Aside from games, the entertainment categories seem to be the most popular to run crowdfunding campaigns for, as indicated by the Parent Category pivot table/chart where Film & Video, Music, and Theater have by far the most crowdfunding campaigns. Of these categories plays are by far the most popular kind of campaign. The data which the campaign started does not seem to largely affect the outcome of the campaign, besides perhaps a small uptick in success during the summer months. The majority of crowdfunding campaigns occur in the US; however this does not seem to greatly impact whether a campaign succeeds or fails. While I might not call this a conclusion, the ratio of successful to failed to canceled campaigns per category seems to be roughly equal, though I would want to run a percentage calculation on that before making that a statement, as I will list down in other possible tables/graphs I would create.

Some of the limitations of the dataset are that the data provided does not provide the actual donation amounts for everyone. It cannot tell you which crowdfunding campaigns are successful simply due to a large donor and which ones are successfully being crowdfunded by many individuals. While it does present the number of donors and you could draw some conclusions based on that data, it would be guessing/assuming that more donors mean there is not one large donor funding most of the campaign. From our point of view a campaign either succeeds or fails and we do not have sufficient data to say why it does.

I would first create a table showing the percentage of campaigns that succeeded vs failed vs canceled per category. This would help me visualize if there is a common trend among each category that only a certain % succeed regardless of category. Also, I would create an additional table showing me how often a “spotlight” was successful in making a campaign succeed. These would give me a better idea as to why some campaigns are failing and why some succeed.

The median is a better summarization of the data because when a data set has a lot of variances, the mean is skewed by its same variance, and can lead to misleading conclusions. Such as with this data set where most of the data points occur under 500, but the mean is all the way up at 851 because it is being skewed by the higher data points having a larger impact on the average.

According to the data and the variance calculations the Successful campaigns have greater variability. This does make sense because most crowdfunding campaigns that fail are likely to have a very small number of backers, as once it gets to a certain point in the campaign and it is not close to being successful less people will back it. It is much rarer to have a campaign which failed and has many backers as at that point people are more likely to want to push it across the goal line. This creates a greater consistency of a low number of backers which would lead to less variance in the failed campaigns as opposed to the successful ones.