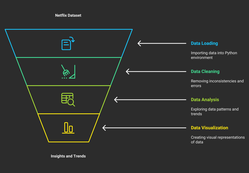
Exploratory Data Analysis (EDA) on Netflix Data

This document presents an exploratory data analysis (EDA) of a Netflix dataset, focusing on the loading, cleaning, analyzing, and visualizing of the data using Python libraries such as Pandas, Matplotlib, and Seaborn. The goal of this analysis is to derive insights from the dataset and prepare it for potential machine learning tasks. The findings from the EDA highlight key trends and patterns in the content available on Netflix, including the distribution of movies and TV shows, rating types, geographical content distribution, release trends, and popular genres and directors.

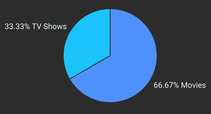
Netflix Data Analysis Process



Findings from EDA

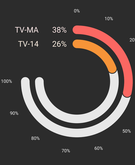
1. Content Distribution: By plotting the data, we observe that the number of movies in our dataset is almost double that of TV shows. This indicates a significant focus on movie content within Netflix's offerings.

Netflix Content Distribution



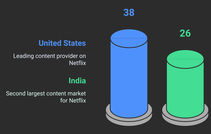
2. Rating Types: The analysis of rating types reveals that the majority of the content is rated TV-MA, accounting for 38% of the dataset. Following this, TV-14 ratings make up 26%, indicating that these two categories dominate the ratings landscape.

Distribution of Netflix Content Ratings

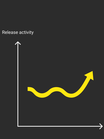


3. Geographical Content Distribution: The United States leads in the amount of content available on Netflix, followed closely by India. This suggests a strong focus on these markets for content production and acquisition.

Geographical Content Distribution on Netflix



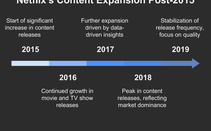
4. Release Trends: Both movies and TV shows experience peak release activity in July, indicating a seasonal trend in content availability. This could be linked to viewer engagement patterns during the summer months.



Peak release activity in July

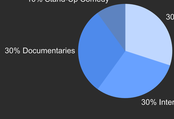
5. Post-2015 Release Frequency: There has been a noticeable increase in the frequency of releases for both movies and TV shows after 2015. This rise can be attributed to the growing binge-watching culture, advancements in content licensing, and Netflix's data-driven approach to production and distribution.

Netflix's Content Expansion Post-2015



6. Top Movie Genres: The analysis identifies the top three movie genres on Netflix as Dramas, International Movies, and Documentaries, with Stand-Up Comedy also being a popular choice among viewers.

Distribution of Top Movie Genres on Netflix



7. Top TV Show Genres: For TV shows, the leading genres include Kids TV, International TV Shows, and TV Dramas, with Crime TV Shows also featuring prominently in viewer preferences.

Top TV Show Genres on Netflix

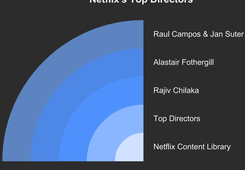
TV Dramas
TV dramas are the third most popular genre on Netflix.

Kids TV
Kids TV shows are the most popular genre on Netflix.

International TV Shows
International TV shows are the second most popular genre.

8. Top Directors: The analysis highlights the top directors who have released the highest number of movies and TV shows on Netflix. These include Rajiv Chilaka, Alastair Fothergill, Raul Campos, and Jan Suter, showcasing their significant contributions to Netflix's content library.

Netflix's Top Directors



This exploratory data analysis provides valuable insights into the Netflix dataset, revealing trends and patterns that can inform future content strategies and machine learning applications.