

## **Proyecto Final**

### **Bootcamp de Análisis de Datos**

# ¿Coinciden la crítica y la audiencia sobre quiénes son los grandes artistas musicales?

*Un análisis en la era del streaming a partir del ranking histórico  
de Rolling Stone 2020 y la popularidad en Spotify.*

**Nombre:** Camila Bozzoletti

**Institución:** Código Facilito

**Programa:** Bootcamp de Análisis de Datos

**Fecha de entrega:** Abril 2025

## 1. Resumen Ejecutivo

El propósito de este proyecto fue analizar el **grado de coincidencia entre la canonización de artistas por parte de la crítica especializada** (a través del Ranking 2020 de la Revista Rolling Stone) y la consagración de la audiencia popular en la era del streaming (medida a través del Índice de Popularidad de Spotify). A lo largo de este análisis, se identificaron las características predominantes de los artistas más destacados según ambas fuentes, con el objetivo de comprender las diferencias en los criterios de evaluación que siguen estos dos actores en el mundo musical contemporáneo.

Los principales hallazgos incluyen que la crítica especializada y la audiencia en Spotify valoran a los artistas según criterios muy distintos, reflejando diferencias históricas, culturales y generacionales. Mientras que la crítica prioriza el legado y la influencia cultural, Spotify tiende a presentar artistas más jóvenes y con trayectorias más cortas. Además, se observó una prevalencia de géneros musicales clásicos en la crítica, como el Soul y el Rock, frente a una mayor predilección por géneros más recientes como el Indie y el Alternative Rock en Spotify. Otro hallazgo relevante fue que ambas métricas muestran una escasa representación femenina y de otras identidades sexo-genéricas entre los artistas más destacados. Además, las diferencias en popularidad y valoración entre la crítica y la audiencia sugieren que Spotify tiene un sesgo hacia lo 'nuevo' y lo más inmediato, mientras que la crítica se enfoca en la consolidación a lo largo del tiempo.

Este análisis no sólo pone en evidencia los contrastes entre los criterios de la crítica y la audiencia, sino también los sesgos inherentes en ambos sistemas, tanto los culturales e históricos como los derivados del algoritmo de Spotify.

## 2. Fundamento temático y marco contextual del problema

### 2.A. Elección de la temática

Esta investigación surge del cruce entre dos mundos que me apasionan: la **lectura crítica de los fenómenos sociales y culturales**, y el potencial del **análisis de datos** para convertir intuiciones en conocimiento accionable. Como Licenciada en Trabajo Social, siempre me ha interesado entender cómo se construyen los discursos, las jerarquías simbólicas y los imaginarios colectivos. Y pocas cosas condensan tanto de eso como la música.

La pregunta que guía este trabajo —¿los rankings más reconocidos reflejan realmente lo que valora la audiencia?— no es solo un interrogante técnico: es también una invitación a revisar cómo se consagra el valor cultural y **quiénes tienen voz en ese proceso**.

Elegí este tema no solo por su valor analítico, sino también por su relevancia estratégica en términos de negocio. En un contexto donde la industria musical necesita comprender con más precisión a su audiencia, conocer si la crítica y el público coinciden en sus valoraciones puede orientar decisiones empresariales, campañas de marketing o descubrimiento de nuevos talentos. Al

mismo tiempo, este análisis busca abrir una conversación más amplia: ¿quién define el valor cultural en una época marcada por el consumo digital masivo?

## 2.B. Contexto del problema

Pese al crecimiento sostenido de la industria musical en los últimos años —especialmente gracias al auge del streaming— aún persisten **desafíos** estructurales que no pueden ser ignorados. Tal como plantea el artículo<sup>1</sup> de *iMusician*, el sector musical enfrenta demandas urgentes por **mayor representación de género, raza y diversidad cultural**, así como también tensiones frente a los **vertiginosos cambios tecnológicos** que transforman las formas de producir, distribuir y consumir música.

En este escenario, comparar rankings como el de *Rolling Stone* con indicadores de popularidad en plataformas como Spotify adquiere una relevancia estratégica. Mientras que la crítica especializada *lleva la batuta* en la definición del canon musical, la masividad de Spotify ofrece una radiografía concreta de lo que la audiencia valora hoy. Todo esto motiva a atender, como intenta realizar esta investigación, qué es lo que busca el público a comparación de lo dictaminado por la crítica, porque si bien esta última siempre será destacada por su alto conocimiento en el sector, el ojo nunca debe alejarse demasiado de lo que la audiencia elige y espera.

Para la industria, conocer estas diferencias tiene un **alto valor estratégico**. Para el campo cultural, aporta nuevas lecturas sobre la **evolución del gusto colectivo**. Y para el análisis de datos, representa una oportunidad única de usar **información concreta** para iluminar un fenómeno social en transformación, a la par de aportar conclusiones a una de las áreas que más está demandando analistas de datos: la **industria del entretenimiento**.

## **3. Pregunta de Investigación y Objetivos**

### 3.A. Pregunta de investigación

¿Coinciden la crítica y la audiencia sobre quiénes son los grandes artistas musicales?

### 3.B. Objetivo general

- Analizar el grado de coincidencia entre la canonización de artistas por parte de la crítica especializada (Rolling Stone) y la consagración de la audiencia popular en la era del streaming.

### 3.C. Objetivos específicos

- Releva a los artistas más reconocidos por la crítica especializada a partir del ranking 2020 de Rolling Stone incluido en el dataset seleccionado para trabajar.

---

<sup>1</sup> IMusician (2023): “State of the Music Industry 2024: On Growth, Challenges, and the Need for Tangible Solutions”.

<https://imusician.pro/en/resources/blog/state-of-the-music-industry-2024-on-growth-challenges-and-the-need-for-tangible-solutions#:~:text=According%20to%20the%20most%20recent,9%25%20of%20the%20global%20market.>

- Identificar a los artistas más valorados por el público en la era del streaming a partir del indicador de popularidad de Spotify incluido en el dataset<sup>2</sup> seleccionado para trabajar.
- Comparar las características predominantes de los artistas reconocidos por la crítica especializada y por la audiencia popular.

## 4. Aspectos metodológicos

Para guiar el desarrollo del presente proyecto se ha optado por utilizar la **metodología CRISP-DM** (*Cross-Industry Standard Process for Data Mining*), dado que constituye un enfoque estructurado, flexible y ampliamente adoptado en proyectos de ciencia de datos. Se trata de una metodología que permite abordar de manera ordenada y cíclica el proceso analítico, facilitando la planificación, ejecución y evaluación de cada etapa del trabajo. Su enfoque iterativo resulta especialmente adecuado para proyectos exploratorios como el presente, en los que la comprensión del problema y de los datos se va profundizando a medida que avanza el análisis.

A su vez, la implementación de esta metodología favorece un tratamiento integral del **ciclo de vida de los datos**, al contemplar de manera sistemática todas las fases que atraviesan los datos: desde su recolección y exploración inicial, pasando por su transformación y análisis, hasta su visualización, almacenamiento o eventual archivo. Este enfoque contribuye a garantizar un proceso analítico riguroso, y alineado con buenas prácticas en ciencia de datos.

En cuanto al enfoque metodológico, este proyecto se inscribe dentro de un abordaje **cuantitativo**, basado en el **análisis exploratorio y descriptivo** de datos estructurados. Se trabajará a partir de un único dataset, utilizando métricas e indicadores disponibles a partir del mismo para describir, comparar y visualizar patrones presentes en los rankings de artistas. El uso de herramientas de programación y visualización como Python y Looker Studio permitirán procesar, transformar y representar gráficamente los datos, sin incorporar técnicas cualitativas propias de otros enfoques. No obstante, algunas interpretaciones finales podrán incorporar **lecturas de tipo contextual o conceptual**, en diálogo con categorías socioculturales vinculadas al campo musical.

### 4.1. Alcances y Decisiones Metodológicas

Este apartado tiene por objetivo explicitar los **criterios metodológicos** adoptados a lo largo del proyecto. Entre ellos se incluyen: la definición del tipo de relación a establecer entre variables; ciertos recortes analíticos realizados en función de la viabilidad del trabajo; los fundamentos detrás de la priorización de determinadas métricas por sobre otras; y algunas aclaraciones conceptuales necesarias para interpretar correctamente las variables en juego. A continuación, se encuentra el detalle de cada uno de los criterios mencionados:

#### 4.1.1. La relación entre variables es de tipo asociativa

---

<sup>2</sup> *Rolling Stone Album Rankings* dataset. Publicado por los usuarios *bradfordjohnson* y *jonthegeek* en el repositorio de GitHub del proyecto #TidyTuesday. Disponible en: <https://github.com/rfordatascience/tidytuesday/tree/main/data/2024/2024-05-07>

El proyecto no busca establecer relaciones de causa-efecto entre las variables analizadas, sino identificar **asociaciones relevantes** que permitan interpretar **patrones significativos**. Esto se debe a que el enfoque adoptado es de tipo exploratorio y descriptivo, y no explicativo. Se parte del supuesto de que ciertos atributos de los artistas (como el género o la edad) podrían estar asociados a mayores niveles de consagración según la crítica o la audiencia, sin afirmar que sean causas directas de ello.

#### 4.1.2. Recorte analítico según la viabilidad

Este trabajo opta por focalizar el análisis sobre el **Top 5 de artistas** presentes en cada variable clave de este proyecto: las cinco figuras más reconocidas por el Ranking 2020 de Rolling Stone, y los cinco artistas más consagrados por el Índice de Popularidad de Spotify. La decisión responde a la necesidad de acotar el universo de estudio para lograr un abordaje más profundo y alineado con los objetivos del proyecto. Se considera que los cinco primeros puestos de cada métrica representan, de forma significativa, los criterios de reconocimiento que poseen tanto la crítica como el público a la hora de valorar a un artista por sobre otro.

Si bien se reconoce que este recorte implica una limitación en términos de representatividad, se lo entiende como una estrategia metodológicamente pertinente para explorar contrastes relevantes entre los valores de la crítica y los de la audiencia. Como posible línea futura de análisis, se propone una ampliación del universo observado (por ejemplo, al top 50), con el fin de fortalecer las conclusiones mediante el reconocimiento de patrones más generalizables.

#### 4.1.3 Ampliación del enfoque de KPI's

Aunque los KPI's (Indicadores Clave de Rendimiento) suelen utilizarse para medir el cumplimiento de objetivos específicos generalmente asociados a performance, impacto o evolución cuantificable —como popularidad, ventas o alcance—, en este proyecto se decidió ampliar dicho enfoque a **indicadores más descriptivos**, como el género, la edad y la trayectoria. Estas métricas, si bien no miden directamente logros o resultados, permiten observar tendencias relevantes en torno a los perfiles de los artistas que acceden a mayores niveles de reconocimiento.

Esta decisión metodológica se fundamenta en la voluntad de enriquecer el análisis **incorporando dimensiones socio-culturales** que contribuyan a interpretar el fenómeno de la canonización y la consagración musical desde una **perspectiva más integral**. Esta ampliación permite trascender una lectura puramente cuantitativa de los rankings, incorporando una mirada crítica orientada a comprender por qué ciertos perfiles artísticos alcanzan el reconocimiento por parte de la crítica o del público.

Asimismo, se consideró relevante incluir atributos como el género musical, la edad del artista o el año de debut, en tanto se trata de características que permiten observar posibles patrones o sesgos en los procesos de consagración crítica y popular. Estas dimensiones aportan una perspectiva

analítica que trasciende el mérito artístico en sí, habilitando la exploración de perfiles que podrían haber sido históricamente privilegiados o relegados en los rankings de legitimación musical.

#### 4.1.4 Marco interpretativo del Índice de Popularidad de Spotify

Si bien el dataset no especifica la fecha exacta de extracción del Índice de Popularidad de Spotify, se parte del supuesto de que fue relevado en un momento contemporáneo al armado del conjunto de datos por parte de los autores. A su vez, se asume que los datos no son posteriores a 2020, dado que dicho año representa el límite cronológico del análisis.

Por otro lado, cabe señalar que Spotify se consolidó como la plataforma dominante del consumo musical global desde principios de la década de 2010, alcanzando plena madurez y representatividad hacia 2020. En este marco, resulta razonable interpretar que el índice de popularidad incluido en el dataset fue recolectado en un punto cercano a esa fecha, garantizando así una coherencia temporal entre las distintas **fuentes de información consideradas en el dataset**.

#### 4.1.5. Priorización de KPI's

Para reconocer a los artistas más valorados por la audiencia en la era del streaming, se decidió trabajar exclusivamente con el **índice de popularidad provisto por Spotify**<sup>3</sup>, evitando la incorporación de otras métricas posibles incluidas en el dataset, como lo eran *Weeks con Billboard*<sup>4</sup> o *Peak Billboard Position*<sup>5</sup>. Esta decisión estuvo fundamentada en dos motivos principales.

En primer lugar, se priorizó contar con una métrica de alcance verdaderamente **global** y representativa del consumo musical contemporáneo a escala mundial, lo cual Billboard —centrado en el mercado estadounidense— no garantiza de la misma manera.

En segundo lugar, se buscó una métrica que reflejara una “fotografía” lo más cercana posible al momento temporal de corte del dataset (2020). A diferencia de Billboard, cuyas métricas acumulativas no permiten distinguir el período exacto de rendimiento de un álbum (y que en muchos casos remiten a décadas pasadas), el índice de popularidad de Spotify ofrece una **medición más actual y contextualizada** dentro de la era del streaming. Por ello, se consideró que esta métrica permitía captar con mayor precisión la valoración contemporánea de los artistas según la audiencia global.

#### 4.1.6 Criterios para la selección de indicadores temporales

---

<sup>3</sup> El Índice de Popularidad de *Spotify* refiere a una puntuación de 0 a 100 que estima cuán popular es un artista en relación con otros, en función del nivel de escucha registrado en la plataforma.

<sup>4</sup> Este indicador refiere a la cantidad de semanas que un álbum musical permaneció en el ranking de *Billboard*, reconocida revista estadounidense que publica semanalmente listados de popularidad de álbumes, canciones y artistas en la industria musical.

<sup>5</sup> Este indicador representa la posición más alta que un álbum alcanzó en el ranking de Billboard durante su permanencia en la lista.

Dentro del conjunto de datos original se disponía de múltiples variables temporales vinculadas a los artistas, pero se optó por utilizar solo aquellas que aportaran **valor concreto al análisis**, priorizando la claridad y la coherencia del enfoque.

En particular, se descartó el indicador *Suma de los años de nacimiento de los integrantes de la banda o artista*<sup>6</sup>, ya que su definición resultaba difícil de interpretar sin realizar cálculos adicionales que permitieran derivar una información más clara, como la edad del artista. En su lugar, se optó por utilizar ***Edad promedio al ingresar al Ranking de los 500 mejores álbumes***<sup>7</sup>, una variable más directa y expresiva para los fines del análisis. Por otro lado, se dejó fuera *Año de lanzamiento del álbum debut*<sup>8</sup>, dado que ***Años transcurridos desde el debut hasta el ingreso al Ranking***<sup>9</sup> resultaba ser un indicador más completo que el anterior, y representativo con respecto al recorrido profesional de cada artista hasta alcanzar un alto grado de reconocimiento por parte de la crítica.

Finalmente, interesa mencionar que esta decisión respondió a un lineamiento metodológico clave: es preferible trabajar con un **número acotado de indicadores bien seleccionados**, que sostengan y fundamenten el análisis, antes que incorporar múltiples métricas cuya abundancia pueda dificultar la interpretación y comunicación efectiva de los resultados.

#### 4.1.7 Interpretación de las variables dependientes

Aunque la definición detallada de variables e indicadores será desarrollada en la sección *Comprensión de los Datos*, resulta pertinente anticipar una distinción conceptual clave que orienta este trabajo: la diferenciación entre el reconocimiento que otorga la crítica especializada y aquel que proviene de la audiencia popular respecto a los artistas musicales.

En este sentido, la primera variable dependiente —el nivel de **canonización** de un artista por parte de la crítica— será entendida como el reconocimiento simbólico que la crítica especializada otorga a ciertos artistas mediante su inclusión y posicionamiento destacado en los rankings históricos elaborados por la revista Rolling Stone.

Por su parte, la segunda variable dependiente —el nivel de **consagración** de un artista por parte del público— será interpretada como el reconocimiento simbólico proveniente de la audiencia, reflejado en el consumo de sus producciones musicales, y observable a través de métricas como el Índice de Popularidad de Spotify.

#### 4.2. Comprensión del Negocio

El presente proyecto busca responder si existe coincidencia entre las valoraciones de la crítica especializada y las preferencias de la audiencia popular con respecto a quiénes son los grandes artistas en la industria musical. La información resultante podría ser útil tanto para **actores del sector**

---

<sup>6</sup> Corresponde a la suma de los años de nacimiento de los miembros del grupo o artista.

<sup>7</sup> Refleja la edad promedio que tenían los miembros del grupo o artista al momento en que su álbum ingresó al ranking de los 500 mejores de Rolling Stone.

<sup>8</sup> Indica el año en que el artista o grupo lanzó su álbum debut.

<sup>9</sup> Refiere la cantidad de años transcurridos entre el lanzamiento del álbum debut del artista o grupo y el momento en que uno de sus álbumes fue incluido en el ranking de los 500 mejores de Rolling Stone.

—en decisiones estratégicas, de marketing o descubrimiento de talentos— como para el **campo cultural**, al ofrecer nuevas lecturas sobre cómo se construye el valor simbólico en un entorno digital. Además, representa una aplicación concreta del **análisis de datos** en un sector en plena transformación.

#### 4.3. Comprensión de los Datos

El análisis del presente proyecto se basa en el dataset *Rolling Stone Album Rankings*, publicado en el repositorio de GitHub como parte del desafío semanal de #TidyTuesday, el 7 de mayo de 2024, por los usuarios @bradfordjohnson y @jonthegeek. Este conjunto de datos fue originalmente desarrollado por el equipo de **The Pudding** para un ensayo visual que explora qué convierte a un álbum en “el mejor de todos los tiempos”, tomando como punto de partida los rankings históricos publicados por la revista *Rolling Stone* en los años 2003, 2012 y 2020.

Según se detalla en la nota metodológica del proyecto, el dataset surgió inicialmente a partir de una hoja de cálculo creada por Chris Eckert, posteriormente evolucionando a un dataset propio desarrollado por los autores. Entre las variables incluidas en su base de datos se encuentran: el nombre del álbum, el género musical, el año de lanzamiento, la posición que ocupa en los rankings de las tres ediciones, el nombre del artista, el año de nacimiento, el género (sexo), los votantes de cada edición, y otros atributos asociados.

Para responder la pregunta de investigación, se requieren datos que representen tanto la **valoración de la crítica especializada** como la **percepción de la audiencia popular**. El dataset incluye las posiciones de los álbumes en tres ediciones históricas del ranking de Rolling Stone, lo que permite construir análisis sobre la canonización de la crítica. A su vez, contiene datos extraídos de indicadores de popularidad en plataformas digitales, permitiendo evaluar el reconocimiento de la audiencia sobre los artistas. Por último, también incorpora **información sobre artistas y álbumes** que será cruzada con los indicadores mencionados, lo cual habilita el análisis comparado que se propone. De este modo, el dataset contiene los tipos de datos necesarios para desarrollar el análisis planteado.

En cuanto a la disponibilidad y condiciones de uso del dataset, éste se encuentra disponible públicamente en el repositorio de GitHub de Tidy Tuesday, bajo licencia **CC0 1.0 (Creative Commons Zero)**, lo que implica que puede ser reutilizado libremente con fines educativos o analíticos. Como ya fue mencionado, fue originalmente desarrollado por el equipo de The Pudding como parte de un ensayo visual interactivo, y posteriormente adaptado<sup>10</sup> y compartido por la comunidad de datos de Tidy Tuesday, manteniendo su accesibilidad y apertura.

#### 4.4. Construcción de Variables y KPI's<sup>11</sup>

En esta sección se detallan las **variables** clave del proyecto, así como los **indicadores** seleccionados para su análisis. Tanto variables como indicadores fueron definidos en función de los

---

<sup>10</sup> Posteriormente, en el apartado correspondiente a *Gobernanza de Datos*, se incluirá un comentario detallado sobre las adaptaciones realizadas por los usuarios de la comunidad de GitHub #TidyTuesday en relación al dataset original.

<sup>11</sup> Key Performance Indicators (indicadores clave de rendimiento).



objetivos específicos y de la pregunta de investigación, buscando capturar la perspectiva de la crítica especializada y, al mismo tiempo, la valoración del público en la era del streaming. Se incluyen también **indicadores descriptivos** que, aunque no constituyen KPIs centrales, amplían la comprensión del fenómeno analizado. Es por ello que se las entiende como parte de un *enfoque ampliado de KPI's*, como fue mencionado en el apartado anterior.

A continuación, se presenta una tabla-resumen de las variables principales del presente proyecto y sus KPIs asociados. La versión completa y detallada se encuentra disponible a través del enlace al final de la tabla.

Variable independiente	Variable dependiente	KPIs asociados	Indicadores descriptivos asociados
Consideración de la crítica especializada	Nivel de canonización de un artista por la crítica	Ranking 2020	Género musical del álbum
			Género del artista
			Edad promedio al ingresar al ranking de los 500 mejores álbumes
			Años transcurridos desde el debut hasta el ingreso al ranking
Valoración de la audiencia popular	Nivel de consagración de un artista por parte del público	Índice de popularidad en Spotify	Género musical del álbum
			Género del artista
			Edad promedio al ingresar al ranking de los 500 mejores álbumes
			Años transcurridos desde el debut hasta el ingreso al ranking

Para acceder a la **versión completa y detallada de esta tabla**, haga [click aquí](#).

#### 4.5. Gobernanza de Datos: decisiones clave para la calidad y el análisis

El presente proyecto tiene como principio rector garantizar las **buenas prácticas que orientan a la gobernanza de datos** en el proceso de lectura, procesamiento, limpieza, análisis, visualización e interpretación de los mismos. En este sentido, es un objetivo de este trabajo asegurar la **calidad, consistencia, integridad y trazabilidad** en el manejo de la información utilizada en el análisis.

A fines de cumplimentar con estos principios, en los apartados siguientes se desarrollan distintos aspectos clave que hacen a una buena gobernanza de datos en el marco de este trabajo. Se abordará, de manera integral, tanto el origen y tratamiento de la información como las decisiones metodológicas adoptadas para garantizar su correcta gestión a lo largo de todo el proyecto.

##### **4.5.1. Origen de los datos**

Este proyecto se construye sobre un **único dataset principal**, desarrollado por The Pudding a partir de los rankings históricos de Rolling Stone (2003, 2012 y 2020) y datos provenientes del Índice

de Popularidad de Spotify. Como fue mencionado, fue extraído del repositorio de **GitHub**, en donde se publicó como parte del desafío semanal de #TidyTuesday el 7 de mayo de 2024, por los usuarios @bradfordjohnson y @jonthegeek. El dataset se encuentra disponible públicamente bajo **licencia CC0 1.0 (Creative Commons Zero)**.

#### 4.5.1.1 Decisiones previas en la limpieza del dataset original

El dataset publicado en GitHub se encontraba acompañado de un **script de limpieza** que incluyó:

- Cambios en los nombres de columnas, reemplazando los espacios en blanco por guiones bajos (*underscores*) y convirtiendo todas las letras a minúsculas (*lowercase*).
- Eliminación de las columnas *Chartmetric Link* y *Album ID Quoted*.
- Reemplazo de los valores *N/A*, *Not on Spotify* y *"-"* por celdas vacías.

#### 4.5.6. Responsabilidad en el manejo de datos

La limpieza, procesamiento, análisis y visualización del dataset estará enteramente a cargo de una sola persona: **Camila Bozzoletti**, autora del presente trabajo. De este modo, se garantiza un criterio único y consistente en la toma de decisiones técnicas a lo largo del proyecto.

Si bien el dataset es de acceso público, durante el desarrollo del análisis solo la autora tendrá acceso directo al archivo de trabajo. Todas las transformaciones realizadas sobre los datos serán documentadas con transparencia (a partir del presente escrito, en el documento de Google Colab y en el archivo de visualización) respetando el sentido original de la información e incluyendo justificaciones para cada decisión tomada.

#### 4.5.7. Herramientas utilizadas

El proceso de limpieza y preparación de los datos, así como el análisis exploratorio (EDA), se llevará a cabo mediante el lenguaje de programación **Python**, utilizando la plataforma **Google Colab**. Una vez completada esta etapa, se procederá a la construcción de un **dashboard en Looker Studio**, con el objetivo de visualizar de manera clara y accesible las principales conclusiones del análisis.

Se adjuntará al presente trabajo el enlace al documento de Google Colab utilizado, de forma que cualquier persona interesada pueda realizar un seguimiento detallado de las acciones efectuadas. Esta decisión se enmarca en un principio de **transparencia y reproducibilidad**, que busca no solo garantizar la trazabilidad del proceso, sino también habilitar su potencial reutilización con fines formativos, académicos o exploratorios.

#### 4.5.8. Documentación de las transformaciones realizadas

Las transformaciones aplicadas a los datos se llevaron a cabo durante la etapa de *Transform* dentro del enfoque ETL (Extract, Transform, Load), en el marco de la fase de preparación de datos. Para preservar la integridad del dataset original, todas las modificaciones se realizaron sobre una **copia** del mismo, almacenada en una variable distinta denominada *df\_copia*.

La **etapa de transformación** incluyó la aplicación de diversas decisiones metodológicas orientadas a garantizar la calidad y consistencia del conjunto de datos. Entre las acciones implementadas se destacan: la **eliminación de columnas no relevantes**, el **ajuste de nombres de columnas**, el **tratamiento de valores nulos y duplicados**, la **conversión de tipos de datos**, la **validación de la integración interna entre los datos de Rolling Stone y Spotify**, y la **verificación de las variables, KPIs e indicadores descriptivos** definidos previamente.

Con el objetivo de asegurar la calidad y la transparencia del análisis, se creó una **nueva versión del dataset original** (*df\_filtrado\_eda*) que contiene únicamente las **filas con valores completos** en las variables clave del estudio. Esta decisión responde a principios éticos fundamentales del presente trabajo, que indican que, para realizar análisis comparativos, es necesario contar con datos completos en todas las columnas involucradas en el análisis. La **única excepción** a esta regla fue la variable **indice\_popularidad\_spotify**, ya que contiene valores como "Not on Spotify" que, si bien no son numéricos, aportan información valiosa sobre la presencia o ausencia de ciertos artistas en esta plataforma, y las implicancias que esto conlleva. Eliminar estos casos implicaría perder información clave para la obtención de insights, por lo que se decidió conservarlos deliberadamente.

El objetivo de estas transformaciones fue asegurar un dataset limpio, estructurado y confiable para el análisis posterior. El detalle completo de los procesos realizados puede consultarse en el apartado 4.7.2.2. de este documento.

#### 4.5.9. Política de Calidad de Datos

Se implementaron decisiones metodológicas orientadas a **garantizar la calidad del dataset**, incluyendo el tratamiento ético y justificado de los valores faltantes, la verificación de duplicados y la estandarización de los nombres de columnas para mejorar su comprensión. A pesar de que el dataset original ya presentaba un buen nivel de limpieza, se documentaron y corrigieron inconsistencias relevantes, asegurando un conjunto de datos ordenado, transparente y adecuado para el análisis.

#### 4.5.10. Visualización de los datos

La etapa de visualización tuvo como propósito central representar de manera **clara y accesible** las conclusiones alcanzadas a partir del análisis. Para ello, se elaboró un **dashboard interactivo** en Looker Studio que permite explorar los resultados y facilitar su interpretación.

Cada decisión tomada durante esta instancia respondió al objetivo de reflejar, de forma íntegra y coherente, los hallazgos obtenidos a lo largo del proceso analítico. En este sentido, se procuró garantizar una **comunicación efectiva** de la información, precisando las **fuentes** utilizadas y cuidando que la forma de exhibir los datos no distorsione su significado ni su contexto.

### 4.6. Ética de Datos

Este proyecto se guía por principios éticos orientados al **uso responsable de los datos**. El dataset utilizado es de acceso público y contiene información sobre artistas y álbumes musicales de carácter público. No se incluyen datos personales ni sensibles, y todas las figuras representadas son

de conocimiento general. A continuación, se destacan los aspectos más relevantes relacionados con la **gestión ética de los datos** utilizados en el presente trabajo.

- **Manejo ético de los datos de popularidad y rankings:** se asegura que el manejo de las puntuaciones vinculadas a los artistas y sus posiciones en los rankings se realizan en el marco de una total transparencia y respeto, sin ejercer manipulación de los datos ni filtrados arbitrarios. Como fue mencionado, todas las transformaciones realizadas en el dataset serán documentadas detalladamente, sin alterar el sentido original de la información.
- **Reconocimiento de limitaciones del dataset:** el dataset utilizado presenta ciertas limitaciones que es importante considerar a la hora de interpretar los resultados. Por un lado, no todos los artistas incluidos en los rankings cuentan con información disponible en Spotify, lo cual puede generar un sesgo a favor de aquellos con mayor presencia en la plataforma. Además, la selección de artistas realizada por Rolling Stone responde a decisiones que pueden (potencialmente) estar influidas por sesgos culturales, geográficos y/o de género. A esto se suma que los datos alcanzan sólo hasta el año 2020, por lo que no contemplan transformaciones recientes en la industria musical ni en las dinámicas de popularidad actuales.
- **Justificación de decisiones metodológicas:** las decisiones tomadas en relación con las métricas elegidas y las transformaciones realizadas responden a criterios estrictamente metodológicos, sin buscar influir a favor ni en contra de determinados artistas o géneros. Por el contrario, se procuró seleccionar indicadores y diseñar tareas de preparación de datos que favorecieran la producción de conclusiones transparentes y representativas de lo que el dataset expresa.
- **Reflexión sobre representación de géneros, razas o regiones:** para aportar al análisis reflexivo, se procuró mantener una mirada crítica respecto del grado de representación de géneros, razas y regiones en los resultados. Esto con la finalidad de analizar si la canonización musical puede estar atravesada por sesgos estructurales o desigualdades históricas que condicionan a quiénes logran mayor visibilidad o reconocimiento. Es importante reflexionar sobre cómo estos factores pueden influir en la percepción pública y crítica de los artistas.
- **Observaciones sobre outliers:** en esta etapa inicial del análisis, aún no se ha realizado una detección técnica exhaustiva de valores atípicos o outliers. No obstante, de forma preliminar, se observó que varias columnas del dataset presentan una desviación estándar elevada, lo cual sugiere la probable presencia de outliers en la fase de EDA. Este fenómeno es comprensible dada la naturaleza del dataset, que incluye rankings con posiciones de artistas que oscilan entre el 1 y el 500.

- **Tratamiento de valores faltantes**<sup>12</sup>: dado el elevado número de valores faltantes en varias columnas, se optó por imputaciones limitadas, como el uso de la categoría 'desconocido' en ciertos casos. En otras columnas, se decidió conservar los valores nulos, ya que su ausencia no impactaba negativamente en los análisis posteriores. De hecho, eliminarlos hubiera supuesto una pérdida significativa de datos (aproximadamente el 30% del total). Para asegurar la calidad y transparencia del análisis, se creó una versión del dataset (*df\_filtrado\_eda*) con únicamente las filas completas en las variables clave, lo cual es esencial para los análisis comparativos. La excepción fue la variable *indice\_popularidad\_spotify*, que conserva valores como 'Not on Spotify' debido a que estos contienen información valiosa sobre la presencia de los artistas en la plataforma y no deben ser descartados.

Este enfoque buscó garantizar que el dataset se mantuviera **representativo** y no distorsionado, a la vez que se evitaba el riesgo de introducir información errónea mediante la imputación, ya que rellenar un dataset con una desviación estándar elevada en varias de sus columnas probablemente no hubiera resultado representativo.

Para concluir el presente apartado, interesa mencionar que todas las decisiones tomadas durante el desarrollo del proyecto responden a criterios éticos orientados a la **transparencia**, la **responsabilidad técnica** y el **respeto por los sujetos** representados en los datos.

#### 4.7. Preparación de datos

Esta etapa tuvo por objetivo obtener un conjunto de datos **depurado**, **organizado** y en condiciones de ser **analizado** para responder la pregunta de investigación planteada sobre el grado de coincidencia entre la crítica especializada y la audiencia popular en la definición de artistas reconocidos.

Para llevar adelante el procesamiento de los datos, se optó por aplicar el **enfoque ETL** (Extract, Transform, Load), ampliamente recomendado cuando se trabaja con datos estructurados y se busca generar visualizaciones, ya que requiere que los datos estén previamente procesados antes de ser cargados.

Brevemente, el enfoque ETL consiste en tres fases:

- **Extract**: lectura y extracción de los datos desde su fuente original.
- **Transform**: limpieza y transformación de esos datos en bruto mediante herramientas como Python, para dejarlos en un formato ordenado y utilizable.
- **Load**: carga de los datos ya tratados a un sistema final —como una base de datos o una herramienta de visualización— donde puedan ser analizados o utilizados por otros servicios.

---

<sup>12</sup> Para obtener más detalles sobre el tratamiento de los valores faltantes, se puede consultar el apartado 4.7.2.2.2. de la sección 'Preparación de Datos' de este informe.

Este apartado detalla el proceso técnico de preparación de los datos. Se explican las transformaciones realizadas sobre el dataset, las herramientas utilizadas y las decisiones metodológicas clave adoptadas a lo largo del proceso.

#### 4.7.1. Ejecución del proceso ETL: Notebook en Google Colab

A continuación, se presenta el **enlace al notebook** desarrollado en **Google Colab**, donde se implementó el proceso ETL correspondiente al proyecto. Allí se detalla el procedimiento completo de extracción, transformación y carga de los datos, junto con las decisiones técnicas adoptadas en cada etapa y su fundamentación.

Para revisar el proceso técnico en detalle, se invita a consultar el notebook del proyecto en Google Colab a través del siguiente [enlace](#).

#### 4.7.2. Documentación de la fase de preparación de datos: proceso ETL

En esta sección se **documenta** el proceso de preparación de los datos, llevado a cabo a través del enfoque ETL (Extract, Transform, Load). Se detallan las acciones realizadas en cada fase: desde la **extracción de los datos**, pasando por su **limpieza y transformación**, hasta su **carga** en un formato adecuado para el análisis. Se explica cómo se manejaron los valores nulos, duplicados y otras transformaciones clave, así como la exportación final del dataset limpio.

##### 4.7.2.1. EXTRACT: exportación y lectura de datos

En esta etapa, se realizó la carga y lectura del archivo CSV original para iniciar el proceso de preparación de los datos. A continuación, se detallan las acciones ejecutadas:

1. **Carga y visualización del archivo CSV:** se importó el dataset original en formato CSV para su análisis posterior.
2. **Visualización y verificación inicial del DataFrame:** se configuró la visualización completa de las columnas del DataFrame, permitiendo observar todas las variables incluidas en el dataset original. Posteriormente, se verificó la estructura del DataFrame mediante la visualización de las primeras y últimas 5 filas y, a través de la función `df.shape`, se confirmó que el dataset constaba de 691 filas, correspondientes a los artistas presentes en las tres ediciones del ranking de Rolling Stone, y 21 columnas. Estas columnas representaban las variables definidas por los creadores del dataset para analizar qué hace a un álbum el mejor de todos los tiempos, según el ensayo visual "What Makes an Album The Greatest of All Times?" de The Pudding. Entre las variables clave para el presente trabajo, que forman parte del dataset original, se incluyen: los nombres de los artistas y sus respectivos álbumes, el ranking histórico del año 2020 de Rolling Stone, el género musical del álbum, el Índice de Popularidad de Spotify, el género del artista, la edad promedio del artista al ingresar al Top 500, y los años transcurridos entre su debut y su ingreso al Top 500.

3. **Comprobación de los tipos de datos:** se registraron algunas inconsistencias en los tipos de datos que requirieron transformación en la siguiente fase. Estas inconsistencias se detallarán en la sección 4.7.2.2. relativa a la fase de transformación.
4. **Detección de valores nulos por columna:** se identificaron las columnas con valores nulos para anticipar el tratamiento que requerirán en la fase de transformación. Sobre un total de 691 entradas, se observó una gran cantidad de valores nulos en las tres columnas correspondientes a los rankings históricos (191 en cada una de las ediciones de 2003, 2012 y 2020), y en la columna de género musical del álbum (164). Aunque las ediciones 2003 y 2012 de los rankings no formen parte del recorte analítico del proyecto, la información restante sobre valores faltantes será considerada en la etapa de transformación. Además, la columna de semanas en el ranking Billboard (*weeks\_on\_billboard*) presentó 119 valores faltantes, pero tampoco será utilizada en este análisis.
5. **Resumen estadístico de variables numéricas:** se generó un resumen estadístico inicial de las variables numéricas para examinar la distribución, dispersión y posibles valores atípicos. Se observó que la **desviación estándar era particularmente alta** en las columnas *rank\_2003*, *rank\_2012* y *rank\_2020*. Esta elevada dispersión no parece ser consecuencia de la presencia de outliers, sino que responde a la naturaleza del ranking, cuyo rango de posiciones varía entre 1 y 500. En este contexto, el elevado nivel de dispersión sugiere que el **promedio no sería una métrica adecuada** para el análisis, ya que podría no reflejar de manera precisa la realidad de los datos.
6. **Revisión de valores únicos y detección de duplicados:** Se realizaron verificaciones de los valores únicos en columnas clave para el análisis, como el nombre del artista (*clean\_name*), el género musical del álbum (*genre*) y el género del artista (*artist\_gender*). Esta revisión **no reveló inconsistencias ni duplicados a nivel de columna**, pero permitió obtener una visión preliminar de las características de las variables. Además, se verificaron registros duplicados a nivel de fila completa, y aunque era esperado que un mismo artista apareciera varias veces por distintos álbumes, **no se detectaron duplicados exactos** en el dataset.
7. **Visión general del DataFrame:** finalmente, se realizó una inspección general de las características del DataFrame para determinar ajustes adicionales y la necesidad de tratamiento de valores faltantes, los cuales se llevarán a cabo en la fase de transformación.

#### 4.7.2.2. TRANSFORM: limpieza, evaluación y transformación de los datos

En esta sección, se detalla el proceso de **limpieza, evaluación y transformación de los datos** utilizados en el análisis. Se describirán las decisiones tomadas respecto a la eliminación de columnas innecesarias, el tratamiento de valores nulos y duplicados, y las conversiones de tipo de datos realizadas para asegurar la integridad y coherencia del dataset. Además, se validará la correcta integración de los datos de Rolling Stone y Spotify, y se explicará cómo se gestionaron los KPI's e indicadores descriptivos para garantizar la calidad de la información que alimenta el análisis posterior.

##### 4.7.2.2.1. Limpieza de datos

El proceso de limpieza del dataset involucró una serie de decisiones metodológicas que tuvieron por objetivo producir una versión de dataset **limpio, claro, y fácil de visualizar** para el proceso de análisis en desarrollo. A su vez, en línea con los principios de gobernanza de datos adoptados en este trabajo, todos los cambios fueron aplicados sobre una **copia del dataset original**, almacenada en una variable diferente denominada **df\_copia**. Esta decisión permitió preservar la integridad del conjunto de datos original, asegurando su disponibilidad para futuras consultas o validaciones.

Las decisiones ejecutadas en la etapa de limpieza de datos se mencionan a continuación:

1. **Eliminación de columnas innecesarias:** se tomó la decisión de eliminar varias columnas del dataset original debido a que no resultaban esenciales para el enfoque de este proyecto. Esta depuración no solo permitió trabajar con un conjunto de datos más ajustado a los objetivos del análisis, sino que también favoreció a una visualización más clara y ordenada de la información utilizada. La eliminación se basó en los siguientes criterios:
  - **Columnas redundantes:** aquellas que repetían información ya contenida en otras variables conservadas.
  - **Columnas no relacionadas con los objetivos del análisis:** columnas que no forman parte de los KPIs ni de los indicadores descriptivos relevantes.
  - **Columnas poco vinculadas a la pregunta de investigación:** datos que, si bien pueden ser de interés en otros contextos, no contribuyen directamente a responder la pregunta de investigación de este proyecto.

De esta manera, las columnas eliminadas fueron: nombre del artista por ordenamiento alfabético ('*sort\_name*'), ranking 2003 ('*rank\_2003*'), ranking 2012 ('*rank\_2012*'), diferencial 2003-2020 que indica la variación de posición de los álbumes entre ambas ediciones del ranking de Rolling Stone ('*differential*'), año de lanzamiento del álbum ('*release\_year*'), tipo de álbum ('*type*'), semanas en el ranking Billboard ('*weeks\_on\_billboard*'), posición máxima en el ranking Billboard ('*peak\_billboard\_position*'), URL del álbum en Spotify ('*spotify\_url*'), cantidad de miembros del artista o banda ('*artist\_member\_count*'), suma de los años de nacimiento de los miembros del artista o banda ('*artist\_birth\_year\_sum*'), año de lanzamiento del álbum debut ('*debut\_album\_release\_year*'), ID del álbum ('*album\_id*').

En consecuencia, el dataset resultante —que constituye la versión final y consolidada sobre la cual se desarrolló todo el análisis de este proyecto— contiene las siguientes columnas:

- **Nombre del artista** ('*nombre\_artista*')
- **Álbum del artista** ('*album*')
- **Ranking histórico 2020** ('*rank\_2020*')
- **Género musical del álbum** ('*genero\_musical\_album*')
- **Índice de popularidad de Spotify** ('*indice\_popularidad\_spotify*')
- **Género del artista** ('*genero\_artista*')



- **Edad promedio del artista al ingresar al top 500** ('edad\_promedio\_ingreso\_ranking')
- **Años transcurridos entre su debut y su ingreso al top 500** ('años\_entre\_debut\_y\_album\_top500')

Por último, en esta etapa se decidió **ajustar los nombres de algunas columnas**, asignándoles nombres más descriptivos y traduciéndolos al español, con el fin de facilitar su comprensión durante el análisis.

#### 4.7.2.2.2. Evaluación y ajuste de calidad de los datos

La etapa de evaluación de datos fue crucial, ya que concentró diversas decisiones metodológicas clave, muchas de las cuales tendrían un **impacto directo en la interpretación de los datos** y en la coherencia del análisis posterior. En particular, esta fase se centró en la gestión de los valores faltantes, la conversión de tipos de datos en algunas columnas del dataset y la verificación de la integración interna de los datos.

Dado que las variables ya habían sido definidas previamente, no fue necesario realizar ajustes adicionales en esta etapa. En cuanto a los indicadores, se evaluó su integridad y se concluyó en que los indicadores **cuantitativos** (KPI's) ya estaban presentes en el dataset, por lo que no fue necesario calcularlos. Los indicadores **descriptivos**, por su parte, son variables categóricas que se utilizarán para análisis posteriores, por lo que tampoco requirieron de cálculos adicionales.

De esta manera, en lo sucesivo se detallan las resoluciones metodológicas adoptadas para esta etapa:

1. **Tratamiento de valores nulos (NaN):** tal y como se destacó previamente en la fase de extracción, se identificaron varias columnas con valores nulos que requirieron ser gestionados en esta etapa. En particular, las tres columnas correspondientes a los rankings históricos presentan una cantidad considerable de valores nulos (191 en cada una), mientras que la columna de género musical del álbum también exhibe una cantidad significativa de valores faltantes (164). Por su parte, la columna semanas en el ranking Billboard ('weeks\_on\_billboard') tiene 119 valores nulos; sin embargo, dado que no será utilizada en este análisis, no será objeto de tratamiento en esta etapa.

Dado el **alto número** de valores nulos y la **variabilidad de las causas de ausencia** de datos según la columna, la decisión sobre cómo abordarlos no solo resultaba clave, sino que también debía tener en cuenta las particularidades de cada caso. En consecuencia, se definió el siguiente tratamiento para los valores NaN:

- **NaN en la columna 'genero\_musical\_album':** se decidió imputar los valores nulos reemplazándolos por la categoría 'desconocido'. Esta elección busca favorecer la posterior visualización de los géneros musicales predominantes en la industria, evitando la pérdida de estos registros en los gráficos y análisis categóricos.
- **NaN en la columna del 'rank\_2020':** se decidió mantener los valores NaN en esta columna, y trabajar sobre el ranking de forma independiente. En análisis donde se necesiten

comparaciones entre variables, se filtrarán las filas que tengan valores completos en todas las columnas relevantes.

Motivos de la ausencia de valores en la columna mencionada<sup>13</sup>:

- *rank\_2020*: NA si el álbum no está en el top 500 de 2020.
- **NaN en la columna '*indice\_popularidad\_spotify*'**: se optó por conservar los valores nulos en esta columna, ya que su presencia no impide realizar análisis significativos sobre los artistas que sí cuentan con este dato, y eliminar dichas filas implicaría perder información valiosa presente en otras variables. Según el diccionario de datos provisto por la fuente original del dataset, la ausencia de este valor se debe a que algunos registros no se encontraban disponibles en Spotify al momento de la recolección, ya sea porque el álbum no estaba cargado en la plataforma, el artista no figuraba en ella, o porque no fue posible recuperar la información sobre su popularidad. En cualquiera de los casos, la consecuencia es la misma: no se dispone del dato de popularidad, por lo que se decidió mantener el valor nulo en esta etapa del análisis.
- **NaN en la columna '*genero\_artista*'**: se imputaron los 5 valores faltantes con la categoría '*desconocido*', a fines de incluir estos casos en el análisis de género sin alterar la representatividad de la muestra.
- **NaN en las columnas '*edad\_promedio\_ingreso\_ranking*' y '*años\_entre\_debut\_y\_album\_top500*'**: se decidió eliminar 5 filas que presentaban valores faltantes en ambas columnas. Dado que representaban una proporción muy pequeña del total y carecían de información clave para el análisis, se optó por su eliminación con el fin de simplificar el análisis sin afectar la calidad del conjunto de datos.

Al finalizar el proceso de limpieza, se realizó una **revisión de los resultados de las estrategias de tratamiento de NaN** para asegurar que las variables esenciales estuvieran completas para el análisis.

2. **Tratamiento de valores duplicados**: si bien este aspecto ya fue revisado en la fase de extracción, se retoma brevemente aquí por corresponder también a esta etapa del proceso. En ese momento, se verificaron posibles duplicados tanto a nivel de columnas clave (como nombre del artista, género musical y género del artista) como de filas completas. No se detectaron inconsistencias ni registros idénticos, por lo que no fue necesario aplicar un tratamiento adicional en esta instancia.

### 3. **Conversión de tipo de dato:**

En esta etapa, se identificaron inconsistencias que requirieron **ajustes en los tipos de datos**. Las columnas correspondientes a la edición 2020 del ranking de Rolling Stone y a la de años

---

<sup>13</sup> La información sobre los motivos de las ausencias fue obtenida del diccionario de datos disponible en la página de GitHub de la fuente del dataset.

transcurridos entre el debut y el ingreso al Top 500, contenían valores en formato '**float**' cuando deberían ser '**int**'.

Algunas columnas con inconsistencias en sus tipos de datos no fueron transformadas, ya que no formaban parte de las variables, indicadores ni KPIs seleccionados para el análisis. Estas fueron finalmente eliminadas en la etapa de transformación.

Con base en lo anterior, se procedió de la siguiente manera:

Se ajustó el tipo de dato de las columnas correspondientes para garantizar su visualización en el formato adecuado. Las columnas numéricas que originalmente estaban en formato 'float' y fueron convertidas a 'int' fueron:

- rank\_2020
- años\_entre\_debut\_y\_album\_top500

Por último, se decidió mantener en formato float la columna `indice_popularidad_spotify`, con el objetivo inicial de evitar que los valores 'Not on Spotify' se transformaran en NaN. Sin embargo, al mantenerse como NaN de todas formas, se optó por conservar ese formato por cuestiones de consistencia. También se mantuvo en float la columna `edad_promedio_ingreso_ranking`, ya que refleja un valor detallado que incluye los meses.

#### 4. Validación de la Integración Interna de Datos (Rolling Stone y Spotify):

Se verificó que los datos provenientes de Rolling Stone y Spotify, ya previamente integrados en el dataset original, estuvieran correctamente asociados. Esta validación consistió en comprobar que los nombres de los artistas coincidieran entre las columnas de rankings de Rolling Stone y la columna del Índice de Popularidad de Spotify. Además, se estimó la pérdida de datos de la siguiente manera:

- **Artistas sin datos de popularidad en Spotify:** se identificaron **33 artistas** sin valores en la columna '`indice_popularidad_spotify`'.
- **Pérdida de datos estimada:** esta pérdida representa aproximadamente un **4.81% del total de artistas** (33 de 686). Aunque la pérdida es pequeña, se optó por utilizar solo las filas con datos completos en los análisis comparativos entre los rankings de Rolling Stone y la popularidad en Spotify, con el objetivo de evitar posibles sesgos en los resultados.

##### 4.7.2.3. LOAD: carga y exportación del dataset limpio

El dataset limpio representa la **versión final** del conjunto de datos, obtenido tras los procesos de limpieza y transformación, evaluación de calidad y verificación de KPI's e indicadores descriptivos. Este conjunto de datos depurado y estructurado contiene la información necesaria para el análisis, incluyendo el ranking histórico de Rolling Stone edición 2020, el Índice de Popularidad de Spotify y otros indicadores descriptivos relevantes.

#### 5. Análisis Exploratorio de Datos (EDA)

##### 5.1. Fases del EDA

En esta etapa, se llevó a cabo el Análisis Exploratorio de Datos (EDA) con el objetivo de profundizar en las variables clave del proyecto y obtener los primeros indicios que permitan abordar la pregunta de investigación. El enfoque principal se orientó a las siguientes variables: **rank\_2020** (canonización por parte de la crítica especializada) e **indice\_popularidad\_spotify** (consagración por parte del público), que constituyen el eje central de este trabajo.

El EDA se organizó en dos fases:

- **Fase 1: análisis individual de las dos variables** mencionadas para comprender su comportamiento, detectar posibles outliers, explorar la existencia de correlaciones y observar cómo la crítica y el público tienden a reconocer a los artistas.
- **Fase 2: caracterización de los artistas mejor posicionados tanto por la crítica como por la audiencia**, a partir del cruce de los top 5 de cada variable con los indicadores descriptivos relevantes del dataset (género musical del álbum, género del artista, edad promedio de ingreso al ranking y años transcurridos entre el debut y el ingreso al top 500).

#### 5.1.1. Fase 1: análisis individual de las variables clave

Para llevar a cabo este análisis, se realizó una **revisión estadística descriptiva** de cada variable con el fin de entender su distribución, identificar cómo se comportan de manera independiente y extraer conclusiones preliminares a partir de las medidas estadísticas obtenidas. La visualización de la distribución de cada variable se realizó mediante gráficos, como **histogramas** y **boxplots**, que resultaron fundamentales para detectar segmentaciones de datos y la presencia de outliers.

El análisis de la variable vinculada al índice de popularidad de Spotify requirió de una atención particular debido a su **distribución heterogénea**. A fines de profundizar su interpretación, se calculó el **rango intercuartílico (IQR)** para determinar dónde se encontraba la concentración del 50% de los artistas dentro de esta métrica. Además, se calcularon los extremos de los bigotes en el boxplot, aspecto que permitió identificar los valores “normales” de la métrica y establecer el umbral a partir del cual los valores serían considerados **outliers**.

Finalmente, se procedió a calcular la **correlación entre las variables**. Para garantizar el rigor ético del análisis comparativo, se filtraron las filas con valores completos en las columnas relevantes para este análisis. A continuación, se calculó la **correlación de Pearson**, cuyos resultados se visualizaron a través de un gráfico de dispersión (scatter plot), permitiendo identificar relaciones significativas entre las variables.

#### 5.1.2. Fase 2: análisis comparativo de los principales artistas reconocidos

Con el fin de asegurar la calidad y transparencia del análisis, se creó una nueva versión del dataset original (*df\_filtrado\_eda*) que solo incluyó las filas con valores completos en las variables clave del estudio. Esta decisión se alineó con los principios éticos del proyecto, que establecen que los análisis comparativos deben basarse en datos completos en todas las columnas relevantes. La única excepción a esta regla fue la variable **indice\_popularidad\_spotify**, que contiene valores como “Not on Spotify”. Estos valores, aunque no numéricos, ofrecen información valiosa sobre la presencia

o ausencia de artistas en la plataforma, lo cual es esencial para comprender el contexto del análisis. Por lo tanto, se decidió mantener estas observaciones en el dataset para preservar información clave.

Con el nuevo dataset filtrado, se identificaron los 5 artistas mejor posicionados según **Rolling Stone** y los 5 con mayor índice de popularidad en **Spotify**. A continuación, se cruzaron estas posiciones con los indicadores descriptivos establecidos: **género musical del álbum, género del artista, edad promedio de ingreso al ranking y años transcurridos entre el debut y el ingreso al top 500**. Este cruce permitió obtener una visión más clara de las características comunes entre los artistas reconocidos por la crítica y la audiencia.

A partir de este análisis, se generaron los principales insights que buscaron responder a los objetivos que orientaron este proyecto. Estos insights serán presentados en el próximo apartado, donde se detallarán las conclusiones y se propondrán posibles líneas de análisis futuro.

## 6. Principales Insights

### 6.1. **Recordatorio acerca del marco analítico del proyecto**

En este proyecto se cruzan los datos del ranking 2020 de Rolling Stone, que clasifica a los artistas según la crítica especializada, con el índice de popularidad de Spotify, que refleja la valoración de los artistas por parte de la audiencia. Es importante destacar que, aunque el dataset no especifica la fecha exacta de extracción del índice de popularidad, se parte del supuesto de que estos datos fueron relevados en un momento cercano a la creación del conjunto de datos por parte de los autores. Además, se asume que los datos no son posteriores a 2020, ya que este año marca el límite temporal del análisis, permitiendo así una comparación coherente con el ranking 2020 de Rolling Stone.

### 6.2. **¿Coinciden la crítica y la audiencia sobre quiénes son los grandes artistas musicales?**

#### 6.2.1. El reconocimiento de la crítica especializada

Los **5 artistas más canonizados por la crítica especializada** de la revista Rolling Stone son:

- **#1 Marvin Gaye**, artista de género masculino. Con su álbum *“What’s Going On”*, enmarcado en el género musical Soul/Gospel/R&B, alcanzó el primer puesto del Ranking 2020 de Rolling Stone. En la plataforma de streaming de música Spotify, su figura logró 68 puntos en la métrica del Índice de Popularidad. La edad promedio del artista al momento de ingresar, por primera vez, al Top 500 de la Revista Rolling Stone fue 32 años, y le tomó 10 años de carrera lograr incluirse en el ranking por primera vez.

- **#2 The Beach Boys**, banda de integrantes de género masculino. Con su álbum *“Pet Sounds”*, enmarcado en el género musical Desconocido, alcanzó el segundo puesto del Ranking 2020 de Rolling Stone. En la plataforma de streaming de música Spotify, los artistas lograron 62 puntos en la métrica del Índice de Popularidad. Su edad promedio al momento de ingresar, por

primera vez, al Top 500 de la Revista Rolling Stone fue 23,17 años, y les tomó 4 años de carrera integrar el ranking por primera vez.

- **#3 Joni Mitchell**, artista de género femenino. Con su álbum "*Blue*", enmarcado en el género musical Country/Folk/Country Rock/Folk Rock, alcanzó el tercer puesto del Ranking 2020 de Rolling Stone. En la plataforma de streaming de música Spotify, la artista no figura, por lo que no ha podido producir puntos ni presencia en la métrica del Índice de Popularidad. Su edad promedio al momento de ingresar, por primera vez, al Top 500 de la Revista Rolling Stone fue 28 años, y le tomó 3 años de carrera lograr incluirse en el ranking por primera vez.

- **#4 Stevie Wonder**, artista de género masculino. Con su álbum "*Songs in the Key of Life*", enmarcado en el género musical Soul/Gospel/R&B, alcanzó el cuarto puesto del Ranking 2020 de Rolling Stone. En la plataforma de streaming de música Spotify, su figura logró 70 puntos en la métrica del Índice de Popularidad. La edad promedio del artista al momento de ingresar, por primera vez, al Top 500 de la Revista Rolling Stone fue 26 años, y le tomó 14 años de carrera lograr incluirse en el ranking por primera vez.

- **#5 The Beatles**, banda de integrantes de género masculino. Con su álbum "*Abbey Road*", enmarcado en el género musical Desconocido, alcanzaron el quinto puesto del Ranking 2020 de Rolling Stone. En la plataforma de streaming de música Spotify, la banda logró 79 puntos en la métrica del Índice de Popularidad. La edad promedio de los artistas al momento de ingresar, por primera vez, al Top 500 de la Revista Rolling Stone fue de 27,75 años, y les requirió 6 años de carrera integrar el ranking por primera vez.

Aunque se trata de los artistas más reconocidos por la crítica especializada, se observa que **no integran el grupo de artistas con mayores puntuaciones en el Índice de Popularidad de la plataforma de streaming Spotify**. La media de los puntajes Índice de Popularidad en Spotify del top 5 canonizado por la crítica es: 69.75 (se debe considerar que Joni Mitchell no se encuentra contemplada en este resultado, dado que no figura como artista en Spotify). Este resultado refleja que los artistas más canonizados por la crítica especializada tienen un grado de consagración por el público de **moderado a elevado**. Como consecuencia, esto refuerza el patrón poco concluyente que arrojó como resultado la correlación de Pearson realizada entre variables: los resultados no responden a una tendencia clara con respecto a cómo se relaciona la opinión de los críticos con respecto a la opinión del público. ¿Por qué decimos esto? Por un lado, nos encontramos con cuatro artistas que tienen un reconocimiento moderado por parte de la audiencia (podría significar que, la opinión de la crítica con respecto a los artistas sí incide en la popularidad de los mismos en Spotify, y viceversa), pero luego, la situación de la artista Joni Mitchell viene a romper con ese patrón. Ella sí figura entre las artistas más reconocidas por la crítica, pero no figura como artista en la plataforma de streaming. Si bien conocer las causas de la no aparición de artistas en la plataforma restarían para un estudio posterior, esto nos permite afirmar que, quizás, no es necesario ser una artista consagrada por la audiencia en los medios de transmisión musical en línea para figurar entre las más alabadas por la crítica especializada. Este último caso demuestra que audiencia y crítica no siempre van de la mano, y refuerza el resultado de la correlación de Pearson entre variables (-0.2), que indica

que, aunque leve, existe una correlación negativa entre el Ranking 2020 de Rolling Stone y el Índice de Popularidad de Spotify (cuando una variable “sube”, la otra “baja”).

En conclusión, aunque los artistas más canonizados por la crítica especializada gozan de un grado moderado a elevado de popularidad en Spotify, la discrepancia de Joni Mitchell en este patrón subraya que **la popularidad en la plataforma de streaming no es un requisito para ser reconocido por la crítica**. Esto refuerza la idea de que la crítica y la audiencia **no siempre coinciden**, lo que queda evidenciado por la **baja correlación negativa** observada entre el ranking 2020 de Rolling Stone y el Índice de Popularidad de Spotify.

#### 6.2.3. El reconocimiento de la audiencia popular

Los **5 artistas más consagrados por la audiencia popular** de la plataforma de streaming Spotify son:

- **#1 Arctic Monkeys**, banda de integrantes de género masculino. Alcanzaron los 91 puntos en la métrica del Índice de Popularidad de Spotify. Por su parte, con su álbum “*AM*”, enmarcado en el género musical Indie/Alternative Rock, lograron el puesto 346 del Ranking 2020 de Rolling Stone. La edad promedio de los artistas al momento de ingresar, por primera vez, al Top 500 de la Revista Rolling Stone fue 27,5 años, y le tomó 7 años de carrera lograr incluirse en el ranking por primera vez.

- **#2 Kanye West**, artista de género masculino. Alcanzó los 87 puntos en la métrica del Índice de Popularidad de Spotify. Por su parte, con su álbum “*Graduation*”, enmarcado en el género musical Hip-Hop/Rap, logró el puesto 204 del Ranking 2020 de Rolling Stone. La edad promedio del artista al momento de ingresar, por primera vez, al Top 500 de la Revista Rolling Stone fue 30 años, y le tomó 3 años de carrera lograr incluirse en el ranking por primera vez.

- **#3 Harry Styles**, artista de género masculino. Alcanzó los 87 puntos en la métrica del Índice de Popularidad de Spotify. Por su parte, con su álbum “*Fine Line*”, enmarcado en el género musical Indie/Alternative Rock, logró el puesto 491 del Ranking 2020 de Rolling Stone. La edad promedio del artista al momento de ingresar, por primera vez, al Top 500 de la Revista Rolling Stone fue 25 años, y le tomó 2 años de carrera lograr incluirse en el ranking por primera vez.

- **#4 Billie Eilish**, artista de género femenino. Alcanzó los 86 puntos en la métrica del Índice de Popularidad de Spotify. Por su parte, con su álbum “*When We All Fall Asleep Where Do We Go?*”, enmarcado en el género musical Indie/Alternative Rock, logró el puesto 397 del Ranking 2020 de Rolling Stone. La edad promedio del artista al momento de ingresar, por primera vez, al Top 500 de la Revista Rolling Stone fue 18 años, y le tomó 0 años de carrera lograr incluirse en el ranking por primera vez.

- **#5 Frank Ocean**, artista de género masculino. Alcanzó los 86 puntos en la métrica del Índice de Popularidad de Spotify. Por su parte, con su álbum “*Blonde*”, enmarcado en el género musical Soul/Gospel/R&B, logró el puesto 79 del Ranking 2020 de Rolling Stone. La edad promedio del artista al momento de ingresar, por primera vez, al Top 500 de la Revista Rolling Stone fue 29 años, y le tomó 4 años de carrera lograr incluirse en el ranking por primera vez.

Estos 5 artistas son los más reconocidos por el público, en el marco de los resultados provenientes de la métrica Índice de Popularidad de la plataforma de streaming Spotify. No obstante, resulta interesante mencionar que tres de estos cinco artistas ocupan la franja de posiciones de **menor reconocimiento por la crítica**. El promedio de las posiciones obtenidas en el ranking 2020 por los 5 artistas más reconocidos por la audiencia es de **303**, mientras que la **mediana** es de **346**. Esta diferencia sugiere una distribución algo dispersa, aunque no extrema. Por lo tanto, ambos valores ayudan a dimensionar que estos artistas, aunque altamente populares en Spotify, en general **no se ubican entre las posiciones más destacadas del ranking elaborado por la crítica especializada**.

Nuevamente, esto vuelve a reforzar lo dicho en el apartado anterior en relación a la correlación de Pearson entre las dos variables. En este caso, la **correlación negativa baja** se vuelve aún más notoria, dado que, en general, quienes han logrado los mayores puntajes de popularidad en Spotify resultan ser los menos reconocidos en el Ranking 2020 de Rolling Stone. La **única excepción** pasa por el artista Frank Ocean, quien logró un alto puntaje en Spotify (86) y una posición (79) dentro del 25% (Q1) de los artistas más reconocidos por la crítica de la revista. No obstante, se trata de un caso aislado, ya que el patrón más representativo lo configuran los otros artistas del top 5.

En síntesis, este apartado refuerza la **correlación negativa baja** observada entre la consagración por parte del público y el reconocimiento de la crítica especializada. A pesar de que los artistas más populares en Spotify cuentan con una gran aceptación de la audiencia, en su mayoría no logran destacarse en el ranking 2020 de Rolling Stone, lo que sugiere que ambos criterios de valoración operan con lógicas parcialmente divergentes (una tendencia general, aunque no una regla absoluta).

### 6.3. Los más admirados por la crítica. ¿son realmente los más populares entre el público?: Caracterización de las diferencias a través de un análisis comparativo<sup>14</sup>

#### 6.3.1. ¿Qué valora la crítica especializada en un artista?

Top 5 Rolling Stone 2020: Análisis de Indicadores Descriptivos		
KPI / Indicador Descriptivo	Medida de Tendencia Central <sup>15</sup>	Resultado
Ranking 2020	Promedio/Mediana	3 (media)
Índice de Popularidad en Spotify	Promedio	69.75
Género musical del álbum	Moda	Desconocido Soul/Gospel/R&B
Género del artista	Moda	Masculino
Edad promedio de ingreso al ranking	Promedio	27.38

<sup>14</sup> Todas las tablas del presente apartado fueron creadas y calculadas a partir del dataset original y del proceso de análisis plasmado en el documento de Google Colab del proyecto.

<sup>15</sup> Para los **indicadores descriptivos**, como el género musical del álbum o el género del artista, se utilizó la **moda** como medida de tendencia central, ya que permite identificar la categoría más frecuente dentro del grupo analizado.

Para los **indicadores cuantitativos**, como la edad promedio de ingreso al ranking o la trayectoria artística, se empleó la **media o promedio**. Esta medida fue considerada adecuada dado que los valores no presentan una dispersión elevada. De todos modos, se reconoce que el promedio puede ser sensible a valores extremos, por lo que su aplicación fue acompañada por una revisión del rango para asegurar su representatividad.



de Rolling Stone		
Años transcurridos entre el debut y el ingreso al Top 500 de Rolling Stone	Promedio	7.40

En el marco del Ranking 2020 de Rolling Stone, los **artistas más canonizados por la crítica especializada** son figuras, de **predominancia masculina**, que han requerido una **trayectoria de 7.40 años en la industria**, en promedio, para lograr integrar el Top 500 de una de las revistas de crítica musical más reconocidas a nivel mundial, considerando que la edad promedio de ingreso de dichos artistas al ranking es de **27.38 años**. En cuanto al género musical de los álbumes, los **estilos más representativos** fueron *Soul/Gospel/R&B* y *Desconocido*, ambos con la misma frecuencia. A su vez, se trata de figuras que cuentan con **un grado de popularidad moderada** en el público consumidor de la plataforma de streaming Spotify.

Respecto al valor *Desconocido*, que aparece como moda en el indicador "género musical del álbum" del Top 5 de la crítica, corresponde a dos artistas: The Beach Boys y The Beatles. Si bien se trata de una categoría no especificada en el dataset original utilizado en este trabajo, se considera útil agregar como información complementaria que ambas bandas pertenecen al género del *rock*, aunque cada una haya desarrollado estilos diferentes dentro de este. Por lo tanto, si se buscara precisar mejor cuál fue el género predominante según esta métrica, se podría hablar del **Rock** y del **Soul/Gospel/R&B**.

#### 6.3.2. ¿Qué valora la audiencia popular en un artista?

Top 5 según Índice de Popularidad de Spotify: Análisis de Indicadores Descriptivos		
KPI / Indicador Descriptivo	Medida de Tendencia Central	Resultado
Ranking 2020	Promedio/Mediana	303.40 (media) 346 (mediana)
Índice de Popularidad en Spotify	Promedio	87.40
Género musical del álbum	Moda	Indie/Alternative Rock
Género del artista	Moda	Masculino
Edad promedio de ingreso al ranking de Rolling Stone	Promedio	25.90
Años transcurridos entre el debut y el ingreso al Top 500 de Rolling Stone	Promedio	3.20

En el marco de los puntajes del Índice de Popularidad de Spotify, los **artistas más consagrados por la audiencia popular** son figuras, de **predominancia masculina**, que han requerido una **trayectoria de 3.20 años en la industria**, en promedio, para lograr integrar el Top 500 de una de las revistas de crítica musical más reconocidas a nivel mundial, considerando que la edad promedio de ingreso de dichos artistas al ranking es de **25.90 años**.

En cuanto al género musical de los álbumes, el estilo más representativo fue el **Indie/Alternative Rock**. Por último, estos artistas cuentan con una gran aceptación de la audiencia, pero en su mayoría **no logran destacarse en el ranking 2020 de Rolling Stone**.

### 6.3.3. Entre el aplauso experto y el hit del público: dos lecturas diferentes del éxito

Análisis Comparativo de KPI's e Indicadores Descriptivos			
KPI / Indicador Descriptivo	Medida de Tendencia Central	Top 5 - Crítica	Top 5 - Audiencia
Ranking 2020	Promedio/Mediana	3 (media)	303.40 (media) 346 (mediana)
Índice de Popularidad en Spotify	Promedio	69.75	87.40
Género musical del álbum	Moda	Desconocido Soul/Gospel/R&B	Indie/Alternative Rock
Género del artista	Moda	Masculino	Masculino
Edad promedio de ingreso al ranking de Rolling Stone	Promedio	27.38	25.90
Años transcurridos entre el debut y el ingreso al Top 500 de Rolling Stone	Promedio	7.40	3.20

Los resultados obtenidos del análisis comparativo de los KPI's y los indicadores descriptivos de cada variable sugieren las siguientes conclusiones:

La crítica especializada de la Revista Rolling Stone y la audiencia popular que se expresa mediante plataformas de streaming como Spotify **no siguen los mismos criterios para valorar a un artista**, porque tanto un actor como otro no se encuentran determinados por los mismos sesgos ni por los mismos principios culturales, sociales, históricos y hasta tecnológicos que los orientan. Esto ya es un factor de peso para entender, en pocas palabras, por qué “no opinan lo mismo”.

El **ranking de los 500 mejores álbumes** de todos los tiempos es una métrica de evaluación que pertenece a la revista Rolling Stone, fundada en 1967 en San Francisco, Estados Unidos, y reconocida como una de las publicaciones más influyentes en el ámbito de la música, la cultura popular y la crítica especializada. Su enfoque es experto: lo integran críticos, artistas, periodistas y productores con alto conocimiento en la materia. Es decir, no se encuentra guiado por niveles de interacción, ventas o popularidad comercial. Y si bien, con el tiempo, ha ampliado los criterios que definen su canon musical, hay algo que permanece constante en sus evaluaciones: se tiene muy en cuenta el contexto sociocultural, el legado artístico, el impacto cultural, la influencia en otros artistas y el aporte a la evolución del género musical que logra alcanzar un álbum.

Considerar estos criterios de reconocimiento lleva a entender por qué, entre las características de sus artistas más reconocidos:

- Se observa una **expectativa de trayectoria considerablemente mayor** (7.40 años en promedio) a la hora de ingresar a su ranking, a comparación de las trayectorias que tienen los artistas más consagrados por la audiencia (y que, quizás no por casualidad, en su mayoría integran las posiciones más bajas del ranking). En consonancia con esto resuena el dato de que **la edad promedio de ingreso al ranking** (27.38) es, aunque por poco, **mayor** que la que se observa entre los artistas más populares.
- Se visualiza que cuentan con **un grado de popularidad moderada** (69.75) en el público consumidor de la plataforma de streaming Spotify. Es decir, resultan ser menos “populares” y “reconocidos” por la audiencia a comparación de las figuras que lideran la popularidad en el mundo del streaming. Esto refuerza la idea de que el factor comercial, de lo inmediato, de lo que dictamina el “termómetro del ahora”, no influye en sus decisiones con respecto a los “mejores de todos los tiempos”.
- Se identifica que pertenecen a géneros musicales como el Soul/Gospel/R&B y el Rock, considerados **más clásicos por su mayor legado y desarrollo histórico**, especialmente entre las décadas de 1950 y 1980, en contraste con géneros más recientes como el Indie/Alternative Rock, que serán abordados más adelante.

Por su parte, el **Índice de Popularidad de Spotify** funciona como una métrica de evaluación completamente diferente al ranking anteriormente descrito. Se trata de una plataforma más joven, nacida en el año 2006 y popularizada a nivel global a partir de la década del 2010, que se mide el éxito de los artistas a partir de observar el **comportamiento real de los usuarios** dentro de la aplicación: cantidad de reproducciones, guardados, inclusión en playlists, compartidos y nivel de interacción general. Puede decirse que su **enfoque es más democrático y abierto**: es una métrica dinámica y actualizada en tiempo real, que refleja las preferencias de la audiencia global, especialmente de los públicos más jóvenes, dentro del ecosistema digital.

Estos aspectos, enfocados en la **aceptación masiva**, permiten que la plataforma consagre a los artistas más populares en su sistema, reflejando el gusto inmediato de la audiencia. De esta manera, el Índice de Popularidad de Spotify valora principalmente la interacción de los oyentes con la música, basándose en una combinación de datos cuantificables, a diferencia de las consideraciones más cualitativas del ranking de Rolling Stone, que tienen ampliamente en consideración el legado y el impacto cultural de una obra musical.

Tomar en cuenta estos parámetros de valoración permite comprender por qué, entre las características de sus artistas más destacados:

- Se identifican varios representantes del Indie/Alternative Rock, el estilo musical más presente entre los artistas preferidos de la plataforma de transmisión musical en línea. Se trata de una **vertiente más joven**, nacida entre las décadas de 1990 y 2000, liderada por bandas que no encajaban en los géneros dominantes. En otras palabras, surgió como una expresión más independiente, con un sonido experimental, letras personales o críticas y una estética alternativa, lo que la volvía más **disidente frente a**

**todo lo considerado “clásico”** hasta entonces. Se posicionó por fuera del circuito comercial tradicional y terminó consolidando una actitud estética y musical propia.

Esto invita a pensar si la **juventud** —público objetivo central de Spotify y, me atrevería a decir, de la revolución digital tan masiva de estas últimas décadas— estará encarnando un ideal más desafiante de lo clásico para abrir paso a nuevas tendencias que, aunque recientes, también son valoradas como representativas y de calidad por una porción significativa de la audiencia.

- **No se visualiza una expectativa de trayectoria sólida en los artistas** (3.20 años en promedio), a diferencia del enfoque de los expertos, quienes tienden a destacar trayectorias más extensas en los artistas mejor valorados por la crítica. En línea con esto, se observa que la edad promedio de ingreso al ranking (25.90) resulta, aunque levemente, inferior a la que presentan los artistas reconocidos por el ranking 2020 de Rolling Stone.
- Se observa que los artistas más populares en Spotify (87.40) cuentan con una **fuerte aceptación por parte del público**, pero en su mayoría **no logran posicionarse entre los destacados del Ranking 2020 de Rolling Stone** (media de 303.40 y mediana de 346). Esto refleja que los artistas preferidos por la audiencia no responden a los mismos criterios de reconocimiento que emplean los expertos, sino a otros factores que, si bien exceden los límites del presente trabajo, pueden vincularse con la **rápida permeabilidad a lo que va ocurriendo**, la **preferencia musical individual**, los **niveles de consumo e interacción en la plataforma**, la **ubicación geográfica**, entre otros aspectos. Si bien resulta complejo condensar en pocas líneas qué espera una audiencia tan masiva de un artista, sí podemos señalar cierta **apertura hacia la incorporación de propuestas musicales más recientes** —aunque aún no cuenten con un legado consolidado—, así como una **mayor valoración de artistas jóvenes**, con trayectorias más breves o en desarrollo.

Desde una **perspectiva más crítica**, considerar el enfoque predominante de la revista implica también tener en cuenta todos los **atravesamientos socioculturales e históricos** que eso conlleva. La **crítica especializada** está conformada por un grupo más limitado de personas, que si bien se ha ido ampliando junto con el canon de la revista, no deja de estar atravesado por ciertos sesgos propios de quienes la integran. Estos sesgos pueden estar influenciados por sus trayectorias profesionales, sus referencias culturales, su formación histórica y una valoración arraigada en lo que fue considerado “increíble” o revolucionario en una cierta época, muchas veces ligados, quizás, a una visión más clásica y dominante del arte. En contraste, **Spotify** propone una **lógica más democrática**: cualquiera puede descargar la aplicación, interactuar y, en cierto modo, “votar”. Esto no significa que el algoritmo funcione del mismo modo para todos los artistas o regiones, ya que también presenta sesgos. Sin embargo, Spotify actúa, me animo a decir, como un **termómetro cultural, sensible a lo que ocurre en la sociedad día a día**. Es mucho más permeable a las variaciones que experimenta la cultura en términos de preferencias y formas de consumo musical, aspectos

constantemente influenciados por los cambios históricos, económicos, sociales, culturales y tecnológicos contemporáneos. Todo esto, potenciado además por el fenómeno de la globalización, que acelera la velocidad de esos cambios y desdibuja sus límites geográficos.

La última nota de análisis se la lleva la cuestión de género, en donde ambas variables, el Ranking 2020 de Rolling Stone y el Índice de Popularidad de Spotify, guardan una coincidencia casi exacta: la **predominancia del género masculino** en ambas variables. En cada grupo de 5 artistas, hubo solamente 1 mujer que logró integrar el top: Joni Mitchell, en el caso del Top 5 de la crítica, y Billie Eilish, en el caso del Top 5 de la audiencia. Queda la pregunta por qué lugar está teniendo, tanto en las plataformas de streaming como en medios más tradicionales como la revista Rolling Stone, la **representatividad de las mujeres y de todas las identidades sexo-genéricas** que el mundo alberga, considerando que el indicador descriptivo *género del artista*, **solo incluye las categorías masculino-femenino**, provenientes de lo dictaminado por el modelo heterocisnormativo que atraviesa de manera histórica, aún hasta hoy, a toda nuestra sociedad.

En conclusión al presente trabajo, me interesa resumir las conclusiones mencionadas en los siguientes **insights principales**:

1. **Diferentes criterios de evaluación:** la crítica especializada en Rolling Stone y la audiencia en Spotify valoran a los artistas según criterios muy distintos. La crítica prioriza el legado y la influencia cultural, mientras que Spotify refleja el gusto inmediato y la interacción de la audiencia.
2. **Trayectoria de los artistas:** los artistas más valorados por la crítica en el ranking 2020 tienen una trayectoria media de 7.40 años, mientras que los más populares en Spotify tienen una trayectoria promedio de sólo 3.20 años.
3. **Género musical predominante:** la crítica valora más géneros clásicos como Soul/Gospel/R&B y Rock, mientras que la audiencia de Spotify prefiere géneros más recientes como Indie/Alternative Rock.
4. **Popularidad y aceptación:** los artistas más destacados por la crítica tienen una popularidad moderada, mientras que los artistas más populares en Spotify tienen una aceptación más masiva, aunque no logran figurar entre los mejores en Rolling Stone.
5. **Diferencias generacionales y culturales:** Spotify refleja un gusto más joven y cambiante, influenciado por la globalización y su permeabilidad a las variaciones culturales, mientras que la crítica especializada tiene una visión más consolidada en las referencias históricas y culturales.
6. **Relevancia de lo "nuevo":** los artistas populares en Spotify son en su mayoría jóvenes con trayectorias más cortas, lo que refleja la apertura de la audiencia hacia propuestas más recientes y desafiantes frente a lo clásico.
7. **Sesgos en la crítica y el algoritmo:** la crítica musical está influenciada por un grupo reducido con sesgos históricos y culturales, mientras que Spotify actúa como un

termómetro cultural más amplio y democrático, aunque también presenta sesgos derivados de su algoritmo.

8. **Discriminación de género:** tanto el ranking de Rolling Stone como el Índice de Popularidad de Spotify muestran una predominancia masculina, con solo una mujer destacada en cada grupo de los 5 artistas más destacados, lo que resalta la falta de representatividad de las mujeres y otras identidades sexo-genéricas en ambas plataformas.

Para concluir este apartado, me gustaría citar una frase del artículo de Rolling Stone *"The 500 Greatest Albums of All Time"*<sup>16</sup>:

***"The classics are still the classics, but the canon keeps getting bigger and better"*** (los clásicos siguen siendo clásicos, pero el canon sigue creciendo y mejorando). Esta afirmación refleja un optimismo social y cultural, ya que sugiere que el espectro de reconocimiento se está ampliando, permitiendo que más artistas, que representan una diversidad de voces y realidades, encuentren su lugar en el canon musical.

## 7. Visualización de los principales insights: dashboard

El dashboard titulado *"¿Coinciden la crítica y la audiencia sobre quiénes son los grandes artistas musicales? Un análisis en la era del streaming a partir de los rankings históricos de Rolling Stone y la popularidad en Spotify"* fue diseñado para representar de forma visual las conclusiones cuantitativas y cualitativas obtenidas a lo largo del análisis. En una primera instancia, se presentan los resultados principales de cada variable por separado —el ranking 2020 de Rolling Stone y el Índice de Popularidad de Spotify— para luego avanzar hacia una síntesis comparativa de ambos enfoques. La visualización busca facilitar la interpretación de los hallazgos, permitiendo identificar coincidencias, contrastes y patrones emergentes que enriquecen la lectura de los datos.

Para su elaboración se utilizó Google Looker Studio, una herramienta de visualización de datos en la nube que permite crear informes y paneles interactivos a partir de fuentes de información. El dataset trabajado fue procesado previamente con Python a través de Google Colab, desde donde se realizó la extracción, limpieza, transformación, y carga de datos, y el Análisis Exploratorio de Datos (EDA) que nutren al dashboard como fuente principal.

Para explorar los principales hallazgos de este análisis de manera visual e interactiva, se puede acceder al dashboard completo haciendo clic en el siguiente [enlace](#).

## 8. Conclusiones finales del trabajo

Ha sido altamente gratificante y enriquecedor realizar este proyecto de análisis de datos. Representó mi primera incursión en el campo de la Ciencia de Datos, un área en la que espero seguir formándome y adquiriendo experiencia continuamente. Entre los aprendizajes que me llevo, destaco la oportunidad de aplicar conceptos teóricos adquiridos a lo largo del bootcamp, logrando plasmar en la práctica lo aprendido, lo cual considero fundamental para el aprendizaje real.

---

<sup>16</sup> Rolling Stone. "The 500 Greatest Albums of All Times". Disponible en: <https://www.rollingstone.com/music/music-lists/best-albums-of-all-time-1062063/jay-z-the-blueprint-3-1063183/>

Proveniente del ámbito de las Ciencias Sociales, la tarea analítica ha sido una de mis favoritas en mi carrera profesional, y poder elegir un tema que me permitiera desarrollar esta habilidad para generar conclusiones significativas ha sido muy valioso. Mi objetivo es combinar las herramientas técnicas de la Ciencia de Datos con la perspectiva teórica y analítica de las Ciencias Sociales, a fines de generar trabajos que realmente aporten.

Realizar un recorte analítico y metodológico fue clave para hacer este proyecto viable. Sin embargo, como horizontes de exploración, considero dos direcciones que serían interesantes de abordar:

1. Analizar la métrica de Índice de Popularidad en función del algoritmo de Spotify, comparándolo con una nueva edición del ranking de la Revista Rolling Stone.
2. Investigar qué variables adicionales influyen en los criterios de evaluación de la crítica y la audiencia. A modo de ejemplo, en el presente trabajo se observó que, entre los artistas más destacados, predominan los estadounidenses, lo que podría justificar un análisis de la variable nacionalidad. Además, profundizar en los algoritmos de Spotify para identificar los factores que contribuyen a consagrar a un artista también sería una exploración valiosa.

Finalmente, mi intención es seguir aportando desde una mirada social y técnica por medio de los datos, buscando generar conocimientos que enriquezcan nuestra comprensión de la realidad y puedan ser utilizados para mejorar diversos ámbitos de la sociedad.