ÍNDEX

[1. Context. 2](#_Toc35927831)

[2. Definir un títol pel dataset. 2](#_Toc35927832)

[3. Descripció del dataset. 2](#_Toc35927833)

[4. Representació gràfica. 2](#_Toc35927834)

[5. Contingut. 2](#_Toc35927835)

[6. Agraïments. 3](#_Toc35927836)

[7. Inspiració. 3](#_Toc35927837)

[8. Llicència. 3](#_Toc35927838)

[9. Codi. 3](#_Toc35927839)

[10. Dataset. 3](#_Toc35927840)

[11. Lliurar. 3](#_Toc35927841)

# Context.

La pàgina de la federació espanyola d'atletisme és [www.rfea.es](http://www.rfea.es/) té un apartat d'estadística on podem veure els diferents rànkings:

<https://www.rfea.es/web/estadisticas/ranking.asp?cod_temporada=21>

Es tracta de extreure de forma global els diferents rankings de les diferents categories, sense haver d’anar executant un per un cada filtra i fer la posterior extracció.

# Definir un títol pel dataset.

Rànkings Atletisme Temporada 2020.

# Descripció del dataset.

Desenvolupar una descripció breu del conjunt de dades que s'ha extret (és necessari que aquesta descripció tingui sentit amb el títol triat).

# Representació gràfica.

# https://www.rfea.es/sirfea/ficheros/comunicados/imagenes/_JON0049Salida.jpg

# Contingut.

Son els rankings de les diferents curses amb els atletes i clubs.

Les dades podrien ser d’l’ultim anys

Es podria analitzar quins clubs tenen millors resultats i que son més actius.

Dataset Detall

<td class="subtitulos\_tabla\_competicion">Rk</td>

<td class="subtitulos\_tabla\_competicion">Marca</td>

<td class="subtitulos\_tabla\_competicion">Apellidos y nombre</td>

<td class="subtitulos\_tabla\_competicion">Fecha nac.</td>

<td class="subtitulos\_tabla\_competicion">CA</td>

<td class="subtitulos\_tabla\_competicion">NºLicen.</td>

<td class="subtitulos\_tabla\_competicion">Federación</td>

<td class="subtitulos\_tabla\_competicion">Club</td>

<td class="subtitulos\_tabla\_competicion">Puesto</td>

<td class="subtitulos\_tabla\_competicion">Lugar</td>

<td class="subtitulos\_tabla\_competicion">Fecha</td>

# Agraïments.

Presentar el propietari del conjunt de dades. És necessari incloure cites de recerca o anàlisis anteriors (si n'hi ha).

# Inspiració.

Sóc el Ricard i faig atletisme des de fa temps. En el club on faig atletisme es va proposar fa un temps de posar al dia els registres dels atletes de l’entitat, per posar al dia les millors marques del club. Aquestes dades estaven desactualitzades ja que fa uns anys es va desvincular del club la persona que estava encarregada i ningú l’havia substituït. El club on estic és prou gran i hi ha prou atletes com per haver fet que en els últims anys hi hagi hagut molts canvis i això fa que mirar els rànquings prova a prova i temporada a temporada sigui una feina molt tediosa i difícil de fer correctament. En canvi, fent web scraping podem obtenir totes les dades del ranking i filtrar-les al gust de les necessitats.

Vaig proposar la web en el fòrum i es va unir la Núria que de seguida va veure altres possibilitats, com la d’explorar el mercat per a les empreses esportives, ja que ens pot donar una orientació d’on es troben els nuclis d’atletes més importants, i per tant on hi ha mercat per a vendre segons quins productes.

# Llicència.

Seleccionar una d'aquestes llicències pel dataset resultant i explicar el motiu de la seva selecció: ○ Released Under CC0: Public Domain License ○ Released Under CC BY-NC-SA 4.0 License ○ Released Under CC BY-SA 4.0 License ○ Database released under Open Database License, individual contents under Database Contents License ○ Other (specified above) ○ Unknown License

# Codi.

Adjuntar el codi amb el qual s'ha generat el dataset, preferiblement en Python o, alternativament, en R.

# Dataset.

Publicar el dataset en format CSV a Zenodo amb una xicoteta descripció.

# Lliurar.

Presentar el treball amb el DOI del dataset a Github