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

Based on the data, three conclusions can be drawn. First, among all categories, “film & video”, “music”, and “theater” have more successful projects than failed projects. Second, the smaller projects are more likely to be successful, especially the projects with a goal less than 20000. The smaller the goal the higher the success rate. Third, the successful projects distributed across the entire year with no apparent seasonal influence except December.

1. What are some limitations of this dataset?

One of the limitations is that the dataset doesn’t have any information on the backers. From an analysis perspective, it would be interesting to know if certain backers tend to pick successful projects. Also each project is unique enough to be an independent data point. Thus, the category or subcategory may have limited utility when doing data analysis. Another issue is the funding for the projects has different funding units. Thus, the goal analysis or average donation analysis can’t be classified or compared without having the currencies converted to the same currency. In addition, the dataset didn’t include any information on reasons for failure. Thus, it is difficult to guide future projects to improve the odds of success.

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

One of the graphs or tables missing is the goal by category. We can also find the trending of the pledged support by category or subcategory. Perhaps there has been a trend for certain categories as the years go by, and this trend can help determine the future projects to increase the chance for a project to be successful. Another graph is to see the total goal by state.

About the mean and median

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

Both”successful” and “failed” sets have outliers that skew the means, thus the median gives more meaningful representations to the center value.

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

On the surface, it seems that successful campaigns have a greater variance than that of unsuccessful ones. This makes sense. Both sets have very high standard deviation relative to the mean, but the standard deviation for successful campaigns is much larger in comparison to the one for unsuccessful campaigns.