



Tipología y ciclo de vida de los datos

Práctica 1

Universitat Oberta de Catalunya

Marta Martínez y
Javier Martínez Arellano

Octubre 2018

Índice

Dataset: Jugadores de futbol profesional es España.....	2
Descripción.....	2
Imagen	2
Contexto.....	2
Contenido.....	3
Agradecimientos.....	3
Inspiración.....	3
Licencia.....	4
Código	4
Dataset.....	4

Dataset: Jugadores de fútbol profesional es España

Descripción

El conjunto de datos creado en este proyecto práctico recopia, en una misma fuente de datos, la información referente a los jugadores de fútbol profesional de primera y segunda división masculino. Agrupa datos de los jugadores, como nombre, edad, fecha de nacimiento, etc, y los relaciona con sus respectivos equipos, categoría y género.

Imagen



Contexto

La información recopilada corresponde con los jugadores de fútbol profesional español.

Aúna y correlaciona tanto información propia de los jugadores como de los equipos a los que pertenecen, así como de la categoría y la división en la que participan.

Contenido

Género: Indica el género del jugador (masculino/femenino).

Categoría: Indica la división (Primera/segunda).

Equipo: Nombre del equipo en el que participa el jugador.

Nombre: Nombre y apellidos del jugador.

Dorsal: Número que aparece en la camiseta del jugador.

Posición: Rol que desempeña (Portero/Defensa/Centrocampista/Delantero).

Fecha: Año de nacimiento.

Altura: Expresada en metros.

Peso: Expresada en kilogramos.

LugarNacimiento: Indica ciudad, provincia y país de nacimiento.

Nacionalidad: Nacionalidad del jugador.

Url Imagen: Url de la imagen de perfil del jugador.

En el directorio `Resultado/imagenes/` se descargan las imágenes correspondientes a estas Url Imagen

Agradecimientos

Los datos obtenidos proceden de la página oficial de fútbol español laliga.es y han sido extraídos por medio del lenguaje de programación Python y técnicas de Web Scraping.

Inspiración

El dataset generado podría utilizarse en aplicaciones de realidad aumentada en la que se detecte al jugador, por su dorsal por ejemplo, y se muestre información

sobreimpresión en tiempo real. Otras aplicaciones prácticas podrían ser estudios analíticos de los jugadores. Como por ejemplo, calcular la media de edad de los jugadores por equipo y, posteriormente, se podrían relacionar los valores generados con los resultados obtenidos por el club a lo largo de la temporada. A continuación, se podrían realizar previsiones de resultados y, de esta forma, poderlos aplicar en apuestas deportivas.

Licencia

La licencia elegida en nuestro caso ha sido CC BY-SA 4.0 License por distintos motivos que benefician a nuestro proyecto como pueden ser los siguientes:

Para copiar y redistribuir nuestros datos se podrán hacer en cualquier medio o formato. Otra ventaja es que estos datos o material pueden tener un uso comercial a diferencia de otro tipo de licencias que esto no lo permitirían.

Otra característica de este tipo de licencias es que si se modifica o se crea otro material a partir de nuestros datos, este tiene que estar guardado con la misma licencia de los datos originales.

Código

Código recogido en el siguiente repositorio de Github:

https://github.com/jmartinezare/TyCVD_Practica1

Además, tiene:

Wiki:

https://github.com/jmartinezare/TyCVD_Practica1/wiki/Inicio

Página web:

https://jmartinezare.github.io/TyCVD_Practica1/

Dataset

Adjuntamos los datos en formato CSV: Directorio Resultado, fichero dataset.csv