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

We can conclude that projects created with goals greater than 49999 have less chance of success than all others. Projects created with goals between 10000 and 14999 also presented a high percentage of failure, however the number of total projects within this range is rather small and may not be robust enough to allow drawing conclusions.

When it comes to analyzing the data by sub-category, we can say that most campaigns were under the sub-category “plays”, but only 54% of them were successful. Similarly, when analyzing the data by category, we can conclude that most campaigns involved film & video, music and theater, with the latter being the most numerous. The quantity of theater campaigns was approximately the double of the other two, when comparing them individually. However, campaigns under all three categories presented a failing rate above 42%.

1. What are some limitations of this dataset?

The number of campaigns under some categories and sub-categories may be too small, which may affect the results if they are underrepresented. The empty fields (missing data) may also bring some inaccuracy to the analysis and make it less reliable as it may introduce bias. For example, the average donation per category and sub-category was not included. Finally, outliers need to be looked at with caution, as they can distort the analysis.

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 table including the average donation and comparing these values against backers\_count and analyze how they vary. For example, this could allow us to analyze the average donation in each category and sub-category over the given period. An additional stacked bar chart could be used to represent the data in such a way that we could visualize how the percentage of successful, cancelled and failed projects change over time.

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

Regarding the bonus statistical analysis, in this case the mean better summarises the data because the count of backers does not present outliers that could significantly impact the mean.

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 with successful campaigns. It does make sense because the population variance and sample variance are higher than the variance with unsuccessful campaigns, which means that the data is more spread out within the successful campaigns.

**I would like to make an observation here and I apologize if this is not the appropriate space to do so. When I upload my Excel file on Google Drive, some sheets, graphs, charts and tables change completely for some reason. I hope this do not happen when I upload them on GitHub. Could you please contact me if this does happen? Please see some examples below. Many thanks!**

How it looks like on my computer:

A screenshot of a computer

Description automatically generated

How it looks like on Google Drive (graph changes!)

A white sheet with blue and orange lines

Description automatically generated

How it looks like on my computer:

A screenshot of a computer

Description automatically generated

How it looks like on Google Drive (the data disappears and lots of errors appear!)

A white sheet of paper with a ladder

Description automatically generated

How it looks like on my computer:

A screenshot of a computer

Description automatically generated

How it looks like on Google Drive (missing the filter):

A screenshot of a computer

Description automatically generated

How it looks like on my computer:

A screenshot of a computer

Description automatically generated

How it looks like on Google Drive (missing the filter):

A screenshot of a graph

Description automatically generated