Part 1

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

Considering the data in all categories and all countries in the dataset, it can be concluded that the highest number of successful outcomes belongs to the theater category, followed by the film & video and music in the third place. This indicates that starting a campaign in these three categories can be beneficial due to their success rate. However, it should also be noted that even though there is a good chance that the projects may succeed in the stated categories, there is also a high risk of failure. The dataset also shows the highest number of failures is dedicated to theater, then music and film & video is in the third place. On the other hand, there are other categories such as photography, publishing, and technology that even though they have a smaller number of successful outcomes compared to the aforementioned three categories, but they also have a lower failed outcomes compared to them. Another point of consideration is the country. The successful and failed outcomes in different categories varies in different countries. For example, it is a very good idea to start a campaign in DK for the theater category since it has a very high number of successful outcomes compared to its failure. However, starting the campaign for technology or music would not be recommended due to their numbers of failure. This is almost exactly the opposite in CA since the number of successful outcomes is higher in technology and music compared to the theater.

Based on the data in all sub-categories and all countries in the dataset, the same conclusion can be derived. Plays which is the sub-category of the theater has the highest number of successful outcomes, followed by rock, the sub-category of music, and documentary, the sub-category of film & video. Even though they have the highest number of successful outcomes, the chance of failure is still high. Different countries still have different outcomes of success and failure which needs to be considered for starting a campaign. Comparing the same two countries as before, starting a campaign for plays and drama in DK is highly recommended and starting the campaign for rock, indie rock, and wearables would not be recommended due to their high rate of failure. However, in CA it is recommended to start the campaign for wearables and indie rock, and it is not recommended to start the campaign for plays and drama.

Dates are also important for having a successful or failed outcome. Comparing the data, it is shown that generally, June and July have the highest number of successful outcomes. January, May, and August have the highest number of failed outcomes. The highest number of canceled outcomes is also in August. Therefore, to start a campaign, it is better to start in July and highly avoid the month of Augst.

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

To have a successful campaign it is also important to know the target audience and gather data for them. Therefore, one of the limitations of this data base is that there is no information dedicated to the target audience. Another limitation of this dataset is that it is comparing the categories and sub-categories for different countries but not all countries have all the sub-categories data, and this makes the comparison more difficult and inaccurate. For example, CA has food trucks sub-category, but DK does not have it. Data should be consistent. One last limitation is that the currencies should have been converted to be the same.

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

Another table and graphs that we could create is to create one that also considers the average donations in different countries considering different categories and sub-categories. This would also help to see how much this would affect the end outcome and the end revenue.

Part 2

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

The variance of our dataset shows that the data has a high variability. In such cases where there is a very high value or a very low value, it is better to use the median to summarize the data. The median does not get affected by those extreme high or low values compared to mean which is exactly affected by them.

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

Based on our variance and standard deviation results, it is shown that the successful campaigns have a higher variability compared to the failed campaigns. It does makes sense since comparing the values it is shown that more backers were involved in the successful campaigns and that makes the data of the successful campaigns to have more variability.