Netflix Movie Analysis & Rating Prediction

Managing Big Data Final Project - Team7

Wei (Will) Jiang, Xiang (Maggie) Meng, Paakhi Srivastava, Yulin Gai, Joe Ebby Karuthedath

Background

Netflix: The Beginning

- Launched its SVOD (Streaming Video on Demand) service in 2007
- Media Clearinghouses: gathering of B-rated movies
- Networks were initially reluctant to give up rights to popular content

Netflix: Now

- Departure from cable, many people started to substitute cable with Netflix
- Netflix launched its android and ios app in 2010
- Rise was aided by technological advancements: Internet speeds, smart phones
- Market cap : $291 billion

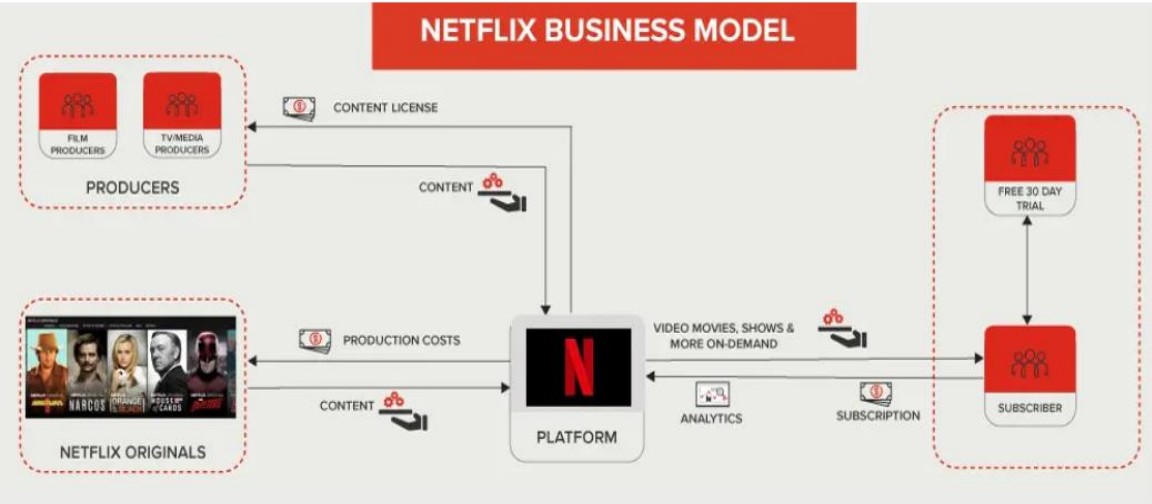
Netflix: The End?

- Challenge posed by new competitors: Amazon Prime video, Disney+
- Eroding market share

NETFLIX

Business Understanding

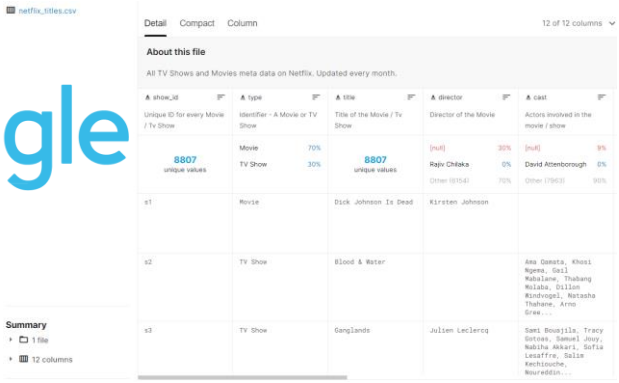
- Major Costs: Acquiring content licenses and producing original content
- Optimizing this cost is essential for profits and maintaining market share
- The model attempts to predict, if a movie is good or bad before the movie is acquired by the platform.



NETFLIX

Data Collection & Storage

kaggle



WIKIPEDIA
The Free Encyclopedia

Feature films [edit]

Title	Genre	Premiere	Runtime	Language
Ghost Stories	Horror anthology	January 1, 2020	2 hours, 24 min.	Hindi
A Fall from Grace	Thriller	January 17, 2020	2 hours	English
Airplane Mode	Comedy	January 23, 2020	1 hour, 36 min.	Portuguese
Horse Girl	Drama	February 7, 2020	1 hour, 44 min.	English
To All the Boys: P.S. I Still Love You	Romantic comedy	February 12, 2020	1 hour, 42 min.	English
Isi & Ossi	Romantic comedy	February 14, 2020	1 hour, 53 min.	German
The Last Thing He Wanted	Political thriller	February 21, 2020	1 hour, 55 min.	English
Yeh Ballet	Drama	February 21, 2020	1 hour, 57 min.	Hindi
All The Bright Places	Romance	February 28, 2020	1 hour, 48 min.	English
Guilty	Thriller	March 6, 2020	1 hour, 59 min.	Hindi
Spenser Confidential	Action comedy	March 6, 2020	1 hour, 51 min.	English
Lost Girls	Crime drama	March 13, 2020	1 hour, 35 min.	English

IMDb

	tconst	averageRating	numVotes	title	id	url	review	rating
0	tt1365050	7.7	78276.0	Beasts of No Nation	tt1365050	tt1365050	[I have seen several documentaries about the i...	6
1	tt1365050	7.7	78276.0	Beasts of No Nation	tt1365050	tt1365050	[Agu is a young boy in an unnamed African nati...	8
		7.7	78276.0	Beasts of No Nation	tt1365050	tt1365050	[Director Cary Fukunaga brings us the turmoil ...	5
		7.7	78276.0	Beasts of No Nation	tt1365050	tt1365050	[Agu (Abraham Attah) is a young lad growing up...	10
		7.7	78276.0	Beasts of No Nation	tt1365050	tt1365050	[You know when you hear other people whining a...	9
		...	...	...	...	...	...	...
		6.4	5442.0	Britney vs Spears	tt15469820	tt15469820	[In some ways, the fascination with Britney Sp...	6
3876	tt15469820	6.4	5442.0	Britney vs Spears	tt15469820	tt15469820	[I gave up half way through..... says it all r...	3
3877	tt15469820	6.4	5442.0	Britney vs Spears	tt15469820	tt15469820	[That documentary was amazing! It revealed so ...	9
3878	tt15469820	6.4	5442.0	Britney vs Spears	tt15469820	tt15469820	[In this documentary, despite the fact that th...	8
3879	tt15469820	6.4	5442.0	Britney vs Spears	tt15469820	tt15469820	[Conservatorships and guardianships are among ...	8

Storage for future analysis

amazon DynamoDB

NETFLIX

Data Preparation & Feature Engineering

season_added
spring, summer, fall, winter

movie_class
Good, Not Good

Threshold: 6.0

	title	country	date_added	release_year	cate_rating	duration	genre	avg_rating	num_votes	is_original
0	¡Ay, mi madre!	Spain	2019-07-19	2019	TV-MA	81	Comedies	3.9	372	0
1	1 Chance 2 Dance	United States	2017-07-01	2014	TV-PG	89	Dramas	4.8	560	0
2	1 Mile to You	United States	2017-07-07	2017	TV-14	105	Dramas	6.3	2473	0
3	10 Days in Sun City	South Africa	2019-10-18	2017	TV-14	87	Comedies	5.1	65	0
4	10 jours en or	France	2017-07-01	2012	TV-14	97	Comedies	5.9	984	0
...	...	...	...	...	...	...	...	...	...	...
3939	Zodiac	United States	2019-11-20	2007	R	158	Cult Movies	8.3	7	0
3940	Zokkomon	India	2018-11-01	2011	PG	104	Children & Family Movies	4.1	396	0
3941	Zombieland	United States	2019-11-01	2009	R	88	Comedies	5.9	8	0
3942	Zoom	United States	2020-01-11	2006	PG	88	Children & Family Movies	6.4	7	0
3943	Zubaan	India	2019-03-02	2015	TV-14	111	Dramas	8.0	10	0

is_US
0/1

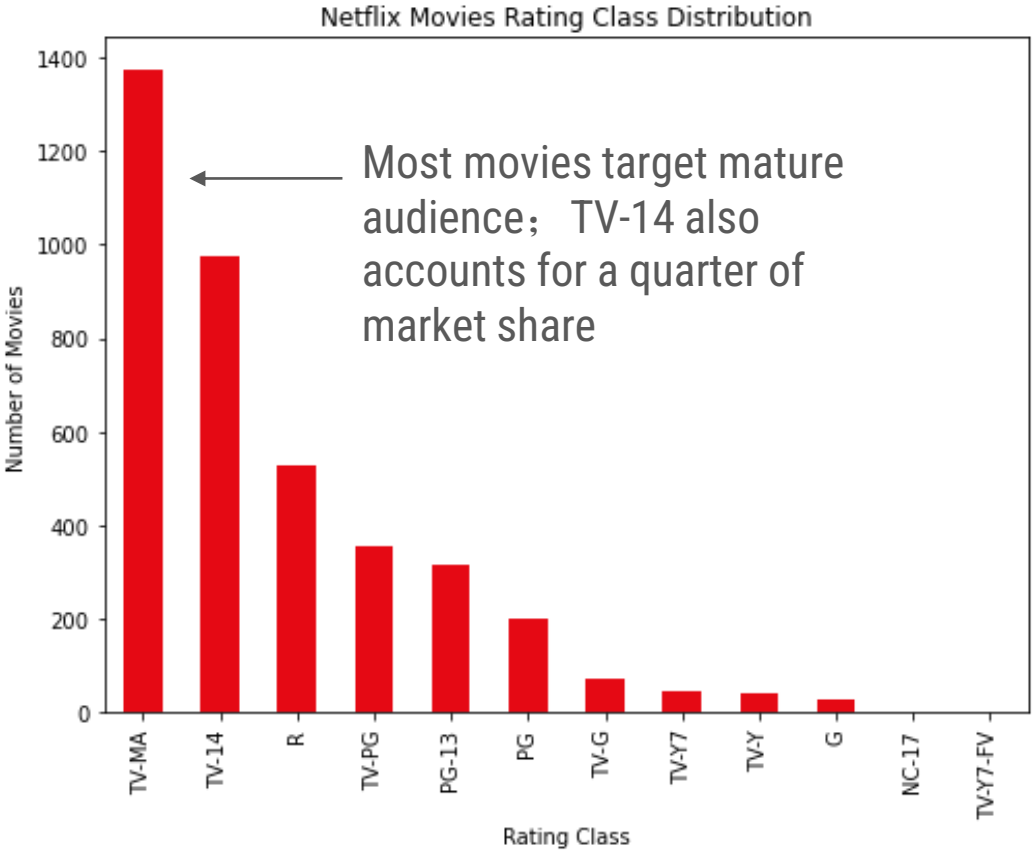
audience_class
1/2/3

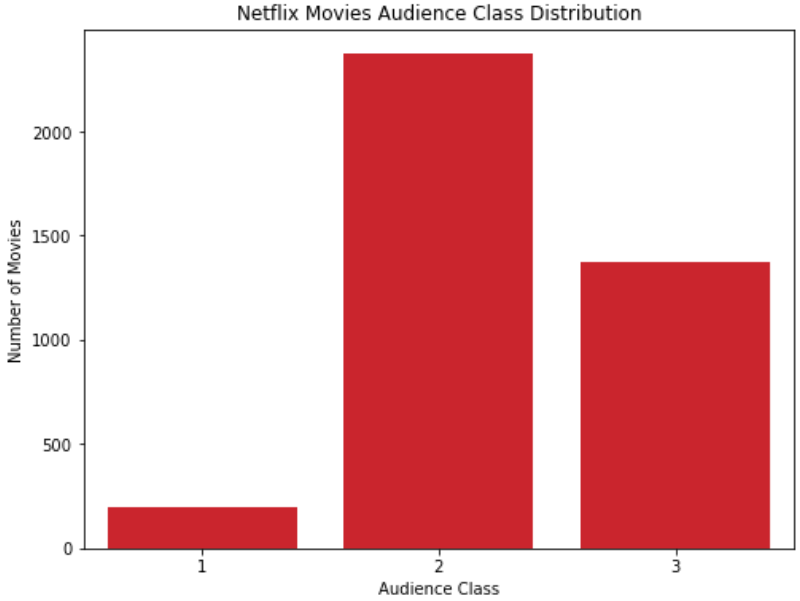
1 = general audience
2 = parent guidance needed
3 = adults only

is_popular
0/1

Threshold: 50% Percentile

Exploratory Data Analysis

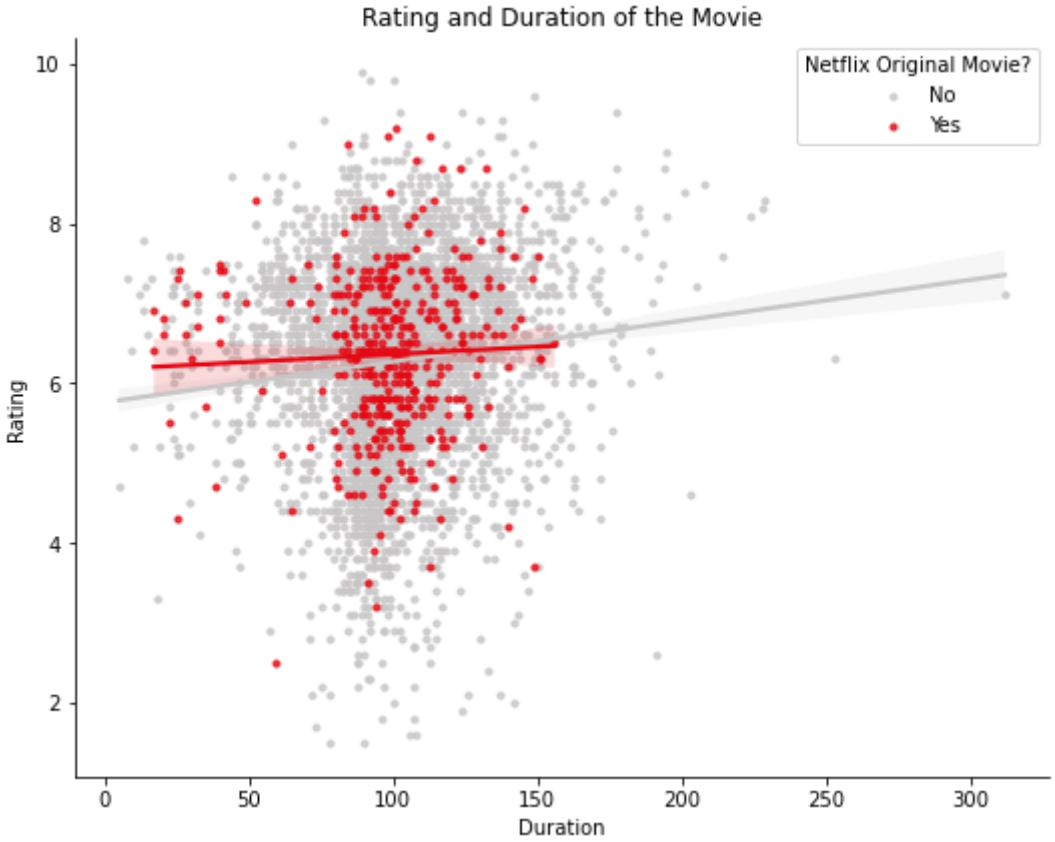




Most movies are in the audience class 2 (need parental guidance), pointing out market opportunities in class 2 movies.

NETFLIX

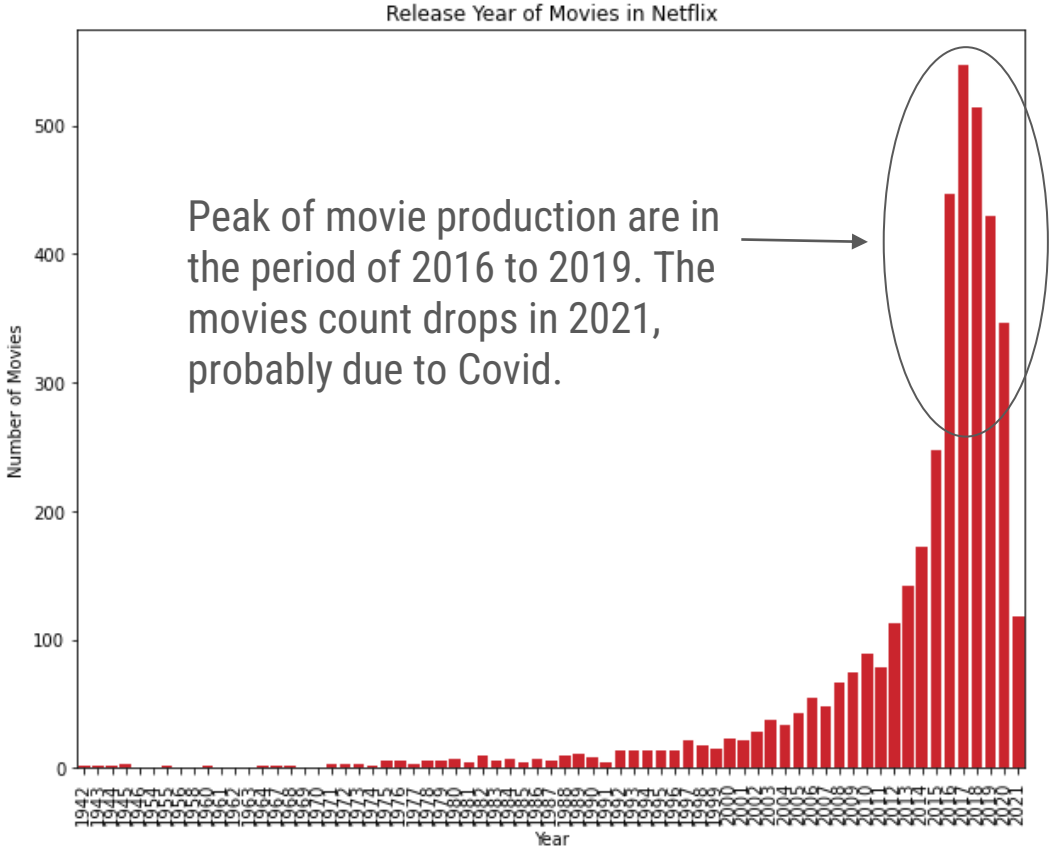
Exploratory Data Analysis



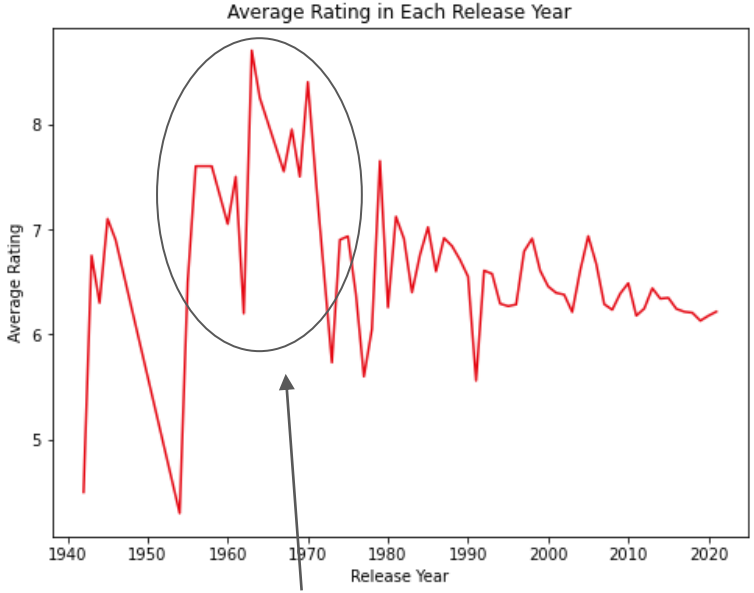
Comparing the regression lines that plot by scatter plots, we can see there is a stronger positive relationship between duration and movie ratings for non-Netflix-original movies, which indicates that we can include movie duration in our model.

NETFLIX

Exploratory Data Analysis



Peak of movie production are in the period of 2016 to 2019. The movies count drops in 2021, probably due to Covid.



However, highest rating scores appear in 1960s, showing that large movie production does not lead to higher scores.

NETFLIX

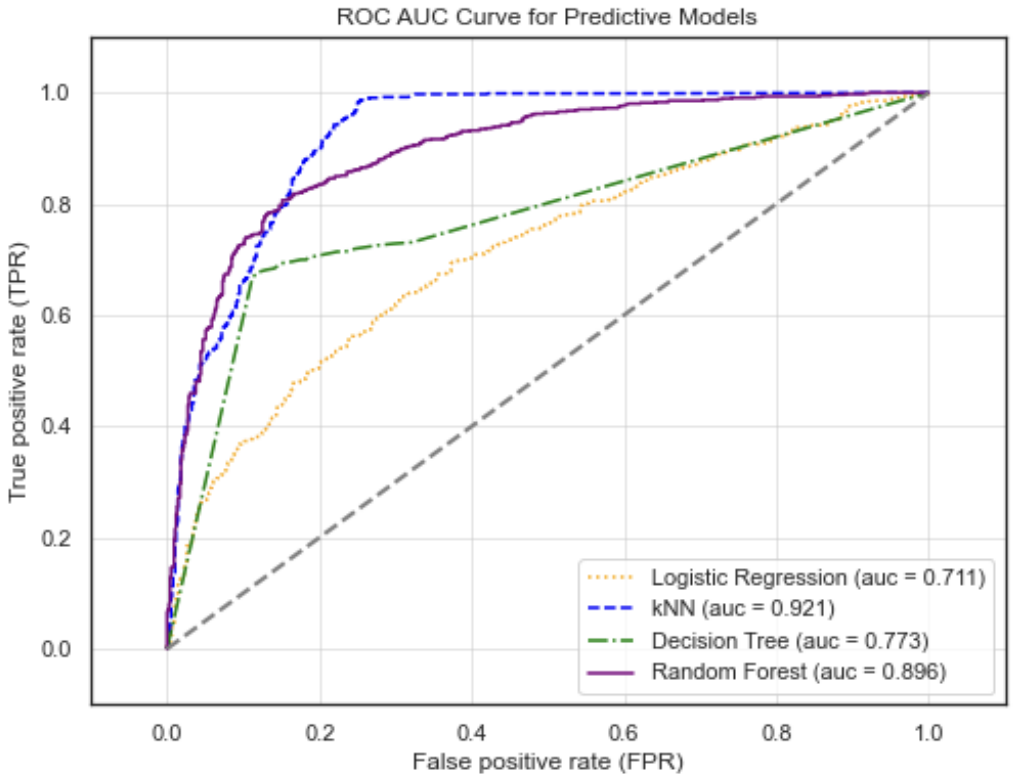
Modeling

- Split the data into test and training sets
- Ran a Logistic Regression, KNN, Decision Tree, and a Random Forest with parameter tuning
- Use the AUC Score to select the best model

	Logistic Regression	**KNN**	Decision Tree	Random Forest
Hyperparameter Grid Search	'lr__C': 10.0, 'lr__penalty': 'l2'	'knn__n_neighbors': 49, 'knn__weights': 'distance'	'criterion': 'gini', 'max_depth': 18, 'min_samples_leaf': 1, 'min_samples_split': 2	'max_depth': 16, 'n_estimators': 100
Non-nested CV accuracy	0.6567	0.7558	0.7504	0.7876
AUC Score	0.71	**0.90**	0.76	0.87

NETFLIX

Model Evaluation



To evaluate our model by using test data, we created an ROC curve on the test data where we can see the KNN model surpasses the others with a AUC of 0.92.

NETFLIX

Business Value

- Netflix Currently spends between $100 and $250 million on blockbuster movies
- The model will aid Netflix executives while making the decision of weather to go ahead producing an in house content.
- And while making a decision to acquire content licenses from a third party producer.

Future Improvements

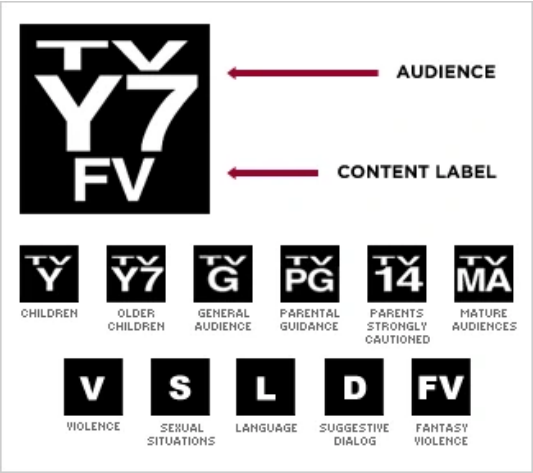
- For Opportunities, it can be said a very friendly period for the development of the streaming platform during the pandemic
- For Threats, however, Netflix also has many competitors competing with them for loyal users， and there are also some further improvements in the script that can be brought about, like using NLP for assessing the scripts of movies to categorize them into different clusters, allowing executives to get a clearer picture

NETFLIX

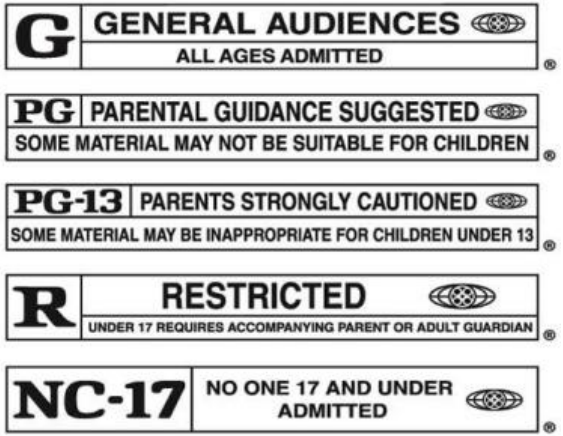
Thank You
&
Questions!

U.S. Movie & TV Rating System

TV Rating System



Movie Rating System



NETFLIX