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

From the data generated, we can conclude that:

* From graph 1, the categories that stood out are film & video, music and theatre. These three categories have the most backers’ interests.
* Secondly, among the 1000 crowdfunding projects, over 700 projects are from the US.
* Thirdly, throughout all years, in the month of June and July, the crowdfunding successful rate are the highest.
* Fourthly, the higher the goal is, the lower the successful rate is.

1. What are some limitations of this dataset?

* The data sample size is too small to determine a more precise trend for instance to predict the future trend and hard to draw a more dependable result
* As mentioned, among the 1000 projects, over 700 projects are from US which the data variability or area is too limited.
* Graphical user interface, application

  Description automatically generated

Parent categories are too limited. Take Indiegogo as an example, the categories are much wider than the collected data which we could analyse a more accurate trend.

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

* Scatter plot with trend line would be one of the options as it can identify a positive or negative relationships between different variables
* The additional value to add is how are the projects funded as this may let us know more about where the backers are from and can conclude another trend.

**Bonus Statistical Analysis**

Use your data to determine whether the mean or the median summarizes the data more meaningfully.

* From the data, median could summarize the data more meaningful as it will not be influenced by extreme data, but the mean will as the difference between the minimum and maximum backers count is quite large.

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

* From the data, there is more variability with successful campaigns. As comparing the standard deviation and variance, successful data partially makes sense.
* When comparing standard deviation, the larger it is, the more variable the data is.
* Meanwhile, comparing variance, the more spread the data, the larger the variance is in relation to the mean.
* The reason for mentioning it is partially making sense is because using interquartile range is a better options to investigate or compare the relations and will be least affected by extreme data (outliers).