Ben Soderberg, Aidan Litt, Joseph Tortorello, and Chase Sonnemaker

Kickstats: What makes a Successful Kickstarter Project?

Motivation

Kickstarter is a funding platform which allows for creators to bring their ideas to reality. It supports a variety of projects, many of which have enjoyed widespread commercial success, including the card game Exploding Kittens and the Pebble Smart Watch. Long before they were even commercially sold, both of these projects had campaigns that not only met their funding amounts, but far exceeded their stretch goals. This is the dream for any entrepreneur who wants to use Kickstarter, but it isn't easily attainable. In the shadow of every high profile project, there are thousands of utter failures that don't even make their funding goals. For projects like The Dual Shower Heads for Two, or the "completely true" *Animated Adventures of Samurai Mary vs Ninja Jews (The Birthing of Christ)* it can seem obvious as to why some of these ridiculous projects didn't get any funding, but often it is not as simple. For example, an annual calendar that helps support cat animal rescues did not make its funding, while a project to help someone make a potato salad for themselves made over $55,000. This begs the question, what makes a successful Kickstarter? It has to be more than just having a killer idea at the right moment. For this project, we wanted to uncover any factors of success behind Kickstarter projects and see if there are any specific measures a project can take to ensure its goals can be met.

Data Collection

Our data was obtained from two csv files posted on Kaggle.com labeled 2018 and 2016. The former had 378,661 entries with 15 categories and the latter 323,750 entries with 12 categories. Upon further inspection, we found that the 2018 dataset contained all of the entries from the 2016 dataset, so the older dataset was discarded. Additional information was planned to be scrapped using a python script which would use the projects name to find the creator's name, the creator's unique ID, if the project was a staff pick, the contents of the project's blurb, and the image URL. This program was only able to return information for 172,941 of our 378,661 entries, but since the information was gathered by searching the randomly generated project ID number, we treated these entries as a sample from our original set for analysis.

Data Cleaning

The dataset had a variety of issues that needed to be fixed before we began our analysis. Some issues involved misprinted or buggy values. One of the first issues we encountered were projects that had launch dates on January 1st, 1970. Since this was long before the founding of Kickstarter, we changed these date entries to NA values. We also found the number of entries for 2018 to be very low when compared to the other years. This was because our dataset was taken from early January of that year and not many projects had been published yet. Since it was misleading to compare this small number of projects to full years,

they were discarded for any analyses comparing years. Some locations for projects also showed up as *N,0"*. We believe this to be a bug from the original scraping program used by the user who uploaded this dataset to Kaggle, and so changed those countries to NA values.

We also faced issues involving unused data and missing data. For example, some projects failed to report if they had made their funding goal, despite the values for their goal and amount pledged being readily available. To fix this, a script was written to manually check if a project had made its goal, and calculate the percent of the goal the project reached when funding closed. Some of the data from certain columns were also found to be redundant and not useful for analysis. For example, there were columns for the amount of funding a project received and another for the goal the project had set for itself, both in the project's native currency. There were already columns that had converted these values to U.S. Dollars so they could be compared, so these redundant columns were deleted from the dataset.

Finally, we had issues within the group about how we should think about the data, and subsequently clean and classify it. We found some projects which had made their funding goals but had been canceled or suspended. Further inspection of these individual projects found that they either contained offensive content or plagiarized materials/products. There was controversy within our group as to whether these projects should be included, however these campaigns did successfully make their goal and the originality of the idea would likely have little effect on the amount of money generated, so these values were not omitted.

By the very nature of crowdfunding, there were many outliers within all of our variables. For example, a band, Vulfpeck, used Kickstarter to organize a "pre-order" of their new album. They set a goal of one dollar and their lowest reward tier was ten dollars for a digital copy of the album. Many of our visualizations purposely omit these numbers so that they can be understood and interpreted better. There are also a myriad of creators who have created multiple projects. Most of these creators have many projects of very similar topics and some are not what most people would think of as a Kickstarter. One of these is a monthly cheese club which only receives funding from a few local participants; this is less of a widespread funding campaign and more just a few friends using the platform to organize around something they want to do. Other repeat creators are established board game companies that use Kickstarter to see if an idea for a board game will sell before they actually invest in the manufacturing process, so their kickstarter is more of a marketing test campaign. This made analysis of repeat creators largely difficult. Combining this with the fact that our scraping was incomplete and would therefore have likely missed some or all of each creator's projects, we mostly dropped the analysis of whether repeat creators have a higher success rate.

Results/Analysis

Figure 1: The Total Amounts of Money Associated with Kickstarter Between 2009 and 2018

To start, lets look at some overarching metrics of our dataset. The combined goal amount for all projects on Kickstarter is about $17 billion, represented the total area of Figure 1. Combined, these projects only generated about $3.4 billion, which already indicates that many projects do not make their funding goals. $3.1 billion of this, however, was invested towards projects which did make their funding goals. For a backer, this shows that investing your money towards a project is usually a safe bet. Only a small portion of these projects however, actually achieved funding over their starting goal. The sheer amount of money associated with Kickstarter shows how important the platform is for us to understand. With $3.1 billion being pledged to new inventions, ideas, performances, events, and products, all which might not have existed without the site, knowing who the money is going to and how others might possibly gain some of this funding for their own ideas would be invaluable.

A. Demographics

Over the course of this project, our group found many unique and interesting projects. Some were funny, others were strange, and yet still some were great ideas. Before we began determining what variables contributed to a successful Kickstarter project, we had to first understand what a typical Kickstarter project looks like.

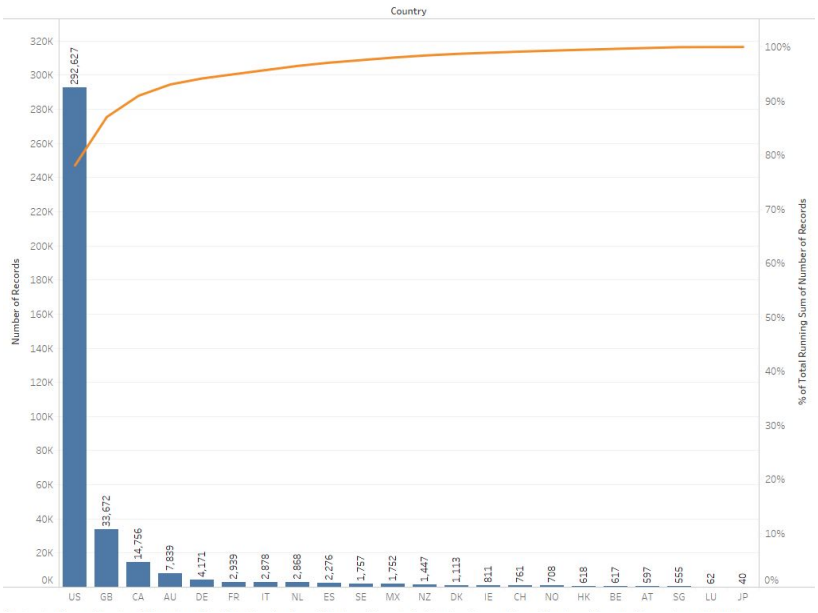

Figure 2: Number of Projects and Percentage of Projects by Country

Beginning with the country of origin, Figure 2 shows that the overwhelming majority of Kickstarter projects in our dataset were launched from the United States of America, with over 290,000 projects. Kickstarter itself is based out of Brookland, New York, so it makes sense that about 80% of the projects released on the website are from the U.S. The top three after the US were the UK, Canada, and Australia. This lead us to the conclusion that most Kickstarter projects originate in english-speaking, Western countries.

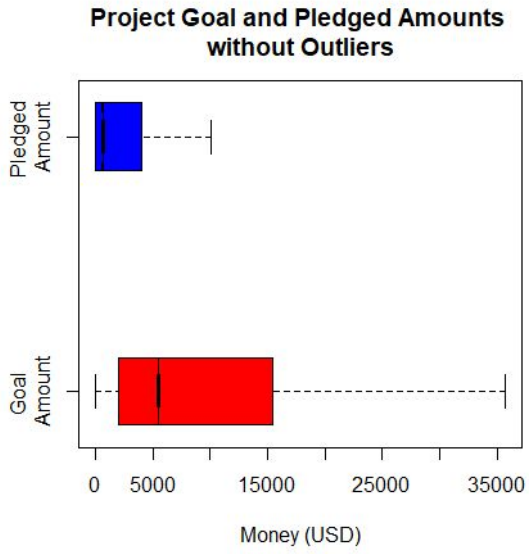

Figure 3: Goal Amount and Pledged Amount of Kickstarter Projects in US Dollars
Outliers were programmatically eliminated from the visualization

The median goal set for all Kickstarter projects is about $5500 with a median pledged of only about a few hundred dollars (Figure 3). This further shows that many projects on the platform fail to receive their funding. All in all, about 36% of projects actually make their allotted goal amount. Interestingly, the number of people who usually support projects also tends to be very low, with about 75% of all projects having 56 backers or less. As mentioned before, there were a considerable number of outliers for the goal and pledged amounts. For example, the maximum goal amount for any project was $166,361,391 while the maximum amount pledged to any one project was $20,338,986, well outside the range of anything from Figure 3. This tells us that while most projects are asking for less than $35,000 and receiving less than $15,000, a select few are receiving and asking for far more, meaning these projects have a strong right skew on these variables. A similar case happens for backers, as there were three projects that had over 100,000 people supporting them.

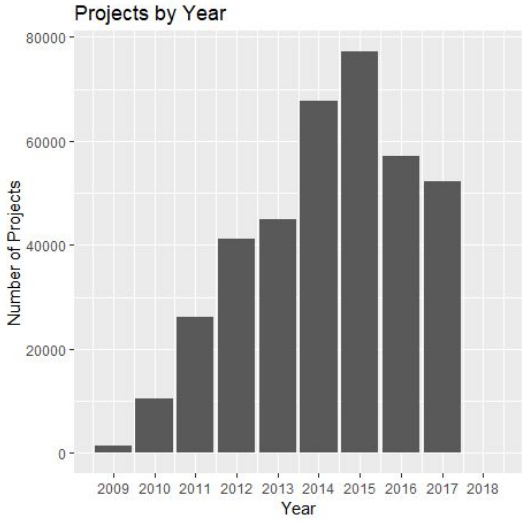

Figure 4: Number of Projects per Year

From the website's launch in 2009, the number of projects has been on the rise, peaking in 2015 at a little less than 80,000 active campaigns (Figure 4). But, after 2015 there was a steep die off to about 60,000 and then 50,000 projects. In further analyses, we will investigate the cause of this dip; however, it is important to understand that the majority of the projects in this dataset will be from between 2014 and 2017.

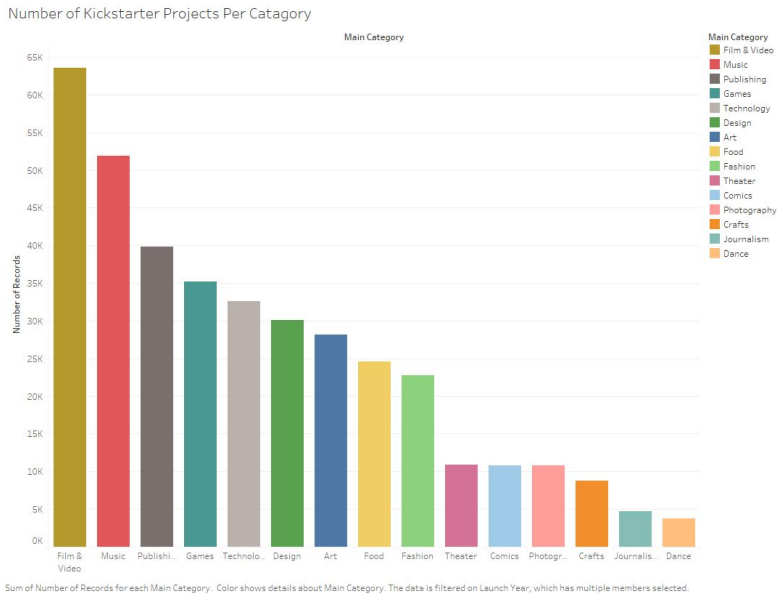

Figure 5: Number of Projects by Main Category

Kickstarter Projects are divided into 15 main categories. Of these, most projects are posted as Film & Video, with about 64,000 entries (Figure 5). This is followed by Music and Publishing, with 54,000 and 40,000 projects, respectively. This means that combined, over 60% of the projects in our dataset were concerned with some form of media entertainment. The lowest number of projects are Crafts (10,000 entries), Journalism (7,000 entries) and Dance (5,000 entries).

B. Main Category

After understanding what the typical Kickstarter project might look like, we moved on to understanding what factors might contribute to a project being successful. The first variable we looked at was the main category of the project. To us, it would make sense if some categories were more popular with backers and therefore generated more funding making those projects more likely to succeed.

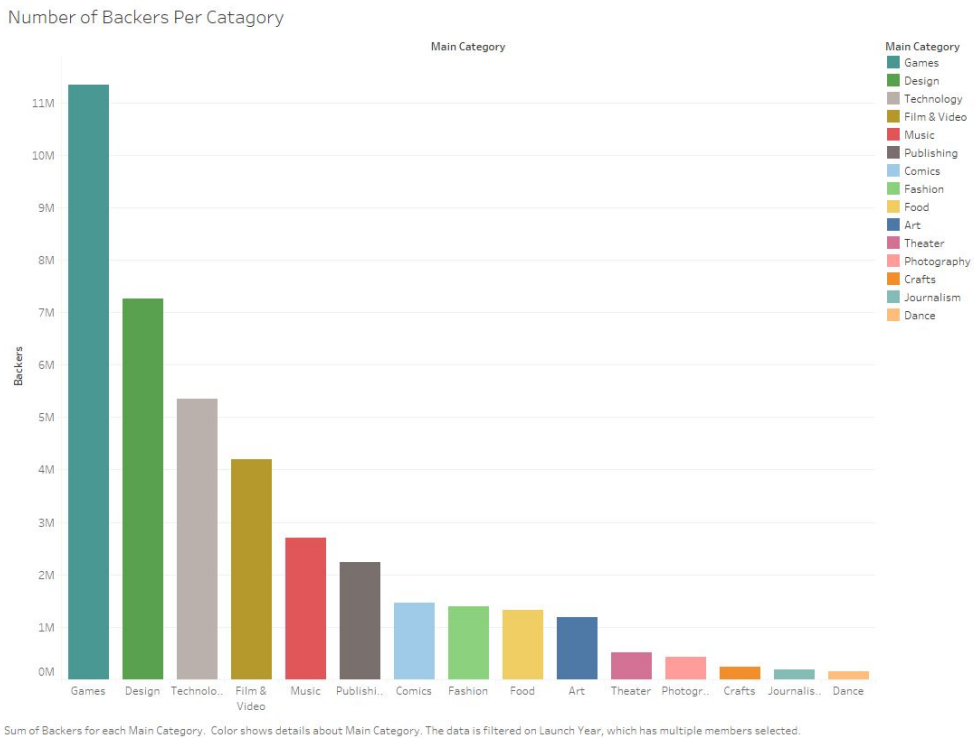

Figure 6: Number of Backers per Category

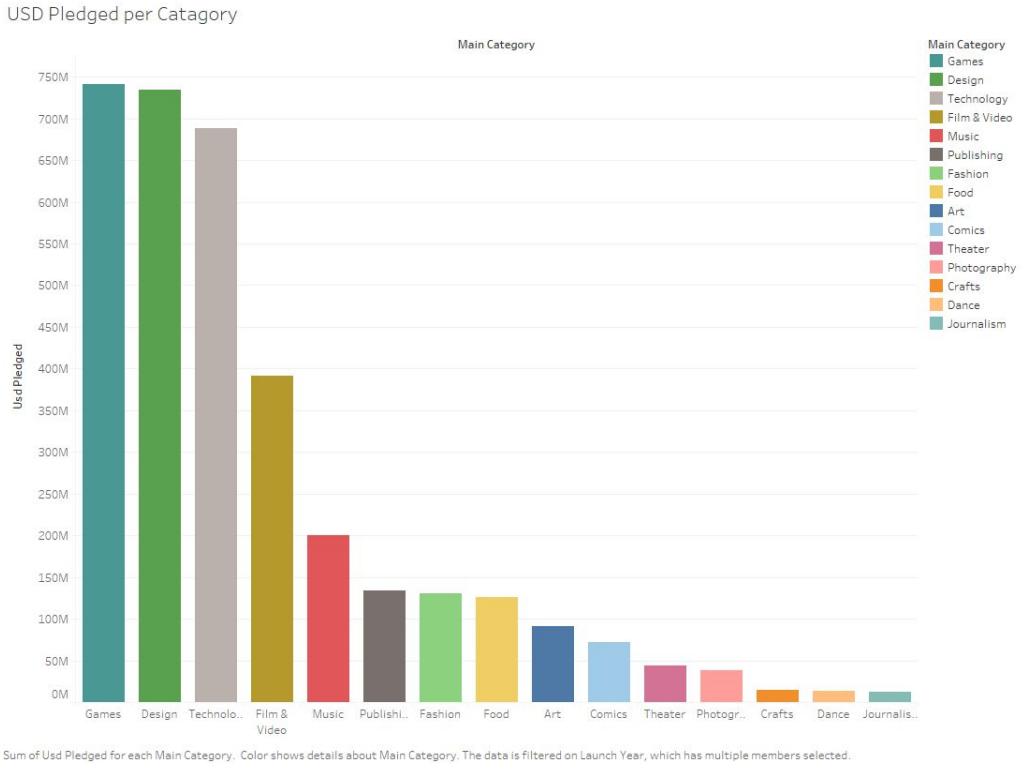

Figure 7: Amount Pledged by Category in USD

Unlike number of projects, the categories with the most backers are Games, Design and Technology (Figure 6). These same three categories also make the most amount of money (Figure 7). Of note, however, is that despite the large differences between the number of backers in the top three categories, all three made between 750 to 700 million dollars. This suggests that Games, as a whole, is receiving their funding from a larger amount of small donations while Design and Technology are making more money on individual projects.

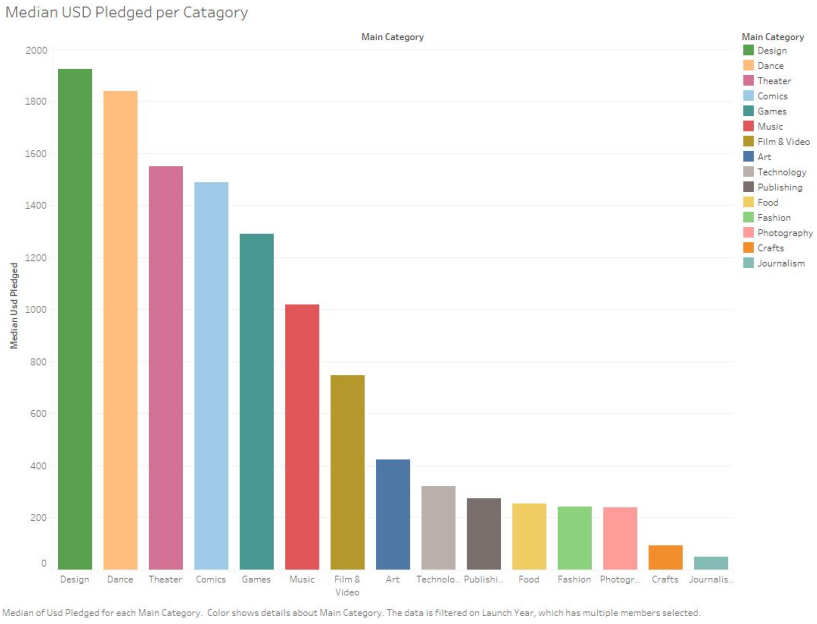

Figure 8: Median USD pledged per Category

For Design, this does appear to be the case, as it has the highest median amount pledged for any category (Figure 8). As expected, Games also has a lower median pledged amount. However, Technology has the lowest median pledged for any of these three values. This suggests that much of the revenue from Technology projects is only being generated by a few massive outliers. The second highest median, interestingly, is actually Dance, which was previously shown to have the lowest number of backers (Figure 6) and the second lowest amount generated (Figure 7). Despite having a much lower number of backers and revenue being invested into this category, Dance projects have a higher amount of money being pledged into them individually.

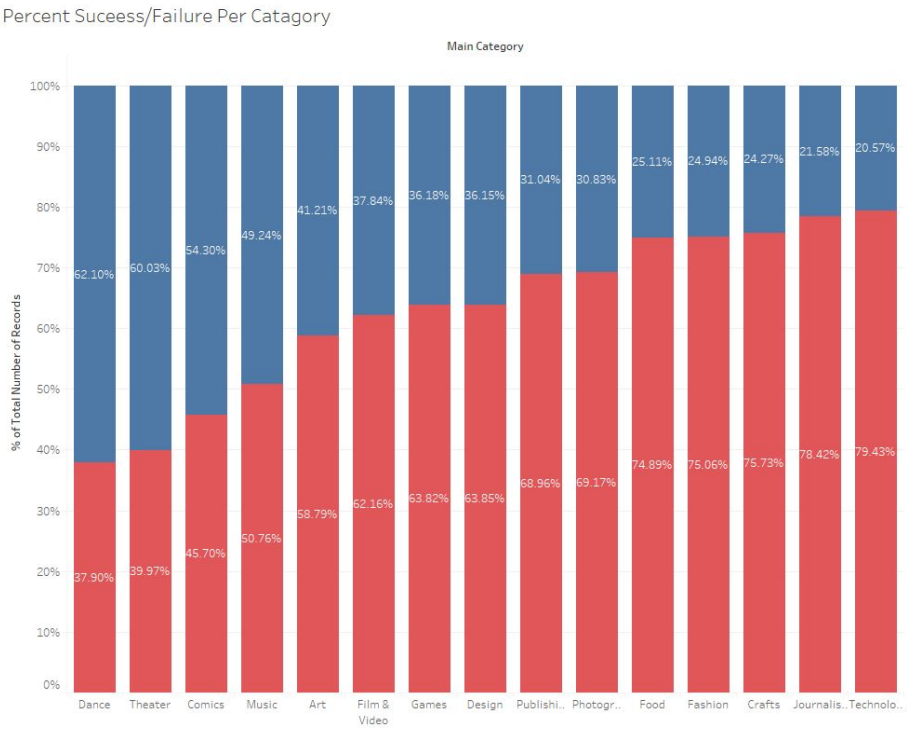

Figure 8: Percent Failure/Success by Category: Red bars represent percent failed, blue represent percent succeeded.

Dance also has a success rate of about 62%, which is much higher than the 36% success rate for all Kickstarters combined (Figure 8). Design and Games, on the other hand, have a success rate of almost exactly 36%, matching the average. Technology has an abysmal 20%. This further solidifies that the revenue being made Technology projects is mostly though well performing outliers, when in reality most projects are failing to even meet their minimal values.

So, in regards to an individual project, it seems like Dance is the most successful. It will not attract a huge number of backers, but it has a high success rate and can get away with asking for a fair amount of money as its goal. From this line of thinking, it seems reasonable to say that Dance's success could come from the lower number of projects, as there is less competition than in a saturated field like Technology. However, Journalism and Crafts, two categories that had a similarly low number of projects (Figure 5), have the lowest median revenues (Figure 8), have a low number of backers (Figure 6), and also have the highest failure rates after Technology (Figure 8). The categories with the highest numbers of projects, (Film & Video, Music, and Publishing) also do not perform either exceptionally well or exceptionally poorly. So, it seems that the number of projects being made in each category does not have much of an effect on an individual project's performance.

C. Country of Origin

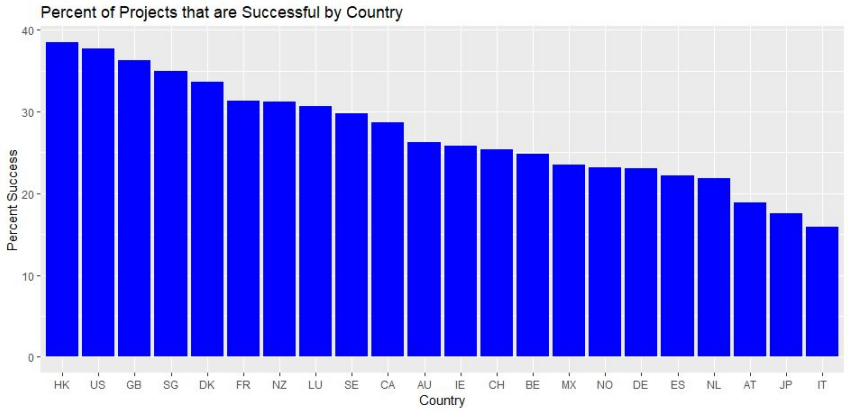

Figure 9: Project Success Rates by Country

After exploring category, we turned to the launch country. We hypothesized that because the site mostly caters to an english-speaking audience, countries like the US and UK would probably have a distinct advantage in making more successful projects. Interestingly though, the United States doesn't actually have the highest success rate out of any country. Instead, that would be Hong Kong, with a rate of about 38% (Figure 9). The other top countries, aside from the US and Hong Kong, are Great Britain and Singapore. The bottom three are Austria, Japan, and Italy. We are unsure as to why these countries in particular are experiencing these successes and failures compared to the others at this current time. We think it may have something to do with the technical prowess of the top countries, since many of them are leaders in technology development. We also attempted to explore categories, thinking that top countries may be contributing more projects from more successful categories compared to bottom countries. However, the top and bottom countries had roughly the same top categories so we ruled this out. Currently, we wish to further understand why we see the trend in Figure 9 and would like to continue this in future iterations of the project.

D. Goal Amount

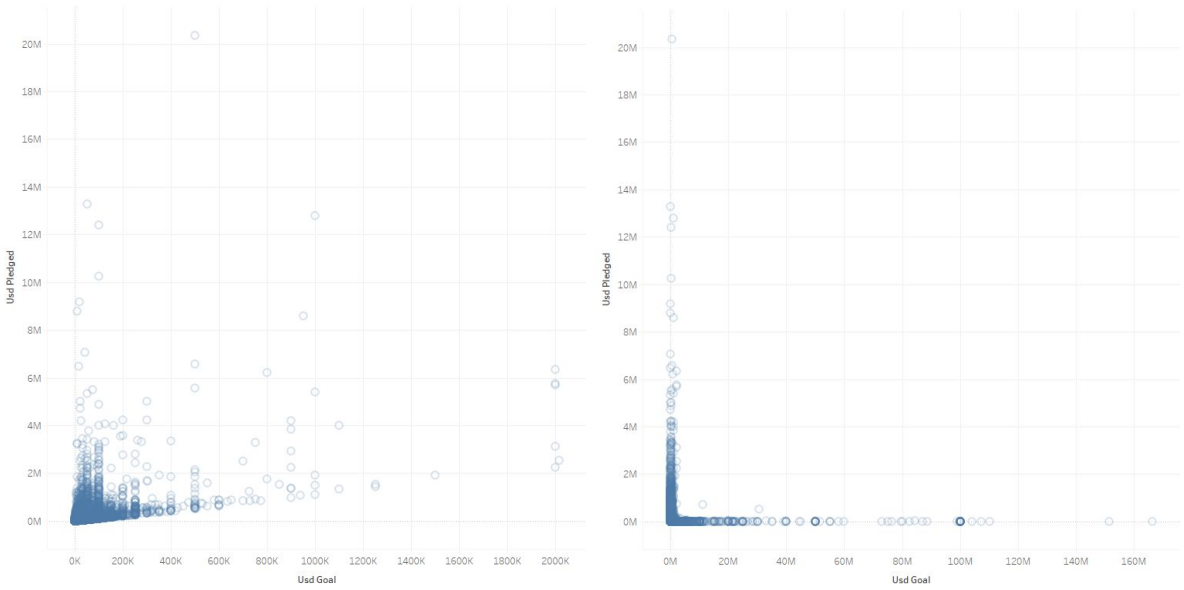

Figure 10: Project Goal vs Pledged Amount for Successful (Left) and Failed Projects (Right). Note the changes in scale between the two graphs

We next looked into the goals set by each project knowing that a lower goal should theoretically be easier to achieve making success easier. We have found that projects appear to be affected by the amount of money they ask for. Figure 10, while difficult to read, is notable due to the difference in the scale of the goal set. The maximum goal of any successful project was approximately $2,000,000 and even then these successful projects with high goals are few and far between. Meanwhile, the scale on the failed scatterplot has several projects requesting more than $2,000,000 all failing to reach their goal.

We attempted to investigate the relationship between the goal of the project and whether it succeeds through a logistic plot. However, due to the massive scale of goals and the lack of successful projects with set goals higher than $2,000,000, our model faced quasi-perfect separation. This means that successful and failed projects were separated by goal amount so well that certain goal values predicted a 0% chance of success.

So, while we do have difficulty visualizing the relationship between goal amount and whether a project succeeds, we do know that projects with higher goals, especially outrageously high goals, are less likely to be successful, while projects with lower more realistic goals are more likely to succeed. We are not sure how accurate this relationship is at lower goal amounts, but it is certainly true for larger goals.

E. Words

An important method for attracting funding towards a specific project is the project's blurb. Figure 11 contains the most commonly used words for all kickstarter projects. The largest are the words "your" and "help", which makes sense since most projects directly ask for the help of their backers to make their funding goals.

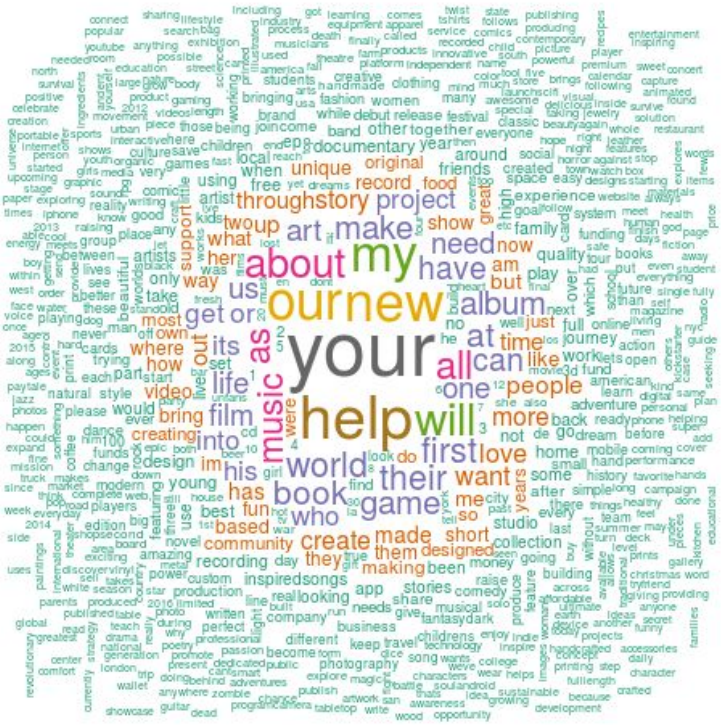

Figure 11: Most Common Words used in Kickstarter Blurbs

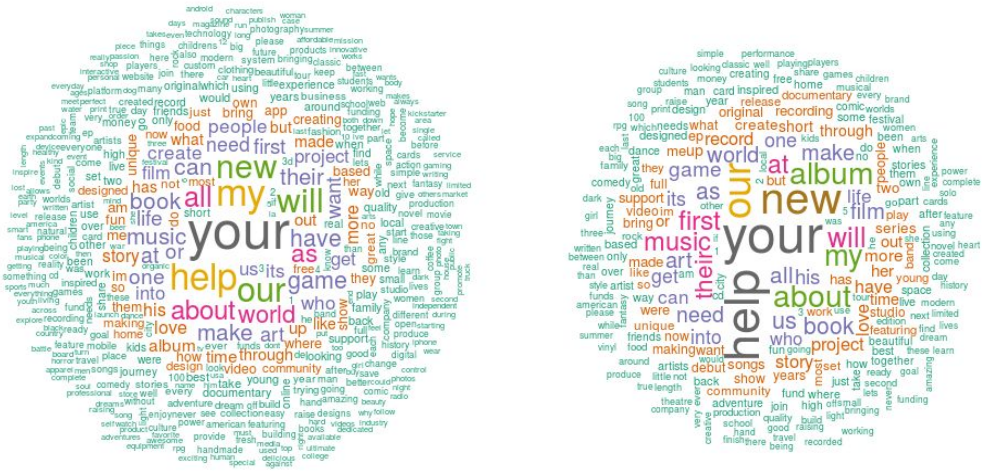

Figure 12: Most Common Words in Failed (Left) and Successful Projects (Right)

For the most part, the same words appear in both failed and successful projects, with some notable exceptions. Your is still the largest word between the two, however successful projects made more use of “help”, “new”, and “our”, words that are associated more with cooperation (Figure 12). Failed projects have “my” as one of the larger words, which might seem more selfish. This analysis suggests that projects with blurbs that are more polite and cooperative are likely more successful compared to those that come across as self-centered.

F. Staff Pick

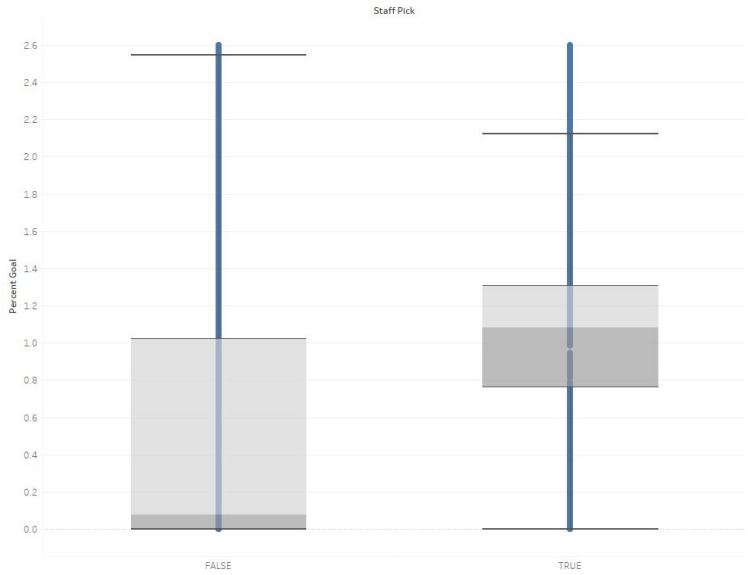

Figure 13: Staff Pick (Left) and Non-Staff Pick (Right) vs Percent Goal.

A Kickstarter Staff Pick is a project that is selected by Kickstarter to be featured on the main page. We hypothesized that this might have an effect on the success rate of a project, since the main page would spread awareness of the project to more potential backers. Staff Pick also appears to have an effect on the success of Kickstarter projects. The median percent of goal values for projects that were staff picks is over 100.0%, meaning that more than half of those projects actually make their funding goals, while the median for non-staff picks is less than 20% of their goal (Figure 13). This would suggest that a project being a staff pick is associated with a higher rate of success, but it is difficult to determine if the reason these projects are succeeding is because they were staff picks or were already going to be successful regardless and were therefore chosen to be a staff pick. In reality, both cases are likely true, where a project starts becoming successful and is then chosen as a staff pick, allowing it to gain even more funding.

G. Time of Launch

As we previously mentioned, Figure 4 shows a distinct and rapid rise and fall in Kickstarter projects between the years of 2013 and 2017. Since the year of the projects release may be associated with how successful a project is and because the trend was unexpected and interesting, we wanted to investigate further. We began by splitting the projects by whether they failed or succeeded in reaching their goal.

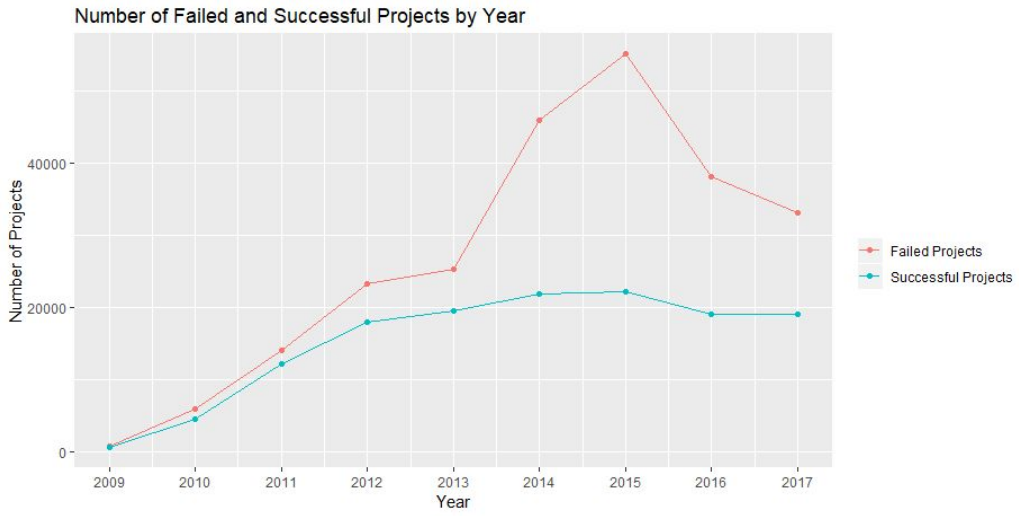

Figure 14: Number of Failed and Successful Projects by Year

From Figure 14 we discovered that between 2009 and 2013, the number of failed and successful projects grew by roughly the same amount; however, from 2013 to 2015 there is a massive rise in the number of failed projects without much of a rise in the number of successful projects. After 2015, the number of failed projects falls while the number of successful projects continues to stay comparatively steady. So, the large uptick and subsequent fall of projects was primarily made up of failed projects. So, why the sudden change in the number of failed projects?

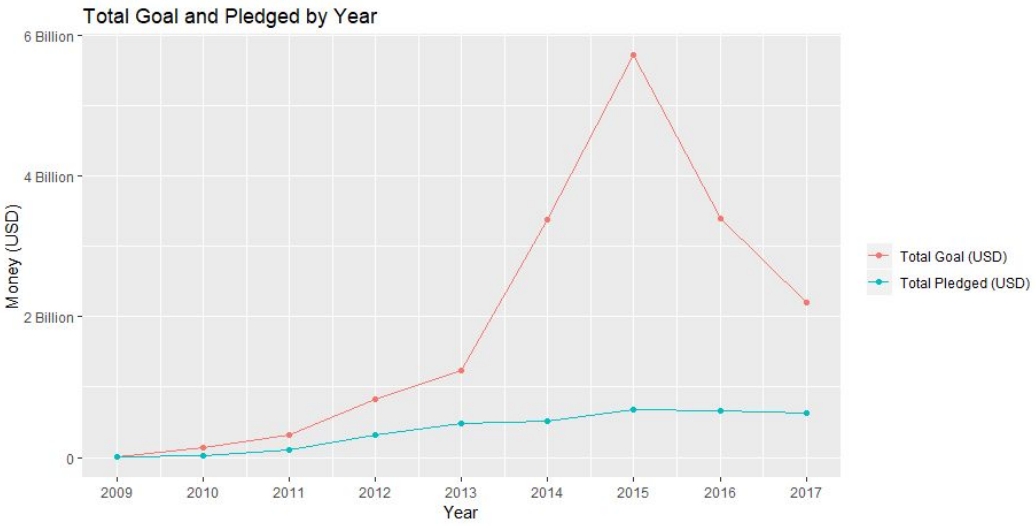

Figure 15: Total Goal and Pledged Amount of All Projects by Year

Moving on to looking at the finances behind the changes, we noticed that the total amount project creators were asking for rose steeply from 2013 to 2015, likely due to the large increase in the number of projects, mentioned above. Unfortunately for the project creators of 2013 through 2015, this increase in demand for money was not met by an increase in money pledged by backers, leading to the large gap seen above in Figure 15. Like before, after 2015, this gap decreases substantially.

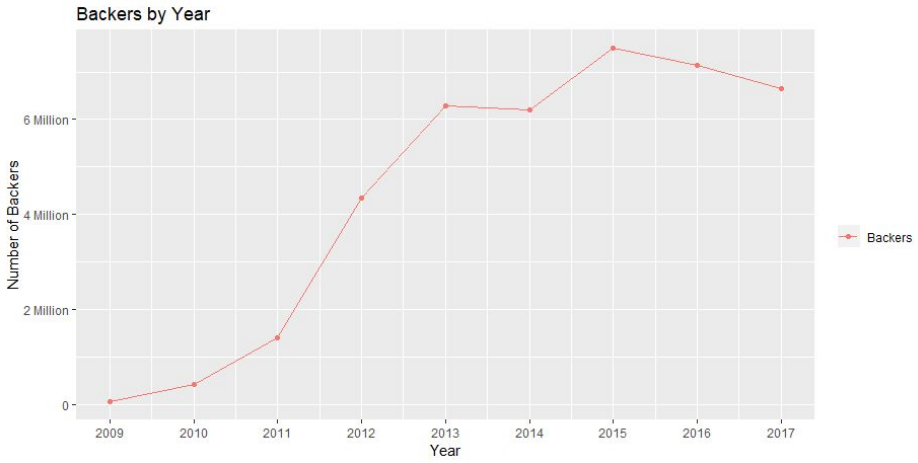

Figure 16: Total Number of Backers by Year

Our theory for the series of changes occuring from 2013 to 2017 is related to what we have dubbed as the “carrying capacity” of Kickstarter. We hypothesize that Kickstarter experienced greater than usual growth in 2013 and 2014, with more projects launched than ever before. This would be in time with several famous Kickstarter events including the Potato Salad project in 2014 and the rise in popularity of certain YouTube channels making fun of silly Kickstarter projects. However, Figure 16 shows that this rise in the number of projects was not

followed by a rise in backers. We believe this is because Kickstarter had reached a limit. People were simply not willing to invest any more money than they had already been investing. This would explain the sudden rise in total goal amount and relative flatline of total pledged amount. As a result, the number of failed projects skyrockets and the number of successful projects remains relatively the same.

While time is not a factor we have control over, one thing we have learned is that timing is associated with whether or not a Kickstarter project fails. The years 2014, 2015, and 2016, clearly have lower success rates compared to previous years, meaning that launching a project at those times would likely put you at a disadvantage at reaching your funding goal. The good news is that since 2015 the number of projects has been falling back to a reasonable level and most of this fall is due to fewer failed projects. This means that success rates will be back on the rise, so the best time to launch a project may be just down the road.

Model

After examining all of the variables possibly associated with the success of a project, we moved on to consolidating our discoveries into something that we could use to predict the outcome of a project. We decided to create a linear model and we used variables that creators might have some level of control over. In the end, we settled on using goal amount, the category of the project, and whether or not the project was a staff pick. Since the model would be digitally accessible, we were comfortable making a slightly more complicated model as we had to use an indicator variable for each category. The model is extremely inaccurate with an R^2 of approximately 0.01; however, with more time we could gather more data using scraping programs to make improvements to the model. Creating the model did provide us with valuable practice creating Shiny Apps and improved our understanding of R, so its creation was not completely invaluable.

Conclusion

So, what is the ideal Kickstarter project? Well, according to our analysis many different variables are associated with successful projects. First, category. Categories like Dance, Comics, and Theater, while not having the most total money pledged or the highest number of backers, have the highest success rates. This means for an individual creator, these three categories are the best ones to try from a statistical standpoint.

In addition to category, we country is associated with successful projects. While we are still having difficulty explaining why, we know that Hong Kong, the US, Singapore, and the UK have higher success rates compared to other countries While it is difficult to control the country of origin for a project, theoretically one of these should be the country of origin for our ideal project. Another difficult-to-control variable is the year of launch. If year was easily manipulated, the best year to launch our ideal project would be between 2009 and 2012; however, our analysis also suggests that years in the near future may also have relatively high success rates. Important to note is that the year matters less than the market saturation, so launching the project when total projects is overall down is key.