Kickstarter “Excel Challenge” Report & Analyses   
 by Ithamar Francois

Within this analysis, of the Kickstarter Excel challenge assignment, we hope to answer some fundamental questions regarding the data that we extrapolated. Firstly, what three conclusions could we draw about Kickstarter campaigns, based on the reported dataset. We will be looking at any trends or patterns that the data shows regarding the different types, or categories, of Kickstarter campaigns that creator’s attempt to run (and their degree of success) but also what do individual backers of projects tend to support and feel passionate about. We will try to draw three conclusions. Secondly, we will touch upon what the limitations are of this dataset and why it could be difficult in assuming the data paints an accurate picture of what is happening with campaigns and backers. And lastly, I will offer some suggestions of other possible graphs or tables that we did not get to utilize, that might have shed some insight on what the data is telling us to perhaps conclude a different analysis.

I will start with the second question posed, as that will lead us into what we can glean from all the data in order to answer the first question. Although we have an abundance of data regarding Kickstarter, the limitations of what we can conclude become apparent in several ways starting in fact with “abundance.” Kickstarter has had well over 300,000 Kickstarter projects that have been attempted by creators, but our dataset only covers about 4000 projects to base any sort of analysis from. This is only about 1% of the total campaigns which is not representative of what might be happening in Kickstarter outcomes. This is issue one. Another observation that we can make is about the years that campaigns were taking place. Our dataset, via the Date Created/Ended Conversion columns (“S” & “T”) in the Excel workbook, only cover projects starting roughly in 2009 and ending roughly in 2017, which is not even a decade in length. That is not a long time to start with, but other limitations can be surmised from there such as: how to factor the length of time a projects stay “active” for (regardless of eventual success or failure etc.), the fact that campaigns can be essentially restarted and copied that may skew data conclusions, not all projects are created equal as some benefit from factors such as word of mouth or front page promotion on the website, or that there may be a correlation between how long a project stays active and say “Kickstarter fatigue” among backers. For example, from the included “KS Launch Date Outcomes” worksheet, a year-by-year filtering of data to show comparison of total “successful” campaigns vs “fails” shows that from 2010 to about 2014, the number of “successful” campaigns dwarfs the “failed” amounts by great amounts. For instance, from 2010 to 2012, there was 401 successful campaigns compared to 100 failed and 14 canceled. But interestingly starting in 2014, the number of successful vs failed campaigns almost evens out, despite having more campaigns occurring. From 2014 to 2016, there were 1500 successful projects vs 1300 fails. What could account for such a drastic change? Market fatigue/saturation, economic factors, changes in laws, cultural shifts? The possibilities are cast, but the data is limiting. As you can see, there are so just too many factors that cannot be accounted for project to project.

So, what 3 things can we conclude tentatively with the data we have? First, according to the Goals Outcome Bonus worksheet, it seems that a correlation can be observed between how much money a project to aiming to raise vs its chance of success. The data seems to indicate that the higher the goal of the campaign (i.e. over 50,000 for instance) the more likely the project will not fare well with the number of successful campaigns dropping drastically (compared to successful campaigns in the less than 10,000 range) but also the number of failed/cancelled projects jumping up. So more expensive projects tend not to do well. Another conclusion, according to the KS Sub-Category Status worksheet, is that projects that are with the realm of the arts tend to be both plentiful and volatile, particularly the field of theatre/plays. Plays alone accounted for roughly 1060 projects with the next highest subcategory being “Rock” music campaigns at 260, for comparison. Of that 1060, roughly 350 failed, another 20 canceled, but 690 were successful. What could this imply? Perhaps plays are under funded in the marketplace so producers seek crowd sourcing to bring pieces to production? Perhaps KS backers really like live theater? Perhaps something else? The last observations come regarding “Food”. This category and subcategory (according to the pivot charts on the respective Status worksheets) is of note due to the large lack of success this field obtains in the data. For some reason, Kickstarter backers and donors, are not fans of food related ventures with crowdsourcing. One would assume that food, unlike say theater, is something everyone has a vested interest in one would think. On top of that, the restaurant industry outside Kickstarter shows there is no shortage of people who like food, but why the lack to success then? Could it be the nature of the projects? Could it be oversaturation in the market? Are the immediate benefits of a successful food campaign more tangible that say an album being made, or some new gadget? Could the subcategory itself not be conducive to an online space?

For the last question in terms of what other possible tables or graphs could we create to help sift through the data, I would propose a pie chart. On top of all the vast amount of data to collect, there was an additional challenge with the exercise of being able to see the data clearly. Sometimes bar graphs and line charts do not give the best visual snapshot of what patterns or trends we are seeing. Bar graphs, such as the worksheet involving category/subcategory, can be hard to parse through due to the limited amount of space and how data can skew, whereas a pie chart would have been a welcome sight (no pun) in order to see exactly how much a particular category’s share of the overall data can be imaged. The right visualization can allow the best analysis, and consequently, the best conclusions to be made.

As you can see, there are many questions that arise from looking at data more closely. This can be a double edge sword in this sense, but these tools that Excel provides are powerful in helping to ask the right questions and understanding the complete picture more accurately.