**Crowdfunding analysis**

From the data provided we can assume that the most successful crowdfunded categories are film & video, music and theater. But besides that, the three categories mentioned are also the ones with the most crowdfunding campaigns out of all the categories. They also have the most failures from the categories, meaning that the arts have the most attempts at crowdfunding.

Looking into subcategories plays have the highest amount of successful and failed campaigns for funding, making them very challenging to fund.

Games have a higher change of failing, with the greatest number of campaigns ending in failure as opposed to successfully funded games.

Some of the limitations that the data presents are the number of campaigns being recorded and the range of time for tracking campaigns, besides that, where we are getting the data from, being able to pull from other data sources might offer a better understanding for crowdfunding.

We could also benefit from creating tables with a focus on countries and which offer the most success for crowdfunding, focusing on categories with a how many successful campaigns where fully funded. Also, by looking at the crowdfunding by range of time it might help to figure out whether a certain time frame will help with gathering funding and if outside of that range, then it would define success or failure. Both examples would help with focusing on geo locations to focus funding and with what category and time frame, leading to better chances of success.

**Which better summarizes the data, the mean or the median?**

The mean shows us the average amount of backers per campaign and shows us a better showcase of the data, especially when looking for the average. It gives us a better understanding of the data gathered if we have move evenly distributed data. if we used the median, it would just show us the middle number in the campaigns, not necessarily the amount comparing it to the maximum number we have in the data. It would work better with more outliers. The standard deviation is not drastically far from the mean so for that reason I would use that in defining the data.

**Variability**

There is more variability with successful campaigns because there is more data and more numbers to gather information from. If compared to the failed campaigns, there is less variability because the data and the numbers aren’t within a similar range. It makes sense because the data does not hold the same number of backers per failed campaigns meaning there is less space for variance.