THE DYNAMICS OF REWARD-BASED CROWDFUNDING

predicting Kickstarter campaign success

R.J. Klaasse Bos

Colofon

Author

R.J. (Roy) Klaasse Bos (0898280)
r.j.klaasse.bos@student.tue.nl
Eindhoven University of Technology (TU/e)

The author declares that the text and work presented in this thesis is original and that no sources other than those mentioned in the text and its references have been used in creating this thesis.

Bachelor End Project (BEP)

In the partial fulfilment of requirements for the degree of Bachelor of Science in *Industrial Engineering and Management Sciences*.

Supervisors & Assessors

M. (Madis) Talmar, TU/e (ITEM - supervisor / assessor)
J. (Joey) van Angeren, TU/e (ITEM - supervisor)
B. (Bob) Walrave, TU/e (ITEM - assessor)

Subject Headings

Crowdfunding, determinants successful fundraising, internet, machine learning, Kickstarter, prediction

Date of Publication

Eindhoven, June 2017

Abstract

Reward based crowdfunding platforms such as Kickstarter allow founders to fund their project by selling perks to a relative large customer base. Drawing on a scraped dataset of 3652 campaigns with combined funding over $51M, this Bachelor End Project (BEP) discusses several underlying dynamics of successful crowdfunding. Empirical analyses show that the mean estimated delivery time of perks, creator's past crowdfunding success rate, the creator's adoption speed, the number of previous projects by the creator and the number of Facebook friends affect this success rate. Using these and other directly scrapeable attributes from Kickstarter, a binary classification model has been built to predict the success or failure of campaigns. Random Forests perform the best achieving an accuracy of 78.8% at campaign launch (20 percentage points above the baseline model). The practical implications of these findings for both Kickstarter and creators are discussed.

Preface

The first thing that comes to mind when staring at this - currently empty - page of which the heading says “Preface”, is that it has literally been a crazy ride, already from the very beginning of my Bachelor End Project (BEP). Things may have not worked out as planned back in the fall of 2016 when I was looking for an internship position. Though, I can happily ensure you that I learned a lot by tackling this fun but challenging project which is way out of my comfort zone and what I am used to do in my regular bachelor program.

I admit it’s a bit of a cliché though it’s utterly true in this case: “*you can’t connect the dots looking forward; you can only connect them looking backwards*”. For example, this is what I wrote down in a reflection form at the start of my BEP:

> *“Even though I am neither a data engineer nor a data scientist I hope to be able to crunch numbers..”*

Although, I would not still assign myself either of those two titles I definitely crunched a lot of numbers (>75K records) and more importantly, acquired an arsenal of new data skills. From hardly knowing R-programming at all, to building various machine learning models. Besides that, I found an answer to a question I raised at the end of the same reflection form:

> *“All in all, during my internship I would like to figure out what I want to do upon bachelor graduation.”*

After all, after the summer break I will continue my education at the Jheronimus Academy of Data Science. In that sense, this BEP has helped me to better orientate in the field of data science and analytics. It also confirmed my preference to work with quantitative data to solve practical business problems. However, this could not have been realized without the help of some others who I would like to thank for that.

In my case, I was supervised by Madis. In first instance, I was not assigned to him though. So, first of all thank you for arranging things internally with your ITEM-colleagues before the start of my project. During my BEP I really appreciated the positive and enthusiastic vibes of your emails. Also, when I sent you another list of questions, I always received an elaborate point by point response. In addition, I want to thank you for the freedom and time you gave me to figure stuff out on the fly. You also involved another PhD student - Joey - in my BEP which turned out to be a great choice, thank you for that Madis! About that, I should of course not forget to thank you, Joey, for dedicating your time to my project, attending the intermediate and final presentation and providing constructive and concrete feedback despite the fact that you are not my assigned supervisor. The same holds for you Bob; your feedback during the final presentation was very helpful to put the finishing touches to this BEP.

On that note, if you - the reader - also have some feedback, I’d love to hear it. Moreover, if you want to experiment with Kickstarter data yourself, I happily invite you to do so. The pre-processed data can be found on my **Github page** and the data cleaning process has been described over **here**.

Then finally, I wish you a lot of reading pleasure and hope you will enjoy both the quantitative and qualitative part of my BEP!

Roy Klaasse Bos.

Table of Contents

1. Introduction

Crowdfunding is a method of raising small amounts of capital from many individuals to fund the start of a new venture or project. The world's largest online crowdfunding platform is Kickstarter which enables individuals across the globe to virtually pitch their ideas to collectively raise the required amount of capital among visitors of the website to bring the idea to life. Since its establishment in April 2009 - but especially in recent years - many scholars have investigated this originally American marketplace which has just exceeded the $3 billion mark in terms of total dollars pledged. Although this figure may give the impression that most Kickstarter campaigns typically raise a lot of money, it should be noted that since its inception 7 years ago on average only 35.82% of all campaigns have been successfully funded (Kickstarter, 2017). In this bachelor thesis it is attempted to identify still unknown predictors to build a model for explaining crowdfunding success. In turn, this information can provide creators with guidelines on how to optimally prepare, launch and market their campaigns on Kickstarter.

1.1 Introduction

Using data mining processes this paper investigates the relationship between a set of attributes available at launch and campaign success. More specifically, the role of attributes related to a creator's trustworthiness such as the past success rate and the mean estimated delivery has been explored. In addition, the adoption speed on Kickstarter has been identified as an indirect cause for creators' crowdfunding success rate via the number of previous projects and the size of their social network on Facebook.

The required data for these empirical analyses have been web scraped from Kickstarter. This data includes a variety of characteristics of Kickstarter campaigns. It turns out some of these project properties are vital for campaign success. Moreover, the results of different machine learning classifiers support the notion that the probabilities of campaign success can be predicted to a certain extent, even before the campaign has been published on Kickstarter. That is why multiple measures are proposed that help creators set-up more effective campaign pages or increase perceptions of trust among backers.

The rest of this thesis is structured as follows: an introduction to crowdfunding and the relevance of this study in the remaining part of this chapter. Chapter 2 presents an overview of attributes which have been previously investigated and discusses how this study fits in. Chapter 3 explains the reasoning of multiple hypotheses and outlines a model of how corresponding attributes relate to the success rate of Kickstarter projects. Chapter 4 and 5 elaborate on the methodology and the emprical analyses performed. Lastly, the results are summarized and the practical implications are discussed in chapter 6.

1.2 Research Context

For two reasons, it is important to set the research context of this study. First, to introduce the phenomenon of crowdfunding to readers who are not familiar with Kickstarter. After all, a recent study showed that 36% of all American consumers is not familiar with crowdfunding yet (Statista, n.d.). Second, these days there are many crowdfunding platforms which offer varying financing options to creators. It turns out some models are more successful than others. Third, different platforms attract backers from different nationalities which in turn can lead to deviating empirical outcomes. For example, an empirical study on Kickstarter data by Xu et al. (2014) suggests that successful campaigns have more updates than unsuccessful campaigns, whereas Shi and Guan (2015) did not find any statistical significant difference for this attribute using data extracted from a Chinese crowdfunding platform, JingDong.

What follows on the next page is a description of Kickstarter based on its characteristics such as the financing model, crowdfunding type and platform specific features and rules.

All-Or-Nothing

Kickstarter is a two-sided online global marketplace: "creators" (i.e. initiators of a campaign) on the one hand and potential "pledgers" on the other hand. Anyone, anywhere can pledge to a project, though currently only residents of specific countries around the world can create a campaign (Kickstarter, 2017). Creators set a funding target and a fixed time frame for reaching this goal. Campaigns that reach or exceed the target amount are considered a "success". Otherwise, the creator does not receive any of the pledged amount and pledgers receive their money back. This goes hand in hand in with the so-called "All-Or-Nothing" (AON) financing model Kickstarter has adopted. Other financing models, such as "Keep-It-All" (KIA) exist, but are typically even less successful than AON-campaigns, because KIA-campaigns create more uncertainty among potential pledgers (Cumming et al., 2015).

Reward Based Crowdfunding Platform

Although not strictly necessary, pledgers receive a reward (or "perk") from the creator in exchange for their financial support. Perks can both be physical (e.g. a table top game) and non-physical (e.g. an e-book), but are never financial (e.g. equity). Therefore, Kickstarter can be characterised as a reward based rather than an equity based crowdfunding platform.

Product Categories

All campaigns are classified into main and subcategories (e.g. "Design" and "Graphic Design" respectively). Besides, campaigns can be featured by Kickstarter which means they are part of an additional category: "Projects We Love". These featured campaigns are prominently displayed on Kickstarter's platform and therefore it is expected that they reach a larger audience.

Comments & Updates

Only backers can comment on the project page to which the campaign creator can respond. Though, comments are publicly visible to anyone. Similarly, creators can release news updates which will automatically be shared with backers but are also visible to the public. Lastly, personal and background information of the creator can be directly retrieved on the creator's Kickstarter bio page.

1.3 Scope

Regardless of the characteristics of the crowdfunding platform, the challenge of building a model for explaining success can be tackled from different angles. The type of crowdfunding, the platform choice and selected features are factors that determine the project scope. These factors and their corresponding scope for this BEP will be explained below.

There are multiple types of crowdfunding: lending-based, donation-based, reward-based, equity-based, hybrid-based and royalty-based. Especially lending-based crowdfunding makes up a large part (73%) of the total funding volume in the crowdfunding industry (CrowdExpert, n.d.). However, as previously mentioned only reward-based crowdfunding will be considered and thus other types of crowdfunding will be out of scope. Within this subcategory alternative crowdfunding platforms exist, for example Indiegogo showcases many reward-based crowdfunding campaigns. Even though it would be interesting to validate whether the success of Kickstarter and Indiegogo campaigns can be explained using the same features, the empirical analysis of this BEP is only based on data scraped from Kickstarter.

With regards to the taken perspective, the data that serves as input for the model and the type of algorithms used for modelling are related to the achieved outcomes. For example, information directly retrievable from the project page as well as external social features such as Twitter (Etter et al., Chen et al., Lu et al.) and YouTube data (Chen et al.) play a role in the modelling process. Also, the way in which text data is processed can be relevant. That is because prior research has shown the effect of language structure on crowdfunding projects' success. For example, Aleyasen (2014) has been able to achieve 73.7% accuracy on language features. This type of analysis is based on natural language processing (NLP) techniques to identify linguistic features in the project and perk texts that affect the success rate.

All in all, the scope of thesis is limited to non-linguistic attributes which can be directly retrieved or derived from the Kickstarter campaign page and the corresponding creator's bio page. In other words, external media sources and linguistic features will be considered out of scope.

1.4 Relevance

It turns out there is a downside to the fact that crowdfunding platforms are getting more and more traction: relatively less and less campaigns succeed. This section elaborates on that to demonstrate the relevance of this study.

Future Market Growth

Although the total funding volume of reward based crowdfunding may be relatively small to the entire crowdfunding industry, it has been growing rapidly in the last decade. According to Statista (n.d.) the transaction value of crowdfunding is expected to show an annual growth rate of 27.3% resulting in a total amount of $18,967 million in 2021. This is in line with the current trend: in 2014 Kickstarter collected a total fund of $529 million compared to $28 million in 2010 (Statista, 2015). This trend is also noticeable on an industry level; from $0.9 billion in 2010 (Statista, 2013) to $16.2 billion four years later (CrowdExpert, n.d.). Note, Kickstarter's growth rate in this time frame is approximately equal to the average industry growth rate.

Declining Success Rate

Contrary to the crowdfunding market growth, the average success rate for Kickstarter campaigns has decreased significantly in the past couple of years. This can be concluded based on the following three facts which can be found on Kickstarter's Stats webpage.

- The average success rate for Kickstarter campaigns in 2011 was: 43.70% (Statista, 2012).
- The aggregated average success rate for Kickstarter campaigns at April 2013 was: 43.56% (Flaherty, 2013).
- The aggregated average success rate up for Kickstarter campaigns at May 2017 is: 35.82% (Kickstarter, 2017).

Given the fact the last two figures have been calculated considering all campaign data until then (i.e. 2009-2013 and 2009-2017), the average success rate in the last 4 years has been well below 43.56%. Remarkably, there is only a slight decrease noticeable for the average success rate (41.1%) of the collected data sample used in this study (n=3652). Please consult Appendix I for an explanation of the sample representativeness.

In short, the total crowdfunding market (including Kickstarter) is expected to continue growing in the next 5 years. However, part of this growth will be inhibited due to a declining success rate for Kickstarter campaigns. This makes the topic of this thesis even more relevant since current improvements will pay off more and more over time.

2. Literature Review

*As can already be derived from the defined scope of this study (*section 1.3*), academics have tackled the classification problem from multiple angles. Summarizing, research has been conducted focusing on one or more of the following three areas: attributes on the Kickstarter campaign page (1), social features (2) and linguistic features (3). It should be noted that some of these features are time-dependent. For example, the total funding at launch is equal to zero, whereas for successful campaigns this figure has exceeded the goal amount. Of course the more information is available, the more accurate the classification of campaign success will be. Therefore it is important to explicitly state the moment of data-collection. In this thesis we focus on the campaign launch (t=0) as well as the moment of campaign completion (t=deadline). Consequently, the number of available attributes for these two phases differs as some of them can only be determined after the deadline (e.g. the average pledge per backer).*

2.1 Overview of Prior Research

In first instance an overview of relevant prior research is given in this section. The focus is on the selected attributes which served as input features for machine learning models in previous studies. Appendix II presents an overview of the attributes included in those studies. They have been classified into one of the following categories and the main conclusions have been summarized:

- ***Generic:** the main campaign properties and stats which can be found on the top of Kickstarter page.*
  Most noticeable is the height of the goal amount which is negatively related to the campaign success (Chen et al., Hussain et al., Lu et al., Mitra & Gilbert, Mollick, Shi & Guan). On the other hand, the number of campaign updates (Mitra & Gilbert, Mollick) and comments (Mollick, Shi & Guan) is positively associated with the success rate. Note that the former two attributes are only available at campaign completion. Lastly, Hussain et al. (n.d.), Mitra & Gilbert (2014) and Mollick (2014) found that shorter campaigns perform generally better.

- ***Campaign Quality:** attributes that give a rough indication of the time and effort put in by the creator while setting up the project page.*
  Prior research shows the importance of having a promotional video: project pages with a video perform significantly better (Chen et al., Hussain et al., Kamath & Kamat, Mitra & Gilbert, Mollick, Shi & Guan).

- ***Social Media:** the potential online reach of the creator and number of references to the campaign on Facebook, Twitter or YouTube.*
  Multiple social media platforms and corresponding metrics have been considered, but the findings are often insignificant or inconsistent. However, Mollick (2014) argues that creators' number of Facebook friends is significantly related to campaign success: the larger the size of their social network the higher the probabilities of success.

- ***Perks and Pricing:** directly and derived perk and pricing properties of the campaign.*
  A relatively common attribute in crowdfunding research is the number of rewards offered by the creator. More rewards are associated with a higher chance of success (Hussain et al., Mitra & Gilbert). Further, derived attributes such as the goal amount divided by the number of rewards significantly affect campaign success (Hussain et al., n.d.).

- ***Text:** linguistic features of the project description or perk texts.*
  Although this category is out-of-scope in this research, other scholars have been able to make predictions about the campaign success rate based on the project, risk and perk descriptions (Desai et al., Mitra & Gilbert, Sawhney et al.).

- ***Creator:** an indication of how active the creator has been on Kickstarter.*
  Chen et al. (n.d.) and Shi & Guan (2015) found that creators who had backed more projects performed significantly better.

2.2 Literature Gap & Relevance

As can be found in Appendix II the most frequently occurring attributes in the selected papers are related to the campaign configuration. The role of creators' trustworthiness is often not considered. However, this is very relevant because there is always a risk associated with crowdfunding. Loria (2016) advices potential backers to look for evidence of past successes or failure as part of their due diligence. He argues that the number of successful crowdfunding campaigns by a creator indicates whether someone knows how to deal with the upcoming challenges. In that sense, the past success ratio (i.e. the number of successful campaigns divided by the total number of campaigns) is a good measure for that. It signals to potential backers that the creator knows how to successfully raise funds among others and thus has gained the trust of other backers. Although Chen et al. (n.d.), Hussain et al. (n.d.) and Shi & Guan (2015) looked at related attributes such as the number of projects created, the share of successful campaigns compared to the number of failed campaigns has not been investigated yet.

One of the upcoming challenges creators face is delivery delays: more than 75% of all creators deliver products later than expected (Mollick, 2014). This finding goes hand in hand with Indiegogo's initiative to offer perk insurance to backers. In 2015 backers could purchase additional insurance so that they would get a refund if the promised product was not delivered within three months of the stated delivery date (Burns, 2015). Prior research has not investigated the effects of the shipping time on the campaign success, which makes it very relevant to determine whether the estimated delivery date affects backers' purchase decision. Especially given the large variations for the shipping time: from immediately available (e.g. digital download) to a delivery time of more than a year (e.g. hardware products).

Further, most papers assume prior actions to the current campaign do not play any role. However, it can be expected that more experienced creators perform better. After all, Chen et al. (n.d.) found that the number of projects created is positively related to the campaign success. Next to that, the number of projects backed by the creator affects the success rate (Chen et al., Shi & Guan). The "adoption speed" may be an underlying cause for these relationships. That is because, in some way it indicates the amount of experience with Kickstarter. Hence, more experienced creators are expected to have created and backed more campaigns and therefore have an advantage over inexperienced creators. Moreover, early adopters of the platform may display other personal characteristics which are relevant for the campaign success.

2.3 Summary of New Attributes

All in all, the literature gap can be summarized as the lack of a study which investigates the role of creators' trustworthiness and their speed of adoption speed. Therefore, the following new attributes will be considered and further discussed in chapter 3:

- *Estimated delivery*: the expected due date for delivering the perk (i.e. shipping time).
- *Past success rate:* this ratio is derived from the number of successful and unsuccessful campaigns by the creator excluding the (live) campaigns part of the study sample.
- *Adoption speed*: the number of days since the registration on Kickstarter until the start date of their latest Kickstarter campaign.

On a final note, a large half of papers on this topic focuses on constructing the most accurate classifier for crowdfunding success. The findings of this research can contribute to that, especially given the fact the new attributes mentioned before are all available at campaign launch. Even more, the empirical analysis also includes a section related to classifying Kickstarter campaign success.

3. Research Question & Hypotheses

The previous chapter pointed out that prior research - into the relationship between attributes directly traceable from the Kickstarter campaign or creator's bio page and the success rate - has been limited to a subset of all available attributes. This study aims to predict the success rate of Kickstarter campaigns on the basis of a more comprehensive set of attributes. The hypotheses corresponding to these unexamined attributes - as mentioned in section 2.3 *- will be discussed in this chapter.*

3.1 Estimated Delivery

Thereby we mean the number of days from the end of the campaign to the expected perk delivery date. Contrary to most ecommerce stores shipping times of rewards can take months or sometimes even more than a year, rather than days. That is because, the creator often needs the financial injection to create or finalize the offering. Note that, the estimated delivery can vary from perk to perk. In that case, the mean estimated delivery time across all perks is determined.

Several eCommerce fulfilment solutions have investigated the importance of fast delivery. In a whitepaper Dotcom (2016) concludes that fast delivery has a significant impact on loyalty: 87% of online shoppers indicate the amount of time it took to receive their order influences his or her decision to shop with a retailer again. Particularly, order intransparency and lack of insight into the status of a package throughout the entire delivery process are detrimental to the consumers' decision to repurchase. These facts can play a role on a returning backer's loyalty to back another project by the same Kickstarter creator.

Also, for first-time crowdfunding creators (76.1% of sample) the estimated delivery is expected to affect a potential pledger's purchase decision. According to the same Dotcom study 63% of all shoppers finds it somewhat important, important or very important to receive their order in the shortest amount of time possible. Moreover, of all examined factors fast delivery has the greatest positive influence on a customer's perception of trust with a brand. The importance of shipping speed is confirmed by a study by comScore (2012). It shows that shipping speed is one of the top-5 factors taken into consideration when shoppers are comparison shopping. In short, the delivery time has a significant effect on the trust in the seller and is considered important by many consumers. In turn, it is expected longer delivery times have a negative impact on the total number of consumer purchases.

Given these expected effects of the estimated delivery time for both new and experienced Kickstarter creators, *it is hypothesised that the mean estimated delivery time of all perks is negatively related to the probabilities of success of Kickstarter campaigns (H1).*

3.2 Past Success Rate

This ratio is determined by dividing the total number of successful projects by the total number of projects created by the creator. Note, that Kickstarter campaigns which are part of our data sample have been excluded in this calculation. In that sense, it is really the success rate prior to starting the considered Kickstarter campaign.

In essence, the past success rate is a combination of two attributes which have been part of prior research: the number of previous projects that had been successful (Husain et al., n.d.) and the total number of projects created by the creator (Chen et al., n.d.; Shi & Guan, 2015).

Husain et al. expected that as users create more projects, they seem to get an idea of what works and thus get more successful. They combined this attribute with the staff picked feature in order to create a composite variable which represents the perceived reliability of the project. For that reason, it is not possible to distinguish the impact of both attributes separately.

On the other hand, Chen et al. and Shi & Guan focused on separate effects of the number of projects created. The underlying idea here is that prior crowdfunding experience is likely related to the creator's trustworthiness which is especially important given the fact Kickstarter has no way of ensuring the creators use their funds in the way advertised. Remarkably, the results of both studies are inconsistent. Chen et al. found that the number of projects created is one of the most important non-temporal features, while Shi & Guan could not confirm this finding. On that note, it should be doubted whether the study by Shi & Guan based on data from a Chinese crowdfunding platform is representative for Kickstarter. After all, they also conclude that backing behaviours (i.e. the number of projects backed by the creator) have a negative effect on raising money, while Chen et al. conclude the opposite. Next to that, Xu et al. (2014), Mollick (2014) and Mitra & Gilbert (2014) argue that the number of project updates is positively related to the probabilities of success, while again Shi & Guan did not find support for this relationship. Because of that, it is assumed their research is indeed not fully representative in this case and will therefore be left out in the comparison. As a result, it is expected that the number of projects created is positively associated with campaign success.

Even though, logical mathematical reasoning is not possible due to a lack of complete information of the significant relationships in the numerator of the past success ratio, *it is hypothesized that the creator's past success rate is positively related to the campaign success (H2)*. That is because, like the total number of projects created a high success ratio is expected to build up trust among potential pledgers and subsequently lead to more funding.

3.3 Adoption

This attribute indicates how long ago a creator registered on Kickstarter and thus started using the crowdfunding platform. Note that this registration date can be long before a creator started his or her first project.

Based on the "Diffusion of Innovations" theory by Rogers (1995) five categories of adopters can be identified: innovators, early adopters, early majority, late majority and laggards. These categories make a distinction based on the speed at which consumers adopt a product or service (i.e. adoption speed). This framework can also be applied to the adoption speed of Kickstarter. For example, individuals who joined the platform right after its launch in 2009 can be considered as innovators.

The characteristics of individuals within these categories differ which in turn may affect their crowdfunding success. Early adopters tend to be most influential and critical, have a high degree of thought leadership, are very active on social media and have a high social status (Understanding Early Adopters and Customer Adoption Patterns, 2016; Schiffman et al., 2012). It is expected that these traits indirectly positively contribute to a creator's success. After all, prior research shows that the number of Facebook friends (Mollick, 2014) and Twitter followers (Lu et al., 2014) is positively related to the campaign success. Moreover, thought leaders are very effective at reaching a large audience because of the trust and credibility they have built up (Schiffman et al., 2012). On the other hand, the late majority typically has lower social status, less interaction with thought leaders and rarely offer any form of thought leadership in a field (Understanding Early Adopters and Customer Adoption Patterns, 2016). That is why it is expected that individuals who joined Kickstarter later (i.e. late majority, laggards) generally do worse than those who joined earlier (i.e. innovators, early adopters, early majority). It might even be an explanation for the declining success rate during the past years (see section 1.3).

Apart from these individuals' traits, the timing - or more precise the experience - is related to the adoption speed. Innovators joined the platform early, so they have probably created and backed more projects than the late majority and laggards. The experience that comes with that, both from own and indirect learning is expected to help creators be more successful with future projects. Chen et al. (n.d.) demonstrated this relationship: the number of projects created and backed has a positive significant relationship with the overall success rate of Kickstarter campaigns.

In short, *it is hypothesized that Kickstarter creators' adoption speed of the platform is positively associated with the number of Facebook friends and the number of projects created and backed which in turn is expected to improve the chances of campaign success (H3)*. In other words, the number of Facebook friends and the number of projects created and backed are mediators in the relationship between the adoption speed and success rate.

3.4 Overview of Hypotheses

The aforementioned hypotheses are visually represented on the next page (Figure 3.4.1).

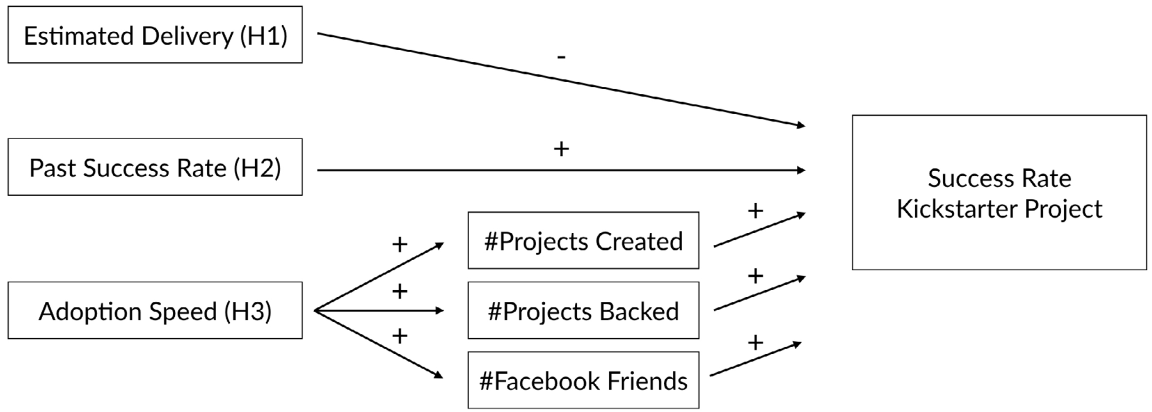


Figure 3.4.1 - Overview of all hypothesized relationships

4. Methodology

Now that the research question and hypotheses have been defined, the next step is to collect the required data and preprocess it so that it can be used for further analyses. Besides, descriptive statistics are presented to get a feeling for the data.

4.1 Data Collection

At the 30th of March 2017 data from 3726 live Kickstarter projects was collected. The start date of projects within this sample ranges from the 29th of January 2017 to the 28th of March 2017 due to diverse project durations (from 4 up to 60 days). Though, all projects would end before the 28th of April 2017. After this deadline, the selected URLs were scraped again, including the creator's bio page.

As a side note, only the first 4000 live Kickstarter campaigns could be web scraped due to technical complications (85% of total number of live projects). Since live projects are ordered by success the average success rates across categories is slightly higher. However, as demonstrated in Appendix I, the frequency distribution of categories and the amount of dollars pledged for successful projects show many similarities with Kickstarter's overall averages. Note that, 274 projects of the initial sample have been lost due to fine-tuning the scraper prior to the first successful attempt.

As mentioned before, the moment of data extraction determines the availability of certain project attributes. Appendix III provides an overview of all attributes part of this study, their definitions and whether they are available at the launch or not. In principle, definitions assume data extraction after campaign completion. However, when only considering attributes available at launch in some cases this assumption might be an oversimplification. For example, creators can introduce new perks throughout the campaign period or update their project page with new information and images. The consequences of this will be discussed in chapter 6.

4.2 Data Preprocessing

Although scraping data from Kickstarter is straightforward thanks to the repeating structure of each webpage, converting raw data to preprocessed data comes with multiple challenges. This conversion step has been outlined step-by-step over **here**. The end result is a dataset containing 3652 records without missing data.

4.3 Descriptive Statistics

As depicted in Table 4.3.1 the dataset contains 1503 successful projects which corresponds to an average success rate of 41.1%. Moreover, all projects combined raised more than $51M. The large majority of that total amount comes from successful campaigns (95%). It follows that the average pledge per backer for successful campaigns is more than 35% higher than failed campaigns. On the contrary, if the project goal is divided by the average perk value, it turns out the average number of required backers to reach the goal is 29.6% lower for successful campaigns.

	Successful	Failed	Total
Campaigns	1503	2149	3652
Proportion	41.2%	58.8%	100%
Backers	469,150	33,462	502,612
Pledged ($)	48,874,303	2,570,287	51,444,590

Table 4.3.1 - Descriptive statistics data sample

The crowdfunding platform of our choice, Kickstarter, is especially popular in the United States. More than 62% of our data sample originates from USA-campaigns. In particular, residents from California and New York start many projects (Figure 4.3.2). Excluding Africa and South America because of their small sample sizes, the average success rate in each continent is close to the 40% mark (Table 4.3.3). Although success rates vary from continent to continent, there are no significant differences (95% CI).

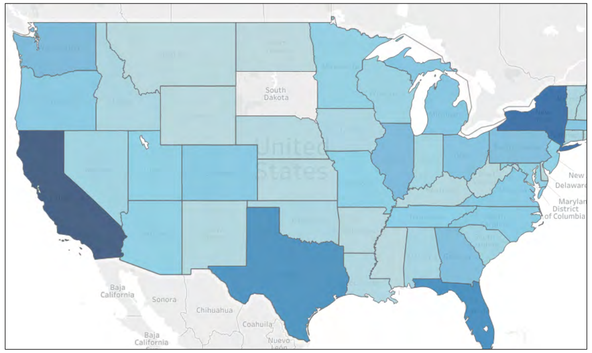


Figure 4.3.2 - Number of Kickstarter campaigns in individual states in the USA. The darker the blue, the higher the frequency.

	#Campaigns	Success rate
Africa	6	16.7%
Asia	96	46.9%
Europe	944	40.6%
North America	2486	41.6%
Oceania	107	35.5%
South America	16	18.8%
Total	3655	41.1%

Table 4.3.3 - Number of campaigns and success rate for all continents

The relative proportion of the 15 main categories are displayed in the pie chart below. The success rates within each category can vary significantly (Appendix IV). For example, 73% of all comics projects succeed versus only 26% of all technology projects. Underlying attributes such as the average height of the crowdfunding goal play a part in this.

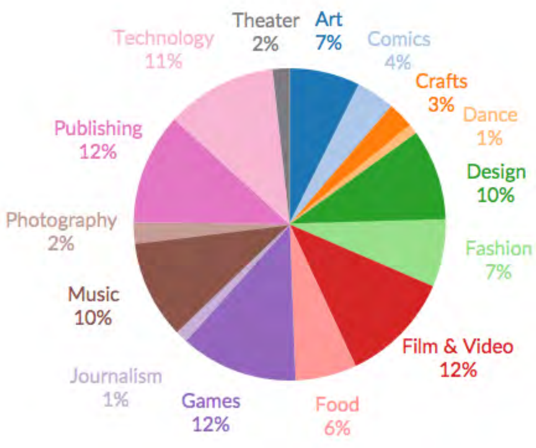


Figure 4.3.4 - Pie chart of main categories Kickstarter (sample)

In addition to the category, other features have been taken into account. Descriptive statistics of all numeric features can be found in Table 4.3.5 as well as a correlation matrix in Appendix V. Note, that outliers have not been removed yet. As a result, the mean and standard deviation of some attributes in the table below (e.g. "goal") may be skewed to the right.

A few observations regarding the descriptive statistics in the table below:

- The low mean number of collaborators can be explained by the fact that only 2% of all creators filled out this optional field on their campaign page.
- About 12% of all perks is immediately available for backers after pledging. This corresponds with an estimated delivery of zero days.
- Kickstarter was launched in April 2009 which explains why the maximum value for "Adoption" attribute can be as high as 7.8 years (2834 days).

Table 4.3.5 - Descriptive statistics Kickstarter dataset

Attribute	Mean	Median	Mode	Std. Dev.	Range
Goal ($)	24307	5378	5000	189,920	1 - 9,000,000
#Backers	138	15	1	704	0 - 26828
#Updates	2.9	1	0	4.7	0 - 50
#Comments	19.3	0	0	177	0 - 8099
Project Duration (days)	33.5	30	30	10.8	4 - 60
#Collaborators	0.037	0	0	0.30	0 - 8
#Images	9.3	4	0	13.9	0 - 117
#Facebook Friends	755	575	575	787	1 - 5000
Average money / reward	5347	947	1000	39,548	0.09 - 1,800,000
First level pledge	39.14	6	5	388.52	0.76 - 10,000
Maximum pledge tier	1322.51	300	1000	2427.10	0.76 - 10,531
Number of rewards	7.5	6	1	5.8	1 - 74
Estimated delivery (days)	95	61	0	115	0 - 1866
#Words in project description	709	512	281	623	1 - 4925
#Projects created by creator	1.7	1	1	2.9	1 - 78
#Projects backed by creator	5.6	0	0	26	0 - 890
Past success rate	0.50	0.50	0.50	0.23	0 - 1
Adoption (days)	479	78	0.07	680	0.02 - 2834
#Direct competitors	74	52	273	76.2	0 - 273

5. Empirical Analysis

This chapter starts off with the results for the hypotheses. Based on these outcomes significant features are selected as the input for machine-learning models. Then, the success of Kickstarter campaigns model is predicted considering data available in two scenarios: the launch and the deadline. The performance of various machine learning models will be compared to the baseline model.

5.1 Results Hypotheses

This section addresses the empirical outcomes for the three hypotheses formulated in chapter 3. In first instance, successful and failed projects will be compared using only descriptive statistics. After that, a logistic regression model is used to validate whether the relationship between the two variables also holds when control variables are included. These control variables are selected based on findings in prior research and Appendix V.

5.1.1 Estimated Delivery

More than 30% of all projects expect to deliver rewards within 30 days after completion. Though, as can be seen in Figure 5.1.1.1 a shorter average shipping time does not necessarily increase the probabilities of success. However, there seems to be a threshold between 150 and 200 days: campaigns of which the average estimated delivery period is later than half a year since the campaign completion (in sample: n=546 or 15%) have an average success rate of only 30%.

Yet, it is important to recall the definition of the estimated delivery which is the *average* number of days between the project its deadline and the estimated delivery for all perks. Accordingly, the estimated delivery between perks can vary. In our data sample 47% of all projects had at least two unique estimated shipping dates. Potentially, an early shipping date can balance out a late one. Therefore, we also need to consider the effects of the separate estimated delivery times.

The significance of the relationship between the average/seperate estimate delivery time and the project success has been tested with multiple logistic regression models (Appendix VI). Based on model 2 and 3 it can be concluded that only the average estimated delivery has a significant negative effect on the success rate. This implies that it does not necessarily harm to offer perks with long delivery times as long as they are complemented with perks with relative shorter delivery times. For example, projects with an average estimated delivery time lower than the median (71 days) which also include one or more perks with an estimated delivery time of more than half a year, have an average success rate of 54%.

5.1.2 Past Success Rate

As the name of this attribute inherently suggests,

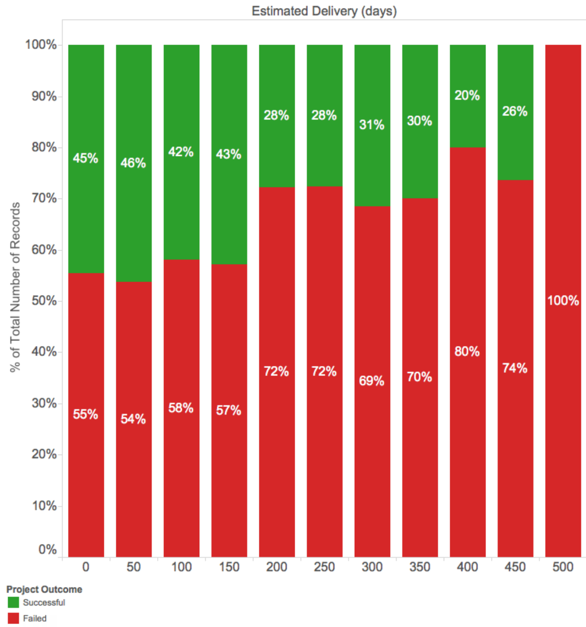


Figure 5.1.1.1 - Success rates vs Average estimated delivery (in days)

the value can only be calculated for creators who started at least one other project prior to their current endeavour. In our data sample that is the case for about 24% of all records. For this subset, the large majority (57%) has created one other project (Figure 5.1.2.1). It turns out the success rate of their first project significantly affects the success rate of their next project: if their first project succeeded (i.e. past success rate = 100%) the average success rate for the second project is 83% vs 33% if the first project failed (i.e. past success rate = 0%). In other words, given their first project was a success, creators who created 2 projects in total are more than 2.5 times as likely to

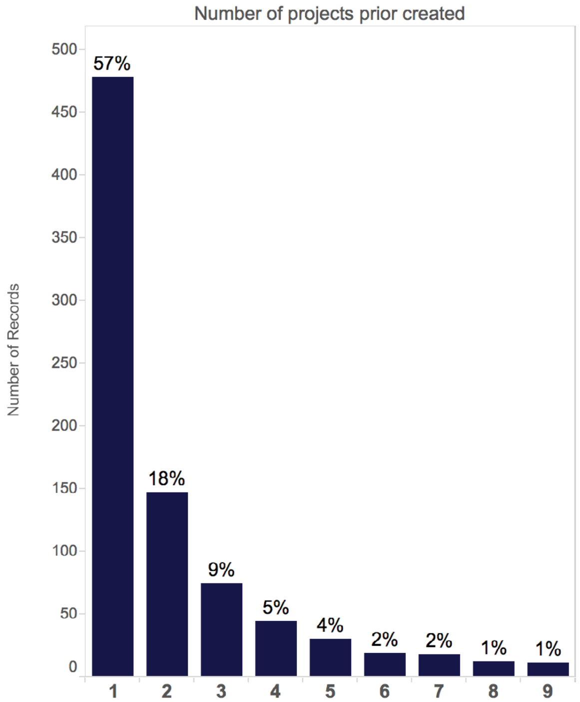


Figure 5.1.2.1 - Histogram number of projects prior created (range: 1-9)

rearch their crowdfunding goal. On the contrary, creators whose first attempt was unsuccessful are almost 4 times as likely to fail their second project as well.

Regardless of the number of prior projects created, it seems like there is a positive relationship between the past success rate and the current campaign success (Figure 5.1.2.2). For example, the success rate for the current campaign is about three times higher for creators with an average past success rate greater than 70% compared to a past success rate lower than 10%. Note, the 10-20%, 20-30% and 40-50% bins contain relatively few data points and therefore the corresponding average success rates are less reliable.

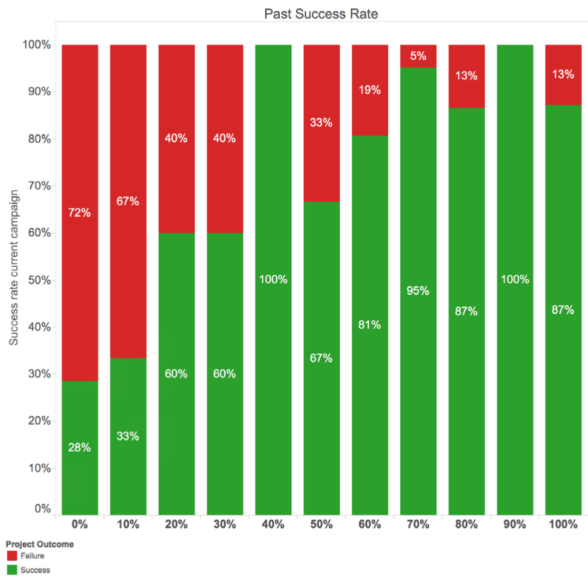


Figure 5.1.2.2 - Past success rate vs success current campaign

As was described in section 3.2 Chen et al. found that the number of projects created is one of the most important features to consider. Likewise, our data sample suggests an upward success rate as the number of projects prior created increases (Figure 5.1.2.3). In general, this trend remains for creators who have initiated more than 9 campaigns in the past (Appendix XIII). This makes it very interesting check the past success rate has still any predictive power in case the number of projects created is considered as a control variable. Put differently, does it matter how successful a creator has been in the past or just the number of projects created?

To answer this question, several logistic regression models have been run (Appendix VI). The results of model 3 show that both the number of projects created and the past success rate positively contribute

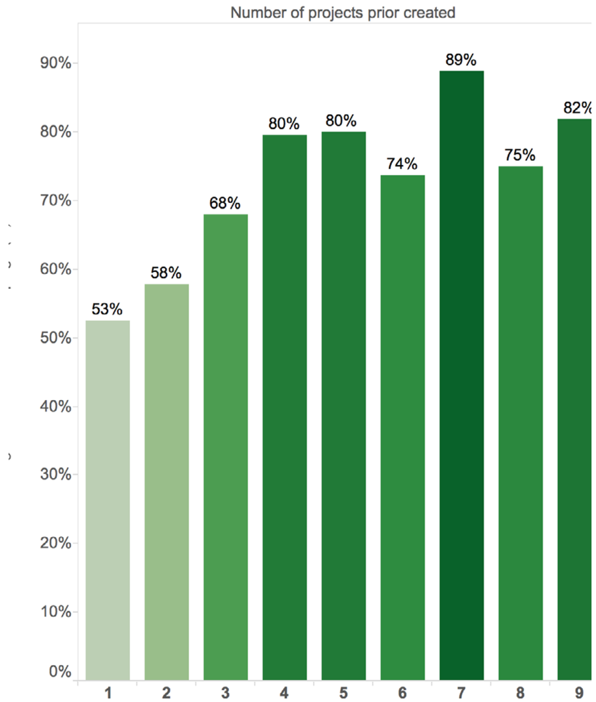


Figure 5.1.2.3 - Number of projects prior created vs success current campaign

to the campaign success. This implies that especially successful projects prior created are significant for future crowdfunding success.

5.1.3 Adoption

The date a Kickstarter user joined the crowdfunding site can be found on the creator's bio page. To determine users' adoption speed the number of days between the start date of their latest campaign and the date on which they joined has been calculated.

To classify users into one of the five categories from Rogers' diffusion curve, the aggregated number of users which constitutes the diffusion curve has been estimated at 377K (between 2009 - 2022). The underlying derivations and assumptions as well as an overview of the years that correspond with each of the adoption categories can be found in Appendix VII.

Creators who currently (2017) register at Kickstarter can be characterised as the late majority. From our hypothesis, it follows that these individuals are expected to indirectly have a lower chance of success than early adopters and innovators who joined the platform respectively 5-6 and 7-8 years ago. Figure 5.2.3.1 is in line with that: creators who joined Kickstarter more than five years ago are on average more than twice as successful than creators who joined in the last 365 days.

As was expected from our hypothesized model, there is a positive relationship between the adoption speed and the number of projects created and backed by the creator as well as the number of Facebook friends (Appendix VI). The latter matches typical characteristics of innovators who are known to be very active on social media. The data shows this category of adopters has more often connected their Facebook account to Kickstarter and has remarkably more Facebook friends. What attribute we did not initially expect but turns out to be significant is the number of comments by the creator. The number of comments is positively associated with the number of projects created and backed. The latter follows from the fact that only backers can comment. In turn, the number of projects backed is positively related to the number of campaigns created. The same holds for the previously introduced attribute "Past success rate". Intuitively this makes sense:

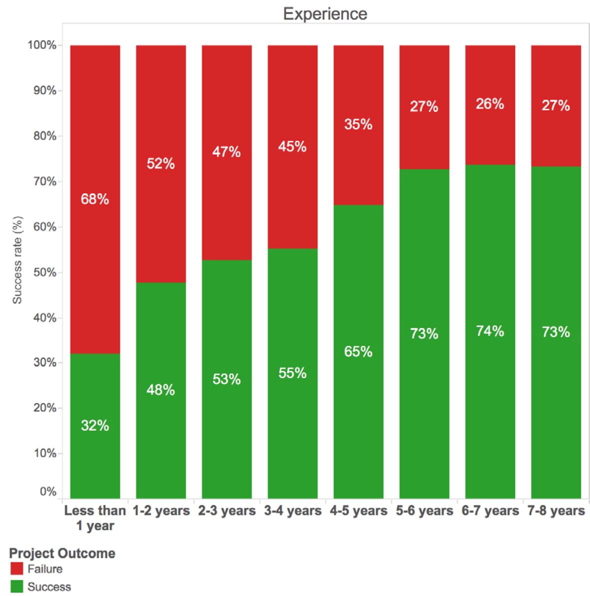


Figure 5.2.3.1 - Number of years experience vs success current campaign

creators may be encouraged by previous successes to start new endeavours.

Turning to the consequences for the campaign success, we find a couple of unexpected relationships. First, the number of projects backed by the creator shows no significant relationship with the success rate. This is contrary to the findings of Chen et al. (n.d.). Second, the adoption speed also directly affects the success rate positively. The same holds for the number of Facebook friends and the number of projects created by a creator. This is in line with the empirical findings of Mollick (2014) and Chen et al. (n.d.) who respectively claim that large networks are associated with successful fundraising and that the number of projects created by the creator is likely related to the creator's trustworthiness.

5.2 Predicting Campaign Success

5.2.1 Input Models

Given aforementioned results, the following attributes are expected to contribute to a more accurate classifier: the estimated delivery (H1), past success rate (H2), adoption speed (H3), number of Facebook friends (H3) and number of projects created (H3).

In addition to these attributes and the ones found in literature (Appendix II), the average number of backers required has a significant effect on the chances of success, even more than the campaign goal. Projects in our data sample which require (on average) 600 backers or more to reach their goal have a success rate of 0% (Figure 5.2.1.1). Given the minimum amount of funding needed to realize their project, creators should consider this information when deciding on their perk pricing. The other way around, campaigns which offer relatively cheap perks should opt for a lower goal.

5.2.2 Baseline

The next section presents several performance measures for a variety of machine learning models. These figures are largely dependent on the distribution of successful and unsuccessful campaigns in the sample. That is why a baseline model for comparison should be introduced.

The baseline model is the priori probability of unsuccessful Kickstarter projects which is 58.8%. After all, if was assumed that all projects were going to fail, this would yield an accuracy of 58.8%. Furthermore, the recall is 58.8% and the precision is 100%, though the latter is not meaningful for baseline models.

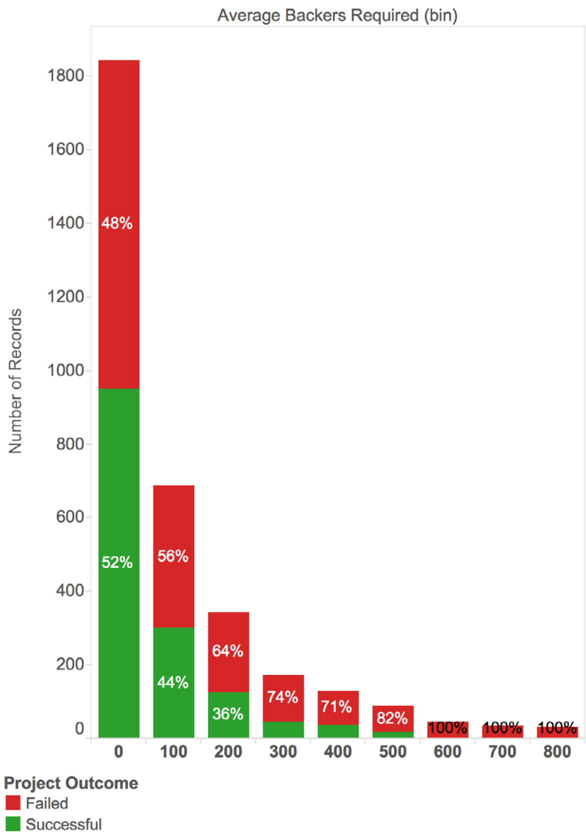


Figure 5.2.1.1 - Odds of success vs average number of backers required

5.2.3 Model Results

In first instance, feature selection was applied to determine significant predictors (Appendix VIII). It should be noted that this selection yields different results dependent on the moment of data extraction: at launch or campaign completion.

Launch

Given the subset of features which are immediately available once the campaign is published, all hypothesized relationships (estimated delivery (H1), past success rate (H2), #projects backed (H3), #projects created (H3), #Facebook friends (H3)) hold true. Moreover, a cheaper maximum pledge tier, more rewards, more collaborators, a project description that counts more words, the presence of a promotional video, more additional videos, more images and a shorter campaign duration are likely to increase the probabilities of success. Remarkably, the height of the goal has not any significant negative relationship with the success rate. This can be explained by the inclusion of a derivative of the goal: the average backers required. In other words, in this case a composite variable created by combining both the height of the goal and perk price is more important than merely the goal.

As can also be found in Appendix VIII and Appendix IX the currency is a significant success predictor. The currency indirectly refers the geography of the project. However, when currency and the continent a creator lives in (e.g. Europe) are part of one model, only the currency has a significant effect on the success rate (Appendix VIII). Further research shows that not the continent, but the creator's country (or state) can explain this relationship. This corresponds to Mollick's (2014) conclusion about the importance of the creator's geographical location in America (i.e. the state) for successful fundraising. In turn, this can be explained by the nature of a population: a proportionally greater creative population is associated with a greater chance of success for founders.

Campaign Completion

As soon as the deadline has been exceeded the following additional attributes are available: total number of comments (by backers), total number of updates and the total number of backers. Especially the effects of the first two attributes are interesting to consider. Like Mitra & Galbert and Mollick, we find that more updates are associated with a higher success rate. Successful campaigns have published on average 5.48 updates vs 1.04 updates for unsuccessful campaigns. With regards to the number of comments no statistical significant relationship has been found, contrary to Mollick's (2014) findings. This can be explained by the attributes which were included in his study: the total number of backers was not part of it. Since only backers can comment on the campaign page, it follows that the total number of backers is a logical mediator between the number of comments and the success rate. Creating the same model without "total number of backers" confirms this statement (Appendix VIII).

Model performance

To maximize model performance, the optimal cut-off level to achieve the highest accuracy has been determined (Appendix X). Tables 5.2.3.1 and 5.2.3.2 outline the average performance on the test set (20%) for the model at launch and campaign completion (Appendix III). Besides that, the corresponding ROC-curves and AUC-values can be found in Appendix XI and Table 5.2.3.3 respectively.

Overall the success of a crowdfunding project can be predicted with an upper bound accuracy of 78.8% at launch. This is an improvement of roughly 20 percentage points over the baseline. Similar to Greenberg et al. (2013) we find that random forest and logistic model trees perform the best. Other

Data available at launch

Model	Accuracy	Recall	Precision
Logistic Regression	76.5 %	75.2 %	71.9 %
Support Vector Machine	75.7 %	61.6 %	77.0 %
K-Nearest Neighbors	72.6 %	63.4 %	69.8 %
Decision Tree	73.4 %	72.5 %	67.7 %
Random Forest	78.9 %	74.5 %	76.1 %
XGBoost	77.4 %	74.5 %	73.4 %

Table 5.2.3.1 - Validation on test set (20%) using 10-fold cross validation (t=launch)

Data available at campaign completion

Model	Accuracy	Recall	Precision
Logistic Regression	92.7 %	90.2 %	92.6 %
Support Vector Machine	87.4 %	78.5 %	91.0 %
K-Nearest Neighbors	76.0 %	67.4 %	74.2 %
Decision Tree	91.3 %	90.3 %	89.6 %
Random Forest	92.7 %	92.8 %	90.5 %
XGBoost	93.6 %	90.4 %	95.3 %

Table 5.2.3.2 - Validation on test set (20%) using 10-fold cross validation (t=campaign completion)

AUC-values

Data input (attributes)	AUC
Campaign launch	0.861
Campaign completion	0.975

Table 5.2.3.3 - Area Under the Curve (AUC) values

than that, XGBoost - an implementation of gradient boosted decision trees - yields a relative high accuracy, especially at campaign completion. Furthermore, Greenberg et al. also looked at the model performance when the number of backers is the only attribute. In our case that leads to an accuracy around 83% which is somewhat higher than what they found (77%). Likewise, when considering all attributes combined Greenberg et al. reached an upper bound around 90% accuracy, whereas our XGBoost model performed slightly better (Table 5.2.3.2).

From a practical point of view, it is more meaningful to predict the campaign success at the launch than after the deadline. That is because just before publishing creators could be recommended to adjust the campaign set-up. For example, to opt for a shorter duration or a lower goal. These settings cannot be changed anymore once the campaign has gone live. This makes it interesting to look how well our model at launch performs compared to other studies. Greenberg et al. (2013) and Chen et al. (n.d.) hit an upper bound of 67%. This is significantly lower than the results in Table 5.2.3.1. On the other hand, Hussain et al. (n.d.) achieved an accuracy between 72-80% dependent on the classification algorithm. In the same way, we derived attributes such as the past success rate and the estimated delivery, Hussain et al. also included more complex features whereas the two aforementioned studies did not. This shows the importance of feature engineering and more generally: having a comprehensive set of attributes to work with.

6. Conclusion

The sixth and last chapter summarises the empirical results related to the hypothesized model. Further, topics for future research are proposed and the limitations and managerial implications for both crowdfunding platforms and creators are discussed.

6.1 Summary Results

The main aim of this study is to validate the hypotheses 1 to 3. The results have been summarized below:

1. The average estimated delivery time of all perks is negatively related to the campaign success. However, this relationship does not hold for the separate delivery time of perks. In other words, an early shipping date can balance out a late one and thus mitigate adverse effects.

2. The creator's past success rate is positively related to the probabilities of success for future Kickstarter campaigns. Though, regardless of the outcome of previous campaigns the experience of creating a crowdfunding campaign contributes positively to future campaign success. In addition, Appendix VIII confirms our hypothesis: a higher past success ratio affects future success.

3. The hypothesis related to the creator's adoption speed consists of multiple relationships. From Appendix VI follows that the adoption speed is positively related to the number of Facebook friends and the number of projects created and backed. Correspondingly, the number of projects created and the number of Facebook friends significantly affect the success rate. The positive relationship between the number of campaigns backed by the creator and the campaign success has been rejected due to the addition of more significant control variables.

6.2 Limitations

There are a couple of limitations of this research. First, the average success rate of campaigns part of the dataset is not fully representative of the overall average success rate on Kickstarter. Second, as can be found in Table 4.3.5 the range for some attributes is very wide. Potentially this can be explained by the presence of outliers which have not been removed from the dataset. Third, the same data have been used for the analysis at campaign launch and completion. In fact, creators can update certain elements of their campaign page throughout the live-period (e.g. number of perks). In this respect, the actual values at launch can deviate from what we assumed in this research. Fourth, the "Projects We Love" campaign property has been ignored because the success rate in our sample (2.1%) is unrepresentative. Fifth, the currency rates have been assumed as a constant. Sixth, the categorisation of adopters (Appendix VII) has been based on discussable assumptions. In particular, there is only one data point that confirms a downward trend for the number of new Kickstarter creators. Consequently, the estimates for 2017 until 2022 are uncertain. Seventh, it is doubtful what share of potential pledgers is aware of creators' past success rate, since it is not directly displayed on the campaign page. Eighth, there are alternative ways to stack the different models and structure the hypotheses which can lead to different outcomes. For example, the role of the number of projects backed by a creator was only significant without a comprehensive set of controls (Appendix VI). Ninth, the number of Facebook friends has been determined at campaign completion, while publishing the Kickstarter project may lead to an increase in the number of Facebook friends, especially for successful creators. In that sense, the presence of a bidirectional relationship has not been investigated. Lastly, it should be noted that our empirical analysis solely relies on scraped data, whereas crowdfunding success is not directly related to scrapeable attributes. For example, subjective concepts such as the quality and contents of the video pitch affect the success rate very likely.

6.3 Future Work

The basis of the conducted empirical research is a series of directly scrapeable attributes of Kickstarter campaigns. Although this approach leads to satisfactory results, the performance of the classifier can probably be improved by taking into account more abstract attributes. Desai et al. (2013), Greenberg et al. (2013), Mitra & Gilbert (2014), Sawhney et al. (n.d.) have already investigated the influence of language in the project and perk descriptions. Though, the effectiveness of the video pitch has not been studied yet. Greenberg et al. suggest hiring Amazon Mechanical Turk workers to evaluate such abstract aspects. Given the recent technological advancements in data science a more scalable solution to analyse the effectiveness of videos would be the use of deep learning.

With reference to the hypotheses of this study, perk characteristics such as the delivery time turn out to be a predictor for success. In the same way, the shipping zones (worldwide or regional), shipping costs and quantity limits may affect the number of pledgers for a specific perk. Moreover, the number of added perks after launch may be a predictor for the total number of recurring pledgers, since Indiegogo (2012) argues that 20% of repeat contributions are for perks that were added after the campaign went live. On that note, it can be interesting to check whether creators with a high past success ratio have relative more returning backers due to the gained trust among past backers. Similarly, the sentiment of previous backers' tweets related to prior campaigns may influence the creators' trustworthiness for future campaigns.

Lastly, the role of equity based crowdfunding is worth mentioning. Especially given the fact Indiegogo has introduced this option for financing only since last year. The corresponding equity based crowdfunding pages have a similar structure which includes a promotional video, images a textual description and a discussion forum. Therefore, attributes which are relevant to Kickstarter campaign pages may also be applicable to Indiegogo's equity based campaigns.

6.4 Managerial implications

Given the aforementioned results, Kickstarter can recommend creators to carefully consider the average estimated delivery time. Although Kickstarter limits the options for the funding duration and provides advice, this is not the case for the estimated delivery (Appendix XII). Following the same reasoning, Kickstarter can display the past success rate more prominently on the campaign page. Even more, Stegmaier (2016) proposes a Kickstarter rating system which showcases the creator's rating at the top of the campaign page. This rating is determined based on the input of backers who have been asked to rate the creator's performance on a scale from 1-5. A poll among Kickstarter community members (n=98) shows that 85.7% of all voters want Kickstarter to implement such a basic rating system for backers. In other words, this is definitely an user interface iteration Kickstarter should look at.

For long-term and sustainable profitability, Kickstarter should investigate how they can increase the success rates for the late majority and laggards creators. After all, the platform only makes money once campaigns reach or exceed their goal amount. By means of a "Creator Handbook" (i.e. help guide) they already educate potential creators on how to set-up a campaign. One topic of interest that is very important according to this study, but is not elaborately mentioned in the help guide is the importance of the average number of backers required to reach the goal. A very direct approach could be to show this figure once the creator has filled out the goal amount and the perk prices. This could be accompanied by an informational text which says that campaigns which require less backers are generally more successful (similar to Appendix XII.I).

From a creator perspective, we find that expensive perks are typically delivered later. A plausible explanation for this relationship can be that those perks are more complex and therefore have a longer production and lead time. Creators who offer such perks should consider offering one or more alternatives with a shorter delivery time. That way, they can target both potential backers who are willing to wait a short and long time. As a rule of thumb: the average estimated delivery time should always be below 180 days.

As a last note, creators should realize the size of their social network (e.g. number of Facebook friends) can indirectly affect their success rate on Kickstarter. Though of course this comes with the assumption that creators know how to optimally leverage the power of social media to effectively promote their campaign.

References

Aleyasen, A. (2014). KickUpper : A Tool For Making Better Crowdfunding Projects. Retrieved from http://sifaka.cs.uiuc.edu/~wang296/Course/IR_Fall/docs/Projects/Samples/38.pdf

Bidaux, T. (2017). Kickstarter in 2016 - Year in review - ICO Partners. Retrieved from http://icopartners.com/2017/01/kickstarter-in-2016-year-in-review/

Building rewards — Kickstarter. (n.d.). Retrieved from https://www.kickstarter.com/help/handbook/rewards

Burns, M. (2015). Indiegogo Testing Way To Refund Money If Crowdfunded Project Does Not Ship | TechCrunch. Retrieved from https://techcrunch.com/2015/02/18/indiegogo-testing-way-to-refund-money-if-crowdfunded-project-does-not-ship/

Chen, K., Jones, B., Kim, I., & Schlamp, B. (n.d.). KickPredict : Predicting Kickstarter Success. Retrieved from http://courses.cms.caltech.edu/cs145/2013/blue.pdf

CrowdExpert. (n.d.). Total crowdfunding volume worldwide from 2012 to 2015 (in billion U.S. dollars). In Statista - The Statistics Portal. Retrieved May 13, 2017, from https://www.statista.com/statistics/620952/total-crowdfunding-volume-worldwide/.

CrowdExpert. (n.d.). Total funding volume in crowdfunding industry in 2015, by source (in billion U.S. dollars). In Statista - The Statistics Portal. Retrieved May 13, 2017, from https://www.statista.com/statistics/620850/funding-volume-in-crowdfunding-instustry-by-source/.

Cumming, D. J., Leboeuf, G., & Schwienbacher, A. (2015). Crowdfunding Models: Keep-It-All vs. All-Or-Nothing *, 1–41.

Desai, N., Gupta, R., & Truong, K. (2013). Plead or Pitch ? The Role of Language in Kickstarter Project Success. Retrieved from https://nlp.stanford.edu/courses/cs224n/2015/reports/15.pdf

Dotcom. (2016). Driving Customer Loyalty With Fast Delivery and Quality Packaging, 1–18.

Etter, V., Grossglauser, M., & Thiran, P. (2013). Launch Hard or Go Home ! Predicting the Success of Kickstarter Campaigns. Retrieved from https://infoscience.epfl.ch/record/189675/files/etter2013cosn.pdf

Greenberg, M. D., Hariharan, K., Gerber, E., & Pardo, B. (2013). Crowdfunding Support Tools: Predicting Success & Failure. CHI 2013, Changing Perspectives, Paris, France, 1815–1820. https://doi.org/10.1145/2468356.2468682

Flaherty, J. (2013). Fine-Tune Your Kickstarter Campaign With These 12 Tools | WIRED. Retrieved from https://www.wired.com/2013/04/12-kickstarter-tools/

How to Set Your Kickickstarter & Indiegogo Rewards and Perks. (n.d.). Retrieved from https://www.shopify.com/guides/crowdfunding/optimizing-crowdfunding-rewards-perks

Hussain, N., Kamel, K., & Radhakrishna, A. (n.d.). Predicting the success of Kickstarter campaigns. Retrieved from https://cseweb.ucsd.edu/classes/wi17/cse258-a/reports/a108.pdf

IndieGoGo field guide [Brochure]. (2015). IndieGoGo field guide [Brochure]. Retrieved from https://learn.indiegogo.com/wp-content/uploads/2016/03/IGG_ER_CampaignerFieldGuide_012115.pdf

Indiegogo Insight: 70% of Campaigns Meeting Funding Goal Have 3-8 Perks - Indiegogo Blog. (n.d.). Retrieved from https://go.indiegogo.com/blog/2011/11/indiegogo-insight-build-5-8-perks.html

Kamath, R.S., Kamat, R.K. (2016). Supervised Learning Model for Kickstarter

Campaigns With R Mining. International Journal of Information Technology, 4(1), 17–28. https://doi.org/10.5121/ijitmc.2016.4102

Kickstarter. (2017). Kickstarter Stats. Retrieved from: https://www.kickstarter.com/help/stats?ref=footer

Kickstarter. (2017). Start your project - Kickstarter. Retrieved from: https://www.kickstarter.com/learn?ref=nav

Kleinman, S. (2012). Online Shopping Customer Experience Study. comScore.

Loria, K. (2016). 7 things to know before you spend a dime on crowdfunding. Retrieved from http://www.businessinsider.com/investing-in-crowdfunding-on-kickstarter-and-indiegogo-2016-5

Lu, C., Xie, S., Kong, X., & Yu, P. S. (n.d.). Inferring the Impacts of Social Media on Crowdfunding Categories and Subject Descriptors. Retrieved from https://www.cs.uic.edu/~xkong/wsdm14_lu.pdf

Mitra, T., & Gilbert, E. (2014). The Language that Gets People to Give : Phrases that Predict Success on Kickstarter, 49–61. Retrieved from http://delivery.acm.org/10.1145/2540000/2531656/p49-mitra.pdf?ip=131.155.212.154&id=2531656&acc=ACTIVE SERVICE&key=0C390721DC3021FF.

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

62% of Campaigns That Reach Their Goal Have Repeat Funders - Indiegogo Blog. (2012). Retrieved from https://go.indiegogo.com/blog/2012/02/62-of-campaigns-that-reach-their-goal-have-repeat-funders.html

Rogers, E.M. (1995). Diffusion of innovations. Macmillian Publishing Co. https://doi.org/citeulike-article-id:126680

Sawhney, K., Tuason, R., Tran, C. (n.d.). Using Language to Predict Kickstarter Success.

Shi, M., & Guan, L. (2015). An Empirical Study of Crowdfunding Campaigns : Evidence From Jing Dong Crowdfunding Platform.

Schiffman, L. G., Lazar, K. L., & Hansen, H. (2012). Consumer behaviour: A European outlook. Harlow: Pearson.

Stam, M. (2016). University of Amsterdam Master Thesis Crowdfunding Success Prediction : From Classification to Survival Regression and back supervised by.

Start your project — Kickstarter. (2017). Retrieved from https://www.kickstarter.com/learn?ref=nav

Statista Survey. (n.d.). Consumer awareness of crowdfunding in the United States in 2016. In Statista - The Statistics Portal. Retrieved May 13, 2017, from https://www.statista.com/statistics/638553/familiarity-with-crowdfunding-usa/.

Statista (2012). Kickstarter - The King of Crowdfunding. Retrieved May 13, 2017, from https://www.statista.com/chart/279/kickstarter---the-king-of-crowdfunding/.

Statista (2013). Kickstarter - Funds Through Crowdfunding in 2012. Retrieved May 13, 2017, from https://www.statista.com/chart/1034/funds-raised-through-crowdfunding-in-2012/.

Statista (2015). Amount of dollars pledged on Kickstarter. Retrieved May 13, 2017, from http://www.statista.com/statistics/249688/amount-of-dollars-pledged-on-kickstarter/.

Statista (n.d.). Crowdfunding - worldwide. Retrieved May 13, 2017, from https://www.statista.com/outlook/335/100/crowdfunding/worldwide#takeaway .

Stegmaier, J. (2016). Why a Kickstarter Reputation System Could Reduce Prejudice – Stonemaier Games. Retrieved from https://stonemaiergames.com/why-a-kickstarter-reputation-system-could-reduce-prejudice/

Understanding Early Adopters and Customer Adoption Patterns | Interaction Design Foundation. (2016). Retrieved from https://www.interaction-design.org/literature/article/understanding-early-adopters-and-customer-adoption-patterns

Xu, A., Yang, X., Rao, H., Fu, W., Huang, S., & Bailey, B. P. (2014). Show Me the Money! An Analysis of Project Updates during Crowdfunding Campaigns. Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems, 591–600. https://doi.org/10.1145/2556288.2557045

Appendices

Appendix I - Sample Representativeness

Category	Share of launched projects (2009-2017)	Share of launched projects (sample)	Average successful dollars pledged per project (2009-2017)	Average successful dollars pledged per project (sample)	Success Rate (2009-2017)	Success Rate (sample)
All	*100%*	*100%*	*21450*	*32518*	*35.8%*	*41.1%*
Games	8.9%	12.2%	54390	37851	34.7%	43.8%
Design	7.6%	9.5%	62258	76485	34.5%	45.5%
Technology	8.3%	11.5%	91996	147848	19.8%	26.3%
Film & Video	17.5%	11.6%	13740	10995	37.2%	41.4%
Music	14.6%	10.2%	6978	8469	49.7%	51.2%
Fashion	5.7%	7.0%	21058	47335	24.1%	31.1%
Publishing	10.5%	11.5%	8996	9018	30.4%	40.5%
Food	6.5%	6.4%	17083	16461	25.0%	30.2%
Art	7.4%	7.4%	6789	7428	40.6%	45.0%
Comics	2.7%	3.9%	11755	9149	53.0%	72.9%
Theater	2.9%	1.8%	5924	4322	60.2%	70.3%
Photography	2.9%	2.2%	9651	25779	30.5%	36.7%
Dance	1.0%	1.0%	5159	4510	62.5%	51.4%
Crafts	2.3%	2.8%	5403	2166	23.8%	28.7%
Journalism	1.3%	1.2%	10146	25803	21.7%	17.8%

A few remarks regarding the representativeness of the data:

- The 2009-2017 columns are based on 357,097 total projects, while our dataset only contains 3652 projects. As a result our outcomes are more prone to outliers. Moreover, the sample is based on projects from 2017 only. The share of launched projects, average successful dollars and the success rate might have changed throughout the years.
- The higher success rate for our sample can be explained by the data scraping process. A maximum of 4000 live campaigns could be scraped. Since successful campaigns are always shown first, this leads to an artificially high success rate. Apart from the performance of the baseline model this should not have too much effect on the overall performance of the machine learning models and the hypotheses tests.

Appendix II - Attributes Included In Studies

Table columns

The table indices (1-11) in the table columns on the coming pages refer to the following authors, years and papers.

Table index	Author	Year	Paper title
1	Chen et al.	n.d.	KickPredict: Predicting Kickstarter Success
2	Desai et al.	n.d.	Plead or Pitch? The Role of Language in Kickstarter Project Success
3	Greenberg et al.	2013	Crowdfunding Support Tools
4	Hussain et al.	n.d.	Predicting the success of Kickstarter campaigns
5	Kamath and Kamat	2016	Supervised Learning Model for Kickstarter Campaigns with R Mining
6	Lu et al.	2014	Inferring the Impacts of Social Media on Crowdfunding
7	Mitra and Gilbert	2014	The Language that Gets People to Give: Phrases that Predict Success on Kickstarter
8	Mollick	2014	The dynamics of crowdfunding: An exploratory study
9	Sawhney et al.	n.d.	Using Language to Predict Kickstarter Success
10	Shi and Guan	2015	An Empirical Study of Crowdfunding Campaigns: Evidence From Jing Dong Crowd nding Platform
11	Stam	2016	Crowdfunding Success Prediction: From Classification to Survival Regression and back

Figure II.I - Legenda table indices

Cell markers

+/-	Significant correlation with success (positive or negative)
0	No significant correlation with success
?	Unknown correlation with success

Figure II.II - Legenda cell markers

Note, authors regularly do not explicitly state the predictive power of attributes separately. This also explains why many cells contain a "?" marker.

Full overview of attributes included in studies*

Category	Attribute	Definition	1	2	3	4	5	6	7	8	9	10	11
Generic	Goal	The minimum self-determined amount of money the creator must collect in a fixed timeframe for the campaign to be successful.	+/-	?	?	+/-	?	+/-	+/-	+/-		+/-	?
Generic	#Backers	The total number of unique backers						+/-		+/-	?	+/-	+/-
Generic	#Pledges	This is the sum of total perk purchases. Note, #pledges can differ from #backers as backers could pledge multiple times.											+/-
Generic	Location creator	The country (for non-US) or state (for US) the creator lives in				0	?			+/-			
Generic	#Updates	The number of updates by the campaign creator						?	+/-	+/-		0	
Generic	#Comments	The sum of the number of comments by backers and responses from the campaign creator							0	+/-		+/-	
Generic	Launch date	The start date of the campaign (i.e. launch date)					?						
Generic	Duration	A fixed timeframe (in days) set by campaign creator at the start that determines how long a campaign is “live”.	0	?	?	+/-	?	?	+/-	+/-			
Generic	Category	The main category of a campaign (e.g. “Design”).			?	+/- **	?		+/- ***				
Generic	Sub category	The sub category of a campaign (subset of a main category, e.g. “Graphic Design”)				+/- **	?		+/- ***				
Generic	Staff-picked	A boolean that indicates whether a campaign is featured by Kickstarter (a.k.a. “Project We Love”)				+/-			0		?		
Campaign quality	Presence of video	If the project page has a video	+/-		?	+/-	+/-		+/-	+/-		+/-	
Campaign quality	Duration of video (if present)	The length of the promotional video (in seconds).								+/-			
Campaign quality	#Images	The number of images present in the project description	0									0	
Campaign quality	#Spelling mistakes	The number of most-common spelling mistakes in the project text								+/-			
Social media	#Facebook friends	The number of Facebook friends of the creator			?					+/-			
Social media	Connected on Facebook	If the Kickstarter account is connected to Facebook	0		?				+/-				
Social media	Connected on Twitter	If the Kickstarter account is connected to Twitter			?								

Category	Attribute	Definition	1	2	3	4	5	6	7	8	9	10	11
Social media	#Twitter followers of the creator	The number of Twitter followers of the creator			?			+/-					0
Social media	#Twitter users the creator follows	The number of users the creator follows on Twitter						?					
Social media	#Tweets with campaign URL	The number of times the project URL (or shortened URL) has been tweeted	0					?					0
Social media	#Tweets referring to a Kickstarter campaign	The number of indirect references to a Kickstarter campaign on Twitter (e.g. “I just backed project_title” @ kickstarter)						?					
Social media	#Retweeted tweets with campaign URL	The number of times a tweet containing the project URL (or shortened URL) has been retweeted											0
Social media	#Favourited tweets with campaign URL	The number of times a tweet containing the project URL (or shortened URL) has been favourited											0
Social media	Presence of YouTube video	Whether or not the project page has an additional YouTube video	0										
Social media	#Views of YouTube video	If a Youtube video is present, the view count of the video over time	0										
Perks and pricing	Average pledge / backer	The total pledge divided by the total number of backers								+/-			
Perks and pricing	Average money per reward level	The goal of the project was divided by the number of reward levels to get an average money per reward level.				+/-							
Perks and pricing	First/Minimum level pledge	The cheapest perk available at the project page	0						0			+/-	
Perks and pricing	Second level pledge	The second cheapest perk available at the project page										+/-	
Perks and pricing	Maximum pledge tier	The most expensive perk available at the project page	0										
Perks and pricing	#Reward levels	The number of pledge tiers	0	?	?	+/-	?	?	+/-				
Text	#Sentences in project description	The total number of sentences the project description contains.		?	?								
Text	Sentiment of project description	An estimation of whether the sentiment of the project description is positive, negative or neutral.			?								
Text	Distance to mean blurb length	The square of the difference between a project’s blurb (i.e. lead of campaign text) length and the mean (18 words).				?							

Category	Attribute	Definition	1	2	3	4	5	6	7	8	9	10	11
Text	URL in blurb	Including a URL that links to relevant information in the blurb				?							
Text	Project description	Linguistic features of the project description		+/-					+/-		+/-		
Text	Risk description	Linguistic features of the risk description		+/-					+/-		+/-		
Text	Perk descriptions	Linguistic features of the perk descriptions		+/-					+/-				
Creator	#Successful projects by creator	The number of previous projects that had been successful by the creator.				?							
Creator	#Projects created by creator	The total number of projects created by the creator (both unsuccessful and successful)	+/-									0	
Creator	#Projects backed by creator	The total number of other projects backed by the creator	+/-									+/-	

Figure II.III - Full overview of attributes considered in other studies

**Based on attributes which are explicitly mentioned by the author(s) of the paper. In a few occasions, this summation will be incomplete because of that. This especially holds for papers focused on natural language processing, since a full overview of control variables is often missing.*

***Provided category and subcategory are both included.*

****The following categories and sub-categories: Graphic Design, Theatre, Food, Games, Documentary, Art, Board & Card Games, Webseries, Fashion, Periodical and Animation.*

Top-5 attributes

The table below displays the five most frequently occurring attributes in aforementioned eleven studies.

Position	Attribute*	Frequency of occurence
1	Goal	9x
2	Duration	7x
3	Number of pledge tiers	7x
4	Presence of video	6x
5	Number of backers	5x

Figure II.IV - Top 5 most frequently considered attributes

Appendix III - Research Design

Category	Attribute	Definition	Available at launch
Generic	Goal	The minimum self-determined amount of money the creator must collect in a fixed timeframe for the campaign to be successful.	Yes
Generic	#Backers	The total number of unique backers.	No
Generic	Location creator	The country (for non-US) or state (for US) the creator lives in.	Yes
Generic	#Updates	The number of updates by the campaign creator.	No
Generic	#Comments	The sum of the number of comments by backers and responses from the campaign creator.	No
Generic	Duration	A fixed timeframe (in days) set by campaign creator at the start that determines how long a campaign is "live".	Yes
Generic	Category	The main category of a campaign (e.g. "Design").	Yes
Generic	Sub category	The sub category of a campaign (subset of a main category, e.g. "Graphic Design").	Yes
Generic	Staff-picked	A boolean that indicates whether a campaign is featured by Kickstarter (a.k.a. "Project We Love").	No
Generic	#Collaborators	The number of additional creators explicitly mentioned on the campaign page.	Yes
Campaign quality	Presence of video	If the project page has a video.	Yes
Campaign quality	#Images	The number of images present in the project description.	Yes
Social media	#Facebook friends	The number of Facebook friends of the creator.	Yes
Perks and Pricing	Average money / reward	The goal of the project is divided by the number of reward levels to get an average money per reward level.	Yes
Perks and Pricing	Average number of backers required	The goal of the project is divided by the mean perk value to get the average number of backers to reach the goal.	Yes
Perks and Pricing	First level pledge	The cheapest perk available at the project page.	Yes
Perks and Pricing	Maximum pledge tier	The most expensive perk available at the project page.	Yes
Perks and Pricing	Number of rewards	The number of pledge tiers.	Yes
Perks and Pricing	Estimated delivery	The average expected due date for delivering the perk.	Yes

Category	Attribute	Definition	Available at launch
Text	#Words in project description	The total number of words the project description contains.	Yes
Creator	#Projects created by creator	The total number of projects created by the creator (both unsuccessful and successful).	Yes
Creator	#Projects backed by creator	The total number of other projects backed by the creator.	Yes
Creator	Past success rate	The number of earlier successful campaigns divided by the total number of earlier campaigns (i.e. excluding the campaigns part of the study sample).	Yes
Creator	Adoption	The number of days since the registration on Kickstarter and their current campaign.	Yes

Figure III.I - Attributes part of research design

Note, unless stated differently the definitions assume data extraction after campaign completion.

Appendix IV - Success Rates Across Categories

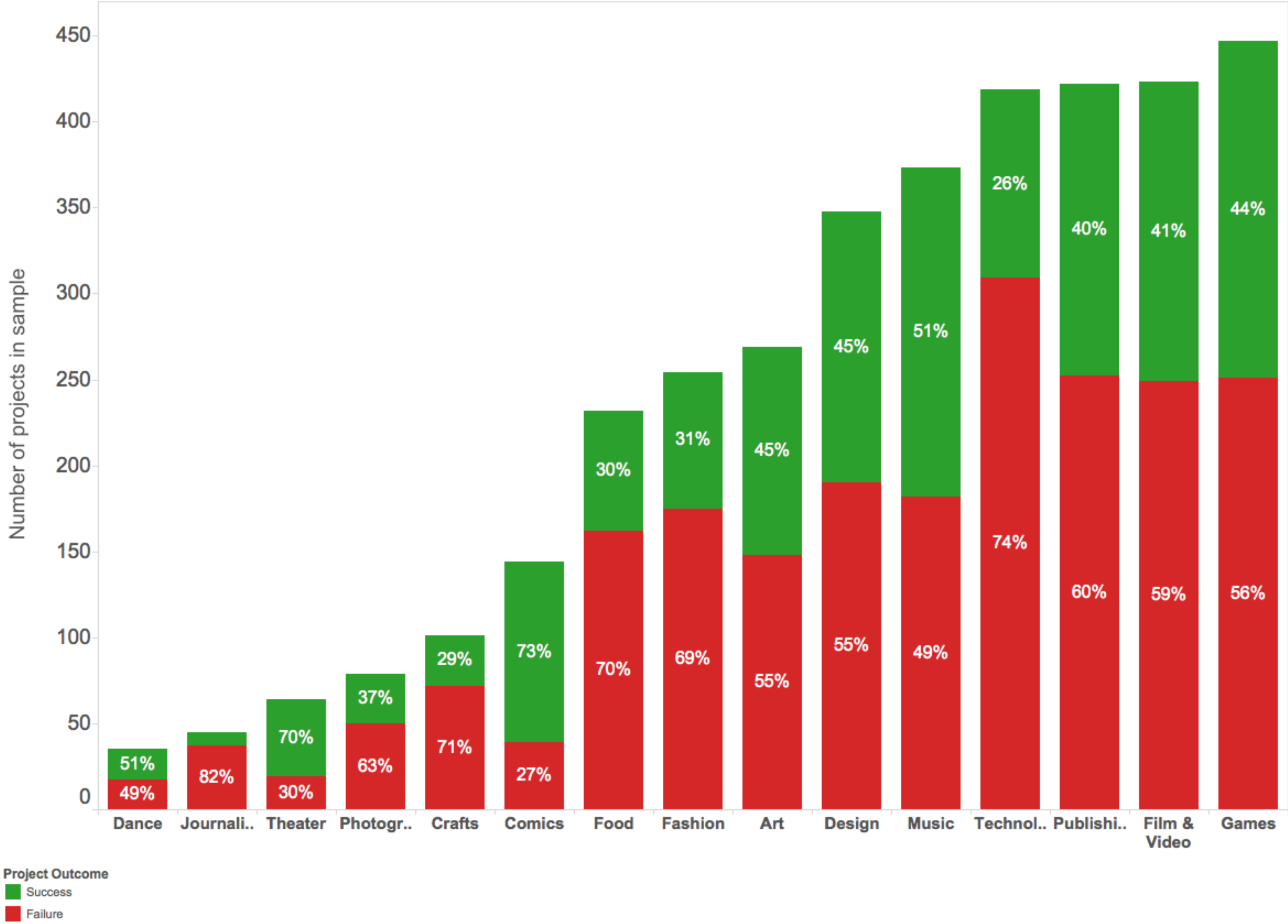


Figure VI.I - Success rate across categories

Appendix V - Correlation Matrix

Attribute	Success	Goal ($)	#Backers	#Updates	# Com-ments	Project Duration	Collabo-rators	#Images	#Face-book Friends	Average money / reward	First level pledge	Maximum pledge tier	Number of rewards	Estimated Delivery	#Words in de-scription	#Projects created by creator	#Projects backed by creator	Past success rate	Adoption (days)	#Direct competi-tors
Success																				
Goal ($)	- 0.06*																			
#Backers	0.21*	0.01																		
#Updates	0.47*	- 0.01	0.35*																	
#Comments	0.12*	0.01	0.86*	0.30*																
Project Duration	- 0.15*	0.10*	0.00	- 0.06*	- 0.01															
Collaborators	0.09*	0.01	0.00	0.10*	0.02	0.00														
#Images	0.29*	0.00	0.04*	0.47*	0.25*	- 0.03	0.13*													
#Facebook Friends	0.07*	- 0.03	0.02	0.00	0.01	0.01	0.00	- 0.02												
Average money / reward	- 0.09*	0.93*	0.02	- 0.05*	0.00	0.10*	- 0.01	- 0.04*	- 0.02											
First level pledge	- 0.05*	0.03	-0.01	- 0.04*	- 0.01	0.06*	- 0.01	- 0.03*	0.02	0.15*										
Maximum pledge tier	0.02	0.09*	0.10*	0.11*	0.04*	0.05*	0.11*	0.13*	0.05*	0.04*	0.14*									
Number of rewards	0.30*	0.03	0.14*	0.39*	0.05*	- 0.01	0.11*	0.40*	0.04*	- 0.07*	- 0.08*	0.38*								
Estimated delivery	- 0.09*	0.17*	0.05*	0.08*	0.05*	0.13*	0.00	0.06*	0.00	0.16*	0.07*	0.15*	0.04*							
#Words in project description	0.24*	0.02	0.18*	0.43*	0.14*	- 0.04*	0.15*	0.57*	- 0.01	- 0.02	- 0.02	0.23*	0.38*	0.12*						
#Projects created by creator	0.18*	-0.02	0.08*	0.13*	0.08*	- 0.06*	-0.01	0.11*	0.07*	-0.02	- 0.01	- 0.05	0.05*	0.00	0.09*					
#Projects backed by creator	0.17*	-0.01	0.13*	0.23*	0.12*	-0.07*	0.01	0.15*	0.05*	- 0.02	- 0.02	- 0.02	0.09*	0.04*	0.16*	0.33*				
Past success rate	0.27*	-0.01	0.11*	0.16*	0.07*	- 0.07*	0.02	0.11*	0.08*	- 0.02	- 0.01	0.01	0.15*	0.00	0.12*	0.14*	0.18*			
Adoption (days)	0.28*	0.00	0.09*	0.29*	0.06*	- 0.13*	0.02	0.13*	0.12*	- 0.01	- 0.02	0.05	0.20*	0.01	0.17*	0.32*	0.34*	0.24*		
#Direct competitors	0.05*	-0.02	0.10*	0.21*	0.09*	- 0.04*	0.00	0.29*	- 0.05*	- 0.03	- 0.01	- 0.06	0.04*	0.04*	0.17*	0.07*	0.10*	0.04*	0.09*	

Figure V.I - Correlation matrix attributes part of research design

*p-value < 0.05

Appendix VI - Hypotheses Validation

Hypothesis 1: Estimated Delivery

For model 2 and 3 duplicate estimated delivery times within one project have been removed. “Individual” refers to the unique estimated delivery of the perks corresponding to a project, whereas “mean” refers to the average estimated delivery date of all perks.

Model 1 - (Only including “Estimated Delivery (mean)”)

Attribute	Estimate	Std. Error	P-value
Estimated Delivery (mean)	-2.23e-03	5.12e-04	1.31e-05***
Goal	-6.41e-05	4.74e-06	<2e-16***
Duration	-2.30e-02	4.29e-03	8.14e-08***
#Updates	2.85e-01	1.85e-02	<2e-16***
#Comments	6.20e-02	6.88e-03	<2e-16***
#Images	-2.20e-03	4.45e-03	0.62
#Perks	9.94e-02	1.05e-02	<2e-16***

Dependent variable: ***project success***

Model 2 - (Only including “Estimated Delivery (individual)”)

Attribute	Estimate	Std. Error	P-value
Estimated Delivery (individual)	-1.21e-04	1.54e-01	0.65
Goal	-5.00e-05	3.17e-06	<2e-16***
Duration	1.89e-02	4.08e-03	3.65e-06***
#Updates	2.03e-01	1.35e-02	<2e-16***
#Comments	7.90e-02	7.09e-03	<2e-16***
#Images	2.09e-03	3.65e-03	0.57
#Perks	8.24e-02	7.81e-03	<2e-16***

Dependent variable: ***project success***

Model 3 - (Including both “Estimated Delivery (mean)” and “Estimated Delivery (individual)”)

Attribute	Estimate	Std. Error	P-value
Estimated Delivery (mean)	-1.03e-03	5.06e-04	0.04*
Estimated Delivery (individual)	3.87e-04	3.65e-04	0.29
Goal	-4.95e-05	3.18e-06	<2e-16***
Duration	-1.86e-02	4.09e-03	5.72e-06***
#Updates	2.04e-01	1.35e-02	<2e-16***
#Comments	7.94e-02	7.14e-03	<2e-16***
#Images	2.40e-03	3.65e-03	0.51
#Perks	8.235e-02	7.839e-03	<2e-16***

Dependent variable: ***project success***

Hypothesis 2: Past Success Rate

Model 1 - (Only including “Past Success Rate”)

Attribute	Estimate	Std. Error	P-value
Past Success Rate	2.49e+00	2.30e-01	<2e-16***
Goal	-6.69e-05	4.84e-06	<2e-16***
Duration	-2.20e-02	4.40e-03	5.77e-07***
#Updates	2.82e-01	1.89e-02	<2e-16***
#Comments	5.84e-02	6.86e-03	<2e-16***
#Images	-8.04e-04	4.51e-03	0.86
#Perks	9.14e-02	1.06e-02	<2e-16***
#Projects Backed	1.76e-03	2.49e-03	0.48

Dependent variable: ***project success***

Model 2 - (Only including “#Projects Created”)

Attribute	Estimate	Std. Error	P-value
#Projects Created	1.45e-01	3.24e-02	7.39e-06***
Goal	-6.30e-05	4.66e-06	<2e-16***
Duration	-2.31e-02	4.31e-03	9.05e-08***
#Updates	2.76e-01	1.85e-02	<2e-16***
#Comments	5.73e-02	6.57e-03	<2e-16***
#Images	-3.60e-03	4.48e-03	0.42
#Perks	9.99e-02	1.05e-02	<2e-16***
#Projects Backed	7.04e-04	2.72e-03	0.80

Dependent variable: ***project success***

Model 3 - (Including both "Past Success Rate" and "#Projects Created")

Attribute	Estimate	Std. Error	P-value
Past Success Rate	2.41e+00	2.30e-01	<2e-16***
#Projects Created	1.31e-01	3.46e-02	1.55e-04***
Goal	-6.46e-05	4.82e-06	<2e-16***
Duration	-2.15e-02	4.43e-03	1.24e-06***
#Updates	2.82e-01	1.88e-02	<2e-16***
#Comments	5.51e-02	6.68e-03	<2e-16***
#Images	1.65e-03	4.55e-03	0.72
#Perks	9.21e-02	1.06e-02	<2e-16***
#Projects Backed	-1.48e-03	2.25e-03	0.51

Dependent variable: ***project success***

Note, the variance inflation factor shows there is no multicollinearity between the past success rate and the number of previous projects created:

Attribute	Tolerance	VIF
Adoption Speed	0.800	1.250
Estimated Delivery	0.998	1.002
#Projects Backed	0.822	1.216
#Facebook Friends	0.982	1.018
Past Success Rate	0.927	1.079
#Projects Created	0.841	1.189

Dependent variable: ***project success***

Hypothesis 3: Adoption Speed

Adoption speed vs #Projects Backed

Attribute	Estimate	Std. Error	P-value
Adoption speed	0.19	0.020	<2e-16***
#Projects Created	0.21	0.018	<2e-16***
#Comments by creator	0.10	0.019	3.29e-07***
Facebook connected	-0.037	0.036	0.293
#Facebook friends	-0.0040	0.018	0.819
Past Success Rate	0.10	0.018	1.36e-08***

Dependent variable: ***#Projects Backed***

Adoption speed vs #Projects Created

Attribute	Estimate	Std. Error	P-value
Adoption speed	0.16	0.020	5.52e-16***
#Projects Backed	0.21	0.019	<2e-16***
#Comments by creator	9.69e-02	1.90e-02	<2e-16***
Facebook connected	0.06	0.036	0.096
#Facebook friends	4.82e-02	1.74e-02	0.0023**
Past Success Rate	0.043	0.018	0.017*

Dependent variable: ***#Projects Created***

Adoption speed vs #Facebook Friends

Attribute	Estimate	Std. Error	P-value
Adoption speed	1.19e-01	2.17e-02	5.27e-08***
#Projects Created	5.68e-02	2.06e-02	0.0057**
#Projects Backed	-1.90e-04	2.06e-02	0.99
#Comments by creator	-6.76e-02	2.08e-02	0.0011**
Past Success Rate	4.43e-02	1.95e-02	0.023*

Dependent variable: ***#Facebook Friends***

Model 1 - (Without comprehensive set of controls)

Attribute	Estimate	Std. Error	P-value
Adoption speed	3.62e-04	6.70e-05	6.65e-08***
#Projects Created	6.25e-02	2.95e-02	0.034*
#Projects Backed	7.88e-03	3.62e-03	0.030*
#Comments by creator	1.08e-02	8.44e-04	<2e-16***
Facebook connected	5.76e-02	7.80e-02	0.46
#Facebook friends	1.32e-04	4.88e-05	0.0067**
Past Success Rate	2.45	2.00e-01	<2e-16***

Dependent variable: ***project success***

Model 2 - (Including control variables)

Attribute	Estimate	Std. Error	P-value
Adoption speed	1.84e-04	7.89e-05	0.019*
#Projects Created	1.11e-01	3.65e-02	0.002**
#Projects Backed	-1.67e-03	2.29e-03	0.46
#Comments by creator	-6.89e-04	1.12e-03	0.54
Facebook connected	2.37e-01	9.42e-02	0.46
#Facebook friends	2.20e-04	5.78e-05	0.00015***
Past Success Rate	2.27	2.33e-01	<2e-16***
Goal	-6.33e-05	5.01e-06	<2e-16***
Duration	-1.98e-02	4.49e-03	1.12e-05***
#Updates	2.87e-01	1.95e-02	<2e-16***
#Comments	5.57e-02	7.12e-03	5.23e-15***
#Images	-1.77e-04	4.62e-03	0.97
#Perks	8.87e-02	1.07e-02	<2e-16***
Estimated Delivery (mean)	-2.41e-03	5.33	5.97e-06***

Dependent variable: ***project success***

Appendix VII - Diffusion Curve

In 2016 a study showed that 36% of all American consumers was not familiar with crowdfunding (Statista, n.d.). At that moment of time 241,873 unique creators had joined Kickstarter since 2009. Therefore, the total market size is estimated at 241,873 / (1-0.36) = 377,927.

Using exponential smoothing (α = 0.15) on available data until 2016, it is expected this total market size will be achieved just after 2022. For simplicity, we assume the diffusion curve starts at 2009 and ends at 2022. Subsequently, the percentage of new creators of the total market size can be determined. Lastly, using the percentages of Rogers' Adoption curve the five adoption categories have been assigned to each year.

This is a very basic model and is based on a couple of assumptions:

- The fact that 64% of all Americans was familiar with crowdfunding in 2016, has been interpreted as the aggregated number of new creators in 2016 is 64% of its total market size.
- It is assumed 76% of the total number of yearly projects has been created by creators who have never created a project before (i.e. "new creators"). This rate is based on our own data sample. For the calculations a constant rate is assumed throughout the years, though this is obviously not the case: in the early phases this rate was probably higher.
- For years indicated with two stars (**) the number of yearly new creators have been forecasted. Due to exponential smoothing the downwards trend in 2016 compared to 2015 repeats itself in the next 6 years. In fact, this might not be the case due to the growing market for crowdfunding.
- Given the normal distribution (bell-shape) it would make sense that the peak is in 2015 and the right tail ends in 2021 (because 2015 - 2009 = 6). However, this would result in a total market size of 364K which is further from our expected value than if it is assumed the diffusion curve ends in 2022. Therefore, the latter has been assumed.

Year	#Pojects*	#New creators	Percent of total	Adopter category
2009	874	874	0.2%	Innovators
2010	8,597	6,534	1.7%	Innovators
2011	23,262	17,679	4.7%	Early Adopters
2012	39,625	30,136	8.0%	Early Adopters
2013	44,094	33,511	8.9%	Early Majority
2014	66,558	50,584	13.4%	Early Majority
2015	76,867	58,419	15.5%	Early / Late Majority
2016	58,074	44,136	11.7%	Late Majority
2017	-	35,373**	9.4%	Late Majority
2018	-	28,753**	7.6%	Late Majority
2019	-	23,447**	6.2%	Late Majority / Laggards
2020	-	19,134**	5.1%	Laggards
2021	-	15,617**	4.1%	Laggards
2022	-	12,747**	3.4%	Laggards
TOT	317,951	376,943	≈ 100%	

Figure VII.I - Number of yearly Kickstarter projects, new creators, percentage relative to aggregated creator amount (2009-2022), and corresponding adopter category.

**Number of yearly projects taken from: (Bideaux, 2017)*

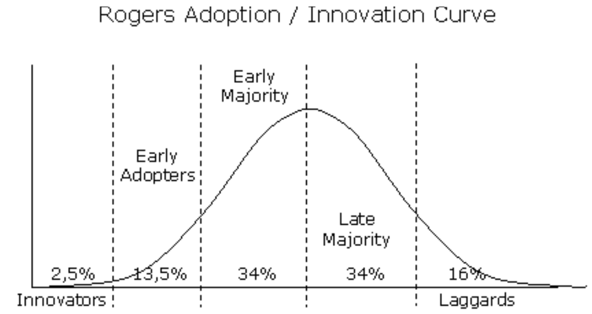


Figure VII.II - Rogers adoption curve and categories

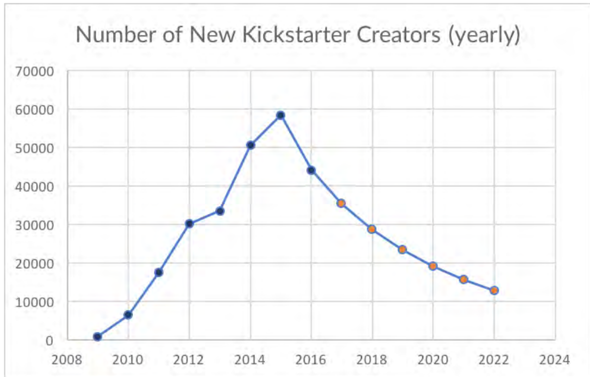


Figure VII.II - Expected number of new Kickstarter creators (yearly). Orange markers are based on data from 2016 and earlier.

Appendix VIII - Predictors Success Rate

The following attributes have been left out from our Research Design (Appendix III): category, subcategory, staff-picked and first level pledge. That is because they don't show any significant relationship with the project success and thus deteriorate the model's performance.

Figure VIII.I - Logistic regression output (t=launch)

Attribute	Estimate	Std. Error	P-value
Goal	-5.58e-06	3.71e-06	0.13
Average number of backers required	-2.27e+01	3.02e+00	6.36e-14***
#Projects backed by creator	4.81e-01	1.71e-01	0.0049**
#Projects created by creator	4.07e-01	1.30e-01	0.0018**
#Facebook friends	1.33e-01	4.90e-02	0.0065**
Currency	9.92e-02	2.47e-02	5.96e-05***
Maximum pledge tier	-1.79e-01	6.02e-02	0.0029**
Number of rewards	5.60e-01	6.81e-02	<2e-16***
Number of collaborators	1.57e-01	6.31e-02	0.013*
Past success rate	4.53e-01	5.49e-02	<2e-16***
#Words in project description	1.67e-01	6.46e-02	0.0010**
Presence of video	2.56e-01	4.88e-02	1.62e-07***
#Additional videos	1.34e-01	6.27e-02	0.03*
#Images	5.32e-01	7.06e-02	4.90e-14***
Duration	-1.93e-01	5.17e-02	0.00019***
Estimated delivery	-1.91e-02	6.06e-02	0.0016**
Adoption	2.26e-01	5.72e-02	7.88e-05***

Dependent variable: ***project success***

Figure VIII.II - Logistic regression output (t=campaign completion) - without continent

Attribute	Estimate	Std. Error	P-value
Goal	-2.16e-04	1.89e-05	<2e-16***
Average number of backers required	-5.65e+01	8.79+00	3.10e-10***
#Projects backed by creator	-8.99e-02	1.40e-01	0.52
#Projects created by creator	3.50e-01	1.63e-01	0.032*
#Facebook friends	1.20e-01	6.93e-02	0.08
Currency	1.40e-01	3.70e-02	0.00016***
Maximum pledge tier	1.57e-01	1.12e-01	0.16
Number of rewards	1.62e-01	1.06e-01	0.13
Number of collaborators	1.13e-01	1.20e-01	0.35
Past success rate	3.22e-01	7.44e-02	1.51e-05***
#Words in project description	4.52e-02	1.07e-01	0.67
Presence of video	1.09e-01	6.90e-02	0.11
#Additional videos	-1.98e-01	1.09e-01	0.068
#Images	2.30e-02	1.15e-01	0.84
Duration	-9.45e-02	7.11e-02	0.18
Estimated delivery	-2.90e-01	9.41e-02	0.0021**
Adoption	-8.35e-02	8.10e-02	0.30
Number of updates	7.83e-01	1.40e-01	2.02e-08***
Number of comments	1.05e+00	2.21e+00	0.63
Total number of backers	3.61e+01	2.18e+00	<2e-16***

Dependent variable: ***project success***

Figure VIII.III - Logistic regression output (t=campaign completion) - with continent

Attribute	Estimate	Std. Error	P-value
Goal	-2.16e-04	1.89e-05	<2e-16
Average number of backers required	-5.59e+01	8.98e+00	4.97e-10***
#Projects backed by creator	-8.50e-02	1.40e-01	0.55
#Projects created by creator	3.44e-01	1.63e-01	0.035*
#Facebook friends	1.24e-01	6.96e-02	0.076
Currency	1.42e-01	3.72e-02	0.00014***
Continent	-1.21e-01	1.01e-01	0.23
Maximum pledge tier	1.61e-01	1.11e-01	0.15
Number of rewards	1.63e-01	1.06e-01	0.12
Number of collaborators	1.08e-01	1.20e-01	0.37
Past success rate	3.27e-01	7.48e-02	1.23e-05***
#Words in project description	3.92e-02	1.07e-01	0.72
Presence of video	1.10e-01	6.91e-02	0.11
#Additional videos	-1.95e-01	1.09e-01	0.07
#Images	2.28e-02	1.15e-01	0.84
Duration	-9.64e-02	7.12e-02	0.18
Estimated delivery	-2.90e-01	9.41e-02	0.0020**
Adoption	-7.40e-02	8.13e-02	0.36
Number of updates	7.81e-01	1.40e-01	2.17e-08***
Number of comments	1.13e+00	2.22e+00	0.61
Total number of backers	3.60e+01	2.18e+01	<2e-16***

Dependent variable: ***project success***

Figure VIII.IV - Logistic regression output (t=campaign completion) - without total number of backers

Attribute	Estimate	Std. Error	P-value
Goal	-12.28	1.70	4.53e-13***
Average number of backers required	-31.72	4.13	1.50e-14***
#Projects backed by creator	0.0069	0.11	0.95
#Projects created by creator	0.19	0.11	0.010**
#Facebook friends	0.21	0.05	0.00016***
Currency	0.012	0.028	5.31e-05***
Maximum pledge tier	-0.0055	0.074	0.94
Number of rewards	0.46	0.077	2.97e-09***
Number of collaborators	0.21	0.074	0.0041**
Past success rate	0.54	0.064	<2e-16***
#Words in project description	0.085	0.080	0.29
Presence of video	0.60	0.12	3.03e-07***
#Additional videos	-0.11	0.079	0.15
#Images	0.040	0.089	0.65
Duration	-0.19	0.058	0.00091***
Estimated delivery	-0.24	0.071	0.00082***
Adoption	0.089	0.063	0.16
Number of updates	1.38	0.11	<2e-16***
Number of comments	15.29	1.83	<2e-16***

Dependent variable: ***project success***

Figure VIII.V - Multicollinearity check (I)

Attribute	Tolerance	VIF
Goal	0.362	2.761
Average number of backers required	0.367	2.728
#Projects backed by creator	0.800	1.250
#Projects created by creator	0.832	1.202
#Facebook friends	0.966	1.035
Currency	0.861	1.162
Continent	0.868	1.153
Maximum pledge tier	0.867	1.154
Number of rewards	0.644	1.552
Number of collaborators	0.963	1.039
Past success rate	0.899	1.112
#Words in project description	0.601	1.665
Presence of video	0.958	1.044
#Additional videos	0.854	1.171
#Images	0.524	1.908
Duration	0.953	1.050
Estimated delivery	0.919	1.088
Adoption	0.688	1.454
Number of updates	0.585	1.709
Number of comments	0.245	4.079
Total number of backers	0.236	4.245

Dependent variable: ***project success***

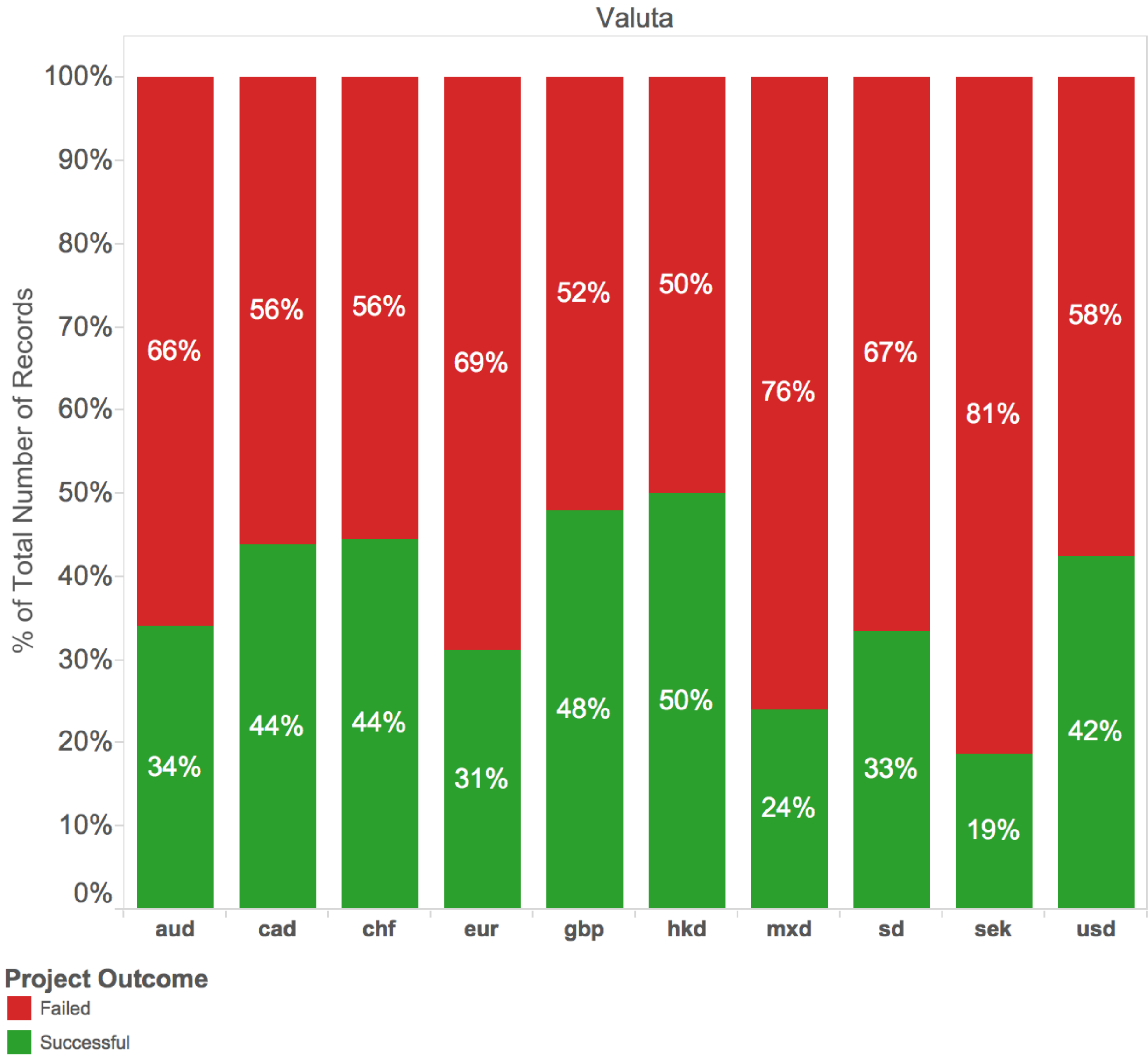
The “average backers required” is indirectly derived from the “goal” amount which explains the high VIF-value for both variables. The same holds for the “number of comments” and “total number of backers”, since only backers can comment on a campaign page. To confirm this, the same model with one of those two variables has been run (Figure VIII.VI).

Figure VIII.VI - Multicollinearity check (II)

Attribute	Tolerance	VIF
Goal	0.959	1.043
#Projects backed by creator	0.798	1.254
#Projects created by creator	0.809	1.237
#Facebook friends	0.965	1.036
Currency	0.863	1.159
Continent	0.873	1.145
Maximum pledge tier	0.789	1.267
Number of rewards	0.646	1.548
Number of collaborators	0.963	1.039
Past success rate	0.904	1.107
#Words in project description	0.601	1.665
Presence of video	0.958	1.044
#Additional videos	0.854	1.171
#Images	0.526	1.900
Duration	0.954	1.048
Estimated delivery	0.922	1.085
Adoption	0.689	1.451
Number of updates	0.588	1.701
Number of comments	0.861	1.162

Dependent variable: ***project success***

Appendix IX - Currency



Abbreviation	Currency
aud	Australian Dollar
cad	Canadian Dollar
chf	Swiss Franc
eur	Euro
gbp	British Pound
hkd	Hong Kong Dollar
mxd	Mexican Peso
sd	Singapore Dollar
sek	Swedish Krona
usd	United States Dollar

Figure IX.I - Success rate across currencies (only currencies with > 20 sample points have been selected)

Appendix X - Optimal Cut-off Levels

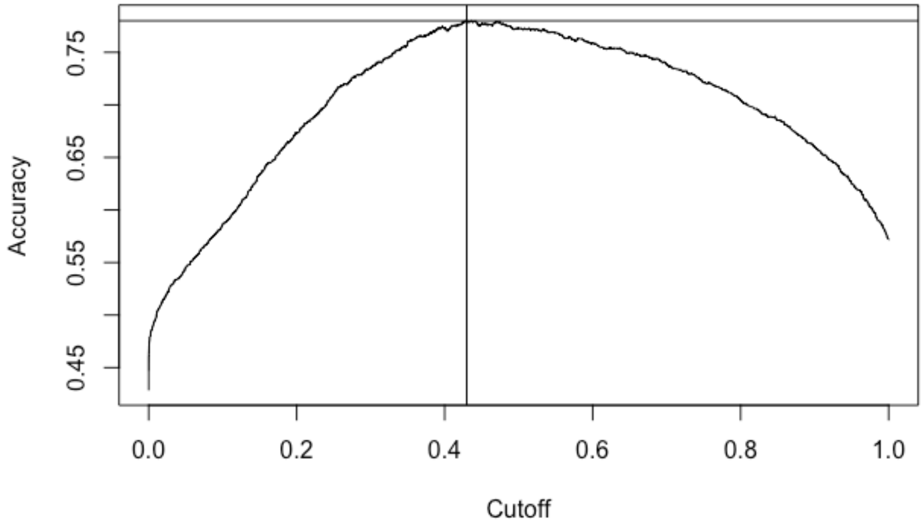


Figure X.I - Optimal cut-off level based on accuracy of model (t= launch) - (0.430, 0.789)

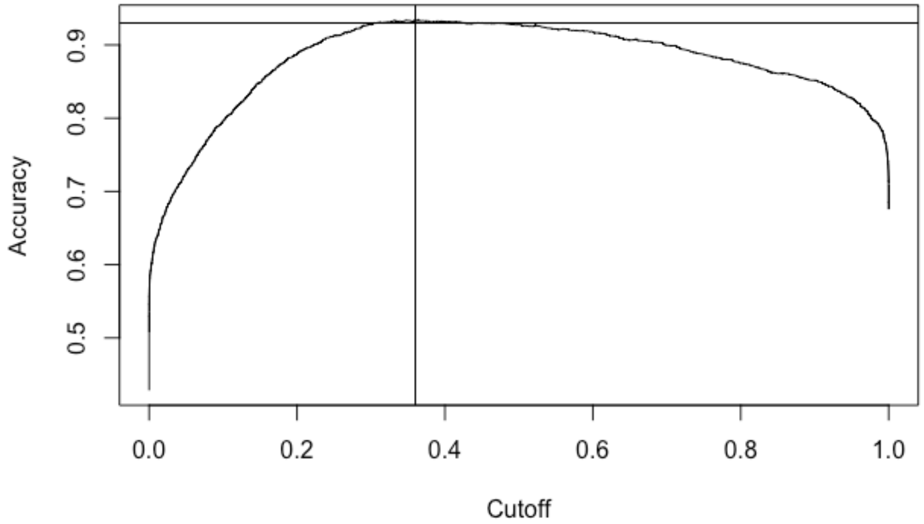


Figure X.II - Optimal cut-off level based on accuracy of model (t=deadline) - (0.364, 0.936)

Appendix XI - ROC curves

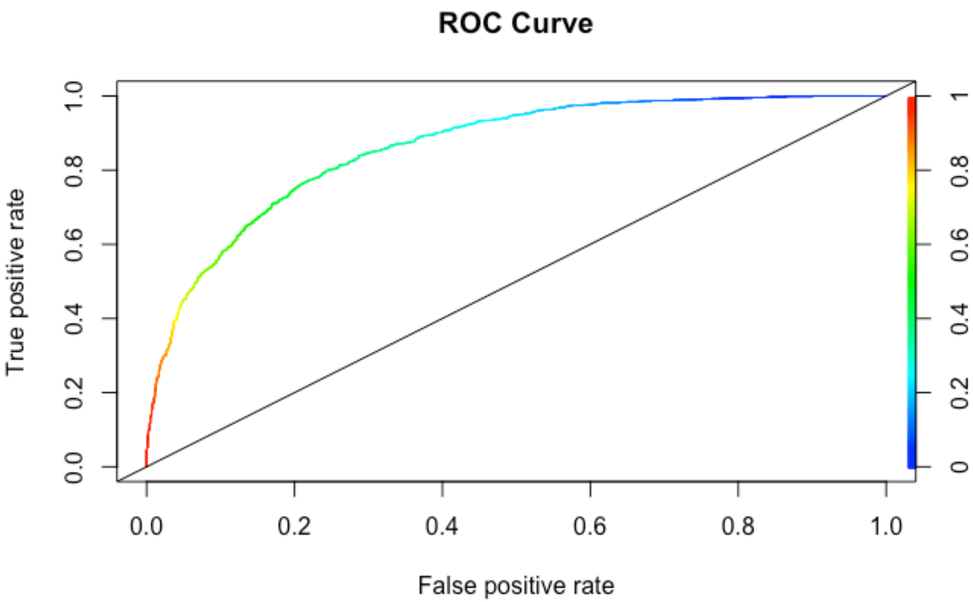


Figure XI.I - ROC-curve (t= launch) - AUC = 0.861

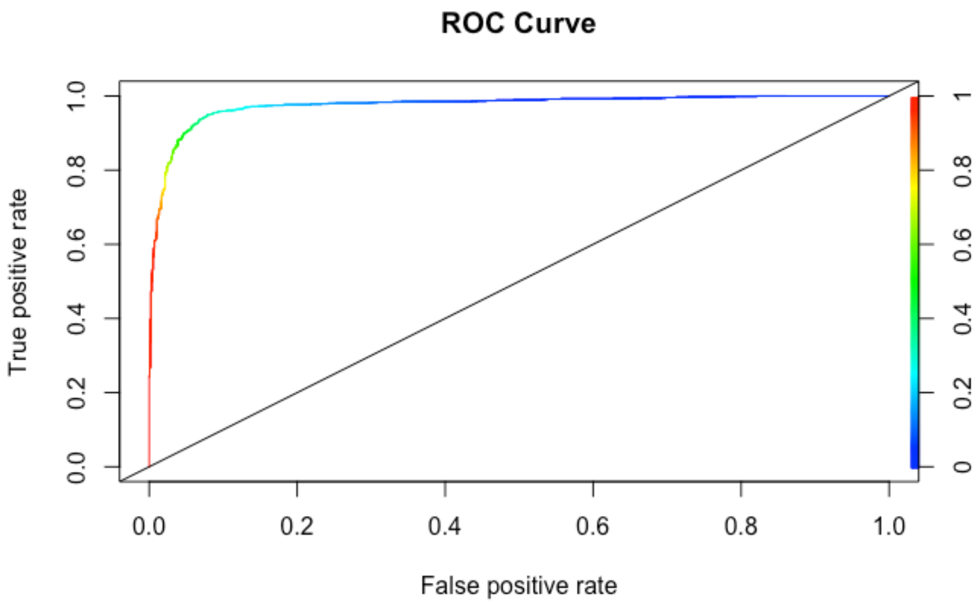


Figure XI.II - ROC-curve (t= deadline) - AUC = 0.975

Appendix XII - Kickstarter Campaign Set-up

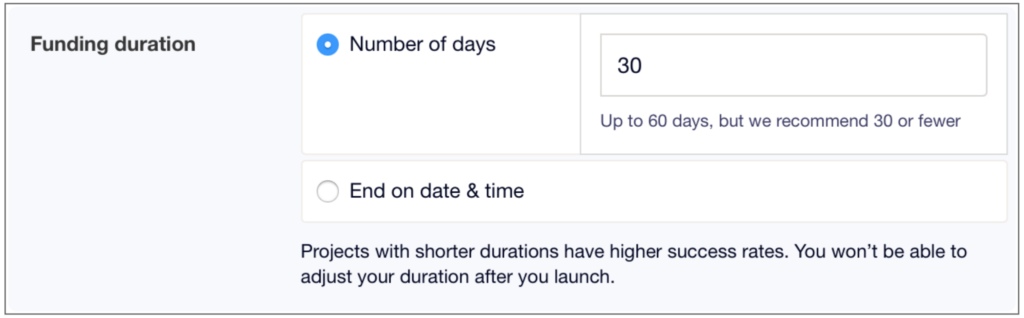


Figure XII.I - Kickstarter funding duration options

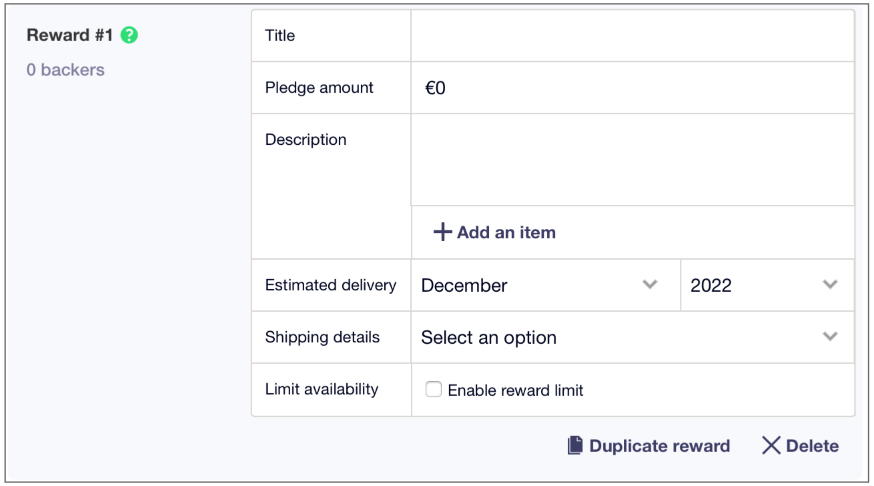


Figure XII.II - Kickstarter reward options

Appendix XIII - #Previous Projects Created vs Current Success