Netflix Data Analytics
›A Case Study

INTRODUCTION

- Netflix, Inc. is an American media company based in Los Gatos, California. Founded in 1997 by Reed Hastings and Marc Randolph in Scotts Valley, California, it operates the over-the-top subscription video on-demand service Netflix brand, which includes original films and television series commissioned or acquired by the company, and third-party content licensed from other distributors. Netflix is a member of the Motion Picture Association—having become the first streaming company to become a member.

- Netflix initially both sold and rented DVDs by mail, but the sales were eliminated within a year to focus on the DVD rental business.

- In 2007, Netflix introduced streaming media and video on demand. The company expanded to Canada in 2010, followed by Latin America and the Caribbean.

- In 2013, the service began to acquire and produce original content, beginning with the political drama House of Cards.

- By 2022, original productions accounted for half of Netflix's library in the United States, and the company had ventured into other categories, such as video game publishing via the Netflix service.

VALUE AND SUBSCRIPTIONS

▶ Netflix's current company valuation is $234 billion. It is currently renowned as the most valued company/media company in the world and transcends even Disney. The success lies in a secret term that is no secret (but the way it's used in a certain way is a secret) — customer retention.

▶ Customer retention may be defined as the process of engaging the customers and appealing to them to use the service or buy the product.

▶ Now, this may look like a simple tactic at first glance but do note that this is considered by many as the most powerful tactic used by any media company. And Netflix used it so intelligently that their customer retention rate is extremely impressive and keeps increasing over the years.

▶ Can you guess the total subscribers of Netflix? Up until December 2020, Netflix subscribers (paid subscription) amounted to a whopping 203.66 million. This is an excellent milestone for Netflix, as it has crossed the 200 million mark for the first time.

CUSTOMER RETENTION

► Netflix's ability to collect and use the data is the reason behind its success. Itresults in better customer retention per year.

► The study says the rate of customer retention is increasing on Netflix because 80%of users follow the recommendation, and the recommended show or movie is streamed.

► Have you ever heard of 'green-light original content'? Green-light means being allowed to do something.

► So, green-lit original content is verified or rated content approved on the basis of various touchpoints taken from the user database.

MARKET POSITION

▶ One of the main differences between Netflix and its competitors is its massive wealth of original content.

▶ Worldwide, Netflix was estimated to spent around 18 billion U.S. dollars on its content in 2022, a value which is expected to expand in the future.

▶ The popularity of shows such as "House of Cards," "Stranger Things," and "Orange is the New Black" have made original programming integral to the company's continued success. Over half of users from the U.S. stated in a February 2021 survey that the announcement of new exclusive programming was their main reason to subscribe to Netflix, and nearly 50 percent of Netflix subscribers stated that the quality of Netflix original shows and movies are better now than they used to be in the past.

DISCRIPTION OF THE PROJECT

- ▶ This project is about Analysing and Visualizing the data from an example data set of Netflix.

- ▶ As Netflix is a subscription-based streaming service that allows subscribers to watch TV shows and movies on an internet-connected device.

- ▶ In this project we can analyze what content is being watched the most in different countries.

- ▶ Comparing the similar content that is screened in different countries of the world.

- ▶ The actors and director worked in a film

Exploratory Data Analysis on Netflix

1. Data Description

- **show_id :** It denotes the unique key for every show listed in Netflix.
- **type :** There is only two type of video content available in Netfilx, i.e. Movie and TV Shows
- **title :** Title is the name of the show/movie.
- **director :** Name of the director(s) who is responsible for production of such specific content(s).
- **cast :** People who acted in such content are given in this column.
- **country :** Country column is given here as some content are contry specific, we can see which content plays in which country.
- **date_added :** This column shows us on which date a content come online in Netflix for the first time.
- **release_year:** Content publishing year on TV or box-office for the first time.
- **rating :** Internal rating from Netflix.
- **duration :** Watch time of the Netflix content.
- **listed_in :** This ia genre, the type of the content, like Documentaries, TV shows, TV Dramas, etc.
- **description :** A short description about the content.

Getting the Data

```
In [2]: import pandas as pd
        import numpy as np
        netflix = pd.read_csv('netflix_titles.csv')
        pd.set_option('display.max_columns',None) # display all the features
        netflix.head(5)
```

Out[2]:

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
0	s1	Movie	Dick Johnson Is Dead	Kirsten Johnson	NaN	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	As her father nears the end of his life, filmm...
1	s2	TV Show	Blood & Water	NaN	Ama Qamata, Khosi Ngema, Gail Mabalane, Thaban...	South Africa	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, TV Dramas, TV Mysteries	After crossing paths at a party, a Cape Town t...
2	s3	TV Show	Ganglands	Julien Leclercq	Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nabi...	NaN	September 24, 2021	2021	TV-MA	1 Season	Crime TV Shows, International TV Shows, TV Act...	To protect his family from a powerful drug lor...
3	s4	TV Show	Jailbirds New Orleans	NaN	NaN	NaN	September 24, 2021	2021	TV-MA	1 Season	Docuseries, Reality TV	Feuds, flirtations and toilet talk go down amo...
4	s5	TV Show	Kota Factory	NaN	Mayur More, Jitendra Kumar, Ranjan Raj, Alam K...	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...	In a city of coaching centers known to train I...

2. Analysing the Data

Data Information

In [3]: `print(netflix.info())`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8807 entries, 0 to 8806
Data columns (total 12 columns):
 #    Column         Non-Null Count   Dtype
----  -------------------------------- ----------------  --------
 1    show_id        8807 non-null    object
 2    type           8807 non-null    object
 3    title          8807 non-null    object
 4    director       6173 non-null    object
 5    cast           7982 non-null    object
 6    country        7976 non-null    object
 7    date_added     8797 non-null    object
 8    release_year   8807 non-null    int64
 9    rating         8803 non-null    object
 10   duration       8804 non-null    object
 11   listed_in      8807 non-null    object
 12   description    8807 non-null    object
dtypes: int64(1), object(11)
memory usage: 825.8+ KB
None
```

- Dataset contains object and intiger type data where 10 features are in object type and 1 in int64 type.

In [4]: `netflix.describe()`

Out[4]:

	release_year
count	8807.000000
mean	2014.180198
std	8.819312
min	1925.000000
25%	2013.000000
50%	2017.000000
75%	2019.000000
max	2021.000000

Checking Duplicates Values

In [5]:
```
print(netflix.duplicated().value_counts())
netflix.drop_duplicates(inplace = True)
print(len(netflix))
```

```
False    8807
dtype: int64
8807
```

Checking Missing Values

```
In [6]: print('Data columns with null values:\n',
               netflix.isnull().sum())
```

```
Data columns with null values:
 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
```

- We can see that there are many null values present in feature: director, cast, country, rating, duration

Checking Unique Values

```
In [7]: netflix.nunique()
```

```
Out[7]: show_id         8807
type               2
title           8807
director        4528
cast            7692
country          748
date_added      1767
release_year      74
rating            17
duration         220
listed_in        514
description     8775
dtype: int64
```

- From the above output we can have an overall idea of how many unique values present in the dataset.

3. Data Filteration & Modification

Dropping unnessesary column(s)

- show_id feature is irrelevant to our analysis so let's remove this feature

#

```
In [9]:   netflix=netflix.drop(['show_id'],axis=1)
```

Handling Null Values

- As we have seen already that in some features the null values are very few so we can remove those rows directly, it will hadrly effect our overall analysis. Below are the features that we are removing:
- 1. date_added
- 2. rating
- 3. duration
- We have already seen that in another three columns, there is a huge chunk on missing values. By dropping those rows from tha dataset will negetively effect our further analysis, so instead of removing those null values we are going to replace them with the keyword "unknown". Below are the features that we are going to replace values:
- 1. director
- 2. cast
- 3. country

```
In [10]:  netflix.dropna(subset=['date_added'],how='any',inplace=True) # droping null value rows of "date_added" column
          netflix.dropna(subset=['rating'],how='any',inplace=True)     # droping null value rows of "rating" column
          netflix.dropna(subset=['duration'],how='any',inplace=True)   # droping null value rows of "duration" column


          netflix['director'].replace(np.nan,'unknown',inplace=True) # replacing NaN value with "unknown"
          netflix['cast'].replace(np.nan,'unknown',inplace=True)   #  replacing NaN  value  with  "unknown"
          netflix['country'].replace(np.nan,'unknown',inplace=True) # replacing NaN value with "unknown"
```

Now let's check again if any null values are present or not.

```
In [11]:  print('Data columns with null values:\n',
                netflix.isnull().sum())

          Data columns with null values:
           type           0
          title           0
          director        0
          cast            0
          country         0
          date_added      0
          release_year    0
          rating          0
          duration        0
          listed_in       0
          description     0
          dtype: int64
```

- So now the data is looking quite good, let's sharpen our data a little bit

Removing white space

```
In [12]: netflix_white_spacefree = netflix.apply(lambda x: x.str.strip() if x.dtype == "object" else x)
```

- After executing the above code we just cleared white spaces before and after of every object data present in the current dataset

Case correction

```
In [13]: netflix_lower = netflix_white_spacefree.apply(lambda x: x.astype(str).str.lower())
         netflix_lower.head(5)
```

Out[13]:

	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
0	movie	dick johnson is dead	kirsten johnson	unknown	united states	september 25, 2021	2020	pg-13	90 min	documentaries	as her father nears the end of his life, filmm...
1	tv show	blood & water	unknown	ama qamata, khosi ngema, gail mabalane, thaban...	south africa	september 24, 2021	2021	tv-ma	2 seasons	international tv shows, tv dramas, tv mysteries	after crossing paths at a party, a cape town t...
2	tv show	ganglands	julien leclercq	sami bouajila, tracy gotoas, samuel jouy, nabi...	unknown	september 24, 2021	2021	tv-ma	1 season	crime tv shows, international tv shows, tv act...	to protect his family from a powerful drug lor...
3	tv show	jailbirds new orleans	unknown	unknown	unknown	september 24, 2021	2021	tv-ma	1 season	docuseries, reality tv	feuds, flirtations and toilet talk go down amo...
4	tv show	kota factory	unknown	mayur more, jitendra kumar, ranjan raj, alam k...	india	september 24, 2021	2021	tv-ma	2 seasons	international tv shows, romantic tv shows, tv ...	in a city of coaching centers known to train i...

- Now all the string type datas present in the dataset are now in small letters.

Drop Duplicates

```
In [14]: netflix_lower.drop_duplicates(subset ="title",keep = False, inplace = True)
         netflix_purified = netflix_lower.copy()
```

- After executing the above code all duplicate values are removed

Feature Specific Data Modification: duration

- In one season there can be 2 episodes or it can be 12 episodes, in average we are taking 6 episodes in one season.
- NETFLIX data says each episodes have 30 minutes(approx) of content, so 6 x 30=180 minutes in average.
- Therefore, in duration column, if 2 or more seasons are there then it will be 180 x 2=360 or 180 x no.of seasons = 180n minutes, if 1 seasons is there then it will be 180 x 1=180 minutes, also if there is 19 seasons then 180 x 19=3420 munites.

```python
In [15]: netflix_purified['duration'] = netflix_purified['duration'].str.replace(' min','', regex=True).str.strip()
         netflix_purified['duration'] = netflix_purified['duration'].str.replace('1 season','130', regex=True).str.strip()
         netflix_purified['duration'] = netflix_purified['duration'].str.replace('2 seasons','360', regex=True).str.strip()
         netflix_purified['duration'] = netflix_purified['duration'].str.replace('3 seasons','540', regex=True).str.strip()
         netflix_purified['duration'] = netflix_purified['duration'].str.replace('4 seasons','420', regex=True).str.strip()
         netflix_purified['duration'] = netflix_purified['duration'].str.replace('5 seasons','900', regex=True).str.strip()
         netflix_purified['duration'] = netflix_purified['duration'].str.replace('6 seasons','1080', regex=True).str.strip()
         netflix_purified['duration'] = netflix_purified['duration'].str.replace('7 seasons','1260', regex=True).str.strip()
         netflix_purified['duration'] = netflix_purified['duration'].str.replace('8 seasons','1440', regex=True).str.strip()
         netflix_purified['duration'] = netflix_purified['duration'].str.replace('9 seasons','1620', regex=True).str.strip()
         netflix_purified['duration'] = netflix_purified['duration'].str.replace('10 seasons','1800', regex=True).str.strip()
         netflix_purified['duration'] = netflix_purified['duration'].str.replace('11 seasons','1980', regex=True).str.strip()
         netflix_purified['duration'] = netflix_purified['duration'].str.replace('12 seasons','2160', regex=True).str.strip()
         netflix_purified['duration'] = netflix_purified['duration'].str.replace('13 seasons','2340', regex=True).str.strip()
         netflix_purified['duration'] = netflix_purified['duration'].str.replace('14 seasons','2520', regex=True).str.strip()
         netflix_purified['duration'] = netflix_purified['duration'].str.replace('15 seasons','2700', regex=True).str.strip()
         netflix_purified['duration'] = netflix_purified['duration'].str.replace('16 seasons','2880', regex=True).str.strip()
         netflix_purified['duration'] = netflix_purified['duration'].str.replace('17 seasons','3060', regex=True).str.strip()
         netflix_purified['duration'] = netflix_purified['duration'].str.replace('18 seasons','3240', regex=True).str.strip()
         netflix_purified['duration'] = netflix_purified['duration'].str.replace('19 seasons','3420', regex=True).str.strip()
```

```python
In [16]: netflix_purified.head(5)
```

Out[16]:

	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
0	movie	dick johnson is dead	kirsten johnson	unknown	united states	september 25, 2021	2020	pg-13	90	documentaries	as her father nears the end of his life, filmm...
1	tv show	blood & water	unknown	ama qamata, khosi ngema, gail mabalane, thaban...	south africa	september 24, 2021	2021	tv-ma	360	international tv shows, tv dramas, tv mysteries	after crossing paths at a party, a cape town t...
2	tv show	ganglands	julien leclercq	sami bouajila, tracy gotoas, samuel jouy, nabi...	unknown	september 24, 2021	2021	tv-ma	130	crime tv shows, international tv shows, tv act...	to protect his family from a powerful drug lor...
3	tv show	jailbirds new orleans	unknown	unknown	unknown	september 24, 2021	2021	tv-ma	130	docuseries, reality tv	feuds, flirtations and toilet talk go down amo...
4	tv show	kota factory	unknown	mayur more, jitendra kumar, ranjan raj, alam k...	india	september 24, 2021	2021	tv-ma	360	international tv shows, romantic tv shows, tv ...	in a city of coaching centers known to train i...

- Here in the dataset we can see that all the unorganised features including duration are now in proper manner and also ready for our futher analysis

```python
In [17]: netflix_purified['release_year']=netflix_purified['release_year'].astype(int)
```

```
In [18]: print(netflix_purified.info())
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 8778 entries, 0 to 8806
Data columns (total 11 columns):
 #   Column        Non-Null Count  Dtype
---  ------------  --------------  --------
 1   type          8778 non-null   object
 2   title         8778 non-null   object
 3   director      8778 non-null   object
 4   cast          8778 non-null   object
 5   country       8778 non-null   object
 6   date_added    8778 non-null   object
 7   release_year  8778 non-null   int32
 8   rating        8778 non-null   object
 9   duration      8778 non-null   object
 10  listed_in     8778 non-null   object
 11  description   8778 non-null   object
dtypes: int32(1), object(10)
memory usage: 788.6+ KB
None
```

4. Exporting the Purified Data

```
In [19]: netflix_purified.to_csv('netfilx_cleaned.csv')
```

- After execution of the above codes now we just make a fresh copy of .csv file with purified data.

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

#

Visualization

- **Introduction**: Great and Warm welcome to everyone present here, I Dhiraj Kr. Choudhary going to continue with Visualization Part.
- **Tools Used**: I have used the Tableau 2022.3
- **Basic Steps**: First off all I received the Clean data after performing EDA from my Team mate then I analyse the NETFLIX Data and tried to figure out the meaningful insights & relationship that can be driven out from the data . Then I preprocess and transform the data according to required situation provided in the problem statement.
- **Observation**: It was found that 2 types of content were available namely MOVIE and TV SHOW.

Continued...

- **Need for KPIs** : I have created few KPIs which will helpful for observation of the insights . The KPIs are as follow:
  - Generic : It will Indicate the sub category of the Content.
  - Rating : It will indicate the content is available for whom.
  - Released year : It will indicate, when the content was released.
  - Duration: It will indicate the time of content in Minute.
  - Upload Year : It will indicate, when the content has been Uploaded at Netflix.
  - Description : It will indicate the description of the selected content.

  These were my KPIs and their role in this project.

Problem Statement

Now moving forward towards the problem statement.

There were 4 different real life scenario in the problem statement, which are given below:

- 1) Identify Which Content are available region or country wise?

Sol: To answer this part I have created a map based on the availability of the types of content. There are basically 2 types of Content available which we can Select from Content type in the Dashboard.

- 2) Identify similar Content on Netflix by matching text based features?

Sol: For this part I have created a Word Art based on the popularity of the content based on time as well as rating. You can visualized by adjusting Date Wise Performance and Rating.

Continued…

- **3) Network analysis of Actors or Directors and find interesting insights?**

Sol: For this part I have created a bubble chart based on Rating of the content along with that I have also provided the option to fetch the popular actors and directors based on time interval. If we hover over the bubble chart the name and the popularity of the actor/director will be Visible.

- **4)Does Netflix has focus on TV Shows than movies in recent year?**

Sol: To answer this part I have created an Area graph based on title and the date added. If we set the date then the trend between Movie Vs. Tv Show could be Visualize. From the graph it can be seen that Netflix has focused on Movies than TV Shows in recent years.

Click here : NETFLIX

NETFLIX DASHBOARD

Genere	Rating	Released Year	Duration	Uploaded Year	Description
dramas, romantic movies, sports movies	tv-14	2017	105	2017	after escaping the bus accident that killed his girlfriend, a high school student channels his grief into running, with the help of a new coach.

Movies Vs TV Show by Region

Type
- TV Show
- Movie

© 2023 Mapbox © OpenStreetMap

Type
(Multiple values)

Title
1 mile to you

Content Available

Movie	6,114
TV Show	2,658

Content popularity

docuseries — romantic tv shows — tv comedies — classic & cult tv — music & musicals — crime tv shows — tv action & adventure — documentaries — dramas — kids' tv — children & family movies — reality tv — international movies — classic movies — tv dramas — romantic movies — international tv shows — stand-up comedy — comedies — horror movies — action & adventure — anime series — sports movies — tv horror — spanish-language tv shows — anime features — thrillers — british tv shows — stand-up comedy & talk shows — movies — tv sci-fi & fantasy — sci-fi & fantasy — independent movies — lgbtq movies — cult movies tv shows

Directors & Actors

raúl campos,
jay karas
rajiv chilaka

Rating

- ☑ (All)
- ☑ Kid 7+
- ☑ nc-17
- ☑ nr
- ☑ pg
- ☑ pg-13
- ☑ r
- ☑ tv-14
- ☑ tv-g
- ☑ tv-ma
- ☑ tv-pg
- ☑ tv-y

TV Shows & Movie by Year

Count of netflix_cleaned1

2008: 2
2009: 1
2010: 13
2011: 3
2015: 174
2016: 349
2017: 410
2018: 835
2019: 1,420
2020: 1,284
2021: 991
589, 595, 504
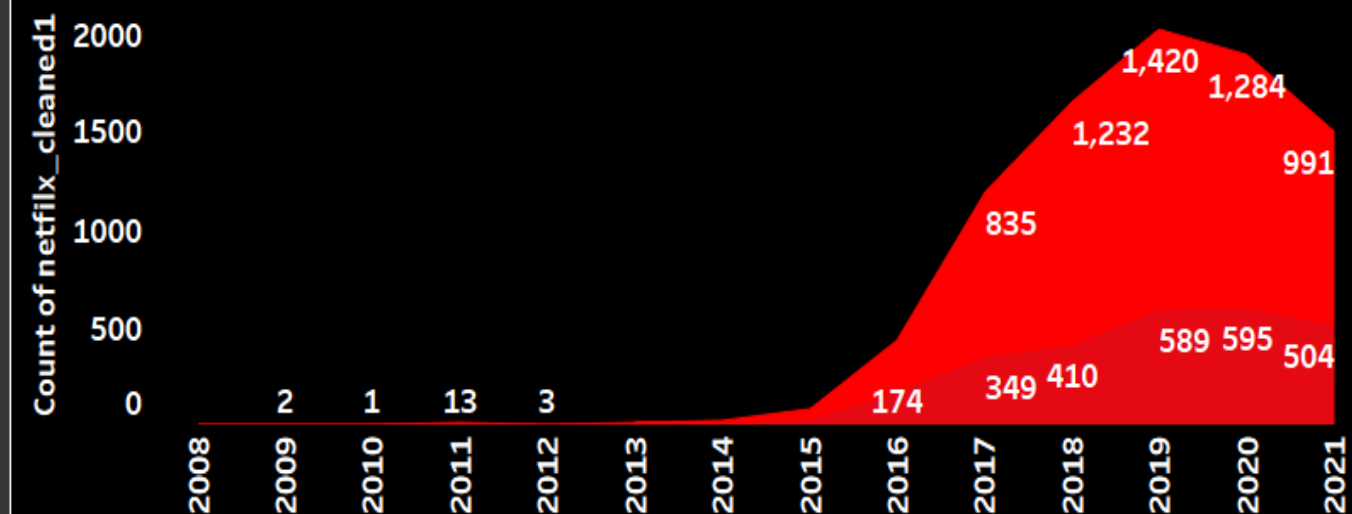
- If we try the different combination of filter we can draw few more interesting insights from the visuals.
- To understand the Insights in detailed and better manner, please hover the cursor at different points of the visuals.
- Link for the Dashboard : LINK

Thank You

In [ ]:

Sentiment Analysis on Netflix

Demonstrated by **BISWARUP DAS**

- **After the visualization performed by my colleague I think its become very important to perform the Sentiment Analysis on Netflix content. So, let's start...**

Slight modification in some datas, keeping the nan values

```
In [19]: netflix_purified['director'].replace('unknown',np.nan,inplace=True) # keeping the NaN values, replace later
         netflix_purified['cast'].replace('unknown',np.nan,inplace=True)  # keeping the NaN values, replace later
         netflix_purified['country'].replace('unknown',np.nan,inplace=True)  # keeping the NaN values, replace later
```

```
In [20]: print('Data columns with null values:\n',
               netflix_purified.isnull().sum())
```

```
Data columns with null values:
 type              0
title             0
director       2617
cast            825
country         829
date_added        0
release_year      0
rating            0
duration          0
listed_in         0
description       0
dtype: int64
```

```
In [21]: import numpy as np
         import pandas as pd
         import plotly.express as px # for data visualization
```

Content Distribution:

To begin the task of analyzing Netflix data, I'll start by looking at the distribution of content ratings on Netflix:

In [22]:
```python
content = netflix_purified.groupby(['rating']).size().reset_index(name='counts')
pieChart = px.pie(content, values='counts', names='rating',
title='Distribution of Content Ratings on Netflix')
pieChart.show()
```

Distribution of Content Ratings on Netflix
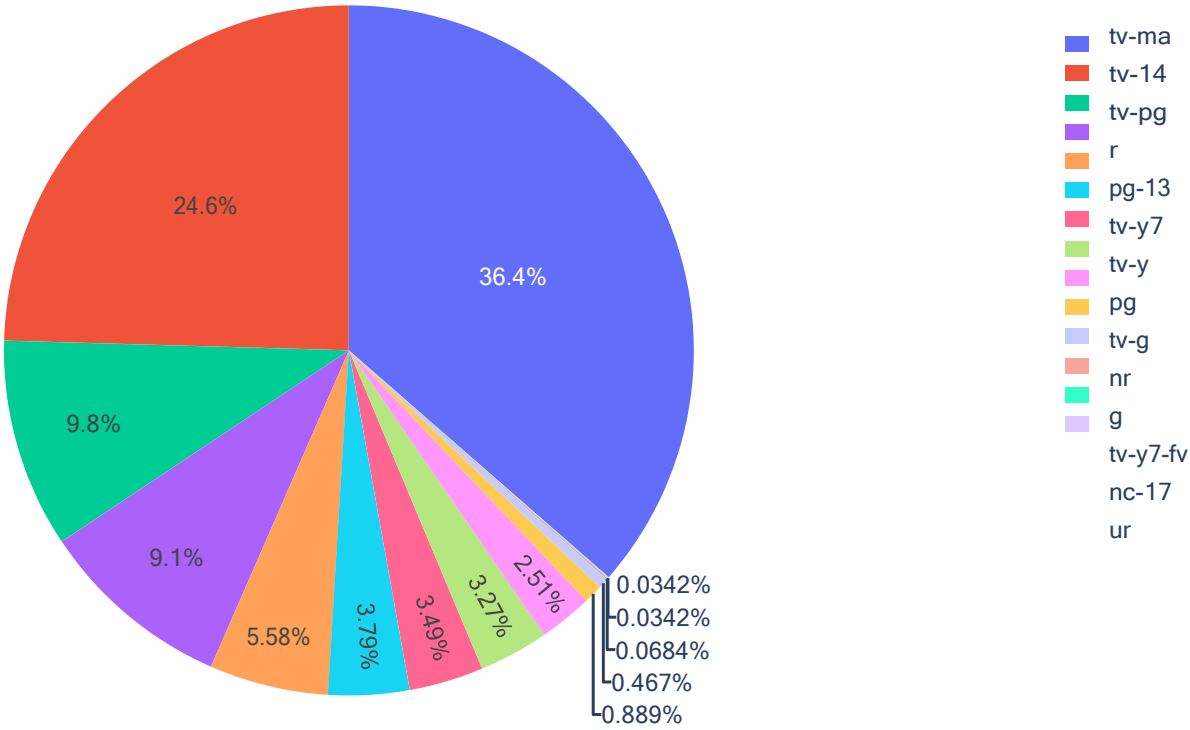

In [23]:
```python
print(content)
```

```
       rating  counts
0          g      41
1      nc-17       3
2         nr      78
3         pg     287
4      pg-13     490
5          r     799
6      tv-14    2155
7       tv-g     220
8      tv-ma    3197
9      tv-pg     860
10      tv-y     306
11     tv-y7     333
12  tv-y7-fv       6
13        ur       3
```

The graph above shows that the majority of content on Netflix is categorized as "TV-MA", which means that most of the content available on Netflix is intended for viewing by mature and adult audiences.

Top 5 Countries:

Now let's see the top 5 countries using Netflix:

```
In [24]:  filtered_c=pd.DataFrame()
          filtered_c=netflix_purified['country'].str.split(',',expand=True).stack()
          filtered_c=filtered_c.to_frame()
          filtered_c.columns=['Country']
          c=filtered_c.groupby(['Country']).size().reset_index(name='Total Content')
          #c=c[c.Country !='No Country Specified']
          c=c.sort_values(by=['Total Content'],ascending=False)
          cTop5=c.head()
          cTop5=cTop5.sort_values(by=['Total Content'])
          fig5=px.bar(cTop5,x='Total Content',y='Country',title='Top 5 Countries on Netflix')
          fig5.show()
```

Top 5 Countries on Netflix
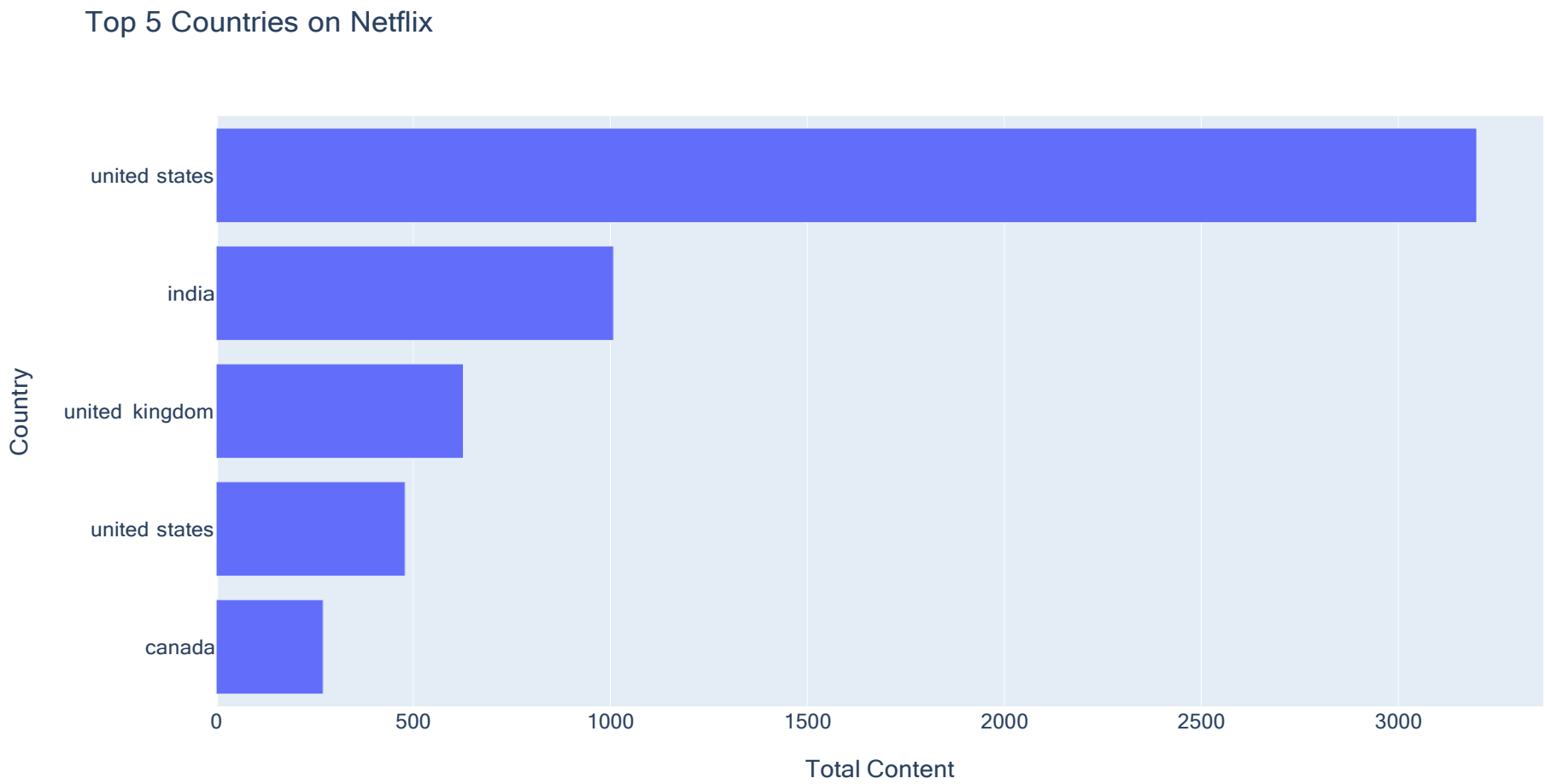
From the above graph it is derived that the top 5 countries on this platform are:

- **USA**

- **India**
- **UK**
- **USA**
- **Canada**

Top 5 Actors and Directors:

Now let's see the top 5 successful directors on this platform:

```
In [25]: netflix_purified['director']=netflix_purified['director'].fillna('No Director Specified')
         filtered_directors=pd.DataFrame()
         filtered_directors=netflix_purified['director'].str.split(',',expand=True).stack()
         filtered_directors=filtered_directors.to_frame()
         filtered_directors.columns=['Director']
         directors=filtered_directors.groupby(['Director']).size().reset_index(name='Total Content')
         directors=directors[directors.Director !='No Director Specified']
         directors=directors.sort_values(by=['Total Content'],ascending=False)
         directorsTop5=directors.head()
         directorsTop5=directorsTop5.sort_values(by=['Total Content'])
         fig1=px.bar(directorsTop5,x='Total Content',y='Director',title='Top 5 Directors on Netflix')
         fig1.show()
```

Top 5 Directors on Netflix
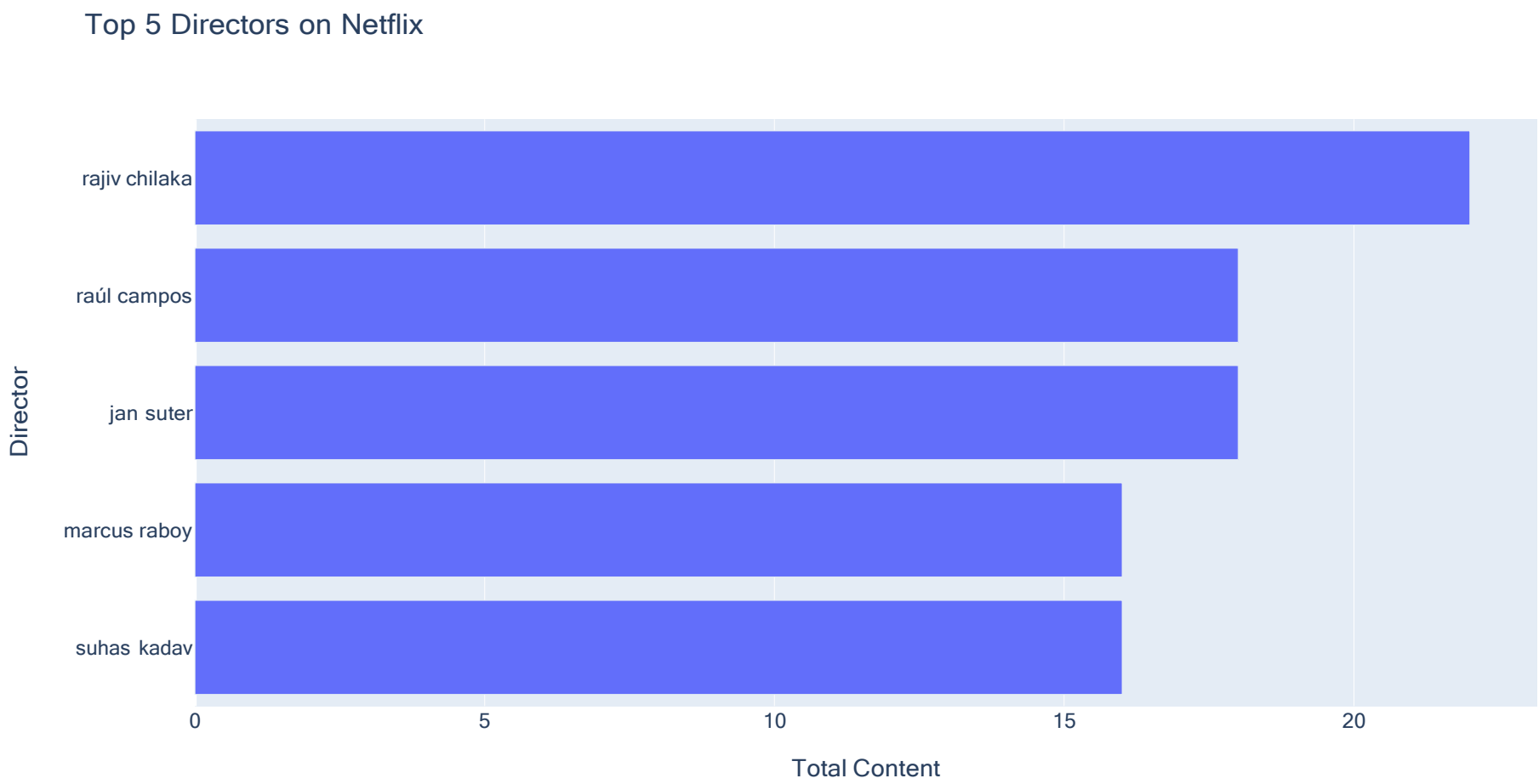
From the above graph it is derived that the top 5 directors on this platform are:

- **Rajiv Chilaka**
- **Raul Campos**
- **Jan Suter**
- **Marcus Raboy**
- **Suhas Kadav**

Now let's see the top 5 successful Actors on this platform:

In [26]:
```python
netflix_purified['cast']=netflix_purified['cast'].fillna('No Cast Specified')
filtered_cast=pd.DataFrame()
filtered_cast=netflix_purified['cast'].str.split(',',expand=True).stack()
filtered_cast=filtered_cast.to_frame()
filtered_cast.columns=['Actor']
actors=filtered_cast.groupby(['Actor']).size().reset_index(name='Total Content')
actors=actors[actors.Actor !='No Cast Specified']
actors=actors.sort_values(by=['Total Content'],ascending=False)
actorsTop5=actors.head()
actorsTop5=actorsTop5.sort_values(by=['Total Content'])
fig2=px.bar(actorsTop5,x='Total Content',y='Actor',title='Top 5 Actors on Netflix')
fig2.show()
```

Top 5 Actors on Netflix
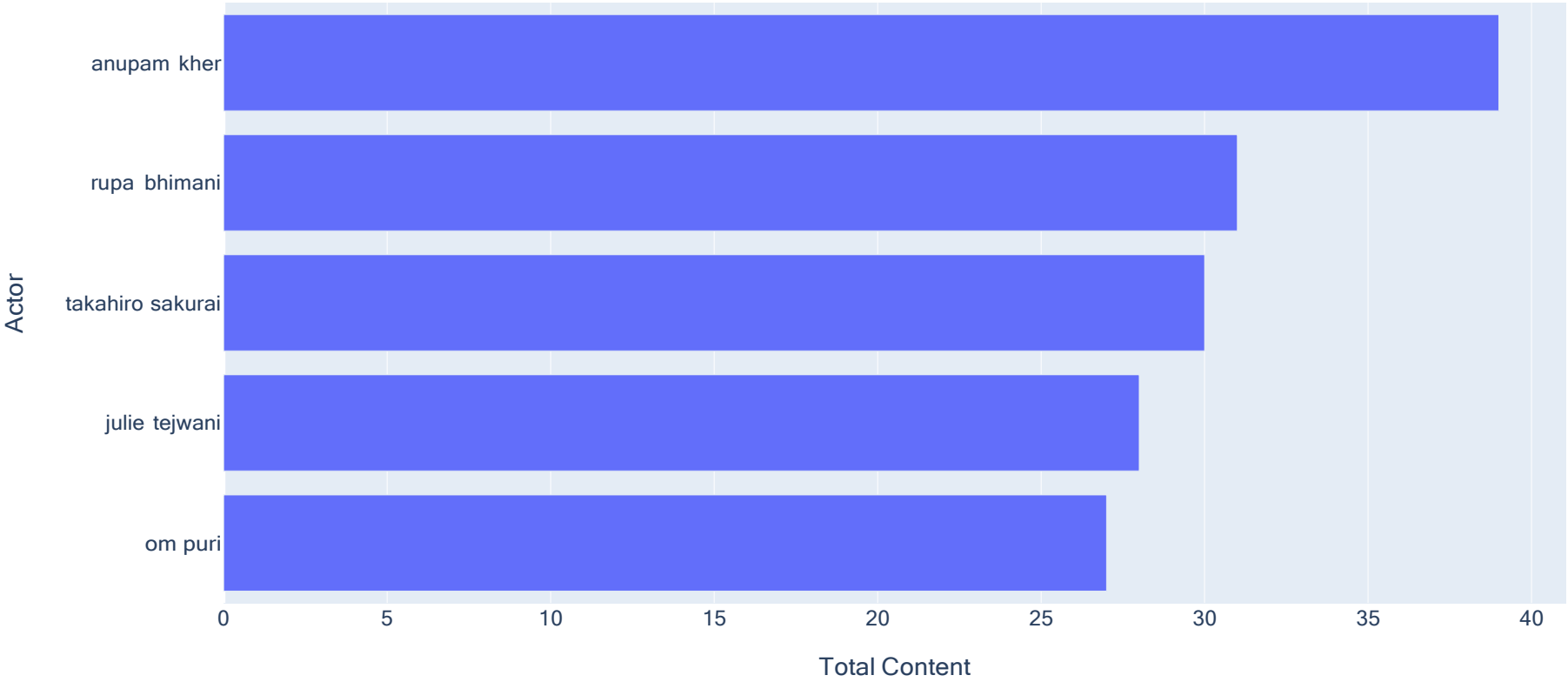
From the above graph it is derived that the top 5 actors on this platform are:

- **Anupam Kher**
- **Rupa Bhimani**
- **Takahiro Sakurai**
- **Julie tejwani**
- **Om Puri**

Analyzing Content on Netflix:

The next thing to analyze from this data is the trend of production over the years on Netflix:

In [32]:
```python
df1=netflix_purified[['type','release_year']]
df1=df1.rename(columns={"release_year": "Release Year"})
df2=df1.groupby(['Release Year','type']).size().reset_index(name='Total Content')
df2=df2[df2['Release Year']>=2010]
print(df2)
fig3=px.bar(df2,x='Release Year',y='Total Content',title='Content Trend on Netflix')
fig3.show()
```

```
     Release Year      type  Total Content
95           2010     movie            151
96           2010   tv show             39
97           2011     movie            145
98           2011   tv show             40
99           2012     movie            173
100          2012   tv show             63
101          2013     movie            225
102          2013   tv show             61
103          2014     movie            262
104          2014   tv show             88
105          2015     movie            396
106          2015   tv show            159
107          2016     movie            658
108          2016   tv show            243
109          2017     movie            763
110          2017   tv show            265
111          2018     movie            767
112          2018   tv show            377
113          2019     movie            633
114          2019   tv show            397
115          2020     movie            517
116          2020   tv show            436
117          2021     movie            277
118          2021   tv show            315
```

Content Trend on Netflix
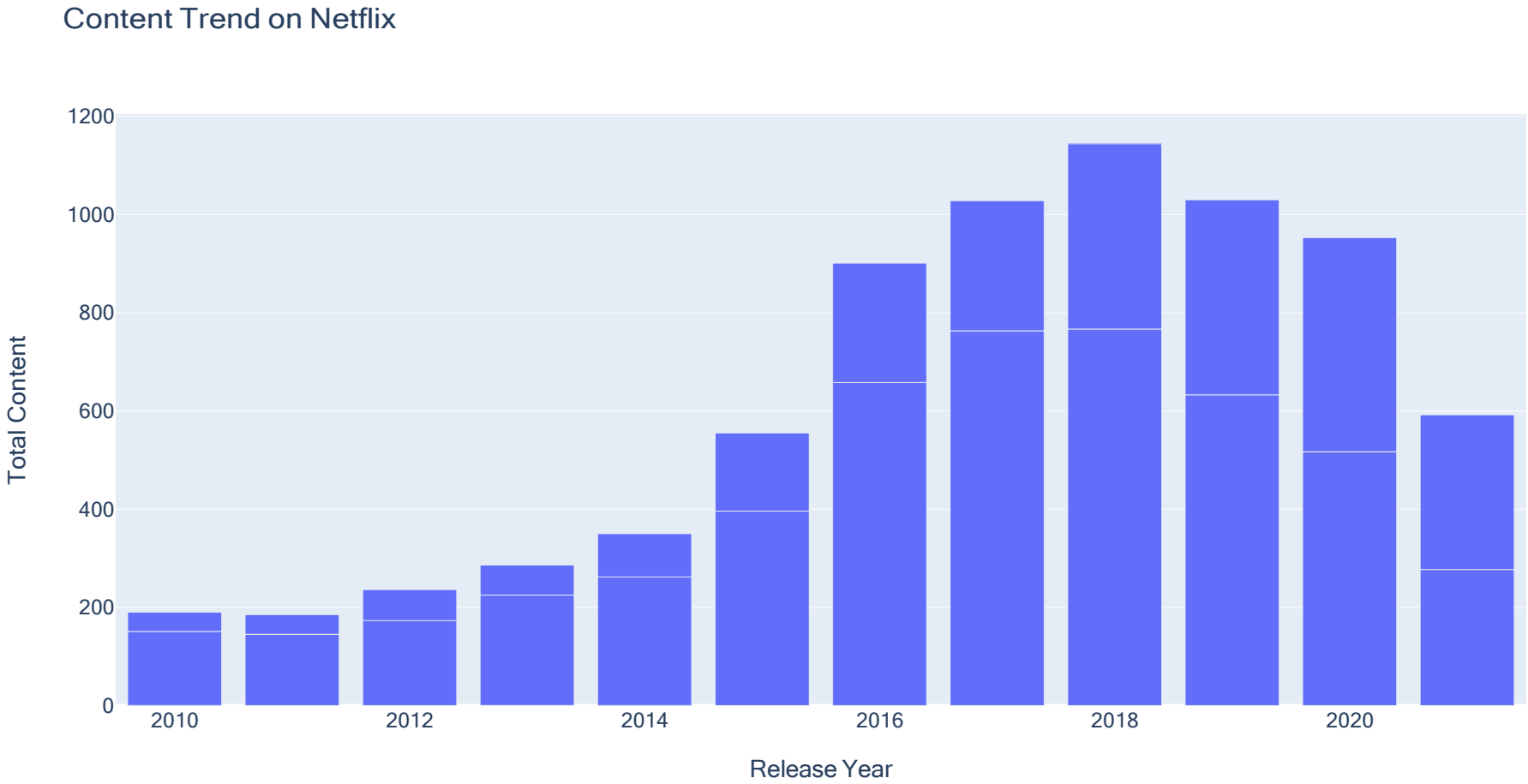
The above line graph shows that from 2011 content addition on Netflix started growing and touches its peak on year 2018 and after that there is a huge fall, it shows that Netfix slow down their process of content addition.

Key Focus of Netflix

In [28]:
```python
df1=netflix_purified[['type','release_year']]
df1=df1.rename(columns={"release_year": "Release Year"})
df2=df1.groupby(['Release Year','type']).size().reset_index(name='Total Content')
df2=df2[df2['Release Year']>=2010]
fig4 = px.line(df2, x="Release Year", y="Total Content", color='type',title='Trend of content produced over the years on Netflix')
fig4.show()
```

Trend of content produced over the years on Netflix
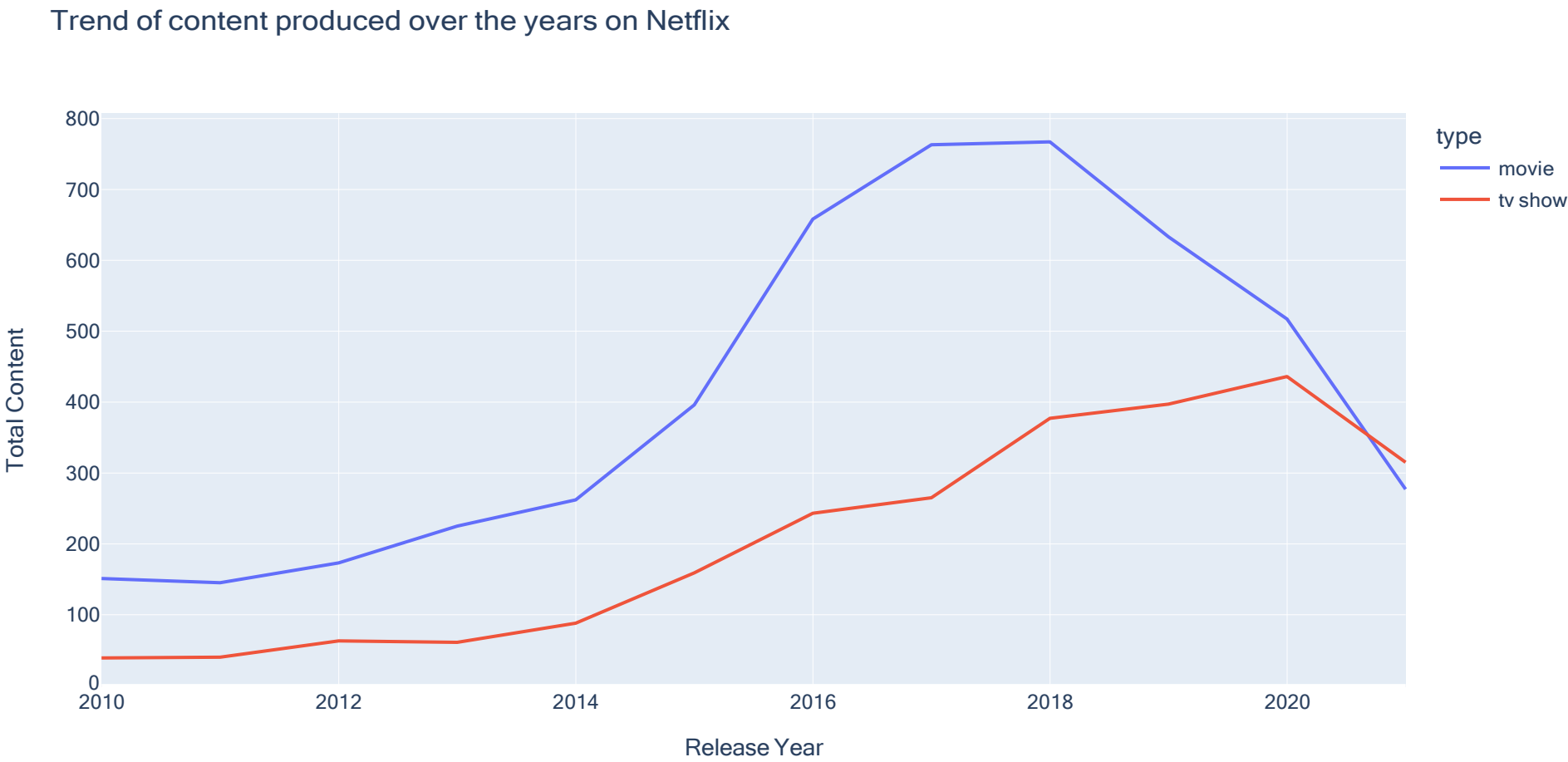
The above line graph shows that there has been a decline in the production of the content for movies since 2018 but for TV shows it gradually increases till 2020 and then there is a sharp decline after 2020.It shows Netfilx has more focus on TV shows. At last, to conclude our analysis, I will analyze the sentiment of content on Netflix:

About TextBlob?

- **TextBlob is a python library and offers a simple API to access its methods and perform basic NLP tasks.**
- **A good thing about TextBlob is that they are just like python strings. So, you can transform and play with it same like we did in python. Below, I have shown you below some basic tasks. Don't worry about the syntax, it is just to give you an intuition about how much-related TextBlob is to Python strings.**

In [29]: `! pip install textblob # This command may not work properly on iOS`

ERROR: Invalid requirement: '#'

In [30]: `from textblob import TextBlob # for sentiment analysis, This command may not work properly on iOS`

Sentiment Analysis

- **Sentiment analysis is basically the process of determining the attitude or the emotion of the writer, i.e., whether it is positive or negative or neutral.**
- **The sentiment function of textblob returns two properties, polarity, and subjectivity.**
- **Polarity is float which lies in the range of [-1,1] where 1 means positive statement and -1 means a negative statement. Subjective sentences generally refer to personal opinion, emotion or judgment whereas objective refers to factual information. Subjectivity is also a float which lies in the range of [0,1].**
- **Let's check the sentiment of our blob.**

In [31]:
```python
# This set of code may not work properly on iOS
dfx=netflix_purified[['release_year','description']]
dfx=dfx.rename(columns={'release_year':'Release Year'})
for index,row in dfx.iterrows():
    z=row['description']
    testimonial=TextBlob(z)
    p=testimonial.sentiment.polarity

    if p==0:
        sent='Neutral'
    elif p>0:
        sent='Positive'
    else:
        sent='Negative'
    dfx.loc[[index],'Sentiment']=sent

dfx=dfx.groupby(['Release Year','Sentiment']).size().reset_index(name='Total Content')
dfx=dfx[dfx['Release Year']>=2010]
fig4 = px.bar(dfx, x="Release Year", y="Total Content", color="Sentiment", title="Sentiment of content on Netflix")
fig4.show()
```
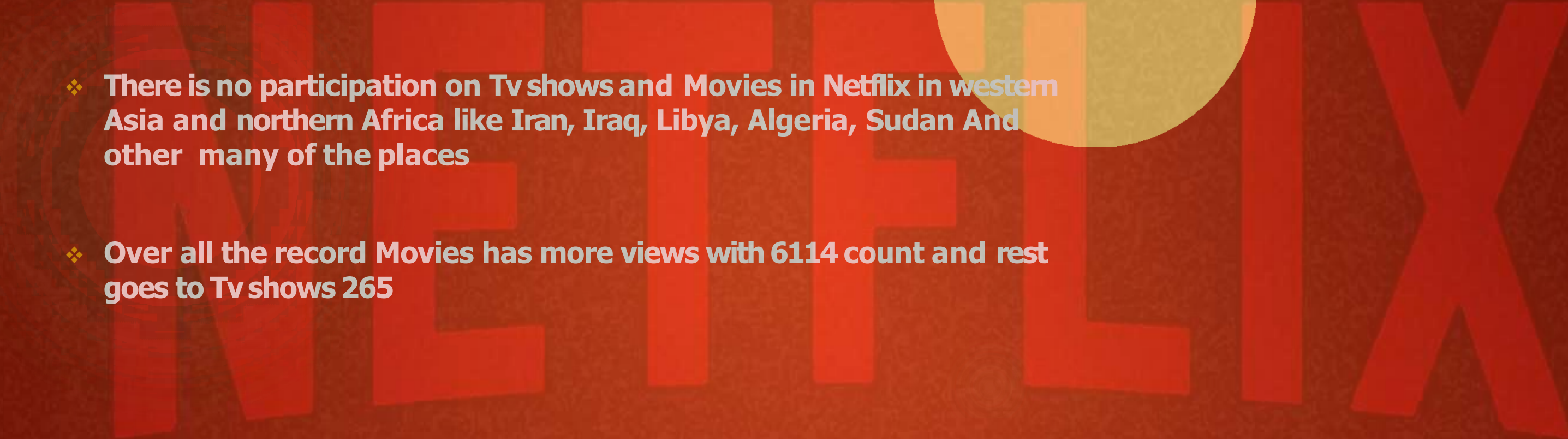
Sentiment of content on Netflix
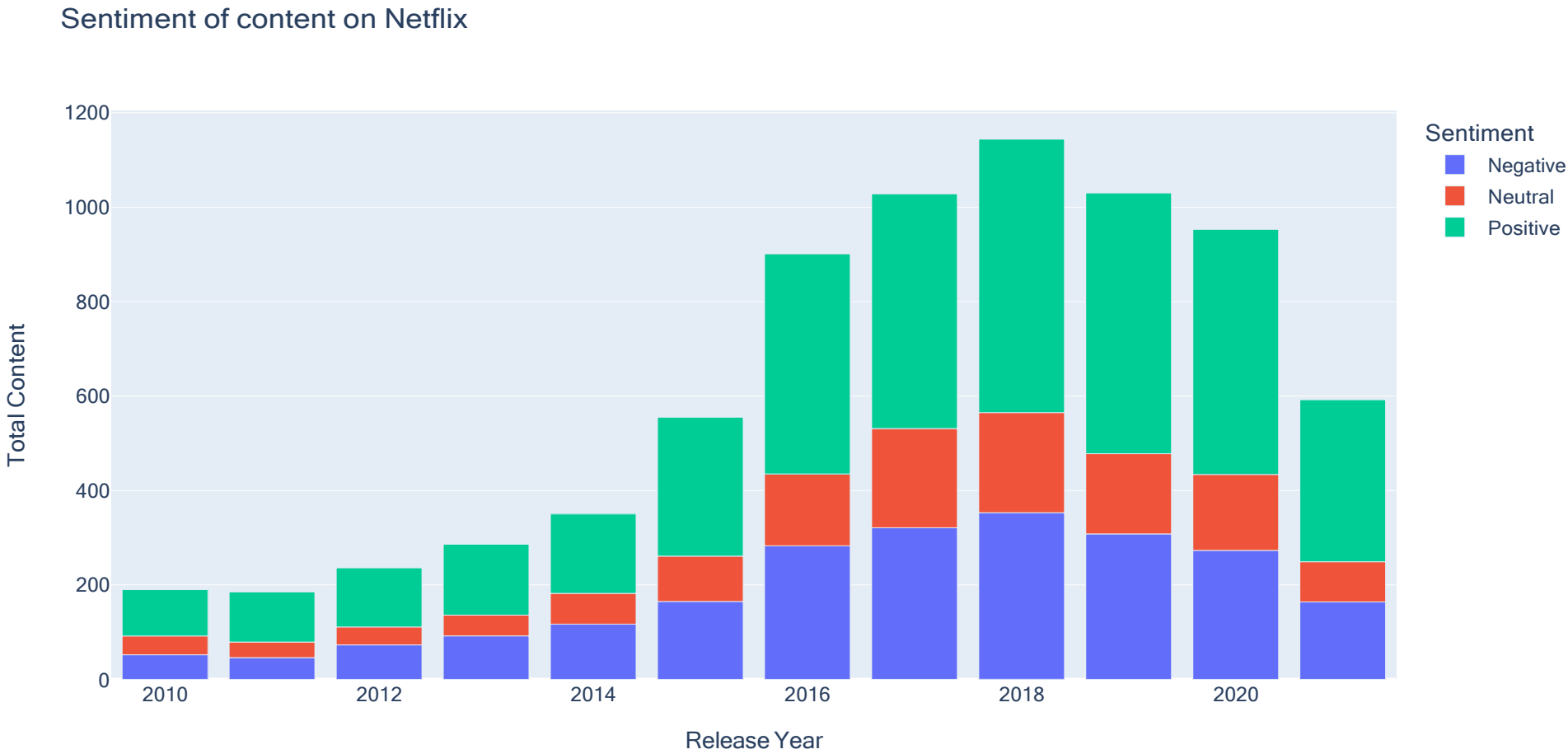
So the above graph shows that the overall positive content is always greater than the neutral and negative content combined.

Observation

- ❖ **In this data we observe is all about the Netflix users**

- ❖ **The Data consist of show id, type, title, director, cast, country, date added, release year, rating, duration, listed in, description**

- ❖ **It has 8808 records**

- ❖ **There are two categories one is Tv shows and the other one is Movies**

- In that movies views are most in 2018, 2019, 2020

- Overall comparing of movies and tv shows, Movies have the most number of views

- In 2019 2020 Most views of Tv shows are there

- There is no participation on Tv shows and Movies in Netflix in western Asia and northern Africa like Iran, Iraq, Libya, Algeria, Sudan And other many of the places

- Over all the record Movies has more views with 6114 count and rest goes to Tv shows 265

- ❖ **Out of cent percentage movies having 69.70% and Tv shows having 30.30%**

- ❖ **Based on the above sentiment Analysis 2018 has high positive response.**

- ❖ **Tv-ma have highest count as comparing with others.**

- ❖ **Top Five countries watching Netflix are United States, India, United Kingdom, and Canada.**

- ❖ **Based on the dataset the top five actors are Anupam Kher, Bhimani Takahiro Sakurai, Om puri, Julie tejwani.**

Conclusion

- Netflix includes a wide variety of familiar network shows and more original series, films, documentaries and special than any of its myriad competitors. We have analyzed the given questions following and filtering the data and data cleaned next we have to original dataset to be create the visualization for different countries. We have to understanding the visualization will be move to contents by matching text-based features. We have calculate for following sets in Movie &Tv Shows by region and year wise, Infant Mortality rate, Observing Directors and Actors.

- We Analyze the visualization for Movies and TV shows the segregate the Region Based vision.

- We did visualise the TV shows& movie by year for take 2009 to 2012 is viewers is not much more than, we looking a customers for high in 2013 to 2019 years day by day increase the customers so widely increase within 1,232 and 1,420 is highest customers for a year. but web has helped Netflix expand this network to an even larger audience.

Shows Contents

- Netflix is a type of business that has the right ideas, allowing their customers to stay inside and being able to watch any movie or TV shows they want.Many more shows will be releasing the children & family movies, International TV shows, stand-up comedies, action & adventures, classic movies thrillers, TV action& adventures.The threat of new entrants that is very small but many companies have tried to enter the market, but Netflix entry barriers are very large which makes it harder for other companies to complete with Netflix.

Case Study demonstrated by

Biswarup Das

D. Kiran Kumar
Deepak kumar
Dhiraj Kumar
Eshwar S
Jokin prabhu
Jyothiswar Sai
Lakshmikanth

Thank you