

431W project

Dataset: Full TMDb TV Shows Dataset 2024 (150K Shows)

Link: <https://www.kaggle.com/datasets/asaniczka/full-tmdb-tv-shows-dataset-2023-150k-shows>

https://www.ict.griffith.edu.au/normalization_tools/normalization/ind.php

<http://raymondcho.net/RelationalDatabaseTools/RelationalDatabaseTools.html>

The TMDB (The Movie Database) is a widely-used resource for movie and TV show data, providing valuable information such as ratings, plot summaries, and more. This dataset contains a collection of 150,000 tv shows from the TMDB database, collected and cleaned.

Use cases:

1. Use Case Name: Show to Genre Mapping

Description: This use case describes the relationship between TV Shows and Genres, allowing users to find shows based on genre. A TV show can belong to multiple genres (at least one), and a genre can categorize multiple TV shows.

Entities Involved: TV Show, Genre

Cardinality: One TV Show to Many Genres (at least one genre per TV show), Many TV Shows to One Genre

Participation: Total for both TV Show and Genre

2. Use Case Name: Language Availability

Description: Describes TV shows available in different languages, supporting searches for shows by spoken language. A TV show can be available in multiple languages (at least two), and a language can be spoken in multiple TV shows.

Entities Involved: TV Show, Language

Cardinality: One TV Show to Many Languages (at least two), Many TV Shows to One Language

Participation: Total for TV Show, Partial for Language

3. Use Case Name: Production Country

Description: Links TV shows to their countries of production. Each TV show is produced in exactly one country, but a country can be the production site for many TV shows.

Entities Involved: TV Show, Country

Cardinality: One TV Show to Exactly One Country, One Country to Many TV Shows

Participation: Total for both TV Show and Country

4. Use Case Name: Show Popularity Over Time

Description: Tracks the popularity of TV shows over time, facilitating analysis of trends. A TV show has many popularity records (at most 10, considering data snapshots), each representing a point in time.

Entities Involved: TV Show, Popularity Record

Cardinality: One TV Show to Many Popularity Records (at most 10), One Popularity Record to Exactly One TV Show

Participation: Total for TV Show, Partial for Popularity Record

5. Use Case Name: Viewer Ratings

Description: Connects TV shows to viewer ratings. Each TV show can have many ratings (at least one rating per TV show), and each rating is associated with exactly one TV show.

Entities Involved: TV Show, Rating

Cardinality: One TV Show to Many Ratings (at least one), One Rating to Exactly One TV Show

Participation: Total for TV Show, Partial for Rating

6. Use Case Name: Content Curation by Network

Description: This use case explores the relationship between TV shows and the networks that broadcast them. A TV show is associated with exactly one network, but a network can broadcast many TV shows.

Entities Involved: TV Show, Network

Cardinality: One TV Show to Exactly One Network, One Network to Many TV Shows

Participation: Total for both entities

7. Use Case Name: Episode Guide

Description: Each TV show consists of multiple seasons, and each season contains multiple episodes (at least one episode per season). This relationship allows for a detailed episode guide for each TV show.

Entities Involved: TV Show, Season, Episode

Cardinality: One Season to Many Episodes (at least one), One TV Show to Many Seasons (at least one)

Participation: Total for Season and TV Show in their respective relationships

8. Use Case Name: Show Creators

Description: Identifies the creators of TV shows. A TV show can have multiple creators (at least two to highlight collaborations), and a creator can be associated with multiple TV shows.

Entities Involved: TV Show, Creator

Cardinality: Many TV Shows to Many Creators (at least two creators per show)

Participation: Total for both TV Show and Creator

9. Use Case Name: Viewer Demographics

Description: Connects viewers to their preferred TV shows, allowing analysis of demographics. A viewer can have many preferred shows (at most 5 for simplicity), and a TV show can be preferred by many viewers.

Entities Involved: Viewer, TV Show

Cardinality: One Viewer to Many TV Shows (at most 5), Many Viewers to One TV Show

Participation: Partial for Viewer, Total for TV Show

10. Use Case Name: Show Tags

Description: Explores how users tag TV shows with descriptive keywords. A TV show can have many tags (at least three to reflect a range of content descriptors), and a tag can be associated with many TV shows.

Entities Involved: TV Show, Tag

Cardinality: Many TV Shows to Many Tags (at least three tags per show)

Participation: Total for TV Show, Partial for Tag

11. Use Case Name: Award Nominations and Wins

Description: This use case links TV shows to the awards they have been nominated for and won, allowing analysis of critical acclaim. A TV show can have multiple award nominations (at least one nomination) and can win multiple awards (at least one win is possible for acclaimed shows).

Entities Involved: TV Show, Award Nomination, Award Win

Cardinality: One TV Show to Many Award Nominations (at least one), One TV Show to Many Award Wins (at least one)

Participation: Total for TV Show, Partial for Award Nomination and Award Win

12. Use Case Name: Cast Members

Description: Associates actors with the TV shows on which they've appeared, facilitating searches based on cast. An actor can be part of many TV shows (at least two for actors with diverse portfolios), and a TV show can feature many actors.

Entities Involved: Actor, TV Show

Cardinality: Many Actors to Many TV Shows (at least two shows per actor)

Participation: Partial for Actor, Total for TV Show

13. Use Case Name: Streaming Platforms

Description: Connects TV shows to the streaming platforms where they are available. A TV show can be available on multiple platforms (at least one platform is mandatory), and a platform hosts many TV shows.

Entities Involved: TV Show, Streaming Platform

Cardinality: Many TV Shows to Many Streaming Platforms (at least one platform per show)

Participation: Total for TV Show, Partial for Streaming Platform

14. Use Case Name: User Reviews

Description: Links user-submitted reviews to TV shows, allowing for a rich dataset of viewer feedback. A TV show can have many reviews (at least one review for popular shows), and a review is associated with exactly one TV show.

Entities Involved: TV Show, User Review

Cardinality: One TV Show to Many User Reviews (at least one), One User Review to Exactly One TV Show

Participation: Total for TV Show, Partial for User Review

15. Use Case Name: Production Companies

Description: Identifies the production companies behind TV shows. A TV show may be produced by multiple companies (at least one company is required), and a production company can produce many TV shows.

Entities Involved: TV Show, Production Company

Cardinality: Many TV Shows to Many Production Companies (at least one company per show)

Participation: Total for TV Show, Partial for Production Company

Entities and Their Attributes

1. TV Show

- Attributes: **id** (PK), **name**, **number_of_seasons**, **number_of_episodes**, **original_language**, **vote_count**, **vote_average**, **popularity**, **first_air_date**, **last_air_date**, **status**
- Key: **id**

2. Genre

- Attributes: **id** (PK), **name**
- Key: **id**

3. **Language**

- Attributes: **id** (PK), **code**, **name**
- Key: **id**

4. **Country**

- Attributes: **id** (PK), **code**, **name**
- Key: **id**

5. **Network**

- Attributes: **id** (PK), **name**
- Key: **id**

6. **Season**

- Attributes: **id** (PK), **tv_show_id** (FK), **season_number**, **episode_count**
- Key: **id**

7. **Episode**

- Attributes: **id** (PK), **season_id** (FK), **episode_number**, **title**
- Key: **id**

8. **Actor**

- Attributes: **id** (PK), **name**, **birthdate**
- Key: **id**

9. **Creator**

- Attributes: **id** (PK), **name**, **birthdate**
- Key: **id**

10. **Streaming Platform**

- Attributes: **id** (PK), **name**
- Key: **id**

11. **User Review**

- Attributes: **id** (PK), **tv_show_id** (FK), **rating**, **review_text**, **review_date**
- Key: **id**

12. Production Company

- Attributes: **id** (PK), **name**
- Key: **id**

13. Tag

- Attributes: **id** (PK), **tag_text**
- Key: **id**

14. Award Nomination

- Attributes: **id** (PK), **tv_show_id** (FK), **award_name**, **year**
- Key: **id**

15. Award Win

- Attributes: **id** (PK), **tv_show_id** (FK), **award_name**, **year**
- Key: **id**

Relationships

1. **TV Show to Genre:** Many-to-Many (A TV show can have multiple genres; a genre can categorize multiple TV shows). Total participation on both sides.
2. **TV Show to Language:** Many-to-Many (A TV show can be available in multiple languages; a language can be spoken in multiple TV shows). Total participation for TV Show, partial for Language.
3. **TV Show to Country:** One-to-Many (A TV show is produced in exactly one country; a country can produce many TV shows). Total participation for TV Show, total for Country.
4. **TV Show to Network:** One-to-Many (A TV show is broadcast by exactly one network; a network can broadcast many TV shows). Total participation for TV Show, total for Network.
5. **TV Show to Season:** One-to-Many (A TV show consists of multiple seasons; a season belongs to exactly one TV show). Total participation for TV Show, total for Season.
6. **Season to Episode:** One-to-Many (A season consists of multiple episodes; an episode belongs to exactly one season). Total participation for Season, total for Episode.
7. **TV Show to Actor:** Many-to-Many (A TV show features multiple actors; an actor can appear in multiple TV shows). Partial participation for Actor, total for TV Show.
8. **TV Show to Creator:** Many-to-Many (A TV show has multiple creators; a creator can create multiple TV shows). Total participation for both TV Show and Creator.

9. **TV Show to Streaming Platform:** Many-to-Many (A TV show can be available on multiple streaming platforms; a platform hosts multiple TV shows). Total participation for TV Show, partial for Streaming Platform.
10. **TV Show to User Review:** One-to-Many (A TV show can have multiple reviews; a review is associated with exactly one TV show). Total participation for TV Show, partial for User Review.
11. **TV Show to Production Company:** Many-to-Many (A TV show can be produced by multiple companies; a company can produce multiple TV shows). Total participation for TV Show, partial for Production Company.
12. **TV Show to Tag:** Many-to-Many (A TV show can have multiple tags; a tag can be associated with multiple TV shows). Total participation for TV Show, partial for Tag.
13. **TV Show to Award Nomination:** One-to-Many (A TV show can have multiple award nominations; each nomination is associated with exactly one TV show). Total participation for TV Show, partial for Award Nomination.
14. **TV Show to Award Win:** One-to-Many (A TV show can win multiple awards; each win is associated with exactly one TV show). Total participation for TV Show, partial for Award Win.
15. **Actor to Episode:** Many-to-Many (An actor can appear in multiple episodes across different TV shows; an episode can feature multiple actors). Partial participation for both Actor and Episode.

ER Diagram

