They do so by playing a crucial role in popularizing projects and growing the campaigner's network through the establishment of multiple new connections (Kaur et al., 2022). Another research on the impact of social media in crowdfunding platforms by Kaur & Gera (2017) confirms the positive effect of social media activity and connection on campaign success.

These studies demonstrate the importance of early pledges and social media in crowdfunding campaigns. Further research shows that the dynamics of pledge behavior on crowdfunding platforms can be split into four categories based on their final funding level to identify characteristics of projects of each category: *Overfunded, Funded, Potential, Low Potential* (Gera et al., 2017). Successful projects falling in the *Overfunded* and *Funded* categories tend to receive a significant portion of their funding goal in the early phases of the campaign (Gera et al., 2017). On the other hand, Kickstarter's in-house statistics claim that 78% of campaigns that reach 20% of their funding goal at any point in their campaign end in success (Kickstarter, 2022). While this is an interesting claim, it doesn't provide the early campaign feedback that campaigners might want to focus on as a predictor of their success. Thus, the first stream investigated in this research aims to provide early funding milestones that can help entrepreneurs and creative individuals understand how their campaigns are performing in their crucial early phases.

The time series data of campaigns provide an important new dimension to analyze. For instance, the gradual accumulation of backers and pledges for individual projects can be graphically represented in hopes of uncovering key characteristics of different campaign profiles. Previous research has successfully used these time-series to make more accurate predictions about a campaign's success (Etter et al., 2013). Important variables have been identified and analyzed in previous research, with a project's *Funding Goal*, *Funding Level*, *Backer Count*, *Category*, *Pledge / Backer* and *Facebook Friend Count* as key variables of interest (Mollick, 2014). Barbi & Bigelli (2017) found multiple instances in the literature where the presence of videos, images and detailed descriptions increased a campaign's chances of success. These types of metrics appear to benefit campaigns by signaling better project quality and preparedness to potential backers (Mollick, 2014). The ability to predict crowdfunding success, while important, does not necessarily inform on the decisiveness of the factors that explain success themselves. Rather than attempt to predict a campaign's success, using this time-series to better understand what allows successful campaigns to achieve their funding goals more efficiently by looking at their distinctive funding profiles could provide a framework for other campaign starters.

Kindler et al. (2021) identifies that success might not be due to virality but rather to social fitness. Mollick's (2014) use of the *Pledge / Backer* variable in his research, which is in fact a calculated variable, provides an interesting case for the use of various KPIs as explanatory variables. Functional clustering methods could be used in this situation since the time series of values for each KPI and campaign is available. These time series could thus be smoothened into their own distinct functions using a spline basis system and then classified into a set of clusters that each have their own distinct characteristics. *The FunFEM* R package developed for this purpose by Bouveyron et al. (2015) to apply functional clustering to the French bike sharing system can be extended to crowdfunding campaigns. As such, the second research stream consists of a functional clustering method for time-series introduced to classify the curves represented by the unique smoothened functions that can be derived from the distinct series of values associated with different Key Performance Indicators (KPI) for each campaign. The goal of this method will be to identify different funding and backer trajectories of campaigns to gain insight about the key characteristics of different types of campaigns, notably based on their success. Previous mobile health data research was able to use this method to identify clusters of patients with different diagnostics to develop cluster-specific therapies (Giordani et al, 2020). Feature engineering is used to create a few additional KPIs to visualize and classify into an optimal number of clusters to identify key characteristics of funding profiles related to the different types of projects.

Further, crowdfunding campaigns, being online ventures, are of international reach to potential backers. This means that despite the geographical distance, backers from all over the world can contribute to a given campaign. While this reduces barriers to potential funding, cultural, geographic and economic variables typically associated with international campaigns could play a role in putting them one step ahead or behind others. Geographic components have been studied previously, showing that the mechanisms of online platforms reduce the economic friction induced by large geographical distances between the project starter and its backers (Agrawal et al., 2011). This research also finds these online mechanisms don't reduce social frictions such as those coming from a campaigner's pre-existing network (Agrawal et al., 2011). Further research by Barbi & Bigelli (2017) found that the mix and concentration of categories associated with the projects echoes the cultural features of the different countries from which they originate. This indicates that there is a potential impact stemming from cultural variables that don't necessarily apply to typical ones sharing the same common attributes that could be seen as default or native to the network. This impact refers to the concept of liability of foreignness, which is novel in crowdfunding research and could shed light on how geographic and cultural biases play into successful funding dynamics.

Liability of foreignness can be defined as the negative impact that being a foreigner or having foreign characteristics can induce in a given situation or network (Zaheer, 1995). This concept was studied in previous research using logistic regression models to quantify the impact in immigrant entrepreneurship situations (Irastorza & Pena, 2013) and in the PGA golf tour (Pastoriza, Plante, Lakhlef, 2021). This is made possible by using an initial model comprised of control variables and adding foreignness factors into the mix to measure their impact on odds ratios of a binary target variable, in this case the campaign's outcome. Given that crowdfunding campaigns are international and borderless by nature but the platform is native to the U.S., the same methodology can be extended to measure the impact of foreignness in crowdfunding networks through variables such as geography, currency and language.

The third research stream analyzes how the concept of liability of foreignness can have an impact on a crowdfunding campaign's odds of success. Geographic components of successful funding in the US have been studied in previous research, mainly finding that the geography of success is quite uneven and that the project mix of a location reflects its cultural background in the US (Mollick, 2014). While Kickstarter is an American platform, it is possible for people from multiple countries to start a campaign, as long as they have an address and bank account from one of Kickstarter's growing list of eligible countries. Research also shows that crowdfunding projects in and outside the US differ, but they have the same determinants of success, particularly the US and UK (Barbi & Bigelli, 2017). Knowing that geography plays an important role in campaign success (Mollick, 2014), one can wonder how being a foreigner on the platform might impact their odds of success.

Given the streams of research mentioned above, the first chapter will first study the importance of early funding milestones and proper goal setting. The second chapter will then look more closely at the time-series of crowdfunding campaigns through the lens of functional data clustering to identify key characteristics of different campaign profiles. Finally, the third chapter will focus on identifying and measuring the potential liability of foreignness in crowdfunding networks using logistic models. However, prior to analysis, dataset acquisition and necessary preparation steps to clean the data of irregularities are detailed below.

Dataset Presentation & Preparation

The datasets used in this research originate from Virginia Tech University and were used by Li et al. (2016) in their research "Project Success Prediction in Crowdfunding Environments". Their data was obtained using the Kickspy web scraper that collected information daily about 18,142 active campaigns on Kickstarter and social media platforms to build a dataset with general information about the campaigns and the time-series of cumulative pledges and backers at a given day. Their data spans 6 months of Kickstarter data from December 2013 to June 2014. The project information dataset contains the general attributes of each project and its creator, including the project's name, category, currency, location, goal amount, video and image count, description word counts, final pledged amount and backer count, as well as social media features such as Facebook connectivity, friend count, and shares. The time-series dataset provides the daily information about each campaign ID in the form of daily cumulative pledges and backers in long format. The complete lists of variables included in the project information and cumulative time-series datasets are reported in Appendix A in the Appendix section.

The average campaigner in the project information has a mean goal of $26,531, along with a final pledged amount of $11,024, 138 backers, and 470 Facebook friends. The maximum amount pledged to a campaign is $6,224,955, while the maximum number of backers obtained is 35,383. In terms of funding, this is very far from the maximum goal of $100M set by the most ambitious campaigner for their project. Most campaigns last 30 days, as shown by the median of 30. Given that there appears to be extreme values in the goal variable, the median can be used to better define a typical campaign. In this initial dataset, the median campaign has a goal of $5,000. Median values for campaigns are the following: $1,722 pledged, 29 backers, and 226 Facebook friends. The dataset has a success rate of 49.8% based on the 9,038 successful campaigns out of 18,142.

Upon inspection, a few steps were required to prepare this dataset for analysis. First, missing values were replaced by 0 for variables when it was implied by their absence i.e., count of Facebook friends if the campaigner did not have a Facebook account. Extreme values were detected upon inspection of the funding goal variable. The minimum goal amount was 100$ as projects with funding goals of less than $100 were removed by Li et al. (2016). However, the maximum goal in the data was $100M, which indicated that some campaigners might have set absurdly high funding goals for themselves. After removing campaigns for which the *Goal* value was extreme by using the 1.5-times Interquartile Range (IQR) Method, a more balanced mix of campaigns was obtained, which is more reflective of typical crowdfunding ventures. The IQR Method uses quartiles to exclude campaigns sitting outside a range that is 1.5x larger than the third quartile and 1.5x smaller than the first quartile to identify extreme values in the goal variable. 1,908 campaigns end up being removed from the dataset from this method. Campaigns for which geocoding information was not available were also excluded.

The summary statistics of this cleaned dataset give a better idea of what a typical campaign really looks like. The new average goal amount in the dataset is $7,240, compared to the $26,531 previously reported. The maximum is now $34,000, while the average final pledged amount and backer count of these campaigns are $6,296 and 101 backers respectively. Campaigners have on average 475 Facebook friends. The success rate of this dataset is 52.7%, slightly higher than before but still a toss-up at approximately 50%. This is driven by the fact that the 1,905 campaigns with extreme funding goal amounts are largely unsuccessful, as reflected by their 25% success rate. This proportion had to be tossed as the scope of this research focuses on campaigns with goals falling in the typical range.

Campaigns can have different durations, but those with a duration of 30 days were selected for analysis for the first two chapters as they are the platform's default duration and have the largest proportion of campaigns opting for this option. 30-Day campaigns have also been identified to be more successful than the ones with longer durations (Barbi & Bigelli, 2017). Furthermore, a common basis for analysis is required to obtain objective proportions in funding milestones and functional data analysis. The resulting dataset contains 2,818 geocoded 30-day campaigns. A version of the resulting dataset with all the possible campaign durations (1 to 60 days) containing 8,658 geocoded campaigns is also kept aside as it is used later in the third chapter of this research on liability of foreignness in crowdfunding networks. This is done to increase the sample size in this chapter since foreign campaigns are of relatively low occurrence to begin with. This also helps provide sufficient observations to look at possible interactions between the variables of foreignness defined for that chapter.

First Chapter:

1. The Importance of Goal Setting and Early Funding Milestones

Crowdfunding campaigns that use the same model as Kickstarter operate with an “all-or-nothing" policy where the backer pledges are only collected if success is reached before the campaign’s deadline (Crosetto & Regner, 2014). While there is a clear value in being able to predict the success of a campaign early on, setting a realistic goal that is achievable could also greatly improve a campaign’s odds of success since it draws the line between getting funded and receiving nothing at all. This also implies that projects that failed to achieve their financial goals might have been viable and might even have succeeded with a slightly lower goal or perhaps a different campaign duration. This is supported by the fact that 1,908 campaign goals were identified to be extreme values and then removed from the original dataset using the Interquartile Range Method (IQR): 75% of these ended up failing to reach their objective.

The figures below show the distribution of funding goal amounts by final outcome (Figure 1.1) and also by category (Figure 1.2) in the project information dataset once extreme goal values were removed.

Figure 1.1 shows that *Successful* campaigns tend to have lower goal amounts than failed ones. The first three quartiles of successful campaigns had a funding goal of less than $10K, ranging from approximately $2K to $8K, while failed campaigns had more varied goal amounts closer to or even surpassing the $10K range. In fact, the typical failed campaign goal ranges from $3K to $12K. Despite lower goals having more success, both groups have some campaigns with large funding goals outside their fourth quartile even after removing extreme values in their fourth quartile, but failed ones remain concentrated farther out near the $30K range. Successful large projects also typically have smaller goals than large, failed projects.

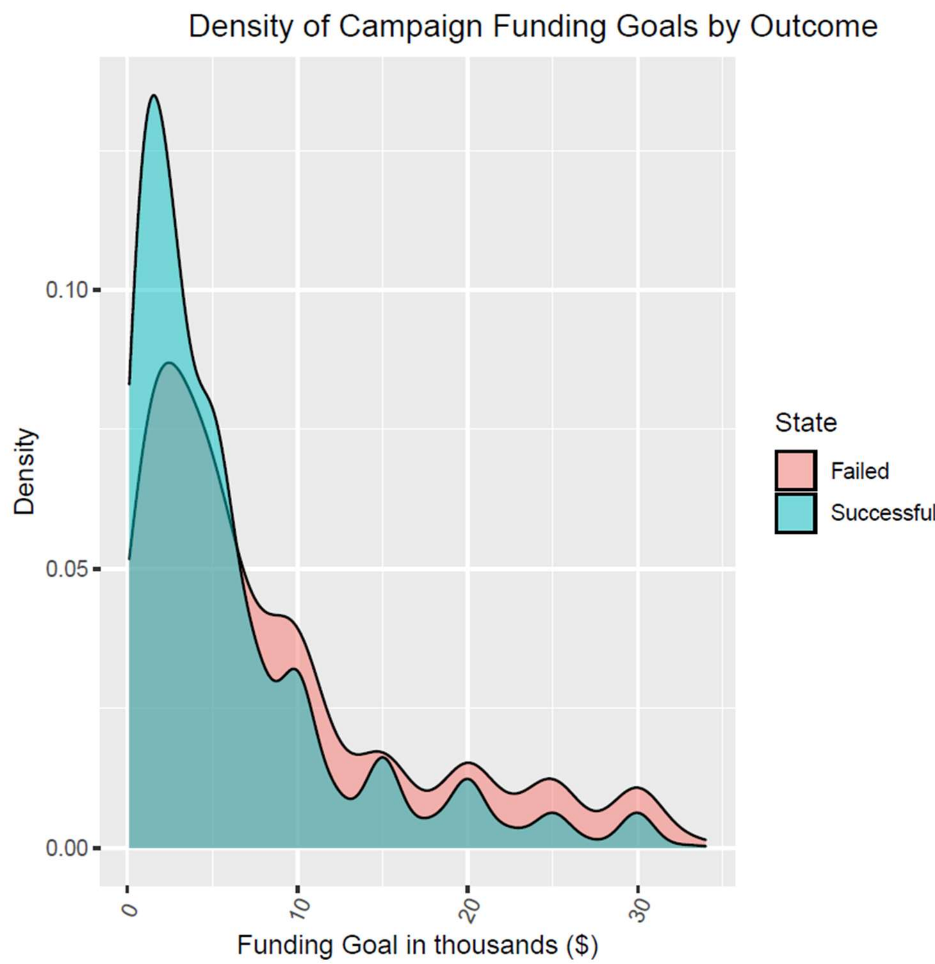

Fig. 1.1 Distribution of campaign funding goals based on final outcome
Obtained using kernel density estimation with the ggplot R package

Figure 1.2 shows the distribution of funding goals of campaigns by outcome for the possible categories to which they are tied. Successful campaigns have much lower funding goals than failed ones. This remains true across all categories except Dance for which it is quite higher than their *Failed* counterparts. On the other hand, successful campaigns from categories such as Teather, Journalism, Games, Film & Video and Design had significantly lower average goal amounts than their failed counterparts from the same category. Based on the charts above, setting a reasonable goal seems to be a key factor towards its outcome. Setting reasonable goals or appropriate scaling of the venture could be indicative of better preparedness by the campaigners, as seen with the presence of videos, images and documentation about the projects provided to potential backers (Mollick, 2014).

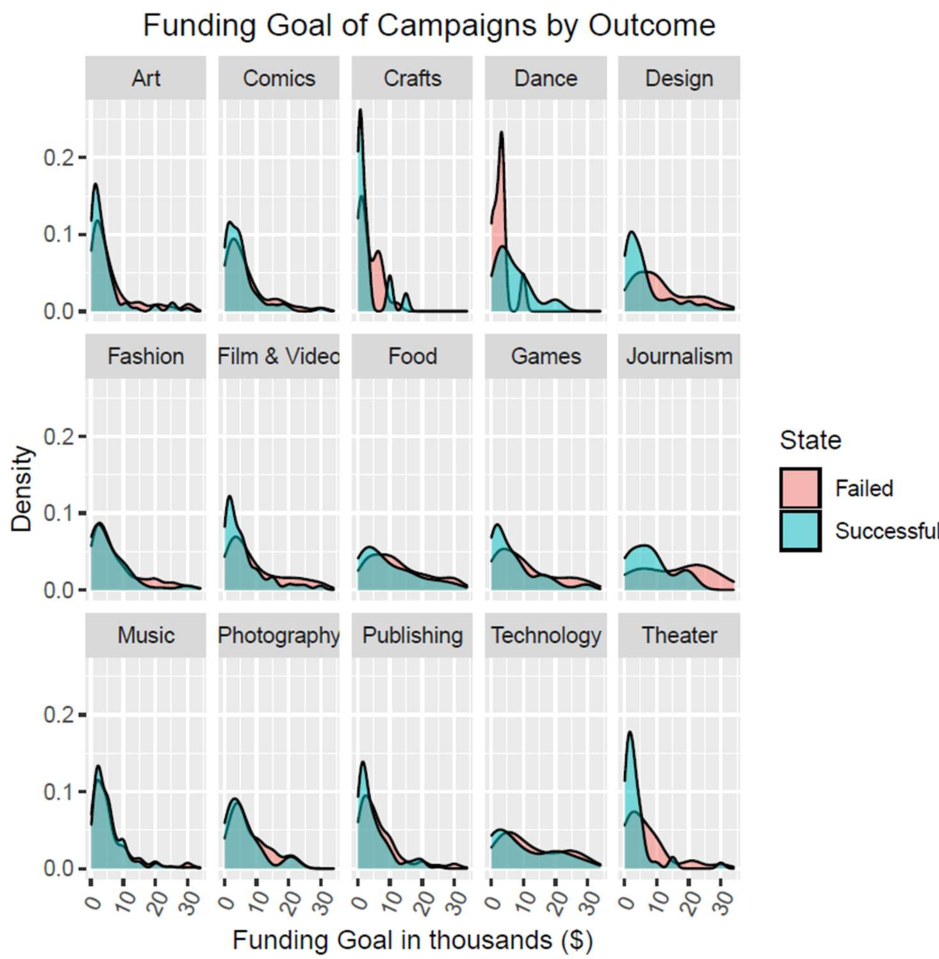

Fig. 1.2 Mean campaign funding goals based on final outcome by category
Obtained using kernel density estimation with the ggplot R package

Successful campaign starters seem better at setting realistic, achievable goals for their projects. This is echoed by the fact that lower goal amounts were associated with higher odds of success (Barbi & Bigelli, 2017). Another key success indicator identified in previous research is related to pledges made in the early days of a campaign. According to Etter et al. (2013), the amount of money pledged to a campaign in the first 15% of its duration was a significant variable in determining its outcome. This ultimately allowed them to accurately predict the outcome of over 85% of campaigns when combined with social predictors. The authors used two methods: the first was a K-Nearest Neighbor (KNN) classifier while they discretized the (time,pledged) space to also develop a time-inhomogeneous Markov Chain model. Both models used partial campaign pledge trajectories to predict their outcome. By definition, the partial series of pledge amount introduced as a predictor is intimately linked to the predicted outcome. However, using partial information provides a more insightful approach that highlights the importance of early pledges. Another interesting fact is Kickstarter's in-house claim that 78% of campaigns that achieve 20% of their goal at any time during their campaign will eventually succeed (Kickstarter, 2022).

How easily they succeed might differ based at least partially on the size of a campaigner's social network and footprint (Kindler et al., 2019). However, the aforementioned methods have certain interpretative drawbacks. The importance of early pledges previously identified does not inform on how much of the funding goal should be achieved. Additionally, Kickstarter's claim provides an interesting threshold but no indication on how timing affects the global success rate claimed. An alternative way to model this problem into a more insightful approach could thus be to create an indicator that dictates if the campaigns in the dataset reached 20% of their funding goal in their first or second week or not.

While this is less inclusive than Kickstarter's claim, which implies that 78% of campaigns that reached 20% at any point before its deadline ended in success, it highlights the importance of early pledges demonstrated by previous research (Li et al., 2016; Etter et al, 2013) and provides campaigners with a meaningful key milestone to reach at a specific point in their campaign to maximize their odds of success. The choice of 20% comes from Kickstarter's claim since a high percentage of success is observed for this threshold. Similar to the work done by Etter et al. (2013), this 20% indicator covariate would also be linked to the outcome and thus would require nuance in its interpretation. However, it would provide more insights to future campaigners by measuring the impact of reaching certain milestones at key moments in their campaign. This new milestone variable could also prove to be a useful general predictor of success if significant. This framework would be helpful to better understand just *how important early funding is* and *determine if it's a dealbreaker or not*.

1.1 Methodology – Benchmarking Milestones

The methodology proposed in this chapter closely follows the one used by Barbi & Bigelli (2017) where they factorize the goal variable into multiple bins each associated to a different level before using logistic regression with a binary target variable to measure how being associated to different goal levels impacts a campaigner's chances of succeeding through the resulting average marginal effects of the logit regression and their levels. Rather than applying it to the goal variable directly, this chapter aims to measure the impact of reaching different funding milestones at different key moments from a campaigner's standpoint.

In order to test these milestones variables, an initial logistic model is developed based on variables available at the beginning of campaigns. The model summary of the initial model highlights a few different elements worthy of mentioning, such as the fact that the variables *Goal*, *Facebook.Friends, Has.Video,* along with *Image, Video, FAQs and Description counts* were primarily the most important variables. Lower goal amount and campaign durations were tied to slightly higher odds of success, as previously identified by Barbi & Bigelli (2017). Final pledged amounts, backers, comments and rewards handed out were excluded as they would provide the models with information typically unavailable at the point in time at which the analysis is made. The binary logistic regression follows the formula below, with y representing the binary dependent variable, X representing the list of explanatory variables and β their respective coefficients.

$$y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n$$

The model parameters are estimated based on Maximum Likelihood Estimation and the Wald test can be used to evaluate the significance of the β coefficients of the different explanatory variables. Exponential β provides Odds Ratios associated to the different explanatory variables, which provides a measure of their impact on the odds of the target variable's binary outcome. The target variable in this case is *State*, a binary indicator for the campaign's outcome, with 1 indicating successful funding and 0 representing failure. The initial logistic regression model used in this research including the following control variables is stated as follows:

State ~ Goal + Category + Facebook.Connected + Facebook.Friends + Has.Video + Description.Word.Count + FAQs.Count + Video.Count + Image.Count

In order to evaluate the importance of early funding that has been identified to carry a lot of weight in previous research, milestone variables were developed that identified if a certain campaign had achieved 20% of their respective goals after the first seven and 14 days. This is inspired by a combination of two previous studies. In the first (Gera et al., 2017), projects were binned into four categories based on final funding levels to then be analyzed for key characteristics. This study also used partial pledge trajectories in an attempt to predict if a given project would end up overfunded, funded, potentially funded, lowly funded (Gera et al., 2017). Predicted classification in this study was done by measuring the Euclidian distance of partial pledges from the median of the four possible funding categories. In the second study, the funding goal variable was binned based on its size and introduced as a factor in a linear probability model to measure its impact on a campaign's chances of success based on the resulting coefficients (Barbi & Bigelli, 2017).

For the purpose of this research, the 20% threshold was chosen based on the Kickstarter claim in order to better understand the dynamics behind it (Kickstarter, 2022). Although dichotomization will decrease the power of our tests compared to using the amount of money raised, using that binary outcome is more representative of reality in this case because of the all or nothing consequence of reaching the campaign goal. In addition, we investigate the role of early funding on the success of a campaign since those two variables have previously been linked. The current proposed method also only uses information available at the time of evaluation, as it is possible to use these milestone variables to position ourselves statically at Day-7 or Day-14 of the campaigns' time-series for the analyses. This decision was made to provide a realistic outlook from the campaigner's point of view, trying to plan for the rest of their campaign based on current results. As such, the *Quick20* variable was created and identified whether the campaigns had achieved 20% of its funding goal after seven days, while the *Good20* variable identified if the campaign had reached the same milestone after 14 days. The newly created variables were then sequentially included to evaluate their significance and impact on a campaign's odds of success using odds ratios and success rate. Despite being linked to the target variables, these milestone indicators provide relevant timely information about this specific situation. However, it is important to interpret the results with a grain of salt since both the effects of being closer to the set funding goal and achieving it early are contained by this metric. To mitigate this limitation, the cumulative backer per count on Day-7 and Day-14 is also introduced and measured as a means to test robustness. Figures 1.3 shows how achieving these milestones by different times impacts on a campaign's odds of success.

1.2 Early Funding Milestones Results

As displayed in Table 1.1, campaigners that reach 20% of their funding goal by Day-7 of their campaign had drastically improved odds of success by a factor of 44.12 within a 95% Wald confidence interval of [33.86, 58.14]. Of the campaigns that had reached at least 20% of their funding goal on Day-7, 90.2% were successful in reaching their goal. This makes achieving this milestone a very strong predictor of success since only 9.8% of campaigns that reached it ended up failing. On the other hand, the success rate of the typical 30-day campaign that hasn't reached 20% of its funding goal by that point in time is only 18%. While this doesn't rule out success, it means that a bit less than one in five campaigns that haven't reached this milestone at Day-7 will eventually succeed, a grim outlook from a campaigner's standpoint. For the sake of giving those campaigns a chance, the same model is rerun with the Good20 milestone variable instead. This will help define what tends to happen if we push the deadline to reach 20% of their funding goal by a week to Day-14, the halfway point in a typical 30-day campaign.

Table 1.1: Impact on odds of success of reaching 20% of a campaign's funding goal by day-7
State is a binary indicator for the campaign's final outcome, where 1 represents success and 0 failure to achieve the funding goal initially set. Quick20 indicates if the campaign had received 20% of its funding goal in pledges by Day-7. Effect of the entire list of control variables is reported in Table 1B of Appendix B. Based on the initial model and control variables with the single addition of the new milestone variable. Confidence intervals are based on Wald tests.

Dependent: State		Level	Failed (0)	Success (1)	Odds Ratio
Variables	Quick20	No	1302 (82.0)	286 (18.0)	-
		Yes	118 (9.8)	1091 (90.2)	44.12 (33.86-58.14, $p < 0.001$)

The results in Table 1.2 indicate that achieving this milestone increases your odds of success even further as a campaigner, this time by a factor of 81.59 within a 95% Wald confidence interval of [60.97, 110.78]. This is achieved despite a slightly lower success rate, with 86.4% of campaigns that had achieved 20% of their funding goal on Day-14 ending up succeeding. This can be explained by the fact that this milestone technically includes campaigns that achieved the milestone within the first week as well. The odds ratio increases nonetheless since there's a combination of a few successful campaigns that achieved 20% of their target funding in the second week and a large amount that didn't achieve it and ended up as failures. This is also expressed by the fact that 93% of campaigns that didn't achieve the milestone despite the additional week ended up as failures.

Table 1.2: Impact on odds of success of reaching 20% of a campaign's funding goal by day-14
State is a binary indicator for the campaign's final outcome, where 1 represents success and 0 failure to achieve the funding goal initially set. Quick20 indicates if the campaign had received 20% of its funding goal in pledges by Day-14. Effect of the entire list of control variables is reported in Table 1B of Appendix B. Based on the initial model and control variables with the single addition of the new milestone variable. Confidence intervals are based on Wald tests.

Dependent: State		Level	Failed (0)	Success (1)	Odds Ratio
Variables	Good20	No	1212 (93.0)	91 (7.0)	-
		Yes	200 (13.6)	1272 (86.4)	81.59 (60.97-110.78, $p < 0.001$)

1.3 Robustness Test Results

In order to interpret the milestone results uncovered with more nuance, the same methodology is used with cumulative backer per day ratios at Day-7 and then at Day-14 where a variable is introduced to the model to measure its impact on odds of success based on resulting odds ratios and significance levels. These ratios helped dissociate two effects from each other: (1) the effect of having accumulated pledges and thus being financially closer to the end goal of the target variable, from (2) the early timing aspect. This was accomplished by using metrics that were not as closely related to the outcome, in this case, backers rather than pledged amounts. Using this ratio puts the campaigns on the same scale and focuses on the early influx level of backers, which also reflects the importance of early campaign performance. While the results cannot be directly extended to the importance of early timing, observing a similar effect solidifies the argument that early campaign performance is important and that the previously developed milestones are relevant in their application.

The results of using the accumulated backer per day ratio at Day-7 and Day-14 are reported in Table 1.3 and show the impact of these two metrics on a campaign's odds of success at these points in time. The resulting odds ratios define the impact on odds of success coming from a change of one in the backer per day ratio maintained over seven and 14 days.

Table 1.3: Impact on odds of success of the backer per day ratio in the campaign's first seven and 14 days

State is a binary indicator for the campaign's final outcome, where 1 represents success and 0 failure to achieve the funding goal initially set. Day-7 indicates the backer per day ratio achieved by the campaign as of Day-7, while Day-14 represents the same but at Day-14. Effect of the entire list of control variables is reported in Table 1B of Appendix B. Based on the initial model and control variables with the single addition of the new milestone variable. Confidence intervals are based on Wald tests.

Dependent: State		**Level**	**Metric**	**Failed (0)**	**Success (1)**	**Odds Ratio**
Variables	**Backer Per Day Ratio**	**Day-7**	**Mean (SD)**	1.3 (2.4)	10.6 (31.7)	1.87 (1.76-2.00, $p < 0.001$)
		Day-14	**Mean (SD)**	0.9 (1.6)	7.4 (22.7)	2.93 (2.64-3.27, $p < 0.011$)

The results displayed in Table 1.3 show the same general effect previously observed for the *Milestone* variables, but on a much smaller scale. The effect of an increase of one unit of the Backer Per Day ratio at Day-7 increased odds of success by a factor of 1.87, while the effect was of a factor of 2.93 if the same increase of one unit was maintained over 14 days. These lower odds ratios were expected given that this variable is not as intimately linked to the target variable *State* which defines success or failure. This means that a campaign that maintains a more elevated Backer Per Day ratio in the early phases of a campaign indeed has amplified odds of success. An elevated Backer Per Day ratio indicates that a campaign is catching the interest of potential backers more successfully and thus maintains stronger early campaign momentum, while being not as directly linked to the total amount. Table 1.3 shows that the ratio is still higher for the Day-14 indicator, which makes sense given that a campaign that manages to keep an elevated ratio of a naturally declining metric implies that it performed better for a longer duration. The ratio naturally declines as it gets divided by a larger and larger number as the days advance.

1.4 Timing Implications behind Kickstarter's Claim

In total, 61.5% of 30-day campaigns reach the 20% of their funding goal. Of those, 80% ultimately succeeded. This *supports Kickstarter's claim* that 78% of campaigns that achieve 20% of their goal at any time end up being successful (Kickstarter, 2022). However, this claim is misleading since the actual timing at which a campaigner reaches that 20% milestone heavily influences how likely it is to succeed. In order to better understand the impact of timing in this case, campaigns are classified into different bins based on the weeks at which they achieved the 20% milestone. The success rate of each bin is then calculated to see how success is distributed for campaigns that achieve this milestone. The results are demonstrated in Figure 1.3, which outlines the success rate associated with achieving the 20% milestone in the different weeks of a 30-day campaign.

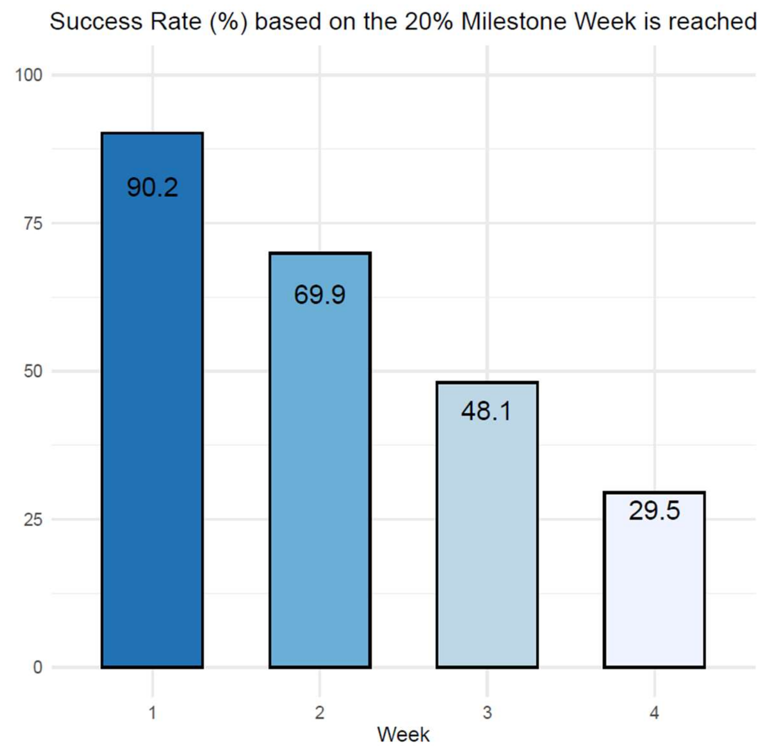

Fig. 1.3: Success rate (%) of 30-day campaigns based on when 20% of funding goal is reached

While 90.2% of campaigns that achieved the milestone in their first week succeeded, the success rate drops to 69.9% for campaigns that achieved it during their second week. In fact, the Figure further drops to 48.1% for campaigns that reach this much funding in their third week, and to only 29.5% for campaigns that do so in their final week. This indicates that reaching 20% of a campaign's funding goal once it's past half of its duration does not have much impact on its odds of success. For instance, achieving 20% of a campaign's funding goal in the third week shows a success rate of 48.1%, which indicates that such a campaign is almost as likely to succeed as a coin toss. The fact that the success rate is just below 50% indicates that the positive impact observed for campaigns that achieve the 20% milestone in the first two weeks fades after the second week. While campaigns that achieve this milestone in the third week still have a chance, they represent a very small fraction (7.1%) of the campaigns that achieve 20% of their funding goal. In fact, most campaigns that achieve this milestone almost all achieve it in the first (71.2%) and second (16.4%) weeks. It is also this 87.6% percent of campaigns that achieve it early that have a very high success rate. On the other hand, achieving it in the final week is simply too late given that the success rate is merely 29.5%. The fact that the success rates are so elevated for the campaigns that achieve the milestone in the first or second week indicate that timing is of the essence and that doing so does indeed put a campaign on track for success, while this effect is lost in the third a fourth weeks.

The median 30-day campaign achieved a total of 63.6% of its funding goal, showing that achieving the 20% milestone is not a guarantee of success and that timing is of the essence. The average 30-day campaign had a funding goal of $7,343, while it was $5,815 for campaigns that achieved 20% of their funding goal in the first week (*Quick20 Milestone*). In comparison, the average successful campaign, regardless of the developed milestones, had a funding goal of $6,022, once again lower than the average. Furthermore, this difference of approximately 20% in average funding goals between both groups highlights a potential burden that campaigns impose on themselves by setting high funding goals. This illustrates the importance of both proper goal setting and achieving certain milestones early on as they both yield much better odds of success. The 9.8% of campaigns that failed despite reaching the *Quick20 Milestone* only reached 28.7% of their funding goal. Importantly, these campaigns had an average goal of $8,590, which is higher than the global average.

Furthermore, social media connectivity and activity is positively linked to campaign success (Kaur & Gera, 2017). It is worthy to note that social success in crowdfunding campaigns tends not to be induced by virality but rather by social fitness (Kindler et al., 2019). This can be defined as social media activity (likes, shares, tweets) and network, which has already been shown to significantly contribute to a campaign's success (Li et al., 2016). The success of the ventures themselves is also influenced by the presence of type-2 backers (Kindler et al., 2019). These backers are characterized as being overly enthusiastic about a particular project and are thus willing to pledge more than their fair share (Kindler et al., 2019). Further, the average campaigner had 455 Facebook friends, while the figure rose to 521 for campaigners that achieved the *Quick20 Milestone*. Additionally, this figure rose to 536 for successful campaigns. Successful campaigns thus have characteristics that are very close to the ones that achieve the *Quick20 Milestone*.

1.5 Early Funding Milestones Takeaway

This chapter's contribution showed that reaching the *Quick20 Milestone* is a strong predictor of success that entrepreneurs could use as a proxy to estimate their odds of success after the first week and subsequently attempt to keep the momentum going or raise additional awareness as needed through their social fitness initiatives in their second week. The milestone may be considered deal-breaking since only 18% of campaigns that don't achieve it will ultimately succeed. If 20% hasn't been reached by Day-14, the rate goes down to a grim 7%. Over the course of four weeks, Figure 1.5 shows that the success rate for campaigns that achieve 20% of their funding level drops by approximately 20% per week as the weeks advance. While 90.2% of campaigns that achieve 20% of their funding goal in the first week ultimately succeed, only 29.5% of those that achieve the same milestone in the fourth and final week ultimately achieve success. Given that timing is key, and that social network size is a positive contributor (Kaur & Gera, 2017), campaigners should also focus on pre-launch and early campaign social media awareness through their network. While it certainly might not always be possible, scaling down a future crowdfunding project to the bare minimum to lower its goal could also be a great tactic to maximize its odds of success. This is particularly important on "all-or-nothing" platforms like Kickstarter since missing a single dollar results in the loss of all pledged amounts.

Second Chapter

2. Clustering Success Through KPI Trajectories

2.1 Methodology – Functional Data Clustering

Time series data are typically high-dimensional due to the nature of high-frequency observations. This creates large amounts of data to sift through as every day becomes its own dimension. This can be an impractical source from which to draw insight as the individual information points might not necessarily be important on their own, but the full or partial series can help find underlying patterns in the data. Such time-series of observations can be turned into individual functions to reduce the number of dimensions while retaining the underlying information. Each time-series then suddenly becomes a distinct, smooth function characterized by the series of values observed. Analyzing this type of data is useful to facilitate the interpretation of patterns and relationships between underlying variables (Bouveyron et al., 2015)

Functional data clustering of time-series is a method that uses a discriminative functional mixture model to cluster such functional data in a discriminative functional subspace (Bouveyron et al., 2015). Part of Functional Data Analysis (FDA), it is a statistical method that turns discrete time-series data into smooth and continuous function-based data objects to identify and compile the general trends in the discrete observations (Giordani et. Al., 2020). The *FunFEM* R package built specifically for such analyses is used to extract the functional nature of the data (Bouveyron et al., 2015), which is represented by the smoothened functions of the various time-varying variables. In this sense, rather than classifying the similarity of individual data points, this method allows for the classification of complete curves, or trajectories, which better capture trends of timely nature in functional data (Bouveyron et al., 2015).

The smoothened function is created by using a set of independent functions that form a basis function system (Giordani et al, 2020). This assumes that the resulting curves can be decomposed in a finite basis of functions (Bouveyron et al., 2015), as per the equation below:

$$x_i(t) = \sum_{j=1}^{p} \gamma_{ij} \psi_j(t) \tag{1}$$

Where $Xi(t)$ represents the set of mathematically independent smoothened functions of the different observations i over a finite set of ordered times represented as a continuum by *t,* the γ_{ij} coefficients are the basis expansion coefficients of the different basis functions ψ_j in the system, which follows a mixture of Gaussian functions (Bouveyron et al., 2015). The basis of this functional data has to be defined and turned into a functional data object before it can be injected into the *FunFEM* R package. B-Spline functions are the most widespread choice of functions for the basis system for non-periodic functional data (Giordani et al, 2020). These basic spline functions are piecewise polynomials curves with minimal support separated by knots used for curve-fitting. A number of basis that creates equidistant knots is chosen as it can be beneficial and more stable when variability is constant. A limited number of functions yield great flexibility in the approximations, with additional knots allowing for greater flexibility at the cost of a more complex model (Giordani et al, 2020).

Once defined, the resulting functional data object can then be processed by the *FunFEM* algorithm. The R package runs a Fisher-Expectation-Maximization Algorithm which considers an additional F step to update the orientation matrix used to map the γ coefficient into the discriminative subspace and requires two inputs to run: (1) a functional data object defined by the *fda* R package, as well as (2) the number of clusters desired (Bouveyron et al., 2015). The F step is required because the particular nature of the functional subspace F doesn't allow for model inference using the Expectation-Maximization algorithm (Bouveyron et al., 2015). Instead, the algorithm maximizes the likelihood over the subspace orientation matrix U which maximizes the projected variance and yields the functional principal component analysis subspace (Bouveyron et al., 2015). This method is based on Fisher's method as the F subspace is such that the variance within any given cluster should be minimized and the variance between the cluster themselves maximized. The algorithm runs through the 12 different discriminative latent mixture models unless one is specified. The 12 models are variants of the general $\sum k\beta k$ discriminative latent mixture model that applies different constraints to the Δk matrix (Bouveyron et al., 2015). The algorithm then identifies the ideal number of clusters based on the specified range and model selection criteria. Model selection can be done based on the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC) or the Integrated Completed Likelihood criterion (ICL), with BIC as default. The model then outputs posterior probabilities, estimated clusters and model parameter estimates. The default initialization type is k-means, but hierarchical clustering is also supported. The mean values of every cluster can then be graphed for the entire duration of the discrete time series to extract the different trends that can be attributed to each cluster.

2.2 Application to Crowdfunding Campaigns

In the crowdfunding environment, functional clustering makes it possible to better investigate time series and effectively classify the cumulative amount pledged and backer trajectories of the individual campaigns into different clusters based on their similarity. This means it is possible to use functional clustering to identify different investment and backer profiles associated to different campaign clusters that might be distinctively more successful than others based on their respective characteristics. These profiles would be represented by different cluster trajectories that each have their own distinct characteristics. While the cumulative pledges and backers at a given day is the only information available in the dataset, it is also possible to create more trajectories by calculating different ratios or KPIs across the entire time-series. By using the initial cumulative pledged and backer metrics' trajectories to develop new ratios and KPIs, this chapter's contribution aims to *identify distinct funding profiles and dynamics typically associated with different degrees of success through functional clustering*. These would have different graphic representations based on their progression that could be picked up by the functional clustering algorithm as investing profiles of campaigns that have different levels of success.

It is noteworthy that Kickstarter campaigns have a maximum duration of 60 days, with most campaigners opting for 30 days as recommended by the platform since it has previously been observed to be more successful than longer campaigns (Barbi & Bigelli, 2017). This is particularly convenient in this chapter to even out the time series and provide a common basis for analysis. By merging the previously used dataset containing general information about the projects with their time series of total pledges and backers, the chronological progression of these variables for the entire duration of the campaigns is obtained.

The progression of the previously identified and calculated KPIs of the individual projects can thus be injected into the *FunFEM* algorithm once a basis has been created and smoothened to fit the data. A B-spline basis was used given the nature of the crowdfunding data that was crawled once daily for their entire 30-day duration (Giordani et al, 2020). This is necessary as the *FunFEM* algorithm requires a functional data object from the *fda* R package to process. An ideal five knots were found to produce balanced results that equally split the campaigns into chunks of six days, each defined by its own function that together form the required basis to initiate the algorithm. This number allows for the division of the timeframe into equidistant knots that closely follows a natural week to week progression, each with an expected constant variability. The next step is to run the algorithm based on the different individual KPIs with the objective of identifying the ideal number of cluster-profiles in the data for each KPI. By performing an initial search of all possible models and a range of clusters going from two to four, the models converge to an ideal number of two or three clusters depending on the KPI used, with BIC being used to perform model selection. The analysis below expresses the results of clustering the projects based on different KPIs.

2.3 Functional Clustering of Crowdfunding Campaigns

2.3.1 Cumulative Metrics

The first set of metrics (KPIs) processed by the *FunFEM* R package are the cumulative *Percent of Goal Achieved over Time* and *Arrival of Backers over Time*. Analyzing these two metrics will provide information about the typical level of achievement in terms of funding and total backers for the various clusters identified. This helps visualize how large the difference is between the clusters purely in terms of total funding and backers. Based on Figure 2.1, a first cluster of high-performance campaigns largely overachieves their goals within a few days, while a second cluster of low-performance campaigns struggles to achieve theirs. The trajectories identified in figures 2.1 and 2.2 show strikingly similar profiles for both performance indicators, with the algorithm identifying the same clusters of campaigns. In fact, the two cumulative metrics could seemingly be used as proxies for each other based on the identical outcomes. This is a similar finding to previous research where large daily pledge signals were used as a proxy for the presence of high-pledge backers (Kindler et al., 2019). The two clusters for each KPI follow a nearly linear progression with the first cluster displaying a much higher lift in the campaigns' funding and backers. This is important as it emphasizes that campaigns that don't display a strong initial momentum might not lift later on either. Another observation is that the trajectories are not strictly increasing in these cumulative KPIs. This indicates that some campaigns lose some of their backers as well throughout their total duration. Following the platform's all-or-nothing model, pledges are only collected once the campaign is over if the funding goal has been reached. Backers can thus cancel their pledge at any time until the end of the campaign, so it is not uncommon to see pledges getting cancelled on the platform.

Percent of Goal Achieved over Time was then analyzed to better understand the pace at which different campaign clusters achieve their goals. This is done to find out if a cluster of campaign that takes more time to succeed can be identified. Finding such a cluster would support the argument that it is still fairly possible to succeed in the third or fourth week of a typical campaign for example. The KPI trajectories were split into two major clusters by the algorithm. These clusters are represented in Figure 2.1. The functional clustering algorithm identified a first cluster of 351 high-performance campaigns that achieved their funding goal after barely a few days and ended up highly overachieving (466% of funding goal reached on average). This cluster contains 12.5% of the 30-day campaigns in the dataset and has a success rate of 100%. The 2,467 campaigns in the second cluster of low-performance campaigns barely make it midway to their funding goal on average (56.6%). This cluster contains a mix of successful (1,035) and failed (1,432) campaigns, totaling a success rate of 42%. Figure 2.1 also shows that most of the campaigns that succeed in this low-performance cluster only do so by a small margin, ending up only a bit above 100% of their objective. The average goal of the high-performance cluster is $5,268, while it is $7,639 for the second low-performance cluster. Once again, the high-performance cluster demonstrates lower goals than average. This is in line with findings from previous research (Barbi & Bigelli., 2017) as well as findings in the First Chapter that also found lower goal amounts to be more successful. The inverse is also true for Facebook friends, which is 466 compared to 422 for the second low-performance cluster. The most popular Categories in the highly successful cluster were Tabletop Games (12.3%), Product Design (7.6%), Hardware (6.5%) and Comics (6.3%).

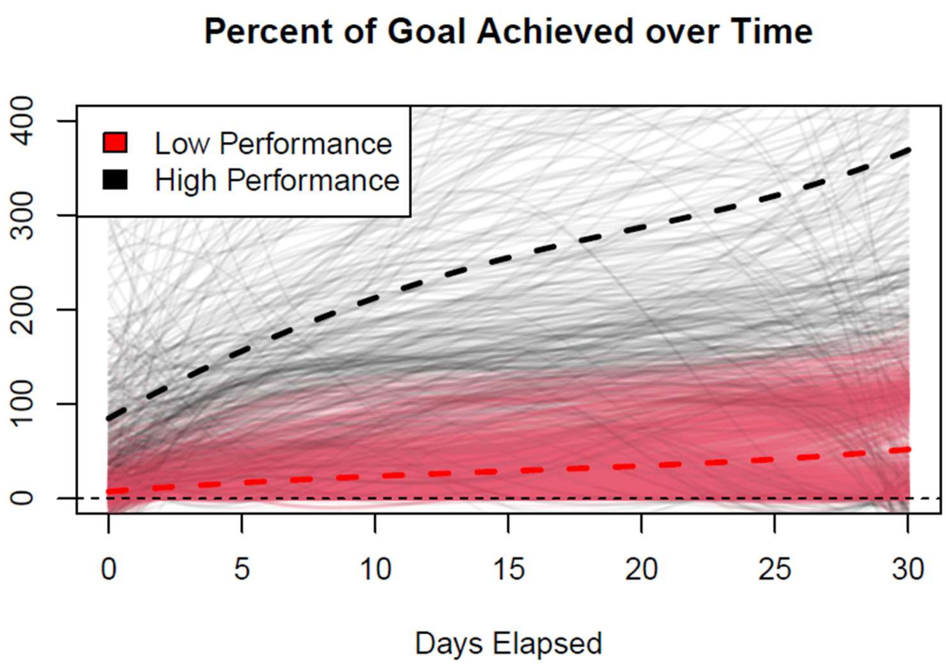

Fig. 2.1 Mean cumulative percent of funding goal achieved by campaign cluster over time

Arrival of Backers over Time is then processed by the algorithm to better understand the relationship between funding and backer trajectories. This KPI also gets separated into two clusters that look identical to the ones identified using *Percent of Goal Achieved* based on Figure 2.2. The mean trajectories of the two clusters identified also appear to closely follow the relatively linear trajectories of the first KPI's clusters. The first cluster contained the same 351 high-performing campaigns previously identified. This cluster was shown to accumulate an early 50 backers in the first five days of the campaign which then continues to increase in an almost linear fashion all the way to 200. On the other hand, the second cluster of previously identified low-performing campaigns barely managed to accumulate 10 to 20 backers over 30 days on average and reached the same mix and success rate of 42%, as the *Percent of Goal Achieved Over Time* KPI. The fact that both cumulative metrics identify the exact same campaigns is an interesting finding that indicates potential interchangeability. However, other calculated metrics could be more insightful or identify different sets of campaigns, which justifies the development of the next set of KPIs.

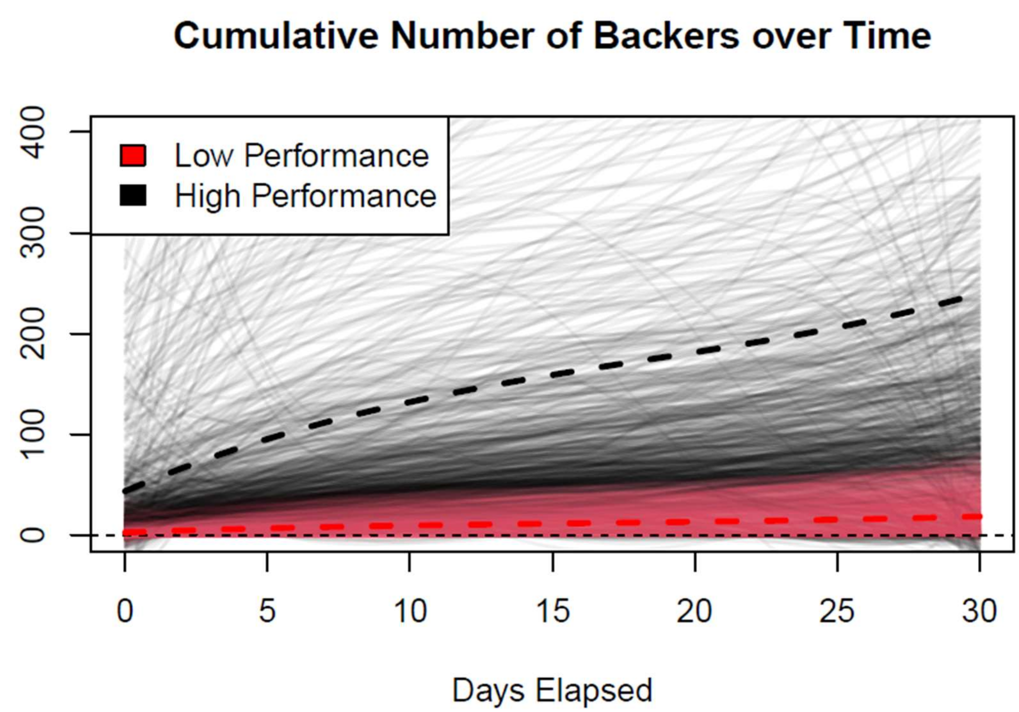

Fig. 2.2 Mean cumulative backer count accumulated by campaign cluster over time

2.3.2 Daily Cumulative Ratios

Within the context of daily cumulative ratios, the algorithm identified three clusters: high-performance, mid-performance, and low-performance campaigns with varying levels of success. The following KPIs offer more insight into important differences between high- and low-performance campaigns while differentiating them from mid-performance campaigns. The KPIs analyzed in this section focused on the progression of the cumulative funding per day and backer per day ratios over the course of a campaign's duration. While the first set in the previous section provided better understanding of the global achievement level of high-performance and low-performance campaigns, this second set will provide more insight into the daily influx level of funding and backers at different stages of a campaign according to three levels of campaign performance. In fact, it might be possible to identify if some campaigns have uptakes of funding and backers that allow them to succeed at key moments in their campaigns better with this KPI. The KPI trajectories of the clusters identified are once again very similar for both ratios in terms of shape, but the content of each cluster differs this time, both in size and proportions of successful campaigns. Nonetheless, both daily funding and backer ratios are typically much higher for very high-performance campaigns in the early stages and then quickly taper off, a phenomenon found to be a great predictor of success in previous research (Etter et al., 2013) and demonstrated to be a deal-breaking measure of success in the first chapter of this research.

Dollar per Day Ratio over Time is the next KPI analyzed to understand if it is better able to identify different clusters of trajectories with distinct characteristics that can help understand their success based on ratio fluctuations. This KPI is calculated using the total amount pledged at any given day divided by the numbers of days elapsed since the beginning of the campaign. Ratios introduces a new point of reference and thus creates new trajectories that could be typical of certain groups of campaigns. This would be represented by a new cluster being identified with new characteristics. The KPI run through *FunFEM* clearly identified high- and low-performance campaigns while there was more nuance in the second mid-performance cluster. This is represented by the success rates and size of the three clusters, which are as follows: the high-performance cluster contained 330 campaigns (12%) and had a success rate of 90.0%, the mid-performance cluster contained 1,326 campaigns (47%) and had a success rate of 68.6%, and the low-performance cluster of 1,162 campaigns (41%) and had a success rate of 15.4%. The high-performance cluster stands out from the rest as its campaigns showed much more momentum early on, as represented by the much higher but rapidly declining ratio. This cluster has an average of 520 Facebook friends, 52% more than the low-performance campaign cluster which had an average of 342.

However, the mid-performance campaign cluster also had an average friend count of 518, nearly identical to the high-performance cluster. This indicates that social media network size is a useful but limited tool in helping campaigns succeed and that other factors are necessary to be highly successful. Further, the high-performance cluster has a mean final pledged value of $31,369, while the mid- and low-performance clusters' values are $4,731 and $373 respectively. Looking at average goal values, it appears that the cluster of high-performance campaigns is comprised of the larger successful campaigns, that is, campaigns with higher goals that were identified in the fourth quartile of the successful campaigns. These 330 high performance campaigns with higher-than-typical goals are mostly comprised of Tabletop Games (45), Product Design (30), Comics (22), Food (19), Hardware (19) and Documentaries (18), which make up nearly half the high-performance cluster (46.4%). This echoes the findings of previous research, which reasons that smaller and larger campaigns could have different models for success (Mollick, 2014). However, Figure 2.3 shows that their curves follow a similar shape but at a different level of amplitude.

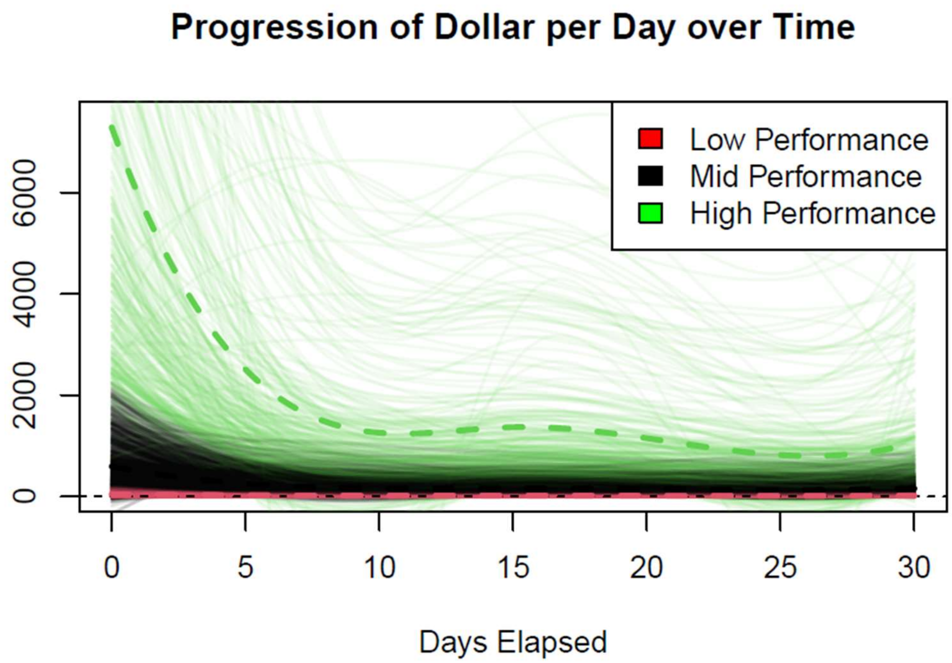

Fig. 2.3 Mean cumulative Dollar per Day Ratio by campaign cluster over time

Backer per Day Ratio over Time is calculated using the same logic and bears a striking resemblance to *Dollar per Day Ratio over Time* as the curves identified in Figure 2.4 are also very similar to the ones in Figure 2.3, with high-performance campaigns cluster also having the same initial high momentum trend. The campaign clusters based on this KPI are broken down as follows: the high-performance cluster of 125 campaigns (4%) had a success rate of 96%, the mid-performance cluster of 640 campaigns (23%) had a success rate of 83.2%, while the low-performance cluster of 2,053 campaigns (73%) had a success rate of 35.7%. The high-performance cluster achieved an impressive average amount of 1,044 backers, while the mid-performance achieved 149 and the low-performance only 26. Unlike the Cumulative Metric KPIs that each identified a cluster of high-performance campaigns with relatively lower goal amounts, clustering on the Daily Cumulative ratio KPIs allowed identification of high-performance campaigns with higher-than-typical goals. For instance, the average funding goal of the low-performance cluster is $6,619, lower than the mid-performance cluster's $8,615 and the high-performance cluster's $12,725. This shows that, surprisingly, the larger successful campaigns are characterized by a high number of backers per day early on that ultimately leads to a high final count of backers. These ratios distinguish between high-, mid- and low-performance campaigns very well based on cluster success rates. This is important since performance indicators and campaign details may subsequently be extracted and compared between clusters to more accurately define characteristics of success.

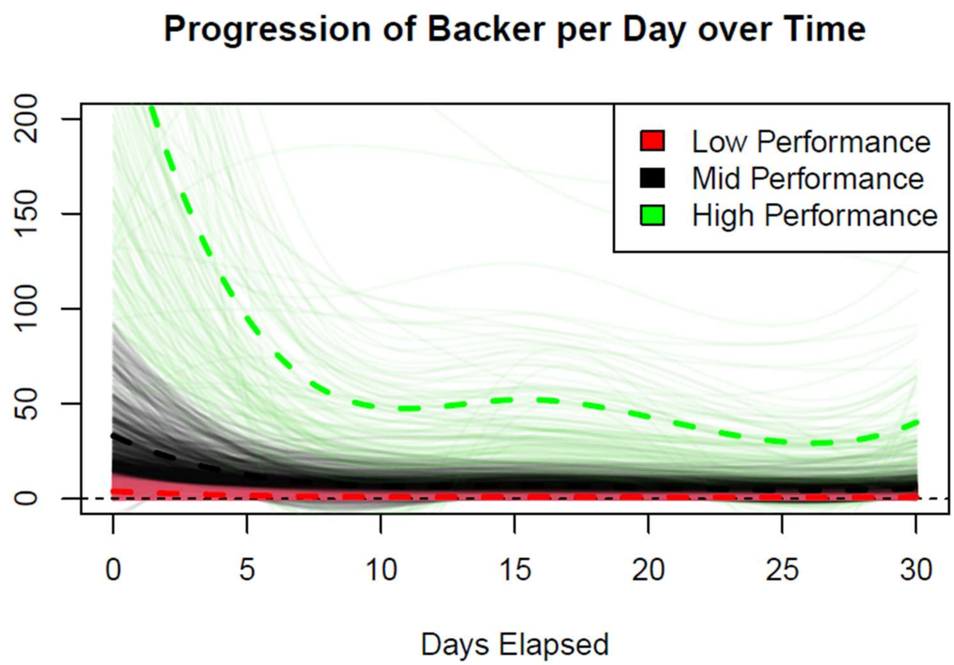

Fig 2.4 Mean cumulative Backer per Day Ratio by campaign cluster over time

2.3.3 Daily Funding Metrics

Next, rather than using the progression of cumulative metrics and ratios applied in the previous sections, the daily incremental values and percentages against the previous day are added to the time series dataset as the calculated metrics *Daily Pledges* and *Daily Percent of Goal Accumulated* over time. The goal of analyzing these metrics is to provide more detail about possible large daily fluctuations or long stagnations that could be characteristic of certain campaigns and ultimately impact their success. Unlike the previous set of KPIs that included Dollar and Backer per Day Ratios and tended to decline quickly as the number of days in the denominator position grew, this new set might reveal different trends. Absolute values might make it possible to better analyze the extent of the early, mid and late phase fluctuations. As seen in Figure 2.5 and 2.6, the two metrics share a similar “W” shape, particularly visible for the *Daily Percent of Goal Accumulated* metric. This “W” shape is more pronounced based on how successful the cluster is, indicating that successful campaigns indeed maintain their momentum better than the others in the half-life and final phase of their campaigns.

Daily Pledges over Time is the next metric analyzed with the intention of finding out if daily pledged amounts by backers can be used to identify trajectories that would be characteristic to different types of campaigns. This will make it easier to pinpoint potential key events in a typical campaign’s timeline. The same qualities previously identified for high-performance campaigns, such as higher early momentum, were shown to gain better support in their middle phase, while also often benefiting from an increase towards the campaign’s final days. This creates a “W” shape that seems to be typical of campaigns that win over the crowd. Three clusters are once again identified in the final output as seen in Figure 2.5. Their composition was as follows: the high-performance cluster contained 285 campaigns (10%) and had a success rate of 95.4%, the mid-performance cluster contained 1,311 campaigns (47%) and had a success rate of 71.2%, while the low-performance cluster contained 1,222 campaigns (43%) and had a significantly lower success rate of only 14.7%. The algorithm once again identified campaigns with higher-than-typical goals as high-performance in the dataset, despite these larger goals being typically less successful (Barbi & Bigelli., 2017). The mid-performance cluster identified with this metric had a higher success rate than with the previous metrics and ratios, identifying most of the successful campaigns that had a funding goal in the typical range of its group. This metric thus appears to identify both types of successful high- and mid-performance campaign clusters, those with: (1) higher-than-typical goals and (2) typical goal sizes. In contrast, the cluster of low-performance campaigns identified campaigns with lower-than-average funding goals.

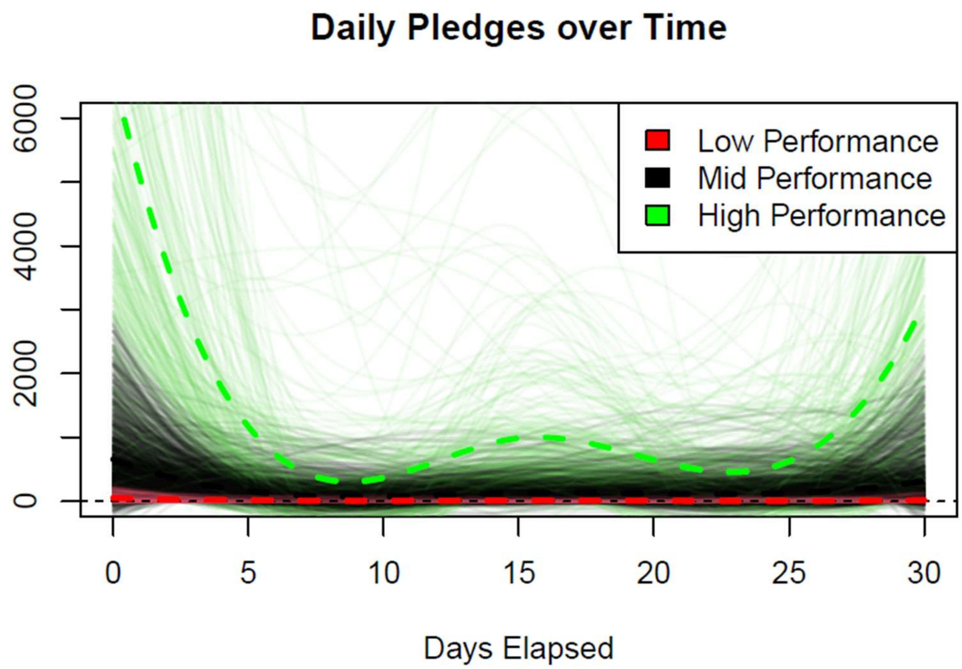

Fig. 2.5 Mean Daily Pledges collected by campaign cluster over time

Daily Percent of Goal over Time is analyzed next to classify the trajectories associated with this KPI in the hopes identifying distinct clusters that might succeed quickly and others that perhaps succeed more slowly with different key characteristics. While the cumulative values helped understand the pace at which campaigns succeed, the daily values will help understand if some campaigns see a sharp increase or decrease in backing in a more specific context. This KPI shows an even more pronounced "W" shape for the identified clusters as seen in Figure 2.6. The final model contained three clusters broken down as follows: the high-performance cluster contained 331 campaigns (12%) and had a success rate of 99.6% as it only contained one failed campaign, the mid-performance cluster contained 752 campaigns (27%) and had a success rate of 85.8%, while the low-performance cluster contained 1,735 campaigns (61%) and had a significantly lower success rate of 23.7%. Based on Figure 2.6, the high-performance cluster achieved their goal within a few days while a good portion of campaigns in the mid-performance cluster also showed strong early campaign performance, but with the predominant middle phase uptake displayed by the high-performance cluster being less pronounced. Looking inside the clusters, the high-performance cluster had a mean Facebook friend count of 499, while the mid-performance cluster had 539 and the low-performance only 394. Given that the mid-performance cluster had the highest mean Facebook friend count, the positive effect generated by the size of one's social network may be limited. Other social attributes might be necessary to explain why some campaigns far outperform the herd.

In fact, despite the lower average friend count, campaigns from the high-performance cluster generated a mean final pledged amount of $22,148, compared to the mid-performance cluster's $7,538 and low-performance cluster's $2,340. This metric of Daily Percent of Goal Over Time identifies successful campaigns with lower funding goals, broken down as follows: a mean of $5,655 for the high-performance, $6,026 for the mid-performance, and $8,236 for the low-performance. Despite the lower mean friend count, the high-performance cluster attracts more total backers than its counterparts (363 compared to 127 and 36), which could be an illustration of better social fitness (Kindler et al., 2019). This could be explained by better reach of promotional activity on social media, as well as external factors such as expertise and credibility. Based on mean Daily Percent of Goal Over Time values for each cluster, the high-performance cluster significantly outperformed the others (48% per day compared to 12% and 1.3%). In short, the high-performance and mid-performance clusters' campaigns have lower funding goals, attract more backers that also pledge more. The mid-performance cluster performs very well even at an average of 12% per day. This sparks an interest to look more deeply into backer related KPIs

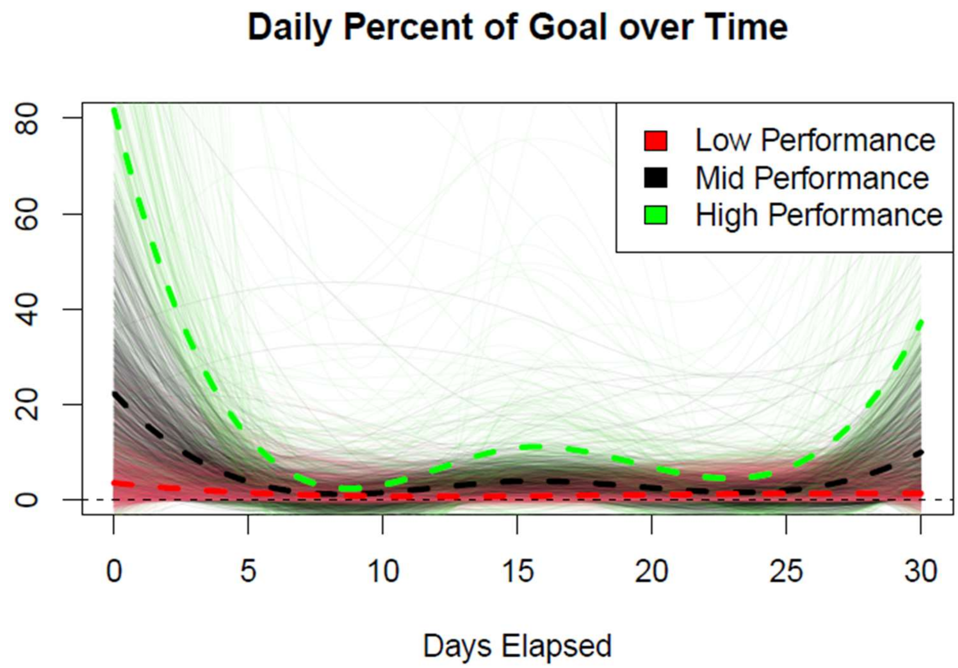

Fig. 2.6 Mean Daily Percent of funding goal collected by campaign cluster over time

2.3.4 Daily Backers & Daily Dollar per Backer Metrics

The next set of KPIs run through the *FunFEM* algorithm were the Daily Backers and Daily Dollar Per Backer metrics. Previous research found that days with large total pledges did not have more backers than usual (Kindler et al., 2019). Thus, they used a large daily pledge signal as a proxy to the presence of type-2 backers and concluded that their presence was a very strong predictor of success (Kindler et al., 2019). This implies that successful campaigns may receive larger pledges from at least some of their early backers compared to less successful or unsuccessful campaigns, which can be interpreted as being the type-2 backers previously mentioned. Only two clusters were identified for each metric this time around, both displaying the “W” shape identified previously: high-performance and low-performance. This is even more pronounced than before for the Daily Dollar Per Backer metric. As such, given the importance of the early stages in a campaign, having type-2 backers at these stages may provide key additional support that can propel a campaign towards success.

Daily Backers over Time produces two clusters that show the qualities previously identified with other metrics in terms of shape and momentum (Figure 2.7). Their composition was as follows: the high-performance cluster contained 784 campaigns (28%) and had a success rate of 87.1%, while the low-performance cluster contained 2,034 campaigns (72%) and had a significantly lower success rate of 34.6%. The mean amount of Daily Backer of the high-performance cluster was 27 per day, which contributed an average of 23.8% of the campaigns' funding goals per day. The low-performance cluster's 1.5 daily backers contributed 4.3% of the campaigns' funding goals daily, which shows that highly successful campaigns attract more backers that also pledge more to their projects than others. Network size in this case is once more a differentiating factor, with high-performance campaigners having on average 581 Facebook friends compared to 393. This metric also identified slightly more ambitious successful projects, with the mean goal being $9,413 for the high-performance cluster and $6,545 for the low-performance cluster.

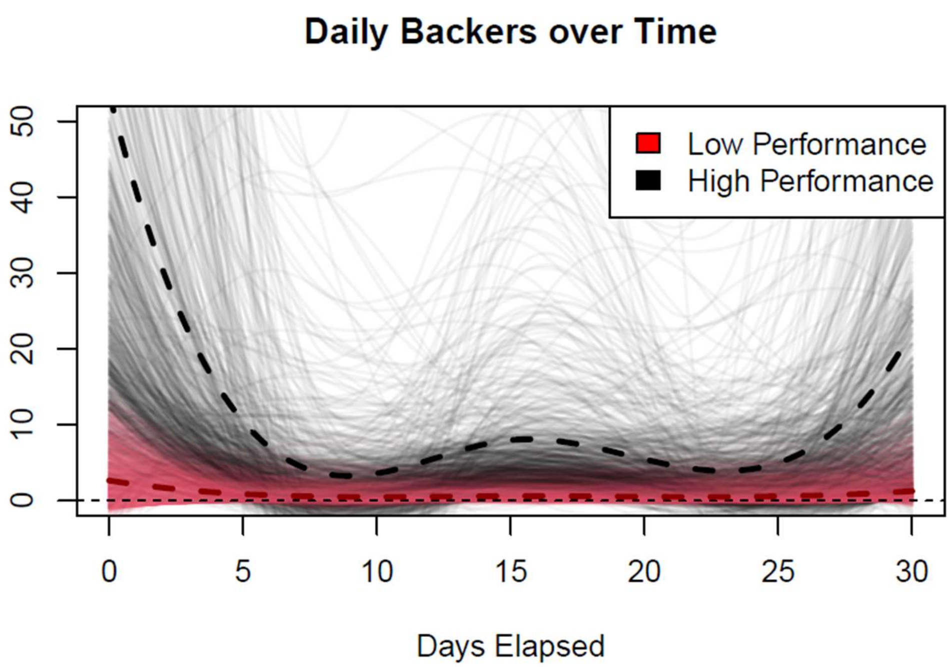

Fig. 2.7 Mean Daily Backers received by campaign cluster over time

Daily Dollar Per Backer over Time is the next metric that was analyzed. Since Funding and Backer KPIs appear to differentiate types of campaigns, looking at a hybrid of the two could be insightful by producing a new set of trajectories. The Daily Dollar Per Backer metric was calculated to illustrate the difference in backer commitment for the different clusters. This could support previous research that showed that high-pledge backers were a strong signal for success (Kindler et al., 2019). The metric was used to split the campaigns into two clusters as illustrated in Figure 2.8. Of all the metrics so far, the high-performance clusters' curve displayed the most pronounced "W" shape, a key characteristic of good momentum. The composition of both clusters are as follows: the high-performance cluster contained only 32 campaigns (1%) and had a success rate of 71.8%, while the low-performance cluster contained 2,786 campaigns (99%) and a success rate of 48.9%. Surprisingly, this metric seems to have performed the worst so far in differentiating the clusters, barely splitting the campaigns into two clusters that do not provide as much information as previous metrics and ratios.

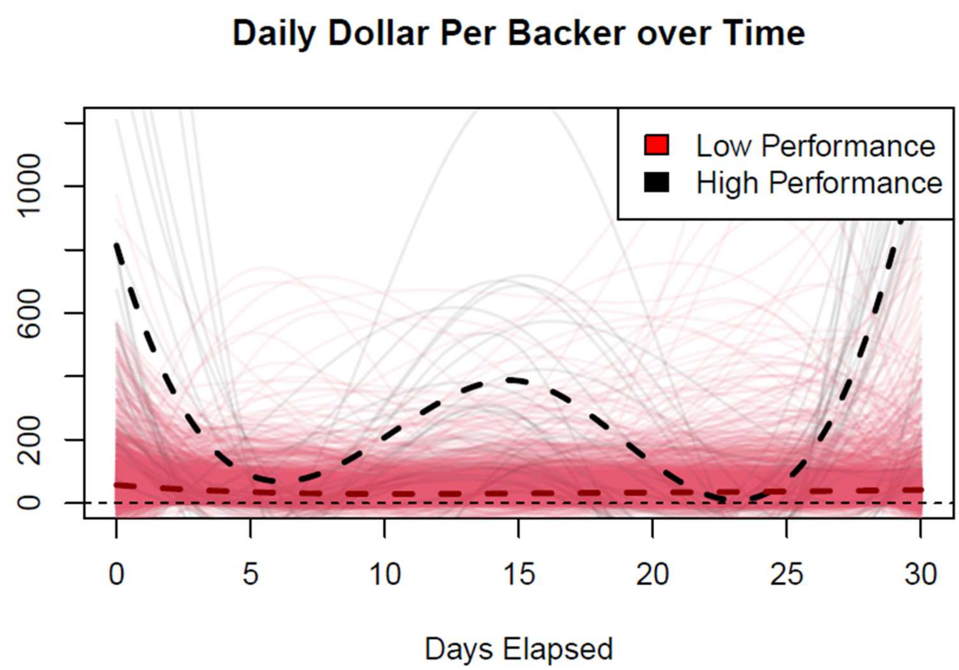

Fig. 2.8 Mean Daily Dollar Per Backer ratio by campaign cluster over time

2.3.5 Difference from Daily Averages Over Time Metrics

Calculating the difference from the average daily values to create new metrics could prove to be useful as it could provide the *FunFEM* algorithm with new perspectives by introducing the daily means as baselines. The new metrics were calculated with the intention of identifying more distinct clusters. For example, for previous daily metrics the final model was able to successfully differentiate high-performance from mid-performance campaigns. This KPI also had the benefit of producing outputs that made it easier to interpret the performance of the various clusters as well by introducing a new level of reference. Thus, the difference from the average daily values were calculated and the Daily Dollar Per Backer metric was reformulated in this way to determine whether it will allow more insight.

Daily Difference from Average Daily Funding was calculated and fed to the algorithm. As visualized in Figure 2.9, there were three clusters identified: high-, mid-, and low-performance campaigns. The high-performance cluster performed significantly better and displayed the same qualities previously identified in terms of early, mid and end phase momentum boosts ("W" shape). The composition of the three cluster is broken down as follows: the high-performance cluster contained 304 campaigns (11%) and had a success rate of only 93.8%, the mid-performance cluster contained 1,408 campaigns (50%) and had a success rate of 67.3%, while the low-performance cluster contained 1,106 campaigns (39%) and had a significantly higher success rate of 13.8%. The success of the high-performance cluster campaigns appears to be due to better social fitness (Kindler et al., 2019). Looking at the three clusters' friend count, high-performance campaigns had the highest count of Facebook friends (578 compared to 502 and 337) and better depth of reach on social media. One indicator of this is the total number of times the campaign was shared on Facebook as it would more likely allow enthusiastic backers that tend to contribute large sums to be exposed to the campaign itself. The mean share counts of high- and mid-performance clusters were indeed higher than for the low-performance cluster. In fact, while high- and mid-performance campaign clusters both had larger network sizes than average, high-performance campaigns appear to be more engaged with their network as suggested by the higher share counts. Once again, this is aligned with previous research that found social media interactions to have a positive effect on crowdfunding success (Kaur & Gera, 2017). Additionally, compared to the Daily Percent of Goal metric that tended to identify smaller projects, Daily Difference from Average Funding provided more insight by also identifying high-performance campaigns with higher-than-typical goals.

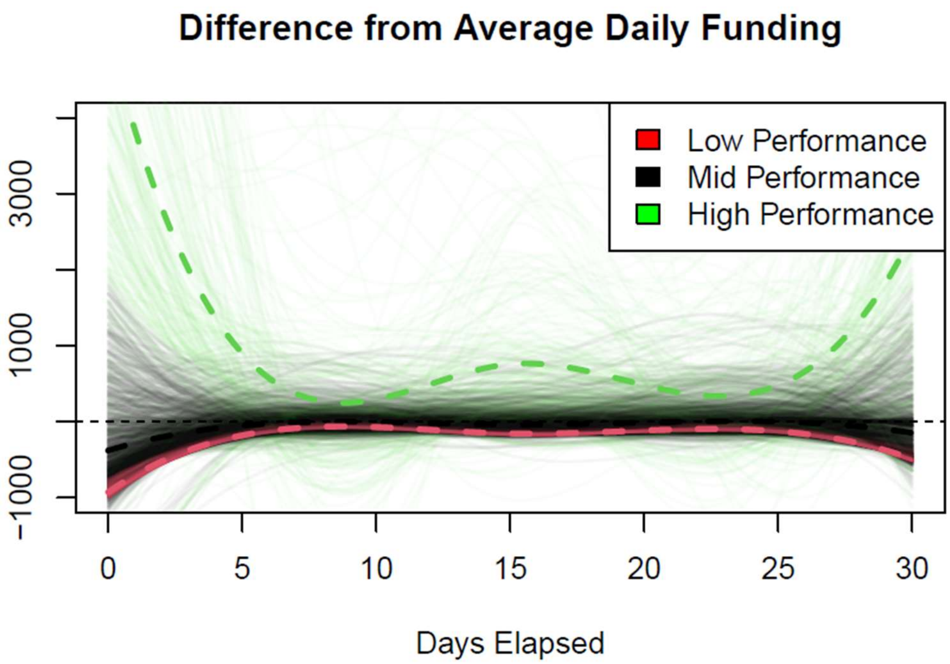

Fig. 2.9 Daily funding Difference from Average Daily Funding by campaign cluster over time

Daily Difference from Average Daily Dollar Per Backer is the last KPI introduced and was found to be more insightful than its previous variant, Daily Doller per Backer Over Time. The algorithm now found an ideal split of three clusters, as seen in Figure 2.10, of which the composition is the following: the high-performance cluster contained 302 campaigns (11%) and had a success rate of 75.5%, the mid-performance cluster contained 466 campaigns (17%) and had a success rate of 64.6%, while the low-performance cluster contained 2, 050 campaigns (72%) and had a significantly lower success rate of 41.8%. A key characteristic of this high-performance cluster is that unlike the other two, the ratio sharply increased from the middle to end phases of the campaigns. These campaigns also ended up slightly overachieved: the mean goal of this cluster being $11,480 but the final pledged amount being $12,218. This could be induced by better depth of reach on social media (Li et al., 2016), as indicated by the higher number of friends and shares these campaigns had. These campaigns might have also benefitted from their own success in that they might have been featured on the Kickstarter website as popular projects given how overachieved they ended up. Depth of reach measures like this could bring a lot more traffic to their project and benefit from the higher exposure to high-pledging, type-2 backers (Kaur et al., 2022). The Daily Difference from Average Daily Dollar Per Backer values for the various clusters range from 24$ for low-performance campaigns, to 64$ for the mid-performance campaigns and all the way to 313$ for high-performance campaigns. As such, these type-2 backers are a key characteristic of successful campaigns that allowed for relatively accurate predictions of success early on (Kindler et al., 2019). However, it should be noted that this metric is limited since the calculated success rates were not as distinct as those of previous metrics. Despite the large range in Daily Difference from Average Daily Dollar Per Backer between the high-, mid-, and low-performance clusters, this is not mirrored by the respective success rates.

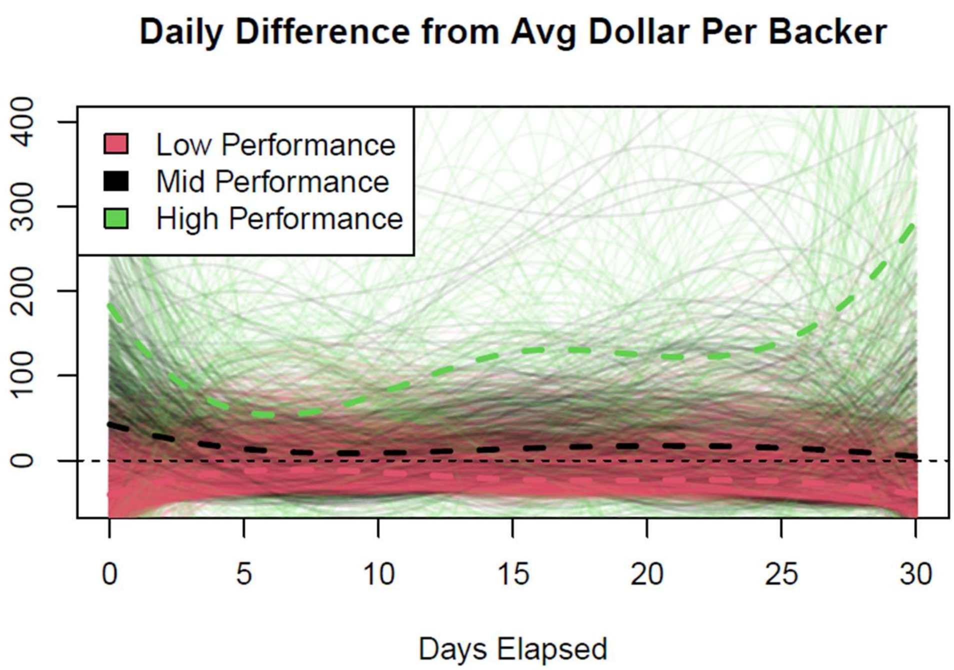

Fig. 2.10 Daily dollar per backer difference from average daily dollar per backer by campaign cluster over time

2.4 Sensitivity Analysis

In order to understand if the peak in the middle phase is an artefact of the number of knots selected in the algorithm, the Daily Pledge over Time KPI was rerun with seven knots instead of five. Figure 2.11 shows that the peak observed with five basis is replaced with a valley between two smaller peaks that are situated just after the initial and final rush of funding and backers observed in the previous KPI trajectories in the first and last days of the campaigns. This shift was also observed with other KPIs when the number of knots was changed with comparable results in terms of mean cluster trajectories and separation, which means that there may not be a spike exactly in the middle of campaigns. In fact, while successful campaign clusters appear to show peaks in performance throughout their campaigns in both scenarios, these peaks are found to be a feature of the number of knots selected. Different numbers of knots place these spurious bumps at different locations. Other than the strong early and late activity that remain consistent, the algorithm was not able to identify a structure common enough to support the presence of peaks in mid-phase activity.

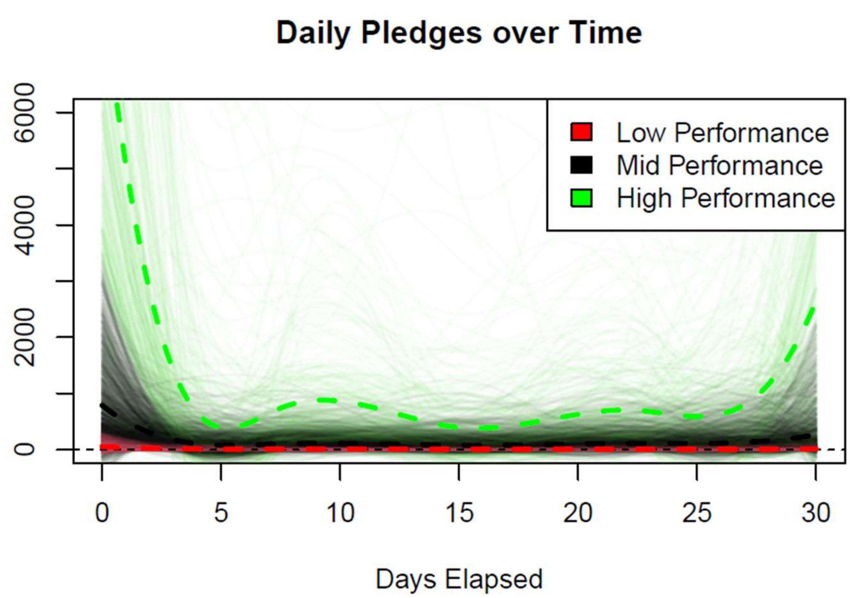

Fig. 2.11 Mean Daily Pledges collected by campaign cluster over time with 7 knots when running the FunFEM algorithm

2.5 Functional Clustering Takeaway

Functional clustering of campaigns made it possible to successfully use certain metrics and ratios to identify various indices of success depending on the metric analyzed. The KPIs analyzed were split into the following sets: Cumulative metrics, Cumulative Daily Ratios, Daily Funding metrics, Daily Backer & Dollar Per Backer metrics, and Differences from Average metrics. Calculated metrics and ratios also lead to different combinations of campaigns being identified. Daily Percent of Goal Achieved, Difference from Average Daily Funding and Daily Pledge and Backer Values led to the most distinct and insightful split of high-performance, mid-performance, and low-performance campaigns. The various metrics were also able to allow the differentiation of clusters with different compositions. For instance, the Daily Percent of Goal metric identified mainly high-performance campaigns with goals on the higher than average while the other two metrics identified high-performance campaigns with lower, more typical goal amounts.

This consolidates reasoning from previous research that large and small projects might be represented by different models altogether (Mollick, 2014). Both these types of campaigns seem to be picked up by the functional clustering algorithm. This would explain why different metrics identify similar but slightly different assortments of campaigns with different goal sizes.

High performance projects were consistently characterized by a ”W” shape which also expressed better social fitness, as indicated by these campaigners having a larger social network and better depth of reach through more successful social media activity. This shape was also present for other clusters, but less pronounced. However, this “W” shape turned out to be an artefact of the functional clustering, with a sensitivity analysis indicating that peaks in pledges and backers throughout the campaign were a feature of knot selection. Only the strong early and late activity were consistently identified by the algorithm. Campaigns that were highly funded and backed also have higher Daily Pledge per Backer. Both a high Backer Count and Dollar Per Backer Per Day were necessary to be highly funded. Campaigns that were shared more attracted a larger number of high-pledging backers and also tended to succeed.

Overall, functional clustering was successfully able to identify distinct clusters with different compositions, each with their respective KPI trajectories. These metrics were also able to allow the identification of differentiating factors of success in crowdfunding campaigns based on the composition and success rates of the clusters themselves.

Third Chapter

3. Liability of Foreignness in Crowdfunding Networks

Geographical information has been shown to be impactful within the analytical context of crowdfunding campaigns (Mollick, 2014). In fact, the distribution of projects and success across geographical space is uneven based on cultural factors (Mollick, 2014). In other words, distinct types of projects are more highly represented and have different odds of success in different locations. With that said, not much is known about the potential liability associated with being a foreigner in the crowdfunding network. Liability of foreignness is a concept that analyzes how being a foreign individual or having foreign characteristics to a particular network might impact the outcome of different situations (Zaheer, 1995). Previous research into immigrant entrepreneurship showed that foreigners can be negatively impacted by the additional barriers that may be associated with their foreign status (Irastorza & Pena, 2013). Within the context of Kickstarter, a platform originally based and most widely used in the US, foreign could be used to define a campaign with non-US typical cultural characteristics. This could be because a campaign originates from outside of the US, uses a different currency than the USD, or for which first language is not English.

These geographical, economic, and cultural factors could contribute to a campaign's success or failure by acting as a bias. As such, one might wonder how does a foreign project's odds of success fare against a similar project in the United States? Are they more or less likely to succeed? What about the currency used to fund the project, or the country's language? This chapter aims to assess the impact of these foreignness variables on a campaign's initial odds of success to determine whether liability of foreignness impacts foreign campaigns in the crowdfunding space. With this goal in mind, hypotheses were built to test the impact of these variables, with the basic assumption being that campaigns with characteristics foreign to Kickstarter's native community will be negatively impacted, as was the case with immigrant entrepreneurs in Irastorza & Pena's research (2013).

The initial hypotheses are as follows:

- ***H1:*** *Campaigns using the USD as currency are more likely to succeed than their foreign counterparts*
- ***H2:*** *Campaigns originating from North America are more likely to succeed than their foreign counterparts*
- ***H3:*** *Campaigns originating from English-speaking countries are more likely to succeed than their foreign counterparts*
- ***H4:*** *Campaigns originating elsewhere than North America are more likely to succeed than their foreign counterparts if they use the USD as Currency*
- ***H5:*** *Campaigns coming from non-English speaking countries will benefit from using the USD as currency compared to other foreign currencies*
- ***H6:*** *Campaigns coming from other continents than North America that also use English as a primary language will perform better than their non-English counterparts*

Descriptive statistics from an initial analysis show that the vast majority of campaigns are started in either the United States or the United Kingdom, both of which are English-speaking countries, as seen in Figure 3.1. In fact, only about 3% of campaigns were started in a country with an official language other than English. Despite that fact, these campaigns have a higher success rate than their English counterparts. Asia and Africa also seem to perform better when it comes to success rates by continent, but currency seems to favor campaigns that were based on the USD or GBP. Figure 3.1 gives a breakdown of the success rate achieved based on the different foreignness variables identified.

Success Rates of different Regions / Languages / Currency

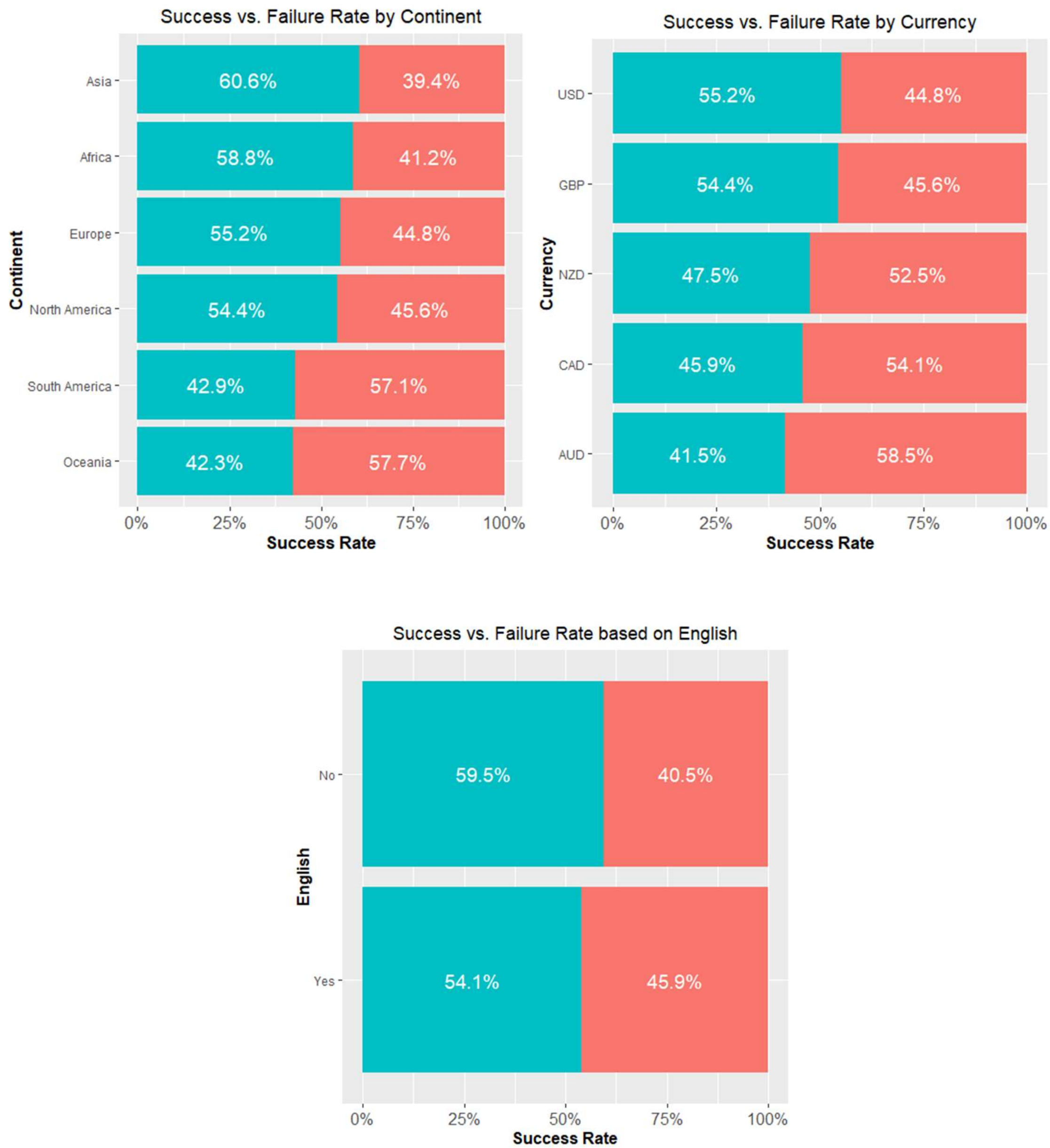

Fig. 3.1 Success rates of campaigns based on geography, currency and language

Counts associated with these proportions are further detailed in Tables 3.1, 3.3 and 3.5.

3.1 Methodology – Liability of Foreignness

The liability associated with foreignness can be assessed using logistic regression models since the target variable is *State,* a binary indicator for the campaign's outcome, with 1 indicating successful funding and 0 representing failure. The initial model using only the key predictors available at the beginning of the campaigns as introduced in the first chapter can thus be used once more. The summary of the same initial model used in the first chapter highlighted that the *Goal*, *Facebook.Friends, Has.Video,* along with *Image, Video, FAQs and Description counts* variables were the most important variables, with lower goal amount and campaign durations being tied to slightly higher odds of success as previously reported in previous research (Barbi & Bigelli, 2017). This initial model summary can be found in Table 1C of Appendix C. As mentioned before, binary logistic regression follows the formula below:

$$y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n$$

Where y represents the binary dependent variable, X different explanatory variables and β their respective coefficients. The estimation of the latter is made based on Maximum Likelihood Estimation. The Wald test can be used to evaluate the significance of the β coefficients. Exponential β provides Odds Ratios, which provides a measure of their impact on the odds of the target variable's binary outcome. The same predictors are used as controls, thus including funding goal amount, project category, Facebook connectivity and friend count, presence and count of videos, images, words in the description and FAQs. Duration is also added in the following analyses since all campaign durations were included in this chapter given the low occurrence of foreign campaigns. As mentioned in the first chapter, these control variables were used as they have been previously shown to be relevant to crowdfunding success. The effect of these control variables is reported in Appendix C along with the outputs of all models developed in this chapter. Key foreignness indicators were then sequentially introduced on their own to measure whether the impact of the *Currency*, *Continent* and *Language* variables play a significant role in a campaign's initial outlook for success.

As such, the initial model used in this chapter is the following:

State ~ Goal + Category + Facebook.Connected + Facebook.Friends + Has.Video + Description.Word.Count + Video.Count + Image.Count + FAQs.Count + Duration

Model outputs allowed for the determination of significance, and odds ratios were then used to measure the impact of the observed foreignness variable compared to the native level. In this case, the native level always used the US characteristics as the default level of these foreignness variables. The methodology followed in this research mirrors previous research, which used logistic regression models given the discrete nature of the events studied (Irastorza & Pena, 2013; Pastoriza, Plante & Lakhlef, 2021). In order to measure the impact of foreignness in crowdfunding networks, it is thus necessary to define and implement foreignness in our models. In the case of Kickstarter, a native campaigner would come from North America, use the USD currency and have English as a first language given that these would be the US characteristics. For this reason, these variables are defined as factors with these native levels set as reference. The currency, continent and language variables were then individually added to the basic model one at a time to measure their impact on campaign outcomes.

3.2 Impact of Currency

Given that the USD currency can be defined as native to the platform and is used by most of the platform’s users, the first hypothesis proposed in this chapter is the following:

H1: Campaigns using the USD as currency are more likely to succeed than their foreign counterparts

To evaluate if this hypothesis is supported, a logistic regression model was developed that added the *Currency* variable to the initial logistic model. The different possible currencies in the dataset were the CAD, USD, AUD, GBP and NZD. Of these, the USD was defined as the reference level in order to measure the impact of the other currencies being used in a campaign and their odds of success. Looking more closely at the outputs of different levels of the *Currency* variable, the results are reported in Table 3.1.

Table 3.1 Effect of Currency on Initial Odds of Success based on Odds Ratios

State is a binary indicator for the campaign's final outcome, where 1 represents success and 0 failure to achieve the funding goal initially set. Effect of the entire list of control variables is reported in Table 1C of Appendix C. Based on the initial model and control variables with the single addition of the new foreignness variable. Confidence intervals are based on Wald tests.

Dependent: State		Level	Failed (0)	Success (1)	Odds Ratio
Variables	Currency	USD	3067 (44.8)	3773 (55.2)	-
		AUD	127 (58.5)	90 (41.5)	0.73 (0.54-1.00, p = 0.048)
		CAD	242 (54.1)	205 (45.9)	0.74 (0.59-0.92, p = 0.007)
		GBP	508 (45.6)	606 (54.4)	0.85 (0.74-0.99, p = 0.034)
		NZD	21 (52.5)	19 (47.5)	0.99 (0.49-1.99, p = 0.967)

As observed in Table 3.1 the CAD, AUD and GBP currencies appear to have a small impact on a campaign's odds of success. This impact was significant for the CAD and almost significant for the AUD and GBP in this model, as shown by the p-values of 0.007, 0.048 and 0.034 respectively. On the other hand, the NZD could not accurately be used as there were insufficient observations, leading to a high p-value and confidence interval. Its impact on the campaign's outcome thus remains unknown. Once converted into odds ratios, the results show that the AUD, CAD and GBP each appear to negatively impact a campaign's odds of success. Campaigns using the AUD fared the worst, being only 0.73 times as likely to succeed as the USD. The CAD fared a little better, being 0.74 times as likely to succeed. Lastly, the GBP outperformed the other foreign currencies and was the closest to the USD with an odds ratio of 0.85 but was also negatively impacted. An interesting finding is that the CAD, which is also in North America, didn't perform as well as the GBP all the way in the United Kingdom. Since foreign currencies seem to be of relatively low occurrence compared to USD, it is worthwhile to consider these results more broadly and generalize the currency variable by reducing it to a binary variable that indicates if the currency used for a specific campaign is the USD or not.

3.3 Generalized Currency

The results in Table 3.2 show that the generalized effect of using a currency that is not the USD is negative. Such campaigns appear to be 19% less likely to succeed in their crowdfunding ventures. In terms of success, 55.2% of campaigns that used the USD ultimately succeeded and 50.6% of campaigns that used a foreign currency succeeded. Though this difference is small, it appeared to have an impact based on the p-values of p = 0.001 for the generalized model reframed to highlight liability of foreignness, as well as all specific currencies significantly impacted negatively except for the NZD for which the impact cannot be accurately measured. This provides support to the initial hypothesis that crowdfunding campaigns would be negatively impacted from the use of foreign currencies compared to the USD. However, this marginal effect will be tested later when it is introduced in a complete model with all the liability of foreignness variables.

Table 3.2 Generalized Effect of Currency on Initial Odds of Success based on Odds Ratios

State is a binary indicator for the campaign's final outcome, where 1 represents success and 0 failure to achieve the funding goal initially set. Effect of the entire list of control variables is reported in Table 1C of Appendix C. Based on the initial model and control variables with the single addition of the new foreignness variable. Confidence intervals are based on Wald tests.

Dependent: State		Level	Failed (0)	Success (1)	Odds Ratio
Variable	USD Indicator	Yes	3067 (44.8)	3773 (55.2)	-
		No	898 (49.4)	920 (50.6)	0.81 (0.72-0.91, p = 0.001)

3.4 Impact of Continent

The next foreignness variable analyzed, *Continent*, was introduced to the model using the same methodology to evaluate how a campaign's continent of origin impacts a campaign's odds of success. As such, the second hypothesis proposed is the following:

H2: Campaigns originating from North America are more likely to succeed than their foreign counterparts

To evaluate if this hypothesis is supported, a logistic model was once again developed that added the *Continent* variable to the initial logistic model. The different possible continents in the dataset were North America, South America, Africa, Asia, Oceania and Europe. Of these, North America was defined as the reference level in order to measure the impact of the other continents being tied to a campaign and their odds of success against it as baseline. Looking more closely at the odds ratios of the different levels of the *Continents* variable, the results are reported in Table 3.3.

Table 3.3 Effect of Continent on Initial Odds of Success based on Odds Ratios

State is a binary indicator for the campaign's final outcome, where 1 represents success and 0 failure to achieve the funding goal initially set. Effect of the entire list of control variables is reported in Table 1C of Appendix C. Based on the initial model and control variables with the single addition of the new foreignness variable. Confidence intervals are based on Wald tests.

Dependent: State		Level	Failed (0)	Success (1)	Odds Ratio
Variables	Continent	North America	3248 (45.6)	3872 (54.4)	-
		Africa	7 (41.2)	10 (58.8)	1.23 (0.44-3.62, p = 0.690)
		Asia	28 (39.4)	43 (60.6)	1.56 (0.92-2.70, p = 0.104)
		Europe	535 (44.8)	660 (55.2)	0.94 (0.81-1.08, p = 0.378)
		Oceania	139 (57.7)	102 (42.3)	0.78 (0.58-1.05, p = 0.104)
		South America	8 (57.1)	6 (42.9)	0.59 (0.17-1.89, p = 0.377)

Campaigns from Africa and Asia seem to benefit from a small advantage compared to the ones in North America, but this impact was not significant. The effect was nearly significant for Asia in this model, sitting just above the 10% level. This is confirmed by the p-values that were 0.690 for Africa and 0.104 for Asia. On the other hand, campaigns originating from Europe, Oceania and South America seemed to suffer from their location compared to campaigns in North America. This impact was however not significant in the model developed. Africa and South America had too few observations to properly evaluate. Once converted into odds ratios, Asia was 1.56 times more likely to succeed, while Oceania was 0.78 times less likely to succeed.

The fact that the continent of origin did not appear to be significant differs from previous research that found geography to be relevant to success (Mollick, 2014). Given the relatively low occurrence of campaigns originating from foreign continents, it is valuable to once again generalize this variable into one that reduces the continent variable to a binary indicator of whether a given campaign originated from North America or elsewhere in the world. This will provide insight into the general impact of being a campaign from a foreign geography. The results are reported in Table 3.4.

Table. 3.4 Generalized Effect of Continent on Initial Odds of Success based on Odds Ratios

State is a binary indicator for the campaign's final outcome, where 1 represents success and 0 failure to achieve the funding goal initially set. Effect of the entire list of control variables is reported in Table 1C of Appendix C. Based on the initial model and control variables with the single addition of the new foreignness variable. Confidence intervals are based on Wald tests.

Dependent: State		Level	Failed (0)	Success (1)	Odds Ratio
Variable	North America Indicator	Yes	3248 (45.6)	3872 (54.4)	-
		No	717 (46.6)	821 (53.4)	0.93 (0.82-1.06, p = 0.283)

Despite this transformation of the continent variable, it appears that from this new standpoint, campaign origin has a varying impact on odds of success. The effect observed was a slightly reduced odds of success (7%) of foreign campaigns compared to North American campaigns. However, this effect was not significant in the model developed. The confidence interval showed that the impact on odds of success was unclear. It could however be interesting to know more about the reasons why campaigns from Asia specifically were more likely to succeed.

Based on the identified p-values, it appears that unlike currency, continent of origin was not a significant differentiator when it came to campaign outcomes. On a global level, this does not support hypothesis two that native campaigns from North America would outperform their counterparts from around the world. Moreover, it is worthy to note that Oceania seemed to have slightly lower odds of success, while Asia had slightly higher odds of success, both of which were nearly significant at the 10% level, as indicated by the p-values of 0.104.

3.5 Impact of Language

The last foreignness variable analyzed, *English*, was introduced to the model using the same methodology to evaluate whether and how originating from countries that list English as one of their first languages impacts a campaign's outcome. Since foreign campaigns might have access to a smaller backer pool and thus be less likely to achieve their funding goal, the third hypothesis proposed regarding the language variable is the following:

H3: Campaigns originating from English-speaking countries are more likely to succeed than their foreign counterparts

In order to evaluate whether this hypothesis is supported, a logistic model was developed that added the *English* variable to the initial model. This binary variable indicated whether a specific campaign originated from a country whose official language is English or not. Of these two possibilities, "Yes" was defined as the reference level in order to measure the impact of originating from a country that does not list English as one of their primary languages on the campaign's outcome. Looking more closely at the proportions and odds ratios of different levels of the *PrimaryEnglish* variable, the results are reported in Table 3.5.

Table 3.5 Effect of Language on Initial Odds of Success based on Odds Ratios

State is a binary indicator for the campaign's final outcome, where 1 represents success and 0 failure to achieve the funding goal initially set. Effect of the entire list of control variables is reported in Table 1C of Appendix C. Based on the initial model and control variables with the single addition of the new foreignness variable. Confidence intervals are based on Wald tests.

Dependent: State		Level	Failed (0)	Success (1)	Odds Ratio
Variabel	Language	English	3867 (45.9)	4549 (54.1)	-
		Other	98 (40.5)	144 (59.5)	1.28 (0.95-1.72, p= 0.102)

It appears that originating from a non-English speaking country is a beneficial factor for a campaign. This impact was almost significant in the model developed, based on a p-value of

0.102. Once converted into odds ratios, the results showed that these campaigns appeared 28% more likely to succeed from the beginning than their peers from English-speaking countries.

This impact only being nearly significant at the 10% level, it does not support hypothesis three that stipulated that campaigns would be negatively impacted by their foreign language in the network. Based on this, coming from a country that does not have English as one of its primary languages isn't a barrier to success for international projects. In fact, it appears to be quite the opposite.

3.6 Interaction: Impact of Generalized Currency x Continent

While these initial models provide specific information about the respective impact of the individual variables, it would be interesting to know more about the effect of potential interactions between these variables. Different combinations of campaigns originating from various geographies might be positively or negatively impacted from using different currencies for their campaign over others. As such, the rest of this chapter will expand the scope of analyses to include interactions between variables. The models developed will follow the same method as before, with any given interaction variable being the only addition to the initial model. For instance, to measure the impact of using Kickstarter's native currency, the USD, for campaigns in foreign geographies. This leads to the following hypothesis that will be evaluated in this section:

H4: Campaigns originating elsewhere than North America are more likely to succeed than their foreign counterparts if they use the USD as Currency

In order to measure these possible combinatory effects, the data was reformatted to include the product of two binary indicators. The first binary variable indicated whether or not a given campaign originated from North America, while the second indicated whether or not the currency used was the USD. By formulating these variables to only answer whether or not a campaign's currency and geography was foreign, it was possible to easily combine them into one interaction variable. This double indicator created four possible combinations that replaced the Currency and Continent variables. Using the same methodology as for the initial models, a new logistic model was developed, its outputs (Table 3.6) this time providing more information on different possible scenarios.

Table 3.6 Interaction Effect of Currency & Continent on Initial Odds of Success based on Odds Ratios

State is a binary indicator for the campaign's final outcome, where 1 represents success and 0 failure to achieve the funding goal initially set. Effect of the entire list of control variables is reported in Table 1C of Appendix C. Based on the initial model and control variables with the single addition of the new foreignness variable. Confidence intervals are based on Wald tests.

Dependent: State		Level	Failed (0)	Success (1)	Odds Ratio
Variable	Currency x Continent	USD-NorthAmerica	3006 (45.1)	3661 (54.9)	-
		NotUSD-NorthAmerica	242 (53.4)	211 (46.6)	0.77 (0.62-0.95, p = 0.016)
		NotUSD-Other	656 (48.1)	709 (51.9)	0.84 (0.74-0.96, p = 0.013)
		USD-Other	61 (35.3)	112 (64.7)	1.76 (1.23-2.52, p = 0.002)

The results show that using the USD in geographically foreign countries indeed appears to net campaigners a significant advantage, making them 76% more likely to succeed than the ones using the USD in North America. This finding could possibly be explained by the argument that it gives these campaigns a door to the larger USD-based backer pool from the platform's native region on top of their own foreign backer pool. This would help them drive and sustain their momentum better than the average campaign. In contrast, not using the USD within North America seems to negatively impact a campaign's odds of success by 23%.

The p-value of 0.002 for the USD-Other group provides support to hypothesis four which speculated that international campaigns using the USD would perform better than other foreign campaigns using their own currency. This indicates that foreign campaigns should consider using the USD for their campaign if possible as it would greatly heighten their odds of success. This is also supported by the fact that foreign campaigns using a currency other than the USD appeared 16% less likely to succeed than the ones using the USD in North America.

3.7 Interaction: Impact of Generalized Currency x Language

The next interaction observed the crossing between Currency and Language to see if the use of different combinations of foreign currencies and primary languages had a significant impact on international campaigns' initial odds of success. The hypothesis in this case is the following:

H5: Campaigns coming from non-english speaking countries will benefit from using the USD as currency compared to other foreign currencies

The same methodology is applied wherein binary indicators were developed and combined to measure the more generalized effect of being a foreigner on the Kickstarter platform. The resulting proportions and odds ratios of this double indicator are expressed in Table 3.7 below and show interesting new dynamics not previously picked up by the individual variables.

Table 3.7 Interaction Effect of Currency & Language on Initial Odds of Success based on Odds Ratios

State is a binary indicator for the campaign's final outcome, where 1 represents success and 0 failure to achieve the funding goal initially set. Effect of the entire list of control variables is reported in Table 1C of Appendix C. Based on the initial model and control variables with the single addition of the new foreignness variable. Confidence intervals are based on Wald tests.

Dependent: State		Level	Failed (0)	Success (1)	Odds Ratio
Variable	Currency x Language	USD-EnglishYes	3005 (45.0)	3667 (55.0)	-
		NotUSD-EnglishNo	36 (48.6)	38 (51.4)	0.80 (0.48-1.34, p = 0.401)
		NotUSD-EnglishYes	862 (49.4)	882 (50.6)	0.82 (0.73-0.93, p = 0.001)
		USD-EnglishNo	62 (36.9)	106 (63.1)	1.49 (1.04-2.14, p = 0.029)

The results indicate two important effects that can be described as liability related to foreignness. The first effect demonstrated that not using the USD in an English-speaking country lowers a campaign's odds of success by 18%. The effect was significant in this model, as shown by the p-value of 0.001. The second effect identified supports hypothesis five by showing that using the USD in a country that doesn't list English as its first language raises a campaigner's odds of success by 49%. This effect was significant as well, as shown by the p-value of 0.029 for this variable. This means that foreign campaigners should take the necessary steps to set the USD as their currency if needed to heavily increase their chances of succeeding.

3.8 Interaction: Impact of Generalized Continent x Language

Finally, the last interaction indicator variable was created by crossing Continent and Language to understand whether the use of different combinations of continents of origin and the country's primary languages had a significant impact on international campaigns' initial odds of success. The hypothesis in this case is the following:

H6: Campaigns coming from other continents than North America that also use English as a primary language will perform better than their non-English counterparts

The resulting double-indicator is used to build a new model once more. This interaction variable also displayed some interesting new dynamics previously unseen from the analyses of individual variables. The results in Table 3.8 showed that campaigns from non-English-speaking countries that are not from North America had 35% higher odds of achieving their goals. Meanwhile, campaigns from primarily English-speaking countries not from North America had 13% reduced odds of success based on the significance of the odds ratios in this model. Additionally, the p-values for these two subgroups were significant at 0.049 and 0.055 respectively. Thus, the findings do not support hypothesis six, since campaigns from non-English speaking countries outside of North America generally performed better than campaigns from English-speaking countries outside of North America. Additionally, North American campaigns that are not from English-speaking countries (I.e., Mexico) appeared to have much lower odds of success. Although the sample was very small, and the effect was not significant.

Table 3.8 Interaction Effect of Continent & Language on Initial Odds of Success based on Odds Ratios

State is a binary indicator for the campaign's final outcome, where 1 represents success and 0 failure to achieve the funding goal initially set. Effect of the entire list of control variables is reported in Table 1C of Appendix C. Based on the initial model and control variables with the single addition of the new foreignness variable. Confidence intervals are based on Wald tests.

Dependent: State		Level	Failed (0)	Success (1)	Odds Ratio
Variable	Continent x Language	NorthAmerica-EnglishYes	3237 (45.6)	3864 (54.4)	-
		NorthAmerica-EnglishNo	11 (57.9)	8 (42.1)	0.51 (0.18-1.41, p = 0.202)
		OtherContinent-EnglishNo	87 (39.0)	136 (61.0)	1.35 (1.00-1.85, p = 0.055)
		OtherContinent-EnglishYes	630 (47.9)	685 (52.1)	0.87 (0.76-1.00, p = 0.049)

3.9 Complete Liability of Foreignness Models

While the partial models presented so far provide insight into the impact of the individual liability of foreignness variables, a final model with all the generalized variables and interaction variables that have previously been created is fitted to clarify the impact of liability of foreignness once all factors are accounted for and find out which remain significant. The summary of this model can be found in Table 1C of Appendix C. We use the generalized variables as they are formulated to focus on the liability of foreignness aspect. A model with the original currency, continent, language and interaction variables is also included in Appendix C for comparison purposes. An interesting finding is that the model with all the generalized variables and interactions has the lowest AIC value of the two complete models based on Appendix C. This also supports the fact that bringing liability of foreignness into the analysis made for a slightly more insightful model.

This final model shows that only the currency indicator is nearly significant based on the p-value of 0.09 and that the impact of not using the USD might be much larger than earlier discovered. In the original model, campaigns that didn't use the USD appeared 19% less likely to succeed, while the combined model reveals that the impact of not using the native USD currency appears to make such campaigns 57% less likely to succeed. This thus provides more support in favor of hypothesis one since all the other variables are now accounted for as well.

3.10 Liability of Foreignness Takeaway

The findings from this chapter show that international crowdfunding campaigns are indeed impacted by liability of foreignness in multiple ways. Various economical, geographic and linguistic variables were analyzed and results showed that using Kickstarter's native currency, the USD, increases a campaign's odds of success over other foreign currencies. This is demonstrated by the general effect that campaigns using foreign currencies were 19% less likely to succeed than the ones using the USD. Furthermore, it was found that campaigns from non-English speaking countries had a 49% increase in odds of success if they used the native currency (USD). Using the USD in foreign continents, regardless of language, was also linked to 76% increased odds of success. Through interaction variables, it was found that not using the USD in North America and in foreign continents were both associated with decreased odds of success. Continent of origin did not appear to induce a liability based on the chapter's findings. While some of these findings were not observed once all the foreignness variables were introduced in the model, not using the USD as currency remained significantly detrimental. This was supported by the fact that these campaigns were 57% less likely to succeed.

Through the validation and invalidation of the developed hypotheses, this chapter strongly supports that liability of foreignness is indeed a variable that impacts campaigns foreign to a crowdfunding platform's native levels. Furthermore, the extent of the impact stemming from the different foreignness variables was measured to better define how large the liability is to foreign campaigners. The method also allowed for the identification of certain solutions such as using the network's native currency.

4. Conclusion

Crowdfunding platforms are becoming more and more popular as a source of alternative funding for entrepreneurs (Fernandez-Blanco et al., 2020). This has led many to try and predict campaign success and identify potential markers of success. While the ability to predict success early on is relevant, understanding what makes these campaigns successful and developing actionable insight allows campaign starters to better understand the path to success and the actions to take in order to maximize one's odds. Proper goal setting, early funding milestones, functional clustering of KPI trajectories and assessing the impact of liability of foreignness in crowdfunding networks were all explored as research streams to discover further insight into the drivers of success. Based on this research's findings, it's possible to conclude that crowdfunding success stems from a combination of strong early campaign momentum, better social fitness of the campaigners and the presence of high-pledge backers early on. Achieving 20% of a campaign's funding in the first week was a very strong indicator of success, while achieving it after the second week was deal-breaking. In fact, 93% of campaigns that did not achieve the 20% funding by that the second week ended up failing. Better social fitness, as expressed by social media network size and better performance from social media activities such as shares also characterized highly successful and typically successful projects identified through functional clustering. This method was able to identify different groups of campaigns based on the various metrics used, providing insight into the characteristics that define success and failure. Crowdfunding campaign trajectories illustrates that stronger early momentum and a strong final push in a campaign's final days are consistently observed in highly successful campaigns. Furthermore, it was observed that the liability of foreignness plays into a campaign's initial odds of success. Using another currency than the USD was linked to decreased odds of success, while using the USD in non-English speaking countries appeared to drastically improve them. Geographic variables alone were not found to cause significant liability, but the interaction between geographic variable and other foreignness variables appeared to play a role that was mitigated once all variables were accounted for.

Future research should explore the progression of social media-based KPIs such as comments, posts or shares for which the time-series were not available and might provide insight into the distinctive social media qualities of crowdfunding campaigns. This could also provide further detail into the social fitness characteristics of different campaigner profiles. The same functional clustering process would also be valid to look more deeply into the characteristics of the different categories of campaigns. Furthermore, different platforms that do not work with an “all-or-nothing" model could also reveal different funding characteristics. Another potential stream of research would be causal analysis as only a few targeted interactions were evaluated in this work.

Reference List

Agrawal, A. K., Catalini, C., & Goldfarb, A. (2011). The geography of crowdfunding (No. w16820). National bureau of economic research.

Barbi, M., & Bigelli, M. (2017). Crowdfunding practices in and outside the US. Research in International Business and Finance, 42, 208–223. https://doi.org/10.1016/j.ribaf.2017.05.013

Bouveyron, C., Côme, E., & Jacques, J. (2015). The discriminative functional mixture model for a comparative analysis of bike sharing systems. *The Annals of Applied Statistics*, *9*(4), 1726-1760.

Crosetto, P., & Regner, T. (2014). Crowdfunding: Determinants of success and funding dynamics. Jena Economic Research Papers.

Diaz, J. R., & Cacheda, B. G. (2016, March). Financing social activism: Crowdfunding and advocatory social movement in Spain in times of crisis. In *Proceedings of the 9th international conference on theory and practice of electronic governance* (pp. 139-148).

Etter, V., Grossglauser, M., & Thiran, P. (2013). Launch hard or go home! Predicting the success of Kickstarter campaigns. Proceedings of the First ACM Conference on Online Social Networks, 177–182.

Fernandez-Blanco, A., Villanueva-Balsera, J., Rodriguez-Montequin, V., & Moran-Palacios, H. (2020). Key factors for project crowdfunding success: An empirical study. Sustainability, 12(2), 599.

Giordani, P., Perna, S., Bianchi, A., Pizzulli, A., Tripodi, S., & Matricardi, P. M. (2020). A study of longitudinal mobile health data through fuzzy clustering methods for functional data: The case of allergic rhinoconjunctivitis in childhood. PloS One, 15(11), e0242197. https://doi.org/10.1371/journal.pone.0242197

Hartmann, F., Grottolo, G., Wang, X., & Lunesu, M. I. (2019, February). Alternative fundraising: success factors for blockchain-based vs. conventional crowdfunding. In *2019 IEEE international workshop on blockchain oriented software engineering (IWBOSE)* (pp. 38-43). IEEE.

Irastorza, N., & Peña, I. (2014). Immigrant entrepreneurship: Does the liability of foreignness matter? Business and Management Research, 3(1), 1–17.

Kaur, H., & Gera, J. (2017). Effect of social media connectivity on success of crowdfunding campaigns. *Procedia computer science*, *122*, 767-774.

Kaur, P., Deshmukh, S., Apoorva, P., & Batra, S. (2022). Analysis and Outcome Prediction of Crowdfunding Campaigns. *International Journal of Information Retrieval Research (IJIRR)*, *12*(1), 1-14.

Kickstarter Stats—Kickstarter. (n.d.). Retrieved August 16, 2022, from https://www.kickstarter.com/help/stats

Kindler, A., Golosovsky, M., & Solomon, S. (2019). Early prediction of the outcome of Kickstarter campaigns: Is the success due to virality? Palgrave Communications, 5(1), 1–6. https://doi.org/10.1057/s41599-019-0261-6

Li, Y., Rakesh, V., & Reddy, C. K. (2016). Project success prediction in crowdfunding environments. Proceedings of the Ninth ACM International Conference on Web Search and Data Mining, 247–256.

Mollick, E. (2014). The dynamics of crowdfunding: An exploratory study. Journal of Business Venturing, 29(1), 1–16. https://doi.org/10.1016/j.jbusvent.2013.06.005

Pastoriza, D., Plante, J.-F., & Lakhlef, N. (2021). Are Foreigners at Disadvantage in a Global Labor Market? Journal of Sports Economics, 22(6), 615–638. https://doi.org/10.1177/1527002521995870

Zaheer, S. (1995). Overcoming the liability of foreignness. Academy of Management journal, 38(2), 341-363.

Zhou, N., & Guillén, M. F. (2015). From home country to home base: A dynamic approach to the liability of foreignness. Strategic Management Journal, 36(6), 907–917. https://doi.org/10.1002/smj.2242

Appendix

Appendix A: Model Variables

Table A1: Complete list of variables in the dataset

Project Features: Information about the project itself

Name: Name of the project

URL: Web link to the project's Kickstarter page

State: Indicates if the project was ultimately "Failed" or "Successful"

Currency: Indicates the designated currency for this project

Top Category: Topical Kickstarter category (15) to which the project is assigned

Category: Subcategory associated to the project

Creator: Name of the project's creator

Location: Location of the project

Updates: Number of updates given throughout the campaign

Comments: Number of comments made throughout the campaign

Rewards: Number of rewards earned by backers

Goal: Financial goal set by the creator

Pledged: Amount pledged to the project at the end of the campaign

Backers: Amount of backers committed to the project at the end of the campaign

Facebook Shares: Number of times the project was shared on Facebook

Start: Start date of the campaign

End: End date of the campaign

Duration in Days: Duration of the campaign selected by the creator (1-60 days)

Has Video: Indicates if a video is available for this project

Latitude: Latitude of the project location

Longitude: Longitude of the project location

Start Timestamp (UTC): Timestamp of when the campaign started

End Timestamp (UTC): Timestamp of when the campaign ended

Videos: Number of videos available for this project

Images: Number of images available for this project

Words (Description): Number of words in the project's description

Words (Risks and Challenges): Number of words in the project's "Risks and Challenges" section

FAQs: Number of Frequently Answered Questions about the project

Creator features: Information about the project's creator

Facebook Connected: Indicates if the creator is connected to Facebook

Facebook Friends: Indicates the number of Facebook friends in the creator's network

Creator -# Projects Created: Number of projects started by this creator

Creator - # Projects Backed: Number of projects backed by this creator

Temporal features: Daily data collection of funds pledged and backer count.

Project ID: Unique identifiers for the different projects

Timestamp (UTC): Daily timestamp associated with this project

Raised: Total amount raised at this specific timestamp

Funders: Total number of backers at this specific timestamp

Appendix B: Early Funding Milestone Model Outputs

Table 1B: Complete early campaign milestone logistic model outputs of Chapter 1

Model Results

	Dependent variable:				
	State				
	Base	Quick20	Good20	Backer Per Day (7)	Backer Per Day (14)
Quick20Yes		44.12^{***}			
		p = 0.00			
Good20Yes			81.59^{***}		
			p = 0.00		
BackerPerDay				1.87^{***}	2.93^{***}
				p = 0.00	p = 0.00
Goal	0.92^{***}	0.96^{***}	0.96^{***}	0.82^{***}	0.79^{***}
	p = 0.00	p = 0.0000	p = 0.0003	p = 0.00	p = 0.00
Facebook.ConnectedYes	0.79^{**}	0.87	0.79	0.88	0.88
	p = 0.04	p = 0.35	p = 0.15	p = 0.35	p = 0.35
Facebook.Friends	1.05^{***}	1.03^{***}	1.03^{***}	1.02^{**}	1.02^{*}
	p = 0.0000	p = 0.004	p = 0.01	p = 0.02	p = 0.08
Duration.in.Days					
Has.VideoYes	2.72^{***}	2.55^{***}	1.90^{***}	2.26^{***}	2.13^{***}
	p = 0.00	p = 0.0000	p = 0.001	p = 0.0000	p = 0.0000
X..Videos	1.09	1.05	1.10	0.96	0.97
	p = 0.17	p = 0.56	p = 0.27	p = 0.64	p = 0.77
X..Images	1.05^{***}	1.02^{**}	1.03^{***}	1.01	1.01
	p = 0.00	p = 0.02	p = 0.005	p = 0.14	p = 0.20
X..Words..Description.	1.04^{***}	1.03^{*}	1.01	1.01	1.01
	p = 0.0002	p = 0.07	p = 0.53	p = 0.32	p = 0.65
X..FAQs	1.24^{***}	1.09^{**}	1.11^{**}	1.07	1.08^{*}
	p = 0.00	p = 0.04	p = 0.03	p = 0.11	p = 0.09
CategoryArt	0.55	0.54	0.42	1.08	1.19
	p = 0.38	p = 0.51	p = 0.42	p = 0.93	p = 0.85
CategoryArt Book	0.97	1.36	0.98	0.99	0.83
	p = 0.97	p = 0.77	p = 0.99	p = 1.00	p = 0.86
CategoryChildren's Book	0.44	0.59	0.55	1.02	1.01
	p = 0.22	p = 0.56	p = 0.57	p = 0.99	p = 1.00
CategoryClassical Music	2.24	3.93	2.76	3.68	3.52
	p = 0.34	p = 0.20	p = 0.44	p = 0.19	p = 0.24
CategoryComics	0.68	0.60	0.36	0.76	0.68
	p = 0.54	p = 0.56	p = 0.32	p = 0.74	p = 0.67
CategoryConceptual Art	0.30	0.86	0.55	1.01	0.94
	p = 0.24	p = 0.91	p = 0.71	p = 1.00	p = 0.97
CategoryCountry	Folk	2.39	3.08	1.92	3.42
	p = 0.22	p = 0.23	p = 0.56	p = 0.17	p = 0.21
CategoryCrafts	0.45	0.44	0.35	1.01	1.05
	p = 0.24	p = 0.39	p = 0.34	p = 1.00	p = 0.97
CategoryDance	1.64	2.35	2.63	4.03	4.08
	p = 0.50	p = 0.38	p = 0.40	p = 0.12	p = 0.15
CategoryDesign	0.61	0.63	0.48	0.67	0.61
	p = 0.48	p = 0.63	p = 0.52	p = 0.67	p = 0.63
CategoryDigital Art	0.34	0.32	0.20	1.52	1.34
	p = 0.27	p = 0.40	p = 0.26	p = 0.71	p = 0.81
CategoryDocumentary	0.88	1.43	1.08	2.37	2.49
	p = 0.85	p = 0.68	p = 0.95	p = 0.29	p = 0.31
CategoryElectronic Music	0.46	0.30	0.32	0.53	0.55
	p = 0.29	p = 0.23	p = 0.33	p = 0.50	p = 0.55
CategoryFashion	0.23^{**}	0.33	0.25	0.60	0.62
	p = 0.02	p = 0.20	p = 0.17	p = 0.53	p = 0.59
CategoryFiction	0.25^{**}	0.29	0.23	0.57	0.62
	p = 0.04	p = 0.17	p = 0.15	p = 0.50	p = 0.60
CategoryFilm	Video	0.32^{*}	0.43	0.40	0.88
	p = 0.08	p = 0.35	p = 0.38	p = 0.88	p = 0.97
CategoryFood	0.44	0.51	0.38	0.83	0.81
	p = 0.19	p = 0.44	p = 0.34	p = 0.82	p = 0.81
CategoryGames	0.14^{***}	0.32	0.45	0.25	0.31
	p = 0.01	p = 0.26	p = 0.50	p = 0.16	p = 0.27
CategoryGraphic Design	0.32	0.18	0.17	0.16	0.16
	p = 0.16	p = 0.11	p = 0.13	p = 0.15	p = 0.19
CategoryHardware	0.52	0.29	0.34	0.57	0.62

	p = 0.34	p = 0.19	p = 0.32	p = 0.55	p = 0.63
CategoryHip-Hop	0.25*	0.46	0.56	0.71	0.77
	p = 0.07	p = 0.45	p = 0.64	p = 0.71	p = 0.80
CategoryIllustration	0.75	0.59	0.52	1.50	1.56
	p = 0.73	p = 0.64	p = 0.61	p = 0.68	p = 0.67
CategoryIndie Rock	2.18	2.90	1.83	3.65	3.50
	p = 0.26	p = 0.25	p = 0.59	p = 0.14	p = 0.18
CategoryJazz	0.94	0.59	0.27	1.57	1.73
	p = 0.94	p = 0.67	p = 0.30	p = 0.65	p = 0.61
CategoryJournalism	0.65	0.47	0.71	1.22	1.34
	p = 0.62	p = 0.55	p = 0.81	p = 0.87	p = 0.83
CategoryMetal	0.44	0.38	0.32	1.16	1.29
	p = 0.32	p = 0.41	p = 0.38	p = 0.88	p = 0.81
CategoryMixed Media	0.47	0.74	0.82	1.19	1.15
	p = 0.30	p = 0.76	p = 0.87	p = 0.85	p = 0.89
CategoryMusic	0.99	1.28	0.72	1.71	1.80
	p = 0.99	p = 0.77	p = 0.75	p = 0.51	p = 0.51
CategoryNarrative Film	0.72	1.51	1.11	1.88	2.13
	p = 0.64	p = 0.65	p = 0.93	p = 0.47	p = 0.43
CategoryNonfiction	0.20**	0.35	0.25	0.43	0.41
	p = 0.02	p = 0.25	p = 0.19	p = 0.32	p = 0.35
CategoryOpen Software	0.16**	0.33	0.37	0.61	0.65
	p = 0.03	p = 0.30	p = 0.43	p = 0.61	p = 0.69
CategoryPainting	0.49	0.55	0.51	1.46	1.58
	p = 0.32	p = 0.54	p = 0.56	p = 0.67	p = 0.63
CategoryPerformance Art	0.29	0.82	0.49	0.83	0.89
	p = 0.12	p = 0.85	p = 0.58	p = 0.85	p = 0.91
CategoryPeriodical	0.35	0.19	0.14	0.51	0.33
	p = 0.23	p = 0.18	p = 0.13	p = 0.55	p = 0.39
CategoryPhotography	0.41	0.58	0.26	1.03	0.99
	p = 0.17	p = 0.55	p = 0.19	p = 0.98	p = 1.00
CategoryPoetry	0.26	0.29	0.46	0.53	0.58
	p = 0.12	p = 0.32	p = 0.57	p = 0.55	p = 0.62
CategoryPop	0.72	1.14	1.05	1.78	1.80
	p = 0.66	p = 0.90	p = 0.97	p = 0.53	p = 0.56
CategoryProduct Design	0.28**	0.24*	0.15*	0.35	0.37
	p = 0.05	p = 0.10	p = 0.06	p = 0.20	p = 0.27
CategoryPublic Art	0.76	1.58	0.84	1.60	1.71
	p = 0.75	p = 0.67	p = 0.89	p = 0.67	p = 0.64
CategoryPublishing	0.46	0.39	0.46	0.75	0.80
	p = 0.27	p = 0.34	p = 0.49	p = 0.76	p = 0.82
CategoryRadio	Podcast	1.50	3.65	1.14	2.56
	p = 0.68	p = 0.26	p = 0.94	p = 0.42	p = 0.48
CategoryRock	1.17	1.63	1.44	2.23	2.15
	p = 0.82	p = 0.59	p = 0.74	p = 0.34	p = 0.40
CategorySculpture	0.39	0.67	0.28	1.02	1.24
	p = 0.28	p = 0.74	p = 0.35	p = 0.99	p = 0.85
CategoryShort Film	1.34	2.43	1.16	3.21	3.43
	p = 0.65	p = 0.30	p = 0.89	p = 0.15	p = 0.16
CategoryTabletop Games	0.64	0.31	0.34	0.28	0.30
	p = 0.49	p = 0.19	p = 0.29	p = 0.14	p = 0.20
CategoryTechnology	0.28*	0.17*	0.20	0.38	0.43
	p = 0.08	p = 0.07	p = 0.14	p = 0.31	p = 0.42
CategoryTheater	1.06	1.27	1.43	2.35	2.43
	p = 0.93	p = 0.79	p = 0.74	p = 0.30	p = 0.33
CategoryVideo Games	0.13***	0.15**	0.15*	0.11***	0.10**
	p = 0.002	p = 0.04	p = 0.07	p = 0.01	p = 0.02
CategoryWebseries	0.39	0.34	0.27	0.91	1.14
	p = 0.17	p = 0.25	p = 0.23	p = 0.92	p = 0.89
CategoryWorld Music	0.45	0.54	0.55	0.80	0.96
	p = 0.33	p = 0.59	p = 0.66	p = 0.83	p = 0.97
Constant	0.77	0.15**	0.10**	0.28	0.28
	p = 0.68	p = 0.03	p = 0.02	p = 0.11	p = 0.15
Observations	2,797	2,797	2,775	2,797	2,775
Log Likelihood	-1,595.46	-994.72	-832.13	-1,127.95	-1,037.66
Akaike Inf. Crit.	3,308.92	2,109.45	1,784.25	2,375.90	2,195.31
Note:					*p<0.1; **p<0.05; ***p<0.01

Appendix C: Liability of Foreignness Model Outputs

Table 1C: Complete liability of foreignness logistic model outputs of Chapter 3

Model Results

	Dependent variable:										
	Initial	Currency	Generalized Currency	Continent	Generalized Continent	English Country	State Currency-Continent Interaction	Currency-Language Interaction	Continent-Language Interaction	All Variables Combined Model	All Liability of Foreignness Model
Goal	0.92*** p = 0.00	0.92*** p = 0.00	0.92*** p = 0.00	0.92*** p = 0.00	0.92*** p = 0.00	0.92*** p = 0.00	0.92*** p = 0.00	0.92*** p = 0.00	0.92*** p = 0.00	0.92*** p = 0.00	0.92*** p = 0.00
CurrencyAUD		0.73** p = 0.05								0.47 p = 0.24	
CurrencyCAD		0.74*** p = 0.01								0.30* p = 0.07	
CurrencyGBP		0.85** p = 0.04								0.43* p = 0.09	
CurrencyNZD		0.99 p = 0.97								0.63 p = 0.53	
USDIndicatorNOT-USD			0.81*** p = 0.001								0.43* p = 0.09
ContinentAfrica				1.23 p = 0.69						1.44 p = 0.63	
ContinentAsia				1.56 p = 0.11						2.04 p = 0.24	
ContinentEurope				0.94 p = 0.38						2.03 p = 0.16	
ContinentOceania				0.78 p = 0.11						1.55 p = 0.50	
ContinentSouth America				0.59 p = 0.38						0.60 p = 0.54	
NorthAmOTHER					0.93 p = 0.29						1.98 p = 0.16
PrimaryEnglishNo						1.28 p = 0.11				0.99 p = 1.00	0.87 p = 0.79
CurrencyContinentNOT-USD_NORTH-AMERICA							0.77** p = 0.02			2.49 p = 0.16	1.80 p = 0.24
CurrencyContinentNOT-USD_OTHER							0.84** p = 0.02				
CurrencyContinentUSD_OTHER							1.76*** p = 0.003				
CurrencyLanguageNOT-USD_No								0.80 p = 0.41		0.98 p = 0.98	1.11 p = 0.86
CurrencyLanguageNOT-USD_Yes								0.82*** p = 0.002			
CurrencyLanguageUSD_No								1.49** p = 0.03			
ContinentLanguageNORTH-AMERICA_No									0.51 p = 0.21	0.51 p = 0.36	0.59 p = 0.46
ContinentLanguageOTHER_No									1.35* p = 0.06		
ContinentLanguageOTHER_Yes									0.87** p = 0.05		
Facebook.ConnectedYes	0.75*** p = 0.0000	0.75*** p = 0.0000	0.75*** p = 0.0000	0.75*** p = 0.0000	0.75*** p = 0.0000	0.76*** p = 0.0000	0.75*** p = 0.0000	0.75*** p = 0.0000	0.75*** p = 0.0000	0.75*** p = 0.0000	0.75*** p = 0.0000
Facebook.Friends	1.06*** p = 0.00	1.06*** p = 0.00	1.06*** p = 0.00	1.06*** p = 0.00	1.06*** p = 0.00	1.06*** p = 0.00	1.06*** p = 0.00	1.06*** p = 0.00	1.06*** p = 0.00	1.06*** p = 0.00	1.06*** p = 0.00
Duration.in.Days	0.98*** p = 0.00	0.98*** p = 0.00	0.98*** p = 0.00	0.98*** p = 0.00	0.98*** p = 0.00	0.98*** p = 0.00	0.98*** p = 0.00	0.98*** p = 0.00	0.98*** p = 0.00	0.98*** p = 0.00	0.98*** p = 0.00
Has.VideoYes	2.25*** p = 0.00	2.22*** p = 0.00	2.22*** p = 0.00	2.24*** p = 0.00	2.24*** p = 0.00	2.25*** p = 0.00	2.22*** p = 0.00	2.21*** p = 0.00	2.23*** p = 0.00	2.22*** p = 0.00	2.22*** p = 0.00
X..Videos	1.11*** p = 0.004	1.11*** p = 0.004	1.11*** p = 0.004	1.10*** p = 0.005	1.10*** p = 0.005	1.10*** p = 0.004	1.11*** p = 0.004	1.10*** p = 0.005	1.11*** p = 0.004	1.11*** p = 0.004	1.11*** p = 0.004
X..Images	1.05***	1.05***	1.05***	1.05***	1.05***	1.05***	1.05***	1.05***	1.05***	1.05***	1.05***

	p = 0.00	p = 0.00	p = 0.00	p = 0.00	p = 0.00	p = 0.00	p = 0.00	p = 0.00	p = 0.00	p = 0.00	p = 0.00
X..Words..Description.	1.04^{***}	1.04^{***}	1.04^{***}	1.04^{***}	1.04^{***}	1.04^{***}	1.04^{***}	1.04^{***}	1.04^{***}	1.04^{***}	1.04^{***}
	p = 0.00	p = 0.00	p = 0.00	p = 0.00	p = 0.00	p = 0.00	p = 0.00	p = 0.00	p = 0.00	p = 0.00	p = 0.00
X..FAQs	1.29^{***}	1.29^{***}	1.29^{***}	1.29^{***}	1.29^{***}	1.28^{***}	1.29^{***}	1.29^{***}	1.28^{***}	1.28^{***}	1.28^{***}
	p = 0.00	p = 0.00	p = 0.00	p = 0.00	p = 0.00	p = 0.00	p = 0.00	p = 0.00	p = 0.00	p = 0.00	p = 0.00
CategoryArt	1.42	1.37	1.37	1.41	1.42	1.40	1.33	1.33	1.38	1.34	1.34
	p = 0.30	p = 0.36	p = 0.36	p = 0.31	p = 0.30	p = 0.33	p = 0.40	p = 0.40	p = 0.34	p = 0.39	p = 0.39
CategoryArt Book	1.64	1.58	1.59	1.63	1.64	1.62	1.56	1.56	1.61	1.56	1.56
	p = 0.22	p = 0.26	p = 0.25	p = 0.23	p = 0.22	p = 0.23	p = 0.28	p = 0.27	p = 0.24	p = 0.28	p = 0.27
CategoryChildren's Book	0.99	0.96	0.95	0.98	0.98	0.99	0.96	0.95	0.98	0.97	0.96
	p = 0.97	p = 0.90	p = 0.89	p = 0.96	p = 0.96	p = 0.98	p = 0.91	p = 0.89	p = 0.95	p = 0.92	p = 0.91
CategoryClassical Music	4.57^{***}	4.32^{***}	4.34^{***}	4.53^{***}	4.53^{***}	4.53^{***}	4.29^{***}	4.26^{***}	4.51^{***}	4.33^{***}	4.34^{***}
	p = 0.0002	p = 0.0003	p = 0.0003	p = 0.0002	p = 0.0002	p = 0.0002	p = 0.0003	p = 0.0004	p = 0.0002	p = 0.0003	p = 0.0003
CategoryComics	1.28	1.23	1.23	1.28	1.27	1.29	1.23	1.23	1.27	1.25	1.24
	p = 0.45	p = 0.53	p = 0.54	p = 0.46	p = 0.47	p = 0.44	p = 0.53	p = 0.54	p = 0.47	p = 0.51	p = 0.52
CategoryConceptual Art	0.47	0.45	0.45	0.47	0.47	0.47	0.45	0.45	0.46	0.45	0.45
	p = 0.21	p = 0.18	p = 0.18	p = 0.20	p = 0.20	p = 0.21	p = 0.18	p = 0.18	p = 0.19	p = 0.19	p = 0.19
CategoryCountry	Folk	3.83^{***}	3.63^{***}	3.62^{***}	3.78^{***}	3.78^{***}	3.84^{***}	3.65^{***}	3.62^{***}	3.75^{***}	3.68^{***}
	p = 0.0003	p = 0.0004	p = 0.0004	p = 0.0003	p = 0.0003	p = 0.0003	p = 0.0004	p = 0.0004	p = 0.0003	p = 0.0004	p = 0.0004
CategoryCrafts	0.84	0.80	0.81	0.84	0.84	0.83	0.79	0.79	0.82	0.80	0.79
	p = 0.63	p = 0.56	p = 0.57	p = 0.64	p = 0.63	p = 0.62	p = 0.52	p = 0.53	p = 0.60	p = 0.54	p = 0.52
CategoryDance	2.65^{***}	2.51^{**}	2.51^{**}	2.62^{***}	2.63^{***}	2.65^{***}	2.49^{**}	2.49^{**}	2.60^{***}	2.50^{**}	2.50^{**}
	p = 0.01	p = 0.02	p = 0.02	p = 0.01	p = 0.01	p = 0.01	p = 0.02	p = 0.02	p = 0.01	p = 0.02	p = 0.02
CategoryDesign	1.11	1.09	1.09	1.12	1.11	1.11	1.09	1.08	1.12	1.12	1.10
	p = 0.78	p = 0.81	p = 0.82	p = 0.77	p = 0.79	p = 0.79	p = 0.81	p = 0.83	p = 0.77	p = 0.76	p = 0.79
CategoryDigital Art	0.48	0.45	0.46	0.48	0.48	0.48	0.44	0.45	0.47	0.44	0.44
	p = 0.16	p = 0.13	p = 0.13	p = 0.16	p = 0.16	p = 0.16	p = 0.12	p = 0.13	p = 0.15	p = 0.12	p = 0.11
CategoryDocumentary	1.99^{**}	1.93^{**}	1.93^{**}	1.96^{**}	1.99^{**}	1.96^{**}	1.90^{**}	1.90^{**}	1.95^{**}	1.92^{**}	1.92^{**}
	p = 0.04	p = 0.05	p = 0.05	p = 0.04	p = 0.04	p = 0.04	p = 0.05	p = 0.05	p = 0.05	p = 0.05	p = 0.05
CategoryElectronic Music	1.00	0.99	0.98	1.02	1.01	1.00	0.97	0.97	1.01	0.99	0.97
	p = 1.00	p = 0.99	p = 0.97	p = 0.97	p = 0.99	p = 1.00	p = 0.95	p = 0.96	p = 0.99	p = 0.99	p = 0.95
CategoryFashion	0.61	0.58^{*}	0.58^{*}	0.60	0.60	0.61	0.59^{*}	0.58^{*}	0.60	0.59^{*}	0.59^{*}
	p = 0.12	p = 0.10	p = 0.10	p = 0.12	p = 0.12	p = 0.12	p = 0.10	p = 0.10	p = 0.12	p = 0.10	p = 0.10
CategoryFiction	0.63	0.60	0.60	0.63	0.63	0.63	0.60	0.60	0.62	0.61	0.61
	p = 0.17	p = 0.13	p = 0.13	p = 0.17	p = 0.16	p = 0.17	p = 0.13	p = 0.13	p = 0.16	p = 0.14	p = 0.13
CategoryFilm	Video	1.13	1.10	1.11	1.13	1.14	1.13	1.10	1.10	1.14	1.11
	p = 0.71	p = 0.78	p = 0.76	p = 0.71	p = 0.70	p = 0.71	p = 0.78	p = 0.78	p = 0.70	p = 0.76	p = 0.77
CategoryFood	1.41	1.33	1.33	1.40	1.39	1.41	1.35	1.34	1.39	1.35	1.35
	p = 0.29	p = 0.38	p = 0.37	p = 0.30	p = 0.30	p = 0.28	p = 0.36	p = 0.37	p = 0.31	p = 0.35	p = 0.35
CategoryGames	0.58	0.57	0.57	0.58	0.58	0.58	0.58	0.57	0.58	0.58	0.58
	p = 0.18	p = 0.18	p = 0.18	p = 0.19	p = 0.18	p = 0.19	p = 0.18	p = 0.18	p = 0.19	p = 0.19	p = 0.18
CategoryGraphic Design	1.24	1.19	1.19	1.22	1.23	1.23	1.18	1.17	1.21	1.18	1.18
	p = 0.59	p = 0.67	p = 0.67	p = 0.62	p = 0.60	p = 0.60	p = 0.69	p = 0.69	p = 0.63	p = 0.69	p = 0.68
CategoryHardware	1.04	1.00	1.00	1.04	1.04	1.04	1.00	0.99	1.03	1.01	1.00
	p = 0.92	p = 1.00	p = 1.00	p = 0.93	p = 0.93	p = 0.91	p = 1.00	p = 0.99	p = 0.94	p = 0.99	p = 1.00
CategoryHip-Hop	0.38^{**}	0.36^{**}	0.36^{**}	0.37^{**}	0.38^{**}	0.38^{**}	0.36^{**}	0.36^{**}	0.37^{**}	0.36^{**}	0.36^{**}
	p = 0.02	p = 0.02	p = 0.02	p = 0.02	p = 0.02	p = 0.02	p = 0.02	p = 0.02	p = 0.02	p = 0.02	p = 0.02
CategoryIllustration	2.25^{*}	2.25^{*}	2.23^{*}	2.28^{*}	2.25^{*}	2.26^{*}	2.24^{*}	2.22^{*}	2.27^{*}	2.28^{*}	2.24^{*}
	p = 0.07	p = 0.07	p = 0.08	p = 0.07	p = 0.07	p = 0.07	p = 0.07	p = 0.08	p = 0.07	p = 0.07	p = 0.07
CategoryIndie Rock	3.74^{***}	3.51^{***}	3.53^{***}	3.71^{***}	3.71^{***}	3.75^{***}	3.53^{***}	3.52^{***}	3.68^{***}	3.54^{***}	3.54^{***}
	p = 0.0003	p = 0.001	p = 0.0005	p = 0.0003	p = 0.0003	p = 0.0003	p = 0.0005	p = 0.0005	p = 0.0004	p = 0.0005	p = 0.0005
CategoryJazz	3.26^{**}	3.06^{**}	3.06^{**}	3.21^{**}	3.21^{**}	3.27^{**}	3.09^{**}	3.07^{**}	3.19^{**}	3.10^{**}	3.09^{**}
	p = 0.02	p = 0.02	p = 0.02	p = 0.02	p = 0.02	p = 0.02	p = 0.02	p = 0.02	p = 0.02	p = 0.02	p = 0.02
CategoryJournalism	0.71	0.69	0.69	0.71	0.71	0.72	0.69	0.69	0.71	0.69	0.68
	p = 0.51	p = 0.47	p = 0.47	p = 0.50	p = 0.50	p = 0.51	p = 0.46	p = 0.47	p = 0.50	p = 0.47	p = 0.46
CategoryMetal	1.06	1.06	1.06	1.07	1.06	1.07	1.06	1.06	1.08	1.07	1.07
	p = 0.91	p = 0.91	p = 0.91	p = 0.90	p = 0.90	p = 0.90	p = 0.90	p = 0.91	p = 0.88	p = 0.89	p = 0.90
CategoryMixed Media	1.11	1.07	1.07	1.10	1.10	1.11	1.06	1.06	1.09	1.09	1.06
	p = 0.79	p = 0.87	p = 0.87	p = 0.81	p = 0.80	p = 0.80	p = 0.88	p = 0.88	p = 0.83	p = 0.84	p = 0.88
CategoryMusic	2.44^{***}	2.32^{***}	2.32^{***}	2.42^{***}	2.42^{***}	2.45^{***}	2.34^{***}	2.33^{***}	2.41^{***}	2.36^{***}	2.35^{***}
	p = 0.01	p = 0.01	p = 0.01	p = 0.01	p = 0.01	p = 0.01	p = 0.01	p = 0.01	p = 0.01	p = 0.01	p = 0.01
CategoryNarrative Film	1.64	1.57	1.58	1.62	1.64	1.63	1.56	1.56	1.62	1.56	1.56
	p = 0.17	p = 0.21	p = 0.20	p = 0.18	p = 0.17	p = 0.17	p = 0.22	p = 0.21	p = 0.18	p = 0.21	p = 0.21
CategoryNonfiction	0.66	0.63	0.63	0.65	0.65	0.65	0.61	0.61	0.64	0.62	0.61
	p = 0.22	p = 0.18	p = 0.17	p = 0.21	p = 0.21	p = 0.21	p = 0.15	p = 0.15	p = 0.19	p = 0.16	p = 0.15

CategoryOpen Software	0.27***	0.27***	0.27***	0.28***	0.27***	0.28***	0.27***	0.27***	0.28***	0.27***	0.27***
	p = 0.01	p = 0.005	p = 0.005	p = 0.01	p = 0.01	p = 0.01	p = 0.005	p = 0.005	p = 0.01	p = 0.005	p = 0.005
CategoryPainting	0.78	0.74	0.74	0.78	0.77	0.77	0.73	0.73	0.77	0.76	0.74
	p = 0.51	p = 0.42	p = 0.43	p = 0.51	p = 0.49	p = 0.50	p = 0.42	p = 0.41	p = 0.49	p = 0.47	p = 0.43
CategoryPerformance Art	1.31	1.27	1.27	1.30	1.30	1.32	1.28	1.27	1.30	1.28	1.28
	p = 0.49	p = 0.55	p = 0.54	p = 0.50	p = 0.50	p = 0.48	p = 0.53	p = 0.54	p = 0.50	p = 0.53	p = 0.53
CategoryPeriodical	1.44	1.42	1.43	1.46	1.45	1.43	1.39	1.40	1.44	1.43	1.39
	p = 0.43	p = 0.45	p = 0.44	p = 0.42	p = 0.42	p = 0.44	p = 0.48	p = 0.47	p = 0.43	p = 0.44	p = 0.47
CategoryPhotography	1.01	0.98	0.98	1.00	1.01	1.00	0.96	0.96	1.00	0.98	0.97
	p = 0.97	p = 0.97	p = 0.96	p = 1.00	p = 0.98	p = 1.00	p = 0.91	p = 0.91	p = 1.00	p = 0.95	p = 0.94
CategoryPoetry	1.05	1.02	1.01	1.04	1.04	1.05	1.01	1.01	1.04	1.01	1.01
	p = 0.91	p = 0.97	p = 0.98	p = 0.93	p = 0.93	p = 0.91	p = 0.99	p = 0.98	p = 0.93	p = 0.98	p = 0.99
CategoryPop	1.55	1.51	1.51	1.55	1.55	1.56	1.52	1.51	1.55	1.52	1.53
	p = 0.24	p = 0.27	p = 0.27	p = 0.25	p = 0.25	p = 0.24	p = 0.26	p = 0.27	p = 0.24	p = 0.26	p = 0.26
CategoryProduct Design	0.74	0.71	0.71	0.74	0.73	0.74	0.71	0.71	0.73	0.72	0.71
	p = 0.35	p = 0.30	p = 0.29	p = 0.35	p = 0.34	p = 0.35	p = 0.29	p = 0.29	p = 0.33	p = 0.31	p = 0.30
CategoryPublic Art	3.00***	2.85***	2.87***	2.97***	2.99***	2.99***	2.86***	2.84***	3.00***	2.91***	2.90***
	p = 0.01	p = 0.01	p = 0.01	p = 0.01	p = 0.01	p = 0.01	p = 0.01	p = 0.01	p = 0.01	p = 0.01	p = 0.01
CategoryPublishing	0.90	0.87	0.87	0.91	0.90	0.90	0.86	0.86	0.89	0.86	0.86
	p = 0.78	p = 0.70	p = 0.71	p = 0.79	p = 0.78	p = 0.77	p = 0.67	p = 0.68	p = 0.76	p = 0.69	p = 0.67
CategoryRadio	Podcast	1.39	1.31	1.31	1.36	1.38	1.38	1.30	1.30	1.35	1.32
	p = 0.47	p = 0.56	p = 0.56	p = 0.50	p = 0.48	p = 0.48	p = 0.57	p = 0.57	p = 0.51	p = 0.54	p = 0.56
CategoryRock	2.84***	2.68***	2.68***	2.81***	2.80***	2.84***	2.70***	2.68***	2.78***	2.73***	2.71***
	p = 0.003	p = 0.005	p = 0.005	p = 0.003	p = 0.003	p = 0.003	p = 0.005	p = 0.005	p = 0.004	p = 0.004	p = 0.005
CategorySculpture	0.70	0.66	0.66	0.70	0.69	0.70	0.67	0.66	0.68	0.67	0.67
	p = 0.44	p = 0.37	p = 0.37	p = 0.43	p = 0.42	p = 0.43	p = 0.38	p = 0.37	p = 0.40	p = 0.38	p = 0.38
CategoryShort Film	2.72***	2.66***	2.69***	2.72***	2.73***	2.71***	2.67***	2.67***	2.74***	2.67***	2.67***
	p = 0.002	p = 0.003	p = 0.002	p = 0.002	p = 0.002	p = 0.002	p = 0.003	p = 0.002	p = 0.002	p = 0.003	p = 0.002
CategoryTabletop Games	1.51	1.43	1.43	1.50	1.49	1.52	1.43	1.42	1.48	1.44	1.43
	p = 0.21	p = 0.28	p = 0.29	p = 0.22	p = 0.23	p = 0.21	p = 0.28	p = 0.29	p = 0.23	p = 0.27	p = 0.28
CategoryTechnology	0.73	0.71	0.71	0.73	0.73	0.73	0.71	0.70	0.72	0.71	0.71
	p = 0.38	p = 0.34	p = 0.34	p = 0.38	p = 0.38	p = 0.38	p = 0.34	p = 0.33	p = 0.37	p = 0.34	p = 0.34
CategoryTheater	3.22***	3.11***	3.14***	3.21***	3.22***	3.22***	3.11***	3.12***	3.24***	3.11***	3.12***
	p = 0.0005	p = 0.001	p = 0.001	p = 0.0005	p = 0.0005	p = 0.0005	p = 0.001	p = 0.001	p = 0.0005	p = 0.001	p = 0.001
CategoryVideo Games	0.32***	0.32***	0.32***	0.33***	0.32***	0.32***	0.32***	0.32***	0.32***	0.32***	0.32***
	p = 0.001	p = 0.001	p = 0.001	p = 0.002	p = 0.001	p = 0.001	p = 0.001	p = 0.001	p = 0.001	p = 0.001	p = 0.001
CategoryWebseries	1.08	1.04	1.04	1.08	1.08	1.08	1.04	1.04	1.07	1.06	1.05
	p = 0.82	p = 0.91	p = 0.91	p = 0.83	p = 0.84	p = 0.83	p = 0.91	p = 0.92	p = 0.84	p = 0.87	p = 0.90
CategoryWorld Music	2.64**	2.52**	2.52**	2.58**	2.62**	2.58**	2.50**	2.44**	2.65**	2.61**	2.59**
	p = 0.03	p = 0.04	p = 0.04	p = 0.03	p = 0.03	p = 0.03	p = 0.04	p = 0.05	p = 0.03	p = 0.03	p = 0.03
Constant	0.77	0.86	0.86	0.80	0.79	0.77	0.85	0.86	0.81	0.85	0.85
	p = 0.43	p = 0.64	p = 0.65	p = 0.48	p = 0.48	p = 0.42	p = 0.63	p = 0.65	p = 0.51	p = 0.61	p = 0.62
Observations	8,658	8,658	8,658	8,658	8,658	8,658	8,658	8,658	8,658	8,658	8,658
Log Likelihood	-4,973.97	-4,967.07	-4,968.07	-4,970.41	-4,973.39	-4,972.62	-4,962.90	-4,965.64	-4,969.05	-4,959.03	-4,962.02
Akaike Inf. Crit.	10,067.93	10,062.14	10,058.14	10,070.82	10,068.78	10,067.24	10,051.80	10,057.27	10,064.10	10,064.05	10,056.04

Note: *p<0.1; **p<0.05; ***p<0.01

[page de garde]