Data Analytics Module 1 Challenge

Module 1: UCI Data Analytics Bootcamp

David Jordan

UCI Data Analytics Bootcamp

**Crowdfunding Goal Analysis**

1. Given the provided data, what are three conclusions that we can draw about crowdfunding campaigns?
2. Theatre/plays are clearly the most popular crowdfunding campaigns by a wide margin. Which skews the aggregate. However, there appears to be correlation in the yearly trend regardless of including or excluding theatre; however, the trend seems to be slightly more volatile without theatre included.
3. The data suggests that over time the success rate of campaigns increased while the failure rate decreased over the same time period. This trend could lead to the conclusion that the crowdfunding sites became more popular and or trusted.

A picture containing plot, line, screenshot, diagram

Description automatically generated

1. The average donation and failure rate trends are somewhat flat regardless of the category. However, for theatre and film/video, there are considerably more campaigns over the other categories. Which could mean the averages for the other categories are not evenly weighted.

A picture containing plot, line, diagram, screenshot

Description automatically generated

What are some limitations of this dataset?

An obvious limitation of the dataset is there is no indicator of what crowdfunding site the data is derived from. By knowing the origin of the data could influence the output based on the subject site. For example, one site such as Kickstarter could have a higher success rate year on year when compared to another site, such as Indiegogo. Further, 1000 sample population may not be enough sample data to fully understand the success, failure, and/or averaging funding.

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

A possible table/graph that would be useful is knowing the ‘success’ rate of crowdfunding [(Cancelled+ failed) + successful campaigns] / successful campaigns. This cut of the data illustrates the growth in the success rate over time, which could be a useful key indicator.



Success rate could be illustrated at the parent category level as well.

A picture containing plot, line, diagram, screenshot

Description automatically generated

**Statistical Analysis**

**A picture containing text, screenshot, font, number

Description automatically generated**

In the above tables, the mean (average) is the more suitable statistical metric to best summarize the dataset because the figure represents actual backer count rather than just the middle count figure. The successful campaign’s variance is higher when compared to failed campaigns; however, the mean of successful campaign backers is approximately 31% greater than the mean of failed campaign backers.