Analysis of the success of various Kickstarter campaigns

MSBX 5420 Project Report: Group 9

Authors: Bella Franz, Danielle Allen, Morgan Likens, Rachel Raifsnider, Samantha Fildish

I. Abstract

Kickstarter provides a way to make dreams or ideas a reality for those who wish to see their ideas come to life but may not have the financial means to do so themselves. When launching a Kickstarter campaign, a creator has the ability to share their ideas and goals with the world. At this time there have been 219,295 successfully funded projects on Kickstarter spanning many different categories and $6,570,743,435 USD pledged [1]. With the many numbers of projects and categories available, we decided to try to find the most successful categories over time and the overall trend of the categories. With this information, we will gather insights into which project categories and main categories may be the most beneficial to invest in.

II. The background and motivation of your questions

When we first started looking at our dataset, some immediate questions popped out based on what was included in the file. While Kickstarter does have a page of basic analytics that could be replicated by our dataset, we wanted to establish a couple key subsets to explore in depth. No one column stood out as being more important for our analysis, so we tried to utilize something from each to give our questions a broad enough scope. We wanted to maintain a balance of being curious about what was in our data as well as try to bring in business value with our research. While we focused less on the individual projects that are represented, there was increased interest in what the general displays of success included. We also had a focus on how time affected success rates because we did have a range of years to observe trends.

If someone was going to start a new Kickstarter project tomorrow, what are some of the topics to consider for a successful campaign? We looked at several factors including average funding, time frame of the launch and deadline, and geographic data to see what, if anything, would stand out as influences that could impact the success of a project. Would any of these features make or break the likelihood of success if a project was fully funded to the goal amount or had enough backers? These questions were interesting ways to investigate trends around the specifics of our dataset that would hopefully offer more insights.

With the data being split out into a main category and subcategory, there was a lot of opportunity to review which was the most successful category overall, as well as see how the success rate was reflected within their subcategories as well. Did the same correlation of success exist between all the subcategories of a category? After looking at the categories and comparing them against one another, we turned our attention to looking at how success rates within the individual category had trended over the years. We wanted to see if they would reflect things like cultural shifts or changes of popularity in different industries. Would successful categories maintain a general level of success?

III. The dataset and analysis methods

Our dataset was pulled from Kaggle and sourced from Kickstarter. It presents 378,661 rows of data that represent different project parameters for campaigns that were posted and tracked on Kickstarter. The projects range from the years 1970 to 2018 (generally the 1970 data we treated as outliers) and represented a range in the "Goal" amount from $0.01 to $100,000,000. We did end up completing a couple preprocessing steps with the dataset after working with it and realizing specific errors that came up. First, we updated the formatting of the

“Launched” and “Deadline” column so that they were all the same between projects. The columns were formatted into a standard YYYY-MM-DD format so they could be read into the DateType column correctly. Second, we removed the commas that were included in the “Name” column so that the text line would not be split into the next column after the comma. It created an offset to all the columns below it, and therefore the columns stopped being consistent after the shift. Overall, we really enjoyed working with the Kickstarter dataset. If we were looking to further expand our data and findings, it would be interesting to look into how the pledge rewards factor into the success of campaigns as well.

After first working with the dataset in Jupyter, we realized that in order to accurately read our data we would need to create a custom schema, so we could correctly identify and work with the different data types needed. Initially, the csv file was being read into the notebook with all columns being converted to StringType columns. This was incorrect based on multiple column types that should have also been included like IntegerType, DateType, or FloatType. After reading in our dataset with the custom schema, we also created our one temporary view for the SQL queries to further prepare for our analysis methods.

The majority of our analysis was done with Spark DataFrame operations as well as Spark SQL. Because we tried to look at the data from multiple angles, we found that we relied heavily on joins. This was helpful so we could present concise outputs based on our business questions without extra information. We also tried to bring in several different operations that used a variety of both basic and more advanced functions. Based on the question we were trying to answer, we spent a lot of time getting the initial format of our data set up correctly with Functions like groupBy, filter, and orderBy. Some of the main operations we used in conjunction relied on Aggregate Functions like count and average, so we could compare data across different groupings. One of our main visualizations from the dataset relied on multiple iterations of Pivot Tables to view each category's success rate year over year and the subsequent trends. We made these more complicated outputs easier to read with the use of Pandas DataFrame.

IV. The results and insights you have obtained

One of the first sections of analysis we completed was to examine the overall success rate of each of the main categories and categories within our dataset. Prior to walking through the results of this analysis, we should clarify how we determined the success rate. We used the following formula to calculate a success rate for each of the main categories; Total Number of Projects per main category divided by Total Number of Projects in Dataset. We used a similar formula to calculate a success rate for each of the categories; Total Number of Projects per category divided by Total Number of Projects in Dataset.

Initially, we completed this analysis on the entire dataset, even though the dataset contained potential outliers in terms of the number of overall projects within each category. With the entire dataset intact, we identified the most successful main category to be Dance, with a 62% success rate. Dance had listed 2,338 projects out of a total of 3,768 projects as "successful". We also found the least successful category to be Technology, with just a success rate of 19%. Technology listed only 6,434 of its 32,569 projects as successful.

After completing the analysis of the main categories, we switched our focus to the Categories within the main categories. The most successful category was "Chiptune" from the main category of Music. This category listed 27 of its 35 projects as successful, giving it a success rate of 77%. The least successful category being Apps from the main category of Technology, with 378 out of 6345 projects marked as successful.

We then decided the next step in our analysis would be to remove the categories that had a small number of projects from the dataset and rerun our analysis. I chose an arbitrary lower limit of 500 projects and removed any category that had less than 500 projects. After removing the data and rerunning the analysis, it was determined that the category that had the highest success rate was Dance, also under the main category of Dance, with a 66% success rate. As we saw earlier in our analysis, the main category of Dance was also the most successful main category, so dropping some of those categories with a low number of projects brought our main category and category analysis results more in line. However, when we re-examine the least successful category, we show that it is still Apps from the main category of Technology, with nearly a 6% success rate.

After we determined the overall most and least successful main categories and categories, we decided to take a look at the success rate of a single category, Design, over the years. When examining Design from 2009, the first year with any successful projects, to 2017, the last year with successful projects, we see that the best year for Design, in terms of success rate, was 2010 with a success rate of 41.5%. Overall, we see that on average the success rate stays within a range of 30% to 40% making this main category's success rate very consistent over the years.

Additionally, we were interested in digging into the data to see if launch and deadline month impacted the success rates of projects. To analyze this, we looked at the number of successful projects per category over the count of total projects per category. We found that the timing does have a small impact on the success of a project. When looking at launch months March has the highest rate of success but is not dramatically different from the success rates of the top 11 months, which all ranged from 32% to 38%. December has the lowest rate of success with a drop of about 8% from March. This is most likely due to December being the last month in the year with many global holidays and expenses for individuals. When analyzing the deadline months, we found that April has the highest rate of success but again there was not much difference between April and the other top 11 months. The top 11 deadline months all range from 33% to 38%. January has the lowest rate of success with a drop of about 9% from April. This could again be attributed to January being the first month of the year and the expenses associated with the prior month.

We also wanted to analyze if geographic location had an impact on the success of a project. We did this by looking at the total number of successful projects per geography over the total number of projects per geography. We found that geographic location does have a small impact on the success rate. The United States has the highest total number of projects on the platform with about 9 times as many projects as the next highest geography ,Great Britain. The United States also has the highest success rate for projects. Although even with the much larger number of projects the US only has a 1.5% higher success rate than the next highest country. Overall, the top 10 countries all have success rates ranging between 28% and 37% that do not appear to be directly tied to the volume of total projects in that category which indicates that geography does have a small impact on the success rate of projects.

After looking at the impact of geography and date on the success rates we analyzed the rate of successful crowdfunding over time per category. We did this by looking at the number of successfully funded projects per category over the total projects per category per year. We found that the entertainment sector, Dance and Theater, had the highest rates of successful crowdfunding over time. We also found that the top 25 categories all had success rates of over 50% of their funding goal. Alongside looking at projects that met their crowdfunding goals we also wanted to assess which project categories overshot their crowdfunding goals. We did this by calculating the amount funded divided by their initial funding goal. The music category took the top 3 spots for overshooting on their funding goals with the project "Vulfpeck" overshooting their goal by about 100,000 times their initial funding goal.

Further analysis directed our focus towards the average goal and average pledge amounts for each main category. We concentrated on the categories with the two highest and two lowest average goals and pledges. This showed us that projects under the Technology category had the highest average goal of $102,289 while Dance categories had the lowest average goal of $9,588. That is a difference of $92,701 between the largest and smallest average goals. Technology likely has the highest average goal amount as those projects tend to need access to larger funds in order to be successful. Looking further into the Tech and Dance categories, we were able to examine their average pledge amounts. Despite having the lowest average goal amount, Dance had an even smaller average pledge of $3,452. Technology projects only had a $21,151 average pledge amount compared to their goal of $102,289.

The highest average pledge amount went to Design with $24,417, but its average pledge was slightly closer to its average goal of $41,871 compared to Technology and Dance. Crafts on the other hand received the smallest average pledge of $1,633 and an average goal of $10,434 which means its average goal was around 6 times as much as its average pledge. Overall, most of the categories show a much higher goal rate than pledge rate. These rather large differences between categories and their pledge/ goal amounts offer great room for further inquiry.

This led us to determine whether a project that had been fully funded, where the pledge amount met or exceeded its goal, was more likely to be successful. Our analytical process seemed to confirm that proper funding meant success. To achieve this conclusion, we grouped all projects by their state, where all projects had a pledge equal to or above their goal. Out of all projects that met or exceeded their goal, the majority were successful. Successful projects accounted for 133,951 out of 137,042 which represents 98% of these projects.

Another insight we established was in regard to the largest pledge amounts and total pledge overall for each main category. Originally, we began our analysis on all projects using the entire Kickstarter data. However, we found outliers existed in the amount pledged that were based on unreasonably successful projects, such as "The Pebble Time Smartwatch" which amassed $20,338,986. The next closest maximum pledges only reached between 6 and 8 million which was still hugely outside of the other 11 categories max pledge amounts. For that reason, we excluded projects that received pledge amounts greater than $500,000, because there were only 488 projects greater than $500,000 and 374,376 projects below that range. After these improvements were made, we found that the category with the highest pledge amount changed to Technology with the largest pledge being $499,168. Interestingly enough, the categories with the largest pledge amount and total pledged amount (Design, Games, Technology) remained the top three for both data frames but changed order. These categories reached roughly above 3 million. The category with the smallest maximum pledge of $146,075 in Dance and also had the smallest total pledge of $1,116,158.

We sought to answer what percentage of the total amount pledged went to each main category. To accomplish this objective, we grouped by main category and then took each categories' total pledge amount and divided it by the total pledge amount for all Kickstarter projects. The first discovery was that over 2 billion dollars in total had been pledged to all Kickstarter projects. The majority of funds were disbursed to Games, Design, Technology and Film & Video categories. These four accounted for approximately 73% of all amounts pledged to Kickstarter. Design captured 19.4%, while Games achieved 21.78% of all pledged amounts. The categories that had the least impact on pledge amounts were Journalism and Crafts which each comprised 0.39%. An important observation is that since the four categories detailed above account for almost three quarters of overall pledge amount, they are categories more likely to have higher impact in the overall dataset and our descriptive statistics.

We also looked into the percentage of funds that went into each category when the major outliers were removed, to better represent the categories. The categories with the highest percentage of total pledge were Design, Games, Technology and Film & Video. They accounted for 65% of the total pledged for Kickstarter. This resulted in Design achieving 17%, Games with 16.9%, Technology with 16.4% and Film & Video with 16%. This shows similar results as before the outliers were removed which helps to substantiate how these categories tend to overall receive larger pledge amounts and interest.

V. Discuss the implications or stories

This study investigated the success rates of the different projects by year, geographic location, launch dates, and crowdfunding goals between the years 1970 and 2018. This allowed us to determine which categories of future projects would be most likely to succeed, and which projects are not as likely to succeed. We believe that the success rate of these projects also helps us determine the financial implication of each project idea. We were able to easily identify Dance as the most successful category and Technology as the least successful category. Further analysis could be made here to determine which traits made each category more successful and which traits made each category least successful. However, we did analyze broad traits such as launch and deadline dates, geographic location, and crowdfunding goals which still allow us to make assumptions about which projects were more successful over others.

The most successful projects had anticipated launch dates earlier in the year, whereas the less successful projects had launch dates during the last two months of the year which is usually consumed by holidays. Similarly, projects in the United States have the highest rates of success over any other country. It would be interesting to analyze why the United States had more successful project outcomes. Was it because there was more funding in the United States, or could it be related to population?

Lastly, analyzing the crowdfunding data allowed us to get a closer look at which projects were anticipated to be most successful based off of their amount pledged. The amount pledged gives us a better idea of which projects people were willing to get behind. Successful projects kept funding goals lower which shows that for a greater chance of success, set goals to the minimum amount it will take to execute your project. Based on these findings, we recommend setting reasonable funding goals that will allow you to complete the campaign in the allotted time. We think that by making your project appealing to as many people as you can you are more likely to take advantage of the crowdfunding. Lastly, we can see that setting a reasonable timeline so that you can meet your deadline at a time of year where people are the least busy.

Furthermore, by using this data and our analysis we can determine whether Kickstarter is the right platform for their project. If it is the right platform for your project, then you could customize your project to be more likely to succeed given the data available. If your project does not fall into one of the more successful categories, you could use this information to alter your project or determine why your project may not be successful. We also noticed that Kickstarter helps to bring different communities together by sharing similar and different ideas for innovation. Kickstarter is also a great outlet of creativity and gives people the tools and opportunity to pursue their ideas on their own terms. Ironically the more creative categories, such as Dance, Theater and Music, are the most successful probably because these categories allow us to be more imaginative and innovative.

VI. References

[1] Kickstarter, PBC © 2022. "Stats." *Kickstarter - Stats*, https://www.kickstarter.com/help/stats?ref=hello.