Gerardo Angulo

March 23rd 2022

Excel Homework: Charting Crowdfunding

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

Looking at the data, the rate of success of crowdfunding over the last ten years has been above 50%, currently at 57%. With technology having the highest rate among parent categories at 68%.

The most popular parent categories are film&video, music, and theater.

June and July have the most successful projects. January has the most failed projects, followed by May and August.

1. **What are some limitations of this dataset?**

First and foremost, we have a data set of 1000 crowdfunding projects spanning over the last 10 years. It is a sample of projects that we don’t know how this data was collected and if it was randomly collected. We also don’t know from what sites this data was collected.

Secondly, when analyzing the data set for the success rate of a projects ability to meet and surpass its goal, I realized a lot of sub-categories don’t have enough data to confidently compare against other sub-categories.

For instance, there are a handful of sub-categories with less than 20 total projects in the dataset. Can we confidently say this is enough data to validate the rate of success for this sub-category?

The sub-category with the best rate of success, with over 20 projects, is animation at 65%. The caveat to this statement is from this data source they only have 34 projects. Plays has the most projects, 344, with a solid success rate of 55%.

1. **What are some other possible tables and/or graphs that we could create, and what additional value would they provide?**

A pivot table that showed the average amount of money donated by category and/or sub-category could show which categories of projects on average get the most money.

This pivot table could also show average “backer\_count” per category and subcategory which could serve as a compliment to the above information.

Bonus Questions:

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

I believe the median summarizes the date more meaningfully because it doesn’t get influenced by outliers with incredibly high backer numbers, how the mean does. This is important in a data set that has a high variance, which indicates how far a set of numbers is spread out from their average value.

 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 with successful projects. This doesn’t make sense because intuitive thinking would assume successful projects would have more backers. More backers should mean more money which in turn would mean a higher chance of success.

From this analysis, it shows the number of backers doesn’t dictate or predict a project’s success or failure.