

Clustering Hands-on Exercise 1

Spotify Clustering of Top Songs

Brock Tibert

2021-10-23

Business Challenge

Your firm has been hired by Spotify to cluster the top songs of 2018. Surprisingly they don't have information on the genre, so perhaps you can make heads and tails of the top songs based on their attributes recorded by Spotify's engineering team.

They have given you access to a modest dataset in Big Query

- `datasets.questrom.spotify_2018`

Data Dictionary

Information about the columns included in the data set can be found here.

[[https://developer.Spotify.com/documentation/web-api/reference/tracks/get-audio-features/](https://developer.spotify.com/documentation/web-api/reference/tracks/get-audio-features/)]

You should review what these attributes are **consider** their datatype. Just because they are numeric, are they really truly numeric data or are some features categorical?

Exercise:

With the data loaded, determine the number of clusters you believe to be present in the Spotify data set.

Considerations:

- What are the benefits to a limited number of estimated genres? A large number?
- What would you recommend to the Executive team? Did you make any assumptions that impacted your recommendation?