**Data Analysis for crowdfunding campaign**

We have crowdfunding data giving us the goal set up for each project and how much it was collected, this will determine the status of these campaigns from being successful, failing or canceled.

The data also provides other elements of information like the number of backers and the average donation by backer, category and sub-category of each project.

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

Given the provided data and the charts made we can see:

* The more extreme the goal is the more risks of failure it has, we can see in the chart Outcome based on Goal that the when the goal is smaller (beginning of Axes lines) the lines of succussed projects and failed projects are close to each other, similarly to when it’s a large number, the lines get closer and crossed at some point.
* Theater projects have the bigger number of projects and still have the largest amount of failure.
* Regarding the country and especially the geographic regions, I see that crowdfunding music projects is more successful in north America while theater is the area where it is more successful in Europe.

The conclusion that may be given is that the smallest /bigger the budget is the smaller are the chances to succeed in raising money in a crowdfunding projects.

* What are some limitations of this dataset?

In the data set as it was given, we are missing the duration of each crowdfunding campaign, we can add that in a new column and make the difference between Date ended and date launched (column P named Duration per day)

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

We can do a chart about the outcome of the campaign and its duration, so we can make a probable link between how long the campaign took and its outcome.

**Statistical Analysis**

In the worksheet number of backers, we can see the box and whisker plots of outcome that there is a lot of outers in the data, meaning that our data is very spread out, so I think that we can’t say Mean is better than median but using both would be the best alternative.