

# Prime Streaming Content

## Business Understanding

Amazon prime is a popular streaming service that offers a vast catalog of movies, TV shows, and original contents. The data consist of contents added to Prime from 2008 to 2021. The oldest content is as old as 1925 and the newest as 2021. The tasks that you need to perform are

- Undertake an exploratory data analysis and produce a 2-page report on your findings/insights
- Clean the data and produce good visualisation for story telling

## Introduction

Amazon Prime Video is an on-demand streaming service with many popular movies and shows. It is home to a wide range of original series and films.

Amazon Prime was launched on September 7, 2006 in United States. Prime Video was added to the Prime subscription in the United Kingdom, Germany and Austria in 2014. In 2016, Prime video was launched worldwide except Mainland China, Cuba, Iran, North Korea and Syria. In 2013, Amazon.com began distributing original television shows through its Amazon Prime Video service.

The given dataset has details about movies that were released in 1925 till 2021. Detailed visualisation was done to identify which countries dominated in terms of maximum number of movies/shows streaming in prime, famous genre of movies, movies/shows with long duration, number of movies/shows added to Prime in each year, etc.

## Initial Findings

- There are 8807 rows and 12 columns in the dataset
- There are totally 4307 null values in 6 columns.
- There are no duplicate entries in the dataset.

## Exploratory Data Analysis

### Data Cleaning

- Datatype of date added column was changed to datetime datatype.
- The Null values in director, cast and country were changed as Unknown.
- Now there are only 17 null values in date added, rating and duration. Since it is approximately 0.2% of the total data, the columns with null values were dropped.
- After null value treatment, there are 8790 rows in the dataset.
- The column show\_id is just an ID column and is not a useful column for EDA. So, this column is dropped from the dataset. Now there are 11 columns in the dataset.

There are 8790 unique movie titles, 4527 different movie directors, 14 different ratings, 513 different genres and 749 different countries from which movies are added to Prime.

Various plots and graphs were created as part of EDA in Tableau. Univariate and bivariate analysis were done using the graphs created.

(For detailed story in Tableau, check the below link:

<https://public.tableau.com/app/profile/chandini.c8189/viz/PrimeStreamingContentVisualization/Story1?publish=yes>)

## **Inference from the graphs**

- The Top 5 countries from which most of the movies and TV shows are added to Prime are United States [US], India, United Kingdom [UK], Japan and South Korea.
- The first movie/show was added in 2008. It was from US. On the whole, maximum movies were added in 2019. Movies from UK and Japan started to stream from 2015 and movies from South Korea from 2016 in Prime.
- Movies/shows belonging to India started to stream from 2016. In 2018, 333 movies and shows were added from India. This is the highest number over the years with respect to the movies belonging from India. There is a downfall after 2018. In 2021, only 95 movies/shows were added.
- In US, movies/shows with 10 different ratings were released. Most of the movies have TV-MA rating. In India, movies/shows with 5 different ratings were released. Most of the movies have TV-14 rating.
- In US and India, number of movies streaming in Prime is more than number of shows. In UK, Japan and South Korea, number of shows is more than that of movies.
- Most of the movies/shows in Prime are of the genre Dramas, International Movies and Documentaries. Another big chunk belongs to Stand-up Comedy.
- Maximum movies from US and UK are of Documentaries genre. The top genre of movies from India are Comedies, Dramas, International movies.
- Maximum number of movies/shows in Prime is of rating TV-MA followed by TV-14.
- Almost 70% of the content in Prime is of movies and the remaining 30% belongs to Tv shows.
- There is a huge rise in the number of movies/shows added to Prime in 2017 compared to previous years. Maximum movies were added in 2019 and maximum shows were added in 2020. Totally 2016 movies were added in the year 2019. This is the highest number of movies/shows added in a year. The number falls after 2019. There is a drastic fall in 2021.
- The movie Black Mirror: Bandersnatch is the lengthiest movie in Prime with duration of 312 minutes. It was released in 2018 and it from US. Grey's Anatomy is the show in Prime with the highest number of seasons – 17 seasons. It is also from US.
- 19 movies directed by Director Rajiv Chilaka is the highest number of movies streaming in Prime from a single director.

## **Conclusion**

The major reason for the decline number of movies added to Prime in all the countries is due to Covid because number of movies/shows releasing has declined. The fall in 2021 is more than that of 2020 because many movies that were completed before Covid were released directly onto Prime without theatrical release in 2020. In 2021, number of movies produced itself was less. So, in Prime a smaller number of movies/shows were added. It is expected that the number will increase in the later years.