Nicholas Wiid

Khaled Karman

UCB-VIRT-DATA-PT-02-2024-U-LOLC-MTTH

5th March 2024

**Berkeley Data Analysis Bootcamp**

**Module 1 Challenge - Crowdfunding Analysis**

## **Conclusions, Dataset limitations and Recommendations for further analysis**

**Conclusions:**

* Outcomes by Category
  + 70% of crowdfunding campaigns were focused on art related projects, specifically in the theater, film & video and music categories. Of all projects, 34% of all campaigns fall in the sub-category “plays”.
  + 76% of all projects were located in the US. Out of a total of 565 successful projects, 436 of these were in the US. More than two thirds (304) of these successful US projects (and therefore more than half of all successful projects) fell under the same art related categories, theater, film & video and music.
* Outcomes by Date
  + If we compare the results when filtering by year, we can see that each individual year of data is so variable from month to month that the total count of project outcomes displayed by month for all the years is not believed to show anything of significant value. In other words, in the table created, there is not believed to be anything special about any specific month in terms of success or failure of campaigns. That said, if we partition by year instead, we can see a few very weak trends: From 2010 to 2019 the number of successful campaigns appears to be trending upwards, while the number of failed projects appears to be trending downwards, as shown in the figure below. However, without further analysis, additional context or a business objective it is not possible to confirm the validity of or infer any further significance about this finding.

A graph of a graph showing the number of the results

Description automatically generated with medium confidence

* Goal Analysis:
  + Of all the projects that set goals between 15000 and 35000, nearly 100% were successful. However, the total number of projects in this range make up less than 5% of the total projects. Looking instead at the total count of projects by goal range, we see that most projects set a goal between 1000 and 10,000, or greater than 50,000. Projects that set goals under 5,000 show the highest number of successful outcomes. Projects that set goals over 50,000 show the highest number of failed outcomes. This finding is close to what we might expect to see, i.e easier success for smaller goals, and more chance of failure for the largest goals. The only exception to this would be the large number of successful projects with goals over 50,000 (114 out of 1000). These are likely to be the truly groundbreaking projects.

**Dataset limitations:**

* Apart from being completely fictional data, there are several data limitations:
  + Small sample size.
  + Heavily skewed towards US crowdfunding campaigns.
  + Too few campaign records to make any clear conclusions about campaigns in other countries or to compare across countries.

**Recommendations for further analysis:**

* We could look at the Percent Funded field in more depth and analyze which categories and sub-categories were the highest or lowest funded. This might give us some insight into the types of successful projects that are the most popular.
* Another interesting analysis would be to determine which categories successful projects with goals over 50,000 fell into. This would give us insight into the most popular and potentially groundbreaking project types.
* We could also look at the difference between date created and date ended to see if the duration of campaigns had any effect on the success/failure or ability to reach a project goal.
* We could also change existing or create new pivot tables to show successful crowdfunding categories by percentage of successful projects. This would show the categories and sub-categories with the highest percentage of successful and unsuccessful projects. For example, we would know that the technology, photography and publishing categories have the highest percentage of successful campaigns.

## **Statistical Analysis**

**Mean or Median:**

* The data is quite skewed towards higher values and contains many outliers, therefore it would be better to use the median to summarize the data. Based on the median values it seems that successful campaigns have a median number of backers around 200, while failed campaigns have nearly half this number.

**Variability:**

* There is more variability with successful campaigns, as shown by the greater overall Range of data, a larger Inter Quartile Range, a greater Standard Deviation, and a larger Variance number. The variability of backers for failed projects is logically lower, because you might expect failed projects to have a lower range of backers compared to successful projects due to the lower number of backers. It could be hypothesized that in general the reasons for failure are more predictable than the reasons for success, meaning that conditions for success are likely more variable. This would of course need further investigation.