Homework #1 – Excel-Challenge

1/ Given the provided data,

1. Given the data provided, campaigns are more likely to be successful (56.5%) than failed (36.4%) or cancelled (5.7%). Majority of the campaigns are Theater/Theater (344/1000). Following are “Film & Video” (178 campaigns) and “Music” (175 campaigns).
2. Campaigns that were started in June or July have a higher chance of success. Except for the campaigns that are still live, there are 55 and 58 successful campaigns in June and July, respectively, while there are less than 50 successful campaigns in any other months. The chart helps us to prove the point since the gap between successful and failed campaigns starts to widen in June and July but immediately decreases In August.
3. When we looked at the 24 sub-categories, “plays” are very popular, however it doesn’t mean that “plays” are guaranteed to be successful (38.4% failure). Even though some sub-categories (audio and world music) have a 100% success rate, the sample is too small to reach a solid conclusion.

2/ Some Limitations of this dataset:

1. The samples are not large enough for all categories and sub-categories. One variable (theater/play) accounts for 34.4% of our samples.
2. We need more information about the creator. If the creator is a celebrity or well-know, or have run a couple campaigns, it could affect the result of the campaign.
3. 237 out of 1,000 campaigns are not from the U.S. The differences in demographics would not help with our analysis.
4. Also, the dataset does not mention if the backer receives anything in exchange for their donations (products, ticket to plays, etc…) since it could impact one’s decision if the backer could receive a benefit.

3/ Some possible tables/graphs that we could create:

1. Graphs and tables that remove outliers. For example: theater/plays, campaigns that have large funding goals.
2. I would be helpful if we analyze the projects by duration conversion using data created conversion and date ended conversion and the success rate. Maybe if a creator has more time to achieve the funding goals, the higher the chance that the project will be successful.
3. Tables and graphs that maybe show the relationship between the success rate and the funding goals. For an example, a lower funding goal might have a higher chance of success.

**BONUS:**

* Use your data to determine whether the mean or the median better summarizes the data. The median better summarizes the data because the distribution is affected by the outliers
* Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not? Based on the standard deviation (1,267.37 vs 961.31), there is more variability with successful campaigns. Most failed campaigns have very low backers or no backers at all (min=0). In other to be a successful campaign, a funding goal needs to be met and there will be high enough number of backers to donate. In some cases, a campaign can have high numbers of backers (in our case max = 7295) that donate over the funding goals. In conclusion, failed campaigns have low number of backers with a small variation and successful campaigns have a lot of backers and the number of backers vary from 1 to 7,295