

■ Netflix Titles Dataset - Exploratory Data Analysis (EDA)

Author: Samad Mehboob

Email: samadmehboob940@gmail.com

1. Objective

The purpose of this project is to perform an in-depth Exploratory Data Analysis (EDA) on the Netflix Titles Dataset, which contains information about all movies and TV shows available on Netflix. The goal is to extract meaningful insights about Netflix's content — such as its distribution by type, year, country, genre, and ratings.

2. Dataset Overview

- Dataset Name: netflix_titles.csv
- Rows: ~8,800 titles
- Columns: 12 (type, title, director, cast, country, release_year, rating, duration, etc.)

Each record represents a Netflix title (movie or TV show) with metadata such as the director, cast, country, genre, release year, and duration.

3. Methodology

The project was implemented using Python (Pandas, Matplotlib, Numpy) in a Jupyter Notebook environment. The workflow includes the following major steps:

Step 1: Data Loading

Dataset imported using `pandas.read_csv()`. Missing values observed in director, cast, country, and rating.

Step 2: Data Cleaning

Removed duplicates, filled missing values with 'Unknown', and standardized formats.

Step 3: Feature Engineering

Created new column 'duration_num' for movie duration or seasons for TV shows.

Step 4: Visualization & Analysis

Generated multiple charts to explore trends and insights.

4. Key Insights

- Netflix focuses more on movies than TV shows (~70% movies).
- Content production rose sharply after 2015, peaking around 2019.
- USA, India, and UK dominate Netflix's content library.
- Most common genres: Drama, Comedy, and Documentary.

- Most frequent rating: TV-MA (Mature Audience).
- Average movie duration: ~90 minutes.

5. Tools Used

Python, Pandas, NumPy, Matplotlib, Jupyter Notebook

6. Conclusion

This analysis reveals how Netflix has evolved into a global entertainment platform with diverse content across multiple genres and countries. By visualizing and understanding this dataset, we gain valuable insights into Netflix's content strategy, focusing on variety, international reach, and modern storytelling trends.