Q. Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?

A: i) The number of successful (565) campaigns exceed failed (364) or cancelled (57) events as it can be seen in Plot 1.

ii) The number of companies in various categories vary quite a bit. There is more investment in Theater, music and film &video categories with large number of successful as well as failed companies. On the other hand, there are not that many companies for journalism, photography and food. There are certain sub-categories that are 100% successful, such as, audio and world music. Although the number is too small, 4 and 3, respectively

iii) The time of the year seem to play a role in the success of a company (Line plot). Companies launched in summer (May – July) seem to have higher chances of success. There were more cancellations for companies created in Feb and Aug, while more failed companies in August.

iv) There is a higher rate of success for goal targets 15000 -35000 (Plot 4).

Q. What are some limitations of this data set?

A: Although the dataset covers a large number of categories and sub-categories overall, but the number of entries for some of the sub-categories is very small to draw any meaningful conclusions about them with confidence. It is difficult to predict if the conclusions from this study are applicable to the smaller sub-categories. Expanding the dataset to include more of the poorly represented sub-categories will make the analysis more meaningful.

Q. What are some other possible tables and/or graphs that we could create and what additional value would they provide?

A: i) It would have been worthwhile looking at campaign outcomes separately for each county. It is possible that certain categories are more popular in some countries and not in others.

ii) Similar to plot generated for campaign date created, it would be interested to look at the outcome on the completion dates.

iii) The data is collected over several years and what we are seeing is the aggregate effect. For campaigns completed in similar timeframe, a separated analysis could be performed.

iv) How does number of backers affect the campaign outcome? Is there a correlation between number of backers and success outcome?

v) It will be worthwhile to see what does staff\_pick and spotlight add to the data? Is there any correlation with the campaign outcome?

Plot 1: Number of successful, live, failed or cancelled campaigns per category.

Plot 2: Number of successful, live, failed or cancelled campaigns per sub-category.

Plot 3: Distribution of companies on the basis of their month of creation.

Plot 4. Percent outcome in various ranges of goal target.

Q: Use your data to determine whether the mean or the median better summarizes the data.

A. Median describes the data as most of the data points are towards median. Averages on the other hand are not in the middle of the data distribution. A few data points on the higher extreme are skewing the average in both the cases.

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

A. Successful campaigns have higher variability. This makes sense because the range of the data is much broader for successful than unsuccessful campaigns. Also, there are more entries for successful than unsuccessful campaigns.