

Tipología y ciclo de vida de los datos

Web Scraping: Instant-Gaming

Sergio Morales Machado

Daniel Sohm

-Contexto:

Se ha recolectado información sobre los precios listados en una pagina web de venta de claves de videojuegos. Estas claves se adquieren directamente a los proveedores (origin, steam, epicgames, etc) siendo así totalmente legales, y se canjean en la respectiva plataforma para adquirir el juego. Esto permite un ahorro significativo respecto a las tiendas físicas o a través de distintos intermediarios y, en consecuencia, ofrecer precios ventajosos a sus clientes.

Este dataset se ha creado para estudiar en más profundidad la cantidad rebajada frente al precio normal.

La página web elegida ha sido: www.instant-gaming.com/es/

-Titulo:

games-2023-04-24_12-13-05.csv

-Descripción:

En el dataset solo aparecen los juegos ofrecidos en la plataforma steam, ya que, si no se generaría un dataset demasiado grande, aunque es sencillo escalarlo a más elementos.

Se ha recolectado información de más de 8000 juegos, y un total de 12 variables. 90 de ellas son de tipo carácter y 2 de ellas de tipo numérico y un boolean (True/False).

Entre otros, se han recogido los nombres de los videojuegos e información relevante sobre ellos: desarrollador, precio, descuento, si están en stock, etc.

Esto nos permitirá realizar un análisis sobre el estado actual del mercado de videojuegos. Si esta operación se repite, se puede realizar un seguimiento temporal del mercado de videojuegos, y observar su evolución.

-Representación gráfica:

IndieGAMING

TendenciasReservasPróximos lanzamientosSoporte 24/7

PCPlaystationXboxNintendo

Minecraft Legends

-20%28.69€

Tendencias

Ver todo

RAVENSWATCH

-31%

Ravenwatch

13.89€

TOTAL WAR: WARHAMMER III

-24%

Total War: Warhammer III - Forge of th...

18.99€

DREDGE

-24%

Dredge

19.09€

<https://www.indie-gaming.com/new/12143-comprar-minecraft-legends-pc-juegos-microsoft-store-es-es/>

-Contenido:

“Title”: Nombre del videojuego

“Price”: Precio del videojuego

“Discount”: Descuento frente al mercado físico.

“Developer”: Estudio que ha desarrollado el juego.

“Publisher”: Quien es el publicador del videojuego.

“tags”: Atributos del juego, como, por ejemplo, de que genero es, si es multijugador, etc.

“release_date”: Fecha de lanzamiento

“valoration”: valoración media del videojuego

“reviews”: número de reseñas que tiene el videojuego

“stock”: Indica si hay disponibilidad/stock

“download”: Tipo de descarga.

“f2p”: Indica si el juego es gratis

Periodo temporal: 2023

-Propietario:

El propietario de los datos es www.instant-gaming.com/es/. Su sede se encuentra en Francia, Paris.

Se han realizado estudios previos similares, como por ejemplo un estudio de la empresa SimilarWeb, que proporciona información sobre el tráfico y las métricas de rendimiento de sitios web.

Otro estudio, de la empresa semRUSH, proporciona información sobre el tráfico orgánico, las palabras clave y la competencia de los sitios web.

Todos los datos que se han recogido están disponibles públicamente, sin necesidad de registrarse ni aceptar ningún tipo de condición.

-Inspiración:

Según la consultora Newzoo, el mercado global de videojuegos generó más de \$175.8 mil millones de dólares en ingresos en 2020, lo que representa un aumento del 20% respecto al año anterior. Tenemos un mercado muy grande y que está en continuo crecimiento (a pesar o gracias al covid-19).

Además, el mercado de videojuegos también ha experimentado un gran cambio en la forma en que se consumen y distribuyen los juegos. En lugar de comprar juegos físicos en tiendas, cada vez más jugadores compran y descargan juegos digitalmente a través de plataformas en línea como Steam, Epic Games Store y GOG.com, entre otras. Estas son las plataformas que vamos a analizar en nuestro estudio. Considerando el tamaño del mercado, cualquier conclusión puede ser valiosa.

A nivel personal, ambos integrantes del grupo estamos interesados en el mundo de los videojuegos, lo que añade un interés adicional en querer estar al tanto de la situación del mercado y su evolución.

-Licencia:

CC BY-NC-SA 4.0 License: Esta licencia permite que el dataset se utilice, copie, modifique y distribuya para cualquier propósito no comercial, siempre y cuando se atribuya al autor original y se mantenga la misma licencia en cualquier uso derivado. Esta licencia es adecuada si desea que el dataset sea compartido y utilizado para fines educativos (como es nuestro caso) o no comerciales, pero no desea que se utilice para fines comerciales.

-Código:

Uno de los problemas a los que nos enfrentamos fue que instant-gaming es una web muy usada, con mucha información respecto a los juegos. Esto ha hecho que nos encontremos con problemas de recolección de datos, pues había que hacer que la araña reconociera los distintos casos, que entre otros eran:

- Ciertas urls entraban en un bucle infinito de redirección, incluso configurando el user-agent de forma correcta.
- Juegos F2P, que por lo tanto nunca tienen stock
- Juegos que no están lanzados, pero se pueden pedir de forma anticipada.
- Configuración del user-agent para peticiones a la página y evitar bucles de redirección

En cuanto al primer punto, tuvimos que cambiar la librería que realizaba los requests, y pasamos de utilizar requests a urllib.request.

Para las diferentes casuísticas, realizamos correcciones en el código a medida que íbamos obteniendo los errores.

Para evitar problemas con las peticiones html relacionados al user agent, los definimos como una constante en un archivo aparte, y así tenerlos disponibles en todo momento.

Contribuciones	Firma
Investigación previa	SMM, DS
Redacción de respuestas	SMM, DS
Desarrollo del código	SMM, DS
Participación en el video	-