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

Based on the graphs created we can see that there were a lot of campaigns for the creative arts that were funded – the top three were theatre, film and music. However, this cannot be the sole metric of success – these three areas also had the highest number of failed campaigns. There was an area that had much fewer campaigns that were all funded e.g. journalism.

There were many campaigns for plays that were funded – reflecting the high level of theatre kickstarters as a parent category, but again it contained a high number of failed campaigns. There were areas that had much fewer campaigns that were all funded – world music and audio. Cancelled campaigns occupied the least space – suggesting that campaigns are either funded or not.

Geography does influence whether a campaign is successful - but it in a sub category by sub category basis for example, the percentage of theatre kickstarters that were funded does not vary too much from country to country. However, for technology, the proportion of successfully funded campaigns in this area from e.g. Italy and Canada were significantly higher.

August appears to be the worst month to launch a Kickstarter as the gap on the graph between successful and unsuccessful campaigns is the smallest.

What are some limitations of the dataset?

There are extremes in the dataset as to what constitutes a successful campaign. Some campaigns required very little investment to get funded (£100) to those that require a six figure level of investment. Is it really appropriate to consider the extremes together as an indicator of overall trends?

One key limitation is the variation in the numbers of campaigns between categories, so caution needs to be applied. If you consider e.g. Journalism on its own with all four campaigns being funded, is this an outlier – or were they four very well run, desirable campaigns so success could be more probable?

The dataset does not give us any context either. We have no way of being able to tell whether for example if a kickstarter failed because it was trying to get funding for a project that enters a well-established marketplace. Many other factors exist such as the actual quality and reputation of the company launching the kickstarter can have an impact on success. From my personal experience, I have backed kickstarter campaigns for companies that produce tabletop gaming pieces. Some are well established, have a reputation for quality and have been successful in multiple previous campaigns. Their campaigns tend to be fully funded in under 24 hours – whereas those companies that do not have this history struggle a lot more. This dataset does not take factors like this into consideration.

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

I would look at turning the first three graphs into percentages, to eliminate the swings in the number of campaigns. If as an investor I could see that e.g. Music had 70% of its campaigns funded whereas Plays had 60% of its campaigns funded, to see a return I would look to back a Music campaign as it is more likely to be realised. This reduces variability in the numbers of campaigns.

I would suggest examining the Percent Funding column further to see which campaigns could be characterised as successful – so examine the level of overfunding which could be used to indicate popularity and demand for a particular industry / area. Does a campaign just hit their backing target? Do certain areas consistently over achieve funding wise? I could use this information to try to judge where risk would be minimised.

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

I would suggest that the Mean summarises the data best. As the Z calculations show, the vast majority of the data (88% successful, 85% unsuccessful) is within 1 Standard Deviation of the mean – so clustered around it at a greater percentage than for a normal distribution. That does not mean that the median cannot be ignored however, as the box and whisker plots show that there are outliers with high data values that will distort the mean – but as so much of the data is clustered around the mean, it seems appropriate to use that.

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

The Z scores are similar – the vast majority are within 1 standard deviation for both, and the outliers have higher Z scores. Neither have any Z scores less than -1.

It does not surprise me that for successful campaigns there are some high positive variance figures – some, if launched at the right time, for a product in demand will be wildly successful, and so above the norm.

It is however surprising that 12% of the unsuccessful campaigns had greater positive variances than +1. This would imply that they were able to attract many backers – but either not enough because they needed a lot of backers – or the backers they did get only donated small amounts. This could be examined further to see which of these hypotheses are correct – if any.

For unsuccessful campaigns this is a surprise as I would have thought that there would have been some highly negative figures. However, what is being compared is the average backers of unsuccessful campaigns rather than campaigns as a whole. As we are comparing the same outcome, it is not too much of a surprise.