Data Science in Netflix

Gregory Sampson S19156605
Word count: 1,188

May 2021

1 Introduction

This report will cover, in this given order:

1. Introduction
2. Motivation and Objectives
3. Results and Findings
4. Summary

I am using data visualisation to help my peers understand my motivation for this project as well as help me answer some questions I have thought of so that I can display and provide a visually stimulating piece of work. As it would be a lot of effort and time to use the data set by its self, Using R and data visualisation help speed up the process of getting the answers I want. Data visualisation can help me make graphs to help me work out the answers. Using data visualisation, I can further facilitate my data analysis using graphical representations as an aid.

The aim of my report is to help people understand how streaming services have grown over the recent years as well as the entertainment industry itself. This type of report could also help in the real world by helping businesses see what type of content is lacking, most popular and also what is being used the most within the service to help them create a more diverse streaming platform. Some achievements of this report are that I've been able to get visual answers for each of my questions and that I have been able to display them in a way that could be used to both educate and inform people about streaming services. I am also happy to see that the answers for these do relate to the real world and you can see that outside factors do affect the data set, such as locations, dates and more.

2 Motivation and Objectives

My data set focuses on content Netflix has released in the past up to the present day. This data set includes the date of release as well as the date it was added to Netflix, genres, directors, length and many more variables. I have chosen this data set as it is quite modern so any finding will be relevant to today, it's also a topic younger generations will find interesting and also it meets all the requirements that were specified and more, giving me a lot of information to use though I did have to clean up and edit some of it so that it was usable in my graphs. This produces some helpful graphs for my questions and also helps me find some of the answers I was looking for.
my four potential questions are:

- How has the number of TV-shows and movies added by Netflix changed over time?
- how has density of content changed over time?
- what is the variation within the genres of movies and TV-shows over all? (is there an equal distribution?)
- Are the number of releases around the world equal?

3 Experimental results

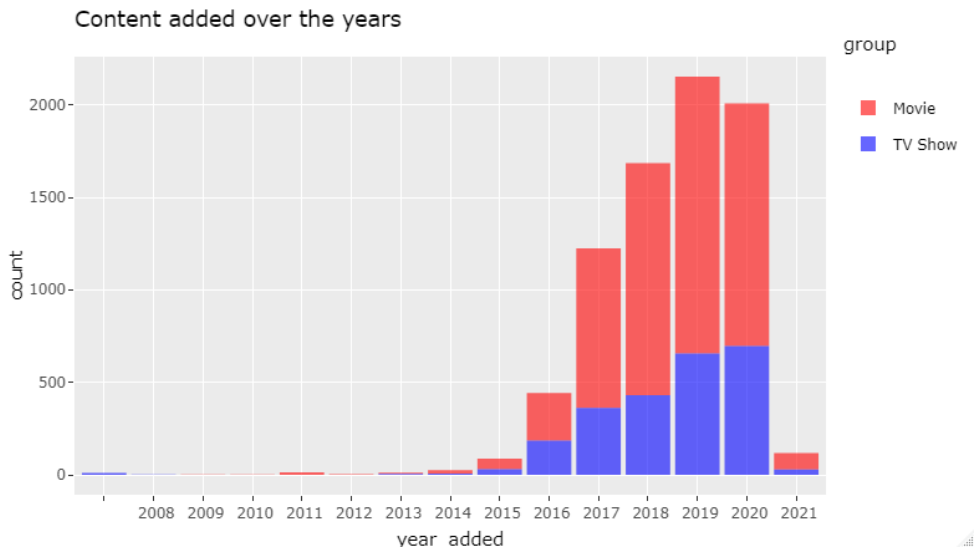

Figure 1: A bar shart to show ontent added to netflix between 2008 and 2021 comparing the number of movies released to the number of TV shows released

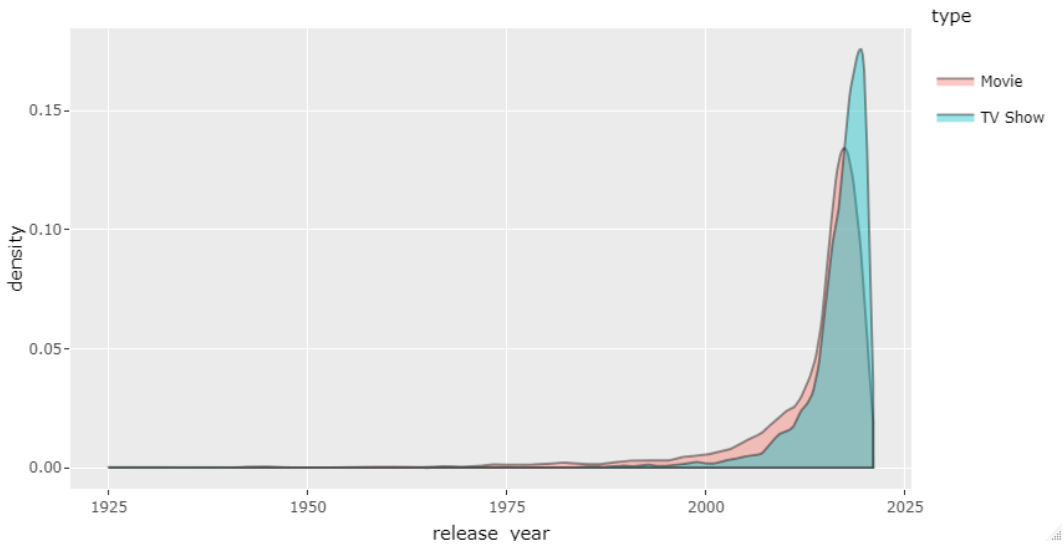

Figure 2: A Density graph to show the density of both movies and TV show releases from 1925 to 2021

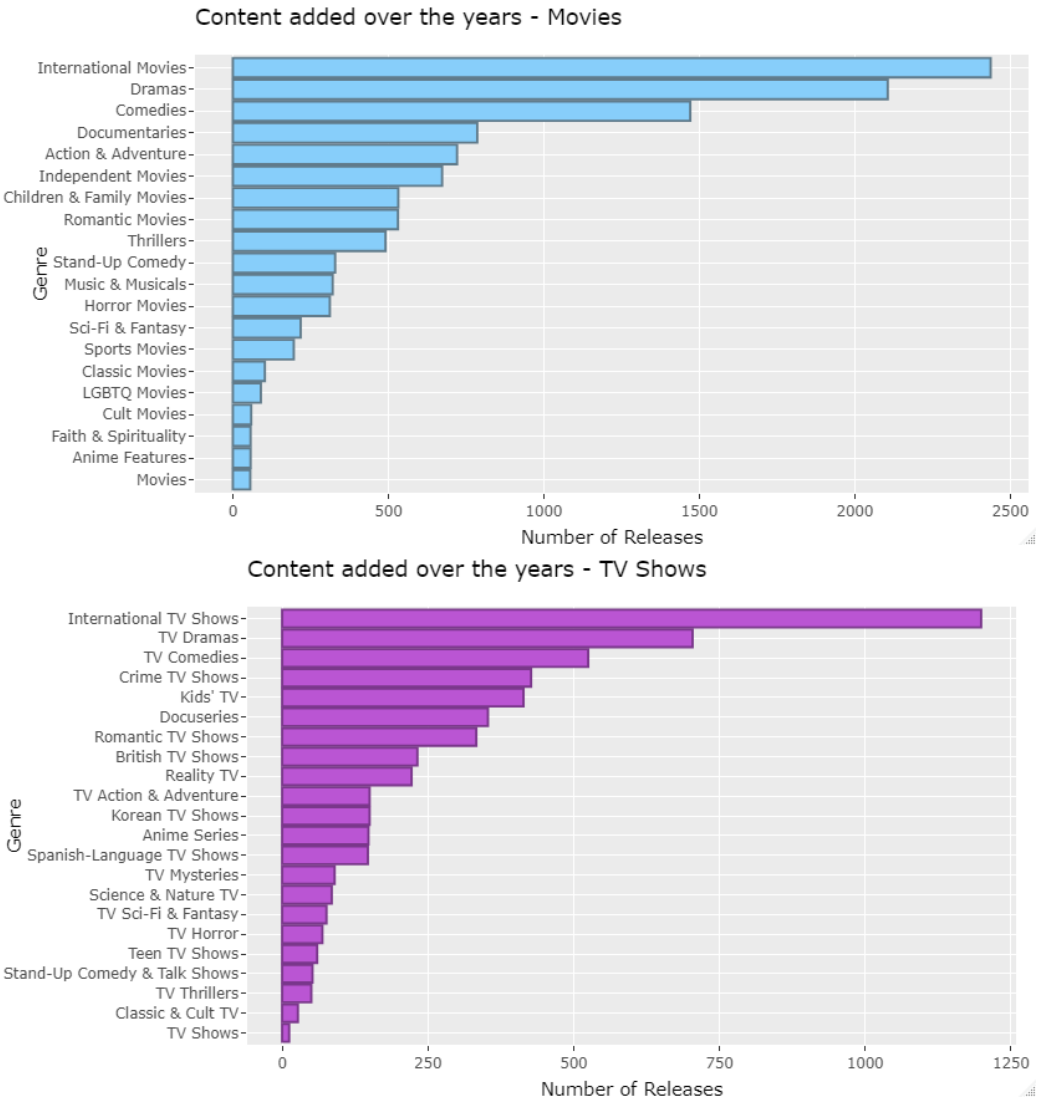

Figure 3: Bar charts showing the number of releases tagged with a certain genre in movies and TV shows

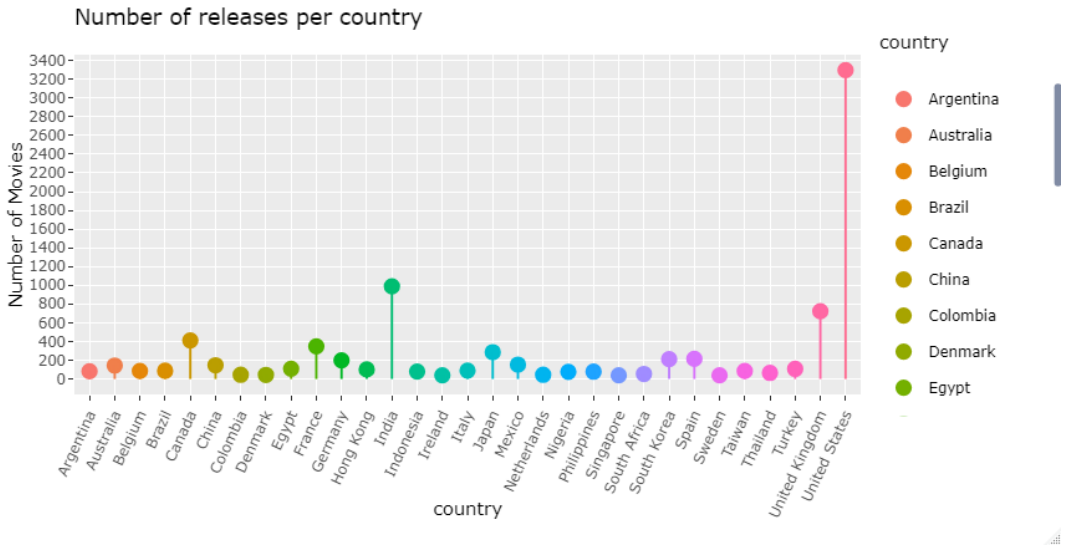

Figure 4: A lollipop chart to show the number of movies each country has released from 1925-2001

4 Summary

4.1 How has the number of TV-shows and movies added by Netflix changed over time? (Figure 1)

From 2008 to 2019 the total amount of content has increased each year, only falling after 2019. This could be due to the coronavirus and fewer new movies and TV shows being created through 2020, meaning there was less of a selection for Netflix to release as well as delays of current shows releases on2020/2021 due to the regulations set in place. Also, 2021 will have a lower amount as this graph was made earlier on in the year, meaning that there is still more time to add movies and Netflix may not have uploaded many in the beginning. We can also see that in most years more movies were added than TV shows and between 2017 and 2020 there was a much larger difference between the two. In conclusion, we can see that releases have increased overall and have only dropped due to the timing of the data-set and external factors.

4.2 how has density of content changed over time? (Figure 2)

subsectionhow has the density of content changed over time? Figure 2 Over time more and more movies and TV shows have been released and the graph shows there was a large increase in density between 2000 and 2021 (The graph goes to 2025 as it fits the scale) for both movies and TV shows, this is definitely due to the growing culture movies and television has had within the more recent years where more and more large industries release many films and TV shows each year. We can also see that TV shows have a higher density which could be due to TV shows having much more overall content as many have multiple seasons which would be counted separately, if seasons were not counted separately I think the distributions would be a lot closer together. In conclusion, as time increases density has also increased. //

4.3 what is the variation within the genres of movies and TV-shows over all?(figure 3)

We can see that although both graphs have different genres in their order, the distribution between genres stays the same with a few larger results at the top then the rest being quite low in comparison, both graphs have a similarity to each other but overall variation is quite low because there are such big differences between the top and the bottom halves of the graphs. This could be because lots of separate content has been tagged with a common genre such as international along with some other less- common ones so overall I would say that variation is low within the genres.

4.4 Is the number of releases around the world equal? (figure 4)

The number of releases could be equal but there are a few countries that do have much higher values, without these, the distribution would be quite equal as the other countries all have a similar amount of releases. But with the top set included, distribution is low. I would say that this is due to the top countries being home to some of the biggest industries in film/TV but In conclusion, the number of releases is not equal. There are many factors that will affect this result such as the number of production companies in the country and how much they earn as lesser earning companies will not be able to produce as much content as higher-earning companies.

4.5 Main findings in this report:

Within my report I have found that as time has increased the overall amount of content has grown rapidly, showing me that up to today the film and TV industry is growing and thriving as of recent times. I can also see that Netflix is also expanding its library of content and will probably continue to grow. Finally looking into the final graph (Figure 4) I can see that there are three countries that provide most of the content in these days and further research shows that they have some of the highest profiting movie industries originating from them. So in conclusion, Netflix has always been growing its content, more and more TV shows and movies are being released as time passes, there are a lot of genres being used but some more than others and finally, there are a top set of countries that are releasing content and without them, the rest are equally releasing their own among the lower numbers.