

Netflix data analysis: trends in content type, duration, release year, and geographic distribution

1. Introduction about the Data:	2
2. Processing of the Data:	3
1. Loading the Dataset	3
2. Handling Missing Values	3
3. Converting Data Types	3
4. Standardizing Data	3
5. Data Segmentation	3
3. Output Graphs and Their Explanation:	4
3.1 Content Added Over the Years:	4
3.2 Distribution of Content Types:	4
3.3 Top 10 Content-Producing Countries:	5
3.4 Movie Duration Distribution:	7
3.5 Most Common Genres:	6

Source: [Netflix Titles Dataset - Kaggle](#)

Prepared By: Hani Bint E Rauf – master's in data science

Tools Used: Python 3.x, Pandas, Matplotlib, Seaborn, Jupyter Notebook

1. Introduction about the Data:

The *Netflix Titles Dataset* from Kaggle contains **8,807** records of movies and TV shows available on Netflix. Key details include title, type, director, cast, country, date added, release year, duration, rating, and genres.

This dataset enables analysis of Netflix's content library by examining type distribution, growth trends, geographic spread, durations, and audience ratings.

2. Processing of the Data:

Before analysis, the dataset required several cleaning and preparation steps to ensure accurate results:

1. Loading the Dataset

The dataset was imported into a Python environment using Pandas for further analysis.

2. Handling Missing Values

- Certain fields like *Director*, *Cast*, *Country*, and *Rating* had missing entries.
- Missing values were either filled with a placeholder ("Unknown") or excluded if they were not essential for a specific analysis.

3. Converting Data Types

- The *Date Added* column was converted from text to **datetime** format to allow sorting and trend analysis.
- Duration values for movies were extracted as integers (minutes) for analysis.

4. Standardizing Data

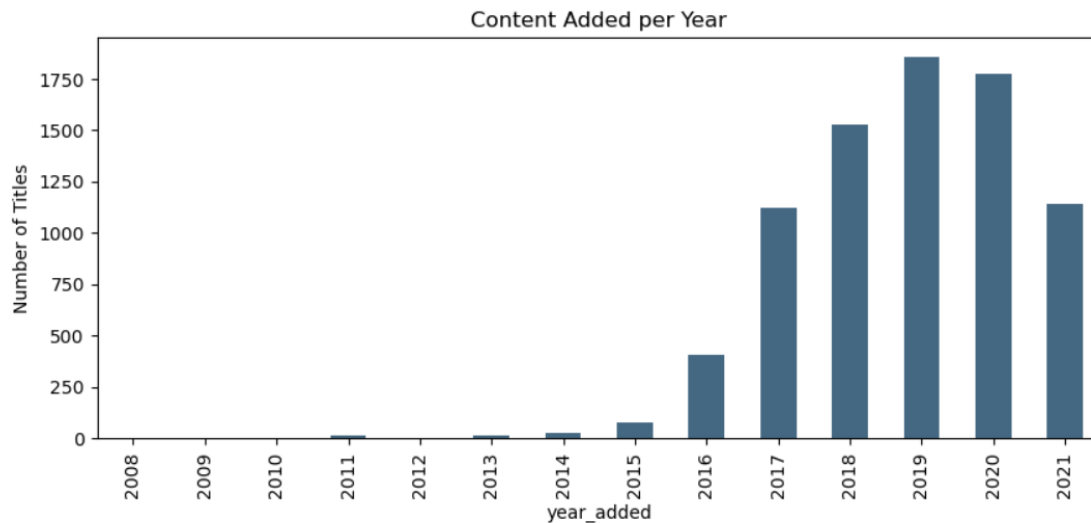
- Country names were kept consistent.
- Extra spaces and formatting inconsistencies were removed from categorical fields.

5. Data Segmentation

- The dataset was split by content type (*Movies* vs *TV Shows*) for comparative analysis.

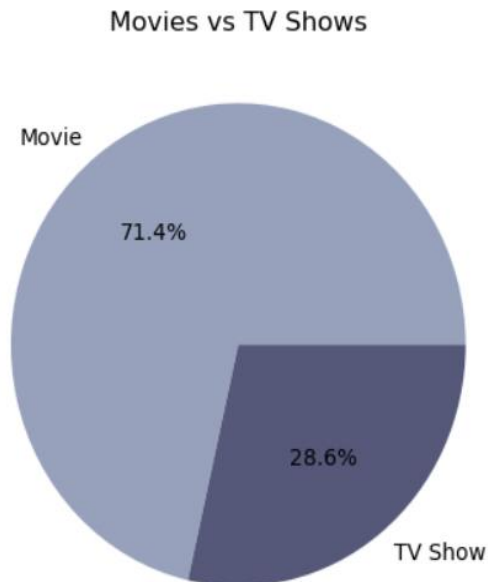
3. Output Graphs and Their Explanation:

3.1 Content Added Over the Years:



The number of titles added to Netflix increased steadily from the early 2010s, peaking around 2019. A slight drop can be seen afterward, possibly due to shifts in licensing strategies or the impact of global production delays during the COVID-19 pandemic.

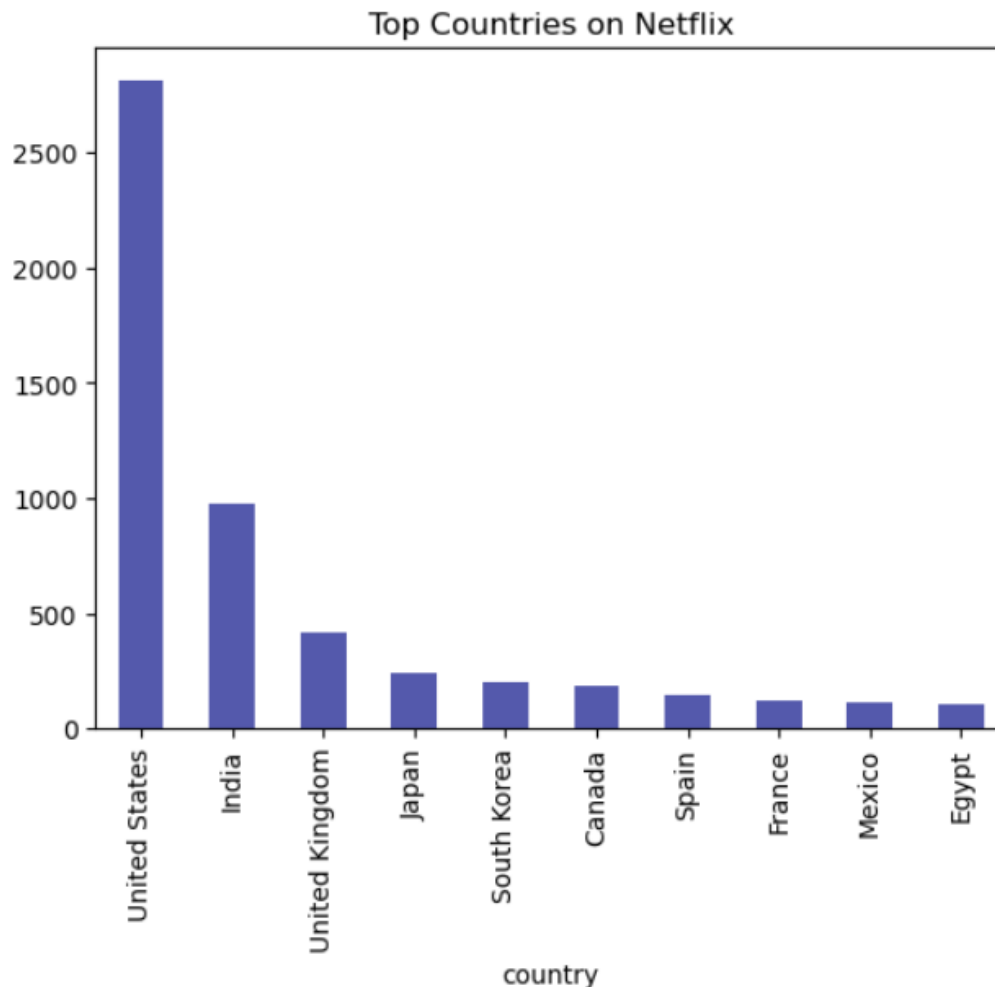
3.2 Distribution of Content Types:



The majority of Netflix titles are movies, with a smaller but significant proportion being TV shows. This indicates that Netflix's content library is primarily film-focused, although episodic content has been growing in recent years.

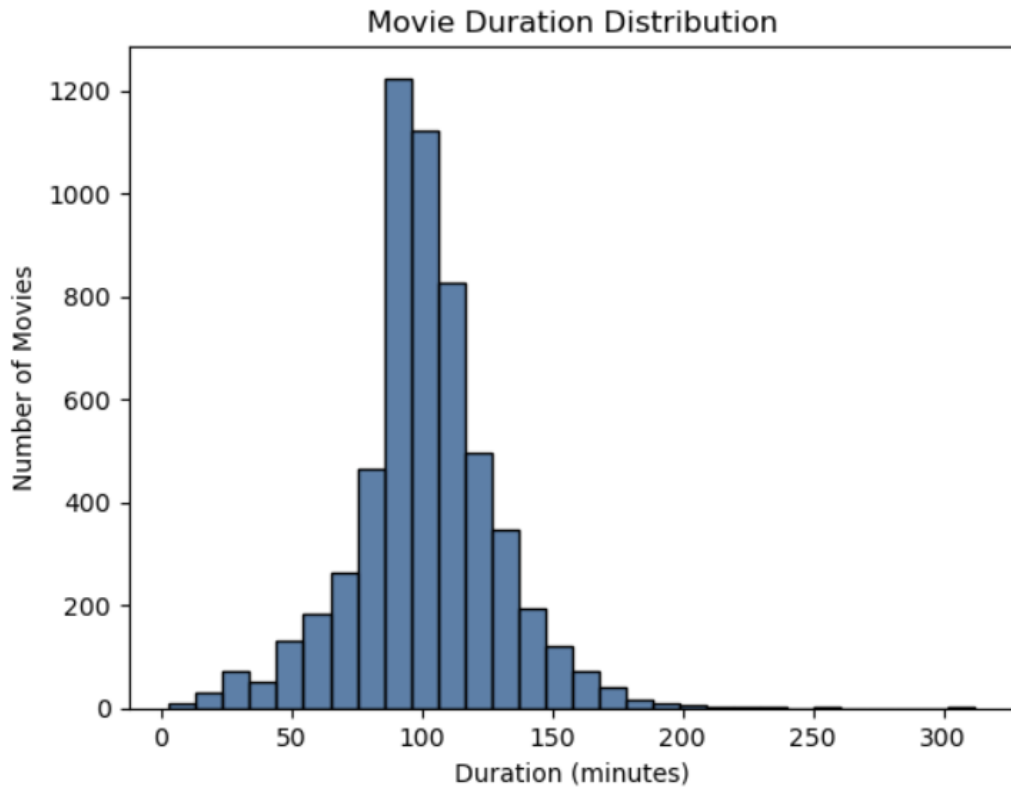
3.3 Top 10 Content-Producing Countries:

The **United States** dominates Netflix's content production, followed by **India**, **United Kingdom**, and other countries. This highlights Netflix's focus on major entertainment markets while also sourcing international content to appeal to a global audience.



3.4 Movie Duration Distribution:

Most movies in the Netflix catalog have a duration between **80 and 120 minutes**, aligning with standard feature-length films. Shorter and longer movies exist but are less common.



3.5 Most Common Genres:

The most frequent genre is “International Movies”, with 2500+ titles. Other popular genres include Dramas, Comedies, and International TV Shows.

