**Week 8 Project: Analysis of Netflix Content between 2008 and 2022**

Kaytie Moyer

Anderson College of Business and Computing

Regis University

MSDS 670: Data Visualization

John Koenig

October 14, 2021

**Summary**

The Week 8 Project showcases data taken from Netflix’s public domains between the years 2008 and 2022. Netflix could use this data on multiple points; to determine success of content released, where categories of content are most popular, and where Netflix focuses on their content releases. The research questions to this end are: is Netflix releasing more content in some countries then others? Does Netflix focus on releasing movies or TV shows? Has there been a year between 2008 and 2022 where Netflix has released more content?

**Data**

The data was sourced from Netflix’s public domains and outlying websites that collect the applicable contents. The author of the csv file used was Shivam Bansal, and is linked in the reference section. The data is updated every thirty days.

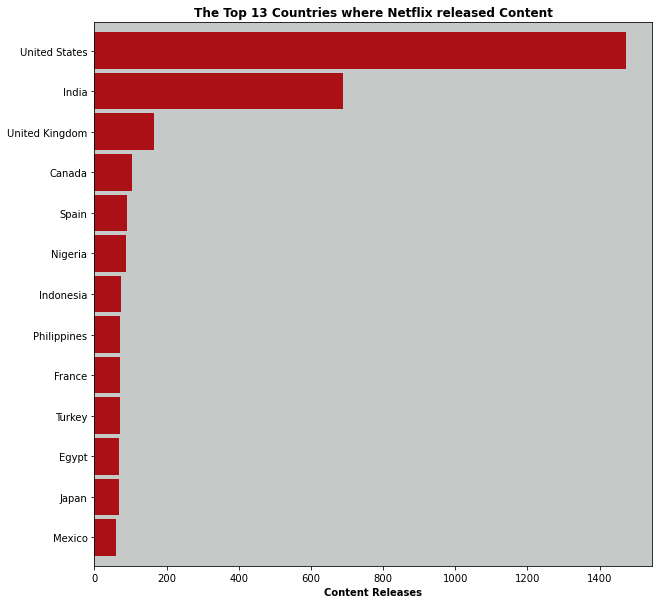
Netflix collects and stores the data on the content they release which includes the directors, categories, types of content, ratings, locations released, and year the content was released. While this is available as public domain data, it can also be viewed on Netflix’s application and the pertaining website as well as other collective domains.

The data was cleaned of NA values, and data from before 2008 was filtered out. Categories of data were placed into specific dataframes for analysis; for example “directors” were value counted and placed under the dataframe “directors”. Dates were also formatted into a pandas dataframe for easier analysis. The following dataframes were used to group data.

* Director – directors involved in content creation with Netflix.
* Category – content categories used for Netflix content.
* Country – the parts of the world where Netflix released content.
* Year\_r – the year that content was released.
* Rating – the rating given to Netflix content upon release.

***Analysis of where Netflix released Content***

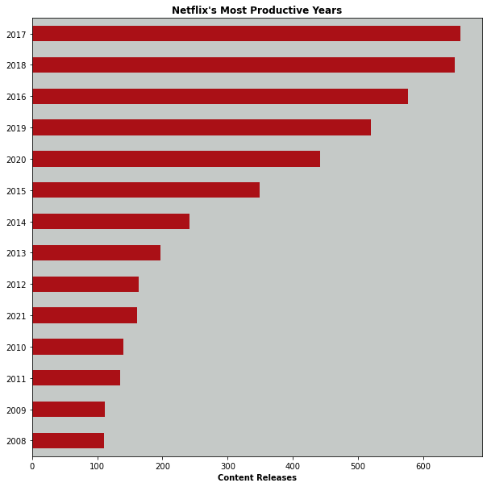
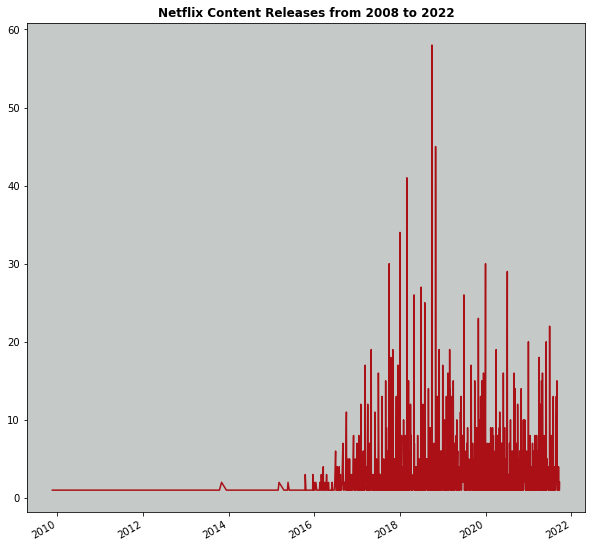
Between the years of 2008 and 2022, Netflix released more content in the United States then any other country. India, by comparison, released half of the content that the United States did, but still released over 100% more content than the other eleven countries included in the chart. Interestingly enough, three of the four countries that released the most amount of content were primarily English-speaking countries. This could possibly mean that language is a correlation to where content is released, but further analysis is needed.



***Netflix Releases throughout the Years***

According to the chart plotting data from 2008 to 2022 of content release from Netflix, the highest spike occurs after 2018, around the time frame of 2019. The chart also shows that before 2014, no content was released and between 2014 and 2016, only a few occurrences of content were released. After 2016, content release increased, spiking during 2019, and then decreasing sharply by 2020. Content releases remained somewhat steady during 2020, beginning to decline towards the end of 2021. This data can be re-evaluated in December of 2021 or January of 2022 in order to determine if this trend carries.

A horizontal bar chart did not confirm these initial results, and instead pointed to 2017 and 2018 as the highest content release years for Netflix, with 2019 coming in fourth on the plot. So far, Netflix has released ¼ the amount of content in 2021 that was released in 2017. 2008, the first year included in the data analysis, was the lowest year for content release.



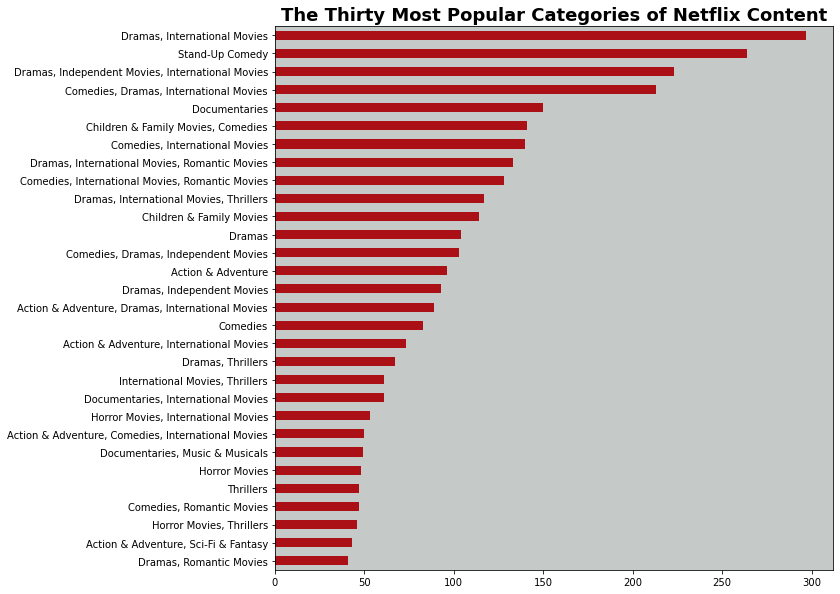
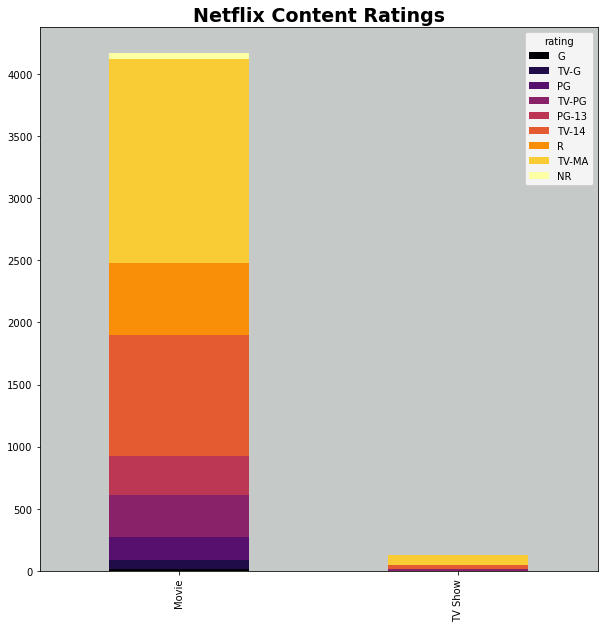
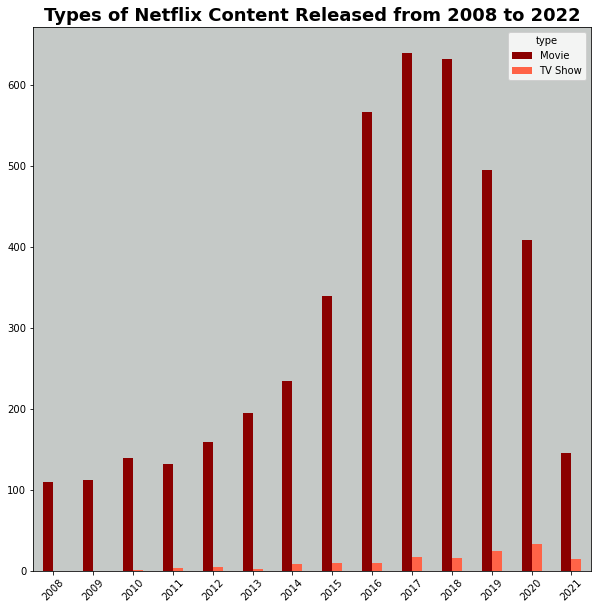
***The Focus of Netflix’s Content Creation***

Between 2008 and 2022, we created a plot that analyzed whether movies or television shows were released more. The chart showed that movies were the overwhelming focus for Netflix’s content creation, with television shows remaining below one-tenth the number of movies. The highest spike for television show release was in 2020, while the highest movie content release was the year 2017. During 2021, movie releases decreased by about 125%, with television show releases decreasing by 100%.

Looking at the ratings for the movies and television shows released, over 50% of the movies that Netflix produced were rated TV-PG, and the second highest rating shown was TV-MA. This indicates that movies from Netflix focused on family ratings and younger audiences, or adults. Television shows on the other hand, were in the majority of TV-MA, with about 25% of that number being TV-14. This indicates that television shows are more heavily marketed towards adolescents and adults.

Finally, we looked at the thirty most popular categories of Netflix content. Each movie and television show is assigned multiple categories to help viewers be as specific as possible when browsing for content. Thus a movie that is placed in the Drama/Romantic category is different than a movie in the Drama/Romantic/International category.

The most popular category of Netflix content released was from the Drama/International category. The second most popular category was stand-up comedy. Interestingly enough, three of the four most popular categories were international movies. Another indicator of the popularity of international films is that Drama/Romantic/International category-related movies are about 100% more popular than the Drama/Romantic category.



**Conclusion**

In conclusion, between the years of 2008 and 2022 Netflix released the most content in 2017 and 2018. The large majority of this content were movies, most of which were international dramas and stand-up comedy specials. Early analysis indicates that this content is produced for multiple age demographics, primarily families and adults for movies, and adolescents and adults for television shows. Since 2018, Netflix has declined in content release, though this trend needs to be re-analyzed after the end of 2021 to determine if it continues.

Areas of future research could look at determining more detailed demographics for released content. For example, areas could be determined to see if the released content is reaching target demographics. Watch times could also be collected and analyzed to see how long viewers remain with chosen content before switching to something else. The streaming platform Youtube does something similar with watch times on videos to show creators how long viewers stay on a particular video. Finally, data could be collected on the number of times viewers return to content after finishing it the first time. This could be used to indicate the “fan-favorites” and popularity of content after initial viewings.

**References**

* <https://translatorswithoutborders.org/language-data-for-india> (data on the popularity of languages in India)
* <https://www.kaggle.com/shivamb/netflix-shows> (the location of the original dataset)
* <https://www.britannica.com/topic/Netflix-Inc> (information on the founding of Netflix and its growth)
* <https://worldclass.regis.edu/d2l/le/content/273433/viewContent/4017190/View> (examples of a previous student’s presentation that was used to help determine the format of my presentation)