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

**ANS.** The 3 main inferences we can draw about Kickstarter campaigns are –

[The sample of 4000 projects is not a true representation of the population (300,000 projects) as the case study suggests that one third of the projects were only successful but, in the sample selected more than half of the projects are successful. These inferences are based on the sample provided]

1. Among the categories Theatre had the maximum projects, followed by music, film and video. All these categories had more than 50% success rate.

Among the categories journalism had all its projects cancelled.

1. Among the subcategories, maximum projects were under plays, and it had a success rate of more than 50%.

Rock, Documentary, and hardware had good number of projects, and all were successful.

Animation, video games and drama had good number of projects, and all were failures.

Audio, art books, science & fiction, and world music had all their projects cancelled.

1. Based on the date of conversion, the projects show maximum live projects happening from January and March.

Maximum successful projects have happened in the month of May.

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

**ANS.**

* 1. Generalizability- The study sample is not representative of the population. The sample of 4000 projects is not a true representation of the population (300,000 projects) as the case study suggests that one third of the projects were only successful but, in the sample selected more than half of the projects are successful. These inferences are based on the sample provided
  2. Lack of information from previous years for comparison and make a proper affirmative inference.
  3. The data is not properly standardized as the fundings are of different currencies.
  4. The same category & subcategory shows failure and success at almost similar rate, different goals and fundings and different backings for the similar categories suggest that the study should be qualitative rather than being quantitative as variables are not well defined.

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

**ANS**

* Percentage of funding to categories/subcategories.
* Backers count to percentage of funding in every category/subcategory.
* States to country

## **Bonus Statistical Analysis – CAMPAIGN**

## **Use your data to determine whether the mean or the median summarises the data more meaningfully.**

**ANS**. The Standard Deviation is greater than its mean, and this indicates high variation between values, and abnormal distribution for data. In such case, it is advisable to use median and interquartile ranges instead of Mean and standard deviation to describe the data.

The normal distribution values calculated in the statistical analysis worksheet shows many outliers and how this data is skewed and an abnormal distribution. The median is the best choice when the distribution of data values is skewed or when there are clear outliers or is abnormally distributed.

## **Use your data to determine if there is more variability with successful or unsuccessful campaigns. Does this make sense? Why or why not?**

ANS. The variability is more in successful projects than in unsuccessful campaigns. This suggests the unsuccessful projects more consistent than successful projects.

But in both cases the variability is showing high difference to central tendency and hence an abnormal distribution of data is observed.