**Assignment 1**

**Person: Vy Nguyen**

**Date: 05/10/2023**

Given the provided data, three conclusions that we can draw about crowdfunding campaigns are:

1. There were a large number of projects in the “theatre”, “film & video” and “music” categories (Figure 1).

Figure 1. Stacked bar chart showing the number of projects per category and the funding outcome per category.

1. There was a significant number of projects related to the “plays” sub-category (Figure 2).

Figure 2. Stacked bar chart showing the number of projects per sub-category and the funding outcome per sub-category.

1. On average, the number of successfully funded projects was still higher than the number of failed projects.

Figure 3. Line chart showing the average number of projects successfully funded or failed to be funded or cancelled.

Some limitations of the dataset are:

* This is a small data set with only 1,000 projects examined.
* Data were collected from only seven countries and therefore might not be suitable to predict the funding trend globally.
* The currency is different depending on the country for “goal”, “pledged” and “Average Donation”, hence cannot use the direct value in these columns to compare between countries.

Some other possible tables and/or graphs that we could create, and the additional value they would provide:

* To identify which country had the largest number of participated projects, we can create a table/ stack bar chart showing the “country” and a “Count of outcome” by category (Figure 4).

Figure 4. Stack bar chart showing the number of the project and the funding outcome by country.

* To identify which country has the highest number of successfully funded projects in each category, we can create a “100% stack bar chart” showing the percentage of successful, failed, live and cancelled projects by country.

Figure 5. “100% Stack bar chart” showing the percentage outcome of the projects by country for the “theatre” category.

Bonus:

The currency is different for different countries and therefore I don't think the analysis suggested for this section is appropriate unless all the value in the "Goal" column is converted to the same currency.

Bonus statistical analysis:

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

The mean and the median number of backers in the successful outcome group are highly different from each other, which indicates there is a skewness in the data, and therefore using the median is better to summarise the data.

Similarly, The mean and the median number of backers in the failed outcome group are highly different from each other, which indicates there is a skewness in the data, and therefore using the median is better to summarise the data.

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 number of backers with successful campaigns (var = 1606216.594) than the unsuccessful campaigns ( var = 924113.455) at p = 7.87555E-09 (F-Test Two-Sample for Variances). The larger recorded number of campaigns in the successful group (565) compared to the failed group (364) could be the reason for the increasing variability of the number of backers observed in successful campaigns.

~~~ END ~~~