



**PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR**

**ESCUELA DE NEGOCIO Y COMERCIO INTERNACIONAL**

**TRABAJO DE INTEGRACIÓN CURRICULAR PREVIO A LA OBTENCIÓN DEL  
TÍTULO DE LICENCIADA EN NEGOCIOS INTERNACIONALES**

**TEMA:**

*“CONSTRUCCIÓN DE UN MODELO DE REGRESIÓN LINEAL PARA LA PROYECCIÓN DE  
VENTAS EN PLATAFORMAS DE STREAMING”*

**AUTOR:**

EMILIA ALEJANDRA CIFUENTES HINOJOSA

**TUTOR:**

MARICELA FERNANDA ORMAZA MOREJÓN

**IBARRA – ECUADOR**

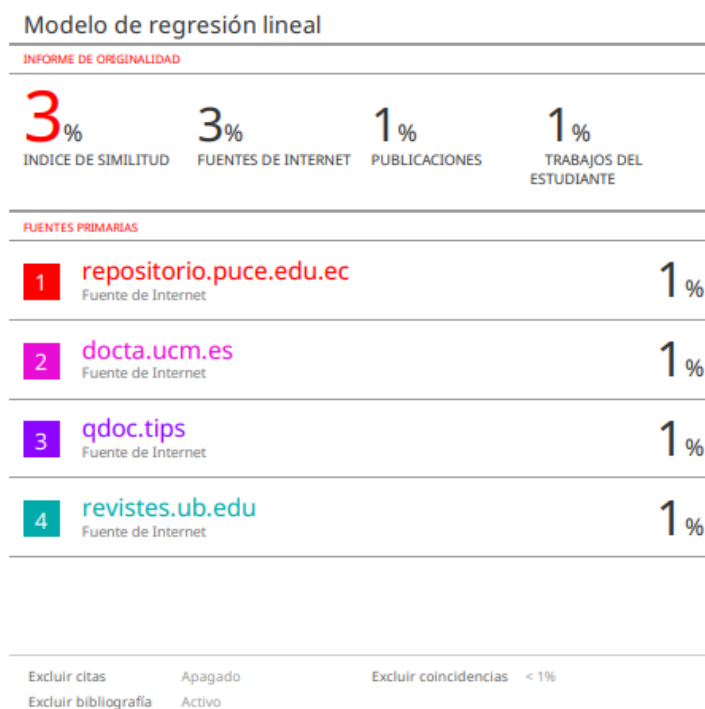
**Junio, 2024**

Ibarra, 22 de junio, 2024

## CERTIFICACIÓN TUTOR

En mi calidad de Tutor del Trabajo de integración curricular titulado: CONSTRUCCIÓN DE UN MODELO DE REGRESIÓN LINEAL PARA LA PROYECCIÓN DE VENTAS EN PLATAFORMAS DE STREAMING, presentado por la estudiante Emilia Alejandra Cifuentes Hinojosa con cédula de ciudadanía N° 1050236510, para obtener el Título de Licenciada en Negocios Internacionales.

Certifico que el trabajo cumple con todos los parámetros establecidos, mediante el cual el estudiante demuestra el desarrollo de competencias en el campo de conocimiento de su profesión con un nivel de argumentación coherente, para ser sometido a la evaluación por parte de los lectores.



(f):

Mgs. Marisela Fernanda Ormaza Morejón  
**TUTOR DE TRABAJO**  
C.C.: 1003324975

## PÁGINA DE APROBACIÓN DEL TRIBUNAL

El tribunal examinador, aprueba el presente trabajo en nombre de la Pontificia Universidad Católica del Ecuador Ibarra:

(f): .....

**Mgs. Maricela Fernanda Ormaza Morejón**

**C.C.: 1003324975**



(f): .....

**Mgs. Sandra Nadyedja Landázuri Espinoza**

**C.C.: 0400898276**



(f): .....

**Mgs. Cristian Xavier Cifuentes Figueroa**

**C.C. : 1001631645**

## ACTA DE CESIÓN DE DERECHOS

Yo, Emilia Alejandra Cifuentes Hinojosa, declaro conocer y aceptar la disposición del Art. 165 del Código Orgánico de Economía Social de los Conocimientos, Creatividad e Innovación, que manifiesta textualmente: “Se reconoce facultad de los autores y demás titulares de derechos de disponer de sus derechos o autorizar las utilidades de sus obras o prestaciones a título gratuito y oneroso, según las condiciones que determinen. Esta facultad podrá ejercerse mediante licencias libres, abiertas y otros modelos alternativos de licenciamiento o la renuncia”.

Ibarra, 22 de junio de 2024

(f): \_\_\_\_\_

Emilia Alejandra Cifuentes Hinojosa

C.C.: 1050236510

## AUTORÍA

Yo, Emilia Alejandra Cifuentes Hinojosa, portadora de la cedula de ciudadanía N° 1050236510, declaro que el presente trabajo de investigación es de total responsabilidad del autor, y eximo expresamente a la Pontificia Universidad Católica del Ecuador Ibarra de posibles reclamos o acciones legales.

(f): .....

Emilia Alejandra Cifuentes Hinojosa

C.C.: 1050236510

## DEDICATORIA

A mi Familia,

Por ser el faro en cada tormenta y el puerto en cada regreso, por sus abrazos que sostienen y sus palabras que alientan, en sus ojos encontré la fuerza incluso cuando sentía que caía, y a mi ángel que me cuida desde el cielo, siempre presente en mi corazón, que nunca me dejó rendirme. Este es el resultado de la confianza que han puesto en mí; lo que soy se los debo.

Gracias por ser la luz, la sabiduría y la fe en este viaje.

Con amor,

Emi.

## AGRADECIMIENTOS

Agradezco profundamente a la Pontificia Universidad Católica del Ecuador – Ibarra, por ser el espacio donde florecieron mis sueños y conocimientos. A toda su planta docente, por su dedicación incansable, sabiduría y pasión, que inspiraron cada paso de este recorrido académico. Gracias por las enseñanzas que van más allá de los libros y por cultivar en mí un espíritu crítico y curioso.

En especial a la Mgs. Maricela Ormaza, cuya guía y paciencia fueron pilares en momentos de incertidumbre. Su compromiso y apoyo constante no solo enriquecieron este proyecto, sino que también dejaron una huella imborrable en mi formación personal y profesional. Su confianza en mis capacidades me impulsó a alcanzar metas que creí inalcanzables.

Este proyecto es el resultado de años de aprendizaje que finalmente rinden frutos. De corazón, les agradezco porque sin su compromiso con la educación y fe inquebrantable en mi desarrollo, este logro no hubiese sido posible.

Con Gratitud,

Emilia Cifuentes

## ÍNDICE DE CONTENIDOS

CERTIFICACIÓN TUTOR .....	ii
PÁGINA DE APROBACIÓN DEL TRIBUNAL .....	iii
ACTA DE CESIÓN DE DERECHOS.....	iv
AUTORÍA .....	v
ÍNDICE DE CONTENIDOS.....	viii
INTRODUCCIÓN .....	1
□ Objetivo General .....	3
□ Objetivos Específicos .....	3
1. ESTADO DEL ARTE .....	4
2. MATERIALES Y MÉTODOS .....	26
2.1. Tipo de Investigación.....	26
2.2. Enfoque.....	26
2.3. Modelo .....	26
2.4. Variables .....	27
2.5. Indicadores .....	27
2.6. Instrumento .....	29
2.7. Muestra .....	30
2.8. Técnicas .....	31
2.9. Procedimiento de Aplicación .....	32

2.9.1. Muestreo.....	32
2.9.2. Exploración .....	33
2.9.3. Modificación.....	34
2.9.4. Modelación.....	35
3.1. Significado de los Coeficientes .....	52
3.2. Interpretación de los Coeficientes .....	53
3.3. Significancia estadística y P-Valores.....	54
3.4. Prueba de Hipótesis .....	55
3.4.1. Resultados de la Prueba de Hipótesis .....	56
3.4.2. Interpretación de los resultados .....	57
3.5. Capacidad de pronosticación.....	58
3.5.1. Metodología de simulación .....	58
3.5.2. Cálculo de los valores de Y.....	59
3.6. Traducción de resultados en Valores Económicos.....	61
3.7. Limitaciones del modelo.....	62
3.8. Aportaciones del modelo .....	64
4.1. Conclusiones .....	66
4.2. Recomendaciones .....	67
5. REFERENCIAS BIBLIOGRÁFICAS.....	70
ANEXOS.....	76

## ÍNDICE DE TABLAS

Tabla 1.....	8
Tabla 2.....	13
Tabla 3.....	15
Tabla 4.....	23
Tabla 5.....	25
Tabla 6:.....	36
Tabla 7.....	38
Tabla 8.....	39
Tabla 9.....	40
Tabla 10.....	49
Tabla 11.....	60
Tabla 12.....	60
Tabla 13.....	61

## ÍNDICE DE FIGURAS

Figura 1.....	7
Figura 2.....	9
Figura 3.....	12
Figura 4.....	16
Figura 5.....	17
Figura 6.....	18
Figura 7.....	20

Figura 8.....	21
Figura 9.....	22
Figura 10.....	24
Figura 11.....	31
Figura 12.....	37
Figura 13.....	41
Figura 14.....	41
Figura 15.....	41
Figura 16.....	42
Figura 17.....	42
Figura 18.....	44
Figura 19.....	45
Figura 20.....	45
Figura 21.....	45
Figura 22.....	46
Figura 23.....	46
Figura 24.....	48
Figura 25.....	57

## RESUMEN

Este Proyecto de Investigación aborda la construcción de un modelo de regresión lineal múltiple para predecir las ventas de artistas ecuatorianos en plataformas de streaming como Spotify y YouTube. El objetivo principal es desarrollar un modelo de regresión lineal que permita proyectar las ventas de música digital de artistas ecuatorianos.

Se identifican variables relevantes como seguidores, oyentes mensuales y popularidad en plataformas de streaming, y se recopilan datos de diversas fuentes. El análisis de regresión se utiliza para entender la relación entre estas variables y predecir ventas de música; el estado del arte muestra la evolución de la industria musical con el surgimiento de redes sociales y plataformas de streaming, que han facilitado la difusión de contenido y creado nuevas oportunidades para los artistas, siendo Spotify el líder emergente en el mercado de streaming, con un crecimiento significativo en usuarios y suscriptores

. La metodología incluye la selección de variables, construcción del modelo de regresión y ajuste mediante el análisis de varianza (ANOVA). El modelo muestra una alta capacidad explicativa ( $R^2$  del 77,7%) y cumple con los supuestos fundamentales: normalidad homocedasticidad, interdependencia, no colinealidad y linealidad. Concluyendo que el modelo es efectivo para predecir la variable dependiente seleccionada, oyentes mensuales en Spotify para artistas ecuatorianos.

**Palabras Clave:** Plataformas de Streaming, Regresión Lineal Múltiple, Industria Musical, Predicción de ventas, Globalización, Artistas Ecuatorianos.

## ABSTRACT

This research project addresses constructing a multiple linear regression model to predict sales of Ecuadorian artists on streaming platforms such as Spotify and YouTube. The main objective is to develop a linear regression model that allows projecting digital music sales of Ecuadorian artists.

Relevant variables such as followers, monthly listeners and popularity on streaming platforms are identified, and data is collected from various sources. Regression analysis is used to understand the relationship between these variables and predict music sales; state of the art shows the evolution of the music industry with the emergence of social networks and streaming platforms, which have facilitated the dissemination of content and created new opportunities for artists, with Spotify being the emerging leader in the streaming market, with significant growth in users and subscribers.

The methodology includes the selection of variables, construction of the regression model and adjustment by analysis of variance (ANOVA). The model shows high explanatory power ( $R^2$  of 77,7%) and meets the fundamental assumptions: normality, homoscedasticity, interdependence, non-collinearity and linearity. They concluded that the model is effective in predicting the selected dependent variable, monthly listeners on Spotify for Ecuadorian artists.

**Keywords:** Streaming platforms, Multiple Linear Regression, Music Industry, Sales Prediction, Globalization, Ecuadorian Artists.

## INTRODUCCIÓN

La transformación digital y la globalización han revolucionado la industria musical en las últimas décadas, reconfigurando sus dinámicas económicas, tecnológicas y culturales. En este contexto, las plataformas de streaming han emergido como actores clave en la distribución y consumo de música. Sin embargo, para Ecuador este nuevo paisaje presenta tanto oportunidades como desafíos, especialmente para artistas ecuatorianos que buscan destacarse en un mercado global saturado. Esta investigación se enfoca en predecir las ventas de artistas musicales ecuatorianos en plataformas de streaming como Spotify y YouTube utilizando un modelo de regresión lineal.

Desde el punto de vista de los negocios internacionales, la industria musical representa una parte importante de la economía global, constituyendo un componente clave de bienes y servicios culturales. Comprender cómo las plataformas digitales están impactando las ventas de música ecuatoriana en un mercado globalizado es fundamental para el desarrollo de estrategias de comercio exterior y promoción de la cultura nacional en el ámbito internacional.

Al comprender las preferencias y tendencias de consumo en estos mercados, las empresas ecuatorianas pueden adaptar sus estrategias de marketing y promoción para llegar de manera más efectiva a audiencias internacionales y aprovechar las oportunidades de crecimiento en el comercio internacional transfronterizo.

La música es un lenguaje universal que nunca ha pasado de moda, y ha sido hasta la fecha un negocio rentable, que ha prosperado con el pasar de los años (UNCTAD, 2022). Vivimos en una era de transformaciones sin precedentes impulsadas por la globalización y digitalización, reconfigurando la naturaleza de la industria musical, y alterando profundamente las dinámicas económicas y de negocio del sector. La proyección de ventas nos proporciona

información sobre cómo la música ecuatoriana se posiciona en este nuevo escenario; no es simplemente un tema de identidad cultural, sino que tiene implicaciones directas tangibles en el ámbito de negocios, y comercio internacional.

El análisis de regresión es una técnica estadística para estudiar la relación entre variables, que puede utilizarse para medir datos cuantificables con fines predictivos (Pereira, 2010). Este tipo de regresiones son utilizadas para predicciones en varios campos económicos, como pueden ser las ventas de música de artistas ecuatorianos. En un mundo digital, el potencial de mercado para la música ecuatoriana es enorme.

Sin embargo, para capitalizar este potencial, es esencial comprender las dinámicas cambiantes de estas plataformas digitales. Esto no solo afectará a artistas individuales, sino que tendrá un impacto en toda la cadena de valor: desde productores y distribuidores hasta empresas tecnológicas y propuestas de marketing. Para las empresas y emprendedores, el entender y adaptarse a estas dinámicas puede significar la diferencia entre el éxito y el fracaso del mercado globalizado.

En el informe anual presentado por la Federación Internacional de la Industria Fonográfica (IFPI), con proyecciones favorables, la institución conformada por 8,000 miembros alrededor del mundo sostiene que el streaming es la principal fuente de ingresos de la industria, representando el 67% del mercado musical (IFPI, 2022). Por lo tanto, predecir las ventas de música de artistas ecuatorianos utilizando datos de búsqueda en estas plataformas puede ser muy eficaz para la previsión del consumo y las ventajas de promocionarse por este medio.

La presente investigación tiene como objetivos:

- ***Objetivo General***

Elaborar un modelo de regresión lineal para la proyección de ventas en plataformas de streaming, centrándose en el consumo de artistas ecuatorianos.

- ***Objetivos Específicos***

- Identificar las variables relevantes que influyen en el consumo de contenido dentro de las plataformas de streaming, con un enfoque específico en el contexto ecuatoriano.
- Minar los datos de diversas plataformas de streaming que alojan contenido de diversos artistas ecuatorianos.
- Construir el modelo de regresión lineal adaptando las características para la proyección de ventas de artistas ecuatorianos en las plataformas de streaming.

## 1. ESTADO DEL ARTE

En los últimos años, la forma en que consumimos contenido multimedia ha presenciado una evidente evolución, no es un misterio que en la industria musical se requieren de estrategias de difusión para conectarse al público adecuado. Con el nacimiento de las redes sociales, estos procesos han logrado simplificarse, aun así, siempre se busca obtener el mayor impacto; la revolución digital impactó a la industria ofreciendo una serie de oportunidades y ventanas de visibilidad para aquellos artistas que deciden impulsar sus proyectos haciendo uso de estas. Con el surgimiento de plataformas digitales como Spotify, Deezer, iTunes Store, YouTube, como menciona Muñoz (2018), que permiten mediante las redes sociales acercar a los artistas hacia nuevos oyentes.

La irrupción de la web 2.0 ha marcado un hito significativo al empoderar a los usuarios de Internet, permitiéndoles una participación en la red. La emergencia de las redes sociales ha facilitado la difusión de contenidos de un usuario a otro, multiplicando su impacto y alcance. En ocasiones, estos contenidos adquieren una popularidad masiva en lapsos de tiempo relativamente cortos, siendo etiquetados como producciones “virales” (Baños-Gonzales et al., 2020). Dentro de la amplia gama de formatos disponibles en la web, el streaming se destaca como el medio que mejor encapsula el fenómeno viral.

El streaming puede definirse como la distribución digital de archivos multimedia a través de la difusión continua por una red informática, como lo menciona Jáuregui (2015). Los usuarios tienen la capacidad de recibir y reproducir contenido de audio o video en tiempo real a través de proveedores sin necesidad de descargarlo previamente, aunque los antecedentes de esta práctica en internet se remontan a mediados de los años 90, el desarrollo de plataformas de música basadas en servicios de streaming se produjo principalmente durante la década siguiente, gracias

al avance en el ancho de banda, la proliferación de conexiones Wi-Fi y el uso extendido de smartphones en todo el mundo.

Específicamente en el ámbito musical, el streaming continúa experimentando un crecimiento sostenido con una base de usuarios de pago que supera los 580 millones de dólares a nivel global. Este crecimiento ha generado un incremento significativo en los ingresos de la industria, que superaron los 17,000 millones de dólares en el 2022 como lo señala Plaza Chicote (2023). Entre las diversas plataformas de streaming musicales, Spotify puede considerarse como líder en el mercado, con más del 30% de los suscriptores a nivel mundial, duplicando las cifras de sus competidores más cercanos como Amazon Music y Apple Music.

Además del crecimiento del contenido en la industria del audio como los podcasts, ha alcanzado más de los 640 millones de oyentes a nivel mundial en 2022. El proyecto inicial de Spotify se gestó en Suecia en 2006, pero su lanzamiento oficial tuvo lugar en 2008, tras asegurar acuerdos de licencia con las principales compañías discográficas después de dos años de pruebas. Pruvost (2014) hace hincapié en que, junto con la introducción de servicios de suscripción de pago, las cuentas gratuitas, inicialmente limitadas a invitaciones, se extendieron gradualmente, llegando al Reino Unido en 2009.

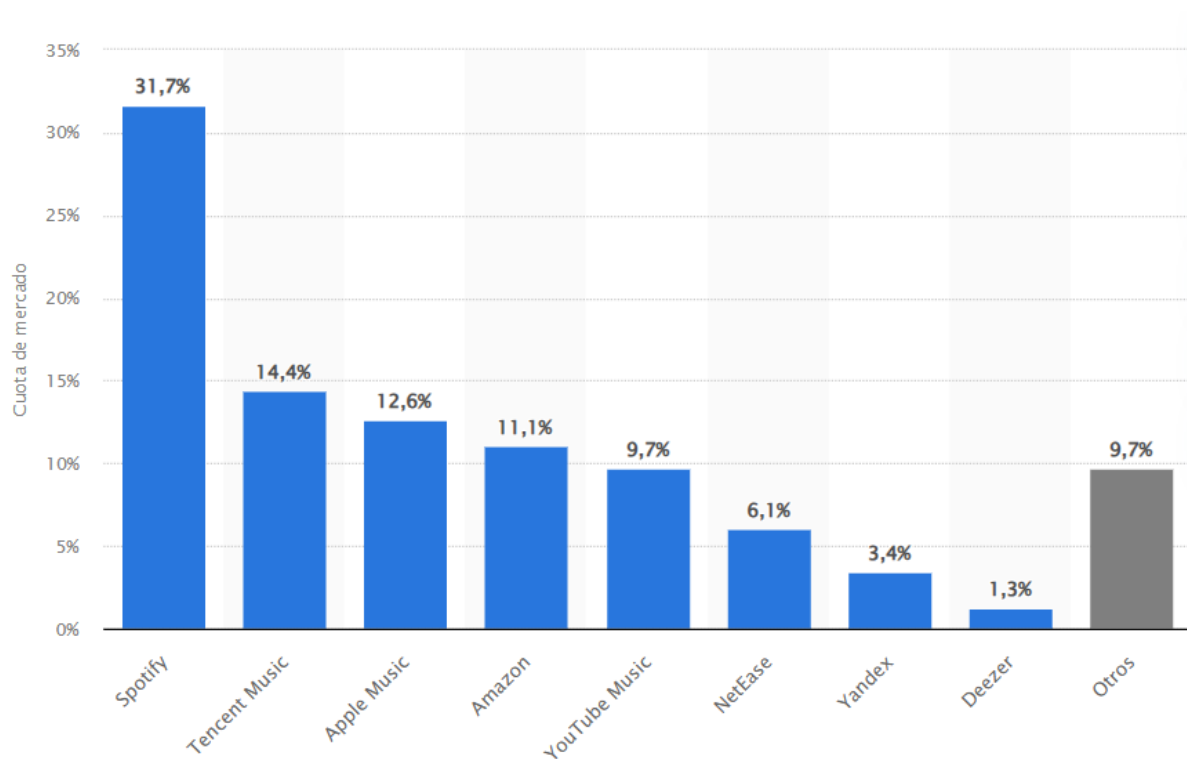
De esta manera, a pesar de enfrentar varios obstáculos, como ser identificado erróneamente como un virus y ser bloqueado en millones de computadoras, en el 2011 Spotify logró establecerse con éxito tras su lanzamiento en Estados Unidos y la transición del modelo de red parcialmente entre pares a uno centralizado; este proceso de adaptación al streaming, reflejó las tensiones inherentes entre la industria discográfica, caracterizada por un sistema centralizado de producción y distribución, y los principios de comunicación en red impulsados por internet.

En la Actualidad, según el Registro Único de Artistas y Gestores Culturales (RUAC), el Ministerio de Cultura y Patrimonio (2024), Ecuador registra más de 23.407 artistas de los cuales 9.137 pertenecen al sector de artes musicales y sonoras. Sin embargo, un gran porcentaje de ellos no cuenta con el apoyo necesario para mantenerse en la industria artística; si bien es cierto que las plataformas digitales permiten la difusión de contenidos, al no estar familiarizados con las redes sociales resulta un verdadero reto conseguir una audiencia permanente que permita establecer una perspectiva clara en el éxito de un lanzamiento musical por medio de estas plataformas.

El músico hoy en día analiza diversas estrategias de comunicación con la intención de sumar seguidores ya que al hacer uso de plataformas digitales buscan promocionarse como artistas. Por otro lado, desde el punto de vista de Naranjo Alvarado (2023), el Ecuador es un país caracterizado por tener una amplia variedad de géneros musicales que reflejan su diversidad cultural y geográfica. Gracias a la privilegiada ubicación donde se encuentra, Ecuador ha recibido influencias musicales de diferentes regiones del mundo, que se han fusionado con las expresiones autóctonas de los pueblos originarios, todos estos estilos se encuentran distribuidos en un nuevo mercado, representado por las plataformas de streaming a nivel mundial, representado en la Figura 1, a continuación.

**Figura 1**

*Distribución porcentual del mercado de la música en streaming a nivel mundial durante el tercer trimestre de 2023, por plataforma.*



*Nota.* El gráfico representa el porcentaje de participación de las distintas plataformas de streaming a nivel mundial a lo largo del 2023. Tomando de Orús (2024).

En Ecuador aún existen grandes falencias dentro del mercado musical, tales como la ausencia de empresas que impulsen al artista, así como la falta de interés por los medios de comunicación de distribuir música producida en el país. Desde la perspectiva de Caicedo (2021), por esta razón el lanzamiento por medios digitales, como plataformas musicales representan una gran oportunidad para crecer en la industria. Actualmente el éxito de un artista puede medirse a través de su participación en la red, es decir que presentar un modelo predictivo de las posibles ventas en estas plataformas resulta ventajoso para aquellos músicos que buscan crecer en la

industria fonográfica, este crecimiento se puede reflejar en la Tabla 1, haciendo hincapié en la situación ecuatoriana.

**Tabla 1**

*Crecimiento de Audiencias en Redes Sociales. Mayo 2023 vs enero 2024*

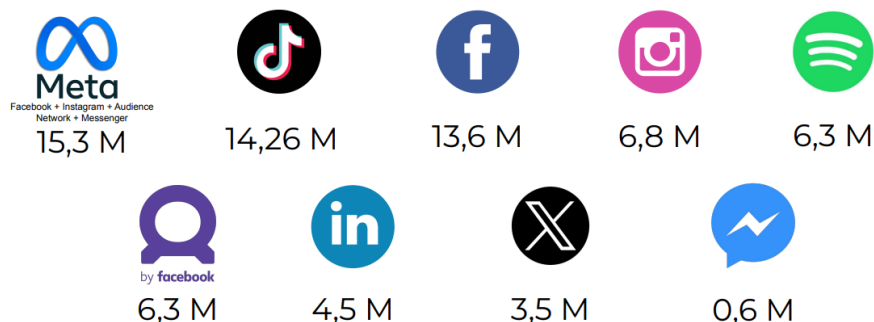
Red Social	Usuarios Mayo 2023	Usuarios Enero 2024
TikTok	11,912,000.00	14,260,000.00
Facebook	14,000,000.00	13,600,000.00
Instagram	7,000,000.00	6,800,000.00
Spotify	6,200,000.00	6,300,000.00
Linkedin	3,900,000.00	4,500,000.00
X (Twitter)	3,500,000.00	3,500,000.00

*Nota.* En la siguiente tabla se evidencia el crecimiento entre la cantidad de usuarios en las diferentes plataformas y redes sociales en el período de mayo de 2023 hasta enero de 2024 en Ecuador. Tomada desde MENTINNO (2024).

Como se puede observar en la tabla realizada por MENTINNO (Inteligencia y Analítica de Negocios), el crecimiento de plataformas digitales tiene sus fluctuaciones en las diferentes redes sociales, sin embargo, Spotify como la principal herramienta de distribución de streaming en la industria musical representa un crecimiento en el número de usuarios registrados. Entre las redes sociales también podemos observar que efectivamente Spotify mantiene el podio entre las plataformas de reproducción de streaming en música que se utilizan en Ecuador, representado en la Figura 2, de igual manera los artistas musicales utilizan YouTube como una herramienta a favor de su impulso y difusión con un mayor alcance y sin ningún costo adicional.

**Figura 2**

*Cuentas-Usuarios de Redes Sociales Alcanzables con Publicidad enero 2024*



*Nota.* El gráfico representa los valores aproximados de las plataformas y redes sociales que más usuarios obtuvieron en Ecuador para el año 2023. Tomada desde el informe anual de MENTINNO (2024).

La industria musical es un sector que aporta un alto nivel económico a las plataformas de streaming, además que es utilizado en sectores como el marketing, publicidad y medios de comunicación. Como lo mencionan Belmares Mendoza y Guevara García (2021) estas plataformas digitales son medios que hoy en día están a disponibilidad de cualquier individuo en dispositivos como celulares, laptops, tabletas, que se han acoplado a la alta demanda de estas plataformas, ya que gracias al internet los consumidores empiezan a escuchar y visualizar material musical de manera rápida y constante.

Según expone Riveros (2018) en las plataformas digitales existe un intermediario entre los artistas y las plataformas, este es el streaming. Son negocios que se interesan por los artistas independientes o intérpretes nuevos que los sellos discográficos y grandes empresas no apoyan. En la actualidad, resulta inevitable el no utilizar plataformas digitales de streaming, a su vez este tipo de plataformas cambian la forma de pensar del consumidor.

Como analiza Masís (2021), tras el debut de Spotify en el año 2014 en el mercado musical, se ha proporcionado como el líder de las plataformas de streaming teniendo como principal competencia a YouTube, una plataforma no solo dirigida al entretenimiento y los videos musicales; Deezer que ofrece un servicio muy parecido a Spotify, permitiendo la reproducción de canciones de manera gratuita y con publicidad; y Apple Music, junto al gigante de la tecnología Apple.Inc.

El contexto en el que se desarrollan los nuevos fenómenos propios de la economía digital o capitalismo contemporáneo y sus transformaciones se pueden encontrar rasgos distintivos, en primera instancia el capitalismo digital, que se tradujo como un proceso de desindustrialización en el que como resultado emerge la economía de servicios, en la que una parte importante de la producción se convierte en conocimiento, datos o información. Como continúa explicándolo Masís (2021), se trata de un proceso de digitalización del conocimiento y la información, en el cual los algoritmos se convierten en los nuevos medios de producción, mientras que la información se transforma en una mercancía sumamente importante que se caracteriza por la inminente facilidad de reproducción.

Por otro lado el capitalismo de plataformas alude a las infraestructuras digitales, que tienen como fin mejorar la rentabilidad a partir de la información, como parte de este sistema surge el llamado capitalismo de vigilancia, buscando básicamente convertir la experiencia humana en mercancía, llegando a la conformación del prosumer (productor y consumidor), como característica fundamental de la web 2.0, ya que al hacer uso de estas plataformas se produce información sobre las preferencias del usuario, que a la vez consume las recomendaciones sugeridas por la misma plataforma mediante algoritmos generados a partir de dicha información. En este sentido, como menciona Masís González (2021), las plataformas digitales son la

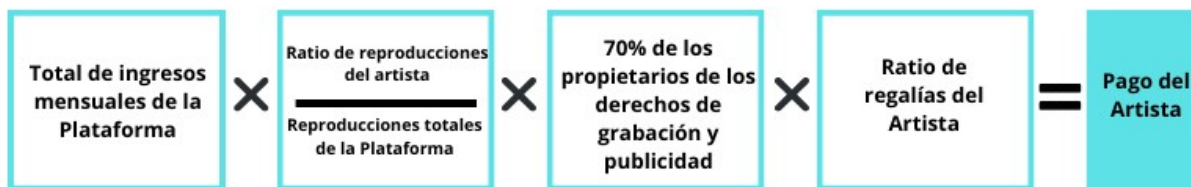
respuesta del capitalismo a la introducción de la era digital; plataformas de streaming como Spotify, Netflix, Disney+, Amazon Prime Video, entre otras, funcionan como mercados lícitos de bienes digitales

Viendo al streaming como un intermediario entre el artista y el consumidor, los artistas pueden agregar su contenido en una variedad de plataformas y redes sociales a través de distribuidores digitales, que, si bien no son canales obligatorios para la publicación de contenido, representan facilidades promocionales y estrategias para garantizar la popularización de los temas. Caicedo (2021), establece que estas compañías actúan como medidores que por un valor generalmente anual o por cada canción o álbum que se suba, cobrará un valor para que este sea agregado a las distintas plataformas para su promoción, además de un porcentaje por las regalías que generen las reproducciones.

En el caso de Spotify, como un ejemplo planteado por Pruvost (2014), la monetización está basada en una modalidad “Freemium”, lo que quiere decir que existe una opción “Free” (gratuita) y una opción “Premium” para usuarios suscriptos, cuyos ingresos provienen de la tarifa anual de usuarios premium y de los ingresos publicitarios de usuarios con cuentas gratuitas. La forma de retribución de derechos se basa en una relación 70/30, de los cuales el 70% de los ingresos corresponde a los titulares de los derechos y un 30% a la plataforma, calculando las regalías mediante una fórmula polinómica que se distribuye como se representa en la Figura 3.

Figura 3

*Fórmula para el cálculo de regalías de artistas en Spotify.*



*Nota.* La figura representa la estructura de la fórmula utilizada generalmente para realizar un cálculo promedio del pago a un artista por las reproducciones de su música en la plataforma de Spotify. Información tomada desde Spotify (2024), y adaptada por la autora.

Sin embargo, el valor final dependerá de las estipulaciones contractuales del artista respectivo con el distribuidor, disquera o participación predispuesta en las leyes aplicables a los convenios realizados en cada país. Como afirma Pruvost (2014), hay que tomar en cuenta que la unidad de medida para el cálculo está representada por las reproducciones (streams), alineando la idea de que la música se percibe como un servicio, más no como un producto, dicha modalidad permite extender ilimitadamente el período de ingresos para los autores.

Algunos factores que influyen de igual manera en el resultado de la monetización de estas plataformas incluyen, la discográfica del artista y especificaciones contractuales, además de otros aspectos que no se controlan, como la cantidad de usuarios premium que escucha la canción, o el país en el que se escucha y el valor de su moneda local. Tomando en cuenta estas consideraciones Caicedo (2021) propuso una representación que se visualiza en la Tabla 2, que representa un aproximado del pago que podría recibir un artista, tomando en cuenta sus reproducciones globales y adaptándolas al contexto nacional del Ecuador en 2024.

**Tabla 2**

*Tasa de ingresos para artistas por servicios de streaming, adaptada a Ecuador.*

<b>Plataforma de Streaming</b>	<b>Reproducciones Para Ganar un dólar</b>	<b>Media de Pago por reproducción</b>	<b>Reproducciones Para Ganar Un salario mínimo</b>
<b>Napster</b>	53	\$ 0,019	23850
<b>Tidal</b>	80	\$ 0,013	36000
<b>Apple Music</b>	136	\$ 0,007	61200
<b>Google Play Music</b>	147	\$ 0,007	66150
<b>Deezer</b>	156	\$ 0,006	70200
<b>Spotify</b>	229	\$ 0,004	103050
<b>Amazon</b>	249	\$ 0,004	112050
<b>Pandora</b>	752	\$ 0,001	338400
<b>YouTube</b>	1449	\$ 0,001	652050

*Nota.* La siguiente tabla representa la cantidad de streams necesaria para ganar 1\$ USD, en las distintas plataformas de streaming musical en el mercado, adaptándola al mínimo de reproducciones para ganar el salario básico en Ecuador. Información tomada desde Caicedo (2021), y adaptada al 2024.

Con el objetivo de predecir las posibles ventas en estas plataformas existen una variedad de modelos que se pueden adaptar a las especificaciones de los datos ofrecidos por las mismas. Según Moreno (2019), la intuición, el conocimiento y la información disponible son requisitos necesarios para intentar establecer pronósticos de ventas, ya que representa una tarea compleja que se facilita cuando se utiliza información histórica. Las técnicas generalmente aceptadas para

la elaboración de estos pronósticos se pueden dividir en cinco categorías, siendo estas: juicio ejecutivo, encuestas, análisis de series de tiempo, análisis de regresión, y pruebas de mercado.

La representación de la relación existente entre dos o más variables a través de un modelo, supone contar con una expresión lógico-matemática que, aparte de resumir esta relación, permita realizar predicciones sobre los valores que pueda tomar una de las variables como lo indican Molina y Rodrigo (2010). En este ámbito el más conocido es el modelo de regresión lineal, siendo el más utilizado a la hora de predecir los valores de una variable cuantitativa, a partir de los valores de otra variable explicativa también cuantitativa, siendo este un modelo de regresión lineal simple, o más de una variable explicativa como en los modelos de regresión lineal múltiple.

Como lo explica Pereira (2010), en un análisis de regresión simple, se pueden visualizar dos tipos de variables; una variable respuesta o dependiente (y) y una variable explicativa o independiente (x), con el propósito de obtener una función sencilla de la variable explicativa que pueda describir lo más ajustadamente posible la posible variación de la variable dependiente. Estas variables vienen representadas o modeladas por la ecuación de una línea recta:

$$\hat{Y} = \beta_0 + \beta_1 * X_1$$

Desde el punto de vista de Núñez et al (2011), el principal supuesto en este caso es la linealidad en la relación ente la variable dependiente, esta debe ser continua, y los predictores; si esto no llegara a cumplirse, se linealiza la relación ya sea transformando la variable o aplicando métodos no paramétricos. En el caso de una regresión lineal múltiple se permite establecer la relación lineal entre variables, definida matemáticamente por un modelo determinístico representado en la siguiente fórmula:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + e$$

Ya que en regresión se trabaja con variables, lo que se hace es calcular promedios y su desviación típica; una vez se calculan ambos se interpretan de manera conjunta, en regresión lineal múltiple suele haber solo una variable endógena con varias variables exógenas, que adoptan dos formas generales, continuas que llenan el espacio y discretas que se mueven a saltos como lo menciona Montero (2016) y se representa en la Tabla 3.

**Tabla 3**

*Modelos de Regresión por tipos de Variables*

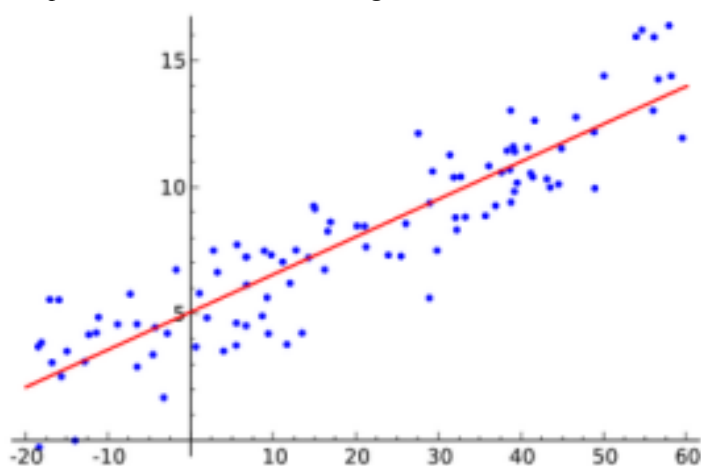
<b>Tipo de Variable</b>	<b>Modelo</b>
Continua	Lineal
Dicotómica	Logit o probit
Recuento	Poisson o Binomial
Factor ordenado	Logit o probit Ordenada
Factor	Logit o probit Multinomial
Porcentaje	Regresión fraccional

*Nota.* La tabla representa los tipos de variables que puedes utilizarse en distintos modelos de regresión, o el modelo recomendable para cada tipo de variable. Información tomada de Montero (2016).

Sin embargo, las variables exógenas solamente exigen una distinta interpretación por lo que basta con el conocimiento de su codificación para identificar los parámetros. Por ejemplo, la representación de una regresión lineal en la Figura 4.

**Figura 4**

*Representación Visual de una regresión lineal*

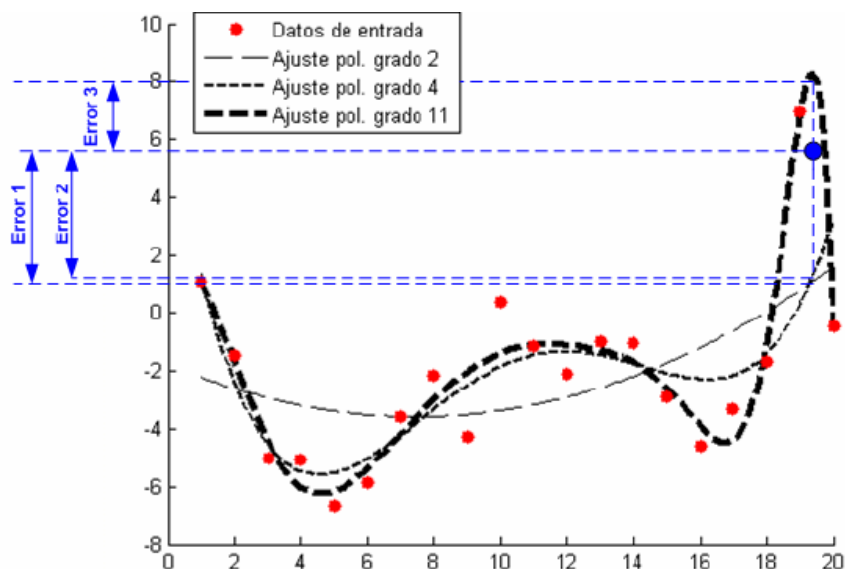


Nasteski (2017), establece que para que los resultados de la regresión sean confiables (insesgados y óptimos), es necesario que la relación entre las variables sea lineal y las perturbaciones de media cero, homocedásticas y no autocorrelacionadas. La regresión lineal, pertenece a una categoría de algoritmos de aprendizaje, es decir que entrenamos el modelo en un conjunto de datos etiquetados (datos de entrenamiento) para predecir datos sin etiquetar (datos de prueba).

Si la relación no es lineal según Pereira (2010), se pueden transformar los valores de una o varias de las variables con la intención de linealizarlas, si esto no fuese posible, se puede comprobar el lado de ajuste de una función polinomial más compleja. Las decisiones no lineales permiten la representación de conceptos más complejos al ajustarse más a los datos, de todos modos, es importante tomar en cuenta que este sobreajuste implica inconvenientes, lo que se conoce instancias de entrenamiento ruidosas que pueden ocasionar una confusión en el sistema de predicción a la hora de predecir nuevas entradas como se evidencia en la Figura 5 con un ajuste polinomial.

Figura 5

*Funciones de ajuste polinomial.*



Otros modelos cuantitativos que se han utilizado tradicionalmente incluyen un método de descomposición en series de tiempo, suavizado exponencial, como autorregresivos y promedio móvil integrado (ARIMA), como lo mencionan, Corres et al (2014). este último quizá es el que mejores resultados ha tenido en términos de desempeño en modelado.

Por otro lado, Martínez (2023) argumenta que se trata de un proceso estocástico, es decir una secuencia de variables aleatorias ordenadas en el tiempo, caracterizados por sus propiedades estadísticas como la media, la varianza y la función de autocorrelación. El proceso de diferenciación integrada autorregresiva (ARIMA), combina un proceso autorregresivo (AR) y un proceso de media móvil (MA), para hacer estacionarias series de tiempo no estacionarias. Consideramos un modelo autorregresivo si la variable endógena de un período (t) es explicada por las observaciones de ella misma correspondientes a periodos anteriores más un término de error, representadas en la siguiente expresión:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + a_t$$

Este error es denominado ruido blanco, que debe cumplir con tres hipótesis básicas tradicionales: una media nula; varianza constante y covarianza nula entre errores correspondientes a observaciones diferentes, representándose de la siguiente fórmula abreviada:

$$\phi_p(L)Y_t = \phi_0 + a_t$$

En donde  $\phi_p(L)$  es el operador polinomial de retardos y el término L es el operador retardo que al aplicar el valor de una variable t, resulte el valor de esa misma variable t-1.

En un modelo de medias móviles, el valor de una determinada variable en un período t se explica en función de un término independiente y una sucesión de errores que corresponden a períodos precedentes, expresados de manera abreviada utilizando un polinomio de retardos:

$$Y_t = \theta_q(L)a_t + \mu$$

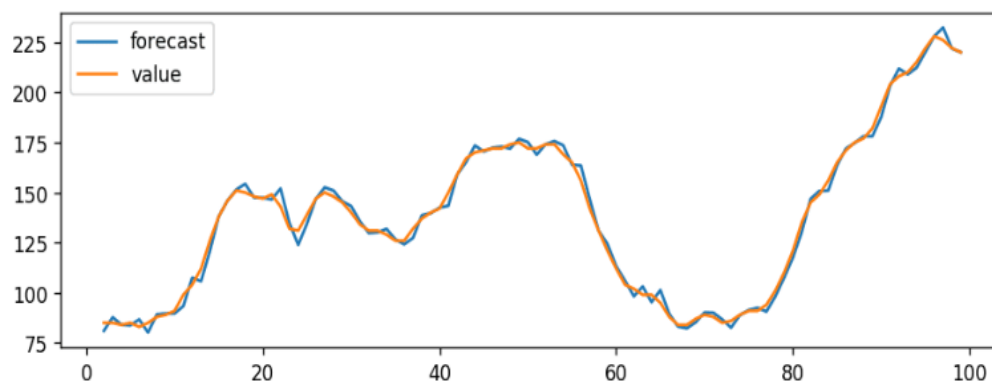
Entonces en su forma más general el modelo ARIMA (p, d, q) podría expresarse:

$$Y_T = \phi_1 Y_{T-1} + \phi_2 Y_{T-2} + \dots + \phi_{p_s+p+d_s+d} Y_{T-p_s-p-sD-d} + \\ + \delta + U_T + \theta_1 U_{T-1} + \dots + \theta_{q_s+q} U_{T-sQ-q}$$

Entendiendo que existe más de un proceso generador de la serie y representando una combinación entre los modelos AR (p) y MA(q) que precisaron una serie de diferenciaciones (d) para ser estacionarios, en una combinación que se representa en la Figura 6.

### Figura 6

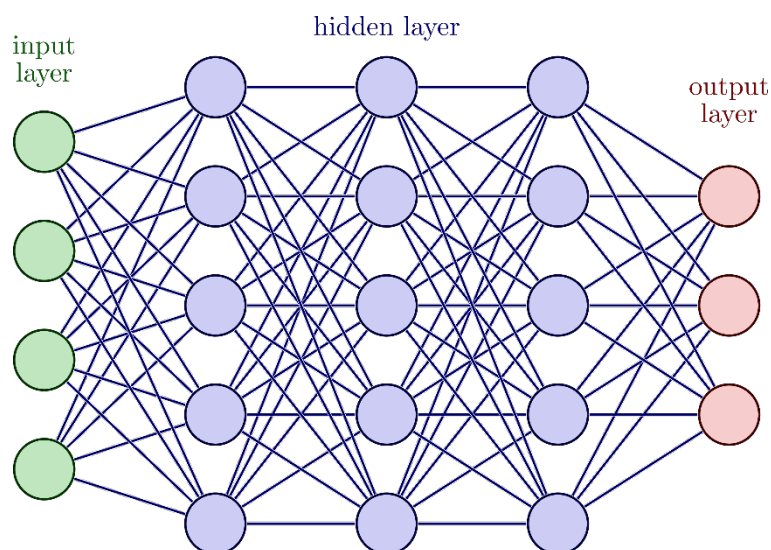
*Representación Gráfica de un modelo ARIMA ajustado*



*Nota.* En el gráfico se pueden evidenciar las dos tendencias del modelo que representan los valores ajustados de los datos en color naranja, mientras que en color azul se siguen las predicciones resultantes del mismo. Información obtenida de Gonzáles (2022).

Por otro lado, asumir que una serie de tiempo es lineal, no sería razonable ya que en el mundo real los sistemas suelen ser no lineales; viéndolo desde esta perspectiva las Redes Neuronales que funcionan como un modelo computacional inspirado en las funciones del cerebro humano, se han aplicado como una herramienta atractiva en el área de pronósticos. Siguiendo este concepto Ruelas Santoyo y Laguna Gonzáles (2014) mencionan que, una red neuronal artificial, se trata de un procesador masivo paralelo que se forma por unidades simples de procesamiento que éstas propensas naturalmente a almacenar conocimiento experimental.

Su ventaja con respecto a la regresión según Plaza (2023), es que son muy útiles para encontrar relaciones no lineales entre variables como se observa en la Figura 7, mientras más variables independientes se presenten, mejor será su funcionamiento. Estas redes permiten modificar parámetros estudiando que opción es la mejor para encontrar el modelo óptimo para minimizar el error, parámetros como “size” (número de nodos), y “decay” (parámetro que ajuste el algoritmo de optimización).

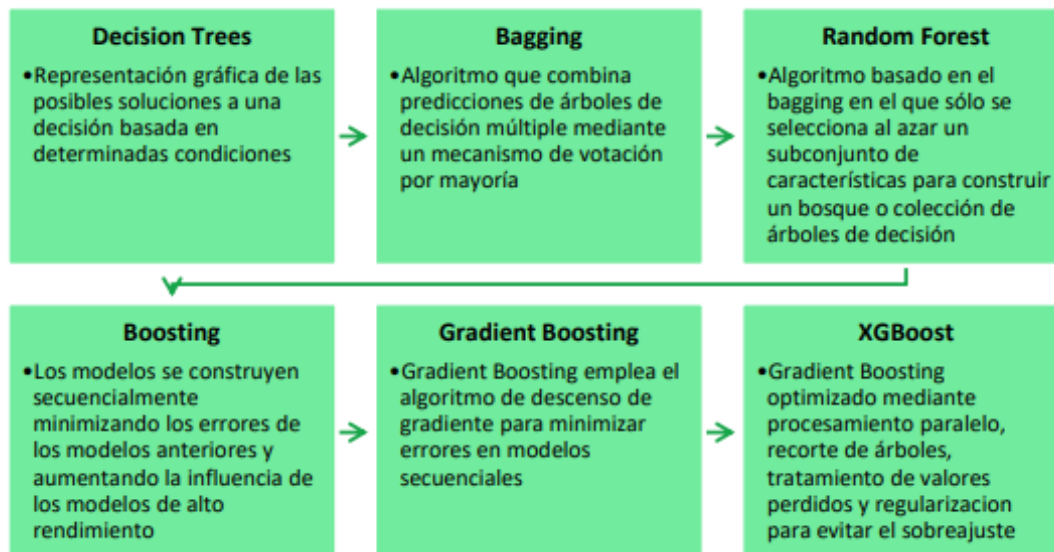
**Figura 7***Estructura de una red neuronal*

*Nota.* El gráfico representa el funcionamiento de una red neuronal, relacionando las variables para encontrar patrones formando nodos. Información obtenida de Nautelings (2021).

Cada neurona calculará una suma ponderada de sus entradas para aplicarle una función de activación, lo cual permite resolver los problemas complejos que surgen de la no linealidad de las variables. Como sugiere Palacios (2020), otra alternativa de análisis también se consideran los árboles de regresión, los cuales son utilizados en la predicción de variables continuas de un conjunto de datos, que son divididos en múltiples particiones de acuerdo con un criterio establecido con el objetivo de obtener predicciones con el error más pequeño posible.

**Figura 8**

*Evolución de los algoritmos de árboles de regresión.*



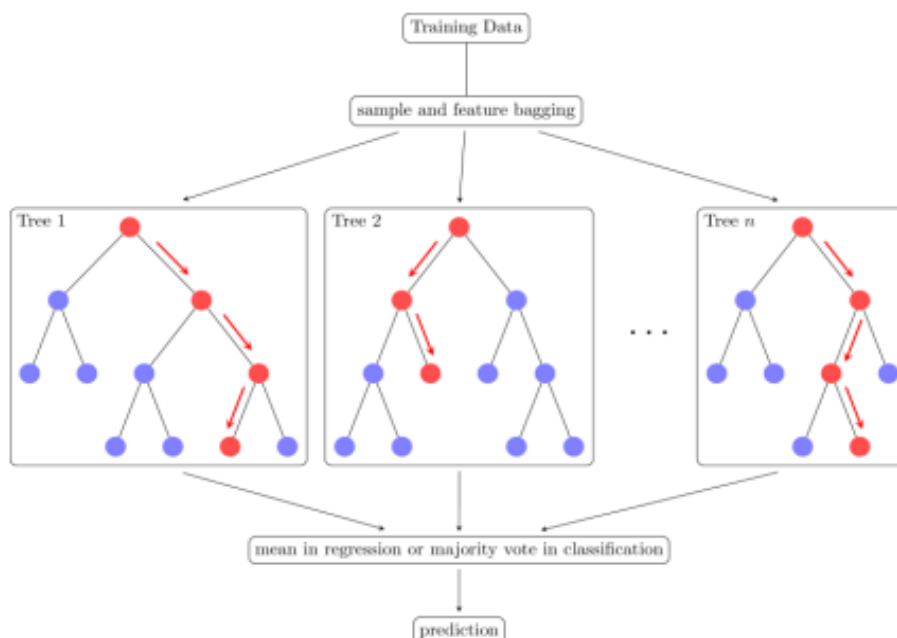
*Nota.* La figura muestra los diferentes modelos de árboles de regresión con las especificaciones de sus variables y algoritmos que se pueden aplicar. Obtenido de Plaza (2023).

De los árboles de regresión, Plaza (2023) sugiere también que el Random Forest resulta un modelo atractivo por su capacidad de manejar un gran número de variables con pocas observaciones. En este método se toman  $N$  submuestras aleatorias con remplazamiento del conjunto de datos original, lo que garantiza la independencia de los árboles, reduciendo el riesgo

de sobreajuste gracias a que aumenta la variabilidad en el modelo como se representa en la Figura 9.

**Figura 9**

*Representación de un modelo Random Forest.*



En una dirección similar Huang y Pai (2020), utilizaron una regresión vectorial de compatibilidad con mínimos cuadrados para pronosticar las ventas de películas con datos de Twitter y bases de datos de películas, alegando que debido a la rápida prominencia y popularidad de las redes sociales, las redes de difusión social con intercambio voluntario de información se han convertido en una de las formas más poderosas de difundir opiniones y por lo tanto se han vuelto más importantes al momento de pronosticar las ventas de productos. Indicando que los resultados del uso de LSSVR (Least Squares Vector Regression) con GA puede dar como resultado el mejor rendimiento de pronóstico en términos de precisión de predicción al utilizar cuatro modelos distintos como se contempla en la Tabla 4.

**Tabla 4***Mediciones Medias de la Predicción*

Fuentes de datos	Mediciones de precisión de pronóstico	Modelos de pronóstico			
		LSSVR	BPNN	GRNN	MLR
Bases de datos de películas	RMSE	33.84	308.86	260.86	266.87
	MAPE(%)	2.46	36.98	28.97	30.89
Tweets	RMSE	63.75	342.28	436.24	374.65
	MAPE(%)	6.18	40.51	49.83	40.12
Datos híbridos	RMSE	4.72	316.98	327.29	331.41
	MAPE(%)	0.53	36.49	39.31	35.72

*Nota.* La siguiente Tabla representa las mediciones medias de predicción generadas por cuatro modelos con tres tipos de datos y validación cruzada de 10 veces.

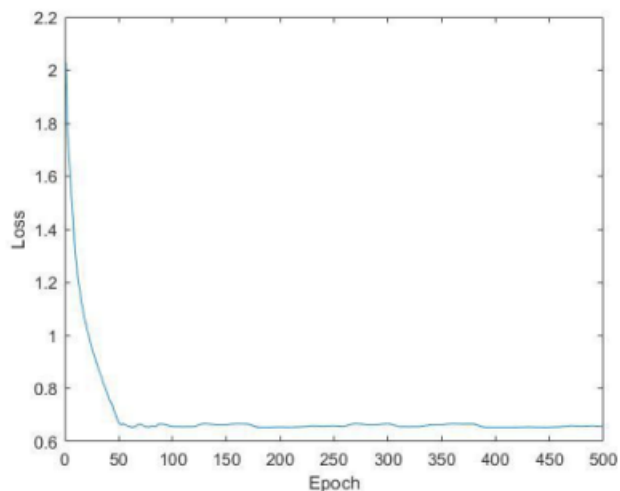
Otra investigación que se relaciona es un sistema de recomendación de música basado en la emoción mediante el uso de diferentes enfoques de machine learning, en la cual Vidyasagar y Karawande (2021) plantean que en los últimos años los algoritmos de recomendación se han vuelto cada vez más populares en las plataformas comerciales para proporcionar a los usuarios recomendaciones personalizadas; recomendando un modelo de dos partes que integra un algoritmo de recomendación que funciona de manera predictiva y filtrado colaborativo.

En primer lugar un enfoque de representación de características basado en regresión polinómica obteniendo características latentes y un modelo de red neuronal profunda para

predecir puntuaciones de calificación, aumentando el rendimiento de las recomendaciones obteniendo los resultados representados en la Figura 10.

**Figura 10**

*Curva de Función de Pérdida*



*Nota.* La figura muestra el proceso de visualización después del entrenamiento de la red, con una precisión del 60%, después de múltiples pruebas de ajuste de parámetros.

Un estudio de la Universidad de Osaka en Japón, estudia la predicción de ventas de artículos de comercio electrónico en un entorno web, en el cual Zhang y Hara (2021) proponen un método de dos pasos, generar las incrustaciones para el producto y el fondo temporal, y luego se predicen las ventas comparando estas incrustaciones. Sugieren distintas técnicas posibles de aprendizaje automático aplicables como la regresión lineal o las máquinas de vectores de soporte, dado a que se esperaba una relación no lineal, se escogió el Random Forest como modelo de aprendizaje, se entrenó el modelo utilizando un conjunto de datos de prueba,

produciendo una predicción del número de ventas para cada producto como se representa en la Tabla 5.

**Tabla 5**

*Precisión Media de los Métodos de Predicción*

	random	RF	item	tweet	item	tweet	item	tweet
		item	now	now	P1	P1	RNN	RNN
Recall@50	0.055	0.053	0.108	0.090	0.065	0.083	0.063	0.070
Recall@100	0.111	0.140	0.185	0.158	0.155	0.158	0.158	0.165
mAP@50	0.001	0.005	0.013	0.007	0.010	0.006	0.009	0.005
mAP@100	0.002	0.008	0.017	0.010	0.014	0.009	0.014	0.009

*Nota.* Los resultados de la tabla señalan que los métodos de predicción son más efectivos que el método aleatorio, indicando que el fondo temporal logra una mejor precisión de predicción.

En relación con la investigación Huang y Pai (2020) sugieren que la expansión en la recopilación de datos tanto estructurados como no estructurados para mejorar el rendimiento de la previsión puede ser una posible dirección. En el caso de los datos estructurados, como pueden ser las bases de datos de plataformas de streaming se podrían incluir indicadores económicos mundiales, pero en el aspecto de los datos no estructurados, se podrían recopilar otras redes sociales, como Instagram, Facebook y YouTube, para pronosticar el rendimiento. Señalando además que es crucial la eficacia de los datos de las redes sociales sobre la mejora de la precisión de las previsiones para diferentes problemas.

## **2. MATERIALES Y MÉTODOS**

### **2.1. Tipo de Investigación**

La presente investigación se fundamenta en una metodología cuantitativa, empleando técnicas estadísticas para recopilar información pertinente para la construcción de un modelo de regresión lineal, con un enfoque predictivo. Este análisis se centra en los datos obtenidos de plataformas de streaming, principalmente Spotify, siendo la más utilizada en este ámbito, complementada con información de otras plataformas de menor impacto en el contexto musical. El objetivo es establecer un pronóstico de las ventas mediante la difusión musical de artistas ecuatorianos en estas plataformas.

El estudio tiene un alcance descriptivo cuantitativo, ya que se detallan las características más relevantes de la industria musical. El carácter no experimental de la investigación asegura que la base de datos contiene información inalterada, obtenida directamente de las plataformas.

### **2.2. Enfoque**

El enfoque cuantitativo se emplea para medir y analizar datos objetivos, permitiendo la identificación de patrones y la construcción del modelo. Por su parte el enfoque predictivo se utiliza para anticipar futuros patrones de consumo mediante el análisis de datos históricos, facilitando proyecciones precisas de las ventas de contenido musical.

### **2.3. Modelo**

Al centrarnos en el objetivo de la predicción de ventas en las plataformas de streaming, la variable dependiente debe ser continua. El modelo por construir será de regresión, y dado que diversos factores influyen en la popularidad de un artista y, por ende, impactan en sus ventas, el modelo adecuado será una regresión lineal múltiple. Este enfoque permite identificar patrones de

comportamiento tomando en cuenta las distintas variables que afectan el rendimiento de los artistas en las plataformas de streaming.

#### **2.4.Variables**

Para la construcción de un modelo de regresión se toman en cuenta dos tipos de variables:

- ***Variable dependiente:*** Representa el resultado que se busca predecir y está directamente influenciada por diversas características del contenido y del mercado, en esta investigación la proyección de ventas de artistas ecuatorianos en plataformas de streaming.
- ***Variables independientes:*** Son los factores que influyen en la variable dependiente, estas variables se seleccionaron con base en su capacidad para explicar las fluctuaciones en las ventas de contenido musical.

#### **2.5.Indicadores**

Algunos de los indicadores que se tomarán en cuenta como variables independientes son los siguientes, centrándose en Spotify como plataforma principal utilizada en la industria musical, acompañada de otras redes sociales que influyen en la popularidad del artista:

- ***Seguidores en Spotify:*** Es una variable continua que representa el recuento de usuarios de la plataforma Spotify que siguen al artista y reciben notificaciones regulares sobre su actividad hasta abril de 2024. Además, es la métrica principal para clasificar al artista en el top 50, donde la posición número 1 corresponde al mayor número de seguidores y así sucesivamente.

- ***Oyentes Mensuales en Spotify:*** Variable continua que representa el número de veces que la música del artista ha sido reproducida en la plataforma por un consumidor durante el mes, también conocido como “streams”.
- **Popularidad en Spotify:** Variable continua medida en una escala de 0 al 100, señalando el grado de popularidad del artista en la plataforma en el mes. Se calcula principalmente a partir de la cantidad de reproducciones en las pistas de un artista considerando la actualidad de esos datos.
- **Suscriptores en YouTube:** Variable continua que representa la cantidad de consumidores registrados en el canal oficial del artista en la plataforma, directamente relacionado con las vistas que puede generar este en la misma mensualmente.
- **Vistas en YouTube:** Variable continua, similar al concepto de streams representa el número de reproducciones de un video en la plataforma de manera mensual.
- **Seguidores en Instagram:** variable continua que representa el recuento de usuarios de la red social que siguen la cuenta oficial del artista mensualmente.
- **Seguidores en TikTok:** Variable continua que representa la cantidad de usuarios de la plataforma que siguen la cuenta oficial registrada del artista por mes.
- **Fans en Deezer:** Variable continua que Representa la comunidad o la cantidad de consumidores que siguen al artista a través de esta plataforma de difusión musical (similar a los suscriptores), mensualmente.
- **Fans en Facebook:** Variable continua asociada al número de usuarios vinculados a la comunidad del artista en su página oficial de Facebook (similar a los seguidores) mensualmente.

- **Seguidores en SoundCloud:** Variable continua que indica el Número de usuarios de la plataforma que siguen a la cuenta oficial del artista.
- **Porcentaje de Crecimiento:** Variable continua que indica la medida porcentual mensual de crecimiento o de decrecimiento experimentado por el artista en un mes desde la última actualización en abril de 2024.

## 2.6. Instrumento

Para la selección de variables, se utilizó la herramienta SOUNDCHARTS, una plataforma de inteligencia de mercados que proporciona información en tiempo real sobre el mercado musical, abarcando todas las áreas de la carrera del artista, desde redes sociales y streaming hasta la radio (Soundcharts, 2024). Esta plataforma organiza la información musical global y la hace universalmente accesible y procesable. SOUNDCHARTS ofrece más de 25 métricas de evaluación, y registra datos actualizado diariamente para más de 8.000.000 de artistas con redes sociales activas.

Al trabajar con el modelo SaaS (Software como Servicio), está construida para la industria de la música y puede utilizarse en las siguientes funciones (Weiszfeld, 2024):

- Supervisar el rendimiento de los lanzamientos de canciones
- Auditar informes de difusión
- Medir el potencial del artista.
- Impacto de las campañas de Marketing
- Localizar oportunidades de crecimiento e identificar mercados de alto potencial
- Optimizar las estrategias en las listas de reproducción.

Tomando en cuenta esta información, escogemos los valores centrándonos específicamente en el mercado ecuatoriano y los artistas presentes en estas plataformas.

Para la realización de las pruebas de normalidad se utilizó el complemento de Excel Real Statistics. Se trata de una herramienta que amplía las capacidades analíticas de Excel, lo que permite al usuario realizar análisis complejos sin la necesidad de un software especializado. Entre sus funcionalidades más destacadas se encuentra la prueba Shapiro-Wilk, la cual es esencial para evaluar la normalidad de los datos, un supuesto crucial en la construcción de modelos de regresión lineal.

Esta prueba permite determinar si los datos de las ventas de música en plataformas de streaming se distribuyen normalmente, lo cual es vital para asegurar la validez de los modelos predictivos. El uso de Real Statistics, como se menciona en el Manual Real Statistics Using Excel (2023), facilita la ejecución de esta prueba directamente en Excel, ahorrando tiempo y simplificando el análisis.

## **2.7.Muestra**

Al filtrar la información en Soundcharts y especificar el bloque de artistas por país, se encontraron un total de 1.320 artistas registrados en Ecuador. Sin embargo, debido a las restricciones de acceso según los planes de pago de la plataforma, se consideraron los primeros 50 artistas con los índices de reproducción más representativos (véase Anexo 1). Se recopiló información de estos artistas en diversas plataformas, no solo de streaming musical, como Spotify, YouTube, Deezer y SoundCloud, sino también en redes sociales como Instagram, Facebook y TikTok. Siguiendo la lógica presentada por Franklin (2015), la inclusión de seguidores en redes sociales se debe a que la mayoría de la audiencia se congrega naturalmente

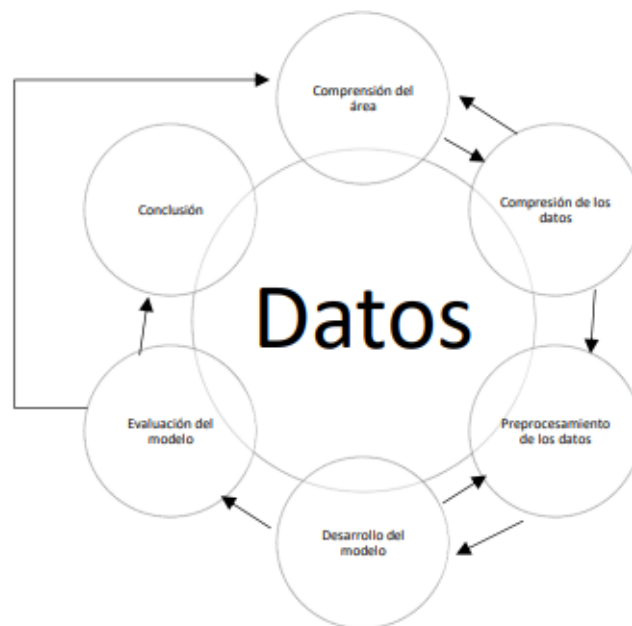
en estas plataformas, creando comunidades y reciprocidad que eventualmente se reflejan en las ventas.

## 2.8. Técnicas

La técnica utilizada fue la minería de Datos, la cual se enfocó en el descubrimiento de patrones, relaciones y tendencias de grandes conjuntos de datos, resultando útil en la toma de decisiones y la generación de conocimiento. Esta técnica permitió recolectar información de las distintas bases de datos de las páginas de streaming. El proceso se llevó a cabo siguiendo los pasos tradicionales de la aplicación de minería de datos, que se especifican en la figura 11:

**Figura 11**

*Proceso de Minería de Datos*



Nota. La figura muestra el proceso básico de una minería de datos para obtener los óptimos resultados de aplicación. Tomado de Palacios (2020).

Este enfoque sistemático permitió extraer información valiosa y desarrollar un modelo de regresión lineal múltiple que anticipara con precisión los patrones de consumo y las ventas de contenido musical. Ya que la técnica de minería de datos se interpreta como un proceso, el instrumento para analizar la información se realizó de acuerdo con la metodología SEMMA (Sample, Explore, Modify, Model and Asses) por sus siglas en inglés, una metodología que divide el proceso en cinco etapas:

- **MUESTRA:** extraer una muestra estadísticamente representativa de datos.
- **EXPLORAR:** aplicar técnicas explicativas, estadísticas y de visualización.
- **MODIFICAR:** para seleccionar y transformar las variables predictivas más significativas.
- **MODELAR:** modelar las variables para predecir resultados.
- **EVALUAR:** para confirmar la precisión del modelo.

## **2.9.Procedimiento de Aplicación**

### **2.9.1. Muestreo**

En primer lugar, antes de la aplicación de un modelo, se seleccionaron las variables por distintos motivos, tales como mejorar la precisión de la predicción eliminando aquellas variables sin relevancia y evitar los sobreajustes. También se buscó describir con parsimonia el conjunto de datos, ajustándolos con la menor cantidad de variables posibles y emplear un menor conjunto de variables predictoras para reducir el esfuerzo computacional en la práctica (Plaza Chicote, 2023). Es decir, se tomaron en cuenta únicamente aquellas variables con un alto nivel de significancia o impacto en la variable dependiente.

### **2.9.2. Exploración**

Para la selección de las variables se realizó una filtración del primer sondeo de datos, eliminando aquellos que no poseían la información suficiente para proporcionar un resultado confiable. Este fue el caso de las Vistas en YouTube, Seguidores en TikTok, Seguidores en SoundCloud y Fans en Deezer, donde no se registraba una cantidad representativa de información o datos suficientes para influir en la predicción del modelo (véase Anexo 2). Al finalizar este proceso, se seleccionaron las siguientes variables:

- **Seguidores en Spotify**
- **Oyentes Mensuales en Spotify**
- **Popularidad en Spotify**
- **Suscriptores en YouTube**
- **Seguidores en Instagram**
- **Fans en Facebook**

Para la construcción del modelo, se pretendió predecir las ventas de artistas mediante su participación en plataformas de streaming, siendo los “streams” o la cantidad de reproducciones lo que se traduce en ganancias para el artista. Es decir, la variable que se buscó predecir fue la cantidad de streams en las plataformas (Oyentes mensuales). Por lo tanto, las variables predictoras seleccionadas fueron aquellas que mejor podían explicar el número de oyentes mensuales, considerando la presencia restante del artista en las diferentes plataformas y redes sociales.

### 2.9.3. Modificación

Para validar la eficiencia del modelo, se aplicaron medidas de evaluación para verificar su precisión, teniendo en cuenta la minimización del error de clasificación de la variable de valoración (Núñez et al., 2011). Estas medidas incluyeron:

- **Error Cuadrático Medio (MSE):** es el promedio de las diferencias al cuadrado entre los valores reales y los valores predichos en un conjunto de datos. Este indicador evalúa la dispersión de los errores, mientras más bajo indica un mejor ajuste del modelo. Se calcula con la siguiente fórmula:

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2$$

Sumando los cuadrados de las diferencias entre los valores reales y los predichos y dividiendo por el número total de observaciones.

- **Raíz del Error cuadrático medio (RMSE):** Como dice su nombre se trata de la raíz cuadrada del MSE y mide la desviación estándar de los errores. Suele preferirse su uso sobre el MSE debido a su interpretación más sencilla ya que se expresa en las mismas unidades que la variable objetivo y se expresa de la siguiente manera:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2}$$

- **Coefficiente de determinación ( $R^2$ ):** Representa la proporción de la varianza en la variable dependiente que es explicada por el modelo de regresión lineal.

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\sum (y_i - \hat{y})^2}{\sum (y_i - \bar{y})^2}$$

Este coeficiente es adimensional y se encuentra siempre entre 0 y 1. Mientras los valores se acerquen más a 1 se indica un mejor ajuste del modelo expresado con la fórmula:

- **Error Absoluto Medio (MAE):** Se trata del promedio de las diferencias absolutas entre los valores reales y los valores predichos en un conjunto de datos, es decir que mide la magnitud promedio de los errores, y se representa con la siguiente fórmula:

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

Regresando a la teoría de la Regresión lineal múltiple, como afirman Vilá Baños et al (2019), los datos deben cumplir 5 supuestos: Linealidad, Independencia, Homocedasticidad, Normalidad y No colinealidad.

- **Linealidad:** que la relación de las variables sea lineal.
- **Independencia:** Que los errores en la medición de las variables explicativas sean independientes entre sí.
- **Homocedasticidad:** Que los errores tengan varianza constante.
- **Normalidad:** Que las variables sigan la Ley Normal.
- **No colinealidad:** Que las variables independientes no estén correlacionadas entre sí.

#### 2.9.4. *Modelación*

La regresión lineal múltiple como mencionan Vilá Baños et al (2019), analiza la relación entre una variable dependiente métrica y varias variables independientes también métricas, para predecir una única variable dependiente, recordando la fórmula:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + e$$

Como primer paso se seleccionó la variable dependiente que se había establecido previamente: las ventas de artistas ecuatorianos en plataformas de streaming, que se deseaba pronosticar. Esta variable se traduce en los streams del artista, específicamente la cantidad de oyentes mensuales en Spotify, identificada como la principal plataforma utilizada en este contexto.

En segundo lugar se seleccionaron las variables independientes o explicativas que se detallaron en el proceso de exploración. Estas variables incluyen Seguidores en Spotify, Popularidad en Spotify, Suscriptores en YouTube, Seguidores en Instagram y Fans en Facebook. Estas fueron especificadas en la sección de exploración y representadas numéricamente en el Anexo 2.

Como siguiente paso se utilizó el complemento de Excel Real Statistics para verificar la normalidad de los datos de las variables seleccionadas, recordando que este es uno de los supuestos del modelo. Se aplicó la prueba de Shapiro-Wilk, que, según la conceptualización planteada por Malato (2023), es una prueba de hipótesis que evalúa si un conjunto de datos sigue una distribución normal. Evalúa los datos de una muestra con la hipótesis nula de que el conjunto de datos se distribuye normalmente. Si el valor de p resultante es bajo, se rechaza la hipótesis nula y se puede concluir que la muestra no se ha generado a partir de una distribución normal.

En este caso se obtuvieron los siguientes resultados:

**Tabla 6:**

*Resultados de la Prueba de Normalidad Shapiro-Wilk*

	OYENTES MENSUALES// CONSUMIDOR	SEGUIDORES # DE PERSONAS	POPULARIDAD %	SUSCRIPTORES// CONSUMIDOR # DE PERSONAS	SEGUIDORES	FANS
	<i>Grupo 1</i>	<i>Grupo 2</i>	<i>Grupo 3</i>	<i>Grupo 4</i>	<i>Grupo 5</i>	<i>Grupo 6</i>
<b>W-STAT</b>	0,96662815	0,96662815	0,95597731	0,96662815	0,88935216	0,95813967

<b>P-VALUE</b>	0,16821558	0,16821558	0,06033549	0,16821558	0,00021782	0,07429216
<b>ALPHA</b>	0,05	0,05	0,05	0,05	0,05	0,05
<b>NORMAL</b>	yes	yes	yes	yes	no	yes

*Nota.* La tabla evidencia que los resultados obtenidos durante la prueba, fueron favorables en la mayoría de las variables seleccionadas con excepción de una, por lo que el modelo es aplicable en este conjunto de datos.

Al comprobar la normalidad de los datos, se procedió a realizar el análisis de regresión utilizando la herramienta de Excel de Análisis de datos con las siguientes especificaciones:

**Figura 12**

*Especificaciones para la aplicación de Regresión Lineal mediante análisis de datos.*

*Nota.* La figura representa las especificaciones para obtener la ecuación de regresión lineal múltiple, con un nivel de confianza del 95%. También se especifican los residuales que se desean analizar para verificar los supuestos del modelo, al igual que el gráfico de probabilidad normal.

A continuación, se presenta la tabla de resumen de regresión lineal obtenida mediante el análisis de datos realizado en Excel con las especificaciones señaladas. Esta tabla resume los resultados del modelo de regresión lineal aplicado para predecir las ventas de artistas ecuatorianos en plataformas de streaming, utilizando como variable dependiente el número de oyentes mensuales en Spotify y las variables independientes ya establecidas y seleccionadas en base a su capacidad explicativa potencial.

**Tabla 7**

*Resumen de Resultados de regresión*

<b>ESTADÍSTICAS DE LA REGRESIÓN</b>	
<b>COEFICIENTE DE CORRELACIÓN MÚLTIPLE</b>	0,88149744
<b>COEFICIENTE DE DETERMINACIÓN R<sup>2</sup></b>	0,77703774
<b>R<sup>2</sup> AJUSTADO</b>	0,75170112
<b>ERROR TÍPICO</b>	289712,316
<b>OBSERVACIONES</b>	50

Al analizar los resultados de regresión, en primera instancia se revela un coeficiente de correlación múltiple de 0,8815 aproximadamente, indicando una fuerte relación entre las variables independientes y la variable dependiente. Este alto valor sugiere que el modelo es eficaz para capturar la variabilidad en la cantidad de oyentes mensuales a partir de las variables predictoras seleccionadas.

El coeficiente de determinación  $R^2$  es de 0,7770, lo que implica que aproximadamente el 77,7% de la variabilidad en la cantidad de oyentes mensuales se puede explicar mediante el modelo de regresión lineal múltiple. El  $R^2$  ajustado, que tiene en cuenta el número de variables predictoras y el tamaño de la muestra, es de 0,7517, confirmando la robustez del modelo al ajustar los datos, por el posible sobreajuste. El error típico de 289712,32 sugiere que, aunque el

modelo es considerablemente preciso, todavía existe un margen para mejorar la precisión de las predicciones, sin embargo. Con el total de 50 observaciones, los resultados proporcionan una base sólida para concluir que las variables seleccionadas son buenas predictoras de la popularidad y ventas de los artistas ecuatorianos en plataformas de streaming.

Se presenta a continuación el análisis de varianza (ANOVA) que permite evaluar la significancia global del modelo al comparar la variabilidad explicada por el modelo con la variabilidad residual. Este análisis cobra importancia para determinar si las variables independientes seleccionadas contribuyen significativamente a la predicción de la variable dependiente, una vez ajustados los datos. Los resultados de ANOVA proporcionan una visión detallada de la adecuación del modelo y la importancia de las variables explicativas utilizadas y resumidas en la siguiente tabla:

**Tabla 8**

*Análisis de Varianza (ANOVA)*

	<b>GRADOS DE LIBERTAD</b>	<b>SUMA DE CUADRADOS</b>	<b>PROMEDIO DE LOS CUADRADOS</b>	<b>F</b>	<b>VALOR CRÍTICO DE F</b>
<b>REGRESIÓN</b>	5	1,2871E+13	2,5741E+12	30,6685637	2,6964E-13
<b>RESIDUOS</b>	44	3,6931E+12	8,3933E+10		
<b>TOTAL</b>	49	1,6564E+13			

Los resultados arrojados por el ajuste son significativos, indicando la eficacia del modelo en predecir la variable dependiente. El valor de F de 30,67, significativamente mayor que el valor crítico de F (2.69E-13), indica que el modelo de regresión lineal múltiple propuesto, explica una proporción significativa de la variabilidad de los oyentes mensuales Spotify, más allá de lo que se esperaría por azar. La suma de cuadrados de la regresión (1,29E+13) y la suma de

cuadrados de los residuos ( $3,69E+12$ ) reflejan que una parte considerable de la variabilidad total es explicada por el modelo.

El promedio de los cuadrados de la regresión y de los residuos proporcionan medidas adicionales de la varianza explicada y no explicada respectivamente. Reafirmando la robustez y significancia estadística del modelo. Con los antecedentes mencionados en la siguiente tabla se resumen los parámetros tomados en cuenta para la construcción final del modelo:

**Tabla 9**

*Coefficientes del Modelo de regresión*

	<i>COEFICIENTES</i>	<i>ERROR TÍPICO</i>	<i>ESTADÍSTICO T</i>	<i>PROBABILIDAD</i>	<i>INFERIOR 95%</i>	<i>SUPERIOR 95%</i>	<i>INFERIOR 95,0%</i>	<i>SUPERIOR 95,0%</i>
<b>INTERCEPCIÓN</b>	-844476,48	199118,74	-4,24	0,011249%	-1245773,94	-443179,03	-1245773,94	-443179,03
<b>SEGUIDORES</b>	1,55	0,24	6,57	0,000005%	1,07	2,03	1,07	2,03
<b>POPULARIDAD</b>	28668,98	5304,89	5,40	0,000251%	17977,66	39360,29	17977,66	39360,29
<b>SUSCRIPTORES// CONSUMIDOR</b>	-0,07	0,08	-0,96	34,31%	-0,23	0,08	-0,23	0,08
<b>SEGUIDORES</b>	0,02	0,02	1,00	32,26%	-0,02	0,05	-0,02	0,05
<b>FANS</b>	-0,18	0,20	-0,89	37,92%	-0,59	0,23	-0,59	0,23

Con los coeficientes obtenidos la ecuación del modelo se plantea de la siguiente manera:

$$Y = -844476,48 + 1,55 X_1 + 28668,98 X_2 - 0,07X_3 + 0,02X_4 - 0,18 X_5$$

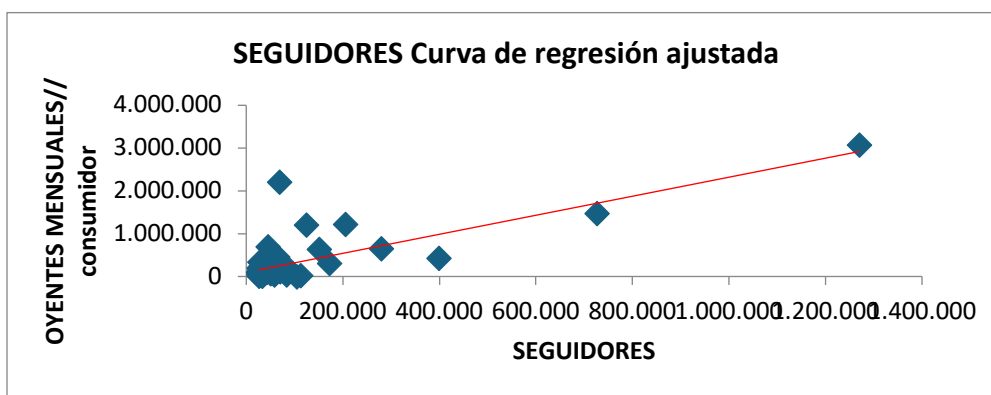
### 2.9.5. Evaluación

La evaluación de la pertinencia del modelo de regresión lineal obtenido se llevó a cabo mediante la verificación de sus supuestos fundamentales asegurando su validez y fiabilidad.

Como primer supuesto se evaluó la linealidad entre las variables predictoras y la variable dependiente, para analizarlo se tomaron en cuenta los gráficos de la curva de regresión ajustada especificados en la herramienta de análisis de datos al momento de realizar la regresión obteniendo las siguientes Gráficas:

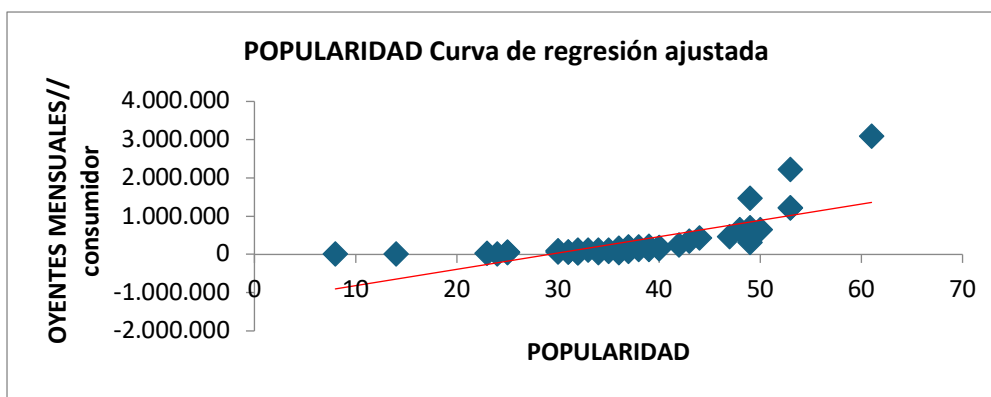
**Figura 13**

*Prueba de linealidad de la variable explicativa: Seguidores en Spotify*



**Figura 14**

*Prueba de linealidad de la variable explicativa: Popularidad en Spotify*



**Figura 15**

*Prueba de linealidad de la variable explicativa: Suscriptores en YouTube*

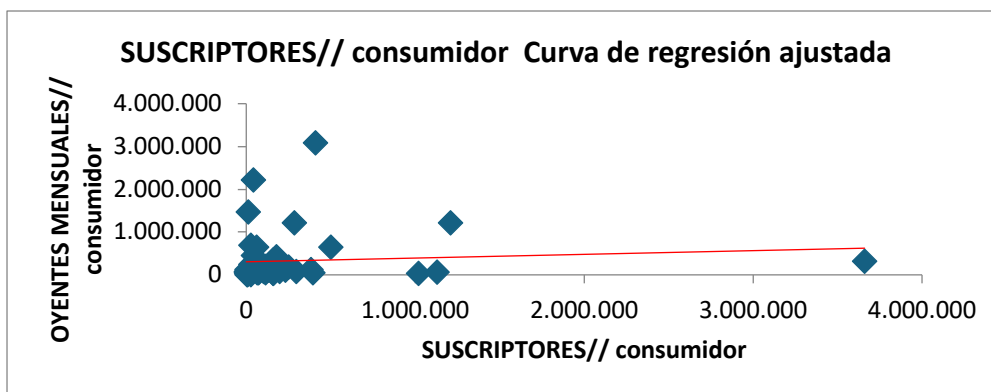


Figura 16

*Prueba de linealidad de la variable explicativa: Fans en Facebook*

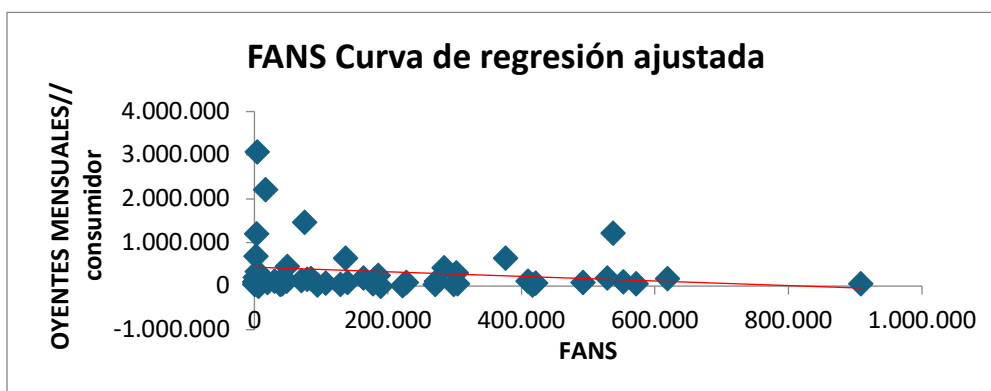
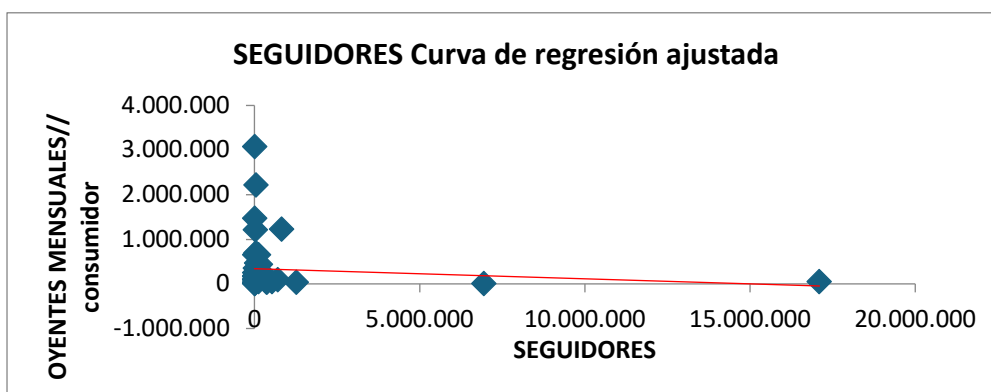


Figura 17

*Prueba de linealidad de la variable explicativa: Seguidores en Instagram*



Los gráficos de regresión ajustada presentados ilustran la relación lineal entre las variables predictoras y la variable dependiente. La presencia de la línea de tendencia resaltada en

rojo permite evaluar visualmente la adecuación del supuesto de linealidad, interpretándose de la siguiente manera:

- ***Seguidores en Spotify:*** el primer gráfico, como bien se muestra en la Figura 13, muestra una tendencia ascendente clara, indicando una relación lineal positiva entre seguidores en Spotify y los oyentes mensuales, lo cual es consistente con los resultados de la regresión que muestran significancia estadística para esta variable.
- ***Popularidad en Spotify:*** este gráfico representado en la figura 14, revela de igual manera una tendencia ascendente notable, sugiriendo una fuerte relación lineal positiva entre las variables. Esta significancia concuerda con la observación del modelo.
- ***Suscriptores en YouTube:*** aunque la dispersión observada es considerable como se evidencia en la figura 15, la línea de ajuste no presenta una tendencia pronunciada, indicando una relación lineal débil que no llega a ser inexistente, resultando consistente con la significancia que esta variable representa en el modelo, relativamente baja en comparación a las anteriores sin llegar a ser nula.
- ***Fans en Facebook:*** Como se evidencia en el gráfico 16, existe una amplia dispersión y una línea de ajuste casi horizontal, indicando una falta de relación lineal significativa que no llega a ser totalmente nula y se explica de igual manera en los resultados de probabilidad del modelo.
- ***Seguidores en Instagram:*** Similar al gráfico anterior la figura 17, evidencia una relación lineal débil, que no llega a ser nula representando una constante, que determina lo explicado por el modelo.

De manera general los gráficos de regresión ajustada apoyan el supuesto de linealidad para las variables que corresponden a la misma plataforma, mientras que las variables restantes, aunque presentan menor significancia, su tendencia no es completamente nula, por lo que el supuesto se cumple de acuerdo con la significancia estadística de cada variable explicativa.

El siguiente supuesto se trata de la interdependencia, implicando que los errores en la medición de las variables explicativas sean interdependientes entre sí ya que las diferentes observaciones de una misma variable pueden crear inercia en el sistema aproximando los valores y disminuyendo su capacidad predictiva. Para ello analizamos el gráfico obtenido de los residuales con la cantidad de observaciones obteniendo el siguiente resultado:

**Figura 18**

*Representación de la interdependencia entre los residuos y el número de observaciones*



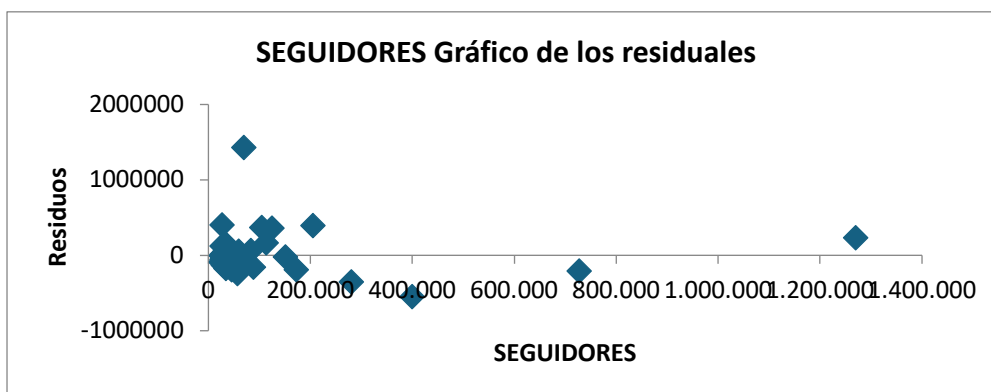
Como se observa en la figura 18, el gráfico no presenta ningún tipo de patrón sistemático, lo que sugiere que no hay autocorrelación significativa entre los residuos, es decir, que esta ausencia de patrones cíclicos o tendencias marcadas indica que los errores son independientes

entre sí. Esto indica que el cumplimiento del supuesto se aplica en el modelo, lo cual resulta ser crucial para su validez.

En una tercera instancia se encuentra la homocedasticidad, que implica que los errores tengan una varianza constante, para evaluarlo se realizó un análisis de las gráficas de residuales especificadas en análisis de datos de Excel al momento de aplicar la regresión. Los resultados fueron los siguientes:

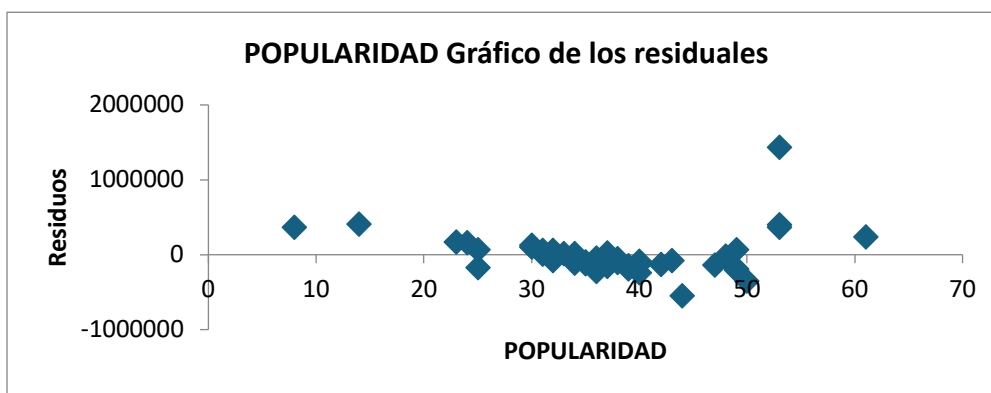
**Figura 19**

*Prueba de homocedasticidad de la variable explicativa: Seguidores en Spotify*



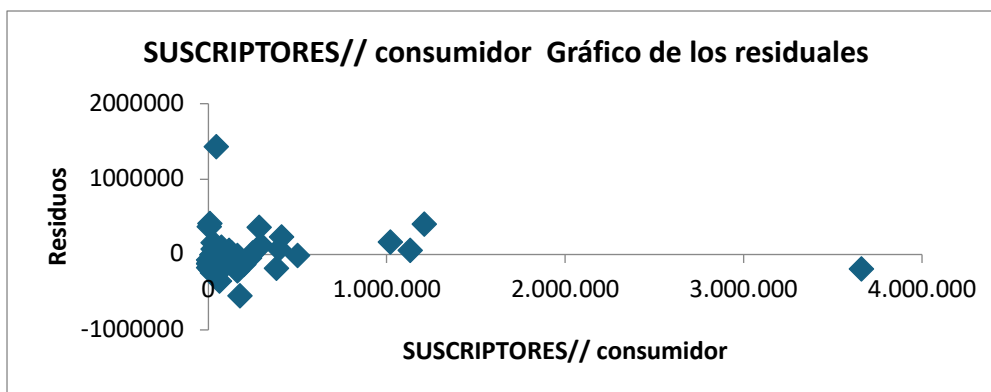
**Figura 20**

*Prueba de homocedasticidad de la variable explicativa: Popularidad en Spotify*



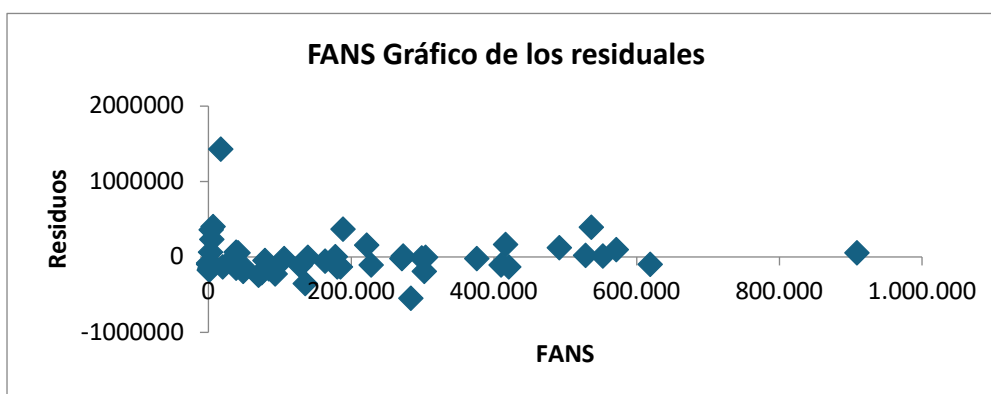
**Figura 21**

*Prueba de homocedasticidad de la variable explicativa: Suscriptores en YouTube*



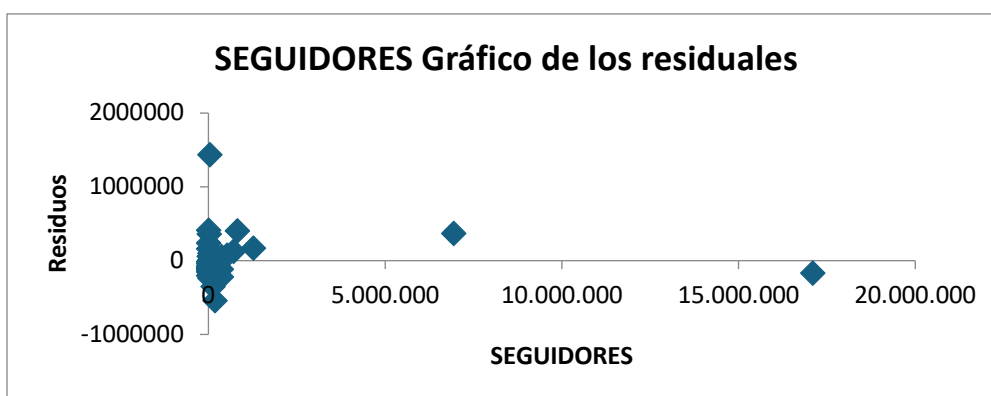
**Figura 22**

*Prueba de homocedasticidad de la variable explicativa: Fans en Facebook*



**Figura 23**

*Prueba de homocedasticidad de la variable explicativa: Seguidores en Instagram*



Al visualizar los gráficos podemos determinar que los residuos de cada variable se distribuyen de manera aleatoria, siendo este a primera vista el requisito principal para que los

datos sean considerados homocedásticos, analizando cada gráfica de manera específica podemos decir que:

- ***Seguidores en Spotify:*** la figura 19 muestra una dispersión aleatoria de los residuos alrededor del eje horizontal, sin formar patrones discernibles, por lo que sugiere que la varianza de los residuos se mantiene constante para los diferentes niveles de seguidores en Spotify, indicando homocedasticidad.
- ***Popularidad en Spotify:*** de manera similar en la figura 20 se puede observar que el gráfico de residuales no revela ningún patrón sistemático, indicando que los errores son homocedásticos. La varianza de los residuos parece ser constante a lo largo de todos los niveles de popularidad.
- ***Suscriptores en YouTube:*** en este gráfico, aunque hay una mayor dispersión en los residuos para los niveles más altos de suscriptores, como representa la figura 21, la mayoría de los puntos se agrupan de manera aleatoria alrededor del origen, sin embargo, no se puede confirmar la completa homocedasticidad en este caso.
- ***Fans en Facebook:*** para esta variable como evidencia la figura 22, el gráfico de residuos muestra una distribución aleatoria alrededor del eje horizontal, sin patrones discernibles, sugiriendo que los errores son homocedásticos para esta variable
- ***Seguidores en Instagram:*** de la misma manera los residuos muestran una dispersión aleatoria sugiriendo homocedasticidad, con una concentración en el origen sin patrones discernibles, como se observa en la figura 23.

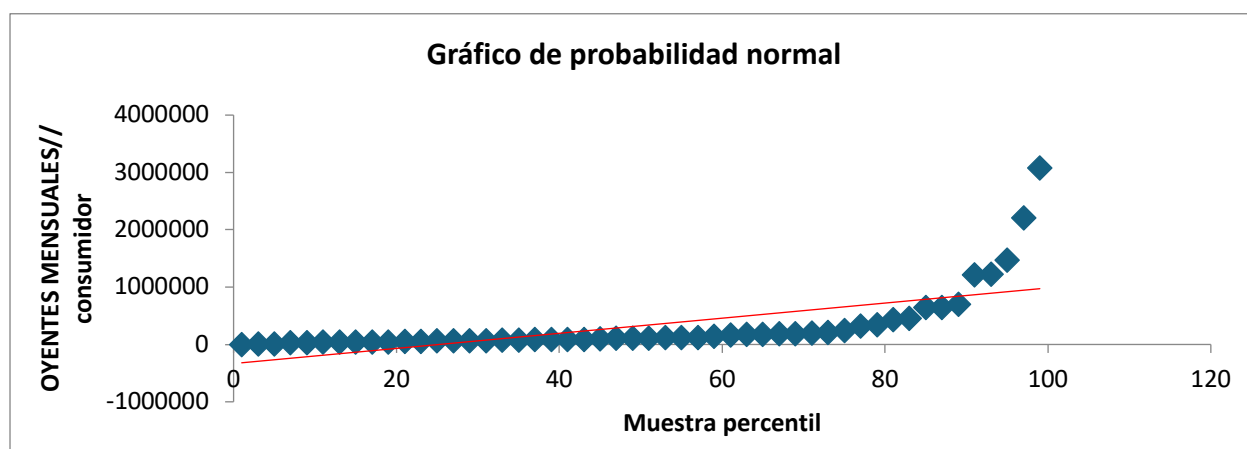
Tomando en cuenta la aleatoriedad de los residuos en las gráficas se puede decir que apoya de manera positiva el supuesto de homocedasticidad en el modelo de regresión, reforzando

la validez de este, asegurando que las inferencias estadísticas son confiables hasta este punto. Para continuar con la evaluación el supuesto a analizar es la normalidad de los residuos, lo cual permite asegurar que las inferencias estadísticas, como pruebas de hipótesis e intervalos de confianza, sean válidas y precisas.

Anteriormente se especificó el uso de la prueba Shapiro-Wilk con la ayuda del complemento de Excel, Real Statistics, para corroborar la normalidad de los datos. Para verificar esta normalidad en el modelo se seleccionó la especificación del gráfico de probabilidad normal mediante la herramienta de análisis de datos al momento de realizar la regresión, obteniendo el siguiente resultado:

**Figura 24**

*Representación de la normalidad de los residuos del modelo*



Como se visualiza en el gráfico la línea de tendencia en rojo representa la distribución normal teórica, para que los residuos se consideren normalmente distribuidos, los puntos deben alinearse aproximadamente a lo largo de la diagonal. La inspección del gráfico revela que, en general, los residuos se alinean a lo largo de la diagonal en la parte central del gráfico, sugiriendo que los residuos se aproximan a una distribución normal en el rango medio. Sin embargo, los

extremos y en particular en el percentil superior, se observa una desviación considerable de los puntos respecto a la diagonal.

A pesar de esta desviación en los extremos, el alineamiento general de los puntos en el rango central del gráfico sugiere que la mayoría de estos cumplen con el supuesto de normalidad. No obstante, la desviación en los extremos puede ser un indicativo de outliers o una posible necesidad de transformación de datos para mejorar la normalidad de los residuos, pero de manera general el complemento ayudaría a confirmar el supuesto y asegurar la robustez del modelo.

El último supuesto de evaluación se trata de la No colinealidad o multicolinealidad de las variables, es decir, implica que las variables explicativas no estén correlacionadas entre ellas, ya que una alta colinealidad puede afectar de manera negativa la estabilidad y la interpretabilidad del modelo. Para evaluar la relación entre las variables, se analizó la matriz de correlación entre variables independientes con la herramienta análisis de datos, obteniendo los siguientes resultados

**Tabla 10**

Matriz de correlación entre las variables dependiente e independientes

	<i>OYENTES MENSUALES// CONSUMIDOR</i>	<i>SEGUIDORES</i>	<i>POPULARIDAD</i>	<i>SUSCRIPTORES// CONSUMIDOR</i>	<i>SEGUIDORES</i>	<i>FANS</i>
<b>OYENTES MENSUALES// CONSUMIDOR</b>	1,00					
<b>SEGUIDORES</b>	0,78	1,00				
<b>POPULARIDAD</b>	0,72	0,50	1,00			
<b>SUSCRIPTORES// CONSUMIDOR</b>	0,08	0,12	0,21	1,00		
<b>SEGUIDORES</b>	-0,10	-0,05	-0,33	-0,05	1,00	
<b>FANS</b>	-0,19	-0,11	-0,09	0,15	-0,10	1,00

Analizando la matriz, podemos decir que la correlación entre las variables predictoras y la variable dependiente se muestra altamente favorable con don de sus variables explicativas, a las cuáles e podría atribuir el hecho de que provienen de la misma plataforma (Seguidores y Popularidad en Spotify), mientras que las otras variables aportan una correlación mucho menor en comparación, indicando su baja capacidad predictora en el modelo, como bien se sugiere en los resultados de regresión obtenidos.

Con respecto a las correlaciones entre variables predictoras, podemos hacer hincapié en las variables seguidores y popularidad en Spotify, ya que su 0,5 de correlación es moderado, pero no lo suficientemente alto para que exista un problema significativo de colinealidad. Las correlaciones entre las demás variables son generalmente bajas (rango de -0,33 a 0,21), indicando una baja colinealidad entre ellas, lo cual resulta beneficioso para el modelo, ya que evita problemas de multicolinealidad que pueden inflar los errores estándar de los coeficientes y dificultar la interpretación.

La verificación de los supuestos fundamentales de la regresión lineal múltiple, han confirmado la validez y robustez del modelo planteado para predecir la cantidad de oyentes mensuales en Spotify, resumiendo que:

- **Normalidad:** La prueba de Shapiro-Wilk y el gráfico Q-Q indicaron que los residuos del modelo se distribuyen aproximadamente de manera normal, especialmente en el rango central, aunque con alguna desviación en los extremos. Este cumplimiento general del supuesto de normalidad asegura la validez de las inferencias estadísticas derivadas del modelo.
- **Homocedasticidad:** El análisis de los gráficos de residuos frente a las variables predictoras mostró que la varianza de los errores se mantiene constante a lo largo

de todos los niveles de las variables independientes. Esta homocedasticidad garantiza que las predicciones del modelo sean fiables y que los intervalos de confianza sean precisos.

- **Interdependencia:** El gráfico de interdependencia de los residuos indicó que no hay autocorrelación significativa entre los errores del modelo, validando el supuesto y reforzando la viabilidad de las predicciones.
- **No Colinealidad:** El análisis de la matriz de correlación confirmó que no existe una colinealidad excesiva entre las variables predictoras. Las correlaciones entre estas fueron generalmente bajas, indicando que el modelo no sufre de multicolinealidad, lo que asegura la estabilidad y la interpretabilidad de los coeficientes del modelo.
- **Linealidad:** Los gráficos de regresión ajustada mostraron que las variables predictoras clave, tienen una relación lineal significativa con la variable dependiente. Esto confirma que el supuesto se cumple adecuadamente.

Al comprobarse los supuestos de regresión lineal múltiple, se ha demostrado que el modelo planteado funciona de manera efectiva al momento de cumplir con los criterios estadísticos necesarios para ser considerado robusto y fiable. Estos resultados proporcionan una base sólida para la aplicación del modelo en la predicción de la cantidad de oyentes mensuales en Spotify, permitiendo realizar inferencias válidas y precisas sobre los factores que influyen en la popularidad de los artistas ecuatorianos en esta plataforma de streaming.

### 3. RESULTADOS Y DISCUSIÓN

Una vez aplicadas las técnicas e instrumentos de análisis, y comprobado los supuestos, se pueden presentar los hallazgos obtenidos mediante los mismos. Este estudio utilizó la regresión lineal múltiple para la construcción de un modelo predictivo sobre las ventas de artistas ecuatorianos en plataformas de streaming, utilizando Spotify como principal referencia y demás plataformas relevantes y redes sociales que influyen en la variable dependiente seleccionada (oyentes mensuales en Spotify), obteniendo mediante la herramienta análisis de datos de Excel el modelo de regresión representado por la siguiente ecuación:

$$Y = -844476,48 + 1,55 X_1 + 28668,98 X_2 - 0,07X_3 + 0,02X_4 - 0,18 X_5$$

Donde:

- Y es la cantidad de oyentes mensuales en Spotify o Streams.
- $X_1$  es el número de seguidores en Spotify.
- $X_2$  es la popularidad en Spotify.
- $X_3$  es el número de suscriptores en YouTube.
- $X_4$  es el número de seguidores en Instagram
- $X_5$  es el número de fans en Facebook.

#### 3.1. Significado de los Coeficientes

- **Intercepción (-844476,48):** Representa el valor esperado de Y cuando todas las variables dependientes son cero. En este contexto, tiene un valor negativo significativo, lo cual sugiere que, en ausencia de seguidores y popularidad, el modelo no predice oyentes mensuales. Sin embargo, es importante señalar que este valor en la práctica no es interpretable ya que es improbable encontrarse en una situación en donde todas las variables señalen el valor de cero.

- **Coefficiente de  $X_1$  (1,55):** Indica que, por cada aumento de seguidor en Spotify, se espera que los oyentes mensuales aumenten en 1,55; manteniendo constantes las otras variables. Este coeficiente positivo y significativo resalta la importancia de la variable en la predicción.
- **Coefficiente de  $X_2$  (28668,98):** Sugiere que un incremento de una unidad en la popularidad de Spotify (en una escala de 0 a 100) resultaría en un aumento de 28668,98 en el aumento de oyentes mensuales, asumiendo la constancia de las demás variables. Este coeficiente es altamente significativo y destaca la relevancia de la popularidad en la plataforma.
- **Coefficiente de  $X_3$  (-0,07):** Indica que un aumento de un suscriptor en YouTube está asociado con una disminución de 0,07 en los oyentes mensuales en Spotify, aunque su efecto es marginal y su significancia es relativamente baja.
- **Coefficiente de  $X_4$  (0,02):** Este coeficiente sugiere que un incremento de un seguidor en Instagram aumenta ligeramente los oyentes mensuales en Spotify. Este efecto es pequeño y no representa gran significancia.
- **Coefficiente de  $X_5$  (-0,18):** Representa una disminución de 0,18 oyentes mensuales por cada aumento de un fan en Facebook. Este coeficiente también muestra una relación negativa marginal y no significativa.

### 3.2. Interpretación de los Coeficientes

Cada coeficiente en la ecuación tiene implicaciones prácticas diferentes:

- **Seguidores en Spotify:** El coeficiente positivo y significativo indica que aumentar el número de seguidores es crucial para incrementar los oyentes

mensuales. Esto es coherente con la literatura que señala la importancia de la base de seguidores para la visibilidad y el engagement en plataformas de streaming.

- **Popularidad en Spotify:** Este coeficiente muestra una fuerte relación positiva, sugiriendo que la popularidad, medida por las reproducciones recientes y la actividad en la plataforma, es un indicador clave de los oyentes mensuales. La popularidad puede reflejar tanto la calidad del contenido como la efectividad de las campañas de marketing.
- **Suscriptores en YouTube:** Aunque se esperaría que tener más suscriptores en YouTube aumentara los oyentes en Spotify, el coeficiente es negativo, aunque pequeño y no significativo, sugiere que los suscriptores en YouTube pueden no traducirse directamente en oyentes en Spotify. Esto podría deberse a diferencias en las preferencias de uso entre plataformas.
- **Seguidores en Instagram y Fans en Facebook:** Estos coeficientes son pequeños y no significativos, indicando que, aunque la presencia en redes sociales es importante, no tiene un impacto directo tan fuerte como los seguidores y la popularidad en Spotify en la predicción de oyentes mensuales.

### 3.3. Significancia estadística y P-Valores

Evaluando los p-valores de cada variable se puede evidenciar que las variables Seguidores y Popularidad en Spotify tienen p-valores extremadamente bajos ( $p < 0,01$ ), lo que indica una alta significancia estadística. Estas variables son altamente predictivas de la variable dependiente y por lo tanto justifican plenamente en el modelo. Mientras que el resto de variable (Suscriptores en YouTube, Seguidores en Instagram y Fans en Facebook), tienen p-valores altos ( $p > 0,05$ ), lo que sugiere que no son significativamente predictivos de los oyentes mensuales en

Spotify. Estos resultados justifican considerar la exclusión de estas variables del modelo para simplificar y mejorar la predicción de este.

Si analizamos los patrones relevantes encontrados, se resalta la importancia de Spotify como plataforma, ya que las variables relacionadas directamente a esta son las más significativas, lo que resalta la importancia de centrarse en esta plataforma para aumentar los oyentes mensuales. Esto alinea los estudios previos que destacan a Spotify como la plataforma líder en streaming musical; desde el punto de vista de las redes sociales, si bien tienen un papel en la promoción y visibilidad de los artistas, su impacto directo en los oyentes mensuales en Spotify es limitado según el modelo. Esto implica que las estrategias de marketing utilizadas deben ser más integradas y no dependientes únicamente de la presencia en redes sociales.

### **3.4. Prueba de Hipótesis**

Para la validación del modelo de regresión lineal múltiple construido, se realizó una prueba de hipótesis, tomando en cuenta los siguientes supuestos:

- **Hipótesis Nula (H0):** Las variables beta son iguales a 0, es decir, ninguna de las variables independientes tiene un impacto significativo en la variable dependiente.
- **Hipótesis Alternativa (H1):** Al menos una de las variables beta es diferente de 0, es decir, al menos una de las variables independientes tiene un impacto significativo en la variable dependiente.
  - **Niveles de Confianza y Significancia**
    - *Nivel de confianza (1- $\alpha$ ):* 95%
    - *Nivel de Significancia ( $\alpha$ ):* 5%

### 3.4.1. Resultados de la Prueba de Hipótesis

El análisis se realizó utilizando el estadístico F y los valores obtenidos en la tabla 8, el análisis de varianza (ANOVA), obteniendo los siguientes resultados:

- **Valor crítico de F ( $\alpha=5\%$ ): 2.43**
- **Estadístico F (ANOVA): 30.67**

Para el cálculo del valor crítico de F se utilizó la función (INV.F) de Excel, que permite encontrar el valor de distribución F inversa para un nivel de significancia dado. Esta función es fundamental en pruebas de hipótesis para determinar el umbral por encima del cual rechazamos la hipótesis nula.

#### - **Funciones Utilizadas en Excel:**

- **INV.F:** Esta función devuelve el valor crítico de la distribución F para un nivel de significancia específico. En este caso, INV.F (0.05, df1, df2) se utilizó para encontrar el valor crítico de F al 5% de significancia, donde df1 y df2 corresponden a los grados de libertad del numerador y denominador, respectivamente.
- **DIST.F. N:** Esta función se utiliza para calcular la distribución acumulativa de F. para la altura del gráfico se utilizó la función con un acumulado de 0, y para la probabilidad acumulada, se usó un acumulado de 1. Esto permitió representar visualmente las probabilidades y determinar las áreas bajo la curva correspondientes a los niveles de confianza y significancia

#### - **Grados de Libertad**

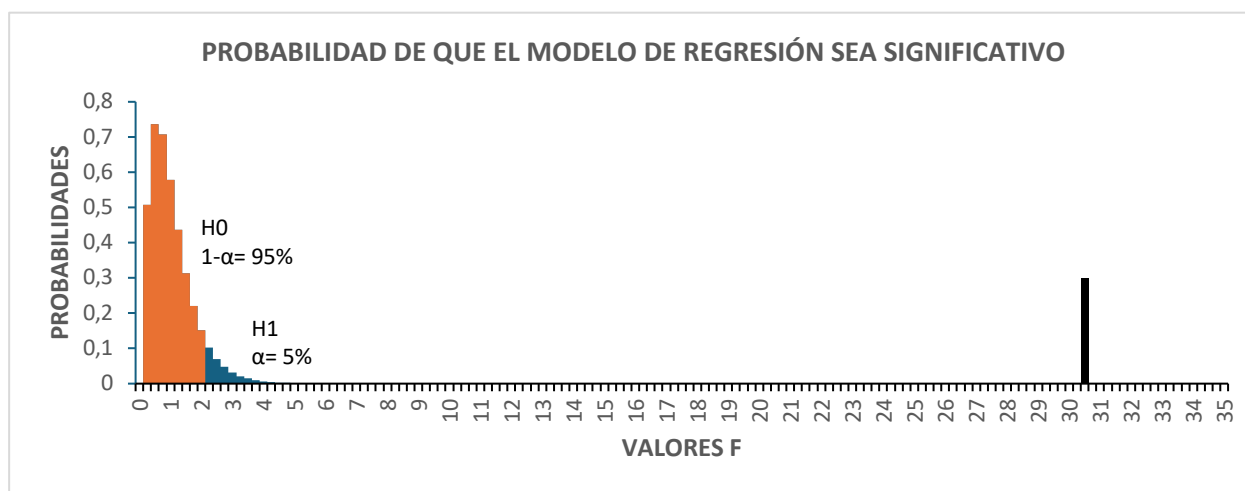
Los grados de libertad utilizados en este análisis se obtuvieron del Análisis de Varianza, en este caso se emplearon 5 grados de libertad para el numerador (número de variables independientes y 44 grados de libertad para el denominador (tamaño de la muestra menos el número de variables independientes menos uno).

### 3.4.2. Interpretación de los resultados

El estadístico de F calculado (30.67) es significativamente mayor que el valor crítico de F (2.43), lo que indica que existe suficiente evidencia para rechazar la hipótesis nula ( $H_0$ ) con un nivel de confianza del 95%. Este resultado sugiere que al menos una de las variables independientes en el modelo tiene un impacto significativo en la variable dependiente, lo que valida la significancia del modelo de regresión utilizado, como se muestra en el siguiente gráfico:

**Figura 25**

Representación gráfica de la prueba de Hipótesis



El gráfico representado (figura 25) muestra la distribución de las probabilidades de que el modelo sea significativo, utilizando los valores representados en el Anexo 5. La región sombreada de azul representa el nivel de significancia ( $\alpha = 5\%$ ), mientras que la región en

naranja representa el nivel de confianza ( $1 - \alpha = 95\%$ ). El estadístico F calculado se encuentra muy por encima del valor crítico, reafirmando que el modelo es estadísticamente significativo.

Los resultados de la prueba de hipótesis confirman que el modelo de regresión lineal múltiple aplicado es significativo, validando su uso para realizar pronósticos sobre la cantidad de oyentes mensuales en Spotify para artistas ecuatorianos. Esta validación es crucial para asegurar la fiabilidad y aplicabilidad del modelo en contextos reales, proporcionando una base sólida para las conclusiones y recomendaciones prácticas del estudio.

### **3.5.Capacidad de pronóstico**

La capacidad de pronóstico de un modelo estadístico es fundamental para evaluar su aplicabilidad y precisión en contextos reales. Para probar la eficacia del modelo, se llevó a cabo una simulación utilizando datos de prueba que no fueron incluidos en el conjunto de datos original. Esta simulación permite validar la capacidad del modelo para predecir la cantidad de oyentes mensuales en Spotify con base en las variables independientes seleccionadas.

#### ***3.5.1. Metodología de simulación***

Para llevar a cabo la simulación y encontrar el punto de equilibrio, como primer paso debemos definir la ecuación presentada con anterioridad:

$$Y = -844476,48 + 1,55 X_1 + 28668,98 X_2 - 0,07X_3 + 0,02X_4 - 0,18 X_5$$

Como segundo paso se generó un rango de valores para cada variable independiente tomando en cuenta 3 factores:

- ***Revisión de Datos Reales:*** La selección de estos rangos se basó en la revisión de datos reales de artistas en plataformas de streaming y redes sociales, asegurando que los valores sean representativos de las condiciones actuales del mercado musical.

- **Cobertura Amplia:** se buscó cubrir un espectro amplio para incluir tanto artistas emergentes como aquellos con una amplia base de seguidores y fans, permitiendo así una simulación que refleje diversas realidades.
- **Consistencia:** Mantener un grupo de rango de 0 a 1.000.000 para la mayoría de las variables garantiza la consistencia en el análisis y facilita la comparación entre diferentes variables independientes.

Los artistas en sus plataformas pueden tener una cantidad muy variable de seguidores, desde unos pocos hasta millones, por lo que el rango escogido puede abarcar tanto a artistas emergentes como aquellos con una gran base de seguidores. Por otro lado, la variable Popularidad en Spotify es una métrica que va de 0 a 100, donde 100 representa la máxima popularidad posible y utilizar el rango completo permite analizar el impacto de diferentes niveles de popularidad, desde artistas menos conocidos hasta los de renombre internacional.

Considerando los factores mencionados, los rangos quedaron distribuidos de la siguiente manera:

- $X_1$  : de 0 a 1'000.000 (Seguidores en Spotify)
- $X_2$  : de 0 a 100 (Popularidad en Spotify)
- $X_3$  : de 0 a 1'000.000 (Suscriptores en YouTube)
- $X_4$  : de 0 a 1'000.000 (Seguidores en Instagram)
- $X_5$  : de 0 a 1'000.000 (Fans en Facebook)

### 3.5.2. Cálculo de los valores de Y

Utilizando la ecuación modelo se calcularon los valores de Y para cada combinación de valores de las variables independientes. Esto permitió observar cómo cambia el número de oyentes mensuales en respuesta a cambios en las variables predictivas.

Para realizar la simulación se seleccionaron dos escenarios:

- **Escenario 1: Valores Cero para todas las variables**

**Tabla 11**

*Simulación del modelo suponiendo que las variables independientes sean cero.*

SIMULACIÓN	
SEGUIDORES	0
POPULARIDAD	0
SUSCRIPTORES	0
SEGUIDORES	0
FANS	0
Y=	-844.476

En este escenario, el valor de Y es igual a la intercepción del modelo, -844.4476, lo cual es lo esperable según la formulación del modelo.

- **Escenario 2: Valores Estimados para encontrar el punto de equilibrio.**

**Tabla 12**

Simulación del modelo para encontrar el punto de equilibrio

SIMULACIÓN	
SEGUIDORES	5.000
POPULARIDAD	30
SUSCRIPTORES	93.705
SEGUIDORES	100.000
FANS	100.015
Y=	0

En este escenario, se logró un punto de equilibrio donde  $Y=0$ . Sugiriendo que con una estimación de 5.000 seguidores en Spotify; una popularidad de 30/100; 93.705 suscriptores en YouTube; 100.000 seguidores en Instagram y 100,015 fans en Facebook, el modelo predice un número de oyentes mensuales que se aproxima a cero.

El análisis de estos dos escenarios proporciona una visión clara sobre como las diferentes variables independientes influyen en la cantidad de oyentes mensuales en Spotify. En el primer

escenario, con todas las variables de cero, el valor de Y es negativo y muy grande en magnitud, lo cual no es realista en un contexto práctico, pero sirve para entender la contribución base en la intercepción del modelo.

En el segundo escenario, los valores realistas de las variables llevaron a un equilibrio donde  $Y=0$ . Esto significa que, para evitar tener oyentes mensuales nulos, los artistas necesitan mantener niveles considerables de seguidores y popularidad en las plataformas.

### 3.6. Traducción de resultados en Valores Económicos

Para traducir los resultados de la simulación en valores económicos, es crucial considerar el modelo de ingresos de plataformas de Streaming, en este caso Spotify, así como los conceptos económicos discutidas en el estado del arte. Como se establece en la Tabla 2 Spotify paga aproximadamente entre \$0,003 y \$0,005 por reproducción, que se promedian en \$0,004 por reproducción. Para ejemplificarlo, utilizamos datos del artista ecuatoriano Julio Jaramillo (Véase Anexo 1). Obteniendo los siguientes resultados según la aplicación del modelo:

*Tabla 13*

*Datos de plataformas de Julio Jaramillo a 2024*

ARTISTA	JULIO JARAMILLO
SEGUIDORES	1.270.000
POPULARIDAD	61
SUSCRIPTORES	410.000
SEGUIDORES	7.100
FANS	4.450
Y=	2.842.807

$$Y = -844476,48 + (1,55 * 1.270.000) + (28668,98*61) - (0,07 * 410.000) + (0,02 * 7.100) - (0,18*4.450)$$

$$Y = 2.842.807$$

Es decir, si multiplicamos el número de oyentes o streams mensuales por la cantidad monetaria que paga Spotify por cada reproducción (\$0,004) obtenemos el ingreso mensual del artista, tomando en cuenta solo sus reproducciones, otros aspectos como derechos de autor, regalías y contratos, deberían considerarse en otro análisis específico para el cálculo del valor total real. Aclarado este punto el resultado fue el siguiente:

$$\mathbf{Ingresos = Streams Totales \times \$0,004}$$

$$\mathbf{Ingresos = 2'842.807 \times \$0,004}$$

$$\mathbf{Ingresos = \$11.371,23}$$

Esto quiere decir que la persona, apoderada de los derechos de la música de Julio Jaramillo, y sus plataformas de streaming, obtiene un ingreso mensual de once mil dólares aproximadamente, este ejemplo fortalece la visión sobre cómo las variables independientes afectan los streams totales, y por ende los ingresos económicos.

### **3.7.Limitaciones del modelo**

A pesar de la robustez y la capacidad predictiva del modelo construido en este estudio, es importante reconocer y discutir sus limitaciones para proporcionar una visión completa y equilibrada del análisis. Estas limitaciones deben tomarse en cuenta al interpretar los resultados y al aplicar las conclusiones en contextos prácticos.

- **Disponibilidad de Datos:** La precisión y la fiabilidad de cualquier modelo de regresión dependen en gran medida de la calidad y la disponibilidad de los datos utilizados. Al trabajar con plataformas de streaming, los datos disponibles públicamente sobre seguidores, popularidad y suscriptores varían constantemente, lo que afecta la representatividad y generalización de los datos.

- **Generalización de Resultados:** El modelo se basa en datos de artistas ecuatorianos, y aunque proporciona insights valiosos, sus resultados pueden no ser generalizables a artistas de otras regiones o géneros musicales. Las tendencias y patrones en el consumo de música pueden variar significativamente según el contexto cultural y regional, lo que limita la aplicabilidad universal del modelo.
- **Supuestos del Modelo de Regresión:** El modelo de regresión lineal múltiple asume la linealidad entre las variables independientes y la variable dependiente, homocedasticidad, interdependencia y normalidad de los errores. Si estos supuestos no se cumplen, la precisión del modelo puede verse comprometida y aunque se realizaron verificaciones para asegurar el cumplimiento de estos supuestos, cualquier turbación puede afectar los resultados y su interpretación.
- **Influencia de Factores Externos:** El modelo no considera factores externos que pueden influir en el número de streams mensuales en Spotify, como cambios en las políticas de la plataforma, variaciones en las preferencias de los oyentes, estrategias de marketing imprevistas, eventos globales o la competencia entre artistas. Estos factores pueden tener un impacto significativo en el comportamiento de los oyentes y, por ende, en las predicciones del modelo.
- **Multicolinealidad:** Aunque se realizó un análisis de multicolinealidad y se encontró que no era excesiva, la presencia de alguna entre las variables puede afectar la estabilidad de los coeficientes del modelo. Esto puede llevar a interpretaciones engañosas sobre la importancia relativa de cada variable.
- **Temporalidad de los datos:** El modelo se basa en datos de un período específico. Las dinámicas en las plataformas de streaming y redes sociales pueden cambiar

con el tiempo, lo que implica que los coeficientes y las relaciones identificadas en el modelo pueden no ser constantes a lo largo del tiempo. Es necesario actualizar el modelo periódicamente para reflejar las nuevas tendencias y patrones en el consumo de música.

### **3.8. Aportaciones del modelo**

El modelo desarrollado ofrece varias aportaciones significativas, tanto para los artistas como para los gestores de la industria musical, con implicaciones importantes en los negocios internacionales. En primera instancia, proporciona una herramienta cuantitativa para predecir el número de streams mensuales, lo que permite a los artistas planificar y ajustar sus estrategias de marketing y promoción de manera precisa y estratégica. Al identificar variables clave que influyen en los streams, los artistas pueden focalizar sus esfuerzos en aumentar sus seguidores y popularidad en Spotify maximizando así su alcance y éxito en la plataforma.

Desde la perspectiva de los negocios internacionales, este modelo proporciona a los artistas ecuatorianos y sus gestores comprender íntegramente como posicionarse en mercados globales, al utilizar los insights proporcionados por el modelo, es posible diseñar campañas de promoción global que trasciendan las fronteras nacionales, captando oyentes de diversos países y culturas. Esto es crucial en la industria musical globalizada donde la capacidad de atraer a una audiencia internacional puede significar un aumento representativo en los ingresos y la popularidad.

Además, el modelo sirve como una base para la toma de decisiones estratégicas en colaboraciones internacionales. Al determinar qué factores tienen mayor impacto en los streams mensuales, los gestores pueden invertir de manera más eficiente en asociaciones con artistas de otros países giras internacionales y campañas publicitarias dirigidas a mercados específicos. Esta

estrategia no solo optimiza los recursos, sino que también facilita la penetración en nuevos mercados, ampliando la visibilidad y el alcance global de los artistas ecuatorianos.

Finalmente, el modelo puede actualizarse con nuevos datos para mantener su relevancia en un entorno musical dinámico. Las herramientas analíticas derivadas del modelo mejoran la capacidad de los artistas para entender y aprovechar las tendencias de consumo musical en diversas regiones del mundo, contribuyendo a una gestión de carrera más científica y eficiente. Esto fortalece la competitividad de los artistas ecuatorianos en el ámbito global, promoviendo una mayor integración y éxito en el mercado internacional de la música.

## 4. CONCLUSIONES Y RECOMENDACIONES

### 4.1. Conclusiones

La investigación ha demostrado que el modelo de regresión múltiple propuesto es un instrumento efectivo para predecir la cantidad de oyentes mensuales en Spotify para artistas ecuatorianos. Este modelo no solo cumple con los supuestos fundamentales de normalidad, homocedasticidad, interdependencia, no colinealidad y linealidad, sino que también presenta un coeficiente de determinación ( $R^2$ ) del 77,7%, lo que indica una alta capacidad explicativa. La robustez del modelo se evidenció en la significancia estadística de las pruebas realizadas, confirmando que las variables seleccionadas contribuyen de manera significativa a la predicción de la variable dependiente.

La investigación reveló que las variables explicativas Seguidores y Popularidad en Spotify son determinantes críticas en la predicción de oyentes mensuales, cuyos coeficientes resaltan la influencia sustancial que estas métricas tienen sobre el éxito de un artista en la plataforma. La popularidad, medida a través de la cantidad de reproducciones recientes y otras actividades en la plataforma, mostró una relación directa y positiva con el aumento de oyentes. Estos hallazgos no solo confirman teorías previas sobre la importancia de construir una base sólida de seguidores, sino que también destacan la necesidad de mantener una actividad constante y relevante en la plataforma para aumentar la visibilidad y el engagement. En contraste, otras variables como suscriptores en YouTube, y los seguidores en Instagram y Facebook, aunque relevantes, no mostraron la misma significancia estadística, indicando que su impacto directo en Spotify es menos pronunciado.

A pesar de la eficacia del modelo, se identificaron ciertas limitaciones que deben considerarse para una interpretación adecuada de los resultados. La disponibilidad de los datos

minados de las distintas plataformas puede variar, lo que llega a influir en la precisión del modelo, tomando en cuenta que no se consideran factores externos como cambios en las políticas de las plataformas y preferencias de los oyentes, o estrategias de publicidad, que pueden tener un impacto significativo en las ventas de música. Por lo tanto, aunque el modelo proporciona una base sólida para la predicción de oyentes mensuales, su aplicación debe ser complementada con un análisis contextual y una actualización periódica de los datos para mantener su relevancia y precisión en un entorno digital y musical dinámico.

La adopción y correcta implementación del modelo predictivo construido no solo beneficia a los artistas en términos de aumentar su número de oyentes, sino que también tiene profundas implicaciones económicas. Un aumento en los oyentes mensuales en plataformas de streaming se traduce directamente en mayores ingresos por regalías y potenciales oportunidades de monetización adicional a través de contratos publicitarios y patrocinios. Asimismo, un conocimiento detallado de las variables clave que influyen en el éxito de un artista en plataformas digitales puede guiar a los gestores y productores musicales en la toma de decisiones estratégicas, optimizando la asignación de recursos para campañas de marketing y promoción. Esta eficiencia en la inversión publicitaria no solo maximiza el retorno sobre la inversión, sino que también puede contribuir significativamente al crecimiento económico del sector musical en Ecuador, posicionándolo en el mercado global.

#### **4.2.Recomendaciones**

Para asegurar la continua validez y precisión del modelo de regresión lineal múltiple construido, es esencial que los datos utilizados sean actualizados regularmente. Dado el dinamismo de la industria musical y las constantes variaciones en las tendencias de consumo en plataformas de streaming, se recomienda que los analistas de datos revisen y actualicen el

modelo trimestralmente. Incorporar nuevas métricas y ajustar las variables existentes según los patrones emergentes permitirá mantener la relevancia del modelo. Además, se recomienda realizar análisis de sensibilidad para evaluar cómo los cambios en diferentes variables afectan las predicciones. Esta práctica garantizará que el modelo permanezca robusto y adaptable, permitiendo a los gestores tomar decisiones informadas y estrategias en sus campañas de marketing y promoción.

Dado que las variables de Seguidores y Popularidad en Spotify han demostrado ser las más influyentes en la predicción de oyentes mensuales, se recomienda a los artistas y sus equipos de gestión centrarse en estrategias que maximicen estas métricas. Invertir en campañas publicitarias dirigidas, como anuncios en redes sociales, colaboraciones con influencers relevantes, puede aumentar significativamente el número de seguidores. Mantener una presencia activa en Spotify mediante lanzamientos regulares, participación en listas de reproducción populares y promoción cruzada en otras plataformas digitales puede elevar la popularidad del artista. Estas acciones no solo incrementarán el tránsito en Spotify, sino que también fortalecerán la relación con la audiencia, lo que se reflejará en un aumento sostenido de oyentes mensuales.

Para mejorar la precisión y aplicabilidad del modelo de regresión, se recomienda diversificar las fuentes de datos y considerar factores externos que puedan influir en el consumo de música. La inclusión de datos de diferentes regiones y géneros musicales permitirá una mejor generalización del modelo. Además, factores externos como cambios en las políticas de las plataformas de streaming, fluctuaciones económicas y tendencias culturales deben ser incorporados en el análisis. La utilización de técnicas avanzadas de minería de datos y aprendizaje automático puede ayudar a identificar patrones no lineales y complejos. Este enfoque

permitirá ajustar las estrategias de marketing y promoción de manera más efectiva asegurando que se adaptan a realidades cambiantes del mercado musical digital.

Aunque la presencia en redes sociales como Instagram y Facebook no mostró una influencia directa y significativa en los oyentes mensuales de Spotify, estas plataformas siguen siendo esenciales para la promoción y visibilidad de los artistas. Se recomienda una estrategia de marketing digital integrada, basada en el análisis de datos, para optimizar el tráfico hacia Spotify; utilizar herramientas de análisis para monitorizar el engagement y las intersecciones en redes sociales puede proporcionar información valiosa sobre las preferencias y comportamientos de la audiencia. Acciones como promociones cruzadas, concursos, anuncios dirigidos y colaboraciones con influencers pueden aumentar la visibilidad y atraer nuevos seguidores a las plataformas.

## 5. REFERENCIAS BIBLIOGRÁFICAS

- Baños-Gonzales, M., Canorea Tiralso, H., y Rajas Fernández, M. (2020). *La difusión del video musical en YouTube. Análisis de la capacidad viral de un videoclip*. Revista Latina de Comunicación Social: <https://www.doi.org/10.4185/RLCS-2020-1452>
- Belmares Mendoza , A., y Guevara García , M. (2021). [www.researchgate.net](http://www.researchgate.net):  
[https://www.researchgate.net/profile/Isabel-Cristina-Flores-Rueda-2/publication/352997066\\_Aplicaciones\\_de\\_herramientas\\_de\\_mercadotecnia/links/60e3341f92851ca944aafc38/Aplicaciones-de-herramientas-de-mercadotecnia.pdf#page=158](https://www.researchgate.net/profile/Isabel-Cristina-Flores-Rueda-2/publication/352997066_Aplicaciones_de_herramientas_de_mercadotecnia/links/60e3341f92851ca944aafc38/Aplicaciones-de-herramientas-de-mercadotecnia.pdf#page=158)
- Bin Ilyas, M., Ikram, A., Aadil Butt, M., y Tariq, I. (2023). *Comparative Analysis of Regression Algorithms used to Predict the Sales of* . Journal of Innovative Computing and Emerging Technologies:  
<https://pdfs.semanticscholar.org/43b4/b040b058cde795a7be290183b5b3d2d4ba37.pdf>
- Caicedo, J. (2021). *La Industria Musical en las Plataformas Digitales en el Año 2021: ¿Cómo emprender con la Música en la Era Digital?* [dspace.uartes.edu.ec](http://dspace.uartes.edu.ec):  
<http://dspace.uartes.edu.ec:8080/bitstream/123456789/1081/1/PROYECTO%20DE%20TITULACION%20-%20Caicedo%20Cano%20Jhonny%20Patricio.pdf>
- Corres , G., Passoni, L., Zárate , C., y Esteban, A. (2014). *ESTUDIO COMPARATIVO DE MODELOS DE PRONÓSTICO DE* . [cloudfront.net](http://cloudfront.net).
- Franklin , K. (2015). *Las Redes Sociales y su revolución de la industria musical*. [www.brandwatch.com](http://www.brandwatch.com): <https://www.brandwatch.com/es/blog/las-redes-sociales-y-su-revolucion-de-la-industria-musical/>

- González Sánchez, G. (2022). *Modelo ARIMA: guía completa para el pronóstico de series de tiempo en Python*. mlstudio.jaol.net: <https://mlstudio.jaol.net/arima-model-time-series-forecasting-python/>
- Huang, Y.-T., y Pai, P.-F. (2020). *Using the Least Squares Support Vector Regression to Forecast Movie Sales with Data from Twitter and Movie Databases*. MDPI: <https://www.mdpi.com/2073-8994/12/4/625#B1-symmetry-12-00625>
- IFPI. (2022). *IFPI*. globalmusicreport.ifpi.org: [https://ifpi-website-cms.s3.eu-west-2.amazonaws.com/GMR\\_2023\\_State\\_of\\_the\\_Industry\\_ee2ea600e2.pdf](https://ifpi-website-cms.s3.eu-west-2.amazonaws.com/GMR_2023_State_of_the_Industry_ee2ea600e2.pdf)
- Jáuregui, J. (2015). *Streaming musical en Spotify: ubicuidad entre géneros y estados de ánimo*. revistas.ort.edu.uy: <https://revistas.ort.edu.uy/inmediaciones-de-la-comunicacion/article/view/2587/2566>
- Malato, G. (2023). *An Introduction to the Shapiro-Wilk Test for Normality*. builtin.com: <https://builtin.com/data-science/shapiro-wilk-test#:~:text=Shapiro-Wilk%20test%20is%20a%20hypothesis%20test%20that%20evaluates,low%20p-value%20indicates%20that%20it%20isn%E2%80%99t%20normally%20distributed.>
- Martinez Osorio, J. S. (2023). *Predicción espacial de ventas mediante enfoques bayesianos y aprendizaje automático utilizando datos de área*. Universidad Nacional de Colombia: <https://repositorio.unal.edu.co/handle/unal/85484>
- Masis González, T. (2021). *Capitalismo de plataformas, producción y consumo de música: el caso de Spotify*. revistas.ucr.ac.cr: DOI 10.15517/es.v81i2.49478
- MENTINNO. (2024). *Ecuador, Estado Digital Febreeo 2024*. www.mentinno.com: <https://www.mentinno.com/acceso-estado-digital-ecuador-2024/>

- Molina, G., y Rodrigo, M. (2010). *El modelo de regresión lineal*. <http://ocw.uv.es/>:  
[http://ocw.uv.es/ciencias-de-la-salud/pruebas-1/1-3/t\\_09nuevo.pdf](http://ocw.uv.es/ciencias-de-la-salud/pruebas-1/1-3/t_09nuevo.pdf)
- Montero Granados, R. (2016). *Modelos de regresión lineal múltiple*. [www.ugr.es](http://www.ugr.es):  
[https://www.ugr.es/~montero/maticas/regresion\\_lineal.pdf](https://www.ugr.es/~montero/maticas/regresion_lineal.pdf)
- Moreno Castro, T. (2019). *La Industria Musical en las Plataformas Digitales en el Año* .  
[repositorio.uautonoma.cl](http://repositorio.uautonoma.cl):  
[https://repositorio.uautonoma.cl/bitstream/handle/20.500.12728/3192/Pronostico\\_de\\_ventas.pdf?sequence=1&isAllowed=y](https://repositorio.uautonoma.cl/bitstream/handle/20.500.12728/3192/Pronostico_de_ventas.pdf?sequence=1&isAllowed=y)
- Muñoz, J. E. (2018). *Streaming: la renovación digital de la industria musical*.  
[repository.javeriana.edu.co](http://repository.javeriana.edu.co):  
<https://repository.javeriana.edu.co/bitstream/handle/10554/35778/TG%20-%20Riveros%20Mu%C3%B1oz,%20Jorge%20Ernesto.pdf?sequence=1>
- Naranjo Alvarado, L. A. (2023). *Exploración de la Música Nacional Divulgada en Producciones Audiovisuales* . <http://dspace.utb.edu.ec/>:  
<http://dspace.utb.edu.ec/bitstream/handle/49000/15142/TIC-UTB-FCJSE-COMUNICACION-000001.pdf?sequence=1>
- Nasteski, V. (2017). *An overview of the supervised machine learning*. [www.researchgate.net](http://www.researchgate.net):  
[https://www.researchgate.net/profile/Vladimir-Nasteski/publication/328146111\\_An\\_overview\\_of\\_the\\_supervised\\_machine\\_learning\\_methods/links/5c1025194585157ac1bba147/An-overview-of-the-supervised-machine-learning-methods.pdf](https://www.researchgate.net/profile/Vladimir-Nasteski/publication/328146111_An_overview_of_the_supervised_machine_learning_methods/links/5c1025194585157ac1bba147/An-overview-of-the-supervised-machine-learning-methods.pdf)
- Nautelings, I. (2021). *Neural networks*. TikZ.net: [https://tikz.net/neural\\_networks/](https://tikz.net/neural_networks/)

- Núñez, E., Steyerberg, E., y Núñez, J. (2011). *Estrategias para la elaboración de modelos estadísticos de regresión*. [www.revespcardiol.org](http://www.revespcardiol.org):  
[https://d1wqtxts1xzle7.cloudfront.net/50934975/S300-libre.pdf?1481982758=&response-content-disposition=inline%3B+filename%3DRegression\\_modeling\\_strategies.pdf&Expires=1712971502&Signature=RVtst12DF-LUxPADRtLcXWnm~rnGQPnjePg~AsAF-1Nido6rNZObHnhiSbZrSJxZoxT](https://d1wqtxts1xzle7.cloudfront.net/50934975/S300-libre.pdf?1481982758=&response-content-disposition=inline%3B+filename%3DRegression_modeling_strategies.pdf&Expires=1712971502&Signature=RVtst12DF-LUxPADRtLcXWnm~rnGQPnjePg~AsAF-1Nido6rNZObHnhiSbZrSJxZoxT)
- OMC. (2019). *Organización Mundial del Comercio*. [wto.org](http://wto.org):  
[https://www.wto.org/spanish/res\\_s/booksp\\_s/00\\_wtr19\\_s.pdf](https://www.wto.org/spanish/res_s/booksp_s/00_wtr19_s.pdf)
- Orús, A. (2024). *Cuota de mercado de las plataformas de música en streaming a nivel mundial en 2023*. [es.statista.com](https://es.statista.com): <https://es.statista.com/estadisticas/698135/cuota-de-mercado-de-las-plataformas-de-streaming-a-nivel-mundial/>
- Palacios Utreras, C. A. (2020). *Análisis y predicción de las tendencias de venta en el mercado usando árboles de regresión*. [repositorio.usfq.edu.ec](http://repositorio.usfq.edu.ec):  
<https://repositorio.usfq.edu.ec/bitstream/23000/8799/1/146132.pdf>
- Pereira Gonzales, A. (2010). *Análisis predictivo de datos mediante técnicas de regresión estadística*. [docta.ucm.es](http://docta.ucm.es): <https://docta.ucm.es/entities/publication/89b543b8-3489-47ef-9cce-a4887509bed2>
- Pereira, A. (2010). *Universidad Complutense Madrid*. [docta.ucm.es](http://docta.ucm.es):  
<https://docta.ucm.es/entities/publication/89b543b8-3489-47ef-9cce-a4887509bed2>
- Plaza Chicote, L. M. (2023). *Análisis predictivo de la popularidad de una canción en Spotify*. Universidad Complutense de Madrid:

<https://docta.ucm.es/rest/api/core/bitstreams/bdc5f5ca-fa9a-4526-9948-2d67d05a8560/content>

Pruvost, A. (2014). *Música, opción legal con Streaming, el caso Spotify*. sedici.unlp.edu.ar:

[https://sedici.unlp.edu.ar/bitstream/handle/10915/41831/Documento\\_completo.pdf?sequence=1](https://sedici.unlp.edu.ar/bitstream/handle/10915/41831/Documento_completo.pdf?sequence=1)

Real Statistics Using Excel . (2023). *Shapiro-Wilk Test for Normality* . [https://www.real-](https://www.real-statistics.com/statistics-tables/shapiro-wilk-table/)

[statistics.com/statistics-tables/shapiro-wilk-table/](https://www.real-statistics.com/statistics-tables/shapiro-wilk-table/)

Riveros Muñoz, J. (2018). *Streaming: la renovación digital de la industria musical* . Universidad

Javeriana: <https://repository.javeriana.edu.co/bitstream/handle/10554/35778/TG%20-%20Riveros%20Mu%c3%b1oz%2c%20Jorge%20Ernesto.pdf?sequence=1&isAllowed=y>

Ruelas Santoyo, E., y Laguna Gonzáles, J. A. (2014). *Comparación de predicción basada en redes neuronales contra métodos estadísticos en el pronóstico de ventas*. redalyc.org:

<https://www.redalyc.org/pdf/2150/215037911008.pdf>

Soundcharts. (2024). *Soundcharts*. help.soundcharts.com:

<https://help.soundcharts.com/en/articles/3188781-what-is-soundcharts>

Spotify. (2024). *Regalías*. support.spotify.com:

<https://support.spotify.com/es/artists/article/royalties/?ref=related>

UNCTAD. (2022). *Perspectivas de la Economía Creativa*. unctad.org:

[https://unctad.org/system/files/official-document/ditctsce2022d1\\_overview\\_es.pdf](https://unctad.org/system/files/official-document/ditctsce2022d1_overview_es.pdf)

Vidyasagar, B. S., y Karawande, V. (2021). *EMOTION BASED MUSIC RECOMMENDATION* .

INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH :

[https://www.ijasret.com/VolumeArticles/FullTextPDF/968\\_44.EMOTION\\_BASED\\_MU](https://www.ijasret.com/VolumeArticles/FullTextPDF/968_44.EMOTION_BASED_MU)

SIC\_RECOMMENDATION\_SYSTEM\_BY\_USING\_DIFFERENT\_ML\_APPROACH.  
pdf

Vilá Baños, R., Torrado Fonseca, M., y Reguant Álvarez, M. (2019). *Reire*. Universitat de Barcelona. Institut de Desenvolupament Professional. ICE:  
file:///C:/Users/emili/Downloads/22704-Text%20de%20l'article-61419-3-10-20190701%20(1).pdf

Weiszfeld, D. (2024). *Soundcharts*. help.soundcharts.com:

<https://help.soundcharts.com/en/articles/3188781-what-is-soundcharts>

Zhang, Y., y Hara, T. (2021). *Predicting E-commerce Item Sales with Web*. 24th International Conference on Business Information Systems :

file:///C:/Users/emili/Downloads/xedig,+22\_37\_BIS\_2021+-+Kopie+-+Kopie.pdf

## 6. ANEXOS

### Anexo 1: Base de datos principal

Nº	ARTISTA	SPOTIFY					YOUTUBE			INSTAGRAM		TWTOK		DEEZER		FACEBOOK		SOUNDCLOUD			
		SEGUIDORES	%DE CRECIMIENTO MENSUAL	VENTAS MENSUALES/ consumidor	%DE CRECIMIENTO MENSUAL	POPULARIDAD	%DE CRECIMIENTO MENSUAL	SUSCRITORES/ consumidor	%DE CRECIMIENTO MENSUAL	VISTAS	%DE CRECIMIENTO MENSUAL	SEGUIDORES	%DE CRECIMIENTO MENSUAL	SEGUIDORES	%DE CRECIMIENTO MENSUAL	FANS/ consumidor	%DE CRECIMIENTO MENSUAL	FANS	%DE CRECIMIENTO MENSUAL	SEGUIDORES/ unidad	%DE CRECIMIENTO MENSUAL
1	LILLO PARRILLO	1.27M	3.94%	3.09M	20.08%	61	-3.67%	-	-	933.73	94.67%	-	-	-	-	298.66	1.13%	4.45	-0.23%	2.86	3.14%
2	PANDAS	726.95M	0.01%	1.47	0.27%	49	-5.77%	-	-	77.18	27.72%	-	-	-	0	-100.00%	74.99	0.00%	0	-100.00%	
3	JUAN FERNANDO VELAZO	399.67M	0.65%	428.6	-0.94%	44	-4.35%	-	-	101.76	1.43%	198.33	3.91%	19.9	0.00%	53.02	0.11%	284.05	-0.02%	178	2.93%
4	CRONIBERS	280.32M	0.54%	647.96	1.74%	53	-5.68%	62.7	1.13%	17.02	-1.49%	93.11	1.92%	-	-	26.97	1.00%	136.12	-0.10%	367	3.67%
5	ALEX PONCE	286.31M	2.23%	1.22	-4.58%	53	-7.02%	-	-	-	-	-	-	-	-	3.06	16.90%	-	-	169	6.96%
6	EDY ROS	172.78M	1.69%	315.27	0.31%	49	-3.92%	3.66	0.63%	255.72	-0.58%	147.98	3.92%	625.7	1.68%	48	9.09%	302.42	1.81%	1.58	2.33%
7	SEBASTIÁN ROSEDO	151.41M	0.93%	644.1	7.29%	48	-5.89%	50.1	7.51%	-	-	7.31	-0.48%	-	-	-	-	376.22	20.89%	-	-
8	MANN	124.78M	2.00%	1.21	9.31%	53	-3.64%	285	5.17%	-	-	18.2	1.99%	287.4	0.07%	1.63	8.46%	3.4	-2.63%	0	-100.00%
9	CHANNIVERA	113.21M	-0.28%	30.76	-3.30%	23	-14.81%	1.02	0.00%	-	-	1.27	-1.65%	2	0.00%	13.45	0.18%	416.57	0.02%	299	0.00%
10	DAVEN	104.33M	-0.11%	1.43	18.56%	8	-27.27%	4.47	-0.67%	-	-	6.94	-1.17%	-	-	780	0.65%	188.78	-0.25%	79.18	-0.01%
11	OLIMPO CARREBAS	88.22M	1.18%	177.44	-4.94%	37	-5.13%	-	-	-	-	-	-	-	-	9.27	0.63%	-	-	8	0.00%
12	ANDINO	83.17M	-0.38%	44.44	-7.27%	23	-23.65%	385	0.26%	-	-	536.84	-1.22%	1.2	0.00%	7.74	0.44%	38.93	4.83%	0	0.00%
13	MATIA	71.06M	10.53%	113.37	-23.20%	32	-8.57%	232	3.11%	-	-	45.59	4.75%	-	-	3.46	1.08%	30.33	0.13%	299	-3.55%
14	MARCELO DEPESIO	68.23M	17.03%	2.21	-0.27%	53	-8.64%	43.2	3.11%	-	-	40.28	4.45%	115.5	8.45%	1.47	8.11%	17.14	45.18%	0	-
15	FLOPEROS	66.15M	9.34%	457.35	0.44%	47	-4.08%	43.2	-0.46%	0	0.00%	63.68	47.79%	79.6	0.00%	6.89	2.56%	49.25	0.49%	0	-100.00%
16	FENIXA	59.16M	13.84%	53.47	-23.10%	32	-3.13%	116	2.66%	-	-	382.98	38.80%	1.3	0.00%	918	6.74%	908.48	3.05%	246	3.80%
17	PALUVALGURE	58.42M	0.10%	35.14	6.77%	38	-7.69%	161	3.21%	-	-	371.59	-21.52%	-	-	2.96	0.03%	93.89	-0.21%	1	-0.50%
18	LEDETO	57.32M	1.94%	136.43	-9.08%	40	2.44%	23.6	1.28%	738	-30.04%	49.59	5.51%	-	-	1.5	1.08%	70.6	0.78%	0	-100.00%
19	JAMAJUNA CAVALÉN	52.96M	3.95%	209.48	10.48%	38	-7.14%	30.9	11.15%	-	-	45.26	2.12%	-	-	-	-100.00%	-	-	75	1.95%
20	SEBASTIÁN CORAN	52.56M	2.39%	181.59	-1.36%	40	-9.09%	43.5	2.11%	-	-	179.81	0.42%	-	-	8.91	0.33%	618.73	-0.02%	35	0.00%
21	ANCEL GUARCA	50.42M	3.90%	75.24	3.65%	37	-5.13%	199	5.29%	246.43	45.19%	37.96	2.43%	-	-	0	-100.00%	-	-	-	-
22	GUARDARVA	45.24M	5.25%	122.05	3.65%	39	-7.14%	382	4.96%	-	-	44.79	4.92%	-	-	5.26	0.75%	49.69	0.16%	194	-2.13%
23	MILONE	45.02M	10.84%	66.1	4.32%	49	-2.00%	28.8	4.28%	-	-	41.57	2.48%	23.9	4.32%	562	15.04%	2.89	0.00%	130	7.44%
24	MARIN OSCALERS	42.89M	1.45%	105.36	-36.30%	34	-15.00%	129	4.03%	-	-	52.61	0.74%	-	-	3.65	0.25%	652.43	1.14%	0	-100.00%
25	AJUD	40.34M	2.22%	188.84	-2.30%	37	-5.13%	249	2.88%	-	-	219.27	0.60%	-	-	3.35	0.45%	528.45	0.93%	-	-
26	PAOLO PLAZA	38.72M	7.37%	117.95	12.66%	37	-5.13%	94.7	1.18%	-	-	114.71	0.28%	-	-	0	-100.00%	38.56	-0.29%	-	-
27	MARIA DE LOS ANGELES	36.18M	4.93%	119.44	10.77%	38	-2.99%	6.05	8.23%	-	-	-	-	-	-	-	-	409.73	1.06%	-	-
28	LIVIA BURBANO	34.92M	1.30%	61.28	12.22%	30	-6.25%	-	-	-	-	49.12	-0.38%	-	-	629	1.23%	-	-	0	100.00%
29	ANCELA	34.66M	2.27%	52.31	-7.79%	23	-10.71%	546	3.19%	-	-	17.1	0.00%	1.1	0.00%	7.81	2.86%	1.18	-0.51%	-	-
30	CLUB M	33.77M	4.53%	83.47	-2.06%	35	-10.28%	200	0.00%	-	-	34.34	3.25%	0	0.00%	2.35	1.34%	19.75	-0.28%	2.16	0.19%
31	MONIC	33.01M	-1.09%	11.11	-6.69%	24	-7.69%	27	1.50%	-	-	12.29	-0.67%	-	-	2.46	1.82%	222.09	-0.59%	720	-0.55%
32	DAPEDON	32.96M	2.12%	62.61	8.10%	31	-6.06%	1.13	0.00%	-	-	30.28	0.73%	-	-	3.99	0.66%	41.88	-0.24%	202	0.00%
33	TERRA CANEA	32.97M	8.87%	244.91	41.51%	42	-2.33%	27.6	1.10%	-	-	3.1	0.00%	-	-	1.65	1.79%	165	6.32%	83	-2.45%
34	MARIBON	32.54M	2.44%	43.91	-2.01%	34	-8.11%	70.7	6.61%	113.48	25.45%	399.22	-0.82%	403.9	0.00%	471	-4.67%	128.94	-0.21%	-	-
35	SEBASTIÁN SCOTO	30.16M	1.89%	73.6	-2.49%	32	-3.03%	103	0.00%	-	-	64.73	2.16%	-	-	0	100.00%	106.35	0.23%	0	0.00%
36	MARIBORTALUJE	30M	4.64%	63.3	12.58%	31	-3.13%	31.2	1.30%	-	-	38.7	2.87%	-	-	778	1.17%	177.66	1.58%	739	0.14%
37	SHIRO	29.92M	2.09%	65.8	-5.81%	36	-7.69%	13.5	-	-	-	6.66	0.16%	-	-	1.16	1.31%	-	-	-	-
38	MARCELES	29.11M	6.43%	177.73	-6.73%	40	-2.44%	149	1.36%	-	-	114.28	2.35%	190.4	0.00%	1	100.00%	180.36	1.92%	0	0.00%
39	ERLLO	28.64M	4.51%	342.87	10.31%	43	0.00%	159	4.61%	-	-	36.03	-0.60%	-	-	0	0.00%	4.25	-0.48%	515	-0.38%
40	BALOS BENS	28.12M	5.37%	43.05	1.99%	32	-5.88%	66	2.64%	-	-	33.55	1.12%	-	-	-	-	271	7.99%	733	0.00%
41	NOEL CRILES	28.03M	1.99%	194.94	4.21%	38	-7.69%	140	0.00%	-	-	-	-	-	-	-	-	78.82	0.00%	107	0.00%
42	MARCEL	27.33M	-0.88%	85.25	-4.44%	30	-9.09%	285	0.34%	-	-	713.58	-0.20%	84	-0.57%	40	8.11%	491.98	-0.27%	191	2.14%
43	CRISTÓBAL MANUEL	27.27M	2.30%	5.96	-16.19%	14	-22.22%	-	-	-	-	-	-	-	-	166	0.00%	-	-	-	-
44	PALUVAL TAMAÑO	26.77M	1.54%	41.22	-4.24%	32	-8.57%	161	2.55%	-	-	122.92	2.63%	-	-	0	-100.00%	288.65	1.25%	-	-
45	FALSTOMINO	26.68M	-0.15%	110.19	-4.16%	33	-8.33%	37.9	2.43%	-	-	85.3	0.89%	-	-	6.57	0.14%	273.03	-0.03%	66	0.00%
46	PAPA DADA	26.49M	6.91%	196.45	-2.30%	38	-13.64%	115	6.38%	-	-	139.7	1.30%	394.8	0.40%	1.66	1.53%	163.05	0.71%	629	-0.32%
47	JANEIRO QUE AMARRA	26.11M	2.66%	87.22	36.30%	32	-8.57%	-	-	51.12	83.51%	47.74	0.00%	-	-	796	0.63%	-	-	-	-
48	MATEO MINGUÁN	26.05M	0.59%	36.93	-2.99%	32	-8.57%	2.74	0.00%	-	-	405	8.00%	-	-	1.9	0.21%	-	-	2.87	-0.38%
49	QUEJANA AMARRA	26.05M	3.01%	59.42	10.21%	32	-8.57%	25.6	0.00%	112.28	133.87%	41.14	0.53%	-	-	1.12	0.63%	-	-	6	0.00%
50	ALEX DÍVEN	25.94M	6.99%	102.17	2.95%	33	-7.69%	29.4	2.80%	-	-	32	-5.88%	18	0.00%	0	-100.00%	-	-	1.01	1.81%

**Anexo 2:** Base de datos con las variables seleccionadas

N°	ARTISTA	VENTAS	SPOTIFY			YOUTUBE	INSTAGRAM	DEEZER	FACEBOOK	SOUNDCLOUD
			SEGUIDORES	OYENTES MENSUALES//consumidor	POPULARIDAD	SUSCRIPTORES//consumidor	SEGUIDORES	FANS // consumidor	FANS	SEGUIDORES//consumidor
1	JULIO JARAMILLO	14.041,96	1.270.000	3.080.000	61	410.000	7.100	218.660	4.450	2.860
2	TRANZAS	5.891,00	726.950	1.470.000	49	11.000	1.800	0	74.990	0
3	JUAN FERNANDO VELAZCO	1.714.578,32	399.870	428.600	44	178.000	189.330	53.020	284.050	178
4	LORY MEYERS	2.591.902,86	280.320	647.960	50	62.700	130.110	26.970	136.120	367
5	ALEX PONCE	6.090,02	205.300	1.220.000	53	1.210.000	829.000	3.060	537.000	169
6	LEO ROJAS	1.265.028,00	172.760	315.270	49	3.660.000	147.980	48	302.420	1.580
7	SEGUNDO ROSERO	2.576.901,05	151.410	644.100	48	501.000	7.310	7.500	376.220	920
8	MUNN	5.125,01	124.760	1.210.000	53	285.000	18.200	1.630	3.400	0
9	JOHANN VERA	124.060,08	113.210	30.760	23	1.020.000	1.270.000	13.450	416.570	299
10	DAXSEN	10.404,47	104.350	1.430	8	4.470	6.940.000	0	188.780	79.180
11	OLIMPO CÁRDENAS	709.804,56	88.220	177.440	37	44.500	0	9.270	85.000	8
12	JANDINO	178.155,05	83.170	44.440	25	395.000	536.840	7.740	38.930	0
13	MATATA	453.712,02	71.060	113.370	32	232.000	45.590	3.460	30.330	299
14	MÚSICA DE FRESTO	8.883,21	69.230	2.210.000	53	43.200	40.280	1.470	17.140	46
15	FLOREROS	1.830.243,24	65.150	457.550	47	43.200	63.680	6.890	49.250	0
16	REN KAI	214.001,51	59.160	53.470	32	116.000	362.930	918	908.480	246
17	PAULINA AGUIRRE	140.721,02	58.420	35.140	36	161.000	371.590	2.960	93.890	1
18	VERDE 70	545.743,61	57.320	136.430	40	23.600	49.590	1.500	70.600	0
19	LA MÁQUINA CAMALEÓN	837.950,90	52.960	209.480	39	30.900	45.260	0	1.700	75

20	GERARDO MORÁN	722.40 3,55	52.560	180.590	40	43.500	179.810	8.910	618.73 0	35
21	ÁNGEL GUARACA	301.15 9,00	50.420	75.240	37	199.000	37.960	0	421.00 0	5.500
22	GUARDARRAY A	488.58 2,03	45.240	122.050	39	382.000	44.790	5.260	48.690	184
23	MILO MAE	2.784. 430,17	45.020	696.100	49	26.800	41.570	562.0 00	2.690	130
24	MÁXIMO ESCALERAS	421.56 9,02	42.890	105.360	34	129.000	92.610	3.650	552.43 0	0
25	AU-D	755.60 9,02	40.340	188.840	37	249.000	219.270	3.350	528.45 0	342
26	PAOLO PLAZA	471.53 4,70	38.720	117.860	37	94.700	114.710	0	38.560	2
27	MARÍA DE LOS ÁNGELES	477.77 2,65	35.080	119.440	38	6.050	114.000	1.100	409.73 0	1.000
28	JUANITA BURBANO	245.19 5,87	34.920	61.280	30	72.100	49.120	629	572.00 0	0
29	ANGÉLICA	209.24 0,69	34.860	52.310	25	646	17.100.0 00	7.810	1.180	11
30	LOLABÚM	333.89 4,30	33.700	83.470	35	200	34.340	2.350	19.750	2.160
31	XION MC	44.467 ,01	33.010	11.110	24	27.000	12.290	2.460	222.09 0	720
32	DA PEÓN	251.57 0,02	32.960	62.610	31	1.130.000	30.280	3.990	41.880	202
33	TIERRA CANELA	979.66 7,61	32.870	244.910	42	27.600	3.100	1.650	185.00 0	83
34	MAR RENDÓN	175.71 3,53	32.540	43.910	34	70.700	359.220	471	128.94 0	4
35	SERGIO SACOTO	294.50 3,00	30.050	73.600	32	103.000	64.730	0	106.35 0	0
36	MARGARITA LUGUE	253.23 5,87	30.000	63.300	31	31.200	39.700	778	177.66 0	739
37	SAHIRO	343.21 3,51	29.920	85.800	36	13.500	636	1.160	228.00 0	11
38	MARQUES	711.06 9,01	29.110	177.730	40	149.000	114.280	1	180.36 0	0
39	GRILLO	1.371. 639,00	28.600	342.870	43	159.000	36.030	0	4.250	515
40	BAJO SUEÑOS	172.26 6,02	28.120	43.050	32	66.000	33.550	3.900	271.00 0	733
41	NOÉ MORALES	659.50 4,77	28.030	164.840	36	140.000	0	795	78.820	107
42	MAYKEL	341.29 5,24	27.350	85.250	30	295.000	713.580	40	491.98 0	191
43	CARLOS MANUEL	23.448 ,48	27.270	5.860	14	7.480	9.000	166	6.500	1
44	PAULINA TAMAYO	165.04 1,00	26.700	41.220	32	161.000	122.920	0	298.65 0	83

4		440.79							273.03	
5	FAUSTO MIÑO	7,94	26.680	110.190	33	37.900	85.300	6.570	0	68
4	PAPAYA	785.91							163.05	
6	DADA	5,01	26.480	196.450	38	115.000	139.700	1.660	0	629
	JAIME									
4	ENRIQUE	348.90							139.00	
7	AYMARA	5,18	26.110	87.220	32	20.400	47.740	796	0	13
4	MATEO	147.33								
8	KINGMAN	4,14	26.050	36.830	32	2.740	405	1.900	0	2.870
4	AZUCENA	237.70							304.00	
9	AYMARA	5,61	26.050	59.420	32	25.600	41.140	1.120	0	6
5	ALEEXX	408.70								
0	JOVEN	9,40	25.840	102.170	35	29.400	32	0	0	1.010

**Anexo 3: Análisis de regresión completo**

<i>Estadísticas de la regresión</i>	
<b>Coefficiente de correlación múltiple</b>	0,881497444
<b>Coefficiente de determinación R<sup>2</sup></b>	0,777037744
<b>R<sup>2</sup> ajustado</b>	0,751701124
<b>Error típico</b>	289712,3163
<b>Observaciones</b>	50

	<i>Grados de libertad</i>	<i>Suma de cuadrados</i>	<i>Promedio de los cuadrados</i>	<i>F</i>	<i>Valor crítico de F</i>
<b>Regresión</b>	5	1,28706E+13	2,57411E+12	30,66856371	2,69642E-13
<b>Residuos</b>	44	3,69306E+12	83933226242		
<b>Total</b>	49	1,65636E+13			

	<i>Coefficientes</i>	<i>Error típico</i>	<i>Estadístico t</i>	<i>Probabilidad</i>	<i>Inferior 95%</i>	<i>Superior 95%</i>	<i>Inferior 95,0%</i>	<i>Superior 95,0%</i>
<b>Intercepción</b>	-							
	844476,48	199118,74	-4,24	0,011249%	1245773,94	443179,03	1245773,94	443179,03
<b>SEGUIDORES</b>				0,000005				
	1,55	0,24	6,57	%	1,07	2,03	1,07	2,03
<b>POPULARIDAD</b>				0,000251				
	28668,98	5304,89	5,40	%	17977,66	39360,29	17977,66	39360,29
<b>SUSCRIPTORES// consumidor</b>								
	-0,07	0,08	-0,96	34,31%	-0,23	0,08	-0,23	0,08
<b>SEGUIDORES</b>								
	0,02	0,02	1,00	32,26%	-0,02	0,05	-0,02	0,05
<b>FANS</b>								
	-0,18	0,20	-0,89	37,92%	-0,59	0,23	-0,59	0,23

**Anexo 4: Análisis de los residuales**

<i>Observación</i>	<i>Pronóstico OYENTES MENSUALES// consumidor</i>	<i>Residuos</i>	<i>Residuos estándares</i>
1	2842807,127	237192,8735	0,863985352
2	1673108,152	-203108,1517	-0,73983027
3	975534,4579	-546934,4579	-1,99223254
4	996530,0508	-348570,0508	-1,26968156
5	821103,6302	398896,3698	1,452997367
6	506777,8625	-191507,8625	-0,69757571
7	661274,6887	-17174,68868	-0,06255955
8	847189,1262	362810,8738	1,321554379
9	-138227,5155	168987,5155	0,615544371
10	-367423,9742	368853,9742	1,343566635
11	334336,5074	-156896,5074	-0,57150235
12	-25570,20315	70010,20315	0,255015208
13	161343,0505	-47973,0505	-0,17474392
14	776735,6846	1433264,315	5,220727573
15	592959,6597	-135409,6597	-0,49323557
16	-2843,82489	56313,82489	0,205125555
17	255735,9879	-220595,9879	-0,80353048
18	377446,5263	-241016,5263	-0,87791317
19	353940,9319	-144460,9319	-0,52620522
20	271137,8586	-90547,85858	-0,32982451
21	203896,4444	-128656,4444	-0,46863669
22	307611,5315	-185561,5315	-0,67591594
23	628375,8239	67724,1761	0,246688255
24	88394,20118	16965,79882	0,06179866
25	168192,1284	20647,87159	0,07521077
26	264333,872	-146473,872	-0,53353745
27	226329,5378	-106889,5378	-0,38934979
28	-38774,96299	100054,963	0,364454552
29	223118,6787	-170808,6787	-0,62217804
30	208182,283	-124712,283	-0,45426991
31	-147412,9633	158522,9633	0,577426785
32	5237,353707	57372,64629	0,208982358
33	374956,0104	-130046,0104	-0,47369825
34	158311,7451	-114401,7451	-0,41671333
35	93733,02305	-20133,02305	-0,07333541

36	56853,72013	6446,279866	0,023480855
37	191537,6657	-105737,6657	-0,38515404
38	305642,2306	-127912,2306	-0,46592586
39	420805,7789	-77935,77886	-0,28388446
40	62960,93699	-19910,93699	-0,07252645
41	206441,9847	-41601,98472	-0,15153704
42	-40790,62182	126040,6218	0,459108444
43	-402401,2398	408261,2398	1,487109313
44	50299,80695	-9079,806955	-0,03307359
45	91991,68889	18198,31111	0,066288139
46	250315,6955	-53865,69547	-0,19620814
47	87457,96386	-237,9638562	-0,00086679
48	113130,2208	-76300,22077	-0,27792687
49	56847,77628	2572,223724	0,009369437

**Anexo 5:** Resultados de datos de probabilidad

<i>Percentil</i>	<i>OYENTES MENSUALES// consumidor</i>
1	1430
3	5860
5	11110
7	30760
9	35140
11	36830
13	41220
15	43050
17	43910
19	44440
21	52310
23	53470
25	59420
27	61280
29	62610
31	63300
33	73600
35	75240
37	83470
39	85250

41	85800
43	87220
45	102170
47	105360
49	110190
51	113370
53	117860
55	119440
57	122050
59	136430
61	164840
63	177440
65	177730
67	180590
69	188840
71	196450
73	209480
75	244910
77	315270
79	342870
81	428600
83	457550
85	644100
87	647960
89	696100
91	1210000
93	1220000
95	1470000
97	2210000

**Anexo 6:** Tabla de significancia de la prueba de hipótesis.

<b>F</b>	<b>Altura del Gráfico</b>	<b>Intervalo de Confianza</b>	<b>Probabilidad Acumulada</b>	<b>Estadístico de Prueba</b>
0	0	0	0	
0,25	0,50815565	0,50815565	0,03060997	
0,5	0,73722441	0,73722441	0,1708686	
0,75	0,70710878	0,70710878	0,36850671	
1	0,57801364	0,57801364	0,55515121	

1,25	0,43578617	0,43578617	0,70182897
1,5	0,31376307	0,31376307	0,80627516
1,75	0,21970946	0,21970946	0,87651323
2	0,15123069	0,15123069	0,92215073
2,25	0,10300894		0,95119664
2,5	0,06973729		0,96946228
2,75	0,04706709		0,98087696
3	0,03173581		0,9879935
3,25	0,02141013		0,99243191
3,5	0,01446784		0,9952063
3,75	0,0098006		0,99694681
4	0,00665921		0,99804373
4,25	0,00454049		0,99873869
4,5	0,00310764		0,9991815
4,75	0,00213552		0,99946536
5	0,00147365		0,99964847
5,25	0,00102128		0,99976734
5,5	0,00071087		0,999845
5,75	0,00049698		0,99989607
6	0,00034897		0,99992987
6,25	0,00024612		0,99995238
6,5	0,00017434		0,99996747
6,75	0,00012403		0,99997764
7	8,8619E-05		0,99998454
7,25	6,3584E-05		0,99998925
7,5	4,5811E-05		0,99999248
7,75	3,3142E-05		0,99999471
8	2,4073E-05		0,99999626
8,25	1,7555E-05		0,99999734
8,5	1,2852E-05		0,9999981
8,75	9,4449E-06		0,99999864
9	6,9671E-06		0,99999902
9,25	5,1582E-06		0,99999929
9,5	3,8328E-06		0,99999948
9,75	2,858E-06		0,99999962
10	2,1385E-06		0,99999972
10,25	1,6056E-06		0,99999979
10,5	1,2095E-06		0,99999985
10,75	9,141E-07		0,99999989

11	6,9306E-07		0,99999992
11,25	5,2711E-07		0,99999994
11,5	4,0213E-07		0,99999995
11,75	3,0771E-07		0,99999996
12	2,3615E-07		0,99999997
12,25	1,8176E-07		0,99999998
12,5	1,4029E-07		0,99999998
12,75	1,0859E-07		0,99999999
13	8,4277E-08		0,99999999
13,25	6,5583E-08		0,99999999
13,5	5,1169E-08		0,99999999
13,75	4,0026E-08		1
14	3,1388E-08		1
14,25	2,4675E-08		1
14,5	1,9445E-08		1
14,75	1,536E-08		1
15	1,2161E-08		1
15,25	9,6508E-09		1
15,5	7,6757E-09		1
15,75	6,1184E-09		1
16	4,8875E-09		1
16,25	3,9126E-09		1
16,5	3,1387E-09		1
16,75	2,523E-09		1
17	2,0322E-09		1
17,25	1,6401E-09		1
17,5	1,3262E-09		1
17,75	1,0745E-09		1
18	8,7213E-10		1
18,25	7,0921E-10		1
18,5	5,7776E-10		1
18,75	4,7152E-10		1
19	3,8549E-10		1
19,25	3,157E-10		1
19,5	2,5898E-10		1
19,75	2,128E-10		1
20	1,7515E-10		1
20,25	1,4439E-10		1
20,5	1,1922E-10		1

20,75	9,8593E-11		1
21	8,166E-11		1
21,25	6,7737E-11		1
21,5	5,6272E-11		1
21,75	4,6816E-11		1
22	3,9005E-11		1
22,25	3,2543E-11		1
22,5	2,719E-11		1
22,75	2,2749E-11		1
23	1,9059E-11		1
23,25	1,5989E-11		1
23,5	1,3431E-11		1
23,75	1,1297E-11		1
24	9,5145E-12		1
24,25	8,0231E-12		1
24,5	6,774E-12		1
24,75	5,7263E-12		1
25	4,8465E-12		1
25,25	4,1068E-12		1
25,5	3,484E-12		1
25,75	2,9591E-12		1
26	2,5162E-12		1
26,25	2,142E-12		1
26,5	1,8254E-12		1
26,75	1,5574E-12		1
27	1,3301E-12		1
27,25	1,1372E-12		1
27,5	9,7334E-13		1
27,75	8,3393E-13		1
28	7,1522E-13		1
28,25	6,1404E-13		1
28,5	5,2769E-13		1
28,75	4,5393E-13		1
29	3,9086E-13		1
29,25	3,3688E-13		1
29,5	2,9063E-13		1
29,75	2,5096E-13		1
30	2,1691E-13		1
30,25	1,8765E-13		1

30,5	1,6248E-13		1	0,3
30,75	1,4082E-13		1	
31	1,2215E-13		1	