111356032 | 王楚雲 | 資管碩一

## 第一題

1. Meaning and range

| Audio feature | Meaning | Range |
|---|---|---|
| Danceability | Danceability describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity. | Float: 0.0 to 1.0<br>A value of 0.0 is least danceable and 1.0 is most danceable. |
| Energy | Energy represents a perceptual measure of **intensity and activity.** Typically, energetic tracks feel fast, loud, and noisy. For example, death metal has high energy, while a Bach prelude scores low on the scale. Perceptual features contributing to this attribute include dynamic range, perceived loudness, timbre, onset rate, and general entropy. | Float: 0.0 to 1.0 |
| Key(調性) | The key the track is in. Integers map to pitches using standard Pitch Class notation. E.g. 0 = C, 1 = C♯/D♭, 2 = D, and so on. If no key was detected, the value is -1. | Integer: >= -1 & <= 11 |
| Loudness | The **overall loudness of a track in decibels (dB)**. Loudness values are **averaged across the entire track** and are useful for comparing relative loudness of tracks. Loudness is the quality of a sound that is the primary psychological correlate of physical strength (amplitude). Values typically range between -60 and 0 db. | Float: 0.0 to -60; |
| Mode(大/小調) | Mode indicates the **modality (major or minor)** of a track, the type of scale from which its melodic content | Integer: 0 or 1 |

| | | |
|---|---|---|
| | is derived. Major is represented by 1 and minor is 0. | |
| Speechiness | Speechiness detects the **presence of spoken word**s in a track. The more exclusively speech-like the recording (e.g. talk show, audio book, poetry), the closer to 1.0 the attribute value. | Float: 0.0 to 1.0;<br>Values above 0.66 describe tracks that are probably made entirely of spoken words.<br>Values between 0.33 and 0.66 describe tracks that may contain both music and speech, either in sections or layered, including such cases as rap music.<br>Values below 0.33 most likely represent music and other non-speech-like tracks. |
| Acousticness | A confidence measure of whether the track is acoustic. | Float: 0.0 to 1.0;<br>1.0 represents high confidence the track is acoustic. |
| Instrumentalness (無人聲) | Predicts whether a track contains **no vocals**. "Ooh" and "aah" sounds are treated as instrumental in this context. Rap or spoken word tracks are clearly "vocal". | Float: 0.0 to 1.0;<br>The closer the instrumentalness value is to **1.0**, the greater likelihood the track contains **no vocal** content. Values above 0.5 are intended to represent instrumental tracks, but confidence is higher as the value approaches 1.0. |
| Liveness | Detects the **presence of an audience** in the recording. Higher liveness values represent an increased probability that the track was | Float: 0.0 to 1.0<br>A value above 0.8 provides strong likelihood that the track |

| | | |
|---|---|---|
| | performed live. | is live. |
| Valence | A measure from 0.0 to 1.0 describing the **musical positiveness** conveyed by a track. Tracks with high valence sound more positive (e.g. happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g. sad, depressed, angry). | Float: 0.0 to 1.0 |
| Tempo | The overall estimated tempo of a track in beats per minute (BPM). In musical terminology, tempo is the speed or pace of a given piece and derives directly from the average beat duration. | Float: 57.967 to 220.29 |
| Time Signature(拍號) | An estimated time signature. The time signature (meter) is a notational convention to specify **how many beats are in each bar** (or measure). The time signature ranges from 3 to 7 indicating time signatures of "3/4", to "7/4". | Integer: 3 to 7 |

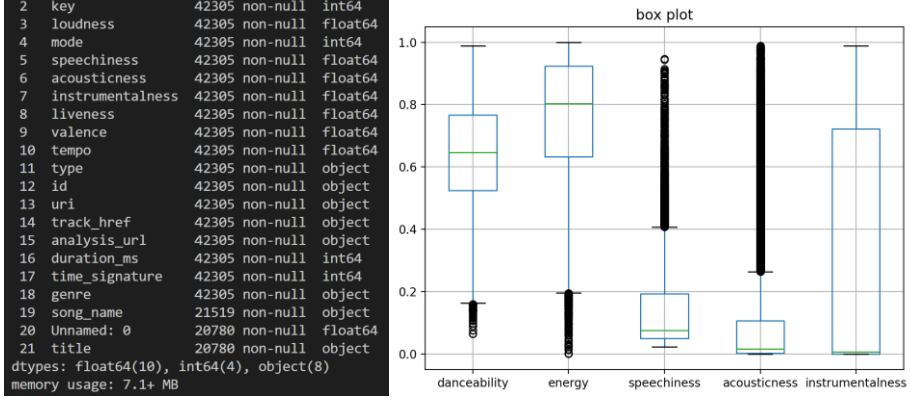
2. missing value and noise

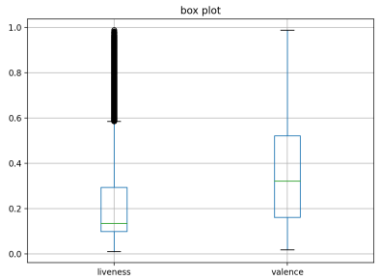
| Audio feature | Missing value | Noise |
|---|---|---|
| Danceability | X | Y |
| Energy | X | Y |
| Key | X | X |
| Loudness | X | Y |
| Mode | X | X |
| Speechiness | X | Y |
| Acousticness | X | Y |
| Instrumentalness | X | X |
| Liveness | X | Y |
| Valence | X | X |
| Tempo | X | Y |
| Time Signature | X | X |

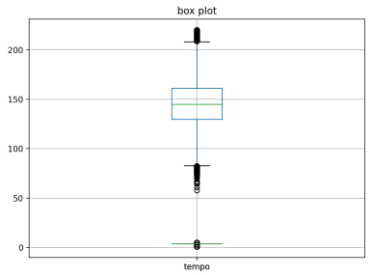
- Missing value: 根據下圖執行結果，總共有 42305 筆資料，所有 audio feature 欄位的非空值皆有 42305 筆，代表沒有空值

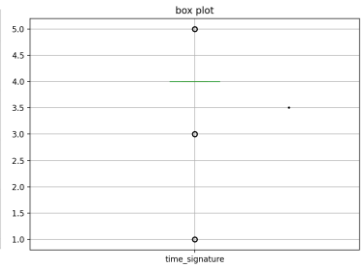
- Noise: 利用 box plot(下圖)，發現除了 key, mode, Instrumentalness, valence 以及 time signature，其他欄位皆有值大於第三四分位距 + 1.5 倍四分位距 (Q3 + 1.5IQR) 或小於 第一四分位距 - 1.5 倍四分位距(Q1 - 1.5IQR)，這些超過的值極為 noise

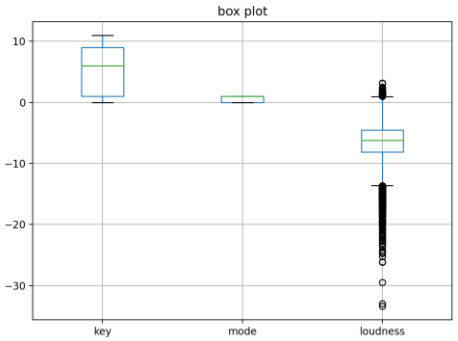
```
RangeIndex: 42305 entries, 0 to 42304
Data columns (total 22 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   danceability      42305 non-null  float64
 1   energy            42305 non-null  float64
 2   key               42305 non-null  int64
 3   loudness          42305 non-null  float64
 4   mode              42305 non-null  int64
 5   speechiness       42305 non-null  float64
 6   acousticness      42305 non-null  float64
 7   instrumentalness  42305 non-null  float64
 8   liveness          42305 non-null  float64
 9   valence           42305 non-null  float64
 10  tempo             42305 non-null  float64
 11  type              42305 non-null  object
 12  id                42305 non-null  object
 13  uri               42305 non-null  object
 14  track_href        42305 non-null  object
 15  analysis_url      42305 non-null  object
 16  duration_ms       42305 non-null  int64
 17  time_signature    42305 non-null  int64
 18  genre             42305 non-null  object
 19  song_name         21519 non-null  object
 20  Unnamed: 0        20780 non-null  float64
 21  title             20780 non-null  object
dtypes: float64(10), int64(4), object(8)
memory usage: 7.1+ MB
```

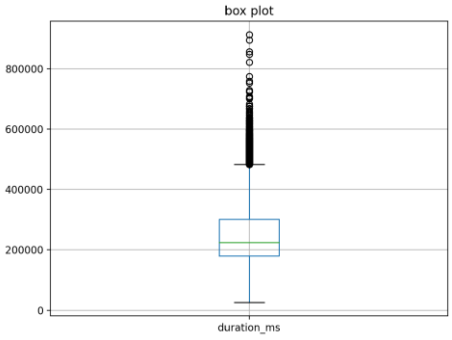


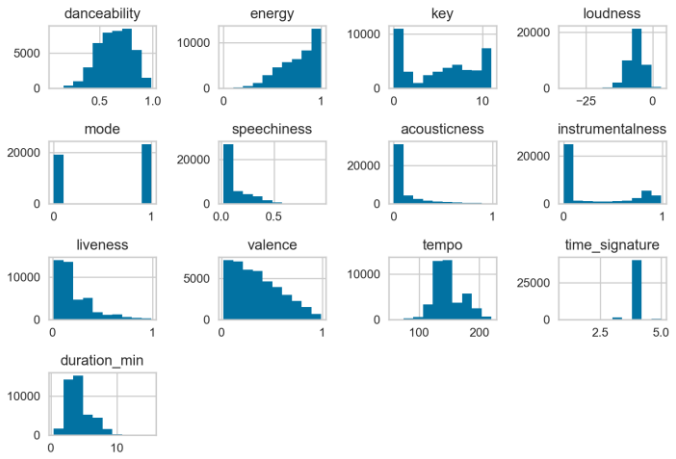




## 第二題 資料前處理

1. 去除不必要欄位

   甲、經過 .value_counts() 或 .unique() 觀察，發現 id, song name, uri, track_href, analysis_url, Unnamed: 0, title 欄位只是為了辨識各個歌曲的名字、type 欄位全部的值都是 audio feature，皆對於歌曲的分群沒有幫助，因此先去除，最後留下 12 種 audio feature

2. 資料標準化



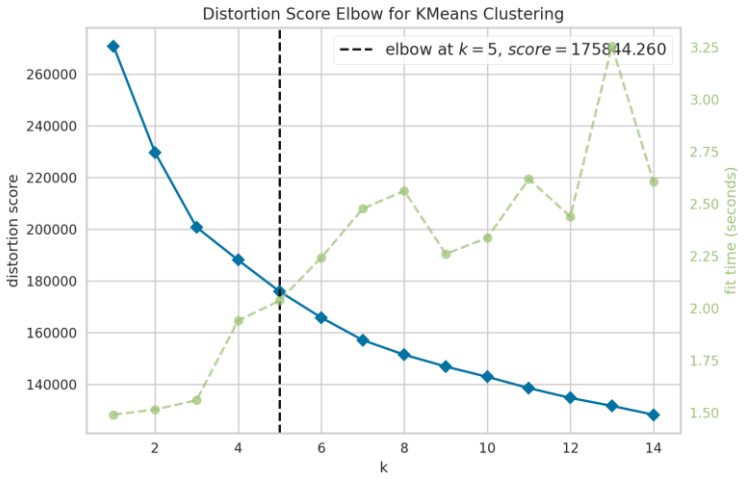

甲、在 describe data 的時候，發現每個 columns 之間的平均數、變異數相差很大，因此利用 StandardScalar() 把資料標準化，避免資料因為值域的不同而產生干擾

3. Null value：經過觀察 (df.isnull().sum()) 在各個欄位都是 0，因此不需要處理空值問題
4. Duration_ms 轉換成分鐘，經過觀察，如果用毫秒為單位，每筆資料的時長就會相差很多，很難找出相似性，也不符合人聽歌的單位習慣(我們聽歌都是以分鐘和秒為單位)，因此轉換為分鐘

## 第三題 kmeans

1. Elbow diagram

甲、如下圖 elbow diagram，發現 sse 最低的點是當 k=5 的時候，因此在後續跑分群時使用 k=5 去跑

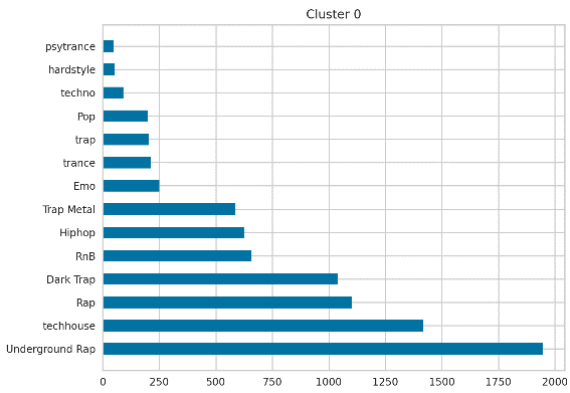


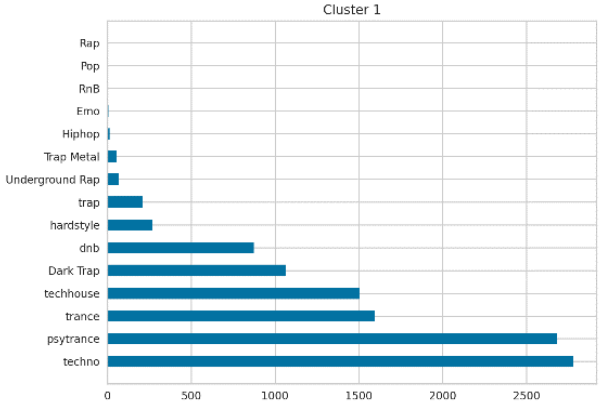
2. 分群結果：參數、群數、每群的數量

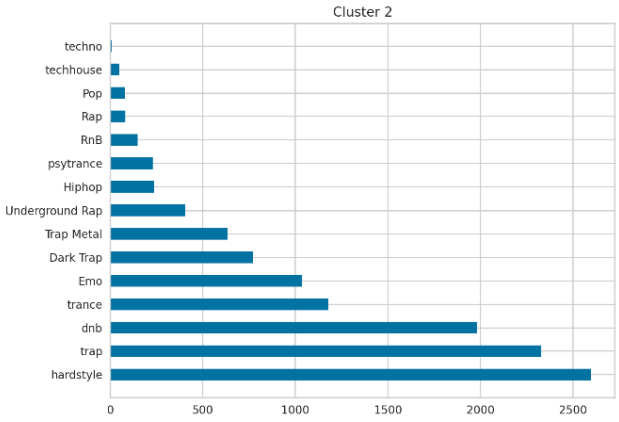
甲、K=5

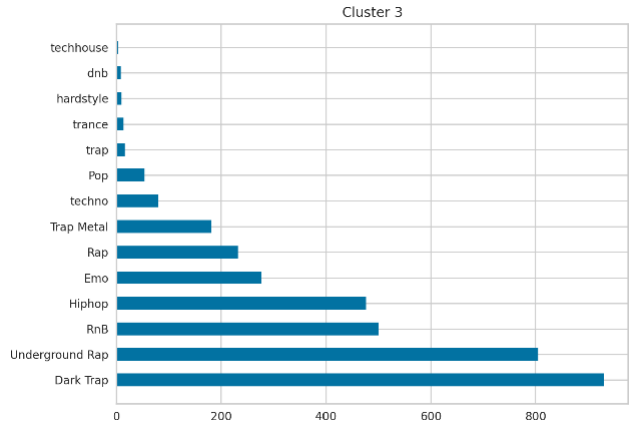
```
2    11765
1    11122
0     8438
4     7394
3     3586
Name: cluster, dtype: int64
```

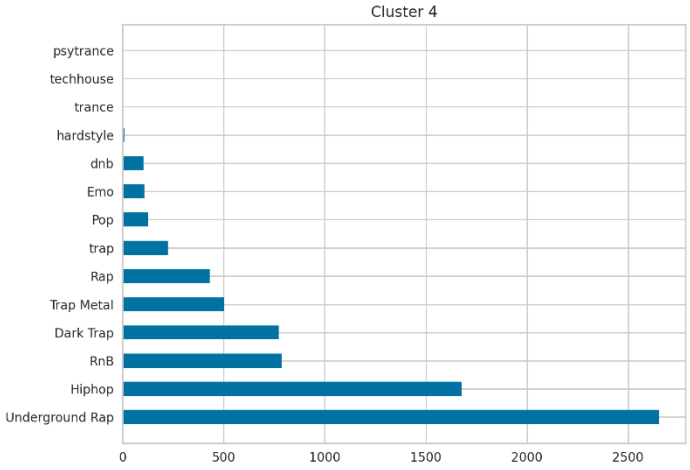
乙、各群主要的音樂類別



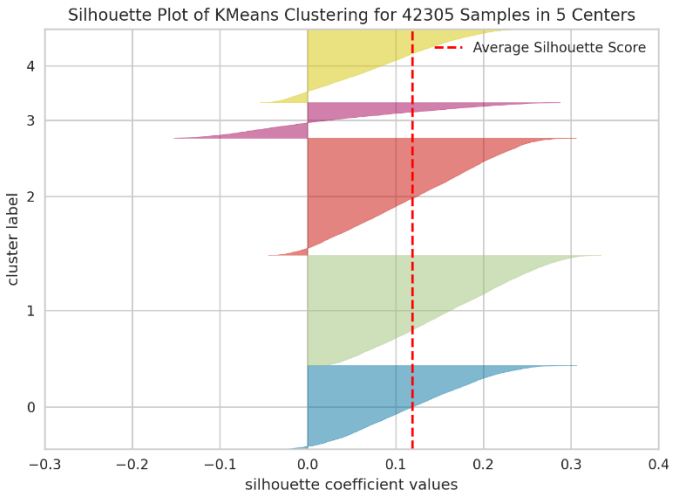



| Cluster | genre |
|---|---|
| 0 | techhouse, rap, trap metal, pop |
| 1 | Techno, psystance, trance |
| 2 | Hardstyle, trap, dnb, emo |
| 3 | Dark trap |
| 4 | Ungrounded rap, hiphop, rnb |

3. Silhouette coefficient: 0.11890265159324606




4. 分群結果評估與 confusion matrix

甲、正確答案設定：因為不同分群演算法會把具有相同性質的資料分到的群不固定，加上相同歌曲類別的歌也不一定會被分在同一群，因此透過該歌曲類別(genre)在哪個 cluster 歌曲數最多，該歌曲類別就屬於該 cluster 作為正確答案，盡可能使不同分群演算法的答案保持一致 e.g. ungrounded rap, hiphop, dark trap 通常會在一群、psytrance, trance, techno 通常會在一群等，下列演算法均採此想法設定正確答案

乙、Rand Index: 0.758936600359646

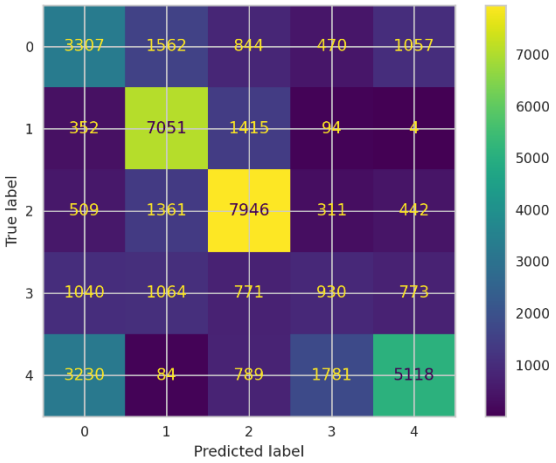
丙、Normalized Mutual Information: 0.2927095575696934

丁、Adjusted Mutual Information: 0.2926235284631347

戊、V Measure: 0.29270955756969347

己、Fowlkes-Mallows Scores: 0.4514631765268758

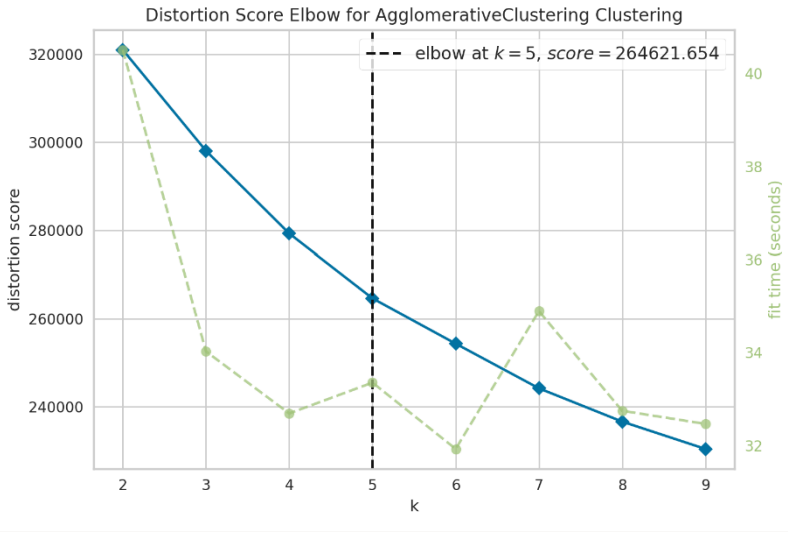
庚、Confusion matrix




## 第四題 Hierarchical Clustering

1. Elbow diagram

甲、如下圖 elbow diagram，發現 sse 最低的點是當 k=5 的時候，因此在後續跑分群時使用 k=5 去跑



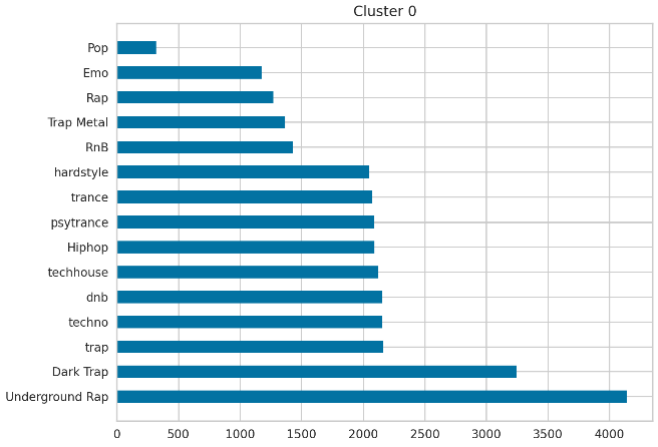

2. 分群結果：參數、群數、每群的數量

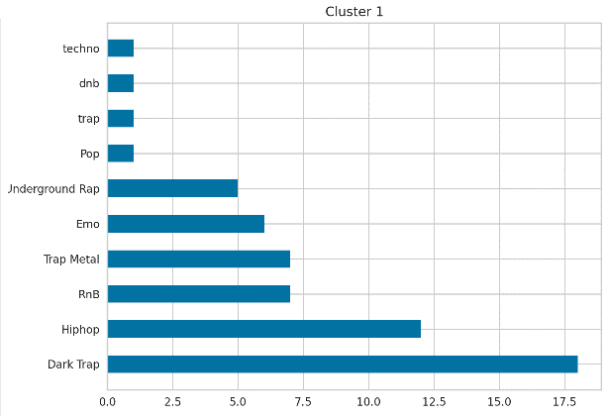
甲、K=5

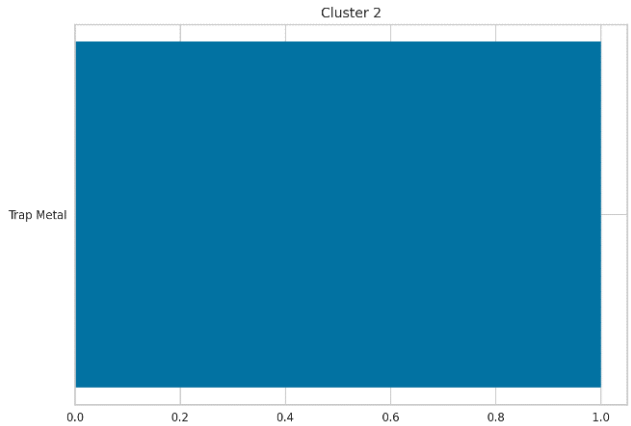
| | |
|---|---|
| 0 | 29843 |
| 3 | 96 |
| 1 | 59 |
| 2 | 1 |
| 4 | 1 |

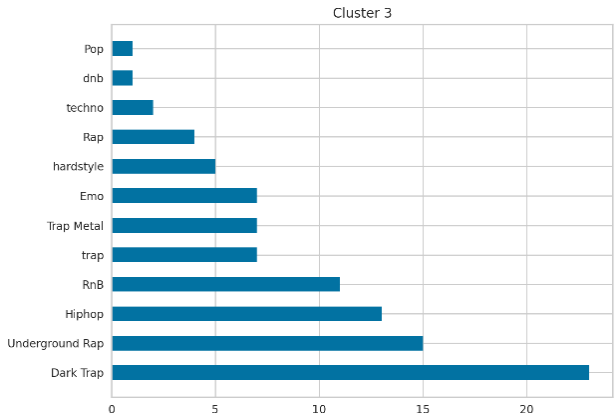
乙、主要的音樂類別

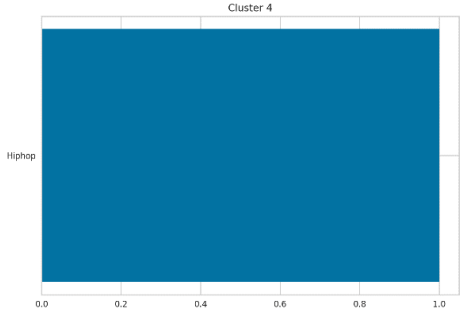
| cluster | Song category |
|---|---|
| 0 | Ungrounded rap, dark trap, trap, techno, dnb, techhouse, psytrance, trance, hardstyle, rap, pop |
| 1 | Trap metal, emo |
| 2 | x |
| 3 | Hiphop, rnb |
| 4 | x |



丙、Silhoutte Coefficient: 0.4127617448798479

3. 分群評估與 confusion matrix

甲、Rand Index: 0.5682127248686067

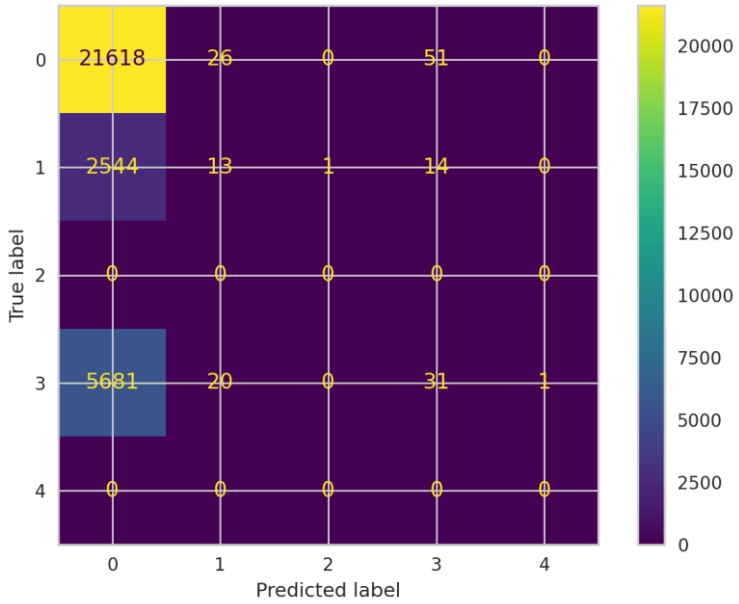
乙、Normalized Mutual Information: 0.0019263572080161437

丙、Adjusted Mutual Information: 0.0016285102060302947

丁、V Measure: 0.0019263572080161434

戊、Fowlkes-Mallows Scores: 0.7507996397742878

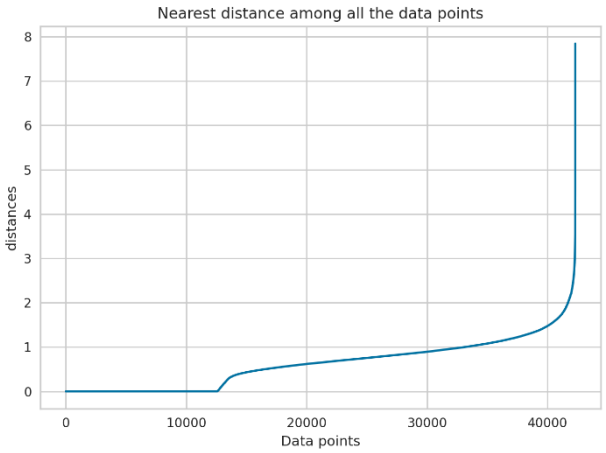
己、Confusion matrix




## 第五題 DBSCAN

- Elbow diagram

甲、因為 DBSCAN 不同於其他演算法是要設定半徑與半徑內的最少點數，因此 DBSCAN 的 elbow diagram 是利用資料點之間的距離去設定參數，由下圖所示，epsilon(半徑) 設為 1.8 較為合適，而 minimum points 依照過往文獻(Sander et al., 1998)，設為 2 * dimension 數，也就是 24



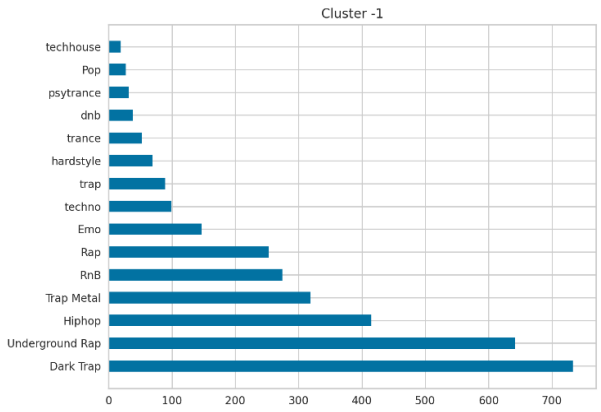

- 分群結果：參數、群數、每群的數量

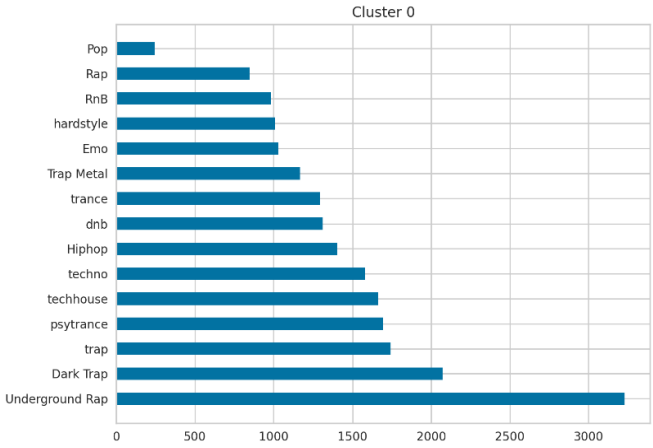
Epsilon= 1.8, minPts=24, K=6

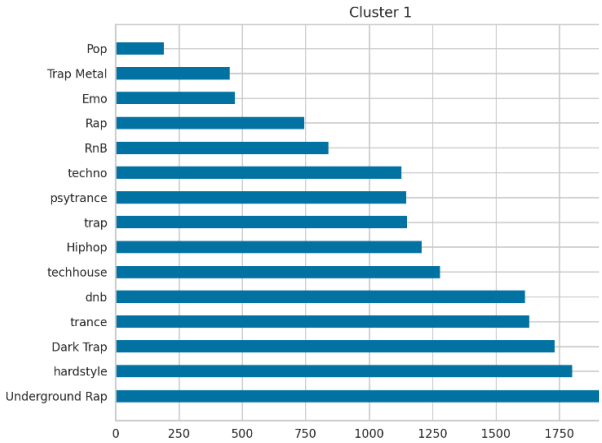
| | |
|---|---|
| 0 | 21272 |
| 1 | 17369 |
| -1 | 3209 |
| 3 | 190 |
| 4 | 183 |
| 2 | 82 |

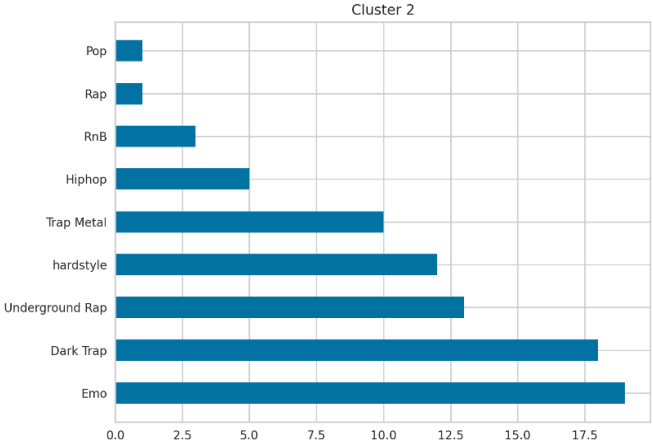
| Cluster | Song Category |
|---|---|
| -1 | hiphop, trap metal |
| 0 | Ungrounded rap, dark trap, trap, trap metal, rap, pop |
| 1 | Dnb, trance, hardstyle |
| 2 | emo |
| 3 | techno, psytrance, techhouse, |
| 4 | rnb |

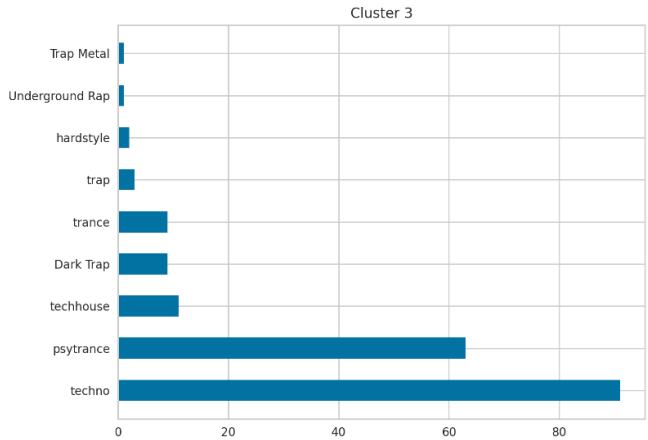
Silhoutte Coefficient: 0.0910426827494032

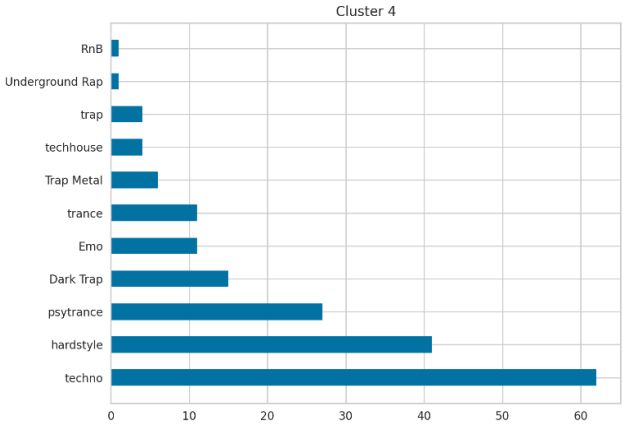



- 分群評估

  甲、Rand Index: 0.5417851927532233

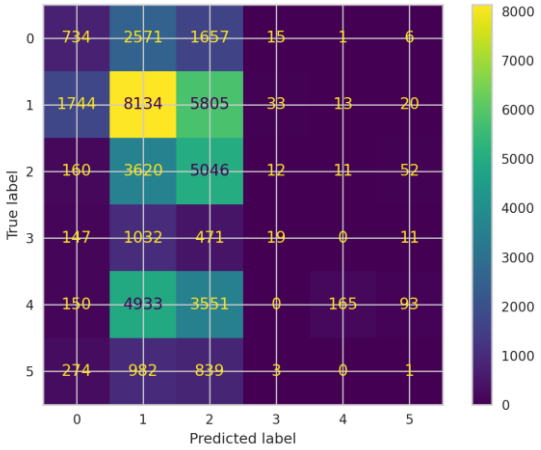
  乙、Normalized Mutual Information: 0.030942770868508062

  丙、Adjusted Mutual Information: 0.030712475893577888

  丁、V Measure: 0.030942770868508062

  戊、Fowlkes-Mallows Scores: 0.3306361302710796

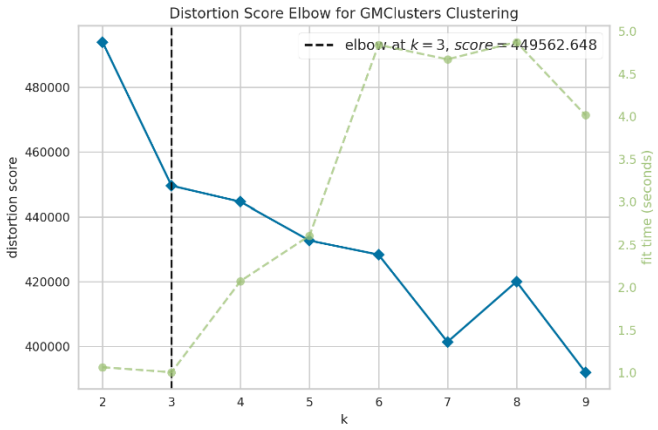
  己、Confusion matrix




## 第六題 GMM

1. Elbow diagram

甲、如下圖 elbow diagram，發現 sse 最低的點是當 k=3 的時候，因此在後續跑分群時使用 k=3 去跑




2. 分群結果：參數、群數、每群的數量

0 23625

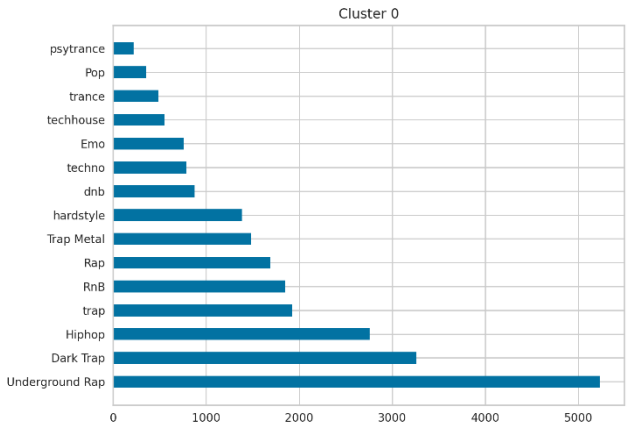
2 16801

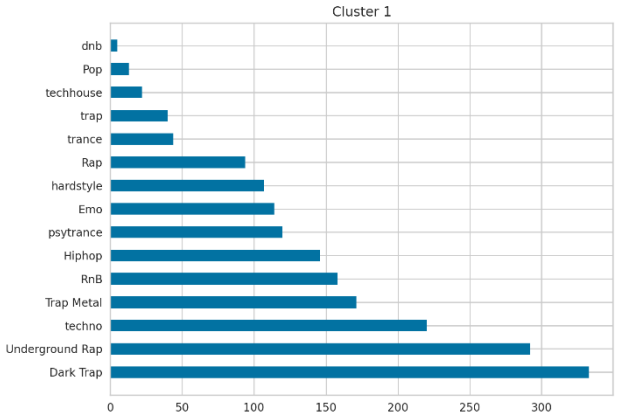
1 1879

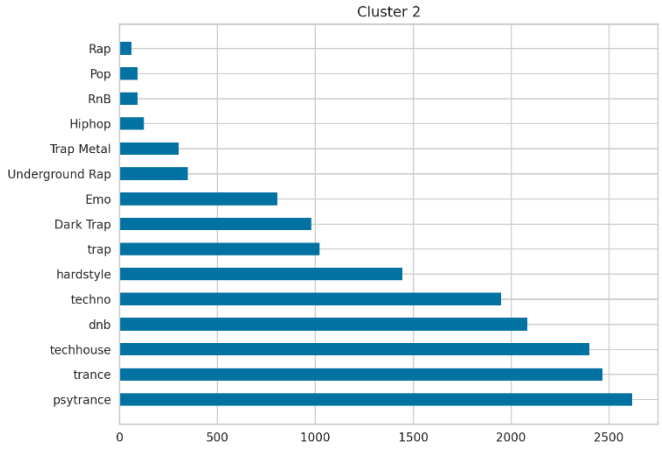
Silhoutte Coefficient: 0.0178239910712015

| Cluster | Song category |
|---|---|
| 0 | Underground rap, dark trap, hiphop, trap rnb, rap, pop |
| 1 | Techno, trap metal |

| | |
|---|---|
| 2 | Psytrance, trance, techhouse, dnb, hardstyle, emo |




3. 分群評估與 confusion matrix

甲、Rand Index: 0.640514526605207

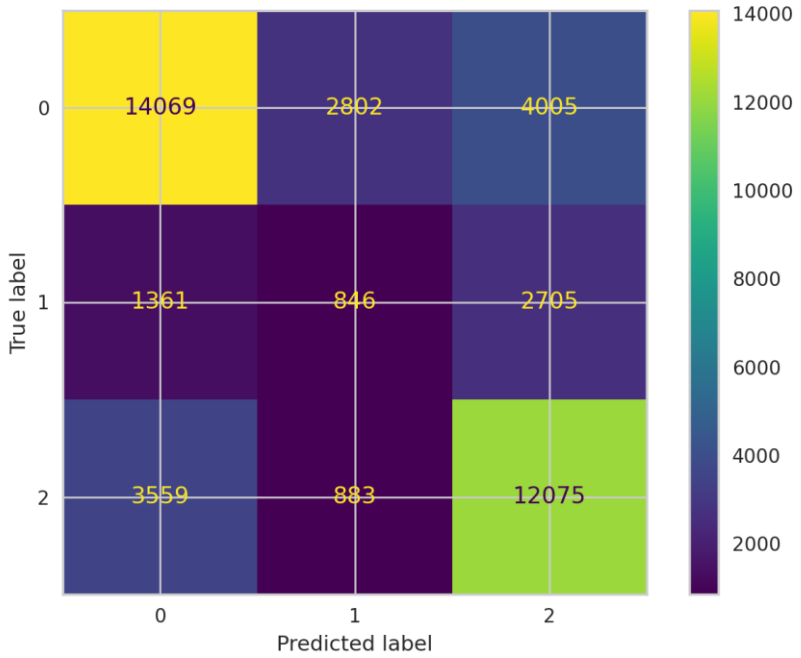
乙、Normalized Mutual Information: 0.18683918042391884

丙、Adjusted Mutual Information: 0.18679630862289792

丁、V Measure: 0.18683918042391887

戊、Fowlkes-Mallows Scores: 0.5944324376055541

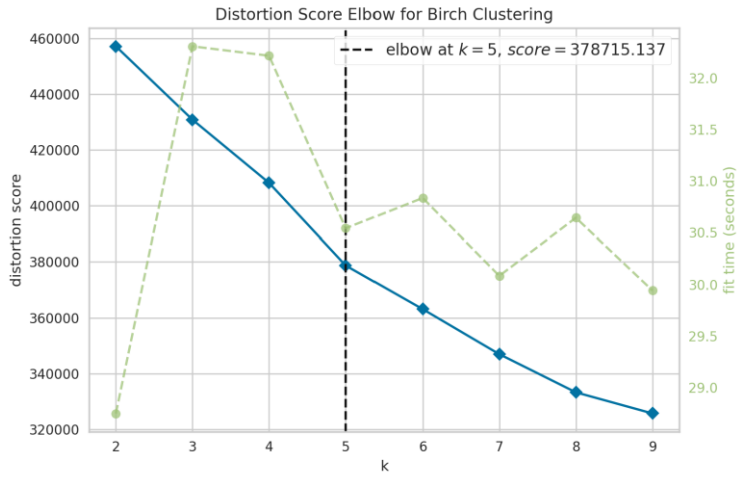
己、Confusion matrix




## 第七題 BIRCH

1. Elbow diagram

甲、如下圖 elbow diagram，發現 sse 最低的點是當 k=5 的時候，因此在後續跑分群時使用 k=5 去跑




2. 分群結果：參數、群數、每群的數量

K=5

0 16694

1 11737

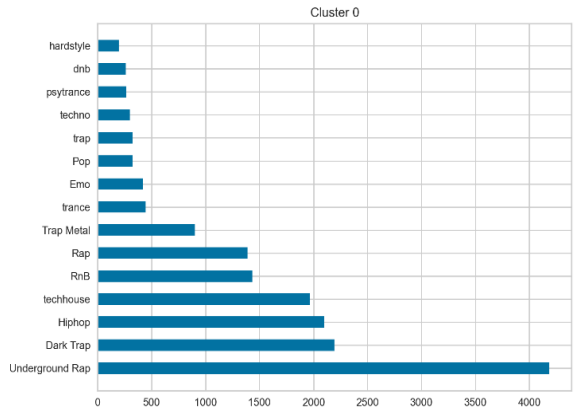
4 9525

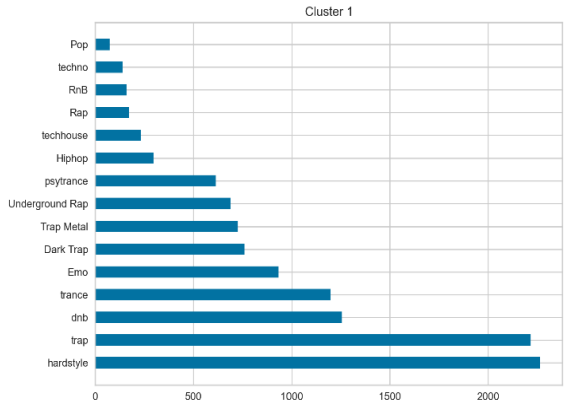
3 3128

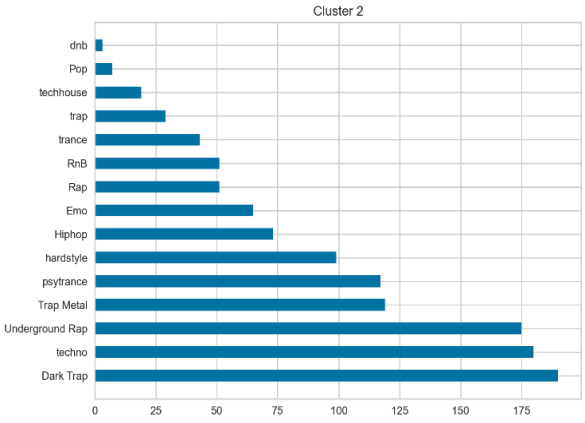
2 1221

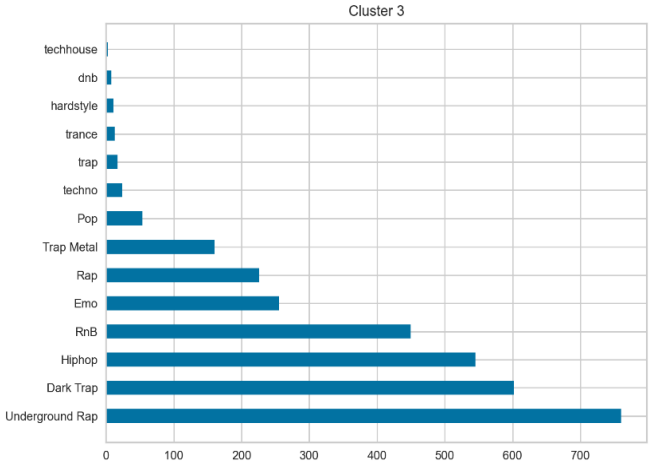
| Cluster | Song category |
|---|---|
| 0 | Ungrounded rap, dark trap, hiphop, techhouse, rap, pop |
| 1 | Hardstyle, trap, dnb, trance |
| 2 | Trap metal |
| 3 | Emo, rnb |
| 4 | Techno, psytrance |

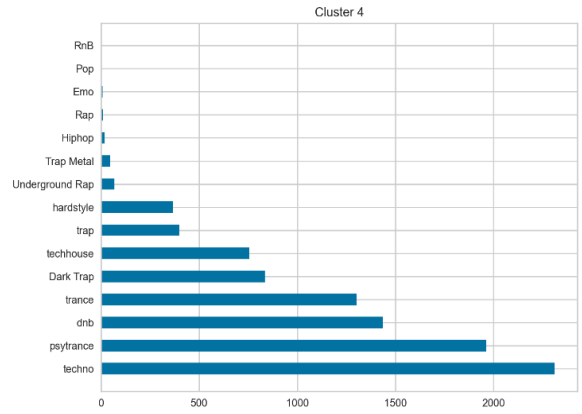
Silhoutte Coefficient: 0.0862082600376029




3. 分群評估與 confusion matrix

甲、Rand Index: 0.682155989007855

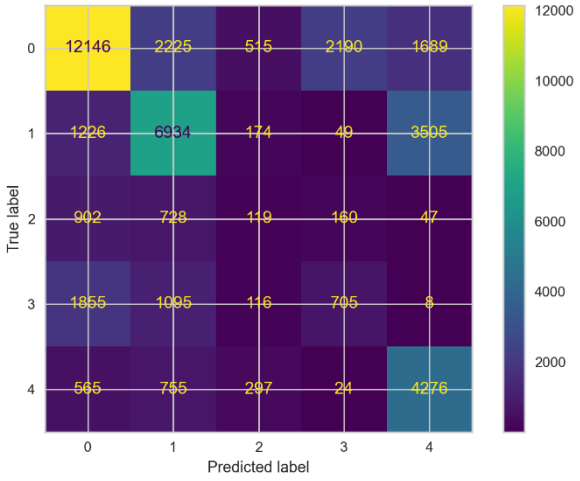
乙、Normalized Mutual Information: 0.22486051990481104

丙、Adjusted Mutual Information: 0.2247519989070999

丁、V Measure: 0.22486051990481104

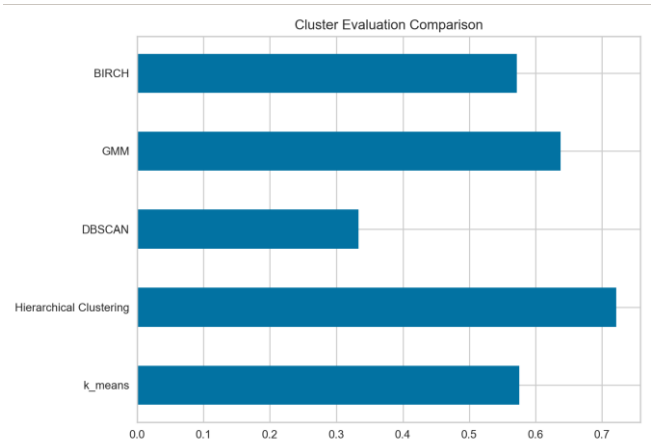
戊、Fowlkes-Mallows Scores: 0.4660091976340529

己、Confusion matrix




第八題 分群效果比較

- 利用 confusion matrix 來計算 5 個分群演算法的正確率




- 如圖所示，在歌曲資料分群上，hierarchical clustering 表現最佳，再來是 GMM，最後是 BIRCH

第九題 效率

- 透過建模、資料分群時間來比較效率高低

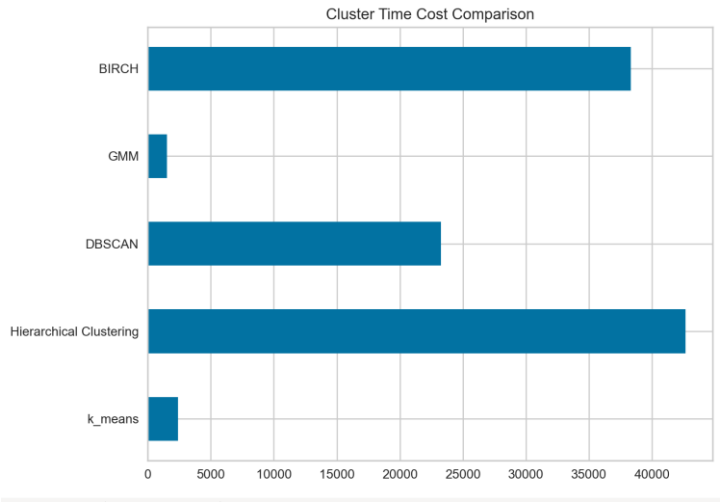
K-Means 2395.153522491455 ms

Hierarchical 42682.422399520874 ms

DBSCAN 23256.197929382324 ms

GMM 1520.604133605957 ms

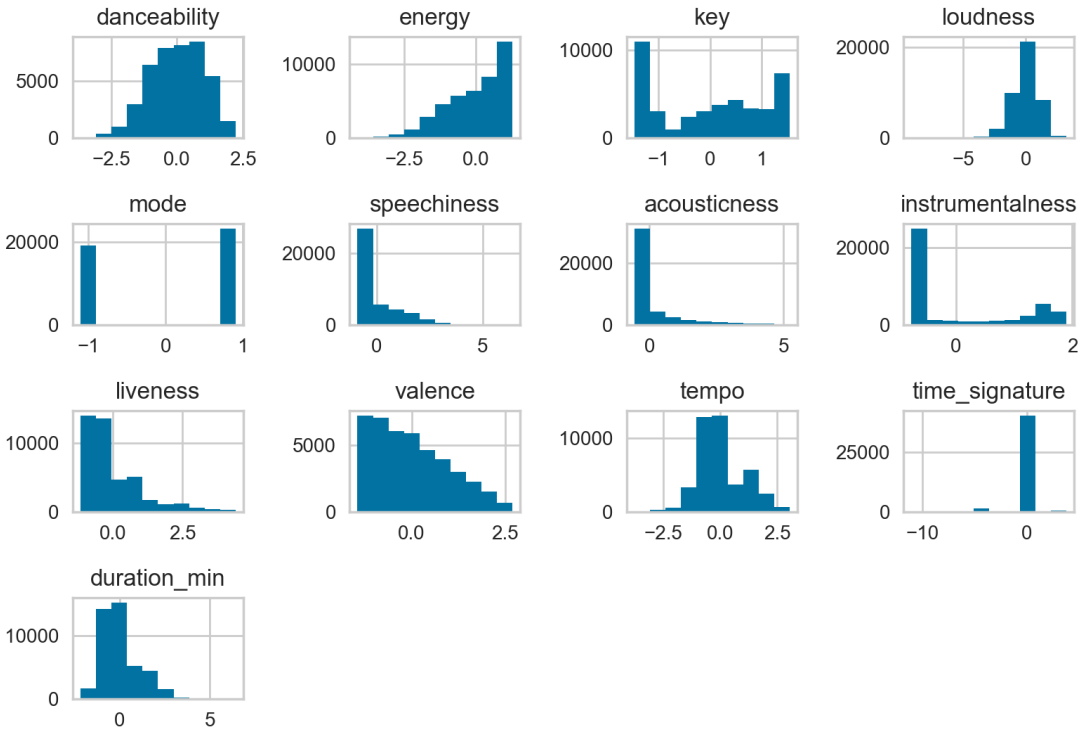
Birch 38306.37454986572 ms




- 如上圖，在分群效率方面 k_means 和 GMM 遠勝於其他分群演算法，而 hierarchical clustering 效率較差

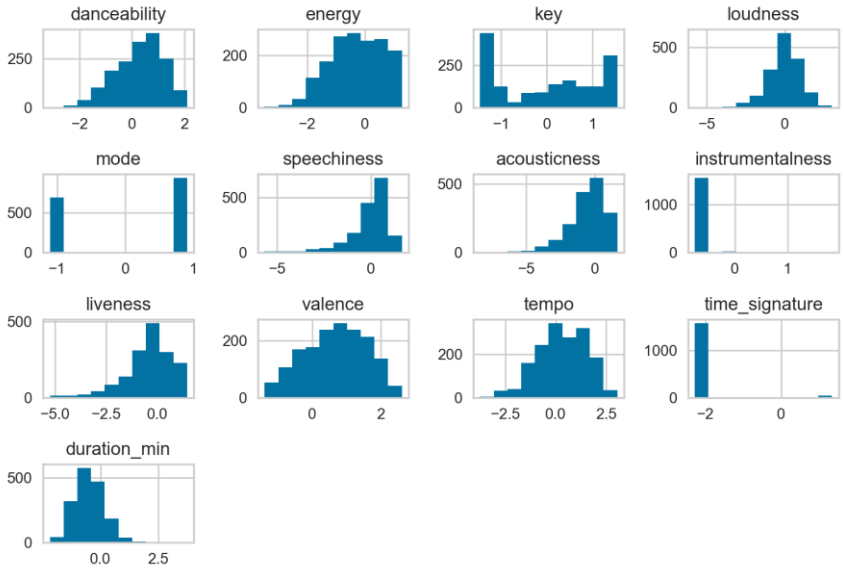
## 第十題 如何改善分群結果

1. 分析：在上述分群過程中，發現有些分群結果不同 cluster 有平分同一個 genre 的情況，例如在 k-means，ungrounded rap 平均出現在 cluster 0 和 cluster 4，代表原先的 22 個 feature 並不能很好地把不同性質的資料分到不同 cluster (separation 不佳)，推測原因可能是 feature 數量過多，因此這邊採用 PCA 進行 feature extraction。
2. 由於要使用 PCA 進行資料降維，其中有用到常態分布假設，因此需要先轉換 skewedness 過於嚴重的欄位，這邊使用 log transformation，但因為有些負的資料，因此先將資料進行平移再轉換。



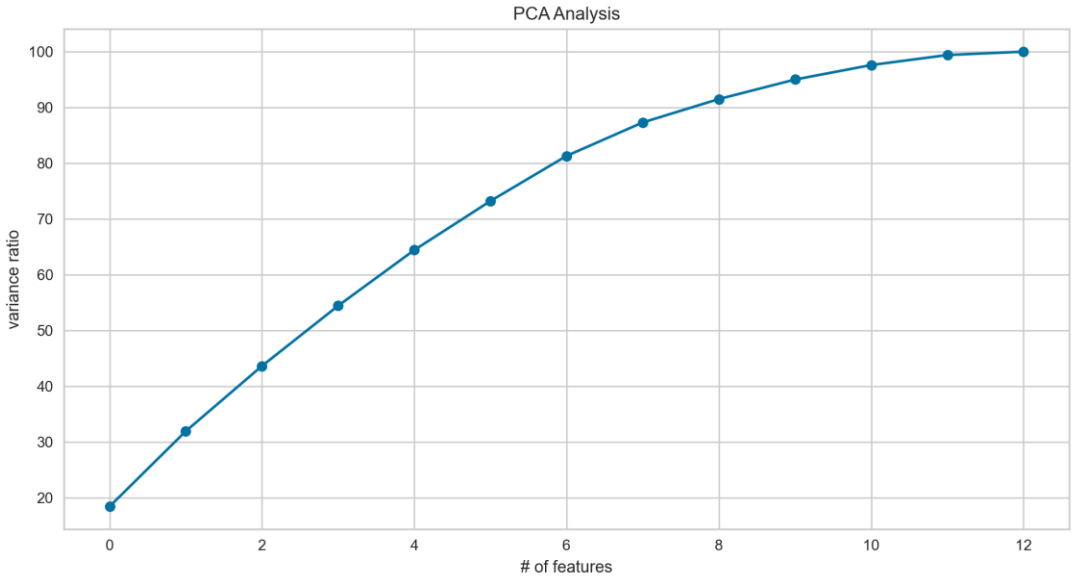

3. 由標準化後的各個 column 的資料分布(上圖)以及 skewedness 計算得知，acousticness、liveness、speechiness 以及 time_signature 的偏態大於 1 或小於 -1(下圖)，代表這些欄位偏態系數高，因此對這些欄位進行 log transformation 讓他們接近常態分配，結果如下下張圖。

```
acousticness          2.534878
liveness              1.782580
speechiness           1.673151
duration_min          0.951910
instrumentalness      0.752695
valence               0.549453
tempo                 0.478337
key                  -0.001146
mode                 -0.198831
danceability         -0.265484
loudness             -0.645511
energy               -0.738103
time_signature       -5.515843
```



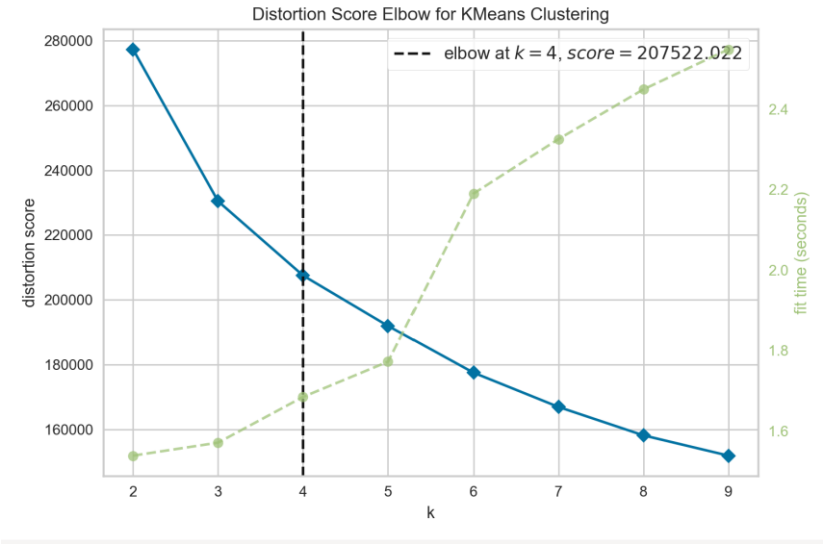

4. 經過 log transformation 後(上圖)，acousticness、liveness、speechiness 以及 time_signature 的資料更為接近常態
5. PCA 維度數

   甲、因為我們的目標是要找出最能夠解釋分群結果變異性的 feature，因此萃取出 6 個較能解釋變異性的 feature 數，也就是 6，我們發現 6 以上的 feature 數在解釋力的提升上開始有顯著下降。




6. 透過 k_means 看 performance 是否有顯著提升

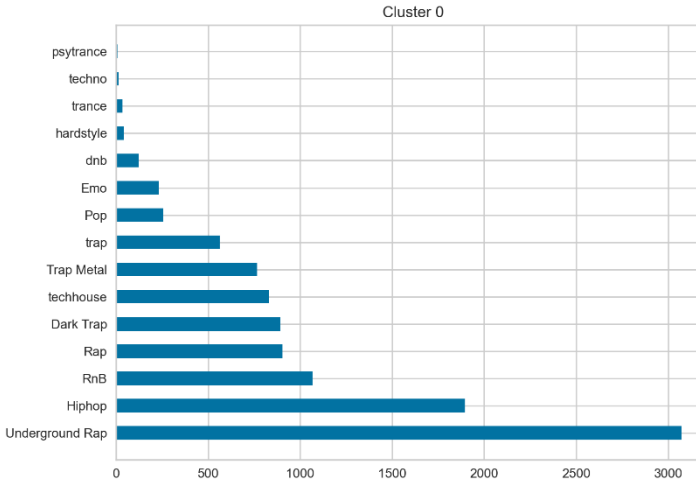
   甲、Elbow diagram 來決定群數，這邊發現 k=4 是肘點

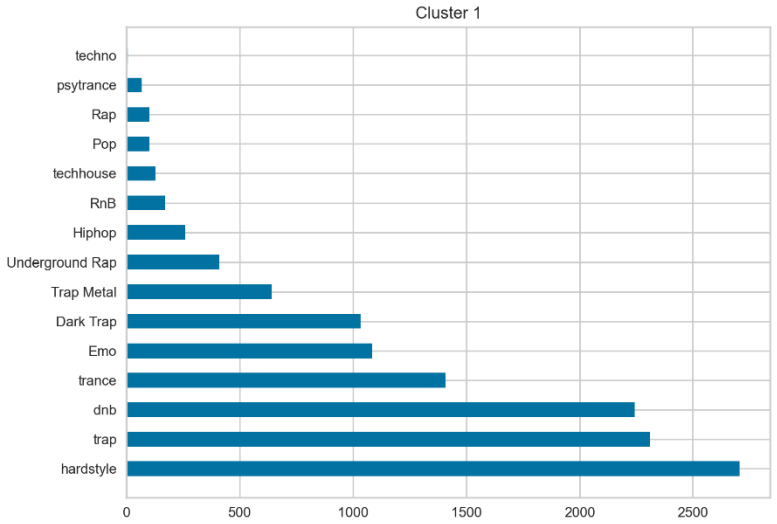


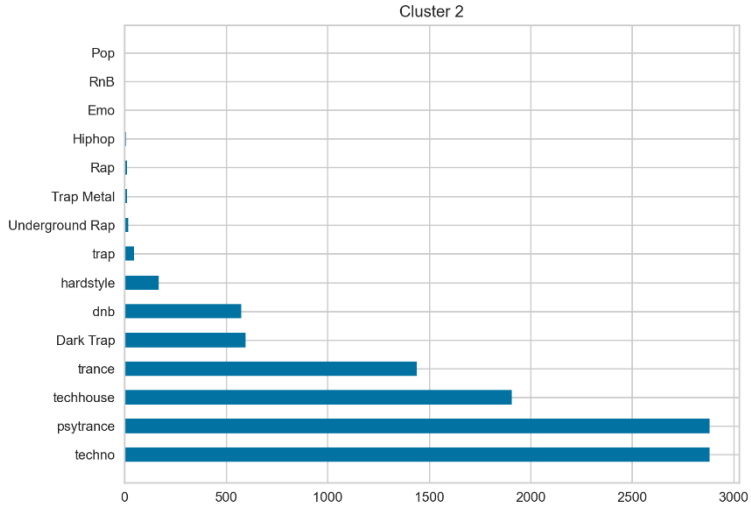

乙、K_means 分群結果：

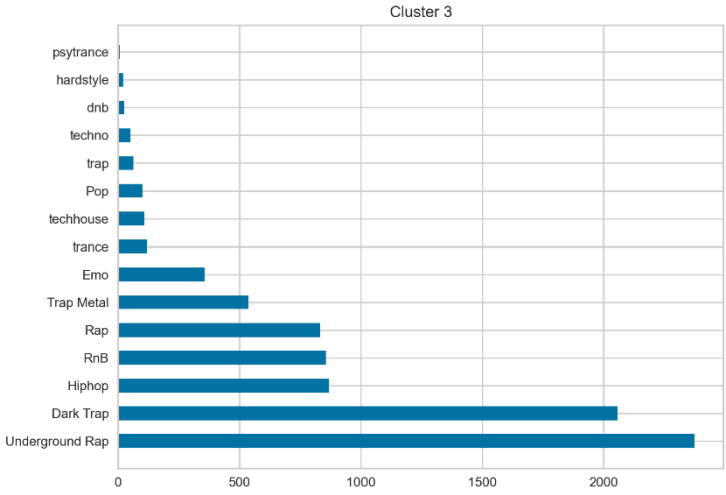
| | |
|---|---|
| 1 | 12673 |
| 0 | 10701 |
| 2 | 10550 |
| 3 | 8381 |

| Cluster | Song category |
|---|---|
| 0 | Ungrounded rap, hiphop, emo |
| 1 | Hardstyle, trap, dnb, emo |
| 2 | Techno, psytrance, techhouse, trance |
| 3 | Dark trap, rnb, rap, trap metal |




**丙、** 由上述分群結果發現，在 feature extraction 後的分群結果來看，**各個群所含的資料點較為平均**

丁、Silhoutte Coefficient: 0.18790573774598407

戊、分群評估

i. Rand Index: 0.7530209892465581

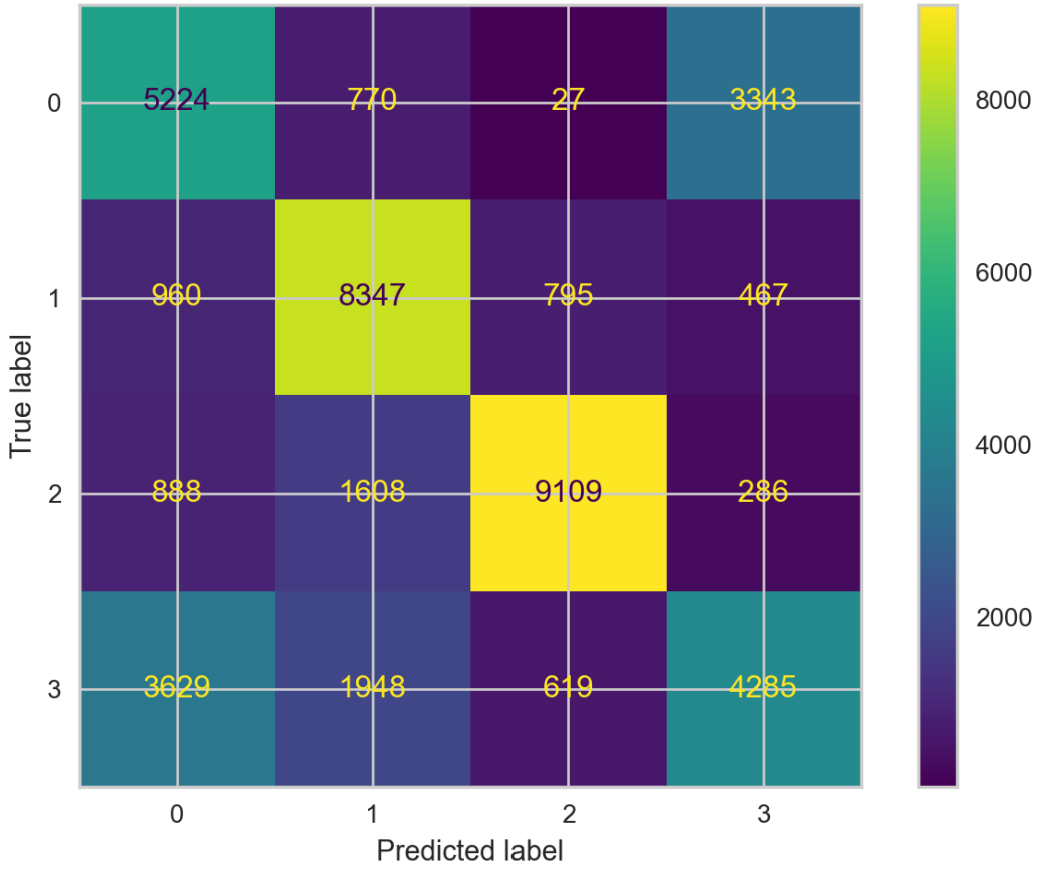
ii. Normalized Mutual Information: 0.34480192524547115

iii. Adjusted Mutual Information: 0.34475138747531725

iv. V Measure: 0.3448019252454711

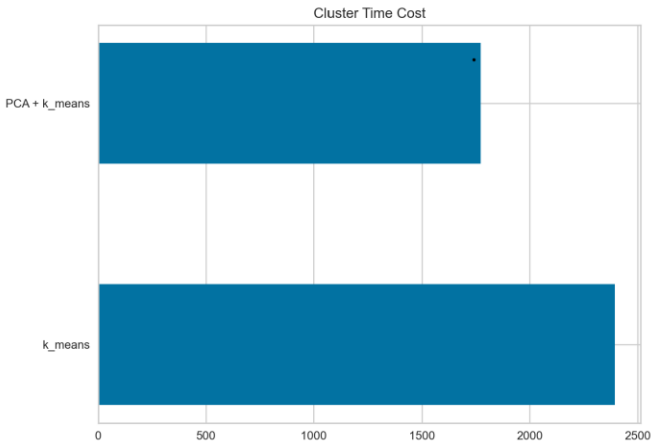
v. Fowlkes-Mallows Scores: 0.5127915062684616

vi. Confusion matrix




7. 與原先 k_means 比較

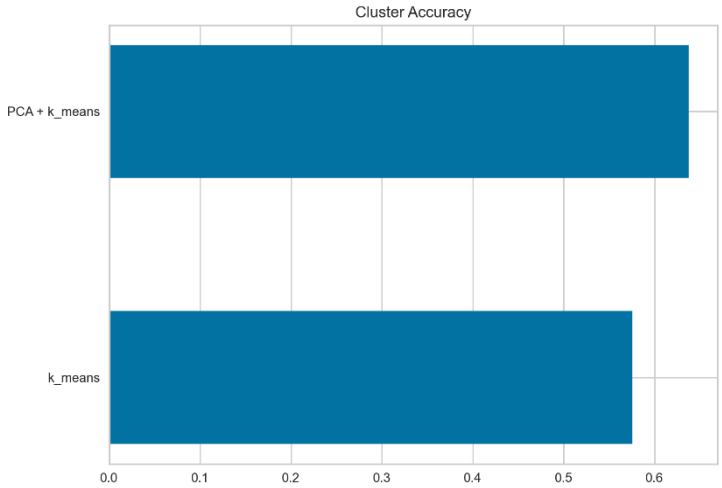
甲、效率




K_means after PCA time: 1772.4387645721436 ms

由上圖我們可以發現，經過 feature extraction 後，k_means 的效率較原先高

乙、效果




由上圖發現，經過 feature extraction 後，PCA+K_means 的正確率能顯著提升