**Crowdfunding Analysis Report**

Libby Berter

Objective

Analyze data provided of 1,000 generated sample projects to determine what increases the likelihood of successfully funded project

Methods

* Conditionally format column for project outcome for successful, failed, cancelled, or live
* Create a % funded column and conditionally format based on %
* Create a column for the average donation per backer
* Split the Category & Subcategory column into 2 columns: Parent category & Sub-category
* Create a pivot table to determine the count of the outcomes of campaigns by Parent Category and use that data to create a stacked column pivot chart
* Create a pivot table to determine the count of the outcomes of campaigns by Sub-Category and use that data to create a stacked column pivot chart
* Convert the launched dates and deadline dates from Unix timestamps to a normal date. Use these dates to create a pivot table and pivot line chart
* Create a new table to calculate the numbers of successful, failed, & cancelled campaigns by goal range as well as calculating the % successful, %failed, & % cancelled campaigns
* Create a new table showing the mean, median, minimum, maximum variance, and standard deviation of the number of backers for both successful and unsuccessful campaigns

Observations from Charts

1. Given the provided data, what are three conclusions we can draw about crowdfunding campaigns?

* The category theater has the most successful campaigns & total number of campaigns. However, it has the 3rd lowest success rate (# of successes/# of submissions)
* Journalism campaigns have a 100% success rate
* For submissions in the Film & video category, TV has the highest success rate of all the subcategories
* June & July are the months with the highest success rate for campaigns while May & August have the highest failure rate for campaigns

2. What are some limitations of this dataset?

* Limited data on campaigns from countries other than the US (US accounted for 76% of the data)
* The blurb column isn’t very clear and doesn’t seem to match the categories

3. What are some other possible tables and/or graphs that we could create, and what additional value would they provide?

* The campaign goal by outcome – are less expensive jobs always more successful?
* The number of backers by Parent Category – are more people interested in supporting/backing a certain type of campaign
* Staff pick related to category and success rate – are campaigns that are staff picks more likely to succeed that campaigns that are not? Are staff picks more heavily favored by Parent Category?
* Spotlight related to category and success rate – are campaigns that are spotlight campaigns more likely to succeed that campaigns that are not? Are spotlight campaigns more heavily favored by Parent Category?

Observations from Statistical Data

1. Use your data to determine whether the mean or the median summarizes the data more meaningfully.

* The median data better represents to the data. Using a box & whiskers plot of the data shows a large number of outliers that skew the mean value

1. Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?

* There is more variability in the successful campaigns. This makes sense to me due to the fact that the unsuccessful projects had significantly less backers compared to the successful projects leaving less potential for variability. There were also roughly 55% more successful projects than failed projects giving more opportunity for variability.