01-Excel-Homework

1. Given the provided data, three conclusions we can draw from the Kickstarter campaigns are:
   1. The category with the most Kickstarters as well as the most successful campaigns is within the realm of theater. Coming up second with the most Kickstarters and the most successful campaigns would be music. If I were to recommend creating a Kickstarter with the highest chance of success, based on the data I would suggest creating a campaign in the category of theater.
   2. Given that the highest category of Kickstarters is theater, it comes to no surprise that the sub-category with the highest count of success and overall total campaigns is plays. 77% of the theater Kickstarter campaigns are from plays and then musicals and spaces at 10% and 13% respectively. While the data is skewed towards the category of theater and the sub-category of plays, many of these campaigns have failed. Within theater, the sub-category of plays has a 50% success rate taking into account the current live campaigns that have no definitive outcome just yet. However, looking at the other sub-categories there are plenty who have 100% rate of success like, classical music, documentary, electronic music, hardware, metal, nonfiction, pop, radio & podcasts, shorts, tabletop games, and television.
   3. Looking at the month pivot table and line chart, the month with the most successes are in May and the month with the most failures is in July. Based off of this information and the previous data, I would draw a conclusion that the best time to start a Kickstarter is in May and in the category of theater with the sub-category of plays.
2. A limitation of the dataset is that it doesn’t show the results of the campaign after the Kickstarter is successful. Sometimes, even though a campaign is successful and has reached the goal amount for funding, the company does not follow through with the product or whatever the campaign was funded for. A good example would be the Kickstarter for The Coolest Cooler. Another limitation is that the data doesn’t have the same currency across all campaigns. Of course I can clean it up and convert the data to have the same currency however, it would be a better overview if it was already converted so that we would be able to have a more holistic view on how much each Kickstarter was asking for.
3. A suggestion to improve the data to tell a better story is to create another pivot table and chart to view the correlation between how much a Kickstarter is asking for and how that corresponds to its success and failure. Similarly, this can also be broken out by category, sub-category, and a time period analysis in the form of a bar chart. I would also suggest building a separate table set to see how many campaigns exceed their goal and see what category and sub-category those Kickstarters are in. That way, we can draw conclusions on which, campaigns have the most enthusiasm and are regarded highly by backers.

01-Excel-Bonus

1. In this case, I believe that the median is more accurate or summarizes the data more meaningfully in terms of central tendency. The median takes into account the zero values in the data set whereas when using mean, the outcome will ignore the zero values which will make it less accurate and skewed. While there is an incremental linear progression within the data, there are a few outliers in the dataset - a sudden jump for example, from 8359 to 20242, making the mean less accurate than the median.
2. Based on the data, I believe the successful campaigns have more variability since there are 2 extreme values versus the failed campaigns which have more of a linear/gradual progression with consistent values. In addition, the successful campaigns have a higher sample size, higher variance, Interquartile range, and standard deviation which supports that if has more variability than the failed campaigns. I believe the data speaks for itself and it make sense since the failed campaigns are more consistent in value and have a smaller sample size where as the successful campaigns are much larger in terms of sample size adding to an increase in variables.