Excel Challenge Written Report:

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

*Conclusion #1:* Categories with the greatest number of campaigns experienced more successful campaigns. For instance, the category “theater” consists of roughly a quarter or 25% of the total campaigns hosted on Kickstarter, which lead the category to experience the most successful campaign. On another note, all campaigns in the “journalism” campaigns were cancelled, which leaves us with insufficient information. The lack of popularity amongst “journalism” may have contributed to failed and cancelled campaigns by the hosts.

*Conclusion #2:* There were only a small number of campaigns hosted outside of the United States, which may have contributed to most campaigns to fail or be canceled outside of the United States. Kickstarter is vastly popular within the United States, as apposed to other countires, which explains why more campaigns for successful within the United States, but unsuccessful elsewhere. To elaborate, more of the general population in the United States knew about Kickstarter, which lead them to donate, but with only 26% of the remainder of the countries hosting campaigns on Kickstarter.

*Conclusion #3:* The launch date outcomes or the date of creation of the campaign contributed the least in determining whether a campaign will be successful or fail. Overall, the sub-category with highest number of campaigns, will have an increased change of successful campaigns. Furthermore, with an increase in successful campaigns per sub-category it is likely to have an increase in failed campaigns. The increase in successful and failed campaigns within a sub-category is likely due to the increase in competition and demand for funding for each sub-category.

1. What are some limitations of this dataset?

Although this dataset provides an adequate amount of quantitative data about the campaigns hosted on Kickstarter there are some limitations. One limitation is the lack of qualitative data that would describe how those who hosted campaigns went about to draw attention to their campaign to gain funding to meet their goal. The host of the successful campaigns likely made additional efforts into surpassing their target goal within the month their campaigns were active.

1. What are some other possible tables and/or graphs that we could create?

Some other possible pivot table and pivot graph we could create include comparing the ‘state’ of the campaign with ‘average donation’ and ‘percentage funded’. With this comparison we can determine if there is any relationship between the donation amounts and the likelihood of the campaign succeeding.

In addition, we could calculate the outliers and create a box and whiskers plot of the ‘average donation’ to check for variabilities. Through this diagram we can group the data points and determine if there are any unusually high or unusually low value points within the average donations made to these Kickstarter campaigns.

Bonus Statistical Analysis:

|  |  |  |  |
| --- | --- | --- | --- |
| **Successful** |  | **Unsuccessful** |  |
| Mean | 194.43 | Mean | 17.71 |
| Median | 62.00 | Median | 4.00 |
| Min | 1.00 | Min | 0.00 |
| Max | 26,457.00 | Max | 1,293.00 |
| Variance | 712,840.99 | Variance | 3,773.22 |
| Standard Deviation | 844.30 | Standard Deviation | 61.43 |

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

The distribution of data for successful and unsuccessful campaigns are positively skewed because the mean is greater than the median in both scenarios. This means that the count of backers of each campaign varies greater, pushing the distribution to the right. In this case, the median summarizes the data more meaningfully because it more accurately describes the general count of backers within this dataset. The distribution is skewed because of potential outliers or inequalities with the data making the mean is disproportionately higher then the median.

* 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 successful campaigns because the standard deviation of this distribution is much more widely spread out compared to the standard deviation of the unsuccessful campaigns. The higher variability makes sense because the backer count of successful campaigns fluctuate vastly with a minimum of 1.00 person donating to the campaign to a maximum of over 26,000 people donating to a single campaign.