

# Dataset: Carreras populares en Galicia

*Luis Antón López Sobrado*

*06 de Abril de 2019*

## Descripción

Los datos extraídos de la página web de la federación gallega de atletismo, contienen información de las carreras populares que se van a celebrar al año en cada lugar de Galicia. Las variables que se recogen son el nombre y tipo de la carrera, la fecha, y el lugar donde se celebrará.

## Imagen identificativa



Figura 1: V carrera nocturna en Santiago de Compostela

## Contexto

Para obtener la información de las carreras populares se dispone de una página web <https://atletismo.gal/competitions> donde se muestra un listado con todas las carreras de Galicia. Esta página muestra la información de cada carrera, con la fecha, el nombre de la carrera, el lugar donde se realizará y el tipo de carrera del que se trata. Para crear el conjunto de datos se ha cogido toda la información de las carreras de Galicia durante el año 2019.

## Contenido

Para cada registro del conjunto de datos, es decir, para cada carrera se recogen las siguientes características:

- Fecha: día el que se va a realizar la carrera en el formato dd/mm/aaa.
- Nombre: edición y nombre de la carrera popular.
- Lugar: lugar donde se realizará la carrera.
- Tipo: los tipos de carreras que se muestran son:
  - Ruta → Ruta
  - Pista AL → Pista Aire Libre
  - PC → Pista Cubierta
  - Marcha → Marcha
  - CT' → Campo a Través
  - Montaña → Montaña
  - Trail → Trail

## Agradecimientos

Los datos han sido recolectados desde la web de la Federación Gallega de Atletismo. Para extraer la información alojada en las páginas HTML, se ha hecho uso del lenguaje de programación Python mediante las librerías Requests y BeautifulSoup.

## Inspiración

El conjunto de datos podría utilizarse para obtener la siguiente información. ¿Cuántas carreras se hacen al mes en Galicia? ¿Cuántas carreras realiza cada ciudad o cada pueblo? ¿De que tipo de carrera? ¿Cuántas realiza cada provincia? Estos son de gran utilidad para el ámbito periodístico y para todos los corredores.

## Licencia

La licencia escogida para la publicación de este conjunto de datos ha sido **Released Under CCo: Licencia de Dominio Publico**. He seleccionado esta licencia porque así se puede copiar, modificar, distribuir los datos y hacer comunicación pública.

## Código fuente y dataset

Tanto el código fuente escrito para la extracción de datos como el dataset generado pueden ser accedidos a través de este enlace.

<https://github.com/luilop/Web-scraping>

## **Recursos**

- David Masip Rodó (2019).El lenguaje Python. Editorial UOC
- Laia Subirats Maté, Mireia Calvo González (2019).Web scraping. Editorial UOC
- Lawson, R. (2015). Web Scraping with Python. Packt Publishing Ltd. Chapter 2. Scraping the Data.