Christina Carr

Assignment 1

**What are three conclusions we can make about Kickstarter campaigns given the provided data?**

1. **Category Insights:**

**Overall 53% of Kickstarter campaigns were successful.**

Top 3 Categories based on total number of successful campaigns:

Theatre, Music, Film & Video

Top Categories based on overall success rate (successful / total number of campaigns):

Music (77%), Film & Video (60%) and Theatre (58%)

Regardless if you would like to look at this analysis by the highest number of successful campaigns or the overall highest percentage success rates chances are if you are a Theatre, Music or Film & Video kickstarter your chances of having a successful campaign are higher than other categories. There is a clear trend that kickstarter donors cater to those 3 categories.

1. **Subcategory Insights:**

Top 3 Subcategories based on total number of successful campaigns:

Plays, Rock & Documentary

Top Subcategories based on overall success rate (successful / total number of campaigns):

Classical Music, documentary, electronic music, hardware, metal, nonfiction, pop, radio & podcasts, rock, shorts, tabletop games and television (all subcategories had a 100% success rate)

The Rock & documentary subcategories both scored in the top highest total number and highest overall percentage success rates.

1. **Campaign seasonality Insights:**

**\*Note: 4064 / 4114 observations are included in this table, the missing 50 observations are a small enough number that they should not influence the overall insights**

Top 3 months with the highest number of successful campaigns:

May, June & February

Top months with the highest proportion of successful campaigns (successful/total campaigns):

May (60%), February (60%), April (59%)

If we wanted to dig deeper we could get a more accurate result by comparing the highest proportion of successful campaigns on a granular basis per year by month as opposed to performing the analysis on all years could slightly skew the data if one year had a particularly high success rate compared to others. We could also look at the proportion of successful to failed kickstarter campaigns by month to see not only chances to succeed are high but also when the chances to fail are low. In any case for a quick and dirty overall analysis May and February are smart picks when it comes to starting a successful kickstarter campaign. Although June scored the highest in total number of successful campaigns April replaced it in the top 3 when it came to the highest proportion of successful campaigns.

**What are some of the limitations of this dataset?**

The first step in understanding the limitation of a dataset is knowing its source. In this case the first limitation is we do not actually know where this dataset came from. Does it come directly from Kickstarter? If it is someone’s observations based off of their knowledge then we should be aware of those limitations so that we can make sure to include any possible gaps in our analysis.

While the average donation column works nicely for practical purposes, ideally it would be nice to have that data in a more granular form (whether that is basing the average donation based on the actual donation breakdown or creating a binary variable with average donation and some parameters like high medium and low) since we needed to make an assumption or best guess on average donation using amount raised and number of donors.

**What are some other possible tables/graphs that we could create?**

The possibilities are endless! Essentially for each column we could go through and take a look for trends between that variable and percentage funded. However, how the analysis was broken down would really depend on the audience. For example, lets say you have a friend in Austria that want to start a kickstarter in the music category with a funding goal of $2,000. You could perform a deep dive analysis on the music category as well as take a look at other kickstarter who had similar funding goals which were successful and filter out those analyses to the Austrian market.

You could also run a linear model with total amount raised through a kickstarter as Y and many of the variables in the table as x to check for any statistically significant correlation between those X variables and total amount raised. I have done this with marketing data before in R-Studio but I am not entirely sure if it would make sense on such a dataset but it may be worth a shot!

Please note the bonus work is also included in my excel file.