## Documentación del modelado para el Trabajo final de Visualización y Storytelling

## **"**Explorando los Sonidos de la Década: Un Viaje Musical a Través de la Data de Spotify"

Equipo 1:

Ana María Soto Orozco

Felipe Diego Lobato

Nicolas Yepes Joven

Yoselin Nieto Gil

1. **Introducción**

Se realizó una exploración a la base de datos con información asociada a la música escuchada en la plataforma Spotify en un rango de tiempo que va desde 1998 hasta 2020, las exploraciones se realizaron respecto a los artistas, emociones transmitidas por la música, ritmos y otras características musicales.

Fuente de los datos: <https://www.kaggle.com/code/varunsaikanuri/spotify-data-visualization/comments>

Forma del tablero

**Gama de colores del tablero:**Se utiliza <https://color.adobe.com/es/> para generar una paleta de colores acorde al logo de spotify. a continuación, se muestra la paleta base del tablero y el logo de Spotify en la parte superior debajo de cada color se encuentra su código de identificación

Imagen que contiene Gráfico

Descripción generada automáticamente

**Fuentes**

* Titulos: la fuente es Oleo Script
* Cuerpo: La fuente es Droid Sans

1. **Variables en la Base de datos original**

|  |  |  |  |
| --- | --- | --- | --- |
| **Nombre original de la variable en la base de datos** | **Nuevo nombre de la variable después del modelado** | **Tipo de variable (texto, número)** | **Descripción o significado** |
| artist | Artista | Texto | Nombre del artista |
| Song | Canción | Texto | Nombre del track o nombre de la canción |
| Duration\_ms | Duracion\_ms | Número | Duracion en milisegundos de la pista |
| explicit | Explicito | Booleano (verdadero o falso) | La letra o el contenido de una canción o un vídeo musical contienen criterios que podrían considerarse ofensivos o inadecuados para los niños. |
| year | Anio | Número | Año de lanzamiento de la pista. |
| popularity | Popularidad | Número | cuanto mayor sea el valor, más popular será la canción. |
| danceability | Bailable | Número | la bailabilidad describe qué tan adecuada es una pista para bailar en función de una combinación de elementos musicales que incluyen el tempo, la estabilidad del ritmo, la fuerza del ritmo y la regularidad general. Un valor de 0,0 es el menos bailable y 1,0 es el más bailable. |
| energy | Energia | Número | La energía es una medida de 0,0 a 1,0 y representa una medida perceptual de intensidad y actividad. |
| key | Clave (variable de codificacion de las notas musicales) y | Número | la clave en la que se encuentra la pista. Los números enteros se asignan a los tonos utilizando la notación estándar de clase de tono. P.ej. 0 = C, 1 = C♯/D♭, 2 = D, y así sucesivamente. Si no se detectó ninguna clave, el valor es -1. |
| nota\_clave (notas musicales contenidas en la variable) | Texto | la clave en la que se encuentra la pista. Los números enteros se asignan a los tonos utilizando la notación estándar de clase de tono. P.ej. 0 = C, 1 = C♯/D♭, 2 = D, y así sucesivamente. Si no se detectó ninguna clave, el valor es -1. |
| loudness | volumen | Número | el volumen general de una pista en decibeles (dB). Los valores de sonoridad se promedian en toda la pista y son útiles para comparar el volumen relativo de las pistas. El volumen es la cualidad de un sonido que es el principal correlato psicológico de la fuerza física (amplitud). Los valores suelen oscilar entre -60 y 0 db. |
| mode | Modo | Número | Modo indica la modalidad (mayor o menor) de una pista, el tipo de escala de la que se deriva su contenido melódico. El mayor está representado por 1 y el menor es 0. |
| speechiness | locuacidad | Número | el habla detecta la presencia de palabras habladas en una pista. Cuanto más exclusivamente hablada sea la grabación (por ejemplo, un programa de entrevistas, un audiolibro, poesía), más cercano a 1,0 será el valor del atributo. Los valores superiores a 0,66 describen pistas que probablemente estén compuestas exclusivamente de palabras habladas. Los valores entre 0,33 y 0,66 describen pistas que pueden contener música y voz, ya sea en secciones o en capas, incluidos casos como la música rap. Los valores inferiores a 0,33 probablemente representen música y otras pistas que no sean de voz. |
| acousticness | Acusticidad | Número | una medida de confianza de 0,0 a 1,0 sobre si la pista es acústica. 1.0 representa una alta confianza en que la pista es acústica. |
| instrumentalness | Instrumentalidad | Número | predice si una pista no contiene voces. Los sonidos "Ooh" y "aah" se tratan como instrumentales en este contexto. Las pistas de rap o de palabra hablada son claramente "vocales". Cuanto más cerca esté el valor de instrumentalidad de 1,0, mayor será la probabilidad de que la pista no contenga contenido vocal. Los valores superiores a 0,5 pretenden representar pistas instrumentales, pero la confianza es mayor a medida que el valor se acerca a 1,0. |
| liveness | Vivisidad | Número | Detecta la presencia de una audiencia en la grabación. Los valores de vida más altos representan una mayor probabilidad de que la pista se haya interpretado en vivo. Un valor superior a 0,8 proporciona una gran probabilidad de que la pista esté activa. |
| valence | Valencia emocional | Número | Una medida de 0,0 a 1,0 que describe la positividad musical transmitida por una pista. Las pistas con valencia alta suenan más positivas (por ejemplo, felices, alegres, eufóricas), mientras que las pistas con valencia baja suenan más negativas (por ejemplo, tristes, deprimidas, enojadas). |
| tempo | Tempo | Número | El tempo general estimado de una pista en pulsaciones por minuto (BPM). En terminología musical, el tempo es la velocidad o el ritmo de una pieza determinada y se deriva directamente de la duración promedio del tiempo. |
| genre | género | texto | Género musical de la pista. |

1. **Variables después del modelado**

A continuación, se indican las variables usadas en la base de datos transformada para el drashboard, esta tabla incluye las variables que fueron calculadas, estos procesos de cálculo se explican más adelante en el proceso de modelado

|  |  |  |
| --- | --- | --- |
| **Variable** | **Tipo de dato** | **Descripción** |
| Artista | Texto | nombre del artista |
| Canción | Texto | Nombre del track o nombre de la canción |
| Duracion\_ms | Número | Duración en milisegundos de la pista |
| Explicito | Booleano (verdadero o falso) | La letra o el contenido de una canción o un vídeo musical contienen criterios que podrían considerarse ofensivos o inadecuados para los niños. |
| Anio | Número | Año de lanzamiento de la pista. |
| Popularidad | Número | cuanto mayor sea el valor, más popular será la canción. |
| Bailable | Número | la bailabilidad describe qué tan adecuada es una pista para bailar en función de una combinación de elementos musicales que incluyen el tempo, la estabilidad del ritmo, la fuerza del ritmo y la regularidad general. Un valor de 0,0 es el menos bailable y 1,0 es el más bailable. |
| Energía | Número | La energía es una medida de 0,0 a 1,0 y representa una medida perceptual de intensidad y actividad. |
| Clave (variable de codificación de las notas musicales) | Número | la clave en la que se encuentra la pista. Los números enteros se asignan a los tonos utilizando la notación estándar de clase de tono. P.ej. 0 = C, 1 = C♯/D♭, 2 = D, y así sucesivamente. Si no se detectó ninguna clave, el valor es -1. |
| volumen | Número | el volumen general de una pista en decibeles (dB). Los valores de sonoridad se promedian en toda la pista y son útiles para comparar el volumen relativo de las pistas. El volumen es la cualidad de un sonido que es el principal correlato psicológico de la fuerza física (amplitud). Los valores suelen oscilar entre -60 y 0 db. |
| Modo | Número | Modo indica la modalidad (mayor o menor) de una pista, el tipo de escala de la que se deriva su contenido melódico. El mayor está representado por 1 y el menor es 0. |
| locuacidad | Número | el habla detecta la presencia de palabras habladas en una pista. Cuanto más exclusivamente hablada sea la grabación (por ejemplo, un programa de entrevistas, un audiolibro, poesía), más cercano a 1,0 será el valor del atributo. Los valores superiores a 0,66 describen pistas que probablemente estén compuestas exclusivamente de palabras habladas. Los valores entre 0,33 y 0,66 describen pistas que pueden contener música y voz, ya sea en secciones o en capas, incluidos casos como la música rap. Los valores inferiores a 0,33 probablemente representen música y otras pistas que no sean de voz. |
| Acusticidad | Número | una medida de confianza de 0,0 a 1,0 sobre si la pista es acústica. 1.0 representa una alta confianza en que la pista es acústica. |
| Instrumentalidad | Número | predice si una pista no contiene voces. Los sonidos "Ooh" y "aah" se tratan como instrumentales en este contexto. Las pistas de rap o de palabra hablada son claramente "vocales". Cuanto más cerca esté el valor de instrumentalidad de 1,0, mayor será la probabilidad de que la pista no contenga contenido vocal. Los valores superiores a 0,5 pretenden representar pistas instrumentales, pero la confianza es mayor a medida que el valor se acerca a 1,0. |
| Vivisidad | Número | Detecta la presencia de una audiencia en la grabación. Los valores de vida más altos representan una mayor probabilidad de que la pista se haya interpretado en vivo. Un valor superior a 0,8 proporciona una gran probabilidad de que la pista esté activa. |
| Valencia emocional | Número | Una medida de 0,0 a 1,0 que describe la positividad musical transmitida por una pista. Las pistas con valencia alta suenan más positivas (por ejemplo, felices, alegres, eufóricas), mientras que las pistas con valencia baja suenan más negativas (por ejemplo, tristes, deprimidas, enojadas). |
| Tempo | Número | El tempo general estimado de una pista en pulsaciones por minuto (BPM). En terminología musical, el tempo es la velocidad o el ritmo de una pieza determinada y se deriva directamente de la duración promedio del tiempo. |
| género | Texto | Género musical de la pista. |
| anio\_mes | número | Año y mes de lanzamiento de la pista |
| Categoría\_instrumentalidad | número | Es un campo calculado que implementa el campo instrumentalidad, si el valor es mayor o igual a 0.3 se cataloga como canción instrumental, en caso contrario es canción vocal |
| Década | número | Es un campo calculado que divide los años presentes en el archivo, en las 4 décadas a las cuáles pertenecen:  1990:1990-1999  2000:2000-2009  2010:2010-2019  2020:2020-2030 |

1. **Proceso de modelado**

**Procesamiento en Notebook de Python**

Inicialmente mediante un notebook de Python se cargó la Base de datos con las transformaciones iniciales("songs\_transformedfinal-songstransformed") y se procedió a realizar la limpieza de la base de datos:

* Se eliminaron los datos vacíos
* Se eliminaron datos duplicados
* Se clasificaron los géneros en 12 categorías: Pop, Hip hop, Rock, Dance, set(), latín, RyB, Country, World y Metal.

Se creó un diccionario de las variables originales para renombrarlas mediante código, el diccionario usado fue el siguiente:

nombres\_columnas = {

'artist': 'artista', 'song': 'cancion', 'duration\_ms': 'duracion\_ms', 'explicit': 'explicito',

'year': 'anio', 'popularity': 'popularidad', 'danceability': 'bailable', 'energy': 'energia',

'key': 'clave', 'loudness': 'volumen', 'mode': 'modo', 'speechiness': 'locuacidad',

'acousticness': 'acusticidad', 'instrumentalness': 'instrumentalidad', 'liveness': 'vivisidad', 'valence': 'valencia emocional', 'tempo': 'tempo', 'genre': 'genero'

}

Se transformó la variable clave bajo una codificación y los resultados se almacenaron en una nueva variable denominada nota\_clave, el diccionario usado para la transformación fue el siguiente:

mapeo\_clave = {

0: 'DO',1: 'DO#', 2: 'RE', 3: 'RE#', 4: 'MI', 5: 'FA', 6: 'FA#', 7: 'SOL',8: 'SOL#', 9: 'LA',10: 'LA#', 11: 'SI'}

En el archivo resultante se obtuvieron las siguientes variables en este orden:

columnas = ['artista', 'cancion', 'duracion\_ms', 'explicito', 'anio', 'popularidad', 'bailable', 'energia', 'clave', 'nota\_clave', 'volumen', 'modo', 'locuacidad', 'acusticidad', 'instrumentalidad', 'vivisidad', 'valencia emocional', 'tempo', 'genero']

finalmente, de la base de datos se decidió eliminar la variable nota\_clave ya que no era necesaria en lo que se plasmaría finalmente en el Tablero.

Finalmente, con esta transformación inicial, se creó la base de datos de trabajo para cargar a looker ("songs\_transformedfinal-songstransformed").