# Netflix Data Exploration Project (Intermediate SQL)

## 1. Introduction

This project analyzes the Netflix dataset obtained from Kaggle using SQL. The goal is to extract business insights such as content trends, regional preferences, and top contributors. The analysis combines data cleaning, aggregation, joins, and window functions. Power BI is later used to visualize and present these findings effectively.

## 2. Data Cleaning & Preparation

* SELECT COUNT(\*) FROM netflix\_titles;

Explanation: Counts total records in the dataset to verify the data load.

Business Insight: Confirms total dataset size before further processing.

* SELECT COUNT(DISTINCT show\_id) AS unique\_ids FROM netflix\_titles;

Explanation: Ensures uniqueness of records by checking duplicate show IDs.

Business Insight: Each show\_id is unique; dataset integrity is confirmed.

* UPDATE netflix\_titles SET rating = 'Unknown' WHERE rating IS NULL;

Explanation: Replaces missing rating values with 'Unknown' to maintain consistency.

Business Insight: Dataset is now clean and ready for analysis.

## 3. Business Insights Queries

* SELECT type, COUNT(\*) AS total FROM netflix\_titles GROUP BY type;

Explanation: Counts total movies and TV shows to identify content distribution.

Business Insight: Netflix has a higher proportion of movies than TV shows.

* SELECT country, COUNT(\*) AS total FROM netflix\_titles GROUP BY country ORDER BY total DESC LIMIT 10;

Explanation: Finds top 10 content-producing countries.

Business Insight: The US and India are leading contributors to Netflix’s library.

* SELECT release\_year, COUNT(\*) AS total FROM netflix\_titles GROUP BY release\_year ORDER BY release\_year;

Explanation: Tracks content release trends over the years.

Business Insight: Content production has increased significantly after 2015.

## 4. Join Queries

* SELECT rc.age\_group, COUNT(nt.show\_id) AS total\_titles FROM netflix\_titles nt JOIN rating\_category rc ON nt.rating = rc.rating GROUP BY rc.age\_group ORDER BY total\_titles DESC;

Explanation: Joins Netflix titles with rating categories to analyze content by age group.

Business Insight: Majority of Netflix content is targeted at adult audiences.

* SELECT ci.region, COUNT(nt.show\_id) AS total\_titles FROM netflix\_titles nt JOIN countries\_info ci ON nt.country = ci.country GROUP BY ci.region ORDER BY total\_titles DESC;

Explanation: Links Netflix data with country-region mapping to see regional content trends.

Business Insight: North America and Asia have the highest volume of Netflix content.

## 5. Window Function Queries

* SELECT director, COUNT(\*) AS total\_titles, RANK() OVER (ORDER BY COUNT(\*) DESC) AS rank\_director FROM netflix\_titles WHERE director <> 'Unknown' GROUP BY director ORDER BY total\_titles DESC LIMIT 10;

Explanation: Uses RANK() to find top directors based on the number of titles produced.

Business Insight: Highlights the most productive directors contributing to Netflix content.

* SELECT release\_year, COUNT(\*) AS yearly\_titles, SUM(COUNT(\*)) OVER (ORDER BY release\_year) AS cumulative\_titles FROM netflix\_titles GROUP BY release\_year ORDER BY release\_year;

Explanation: Applies SUM() OVER() to compute cumulative content growth by year.

Business Insight: Reveals Netflix’s consistent catalog expansion over time.

## 6. Power BI Dashboard Plan

To extend this SQL analysis, Power BI can be used for interactive visualization. Key visuals include:  
- Bar chart: Movies vs. TV Shows distribution  
- Map: Top content-producing countries  
- Line chart: Yearly release trends  
- Donut chart: Age group content ratio  
The Power BI dashboard will connect directly to SQL queries or use exported CSVs, allowing users to dynamically explore insights.

## 7. Conclusion & Next Steps

This SQL project demonstrates intermediate-level skills in data cleaning, aggregation, joins, and window functions. It provides meaningful insights into Netflix’s content distribution, regional focus, and production trends. The next step involves integrating Power BI visuals for better storytelling and presentation in data analyst portfolios.