

Web Scraping Project

Ryan Park

NYCDSA 2020











Overview

Motivation

- There are 200+ Streaming Services including Netflix, HBO Max, Hulu etc.
- It is hard to access the content catalogue from the websites without subscription

Objective

- Compare the content (movies and TV series) from 5 major streaming services
 - (Amazon Prime, Disney plus, HBO Max, Hulu, and Netflix)
- Help the audience understand which streaming service is the best fit for them

Target Audience

People low on budget looking for one streaming service to subscribe





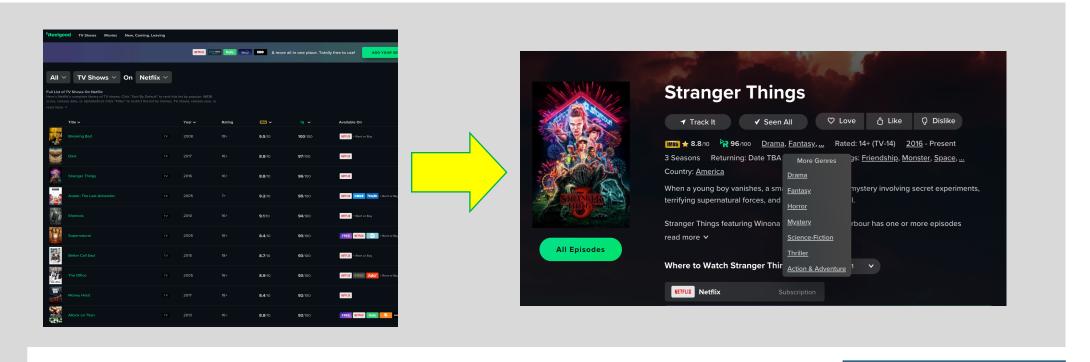








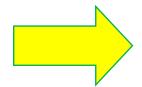
Data Collection



Information Collected

- Title
- Service Name
- Content Type
- Year
- # of Seasons

- IMDB Score
- Reelgood Score
- Film Rate
- Genre



27,305 Movies and TV Series in total











Questions of Interest

- Which service has the most content?
- How does the distribution of the imdb scores look for each service?
- How many high quality movies and TV series are there for each service?
- Is there a trend in the count of high-quality content?
- How does the genre breakdown look for high-quality content?
- How many content do you get for one dollar spent on subscription?



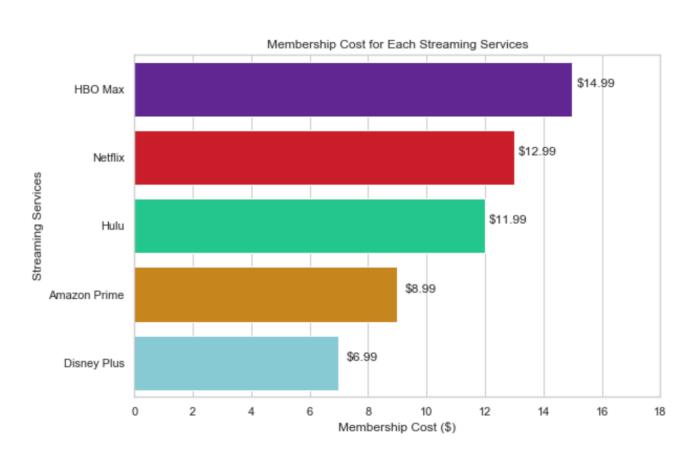








Baseline Information: Membership Cost



 While membership costs vary extensively, more factors need to be taken into account to decide which platform to subscribe.





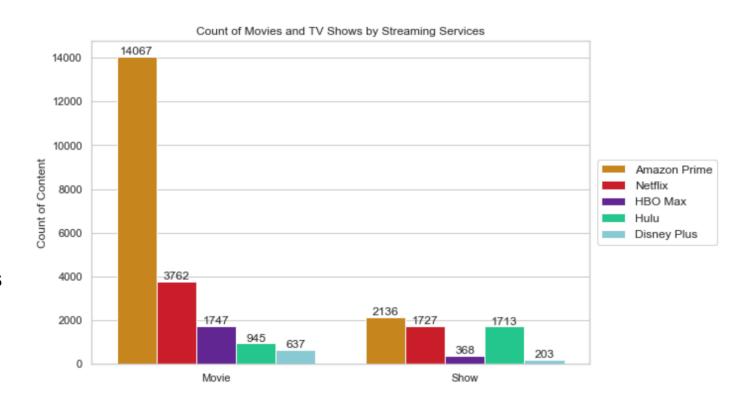






Q. Which service has the most content?

- Amazon Prime has the most movies
- Similar counts of TV shows for Amazon Prime, Netflix, Hulu
- Hulu has more TV shows than movies





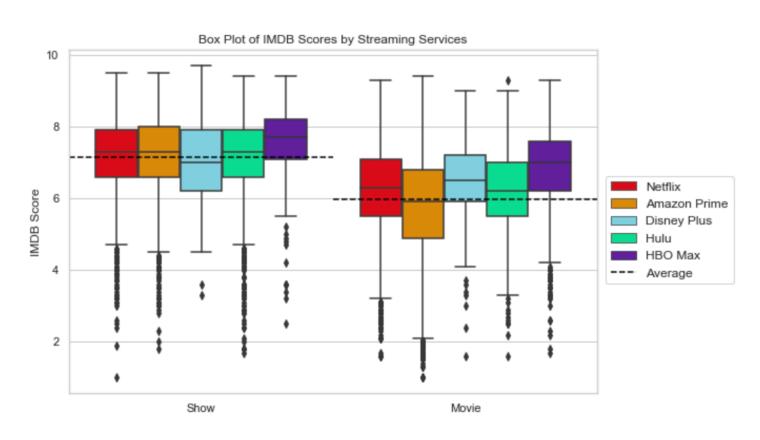








Q. How does the distribution of the IMDb scores look?



- TV shows tend to be rated higher than movies
- There are many badly rated content



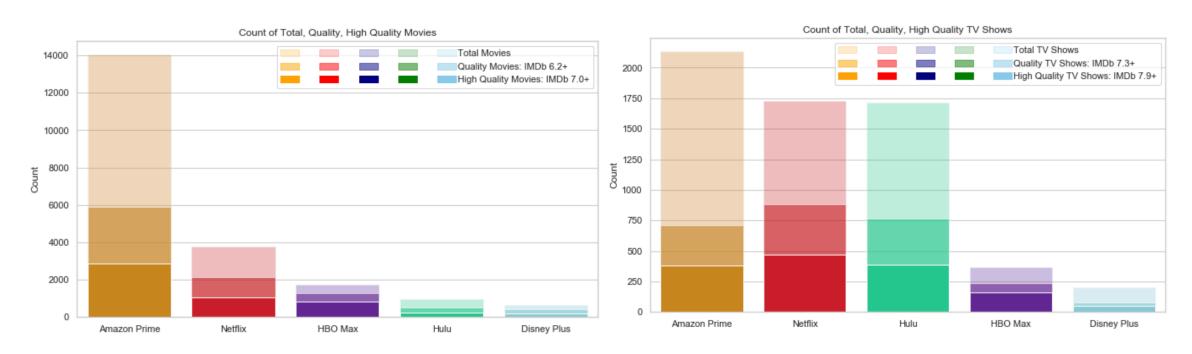








Q. How many high-quality content are there?



- Quality Content: above 50%
 High Quality Content: above 75% (IMDb Scores)
- When it comes to quality and high-quality movies, Amazon Prime is still the leading platform.
- However, Netflix wins the first place for having the most quality and high-quality TV shows.
- Note that, percentage-wise, HBO Max is a safe choice, if you don't like checking ratings



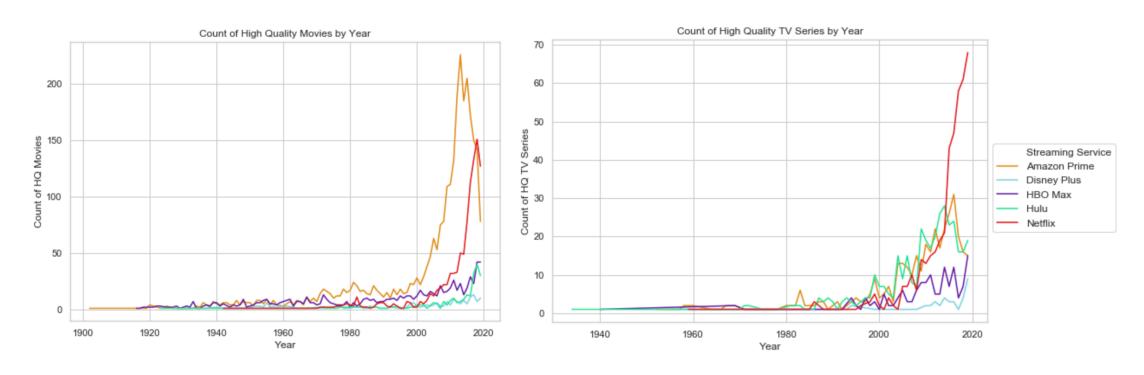








Q. Is there a trend in the count of high-quality content?



- High-quality content are mostly made in the 2000s.
- General public may be more interested in relatively newer content.



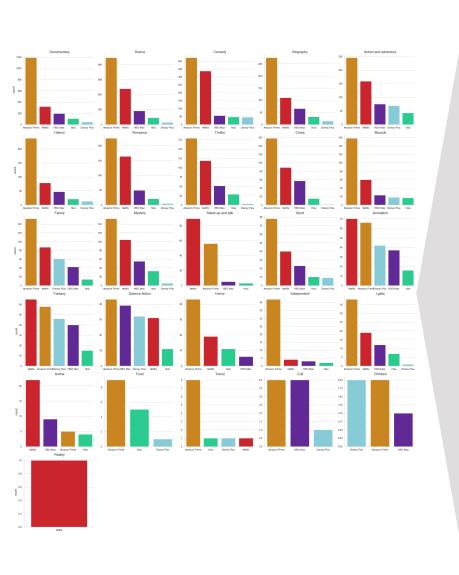


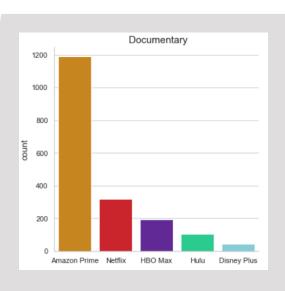


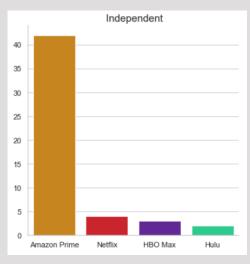


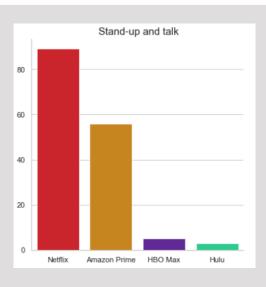


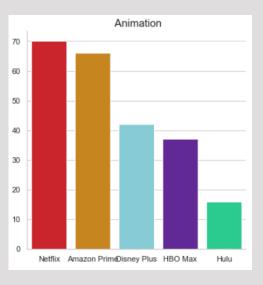
Q. How does the genre breakdown look for HQ movies in 2000s?

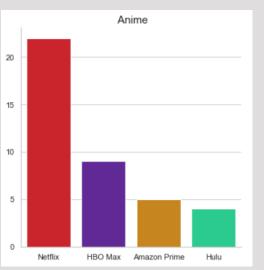


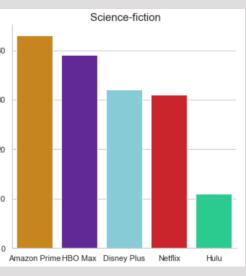














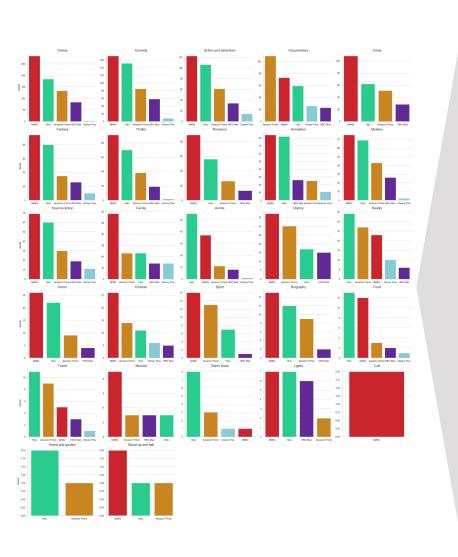


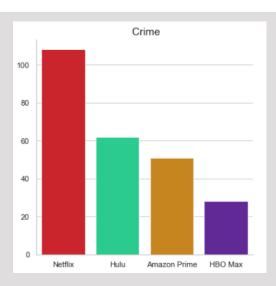


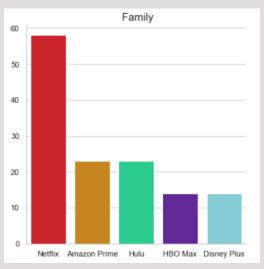


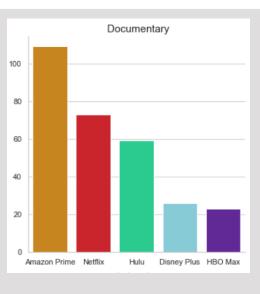


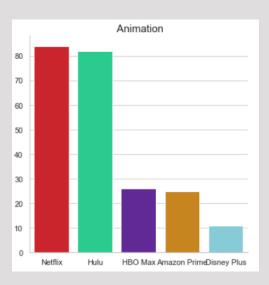
Q. How does the genre breakdown look for HQ TV series in 2000s?

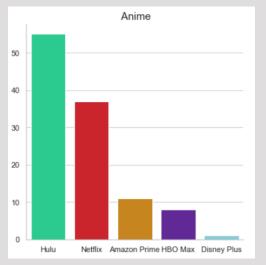


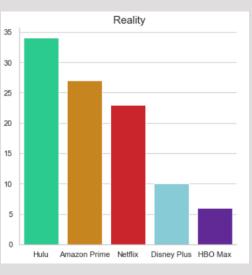
















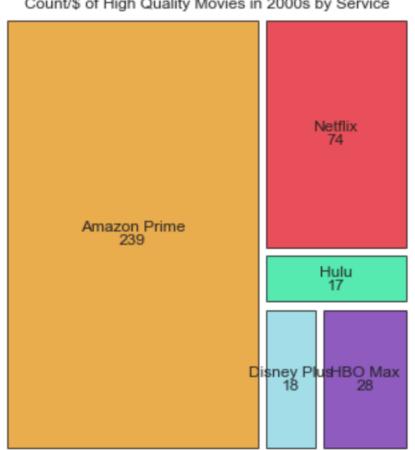




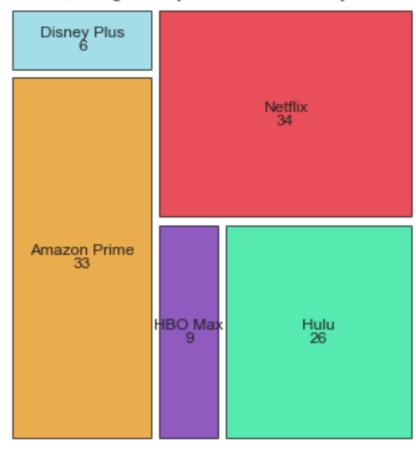


Q. How many content do you get for one dollar spent on subscription?





Count/\$ of High Quality TV Shows in 2000s by Service













Recap: Key Findings and Insights

Movies

- In general, Amazon Prime has the most high-quality movies, followed by Netflix and HBO Max.
- Specifically, Amazon Prime is well-known for documentary, history, and independent movies.
- Yet, Netflix is the leading service with stand-up and talk, fantasy, animation, and anime movies.
- HBO Max is popular for its science-fiction movies.

TV Shows

- Netflix has the most high-quality TV series, followed by Hulu, and Amazon Prime.
- Netflix is especially well-rated with its crime and family TV series.
- Hulu has a focus on its TV series and is popular for its reality and anime TV series.

Cost & Benefit

- If you are into well-rated movies, you should subscribe to Amazon Prime.
- TV series lovers should subscribe to Netflix. But it has a tight match with Hulu and Amazon Prime for TV series. Therefore, further checking the count of content by genres is recommended.











Further Approach

- Analysis of movies and TV series that appear in multiple services
- Comparison of classic and new content
- Analysis of number of seasons for TV series
- App that helps users search and filter content by different criteria