**Frank McKenzie-Stripp**

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

It is difficult to come up with any solid conclusions regarding the data without having performed statistical analyses, but it appears that:

1. Theatre is the most popular primary category – as there are far more projects within this category than any other. (In general, it appears that entertainment (e.g. music, film & video, theatre) crowdfunding projects are the most common.)
2. Plays are by far the most common sub-category for crowdfunding projects.
3. July is the month in which projects are most likely to be successful, followed by June. (Perhaps this may be related to this period being summer in the northern hemisphere?)
4. **What are some limitations of this dataset?**

Not all successful crowdfunding campaigns actually go on to be successful products – perhaps including follow up data relating to this could be useful.

Calculating the average donation is good, but it is still definitely possible that for some of these projects, a few people funded the majority of it (a handful of very large donations) and the rest were many minor donations.

Including the specific site each of these projects was posted on (e.g. Kickstarter, Indiegogo) would be useful, as any potential differences in the success rate between these sites may help with deciding which site to post future projects on.

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

Comparing the ratio of projects that were successful vs unsuccessful across different categories, countries and years would interesting. This may provide some insight into whether projects of a certain type, or in a certain country are more likely to be successful, as well as whether projects have become more or less likely to succeed over time.

Also, visualising the amount of successful campaigns that were either staff picks and/or in the spotlight vs how many were successful that were neither of these would provide an idea of how much of an impact the staff picks and spotlight has on what does/doesn’t get funded.

**Bonus statistical analysis –**

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

In this case the median appears to more accurately represent the data, as this dataset is fairly skewed and the mean is being affected by the presence of some extremely high outliers. As the median is essentially the value at which 50% of the values are below it and 50% are above it, it is less susceptible to the presence of outliers and is better used to summarise the data in cases such as this, where it is skewed. We can tell that this dataset is skewed and that a few extremely large outliers are present here based on the significant difference between the median and the mean, as well as the extremely high maximum values, which are nowhere near the median or mean.

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

Successful campaigns have greater variability. This makes sense logically, as we would expect successful campaigns to have on average received more backers, which we can see is the case in the mean and median. As a result, the range of the number of backers is likely going to be more variable for successful projects, as though we may see some projects with which only a few backers were needed to reach the goal (whether that be because they each donated a lot, or the goal was just very low), we will see many more of the very popular projects included in this data, the ones to which an extremely large number of people donated to (this is seen in the higher max value, as well as the higher mean and median).