**Conclusions**

One conclusion that can be drawn from the data set is that, overall, the categories in which a project would have the greatest chance of success are film & video, music, and theater, with success rates of projects within the category of 57.7%, 71.4% and 62.2% respectively.

Another conclusion is that Kickstarter campaign successes vary depending on the month of their creation date. There are generally spikes in success in the spring and summer months, with a downward trend at the end of most years. The data would suggest that creating a project in the middle of the year would make it more likely to be successful.

A final conclusion is that the chance a project will meet their funding goal is higher the lower the amount requested is. This could be because the lower the amount requested, the quicker it would be to meet the goal, all other factors being equal. It could also be because a potential backer feels like their money could make a bigger difference for a lower goal, or conversely, that there is less pressure to donate a larger amount in order to make an impact.

**Potential Limitations**

One limitation is that the sample is not representative of the rate of success of overall Kickstarter projects. Only a third of Kickstarter projects end with a successful outcome (Columbia Bootcamp GitLab Repository, 2019); the data set given has 2185 successful projects out of 4114, which is around 53% of the projects. The difference in makeup could lead to skewed results.

Another limitation is that the currencies aren’t standardised. While this is not a problem for percent funded, since each project has only one currency for ‘goal’ and ‘pledged’, it would create problems for trying to compare projects to each other, by metrics such as ‘average donation’. The currencies would have to be converted to one standard currency before making such comparisons.

Related to both of the previously stated issues is the geographical source of the data. The majority of data (over 3000 data points) is from the United States, with only the UK and Germany having over 100 data points, and the majority of the rest under 20 data points. With such small sample sizes, it is difficult to draw conclusions about whether the percentage successes by category are reflective of enthusiasm for certain types of projects within a country, or whether it only seems that way because those are the only projects within the data set. For example, there is a 100% success rate of photography projects, and a 100% failure rate of food projects on Kickstarter in Belarus, given that there was only one project per each category.

**Suggested Graphs/Tables**

A table/graph that could be useful is measuring campaign categories vs. percent funded, rather than the binary ‘successful’/’unsuccessful’. Looking at the data this way could reveal campaigns that were close to hitting their goals but didn’t. These campaigns that were close (80% and above) could then be analysed to see what makes them different both from campaigns that failed more dramatically and campaigns that were successful.

Methods of promotion such as ‘staff pick’ could be compared to ‘backers\_count’ or ‘average donation’ to see if those methods of promotion might affect the willingness of people to give money to the campaigns.

Average donation and backers\_count could also be compared against the success or failure of a campaign to see which of these values are more important to the success of a project. Comparing these values could suggest where further research could better be pointed towards: getting more backers, or getting backers to donate larger sums.

**Bonus Statistics**

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

The median would be a better summary of the data, as there is a large amount of variation within the data. Using the definition of outliers as being values less than the difference between the first quartile and 1.5 multiplied times the interquartile range; or values more than adding the third quartile and 1.5 multiplied times the interquartile range, there are 244 values within the successful projects’ backers\_count, and 192 values within the unsuccessful projects’ backers\_count that qualify as outliers. With this great number of outliers, the mean would be skewed upwards because of these extreme high values. The median would be less affected by these outliers, and thus would be a more accurate picture of the typical value of the data set.

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 within the successful campaigns data set. The difference between the max and min values, the variance, and the standard deviation for the successful campaigns set are all higher. All the unsuccessful campaigns had the number of backers that would be insufficient to meet their money goal, while the successful campaigns had enough backers to meet and exceed those goals. While each of the individual campaigns have different goal amounts, with different amounts of backers donating different averages of money amounts, one would expect the upper amount of the successful campaigns’ goal amounts, and along with it, backers, to be higher due to their meeting or exceeding their goals.This assumption seems supported by the data, in that the minimum values are fairly similar for both campaigns (0 for unsuccessful campaigns and 1 for successful ones), but the maximum value for the successful campaigns is nearly twice that of the unsuccessful ones.

References

*Excel Homework - Kickstart My Chart*. (2019) Columbia Bootcamp GitLab Repository. Retrieved November 4, 2020 from https://columbia.bootcampcontent.com/columbia-bootcamp/cu-nyc-data-pt-11-2020-u-c/tree/master/02-Homework/01-Excel/Instructions