**UC Berkeley Extension Data Analytics Program**

**Homework 1 Report**

Submitted by: Alejandro Montesinos

February 16, 2019

This report summarizes the conclusions from the data analysis exercise presented in the assignment of Unit 1 – Excel. The exercise consisted of analyzing data of projects launched by the crowdfunding service Kickstarter and identify trends using Microsoft Excel tools such as conditional formatting, pivot tables and functions.

The report presents the main conclusion in section 1. Section 2 discuss the limitations of the data set and finally Section 3 presents some suggestion for further analysis. The data analysis in excel is available in [GitHub](https://github.com/alejandromontesinos/UCB_DATA_H1_Excel).

1. **What are three conclusions we can make about Kickstarter campaigns given the provided data?**

Theater is the category with most number projects, 1,393 out of 4,114 (34%) and the one that contributes the most successful projects, 893 out of all 2,185 successful projects (41%) are in the theater category. However, in relative terms the category of Music has the highest successful rates. 77% of music projects are successful, followed by theater 60% and film and view 58%

In terms of subcategories, Plays is by far the one with the greatest number of projects, 1,066 out of 4,114 (26%) and the one with the greatest number of successful projects, 694 out of 2,185 () successful projects (32%). In relative terms, Rock, Documentary, Hardware, Tebletop Games, Non-Fiction, Shorts, Television, Classical Music, Electronic Music, Pop, Metal and Radio & Podcasts are very successful category showing 100% successful rate. It’s wort noting however that these categories are small in terms of number of projects (20-260). On the other end, Subcategories Animation and Video Games with 100 projects each are among the worst performers. All the projects in those categories failed (100% failure rate).

Finally, the size of the goal is inversely correlated with the rate of success. Of the projects with a goal of less than one thousand dollars, 71% of the projects were successful. In contrast, among projects with goals of more than 50 thousand dollars, only 19% were successful.

Based on the data analyzed so far, projects with modest goal amounts and in the Music category are the most likely to succeed.

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

Though the data does provide the name of the project and a blurb it does not provide the name of the project owner. The successful rate of the Music category could suggest that people may already know the work of the musician and decide to fund the project if they like his previous work.

Therefore, having data on the project owner would be useful as it could help understand additional trend associated with success. For example, whether the project owner has already work that is known to the public, how the public rate past work, whether the project owner has media exposure (youtube, facebook), if the project owner had previous projects funded in Kickstarter, whether it succeeded o failed to complete the project as initially planned.

Finally, information on the backers would be useful, such as demographics. If a particular demographic group dominated the backer populations, it is very likely that projects in the topics that are of interest of that demographic groups also attract more funding.

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

Spotlight, Percent funded, and length of fundraising period are promising variables to explore further the determinants of success.

From a quick vies it seems most successful projects have spotlight = TRUE. Therefore, a table that provides counts of Outcome by spotlight will provide and insight of the relationship of these variables.

By tabulating the percent funding by category and sub category will help identify what are those one that usually exceeds the goal by a range of margins. On the opposite direction, it will help identify whether there are specific categories/subcategories that usually fail to meet the goal.

Finally, by combining the information on the launch date and the deadline date we can compute the length of the fundraising period. And tabulate that against the state of the outcome and also the range of the goal. Studies in fundraising suggest that shorter periods of fundraising, generally 30 days, combined with relatively small goal amounts is associated with higher odd of success. We could test if such a pattern is observed in this dataset.