MyAnimeList Data Project

Max Yang
Wing Yi Ma (Maggie)

Domain

Anime (Japanese animated film/tv/etc),
User,
Users' review data

From the website - MyAnimeList.net

Investigative Questions

1. **What are the *best media/production* (type, source, episodes, rating) *forms* for a popular anime?**

    => Do results change when filtering to only take animes that have been scored by >= 1000 users

2. **For each scoring category, what are the *best* and *worst* genres?**

    => Do female users tend to give higher scores?

3. **What are users' *most favorite* animes by their attributes (gender, age)?**

    => Do these resulting top anime actually correspond to their age/gender demographic genres

Results

➔ Light novel based anime tend to be the most popular (by the metrics in our queries)

◆ Many more Manga than Light Novel based anime
3024 Manga vs **540** Light Novel vs Filtered **2119** Manga vs **495** Light Novel

➔ Josei (targeted towards teenage/older women) anime tend to be the most highly rated by users

◆ Female users tend to give higher scores on average.
102190 Male Users : **32795** Female Users : **1270** Non-Binary Users
47 Josei anime / **4750** anime with reviews

-- BIAS --

Challenges

Finding an appropriate dataset

◆   Designing the schema

Linking the two data sources of our dataset

◆   Cleaning real user data - found a lot of

inappropriate or nonsense data

Lessons

How non-decisive data queries can be for answering questions/research

Have to be SMART and CAREFUL when working with databases

Questions?