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

- The conclusion that can be drawn depends heavily on the uestion one wants to ask. Let us assume the inquiry us from a potential business looking for alternatives to traditional funding streams.

A category-level analysis (see tab 01-Per Category) would show that most Kickstarter projects are in the entertainment sphere: games, music, theater, and film & video collectively account for 2833 listed projects or 68.86 percent of projects on Kickstarter. Distilled, these categories account for 1759 successful campaigns or 80.50 percent of successful campaigns. This would suggest that popular culture-adjacent projects represent a higher percentage of successful projects than total projects overall which is consistent with finding that their success rate (62.08%) is better than for all Kickstarter campaigns recorded in the data (53.11%). Graph 1,1 supports this as it clearly shows pop culture categories becoming smaller proportion of failed and cancelled campaigns than successful campaigns.

Analyzing at the Subcategory level (refer to Tab 2) can support this supposition. When we filter Graph 2.1 to focus on the four “popular culture” categories, we see the same big difference in success and failure rates at the subcategory level. For example, the categories of Plays (under theater), and rock (under music) are the largest subcategories; and their stacked columns show substantially more successful campaigns than failed campaigns. Of course, certain pop culture subcategories like animation video)(under film & video)and world music (under music) also have near 100 percent failure or cancellation rates which suggests that there is nuance to be found in subcategory-level analysis.

Further analysis, as shown in Tab 3, suggests that the Month of a campaign’s launch has no significant impact on success rate. While Spring and Summer campaigns appear show more successful and failed campaigns as compared to fall and winter campaigns, as illustrated in Graph 3.1, where successful campaigns outnumber failed campaigns overall throughout the year except for December. If broken down at the category-level, however, the data suggests that various categories of campaigns have varying levels of success throughout the year. For the popular culture entertainment categories, the data suggests that campaigns started early or in the middle of the year tend to have higher success rates than later in the year.

Overall, the data suggests that Kickstarter can provide alternative funding for popular culture; but it might nt necessarily displace the role of more sophisticated funding models like Venture Capitalists in technology fields.

When we break down summary statistics (refer to Tab 4 – Summary Table), Table 4.2 shows that that there is a notable difference between failed (µ=17.71, б=82.43)and successful (µ=194.43, б=844.30) campaigns. Judgubg by their standard deviations alone (б =844.30, 82.43), the two groups do not even meet the assumption of homogeneity of variances that is necessary for conducing parametric tests for inferential statistics. However, it is enough to support the intuitive idea that successful campaigns on Kickstarter will, on average attract more individual pledges than unsuccessful campaigns.

Table 4.1 shows a breakdown for Kickstarter campaigns and their outcomes according to their original funding goal. This data supports the assertion that the majority of Kickstarter campaigns end up reaching their funding goals (Success Rate = 53%) with few campaigns setting their funding goals above $20,000.

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

For one, the data only concerns the Kickstarter phase of the listed ventures. For an entrepreneur considering alternative methods of financing on Kickstarter, the dataset offers no illumination on how sustainable the subsequent ventures become once they receive Kickstarter funding. For a prospective donor, the data offers no guidance on how likely they are to see a return on their pledges. For Kickstarter, this data offers no insight on the demographics of Kickstarter pledgers; or the average pledge. Admittedly, the anal

**3. What are some other possible tables and/or graphs that we could create?**

It is my view that Excel and Pivot Tables are not the proper tool for analyzing this data. It is interesting to know that success rates for “Pop Culture related campaigns are higher than those for other categories; but the information a decision maker needs can only come from a logistic regression analysis that shows how much the odds of a successful campaign improve when particular variables are increased. A table compiling summary statistics per category and subcategory might also be desirable for the purposes of comparing how well these campaigns do outside the narrow field of whether they meet their goal or not. With summary statistics, we can perform 2-sample t-tests for statistical significance which only require the means and standard deviations of 2-groups to perform. Additionally, it seems desirable to make more stacked-column charts breaking down data between failed and successful campaigns to see whether the groups are so different after all.

A scatterplot examining the relationship between the stated goal for a campaign and the amount it receives in pledges may also be worth doing. A similar scatterplot can also see if there is a relationship between the total amount of pledges in a Kickstarter campaign and the number of individual pledges

Another table can possibly examine whether the average donation is different in campaigns in a certain category than in another category. Lastly, it might be a good idea to breakdown the dataset into quartiles based on any f the quantitative data points such as amount of total pledges, original goal, etc. One such analysis based on the number of backers (see Tab 5) shows a significant amount of variability in the data. It can almost be argued that specific portions of the dataset behave like a totally different population than the bulk of the dataset. A more granular analysis can provide us more insight into these segments of the data.

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

To make this analysis, the data for successful and failed campaigns were each divided into quartiles based on their number of backers. In a population where the Mean accurately reflects the overall data, one expects the data points to follow a normal distribution where 68% of data points fall within one standard deviation of the mean. When a confidence interval around the mean number of donors for both successful (µ = 194.43, б = 844.30; 88%CI: = -649.47 -1,048.72) and failed campaigns (б=17.71, б= 61.43; 68% CI: -43.72 – 17.71) is calculated, both return lower bounds in the negative. A population where the mean is the best central value shouldn’t be so badly skewed as to return a confidence interval that is partly in the negative. Furthermore, division into quartiles reveals that the means reported for both groups puts the mean alues well outside the 4th quartile. Thound. This suggests that the data has a sf=significant positive ske, as demonstrated by Charts 5.1 and 5.2. Ina positive skew, you will find most data point clustered around the mean closer to zero; and a long tail of more dispersed values leading ever further. This long tail will tend to contain many extreme values that earp the mean, and standard deviation. Hence, the moreconservative median appears to be better represent this data set.

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

The standard measure of statistical variability is the variance (or mean sum of squares). Based on that measure, it is easy to make the case that the population of successful campaigns (var = 71,2840.99) has more variability than that of failed campaigns (var = 3,773.22.)

This makes sense intuitively as the range, the difference between the minimum and maximum values, is much larger in the former group (range = 26,456) as it is in the latter (range = 1293 . On a technical level, the variance is the sum of squared deviations from the mean do it can be expected to become inflated for larger ranges. On a more intuitive level, there are many paths to success and fewer paths to failure so successful campaigns might be expected to be different from one another.