



## Data Visualization

-

### Discovering the movies data

---

Ben Hassen Mahdi

Group MNS

## Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Data analysis</b>	<b>3</b>
2.1	Data Provider . . . . .	3
2.2	Content of data . . . . .	3
<b>3</b>	<b>The path and challenges for the final result</b>	<b>3</b>
<b>4</b>	<b>Peer assessment</b>	<b>5</b>

## 1 Introduction

Films are cultural artifacts created by specific cultures. They reflect those cultures, and, in return, affect them. Film is considered to be an important art form, a source of popular entertainment, and a powerful medium for educating—or affecting—citizens. The visual basis of film gives it a universal power of communication. Some films have become popular worldwide attractions through the use of dubbing or subtitles to translate the dialog into other languages. That's why for this project, we decided to use IMDb data set to get more insight about the history of movies and try to visualize it in an interactive way.

## 2 Data analysis

### 2.1 Data Provider

The data set was provided from IMDb, the most popular movie website. It combines movie plot description, Metastore ratings, critic and user ratings and reviews, release dates, and many more aspects. The website is well known for storing almost every movie that has ever been released (the oldest is from 1874 - "Passage de Venus") or just planned to be released (newest movie is from 2027 - "Avatar 5"). More than 6 million titles information are stored (of which almost 500,000 are featured films).

### 2.2 Content of data

The movies dataset includes 85,855 movies with many attributes such as:

- Title: the title of each existing movie
- Year: the year of production.
- Genre
- Duration
- Country
- Director and writer
- Production company
- Actors: the principal ones
- avg\_vote: the mean rating

## 3 The path and challenges for the final result

First of all, we started collecting the data and cleaning it to make it understandable and reliable. Then we were interested in the distribution of the genre of all movies in the world. This may clearly describe the interest of the most of viewers on drama mostly, comedy second and romance in third position.

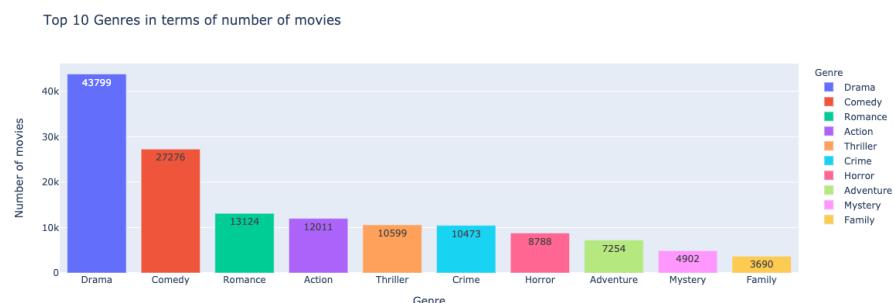


Figure 1: movies By Genre

After that, we decided to add another attribute and focus on classifying all movies based on their mean rating and proceeding with the production companies.

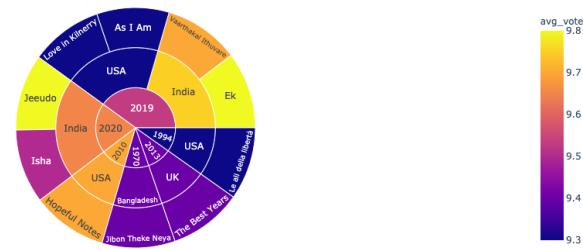


Figure 2: Best movies

The challenge here was about the idea of presenting all these informations in an interactive way in our website. At the beginning, we decided to create a map, since we were interested in analyzing each country and this was the first challenge. Starting with a blank page was pretty hard, but then i tried to implement what we learned in class about html, css and javascript. Forums were also my best friends for making this project run as it should. So the first map didn't look good for me and the way of presenting the number of movies for each country by some bubbles was not a good idea at the end. In addition to that, i was interested about clicking on each country and plotting some details, and that implementation didn't allow that at all. So i changed it by a d3 map to make it more useful. Second problem was connecting informations to each country id of the map since countries id of the data set were different from those used by the json file of the map. So i tried to correct the ids of our data and add all the attributes that i need the same json file so that it is easier to access and manipulate.

After that, i thought about coloring the map based on the number of movies per country to make it look nicer. A bar of color gradient was also put on the left to describe the distribution. For the rest of informations, i decided to make a popup on clicking on country, and showing all what i want. This was also challenging at first but i came up with a solution using tingle.



Figure 3: The map page before and after

## Russia

**Most produced genre:** Drama  
**Director with most movies:** Sergey A.  
**Writer with most movies:** Sergey A.  
**Most popular production company:** CTB Film Company  
**Most popular actor:** Sergey Garmash

Figure 4: Popup with all the informations when clincking on country

When i finished with the map page, i wanted to do something more challenging, such as creating some small animation and that's how i decided to build a moving figure consisting of a droid. Making all the part of the body synchronized in the rotation and fixing them together was the main issue. After that, I Decided to map each droid to a specific genre of movie, and make it rotate and change the color by just a simple click and finally plotting the best rated movies. I had to reorder the rating each time when clicking on two or more droids and that a bit difficult to handle.



Figure 5: The droids before and after interactions

Since CSS was the most amusing part, so even though i coded a lot of it in the two previous pages, i thought it would be great if i could add a first page with a nice changing page button consisting of some arrows.



Figure 6: First page

## 4 Peer assessment

At the beginning of the project we were a group of three talking about the project and deciding together on the visual part of the website. But just after the second milestone, my two team mates decided to drop the course just before the coding part, for their reasons that i understand. Since i was interested on web development and i didn't want to abandon the project, i pushed harder to come up with a good looking and working pages.