**Camilla Inhapim. Challenge 1:**

1. **Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?**
2. The theater category has 34% of the projects, being the highest category in quantity of projects but with a rate of 54% success projects, it’s only the 7 º highest success rate. The journalism category, on the other hand, has a low quantity of projects, 4 only, but they have a 100% of success rate. That way, we can infer that a high quantity of projects in a category does not mean a necessarily high successful rate and vice-versa.
3. The plays correspond to 34% of the sub-category projects, above the average of 4% of contribution of the other sub-categories. The US has the most projects per sub-categories, 76%. It has the 2º best success rate among the countries, with 57%. The highest success rate among the countries is GB and the lowest CA (58% and 50%, respectively).
4. Overall, 57% of the projects are successful. June, July and September are the months with the highest rate of success (63%,62% and 62%, respectively). The average (mean) of projects per year is 90.90. That means that 2020, with only 2 projects, is way far from the mean and it could be considered abnormal, an outlier.
5. **What are some limitations of this dataset?**

The dataset is comparing 7 currencies from different countries, neglecting their difference of value and purchasing power. Also, the dataset has been captured throughout 10 years and the value of the currencies can be impacted by inflation in this time.

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

We could create a chart displaying the different outcomes per years and per countries as to have more details of which country participation on the successful and failed outcome throughout the years. We could also create control charts or box plots to display the outliers on the data.

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

Since the dataset has too much noise (as per the high variance and high standard variation), it’s not symmetrical and have clear outlines that can mislead the mean calculation, it would be better to use the median to summarize the data (Bobbitt, 2021).

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

There is more variability in the successful projects since the variance and standard deviation are higher there than in the unsuccessful projects. Drawing a control chart, we can observe that the successful projects have more outliers than the unsuccessful projects, therefore the successful projects have more variability. Adding context to the data analysis, we could infer that people would have several different reasons to invest in crowdfunding projects throughout the years, particularly if we consider the numbers of countries, categories and sub-categories available, leading to a large variation in the successful projects.

# **Works Cited**

Bobbitt, Z. (2021, May 12). *When to Use Mean vs. Median (With Examples)*. Retrieved from Statology: https://www.statology.org/when-to-use-mean-vs-median/#:~:text=It's%20best%20to%20use%20the%20mean%20when%20the%20distribution%20of,when%20there%20are%20clear%20outliers.