

# Práctica 1: Web scraping

Tipología y ciclo de vida de los datos



Alumnos: Amanda Iglesias Moreno / Fernando García Dorador

## Contexto

Desde hace ya unos años, la venta de videojuegos de forma digital se ha convertido en el estándar de la industria, incluso en plataformas como PlayStation y Xbox donde tradicionalmente se vendían físicamente. En este contexto, surgieron páginas web que ofrecen copias digitales a precio reducido de juegos en diferentes plataformas, que se activan mediante un código. Instant-Gaming es una de las páginas populares que ofrece estos servicios.

## Definir un título para el dataset

Videogame data from Instant-Gaming

## Descripción del dataset

El conjunto de datos contiene un listado de las características principales de los juegos disponibles en la página web de videojuegos online instant gaming. Entre las características disponibles en el dataset encontramos variables relativas al precio y tipo de juego así como referentes a la valoración por parte de los usuarios.

## Representación gráfica



Imagen de Florian Olivo en Unsplash <https://unsplash.com/photos/Mf23RF8xArY> (Nota: Unsplash permite el uso libre de las fotos de la plataforma sin una autorización previa del autor)

## Contenido

Los campos que incluye el dataset son los siguientes:

- **Id:** Identificador del videojuego, campo generado manualmente.
- **Title:** Título del videojuego.
- **Price\_ig:** Precio del videojuego en *Instant-Gaming* una vez aplicado el descuento.
- **Discount:** Descuento ofrecido por *Instant-Gaming*.
- **Price\_retail:** Precio del videojuego sin el descuento.
- **Price\_steam:** Precio del videojuego en la plataforma Steam.
- **Region:** Región en la que se puede activar el juego.
- **DLC:** Variable binaria que indica si es downloadable content, es decir, si requiere de un juego base, o no.
- **Platform:** Plataforma en la que se puede jugar el videojuego.
- **Link:** link de compra del videojuego.
- **Availability:** Indica si el juego está disponible o no.
- **Genre:** Géneros del videojuego.
- **Languages:** Idiomas disponibles del videojuego.
- **Release\_date:** Fecha en la que el juego ha sido publicado.
- **User\_rating:** Puntuación otorgada por los compradores de *Instant-Gaming*.
- **Number\_coments:** Cantidad de comentarios sobre el videojuego.
- **Review:** Comentario con más votos de la página.
- **Rating\_review:** Puntuación del videojuego que otorga la persona con el comentario más popular.

Los datos obtenidos corresponden a la fecha de raspado de estos. Los datos más cambiantes serían los precios y los comentarios, que pueden variar ligeramente con carácter diario/semanal.

Todos los datos han sido obtenidos utilizando BeautifulSoup e identificando la clase a la que pertenecen, a excepción de *Price\_steam*, en la que se ha utilizado Selenium, obteniendo el precio que aparece en la caja de sugerencias.

## Agradecimientos

Los datos han sido recolectados de las siguientes páginas web (información obtenida mediante la librería whois de Python):

- <https://www.instant-gaming.com>, creada el 9 de diciembre de 2005 y con fecha de expiración el 2 de diciembre de 2020, registrada en el servidor *godaddy.com*, con sede en Arizona, Estados Unidos.
- <https://store.steampowered.com>, creada el 9 de agosto del 2000 y con fecha de expiración el 9 de agosto de 2022, registrada en el servidor *networksolutions.com*, propiedad de Valve Corporation, con sede en Washington, Estados Unidos.

## Inspiración

La reciente **pandemia del covid** ha reestructurado nuestros hábitos de vida y la forma en la que empleamos nuestro tiempo libre. Las limitaciones sociales fruto del confinamiento nos han obligado forzosamente a buscar nuevas alternativas de diversión o a recuperar viejas

aficiones ya olvidadas, obligandonos ha adentrarnos en muchos casos en el **mundo del entretenimiento digital**.

En este contexto, una industria que ha crecido notablemente es el **mercado de los videojuegos online**. Según indica Acuity Ads, el consumo de videojuegos en línea ha aumentado un 75 por ciento en España. Por este motivo, ante el creciente interés del sector, hemos pensado que sería realmente interesante proporcionar al usuario un conjunto de datos que contenga información sobre los juegos disponibles en una de las páginas web líderes en la venta de juegos online.

La página **instant gaming** ofrece múltiples características sobre los juegos disponibles en si misma, lo que permitirá posteriormente obtener análisis detallados que pretendan no sólo responder a la **estrategia de precios de la compañía** sino también a los **gustos de los consumidores** y a la **calidad de los juegos**.

Algunas de las preguntas que se podrían responder tras el análisis de los datos recolectados son las siguientes:

- ¿Cuáles son los descuentos más comunes en el sitio web instant gaming?
- ¿En qué idiomas están disponibles los juegos?
- ¿Cuáles son los juegos mejor valorados?
- ¿Existe una relación entre la valoración de los juegos por parte de los usuarios en la página web y los descuentos que esta aplica?
- ¿Cuáles son los juegos que más debate suscitan (mayor número de comentarios)?
- ¿Qué género está mejor valorado entre los usuarios de la página web?
- ¿Son los juegos más antiguos (los que llevan más tiempo en el mercado) más económicos?
- ¿Se ha incrementado la calidad de los juegos con el tiempo o por el contrario los juegos antiguos están mejor valorados por los usuarios?

Por otra parte, la recolección de los datos durante un periodo de tiempo (p. ej. diariamente durante un año) permitiría incluir la variable tiempo en el análisis, ampliándose de esta manera el número de preguntas que se podrían responder a partir del conjunto de datos, por ejemplo, se podría analizar si la página web ofrece mayores descuentos en determinadas épocas del año o cuántos juegos se han incorporado y eliminado de la página web en dicho periodo de tiempo.

## Licencia

Los datos se han publicado bajo la licencia **Creative Commons BY-NC-SA 4.0**. Esta licencia **permite**:

- **compartir** - copiar y redistribuir el material en cualquier medio o formato.
- **adaptar** - mezclar, transformar y crear a partir del material.

Sin embargo, no autoriza emplear el material con **objetivos comerciales**. Además, **garantiza la atribución y la divulgación bajo la misma licencia**.

Esta se trata de una de las licencias Creative Commons más restrictiva y la hemos escogido para garantizar en todo momento el uso del material con una **finalidad no comercial**, asegurando a instant gaming que ningún individuo u organización pueda utilizar los datos con fines que atenten contra sus intereses comerciales. Por otra parte, bajo esta licencia, se debe reconocer la **autoría original del material**, así como garantizar **las obras derivadas se distribuyen utilizando la misma licencia**, es decir, con fines no comerciales.



*Icono de la licencia*

Contribuciones	Firma
Investigación previa	FGD, AIM
Redacción de las respuestas	FGD, AIM
Desarrollo código	FGD, AIM