**Report**

**1. What are 3 conclusions we can make about Kickstarter campaigns given the dataset?**

* Out of 4114 total projects, there are 2185 successful projects and 1530 failure projects. The successful rate overall is higher than failure rate, and this is also true for every single category. Investigating further by looking at the outcome vs goal chart, the successful project rate tends to go down as goal amount increases. This explains why the failure and canceled rates follow similar trend. They tend to go up as goal amount increases. There is an interesting observation about the successful rate for projects with goal in the range from 30,000 to 45,000 as projects in this goal range don’t follow a general success/failure trend. One can study this exception further by looking into several attributes such as: project category, date launched or number of backer etc…
* Most of Kickstarter projects fall into 4 major categories: theater (1393 projects), music (700 projects), technology(600 projects ) and film&video (520 projects). Most of these projects start in the US. Among the sub-category, plays seem to be the most popular one among Kickstarter projects
* The average number of kickstart project is similar for each month. By looking at date created vs outcome, projects which start early in the year tend to have higher chance of being successful compared to projects that started later in the year.

**2. What are some of the limitations of this dataset?**

* There are no clear dependencies among the factors in the dataset. Therefore, it’s hard to tell which factors do contributes the most or least to the success or failure of a project.
* The currency column does not have the same currency units – Data pre-processing needs to be performed to make sure the goal amount follows the same type of currency or scale when we need to make comparison among different projects ( or simply use the normalized funded ratio)
* Big number of project start in the US, so it may not be appropriate to make inference in global scale. Also, most of the projects belong to 4 big categories, so using the prediction result from this dataset to make prediction for projects with similar goal amount may not achieve high accuracy.

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

* Scatterplot of successful rate vs funded ratio or failure rate vs funded ratio to look for dependency between the outcome and funded ratio
* Scatter plot number of backer vs state – to see if number of backer do affect the outcome
* Staff pick or spotlight vs state – Looking for accuracy staff prediction vs projection