

# **IMDB Analytics**

## **Group ID 7**

Ahmed Ibrahim

Etiene Effiong

Rupinen Kristen

Asamoah Paul

## **Internet Movie Database (IMDB)**

The IMDb Movies dataset is a rich and informative collection of data that has been meticulously compiled from the Internet Movie Database (IMDb), one of the world's most popular and authoritative sources for movie, TV, and celebrity content.

Each movie in this dataset is a representation of the diverse and dynamic world of cinema, spanning different eras, styles, and cultural backgrounds. The dataset includes both critically acclaimed masterpieces and popular mainstream films, providing a comprehensive view of the movie industry. With data points like ratings and Metascores, the dataset offers a quantitative glimpse into the reception and critical acclaim of the movies. Additionally, the inclusion of details like directors and star casts helps in understanding the creative forces behind these cinematic works.

This dataset serves as an invaluable tool for movie enthusiasts, industry analysts, and researchers alike. It offers a foundation for various analytical projects, from exploring trends in movie genres and audience preferences to studying the commercial success of films. By delving into this dataset, one can uncover patterns and insights that reflect both the artistry and the business of filmmaking. Whether used for academic research, market analysis, or simply exploring the vast world of movies, the IMDb Movies dataset stands as a testament to the rich and evolving landscape of cinema.

### **Data Collection**

The datasets were generated through the utilization of Python's scraping library BeautifulSoup. The approach entailed gathering movie data from diverse users on IMDb. Hyper Text Markup Language (HTML) tags and elements were strategically employed, utilizing Google Chrome's page inspection feature to curate the necessary data.

To achieve this, the code executed a loop through the IMDb's movie pages of five users, individually extracting various details for each movie. Subsequently, the datasets obtained from each page were merged to form a comprehensive dataset that could be exported as a Comma Separated Values (CSV) file named "merged\_movies\_dataset".

The details extracted from the links include Title, Year, Genre, Runtime, Imdb Rating, Metascore, Summary, Director, Stars, Votes, and Gross Revenue.

The dataset was extracted from the links below:

[User One](#)

[User Two](#)

[User Three](#)

[User Four](#)

[User Five](#)

## Data Structure and Variables

The dataset includes 8992 samples and 11 variables of movies from IMDb's website.

Column Names	Description	Data Type
Title	The title of the movie	String
Year	The release year of the movie	Integer
Genre	The genre(s) of the movie	String
Runtime	The duration of the movie in minutes	Integer
Ratings	The IMDb rating of the movie from a scale from 1 to 10	Float
Metascore	The Metascore of the movie in percentage	Float
Summary	A brief description of the movie	String
Director	The name of the director of the movie	String
Stars	The main stars of the movie	String or List
Votes	The number of votes the movie has received on IMDb based on the popularity or viewer engagement of the movie	Integer
Gross Revenue	The gross revenue of the movie in million dollars	Float