1. **Kickstart My Chart - Conclusions**

Based on the results of the analysis conducted on a database of 4,000 Kickstarter projects, the following conclusions can be drawn.

1. Given the provided data, the following three main conclusions can be drawn about the Kickstarter campaigns.

It is important to note that the results described below where also drawn from an evaluation of the percentage of successful and unsuccessful projects in each category and subcategory. While these percentages were extrapolated from the data provided, they were not part of the analysis required as part of this homework.

1. Results regarding categories, extrapolated from PivotTable 1 – Category Statistics, in the attached spreadsheet:
2. Accounting for all countries, the category Theater has the highest number of overall projects: 1393, or 34% of the total number of projects. This category has also the highest number of successful projects (839) and failed projects (493), corresponding to 60% and 35% of the total Theater projects.
3. Music has the second highest number of overall projects (700, or 17% of the total number of projects) and successful projects (540, or 77% of the total music projects). The number of failed projects is 120, or 17% of the total music projects. This percentage represents the lowest percentage of failed projects among the various categories.
4. Film & Video projects are fewer than music projects (total of 520 versus 700, or 13% versus 17% of the total projects). However, once normalized, the percentage of successful and failed projects are comparable. The number of successful and failed film & video projects is, respectively, 58% and 35% of the total number of projects in this category.
5. For comparison purposes, the remaining categories represent between 1 and 15% of the total number of projects.
6. US and GB have the highest percentage of overall projects (700, or 17% of the total number of projects) and successful projects (30% for the US and 59% for GB of the total number of projects). For the US, music, film & video and theater score the higher percentage of successful projects in their respective categories (in that order). For GB instead, the order is for music, theater and games.
7. Results regarding categories, extrapolated from PivotTable 2 – Sub-category Statistics, in the attached spreadsheet:
8. Among their respective sub-categories, plays has by far the highest number of overall projects: 1066, or 26% of the total projects. Further, this category has the highest number of successful projects (694) and failed projects (353), corresponding to 65% and 33% of the total play projects. Though a far distant second, rock has the second highest number of overall projects (260), all successful.
9. Of all the sub-categories, 12 have 100% successful projects falling within the following parent categories: film & video, games, journalism, music, publishing and technology. Each of these subcategories represents between 1 and 6% of the total number of projects.
10. Results regarding project outcome, extrapolated from PivotTable 3 – Project Outcome Based on Launch Date, in the attached spreadsheet:
11. The highest percentage of overall successful projects was launched in May (234, or 61% of the overall projects per month). Following May, the overall number and percentage of successful projects decreased steadily until September (147 successful projects, or 49% of the total projects in that month). A second, lower peak was reached between October and November (183 successful projects per month, 44% of the total projects in those months).
12. The lowest percentage of failed projects was launched between February and April (30-33% of the total projects per month) and in November (34%). The highest percentage per month of failed projects were launched in December (47%), followed by the projects launched between June and October (38 to 42% of the total number of projects per month).
13. Limitations of the dataset
14. The database evaluated herein is based on 4,000 past projects. It is unknow how this database was selected. Importantly, it corresponds to only 1.3% of the overall number of projects launched on Kickstarter.
15. The dataset was compiled between May 2009 and March 2017. Historical trends up to 2020 may include relevant and up to date information that is not reported in the current dataset.
16. The percentage of successful projects reported is 53% of the total number of projects. The background note to this assignment indicates that only a third of the 300,000 projects launched on Kickstarter made it through the funding process a positive outcome. This discrepancy suggests that the dataset may be too small or selected based on parameters that do not provide results consistent with the overall trends of the complete set of projects launched on Kickstarter.
17. Additional possible tables and/or graphs that could be created.
18. Additional tables/graphs not directly evaluated as part of this analysis include the following:
    1. Analysis of the percentage of successful and failed projects for each category and subcategory and associated graphs. This allows for a normalization of the data, rather than looking at the total number of successful projects in each category and sub-category. This evaluation was conducted in part A above.
    2. Statistical analysis, as those covered in the bonus part of this homework, including the following: number of backers of successful and unsuccessful campaigns; mean, median, minimum, maximum, variance and standard deviation of number of backers. This evaluation is meaningful to better analyze successful campaigns and metrics that are successful in this evaluation. A statistical evaluation of the timing at which the campaigns are launched would also provide insight, as suggested by the evaluation in part A, possibly normalized to the amount of money donated in each month for each.
    3. Additional statistical analyses that would be helpful with this dataset and the large variability of the maximum data from the median relate to qualitative and quantitative evaluations using IQR interquantile ranges and first/third quartile analyses.

**01. Kickstart My Chart Bonus - Statistical Analysis Conclusions**

1. The statistical analysis shows the presence of outliers in the dataset. In particular, the max number of backers of successful and failed campaigns (26457 and 1293, respectively) far exceeds the average, or mean, number of backers.

Therefore, basing the evaluation on “arithmetic” average of the dataset can be misleading. The median identifies the middle value of the data set and can therefore be a more meaningful measurement of the center of the data set, with respect to the number of backers of successful and unsuccessful campaigns.

1. The standard deviation is more meaningful than the variance as it shows how spread out the dataset is from the mean. It also uses the same units of measurement as the mean, while the variance is unit-less. Given the outliers in the dataset, the variance is less meaningful, as it describes how far the values in the data set are from the mean. And as discussed, the median is more indicative of the mean for this dataset.

The standard deviation indicates that there is greater variability in the successful campaigns than in those that failed. This is expected based on other metrics, such as the higher median and maximum number of backers of successful campaigns compared to failed ones, as well as the largest number of backers of successful campaigns.