

# Dataset: Clasificación e información de los participantes de los opens 2021 de Crossfit

Ivan Aguilar Nieto & Alexander Holler Rodriguez

Abril de 2021

## Contents

<b>1</b>	<b>Introducción</b>	<b>2</b>
<b>2</b>	<b>Contexto</b>	<b>2</b>
<b>3</b>	<b>Contenido</b>	<b>2</b>
<b>4</b>	<b>Agradecimientos</b>	<b>3</b>
<b>5</b>	<b>Inspiración</b>	<b>3</b>
<b>6</b>	<b>Licencia</b>	<b>4</b>
<b>7</b>	<b>Código</b>	<b>4</b>

# 1 Introducción

Para la realización de esta práctica hemos utilizado diferentes librerías para poder analizar una web que insertaba los datos mediante JavaScript, justo después de que la página haya sido ya cargada.

Estas librerías han sido: \* selenium: Encargada de lanzar un explorador web que cargara la url de la página solicitada. \* csv: Nos permite crear y leer archivos CSV. \* bs4: Permite hacer el código HTML más fácil de leer y tratar \* urllib: Biblioteca encargada de lanzar peticiones http y obtener los datos de la pagina (No recoge los datos generados con javascript) \* json: Nos permite transformar los datos en un json y tratarlos.

## 2 Contexto

La página de la que queremos obtener los datos es la página oficial de los CrossFit Games, donde están las puntuaciones de todos los atletas que se han clasificado, cuenta con más de 3000 registros.

## 3 Contenido

Nos encontraremos con las siguientes columnas:

- **CompetitorId**: Id del competidor
- **Ranking**: Posición en la competición
- **Name**: Nombre del atleta
- **Lastname**: Apellidos del atleta
- **Country**: País por el que compite el atleta
- **Points**: Puntos totales de la competición
- **21.1**: Resultados del entrenamiento 21.1
- **21.2**: Resultados del entrenamiento 21.2
- **21.3**: Resultados del entrenamiento 21.3
- **21.4**: Resultados del entrenamiento 21.4
- **Division**: Division en la que compiten
- **Age**: Edad del atleta
- **Height**: Altura del atleta
- **Weight**: Peso del atleta
- **Affiliate**: Gimnasio donde esta afiliado
- **Affiliate\_code**: Código del gimnasio que esta afiliado
- **Back\_squat**: Marca personal en el ejercicio de Back Squat
- **Clean\_and\_jerk**: Marca personal en el ejercicio de Clean and jerk
- **Snatch**: Marca personal en el ejercicio Snatch
- **Deadlift**: Marca personal en el ejercicio Deadlift
- **Fight\_gone\_bad**: Marca personal en el ejercicio Fight Gone Bad
- **Max\_pull\_ups**: Marca personal en el ejercicio Pull Ups
- **Fran**: Marca Personal en el ejercicio Fran
- **Grace**: Marca Personal en el ejercicio Helen
- **Filthy\_50**: Marca Personal en el ejercicio Filthy 50
- **Sprint\_400m**: Marca Personal en el ejercicio Sprint 400m
- **Run\_5k**: Marca Personal en el ejercicio Run 5k
- **Affiliate\_country**: País de donde es el gimnasio que esta afiliado
- **Localitation**: Localizacion del gimnasio afiliado
- **Phone**: Telefono del gimnasio afiliado

```
# install.packages("readr")
library(readr)
path <- "../csv/dataset.csv"
data <- read_csv(path)
head(data)

## # A tibble: 6 x 31
##   CompetitorId Rank Name   Lastname Country Points `21.1` `21.2` `21.3` `21.4`
##         <dbl> <dbl> <chr>   <chr>     <chr>    <dbl> <chr>  <chr>  <chr>  <chr>
## 1         469656     1 Jeffr~ Adler    Canada     101 20 (1~ 8 (9:~ 27 (8~ 46 (3~
## 2          34796     2 Scott Panchik United~    141 33 (1~ 47 (9~ 5 (7:~ 56 (3~
## 3         105875     3 Travis Mead    United~    165 87 (1~ 24 (9~ 48 (8~ 6 (34~
## 4         310970     4 Saxon Panchik United~    217 5 (11~ 68 (9~ 87 (8~ 57 (3~
## 5          11435     5 Richa~ Froning ~ United~    254 58 (1~ 91 (1~ 5 (7:~ 100 (~
## 6           2725     6 Noah  Ohlsen   United~    272 11 (1~ 21 (9~ 2 (7:~ 238 (~
## # ... with 21 more variables: Division <chr>, Age <dbl>, Height <chr>,
## #   Weight <chr>, Affiliate <chr>, Affiliate_code <dbl>, Back_squat <chr>,
## #   Clean_and_jeark <chr>, Snatch <chr>, Deadlift <chr>, Fight_gone_bad <chr>,
## #   Max_pull_ups <chr>, Fran <chr>, Grace <chr>, Helen <chr>, Filthy_50 <chr>,
## #   Sprint_400m <chr>, Run_5k <chr>, Affiliate_country <chr>,
## #   Localitation <chr>, Phone <chr>
```

Estos son los datos en referencia a cada atleta, donde estan los datos de las competiciones entrados por un juez certificado que ha estudiado a cada uno de los atletas, ademas de toda la información referente con el atleta y el lugar donde esta afiliado.

## 4 Agradecimientos

La base de datos ha podido ser generada gracias a CrossFit, No Bull, Rebook, sus colaboradores y los propios atletas.

## 5 Inspiración

Esto podría ser interesante para analizar la tendencia de los mejores atletas y obtener predicciones sobre futuras competiciones. Así como los patrocinadores podrían priorizar a sus atletas para ganar en su inversión.

## 6 Licencia

La licencia que hemos decidido usar es la Atribución-CompartirIgual 4.0 Internacional es una licencia que permite compartir y adaptar todo el código como el Dataset y permite el uso comercial, pero estará bajo los siguientes términos:

- Atribución: Deberán dar crédito de manera adecuada, brindar un enlace a la licencia, e indicar si se han realizado cambios. De esta manera se reconocerá el trabajo del autor y el que haga uso y aporte sobre el proyecto.
- Compartir igual: Deberán compartir el código de cualquier modificación con la misma licencia, con esto beneficiamos a la creación de un mejor proyecto y que sea de dominio público.

## 7 Código

El código para acceder a este proyecto es Web Scraping Crossfit leaderboard

Contribuciones	Firma
Investigación previa	Ivan Aguilar y Alexander Holler
Redacción de las respuestas	Ivan Aguilar y Alexander Holler
Desarrollo código	Ivan Aguilar y Alexander Holler