# Crowdfunding Capital: A Study on Factors Related to Kickstarter Success

Daniel Awotundun\*, Nicholas Wilger\*

Webber Academy
*\*Both authors contributed equally to the writing and research in this study. Their names are listed in alphabetical order.*

February 4, 2019

## Abstract

As crowdfunding becomes an increasingly popular method to start and grow businesses around the world, more insight into its intricacies is critical. To date, the Kickstarter platform alone has raised over $4.07 billion USD, spread across over 430,000 projects [1]. In this context, our study aims to analyze the factors most closely correlated to the eventual success of Kickstarter projects. Data from 42,951 unique campaigns, scraped by WebRobots.io [7], was used to investigate the relationship between these attributes and the target. Data analysis and manipulation tools, such as Orange, Tableau, and Excel, were used to aggregate, model, and visualize the dataset. Attributes were then ranked by their information gain and modeled using kNN, constant, and logistic regression algorithms. These findings were subsequently visualized using multiple methods, with data accuracy and clarity at the forefront of our investigation. It was concluded that factors like spotlight and backer penetration had the greatest influence on success, while the project category and country of origin attributes, among others, also played a significant role. If development is truly turning towards crowdfunding as a means of raising capital, the success of these projects will increase in importance. Using the findings of our paper, prospective project creators will have more information going into a crowdfunding campaign. Hopefully, this information will be used by entrepreneurs worldwide as they aim to innovate and create.

**Keywords** crowd-funding; entrepreneurship; factor analysis; alt-metric data

## Introduction

One major obstacle that any new business venture faces is the lack of sufficient capital and collateral. The traditional way to overcome this is to approach financial investors such as banks, venture capital funds, or angel investors. However, many entrepreneurs are turning to the Internet to gain the funds they need directly from the general public. This technique, called "crowdfunding," has made it possible for any entrepreneur with a vision to carry out their plan and gain the support they need. Kickstarter.com is the most popular crowdfunding platform of this kind [9]. The site follows a similar model to many of its competitors: creators submit projects, varying from small art installations to high-tech startups, and contributors pledge money to see the project become successful. In return, contributors are offered benefits based on their pledge, such as early access to products. Although Kickstarter is not available to entrepreneurs internationally, contributors from all over the world are able to pledge money towards their favourite creations. The fundraising industry extends beyond Kickstarter as well. Crowdfunding as a whole is growing at a blistering pace. As of 2019, the industry has created over 270,000 jobs, raised over $34 billion USD, and added over $65 billion USD to the global economy [3]. On top of that, approximately 60% of successful campaigns are in business within a year [5]. These staggering statistics show that crowdfunding truly is the future for startups. Because widespread crowdfunding is a relatively new concept and has only grown popular in the past decade, there has been little research into what makes a project successful. The few papers that do have maximum sample sizes

of less than 2,500 projects [2]. Our dataset consists of 42,951 unique campaigns, a number significantly larger than those of previous studies, allowing us to create more accurate conclusions. Our analysis aims to examine factors, such as the length or goal of the campaign, to provide future project creators with the necessary information to increase their chances of success, represented by the total funds raised as well as the pledge multiplier. Data analysis and modeling software such as Orange and Tableau were used to accomplish this goal with clear charts and diagrams to visualize our findings. In the coming decades, crowdfunding platforms will play a much greater role in the financing and creation of startups, especially as mankind looks increasingly towards the skies for growth and exploration. The crucial first step in this process is ensuring entrepreneurs and startups have the resources they need. More research into crowdfunding, such as ours, is necessary to ensure their success and to continue innovation.

## Materials & Methods

To understand why some crowdfunding projects are successful while others are not, twenty datasets from the years 2017 and 2018 were extracted from Kickstarter.com using Webrobots.io [7], with a total of 42,951 unique projects included. Using Microsoft Excel, these datasets were aggregated into two separate master sets, one for training (made from sixteen of the original datasets) and one for testing (made from four of the original datasets). As well, of thirty original attributes, ten of the most relevant were selected for analysis, including project category, set goal, campaign length, raised funds, country of origin, spotlight, and staff pick. New attributes were calculated or created, including campaign length and pledge multiplier (the factor by which a project exceeds its goal). Any Kickstarter projects not yet completed, missing information, or found to be duplicates were excluded from both datasets to ensure our findings were as accurate as possible. Orange, an open-source data visualization, machine learning, and data mining toolkit created by the University of Ljubljana [11], was used to analyze the dataset. Twelve different widgets were utilized, including one to rank feature attributes in relation to the target attribute, and one to test and score different models in relation to the master test file. In the end, we modeled the data using kNN, constant, and logistic regression algorithms. The software was also useful in creating the tables shown below. Using the information obtained from Orange, the Tableau

desktop software was used to further organize the dataset and visualize correlations. Feature attributes were compared using multiple visualization methods (such as scatter plots and bar charts) and were also compared to the target attributes of success and final funding. Each chart was optimized for aesthetics as well as data clarity. Outliers in the dataset were sometimes excluded in our visualizations.

## Results

In Orange, each attribute was ranked based on the information gain each produced in predicting success. For example, it was found that whether a project was Spotlighted or not had an information gain of 0.9855596198534557, while the length of the campaign only had an information gain of 0.01420938108037828. This is shown below in Figure 1.

**Target: State**

**Ranks**

| | # | Info. gain |
|---|---|---|
| Spotlight | 2.0 | 0.9855596198534557 |
| Percentage Pledged | | 0.7548888951251964 |
| Backers Count | | 0.4780364717101253 |
| Pledged Amount (USD) | | 0.4502930918189336 |
| Category | 15.0 | 0.08678950349404846 |
| Staff Pick | 2.0 | 0.050069703605189586 |
| Goal | | 0.044152884037910844 |
| Country | 22.0 | 0.01463606632652592 |
| Campaign Length | | 0.01420938108037828 |

Figure 1: This table shows the ranks of different attributes based on information gain (relating to the target success attribute).

The strongest pairs of attributes were also found and ranked using both Pearson and Spearman Correlations. Unsurprisingly, both indicated that the reach of the project, as well as the amount of funding raised, had the greatest effect on the success of any given project, with values 0.752 and 0.941 respectively (Figures 2 and 3).

**Correlations**

**Pearson correlation**

| | | |
|---|---|---|
| 1 | +0.752 | Backers Count, Pledged Amount (USD) |
| 2 | +0.261 | Backers Count, Percentage Pledged |
| 3 | +0.245 | Percentage Pledged, Pledged Amount (USD) |
| 4 | +0.030 | Campaign Length, Pledged Amount (USD) |
| 5 | +0.027 | Campaign Length, Goal |
| 6 | +0.021 | Backers Count, Campaign Length |
| 7 | -0.008 | Campaign Length, Percentage Pledged |
| 8 | -0.007 | Goal, Percentage Pledged |
| 9 | +0.007 | Goal, Pledged Amount (USD) |
| 10 | +0.005 | Backers Count, Goal |

Figure 2: This table shows the Pearson Correlation between different pairs of attributes and the target success attribute.

**Correlations**

**Spearman correlation**

| | | |
|---|---|---|
| 1 | +0.941 | Backers Count, Pledged Amount (USD) |
| 2 | +0.780 | Backers Count, Percentage Pledged |
| 3 | +0.730 | Percentage Pledged, Pledged Amount (USD) |
| 4 | -0.285 | Goal, Percentage Pledged |
| 5 | +0.274 | Goal, Pledged Amount (USD) |
| 6 | +0.227 | Campaign Length, Goal |
| 7 | +0.166 | Backers Count, Goal |
| 8 | -0.072 | Campaign Length, Percentage Pledged |
| 9 | +0.071 | Campaign Length, Pledged Amount (USD) |
| 10 | +0.039 | Backers Count, Campaign Length |

Figure 3: This table shows the Spearman Correlation between different pairs of attributes and the target success attribute.

Finally, the kNN, constant, and logistic regression algorithms used were tested and ranked by their classification accuracy. The k-nearest neighbour model had the highest classification accuracy with a score of 0.634, followed by the logistic regression model with a score of 0.619, and finally the constant model with a score of 0.571 .

**Test & Score**

**Settings**

Sampling type: No sampling, test on testing data
Target class: Average over classes

**Scores**

| Method | AUC | CA | F1 | Precision | Recall |
|---|---|---|---|---|---|
| Constant | 0.500 | 0.571 | 0.415 | 0.326 | 0.571 |
| Logistic Regression | 0.680 | 0.619 | 0.542 | 0.665 | 0.619 |
| kNN | 0.595 | 0.634 | 0.598 | 0.640 | 0.634 |

Figure 4: This table shows the tested scores of each model used. The k-nearest neighbour algorithm yielded the highest classification accuracy with a value of 0.634.

Tableau was then used to compare the different attributes and visualize them. Many varied visualization techniques were used, ranging from bar charts to packed bubbles visualizations. These visualizations are shown in the appendix.

# Discussion

In any crowdfunding scenario, the actual product or idea itself is the single most influential factor in predicting its possible success. However, the influence of multiple altmetric factors also has an undeniable role in shaping the success of the project. One surprising find is that the length of the campaign actually had a very small effect on the overall success of the project. In fact, using the rank widget in Orange, campaign length actually has the least information gain in trying to predict project success (with a value of only about 0.014). On the other hand, whether a project was put in the "Spotlight" section of Kickstarter or not had an information gain of nearly 0.99 when attempting to predict project success. Another interesting factor influencing the funding a project receives, and its eventual success or failure, is the category that the project is in. Though the music, film, publishing, and art categories see the highest number of projects, they also receive some of the lowest total funding. One possible explanation is that the funding goals in these categories are lower, on average than the average project in the technology category, but this possibility is discounted if you look at the average pledge multiplier. Instead, a more likely explanation is that the popularity of these categories limits the amount of funding that can be allocated to each individual project. Consumer interest in these categories, as well as the lack of physical "perks" for contributors, also likely has an effect. However, future research into this phenomenon may prove useful in determining what exactly causes this difference in funding between the categories. This observation is also part of a larger discovery. Project success, pledged money, backers, and other key attributes vary drastically depending on seemingly less important factors. The category that the project is in has great effects, but the variance is not completely unexpected. However, looking at the country each project is created from reveals even more differences that are much harder to pin down. A project started in Hong Kong receives an average pledge of nearly $160 USD, nearly quadruple the average pledge that a Belgian or Japanese project receives ($45.50 and $38.90 respectively, the lowest values in our analysis). At the same time, an examination of the average number of backers different projects receive now shows Japan climbing from the lowest in average pledges to the third. This

divergence is more than likely based on many different factors, possibly including the different cultures of regions of the world. More insight into this aspect of crowdfunding would also prove interesting in future analyses. As for our specific methodologies, the use of Orange and Tableau proved instrumental in crafting our conclusions. Unlike alternative data analysis software like Weka, Orange uses intuitive "widgets" to perform all tasks, ranging from viewing data to modeling it. At the same time, its robust tools allow us to look at our dataset from different points of view, creating the best possible conclusions. To visualize our findings, we decided to use Tableau instead of Excel. Though the software was initially more difficult to use, the possibilities it provided outweighed its drawbacks. Many different tools, ranging from simple scatter plots to heat maps and tree maps allowed us to clearly portray our data while also going more in-depth than other software would have allowed. Despite the software at our disposal, it would be foolish to think our conclusions were completely accurate. For one, our dataset came from only one crowdfunding platform. Each has its own unique model for project funding, so our analyses may not apply to other platforms. As well, our data-set, though it included over 40,000 unique projects, was still a small sample size. Attributes with relatively rare occurrences, such as Kickstarter projects originating from Belgium or in the "dance" category would have likely skewed our results. The relatively small sample size would also affect our "average pledge" and "average backers" values. Crowdfunding will undoubtedly play a major role in the future of finance and business. For that reason, more investigation into different platforms and the inclusion of larger sample sizes, along with recording a greater number of attributes, will give entrepreneurs further control over their projects.

## Conclusions

Our findings are more important today than they have ever been. In an era of very diverse levels of education and expertise as well as online educational resources, it is imperative that we do not neglect the impact a small group of innovators can have in the industry. The eventuality of long term and long-distance space travel is aided by the tendency backers have towards technological campaigns shown in figure. The potential advantages are limited when considering that most successful campaigns receive funding within 76 weeks according to figure. The efficacy of this platform allows for prolific innova-

tion that will create development in several technological domains. With our predictive models that characterize the optimal name length , pledge amount , campaign length , and optimal advertising , 3rd parties are able to pitch the perfect product within optimal parameters receiving funding in a 76-week time frame. Potentially, a formula that incorporates all parameters could be developed to determine the optimal project. Subsequently, an index could be developed that quantifies any restrictions we have due to parameters being subjective. If titles and descriptions can be quantified numerically, then the potential discrepancies created by omitting those parameters can be removed. Crowdfunding will undoubtedly play a major role in the future of finance and business. For that reason, more investigation into different platforms and including a larger sample size, along with recording a greater number of attributes, would only give entrepreneurs further control over their projects. Using the findings of our paper, prospective project creators will have more information going into their next crowdfunding project. Seemingly less relevant factors like the project name, campaign length, and country of origin do indeed play a role in influencing the success of a project, however incremental. A deeper understanding of these attributes cannot be ignored in the future of crowd-funding.

## Acknowledgements

## References

[1] Bidaux, T. (2018, February 01). Kickstarter in 2017- Year in Review. Retrieved from http://icopartners.com/2018/01/kickstarter-2017-year-review/

[2] Crosetto, P., Regner, T. (n.d.). Crowdfunding: Determinants of success and funding dynamics [Abstract]. EconStor. Retrieved January 28, 2019.

[3] Crowdfunding Statistics: The Facts About the LAtest Fundraising Craze. (n.d.). Retrieved from

https://blog.fundly.com/crowdfunding-statistics/

[4] Duncan, E.(n.d.). Topic: Kickstarter. Retrieved from www.statista.com/topics/2102/kickstarter/

[5] Equity Crowdfunding Statistics 2018. (2018, July 09). Retrieved from https://www.crowdcrux.com/equity-crowdfunding-statistics/

[6] Giudici, G., Guerini, M., & Rossi-Lamastra, C. (n.d.). Why Crowdfunding Projects can Succeed: The Role of Proponents' Individual and Territorial Social Capital [Abstract]. Social Science Research Network. Retrieved January 28, 2019.

[7] Kickstarter Datasets. (n.d.). Retrieved from https://webrobots.io/kickstarter-datasets/

[8] Kickstarter: Projects and dollars 2019 Statistic. (n.d.). Retrieved from https://www.statista.com/statistics/251727/projects-and-dollars-overview-on-crowdfunding-platform-kickstarter/

[9] Kim, L., & Kim, L. (2018, August 06). Top 10 Crowdfunding Platforms of 2018. Retrieved from https://www.inc.com/larry-kim/op-10-crowdfunding-platforms-of-2018.html

[10] Moss, A., & Moss, A. (2018, April 30). More startups are meeting success through crowdfunding. Retrieved from https://medium.com/swlh/more-startups-are-meeting-success-through-crowdfunding-c1292b2cad47

[11] Orange (software). (2019, January 06). Retrieved from https://en.wikipedia.org/wiki/Orange

[12] Our mission is to help bring creative projects to life. (n.d.). Retrieved from https://www.kickstarter.com/about?ref=global-footer

[13] Tableau: Business Intelligence and Analytics Software. (n.d.). Retrieved from https://www.tableau.com/

[14] Tableau Training & Tutorials. (n.d.). Retrieved from https://www.tableau.com/learn/training

[15] 10 Crowdfunding Statistics to Raise More Money. (2018, September 09). Retrieved from https://www.crowd101.com/crowdfunding-success-statistics-raise-money-online/

[16] University of Ljubljana. (n.d.). Orange – Data Mining Fruitful & Fun. Retrieved from https://orange.biolab.si/

# Appendix

Average Pledge Multiplier vs. Category
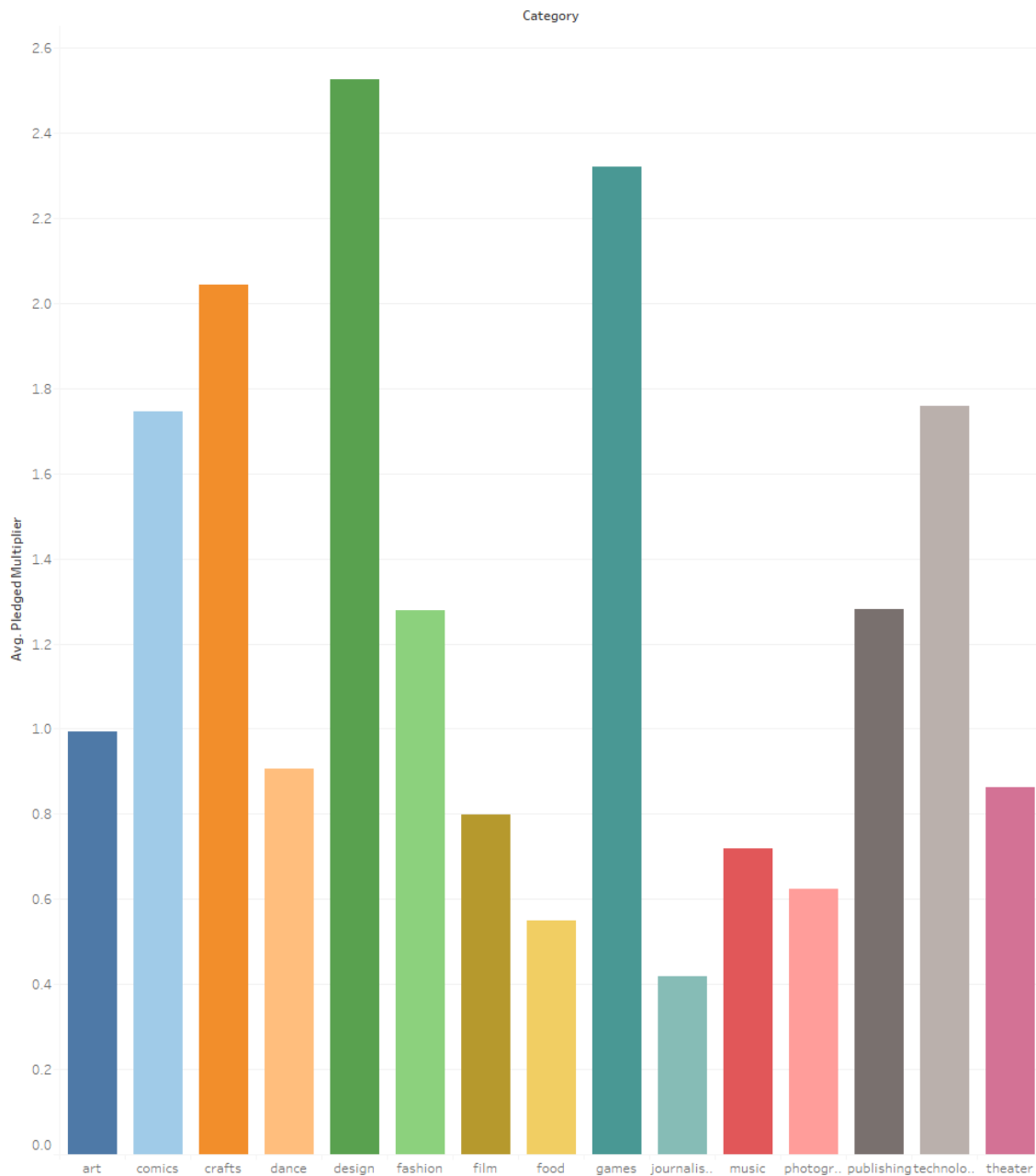


Average of Pledged Multiplier for each Category. Color shows details about Category as an attribute.

Figure 5: This bar chart compares the average pledge multiplier of different project categories. The average project in the design category exceeded its goal by a factor of 2.526, while the average project in the journalism category only managed to receive 41.9% of its initial goal.
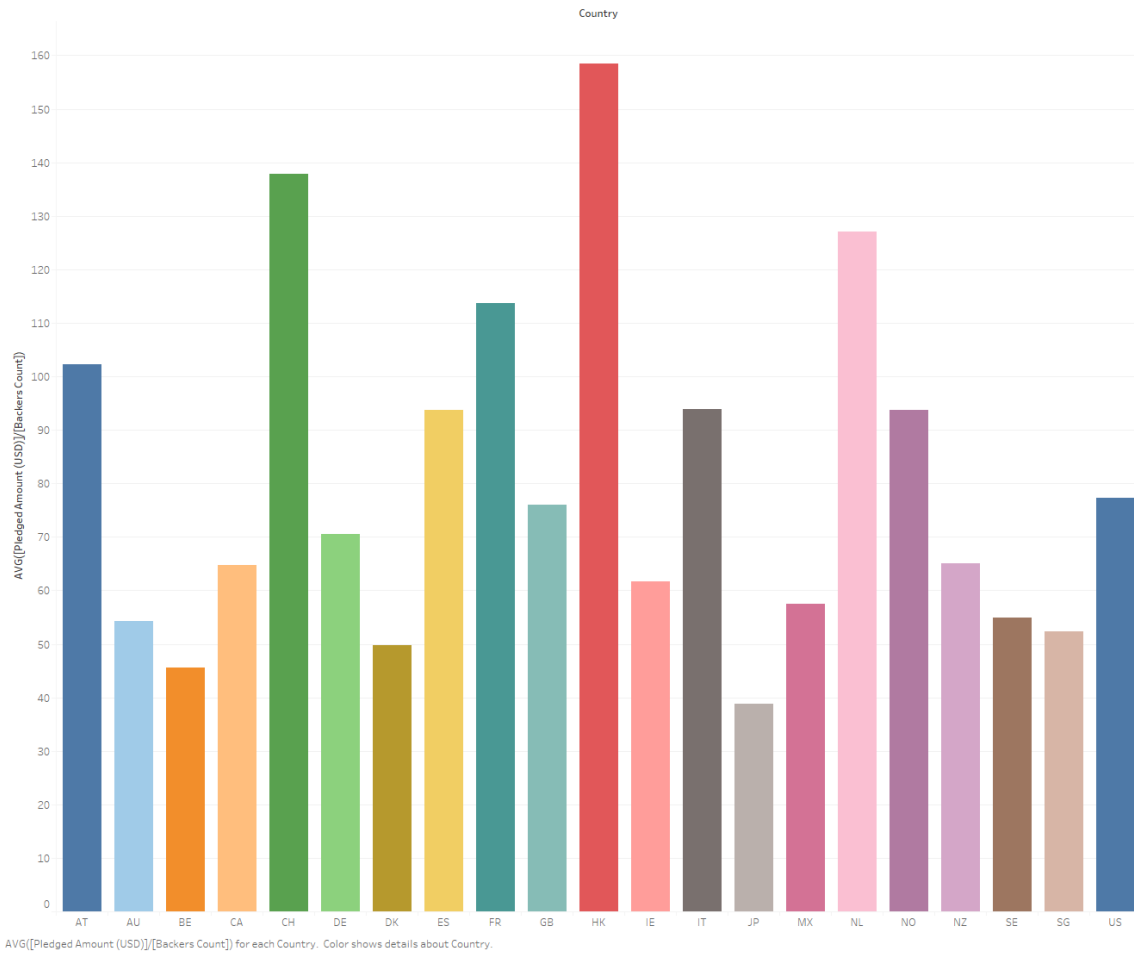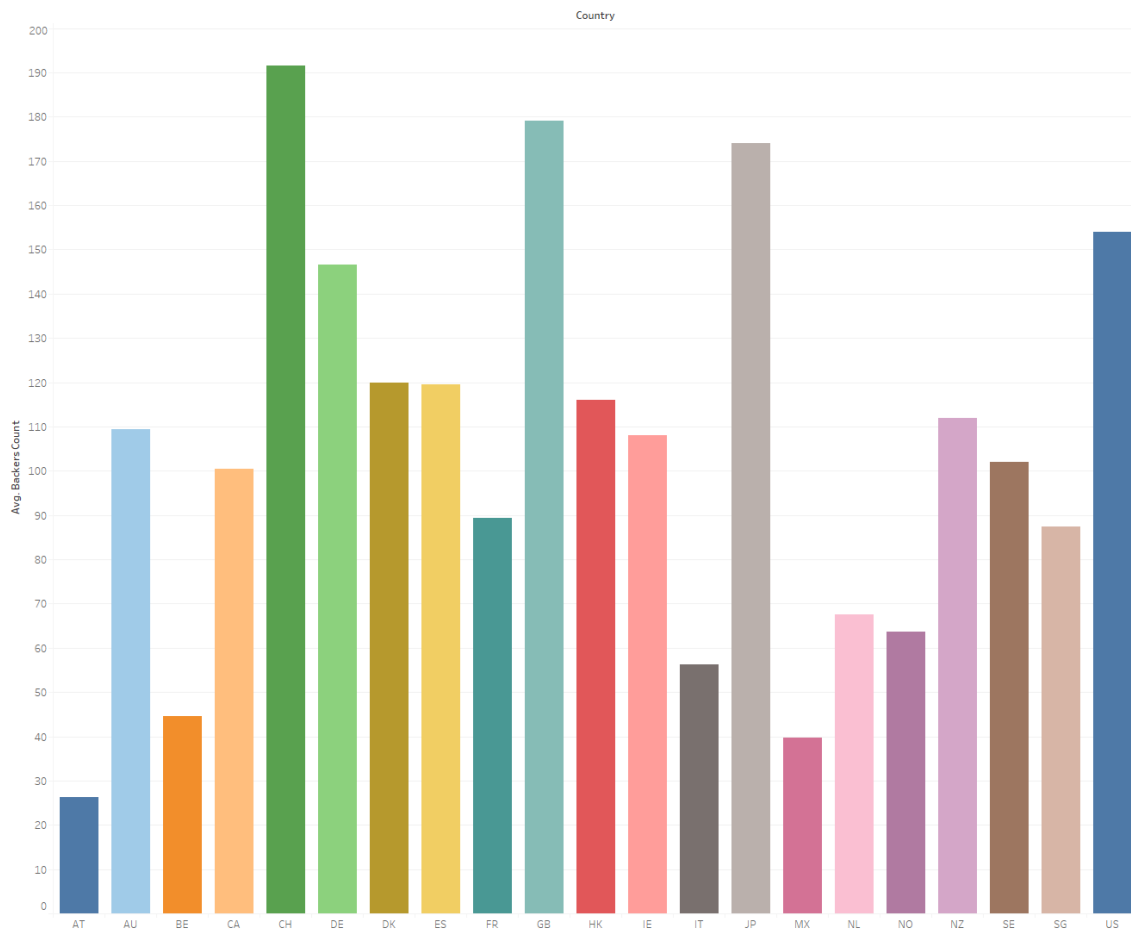
Figure 6: This bar chart compares the average pledge a project receives based on its country of origin. The average project from Hong Kong received an average pledge of $158.40 USD, while the average project from Japan received an average pledge of $38.90 USD. It is important to note that the project success cannot be determined from this graph, as some projects may receive a small number of highly valuable pledges, while another receives many low-value pledges.

Figure 7: This bar chart compares the average number of backers a project receives based on its country of origin. The average project from Switzerland received 192 backers, while the average Austrian project received 26 backers. It is important to note that the project success cannot be determined from this graph, as some projects may receive a small number of highly valuable pledges, while another receives many low-value pledges.

Category Popularity
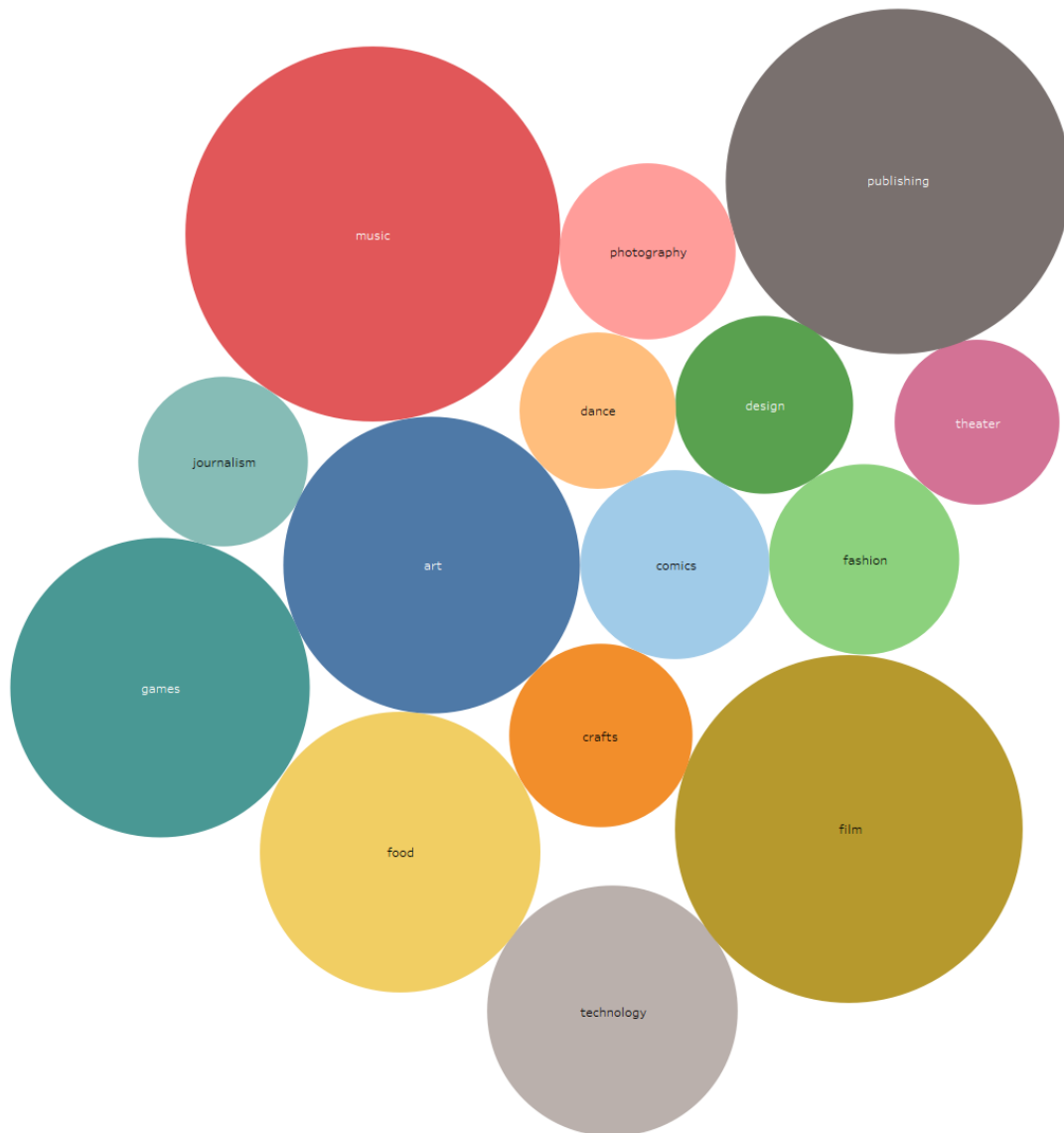


Category. Color shows details about Category. Size shows count of Category. The marks are labeled by Category.

Figure 8: This packed bubbles visualization shows the number of projects under each category. The larger bubbles signify a greater number of projects in that category, while smaller bubbles signify a relatively fewer number of projects.

Category Popularity

Figure 9: This bar chart shows the number of projects under each category. Music has the most, with 1569 recorded projects, while dance had the least, with 273 recorded projects.

Failed and Successful Project Count
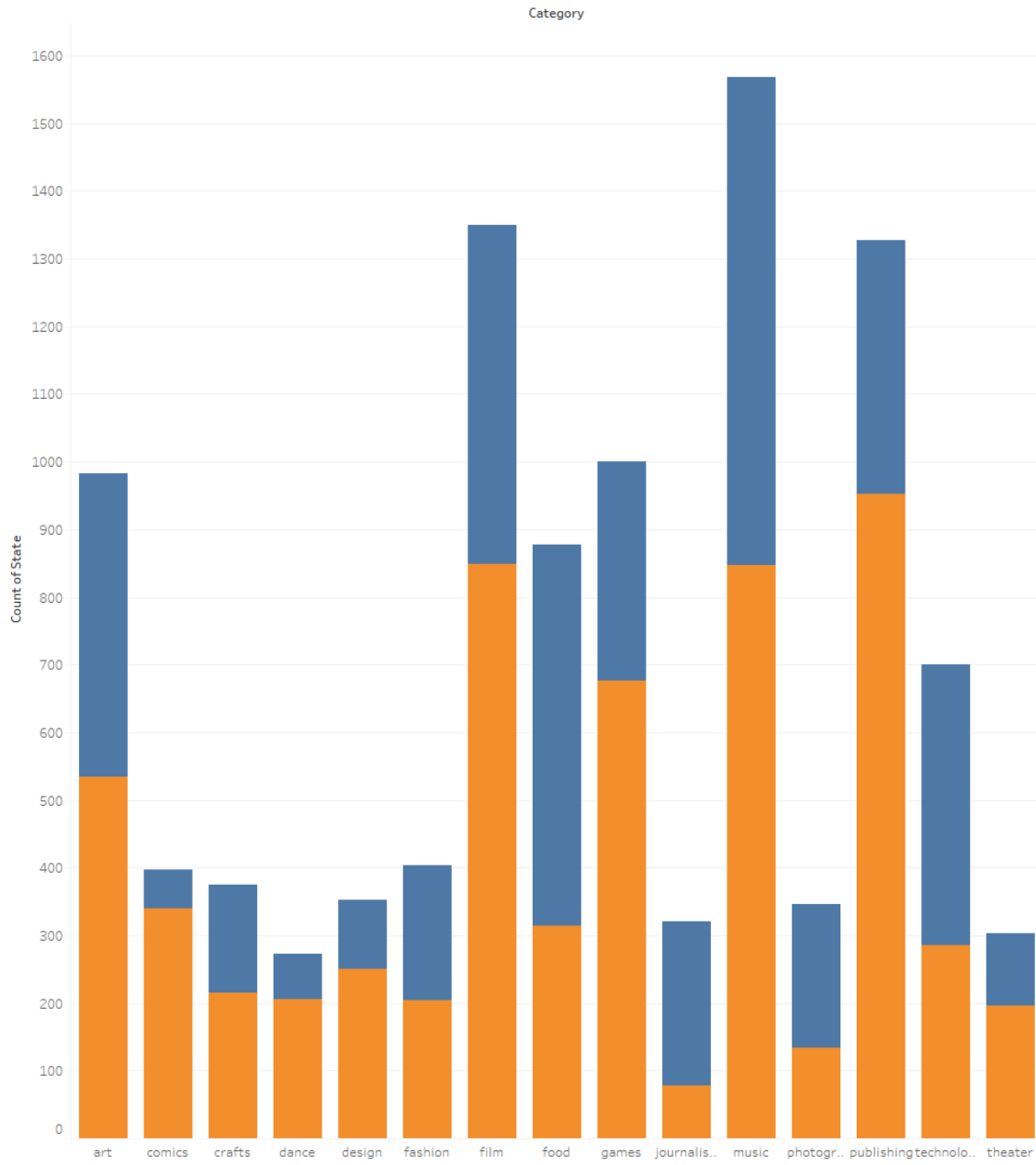
Category



Count of State for each Category. Color shows details about State.

Figure 10: This stacked bar chart shows the proportion of successful and failed projects by category. The orange area signifies successful projects, while the blue area signifies failed projects. This chart adds a bit more detail to the project count of each category.

Name Length Prevalence



The trend of count of Name Length for Name Length. The marks are labeled by average of Name Length.
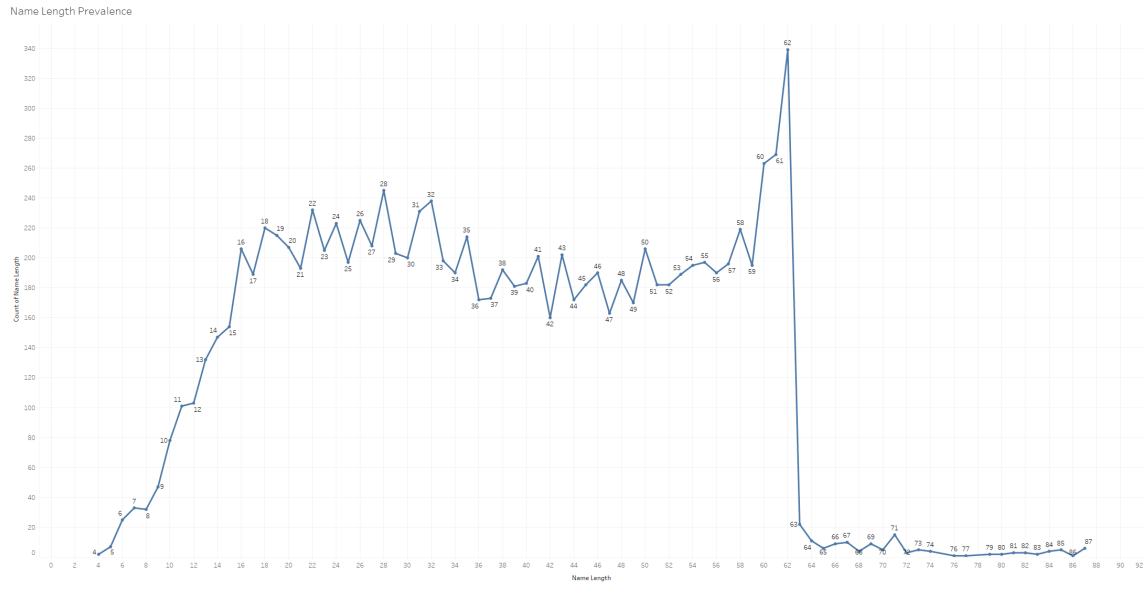
Figure 11: This continuous line graph shows the prevalence of project name lengths by number of characters. The drastic drop from the peak (62 characters) is caused by a revision in Kickstarter policy, allowing longer name lengths since early 2017.
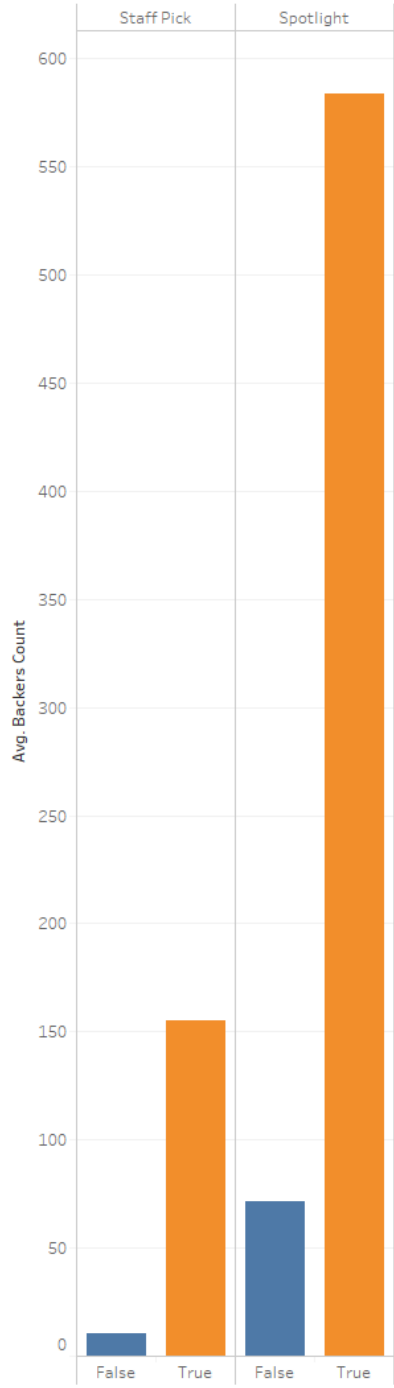
Figure 12: This double bar chart shows the impact having Staff Pick or Spotlight has on the average number of backers a project receives. Both have a significant effect.

Staff Pick/Spotlight vs. Pledge Multiplier
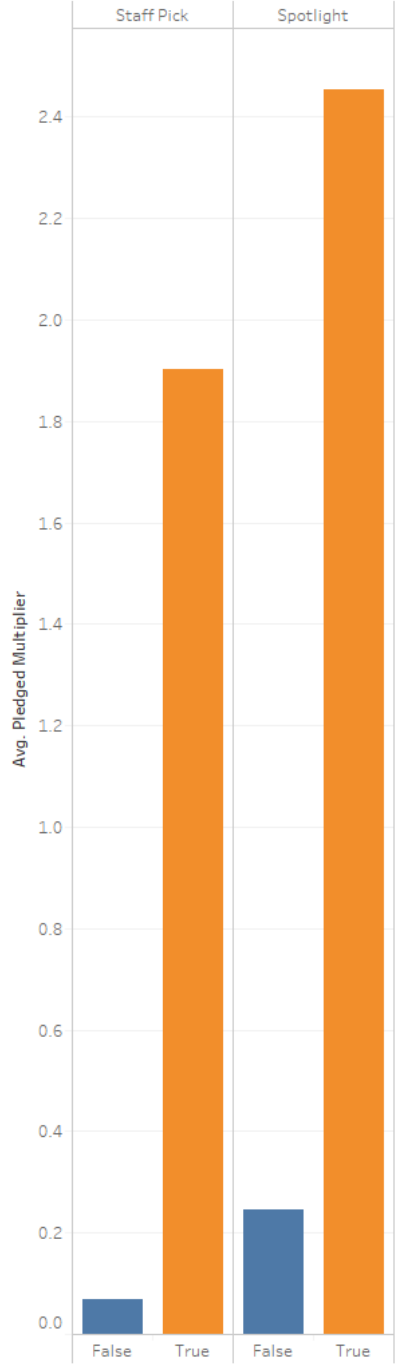
Figure 13: This double bar chart shows the impact having Staff Pick or Spotlight has on the average pledge multiplier a project receives. Both have a significant effect.

Pledged Amount vs Blurb Length
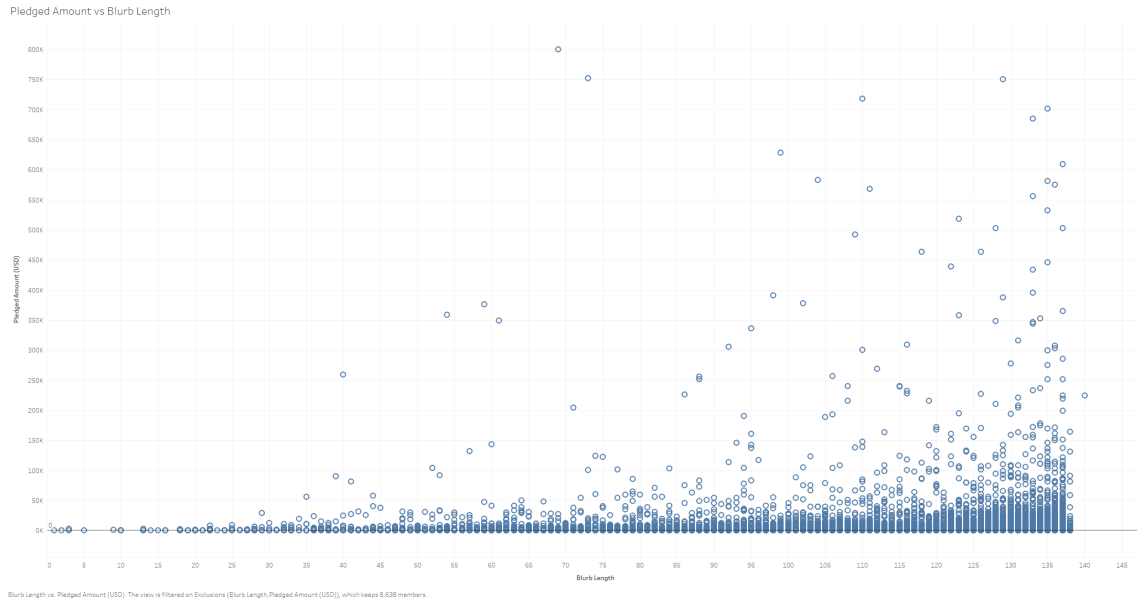
Figure 14: This scatter plot shows the relationship between blurb length (in characters) and the total pledged amount. It can be seen that longer blurbs tend to correlate with higher funds raised, likely because it is an indicator of the effort put into a project. The graph was shortened for easier viewing.
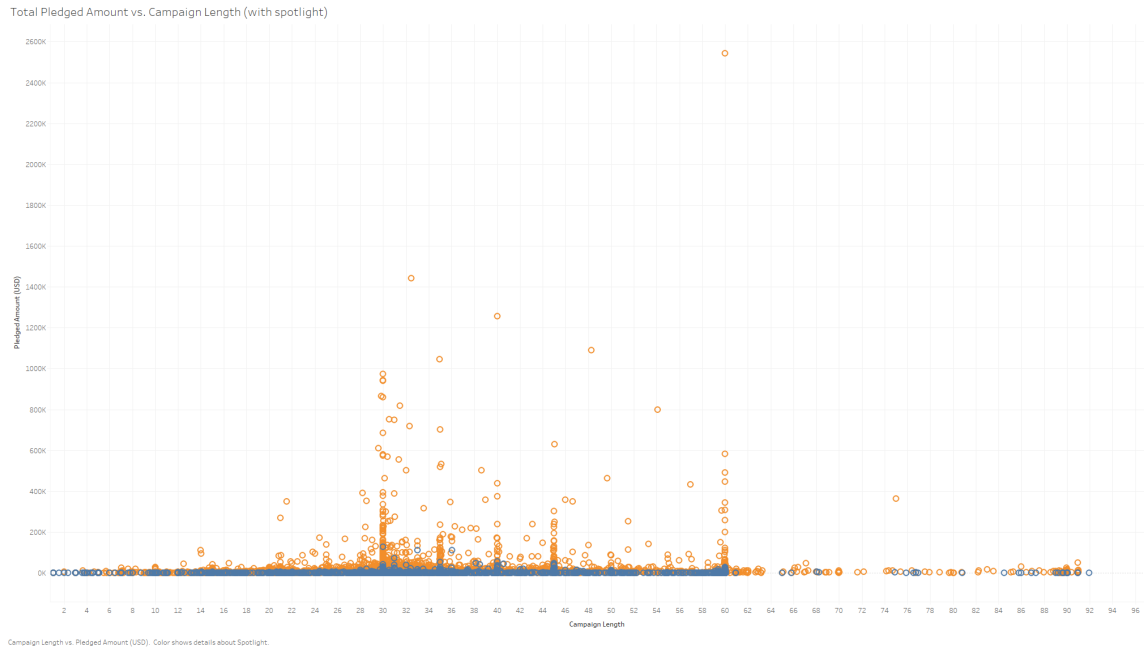
Figure 15: This scatter plot shows the relationship between the total pledged amount and the campaign length. Orange data points indicate that the project had a Spotlight.
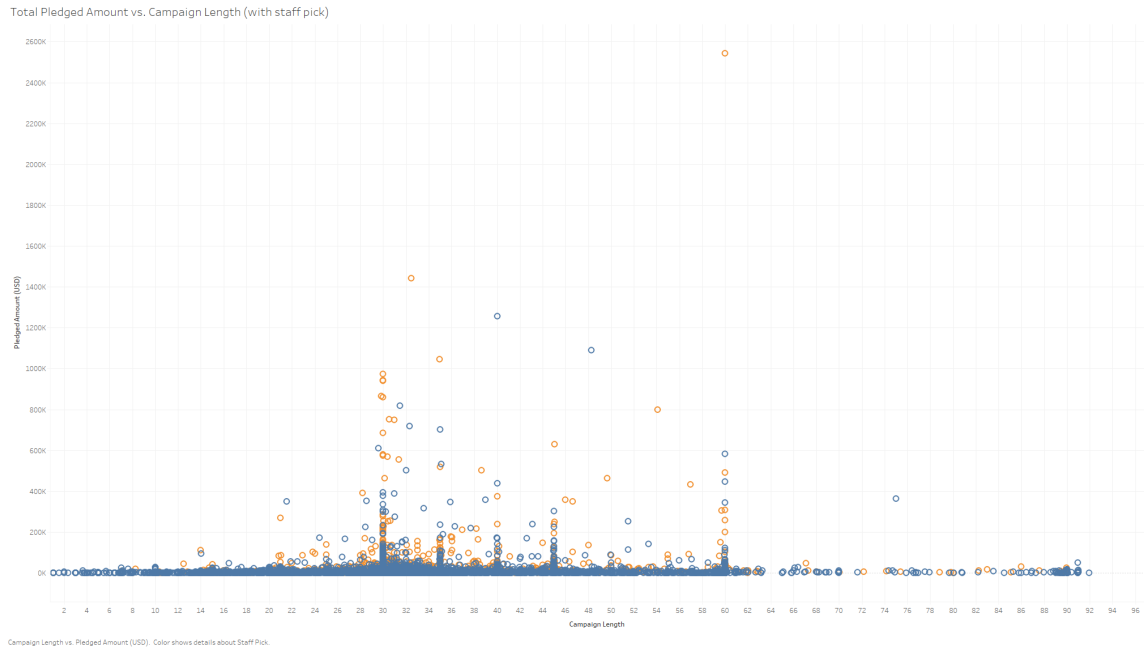
Figure 16: This scatter plot shows the relationship between the total pledged amount and the campaign length. Orange data points indicate that the project had a Staff Pick.
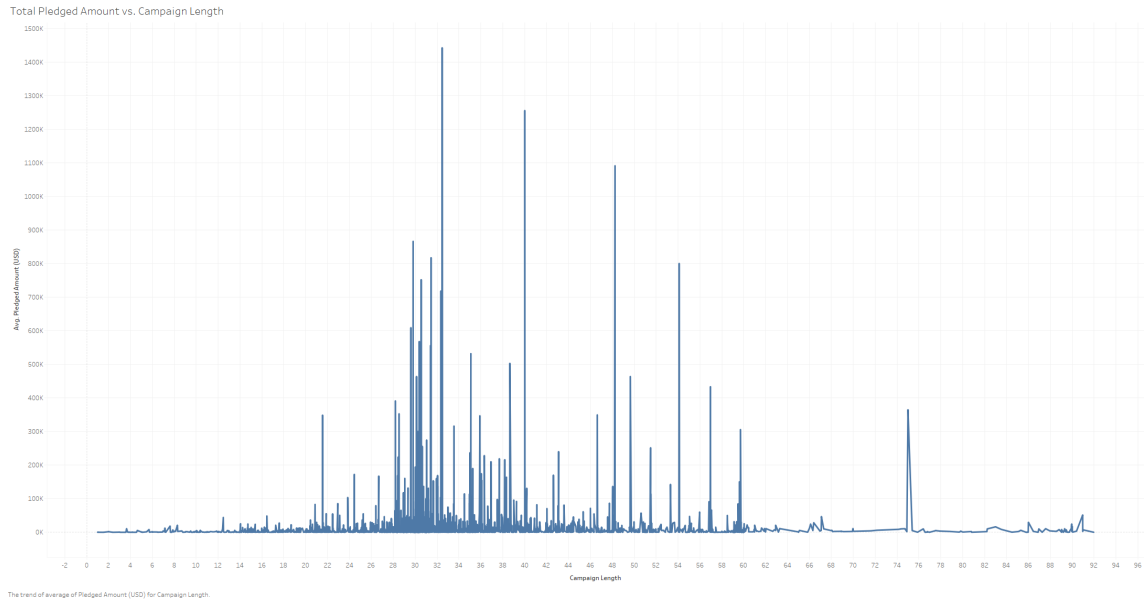
Figure 17: This line graph shows the relationship between the total pledged amount and the campaign length.

## Average Total Funding vs. Category
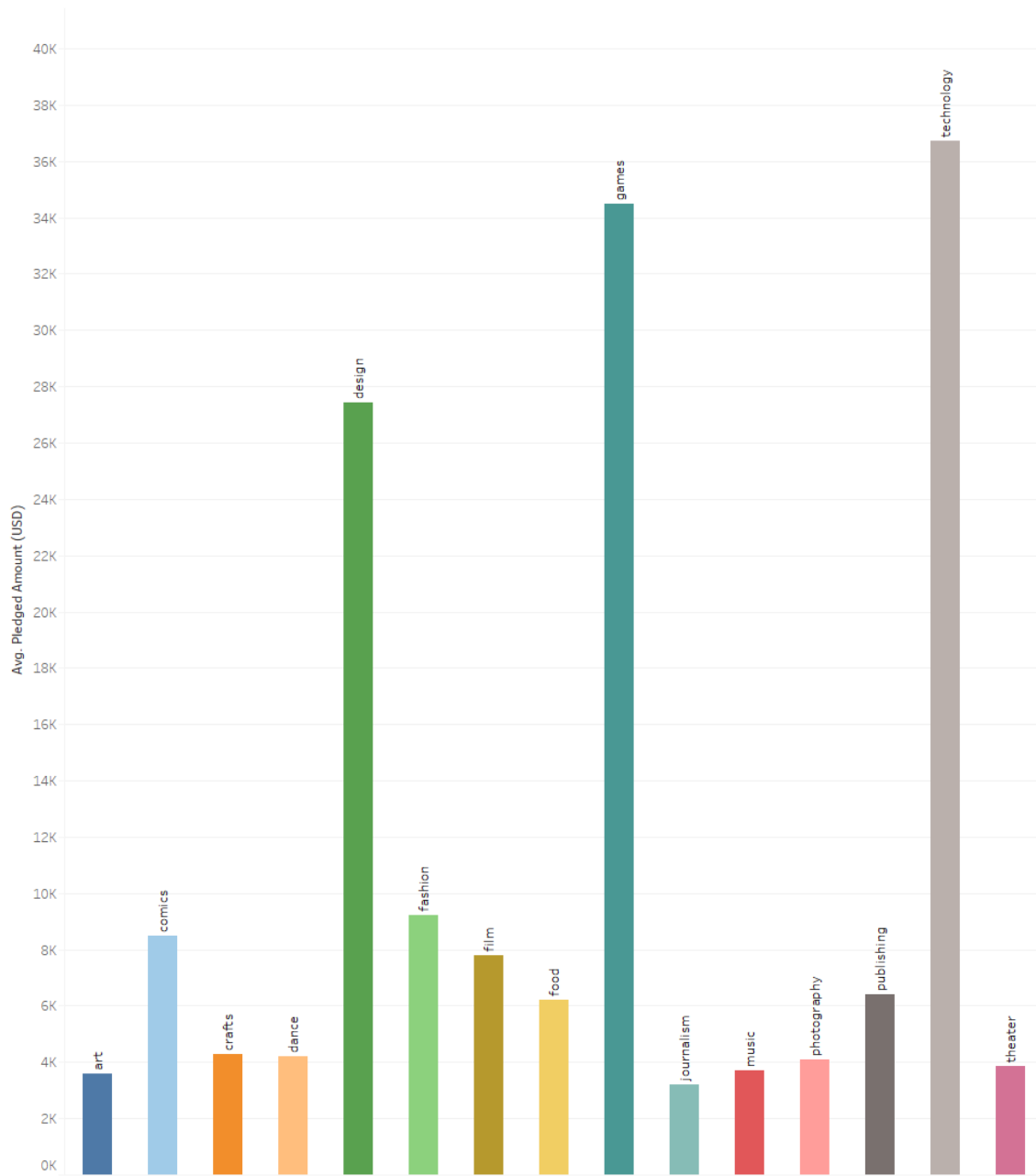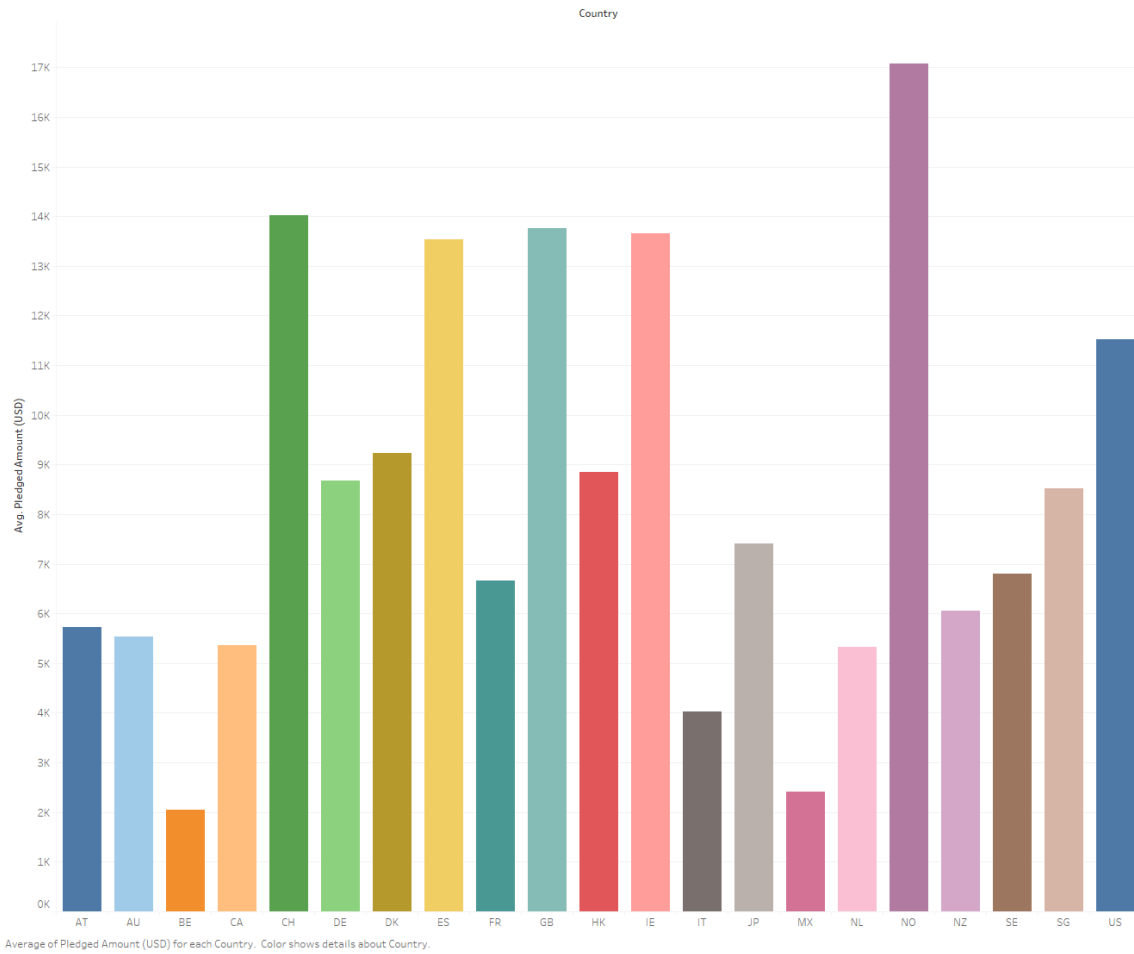


Figure 18: This bar graph shows the average total funding a project receives in relation to its category. The average technology project received over $36000 USD, while the average journalism project received less than $3300 USD.

Figure 19: This bar graph shows the average total funding a project receives in relation to its country of origin. The average Norwegian project received over $17000 USD, while the average Belgian project received just over $2000 USD. These values are likely influenced by the small number of projects from certain countries, leading to skewed results.
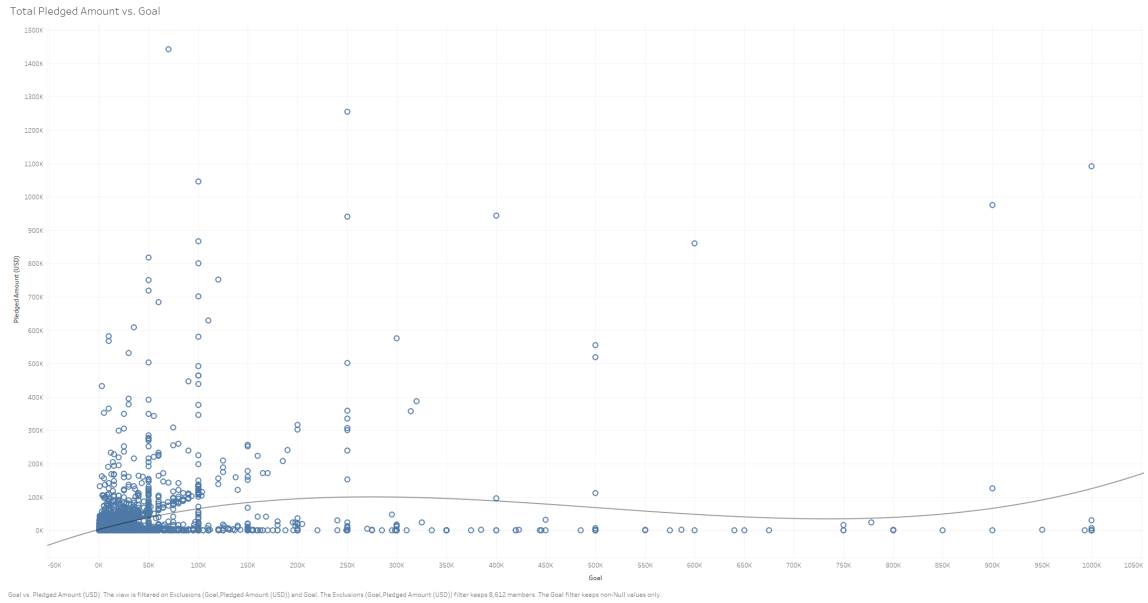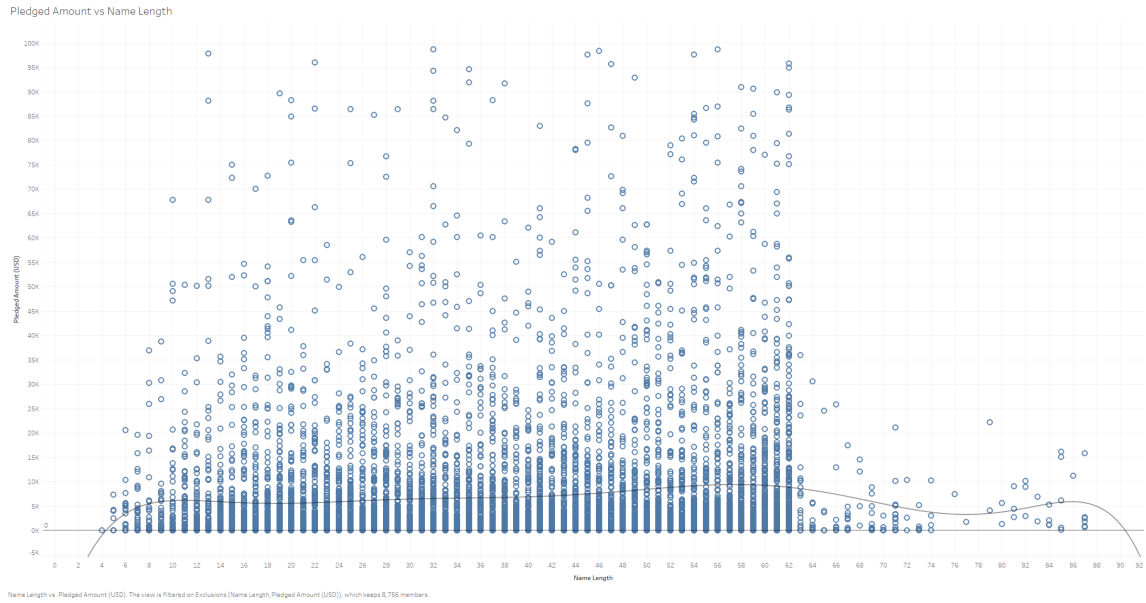
Figure 20: This scatter plot relates the total pledged amount a project receives and its initial goal. A polynomial line of best fit shows that a higher goal is somewhat correlated to higher final funding.

Figure 21: This circle views chart shows the total pledged amount a project receives and its name length in characters. A polynomial line of best fit shows a small correlation between the two values. The rapid drop-off at 63 characters is due to Kickstarter changing its name length policy in early 2017.