Netflix : Visualization and Recommendation System

Andi Anugrah Mahyuddin

NETFLIX IS A STREAMING SERVICE THAT OFFERS A WIDE VARIETY OF AWARD-WINNING TV SHOWS, MOVIES, ANIME, DOCUMENTARIES, AND MORE ON THOUSANDS OF INTERNET-CONNECTED DEVICES.
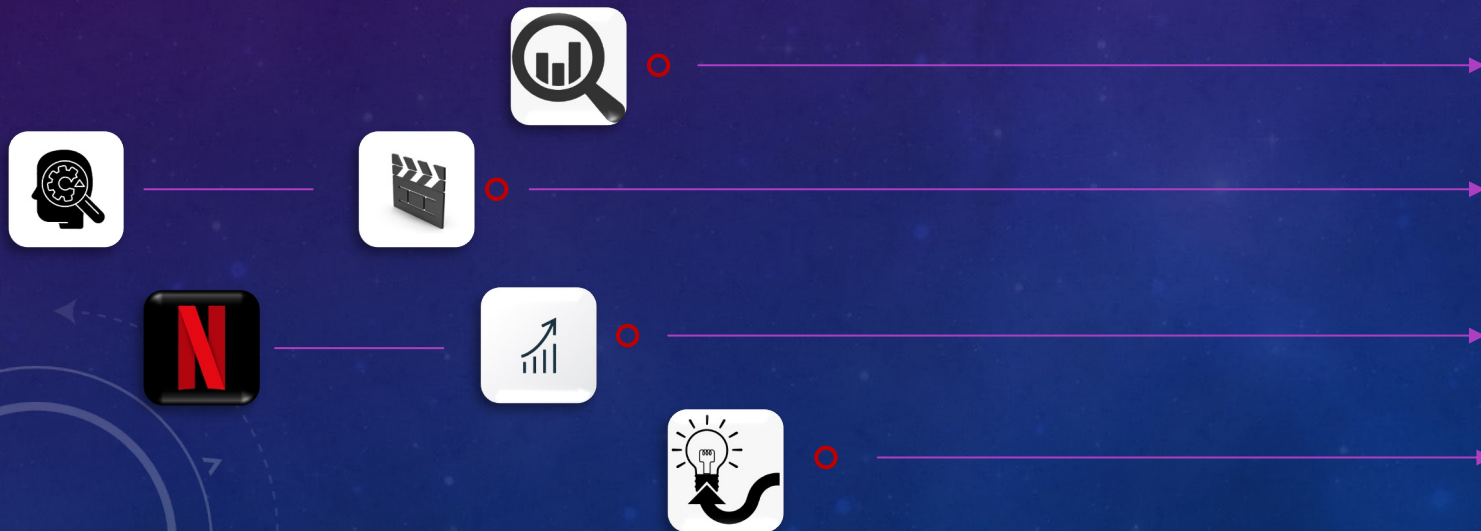
NETFLIX

2022

Problem Statement

This dataset consists of tv shows and movies available on Netflix as of 2019. The dataset is collected from Flixable which is a third-party Netflix search engine.

In 2018, they released an interesting report which shows that the number of TV shows on Netflix has nearly tripled since 2010. The streaming service's number of movies has decreased by more than 2,000 titles since 2010, while its number of TV shows has nearly tripled. It will be interesting to explore what all other insights can be obtained from the same dataset.

This project required to do :

1. Exploratory Data Analysis

2. Understanding what type content is available in different countries

3. Is Netflix has increasingly focusing on TV rather than movies in recent years.

4. Recomendation System

ABOUT DATASET

8807 ROWS

12 COLUMNS

1. There are 8807 rows and 12 columns
2. 11 of the columns are object
3. 1 column is an integer
4. There are missing values in some columns

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8807 entries, 0 to 8806
Data columns (total 12 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   show_id       8807 non-null   object
 1   type          8807 non-null   object
 2   title         8807 non-null   object
 3   director      6173 non-null   object
 4   cast          7982 non-null   object
 5   country       7976 non-null   object
 6   date_added    8797 non-null   object
 7   release_year  8807 non-null   int64
 8   rating        8803 non-null   object
 9   duration      8804 non-null   object
 10  listed_in     8807 non-null   object
 11  description   8807 non-null   object
dtypes: int64(1), object(11)
```

DATA PREPROCESSING

Data Cleaning

Exploratory Data Analysis

First null-count summary:

```
show_id           0
type              0
title             0
director       2634
cast            825
country         831
date_added       10
release_year      0
rating            4
duration          3
listed_in         0
description       0
dtype: int64
```

Second null-count summary:
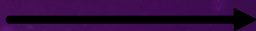
```
show_id           0
type              0
title             0
director          0
cast              0
country           0
date_added        0
release_year      0
rating            0
duration          3
listed_in         0
description       0
dtype: int64
```
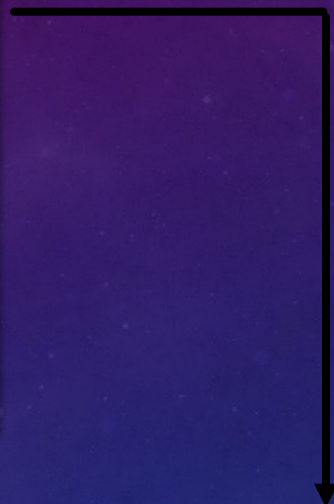
Result:

rating	duration
	74 min
	84 min
	66 min

|	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
5541	s5542	Movie	Louis C.K. 2017	Louis C.K.	Louis C.K.	United States	April 4, 2017	2017	74 min	NaN	Movies	Louis C.K. muses on religion, eternal love, gi...
5794	s5795	Movie	Louis C.K.: Hilarious	Louis C.K.	Louis C.K.	United States	September 16, 2016	2010	84 min	NaN	Movies	Emmy-winning comedy writer Louis C.K. brings h...
5813	s5814	Movie	Louis C.K.: Live at the Comedy Store	Louis C.K.	Louis C.K.	United States	August 15, 2016	2015	66 min	NaN	Movies	The comic puts his trademark hilarious/thought...

VISUALIZATION

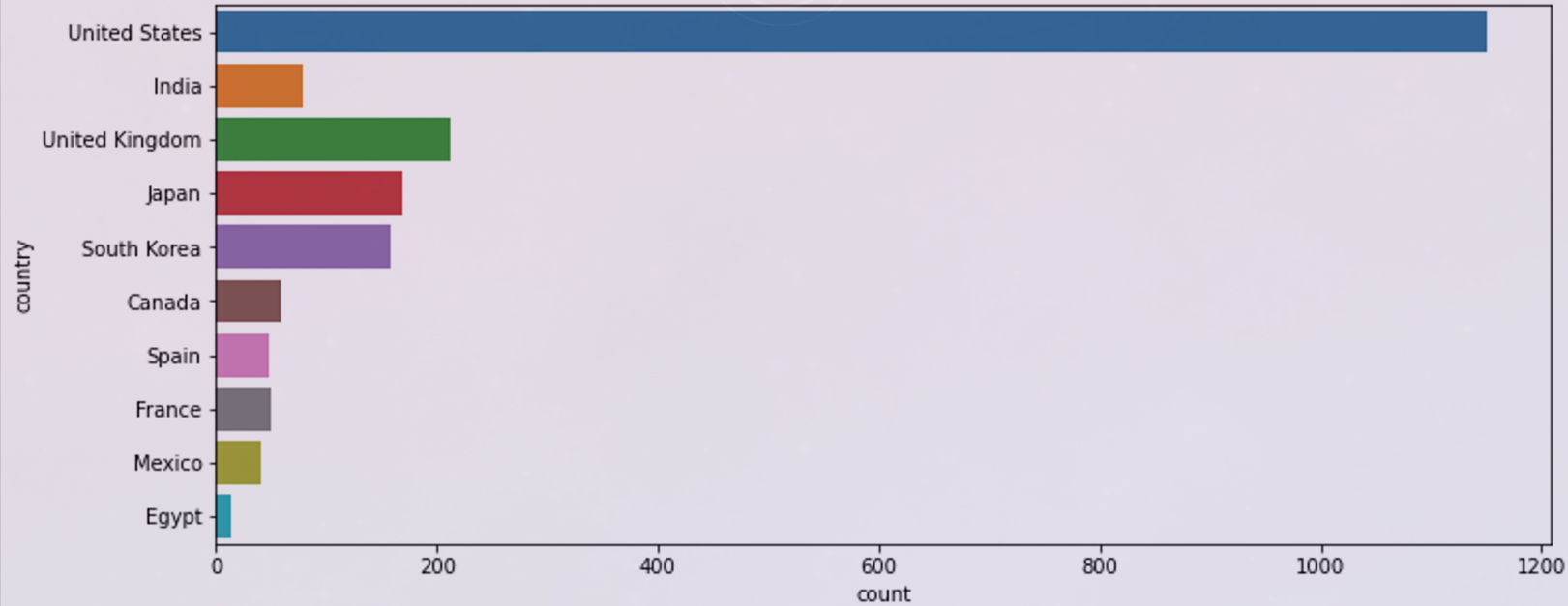Netflix has more movie than TV Show
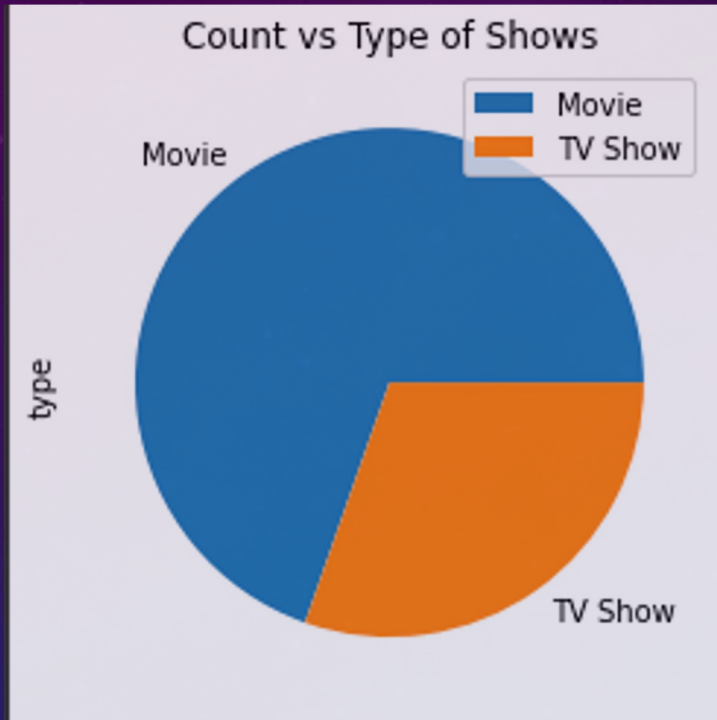
Movie : 6131
TV Show : 2676

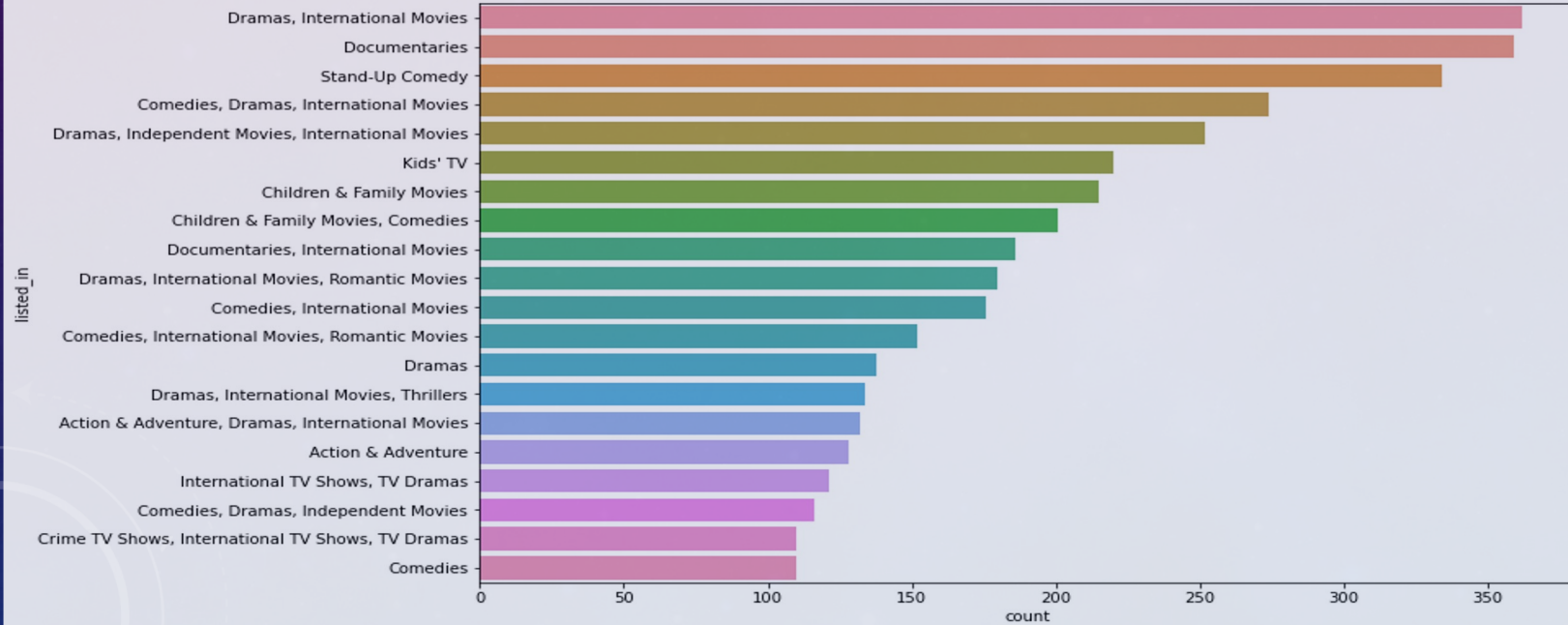United States has producing Movies in Netflix more than 2000
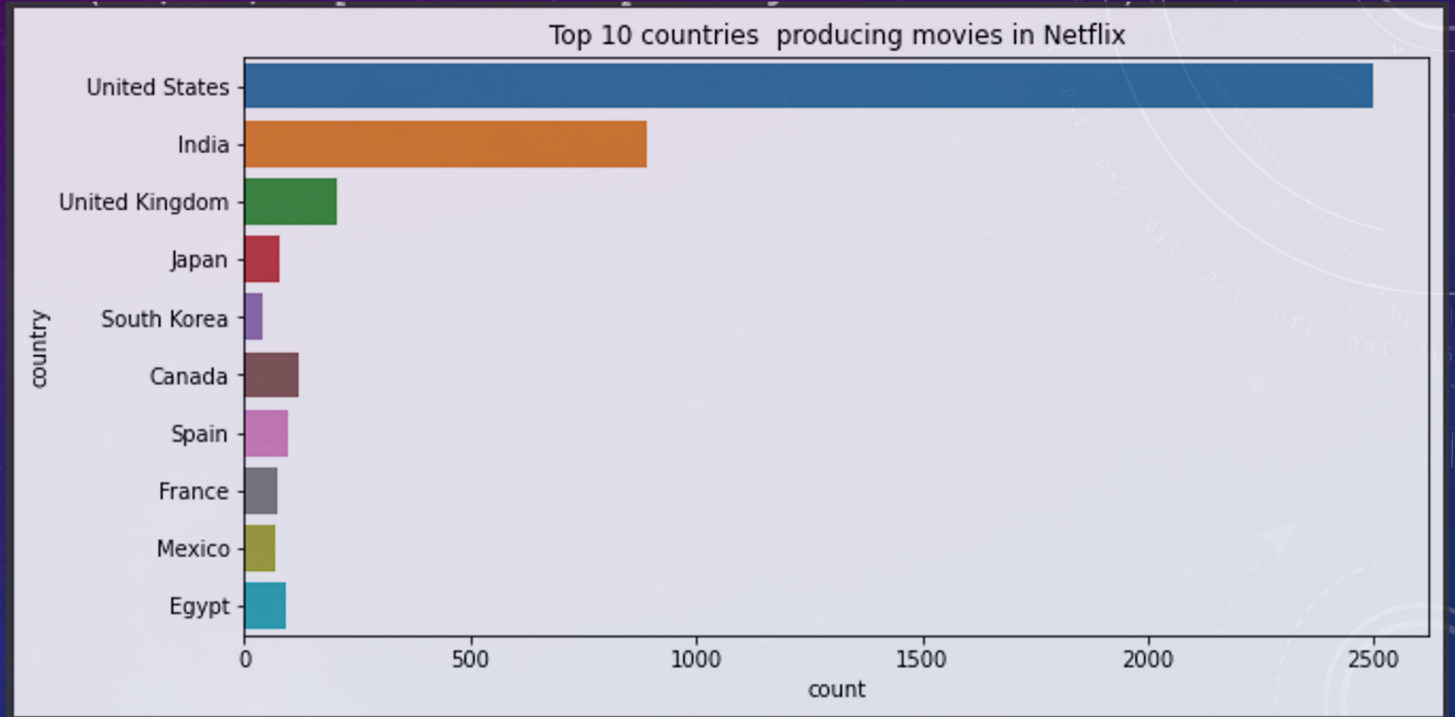
United States : 2818
India : 972
United Kingdom : 419

Top 10 Countries Producing TV shows in Netflix

United States has producing TV Show in Netflix more than 1000

Top 20 Genres on Netflix

The most popular genre in Netflix are International Drama Movies and Drama
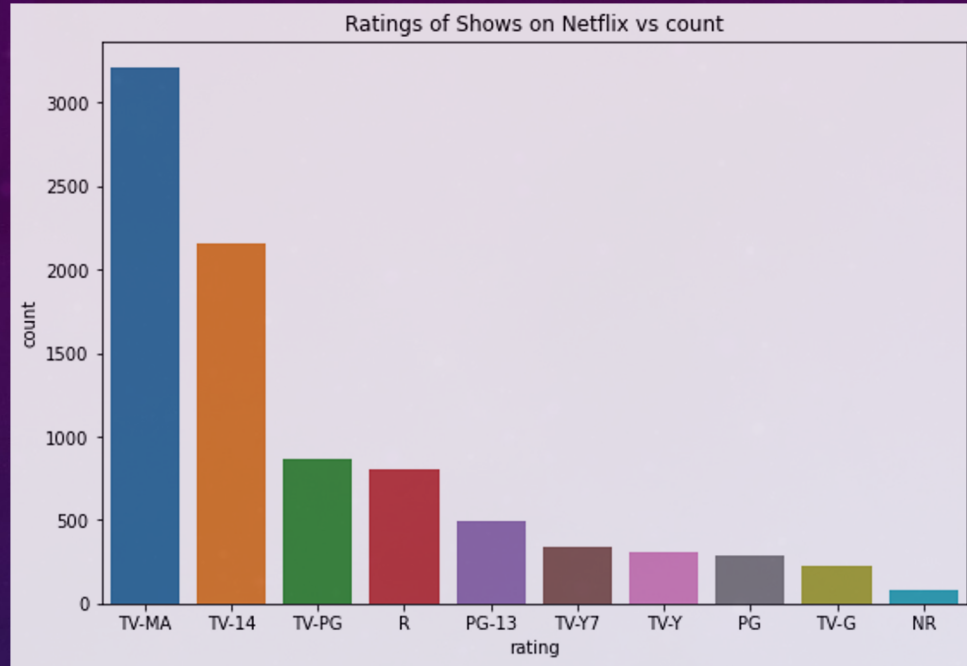
Dramas, International Movies : 362
Documentaries : 359
Stand-Up Comedy : 334
Comedies, Dramas, International Movies : 274

Most of The Shows has TV-MA and TV-14 ratings

Ratings of Shows on Netflix vs count

TV-MA : 3207
TV-14 : 2160
TV-PG : 863
R : 799

Content release in years on Netflix is in 2018

Content Release in Years on Nteflix vs Count
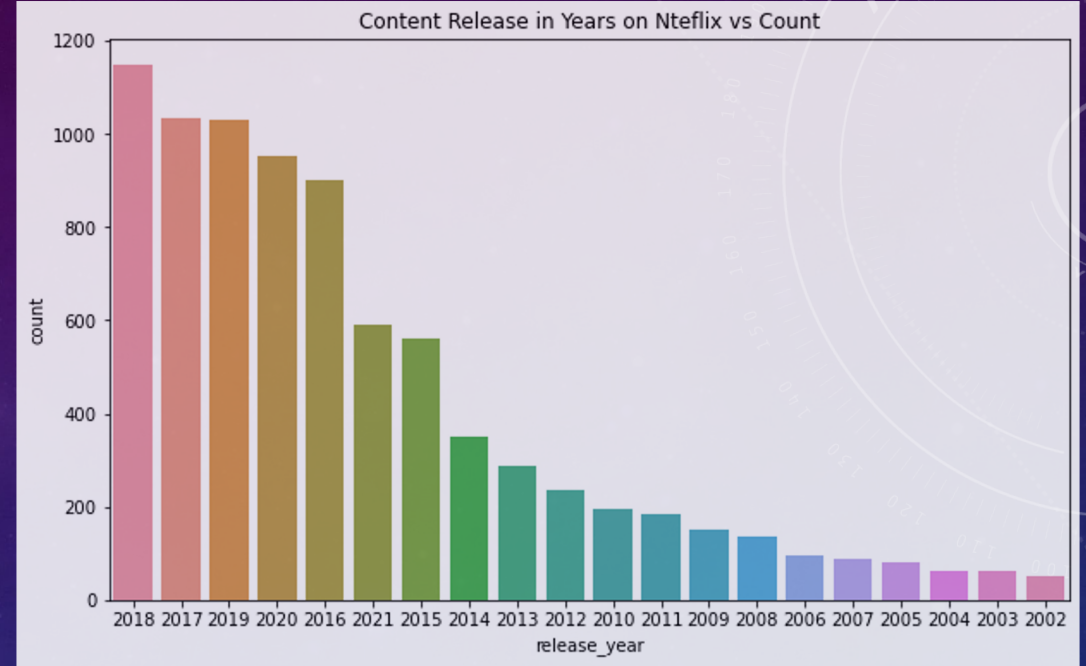
2018 : 1147
2017 : 1032
2019 : 1030

RECOMMENDATION SYSTEM MODEL

CONTENT BASED FILTERING

TF-IDF(Term Frequency-Inverse Document Frequency

Consine Similarity

$$TF = \frac{\text{number of times the term appears in the document}}{\text{total number of terms in the document}}$$

$$IDF = log(\frac{\text{number of the documents in the corpus}}{\text{number of documents in the corpus contain the term}})$$

$$TF\text{-}IDF = TF * IDF$$

The cosine similarity method is a method for calculating the similarity between two objects expressed in two vectors by using keywords from a document as a measure.
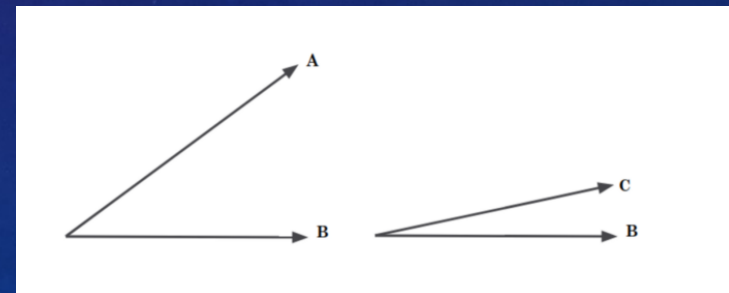
```
get_recommendation('Blood & Water')

5344        Message from the King
1884          Walk Away from Love
4285                        Lilli
4271                    Lion Pride
4209                    Next Enti?
613                      Voiceless
108                      Dive Club
1905                   Cold Harbour
6289                        Bewafaa
5485                      Ram Jaane
Name: title, dtype: object
```

```
get_recommendation('Peaky Blinders')

7683                    Our Godfather
2646                  My Stupid Boss
3133                            Don
8293                        The Fear
7140    Jonathan Strange & Mr Norrell
7785              Power Rangers Zeo
8467                      The Prison
8539                      The Tudors
1510                  The Con Is On
8391    The Legend of Michael Mishra
Name: title, dtype: object
```

CONCLUSION

1. NETFLIX HAS MORE MOVIES THAN TV SHOWS
2. MOST NUMBER OF MOVIES AND TV SHOWS ARE PRODUCED BY UNITED STATES AND INDIA
3. MOST OF THE CONTENT ON NETFLIX MOVIES AND TV SHOWS IS FOR MATURE AUDIENCE
4. IN 2018 IS THE YEAR IN WICH NETFLIX RELEASE A LOT MORE CONTENT AS COMPARED TO OTHER YEARS
5. INTERNATIONAL MOVIES AND DRAMA ARE THE MOST POPULAR GENRE ON NETFLIX
6. ON THE SYSTEM RECOMMENDATION WE CAN SEE A LIST OF MOVIES THAT HAVE SIMILARITIES TO THE MOVIE YOU WANT TO WATCH

THANK YOU