May 16, 2020 Homework - Excel

Kickstarter Campaign Analysis submitted by J Kurt Baltero

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

* The three most successful campaigns include **Theater**, **Music**, and **Film & Video** categories. Whereas all **Journalism**-related categories have been cancelled.
* Among its sub-categories, **Plays** belonging to Theater have the greatest number of projects. The successful outcomes of Plays were almost doubled than its failures.
* Fewer projects were launched or created during the month of **December**, while the peak months were **May to July**. Lastly, a crowdfunding goal that amounts to **1000 or less** had more chances of success.

What are some limitations of this dataset?

* Some data cannot be presented directly and needs to be converted first (e.g. timestamp, further sub-categorizing data, currency conversions).
* The number/value of monthly campaigns varies over time and different over the years in the dataset.
* Possible biased in each country.

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

* We can create another table/graph to represent the relationship between “state” and “percent fund”, in order to determine the minimum percentage funding to go “live”. Also include Pie charts to describe proportions of categories and outcomes as whole.
* Histogram graphs, Box and Whisker Plots can also be created to identify outliers when analyzing table statistics.

Bonus:

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

The **Median** summarizes the data more meaningfully since the data and charts herein depict a "skewed distribution" (not symmetric). The extreme values have a substantial impact on the mean as it pulls the mean away from the center. Hence, median would be better to represent the data generally.

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 than unsuccessful campaigns. It makes sense because the values of successful dataset are more dispersed, while the unsuccessful dataset appear closer together.