**Mark Sutton**

**May 26, 2020**

**Kickstart My Chart**

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

Campaigns by Date

There appears to be a seasonal trend that occurs twice per year, a mid-year bump and a holiday bump, of campaigns launched. This can be seen in the total campaigns by month in the pivot table on the “Campaigns by Date” tab. As we see on the graph on the “Campaigns by Date” graph, the peaks of the failed and canceled projects show a lagging effect after the peaks of the successful campaign. This indicates that timing of the launch of a Kickstarter can affect the chances of success.

Outcomes Based on Goal

Looking at the “Outcomes Based on Goal” graph, we see a clear pattern that the smaller the monetary goal, the more likely the project is to succeed. And conversely, the larger the goal, the more likely a project is to fail or be canceled.

Sub-Category

In looking at the Sub-category pivot table and chart, it is interesting to note that some sub-categories in the sample had either a 100% or 0% success rate. For instance, Kickstarters for animation and video games failed 100 out of 100 times. Meanwhile, tabletop games and television project were successful 100% of the time. Therefore, it would be worthwhile for anyone considering a Kickstarter campaign to see the odds of other similar projects before deciding if it’s worthwhile.

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

This is a set of 4,000 projects out of a total of 300,000 that have been launched at Kickstarter. While this is a fairly large sample size, we don’t know how truly random it is. For instance, in looking at the Sub-categories chart, we can see that there are over 1,000 plays, far more than any other Sub-category. Is this indicative of the overall mix, or is our sample leaning toward plays?

Only one third of the total 300,000 projects are successful, yet in our sample, 2185 out of 4114 succeeded, or over 50%. Therefore, it was not a typical sample in that regard.

Also, while he have the information of total number of donors, and total raised, we don’t know if successful campaigns were likely to have a small number of big donors, with the rest donating modust sums, or if instead the successful campaigns drew larger donations from everyone. That is, we don’t have the standard deviation of individual donations.

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

We could create a box and whisker plot to see if there is any outlier data that affecting the numbers, particularly on the high side. For instance, were there a few projects that raised a large amount of money that skewed the average?

**Bonus Statistical Analysis**

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

In looking at the data from the pivot table and calculation on the tab labeled “OutcomeComps,” both the mean (194 vs. 18) and median (62 vs. 4) seem to indicate a clear trend in favor of the successful campaigns being higher.

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

We see more variability in the successful campaigns (STDEV of 844 vs 61), which makes sense. A successful campaign could be a small campaign that just achieves success, or it could be a large campaign that succeeds wildly beyond its goals. However, most failed campaigns would tend to have fewer backers, thus limiting the variability.