Netflix Data Analysis (SQL Portfolio project)

Netflix is a leading global provider of streaming entertainment, offering a vast library of movies, TV shows, and original content to its subscribers. The platform has transformed the way people consume entertainment, enabling on-demand streaming from various devices. In this data analysis project, we explore Netflix data to gain insights into its content, user interactions, and trends.

Through **data analysis and SQL** queries, we can gain a deeper understanding of Netflix content and user behavior. This project showcases the ability to extract meaningful insights from real-world data and highlights proficiency in using SQL for data manipulation.

- Cleaning the data is crucial for any data analysis project. For this project, I used a Kaggle's Netflix dataset, cleaned it ensuring it contains no null values, blank columns, or rows, also checking for consistency in values in specific columns.
- Created two CSV files with the help of Kaggle Netflix dataset Netflix_Contentdata and NetflixUserdata and each file has its own fields.
- The Netflix_Contentdata consists of the following fields such as: “Show_id” ,”Type”, “Title”,” Director”, “Date_added” ,”Release_year”, “Duration”,”Genre”.
- The NetflixUserdata has the following fields such as: "User_id," "User_name," and "Show_id".
- The two tables can be linked using “Show_id” column to establish a relationship between them.
- With the cleaned CSV files derived from the Kaggle Netflix dataset, it is easier to perform various analyses, generate insights, and run SQL queries to extract valuable information.

Queries to get insights from the data:-

- show databases;
- use netflixdata;
- show tables;
- describe netflix_contentdata;
- describe netflixuserdata;

```
1 •  show databases;
2 •  use netflixdata;
3 •  show tables;
4 •  describe netflix_contentdata;
5 •  describe netflixuserdata;
```

To display the records in the netflix_contentdata :

- select * from netflix_contentdata;

To count the number of records we use:-

- select count(*) from netflix_contentdata;

```
6 • select * from netflix_contentdata;
7 • select count(*) from netflix_contentdata;
8 • select * from netflixuserdata;
9 • select count(*) from netflixuserdata;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Show_id	Type	Title	Director	Date_added	Release_year	Duration	Genre
s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	9/25/21	2020	90 min	Documentaries
s2	TV Show	Ganglands	Julien Leclercq	9/24/21	2021	1 Season	Crime TV Shows
s3	TV Show	Midnight Mass	Mike Flanagan	9/24/21	2021	1 Season	TV Dramas
s4	Movie	Sankofa	Haile Gerima	9/24/21	1993	125 min	Dramas

```
6 • select * from netflix_contentdata;
7 • select count(*) from netflix_contentdata;
8 • select * from netflixuserdata;
9 • select count(*) from netflixuserdata;
```

esult Grid | Filter Rows: | Export: | Wr

count(*)
550

To display the records in the netflixuserdata:

- select * from netflixuserdata;

To count the number of records we use:

- select count(*) from netflixuserdata;

```
7 •  select count(*) from netflix_contentdata;
8 •  select * from netflixuserdata;
9 •  select count(*) from netflixuserdata;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

User_id	User_name	Show_id
1	Alicesmith	s3
2	John	s4
3	Garima	s5
4	Alicesmith	s4
5	Garima	s5

```
9 •  select count(*) from netflixuserdata;
```

Result Grid | Filter Rows: | Export:

count(*)
550

Retrieve all TV shows from the "Netflix_Contentdata" table:

- SELECT Title FROM Netflix_Contentdata WHERE Type = 'TV Show';
- SELECT count(Title) FROM Netflix_Contentdata WHERE Type = 'TV Show';

```
10 • SELECT Title FROM Netflix_Contentdata WHERE Type = 'TV Show';
11 • SELECT count(Title) FROM Netflix Contentdata WHERE Type = 'TV Show';
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Title
Ganglands
Midnight Mass
The Great British Baking Show
Bangkok Breaking
Monsters Inside: The 24 Faces of Billy Milligan

```
9  • select count(*) from netflixuserdata;
10 • SELECT Title FROM Netflix_Contentdata WHERE Type = 'TV Show';
11 • SELECT count(Title) FROM Netflix_Contentdata WHERE Type = 'TV Show';
12 • SELECT Title FROM Netflix Contentdata WHERE Type = 'Movie';
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

count(Title)
32

Retrieve all Movies from the "Netflix_Contentdata" table:

SELECT Title FROM Netflix_Contentdata WHERE Type = 'Movie';

SELECT count(Title) FROM Netflix_Contentdata WHERE Type = 'Movie';

```
12 • SELECT Title FROM Netflix_Contentdata WHERE Type = 'Movie';
13 • SELECT count(Title) FROM Netflix Contentdata WHERE Type = 'Movie';
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Title
Dick Johnson Is Dead
Sankofa
The Starling
Je Suis Karl
Confessions of an Invisible Girl

```
13 •  SELECT count(Title) FROM Netflix_Contentdata WHERE Type = 'Movie';
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

count(Title)
518

Find movies released in or after the year 2020:

SELECT Title, Release_year FROM Netflix_Contentdata WHERE Type = 'Movie' AND Release_year >= 2020;

```
14 •  SELECT Title, Release_year FROM Netflix_Contentdata WHERE Type = 'Movie' AND Release_year >= 2020;
15
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Title	Release_year
Dick Johnson Is Dead	2020
The Starling	2021
Je Suis Karl	2021
Confessions of an Invisible Girl	2021
Intrusion	2021

List the distinct Genres from the "Netflix_contentdata" table:

SELECT DISTINCT Genre FROM Netflix_Contentdata;

SELECT count (DISTINCT Genre) FROM Netflix_Contentdata;

```
15 •  SELECT DISTINCT Genre FROM Netflix_Contentdata;
16 •  SELECT count(DISTINCT Genre) FROM Netflix_Contentdata;
17 •  SELECT User_name FROM NetflixUserdata;
18
19
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Genre
Documentaries
Crime TV Shows
TV Dramas
Dramas
British TV Shows

```
15 •  SELECT DISTINCT Genre FROM Netflix_Contentdata;
16 •  SELECT count(DISTINCT Genre) FROM Netflix_Contentdata;
17 •  SELECT User_name FROM NetflixUserdata;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

	count(DISTINCT Genre)
▶	21

To find the top 5 genres that has the maximum number of views (i.e., the most-watched genre)

SELECT Genre, COUNT(*) AS GenreViews FROM Netflix_Contentdata C JOIN NetflixUserdata U ON C.Show_id = U.Show_id GROUP BY Genre ORDER BY GenreViews DESC LIMIT 5;

```
91 •  SELECT Genre, COUNT(*) AS GenreViews
92    FROM Netflix_Contentdata C
93    JOIN NetflixUserdata U ON C.Show_id = U.Show_id
94    GROUP BY Genre
95    ORDER BY GenreViews DESC
96    LIMIT 5;
```

Result Grid | Filter Rows: | Export: | Wrap Cell

Genre	GenreViews
▶ Action & Adventure	211
Children & Family Movies	124
Dramas	57
Comedies	44
Thrillers	27

List the user names from the "NetflixUserdata" table:

SELECT User_name FROM NetflixUserdata;

```
17 •  SELECT User_name FROM NetflixUserdata;
18
19
```

Result Grid | Filter Rows: | Export: | Wrap Cell Co

User_name
▶ Alicesmith
John
Garima
Alicesmith
Garima

List users who have watched TV shows:

SELECT DISTINCT U.User_name FROM NetflixUserdata U INNER JOIN Netflix_Contentdata C ON U.Show_id = C.Show_id WHERE C.Type = 'TV Show';

```
18 •  SELECT DISTINCT U.User_name
19    FROM NetflixUserdata U
20    INNER JOIN Netflix_Contentdata C ON U.Show_id = C.Show_id
21    WHERE C.Type = 'TV Show';
22
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

User_name
Teena
Bhumi
rohith
Jones
Micky
Jenny
Kevin

List users who have watched Movies:

SELECT DISTINCT U.User_name FROM NetflixUserdata U INNER JOIN Netflix_Contentdata C ON U.Show_id = C.Show_id WHERE C.Type = 'Movie';

```
22 •  SELECT DISTINCT U.User_name
23    FROM NetflixUserdata U
24    INNER JOIN Netflix_Contentdata C ON U.Show_id = C.Show_id
25    WHERE C.Type = 'Movie';
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

User_name
John
Alicesmith
Jennifer
Amanda
Lili
Edward
Meera

Find the year with the highest number of TV shows released:

SELECT Release_year, COUNT(*) AS Number_of_TV_Shows FROM Netflix_Contentdata

WHERE Type = 'TV Show' GROUP BY Release_year ORDER BY Number_of_TV_Shows DESC LIMIT 1;

```
26 •  SELECT Release_year, COUNT(*) AS Number_of_TV_Shows
27    FROM Netflix_Contentdata
28    WHERE Type = 'TV Show'
29    GROUP BY Release_year
30    ORDER BY Number_of_TV_Shows DESC
31    LIMIT 1;
32
```

Result Grid | Filter Rows: | Export: | Wrap Cell Conten

Release_year	Number_of_TV_Shows
2021	15

Retrieve usernames and the titles of movies they've watched:

SELECT U.User_name, C.Title FROM NetflixUserdata U INNER JOIN Netflix_Contentdata C ON U.Show_id = C.Show_id WHERE C.Type = 'Movie';

```
32 •  SELECT U.User_name, C.Title
33    FROM NetflixUserdata U
34    INNER JOIN Netflix_Contentdata C ON U.Show_id = C.Show_id
35    WHERE C.Type = 'Movie';
36
37
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

User_name	Title
Mohit	Grown Ups
rohith	Dark Skies
rohith	Paranoia
Richard	Ankahi Kahaniya
Lili	The Father Who Mo...
Lili	Birth of the Dragon
Edward	Jaws
Edward	Jaws 2
Samuel	Jaws 3

Find the directors of TV shows watched by users:

SELECT U.User_name, C.Director FROM NetflixUserdata U INNER JOIN Netflix_Contentdata C ON U.Show_id = C.Show_id WHERE C.Type = 'TV Show';

```
36 •  SELECT U.User_name, C.Director
37    FROM NetflixUserdata U
38    INNER JOIN Netflix_Contentdata C ON U.Show_id = C.Show_id
39    WHERE C.Type = 'TV Show';
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

User_name	Director
Alicesmith	Mike Flanagan
Garima	Andy Devonshire
Garima	Andy Devonshire
William	Mike Flanagan
William	Andy Devonshire
Jennifer	Mike Flanagan
Richard	Andy Devonshire
Amanda	Andy Devonshire
Richard	Mike Flanagan

Find the directors of Movies watched by users:

SELECT U.User_name, C.Director FROM NetflixUserdata U INNER JOIN Netflix_Contentdata C ON U.Show_id = C.Show_id WHERE C.Type = 'Movie';

```
40 •  SELECT U.User_name, C.Director
41    FROM NetflixUserdata U
42    INNER JOIN Netflix_Contentdata C ON U.Show_id = C.Show_id
43    WHERE C.Type = 'Movie';
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

User_name	Director
John	Haile Gerima
Alicesmith	Haile Gerima
Jennifer	Haile Gerima
Amanda	Haile Gerima
Lili	Haile Gerima
Edward	Haile Gerima
Meera	Haile Gerima
Bobby	Haile Gerima
rohith	Kirsten Johnson

Find the top 5 most-watched TV shows and their total viewership:

SELECT C.Title, COUNT(U.User_id) AS Total_Viewers FROM Netflix_Contentdata C

JOIN NetflixUserdata U ON C.Show_id = U.Show_id WHERE C.Type = 'TV Show'

GROUP BY C.Title ORDER BY Total_Viewers DESC LIMIT 5;

```
40 •  SELECT C.Title, COUNT(U.User_id) AS Total_Viewers
41    FROM Netflix_Contentdata C
42    JOIN NetflixUserdata U ON C.Show_id = U.Show_id
43    WHERE C.Type = 'TV Show'
44    GROUP BY C.Title
45    ORDER BY Total_Viewers DESC
46    LIMIT 5;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

Title	Total_Viewers
Midnight Mass	13
The Great British Baking Show	13
Ganglands	6
Bangkok Breaking	1
Monsters Inside: The 24 Faces of Billy Milligan	1

Find the top 15 most-watched Movies and their total viewership:

SELECT C.Title, COUNT(U.User_id) AS Total_Viewers FROM Netflix_Contentdata C

JOIN NetflixUserdata U ON C.Show_id = U.Show_id WHERE C.Type = 'Movie' GROUP BY C.Title

ORDER BY Total_Viewers DESC LIMIT 15;

```
47 •  SELECT C.Title, COUNT(U.User_id) AS Total_Viewers
48    FROM Netflix_Contentdata C
49    JOIN NetflixUserdata U ON C.Show_id = U.Show_id
50    WHERE C.Type = 'Movie'
51    GROUP BY C.Title
52    ORDER BY Total_Viewers DESC
53    LIMIT 15;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch r

Title	Total_Viewers
Sankofa	12
Dick Johnson Is Dead	4
Chhota Bheem aur Krishna	4
Chhota Bheem in African Safari	4
Chhota Bheem: Bheem vs Aliens	4
Chhota Bheem: Dholakpur to Kathmandu	4
Chhota Bheem: Journey to Petra	4
Chhota Bheem: Master of Shaolin	4
Chhota Bheem: The Rise of Kirmada	4

Find the user(s) who watched the most content on Netflix (movies and TV shows combined):

SELECT U.User_name, COUNT(U.Show_id) AS Total_Watched FROM NetflixUserdata U

GROUP BY U.User_name ORDER BY Total_Watched DESC LIMIT 1;

```
54 •  SELECT U.User_name, COUNT(U.Show_id) AS Total_Watched
55    FROM NetflixUserdata U
56    GROUP BY U.User_name
57    ORDER BY Total_Watched DESC
58    LIMIT 1;
59
60
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

User_name	Total_Watched
William	68

List users who watched Movies from the "comedies" genre and the titles of those TV shows:

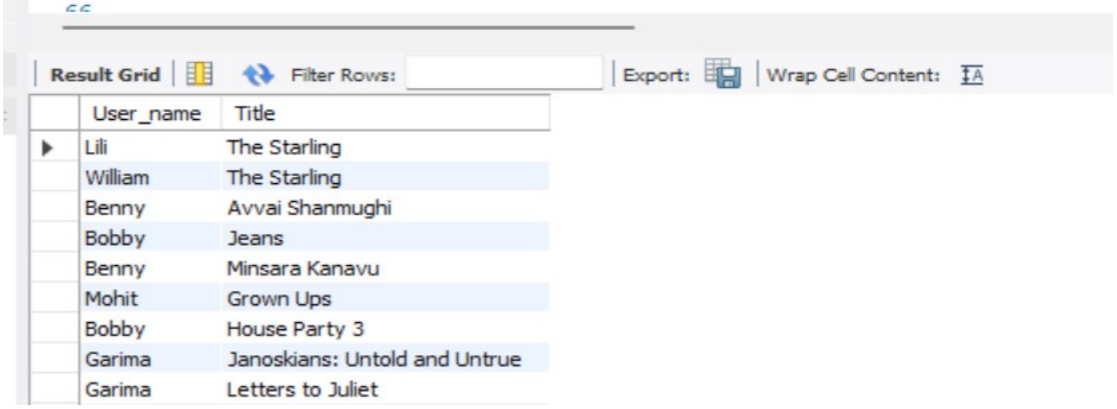
SELECT U.User_name, C.Title

FROM NetflixUserdata U

JOIN Netflix_Contentdata C ON U.Show_id = C.Show_id

WHERE C.Type = 'Movie' AND C.Genre = 'Comedies';

```sql
SELECT U.User_name, C.Title
FROM NetflixUserdata U
JOIN Netflix_Contentdata C ON U.Show_id = C.Show_id
WHERE C.Type = 'Movie' AND C.Genre = 'Comedies';
```



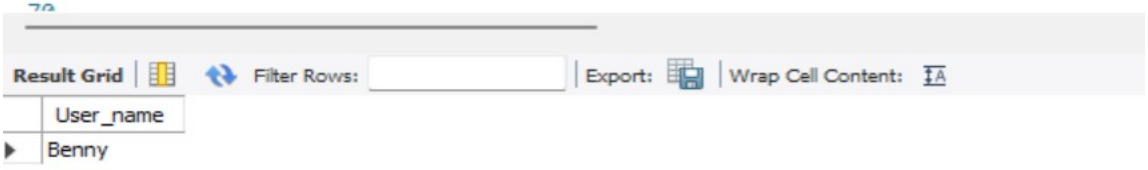
Result Grid | Filter Rows: | Export: | Wrap Cell Content:

User_name	Title
Lili	The Starling
William	The Starling
Benny	Avvai Shanmughi
Bobby	Jeans
Benny	Minsara Kanavu
Mohit	Grown Ups
Bobby	House Party 3
Garima	Janoskians: Untold and Untrue
Garima	Letters to Juliet

Find the user(s) who watched content directed by "Rajiv Menon":

SELECT U.User_name FROM NetflixUserdata U JOIN Netflix_Contentdata C ON U.Show_id = C.Show_id WHERE C.Director = 'Rajiv Menon';

```sql
SELECT U.User_name
FROM NetflixUserdata U
JOIN Netflix_Contentdata C ON U.Show_id = C.Show_id
WHERE C.Director = 'Rajiv Menon';
```



Result Grid | Filter Rows: | Export: | Wrap Cell Content:

User_name
Benny

List Movies with the same directors watched by multiple users:

SELECT C.Title, C.Director, COUNT(U.User_id) AS Viewers FROM NetflixUserdata U

JOIN Netflix_Contentdata C ON U.Show_id = C.Show_id WHERE C.Type = 'Movie'

GROUP BY C.Title, C.Director HAVING COUNT(U.User_id) > 1;

```
67 •  SELECT C.Title, C.Director, COUNT(U.User_id) AS Viewers
68    FROM NetflixUserdata U
69    JOIN Netflix_Contentdata C ON U.Show_id = C.Show_id
70    WHERE C.Type = 'Movie'
71    GROUP BY C.Title, C.Director
72    HAVING COUNT(U.User_id) > 1;
73
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Title	Director	Viewers
Sankofa	Haile Gerima	12
Dick Johnson Is Dead	Kirsten Johnson	4
The Starling	Theodore Melfi	2
Birth of the Dragon	George Nolfi	2
Chhota Bheem aur Krishna	Rajiv Chilaka	4
Chhota Bheem aur Krishna vs Zimbara	Rajiv Chilaka	4
Chhota Bheem in African Safari	Binayak Das	4
Chhota Bheem: Bheem vs Aliens	Rajiv Chilaka	4
Chhota Bheem: Dholakpur to Kathmandu	Rajiv Chilaka	4

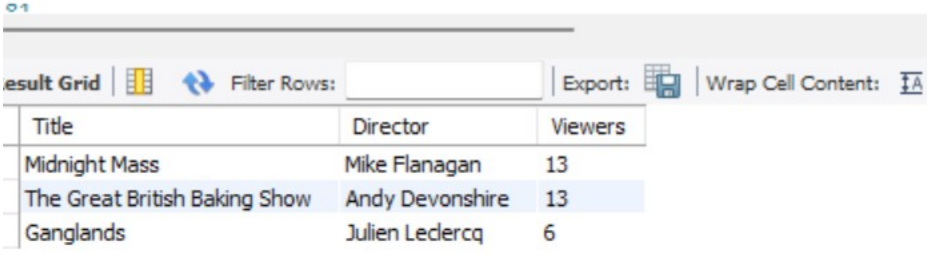
List TV shows with the same directors watched by multiple users:

SELECT C.Title, C.Director, COUNT(U.User_id) AS Viewers FROM NetflixUserdata U

JOIN Netflix_Contentdata C ON U.Show_id = C.Show_id WHERE C.Type = 'TV Show'

GROUP BY C.Title, C.Director HAVING COUNT(U.User_id) > 1;

```
73 •    SELECT C.Title, C.Director, COUNT(U.User_id) AS Viewers
74      FROM NetflixUserdata U
75      JOIN Netflix_Contentdata C ON U.Show_id = C.Show_id
76      WHERE C.Type = 'TV Show'
77      GROUP BY C.Title, C.Director
78      HAVING COUNT(U.User_id) > 1;
79
80
```



Title	Director	Viewers
Midnight Mass	Mike Flanagan	13
The Great British Baking Show	Andy Devonshire	13
Ganglands	Julien Leclercq	6

List all users and the titles of movies they've watched, including users who haven't watched any movies:

SELECT U.User_name, C.Title FROM NetflixUserdata U LEFT JOIN Netflix_Contentdata C ON U.Show_id = C.Show_id WHERE C.Type = 'Movie';

```
79 •   SELECT U.User_name, C.Title
80     FROM NetflixUserdata U
81     LEFT JOIN Netflix_Contentdata C ON U.Show_id = C.Show_id
82     WHERE C.Type = 'Movie';
83
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

User_name	Title
John	Sankofa
Alicesmith	Sankofa
Jennifer	Sankofa
Amanda	Sankofa
Lili	Sankofa
Edward	Sankofa

Display all movies and their user viewers, including movies that haven't been watched by any users:

SELECT U.User_name, C.Title FROM NetflixUserdata U RIGHT JOIN Netflix_Contentdata C ON U.Show_id = C.Show_id WHERE C.Type = 'Movie';

```
3 •   SELECT U.User_name, C.Title
4     FROM NetflixUserdata U
5     RIGHT JOIN Netflix_Contentdata C ON U.Show_id = C.Show_id
6     WHERE C.Type = 'Movie';
```

sult Grid | Filter Rows: | Export: | Wrap Cell Content:

User_name	Title
Amanda	Sankofa
Jennifer	Sankofa
Alicesmith	Sankofa
John	Sankofa
William	The Starling

list users who watched movies released after 2020, including the users who haven't watched any such movies:

SELECT U.User_name FROM NetflixUserdata U LEFT JOIN Netflix_Contentdata C ON U.Show_id = C.Show_id WHERE C.Type = 'Movie' AND C.Release_year > 2020;

```
87 •    SELECT U.User_name
88      FROM NetflixUserdata U
89      LEFT JOIN Netflix_Contentdata C ON U.Show_id = C.Show_id
90      WHERE C.Type = 'Movie' AND C.Release_year > 2020;
91
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

User_name
Lili
William
William
Meera
Meera
Richard
Lili
Samuel
Samuel
Bobby
William
Bhumi

Insights:

1. **Content Popularity:** Users have engaged with both movies and TV shows, with 518 movies and 32 TV shows in the dataset.
2. **Genre Diversity:** There are 21 different genres in the dataset, with "Action" and "Adventure" being highly viewed by users.
3. **Release Year:** 2021 had the highest number of TV shows released, indicating a potential growth in TV show content.
4. **Top TV Shows:** The top 3 most-watched TV shows are "Midnight Mass," "The Great British Baking Show," and "Ganglands."
5. **Top Movie:** The top-watched movie is "Sankofa."
6. **User Engagement:** The user "William" has watched the most content on Netflix, suggesting a highly engaged user.

Recommendations:

1. **Content Diversity:** It is good to continue investing in a diverse range of content, including both movies and TV shows.
2. **Genre-Specific Promotion:** It is good to consider targeted marketing and promotions for highly viewed genres.
3. **Content Quality:** Focus on producing high-quality content, especially in genres with high user engagement.
4. **User Retention:** Analyze user activity patterns to retain engaged users.
5. **Data Quality Improvement:** Ensure data quality by addressing null values and inconsistencies if any.

These recommendations aim to enhance user satisfaction, increase user engagement, and make data-driven decisions to improve the Netflix platform and user experience.

Thank you!