

Analyzing Netflix Shows & Movies with SQL

Objective

To help Netflix analyze and draw meaningful insights from a large dataset of shows and movies using SQL. The goal is to identify content trends, viewer preferences, and top-performing media to improve content strategy and user engagement.

Dataset Overview

- Source: Kaggle
- Files Used: `titles.csv`, `credits.csv`
- Total Rows: ~82,000 combined
- Columns Include:
 - `title`, `type`, `release_year`, `imdb_score`, `genres`, `age_certification`, etc.
 - `credits.csv` includes actor/director roles and their associations with shows/movies.

Dataset Used : Netflix TV Shows and Movies

Tools Used

- **SQLite:** To create tables, import data, and write queries.
- **VS Code + SQLite extension:** SQL editing and terminal-based querying.

Q1.What were the top 10 movies according to IMDB score?

```
SELECT title,  
type,  
imdb_score  
FROM shows_movies.titles  
WHERE imdb_score >= 8.0  
AND type = 'MOVIE'  
ORDER BY imdb_score DESC  
LIMIT 10
```

title	type	imdb_score
Chhota Bheem & Krishna vs Zimbara	MOVIE	9.1
Major	MOVIE	9.1
David Attenborough: A Life on Our Planet	MOVIE	8.9
C/o Kancharapalem	MOVIE	8.9
Inception	MOVIE	8.8
Forrest Gump	MOVIE	8.8
Chhota Bheem & Krishna in Mayanagari	MOVIE	8.7
Bo Burnham: Inside	MOVIE	8.7
A Lion in the House	MOVIE	8.7
Chhota Bheem Neeli Pahaadi	MOVIE	8.7

2)What were the top 10 shows according to IMDB score?

```
SELECT title,  
type,  
imdb_score  
FROM shows_movies.titles  
WHERE imdb_score >= 8.0  
AND type = 'SHOW'  
ORDER BY imdb_score DESC  
LIMIT 10
```

title	type	imdb_score
#ABtalks	SHOW	9.6
Khawatir	SHOW	9.5
Breaking Bad	SHOW	9.5
Our Planet	SHOW	9.3
Avatar: The Last Airbender	SHOW	9.3
Reply 1988	SHOW	9.2
Kota Factory	SHOW	9.1
The Last Dance	SHOW	9.1
My Mister	SHOW	9.1
Okupas	SHOW	9

3)What were the bottom 10 movies according to IMDB score?

```
SELECT title,  
type,  
imdb_score  
FROM shows_movies.titles  
WHERE type = 'MOVIE'  
ORDER BY imdb_score ASC  
LIMIT 10
```

title	type	imdb_score
Aerials	MOVIE	1.5
Me Against You: Mr. S's Vendetta	MOVIE	1.6
Himmatwala	MOVIE	1.7
Kyaa Kool Hain Hum 3	MOVIE	1.9
FRED 3: Camp Fred	MOVIE	2
Hajwala: The Missing Engine	MOVIE	2
Dostana	MOVIE	2.1
Time to Dance	MOVIE	2.2
B. A. Pass 2	MOVIE	2.2
Nothing to Lose	MOVIE	2.3

4) What were the bottom 10 shows according to IMDB score?

```
SELECT title,  
type,  
imdb_score  
FROM shows_movies.titles  
WHERE type = 'SHOW'  
ORDER BY imdb_score ASC  
LIMIT 10;
```

title	type	imdb_score
He's Expecting	SHOW	2
Thomas & Friends: All Engines Go!	SHOW	2
Hype House	SHOW	2.1
A House of Blocks	SHOW	2.3
Until Dawn	SHOW	2.4
The Goop Lab	SHOW	2.5
Byron Baes	SHOW	2.6
First Class	SHOW	2.8
Bonus Family	SHOW	2.9
Richie Rich	SHOW	3

5)Count of movies and shows in each decade

```
SELECT CONCAT(FLOOR(release_year / 10) * 10, 's') AS decade,  
        COUNT(*) AS movies_shows_count  
FROM shows_movies.titles  
WHERE release_year >= 1940  
GROUP BY CONCAT(FLOOR(release_year / 10) * 10, 's')  
ORDER BY decade;
```

decade	movies_shows_count
1940s	1
1950s	5
1960s	8
1970s	18
1980s	52
1990s	121
2000s	369
2010s	3304
2020s	1972

6)What were the average IMDB and TMDB scores for each age certification for shows and movies?

```
SELECT DISTINCT age_certification,  
ROUND(AVG(imdb_score),2) AS avg_imdb_score,  
ROUND(AVG(tmdb_score),2) AS avg_tmdb_score  
FROM shows_movies.titles  
GROUP BY age_certification  
ORDER BY avg_imdb_score DESC
```

age_certification	avg_imdb_score
TV-14	6.71
TV-MA	6.62
TV-PG	6.37
TV-Y7	6.32
PG-13	6.2
TV-G	6.01
R	6
TV-Y	6
PG	5.96
NC-17	5.76
N/A	5.58
G	5.09

7)What were the 5 most common age certifications for movies?

```
SELECT age_certification,  
COUNT(*) AS certification_count  
FROM shows_movies.titles  
WHERE type = 'Movie'  
AND age_certification != 'N/A'  
GROUP BY age_certification  
ORDER BY certification_count DESC  
LIMIT 5;
```

age_certification	certification_count
R	556
PG-13	451
PG	233
G	124
NC-17	16

8)Which genres had the most movies?

```
SELECT genres,  
COUNT(*) AS title_count  
FROM shows_movies.titles  
WHERE type = 'Movie'  
GROUP BY genres  
ORDER BY title_count DESC  
LIMIT 10;
```

genres	title_count
['comedy']	384
['documentation']	230
['drama']	224
['comedy', 'documentation']	100
['comedy', 'drama']	84
['drama', 'romance']	76
['drama', 'comedy']	65
['comedy', 'romance']	60
['romance', 'comedy']	48
['comedy', 'drama', 'romance']	44

9)Which genres had the most shows?

```
SELECT genres,  
COUNT(*) AS title_count  
FROM shows_movies.titles  
WHERE type = 'Show'  
GROUP BY genres  
ORDER BY title_count DESC  
LIMIT 10;
```

genres	title_count
['reality']	113
['drama']	104
['comedy']	100
['documentation']	99
['comedy', 'drama']	51
['drama', 'romance']	48
['drama', 'comedy']	44
['documentation', 'crime']	35
['animation']	30
[]	25

10) What were the top 3 most common genres?

```
SELECT t.genres,  
COUNT(*) AS genre_count  
FROM shows_movies.titles AS t  
WHERE t.type = 'Movie'  
GROUP BY t.genres  
ORDER BY genre_count DESC  
LIMIT 3;
```

genres	genre_count
['comedy']	484
['documentation']	329
['drama']	328

Conclusion

By using SQL to query and explore Netflix's vast dataset, valuable trends and patterns were identified that could help Netflix make data-driven decisions for content curation. The approach provides a scalable way to analyze entertainment data and can be extended to other streaming platforms.

THANK YOU