TV Series Performance & Reviewer Behavior Analysis

by Bisma Putra Atallah

https://www.linkedin.com/in/bismaputraatallah/

github.com/bismaputraatallah/Portofolio-Data-Analytics

Overview

The TV series dataset in this project consists of television series information (title, release year, and genre), reviewer data, and rating-based review records, all connected through a **relational database structure**.

The analysis in this project focuses on **TV series performance and reviewer behavior**, including average ratings, review volume, popularity versus quality, and genre-level insights using SQL-based analytical techniques.

!Disclaimer!

- This analysis is based on the The Ultimate MySQL Bootcamp: Go from SQL Beginner to Expert Udemy course dataset.
- The TV series dataset is fictional and used for educational purposes, with minor improvisations by the author.

Dataset Overview

- Key Tables:
  - Series
  - Reviewers
  - Reviews
- Relevan Fields:
  - Series Title
  - Release Year
  - Genre
  - Rating
  - Series ID
  - Reviewer ID

Project Goals

- To analyze TV series performance based on ratings and popularity.
- To examine reviewer behavior and rating patterns.
- To generate actionable insights from TV series review data.

Methodology

Analysis Approach

- Data exploration and analysis using SQL on a TV series dataset.
- Insight generation on series performance, popularity, genre, and reviewer behavior.
- Tools used:
  - MySQL – for query
  - Canva – for reporting

Key SQL Queries

- Average rating per TV series
- Highest and lowest rated series
- Most active and inactive reviewers
- Series performance by genre
- Popularity vs quality comparison
- Top-rated series within each genre
- Reviewer genre diversity analysis

Data Analysis and Insight

1. Identify TV series with above-average ratings

```sql
SELECT
    s.id,
    title,
    ROUND(AVG(rating), 2) AS avg_rating,
    (SELECT ROUND(AVG(rating), 2) FROM reviews) AS overall_average
FROM series s
    JOIN reviews r ON s.id = r.series_id
GROUP BY 1,2
HAVING AVG(rating) > (SELECT ROUND(AVG(rating), 2) FROM reviews)
ORDER BY 3 DESC;
```

Query Explanation

- This query evaluates **TV series performance** by calculating the average rating for each series and comparing it to the **overall average rating (8.03)** as a benchmark.
- The results highlight **higher- and moderate-performing** series, showing differences in audience reception among TV series that all perform above the overall average.

Insight

- Top-rated TV series (Halt and Catch Fire, Fargo, and Breaking Bad) clearly outperform other above-average series, while the remaining titles show **more moderate but still strong audience reception** compared to the overall average rating.

id	title	avg_rating	overall_average
10	Halt and Catch Fire	9.90	8.03
7	Fargo	9.40	8.03
5	Breaking Bad	9.36	8.03
14	Stranger Things	8.77	8.03
8	Freaks and Geeks	8.60	8.03
1	Archer	8.12	8.03
6	Curb Your Enthusiasm	8.12	8.03
2	Arrested Development	8.08	8.03

Data Analysis and Insight

2. Compare highest vs lowest average-rated TV series

```
(SELECT
    'hisghest-rating' AS series_type,
    title,
    ROUND(AVG(rating), 2) AS avg_rating
FROM reviews r
    JOIN series s ON r.series_id = s.id
GROUP BY 1,2
ORDER BY 3 DESC
LIMIT 1)
UNION
(SELECT
    'lowest-rating' AS series_type,
    title,
    ROUND(AVG(rating), 2) AS avg_rating
FROM reviews r
    JOIN series s ON r.series_id = s.id
GROUP BY 1,2
ORDER BY 3 ASC
LIMIT 1);
```

Query Explanation

- This query calculates the **average rating for each TV series** and retrieves the **highest- and lowest-rated series** to compare performance extremes based on audience ratings.

series_type	title	genre	avg_rating
hisghest-rating	Halt and Catch Fire	Drama	9.90
lowest-rating	General Hospital	Drama	5.38

Insight

- Halt and Catch Fire (Drama) stands out as the **highest-rated series** with an average rating of **9.90**, while General Hospital (Drama) records the **lowest average rating** at **5.38**, highlighting a significant difference in audience reception within the same genre.

Data Analysis and Insight

3. Identify the release year with the highest number of reviews & Analyze the average number of reviews per release year

```sql
SELECT
    s.released_year,
    COUNT(r.id) AS num_reviews
FROM series s
LEFT JOIN reviews r ON s.id = r.series_id
GROUP BY 1
ORDER BY 2 DESC;
```

released_year	num_reviews
2014	8
2009	5
2003	5
2011	5
2008	5
2000	5
1963	5
1999	4
2016	3
1989	2
2007	0

```sql
SELECT
    s.released_year,
    ROUND(COUNT(r.id) * 1.0 / COUNT(DISTINCT s.id), 2) AS average_reviews_per_year
FROM series s
LEFT JOIN reviews r ON s.id = r.series_id
GROUP BY 1
ORDER BY 2 DESC;
```

released_year	average_reviews_per_year
1963	5.00
2003	5.00
2008	5.00
2009	5.00
2011	5.00
1999	4.00
2016	3.00
2014	2.67
2000	2.50
1989	2.00
2007	0.00

Query Explanation

- This query **groups TV series** by release year and **counts the total number of reviews** to identify **which year received the most reviews**. It also **calculates the average number of reviews per series** in each release year by dividing total reviews by the number of distinct series.

Insight

- The year **2014** has the **most reviews**, showing **high audience engagement**, while **earlier years** have **fewer series but higher average reviews**, indicating **more consistent attention per series**.

Data Analysis and Insight

4. Evaluate reviewer activity and rating behavior

```sql
SELECT
    first_name,
    last_name,
    COUNT(rating) AS count,
    IFNULL(MIN(rating), 0) AS min,
    IFNULL(MAX(rating), 0) AS max,
    ROUND(IFNULL(AVG(rating), 0), 2) AS average,
    IF(COUNT(rating) > 0,
        'ACTIVE',
        'INACTIVE') AS status
FROM
    reviewers
        LEFT JOIN
    reviews ON reviewers.id = reviews.reviewer_id
GROUP BY first_name , last_name;
```

Query Explanation

- This query summarizes each reviewer's activity by counting how many ratings they have given.
- It calculates the **minimum, maximum, and average** rating for each reviewer.
- Reviewers are labeled **ACTIVE** if they have at least one rating and **INACTIVE** if they have none.

reviewers	count	max	min	average	statues
Thomas Stoneman	5	9.5	7.0	8.02	ACTIVE
Wyatt Skaggs	9	9.3	5.5	7.80	ACTIVE
Kimbra Masters	9	9.0	6.8	7.99	ACTIVE
Domingo Cortes	10	9.1	5.8	7.83	ACTIVE
Colt Steele	10	9.9	4.5	8.77	ACTIVE
Pinkie Petit	4	8.8	4.3	7.25	ACTIVE
Marlon Crafford	0	0.0	0.0	0.00	INACTIVE

Query Explanation

- Active reviewers show different rating behaviors, with varying averages and rating ranges.
- Inactive reviewers have no ratings, which explains why their values appear as zero.
- This helps identify which reviewers are contributing most to the review data and how consistent their ratings are.

Data Analysis and Insight

```sql
WITH cte_count_avg AS (
    SELECT
        s.title,
        COUNT(r.id) AS popularity,
        ROUND(AVG(rating), 2) AS quality
    FROM series s
        JOIN reviews r ON s.id = r.series_id
    GROUP BY 1),
    cte_rank AS (
    SELECT
        title,
        popularity,
        quality,
        DENSE_RANK() OVER (ORDER BY popularity DESC) AS rank_popularity,
        DENSE_RANK() OVER (ORDER BY quality DESC) AS rank_quality
    FROM cte_count_avg)
SELECT
    title,
    popularity,
    quality,
    CASE WHEN rank_popularity <= 3 AND rank_quality <= 3 THEN 'hight popularity & quality'
        WHEN rank_popularity > 3 AND rank_quality <= 3 THEN 'low  popularity but hight quality'
        WHEN rank_popularity <= 3 AND rank_quality > 3 THEN 'hight popularity but low quality'
    ELSE 'low popularity & quality' END AS category
FROM cte_rank
ORDER BY popularity DESC;
```

5. Classify TV series based on popularity and quality

Query Explanation

- The query first counts the total number of reviews for each TV series to measure its **popularity** and calculates the **average rating** to represent its quality.
- It then ranks all series separately by popularity and by quality using **DENSE_RANK**, where higher values indicate better performance.
- Based on these rankings, each series is classified into one of four categories: **high popularity & high quality, high popularity but low quality, low popularity but high quality, or low popularity & low quality**.
- Finally, the results are displayed with the series title, popularity score, quality score, and its assigned category, ordered by popularity.

Data Analysis and Insight

```sql
WITH cte_count_avg AS (
    SELECT
        s.title,
        COUNT(r.id) AS popularity,
        ROUND(AVG(rating), 2) AS quality
    FROM series s
        JOIN reviews r ON s.id = r.series_id
    GROUP BY 1),
    cte_rank AS (
    SELECT
        title,
        popularity,
        quality,
        DENSE_RANK() OVER (ORDER BY popularity DESC) AS rank_popularity,
        DENSE_RANK() OVER (ORDER BY quality DESC) AS rank_quality
    FROM cte_count_avg)
SELECT
    title,
    popularity,
    quality,
    CASE WHEN rank_popularity <= 3 AND rank_quality <= 3 THEN 'hight popularity & quality'
        WHEN rank_popularity > 3 AND rank_quality <= 3 THEN 'low  popularity but hight quality'
        WHEN rank_popularity <= 3 AND rank_quality > 3 THEN 'hight popularity but low quality'
    ELSE 'low popularity & quality' END AS category
FROM cte_rank
ORDER BY popularity DESC;
```

5. Classify TV series based on popularity and quality

title	popularity	quality	category
Breaking Bad	5	9.36	hight popularity & quality
Archer	5	8.12	hight popularity but low quality
Curb Your Enthusiasm	5	8.12	hight popularity but low quality
Arrested Development	5	8.08	hight popularity but low quality
Bojack Horseman	5	7.94	hight popularity but low quality
Bob's Burgers	5	7.52	hight popularity but low quality
General Hospital	5	5.38	hight popularity but low quality
Freaks and Geeks	4	8.60	hight popularity but low quality
Stranger Things	3	8.77	hight popularity but low quality
Fargo	2	9.40	low popularity but hight quality
Seinfeld	2	7.60	low popularity & quality
Halt and Catch Fire	1	9.90	low popularity but hight quality

Query Explanation

- Series like **Breaking Bad** fall into the **high popularity & high quality category**, meaning they are both widely reviewed and highly rated.
- Some series, such as **Halt and Catch Fire**, have **low popularity but high quality**, indicating strong ratings despite fewer reviewers.
- Meanwhile, shows like **General Hospital** are **high in popularity but low in quality**, suggesting wide audience reach but weaker viewer satisfaction.

Data Analysis and Insight

6. Genre-Level Performance and Top Reviewer Analysis

```sql
WITH genre_summary AS (
    SELECT
        s.genre,
        COUNT(DISTINCT s.id) AS num_series,
        COUNT(r.id) AS total_reviews,
        ROUND(AVG(r.rating), 2) AS avg_rating
    FROM series s
        LEFT JOIN reviews r ON s.id = r.series_id
    GROUP BY 1),
    reviewer_rank AS (
        SELECT
            s.genre,
            r2.reviewer_id,
            CONCAT(rv.first_name, ' ', rv.last_name) AS reviewer_name,
            COUNT(r2.id) AS review_count,
            ROW_NUMBER() OVER (PARTITION BY s.genre ORDER BY COUNT(r2.id) DESC) AS rank_in_genre
        FROM series s
            JOIN reviews r2 ON s.id = r2.series_id
            JOIN reviewers rv ON r2.reviewer_id = rv.id
        GROUP BY 1, 2, 3)
SELECT
    g.genre,
    g.num_series,
    g.total_reviews,
    g.avg_rating,
    review_count
FROM genre_summary g
    JOIN reviewer_rank rr ON g.genre = rr.genre
WHERE rr.rank_in_genre = 1
ORDER BY g.avg_rating DESC;
```

Query Explanation

- The query summarizes each **genre** by counting the number of TV series, total reviews, and calculating the average rating.
- It then identifies the **top reviewer in each genre** based on who wrote the most reviews.
- Finally, both results are combined to show genre performance along with its most active reviewer.

genre	num_series	total_reviews	avg_rating	review_count
Comedy	6	16	8.16	4
Drama	5	16	8.04	4
Animation	3	15	7.86	3

Insight

- Comedy and Drama have the highest total reviews, showing strong audience engagement.
- Comedy has the highest average rating, indicating better overall viewer satisfaction.
- Each genre is influenced by a key active reviewer, suggesting reviewer activity plays an important role in shaping genre insights.

Insight Summary & Recommendation

Insight	Recommendation	Priority
Certain TV series have **high quality (ratings) but low popularity**, meaning they are underexposed.	Increase promotion and visibility for high-quality but low-popularity series to maximize their potential audience.	**High**
Popular series do not always have the best ratings, showing popularity is driven by exposure rather than quality.	Balance content strategy by combining popular series with high-quality recommendations to improve overall viewer satisfaction.	**High**
Specific release years (e.g., 2014) generate the **highest number of reviews**, indicating strong audience engagement.	Focus content acquisition and promotion on patterns from high-engagement release years.	**Medium**
Drama and Comedy dominate in both **number of reviews and average ratings**.	Prioritize investment and marketing efforts in Drama and Comedy genres.	**High**
A small group of reviewers contributes a large portion of reviews, while others remain inactive.	Encourage inactive reviewers through incentives and recognize active reviewers to sustain engagement.	**Medium**

Insight Summary & Recommendation

Final Business Takeway

The TV series platform can improve **audience engagement** and **content performance** by strategically promoting **high-quality but underexposed series**, prioritizing **high-performing genres such as Drama and Comedy**, and leveraging reviewer engagement patterns. By **balancing popularity with quality** and e**ncouraging consistent reviewer participation**, the platform can enhance **viewer satisfaction, optimize content value,** and drive **sustainable growth**.

THANK YOU

https://www.linkedin.com/in/bismaputraatallah/

github.com/bismaputraatallah/Portofolio-Data-Analytics