

BRecSys

BRecSys is a suitable dataset to study systems that aim at offering recommendations that consider both sensitive and non-sensitive attributes of the users and items. The ultimate goal of BRecSys is to provide a common source of data for benchmarking algorithms that covers equity from the user and item side.

Data Content

Our dataset consists of 3,317,428 ratings from around 31,013 users for 646,155 songs and 36,541 artists.

The BeyondDATA dataset contains four files:

- *user_info.csv*. This file contains specific users' information depicted by the attributes <user_id, country, continent and gender>.
- *tracks_info.csv*. This file contains information on the music items (i.e., tracks or songs). The attributes in this file are <track_id, artist_id, styles (i.e., the style of music), and category_styles (i.e., music style categories)>.
- *artists_info.csv*. This file contains artists' information. The attributes in this file are <artist_id, gender, country, continent, styles (i.e., the style of music), and category_styles (i.e., music style categories)>.
- *ratings.csv*. This file contains the users' music preferences represented by a rating. In particular, the attributes in each tuple are <user_id, track_id, and rating>.

Some Statistics

Gender	Users		Artists	
	Number	Percentage	Number	Percentage
N/A*	0	0,000%	25512	69,817%
Male	19633	63,306%	7638	20,903%
Female	11380	36,694%	3325	9,099%
Non-binary	0	0,000%	29	0,079%
Trans woman	0	0,000%	14	0,038%
Trans man	0	0,000%	2	0,005%
Genderfluid	0	0,000%	7	0,019%
Male organism	0	0,000%	5	0,014%
Female organism	0	0,000%	1	0,003%
Genderqueer	0	0,000%	4	0,011%
Cisgender female	0	0,000%	1	0,003%
Bigender	0	0,000%	1	0,003%
Agender	0	0,000%	1	0,003%
Transgender	0	0,000%	1	0,003%
Totals	31013	100%	36541	100%

Table 1. Gender distribution for Users and Artists

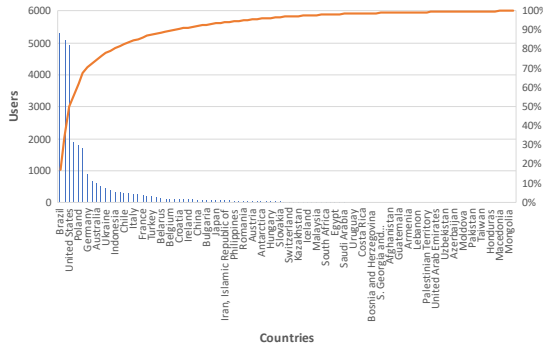


Figure 1. Countries' representativeness by users

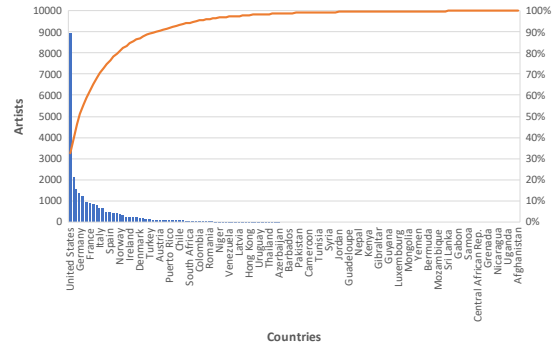


Figure 2. Countries' representativeness by artists

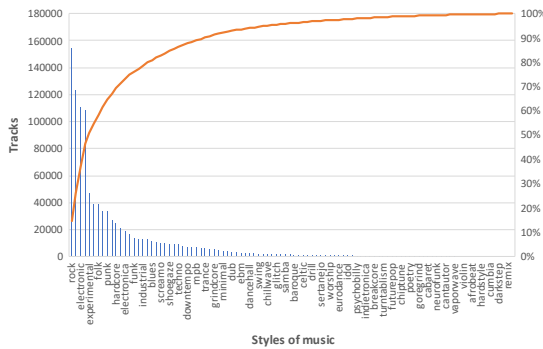


Figure 3. Styles' representativeness by tracks

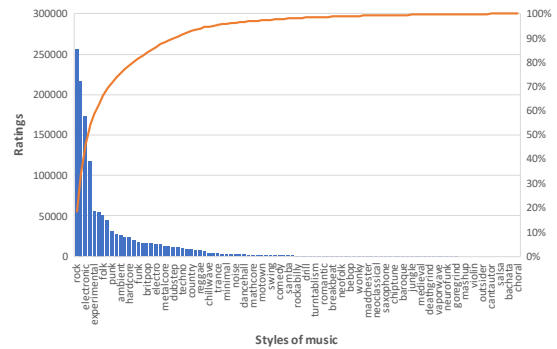


Figure 4. Styles' representativeness by ratings

Download

Click here to download the full dataset ([link](#)).