

# Data Exploration through Tableau

Giulio Cappellani — 20599547

The dataset for this assignment can be found at:  
<https://www.kaggle.com/kuntalmaity/complete-list-of-movies-on-netflix>

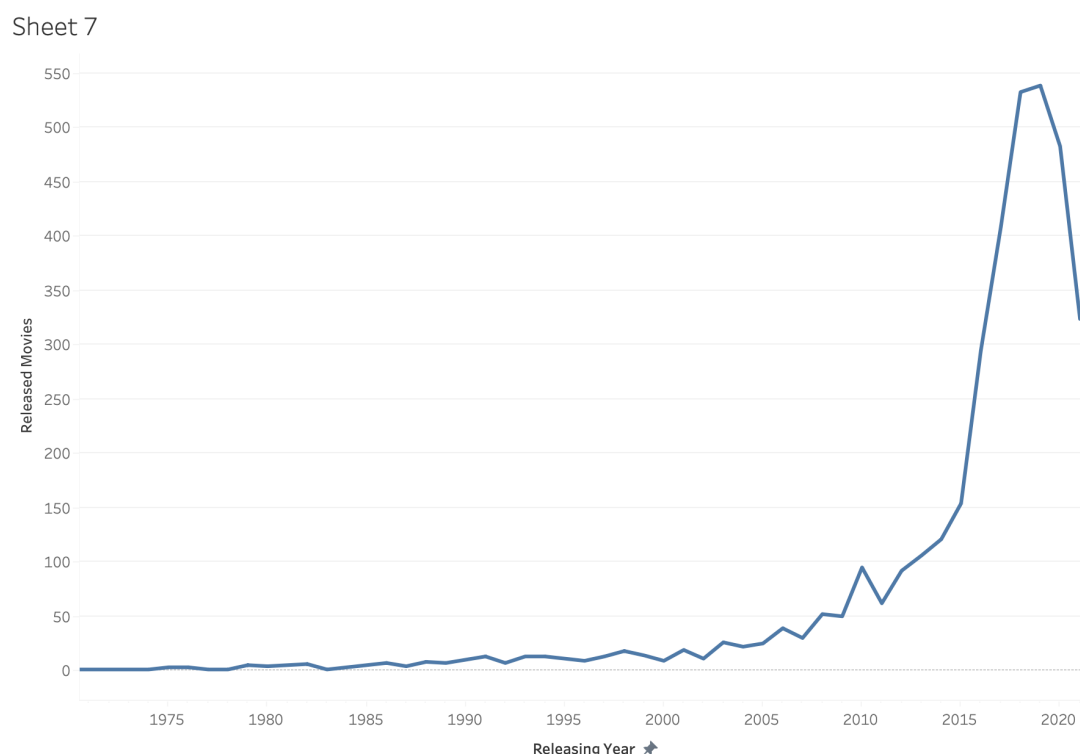
Such dataset contains the following features:

- Title (not used)
- Language
- Released Year
- Rating (not used)
- IMDB Rating (converted from string to a continuous variable)

Out of 3681 records 40% contains 'null' values for the 'IMDB Rating' feature. Since the remaining 60% (2110) still represent a fair amount we decide to ignore them when dealing with such a feature.

## Finding #1

The first thing we would like to play with is the number of movies released between 1970 and 2021. What we would like to find out here is how much COVID has influenced the industry and how production behaves overtime in general.

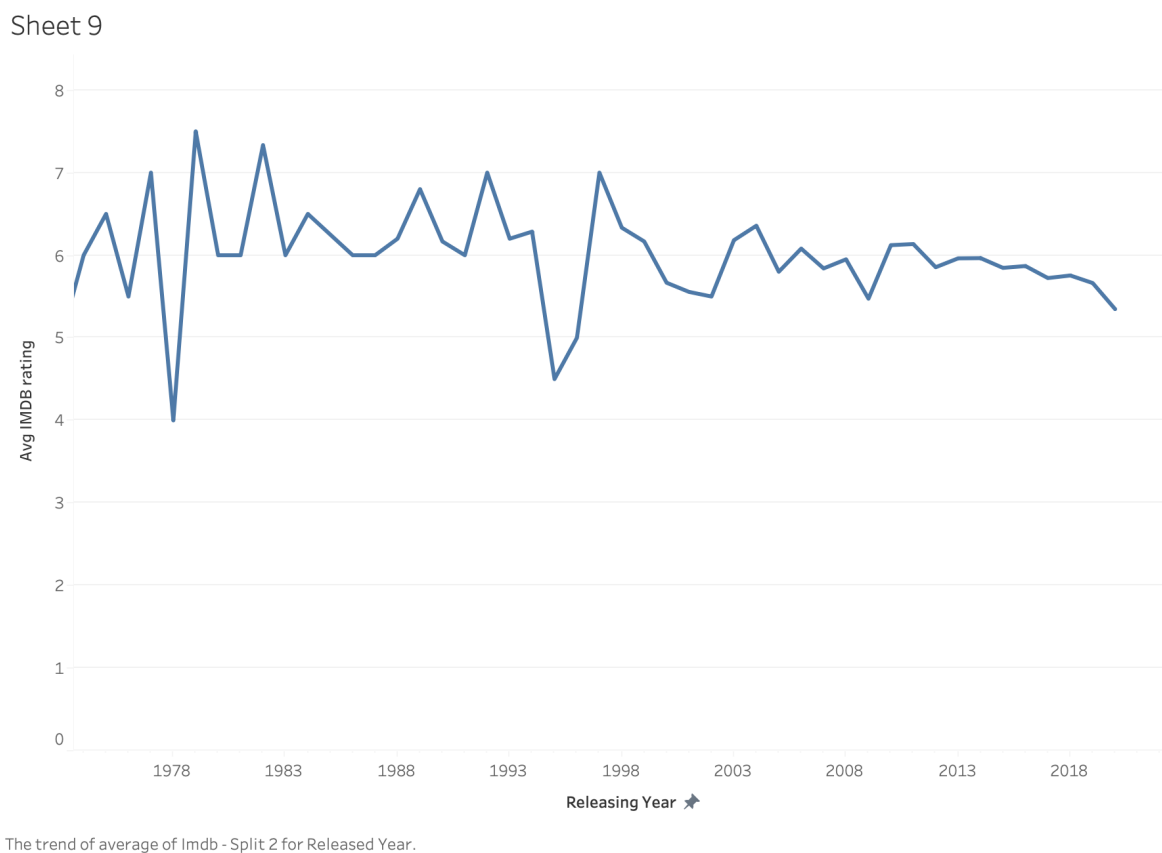


The trend of count of NETFLIX.csv for Released Year.

As we can see production seemed increasing, peaking between 2015 and 2019. Notice the huge drop in production that took in place between 2019 and 2020, this has to be related to COVID.

## Finding #2

With millions of active users, IMDB is always been one of the most reliable site when it comes to movie ratings. So the thing we would like to analyze at this point is how much the quality of movies is changing overtime, are we getting better at making movies?. In order to get a sense of this we will be using the average of the IMDB ratings (and not the sum) as the number of released movies drastically changes starting from 2012 as we have seen before.



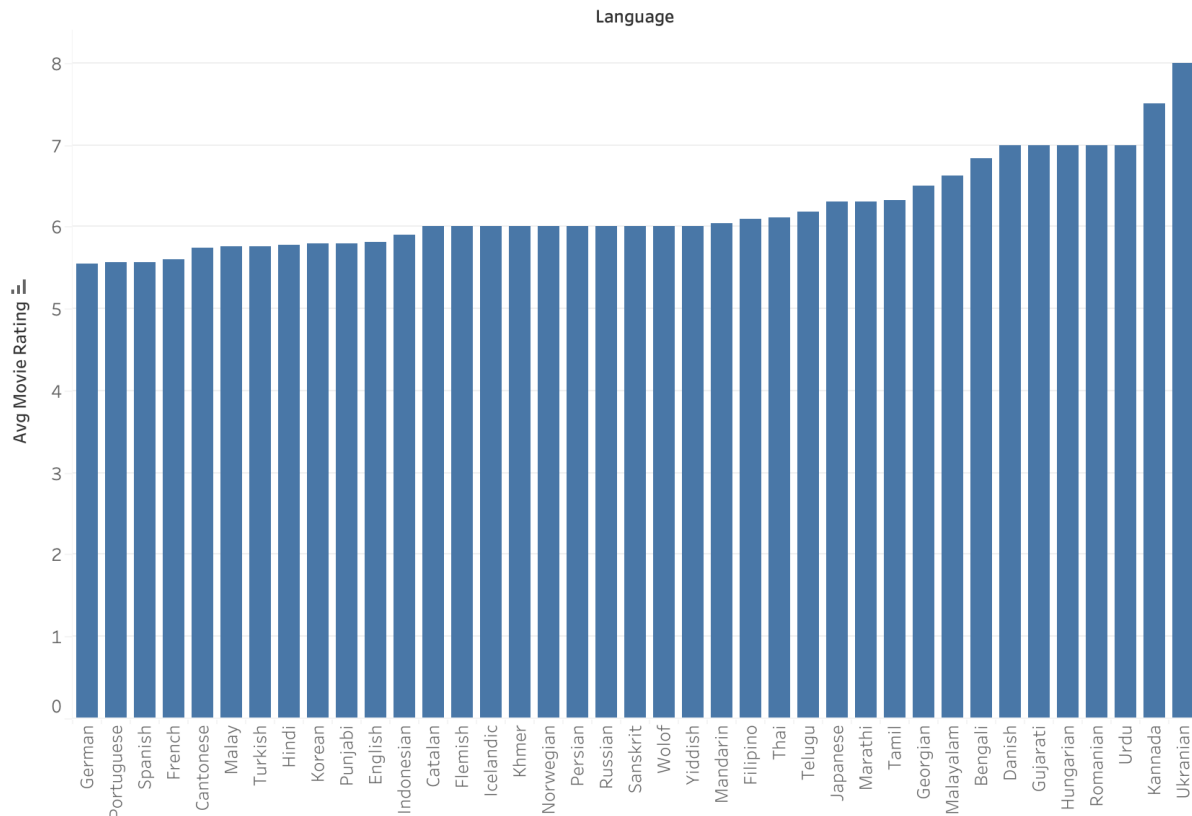
Though the changing rate is getting more stable overtime (which has to be related to the fact that fewer movies were released each year before 2006), it seems that we are slowly getting worse at making movies!

Again, this is just a weak consideration given how many records we are using and how they are spread overtime.

## Finding #3

So what countries are doing best? We will actually consider 'Languages' for this, as we have no informations on the actual countries that have released a movie so, for example, 'English' will involve all countries that speak that language (UK, US, Australia...)

Sheet 10



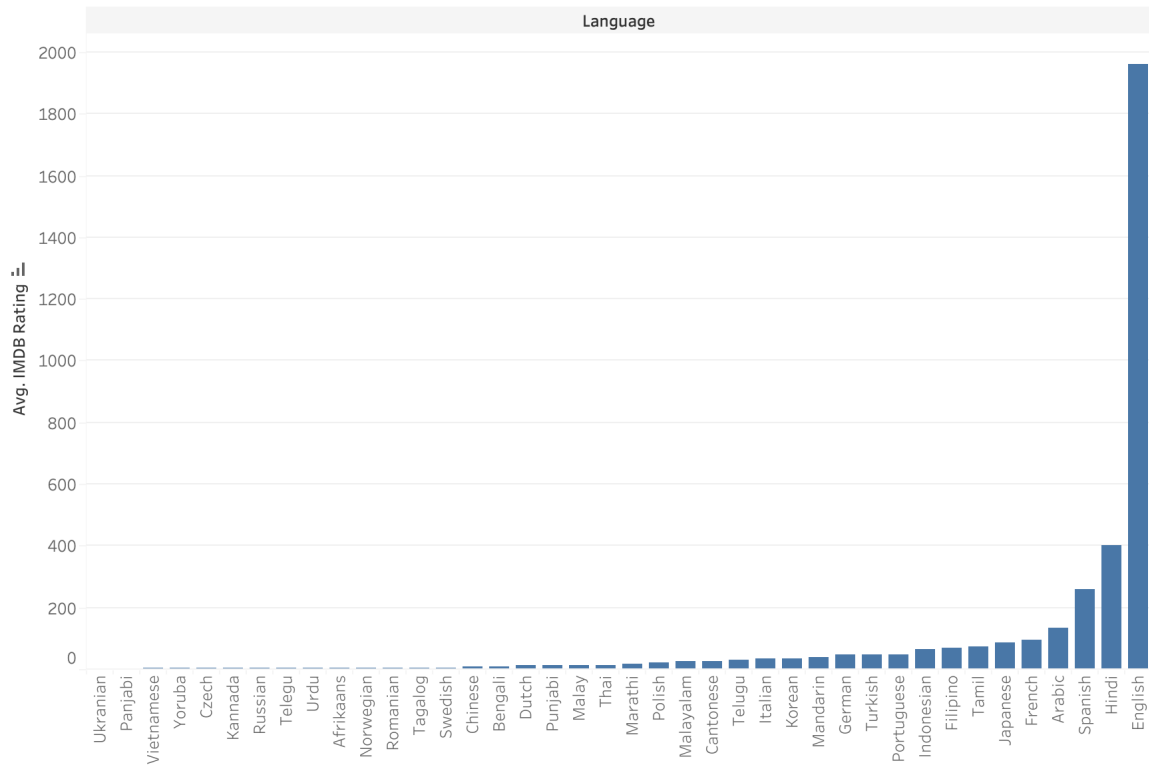
Well, the average movie rating seems increasing slightly, even though there is quite a big difference between the last ones and the first ones.

## Finding #4 (Bonus)

But, again, we have to contextualize this, as what if a group of countries sharing the same language have released just a single good (or bad) movie? That would not be a fair game.

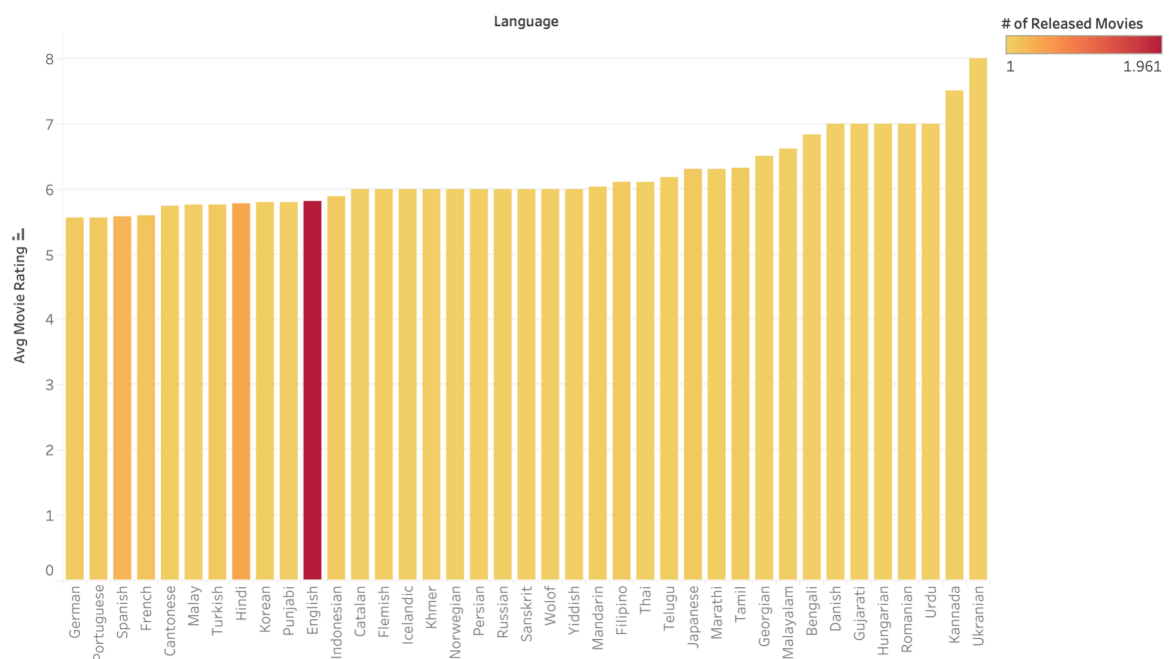
So what we would like to see now is how many movies have been released per each language (see next page) and what is the one who released the more movies.

Sheet 14



As we thought there are some group of countries that have released a very few numbers of movies compared to countries speaking 'Spanish' or 'English' (which is the one who have released the more movies), so let's combine the two graphs together with using colors.

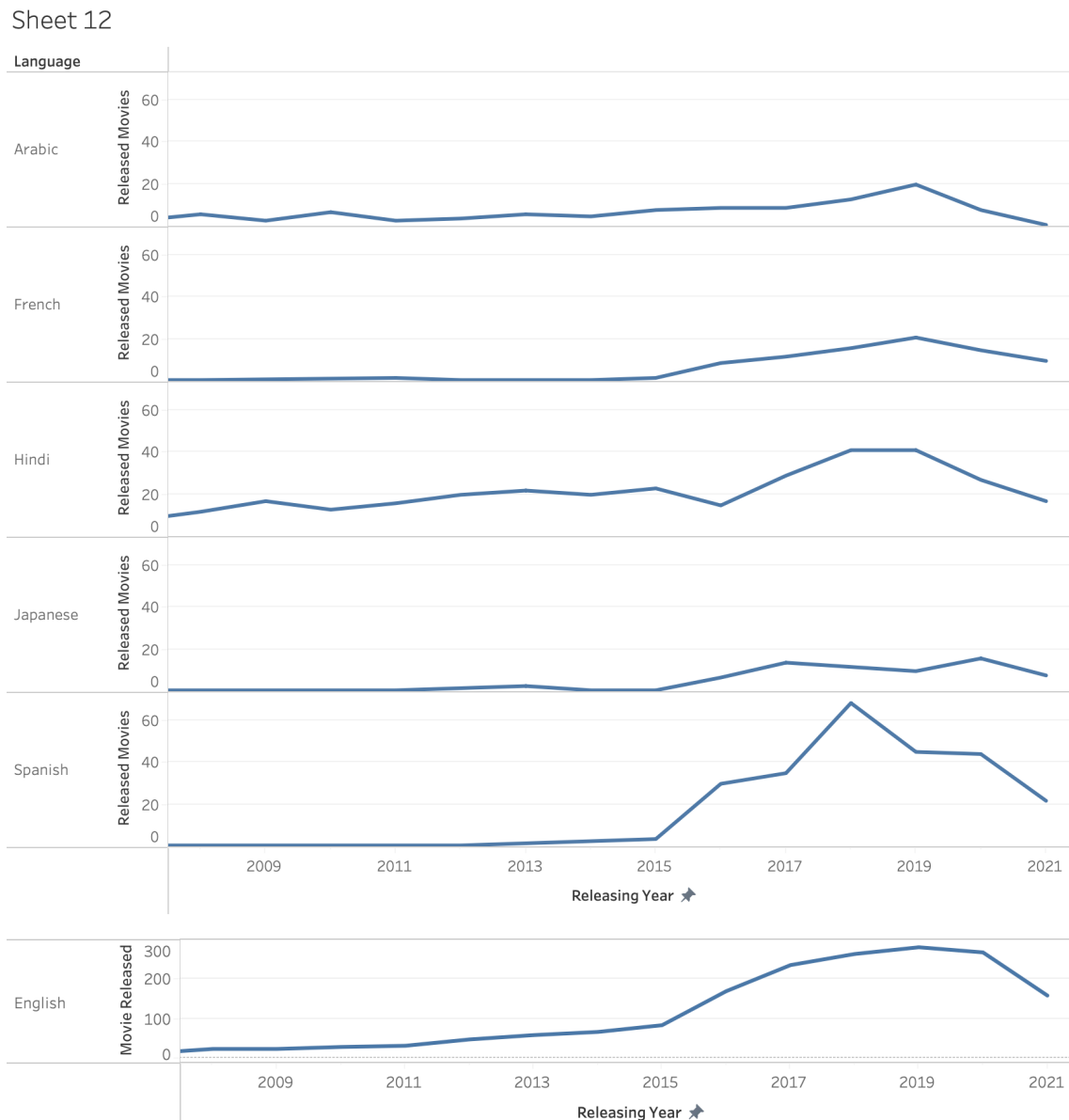
Sheet 10



## Finding #5 (Bonus)

Back to COVID impact on the movie industry, for our last finding, we would like to see how it affected the languages having released the more movies. In the figure below we decided to exclude 'English', showing it apart, as its number of released movies is way higher than the other ones, thus making difficult to notice the differences.

Also, since COVID happened in 2019, and given the trend until 2007, we decided to set the values on the x-axis to 2007-2021.



As we can see COVID had an impact on everyone, on someone more and on someone less, for example, the Japanese movies, seemed not to have had a huge drop compared to its previous trending while Spanish movies, already dropping in numbers before COVID happened, still had a significant drop in number. For all the other languages the difference between pre and post COVID is quite evident.

