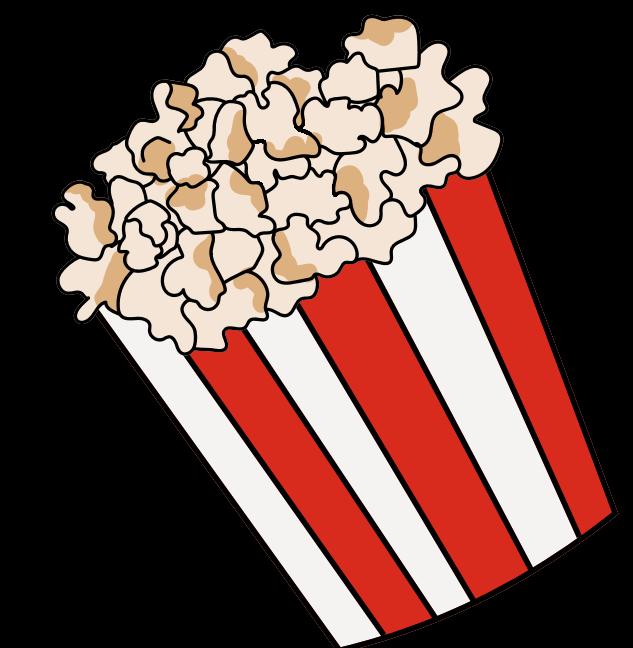
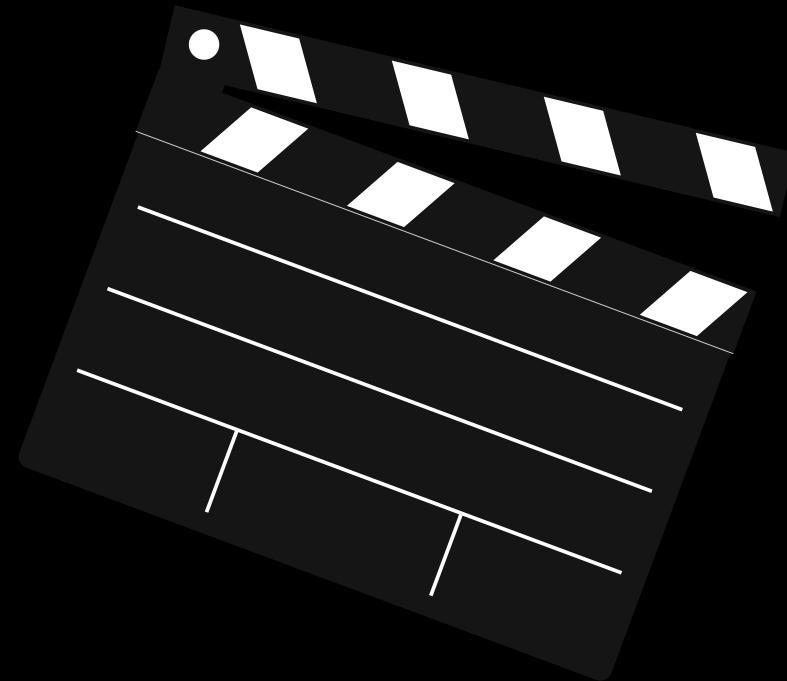
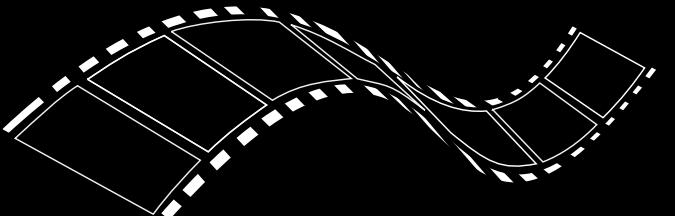
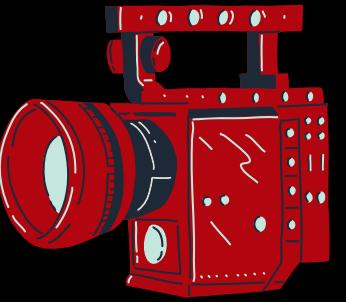


# NETFLIX DATA ANALYSIS SQL PROJECT

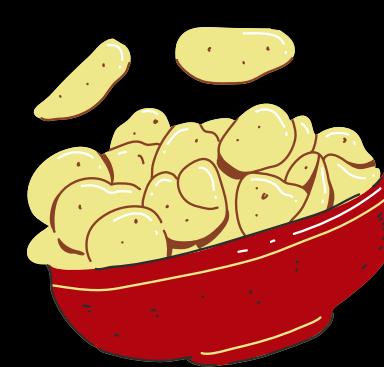
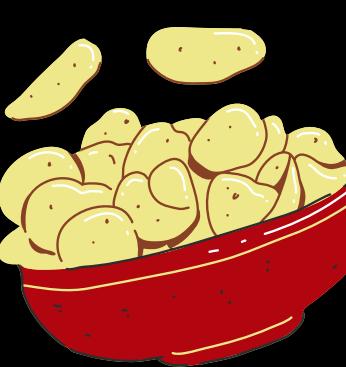
BY VEER SINGH  
IIM AMRITSAR



# INTRODUCTION



This project, titled "Netflix Movies and TV Shows Data Analysis Using SQL", aims to perform an in-depth exploration of Netflix's vast catalog of movies and TV shows. By leveraging SQL, we will extract key insights from the dataset to help answer a variety of business questions that reflect the streaming giant's content distribution, trends, and patterns.



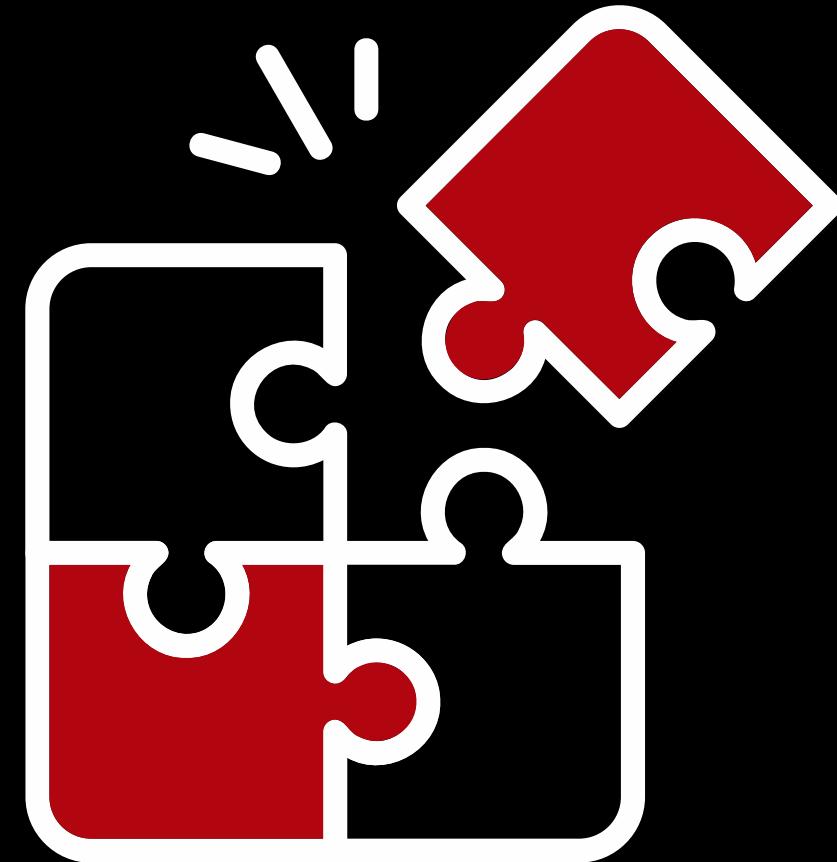
The primary goal is to analyze the Netflix content library to discover meaningful insights, such as identifying the distribution between movies and TV shows, understanding content ratings, categorizing content by keywords, and assessing geographical representation across the platform. This analysis is crucial for businesses, as it helps drive strategic decisions around content acquisition, marketing, and audience engagement.

In this project, SQL will be used to analyze content across multiple dimensions such as content type, release year, duration, genre, and country of origin. The analysis will reveal patterns and trends that provide a better understanding of Netflix's global library, giving us a clear picture of content diversity, viewer preferences, and release frequency over time.

# BUSINESS PROBLEMS

The analysis addresses a series of business problems, including the following key questions:

1. Counting the number of Movies vs. TV Shows on Netflix.
2. Identifying the most common ratings for both movies and TV shows.
3. Listing all movies released in a particular year (e.g., 2020).
4. Identifying the top 5 countries with the most content on Netflix.
5. Finding the longest movie available on the platform.
6. Analyzing content added in the last 5 years.
7. Finding all movies or TV shows directed by 'Rajiv Chilaka.'
8. Listing all TV shows with more than 4 seasons.
9. Counting the number of content items in each genre.
10. Find number of content added in india every year.
11. Listing all documentary films.
12. Counting the number of movies actor 'Salman Khan' appeared.
13. Categorizing content based on the presence of the keywords 'kill' and 'violence' in descriptions and labeling them as 'Bad' or 'Good.' Count how many items fall into each category



# SCHEMA

CREATE TABLE NETFLIX

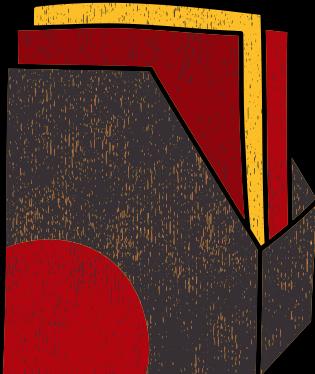
(

**SHOW\_ID VARCHAR(5) PRIMARY KEY,**  
**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)**

);



# DATASET



# COUNT THE NUMBER OF MOVIES VS TV SHOWS

```
select type, count(type) as count  
from netflix  
group by type  
order by count desc;
```

|   | type<br>character varying (10) | count<br>bigint |
|---|--------------------------------|-----------------|
| 1 | Movie                          | 6131            |
| 2 | TV Show                        | 2676            |



# FIND THE MOST COMMON RATING FOR MOVIES AND TV SHOWS

```
select t,r,c  
from  
(select t,r,c, rank() over(partition by t order by c desc) as rn  
from  
(select type as t, rating as r, count(rating) as c  
from netflix  
group by t,r) as a) as b  
where rn=1;
```

|   | t<br>character varying (10) | r<br>character varying (15) | c<br>bigint |
|---|-----------------------------|-----------------------------|-------------|
| 1 | Movie                       | TV-MA                       | 2062        |
| 2 | TV Show                     | TV-MA                       | 1145        |

# LIST ALL MOVIES RELEASED IN A SPECIFIC YEAR (E.G., 2020)

```
select title  
from netflix as n  
where n.type='Movie' and release_year=2020;
```

|   | title<br>character varying (250)                    |
|---|---|
| 1 | Dick Johnson Is Dead                                |
| 2 | Europe's Most Dangerous Man: Otto Skorzeny in Spain |
| 3 | Tughlaq Durbar                                      |
| 4 | Omo Ghetto: the Saga                                |
| 5 | Shadow Parties                                      |
| 6 | Here and There                                      |
| 7 | Shikara   |
| 8 | Really Love   |

Total rows: 517 of 517 Query complete 00:00:00.120



# FIND THE TOP 5 COUNTRIES WITH THE MOST CONTENT ON NETFLIX

```
select
    unnest(STRING_TO_ARRAY(country, ',')) as new_country, count(show_id) as count
from netflix
where country is not null
group by new_country
order by count desc limit 5;
```

|   | new_country    | count |
|---|----------------|-------|
| 1 | United States  | 3211  |
| 2 | India          | 1008  |
| 3 | United Kingdom | 628   |
| 4 | United States  | 479   |
| 5 | Canada         | 271   |

## IDENTIFY THE LONGEST MOVIE

```
select title, duration
from netflix as n
where n.type='Movie' and duration is not null
order by split_part(duration, ' ',1)::INT DESC limit 1;
```

|   | title                      | duration |
|---|----------------------------|----------|
| 1 | Black Mirror: Bandersnatch | 312 min  |

## FIND CONTENT ADDED IN THE LAST 5 YEARS

```
select title, date_added
from netflix
where to_date(date_added,'Month DD,YYYY')>= current_date - interval '5 years';
```

|   | title                            | date_added         |
|---|----------------------------------|--------------------|
| 1 | Dick Johnson Is Dead             | September 25, 2021 |
| 2 | Blood & Water                    | September 24, 2021 |
| 3 | Ganglands                        | September 24, 2021 |
| 4 | Jailbirds New Orleans            | September 24, 2021 |
| 5 | Kota Factory                     | September 24, 2021 |
| 6 | Midnight Mass                    | September 24, 2021 |
| 7 | My Little Pony: A New Generation | September 24, 2021 |
| 8 | Sankofa                          | September 24, 2021 |

Total rows: 1000 of 4080    Query complete 00:00:00.202    Ln 39, Col 1

# FIND ALL MOVIES/TV SHOWS BY DIRECTOR 'RAJIV CHILAKA'

```
select director, type, title  
from netflix  
where director LIKE '%Rajiv Chilaka%';
```

|   | director<br>character varying (550) | type<br>character varying (10) | title<br>character varying (250)                      |
|---|-------------------------------------|--------------------------------|---|
| 1 | Rajiv Chilaka                       | Movie                          | Chhota Bheem - Neeli Pahaadi                          |
| 2 | Rajiv Chilaka                       | Movie                          | Chhota Bheem & Ganesh                                 |
| 3 | Rajiv Chilaka                       | Movie                          | Chhota Bheem & Krishna: Mayanagar                     |
| 4 | Rajiv Chilaka                       | Movie                          | Chhota Bheem & Krishna: Pataliputra- City of the Dead |

Total rows: 22 of 22 Query complete 00:00:00.220 Ln 42, Col 39

# LIST ALL TV SHOWS WITH MORE THAN 4 SEASONS

```
select title, type, duration  
from netflix as n  
where n.type='TV Show' AND split_part(duration, ' ', 1)::INT >=5;
```

|   | title<br>character varying (250)       | type<br>character varying (10) | duration<br>character varying (15) |
|---|--|--------------------------------|------------------------------------|
| 1 | The Great British Baking Show          | TV Show                        | 9 Seasons                          |
| 2 | Resurrection: Ertugrul                 | TV Show                        | 5 Seasons                          |
| 3 | Nailed It                              | TV Show                        | 6 Seasons                          |
| 4 | Numberblocks                           | TV Show                        | 6 Seasons                          |
| 5 | Saved by the Bell                      | TV Show                        | 9 Seasons                          |
| 6 | Jack Whitehall: Travels with My Father | TV Show                        | 5 Seasons                          |

Total rows: 164 of 164 Query complete 00:00:00.197 Ln 47, Col 1

# COUNT THE NUMBER OF CONTENT ITEMS IN EACH GENRE

```
select  
    unnest(STRING_TO_ARRAY(listed_in, ',' )) as genre, count(show_id) as c  
from netflix  
where listed_in is not null  
group by genre  
order by c desc;
```

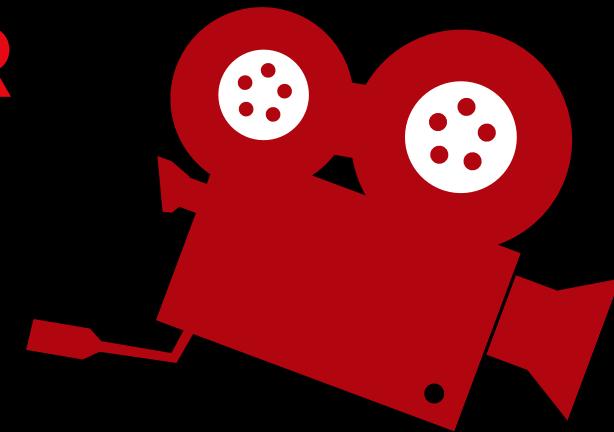
|   | genre<br>text          | c<br>bigint |
|---|------------------------|-------------|
| 1 | International Movies   | 2752        |
| 2 | Dramas                 | 2427        |
| 3 | Comedies               | 1674        |
| 4 | International TV Shows | 1351        |
| 5 | Documentaries          | 869         |
| 6 | Action & Adventure     | 859         |
| 7 | TV Dramas              | 763         |

Total rows: 42 of 42 Query complete 00:00

# FIND NUMBER OF CONTENT ADDED IN INDIA EVERY YEAR

```
select extract(year from to_date(date_added,'Month DD,YYYY')) as year, count(show_id)
from netflix
where country='India'
group by year
order by c desc;
```

|   | year<br>numeric | c<br>bigint |
|---|-----------------|-------------|
| 1 | 2018            | 333         |
| 2 | 2019            | 203         |
| 3 | 2020            | 189         |
| 4 | 2017            | 142         |
| 5 | 2021            | 95          |
| 6 | 2016            | 10          |



## LIST ALL MOVIES THAT ARE DOCUMENTARIES

```
select title
from netflix
where listed_in LIKE '%Documentaries%';
```

|   | title<br>character varying (250)                     |
|---|--|
| 1 | Dick Johnson Is Dead                                 |
| 2 | Europe's Most Dangerous Man: Otto Skorzeny in Spain  |
| 3 | My Heroes Were Cowboys                               |
| 4 | Schumacher   |
| 5 | Blood Brothers: Malcolm X & Muhammad Ali             |
| 6 | The Women and the Murderer                           |
| 7 | If I Leave Here Tomorrow: A Film About Lvnrd Skvnrvd |

Total rows: 869 of 869    Query complete 00:00:00.220

## FIND HOW MANY MOVIES ACTOR 'SALMAN KHAN' APPEARED

```
select n.type,title,casts
from netflix as n
where casts LIKE '%Salman Khan%' and n.type='Movie';
```

|   | type<br>character varying (10) | title<br>character varying (250) | casts<br>character varying (1050)                    |
|---|--------------------------------|----------------------------------|--|
| 1 | Movie                          | Hello Brother                    | Salman Khan, Arbaaz Khan, Rani Mukerji, Johny Le     |
| 2 | Movie                          | Bodyguard                        | Salman Khan, Kareena Kapoor, Raj Babbar, Asrani,     |
| 3 | Movie                          | Prem Ratan Dhan Payo             | Salman Khan, Sonam Kapoor, Anupam Kher, Neil N       |
| 4 | Movie                          | Andaz Apna Apna                  | Aamir Khan, Salman Khan, Raveena Tandon, Karis       |
| 5 | Movie                          | Kuch Kuch Hota Hai               | Shah Rukh Khan, Kajol, Rani Mukerji, Farida Jalal, P |
| 6 | Movie                          | Paharganj                        | Lorena Franco, Bijesh Jayarajan, Neet Chowdhary,     |

Total rows: 20 of 20    Query complete 00:00:00.193    Ln 71, Col 51



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

```
With new_table
as
(
select *,
  case
  when
    description ilike '%kill%' or
    description ilike '%violence%'
  Then 'Bad'
  else 'Good'
  END Category
from netflix)
Select Category, count(show_id) as total_count
From new_table
group by category;
```

|   | category | total_count |
|---|----------|-------------|
|   | text     | bigint      |
| 1 | Good     | 8465        |
| 2 | Bad      | 342         |

THANK YOU! N

