1. One conclusion that can be drawn after analyzing the data is that the largest number of crowdfunding campaigns (344 out of 1000) were skewed towards the Theater category. The Theater category takes up more than a third of the data sample. Within the sub-category analysis the difference between the count of Plays in the Theater sub-category is even more drastic than the rest of the sub-categories.

Also, judging from the outcomes of each crowdfunding campaign, if the campaign does not have a full 100% of funding pledged compared to their goal funding, that campaign was failed. Even the 3 campaigns that rounded up to 100% funding but were not fully 100% funded failed. If the campaign’s pledged funds were greater than their goal, the campaigns were successful.

Third, on an overall basis, it seems that the number of crowdfunding campaigns started starts off strong at the beginning of the year with 91 total campaigns, slows down a little, then picks up to hit a peak of 93 in July again. Surprisingly, the number of successful campaigns also hits a peak of 58 in July before they fall back down again. It seems that there may be a trend in a large number of campaign starts at the beginning of the year and then again a large number of campaign starts towards the middle point of the year.

1. One limitation of the dataset would be that the goal category is limited to 12 different ranges in which 2 of the ranges encompass a drastically different quantity of funding goals. The less than 1000 range and greater than or equal to 50000 range do not equal the range of the other 10 categories, and the greater than or equal to 50000 contains almost a third of the sample set. A better option to encompass the data more equally might be to identify the highest goal amount and then use that to equally distribute the ranges for the goals of the campaigns.

Another limitation of the dataset may be that there is no way to know how meaningful the Average Donation column is because the formula only identifies the average donation made using amount pledged divided by backers\_count. Due to this, there is no way we can know if the average donation amount is being skewed by small donation amounts or large donation amounts. The ability to see specific donation amounts might provide better insight to the meaningfulness of the average donation amount.

1. An interesting table/graph to see would be how or if the origin country of the campaign startup relates to the outcome of the campaign. So, for the rows, you can input the country and for the columns, you can do the outcome. This would be similar to the Outcomes by Category and Sub-Category tables, just instead of category and sub-category, the rows are the countries. This specific table could provide insight as to whether the country of origin has a potential effect on the outcome of the campaign.

Another possible table could be to see if the backer\_count has any relationship with the outcome of the campaign. Similar to the table above as well as the Outcomes by Category and Sub-Category, the outcomes would still be in the columns, but the rows would now be the backer count. The best way to capture this would be to use a range for the backer values similar to the goal ranges for Outcomes Based on Goals. This comparison table would be interesting to see because it may provide insight into the potential effect the number of backers a campaign has on the outcome of the campaign.

Bonus:

The Median is most likely a better summary of the data because the large difference between both means and medians for successful and failed outcomes indicates that the data may be skewed, meaning they may not have a normal distribution. So, judging from the fact that the means are much larger than the medians, the data may be skewed upward as a larger number of backers may affect the mean upward.

There is more variability with the successful campaigns than there is with failed campaigns because the standard deviation is higher for the successful outcomes (1267.37) than it is for the failed outcomes (961.31). This makes sense because the range for the successful outcomes is higher than the failed outcomes and the successful outcomes also have a higher variance than the failed outcomes.