# Informe TPE Fundamentos de la Ciencia de Datos

Integrantes:

* Dimuro, Agustín Nicolás
* Grillo, Agustín German
* Padilla, Tomás Agustín

Introducción

Durante el transcurso del siguiente informe trabajaremos sobre un conjunto de datos proveniente de Amazon Music, el cual contiene observaciones sobre canciones de la década de los años 1970 al 1979. Realizaremos sobre este conjunto el filtrado de datos los cuales puedan ser erróneos para luego poder utilizarlos con el fin de buscar respuestas a ciertas hipótesis que puedan ser útiles para obtener información a cerca del comportamiento de las canciones en esa época.

Las hipótesis que planteamos y que vamos a estar analizando durante el transcurso del siguiente informe vamos a estar explicándolas a continuación.

* Las canciones en la primera mitad de la década tienen diferencias significativas con respecto al estilo, contra las de la segunda mitad. Poniendo el foco en el análisis sobre sus valores de energía, grados de acusticidad, pulsaciones por minuto y volumen promedio de la canción en decibeles.
* Las canciones compuestas con 4 pulsaciones por compas durante la década del 1970 actualmente son más populares que aquellas que fueron compuestas con un distinto número.
* Las canciones que tienen valores mas altos de positividad y alegría hoy en día son las más populares.
* Existe una relación entre las canciones que son bailables y las que son instrumentales.
* Las canciones que son más habladas con respecto a las demás son menos instrumentales.
* Las canciones que fueron grabadas durante un concierto en vivo tienen valores de energía mas altos que las demás.
* Las canciones que tienen mayor grado de instrumentalidad son las que más duración tienen comparando con las demás canciones de la década.

Materiales

Encontramos la siguiente explicacion de la columna ya que nos resultaba extraña

popularity: The popularity of a track is a value between 0 and 100, with 100 being the most popular. The popularity is calculated by algorithm and is based, in the most part, on the total number of plays the track has had and how recent those plays are. Generally speaking, songs that are being played a lot now will have a higher popularity than songs that were played a lot in the past. Duplicate tracks (e.g. the same track from a single and an album) are rated independently. Artist and album popularity is derived mathematically from track popularity.

sacado de un dataset de spotify de kagle tiene sentido

Métodos / Resultados

Conclusiones

Referencias