

Pràctica 1 - Web Scraping

OK Liga Taula Golejadors 2016-2021

Genís Bosch Bernat Armengol

Març 2021

1 Context

El conjunt de dades escollit per fer la nostra pràctica és la taula de golejadors de l'OK Liga de les últimes 5 temporades. Conseqüentment, el lloc web triat és el lloc web de la Real Federación Española de Patinaje[1] ja que l'OK Liga pertany en aquesta federació.

2 Inspiració

Primerament, els dos som uns grans amants de l'esport i particularment un de nosaltres és un fidel aficionat de l'hoquei.

El principal motiu pel qual vam decidir fer-ho sobre les estadístiques de l'Ok Liga és bàsicament perquè el lloc web de la mateixa federació és bastant senzill, arcaic i no es pot treure "suc" de les dades. Així mateix, les dades trobades en la wikipedia segueixen el mateix format.

Per tant, vam decidir crear un dataset on es pogués comparar les estadístiques dels golejadors en diferents temporades i, d'aquesta manera, no tenir que anar canviant el filtre de temporada ni de quina taula visualitzar de forma lenta i rudimentària.

3 Descripció

El dataset està format per tots els jugadors que han marcat gols en aquesta lliga durant les últimes 5 temporades, és a dir, des de la temporada 2016/17 fins la temporada actual.

4 Representació gràfica

La taula que mostrem a continuació està formada pels golejadors de la temporada actual. No obstant, la nostra pràctica consistirà en unir els registres de les últimes 5 temporades.






























					G	PJ	Gpp	As	App	Pen	Pen %	FD	FD %	Az	Azpp	Rj	Rjpp
1	REU		RODRIGUEZ DALMAU ALEX		35	22	1.59	5	0.23	7/11	63.64	0/2	0.00	5	0.23		
2	LIC		ADROHER MAS JORDI		32	22	1.45	5	0.23	3/5	60.00	16/39	41.03	3	0.14		
3	BAR		RODRIGUES BRAZÃO JOÃO MIGUEL		31	23	1.35	5	0.22	4/7	57.14	1/4	25.00				
4	BAR		ALVAREZ VERA PABLO FEDERICO		30	22	1.36	14	0.64	1/1	100.00	1/9	11.11	2	0.09		
5	BAR		NUNES PEREIRA HELDER		29	24	1.21	16	0.67	5/8	62.50	2/4	50.00	2	0.08		
6	NOI		COSTA FERRE XAVIER		28	21	1.33	1	0.05	9/16	56.25			6	0.29		
7	LLO		GONZALEZ ROVIROSA MARC		27	22	1.23	3	0.14			11/19	57.89	4	0.18		
8	GIR		GELMÀ PAZ DAVID		26	23	1.13	4	0.17	6/13	46.15	0/5	0.00	6	0.26		
9	CAL		MITJANS COLS ELOI		26	24	1.08	2	0.08	5/10	50.00	5/15	33.33				
10	REU		MARIN MARTIN RAUL		26	19	1.37	5	0.26	4/8	50.00	6/16	37.50	7	0.37		
11	REU		JULIÀ SOLER MARC		23	23	1.00	4	0.17			8/22	36.36	2	0.09		
12	VOL		TEIXIDO ABAT GERARD		23	23	1.00	11	0.48	2/6	33.33	8/18	44.44	5	0.22		
13	TAR		RODRIGUEZ DALMAU DANIEL		21	24	0.88	4	0.17	2/4	50.00	13/24	54.17	2	0.08		
14	PAL		CANET VILA SERGI		20	22	0.91	1	0.05	1/2	50.00			3	0.14		
15	BAR		BARGALLÓ POCH PAU		19	22	0.86	20	0.91	0/1	0.00	6/15	40.00	4	0.18		
16	VOL		CANAL TRUY ARNAU		19	24	0.79	4	0.17	3/5	60.00	0/1	0.00	3	0.13		
17	LLE		VIVES CLOS ORIOL		19	23	0.83	2	0.09	4/6	66.67	1/6	16.67	2	0.09		
18	CAD		ROSA CAPELLA FERRAN		18	19	0.95	4	0.21	0/1	0.00	7/17	41.18	4	0.21		
19	TAR		CRESPO OLIVA VICTOR		17	23	0.74			4/6	66.67	1/10	10.00	4	0.17		
20	LLO		THIEL MAX		16	22	0.73	3	0.14	2/6	33.33	0/6	0.00	3	0.14		
21	NOI		MANRUBIA PERÓ POL		16	22	0.73	10	0.45			7/16	43.75	2	0.09		
22	CAD		MIRAS VALERO SERGIO		16	22	0.73	2	0.09	6/11	54.55	0/1	0.00	2	0.09		
23	LLE		TOMÁS TOHÀ ANDREU		15	23	0.65	2	0.09	5/10	50.00	4/10	40.00	3	0.13		
24	PAL		FIGA RENJIFO MARC		15	20	0.75	1	0.05	4/9	44.44	1/4	25.00				
25	LLE		SELVA CRISTIA JOSEP MARIA		15	22	0.68	4	0.18	1/2	50.00	4/12	33.33	4	0.18		
26	CAD		GIMENEZ ADMIRABLE ALVARO BORJA		15	21	0.71	6	0.29			0/3	0.00	5	0.24		
27	BAR		PASCUAL MATÍAS JOSÉ		14	22	0.64	6	0.27					3	0.14		
28	LIC		TORRES PASTORIZA DAVID		14	20	0.70	3	0.15	1/6	16.67	0/2	0.00	2	0.10		
29	VEN		ESCALA LÓPEZ JOAN SALVADOR		14	21	0.67	3	0.14	5/7	71.43	1/3	33.33	4	0.19		

Fig. 1: Taula Golejadors Ok Lliga Temporada 20/21

5 Contingut

Cada registre és un jugador únic per cada temporada, és a dir, no pot haver-hi dos registres del mateix jugador a la mateixa temporada.

Com ja hem dit, el període de temps són les últimes 5 temporades. Tot i això, ens agradaria puntualitzar que només recollim aquest període de temps ja que només hi ha registres a partir de la temporada 2016/17.

El dataset té els següents camps:

- Temporada: Temporada en la qual el goleador ha fet els gols
- Rank: Ranking ordenat de golejadors per temporada. Per tant, aquest índex no serà únic sino que hi haurà 5 rankings simultanis, és a dir, per exemple hi haurà 5 índexs amb valor 1 per cada pitxitxi de cada temporada.

- Equip: Les sigles del club del qual forma part el golejador.
- Escut: Imatge de l'escut del club del golejador.
- Nacionalitat: Bandera nacional del país del golejador corresponent.
- Jugador: Cognoms i nom del golejador corresponent, en aquest ordre.
- Gols: Número de gols marcats pel jugador en aquella temporada.
- PJ: Número de partits jugats pel jugador en aquella temporada
- Gpp: Gols per partit del golejador en qüestió en aquella temporada.
- Asist: Número d'assistències del jugador en aquella temporada.
- App: Assistències per partit del jugador en aquella temporada.
- Pen: Proporció d'encert en penaltis del jugador en aquella temporada. Aquesta variable és una fracció de valors enters.
- Pen %: Percentatge d'encert en penaltis del jugador corresponent en aquella temporada.
- FD: Proporció d'encert en faltes directes del jugador en aquella temporada. Aquesta variable és una fracció de valors enters.
- FD %: Percentatge d'encert en faltes directes del jugador corresponent en aquella temporada.
- Az: Número de targetes blaves del jugador en aquella temporada.
- Azpp: Número de targetes blaves per partit del golejador en aquella temporada.
- Rj: Número de targetes vermelles del jugador en aquella temporada.
- Rjpp: Número de targetes vermelles per partit del jugador en aquella temporada.

6 Agraïments

Com hem comentat anteriorment, el propietari del conjunt de dades és la Real Federación Española de Patinaje.

Per una banda, hi ha hagut històricament anàlisis d'aquest lloc web per part de diaris esportius i/o clubs que formen part d'aquesta lliga, però mai sobre la taula golejadors.

Únicament, hem pogut trobar aquesta entrada a la Wikipedia de la temporada 2017/18 (temporada que nosaltres també analitzem), on podem veure que

hi ha una taula dels màxims golejadors[2].

<https://en.wikipedia.org/wiki/2017>

Top goalscorers [[edit](#)]

Raül Marín beat the record of goals in a season with 58. ^[2]

Rank	Player	Team	Goals
1	 Raül Marín	Reus Deportiu	58
2	 Pablo Álvarez	Barcelona Lassa	45
3	 Carlo di Benedetto	Liceo	28
4	 Darío Giménez	ICG Software Lleida	27
	 Maximiliano Oruste	PAS Alcoy	
6	 Jordi Ferrer	Vendrell	17
	 Raül Pelicano	Citylift Girona	
	 Sergi Pla	Igualada Rigat	
	 Marc Torra	Reus Deportiu	
	 David Torres	Liceo	

Fig. 2: Entrada Golejadors OK Lliga Wikipedia

7 Llicència

Acordem publicar el dataset sota una llicència CC BY-NC, amb la que es permet l'adaptació i ús del dataset per part de tercers sempre i quan se'ns dongui crèdit per la creació d'aquest i no s'utilitzi per propòsits comercials.

Assignem aquesta llicència perquè, tot i haver extret les dades mitjançant un scraper de la nostra autoria i no haver trobat un fitxer robots.txt que limités el web scraping a la seva web, les dades pertanyen a la Real Federación Española de Patinaje i s'hauria de preguntar a aquesta organització abans de donar-les-hi un ús comercial.

Referències

- [1] 2017–18 OK Liga. https://en.wikipedia.org/wiki/2017-18_OK_Liga. Consultat : 2021 – 03 – 25.
- [2] RFEP Hockey Patines. <http://www.hockeypatines.fep.es/>. Consultat: 2020-03-25.