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

From the data provided, many conclusion can be made. First conclusion that can be seen is by looking at the campaign outcome based on the categories. It is apparent that the most successful campaigns were in theatre, however, the most failed were also in theatre. It can be inferred that theatre can be very risky but have a high yield. This is followed by music and film. Looking deeper into the graphs, it can be determined that technology, publishing and photography have a high success rate compared to failed outcomes. This suggests that these campaigns are low risk but do not yield a bigger outcome in total. If there is less room for risk, campaigning in technology, publishing and photography might be a better idea. Diving deeper into data by filtering through the countries, it can be seen that different countries yield a better result than others for example, journalism in US or games in GB. Similarly, campaigning for publishing in GB would be a bad idea due to a very high failed rate and almost no success rate.

Second conclusion can drawn from examining the outcome of the campaigns by sub category. Similar conclusion can be taken from the sub categories as the trend shows that the highest success rate is in plays as well as the highest failure rate. Therefore, plays are risky for campaigns. DK and IT are the best places to campaign for plays due to their higher success rate. Web, wearables, translations, television, and shorts are the safest sub categories to run a campaign in.

Lastly, a conclusion will be drawn from the dates the campaigns were created. Overall, it can be seen that in Feb, a lot of the campaigns are cancelled as their success and failure drops. In May, overall campaigns fail more. In June, there is a high success rate and campaigns fail less and get cancelled less. This trend is followed by Aug where the success rate significantly dips and failure rate increases significantly along with cancelled rate. There it would be a bad idea to start a campaign in Aug.

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

One of the biggest limitations that can be seen is that the dataset contains multiple currencies which may not equate to the same amount giving an inaccurate measure when it comes to comparing pricing, profits and revenues. Another limitation that can be seen is that not all categories and sub categories are equally represented in all countries. For example, journalism shows a good success rate but the sample was only collected in US. This can mislead analysis into thinking that journalism might be profitable but might not be in any different country. Similarly, the amount of data collected is not uniform across the board, meaning that there is a very large sample of plays but a very small sample for audio. Another limitation is that the sub categories are not all equal to each category. There are a lot more sub categories for film and video but only 1 sub category for food.

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

Other graphs we could create and analyze would include, count of backers to each category and sub category which could tell us if there is a higher demand for a sub category. This could be coupled with average donation to further analyze the revenue margin. This could be filtered by countries to show which countries would yield a better campaign.

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

Looking at the data set, choosing the median would be a better choice. The reason for this is when we look at the mean with respect to the max value and the min value, there is a much bigger gap. Which means that unusually high and low data will influence the mean and the result will be skewed. For example, there are campaigns that received many backers and many campaigns that received very few backers. In this sense, choosing the median would be a better idea since it takes away unusual high or low data and doesn’t let them affect the final score when making a decision. Looking at this statistically, the standard deviation for both successful and failed campaigns is close to 1000. This suggests that the data is far spread around the mean and has a lot of variance. Therefore, the mean would not be reliable to determine a statistical analysis. The median would be more robust.

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

The successful campaigns have more variability due to a much higher variance. The maximum number of backers is much higher in successful campaigns compared to failed campaigns. This suggests that the data is skewed to a higher value for successful campaigns.