

Informe Técnico – Proyecto Spotify

Santiago Arboleda

2025-09-01

Cargar datos

```
# Lee el archivo CSV ubicado en la misma carpeta que este .Rmd
# Si está en otra ruta, actualiza el path, por ejemplo: "data/Spotify2024_clean.csv"
spotify <- readr::read_csv("Spotify2024_clean.csv", show_col_types = FALSE)

# Verificación de columnas requeridas
req_cols <- c(
  "cancion", "album", "artista", "fecha_lanz", "isrc", "ranking_hist", "puntaje",
  "spotify_streams", "spotify_playlists", "spotify_alcance", "spotify_pop",
  "youtube_views", "youtube_likes", "tiktok_posts", "tiktok_likes", "tiktok_views",
  "youtube_alcance", "apple_playlists", "airplay_spins", "sirius_spins",
  "deezer_playlists", "deezer_alcance", "amazon_playlists", "pandora_streams",
  "pandora_stations", "soundcloud_streams", "shazam", "tidal_pop", "explicito"
)

faltan <- setdiff(req_cols, names(spotify))
stopifnot("Faltan columnas requeridas en 'Spotify2024_clean.csv'." = length(faltan) == 0)
```

Resumen del Proyecto y Objetivos

Este **informe técnico** breve resume los hallazgos principales del dataset **Spotify 2024** en relación con los objetivos de:

- 1) **Cuantificar** el consumo y popularidad de canciones,
- 2) **Explorar** la relación entre plataformas (Spotify y YouTube), y
- 3) **Sintetizar** indicadores clave para orientar decisiones.

Tabla de Indicadores (Resumen)

```
resumen <- spotify %>%
  summarise(
    Total_Canciones      = dplyr::n_distinct(cancion),
    Total_Artistas       = dplyr::n_distinct(artista),
    Promedio_Streams_Spotify = mean(spotify_streams, na.rm = TRUE),
    Max_Streams_Spotify    = max(spotify_streams, na.rm = TRUE),
```

```

Popularidad_Spotify_media = mean(spotify_pop, na.rm = TRUE),
Vistas_YT_media          = mean(youtube_views, na.rm = TRUE),
Vistas_TikTok_media      = mean(tiktok_views, na.rm = TRUE)
)

knitr::kable(resumen, caption = "Tabla 1. Indicadores resumidos - Spotify 2024")

```

Table 1: Tabla 1. Indicadores resumidos – Spotify 2024

Total_Canciones	Total_Artistas	Promedio_Streams	Max_Streams_Spotify	Popularidad_Spotify_media	Vistas_YT_media	Vistas_TikTok_media
4365	2065	437916567	4281468720	651429.5	396245079	1144741630

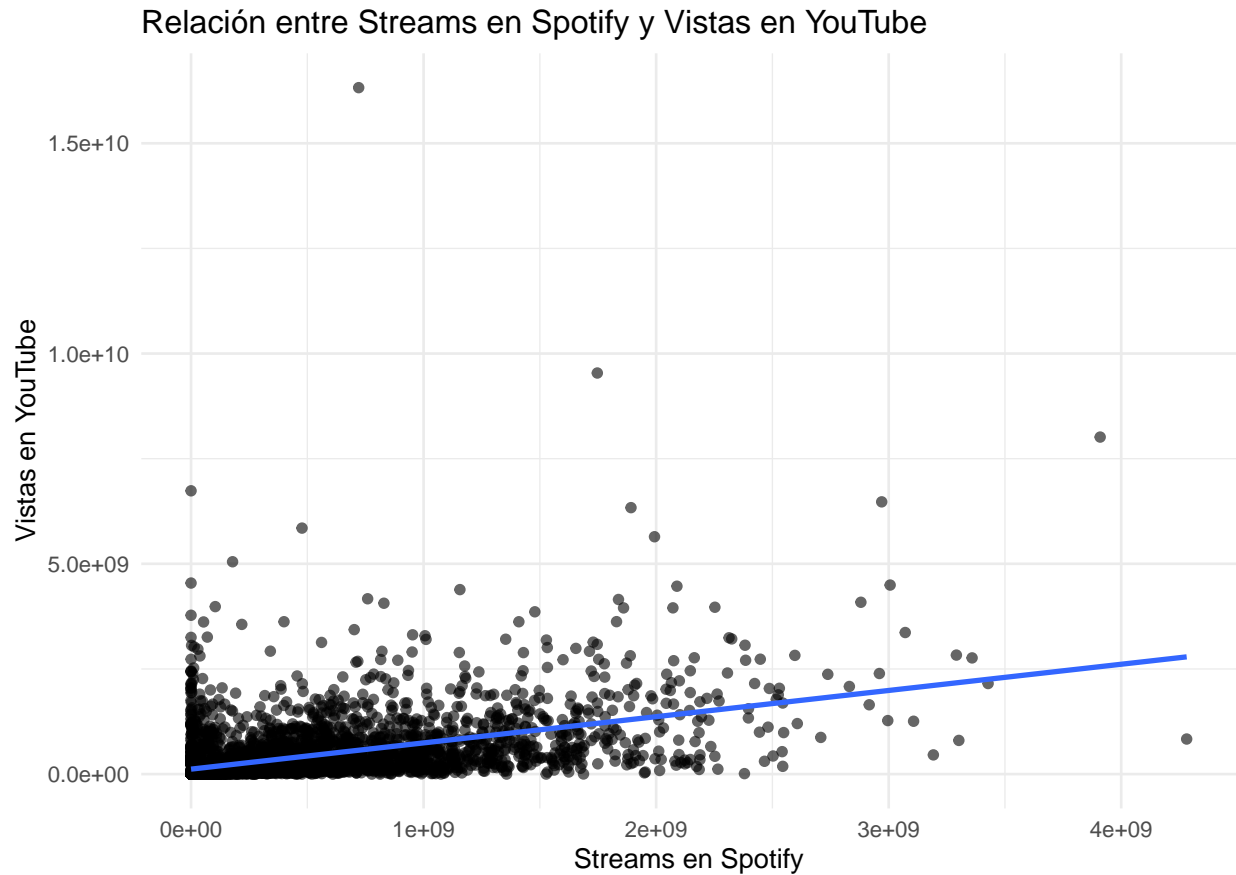
El dataset contiene **4365** canciones de **2065** artistas.
 En promedio, se registran **437,916,567 streams en Spotify**,
 con un **máximo** de **4,281,468,720**.
 La **popularidad media** en Spotify es 6.514295×10^5 puntos.

Gráfico de Resultados (Principal)

```

ggplot(spotify, aes(x = spotify_streams, y = youtube_views)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", se = FALSE) +
  labs(title = "Relación entre Streams en Spotify y Vistas en YouTube",
        x = "Streams en Spotify",
        y = "Vistas en YouTube") +
  theme_minimal()

```



La correlación entre **streams en Spotify** y **vistas en YouTube** es **0.472**.
 Este resultado **apoya** la expectativa de una asociación positiva entre plataformas.

Explicación e Interpretación (con resultados de R)

Para cuantificar la relación anterior, se ajusta un **modelo lineal simple**:

```
modelo <- lm(youtube_views ~ spotify_streams, data = spotify)
summary(modelo)
```

```
##
## Call:
## lm(formula = youtube_views ~ spotify_streams, data = spotify)
##
## Residuals:
```

	Min	1Q	Median	3Q	Max
	-1.954e+09	-2.084e+08	-1.192e+08	1.978e+07	1.576e+10

```
##
## Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.183e+08	1.224e+07	9.664	<2e-16 ***
spotify_streams	6.235e-01	1.797e-02	34.705	<2e-16 ***

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 611600000 on 4209 degrees of freedom
## (387 observations deleted due to missingness)
## Multiple R-squared:  0.2225, Adjusted R-squared:  0.2223
## F-statistic: 1204 on 1 and 4209 DF,  p-value: < 2.2e-16
```

El coeficiente sugiere que, por cada unidad adicional de **streams en Spotify**, las **vistas en YouTube** cambian en

0.6235 unidades (aprox.).

El **R²** del modelo es **0.2225**, lo que indica el grado de variación explicada.

Conclusiones

- Los **indicadores clave** sintetizan el nivel de consumo y popularidad (Tabla 1).
- El **gráfico principal** y el **modelo lineal** evidencian la relación entre plataformas, útil para promoción y análisis de rendimiento.
- Los resultados están alineados con los **objetivos del proyecto**, proporcionando una base para decisiones informadas.