

# Videogames genre dataset

## Minería de datos

Aingeru García Blas

## Índice

<b>0. Introducción</b>	<b>2</b>
<b>1. Introducción a RAWG</b>	<b>3</b>
<b>2. Descripción del Dataset</b>	<b>3</b>
<b>3. Variables del Dataset</b>	<b>3</b>
<b>4. Relevancia del Análisis</b>	<b>3</b>
<b>5. Recopilación de datos</b>	<b>4</b>
5.1. Extracción . . . . .	4
5.2. Procesamiento de Datos . . . . .	4
5.3. Etiquetando; determinación del éxito . . . . .	4
5.4. Almacenamiento de Datos . . . . .	4
<b>6. Recursos</b>	<b>4</b>
6.1. Repositorio del Proyecto . . . . .	4
6.2. Dataset . . . . .	4
6.3. Script de Análisis . . . . .	5
<b>7. Conclusión</b>	<b>5</b>

## 0. Introducción



Muchos videojuegos son verdaderas obras de arte, pero desgraciadamente hay un sesgo y prejuicios bastante marcados, y tienden a ser subestimados en nuestra sociedad. Sin embargo, gran parte de ellos disponen de un impresionante arte visual creado por diseñadores y artistas de gran talento. Qué decir de la música, se suelen acompañar de música compuesta por orquestas de grandísimo nivel y compositores famosos. Sus historias, a veces son concebidas por increíbles loremasters que ofrecen una experiencia narrativa excepcional creando universos que podrían compararse al de Tolkien. Todo esto se combina en un mundo donde el jugador puede interactuar a su antojo y tomar decisiones; los videojuegos pueden llegar a llevar el arte a un nivel superior de inmersión y creatividad.

Pero claro, existen tantos tipos de géneros de juegos diferentes... no es lo mismo un RPG (Role Playing Game) que un Pinball o un simple juego de golf (que, irónicamente, el programarlo, no es tan simple), no es lo mismo un Mario (buenísimo, pero sin historia realmente, más que la nostalgia que podamos tener acumulada), que Child of Light o The Witcher 3. No es lo mismo un 'mata-mata' que World of Warcraft o incluso que un Final Fantasy - cuyas complejas y embaucadoras historias dejan la mayoría de las películas o series como historias simples y 'sin chicha'. Con lo que, dado el impresionante interés que suscita esta industria, vamos a diseñar un pequeño dataset que nos permita analizar las preferencias de los usuarios desde los 'años retro', con respecto a los diferentes géneros que se han ido forjando con el tiempo.

## 1. Introducción a RAWG

RAWG es una de las bases de datos de videojuegos más grandes y completas disponibles. Siempre ha sido famosa por disponer de información muy extensa y dispar, incluyendo detalles de todo tipo, como géneros, plataformas, calificaciones, fechas de lanzamiento, información sobre desarrolladores, tipos de empresa (AAA, indies)...etc. Es una muy buena herramienta para analizar tendencias y patrones en esta industria. Se puede consultar la [extensa documentación de su API](#) para conocer las diferentes consultas que se pueden realizar

## 2. Descripción del Dataset

El dataset producido contiene datos extraídos de RAWG. Tiene información juegos lanzados desde 1980 hasta 2023, poniendo el foco en algo simple; aspectos como el género del juego, el año de lanzamiento, las calificaciones o ratings y la cantidad de juegos lanzados por año. Pero, pese a su simpleza, podemos extraer información bastante interesante para posteriormente poder analizar.

## 3. Variables del Dataset

El dataset AARFF (Attribute-Relation File Format) producido, contiene las siguientes variables:

---

**genre** (*STRING*): El género del videojuego. Representa la categoría o tipo del juego: Acción, RPG, Aventura, etc.

**year** (*NUMERIC*): El año de lanzamiento.

**average\_rating** (*NUMERIC*): La calificación o rating de ese género para ese año en concreto.

**game\_count** (*NUMERIC*): La cantidad de juegos del género en cuestión que fueron lanzados ese mismo año.

**genre\_successful** (*NUMERIC*): Etiqueta o clase, valor binario (1 o 0) que muestra si un género es considerado exitoso en un año basado en el umbral de calificación promedio.

## 4. Relevancia del Análisis

El análisis de estos datos puede mostrarnos tendencias en la popularidad de diferentes géneros de videojuegos, así como cambios en cómo los usuarios han recibido ciertos géneros a lo largo de los años. Puede parecer algo suérfuo, pero este tipo de información es realmente útil, nos ayuda a comprender la dinámica del mercado e incluso ver ciertos patrones; de hecho tanto desarrolladores indie como empresas más grandes hacen estudios de mercado potentes donde incluyen éste tipo de análisis (y muchísimo más complejos, claro está).

## 5. Recopilación de datos

### 5.1. Extracción

El proceso comienza con la extracción de datos desde la API de RAWG - se necesita una [API key](#) de developer que se puede conseguir simplemente registrándose y solicitándolo. Se utilizan solicitudes HTTP, se recopilan datos como el nombre del juego, género, fecha de lanzamiento y calificaciones de los usuarios. Estos datos son recogidos para juegos lanzados desde 1980 hasta 2023.

### 5.2. Procesamiento de Datos

Una vez recopilados, se procesan para su análisis. Esto incluye:

- Convertir la fecha de lanzamiento a un formato de fecha y hora para extraer el año.
- Transformar las calificaciones a valores numéricos para facilitar cálculos posteriores.
- Agrupar los datos por género y año, y calcular el rating medio y la cantidad de juegos lanzados cada año.

El script para realizar este proceso puede ser encontrado aquí: [game\\_data\\_analyzer.py](#)

### 5.3. Etiquetando; determinación del éxito

Para determinar si un género de juego ha sido exitoso en un año específico, se establece un umbral de éxito. En este caso, se considera que un género es exitoso si la calificación promedio de sus juegos supera el **3.2**. Así etiquetamos cada instancia (género por año).

### 5.4. Almacenamiento de Datos

Una vez procesados, los datos se almacenan en un archivo en formato **ARFF**, que incluye la definición de los atributos (género, año, rating promedio, cantidad de juegos, éxito del género) y los datos organizados (**462 instancias**)

## 6. Recursos

### 6.1. Repositorio del Proyecto

El repositorio completo del proyecto está disponible en GitHub. Aquí se puede encontrar toda la información relacionada con el análisis, incluyendo documentación ([README.md](#)), scripts y datos procesados.

- URL del Repositorio: [github.com/geru-scotland/videogames-dataset-analysis](https://github.com/geru-scotland/videogames-dataset-analysis)

### 6.2. Dataset

El dataset procesado en formato ARFF, listo para ser utilizado en herramientas de machine learning, se encuentra en el siguiente link.

- URL del Dataset: [github.com/geru-scotland/videogames-dataset-analysis/data/arff/video\\_games\\_data.arff](https://github.com/geru-scotland/videogames-dataset-analysis/data/arff/video_games_data.arff)

### 6.3. Script de Análisis

Y aquí el script principal desarrollado para la obtención, procesamiento y almacenamiento:

- URL del Script: [github.com/geru-scotland/videogames-dataset-analysis/game\\_data\\_analyzer.py](https://github.com/geru-scotland/videogames-dataset-analysis/game_data_analyzer.py)

## 7. Conclusión

Es un dataset simple al que se le ha aplicado un proceso y visión muy simplistas, por 'jugar' un poco y disfrutar de trabajar en ello. No obstante, se puede llegar a sacar detalles y patrones jugosos si se analizan bien los datos.

Esta industria, es una industria demoledora, así que, futuros desarrolladores de juegos, ¡hacedlo por amor y no por dinero!

«No somos fabricantes de juegos, somos fabricantes de mundos, creadores de historias y forjadores de sueños.»

–Hironobu Sakaguchi (Creador de la saga Final Fantasy)

31 de Diciembre de 2023