# **Module 1 Analysis**

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

- We can analyse that theater campaigning funding through the theater was the most successful and the most used source of campaigning

- The most successful and used sub-category was plays

- There were more successful campaigns than cancelled or failed throughout the years and July being the most successful for campaigning and January being the least successful

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

- The amounts pledged are in different currency. Therefore, we do not accurately know how much total was pledged unless all pledged amounts are converted to the same currency

- This is a sample data set of 1000 therefore the dataset is not large enough and could cause some discrepancies based on the campaigns chosen

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

- Country/outcome: By comparing the country and outcome we are able to see where most of the funding is coming from and which country crowd funding is most successful in. This could be done through a heat map or bar chart.

- Average Donation/parent category & sub-category: By doing this we are able to see which type of crowd funding generates the most pledges per person. This could be done through a a table or bar chart.

- Outcome/Goal: By doing this we can see the success rate with the goal set for that campaign, telling us if certain goals are more achievable than others

- Add Duration of campaign/Outcome: This will show us if certain lengths of a campaign will effect the success rate. This can be done through a scatter plot.

# 

# **Statistical Analysis**

- In this case, I believe that median will better summarize the data as the sums of the median and mode are drastically different. This means that there are outliers that are skewing the data in both successful and unsuccessful data, there are both very high and very low numbers effecting the mean. Therefore the median will better accurately summarize the data, giving us a midpoint of the data.

- Looking at the variance and standard deviation, we see that successful campaigns have more variability compared to unsuccessful campaigns due to the high number of successful campaigns for the variance and standard deviation. This makes sense as successful campaigns would have a significantly more number of backers as the campaign has drawn in a higher number of backers in order to be successful in reaching their goal.