Unit 1 – Homework: Kick start my chart

# Mitsuko Kitazawa

June 14th, 2019

**Purpose:** Analyze data in StarterBook.xls

In this homework, we analyzed campaigns status per categories (fig 1) and per sub-categories (fig 2).

## Question 1: Give three conclusions

Here are the conclusions:

1. The most successful categories are “theater”, “music” and “film & video”. None of campaign in category “Journalism” was successful.

Figure 1: Status of campaigns per categories

1. Within a same category, some sub-categories had all failed or cancelled whereas sub-categories were all successful. For example, in “Film and video”, all campaigns in “science fiction” were cancelled, all campaigns in “animation” and “drama” failed. Successful campaigns are in the “Documentary”, “short” and “television”. It suggests that if campaigns are submitted in one of these three sub-categories, they may have better chance to be successful. Similar conclusion can be made for “rock”, “classical music”, “electronic music”, “hardware”,” metal”, “nonfiction”, “radio & podcasts”. For “Theater”, campaigns in “plays” are more successful than other sub-categories. None of the three sub-categories have a 100% rate of success nor of failure.

Figure 2: Status of campaigns per sub-categories

1. The campaign submitted in May are more successful that the one submitted in December. The cancellation rate is regular and does not depend on the submission month.

Figure 3: Status of campaigns in function of submission month

## Question 2: Limitation of the dataset

The datasheet is handling campaigns from different countries, and it uses different currencies. It will good to convert all the budget information in one same currency (like USD). It will make the budget comparison easier.

“Successful” and “Failed” are not defined. It is very subjective to make conclusion. Does “Successful” mean the project was financed, or it requires a good feedback?

## Question 3: other possible tables and graphs

I would like to look if the country of the project does impact the campaign status. I would like to look the financial aspect. Is it easier to get a project with a smaller budget?

# Statistical analysis

Median is the value at the middle of the dataset. Mean is the average value of the dataset. Both are representing different aspect of the dataset. For the “successful” campaigns, the median is 62, meaning that half of the backer count is less than 62, whereas the average backers count is 194. This indicates that the dataset is scattered, and it is confirmed by the standard deviation, the variance, the minimum and maximum values. The “Failed” campaigns are less scattered, but values are generally lower than numbers for the “Successful” projects.

In my opinion, both median and mean are needed to understand the dataset.