Netflix Dataset Analysis

Netflix is one of the world's leading streaming platforms, offering a vast collection of TV shows, movies, documentaries, and more. Understanding its content distribution, trends, and insights can provide valuable information about viewing patterns, genre popularity, and production decisions.

In this project, we analyze a Netflix dataset using SQL to extract meaningful insights. The dataset contains information such as titles, genres, release years, duration, ratings, and types (TV Show or Movie). By writing SQL queries, we perform various analyses.

This analysis helps in understanding Netflix's content strategy and how its library has evolved over the years. Through SQL queries, we efficiently retrieve, manipulate, and analyze data to derive actionable insights.

Introduction and Goals

Project Overview

Deep dive into Netflix's dataset to uncover key insights and trends in the streaming industry.

Primary Objective

Identify patterns in content performance, user engagement, and market trends.

Methodology

Employ data exploration, statistical analysis, and predictive modeling techniques.

Table of Contents

1 **Introduction**

Understanding Aim of the project ,dataset structure and key metrics.

2 **Technology Used**

Tools and technologies used for analysis.

3 **Insights**

Discoveries from Netflix's content and user data.

4 **Conclusion**

Overall conclusion of this project.
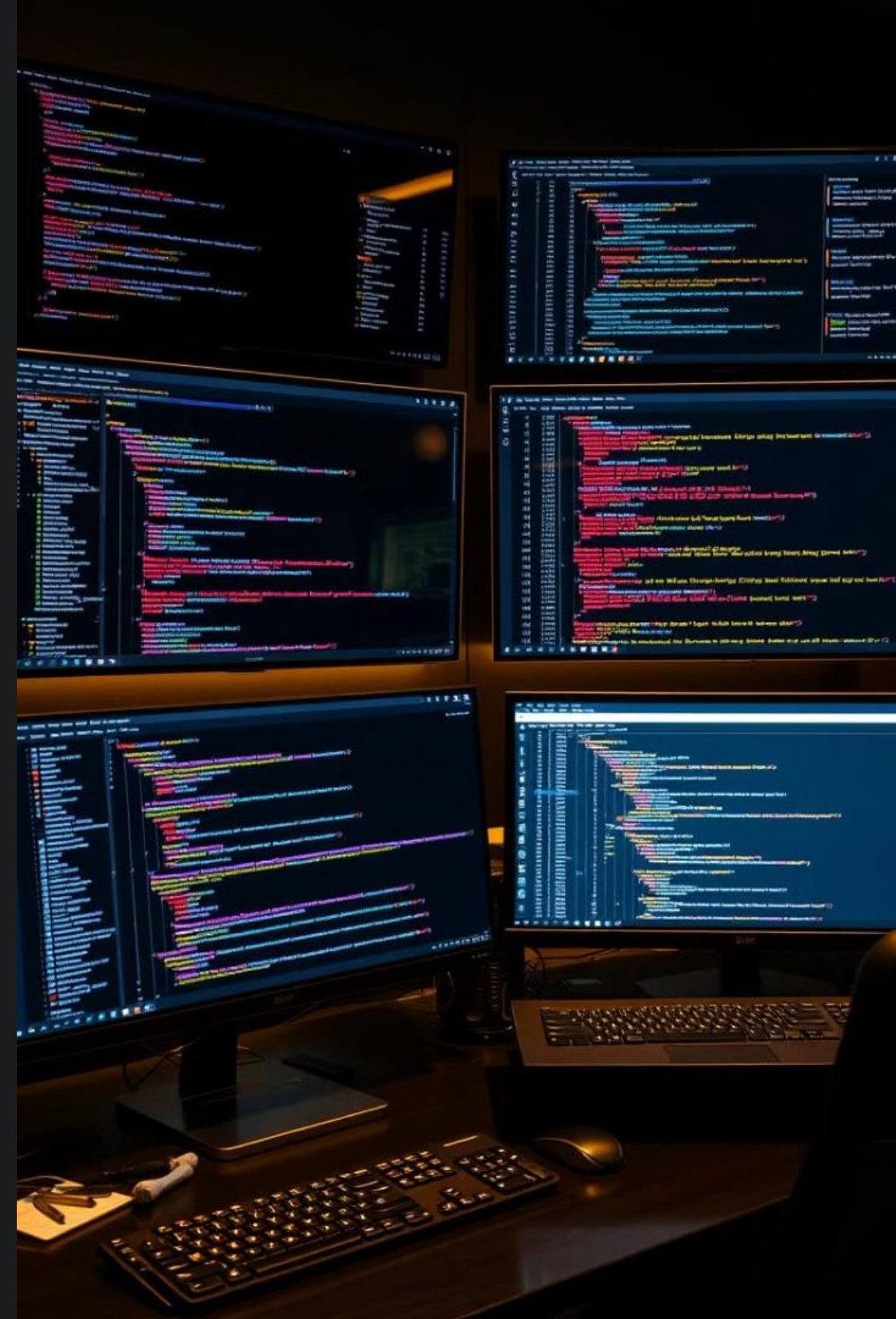
Technology Stack

- I used PostgreSQL for querying and analyzing the dataset .

- I used Gamma for generating the presentation of this project.

- I used Microsoft Power-Point for modifying the presentation of this project manually.

Business Problem

1

Content Performance Analysis

Identifying which genres, countries, and types (Movies vs. TV Shows) perform best.

2

User Engagement & Retention

Analyzing viewership trends to improve recommendations and reduce churn.

3

content Acquisition Strategy

Deciding what type of content to license or produce based on historical data.

4

Regional Market Expansion

Understanding content distribution across different countries for localization.

Made with Gamma

Table Creation

```sql
-- Netflix project
drop table if exists netflix;
Create table netflix(
show_id Varchar(10),
type Varchar(10),
title Varchar(150),
director Varchar(220),
casts   varchar(1000),
country varchar(150),
date_added varchar(50),
release_year int,
rating  varchar(10),
duration varchar(15),
listed_in   varchar(105),
description varchar(250)
);
```

1. Count the number of Movies vs TV Shows

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

```sql
40 ∨  select type ,rating
41    from
42    (
43    select type,
44    rating,
45    count(*),
46    Rank() over (partition by type order by count(*) DESC) as "ranking"
47    from netflix
48    group by type, rating
49    order by 1,3 desc
50    ) as t1
51    where ranking = 1;
52    |
```

Data Output   Messages   Notifications

	type character varying (10) 🔒	rating character varying (10) 🔒
1	Movie	TV-MA
2	TV Show	TV-MA

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

```
54 ∨  select *
55     from netflix
56     where type = 'Movie' and
57     release_year = '2020';
58
```

```
63     group by country
```

Data Output | Messages | Notifications

Showing rows: 1 to 517   Page No: 1   of 1

	show_id character varying (10)	type character varying (10)	title character varying (150)	director character varying (220)	casts character var
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 Sethupa
4	s85	Movie	Omo Ghetto: the Saga	JJC Skillz, Funke Akindele	Funke Akinde
5	s104	Movie	Shadow Parties	Yemi Amodu	Jide Kosoko,
6	s120	Movie	Here and There	JP Habac	Janine Gutier
7	s127	Movie	Shikara	Vidhu Vinod Chopra	Aadil Khan, S
8	s228	Movie	Really Love	Angel Kristi Williams	Kofi Siriboe,
9	s230	Movie	The Old Ways	Christopher Alender	Brigitte Kali (
10	s251	Movie	A Faraway Land	Veronica Velasco	Paolo Contis,

Total rows: 517   Query complete 00:00:00.283   CRLF   Ln 61, Col 36

Made with Gamma

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

```sql
60  select UNNEST(string_to_array(country, ',')) as new_country,
61      count(show_id) as total_content
62  from netflix
63    group by country
64    order by 2 desc
65  limit 5;
66
```

Data Output   Messages   Notifications

	new_country text	total_content bigint
1	United States	2818
2	India	972
3	United Kingdom	419
4	Japan	245
5	South Korea	199

Made with Gamma

5. Identify the longest movie

```sql
71 v  select title,type, duration
72       from netflix
73     where type ='Movie'
74       and
75     duration = (select max(duration) from netflix);
76
```

Data Output   Messages   Notifications

title character varying (150)	type character varying (10)	duration character varying (15)
InuYasha the Movie 2: The Castle Beyond the Looking Glass	Movie	99 min
InuYasha the Movie 3: Swords of an Honorable Ruler	Movie	99 min
Here and There	Movie	99 min
Good Luck Chuck	Movie	99 min
My Girl 2	Movie	99 min
Cousins	Movie	99 min
Rogue Warfare: Death of a Nation	Movie	99 min
Copenhagen	Movie	99 min
Black Holes	The Edge of All We Know	Movie	99 min
Ninja Assassin	Movie	99 min
Swordfish	Movie	99 min
678	Movie	99 min

Total rows: 118   Query complete 00:00:00.147

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

```
94 ∨  select *
95         from netflix
96     where director
97         ILike '%Rajiv Chilaka%'
98
```

Data Output   Messages   Notifications

Showing rows: 1 to 22   Page No:

	show_id character varying (10)	type character varying (10)	title character varying (150)	director character varying (220)
1	s407	Movie	Chhota Bheem - Neeli Pahaadi	Rajiv Chilaka
2	s408	Movie	Chhota Bheem & Ganesh	Rajiv Chilaka
3	s409	Movie	Chhota Bheem & Krishna: Mayanagari	Rajiv Chilaka
4	s410	Movie	Chhota Bheem & Krishna: Pataliputra- City of the Dead	Rajiv Chilaka
5	s411	Movie	Chhota Bheem And The Broken Amulet	Rajiv Chilaka
6	s412	Movie	Chhota Bheem And The Crown of Valhalla	Rajiv Chilaka
7	s413	Movie	Chhota Bheem and the Incan Adventure	Rajiv Chilaka
8	s414	Movie	Chhota Bheem and The ShiNobi Secret	Rajiv Chilaka
9	s415	Movie	Chhota Bheem Aur Hanuman	Rajiv Chilaka
10	s416	Movie	Chhota Bheem aur Krishna	Rajiv Chilaka
11	s417	Movie	Chhota Bheem aur Krishna vs Zimbara	Rajiv Chilaka
12	s420	Movie	Chhota Bheem: Bheem vs Aliens	Rajiv Chilaka
13	s421	Movie	Chhota Bheem: Dholakpur to Kathmandu	Rajiv Chilaka

Total rows: 22   Query complete 00:00:00.151

Made with Gamma

8 List all TV shows with more than 5 seasons

```sql
101  SELECT *
102    FROM netflix
103  WHERE type = 'TV Show'
104      AND
105  SPLIT_PART(duration, ' ', 1)::NUMERIC > 5;
106
107
```

	show_id character varying (10)	type character varying (10)	title character varying (150)	director character varying (220)	casts character varying (1000)
1	s9	TV Show	The Great British Baking Show	Andy Devonshire	Mel Giedroyc, Sue Perkins, Mary Berry, Paul Hollywood
2	s56	TV Show	Nailed It	[null]	Nicole Byer, Jacques Torres
3	s66	TV Show	Numberblocks	[null]	Beth Chalmers, David Holt, Marcel McCalla, Teresa Gallagher
4	s68	TV Show	Saved by the Bell	[null]	Mark-Paul Gosselaar, Tiffani Thiessen, Mario Lopez, Lark Voorhies, Elizabeth Berkley, D
5	s83	TV Show	Lucifer	[null]	Tom Ellis, Lauren German, Kevin Alejandro, D.B. Woodside, Lesley-Ann Brandt, Scarlett E
6	s275	TV Show	Grace and Frankie	[null]	Jane Fonda, Lily Tomlin, Martin Sheen, Sam Waterston, June Diane Raphael, Brooklyn De
7	s324	TV Show	30 Rock	[null]	Tina Fey, Alec Baldwin, Tracy Morgan, Jane Krakowski, Jack McBrayer, Scott Adsit, Juda
8	s339	TV Show	Hunter X Hunter (2011)	[null]	Megumi Han, Mariya Ise, Keiji Fujiwara, Miyuki Sawashiro, Daisuke Namikawa
9	s381	TV Show	The Flash	Glen Winter	Grant Gustin, Candice Patton, Danielle Panabaker, Carlos Valdes, Tom Cavanagh, Jesse
10	s391	TV Show	The Walking Dead	[null]	Andrew Lincoln, Steven Yeun, Norman Reedus, Chandler Riggs, Melissa McBride, Lauren
11	s526	TV Show	My Little Pony: Friendship Is Magic	[null]	Ashleigh Ball, Tabitha St. Germain, Tara Strong, Andrea Libman, Cathy Weseluck, Nicole
12	s527	TV Show	Orange Is the New Black	[null]	Taylor Schilling, Kate Mulgrew, Laura Prepon, Jason Biggs, Natasha Lyonne, Michael Ha
13	s538	TV Show	Terrace House: Opening New Do...	[null]	You, Reina Triendl, Yoshimi Tokui, Azusa Babazono, Ryota Yamasato, Shono Hayama
14	s549	TV Show	Grey's Anatomy	[null]	Ellen Pompeo, Sandra Oh, Katherine Heigl, Justin Chambers, Patrick Dempsey, Chandra

Total rows: 99   Query complete 00:00:00.211   CRLF   Ln 102, Col

9. Count the number of content items in each genre

```sql
SELECT
    UNNEST(STRING_TO_ARRAY(listed_in, ',')) AS genre,
    COUNT(*) AS total_content
FROM netflix
GROUP BY 1
ORDER BY 2 DESC;
```

	genre (text)	total_content (bigint)
1	International Movies	2624
2	Dramas	1600
3	Comedies	1210
4	Action & Adventure	859
5	Documentaries	829
6	Dramas	827
7	International TV Shows	774
8	Independent Movies	736
9	TV Dramas	696
10	Romantic Movies	613
11	Children & Family Movies	605
12	International TV Shows	577
13	Thrillers	512
14	Comedies	464

Total rows: 73   Query complete 00:00:00.143

10.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
    country,
    release_year,
    COUNT(show_id) AS total_release,
    ROUND(
        COUNT(show_id)::numeric /
        (SELECT COUNT(show_id) FROM netflix WHERE country = 'India')::numeric * 100, 2
    ) AS avg_release
FROM netflix
WHERE country = 'India'
GROUP BY country, release_year
ORDER BY avg_release DESC
LIMIT 5;
```

Data Output   Messages   Notifications

	country character varying (150)	release_year integer	total_release bigint	avg_release numeric
1	India	2017	101	10.39
2	India	2018	94	9.67
3	India	2019	87	8.95
4	India	2020	75	7.72
5	India	2016	73	7.51

Made with Gamma

11. List all movies that are documentaries

```sql
select *
from
(SELECT title,
    UNNEST(STRING_TO_ARRAY(listed_in, ',')) AS genre
FROM netflix)
  where  genre = 'Documentaries' ;


  select * from netflix
  where listed_in ILike '%Documentaries%';
```

Data Output   Messages   Notifications

Showing rc

	title character varying (150)	genre text
1	Dick Johnson Is Dead	Documentaries
2	Europe's Most Dangerous Man: Otto Skorzeny in Spain	Documentaries
3	My Heroes Were Cowboys	Documentaries
4	Schumacher	Documentaries
5	Blood Brothers: Malcolm X & Muhammad Ali	Documentaries
6	The Women and the Murderer	Documentaries
7	If I Leave Here Tomorrow: A Film About Lynyrd Skynyrd	Documentaries
8	Untold: Breaking Point	Documentaries
9	Final Account	Documentaries

Total rows: 829   Query complete 00:00:00.202

12. Find all content without a director

```sql
select *
    from netflix
where
    director Is null
```

Data Output | Messages | Notifications

Showing rows: 1 to 1000 | Page No: 1

	show_id character varying (10)	type character varying (10)	title character varying (150)	director character varying (220)	casts character varying (1
1	s2	TV Show	Blood & Water	[null]	Ama Qamata, Khosi
2	s4	TV Show	Jailbirds New Orleans	[null]	[null]
3	s5	TV Show	Kota Factory	[null]	Mayur More, Jitendr
4	s11	TV Show	Vendetta: Truth, Lies and The Mafia	[null]	[null]
5	s15	TV Show	Crime Stories: India Detectives	[null]	[null]
6	s16	TV Show	Dear White People	[null]	Logan Browning, Bra
7	s369	TV Show	Khawatir	[null]	[null]
8	s18	TV Show	Falsa identidad		

✓ Successfully run. Total query runtime: 230 msec. 26

Total rows: 2634 | Query complete 00:00:00.230

13. Find how many movies actor 'Salman Khan' appeared in last 10 years!

```sql
select * from netflix
where
        casts ILIKE '%Salman Khan%'
and release_year > Extract(year from current_date) - 10
```

Data Output Messages Notifications

show_id character varying (10)	type character varying (10)	title character varying (150)	director character varying (220)	
1	s2757	Movie	Paharganj	Rakesh Ranjan Kumar	

Total rows: 1   Query complete 00:00:00.207

14. Find the top 10 actors who have appeared in the highest number of movies produced in India.

```sql
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 ;
```

	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
8	Amitabh Bachch…	20
9	Kareena Kapoor	20
10	Asrani	17

Data Output Messages Notifications

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.

```sql
SELECT
    category,
    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 category;
```

Data Output | Messages | Notifications

	category text	content_count bigint
1	Good	8465
2	Bad	342

Insights

➢ I found that there are 8807 total content is on Netflix.

➢ There is two different types of content on Netflix such as 'Movies' and 'TV show' out of this there are 6132 are Movies and 2676 are TV shows.

➢ There are 517 Movies have released in '2020'.

➢ Longest movie duration on Netflix is 99 min.

➢ There are 99 TV shows having more than 5 seasons.

➢ Top 3 Genres on Netflix are International Movies , Dramas and Comedies.

➢ In 2017 India has released 101 Content which was the highest .

➢ There are 869 Movies of Documentary Genre on Netflix.

➢ 'Anupam Kher' is the one actor who have appeared in the most of movies released in INDIA.

Conclusion and Recommendations

Diversify Content

Invest in unique, culturally relevant Originals to differentiate from competitors.

Improve Personalization

Enhance recommendation algorithm to reduce churn and increase engagement.

Pricing Strategy

Explore tiered pricing models to retain price-sensitive subscribers while maintaining premium options.

Expand Markets

Focus on content localization and regional partnerships for growth in emerging markets.

Project made by – Gitanjali Pekamwar