

Full Name: Felicia Huffman

DATA1201 Final Exam

Date: November 30, 2023

Topic: The qualities of highly-rated Netflix films

Question: What are common qualities of highly rated Netflix films?

Question

Netflix is a popular streaming service that many people around the world use to find movies, shows, and other streaming content. Netflix constantly has a top ten list of the top ten movies being watched on any given day. Many of these films are popular and have very high ratings, but why? What makes some movies better than others? This question will be pondered and analyzed throughout this portfolio.

My scientific goal for this project was to find and explain the similarities between highly rated Netflix films in order to explain why some the pattern of films that do significantly better than others. I execute this goal by finding a dataset related to my question and using it to answer it.

The question I ask and explain in the remainder of this portfolio is “What are qualities of highly rated Netflix films.” However, this question was derived from my initial question: “What is the most watched show on Netflix this week and how does that compare to previous most-watched shows.” One of the main reasons my question evolved so much was because of the lack of a single dataset to answer my initial question. I was able to find a Netflix dataset on Kaggle with ratings along with traits such as genre, certificate, and duration for Netflix films. This data was good for analyzing what qualities make a highly rated movie but did not necessarily provide data on the most-watched shows, so I decided to answer a smaller portion of my question that could be answered using a single dataset. This data could be used to help answer my initial question by providing information on what traits make a highly-rated movie. I could have answered my initial question using multiple datasets, but I believe it was outside the scope of this project.

I came up with my initial question out of curiosity; I wanted to know what makes some shows and movies rise to the top of the watch lists and if there were any similarities between them. I thought the information from this dataset could be particularly useful to someone in the film industry such as a Netflix director looking to make a good profit on a film. This data could help them decide what kind of film to produce, what duration time is best, and what audience they should target in order to be as successful as possible.

I want to predict what traits a movie should have in order to have the greatest chance at being successful. I do that by creating visuals that show which qualities tend to be rated the highest. I started by finding a dataset that was appropriate for answering my question, manipulating that dataset to include information that is necessary to answer the question, and then creating tables and visualizations to help answer my question. After viewing the visualizations, I was able to use the data to answer my question and make my prediction.

Getting the Data

In order to answer my question, I need data regarding qualities of highly rated Netflix movies. This includes variables such as rating, duration, genre, and certificate. I got this data from an Excel sheet on Kaggle, which received it from a Netflix database. I am looking specifically for any correlations between ratings and other factors, such as genre and year. I will have to modify my data to be more useful for my application. For example, I need to separate the unit min from the duration, and separate the genres. I can do this by manipulating in an Excel spreadsheet.

The ethics associated with working with and deducing from a smaller subset of data are simple: make sure what you are doing to wrangle your data is not producing results that are outside of results if you were to use the entire dataset. For example, I had to select the data in which none of the ratings were blank, because if some were blank python did not know how to handle it. Even worse, if it had calculated it as a zero, they would have majorly affected the average ratings. So, while I did pull out multiple films for having no rating, it was a critical step to ensuring my data was as accurate as possible.

The other thing you must keep in mind when data wrangling is that you can not data wrangle based on bias. For example, I can not pull out all movies rated R just because I do not prefer watching those. You can only wrangle your data to be more accurate for answering your question, such as removing blank data that will alter your calculations.

There are always concerns about what kind of result you might get if you used the entire dataset instead of a subset. In order to minimize these concerns, you should always report what kind of changes you made to the dataset before explaining your results. Keeping copies of your changes using the command line can be very helpful when it comes to keeping people well informed about what changes you make, which helps them feel more confident in your results.

Before & After:

title	year	certificate	duration	genre	rating	description	stars	votes
Cobra Kai (2018â€“)		TV-14	30 min	Action, Comedy, Drama	8.5	Decades after the	['Ralph Ma	177,031
The Crown (2016â€“)		TV-MA	58 min	Biography, Drama, History	8.7	Follows the politi	['Claire Fo	199,885
Better Call (2015â€“2022)		TV-MA	46 min	Crime, Drama	8.9	The trials and trit	['Bob Ode	501,384
Devil in Oh	-2022	TV-MA	356 min	Drama, Horror, Mystery	5.9	When a psychiatr	['Emily De	9,773
Cyberpunk (2022â€“)		TV-MA	24 min	Animation, Action, Adventure	8.6	A Street Kid tryin	['Zach Agu	15,413
The Sandrr (2022â€“)		TV-MA	45 min	Drama, Fantasy, Horror	7.8	Upon escaping af	['Tom Stur	116,358
Rick and M (2013â€“)		TV-MA	23 min	Animation, Adventure, Comedy	9.2	An animated seri	['Justin Ro	502,160
Breaking B (2008â€“2013)		TV-MA	49 min	Crime, Drama, Thriller	9.5	A high school che	['Bryan Cr	1,831,340
The Imper (2022â€“)		TV-MA	45 min	Action, Adventure, Drama	6.3	After an experim	['Morgan T	3,123
Blonde	-2022	NC-17	166 min	Biography, Drama, Mystery	6.2	A fictionalized ch	['Andrew C	935
Stranger Ti (2016â€“)		TV-14	51 min	Drama, Fantasy, Horror	8.7	When a young bc	['Millie Bol	1,149,889
End of the (II) (2022)		R	89 min	Action, Crime, Drama	4.7	In this high-octan	['Millicent	5,122
The Walkir (2010â€“2022)		TV-MA	44 min	Drama, Horror, Thriller	8.1	Sheriff Deputy Ri	['Andrew L	970,060
Glass Onio	-2022	PG-13	139 min	Crime, Drama, Mystery	8.1	Famed Southern	['Rian Johr	547
Do Reveng	-2022	TV-MA	118 min	Comedy	6.4	Drea and Eleanor	['Jennifer	13,457
I Came By	-2022	TV-MA	110 min	Thriller	6.1	Follows a young	['Babak An	16,230
No Limit	-2022	TV-MA	118 min	Drama, Romance, Sport	5.8	Roxana Aubrey d	['David M.	1,695
1899 (2022â€“)			60 min	Drama, History, Horror	9.6	Multinational imr	['Ben Ashe	853
Peaky Blin (2013â€“2022)		TV-MA	60 min	Crime, Drama	8.8	A gangster family	['Cillian M	531,058
Fate: The V (2021â€“)		TV-MA		Action, Adventure, Drama	6.9	A live-action ada	['Sadie Sov	43,179
The Lord c	-2001	PG-13	178 min	Action, Adventure, Drama	8.8	A meek Hobbit fr	['Peter Jac	1,844,055
Grey's Ana (2005â€“)		TV-14	41 min	Drama, Romance	7.6	A drama center	['Ellen Pon	303,617
Supernatu (2005â€“2020)		TV-14	44 min	Drama, Fantasy, Horror	8.4	Two brothers foll	['Jared Pac	439,597
American I (2011â€“)		TV-MA	60 min	Drama, Horror, Sci-Fi	8	An anthology seri	['Lady Gag	312,970
The Accide (2022â€“)		TV-MA	50 min	Action, Biography, Crime	7.4	A civilian busines	['Ha Jung-v	2,056
NCIS (2003â€“)		TV-14	60 min	Action, Crime, Drama	7.8	The cases of the	['David Mc	145,895
Friends (1994â€“2004)		TV-14	22 min	Comedy, Romance	8.9	Follows the persc	['Jennifer	979,424
Shameless (2011â€“2021)		TV-MA	46 min	Comedy, Drama	8.6	A scrappy, feisty,	['Emmy Rc	239,541
Modern Fa (2009â€“2020)		TV-PG	22 min	Comedy, Drama, Romance	8.5	Three different b	['Ed O'Nei	423,963
Ozark (2017â€“2022)		TV-MA	60 min	Crime, Drama, Thriller	8.5	A financial advic	['Jason Ba	309,552
Echoes	-2022	TV-MA	326 min	Drama, Mystery, Thriller	5.9	Leni and Gina are	['Michelle	9,380
The Gray M	-2022	PG-13	122 min	Action, Thriller	6.5	When the	['Anthony	172,136

title	year	certificate	duration	unit	genre	g2	g3	rating	descript	stars	votes
Cobra Kai	2018	TV-14	30 min		Action	Comedy	Drama	8.5	Decades a	['Ralph Ma	177,031
The Crown	2016	TV-MA	58 min		Biography	Drama	History	8.7	Follows th	['Claire Fo	199,885
Better Call Saul	2015	TV-MA	46 min		Crime	Drama		8.9	The trials	['Bob Ode	501,384
Devil in Ohio	2022	TV-MA	356 min		Drama	Horror	Mystery	5.9	When a ps	['Emily De	9,773
Cyberpunk: Edgerunn	2022	TV-MA	24 min		Animation	Action	Adventure	8.6	A Street Ki	['Zach Agu	15,413
The Sandman	2022	TV-MA	45 min		Drama	Fantasy	Horror	7.8	Upon esca	['Tom Stur	116,358
Rick and Morty	2013	TV-MA	23 min		Animation	Adventure	Comedy	9.2	An animati	['Justin Ro	502,160
Breaking Bad	2008	TV-MA	49 min		Crime	Drama	Thriller	9.5	A high sch	['Bryan Cr	1,831,340
The Imperfects	2022	TV-MA	45 min		Action	Adventure	Drama	6.3	After an e	['Morgan T	3,123
Blonde	2022	NC-17	166 min		Biography	Drama	Mystery	6.2	A fictiona	['Andrew C	935
Stranger Things	2016	TV-14	51 min		Drama	Fantasy	Horror	8.7	When a yc	['Millie Bol	1,149,889
End of the Road	2022	R	89 min		Action	Crime	Drama	4.7	In this high	['Millicent	5,122
The Walking Dead	2010	TV-MA	44 min		Drama	Horror	Thriller	8.1	Sheriff Def	['Andrew L	970,060
Glass Onion: A Knives	2022	PG-13	139 min		Crime	Drama	Mystery	8.1	Famed Sou	['Rian Johr	547
Do Revenge	2022	TV-MA	118 min		Comedy			6.4	Drea and E	['Jennifer	13,457
I Came By	2022	TV-MA	110 min		Thriller			6.1	Follows a	['Babak An	16,230
No Limit	2022	TV-MA	118 min		Drama	Romance	Sport	5.8	Roxana Au	['David M.	1,695
1899	2022		60 min		Drama	History	Horror	9.6	Multinatio	['Ben Ashe	853
Peaky Blinders	2013	TV-MA	60 min		Crime	Drama		8.8	A gangster	['Cillian M	531,058
The Lord of the Rings	2001	PG-13	178 min		Action	Adventure	Drama	8.8	A meek Hc	['Peter Jac	1,844,055
Grey's Anatomy	2005	TV-14	41 min		Drama	Romance		7.6	A drama c	['Ellen Pon	303,617
Supernatural	2005	TV-14	44 min		Drama	Fantasy	Horror	8.4	Two broth	['Jared Pac	439,597
American Horror Stor	2011	TV-MA	60 min		Drama	Horror	Sci-Fi	8	An antholc	['Lady Gag	312,970
The Accidental Narcc	2022	TV-MA	50 min		Action	Biography	Crime	7.4	A civilian b	['Ha Jung-v	2,056
NCIS	2003	TV-14	60 min		Action	Crime	Drama	7.8	The cases	['David Mc	145,895
Friends	1994	TV-14	22 min		Comedy	Romance		8.9	Follows th	['Jennifer	979,424
Shameless	2011	TV-MA	46 min		Comedy	Drama		8.6	A scrappy,	['Emmy Rc	239,541
Modern Family	2009	TV-PG	22 min		Comedy	Drama	Romance	8.5	Three diffe	['Ed O'Nei	423,963
Ozark	2017	TV-MA	60 min		Crime	Drama	Thriller	8.5	A financial	['Jason Ba	309,552
Echoes	2022	TV-MA	326 min		Drama	Mystery	Thriller	5.9	Leni and G	['Michelle	9,380
The Gray Man	2022	PG-13	122 min		Action	Thriller		6.5	When the	['Anthony	172,136

Exploring the Data

See data exploration in Jupyter Notebook.

Results/Data Visualization

See Notebook for full results and visualization

According to the visualizations, the most highly rated Netflix films are long in duration, animated, and rated TV-PG. This could mean Netflix aims to target families with the popular shows that they provide, which makes sense given the millions of families with Netflix around the world. I am sure that many families are happy to put on the animated films for their children, and are willing to rate them highly when they are also able to enjoy it.

Conclusion

The results tell us a few things about Netflix. Firstly, the films that have the highest average rating tend to be long in duration, which could mean that the longer your film is the more likely you are to receive a higher rating. Secondly, the most highly-rated films also are rated TV-PG, which is typically for children of all ages, along with "Parental Guidance." Lastly, Animation is the genre with the highest average rating, which makes sense considering TV-PG is the certificate rated the highest. These three conclusions all lead me to believe that Netflix continues to target families as their main consumers, or maybe there is an underlying reason that Animated/TV-PG films tend to be rated higher than other films. They could remind adults of their childhood, or enjoy having their child distracted by something other than themselves. There are many reasons that the results could be displayed as they are, and further data manipulation on a larger scale would be needed to narrow them down.

Based on this data, if a producer wants to make it big I would recommend creating a long, Animated, TV-PG film in order to have the highest chances at being successful.