



**Actividad:**

**Proyecto Final  
Ventas Globales de Videojuegos de 1980 a 2016**

**Asignatura:**

**Ciencia de Datos**

**Docente:**

**Dr. Farid Krayem Pineda**

**Alumna:**

**Karen Alexxa Hernández Preza**

## **I. Contexto del Problema**

La industria de videojuegos ha evolucionado drásticamente desde 1980, generando millones de dólares al año y siendo altamente competitiva. Cada año se lanzan cientos de títulos, pero solo una fracción logra ventas significativas, mientras que el 75.7% “fracasan” comercialmente (en este análisis se considera fracaso cuando no se supera el medio millón de títulos vendidos).

Encontrar qué factores ayudan al éxito comercial de un videojuego es clave para estudios, distribuidoras y equipos de marketing al momento de tomar decisiones de desarrollo y lanzamiento.

En este proyecto, desarrollé un dashboard interactivo con Streamlit y un modelo de Machine Learning para clasificar el nivel de éxito esperado para nuevos juegos con base en características como género, plataforma, editorial y año de lanzamiento.

## **II. ¿Por qué este análisis?**

La industria de los videojuegos es muy interesante y me imagino que las grandes empresas de entretenimiento se preguntarán:

- ¿En qué género invertir?
- ¿Qué plataforma priorizar?
- ¿Qué factores predicen el éxito?

Ya que errar en estas decisiones puede significar pérdidas millonarias en desarrollo, marketing y distribución. Por ello, este análisis busca predecir el éxito comercial utilizando datos históricos y técnicas de Machine Learning. Estoy consiente de que esto es un prototipo y necesita afinación para mejorar la predicción, pero es una base para continuar la investigación.

## **III. Preguntas de Investigación**

Principalmente: ¿Qué características incrementan la probabilidad de que un juego sea un éxito y es posible predecir el desempeño de un nuevo título antes de su lanzamiento?

- ¿Qué géneros tienen mayor potencial de ventas?
- ¿Qué plataformas generan mayores ventas globales?
- ¿Qué editoriales dominan el mercado?
- ¿Cuáles son los juegos más exitosos históricamente?
- ¿Podemos crear un modelo ML que estime el éxito de un juego nuevo?

## **IV. Metodología**

### **1. Procesamiento de datos**

**Fuente:** Dataset público de ventas de videojuegos (1980-2016)

**Volumen:** 16,326 registros

**Variables:** Nombre, Plataforma, Año, Género, Editorial, Ventas por región (NA, EU, JP, Otros)

**Tratamiento:**

- Limpieza de nulos y estandarización de variables.
- Transformación de variables categóricas mediante Label Encoding.
- Creación de categorías de éxito basadas en ventas globales.
- Generación de dataset procesado para análisis.

## 2. Análisis Exploratorio

A través del dashboard se analizaron:

- Ventas por género
- Ventas por plataforma
- Distribución por editorial
- Tendencias por año
- Juegos más vendidos
- Correlación entre variables
- Dispersión Año–Ventas

## 3. Machine Learning – Clasificación (Random Forest)

**Objetivo:** Predecir categoría de éxito (Fracaso, Moderado, Éxito, Blockbuster).

**Categorización:**

- a. Fracaso: < 0.5M copias (75.7% de juegos)
- b. Moderado: 0.5M - 1M copias (11.7%)
- c. Éxito: 1M - 3M copias (9.8%)
- d. Blockbuster: > 3M copias (2.9%)

Modelo entrenado y guardado en modelo\_entrenado.pkl.

**Variables predictoras:** género, plataforma, editorial y año.

**Métricas principales:** Accuracy e Importancia de variables.

**Validación:** Train/test split (75/25) y Cross-validation.

**Resultado:** 68.18% de precisión.

## 4. Análisis de Clustering (K-Means)

**Variables:** Género, Plataforma, Ventas por región.

**K óptimo:** 5 clusters (método del codo).

**Resultado:** 5 perfiles diferenciados:

**Cluster 3 (Mega-Blockbusters):** 28 juegos, 25.39M promedio

- Como Wii Sports, Super Mario Bros., Mario Kart
- Características: franquicias establecidas y plataformas masivas (Wii, NES)

**Cluster 4 (Éxitos AAA):** 429 juegos, 5.34M promedio

- Como GTA San Andreas, GTA V, God of War
- Características: Género Acción. PS2 predominante y juegos con alta inversión.

**Clusters 1, 2, 5 (Mainstream/Nicho):** 15,869 juegos combinados

- Ventas promedio: 0.30M - 0.42M
- Diversificados en Sports, Action y géneros casuales.

## 5. Herramientas

**Streamlit** para dashboard interactivo

**Random Forest Classifier** (200 árboles, profundidad 20)

**K-Means Clustering** (5 clusters, 10 n\_init)

## V. Resultados

- Los géneros con mayores ventas globales destacan son: acción (género más lucrativo), disparos (alto promedio por título: 0.80M) y deportes.
- Consolas icónicas: PS2, Wii, PS3 y Xbox 360 concentran gran parte de ventas históricas.
- **La distribución refleja ciclos tecnológicos y preferencia del usuario.**
- El mercado tuvo su pico entre 2006–2011, coincidiendo con la era dorada de las consolas de séptima generación (La séptima generación de consolas de videojuegos, que comenzó en 2005, incluye la Xbox 360, PlayStation 3 y Wii).
- PS2: 1,233M (plataforma más exitosa históricamente)
- Después del 2011 hay una tendencia a la baja.
- El mercado está altamente concentrado: pocos juegos generan la mayoría de las ventas globales y **Editoriales** como Nintendo (Nintendo lidera con 1,784M ventas totales y 2.56M promedio), Electronic Arts y Activision registran mayor marketshare.
- El **género** es el factor más determinante en el éxito de un videojuego.
- Las **plataformas** con bases de usuarios grandes elevan las ventas por título.
- Las **editoriales** consolidadas logran mayor impacto comercial.
- Las ventas históricas por año permiten identificar picos de demanda y tendencias temporales.

### Modelo Predictivo (ML)

Variables más influyentes:

- ♦ **Género**
- ♦ **Plataforma**
- ♦ **Editorial**
- ♦ Año (creo que quitaré esta variable, ya que en mi gráfica de correlación no parece ser importante).

El sistema predice la probabilidad de éxito de un nuevo juego, con resultados visuales y gráficos.

## VI. Conclusiones

- ✓ Alinear nuevos proyectos a géneros de alta demanda aumenta la probabilidad de éxito.

- ✓ Priorizar el lanzamiento en plataformas dominantes (según año y tendencias actuales) incrementa el alcance.
- ✓ Asociarse con editoriales reconocidas mejora la visibilidad del juego.
- ✓ El dashboard permite identificar oportunidades de negocio y planear estrategias de marketing por segmento.
- ✓ El predictor ML es una herramienta clave para: estimar ventas tempranas, planear presupuestos de desarrollo y tomar decisiones estratégicas sobre plataformas y ventanas de lanzamiento.
- ✓ Si se quiere lanzar un juego de acción se debería buscar respaldo de alguna de las top 10 editoriales (Nintendo, EA, Activision)
- ✓ Como Norteamérica y Europa tienen el 76% del mercado, deberían priorizarse estos mercados.

**Limitaciones:**

1. Tenemos datos hasta 2016: No incluye era actual de Game Pass, F2P, móviles y la influencia de los streamers.
2. Solamente estamos contando las ventas físicas únicamente: Ignora descargas digitales (creciente a partir del 2010)
3. Variables faltantes: Calidad del juego, presupuesto de marketing, reviews.
4. Sesgo: Modelo identifica 68% de fracasos potenciales antes del lanzamiento.

streamlit run app3.py