Netflix Movies & TV Shows
End-to-End Data Analytics Project Report

Name: Abha Garg

Course: Strategic business Management

Roll.No: C1-003

Course: How do big data tools uncover trends?

Assignment: Data Analytics

Tools Used: Python | SQL | Power BI

Dataset: Netflix Movies & TV Shows

Submission Type: Individual Project

1. Introduction

This project presents an end-to-end data analytics study on the Netflix Movies and TV Shows dataset. The objective of the analysis is to transform raw content-level data into meaningful insights using Python for data cleaning and exploratory analysis, SQL for structured querying, and Power BI for interactive visualization. The project demonstrates how data-driven insights can support strategic decision-making in the digital entertainment industry.

2. Objectives of the Analysis

The key objectives of this analysis are:

- To understand the overall composition of Netflix's content library.
- To analyse trends in content addition over time.
- To identify dominant genres, ratings, and countries contributing to Netflix content.
- To build an interactive dashboard for exploratory and decision-driven analysis.

3. Dataset Description

The dataset contains information about movies and TV shows available on Netflix. Each record represents a unique title and includes attributes such as content type, title, director, cast, country, date added to Netflix, release year, rating, duration, genre, and description.

After data cleaning, the final dataset consisted of approximately 8,800 records and 15 analytical columns, making it suitable for large-scale content analysis.

4. Tools and Technologies Used

The following tools and technologies were used in this project:

- Python (Pandas, NumPy): Data cleaning and exploratory analysis
- SQLite (SQL): Structured querying and business analysis
- Power BI: Interactive dashboard and visualization
- GitHub: Project version control and structured submission

5. Data Cleaning and Preparation

Data cleaning was a critical phase of the analysis. The following steps were performed using Python:

- Missing values in columns such as director, cast, and country were handled using logical placeholders.
- The date_added column was converted into a datetime format, and new features such as year_added were derived.
- The duration column was transformed to extract numeric values for meaningful comparison.

• Invalid and inconsistent records were identified and removed.
• A cleaned dataset was saved and reused consistently across SQL and Power BI.

6. Exploratory Data Analysis and Key Findings

The exploratory analysis revealed several important insights:

• Netflix hosts approximately 8,800 titles in total.
• Movies dominate the platform with around 6,100 titles, while TV shows account for approximately 2,600 titles.
• Content addition increased sharply after 2016, with peak additions observed between 2018 and 2020.
• Drama-related content is the most prevalent genre, followed by Documentaries and Stand-Up Comedy.
• The United States is the leading content producer with around 2,800 titles, followed by India (~970 titles) and the United Kingdom (~420 titles).

7. SQL-Based Business Analysis

SQL was used to perform structured queries on the cleaned dataset. Key business questions addressed through SQL included:
• Distribution of movies versus TV shows
• Top 10 genres by content volume
• Top contributing countries
• Number of movies released after 2015

SQL analysis validated the findings obtained from Python and ensured accuracy and consistency of the insights presented.

8. Power BI Dashboard Overview

An interactive Power BI dashboard was developed to visualize insights derived from the analysis. The dashboard consists of three pages:

• Executive Overview: High-level KPIs such as total titles, movies, TV shows, rating distribution, and content growth trends.
• Content Deep Dive: Genre-wise, country-wise, and format-based analysis.
• Interactive Exploration: User-controlled filters enabling dynamic exploration by type, year, rating, and country.

The dashboard enables decision-makers to explore Netflix's content strategy visually and interactively.

9. Business Insights and Recommendations

Based on the analysis, the following insights and recommendations are proposed:

• Netflix should continue investing in drama and documentary genres, as they consistently attract high content volume and user engagement.
• Regional content strategies, especially in markets such as India and the United Kingdom, can drive subscriber growth and localization.
• Maintaining a balanced mix of movies and TV shows is essential to cater to diverse audience preferences.
• Monitoring content addition trends can help Netflix optimize acquisition and production strategies.

10. Conclusion

This project successfully demonstrates an end-to-end data analytics workflow using Python, SQL, and Power BI. By transforming raw Netflix data into actionable insights, the analysis highlights the importance of data-driven decision-making in shaping content strategies and sustaining competitive advantage in the streaming industry.