ÍNDEX

[1. Context. 1](#_Toc36452831)

[2. Definir un títol pel dataset. 1](#_Toc36452832)

[3. Descripció del dataset. 1](#_Toc36452833)

[4. Representació gràfica. 2](#_Toc36452834)

[5. Contingut. 2](#_Toc36452835)

[6. Agraïments. 3](#_Toc36452836)

[7. Inspiració. 3](#_Toc36452837)

[8. Llicència. 3](#_Toc36452838)

[9. Codi. 3](#_Toc36452839)

[10. Dataset. 3](#_Toc36452840)

[11. Lliurar. 3](#_Toc36452841)

# 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 filtre i fer la posterior extracció.

# Definir un títol pel dataset.

Rànkings Atletisme últims 5 anys.

# Descripció del dataset.

Ens aportarà tota la informació necessària per tenir diferents punts de vista a diferents nivells sigui per localitzacions, clubs, atletes...

Amb aquesta informació pretenem tenir diferents punts de vista per tractar:

* Localització de clubs més participatius
* Marques dels atletes al llarg dels anys per comprovar per edats quan estan en millor rendiment
* Atletes adolescents (explotar les seves dades)
* Informació de rellevància per als clubs d’atletisme (marques dels atletes de cada entitat al llarg del temps, comparació amb altres clubs...)
* Informació de rellevància per a entrenadors (actualment la majoria d’entrenadors copien a mà els resultats dels seus atletes en taules pròpies, excels, documents manuscrits...)

# Representació gràfica.



# Contingut.

Son els rankings de les diferents proves d’atletisme amb els atletes i clubs.

Les dades podrien ser dels últims 15 anys

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

Dataset Detall

Tipus de pista, sexe, descripció de la prova, rànquing, marca, vent, atleta, data de naixement, categoria, llicència del atleta (identifica al atleta), federació, club, posició, localització, data de la marca.

<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.

A la federació d’atletisme per poder accedir a la seva web

# Inspiració.

Sóc el Ricard i faig atletisme des de fa més de 10 anys. 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 succeït molts canvis en aquests últims anys. 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 rànquing 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, veure on hi ha mercat per a vendre segons quins productes.

# Llicència.

La Llicència pública general de GNU és una llicència de dret d'autor àmpliament usada en el món del programari lliure i codi obert, garanteix als usuaris finals (persones, organitzacions, companyies) la llibertat d'utilitzar, estudiar, compartir (copiar) i modificar el programari. El seu propòsit és doble: declarar que el programari cobert per aquesta llicència és lliure, i protegir-(mitjançant una pràctica coneguda com copyleft) d'intents d'apropiació que restringeixin aquestes llibertats als nous usuaris cada vegada que l'obra és distribuïda, modificada o ampliada. Aquesta llicència va ser creada originalment per Richard Stallman fundador de la Free Software Foundation (FSF) per al projecte GNU.

# Codi.

Adjuntem el codi en el document de Python 🡪 WebScrapingRankingRFEA.py

# Dataset.

Adjuntem el document CSV 🡪 resultats.csv

Link a github 🡪 <https://github.com/nuria-pont/UOC_M2951_Practica1>

Publiquem el dataset en format CSV a Zenodo:

<https://zenodo.org/record/3751072#.XpWbGsgzZPY>