**Netflix Data Analysis using SQL**

**Objective**

The objective of this project is to perform Exploratory Data Analysis (EDA) on Netflix's content dataset using SQL. The analysis focuses on identifying content types, common ratings, country-wise distribution, actor appearances, and content trends over time using structured queries.

**Dataset Summary**

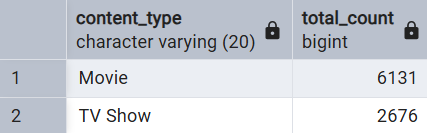
* **Source**: Netflix dataset from Kaggle
* **Records**: 8000 Rows
* **Columns** include:
  + title, type, director, cast, country, date\_added, release\_year, rating, duration, listed\_in, description

**Tools Used**

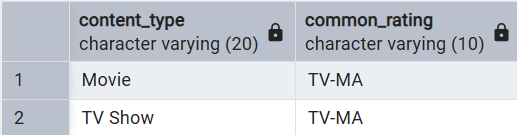
* SQL (PostgreSQL)
* Microsoft Word for documentation

**Exploratory Data Analysis (EDA)**

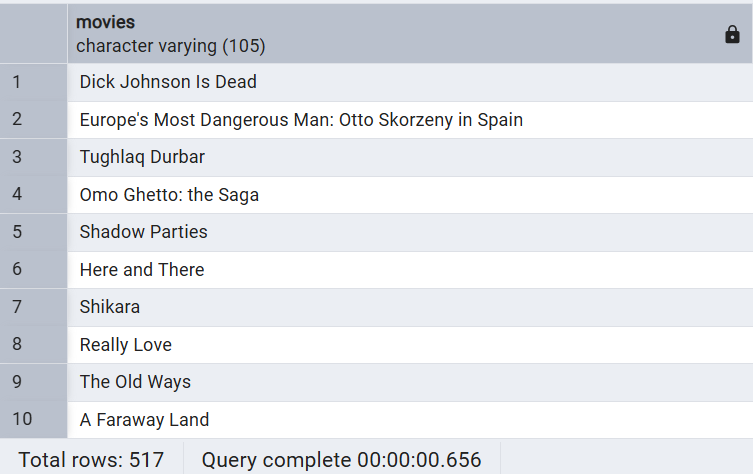
1. Counted total number of Movies and TV Shows



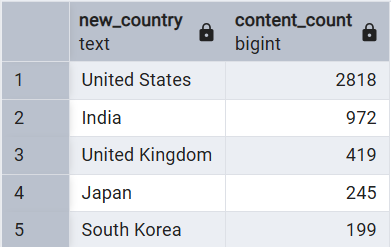
1. Found the most common rating for each content type



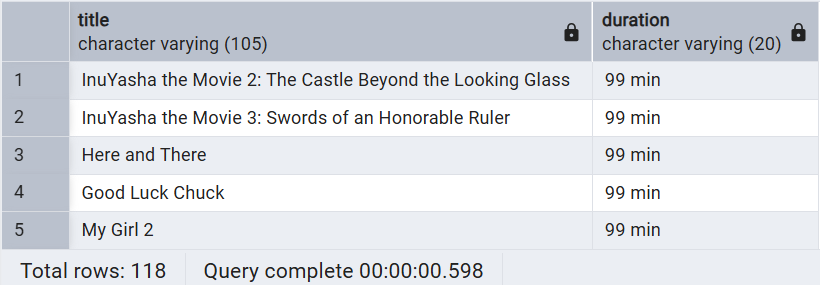
1. Listed all Movies released in 2020



1. Identified Top 5 Countries with the most Netflix content



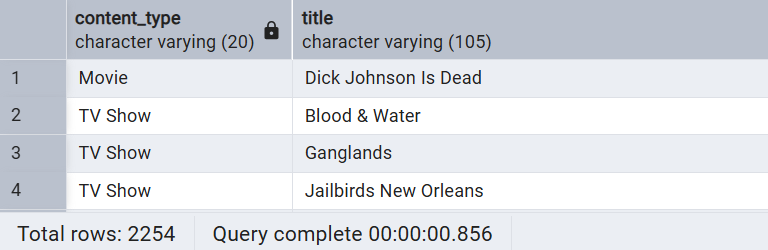
1. Found the longest Movies



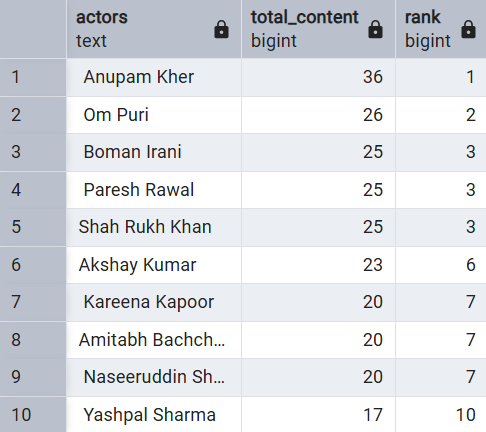
1. Listed TV Shows with more than 5 seasons



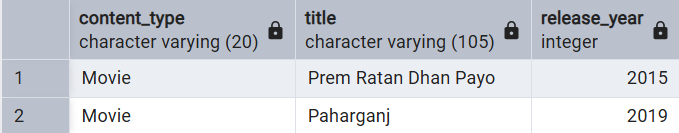
1. Identified content added in the last 5 years



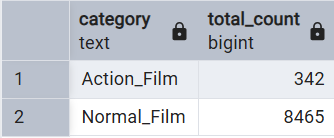
1. Listed Top 10 actors with the most appearances in Indian content



1. Found the number of Movies where actor *Salman Khan* appeared in the last 10 years



1. Categorized content as *Action\_Film* or *Normal\_Film* based on keywords in the description



**Conclusion**

This project demonstrates the power of SQL for data analysis on real-world datasets. It highlights patterns in Netflix's catalog across content types, genres, countries, and user-focused classifications.

**Author**

**Name:** Shubham Dalvi  
**Role:** Aspiring Data Analyst   
**GitHub:** https://github.com/Shubham-Dalvii  
**LinkedIn:** https://www.linkedin.com/in/shubham-dalvi-031939376/