# Module 1 Challenge

## Background

Crowdfunding platforms like Kickstarter and Indiegogo have been growing in success and popularity since the late 2000s. From independent content creators to famous celebrities, more and more people are using crowdfunding to launch new products and generate buzz, but not every project has found success.

To receive funding, the project must meet or exceed an initial goal, so many organisations dedicate considerable resources looking through old projects in an attempt to discover “the trick” to finding success. For this week's Challenge, you will organise and analyse a database of 1,000 sample projects to uncover any hidden trends.

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

* Most crowdfunding campaigns belong to the categories of Theater (34%), Film and Video (17.8%), and Music (17.5%); with a combine success rate of 60%.
* Among the 1,000 crowdfunding campaigns studied, a notable 763 originated from the United States. Notably, 436 of these campaigns achieved success.
* The initiation of crowdfunding projects exhibits consistency, averaging at 83 projects per month. Furthermore, July emerges as the month with the highest incidence of successful campaigns, as discerned from the data analysis.

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

The analysis is highly influence by the high number of projects in the United States, and the conclusions drawn from the results could not necessarily extend to other demographics. Projects created on countries that are outside the evaluated scope could exhibit a different behaviour.

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

* Create a Pivot Table to analyse the daily average goal based on the creation and deadline dates to determine if the amount of required funds has an influence on the probability of success/failed.
* Create a 100% Stacked Column Chart to evaluate if the “spotlight” campaigns are more propense to succeed.
* Create a pivot table to analyse the success percentage of a campaign by Sub-category and determine which category would have a greater probability of success.

## Bonus Statistical Analysis

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

Given the high variability (high standard deviation), the amount of outliers present, and the skewness of the data distribution, the median is a better suited statistical measure of central tendency.

**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 on the backers count for the successful campaigns (higher variance and standard deviation). This variability could be explained by the initial assumption of crowdfunding in which everyone can donate to every project (there is no upper limit for the backers count), which translates in a distribution that have high probability of outliers (specially for successful campaigns).