Dawn Brenner

Data Analytics Bootcamp – June 10 Class

Homework 1

Analysis of Kickstarter program

1. Three conclusions that can be drawn about Kickstarter campaigns are:
   * Campaigns for Film and Music have significantly greater chances of success, 100% and 89%, than campaigns in other categories. (See Category Analysis tab in Excel file “Homework 1.”)
   * Kickstarter campaigns are used significantly more often for plays than any other sub-category and have about a 50 percent chance of success. (See Sub-Category Analysis tab.)
   * The lower a campaign’s goals are, the more likely it will succeed. (See Goal Line Graph tab.)
2. Limitations of this dataset include:
   * The dataset ends with a creation date of 3/15/2017. The missing most recent two years of data could provide inciteful information.
   * I would like more information on what constitutes a canceled versus failed campaign state. If most canceled campaigns are simply campaigns that were headed in the direction of failing, it would be useful to combine those outcomes/states.
   * If someone is looking at this information to estimate a new projects’ likelihood for success, they should know if the campaign was advertised outside the Kickstarter program or if the project was listed as a “Featured Project” or “Fresh Favorite.”
   * Define Staff\_Pick and Spotlight in the data. In this data, “Staff\_Pick” fields are FALSE for all data, but Spotlight is not. If I knew what it meant, I might glean information from this data (especially if Spotlight campaigns were “Featured Projects” on the Kickstarter website.)
   * One hundred twenty-three successful campaigns had 0 backers. How does that work?
3. Other possible tables and/or graphs that would be interesting include:
   * I’m curious if Kickstarter continues to grow or decline in popularity overall and in specific categories over time. A stack bar chart of year vs campaign count stacked by category would provide this information.
   * Per response number 2 above, I would like to see the outcome charts plotted with the failed and canceled categories combined. This would provide a more accurate picture of a go or no go campaign.
   * The Launch Date Line Graph would be more interesting to me if the X axis was the duration of the campaign (Date Ended less Date Created) rather than month or just a timeline since Kickstarter was started.
   * If I were interested in using Kickstarter for a particular category, I would like to see more graphs about what makes that particular category more or less successful such as a general theme to entertainment campaigns or target audience. (Keywords would have to be taken from the description.)
   * I would like to see more visual data from the Average donation such as average donation per category and average donation versus outcome/state.

Bonus Analysis Question: Mean vs. Median

In this example, I believe the median summarizes the data more meaningfully than the mean. The standard deviation for backer counts of both successful and failed campaigns are high (especially for successful campaigns) indicating a high number of outlier data. This outlier data is skewing the representation of the data. (The Distribution charts show data more evenly spread on the other side of the median versus mean (the leftmost group of each chart).)