GW Data Analytics Bootcamp – Homework 1

Fiona Nguyen

***1. What are three conclusions we can make about Kickstarter campaigns given the provided data***

On Kickstarter, campaigns in the theater are the most popular campaign, followed by music and technology. The film & video category leaps over technology to be in the third place with the most successful campaigns. Campaigns in the journalism category are the least popular, with most of them being cancelled.

While the sub-category “plays” has the greatest number of successful campaigns (almost 700 campaigns), it also has the highest number of failed campaigns (353 campaigns).

From 2009 to 2017, the number of successful campaigns peaked in May, but the trend headed downwards as the year progressed. This could be because less campaigns were initiated during the last months of the year. During the same period, the highest numbers of failed campaigns fall during the summer months, and then again in October. Throughout the years, the number of cancelled campaigns remained between the range of 20-43 campaigns.

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

The trends shown here only represents Kickstarter, but there are other crowdfunding websites, such as GoFundMe and CircleUp.

While the dataset contains over 4000 entries, this only gives a snapshot of the whole picture, as Kickstarter sees over 70,000 campaigns each year. To get a better picture, we should analyze all entries.

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

We could create:

* a time series plot (line chart) of a parent category,
* a table of percentage of successful/failed/cancelled campaigns over time,
* or a chart with numbers of successful/failed/cancelled campaigns by goal amount, filtered by category.