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

With the provided data, we can conclude that Kickstarter campaigns yield the most successful campaigns in the Spring season, the most successful launch date month being May. Secondly, we can conclude that certain categories and subcategories perform much better on Kickstarter than others: the best performing categories include Film and Video, Technology, and Music. Within those categories, Film and Video/Documentaries, Technology/Wearables, and Music/Rock performed the best in our sample of campaigns. Finally, we can conclude that there are Categories and Subcategories that fail most of the time on Kickstarter. Food and Photography categories do not seem to perform well. Within the food category, most food trucks fail their campaign. Kickstarter backers also do not seem to back jazz music at a high rate, or world music at all.

1. What are some limitations of this data?

Some limitations of the data are that this data is only through 2017. Kickstarter campaign formats and backers may have changed since this data was pulled and different variables may affect success and failures of different campaigns. Additionally, approximately 53% of this sample of campaigns met their funding goal. As stated in the introduction, approximately 1 in 3 Kickstarter campaigns are successful, or about 33%. This makes the Kickstarter data look as though there is more success than there actually is.

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

Other tables and/or graphs that could be created are:

* Average Donation of Successful, Failed, Live and Canceled campaigns.
* Average Donation of Campaigns according to Parent Category and Subcategory.
* Average Donation of Campaigns according to month and year of launch date as well as month and year of deadline.
* Amount of Successful Categories and Subcategories according to month and year of launch date as well as month and year of deadline.

Bonus:

1. Does the mean or median summarize the data more meaningfully?

With the current campaign sample, the median summarizes the data material more meaningfully. Looking at the maximum number of backers, there are two campaigns with abnormally large numbers of backers (>20000), and only about 60 campaigns of greater than 1000 backers. The mean of 194 backers takes all of these numbers into account. However, many more campaigns are funded with less than 100 backers. Because of this, the median appears to accurately reflect the reality of this sample of Kickstarter campaigns, rather than the mean. With a different sample, and if the mean and median were closer to each other, the mean would probably be more expressive of the data set.

1. Is their more variability with successful or unsuccessful campaigns? Does this make sense? Why or why not?

There is more variability with successful campaigns, as the variance equation for the successful campaigns yielded a higher number than the variance for failed campaigns. This makes sense, as successful campaigns have a large number of backers into the thousands. However, failed campaigns almost universally have fewer backers than successful campaigns, and therefore cannot meet their goals. In addition, when comparing the mean and the median for each set of campaigns, the mean and median are much closer together- and much lower overall- for the failed campaigns vs. successful campaigns. Successful campaigns may only have one backer or tens of thousands of backers, but most failed campaigns have about 25 or fewer backers.