**Crowdfunding Goal Analysis Report**

Since the late 2000s, crowdfunding has been growing in success and popularity. As a result, many people are now using this practice to launch new products and generate publicity. In this challenge, I analyze crowdfunding data from a database of 1,000 sample campaigns to better understand its trends. As the first step, I implemented the spreadsheets and their accompanying tables and graphs necessary for the challenge: Crowdfunding, Sub-Category by Outcomes, Primary Category by Outcomes, Months by Outcomes, Crowdfunding Goal Analysis, and Statistical Analysis.

Subsequently, I merged the pivot tables from the Primary Category by Outcomes and Sub-Category by Outcomes and calculated the likelihood of a successful campaign per category; but omitted primary categories and sub-categories with less than five campaigns: due to low numbers, these categories provide an insufficient sample space for evaluation (see spreadsheet, Both Cat. By % Success). From my observations, a crowdfunding campaign in a primary category has a 56% chance of success with Technology, Photography, and Publishing at the top; sub-category campaigns also have a 56% chance of funding where Web (Technology), Translations (Publishing), and Television (Film & Video) are the leaders. In terms of sheer numbers, Theater, Film & Video, and Music are the most popular primary categories by significant margins; Plays (Theater), Rock (Music), and Documentary (Film & Video) are the most sought-after sub-categories. Although this effort divulged some insights, there was still an incomplete picture.

To inquire further, I created a pivot table with categories versus the sum of the campaign goals, the sum of the pledged contributions, and the percent of the pledged contributions over the goals (see spreadsheet, Both Cat. By % Funded). Percent funded above the goal for all campaigns regardless of status or outcome is 97% with Technology, Music, and Theater at the top; the highest sub-categories are Metal (Music), Wearables (Technology), and Jazz (Music). For successful projects, the overall funding rate is 191% with the primary categories of Food, Games, and Photography and the sub-categories of Metal (Music), Jazz (Music), and Science Fiction (Film & Video) being the highest: in particular, among the primary categories, successful Games and Food campaigns receive 2.1 and 2.3 times their goals respectively; among the sub-categories, successful Jazz (Music) and Metal (Music) campaigns receive funding at 3.5 and 4.4 times their goals.

A cursory glance at the spreadsheet, Months by Outcomes, yielded no immediate understanding: the static nature of these metrics called for transformation into a more apprehensible form. To this end, I created the spreadsheet, Months by % Success, with the success rate per month included with the original information. These rates vary from 52% to 64% on a month-to-month basis where they rise from January to June, plateau until September, then steadily decline. This trend has glaring exceptions in May and August where the rates plummet 5% and 13% from the previous month. In this context, it appears that the possibility of a successful campaign can be increased through timing.

In the spreadsheet, Crowdfunding Goal Analysis, I examined outcomes based on ranges of goals only to encounter specious results. In the graph, the success rate’s trend increases as the funding moves from $0 to $15,000, levels off between $15,000 and $34,999, and then declines from $35,000 on. Overwhelmingly, the lowest three categories have the highest number of projects: all the categories equal to or above $10,000 have less than 15 projects compared to those below $10,000, which have two categories with hundreds of projects and one with about 50. For the larger goals, these low numbers of campaigns per category provide an insufficient sample space, which challenges the integrity of the analysis results; unfortunately, these circumstances are present in other parts of the dataset as well.

For successful and unsuccessful campaigns, the Statistical Analysis spreadsheet exhibits distributions for the number of financial backers with significant ranges between minimum and maximum values, large variances, and considerable differences between average and median values. When the average and median values deviate as they do in this case, the distribution is skewed, and the average no longer is an adequate representation of the typical or middle value: in this instance, the median becomes the more suitable measure and better summarizes the data. Regrettably, the calculations in this analysis employ averages when they should apply medians. Moreover, the variances for successful campaigns are larger and, consequently, indicate more variability than those for unsuccessful ones. Clearly, variance measures heterogeneity: a higher variance signifies the values are more spread out, and a lower variance denotes that the values are closer together. This determination makes sense because there is a large spectrum of backers for successfully funded campaigns while unsuccessful ones have concentrated numbers of backers closer to and including zero.

Therefore, along with obvious deductions about success and funding trends, I drew the following three conclusions about crowdfunding campaigns.

* The probability of a successful campaign in a certain category is not related to that category’s popularity: the categories with the most campaigns are not necessarily the most successful ones.

* The likelihood of a successful campaign can be increased through timing.
* The variability in successful campaigns is greater than that of unsuccessful ones because the values in the distribution are more spread out in the former and congregated and closer to zero in the latter.

Additionally, the two distributions of backers, for successful and unsuccessful campaigns, are heavily skewed, which the egregious deviation in average and median values conveys. Hence, in this case, the median is a better measure of central tendency than the average, but the average is applied instead. This situation coupled with insufficient sample spaces discussed in the report limits the utility of this dataset.

Above and beyond the analysis conducted, there are numerous other tables and/or graphs that we could create. For the sake of brevity, I will mention only three: countries vs. percent of successful campaigns, year vs. percent of successful campaigns, and outcomes vs. standard deviations of goals and pledged contributions. In countries vs. percent of successful campaigns, the table and graph communicate if campaigns are more likely to succeed in one country over another (see spreadsheet, Countries by % Success). For year vs. percent of successful campaigns, the calculations demonstrate whether crowdfunding is becoming increasingly more popular or following a cyclical pattern (see spreadsheet, Years by % Success). Finally, outcomes by the standard deviations of goals and pledged contributions will reveal if anything differentiates distributions for a successful campaign over the alternatives; there is some anomaly: the standard deviation of campaign goals is greater than the standard deviation of pledged contributions in all cases except successful campaigns where the reverse occurs. What’s more, the standard deviation for a campaign goal is approximately the same for all outcomes except successful campaigns where it drops precipitously (see spreadsheet, Outcome by StdDev Goal-Pledge).

Altogether, this investigation yielded practical results, but the limitations of this dataset render those results suspect. In the future, I would suggest a similar analysis with more tables and graphs on a much larger dataset – preferably, one to two orders of magnitude larger – using medians for calculations instead of averages.