

# Streaming web industry

Business Statistics and Data Modelling

Jade Magbanua

MANUKAU INSTITUTE OF TECHNOLOGY BACHELORS IN DIGITAL TECHNOLOGY



## Contents

DATA UNDERSTANDING .....	2
Datasets: .....	2
Descriptive Analysis .....	2
Prescriptive analysis.....	6
Challenge .....	6
Solution .....	6
Time Series .....	7
Conclusion .....	9
<b>REFERENCES:</b> .....	10

## DATA UNDERSTANDING

- Before doing any pre-analysis for the report, we have to understand the data that we are going to use first for this report.
- The dataset that I will be analysing is the streaming websites or software in terms of shows that are available only in certain streaming platforms.
- I used 2 data for the analysis merging those 2 data can create more analytical process for the report meaning that having more data can led to a more accuracy for the statistical processing analysis of the dataset.
- These 2 data came from Kaggle and github.

Datasets:

Name	Date modified	Type	Size
 MoviesOnStreamingPlatforms_updated	30/06/2023 6:25 pm	Microsoft Excel C...	1,898 KB
 tv_shows	30/06/2023 5:36 pm	Microsoft Excel W...	342 KB

## Descriptive Analysis

- Descriptive analysis is a way of examining and summarizing data to understand its key characteristics. It involves organizing, describing, and presenting data in a meaningful manner, without making any statistical inferences or drawing conclusions beyond what the data itself reveals.
- To create this type of analysis, I used python to code and merge both data for further analysis.

```
In [45]: r1 = df_movies.sort_values(by='IMDb', ascending=False)
r1 = r1[0:10]
r1['Movies'] = 'Movies'
fig = px.treemap(r1, path=['Movies', 'Title', 'Genres', 'Language'], values='IMDb', color='IMDb', color_continuous_scale='Purp')
fig.show()
```



- 
- Python code:
  - `r1 = df_movies.sort_values(by='IMDb', ascending=False)`

```

r1 = r1[0:10]
r1['Movies']='Movies'
fig = px.treemap(r1, path=['Movies','Title', 'Genres','Language'],
values='IMDb',color='IMDb',color_continuous_scale='Purp')
fig.show()

```

- IMDb is a rating website where critics rate certain movies for future viewers and the producers of a certain movie.
- This helps directors and producers to get feedbacks from the people that wants to watch the movie they produced.
- This graph tells us that there are certain preferences that some people loves, in this context it shows that the highest rating is approximately 10 and one of the most picked or best rated movies are in genres of action, adventure and comedy. Their preferred language as well is English but we have to consider that language is rather based on the culture of the person who is watching the movie.
- These types of scenario tell us that there is also a biased side when it comes to rating a movie or a tv show it might be because of the size or population of the critiques only watches English movies and not different languages. So, this means that in terms of critiques we have to consider options and preferences not just based on the language but a true critic that will consider genre, statistical survey, movie content, as well as acting of the actors from the movie.

- Python code

```

○ df_t = df_tv.copy()
df_t = df_t[df_t['Age'].notna()]
df_t['Age']=df_t['Age'].str.replace('+',"")
df_t['Age']=df_t['Age'].str.replace('all','0')
df_t['Age']=df_t['Age'].astype(str).astype(int)
def barplot(i,dataframe,platform,p):
    plt.subplot(i)
    dataframe=dataframe.loc[dataframe[platform] == 1]
    dataframe = dataframe['Age'].value_counts().reset_index()
    dataframe.columns=['age','count']
    sns.barplot(x="age", y="count", data=dataframe,palette=p)
    plt.xlabel('Age', fontsize = 15)
    plt.ylabel(platform, fontsize = 15)

```

```

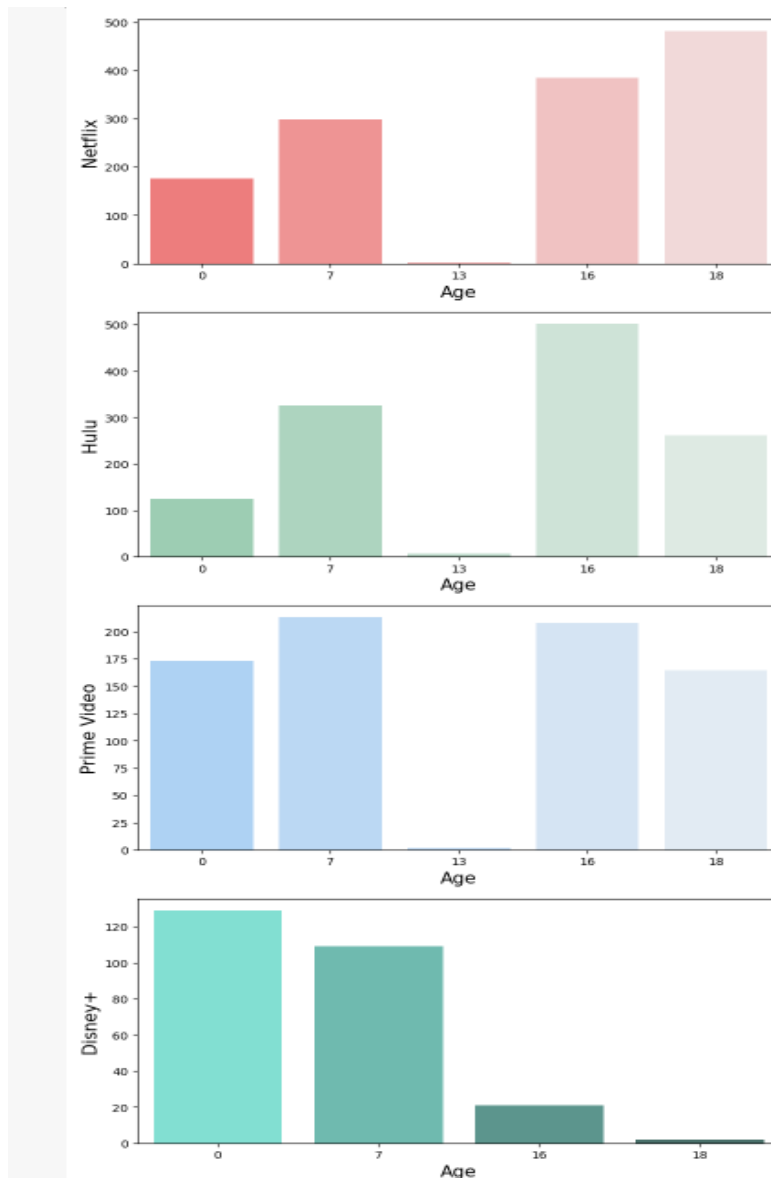
plt.figure(figsize = (20, 20))

```

```

barplot(421,df_t,'Netflix',netflix_p)
barplot(423,df_t,'Hulu',hulu_p)
barplot(425,df_t,'Prime Video',prime_p)
barplot(427,df_t,'Disney+',disney_p)

```



- Another form of a descriptive analysis are the bar charts. In this context we have age restrictions and the streaming platforms that are being used for this report.
- We can see that a large quantity of tv shows that are rated for 18+ people came from Netflix. Which means that these kinds of tv shows or movies are only exclusive for people that are aged from 18 and above.
- This also applies that parental guidance is heavily required for young people to use this type of platform.
- We have to consider an age restriction control or parental control to restrict kids or young people from accessing these kinds of movies and tv shows.
- This also shows that Netflix is not quite a family friendly platform which means that in terms of adult people that are aged 18 above can enjoy this type of platform.
- Speaking of Family friendly platforms Prime video and Disney+ is a great platform for a family to watch from, especially Disney+ since most of its movies are cartoon based this will give the relief for the family to use this platform for the entire household.
- Python code:

```

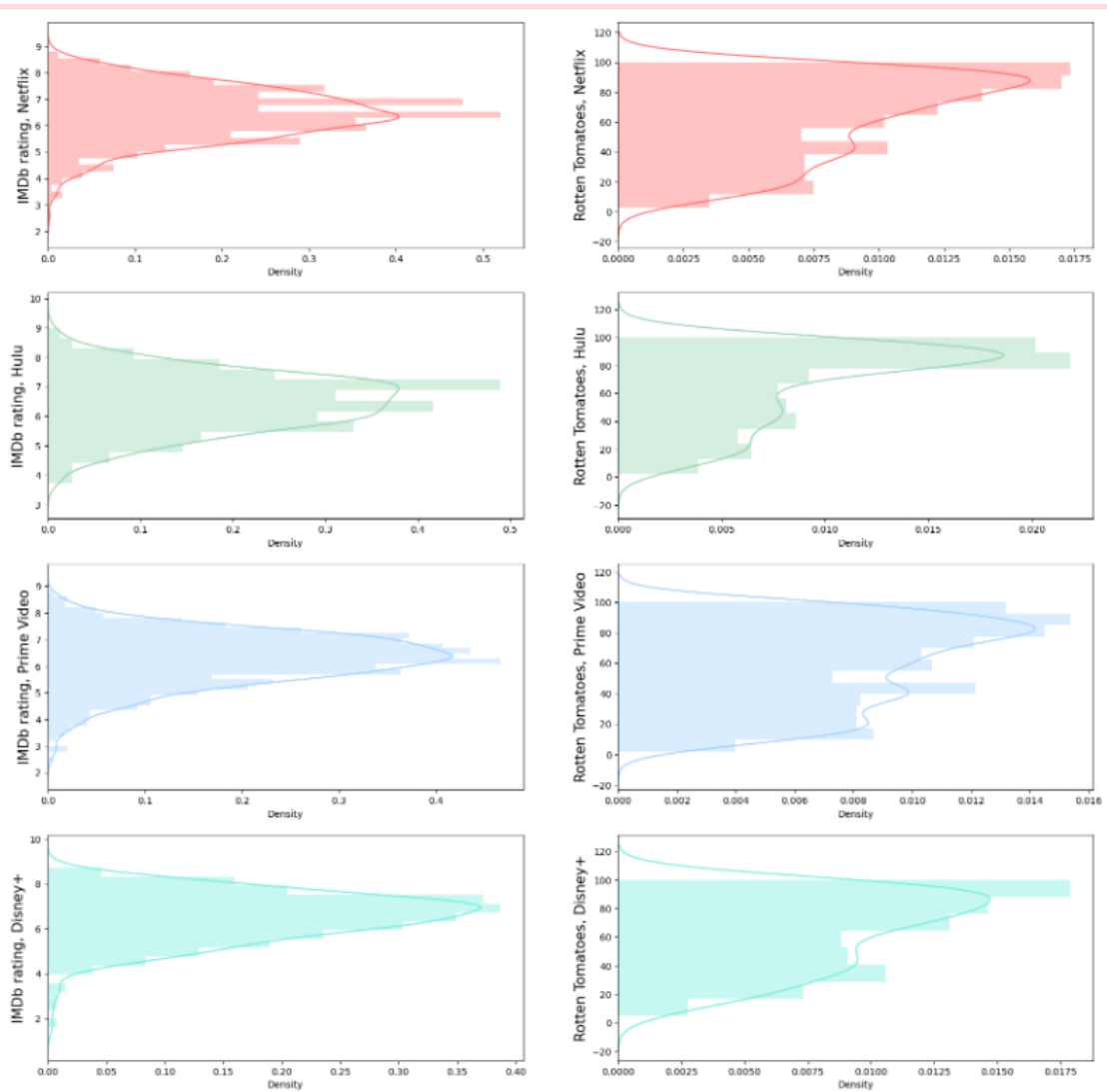
○ def dist(i,dataframe,platform,c):
    plt.subplot(i)
    dataframe=dataframe.loc[dataframe[platform] == 1]
    sns.distplot(dataframe['IMDb'], color=custom_colors[c],vertical=True)
    plt.ylabel('IMDb rating, '+platform, fontsize = 15)
    plt.subplot(i+1)
    sns.distplot(dataframe['Rotten Tomatoes'],
    color=custom_colors[c],vertical=True)
    plt.ylabel('Rotten Tomatoes, '+platform, fontsize = 15)
    plt.figure(figsize = (20, 20))

```

```

dist(421,df_m,'Netflix',0)
dist(423,df_m,'Hulu',1)
dist(425,df_m,'Prime Video',2)
dist(427,df_m,'Disney+',3)

```



- This graph compares rating of platforms using 2 famous critics websites. IMDb and Rotten tomatoes are one of the most popular websites that mostly critics movies and tv shows.
- This graph shows different areas between the 4 platforms which is not really too much of a big deal since rating is not the one that drives the streaming company. It helps but it's still dependent based on the user and the subscribers for the streaming web or application.

## Prescriptive analysis

- Prescriptive analysis is a type of data analysis that goes beyond descriptive and predictive analysis to provide recommendations or actions to optimize outcomes or solve problems. It takes into account various factors and constraints to suggest the best course of action.
- This type of analysis creates the most important thing for a business which is creating a possible solution based on the descriptive analysis and predictive analysis results.
- In this context we have to give possible business challenges that are based from this data and give a potential solution for these challenges.

## Challenge

- There is an existing competition between multiple platforms in terms of users, movie and tv show productions, monthly subscription prices, content, platform-based feature and region-based preferences (Region exclusive movies and tv shows)

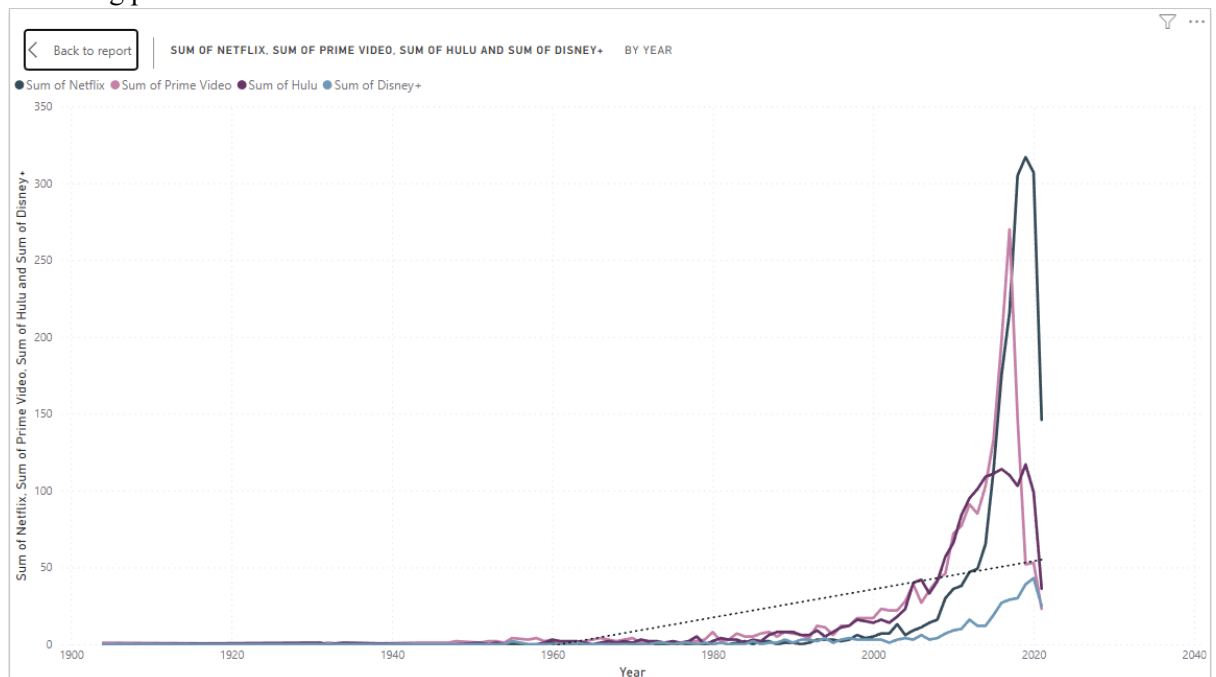
## Solution

- Based on the data that I am using for this report, we can consider stages of solutions if a certain platform needs to act on these factors that might affect sales for the company.
  - To enhance user experience on the streaming platform, we will improve the user interface for easy navigation and seamless content discovery. We will also implement personalized recommendations based on users' viewing history and preferences, introduce user profiles and multiple device synchronization for a personalized experience across devices, and offer offline viewing options to allow users to download content and watch it later without an internet connection.
  - To enhance content library, acquire licensing agreements with studios, produce original exclusive content, curate diverse genres, and update regularly.
  - To improve your pricing strategies, conduct market research to understand what your target audience expects and what your competitors are charging. Offer different subscription plans with various features and prices to accommodate different users. Consider offering a free trial or limited access version to attract new users. Continuously monitor user feedback and adjust your pricing strategy to balance affordability and value.
  - To enhance user experience on a streaming platform, introduce social and interactive features, advanced search filters, and high-quality streaming technology.

- To cater to different regions, do market research to understand their preferences. Offer content in multiple languages and collaborate with local providers. Personalize recommendations and promotions to increase user satisfaction.
- Promote the streaming service through targeted marketing campaigns, social media, online ads, and influencer partnerships. Collaborate with streaming device, smart TV, and telecom companies to pre-install or promote the service. Offer special deals or discounts with partner companies to attract new users and encourage subscriptions.
- To continuously improve a streaming platform, gather user feedback through surveys, ratings, and reviews, monitor viewing patterns, and stay up-to-date with emerging technologies and industry trends to introduce new features and innovations.

## Time Series

- The data produced a significant time series data. These data are a good comparison on which streaming platform has the most tv shows as well as movies.



- Based on the time series graph, Netflix tv show has a really great number of tv shows and movies, these data show the growth of some platforms starting from year 2000 then continuously increase until 2020.
- This graph shows a plot line indicating that these platforms continuously increase yearly and since this growth is also based on the production of the movies and its qualities.
- This graph has been produced from power BI.

Python code:

```
post_1980 = df_movies[df_movies.Year >= 1980]
year_movies = post_1980.groupby('Year')[['Netflix','Hulu','Prime Video','Disney+']].sum()
plt.figure(figsize = (16, 8))
sns.lineplot(x=year_movies.index,y=year_movies['Netflix'])
```



```

sns.lineplot(x=year_movies.index,y=year_movies['Hulu'])
sns.lineplot(x=year_movies.index,y=year_movies['Prime Video'])
sns.lineplot(x=year_movies.index,y=year_movies['Disney+'])
plt.xlabel('Release Year', fontsize = 15)
plt.ylabel('Count', fontsize = 15)
plt.show()

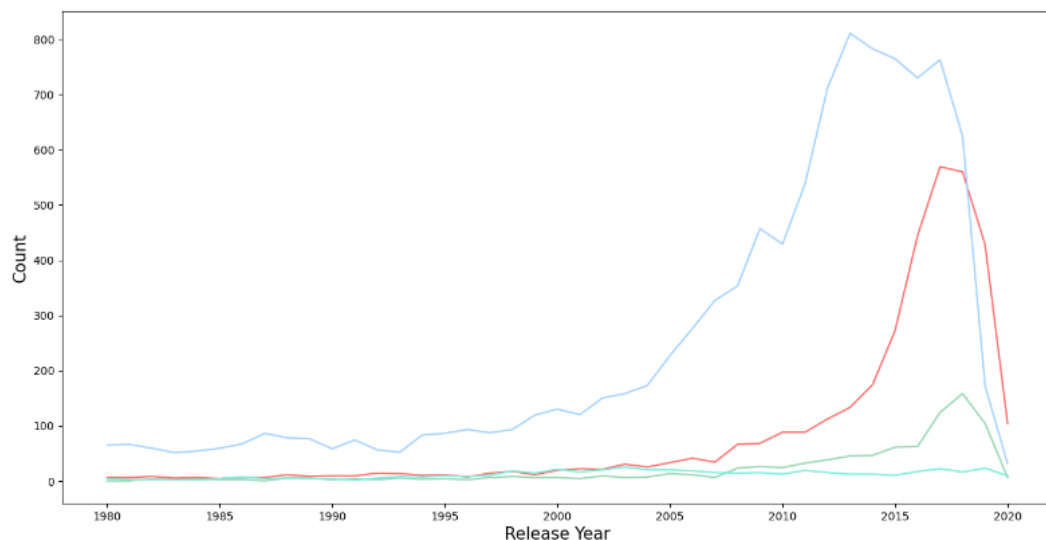
```

```

In [37]: post_1980 = df_movies[df_movies.Year >= 1980]
year_movies = post_1980.groupby('Year')[['Netflix','Hulu','Prime Video','Disney+']].sum()

plt.figure(figsize = (16, 8))
sns.lineplot(x=year_movies.index,y=year_movies['Netflix'])
sns.lineplot(x=year_movies.index,y=year_movies['Hulu'])
sns.lineplot(x=year_movies.index,y=year_movies['Prime Video'])
sns.lineplot(x=year_movies.index,y=year_movies['Disney+'])
plt.xlabel('Release Year', fontsize = 15)
plt.ylabel('Count', fontsize = 15)
plt.show()

```



- This graph from python is not as far from power BI although this graph shows significant detailed data compared to power BI
- In this scenario, we can see a data incompatibility, meaning that both developed graphs are not quite similar and I think this is because of the data set since I only used 1 table for the power BI and 2 datasets in python.
- This issue needs a comparative research to plot data lines that can help us choose which graph is more accurate.
- Based on my research, both platforms are not that far in terms of users, but Netflix is quite the largest competitor in terms of movies and tv shows compared to prime video. According to one of the articles comparing these 2 platforms “Netflix has 209 million users, as of July, while over 175 million of Amazon Prime Video's 200 million+ users had streamed content within the past 12 months leading up to April 2021.” (Roberts, 2021)
- Both of these platforms are quite significant towards the public since the streaming industry is pretty popular due to other factors as well, such as the lockdown that happened, forced people to stay in their house which led to watching tv shows and movies from these streaming platforms resulting to an increase of users.

## Conclusion

- This database is pretty helpful for data analysis, since most of these data are very relevant for the customers or users.
- Streaming industries are still businesses so the sales and income are still one of the top priorities from these companies.
- Creating these models and analysis shows different insights in terms of dealing a company problem. We have to consider different factors that affects these changes as well as the setbacks that the company makes. Considering that there are thousands of data that has been gathered from these streaming platforms, we can create new sales logic and offer different perspectives that may or may not benefit the company.

## REFERENCES:

*Delivering actionable, objective insight to executives and their teams.* (n.d.).

Gartner. <https://www.gartner.com/en>

*Forrester.* (2022, September 22). Forrester. <https://www.forrester.com/bold>

(n.d.). Harvard Business Review - Ideas and Advice for Leaders. <https://hbr.org>

*Journal of marketing.* (2023, June 28). American Marketing

Association. <https://www.ama.org/journal-of-marketing/>

(n.d.). Kaggle: Your Machine Learning and Data Science Community. <https://www.kaggle.com>

(2023, June 28). Motion Picture Association. <https://www.motionpictures.org>

Roberts, S. (2021, August 21). *Netflix vs Amazon Prime video: Which streaming service is winning*

*in 2021?* TechRadar. <https://www.techradar.com/news/netflix-vs-amazon-prime-video-2021>

*What is prescriptive analytics? 6 examples.* (2021, November 2). Business Insights

Blog. <https://online.hbs.edu/blog/post/prescriptive-analytics>