**Written Report**

*Excel Homework: Charting Crowd Funding*

**Part 1: Three Conclusions About Crowdfunding Campaigns**

In general, most of the crowdfunding campaigns were successful. Out of 1000 campaigns launched, 565 (~57% of campaigns) were successful, 364 (~36% campaigns) failed, 57 (~6% campaigns) were canceled, and 14 (~1% campaigns) are live.

* 1. **Conclusion for Outcome Based on Category Pivot Table/Chart:**

When looking at the outcomes of crowdfunding campaigns by category, the majority of crowdfunding campaigns by category are:

* Theater (54% successful 38% failed, 344 campaigns),
* film & video (57% successful, 34% failed 178 campaigns)
* Music (57% successful, 38% failed, 175 campaigns).

The three campaign categories with the highest success rates:

* journalism (100% success, 4 campaigns)
* technology (67% success, 29% failed, 96 campaigns)
* photography (62% success, 26% failed, 42)

The three campaign categories with the highest failure rate (and highest success rate):

* games (44% success, 48% failure, 48 campaigns)
* food (48% success, 43% failure, 46 campaigns)
* theater (54% success, 38% failed, 344 campaigns)

Conclusion 1: Journalism has the highest rate of success (100%) out of any other campaign by category. However, there have been only 4 crowdfunding campaigns for journalism in our dataset. Due to the small sample size, we can draw limited for insight for future campaigns. Technology has a high success rate (67%) with 96 crowdfunding campaigns. This category may be worth investing in future campaigns due to its high success rate.

* 1. **Conclusion for Outcome Based on Subcategory Pivot Table/Chart:**

When looking at the outcomes of crowdfunding campaigns by subcategory, the highest number of crowdfunding campaigns by subcategory are in:

* Plays (54% successful, 38% failed, 344 campaigns) – subcategory of theater.
* Rock, music (57% successful, 35% failed, 85 campaigns) – subcategory of music.
* Documentary (57% successful, 35% failed, 60 campaigns) – subcategory of film.

The three campaign subcategories with the highest success rates:

* Audio (100%, 4 campaigns) – the only subcategory of journalism
* World music (100%, 3 campaigns) - subcategory of music.
* Web (71%, 24%, 51 campaigns) – subcategory of technology

The three campaign subcategories with the highest failure rate (and highest success rate):

* Mobile games (30%, 62%, 13 campaigns) - subcategory of games
* science fiction (64%, 31%, 14 campaigns) - subcategory of publishing
* radio and podcasts (50%, 50%, 8 campaigns) - subcategory of publishing

Conclusion 2: The only subcategory of journalism is audio which has the highest success rate (100%) along with world music of all the subcategories. Respectively, journalism has only 4 crowdfunding campaigns and world music only has 3 crowdfunding campaigns. Due to the small sample size, we can draw limited for insight on future campaigns. The Web, subcategory to technology has a high success rate (72%) with 51 crowdfunding campaigns. This category may be worth investing in future campaigns due to its high success rate.

* 1. **Conclusion for Outcome Based on Date Pivot Table/Chart:**

Conclusion 3: When looking at outcomes based on all years, July had the highest number of campaigns created (58) and had the most number successful campaigns out of all the months (93). However, June had highest rate of successful campaigns (64%) while 62% of campaigns in July were successful.

Conversely, based on all years documented in the dataset, August had the lowest number of successful campaigns (41), most failed number of campaigns (35), and the lowest rate of success (49%).

**Part 2: Limitations of this Dataset**

**There are a few limitations to the data set:**

* The duration of crowdfunding campaigns is not considered in our data when looking at outcomes. Successful crowdfunding campaigns may have had more time to reach their goal while failed crowdfunding campaigns may not have had as much time to reach their goal.
* Most crowdfunding campaigns were in the United States (763 out of 1000), therefore looking at location based on country may be too broad. Additionally breaking down location by state or region may provide better insight since the success and failure in campaigns may differ in different parts of the United States.
* Canceled and live campaigns may change the data if they were completed or to be completed, respectively. For example, were some campaigns canceled because they were expected to fail? If so, there may be a larger proportion of failed campaigns than is already documented.

**Part 3: Other Possible Tables and/or Graphs**

* Creating a pivot table/chart (line) of the outcome of a campaign based on the duration (in weeks) would provide valuable insight on if there is a difference in the duration of successful versus failed campaigns.
  + Duration of a campaign can be created by adding a new row to the Crowdfunding dataset and subtracting date\_created by date\_ended.
* Creating a pivot table/chart on outcome based on goal of a campaign would provide valuable insight on if the goal of a campaign factors into its success versus failure. Adding a category filter would provide additional insight on outcome based on goal per category. For example, perhaps audio campaigns are very successful because the goal amount of the campaign is low while theater campaigns have a high goal amount to reach. (Note, this pivot table/chart are created in the bonus section).
* Creating a pivot table/chart on outcome based on average\_donation would provide valuable insight on if there is a difference in the duration of successful versus failed campaigns. Adding a category filter would provide additional insight on outcome based on goal per category. For example, perhaps successful campaigns have large donation averages. Looking closer, perhaps technology campaigns have large donations and that is a factor for the high rate of success.

**Bonus: Summary Statistic Questions**

|  |
| --- |
| * Use your data to determine whether the mean or the median summarizes the data more meaningfully.   The median would best summarize the data for both successful and unsuccessful datasets. There are many outliers creating a right (positive) skew in both outcome datasets where the mean is significantly greater than the median. There is a high level of variability from the mean when looking at standard deviation (sd) and variance (var). Therefore, the mean may not be the most accurate measure of central tendency when compared to the median. |
| * Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?   Based on the standard deviation and variance there is more variability in number of backers in successful campaigns (var = 1606217, sd = 1267) compared with unsuccessful campaigns (var = 924113, sd = 961). This makes sense when looking at the boxplot and mean versus median of both outcome datasets. The boxplot of the successful backer\_counts appears to have a higher range and the amount of outliers appear to be further from the median when comparted to unsuccessful backer\_counts. In addition, the difference between the mean and median is greater to successful backer\_counts than to unsuccessful backer\_counts. |