**Assignment 1**

**Data Analysis for KickStarter**

Out of the 4,114 records for projects included in the KickStater dataset, the top four categories of projects submitted included theater (34%), music (17%), technology (15%), and film and video (13%).

Overall, the proportion of projects that were successful is 53%, whereas 37% of the projects failed, 8% were cancelled, and currently only 1% is live. This means that overall about 1 in 2 projects gets funded, close to 4 in 10 fail, and 8 in 100 projects get cancelled across all countries (fig 1).

Similarly, across all countries, the categories of projects with higher proportion of successful endeavors are theater (20%), music (13%), and film and video (7%). In contrast, the categories with the larger proportion of failed projects are theater (12%), technology (5%), and film and video (4%). Half of the projects that get cancelled follow under the technology category (4%) (fig 1).

In the USA, the proportion of projects in music (77%), theater (57%), and film and video (62%) is relatively high. Similarly, theater (38%) and film and video (30%) also show relatively high proportions of failing (fig 2).

Taking a closer look at the data by subcategories will shed light on what specific type of projects may be more viable over others across all countries. For instance, plays (17%) have a higher proportion of success on being funded through this platform. Other projects like documentaries (4%), hardware (3%), indie rock (#%), and photobooks (3%) had higher proportions of being successfully funded (fig 3).

In contrast, the top three subcategories of projects that failed were plays (9%), food trucks (3%), and wearables (3%). While the proportion of projects in subcategories such as classical music, hardware and rock is rather small, thus far these projects have a perfect rate of being funded (fig 3).

When looking at the dates in which these projects are being created, data suggest successful endeavors take place in the late spring around May. This trends down during the summer and it seems to pick up around the time the school year begins, falling again as the holiday season approaches in December. Failed projects seem to follow a similar pathway and cancelled projects seem to be very steady throughout the year across all countries (fig 4).

About 70% of projects whose goal was to collect less than $1,000 were successful. As the monetary goal increased, the proportion of successful projects decreased to about 40% at around the $30,000 to $34,000 range. For goals between $35,000 to $44,999 the proportion of successful projects increases to almost 50%. However, for projects with more ambitious goals with $45,000 and beyond, the proportion of successful projects decreases to about 20% (fig 5).

Inversely, the proportion of failed projects with a goal of $1,000 starts at around 25%, and steadily increases to 50% at the $25,000 point. Moreover, this proportion decreases about 10 percent points at the $45,000 mark. The proportion of projects that fail takes a big hike for projects who wish to collect over $45,000 reaching almost 60%.

2) Limitations of dataset

While I have not identified many limitations, I believe in order to conduct a robust analysis, variables such as currency must be normalized to a reference currency. Currently, the dataset show variables in US dollars, pounds, Australian dollars among others. The analyst must be aware of this matter and convert all currencies to a reference one.

Furthermore, when looking at trends for periods in which projects are more likely to be funded, we cannot generalize across countries given that seasons are inverse among countries located in the Northern and Southern hemispheres.

Last, it was not clear to me from this dataset as to whether the backers were all local or if this platform allows for projects to be funded by individuals abroad. Perhaps getting further information as to where the funding came from and the amount of money they were giving to these projects, may allow organizations to target specific potential donors.

3) It seems we can create multiple analyses to further inform decisions. For instance, bivariate analysis between countries and categories and subcategories, proportion of successful, failed, cancelled and live projects. We looked at counts, but to better inform and make decisions, proportions will be a better outcome to look at. Furthermore, regression analysis looking at potential correlations between countries and types of projects, projects and size of monetary goal, or goal and season of the year may be other interesting graphs that could inform organizations in making decisions.