**UC Davis Data Analytics Bootcamp**

**Matt Jensen**

**Excel Homework #1: Kickstarter**

1. Given the provided data, what are three conclusions we can draw about Kickstarter campaigns?
   1. Theater productions are the largest category of campaigns and the second most successful in terms of percent of campaigns meeting funding goals.
   2. It appears to be easier to receive Kickstarter funding during the first half of the year, and much harder in December (likely when people are distracted with the holidays)
   3. Later years have many more Kickstarter campaigns than earlier years. This may be due to Kickstarter being newer to market and not as well known to the general population or perhaps due to the earlier years being in the midst of the financial crisis.
2. What are some limitations of this dataset?
   1. It appears that our dataset is just a small subset of the overall Kickstarter campaigns. The instruction set mentions “more than 300,000 projects launched on Kickstarter” and also mentions that only one third make it through the funding process. However, our dataset is limited to just over 4,100 campaigns. Further, there may be some bias or sampling issues with our data, as we show 53% of the campaigns in our dataset have been “successful” in reaching funding goals.
   2. One limitation is that Kickstarter campaigns often have varying donation/ reward combinations. It would be nice to have more information on these pairings (maybe lowest donation amount that includes a reward (i.e., product being launched).
   3. Another limitation is this dataset doesn’t contain data regarding campaigns post funding. Many campaigns meet their funding goal, but ultimately aren’t able to launch product, and it would be interesting to analyze data related to “successful” campaigns and whether they ultimately are able to ship product or not.
3. What are some other possible tables and/or graphs that we could create?
   1. Percent of campaigns that are successful by category
   2. Comparison of percentage of successful campaigns to the average donation amount (hypothesis being that campaigns with lower average donations per backer might have higher success rates because backers have less to lose if they back a campaign that ultimately fails).
   3. The deciles and quartiles of the number backers of successful and unsuccessful campaigns
   4. Interquartile range of the number backers of successful and unsuccessful campaigns
   5. Analysis of the dataset excluding clear outliers in the data to get a more reasonable picture of the typical campaigns.
   6. Analysis of data by year (some review of the data shows many fewer campaigns in the earlier years vs. later years in the dataset).