7/22/2021

**Data Analysis Report on Kick Starter Projects**

By

Nimmy Jose

Contents

[Scope: 1](#_Toc77847504)

[Analysis Performed: 1](#_Toc77847505)

[Conclusion: 1](#_Toc77847506)

[Limitation of Data Set: 2](#_Toc77847507)

[Possible Additional Tables/Graphs: 2](#_Toc77847508)

[Summary Statistics Analysis 2](#_Toc77847509)

# Scope:

Perform data analysis on given dataset of 4000 Kickstarter projects and its performance from the past to provide data insights on various trends and help in decision making.

# Analysis Performed:

Below data analysis was performed on the given data using Excel:

1. Added few derived metrics like Percent Funded, Avg Donation and formatted few date columns for easy representation.
2. Applied Conditional Formatting on Outcome and Percent funded to visually highlight the key categories.
3. Plotted Charts on Project Outcome against Category , Sub-Category and Month on which project started
4. Also plotted the projects outcome by bucketing the projects based on the goal size; it gives the user a view of performance among different sized projects.

# Conclusion:

Based on the above analysis below are my findings:

* “Theater “ category attracted maximum number of projects but considering the success percentage “Music” category had better outcome. Data also shows that least successful categories were those projects started under Food and Journalism
* Trends shows and inverse relation between success outcome and goal size of the projects. Shows projects above 45K goal have less than 30% success outcome
* Analyzing the Sub categories; ‘Hardware’ , ’Rock’ and ‘Documentary’ are having 100% success rate.

# Limitation of Data Set:

* The given data set is relatively a small sample set of 4000 projects out of which 80% of data is based on projects from US/UK. Equal sample data set from other countries could help confirming the trends. It could eliminate any regional or cultural influence the current data set may have on selecting the successful project categories.
* Project goal size is captured in different currencies; project size are not on same scale. This will impact the trends.
* Data does not capture any details about success/failure reason. Insights in to such details could help to get additional insights.

# Possible Additional Tables/Graphs:

1. Could plot success percentage in different geo locations.
2. Plot Backers count against Category and Sub category to show which of those attract more backers.

# Summary Statistics Analysis

* Median summarizes the data more, as the backer's count data set is skewed.
* From the Summary Statistics, we could conclude that variability of successful campaigns is more than that of unsuccessful ones. The calculated  *'Variance'* value of the campaigns implies that the data of successful campaigns is more deviated from the mean than the Unsuccessful projects.