Sarah Zachrich

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

* There are more successful campaigns than failed campaigns (including those that were cancelled).
* The campaigns that have the most success appear to be those relating to entertainment such as music, theater, film & video (best subcategories plays, documentaries, and rock music).
* The time of year doesn’t really seem to have much influence as there isn’t really much change throughout the year, although an argument can be made that the first half of the year looks a bit more favorable to succeed and I would avoid September and December. The cancelled and failed campaigns are fairly steady.
* Based off the bonus analysis it appears campaigns are more likely to be successful if they have a lower goal (less than ~45000 and especially less than 10000).

1. What are some limitations of this dataset?

I think the limitations depend on who is doing the analysis and what questions they are asking but in general there are a few that are widespread.

One is that is data from various countries. While this is also great to give us a bigger and less biased sample it could also cause issues depending on the type of analysis you want to do. If you are in a small European country, campaigns that are successful in the USA for example may not be a good basis for what campaigns will be successful locally. This greatly limits the amount of data you could use. Another issue the countries cause is with any financial data as the currency differs meaning currency conversions and much extra work would be needed.

There are some large outliers in this dataset that while I believe they are accurate due to the type of data could cause any analysis and charts to be skewed to one end or the other and may not be properly representative of the data.

Another limitation is the high variability of the data. We have very small campaigns up to very large campaigns, all with different goals, categories and time frames. While being a great data set for overall analysis like we have been doing for this homework, if it were a team with a small budget and time frame I feel they would have much less relevant data to work with.

Some of the data may not be as useful. For example all the live campaigns don’t yet have complete data. Other than to see what is currently running, this live data may not be very useful. This also holds for some of the other data such as the name and blurb categories as there is less you can do with this data.

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

I feel there are hundreds of other tables and graphs we could create with this dataset. These are some I feel may be more helpful or interesting than others.

* A graph comparing how long the campaign ran versus the success count would be interesting. In theory the longer the campaign ran the higher the success rate would be.
* Some charts looking at different variables (success, average donation, backers) versus country would be interesting to see how various countries compare
* A comparison of average donation or percent funded versus category or subcategory could give us an idea of what type of campaign people are willing to give more money toward. Something similar could also be done with backers count, average donation and categories to take it even more in depth and see which campaigns have less backers yet a higher average donation.