Assignment 1 EXCEL **October 27th, 2022**
Natasha R. Girdharry

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

Three conclusions that can be made from this data include the following:

I. Theatre events were the largest successful crowdfunding initiatives with 321 plays held and 23 cancelled. Of the plays that occurred, 58% (187 plays) were successful in generating the amount of required donations, and 41% did not raise the donation goal for that crowdfunding event.

II. Film and video events were the second largest successful crowdfunding initiatives with 167 held, and 11 cancelled. Of the events that occurred, 61% were successful and 36% failed in raising the donation goal. Data from the sub-category analysis indicates that documentaries made up 33% of the successful film and video crowdfunding campaigns, with science fiction making up only 5% of the successful campaigns, and 15% of the failed ones.

III. The crowdfunding campaign data extended across 10 years (2010-2020). Based on the data from the campaign outcomes by month it appears that the best time to launch a crowdfunding campaign is in July as it resulted in the greatest number of successful outcomes (58) on average. Campaigns launched in August had the most events cancelled (8) on average, the second greatest number of failed events (35) and the lowest number of successful campaigns (41) on average. Therefore, campaigns launched in August are least favourable to achieve crowdfunding goals.

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

I. One limitation of this dataset is that 76.3% of this data is coming from campaigns launched in the United States. The remaining countries represented including Australia (43), Canada (44), Switzerland (23), Denmark (31), Great Britain (48), and Italy (48) did not have 50 crowdfunding campaigns. Therefore, the trends observed in this dataset may not be representative of the effectiveness of crowdfunding campaigns, and the campaign preferences in these countries. As well, due to the limited data in these countries, the successful of campaigns launched based on month or date may also not be applicable in these countries.

II. Another limitation of this dataset is the broad distribution of the sub-categorical data among the parent categories. There is more data available for film and video (ex. Science fiction, documentaries etc.) however, there is fewer for journalism or theatre plays. There isn't enough data within these subcategories to make meaningful conclusions to ascertain why one category might be more popular than another. For example, which plays are most popular and which medium is preferred? This

information can become important in this analysis of understanding crowdfunding trends among donators.

III. An additional limitation in this analysis is that we don't have demographic information of the individuals donating including, gender, age, income levels, and educational backgrounds. This could help explain some trends observed based on donators preferential activities.

IV. An additional limitation is the lack of uniformity in the dollar amounts provided as the goals and pledged amounts currency varies. To make the data more comparable one should use the exchange rate to convert the non-USD campaigns into USD currency.

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

I. To create a better visual representation of the successful and failed campaigns one could use a pie chart to represent these by parent category.

II. One could calculate the summary statistics to get the mean, and median length of campaigns by parent category, and by sub-category. This could provide helpful information in the future or to perspective clients to understand how long the campaign may take dependent on the type of crowdfunding campaign they decide to do.

III. To create a visual representation of the campaigns that generate the most money one could first ensure that the dollar amounts are all in one currency (USD preferably given the number of campaigns in the US). Then they could create a line graph, with the parent categories, and eventually sub-categories on the X axis, and average amount of donations on the Y axis.

Bonus Statistical Analysis Questions

Summary Statistics of Successful Campaigns	
Mean	**851.1**
Standard Error	53.3
Median	**201.0**
Mode	85.0
Standard Deviation	**1267.4**
Sample Variance	**1606216.6**
Kurtosis	5.0
Skewness	2.2
Range	7279.0
Minimum	**16.0**
Maximum	**7295.0**
Sum	480898.0
Count	565.0
Largest(1)	7295.0
Smallest(1)	16.0

Summary Statistics of Failed Campaigns	
Mean	**585.6**
Standard Error	50.4
Median	**114.5**
Mode	1.0
Standard Deviation	**961.3**
Sample Variance	**924113.5**
Kurtosis	8.8
Skewness	2.7
Range	6080.0
Minimum	**0.0**
Maximum	**6080.0**
Sum	213164.0
Count	364.0
Largest(1)	6080.0
Smallest(1)	0.0

1. **Use your data to determine whether the mean or the median better summarizes the data.**

The measures of central tendencies that best represent this data of successful and failed campaigns is the median (IQR). As demonstrated in the summary statistics above, the mean for successful campaigns is 851 ± 1267.4. The standard deviation is larger than the mean indicating that the data is more spread out which can be observed using the range of 7,295. Similarly, the statistics from the failed campaigns is 586 ± 961 with a range of 6,080. The median for each campaign is much lower than the mean; successful campaigns – 201 (1152), failed campaigns – 115 (747) and is a more robust measure of central tendencies since the mean is skewed easily by outliers.

2. **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 with successful campaigns in this crowdfunding data than in unsuccessful campaigns. The data from successful campaigns indicates that they have a larger range (7,279), standard deviation (1267), sample variance (1,606,217), and interquartile range (1152) compared to the failed campaigns as displayed above in the summary statistics. This makes sense because there are likely fewer backers for the failed campaigns, hence why these campaigns did not reach their fundraising goal, and don't have as much variability as the successful campaigns. Some successful campaigns included as much as 7000 backers whereas the failed campaigns largest number of backers reaches a little over 5000. Therefore, because the failed campaigns were likely to have fewer backers it makes sense that there is less variability for these campaigns.