

[様式第4号の1]

令和5年3月30日

## 令和4年度 学生自主研究成果報告書

教 育 本 部 長 様

学生自主研究グループ名	Akita note	
研究課題名	Spotify APIを用いたデータ分析	
研究代表者（学生）	学籍番号	B24D034
	氏 名	前川恵吾
指導教員	学 科	経営システム工学科
	氏 名	松原佳亮

学生自主研究の報告書を別紙のとおり提出します。

## Spotify API を用いたデータ分析

システム科学技術学部 経営システム工学科

前川恵吾

指導教員 システム科学技術学部 経営システム工学科

松原佳亮

### 1. 研究目的と背景

音楽のサブスクリプションサービスができたことにより、外国の曲やマニアックな曲を聴くことが容易になった。また一方で、サブスクリプションサービスの API (application programming interface) を利用することで、楽曲の特徴や視聴頻度などの情報を容易に取得し、分析できるようになっている。例えば、大手音楽サブスクリプションサービスである Spotify の API では踊りやすさやエネルギー (激しさ) など独自の指標を取得できる。

その中で、国や言語によって曲調に違いがあるのではないかと仮説を立てた。また売れている曲には特徴があるのではないかと仮説を立てた。

そこで、楽曲及びその再生回数等のデータを分析して共通点や傾向を調査し、上記の仮説を検証する。国や言語による曲調の違いや売れている曲の特徴を明らかにすることで、より個人の興味にあった楽曲の推薦などへの活用が期待できる。

### 2. 研究対象

言語による音楽性の特徴・違いを明らかにするために、世界で使用している人口が多い言語を使用している国、音楽の文化が成熟している国のデータを使用する。

- ・アメリカ 英語-15億3100万人

- ・スペイン スペイン語-5億1600万人

- ・日本 日本語-1億2600万人

### 3. 使用するデータ

使用するデータは、Spotify Charts から取得した週間のトップ100のデータである。探索的データ分析・可視化には、2021年7月第1週(2日~8日)のデータを用いた。検定には、2022年7月第1週(1日~7日)のデータを用いた。

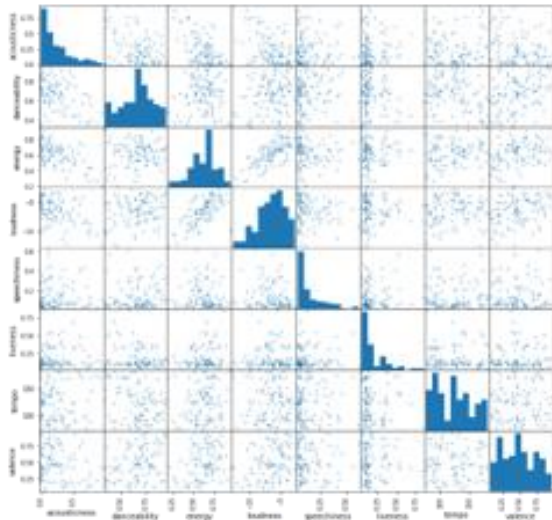
### 4. 分析ツール

探索的データ分析・データの可視化には Python を用いた。データの取得には Spotify API を用いた。

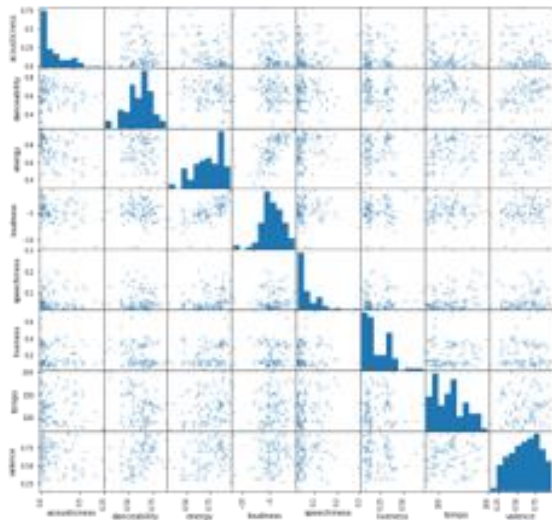
### 5. 探索的データ分析

まず、Spotify API で得られる指標の特徴を捉え、特に国ごとの顕著な特徴が現れている指標を抽出することを目的に探索的データ分析を行った。

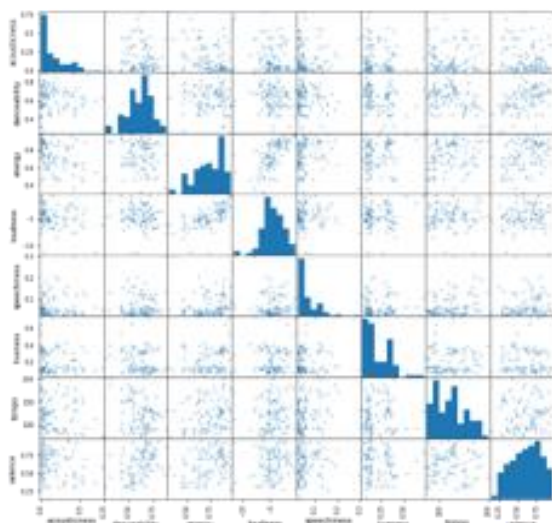
探索的データ分析を行った結果、energy と valence の指標において、国によって数値や分布に特徴や違いがみられた。特に、スペインと日本では energy と valence で大きい値を取る傾向にあった。



＜アメリカ＞



＜スペイン＞



＜日本＞

図1. 散布図行列

＜energy＞

曲の激しさ

曲が高揚していて、音が大きく、騒がしければ1に近づく。

＜範囲＞0.0-1.0

＜valence＞

曲の明るさ・暗さ

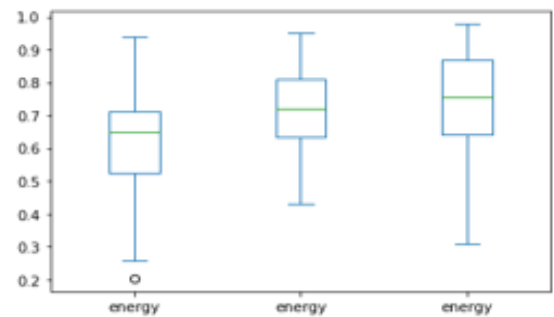
1が明るい（ハッピー・陽気さ・多幸福感を感じる時）

0が暗い（悲しみ、憂鬱さ、怒りを感じる時）

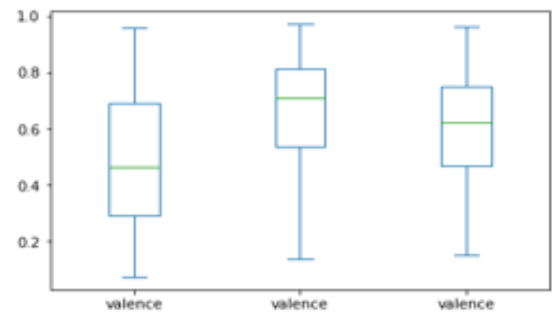
＜範囲＞0.0-1.0

6. 国ごとの指標の比較

国ごとの指標の比較では、energy と valence のデータ（2021年）を plot 関数で可視化した。可視化するには、箱ひげ図（図2）とヒストグラムで比較（図3）した。



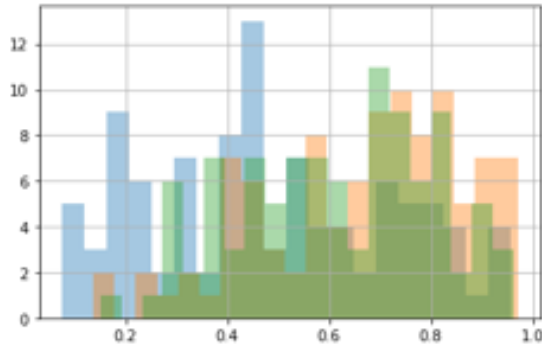
＜energy＞



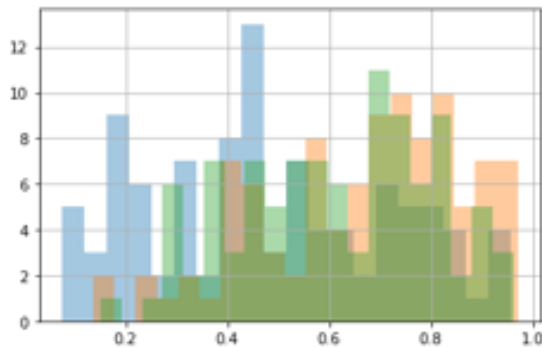
＜valence＞

図2. 箱ひげ図

（左からアメリカ、スペイン、日本）



<energy>



<valence>

図3. ヒストグラム

青:アメリカ オレンジ:スペイン 緑:日本

国によって音楽性に本当に差があるのかを調査するため、検定を行う。検定を行う際、検定の多重性の問題を回避するために、チューキーの方法を用いる。可視化する際は2021年のデータを使っているのに対して、検定を行うときは2022年のデータを使用する。

帰無仮説 H0:energy/valance に違いがない  
 対立仮説 H1:energy/valance に違いがある  
 有意水準:0.01

	アメリカ	スペイン	日本
energy	0.61±0.15	0.72±0.12*	0.74±0.16*
valence	0.48±0.24	0.67±0.20*	0.61±0.19*

\* p<0.01 for vs. アメリカ

表4. 平均値

各指標の国別の違いと結果に対する考察については以下の通りである。

<energy>

表4より、日本の平均値が0.74付近で最も高い。図3より、スペインは大部分が0.5以上である。図2より、アメリカは平均値が一番低く、範囲が広い。スペインと日本は激しい曲が多い傾向にある。

Energyにおいて、アメリカ vs スペインには違いがあると言える。また、スペイン vs 日本では違いがあるとまでは言えない。よって、スペイン vs 日本は曲調について共通した要素を持っていると言える。

<valence>

表4より、スペインの平均値が一番高い。明るい曲が最も多い。図3より、日本は範囲が最も狭く似たような曲が多い傾向にある。表4より、アメリカの平均値0.48と一番低い。図3より、アメリカは0.2, 0.4付近に曲が集中して暗い曲が多い。図2より、アメリカは、範囲が最も広く曲調の多様性がある。

valenceにおいてアメリカ vs スペイン、アメリカ vs 日本には違いがあると言える。また、スペイン vs 日本では違いがあるとまでは言えない。よって、スペイン vs 日本は音楽性について共通した要素を持っていると言える。

## 7. 結論

アメリカ vs スペイン, アメリカ vs 日本の間には曲調や音楽性に違いがあることが明らかになった. 一方で, スペイン vs 日本の間には曲調や音楽性に共通した特徴があることが示唆された. 特に, 日本は激しく・明るい曲調の曲が多い傾向がみられた.

今回の研究では, 国や言語による曲調の違いや特徴を明らかにしたが, 売れている曲のデータや年齢別のデータを使うことで, より個々人の興味にあった楽曲の推薦などへの活用が期待できる.