**Tipologia i cicle de vida de les dades  
Pràctica 2**

**Celio Sánchez Bañuls  
José V. Grimalt Aranda**

**Màster de Ciència de Dades  
set.2021 - gen.2022**

**Taula de continguts**

[Integrants del grup 3](#_30j0zll)

1[. Descripció del *dataset*](#_2et92p0) [4](#_1fob9te)

[2. Integració i selecció de les dades d’interès a analitzar 4](#_3znysh7)

[3. Neteja de les dades. 5](#_2et92p0)

[4. Anàlisi de les dades. 5](#_tyjcwt)

[5. Representació dels resultats a partir de taules i gràfiques. 6](#_3dy6vkm)

[6.Resolució del problema. 7](#_1t3h5sf)

7[.](#_2s8eyo1) [Codi](#_17dp8vu) [11](#_2s8eyo1)

[Taula de contribucions al treball 12](#_26in1rg)

## Integrants del grup

Aquest són els integrants del grup que ha realitzat la pràctica:

|  |  |  |
| --- | --- | --- |
| **Nom** | **E-mail** | **Inicials** |
| Celio Sánchez Bañuls | [celiosanchez@uoc.edu](mailto:celiosanchez@uoc.edu) | CSB |
| José V. Grimalt Aranda | [jgrimalt@uoc.edu](mailto:jgrimalt@uoc.edu) | JGA |

## 1. Descripció del *dataset*

El conjunt de dades està compost per un únic fitxer amb format CSV, que conté informació de valoració de mercat dels 500 millors jugadors de futbol en cada posició segons els criteris del lloc web [transfermarkt.com](https://www.transfermarkt.com/). Aquest *dataset* inclou diferents tipus de dades:

* Identificació del jugador: nom i URL en [transfermarkt.com](https://www.transfermarkt.com/).
* Dades personals: edat, posició, nacionalitat, club.
* Dades sobre valoració: valor de mercat, *ranking* per posició.
* Dades estadístiques de la temporada en curs: gols marcats, targetes, substitucions...

Aquest és l'enllaç al repositori Git del codi amb el qual s'ha generat el *dataset*:

* URL: <https://github.com/Celiosanchez/TCVD-Practica2>

El codi s'ha desenvolupat en R, utilitzant les següents llibreries:

## 2. Integració i selecció de les dades d’interès a analitzar

Atès que es tracta d'un conjunt de dades amb informació no lligada a cap idioma, escollim un nom en anglès que descriu el contingut. Considerem important incloure el nom de l'empresa que proporciona la informació per donar crèdit i contextualitzar l'origen de les dades.

**transfermakt-com-most-valuable-football-players-YYYYMMDD.csv**

## 3. Neteja de les dades.

El fitxer conté 6500 registres i 18 variables.

Conte 6500 jugadors dividits en grups de 500 segons posició.

* Value/position
* Value/matches
* Value/goals
* Value/age
* Value/assists
* Value/yellow2cards
* Value/redcards

Dades recollides en la taula del conjunt de dades:

|  |  |  |
| --- | --- | --- |
| **Camp** | **Tipus** | **Descripció** |
| **position\_id** | *Numèric* | Identificador de la posició del jugador |
| **position\_ranking** | *Numèric* | Classificació del jugador segons la posició |
| **player** | *Text* | Nom del jugador |
| **url** | *Text* | URL relativa del jugador en el web [transfermarkt.com](https://www.transfermarkt.com/) |
| **position** | *Text* | Nom descriptiu de la posició |
| **age** | *Numèric* | Edat |
| **country** | *Text* | Nacionalitat |
| **club** | *Text* | Club o equip en el que juga actualment |
| **value** | *Numèric* | Valoració de mercat en € |
| **matches** | *Numèric* | Partits disputats |
| **goals** | *Numèric* | Gols marcats |
| **owngoals** | *Numèric* | Gols marcats en pròpia porteria |
| **assists** | *Numèric* | Assistències |
| **yellowcards** | *Numèric* | Targetes grogues |
| **yellow2cards** | *Numèric* | Expulsions per dues targetes grogues |
| **redcards** | *Numèric* | Expulsions per targeta roja directa |
| **subston** | *Numèric* | Nombre partits jugats sense ser titular |
| **substoﬀ** | *Numèric* | Nombre de vegades que el jugador és substituït |

Notes:

* Tots els camps de tipus numèric són de nombres enters.
* Les dades estadístiques (partits, gols, targetes, etc.) fan referència a la temporada en curs.

## 4. Anàlisi de les dades.

## 5. Representació dels resultats a partir de taules i gràfiques.

## 6. Resolució del problema.

El propietari

## 7. Codi

## Taula de contribucions al treball

|  |  |
| --- | --- |
| **Contribucions** | **Signatura** |
| Investigació prèvia | CSB, JGA |
| Redacció de les respostes | CSB, JGA |
| Desenvolupament del codi | CSB, JGA |