

# Data Nutrition Label

**DATASET:** ANNUAL COUNTRY-SPECIFIC NETFLIX RELEASES AND PER-CAPITA GDP VALUES

**CREATED BY:** INFO 201 GROUP 55  
Aiman Shamsul, Alisa Tcurko, and Isha Kshirsagar

**CREATED ON:** 5/7/2023

## ABOUT DATASET

This dataset was created for our INFO 201 Final Project, in which we will investigate Netflix's global expansion strategy and answer the question, "**Do countries with the fastest growing economies correspond with the global regions from which Netflix is investing in the most streaming content?**". To help us answer this question, we created a merged dataset that combines the country-specific numbers of new Netflix movie and television series releases per year (based on the country each new release was produced in) with annual per-capita GDP (gross domestic product) values for each country. This dataset contains 5,490 original data points representing the number of new Netflix releases and per-capita GDP values of each country per year from 2000 to 2022. With this data, we seek to determine whether it is possible to predict future loci of high economic growth based on current trends regarding Netflix's global content investments.

## HOW THE DATA WAS SOURCED

We sourced our data from two original datasets, one compiled and released onto the Kaggle open-source data platform by a private citizen about annual country-specific Netflix releases (netflix\_titles.csv), and another about annual, country-specific per-capita GDP values compiled by the U.S. Department of Agriculture's Economic Research Service (GDP\_growth.csv). After merging together the two source datasets and cleaning them to remove any null, unwanted, or irrelevant values, we produced our novel dataset.

### Sources:

<https://www.kaggle.com/datasets/shivamb/netflix-shows>  
<https://ers.usda.gov/data-products/international-macroeconomic-data-set.aspx>

## DATASET GREEN FLAGS (PROS)

- Composition:**
  - Dataset was cleaned and wrangled prior to being released
  - Data is mostly complete and void of any redundancies or copying errors
  - Does not include sample data
  - Mostly includes raw data, but some instances have been cleaned and modified for ease of use
- Provenance:**
  - GDP data came from a highly verified, government-funded source (USDA Economic Research Service)
  - Raw data is from the public domain and is not commercially licensed
- Collection:**
  - Doesn't include any confidential data or data that identifies individual humans or sub-populations (age, race, gender, etc.)
  - Doesn't violate any privacy concerns - includes broad, country-level data
- Management:**
  - Data will be updated until June 2023, but only for instances where errors in the data need to be cleaned.

## DATASET RED FLAGS (CONS)

- Composition:**
  - No data dictionary or metadata repository
  - Not publicly available for peer review
  - Is dependent on data from upstream sources
  - Dataset will not be updated regularly, even if upstream sources are updated over time
- Provenance:**
  - Netflix data came from an unverified, private user from the online platform Kaggle. There is no indication on how this data was collected by said user.
- Collection:**
  - Data was not reviewed for quality
  - No ethical review processes were conducted
- Management:**
  - Data cannot currently be accessed by anyone besides those in the INFO 201 course at the University of Washington
  - Data will not be updated for any reason besides cleaning errors in the data. Data cleaning will end in June 2023.

## INTENDED TO BE USED FOR

- Academic research in the UW INFO 201 course
- Market / consumer research by Netflix and related companies
- Personal research by Netflix stakeholders and investors

## EXAMPLE USE CASES

- Comparing annual geographic Netflix content growth
- Comparing annual geographic GDP growth
- Predicting future geographic trends in Netflix content growth

## POTENTIALLY HARMFUL USES

Making potentially harmful or consequential generalizations about a given country, such as its level of global advancement based on how much content Netflix is releasing from that country

Besides the relative amount of economic opportunities a country may provide for Netflix, there could be many unrelated reasons why Netflix may or may not invest into a certain country that aren't represented by this dataset - such as, but not limited to, the historical size of a country's filmmaking industry or the legal and geopolitical restrictions in a specific country.

## OVERALL RISK ASSESSMENT

- Green Flags: +9 points total
- Red Flags: -9 points total

**TOTAL SCORE: NET ZERO**

This dataset is not obviously skewed towards either the risks or rewards of using it. Users of this dataset should exercise caution and evaluate the pros and cons of the data to determine whether it is appropriate for their individual needs

