

DATA DICTIONARY - SCRAPED MYANIMELIST DATA

GENERAL INFORMATION

The following data dictionary refers to data scraped from <https://myanimelist.net/anime/season/archive> using a tool called <https://www.webscraper.io/>. The data was scraped on 29th March 2022 and is around 1MB in size. The data is scraped by Chanyut Yuvacharuskul, and is part of the 2143488 Big Data and Artificial Intelligence free elective course for Chulalongkorn University ISE students.

HOW TO READ THE DICTIONARY

The following is a data column shown in the dictionary:

rank

Feature : Nullable

Ranking of an anime, based on the score. The higher the score, the higher the rank. Calculated with respect to all anime in MyAnimeList.

#\d+

#12665

Ranked 12,665th in MyAnimeList.

The unindented line shows the column name.

The first line indicates whether it is a feature, an identifier of the data, or a metadata. After the colon, the information of whether all data are non-null or nullable is shown.

The second line describes the meaning of the column.

The third line is a regex describing the patterns of the data in that column. To explore more about regex, go to <https://regex101.com/>.

The fourth line is an example of data in that column.

The fifth and final line explains the meaning of the example data shown in the fourth line.

DICTIONARY

web-scraper-order

Identifier : Non-Null

An ID for the dataset. Composed of a unix timestamp followed by the order it is scraped. The order is not sorted.

\d+-\d+

1648530274-1492

Scraped on unix time 1648530274 (March 29th 2022, 12:04 PM GMT+07:00). Is the 1,492th anime that is scraped.

web-scraper-start-url

Metadata : Non-Null

The root of the scraped website. Indicates the webpage where all data in the dataset is accessible from.

<https://myanimelist.net/anime/season/archive>

<https://myanimelist.net/anime/season/archive>

The URL of MyAnimeList. The value is the same for all rows.

season

Feature : Non-Null

The season an anime starts airing on.

(Winter|Spring|Summer|Fall) \d{4}

Spring 2004

The anime started airing around April 2004.

season-href

Metadata : Non-Null

The URL of MyAnimeList listing anime on the season an anime starts airing on.

[https://myanimelist.net/anime/season/\d{4}/\(winter|spring|summer|fall\)](https://myanimelist.net/anime/season/\d{4}/(winter|spring|summer|fall))

<https://myanimelist.net/anime/season/2011/summer>

The URL of MyAnimeList listing anime that started airing in the summer of 2011.

anime

Feature : Non-Null

The name of an anime.

.+

Hikaru No Go

The name of the anime is "Hikaru No Go".

score

Feature : Nullable

The weighted average of the score given by the community.
Ranges from 1 to 10, inclusive. Exactly how MyAnimeList
calculates the score is described in
<https://myanimelist.net/info.php?go=topanime>.

(N\A\d.\d{2})

8.08

The score of this anime is 8.08/10.

rank

Feature : Nullable

Ranking of an anime, based on the score. The higher the score,
the higher the rank. Calculated with respect to all anime
in MyAnimeList.

(N\A#\d+)

#12665

Ranked 12,665th in MyAnimeList.

popularity

Feature : Non-Null

Ranking of the popularity of an anime. The higher the number of
viewers, the higher the rank. Calculated with respect to
all anime in MyAnimeList.

#\d+

#16458

16,458th most popular in MyAnimeList.

members

Feature : Non-Null

Numbers of viewers in an anime. Calculated by how many accounts
have listed an anime as viewed.

\d{1,3}(,\d{3})*

123,926

123,926 accounts watch this anime.

season_repeat

Metadata : Non-Null

The season as shown in the webpage of an anime. Should be of
the same value as the "season" column.

(Winter|Spring|Summer|Fall) \d{4}

Fall 2001

The season the anime first aired is Fall 2001.

type

Metadata : Non-Null

The type of anime as shown in the webpage of an anime. Should
be listed as "TV".

TV

TV

The anime is a TV show.

studio

Feature : Nullable

The studio that produced the anime.

(.+|null)

Toei Animation

The anime was produced by Toei Animation.