

# HAY VIDA DESPUÉS DE MODRIC Y KROOS ? EL BIG DATA ENCUENTRA SUS SUSTITUTOS



# ÍNDICE

- **Objeto**
- **Algoritmo de similitud: búsqueda de perfiles**
- **Métricas utilizadas**
- **Luka Modrić**
- **Toni Kroos**
- **Conclusiones**

**Dado que la carrera tan exitosa de la pareja de centrocampistas del Real Madrid, Luka Modrić y Toni Kroos, está llegando a su fin, este análisis tiene como objetivo encontrar perfiles similares en las 5 grandes ligas.**

**Para ello, con la ayuda de los datos y una serie de conocimientos adquiridos en el Máster en Big Data aplicado al Scouting en Fútbol, se tratará de obtener distintas opciones que sean accesibles de incorporar en las temporadas venideras.**

# ALGORITMO DE SIMILITUD

- **Un algoritmo en términos de Big Data es básicamente la obtención de patrones y relaciones entre variables, en este caso, métricas de fútbol. Son aquellos que ejecutan cálculos matemáticos para determinar los niveles de semejanzas o compatibilidades que existen entre conjuntos de datos.**
- **El algoritmo de similitud nos indica los nombres de los jugadores que son similares al que estamos buscando y con qué porcentaje de similitud**
- **Debemos de escoger unas reglas de entrada como el nombre del jugador al que queremos buscar similitudes, ligas, edad máxima, mínimo de minutos jugados, posición, métricas y tipo de distancia (euclidea o coseno)**

- **Las distancias coseno o euclidea son las más utilizadas para estos casos y no entraremos a valorar cuál es mejor en cada caso ya que ambas funcionan correctamente y son muy similares**
- **Para que el algoritmo sea lo más completo posible, se han elegido 24 métricas divididas en tres bloques de 8: métricas ofensivas, métricas de construcción del juego y métricas defensivas.**

**Para este análisis, se han utilizado 8 métricas de cada uno de los aspectos defensivo, construcción del juego y ofensivo.**

➤ **RENDIMIENTO DEFENSIVO**

➤ **CONSTRUCCIÓN DEL JUEGO**

➤ **RENDIMIENTO OFENSIVO**

- **RENDIMIENTO DEFENSIVO: Presiones con éxito (%), Presiones realizadas por cada 90', Recuperaciones por cada 90', Faltas por cada 90', Entradas con éxito por cada 90', Intercepciones cada 90', Duelos Aéreos ganados (%), Robos con éxito frente a regates (%)**

- **CONSTRUCCIÓN DEL JUEGO: Pases cortos con éxito (%), Pases medios con éxito (%), Pases largos con éxito (%), Pases clave (pases que dan lugar a tiro) por cada 90', Pases efectuados en el último tercio del campo por cada 90', Regates efectuados con éxito (%), Pases progresivos (pases que avanzan 9 metros hacia la portería rival o entran dentro del área) por cada 90' y Pases efectuados al área por cada 90'.**

- **RENDIMIENTO OFENSIVO:  $G - xG$  (Diferencia entre goles y goles esperados, es decir, eficacia goleadora),  $xA$  (Asistencias esperadas) por 90', Goles por 90', Asistencias por 90', Toques en área rival por 90', Pases que acaban en tiro por 90', Conducciones progresivas (que avanzan al menos 5 metros hacia el área rival o entran dentro del área) por 90' y Tiros a puerta (%).**

**Ejecutamos el algoritmo de similitud con las siguientes variables de entrada:**

**Jugador = Luka Modrić**

**Mínimo de minutos jugados = 1000**

**Distancia = coseno**

**Métricas = Las 24 descritas anteriormente**

**Ligas = 5 grandes ligas (La Liga, Serie A, Premier League, Ligue 1 y Bundesliga)**

**Posición = Centrocampista**

**Edad máxima = 25 años**

**Número de jugadores similares = 10**

# LUKA MODRIĆ

**El resultado que saca el algoritmo es el siguiente:**

|   | Jugador                | Similitud |
|---|------------------------|-----------|
| 0 | Mattéo Guendouzi       | 94.597    |
| 1 | Renato Sanches         | 88.449    |
| 2 | Frenkie de Jong        | 84.092    |
| 3 | Nicolò Barella         | 80.925    |
| 4 | Bruno Guimarães        | 79.753    |
| 5 | Maxime Lopez           | 79.311    |
| 6 | Federico Valverde      | 75.108    |
| 7 | David Pereira da Costa | 74.856    |
| 8 | Cheick Doucouré        | 74.057    |
| 9 | Khéphren Thuram-Ulie   | 73.382    |

**Utilizaremos los tres primeros candidatos ya que son los que mayor porcentaje de similitud presentan, a excepción de Frenkie de Jong, ya que será prácticamente imposible robárselo al FC Barcelona. En su lugar, analizaremos a Nicolás Barella.**

# LUKA MODRIĆ

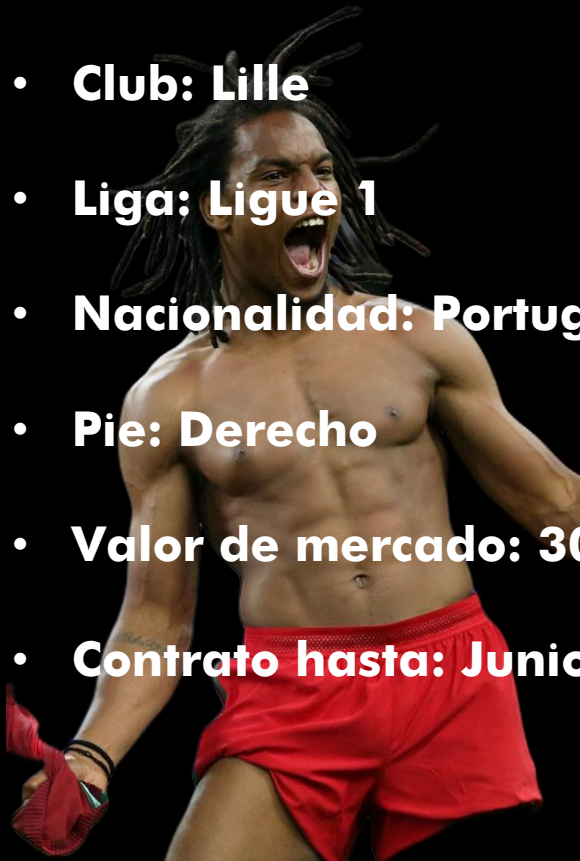
## Mattéo Guendouzi

- **Edad: 23 años (1999)**
- **Club: Olympique de Marsella (cedido de Arsenal FC)**
- **Liga: Ligue 1**
- **Nacionalidad: Francés / Marroquí**
- **Pie: Derecho**
- **Valor de mercado: 22M**
- **Contrato hasta: Junio 2023**



## Renato Sanches

- **Edad: 24 años (1997)**
- **Club: Lille**
- **Liga: Ligue 1**
- **Nacionalidad: Portugués**
- **Pie: Derecho**
- **Valor de mercado: 30M**
- **Contrato hasta: Junio 2023**



## Nicolò Barella

- **Edad: 25 años (1997)**
- **Club: Inter de Milán**
- **Liga: Serie A**
- **Nacionalidad: Italiano**
- **Pie: Derecho**
- **Valor de mercado: 70M**
- **Contrato hasta: Junio 2026**

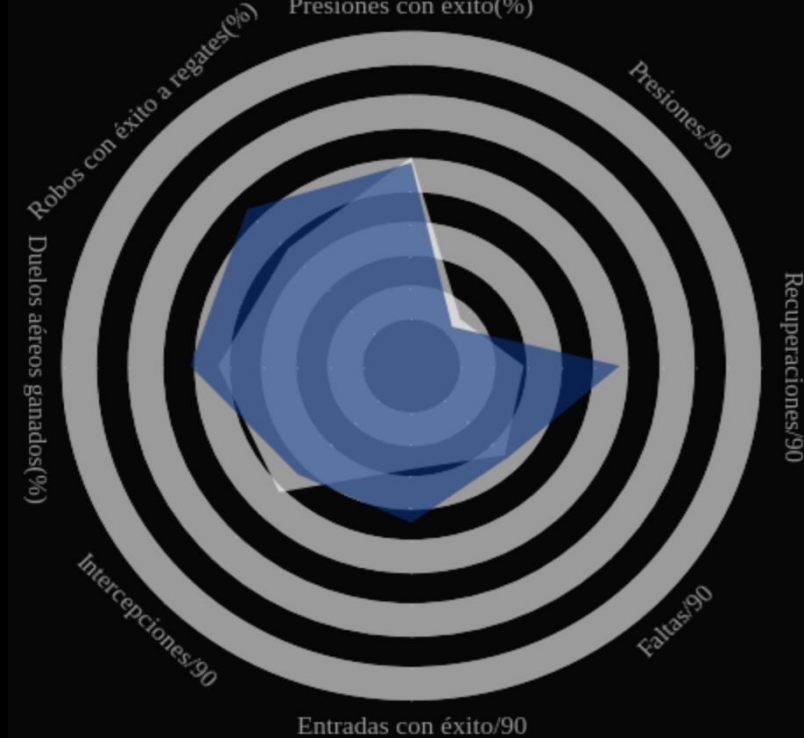
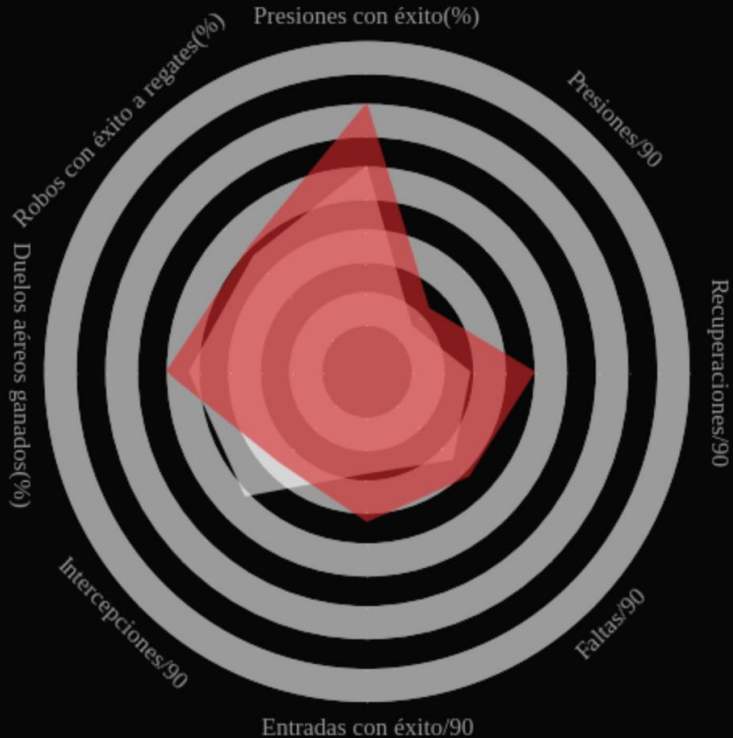
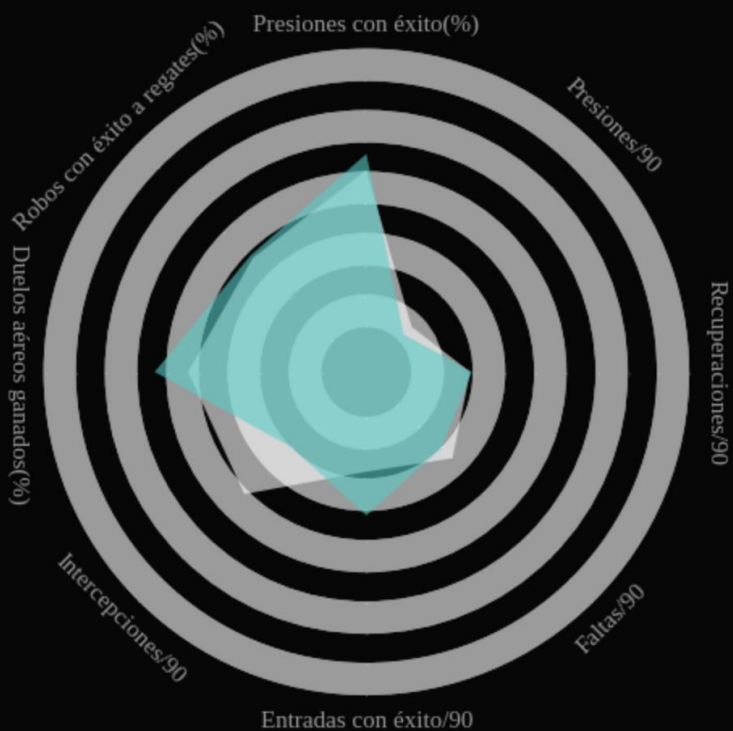


# LUKA MODRIĆ - DEFENSIVO

**Mattéo Guendouzi**

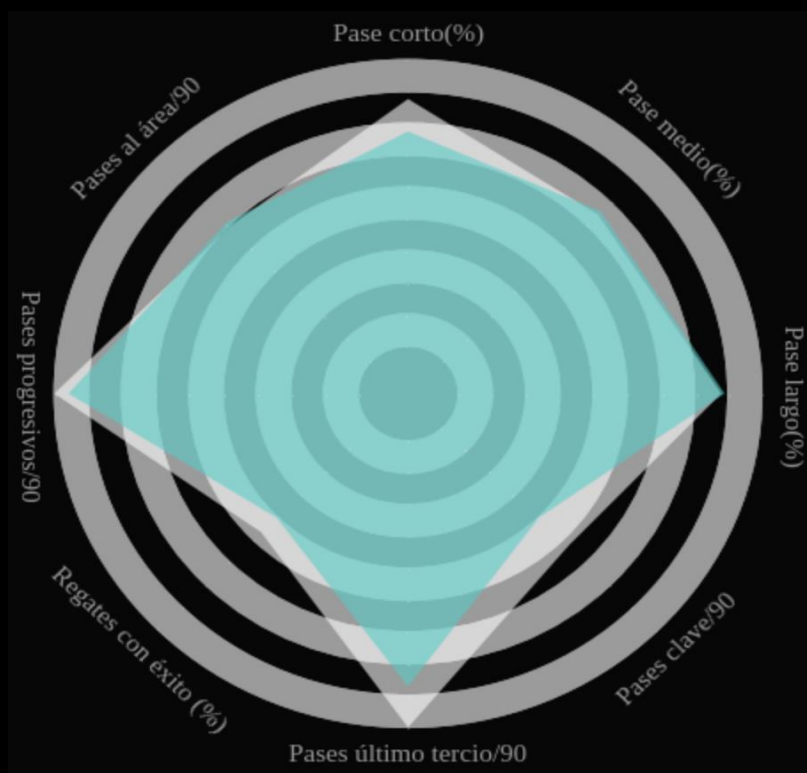
**Renato Sanches**

**Nicolò Barella**

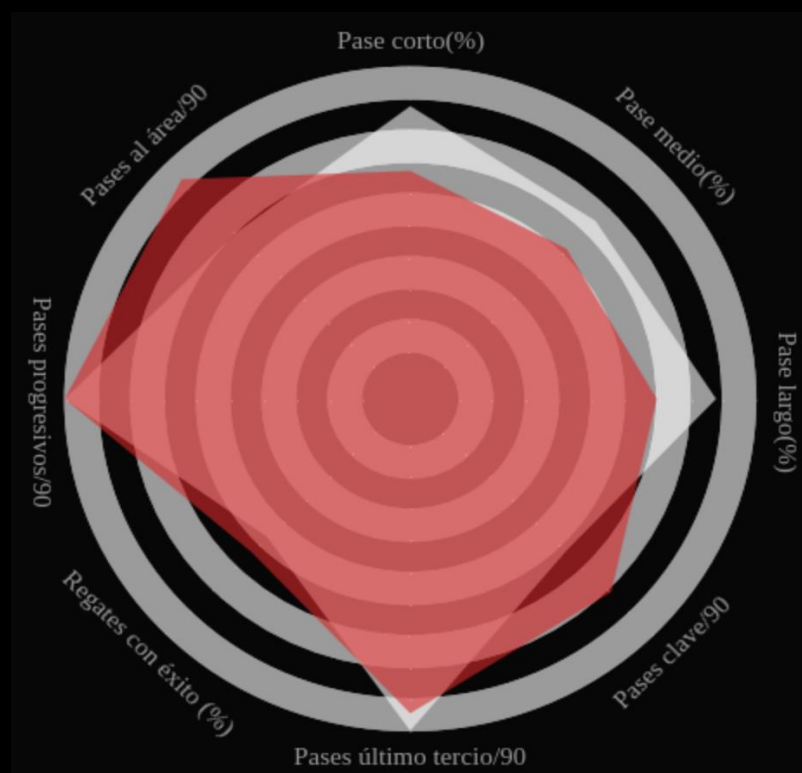


# LUKA MODRIĆ - CONSTRUCCIÓN

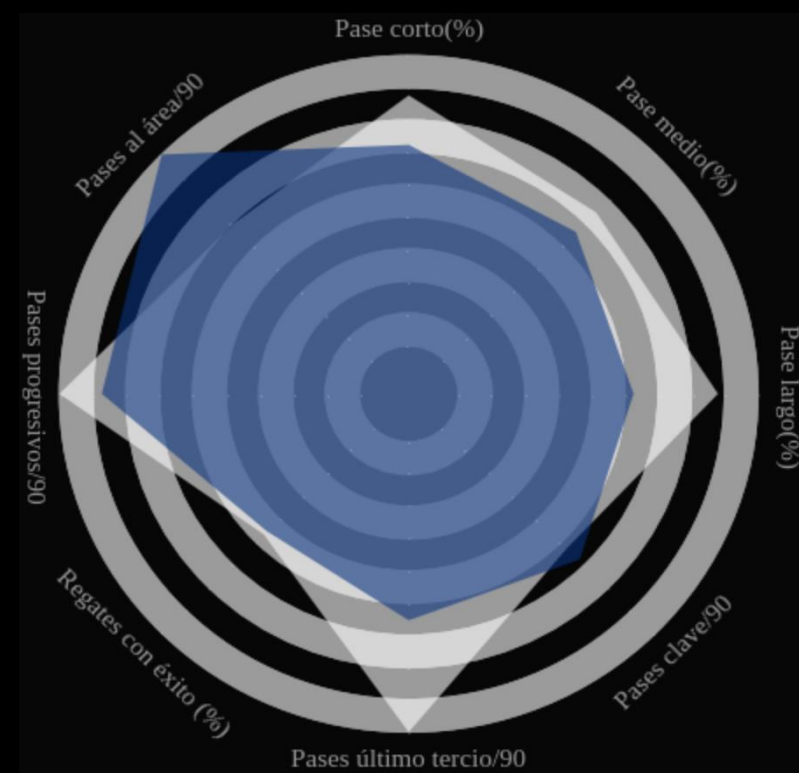
**Mattéo Guendouzi**



**Renato Sanches**

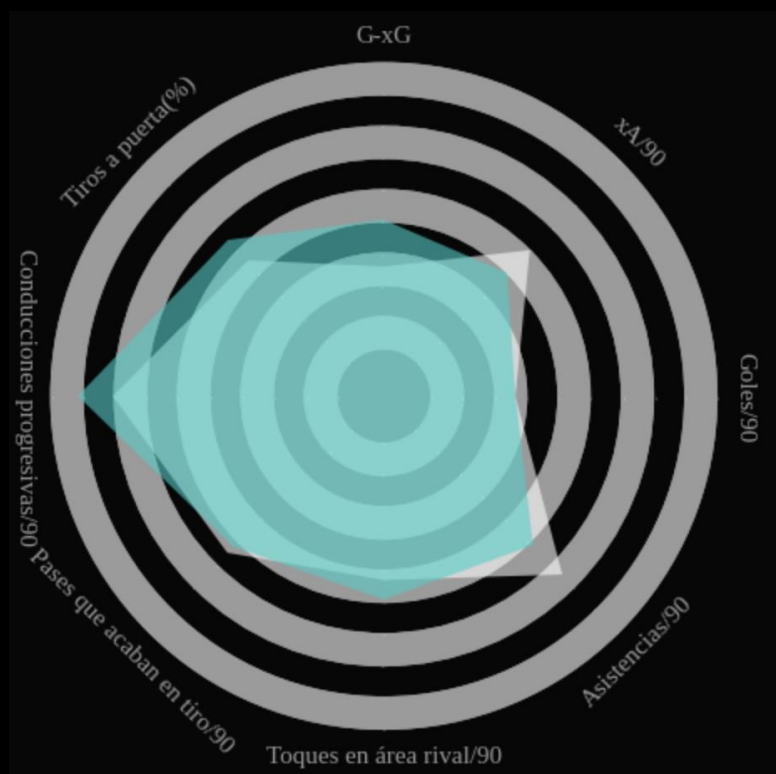


**Nicolò Barella**

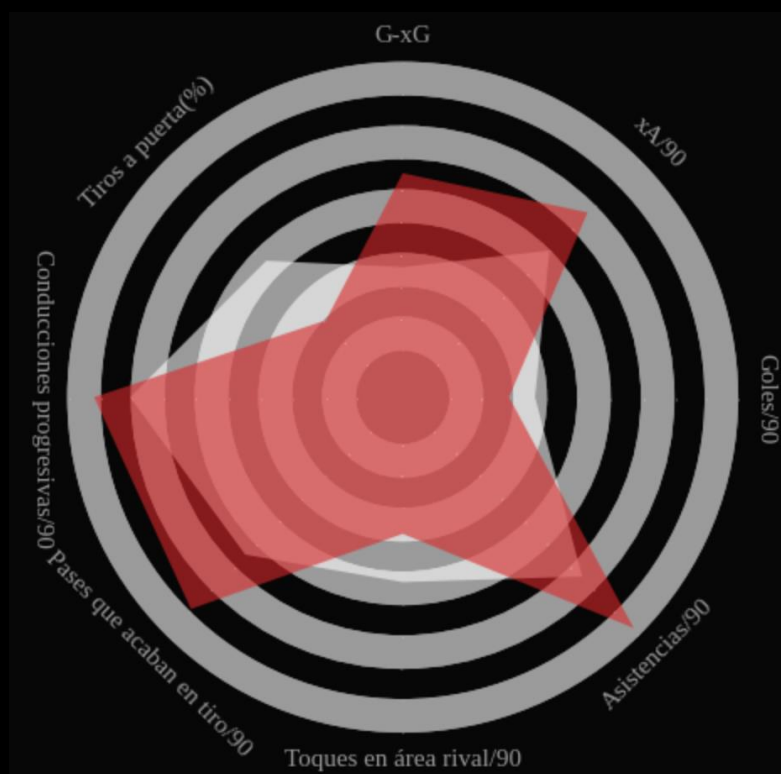


# LUKA MODRIĆ - OFENSIVO

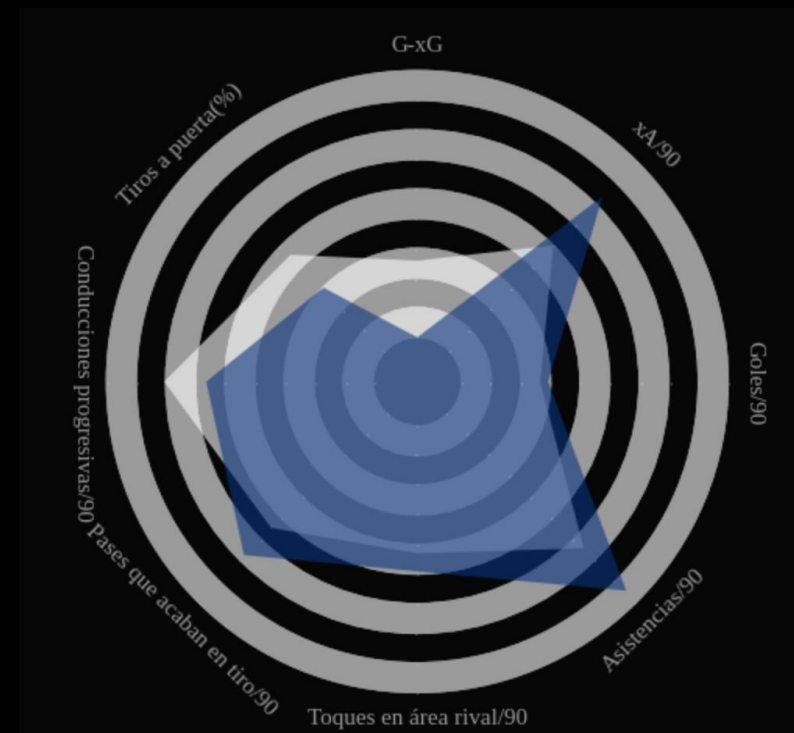
**Mattéo Guendouzi**



**Renato Sanches**



**Nicolò Barella**



**Ejecutamos el algoritmo de similitud con las siguientes variables de entrada:**

**Jugador = Toni Kroos**

**Mínimo de minutos jugados = 1000**

**Distancia = coseno**

**Métricas = Las 24 descritas anteriormente**

**Ligas = 5 grandes ligas (La Liga, Serie A, Premier League, Ligue 1 y Bundesliga)**

**Posición = Centrocampista**

**Edad máxima = 25 años**

**Número de jugadores similares = 10**

# TONI KROOS

**El resultado que saca el algoritmo es el siguiente:**

|   | Jugador              | Similitud |
|---|----------------------|-----------|
| 0 | Maxime Lopez         | 85.639    |
| 1 | Bruno Guimarães      | 84.705    |
| 2 | Mattéo Guendouzi     | 83.515    |
| 3 | Renato Sanches       | 80.061    |
| 4 | Rodri                | 74.243    |
| 5 | Nicolò Barella       | 73.664    |
| 6 | Albert Sambi Lokonga | 70.871    |
| 7 | Pedro Chirivella     | 70.235    |
| 8 | Cheick Doucouré      | 68.332    |
| 9 | Frenkie de Jong      | 68.179    |

**Utilizaremos los tres primeros candidatos como hemos hecho con el ejemplo anterior. Sin embargo, obviaremos aquellos nombres que nos salieron similares a Luka Modric**

## TONI KROOS

## Maxime López

- Edad: 24 años (1997)
- Club: US Sassuolo
- Liga: Serie A
- Nacionalidad: Francés / Argelino
- Pie: Derecho
- Valor de mercado: 14M
- Contrato hasta: Junio 2025



## Bruno Guimarães

- Edad: 24 años (1997)
- Club: Newcastle United
- Liga: Premier League
- Nacionalidad: Brasileño
- Pie: Derecho
- Valor de mercado: 30M
- Contrato hasta: Junio 2026



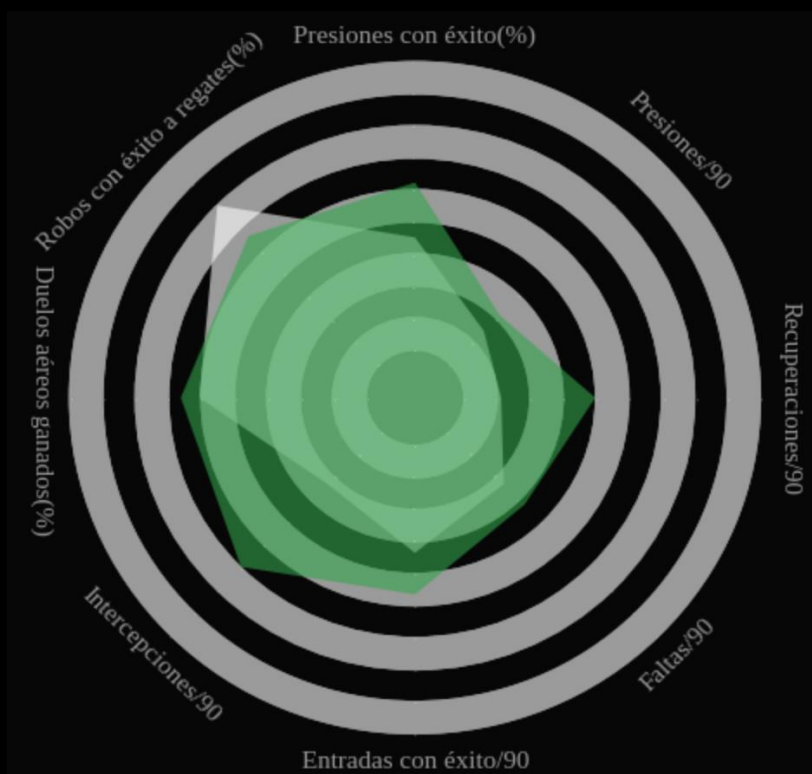
## Rodri Hernández

- Edad: 25 años (1996)
- Club: Manchester City
- Liga: Premier League
- Nacionalidad: Español
- Pie: Derecho
- Valor de mercado: 70M
- Contrato hasta: Junio 2024

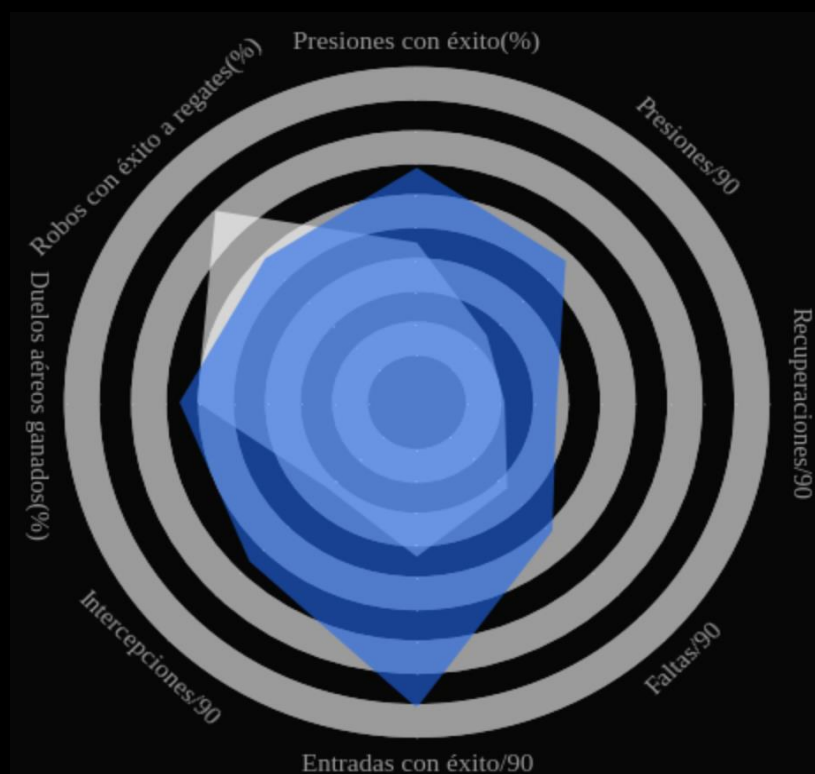


# TONI KROOS - DEFENSIVO

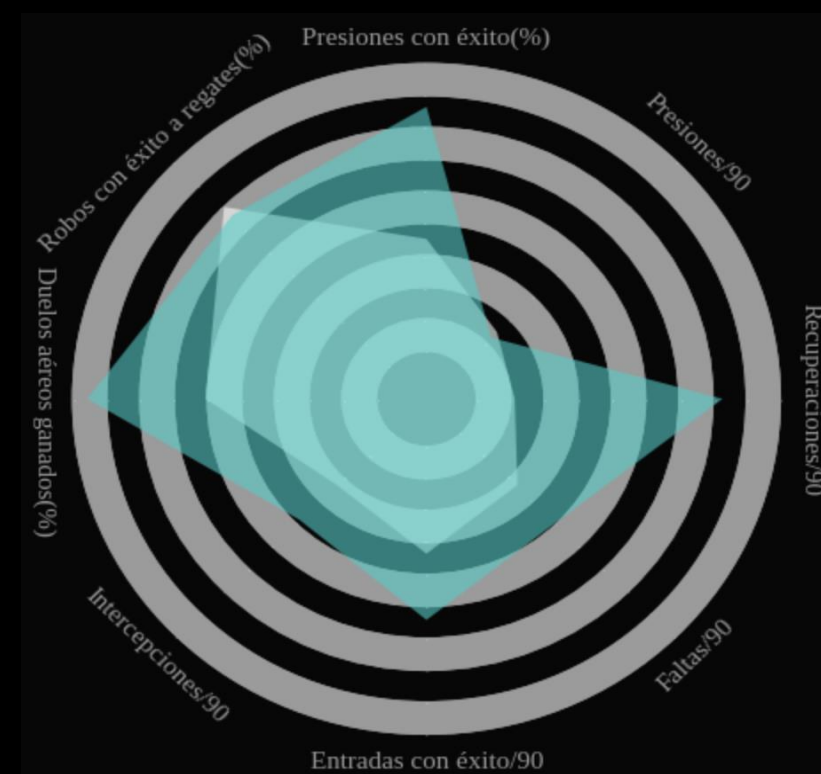
**Maxime López**



**Bruno Guimarães**

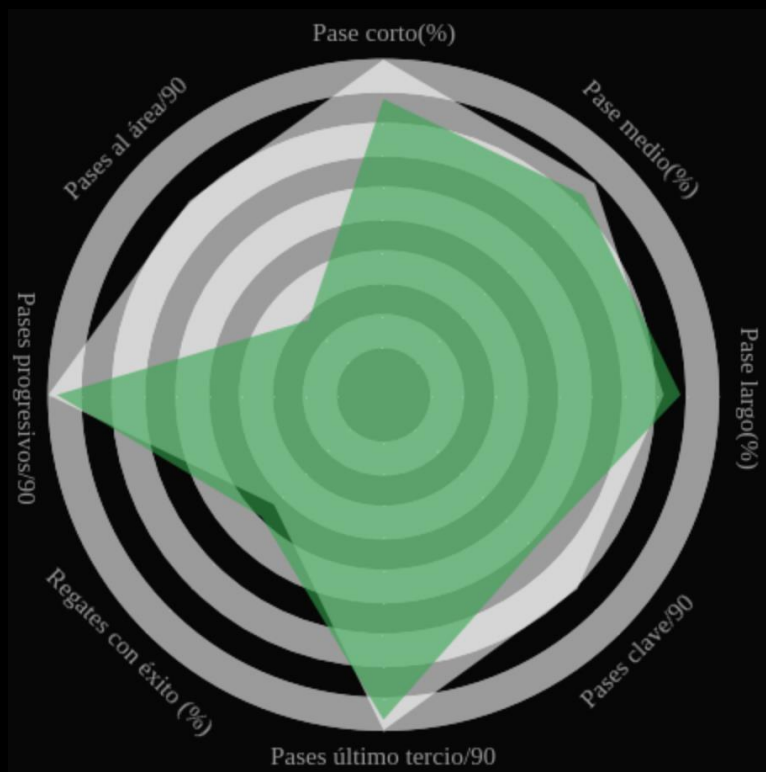


**Rodri Hernández**

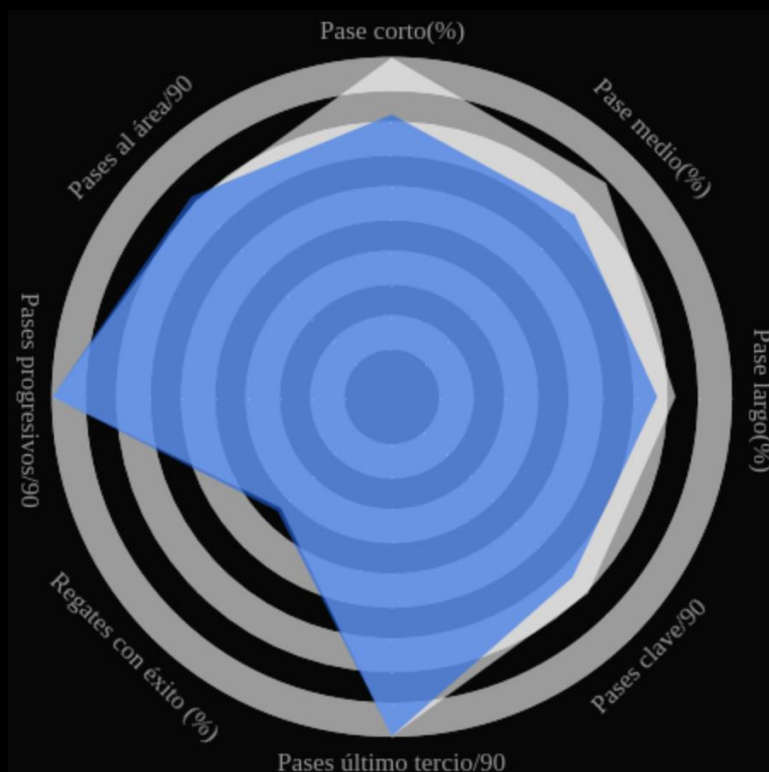


# TONI KROOS - CONSTRUCCIÓN

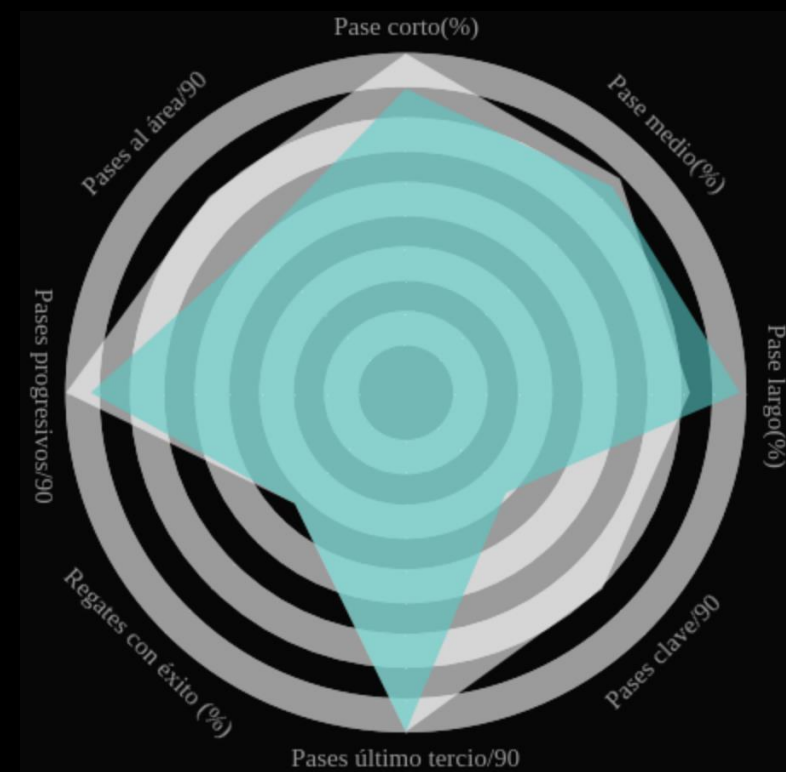
**Maxime López**



**Bruno Guimarães**

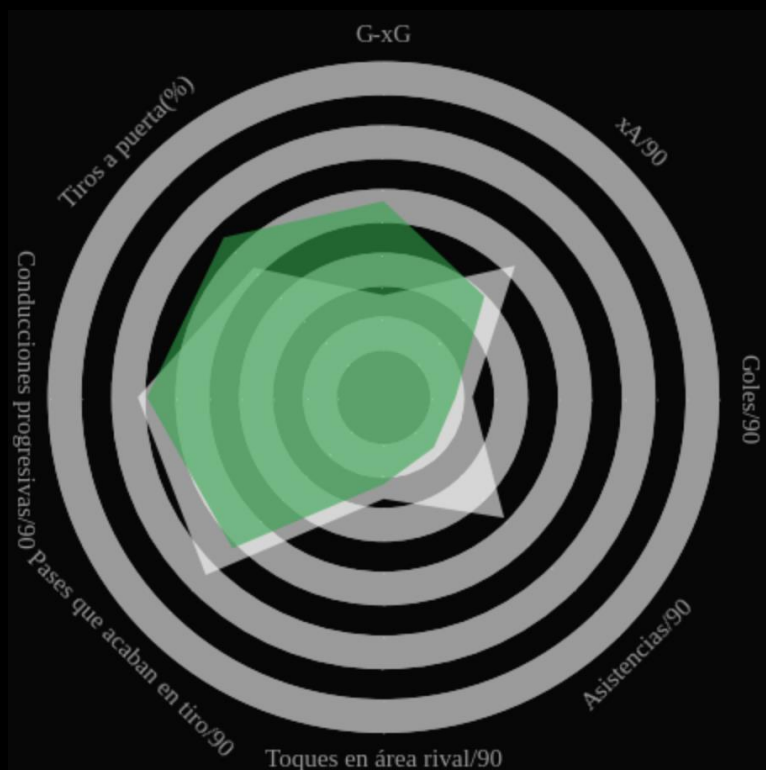


**Rodri Hernández**

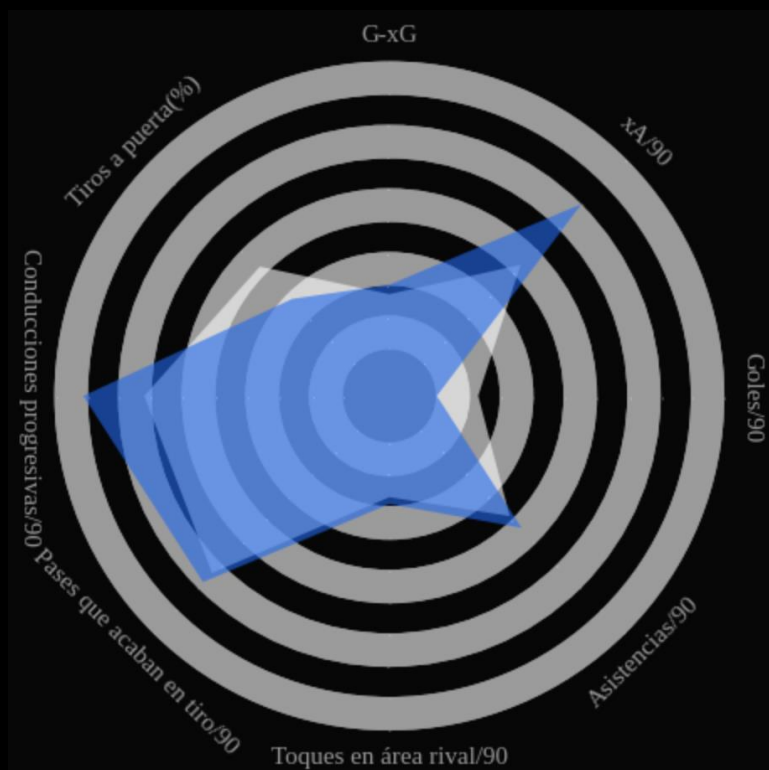


# TONI KROOS - OFENSIVO

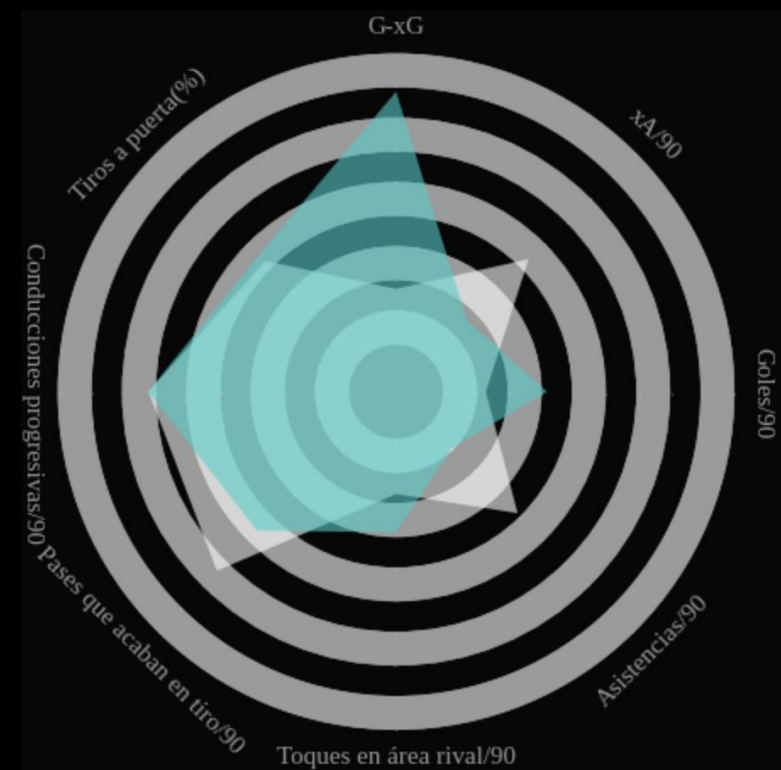
Maxime López



Bruno Guimarães



Rodri Hernández



# CONCLUSIONES

- A nivel defensivo, es Rodri el que se impone, aunque tanto Bruno Guimaraes como Maxime López también tienen mejores números que los dos jugadores madridistas
- En cuanto a la creación del juego, los más similares se podría decir que son Guendouzi, aunque con menos pases comprometidos, y Bruno, con menor acierto en el pase. Sin embargo, cabe destacar también a Renato Sanches ya que, aunque tenga menor acierto en el pase, tiene mayor influencia en pases clave (que acaban en tiro) y pases dentro del área
- Ofensivamente, sin duda Renato es el que más aspectos abarca y genera. A pesar de ello, el centrocampista español es que más gol tiene

# CONCLUSIONES

- **Guendouzi es el más similar, así lo refleja el algoritmo (94% con Modric, 85% con Kroos) y se ve claramente en las comparaciones con gráficas de radar**
- **Rodri es el más defensivo y quien más gol tiene**
- **Renato, a pesar de no tener ese porcentaje de éxito tan alto en el pase, aporta mayor rendimiento en el área rival**
- **Maxime supera en números defensivos al croata y alemán, pero genera menos situaciones de gol**
- **Barella es ligeramente superior en acciones defensivas y situaciones con balón cercanas al área, pero tiene peor porcentaje de éxito en pases**
- **Guimaraes tiene un perfil similar en creación del juego (ligeramente peor en % de pases) y supera a ambos en acciones defensivas y ofensivas**

*Aritz Zabaleta Ibero*

# Máster en Big Data Aplicado al Scouting en Fútbol



@aritzzabaleta14



[https://www.linkedin.com/in/  
aritz-zabaleta-ibero-  
652721b1](https://www.linkedin.com/in/aritz-zabaleta-ibero-652721b1)



aritzzabaleta14@gmail.com