Assignment 1 – Crowd Funding analysis

Conclusions

1. The population of people who contribute to crowd funding campaigns are primarily interested in film & video, music, and theater
   1. Of those categories plays are the most popular, by a wide margin
2. There is a spike in successful campaigns in June and July. Cancellations rise steadily over the summer months. Failures hold steady over the course of the year, with a drop off in fall correlating with a drop in successful campaigns and a peak of cancellations.
3. The US is the majority source of the crowdfunding population at about 76%

Limitations

the campaigns are divvied up by geography of what countries the campaigns are located in, but lack information on the individual backers themselves. Places as large and diverse as the US can have a wide spread of folks who can contribute to campaigns and it's important examine what demographics may involve themselves with particular categories and subcategories of campaigns.

In addition the data at hand don't tell us whether or not there was any large variance in the donations themselves - ie in an extreme case a wealthy individual funding $250k of a $500k goal campaign with normal sized donations filling the gap.

Suggestions

An examination of staff\_pick/spotlight campaigns to see how those may have an effect on success rate. Something like a median success rate for staff\_pick/spotlight campaigns matched up against all other types of campaigns

A scatterplot of the sum average\_donation to outcome to see if there’s any correlation between the average donation size and success/failure rate of a given campaign.

A line graph of the number of campaigns of a given category or sub-category over time could show if there are any trending topics that come up over time, or are possibly even seasonal.

Bonus analysis

The median is a more meaningful summary of data as the mean appears to be skewed heavily towards the first quartile for both successful and failed campaigns

Variance is greater in successful campaigns than failed campaigns. This is understandable as a successful campaign with it’s higher median and maximum has more “room” for datapoints to stray from the mean.