# Kickstarter Challenge- Week 1

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

Crowdfunding overall has about a 50% success rate.

Theater/Plays category are more often successful than other categories. It had a significantly higher success rate than other categories.

The month of August has one of the the highest number of failed and the lowest number of successful kickstarters. This adds a question of why is August not a month selected by campaigns to end, and why do so many of those that do choose this month to fail?

1. **What are some limitations of this dataset?**

The dataset includes seven countries, which can portray a skewed results due to the country. As an example, there are internet restrictions on a country. Then the number of kickstarters for this country will be lower than a country without limitations.

Data is less recent. Being a web-based fundraising tool, analyzing data from 2015 to early 2020 can give a historical view of the business. The lack of recent data would only allow the business to make a dated decision, which could result in financial ruin.

The few categories cause a narrow picture of the data. Without a sample of all categories, a decision can only reliably made if we are making direct comparisons between these categories. Any business-wide conclusions drawn would not encompass those missing categories.

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

A relational discovery graph to determine if the “Staff-Pick” gives higher numbers of backers than those not listed as a “Staff-Pick.” A scatterplot to determine if the total backer count for each kickstarter is positively or negatively correlated to “Staff-Pick.”

A graph to determine if the “Spotlight” gave a kickstarter a better chance of success or not. A bar chart to show the spotlight and non-spotlight kickstarters successes. A follow-up calculation to determine the percentage of increased successes for spotlighted kickstarters. These items would provide value in that they give the business an idea if their “Spotlight” and “Staff-Picks” are helping the campaigns.

1. **Bonus Statistical Analysis: Use your data to determine whether the mean or the median summarizes the data more meaningfully.**

The median will more closely represent the data than the mean. The distribution of the data is mostly between the 0-400 on the chart, but the mean is about 851 and 585 for successes and fails, respectively. The median of 200 and 114 for success and fail better represents the data that falls under four hundred.

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

After performing a box and whisker plot on the successes and failures and checking the Upper and lower bounds manually, a higher variability in successful campaigns over unsuccessful is present. It was an expected outcome due to the backer counts found in the campaign types. Failed backer counts tended to have few to no backers, which brought the variability down. There were outliers with high numbers of backers, but they were rare. While successful campaigns had more occurrences of higher backer counts as well as small backer counts. Successful campaigns received their funding from both few and many backers in a consistent manner, causing the variances seen in the data.