**Discussion Questions**

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

Using the information in the Kickstarter Campaigns data set, we could conclude which category of campaign is the most common, which category will most likely succeed, and infer which categories are generally supported by wealthier backers.

Using a pivot table and corresponding pivot chart broken out by category, the most common Kickstarter campaign is listed in the theater category. This is followed by music and then film and video. The pivot bar chart shows us the combined number of campaigns submitted in each category, regardless of success or failure. The least listed category for a Kickstarter campaign is journalism, which includes only cancelled campaigns. This information could be useful for either a person starting a campaign or for a potential investor. In aggregate, the most *common* campaigns are for performing arts.

The pivot table associated with this chart gives more specific detail about the campaign counts. By adding an additional column to the pivot table to calculate the “Percent Successful” for each category, a potential poster or investor could see which category is most likely to return a successful campaign. Looking that the chart, it is obvious that theater campaigns are the most common, however, the “Percent Successful” column on the table shows that music campaigns are the most *successful,* with theater being a close second and film and video third. Again, the most successful campaigns are for the performing arts. Another metric to determine the success of campaign categories could be comparing the number of cancelled campaigns between categories. There are many possible reasons why a Kickstarter creator may cancel a campaign, and based on Kickstarter’s rules, there is no penalty for cancellation. It is not possible to know from this data *why* some creators cancelled their campaigns, but an investor may be able to use the cancellation data to direct his or her money to a specific category. (That is, he or she could determine which category is the most likely to have a cancelled campaign and avoid investing in that area.) According to this data set, 100% of journalism campaigns are cancelled, while games and photography categories have failed campaigns, but no cancelled campaigns. Of course, this same strategy could be applied to this data to determine which category is most likely to have a *failed* campaign as well. The data presented in this paragraph is assuming looking at all countries. However, the pivot table and chart are filterable by country, so a potential backer could view the data just for their country (or country of investment.)

Another potential use for this data set is determining which category of campaign attracts the wealthiest backers. One of the columns added to this data set was “Average Donation” which is the average amount each donor contributed to a particular campaign. This is determined by dividing amount donated (“Pledged”) by the number of backers (“Backers\_Count.”) Sorting the data sheet in descending order by “Average Donation” and then filtering out the null values shows that 9 of the top 15 campaigns with the highest individual donations belong to the technology category. (1 of the top 15 entries was for film and video and 5 of the top 15 were for theater.) The campaign with the largest individual average donation was a technology campaign at $3,304.

1. What are some limitations of this dataset?

While this data set shows *whether* a campaign was successful, it cannot explain why campaign would succeed or fail or why a campaign would be cancelled. This data set also only considers whether a campaign was successful or failed (or cancelled) based on whether or not the campaign reached its funding goal. There is no column that indicates whether the product or service promised was delivered and to what extent the product or service itself was successful. In terms of just data, the inclusion of the “live” campaigns could be distorting the information. If attempting to make decisions based on historical data, the live campaigns would be excluded.

A problem with this data set is apparent in the conditional formatting of the “Percent Funded” category. There are some campaigns that had very small funding goals (example: goals of 1 and 10) which caused the percent funded calculation to show that these campaigns were several thousand percent funded. It seems unlikely (though possible) that a campaign that asked for $10 would receive $30,383.32 (ID number 1253) which correlates to 303,833% percent funded project.

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

There are many additional tables or graphs that could be created depending on what information was necessary for a particular analysis. Previously discussed (in question 1) were additions to the Category Pivot Chart. A separate table could have been created that calculates the percentage of successes, failures and cancellations for each category. Another table that compares the category and whether or not the campaign earned a “Staff Pick” could demonstrate what difference the “Staff Pick” designation makes to the chances of success for any particular campaign. An additional column on the data sheet could be added that provides the number of days each campaign was active. This could be incorporated into a table of successful campaigns by category to determine how many days the average campaign required to gain “successful” funding.

**Bonus Statistical Analysis**

1. Use your data to determine whether the mean or median summarizes the data more meaningfully.

In a total of 2,185 successful campaigns, the full range of backers was 1 to 26,457 investors. The mean number of backers is 194, while the median number of backers is 62. In a total of 1,530 failed campaigns, the full range of supporters was 0 to 1,293 investors. The mean number of backers was 18 and the median number of backers was 4. Since the median represents the middle value of a data set ordered sequentially, this means that 1,092 successful campaigns had 62 or fewer backers and 765 failed campaigns had 4 or fewer backers. The large numbers of backers at the top of the ranges is skewing the mean higher, making the median a more meaningful summary of each data set.

1. Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?

The successful campaign data set has a variance of 712,841 and a standard deviation of 844.3. The failed campaign data set has a variance of 3773.2 and a standard deviation of 61.43. The successful campaign data set has a variance that is over 188 times higher than the variance of the failed campaigns. The fact that there is a wider range of backers for successful Kickstarter campaigns makes sense. There are many reasons why a successful campaign could have very few or very many backers, including low goal funding amounts, high numbers of supporters that can only afford small donations, and few backers that are passionate enough to fund all or a majority of a project. A smaller variance among failed campaigns makes sense. The median number of backers in the failed set is 4 and the mean is 18 backers. This shows that for the failed campaigns, many of the projects had very few supporters. These projects, for many possible reasons, were not able to attract the number of supporters necessary to fully fund a project. It’s also important to note Kickstarter rules here. Kickstarter has an “all-or-nothing” funding model which requires a campaign to have at minimum the “goal” amount in pledges in order to fund a project. This means that while some projects had support, they were unable to reach their funding goal during a prescribed amount of time and were deemed “failed” campaigns. In that case, it makes sense that a project with few backers would not garner the necessary attention to attract additional backers, thus keeping the number of backers for a campaign smaller.