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

1. The most amount of crowdfunding campaigns are for plays. We can see a large amount of success from these campaigns but there are also a lot of failed campaigns. This makes sense the theater industry has a lot of risk associated with it. Raising money for plays can be daunting, which is why crowdfunding has a crucial role in making small productions happen.

2. We can see that there are more successful campaigns in the summer months as opposed to the rest of the year. Sense theater is one of the big driving forces behind the success of crowdfunding campaigns, we can easily attribute the summer to be the most popular. With the climate being milder for outdoor venues and school being out there is a clear reason for more success in the summer.

3. Audio and world music have the least amount of campaigns. However, they both have 100% success on their crowdfunding campaigns.

What are some limitations of this dataset?

We are limited to the size of the dataset. Although the dataset is 1000 campaigns, the more we can track the better our analysis can be. We are also limited to some deficiencies in the data itself. For example, it would be helpful to know where some of these campaigns took place. Then we can pinpoint maybe why certain campaigns failed.

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

I think an interesting graph that could tell us some more information would be average donation vs. outcome. Perhaps we could see if the higher the average donation the more success the campaign has. If a few people can make a campaign successful or does it take a lot of people to make a campaign successful?

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

Since we have some campaigns with a large amount of backers and some with zero backers the median would be the best to summarize the data.

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

Variance tells you the degree of spread in our set. Since the successful backers produced a larger variance this means that the data is more spread. I suppose this might answer the question I had above: it doesn’t matter how many backers a campaign has. There is still a chance of success with a few people or thousands.