**Kick Start My Chart**

In this assignment, we are asked to analyze a set of data in Microsoft Excel based on the failure, success, canceled and live states of 4,000 Kickstarter projects from their conception dates to their deadline dates. By categorizing and creating graphical representation of the data we conclude how well each project performs based category, subcategory, and launch date. Furthermore, we can explore the limitations are within the dataset and how other graphs and/or tables can help better visualize the dataset.

Based on category we can see from pivot table our “Category Stats” that the “theater” category has higher number of successful Kickstarter projects (839) than other categories but with also with the highest number of failed projects (493). We see that “music” has the second highest number of successful projects (540). “Music” failed projects (120). We can visually see from the stacked column pivot chart that the “music” category has greater percentage of success when comparing to its other states of failure, canceled, and live. Although “theater” seems to have a larger number of projects on the table, the chart shows that the ratio of failed projects to successful projects is significantly larger with “music” or even some of the other categories. Why are there people even trying to Kickstart something in “journalism”? Twenty-four canceled projects? Pathetic.

If we look at the “Subcategories Stats”, we can further see which specific category contributes to overall success of the main category when it comes to Kickstarter projects. If filter the pivot table to “theater” we can see that plays contribute to most projects under “theater”. The pivot stacked column chart shows that if we were to take the total number failed play projects, it would still exceed the stack total number of “musical” projects and “spaces” projects. We can also conclude that “spaces” and “musical” projects have roughly the same number of failed, successful, and canceled states across all countries. From the pivot stacked column chart if we were to select all for “Country” and all for “Parent Category”, we can see that “plays” outshines all subcategories when it comes to total number of projects. However, taking a closer look, we can see that subcategories “rock”, “hardware”, and “documentary” are clearly 100% successful.

For our Excel sheet, “Outcomes Based on Launch Date”, we can see that if we filter “All” for “Years” and “Parent Category” there is no real significance between the success and fail rate for which month of the year a project is launched. Although projects do seem to be a bit more successful when launch around may, and less successful around a December launch; could this be a correlation with tax returns and holiday spending, respectively?

There are some limitations to the data provided. Some of the questions that could be asked to reveal as to why certain categories fail and some succeed are: does taxes or holiday influence the success of a project? Why was a project canceled? Does the marketing of the product or service entice the audience? Is this controversial/safe product or service? Was there a pandemic that year? Was there a recession that year? The data provided only considers the launch and deadline data and categories as criteria for success. As there could be many factors to the success or failure of a project.

We could get other and better insights by graphing and tabulating other parameters, such as:

* goal versus pledged versus state
  + Does the goal size affect the successful of the Kickstarter?
  + Use a scatter plot to determine how spread out the goal and pledge data when filtered by Category or Subcategory
* country versus state
  + Could be used to determine how successful each category is in each country
  + Use a Map to Illustrate which category and subcategory is successful when filtered by year

**Statistical Analysis**

From the statistical numbers on the number of backers of campaigns that were successful, we see that since there is a such a large difference between the maximum and minimum backers the mean (194) can become misleading, whereas the median (62) is less affected by the outliers. In order to measure central tendency of this data set, it is better to use the median value. It could give the campaigner more insight on how a campaign can be successful. If a new campaign can get the number of backers equal or more than the mean of the past successful backers, the more likely that campaign will succeed. This also applies to the how likely a campaign will fail. If we look at the mean (3) for unsuccessful campaigns, we can see that the minimum number of backers for a unsuccessful campaign is 0 and the maximum is 1501. This will skew the mean, 19. If the campaign has 1501 backers and still fails but a successful campaign only 1 backer, we can see how misleading the mean can be.

There is more variability between backers of successful and unsuccessful campaigns. Unsuccessful campaign backer variance is around 5243; whereas the variance for successful campaigns is 712840. This is likely due successful projects meeting or even exceeding their goals, whereas unsuccessful backers never do. Not all campaign has the same goal levels. But if a campaigned is canceled (keeps the backer numbers low) or fails to reach its goal that makes the variance for the number of backers for unsuccessful campaigns even smaller. We can see that one successful campaign can succeed with 1 rich individual but also succeed with many backers.