Joel Sotelo Llamas
Nov 16, 2018

Unit 1 | Assignment - KickStart My Chart

https://tec.bootcampcontent.com/Tecnologico-de-Monterrey-Coding-Boot-Camp/TECMC201811DATA2/blob/master/week-1/HOMEWORK/README.md

Background

Over two billion dollars have been raised using the massively successful crowdfunding service, Kickstarter, but not every project has found success. Of the over 300,000 projects launched on Kickstarter, only a third have made it through the funding process with a positive outcome.

Since getting funded on Kickstarter requires meeting or exceeding the project's initial goal, many organizations spend months looking through past projects in an attempt to discover some trick to finding success.

Problem

To organize and analyze a database of four thousand KickStarter projects in order to uncover any hidden trends that could help organizations to find success.

Dataset

- **Id [Integer]:** Unique Id of the KickStarter campaign.
- **Name [Text]:** Name of the KickStarter project.
- **Blurb [Text]:** Description of the KickStarter project.
- **Goal [Float]:** Money expected for the success of the project.
- **Pledged [Float]:** Money collected from the KickStarter campaign.
- **State [Text]:** Current state of the KickStarter campaign.
- **Country [Text]:** Origin country of the KickStarter project.
- **Currency [Text]:** Base currency for the campaign's Goal.
- **Deadline [UNIX timestamp]:** Deadline of the campaign.
- **launched_at [UNIX timestamp]:** Starting date of the campaign.
- **staff_pick [Boolean]:** True if the campaign is promoted by the Kickstarter Staff.
- **backers_count [Integer]:** Number of sponsors of the campaign.
- **spotlight [Boolean]:** True if the project has its own webpage inside KickStarter's Spotlight
- **Category and Sub-Category [Text]:** Category/SubCategory of the project on KickStarter.

*Note: This DataSet contains dummy information for educational purposes.

Joel Sotelo Llamas
Nov 16, 2018

Data Preparation Process

In order to find some trends and understand the behavior of the funding, first I added a column over the dataset to know the Campaign Funded Percentage.

The **Percent Funded** was obtained by dividing the amount of pledged money by the campaign's Goal.

$$Percent\ Funded = \frac{Pledged}{Goal}$$

With the Percent Funded we can understand more about the success or failure of a campaign.. letting us know if a campaign pledged way more money than expected.

For example, *The Universe in a Sphere (Relaunch)* project, a German project backed by only 52 persons, pledged 306,970 EUR, 27 times more than its original goal: 11,000 EUR.

This kind of information can give us a clue of where the success of a project can be.

Another column that can be useful is the average amount of money donated (**Average Donation**).

This can be calculated by dividing the amount of Pledged money by the number of donors of the project.

$$Average\ Donation = \frac{Pledged}{backers\_count}$$

The Practical Meter: Know your power!, a project with the purpose of improving the charging process of phones, was able to Pledge $167,820.60 USD by getting only $20 USD on average by his donors.

Formatting.

During the Data Preparation process, sometimes is necessary to give format to the dataset to find the best way to analyze the data.

In this case I made two format modifications:

- **Category and Sub-Category** field was split in order to have more depth in the analysis.
- The **Deadline** and **Launched_at** fields were formatted from the UNIX timestamp to a MM/DD/YY date format.

Analysis Process

To analyze this information and to obtain insights, we will create some charts.

Chart 1.- Projects by Category x State

This chart allows us to get an idea of how successful the projects are in correlation to their category.

An insight that can be obtained from this chart is that, in average, *Music* projects are more successful than other type of projects, even more successful than the category with the higher number of projects: *Theater*.
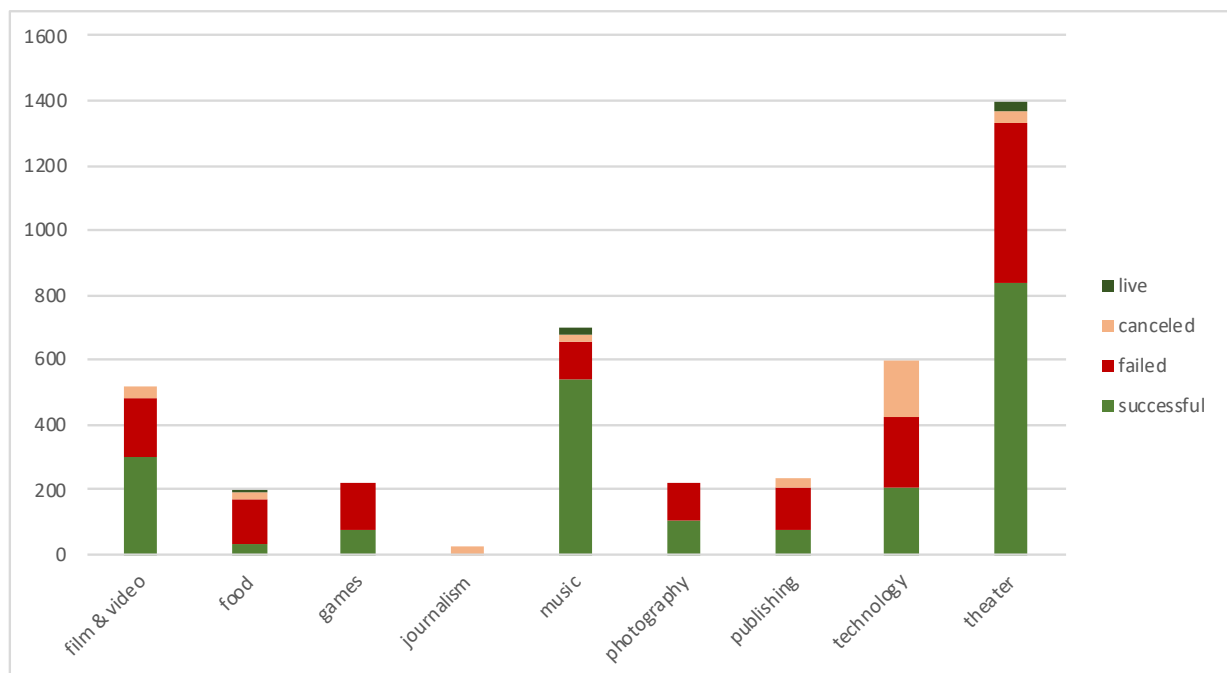
Chart 2.- Projects by Sub-Category x State

This chart is a drop down of the Chart 1 that allows us to really understand, no matter the Parent Category, what kind of projects are more successful in general and what kind are more prone to fail.

In this case, by filtering the *Film & Video*, *Games* and *Tech* categories, we can see that within *Tech* the most successful kind of projects are the *Hardware* related, winning over *Wearable* projects despite being more in number.
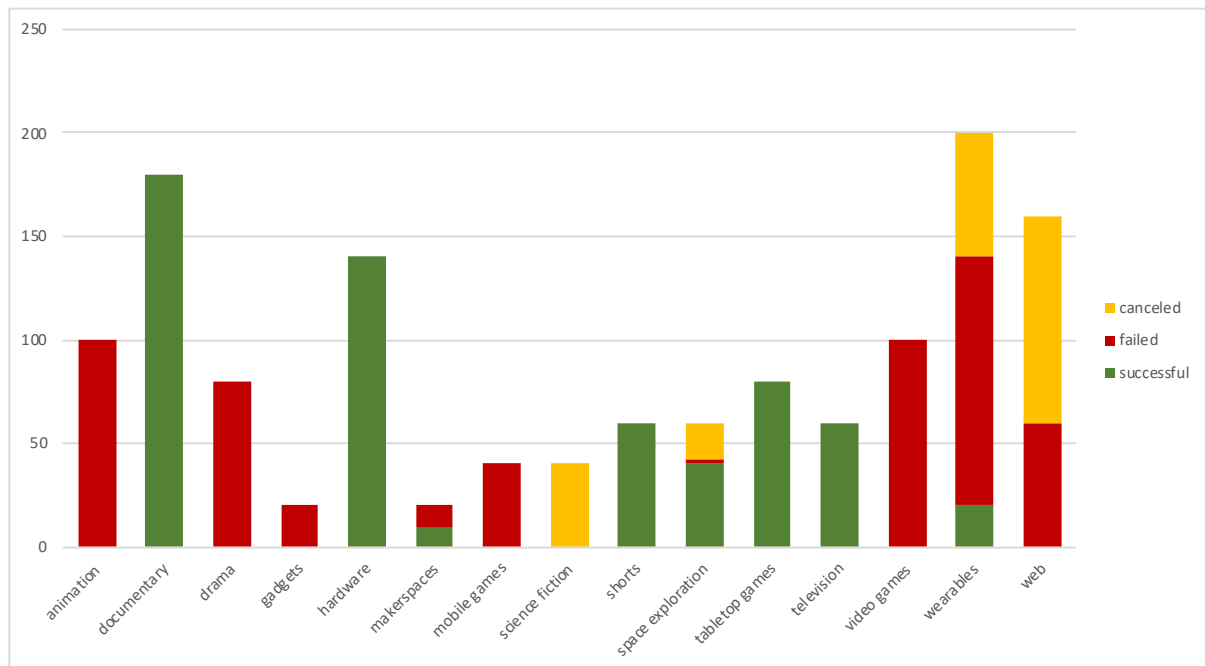
Chart 3.- Projects by Current State x Month Started.

This chart allows us to know which months are the best for launching a project on KickStarter.

According to this chart, the best months to launch a KickStarter campaign are *May*, followed by *February* and *April*. However, we have to understand that there is a correlation also between the number of projects launched each month. As we can see, the number of projects cancelled also raised from April to May.

A way to analyze this chart is to see the distance between the successful campaign line and the failed campaign line.. the more distanced they are from each other (as long as the successful campaign line is above), the better is the month.
Under this analysis, we could say that the best season to launch a campaign is *Spring* (from *February* to *June*), and the worst season to launch a KickStarter campaign is *Winter*.
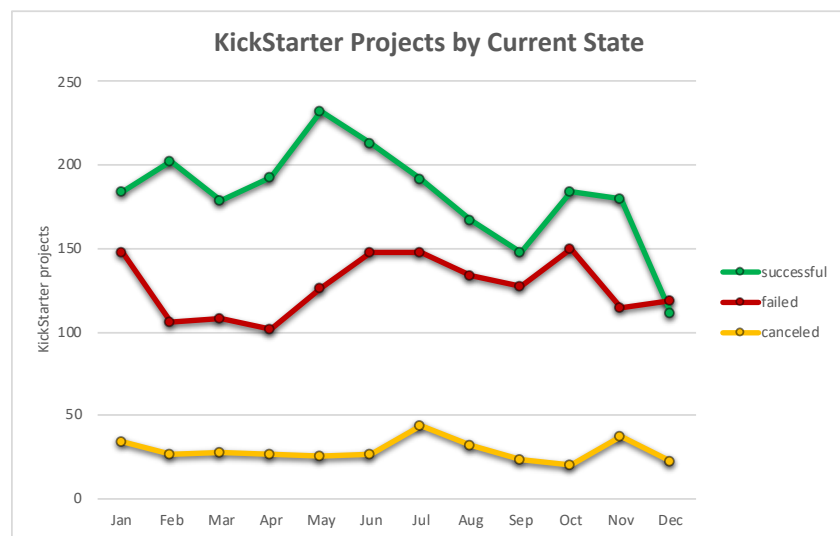
Chart 4.- Outcomes based on Goal.

This chart can give us an idea of how successful a campaign can be according to the amount of money set as Goal.

Just by looking at the chart we can see that the "sweet spot" is setting a goal between 10,000 and 20,000, this is when the percentage of successful campaigns is over the percentage of failed campaigns and the amount of money is higher (for this case I will exclude the range from 35,000 to 45,000, because the amount of projects in that range is low and it can mislead the trend: 98 projects out of 4064).

Obviously the lesser the Goal is, the higher the chance of hitting the Goal: 71.1% of projects with less than $1,000 as Goal were successfully funded.
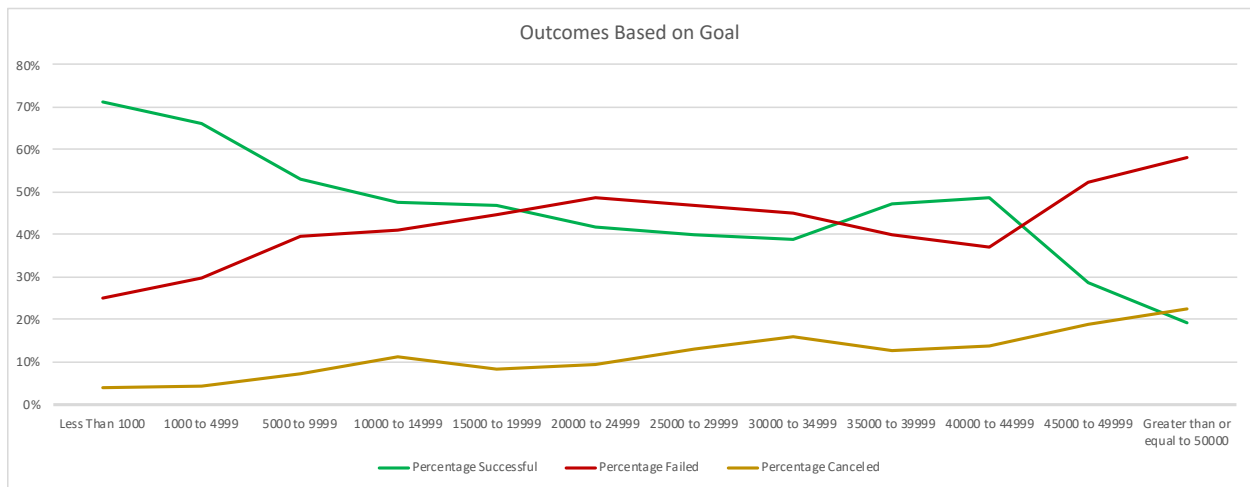

Insight/Note: This chart can be misleading because some projects have different currency. In order to accurately represent the information, we should convert the Goal and the Pledged fields to USD.

Analysis Conclusion

What are three conclusions we can make about Kickstarter campaigns given the provided data?

1.- At least 50% percent of the projects in the *Music*, *Theater* and *Film&Video* categories are successfully funded; being *Music* the most reliable category to invest, with over 75% of its projects successfully funded in proportion to the total number of projects by category.

2.- The KickStarter platform is starting to lose users. If we see the KickStarter projects by Year/Quarter we will see that the amount of campaigns on KickStarter has been decreasing since Mid 2015, having for Q1 2016 over 250 projects (100 projects less than Q1 2015).
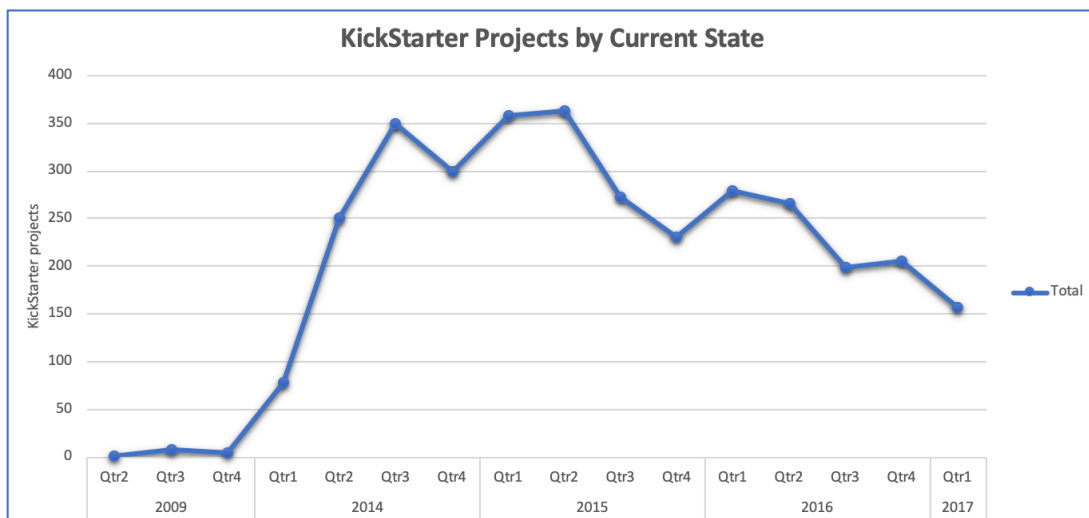
3.- KickStarter is not popular in Mexico. Out of 12 KickStarter projects Started in Mexico, 8 have failed, 3 are still Live and 1 got cancelled.
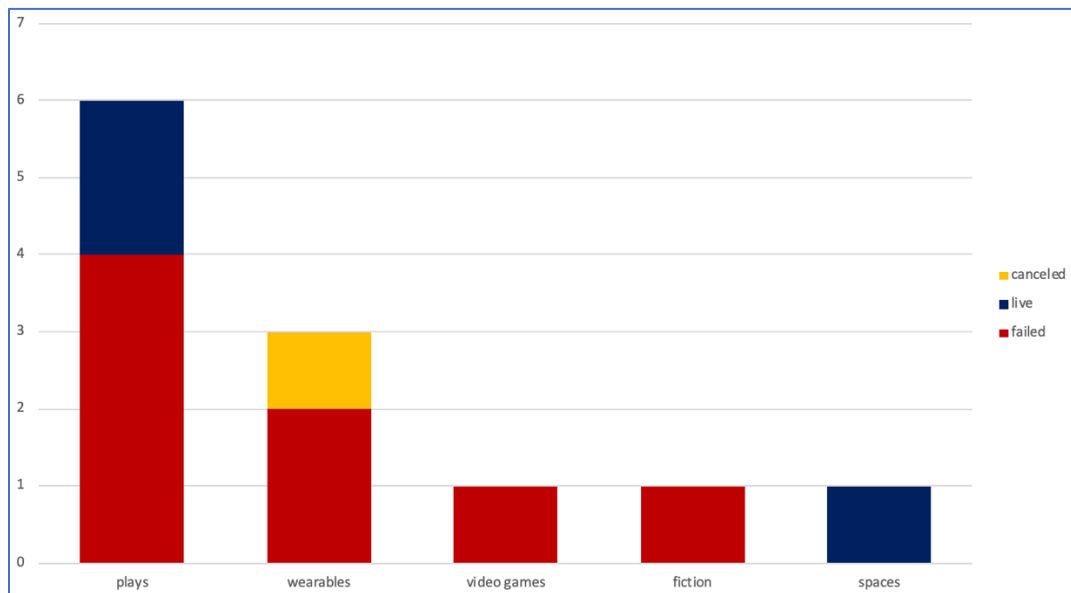
However, there is a light at the end of the tunnel: Of the three Live Mexican projects, the *Bisagra Teatro: Foro Multidisciplinario* project has already reached its goal of getting $40,000 Mexican Pesos.

name	blurb	goal	pledged	state	country	Category and Sub-Category	percent funded
Bisagra Teatro: Foro Multidisciplinario	¡Tu nuevo espacio cultural multidisciplinario en el centro de Pachuca, Hidalgo	40000	41500	live	MX	theater/spaces	104%
Make 100 | Geek & Chic: Smart Safety Jewelry.	Geek & Chic Smart Jewelry Collection, Wearables Meet Style!	30000	14000	failed	MX	technology/wearables	47%
Los Tradicionales	Somos una compañía de teatro independiente. Y en el 2017 queremos arrancar con el montaje de 3 obras. 3 elencos, 3 espacios.	38000	7500	failed	MX	theater/plays	20%
Soñé una ciudad amurallada	Somos... Podemos... Amamos... Nuestra muralla, nuestra utopía. Que el amor sea el límite	50000	5250	live	MX	theater/plays	11%
¡LlÉvame!	Buscamos finalizar el proceso de producción de un espectáculo de payaso y con Él, activar espacios públicos para la escena clown.	33000	2300	failed	MX	theater/plays	7%
Casa Calabaza, Premio Nacional de Teatro Penitenciario.	Conoce y apoya el teatro de calidad que se escribe desde los centros penitenciarios, como es el caso de Casa Calabaza, de Maye Moreno.	50000	2700	live	MX	theater/plays	5%
	Power Go es una línea de cargadores solares para						

## What are some of the limitations of this dataset?

I think it doesn't really have limitations. Maybe more information about the projects could be useful, for example, the city of the projects, cancelled date (to exclude projects that started and got cancelled during the same week), etc.

It could be really good to have the data by donors, by this way we could know which donors gave more money, which donated to different projects and also to know if the donors were from the same country of the project's country.

Joel Sotelo Llamas
Nov 16, 2018

What are some other possible tables/graphs that we could create?

Number of Donors by Category. This chart can give us an insight of which categories get more donors independently of the amount of money they donate.

Number of Backers by Category
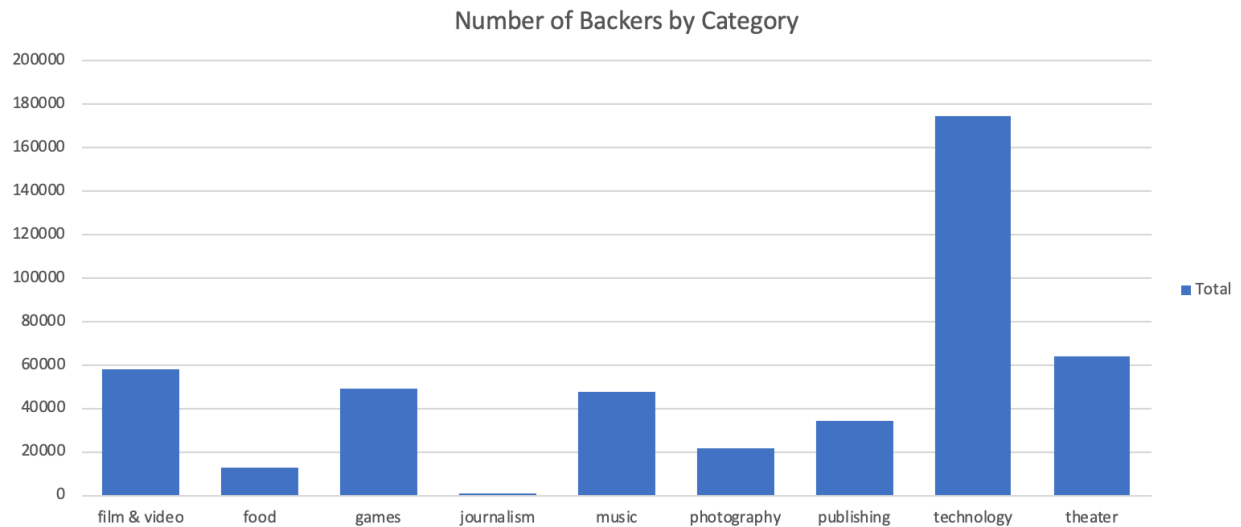
Other charts that can be useful are:
- Campaigns by country.
- Campaigns by country by Category. To know if it's easier to get funded in another country.
- Staff_Pick vs Success rate would be useful to know if getting promoted by KickStarter helps. However, there are no projects with Staff Pick field in "True".
- Top 1 Successful Sub-Category by Category.

Link to the project: https://github.com/joelsotelods/kickstarter-success-analysis