

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | | | |
|  | | Module 1 Challenge | | | | |  | |
|  |  | | | | | | |  |
|  | | | |  |  | | | |
|  | | | | John Teasley |  | | | |
|  | | | | 3/18/2024—KU Data Analytics Boot Camp—Benjamen |  | | | |
|  | | |  | | |  | | |

# Question 1

### Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?

1. Overall, the parent category “theater” has the largest number of crowd funding campaigns. The results change when looking at individual countries, with Great Britain having more campaigns for “film and video” and Switzerland having more campaigns for “music”.
2. Within the parent category, “theater”, the subcategory “plays” has the largest number of crowd funding campaigns for all countries.
3. More campaigns succeed than fail. Over the past 10 years, July has had the largest number of successful campaigns, with the years of 2014 and 2018 having the highest count of 9 successful campaigns in July.

# Question 2

### What are some limitations of this dataset?

1. The currency is based on the country and could use a conversion to a single exchange rate.
2. We do not have demographics for the backers.
3. The columns ‘staff\_pick’ and ‘spotlight’ are ambiguous.

# Question 3

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

1. We could look at the average donation and see how it relates to each category. Using the outcome status, we may be able to see if there is a correlation between the average pledge and success rate.
2. We can look at the over and underfunded currency value for each category to determine how to maximize and improve the success rate for future endeavors.

# Statistical Analysis

### Use your data to determine whether the mean or the median better summarizes the data.

1. Due to the high variance of the data, the median would be the most accurate use for the data summarization.

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

1. The successful backer count has a larger variability than the failed backer count.
2. The variance based on backer count alone does not provide enough information to make a sound judgement. Predictions for future campaigns cannot be made due to the lack of constraint. More exploratory data analysis may provide a better understanding of the backer count.