The second data source we used was Rotten tomatoes(https://www.rottentomatoes.com). Rotten Tomatoes consisted of reviews from a wide range of audiences. All the movies were categorized into different Genre and we found it is something similar to IMDB and it was being used to compare with the initial data scrapped from IMDB. We scrapped the movie name, date, and the rotten tomatoes rating from the location and used it to compare it with the IMDB data.

We found that the Rotten Tomato rating was much higher than the IMDB rating. It is because IMDB has more users to rate capered with Rotten Tomato’s. Initially scrapping the data was easy but it was not tidy. We had to remove the blank spaces, unwanted brackets and we also had to separate the data which was attached to the movie name. The other difficulty we faced was converting the movie rating from percentage to decimal values. We used packages like rvest, tidytext and tidyverse to get the necessary functions to clean the data. The majorly used function was “gsub” to clean the data obtained and “write.csv” to store the data.

One of the main difficulties we faced was we were not able to scrape large amount of data from Rotten Tomatoes as it was not permitting as the website was not permitting us to collect the whole data. Therefore, we had to stick with the top 100 movies based on the Genre.