PacalineN /
Netflix-Data-Analysis

`<>` **Code**   Issues   Pull requests   Actions   Projects   Security   Insigh

Netflix-Data-Analysis / **README.md**

PacalineN **Update README.md**   5 months ago

68 lines (40 loc) · 4.08 KB

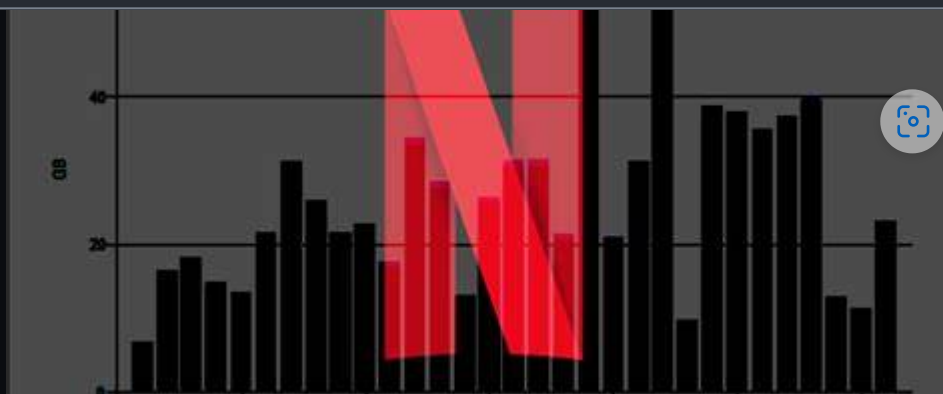**Netflix-Data-Analysis** / **README.md**   ↑ Top

Preview | Code | Blame   Raw

Netflix Data Analysis with Python : Unveiling Insights and Trends

Table of Content

- Project Overview

- Data sources

- Tools Used

- Data Cleaning/Preparation

- Exporatory Data Analysis

- Results/Findings

- Recommendation

- [Reference](#)

Project Overview

Welcome to the Netflix Data Analysis project! In this repository, I explore the Netflix dataset using Python for data analysis. My goal is to uncover interesting insights and trends within the vast collection of movies and TV shows available on Netflix.

Data sources

This Netflix dataset has information about the tv shows and Movies available on Netflix till 2021
The dataset is collected from Flixable which is a third-party Netflix search engine,and available on Kaggle website for free

Tools Used

- Python
- Pandas
- Matplotlib and Seaborn for data visualization

Data Cleaning

In the initial data preparation phase,I performed the following tasks :

1. Data loading and inspection
2. Handling missing values
3. Data cleaning and formatting

Exporatory Data Analysis

Exploratory Data Analysis (EDA) is crucial for understanding the Netflix dataset, uncovering patterns, trends, and insights. In this section, I explore various aspects of the dataset to gain a comprehensive understanding of the content available on Netflix, addressing key questions:

1. **Release Trends Over the Years:**

   - In which year did the highest number of TV shows and movies release? (Visualized using a Bar Graph)

2. **Movies and TV Shows by Top 10 Directors:**

- Show all records where the category is "Movie" and type is "Comedians" or the country is "United Kingdom." Additionally, display the top 10 directors who gave the highest number of TV shows and movies to Netflix. (Visualized using a Bar Graph)

3. **Different Ratings Defined by Netflix:**

   - What are the different ratings defined by Netflix?

Results

The Netflix dataset analysis reveals significant trends. The year 2019 stands out with the most releases, marking a pivotal moment in Netflix content production. Raúl Campos, Jan Suter, and Marcus Raboy lead in directorial contributions, emphasizing their impact on Netflix's diverse content landscape. The dataset's ratings, encompassing 'TV-MA,' 'R,' 'PG-13,' etc., showcase the platform's commitment to varied content suitability. The United States dominates TV show contributions, emphasizing its global influence. A concise bar graph illustrates the distribution of movies and TV shows, highlighting insights into content diversity.
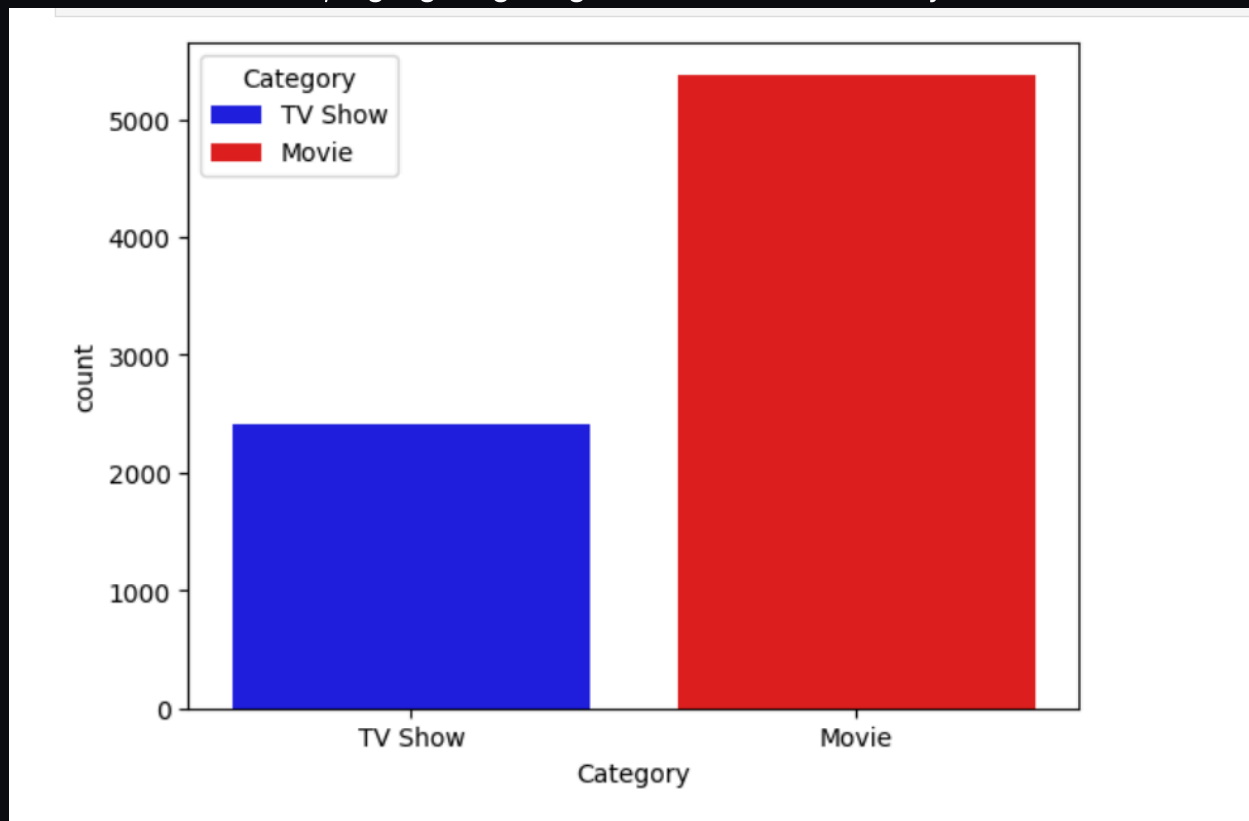
Recommendation

For those intrigued by the dynamics of Netflix content, this project offers a compelling exploration of key facets. The detailed analyses and visualizations provide a nuanced understanding of release patterns and directorial contributions. The inclusion of content ratings and a focus on top directors enhance the depth of insights. The project's use of a bar graph to represent distribution adds clarity. Overall, this analysis serves as a valuable reference for individuals interested in content consumption trends on streaming platforms, providing a comprehensive overview of Netflix's evolving content landscape.

Reference

The dataset used in this project is sourced from Flixable, a third-party Netflix search engine. Flixable provides comprehensive information about TV shows and movies available on Netflix up until 2021. The dataset is freely accessible on Kaggle, a platform for data science and machine learning enthusiasts. The data from Flixable serves as a valuable resource for exploring trends, patterns, and insights within the Netflix content landscape. The original dataset can be found on Kaggle, and credit goes to Flixable for compiling and making the data publicly available.