**Homework #1 – Kickstart My Chart**

**Andrea Kwong**

1. **What are three conclusions we can make about Kickstarter campaigns given the provided data?**

There were more successful campaigns than failed campaigns, with the majority of campaigns typically launching from May through August. The success to failure ratio of kickstarter projects was greater before 2014 (3.51) than after 2014 (1.16). While there were a larger number of projects launched after 2014, the percentage of successful projects is less compared to before 2014.

The US had the largest number of campaigns out of all the countries with 3,038, which represented almost 73.8% of all campaigns. GB, the country with the second largest number of campaigns with 604, which was just 14.7% of all campaigns. Whereas almost 60% of GB’s campaigns were from the theater, the US had more categorical balance in its campaigns, with 30% of campaigns from theater, 20.9% from music, 13.9% from film & video.

Theater, specifically plays, had the largest number of campaigns (1066), both successful (694) and failed (353). Music, had the best success to failure ratio of any category, with the majority of subcategories (rock, pop, metal, classical music, electronic music) having only successful campaigns.

1. **What are some of the limitations of this dataset?**

The dataset may not provide all the variables that may influence the final state or outcome of a campaign. Variables such as updates on campaigns, use of social media, the story behind the campaign, whether or not there was a video are all other factors that could affect the outcome of a campaign and are not present in this dataset. Another limitation is that the state, or dependent variable is categorical and not a continuous quantitative variable. In order to perform a linear regression analysis to determine explanatory variables for whether a campaign is successful wouldn’t be possible using the linear regression model. We would need to use other types of models. Another limitation is that we don’t know if this sample was a random sample or how this sample was selected, which can create selection bias that can skew data results. We also have not accounted for outliers which may also affect results.

1. **What are some other possible tables/graphs that we could create?**

We could look at the difference between ‘date created conversion’ and ‘date ended conversion’ and put them in different ‘days or month bucket intervals’ and create a stacked bar chart to count how many campaigns were successful, failed, or canceled. This would show whether the amount of time that campaign ran for determined success.

Similarly, we could also create bucket of intervals for the ‘backer\_count’ and create a clustered column chart to see trends on whether the more backers for a campaign means a greater likelihood that a campaign will be successful.

We could also create a clustered column chart on whether a ‘staff\_pick’ (true or false) had an impact on the outcome.

Another graph we can make is looking at the outcomes by country with a stacked column chart which would show us where these campaigns originate and their outcomes.