
DASHBOARD

ARTUR XARLES ESPARRAGUERA
NIKOLINA LIPOVCIC
ENRIC AZUARA OLIVERA

Master in Fundamental Principles of Data Science
Universitat de Barcelona
2021

1 Chosen dataset

The dataset we will work with was found in one of the PDFs provided by the course, which is a public repository from Tableau[2] that references a Kaggle dataset[1].

This dataset consists of 7787 TV shows and movies available on Netflix. It was collected from “Flixable” which is a third-party Netflix search engine. It is a tabular dataset which should be updated at least annually. The dataset contains mainly categorical variables, but it also contains a temporal one and a numerical variable corresponding to the duration of the film/TV show. We will describe each variable on next section.

Given the raise this type of platforms had during the last years, this dataset provides a lot of opportunities to explore the evolution of Netflix, analyze its catalogue, among others.

1.1 Variable description

- **Type:** This variable divides the entries in two categories: movie or TV show. (69% movies and 31% TV show)
- **Title:** The variable that gives the name of the movie/TV show. Each item has a different title.
- **Director:** The variable that gives the full name of the director of the movie/TV show. There are 4049 different directors. Each director has done from 1 to 18 films.
- **Cast:** The list of the actors that perform in the movie/TV show. Each film has many different actors.
- **Country:** The variable that gives the country of origin of the movie/TV show, which is the country where the production company of the title is based. There are 82 different countries.
- **Date_added:** The date (day, month and year) when the movie was uploaded on the Netflix website.
- **Release_year:** The year when the movie was first released for the public to see. Years are between 1925 and 2021.
- **Rating:** TV Rating of the movie/show. This variable has different categories depending on which audience is suitable for the given content. For instance, TV-MA programs are intended to be viewed by a mature audience, TV-14 may be unsuitable for children under 14, or TV-PG programs, which may contain material that parents could find inappropriate. Here we show all the different categories: TV-MA, TV-14, TV-PG, R, PG-13, TV-Y, TV-Y7, PG, TV-G, NR, G, TV-Y7-FV, UR, NC-17. They have been classified into 4 groups, depending if the content is suitable for children, teenagers, young adults or adults.
- **Duration:** Indicates duration in minutes for films or number of seasons for TV shows.
- **Listed_in:** Variable that makes reference to the genre of the movie/show. Each film can have different genres. There are 36 different genres: Anime Features, Children & Family Movies, Classic & Cult TV, Classic Movies, Comedies, Crime TV Shows, Cult Movies, Documentaries, Docuseries, Dramas, Faith & Spirituality, Horror Movies, etc.
- **Description:** Brief summary about the movie/show.

1.2 Brief visual exploration

In Figure 1 we can see that we have more movies than TV shows in our dataset. Concretely, we have around 5400 movies while we only have 2400 TV shows.

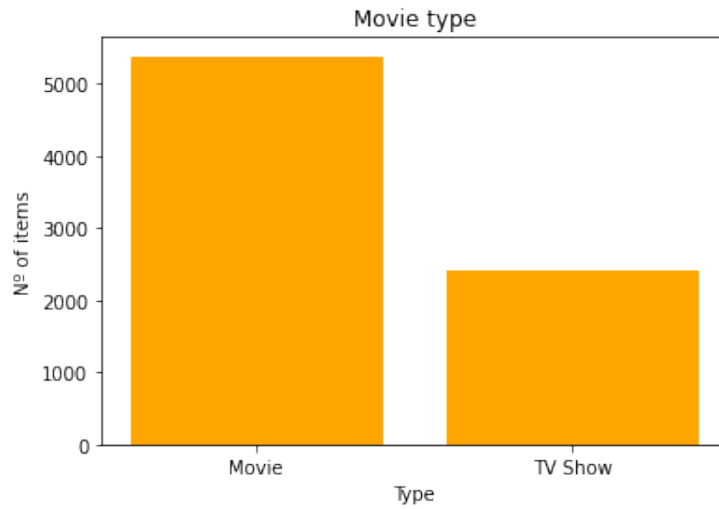


Figure 1: Amount of movies and TV shows available.

We can observe at left graphic of Figure 2 that most of the movies / TV shows available on Netflix are from the United States, which has twice the number of films of the following country, India. This result makes sense with right graphic of same Figure, since six of the ten directors with more films in the dataset are American.

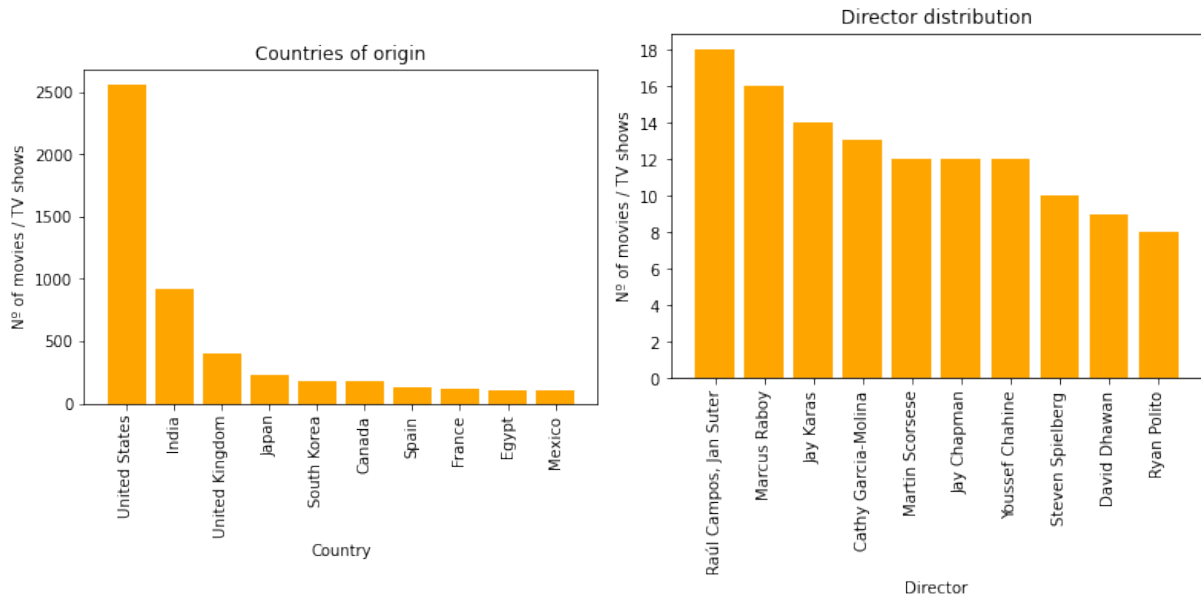


Figure 2: Countries and directors with more appearances. Left plot shows number of films and TV shows per country and right plot directors with the most number of films/TV shows.

As we can see at Figure 3, most of the movies / TV shows are made for a mature audience, or people over a certain age, while only a few of them are for a general audience (G).

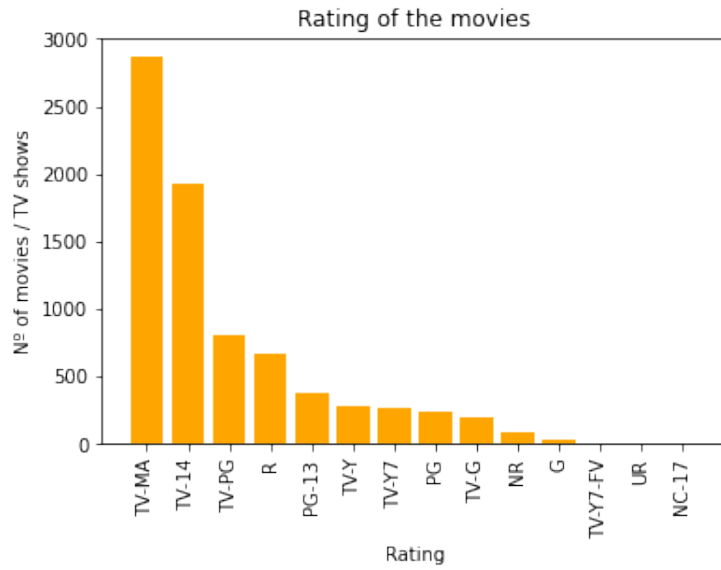


Figure 3: Frequency of rating category for films and TV shows.

To conclude, Figure 4 shows that the number of movies / TV shows released per year. As we can see at Figure 4, the production of visual content has increased during the years but on 2020 it decreased because of COVID.

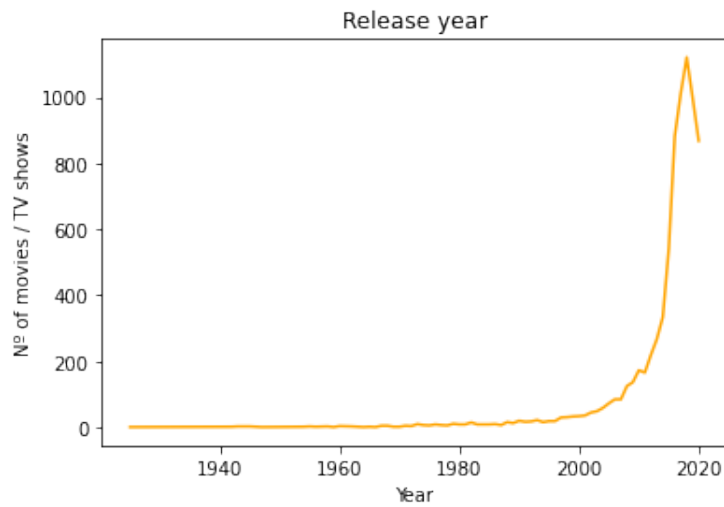


Figure 4: Movies/TV shows released per year.

2 Audience (top-bottom): Who is your audience?

2.1 Audience

Mostly non-experts, Netflix subscribers, or potential subscribers who want to explore what Netflix offers. The dashboard has been created to visualize it via laptop. They might consult the dashboard only once or every year for the curiosity of Netflix policies about which types of films/tv shows are available and maybe see the evolution of it.

2.2 Persona

The persona is any user who wants to check which type of content is available on Netflix taking into account different filters (genre, TV rating, country, etc.) regardless of whether this person is already a subscriber or not. Thus, it would be any person who is willing to watch visual content . A scenario of use could be a person who is deciding if he wants to join (or resubscribe) at Netflix and uses this dashboard to decide. Although it's not the main goal, this dashboard can be used to compare with other platforms, but it will only show information about Netflix. In addition, the persona could also be an internal customer (Netflix itself) who wants to have an overview of the titles they offer.

2.3 Message and questions to answer

The main message of the dashboard is showing the different types of content that Netflix offers taking into account different filters. Thus, the main goal for the dashboard is attracting new potential users, retain the ones that they already have, or just give general information about the content available at Netflix. Questions to answer are the following:

1. Number of old films (1990>) available at Netflix? This way, people who enjoy old or new films will know if there is that type of content on Netflix.
2. What is the movie/TV show distribution at Netflix at an specific country? So that users and potential users can know which type of content is more common in their country.
3. What is the distribution of the movies by their rating? A potential subscriber might want to know how many titles are appropriate for children or if it's more adult-oriented.
4. What is the movie/TV show genre distribution? A potential or actual subscriber might be interested in specific genres.
5. General information about an specific film or actor. Users might like certain actors and directors, so they could consult this Dashboard in order to filter out the movies and TV shows that are acted by or directed by these people and also look for the most rated movies by an specific filter.

3 Choose charts. 1st napkin design

The 1st napkin design is the following one:

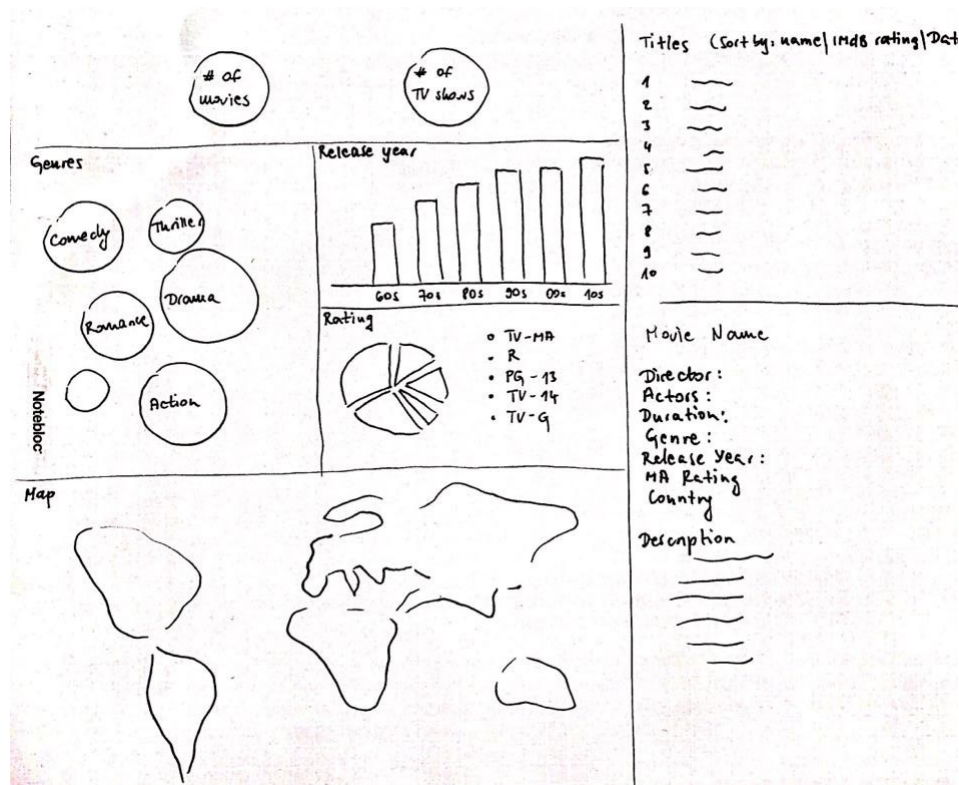


Figure 5: First napkin design

On the left part of the dashboard, we have charts for the following variables: Type, Genre, Release year, Rating and Country. For the type of the films, we show the number of movies and TV shows. Below, we show via bubble chart the genres, we use a barplot for the release year and a pie chart to show the different ratings. To conclude, on the bottom part of the left part, we have a country map colored depending on the number of cases per country. All charts described are interactive, so you can select between the different labels of each plot. After that, all the other charts would be actualized depending the selected filters.

On the right part of the dashboard, we have a list of 10 movies/TV shows that are from the labels selected on the left part (the first 10 ordered by name). Moreover, you can choose between the 10 movies and get information about them (Director, Actors, duration...). We might consider adding a search bar to look for an specific film or TV show.

4 Second design and implementation

After considering the options we had, we redesigned the dashboard into 2 screens:

1. **First screen:** Allows you to interact between different properties of the data as genre, TV rating, Country, date added, etc.
2. **Second screen:** Gives detailed information about the films that fits the filters selected and also gives their rating on IMBD, the actors on it, the director of the film or TV show and its poster.

4.1 First screen

On the first screen we added a total of 5 charts, where all of them are interactive, so you can look for specific information taking into account different properties. All the charts have a different color attached, to not confuse the meaning of them, but more or less they follow the same palette for design reasons. The design palette chosen for the Dashboard is the red-brown, but it also includes white and gray borders for visual design reasons. This is the chosen color trying to match the known colors for Netflix, which are red over white or black background.

- **Genre of the movie chart:** The first chart we are going to talk about is one that tells you the amount of content depending on its genre. To visualize it we have chosen a table because the amount of different genres is too high, so this way we can see all the available genres. The palette used for this table is a gray-white one, in order to make it more friendly to look at. The color has no meaning and we have chosen this one due to design reasons.

We placed this chart of the top-left part of the Dashboard because we consider it one of the most important for the user.

- **Type of content chart:** The second chart we will talk about is the one that tells you the number of TV shows or films available. The chart chosen is a bubble chart because we want to show the amount of content that is a TV show or a movie. The colors used are pink and flesh. Since the size of the bubble gives you an idea on the amount of content for each category, the only purpose of the color is design wise.

This chart is situated on the middle top part of the Dashboard, since we consider it one of the most important.

- **TV-Rating chart:** The third chart we have on the Dashboard is the one that tells you the TV-Rating of the content. The type of chart chosen is a stacked column/bar because we don't have a lot of categories and it is a good way of visualizing in a fast way how many content you can watch given a certain age. The color chosen is orange. It has no meaning and the only purpose of it is following the same palette of the Dashboard without superposing with a color from another plot.

We positioned this chart on the middle bottom part because we consider that it has overall less importance than the other two, but it is still an important one.

- **Country chart:** The fourth chart we can see on the Dashboard is the one that tells you the country of origin of the content. The chart shows a map where you can visualize via the intensity of the color the amount of content from that specific place. The color chosen is a palette of brown, where the most intense brown determines more amount of content and a lower intensity of it tells you that there is less content. The intensity of the number of

cases is determined by the logarithmic scale of the cases in order to intensify the difference between countries that produces less content.

This chart is positioned on the top-right part of the Dashboard because we think it might be important for some users knowing the origin of the content, but others might not care as much.

- **Release year chart:** The last plot on the Dashboard is the release year of the visual content. The chart chosen is a barplot with 10 columns that tells you the amount of content available on a range of years. The first category range is before 1980, since there are few cases released previously that year. The color chosen is a light yellow. It has no meaning since the height of the bar tells you the amount of content on each year range and the main reason to chose that color is to follow the design palette established before.

The position of this plot is bottom-right. Although some people is only interested in the classical movies or only the newest ones, we think that the year of the content is the least important feature to look at.

As said before, all this charts are interactive and allows you to explore among them the available content taking into account all this properties.

The first screen also includes information for each chart, instructions on how to proceed to interact with the charts and a reset filter button in order to clear all the conditions applied. The last thing this screen includes is a button to go into the second screen.

4.2 Second screen

Second screen is an interactive screen that will show a list of the content that matches the filters applied on the previous screen. This screen can be divided into 3 parts:

1. **Search bar:** The search bar is situated on the top of the screen and is the smallest part of it. The only feature it holds is a search bar. The only purpose of it is to look for a movie when not a lot of filters have been applied. This way, the user can fast look by searching on this bar if the movie or TV show is available on Netflix. The search allows the user to look for film, actor or director. Once the search is done, it will appear all the content that matches the search on the left part.
2. **Search results:** It is situated on the left part of the screen. It will show the name of the content that matches the search done on the search box and the content that passes the filters applied on Screen 1. On this search box the user is also able to look for Directors or Cast.

On top of everything, we get a list with the filters applied, followed by the Title of the content, the IMDb Rating attached to it and the date added to Netflix. If you click on a certain movie or TV show, it will show more information on the right part of the screen.

3. **Extra information:** The right part of the screen shows extra information of the selected content on the left part. At the top it shows the type of content, the year added, the number of seasons (in the case of the TV series) and the TV-Rating of it. After that, we get some general information as the Director, the Cast, genre, Country of origin and summary of the content. Below all this information, we get a poster of the content, which helps the user to familiarize with the TV show or film.

This screen also has a button to return to Screen 1 in order to change some filters or revisit again the charts.

4.3 Final implementation

After having designed the type of Dashboard we want to implement, we used Tableau to do so. We have chosen Tableau because of it's simplicity to build the charts and for design reasons.

The dashboard can be visualized on the html attached on the folder delivered on the campus or on the following link to access at the GitHub where we have all the sources, code and the Dashboard. Link to the Github: <https://github.com/arturxe2/Visualization-Dashboard-Netflix>.

When accessing the dashboard in the HTML format you may see that the poster can't be visualized as a whole. That's caused by a common problem in Tableau Public that is not able to resize the images. However, the posters are correctly seen in Tableau Desktop. If you open the HTML with Mozilla Firefox, there is an improvement in the visualization of the posters, which appear in full size. However, they are still not correctly located (not centred).

At Figure 6 and Figure 7 can be visualized both screens of the Dashboard.

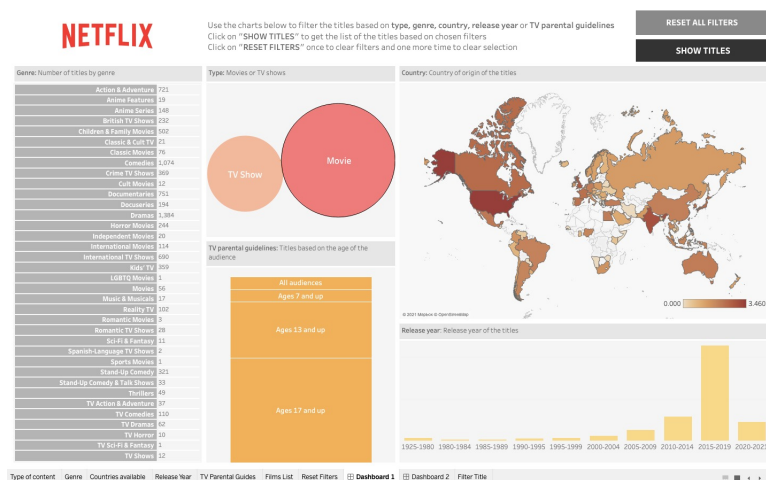


Figure 6: First screen of the Dashboard

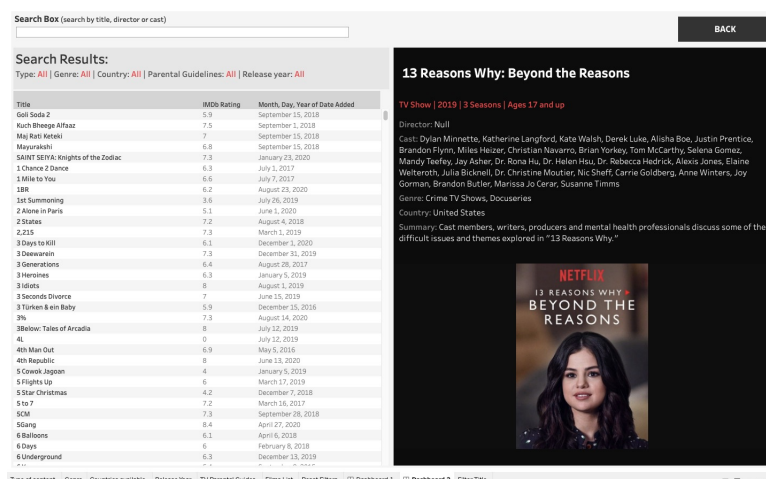


Figure 7: Second screen of the Dashboard

References

- [1] TV Shows and Movies listed on Netflix <https://www.kaggle.com/shivamb/netflix-shows>
- [2] Tableau public repository https://public.tableau.com/s/resources?qt-overview_resources=1#qt-overview_resources
- [3] TV Parental Guidelines https://en.wikipedia.org/wiki/TV_Parental_Guidelines