

**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 .....	4
7. Inspiración.....	4
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.**

Aunque el Padel es un deporte en crecimiento y con un gran potencial, que se está expandiendo de manera muy rápida entre los diferentes países Europeos, aún no tiene la misma popularidad que otros deportes mayoritarios lo que provoca que no sea tan fácil la obtención de la información de los jugadores como lo es en otros casos como el fútbol.

La organización principal que está detrás del Padel mundial e impulsando dicho deporte es World Padel Tour; empresa privada que ha creado un circuito de Padel profesional y que a día de hoy es el más reconocido y que más fuerza está cogiendo pudiéndose comparar con lo que sería la primera división en el fútbol español.

Con el objetivo de ayudar a la difusión de este deporte me ha parecido interesante la creación de un dataset que permita a la gente crear sus propias visualizaciones y hacer sus propios análisis. Para ello he accedido a la página web del World Padel tour (WPT) para descargar la información relacionada con los jugadores y crear un dataset público.

Página web del World Padel Tour: <https://www.worldpadeltour.com/>

## 2. Definir un título para el dataset

**Elegir un título que sea descriptivo.**

El título elegido para el dataset creado es: WPT\_players\_info

## 3. Descripción del dataset

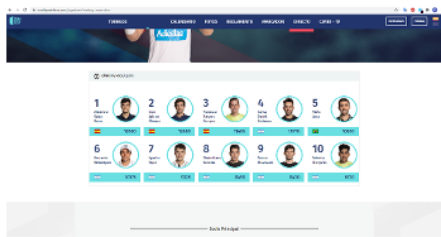
**Desarrollar una descripción breve del conjunto de datos que se ha extraído (es necesario que esta descripción tenga sentido con el título elegido).**

El dataset creado contiene información tanto demográfica como del rendimiento de los jugadores del ranking World Padel Tour (WPT).

## 4. Representación gráfica

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

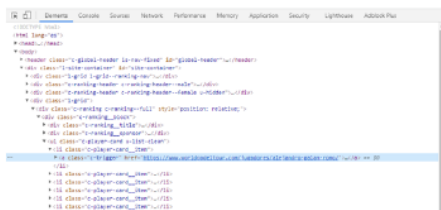
### Paso 1: Abrir URL del Ranking de jugadores World Padel Tour



### Paso 2: Simular el scroll down para cargar todos los jugadores



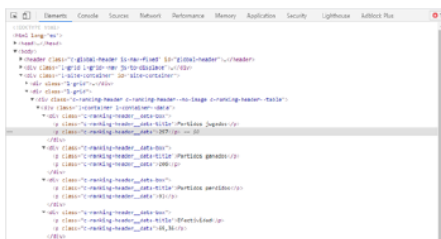
### Paso 3: Descargar las URLs de la ficha de cada jugador



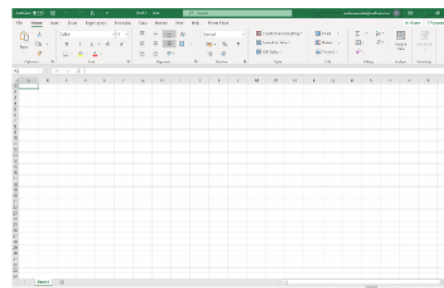
### Paso 4: Acceder a cada una de las URLs de cada jugador



### 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.**

El dataset creado contiene la fecha en la que se ha extraído la información de manera que podamos ir guardando fotos del mismo a lo largo del tiempo.

Los campos seleccionados para la creación de la primera version del dataset han sido:

- Nombre: Contiene el nombre completo del jugador
- Origen:
- País de residencia:
- Altura:
- Ranking actual: Contiene la posición del ranking WPT a fecha de extracción de la información
- Partidos jugados
- Partidos ganados
- Victorias consecutivas

- Efectividad: Cálculo del número de partidos ganados entre el número total de partidos jugados
- URL: enlace a la ficha del jugador de la cual se ha extraído la información

## 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.**

Este dataset es solo un primer paso para la ampliación del mismo con información más técnica de cada uno de los jugadores que permita un análisis en profundidad de su rendimiento.

Aunque ahora solo incorpora datos como el total de partidos ganados o la mejor racha de partidos sin perder, será interesante la inclusión de datos como tipo de golpes, efectividad de cada uno de ellos, etc.

## 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.**

El código del proyecto se puede ver en el siguiente enlace: (Añadir enlace de github)

## 10. Dataset

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

El data set se puede consultar en Zenodo siguiendo el siguiente enlace: (Añadir enlace de Zenodo)

<https://zenodo.org/>

