Week1 Challenge

By: Vicky L

Crowdfunding REPORT

Vicky Lasota, Data Science Student

Northwestern
DATA SCIENCE & VISUALIZATION
BOOT CAMP

CONTENTS

Crowdfunding Campaign Results

Conclusions - Executive Summary

The analysis conducted on the Crowdfunding dataset allowed investigation of the types of crowdfunded campaigns by category, success rate, time of delivery, and country of origin. The US accounted for 75% of all reviewed projects and took second place in the overall campaign success rate at 69%. While projects originating in Great Britain constituted about 5% of the dataset, they were the most successful – showing a 74% success rate.

Based only on the prescribed analysis, the projects in categories theater, film & video, and music constituted 70% of all projects with “plays” leading among all subcategories (35%). It is worth noting that while the most initiated, the success rate of the three winning categories (56% average) was proportionally lower than the success rate of some of the less popular categories such as technology and photography (64% average).

A general overview of the project success rate against month of the launch showed July and January as the most popular months to launch a campaign - September and November as the least popular. Interestingly, projects initiated in June have the highest (64%) success rate while those initiated in August the lowest (49%). In addition, projects launched in August have the highest risk of cancellation (11%), while those launched in April have the lowest, 1%.

Maximize your chances of success – based on data drill down only:

1) Launch in the US or GB
2) Try your luck in tech
3) Start in July

Further Analysis - Recommendations

For further analysis it would be beneficial to review more information relating to funding and overall financial support for different projects. The analysis would allow to gauge interest and niche opportunities based on the provided data set.

Some of the recommended graphs/tables could include:

- 100% stacked column charts to visualize the proportion of successful projects per category with a country filter and subcategory drilldown. This would allow easy comparison of the proportion of successful projects for each category.
- Line graph showing dollar/backer spent across different categories with a country filter and subcategory drilldown – comparison would first require currency exchange computation within the dataset. This graph could potentially show which projects garner highest per person donations potentially revealing something about populations supporting specific categories and direct marketing strategy.
- Boxplot charts showing outlier projects in terms of per backer spending, and/or size of the project. This would allow better sampling to answer specific question on the nature of populations and their preferences.
- Scatterplot to search for dependencies between the project success and length of the campaign. This could serve as a way of finding other outliers in each category, but could also provide information on the optimal length of the campaign.
- Scatterplot showing dependency between dollar per backer spent and number of backers to find projects and categories with the highest per backer spent with the highest number of backers.

Dataset Limitations

The data is relatively general and wide in scope with limited drilldown options. For a more detailed drilldown related to further subcategories, types of backers, marketing and promotion, more granular information or an additional supporting dataset would be needed. To learn more about the limitations, it would be necessary to learn more about how the data was obtained and its intended use.

At glance, the dataset seems to be largely skewed toward the US, which might be due to the source, type of data, or just the size difference between the countries included in the set. In an attempt to normalize the data it might be a good idea to translate some of the information provided into per capita or GDP-terms, or at least source more information regarding the state-level breakdown for the US subset.

The dataset also largely favors theater and film & video parent categories which again skews the data considerably.

Because of the skewing, it is hard to compare campaign categories in a meaningful way. Further analysis, potential exclusion of certain categories, or sourcing additional data might be necessary.