MovieDB - A website to help the user choose a movie platform

Ana Sofia Moniz Fernandes, 88739 | Pedro Miguel Oliveira, 89156

Information Visualization, 2020 (Master's Degree in Informatics Engineer, University of Aveiro)

Abstract

The present report aims to explain the implementation of the platform developed for the mini-project of the Visualization Information course - Development of a Visual Data Exploration Application.

This application uses D3.js library to show, in a clear and concise way, information about movies that exist in four platforms - *Netflix, Amazon Prime, Hulu* and *Disney+*. Thus, the user must be able, after using the ***MovieDB*** platform, to choose the movie streaming platform that suits his tastes the most.

Motivation and objectives

In the scope of the course Information Visualization, the students were asked to choose a theme (from the suggested list or another, if agreed with the professors) to build a platform that would clearly show information.

Thus, the chosen theme, presented in this report, is "Movies on Netflix, Prime Video, Hulu and Disney+" - a very complete dataset, that contains various information about the movies presented in these four platforms; it included over 16 thousand entries, with no missing values.

The main goal, besides explicitly showing information, is to have a minimal application, easy to use, that will allow the user to choose a movie streaming platform that suits his personal tastes. The user should leave the website with no doubts about which movie platform to choose.

Users and the Questions

In order to build an application easy to use, the target users and questions to be answered by the movie platform must be known.

Characterization of the users and their context

Nowadays, having a movie streaming service easily accessible is an important question. This type of service has become essential in most people's life. People don't buy/book movies anymore, and the habit of going to the cinemas has been decreasing over the years.

However, people find it hard to choose a movie platform - the platforms have different movies, with different genres, reviews, runtime, age range, etc.

Therefore, the main users are people who are looking for a streaming platform but still don't know which one to choose. First, the user needs to be sure which platform suits his tastes the most, once it is a paid service- the action of subscribing to a movie streaming platform is a reasoned decision.

Questions to Answer

In the process of choosing a platform, several questions may appear to the user:

- Which movie streaming service should I subscribe to?
- Which movie streaming service meets my personal taste?
- Which movie streaming service has more movies?
- Which movie streaming service has the best-reviewed movies?
- Which movie streaming service is more appropriate to my age?
- Which movie streaming service has the longest movies?
- Which movie streaming service has more movies in my native language?
- Which movie streaming service has more movies from my native country?

Dataset

The chosen dataset contains 16743 entries and 17 columns and can be used to get information about the movies in each platform, for example:

- Know in which streaming platform(s) a movie can be found
- IMDb and Rotten Tomatoes scores for the movies of each platform
- Target age group movies
- Year of production of each movie
- Who directed the movie
- Get to know the genres of movies in each platform

Besides the movie dataset, it was used another one to supply countries' coordinates- these coordinates were used in a visualization that will be shown later in this report.

Visualization Solution

Taken all this into account, it is easily seen that a minimal website, easy to use, is needed to support the decision of which movie platform to subscribe.

Thus, ***MovieDB*** is the visualization solution for this problem.

Low fidelity prototype and user feedback

Before implementing the website itself, it was created a low fidelity prototype, using proto.io, and it was tested with colleagues in the class of Information Visualization course, on 18th November 2020.

In this class, heuristics evaluation and user usability tests were made. These evaluations were very important in the process, once it allowed obtaining important feedback:

1. Main feedback - directors chart is not perceptible. It doesn't make sense in the platform once this information is not relevant for the user (Figure 1);

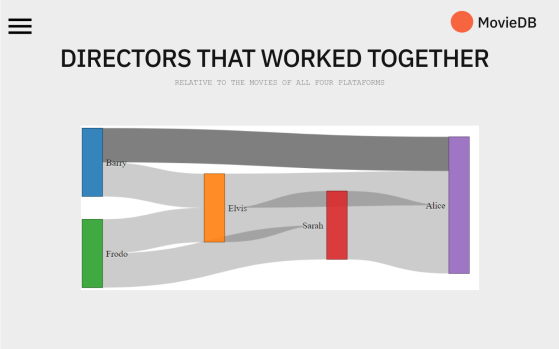

Figure 1: Prototype page for directors who worked together

2. In the "Movies Released Per Year" section, it would be better to use a chart with colours instead of having a tab bar (Figure 2);

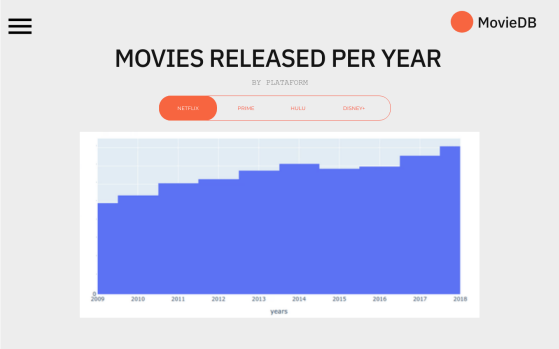

Figure 2: Prototype page for movies released per year

3. Having the common/not-common movies between platform - eg. there can be an important movie for the user that is present in a platform and not in the other;
4. In some charts, the tab bar should be centred;
5. Beside the wordcloud of each platform should exist a chart that would compare the genres between platforms (Figure 3);

Figure 3: Prototype page for movie genres

6. The movie runtime chart should be coloured too (Figure 4).

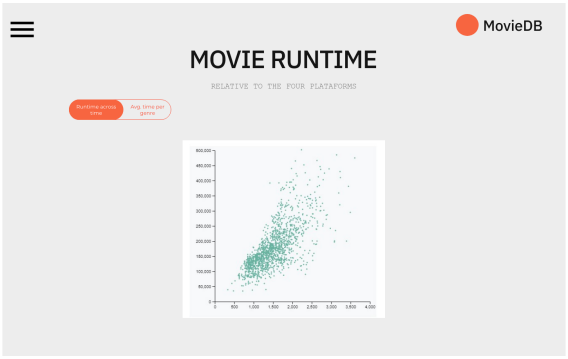

Figure 4: Prototype page for movies runtime

Functional prototype

After receiving feedback from the user, it was time to implement the functional prototype.

A.Information visualization

1. Main page: on the main page, the goal was to give the user some global information. This way, the user can:
    a. Discover the number of movies per platform- the piechart (Figure 5) allows the user to pass the mouse over the chart and see a tooltip informing how many movies are present in that platform;
    b. Discover the exclusive movies per platform - the barplot (Figure 5) shows a tooltip informing how many exclusive movies each platform has if the user passes the mouse over the chart;

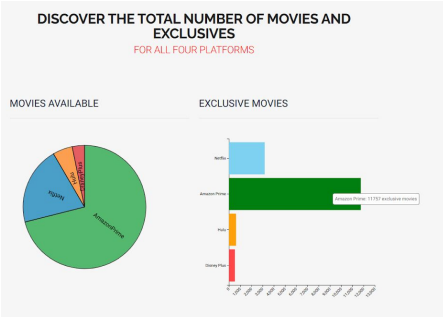

Figure 5: Total number of movies available (piechart) and exclusive movies (barplot), both for each platform

c. The top 10 genres for all four platforms - a doughnut chart was used (Figure 6);
d. The top 10 films released year for all four platforms- a doughnut chart was used (Figure 6);

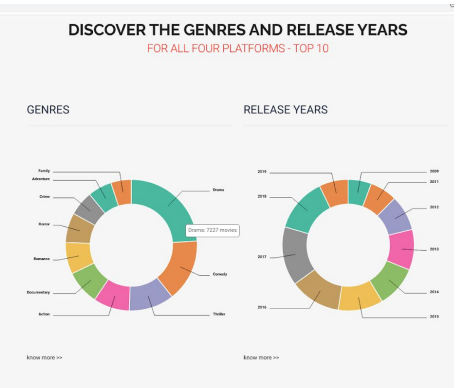

Figure 6: Top 10 genres and release years for all four platforms - doughnut chart

It should be noted that once the data was too big, it was only perceptible to show the top 10 of genres and movies released per year.

2. Movies Reviews: an important aspect, when choosing a platform, is to know if the movies present in there are well classified. It is important to note that the reviews were made by *IMDb* and *Rotten Tomatoes*. Thus, the "Movie Reviews" page allows the user to navigate between three charts, discriminated by platform:
    a. Movies by Classification Range: the user can discover, for the platform he chose (through the navbar), how many movies exist for each classification range (eg. for the range 6.2-6.4 there might be 140 movies rated by *IMDb* and 120 rated by *Rotten Tomatoes* (Figure 7));

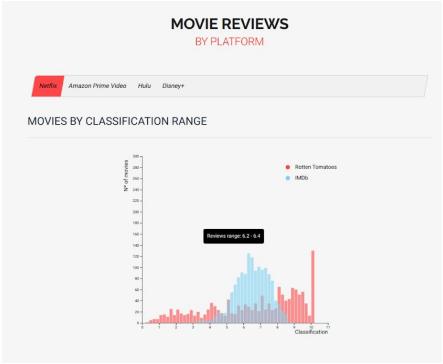

Figure 7: Double histogram- movies by classification range for Netflix

b. Avg Classification by Age Group: with this Cleveland dot plot, the user can discover, for each age group, which one is better classification. Each circle represents the reviewer - IMDb or Rotten Tomatoes (Figure 8). If a mouseover is done on the circles, the classification for that reviewer and age group is shown in a tooltip;
c. Top 10 movies: with this lollipop chart (Figure 8), the user can discover which are the top 10 movies with a higher classification of that platform. If a mouseover is done on the circles, the classification value is shown in a tooltip.

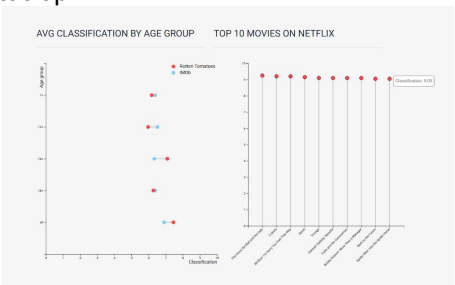

Figure 8: Average classification by age group (Cleveland dot plot) and top 10 movies on Netflix (lollipop chart), respectively

3. Movies Runtime: In this page, the client can see, through a Connected scatter plot, the average length (in minutes) for each genre of movies (Figure 9).

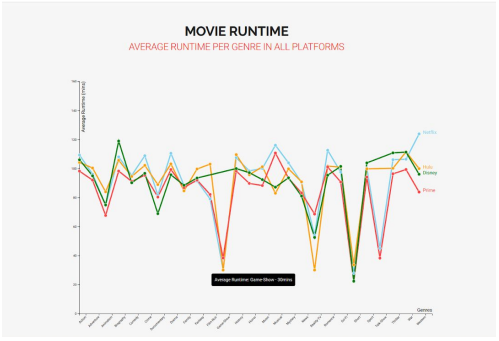

Figure 9: Connected scatter plot with multiple series- Average time per genre, for the four platforms

4. Genres by platform: here, the user can navigate between the four platforms and discover the predominant movie genres of each one. The wordcloud is combined with a lollipop chart (Figure 10) so that the user can know, precisely, how many movies there are in each genre;

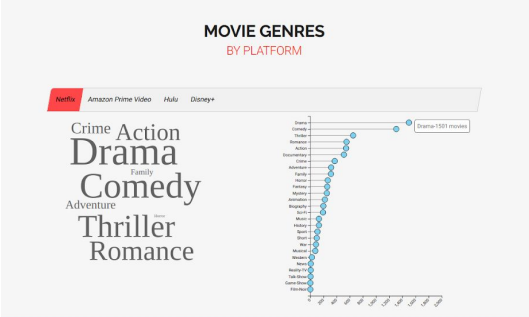

Figure 10: Movie genres on Netflix (wordcloud on the left side and lollipop chart on the right side)

5. Movies per year, by platform: This page has two main purposes - it lets the user keep track of how many movies a platform will have of a period of time (e.g. for users that may prefer movies from the *90s*) as well as let the users compare directly between platforms about how many movies each platform (Figure 12) has released over the past 20 years, letting them know if some platform is falling off or if it's getting better. Besides that, the user also has access to a grouped barplot (Figure 11) that compares the four platforms side by side, so that he doesn't need to navigate between the tabs multiple times when comparing the platforms between them.

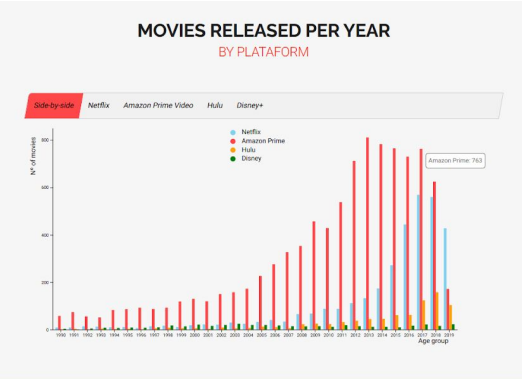

Figure 11: Grouped barplot - comparison of movies released per year on each platform

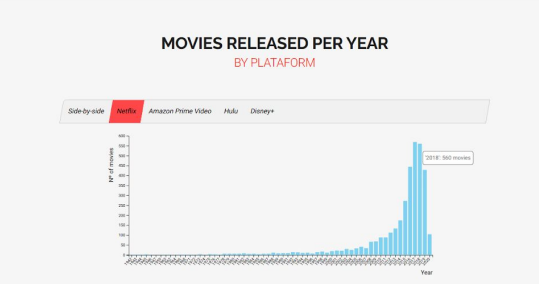

Figure 12: Barplot - movies released per year on Netflix

6. Movies per country/language: here, the user can navigate between:
   a. Countries - the user can discover the origin country of the movies of each platform. The information is represented in a map (Figure 13) that, with mouseover, shows a tooltip for each country where movies were made;

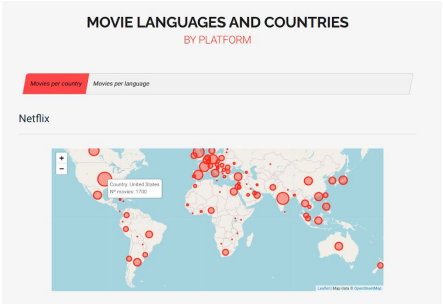

Figure 13: Map showing from where the Netflix movies are

   b. Language - the user can discover the native language of the movies of each platform. The information is represented in a circular barplot (Figure 14) that, with mouseover, shows a tooltip for each language (indicating how many movies there are for that language). Due to a large amount of data, it was only perceptible to show the top 50 languages. It should be noted that each barplot's color differs with the platform that is being observed;

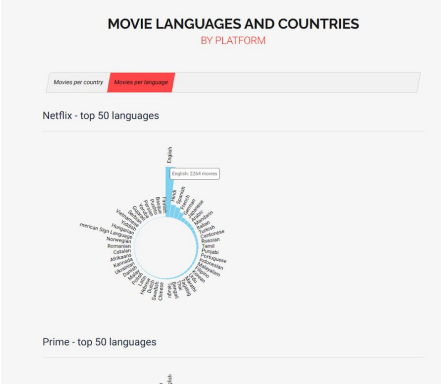

Figure 14: Circular barplot - Netflix top 50 languages

7. Movies by age range: the user can navigate between the four platforms and discover the age group predominant of each one, through barplots (Figure 16). It might be important to a user that, for example, seeks for movies to watch with the whole family). Besides that, it is possible to compare the four platforms in one unique grouped barplot (Figure 15) It should be noted that, once again, each barplot's color differs with the platform that is being observed.

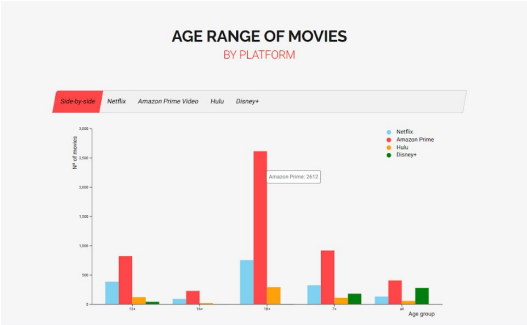

Figure 15: Grouped barplot - Comparison of amount of movies per age group per platform

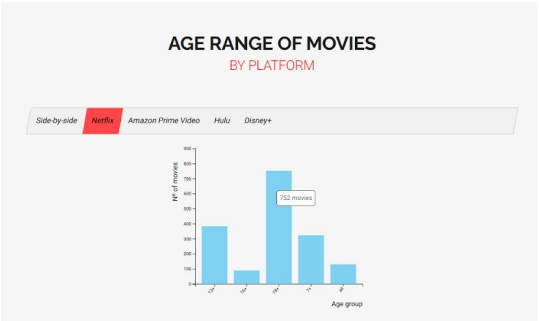

Figure 16: Barplot - Ammount of movies per age group

Implementation challenges

Due to being a very extensive dataset, it was difficult to maintain some of the initial ideas - some information would be very little and not perceptible in the charts. It was the case with doughnut charts on the main page, for genres and release years, as well as with the main languages of the movies of each platform.

Besides that, the team tried to implement all the feedback that was given, once it was transversal between all the users that tested it in the interface tests. Some difficulties were found in the moment of choosing the best d3 chart to represent these new ideas.

Evaluation and changes in the prototype

To have a website that would answer all the user questions relative to which platform to chose the team tried to take into account all the user feedback and implement all changes and new features suggested during the interface evaluation.

Hence, the main differences between the website itself and the prototype are:

1. The directors' page (Figure 1) doesn't exist on the website, once it wouldn't interfere with the user choice;
2. The genres wordcloud (Figure 3) is now combined in a lollipop chart so that the user can have a better notion about the genres in each platform (Figure 10);
3. A colour-coded chart was added to the movies release section (Figure 11), as well as to the movies age range section (Figure 15);
4. Due to a large amount of data, it could only be shown the top 50 languages of each platform (Figure 14), as well as the top 10 genres and released years of all four platforms, in the home page (Figure 10);
5. It was added a chart with the exclusive movies counting of each platform (Figure 5).
6. In the prototype, the idea was to have two Cleveland dot plot, one for the average classification by age group (Figure 8) and a second one for the average classification by genre. The team concluded that the second one wouldn't be possible, once there are two many genres. Once there is already a genre section on the website, it was decided to implement a lollipop chart showing the top 10 movies with a higher qualification, in each platform.

Conclusion and Future Work

It was possible to infer that the choice of subscribing to a movie platform is very important for people nowadays. It is a very weighted decision, that comes with lots of questions and doubts. This way, ***MovieDB*** platform is the solution for the people who struggle with these questions, once it shows information clearly and explicitly.

For future work, the idea is to implement a mobile application so that the user can see the information through a smartphone. Some charts will have to be tailored once it might be hard to read some information screens with reduced dimension.

References

[1] "Movies on Netflix, Prime Video, Hulu and Disney+", [Online]. Available: https://tinyurl.com/y55aj773 . [Accessed on 20th December 2020]

[2] "Visualization, analysis and Recommendation System", [Online]. Available: https://tinyurl.com/yxwytoyy . [Accessed on 20th December 2020]

[3] "Movies around the world", [Online]. Available: https://tinyurl.com/y27zgnmw . [Accessed on 20th December 2020]

[4] "The D3.js Graph Gallery", [Online]. Available: https://tinyurl.com/yykx3p7w . [Accessed on 20th December 2020]

[5] "How to create a grouped bar chart in D3.js", [Online]. Available: https://tinyurl.com/y4tc6t7y . [Accessed on 20th December 2020]

[6] "Color Schemes", [Online]. Available: https://tinyurl.com/yxpp2v6v . [Accessed on 20th December 2020]

[7] "DIGITAL MARKETING FREE CSS TEMPLATE", [Online]. Available: https://tinyurl.com/y3bgz7xy . [Accessed on 20th December 2020]