Module 1- Excel challenge

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

* One conclusion that we can draw is the total number of successful crowdfunding campaigns exceeds the total number of failed crowdfunding campaigns and the total number of cancelled campaigns at each month. Another conclusion that we can draw is the total number of successful crowdfunding campaigns is at its highest during the summer (June and July). A third conclusion we can draw is the total number of successful, failed, and cancelled crowdfunding campaigns are inconsistent from month to the month, as the trend for each outcome rotates between increasing and decreasing about seven times.

What are some limitations of this dataset?

* One limitation of this dataset is we can’t really generalize the fact that there are more successful/failed/cancelled crowdfunding campaigns in one month compared to another month. This is because our data set features the total number of crowdfunding campaigns from 2010 – 2020 in a specific month, instead of the total number of crowdfunding campaigns per month of a specific year. For example, there could be just slightly more successful crowdfunding campaigns in February than in January for ten out of the eleven years, but for the remaining one year, there could be significantly more successful crowdfunding campaigns in January than in February, causing the total from 2010-2020 being higher in January than February. This limitation thus leads to a misrepresentation of data. Another limitation of this dataset is the total number of crowdfunding campaigns for a specific outcome can be deceiving when comparing it between different months. For example, looking at the total number of successful crowdfunding campaigns for January and March, you would think the success is the same because of the number (49 for each), but it’s not because the total number of crowdfunding campaigns for January exceeds the amount for March. And so, the success rate in January is lower than in March. This limitation thus also leads to a misrepresentation of data.

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

* To avoid the two limitations described above, we could create a segmented bar graph (11 years with each year consisting of three bars and with each bar of a specific year representing a different outcome – successful, failed, and cancelled). The segments of each bar would represent the rate for every month. For example, in 2010, the successful bar would consist of the rate of success for each month in 2010. And the failed bar would consist of the rate of failed crowdfunding campaigns for each month in 2010. The additional values we would need are 1) the total number of crowdfunding campaigns at each outcome, at every month, at every year and 2) the grand total number of crowdfunding campaigns at every month, at every year.

Statistical Analysis

Statistical Analysis -1

* According to the box and whisker plot, the distribution of the number of successful campaign backers is skewed to the right. Because the distribution is skewed right, it’s better to use the median than the mean to represent the data. This is because the median is not affected by outliers while the mean is affected by outliers. We don’t want to use the mean because since the distribution is skewed right, the mean would overestimate the data (as it is affected by outliers). Because the distribution of the number of failed campaign backers is also skewed to the right, it’s better to use the median than the mean to represent the data for the same reasons.

Statistical Analysis -2

* According to the box and whisker plot, there is more variability with successful campaigns. We can tell because the data is more spread out in the distribution of successful campaigns (i.e. the distribution of successful campaigns features a bigger range than the distribution of unsuccessful campaigns). According to the summary statistics, this does make sense because the standard deviation, which is a measure of variability, of successful campaigns is bigger than the standard deviation of unsuccessful campaigns.