**HW1**

Marc Omar Haddad

UNCC-Data Bootcamp

September 21, 2019

Q1)

*Given the provided data, what are three conclusions we can draw about Kickstarter campaigns?*

Conclusions:

1. Campaigns that start between Feb. and Jun. during any year generally enjoy a better success-to-failure ratio. Conversely, campaigns that begin in the months of Sep. or Dec. are more likely to fail. Cancellation rate remains fairly consistent throughout the year.
2. Music campaigns have the best success-to-other ratio (“other” being a combination of “cancelled” and “failed”), whereas Theater campaigns have the largest absolute number of successes at the cost of higher failure and cancellation rates.
3. After factoring out goal numbers that are essentially guaranteed to succeed (which I have determined to be between $1-$100), the success rate of any campaign appears to be determined in significant part by the average percent of funding. I found that most successes have an average percent funding of 160% of requested goal. Conversely, I found that most failed campaigns have an average of 9.4% funded of requested goal. Interestingly, cancelled campaigns appear to be more frequent when the relative average funding is around 69.4% of the campaign goal.

Q2)

*What are some limitations of this dataset?*

Limitations:

* “Joke” campaigns appear to be included in the data (e.g. campaigns with $1 goals, or campaigns with ludicrously high goals that are not meant to be attained), thus significantly skewing averages and success/failure ratios. Arguably these outliers can be filtered out and accounted for, but the question still remains if there are other “joke” campaigns our filtrations may have missed.
* The dataset is also out of date (latest figures appear to be 2017). This is a small limitation, but a limitation, nonetheless.
* This dataset does not include peripheral-yet-significant information that may affect people’s willingness to spend money on campaigns (e.g. economic indicators, political environment, backer demographics, financial status, money spent on marketing, etc.). Therefore, the dataset may be missing a key piece of information that could explain why backers are more likely or less likely to invest money into campaigns.
* An inherent issue with any dataset with descriptive subcategories is the fact that there are a multitude of valid methods to subcategorize qualitative data. Few things in reality can fit into perfectly neat categories, and thus subjective calls have to be made that may result in fundamentally differing conclusions dependent upon the individuals doing the concluding.
* This dataset also does not show the volume of backers during any one point in a campaign’s lifespan. Therefore, we cannot answer questions like “do campaigns mostly gain backers when they are announced? Or do they see more success as their campaigns progress?”

Q3)

*What are some other possible tables and/or graphs that we could create?*

Tables/Graphs we could create:

* A table that analyzes the effects of “spotlights” and “staff picks” on the state of a campaign.
* A statistical analysis on the effects of “mega-donors” vs. “micro-donors” on rates of success.
* A graph that would visualize the average donation per backer according to their geographic location.
* A table comparing the rate of backer acquisition and campaign subcategories.