# Audio Tracks Segmentation

spotify

Problem definition:

Examine a set of audio tracks extracted from spotify and using clustering techniques to identify patterns and determine the potential genre of a track. Investigate how different the segments are by exploring their characteristics.

Approach:

- Extract data using spotify API
- Delete duplicates if applicable, scale data
- EDA using dimensionality reduction techniques
- Identify patterns and build a model
- **Present results**
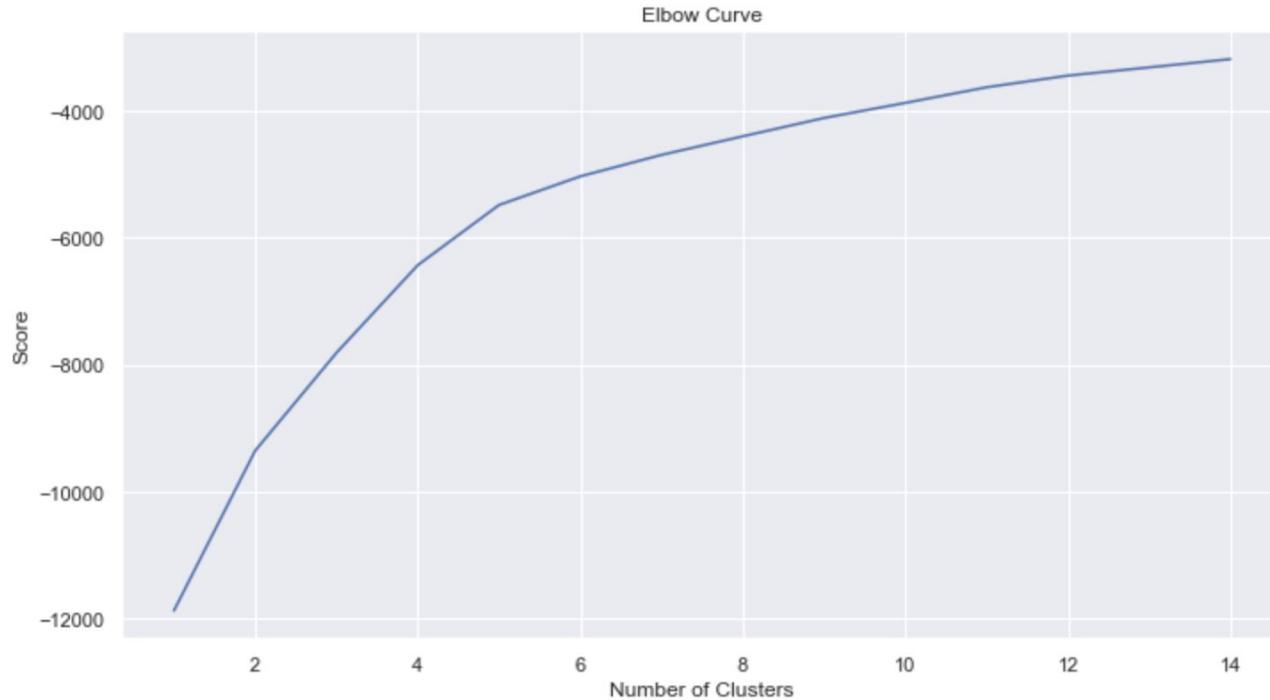
Data Analysis PCA
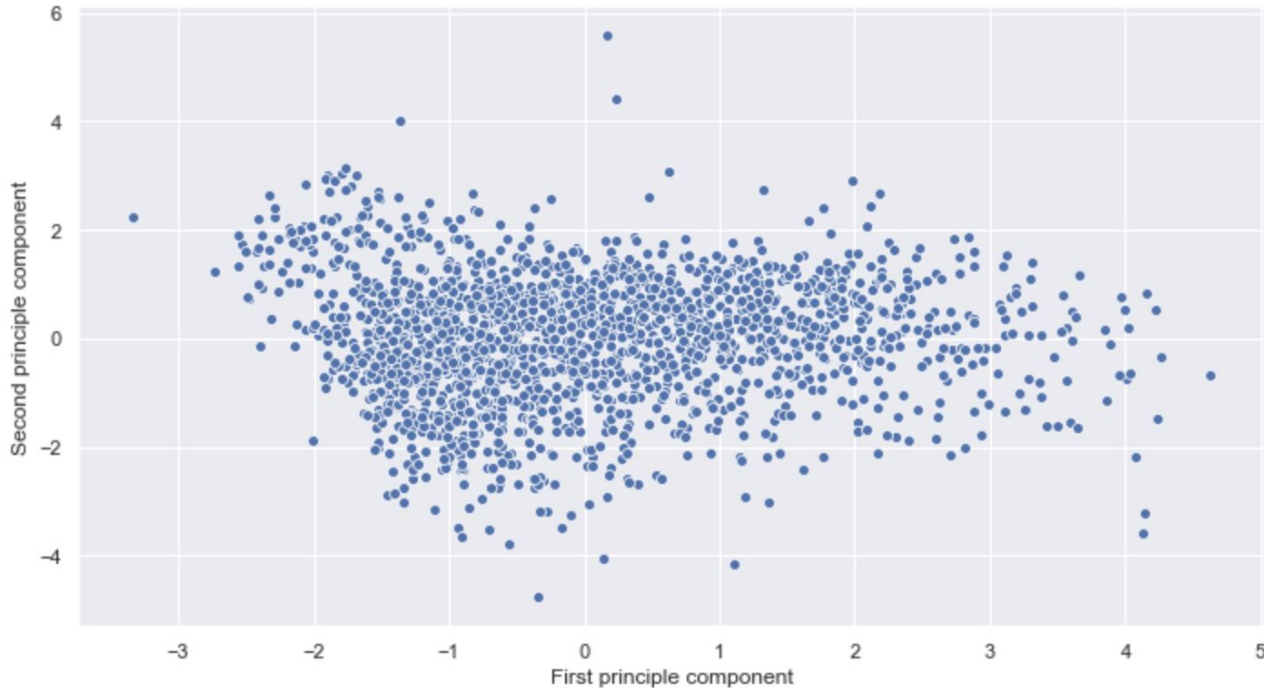
Visualize scaled dataset using dimensionality reduction technique. By reducing from **6** dimensions to **2** dimensions,
50% information of the data is preserved
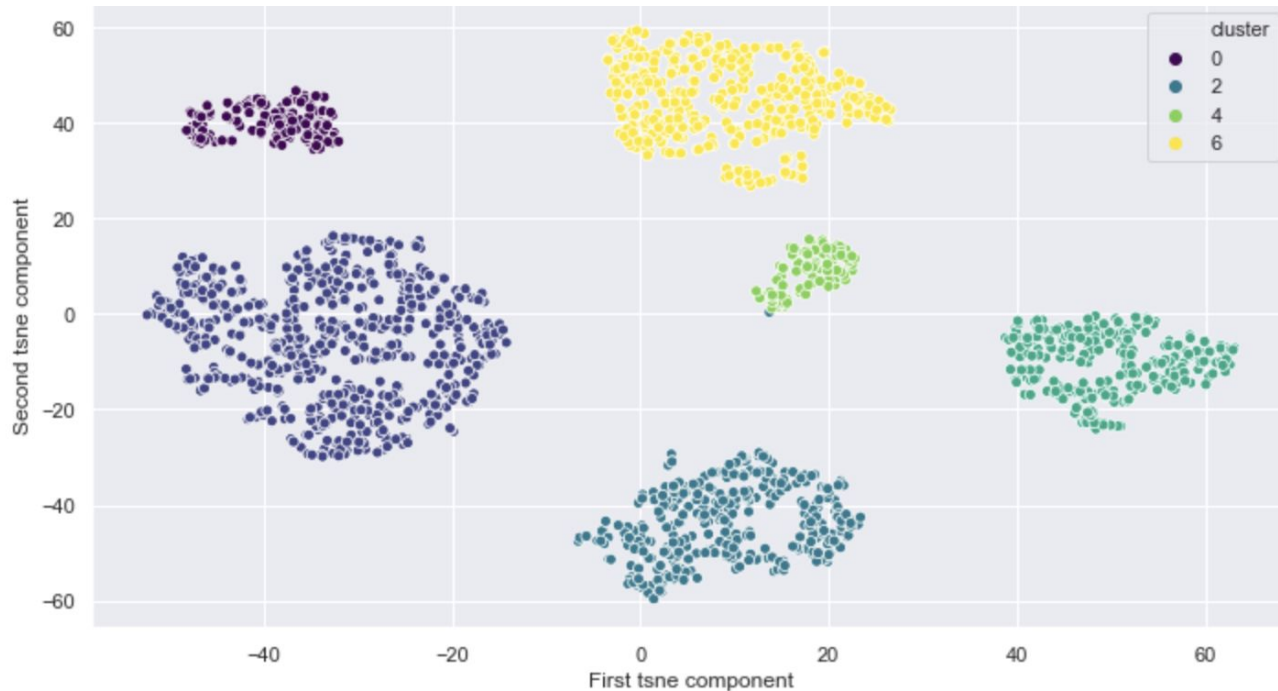ratio: ([**0.300, 0.218**])

Data Analysis TSNE
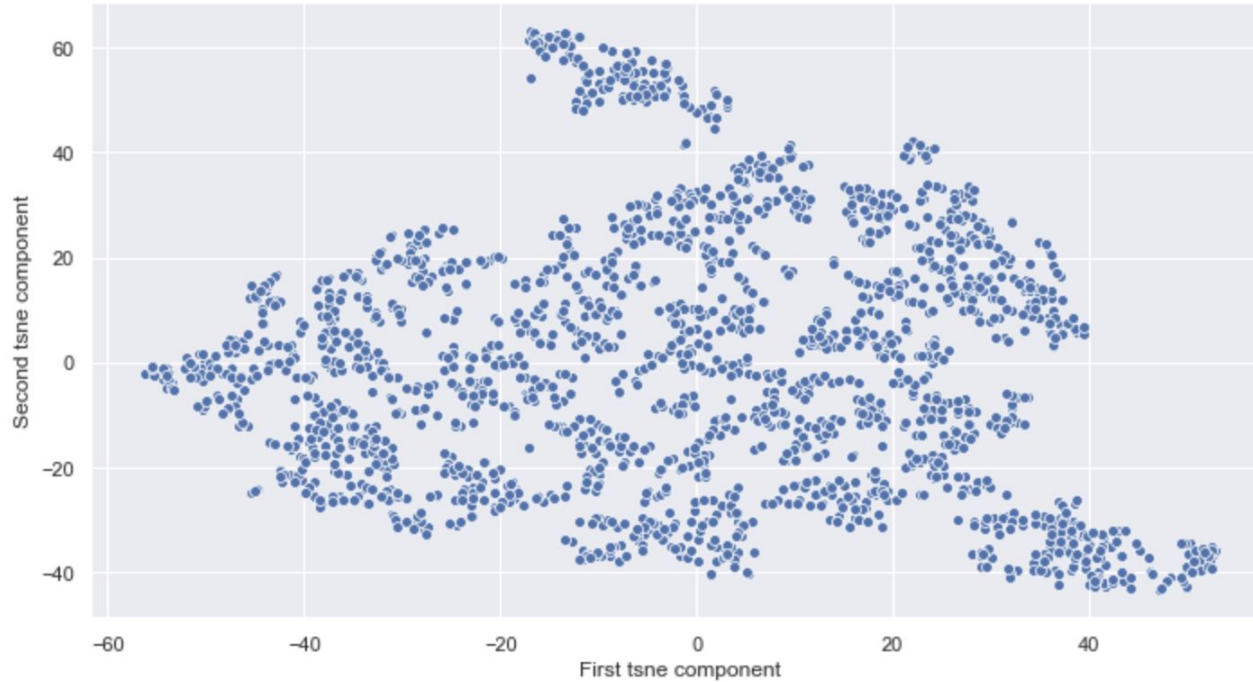
Visualization scaled data using TSNE technique.
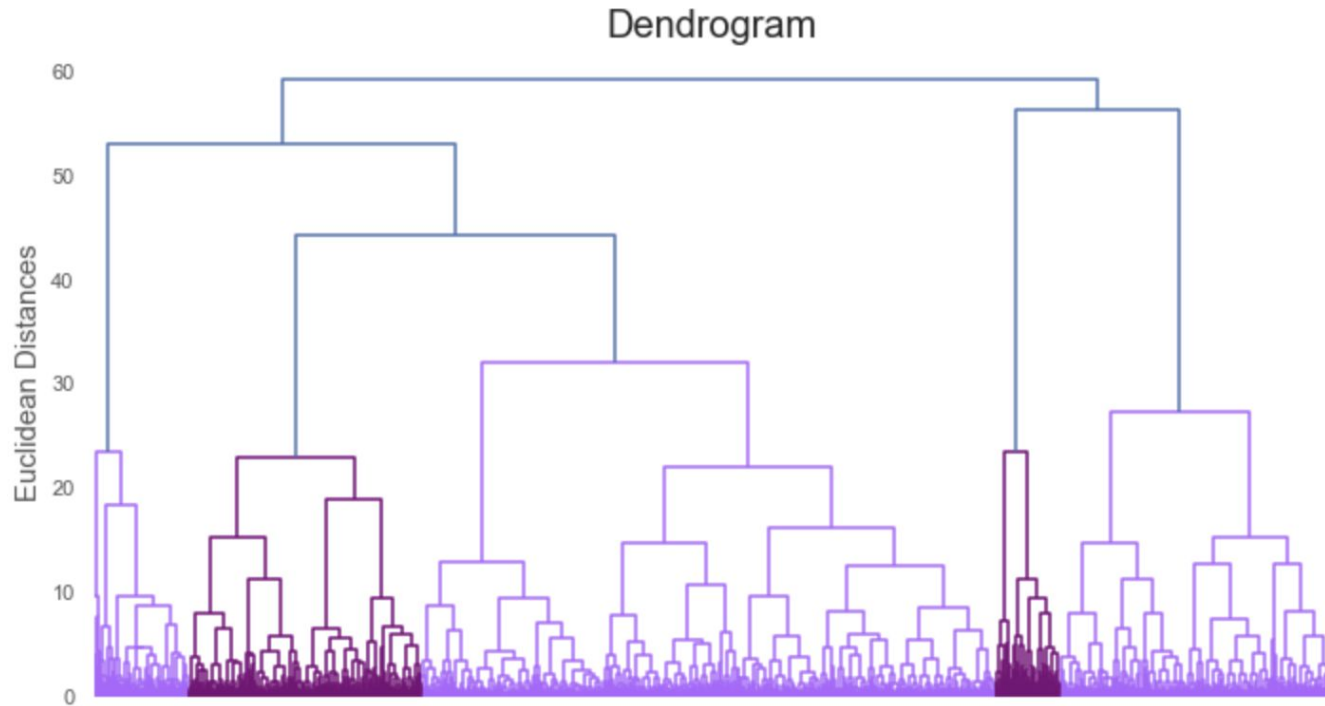
K-Means Elbow Method

Elbow Curve

Based on the elbow method, the number of clusters 6 looks most reasonable because the line became gentle.

K-Means TSNE

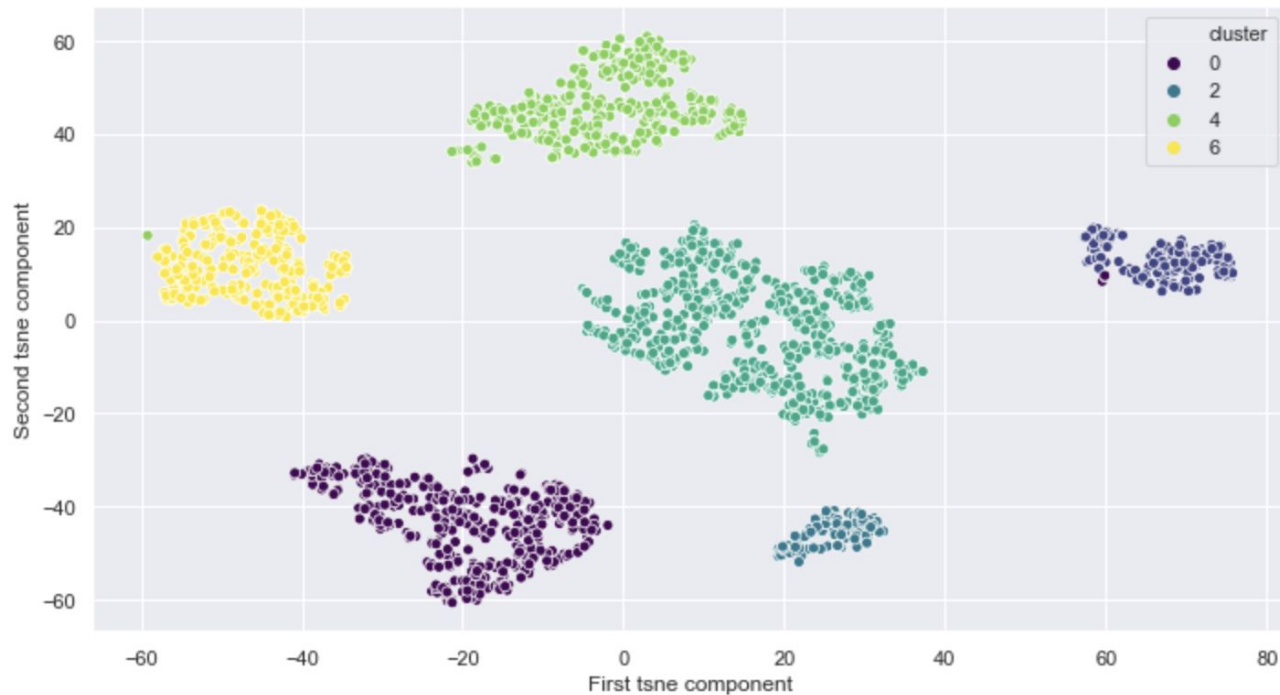There are three big clusters that could represent Pop, Rock and Dance musical class. Other segments could be Classical, Jazz genres.
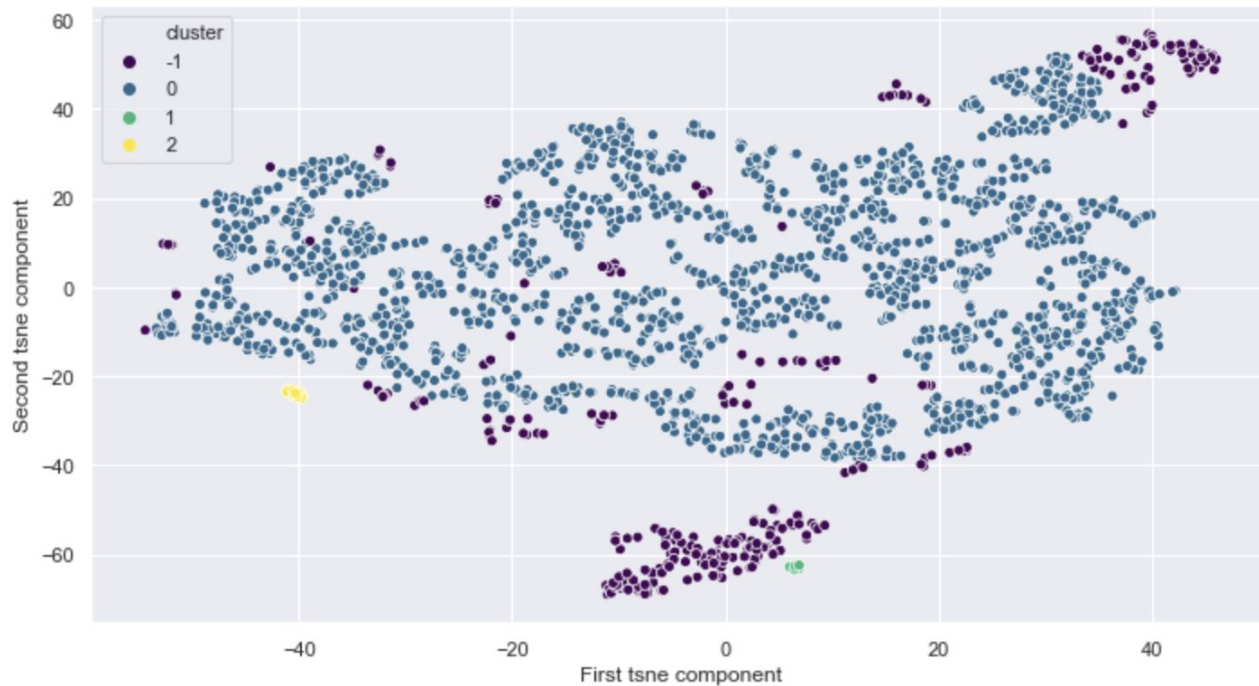
Dendrogram

Dendrogram

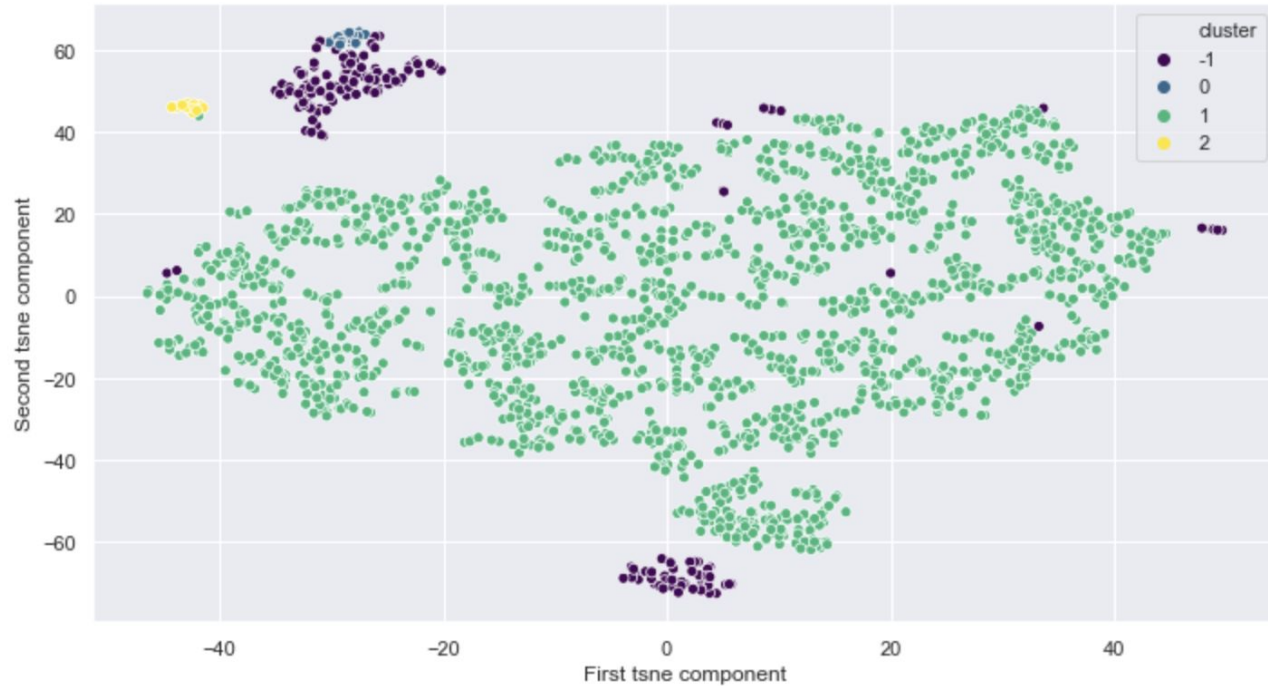According to Dendrogram I chose 6 number of clusters

Agglomerative Clustering TSNE

DBSCAN TSNE

HDBSCAN TSNE

Discovered

Cluster 1 HIGH: acousticness, instrumentalness; LOW: energy, speechiness => **CLASSICAL**

Cluster 2 MID: energy, danceability

Cluster 3 LOW: energy, tempo

Cluster 4 HIGH: energy, danceability; MID: acousticness, speechiness => **POP**

Cluster 5 HIGH: acousticness, instrumentalness; MID: energy => **JAZZ**

Cluster 6 HIGH: energy, tempo; MID: acousticness, speechiness => **ROCK**

Artists genre

	genre	count
94	pop	71
97	rock	64
60	folk	18
42	dance	14
69	hip hop	13
79	metal	11
32	classical	10
72	jazz	6
62	funk	4
54	electronic	3

The most popular genres: Pop, Rock, Folk, Dance, Hip hop, Metal, Classical

But we also could see some representatives from Jazz, Funk and Electronic.

Plot provided by Spotify
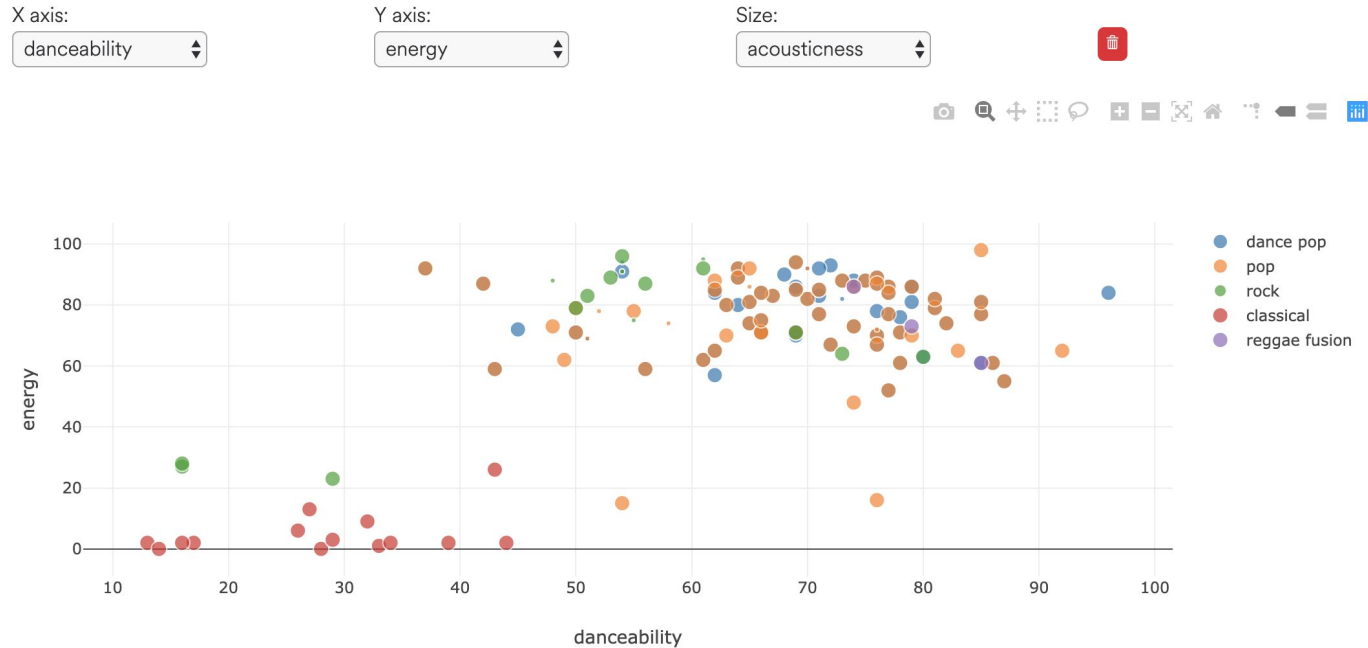
Classical genre has low danceability and energy, but high acousticness.

Rock genre has lower acousticness.

All other genres have similar parameters

Results

After clustering analysis and visualization I got a rough idea of audio track genres that were in 1996. I was able to identify only four clusters: **Classical, Pop, Rock, Jazz.**

I checked my results by extracting artists genres of the audio tracks from 1996. The most common genres were defined by grouping of music type. **Pop, Rock, Folk, Dance, Metal and Classical**

In the plot, that provided spotify there is a significant difference only for **Classical** music (low: danceability, energy) and some **Rock** music (lower: acousticness)

Popular music in the 1990s (wiki): The Red Hot Chili Peppers, Nirvana, No Doubt, Green Day, The Offspring, Marilyn Manson, Aerosmith, Bon Jovi,Backstreet Boys, *NSYNC, Christina Aguilera, Britney Spears, Jennifer Lopez and Destiny's Child, Michael Jackson