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

Below are the three conclusions from the Kickstarter campaigns:

* ParentCategory sheet: Music, theatre and film/video have a good success rate compared to other categories (as shown in the graph)
* LaunchDateOutcomes sheet: As illustrated in the line graph, for projects launched in December there are more projects which failed than were successful over time.
* GoalOutcomes sheet: As observed in the table, projects with lower goal amounts have higher success percentage. Conversely, the failure rate is higher for projects with larger goals. However, we do not have any data which could speak to why the big projects might have failed.

1. What are some limitations of this dataset?

* Limited data set:

Out Of 300,000 projects funded by Kick Starter, the current dataset is an analysis of only 4000 projects which is 1.34% of the total data set. This may not be representative of the full data set. Also, we don’t know how this sample data set was chosen and if it was a biased selection.

* Number of projects in each category is not uniform:

Some categories seem to have way more projects (music, theatre etc.) than others in the sample, thus for the smaller sample sized data (category wise), we are creating more margin of error in the analysis in the bigger picture.

1. What are some other possible tables and/or graphs that we could create?

A table and/or graph could be used to show the relation between the percentage distribution of project funding in regard to the parent categories, e.g., Music. For instance, the graph within sheet “Category\_PercentFunded” illustrates how funding impacted the state of the projects.

Additionally, the table within sheet “SubCat\_PercentFunded” delves further to show how funding impacted each subcategory via slicing the data between years and subcategories, along with their state.

Specifically, the above 2 tables show a relation between funding (or reaching the fund goal) and the success rate of a project. Furthermore, the tables and graphs show that every successful project in most of the parent categories have achieved more than 100% funding.

Despite the link between project success rate and funding percentage, there is what appear at first to be contradictions contained within the data. For example, projects related to subcategory ‘wearables’ under technology, were cancelled despite having over 700% funding in the year 2016. However, when filtering the records for 2016, we can see that 1 project is offsetting the entire data set, i.e., the “HALLAM new york SMART JACKET 2.0 for TRAVEL with 29 FUNCTION”. This project was cancelled, however it had 21535% funding, which is an outlier when compared to the state of other projects within the same category. Unfortunately, the data in the excel is insufficient to determine what led to the cancellation of this project.

Lastly, another limitation to the data set concerns the amount of funding a project receives. Though majority of the ‘successful’ projects are fully funded, the data shows that a lot of these projects are heavily funded. We do not have any data to support why a percentage of these projects were over funded, moreover if overfunding contributed to their success rather than them meeting their goals.