HW 1: Excel Challenge

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

* From the data of 2009-2017, category of theater and subcategory of plays have the greatest number of projects started.
* For the 8 years of data we have, we can see a trend of increasing percentage of cancelations on campaigns with the increasing goal amounts.
* The group of campaigns with goal less than 1000 has the highest successful rate and lowest failure rate and cancelation rate (limitation: the goal amount in this dataset contains different currency format).

1. What are some limitations of this dataset?

* The amount of pledged is in different currency formats which makes the field incomparable between countries. Also, the data seems less intuitive for users.
* The geographic of the dataset is across difference countries. This might affect the results of the general public around the world. For example, the United States had the most campaigns in film & video field. However, we can’t conclude the successful rate of film & video campaigns are high (>50%) in any country because the US created 85% of the film & video campaigns.

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

* We could create scatter plot of ‘backers\_count’ of failed campaigns and successful campaigns to see if there are any outliers of the two groups that might affect the means of these 2 groups.
* We could create a column chart of count of states (successful, failed, and canceled) by country to see how the campaigns perform in each country.

Bonus2 Questions:

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

Mean is better to summarize the number of backers for successful campaigns because the variance is large and there are extreme values. However, median is better to summarize unsuccessful campaigns because the variance is smaller and less extreme values.

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 ‘backers\_count’ of successful group has a larger variance. That makes sense because there’s no upper limit of the numbers of backers of a successful campaign. People can still donate even though the goal has been reached. Also, the count of backers of a successful campaign can be as low as one when this backer donates just enough to reach the goal. However, for unsuccessful campaigns, one of the factors to fail is having not enough backers. Therefore, it makes sense ‘backers\_count’ of unsuccessful campaigns won’t go as high as successful campaigns.