**Executive Summary**

The Kickstarter dataset covers the time range 2009 to 2017 inclusive. Therefore, it is important to identify the limitation of the analysis, which is that it does not consider the last years and therefore acquiring that data would enable to provide more meaningful recommendations for your Kickstarter campaign. The conclusions of my analysis based on this time range are (i) do not begin a campaign between May to September and November to December. These periods have shown an increasing of failed campaigns. (ii) Aim for goals between $0-$20000 if possible, this have the highest percentage of goals achieved. Lastly, (iii) campaigns in the music category are most popular with a high percentage of success, particularly the ‘Rock’ subcategory. “Plays and Documentary” are other high Success rate subcategories and so is Hardware. All things considered, Food category has been the worst performing category and therefore it would be advisable to stay away from it on your first campaigns.

Summary statistics on for backer counts on both successful and failed campaigns reveal that the median is a more appropriate metric than the mean as there is a major discrepancy between both values that demonstrates that the data is skewed. The variance for successful campaigns is greater than for failed campaigns, which is expected. It is more likely to perceive greater changes on successful campaigns as they will attract more backers the better they are performing but will become less attractive as the campaign underperformed, hence failed campaigns behave more uniformly.

**Introduction**

On this report I am analysing the dataset for Kickstarter campaigns for the date range 2009 to 2017. Kickstarter aims to bring ideas into reality through a crowd founding model with a diverse range of categories and sub-categories to choose from.

On the following report I am going to provide three data-driven recommendations based on my analysis, provide summary statistics for the backers count, following I will state the limitations I discovered and provide which additional graphs could have been utilised to enhance the understanding of the data. I will conclude with re-stating the recommendations as conclusions of the analysis.

**Analysis and Recommendations**

Visualizing how campaigns states - ‘succesful’, ‘failed’ and ‘cancelled’ - fluctuate over a timeline(graph 1), allows us to quickly identify which are the months that are more likely to provide a successful outcome.

Graph 1

We can therefore conclude that successful campaigns decrease between May to October and then every more so drastically between November and December. It would be advisable, without zooming into categories at this point, to aim to run kickstart campaigns outside of these time ranges.

Moving on, it would be relevant to understand which “goals” are more likely to be met when you are strategizing your Kickstarter campaign. Graph 2 depicts the percentages on the “y” axis against goal ranges for each tick on the “x” axis.

Graph 2

A limitation of this graph is that the ranges act as categories, but this has been done for data manipulation purposes to provide an enhanced visual representation. Summary statistics over the original data set would be a more comprehensive approach and would hold more insightful answers. However, this graph allows to quickly perceive an overview for higher percentages of goals met between 0 and 20000 and then spiking slight at around 45000.

Thirdly and lastly, drilling deeper into categories and sub-categories, we can deduct from graph 3 that “Music” holds the highest amount of successful campaigns versus failed campaigns. Zooming in deeper into this category (graph 4), “Rock” seems to be the most successful sub-category within “Music”, therefore keeping this information in mind might enable higher propabilities for success

Graph 3

Graph 4

**Summary Statistics for Backers Count**

|  |  |  |
| --- | --- | --- |
|  | **Successful** | **Failed** |
| **Mean Number Backers** | 194.43 | 17.71 |
| **Median Number Backers** | 62.00 | 4.00 |
| **Minimum Number Backers** | 1.00 | 0.00 |
| **Maximum Number Backers** | 26457.00 | 1293.00 |
| **Variance Number Backers** | 712840.99 | 3773.22 |
| **Standard deviation Number Backers** | 844.30 | 61.43 |

For both columns it seems like the median is a more appropriate metric as they are values far lower than the mean. The median state the middle values therefore having a considerable difference between median and mean, gives us the insight that the data is skewed.

We ca also see that there is more variability with successful campaigns which makes sense because the more momentum and successful a campaign becomes, the more backers it will attract. Conversely, if the campaign is not evolving steadily towards its goal and looking like it will surpass it, then most likely it will not be an attractive campaign. This shows a more abrupt behaviour of change for successful campaigns whereas it is less likely than a campaign going bad will turn for the better, it will most likely fail and therefore there is less variance amongst failed campaigns.

**Limitations**

Identifying limitations of the analysis is a key step to provide a well-rounded understanding of how useful the information presented is. Therefore, it is paramount to understand that the last three years of data are missing, hence the following conclusions cannot be generalised to the present time. It would be highly relevant to see what has happened over 2020 with the global pandemic and how that affected crowd funding projects.

Another limitation on the dataset is that there are many currencies. Having had one uniform currency would be more valuable because by applying currency filters then comprehensive data is lost and therefore it diminishes the value derived from conclusions.

**Additional Graphs**

Additional graphs that could enhance this analysis are a box plot for the backs count for each state a scatterplot for each category illustrating goals vs backer count to see a understand the relationship between these continuous variables.

**Conclusion**

In conclusion, there is more analysis to be done on this data set and the recommendations I have provided are to be taken lightly, as a point reference, until the last three years of data are added onto it.