Netflix Movies Analytics Project (SQL)

Project Overview
This project analyzes a detailed Netflix movie dataset containing metadata such as budget, revenue, popularity, ratings, cast, genres, and more. The dataset is imported into MySQL using Python (SQLAlchemy) for advanced analytical queries and business intelligence insights.

Dataset Description
- show_id – Unique ID for each movie
- type – Content type (Movie/Show)
- title – Title of the movie
- director – Director(s) of the movie
- cast – List of cast members
- country – Country of origin
- date_added – Date added to platform
- release_year – Year the movie was released
- rating – Rating score
- duration – Duration of movie
- genres – List of genres
- language – Original language
- description – Summary of the movie
- popularity – Popularity index
- vote_count – Number of votes
- vote_average – Average user rating
- budget – Production budget
- revenue – Box-office revenue
- date_N – Normalized date format
- Rev – Processed revenue column
- Year – Extracted release year

Technologies Used
Python Libraries:
- - pandas
- - SQLAlchemy
- - PyMySQL
- - urllib.parse (quote_plus)

Database:

- - MySQL 8.x
- - MySQL Workbench

Database Setup:

1. Create Database:
CREATE DATABASE netflix;

2. Upload Data Using Python:

```
from sqlalchemy import create_engine
from urllib.parse import quote_plus
import pandas as pd

username = "root"
password = "password"
host = "localhost"
database = "netflix"

password_encoded = quote_plus(password)
engine =
create_engine(f"mysql+pymysql://{username}:{password_encoded}@{host}/{database}")

df = pd.read_csv("netflix_movies.csv")
df.to_sql("movie", con=engine, if_exists="replace", index=False)
```

Advanced SQL Queries

Top 10 ROI Movies

SELECT title, budget, revenue, ROUND((revenue - budget) / budget, 4) AS ROI FROM movie WHERE budget > 0 AND revenue IS NOT NULL ORDER BY ROI DESC LIMIT 10;

title	budget	revenue	ROI
The Beatles: Eight Days a Week - The Touring Y...	5	12283800	2456759.0000
No Dogs or Italians Allowed	5	1395922	279183.4000
Scrapper	84	1234483	14695.2262
The Good Neighbor	105	94909	902.8952
Animal World	125000	69885700	558.0856
Mr. Six	250000	139191345	555.7654
One Cut of the Dead	52406	27590180	525.4699
Secret Superstar	286284	137416709	479.0014
The Gallows	100000	42964410	428.6441
Freaks	2000	368410	183.2050

High-Performing Directors

SELECT director, COUNT(*) AS total_movies, ROUND(AVG(vote_average), 2) AS avg_rating FROM movie WHERE director IS NOT NULL AND director <> '' GROUP BY director HAVING AVG(vote_average) > 7.5 ORDER BY avg_rating DESC LIMIT 10;

director	total_movies	avg_rating
J. Gomes	1	10
Ji Hyun-sook	1	10
Sadao Sadaoka	1	10
Anouk de Clercq, Tom Callemin	1	10
Park Eun-soo-I	1	10
Choi Eun-jung	1	10
Yiannis Stravolaimos	1	10
Yoon Kyung-sik	1	10
Somaratne Dissanayake	1	10
Nuel C. Naval	1	10

Genre With Highest Average Revenue (in Million)

SELECT genres, AVG(revenue) AS avg_revenue FROM movie GROUP BY genres ORDER BY avg_revenue DESC;

genres	avg_revenue
Adventure, Drama, Family, Animation	166
Adventure, Science Fiction, Action	151
Action, Adventure, Science Fiction, Thriller	149
Family, Fantasy, Romance	127
Adventure, Fantasy, Romance, Family	105
Action, Animation, Comedy, Family, Adventure	103
Adventure, Comedy, Animation, Family	88
Action, Adventure, Science Fiction, Drama	88
Animation, Family, Adventure, Drama, Comedy	86
Adventure, Action, Science Fiction	85
Animation, Family, Adventure, Fantasy, Comedy	79
Fantasy, Adventure, Action, Family, Romance	76
Adventure, Fantasy, Drama, Romance	76
Comedy, Adventure, Fantasy, Animation, Family	75

High Popularity but Low Ratings

SELECT title, popularity, vote_average FROM movie WHERE popularity > (SELECT AVG(popularity) FROM movie) AND vote_average < (SELECT AVG(vote_average) FROM movie) ORDER BY popularity DESC;

Revenue Trend Over Years

SELECT Year, COUNT(*) AS movies_released FROM movie GROUP BY Year ORDER BY Year;

-Which factors (budget, popularity, vote_count) most strongly predict high revenue?

```
SELECT
   CASE
      WHEN revenue < 100000000 THEN 'Low Revenue'
      WHEN revenue BETWEEN 100000000 AND 500000000 THEN 'Medium Revenue'
      ELSE 'High Revenue'
   END AS revenue_category,
   ROUND(AVG(budget), 0) AS avg_budget,
   ROUND(AVG(popularity), 2) AS avg_popularity,
   ROUND(AVG(vote_count), 0) AS avg_vote_count,
   ROUND(AVG(vote_average), 2) AS avg_rating
FROM movie
GROUP BY revenue_category
ORDER BY avg_budget DESC;
```

revenue_category	avg_budget	avg_popularity	avg_vote_count	avg_rating
High Revenue	164925393	123.25	12134	7.08
Medium Revenue	70540430	57.04	4849	6.71
Low Revenue	3989931	17.43	388	5.91

Key Insights

- ROI shows how profitable each movie is relative to its budget.
- Directors with consistently high ratings indicate content quality trends.
- Genres vary significantly in revenue potential.
- Popularity does not always correlate with good ratings.
- Release year trends show production and performance patterns.