From: Yi An Pan (Emma)

Subject: Excel Homework

Date: 13th of June 2020

*Questions*

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

- By analyzing through Category, theater had the highest successful/failed projects and technology had highest canceled projects. However, music had higher successful rate (~77%) and food had higher failed rate (~70%) compared to other categories.

- By analyzing through Sub-category, plays from theater section had highest successful and failed projects and web from technology section had highest canceled projects.

- By analyzing with monthly trend, data showed that February, April, and May had higher successful rate (~60%) and December had higher failed rate (the only month that canceled projects are higher than successful projects). Average successful and failed rate are ~54% and ~38% respectively.

**2. What are some limitations of this dataset?**

- data insufficiency: For example, most of the dataset projects are from united states. If data showed more projects from other countries, it is possible to lead to different results.

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

- Average donation and category/subcategory: to see if there’s correlation to influent successful/canceled rate

- Countries and Successful/Failed/Canceled/Live: to see the detailed of database and compare to the category successful/failed pivot tables.

*## Bonus Statistical Analysis*

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

Mean is more referable for unsuccessful backers count because standard deviation is too high on successful backers count. However, mean is more referable to successful backers count because variance is too high on successful bakers count.

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

Successful campaigns had more variability as minimum and maximum on backers count had much larger difference compared to unsuccessful. Also, variance/standard deviation are also high.