Project 1 – Using IMDB Data to Study Differences in Film Genres

Covering Descriptive Statistics, Metrics, and Aggregations

Utilizing IMDB Dataset

**Topics to be Reviewed/Studied in this Project**

* The Following Descriptive Statistics
  + Average
  + Median
  + Mode
  + Standard Deviation/Variance
  + Count
  + Maximum
  + Minimum
  + Count Distinct
  + Percentiles
* Creating Aggregated Statistics Based on Grouped Values
* Creating Derived Metrics to Further Understand Data
* Topics Covered in All Analyses
  + Data Review & Preparation
  + Best Practices
  + Proper Project Organization and Planning
  + Writing and Understanding Research Questions
  + Creating Data Visualizations and Presentations

**Hypothetical Scenario**

A major streaming service has an upcoming launch and is unsure what kind of programing it should place on its platform. They want to know what Genres are most popular, and the most well liked. Since they are just starting out, they want to focus on genres with consistent scores to minimize risk. Considering the platform needs to have content available for all ages, it needs to have a diversity of different age ratings. They want to know how this should impact which genres they chose to produce, and what factors affect the age rating of different genres. They also want to understand what audiences expect in terms of film length or the number of episodes in a series. Finally, they are interested in understanding which films and TV shows they should try to purchase outright for the platform and need a list of the top 50 films and tv serries respectively. They would like to see a tableau dashboard displaying this information

**Research Questions**

* What are the top 20 films and tv shows with the most votes?
* How many different movies are in each genre?
* What are the top 10 most common genre combinations?
* What are the top 5 Genres with the most films and shows, what are the bottom 5?
* What is the median number of votes for films in each genre?
* What is the average number of votes for films in each genre? Percentiles?
* What is the mean and Standard deviation of film and serries lengths? Percentiles?
* What percent of films in each genre belong in each age rating category?
* What is the relationship between different ration category attributes, and different genres?
* What recommendations should be made to the streaming platform?
* What potential flaws or areas of further research should be done?