**Data Science Bootcamp – Module Challenge #1**

**Part 1**

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

* Based on the stacked column chart by parent category, Theatre is the category which received the most funding from the campaigns and more than 50% of it was successful. Almost 40% of the Theatre campaigns did fail tough. This seems to be the most popular crowdfunding campaign amongst the backers.

Film/video and music are the next set of categories which are the most popular as they both had more than 50% of successful campaigns and had about 33% of campaigns that were unsuccessful.

The category which had the least funding was Journalism but it had a 100% success rate with campaigns. One outcome is that even though a category might have a lot of funding, its success rate may not be as high as a campaign that had a limited number of backers.

* As per the second stacked column chart which categorizes based on sub-category, the plays definitely had the most campaigns funded, and they outperformed all the other sub-categories. This is in trend with the Theatre category receiving the most funding. It also shows that the plays have more than 50% successful campaigns but also a great number of failed campaigns. This could owe to the demographic of backers/investors in the certain countries and years that the campaigns were run.
* The third conclusion that could be made is that crowdfunding campaigns are much higher from May to July, and then drops down from August and goes up around December. If we look at the categories of campaigns that were highest which were Theatre, film/video and music, these are more prevalent in the late spring and summer months (May-August) by having festivals, concerts, shows and new programs and releases, which could explain why they are trending higher in these months. There would also be more shows and entertainment programs start around the holidays

1. *What are some limitations of this dataset?*

* One limitation of this dataset could be the type of investors that each category had for funding based on the region in a country, rather than the country as a whole. Certain countries might have investors or backers of a specific field, which would account for having more funding of a specific category.
* Another limitation is the specific types of campaigns that were run. We do not have too much information on what the individual campaigns were, just a blurb, which might not give much in site as to what types of campaigns within each sub-category received more funding. More details on the campaign purpose and how they were promoted and the types of people supporting these campaigns would provide in sites for future campaigns in those categories. It could also help people that want to crowdfund for the categories that did not receive much funding to determine what tactics they should use and when they should go ahead with the campaigns, to increase the chances of receiving funding.

1. *What are some other possible tables and/or graphs that we could create, and what additional value would they provide*

* A graph that compares the date the campaign was created versus country could be beneficial to see how many specific types of campaigns were started in which country, based on the demographic and the core industries that are present in that country. If the specific region of the country is mentioned in the table or include in the graph, it could help us understand where to go for specific types of campaigns, rather than just reaching out to the backers in your current region. This would also help us understand when specific countries funded certain types of campaigns, based on their economic situation within.

**Part 2**

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

* As per the statistical analysis, the mean better summarizes the data because there is a lot of variation between the backers count for the successful and unsuccessful campaigns. Since there is a great range between the min and max values, the mean would most likely provide the result that better summarizes the data for the campaigns.

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

* As per the statistical analysis, there is more variability with the successful campaigns because the variance for it is 1,603,373.732, which is much greater than the variance for the unsuccessful campaigns, 921,574.682. This does also make sense because there are a lot more data points for the successful campaigns, so the chances of having variability would potentially be with the campaign that has more data points.