

M114020014 林詠翔

M114020039 張家晧

M114020022 高雲揚

M114020052 王渙鈞



In recent years, the use of OTT (online streaming media) as a way of entertainment for young people has gradually become mainstream, with Netflix and Disney+being the largest players in the market.

Disney+ has many exclusive Marvel licensed content, while Netflix has a lot of original drama.

The topic of this assignment is to analyze and compare the Netflix and Disney+ series datasets on Kaggle via data visualization.



Relevant Data (Netflix titles.csv)

This data was acquired in July 2022 containing data available in the United States.

Missing Value: 10% Data size: 1.93MB Source: Kaggle

- ✓ description: A brief description.
- ✓ release_year: The release year.
- → age_certification: The age certification.
- ✓ runtime: The length of the episode (SHOW) or movie.
- ☑ genres: A list of genres.

- ✓ title: The name of the title.
- ✓ show_type: TV show or movie.
- production_countries: A list of countries that produced the title.
- ✓ seasons: Number of seasons if it's a SHOW.

- ∠ imdb_id: The title ID on IMDB.
- ☑ imdb_score: Score on IMDB.
- → tmdb_popularity:
 - Popularity on TMDB.



Relevant Data (Netflix_credit.csv)

This data was acquired in July 2022 containing data available in the United States.

Missing Value: 3.3%

Data size: 3.63MB

- person_ID: The person ID on JustWatch.
- id: The title ID on JustWatch.
- → name: The actor or director's name.
- □ Language State Stat
- ✓ role: ACTOR or DIRECTOR.











Relevant Data (Disney+ titles.csv)

This data was acquired in May 2022 containing data available in the United States.

Missing Value: 13% Data size: 610KB Source: Kaggle

- ✓ description: A brief description.
- ✓ release_year: The release year.
- → age_certification: The age certification.
- ✓ runtime: The length of the episode (SHOW) or movie.
- ☑ genres: A list of genres.

- ✓ Loreid: The title ID on JustWatch.
- → title: The name of the title.
- ✓ show_type: TV show or movie.
- production_countries: A list of countries that produced the title.
- → seasons: Number of seasons if it's a SHOW.

- ☑ imdb_id: The title ID on IMDB.
- ☑ imdb_score: Score on IMDB.
- → tmdb_popularity:
 - Popularity on TMDB.



Missing Value: 1.5%

Data size: 1.31MB

Relevant Data (Disney+ credit.csv)

This data was acquired in May 2022 containing data available in the United States.

- → person_ID: The person ID on JustWatch.
- ☑ id: The title ID on JustWatch.
- ☑ name: The actor or director's name.
- □ character_name: The character name.
- ☑ role: ACTOR or DIRECTOR.











- 1. Share of Show and Movie in the top 20 of Netflix and Disney+ IMDB ratings?
- 2.Netflix and Disney+ movies, which year has the largest selection?
- 3. Comparison of the Top 20 production countries ranked by IMDB in Netflix and Disney+?
- 4. Which country has the highest frequency Ristricted on Netflix and Disney+?
- 5.Is there a big drop in the rating of the same film in Netflix and Disney+ on IMDB and TMDB?



6.Is the credibility of the imdb score on Netflix and Disney+ the same?

7. Which movie type is the most at the top 20 movie on Disney+ and Netflix according to IMDB score?

8.Is the relationship between tmdb score and imdb score the same in Disney+ and Netflix?

9. The most popular actors on Netflix and Disney+?

10. Which genre of show is the most at top 20 ranked by IMDB score on Netflix and Disney+?



Music not only soothes the soul, but sometimes also serves as a driving force for people to move forward. The source and content of Theme 2 dataset are as follows:

- The dataset is a merger of two datasets:
 Billboard and Spotify's popular songs dataset.
- Billboard: Web Crawler
- Spotify: Kaggle
- Using this dataset to perform data visualization analysis, we can explore more interesting stories hidden behind popular songs.



Relevant Data (Spotify.csv)

Missing Value: 22% Data size: 19.4MB Source: Spotify, billboard

- id: Spotify track ID
- → name: Track title
- popularity: The popularity of the track. The value will be between 0 and 100, with 100 being the most popular.
- ☑ duration ms: Duration of a track, in milliseconds (ms)

- ≥ xplicit: Whether or not the ≥ danceability: How track has explicit lyrics (true = yes it does; false = no it does not OR unknown).
- → artists: List of artist names
- id artists: List of Spotify artist IDs
- ✓ release date: Full release date of a track, usually in YYYY-MM-DD format

- suitable a track is for dancing
- → energy: How intense and active a track is
- the track
- ☑ loudness: Overall loudness of the track, in decibels (dB)



Relevant Data (Spotify.csv)

- track is in major mode (1) or minor (0)
- of spoken words in the track
- → acousticness: Confidence measure of whether a track is acousticScore on TMDB.

- **≥** instrumentalness: Proportion of instrumental parts in a track
- → liveness:Detects live audience in a track. Represents the probability time signature: Overall that a track was performed live
- → valence: Measures how positive a track sounds, from 1 (extremely positive) to 0 (extremely negative)
- track, in beats per minute (BPM)
 - time signature of a track
 - certification.



- 1. The proportion of songs with explicit lyrics (expressed as 1) that enter the HOT100.
- 2. The distribution of notes in different syllables among songs that enter the HOT100.
- 3. The distribution of songs in major key (1) and minor key (0) among songs that enter the HOT100.
- 4. The distribution of the proportion of original instrumental music in songs that enter the HOT100.
- 5. The distribution of popularity scores (computed by the algorithm) among songs that enter the HOT100.



- 6. The distribution of songs presented in a rap style among songs that enter the HOT100.
- 7. The distribution of songs that convey positive energy among songs that enter the HOT100.
- 8. The distribution of different dB levels among songs that enter the HOT100.
- 9. The distribution of the level of rhythmic intensity among songs that enter the HOT100.
- 10. The distribution of the Pitch Class notation among songs that enter the HOT100.



Contribution

25%

25%

25%

25%

M114020014 林詠翔 contribution Data collecting. M114020022 高雲揚 contribution Data collecting. M114020039 張家皓 contribution PPT M114020052 王渙鈞 contribution PPT

