

1. Context

Aquesta pràctica ha estat realitzada com a Pràctica 1 de l'assignatura Tipologia i cicle de vida de les dades pel Màster Universitari en Ciència de Dades de la UOC.

Quan es treballa sobre dades que generen interès personal, es treballa amb més ganes, per aquest motiu vaig estar buscant un parell de dies sobre quina temàtica podria interessar-me treballar i a més, que tingués dades on poder implementar un web *scraper* per utilitzar-les.

Després de recerca entre *datasets* del ministeri, *datasets* dels exemples de Github proporcionats en l'enunciat etc., vaig arribar a la web del Fifa, què dir? No m'interessa massa el futbol, però em va donar la idea definitiva, buscar dades d'un esport que sí m'agrada, m'interessa i practico, el voleibol.

Amb aquesta idea al cap, vaig arribar a la web de la [Real Federació Española de Voleibol](#), on es mostra informació de diferents seleccions i divisions. Així que vaig elegir fer un web *scraper* sobre dades de les Jornades dutes a terme fins a dia de l'entrega d'aquesta pràctica de la [Primera Divisió Femenina Grup A Classificació](#).

2. Títol pel dataset

He creat dos datasets: dataset_clasificacion i dataset_resultado.

El dataset_clasificacion representa la taula de classificacions general de cada jornada de la primera divisió femenina grup a de voleibol i el dataset_resultado mostra els resultats de cada partit en concret.

3. Descripció del dataset

El primer *dataset*, dataset_resultados, extreu les dades dels partits jugats, aquestes dades són els dos equips, la data i hora, sets que ha guanyat cada equip i puntuacions dels 5 possibles sets del partit.

El segon *dataset*, dataset_clasificacion, conté les dades de la classificació de la lliga de les jornades 1 a la 5 que són les jornades que s'han jugat fins a dia d'avui, 8 de novembre de 2020, conté dades com els punts que porta a la lliga o les sancions.

4. Representació gràfica

En les següents imatges es poden veure part dels *datasets* que he generat a partir de les dades de la web.

Numero	Equipo1	Equipo2	Fecha	Hora	Sets_ganados_equipo1	Sets_ganados_equipo2	Resultado_set1	Resultado_set2	Resultado_set3	Resultado_set4	Resultado_set5
2	CD Voleibol Astillero	Universidad de Salamanca	11/10/20	11:30	3	0	25-20	25-23	25-18	0-0	0-0
4	CV Bruxas	Calasancias Coruña	10/10/20	19:30	3	0	25-10	25-16	25-13	0-0	0-0
5	CV Oviedo	D,Elhuyar	10/10/20	17:00	3	0	25-13	25-15	25-14	0-0	0-0
8	D,Elhuyar	Aceites Abril CVO	17/10/20	16:30	0	3	14-25	18-25	21-25	0-0	0-0
9	Calasancias Coruña	CV Oviedo	17/10/20	13:00	0	3	16-25	17-25	16-25	0-0	0-0
12	Universidad de Valladolid	CD Voleibol Astillero	17/10/20	20:30	3	0	26-24	25-15	26-24	0-0	0-0
15	CV Bruxas	Universidad de Salamanca	24/10/20	19:30	3	0	25-12	25-17	25-22	0-0	0-0
17	Aceites Abril CVO	Calasancias Coruña	25/10/20	11:30	3	0	25-21	25-14	25-15	0-0	0-0
18	Curtidora Universidad de Oviedo	D,Elhuyar	24/10/20	19:30	3	1	15-25	25-13	25-14	26-24	0-0
6	Aceites Abril CVO	Curtidora Universidad de Oviedo	31/10/20	17:00	3	2	25-19	21-25	18-25	25-20	15-12
20	Calasancias Coruña	Curtidora Universidad de Oviedo	01/11/20	12:00	0	3	24-26	17-25	21-25	0-0	0-0
22	Universidad de Salamanca	CV Oviedo	31/10/20	17:00	1	3	12-25	25-20	11-25	20-25	0-0
24	CD Voleibol Astillero	Fertiberia Los Campos	31/10/20	16:00	2	3	22-25	25-22	25-27	25-14	7-15
27	CV Oviedo	Universidad de Valladolid	07/11/20	17:00	3	0	25-13	25-18	25-12	0-0	0-0
28	Aceites Abril CVO	Universidad de Salamanca	08/11/20	11:30	3	0	25-22	26-24	25-12	0-0	0-0
30	D,Elhuyar	Calasancias Coruña	07/11/20	16:00	1	3	25-15	23-25	16-25	25-27	0-0

Il·lustració 1: Mostra del dataset_resultados

jornada	posicion_equipo	ptos	j	g3	g2	p1	p0	sf	sc	pf	pc	sanciones
1	1. CV Bruxas	3	1	1	0	0	0	3	0	75	39	0
1	2. CV Oviedo	3	1	1	0	0	0	3	0	75	42	0
1	3. CD Voleibol Astillero	3	1	1	0	0	0	3	0	75	61	0
1	4. Universidad de Salamanca	0	1	0	0	0	1	0	3	61	75	0
1	5. D,Elhuyar	0	1	0	0	0	1	0	3	42	75	0
1	6. Calasancias Coruña	0	1	0	0	0	1	0	3	39	75	0
2	1. CV Oviedo	6	2	2	0	0	0	6	0	150	91	0
2	2. CV Bruxas	3	1	1	0	0	0	3	0	75	39	0
2	3. Aceites Abril CVO	3	1	1	0	0	0	3	0	75	53	0
2	4. Universidad de Valladolid	3	1	1	0	0	0	3	0	77	63	0
2	5. CD Voleibol Astillero	3	2	1	0	0	1	3	3	138	138	0
2	6. Universidad de Salamanca	0	1	0	0	0	1	0	3	61	75	0
2	7. D,Elhuyar	0	2	0	0	0	2	0	6	95	150	0
2	8. Calasancias Coruña	0	2	0	0	0	2	0	6	88	150	0
3	1. CV Bruxas	6	2	2	0	0	0	6	0	150	90	0
3	2. CV Oviedo	6	2	2	0	0	0	6	0	150	91	0
3	3. Aceites Abril CVO	6	2	2	0	0	0	6	0	150	103	0
3	4. Universidad de Valladolid	3	1	1	0	0	0	3	0	77	63	0
3	5. Curtidora Universidad de Oviedo	3	1	1	0	0	0	3	1	91	76	0
3	6. CD Voleibol Astillero	3	2	1	0	0	1	3	3	138	138	0

Il·lustració 2: Mostra del dataset_clasificacion

5. Contingut

El contingut del dataset té dades des de la jornada 1 a la 5 d'enguany de la primera divisió femenina, grup a, de voleibol segons la Real Federació Espanyola de Voleibol. He elegit les jornades fins a la 5 ja que l'últim partit jugat escau en el dia anterior al màxim d'entrega d'aquesta pràctica, les següents jornades encara no tenen totes les dades omplertes. S'han recollit les dades utilitzant el web scraper que podem trobar en el punt 9. Codi d'aquest document.

El dataset_resultados conté els següents camps:

1. Numero: Numero que defineix el partit a jugar entre l'equip1 i equip2
2. Equipo1: Equip que juga a casa
3. Equipo2: Equip contrincant
4. Fecha: Dia establert pel partit
5. Hora: Hora establerta pel partit
6. Sets_ganados_equipo1: Nombre de sets guanyats per l'equip1
7. Sets_ganados_equipo2: Nombre de sets guanyats per l'equip2
8. Resultados_set1: Resultats del primer set, essent el primer nombre el resultat de l'equip 1 i el segon el de l'equip2
9. Resultados_set2: Resultats del segon set
10. Resultados_set3: Resultats del tercer set
11. Resultados_set4: Resultats del quart set
12. Resultados_set5: Resultats del cinquè set

El dataset_clasificación conté els següents camps:

1. Jornada: Jornada de la classificació de la que es representen les següents dades
2. Posicion_equipo: numero de la posició en la classificació seguit del nom de l'equip
3. Ptos: punts aconseguits per l'equip
4. J: Partits jugats
5. G3: Partits guanyats els 3 sets
6. G2: Partits guanyats amb 2 sets
7. P1: Partits perduts amb 1 set guanyat

- 8. P0: Partits perduts amb 0 sets guanyats
- 9. SF: Sets a favor
- 10. SC: Sets en contra
- 11. PF: Punts a favor
- 12. PC: Punts en contra
- 13. Sanciones: Nombre de sancions de l'equip

6. Agraïments

En aquest apartat he d'agrair a la Real Federación Española de Voleibol per tenir i publicar les dades actualitzades a la seva web.

7. Inspiració

Continuant amb el que explicava al 1.Context, he vist que la web de la federació podria ajudar-me a extreure dades dels partits jugats de la primera divisió femenina i com evolucionava la classificació en cada jornada.

Els conjunts de dades extrets penso que son interessants perquè el de resultats pot donar dades que es poden comparar amb altres lligues, també pot donar informació de si un equip ha anat millorant partit darrera partit, mirant quants punts s'han fet en els sets en cada un dels partits i per ordre cronològic.

El conjunt de dades de la classificació per cada jornada pot respondre a la pregunta de com ha evolucionat un equip en la posició de la classificació i en cada jornada.

8. Llicència

He elegit una llicència de codi obert ja que d'aquesta manera qualsevol persona pot consultar les dades dels *datasets* resultants, explicant que es tracta d'un exercici per aquesta pràctica d'aquesta assignatura de la UOC, també perquè les dades extretes es poden consultar en la web de la federació en qualsevol moment, i són les mateixes, guardades de diferents maneres.

9. Codi

El *dataset* s'ha generat amb el codi propi que es pot trobar al següent repositori de Github:

<https://github.com/jorginaarres/volleyballScraper>

10. Dataset

Els *datasets* generats estan publicats a Zendoo i a Github conjuntament amb el codi. A Zendo es poden consultar a través del següent enllaç:

<https://doi.org/10.5281/zenodo.4263053>

11. Contribucions al treball

Aquest treball ha estat desenvolupat de forma completament individual , per tant la taula de contribucions queda d'aquesta manera:

Contribucions	Signa
Recerca prèvia	J.A.C
Redacció de les respostes	J.A.C
Desenvolupament codi	J.A.C