

El objetivo de esta actividad será la creación de un dataset a partir de los datos contenidos en una web. Para su realización, se deben cumplir los siguientes puntos:

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

Se ha decidido recolectar información referente a todas las especies de pokémon para poder realizar un estudio de habilidades, atributos y clases de estos pequeños monstruos. El sitio web seleccionado es:

<https://pokemondb.net/pokedex/all>

El sitio web, tiene una sección pokedex. La pokedex en el mundo Pokémon, es una enciclopedia electrónica portátil que los entrenadores Pokémon llevan consigo con las fichas de todas las diversas especies Pokémon, por tanto, es el lugar idóneo para obtener la información.

2. Definir un título para el dataset. Elegir un título que sea descriptivo.

El nombre seleccionado para el nombre del dataset es pokemon.csv ya que se almacenan valores específicos de cada Pokémon.

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 contiene distintos valores de cada especie de Pokémon que permiten identificar sus habilidades, atributos, clases y condiciones físicas.

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



Imagen de la pokeball (Foto de Vincent M.A. Janssen en Pexels), que es un aparato especial para capturar y transportar Pokémon

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

El dataset incluye los siguientes datos:

- Name: Nombre del pokémon
- Evolution: Evolución del pokémon
- Type: Clase del pokémon
- Total: Puntos totales
- HP: Puntos de vida (Hit Points)
- Attack: Mide la fuerza de los movimientos
- Defense: Mide la habilidad de recibir ataques
- Sp.Atk: Mide la fuerza de los movimientos (ataques especiales)
- Sp.Def: Mide la habilidad de recibir ataques (ataques especiales)
- Speed: Define que pokémon hará el primer movimiento en la batalla
- Species: La especie a la que pertenece
- Height: Altura
- Weight: Peso
- abilities: Habilidades
- ev\_yield: Valor de esfuerzo producido (effort value yield)
- catch\_rate: probabilidad de atraparlo
- base\_friendship: Amistad base
- base\_experience: Experiencia base
- growth\_rate: velocidad de crecimiento
- eggs\_groups: Grupos de huevos
- gender\_rate: porcentaje de género
- eggs\_cycles\_rate: ciclo de huevo hasta que eclosiona

Los datos se han recogido implementando un programa de web scraping en Python que va recolectando los atributos que definen a cada pokémon y acaba generando un dataset en formato .csv para su posterior análisis.

En cuanto al periodo de tiempo de los datos, son datos que no se ven alterados en función del momento que se han 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.

Ha sido posible generar el dataset gracias a la información almacenada referente a los Pokémon de la web <https://pokemondb.net/> mediante en lenguaje de programación Python y las técnicas de web scraping.

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.

Me ha resultado interesante obtener el siguiente conjunto de datos ya que mi idea es conseguir estudiar mejor el mundo Pokémon:

¿Que clase de pokémon son mas fuertes y mas débiles? ¿Cuales son mas difíciles de encontrar? Y la respuesta mas interesante que me gustaría obtener es ¿que set de Pokémons puede ganar mas batallas en función de sus atributos?

También podríamos aplicar técnicas de clustering para generar conocimiento a partir de los datos obtenidos.

8. Licencia. Seleccione una de estas licencias para su dataset y explique el motivo de su selección:

**La licencia seleccionada es Creative Commons Attribution 4.0 International** ya que permite:

- Compartir: copiar y redistribuir el material en cualquier medio o formato
- Adaptar: remezclar, transformar y construir a partir del material para cualquier propósito, incluso comercialmente.

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

Dataset almacenado en el repositorio de github en el archivo main.py:

[https://github.com/ujaque/Pokemon\\_scraper/blob/master/main.py](https://github.com/ujaque/Pokemon_scraper/blob/master/main.py)

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

**10.5281/zenodo.4661775**

| Contribuciones              | Firma       |
|-----------------------------|-------------|
| Investigación previa        | David López |
| Redacción de las respuestas | David López |
| Desarrollo del código       | David López |