

NETFLIX

INTRODUCTION

This project uses SQL to clean and analyze Netflix data. By organizing and exploring the dataset, we answer key questions about content types, popular genres, ratings, and trends across countries. Key steps include:

- Cleaning data by handling missing values.
- Counting movies vs. TV shows.
- Identifying popular genres and top-rated content.
- Analyzing content trends by year and region.

BACKGROUND

Netflix, a global streaming platform, offers a wide variety of movies, TV shows, and documentaries. Understanding its content library can reveal trends in popular genres, regional content, and viewer preferences.

In this project, we use SQL to clean and analyze Netflix's data. By organizing and exploring the dataset, we aim to uncover useful insights that can help guide content decisions and improve recommendations.

GOALS

The goal of this project is to clean and analyze Netflix's content data using SQL to uncover useful insights. We aim to identify popular content types, genres, and ratings, as well as trends in content release by country and year. Additionally, we'll explore which directors and actors appear most often and categorize content based on keywords in descriptions. Overall, this analysis will provide a clearer understanding of Netflix's catalog and help guide content recommendations and strategy.

PROBLEMS

This project addresses key questions about Netflix's content library, including:

1. Count the number of Movies vs TV Shows
2. Find the most common rating for movies and TV show
3. List all movies released in a specific year (e.g., 2020)
4. Find the top 5 countries with the most content on Netflix
5. Identify the longest movie
6. Find content added in the last 5 years
7. Find all the movies/TV shows by director 'Rajiv Chilaka'!

PROBLEMS

8. List all TV shows with more than 5 seasons.
9. Find each year and the average numbers of content release in India on netflix.
return top 5 year with highest avg content release.
10. List all movies that are documentaries.
11. Find all content without a director.
12. Find how many movies actor 'Salman Khan' appeared in last 10 years.
13. Categorize the content based on the presence of the keywords 'kill' and 'violence' in the description field.

FileEditViewQueryDatabaseServerToolsScriptingHelp

Navigator

SCHEMAS

Filter objects

netflix

Tables

netflix

Columns

Indexes

Foreign Keys

Triggers

Views

Stored Procedures

Functions

pizzahut

sakila

sys

AdministrationSchemas

Information

Schema: netflix

SQL File 4*SQL File 5*SQL File 6*SQL File 7*netflixSQL File 7*SQL File 8*SQL File 9*SQL File 10*SQL File 12*SQL File 13*SQL File

Limit to 1000 rows

1-- Count the number of Movies vs TV Shows

2• SELECT type,count(*) AS Total_content

3FROM netflix

4GROUP BY Type;

Output

Action Output

#	Time	Action	Message	Duration / Fetch
52	06:57:21	SELECT CASE WHEN description LIKE "%kill%" OR description LIKE "%viol..."	2 row(s) returned	0.000 sec / 0.000 sec

FileEditViewQueryDatabaseServerToolsScriptingHelp

Navigator

SCHEMAS

Filter objects

netflix

Tables

netflix

Columns

Indexes

Foreign Keys

Triggers

Views

Stored Procedures

Functions

pizzahut

sakila

sys

AdministrationSchemas

Information

SQL File 4*SQL File 5*SQL File 6*SQL File 7*netflixSQL File 7*SQL File 8*SQL File 9*SQL File 10*SQL File 12*SQL File 13*SQL File

Limit to 1000 rows

1-- Find the most common rating for movies and TV shows

2

3• SELECT rating, count(*) AS count_data

4from netflix

5group by rating

6order by count_data desc

7limit 1;

Output

Action Output

#	Time	Action	Message	Duration / Fetch
---	------	--------	---------	------------------

EMAS

filter objects

netflix

- Tables
 - netflix
 - Columns
 - Indexes
 - Foreign Keys
 - Triggers
- Views
- Stored Procedures
- Functions

pizzahut

sakila

sys

Administration Schemas

Schema: netflix

SQL File 4* SQL File 5* SQL File 6* SQL File 7* x netflix SQL File 7* SQL File 8* SQL File 9* SQL File 10* SQL File 12* SQL File 13* SQL File

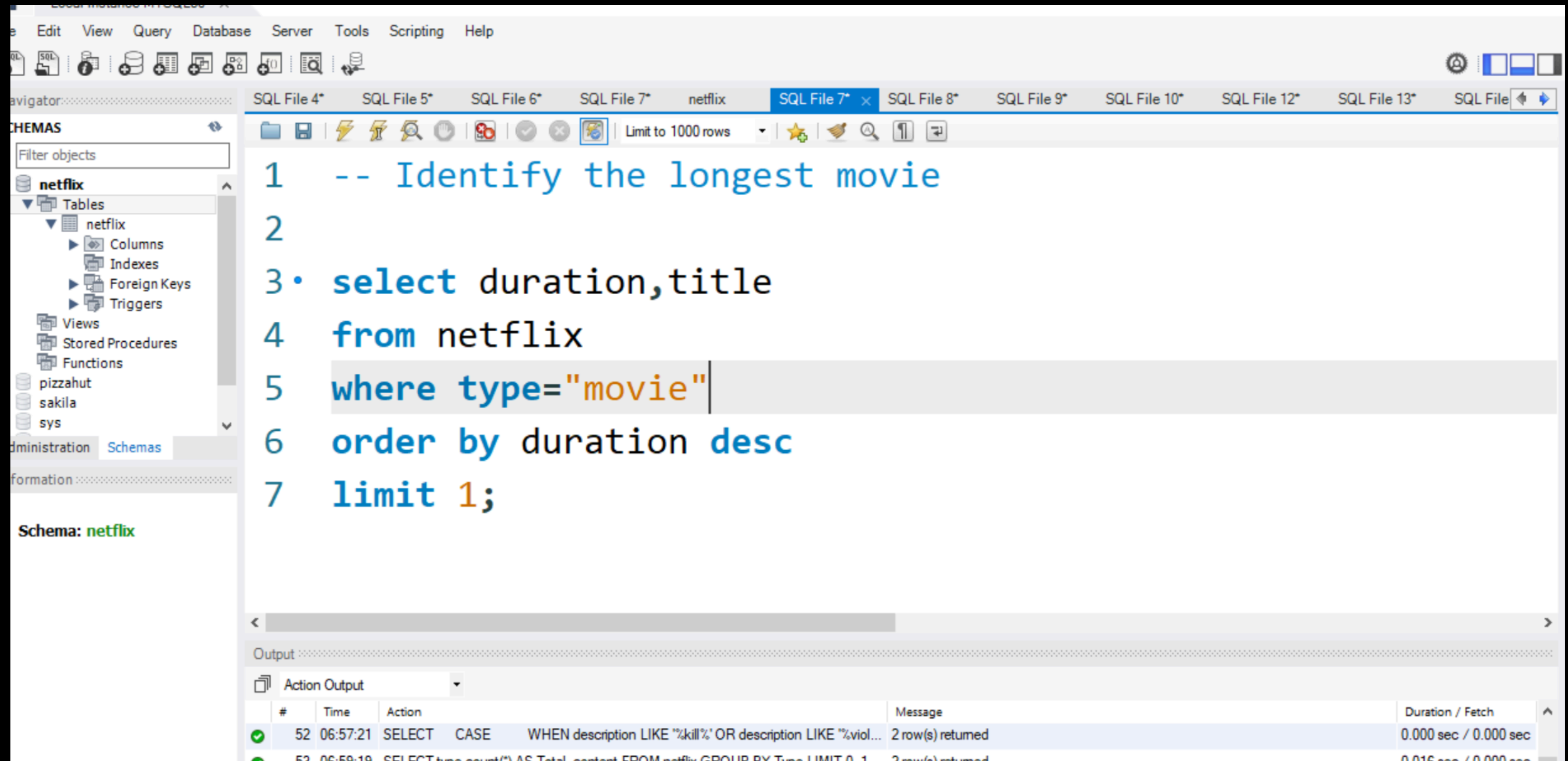
Limit to 1000 rows

```
1 -- Find the top 5 countries with the most content on Netflix
2
3 • select country, count(*) as count
4 from netflix
5 group by country
6 order by count desc
7 limit 5;
```

Output

Action Output

#	Time	Action	Message	Duration / Fetch
52	06:57:21	SELECT CASE WHEN description LIKE "%kill%" OR description LIKE "%viol..."	2 row(s) returned	0.000 sec / 0.000 sec
53	06:59:19	SELECT type, count(*) AS Total_content FROM netflix GROUP BY Type LIMIT 0, 1	2 row(s) returned	0.016 sec / 0.000 sec



The screenshot shows a SQL IDE interface. The top menu bar includes 'Edit', 'View', 'Query', 'Database', 'Server', 'Tools', 'Scripting', and 'Help'. The toolbar contains various icons for file operations and execution. The left sidebar shows a tree view of database objects, including 'netflix' under 'Tables'. The main editor window displays a SQL query:

```
1 -- Find content added in the last 5 years
2
3 select title,type,release_year
4 from netflix
5 where release_year>= year(curdate())-5;
```

The query is highlighted in blue. Below the editor, the 'Output' tab is selected, showing a table with columns: '#', 'Time', 'Action', 'Message', and 'Duration / Fetch'. The table contains one row:

#	Time	Action	Message	Duration / Fetch
52	06:57:21	SELECT CASE WHEN description LIKE "%kill%" OR description LIKE "%viol..."	2 row(s) returned	0.000 sec / 0.000 sec

Local instance: MySQL 5.6.23

Edit View Query Database Server Tools Scripting Help

SQL File 4* SQL File 5* SQL File 6* SQL File 7* netflix SQL File 7* SQL File 8* SQL File 9* x SQL File 10* SQL File 12* SQL File 13* SQL File

netflix

- Tables
 - netflix
 - Columns
 - Indexes
 - Foreign Keys
 - Triggers
 - Views
 - Stored Procedures
 - Functions
- pizzahut
- sakila
- sys

Administration Schemas

Schema: netflix

```
1 -- Find all the movies/TV shows by director 'Rajiv Chilaka'!  
2  
3 • select type, title  
4 from netflix  
5 where director="Rajiv Chilaka";
```

Output

Action Output

#	Time	Action	Message	Duration / Fetch
✓ 52	06:57:21	SELECT CASE WHEN description LIKE "%kill%" OR description LIKE "%viol..."	2 row(s) returned	0.000 sec / 0.000 sec
✓ 53	06:59:19	SELECT type,count(*) AS Total_content FROM netflix GROUP BY Type LIMIT 0, 1...	2 row(s) returned	0.016 sec / 0.000 sec

Info Session

Local instance: MySQL 5.7.17

Edit View Query Database Server Tools Scripting Help

SQL File 4* SQL File 5* SQL File 6* SQL File 7* netflix SQL File 7* SQL File 8* SQL File 9* SQL File 10* x SQL File 12* SQL File 13* SQL File 14*

netflix

filter objects

netflix

- Tables
 - netflix
 - Columns
 - Indexes
 - Foreign Keys
 - Triggers
 - Views
 - Stored Procedures
 - Functions
- pizzahut
- sakila
- sys

Administration Schemas

Schema: netflix

```
1 -- List all TV shows with more than 5 seasons
2
3 • SELECT title
4 FROM netflix
5 WHERE type = 'TV Show' AND CAST(duration AS UNSIGNED) > 5;
```

Output

Action Output

#	Time	Action	Message	Duration / Fetch
52	06:57:21	SELECT CASE WHEN description LIKE "%kill%" OR description LIKE "%viol..."	2 row(s) returned	0.000 sec / 0.000 sec
53	06:59:19	SELECT type, count(*) AS Total_content FROM netflix GROUP BY Type LIMIT 0, 1	2 row(s) returned	0.016 sec / 0.000 sec

Local instance MySQL80

Edit View Query Database Server Tools Scripting Help

SQL File 4* SQL File 5* SQL File 6* SQL File 7* netflix SQL File 7* SQL File 8* SQL File 9* SQL File 10* SQL File 12* x SQL File 13* SQL File

netflix

filter objects

netflix

Tables

Columns

Indexes

Foreign Keys

Triggers

Views

Stored Procedures

Functions

pizzahut

sakila

sys

Administration Schemas

Schema: netflix

```
1 -- Find each year and the average numbers of content release in India on netflix.
2 -- return top 5 year with highest avg content release!
3
4 • SELECT release_year, AVG(content_count) AS avg_content_count
5 FROM (
6     SELECT release_year, COUNT(*) AS content_count
7     FROM netflix
8     WHERE country = 'India'
9     GROUP BY release_year
10 ) AS yearly_content
11 GROUP BY release_year
12 ORDER BY avg_content_count DESC
13 LIMIT 5;
```

Output

Action Output

#	Time	Action	Message	Duration / Fetch
✓ 52	06:57:21	SELECT CASE WHEN description LIKE "%kill%" OR description LIKE "%viol..."	2 row(s) returned	0.000 sec / 0.000 sec

Local instance MySQL80

Edit View Query Database Server Tools Scripting Help

SQL File 4* SQL File 5* SQL File 6* SQL File 7* netflix SQL File 7* SQL File 8* SQL File 9* SQL File 10* SQL File 12* SQL File 13* x SQL File

netflix

Filter objects

netflix

- Columns
- Indexes
- Foreign Keys
- Triggers
- Views
- Stored Procedures
- Functions

pizzahut sakila sys

ministration Schemas

Schema: netflix

```
1 -- List All Movies that are Documentaries
2
3 • SELECT *
4 FROM netflix
5 WHERE listed_in LIKE '%Documentaries';
```

Output

Action Output

#	Time	Action	Message	Duration / Fetch
52	06:57:21	SELECT CASE WHEN description LIKE "%kill%" OR description LIKE "%viol..."	2 row(s) returned	0.000 sec / 0.000 sec
53	06:59:19	SELECT type count(*) AS Total content FROM netflix GROUP BY Type LIMIT 0, 1	2 row(s) returned	0.016 sec / 0.000 sec

Local instances in PostgreSQL

File Edit View Query Database Server Tools Scripting Help

SQL File 4* SQL File 5* SQL File 6* SQL File 7* netflix SQL File 7* SQL File 8* SQL File 9* SQL File 10* SQL File 12* SQL File 13* SQL File 14*

Limit to 1000 rows

netflix

Tables
netflix
Columns
Indexes
Foreign Keys
Triggers
Views
Stored Procedures
Functions
pizzahut
sakila
sys

Schemas

Schema: netflix

1 -- Find All Content Without a Director

2

3 • SELECT *

4 FROM netflix

5 WHERE director IS NULL;

Output

Action Output

#	Time	Action	Message	Duration / Fetch
52	06:57:21	SELECT CASE WHEN description LIKE "%kill%" OR description LIKE "%viol..."	2 row(s) returned	0.000 sec / 0.000 sec

EMAS

Filter objects

netflix

Tables

- netflix
 - Columns
 - Indexes
 - Foreign Keys
 - Triggers
- Views
- Stored Procedures
- Functions

pizzahut

sakila

sys

Administration Schemas

Schema: netflix

SQL File 7* netflix SQL File 7* SQL File 8* SQL File 9* SQL File 10* SQL File 12* SQL File 13* SQL File 14* SQL File 15* x SQL File 17* ne

Limit to 1000 rows

```
1 -- Find How Many Movies Actor 'Salman Khan' Appeared in the Last 10 Years
2
3 • SELECT COUNT(*) AS count
4 FROM netflix
5 WHERE type = 'Movie' AND cast LIKE '%Salman Khan%' AND release_year >= YEAR(CURDATE()) - 10;
```

Output

Action Output

#	Time	Action	Message	Duration / Fetch
52	06:57:21	SELECT CASE WHEN description LIKE '%kill%' OR description LIKE '%viol...'	2 row(s) returned	0.000 sec / 0.000 sec

Local instance MySQL80

Edit View Query Database Server Tools Scripting Help

SQL File 7* netflix SQL File 7* SQL File 8* SQL File 9* SQL File 10* SQL File 12* SQL File 13* SQL File 14* SQL File 15* SQL File 17* x ne

netflix

Tables

- netflix
 - Columns
 - Indexes
 - Foreign Keys
 - Triggers
- Views
- Stored Procedures
- Functions

pizzahut

sakila

sys

Administration Schemas

Schema: netflix

```
1  -- Categorize Content Based on the Presence of 'Kill' and 'Violence' Keywords
2
3  • SELECT
4      CASE
5          WHEN description LIKE '%kill%' OR description LIKE '%violence%' THEN 'Bad'
6          ELSE 'Good'
7      END AS category,
8      COUNT(*) AS count
9  FROM netflix
10 GROUP BY category;
```

Output

Action Output

#	Time	Action	Message	Duration / Fetch
52	06:57:21	SELECT CASE WHEN description LIKE "%kill%" OR description LIKE "%viol..."	2 row(s) returned	0.000 sec / 0.000 sec

ANALYSIS

This project analyzes Netflix's content data by cleaning and exploring it using SQL. We answer key questions like how many movies vs. TV shows are available, what genres and ratings are most common, and which countries have the most content. We also look at trends in content released over the years, the most popular directors and actors, and classify content based on keywords in descriptions. The goal is to uncover patterns in Netflix's catalog to help guide content decisions and improve recommendations.

CONCLUSION

This project successfully analyzed Netflix's content data using SQL to uncover key patterns and insights. By cleaning the data and answering important questions, we gained a better understanding of content distribution, popular genres, ratings, and regional trends. We also identified trends over the years, as well as the influence of specific directors and actors. These insights can be used to improve content strategies and recommendations on the platform, providing valuable information for content creators and business decisions.



THANK YOU

NETFLIX

BY
RAHUL YADAV