Netflix Dataset Analysis Project

# Introduction

This project aims to analyze the Netflix dataset to uncover interesting insights and trends. The dataset contains information about various Netflix titles, including their type, title, director, cast, country, date added, release year, rating, duration, genres, and description.

# Import Process

To import the dataset into MySQL Workbench, the following steps were taken:  
  
1. Create a Database in MySQL Workbench:  
  
CREATE DATABASE netflix\_db;  
USE netflix\_db;  
  
  
2. Create a Table to Match the Dataset Structure:  
  
CREATE TABLE netflix\_titles (  
 show\_id VARCHAR(10) PRIMARY KEY,  
 type VARCHAR(10),  
 title VARCHAR(255),  
 director VARCHAR(255),  
 cast TEXT,  
 country VARCHAR(255),  
 date\_added VARCHAR(50),  
 release\_year INT,  
 rating VARCHAR(10),  
 duration VARCHAR(50),  
 listed\_in TEXT,  
 description TEXT  
);  
  
  
3. Import the Data:  
 - Use the MySQL Workbench import wizard to import the CSV file into the `netflix\_db` database.  
  
Difficulty Encountered: Issues with data types or large text fields, which were resolved by adjusting column types to `TEXT`.  
Interesting Finding: The dataset includes a variety of genres and countries, indicating a diverse content library.

# Cool Facts

Using SQL queries, two cool facts were discovered about the Netflix dataset:  
  
1. Number of Titles by Type: The dataset contains both movies and TV shows, with a significant number of entries in each category.  
  
SELECT type, COUNT(\*) AS count  
FROM netflix\_titles  
GROUP BY type;  
  
  
2. Average Release Year of Movies: The average release year of movies in the dataset provides insight into the era of the content available on Netflix.  
  
SELECT AVG(release\_year) AS avg\_release\_year  
FROM netflix\_titles  
WHERE type = 'Movie';

# Questions and Answers

Two questions were formulated about the data, and SQL queries were used to find the answers:  
  
1. Most Common Genres: This query identifies the most frequently listed genres in the dataset.  
  
SELECT listed\_in, COUNT(\*) AS count  
FROM netflix\_titles  
GROUP BY listed\_in  
ORDER BY count DESC  
LIMIT 10;  
  
  
2. Number of Titles Added Each Year: This query shows the trend of titles added to Netflix over the years.  
  
SELECT YEAR(STR\_TO\_DATE(date\_added, '%M %d, %Y')) AS year\_added, COUNT(\*) AS count  
FROM netflix\_titles  
GROUP BY year\_added  
ORDER BY year\_added;

# Charts

To visualize the findings, three charts were created:  
  
1. Bar Chart: Number of Titles by Type  
![Number of Titles by Type](number\_of\_titles\_by\_type.png)  
  
2. Pie Chart: Distribution of Titles by Rating  
![Distribution of Titles by Rating](distribution\_of\_titles\_by\_rating.png)  
  
3. Line Chart: Number of Titles Added Each Year  
![Number of Titles Added Each Year](number\_of\_titles\_added\_each\_year.png)

# Summary

This project provided insights into the diversity and trends in Netflix's content library. Using SQL queries and visualizations, interesting patterns and facts were uncovered, highlighting the richness and variety of content available on the platform.