



**Tecnológico
de Monterrey**

**Statistical and Probabilistic Analysis of Spotify
Popularity**

Alumnos:

Marco Miloslavich Airola - *A01424280*

Alejandro Miloslavich Airola - *A01424120*

Samuel López Araiza Cadena - *A01026507*

Andrés Pi González - *A01664155*

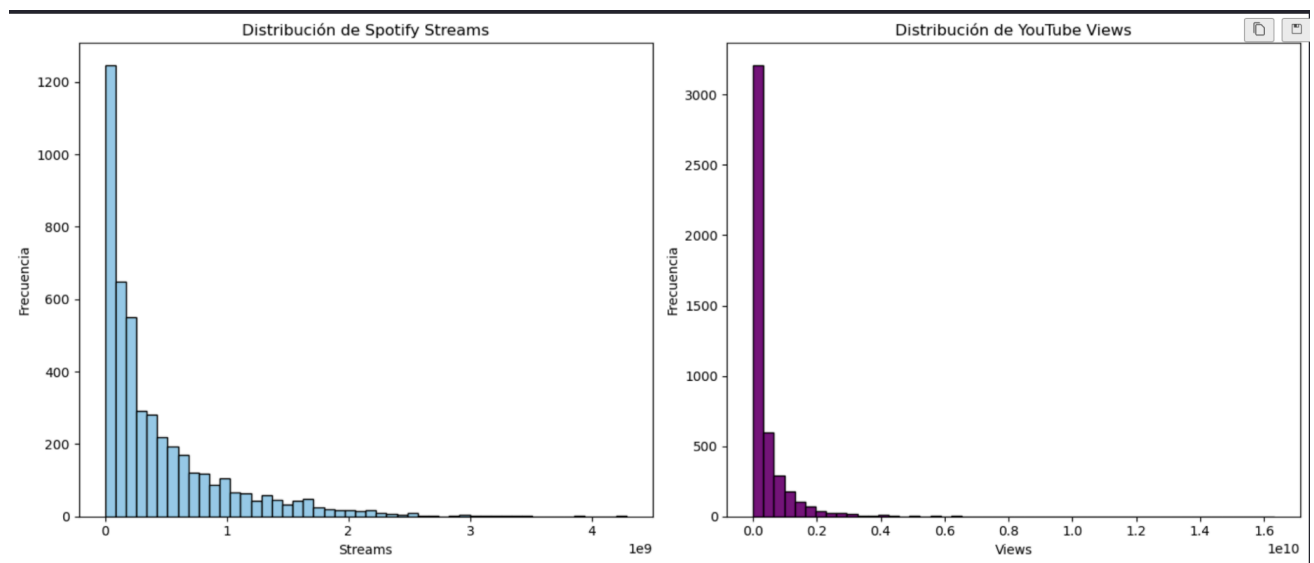
Luis Navarro Vivas - *A01663776*

Andrew Steven Williams Ponce - *A01662814*

Bruno Manuel Zamora García - *A01798275*

Fecha: 10 de Septiembre, 2025

Para la primera parte de la visualización de la distribución de la información, tenemos los histogramas de streams de Spotify y de Youtube, donde podemos apreciar una larga cola o “long tail” de datos en ambos histogramas (img. 1.0)



Img. 1.0

Donde nosotros podemos concluir que esto hace referencia a que, tanto dentro del mundo de Spotify como en YouTube, la industria musical está dominada por unos pocos éxitos globales, mientras que la mayoría de canciones acumulan números relativamente bajos.

Este fenómeno nos demuestra cómo el éxito no se reparte de manera uniforme, sino que está concentrado en unos pocos “hits” que dominan el consumo global y dando a entender mediante la columna más larga, que existen miles de canciones que, aunque no llegan a ser masivas, siguen formando parte del ecosistema musical y contribuyen a la diversidad de la oferta. Es decir, nos indica la fuerte desigualdad existente en la distribución de la atención, donde son pocos los temas que capturan la mayor parte del mercado, dejando a la mayoría en niveles modestos de popularidad.

Ahora bien, al vincular este análisis con las medidas de tendencia central, se observa que la media se ve fuertemente influenciada por estos valores extremos, lo que provoca que se desplace hacia la derecha de la distribución. Esto ocurre porque unos pocos éxitos masivos, con cifras extraordinarias de reproducciones y vistas, elevan el promedio de manera considerable. En contraste, la mediana permanece mucho más cercana al centro de la concentración principal de datos, representando mejor a la mayoría de las canciones con niveles bajos o moderados de popularidad.

Ahora bien, si extendemos el análisis hacia la variabilidad de los datos mediante el coeficiente de variación (CV), es posible comparar de manera relativa la dispersión

existente entre las distintas columnas analizadas. El CV, tal y como se sabe que se obtiene, lo obtuvimos tras dividir la desviación estándar entre la media, permitiéndonos así dimensionar cuánto varían los datos en relación con su promedio.

Coeficiente de Variación	
Spotify Popularity	0.230274
Spotify Streams	1.204546
YouTube Views	1.765289

Al calcular este indicador para las tres columnas principales, podemos ordenarlas de la más consistente a la más volátil. Generalmente, una columna como **Spotify Popularity** tiende a ser más consistente, ya que sus valores están acotados y muestran menor dispersión relativa. Cosa contraria ocurre con la de **YouTube Views**, la cual suele ser la más volátil debido a la enorme diferencia entre los videos con pocas vistas y los que alcanzan cifras millonarias o incluso billonarias.

En cuanto al contexto del éxito de una canción, esto significa que:

- **Popularidad en Spotify (más consistente):** refleja una métrica más estable y menos sensible a valores extremos, mostrando una tendencia uniforme en cómo la plataforma califica los temas.
- **Streams en Spotify (intermedio):** indican cierta diversidad en el consumo, pero con menos variaciones extremas que en YouTube.
- **Vistas en YouTube (más volátil):** evidencian un entorno mucho más desigual, donde unos pocos videos concentran gran parte de la atención, generando una variabilidad enorme frente a la mayoría de canciones que reciben cifras mucho menores.

Ahora bien, para poder profundizar un poco más en nuestro análisis, decidimos aplicar el Teorema de Bayes, el cual se formula como:

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)}$$

Donde:

- **A:** evento de que una canción tenga “Alta Popularidad”
- **B:** evento de que una canción sea “explícita”
- **P(B):** la probabilidad previa de que una canción sea explícita
- **P(A|B):** la probabilidad de que una canción tenga alta popularidad dado que es explícita
- **P(A):** La probabilidad total de que una canción tenga alta popularidad

Y, para nuestro caso, los resultados muestran que, en comparación con la probabilidad previa **P(B)**, la posterior dice ser menos probable, lo cual, en otras

palabras, nos dice que el saber que la canción es MUY popular reduce la probabilidad de que sea explícita.

Finalmente, podemos concluir que el análisis confirma cómo el éxito musical está fuertemente concentrado en unos pocos “hits” globales, mientras que la mayoría de las canciones se mantiene en niveles modestos de popularidad, tal como se refleja en la long tail observada en las distribuciones de Spotify y YouTube. Esta concentración también explica por qué la media se ve desplazada hacia la derecha por los valores extremos, mientras que la mediana permanece cercana al centro de la mayoría de los datos.

Asimismo, el coeficiente de variación nos permitió evidenciar que no todas las métricas presentan la misma estabilidad: la popularidad en Spotify resulta más consistente, los streams muestran una variabilidad intermedia, y las vistas en YouTube se caracterizan por una mayor volatilidad. Esto indica que cada plataforma refleja patrones distintos en la manera en que los usuarios consumen música.

A su vez, al aplicar el Teorema de Bayes, observamos que la probabilidad de que una canción sea explícita cambia al condicionarla con un nivel alto de popularidad. Este resultado demuestra que la popularidad no es independiente de las características del contenido, y que el análisis estadístico puede revelar relaciones más profundas entre los atributos de las canciones y su éxito.

En conjunto, estos hallazgos muestran que la industria musical no sólo es desigual en la distribución del éxito, sino también compleja en los factores que lo determinan, lo que exige un análisis cuidadoso de las métricas para interpretar correctamente el alcance real de una canción.