Abigail Herrup

HW 1: Kickstarter Data

Background and Overview:

Over $2 billion has been raised using the massively successful crowdfunding service, Kickstarter, but not every project has found success. Of the more than 300,000 projects launched on Kickstarter, only a third have made it through the funding process with a positive outcome.

Getting funded on Kickstarter requires meeting or exceeding the project's initial goal, so many organizations spend months looking through past projects in an attempt to discover some trick for finding success. (Source: Background in Git).

Below, is an analysis of the data from Kickstarter. Trends identified will be provided in the following analysis. An overview of how the data was cleansed and a final recommendation will be provided from analysis preformed. At the end, possible limitations and biases of the data will be discussed, along with analysis that should be done in the future.

Analysis:

From the data there were trends that became very evident. The category will the most Kickstarter campaigns, by count, was Theater. They had almost double the amount of Kickstarter campaigns compared to the count of other categories (as shown in Chart A below). Not only did they Theater have the greatest number of Kickstarter Campaigns, they also had the largest number of successful campaigns (Chart B). In the Theater category, the greatest number of campaigns were done for the sub-category of Plays (Chart C).

When examining all categories, the month of the year with the largest success rate was May (Chart D). In Comparison, the month with the largest failure rate was October (Chart E).

When looking at Kickstarter Campaign Goals, the average successful project had a goal of approximately $9.9K. The largest number of all Kickstarter Campaigns and successful Campaigns had goals between $1k and $5K. The average of successful projects is higher than the goal range for the highest number of total counts due to the second largest group of successful campaigns having a goal range of $5k to 10K

Data Cleansing:

To cleanse the data, a few activities were done. First, was the average donation and percent of the campaigns funded were calculated. Then, the dates that of when the campaigns started and ended were converted from the Unix timestamp to a calendar standard view of the dates. Lastly, we separated the category and sub-categories into different columns in order to create views using one or the other in exclusivity.

Recommendations:

Based on this analysis, a couple recommendations have been created. Firstly, the best month to have the Kickstarter Campaign begin is May. One will have the largest change at success when

beginning in that month. Also, when setting a goal for one's campaign, it is recommended to set your goal in the range of 1K to 10K. By having a goal in this range, you will have a higher likelihood of success.

Data Limitations and Biases:

There are a few large limitations in this data. The first is that the majority of the data is taken from the US, so if one was launching a campaign in a different country, this data would not be relevant. This data is also mainly collected from Theater Campaigns, having a total of ~1.4K data points from that category of Campaign (a quarter of all the data), while they only have 24 data points from journalism. If one is looking to start a Campaign for a category that is not widely represented in this data, they should preform another analysis.

Future Work:

Given more time, there are a couple of additional analyses that could be performed. One could look at the likelihood of success by country. Also, an analysis of the average donation compared for each category could prove useful. One could also collect additional data from other countries, to get an even representation of the world in the data. A model to predict the likelihood of success based on the category/sub-category of campaign, the goal, and the country the campaign is from could be created.

Chart A:

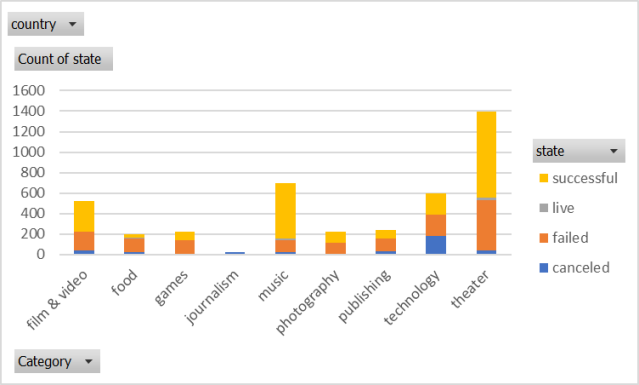

Chart B:

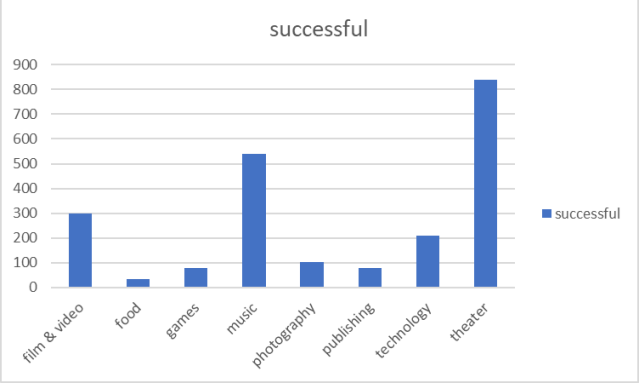

Chart C:

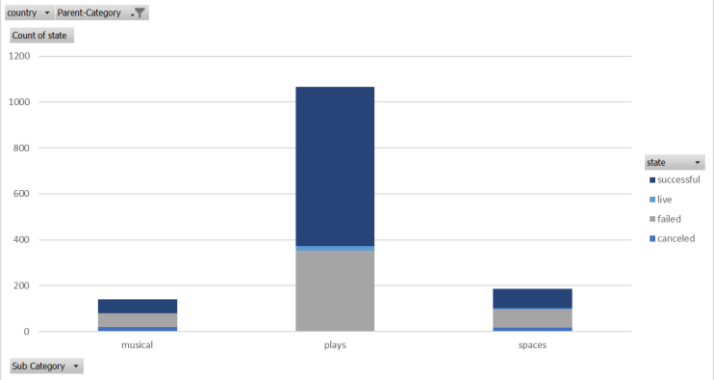

Chart D:

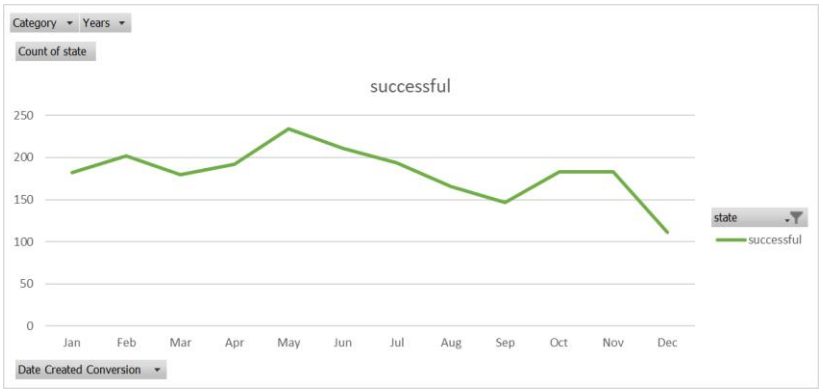
Category
Years
Count of state
successful
250
200
150
100
50
0
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Date Created Conversion
state
successful

Chart E:


Category
Years
Count of state
failed
160
140
120
100
80
60
40
20
0
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Date Created Conversion
state
failed