

Pràctica 1. Web scraping

| | |
|----------------------------|---|
| Títol del dataset..... | 1 |
| Subtítol del dataset | 1 |
| Imatge identificativa..... | 1 |
| Context | 2 |
| Contingut..... | 2 |
| Agraïments | 4 |
| Inspiració | 4 |
| Llicència | 5 |

Títol del dataset

“Jugadors actuals de l’NBA americana”

Subtítol del dataset

“Característiques rellevants dels jugadors actuals de l’NBA americana”

Imatge identificativa



Font: <http://www.espn.com/nba/>

Context

El conjunt de dades seleccionat conté informació referent a tots els jugadors que actualment participen en la lliga de bàsket americana, més coneguda pel seu acrònim d'**NBA** (*National Basketball Association*). La temporada, per tant, de la qual s'han obtingut dades és la 2017-2018.































Aquesta extracció s'ha realitzat amb el llenguatge **Python 3**, sobre una font de dades en **format HTML**.

Contingut

La informació detallada que s'ha extret per cada un dels jugadors ha estat la següent:

- **TEAM:** Equip de l'NBA al qual pertany el jugador
- **NO.:** número de dorsal que porta el jugador actualment
- **NAME:** Nom i cognoms del jugador
- **POS:** Posició oficial dins l'equip del jugador. Aquestes posicions s'especifiquen amb els següents acrònims:
 - **PG:** point guard
 - **SG:** shooting guard
 - **SF:** small forward
 - **PF:** power forward
 - **C:** centre
 - **F:** forward
 - **G:** guard
- **AGE:** Edat actual del jugador
- **HEIGHT:** Alçada actual del jugador (en peus i polzades)
- **WEIGHT:** Pes actual del jugador (en lliures)
- **COLLEGE:** Universitat en la que s'ha format com a jugador
- **SALARY:** Sou anual del jugador (en dòlars)

Els jugadors dels quals s'han extret les dades pertanyen als següents equips:

| NBA Players | |
|--|--|
| Search By: Last Name ▼ | GO |
| Atlantic | Pacific |
|  Boston Celtics |  Golden State Warriors |
|  Brooklyn Nets |  LA Clippers |
|  New York Knicks |  Los Angeles Lakers |
|  Philadelphia 76ers |  Phoenix Suns |
|  Toronto Raptors |  Sacramento Kings |
| Central | Southwest |
|  Chicago Bulls |  Dallas Mavericks |
|  Cleveland Cavaliers |  Houston Rockets |
|  Detroit Pistons |  Memphis Grizzlies |
|  Indiana Pacers |  New Orleans Pelicans |
|  Milwaukee Bucks |  San Antonio Spurs |
| Southeast | Northwest |
|  Atlanta Hawks |  Denver Nuggets |
|  Charlotte Hornets |  Minnesota Timberwolves |
|  Miami Heat |  Oklahoma City Thunder |
|  Orlando Magic |  Portland Trail Blazers |
|  Washington Wizards |  Utah Jazz |

Font: <http://www.espn.com/nba/players>

Aquest són tots els equips que durant la temporada 2017-2018 participen en la NBA. Com podem observar, la lliga americana es divideix en 6 conferències, cada qual es compon a la seva vegada de 5 equips.

Veiem un exemple de la llista (*roster*) de jugadors de l'equip "Boston Celtics":

| Boston Celtics Roster | | | | | | | |
|-----------------------|--------------------|-----|-----|------|-----|------------------|------------------|
| TEAM ROSTER | | | | | | | |
| NO. | NAME | POS | AGE | HT | WT | COLLEGE | 2017-2018 SALARY |
| 46 | Aron Baynes | C | 31 | 6-10 | 260 | Washington State | \$4,328,000 |
| 7 | Jaylen Brown | SG | 21 | 6-7 | 225 | California | \$4,956,480 |
| 60 | Jonathan Gibson | PG | 30 | 6-2 | 185 | New Mexico State | \$44,495 |
| 20 | Gordon Hayward | SF | 28 | 6-8 | 226 | Butler | \$29,727,900 |
| 42 | Al Horford | PF | 31 | 6-10 | 245 | Florida | \$27,734,406 |
| 11 | Kyrie Irving | PG | 26 | 6-3 | 193 | Duke | \$18,868,626 |
| 8 | Shane Larkin | PG | 25 | 5-11 | 175 | Miami (FL) | \$1,471,382 |
| 55 | Greg Monroe | C | 27 | 6-11 | 265 | Georgetown | \$5,000,000 |
| 13 | Marcus Morris | PF | 28 | 6-9 | 235 | Kansas | \$5,000,000 |
| 28 | Abdel Nader | SF | 24 | 6-6 | 230 | Iowa State | \$1,167,333 |
| 37 | Semi Ojeleye | PF | 23 | 6-7 | 235 | SMU | \$1,291,892 |
| 12 | Terry Rozier | PG | 24 | 6-2 | 190 | Louisville | \$1,988,520 |
| 36 | Marcus Smart | PG | 24 | 6-4 | 220 | Oklahoma State | \$4,538,020 |
| 0 | Jayson Tatum | SF | 20 | 6-8 | 205 | Duke | \$5,645,400 |
| 27 | Daniel Theis | PF | 26 | 6-9 | 243 | | \$815,615 |
| 30 | Guerschon Yabusele | PF | 22 | 6-8 | 260 | | \$2,247,480 |

Font: http://www.espn.com/nba/team/roster/_name/bos/boston-celtics

Agraïments

Aquesta informació s'ha obtingut d'un dels portals més importants i amb més reputació d'esports americà, l'**ESPN** (*Entertainment and Sports Programming Network*), propietat del grup **ABC** i amb web oficial <http://www.espn.com>.

Inspiració

Un dels motius pels quals s'ha decidit fer web scraping sobre una informació referent a esports, i més en concret de l'**NBA**, és pel gran interès que desperta, no només en una gran quantitat d'espectadors a tot el món, sinó també en multitud d'empreses de tot tipus que obtenen importants ingressos amb aquesta lliga: agències publicitàries, canals de televisió, canals de ràdio, empreses de material esportiu, companyies aèries, cadenes hoteleres, etc.

L'**NBA** és un referent a nivell mundial, la millor lliga de bàsket amb diferència i on per descomptat juguen els millors jugadors del món.

Considerem, per tot això, que podria ser un conjunt de dades molt interessant per tasques de mineria de dades. A mode d'exemple, el salari del jugadors es podria comparar amb salaris d'altres països o inclús comparar-lo amb els salaris dels jugadors de la lliga americana universitària (**NCAA**), que és la lliga anterior a l'**NBA**, per veure si hi ha diferències considerables. O bé comprovar si els salaris ténen alguna relació directa amb la conferència o ciutat on juga un equip.

Una altra possible anàlisi seria verificar si el pes i l'alçada dels jugadors estan directament lligats a la posició que generalment ocupen. Se suposa que quan un jugador té una posició assignada prop de la cistella ha de ser més corpulent que no pas un que està més allunyat. Hi hauria algun cas que no compliria aquest patró? Quants serien?

Es podria també fer scrapping de les mateixes dades però per anys anteriors, en cas d'estar disponibles, i veure si hi ha alguna tendència rellevant, com per exemple, que el major nombre de jugadors sempre pertanyen o no a les mateixes universitats, i així fer un rànking amb les universitats que ténen millor preparació de jugadors.

Llicència

La llicència escollida per publicar aquest conjunt de dades ha estat **CC BY-NC-SA 4.0**. Les condicions d'aquesta llicència s'adapten perfectament al nostre objectiu:

- **Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made:** indicarem quina és la font original de les dades, un enllaç a les mateixes, i si hem realitzat alguna modificació sobre aquestes.
- **NonCommercial - You may not use the material for commercial purposes:** considerem que aquest conjunt de dades només tindrà com a destinatari el món acadèmic. Per tant, no volem que sigui utilitzat amb una finalitat comercial.
- **ShareAlike — If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original:** permetrem la modificació de les dades. Pot ocórrer que alguns camps com per exemple el pes o l'alçada es vulguin convertir a altres unitats equivalents, com centímetres i kilograms. Però es garanteix que es faria una nova distribució amb la mateixa llicència que hem escollit.

Font d'informació: <https://creativecommons.org/licenses>