
NETFLIX

ABOUT US

TO HELP YOU DISCOVER SOMETHING GREAT EVERY TIME, NO MATTER WHAT YOU'RE LOOKING FOR. FANDOM. WHEN OUR SERIES AND FILMS BECOME CULTURAL MOMENTS

ABOUT ME

MITALI GUPTA, A PROFICIENT DATA ANALYST, DEVELOPED AN SQL PROJECT ON NETFLIX USING POSTGRESQL TO ANALYZE STREAMING DATA AND USER BEHAVIOR. THE PROJECT INVOLVED QUERYING LARGE DATASETS TO EXTRACT INSIGHTS ON CONTENT TRENDS, VIEWER PREFERENCES, AND PLATFORM PERFORMANCE. SHE UTILIZED ADVANCED SQL TECHNIQUES SUCH AS WINDOW FUNCTIONS, CTES, AND JOINS TO ANALYZE MOVIE GENRES, RATINGS, AND REGIONAL POPULARITY. THE PROJECT PROVIDED VALUABLE INSIGHTS INTO NETFLIX'S CONTENT STRATEGY, HELPING TO UNDERSTAND FACTORS INFLUENCING USER ENGAGEMENT AND RETENTION.

COUNT THE NUMBER OF MOVIES VS TV SHOWS

```sql
select show_type,
    count(show_id) as "total_of content"
    from netflix
    group by 1
```

FIND THE MOST COMMON RATING FOR MOVIES AND TV SHOWS

```sql
select * from
    (select show_type,
    rating,
    count(show_id) as "total_content",
    rank() over(partition by show_type order by count(show_id) desc) as "ranking"
    from netflix
    group by 1,2)
    where ranking = 1
```

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

```sql
select * from netflix
    where show_type = 'Movie'
    and release_year = 2020
```

FIND THE TOP 5 COUNTRIES WITH THE MOST CONTENT ON NETFLIX

```
select unnest(string_to_array(country, ',')),
    count(show_id) as "total_content"
    from netflix
    group by 1
    order by 2 desc
    limit 5
```

IDENTIFY THE LONGEST MOVIE

```sql
select show_type,
title,
duration
from netflix
where show_type = 'Movie'
and duration = (select max(duration) from netflix)
```

FIND CONTENT ADDED IN THE LAST 5 YEARS

```
select show_type,
title,
date_added:: date from netflix
where date_added:: date > current_date - interval '5 years'
```

FIND ALL THE MOVIES/TV SHOWS BY DIRECTOR 'RAJIV CHILAKA'!

```sql
select show_type,
director
from netflix
where director like '%Rajiv Chilaka%'
```

LIST ALL TV SHOWS WITH MORE THAN 5 SEASONS

```sql
select * from
    (select show_type,
    title,
    duration,
    split_part(duration,' ',1):: int as "season"
    from netflix
    where show_type = 'TV Show')
where season > 5
```

COUNT THE NUMBER OF CONTENT ITEMS IN EACH GENRE

```sql
select show_type,
unnest(string_to_array(listed_in,',')) as "type_of_content",
count(show_id)
from netflix
group by 1,2
order by 3 desc
```

FIND EACH YEAR AND THE AVERAGE NUMBERS OF CONTENT RELEASE IN INDIA ON NETFLIX RETURN TOP 5 YEAR WITH HIGHEST AVG CONTENT RELEASE!

```sql
select Extract('Year' from date_added:: date),
    count(),
    round(count()::numeric/(select count(*)
    from netflix
    where country = 'India'),2) * 100 as "Avg_content"
    from netflix
    where country = 'India'
    group by 1
```

LIST ALL MOVIES THAT ARE DOCUMENTARIES

```
select * from
    (select show_id,
    show_type,
    title,
    unnest(string_to_array(listed_in,',')) as "genre"
    from netflix
    where show_type = 'Movie')
where genre = 'Documentaries'
```

FIND ALL CONTENT WITHOUT A DIRECTOR

```
select show_type,
title,
director
from netflix
where director is null
```

FIND HOW MANY MOVIES ACTOR 'SALMAN KHAN' APPEARED IN LAST 10 YEARS!

```
select title,
    show_cast,
    release_year
    from netflix
    where show_cast like '%Salman Khan%' and
    release_year :: numeric >= Extract('Year' from current_date - interval '10 years' )
```

FIND THE TOP 10 ACTORS WHO HAVE APPEARED IN THE HIGHEST NUMBER OF MOVIES PRODUCED IN INDIA.

```sql
select unnest(string_to_array(show_cast,',')),
count(show_id)
from netflix
where country like 'India%'
group by 1
order by 2 desc
limit 10
```

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.

```sql
select content_based,
    count(*) from
    (select show_type,
    title,
    descriptions,
        case
            when descriptions Ilike '%kill%' then 'bad'
            when descriptions Ilike '%violence%' then 'bad'
            else 'good'
        end as content_based
    from netflix)
    group by 1
```

THANK YOU FOR YOUR ATTENTION