Verónica Lizbeth Castillo Morales

Data Bootcamp 2021

Lesson 1: Excel

**Excel Challenge: Kickstarter Activity**

**Report**

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

*One of the conclusions that can be drawn from the first chart is that, in the global scale, while there are more theater campaigns than any other, they don’t have the highest success rate. It can be concluded that music campaigns account for the highest success rate among any category, while food campaigns account for the lowest success rate. By looking at the second chart, some other conclusions can be drawn. For instance, some sub-categories like documentaries, television and shorts (film & video), and classical, pop, electronic and rock (music) have a 100% success rate. From the same categories, animation, drama and faith music have a 0% success rate, so there is still a lot of variability in terms of how successful a campaign is inside the same category. Finally, from the third chart it can be concluded that the second trimester of the year had the highest success rate, while a campaign was more likely to fail in the last months of the year.*

1. What are some limitations of this dataset?

*A limitation is that, like any sample, it is limited. If we had a number of total kickstarter campaigns and the average distributions of each category, funding, etc, we could determine if the sample is representative of the population. Other limitation is that the table has different currencies, and if we were to analyze funding, we would need to convert all of the data to one currency. Additionally, the column “blurb” that contains descriptions of each campaign is unnecessary in my opinion, as it takes up space and if we were to have a larger sample, it would take both space and memory.*

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

*Some tables and graphs that could be good to analyze further are:*

* *Goal vs. state. It would be interesting to find if there is a correlation between how high the goal of the campaign is and how likely it is to fail. For instance, if expensive campaigns usually fail more that cheaper ones.*

*Note: this was achieved in the first bonus of the homework.*

* *Average donation vs. state. This would be useful to get insight in regard to the prices of the packages in the campaign. For instance, if having a lot of small donations yields a higher success probability that a few large donations, or if it is irrelevant.*
* *Average donation vs. category. I think this would be an interesting insight, as this could show if the category of the campaign influences how much are people likely to donate. For instance, if social projects have higher average donations than food campaigns, etc.*

**Bonus**

Use your data to determine whether the mean or the median summarizes the data more meaningfully and determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?

*By graphing the results, it can be concluded that in both cases, the median represents the data better, as both sets of data are skewed to one side, and relatively high values move the mean from a value that can be deemed as representative. At the same time, there is more variability in successful campaigns, and in my opinion, it makes sense because campaigns usually fail because of an inability to reach their goal, meaning there are less donations and thus, less bankers, which means the values would tend to be small. In the successful case, the goals for each project go from less than 1000 to more than 50000, and having successful projects in every case with different average donations, a wide range of number of bankers also makes sense.*