

Data Dictionary

Overview

This Data Dictionary provides structured definitions and descriptions of the dataset used in this project.

The dataset is sourced from IMDb and contains detailed information about movies, ratings, and financial performance.

Table of Contents

IMDB Movie Dataset

Name: imdb_movie_dataset

Description:

A cleaned movie dataset containing metadata, performance indicators, and audience engagement metrics.

The dataset has been preprocessed to remove inconsistencies and ensure accurate analysis.

Column Details

1. Movie_ID

Data Type: String

Description: Unique identifier assigned to each movie.

Relevance: Ensures record uniqueness and prevents duplication during analysis.

2. Title

Data Type: String

Description: Official name of the movie.

Relevance: Used for identification, reporting, and labeling visualizations

3. Release_Year

Data Type: Integer

Description: Year in which the movie was released.

Relevance: Enables trend analysis and time-based comparisons.

4. Genre

Data Type: String

Description: Category or categories assigned to the movie (e.g., Drama, Action).

Relevance: Used for genre-based performance and rating analysis.

5. Duration_Minutes

Data Type: Integer

Description: Total runtime of the movie in minutes.

Relevance: Helps analyze the relationship between duration and audience ratings.

6. Country

Data Type: String

Description: Country where the movie was produced.

Relevance: Enables country-wise comparison and regional performance analysis.

7. Content_Rating

Data Type: String

Description: Age classification of the movie (e.g., PG, R).

Relevance: Useful for audience segmentation and content analysis.

8. Director

Data Type: String

Description: Name of the movie director.

Relevance: Helps analyze director impact on movie ratings and income.

9. Income_USD

Data Type: Float

Description: Total box office revenue generated by the movie (in USD).

Relevance: Used for financial performance evaluation and profitability trends.

10. Votes

Data Type: Integer

Description: Total number of audience votes received.

Relevance: Indicates popularity and audience engagement.

11. IMDB_Score

Data Type: Float

Description: IMDb rating score of the movie (scale 0–10).

Relevance: Primary performance metric for evaluating movie quality.

IMDB Movie Dataset - Data Dictionary

This Data Dictionary provides structured definitions and descriptions of the dataset used in this project. The dataset is sourced from IMDb and contains detailed information about movies, ratings, and financial performance.

IMDB Movie Dataset

Name: imdb_movie_dataset

Description: A cleaned movie dataset containing metadata, performance indicators, and audience engagement metrics. The dataset has been preprocessed to remove inconsistencies and ensure accurate analysis.

| Column Name | Data Type | Description | Relevance to Analysis |
|------------------|-----------|---|---------------------------------------|
| movie_id | String | Unique identifier for each movie | Helps avoid duplicate records |
| title | String | Name of the movie | Used for identification and reporting |
| release_year | Integer | Year the movie was released | Used for trend analysis over time |
| genre | String | Category of the movie (Drama, Action, etc.) | Analyze rating and income by genre |
| duration_minutes | Integer | Length of the movie in minutes | Study duration vs rating trends |
| country | String | Country of production | Country-wise comparison |
| content_rating | String | Age classification (PG, R, etc.) | Audience segmentation analysis |
| director | String | Name of the director | Director performance trends |
| income_usd | Float | Total box office revenue in USD | Financial performance analysis |
| votes | Integer | Number of audience votes | Measures popularity |
| imdb_score | Float | IMDb rating score (0–10) | Main performance indicator |