**Advanced SQL – Reinforcement Project – IMDB Dataset**

The dataset provided is a simplified version of the IMDb database, structured to capture essential information about movies, their genres, actors, directors, ratings, and more. This database consists of several tables that contain various details such as:

1. **Movie**: Contains basic information about each movie, including title, release year, duration, country, income, languages, and production companies.

2. **Genre:** Describes the genres associated with each movie.

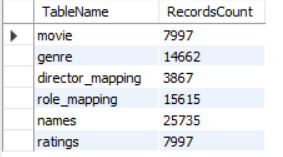
3. **Director Mapping:** Maps movies to their directors.

4. **Role Mapping:** Maps actors/actresses to movies and specifies the role category (e.g., actor, director, producer).

5. **Names:** Stores information about people (actors, directors, etc.), including their birthdates, heights, and known movies.

6. **Ratings:** Contains ratings information for movies, including the average rating, total votes, and median rating.

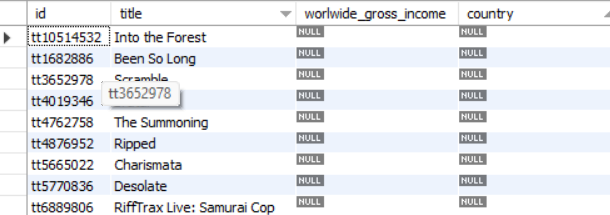
1. **Count the total number of records in each table of the database**

****

**Explanation:**

* **Purpose**: This query returns the total number of records in each table of the database.
* **Use Case**: Helps database administrators or analysts to get an overview of the size of each table in the database.
* **Expected Output**: A table showing the name of each table (Table Name) and the count of records (Records Count).

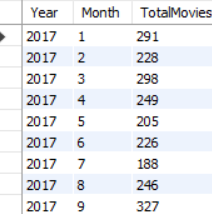
1. **Identify which columns in the movie table contain null values**

****

**Explanation:**

* **Purpose**: This query checks for null values in the world wide\_gross\_income column of the movie table.
* **Use Case**: Helps identify missing or incomplete data in the specified column.
* **Expected Output**: Lists the rows where the worlwide\_gross\_income is null, and counts the number of null values in the column.

1. **Determine the total number of movies released each year and month**

****

**Explanation:**

* **Purpose**: This query groups the movies by year and month to calculate how many movies were released each month.
* **Use Case**: Useful for trend analysis, to track the number of movie releases over time.
* **Expected Output**: A table showing the year, month, and total number of movies released in that period.

1. **Count movies produced in either the USA or India in 2019**

**A screenshot of a number

Description automatically generated**

**Explanation:**

* **Purpose**: This query counts the number of movies produced in either the USA or India in the year 2019.
* **Use Case**: Helps in understanding movie production trends in different countries.
* **Expected Output**: A table showing the country and the total count of movies produced in 2019.

1. **List the unique genres in the dataset and count movies exclusively belonging to one genre**

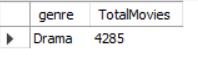
**A screenshot of a computer

Description automatically generated**

**Explanation:**

* Purpose: This query lists genres and counts the number of movies in each genre that have more than one movie.
* Use Case: Useful to identify the distribution of movies across genres.
* Expected Output: A list of genres and the corresponding number of movies.

1. **Determine the genre with the highest total number of movies**

****

**Explanation:**

* Purpose: This query identifies the genre with the highest number of movies.
* Use Case: Helps in understanding which genre has the most movies in the database.
* Expected Output: A single genre with the highest movie count.

1. **Calculate the average movie duration for each genre**

**A screenshot of a computer

Description automatically generated**

**Explanation:**

* Purpose: This query calculates the average duration of movies for each genre.
* Use Case: Helps analyze the length of movies in different genres.
* Expected Output: A list of genres with their corresponding average movie durations.

1. **Identify actors/actresses who have appeared in more than three movies with an average rating below 5**

**A screenshot of a computer screen

Description automatically generated**

**Explanation:**

* Purpose: This query identifies actors/actresses who have acted in more than three movies with an average rating below 5.
* Use Case: Useful for finding actors/actresses associated with poorly-rated films.
* Expected Output: A list of actors/actresses with their movie count and average rating.

1. **Find the minimum and maximum values for each column in the ratings table**

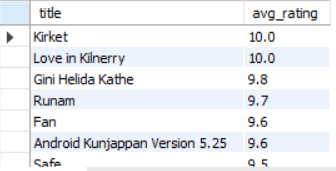
**A screenshot of a computer

Description automatically generated**

**Explanation:**

* Purpose: This query finds the minimum and maximum values for average rating, total votes, and median rating in the ratings table.
* Use Case: Helps in analyzing the range of ratings and votes.
* Expected Output: A table showing the minimum and maximum values for each of the columns.

1. **List the top 10 movies based on their average rating**

****

**Explanation:**

* Purpose: This query lists the top 10 movies based on their average rating.
* Use Case: Helps to identify the highest-rated movies.
* Expected Output: A list of top 10 movies with their average ratings.

**11). Summarize the ratings table by grouping movies based on their median ratings**

**A screenshot of a computer

Description automatically generated**

**Explanation:**

* Purpose: This query summarizes the ratings table by grouping movies based on their median ratings.
* Use Case: Useful to see how many movies have specific median ratings.

**12). How many movies, released in March 2017 in the USA within a specific genre, had more than 1,000 votes?**

A screenshot of a computer

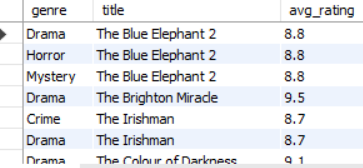
Description automatically generated

**Explanation:**

 **Purpose**: The genre of the movies.

 **Use Case**: The count of movies that match the criteria for each genre.

**13. Find movies from each genre that begin with the word "The" and have an average rating greater than 8**

****

**Explanation:**

* Purpose: This query retrieves movies from each genre that start with the word “The” and have an average rating greater than 8.
* Use Case: Helps identify highly rated movies with titles starting with "The" across different genres.
* Expected Output: A list of movies from different genres that start with "The" and have an average rating higher than 8.

**14. Count how many movies released between April 1, 2018, and April 1, 2019, received a median rating of** **8**

****

**Explanation:**

* **Purpose**: This query counts the number of movies released between April 1, 2018, and April 1, 2019, that received a median rating of 8.
* **Use Case**: Helps identify high-rated movies within a specific time frame.
* **Expected Output**: The count of movies that meet the criteria of being released within the given period and having a median rating of 8.

**15. Do German movies receive more votes on average than Italian movies?**

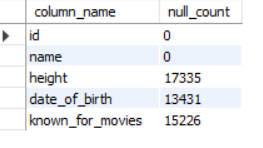
**A screenshot of a computer

Description automatically generated**

**Explanation:**

* **Purpose**: This query compares the average total votes between German and Italian movies.
* **Use Case**: Useful for comparing the popularity (in terms of votes) of German and Italian movies.
* **Expected Output**: A comparison showing the average number of votes for movies from Germany and Italy.

**16. Identify the columns in the names table that contain null values**

****

**Explanation:**

* **Purpose**: This query identifies the columns in the names table that contain null values by checking each relevant column (id, name, height, date\_of\_birth, known\_for\_movies).
* **Use Case**: Useful for data cleaning or analysis to check for missing values in important fields.
* **Expected Output**: A list of columns and the count of null values in each.

**17. Identify the top two actors whose movies have a median rating of 8 or higher**

**A screenshot of a computer

Description automatically generated**

**Explanation:**

* **Purpose**: This query finds the top two actors who have acted in movies with a median rating of 8 or higher.
* **Use Case**: Helps identify actors who are associated with highly-rated movies.
* **Expected Output**: The top two actors with the highest number of movies that have a median rating of 8 or higher.

**18. Identify the top three production companies based on the total number of votes their movies received**

**A screenshot of a computer

Description automatically generated**

**Explanation:**

* **Purpose**: This query identifies the top three production companies based on the total number of votes their movies received.
* **Use Case**: Useful for identifying the most popular production companies based on user votes.
* **Expected Output**: A list of the top three production companies with their corresponding total vote counts.

**19. How many directors have worked on more than three movies?**

**A close-up of a white background

Description automatically generated**

**Explanation:**

* **Purpose**: This query counts how many directors have worked on more than three movies.
* **Use Case**: Useful to find prolific directors who have contributed to a larger number of movies.
* **Expected Output**: A single value representing the count of directors who have directed more than three movies.

**20. Calculate the average height of actors and actresses separately**

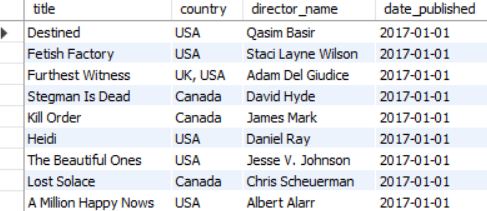
**A screenshot of a computer

Description automatically generated**

**Explanation:**

* **Purpose**: This query calculates the average height of actors and actresses separately.
* **Use Case**: Helps in analyzing physical attributes (like height) of actors and actresses.
* **Expected Output**: A table showing the average height for both actors and actresses.

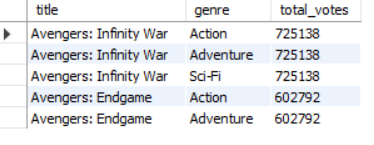
**21. List the 10 oldest movies in the dataset along with their title, country, and director**

****

**Explanation:**

* **Purpose**: This query lists the 10 oldest movies, along with their title, country, and director.
* **Use Case**: Useful for identifying the oldest movies in the dataset along with their related details.
* **Expected Output**: A list of the 10 oldest movies, including the title, country, director, and release date.

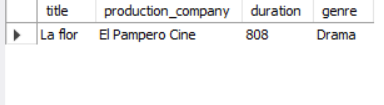
**22. List the top 5 movies with the highest total votes, along with their genres**

****

**Explanation:**

* **Purpose**: This query lists the top 5 movies with the highest total votes and includes their genres.
* **Use Case**: Helps identify the most popular movies based on the total number of votes received.
* **Expected Output**: A list of the top 5 movies with the highest total votes, along with their genre.

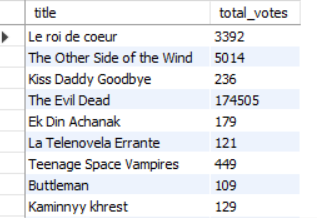
**23. Identify the movie with the longest duration, along with its genre and production company**

****

**Explanation:**

* **Purpose**: This query identifies the movie with the longest duration, along with its genre and production company.
* **Use Case**: Useful for finding the longest movie in the dataset and related details.
* **Expected Output**: A single movie with the longest duration, along with its genre and production company.

**24. Determine the total number of votes for each movie released in 2018**

****

**Explanation:**

* **Purpose**: This query calculates the total number of votes for each movie released in the year 2018.
* **Use Case**: Useful for analyzing the popularity of movies released in 2018 based on votes.
* **Expected Output**: A list of movies released in 2018 with their corresponding total vote count.

**25. Find the most common language in which movies were produce**

****

**Explanation:**

* **Purpose:** This query identifies the most common language in which movies were produced.
* **Use Case:** Useful for analyzing language trends in the production of movies.
* **Expected Output:** A single language with the highest count of movies produced in that language.

Top of Form

Bottom of Form