



STRANGER  
THINGS

PURWADHIKA DATA SCIENCE FINAL PROJECT:

# NETFLIX MOVIES & TV SHOWS

EDA & CONTENT-BASED RECOMMENDATION

BY ELSA LEVINA

# **PRES**ENTATION **O**VERVIEW

## **I**NTRODUCTION

Project Introduction  
Goals

## **E**DA

Data Cleaning  
Data Visualization & Analysis

## **R**ECOMMENDATION SYSTEM

Creating Content-Based  
Recommendation System

## **C**LOSING

Results  
Conclusion

# INTRODUCTION

# INTRODUCTION

**Netflix** is subscription-based streaming service which offers online streaming of a wide variety of movies and television programs.

As of April 2020, Netflix had over 182 million paid subscriptions worldwide. It seems that Netflix's popularity are increasing day by day especially on this 'stay-at-home' situation.

Therefore I thought it would be interesting to do some analysis on Netflix's contents as well as to build a content-based recommendation system to give Netflix Movies and TV shows suggestions for prospective users.

# GOALS

- Finding what contents are available on Netflix
- Making a simple content-based recommendation system by matching text-based features to identify similar movies / TV shows

# EXPLORATORY DATA ANALYSIS

# DATA CLEANING



# ABOUT THE DATASET

The data used on this project is from **Netflix Movies and TV Shows** dataset founded on Kaggle (<https://www.kaggle.com/shivamb/netflix-shows>).

This dataset consists of TV Shows and Movies available on Netflix as of 2019. It also includes some datas from early 2020. The dataset is collected from Flixable which is a third-party Netflix search engine.

# ABOUT THE DATASET

The dataset has 12 columns as described below:

- **show\_id** = Unique ID for every Movie / Tv Show
- **type** = Identifier - A Movie or TV Show
- **title** = Title of the Movie / Tv Show
- **director** = Director of the Movie
- **cast** = Actors involved in the Movie / TV Show
- **country** = Country where the Movie / TV Show was produced
- **date\_added** = Date it was added on Netflix
- **release\_year** = Actual Release year of the Movie / TV Show
- **rating** = Rating of the Movie / TV Show
- **duration** = Total Duration - in minutes or number of seasons
- **listed\_in** = Genre of the Movie / TV Show
- **description** = The summary description

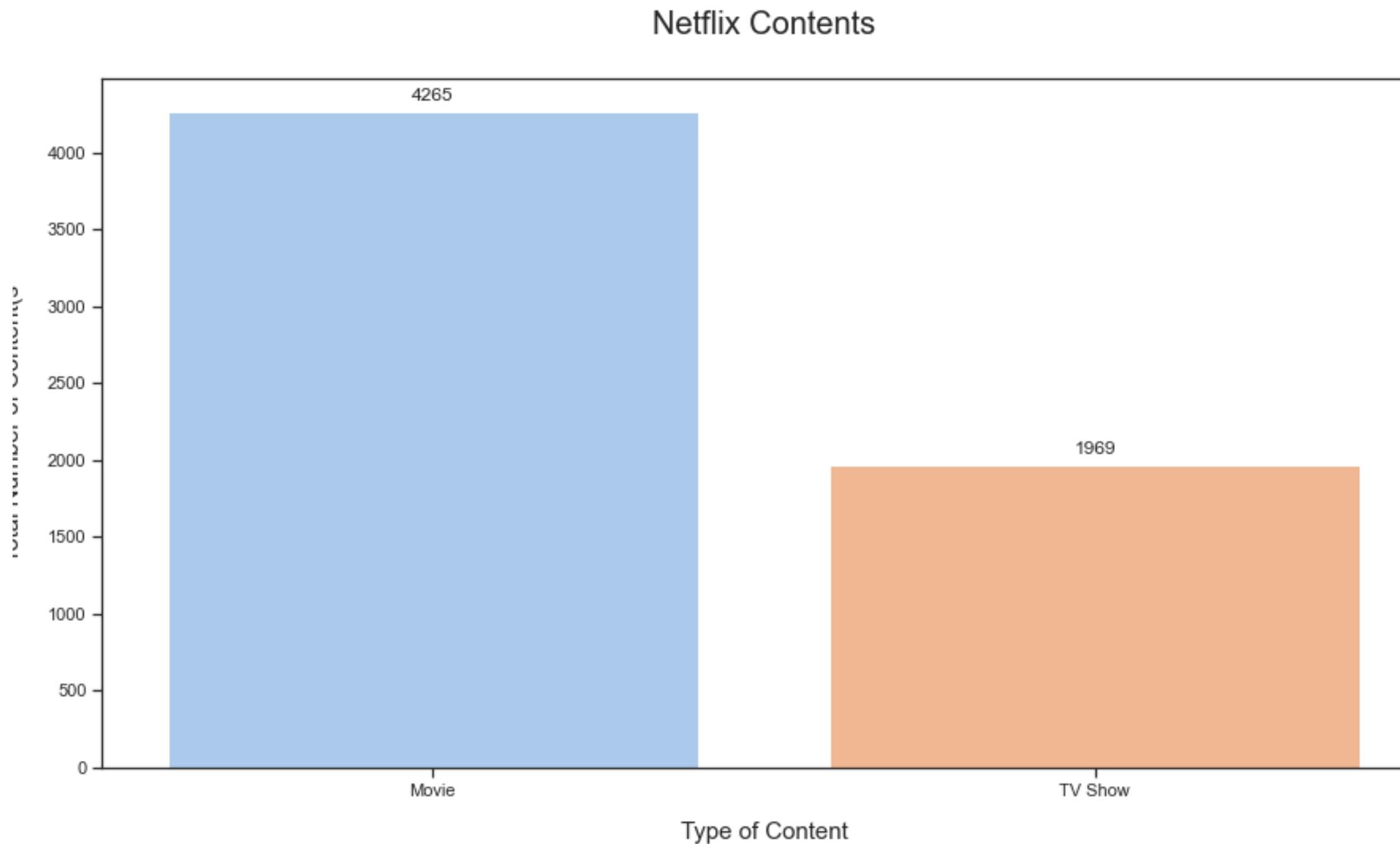
# DATA CLEANING

[ STEP BY STEP EXPLANATION ON JUPYTER NOTEBOOK ]

# DATA VISUALIZATION & ANALYSIS

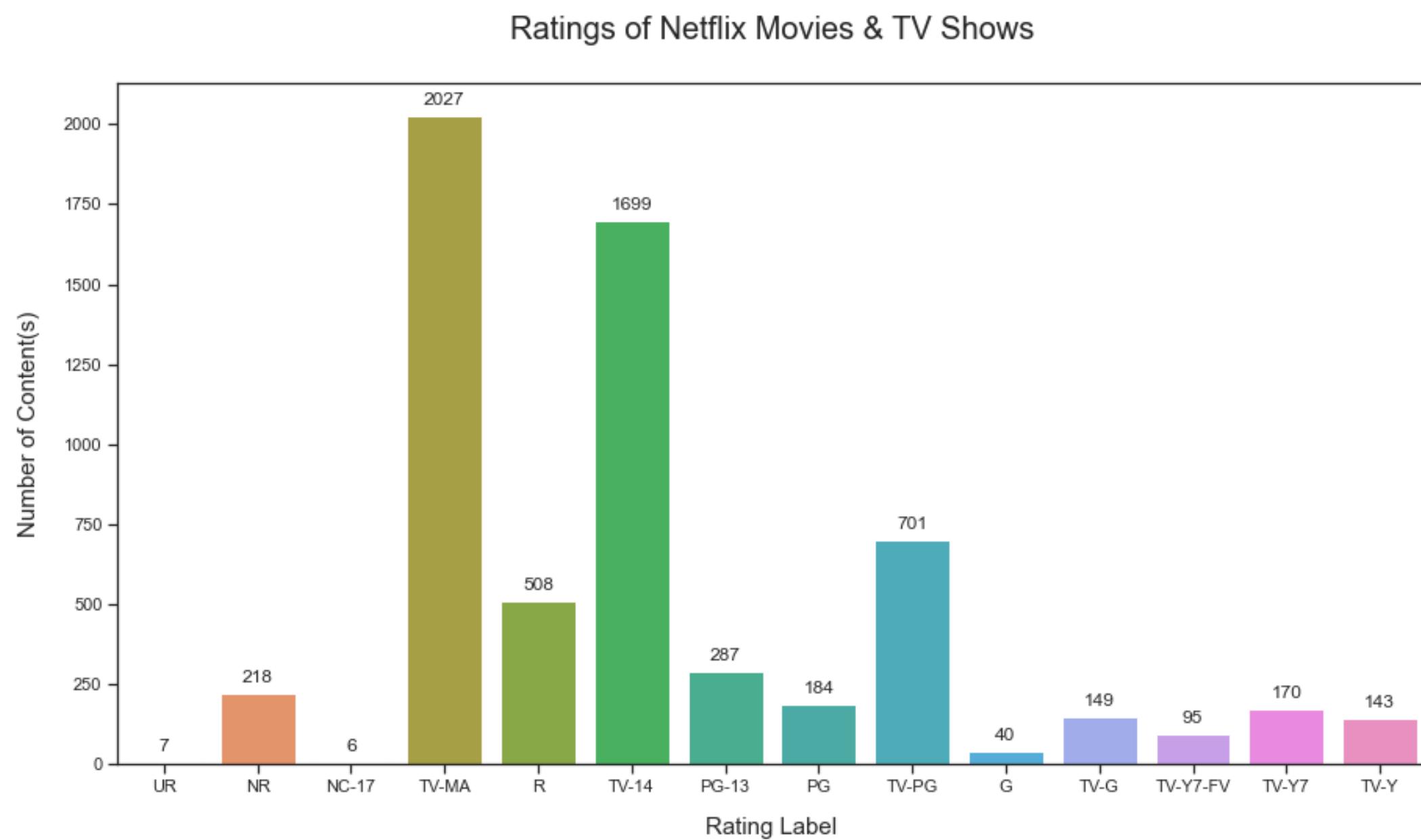


# NETFLIX CONTENT TYPE



- Contents founded on Netflix: Movies & TV Shows
- Movies > TV Shows
- Movies = 68.4%
- TV Shows = 31.6%

# CONTENT RATING



Top 4 most frequent rating:

1. **TV-MA**

(for adults, unsuitable for children under 17)

2. **TV-14**

(some material unsuitable for children under 14)

3. **TV-PG**

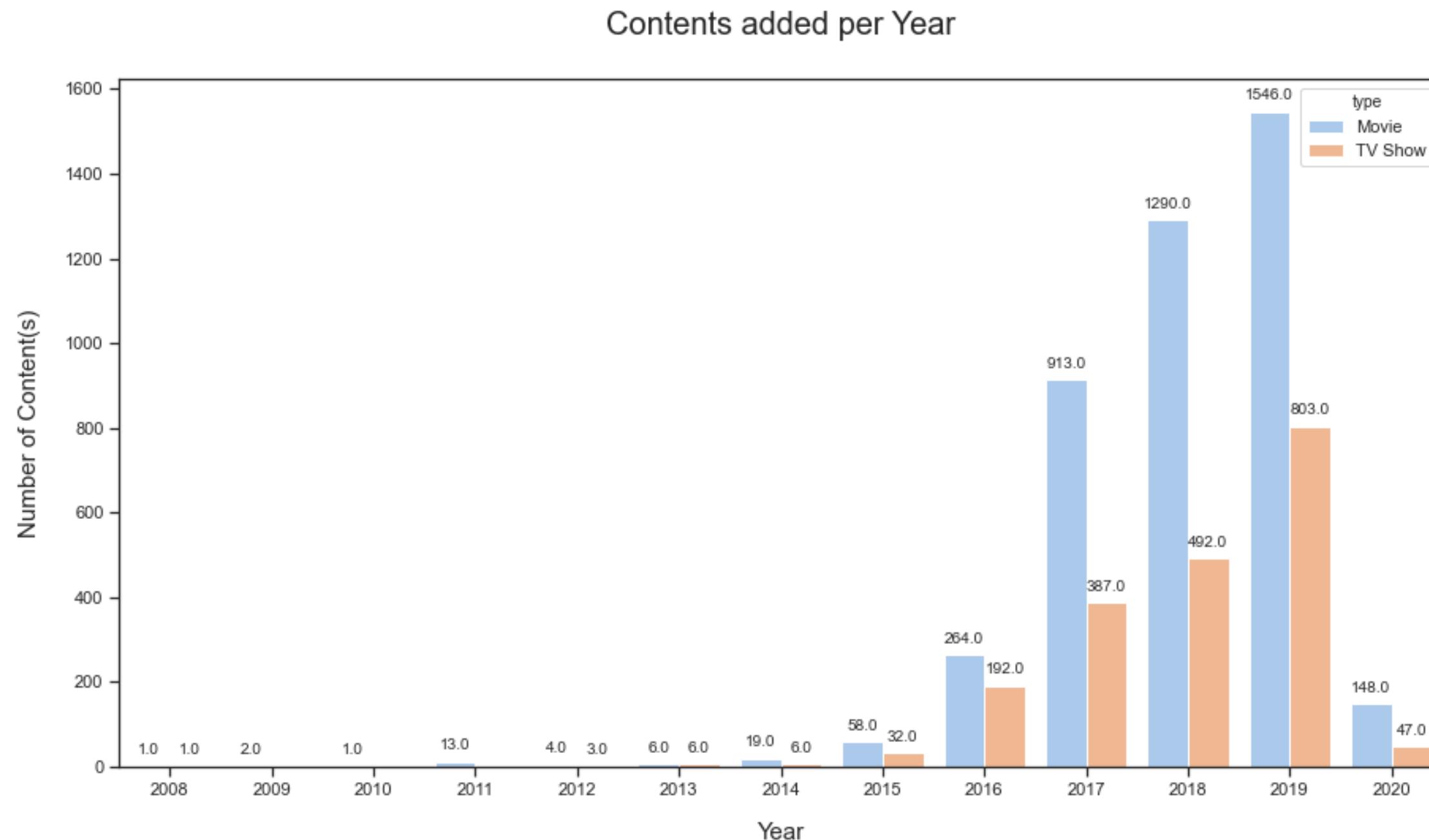
(some material unsuitable for younger children)

4. **R**

(restricted, under 17 requires accompanying parent)

Most contents are designed for older audiences (adults & teenagers / older kids)

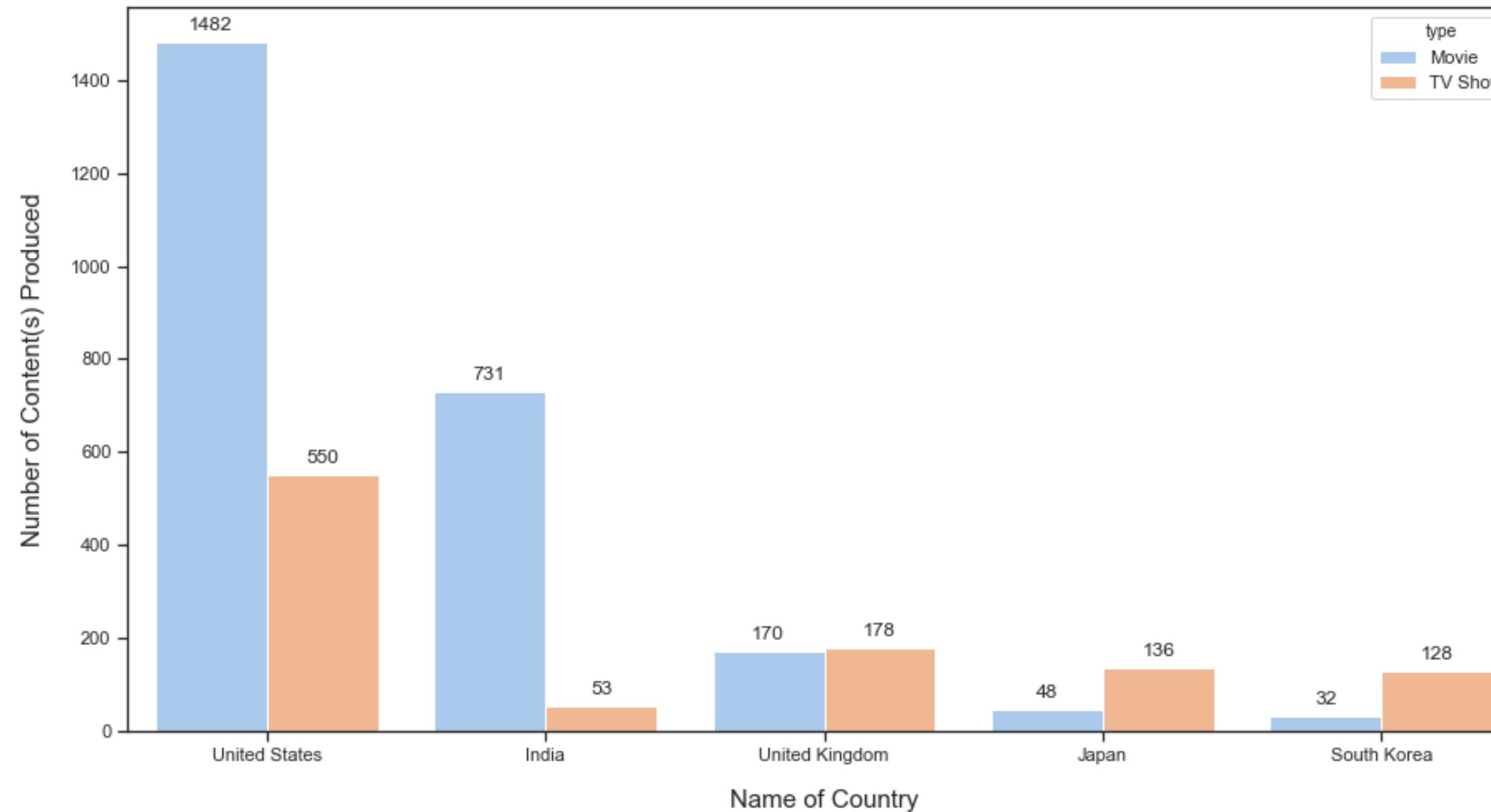
# ADDED CONTENTS PER YEAR



- Netflix is adding more and more contents each year
- Information on year 2020 is not yet fully updated, that's why it seemed low in number
- Seeing the pattern, this numbers are expected to increase with time

# CONTENT PRODUCING COUNTRIES

Top 5 Content Producing Countries



Top 5 content producing countries:

- US

(2032 contents: 72.9% movies, 27.1% TV shows)

- India

(784 contents: 93.2% movies, 6.8% TV shows)

- UK

(348 contents: 48.9% movies, 51.1% TV shows)

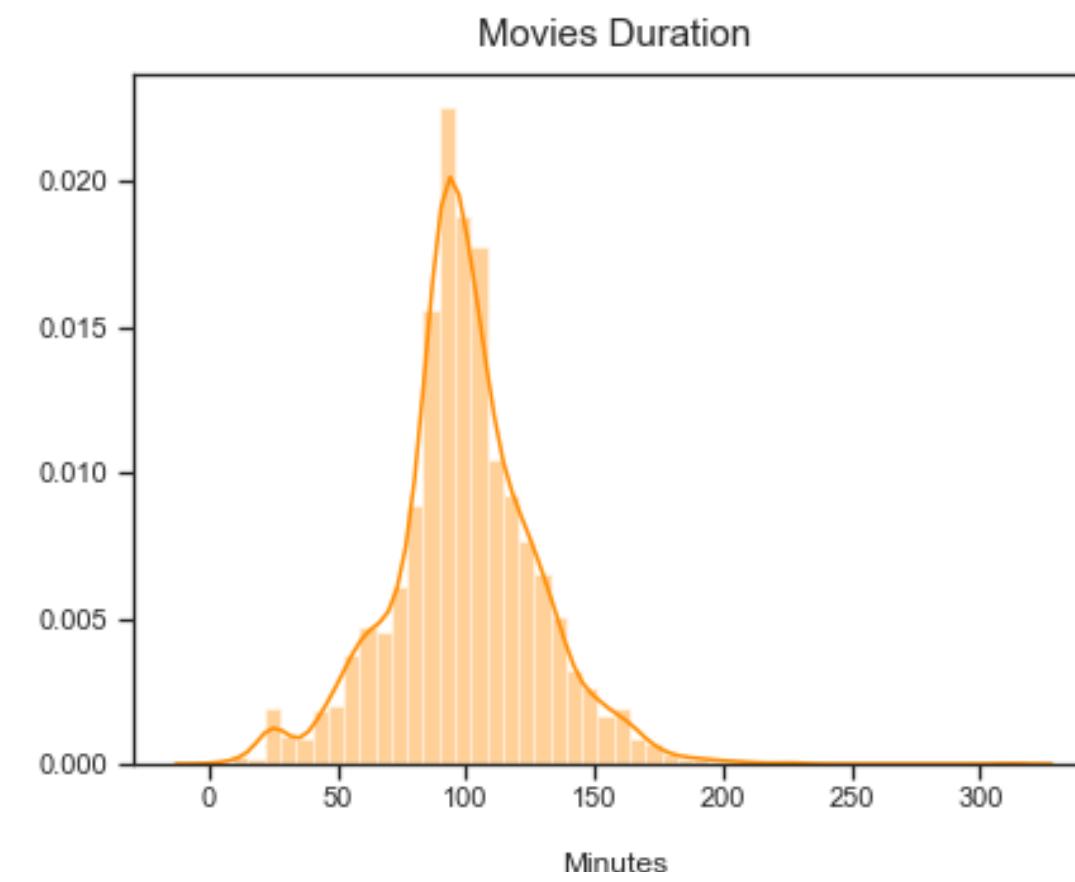
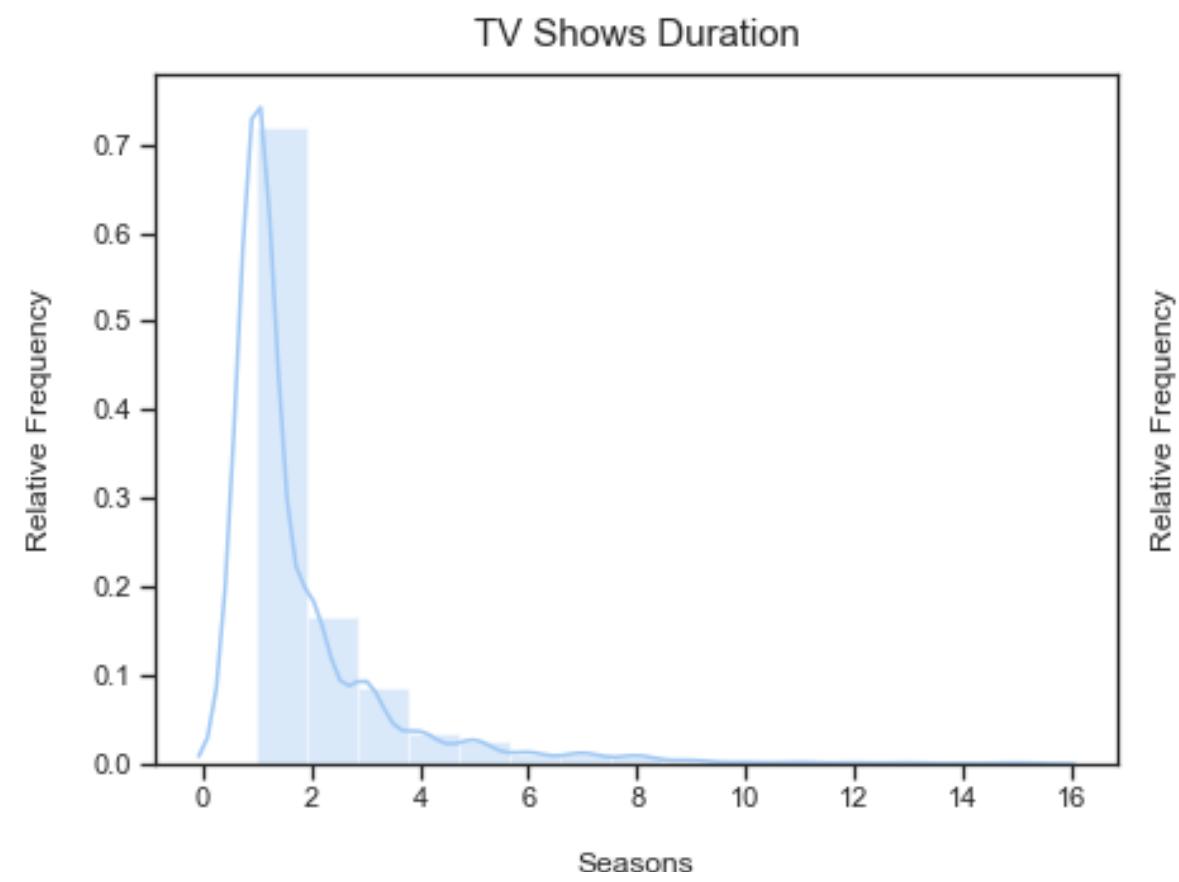
- Japan

(184 contents: 26.1% movies, 73.9% TV shows)

- South Korea

(160 contents: 20% movies, 80% TV shows)

# CONTENT DURATION



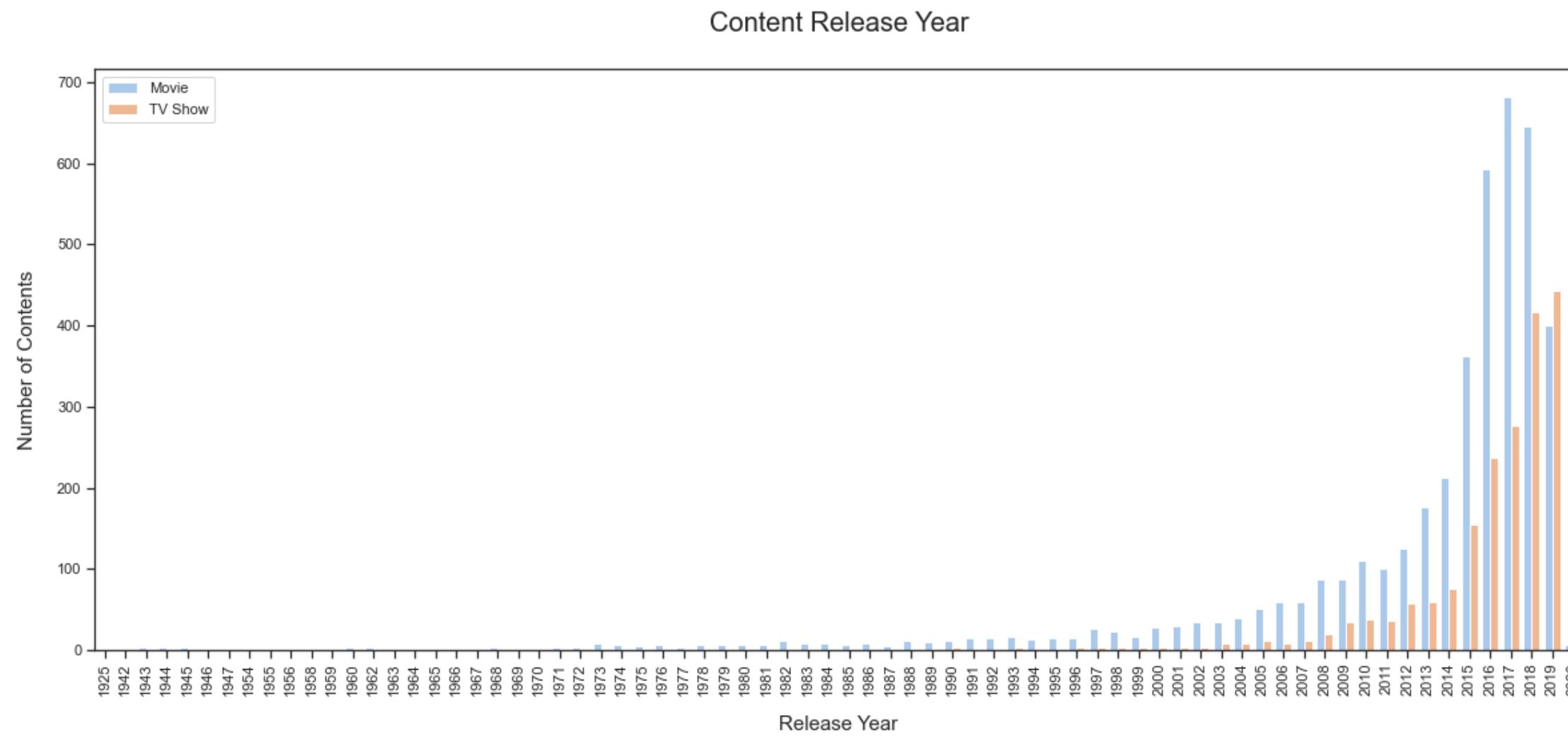
## TV Shows

- Shortest: 1 season
- Longest: 15 seasons
- Most frequent: 1 season
- Average: 1.779 ( 1 - 2 seasons )

## Movies

- Shortest: 3 mins
- Longest: 312 mins
- Most frequent : 90 mins
- Average = 99.1 mins

# CONTENT RELEASE YEAR

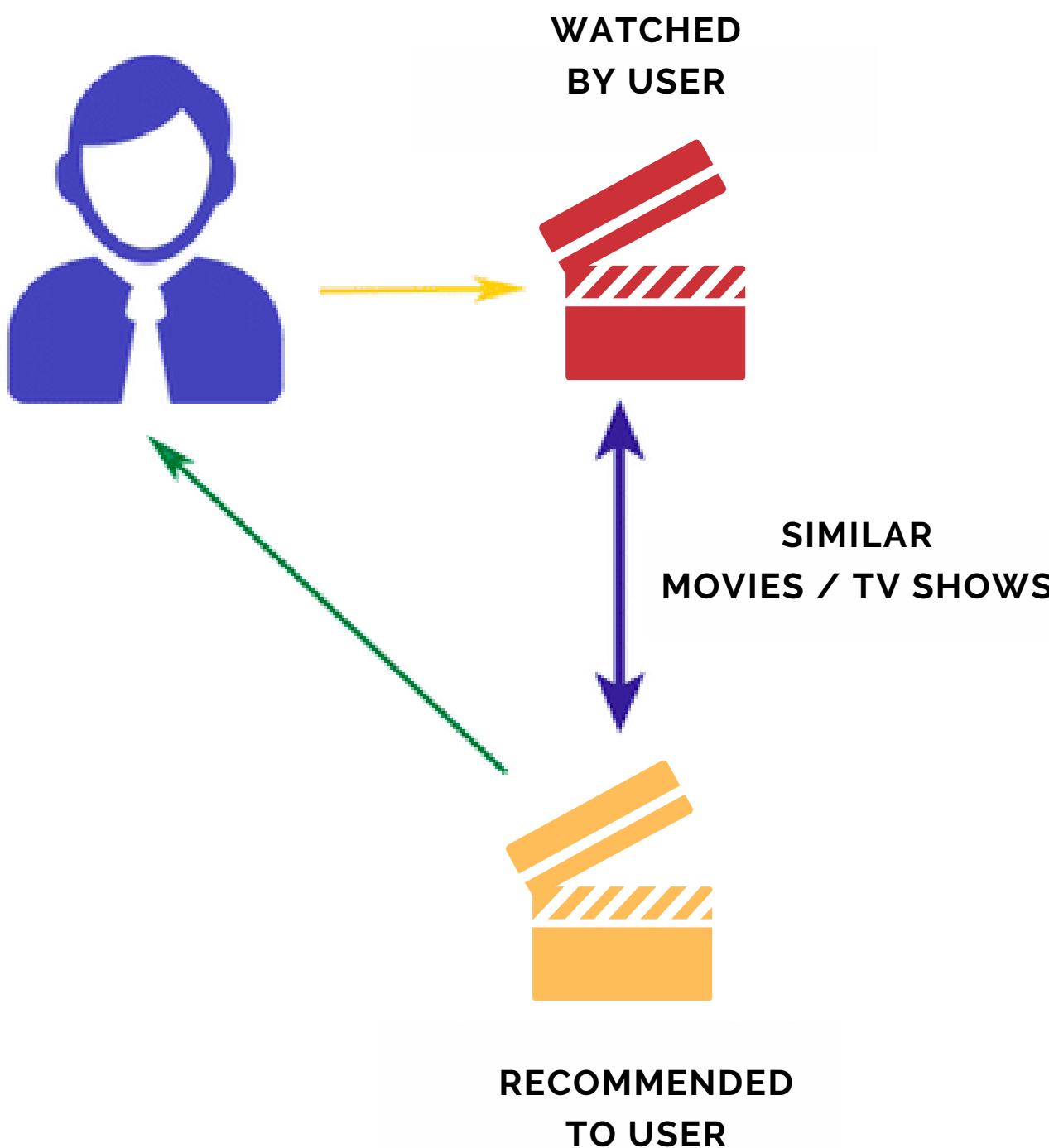


- Oldest content release: 1925
- Latest content release: 2020
- Most frequent content release: 2018

Most of the contents on Netflix are newly released.

# CONTENT-BASED RECOMMENDATION SYSTEM

## CONTENT-BASED FILTERING



# CONTENT-BASED RECOMMENDATION

- Based on a description of the item and user's preferences.
- Best suited to situations where there is known data on an item (name, location, description, etc.), but not on the user.
- Content-based recommendation focused on the similarity between items to create a recommendation for the user.

**[ STEP BY STEP EXPLANATION ON JUPYTER NOTEBOOK ]**

# RESULTS & CONCLUSION

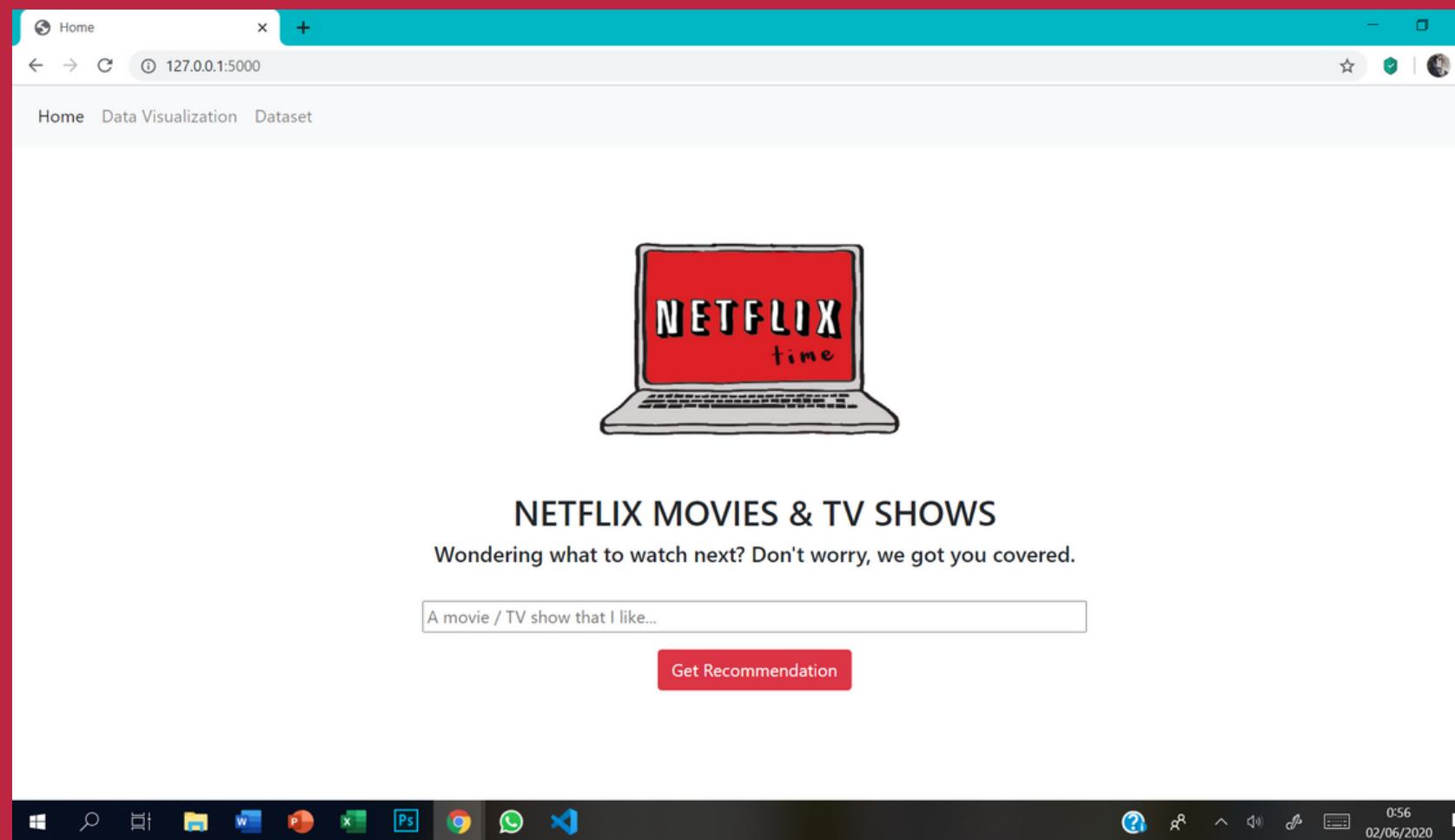
# RESULTS

## WHAT CONTENTS ARE AVAILABLE ON NETFLIX?

- We can find movies and TV shows on Netflix. There are more movies than TV shows, being 68.4% of the contents are movies and 31.6% are TV shows.
- Based on the ratings, the contents are mostly targeted on adults and teenagers/older kids.
- Most contents are produced in the US, India, UK, Japan and South Korea respectively.
- Most of the TV shows have around 1-2 seasons and most of the movies are around 90 minutes long.
- Although there are some classic contents available, most of the contents on Netflix are newly released.
- Netflix adds more and more contents each year, the pattern shows that it's expected to increase in the future.

# RESULTS

## CONTENT-BASED RECOMMENDATION SYSTEM

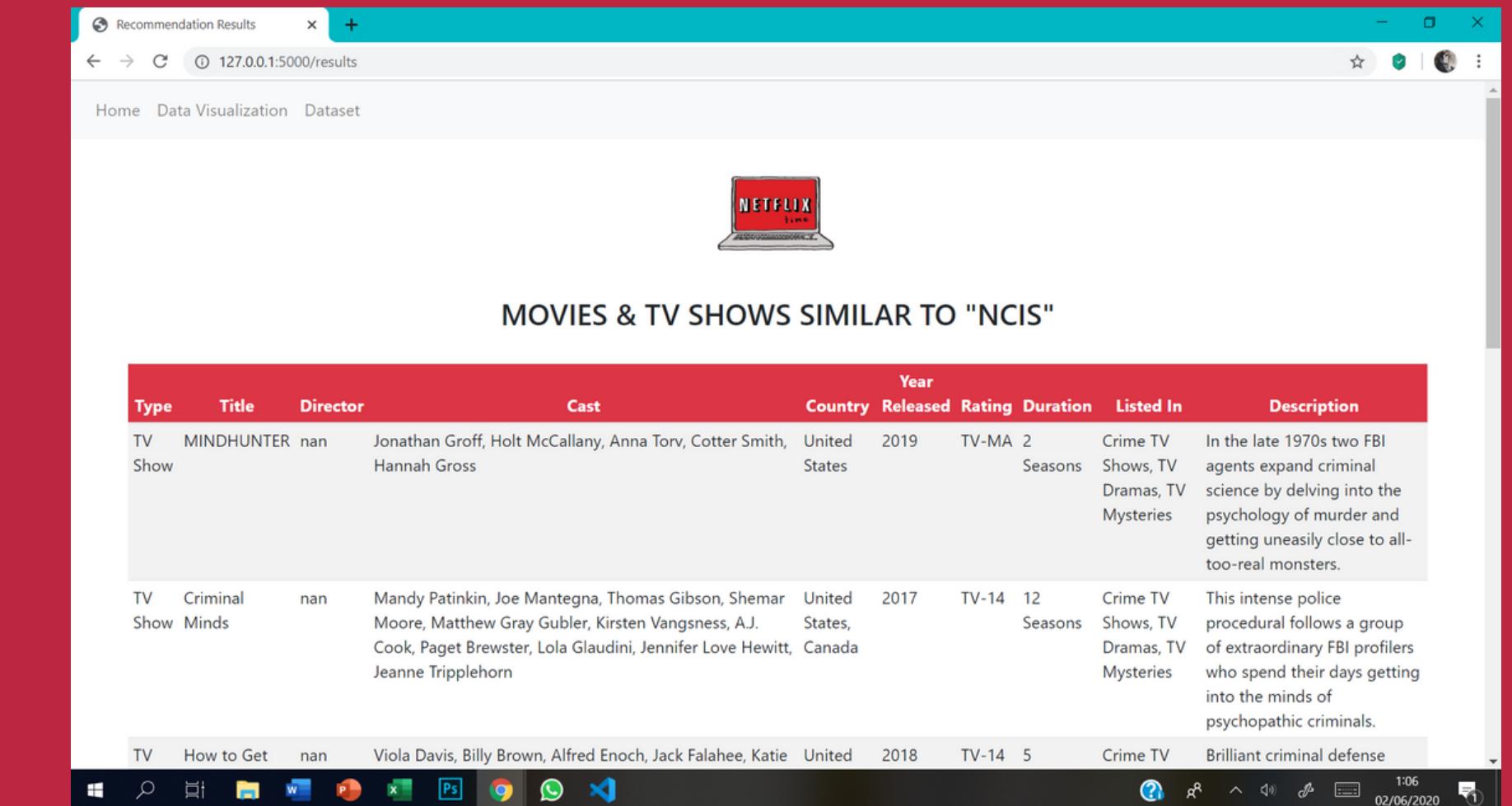


NETFLIX MOVIES & TV SHOWS

Wondering what to watch next? Don't worry, we got you covered.

A movie / TV show that I like...

Get Recommendation



MOVIES & TV SHOWS SIMILAR TO "NCIS"

Type	Title	Director	Cast	Year	Country	Released	Rating	Duration	Listed In	Description
TV Show	MINDHUNTER	nan	Jonathan Groff, Holt McCallany, Anna Torv, Cotter Smith, Hannah Gross	United States	2019	TV-MA	2 Seasons	Crime TV Shows, TV Dramas, TV Mysteries	In the late 1970s two FBI agents expand criminal science by delving into the psychology of murder and getting uneasily close to all-too-real monsters.	
TV Show	Criminal Minds	nan	Mandy Patinkin, Joe Mantegna, Thomas Gibson, Shemar Moore, Matthew Gray Gubler, Kirsten Vangsness, A.J. Cook, Paget Brewster, Lola Glaudini, Jennifer Love Hewitt, Jeanne Tripplehorn	United States, Canada	2017	TV-14	12 Seasons	Crime TV Shows, TV Dramas, TV Mysteries	This intense police procedural follows a group of extraordinary FBI profilers who spend their days getting into the minds of psychopathic criminals.	
TV	How to Get Away With Murder	nan	Viola Davis, Billy Brown, Alfred Enoch, Jack Falahee, Katie Lowes, Liza Weil, Matt Craven, Laurel Holloman, Tony Goldwyn, Cressida Dickson, Kaitlyn Dever, Sage Brocklebank, Isabellaapostoli, Daniel P. Johnson, Michael Ely, J. R. Cacia, Leah C. Williams, Samira Wiley, Aja Naomi King, Yara Shahidi, Holly Hunter, Lela Loren, Liane Balaban, Lorraine Bracco, Viola Davis, Billy Brown, Alfred Enoch, Jack Falahee, Katie Lowes, Liza Weil, Matt Craven, Laurel Holloman, Tony Goldwyn, Cressida Dickson, Kaitlyn Dever, Sage Brocklebank, Isabellaapostoli, Daniel P. Johnson, Michael Ely, J. R. Cacia, Leah C. Williams, Samira Wiley, Aja Naomi King, Yara Shahidi, Holly Hunter, Lela Loren, Lorraine Bracco	United States	2018	TV-14	5 Seasons	Crime TV	Brilliant criminal defense	

# RESULTS

## CONTENT-BASED RECOMMENDATION SYSTEM

The screenshot shows a web browser window with a teal header bar. The address bar displays "127.0.0.1:5000". The main content area features a large illustration of a laptop screen with the Netflix logo. Below the illustration, the text "NETFLIX MOVIES & TV SHOWS" is centered. A subtext below it reads "Wondering what to watch next? Don't worry, we got you covered." At the bottom, there is a search input field with the placeholder "A movie / TV show that I like..." and a red "Get Recommendation" button.

The screenshot shows a web browser window with a teal header bar. The address bar displays "127.0.0.1:5000/results". The main content area features a large illustration of a laptop screen with the Netflix logo. Below the illustration, the text "OH NO!" is displayed in large, bold, black letters. A subtext below it reads "The movie / TV show you inputted is not available on Netflix. Please try again with another title." At the bottom, there is a red "Back to Home" button.

# CONCLUSION

Netflix contents are consisted of movies and TV shows. A lot of them are new-released, not very long in duration and mostly targeted for older audiences (adults & teenagers). There are contents from a lot of different countries but the majority of them are produced in the US, India, UK, Japan and South Korea. New contents are added more by each year and the pattern shows that the numbers are expected to increase in the future.

# CONCLUSION



After tested by inputting various movie and TV show titles, it can be said that the model works nicely. It **gives appropriate movies and TV shows recommendation with similar genres, plot, casts and/or directors** as the one inputted.

This simple content-based recommendation system uses NLP and Bag of Words to find similar movies / TV shows to recommend by matching text-based features. It still has rooms for further developments to give better product recommendation to prospective users.

STRANGER  
THINGS

THANK  
YOU