



Final Project - A Data-Driven Lens on OTT Platforms



By - Avni Gujral, Shagun Modi, & Noopur Joshi

M.S. Marketing Analytics & Insights, Rutgers Business School

Data Analysis and Visualization

Professor Michail Xyntarakis

May 13, 2025

A Data-Driven Lens on OTT Platforms

1. Introduction and Methodology

During the pandemic, the world saw a sudden rise in the number of OTT platforms, the content available, and the consumption patterns. Many companies saw a rise in revenues, users, and profits. Through this dataset, we aim to explore the different aspects of OTT platforms and their content. The goal is to employ a data-driven methodology to identify trends in regional presence, audience targeting, and content strategy.

The dataset used is fictional but resembles real-world OTT data and includes more than 22,000 movies from Netflix, Prime Video, Hulu, and Disney+. It includes attributes such as genre, language, age rating, country, release year, and IMDb and Rotten Tomatoes ratings. Basic preprocessing involved cleaning the data and splitting fields such as genres, and merging multi-label fields such as platforms and languages.

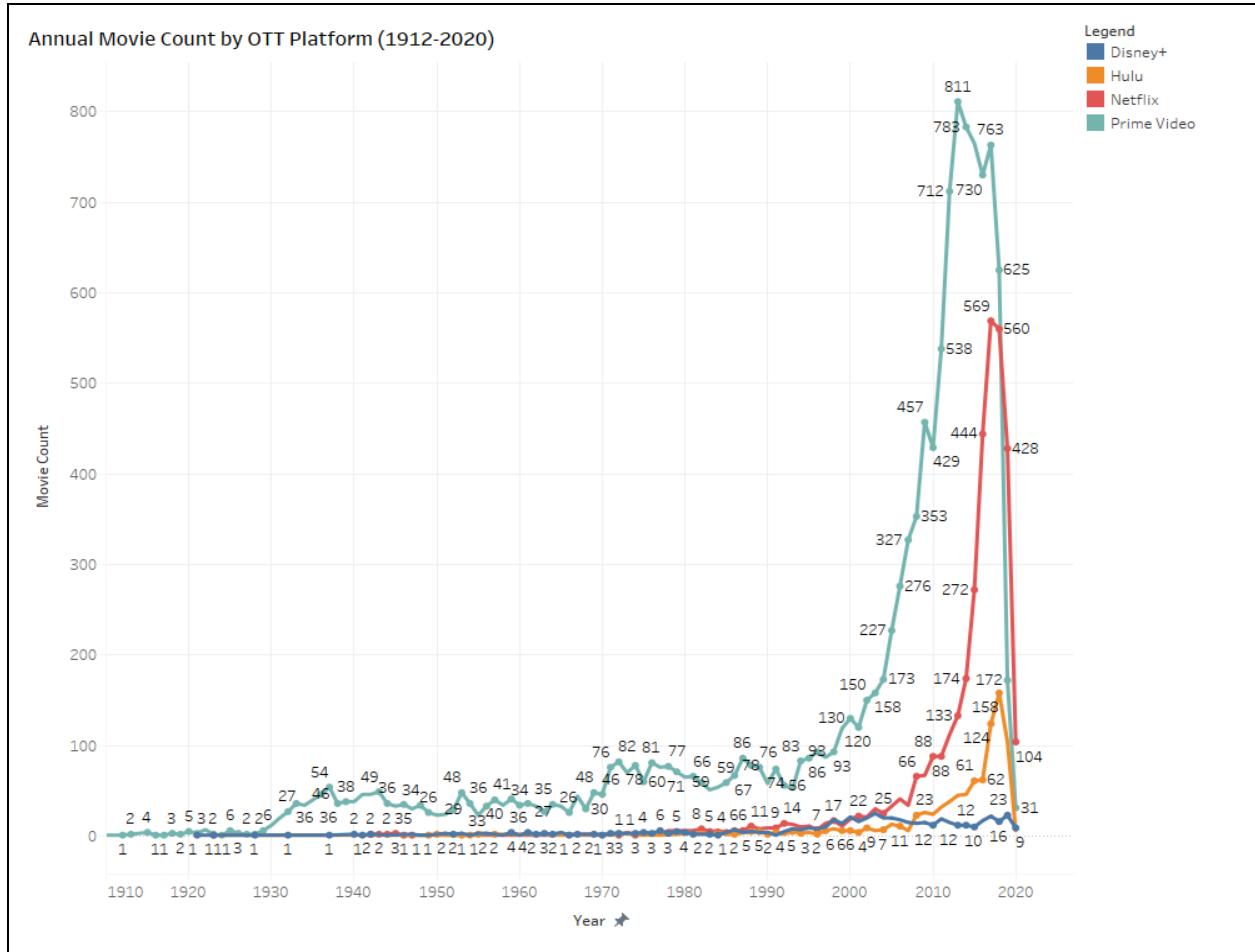
In our process, we have used Python for data processing and advanced visualisations and Tableau to build interactive dashboards and explore geographic and platform-wise trends. We have created visualizations such as line graphs, choropleth maps, bar charts, scatter plots, and t-SNE clustering.

Here is the raw data at a glance, with the columns in the dataset:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	ID	Title	Year	Age	IMDb	Rotten Tomatoes	Netflix	Hulu	Prime Video	Disney+	Type	Directors	Genres	Country	Language	Runtime
2	1	Inception	2010	13+	8.8	87%	1	0	0	0	0	0	Action,Adv	United Sta	English,Ja	148
3	2	The Matrix	1999	18+	8.7	87%	1	0	0	0	0	0	Sci-Fi	United Sta	English	136
4	3	Avengers: Infinity War	2018	13+	8.5	84%	1	0	0	0	0	0	Action,Adv	United Sta	English	149
5	4	Back to the Future	1985	7+	8.5	96%	1	0	0	0	0	0	Adventure	United Sta	English	116
6	5	The Good, the Bad and the Ugly	1966	18+	8.8	97%	1	0	1	0	0	0	Western	Italy,Spain	Italian	161
7	6	Spider-Man: Into the Spider-Verse	2018	7+	8.4	97%	1	0	0	0	0	0	Animation	United Sta	English,Sp	117
8	7	The Pianist	2002	18+	8.5	95%	1	0	1	0	0	0	Biography	United Kin	English,Ge	150
9	8	Django Unchained	2012	18+	8.4	87%	1	0	0	0	0	0	Drama,We	United Sta	English,Ge	165
10	9	Raiders of the Lost Ark	1981	7+	8.4	95%	1	0	0	0	0	0	Action,Adv	United Sta	English,Ge	115
11	10	Inglourious Basterds	2009	18+	8.3	89%	1	0	0	0	0	0	Adventure	Germany,I	English,Ge	153
12	11	Taxi Driver	1976	18+	8.3	95%	1	0	0	0	0	0	Crime,Dra	United Sta	English,Sp	114
13	12	3 Idiots	2009	13+	8.4	100%	1.01	0	1	0	0	0	Comedy,D	India	Hindi,Engli	170
14	13	Pan's Labyrinth	2006	18+	8.2	95%	1	0	0	0	0	0	Drama,Fai	Mexico,Sp	Spanish	118
15	14	Room	2015	18+	8.1	93%	1	0	0	0	0	0	Drama,Thr	Ireland,Ca	English	118
16	15	Monty Python and the Holy Grail	1975	7+	8.2	97%	1	0	0	0	0	0	Adventure	United Kin	English,Fr	91
17	16	Once Upon a Time in the West	1968	13+	8.5	95%	1	0	1	0	0	0	Western	Italy,Unite	Italian,Eng	165
18	17	Indiana Jones and the Last Crusade	1989	13+	8.2	88%	1	0	0	0	0	0	Action,Adv	United Sta	English,Ge	127
19	18	Groundhog Day	1993	7+	8	96%	1	0	0	0	0	0	Comedy,F	United Sta	English,Fr	101
20	19	The King's Speech	2010	18+	8	95%	1	0	0	0	0	0	Biography,United Kin	English	English	118
21	20	Her	2013	18+	8	95%	1	0	0	0	0	0	Drama,Roi	United Sta	English	126

2. Visualizations

Annual Movie Count Analysis based on OTT Platforms



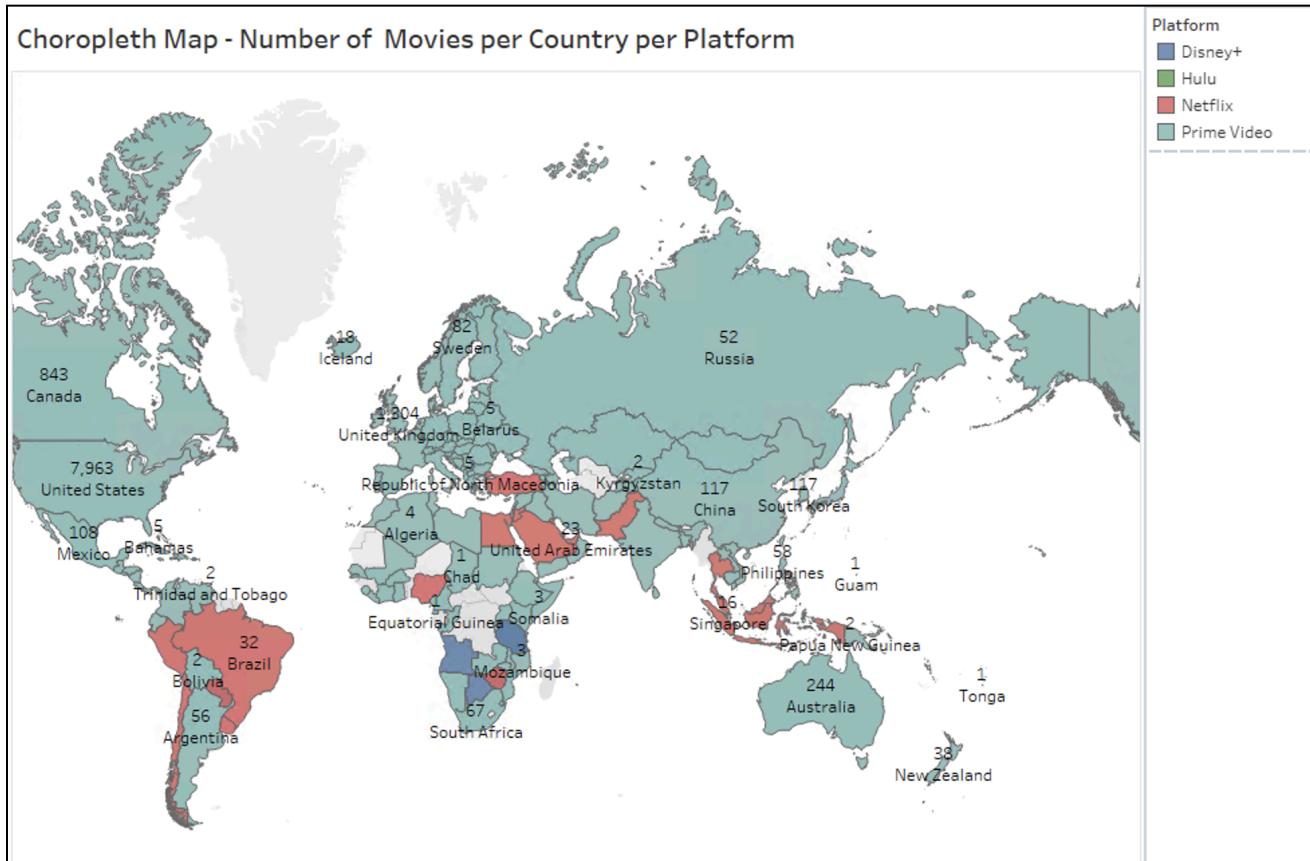
The line chart analyses the growth of each platform based on the annual movie count. Prime Video stands as the platform producing the most number of movies annually which is followed by Netflix. Disney+ and Hulu both cumulatively have a very low movie count annually.

Prime Video initially saw a growth in its annual movie count compared to the other platforms in 1931. The peak time for the platform was in the year 2013. Prime Video has consistently been one of the top platforms in releasing a large number of movies per year. This shows that the platform dominates the OTT platform based on the annual movie count.

Netflix, on the other hand, remains strong and releases a good amount of movies per year, however, falls short of Prime Video. Hulu and Disney+ show a very low count of movies annually, as they were introduced in the market much later than the other two platforms.

Content Distribution and Audience Targeting Across Platforms

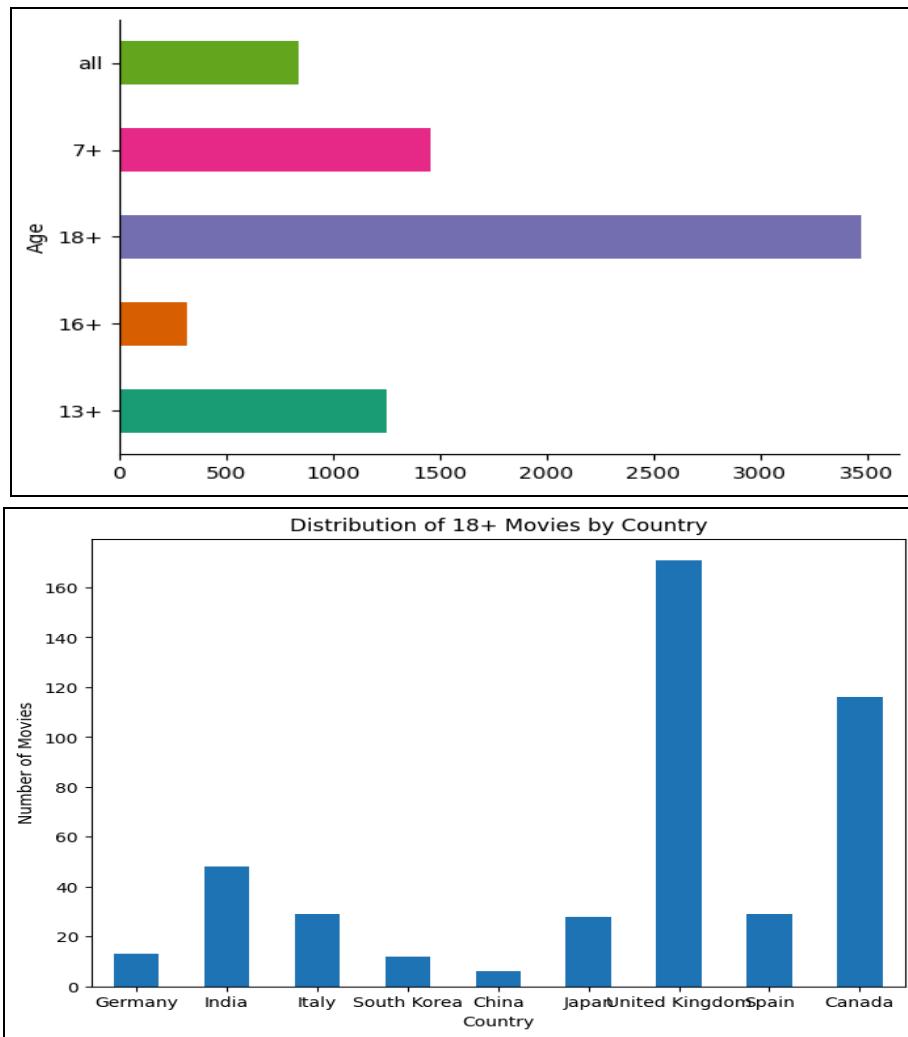
- World Map - Number of Movies Per Country Per Platform



This Choropleth chart displays the number of movies from each country listed on major OTT platforms. The graph shows that America tops the OTT content market with over 7,900 titles. This is not an unexpected lead, considering that all four platforms are based in America and invest significantly in American-based content. Canada and the UK also figure high because they have large English-language film and television industries. Countries like India, Australia, and China contribute significantly. India is a key market for Prime Video and Netflix, as they invest substantially in regional and Bollywood content. However, Africa and parts of South America have fewer titles due to smaller film industries or lower streaming demand.

Each platform also has regional strengths. Netflix is a global leader in localized content and is available in more than 190 countries, with multilingual support. Although Prime Video is accessible in more than 240 regions, the depth of its catalog varies heavily by country. Hulu has less of an international presence since it is mainly U.S.-based. Meanwhile, Disney+ is present in many countries but specializes in content from its core franchises, which lowers its regional content variety. The colour shading in the map corresponds to the dominant platform for each country. This color coding helps highlight which platform leads in content volume within each country.

- **Age Analysis- Distribution of the Movies By Age and Country**



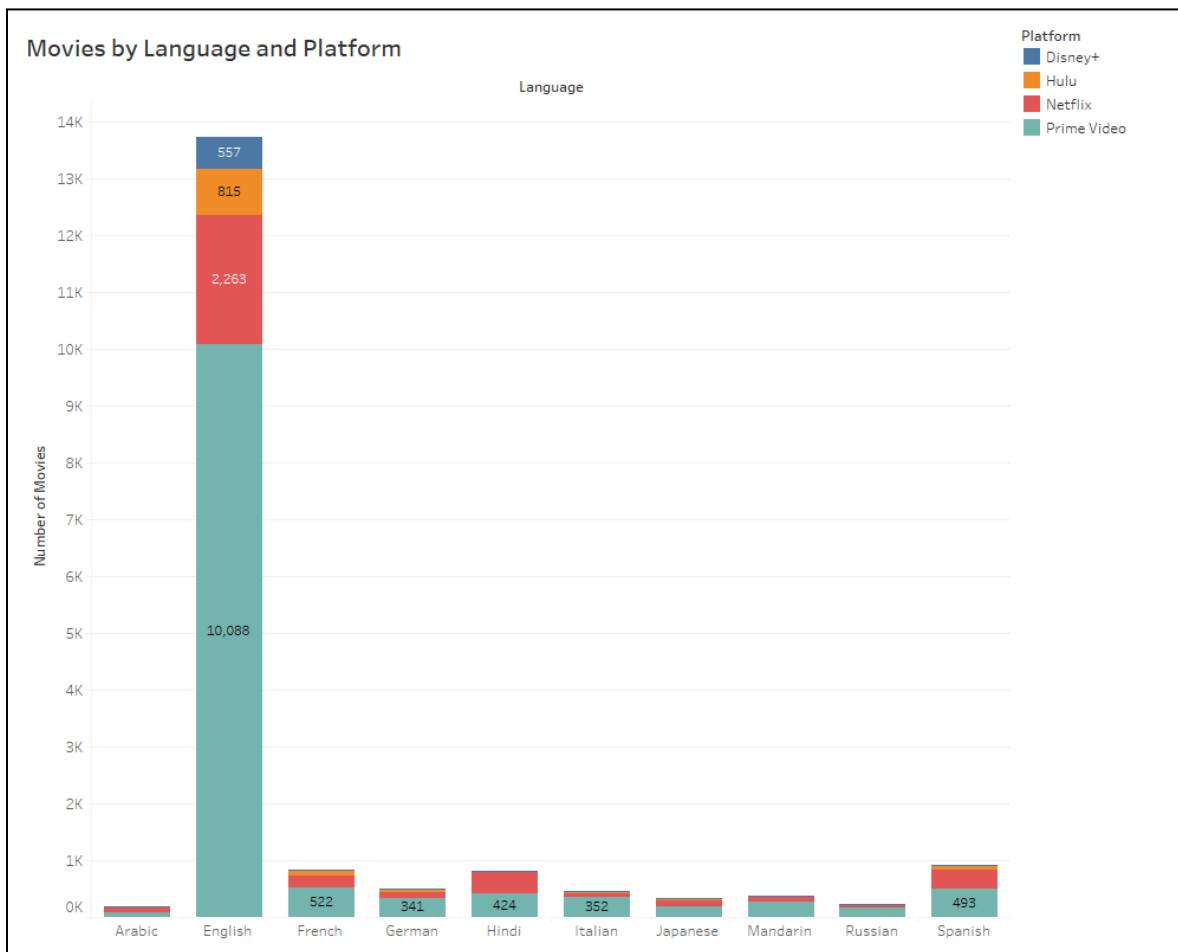
The above bar charts represent the distribution of movies by age rating. The leading category is the adult 18+ movies with around 3,500 movies. The second highest category falls in the kids category which is the 7+ aged movies, which show in the categories of family, drama, adventure, or comedy kind of movies.

Based on the geographics, the 18+ movies are primarily in the United Kingdom releasing over 170 categories. This is followed by Canada with around 120 categories, and India with 50 categories of movies. This shows that these regions focus on adult content. This could be a strategic decision by the platforms in order to generate more subscriptions, traffic, and viewing on their sites.

Meanwhile, China has the lowest number of 18+ movies airing across all the platforms. This could be due to stricter censorship laws. Additionally, in countries such as China and South

Korea, the OTT platforms are controlled by the government due to which it limits such categories.

- **Movies by Language & Platform**



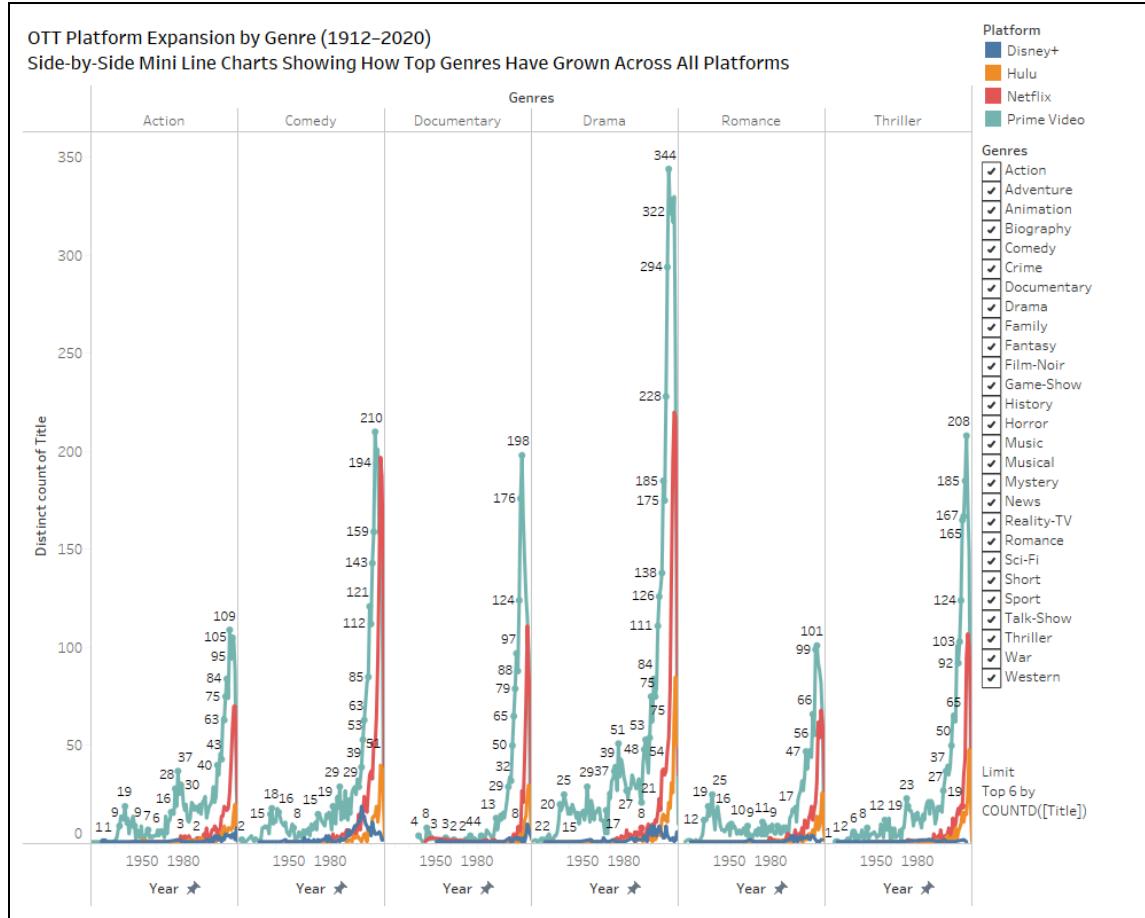
This stacked bar chart shows the number of movies available in different languages on the four major OTT platforms.

The majority of the content is in English, with Prime Video alone providing 10,088 English movies, followed by Netflix, Hulu, and Disney+. Among non-English languages, French, Spanish, Hindi, and German are the most widely supported. Netflix and Prime Video contribute the most to multilingual content, with many movies in these languages. Hulu and Disney+ offer comparatively limited content in non-English languages.

Prime Video leads in language variety and volume, making it a strong player in international content. Netflix also showcases a commitment to global reach by having content available in multiple languages.

Genre

- **OTT Platform Expansion Across Genres Evolution Over Time**



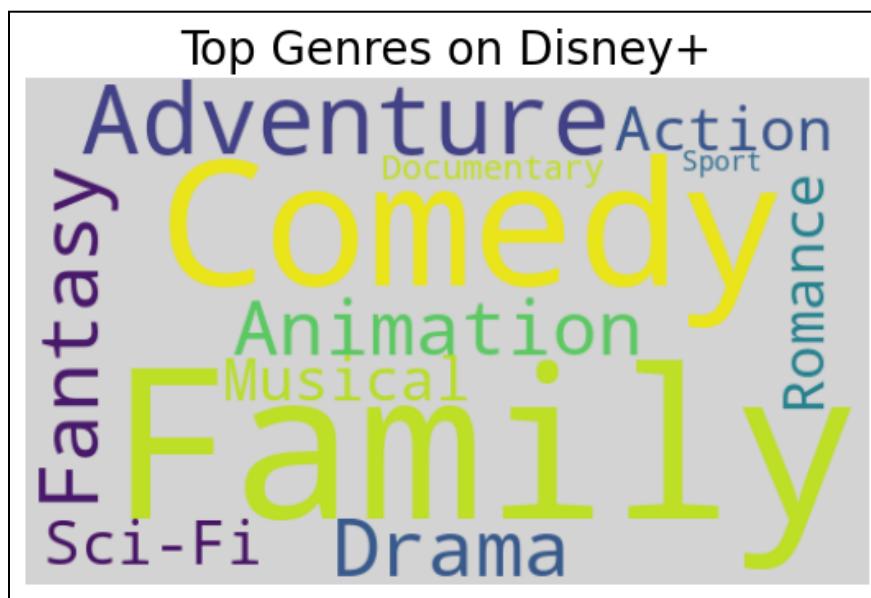
This graph illustrates how leading OTT platforms have expanded their content across different genres from 1912 to 2020. Each genre has its mini-line chart with separate lines for Netflix, Prime Video, Hulu, and Disney+, making it easy to compare how each has evolved.

Drama and comedy have consistently been the most produced genres, with a substantial increase starting in the 2000s. In particular, Netflix and Prime Video display the most significant increases, particularly after 2010, which suggests they are actively expanding their content. Prime Video, for example, leads in producing documentaries and thrillers, while Netflix has a

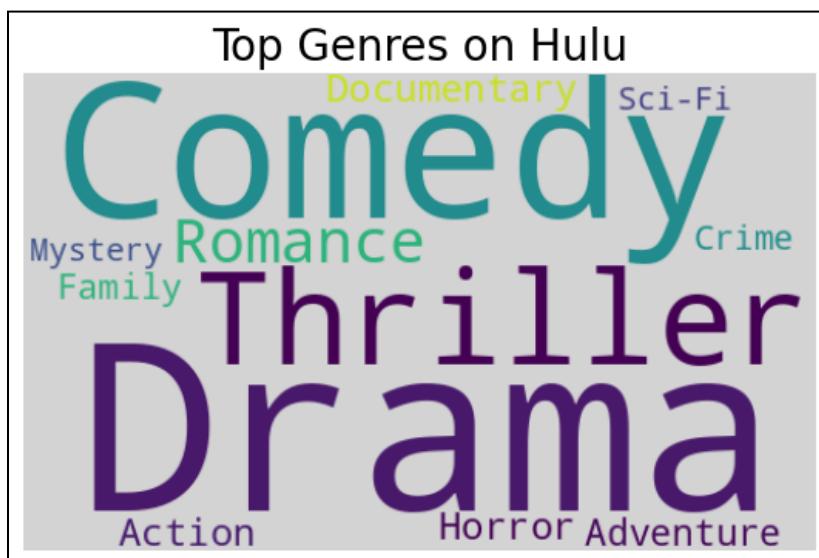
strong presence in nearly all genres. Disney Plus and Hulu make more modest contributions, with Disney Plus prioritizing family-friendly genres.

The chart generally shows how platforms prioritise different genres and how content production has increased significantly in the last two decades. This makes it a powerful visual for understanding platform focus and genre strategy in the streaming space.

- **Most Common 12 Genres**



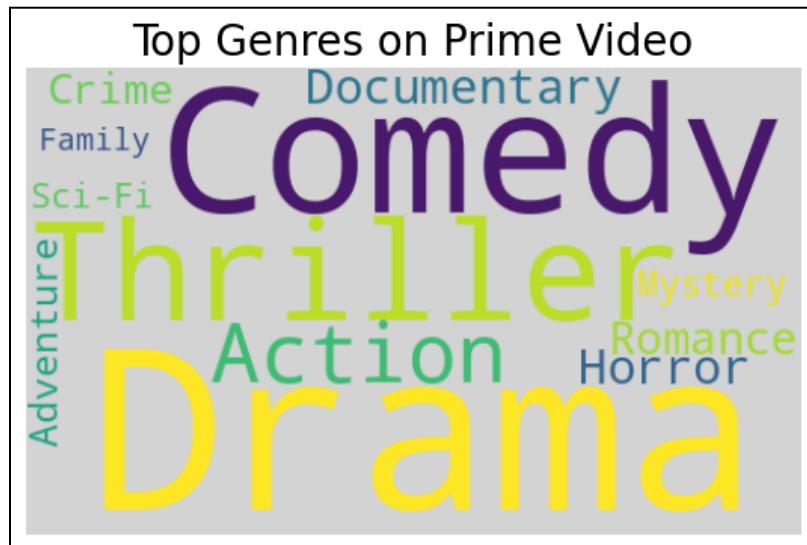
The common genres on Hulu are: Comedy, Thriller and Drama, and some of the less common genres are: Horror, Crime and Action.



The common genres on Netflix are: Comedy, Thriller and Drama, and some of the less common genres are: Horror, Family and Mystery.



The common genres on Prime Video are: Comedy, Thriller and Drama, and some of the less common genres are: Sci-Fi, Adventure and Documentary.



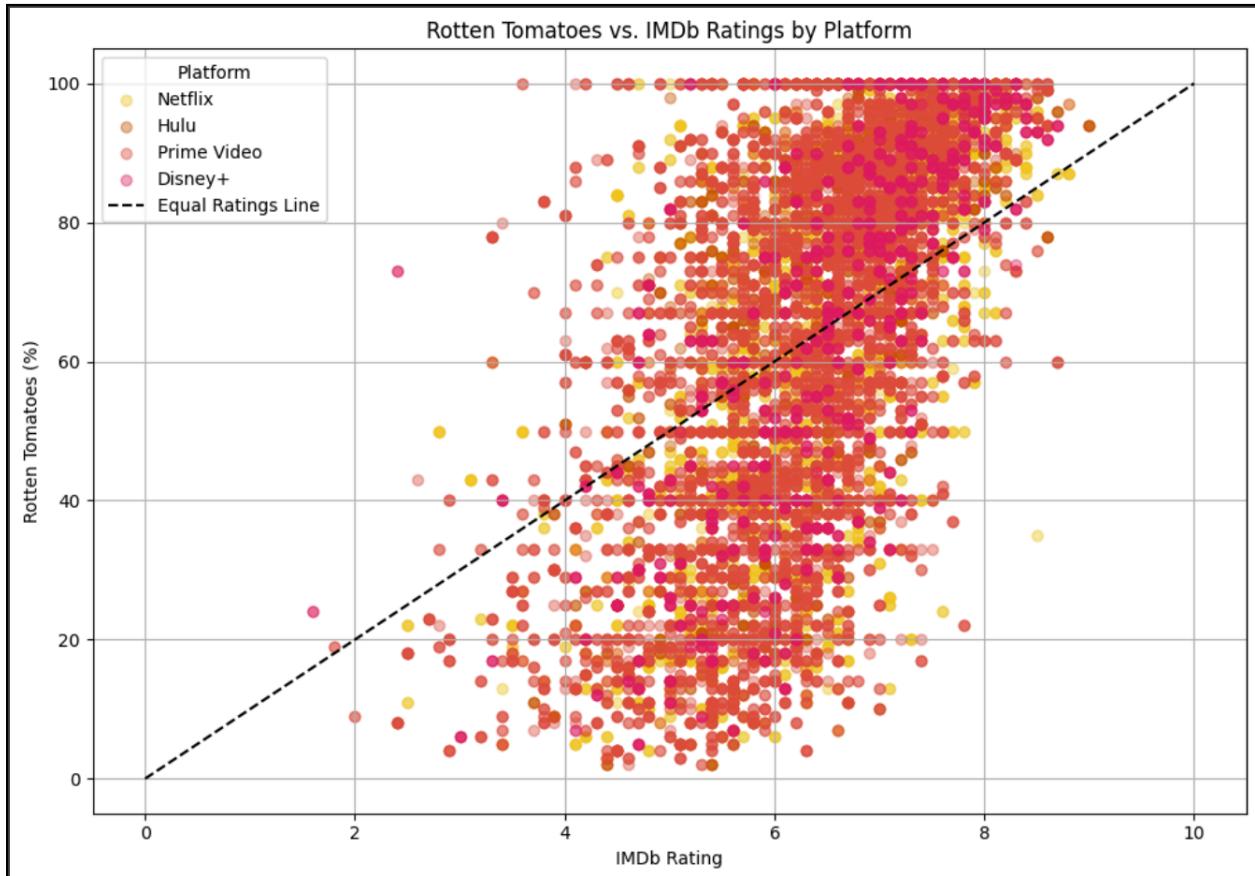
As part of a deep dive into the genres, the team created word clouds for each platform to visualize genre frequency. A consolidated word cloud to highlight the 12 most common genres across platforms in terms of how frequently they repeated.

To make this visualisation in its most accurate form - the genres were split into strings using the **str.split (',')** command. This helps in segregating a mix bag into multiple buckets.

The main idea of a word cloud is to show the frequency, hence bigger the font size, the more common that particular genre is.

Evidently so, the most common themes on Disney+ are: Adventure, Comedy and Family, and lesser ones were Sport, Action or Romance.

Platform Ratings



Prime Video and Disney+ mainly lie above the line which indicates Rotten Tomatoes rating, signifying that these platforms favor the critically acclaimed content more than that of the audience rated content. These platforms have more plots towards the Rotten Tomatoes region ranging from 40% to 100% than the IMDB region which ranges from 4 to 6 rating suggesting the critics rate the content much higher than the audience.

Many ratings from the platforms Netflix and Hulu lie in the IMDB region suggesting that the platform releases content that caters more to the audience than the critics, perhaps from analyzing the previous reviews of similar content or genre. For these two platforms, the audience has rated the content higher than the critics.

The most noticeable evidence of the difference between the critic reviews vs audience reviews is seen on the top horizontal line where the critics reviewed some of the content as 100%, which is quite high. However, the IMDB ratings fall mostly between 5 to 8, which shows that the audience thought it was alright but not quite as great as the critics claimed. This highlights the discrepancy between the two reviews.

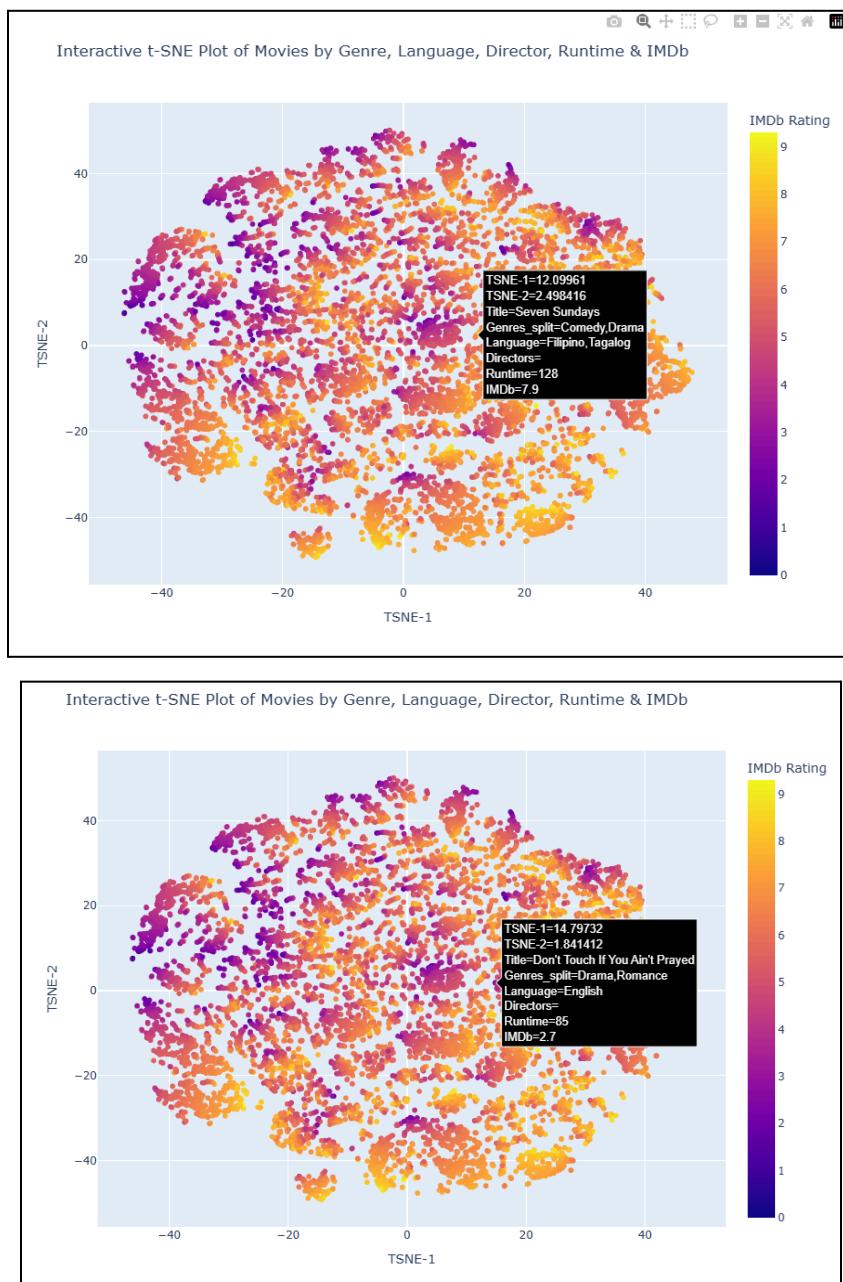
tSNE - Similarity Analysis



This plot is a tSNE plot for the similarity analysis of each movie on the platform, categorized by multiple factors, first the features were run using the one-hot encoding for the categorical data, then scaled the numerical data, then using the PCA method for bringing the features down to 2D, and the IMDb rating as a means to color code the cluster for better visual understanding. For the PCA method multiple features were combined such as the genre, directors, language, runtime. The categorical data was made into numerical, and standardized using the scaler to make sure that when the clusters are created no feature is overpowering

The tSNE plot works to preserve the local similarity, so movies that are together are similar in the features used.

Each dot in this plot represents a movie from the dataset. As the legend mentions, these movies are coloured-coded on IMDb ratings, so movies with each 2 point bracket move to a better rating, the lightest green dots represent the highest ratings.



Furthermore, in pursuit to make this visualisation interactive, this plot was created (using) a few lines of code. Hovering over any point will give details about the movie. The data was reduced to use only 50 directors in order to avoid thousands of columns.

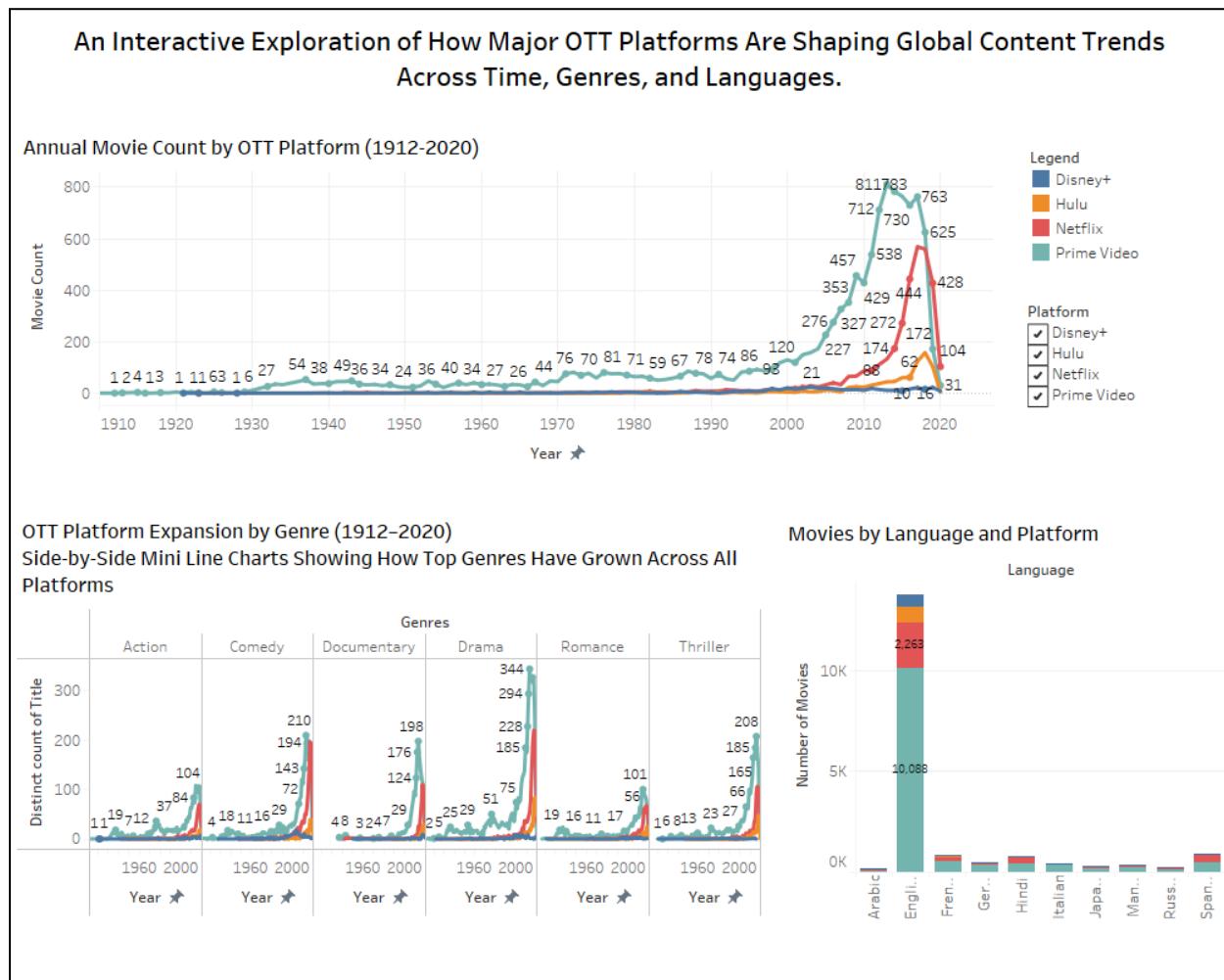
From the colour scaler it is evident that lower scored movies are coded darker shades in the interactive plot.

When trying to identify the clusters it is evident that:

- High IMDb, genre rich movies are clustered together (bright yellow spots)
- Low IMDb, niche// poorly rated content are frills of dark pink-purple areas

It is important to note that IMDb is not defining the clusters, it's the features that are defining, IMDb just helps in indicating the quality (ratings) concentration in the regions of the plot.

Interactive Dashboard



The dashboard shows the interactivity between all the platforms- Disney+, Hulu, Netflix, and Prime Video. The aim of the dashboard was to show all the elements including the annual movie count, genre expansion for each platform through the years, and the languages that the content is available in on all the platforms without having to navigate through them individually. The dashboard eases the process of making a thorough analysis by navigating smoothly through the platform filters to analyze each platform individually. Additionally, the color legend remaining consistent among all the charts allows for a smoother transition.

The dashboard also allows us to create a connection among all the charts. For example, in the movie count chart, Amazon Prime has the highest number of movies and these are typically in the thriller genre in the English language, as seen by the most recent years. This information can later be used to drill down even more data by researching deeper into it.

3. Conclusion

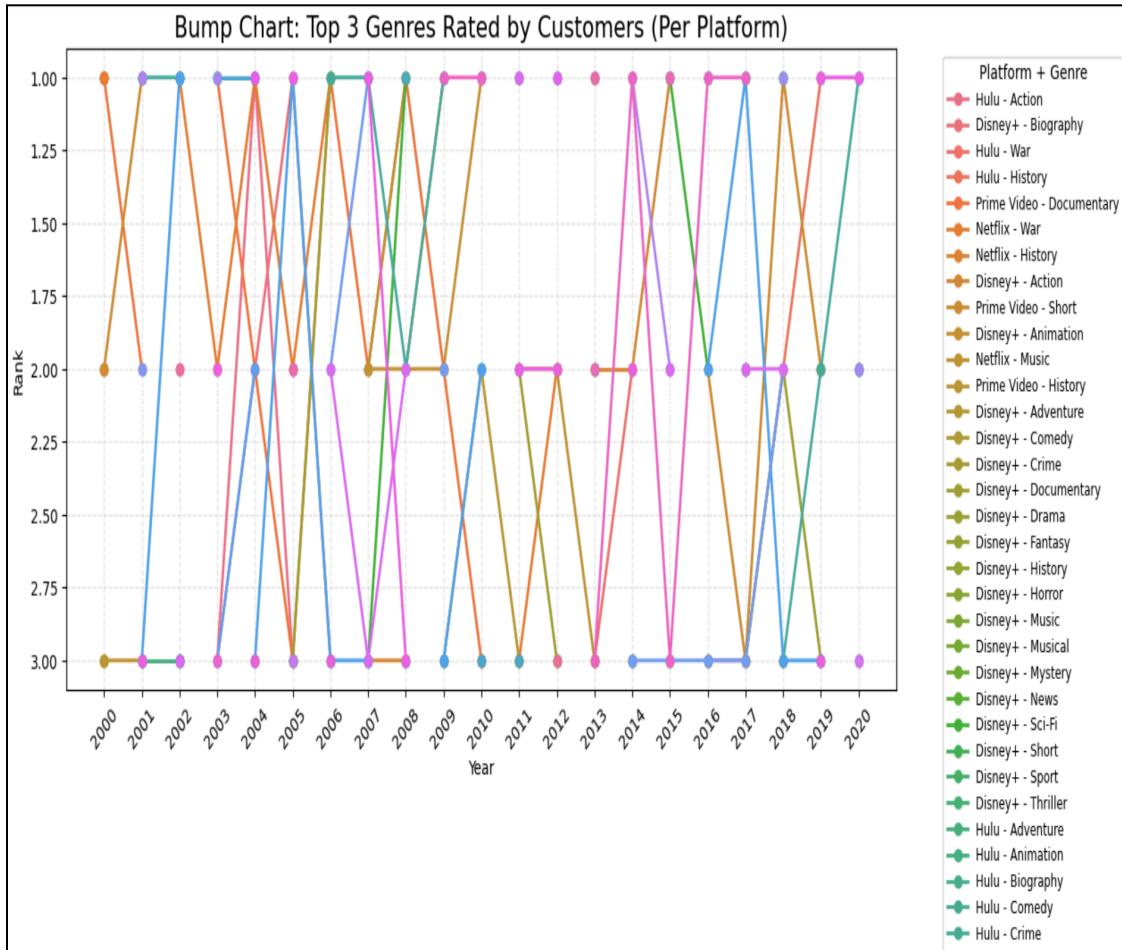
Through this comprehensive analysis through various chart types, we have discovered various patterns and interesting elements from the dataset. The dataset compares the platforms Netflix, Prime Video, Hulu, and Disney+ across the content, country, age, language, genre, ratings, and similarity analysis. The report clearly highlights the unique positioning and strategies of the different platforms across all the various categories.

Whether it be the genre or the annual movie count, the overall analysis of the platforms shows the revolution of the platforms over time as well as the revolution of the OTT industry through the years. It shows how the OTT platforms are shaping and responding to global entertainment consumption patterns.

Additionally, the use of a dashboard also allows us to make a comparative analysis to examine which platforms are doing the best and the ones that still have scope to improve and create a significant impact on the market.

As the OTT industry is becoming more and more competitive due to heavy consumer consumption, it is very important to conduct such visual and analytical approaches to gain insights and make data driven and informed decisions and stay ahead of the competitive market.

4. Appendix (Additional Chart)



This bump chart aimed to analyze the top three genres that generated the most traffic on the platforms.

However, this chart was not successful as the dataset did not contain any information on traffic, popularity, user behavior, or the streaming activity which makes it difficult to visualize the top rated genres. However, we used the IMDb ratings in order to attain the same.

Although we were partially able to achieve our vision, it was quite difficult to interpret the insights from the chart as the dots connecting the lines do not show the same color signifying that they have been incorrectly loaded despite several efforts. Additionally, the legend list as well is quite long which makes it difficult to analyze the chart.