

Tipología y ciclo de vida de los datos

Práctica 1: Web Scraping



Alicia Rodríguez Gómez
6 de noviembre de 2021

Contexto

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

El objetivo era recoger datos sobre la Premier League (liga fútbol inglesa) en el año 2021. Tras una búsqueda se ha decidido elegir la web <https://es.soccerway.com/players/> para obtener estos datos. Se trata de una web con gran cantidad de datos referentes a cualquier liga de fútbol y que recoge los datos sobre traspasos de cualquier liga de futbol profesional, en particular, de la Premier League que es lo que se perseguía en esta práctica.

Título

Definir un título que sea descriptivo para el dataset.

Traspasos de la Premier League en 2021

Descripción del dataset

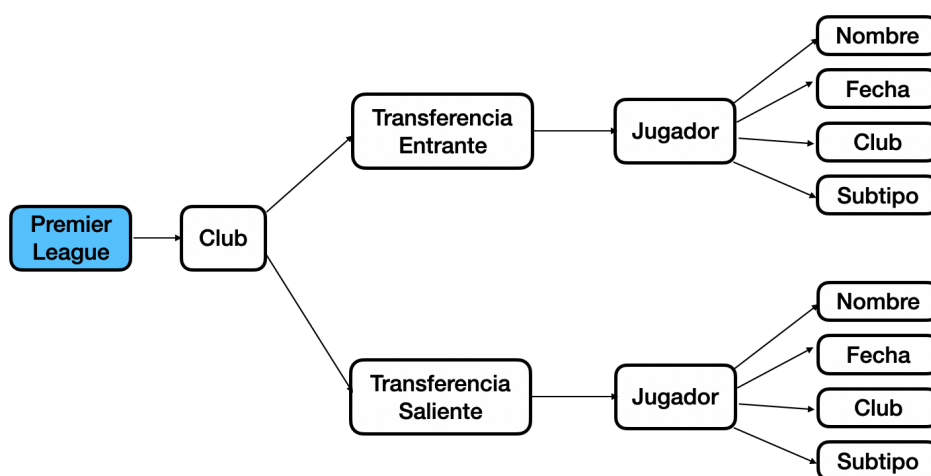
Desarrollar una descripción breve del conjunto de datos que se ha extraído. Es necesario que esta descripción tenga sentido con el título elegido.

El conjunto de datos extraído está compuesto por información relativa a los traspasos realizados en el años 2021 en la liga inglesa de fútbol, la Premier League. En particular, se tiene la información de todos los jugadores que han causado baja o alta en cualquier equipo de la Premier League. Para ello, se clasifica por equipos y para cada uno de estos se tiene la siguiente información de todos los jugadores tanto que causaron alta como causaron baja durante el 2021.

En particular, para cada jugador se obtiene el nombre, fecha del traspaso, club de origen o destino (según corresponda) y tipo de traspaso (cesión , venta con precio).

Representación gráfica

Dibujar un esquema o diagrama que identifique el dataset visualmente y el proyecto elegido



Contenido

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

Los campos por los que está compuesto el dataset son los siguientes:

- **Club:** Equipo actual al que pertenece el jugador tras el período de traspasos.

- **Tipo Transferencia:** Este campo puede tomar dos valores: Transferencias entrantes o transferencias salientes. Indica el tipo de transferencia del jugador identificados por *Nombre Jugador* respecto del club identificado por *Club*.

- **Nombre Jugador:** Nombre completo del jugador transferido.
- **Fecha Traspaso:** Fecha de la ejecución del traspaso del jugador al club.
- **Equipo:** Este campo indica el equipo de procedencia del jugador en el caso de que el tipo de transferencia sea entrante. Si la transferencia es saliente, indica el equipo de destino del jugador.
- **Subtipo Traspaso:** Indica detalles sobre el tipo de traspaso: cesión o venta. En caso de venta, indica el precio de esta.

Los datos almacenados en el CSV corresponden a los traspasos de la Premier League (liga inglesa) del periodo del año 2021. Estos datos han sido capturado con la técnica de web scraping gracias a la librería de Python, BeautifulSoup.

Agradecimientos

Presentar al propietario del conjunto de datos. Es necesario incluir citas de análisis similares. Justificar qué pasos se han seguido para actuar de acuerdo a los principios éticos y legales en el contexto del proyecto.

El propietario de los datos el Perform Media, un grupo de empresas que posee, opera o representa diversas propiedades de medios de comunicación deportivos digitales de uso gratuito, consistentes en sitios web o apps móviles.

No se han encontrado análisis similares sobre el conjunto de datos obtenidos, por lo que se han podido incluir ningún análisis.

Inspiración

Explicar 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.

Este conjunto de datos es interesante para observar los movimientos del mercado de fichajes y puede ayudar a pronosticar futuros resultados de los encuentros a lo largo de la temporada. La plantilla por la que esté formada un equipo es vital tanto para el resultado de sus encuentros como para la clasificación al final de la temporada, es por ello que hacer buenos fichajes te debe permitir acabar en mejor posición de la clasificación al final de la temporada.

Por otra parte, este conjunto de datos también es útil para observar los movimientos del mercado y responder a preguntas como ¿En qué época del mercado de fichajes se hacen la mayoría de traspasos? ¿Cuánto gastan de media los equipos grandes en traspasos? ¿Y los equipos más humildes? ¿Los equipos de la Premier League fichan o venden jugadores a equipos de su propia liga (futuros rivales) o emigran a otras ligas?

No tenemos análisis previos de este conjunto de datos, por lo que no podemos comparar con estos, pero en futuros análisis se podrían dar respuesta a las preguntas anteriormente planteadas.

Licencia

Seleccionar una de estas licencias para el dataset resultante y justificar el motivo de su selección:

La licencia elegida para el dataset resultante es **Unknown License**. El motivo de elegir esta licencia es que no he conseguido obtener el tipo de licencia que tienen los datos de origen, por lo que no me da la capacidad de elegir otro tipo de licencia para estos datos.

Código

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

<https://github.com/aliciarg6/WebScraping.git>

Dataset

Publicar el dataset obtenido en formato CSV en Zenodo con una breve descripción. Obtener y adjuntar el enlace del DOI.

Enlace DOI: <https://doi.org/10.5281/zenodo.5655248>