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

Based on this bar graph of number of each outcome per Parent Category, We can see that Theater has the highest number of projects with the most successes as well as the most failures whereas journalism has the least number of projects (4 total) with all of them being successful. According to this data, you would probably have the highest chance of succeeding or gaining traction if you start a theater related project.

Based on this graph with the number of each type of outcomes per sub category, we can see that a majority of projects are within the Radio & Podcast category, containing the most successes and failures while all the other sub-categories have less than 55 projects attempted. Starting a project in Radio & Podcast will give you the highest chance of gaining traction but also a high chance of failing since it is such a populated category.

A picture containing text, screenshot, font, plot

Description automatically generated

Based on this graph showing the outcome over time, we can see that the most successes happen in July and there is a big drop to the lowest number of successes in August. August also has the highest number of failed and cancelled projects too which means August is probably the worst Month of start a project on Kickstarter.

1. What are some limitations of this dataset?

I don’t know how the data was collected. It should be a random representative sample. If not then that is one of the limitations.

We also don’t know the demographic of the people who use kickstarter. So Age, Gender, Nationality, Ethnicity and other demographic information like that may have an effect on the data.

Kickstarter claims to be a global platform but maybe, specific regions get more recognition than others which may affect our data/analysis.

Also, in kickstarter, you only get the money people have pledged if you reach your goal so some people may put up smaller goals so they can reach them and get the pledged money which will make them appear successful but the project itself may not have become successful.

We also don’t have the same number of projects under each goal category. For example, we only have 7 projects under the “20000-24999” goal category and all of them have been successful which shows a 100% success rate and 0% chance of being cancelled or failed which seems extreme and unrealistic. We have 315 projects in the “5000 to 9999” goal category which showed a 52% success rate with 164 of the projects being successful. It would be better if we had a lot of projects from each goal sub category so that we don’t have possibly skewed success rates.

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

A box plot could be used for the statistical analysis section so we can visualize the spread of the data when showing the number of backers in successful and failed kickstarters. This would make it easier to see if there is any data skewing the numbers.

It would be beneficial to look into the average goal per category and sub category and the outcomes next to that to see what is a normal goal range for each category. This could show us if some projects failed because they set their goal too high for the normal goal amount for their category.

**Statistical Analysis**

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

The median better represents the data because the Standard Deviation being so high tells us that the data is very spread out. It is very inconsistent and has very high variability so there is a high chance that there are a few outliers skewing the data. The number of backers for failed projects ranges from 0 to 6080 but the median is 114. This means there is a very concentrated number of projects with less than 114 backers and a bunch of projects with a lot more than 114 backers. The same goes for the successful projects whos number of backers range from 16 to 7295 but the median is 201.

1. 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 projects. I think this makes sense because there is a lot more successful projects than unsuccessful and different categories of projects probably demand very different number of backers which causes the high variability. The smaller sample size of failed projects is probably has more projects with a very low number of backers which is causing them to fail.