Task 1 Report:

SRI HARSHITHA P,Kumaraguru College of Technology

Steps:

Step 1: Basic Exploration:

Importing Pandas for data management, NumPy for numerical operations, Matplotlib for creating visualizations, and Seaborn for enhancing the aesthetics of our plots.

The dataset comprises 8,807 rows and 12 columns, containing information about various TV shows and movies. The columns include 'show\_id', 'type', 'title', 'director', 'cast', 'country', 'date\_added', 'release\_year', 'rating', 'duration', 'listed\_in', and 'description'.

Step 2:Null value treatment

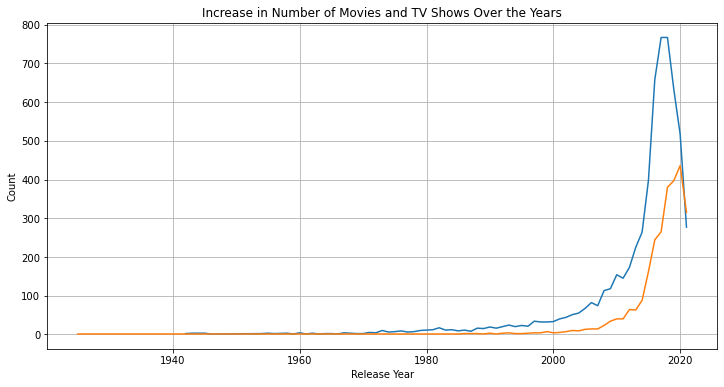
The dataset is examined for missing values, and the 'director', 'cast', 'country',

'rating', and 'duration' columns are imputed with "Unknown" as the columns

were having considerable number of missing values.

Exploring the column “type”:

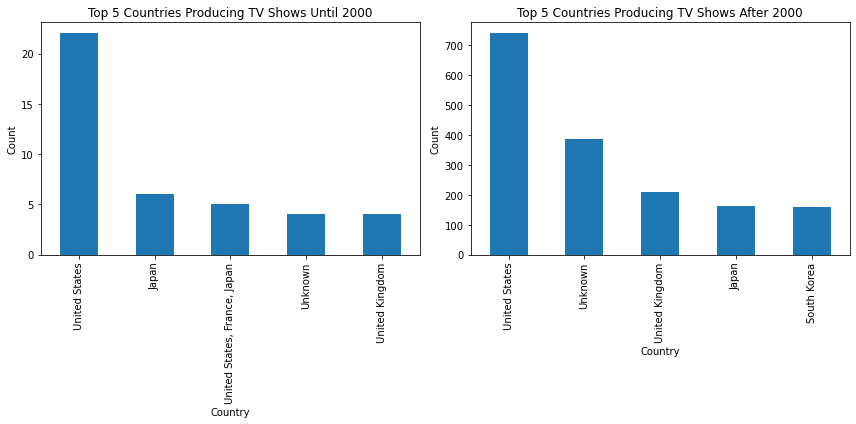
* We see that movie has more entries in this dataset than TV shows. A countplot was created to visualize the distribution of content types in the dataset. It revealed that there are 6,131movies and 2,676 TV shows.
* A countplot was generated to display the distribution of content types in the dataset for entries with a release year before and after 2000. We see that count of TV shows have **increased significantly since 2000**.
* A line plot was generated to visualize the growth in the number of movies and TV shows over the years.We see a gradual increase in number of TV Shows and movies over the years.



* Two separate plots are created, displaying the top 5 countries producing movies until and after 2000. Similarly, it filters and creates separate plots for TV shows released until and after 2000, showcasing the top 5 countries in each category.
* We see that United States stands first among all the countries to produce movies and TV Shows since 1925.India finds a place among top 5 in movie production but not TV Show production. United Kingdom finds a place among top 5 in both since 1925.Hongkong has been good producer of movies until 2000 and Canada evolved among top movie producers since 2000.Japan and South Korea have stroke a chord in TV Shows.

A graph with blue bars

Description automatically generated with medium confidence



A blue and white bars

Description automatically generated with medium confidence

* We observe that mostly before and after 2000, both movies and TV shows strongly recommend watching them with parental consent. Some TV shows before and after 2000 are designed for general audiences and children older than 7 years.
* On the whole, We see India is the second largest movie producer among all. US stands first.
* India doesn’t hold any of the top 5 positions in TV Shows. The countries in bar chart are the top 5 countries.