DATA ANALYTICS BOOTCAMP

HOMEWORK #1 - EXCEL

MARIO VICENTE MARTINEZ ESTRADA

2020-03-24

# Given the provided data, what are three conclusions we can draw about Kickstarter campaigns?

Different conclusions could be drawn from the provided data, some of them could be:

Most Kickstarter campaigns are in “theater” category and “plays subcategory”.

It seems that the month of the year in which a Kickstarter campaign is launched has an impact in its success rate mostly for “theater”. This category has a higher success rate during May-Jul.

According to the sample if you were a backer, “journalism” is the worst choice to support, technology has a high risk and the category that have better odds as succeeding is music.

Kickstarter with high “amount goals” seems to fail more often than Kickstarter with smaller expectations.

# What are some limitations of this dataset?

Small sample that may not show the reality of the whole data. We have data for 4,000 projects while the whole population is over 300,000, it’s unknown if we have a well distributed sample with all the categories and dates.

Sample is biased towards successful kickstarts, in all the sample we have around 2185 successful kickstarts which its around 50% of the total, however according to the task only 1/3 of Kickstarter have made it through the funding process.

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

We could analyze different scatterplots showing correlation among quantitative variables for example, does the goal amount has any correlation with the pledged amount? Does the quantity of backers have a correlation with the amount goal/pledged amount?

We could measure time how long a Kickstarter campaign lasts according to the data? Does the time period have any correlation with the amount pledged? Or does the time period have any correlation with the success/failure of a Kickstarter?

# Bonus Question 1: Use your data to determine whether the mean or the median summarizes the data more meaningfully

As we could observe in both datasets for “successful” and “failed” campaigns the median and mean are very different from each other’s, this is showing that our data are skewed or in other words very fa away from a “normal distribution”.

Due to the above I would take the median as a more representative measurement.

# Bonus Question 2: Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?

According to the data there is more variability with successful campaigns and less with unsuccessful. This makes partial sense, for example this show us that usually unsuccessful campaigns have a low quantity of backers, however it does not explain totally the successful campaigns.

Some successful campaigns are successful even though they have low quantity of backers.

The main problem here is that the dataset contains a lot of low goal kickstarts that are successful with a low amount that can be financed with a small number of persons, therefore in order to make a better analysis it may be necessary to split the dataset in ranges of goal amount as it was done in the Bonus 1 exercise.

# If you were a Kickstarter Is being more ambitious in your goal will yield more pledged amount?

|  |
| --- |
| Does goal amount influences pledge amount? Or its not related at all in other word, would people support bigger project with big goal amounts than a small project? |
| Graph suggests that there is positive correlation between goaled amount and pledged amount, however other factors may come into play: big Kickstarter usually promise more than smaller ones; therefore they get more people. |
| We could say that from 0 to around 18,000 the positive correlation is strong and maybe the goal amount could explain the pledged amount, therefore if you have a small Kickstarter it maybe be wise to aim for around 16,000-18,000 of a goal |
| for the amounts higher than 18,000 data is inconclusive and we would be "extrapolating" |