**KICKSTARTER CAMPAIGN DATA ANALYSIS REPORT**

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

**COUNT PER CATEGORY (Conclusion 1A)**

According to the data visualized in the chart, *Count Per Category* (in the Excel sheet with the same title), three categories show a greater quantity of successes than failures, based on absolute counts—these include *music* (which shows the highest percentage of success)*, theater,* and *film & video*.

In contrast, five categories show the opposite trend and show a greater quantity of failures than success, based on absolute counts. These include *food* (which shows the highest ratio of failures to successes), *photography*, *games*, *publishing*, and *technology*. The remaining category, *journalism*, could not be analyzed in such a fashion since every count instance had been canceled.

Note that this analysis is based upon the global complete count instance per category. The United States followed this same trend when it was selected/filtered using the Pivot Table. Additional insight for a given location can be elicited by choosing the country of interest and subsequent chart analysis.

**STATE COUNT PER SUBCATEGORY (Conclusion 1B)**

Additional insight (global, not country specific) into the categories listed above were analyzed using the visualization, *Count Per Subcategory*:

* When the category of *music* was analyzed, there existed five subcategories that showed a 100% success rate--*rock*, *pop*, *metal*, *electronic music*, and *classical music*. *Indie rock* showed a success rate of approximately 90%, while the two subcategories *jazz* and *faith* showed a 100% failure rate.
* When the category of *theater* was analyzed, the *play* subcategory showed a 66% success rate, while the other two subcategories, *spaces* and *musicals*, both showed a success rate of approximately 50%.
* When the category of *film & video* was analyzed, there were two subcategories that showed a 100% success rate: *television* and *documentary*. This was contrary to the 100 % failure rate of the two subcategories, *drama* and *animation*.
* When the category of *food* was analyzed, the subcategory of *small batch* showed a success rate of 100% while the other two subcategories showed a 100% failure rate (*restaurants* and *food trucks*).
* When the category of *photography* was analyzed, the subcategories of *places*, *people* and *nature* each showed a failure rate of 100%, while the subcategory *photobooks* showed a success rate of only 62%.
* When the category of *games* was analyzed, the subcategories of *video games* and *mobile games* showed a failure rate of 100% while the subcategory of *tabletop games* had a success rate of 100%.
* When the category of publishing was analyzed, the subcategories of *translations*, *fiction* and *children’s books* all had a 100% failure rate (*art books* projects were canceled), while the subcategories of *radio & podcasts* and *nonfiction* had a 100% success rate.
* When technology was analyzed, the *web* subcategory consisted of a 100% failure and/or cancellation rate. *Wearables* showed an approximate 90% failure and/or cancellation rate, while *makerspaces* and *space exploration* showed an approximate 50% and 66% success rate, respectively. *Hardware* had a 100% success rate.

It can be seen from this chart that regardless of the ratio of successes to failures indicated in the chart discussed above, *Count Per Category*, further drilling down into the data reveals subcategories that have shown a 100% success rate or a 100% failure rate that can guide future decisions on potential projects.

**STATE COUNT PER LAUNCH DATE (Conclusion 2)**

Two charts were created, with the first chart showing state count per monthly launch date, and the second showing state count per quarterly launch date. According to the data visualized in the chart, *State Count Per Launch Date (Quarterly)*, one can see a trend of an increasing success to failure ratio over the first quarter with a maximum during the second quarter (this occurs during the month of May according to the data visualized in the chart, *State Count Per Launch Date (Monthly)*. This ratio increasingly decreases over the third and fourth quarters, showing a ratio of success to failure at a minimum in December with an approximate value of 1:1. From a broader perspective, when all categories and years are taken into account, it appears that launch dates during the first half of a given year resulted in an increase in the ratio of successes to failures, with this ratio decreasing over the second half of the year.

**PERCENT STATE PER GOAL (Conclusion 3)**

According to the data visualized in the chart, *Percent State Per Goal*, a general trend is revealed that shows as the goal amount increases from $0 to over $50,000, the percent of successful projects decreases from approximately 70% to 20%, while the percent failing projects increases from approximately 25% to 60%.

**2) What are some limitations of this dataset?**

* The data set that was provided contained only 4,000 projects (there are approximately more than 300,000 projects launched on Kickstarter).
* The data set is not an accurate representative sample of the Kickstarter database—only 33% of the projects are successful but the data set that was given had a project success rate of 53%.
* The data set included canceled and live projects—because these are neither ‘failures’ nor ‘successes,’ they did not contribute to the data set in a way that allowed for an increased discernment of hidden trends.

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

* Create a Pivot Table & Chart that shows *Launch Date Per Category* or *Launch Date Per Subcategory*
* Create a Pivot Table & Chart that shows *Goal Per Category* or *Goal Per Subcategory*

**BONUS STATISTICAL ANALYSIS**

See Summary Statistics Table in Excel Sheet titled *Summary Statistics*.

1. **Which measure of central tendency best summarizes the data?**

In the case of the successful projects, the mean is 194.42 while the median is 62. Because the mean is greater than the median, the data set is positively skewed, which is backed up the fact that the value for this property in the summary statistics table is 21.4. If this followed a normal distribution, the mean would be the best indicator of central tendency and skew value would fall between +1 and -1. Because of this, the median should be used as the best measure of central tendency.

In the case of the failed projects, the mean is 17.71 while the median is 4. Because the mean is greater than the median, the data set is positively skewed, which is backed up by the fact that that the value for this property in the summary statistics table is 11.8. If this followed a normal distribution, the mean would be the best indicator of central tendency and skew value would fall between +1 and -1. Because of this, the median should be used as the best measure of central tendency.

1. **Which category of campaigns displays greater variability?**

There is more variability in the data set for the successful campaigns than for the failed campaigns. First, the range for the successful campaigns is 26456 compared to the range of 1293 for the failed campaigns (range is approximately 26 times larger for the former campaign). Second, the variance (and thus the standard deviation) is larger for the successful campaign data as compared to the failed campaign data—the standard deviation is approximately 14 times larger for the former campaign. This makes sense because campaign success correlates with the number of backers—the failed campaign dataset would have many values that were low, with a number of values of ‘0’ (the mode of the failed campaign data set). Successful campaigns would need at least one or more backers in order to achieve the overall goal for a given project.