

## ENTREGA 1

2. La industria elegida es el entretenimiento, en la que Spotify tiene una posición bastante importante en el streaming de música. La elección de este sector se basa en que muestra cómo la digitalización ha cambiado las tendencias de consumo cultural al posibilitar que los usuarios tengan acceso instantáneo a música y grabaciones de audio en general. El entretenimiento digital, por otra parte, es un área con alto crecimiento e impacto a nivel mundial; por lo tanto, se vuelve un sector importante para llevar a cabo análisis de tendencias y datos.

3. Spotify es una compañía sueca de carácter multinacional que trabaja en el ámbito del entretenimiento digital, centrada en la distribución de música, podcasts y más contenidos de audio por medio de internet. Ofrecen a los usuarios acceso bajo demanda a un amplio catálogo de música, utilizando un modelo con una versión premium que requiere la suscripción mensual y otra gratuita con anuncios. La elaboración de listas de reproducción automatizadas o a la medida, la posibilidad de encontrar tendencias locales o globales y la personalización de sugerencias musicales fundamentadas en algoritmos son algunos de sus servicios más destacados. Spotify, por otra parte, representa un actor esencial en la industria de la música, no solo porque conecta a una gran cantidad de oyentes con artistas globales, sino también porque ha cambiado el modelo comercial musical al ofrecer nuevas vías de ingresos a los artistas.

5.

Atributo (Variable )	Tipo de Dato	Direcc ión de orden	Descripción
user_id	Cuantitativa	Secuencial	Identificador único para cada usuario.
gender	Categórica	-	Género del usuario (Male, Female, Other).
age	Cuantitativa	Divergen	Edad del usuario en años.
country	Categórica	-	País de residencia del usuario.
subscription_type	Categórica	-	Tipo de plan de Spotify (Free, Premium, Family, Student).
listening_time	Cuantitativa	Divergen	Tiempo total de escucha en horas por semana.
songs_played_per_day	Cuantitativa	Divergen	Promedio de canciones reproducidas diariamente.

skip_rate	Cuantitativa	Divergen	Tasa de omisión de canciones (proporción de canciones saltadas).
device_type	Categórica	-	Dispositivo principal de escucha (Desktop, Mobile, Web).
ads_listened_per_week	Cuantitativa	Divergen	Número de anuncios escuchados por semana.
offline_listening	Categórica (Binaria)	-	Indica si el usuario utiliza la función de escucha sin conexión.
is_churned	Categórica (Binaria)	-	Variable objetivo. Indica si el usuario canceló la suscripción.

6. Diagnosticar y priorizar los factores asociados al *churn* (variable *is\_churned*) en usuarios de Spotify, segmentando por tipo de suscripción (*subscription\_type*), país (*country*), dispositivo (*device\_type*) y rasgos de uso/engagement (*listening\_time*, *songs\_played\_per\_day*, *skip\_rate*, *ads\_listened\_per\_week*, *offline\_listening*), para identificar segmentos de alto riesgo y definir palancas de intervención que reduzcan la tasa de *churn* en la muestra analizada.

7.

1. Resumir la tasa de churn y su distribución por plan, país, dispositivo, rangos de edad y género; mostrar porcentajes, conteos y un ranking de severidad.
2. Comparar retenidos vs. churners en *listening\_time*, *songs\_played\_per\_day*, *skip\_rate*, *ads\_listened\_per\_week* y uso offline por plan y país, para detectar brechas accionables.
3. Derivar umbrales operativos de riesgo (p. ej., *skip\_rate* > P75 y *listening\_time* < P25) y una regla de alerta clara por plan/país.
4. Localizar segmentos de alto riesgo donde la tasa de churn supere un umbral (ej. 30%) en el cubo plan×país×dispositivo×edad, y priorizarlos por tamaño × tasa.
5. Descubrir outliers y patrones raros (picos por país/dispositivo, combinaciones poco obvias) que expliquen rupturas o casos extremos.

8.

Núm. tarea	Indicador/métrica	Formula	Variables	Segmentación
Tarea 1	Churn por segmento	$churn\_rate\_s = (\# \text{ de usuarios con } is\_churned=1 \text{ en } s) / (N\_s)$	<i>is_churned</i> + segmentador	por <i>subscription_type</i> , <i>country</i> , <i>device_type</i> , <i>age_bin</i> , <i>gender</i> . Ranking Top-N por severidad.

Tarea 2	Brecha de listening_time	$\Delta LT = \text{mean}(\text{listening\_time})$	retenidos – $\text{mean}(\text{listening\_time})$	is_churned=1
Tarea 3	Riesgo relativo	Definir alerta: (skip_rate > P75_plan,país) & (listening_time < P25_plan,país). Luego RR = P(churn	(alerta) / P(churn)	Grupo de referencia
Tarea 4	Prioridad del segmento	prioridad = N_s × churn_rate_s	Derivadas de T1	cubo subscription_type × country × device_type × age_bin. Ordena backlog por impacto potencial.
Tarea 5	% de outliers	%outliers = (# valores fuera de [Q1–1.5·IQR, Q3+1.5·IQR]) / N	Métricas continuas (LT, songs/day, skip_rate, ads/week)	por subscription_type, country, device_type, age_bin, gender. Para descartar casos extremos.