Excel Challenge

Homework 1

Keyur Patil

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

The given data provides us an insight into approximately 4000 Kickstarter projects. With this sample size we can conclude on the following:

* + Based on campaigns per category, we can conclude:
    1. The *“Theater”* category contains the largest amount of campaigns with approximately two-thirds of them being successful.
    2. The *“Journalism”* category has the poorest scores with all of them being canceled, and thusly can be concluded that Kickstarter is not a proper platform for campaigns in this category.
    3. The *“Music”* category has the highest success rate, the *“Food”* category has the Lowest success rate, and the *“Technology”* category is split evenly in terms of success, failure, and canceled.
  + Based on the campaigns per sub-category, we can conclude that the sub-category *“plays”* contains the largest sample group with an approximate 60% success rate.
  + Based on the Date conversion table and graph, we can conclude that campaigns are more likely to be succeed and fail during the summer months (May, June, July).

Based on the provided data, we can also conclude that campaigns are far more likely to succeed on Kickstarter if “Spotlight” is given; most campaigns that succeed were given the spotlight, and most campaigns that failed were not.

1. What are some limitations of this dataset?

Limitations of the dataset include small sample size, and limited categories and sub-categories. The data set does not include reasons of success, failure, or cancelation, as this would allow us to find any outliers in the data set. Another limitation of the data set is location of the campaigns as a majority are from the United States.

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

Additional table and graphs that would provide additional useful information include:

* + A percentage table of successful, failed, canceled, and live for both categories and sub-categories; a pie-chart would give us a visual to compare this information.
  + A table that looks at the success, failed, canceled, and live rates verses how long the campaign was on Kickstarter for. This would further help us analyze what type of campaigns do well, in correlation to their duration Kickstarter.

Bonus

1. Use your data to determine whether the mean or the median summarizes the data more meaningfully.
2. Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?
3. The **Median** summarizes the data more meaningfully, as there are a few outliers in the successful and failed backer-count data set. These outliers skew the data, so the median is more meaningful than the mean to summarize the data set
4. By looking at the variance and standard deviation numbers of both campaigns, it can be concluded that there is more variability in the **successful campaigns**. This makes sense because of the larger spread with the successful campaigns via the standard deviation number.