

|Unlocking Content Strategy with Data-Powered Insights

NETFLIX

Content Analysis with SQL

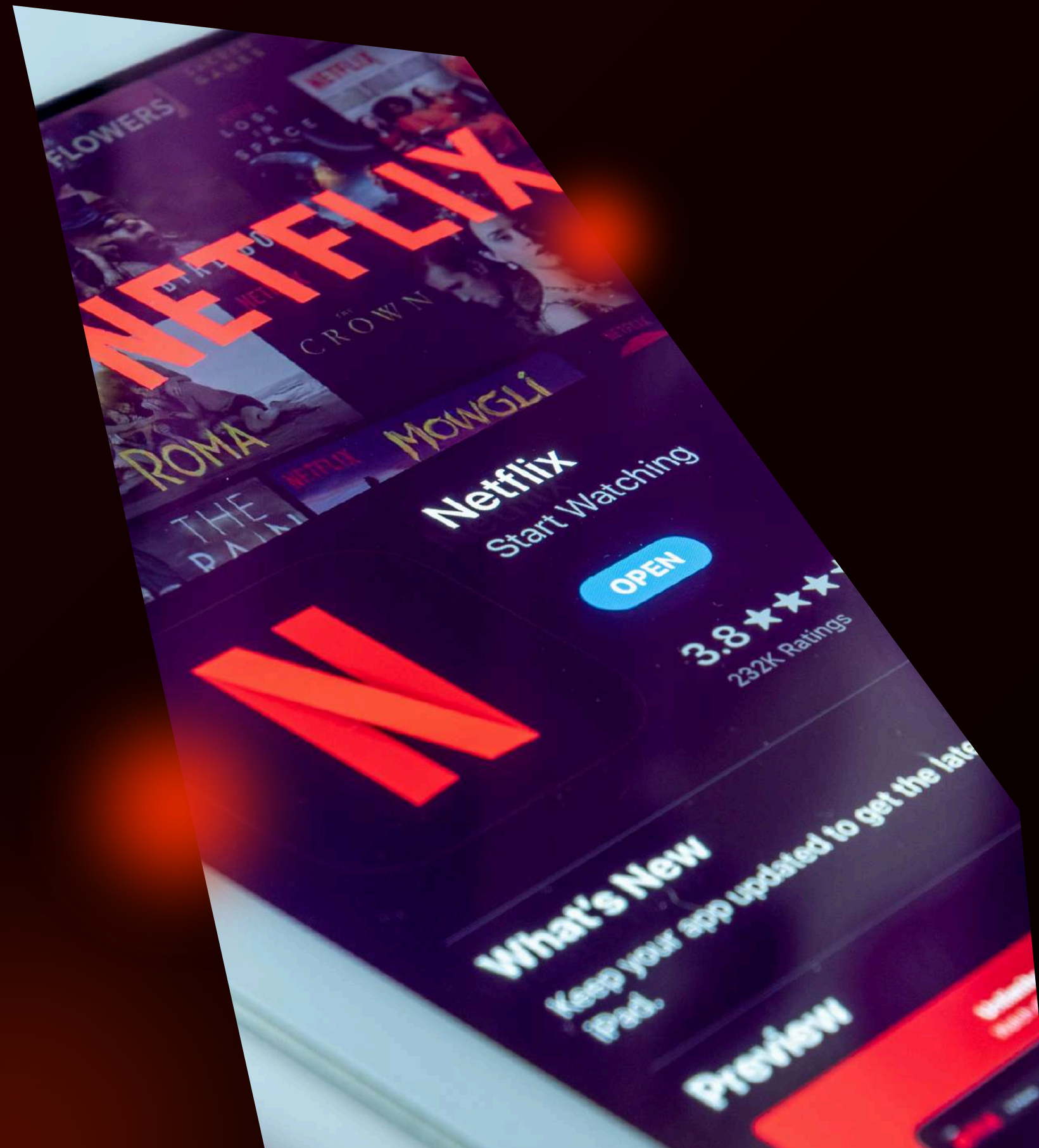
PROJECT OVERVIEW

- Comprehensive analysis of Netflix's movie and TV show library.
- Dataset from Kaggle with 8,000+ records.
- Used advanced SQL queries for pattern extraction and business insights.
- Aimed at informing data-driven decisions for content strategy.

PROJECT GOALS

- Analyze distribution of movies and TV shows.
- Identify popular content ratings.
- Understand regional content production and genre preferences.
- Flag sensitive content using keyword analysis.





TOOLS & TECHNOLOGIES

01

PostgreSQL: Primary query engine for complex data analysis.

02

SQL: Core language for all data transformations and insights.

03

Excel: Data preview and basic analysis of dataset. It helps to understand.

DATASET SUMMARY

- Sourced from Kaggle website
- Over 8,000 entries with 12 fields.
- Covers:
 - Show Type, Title, Director, Cast
 - Country, Date Added, Release Year
 - Rating, Duration, Genre, Description



DATASET FEATURES

- show_id - Unique identifier for each show.
- type - Either "Movie" or "TV Show".
- title, director, casts - Descriptive metadata.
- country, date_added, release_year - Contextual data.
- rating, duration - Classification & length.
- listed_in - Genre tags.
- description - Synopsis used for keyword-based classification.



SQL QUERIES

01

Count of Movies vs TV Shows

Objective: Determine the distribution of content types on Netflix.

```
SELECT
    type,
    COUNT(*)
FROM netflix
GROUP BY TYPE;
```

	type character varying (10) 🔒	count bigint 🔒
1	Movie	6131
2	TV Show	2676

SQL QUERIES

02

Most Common Ratings by Type

Objective: Identify the most frequently occurring rating for each type of content.

```
SELECT
  type,
  rating,
  cnt
FROM (
  SELECT
    type,
    rating,
    COUNT(*) AS cnt,
    RANK() OVER (PARTITION BY type ORDER BY COUNT(*) DESC) AS ranking
  FROM netflix
  GROUP BY type, rating
) AS t1
WHERE
  ranking = 1;
```

	type	rating	cnt
	character varying (10)	character varying (10)	bigint
1	Movie	TV-MA	2062
2	TV Show	TV-MA	1145

SQL QUERIES

03

Movies Released in 2020

Objective: Retrieve all movies released in a specific year.

```
SELECT * FROM netflix
WHERE
    type = 'Movie'
AND
    release_year = 2020;
```

Showing rows: 1 to 517					
Page No: 1 of 1					
	show_id character varying (6)	type character varying (10)	title character varying (150)	director character varying (208)	casts character varying
1	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	[null]
2	s17	Movie	Europe's Most Dangerous Man: Otto Skorzeny in Spain	Pedro de Echave García, Pablo Azorín Williams	[null]
3	s79	Movie	Tughlaq Durbar	Delhiprasad Deenadayalan	Vijay Sethupathi,
4	s85	Movie	Omo Ghetto: the Saga	JJC Skillz, Funke Akindele	Funke Akindele, /
5	s104	Movie	Shadow Parties	Yemi Amodu	Jide Kosoko, Om
6	s120	Movie	Here and There	JP Habac	Janine Gutierrez,
Total rows: 517 Query complete 00:00:00.157 CRLF Ln 59, Col 1					

SQL QUERIES

04

Top 5 Countries by Content

Objective: Identify the top 5 countries with the highest number of content items.

```
SELECT
  UNNEST(STRING_TO_ARRAY(country, ',')) AS new_country,
  COUNT(show_id) AS total_content
FROM netflix
GROUP BY 1
ORDER BY 2 DESC
LIMIT 5;
```

	new_country text	total_content bigint
1	United States	3211
2	India	1008
3	United Kingd...	628
4	United States	479
5	Canada	271

SQL QUERIES

05

Longest Movie

Objective: Find the movie with the longest duration.

```
SELECT * FROM netflix
WHERE
    type = 'Movie'
    AND
    duration = (SELECT MAX(duration) FROM netflix);
```

Showing rows: 1 to 118 | Page No: 1 of 1

	show_id character varying (6)	type character varying (10)	title character varying (150)	director character varying (208)	casts character varying (1000)
1	s52	Movie	InuYasha the Movie 2: The Castle Beyond the Looking Glass	Toshiya Shinohara	Kappei Yamaguchi, Satsuki Yukino, Mieko Harada, Koji Tsuji
2	s53	Movie	InuYasha the Movie 3: Swords of an Honorable Ruler	Toshiya Shinohara	Kappei Yamaguchi, Satsuki Yukino, Koji Tsujitani, Houko Ku
3	s120	Movie	Here and There	JP Habac	Janine Gutierrez, JC Santos, Victor Anastacio, Yesh Burce, L
4	s338	Movie	Good Luck Chuck	Mark Helfrich	Dane Cook, Jessica Alba, Dan Fogler, Ellia English, Sasha Pi
5	s345	Movie	My Girl 2	Howard Zieff	Anna Chlumsky, Austin O'Brien, Dan Aykroyd, Jamie Lee Cur
6	s427	Movie	Cousins	Ainsley Gardiner, Briar Grace-Smith	Rachel House, Briar Grace-Smith, Tanea Heke, Tioreore Nga

Total rows: 118 | Query complete 00:00:00.101 | CRLF | Ln 77, Col 1

SQL QUERIES

06

Content Added in Last 5 Years

Objective: Retrieve content added to Netflix in the last 5 years.

```
SELECT
    *
FROM netflix
WHERE
    TO_DATE(date_added, 'Month DD, YYYY') >= CURRENT_DATE - INTERVAL '5 years';
```

Data Output Messages Notifications				
Showing rows: 1 to 1000 Page No: 1 of 3				
	show_id character varying (6)	type character varying (10)	title character varying (150)	director character varying (208)
1	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson
2	s2	TV Show	Blood & Water	[null]
3	s3	TV Show	Ganglands	Julien Leclercq
4	s4	TV Show	Jailbirds New Orleans	[null]
5	s5	TV Show	Kota Factory	[null]
6	s6	TV Show	Midnight Mass	Mike Flanagan
Total rows: 2601 Query complete 00:00:00.191 CRLF Ln 94, Col 1				

SQL QUERIES

07

Content by Director 'Rajiv Chilaka'

Objective: List all content directed by 'Rajiv Chilaka'..

```
SELECT * FROM netflix
WHERE director ILIKE '%Rajiv Chilaka%';
```

Showing rows: 1 to 22 Page No: 1 of 1					
	show_id character varying (6)	type character varying (10)	title character varying (150)	director character varying (208)	casts character varying (1000)
1	s407	Movie	Chhota Bheem - Neeli Pahaadi	Rajiv Chilaka	Vatsal Dubey, Julie Tejjwani, Rupa Bhimani, Jig
2	s408	Movie	Chhota Bheem & Ganesh	Rajiv Chilaka	Vatsal Dubey, Julie Tejjwani, Rupa Bhimani, Jig
3	s409	Movie	Chhota Bheem & Krishna: Mayanagari	Rajiv Chilaka	Vatsal Dubey, Julie Tejjwani, Rupa Bhimani, Jig
4	s410	Movie	Chhota Bheem & Krishna: Pataliputra- City of the Dead	Rajiv Chilaka	Vatsal Dubey, Julie Tejjwani, Rupa Bhimani, Jig
5	s411	Movie	Chhota Bheem And The Broken Amulet	Rajiv Chilaka	Vatsal Dubey, Julie Tejjwani, Rupa Bhimani, Jig
6	s412	Movie	Chhota Bheem And The Crown of Valhalla	Rajiv Chilaka	Vatsal Dubey, Julie Tejjwani, Rupa Bhimani, Jig
Total rows: 22 Query complete 00:00:00.072 CRLF Ln 95, Col 1					

SQL QUERIES

08

TV Shows with More than 5 Seasons

Objective: Identify TV shows with more than 5 seasons.

```
SELECT * FROM netflix
WHERE
    type = 'TV Show'
    AND
    SPLIT_PART(duration, ' ', 1)::numeric > 5;

-- 9. Count the number of content items in each genre.
```

Data Output

Messages

Notifications

<

SQL QUERIES

09

Genre Distribution

Objective: Count the number of content items in each genre.

```
SELECT
  UNNEST(STRING_TO_ARRAY(listed_in, ',')) AS genre,
  COUNT(show_id) AS total_content
FROM netflix
GROUP BY 1;
```

	genre text	total_content bigint
1	International TV Shows	774
2	Stand-Up Comedy	9
3	Spanish-Language TV Shows	172
4	Romantic Movies	613
5	Anime Features	21
6	TV Horror	11
7	Stand-Up Comedy & Talk Shows	34
Total rows: 73		Query complete 00:00:00.064

SQL QUERIES

10 Avg Content Release in India (Top 5 Years)
Objective: Calculate and rank years by the average number of content releases by India.

```
SELECT
  EXTRACT(YEAR FROM TO_DATE(date_added, 'Month DD, YYYY')) AS year,
  COUNT(*) AS yearly_content,
  ROUND(
    COUNT(*)::numeric/(SELECT COUNT(*) FROM netflix WHERE country = 'India')::numeric * 100
  ,2) AS avg_count_per_year
FROM netflix
WHERE country = 'India'
GROUP BY 1;
```

	year numeric	yearly_content bigint	avg_count_per_year numeric
1	2018	333	34.26
2	2016	10	1.03
3	2019	203	20.88
4	2021	95	9.77
5	2020	189	19.44
6	2017	142	14.61
Total rows: 6 Query complete 00:00:00.063			

SQL QUERIES

11

Documentary Movies

Objective: Retrieve all movies classified as documentaries.

```
SELECT * FROM netflix
WHERE
    listed_in ILIKE '%documentaries%';
```

Data Output Messages Notifications				
Showing rows: 1 to 869 Page No: 1 of 1				
	show_id character varying (6)	type character varying (10)	title character varying (150)	director character varying (208)
1	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson
2	s17	Movie	Europe's Most Dangerous Man: Otto Skorzeny in Spain	Pedro de Echave García, Pablo Azorín Williams
3	s46	Movie	My Heroes Were Cowboys	Tyler Greco
4	s69	Movie	Schumacher	Hanns-Bruno Kammertöns, Vanessa Nöcker, Michael '
5	s89	Movie	Blood Brothers: Malcolm X & Muhammad Ali	Marcus Clarke
6	s92	Movie	The Women and the Murderer	Mona Achache, Patricia Tourancheau
Total rows: 869 Query complete 00:00:00.134 CRLF Ln 127, Col 1				

SQL QUERIES

12

Entries Without a Director
Objective: List content that does not have a director.

```
SELECT * FROM netflix
WHERE
    director IS NULL;
```

Data Output Messages Notifications				
Showing rows: 1 to 869 Page No: 1 of 1				
	show_id character varying (6)	type character varying (10)	title character varying (150)	director character varying (208)
1	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson
2	s17	Movie	Europe's Most Dangerous Man: Otto Skorzeny in Spain	Pedro de Echave García, Pablo Azorín Williams
3	s46	Movie	My Heroes Were Cowboys	Tyler Greco
4	s69	Movie	Schumacher	Hanns-Bruno Kammertöns, Vanessa Nöcker, Michael
5	s89	Movie	Blood Brothers: Malcolm X & Muhammad Ali	Marcus Clarke
6	s92	Movie	The Women and the Murderer	Mona Achache, Patricia Tourancheau
Total rows: 869 Query complete 00:00:00.134 CRLF Ln 140, Col 22				

SQL QUERIES

13

Salman Khan Movies (Last 10 Years)

Objective: Count the number of movies featuring 'Salman Khan' in the last 10 years.

```
SELECT * FROM netflix
WHERE
    casts ILIKE '%Salman Khan%'
    AND
    release_year > EXTRACT(YEAR FROM CURRENT_DATE) - 10;
```

Data Output Messages Notifications								
Showing rows: 1 to 1 Page No: 1 of 1								
	show_id character varying (6)	type character varying (10)	title character varying (150)	director character varying (208)	casts character varying (1000)	country character varying (150)	date_added character varying (50)	release_year integer
1	s2757	Movie	Paharganj	Rakesh Ranjan Kumar	Lorena Franco, Bijesh Jayar...	India	March 31, 2020	2019










SQL QUERIES

14

Top 10 Indian Actors

Objective: Identify the top 10 actors with the most appearances in Indian-produced movies.

```
SELECT
UNNEST(STRING_TO_ARRAY(casts, ',')) AS actors,
COUNT(*) AS total_content
FROM netflix
WHERE country ILIKE '%India%'
GROUP BY 1
ORDER BY 2 DESC
LIMIT 10;
```

Data Output		Messages	Notifications
			
			
	SQL		
	actors text	total_content bigint	
1	Anupam Kher	36	
2	Om Puri	26	
3	Boman Irani	25	
4	Paresh Rawal	25	
5	Shah Rukh Khan	25	
6	Akshay Kumar	23	
7	Naseeruddin Sh...	20	
Total rows: 10		Query complete 00:00:00.069	

SQL QUERIES

15

Good vs Bad Content Classification

Objective: Categorize content as 'Bad' if it contains 'kill' or 'violence' and 'Good' otherwise. Count the number of items in each category.

```
WITH new_table
AS
(
SELECT
*,
CASE
WHEN
description ILIKE '%kill%' OR
description ILIKE '%violence%' THEN 'Bad Content'
ELSE 'Good Content'
END category
FROM netflix
)
SELECT
category,
COUNT(*) AS total_content
FROM new_table
GROUP BY 1;
```

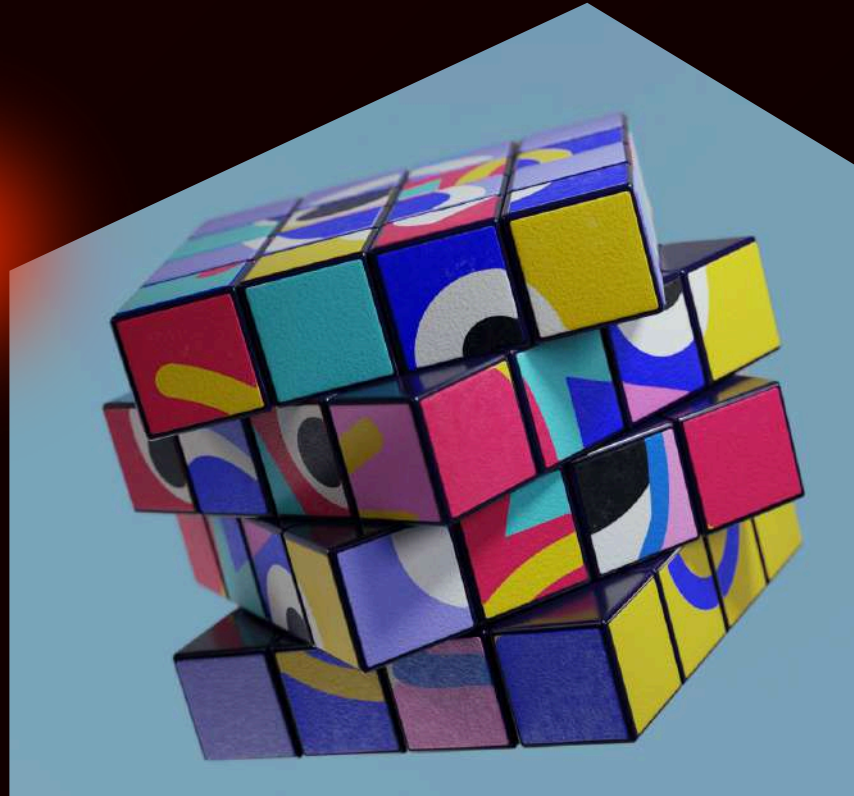
	category	total_content
	text	bigint
1	Good Content	8465
2	Bad Content	342



BUSINESS INSIGHTS

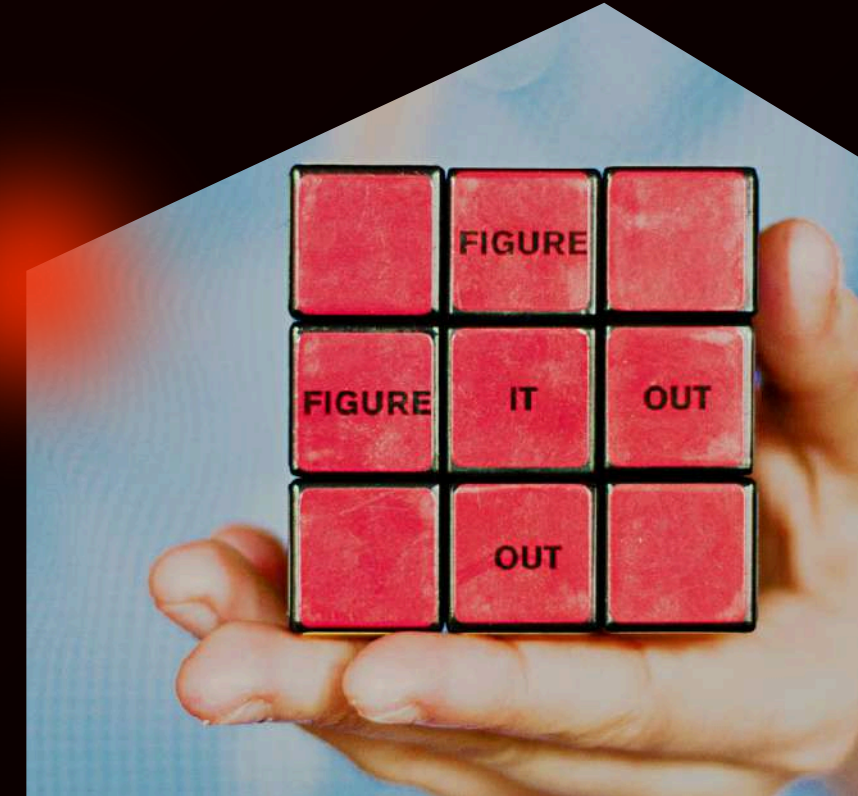
- Movies dominate Netflix's content catalog.
- Most common rating: TV-MA and TV-14.
- India and the US are the biggest contributors of Netflix content.
- Documentaries and dramas rank high in genre popularity.
- Regional and director-specific insights reveal niche content production patterns.

CHALLENGES & SOLUTIONS



Challenges

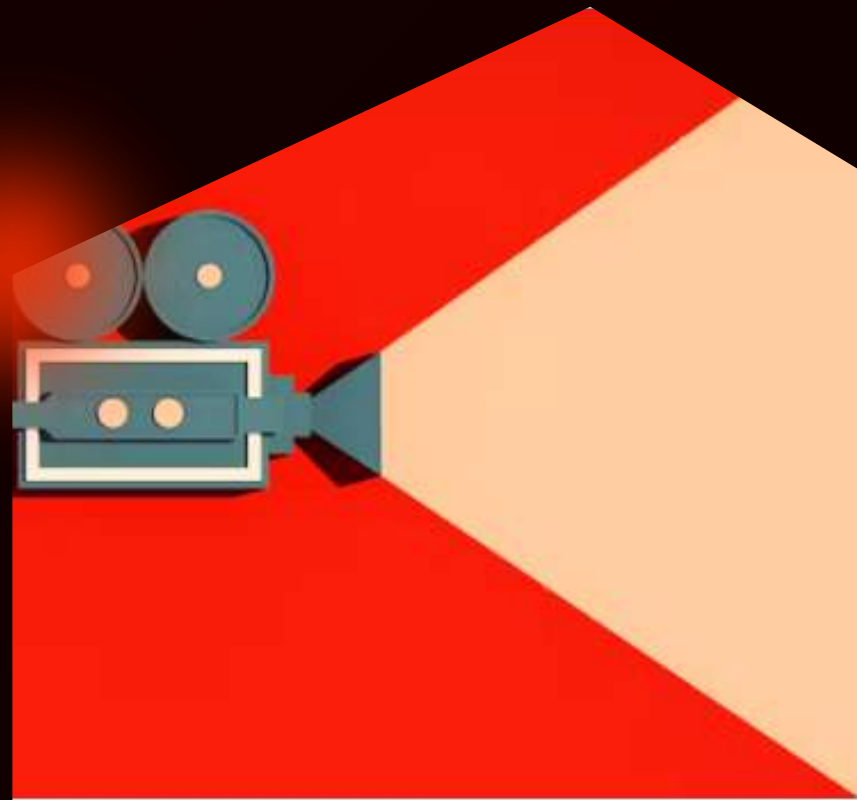
- Multi-value fields (e.g., country, genre) required string parsing.
- Inconsistent date formats and missing data.
- Complex keyword filtering for content categorization.



Solutions

- Used UNNEST, STRING_TO_ARRAY for field normalization.
- Applied TO_DATE for parsing inconsistent dates.
- Implemented CASE logic with ILIKE for content labeling.

KEY OUTCOMES & CONCLUSION



Key Outcomes

- Clean and enriched dataset for future reporting.
- Predefined queries to drive BI dashboards.
- Established framework for keyword-based content labeling.
- Identification of high-volume production regions and actors.



Conclusion

- SQL is an efficient tool for structured data exploration.
- Content insights can guide business, marketing, and acquisition strategies.
- The project delivers a ready-to-integrate query suite for analytics teams.



THANK YOU



 bhaskarpal.official@gmail.com