Analysis:

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

1. The most number of ventures are done in Theater, of which the sub category is Plays and hence it has the highest count of successful and failed projects, which is 187 and 132 respectively with the success rate of 70.588%. On the other hand Journalism had the least attempted projects but the success rate is 100% as all the 4 projects were successful.

2. For the projects overall, the month of June and July have been very successful but for the ventures in Theatre June, sept and October have been more successful

3. The country with the highest number of projects is US. It has a total of 763 projects out of which 436 were successful and 274 which is 57% and 35.9% respectively. The most of these projects were Plays which is the subcategory of Theatre.

* What are some limitations of this data set?

The data does not describe when the funds were made available to each project helping us determine the role of funds on each stage of the projects. Also there is no information about the Backer’s profile. The factors such as age, gender, financial strength, familiarity in industry, portfolio etc often influences the success rate of the project. Across the data for the countries it states which countries were successful in which data but there is not enough data backing the reasons for success of those projects in those countries except the funds available.

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

Tables including some information about the Backer’s portfolio such as industries they belong too, success rate as backers and age could help determining their influence on the project.

Some information of when the funds were introduced in the projects across its life will help us judge which stage of the project needs most cash inflow. A line graph across the projects age will help determine that

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

The data is right skewed so Median would be better than mean to summarize the data

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

In the data the successful projects have higher interquartile range and a higher variance. Standard deviation for the successful projects is 1266.24 and it shows more variation compared to failed cases which has standard deviation of 959.99.