Netflix Data Analysis and Visualization

Overview

This project involves analysing and visualizing Netflix data to uncover trends, patterns, and insights. Netflix is a popular streaming service offering a vast catalogue of movies, TV shows, and original content. The dataset used here spans content added to Netflix from 2008 to 2021, with the oldest content dating back to 1925. The primary goal of this project is to demonstrate data cleaning and visualization skills using tools like PostgreSQL, Python, and Power BI.

Dataset

The dataset used in this project is a cleaned version of the original dataset, which can be accessed via the following link: Netflix Dataset on Kaggle

Key Features of the Dataset:

- **Time Period**: 2008 to 2021

- **Oldest Content**: 1925

- **Attributes**: Show ID, type, title, director, country, date added, release year, rating, duration, and genres

Tools and Technologies

- **Data Cleaning**: PostgreSQL

- **Visualization**: Power BI

- **Programming**: Python (Pandas, Matplotlib, Seaborn)

Data Cleaning Process

The data cleaning steps included:

1. Treating null values.

2. Removing duplicates.

3. Populating missing rows.

4. Dropping unnecessary columns.

5. Splitting columns where necessary.

Key Insights

Exploratory Data Analysis (EDA)

Content Type Distribution

- Distribution of Movies vs. TV Shows available on Netflix.

- Movies account for a larger proportion of the content.

Genre Analysis

- Identified the most popular genres on Netflix.

- Analysed the frequency of genres in movies and TV shows.

Temporal Trends

- Content added over time was analysed by year and month.

- Noted peaks in content addition trends.

Top Contributors

- Highlighted the top directors with the most content on Netflix.

Ratings Distribution

- Visualized the frequency of different content ratings.

Power BI Dashboard

The Power BI dashboard provides interactive insights into Netflix content, such as:

- Overview of content types and their distribution.

- Trends in content addition over the years.

- Popular genres and directors.

Visualizations

Sample visualizations include:

1. Bar charts showcasing the distribution of genres.

2. Line graphs illustrating content added over time.

3. Word clouds of movie and TV show titles.

How to Use

1. Download the Power BI file NetflixReport.pbix from this repository.

2. Open the file in Power BI Desktop.

3. Explore the interactive dashboards to gain insights.

Conclusion

This project showcases the effective use of data cleaning and visualization tools to extract meaningful insights from a dataset. The Power BI dashboard is designed to make Netflix data more accessible and informative for stakeholders.

Dataset Credits

- **Original Dataset**: [Netflix Shows on Kaggle](#)
- **Tableau Dashboard**: [Netflix Tableau Dashboard](#)