Module 1 Challenge: Excel

Christopher Yang, Class A

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

The top 3 most crowdfunded categories are: Film and Video, Music, and Theater. When broken down into subcategories, the majority of crowdfunded projects are Theater Plays specifically. In any given month, there are more successful crowdfunding projects than failures.

## What are some limitations of this dataset?

The data provided does not explain why certain campaigns are significantly more funded compared to other campaigns. It does not explain why certain campaigns are much more 'popular' or exceeds their funding goals by a larger margin than other campaigns.

The dataset does not include data about crowdfund backer rewards (products or services for backers who donate a certain amount to the crowdfund campaign). This factor influences the outcome of the project by providing incentive for backers to join a crowdfund campaign and to donate a certain amount of money to receive the campaign rewards (like products and services from successful crowdfunds). It is also important because the cost of these rewards also factors into the minimum goal amount that crowdfund campaigns require in order to achieve a successful outcome.

Finally, the dataset does not include data about crowdfund campaign advertisement efforts in terms of cost, platform, etc. Similar to campaign rewards, the amount of advertising can affect the outcome of a project. The cost of advertising could be another factor in the minimum goal required by the crowdfund campaign to achieve a successful outcome.

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

* Pivot Table and line chart showing Date created compared to Date ended and outcome
  + This can show whether a lengthier campaign tends to be more successful
* Pivot Table and bar chart comparing monetary Goal to subcategories
  + This would require normalizing the various currencies into one, most likely USD. Comparing the monetary goal of subcategories can tell us what subcategories are most expensive to fund.
* Pivot table and line chart comparing Number of backers vs average donation, filtered by outcome (successful, failed)
  + This can tell us if campaigns are reliant on only a few backers with deep pockets, or many backers with each making small contributions
* Pivot table and scatter plot showing spotlighted or staff-picked, campaigns vs. Outcome
  + This would tell us if exposure/advertisement helps campaigns to more likely succeed

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

The standard deviation and variability in the dataset are very high compared to the mean, implying that the data set is widely skewed. When taking into consideration how low the median is compared to the mean, this suggests that most campaigns have relatively low number of backers, with a relatively fewer campaigns with large number of backers that’s skewing the dataset.

This being the case, the median is a likely better summarization of the data set, since it more closely reflects the number of backers for most campaigns. The mean would do a worse job because there’s a risk that it will overstate the number of backers for most campaigns.

## 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 with successful campaigns, both from a standard deviation standpoint as well as variance. This makes sense.

The data shows that successful campaigns have higher median and mean number of backers compared to unsuccessful campaigns. Another factor is that the minimum number of backers required for some campaigns to be successful are very low- as low as 16. But successful campaigns also include ones that overachieve their number of required backers, with some campaigns overachieving their goals by a large degree. This large degree of difference between the minimum required for campaign success vs. the high numbers that some popular campaigns achieve leads to a high variance.

For unsuccessful campaigns, the mean and median number of backers is comparatively lower than successful campaigns. As explained earlier, successful campaigns include ones that reach a minimum required number of backers, as well as ones that overachieve their number of backers; unsuccessful campaigns will always fail to achieve their required number of backers. This means variance can be expected to be lower compared to unsuccessful campaigns.