

# Milestone 1 - Video streaming platforms

## Dataset

We found multiple datasets on Kaggle that consisted of the main video streaming providers; Netflix<sup>1</sup>, Amazon Prime<sup>2</sup>, Disney+<sup>3</sup>, and Hulu<sup>4</sup>. The preprocessing part is fairly simple as the quality of the dataset is quite good. However, there are some of the rows that consist of empty cells. In other words, lack of information. If time permits, we plan to add information from the IMDb database to those datasets to get information about ratings and genres.

## Problematic

- What am I trying to show with my visualization?
  - The amount of content each platform contain
  - Ratings
  - Amount of seasons
  - Frequently new releases
  - Find Actors/Directors focusing on certain streaming services
- Think of an overview for the project, your motivation, and the target audience.
  - People that want to decide which streaming platform they should choose
  - Give information about which streaming service offers the best value
  -

Our work should

## Exploratory Data Analysis

- Show some basic statistics and get insights about the data

## Related work

Kaggle provides a feature to share and search code linked to a dataset. We leverage this feature to find previous work. With Netflix being in business for a long time, it is no surprise that there is a lot of work done with the dataset. Most of them are indeed concerned with visualization. However, when moving on to the other datasets, much less projects leverage them. The Hulu set has only 2 code projects available. It needs to be mentioned that there has been work done on the Disney+ dataset to compare it with Netflix. However, our goal is to compare all four of them to find the one streaming service to rule them all. Some codes<sup>5</sup>

---

<sup>1</sup> <https://www.kaggle.com/datasets/shivamb/netflix-shows>

<sup>2</sup> <https://www.kaggle.com/datasets/shivamb/amazon-prime-movies-and-tv-shows>

<sup>3</sup> <https://www.kaggle.com/datasets/shivamb/disney-movies-and-tv-shows>

<sup>4</sup> <https://www.kaggle.com/datasets/shivamb/hulu-movies-and-tv-shows>

<sup>5</sup> <https://www.kaggle.com/code/subinium/storytelling-with-data-netflix-ver>

have really inspiring visualisations done with one dataset. However, our goal would be to focus on visualizations which compare the services directly instead of putting the same visualizations for all providers side by side.

The fifth dataset<sup>6</sup> has been leveraged to do some side to side comparison<sup>7</sup>. However, as mentioned, we'd like to focus on doing single visualisations comparing the services instead of putting facts & figures side by side. Additionally, the visualisations we've found are fairly static. We'd like to make our website more adaptive allowing the user to get info about his favourite actors or about his most beloved years.

Of course, there are websites trying to answer the question which streaming service will be the best. However, as far as we can tell, all of them seem to focus on doing a textual comparison without visualization. Our goal is to go the opposite. Compare them visually and interactively with text to explain the visualizations if needed.

- What others have already done with the data?
- Why is your approach original?
- What source of inspiration do you take? Visualizations that you found on other websites or magazines (might be unrelated to your data).
- In case you are using a dataset that you have already explored in another context (ML or ADA course, semester project...), you are required to share the report of that work to outline the differences with the submission for this class.

---

<sup>6</sup> <https://www.kaggle.com/datasets/ruchi798/movies-on-netflix-prime-video-hulu-and-disney>

<sup>7</sup> <https://www.kaggle.com/code/ruchi798/movies-and-tv-shows-eda>