Justine Sbarra

**Homework #1 - Excel**

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

From the first graph, you can see that Kickstarter projects are generally more likely to be successful than to fail or get cancelled. About 53% of all Kickstarter projects are successful, whereas about 37% fail, and about 8.6% are cancelled. If you look at the graph on a more granular level as opposed to a general level, it examines the relationship between category and state. From the graph you can see that certain categories are more likely to be successful than other categories. For example, theater, music and film and video have the highest success rates, but also the highest number of projects in that category. Food, games and publishing have the lowest success rates, but also the lowest number of projects in that category. It Is also interesting to note that all journalism projects were cancelled from this data, and in the technology sector, more than half the projects failed. All interesting information that can be drawn from this data and would help backers and creators gain insight into the different categories. For example, someone may interpret this data to show that some categories like theater are likely to be more successful, whereas categories like technology look like they are almost half as likely to fail as succeed, and why that may be.

The second graph shows the relationship between sub-category and state. There are many interesting points to highlight in this one. First, when you filter the graph and look at successful and unsuccessful projects for each sub-category, there are many that stand out. There are some sub-categories in which all projects were unsuccessful – children’s books, animation, drama, faith, fiction, food trucks, gadgets, jazz, mobile games, etc, There are also some sub-categories in which all projects were successful – classical music, documentary, electronic music, hardware, indie rock, metal, nonfiction, etc. Interesting to see that a lot of the music sub-categories are 100% successful, which makes sense because music was one of the higher success rated categories from our first analysis. You can also see from the subcategories, that one particular category is extremely popular – plays. Plays are very popular, and have a pretty high success rate.

The last graph, the line graph, visualizes the outcome of the projects in each month based on the date the project was created. One of the main things to note on this graph, is that the end of May is the peak of the graph where the number of successful projects is the greatest. It is also when the the count of successful projects decreases significantly as the months progress until September. However, during this time is when there is an influx of projects, and although there is a stark drop in the success count, a lot of projects ARE still successful. Also, during this time, the difference between the successful and unsuccessful projects gets narrower, and in December, the amount of successful projects is actually lower than the amount of unsuccessful projects. You can see that the amount of canceled projects pretty much stays at a constant throughout most of the year, with little change.

**2.What are some limitations of this dataset?**

One of the limitations of this is that the dataset may not be large enough so may not be statistically significant. You need to have a large dataset in order to draw solid conclusions of the data at hand. Also, this data only represents data on Kickstarter, so in order to be able to draw deeper conclusions about crowdfunding as a whole, more data would need to be collected from other places, not just KickStarter. Also, 74% of this data is accounted for by the US, so maybe being able to break down the data into cities in the US could have provided some more information that would have been beneficial as opposed to seeing it by country.

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

One graph that could be insightful and interesting would be seeing the relationship between the duration of the projects (date created to date ended) vs the state of the project. It would be interesting to see if there is some sort of relationship there. Another graph, which I mentioned above, would’ve been analyzing the actual states in the US that contributed to these projects, since 74% of the projects were from the US. Another graph that might be insightful would be looking at the relationship between the average donation vs category/subcategory. Is there a trend in which projects receive higher donations compared to others?