Robert Rosenbleeth

December 18, 2021

Assignment 1

1. Given the provided data, what are three conclusions we can draw about Kickstarter campaigns? Explain the reasoning behind your answers.

Three conclusions we can draw about the data are:

* The most popular category of campaign is theater.
  + It is the category with the most campaigns.
  + Number of campaigns is assumed to be a measure of popularity.
* The most popular sub-category of campaign is plays.
  + It is the sub-category with the most campaigns.
  + Number of campaigns is assumed to be a measure of popularity.
  + Plays falls in theater, which aligns with the finding above.
* Campaigns are relatively evenly spaced throughout the year, with a slight spike in the summer months (May-August)
  + The date created of campaigns is relatively evenly spaced throughout the year
  + The slowest month for new campaigns is December, which makes sense given the holiday season.

1. What are some limitations of this dataset?

This data set has some limitations that may keep conclusions from being broadly applicable. First, the data set is only about Kickstarter campaigns. This factor means that conclusions may only apply to certain types of campaigns that use Kickstarter, rather than all campaigns. Moreover, the data are limited to when Kickstarter has existed and may not be applicable to the distant past or future. The data also do not account for rewards promised to participants. These rewards may influence whether the campaign is successful and may also have an economic impact. Moreover, the data do not have any measure of the level polish in the campaign. Well-polished campaigns may have a higher level of success regardless of category.

1. What are some other possible tables and/or graphs that we could create, and what additional value would they provide?

It is possible to show percent funded by category and sub-category. This data would highlight the relative level of success across project focus area. In addition, it would be helpful to show state as function of staff\_pick. This graph would show the value of receiving a staff pick. Moreover, it could be useful to graph average donation and percent funded. This data would show if larger donations signify whether a project is more likely to be funded. Moreover, it could shed light on whether the project has broad support or support from just a few large donors.

*Statistical Bonus*

The median appears to be more meaningful. There is a relatively large difference between the mean and median, suggesting that the data is skewed rightward. The large maximum for each dataset further supports this conclusion.

There appears to be more variability with successful campaigns since both the variance and standard deviation are larger. This makes sense given that successful campaigns may attract a few large donations and many small donations, whereas unsuccessful campaigns do not attract as many donations.