

TIPOLOGÍA Y CICLO DE VIDA DE LOS DATOS

PRÁCTICA 1: WEB SCRAPING

1. **Contexto:** se pretende obtener datos sobre el valor de mercado de los jugadores de los equipos de la Liga Santander, así como alguna información adicional que aparezca con el dato que queremos. El conjunto de datos se ha obtenido de una página dedicada a mostrar información sobre el valor de mercado de jugadores de fútbol.
2. **Título:** Información sobre el valor de mercado de los jugadores de la Liga Santander.
3. **Descripción:** el conjunto de datos contiene información sobre los jugadores de la Liga Santander, en concreto: equipo al que pertenece, su edad y fecha de nacimiento, su valor de mercado, su posición, su dorsal y su nacionalidad. Se han descargado los jugadores según el orden que aparecen los equipos en la web (orden alfabético) y dentro de cada equipo por posición. En total se han obtenido datos de 495 jugadores. La fecha de descarga ha sido el 6 de abril de 2019, con lo cual el valor de mercado corresponde a dicha fecha. El script se puede ejecutar en cualquier momento para actualizar dichos valores, ya que el archivo se sobrescribirá.
4. **Representación gráfica:**



Referencia: <https://www.marca.com/futbol/primeradivision/especial/2018/08/16/5b7191b3268e3ea22c8b45ac.html>

5. **Contenido:** para cada jugador se recogen los siguientes datos:
 - a. **Equipo:** equipo al que pertenece.
 - b. **Dorsal:** dorsal que porta en la temporada en curso en los partidos de la Liga Santander.
 - c. **Posición:** ubicación que tiene dicho jugador en el terreno de juego.
 - d. **Nombre:** nombre del jugador.
 - e. **Fecha de nacimiento:** fecha de nacimiento del jugador.

- f. **Edad:** edad del jugador.
- g. **Nacionalidad:** nacionalidad del jugador
- h. **Valor de mercado:** valor actual de mercado del jugador (en millones de euros).

6. **Agradecimientos:** el conjunto de datos ha sido extraído de la web de Transfermarkt (<https://www.transfermarkt.es/>). Se ha hecho uso del lenguaje Python y de las librerías BeautifulSoup y request para realizar web scraping y obtener la información alojada en las páginas HTML de cada equipo.

7. **Inspiración:** como amante del fútbol, me ha parecido muy interesante este conjunto de datos, ya que proporciona información acerca de todos los jugadores de la Primera División del fútbol español.

Con este dataset se puede, por una parte, conocer el dorsal que tiene cada jugador en su equipo, lo cual puede ser útil a la hora de elaborar camisetas o de tener identificados a los jugadores.

También podremos saber su posición, lo que nos puede servir para conocer en qué zona del campo se desenvuelven.

Podremos conocer también su nacionalidad, lo que es de utilidad cuando hay campeonatos de selecciones nacionales saber por cuál de los países puede ser seleccionado cada jugador.

Por último, se obtiene el valor de mercado dado por la web, el dato que me parece más interesante. Si se planifica una descarga automática de estos datos, se podría tener un histórico de la evolución temporal del valor de mercado de los futbolistas. Estos datos podrían posteriormente ser analizados para detectar algún tipo de tendencia, como por ejemplo una época del año en la que el valor de mercado suba o baje. Al tener además la posición y la edad de cada jugador, también podremos ver si existe una relación entre el valor de mercado de cada jugador y su puesto sobre el campo o los años que tiene.

8. **Licencia:** la licencia escogida es Released Under CC BY-NC-SA 4.0 License por los siguientes motivos:

- a. **Atribución:** de esta forma se reconoce el trabajo original y los cambios que se han realizado, permitiendo adaptar el dataset y mejorarlo con otras ideas distintas.
- b. **Uso no comercial:** este trabajo me parece más didáctico y dedicado a la investigación que algo comercial. Pienso que en el mundo del fútbol ya hay demasiados elementos comerciales como para añadir más.
- c. **Misma licencia:** todos los trabajos realizados a partir de este conjunto de datos deberán mantener la misma licencia, respetando de este modo los motivos que llevaron a escogerla.

9. **Repositorio:** el enlace al repositorio es el siguiente:

<https://github.com/gquintairos/transfermarketLigaSantander1819>

10. Contribuciones:

Contribuciones	Firma
Investigación previa	Gabriel Quintairos Rial
Redacción de las respuestas	Gabriel Quintairos Rial
Desarrollo código	Gabriel Quintairos Rial