WebScrape Insights Project Proposal

1. Project Overview

This project is designed to collect movie data from the [metacritic.com](https://www.metacritic.com) website through web scraping technology, and then clean, analyze, and visualize this data to gain insights into movie ratings, genre distribution, director works, and other aspects.

2. Project member

Chuyi Wang
chuyi@usc.edu
5458365139

3. Project Objectives

This project aim to find movie data and to analyze the relationship among film critics' ratings, user ratings and the final rating.
This project will focus on these questions:
1. What is the relationship between critics' ratings, user ratings and the final Metacritic score? Are they consistent?
2. Do the characteristics of a film, such as its genre, production company, and release year, present certain patterns? Could these features possibly affect a movie's rating?
3. Is there consistency within the internal system of film critics and user ratings?

3. Data Sources

Data will be collected from metacritic.com, including basic attributes such as title, year, type, and production company I will use request to obtain content and use BeautifulSoup parse structured information. Also this project obtain the rating data and sentiment ratios from both film critics and users, and record the final Metacritic comprehensive score. Web crawlers use pagination to capture multiple pages of content and incorporate deduplication mechanisms and random delays during the process to enhance the stability and reliability of data collection. It is expected to collect data from 50 to 100 films, depending on the number of pages on Metacritic.

4. Data analyze and visualization
To present these structures and patterns more intuitively, this project will display the main scores, feature distributions and correlation results in the chart.

Data analysis content

Descriptive statistics of the scoring data
A comparison of the differences between film critics' and users' ratings
Correlation analysis (variables such as the number of comments, the proportion of emotions, and ratings)
Observation on the distribution and trend of types, production companies and years

Visual charts

Score statistics chart
Distribution map of types and production companies

Trend chart of release year
Full variable correlation heat map