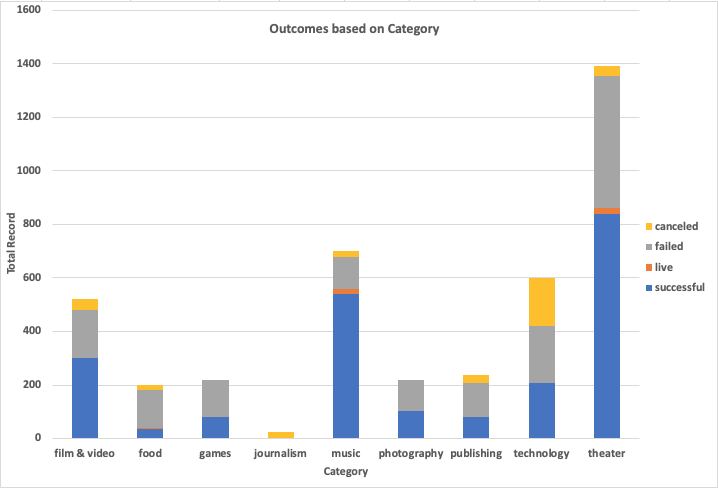
**Report**

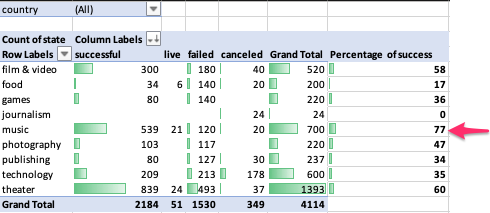
1. **Given the provided data, what are three conclusions we can draw about Kickstarter campaign**

**Answer:**

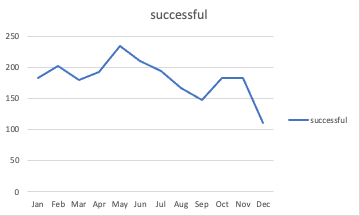
**Conclusion 1.** The Kickstarter campaign has shown a higher frequency in the theater category compared to other groups. Not only theater but entertainment in general, is the best place to get more people’s attention, in the data next to theater music and film & video is most prevalent. These may be due to theater is the best-suited place to get a lot of people at once and conduct a campaign.



**Conclusion 2**. In general, the Kickstarter campaign has greater success than failed, canceled, or live. However, the sample size is yet limited to generalize. When we come to successful campaign, music is the top. The percentage of successful music campaigns from the total is higher than the other groups. Music has a 77 % success rate followed by theater, which has a 60% success rate. Music also has a lower failed and canceled value.

****

**Conclusion 3.** The time serious analysis of Kickstarter campaign has shown successful result on May, however if we filter theater from the category February is the peak. This may due to summer is on the way people start to go out for entertainment or theater. In the other category the binging’s of the year are the peak it may be due to a new year resolution started to be implemented.

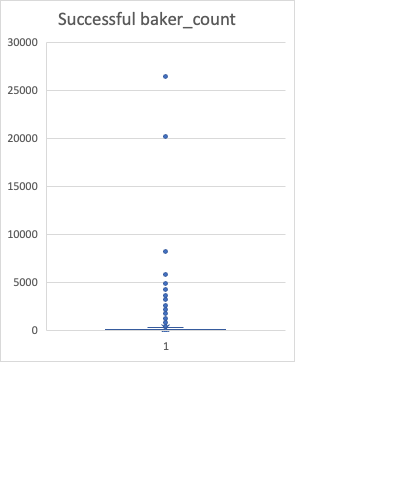
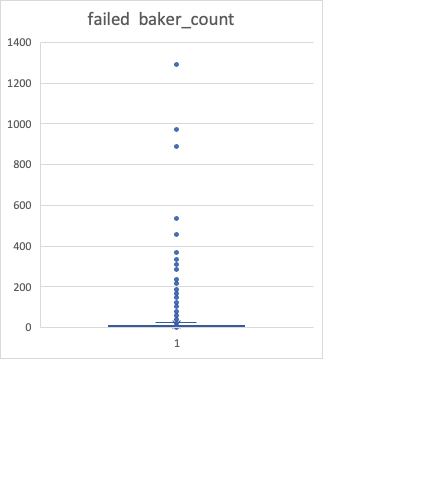


**Excluding theater**



**Conclusion 4.** The mean, Z-score, and the outlier result on the successful vs failed backers count showed that there is more variability in a successful group than the unsuccessful group. However, for both groups the data is not normally distributed, it is positively skewed (left-skewed). The probability distribution graph also looks identical to the data graph.

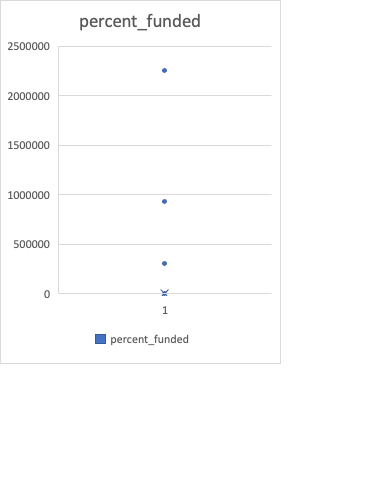
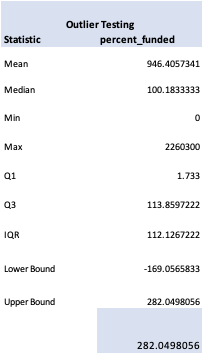
The median, the quartile, and outlier tests also showed an upper bound outlier for both successful and failed groups. There is higher variability in a successful group than the failed group. Based on the given data the median summarizes the data more meaningfully.

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

**Answer:**

* This dataset only includes a third of the population from more than 300,000 projects. To have a better insight, we need more samples.
* The data sourced from multiple countries, and their currency, in this case, I don't get how the data normalized those currencies into one and compared the outcome.
* The data is not normally distributed we need more samples to generalize about the data set.
* The dataset has outliers. For example, the outlier testing of the percent funded column has upper bound outliers. In addition to that backers count for both successful and failed campaign outcome showed upper-bound outlier.

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

* Make outlier testing
* Creating a table and graph by excluding outliers
* Creating a graph that shows the campaign result quarterly, and yearly not only by month
* Conducting some statistical analysis, creating a table and graph that shows the analysis difference across outcomes, categories and time
* Creating a graph and table by making a comparison across different countries on the campaign being successful or failed.
* Creating a comparison table or graph which shows average donation across campaign outcome