

Distribución Digital de la Música

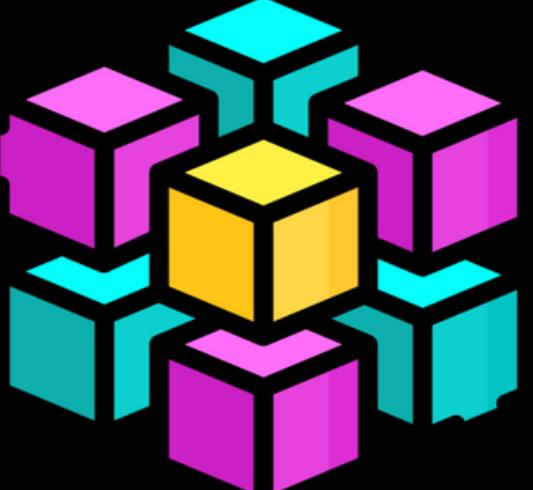
01 / Introducción

Analizar la distribución digital
de la música en



Identificar patrones de
popularidad y su relación con
características musicales

DATASET



20.718



2.079



02/Variables

Características musicales

Bailabilidad



Tempo
Estabilidad del ritmo

Energía



Intensidad
Velocidad

Emoción



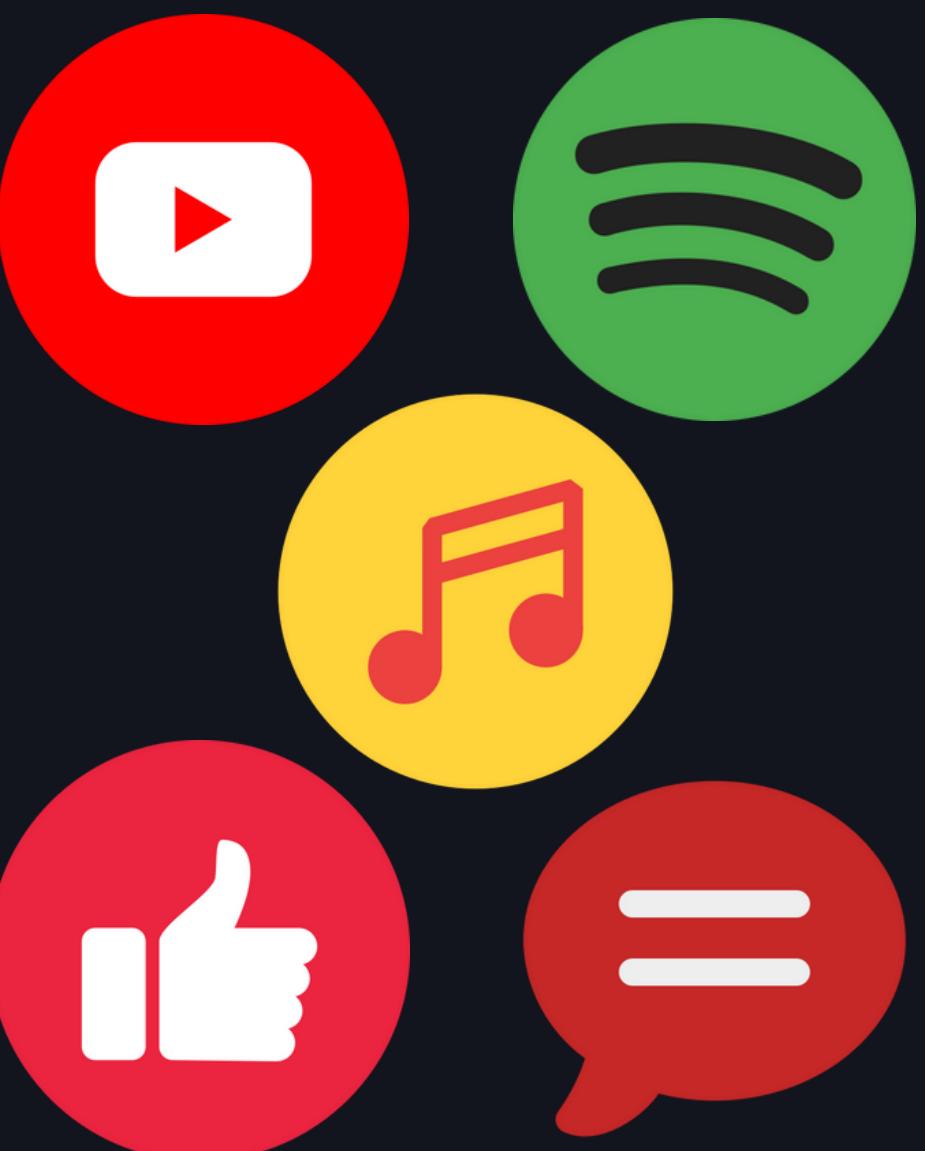
Positivas - Alegres
Negativas - Tristes

Sonoridad



Percepción del
nivel de sonido

Métricas



Duración



Duración de la
canción en min

Acústica



Probabilidad de
que se acústica

En vivo



Presencia de
audiencia

Hablabilidad



Prsencia de
palabras habladas

03/Hipótesis

Hipótesis 1

Las canciones más populares son las más bailables, alegres y enérgicas



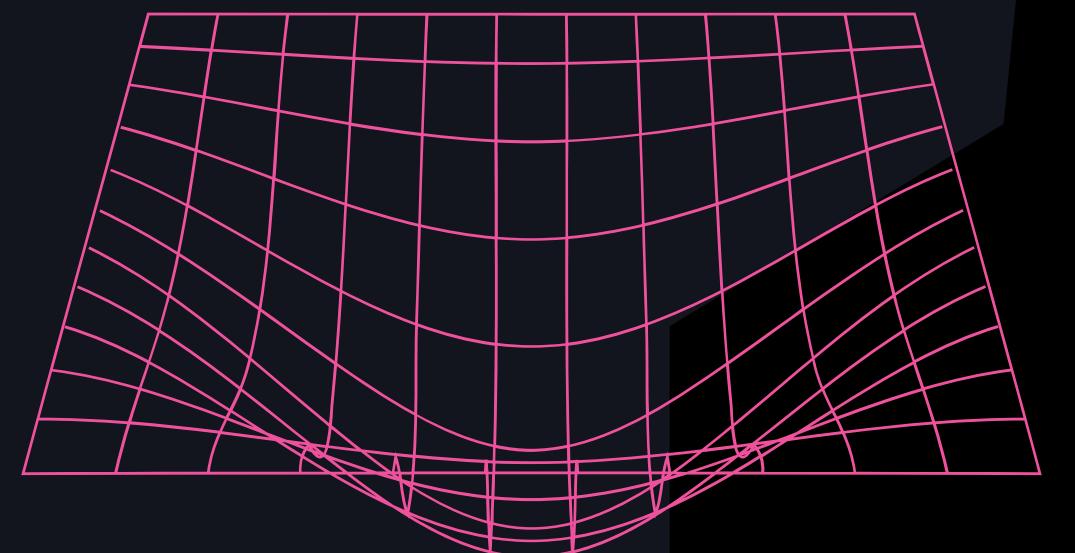
Hipótesis 2

Las canciones más escuchadas no pueden tener una duración muy larga.



Otras consultas

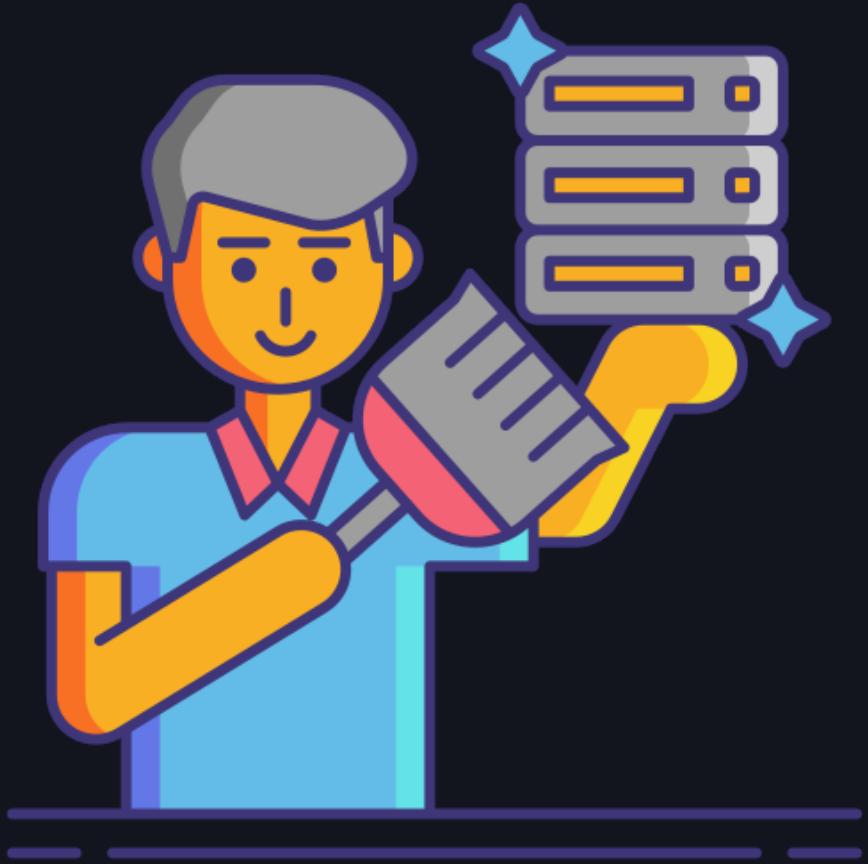
- Artistas más populares en Spotify
- Artistas más populares en Youtube
- Canciones más escuchadas en ambas plataformas
- Canciones con mayor interacción (likes, comentarios)



04 / Limpieza y tratamiento de Datos



Muy baja -> White Noise
Muy alta -> Audiolibros



Canción - Artista - Álbum

Meditation
Motion Picture
Movie Soundtrak
HBO - GOT
Sleep
Minecraft



Métricas nulas

Completamos con la media de la misma variable
por **artista**

No hay datos?
Views y likes = 0

Reproducciones y visualizaciones -> Estimador

05/ Análisis de Datos Univariantes

Bailabilidad

Resumen estadístico

Cantidad: 19537

Media: 0.625

Desvío: 0.159

Cuartiles:

min 0.064

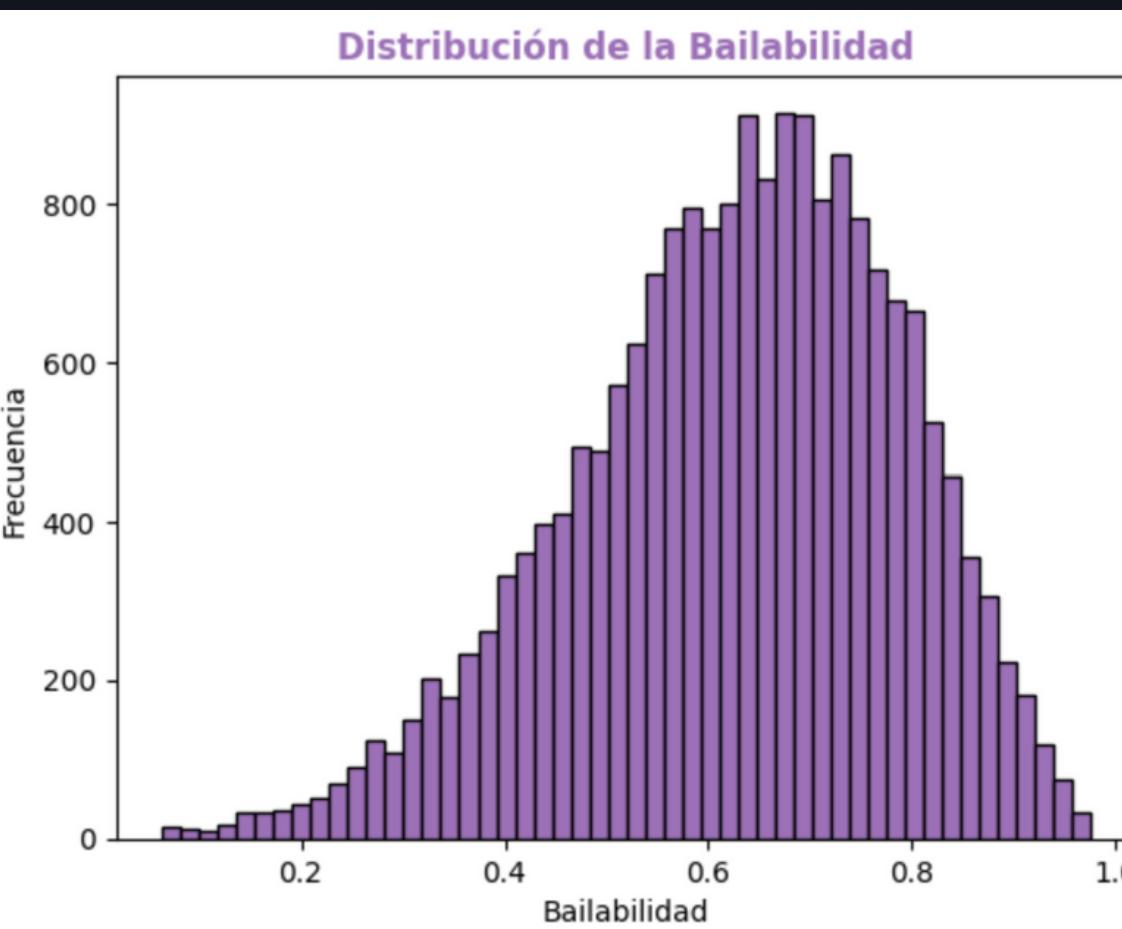
25% 0.524

50% 0.641

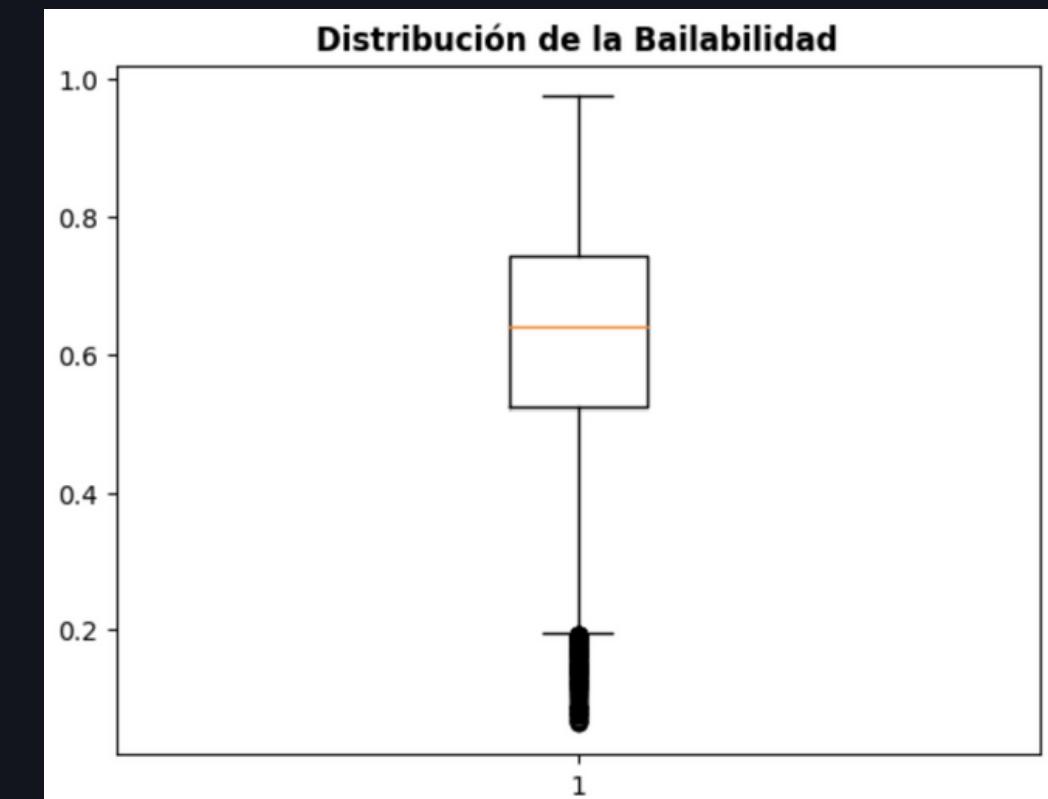
75% 0.744

max 0.975

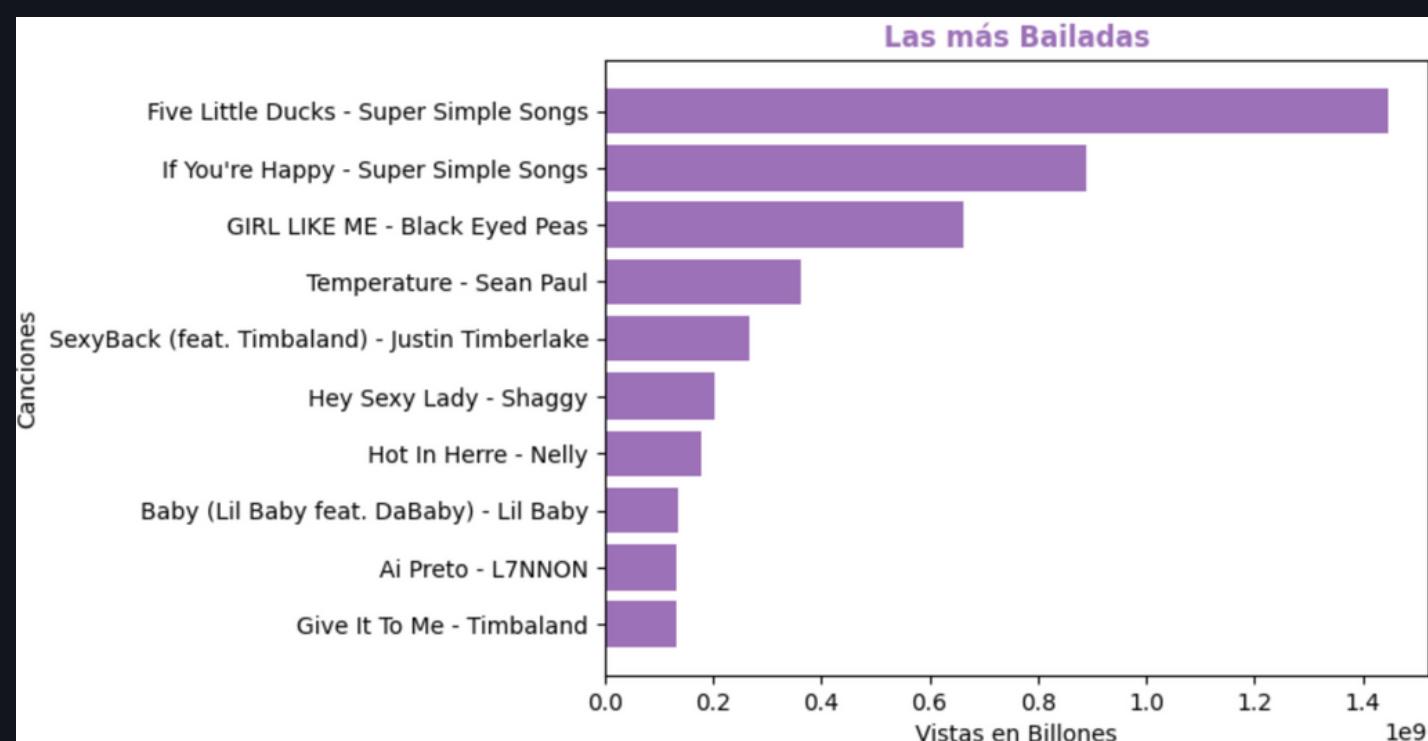
Histogramas



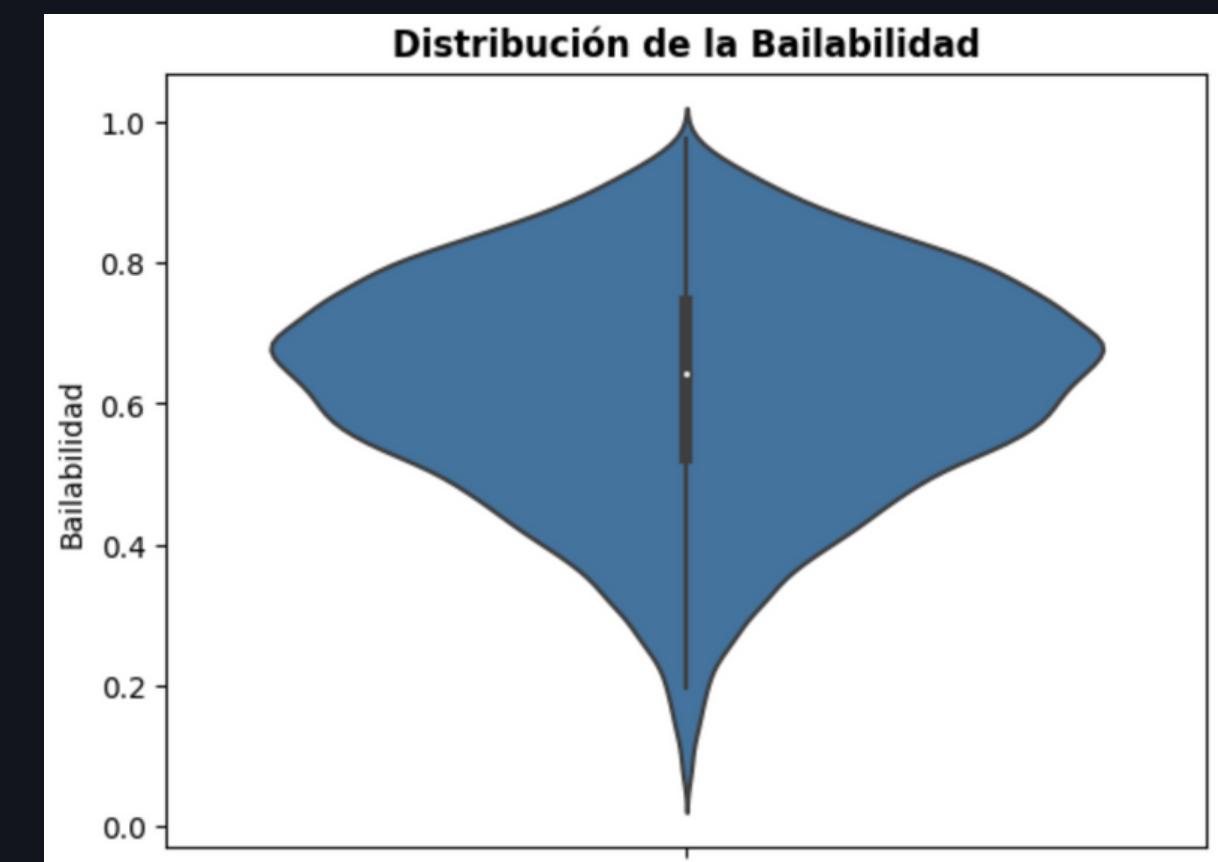
Gráficos de caja



Top 10 bailables



Gráficos de violín



05/ Análisis de Datos Univariable

Prueba de Normalidad

H₀: Distribución Normal

H₁: Distribución no Normal

Anderson Darling

Estadístico de Prueba=67.27 > Valor crítico = 0.78

Nivel significancia = 5 %

Se rechaza H₀ X

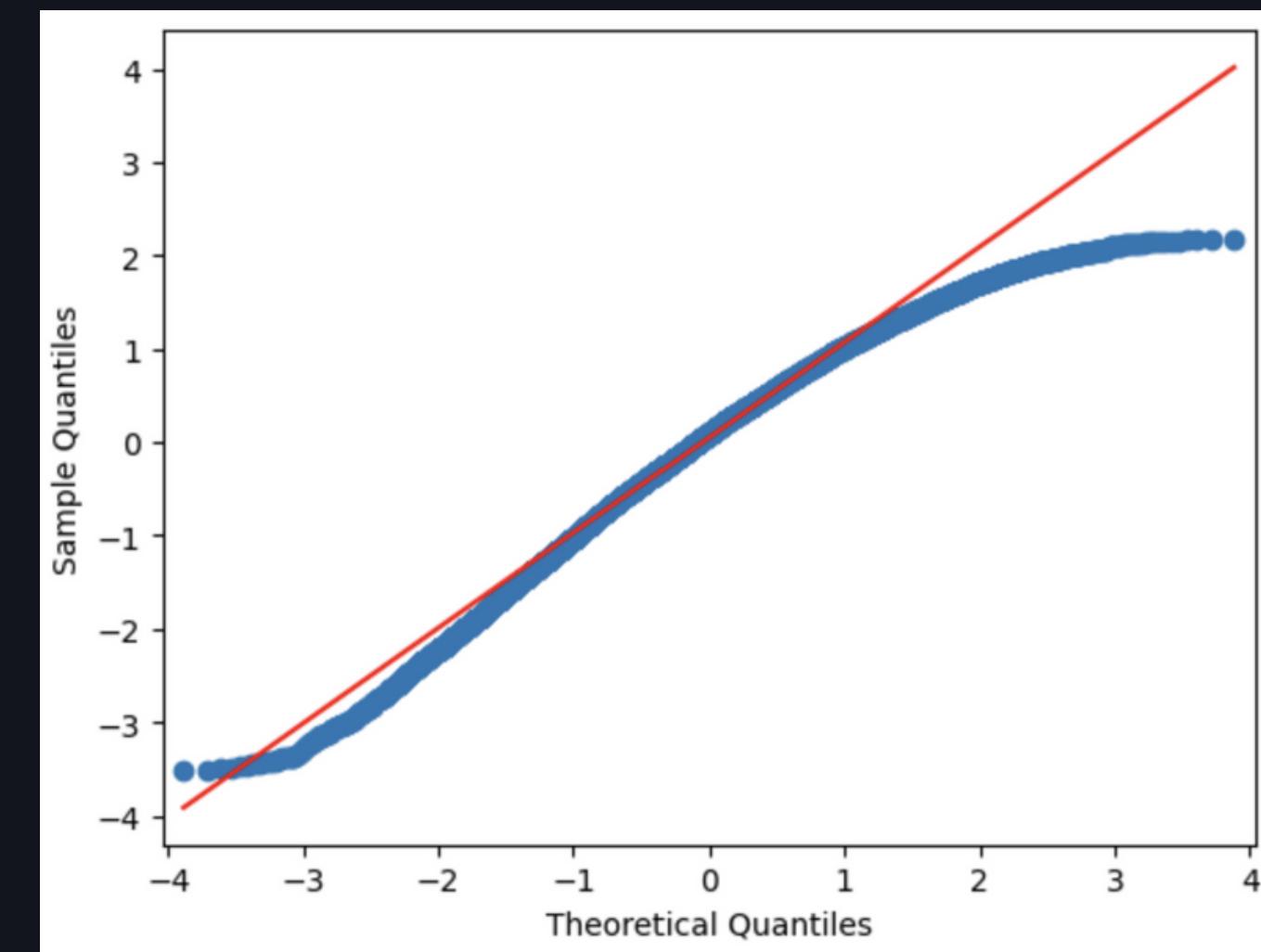
Normal Test

p_valor < Nivel de significancia

Se rechaza H₀ X

Gráfico de Normalidad QQPLOT

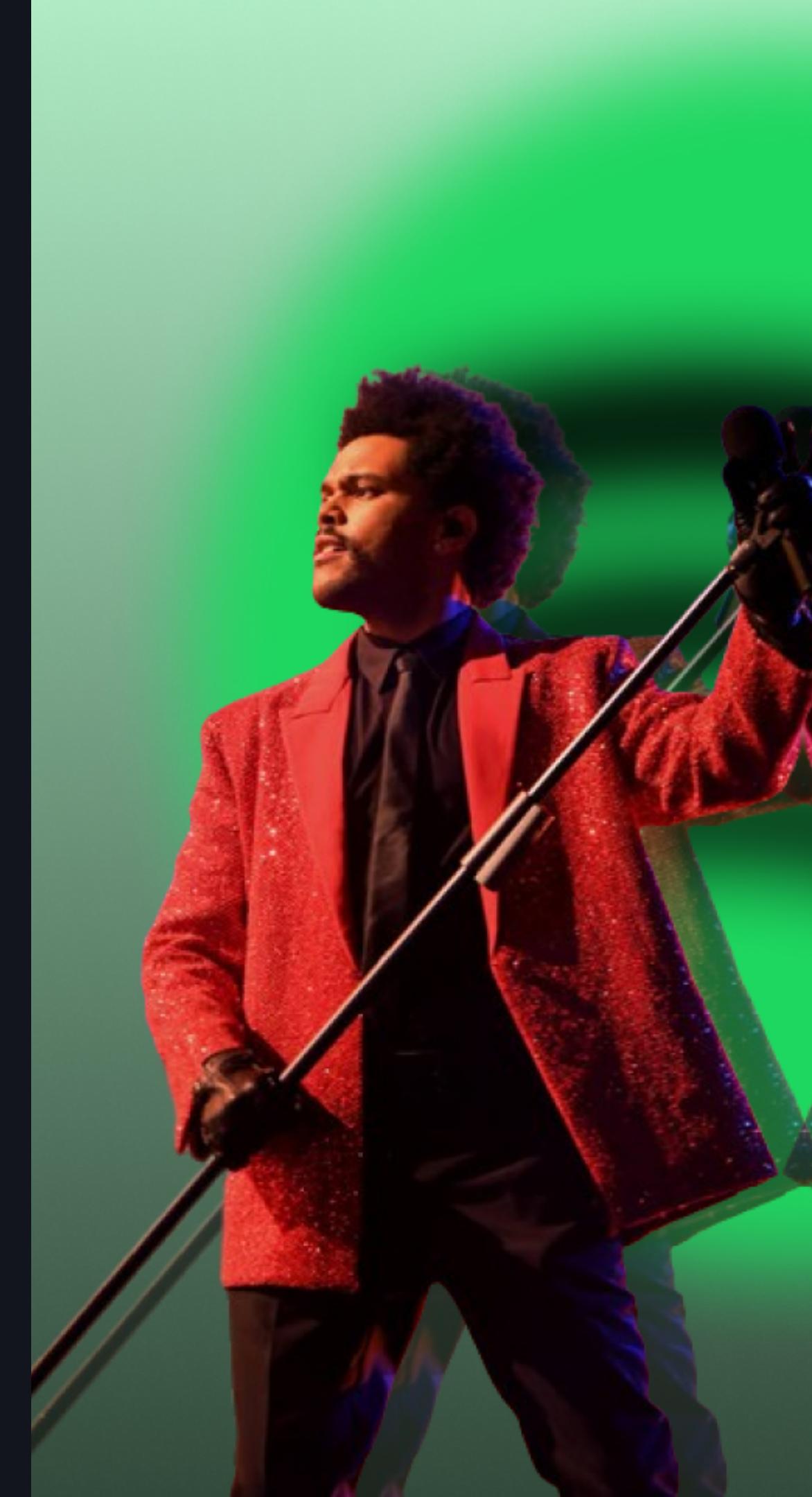
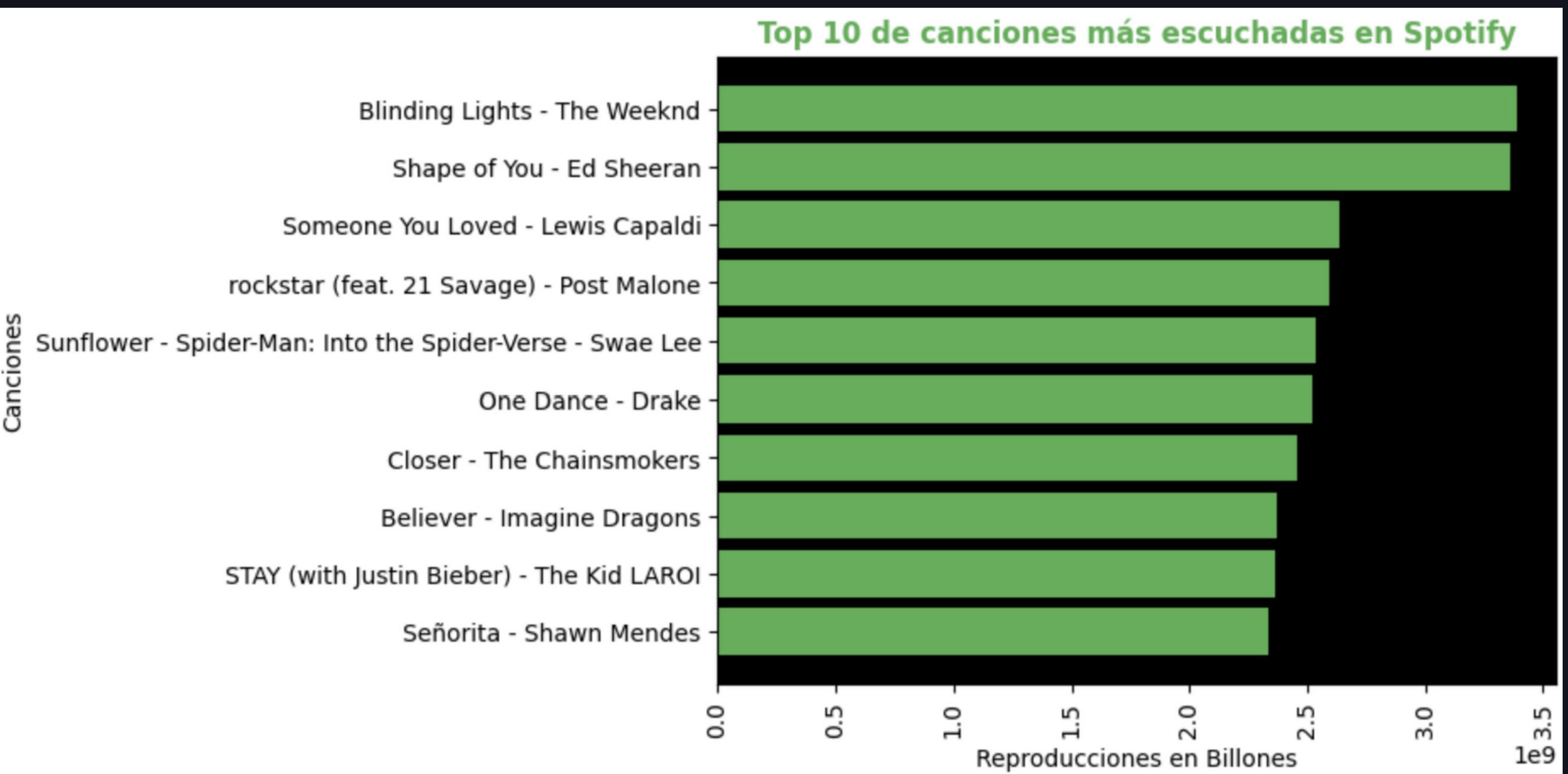
Bailabilidad



Valores se alejan de la linea roja
Distribución NO Normal

05/ Análisis de Datos

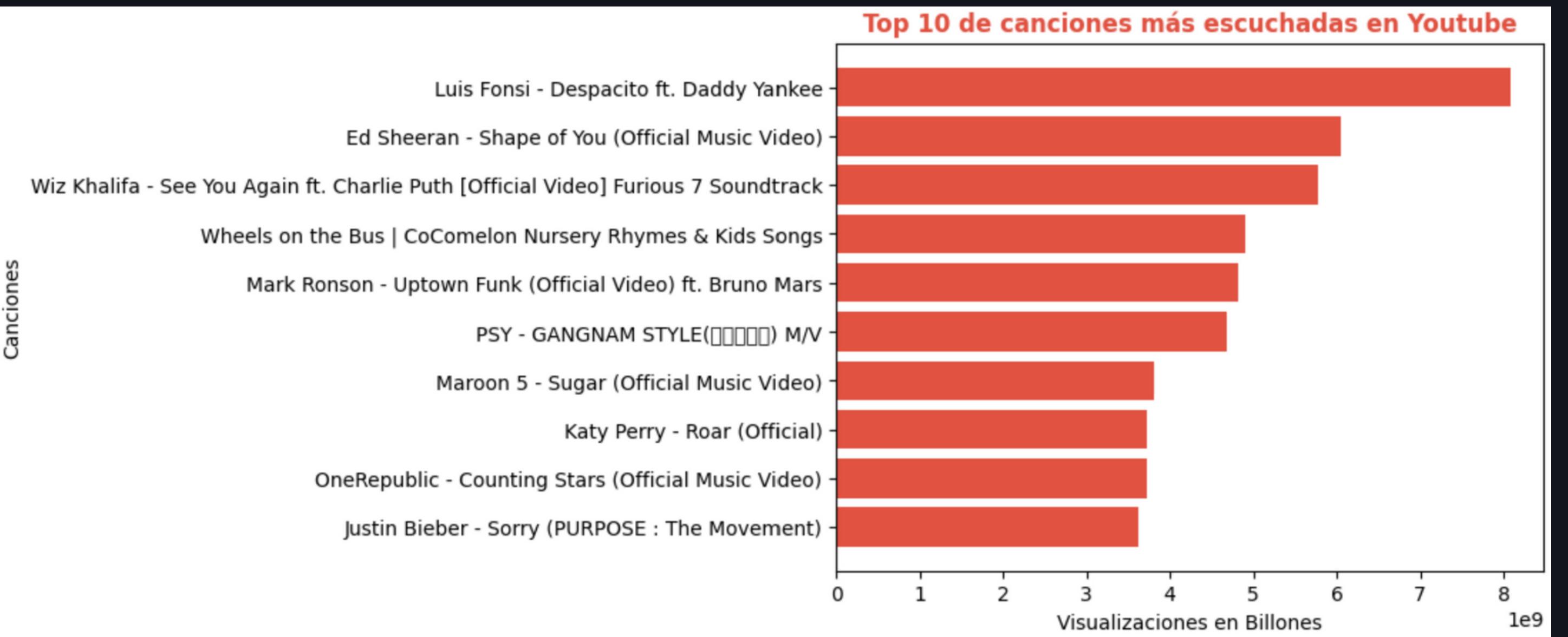
Métricas canciones



05/ Análisis de Datos

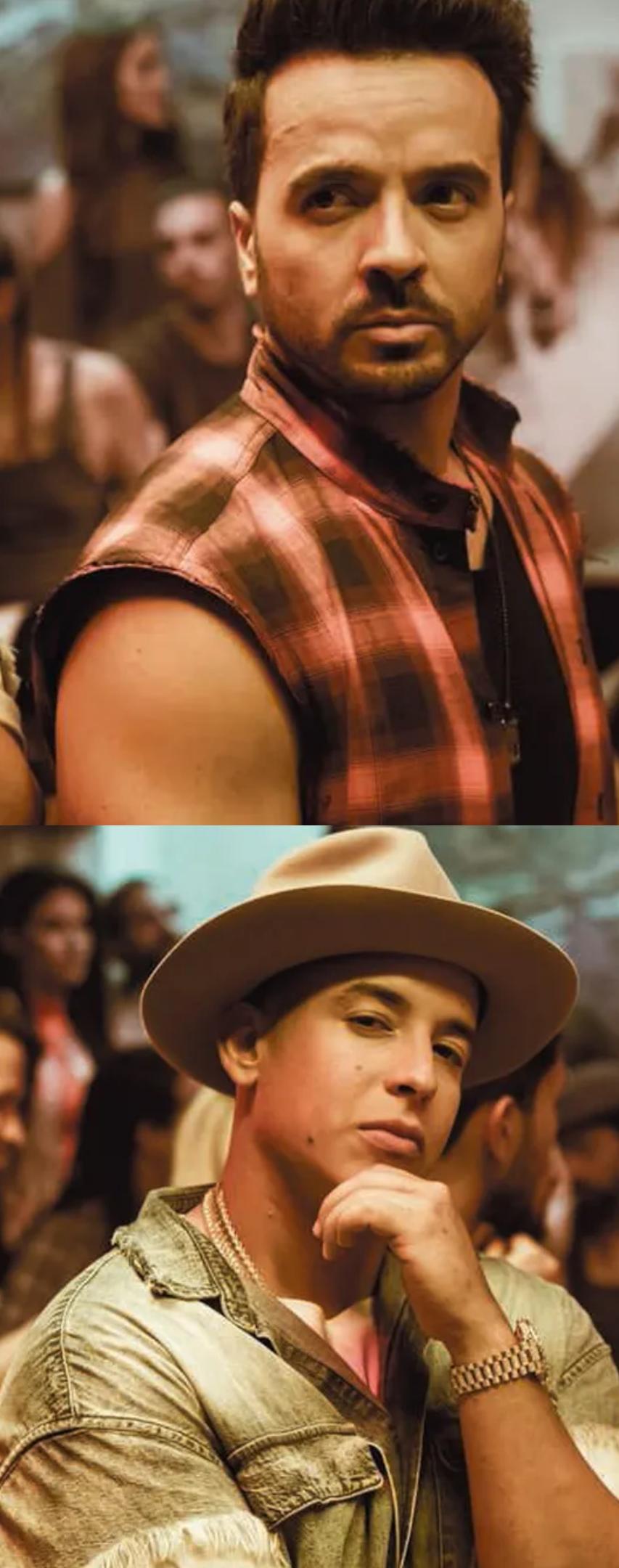
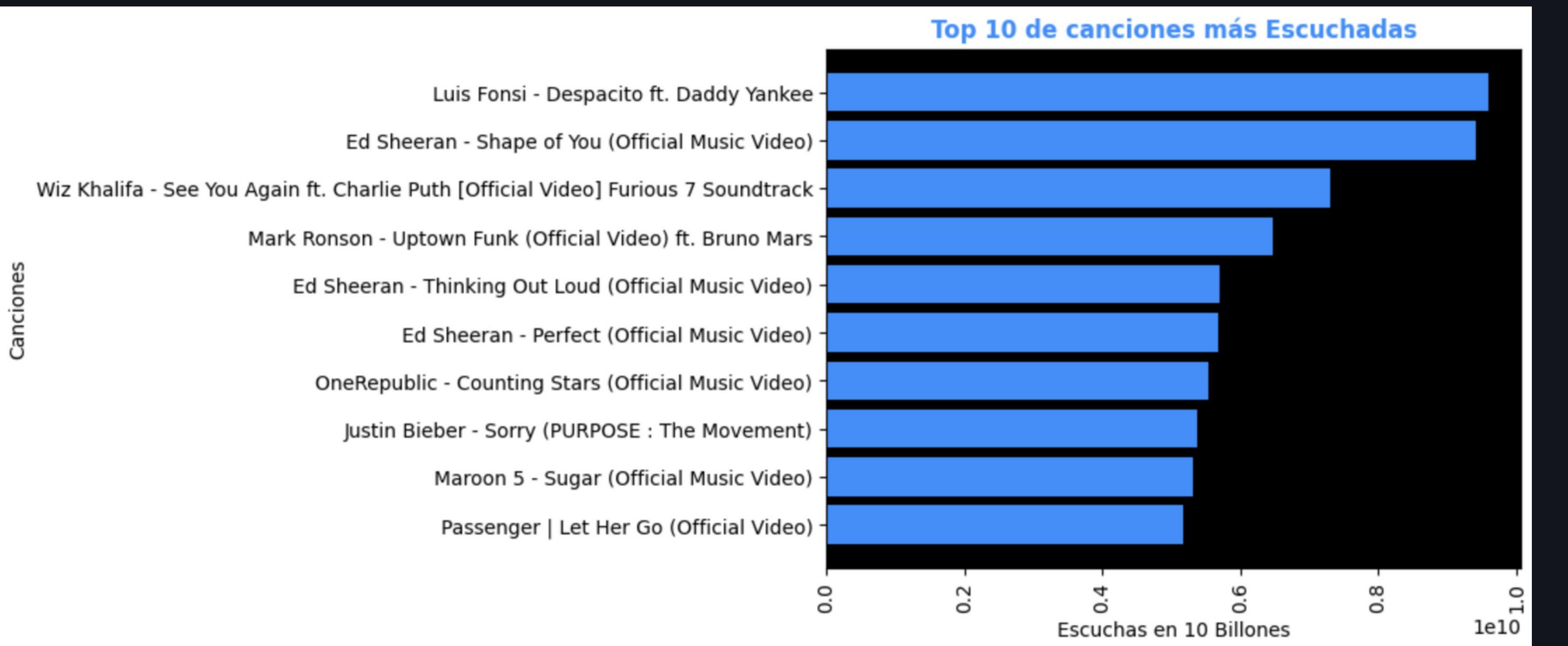


Métricas canciones



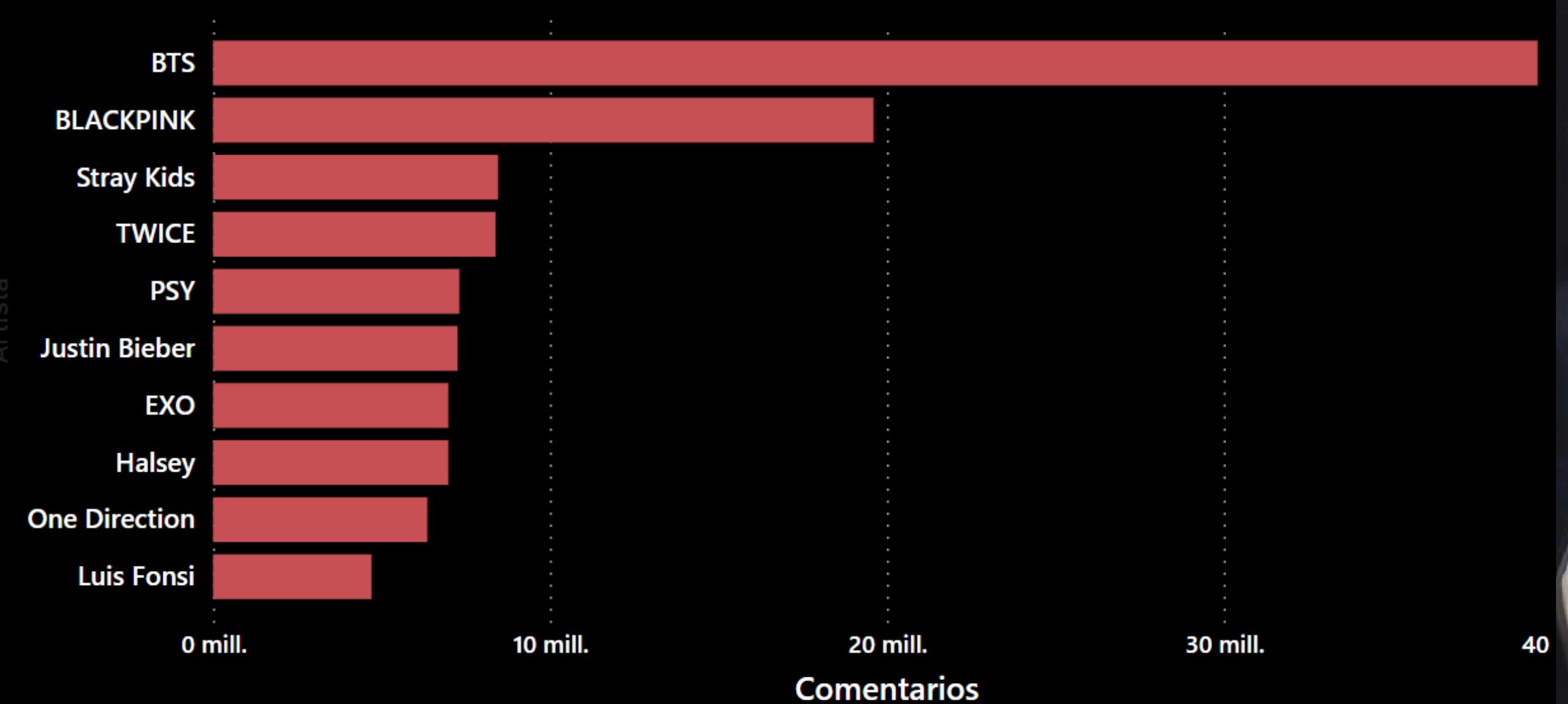
05/ Análisis de Datos

Métricas canciones



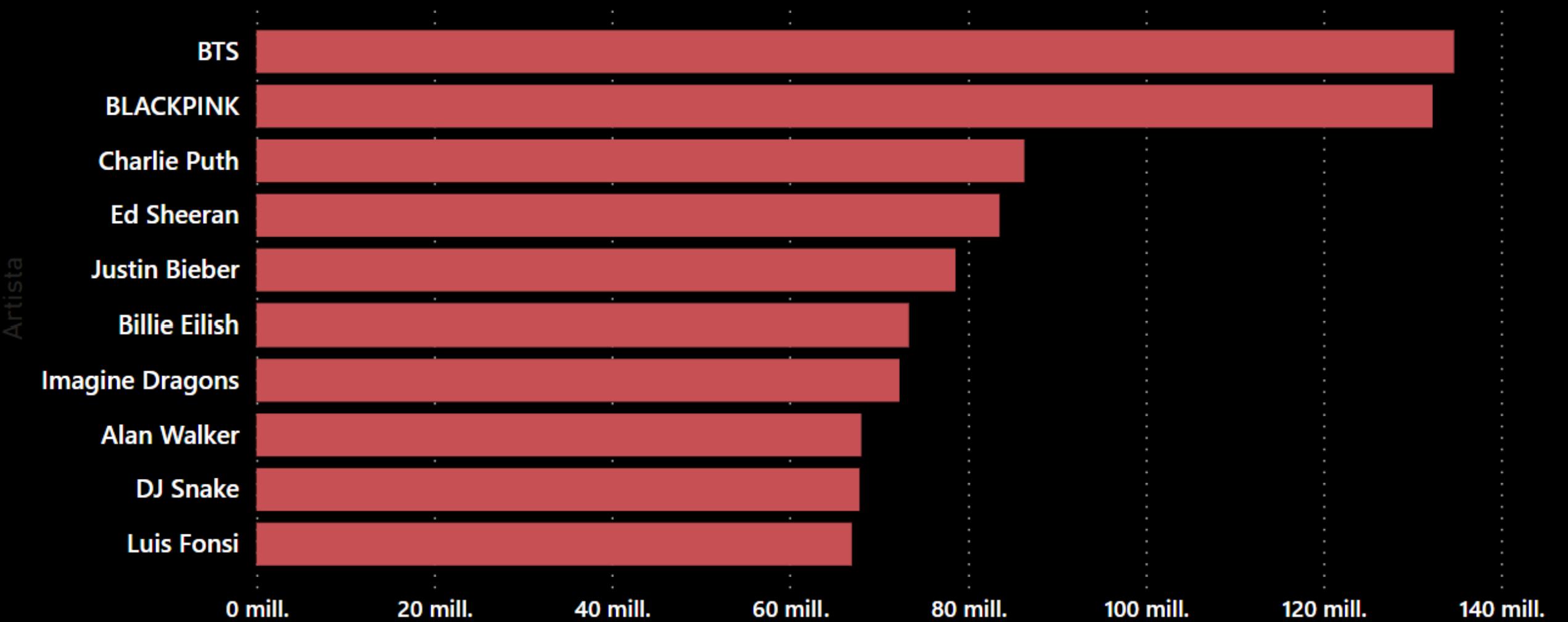
05/ Análisis de Datos

Métricas canciones



05/ Análisis de Datos

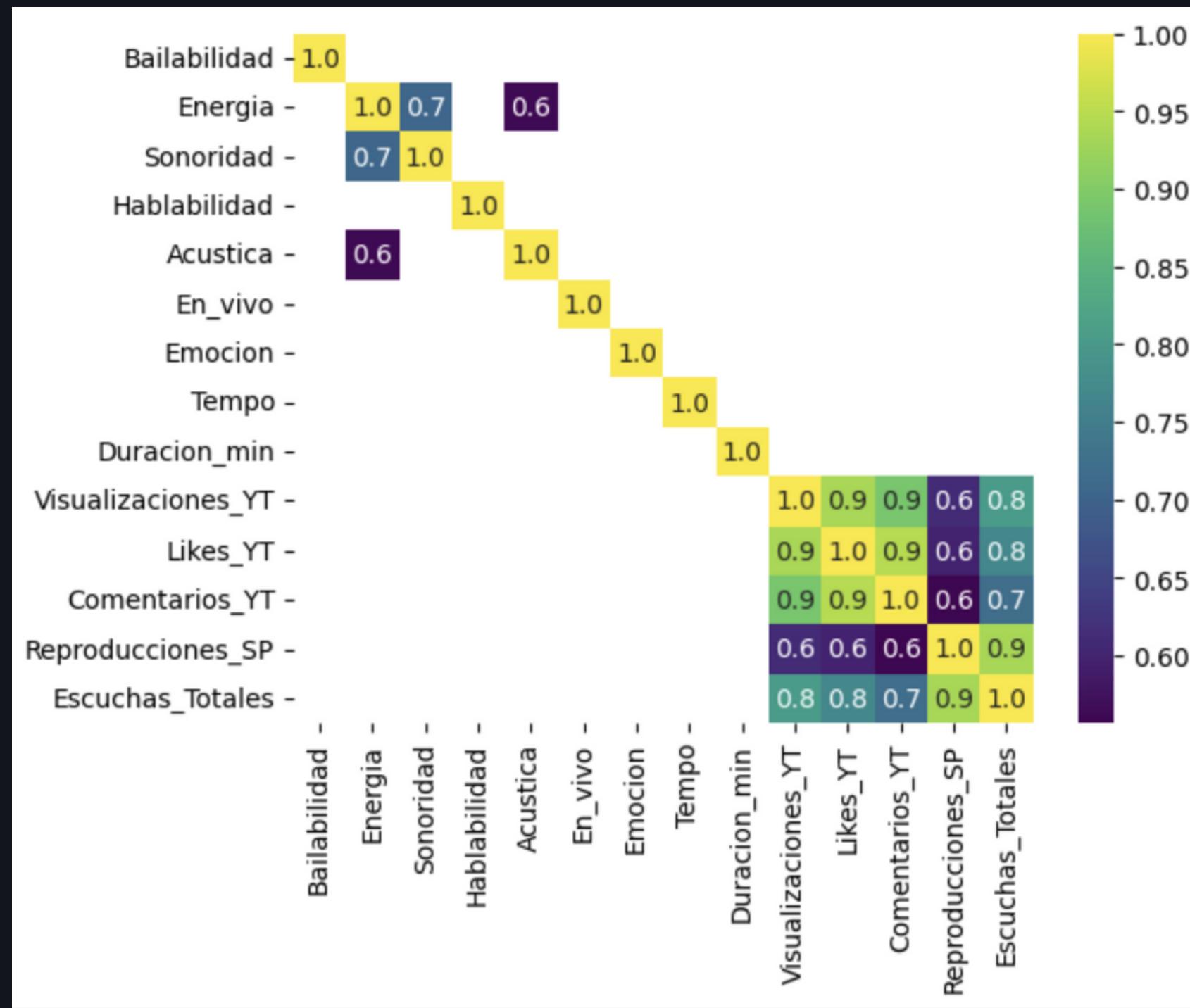
Métricas canciones



05/ Análisis de Datos

Correlación

Métricas alta correlación



Características Musicales
Energía - Sonoridad



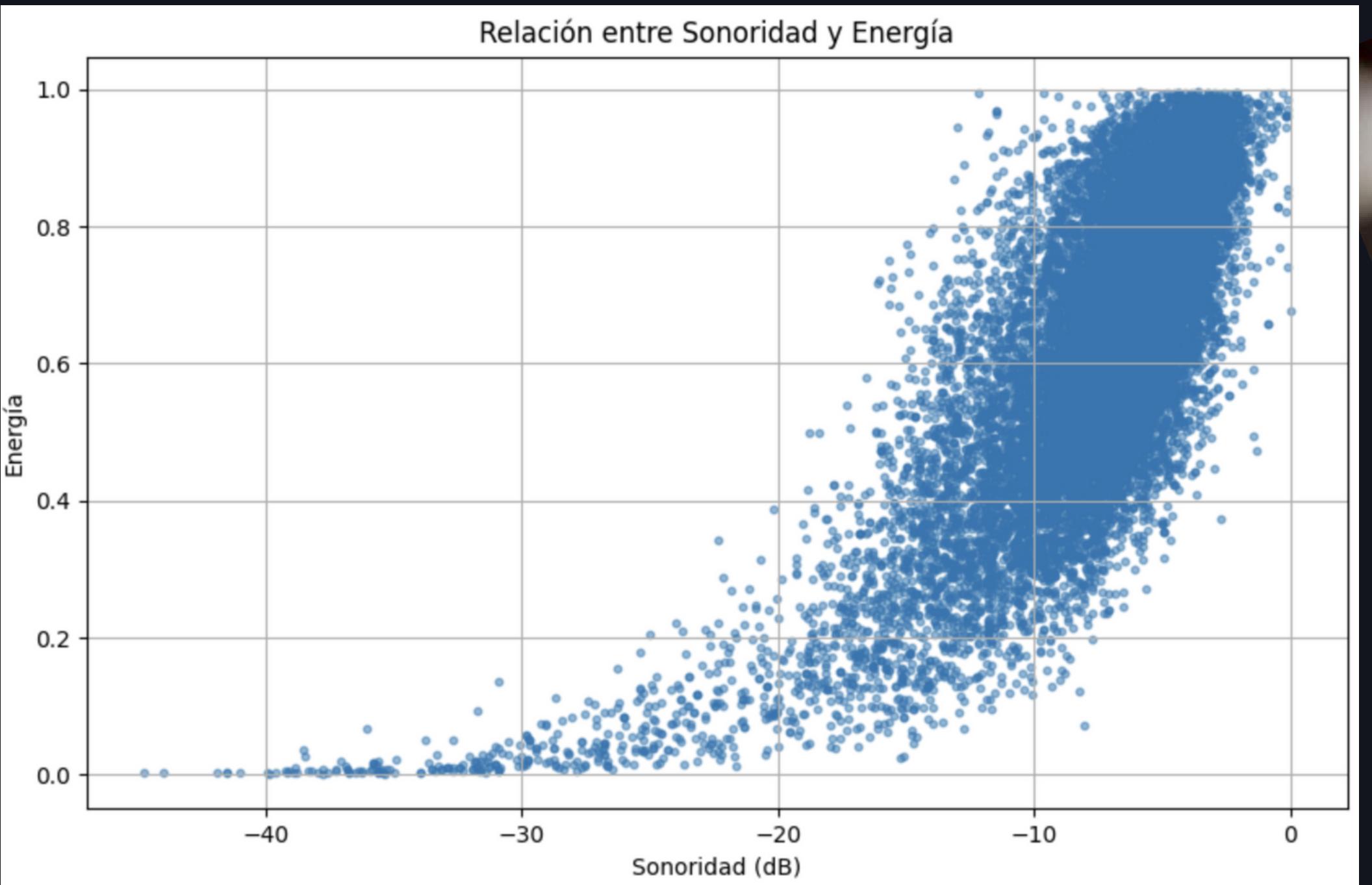
Energía - Acústica



05/ Análisis de Datos

Gráfico de dispersión

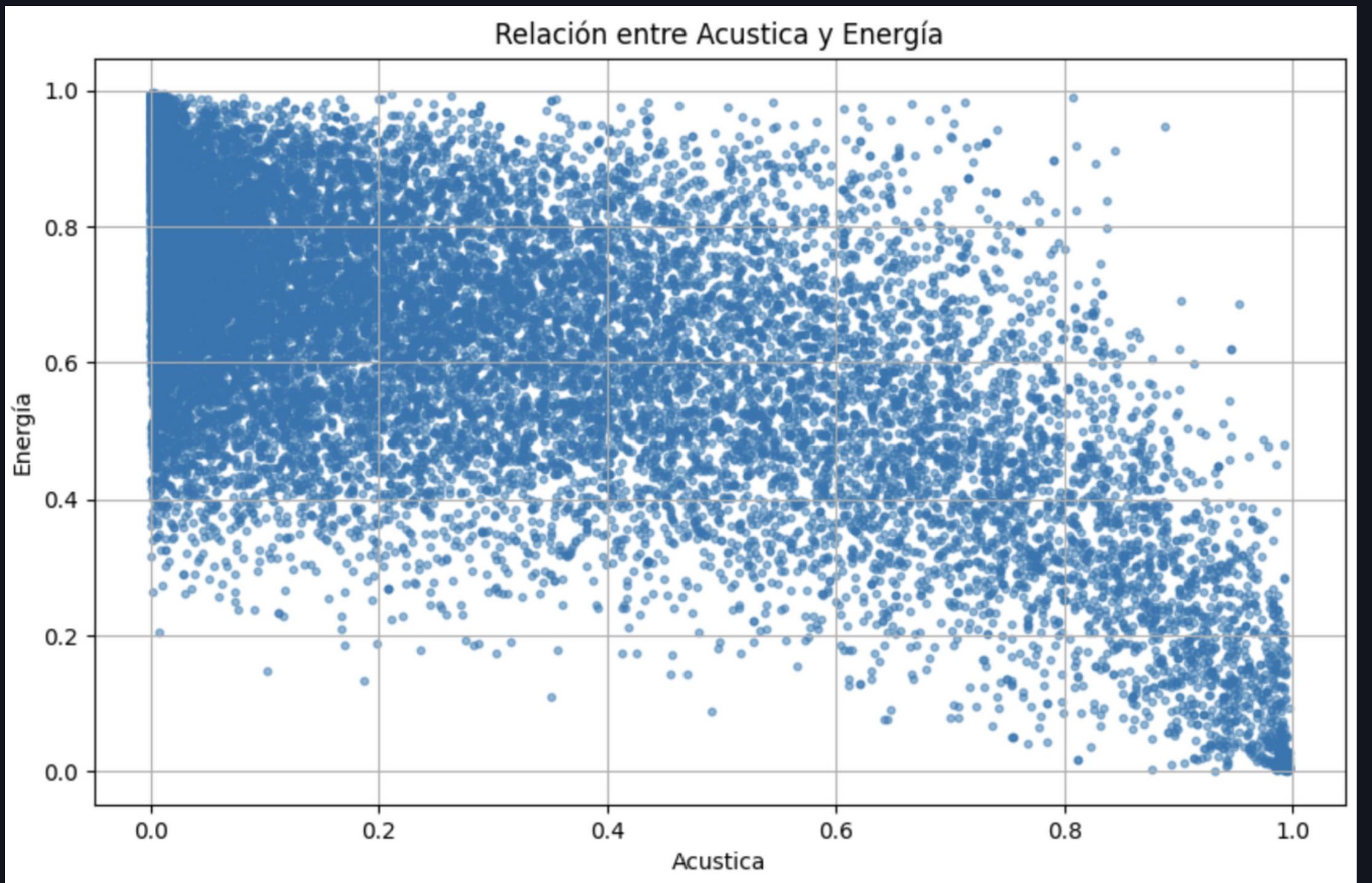
Cuanto más ruidosa, más energica



05/ Análisis de Datos

Grafico de dispersión

Cuanto más acústica, menos enérgica



06 / Test de Hipótesis 1

Hipótesis:

Las canciones más populares son las más bailables, alegres y enérgicas

Estudiamos si las medias poblacionales son iguales a las medias muestrales de las variables elegidas

μ = Media todas las canciones

X = Media Top 10



Prueba de Mann-Withney

Nivel de significación = 0.05

En las 3 variables **NO** podemos rechazar la Hipótesis nula

$$H_0: \mu = X$$

$$H_1: \mu \neq X$$

Bailabilidad: P_valor = 0.13 > 0.05
Emoción: P_valor = 0.65 > 0.05
Energía: P_valor = 0.49 > 0.05

$$H_0: \mu \geq X$$

$$H_1: \mu < X$$

En las 3 variables **NO** podemos rechazar la Hipótesis nula

Bailabilidad: P_valor = 0.06 > 0.05
Emoción: P_valor = 0.32 > 0.05
Energía: P_valor = 0.75 > 0.05

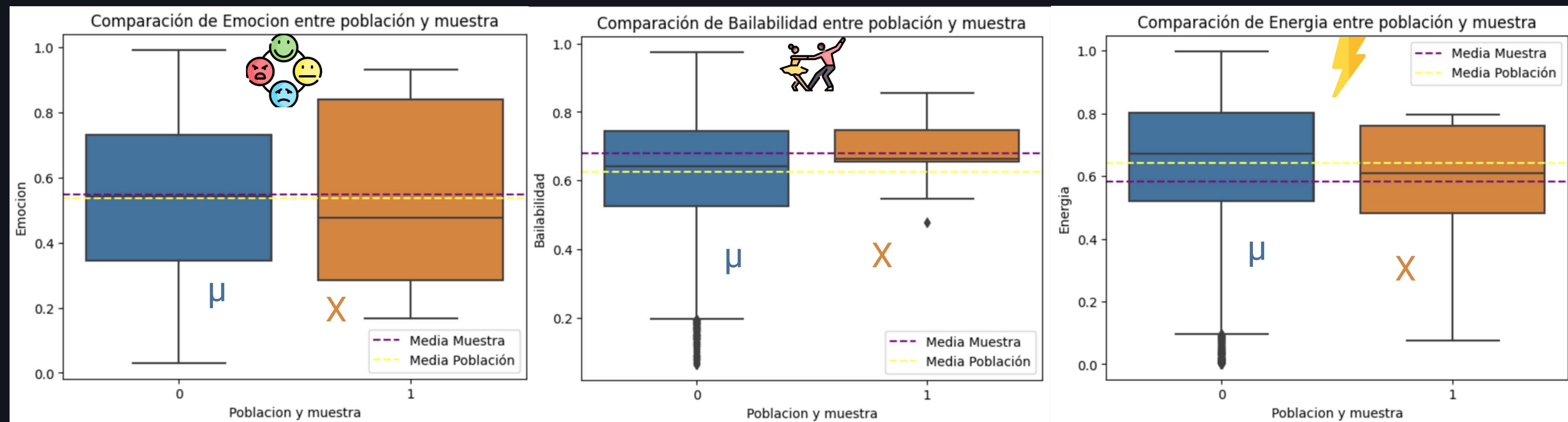
Conclusión: No podemos decir que las medias son significativamente **distintas**

Conclusión: No podemos decir que las medias poblacionales son significativamente **menores**

06 / Test de Hipótesis 1

Comparación graficos caja

Gráfico de caja para comparar las medias poblacionales y muestrales



Conclusión Hipótesis 1:

Las canciones populares tienden a ser más bailables, más alegres y menos enérgicas

06 / Test de Hipótesis 2

Hipótesis:

Las canciones más escuchadas no pueden tener una duración muy larga.

Agrupamos las canciones por su duración en minutos

Grupo 1: Canciones de 1 a 3 minutos.

Grupo 2: Canciones de más de 3 a 6 minutos.

Grupo 3: Canciones de más de 6 a 10 minutos.

Prueba de Kruskal-Wallis

H₀: las medias de los 3 grupos son iguales

H₁: Las medias de los 3 grupos son distintas

$$H_0: X_1 = X_2 = X_3$$

$$H_1: X_1 \neq X_2 \neq X_3$$

P_valor < 0.05

Hipótesis Nula - H₀

1 - 3

3 - 6

6 - 10

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

00:00

0

Prueba de Mann-Withney

$H_0: X_1 = X_2$ $H_0: X_1 \geq X_2$

$H_1: X_1 \neq X_2$  $H_1: \mu_1 < X_2$

Los grupos 1 y 2 son diferentes

El grupo 1 es superior a 2

Los grupos 1 y 3 son diferentes

El grupo 1 es superior a 3

Los grupos 2 y 3 son diferentes

El grupo 2 es inferior a 3

Conclusión Hipótesis 2:

Las canciones con la duración entre 1 y 3 min son las más escuchadas.

Sin embargo, las canciones con más de 6 minutos tienen más escuchas que las que tienen entre 3 y 6 minutos

08/CONCLUSIONES

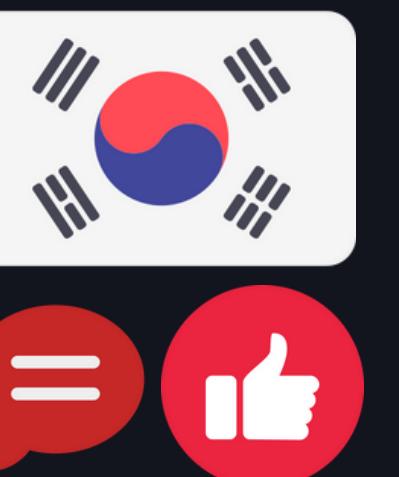
Las canciones populares tienden a ser más bailables, más alegres y menos enérgicas

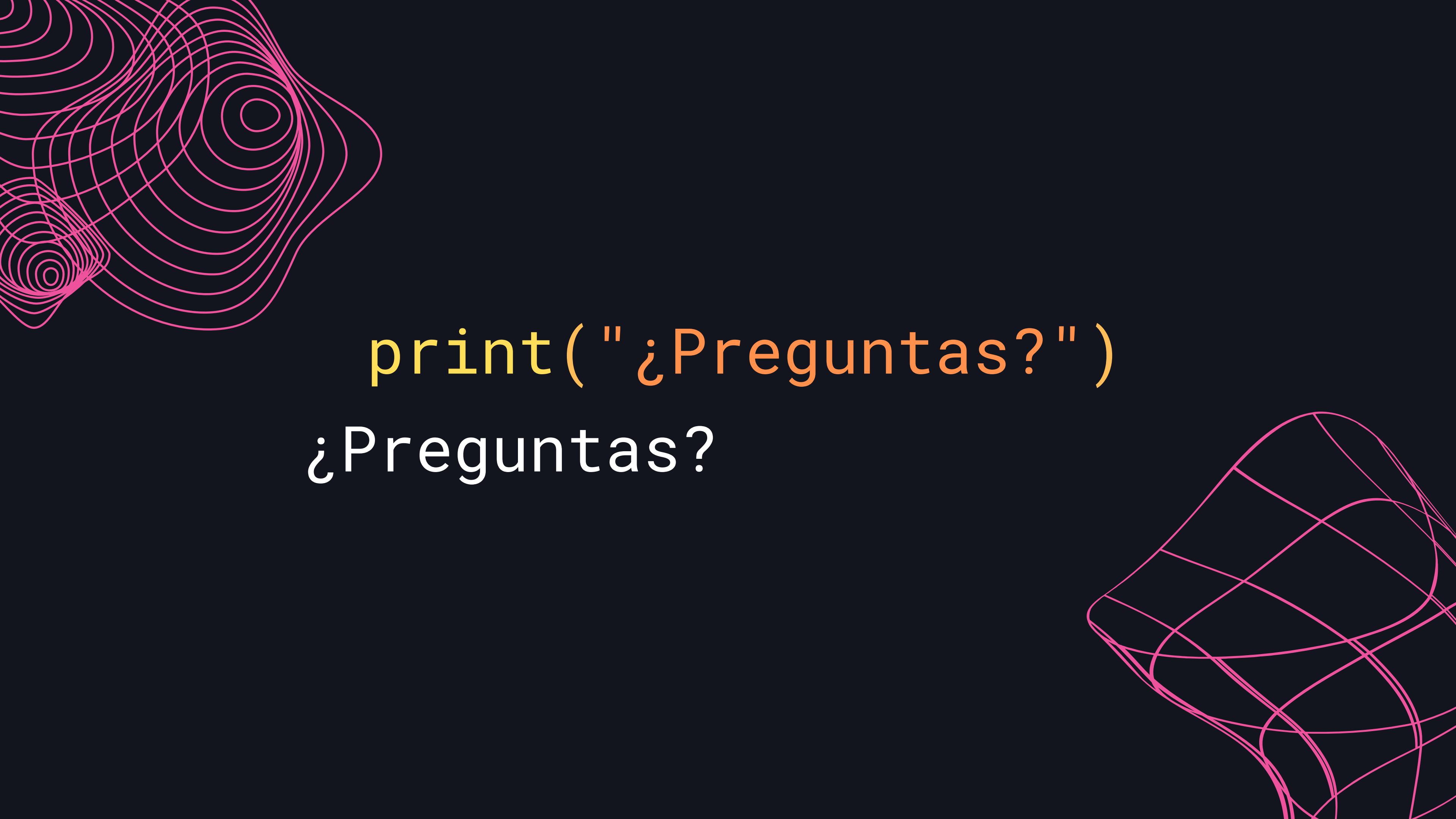


No garantiza que un artista o canción sea igualmente popular en ambas.



Las canciones con duración entre 1 y 3 min son las más escuchadas.





```
print("¿Preguntas?")
```

¿Preguntas?

Gracias.

Sebastián Pomi

