

## Module 1 Short Answers

1. Based on the data, three conclusions we can draw about crowdfunding campaigns are that theater and plays are the most widespread category with a grand total of 344 campaigns; with 187 campaigns being successful, the most being in the months of May through July. Another conclusion that we can draw is that there were more successful campaigns starting from 2017 and on. The number of successful campaigns from 2017 is above 60, while the number would be below 60 in the years leading up to 2017. A third conclusion that can be drawn is that the theater category had the most successes because of how much funding it had. Theater had the most funding out of all the parent categories with \$15,763,227, and this could be a reason why they had the most successful campaigns.
2. Some limitations of this data set could include that it only has data from the years 2010 to 2020. Because of this, it is not a complete and accurate representation of all Kickstarter campaigns, and I think that it would be interesting to see data from 2020 and on to know how the pandemic influenced Kickstarter campaigns. Another limitation of this data set is that there isn't a lot of detail about what kind of projects were launched, besides what category/sub category. It would be helpful to see which campaigns launched what kind of project, whether it be a tangible product or something more concept-like. This data could be useful in figuring out which projects had more success.
3. Another possible table that could be created is one that shows the mean, median, max, and min of the length of the campaigns per category, and if they were successful or not. This would help give insight on the timeline of the campaigns, and whether or not one category was more successful over another due to the amount of time they had for the campaign. One graph that could be included is one that shows that shows the outliers of the funding goals or how much was actually pledged per category. This can help visualize how the data can be skewed from the outliers.

## Statistical Analysis

1. For both successful and failed campaigns, the median summarizes the data better because the data is not normally distributed. With such a high variance for both, the data points

are much more spread out from the mean and each other. Because of this, the median is a better representation of the central tendency, as it is not affected by outliers as much as the mean is.

2. Based on the minimum, maximum, and standard deviations of successful and failed campaigns, there is more variability with successful campaigns and this can be explained by the number of campaigns that are successful due to how many backers they have. The number of backers for successful campaigns ranges from 16 to 7,295, with a notable variation from campaign to campaign. In comparison, the number of backers from failed campaigns ranges from 0 to 6,080, which results in less variability than the successful campaigns. This makes sense because successful campaigns could have been successful due to high funding from its many backers, while failed campaigns could have failed due to low backers and not enough funding.