**Conclusions**

* 1. Based on the data provided, we can conclude that there are more successful crowdfunding campaigns than failed ones. Moreover, there is only a low number of canceled campaigns. If we filter out campaigns that have been canceled or are still live, we can see that about 61% of the 929 completed campaigns were successful. The lowest success rates are Games (48%) and Food (52%). The highest success rates were Journalism (100%), Photography (70%) and Technology (70%). It is important to note that there were only 4 total campaigns for Journalism. It is important to note that Food and Photography had about a 10% cancellation rate. Investors/backers might be influenced by the 10% cancellation rate and 52% success rate in the Food category. Another conclusion that we can draw is that ‘Games’ is the only category that has more failed campaigns than successful. This category includes video games and mobile games.
  2. Looking closer at specific sub-categories, the largest number of campaigns clearly involved ‘Plays’. ‘Plays’ had a total of 344 campaigns whereas the next highest category, ‘Rock’, only had 85. Plays had more successful campaigns and failed campaigns than any other category had campaigns in total. Other than the campaigns involving Plays, the statistics need to be taken with a grain of salt because of the small sample size. When only looking at completed campaigns, it is interesting to note that Audio and World Music have a 100% success rate. However, since their sample size is so small, it is hard to draw a conclusion based on this. Photography Books, Television, and Web, all had over 70% success rates, while Mobile Games, Science Fiction, Food Trucks, Radio & Podcasts, and Video Games, all had under a 55% success rate.
  3. In June and July, there is a larger gap between the number of successful and failed campaigns. We can make some inferences based on this but would need more information as to why campaigns are more successful in those months. My inference is that since about 75% of the campaigns are in the US, it has something to do with these months being summer months (no school, less work, etc.). There is no month where the number of failed campaigns is higher than successful campaigns, however, the closest gap are in August and December.

**Limitations of the dataset**

* Sample size. Some categories and sub-categories have very little data. This tells us that certain aspects of the data is statistically insignificant.
* This data is not up to date. Kickstarter was founded in 2009 but this data only goes up to 2019. Given the exponential popularity and growth of smart phones, social media, and other aspects of technology, there is a good chance that the data from the last four years would be very valuable.
* There are other crowdfunding websites other than Kickstarter. It would be interesting to compare this data to that of those companies.

**Other Possible Tables and/or Graphs and Value**

I would be interested to see the number of backers and average donation for each category. This would show if certain categories were getting more people to donate and/or more money per donation.

I would also like to explore the average donation in each month. As I was analyzing the data, I kept thinking if there are certain months that people are donating more money. This could help determine when to start a campaign to target those months.

I noticed two columns in the main table titled “staff\_pick” and “spotlight”. I would like to see if there is any correlation between these and the success rate of a campaign.

It would be interesting to compare the success rates in different counties. This could show the popularity of Kickstarter internationally.

For me, the most important table or chart to look at would be involving goal size. Is there a difference between the goal sizes of successful campaigns and unsuccessful campaigns (mean, median, maximum, minimum, variance, standard deviation)?

**Statistical Analysis**

Based on the data surrounding the number of backers, I believe it would be more beneficial to use the median. In general, medians are better to use than the mean when there is high variability and/or extreme values. The mean is more effected by outliers as well. Therefore, the median would be a better representation of a “central value” since there is a large variability.

There is more variability with the number of successful backers (higher variation, higher standard deviation). This does not make sense. I would have expected there to be a higher and more importantly, more consistent, number of backers for successful campaigns. Just looking at the number of backers does not take into account other factors such as average donation, which I think is needed to tell the whole story.

**Additional Analysis**

I started to wonder if there was a relationship between goal size and the success of a campaign. I found the mean goal size of successful campaigns was almost $40,000 less than the mean goal size of unsuccessful campaigns, the maximum goal for successful and unsuccessful campaigns was very similar. There was also much less variability among the goals for successful campaigns (successful campaigns had a standard deviation of about $30,000 less). This shows that there the success rate of a campaign could be related to the goal.