

Proyecto: EDA aplicado a Twitch

Durante la pandemia, Twitch experimentó un boom de creadores y espectadores, y mi mejor amigo fue uno de ellos. Empezó como un hobby, transmitiendo juegos y creando una pequeña comunidad que, con el tiempo, ha ido creciendo. Aunque aún no es famoso, sueña con vivir de esto y convertirlo en su carrera.

Siempre me ha interesado cómo funciona Twitch, y viendo su esfuerzo diario, decidí ayudarlo. Quiero analizar datos de la plataforma para identificar qué estrategias, comportamientos o patrones pueden hacer que un streamer destaque y crezca más rápido. Mi objetivo es responder a la pregunta: **"¿De qué forma sería más fácil que mi amigo se haga famoso y pueda vivir de Twitch?"** y ofrecerle recomendaciones claras basadas en evidencia.

Paso 0: Limpieza de los datos:

- Eliminar duplicados.
- Rellenar/descartar valores nulos.
- Transformar variables categóricas (si aplica).
- Convertir fechas a un formato estándar (si el dataset incluye fechas).

Paso 1: Entender el problema

En los últimos años, Twitch ha crecido exponencialmente como una de las plataformas más populares para creadores de contenido, pero también se ha vuelto altamente competitiva. Mi mejor amigo, un streamer apasionado que ha construido una comunidad pequeña pero fiel, aspira a convertir su hobby en una carrera profesional y alcanzar un nivel de fama que le permita vivir de ello. Sin embargo, destacar en Twitch requiere más que dedicación: es fundamental entender qué factores influyen en el crecimiento de un canal y cómo optimizarlos para atraer más espectadores, ganar seguidores y aumentar la interacción en la plataforma.

Problema central:

¿Cómo puede un streamer, especialmente en sus primeros años, identificar y priorizar las estrategias más efectivas para maximizar su crecimiento y escalar más rápido en Twitch?

La solución a este problema permitirá a mi amigo, y potencialmente a otros streamers emergentes, desarrollar un enfoque basado en datos para alcanzar el éxito en un entorno tan competitivo.

Preguntas secundarias:

¿Cuánto tiempo deberían dedicar al streaming por semana para maximizar su alcance?

¿Qué juegos o categorías son más efectivos para atraer viewers y seguidores?

¿Qué idioma es mejor para captar audiencias específicas?

Factores clave de crecimiento:

- ¿Qué juegos generan más viewers para streamers emergentes?
- ¿Cuánto influye el idioma del stream en la cantidad de viewers promedio?
- ¿Cómo impacta la duración promedio de los streams en el crecimiento del canal?

- ¿El número de días activos por semana está relacionado con un mayor incremento de seguidores?

Patrones de comportamiento exitosos:

- ¿Cuáles son los días de la semana más efectivos para ganar seguidores?
- ¿Transmitir más de un juego por stream (variedad) favorece el crecimiento o es mejor especializarse en un solo juego?
- ¿Qué tipo de contenido (gameplay, charlas, colaboraciones) tiene más impacto en la audiencia inicial?

Impacto del engagement:

- ¿Cuál es la relación entre la interacción en vivo (chat, respuestas, actividades) y el número de viewers promedio?
- ¿El uso de redes sociales externas afecta positivamente el crecimiento en Twitch?

Estrategias de optimización:

- ¿Es más efectivo aumentar el tiempo total transmitido o la calidad/variedad de los streams?
- ¿Cuáles son los juegos más efectivos para atraer seguidores en el idioma principal del streamer?
- ¿Qué prácticas de streamers famosos podrían replicarse en canales pequeños para acelerar su crecimiento?

Paso 2: Descripción del set de datos

- Descripción y definición de las variables del set de datos.

Paso 3: Lectura del dataset

- Carga y lectura de la base de datos a utilizar.

Paso 4: Análisis Exploratorio

4.1 Análisis de cada variable de manera individual

1. Entender las variables numéricas:

- o Distribución de Viewers, Followers o Duración.
- o Detectar outliers o datos atípicos.
- o Visualización sugerida: histogramas y boxplots.

2. Entender las variables categóricas:

- Frecuencia de categorías (Juegos, Idioma, Días).
- Visualización sugerida: gráficos de barras.

4.2 Análisis univariado

1. Variables categóricas:

- o Comparar promedios de la variable objetivo por categorías.
- o Visualización sugerida: gráficos de barras agrupados.

2. Variables numéricas:

- Relación entre las variables numéricas.
- Visualización sugerida: diagramas de dispersión.

4.3 Análisis bivariado (Interacción entre predictores)

Investigar cómo pares de variables interactúan entre sí para influir en la variable objetivo.

Paso 5: Conclusiones preliminares

Resumir patrones detectados.

Recomendaciones para streamers emergentes o que ya tienen una comunidad.

Ideas para enriquecer el proyecto u otros tópicos a analizar.