# Pràctica 1. Tipologia i cicle de vida de les dades

## Autors:

* Álvaro Díaz
* David Leiva

## Respostes:

### 1 Context

Per tal de poder estudiar les trajectòries dels jugadors de Basket així com la composició de cada equip que participa a la Euroleague a la temporada 2020/21, vam decidir extreure la informació de la web oficial d’aquesta competició (<https://www.euroleague.net/>). Aquesta web proporciona totes les estadístiques tant del jugadors, com dels equips així com del diferents partits jugats.

Aquesta web no disposa d’una API ni es pot descarregar fitxers amb aquesta informació, pel que van haver de crear un script per fer web scraping.

### 2 Títol del dataset

Es generen dos datasets diferents:

* **euroleaguePlayers\_season**
* **euroleaguePlayers\_average**

### 3 Descripció del dataset

A ambdós datasets s’extreu la informació de les estadístiques dels jugadors en actiu la temporada 2020/21 a l’Euroleague de Basket.

* **euroleaguePlayers\_season**: S’extreu la informació desglossada per temporada jugada per cada jugador
* **euroleaguePlayers\_average:**  S’extreu les dades mitjanes proporcionades per cada jugador

### 4 Representació gràfica

### 5 Contingut

Los datasets conté la informació dels jugadors en actiu de la temporada actual i de les temporades anteriors jugades per cada jugador, amb dades que rang de dades que depèn de cada jugador, sent la més antiga de la temporada del 2002.

El lloc web permet el web scraping en general només limitant a certs rastrejadors i amb la recomendació de pauses de 15 segons entre consultes a la web.

El camps que contenen els respectius datasets són:

* **euroleaguePlayers\_season**
  + **Season:** Temporada
  + **Team:**  Equip on jugava
  + **G:** Partits jugats
  + **Pts:** Punts anotats
  + **Avg:** Mitja de punts per partit
  + **2FG:** Tirs de 2 punts en format “encerts/llançaments”
  + **2FG\_%:** Percentatge de tirs de 2 punts anotats
  + **3FG:** Tirs de 3 punts en format “encerts/llançaments”
  + **3FG\_%:** Percentatge de tirs de 3 punts anotats
  + **FT:** Tirs lliures en format “encerts/llançaments”
  + **FT\_%:** Percentatge de tirs lliures anotats
  + **Reb:** Promedi de rebotsaconseguits per partit
  + **St:** Promedi de **r**obos de pilota per partit
  + **As:** Promedi d’assistències per partit
  + **Bl:** Promedi debloqueig per partit
  + **Full\_name:** Nom complet del jugador en format “Cognom, Nom”
* **euroleaguePlayers\_average** (s’omet la descripció de les variables comunes als dos datasets)
  + **Full\_name**
  + **Surname:**  Cognom del jugador
  + **Name:** Nom del jugador
  + **Club:** Equip on juga la temporada actual
  + **Dorsal:** Número que porta el jugador
  + **Position:**  Posició que ocupa sent [“Guard”, Center”,”Forward”]
  + **Height:** Altura en metres
  + **Born:** Data de naixement en format “dd Month, Year”
  + **Nationality:** Nacionalidad
  + **G**
  + **Pts**
  + **Avg**
  + **2FG**
  + **2FG\_%**
  + **3FG**
  + **3FG\_%**
  + **FT**
  + **FT\_%**
  + **Reb**
  + **St**
  + **As**
  + **Bl**

### 6 Agraïments

Agrair la informació al propietari de les dades, el propietari del domini ‘EUROLEAGUE.NET’ , l’organització ‘Euroleague Properties S.A’ amb seu a Luxemburg.

Altres estudis utilitzant les dades de la web és:

* Vinué, Guillermo. “A Web Application for Interactive Visualization of European Basketball Data.” *Big Data*, vol. 8, no. 1, Jan. 2020, pp. 70–86. *liebertpub.com (Atypon)*, doi:10.1089/big.2018.0124.
* Horvat, Tomislav, et al. “Prediction of Euroleague Games Based on Supervised Classification Algorithm K-Nearest Neighbours:” *Proceedings of the 6th International Congress on Sport Sciences Research and Technology Support*, SCITEPRESS - Science and Technology Publications, 2018, pp. 203–07. *DOI.org (Crossref)*, doi:10.5220/0006893502030207.

### 7 Inspiració

### 8 Llicència

### 9 Codi

Link al repositori de GitHub

<https://github.com/dleiva29/UOC_Tipologia_Euroleague.git>

### 10 Dataset