**Kickstart My Chart: Analyzing Trends in Kickstarter Projects**

Jerome Thresher Olsson1

1University of California, Davis Data Visualization Boot Camp, Part Time

**Abstract**

Given the provided data, three insights are immediately apparent. First, that theatrical productions, specifically plays, constitute a large majority of Kickstarter projects, both successful and failed. Second, that the size of goal only seems to decrease the likelihood of success at the highest ranges. Third, that by any measure a higher number of backers seems to increase the likelihood of success for Kickstarter projects. Some important limitations on the data set, including its small size and limited time in which it was collected, means that any conclusions will be difficult to generalize for all Kickstarter projects. Some other possible graphs and tables that could be created include the average number of backers per subcategory, average donation per subcategory, the minimum count of backers per subcategory, and the minimum average donation per subcategory. When considering the central tendency for the number of backers for successful and failed Kickstarter projects, the median is more useful than the mean because of the number of backers. In addition, successful campaigns have a higher variance and standard deviation in their number of backers than failed campaigns, which indicates that after attaining a minimum number of backers for viability, there is a greater allowance for successful projects in terms of how many backers they have.

*Keywords:* Microsoft, Excel, data analysis, data visualization, Kickstarter.

**Kickstart My Chart: Analyzing Trends in Kickstarter Projects**

**Examining the data: first steps**

First the Excel table provided, modify and analyze the data of 4,000 past Kickstarter projects as you attempt to uncover some market trends. Use conditional formatting to fill each cell in the state column with a different color, depending on whether the associated campaign was successful, failed, or canceled, or is currently live. Create a new column O called Percent Funded that uses a formula to uncover how much money a campaign made to reach its initial goal. Use conditional formatting to fill each cell in the Percent Funded column using a three-color scale. The scale should start at 0 and be a dark shade of red, transitioning to green at 100, and blue at 200. Create a new column P called Average Donation that uses a formula to uncover how much each backer for the project paid on average. Create two new columns, one called Category at Q and another called Sub-Category at R, which use formulas to split the Category and Sub-Category column into two parts.

**Initial data analysis: state of projects per category and subcategory,**

Create a new sheet with a pivot table that will analyze your initial worksheet to count how many campaigns were successful, failed, canceled, or are currently live per category. Create a stacked column pivot chart that can be filtered by country based on the table you have created. Create a new sheet with a pivot table that will analyze your initial sheet to count how many campaigns were successful, failed, or canceled, or are currently live per sub-category. Create a stacked column pivot chart that can be filtered by country and parent-category based on the table you have created. The dates stored within the deadline and launched\_at columns use Unix timestamps. Fortunately for us, there is a formula that can be used to convert these timestamps to a normal date. Create a new column named Date Created Conversion that will use this formula to convert the data contained within launched\_at into Excel’s date format. Create a new column named Date Ended Conversion that will use this formula to convert the data contained within deadline into Excel’s date format. NEW COLUMN T but is this supposed to be the same sheet? Create a new sheet with a pivot table with a column of state, rows of Date Created Conversion, values based on the count of state, and filters based on parent category and Years. Now create a pivot chart line graph that visualizes this new table.

**Data analysis continued: state versus number of backers, central tendency and variance of number of backers by state**

Using the COUNTIFS() formula, count how many successful, failed, and canceled projects were created with goals within the ranges listed above. Populate the Number Successful, Number Failed, and Number Canceled columns with this data. Add up each of the values in the Number Successful, Number Failed, and Number Canceled columns to populate the Total Projects column. Then, using a mathematical formula, find the percentage of projects that were successful, failed, or canceled per goal range. Create a line chart that graphs the relationship between a goal’s amount and its chances at success, failure, or cancellation. If one were to describe a successful crowdfunding campaign, most people would use the number of campaign backers as a metric of success. One of the most efficient ways that data scientists characterize a quantitative metric, such as the number of campaign backers, is by creating a summary statistics table. For those looking for an additional challenge, you will evaluate the number of backers of successful and unsuccessful campaigns by creating your own summary statistics table. Create a new worksheet in your workbook, and create a column each for the number of backers of successful campaigns and unsuccessful campaigns. Use Excel to evaluate the following for successful campaigns, and then for unsuccessful campaigns. The mean number of backers. The median number of backers. The minimum number of backers. The maximum number of backers. The variance of the number of backers. The standard deviation of the number of backers. Use your data to determine whether the mean or the median summarizes the data more meaningfully. Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not? Higher variance and standard deviation with successful campaigns which makes sense because larger numbers of backers are generally required for successful campaigns.

**References**