

1. Contexto. Explicar en qué contexto se ha recolectado la información. Explique por qué el sitio web elegido proporciona dicha información.

La Euroliga es la máxima competición de clubes de baloncesto a nivel europeo. En la página web oficial <https://www.euroleague.net/> podemos encontrar toda la información oficial acerca de jugadores, equipos, noticias y resultados. Navegando por la web podemos ver que por cada partido disputado se muestran las estadísticas de los jugadores de ambos equipos.

2. Definir un título para el dataset. Elegir un título que sea descriptivo.

Estadísticas de los partidos de la Euroliga de baloncesto de la temporada 2019.

3. Descripción del dataset. Desarrollar una descripción breve del conjunto de datos que se ha extraído (es necesario que esta descripción tenga sentido con el título elegido).

El dataset contiene las estadísticas de todos los partidos de la temporada 2019 de la Euroliga. Estas estadísticas incluyen la información detallada sobre los minutos jugados, puntos, rebotes, bloqueos y faltas de cada jugador para todos los partidos.

4. Representación gráfica. Presentar esquema o diagrama que identifique el dataset visualmente y el proyecto elegido



Shane Larkin y Walter Tavares disputando un balón en un partido de la Euroliga

#	Player	Min	Pts	2FG	3FG	FT	O	D	T	As	St	To	Fv	Ag	Cm	Rv	PIR	season	game	team	gameid
1	SHVED, ALEXEY	30:19	22	4/7	2/10	8/10		2	2	6		3	1		1	5	19	2019	Khimki Moscow Region-Maccabi FOX Tel Aviv	Khimki Moscow Region	1
5	BOOKER, DEVIN	19:21	4	1/4		2/2	3	3	6		2	1			2	1	7	2019	Khimki Moscow Region-Maccabi FOX Tel Aviv	Khimki Moscow Region	1
6	TIMMA, JANIS	33:53	17	1/3	5/12		1	3	4	1	4	1			5	2	13	2019	Khimki Moscow Region-Maccabi FOX Tel Aviv	Khimki Moscow Region	1
8	ZAYTSEV, VYACHESLAV	10:23					1	1	2	2	1	1			2	1	3	2019	Khimki Moscow Region-Maccabi FOX Tel Aviv	Khimki Moscow Region	1
9	VIALTSEV, EGOR	4:58							1						2		-1	2019	Khimki Moscow Region-Maccabi FOX Tel Aviv	Khimki Moscow Region	1
10	DESIATNIKOV, ANDREI	DNP	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2019	Khimki Moscow Region-Maccabi FOX Tel Aviv	Khimki Moscow Region	1
11	JEREBKO, JONAS	24:29	14	2/7	2/5	4/6	1	6	7		2	1		2	2	3	11	2019	Khimki Moscow Region-Maccabi FOX Tel Aviv	Khimki Moscow Region	1
12	MONIA, SERGEY	19:12	6		2/2		2	2	2		1				1		8	2019	Khimki Moscow Region-Maccabi FOX Tel Aviv	Khimki Moscow Region	1
13	GILL, ANTHONY	23:37	18	2/6	3/3	5/6	3	3	6	4				1	5	5	22	2019	Khimki Moscow Region-Maccabi FOX Tel Aviv	Khimki Moscow Region	1
24	JOVIC, STEFAN	17:00			0/2		1		1	4		1			1	2	3	2019	Khimki Moscow Region-Maccabi FOX Tel Aviv	Khimki Moscow Region	1
40	EVANS, JEREMY	16:48	8	3/5	0/1	2/3	3	3	6			1			4	2	7	2019	Khimki Moscow Region-Maccabi FOX Tel Aviv	Khimki Moscow Region	1
0	BRYANT, ELIJAH	16:58	10	2/4	1/2	3/4	1		1	1		2			5	2	3	2019	Khimki Moscow Region-Maccabi FOX Tel Aviv	Maccabi FOX Tel Aviv	1
1	WILBEKIN, SCOTTIE	25:52	9	1/2	2/8	1/1		1	1	3	1	2			4	4	5	2019	Khimki Moscow Region-Maccabi FOX Tel Aviv	Maccabi FOX Tel Aviv	1
2	ACY, QUINCY	17:16	4	2/2	0/2		1	7	8	1	1	2			2	1	9	2019	Khimki Moscow Region-Maccabi FOX Tel Aviv	Maccabi FOX Tel Aviv	1
5	HUNTER, OTHELLO	18:15	11	5/6		1/3	3	5	8		1				1	3	19	2019	Khimki Moscow Region-Maccabi FOX Tel Aviv	Maccabi FOX Tel Aviv	1
7	CASSPI, OMRI	22:44	13	4/5	0/2	5/6		6	6	3	1	1			1	3	20	2019	Khimki Moscow Region-Maccabi FOX Tel Aviv	Maccabi FOX Tel Aviv	1
8	AVDIJA, DENI	10:34			0/1			1	1	3		1	2		2		2	2019	Khimki Moscow Region-Maccabi FOX Tel Aviv	Maccabi FOX Tel Aviv	1
11	DORSEY, TYLER	10:03	2	1/5				1	1		1	2		1			-3	2019	Khimki Moscow Region-Maccabi FOX Tel Aviv	Maccabi FOX Tel Aviv	1
12	DIBARTOLOMEO, JOHN	10:34	6		1/2	3/3		1	1	1		2			2	2	5	2019	Khimki Moscow Region-Maccabi FOX Tel Aviv	Maccabi FOX Tel Aviv	1
14	WOLTERS, NATE	18:49	7	3/4	0/1	1/1		1	1	1		1			1		5	2019	Khimki Moscow Region-Maccabi FOX Tel Aviv	Maccabi FOX Tel Aviv	1
15	COHEN, JAKE	DNP	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2019	Khimki Moscow Region-Maccabi FOX Tel Aviv	Maccabi FOX Tel Aviv	1
28	BLACK, TARIK	28:44	17	6/8		5/6	2	1	3	1			1		2	6	23	2019	Khimki Moscow Region-Maccabi FOX Tel Aviv	Maccabi FOX Tel Aviv	1
50	ZOOSMAN, YOVEL	20:11	4	2/4	0/4	0/1	1	1	2	1	1	1			2	2		2019	Khimki Moscow Region-Maccabi FOX Tel Aviv	Maccabi FOX Tel Aviv	1
7	KING, ALEX	DNP	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2019	FC Bayern Munich-AX Armani Exchange Milan	FC Bayern Munich	2
8	KOPONEN, PETTERI	11:19	3	0/2	1/2			2	2	2		1		1	2			2019	FC Bayern Munich-AX Armani Exchange Milan	FC Bayern Munich	2
10	MONROE, GREG	22:36	6	3/9			1	9	10	4		1		1	3	2	11	2019	FC Bayern Munich-AX Armani Exchange Milan	FC Bayern Munich	2
11	LUCIC, VLADIMIR	30:29	18	6/7	1/5	3/3	2	3	5		2				1	7	26	2019	FC Bayern Munich-AX Armani Exchange Milan	FC Bayern Munich	2
12	LO, MAODO	22:54	12	3/4	2/4			2	2	3		3	1		3	2	11	2019	FC Bayern Munich-AX Armani Exchange Milan	FC Bayern Munich	2
14	DEDOVIC, NIHAD	29:52	16	6/10	0/2	4/4		3	3	3	4	1		1	2	3	19	2019	FC Bayern Munich-AX Armani Exchange Milan	FC Bayern Munich	2

Primeras filas del dataset generado

5. Contenido. Explicar los campos que incluye el dataset, el periodo de tiempo de los datos y cómo se ha recogido.

El contenido del dataset son las estadísticas de todos los partidos y de todos los equipos de la Euroliga de baloncesto de la temporada 2019.

Las estadísticas incluyen el detalle de todos los jugadores, así como el total por equipo/partido.

Podemos observar a continuación ejemplo de BoxScore de un partido:

- BoxScore equipo local

KIROLBET BASKONIA VITORIA-GASTEIZ																		
#	Player	Min	Pts	2FG	3FG	FT	Rebounds					Blocks		Fouls		PIR		
							O	D	T	As	St	To	Fv	Ag	Cm		Rv	
1	PENAVA, AJDIN	DNP	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
5	GONZALEZ, MIGUEL	17:11			0/2		1	2	3							2	3	
6	CHRISTON, SEMAJ	30:37	14	6/9	0/3	2/5	1	1	2	5	1	2			3	5	13	
7	HENRY, PIERRIA	24:59	2	1/3	0/1			2	2	6	1				2	1	7	
10	LOPEZ, LAUTARO	DNP	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
11	JANNING, MATT	33:36	10	2/6	2/3		4	4	3	2	1			1	1	2	13	
12	DIOR, ILMANE	16:47	11	4/6	0/1	3/3		2	2			1			2	4	11	
17	GARCIA, SERGI	DNP	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
19	FALL, YOUSSEUPHA	3:07					1	1				3			2		-4	
23	SHENGELIA, TORNIKE	34:26	24	8/13	1/4	5/6	4	3	7	5	3	3	1	2	3	5	28	
33	POLONARA, ACHILLE	25:08	5	1/3	1/1		1	2	3	1			1	1	2	1	6	
50	ERIC, MICHAEL	14:09	7	3/5		1/3	1	1	2	1	2	2			3	2	5	
Team							2	4	6			2					4	
Totals		200:00	73	25/45	4/15	11/17	10	22	32	21	9	14	2	4	18	22	86	
				55.6%	26.7%	64.7%												
Head coach: IVANOVIC, DUSKO																		

- BoxScore equipo visitante

ALBA BERLIN																		
#	Player	Min	Pts	2FG	3FG	FT	Rebounds					Blocks			Fouls			
							O	D	T	As	St	To	Fv	Ag	Cm	Rv	PIR	
5	GIFFEY, NIELS	23:46	14	3/5	2/4	2/2	1	1	1	1	1	2		4	1	12		
6	DELOW, MALTE	5:27	4	2/2	0/1					1						4		
8	ERIKSSON, MARCUS	34:02	16	2/4	4/8		1	1	2	1	1	1		1	2	14		
9	MATTISSECK, JONAS	15:05	8		2/3	2/2	1	1	2	2		3			4	2		
15	HERMANNSSON, MARTIN	30:20	5	1/3	1/5			3	3	6		4	1	1	3	5		
16	NIKIC, KRESIMIR	5:03										1			1	1		
25	OGBE, KENNETH	12:38	5	1/2	1/2						1				1	3		
32	THIEMANN, JOHANNES	15:25	6	0/1	2/2			3	3	1		2			3	2		
34	CAVANAUGH, TYLER	18:20	2	1/1	0/2		2	6	8	2	1	2	1		3	3		
35	NNOKO, LANDRY	13:31	3	1/4		1/2	1	4	5	1		2			2	3		
43	SIKMA, LUKE	26:23	9	3/7	1/1		3	3	6	6	2	3	1	2	1	14		
Team							2	2									2	
Totals		200:00	72	14/29	13/28	5/6	8	24	32	22	4	18	4	2	22	18	79	
				48.3%	46.4%	83.3%												
Head coach: GARCIA RENESES, AITO																		

Como podemos ver, tenemos el detalle de todas las estadísticas para cada jugador. Concretamente:

- #: dorsal del jugador
- Player: nombre del jugador
- Min: minutos jugadores
- Pts: puntos anotados

- 2FG: tiros de dos puntos anotados y lanzados
- 3FG: tiros de tres puntos anotados y lanzados
- FT: tiros libres anotados y lanzados
- Rebounds O: rebotes ofensivos capturados
- Rebounds D: rebotes defensivos capturados
- Rebounds T: rebotes totales capturados
- As: asistencias
- St: robos de balón
- To: pérdidas de balón
- Blocks Fv: tapones realizados
- Blocks Ag: tapones recibidos
- Fouls Cm: faltas realizadas
- Fouls Rv: faltas recibidas
- PIR: valoración global

Ya que el dataset tendrá información de todos los partidos de la temporada hemos añadido la siguiente información a cada fila:

- Season: temporada
- Game: nombre del equipo local – nombre del equipo visitante
- Team: equipo del jugador
- GameId: identificador del partido

6. Agradecimientos. Presentar al propietario del conjunto de datos. Es necesario incluir citas de análisis anteriores o, en caso de no haberlas, justificar esta búsqueda con análisis similares.

El propietario de los datos es Euroleague, entidad privada organizadora de la competición. Los datos están disponibles en su [página web](#) para cualquiera interesado en ellos.

Un análisis similar al que hemos realizado es el que hizo Sergio Olmos y publicó en su Github, donde obtenía los datos de todos los partidos de la Euroliga para el conjunto de PlayByPlay, es decir, para los eventos registrados en cada una de las jugadas de todos los partidos.

Github de Sergio Olmos disponible en este link: <https://github.com/solmos/eurolig>

7. Inspiración. Explique por qué es interesante este conjunto de datos y qué preguntas se pretenden responder. Es necesario comparar con los análisis anteriores presentados en el apartado 6.

Dentro del mundo del baloncesto (y del deporte en general) cada vez se utilizan más los datos para ayudar en la toma de decisiones. Algunos de los usos de datos dentro del mundo del baloncesto son:

- Scouting de jugadores para evaluación de rendimiento del equipo y jugadores individualmente.
- Scouting de jugadores para posible fichaje.
- Scouting de equipos rivales.
- Estudios de las tendencias generales del juego y su evolución a lo largo de los años.
- Uso periodístico.

Considerando este contexto, tener el detalle de todas estadísticas de todos los partidos para una temporada de la Euroliga es muy interesante tanto desde el punto de vista aficionado como desde el punto de vista del profesional del baloncesto.

En comparación con el análisis de Sergio Olmos, presentado en el apartado 6, la principal diferencia es el tipo de datos que se extraen: nosotros extraemos BoxScore y Sergio Olmos extrae PlayByPlay. Con el PlayByPlay se pueden obtener detalles que no están disponibles en el BoxScore (relación entre jugadores/asistencias, tiempos, etc..) y el BoxScore nos da una versión de datos ya consolidada con la que para la mayoría de los análisis será suficiente. El uso de uno u otro dependerá del tipo de análisis que necesitemos hacer.

8. Licencia. Seleccione una de estas licencias para su dataset y explique el motivo de su selección: ☐ Released Under CC0: Public Domain License ☐ Released Under CC BY-NC-SA 4.0 License ☐ Released Under CC BY-SA 4.0 License ☐ Database released under Open Database License, individual contents under Database Contents License ☐ Other (specified above) ☐ Unknown License

No queremos que el uso de estos datos tenga ninguna limitación y nos gustaría que cualquier persona pudiera utilizarlos. Por lo tanto, utilizaremos la licencia *“Released Under CC0: Public Domain License”*.

9. Código. Adjuntar el código con el que se ha generado el dataset, preferiblemente en Python o, alternativamente, en R.

https://github.com/OscarAlvarezAlvarez/UOC_PRA1_WebScraping

10. Dataset. Publicación del dataset en formato CSV en Zenodo (obtención del DOI) con una breve descripción.

<https://zenodo.org/record/4679353#.YHLVw-gzaUk>

DOI: 10.5281/zenodo.4679353

Contribuciones	Firma
Investigación previa	GGC, OAA
Redacción de las respuestas	GGC, OAA
Desarrollo código	GGC, OAA