

---

# Course Project on Analysis of Video Platforms

---

**ZHANG Yuxuan**

221041056

School of Data Science

The Chinese University of Hong Kong, Shenzhen

Longgang, Shenzhen

221041056@link.cuhk.edu.cn

## Abstract

This course project focus on analysis of video stream data. Based on 4 tabular datasets consisting of information of movies or TV shows available on 4 top-tier video streaming platforms, analysis of characteristics of different platforms, trend of different country of production, proportion of different kind of contents, etc., are presented Using Python data analysis techniques. Specifically, data processing package Pandas, data visualization package Matplotlib etc. are used.

## 1 Introduction

Media and Video streaming platforms are getting more and more popular in recent years and many companies are following this trend. In this project, we use 4 tabular datasets consisting of information of movies or TV shows available on 4 top-tier video streaming platforms, including Amazon Prim, Disney+, Hulu, Netflix. Analysis on these four platforms are made using given datasets and compare them with each other.

## 2 Information of Datasets

### 2.1 Information of the platforms

Hulu is an online movie and tv shows streaming platform owned by The Walt Disney Company, launched on 2007. Currently it is only available in the United States. As of the fourth quarter of 2021, Hulu has 43.8 billion subscribers. Disney+ is an American subscription video on-demand over-the-top streaming service launched on 2019. It is available in America, Europe, India and Australia, and owns 118.1 million global subscribers as of October 2, 2021. Amazon Prime is a paid subscription program from Amazon which is available in various countries and gives users access to additional services otherwise unavailable or available at a premium to other Amazon customers. In April 2021, Amazon reported that Prime had more than 200 million subscribers worldwide. Netflix is an American subscription streaming service and production company. Launched on August 29, 1997, it offers a library of films and television series through distribution deals as well as its own productions, known as Netflix Originals. As of October 2021, it has over 214 million subscribers worldwide.

### 2.2 Dataset Overview

The given datasets are contained in four CSV files, each row has information of one movie/TV show. The information includes: the unique id of the item, type of the item(Movie or TV show), title, director, casts, country of production, date added on the platform, original release year, rating, duration, the genre(s) it belongs to, brief description of the item. The Amazon dataset have 9668

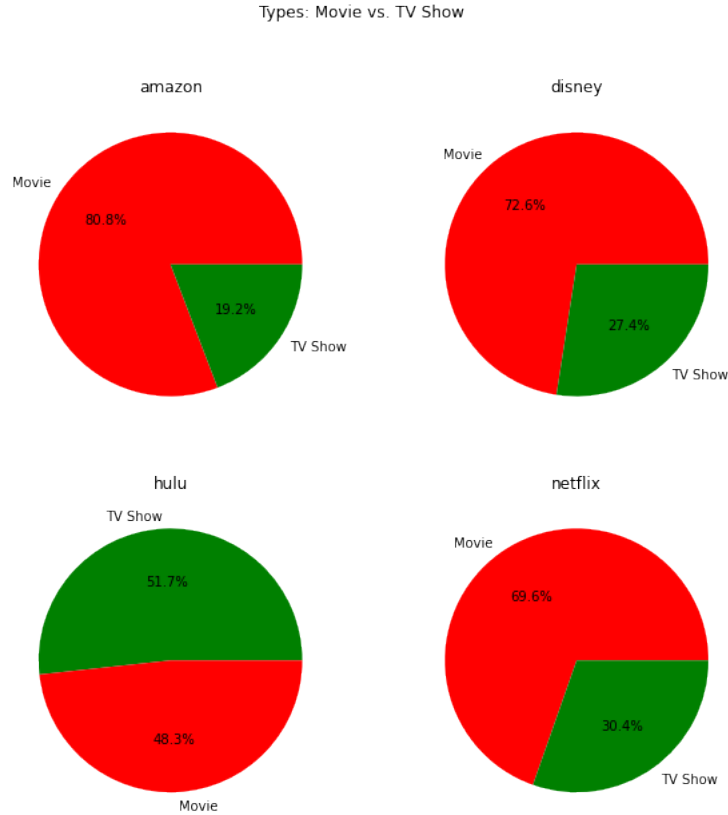


Figure 1: Movie vs. TV Show

records, the Disney+ dataset have 1450 records, the Hulu dataset have 3073 datasets, and netflix dataset have 8807 records, thus we have 22998 records in total.

### 3 Comparison of the Platforms

In this section, comparisons are made between the 4 platforms.

#### 3.1 Types: Movie vs. TV Show

Figure.1 illustrate the proportion of Movies and TV shows respectively in 4 platforms. Most platforms have more movies than TV shows, among which movies takes 80.8% of all contents on Amazon, which is the greatest. However, Hulu have more TV shows than movies.

#### 3.2 Release Year of Items

Figure.2 illustrate the distribution of Movies and TV shows with respect to release years respectively in 4 platforms. The distribution generally follows the exponential growth of contents production. However, contents on Hulu and Netflix showed decreasing trend after 2020. Also, comparing to other platforms, disney+ have more contents from earlier years.

#### 3.3 Added Year of Items

Figure.3 illustrate the distribution of Movies and TV shows with respect to added years respectively in 4 platforms. The distribution of Hulu and Netflix generally follows the exponential growth of contents production. However, the disney+ shows decreasing trend since we only have data starting with 2019, and the following year are impacted by COVID. The result of amazon do not indicate

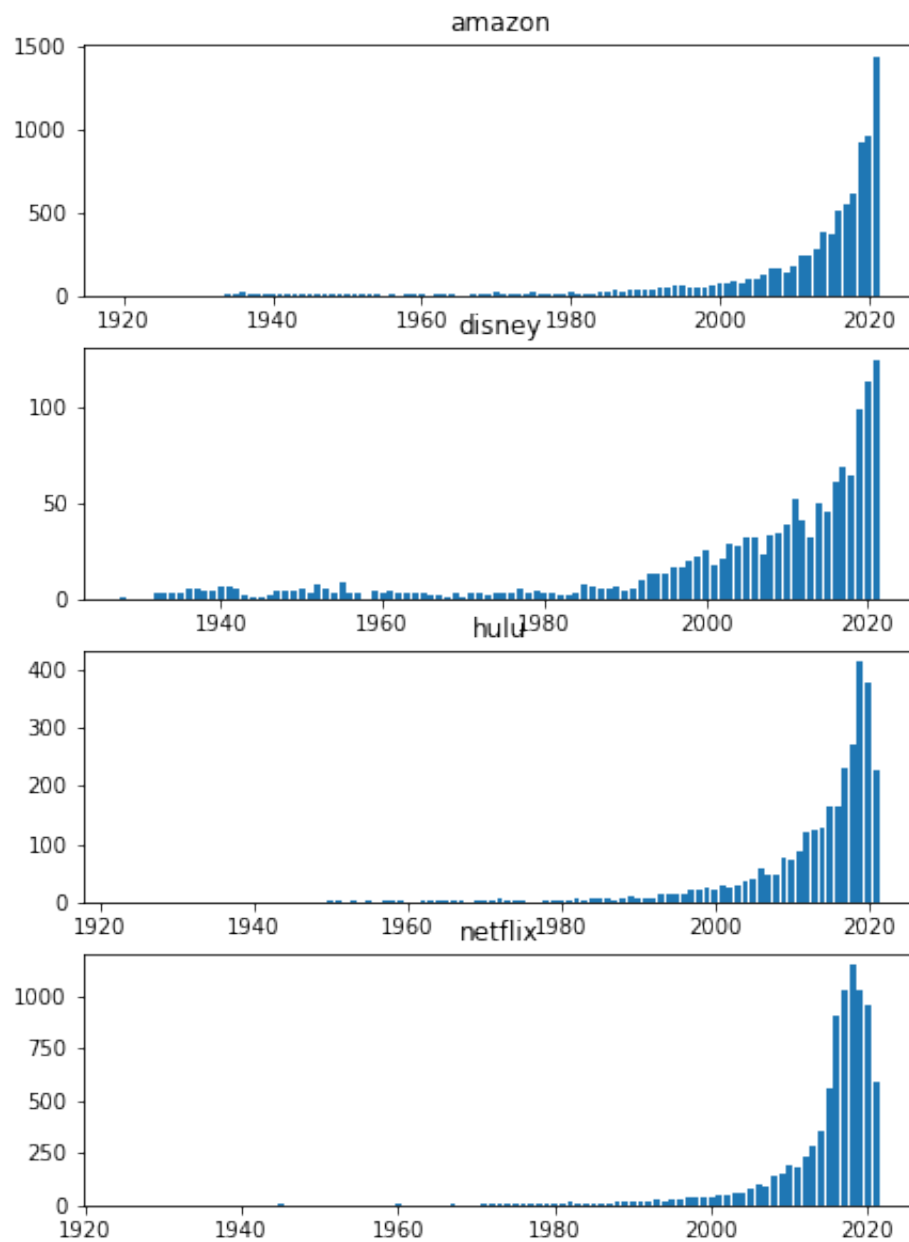


Figure 2: Movie vs. TV Show

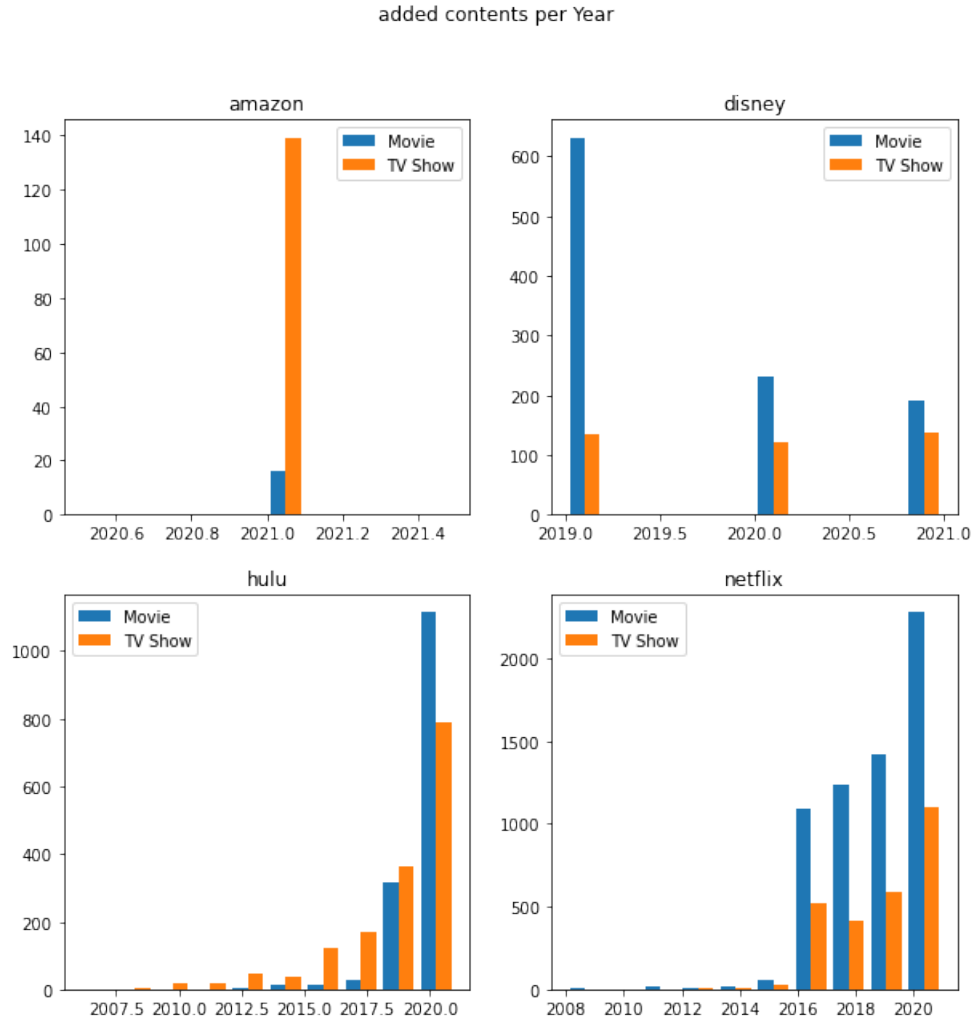


Figure 3: Year of Contents being Added Online

more information because in fact, in the amazon dataset, only 155 of the 9668 records show the information of added year.

### 3.4 Genre of Contents

Figure.4 shows the wordclouds of genres of contents in 4 platforms, which clearly indicates the type of customers that each platform aims at. Here, the general classification such as comedy, drama, TV, Movies, etc., are not considered. Amazon have more contents of action, suspense type; Disney+ have more animation, family contents; Hulu focus on action, adventure; and netflix have more international contents.

### 3.5 Rating of Contents

Figure.5 shows the distribution of ratings of contents in 4 platforms, which more clearly indicates the type of customers that each platform aims at. For contents on Disney+, the greatest proportion of content is rated as TV-G, which means most parents would find it suitable for all ages. Since Disney is a company that also produces animation for children, it is reasonable that its video platform aims at family users. As for other platforms, most contents have age limitations. In fact, all-age contents takes a very small proportion of contents in platforms other than Disney+.

[illegible]

5

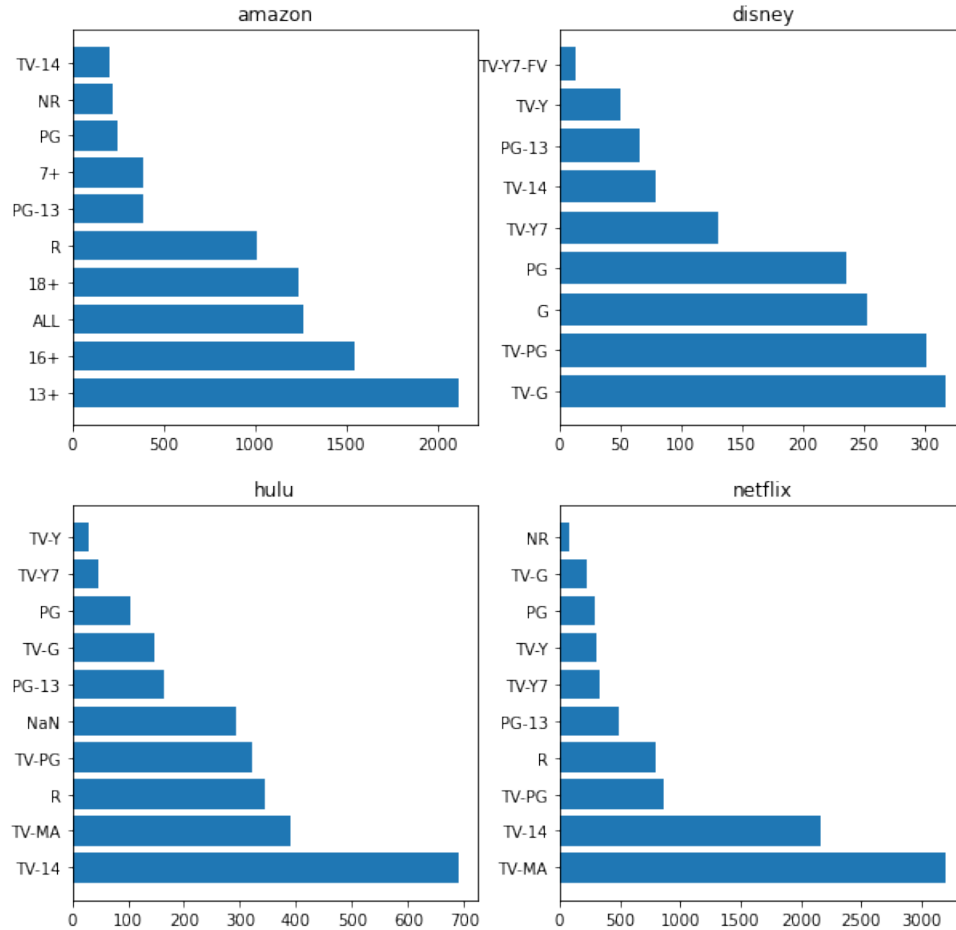


Figure 5: Distribution of Rating

## 4 Comparison of the Countries

In this section, comparisons are made between main countries of production.

### 4.1 Country of Production in Each Platform

Figure.6 shows the distribution of country of production in 4 platforms. All 4 platforms have most content from the United States, especially for Disney+, where international contents takes very small proportion, whereas amazon have more international contents. This statistic indicates that Disney+ focus on domestic users, and other platform probably have more global market share.

### 4.2 Country of Production in Each Platform

Figure.7 shows the distribution of Top 10 country of production of all platforms in total. The Top 6 countries with the most productions are respectively United States, India, United Kingdom, Japan, Canada and South Korea. Figure 8 showed the trend of production per year of the 6 countries respectively. All countries experienced exponential growth of content production, and the US is more advanced in this process. However, the production of all 6 countries dropped in 2020, mostly because of the impact of COVID-19.

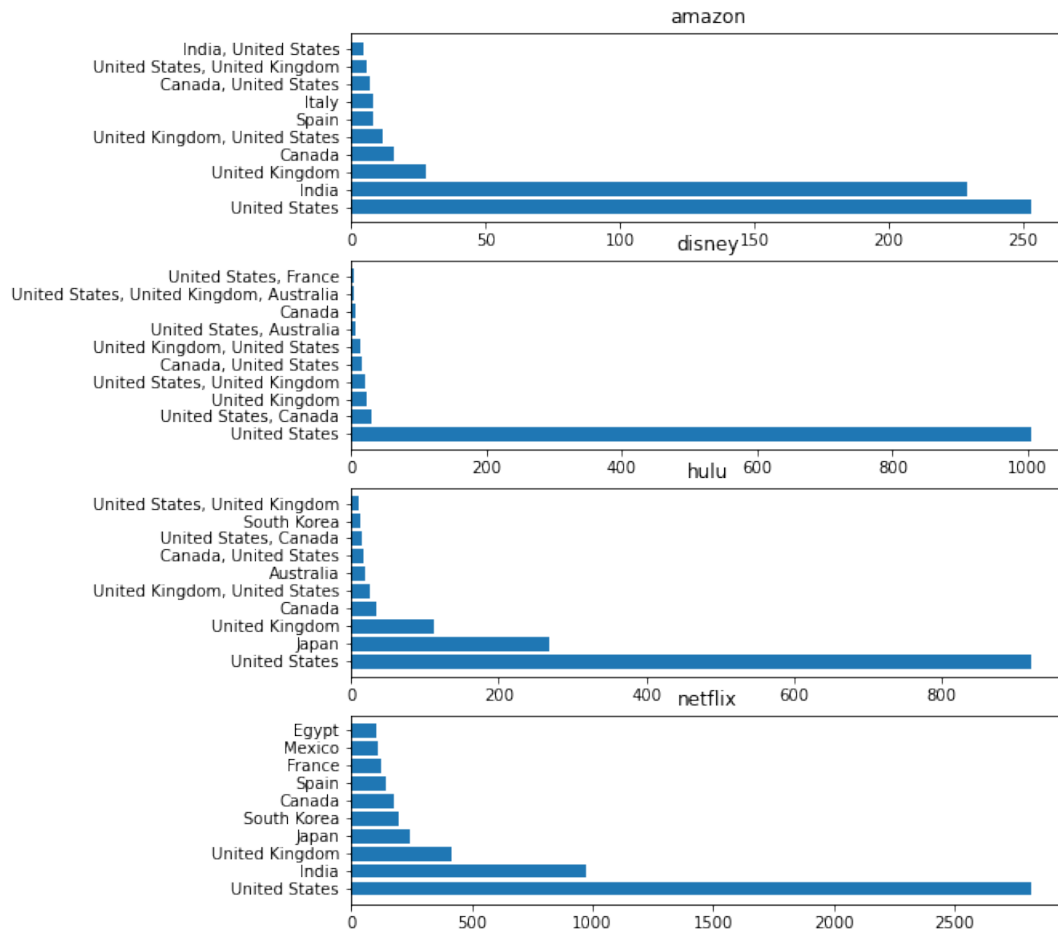


Figure 6: Distribution of Country of Production

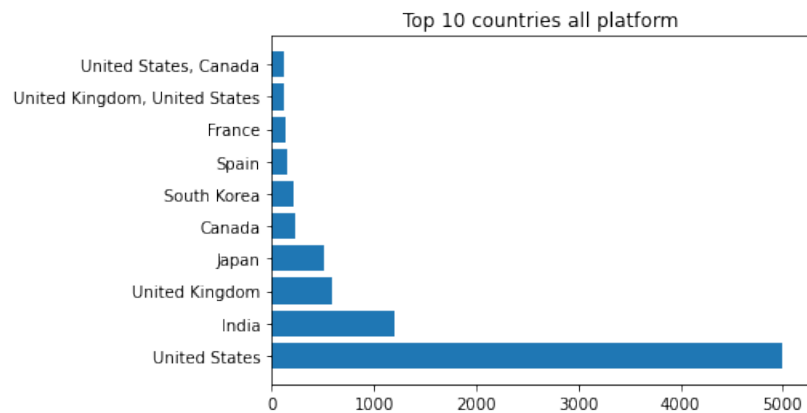


Figure 7: Distribution of Top 10 Country of Production of all Platforms

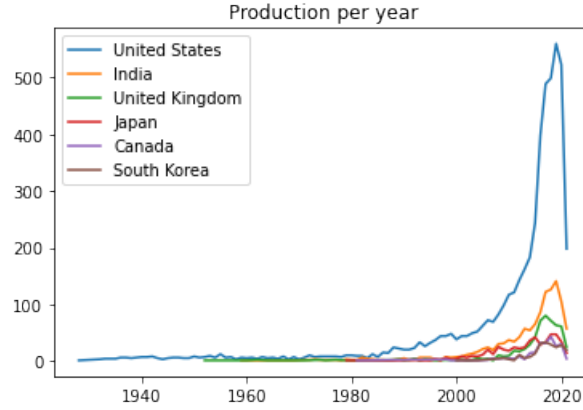


Figure 8: Trend of Top 6 Country of Production per year

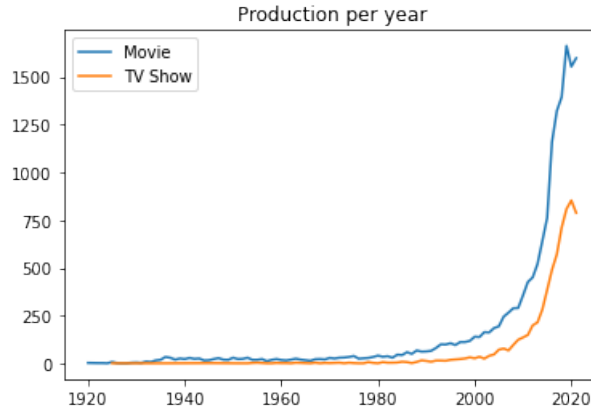


Figure 9: Trend of Top 6 Country of Production per year

## 5 Trend of Movies and TV Shows in Online Platforms

Figure 9 shows the annual growth of movies and TV show in all platforms. The general trend is exponential, and production of both movies and TV shows dropped because of the impact of COVID-19.

## 6 Conclusion

In this project, analysis are made using 4 tabular datasets consisting of information of movies or TV shows available on 4 top-tier video streaming platforms, including Amazon Prim, Disney+, Hulu, Netflix. First, comparison are made between the 4 platforms to get their characteristics, and their operation strategies can be analyzed. Then, analysis are made based on different countries of production. Finally, trend of movies and TV shows are given. The project shows that useful information can be obtained based on analysis of raw data.