**Create a report in Microsoft Word and answer the following questions:**

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

**Answer:**  Three conclusions that we can draw from the provided dataset of 1,000 sample projects are briefly explained as follows:

1. **Crowdfunding is still a popular tool for product launching:** Based on the provided sample data, we found that successful projects have more supporters overall than failed ones. Also, some successful campaign projects have significantly more supporters than others. Primarily, the " Fox-Williams " project has the highest number of backers, i.e., 7295.

Likewise, out of 1,000 total sample campaigns, we found 565 counts of successful campaigns, whereas unsuccessful or failed campaigns are much less, which is 364. Based on this, we can infer that more than 50% of the sample campaigns were able to succeed. However, this rate looks a little small, but with suitable projects, planning, and execution, the success rate can be drastically increased.

1. **The number of outcomes does not alone reflect the flourishing campaign category and sub-category:** We can note that some projects under specific categories and sub-categories may give us the wrong impression of being the most successful ones if we only look at their counts of outcomes; however, it is not an accurate picture. For example, the category ‘theater’ has 187 successful outcomes out of 344 grand total outcomes of this category in all countries. However, let's compare that data to all other categories regarding the percentage of successful campaigns from the grand total of campaigns under that specific category. We will then realize a different picture here, as now we will see only 54% of the total campaigns under the ’theater’ category have been successful. The same thing goes for the outcomes as per the sub-categories as well. Another category, “Film & Video, " shows misleading pictures too. So far, the category “journalism” is the one that has secured a whopping 100% success rate among all categories, even though the outcomes quantitively seem very low.

Hence, the analysis based on this success rate suggests that campaigns based on the “journalism” category are more likely to be successful. The categories, like theatre and film & video, maybe more popular as more projects are coming up, but they cannot guarantee 100% success. Likewise, the sub-categories “world music” and “audio” look very promising, with 100% success rates. It would be a great idea to explore these sub-categories further.

1. **Campaigns during summer are more successful:** If we see the outcomes of campaigns, we see they are more successful during the months of June and July, which are the summer seasons of the year. That does make sense as people are more accessible during the summer season as they have longer days, most children are on summer breaks, and people tend to go for vacations and enjoy more free time with their family members and friends, hence a perfect time to launch a new “popular” campaign.

Moreover, if we dive deeper into the data and see which month has the highest success rate, we find the month of June is 64%, which is the highest, and the second highest is the month of July. One point to be noted here is that immediately following the month of July, during the month of August, the line of outcomes significantly drops to 49%, suggesting people are less likely to invest their time and their resources into any campaigns for some time being which again would incline above, during the month of September gradually and then again go down and then again up, suggesting a change of pattern during different months of a single year.

Based on this trend, we can decide when to start working on our projects and initiate them to get the maximum successful outcomes.

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

Answer: There are many limitations of this dataset. Some of the main limitations are the following:

* We do not know how this dataset was collected or what process and resources were used while collecting this dataset.
* Also, this dataset may have gone through some cleaning while being processed, and we do not know if any of these data were duplicated, misplaced, or accidentally deleted before this.
* There may be certain guidelines or policies that may have unknowingly or knowingly affected the data collection and cleaning procedures.
* Within this dataset, we do not have all the data for every category and subcategory if we see some data are unavailable. For example, we don’t have data here for the journalism category from the country Australia.

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

Answer: All campaigns may have different outcomes if launched within different countries. Hence, we need to make tables and/or graphs that show outcomes per country. For example, we can make a histogram showing how different countries may give different outcomes based on their cultural, religious, or demographic features. Also, we need to make one more table and graph showing yearly outcomes. This will help us understand the trend of outcomes over the years. For example, a bar graph with a trend line showing how the outcomes are at inclining or declining rates over the years will give us some idea regarding its growing or declining popularity. Also, we need to calculate the interquartile ranges (IQRs) and make box and whisker plots for both sets of tables of successful and failed campaigns. This would help us identify the possible outliers on the counts of backers for both successful and unsuccessful campaigns, making our statistical analysis more accurate.

(Disclaimer: I have made extra statistical calculations and graphs here to summarize my datasets, such as IQR calculations and box-whisker plots).

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

Answer: For the successful campaigns, the mean, i.e., 851, is significantly higher than the median, i.e., 201. Likewise, the variance and the Standard deviation of counts of backers of successful campaigns are respectively 1603374 and 1266, which are also very high. On the other hand, for the unsuccessful campaigns, the mean is, i.e., 586, which is again significantly higher than the median, which is 115. The same is the variance and the standard deviation of unsuccessful backers, which are respectively 921575 and 960, which are high numbers.

The above calculations for both successful and unsuccessful backers suggest that our data distribution for both tables is skewed away to the right, indicating high values on both tables on the right side. For some better analysis, I went a little further ahead and created box plots for both tables to ascertain the possible outliers and was able to locate them as shown in the box plots for both tables.

Using mean in such a skewed dataset can give us the wrong picture as it is sensitive to extreme values and skewed data. Considering this, it is fair to say that in such conditions, the median would be the better and safest statistical tool to summarize the data as it is less affected by the outliers and gives us only the center number between data values.

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

Answer: The variance and the Standard deviation of counts of backers for successful campaigns are, respectively, 1603374 and 1266. Whereas, for the unsuccessful or failed campaigns, the variance and the standard deviation for the counts of backers, respectively, are 921575 and 960. In both cases, the numbers are very high. The variance shows how much the data points in a particular dataset vary from the mean. In contrast, standard deviation, which is only the square root of the variance, measures the same thing, like how far apart numbers are in a data set.

Based on the calculations as shown in the Excel sheet, it tells us here that there is more variability with successful campaigns, which have higher variance (1603374>921575) and higher standard deviation (1266>960) than unsuccessful or failed campaigns. This makes sense to us because any successful project may have different varieties of backers in different numbers. Some projects may be much more popular than others altogether. In this case, for example, the project named “Fox-Williams”, which has the highest number of backers, i.e., 7295, may have led to higher variability for the entire table of backers’ count for the successful campaigns compared to the table of the backers count for unsuccessful campaigns.