**IMDb Data Processing and Web Scraping Project**

This project aimed to collect and preprocess movie data from IMDb, starting from web scraping to the preparation of a clean and structured dataset. The following steps summarize the workflow:

-A web scraping script was implemented using Python libraries such as requests and BeautifulSoup. It targeted the IMDb "Most Popular Movies" page to extract key details about movies, including titles, descriptions, ratings, genres, and durations.

-The scraped data was stored in a structured format (CSV) to facilitate easy access and further analysis. The raw dataset contained inconsistencies such as missing values and non-uniform formats.

-An initial dataset of 100 movies was created, capturing essential fields like movie title, URL, description, best/worst ratings, average ratings, genres, and duration.

-A cleaned dataset was produced by addressing missing values, formatting durations into a consistent ISO 8601 format, and ensuring all fields were correctly parsed.

-Scripts were written to standardize and preprocess the data. This included handling missing values, converting ratings to numerical formats, and ensuring genre lists and durations were uniform.

-Advanced cleaning scripts further refined the data by:

* Removing incomplete records (e.g., missing titles or essential ratings).
* Transforming textual duration formats into numerical equivalents for analysis.
* Validating the integrity of URLs and descriptions.

-The final dataset was prepared as a ready-to-use resource for further analysis or machine learning tasks. It contains clean, standardized movie information.

-The entire workflow was documented for reproducibility. This includes:

* Clear explanations of each function in the scripts.
* Usage instructions for scraping and preprocessing scripts.
* Details on handling edge cases and common errors during execution.

**Key Achievements**

* Successfully extracted, cleaned, and stored IMDb movie data.
* Delivered a well-documented workflow for future scalability and usability.
* Prepared a high-quality dataset ready for advanced data analysis.