Mariana Geffroy

16/03/2021

# Excel Homework Report

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

* Most campaigns are inside the theater category, followed up by music and then by the technology category. The first two also have the highest number of successful campaigns, but in this case the technology category is replaced by film & video.
* There are 12 subcategories with a 100% success rate. In contrast there are 11 sub-categories that have a 100% failure rate.
* Throughout the year there are different trends concerning the number of successful, failed and canceled campaigns. For example, the number successful campaigns drops towards the end of the year, having their peak in May, whereas failed and canceled campaigns don’t have mayor changes during the year.
* Bonus: In the outcome depending on goal graph. The amount of money needed to reach the goal is related to the success of campaigns. The less amount of money is needed there is a higher chance the Kickstarter campaign is successful. For example, if the needed money is less than 1000 there is a 70% success rate compared to if the money needed is more than 50000 the success rate drops to 20%.

## What are some limitations of this dataset?

* Information from this dataset is not clear and tags are somehow confusing at times. It would be nice having a dictionary of the data in each column, and an explanation of what they mean, what and how this information was measured.
* We don’t actually know how much actual money was raised for each campaign, because it is in different currencies. Making this data homogenous and fixed to just one currency (USD, GBP, EUR) might give us an outlook on the real amounts of money needed for each campaign and raised. Conversion might be difficult to calculate due to different times and changing conversion rates. Maybe everything is in the same currency, but it is not explained anywhere so there is no way to know.

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

* A general pie chart to see the % of the 4 possible outcomes
* A table that could be created could be the state(outcome), and see if the outcome is related to “spotlight” and also to “staff pick”.
* A table and bar chart could be created using the dates to see if depending on the year there were different outcomes concerning the campaigns

## Statistics BONUS Questions:

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

In the case of both successful and failed campaigns I think that the median better summarizes the data because the variance in both and the max values are quite high, inflating the mean (more in the successful campaigns).

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

There is more variability in successful campaigns that in failed ones. This can be explained because the max number in successful campaigns is around 20 times higher than the max number in failed ones. Also the mode in failed campaigns is 0, and the median is 4 so half of the numbers are in a 4 number range.