Jennie Cinelli

University of Minnesota

Data Analytics and Visualization Bootcamp

Homework 1 – Excel

1. **Given the provided data, what are three conclusions we can draw about Kickstarter campaigns?**
   1. Given the categories of campaigns in comparison to the state of the campaign, out of 9 parent categories, 53% were successful. The most successful categories are music, theater, and film & video. The category music was the most successful category at 77% including live campaigns. The second most successful category is theater at 60%. Film & video comes into third place at a 58% success rate. Overall, the category that had the poorest performance was the category Journalism, which had 0% success rate with 100% of their campaigns cancelled.

Each category can be analyzed further by country. For example, when looking at the overall most successful category (music) by individual countries, you can see that the percentage of success, failure, and cancellation rate varies. When filtering on the country of the United States, the category music has a 77% success rate, in Australia it has a 100% fail rate, and in China the category has a 100% cancel rate.

When filtering, you can get an idea that there may be some limitations to the data. For example, the lowest preforming category (journalism) is skewed as a result of only having campaigns in the United States and none in the other 20 countries. We will discuss limitations of the data further on in question 2.

* 1. The sub-categories of campaigns in comparison to the state of campaigns shows that out of the 41 sub-categories, 53% of campaigns were successful. 12 of the sub-categories were 100% successful with a 0% fail or cancel rate. Five of these sub-categories are from the successful parent category of music which includes classical music, electronic music, metal, pop, and rock. Three of the sub-categories are from the parent category of film & video, which include documentary, shorts and television. Two are from the parent category publishing: non-fiction and radio & podcasts. And the last two hardware (technology) and tabletop games (games).

The sub-categories: art, audio, science fiction, and world music all had a 0% success rate with all of their campaigns being 100% canceled.

These categories can be further analyzed by parent category and country. For example, if you filter the data to show only the United States and the subcategories of the parent category music, you can see that rock dominates the music campaigns with 240 of the campaigns being 100% successful. It also shows that faith, jazz and world music all have 100% cancelation and/or failure in the United States.

* 1. An overall analysis of launch date of campaigns in comparison to state of campaigns for all parent categories during the time period from 2009 to 2017 show that while 53% of all campaigns were successful, the campaigns were most successful when they were launched in the month of May at a rate of 61%. However, it is observed that those campaigns launched in May also has a growth of failed campaigns with the chances of failure at 33%. This may indicate that the best month of the year to launch a campaign would be February when the success rate is increasing, and the failure rate is declining. This data also shows that the worst time of the month to launch a campaign would be the month of December where the rate of success (44%) is lower than the rate of failure (47%).

1. **What are some limitations of this dataset?**

There was one immediate limitation I encountered early on in the process of analyzing the data; some category data was only available from certain countries and not others. For example, in the first pivot table that was created to compare categories of campaigns to the state of campaigns, the category journalism (which was the lowest scoring category) was only as a result of campaigns ran in the United States. This calls into question of the validity of the data. A future study that would make the data more reliable should only analyze data from countries that contain data from all categories being analyzed and exclude countries/categories where data is not available for comparison.

Another limitation would be that the data is a few years dated. A future analysis that would make the data more reliable would be to include data from the most recent years of 2018 to 2020.

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

A table and graph that would be useful to create would be a table that compares the length of the campaign in comparison to the state/outcome of the campaign. Another table that would be beneficial to analyze would be the number of campaigns by country. The last table that could be created would be the relationship between the goal of a campaign and the amount pledged.

**Bonus Statistical Analysis**

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

In both cases, the will median will summarize the data more meaningfully as the data has a large variance of numbers, meaning that the median will be a more representative number for the sample.

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

According to the standard deviation of the successful and failed campaign data, there is more variability with successful campaigns. Most of campaigns are successful because of the large number of backers who pledge enough money for the campaigns to succeed. In many cases, the successful campaigns overachieve their goals due to large number of backers, but this creates a significant variation at an individual level. Campaigns fail due to the low count of backers or no backers at all. For failed campaigns, nearly all of them have a low backer count with a small variation.