**Rice University Boot Camp | Data Analytics**  
**Unit 1 Homework: Kickstart My Chart; by Brian Labelle ( 2019-06-15 )**

1. **Given the provided data, what are three conclusions we can draw about Kickstarter campaigns?**
   1. **IF** this small sample set represents the overall picture of Kickstarter funding and it does seem to mirror just logic that there is an elevated proportion of successful campaigns versus lower funding goals. The sweet spot seems to be the $1k to 5k goal bracket. With the obvious higher percentage of failure with over $50k campaigns and drop off of successful campaigns below the 20% mark.
   2. Again, **IF** this small sample set represents the overall picture of Kickstarter funding, is really interesting to see the elevated success rates in the music and theater categories. It seems that the creative arts have higher success rates then other types.  
      * 31.8 % of all of the successful Kickstarter campaigns were plays.
      * 25.9 % of the total Kickstarter campaigns were plays
      * 65.1 % of all of the Kickstarter plays were successful
        + 412 US plays were successful & 238 UK plays were successful
   3. If we could analyze the entire dataset from Kickstarter it would be interesting to see if seasonal success rate plunging between May to September is across the board every year. We can draw a possible quick conclusion of the successful drop off rate in December due to the holidays and potential backers focused on other priorities. If we believe what Mashable said in 2012 about the biggest factors is the campaign’s timeline. They claim that projects lasting 30 days or less have the highest success rates. So it’s probably best to start your Kickstarter campaign in April to hit the 30 day peak in May. ( [link](https://mashable.com/2012/05/13/kickstarter-tips/) )
2. **What are some limitations of this dataset?**

According to the information provided to us in the exercise at the time of the published data, we are only analyzing 1.3% of the projects some 4000 out of 300 000. We are uncertain if the sampled datasets were skewed in anyway by the originator; there seems to be a penchant towards plays (theater) in this dataset. This observation also mirrors the small sampling of only analyzing 2.3% of the total raised at $46m versus the touted 2 billion raised. We could easily draw better truths and have better story telling if we were able to have the entire datasets from Kickstarter. It would also be interesting to compare the entire data from Kickstarter to other crowdsourcing services and analyze if they share similar success rates. There are several outliers that really need to be cleaned out, having a goal of $1 and raising $22k or having a goal of $10 and raising over 30k skews the numbers.  
  
After reading a little about coefficient of variation ( CV = standard deviation / mean ) by [Joshka Kauffmann](https://www.researchgate.net/post/What_do_you_consider_a_good_standard_deviation) ( University College Cork ).   
  
He states: *“ For an approximate answer, please estimate your coefficient of variation (CV=standard deviation / mean). As a rule of thumb, a CV >= 1 indicates a relatively high variation, while a CV < 1 can be considered low. This means that distributions with a coefficient of variation higher than 1 are considered to be high variance whereas those with a CV lower than 1 are considered to be low-variance.”*So, if I am to believe that I calculated the variance and standard deviation correctly, I believe that my coefficient of variation is 4.34 for successful campaigns and 3.47 for failed campaigns. So I believe that the data is considered to be in high variance.

1. **What are some other possible tables and/or graphs that we could create?**According to Mashable 2012 KickStarter tips, ( [**link**](https://mashable.com/2012/05/13/kickstarter-tips/) ) "Kickstarter projects with a video component succeed at a higher rate than those without (50% vs. 30%). " - It would be an interesting metric to see how many of these successfully campaigns out of the 4000 we analyzed used a video.   
     
   I would like to dive deeper in to analyze why none of the journalism campaigns were successful. I would assume that in this day and age a comparison of old-world media type projects versus new digital media campaigns would totally be biased towards the newer modern media content.   
     
   It would also be interesting to see how the duration of the campaign impacted the over all success rate.