

# NETFLIX ANALYSIS USING SQL



# Overview

This project involves a comprehensive analysis of Netflix's movies and TV shows data using SQL. The goal is to extract valuable insights and answer various business questions based on the dataset. The following details the project's objectives, business problems, solutions, findings, and conclusions.



# Objectives

1

Analyze the distribution  
of content types  
(movies vs TV shows).

2

Identify the most  
common ratings for  
movies and TV shows.

3

List and analyze content  
based on release years,  
countries, and durations.

4

Explore and categorize  
content based on specific  
criteria and keywords.





# SCHHEMA



-- SCHEMAS of Netflix

**CREATE TABLE** netflix

(  
    show\_id **VARCHAR(5)**,  
    **type** **VARCHAR(10)**,  
    title **VARCHAR(250)**,  
    director **VARCHAR(550)**,  
    casts **VARCHAR(1050)**,  
    country **VARCHAR(550)**,  
    date\_added **VARCHAR(55)**,  
    release\_year **INT**,  
    rating **VARCHAR(15)**,  
    duration **VARCHAR(15)**,  
    listed\_in **VARCHAR(250)**,  
    description **VARCHAR(550)**  
);

**SELECT** \* **FROM** netflix;



# Business Problems and Solutions



# Problem # 1. Count the number of Movies vs TV Shows

```
SELECT  
    type,  
    COUNT(*)  
FROM netflix  
GROUP BY 1
```



type	count
Movie	6131
TV Show	2676



# Problem # 2. Find the most common rating for movies and TV shows

Select type, rating

From

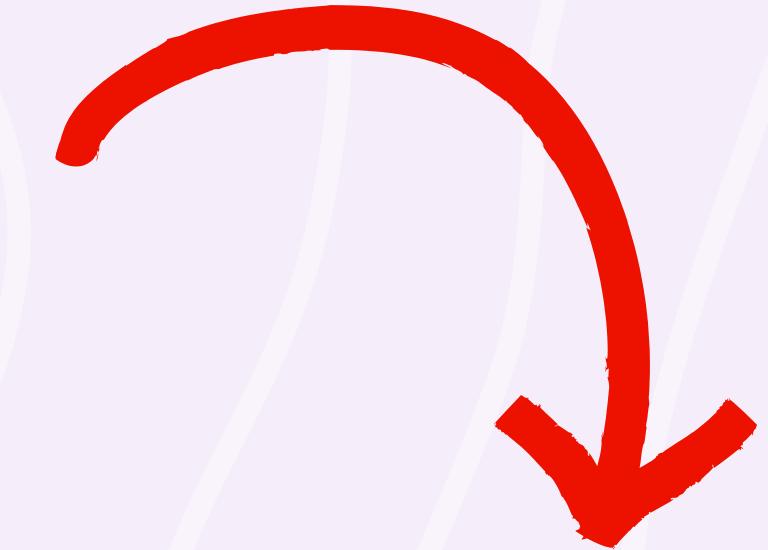
(

Select type, rating, Count(\*),  
Rank() OVER(PARTITION BY type Order BY  
COUNT (\*) DESC ) as ranking

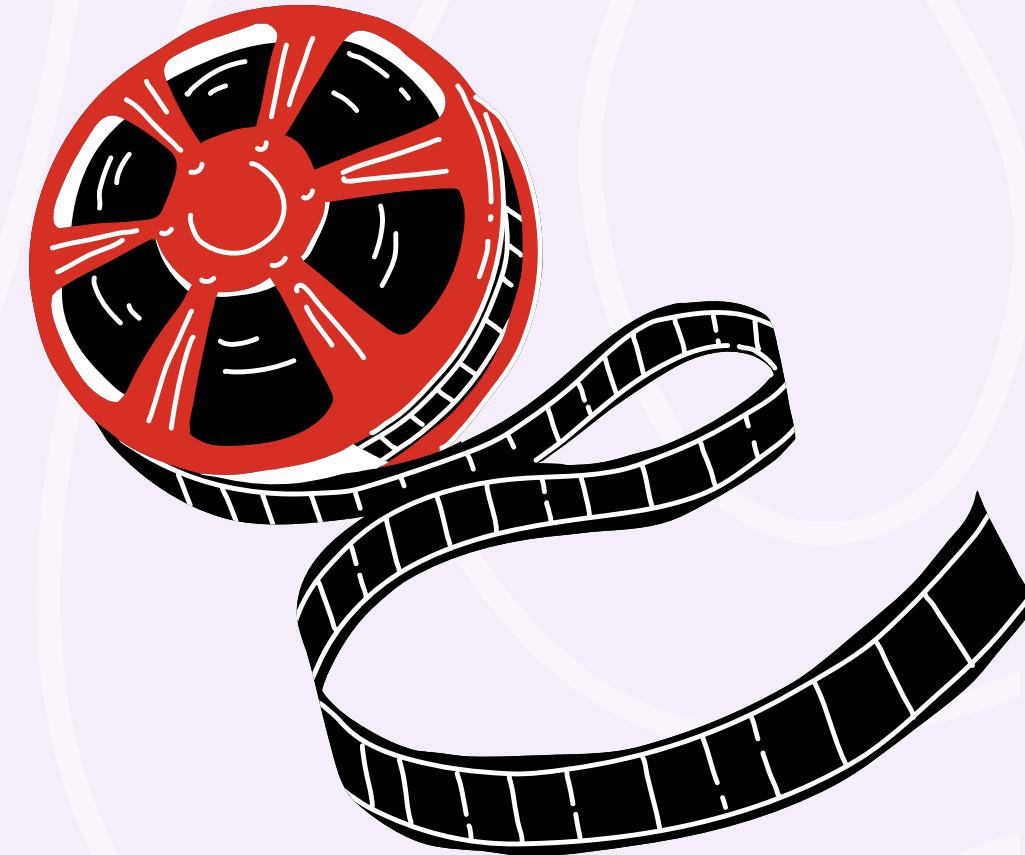
From netflix Group BY 1,2

) as t1

where ranking = 1



type	rating
character varying (10)	character varying (15)
Movie	TV-MA
TV Show	TV-MA

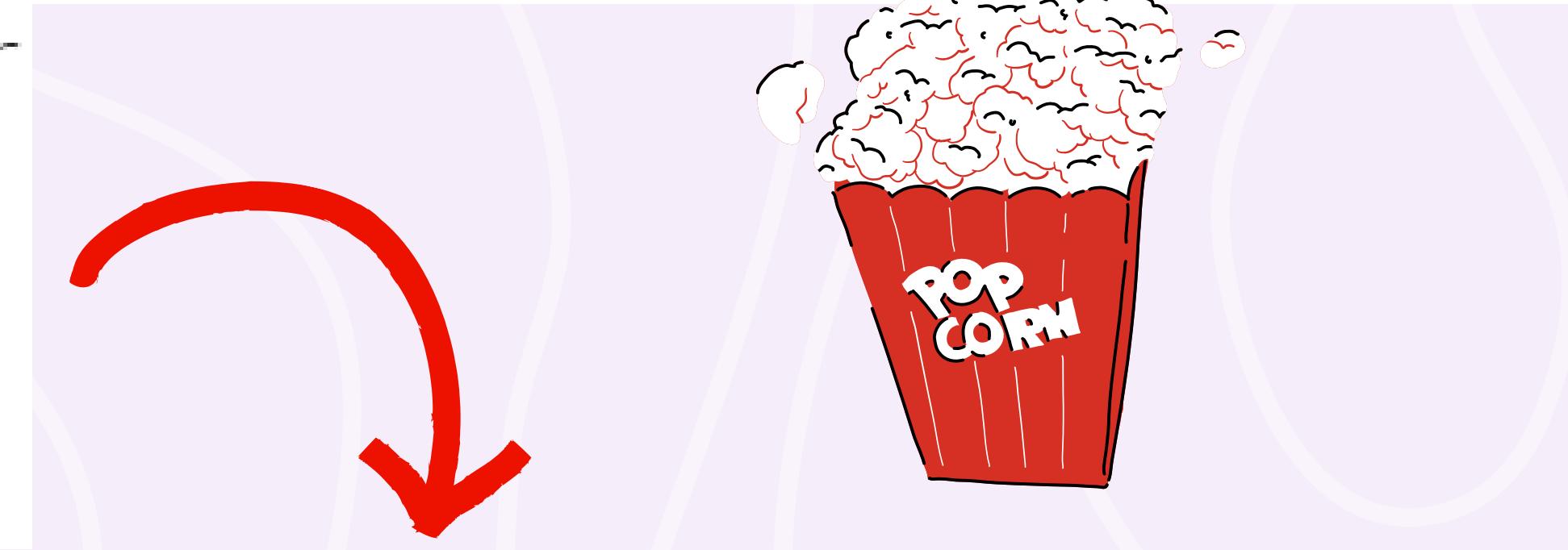


# Problem # 3. List all movies released in a specific year (e.g., 2020)

```
SELECT type, title, release_year  
FROM netflix  
WHERE type = 'Movie'  
AND  
release_year = 2020
```



type character varying (10) 	title character varying (250)	 release_year integer 
Movie	Dick Johnson Is Dead	2020
Movie	Europe's Most Dangerous Man: Otto Skorzeny in Spain	2020
Movie	Tughlaq Durbar	2020
Movie	Omo Ghetto: the Saga	2020
Movie	Shadow Parties	2020
Movie	Here and There	2020
Movie	Shikara	2020
Movie	Really Love	2020
Movie	The Old Ways	2020



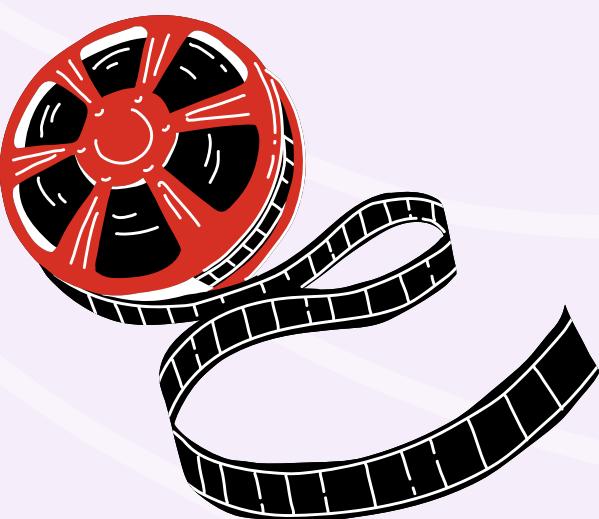
# Problem # 4. Find the top 5 countries with the most content on Netflix

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	total_content
United States	3211
India	1008
United Kingdo...	628
United States	479
Canada	271

# Problem # 5. Identify the longest movie

```
SELECT type, title, duration  
FROM netflix  
where  
type = 'Movie'  
AND  
duration = (Select MAX(duration) from netflix)
```



type	title	duration
character varying (10)	character varying (250)	character varying (15)
Movie	InuYasha the Movie 2: The Castle Beyond the Looking Glass	99 min
Movie	InuYasha the Movie 3: Swords of an Honorable Ruler	99 min
Movie	Here and There	99 min
Movie	Good Luck Chuck	99 min
Movie	My Girl 2	99 min
Movie	Cousins	99 min
Movie	Rogue Warfare: Death of a Nation	99 min
Movie	Copenhagen	99 min
Movie	Black Holes   The Edge of All We Know	99 min

# Problem # 6. Find content added in the last 5 years

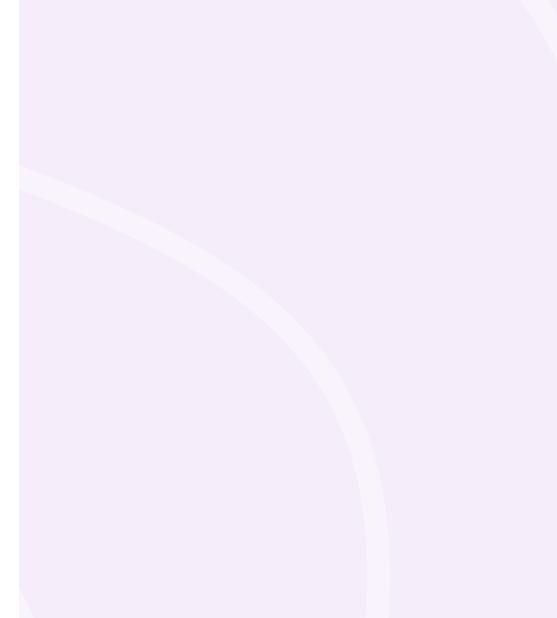
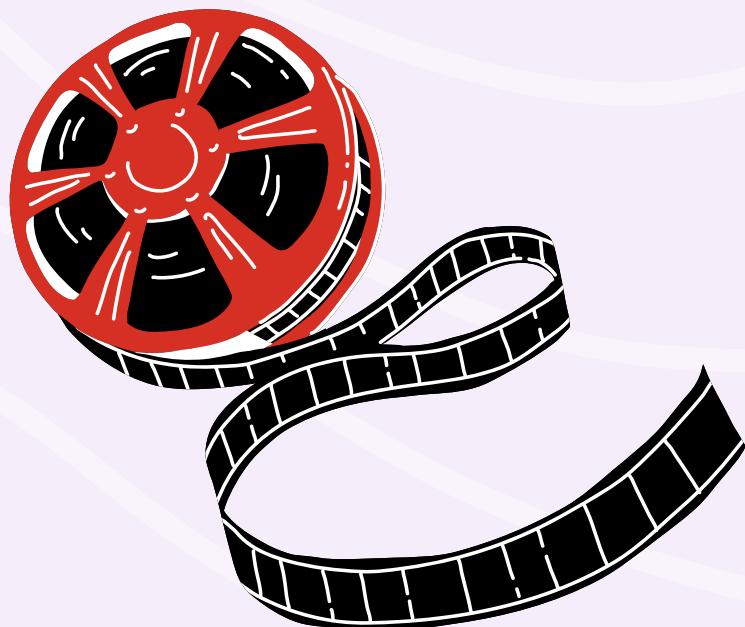
```
SELECT title, date_added, type FROM netflix  
WHERE  
TO_DATE(date_added, 'Month DD, YYYY')  
>= CURRENT_DATE - INTERVAL '5 years'
```



title	date_added	type
character varying (250)	character varying (55)	character varying (10)
Dick Johnson Is Dead	September 25, 2021	Movie
Blood & Water	September 24, 2021	TV Show
Ganglands	September 24, 2021	TV Show
Jailbirds New Orleans	September 24, 2021	TV Show
Kota Factory	September 24, 2021	TV Show
Midnight Mass	September 24, 2021	TV Show
My Little Pony: A New Generation	September 24, 2021	Movie
Sankofa	September 24, 2021	Movie
The Great British Baking Show	September 24, 2021	TV Show
The Starling	September 24, 2021	Movie
Vendetta: Truth, Lies and The Mafia	September 24, 2021	TV Show

# Problem # 7. Find all the movies/TV shows by director 'Rajiv Chilaka'!

```
SELECT title, type, director  
FROM  
(SELECT *,  
    UNNEST(STRING_TO_ARRAY(director, ', '))  
    as director_name FROM netflix )  
  
WHERE  
    director_name = 'Rajiv Chilaka'
```



title	type	director
character varying (250)	character varying (10)	Rajiv Chilaka
Chhota Bheem - Neeli Pahaadi	Movie	Rajiv Chilaka
Chhota Bheem & Ganesh	Movie	Rajiv Chilaka
Chhota Bheem & Krishna: Mayanagari	Movie	Rajiv Chilaka
Chhota Bheem & Krishna: Pataliputra- City of the Dead	Movie	Rajiv Chilaka
Chhota Bheem And The Broken Amulet	Movie	Rajiv Chilaka
Chhota Bheem And The Crown of Valhalla	Movie	Rajiv Chilaka
Chhota Bheem and the Incan Adventure	Movie	Rajiv Chilaka
Chhota Bheem and The ShiNobi Secret	Movie	Rajiv Chilaka
Chhota Bheem Aur Hanuman	Movie	Rajiv Chilaka
Chhota Bheem aur Krishna	Movie	Rajiv Chilaka
Chhota Bheem aur Krishna vs Zimbara	Movie	Rajiv Chilaka

# Problem # 8. List all TV shows with more than 5 seasons

```
SELECT title, duration FROM netflix
```

```
WHERE
```

```
    TYPE = 'TV Show'
```

```
AND
```

```
SPLIT_PART(duration, ' ', 1)::INT > 5
```

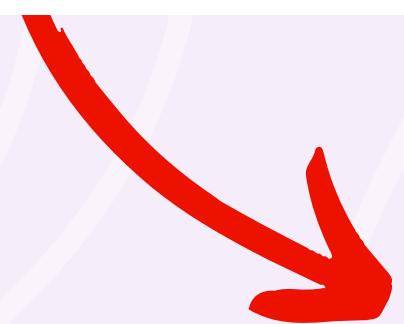
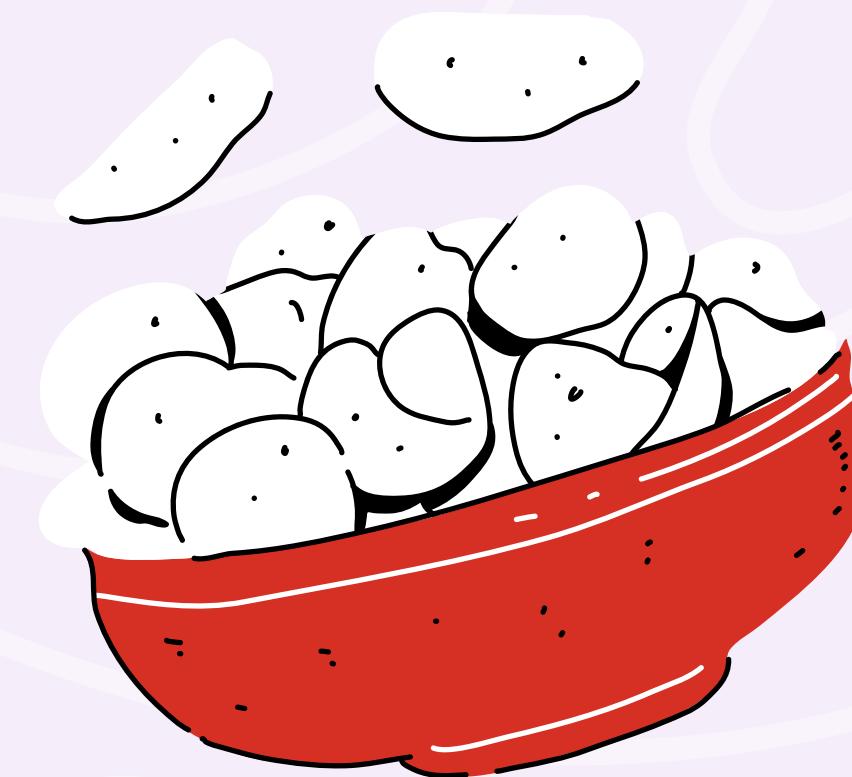


title	duration
character varying (250)	character varying (15)
The Great British Baking Show	9 Seasons
Nailed It	6 Seasons
Numberblocks	6 Seasons
Saved by the Bell	9 Seasons
Lucifer	6 Seasons
Grace and Frankie	7 Seasons
30 Rock	7 Seasons
Hunter X Hunter (2011)	6 Seasons
The Flash	7 Seasons
The Walking Dead	10 Seasons

# Problem # 9. Count the number of content items in each genre

SELECT

```
UNNEST(STRING_TO_ARRAY(listed_in, ','))  
AS genre,  
COUNT(*) AS total_content  
FROM netflix  
GROUP BY genre  
ORDER BY total_content DESC  
LIMIT 5;
```



genre	total_content
text	
International Movies	2624
Dramas	1600
Comedies	1210
Action & Adventure	859
Documentaries	829

# Problem # 10. Find each year and the average numbers of content release by United States on netflix. Return top 5 year with highest avg content release !

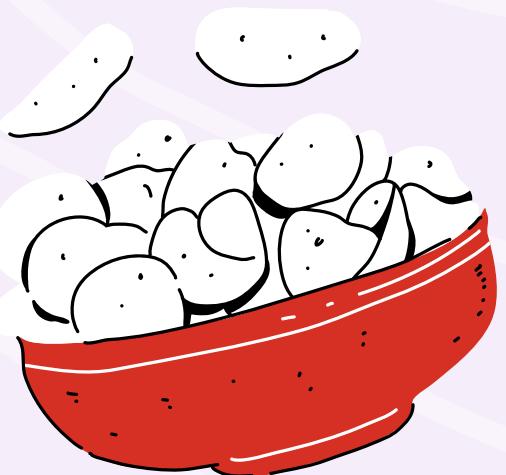
```
SELECT country, release_year,  
       COUNT(show_id) as total_release,  
       ROUND( COUNT(show_id)::numeric/  
              (SELECT COUNT(show_id) FROM netflix  
               WHERE country = 'United States')::numeric * 100  
              ,2) as avg_release  
  
FROM netflix  
WHERE country = 'United States'  
GROUP BY country, 2  
ORDER BY avg_release DESC  
LIMIT 5
```



country character varying (550)	release_year integer	total_release bigint	avg_release numeric
United States	2018	356	12.63
United States	2017	352	12.49
United States	2019	351	12.46
United States	2020	336	11.92
United States	2016	263	9.33

# Problem # 11. List all movies that are documentaries

```
SELECT title, listed_in FROM netflix  
WHERE type = 'Movie'  
AND  
listed_in LIKE '%Documentaries'
```



title	listed_in
character varying (250)	Documentaries
Dick Johnson Is Dead	Documentaries
My Heroes Were Cowboys	Documentaries
Final Account	Documentaries
NOVA: Chinese Chariot Revealed	Documentaries
Bob Ross: Happy Accidents, Betrayal & Greed	Documentaries
Lady Boss: The Jackie Collins Story	Documentaries
NOVA: Day the Dinosaurs Died	Documentaries
Fantastic Fungi	Documentaries
9to5: The Story of a Movement	Documentaries
This Changes Everything	Documentaries
NOVA: Death Dive to Saturn	Documentaries

# Problem # 12. Find all content without a director

```
SELECT title, director FROM netflix
```

```
WHERE director IS NULL
```



title	director
character varying (250)	[null]
Blood & Water	[null]
Jailbirds New Orleans	[null]
Kota Factory	[null]
Vendetta: Truth, Lies and The Mafia	[null]
Crime Stories: India Detectives	[null]
Dear White People	[null]
Khawatir	[null]
Falsa identidad	[null]
Jaguar	[null]
Resurrection: Ertugrul	[null]
Love on the Spectrum	[null]
Chicago Party Aunt	[null]
Sex Education	[null]
Squid Game	[null]
Tayo and Little Wizards	[null]

# Problem # 13. Find how many movies actor 'Keanu Reeves' (JOHN WICK) appeared in last 10 years!

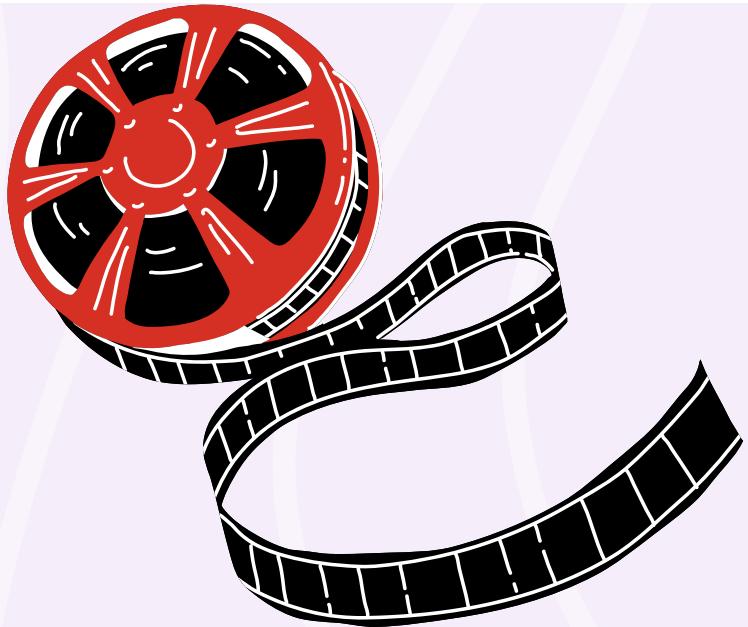
```
SELECT type, title, release_year FROM netflix  
WHERE  
    casts LIKE '%Keanu Reeves%' -- JOHN WICK  
    AND  
    release_year > EXTRACT(YEAR FROM CURRENT_DATE) - 10
```



type	title	release_year
character varying (10)	character varying (250)	integer
Movie	The Whole Truth	2016
Movie	SPF-18	2017
Movie	The Bad Batch	2016
Movie	To the Bone	2017
Movie	Knock Knock	2015

# Problem # 14. Find the top 5 actors who have appeared in the highest number of movies produced in United States.

```
SELECT  
    UNNEST(STRING_TO_ARRAY(casts, ',')) as actor,  
    COUNT(*)  
FROM netflix  
WHERE country = 'United States'  
GROUP BY 1  
ORDER BY 2 DESC  
LIMIT 5
```



actor	count
Adam Sandler	20
Fred Tatasciore	15
Molly Shannon	14
Erin Fitzgerald	13
Samuel L. Jackson	13

# Problem # 15

Categorize the content based on the presence of the keywords 'kill' and 'violence' in the description field. Label content containing these keywords as 'Bad' and all other content as 'Good'. Count how many items fall into each category.



```

SELECT category, TYPE,
       COUNT(*) AS content_count
FROM (SELECT *, CASE
      WHEN description ILIKE '%kill%' OR
      description ILIKE '%violence%' THEN 'Bad'
      ELSE 'Good'
      END AS category
      FROM netflix
    ) AS categorized_content
GROUP BY 1,2
ORDER BY 2

```



category	type	content_count
Bad	Movie	251
Good	Movie	5880
Bad	TV Show	91
Good	TV Show	2585





# Thank You

---



+92 3176063654



[linkedin.com/in/waleed-ilyas-664839213](https://www.linkedin.com/in/waleed-ilyas-664839213)



[waleedilyas99@gmail.com](mailto:waleedilyas99@gmail.com)

