**Arielle Bell (Eagan)**

**NU-CHI-DATA-PT-06-2021-U-C-MW**

**1. Excel Homework**

Due: June 26, 2001

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

We can draw several conclusions about Kickstarter campaigns from the provided data:

* Campaigns’ success varied greatly across categories and subcategories, with many campaigns within each category showing more homogenous success or failure and only a number showing more heterogenous states. Certain campaigns were overwhelmingly successful (ex: the technology/hardware category was 100% successful with all 140 campaigns being marked as “successful”) while others were overwhelming unsuccessful (ex: the technology/gadgets category was 100% failed, with all 20 of the campaigns being marked as “failed”). This suggests that the category and subcategory of campaign in an important variable in its successfulness.
* More campaigns succeed compared to those that failed; however, if canceled campaigns are added to the “unsuccessful” category, then the number of campaigns that were successful (“successful”) versus unsuccessful (“failed” or “canceled”) were similar. Campaigns were most successful when created between January and July.
* Campaigns’ goal amount and success seem to be inversely related; the lowest goals (<$1,000 to $4,999) were most successful (50-100%), mid goals ($5,000 to $44,999) were ~50% successful, and highest goals ($45,000+) were under only 20-50% successful. Campaigns with goals of $40,000+ seemed to fail (40-60%) more often than succeed (20-50%) as the goal grew.

1. What are some limitations of this dataset?

One limitation so the dataset is that the variable “live” was not well defined, as it was listed separately than “successful,” leading the reader to question when and how a campaign it determined to be “successful” versus “live.” Similarly, a campaign being “canceled” is listed separate from “failed” but is not well described. Despite this, I chose to include “canceled” campaigns in the “unsuccessful” grouping, as it was unsuccessful regardless of reason. Another limitation is that some categories had subcategories while others did not. This could bias results connecting to categories, as those without subcategories may be grouped together in error. A further limitation it that campaigns listed as “live” only A final limitation of the dataset was that the backers’ individual contributions were not listed. Though we could only find the min, max, mean, median, variance, and standard deviation of the backers based on the amount pledged, it would have been more helpful to have the individual contribution amounts of each backer for each project.

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

Other possible tables and/or graphs that could be created include:

* A stacked column chart to show campaigns’ status (successful, failed, canceled, live) by category and then by subcategory. Particularly if the colors in the stacked columns were used to highlight the highest likely state within each category and subcategory, this would create an easy visual for the reader to see which categories and subcategories were the most successful, unsuccessful, etc.
* A column chart to show the campaigns by state (successful, failed, canceled, live) by date created to see if there was a possible time-related aspect relating to campaigns’ success. Though we used a line graph to see this, the column graph could be an alternative visual.

**Bonus:**

\* Use your data to determine whether the mean or the median summarizes the data more meaningfully.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Row Labels** | **Count of name** | **Sum of backers\_count** | **Average of backers\_count2** | **Median of backers\_count2** |
| successful | 2,185 | 424,819 | 194 | 609 |
| live | 50 | 1,801 | 36 | 30 |
| failed | 1,530 | 27,096 | 18 | 179 |
| canceled | 349 | 9,530 | 27 | 70 |

The table above shows the number of campaigns by state, the sum of backers within each of these campaign states, and both the average and the mean of these backers. All are useful data pointes. However, based on this table, I would say that the median summarizes the data more meaningfully, as it shows us that several successful campaigns were able to reach success with few numbers of backers (thus pulling the average down from the median number of backers). It may be helpful to use this information to learn more about the backers of these specific campaigns that were backed by fewer people to understand the profile of the backer and thus target this backer audience in future fundraising and marketing.

\* Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Row Labels** | **Count of name** | **Sum of backers\_count** | **Min of backers\_count2** | **Max of backers\_count2** | **Var of backers\_count2** | **StdDev of backers\_count2** |
| successful | 2185 | 424,819 | 1 | 26,457 | 713,167.38 | 844.49 |
| live | 50 | 1,801 | 0 | 884 | 15,648.67 | 125.09 |
| failed | 1530 | 27,096 | 0 | 1,293 | 3,775.69 | 61.45 |
| canceled | 349 | 9,530 | 0 | 1,501 | 11,648.45 | 107.93 |

By the data above, there is much more variability in the support of backers of successful campaigns versus those that were unsuccessful (“failed” and “canceled”). This makes sense to me because successful campaigns have a higher max number of backers, meaning that there is likely a higher probability of having some high-pledging backers among the large sample of backers. Even if many of the 26,457 backers donated small amounts of money, the variety and standard deviation lead us to understand that there were also several very high pledges that were given, which could increase the pledged total up to or beyond the goal amount at a faster rate. Campaigns that failed or were canceled had lower numbers of backers (1,293 and 1,501 respectively) and had far less variability and lower standard deviations. This suggests that their backers were more homogenous and thus there was less chance for individual backers to give large amounts, thus keeping both the mean and median relatively low.