***Analyzing basic metrics* :**
we have been provided the Netflix dataset with columns like show id,title,cast,director,release year etc. we have to analyze the data based on the given data and produce some basic insights on the basis of release year, no. of movies or tv shows produced in a single year, most famous director,most famous genre and many more.

Basic Observation on the dataset:

- ***shape of dataset****:* The original dataset contains 8807 rows and 12 columns which may be changed after we perform the data cleaning.
- ***datatypes of columns:*** All the columns have datatype of object except the release_year column which has integer datatype. but this has to be changed as we can see that the release_year column is in integer but it has to be datetime format. and like this many more column's datatype needs to be changed.

Missing Values:There are some columns which has some missing values like director,cast,country,date_added etc.

```
df.isna().sum()

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
```

This image shows the null values in the original data frame of netflix.

Non Graphical Analysis:(After Data Cleaning and filling missing values)

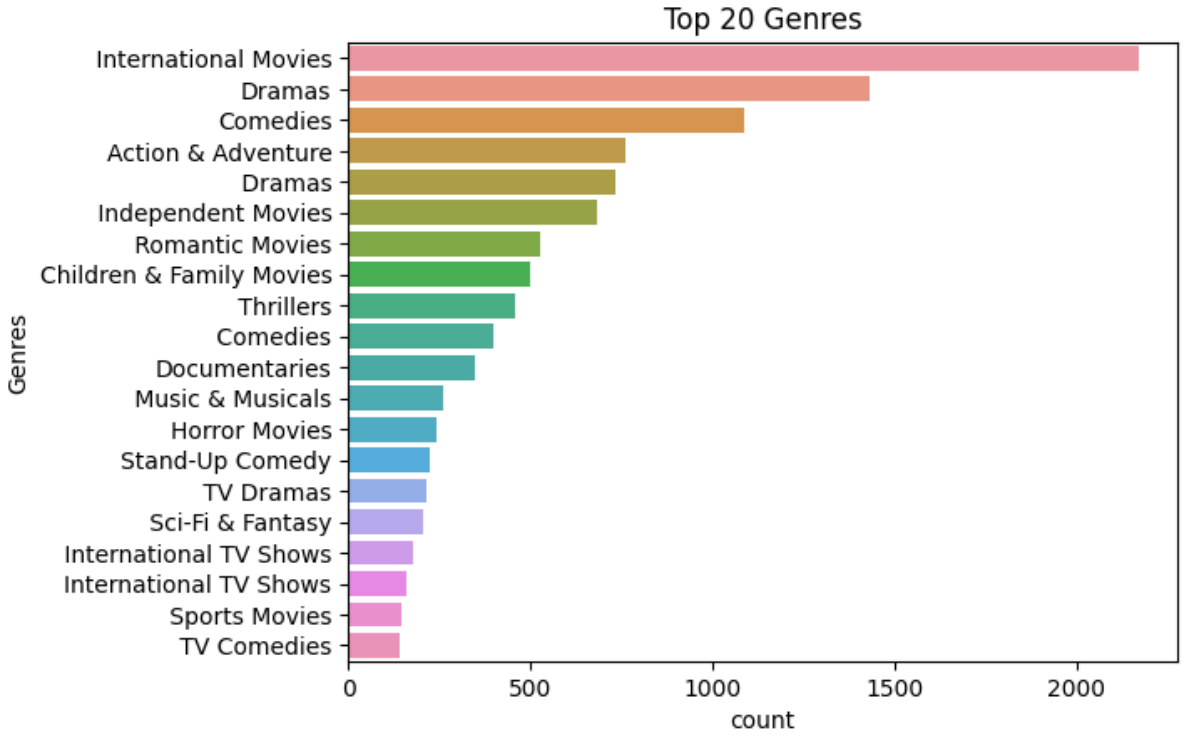
1. There are a total of 807 unique movies and tv shows, In which there are 6131 Movies and 2676 Shows, which clearly depicts that Netflix has focused more on Movies than Shows.
2. ANUPAM KHER is the most frequent actor who has done 39 movies.

3. RAJIV CHILAKA is the director who has done maximum movies which is 23.
4. United States(2296) has produced most of the movies on netflix followed by India(928) and United Kingdom(398).
5. Most famous Director and Cast pair who has done 23 movies together is RAJIV CHILAKA and JULIE TEJWANI.

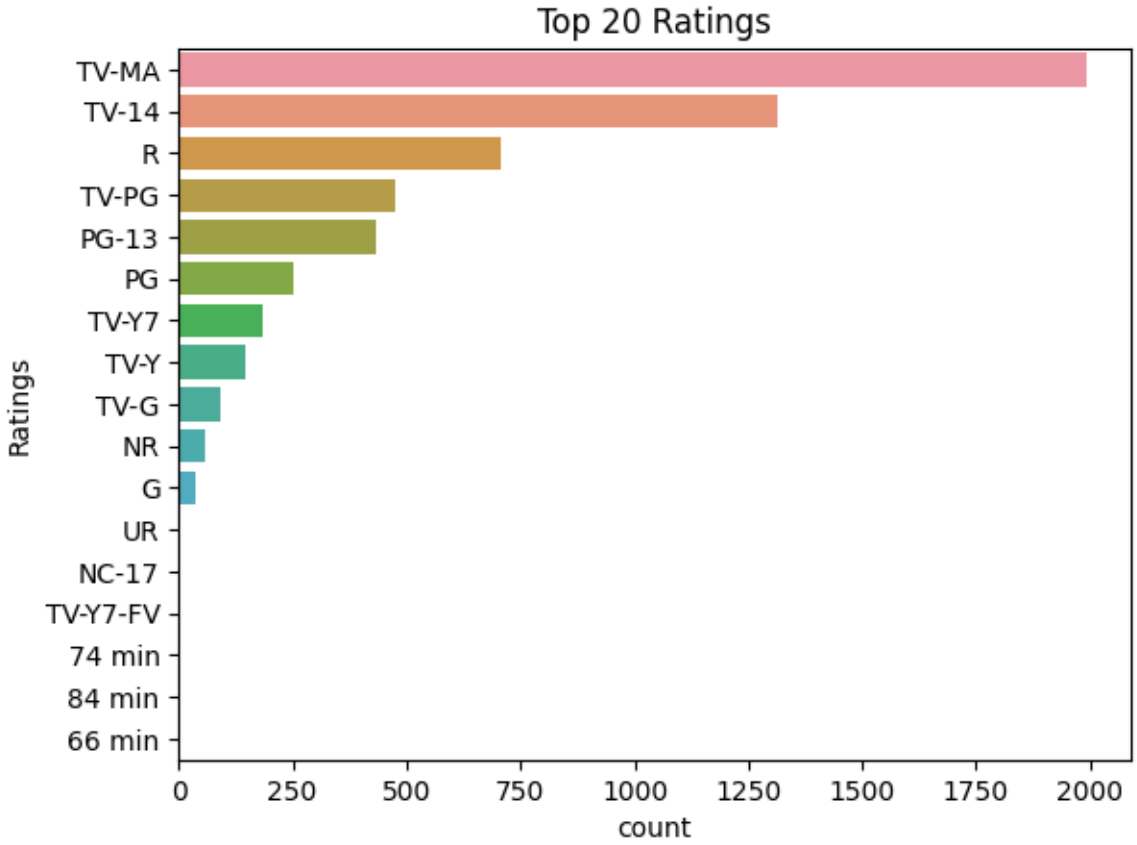
Movies for which release year and year_added are same are most probable that they are produced originally my Netflix, which is 1862 number of movies.

Visual Analysis:

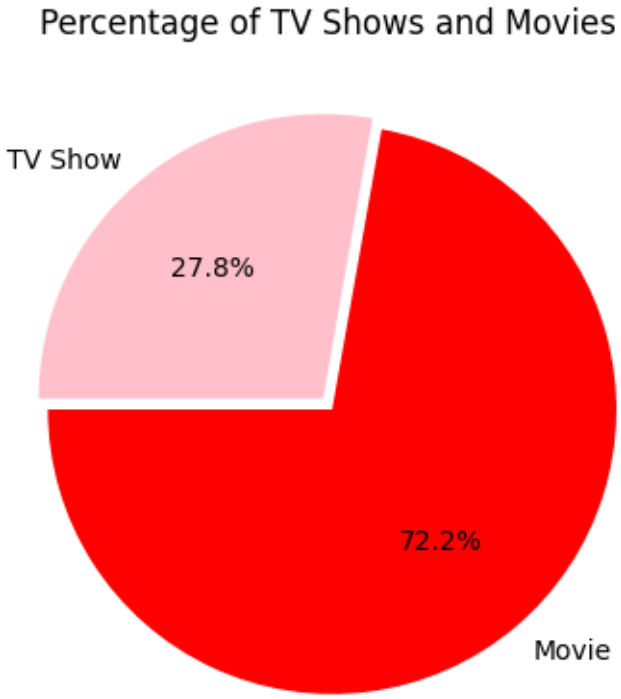
1. **Top 20 genres:In** all the genres we have the most famous genre is the 'International Movies', followed by 'Dramas' and 'Comedies'.



2. Top 20 Ratings: In all the ratings we have most of the movies have the ratings 'TV-MA',followed by 'TV-14'and'R'. .

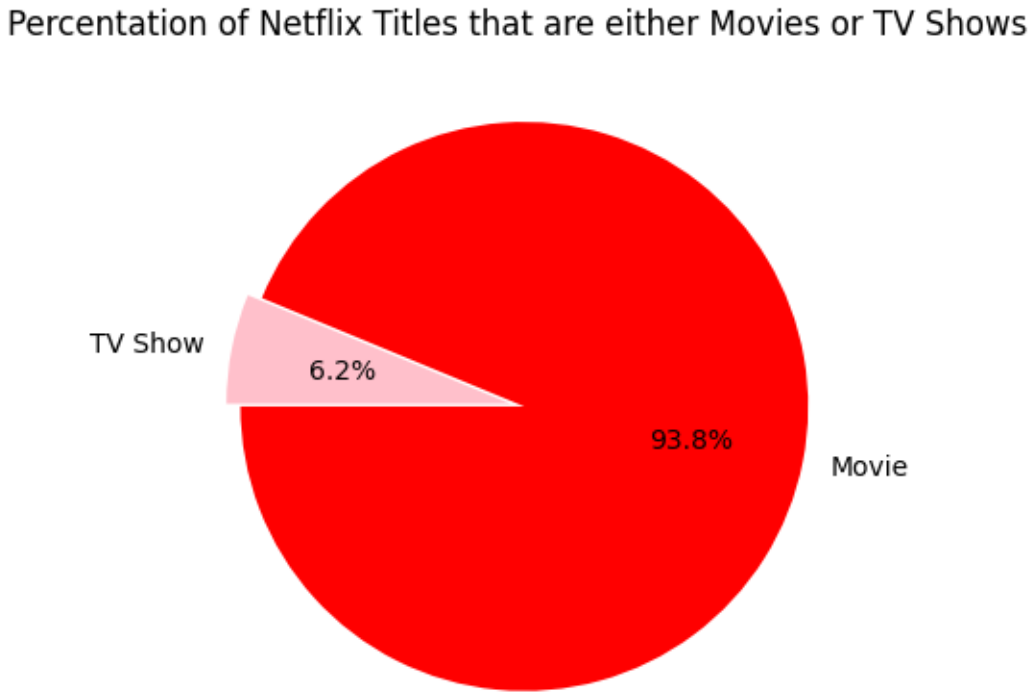


3. Movies and TV shows Internationally: Netflix has focused more on movies

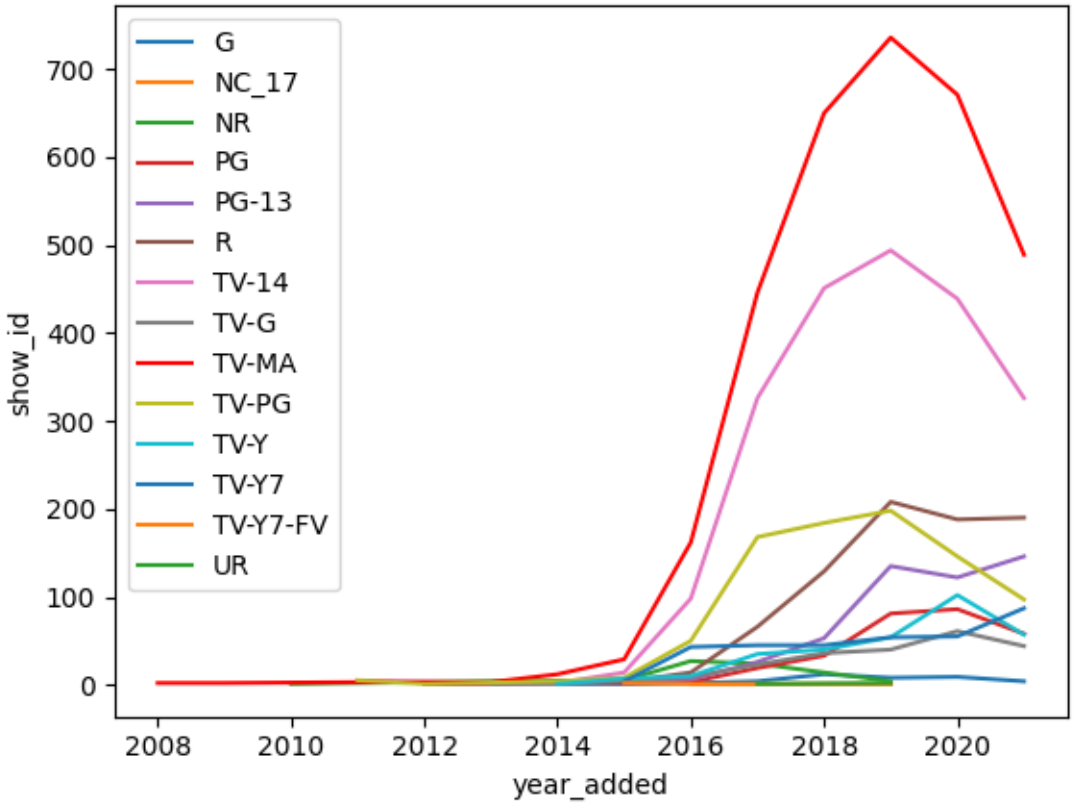


than Tv shows.

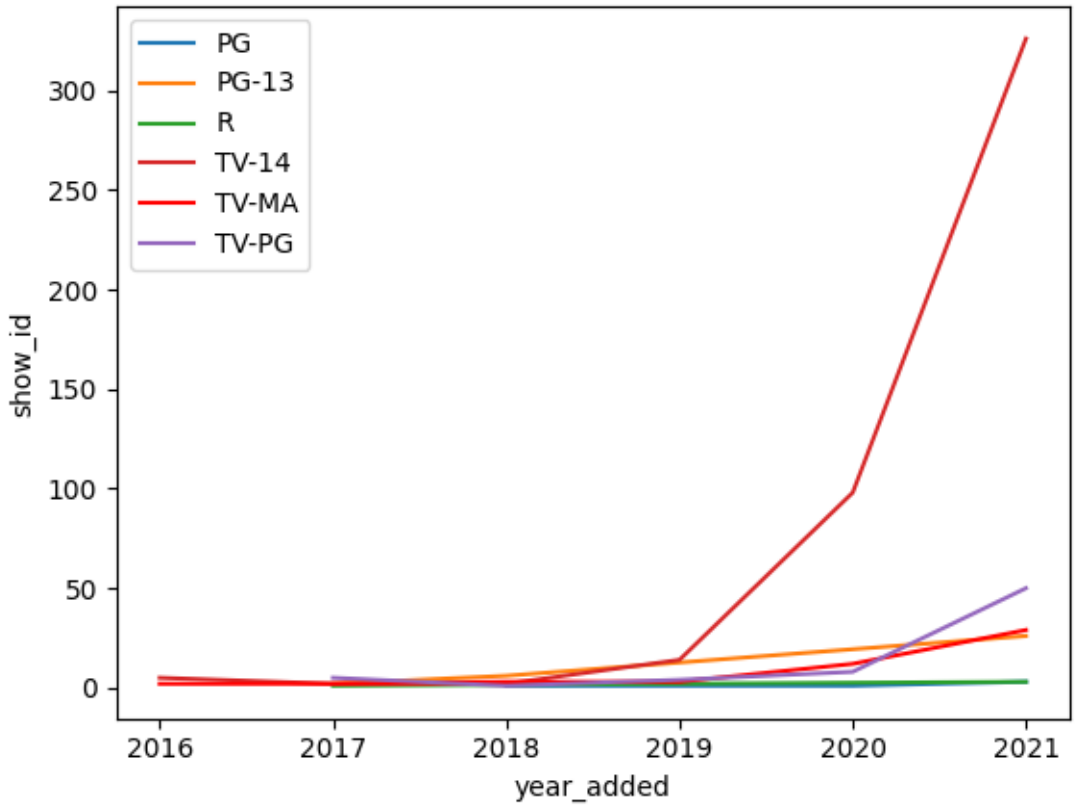
4. Movies and Tv shows in India: The percentage of movies and tv shows in india have much shocking percentage distribution.



5. Change of Ratings over the year: we can see that there is increasing popularity of TV-MA rating content over the years.

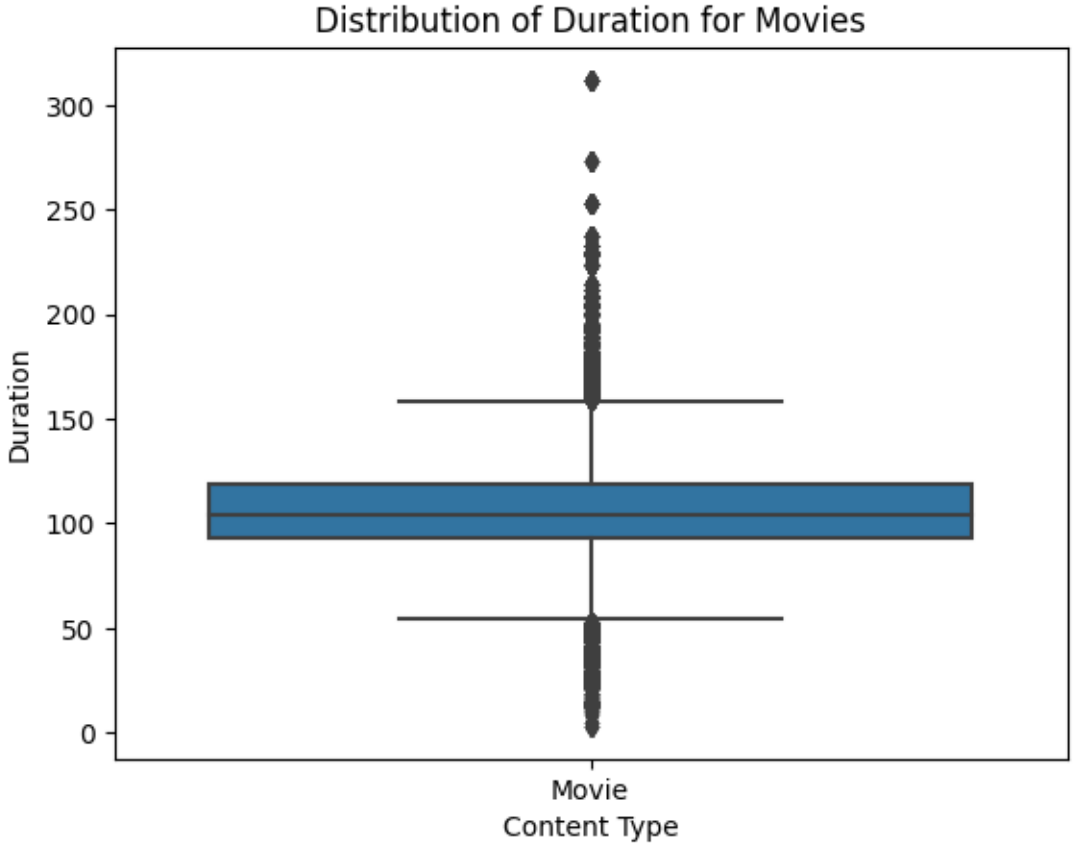


6. Ratings over the year in India:India also shows the same trend for the ratings.

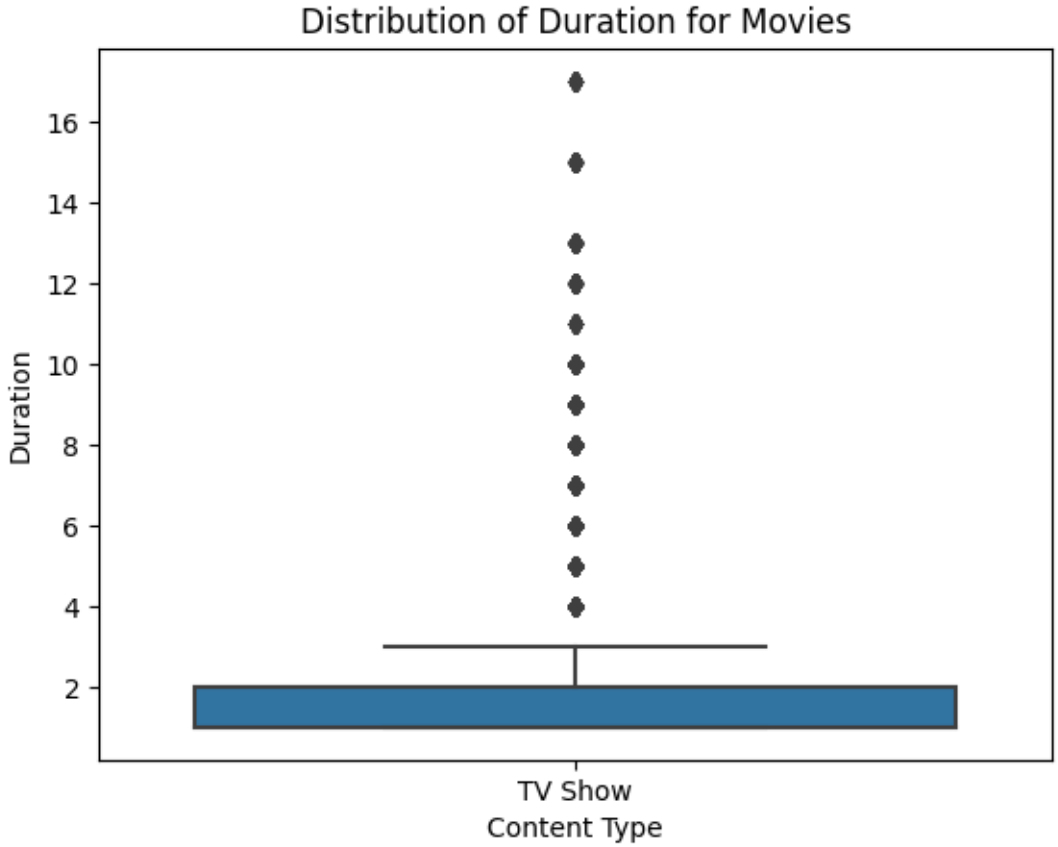


7. Movies Duration: we can see that most movies fall within a reasonable duration range, with few outliers exceedingly approximately 2.5 hours. This suggests that most movies on Netflix are designed to fit within a standard

viewing time.



TV Show Duration: For TV shows, the box plot reveals that most shows have one to four seasons, with very few outliers having longer durations.



Missing Values and Outlier:

In a dataset, we often see the presence of empty cells, rows, and columns, also referred to as Missing values. They make the dataset inconsistent and unable to work on. Many machine learning algorithms return an error if parsed with a dataset containing null values. Detecting and treating missing values is essential while analyzing and formulating data for any purpose.

how to fill missing values: we will fill the missing values by taking a calculated guess.

for missing values in director column we will find the mode of director for each cast and after that for the missing values we see the cast and fill the missing value with the mode of diretor for that cast.

BUSINESS INSIGHTS:

1. Quantity: Our analysis revealed that Netflix had added more movies than TV shows, aligning with the expectation that movies dominate their content library.
2. Content Addition: July emerged as the month when Netflix adds the most content, closely followed by December, indicating a strategic approach to content release.

3. Genre Correlation: Strong positive associations were observed between various genres, such as TV dramas and international TV shows, romantic and international TV shows, and independent movies and dramas. These correlations provide insights
into viewer preferences and content interconnections.
4. Movie Lengths: The analysis of movie durations indicated a stabilization around 100 minutes, highlighting a trend in movie lengths over time.
5. TV Show Episodes: Most TV shows on Netflix have one season, suggesting a preference for shorter series among viewers.
7. Rating Distribution: The distribution of ratings over the years offers insights into the
evolving content landscape and audience reception.
8. Data-Driven Insights: Our data analysis journey showcased the power of data in unravelling the mysteries of Netflixâ€™s content landscape, providing valuable insights
for viewers and content creators.
9. Continued Relevance: As the streaming industry evolves, understanding these patterns and trends becomes increasingly essential for navigating the dynamic landscape of Netflix and its vast library.

RECOMMENDATIONS:

1. Netflix has to focus on TV Shows also because there are people who will like to see
tv shows rather than movies
2. By approaching the top director we can plan some more movies/tv shows in order to
increase the popularity
3. We have seen most no of international movies genre so need to give priority to other
genres like horror,comedy..etc.
4.Netflix can focus more on R rated or PG rated content as they are quite popular among audience.
5. Most of the movies released in ott is in a year 2019 so we need to go on increasing
this value in order to attract people by showing that getting subscription
is useful as netflix is releasing more movies per year.
6.Focus more on Local contents for the countries as it will attract the viewership.
7. Some movies can be released directly into ott which has some positive talk which
may help in improving subscriptions
8.Should focus on a actor who has immense following and make use of it by doing a TV Shows or web series

9.Advertisement in the country which has very less movies released should be increased and attract people of that country by making their native TV Shows