

Práctica 1

Dataset: Scraping de las carreras IMD 2018

Extracción de la temporada de carreras del circuito de Sevilla de 10Km en 2018

Realizado por Álvaro Pavón Díaz.

1. ImdRunnerScraping
2. Extracción de los datos de la temporada de carreras del circuito de Sevilla de 10Km en 2018.
3. Imagen:



4. La materia del conjunto de datos obtenidos es la recuperación de los listados de las carreras del circuito de 2018. Se genera un conjunto de datos por cada carrera.
5. Los campos que incluyen son:
 - Posición: en la carrera
 - Dorsal: de la carrera
 - Nombre: nombre del corredor
 - Club/Entidad: si tiene aparece el nombre, en caso contrario nada
 - Tiempo Km.5: tiempo de paso por el kilómetro 5
 - Tiempo Real: desde que se pasa por la línea de salida hasta que llega a la meta
 - Tiempo Oficial: desde que se da la salida hasta que se pasa por meta
 - Media por Km.: tiempo medio en hacer un kilómetro

Los tiempos se expresan con la estructura h:mm:ss, siendo h un dígito de hora, los mm dos dígitos que expresan los minutos y ss que expresan en dos dígitos los segundos. La media por Km. se expresa con la estructura m:ss.

Los datos son obtenidos durante las carreras que se produjeron en 2018. Las fechas de las carreras quedan reflejadas en la imagen obtenidas.

6. El propietario del conjunto de datos es el IMD (Instituto Municipal de Deportes) de Sevilla que es la organización que realizan las carreras. Otros conjuntos de análisis anteriores serían los conjuntos de carreras que hay en la web o el mismo circuito, ya que podemos consultar del circuito las carreras desde el año 2008.
7. Este conjunto de datos es interesante para personas que quieran hacer análisis de evolución de las carreras, como pueden ser entrenadores o los organizadores para buscar nuevos inversores en publicidad. Las carreras en los últimos años también mueven dinero en turismo para correr grandes maratones, así que también se pueden utilizar si alguna carrera es especialmente rápida para publicitarla para atraer a corredores profesionales como reclamo.
8. La licencia elegida es CC BY-NC-SA 4.0 debido a que permite distribuir el contenido comentando siempre quien ha realizado el trabajo y reutilizarlo sin fines comerciales. Siempre que se reutilice quedará bajo los mismos derechos los otros trabajos.
9. <https://github.com/apavond/lmdRunnerScraping>