

**COURSE TITLE: EXPLORATORY DATA ANALYSIS**

**COURSE CODE: 21ECSC210**

**PROJECT TITLE: ANALYSIS ON OTT PLATFORM**

**Submitted by:**

TEAM-09

ANIKET 01fe20bcs044 144

YAJAS 01fe20bcs048 148

KARTHIK 01fe20bcs055 154

SANGAMESH 01fe20bcs060 159

Contents:

1. Abstract
2. Introduction
3. Problem Statement
4. Dataset Description
5. Data Pre-processing
6. Data Integration
7. Questions and Analysis
8. Conclusions
9. References

Abstract:

Over-the-Top (OTT) video platforms, once considered a luxury is today a commodity. In India, there is an increasingly growing number of consumers adapting to it. While Indian streaming services like Hotstar and Jio Cinema has gained a stronger foothold, global players like Netflix and Amazon Prime have steadily grown their market share in India. This paper explores the emergence, advantage, and future of streaming service in India through analytical research. We also present the various OTT services, their growth factors, technology background, audience characteristics, content, censorship and future developments expected in the industry.

**Introduction:**

Traditionally, the consumption of movies and other audio and video content has always been in the form of mediums like theatre and television. As the technology developed, it was easily accessible at home and whenever required with the introduction of VHS, DVDs, Blu-rays and disc rental services. Further, cable television brought the content through Co-axial cables and fibre optic cables. Recently, technological advancements have made the movie or TV watching more convenient through online streaming or Video on Demand (VoD) services. VoD refers to streaming of video content over the Internet, through applications typically referred to as Over-The-Top (OTT). Viewers can access video content through OTT apps in any Internet-connected device like a Smartphone, smart TV, tablet, desktop computer, laptop, etc. Unlike traditional media, streaming services tell varied stories that are not restricted by censors, box office or demographic. It gives a viewing experience with greatly improved sound and visual quality, provided the consumers have a stable Internet.

**Problem Statement:**

Analysis of the Covid 19 Pandemic on the OTT Platform

**Dataset Description:**

Our data consists of 6 datasets namely:

* **Netflix Dataset:** This dataset consists of **8807** rows x **13** columns of movie and tv shows, directors, cast, country, release year, duration, rating, listed in and description.
* **Amazon Prime Dataset:** This dataset consists of **9668** rows x **13** columns of movie and tv shows, directors, cast, country, release year, duration, rating, listed in and description.
* **Disney+ Hotstar Dataset:** This dataset consists of **1450** rows x **13** columns of movie and tv shows, directors, cast, country, release year, duration, rating, listed in and description.
* **Hulu Dataset:** This dataset consists of **3073** rows x **13** columns of movie and tv shows, directors, cast, country, release year, duration, rating, listed in and description.
* **Movie Dataset:** This dataset consists of **9515** rows x **11** columns of movies, year, age, rotten tomatoes, type and OTT platforms.
* **Tv show Dataset:** This dataset consists of **5368** rows x **11** columns of movies, year, age, rotten tomatoes, type and OTT platforms
* **Revenue Dataset:** This dataset consists of **40** rows X **3** columns of area, years, revenue of Netflix platform.

**Attributes:**

|  |  |  |
| --- | --- | --- |
| **Attribute Name** | **Description** | **Type of Data** |
| Type | Movie or Tv show | Categorical Data |
| Title | Title of the content | Nominal Data |
| Release year, Year | Year in which the content was released | Ordinal Data |
| Listed in | Genre of the content | Nominal Data |
| Ott | Name of the Ott platform | Categorical Data |
| Age | Age group | Categorical Data |
| IMDb | IMDb ratings of content | Discrete Data |
| Rotten Tomatoes | Rotten Tomatoes ratings of content | Discrete Data |
| Area | Area /Region | Categorical Data |
| Years | Year in which revenue is collected | Ordinal Data |
| Revenue | Revenue of Netflix | Discrete Data |

**Data Pre-processing:**

* **Netflix Dataset:** This dataset consists of total **4307** null values, we dropped few unwanted columns in order to remove the null values.
* **Amazon Prime Dataset:** This dataset consists of total **22161** null values, we dropped few unwanted columns in order to remove the null values.
* **Disney+ Hotstar Dataset:** This dataset consists of total **888** null values, we dropped few unwanted columns in order to remove the null values.
* **Hulu Dataset:** This dataset consists of total **8627** null values, we dropped few unwanted columns in order to remove the null values.
* **Movies Dataset:** This dataset consists of total **4184** null values, replaced the null values by mode of the attribute in order to remove the null values.
* **Tv shows Dataset:** This dataset consists of total **3089** null values, replaced the null values by mode of the attribute in order to remove the null values.
* **Revenue Dataset:** This dataset consists of **zero** null values.

**Data Integration:**

* We have integrated Netflix, Amazon Prime, Disney+ Hotstar and Hulu datasets into one dataset.
* We have integrated Movies and Tv shows datasets into one dataset

**Questions and Analysis:**

1. Growth of OTT platforms year wise.

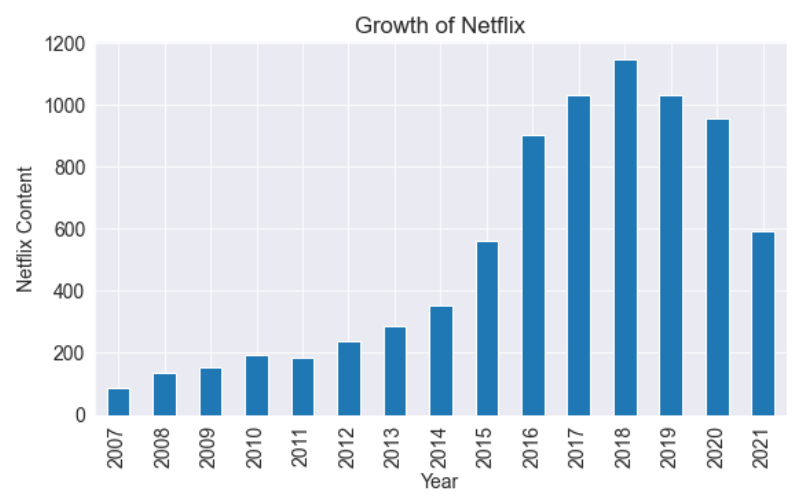


Figure 1.a: Total no. of content present in Netflix over the years.

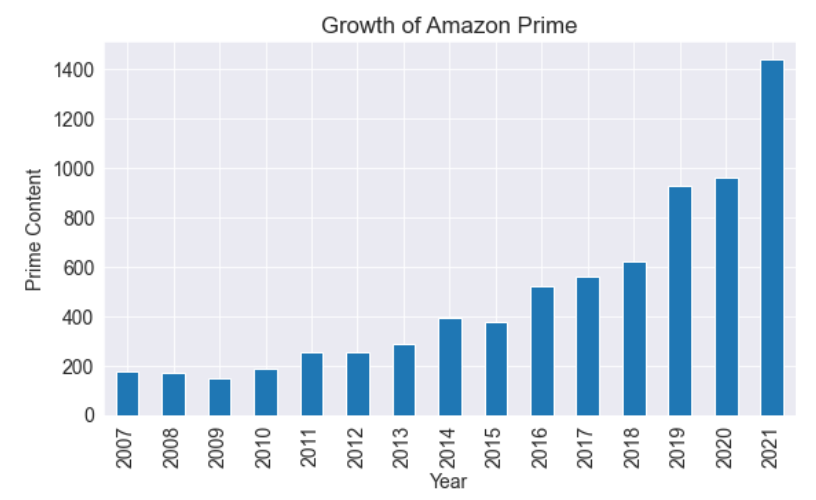


Figure 1.b: Total no. of content present in Amazon Prime over the years.

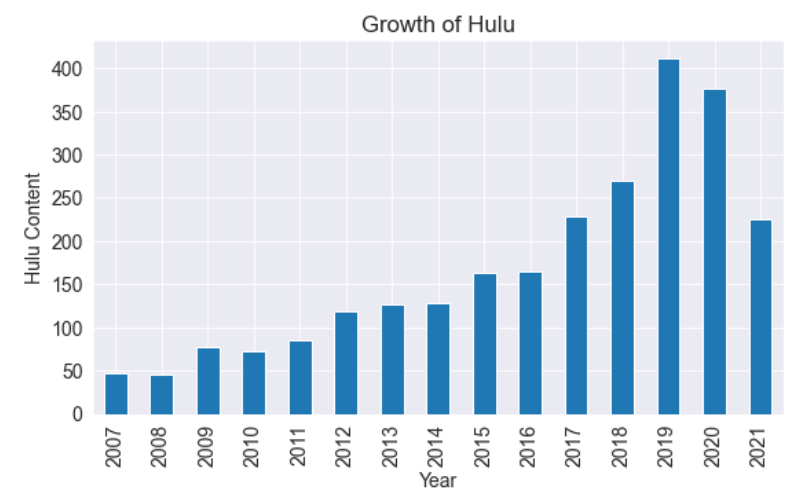


Figure 1.c: Total no. of content present in Hulu over the years.

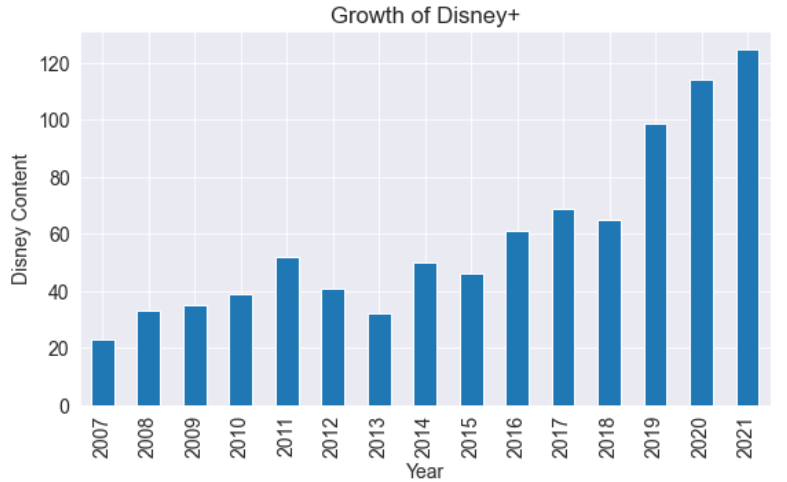


Figure 1.d: Total no. of content present in Disney+ hotstar over the years.

**Inference:**

* Figure 1.a says that Netflix in the year 2020 content decreased because of lack of new movies and tv shows.
* Figure 1.b says Amazon prime in the year 2020 content got increased as they added existing movies and tv shows.
* Figure 1.c says that Hulu content in the year 2020 decreased because of lack of movies and tv shows.
* Figure 1.d says Disney + Hotstar in the year 2020 content got increased as they added existing movies and tv shows.

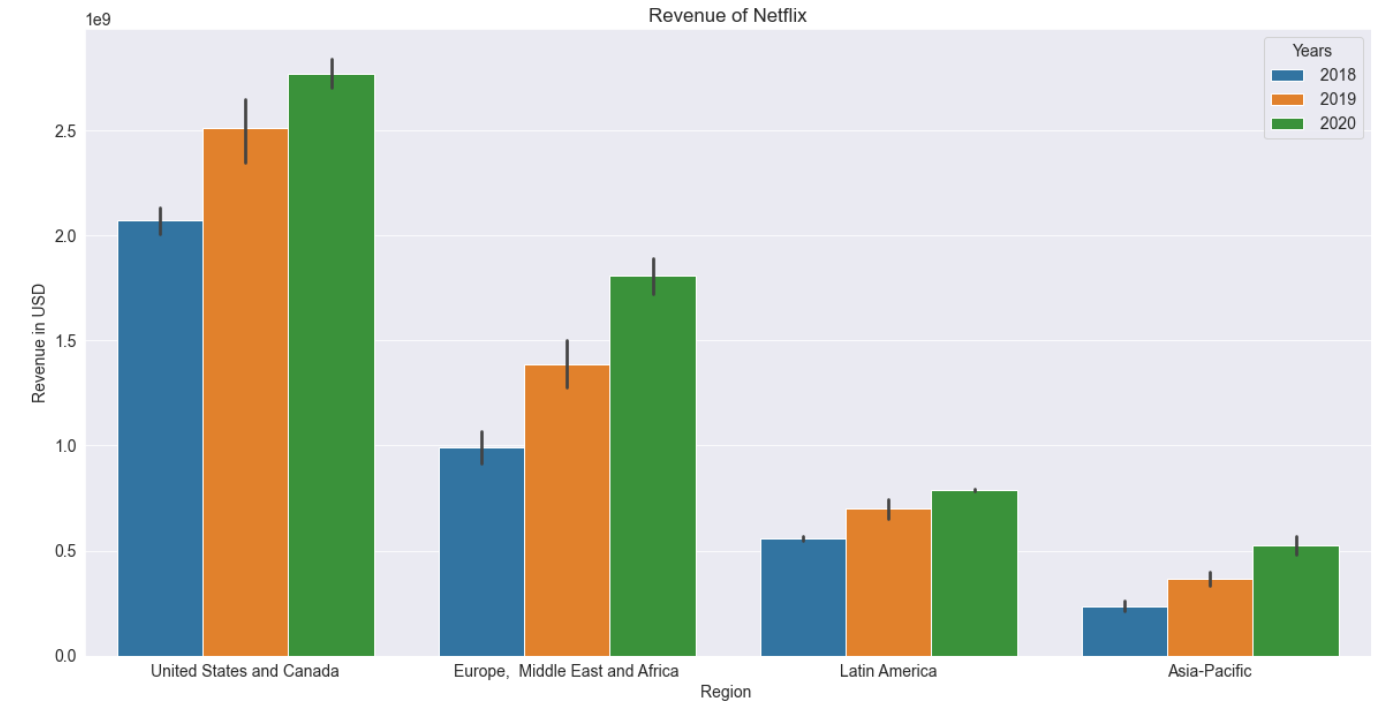
2. Revenue of Netflix over the years.

Figure 2: Revenue of Netflix in 2018, 2019 and 2020

**Inference:**

* From figure 2, revenue of Netflix has increased in all regions, comparing the revenue of the year 2018 and 2019.

# 3. The no. of content produced by OTT platforms in the year 2020 and 2021.

# 

Figure 3.a: No. of content in different OTT platforms in year 2020.

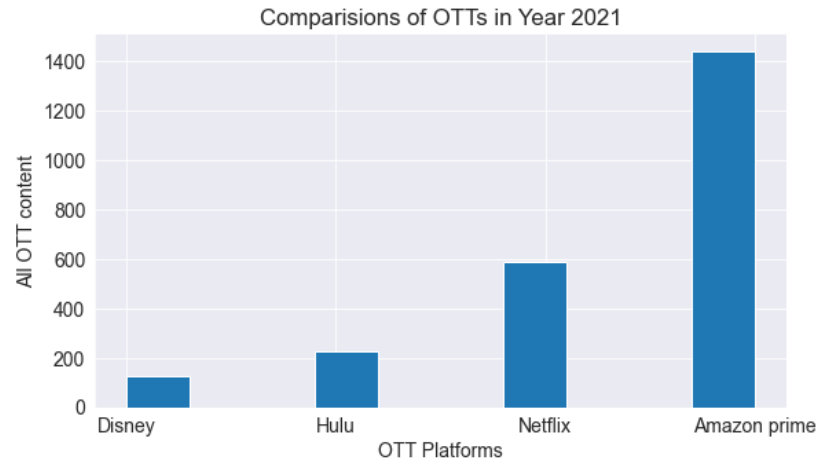


Figure 3.b: No. of content in different OTT platforms in year 2021.

**Inference:**

* Figure 3.a shows the number of movies and tv shows present in each OTT platform in the year 2020.
* Figure 3.b shows the number of movies and tv shows present in each OTT platform in the year 2021.

4.The impact of COVID-19 on IMDb ratings.

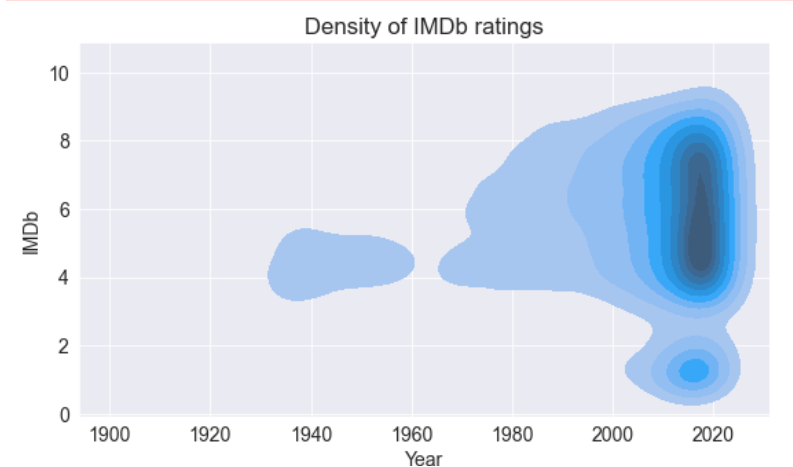


Figure 4: Density of IMDb ratings over the years.

**Inference:**

* Figure 4 shows that how density of IMDb ratings is high near the year of 2020.
* It also says that people watched more number of movies and tv shows and rated them in the year 2020.

# 5. Type of Content that is added by Netflix in the year 2020.

# 

Figure 5: Type content released by Netflix in the year 2020.

**Inference:**

* Figure 5 says that what are movies and content added by Netflix based on the maturity ratings. TV-MA content is added the highest amongst all.

**Conclusions:**

* By this data analysis we can conclude that the growth of OTT platforms as been increased during the COVID-19 lockdown phase.
* Revenue of Netflix has increased in all regions, comparing the revenue of the year 2018 and 2019.
* Netflix and Amazon Prime has grown more among all the OTT platforms in terms of content added.
* As people had free time, people started to watch movies and tv shows and also give them ratings, resulting in the increase of IMDb ratings count.
* The number of content added by each OTT platforms in the year 2020 and 2021, in which Amazon Prime has added more content compared to all other platform.
* As Netflix is more popular among the OTT platforms, so the content added by Netflix in year 2020 is of TV-MA maturity rating.

**References:**

1. [www.researchgate.net/publication/341558182\_Emergence\_and\_future\_of\_Over-the-top\_OTT\_video\_services\_in\_India\_an\_analytical\_research](http://www.researchgate.net/publication/341558182_Emergence_and_future_of_Over-the-top_OTT_video_services_in_India_an_analytical_research)
2. [www.kaggle.com/datasets](http://www.kaggle.com/datasets)