**Excel-Challenge**

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

With the provided data, we can conclude that:

1). The large number of backers does not guarantee that the campaign will be successful in raising money. Refer to Statistical Analysis tab in the excel workbook, even the campaign had 6080 backers, it still failed.

2). From the graph on Outcome by Month tab of the workbook, the campaigns that were launched in June and July seems to have a higher rate of successful than the campaigns that were launched in other months. Also, August seems to be the months that have less rate of successful and higher rate of failure. However, we do not have enough information to tell what the reasons are.

3). Projects type that are 100% successful are small, niche projects that seems unpopular. Refer to Outcome by Category and Outcome by Sub-Category tabs on the workbook, Journalist and World Music projects are those that have 100% successful rate. There are 4 Journalist projects and there are 3 World Music projects, which is small in number compare to the other type of projects in this sample data. The goal range for the Journalist is between $2,400 - $9,700, and the goal range for the World Music is between $4,500 – 88,400.

The most popularity type of the project is the theater follow by film & video, and music. More than 50% of these 3 types of projects are successful in crowdfunding.

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

From looking at this dataset, there are 3 limitations related to this dataset.

1). The sample data was collected in 10 years range, from 2010 – 2020. The old data that are dated back 5 to 10 years ago may not represent the current situation where the public nowadays is more aware of this kind of internet fund raising for startup businesses.

2). There are 763 campaigns from the US and 237 campaigns from other countries. This may not give a true finding if this dataset is intended to find the trend in the US. The dataset would yield more accurate result if it only focuses on the sample from US only. On the other hands, if this dataset is intended to find the trend in global internet funding projects, it should equally weight the sampled from each country and not just use a majority of sample from the US. In this case, the data is bias to study the global trend since it weighs more than 70% on US samples.

3). From the Statistical Analysis tab on the workbook, the huge variance on both successful and failed projects shows that the dataset contains samples that are widely spread out. The study may not yield an accurate result when there is a huge difference in the size of the data.

4). There is no information regarding the specific internet platform for these campaigns from the sample are from. The popularity and the credibility of the crowdfunding platforms may have a significant relationship to the successful and failure in fund raising of these campaigns.

5). We don’t have the information regarding the credibility of the campaign owners. It is possible that some of the successful campaigns were launched by a well-known person or company.

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

The other possible tables and or graphs that we could create for this analysis is a table that shows the duration of time launch and the number of successful and failure. This is to find out whether or not the duration of campaign launch has any significant relationship with the rate of successful or fail. It is possible that the longer time available to raise money may lead to more successful outcome; and the less time have to raise money, the less successful outcome. Given more time to raise money, it may increase more awareness toward the projects. We can prepare a new table showing the range of campaign launch duration, like 0 – 10 days; 11-20 days; 21-30 days in the first column and have other columns show the number of success/fail projects, and percentage of success/fail projects. Then plot the graph to show the result.

(Continue on next page)

**Statistical Analysis**

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

From the sample of 1000 campaigns, there are 565 successful campaign and 364 failed campaigns. The average number of backers of the successful campaigns is 851 and the average number of backers of the failed campaigns is 586. The median number of backers for the successful campaign is 201 and the median number of the backers of the failed campaign is 114.5.

For this sample set of data, using mean gives a better summarization of the data since the mean shows the average value of the sample. The median only shows the observation of what the middle value of the sample is, by listing the data from the lowest to the highest. It is just an eye gazing on the data and the midpoint without taking any weight into calculation. Thus, it would yield the less accurate result.

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 with the successful campaigns than the unsuccessful campaigns. However, the variable for both successful and unsuccessful campaigns is very large. The large variance means the data are very spread out and less close to the mean. With this high variability, it means that the data in this dataset is less consistent; thus, make it more difficult to make prediction about the population.

From this dataset, I think it makes sense that both successful and unsuccessful campaigns have a huge variance. This is because the dollar amount of the smallest goal and the dollar amount of the largest goal is very far apart.

The dollar amount of the lowest goal is $100 while the dollar amount of the highest goal is $199,200. The smallest goal of $100 have zero or 1 backer while the largest goal of $190K + have about 2 to 4 thousand backers. There is a huge gap from zero-dollar goal to almost 200,000 dollars goal. And because of this, it causes a huge gap in the number of backers on the lowest dollar campaigns and the largest dollar campaigns.

The mean number of backers for the successful campaign is 851, and the mean number of backers for the unsuccessful campaigns is 586. This may not represent the true population due the extreme difference of the lowest and the highest number of backers from the dataset. If the campaigns goals were from those with closer goal range, the variance and standard deviation will be smaller making them more accurate for the research.