There are some very interesting conclusions that can be determined with this data regarding the ultimate success and failure of this sample of Kickstarter campaigns. While I would have assumed that funding campaigns for film, music and technology would be among the majority, I was not expecting Theater to not only take up the top spot but have double of any of the previously mentioned three. This was quite surprising at first until I realized that most theater/play productions usually have smaller budgets and are only held locally or in a small number of venues. The data confirms this intuition by showing that 71% of all theater campaigns have a goal of less $10,000. There were 68 however that had larger budgets of over $50,000. Less than 12% of these were successful compared to 73% if the goal was set under $10,000.

The data regarding which months generated success and failures was a bit over the place. While I could have given an educated guess that December would have been one of the worst months to start a Kickstarter, I would have not thought that May would be the best month for one to be successful. Looking at this, December makes sense as money is usually allocated to other means both personally and commercially during the holiday season. Nothing of significance generally happens in May. To make this even more confusing, May also had around an average number of failures. Seems like May would have been among the highest as it was the 2nd highest in total campaigns started (more overall campaigns started = greater probability of total number of failures that month). This would be something I would love to look further into, perhaps beyond the data itself, to discover the reasons why May presents itself as the hot month for a successful Kickstarter.

The data also presents an interesting conundrum regarding how successful a campaign will be with regards to its goal. Common sense would say that as a campaigns goal increases, so does its chance for failure and cancellation which the data generally confirms. However, the data takes a 180 when looking at the ranges from $35,000 to $39,999 and $40,000 to $44,999. Successes decline and failures increase as goal total increases until they hit these two ranges which both increase. Success increases from 39% to 47% and then from 47% to 49% while failures decrease from 45% to 40% and then from 40% to 37%. After this point, success rate plummets while failure rate skyrockets.

Perhaps my biggest complaint that I have with this data is that it is a bit too organic to determine a useful conclusion regarding a Kickstarter success or failure. With regards to the outcome of theater campaigns, there are many questions that are not answered. Each theater has its idiosyncrasies that would require further research beyond this data set. Questions like “Is this play Local”, “Is it National?”, “Does it have any advertising? If so, is it only locally? Nationally?”, “Is the material something that many people have seen before or is it a new production?”. Conclusions in this data set feel like they would be a great springboard into these questions as information with a binary conclusion often cannot answer them or only partially.

There are tons of further details that can be discovered within this data set. I did some further work to compare the total number of campaigns within the specified ranges for Film and movies, Theater, Music and Technology. Not surprising, three of the four resembled power functions (where a vast majority of data points are in the first couple ranges). Both theater and music look like they could easily be fitted with the gamma distribution that would allow for some predictions and further analysis if campaign count was all we were concerned about.

Film & movies and technology look like they would require other methods for further analysis. Film & movies resembles a gamma distribution but has a very strong weight for the $50,000+ that is affecting the data. Technology is something totally different as it almost looks like a reverse power function as it is extremely affected by the $50,000+ range as nearly half of the date points come from that area. Much further work such as a transform or expanding the model with more variables would need to be done to be able to use this data for predictions with Kickstarter Technology campaigns.

Lastly, I figured trying some regression for the Theater data would be intriguing. I was interested in predicting the total amount pledged to a campaign so I set up 2 initial models; one using number of backers, average donation, and goal with the other using these 3 plus a dummy variable assigning a 1 for a successful campaign and 0 for every other outcome. After running both, I discovered that goal was not significant in either model, so I dropped it and ran a final version with number of backers, average donation and the success/failure dummy variable. Not surprising, the predicted values were all other the place and all models yielded an R score under .38. None of the three models would be good for further analysis which was inline with the problems discovered earlier using just total number of campaigns. Further methods would have to tested.