

Dataset: Carreras deportivas en España para el año 2019

Eugenio Carmona Soriano - Antonio Ruiz Falco Rojas

7 de abril de 2019

Descripción

El conjunto de datos generado como parte de esta actividad práctica recoge los eventos de carreras deportivas que se realizan en España el año 2019. Las variables que se recogen en el conjunto de datos son la fecha, el nombre, el lugar, el tipo de carrera, la distancia, la página web de la carrera y si incluye categorías infantiles.

Imagen identificativa



Figura 1: Cartel de la media maratón de Barcelona

Contexto

El conjunto de datos se corresponde con las competiciones deportivas, más concretamente carreras deportivas, que se realizan en España el año 2019. Se incluyen varios tipos de carreras entre los que encontramos: Rutas, Trails, Triatlón, Ciclismo, Duatlón, Obstáculos y Cross/Tierra.

La página web elegida para obtener la información la ofrece a título informativo a todo aquel que quiera conocer las diferentes competiciones que se realizan, el lugar y la fecha. Además de ofrecer el enlace a la página web oficial de cada carrera.

Contenido

Para cada carrera, se crea un registro en el conjunto de datos que recogen los siguientes campos para todo el año 2019:

- **Fecha:** el día en el que se realiza la carrera en formato dd/mm/aaaa.
- **Carrera:** Nombre de la carrera.
- **Ciudad:** Ciudad en la que se realiza la carrera.
- **Provincia:** Provincia en la que se realiza la carrera.
- **Tipo:** Tipo de carrera. Ruta, Trail, Triatlón, Ciclismo, Duatlón, Obstáculos y Cross/Tierra.
- **Distancia:** La distancia a recorrer en la carrera:
- **Web:** Enlace a la página web de la carrera.
- **Infantil:** Si o No. Según incluya o no categorías infantiles.

Los autores de la página corriendovoy recopilan información de carreras y ofrecen servicios de grabación de corredores, además de noticias, el calendario de carreras y sorteo de material deportivo.

Los datos han sido recolectados desde la página web de corriendovoy mediante técnicas de Web Scraping con el lenguaje de programación Python para extraer la información en html.

Agradecimientos

A los autores de la página web corriendo voy por poner los datos al alcance de los usuarios.

Inspiración

La utilidad de este conjunto de datos es muy amplia, desde la creación de mapas visuales con las carreras hasta el análisis sobre el conjunto de datos que puedan por ejemplo ofrecer a una empresa de material deportivo tomar decisiones a la hora de publicitar/vender sus productos deportivos o patrocinar aquellos eventos que puedan repercutirle mayores beneficios.

Licencia

La licencia escogida para la publicación de este conjunto de datos ha sido CC BY-SA 4.0 License. Los motivos que han llevado a la elección de esta licencia tienen que ver con la idoneidad de las cláusulas que esta presenta en relación con el trabajo realizado:

- Se debe proveer el nombre del creador del conjunto de datos generado, indicando los cambios que se han realizado. De esta manera, se reconoce el trabajo ajeno y en qué medida se han realizado aportaciones en relación con el trabajo original.
- Se permite un uso comercial. Esto haría que incrementen las probabilidades de que una empresa utilice los datos generados y realicen trabajos de calidad que reporten cierto reconocimiento al autor original.
- Las contribuciones realizadas a posteriori sobre el trabajo publicado bajo esta licencia deberán distribuirse bajo la misma. Esto hace que el trabajo del autor original continúe distribuyéndose bajo los términos que él mismo planteó.

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/ujinuoc/runnings-ws>.

Recursos

1. Masip, D. El lenguaje Python. Editorial UOC.
2. Lawson, R. (2015). Web Scraping with Python. Packt Publishing Ltd. Chapter 2. Scraping the Data