



NETFLIX ORIGINALS DATA ANALYSIS USING MYSQL

(PROJECT REPORT)

PREPARED BY:-

Amandeep Kaur

BBA STUDENT



Introduction

Netflix Originals have become a significant part of the global entertainment landscape, offering a diverse range of movies, series, and documentaries. Understanding the trends, viewer preferences, and quality of content within these Originals is crucial for content creators, marketers, and data analysts. By analyzing key attributes such as IMDb scores, genres, and runtimes, this project aims to uncover valuable insights into the performance and reception of Netflix Original content. Leveraging SQL for data querying and analysis provides a practical approach to extracting meaningful patterns from the datasets, which can inform strategic decisions and improve content curation.

Objective

The objective of this project is to analyze the Netflix Originals dataset using MySQL queries to extract meaningful insights and trends. Through this analysis, various SQL operations such as GROUP BY, HAVING, ORDER BY, LIMIT, JOINS, WINDOW FUNCTIONS, and SUBQUERIES are applied to understand viewer preferences, genre performance, and overall content quality.

Dataset Description:-

- . **Netflix_Originals:** Contains detailed information about Netflix Originals including title, genre ID, runtime, IMDb score, language, premiere date, etc.
- . **Genre_Details:** Contains genre ID and genre name mappings.

Methodology

This project uses SQL-based data analysis techniques to explore and derive insights from the Netflix Originals dataset. The following approach was taken:

1. **Data Preparation:**

The two datasets—Netflix_Originals and Genre_Details—were joined using the genre_id field to combine title information with genre names.

2. **Aggregation and Grouping:**

Aggregate functions such as AVG() and COUNT() were used along with GROUP BY to calculate metrics like average IMDb scores and count of titles per genre.

3. **Filtering Data:**

The HAVING clause was applied to filter grouped data based on conditions like minimum average IMDb score or minimum count of titles meeting a score threshold.

4. **Ordering and Limiting Results:**

To identify top entries, the ORDER BY clause was used to sort data by IMDb scores or runtime, and LIMIT was applied to restrict output to the top N results.

5. **Ranking Within Groups:**

Window functions such as RANK() or ROW_NUMBER() partitioned by genre were used to rank Netflix Originals based on their IMDb scores within each genre.

6. **Subqueries:**

Subqueries helped compare individual IMDb scores against the overall average IMDb score for all Netflix Originals, enabling identification of above-average titles.

This structured SQL approach allowed for a comprehensive analysis of Netflix Originals' content quality, genre performance, and runtime trends.

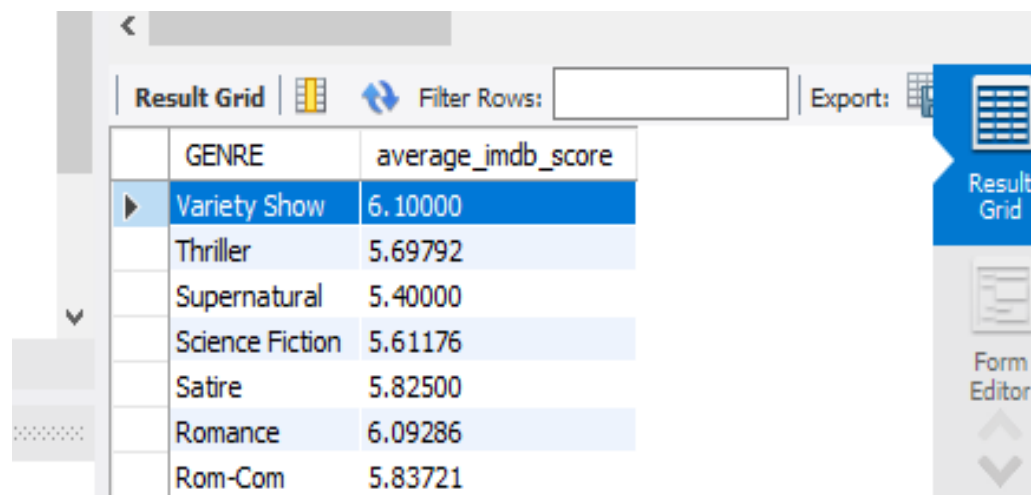
Analysis and results

!.What are the average IMDb scores for each genre of Netflix Originals?

QUERY:-

```
SELECT g.GENRE,AVG(n.IMDBSCORE) AS average_imdb_score
FROM netflix_originals n
JOIN genre_details g ON n.GENREID = g.GENREID
GROUP BY g.GENRE
ORDER BY average_imdb_score DESC;
```

RESULT:-



The screenshot shows a database query result grid with the following data:

GENRE	average_imdb_score
Variety Show	6.10000
Thriller	5.69792
Supernatural	5.40000
Science Fiction	5.61176
Satire	5.82500
Romance	6.09286
Rom-Com	5.83721

The interface includes a 'Result Grid' tab, a 'Filter Rows' input field, an 'Export' button, and a 'Form Editor' button on the right side.

2.Which genres have an average IMDb score higher than 7.5?

QUERY:-

```
SELECT g.genre, avg (n.IMDBSCORE)AS Highest_average_imdb_score
FROM netflix_originals n
JOIN genre_details g ON n.GENREID = g.GENREID
GROUP BY g.GENRE
HAVING avg (n.IMDBSCORE)>7.5;
```

RESULT:-

9

<		
Result Grid		
Filter Rows:		
Export:		
	genre	Highest_average_imdb_score
▶	Concert Film	7.63333



3. List Netflix Original titles in descending order of their IMDb scores?

QUERY:-

```
SELECT TITLE,IMDBSCORE FROM NETFLIX_ORIGINALS
```

```
ORDER BY IMDBSCORE DESC;
```

RESULT:-

Result Grid   Filter Rows: <input type="text"/>		
	TITLE	IMDBSCORE
▶	David Attenborough: A Life on ...	9.0
	Emicida: AmarElo - It's All For Y...	8.6
	Springsteen on Broadway	8.5
	Ben Platt: Live from Radio City ...	8.4
	Taylor Swift: Reputation Stadiu...	8.4
	Winter on Fire: Ukraine's Fight ...	8.4
	Cuba and the Cameraman	8.3

NETFLIX_ORIGINALS 4 ...

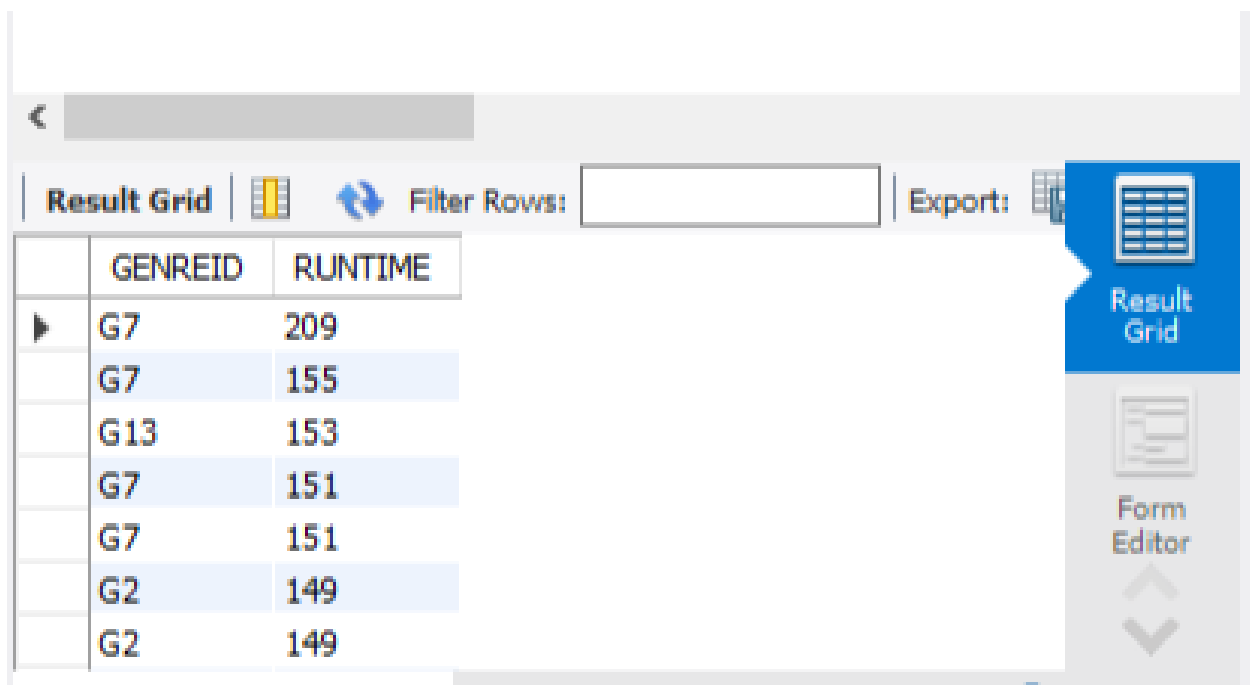
Retrieve the top 10 longest Netflix Originals by runtime.

QUERY:-

SELECT GENREID,RUNTIME FROM NETFLIX_ORIGINALS

ORDER BY RUNTIME DESC LIMIT 10;

RESULT:-



	GENREID	RUNTIME
▶	G7	209
	G7	155
	G13	153
	G7	151
	G7	151
	G2	149
	G2	149

Retrieve the titles of Netflix Originals along with their respective genres

QUERY:-

SELECT

N.Title,

G.Genre_Name

FROM

Netflix_Originals AS N

JOIN

Genre_Details AS G

ON

N.Genre_ID = G.Genre_ID;

RESULT:-

Result Grid			Filter Rows:	Export:
	TITLE	GENRE		
▶	Enter the Anime	Documentary		
	Dark Forces	Thriller		
	The App	Science Fiction		
	The Open House	Horror		
	Kaali Khuhi	Mystery		
	Drive	Action		
	Leyla Everlasting	Comedy		

Result 9 x Read O

- RankNetflix Originals based on their IMDb scores within each genre

QUERY:-

SELECT

no.title,

gd.genre,

no.imdbscore,

```

DENSE_RANK() OVER (PARTITION BY gd.genre ORDER BY
no.imdbscore DESC) AS imdb_rank_in_genre
FROM
netflix_originals no
JOIN
genre_details gd ON no.genreid = gd.genreid;

```

RESULT :-

Result Grid				
Filter Rows:				
	title	genre	imdbscore	
▶	Okja	Action	7.3	1
	The Siege of Jadotville	Action	7.2	2
	Extraction	Action	6.7	3
	The Old Guard	Action	6.7	3
	Mowgli: Legend of the Jungle	Action	6.5	4
	Polar	Action	6.3	5
	Sand Castle	Action	6.3	5

Result Grid
Form Editor

Result 22 ×
Read Only


Which Netflix Originals have IMDb scores higher than the average IMDb score of all titles?


QUERY:-

```
SELECT TITLE, GENREID, IMDBSCORE AS HIGHER_SCORES FROM  
NETFLIX_ORIGINALS WHERE IMDBSCORE > (SELECT AVG(IMDBSCORE) FROM  
NETFLIX_ORIGINALS);
```

RESULT:-


Result Grid






Filter Rows:


Export:




	TITLE	GENREID	HIGHER_SCORES
▶	1922	G9	6.3
	A Tale of Two Kitchens	G1	6.3
	Alex Strangelove	G17	6.3
	Apostle	G9	6.3
	Benji	G7	6.3
	Bright	G7	6.3
	Cargo	G7	6.3




Result Grid




Form Editor





NETFLIX_ORIGINALS 15

✕

 Read Only

C

- How many Netflix Originals are there in each genre?

```
SELECT g.genre ,count(*) as total_titles from
```

```
netflix_originals n
```

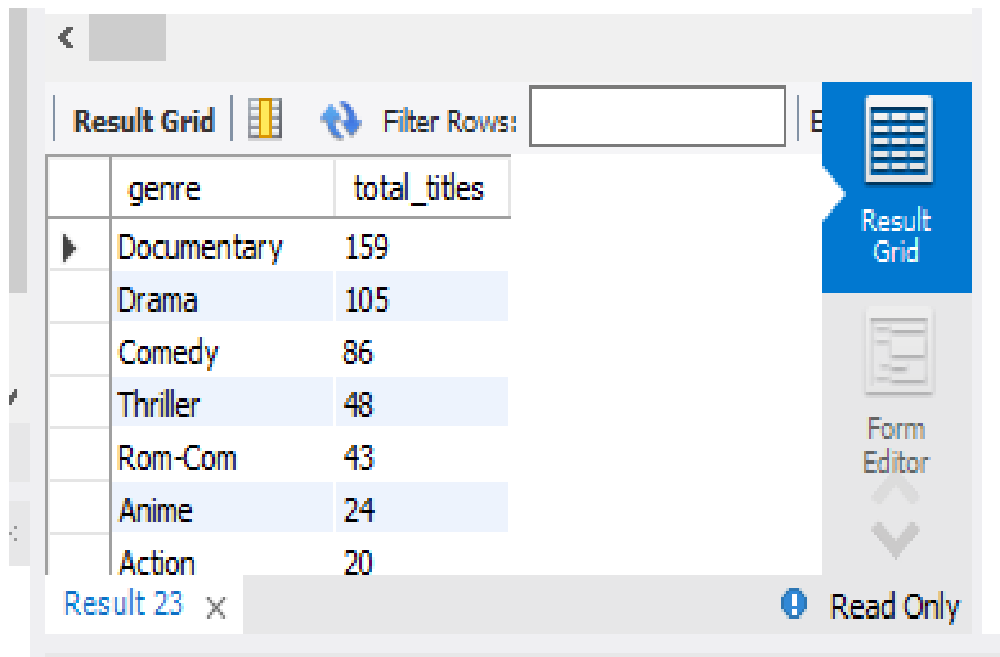
```
JOIN
```

```
genre_details g ON n.genreid = g.genreid
```

```
GROUP BY g.genre
```

```
ORDER BY
```

```
total_titles DESC;
```



The screenshot shows a database interface with a 'Result Grid' tab selected. The grid displays the results of a SQL query, showing the number of Netflix Originals for each genre. The genres are listed in descending order of total titles. The interface includes a 'Filter Rows' input field, a 'Result Grid' button, and a 'Form Editor' button. The status bar at the bottom indicates 'Result 23' and 'Read Only'.

genre	total_titles
Documentary	159
Drama	105
Comedy	86
Thriller	48
Rom-Com	43
Anime	24
Action	20

Which genres have more than 5 Netflix Originals with an IMDb score higher than 8?

QUERY:-

```
SELECT GENREID, COUNT(*) AS HIGH_RATED_TITLES
```

```
FROM NETFLIX_ORIGINALS
```

```
WHERE IMDBSCORE > 8
```

```
GROUP BY GENREID
```



```
HAVING COUNT(*) > 5
```

```
ORDER BY HIGH_RATED_TITLES DESC;
```

RESULT:-

12


<

Result Grid |  Filter Rows: | Export: 

	GENREID	HIGH_RATED_TITLES
▶	G1	12

Result Grid

Form Editor

Result 17 x  Read Only

What are the top 3 genres with the highest average IMDb scores, and how many Netflix Originals do they have

QUERY:-

SELECT

G.Genre_Name,

ROUND(AVG(N.IMDb_Score), 2) AS Avg_IMDb_Score,

COUNT(N.Title) AS Total_Originals

FROM

Netflix_Originals AS N

JOIN

Genre_Details AS G

ON

N.Genre_ID = G.Genre_ID

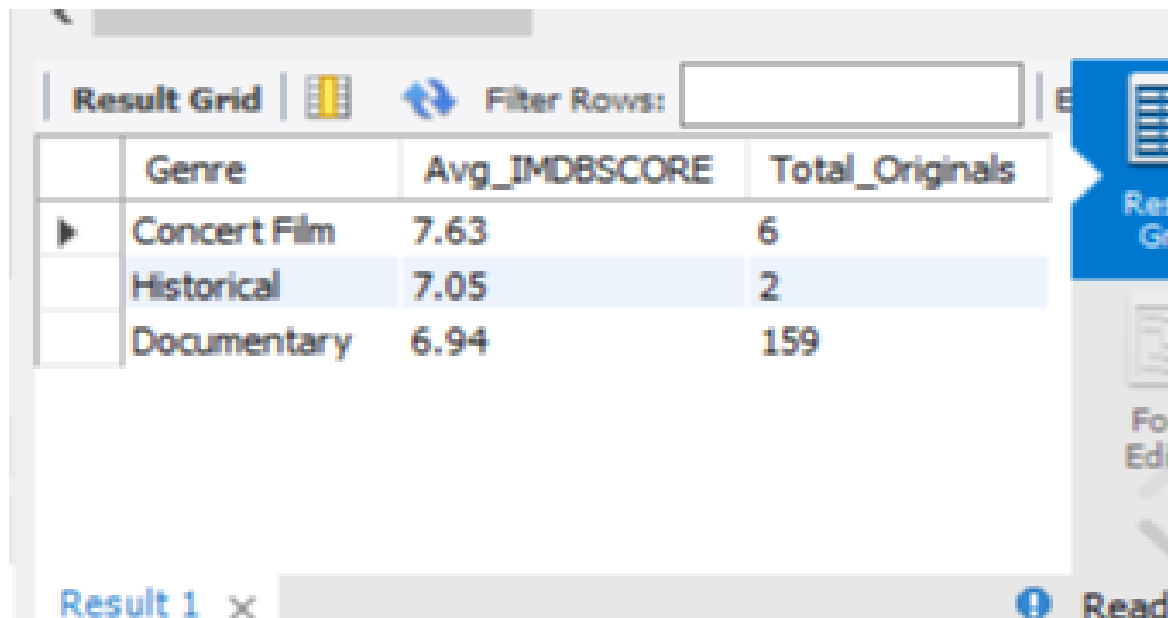
GROUP BY

G.Genre_Name

ORDER BY

Avg_IMDb_Score DESC

LIMIT 3;



The screenshot shows a database query result grid with the following data:

	Genre	Avg_IMDBSCORE	Total_Originals
▶	Concert Film	7.63	6
	Historical	7.05	2
	Documentary	6.94	159

The interface includes a 'Result Grid' tab, a 'Filter Rows' input field, and a status bar at the bottom indicating 'Result 1' and a 'Read' button.

KEY INSIGHTS:-

After analyzing the dataset of Netflix Originals, several insights were observed:

- **Average IMDb Scores by Genre:**
The analysis shows how each genre performs in terms of audience ratings. Genres such as *Documentary* and *Drama* generally have higher IMDb scores compared to *Comedy* or *Romance*.
- **Genres with High Average Ratings:**
The *Concert Film* genre has the highest average IMDb score (7.63), followed by *Historical* (7.05) and *Documentary* (6.94). This shows that audiences appreciate factual, artistic, and educational genres.
- **Top-Rated Titles:**
When ranked by IMDb scores, Netflix Originals show consistent quality in genres like *Drama* and *Documentary*, with several titles exceeding the platform-wide average rating.
- **Runtime Analysis:**
The longest Netflix Originals are mostly dramas.
- **HIGH RATED TITLES:-**
- Documentaries has more than 5 Netflix Originals with an IMDb score higher than 8.
- **High IMDb Scoring Titles:**
Multiple titles have IMDb scores higher than the dataset average, proving Netflix's consistent delivery of well-rated Originals.

Conclusion:-

The analysis highlights that Netflix's strongest genres are Concert Films, Historical, and Documentaries, based on average IMDb ratings. While *Documentary* has the largest volume of Originals, *Concert Film* content receives the highest audience appreciation.

This suggests that Netflix should continue investing in quality factual and cultural content, while also enhancing viewer engagement in other genres to improve their overall IMDb performance.

REFERENCE :-

**NETFLIX DATASET -INTERNSHALA
INTERNSHALA TRAINING COURSE CONTENT**