

EXCEL HOMEWORK: KICKSTART MY CHART

QUESTION 1: WHAT ARE THE THREE CONCLUSIONS WE CAN DRAW ABOUT KICKSTARTER CAMPAIGNS?

From the dataset about Kickstarter campaigns, we can conclude that more than 50% of the projects will raise the funds to move forward as successful projects (53.11%). On the opposite side we have around 37% of failed projects, the remain amount are divided into cancelled and live projects.

Another interest point to observe is the fact that the majority of the projects are under one of these four categories (Theater, Music, Film & Video and Technology). Together they account for more than 78% of the total Kickstart Projects.

Looking into the detailed information about Categories (as per figure below), we can conclude the Category with the higher numbers of projects is Theater, followed by Music and Technology. On the other hand, categories like Journalism, Food, Games and Publishing are the ones with a lower number of projects.

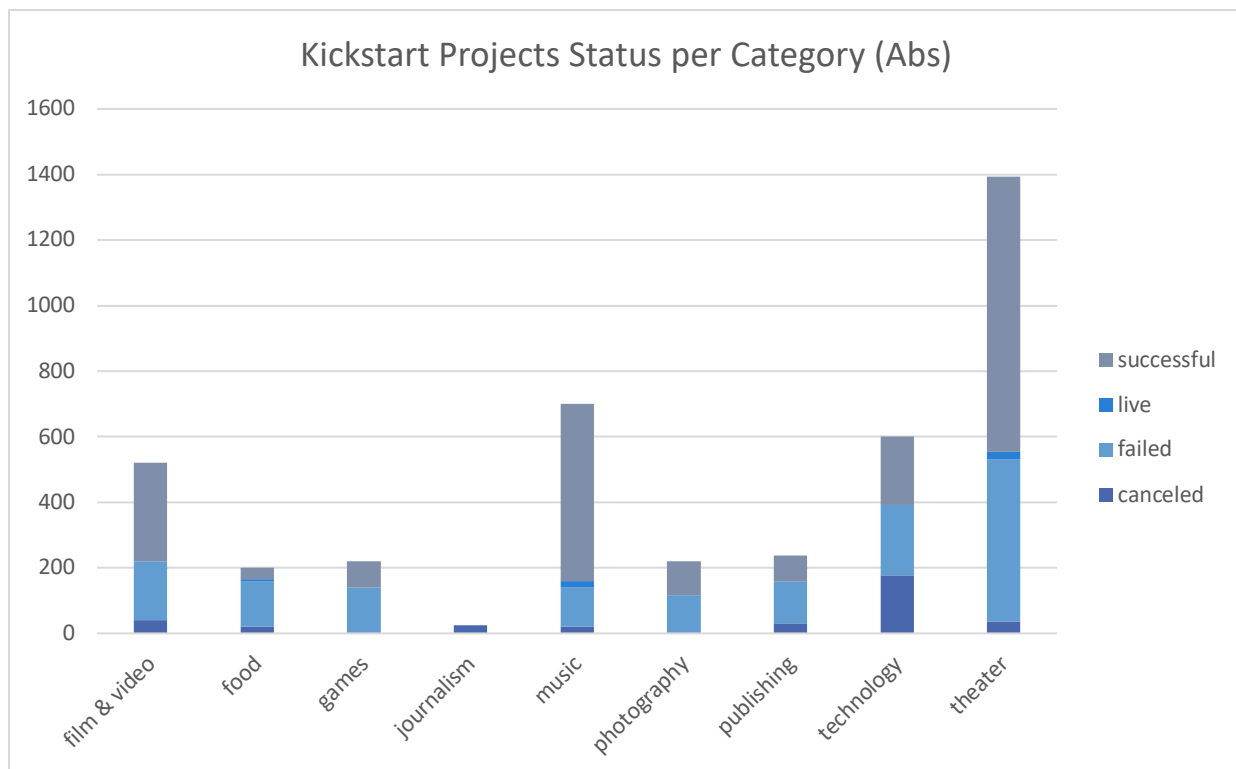


FIGURE 1 - KICKSTART PROJECTS: CATEGORY PER. STATUS

Despite the number of projects by category, the major insight we may have is to understand what categories are more successful compared with others. As you can see on the below table, Music is the most successful category since almost 80% of their projects are or will be funded. On the opposite side, Food and games projects are the most difficult ones to be funded.

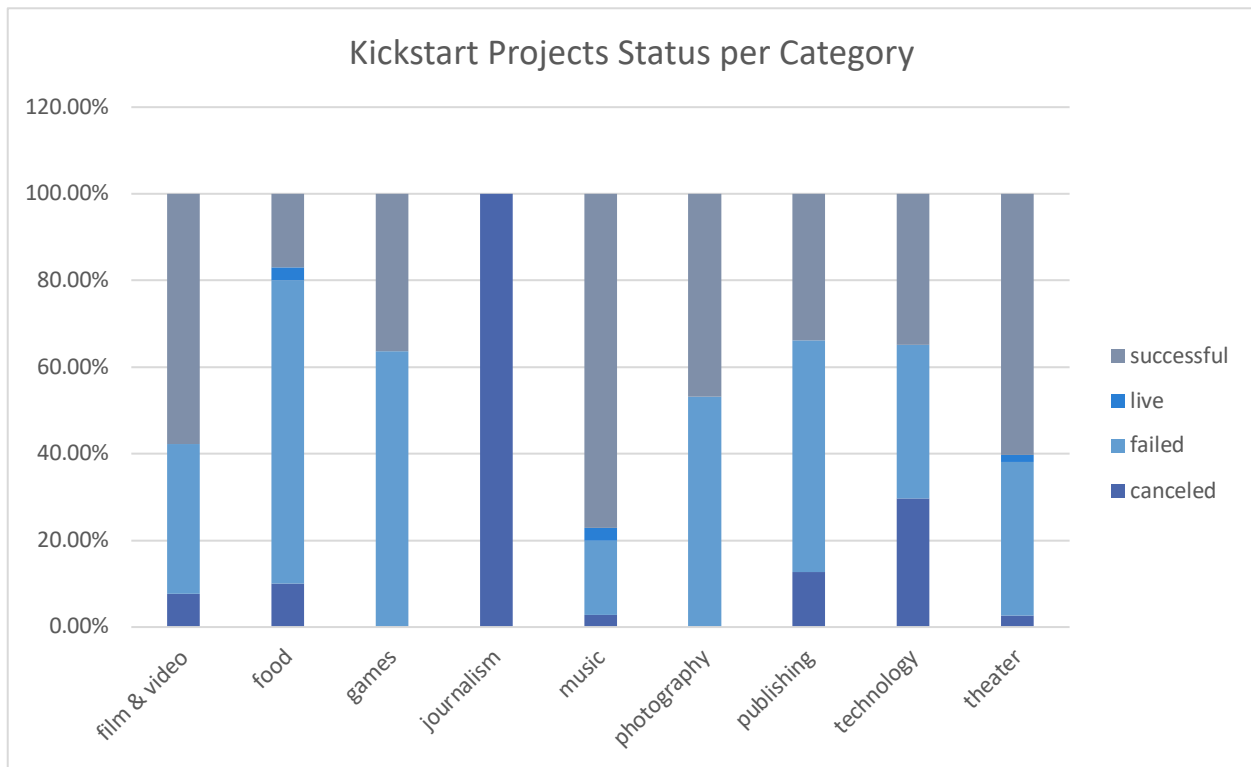


FIGURE 2 - KICKSTART PROJECTS: COMPARISON OF SUCCESSFULNESS WITHIN CATEGORY

Going one step down into the analysis, we clearly identify that Sub-Categories Plays and Rock clearly outperform other subcategories in what concerns number of projects. Nevertheless, the sub-categories which % of success have a significant amount are Hardware, Documentary and Rock.

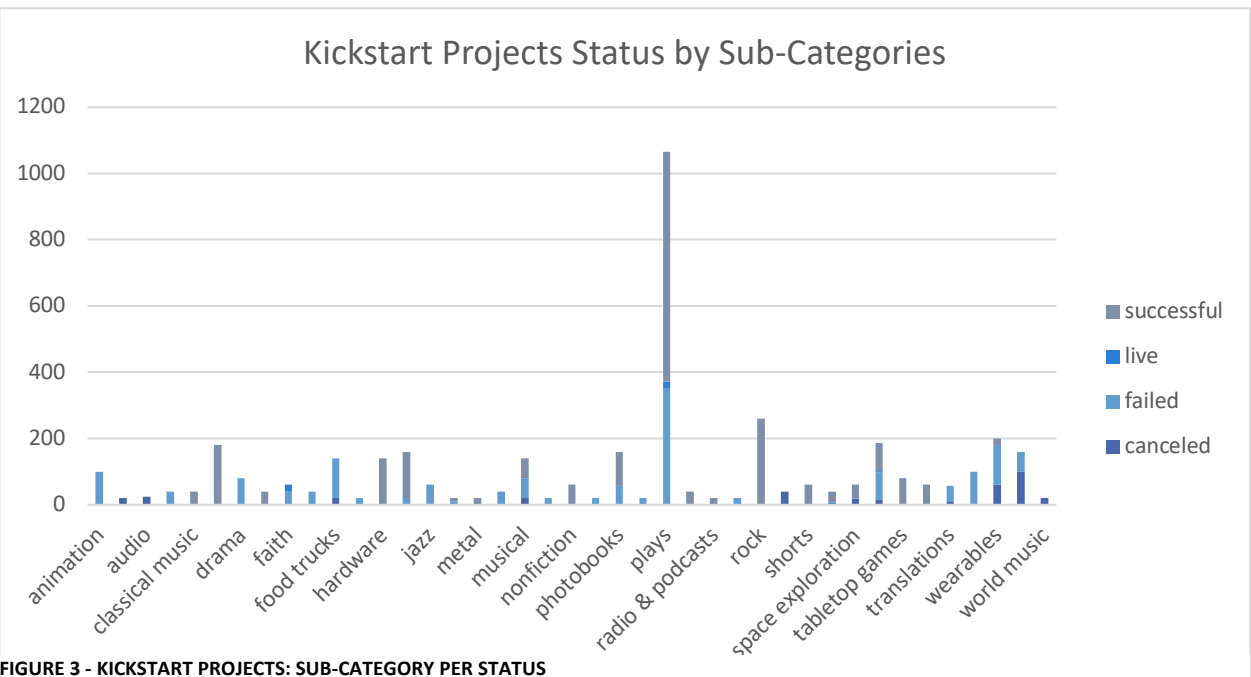


FIGURE 3 - KICKSTART PROJECTS: SUB-CATEGORY PER STATUS

Tecnologico de Monterrey

Another interesting point to cover is the seasonality of the projects and how this can influence reach the funding goal. The picture below highlights that seasonality. As you can see projects launched at the beginning of the year has a higher probability of get funded, what was expected since the people are more willing to spend some money since they may have more time, through the year, to get it back. Another main takeout of this graph is the fact the total number of projects is decreasing with the running of the year. In what concerns all canceled projects, they are not that dependent of any seasonality – very flat line.

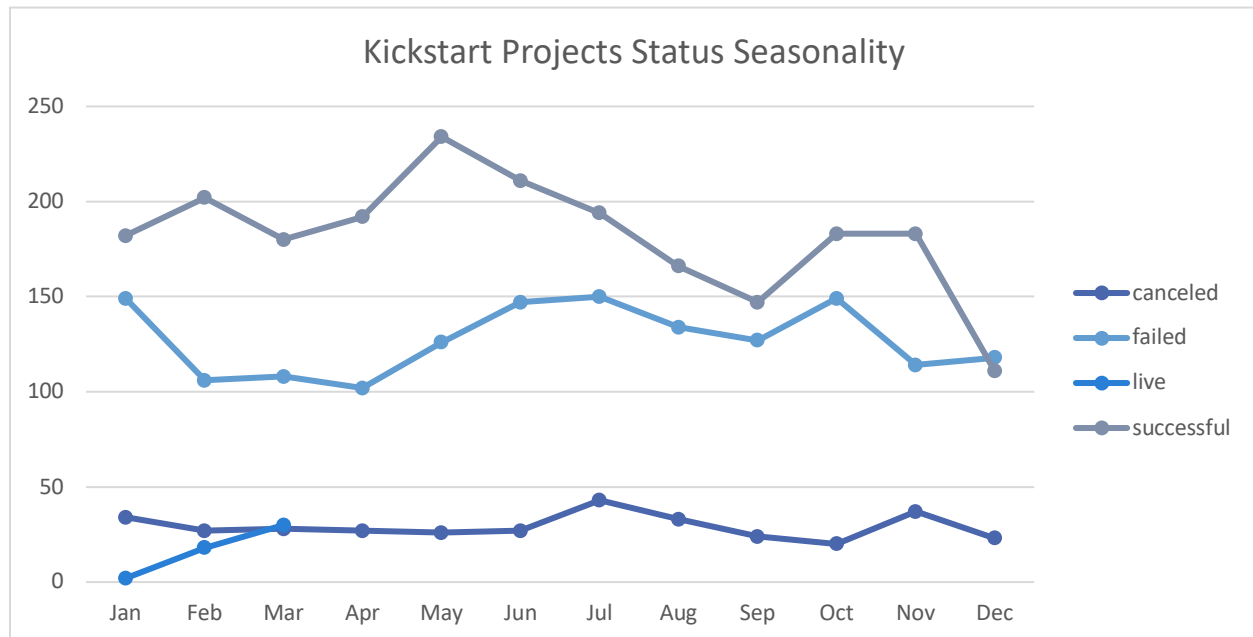


FIGURE 4 - KICKSTART PROJECTS: STATUS SEASONALITY

In summary, we can understand if we want to increase the probability of success we clearly should focus on projects around music, mainly rock, and launching them at the beginning of the year.

QUESTION 2: WHAT ARE SOME LIMITATIONS OF THIS DATASET?

Although the number of canceled projects is not that significant when compared with Successful and Failed Projects, one of the biggest limitations of the current dataset is not provide any explanation about why the projects were cancelled.

Being a funding website, it was expected to have some digital figures or variables behind this dataset.

Another point relevant to add to this discussion is the fact we are looking at aggregated data for each project, what can skew some information on the funding profile (a lot of investors in small amounts or a big player with a high investment amount).

QUESTION 3: WHAT ARE SOME OTHER POSSIBLE TABLES AND/OR GRAPHS THAT WE COULD CREATE?

There are a couple of other variables and relationships we can analyze. From all the possibilities we could run:

Tecnologico de Monterrey

1. A relationship between the Goal vs. State
2. A relationship between Goal vs. % Funded
3. Understand the relationship between how long a project is open and their status
4. A relation between Status and the Average Donation assessed by Category/Sub-Category
5. Understand the Year seasonality in terms of status and Categories/Sub-Categories.
6. Evaluate projects based on their currency
7. Evaluate projects based on their countries
8. How Spotlight influence the Success of a Project

There are a couple others of tables/graphs and insights we can run out of this dataset like more analysis about any kind of Clustering based on the % of funding each project gets.

BONUS: COMMENT THE RELATIONSHIP BETWEEN A GOAL'S AMOUNT AND ITS CHANCE OF SUCCESS, FAILURE OR CANCELATION.

There is a strict relationship between the goal's amount and the chances of success, as illustrates at the picture below. Thus, as we increase the Goal's amount, the chances of success will decrease considerable, from around 70% (on projects near \$1000) to 20% for projects above \$50,000. Somehow, it was expected on these kinds of platforms

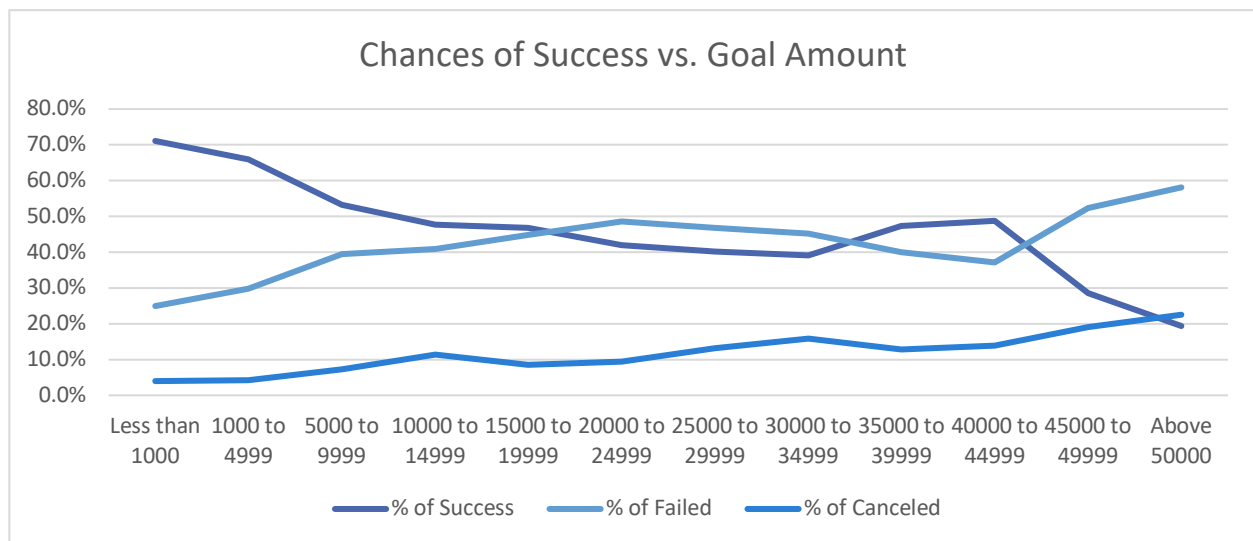


FIGURE 5 - KICKSTART PROJECTS: GOAL FUNDING EFFECTIVENESS

since the Capital raising is depending of considerable small amounts across a variety of persons, what makes difficult to achieve success on projects where the amount to invest is too high. Moreover it can also be connected with the level of risk/stage of each project, since, probably, people are more willing to invest on projects that has a shorter return (which typically will need less funding) compared with other projects where the payback may happen on a latter phase.

The opposite happens for Failed and Cancelled projects, they % will grow as we grow the goal's amount.

BONUS STATISTICAL 1: DEFINE MEAN, MEDIAN, VAR, STD. DEV, MIN AND MAX FOR BOTH SUCCESSFUL AND FAILED CAMPAIGNS.

The table below highlight the statistics of both distributions – Successful and Failed Campaigns.

TABLE 1 - SUCCESSFUL & FAILED CAMPAIGNS STATISTICS

Success	
Mean	194.43
Median	62.00
Var	712,840.99
Std Dev	844.30
Max	26457
Min	1

Fail	
Mean	17.71
Median	4.00
Var	3,773.22
Std Dev	61.43
Max	1293
Min	0

BONUS STATISTICAL 2: WHICH METRIC CHARACTERIZES BETTER THE DATA FOR BOTH SUCCESSFUL AND FAILED CAMPAIGNS?

Looking for Table 1 - Successful & Failed Campaigns Statistics we can understand, for both status, Mean and Median are very disconnected (not closed) and also the range of # of Backers varies from:

- 1 – 26457: Successful Projects, and
- 0 – 1293: Failed Projects

This shows that the range of each distribution is too wide when compared with any distribution central measure. It can also be observed on the following images.

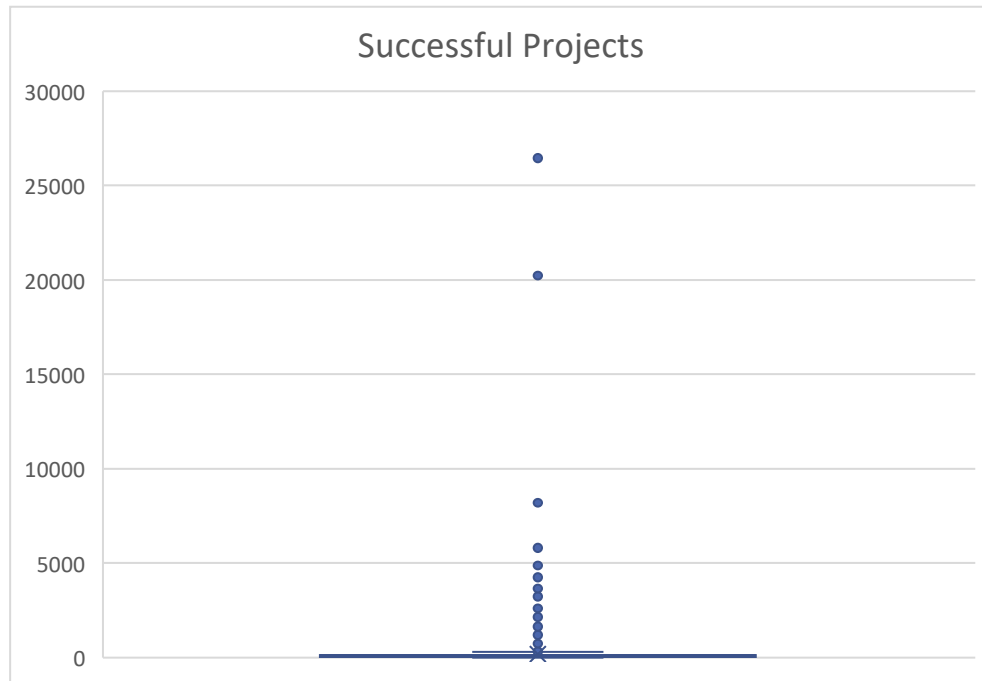


FIGURE 6 - SUCCESSFUL CAMPAIGNS BOXPLOT

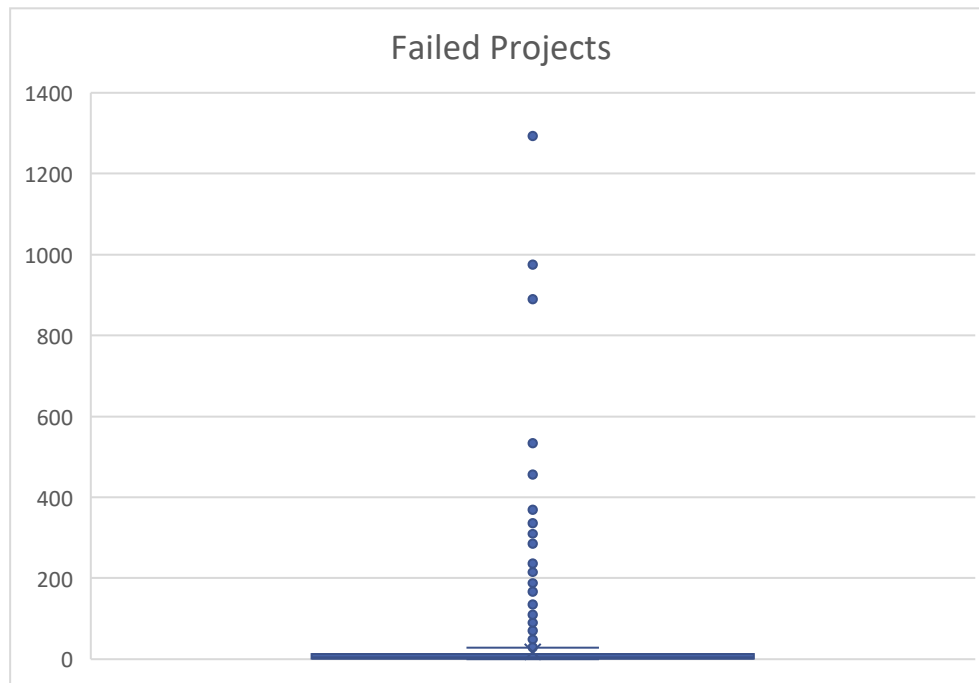


FIGURE 7 - FAILED CAMPAIGNS BOXPLOT

Here, we can clearly see that the number of outliers is very high on both distributions and therefore they skew the data a lot to consider the MEAN as one measure of centrality of the distribution. Thus, we should use MEDIAN as the most accurate measure of centrality for those 2 distributions.

BONUS STATISTICAL 3: DETERMINE IF THERE IS MORE VARIABILITY WITH SUCCESSFUL OR FAILED CAMPAIGNS. DOES THIS MAKE SENSE? WHY OR WHY NOT?

Clearly, there is more variability in Successful Campaigns than in Failed ones ($\text{Var Successful} \gg \text{Var Failed}$). In my personal opinion this makes sense, since the campaigns failed due to the lack of investment, getting as much money as possible from a short number of persons or in getting adherence from a vast number of persons.

Taking into account the nature of Kickstarter projects, raising funds depends on small investment amounts across a huge number of persons, therefore, to be successful the project needs to guarantee a specific number of investors – depending on their Goal amount. Thus, it is expected that Successful Projects will have a higher variance of # of Backers.

On the other hand, we may also conclude that missing to reach a certain initial number of investors will limit the opportunity to get your project funded, reason why the VAR around failed projects are smaller.