1.

* We can clearly see with the pivot charts that higher number of successes does not necessarily mean high probability of success. Whilst, for example, theater productions have the highest number of successes (839 successful campaigns), this number has only been achieved because theater has a much larger number of total campaigns (1393 total campaigns). This means that theater also has the largest number of failed campaigns (493 failed), leaving a success ratio of ~60%. The category with the highest percentage of successful campaigns would be music, with a ratio of ~77% (540 successful campaigns out of 700). These are not conclusions we could have easily found had we been simply looking at the counts on the table.
* As we adjust the table to individual nations, we can see similar looking figures throughout the world. The Category which appears to have the more diversity of invest as we change countries, however, is technology, with some countries being their largest investment (Germany, Netherlands, France) whereas some others have zero investments (Hong Kong, Norway, Luxembourg). This makes sense, as the presence of a technology sector varies widely across the world. Countries with high numbers of overall campaigns will have more diversity in their campaign Categories and Sub-Categories, whereas those with lower numbers will have more constricted Categories.
* Our Start-Date Pivot-Chart shows that the count of successes, failures, and cancelations year is fairly stable throughout the, with the largest visible variance being with the successful campaigns. Successes find their peak in May, with a count of 234 campaigns, with a steady decline through the summer until September with 147 campaigns. Successful campaigns make a dramatic drop to 111 campaigns in December. Failed campaigns reach a jump upward in January, with 149 failed campaigns, and reach a max in June and July, figures 147 and 150, respectively. Aside from this, there do not appear any significant patterns in the data. The successful campaigns nearly always outnumber the failed, with the exception of December, wherein successes just fall short of failures by only 7 campaigns.

2.

There is not enough adequate information about individual campaigns, to extract reasoning behind its status being success or failure. Yes, there is the campaign name and a short blurb, and one could attempt to look at individual blurbs and attempt to use common sense to guess its failure probability, but this is not information which we could analyze with a graph, in which we could more accurately predict any individual project’s chance of success. We have no way of knowing, for example, which of our Theater Plays, as shown on our sub-category pivot table, are representative of successes, and which are of failures. This leads us to make the only possible conclusion that a play’s chance of success is simply random, roughly 50%, which is not a conclusion that we data analysts are satisfied with finding.

3.

I found, as I was working on this project, that an important attribute to understand from this data was the percentage of a Category/Sub-Category’s successes of its total. For this reason, any chart with the purpose of displaying percentages, such as a pie-chart or a 100% stacked column, would be useful.

Bonus Statistical Analysis

Of both the mean and median, the values for successful campaigns are significantly larger than those for unsuccessful campaigns. For successful campaigns, I would say the mean summarizes the data better, as it shows how the large data points of individual campaigns garnering large numbers of can heavily impact the likelihood of a campaign succeeding. For the unsuccessful campaigns, the median summarizes the data more meaningfully, as it emphasized just how many failed campaigns had very few backers. However, for both successful and unsuccessful campaign, I would say the most meaningful aggregate value to summarize the dataset would actually be the mode, as for both datasets it displays both the fact that successful campaigns need significant numbers of backs (17 backers), and the fact that unsuccessful campaigns are unlikely to have support at all (0 backers).

Both datasets have significantly large variances, most likely cause by large counts and ranges in our values. This leads variance to be somewhat unuseful. The standard deviation, which is typically more usable for looking for data, is more practical in this exercise. Of our two standard deviations, 844 backers for successful campaigns and 61 for unsuccessful, the standard deviation for successful campaigns is more variable as it is larger. This makes sense, as the dataset for successful campaigns is both larger in count and has more variety in the data values than that of the failed campaigns.