



DATA ANALYSIS PROJECT USING SQL



Project Overview

This project entails an in-depth analysis of Netflix's movies and TV shows dataset using SQL. The aim is to extract valuable insights and address key business questions derived from the data. This README provides a comprehensive overview of the project's objectives, the business challenges tackled, the solutions implemented, significant findings, and concluding thoughts.

Objectives

1. Analyze the distribution of content types (movies vs. TV shows).
2. Identify the most prevalent ratings for both movies and TV shows.
3. Examine and categorize content by release years, countries, and durations.
4. Explore and classify content based on specific criteria and keywords.

❖ The data for this project is sourced from a Kaggle dataset:

[Kaggle Netflix Dataset](#)

Data Queries for Netflix Content Analysis: A SQL Approach

1. Count the total number of Movies and TV Shows.
2. Top 3 most common Ratings for Movies and TV Shows.
3. List of all movies released in the year of 2020 and 2021.
4. Top 5 countries with the most content on Netflix.
5. Identification of the top 10 longest-duration movies.
6. Content added in the last 5 years.
7. All the movies and TV shows by the renowned animation director 'Rajiv Chilaka'!
8. All the TV shows with more than 5 seasons.
9. The average number of content released in India on Netflix for each year.
10. All the movies that are listed as documentaries.
11. All the content without a director.
12. Top 10 Directors with the most titles, who have the most movies or shows on Netflix.
13. Content released trend: The number of movies released each year.
14. The number of titles featuring specific well-known actors, like 'Leonardo DiCaprio' or 'Scarlett Johansson'.
15. Content with the longest descriptions: display the title and description to understand content marketing trends.

❑ Database Creation:

```
create database netflix_db;  
use netflix_db;
```

❑ Schema:

```
• create table netflix  
(  
    show_id varchar(10),  
    type varchar(10),  
    title varchar(150),  
    director varchar(250),  
    casts varchar(1000),  
    country varchar(150),  
    date_added varchar(50),  
    release_year int,  
    rating varchar(15),  
    duration varchar(15),  
    listed_in varchar(100),  
    description varchar(270)  
);
```

❑ Data Import:

```
LOAD DATA INFILE 'D:/SQL projects YT/Netflix SQL Project (Zero Analyst)/netflix_titles.csv'  
INTO TABLE netflix_db.netflix  
FIELDS TERMINATED BY ','  
ENCLOSED BY ''''  
LINES TERMINATED BY '\n'  
IGNORE 1 ROWS;
```

1. Count the total number of Movies and TV Shows.

Query:

```
select
    type, count(*) as total_number
from netflix
group by type;
```

Output:

	type	total_number
▶	Movie	6131
	TV Show	2676

2. Top 3 most common Ratings for Movies and TV Shows.

Query:

```
select
    type, rating
from
    (select
        type, rating, count(*),
        rank() over(partition by type order by count(*) desc) as position
    from netflix
    group by type, rating) as by_rank
where position in (1, 2, 3);
```

Output:

	type	rating
▶	Movie	TV-MA
	Movie	TV-14
	Movie	R
	TV Show	TV-MA
	TV Show	TV-14
	TV Show	TV-PG

3. List of all movies released in the year of 2020 and 2021.

Query:

```
Select
    title, type, release_year
from netflix
where
    type = "movie" and release_year in (2020, 2021)
order by release_year desc;
```

Output:

	title	type	release_year
▶	My Little Pony: A New Generation	Movie	2021
	The Starling	Movie	2021
	Je Suis Karl	Movie	2021
	Confessions of an Invisible Girl	Movie	2021
	Intrusion	Movie	2021
	Go! Go! Cory Carson: Chrissy Takes the Wheel	Movie	2021
	Ankahi Kahaniya	Movie	2021
	The Father Who Moves Mountains	Movie	2021

794 row(s) returned

4. Top 5 countries with the most content on Netflix.

Query:

```
select
    country, count(show_id) as total_content
from netflix
where country <> ''
    group by country
    order by total_content desc
limit 5;
```

Output:

	country	total_content
►	United States	2818
	India	972
	United Kingdom	419
	Japan	245
	South Korea	199

5. Identification of the top 10 longest-duration movies.

Query:

```
select
  show_id,
  title as movie_name,
  duration as time_in_database,
  cast(substring_index(duration, " ", 1) as unsigned) as movie_time
from netflix
where type = "movie"
order by movie_time desc limit 10;
```

Output:

	show_id	movie_name	time_in_database	movie_time
▶	s4254	Black Mirror: Bandersnatch	312 min	312
	s718	Headspace: Unwind Your Mind	273 min	273
	s2492	The School of Mischieif	253 min	253
	s2488	No Longer kids	237 min	237
	s2485	Lock Your Girls In	233 min	233
	s2489	Raya and Sakina	230 min	230
	s167	Once Upon a Time in America	229 min	229
	s7933	Sangam	228 min	228
	s1020	Lagaan	224 min	224
	s4574	Jodhaa Akbar	214 min	214

6. Content added in the last 5 years.

Query:

```
select *,
       str_to_date(date_added, "%M %d, %Y") as Formatted_date
from
  netflix
where
  str_to_date(date_added, "%M %d, %Y") >= date_sub(curdate(), interval 5 year);
```

Output:

show_id	type	title	director	casts	country	date_added
s1	Movie	Dick Johnson Is Dead	Kirsten Johnson		United States	September 25, 2021
s2	TV Show	Blood & Water		Ama Qamata, Khosi Ngema, ...	South Africa	September 24, 2021
s3	TV Show	Ganglands	Julien Lederqc	Sami Bouajila, Tracy Gotoas, ...		September 24, 2021
s4	TV Show	Jailbirds New Orleans				September 24, 2021
s5	TV Show	Kota Factory		Mayur More, Jitendra Kumar, ...	India	September 24, 2021
s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, Zach Gilford, Ha...		September 24, 2021
s7	Movie	My Little Pony: A New Gene...	Robert Cullen, José Luis...	Vanessa Hudgens, Kimiko Gle...		September 24, 2021
s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba, Oyafunmike O...	United States, Ghana, Burk...	September 24, 2021
s9	TV Show	The Great British Baking Show	Andy Devonshire	Mel Giedroyc, Sue Perkins, M...	United Kingdom	September 24, 2021
s10	Movie	The Starling	Theodore Melfi	Melissa McCarthy, Chris O'Do...	United States	September 24, 2021
s11	TV Show	Vendetta: Truth, Lies and T...				September 24, 2021
s12	TV Show	Bangkok Breaking	Kongkiat Komesiri	Sukollawat Kanarot, Sushar M...		September 23, 2021
s13	Movie	Je Suis Karl	Christian Schwochow	Luna Wedler, Jannis Niewöhn...	Germany, Czech Republic	September 23, 2021
s14	Movie	Confessions of an Invisible Girl	Bruno Garotti	Klara Castanho, Lucca Picon		September 22, 2021

1000 row(s) returned

7. All the movies and TV shows by the renowned animation director 'Rajiv Chilaka'!



Query:

```
select
  show_id,
  director,
  title
from netflix
where director like "%Rajiv Chilaka%";
```

Output:

	show_id	director	title
▶	s407	Rajiv Chilaka	Chhota Bheem - Neeli Pahaadi
	s408	Rajiv Chilaka	Chhota Bheem & Ganesh
	s409	Rajiv Chilaka	Chhota Bheem & Krishna: M...
	s410	Rajiv Chilaka	Chhota Bheem & Krishna: P...
	s411	Rajiv Chilaka	Chhota Bheem And The Bro...
	s412	Rajiv Chilaka	Chhota Bheem And The Cro...
	s413	Rajiv Chilaka	Chhota Bheem and the Inca...
	s414	Rajiv Chilaka	Chhota Bheem and The ShiN...
	s415	Rajiv Chilaka	Chhota Bheem Aur Hanuman
	s416	Rajiv Chilaka	Chhota Bheem aur Krishna
	s417	Rajiv Chilaka	Chhota Bheem aur Krishna v...
	s420	Rajiv Chilaka	Chhota Bheem: Bheem vs Al...
	s421	Rajiv Chilaka	Chhota Bheem: Dholakpur t...
	s422	Rajiv Chilaka, Owl Mina	Chhota Bheem: Dus Pe Dus
	s423	Rajiv Chilaka	Chhota Bheem: Journey to

22 row(s) returned

8. All the TV shows with more than 5 seasons.

Query:

```
select
  show_id, type, title, duration as default_duration,
  cast(substring_index(duration, " ", 1) as unsigned) as formatted_duration
from netflix
where
  type = "tv show"
and
  cast(substring_index(duration, " ", 1) as unsigned) > 5
order by formatted_duration asc;
```

Output:

	show_id	type	title	default_duration	formatted_duration
	s6743	TV Show	Father Brown	6 Seasons	6
	s4748	TV Show	Kath and Kim	6 Seasons	6
	s4085	TV Show	Chef's Table	6 Seasons	6
	s5222	TV Show	West Coast Customs	6 Seasons	6
	s3109	TV Show	Private Practice	6 Seasons	6
	s5181	TV Show	Project Mc²	6 Seasons	6
	s7629	TV Show	Nurse Jackie	7 Seasons	7
	s275	TV Show	Grace and Frankie	7 Seasons	7
	s324	TV Show	30 Rock	7 Seasons	7
	s381	TV Show	The Flash	7 Seasons	7
	s6506	TV Show	Club Friday The Series 7	7 Seasons	7
	s7369	TV Show	Mad Men	7 Seasons	7
	s527	TV Show	Orange Is the New Black	7 Seasons	7
	s7710	TV Show	Parks and Recreation	7 Seasons	7
	s1044	TV Show	Gilmore Girls	7 Seasons	7

99 row(s) returned

9. The average number of content released in India on Netflix for each year.

Query:

```
select
    released_year,
    round(avg(total),2) as average_movie_count
from
    (select
        release_year as released_year,
        count(show_id) as total
    from netflix
    where country like "%India%"
    group by release_year)
as average_for_India
group by released_year
order by released_year desc;
```

Output:

	released_year	average_movie_count
▶	2021	35.00
	2020	77.00
	2019	93.00
	2018	101.00
	2017	111.00
	2016	80.00

57 row(s) returned

10. All the movies that are listed as documentaries.

Query:

```
select
  show_id,
  title,
  listed_in
from netflix
where listed_in like "%Documentaries%";
```

Output:

	show_id	title	listed_in
▶	s1	Dick Johnson Is Dead	Documentaries
	s17	Europe's Most Dangerous M...	Documentaries, International Movies
	s46	My Heroes Were Cowboys	Documentaries
	s69	Schumacher	Documentaries, International Movies, Sports M...
	s89	Blood Brothers: Malcolm X & ...	Documentaries, Sports Movies
	s92	The Women and the Murderer	Documentaries, International Movies

869 row(s) returned

11. All the content without a director.

Query:

```
select *  
  from netflix  
 where trim(director) = "" or director is null limit 3000;
```

Output:

	show_id	type	title	director	casts	country	date_added
▶	s2	TV Show	Blood & Water		Ama Qamata, Khosi Ngema, ...	South Africa	September 24, 2021
	s4	TV Show	Jailbirds New Orleans				September 24, 2021
	s5	TV Show	Kota Factory		Mayur More, Jitendra Kumar, ...	India	September 24, 2021
	s11	TV Show	Vendetta: Truth, Lies and T...				September 24, 2021
	s15	TV Show	Crime Stories: India Detectives				September 22, 2021
	s16	TV Show	Dear White People		Logan Browning, Brandon P. ...	United States	September 22, 2021
	s18	TV Show	Falsa identidad		Luis Ernesto Franco, Camila S...	Mexico	September 22, 2021
<

2634 row(s) returned

12. Top 10 Directors with the most titles, who have the most movies or shows on Netflix.



Query:

```
select
  director,
  count(title) as total_content
from netflix
where director <> ""
group by director
order by total_content desc
limit 10;
```

Output:

	director	total_content
►	Rajiv Chilaka	19
	Raúl Campos, Jan Suter	18
	Suhas Kadav	16
	Marcus Raboy	16
	Jay Karas	14
	Cathy Garcia-Molina	13
	Jay Chapman	12
	Martin Scorsese	12
	Youssef Chahine	12
	Steven Spielberg	11

13. Content released trend: The number of movies released each year.

Query:

```
select
    release_year,
    count(title) as total_movies
from netflix
    where release_year is not null
group by release_year
order by release_year desc;
```

Output:

	release_year	total_movies
▶	2021	592
	2020	953
	2019	1030
	2018	1147
	2017	1032
	2016	902
	2015	560
	2014	352
	2013	288
	2012	237
	2011	185
	2010	194
	2009	152
	2008	136

74 row(s) returned

14. The number of titles featuring specific well-known actors, like 'Leonardo DiCaprio' or 'Scarlett Johansson'.

Query:

```
select
  count(title) as total_movies
from netflix
where
  casts like "%Leonardo DiCaprio%"
  or
  casts like "%Scarlett Johansson%";
```

Output:

	total_movies
▶	15

15. Content with the longest descriptions: display the title and description to understand content marketing trends.

Query:

```
select
  title,
  description,
  length(description) as length_of_description
from netflix
where length(description) =
  (select max(length(description)) from netflix);
```

Output:

	title	description	length_of_description
►	Billu	An Indian village is abuzz with news that a major movie sta...	250

Conclusion

- ❖ **Content Distribution:** The dataset features a diverse collection of movies and TV shows, showcasing a variety of genres and ratings.
- ❖ **Popular Ratings:** Analyzing the most frequent ratings reveals insights into Netflix's target audience and the type of content that appeals to them.
- ❖ **Geographical Trends:** Highlighting top countries, such as India, and the average number of content releases provides a regional perspective on Netflix's distribution strategy.
- ❖ **Content Categorization:** Grouping titles by specific keywords offers a deeper understanding of the types of content available on the platform.

Overall, this analysis presents a comprehensive overview of Netflix's content landscape, providing valuable insights that can guide content strategy and decision-making.

*Thank
you*