Iván M. Velázquez Martínez

Data Analytics Bootcamp

Instructor: José Manuel Barrera

Teachers Assistants: Alejandro Pimentel & Hector Padilla

Homework week 1: Kickstart my chart

1. Given the provided data, what are three conclusions we can draw about Kickstarter campaigns?

According to the problem stated at the beginning, many organizations look through past projects to discover a way for finding success, that is, getting funded. Having said this, some important conclusions from the analysis are the following:

1. According to Pivot chart 1, film & video, music and theatre categories are the ones that have more successful campaigns. Also, these categories have a positive balance between successful campaigns and unsuccessful campaigns. Then, technology is a category that has the most unsuccessful projects. So, it is more likely that projects related to the first three categories mentioned at the beginning get funded.
2. Pivot chart 2 shows that documentary, hardware, rock, classic music, electronic music, pop, radio & podcasts, short, tabletops and television subcategories are 100% successful. On the other hand, the subcategories that may fail or be cancel are 20. Within these, animation, drama, jazz, food trucks, videogames, wearables and web are founded, among others. It appears that if people focus on developing projects within the 9 subcategories mentioned at the beginning they will have a higher success rate.
3. Pivot chart 3 shows that May is the month with the most successful projects and July is the month with the greater amount of failed projects. Also, December is the month when there are more failed projects than successful projects. Having said this, it is more likely getting funded in months like February, April, May and June than in months like August, September and December.
4. What are some limitations of this dataset?
5. The presence of potential outliers is something worth saying. Even though these can be removed, the reliability of the data set can be questioned.
6. It is not known with certainty if the data are trustworthy and also, if the data used is a representative sample to discover hidden trends.
7. It cannot be said with certainty if all the data or variables presented are useful, in this case to found trends that make projects succeed.
8. What are some other possible tables and/or graphs that we could create?
9. It would be useful for future entrepreneurs to know the average of donations per category and subcategory. Even though the numbers of backers is already known, the average of donations will provide an extra overview.
10. The average time taken to reach the goal per category could be another useful information. There are some categories that reach it before long the deadline and others that that take more time.
11. Knowing if there is relation between the number of comments per category and the success/failure of the campaign could be created in order to have extra information.
12. Also, it is relevant to know how many people contribute with and without a reward to the project per category. And if this was relevant to defined the success or failure of it.
13. Browsing through Kickstarter page, It can be noted that projects do not have a rating. The latter could be a good piece of information when deciding the approach of the project.

Bonus statistical analysis

1. Use your data to determine whether the mean or the median summarizes the data more meaningfully

I think the median summarizes the data more meaningfully. First, it has to be mentioned that both data sets have outliers; this was known making a box plot. Also, both data sets are asymmetric because there is a big difference between the values of mean and median. Having said this, due to the outliers affect the mean more than the median, the median appears to be the measure that best summarizes the data.

1. Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?

According to the variance calculated from both successful and unsuccessful projects, the first one present more variability. This situation seems to be correct because certain projects, in this case, the successful ones, result in a considerable number of backers even if the goal has already been reached. It means that this can continue to increase before the deadline and the range of backers could be very wide. Of course, this is related to the product being offered, which, in people's eyes, has a lot of potentials. On the other hand, those that were not successful do not have a very wide range of sponsors. That is, the numbers of backers is limited and most of the time do not vary.