Executive Summary

SQL – RSVP Movies Case Study

The provided SQL queries cover a wide range of analyses on a movie-related database. Let's summarize the key insights derived from these queries.

1. Table and Data Overview:

The initial queries provide a comprehensive view of the database by counting the number of records in various tables (DIRECTOR_MAPPING, GENRE, MOVIE, NAMES, RATINGS, ROLE_MAPPING). This helps understand the scale and distribution of data across different entities.

2. Movie Data Quality:

The queries examining the movie table focus on assessing data quality. They check for NULL values in critical columns such as id, title, year, date_published, duration, country, worldwide_gross_income, languages, and production_company. These checks are crucial for ensuring data integrity and identifying potential areas for improvement.

3. Movie Counts and Distribution:

Queries related to movie counts and distribution by year and month offer insights into the temporal trends of movie production. The counts provide

an overview of how many movies were released each year and each month, helping to identify any patterns or anomalies.

4. Genre Analysis:

The SQL queries delve into the genres of movies, exploring both the most popular genre and genres with specific characteristics. This includes finding the genre with the highest movie count, movies with only one genre, and average movie duration by genre.

5. Ratings and Movie Success:

Queries related to ratings analyze the success of movies based on various criteria. This includes finding movies with the highest average ratings, assessing the impact of genres on ratings, and exploring the distribution of movies based on median ratings.

6. Production Company Performance:

The SQL queries explore the performance of production companies, focusing on those with hit movies. Production companies are ranked based on the number of highly-rated movies they have produced. Additionally, there is an analysis of production companies with the most successful movies based on total votes.

7. Actor and Actress Analysis:

The SQL queries provide insights into the performance of actors and actresses. This includes identifying actors with the highest number of movies and actresses with successful careers in specific genres. There's also an examination of actresses in Hindi movies with a focus on their average ratings.

8. Time Analysis for Directors:

The queries related to directors offer a temporal analysis of movie releases. This includes identifying directors with the highest number of movies, their average inter-movie days, and total movie duration. These metrics provide a glimpse into the productivity and efficiency of directors.

9. Thriller Movies and Average Ratings:

A specific analysis is conducted on thriller movies, categorizing them based on average ratings. This helps in understanding the distribution of ratings within the thriller genre.

10. Moving Average Duration by Genre:

The final query calculates the moving average duration for movies across different genres. This dynamic analysis provides insights into how the average duration of movies changes over time within each genre.

In summary, these SQL queries offer a comprehensive exploration of the movie-related database. The analyses cover data quality checks, temporal trends, genre characteristics, ratings distribution, production company performance, and insights into the careers of actors, actresses, and directors. The results provide valuable information for stakeholders in the movie industry, helping them make informed decisions and understand the dynamics of movie production, success factors, and audience preferences.

Submitted By :

Praasuk Jain

Sujata Pawar

Nidhi Singh