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

One of the conclusions I was able to draw about the Kickstarter campaigns is the category theater has the most successful campaigns compared to the other successful campaigns of the other category ones. The second conclusions are the sub category plays have the most successful campaigns compared to the other sub category ones. The third conclusions I was able to draw is that the Month of December is the worst month to start a Kickstarter campaigns because there are more failed campaigns than successful ones.

2. What are some limitations of this dataset?

One of the limitations of this dataset is it doesn’t show the amount of people who were told about these campaigns and how these campaigns were told to people. For example, if someone or a group of people have started a Kickstarter campaigns and they have a lot of friends, family members, and connections who could spread the word and post it on social media, it makes it easier to reach the goal and have a successful Kickstarter campaign. For someone who doesn’t have a lot of friends, family members, and connections it makes it harder to reach the goal. Another limitation is in what city are these campaigns started from because in the US for the music category they have a higher chance of success if these campaigns are started in LA, Atlanta, and Nashville. I believe these three cities are the place you would go if you want to make music. Plus, these three cities have a good amount of people who are in the music industry but if the campaign was started in Albuquerque it has a lesser chance of being successful. The dataset also doesn’t show the gender and demographic information of the people who started these campaigns because this will provide us with a better picture of which gender and demographic have a higher success.

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

For the Pivot Table that created the Line Graph we could have instead used a 2-D Bar graph which we could have the y-axis to start with January from the top going down and ending with December. It will still show that there are more failed campaigns than successful ones in the month of December. We could also create a pie chart and it will contain the total amount of campaigns for each category or contain the total amount of campaigns for each sub category.

**Bonus Statistical Analysis**

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

I believe the median summarizes the data more meaningfully because the gap between mean and median is very large. The mean just calculates the average of all the backers and if there are a couple large numbers it will push the mean to be a higher number while the median finds the middle number between all the backers and doesn’t get influence if there are a couple large numbers.

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

There is more variability with successful campaigns and this makes sense because if you look at the standard deviation, successful campaigns have a standard deviation of 844.49 compared to unsuccessful campaigns 61.44 standard deviation. The mean and median of successful campaigns has a large gap between them compared to the unsuccessful campaigns mean and median gap.