**Crowdfunding Summary**

Name: Yuntian Xue

* Given the provided data, what are three conclusions we can draw about crowdfunding campaigns?
  1. For the total 1000 crowdfunding, theater is in the majority. The total number of theater-related crowdfunding is 344. While the number of successful, failed and canceled theater crowdfunding campaigns is also the largest compare to other categories. The journalism has the least number with only 4 campaigns, but all of them were successful.
  2. For the sub-categories, theater has only one sub-category – plays, which has the total number of crowdfunding far more than other categories. For the rest categories, rock music and documentary have the second and the third most number of crowdfunding, respectively. Most categories have more successful campaign than failed ones except mobile games (8 failed and 4 successful).
  3. Through all years, when categorize depending on the date that the campaigns were created, each month the number of canceled, failed and successful crowdfunding campaigns stayed within certain ranges: canceled – 0~19; failed (20~40); successful (40~60). For each month there were more successful campaigns than failed and canceled ones.
  4. Goals will affect the percentage of successful campaigns, for the ones had the goal between 15k to 25k and 30k to 35k, the percentage successful is 100%. Campaigns with goal greater than 50k had more percentage failed and canceled (53% and 9%). Most campaigns had the goals between 1k to 5k and 5k to 10k (231 and 315 total projects), while more campaigns with the goal of 1k to 5k succeeded (83% vs. 52%).
* What are some limitations of this dataset?
  1. The currency was not unified, which made the goal and pledged sections not accurate.
  2. There can be more sub-categories for theater so that the comparison among sub-categories will be more balanced. The current table has some parent categories divided to very specific sub-categories, such as music and technology, while the theater only contains one sub-category – plays. The plays can be further categorized based on their themes.
  3. The years should be an important filter for analyze of trend. Since the dataset is across 10 years and some categories might have gained less popularity and lose their attractions for crowdfunding.
  4. The analysis only considered category and years, it didn’t do anything with the percent funded, backers count, average donation and country.
* What are some other possible tables and/or graphs that we could create, and what additional value would they provide?
  1. Unify the currency for the goal, pledged and average donation columns.
  2. Calculate, compare and summarize the percentage of successful, failed and canceled crowdfunding campaigns among different categories. This will provide a better overview of which categories have the most/least interests.
  3. Calculate the duration of each crowdfunding campaign and see if there is some relation between campaign duration and percentage of successful campaigns.
  4. Filter the data by years and see if the trend has changed over years.
  5. Look at how other variables such as percent funded, average donation, backers count can affect the outcomes.
* Use your data to determine whether the mean or the median summarizes the data more meaningfully.

The median summarizes the data more meaningfully. Since the range of number of backers is very large, from 16 to 7295, the mean is just a numeric value and will likely skew to larger values, and not reflecting the truth. Instead, median showed the number in the middle, which reflects the backer situation for most campaigns.

* 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 data calculated, there is more variability in backer number with successful campaigns. This makes sense. The successful campaigns are dependent on a lot of other factors, even if there is lacking backers, the advantages in other aspects may still make things work. While for failed campaigns, lack of backers is one of the very important reasons for failure and thus the variance would be smaller.