**Report for MS Excel Challenge Sajid Anjum**

1. Given the provided data, what are three conclusions we can draw about crowdfunding campaigns?
   1. The majority of kickstarter campaigns are in the general categories of fine arts – music, theater, and film/video. Out of the total sample of 1000 kickstarter campaigns, 697 are from these three categories, of which 388 were successful.
   2. The probability of success did show some variability with the size of the fundraising goal. Goals less than $15,000 had between a 40 and 60 percent chance of success, while goals set between $15,000 and $50,000 had between a 60 and a 100 percent chance of success. Goals set higher than this amount had a lower chance of success.
   3. The category with the least chance of success was **games** with 44%, while the category with the highest chance of success was **journalism** with 100%. However, since only four kickstarter campaigns were from the **journalism** category, this data must not be taken very seriously. Discarding **journalism**, the category with the highest success was **photography**, with a rate of 62%
2. What are some limitations of this dataset?
3. No information was given as to how the data in this dataset was selected. Was the data taken randomly, or were the 1000 students all graduates from a fine arts institution? This would explain why so many campaigns were in the categories of music, theater, and film/video
4. The time of the data collected ranges over an entire decade from 2010 to 2020 even though most campaigns only lasted a few months. The sample size of 1000 may be significant, but it is not large enough to account for how preferences may have evolved over time. It maybe that publishing was a relatively more popular campaign in 2010 compared to technology than in 2020, but dividing the samples of both these categories by a factor of 10 would make the sample sizes too small for them to be meaningful.
5. The overwhelming majority of the campaigns (more than 75%) were taken from the US.
6. What are some other possible tables and/or graphs that we could create, and what additional value would they provide?
7. A graph that divides the campaigns by length of time the campaign lasted would be quite useful. Some campaigns were only a couple of days long while many lasted for weeks. Did this have any effect on success?
8. Dividing the campaigns by year would also be useful. Was 2010 a better environment to start a kickstarter campaign than 2020? This information could be quite useful for people who want to determine whether kickstarters are a fad that has already peaked or whether the phenomenon will sustain itself.

Bonus

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

The **mean** is far larger than the **median**, which means that the data is heavily skewed to the right. A few campaigns with a large number of backers are affecting the statistics. If a person selected randomly was going to start a campaign, the **median** would be a much better representative of that campaign would look like. However, the **mean** is a better representative of the total number of backers engaged in kickstarter campaigns.

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

The standard deviation of the number of backers of **successful** campaigns is **1266**, while the standard deviation of the number of backers of **unsuccessful** campaigns is **960**. One reason for this is that both the mean and the median number of backers is higher for successful than for unsuccessful campaigns, and variation increases proportionally with the size of the random variable. In fact, adjusted for size, the variation of the backers of unsuccessful campaigns is actually larger.