

The newly merged dataset with 19 variables and 16368 rows

1. uid

This is a unique identifier for each entry or anime in the dataset.

2. title

The title of the anime.

3. synopsis

A brief summary or synopsis of the anime's plot.

4. genre.x

genre of the anime. There is originally a genre.y, but they are same idea. To clean data, I set genre.y to null

5. aired

The original air date(s) of the anime. It may contain a range of dates or a single date.

6. episodes.x

The number of episodes the anime has

7. members.x:

The number of members for a specific anime

8. popularity

A ranking or score indicating how popular the anime is among viewers.

9. ranked:

The anime's rank in comparison to other anime, likely based on some sort of scoring or rating system.

10. score

The overall score or rating given to the anime, probably averaged from user reviews or ratings.

11. img_url

A URL link to an image of the anime, possibly a thumbnail or cover art.

12. link:

A direct link to a webpage containing more information about the anime.

13. name:

There are some animes showed in one dataset but not in second one. So we need to save both name and title

14. type:

■ The format or type of the anime, such as TV series, movie, OVA, etc.

15. episodes.y:

■ Similar to 'episodes.x', indicating the number of episodes from the other dataset involved in the merge.

16. rating

■ Similar to score, but with another calculating system

17. members.y:

■ Similar to 'members.x', indicating the number of members or viewers from the other dataset.

18. year:

■ The year extracted from the 'aired' column, indicating the last year the anime aired.

19. new_category

■ This is a new column we made to compare the rating situation. When the score is larger than 6, we consider it as high. Otherwise, we consider it as low