A data scraping exercise was performed in Jupyter Notebook (RT100\_analysis.ipynb).

The purpose of this exercise is to understand how the critic ratings weigh against audience ratings for the top 100 films rated on Rotten Tomatoes and whether critic ratings should be the ultimate truth metric for a movie's success rating

The exercise involved multiple steps:

1. Data Wrangling – Data was scrapped from multiple webpages
2. Data Cleaning
3. Data Analysis & Visualization
4. Data Wrangling:

Data was scrapped from multiple webpages:

1. The URLs and rankings for the RT top 100 movies were scrapped from the homepage: <https://www.rottentomatoes.com/top/bestofrt/>
2. The URLs were used to extract the following information:
   1. Audience Count
   2. Audience Score
   3. Critic Count
   4. Critic Score
   5. Box Office Earnings
3. The review of the top-rated movie was scrapped off the Roger Ebert Review website

Media Wiki API, wptools, was used to access Wikipedia data for the top ranked movie.

1. Data Cleaning

Data was cleaning using various tools available in the pandas library. All the data was combined into a single pandas DataFrame for analysis

1. Analysis and Visualization

Data was analyzed using pandas. A confidence interval test was conducted by bootstrapping the data points and multiple exploratory and explanatory visualizations were plotted using:

1. Matplotlib
2. Tableau