

MBTI vs Spotify playlists

1. Introduction

The aim of this project is to investigate associations between Myers–Briggs Type Indicator (MBTI) types and music taste according to personalized spotify playlists.

The MBTI type consists of the following four aspects:

Extroversion (E) - Introversion (I)

Intuition (N) – Sensing (S)

Thinking (T) – Feeling (F)

Judging (J) – Perceiving (P)

A person's MBTI type is determined by the combinations of four letters, where each letter is dominant in each of the four pairs. For example, the author's MBTI is INTP, meaning that the author's personality is introverted, intuitive, thinking, and perceiving.

In this dataset from Kaggle, 4072 Spotify playlists were collected and analyzed by the users' MBTI and the music features.

Some music features of interests are described as follows:

Acousticness: confidence level of whether a track is acoustic (0 - 1)

Danceability: how suitable a track is for dancing (0 - 1)

Energy: perceptual measure of intensity and activity (0 - 1)

Instrumentalness: confidence level of whether a track contains no vocal (0 - 1)

Liveness: confidence level of a track being live and have audience presence (0 - 1)

Loudness: overall loudness in decibels(dB) (-60 - 0)

Mode: modality of a track (major=1, minor=0)

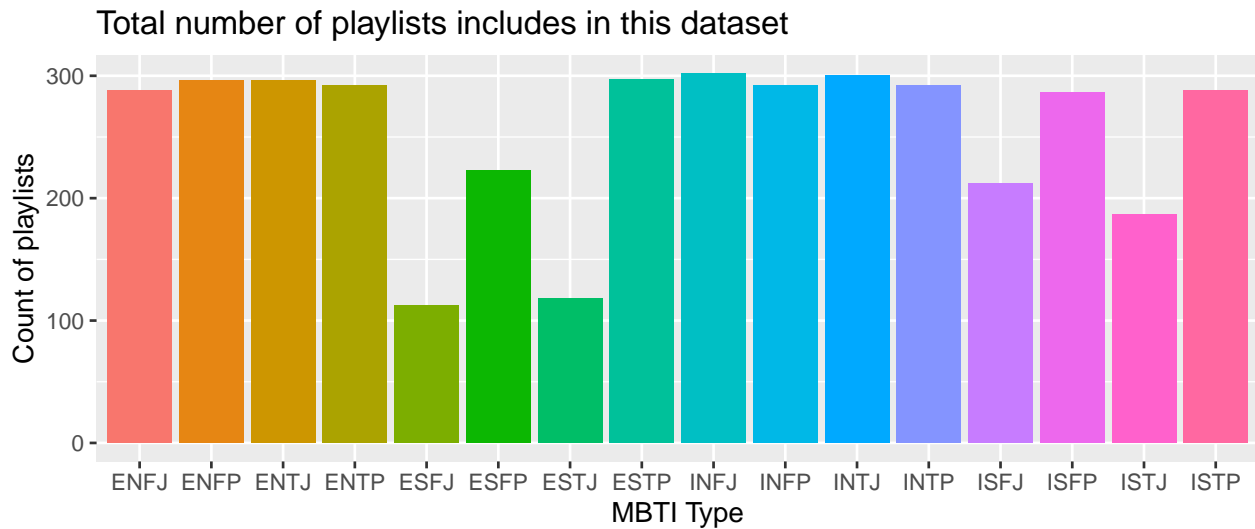
Speechiness: proportion of spoken-words (0 - 1)

Valence: Musical positiveness (0 - 1)

We are interested to see if a particular MBTI or any of the four personality categories are more likely to appreciate specific music features. We will generate various figures to aid our investigations.

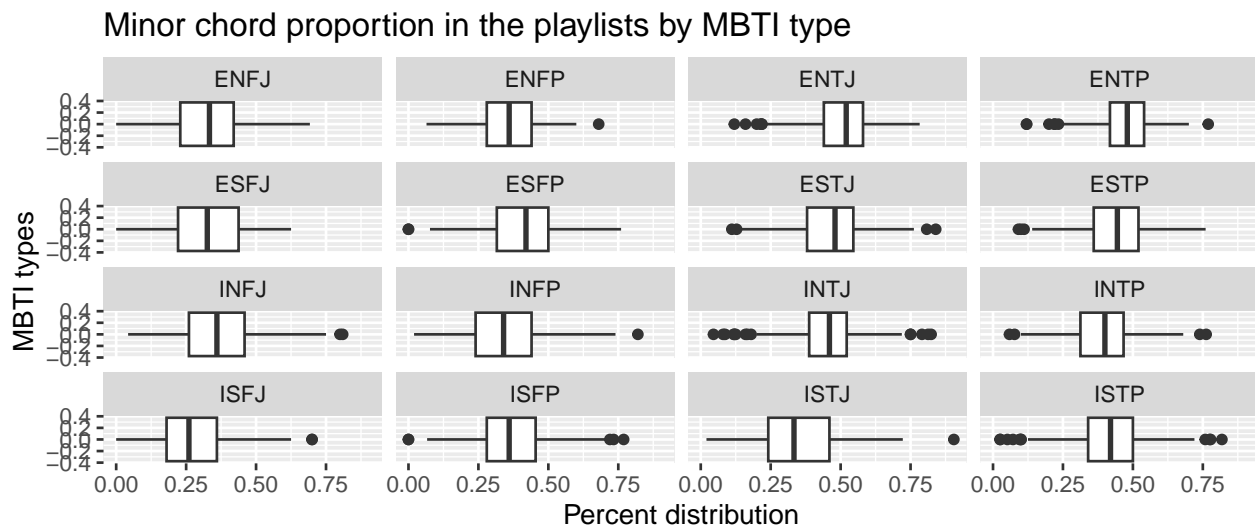
2. Raw Data

First, we counted the number of playlists collected in this dataset by each MBTI type.

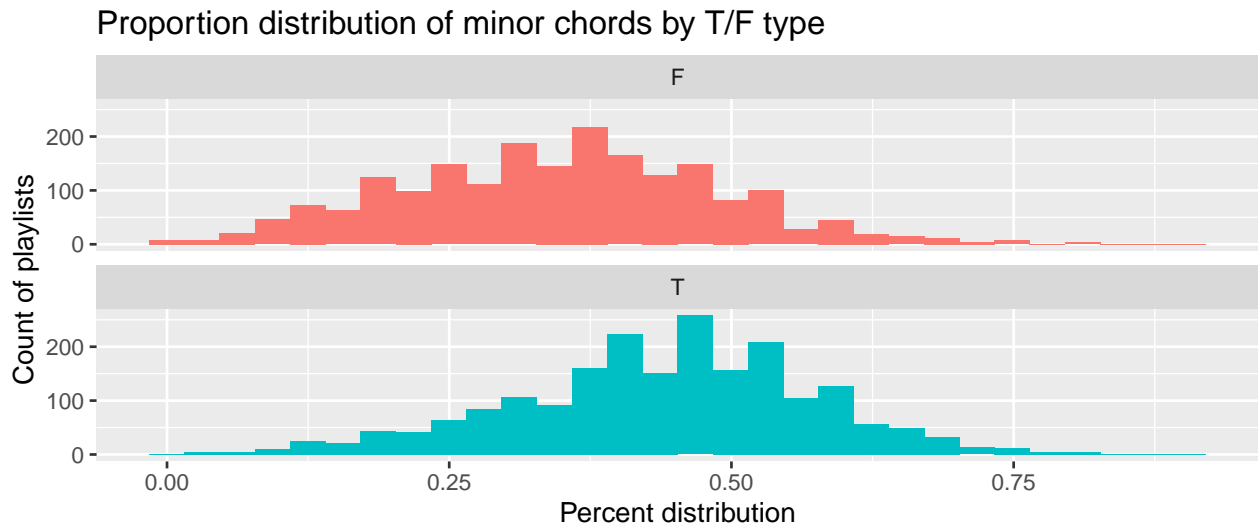


It is interesting how the number of playlists by people with SJ function pair are much lower than the other function pairs. However, the author of the dataset did not specify how data were collected, but the counts of other MBTI types are all around 300, so we suspect that this difference is due to lack of data. Are SJ people are less likely to listen to music? Or are they perhaps reluctant to provide personal information for survey purposes?

3. T/F vs Minor Modality



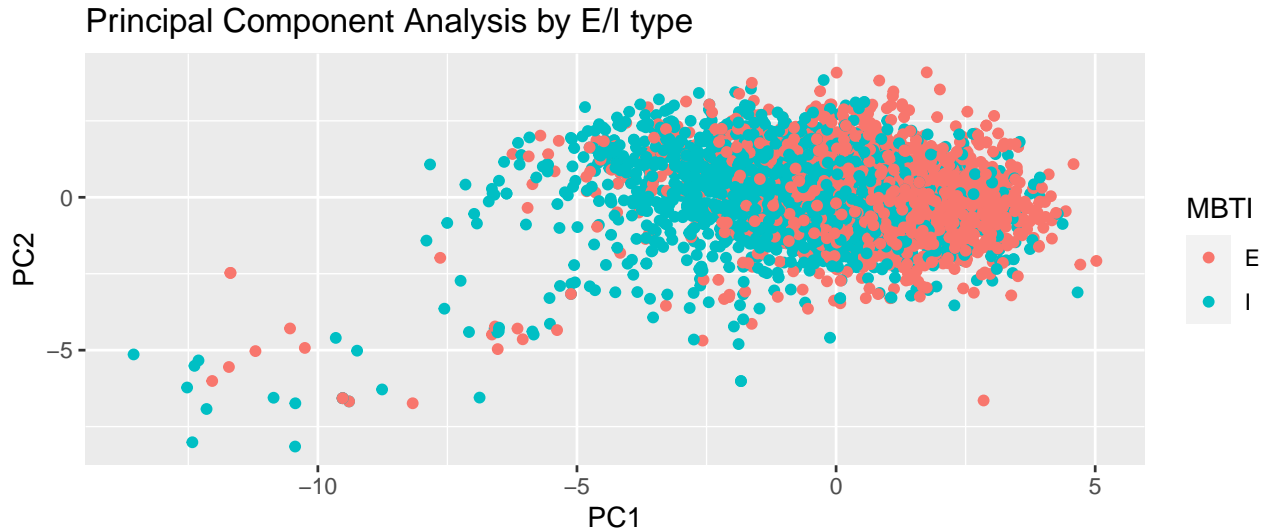
We also plotted the proportion of songs with minor modality in the playlists of each MBTI type. It can be observed that the left two columns seem to have a lower average percent than the right two columns, which corresponds to the T/F distinction. We would like to delve deeper by making a second plot comparing the histograms of the minor chord proportion distribution for Thinking(T) vs Feeling(F)) personality types.



There seems to be a significant shift in the distribution of minor chords among the T and F playlists. This is unexpected for me because minor chords are usually associated with sadness or moodiness, which are emotions that F people will more likely to empathize. However, the plots indicates that T people, well-known for their preference for logs over emotions, are more likely to listen to minor chords.

4. E/I vs Music preference

To delve deeper into the high dimension of the dataset, we conducted a principal component analysis of the various music attributes for dimension reduction. The scatterplot of the first two components only showed a relatively clear boundary in the following plot group by E/I type.

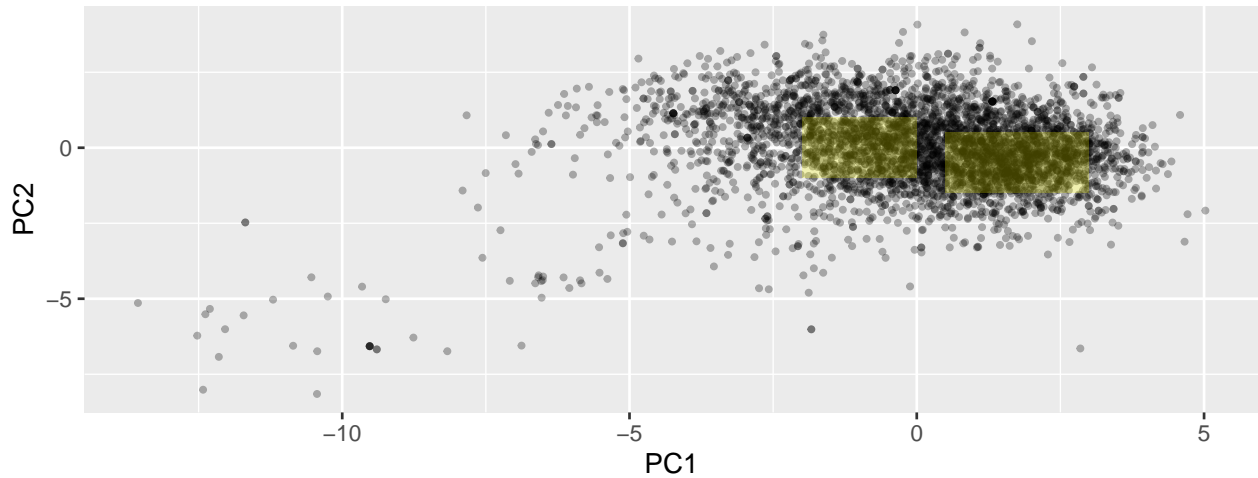


It looks like the playlist of people with Extroverted personality and those with Introverted personality tends to have slightly different preferences in music. This is also fascinating because it implies that music taste is associated with sociability. However, due to dimensional reduction, it is not clear which music features are the most determining factors that E/I types prefer differently.

5. Delving deeper

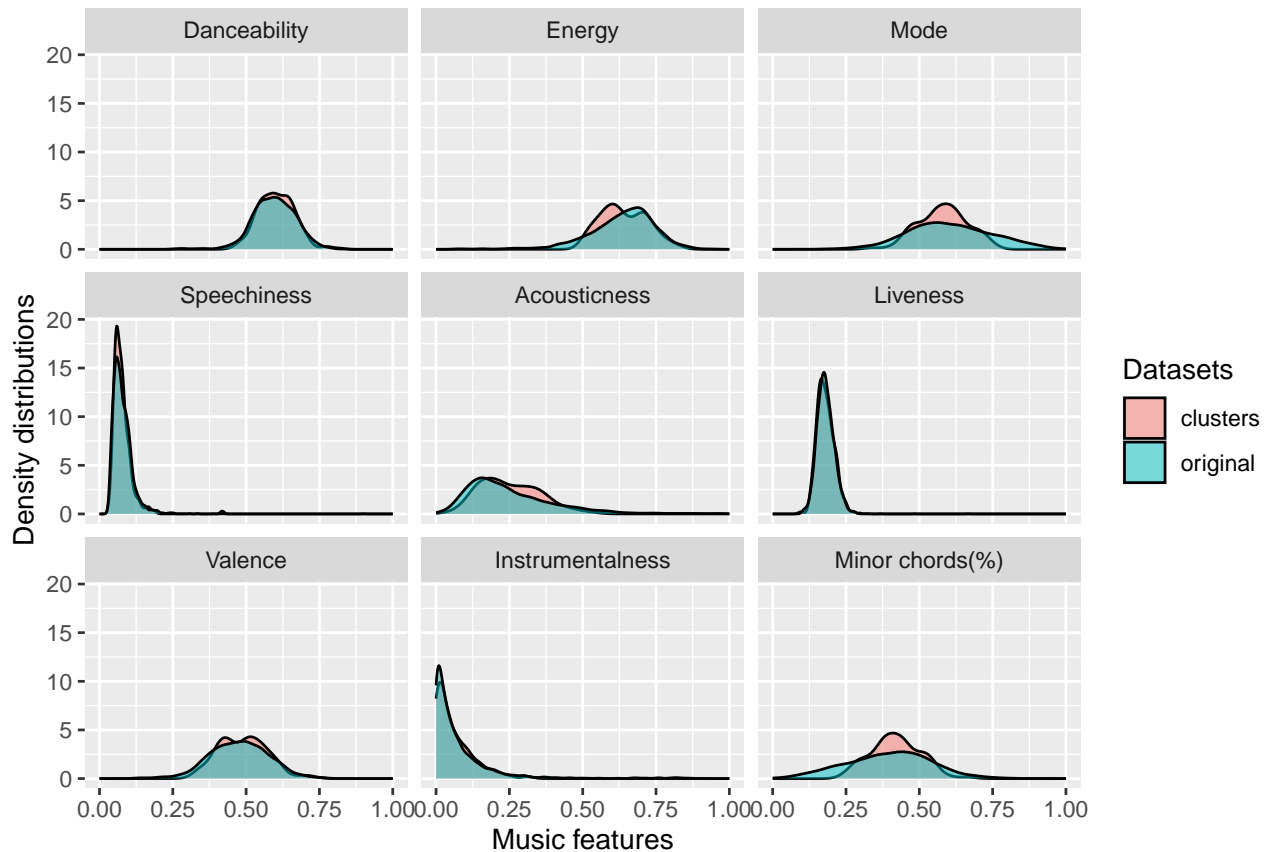
We proceed to perform a cluster analysis on the two clusters highlighted in yellow below.

Selecting two clusters in the PCA plot



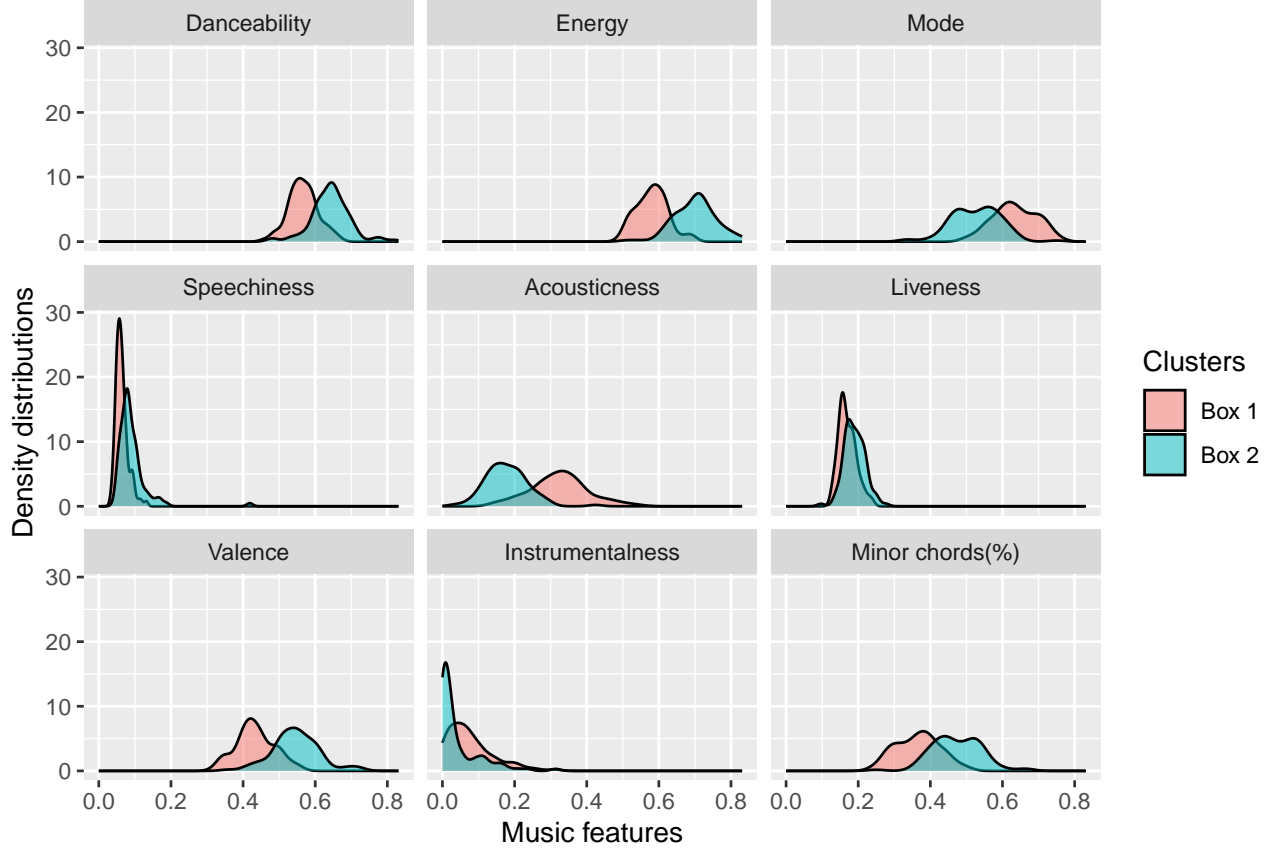
Define the box on the left in the I type cluster as Box 1 and the box on the right in the E type cluster as Box 2.

Comparing the distributions for Original vs Clustered data



Comparing the distribution in the clusters with the original dataset, we can see some disparities in the density distributions for Energy, Mode, and Acousticness. Overall, the distributions are very similar since the two clusters represents a good proportion of data in the clump. Thus we proceed to compare the density distributions among the two clusters.

Comparing the distributions for the two Clusters



We can clearly see that the distributions for danceability, energy, mode, acousticness, and valence are significantly shifted. Since the two clusters are side-by-side, the perceived differences are mostly caused by their relative positions on the x-axis, which is the first component from PCA, representing the axis in the data state space with minimized variations in other dimensions. Therefore, we can hypothesize that the music features with large shifts are the most determining factors in explaining the differences of music tastes for different MBTIs.

6. Discussion

This study showed the following associations between MBTI and Spotify music preference:

1. T types are more likely to listen to songs with minor chord than F types.
2. E and I types have different music preferences, especially for danceability, energy, acousticness, and valence.

It is worth noting that these are only observed associations from the data set, and there can be other unknown confounding variables that have effects on both MBTI types and music features. However, a causal effect is not necessary to be established if the music artists want to cater to specific MBTIs.

In the future, it would also be interesting if the music artists's MBTI can be included in the dataset. A lot of Spotify users follow their favorite artists and listen to their works more than other music, which can also play a role in determining their music preferences.

Additionally, we can construct a machine learning model using this dataset. The model can use supervised learning to train using the music features and predicts the MBTI types that will listen to them (thus adding to their personal playlist). Hence, Spotify can improve their music recommendation algorithm to suggest songs with specific music features that fit the user's personal preference based on their MBTI type.