HW 1 Report: Kickstarter Campaigns

Kickstarter has seen varying results in campaign successes and failures. Different variables such as category and subcategory, start and end date, and goal amount effect the success or failure of the campaigns. Using the excel worksheet given, this report attempts to analyze the collection of data to discover trends on Kickstarter campaign successes and failures.

The data set is a collection of 4114 Kickstarter campaigns. It consists of the name and category/subcategory, as well as its state (successful, failure, live, or cancelled), the pledged amount and goal, a short description of the project, country and currency, launch and deadline dates encoding, staff pick, and spotlight. This data is a collected sample of campaigns from 2009 to 2017, spanning almost a decade, and goes over an array of categories; film and tv, food, games, journalism, music, photography, publishing, technology, and theater. Each of these categories have multiple sub-categories. This is a sufficient amount of information to run an analysis on the potential variables that affect that correlate to the success or failure of a project. However, there are some limitation to the document that would otherwise add further to the analysis and eliminate other external factors. I feel the most significant element not considered by the data set is the spotlight and what elements effected a campaign from being admitted into the spotlight category. A look at the data can show that there is a clear correlation between spotlight picks and successful campaigns, however there isn’t enough data to show why these specific campaigns were entered into the spotlight category. Another limitation are the incentives that each campaign offers. This can potentially be a significant factor in the success of a campaign, as many individuals will likely go for a project only for the incentives it provides. However, this sort of data qualitative and would likely need user survey results to evaluate. Even so, it is still a factor that could affect the success of the project and should be included to cover all variables

I used a mixture of functions and pivot tables with charts to analyze the data. First thing was to create a visual representation of the of the state of each campaign. To do so, I created conditional formatting rule for column F that would color the cells based on the words; green for successful, red for unsuccessful, yellow for cancelled, and blue for live. Then I calculated the percent funded for each of the campaigns in column O using the function =E#/D# (# representing the current row) and used a conditional format based on the scale of 0% to 200% (red for 0%, green for 100%, and blue for 200%). I then calculated the average donation in column P using the function =IF(E#=0,0,IF(E#>0,E#/L#)). Then I split the category and sub-category in column H into category in column Q and sub-category in column R using the Text to Column data tool. Next I created 2 sheets containing pivot tables, for category and for subcategory. For the category pivot table, the rows are the categories, the columns are the state, the values are the count of the states, and a bar graph was created using the data and created a filter by country. For the sub-category, the rows are the sub-categories, the columns are again the state, the values are the count of the states, and a bar graph was created using the data and created a filter by country and category. Moving back to the main sheet, I calculated the start date in column S and end date in column T to translate the timestamp from their respective columns using formula =(((##/60)/60)/24)+DATE(1970,1,1). I then created another pivot table looking at the correlation of date and success; the rows are the date created by month, the columns are the state, the value are the count of the state, and created a line graph adding filters for category and years. Finally, I made another sheet with the table analyzing the data for every 5000 currency intervals and the COUNTIFS function to calculate the number successful and failed along with the percent (i.e. =COUNTIFS('Kickstarted Table'!$F:$F,"successful", 'Kickstarted Table'!$D:$D,">45000", 'Kickstarted Table'!$D:$D,"<49999"). Using that data, I created a bar graph.

A few assumptions can be made from this data. Looking at the category statistics pivot table, we can see that the theater category has the largest amount of campaigns and journalism has the least amount of campaigns, 1393 and 24 respectively. This correlates to the total amount of successes and failures in each category. Theater campaigns have the highest amount of successes at 893 campaigns, while journalism on the other hand have all 24 campaigns cancelled. This would suggest that theater campaigns have the best chance of success due to its great amount of campaigns and its great percentage success to all rate. Next looking at the monthly and annual statistics pivot table, there have generally been a greater amount of successes than failures or cancellations; the highest disparity in the month of May and the point in which failures outweigh successes in December. Filtering through the years, we can see that failures outweigh successes in specific months throughout 2010 and from 2014 onwards. This suggests that while there has been an overall net success, the rate of failure has increased over the years. Now looking at the goal cost statistics chart, we can see the success rate and percentage. The highest amount of campaigns is within the 1000 to 4999 range; it’s not a surprise then to see it has the greatest number of successes by far out of the other categories. The least amount of campaigns is within the 40000 to 44999 range. By percentage, the campaigns under 1000 have the highest success rate at 71% and the campaign in the 35000 to 39999 range has the lowest success rate at 29%. This in total suggests that the lower the campaign funding goal, the greater chance it has for success.

Overall, we can see the elements for success in a Kickstarter campaign. Theater campaigns with a goal of under 1000 has the greatest chance for success, though success rates have decreased over time. Something to consider for future studies is how these specific elements, category and sub-category, date of entry and goal, goal amount, etc., interact with each other on a greater level. In that pursuit, further studies would require a more recent and a greater sample of data.