**Homework 1 - Kickstarter – Conclusions**

# **What are three conclusions we can make about Kickstarter campaigns given the provided data?**

* “Theater” is the most popular Kickstarter category. About 34% of the campaigns fall into this category. After “Theater”, the next most popular category seems to be “Music” (17%).
* The success percentage (77%) of the projects in the “Music” category was the highest and the success percentage of the “Journalism” projects was the lowest (0%).
* All the projects under “Journalism” category were cancelled.
* Of all sub-categories, “Plays” is the most popular and the projects applied to this sub-category seem to have a high success rate (65%).
* Kickstarter campaigns with larger fund-raising campaigns (greater than 10000$) are less likely to be successful than those with smaller pledge amount requirements.
* About 75% of Kickstarter campaigns are from the US and 15% of them are from GB.

# **What are some of the limitations of this dataset?**

* The dataset is a very small subset and due to this it is not possible to make assertive conclusions about the Kickstarter campaigns.
* More information about the demographics (age range, individual or institution, etc.) of the backers will help in observing the trends of the Kickstarter campaigns. For example, we can test the hypothesis - ‘the projects in the technology category have more institutions than individuals as backers’.
* When analyzing market trends, one of the most important analysis is with regards to time (eg: forecasting, trends from moving averages, etc.). In the given dataset, we have information about when the project was created and what was its deadline. Information like when did the project reach its 50% mark or 75%mark would be useful in drawing more actionable insights about Kickstarter campaigns.
* We do not have a complete understanding of the variables in the given dataset as the descriptions of the observed data are not given. From the names of the variables, it is possible only to take a guess about what the variable possibly denotes.
* From the dataset, we can calculate the average funding a project received. However, it would be useful to have standard deviation as a parameter as it would help in understanding the variability in the funding the projects received.

# **What are some other possible tables/graphs that we could create?**

* A two-way table of *state* (was "successful," "failed," "cancelled") versus *staff\_pick* (“True”,”False”) will be useful in seeing association between the staff’s choice and success of the project (backer’s choice).
* A table/chart comparing the total number of *backers* in for each *Kickstarter category/sub-category* would be useful. This way we can analyze if there’s a correlation between overall interest in a category/subcategory and its success.
* A table with *countries* as rows and *category* as columns would be useful to see if there are any country-wise trends that exists.
* A table with *years* as rows and *category* as columns will be useful in analyzing the change in popularity in the Kickstarter campaigns over the years.