KICKSTARTER ANALYSIS

# Objective

To determine market trends from past Kickstarter projects that could give a project a better chance for successful outcome.

# Assumptions

Below are the assumptions that have been applied to this analysis:

* IQR method, with percent-funded, have been used to identified outliers (amount > 250%). Outliers have been excluded when determining all the conclusions in this analysis.
* The dataset is assumed to be a sample data set (4,114) of all the campaigns in Kickstarter (more than 300,000).
* ‘Live’ state was introduced from January 2017 while ‘successful’ state declined from this period. Therefore, it is assumed that there was a change of state description and both ‘live’ and ‘successful’ should be considered as ‘successful’.

# Initial Conclusion

Base on the initial analysis, below are the three conclusions that can be drawn about Kickstarter campaigns:

1. Theatre category is most popular and has highest level of successful outcomes (refer to ‘Pivot Count per Category’)
2. The sub-category under theatre campaigns that has the highest successful outcomes is ‘plays’ (refer to Pivot Count per Sub Category’)
3. The number of campaigns seem to increase dramatically from April 2014. However, due to a more competitive market, the number of successful campaigns were overtaken or closely matched by failed campaigns, and the number of cancelled campaigns also increased. Compare to the campaigns prior to April 2014 when the market was assumed to be less competitive, the rate of successful campaigns is constantly higher than failed campaigns. (refer to ‘Pivot Date Created vs State’)

# Limitations

The main limitation of this dataset is that there is no information on whether the dataset is a total population for given period or only sample selection. If it is a sample dataset, more information on sampling method is required in order to determine how confident and accurate the conclusions are. Information such as:

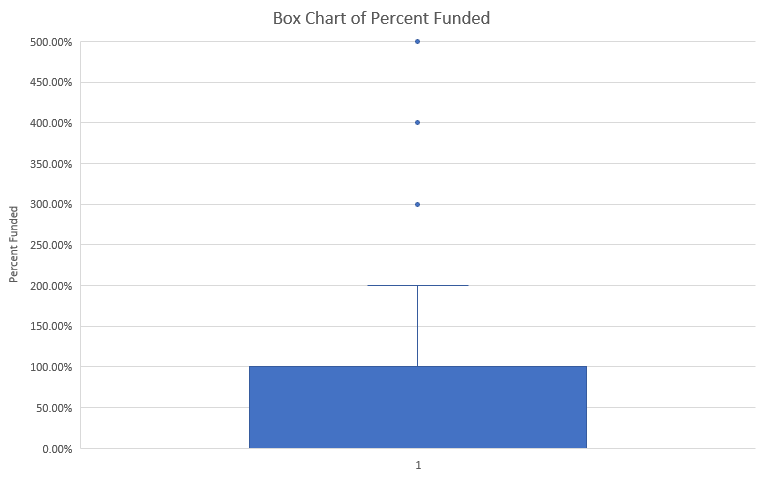
* Were the samples randomly selected?
* Was there cluster sampling e.g. by category or by country?

The other limitation of the data is around the type of pledge. In Kickstarter, the project can have pledging by donation or by pledging as a purchase for final project product or some kind of rewards. Trend in product price can be analysed to determine what is a price range most backers would likely to contribute.

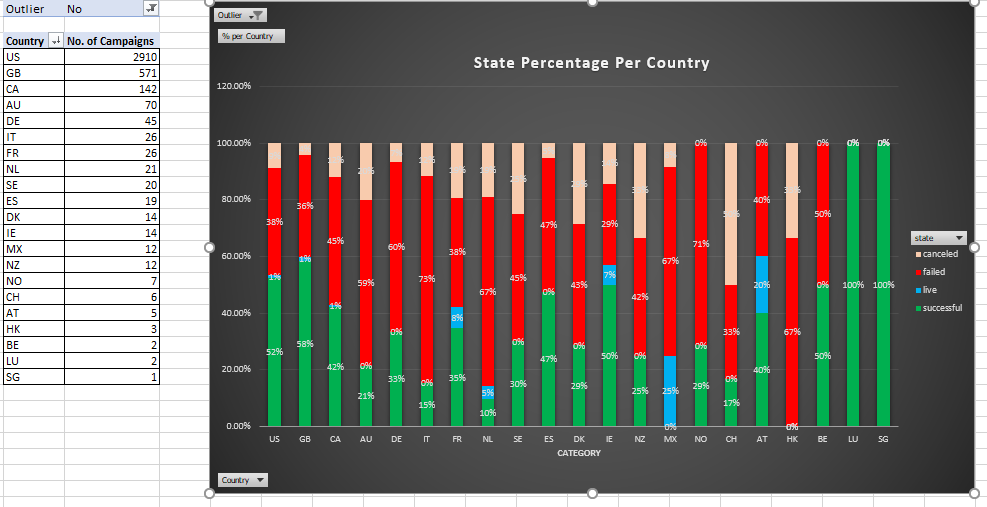
# Other Charts and Analysis

Below are other charts and tables that can be used for more insights.

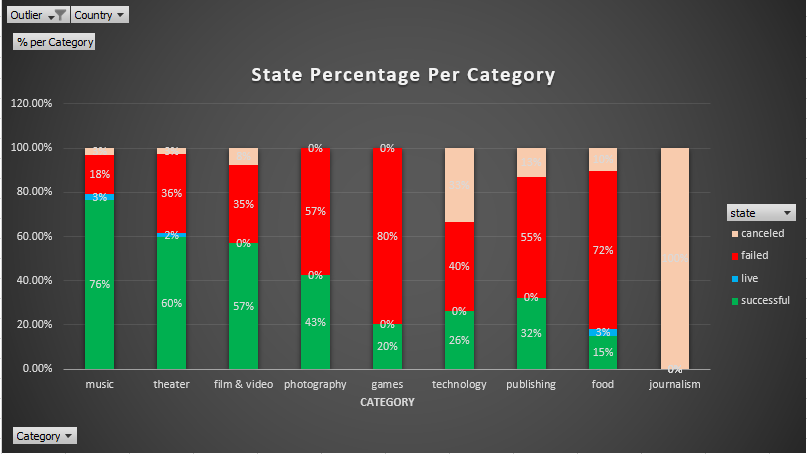
1. Box plot to identify if there are any outliers (refer to “Outliers’)



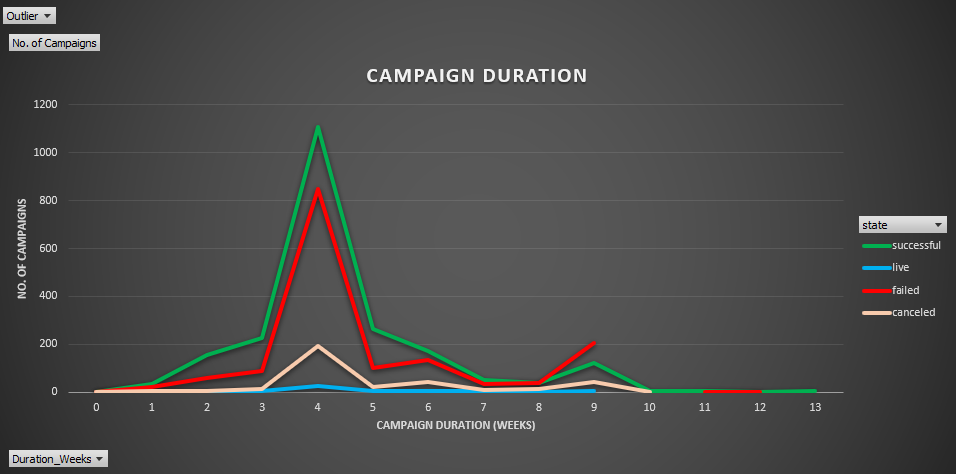
1. State percentage of total per country stacked bar chart (refer to ‘Pct per Country (Extra)’). Also need to consider the number of campaigns located in each country to decide if the percentage is worth considered i.e. 100% of country where there is only 1 campaign should not be considered. From this chart, the top three countries with highest percentage of successful projects are GB, US and CA.



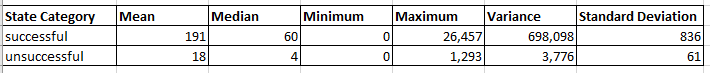
1. State percentage of total per category stacked bar chart provides a different conclusion as to which Category have better chance of success (refer to ‘Pct per Category (Extra)’). When looking at the number of campaigns by category chart (‘Pivot Count per Category’), ‘theater’ seems to be the top category to have better chance of success. However, if we compare within each category, ‘music’ has a better chance of success.



1. The other chart worth considering is the line chart of campaign duration (refer to ‘Campaign Duration (Extra)’). The conclusion of this chart is that projects with duration of 4 weeks have better chance of being successful.



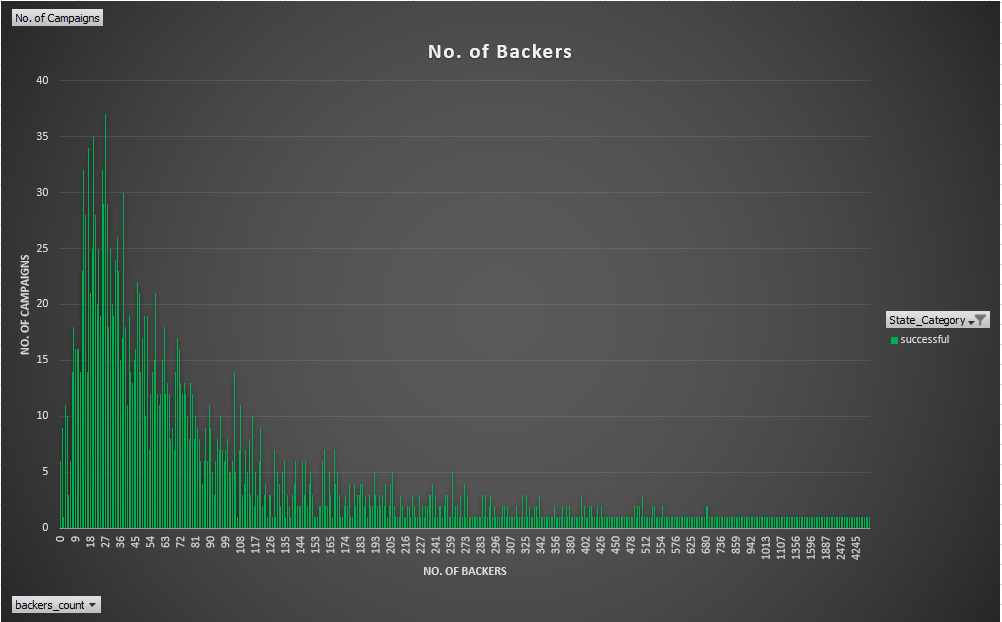
# Statistical Analysis

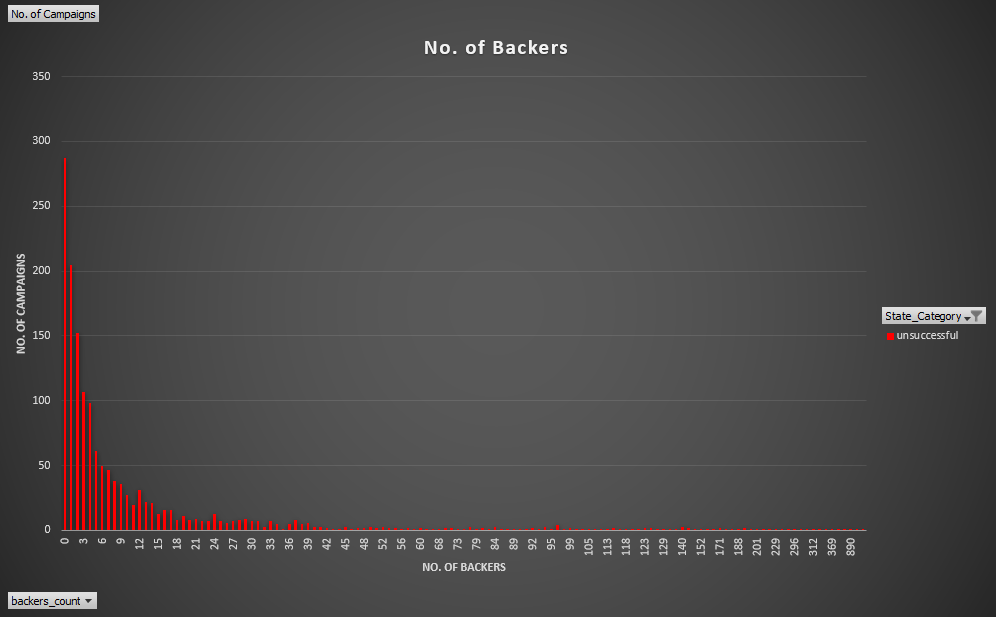


*Refer to ‘Bonus Statistical Analysis’*

**Mean or Median?**

Given that the range (minimum/maximum) of values is quite large, the mean calculation can be skewed by high values and potentially dominated by outliers. Therefore, it is more meaningful to use the median, the most central value(s). This is supported by the chart below where the data is skewed by high values.





**Variability**

From the statistical analysis above, ‘successful’ campaigns have more variability. This make sense because in order for a campaign to be successful, high number of backers is required, and some campaigns may require lot more backers depending on how the pledge values are setup. If the amount is small, then the campaign would need more backers to achieve set goal amount.

Compare to ‘unsuccessful’ campaigns, the variability is less because the number of backers is expected to be low. However, it is expected that high number of backers do exist for some unsuccessful campaigns because ‘cancelled’ campaigns are classified as unsuccessful.