

Tipología y ciclo de vida de los datos
M2.851 - Aula 3

Práctica 1: Web Scraping

World Padel Tour

Rubén Gonzalo Coll Menéndez

Índice

1. Contexto.....	2
2. Definir un título para el dataset.....	2
3. Descripción del dataset.....	2
4. Representación gráfica	2
5. Contenido.....	3
6. Agradecimientos	3
7. Inspiración.....	3
8. Licencia.....	4
9. Código	4
10. Dataset	4

1. Contexto

Explicar en qué contexto se ha recolectado la información. Explique por qué el sitio web elegido proporciona dicha información.

2. Definir un título para el dataset

Elegir un título que sea descriptivo.

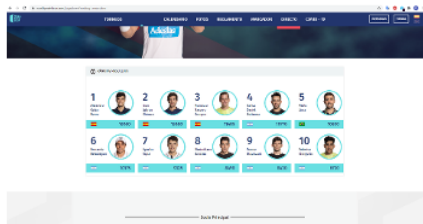
3. Descripción del dataset

Desarrollar una descripción breve del conjunto de datos que se ha extraído (es necesario que esta descripción tenga sentido con el título elegido).

4. Representación gráfica

Presentar esquema o diagrama que identifique el dataset visualmente y el proyecto elegido.

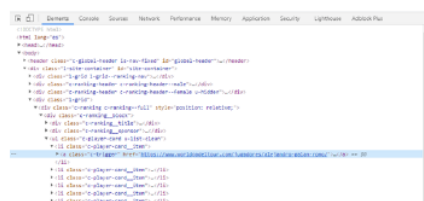
Paso 1: Abrir URL del Ranking de jugadores World Padel Tour



Paso 2: Simular el scrolldown para cargar todos los jugadores



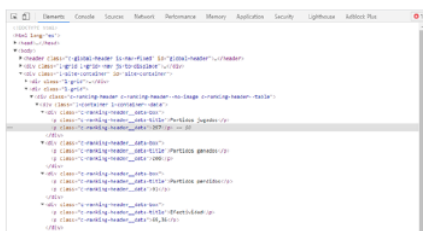
Paso 3: Descargar las URLs de la ficha de cada jugador



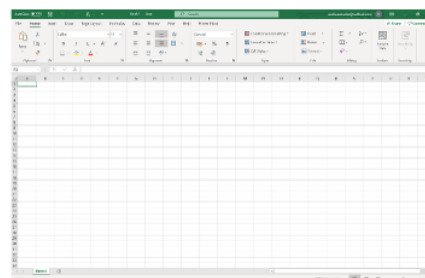
Paso 4: Acceder a cada una de las URLs de cada jugador



Paso 5: Descargar la información de cada una de los jugadores



Paso 6: Crear un fichero CSV con la información descargada



5. Contenido

Explicar los campos que incluye el dataset, el periodo de tiempo de los datos y cómo se ha recogido.

6. Agradecimientos

Presentar al propietario del conjunto de datos. Es necesario incluir citas de análisis anteriores o, en caso de no haberlas, justificar esta búsqueda con análisis similares.

7. Inspiración

Explique por qué es interesante este conjunto de datos y qué preguntas se pretenden responder. Es necesario comparar con los análisis anteriores presentados en el apartado 6.

8. Licencia

Seleccione una de estas licencias para su dataset y explique el motivo de su selección: ☐ Released Under CC0: Public Domain License ☐ Released Under CC BY-NC-SA 4.0 License ☐ Released Under CC BY-SA 4.0 License ☐ Database released under Open Database License, individual contents under Database Contents License ☐ Other (specified above) ☐ Unknown License

9. Código

Adjuntar el código con el que se ha generado el dataset, preferiblemente en Python o, alternativamente, en R.

10. Dataset

Publicación del dataset en formato CSV en Zenodo (obtención del DOI) con una breve descripción.

