

## Anexo I. Descripción de las bases de datos empleadas

A continuación se explica con detalle el origen de las bases de datos empleadas en nuestro estudio.

### [Spotify\\_dataset\\_1921-2020](#)

Es un fichero relacionado con la música de la plataforma de *Spotify*, subida por el usuario Yamac Eren Ay. Este fichero contiene 172.230 canciones con 20 atributos diferentes, recolectadas directamente desde la propia API de *Spotify*. Además del nombre de la canción, contiene el artista, la fecha de lanzamiento, el identificador de la canción dentro de la API y muchas otras variables numéricas que explican características de la propia canción, como energía, popularidad, tempo, sonido. Las variables son:

- id
- acousticness
- danceability
- energy
- duration\_ms
- instrumentalness
- valence
- popularity
- tempo
- liveness
- loudness
- speechiness
- mode
- explicit
- key
- time\_signature
- artists
- artists\_id
- release\_date
- name

### [Spotify 1.2M+ songs](#)

Al igual que el anterior dataset, este contiene información de canciones de la plataforma de *Spotify*. Fue subida a la plataforma Kaggle por el usuario Rodolfo Figueroa. Al contener información muy parecida, los dos conjuntos son complementarios y podemos aumentar la cantidad de datos que tenemos a disposición para el proyecto. La única diferencia que se puede encontrar entre ellos es que este no contiene el atributo 'popularity' y 'time\_signature' (que puede resultarnos muy útil para un sistema de recomendación) pero sí tiene algunos identificadores (de disco y álbum) que no se contemplan en el otro fichero. Todos los demás atributos se encuentran en ambos ficheros. Contiene 1,204,025 canciones y 24 variables diferentes:

- id
- name
- album
- album\_id
- artists
- artists\_id
- track\_number
- disc\_number
- explicit
- danceability
- energy
- key

- loudness
- mode
- speechiness
- acousticness
- instrumentalness
- liveness
- valence
- tempo
- duration\_ms
- time\_signature
- year
- release\_date