**Logotipo

Descripción generada automáticamente**

**Evaluación y Desarrollo de Modelos Predictivos en el Reto Spotify Million Playlist Dataset Challenge**

Sara Durango Ceballos  
Cristian Camilo Caballero Ayala

Monografía presentada para optar al título de Especialista en Analítica y Ciencia de Datos

Asesor  
David Manuel Villanueva

Universidad de Antioquia  
Facultad de Ingeniería

Especialización en Analítica y Ciencia de Datos

Medellín, Antioquia, Colombia

2023

|  |  |
| --- | --- |
| **Cita** | (Sara Durango Ceballos & Cristian Camilo Caballero Ayala, 2023), |
| **Referencia**  **Estilo APA 7 (2020)** | Durango Ceballos, S., & Caballero Ayala, C. C. (2023). *Título del trabajo* Trabajo de grado especialización]. Universidad de Antioquia, Medellín, Colombia. |

**** 

Especialización en Analítica y Ciencia de Datos, CohorteVI.

Centro de Investigación Ambientales y de Ingeniería (CIA).

|  |  |
| --- | --- |
|  | Diagrama  Descripción generada automáticamente con confianza media |

Centro de Documentación Ingeniería (CENDOI)

**Repositorio Institucional:** http://bibliotecadigital.udea.edu.co

Universidad de Antioquia - www.udea.edu.co

Rector: John Jairo Arboleda Céspedes.

Decano: Julio Cesar Saldarriaga Molina

Jefe departamento: Diego José Luis Botia Valderrama

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

**Dedicatoria**

A nuestras familias por el apoyo recibido

**Agradecimientos**

A los profesores que nos formaron durante esta especialización

**Tabla de contenido**

[Resumen 9](#_heading=h.3rdcrjn)

[Abstract 10](#_heading=h.26in1rg)

[1.](#_heading=h.35nkun2) Descripción del problema 11

[1.1.](#_heading=h.1ksv4uv) Problema de negocio 11

[1.2.](#_heading=h.44sinio) Aproximación desde la analítica de datos 11

[1.3.](#_heading=h.z337ya) Origen de los datos 11

[1.4.](#_heading=h.3j2qqm3) Métricas de desempeño 11

[2.](#_heading=h.4i7ojhp) Objetivos 12

[2.1.](#_heading=h.2xcytpi) Objetivo general 12

[2.2.](#_heading=h.3whwml4) Objetivos específicos 12

[3.](#_heading=h.2bn6wsx) Datos 13

[3.1.](#_heading=h.qsh70q) Datos originales 13

[3.2.](#_heading=h.3as4poj) Datasets 13

[3.3.](#_heading=h.1pxezwc) Analítica descriptiva 13

[4.](#_heading=h.49x2ik5) Proceso de analítica 14

[4.1.](#_heading=h.2p2csry) Pipeline principal 14

[4.2.](#_heading=h.147n2zr) Preprocesamiento 14

[4.3.](#_heading=h.3o7alnk) Modelos 14

[4.4.](#_heading=h.23ckvvd) Métricas 14

[5.](#_heading=h.ihv636) Metodología 15

[5.1.](#_heading=h.32hioqz) Baseline 15

[5.2.](#_heading=h.1hmsyys) Validación 15

[5.3.](#_heading=h.41mghml) Iteraciones y evolución 15

[5.4 Herramientas 15](#_heading=h.2grqrue)

[6.](#_heading=h.vx1227) Resultados y discusión 16

[6.1.](#_heading=h.3fwokq0) Métricas 16

[6.2.](#_heading=h.1v1yuxt) Evaluación cualitativa 16

[6.3.](#_heading=h.4f1mdlm) Consideraciones de producción 16

[7.](#_heading=h.kgcv8k) Conclusiones 20

[8.](#_heading=h.34g0dwd) Recomendaciones 21

[Referencias 22](#_heading=h.1jlao46)

[Anexos 23](#_heading=h.43ky6rz)

[Anexo 1. Autoarchivo en Repositorio y documentos de interés 24](#_heading=h.xvir7l)

[Anexo 2. Gestor de citas y referencias de Microsoft Word Microsoft Word - Wikipedia 25](#_heading=h.1baon6m)

[Anexo 3. Citas y referencias de material legal (leyes, decretos, sentencias, etc.) 27](#_heading=h.3vac5uf)

[Anexo 4. Ortografía y gramática 30](#_heading=h.2afmg28)

[Anexo 5. Buscar, reemplazar y eliminar espacios (o palabras) 32](#_heading=h.pkwqa1)

[Anexo 6. Atajos de teclado útiles en Microsoft Word 33](#_heading=h.39kk8xu)

[Anexo 7. Sinónimos y antónimos 34](#_heading=h.1opuj5n)

[Anexo 8. Copiar y pegar sin formato 35](#_heading=h.48pi1tg)

[Anexo 9. Comparar dos documentos 36](#_heading=h.2nusc19)

[Anexo 10. Control de cambios 37](#_heading=h.1302m92)

[Anexo 11. Insertar salto de página 39](#_heading=h.3mzq4wv)

[Anexo 12. Recortar y abreviar direcciones web largas 40](#_heading=h.2250f4o)

**Lista de tablas**

[**Tabla 1**](#_heading=h.2u6wntf). Descripción de las variables, base de datos cruda [1](#_heading=h.2u6wntf)7

**Lista de figuras**

[**Figura 1** Portada Normas APA séptima edición 2020 en inglés](#_heading=h.2r0uhxc)

[**Figura 2** Logo Universidad de Antioquia](#_heading=h.1664s55)

**Figura 3** Tipos de datos asociados a las variables del dataset 21

# Figura 4. Distribución de variables numéricas 22

**Figura 5.** Correlación entre variables numéricas 22

**Figura 6.** Ranking de Artistas en Playlist de Spotify 23

**Figura 7**. Top de Álbumes en Playlist de Spotify 23

**Siglas, acrónimos y abreviaturas**

**APA** American Psychological Association

**Cms.** Centímetros

**ERIC** Education Resources Information Center

**Esp.** Especialista

**MP** Magistrado Ponente

**MSc** Magister Scientiae

**Párr.** Párrafo

**PhD** Philosophiae Doctor

**PBQ-SF** Personality Belief Questionnaire Short Form

**PostDoc** PostDoctor

**UdeA** Universidad de Antioquia

# Resumen

La constante evolución de las plataformas digitales, como Spotify, ha generado desafíos en la mejora continua de sus sistemas de recomendación musical. Este problema se ve exacerbado por la diversidad de gustos de los usuarios, los cambios en el estado de ánimo y contexto de éstos, la necesidad de descubrir nueva música y la competencia en la industria del streaming. Con el objetivo de abordar estos desafíos, Spotify ha lanzado el concurso "The Spotify Million Playlist Dataset Challenge", promoviendo la investigación en algoritmos de recomendación para aumentar la retención de usuarios, la satisfacción del cliente y el tiempo de uso en la plataforma.

En el marco del concurso, este estudio se apoya en los datos proporcionados por Spotify para desarrollar un modelo predictivo que permita mejorar la continuación automática de listas de reproducción al incluir automáticamente canciones y artistas relacionados a partir de patrones de escucha de los usuarios, teniendo en cuenta variables como la duración de las canciones, el nombre del artista, el álbum musical y los nombres de las listas de reproducción, entre otros; con el propósito de mejorar la personalización y satisfacción del usuario al usar la plataforma.

Los datos provienen del dataset proporcionado por Spotify para el concurso, el cual comprende más de un millón de listas de reproducción y dos millones de pistas. Para abordar esta problemática, se realizó un análisis exploratorio inicial de los datos, se transformaron y prepararon los datos crudos, y se crearon diferentes modelos de machine learning, permitiendo definir cuál de ellos se ajusta mejor a los datos entrenados vs predichos.

Este trabajo busca contribuir a la mejora del sistema de recomendación de Spotify, brindando una experiencia más enriquecedora y adaptada a las preferencias individuales de los usuarios.

<https://github.com/SaraDurango/spotify_dataset_challenge/tree/main>

***Palabras clave*: Spotify Million Playlist Challenge, predicción, playlist, datos .**

# Abstract

The constant evolution of digital platforms, such as Spotify, has presented challenges in the continuous enhancement of their music recommendation systems. This issue is further exacerbated by the diversity of user preferences, changes in their mood and context, the need to discover new music, and the competition in the streaming industry. In addressing these challenges, Spotify has initiated "The Spotify Million Playlist Dataset Challenge," promoting research in recommendation algorithms to increase user retention, customer satisfaction, and platform usage time.

Within the framework of the competition, this study relies on data provided by Spotify to develop a predictive model aimed at improving the automatic continuation of playlists by incorporating songs and related artists based on user listening patterns. Key variables considered include song duration, artist name, music album, and playlist names, among others, with the purpose of enhancing user personalization and satisfaction while using the platform.

The data originates from Spotify's extensive dataset, comprising over a million playlists and two million tracks. To address this issue, an initial exploratory analysis of the data was conducted, raw data was transformed and prepared, and various machine learning models were created to determine which best fits the trained versus predicted data.

This work seeks to contribute to the enhancement of Spotify's recommendation system, providing a more enriching and tailored experience based on individual user preferences.

***Keywords: Spotify Million Playlist Challenge, prediction, playlist, data****.*

# Descripción del problema

Las plataformas digitales como Spotify enfrentan un desafío constante para mejorar su

sistema de recomendación de canciones para los usuarios debido la diversidad de gustos,

los cambios de ánimo y de contexto en el que se desenvuelven los usuarios, el

descubrimiento de nueva música, la necesidad de generar una experiencia cada vez más

personalizada y la competencia de la industria. [2]

Dado esto, Spotify propone promover la investigación en los algoritmos de

recomendaciones musicales mediante un concurso abierto al público, con el fin de mejorar

el sistema de recomendación de canciones buscando aumentar la retención de usuarios,

la satisfacción del cliente, el tiempo de uso de la plataforma y el éxito continuo en un

mercado altamente competitivo y en constante evolución. [1]

## Problema de negocio

Spotify enfrenta un desafío constante para mejorar su sistema de recomendación de canciones para los usuarios debido a varias razones:

• Diversidad de gustos: Los usuarios de Spotify tienen una amplia gama de preferencias musicales. Algunos pueden disfrutar de varios géneros y artistas, lo que hace que sea un desafío proporcionar recomendaciones precisas que se adapten a gustos cambiantes y diversos.

• Cambios en el estado de ánimo y contexto: La música que una persona prefiere puede variar según su estado de ánimo, actividad o contexto. Por lo tanto, es fundamental que las recomendaciones sean sensibles a estos factores cambiantes.

• Descubrimiento de nueva música: Los usuarios esperan que el sistema de recomendación no solo les ofrezca música que ya conocen y les gusta, sino que también les ayude a descubrir nueva música que sea relevante para sus gustos.

• Experiencia personalizada: La competencia en el mercado de la transmisión de música es intensa, y los servicios como Spotify buscan diferenciarse ofreciendo una experiencia altamente personalizada. Esto implica la necesidad de refinar constantemente los algoritmos de recomendación para adaptarse a las preferencias de cada usuario.

• Evitar la fatiga del usuario: Si un usuario recibe recomendaciones irrelevantes con frecuencia, puede llevar a la insatisfacción y, en última instancia, a la pérdida de interés en el servicio. Por lo tanto, mejorar la precisión de las recomendaciones es fundamental para mantener a los usuarios comprometidos.

• Competencia en la industria: La competencia en el mercado de la música en streaming es feroz. Para retener y atraer a más usuarios, Spotify debe ofrecer un sistema de recomendación superior que no solo se base en lo que ya conocen los usuarios, sino que también brinde valor al descubrir nueva música de manera efectiva.

Dado esto, mejorar el sistema de recomendación de canciones es esencial para la retención de usuarios, la satisfacción del cliente y el éxito continuo de Spotify en un mercado altamente competitivo y en constante evolución.

## Aproximación desde la analítica de datos

Un modelo de machine learning puede mejorar la predicción de canciones a añadir automáticamente a una lista de reproducción de acuerdo con diversos patrones de comportamiento del usuario en la plataforma. Para desarrollar dicho modelo se pueden explorar diferentes técnicas hasta encontrar la que mejor pronostique la recomendación de reproducción automática de acuerdo con los gustos de los usuarios.

Para esto se desarrollarán las siguientes fases:

• Preparación de datos:

Utilizar el conjunto de datos de listas de reproducción proporcionado por Spotify, que contiene información sobre las listas de reproducción y las pistas que contienen. Extraer características relevantes de las listas de reproducción y las pistas, como género musical, tempo, popularidad, artistas, etc.

• Creación de modelo Machine Learning:

Entrenar un modelo de ML utilizando las características de las listas de reproducción y las pistas como datos de entrada.

• Generación de recomendaciones Cuando un usuario proporciona una lista de reproducción específica, el algoritmo de ML puede identificar las listas de reproducción más similares en función de las características. Las pistas de las listas de reproducción similares pueden recomendarse como continuación automática para la lista de reproducción del usuario.

• Ajuste de hiperparámetros:

Es importante ajustar los hiperparámetros para mejorar el rendimiento y la capacidad de generalización del modelo

• Evaluación y validación:

Evaluar el rendimiento del modelo utilizando métricas como precisión, recall o F1-score para verificar la calidad de las recomendaciones generadas.

• Optimización continua:

El modelo de ML puede ser reentrenado periódicamente para adaptarse a las tendencias cambiantes de las listas de reproducción y las preferencias de los usuarios.

## Origen de los datos

La fuente de datos es el propio dataset de Spotify de acceso público para el reto de

reproducción automática. Dicho dataset comprende 1 millón de listas de reproducción, consta de más de 2 millones de pistas únicas de casi 300 000 artistas y representa el conjunto de datos públicos de listas de reproducción de música más grande del mundo. Además, incluye listas de reproducción públicas creadas por usuarios de Spotify de EE. UU y se compone de datos relevantes como el título de la lista de reproducción, la lista de pistas (incluidos los ID de las pistas) y otros campos de metadatos (hora de la última edición, número de ediciones de la lista de reproducción, entre otros). Todos los datos son anonimizados para proteger la privacidad del usuario. Las listas de reproducción se muestrean con cierta aleatorización, se filtran manualmente para determinar la calidad de la lista de reproducción y eliminar contenido ofensivo, y se les agregan algunas pistas ficticias y vacilantes.

Spotify proporciona adicionalmente, un conjunto de datos de prueba, con el cual se busca validar los pronósticos del modelo. En dicho conjunto de datos se incluyen listas de reproducción incompletas, con la finalidad de completarlas usando el modelo desarrollado, para, posteriormente, evaluar la asertividad en dicha predicción.

## Métricas de desempeño

El modelo presentado se evaluará, tanto a nivel de pista (coincidencia exacta de pista,) como a nivel de artista (cualquier pista del mismo artista es una coincidencia), utilizando la siguiente

métrica:

• Precisión R:

La precisión R es el número de pistas relevantes recuperadas dividido por el número de pistas

relevantes conocidas (es decir, el número de pistas retenidas):

La métrica se promedia en todas las listas de reproducción del conjunto de desafíos. Esta

métrica recompensa el número total de pistas relevantes recuperadas (independientemente del

orden)

Como parte del desafío, Spotify otorga adicionalmente, un conjunto de datos de prueba que

consta de 10,000 listas de reproducción con información incompleta. Éste tiene muchos de los

mismos campos del dataset de entrenamiento, además de seguir la misma estructura, pero las

listas de reproducción pueden incluir metadatos incompletos (sin título) y solo incluyen K pistas.

Más específicamente, el conjunto de datos del desafío se divide en 10 escenarios, con 1000

ejemplos de cada escenario:

• Sólo título (sin pistas)

• Título y primera pista

• Título y primeras 5 pistas.

• Solo las primeras 5 pistas

• Título y primeras 10 pistas.

• Solo las primeras 10 pistas

• Título y primeras 25 pistas.

• Título y 25 pistas aleatorias.

• Título y primeras 100 pistas.

• Título y 100 pistas aleatorias

Se busca entonces, que para cada escenario del conjunto de datos prueba, se obtenga un

pronóstico de lista de canciones recomendadas. [3]

# Objetivos

## Objetivo general

Desarrollar un modelo predictivo utilizando los datos de Spotify del concurso "The Spotify Million Playlist Dataset Challenge". para continuar automáticamente las listas de reproducción con canciones y/o artistas relacionados a partir de patrones de escucha de los usuarios teniendo en cuenta variables como género, duración, artista, entre otras, fomentando una experiencia de usuario más personalizada y satisfactoria.

## Objetivos específicos

* Estructurar el dataset inicial a partir de los diccionarios proporcionados por el desafío, garantizando su procesabilidad y comprensión.
* Seleccionar e implementar algoritmos que optimicen la precisión y relevancia de las recomendaciones.
* Establecer un patrón de recomendación a partir de las variables de usuario que permita tener una experiencia personalizada.

.

# Datos

## Datos originales

La fuente de datos es el propio dataset de Spotify de acceso público para el reto de reproducción automática. Dicho dataset comprende 1 millón de listas de reproducción y consta de más de 2 millones de pistas únicas de casi 300 000 artistas. Además incluye listas de reproducción públicas creadas por usuarios de Spotify de EE. UU y se compone de datos relevantes como el título de la lista de reproducción, la lista de pistas (incluidos los ID de las pistas) y otros campos de metadatos (hora de la última edición, número de ediciones de la lista de reproducción, entre otros). Cabe resaltar que estos datos presentados por Spotify fueron previamente anonimizados para proteger la privacidad de los usuarios y que las listas de reproducción fueron previamente muestreadas con cierta aleatorización; además se filtraron manualmente para determinar la calidad de éstas y eliminar contenido ofensivo.

| **Variable** | **Definición** |
| --- | --- |
| Nombre Playlist | Nombre de una playlist de Spotify |
| Número de Pistas | Número de canciones dentro de una playlist de Spotify |
| Duración Playlist | Duración total de toda la lista de reproducción (en ms) |
| Número de Seguidores | Número de seguidores de la playlist |
| Nombre Pista | Nombre de la canción o pista de la playlist |
| Nombre Artista | Nombre del artista |
| Nombre Álbum Musical | Nombre del Álbum Musical al que corresponde la pista |
| Duración Pista | Duración de una pista de la lista de reproducción (en ms) |
| Posición | Posición de la pista dentro de la playlist |
| Track URI | Identificación o ID de la pista dentro de Spotify |
| Artist URI | Identificación o ID del artista dentro de Spotify |
| Álbum URI | Identificación o ID del álbum musical dentro de Spotify |

* **Descripción de variables**

En la siguiente tabla se describen las variables que comprenden los datos proporcionados por Spotify

*Tabla 1. Descripción de las variables, base de datos cruda*

A continuación, se muestra un ejemplo de una entrada típica de una lista de reproducción:

{

"nombre": "musical",

"colaborativo": "falso",

"pid": 5,

"modificado\_at": 1493424000,

"num\_álbumes": 7,

"núm\_pistas": 12,

"num\_seguidores": 1,

"num\_edits": 2,

"duración\_ms": 2657366,

"num\_artistas": 6,

"pistas": [

{

"pos": 0,

"artist\_name": "Degiheugi",

"track\_uri": "spotify:pista:7vqa3sDmtEaVJ2gcvxtRID",

"artist\_uri": "spotify:artista:3V2paBXEoZIAhfZRJmo2jL",

"track\_name": "Final",

"album\_uri": "spotify:álbum:2KrRMJ9z7Xjoz1Az4O6UML",

"duración\_ms": 166264,

"album\_name": "Acordes danzantes y luciérnagas"

},

{

"pos": 1,

"artist\_name": "Degiheugi",

"track\_uri": "spotify:pista:23EOmJivOZ88WJPUbIPjh6",

"artist\_uri": "spotify:artista:3V2paBXEoZIAhfZRJmo2jL",

"track\_name": "Betty",

"album\_uri": "spotify:álbum:3lUSlvjUoHNA8IkNTqURqd",

"duración\_ms": 235534,

"album\_name": "Sonrisa sin fin"

},

{

"pos": 2,

"artist\_name": "Degiheugi",

"track\_uri": "spotify:pista:1vaffTCJxkyqeJY7zF9a55",

"artist\_uri": "spotify:artista:3V2paBXEoZIAhfZRJmo2jL",

"track\_name": "Algo de ritmo en mi cabeza",

"album\_uri": "spotify:álbum:2KrRMJ9z7Xjoz1Az4O6UML",

"duración\_ms": 268050,

"album\_name": "Acordes danzantes y luciérnagas"

},

// 8 pistas omitidas

{

"pos": 11,

"artist\_name": "Mo' Horizons",

"track\_uri": "spotify:pista:7iwx00eBzeSSSy6xfESyWN",

"artist\_uri": "spotify:artista:3tuX54dqgS8LsGUvNzgrpP",

"track\_name": "Fiebre 99\u00b0",

"album\_uri": "spotify:álbum:2Fg1t2tyOSGWkVYHlFfXVf",

"duración\_ms": 364320,

"album\_name": "Ven a tocar el sol"

}

],

}

Spotify proporciona adicionalmente, un conjunto de datos de prueba, con el cual se busca validar los pronósticos del modelo. En dicho conjunto de datos se incluyen listas de reproducción incompletas, con la finalidad de completarlas usando el modelo desarrollado, para posteriormente evaluar la asertividad en dicha predicción.

## Datasets

Para participar en el concurso, Spotify proporcionan dos bases de datos: la base de datos completa que mencionamos anteriormente, cuyo alcance abarca un millón de listas de reproducción; y una base de datos de entrenamiento que comprende una muestra con 10.000 registros, en los que cada registro representa una playlist con información detallada como nombre, duración, pistas, número de seguidores, entre otras variables presentadas anteriormente.

Para preparar los datos y obtener el dataset inicial o tabla origen, se realizaron los siguientes pasos:

● Eliminación de columnas que no aportan valor, como la fecha de última actualización de la playlist, el número de modificaciones y columnas en blanco

● Renombrado de algunas columnas para hacerlas más legibles.

● Eliminación de listas de reproducción vacías o sin ninguna pista añadida

● Conversión de las variables de tiempo de milisegundos a segundos

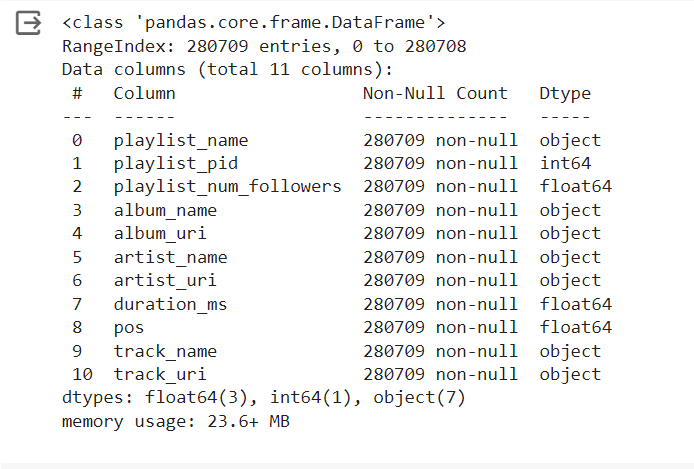
● Escalamiento de variables numéricas

● Conversión de variables categóricas en dummies

## Analítica descriptiva

Inicialmente se identificaron los tipos de variables y se hicieron algunas transformaciones necesarias para mejor manejo de los datos. Por ejemplo, la variable asociada al número de seguidores de las playlist se tomaba como texto y no como float 64.

En la figura 1 se ilustran los tipos de datos asociados a cada variable del dataset

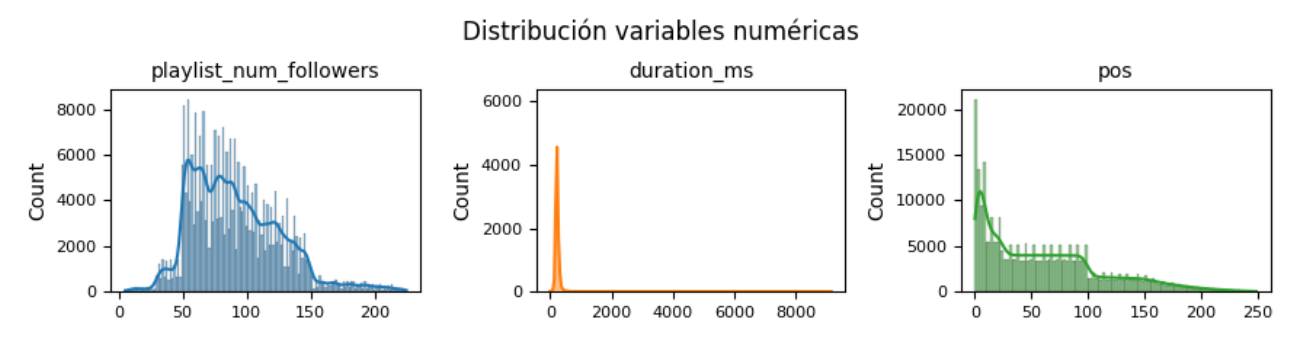


*Figura 3. Tipos de datos asociados a las variables del dataset*

El análisis exploratorio mostró que en promedio, una playlist de Spotify puede tener alrededor de 91 seguidores.

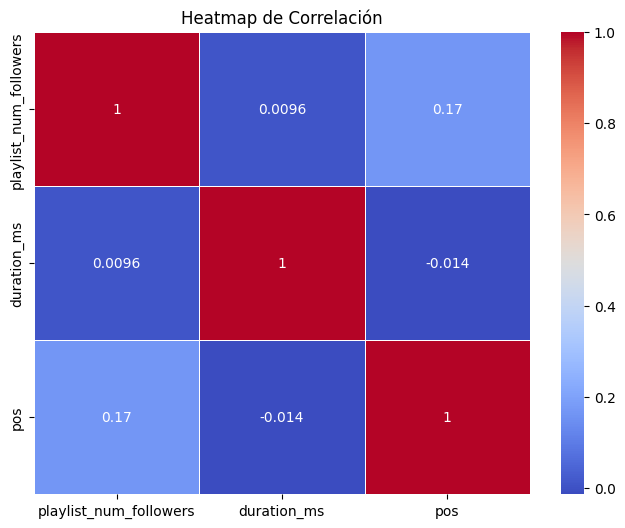
La duración promedio de las canciones es aproximadamente de 3.88 minutos. Además, la dispersión en las duraciones de las canciones es moderada, ya que el IQR va de 3.33 minutos a 4.27 minutos.

Por otro lado, la cantidad de canciones de una playlist es aproximadamente 59 canciones, pero se han presentado playlist hasta de 259 pistas.

**Distribución de variables numéricas**

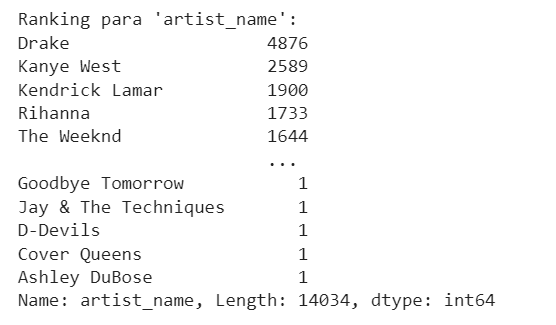
# *Figura 4. Distribución de variables numéricas*

De acuerdo a las distribuciones observadas, puede inferirse que las variables numéricas no tienen una distribución normal y podría contemplarse alguna transformación si se requiere posteriormente.



*Figura 5. Correlación entre variables numéricas*

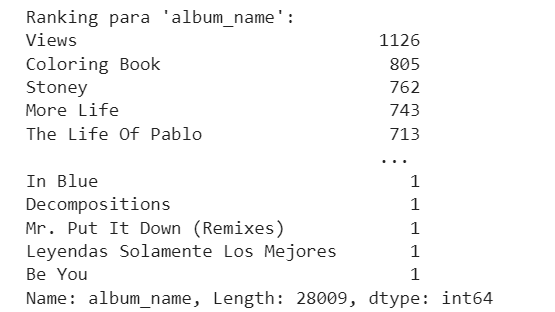
El gráfico de correlación muestra que no existe una correlación lineal fuerte entre las variables numéricas, lo que sugiere una relación de independencia entre las mismas e indica que no son variables redundantes.



*Figura 6. Ranking de Artistas en Playlist de Spotify*

De acuerdo con la figura 4, el artista con mayor participación en las playlist de Spotify es Drake, seguido de Kanye West, Kendrick Lamar, Rihanna y The Weekend.

Mientras que los álbumes cuyas canciones tienen mayor presencia en playlist de Spotify son Views del artista Drake, Coloring Book de Chance the Rapper, Stoney More Life de Post Malone y The Life of Pablo de Kanye West, como se ilustra en la figura 5.



*Figura 7. Top de Álbumes en Playlist de Spotify*

# Referencias

[1] S. Antenucci, S. Boglio, E. Chioso, E. Dervishaj, S. Kang, T. Scarlatti, and M. F. Dacrema. Artist-driven layering and user’s behaviour impact on recommendations in a playlist continuation scenario. In Proceedings of the 2018 ACM Recommender Systems Challenge, RecSysChallenge ’18, Vancouver, BC, Canada, 2018

[2] Arango Archila, F. (2016). El impacto de la tecnología digital en la industria discográfica. Dixit, 24(1), ISSN 1688-3497 (versión impresa) / ISSN 0797-3691 (versión en línea). Pontificia Universidad Javeriana, Bogotá, Colombia.

[3] CW Chen, P. Lamere, M. Schedl y H. Zamani. Recsys Challenge 2018: Continuación automática de listas de reproducción de música. En actas de la 12.ª Conferencia ACM sobre sistemas de recomendación (RecSys '18), 2018