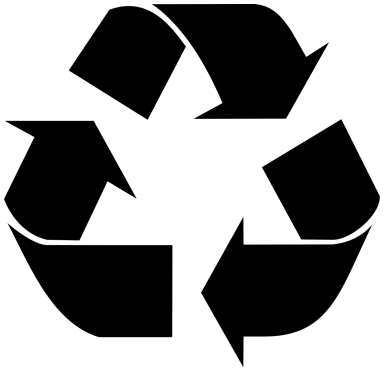
Reggaetón Recommending System

José E. Liquet y Gonzalez
January 03, 2024

Music is important to me, but finding new music is hard

- Reggaetón is my favorite genre
  - I grew up with it
  - I know most of the original songs
- Spotify recommendation algorithm does a bad job a finding new music:
  - It recommends you a handful of songs of popular artists from the same genre
  - It recommends you randomly saved songs from same genre
  - It recommends you music you've interacted positively before



How can I find new music related to my favorite genre?

Impacts:

- “Algorithms on large scale platforms once super-served users, encouraging them ever closer to their respective niches. Now algorithms are increasingly pushing users to the content that supports platform monetisation priorities over user priorities. Users end up feeling that the algorithm is not listening to them anymore. **This trend will accentuate in 2024 among the world’s biggest consumer platforms, resulting in user dissatisfaction and creating a window of opportunity for new, user-need-focused platforms, starting the cycle all over again.**” - Mark Mulligan, “MIDiA’s 2024 predictions: The algorithm is not listening” Dec.18, 2023

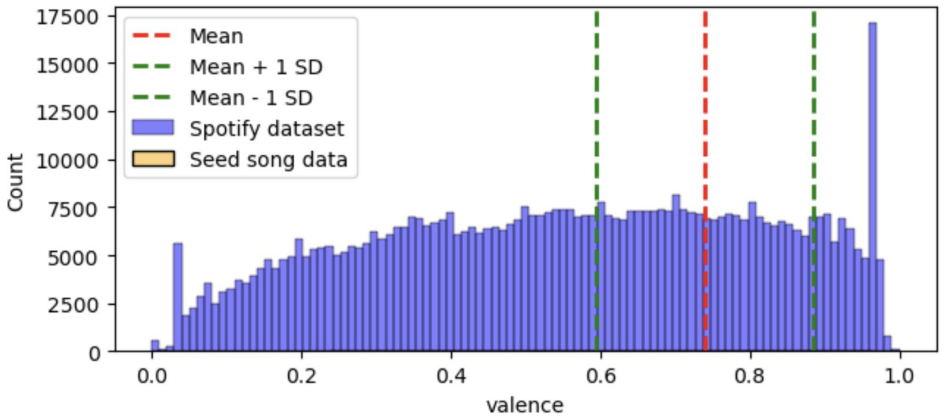
Sprint 1 flow

I like a song -> Find song's URI (Spotify Web Player) -> Query Spotify Web API with Track's URI -> Get metadata -> Perform EDA

index	danceability	energy	key	loudness	mode	speechiness	acousticness	instrumentalness	liveness	valence	tempo	type	id
0	0.82	0.788	6	-5.478	0	0.191	0.272	1.65e-06	0.0404	0.648	93.961	audio_features	0tDSgSmZsbxCkdkfUPjg59

The dataset songs' valence graphed with my seed songs' mean and standard deviation energy represented by red and green color lines, respectively.

A high Valence value represents a "positive" song.



Feature Selection and model training

- In a dataset with over 1000 songs, I made a column for “Reggaeton songs I like” and used this as the target variable to train models and test accuracy.

Random Forest Feature Importance:

	Feature	Importance
0	bpm	0.228172
1	danceability_%	0.175050
3	energy_%	0.131409
2	valence_%	0.122113
4	acousticness_%	0.119642
7	speechiness_%	0.117827
6	liveness_%	0.093578
5	instrumentalness_%	0.012209

Model Accuracy: 0.8901
Model Recall: 0.1429
Model Precision: 0.5000

Gradient Boosting Feature Importance:

	Feature	Importance
0	bpm	0.391494
1	danceability_%	0.235086
3	energy_%	0.102232
2	valence_%	0.082278
4	acousticness_%	0.079896
7	speechiness_%	0.060449
6	liveness_%	0.046181
5	instrumentalness_%	0.002383

Model Accuracy: 0.8953
Model Recall: 0.2381
Model Precision: 0.5556

Neural Network Feature Importance:

	Feature	Importance
1	danceability_%	0.173602
2	valence_%	0.170540
3	energy_%	0.162433
4	acousticness_%	0.162214
5	instrumentalness_%	0.159262
0	bpm	0.155945
7	speechiness_%	0.136558
6	liveness_%	0.131220

Model Accuracy: 0.8901
Model Recall: 0.0000
Model Precision: 0.0000

Next steps

1. Try different features
2. Test accuracy of models
    a. Need to look into how to improve Model accuracy, recall and precision
    b. Overall, the important features are BPM, Danceability, and Energy
3. Use different distance and clustering methods
4. Look up ways to compress and vectorize all song features and cluster