Case Western Reserve Bootcamp, Data Analytics Homework

Due 5/11/2019

Kickstarter Word Document

Carol M. Pietro

Answer the following three questions:

What are three conclusions we can make about Kickstarter campaigns given the provided data?

What are some of the limitations of this dataset?

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

The Kickstarter data covers 4114 projects from dates launched ranging from 2009 to 2017. The data set tells us that 349 of the projects were cancelled, 1,530 failed, 50 were live, and 2185 were successful. A campaign on Kickstarter, during this date range, had an 54% chance at success / live status versus a 46% chance of failure / cancellation. We know the max average donation was more than $3,300 and the minimum average was at least $3. The lowest goal for a campaign was only $1 and there were three of these, each listed as successful with average pledges of $10,656, so we also know you don’t need to have a lofty goal to meet with success on Kickstarter. We also know the highest goal was $100,000,000 with $0 in pledges and a cancelled status, maybe there is a limit to the goal setting for a chance at success. We can also see at a glance that the category of Theater has the highest number of campaigns total, 1,393 with 839 successful. A healthy, 60% chance at success. While Music may not have had as many total, 700 or total successes 540, the chances at success were greater, at a healthier 77%. An overall conclusion would be that Kickstarter is a good place to start!

There were some data points within this data set that prevent us from telling a full story. For example, there are no clear definitions of listed values of “state”. We are making assumptions that the terms success, live, cancelled or failure, are in related to the total goal and pledged amounts, but that still leaves us unclear about why a project was cancelled. The funds are listed as being in various currencies, but it is unclear if the values listed under goal and pledge have been converted so that a comparison can be made accurately. Additionally, there are several variables such as spotlight and staff-pick that we have no definition of, therefore determining any weight or value from those data points are all assumptions. And lastly, the data is old, ending in 2017 but there is a category of live for 50 of the projects which limits the value of that data point, since at this point in time they are not live and we do not know when the data was pulled to know how likely they were to have met with success or failure, base on the other data points we have.

The assigned charts provided give a good review of the data set however, I think it would be interesting to create a comparison of the campaigns that received a spotlight, had any greater chance at success as well as a graph comparing those who were a staff pick. There is also comparison to be had within the country range, for example, graphing out if those in the US did better or worse than those in Europe, or just a general line graph by country to see which met with the most success. And lastly, I would think it interesting to do a follow-up study to see how the projects were fairing now, compared to the status they were given in 2017.