◆ Scenario 1: Data Cleaning and Standardization

Business Need: The marketing team wants to use clean and standardized movie data for campaign targeting.

ETL Tasks:

- **Extract**: Load the raw movie dataset (CSV, JSON, or from cloud storage like S3 or ADLS).
- **Transform**:
  - Trim strings (e.g., `Title`, `Director`, `Lead_Actor`).
  - Fill missing `Content_Rating` with `"NR"` (Not Rated).
  - Standardize `Language` and `Country` (e.g., "english", "English" → "English").
  - Convert `Rating` to 1 decimal place.
- **Load**: Write the cleaned dataset to a `CleanedMovies` table in a database or data lake.

◆ Scenario 2: Calculate Movie Profit and Profitability

Business Need: Finance team needs to understand which movies are profitable.

ETL Tasks:

- **Extract**: Use the cleaned dataset.
- **Transform**:
  - Calculate `Profit = BoxOffice_USD - Budget_USD`.
  - Create `Profitability = Profit / Budget_USD` (handle divide-by-zero).
  - Categorize:
    - `"High"` if `Profitability > 1.5`
    - `"Moderate"` if between `0.5 - 1.5`
    - `"Low"` otherwise
- **Load**: Write to a table `MovieFinancials` with columns: `Title`, `Year`, `Profit`, `Profitability`, `Profit_Category`.

◆ Scenario 3: Top Directors and Actors by Awards

Business Need: Awards data is needed for a partnership campaign.

ETL Tasks:

- **Extract**: Pull from `CleanedMovies`.
- **Transform**:

    - Aggregate `Num_Awards` by `Director` and `Lead_Actor`.
    - Rank them and select the top 5 in each category.
- **Load**: Store in `TopAwardedDirectors` and `TopAwardedActors` tables.

◆ Scenario 4: Genre-Wise Performance Summary

Business Need: Analytics team wants insights by genre.

ETL Tasks:

- **Extract**: Use clean movie data.
- **Transform**:
    - For each `Genre`, compute:
        - Average `Rating`, `Votes`, `Profit`
        - Total number of movies
    - Handle multi-genre (e.g., "Action|Thriller") if applicable.
- **Load**: Save results in a table `GenrePerformance`.

◆ Scenario 5: Yearly Movie Trends

Business Need: The business wants to understand trends over time.

ETL Tasks:

- **Extract**: Use main movie dataset.
- **Transform**:
    - Group by `Year`
    - Calculate:
        - Count of movies released
        - Avg `Rating`, `Votes`, `Profit`
        - Most common genre (optional)
- **Load**: Store in `YearlyMovieTrends`.

◆ Scenario 6: Rating & Review Quality Assessment

Business Need: Data science team is building a model to predict box office success.

ETL Tasks:

- **Extract**: Use cleaned movie data.
- **Transform**:
  - Create a feature called `Review_Quality = Critic_Reviews / Votes` (normalize reviews)
  - Create label: `Hit` if `BoxOffice_USD > 1.5 * Budget_USD`, else `Flop`
- **Load**: Write to `ModelTrainingDataset`.

◆ Scenario 7: Production Company Insights

Business Need: Management wants to track top-performing production companies.

ETL Tasks:

- **Extract**: Use clean dataset.
- **Transform**:
  - Group by `Production_Company`
  - Compute:
    - Avg `Rating`, `Profit`, `BoxOffice_USD`
    - Total movies produced
  - Filter companies with at least 3 movies.
- **Load**: Save to `ProdCompanyInsights`.