|  |
| --- |
| **Refer to ‘Overview data’ excel sheet for analysis supporting the following analysis:**  In this exercise we have analysed a total of 4114 projects that have been launched in Kickstarter between the years 2009 and 2017 and have raised over 46 million dollars. |
| These data **represent around 1% of all projects** launched in Kickstarter according to the number provided in the instructions and around **2% of all the total money raised**, thus represent a **very small sample** of the total data and any statistical analysis/trends identified should acknowledge the limitations of the data sample size. |
| Analysing overall trends, we see that the majority of projects within this period have been successful (53%) followed by over 37% of failed projects. The percentage of canceled (8%) or live (1%) constitute a very small part of the dataset, thus most analysis will focus and compare successful vs failed in order to simplify representation and analysis. |
| We first calculated the percent funded (which is the amount pledged over the amount targeted). This revealed that what defines a successful project is having reached 100% of the targeted amount (all successful projects have percent funded in the green/blue zone). While all failed projects had a percent funded below 100% (on average 9% with 89% being the maximum). |
| In some instances, the canceled or live projects reached 100% funded, but on average had a percent funded of 69% and 31% respectively, below 100%. |
| In conclusion there is a trend in which **successful projects reached their money target**, while failed projects didn't. |
| We also calculated the average donation per backer, however there wasn't a strong relationship between this and the success of a project. Moreover, the average donation between failed or successful projects were not too different (1530 and 2185 respectively). This makes sense and probably speaks to the diversity of projects and goals that required different amounts of donation to reach the target.  **Refer to ‘Category stats’, ‘Category stats\_graph’ and ‘Subcategory stats\_graph’ excel sheet for analysis supporting the following analysis:**   |  | | --- | | We then analysed the data by countries and categories. | | The majority of the projects come from the US (3038), followed by GB (604) and CA (146). | | Similarly, the majority of the projects belong to the theatre (1393), music (700), technology (600) and film&video (520). | | The majority of successful, but also failed projects, are from the US, BG and CA. This probably just refers to the fact that these countries had the highest project proposals. | | Regarding categories, theatre and music have the most success but **music seems to have very little failures in proportion, so it seems like a good category**. | | This is readily visible in the following graph that shows a breakdown of the state of projects within the main countries and categories. | | When analysing subcategories:   * within theatre: plays have the most successful and failed projects; * within music: indie rock and rock have the most successful projects **(remarkably rock hasn’t had failed projects)**; * within technology: hardware has the most success and **wearables the most failures**; * and within film: documentaries have the most success, while animation and drama have the most failures.   **Refer to ‘Outcomes based on launched date’ excel sheet for analysis supporting the following analysis:**   |  | | --- | | When analysing the outcomes per year in all categories, it's apparent that between the years 2009-2013 there was a steady increase of project proposals and the chances of getting it successful were high. | | Between the years 2013-16 there was a sharp increase in project proposals and the number of failed projects also spiked up, indicating that **it's harder to get projects funded in recent years**. | | The year 2017 is not complete (only months Jan-March), thus we won't account for it in the rest of the analysis. | | When looking at the most popular categories, it seems like **music, followed by theatre, are the most popular sectors that have maintained a higher rate of success over failures throughout the years**. | | While technology and film have had a close number of failures to success in recent years. | | When looking at all the years and categories combined throughout Jan-Dec, it seems like there is an overall decline of successful projects as the year moves forward, so it might be recommended to launch the project early on. | | |

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

Based on all the above, we can conclude that:

* Having a **project within the music category**, more specifically within indie rock or **rock subcategory**, has a high chance of being successful.
* Do **not have a project in wearables technology**, since the failure rate is higher than successful projects.
* It was **easier to get a successful project when launching between 2009-13** but still **music and theatre have maintained a higher success rate in recent years**, this doesn’t happen within technology or film, so it’s not recommended to launch a project in the latter sectors.

1. **What are some limitations of this dataset?**

As mentioned at the start, the data sample represents only 1 and 2% of the amount of projects and total money raised within Kickstarted, thus its **size is very small and might not be representative**.

The data might also be **skewed towards particular countries** which used it more (US, GB, CA) during the period analysed (2009-2017), however in recent years it is possible that other countries have used the platform more and thus the data might not be representative for those.

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

We could plot the **outcomes per year according to countries** to see whether in recent years there has been an increase of the “not so popular” countries.

We could also focus on a specific category of interest (despite this not being one of the major four that I plotted in the ‘outcomes based on launched date’ sheet), and plot the outcomes over time to see whether it has had recent high success rate.

**BONUS ANALYSIS**

|  |
| --- |
| From this plot, it can be observed that there is an inverse correlation between goal amounts and success. |
| Where the projects with the lowest target amounts ($1000 to $15000) reached the highest success and the least failures or cancelations. Whereas the highest targets (over $45000), reach the lowest success, and highest number of failures and cancellations. |
| **BONUS STATISTICAL**   1. **Use your data to determine whether the mean or the median summarizes the data more meaningfully.**   The median is a better representation of the data since the sample is not normally distributed. This means that the dataset contains many outliers and thus an average calculation can't represent well the data set when it’s so spread.   1. **Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?**   There is more variance within successful campaigns. This makes sense since the range of number of backers is also broader in successful campaigns (1-26457) compared to failed campaigns (1-1293). |