Homework: Unique Randomness

Dataset Link: https://www.kaggle.com/datasets/ritiksharma07/imdb-top-1000-movies-dataset

For this homework assignment, I selected a relatively up-to-date dataset on IMDb's top 1000 movies. IMDb (Internet Movie Database) is an well-known online database that provides information on movies and TV shows, as well as other popular forms of media. Since the dataset I selected was scraped directly from IMDb, I was able to uncover significant details on each movie, including its year of release, duration, age rating, rating, number of ratings, metascore, description, and name. Although I did not use all of these attributes in my final visualization, I found exploring the various columns of data both fascinating and informative. I specifically selected a dataset about movies as I thoroughly enjoy watching movies in my free time. I felt that analyzing randomness within this dataset would be both engaging and enjoyable, and allow me to connect my academic goals with a hobby I'm passionate about. In regards to the actual data, I sourced the dataset from Kaggle, which is a well-known online platform that hosts a wide range of datasets. Having taken several data science courses in the past, I was already familiar with the platform and thus was able to find a suitable dataset relatively quickly. One of the key advantages of using Kaggle was that I was able to directly download the dataset as a .csv file, which made the initial setup process relatively straightforward. However, one of the challenges I found was that several of the datasets I initially explored were not fully cleaned. For example, before finding my final dataset, some datasets I encountered were composed of incorrect or improperly formatted data, which would have led to errors during analysis. Furthermore, some of the datasets were also outdated or were missing columns of data that I would have liked to explore. Thus, although I was able to access the data relatively smoothly, there were some challenges that I had to navigate before arriving at my final dataset.