

SDS 313 - Project 2 Data Proposal

1. A brief description of the main topic of your project and why you chose it.
 - a. I chose this dataset because it provides an interesting look at the titles that have resonated the least with audiences, which could lead to insights about what makes a movie or TV show unpopular or poorly received. Analyzing this data could reveal patterns or common factors among the worst-rated titles, which could be valuable for understanding audience preferences and the factors that contribute to a title's critical reception.
2. A link to the source(s) of data you want to use with a brief description of where the data comes from.
 - a. <https://www.kaggle.com/datasets/octopusteam/imdb-top-1000-worst-rated-titles?resource=download>
 - b. The data is scraped from the IMDb website.
3. A brief description of how you plan to import the data into R and any merging/processing/reshaping you expect to have to carry out before you can begin analyzing it.
 - a. I believe that the data is already set to be investigated as is, and it has a usability rating on Kaggle of 10.0/10.0 meaning it generally is good as is.
4. A list of some of the variables of interest you plan to investigate.
 - a. Something that could be interesting to investigate is the number of votes different genres get and if maybe that has an influence on the average rating (for example, does comedy usually get more votes leading to a higher average rating as compared to horror that might get less votes that lead to a lower average rating)