Carlos Soriano

Wash U Data Boot Camp – Homework 1

This is an analysis of Kickstarter Project Data from 2009 to 2017.

The following Charts and tables show that there are certain countries in which Kickstarter campaigns are more successful, and, categories and sub-categories are more diverse. These countries are the US, UK, AU, and GB. These are all English-speaking countries and it is no surprise that they make up the bulk of material on a platform largely operating in that language. The trend that is visualized on charts that don’t include data from these top countries is different from the ones that include all the data supporting the conclusion that it makes up a different panorama of how the platform is viewed and used in non-top countries.

Further analysis shows that certain categories are most successful and make up the bulk of the campaigns. These top categories are, in order, Theater, Music, Technology, and, Film & Video. Similarly to the country trends, the data shows a very different profile when the top categories are removed.

Data and charts on ranges of average donations and goal amounts show that the most successful projects have more backers, and fall within specific ranges of goal amounts and average donations, perhaps obviously.

There seems to be a correlation between the bump in popularity that Kickstarter enjoyed with the amount of theater projects. While data shows that Staff Picks leaned towards the Film & Video and Technology categories, the bulk of successful campaigns were theater campaigns. While the bulk of campaigns were theater campaigns, it would be interesting to see where the bulk of the money raised is to be found. If goal and pledge amounts were to be ranged within the main data set, we would be able to chart them against the rest of the data more effectively. The money should tell a good part of the story as well.

Kickstarter projects follow predictable trends. Given more time to reclassify the data according to certain ranges and perform regression analysis on the statistical significance of the data would make for stronger conclusions.

Diagram 1 shows that a plurality of the projects and diversity of the sub-categories are to be found in just a few countries. Notably, the US, GB, and CA, and AU all have the majority and largest diversity of projects. This is most likely due to English language projects occupying most of the landscape.

Diagram 2 shows that theater, music, film & video, and technology account for the majority of projects.

Diagram 2b shows the distribution categories is different with the removal of data from the US, CA, GB, and AU.

Diagram 3 shows that a disproportionately large number of projects are in the “Plays” sub-category.

Notice again how in Diagram 3b the distribution of sub-categories changes markedly with the removal of data from the US, CA, GB, and AU.

Diagram 4 shows the trend of successful vs failed vs cancelled vs live projects over time.

Diagram 4b shows the trend without the largest category, theater.

Diagram 4c shows the trend without the largest and second largest, theater and music.

Diagram 4d shows the trend without 1, 2, and 3, the third being technology.

Diagram 4e finally removes the 4th largest category as well, which is film & video. I wish I could have shown this as a GIF-type animation, but you can see the trend change over time. The failed line is pretty similar but with steeper/less curvy slopes, probably in part due to less data. The cancelled line shows a marked drop after the removal of the 2 most successful categories. The successful line changes the most as is to be expected with the removal of the most popular/effective campaign categories.

Diagram 4f is sort of an inverse of the previous diagrams in the series. It shows only the theater projects. This shape is the rough shape of the general trend in the success chart, and a few of the other ones. It also follows the sharp growth and decline of the general trend. Both these things would lead me to conclude that theater projects played a major part in the rise and fall of Kickstarter.

Diagram 5 shows the percentage of results within the goal range. Success rates are noticeably higher when projects have lower funding goals, with noticeable drops around the $10k and $50k marks. Cancellation percentage increases steadily showing only a slight bias against more expensive projects.

Diagram 5b shows a similar trend as diagram 5. By charting the percentages within each range against the total amount of campaigns, we can see that most of the action is in the $2500-$10000 range. It stays flat until it spikes at the larger end of the spectrum. Notably, failure rates spike at the end more than success. This is probably a sample size limitation as well on the larger end.

Diagram 6 shows average donations on the x-axis stepped up by $10 increments, until $100, then $500 and above, with percentage of results within that range on the y-axis.

Diagram 6b shows the same data as 6 in a different view. It better illustrates the trend of successful campaign having a certain range of average donations.

Diagram 6c is a similar change of point of view as from diagrams 5 to 5b, while maintaining the axes of reference as in diagram 6. The percentages are calculated off the total amount of campaigns instead of within each range.

Diagram 6d is the second point of view of the data in 6c for visualization purposes.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Row Labels** | **Count of state** | **Average of percent funded** | **Sum of backers\_count** | **Average of backers\_count2** |
| canceled | 349 | 0.691805905 | 9530 | 27.30659026 |
| failed | 1530 | 0.093434488 | 27096 | 17.70980392 |
| live | 50 | 0.313363629 | 1801 | 36.02 |
| successful | 2185 | 17.63618704 | 424819 | 194.4251716 |
| **Grand Total** | **4114** | **9.464057341** | **463246** | **112.6023335** |

Diagram 7 is a chart that shows just how many more backers, on average, successful campaigns have.

Diagram 8 shows the difference in result rates between projects designated “Staff Pick” (TRUE) or those that are not (FALSE). Those that are not staff picks follow the general trend of success results over time as in Diagrams 4(x). Those that are staff picks show a markedly different trend. The following series of diagrams will show staff pick data over time with the successive removal of the 4 most successful categories respectively, theater, music, technology, and film & video. The success rates drop with each one, leading to a conclusion that the “Staff Pick” designation does have an impact on success. While the impact is noticeable, the overall trend of non-staff picks is still closer to the general trend.

Diagram 8b

Diagram 8c

Diagram 8d

Diagram 8e

Diagram 8f shows that staff picks had a sharp effect in the technology category over time.

Diagram 8g shows that the effect on the Film & Video category was the most noticeable.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Count of state** | **Column Labels** |  |  |  |  |
| **Row Labels** | **canceled** | **failed** | **live** | **successful** | **Grand Total** |
| **animation** |  | **100** |  |  | **100** |
| FALSE |  | 100 |  |  | 100 |
| **art books** | **20** |  |  |  | **20** |
| FALSE | 20 |  |  |  | 20 |
| **audio** | **24** |  |  |  | **24** |
| FALSE | 24 |  |  |  | 24 |
| **children's books** |  | **40** |  |  | **40** |
| FALSE |  | 40 |  |  | 40 |
| **classical music** |  |  |  | **40** | **40** |
| FALSE |  |  |  | 40 | 40 |
| **documentary** |  |  |  | **180** | **180** |
| FALSE |  |  |  | 60 | 60 |
| TRUE |  |  |  | 120 | 120 |
| **drama** |  | **80** |  |  | **80** |
| FALSE |  | 80 |  |  | 80 |
| **electronic music** |  |  |  | **40** | **40** |
| FALSE |  |  |  | 33 | 33 |
| TRUE |  |  |  | 7 | 7 |
| **faith** |  | **40** | **20** |  | **60** |
| FALSE |  | 40 | 20 |  | 60 |
| **fiction** |  | **40** |  |  | **40** |
| FALSE |  | 40 |  |  | 40 |
| **food trucks** | **20** | **120** |  |  | **140** |
| FALSE | 20 | 120 |  |  | 140 |
| **gadgets** |  | **20** |  |  | **20** |
| FALSE |  | 20 |  |  | 20 |
| **hardware** |  |  |  | **140** | **140** |
| FALSE |  |  |  | 60 | 60 |
| TRUE |  |  |  | 80 | 80 |
| **indie rock** |  | **20** |  | **140** | **160** |
| FALSE |  | 20 |  | 120 | 140 |
| TRUE |  |  |  | 20 | 20 |
| **jazz** |  | **60** |  |  | **60** |
| FALSE |  | 60 |  |  | 60 |
| **makerspaces** |  | **11** |  | **9** | **20** |
| FALSE |  | 5 |  | 9 | 14 |
| TRUE |  | 6 |  |  | 6 |
| **metal** |  |  |  | **20** | **20** |
| FALSE |  |  |  | 17 | 17 |
| TRUE |  |  |  | 3 | 3 |
| **mobile games** |  | **40** |  |  | **40** |
| FALSE |  | 40 |  |  | 40 |
| **musical** | **20** | **60** |  | **60** | **140** |
| FALSE | 20 | 60 |  | 60 | 140 |
| **nature** |  | **20** |  |  | **20** |
| FALSE |  | 20 |  |  | 20 |
| **nonfiction** |  |  |  | **60** | **60** |
| FALSE |  |  |  | 60 | 60 |
| **people** |  | **20** |  |  | **20** |
| FALSE |  | 20 |  |  | 20 |
| **photobooks** |  | **57** |  | **103** | **160** |
| FALSE |  | 11 |  | 64 | 75 |
| TRUE |  | 46 |  | 39 | 85 |
| **places** |  | **20** |  |  | **20** |
| FALSE |  | 20 |  |  | 20 |
| **plays** |  | **353** | **19** | **694** | **1066** |
| FALSE |  | 342 | 19 | 591 | 952 |
| TRUE |  | 11 |  | 103 | 114 |
| **pop** |  |  |  | **40** | **40** |
| FALSE |  |  |  | 40 | 40 |
| **radio & podcasts** |  |  |  | **20** | **20** |
| TRUE |  |  |  | 20 | 20 |
| **restaurants** |  | **20** |  |  | **20** |
| FALSE |  | 20 |  |  | 20 |
| **rock** |  |  |  | **260** | **260** |
| FALSE |  |  |  | 220 | 220 |
| TRUE |  |  |  | 40 | 40 |
| **science fiction** | **40** |  |  |  | **40** |
| FALSE | 40 |  |  |  | 40 |
| **shorts** |  |  |  | **60** | **60** |
| FALSE |  |  |  | 60 | 60 |
| **small batch** |  |  | **6** | **34** | **40** |
| FALSE |  |  | 6 | 20 | 26 |
| TRUE |  |  |  | 14 | 14 |
| **space exploration** | **18** | **2** |  | **40** | **60** |
| FALSE | 14 | 2 |  | 20 | 36 |
| TRUE | 4 |  |  | 20 | 24 |
| **spaces** | **17** | **80** | **5** | **85** | **187** |
| FALSE | 17 | 79 | 4 | 68 | 168 |
| TRUE |  | 1 | 1 | 17 | 19 |
| **tabletop games** |  |  |  | **80** | **80** |
| FALSE |  |  |  | 77 | 77 |
| TRUE |  |  |  | 3 | 3 |
| **television** |  |  |  | **60** | **60** |
| FALSE |  |  |  | 60 | 60 |
| **translations** | **10** | **47** |  |  | **57** |
| FALSE | 10 | 45 |  |  | 55 |
| TRUE |  | 2 |  |  | 2 |
| **video games** |  | **100** |  |  | **100** |
| FALSE |  | 100 |  |  | 100 |
| **wearables** | **60** | **120** |  | **20** | **200** |
| FALSE | 60 | 120 |  | 20 | 200 |
| **web** | **100** | **60** |  |  | **160** |
| FALSE | 100 | 60 |  |  | 160 |
| **world music** | **20** |  |  |  | **20** |
| FALSE | 20 |  |  |  | 20 |
| **Grand Total** | **349** | **1530** | **50** | **2185** | **4114** |

Diagram 8h shows the tally of staff pics by category to further illustrate the trend.