Project 2
Spotify Analysis

Tyler Norlund & Edwin Ramirez

Spotify

Spotify

- More than 40 Million songs

- Pays royalties based on the number of artists's streams

- Operates under a *freemium* business model

Objective

- Classify similar songs to recommend new music

  - Songs can be characterized by many features

  - Spotify offers an API using REST principles for JSON metadata

The Data

- Using the Spotipy Library we collected data that included:

    - 7,564 artists
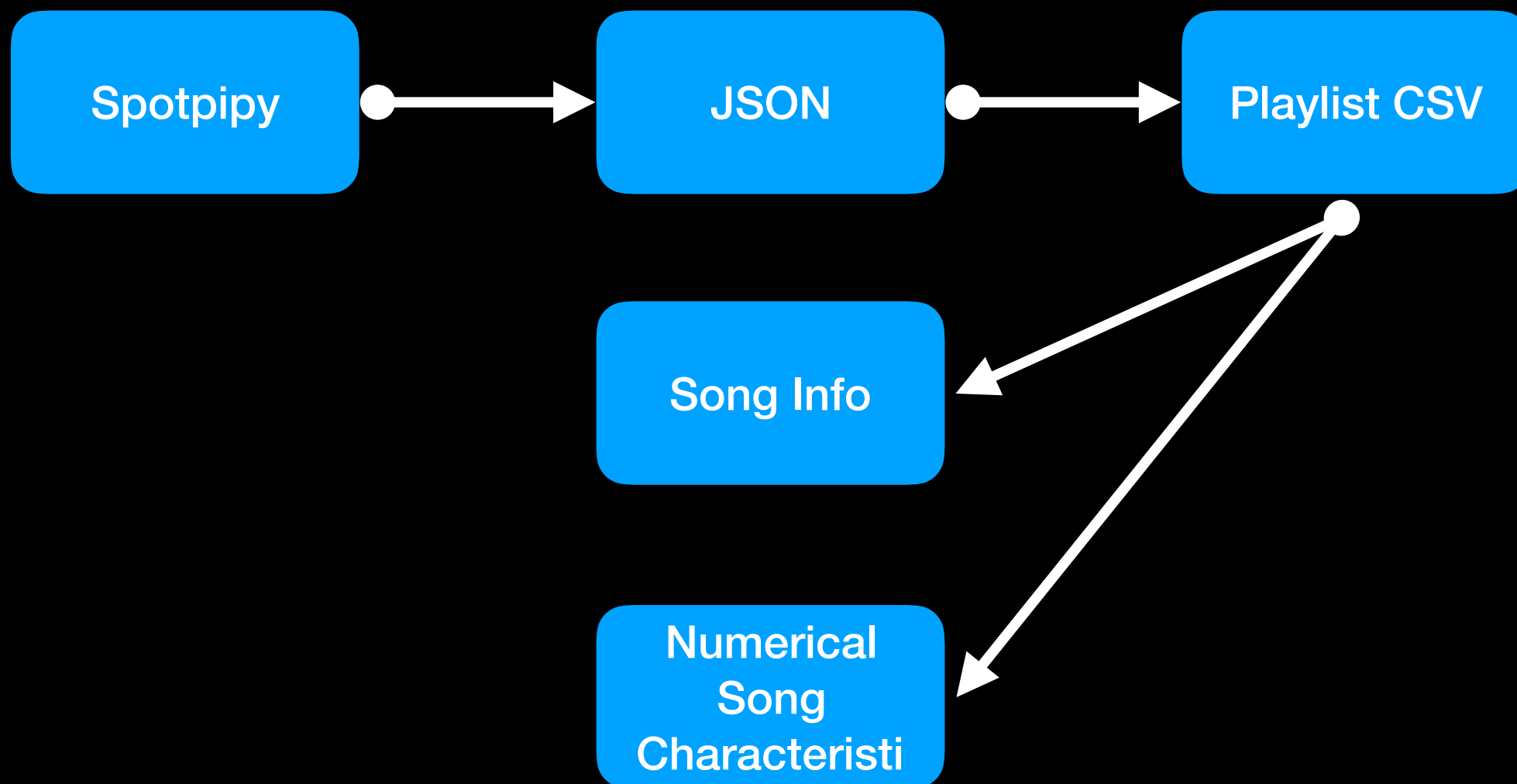
    - 18,835 songs

    - 300 playlists

The Data

- Song attributes of the data:

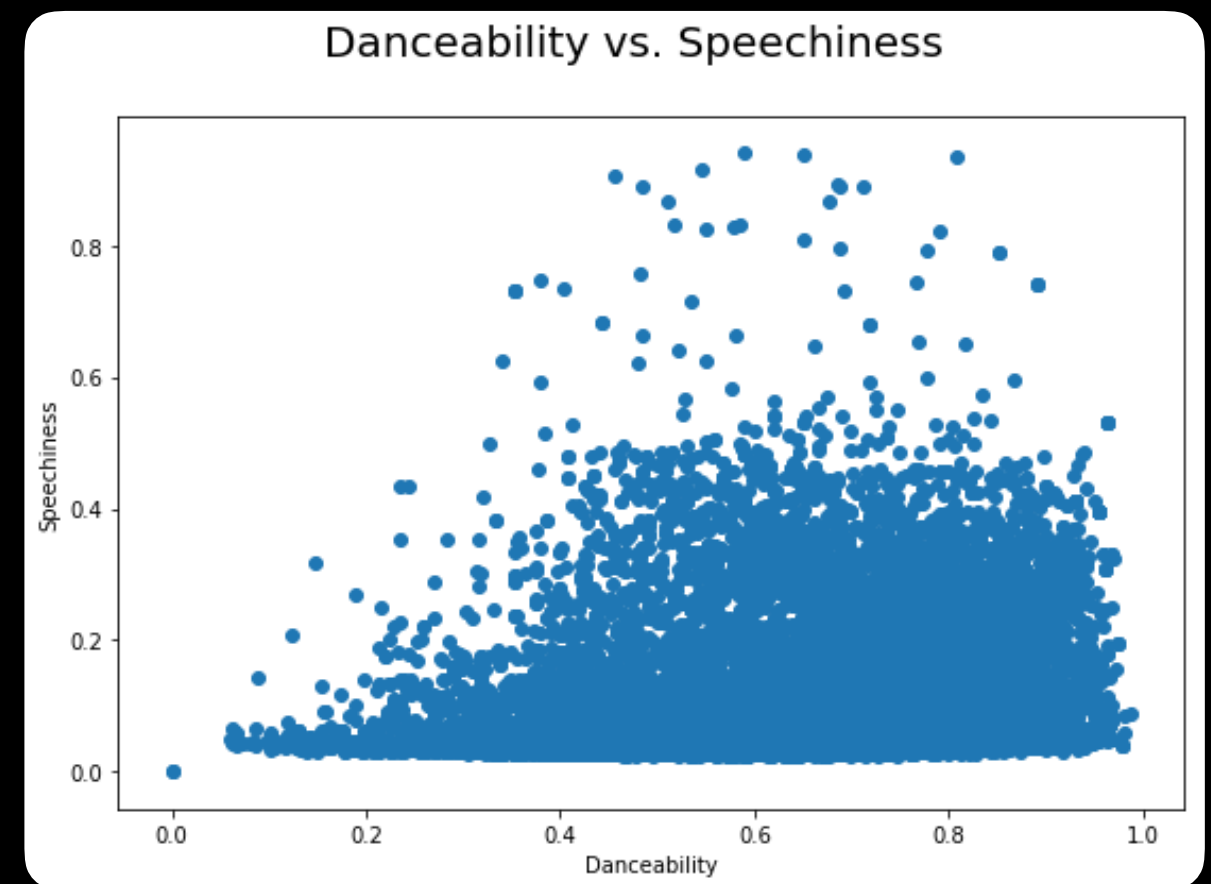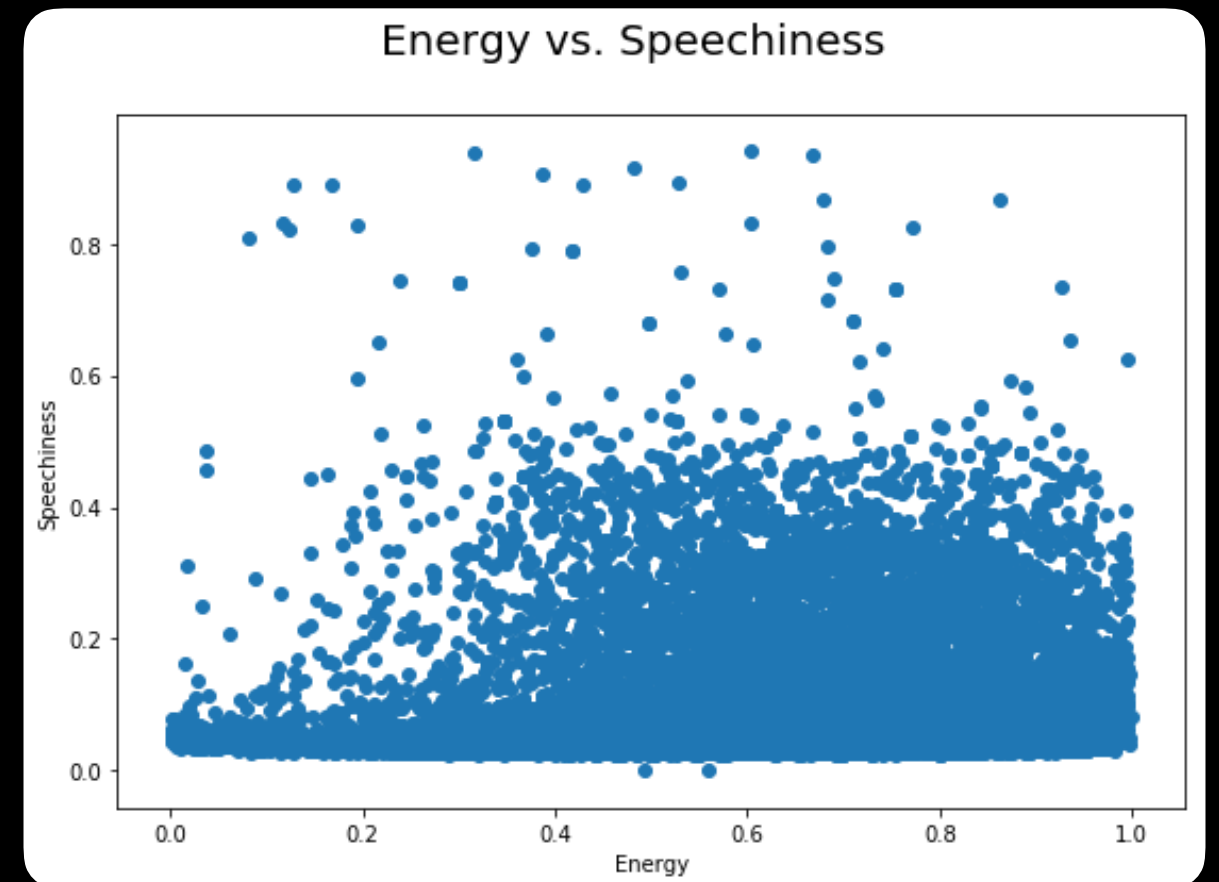|		
Song Popularity	Energy	Audio Mode
Song Duration (ms)	Instrumentalness	Speechiness
	Key	Tempo
Acousticness	Liveness	Time Signature
Danceability	Loudness	Audio Valence

Data Adaptation

Visualization

Danceability & Speechiness

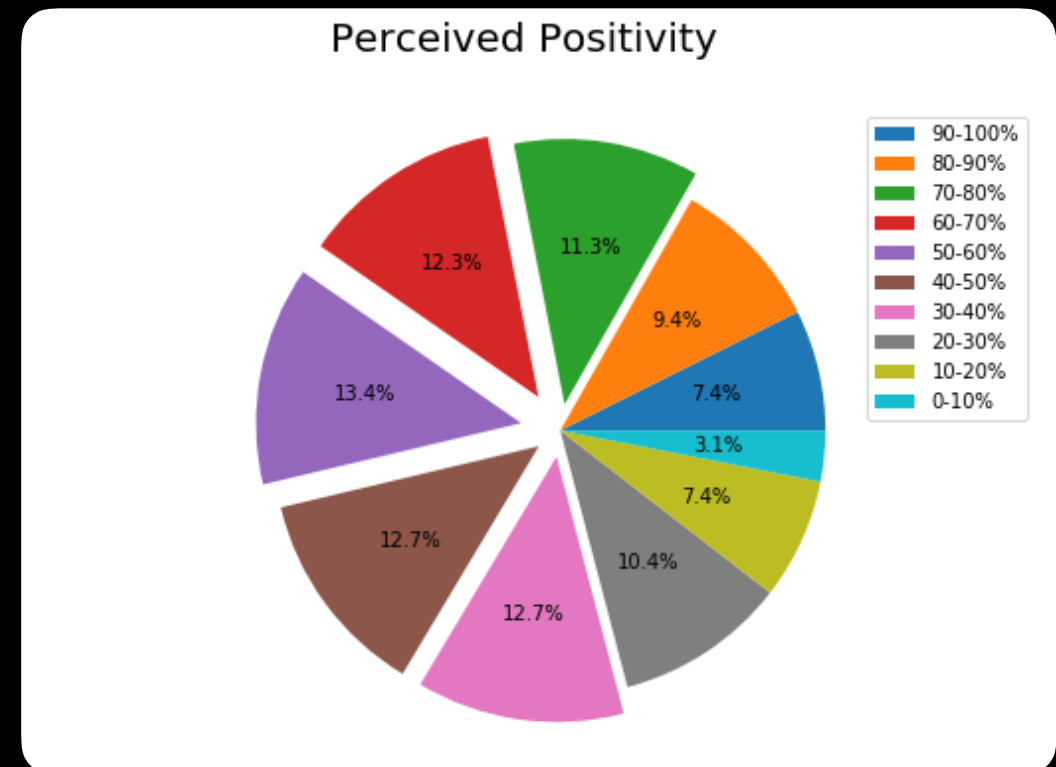- Speechiness is the ratio of vocals present in a song.

Energy & Speechiness

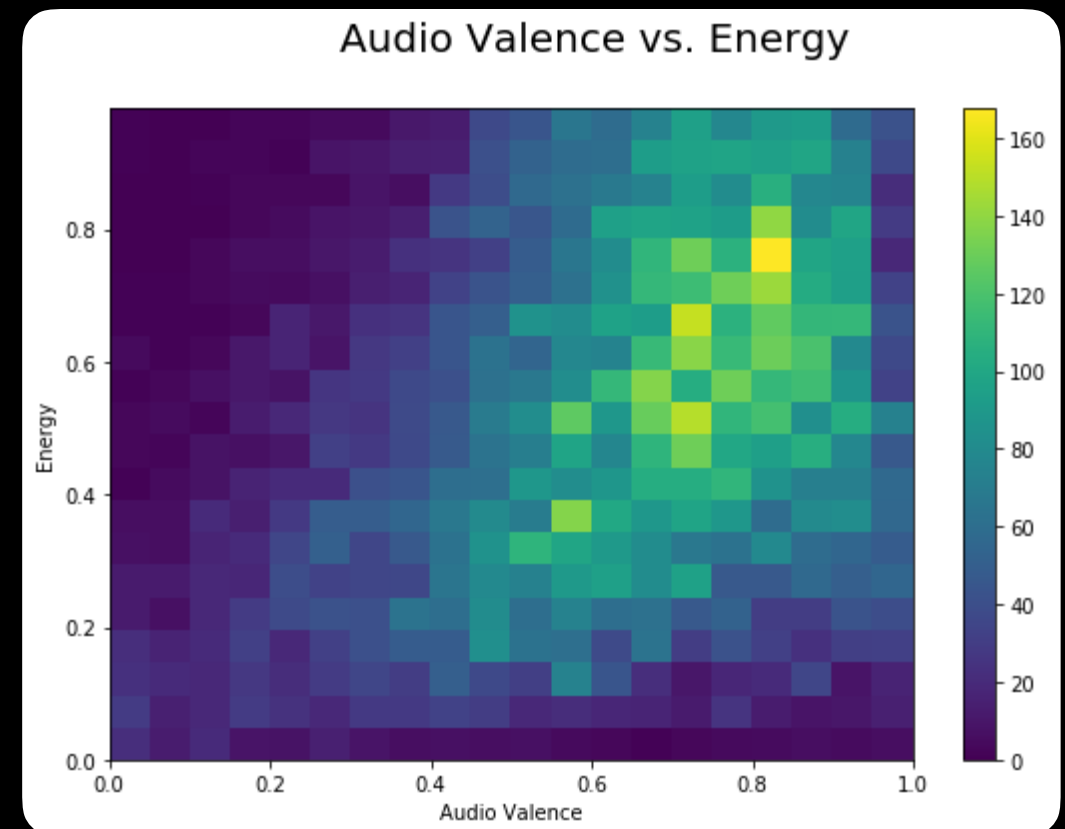- Energy and Speechiness are not correlated

Energy vs. Speechiness

Audio Valence

- Most songs are found between 30-80% "Happiness"

Perceived Positivity

Legend:
- 90-100%
- 80-90%
- 70-80%
- 60-70%
- 50-60%
- 40-50%
- 30-40%
- 20-30%
- 10-20%
- 0-10%

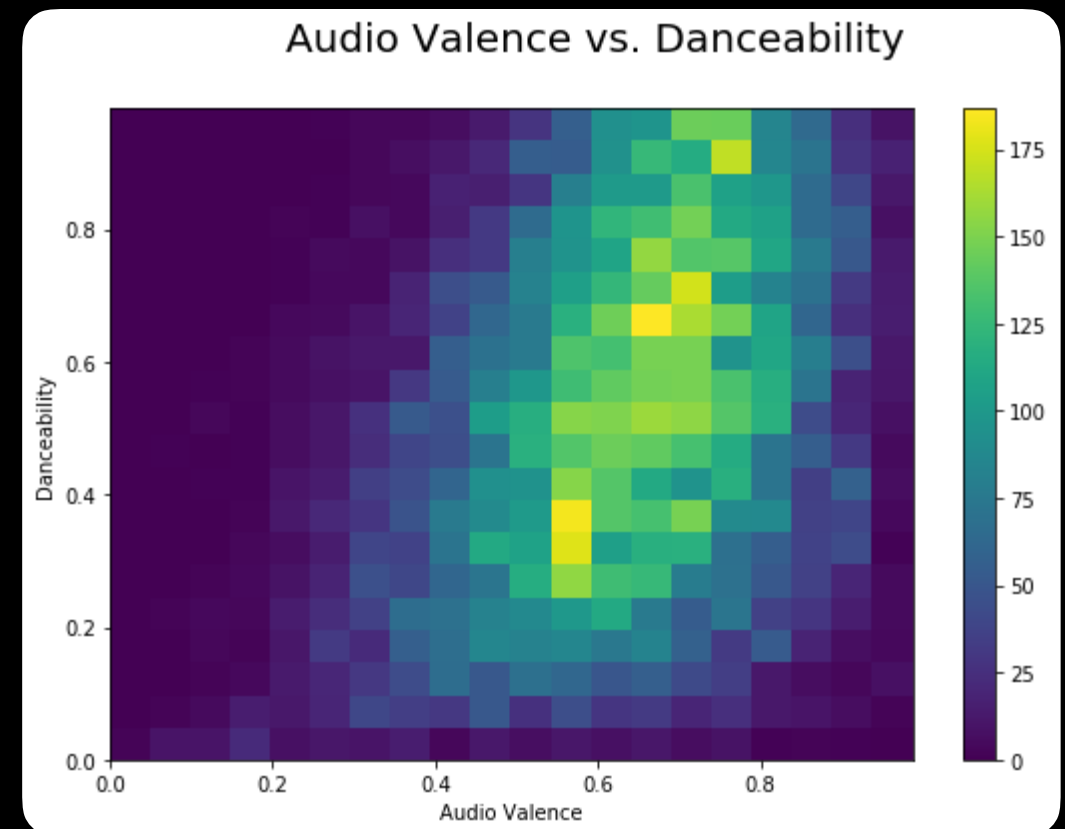Values shown: 11.3%, 9.4%, 7.4%, 3.1%, 7.4%, 10.4%, 12.7%, 12.7%, 12.3%, 13.4%

Audio Valence & Energy

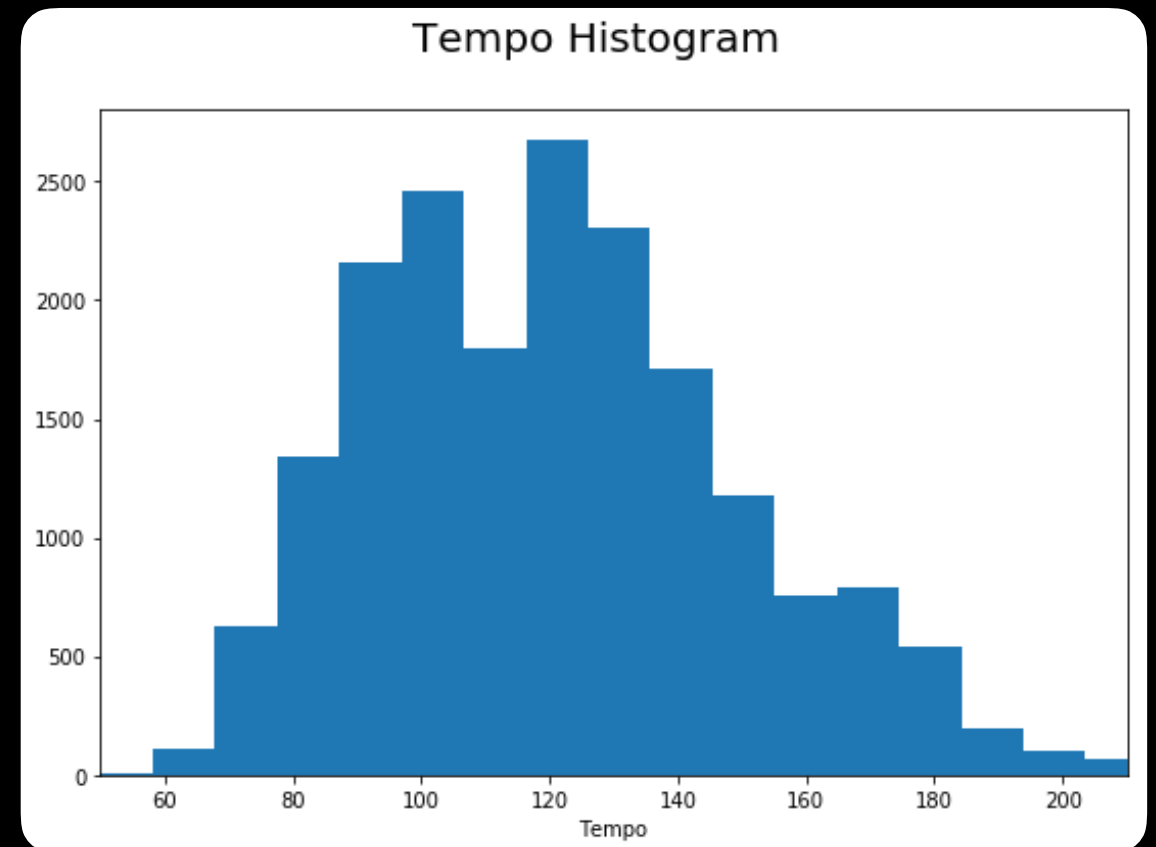- "Happier" sounding songs have higher energy

Audio Valence & Danceability

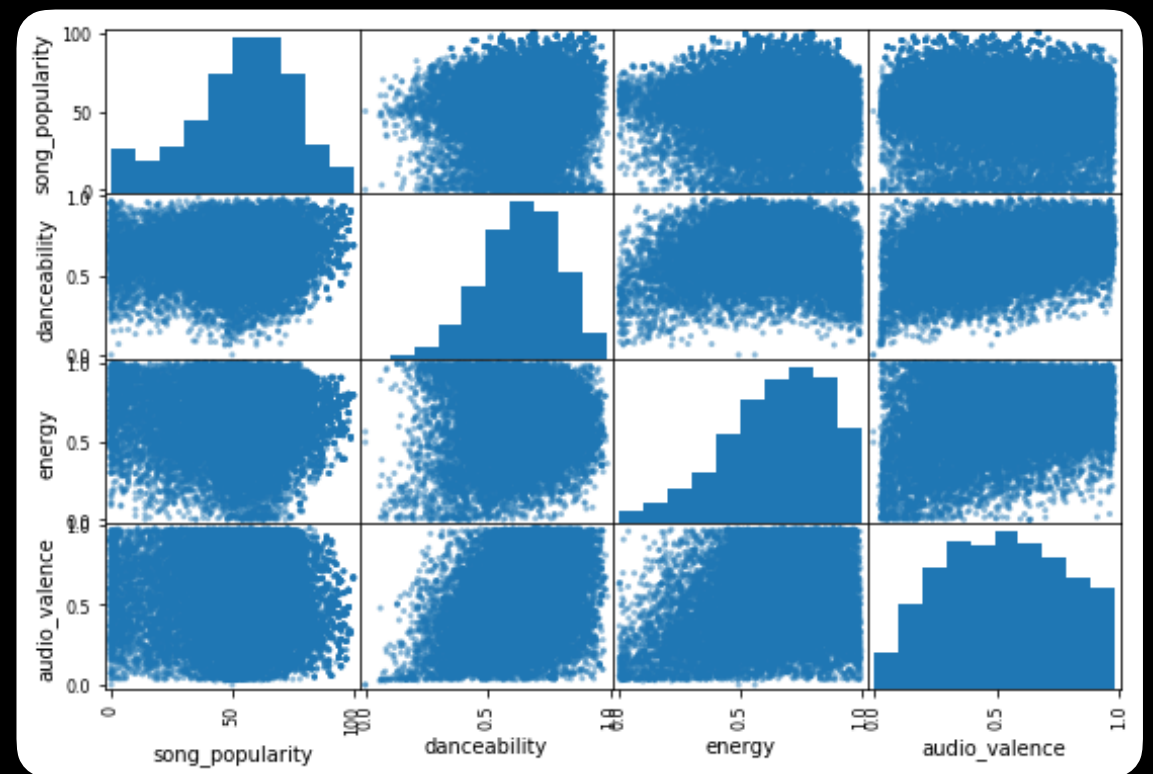- "Happier" sounding songs are fun to dance to

Tempo

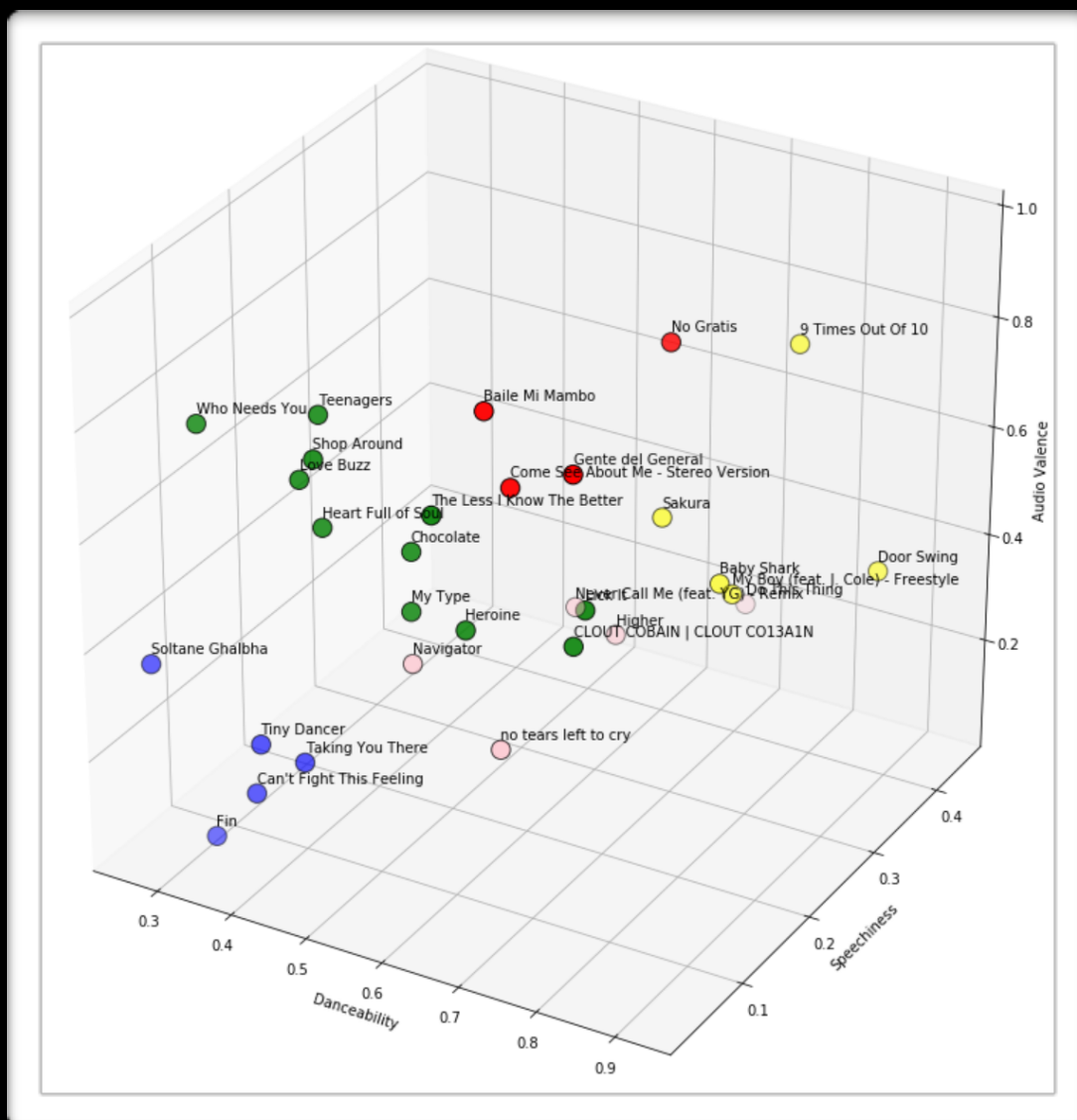- Most songs have 120 BPM, 2 beats per second

Scatter Matrix

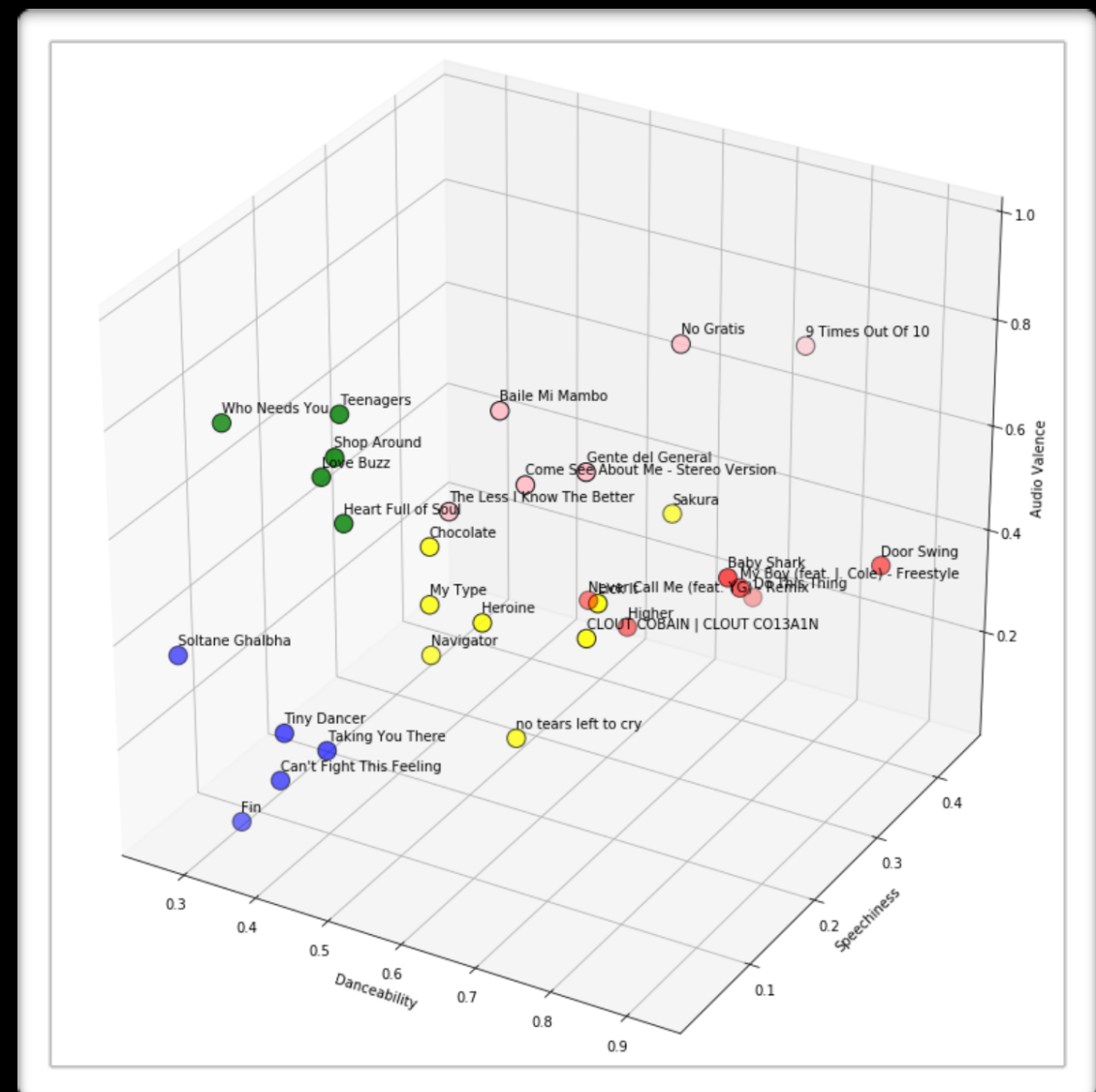- Song Popularity, Danceability, Energy, and Audio Valence is not correlated

Clustering

Clustering with 3 Features

kNN Clustering

kMeans Clustering

Predictions with Model 1

Features Included:

- Danceability

- Speechiness

- Audio Valence

Results of Cluster:

- Fly Me To The Moon

- Love Never Felt So Good

- I Walk the Line

Predictions with Model 2 with All Features

Results of Cluster Rock/Metal:

- Spit Out the Bone

- The Stage

Predictions with Model 2 with All Features

Results of Cluster Rap/Rock/SubGenre:

- Dark Necessities

- Cinderella Man

Summary

Summary

- Spotify API allows access to a lot of songs

- Danceability, speechiness, and audio valence are not enough to make a good model

- When all features are combined, clusters have higher accuracy

- Similar sounding songs suggested can keep people listening to the music they like