**Homework – Excel**

Mika Nagamine

**Create a report in Microsoft Word and answer the following questions.**

**Given the provided data, what are three conclusions we can draw about Kickstarter campaigns?**

* Categories that a campaign belongs to seem to be one of the factors determining whether it can be successful: Kickstarter campaigns in some categories (such as theater, music, and film & video) tend to be successful more often than the campaigns in other categories such as food, games.
* Sub-categories that a campaign belongs to seem also impact whether it can be successful: Success for the campaign is not guaranteed even if they belong to a parent category that have many successful campaigns (such as music). We have to look at what sub-categories under the parent categories (such as music) that the campaign belongs to. For example, many campaigns about music (parent category) tend to be successful, but no campaigns about jazz and faith music (sub-categories under music) are successful.
* Timing that a campaign is launched also seems to impact whether it can be successful: The number of successful campaigns launched in the first half of the year seems to be larger than the numbers of successful campaigns launched in the latter half of the year.

**What are some limitations of this dataset?**

“Goals” and “pledged” are presented using various currencies, which could limit what analysis we can perform. The values of money are not the same across different currencies. Thus, unless we convert various currencies to the value of one specific currency such as USD, we cannot compare pledged amounts, calculate overall statistics such as the grand total of pledged money for all the campaigns, average amount pledged per campaign. With the same reason, the average donation in USD cannot be compared with the average donation based on the other currencies.

**What are some other possible tables and/or graphs that we could create?**

I would create a table that presents, by category and sub-category, average goals and the lengths that the campaigns need to be run. Please see pivot tables in “my\_table” tab.

One table (left one) shows the successful campaigns within a month. The length of campaign was calculated as (Date Ended Conversion - Date Created Conversion). Currency of choice is USD (but people can change it) to avoid the limitation of the dataset mentioned above (mixing the different currencies and their values). The table shows the sum of pledged amount and average pledged amount by categories/subcategories. For comparison, campaigns that failed within a month are presented in the table at the right side.

Suppose a person wants to launch a campaign. She has never used Kickstarter. She knows her campaign is about pop music and would like to run her campaign relatively short period of time. Beyond that, she does not have a clear idea about how long she needs to run her campaign and what is the reasonable goal she can expect from her relatively short campaign. The table might potentially help a person like her who is new with Kickstarter. It might help them estimate the length of campaign and approximate goal before actually launching their campaign.

**Bonus Statistical Analysis**

|  |  |  |
| --- | --- | --- |
|  | Successful | Unsuccessful |
| Mean | 194 | 18 |
| Median | 62 | 4 |
| Minimum | 1 | 0 |
| Maximum | 26,457 | 1,293 |
| Variance | 712,840.987 | 3,773.222 |
| Standard Deviation | 844.299 | 61.427 |

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

Successful Campaigns

I would use the median for summarizing the number of backers

The variance and standard deviation of backer\_count seem to be big with the successful campaigns, which indicates that the data for successful campaigns might include outliers. A large difference between the minimum and the maximum also indicates that the maximum can potentially be an outlier. The mean is sensitive to the outliers. If 26,457 is an outlier, the mean is overestimated. Unlike the mean, the median is less likely impacted by outliers. In other words, it is better to use the median to summarize overall characteristics of backer\_count data when the data seems to include outliers.

Unsuccessful Campaigns:

I would use the mean for summarizing the number of backers.

The variance and the standard deviation are not as large as those with the successful campaigns. In addition, difference between the minimum and the maximum is smaller than that of successful campaigns. These indicate that backer\_count for unsuccessful campaigns may not have outliers, and thus, it is probably safe to use the mean to summarize the data.

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

As mentioned above, the variance and standard deviation for the successful campaigns seem to be large, which indicates that there is more variability in the successful campaign data. This makes sense to me because the successful campaigns have a larger number of backers (median=62). Having a large number of backers can increase a chance that this group gets some backers that can donate larger amounts than other backers.