**Similarity Analysis of Movie Reviews Across IMDB, Metacritic and Rotten Tomatoes**

**Name: Prajwal Prasad**

**ID: 1001750483**

This project focused on collecting movie reviews of top 50 movies across IMDB, Metacritic and Rotten Tomatoes websites and calculating the cosine similarity between the 50 most frequently used words in each of those sites.

The review data collected for each website is as follows:

* **IMDB**: 50 most popular user reviews, these are pretty lengthy compared to reviews on other sites.
* **Metacritic**: I have considered 50 most popular user reviews or how many ever present, if there were no user reviews for a movie, I have opted for critic review instead of skipping.
* **Rotten Tomatoes**: I have considered critic reviews for this since some movies did not have any user reviews or less, also the critic reviews were easier to access.

**Similarity scores for reviews:**

Similarity score for keywords between imdb and metacritic is: 0.953053201178429

Similarity score for keywords between imdb and rotten tomatoes is: 0.8274495152606323

Similarity score for keywords between rotten tomatoes and metacritic is: 0.803066869486583

The cosine similarity score ranges from 0 (least similar) -1 (most similar), and based on my results, we see that the scores are closer to 1 but IMDB and Metacritic are extremely close with 0.95, whereas rotten tomatoes is a little less similar showing very close results with both IMDB and Metacritic around 0.8.

**Similarity scores for genres:**

Similarity score for genres between imdb and metacritic is: 0.8857105210415298

Similarity score for genres between imdb and rotten tomatoes is: 0.7707048953466868

Similarity score for genres between rotten tomatoes and metacritic is: 0.6824329998331107

Here, the cosine similarity scores are a bit varied. We see that IMDB and Metacritic use pretty much the same genre tags for movies (very close to 1) and again rotten tomatoes give a bit different genre tags as we can see from the score dropping when comparing with it (around 0.77 with imdb and 0.68 with Metacritic).

**I have used Website, Rank, Movie Title, Genre as my columns for both DB and CSV.**

**-Note: I have used ‘Metacritics’ instead of ‘Metacritic’ in my code, please ignore this grammatical mistake.**