**Netflix Case Study**

**Exploratory data analysis (EDA)** is an approach to:

-analyzing and understanding data that is focused on finding patterns and relationships in the data, rather than on testing hypotheses or making predictions

-visualizing the data and looking for trends, patterns, and anomalies, as well as summarizing the main characteristics of the data.

**EDA** is an important step in the data science process because it helps you get to know your data, identify any problems or issues with the data, and formulate hypotheses for further analysis.

**Business Problem:**Analyze the data and generate insights that could help Netflix in deciding which type of shows/movies to produce and how they can grow the business in different countries.

**Approach:** We are doing EDA on Netflix data, employing various Python libraries (NumPy, Pandas, Matplotlib, Seaborn), with a primary objective to unearth valuable insights and offer actionable recommendations.

**Collab link:**https://colab.research.google.com/drive/1D83zleayz\_SLtDuL7nV1xcwXDttNF7L\_?usp=sharing

**Exploring the data**

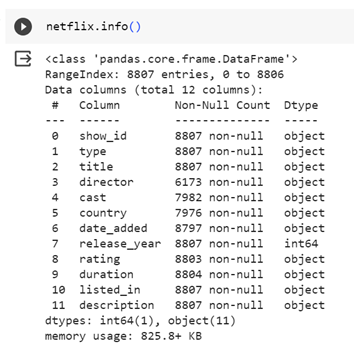
-netflix.head()

We can check the different attributes present in our dataset, and the type of the data for each column.



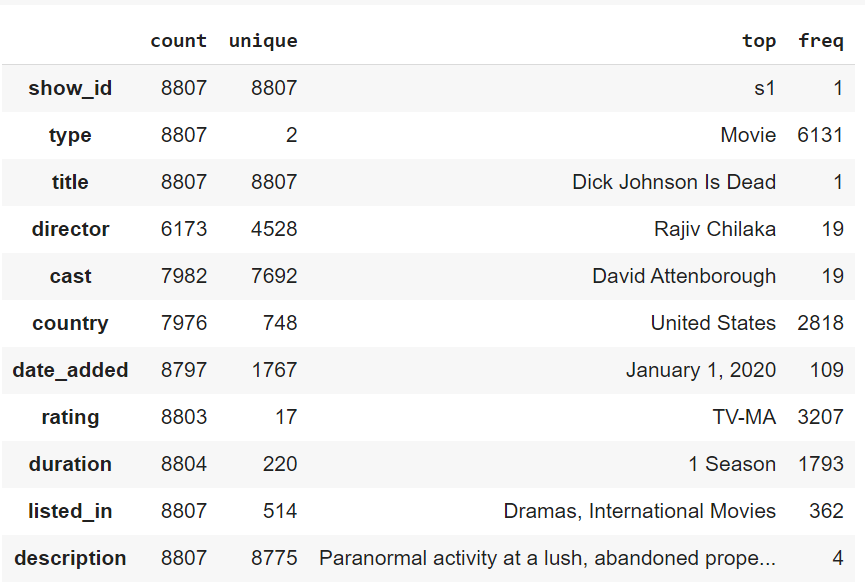
 -netflix.info()

Gives the idea on number of rows present in the dataset, data type of each column, and number non null values present.



-netflix.describe(include='object').T

Returns the descriptive statistics summary of the dataframe.



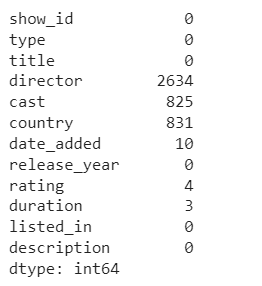
-netflix.shape

Returns shape of the dataframe.

(8807, 12)

-Checking for missing values in each column.

netflix.isnull().sum()



On exploring the dataset, we notice a few problems:

1. Missing values: We will be imputing the missing values, by using different techniques.

2.Nested values: Nested columns such as listed\_in, cast and country ,will be unnested for better analysis.

3.date\_added: date\_added column will converted to date type from object.

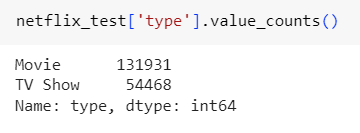
4.Data Inconsistency: We can see values for few values duration column present in rating column, we will correct this .

After employing all the fixes on our dataset, we start to do some analysis.

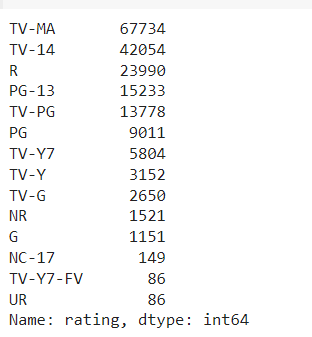
**Non-Graphical Analysis:**

We check the value counts for each of the categorical attributes:

Show Type:



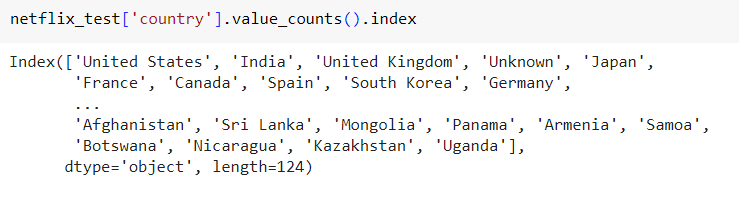
Rating:



Genre:



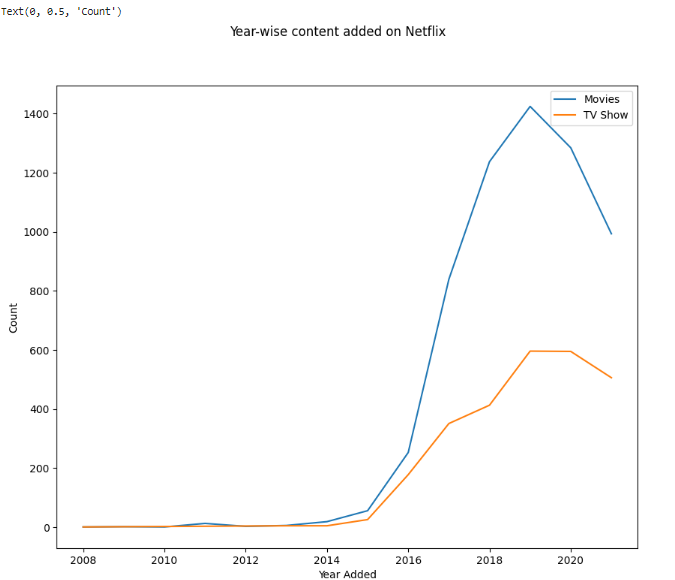
Country:



**Graphical Analysis:**

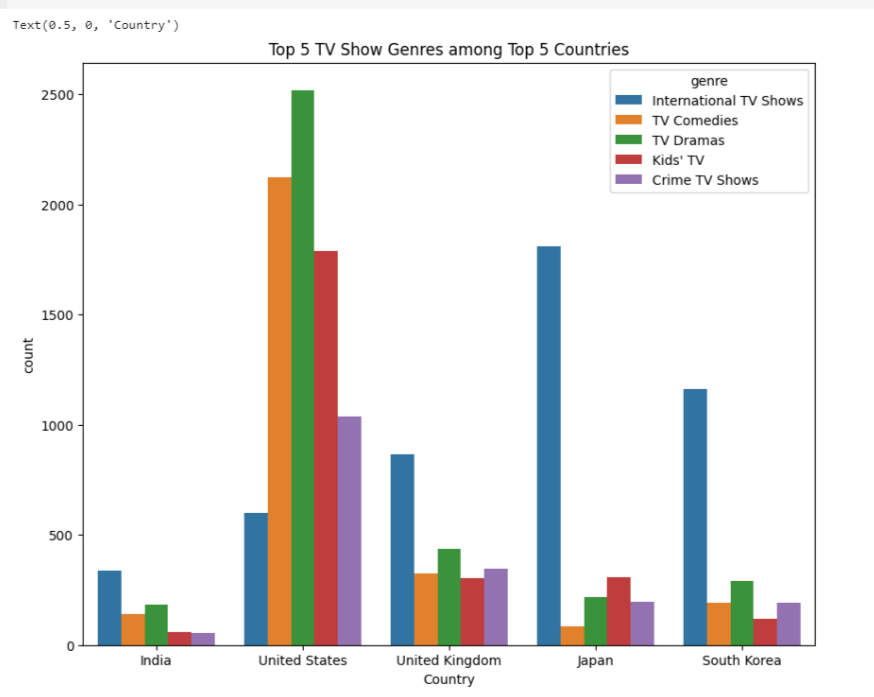
We try to plot some graphs to gain some insights.

1. Year-wise content added to Netflix



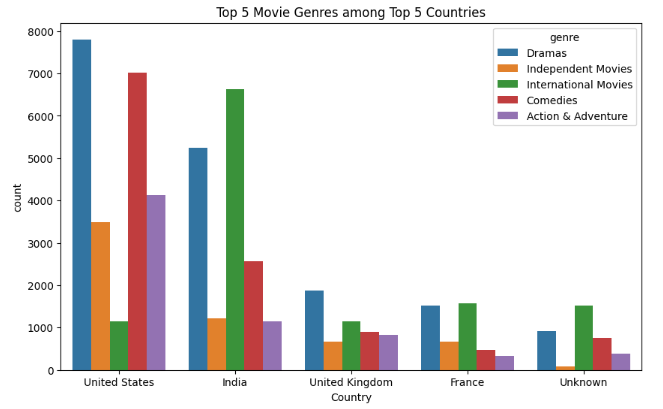
From the plot we can infer, although the movies content is higher than TV shows, we can see a decline in the no. of movies added in recent years , whereas the no. of TV series added is more stable from past few year, possibly due to the increase in popularity of TV shows.

2. Top 5 TV Show Genres among Top 5 Countries



From the plot we can infer, United States has the highest TV show content added , with TV dramas being the most popular genre. Whereas other countries, mostly prefer International TV Shows.

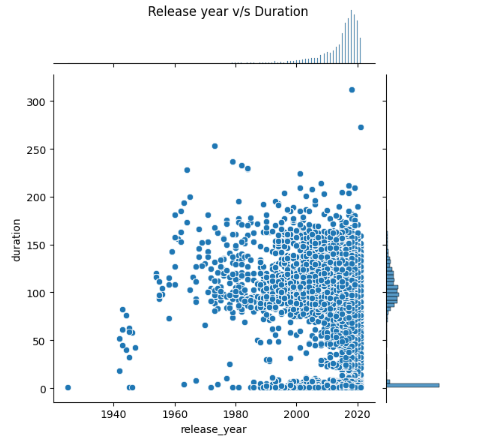
3.Top 5 Movie Genres among Top 5 Countries



From the plot we can infer, United States has the highest Movie show content added , with Dramas and Comedies being the most popular genre. Whereas other countries, mostly prefer International Movies and Dramas.

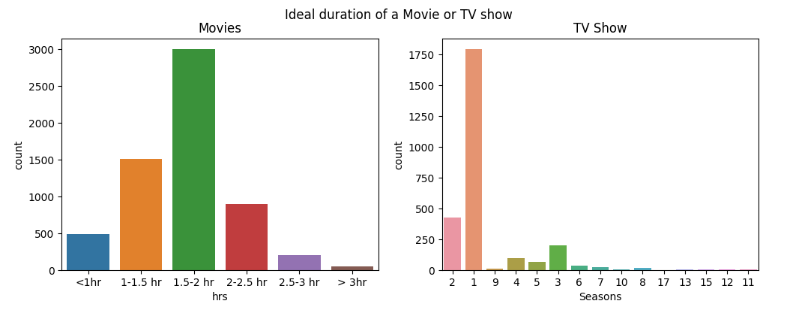
4. Finding the Ideal Duration of a Movie

Plot1: Duration time over the years



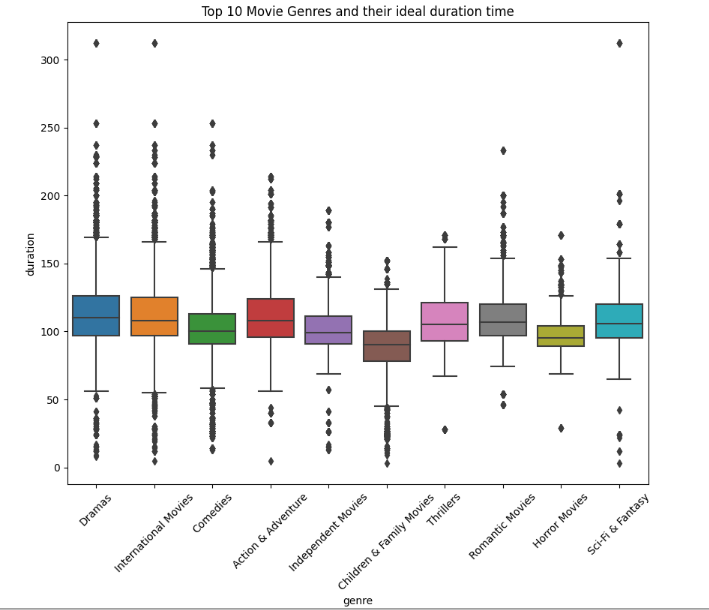
We can infer, before year 2000 there was no ideal duration time for a Movie. But post 2000, we can observe a lot of movies fall under 90-120 min range.

Plot 2: Ideal Duration of Movie or TV show.



We get further evidence that most movies have duration of 1.5-2 hrs.

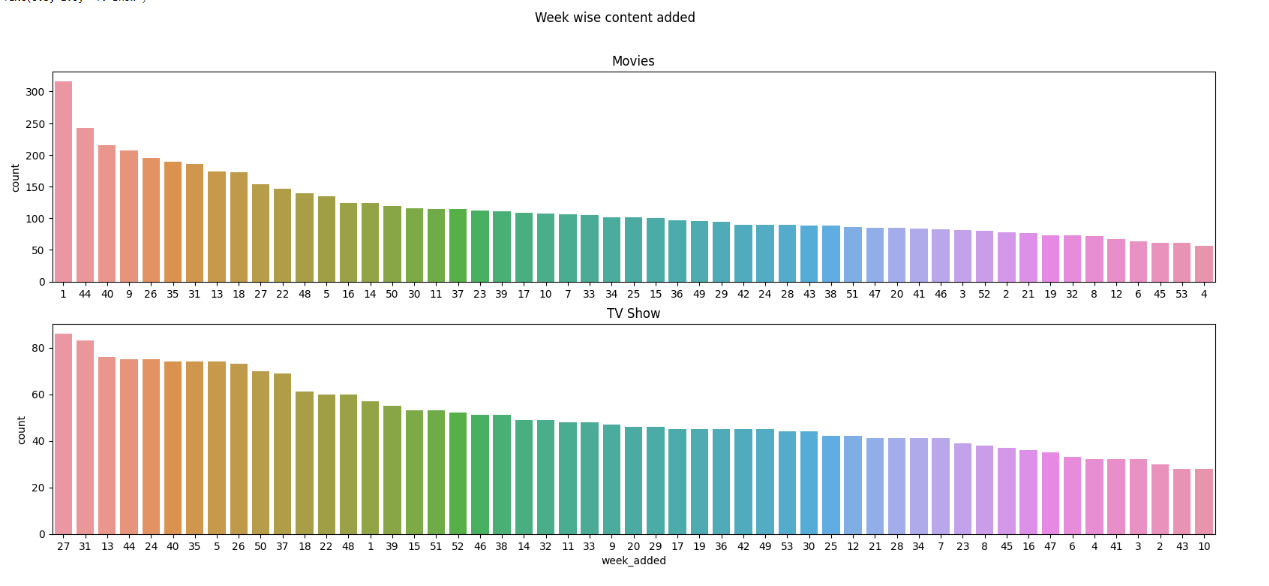
Plot 3: Top 10 Movie Genres and their ideal duration time



 We can further observe that for top 10 movie genres, the median duration time for most genres fall at 120min mark.

We can conclude the ideal duration for a movie on Netflix is 1.5-2hrs.5.

 5. Identifying which week has highest content added



 For Movies, we can see that week 1 has highest content added.

For Movies, we can see that week 27 and 31  has highest content added.

**Business Insights:**

1. We observed that TV shows had a steady growth in popularity over the recent years.

2. United States has the highest TV show content produced , with TV dramas being the most popular genre. Whereas other countries in the top 5 TV shows produced, mostly prefer International TV Shows. Other popular genres are TV Comedies, Crime TV shows and Kids TV.

3. United States has the highest Movie show content produced , with Dramas and Comedies being the most popular genre. Whereas other countries in the top 5 TV shows produced, mostly prefer International Movies and Dramas. Other popular genres are Action and Adventure, Independent Movies.

4. The Ideal watch time of a Movie to keep the viewer interested is 1.5-2hrs.

5. We observed that most movies are added on first week of the year (Jan) and also in October (week no. 40,44), and most TV shows are added in the months of July(week 27) and August(week 31).

**Recommendations:**

1. Add/produce movies that are ideally in range of 90-120mins.

2. Add/produce TV shows of the following genre- TV dramas, International TV Shows, TV Comedies, Crime TV shows and Kids TV.

3. Add/produce Movies of the following genre- Dramas, Comedies, International Movies, Action & Adventure, Independent Movies.