1. **What are the three conclusions we can draw about crowdfunding campaigns?**

The first thing noticed in the data provided is that the number of campaigns remained relatively stable during the years. This moved from 106 in the first year to a minimum of 81 in 2013 and rebounding to a mean of 103 in the following 6 years, until 2019. Year 2020 is not comparable cause the database only has data for the first two months.

Crowdfunding campaigns seemed like an attractive option for people who needed funds but, based in this sample, we can expect an overall successful rate of roughly 60%. Audio journalism and world music subcategories show the highest results, both categories surpassing their goals in all their campaigns. In contrast campaigns aiming to fund mobile games and science fiction projects are a long shot, with only 31% and 36% chance of success. Thus, this method, as a financing option would have to be approached with caution in those cases.

It is easy to notice that categories on the entertainment business (theater, music and film and video) launched the majority of the campaigns in this format, accounting for 70% of the total number of campaigns (not necessarily in amount) in this period of time.

Lastly, looking at the data, it stands out that the US is the country where most of the campaigns took place. This can be either because more US people and/or companies seek this kind of financing option or the site is focused on the US markets.

Summarizing:

1. The number of crowdfunding campaigns has remained relatively stable in the last years.
2. Seeking to capitalize through crowdfunding is not a secure option.
3. Depending on the category chances of a success campaign can vary widely.
4. The entertainment industry sought the most this kind of financing tool.
5. The US is the country that launches more of the campaigns.

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

Some of the limitations of the database include:

1. The sample contains data from 7 countries, leaving out the rest of the world.
2. In addition, it only focuses in 9 different categories, excluding industries as education, sports, agriculture, etc.
3. The dataset doesn’t consider other valuable indicators as: the kickoff money invested for the launching and the publicity done for the project, to mention a few.
4. There are some projects labeled as “live”, even though they were begun and closed on different periods of time. There is no other outcome stated for these campaigns.
5. It is not clear the meaning of some of the columns in the set (staff pick and spotlight, for example).
6. **What are some other possible tables and or graphs that we could create and what additional value would they provide?**

Besides the tables done in the first part of this exercise we could calculate the average pledge per category (and/or subcategory) and look if there is any relation between these concepts. In case it is positive this can be used to estimate the average pledge needed for success. In the same tone looking into the relation between number of backers and outcome could be interesting.

We can differentiate projects by goals and comparing the amount of money pursue to their possibility of success. Also comparing duration of the campaigns and outcome can be another useful indicator.

**STATISTICAL ANALYSIS**

1. **Determine whether the mean or the median better summarizes the data.**

Looking at the data with a Whisker box it is evident the presence of many outliers, making this a asymmetric set of data. In fact, in this dataset, there are very few campaigns with more than 2,500 backers. As the difference between the min and max number of backers is so big Mean calculations are too high. Based on this a good option is to look to the histogram for the data, where we can see that 332 out of 565 successful campaigns had less than 250 backers each (which means 59% of the data).

Using the MEAN of 851 would be overestimating the number of people backing these campaigns and therefore using the MEDIAN is a better estimate, making more sense for this analysis.

The same situation is observed for the Failed campaigns where 222 out of 364 campaigns have less than 250 backers (61%).

In both cases the MEDIAN summarizes the data in a better way than the MEAN.

1. **Determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?**

At first sight Failed campaigns have less variability that the successful ones. For these campaigns the number of backers runs from 0 to 6,080 (with a range of 6,080), while the Successful ones have supporters from 16 to 7,295 (with a range of 7,279).

Even that we have other central tendency indicators for this sample: like mean, median and variance it is difficult to make conclusions with this information as the distribution of the data is not normal. Therefore, we cannot apply parametric tests to compare the discrepancies on both samples and we cannot conclude anything for certain.