Excel HW

* Data Observations

1. Three Conclusions
   1. The KickStart Program data set enables various observations to be made, some more prominent than others. For instance, the most successful programs come from theatre, music, and film/video. Out of these three categories, theatre has launched more programs and had more successful programs as a result. However, the music category has a better percentage of successful programs even though their actual startup program count is less than the other two categories. We can therefore conclude that these three categories had more success than others.
   2. Many of these programs were arranged in broad categories but they had many sub-categories. It is vital to acknowledge these sub-categories due the differences in “state” status of these sub-categories. Some sub categories did very well while others failed. In the case of the three most popular successful areas mentioned in point “a” (theatre, music, and film/video) the most successful sub-categories were; plays, rock music, and documentaries. Therefore, it points out that your campaign is not guaranteed to be a success just because you are in a certain “parent category”
   3. From the observations stated above, it’s imperative to state the correlation between success and percentage funded. Those programs that had higher percentage funded, showed more success and usually had more longevity than the unsuccessful campaigns. Furthermore, more often than not, the successful categories/sub-categories had more backer support as well.
2. Some of the limitations in the data set is the “Staff Picked” and “Spotlight” data. There seemed to be a relationship between these two data sets and the status of the programs. Programs that were staff picked and in spotlight were most likely successful. However, if there was more information on what these data entailed, I feel they would serve a better purpose in uncovering trends within various data these projects. I also believe the date created and ended have some limitations for it doesn’t reflect the success of the programs as one may think. For example, many successful programs are short lived based on the date created and ended data. There needs to be more data to supplement these dates so there can be meaningful information extracted to support the discoveries of hidden trends. A way to amend this ,maybe, information showing the duration of campaigns for each respective category so one can expect to see the average time a successful campaign is running compared to that of an unsuccessful one.

In addition to the limitations above, I also think the countries need to be broken down into subcategories as well, most likely states (California for example) or regions (Northwest, West, Mid-West, East Coast, West Coast, and more). I believe this could show preferences and trends of campaigns by region. Some campaigns would be more primitive in one region as opposed to another. With this information, future campaigns would know what locations give them the best chance of opening a successful project. The regional data collections could further unveil trends that can be applied to other regions for their benefit. For example, if some regions display better stats than other regions (i.e. backer support and funds), then there could be further analysis performed on those advantages. As a result, trends can be revealed which can be applied in struggling regions with similar campaigns.

1. Some graphs that could add value are tables showing relationships between spotlights and staff picked. It seems that these categories are associated with favorable project “states”. Moreover, tables combined with data metrics (mean, variance, etc..) on percentage funded and average donation could reveal some favorable trends. Conclusions can be made about funds and average donations. The data can indicate the average amount of support needed to make a certain campaign successful. This will allow potential campaigns to have proper awareness of the support they will need to have a fruitful campaign.
2. Bonus Statistical Analysis
   1. The mean would have more meaning in summarizing the data as opposed to the median. The median simply states the count of the backers that is located in the middle of that data set. While the median offers to show what is the average amount of backers a successful project got as opposed to the failed campaigns. This would give us some meaningful data to see if the number of backers have significant meaning or value to the “state” (success or fail) of a campaign.
   2. There is more variability within the successful campaigns. This variability makes sense due to the different types and aspects of each campaign. Some campaigns may require more backers than others. However, the disparity of the mean when it comes to overall comparison with the number of backers amongst the successful and failed projects is very apparent. Even though there is less spread among the failed projects, there is no denying that failed campaigns have far less backers than the successful campaigns. These difference do reveal that successful projects tend to have more campaign backers than less successful projects.