

## **What Netflix Offers and IMDb Score Prediction**

The global video streaming market was valued at 229.19 billion dollars in 2018 and is expected to reach 742.74 billion dollars by 2025 with projected growth of 18.29 percent year over year.<sup>i</sup> This projected growth can be attributed to viewers' increased affinity for streaming entertainment and the wide field of streaming options available. While the industry is quickly becoming crowded with new competitors, Netflix Inc has been the pioneer of this field since it was founded in 2007. Netflix is the world's largest streaming service with over 207 million global subscribers.<sup>ii</sup> Netflix has such a pervasive reach that researchers estimate 23% of adults in the United States watch Netflix daily.<sup>iii</sup>

### **Description of Project Goals**

A dataset on Netflix Movies and TV Shows 2021 provided by Kaggle.com was investigated. The dataset included 5,967 observations and 13 variables: Show Id, Title, Description, Director, Genres, Cast, Production Country, Release Date, Rating, Duration, IMDb Score, Content Type, and Date Added.

The questions being investigated include: (1) Are there any identifiable trends in the content Netflix offers to subscribers? (2) What factors in the dataset can affect the IMDb score of a movie/TV show? The goal was to explore trends in the types of content that Netflix offers by looking at Genre, Content Type, Production Country, Rating, etc. IMDb scores were used to determine if any of the variables were suitable predictor variables to predict how viewers would rate movies and TV shows. This report will provide interesting insights on a topic that is relatable to over 20% of the United States population that are Netflix subscribers. The insights gained could help current and prospective Netflix users better understand the array and quality of content that Netflix offers to them.

### **Exploratory Analysis**

#### **Data Cleaning**

The variables of interest included Genre, Rating, Duration, Production Country, IMDb score, Release Time, and Date Added. Several steps were then taken to clean the dataset including: removing missing values, converting data types, categorizing instances where multiple production countries were listed, separating composite data, and creating new variables (Appendix 1).

Finally, the dataset was separated into different data frames, one consisting of Movies and the other consisting of TV Shows since they show many obvious differences in aspects such as Genre categories and Rating Scales. This was to ensure that both content types could be analyzed independently.

### **Trends in the content Netflix offers to subscribers**

The dataset showed that as of 2021, there are 3,497 unique movies and 716 unique TV shows offered on the Netflix platform. Based on this information a series of questions were proposed to dive deeper into the trends and characteristics of the content on Netflix. Genre, Production Country, and Release Time were the three specific variables of initial interest. These three categories showed the potential to provide interesting insights and present trends in the content to be explored through modeling. The exploratory analysis into Genre showed that the largest Movie genres offered on Netflix consist of Dramas, Comedies, and Action movies (Appendix 2). The same analysis of the TV Show data found that Kids TV, International TV, and Crime TV shows are the largest genres offered by Netflix (Appendix 3). The Genre analysis prompted exploration of the lag time between when entertainment content is released, to when the content is added by Netflix. Interestingly, even though Dramas and Comedies are the most abundant on Netflix, release lag time for these categories is above the median average lag time of 2.65 years (Appendix 4). The release lag time for Kid's TV and International TV Shows fell below the median average lag time for TV Shows of 2.0 years (Appendix 5). While Netflix offers a large number of Dramas and Comedies, these genres are less likely to have newer releases. This insight is the opposite for the TV Show market, where Netflix seems to be making an effort to ensure that its largest TV Show genres are also newly released.

### **Trends in internationally produced offerings**

Facing increased competition from Hulu, Amazon, and Disney+, Netflix sought to further expand its presence in the international markets and win subscribers.<sup>iv</sup> To achieve this goal, Netflix has offered internationally produced content at a much faster rate than US-produced ones (Appendix 6). In general, Dramas, Comedies, and Documentaries took up the majority of the internationally-produced offerings while those from Japan had a heavy focus on Anime-related Genres (Appendix 7). Digging deeper into the Genre distribution of international and US-produced offerings, the genres of US-produced offerings are more evenly distributed compared to offerings produced by other countries (Appendix 8). Specifically, looking at the Genre distribution comparison between the US and India, the largest contributor to Netflix's internationally produced offerings. Both the US and India had Dramas and Comedies as one of their major genres. However, these two genres only

took up 27% of the US production while they took up over 70% of the Indian production, showing internationally produced content was heavily skewed towards a few genres (Appendix 9). Historically, Netflix started offering a few genres when introducing internationally produced content. Then, they diversified genres as offerings increased. For example, there were only two genres when Netflix began offering Indian-produced content in 2016, by 2020 the genres offered had increased to 12 (Appendix 10). Netflix is expected to continue expanding its international content portfolio in 2021 as they continue to be involved with global product offerings to fend off competition.<sup>v</sup>

### **Factors that can affect IMDb score**

IMDb Score is a very important criterion for a movie/TV show as many people select what to watch based on IMDb Score. The top and bottom 10 movies and TV shows were chosen to get a high-level understanding of the distribution of IMDb Score and its relationship with the other factors. In the movie field, Documentaries are inclined to receive higher IMDb Scores, while Action & Adventure movies receive lower scores. Regarding TV shows, higher scores are evenly distributed among more than 3 genres, while lower scores are concentrated on Kid's TV (Appendix 11).

However, the top and bottom 10 utilized a small sample, and information extracted from a small sample can be biased. For instance, there was no obvious difference in Rating and Production Country when comparing the top and bottom 10. Thus, Production Country, Genre, and Ratings with the highest average IMDb Score were further screened out and plotted against Duration, Age, and IMDb Score, to obtain more insights (Appendices 12 and 13).

Based on the above, the preliminary assumptions were made: (1) Genre, Ratings, and Age could be potential statistically significant predictor variables for IMDb Score; (2) for TV shows, Duration may also contribute to IMDb Score; (3) Production Country does not provide significant information for IMDb Score prediction.

### **Model Establishment**

Running a regression on IMDb Score and content type, it was shown that there was a statistical difference between TV and movies. Both models started with all variables available in order to predict IMDb Scores. Based on the model summary all continuous variables that were not statistically significant were removed. The regressions were then rerun after removing one of the categorical variables and evaluating the change in adjusted R<sup>2</sup>. If the adjusted R<sup>2</sup> lowered significantly it was added back in. The full TV model started with an adjusted R<sup>2</sup> of 0.187 with the final model ending with an R<sup>2</sup> of 0.187 using

Genre, Ratings, and Duration. This means that those three variables can account for 18.7% of the variance in IMDb Scores. Based on an F-statistic of 9.2 (p-value =  $7.18e-25$ ) it is statistically better than a null model. The full movie model started with an adjusted R<sup>2</sup> of 0.233 with the final model ending with an R<sup>2</sup> of 0.230 using Genre, Production Country, Age, number of words in Title, and Duration. This means that those five variables can account for 23% of the variance in IMDb scores. Based on an F-statistic of 13.7 (p-value =  $2.72e-154$ ) it is statistically better than a null model. While both models are better than a null model, the test statistics indicate that these models may not be that much better.

### **Conclusion and Insights**

In the last 5 years, the growth in the number of movies on Netflix is much higher than that of TV shows, indicating that Netflix has increasingly focused on movies rather than TV shows.

Netflix offers many Dramas and Comedies in the Movie field, however, none of them have received a high average IMDB score. When looking at TV shows, the largest Genre offered is Kid's TV, but the quality is relatively poor.

International content offerings on Netflix are rising much faster, but heavily skewed towards Dramas and Comedies. However, Netflix is expected to continue expanding its international content portfolio in 2021 to fend off market competition. In the coming years, there will not only be more international TV shows and movies to watch but also more genres available to choose from.

Finally, based on the model, Parental Rating and Genre are relatively strong predictors of IMDb Score. If a subscriber is looking for high-quality content on Netflix they should look towards the Documentaries Genre for movies, and stay away from the Kid's TV Show offerings.

## **Appendix 1: Data Cleaning Process**

Data cleaning is the process of ensuring data is correct, consistent, and usable. To achieve the project goals and perform quantitative analysis, our data cleaning approach includes the following steps:

- Dropped the "ID", "Cast", "Director", and "Description" columns since these columns contain most missing values and unique information which is not beneficial for data exploratory and quantitative analysis.
- Removed entries containing null values based on the first step, therefore, most of the information is preserved.
- Converted "Release Date" and "Date Added" into time series variables.
- Stripped fraction notation from "IMDb Score", so that we can perform functions and methods such as filtering, sorting, grouping, etc.
- Edited "Genre" to only include the first Genre listed.
- Categorized instances with multiple production countries listed into "Multi-Country Production".
- Created "Age" variable to represent the time elapsed between "Release Date" and "Date Added".
- Cleaned Duration column to remove minute and season notation.
- Split "Title" into a list of strings and created "num\_words\_title" variable to count the number of words in titles.

## Appendix 2: Top 10 Largest Movie Genres offered on Netflix

Movie Genre	Frequency
Dramas	953
Comedies	758
Action & Adventure	468
Documentaries	391
Children and Family Movies	295
Stand Up Comedy	285
Horror Movies	143
International Movies	82
Thrillers	32
Classic Movies	19

### Appendix 3: Top 10 Largest TV Show Genres Offered on Netflix

TV Show Genre	Frequency
Kid's TV	149
International TV Shows	111
Crime TV Shows	103
British TV Shows	74
TV Comedies	64
Anime Series	57
Docuseries	32
Reality TV	32
TV Dramas	29
TV Action & Adventure	20

#### Appendix 4: Average Number of Years between Movie Release Time and Date Added

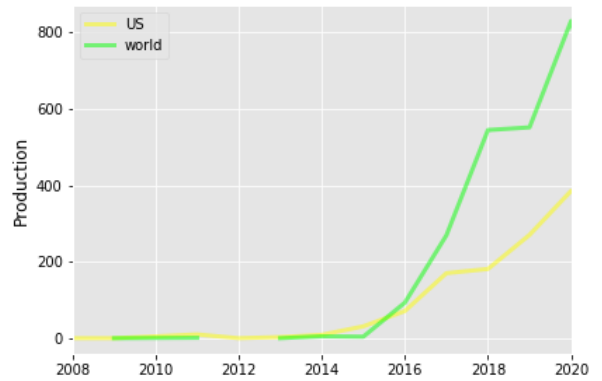
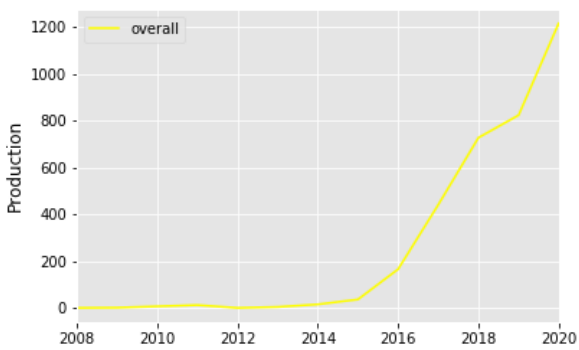
Movie Genre	Average Number of Years
Classic Movies	40
Cult Movies	26.714286
Action & Adventure	8.307692
Thrillers	5.65625
Comedies	4.889182
Dramas	4.027282
Movies	3.666667
Children & Family Movies	3.583051
Horror Movies	2.734266
International Movies	2.585366
Independent Movies	2.583333
Stand-Up Comedy	2.357895
Anime Features	2.117647
LGBTQ Movies	2
Romantic Movies	2
Music & Musicals	1.928571
Documentaries	1.810742
Sci-Fi & Fantasy	1.5
<b>Median</b>	<b>2.659815</b>



## Appendix 5: Average Number of Years between TV Show Release Time and Date Added

TV Show Genre	Average Number of Years
Classic & Cult TV	9.181818
Anime Series	5.035088
Romantic TV Shows	3.533333
TV Action & Adventure	3.35
Reality TV	3.34375
TV Comedies	2.59375
British TV Shows	2.189189
Kids' TV	2
International TV Shows	1.495495
TV Dramas	0.965517
Crime TV Shows	0.660194
Stand-Up Comedy & Talk Shows	0.583333
TV Horror	0.5
Docuseries	0.46875
TV Sci-Fi & Fantasy	0
<b>Median</b>	<b>2.0</b>

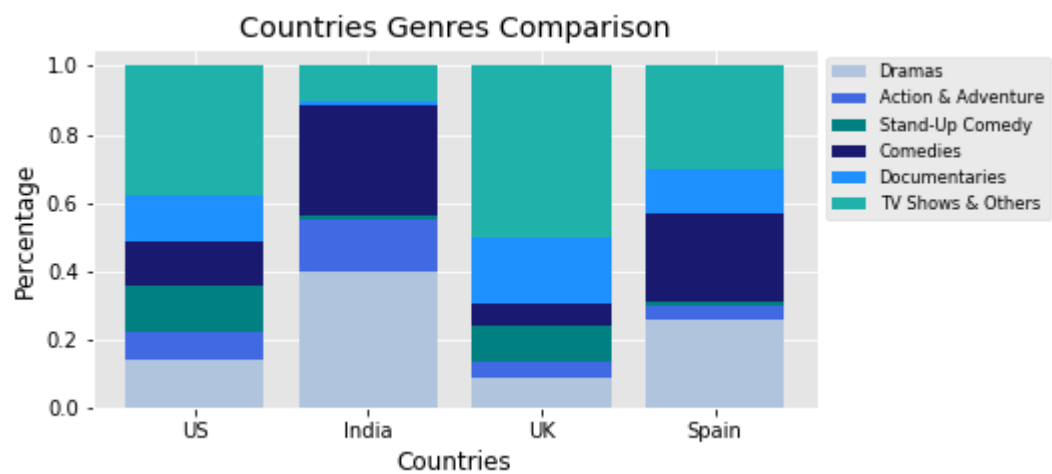
## Appendix 6: Overall offerings trend, US-produced offerings vs world-produced offerings trends



## Appendix 7: Top genres of 5 countries that produced the most Netflix offerings

US	India	UK	Japan	Spain
Dramas	Dramas	British TV Shows	Anime Series	Comedies
Documentaries	Comedies	Documentaries	Action & Adventure	Dramas
Stand-Up Comedy	Action & Adventure	Stand-Up Comedy	Anime Features	Documentaries
Comedies	International Movies	Dramas	Dramas	International Movies
Children & Family Movies	Horror Movies	Comedies	International TV Shows	Crime TV Shows

**Appendix 8: Genre distribution of internationally produced content**



## Appendix 9: Production Comparison: US vs India

Content Type	US	India
Movie	79%	99%
TV Show	21%	1%

Genres	US	India
Dramas	14.3%	39.9%
Comedies	12.9%	32.4%
Action & Adventure	7.9%	15.1%
International Movies	0.0%	4.2%
Horror Movies	3.0%	2.4%
Children & Family Movies	10.1%	2.0%
Documentaries	13.7%	1.4%
Stand-Up Comedy	13.6%	1.0%
Classic Movies	0.3%	0.3%
Docuseries	1.6%	0.3%

## Appendix 10: Genre distribution trend for Indian produced content

2016		2017		2018		2019		2020	
Genre	#	Genre	#	Genre	#	Genre	#	Genre	#
Dramas	8	Action & Adventure	19	Dramas	73	Dramas	55	Comedies	63
Horror Movies	1	Dramas	19	Comedies	63	Comedies	39	Dramas	55
		Comedies	11	Action & Adventure	31	Action & Adventure	14	Action & Adventure	20
		Stand-Up Comedy	1	Horror Movies	5	International Movies	7	International Movies	10
		Children & Family Movies	1	International Movies	5	Horror Movies	2	Horror Movies	6
		Classic Movies	1	Children & Family Movies	5	Documentaries	2	Stand-Up Comedy	5
				Documentaries	2	Children & Family Movies	1	Children & Family Movies	5
						International TV Shows	1	Docuseries	2
								Documentaries	2
								Kids' TV	1
								Cult Movies	1
								Classic Movies	1
<b>Total</b>	<b>9</b>		<b>52</b>		<b>184</b>		<b>121</b>		<b>171</b>

## Appendix 11: Overview of Top/Bottom 10 Movies/TV Shows

Top 10 Movies	Genres	Production Country	Rating	Duration	Age	Imdb Score
Bo Burnham: Inside	Stand-Up Comedy	United States	TV-MA	88	0	9.0
No Longer kids	Comedies	Egypt	TV-14	237	42	9.0
David Attenborough: A Life on Our Planet	Documentaries	Multi Country Production	PG	84	1	9.0
In Our Mothers' Gardens	Documentaries	United States	TV-MA	85	1	8.9
Care of Kancharam	Comedies	India	TV-14	142	3	8.9
Sky Tour: The Movie	Documentaries	Vietnam	TV-G	93	1	8.8
Bye Bye London	Comedies	Kuwait	TV-14	177	40	8.8
Unbroken	Documentaries	India	TV-14	109	2	8.8
KD (A) Karuppudurai	Comedies	India	TV-14	116	2	8.7
Merku Thodarchi Malai	Dramas	India	TV-MA	115	3	8.6

Items	Minimum	1 <sup>st</sup> Quartile	Median	3 <sup>rd</sup> Quartile	Maximum
<b>IMDb Score</b>	8.60	8.80	8.85	8.975	9.00
<b>Duration</b>	84.00	89.25	112.00	135.50	237.00
<b>Age</b>	0.00	1.00	2.00	3.00	42.00

### Appendix 11: Overview of Top/Bottom 10 Movies/TV Shows (con't)

Bottom 10 Movies	Genres	Production Country	Rating	Duration	Age	Imdb Score
Aerials	International Movies	United Arab Emirates	TV-14	90	5	0.3
Time to Dance	Dramas	India	TV-PG	107	1	0.7
Himmatwala	Action & Adventure	India	TV-14	140	8	0.8
Cross: Rise of the Villains	Action & Adventure	United States	R	100	2	0.9
5Gang	Comedies	Romania	TV-14	96	2	0.9
Mythily Veendum Varunnu	Horror Movies	India	TV-PG	88	4	2.0
Hajwala: The Missing Engine	Action & Adventure	United Arab Emirates	TV-PG	97	5	2.2
B.A. Pass 2	Dramas	India	TV-14	126	4	2.3
Deewana Main Deewana	Comedies	India	TV-14	107	8	2.3
Let's Dance	Comedies	Egypt	TV-14	116	15	2.4

Items	Minimum	1 <sup>st</sup> Quartile	Median	3 <sup>rd</sup> Quartile	Maximum
<b>IMDb Score</b>	0.30	0.825	1.450	2.275	2.40
<b>Duration</b>	88.00	96.25	103.50	113.75	140.00
<b>Age</b>	1.00	2.50	4.50	7.25	15.00



### Appendix 11: Overview of Top/Bottom 10 Movies/TV Shows (con't)

Top 10 TV Shows	Genres	Production Country	Rating	Duration	Age	Imdb Score
Breaking Bad	Crime TV Shows	United States	TV-MA	5	8	9.4
Avatar: The Last Airbender	Classic & Cult TV	United States	TV-Y7	3	14	9.3
The Last Dance	Docuseries	United States	TV-MA	1	1	9.2
Leah Remini: Scientology and the Aftermath	Docuseries	United States	TV-14	3	3	9.0
Hunter X Hunter (2011)	Anime Series	Japan	TV-14	4	8	9.0
Fullmetal Alchemist: Brotherhood	Anime Series	Japan	TV-14	5	11	9.0
Leyla and Mecnun	International TV Shows	Turkey	TV-MA	3	7	9.0
Monty Python's Flying Circus	British TV Shows	United Kingdom	TV-14	4	47	8.9
Black Mirror	British TV Shows	United Kingdom	TV-MA	5	2	8.9
Dark	Crime TV Shows	Multi Country Production	TV-MA	3	1	8.9

Items	Minimum	1 <sup>st</sup> Quartile	Median	3 <sup>rd</sup> Quartile	Maximum
<b>IMDb Score</b>	8.90	8.9250	9.00	9.15	9.40
<b>Duration</b>	1.00	3.00	3.50	4.75	5.00
<b>Age</b>	1.00	2.25	7.50	10.25	47.00

## Appendix 11: Overview of Top/Bottom 10 Movies/TV Shows (con't)

Bottom 10 TV Shows	Genres	Production Country	Rating	Duration	Age	Imdb Score
Richie Rich	Kids' TV	United States	TV-G	2	6	3.2
Bill Nye Saves the World	Stand-Up Comedy & Talk Shows	United States	TV-14	3	3	3.4
Droppin' Cash: Los Angeles	Reality TV	United States	TV-MA	1	2	3.4
Little Singham	Kids' TV	India	TV-Y7	2	1	3.5
Some Assembly Required	Kids' TV	Canada	TV-Y	2	6	3.5
Cloudy with a Chance of Meatballs	Kids' TV	Multi Country Production	TV-Y7	1	4	3.6
Reboot: The Guardian Code	Kids' TV	Canada	TV-Y7	2	3	3.6
Millennials	International TV Shows	Argentina	TV-MA	3	1	3.8
Yummy Mummies	International TV Shows	Australia	TV-MA	2	2	3.9
Teen Mom 2	Reality TV	United States	TV-14	2	10	4.0

Items	Minimum	1 <sup>st</sup> Quartile	Median	3 <sup>rd</sup> Quartile	Maximum
<b>IMDb Score</b>	3.20	3.425	3.55	3.75	4.00
<b>Duration</b>	1.00	2.00	2.00	2.00	3.00
<b>Age</b>	1.00	2.00	3.00	5.50	10.00

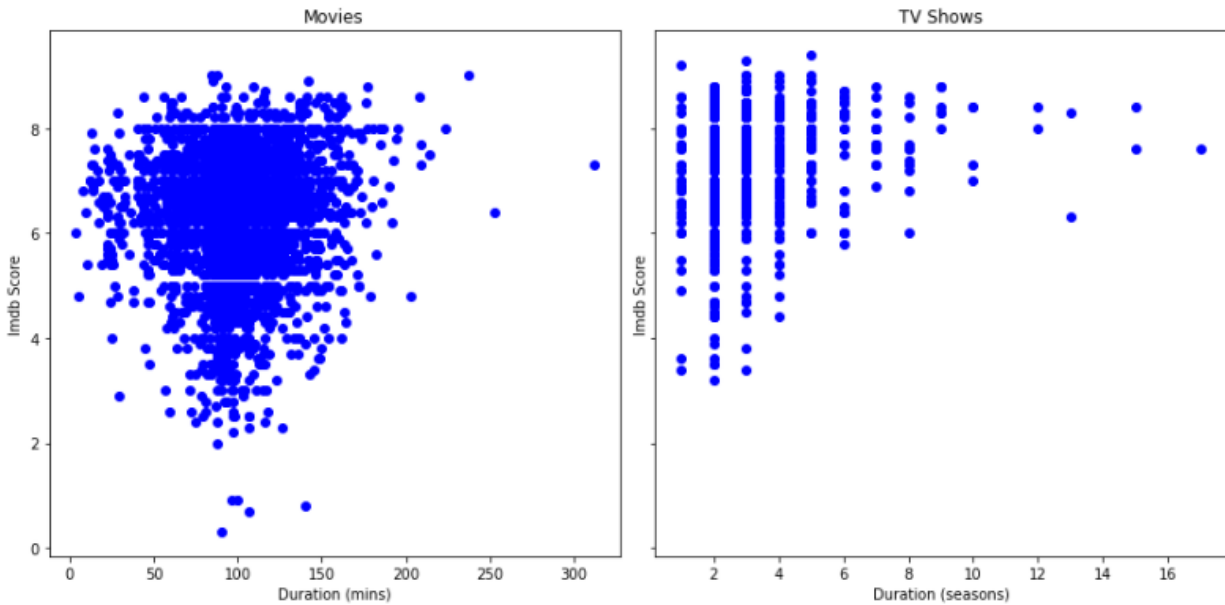
## Appendix 12: Top 5 Countries, Genres, and Ratings with Highest Average IMDb Score

Movie	Production Country	Genres	Rating
1	Iceland	Music & Musicals	TV-PG
2	Palestine	Classic Movies	PG-13
3	West Germany	Documentaries	TV-MA
4	Israel	Anime Features	TV-14
5	Ireland	Movies	PG

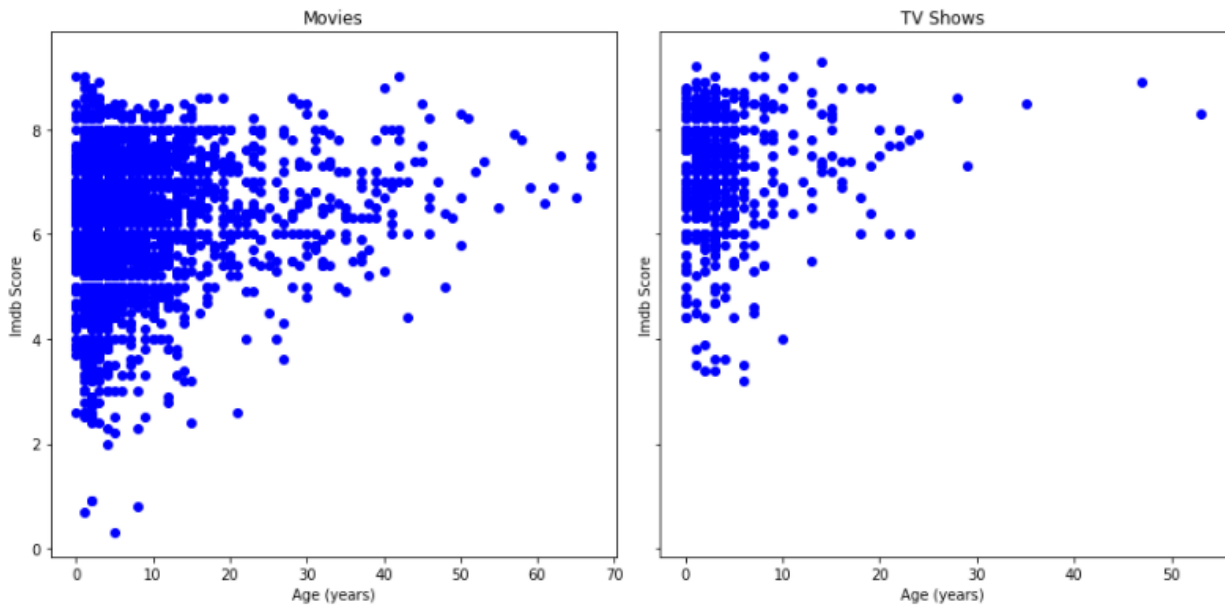
TV Show	Production Country	Genres	Rating
1	Malaysia	Classic & Cult TV	TV-MA
2	Chile	Docuseries	TV-14
3	Lebanon	Crime TV Shows	TV-PG
4	Indonesia	TV Sci-Fi & Fantasy	TV-Y7
5	Croatia	British TV Shows	TV-Y

## Appendix 13: Scatter Plots

- Relationship between Duration and IMDb Score



- Relationship between Age and IMDb Score



## End Notes

---

<sup>i</sup> Brand Essence Research., "Video Streaming Market 2021 by Type of Video Streaming," (April 2020), <https://brandessenceresearch.com/technology-and-media/video-streaming-market-size> , accessed August 2021.

<sup>ii</sup> Spencer, Samuel., "How Many Subscribers Do Netflix, Disney+, and the Rest of the Streaming Services Have?," *Newsweek*, (May 2021), <https://www.newsweek.com/netflix-amazon-hulu-disney-most-subscribers-streaming-service-1590463>, accessed August 2021.

<sup>iii</sup> "Netflix by the Numbers: Statistics, Demographics, and Fun Facts" *MuchNeeded*, (January 2021), <https://muchneeded.com/netflix-statistics/>, accessed August 2021.

<sup>iv</sup> Strain, Mike., "Two Challenges Netflix Faces From Rising Competition," *The Motley Fool*, (August, 2019), <https://www.fool.com/investing/2019/08/26/two-challenges-netflix-faces-from-rising-competiti.aspx>, accessed August 2021.

<sup>v</sup> Zacks Equity Research., "Netflix (NFLX) Rides on International Content Growth in 2021," *Nasdaq*, (May, 2021), <https://www.nasdaq.com/articles/netflix-nflx-rides-on-international-content-growth-in-2021-2021-05-27>, accessed August 2021.