

# MEMORIA PROYECTO REALIZADO POR ALVARO JAÉN:

## Visualizaciones de los canales de videojuegos en la plataforma Twitch

### INTRODUCCION:

El principal motivo de esta elección como tema de mi EDA es debido al aumento en el interés, seguimiento y uso de la plataforma Twitch desde el inicio de la pandemia por COVID-19.

Desde los 14 años llevo jugando a videojuegos con mis amigos, después de 20 años habiendo pasado por los juegos de consolas: Nintendo, Super Nintendo o Nintendo 64, como juegos de ordenador desde un jugador, multijugador o juegos online masivos, ha habido un cambio en el concepto del seguimiento de videojuegos a nivel mundial.

Hace 10 años había eventos en las ciudades dónde se jugaban los torneos e iban los jugadores más “Top” a nivel mundial que solo lo conocías si eran muy fan de esos juegos.

Desde hace pocos años este “seguimiento” lo tienes a un “click” desde tu casa. Con una variedad de juegos infinita puedes ver desde canales oficiales con presentadores a personas jugando desde su casa, siendo el mejor del mundo o tu vecino de enfrente.

El aumento de esta plataforma ha sido tan exitosa que ya es conocida mundialmente en personas de todas las edades, ya que es el “sustituto” de la televisión junto a otras plataformas de series y películas.

Dado que el tema a elegir era abierto, he decido hacerlo por primera vez en mi vida sobre mi hobby con el que disfruto diariamente.

### HIPÓTESIS

La pregunta principal que me llevo haciendo hace un par de años es si el concepto de las visualizaciones a mundiales de juegos en Twitch puedes ser comparable con eventos deportivos como finales de champions o finales de mundiales de futbol. La lógica me dice que no pero vamos a ver que resultado obtenemos.

Durante el comienzo de este EDA y antes de sacar los datos, se me han ocurrido otras dos preguntas acerca de esta plataforma como:

- Cuáles han sido los juegos más seguidos cada año.
- Las visualizaciones, han aumentado en el tiempo o no.

Por último he sacado “hipótesis” terciarias tras obtener los datos como:

- Qué tipo de juegos son los más seguidos.
- Hay variación de visualizaciones según el mes del año

## PREPARACIÓN DATOS:

Para obtener los datos comencé por dos vías:

1. La primera era obtener los datos directamente desde Twitch a través de su API. Viendo la dificultad y la cantidad de horas dedicado a ello sin avanzar, investigué la segunda opción siendo la elegida finalmente.
2. Base de datos de Kaggle. Encontré varias bases de datos, que tras revisarlas enteras, vi que eran todas incompletas o no las entendía para poder analizarlas exceptuando una base de datos de 17.000 filas y con unos datos muy limpios y claros.

Esta base de datos se trata de la recopilación de datos desde el año 2016 hasta principio de 2023 de los 200 primeros canales que hay en Twitch de videojuegos seleccionados cada mes y cada año. Filtré los datos que me interesaban en función de mis hipótesis con los conocimientos aprendidos en clase hasta el día de hoy.

## REPARTO DE TIEMPO DEDICADO AL PROYECTO

Aunque haya dedicado muchas horas (entre 6-8 horas) al intentar entender la API de Twitch o buscar bases de datos, el resto del trabajo se ha basado en dos bloques:

- A. Tratamiento de datos: Por suerte el CSV tenía unos datos bastante limpio, por ello el orden del tratamiento ha sido:
  - a. Existencia de símbolos junto a los datos que impedían su lectura. Por ello he hecho varios replaces para quitar → `"",%,(),[],{} ,etc.`
  - b. Eliminar las columnas que no me interesaban.
  - c. Al estar los datos por los 200 juegos más vistos al mes, ordenados por meses 1-12 y por años 2016-2023, tenía que agrupar por año/meses según me interesaba para la hipótesis .
- B. Streamlit: Personalmente me llama la atención y me gusta el tema de "Visualizaciones". Por ello he dedicado casi la mitad del tiempo a hacer una buena presentación en Streamlit. Me he apoyado un poco "Estructuralmente" en el EDA de Alejandro en las partes que me gustaban y mirando principalmente en la documentación oficial de Streamlit y Lottie para Gift e iconos.

## **BIBLIOTECAS UTILIZADAS**

- Pandas
- Numpy
- Streamlit
- Streamlit Lottie
- Json
- PIL
- Plotly.express
- Plotly.graph
- Plotly.offline
- Wordcloud