

What makes songs popular?

Analysis on Spotify music



林修同 吳彬睿 蔣思陽 何欣育

Outline

- 1 Motivation
- 2 Dataset description
- 3 Data Preprocessing
- 4 Analysis
- 5 Conclusion
- 6 Future work

Motivation



Idea

1. 愛音樂
2. 3C世代，獲取數據更便利

Interest

1. 每年流行的歌曲是否有所差異？
2. 是否音樂特徵會影響歌曲熱門度？

Dataset Description

Spotify web API

2000年～2019年歌曲資料

每年各10,000筆數據

共20萬筆

The screenshot shows the Spotify for Developers console interface. At the top, there is a navigation bar with the Spotify logo and the text 'Spotify for Developers'. To the right of the logo are navigation links: 'DISCOVER', 'DOCS', 'CONSOLE' (which is underlined), and 'COMMUNITY'. On the left side, there is a sidebar menu with the following items: 'CONSOLE', 'Albums', 'Artists', 'Browse', 'Follow', 'Library', 'Personalization', 'Player', 'Playlists', 'Search', 'Tracks', and 'Users Profile'. The main content area features a blue banner with the text: 'Psst! Check out our brand-new Web API Reference in beta! And be sure to tweet us your feedback at @SpotifyPlatform on Twitter!'. Below the banner, the title 'Search for an Item' is displayed in a large, bold font. Underneath the title, the text 'Search for an Item' is repeated. A table provides details about the endpoint:

Description	Search for an Item	DOCS
Endpoint	https://api.spotify.com/v1/search	
HTTP Method	GET	
OAuth	Required	

Dataset Description

18個Attributes

Song Name	Tempo	Loudness
Artist Name	Duration(ms)	Speechiness
Release Date	Danceability	Valence
Key	Acousticness	Instrumentalness
Mode	Energy	Artist Popularity
Genres	Liveness	Popularity

Data Preprocessing

Remove

Missing Value, Duplicate Sample 22,328筆

Variety Artist 無法分辨來自哪個歌手 193筆

New Song 尚未有popularity資訊 1,432筆

⇒ 剩餘 **176,047** 筆

Check 是否有資料錯誤

Analysis

1. Popular Songs & Popular Artists

定義Popularity > 75% 的歌曲為Popular

定義Popularity < 25% 的歌曲為Unpopular



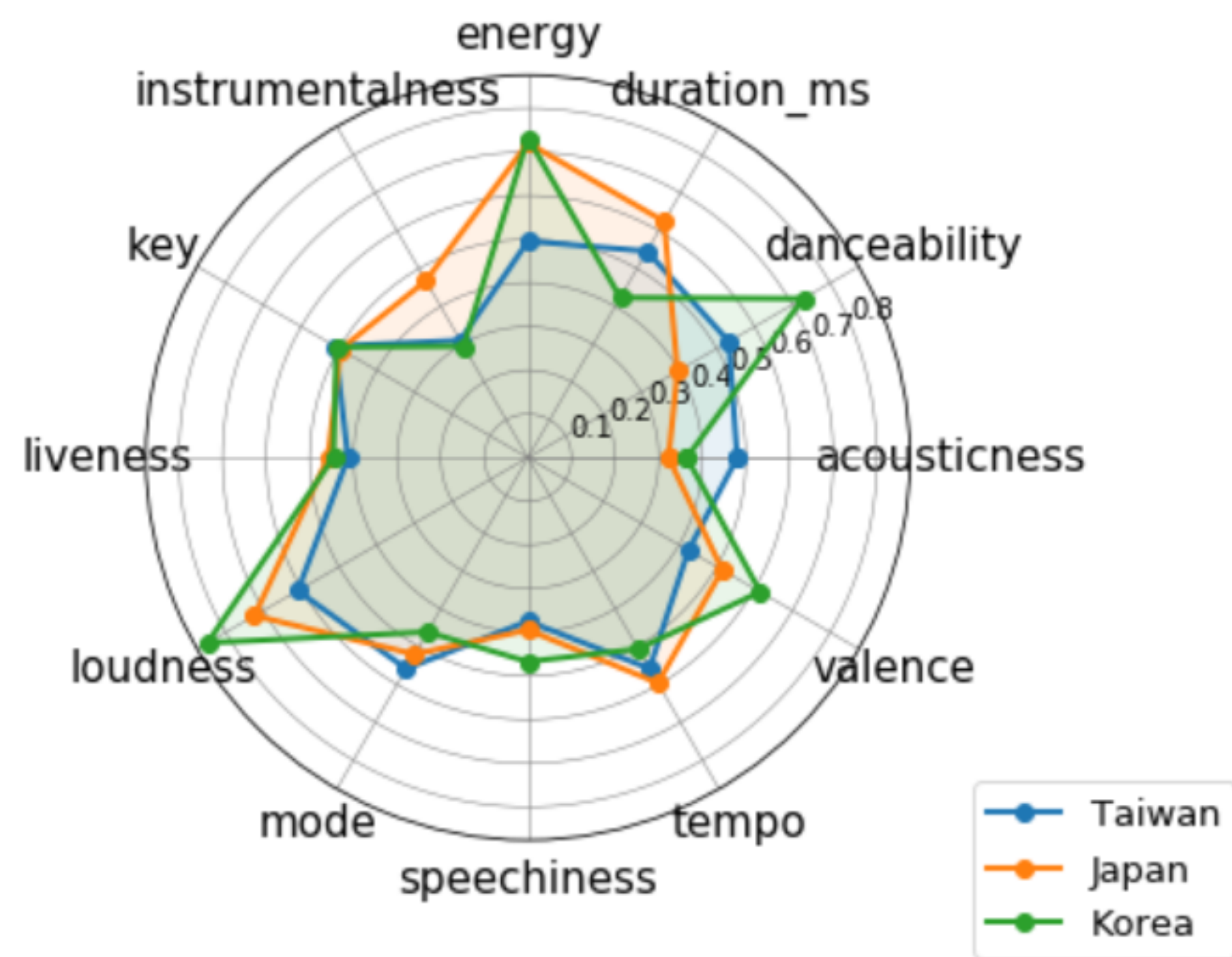
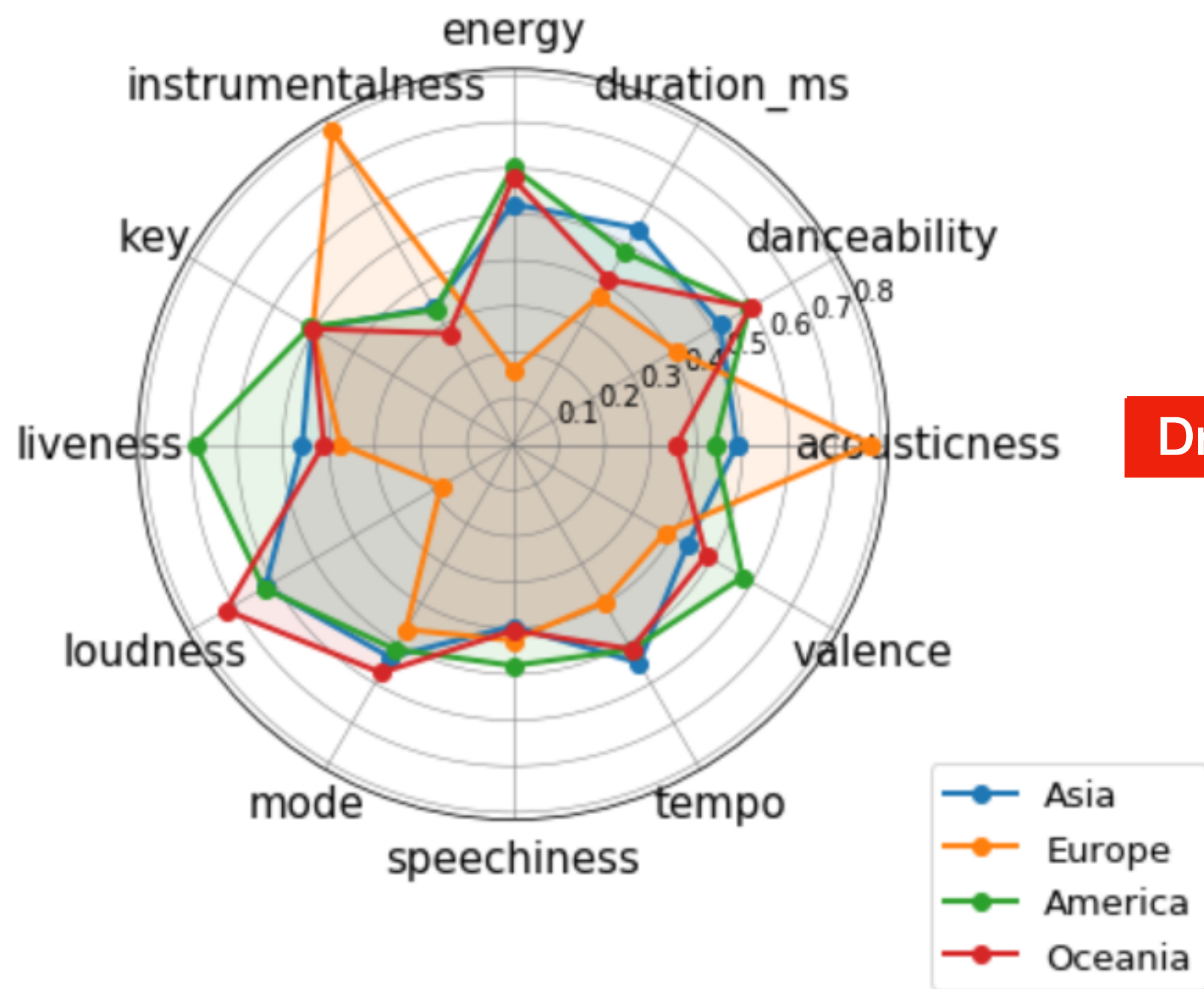
Song



Artist

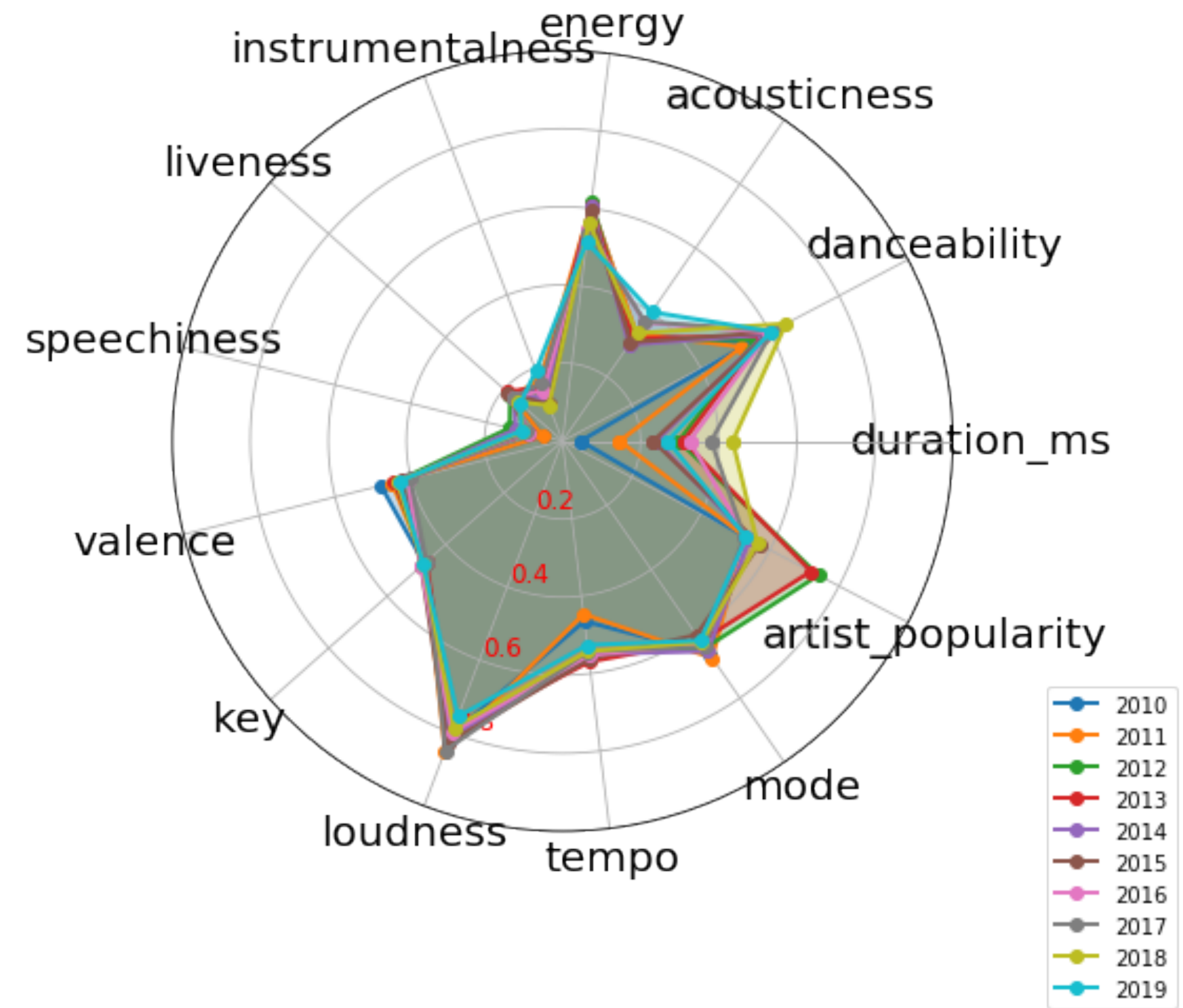
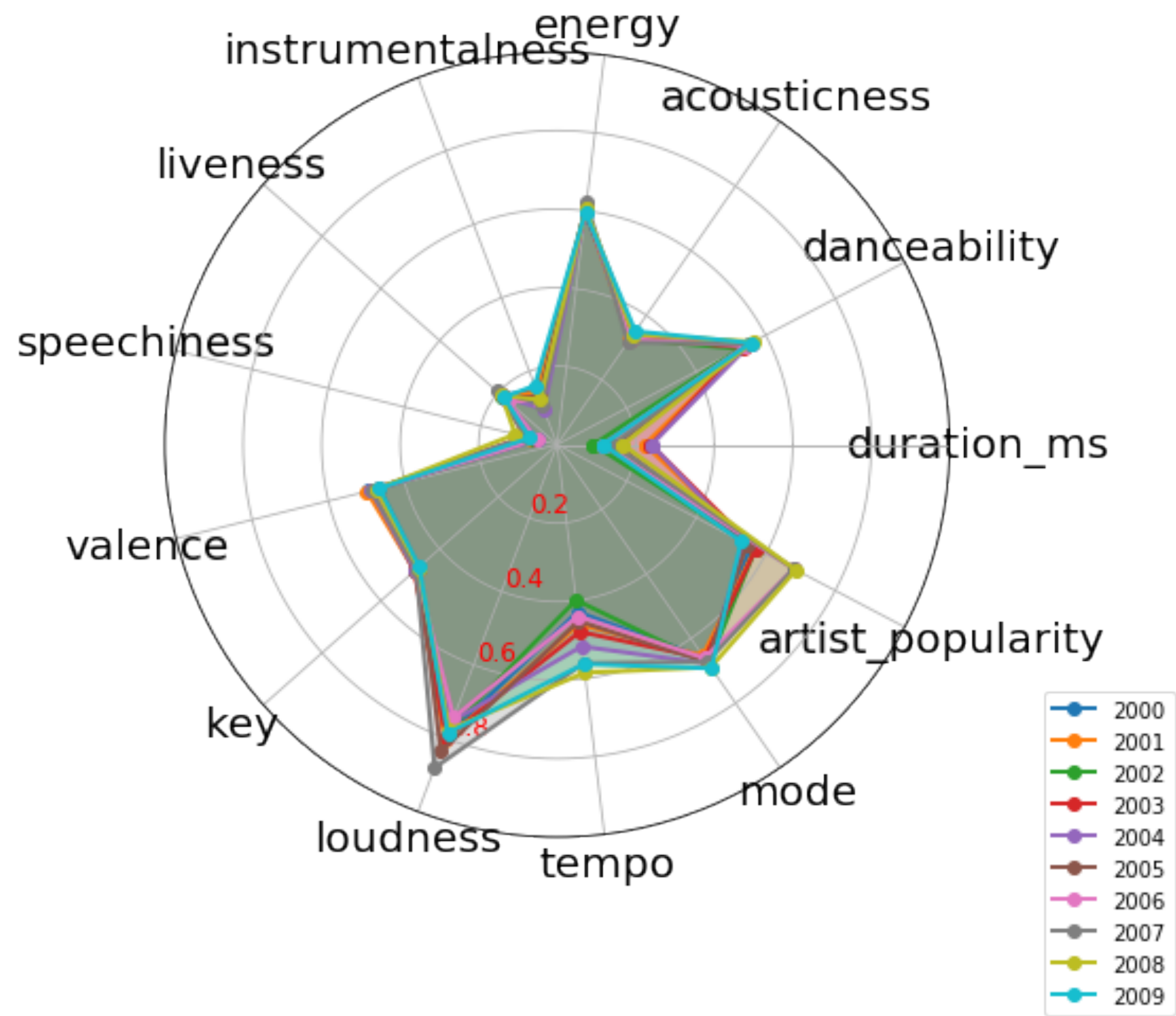
Analysis

2. 以地區來看，歌曲的特徵差異



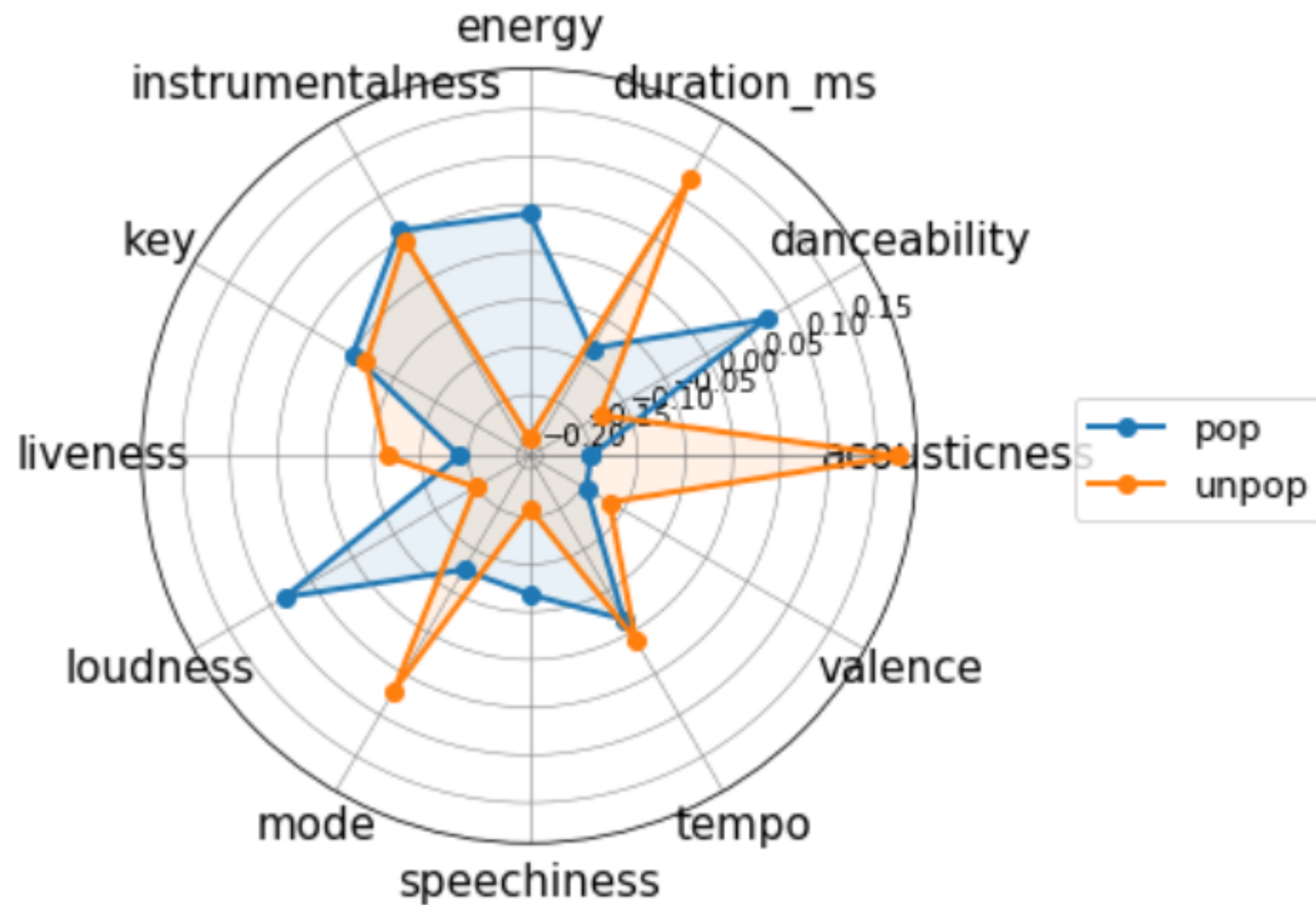
Analysis

3. 以年度來看，流行歌曲的特徵差異



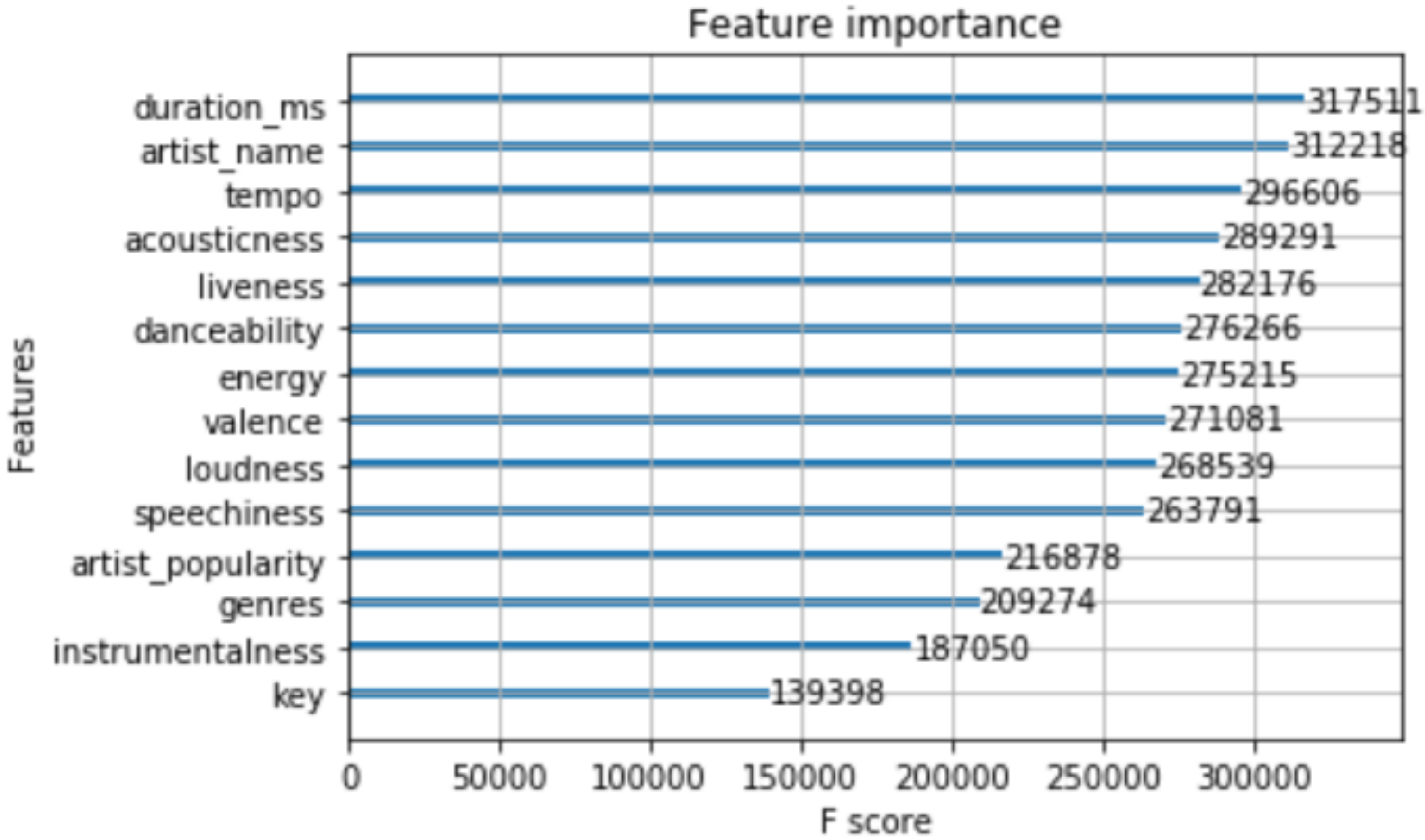
Analysis

4. 2000-2019 熱門跟冷門歌曲的特徵差異



Analysis

5. Important Factor 針對Global Music



透過XGBoost model,
找出較影響popularity
的音樂特徵

RMSE: 9.736615

6. Clustering

Concept

Discover the characteristics of Mandarin popular songs in each year

⇒ recommend similar songs

How

Construct clusters of each year by K-means

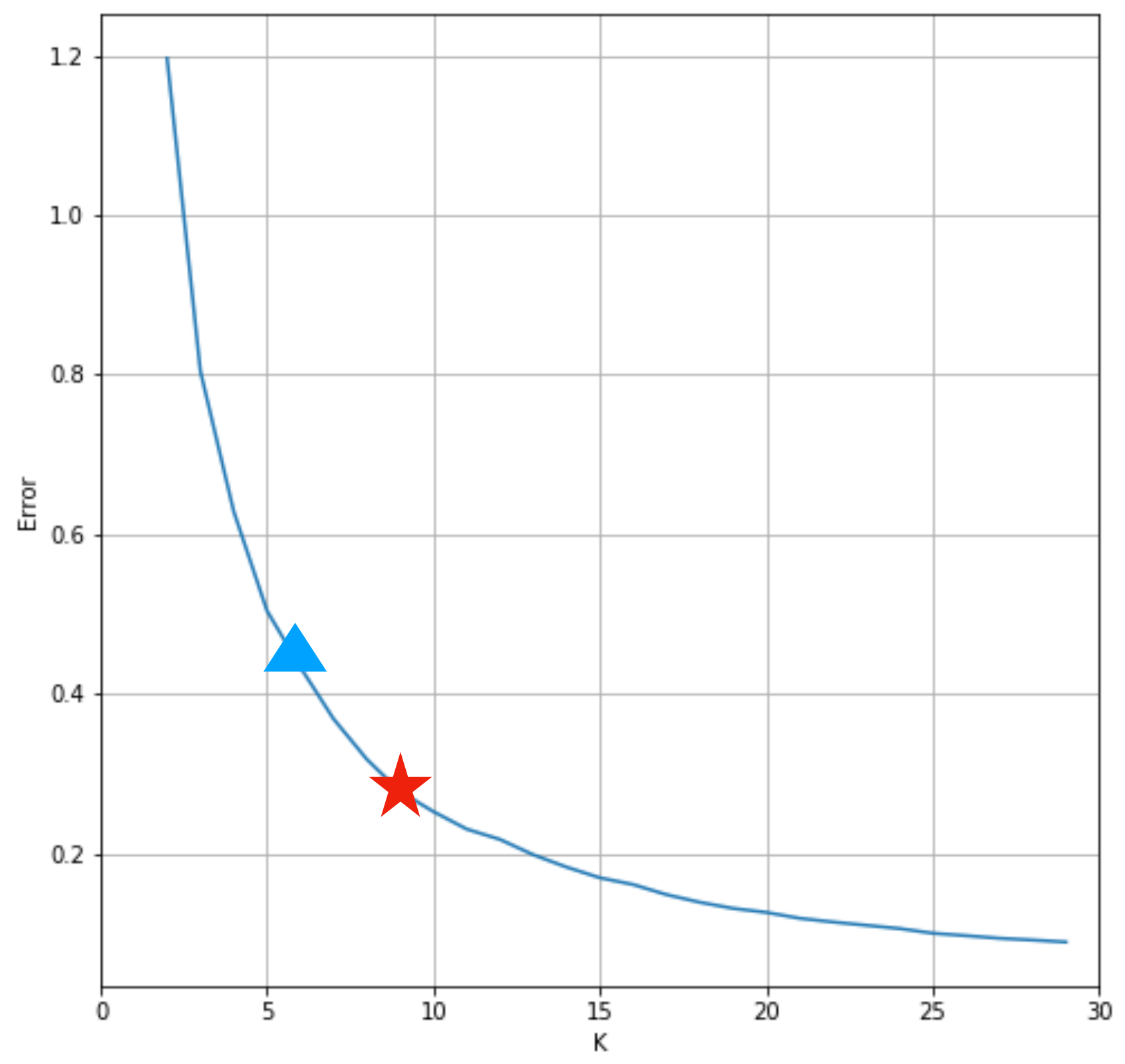
Following example

1. Randomly choose a song released in 2018
2. Find out which cluster of 2019 to be belong to
3. Recommend 15 songs of 2019 that are most similar to this song

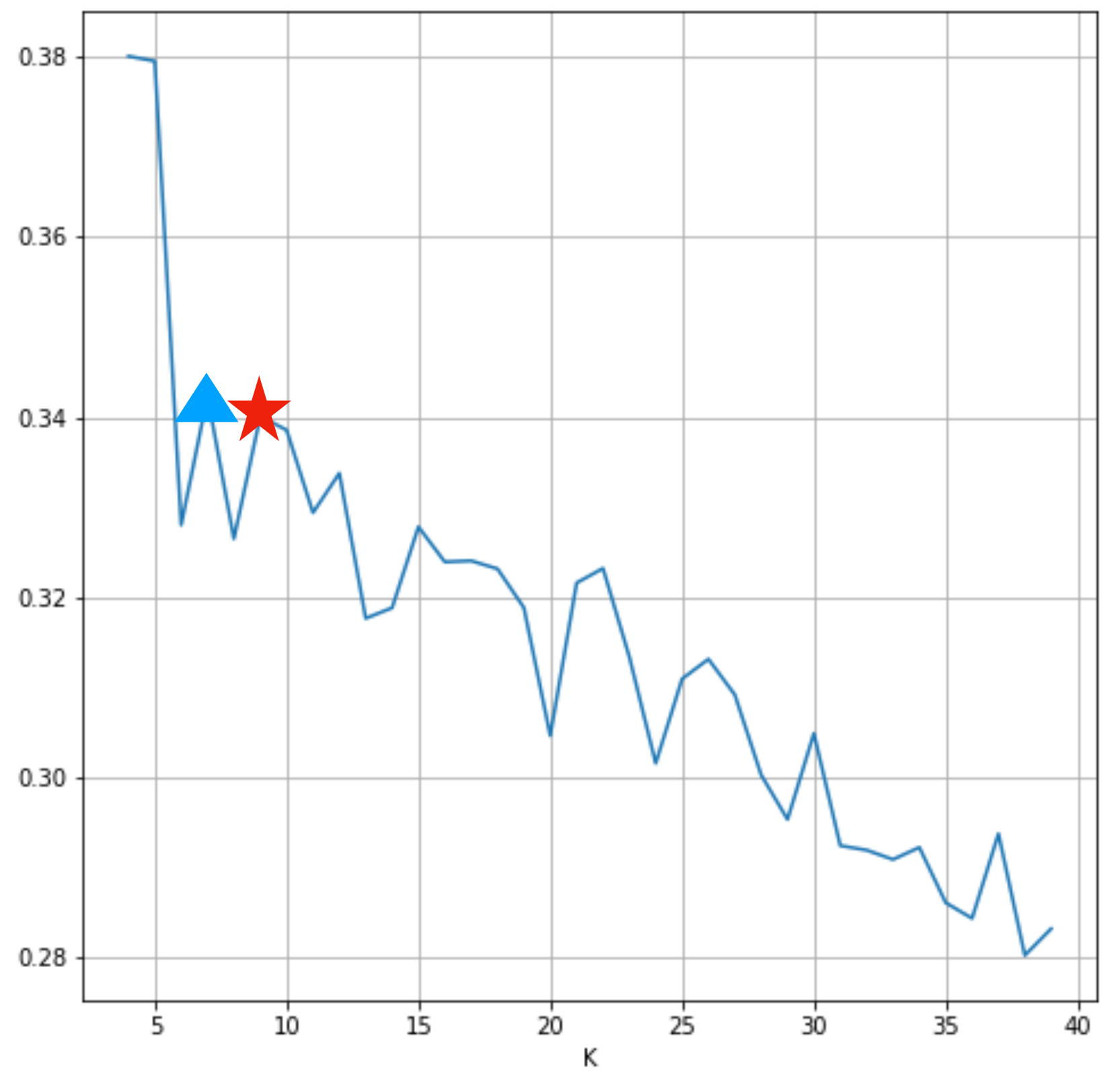
Analysis

6. Clustering— How

Elbow Method 2019



Silhouette Method 2019



Analysis

6. Clustering—Example

Randomly select a song released in 2018

不染 (電視劇《香蜜沉沉燼如霜》主題曲)

loudness	0.750248
acousticness	0.502049
danceability	0.348613
energy	0.45815
instrumentalness	0
liveness	0.0804355
speechiness	0.0199081
tempo	0.504866
valence	0.329091
key	0.636364

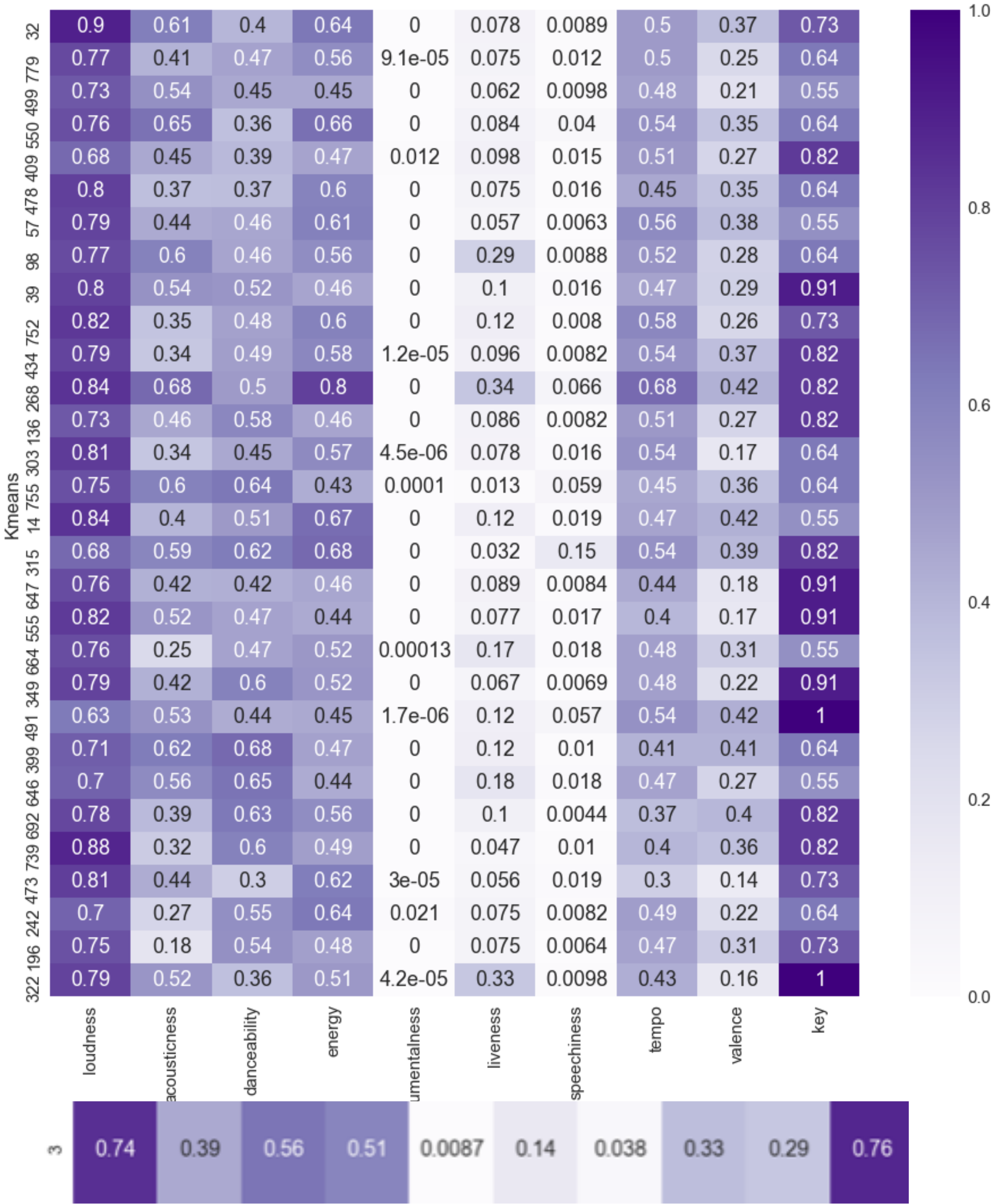


Analysis

6. Clustering—Example

推薦歌曲

32	將故事寫成我們
779	水深火熱
499	暫時愛著我
550	時間停止
409	不敢
478	清醒 (戲劇《淺情人不知》片尾曲)
57	我比從前想你了 - 電視劇《我們不能是朋友》片尾曲
98	寂寞考 - 大人中演唱會 LIVE
39	囚
752	給親愛的你
434	平行時空
268	別再叫我哥
136	初戀
303	失憶鎮
755	Hold On





12月5日好日子眾多歌手今天發片 TVBS新聞網

<https://news.tvbs.com.tw> > other ▼

2003年12月5日 - ... 發片，交出自己在這一年裡的最後一張成績單，當然今年也不例外，而且更巧的是大夥都不約而同地，選中今天12月5日這個黃道吉日發行新專輯。

[【聽歌懶人包】蔡依林、蔡健雅等7大女歌手同天發片網驚「神仙...](https://www.kkbox.com)

<https://www.kkbox.com> > column > showbiz-0-6936-1 ▼

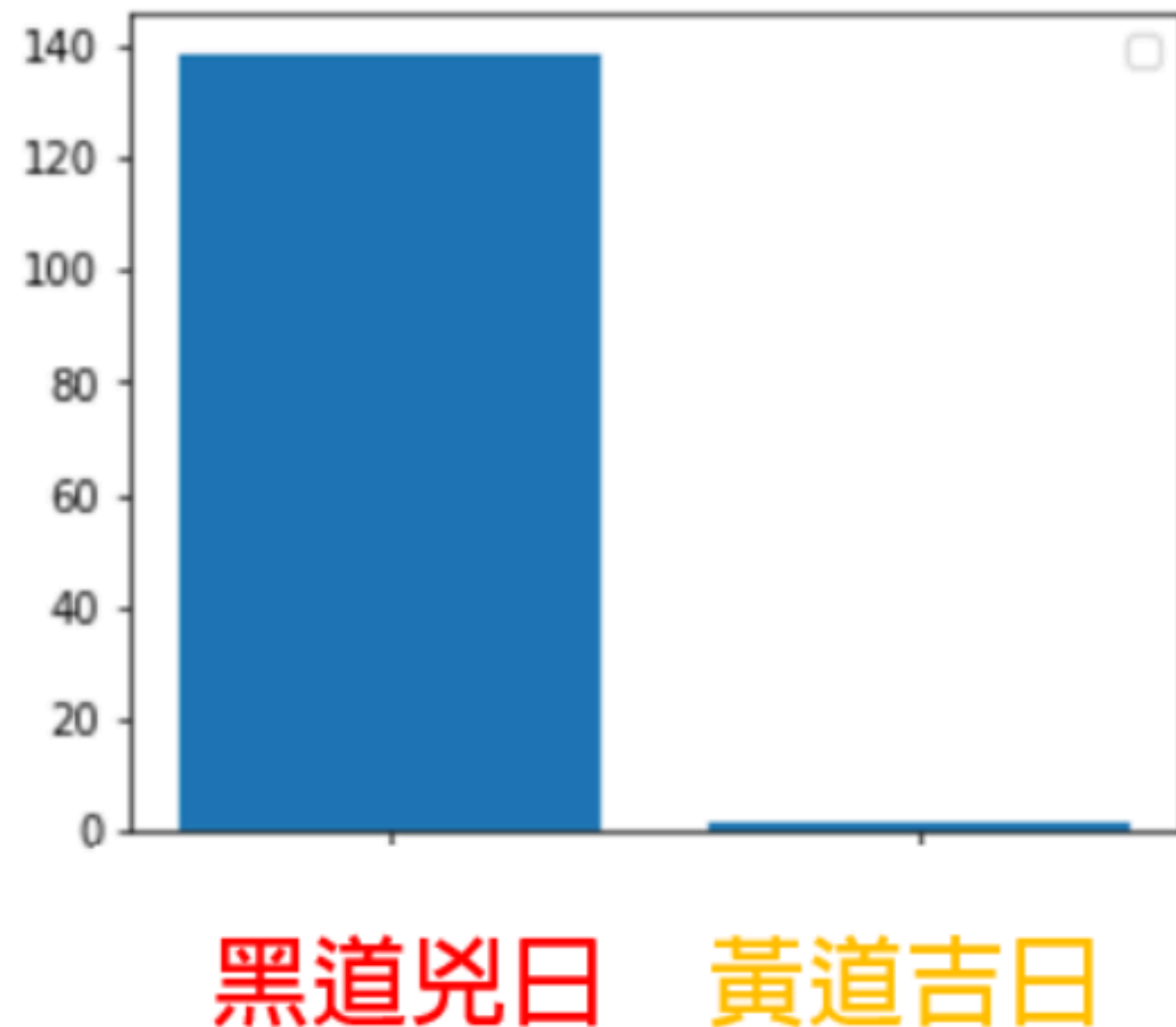
2018年12月21日 - 12/21 的今天究竟是什麼良辰吉日？... 包括蔡依林、林憶蓮、艾怡良、蔡健雅、Karencici、江美琪以及李千那七人的新歌、新專輯同時問世，讓歌...

挑定良辰吉日倪安東打歌又打片 | ETtoday星光雲 | ETtoday新聞雲

<https://star.ettoday.net> > ETtoday星光雲

2012年2月6日 - 陳漢典(左)助陣倪安東新專輯造勢。(圖/華研). 記者盧怡秀/台北報導. 「情歌王子」倪安東即將發行第二張個人專輯《Wake Up》，昨他率先出席預購...

Lunar Calendar



一般而言：

黃道吉日適合辦理民間喜事，黑道日適合軍事，戰略，策略等事宜

建日：諸事可為，但不宜動土修建。

除日：除舊迎新大吉。

滿日：祭祀吉，其它不吉。最忌嫁娶。

平日：萬事大吉。

定日：作灶台、宴會、定協議大吉，不宜打官司、醫療、出師。

執日：建房、種植吉，搬家、旅行、開市不吉。

破日：萬事不利，但可做拆毀之事。

危日：萬事不利。

收日：收藏、作灶台吉，不宜開業、舉事。

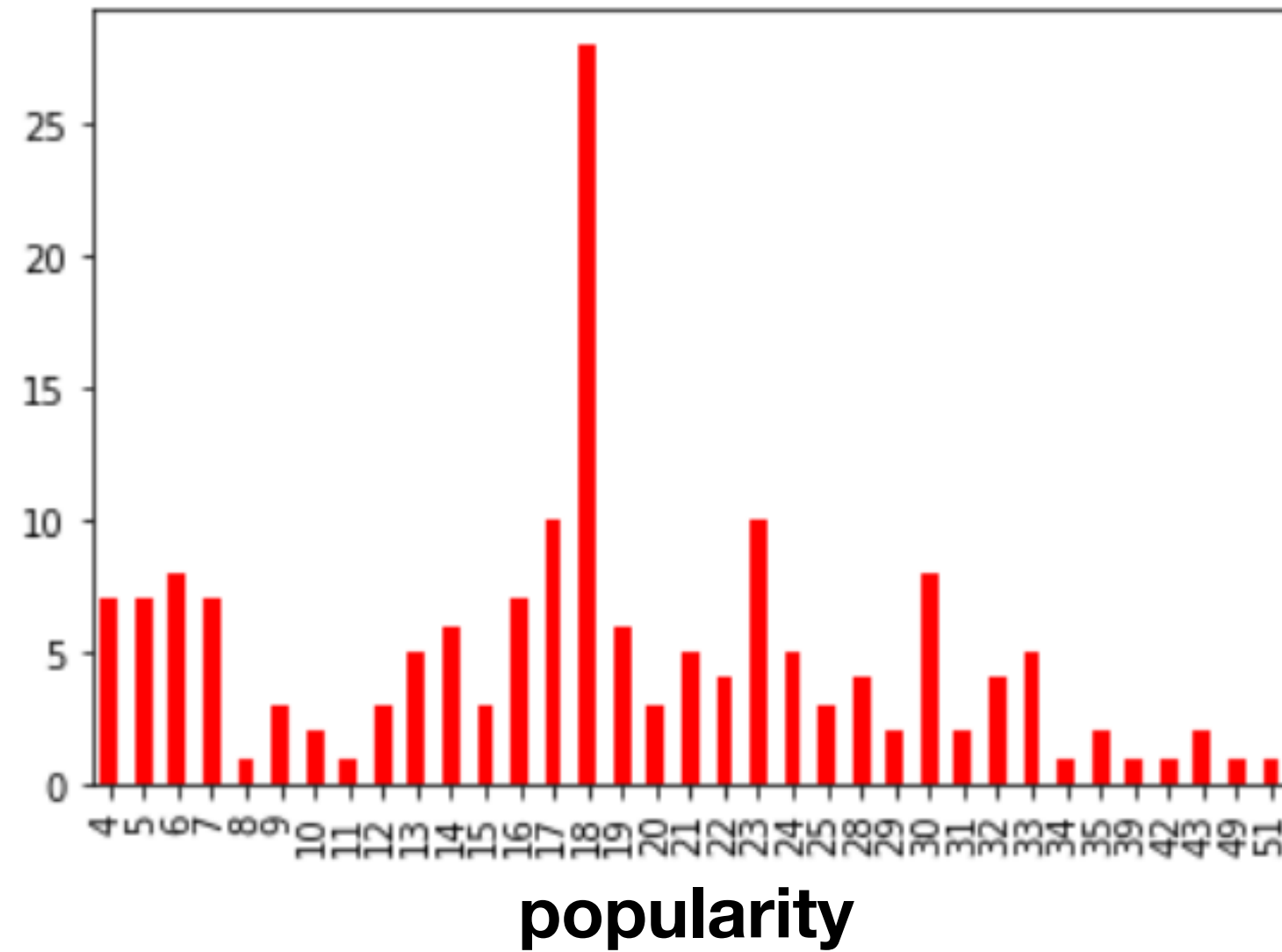
閉日：萬事皆凶，但宜修築。

Analysis

7. 突破迷信 歌曲發行日期大哉問

台灣地區歌曲

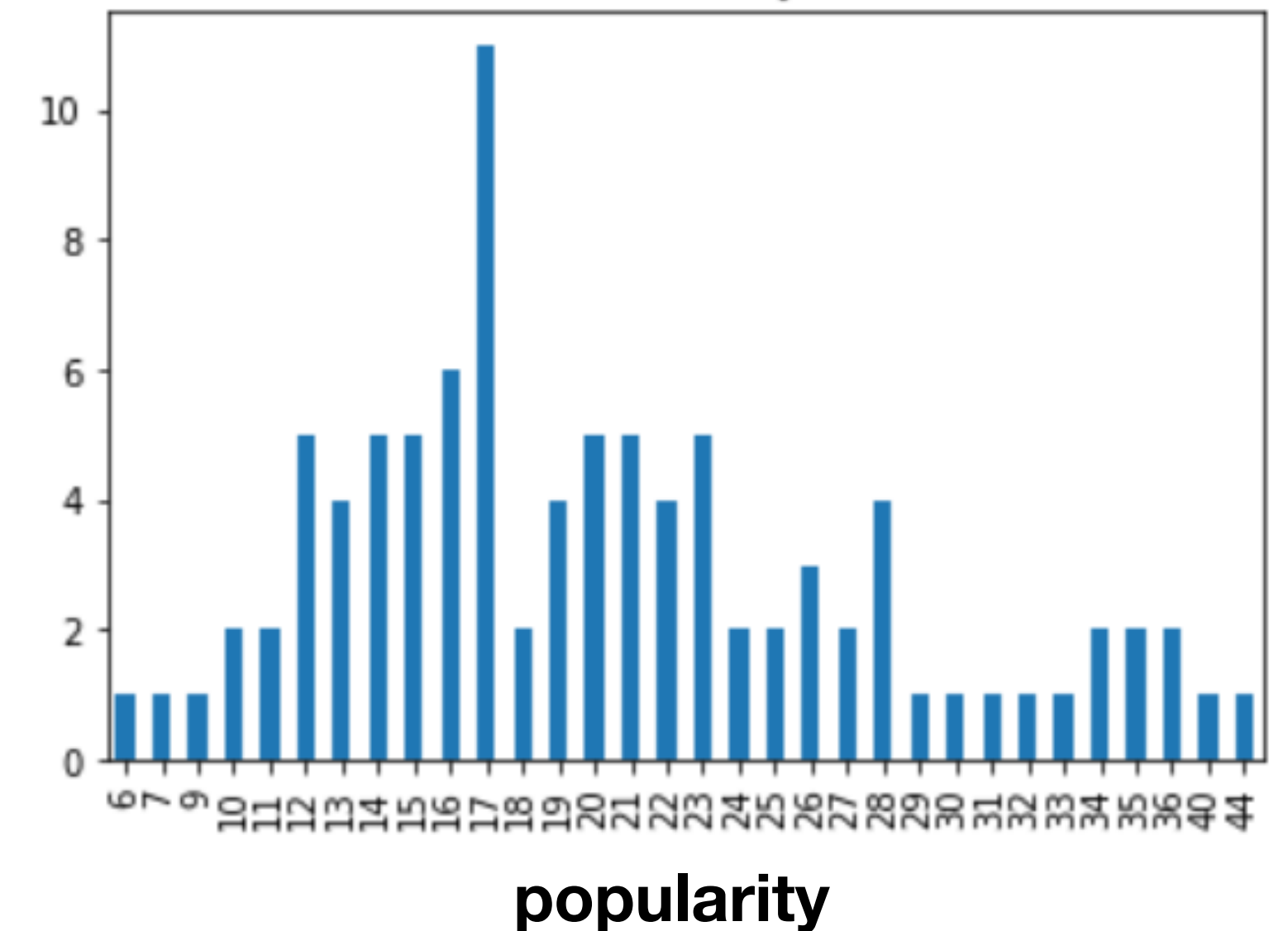
Distribution of 閉日 危日 破日



count	168.000000
mean	18.869048
std	9.551975
min	4.000000
25%	13.000000
50%	18.000000
75%	23.250000
max	51.000000

count	94.000000
mean	20.372340
std	7.632675
min	6.000000
25%	15.000000
50%	19.000000
75%	24.750000
max	44.000000

Distribution of 收日 執日



Conclusion

1. 各個地區音樂類型差異很大，即使是地緣關係相近，風格依舊迥異。
2. 透過Regression，發現歌曲長度、歌手姓名會影響歌曲是否popular，而歌曲的key、instrumentalness較無相關。
3. 透過Clustering(K-Means)，我們可以得知華語地區每年度流行音樂的特徵分佈，進而做到推薦歌曲的功能。
4. 台灣藝人偏好在「黑道凶日」發歌，但發行日期並不會影響其歡迎程度。

Future Work

Cont'd of clustering

