

Reporte de análisis de datos

Datos generales:

El conjunto de datos seleccionado trata sobre una colección de información acerca de canciones de Spotify, cuantificando diversos aspectos de cada una como su ritmo, volumen, energía, etc. De modo que se cree un modelo predictivo a partir de este. Este conjunto se obtuvo de la página Kaggle ¹, donde se almacenan y publican diversos archivos para análisis de datos de varios temas.

Descripción de datos:

Como se puede ver en el archivo *Estadísticas.py*, se registraron 195 registros dentro de 14 variables distintas, cada una teniendo un rango que varía de 0 a 1, las cuales se describen a continuación:

Tabla 1. Variables y sus tipos de datos

<i>Nombre</i>	<i>Tipo de dato</i>
Danceability	Float
Energy	Float
Key	Int
Loudness	Float
Mode	Int
Speechiness	Float
Acousticness	Float
Instrumentalness	Float
Liveness	Float
Valence	Float
Tempo	Float
Duration_ms	Int
Time_signature	Int
Liked	Int

Análisis de datos:

Las dos variables que escogí para su análisis fueron Danceability y Energy, ya que me parece interesante descubrir alguna posible relación que tengan entre sí. El rango de estas variables que utilicé para el análisis fue el de todos los registros, con el fin de obtener valores más precisos en cuanto a los datos estadísticos.

¹ Vergnou, B. (2021) *Spotify Recommendation*. Kaggle <https://www.kaggle.com/bricevergnou/spotify-recommendation>

Tabla 2. Datos estadísticos de variables

<i>Nombre</i>	<i>Valor máximo</i>	<i>Valor mínimo</i>	<i>Media</i>	<i>Mediana</i>	<i>Desviación estándar</i>
Danceability	0.996	0.0024	0.638	0.659	0.260
Energy	0.946	0.130	0.637	0.705	0.217

Para empezar, todos los valores obtenidos parecen estar dentro de los rangos esperados, ya que la mayoría de los datos de la tabla contienen valores float que van desde 0 a 1.

Posteriormente, dado el análisis que regresa el programa de Python, podemos saber por la media de ambas variables que el conjunto de canciones dentro del csv son más favorables a que sean música para bailar de alta energía, aunque el valor de la media parece estar dentro del rango medio entre ambos extremos; además, el valor de la desviación estándar para ambos es relativamente pequeño, lo que quiere decir que, tanto la mayoría de canciones se encuentran dentro de este rango, así como que la persona que seleccionó estas canciones normalmente escoge canciones bastante parecidas entre sí. De momento, con la información obtenida con estas dos variables, se puede ver que los resultados son muy parecidos entre sí, lo que quiere decir que hay una posible relación entre ellos, aunque faltaría ver el comportamiento de las demás variables para ver si esto es cierto solo para estas o hay un comportamiento distinto todavía.