

Unbetting Football, Capítulo II

La Caja de Pandora

Para elaborar el siguiente estudio pedí permiso a Joseph Buchdahl de Football-data.co.uk para emplear sus datos sobre resultados deportivos y apuestas vinculadas a ellos. Se trata de una base de datos de alrededor de 150,000 partidos ocurridos entre 2000 y 2021 en 28 ligas europeas, a lo que añadí resultados y datos de la Liga MX. Todos los datos presentados en este capítulo involucran solo resultados de partidos de liga o temporada regular. El tratamiento de datos fue hecho con Python 3.9.7.

El 24 de octubre de 2020 ocurrió el partido Barcelona vs Real Madrid y el escenario del partido, según las casas de apuestas, fue el siguiente:

Barcelona	Empate	Real Madrid
48%	26%	26%

Antes de continuar debemos tener en mente que una probabilidad no puede decirnos qué ocurrirá en un evento en específico, más bien:

La Probabilidad es una descripción de lo que tendería a ocurrir en el largo plazo

Continuando con la narración, supongamos que fuese posible que, en octubre de 2020, en ese momento en el tiempo, ese Barcelona y ese Real Madrid decidieran zanjar de una vez por todas quién es el mejor entre esos dos clubes, para lo que jugarían una serie de 100 partidos entre ellos. Si jugasen un partido por semana tardarían poco menos de 2 años en cumplir la centena de partidos. Pero ¿y si observamos qué ocurrió en los otros partidos que tenían el mismo escenario? Por **escenario** me refiero a los partidos que fueron clasificados por las casas de apuestas bajo el siguiente esquema:

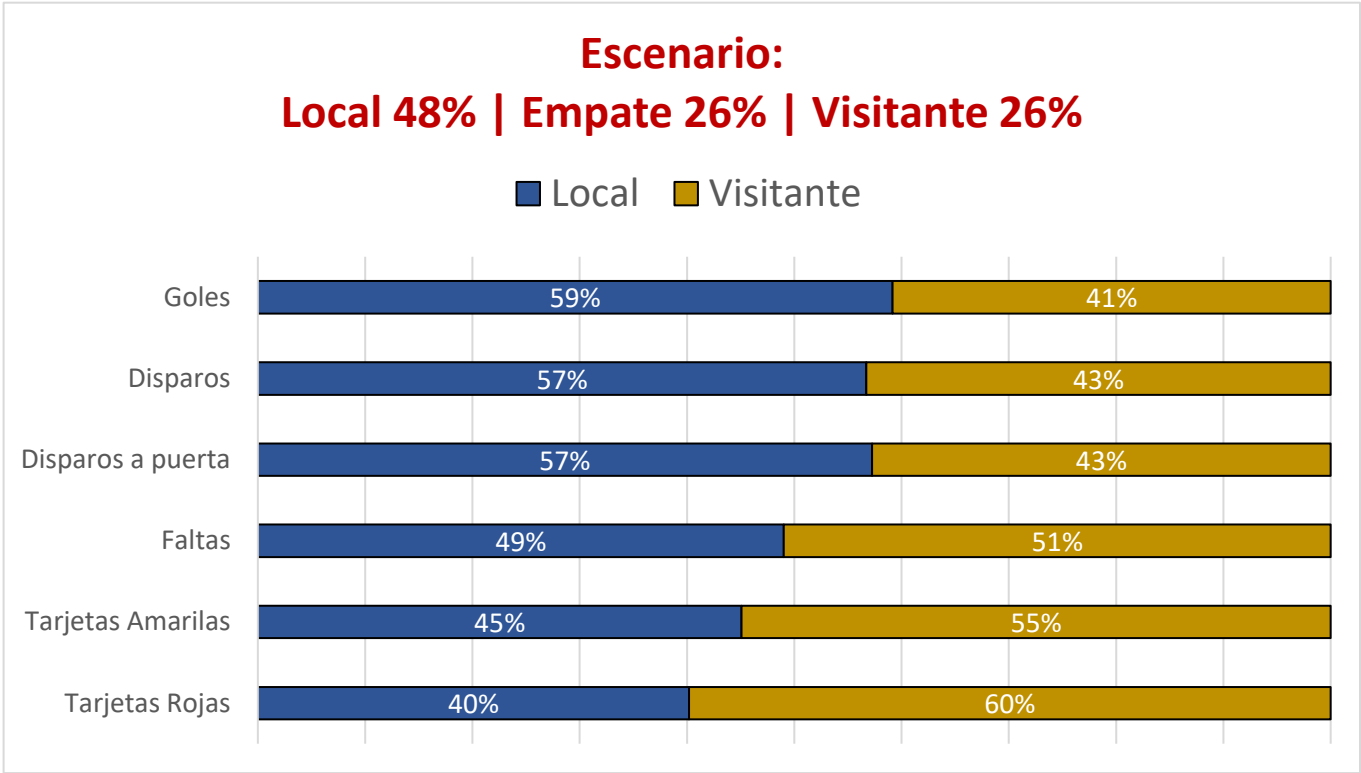
Victoria del Local	Empate	Victoria del Visitante
48%	26%	26%

Supongamos que esa selección de partidos resultó ser, exactamente, de 100 encuentros. ¿Cómo se distribuirían los resultados de esos 100 partidos? ¿Cuántas veces ganarían los equipos locales? ¿cuántas ganarían los visitantes o cuántas empatarían? Las cantidades tenderían a ser las siguientes:

Victorias de los Locales	Partidos Empatados	Victorias de los Visitantes
48	26	26

Por supuesto, los partidos que caen en el escenario no son solo 100, en realidad fueron 2075, pero si tomásemos muestras aleatorias de 100 partidos de ese escenario, los promedios de los resultados tenderían a lo expresado en la tabla anterior. Reuniendo esos 2075 partidos se obtuvieron los promedios de las estadísticas registradas. Por ejemplo, se encontró que los equipos locales anotaron en promedio 1.55 goles por partido, mientras que los visitantes promediaron 1.07 goles por partido, pero ¿qué nos dice esto? Una manera de interpretar los resultados podría ser pensarlos en términos de 100 partidos. Si dos equipos se enfrentasen en estas mismas circunstancias, al cabo de 100

partidos entre ellos esperaríamos que el local tendiese a anotar 155 goles (59% del total), mientras que el visitante tendería a 107 (41% del total). Observemos el resumen presentado en la siguiente gráfica:



Una manera de interpretar el gráfico anterior es que, si pudiésemos extender en el tiempo los partidos, es decir, aumentar la cantidad de partidos entre el Barcelona y el Real Madrid del 24 de octubre de 2020, los resultados tenderían a lo que se muestra en la gráfica. Ahora bien, por supuesto que este tipo de competición (100 partidos seguidos contra el mismo rival) sería impracticable en casi cualquier deporte, en especial en el caso del fútbol. Por ejemplo, en el caso de una temporada de LaLiga, Serie A, Ligue 1 o Premier League, cada equipo juega 38 partidos que representan 38 escenarios distintos. Imaginemos un caso más simple y, después, veamos si a partir de ahí llegamos a algo interesante (y llegaremos).

Imaginemos que hay una liga de fútbol en la que un equipo al que llamaremos ELE KIPAZO, se enfrentará solo a dos rivales a los que únicamente recibirá como local y, además, deberá enfrentarse 10 veces a cada uno, dando un total de 20 partidos. Clasificando cada partido según la Probabilidad de Victoria del Local tendríamos los siguientes dos escenarios:



Como seguro estarás intuyendo, existe una nutrida muestra disponible para cada uno de esos escenarios (algunos miles, de hecho, entre 2000 y 2021). Por ahora pensemos solo en términos de otra estadística, consideremos primero los **Puntos Promedio** que, para esos 3 escenarios, serían los siguientes:

Escenarios	Puntos Promedio de ELE KIPAZO
Victoria 60%	2
Victoria 40%	1.5

Si consideramos que en cada escenario se jugarían 10 partidos, podríamos solo **multiplicar por 10 cada uno** de los promedios y sacar sus totales sumándolos, para encontrar los **Puntos Esperados**, como se muestra aquí:

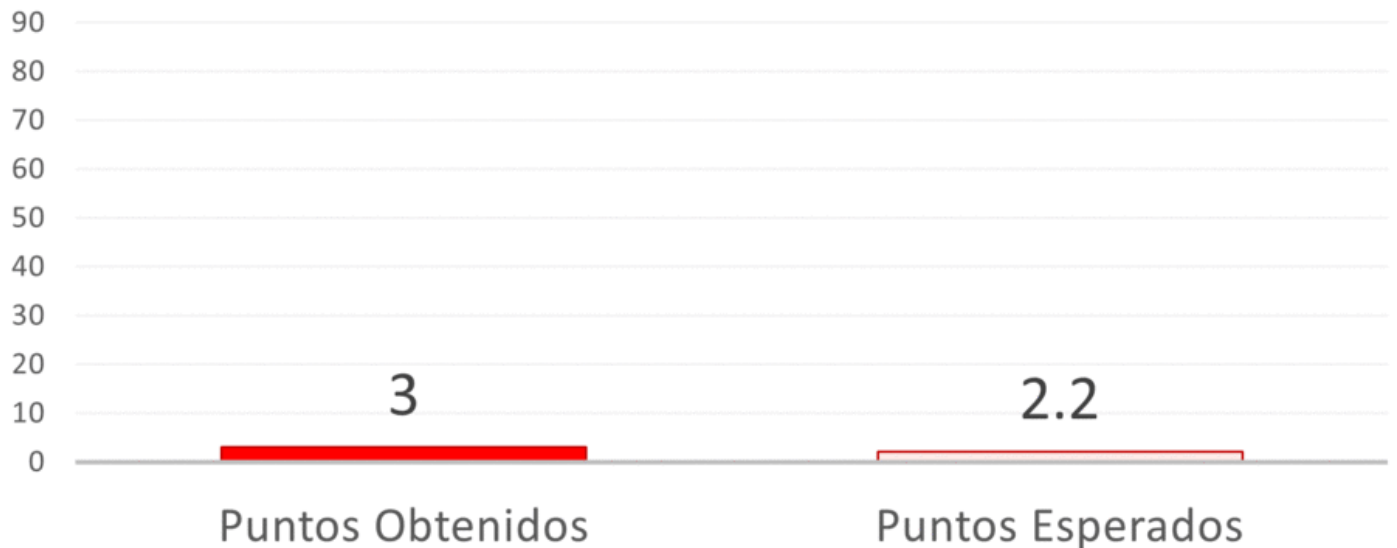
Escenarios	Puntos Esperados de ELE KIPAZO
Victoria 60%	$2 \times 10 = 20$
Victoria 40%	$1.5 \times 10 = 15$
TOTALES	35

Cuando llegase el final del torneo ELE KIPAZO podría medir su desempeño en cuanto a los puntos que obtuvo en el torneo mediante una simple división; por ejemplo, si se esperaba que ELE KIPAZO obtuviera 35 puntos, pero al final del torneo ese equipo alcanza los 42 puntos, su desempeño estaría 20% por arriba de lo esperado ($42 \div 35 = 1.2$). Lo mismo se puede hacer para establecer una referencia de desempeño para cualquier otra estadística (por ejemplo, si ELE KIPAZO hubiera cometido un total de 390 faltas y se esperaba que cometiese 300, tendría 30% más faltas de lo esperado). Por supuesto, este sencillo ejercicio: **medir las variaciones respecto a Lo Esperado**, podemos replicarlo en cualquier liga e, incluso, desde diferentes perspectivas, lo que necesitaríamos sería, primero reunir una cantidad considerable de partidos con sus respectivos datos, para completar nuestra anterior tablita incluyendo más escenarios y los promedios de los datos que se necesite (goles, disparos, faltas, tarjetas, fueros de juego, etcétera).

Registrando los puntos esperados para cada equipo, jornada a jornada podemos comparar el desempeño de los equipos mediante pequeñas sumas y, llegado una cantidad suficiente de partidos (más bien, temporadas completas), empezar a hacer observaciones. Pero mejor veamos algunos ejemplos. Empecemos con el Atlético de Madrid Campeón de **LaLiga en la temporada 2020-2021**, en el que resultaron campeones con la suma de 86 puntos en 38 partidos de liga. ¿Es esto mucho o poco? ¿En qué medida cumplieron con su expectativa?

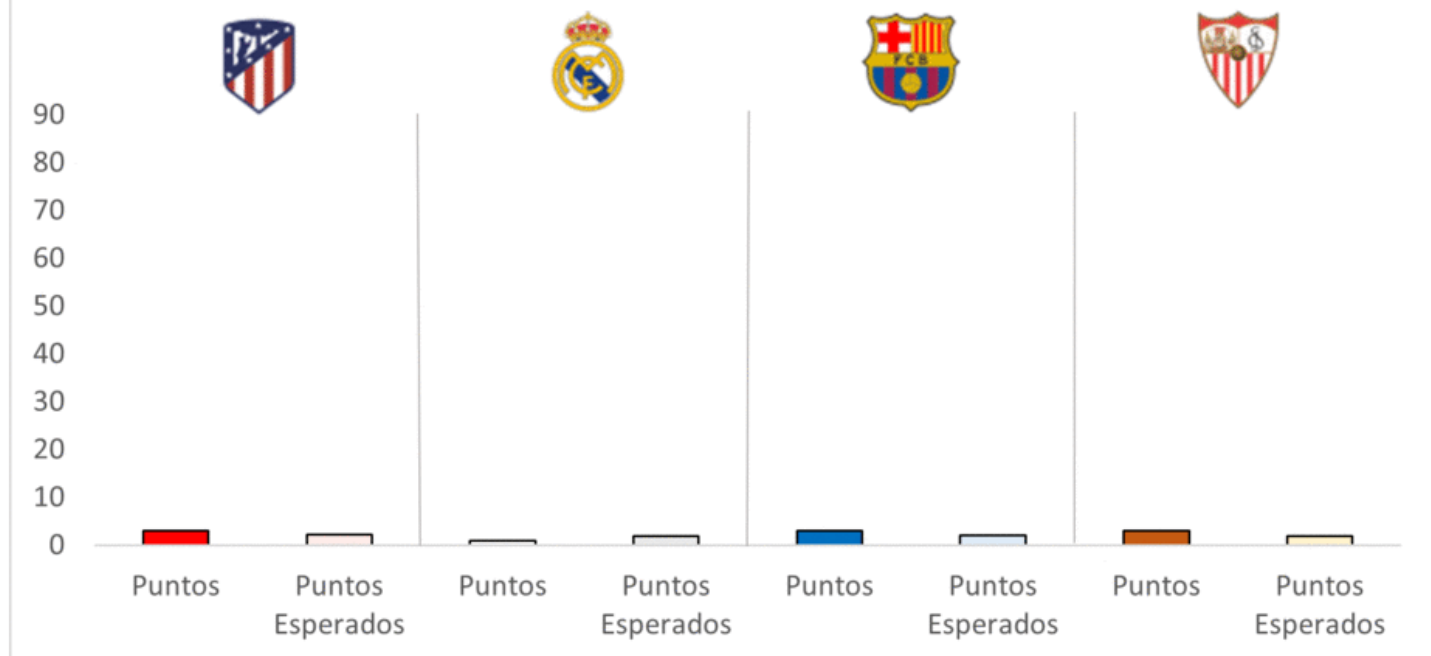


Atlético de Madrid Jornada 1



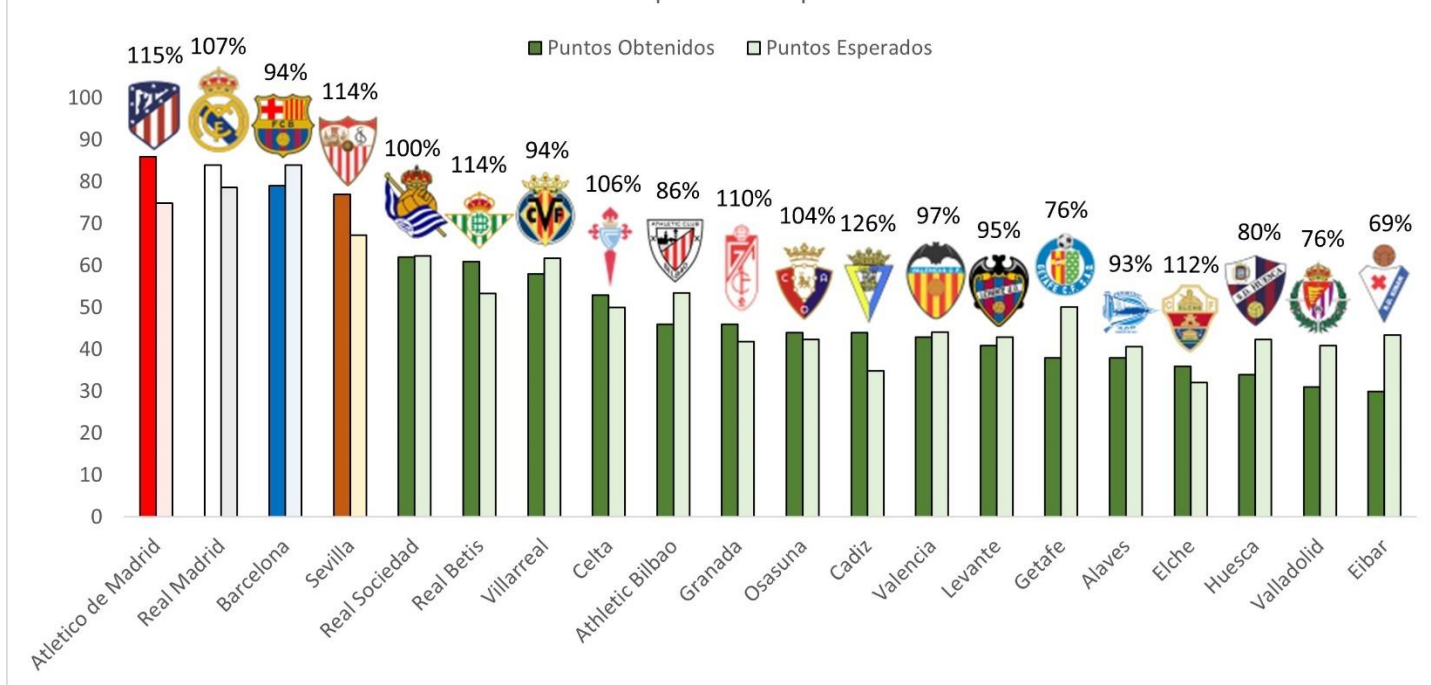
Observando la diferencia entre los Puntos Obtenidos y los Esperados, vemos que el Atlético de Madrid obtuvo 11.1 puntos más de lo que se esperaba, es decir, 15% más de los puntos que se esperaban de ese equipo en ese torneo ($86 \div 74.9 = 1.15$). Sin embargo, si consideramos que, como suele ocurrir en LaLiga, a inicio del campeonato los principales favoritos eran Real Madrid y Barcelona, entonces ¿qué tuvo que pasar para que el Atlético de Madrid ganase LaLiga? ¿solo fue mejor el Atlético de lo esperado o hubo una combinación de variaciones en los desempeños? Veamos la siguiente gráfica del Desempeño en Puntos de los primeros cuatro lugares de la temporada 2020-2021 de LaLiga:

Desempeño en Puntos 2020-2021 Jornada 1



Resulta muy ilustrativo que Sevilla y Atlético de Madrid tuviesen casi el mismo desempeño y, sin embargo, uno fue cuarto lugar y el otro campeón, esto se debe a que **el desempeño depende de lo que se espera de cada equipo en los escenarios a los que se enfrenta**. Veamos la siguiente tabla para LaLiga en la misma temporada 2020-2021 que incluye a todos los equipos.

LaLiga temporada 2020-2021 Desempeño en puntos



Resulta curioso que el equipo con mejor desempeño en el torneo no solo no fue el campeón, sino que estuvo lejos de serlo, Cádiz obtuvo 126% (44 puntos de 35 esperados) alcanzando apenas el 12vo puesto en la tabla general. Mientras que Barcelona, a pesar de tener un desempeño de apenas 94% (79 puntos de 84 esperados), le alcanzó para ser tercer lugar general. Observemos que Getafe, a pesar de tener el segundo peor desempeño del campeonato, logró salvar la categoría.

Las expectativas sobre un equipo dependen del propio equipo (con todos sus actores) y del entorno en el que competirá. La métrica presentada en este capítulo es precisamente eso: una cuantificación de las expectativas y su relación con el entorno. Por ejemplo, solo observando el número de puntos obtenidos por un equipo, ¿en qué medida es sensato exigir al cuerpo técnico de un equipo que obtenga tal o cual cantidad? ¿sería sensato exigir al Osasuna que obtenga 60 puntos, si sus puntos esperados terminan en 43? ¿Podríamos considerar como bueno el torneo de un equipo que solo alcanza 86% de su desempeño, como fue el caso del Athletic de Bilbao? ¿o habrá algún factor añadido, como una desventaja, que influya de manera sostenida en el rendimiento del Athletic de Bilbao? En páginas posteriores pretendo hacer un comentario sobre los equipos con restricciones en sus pools de talento y me basaré en lo que expresa Jared Diamond en su libro *Armas, Gérmenes y Acero*, pero en el caso del deporte le llamaré: El Problema del Volumen.

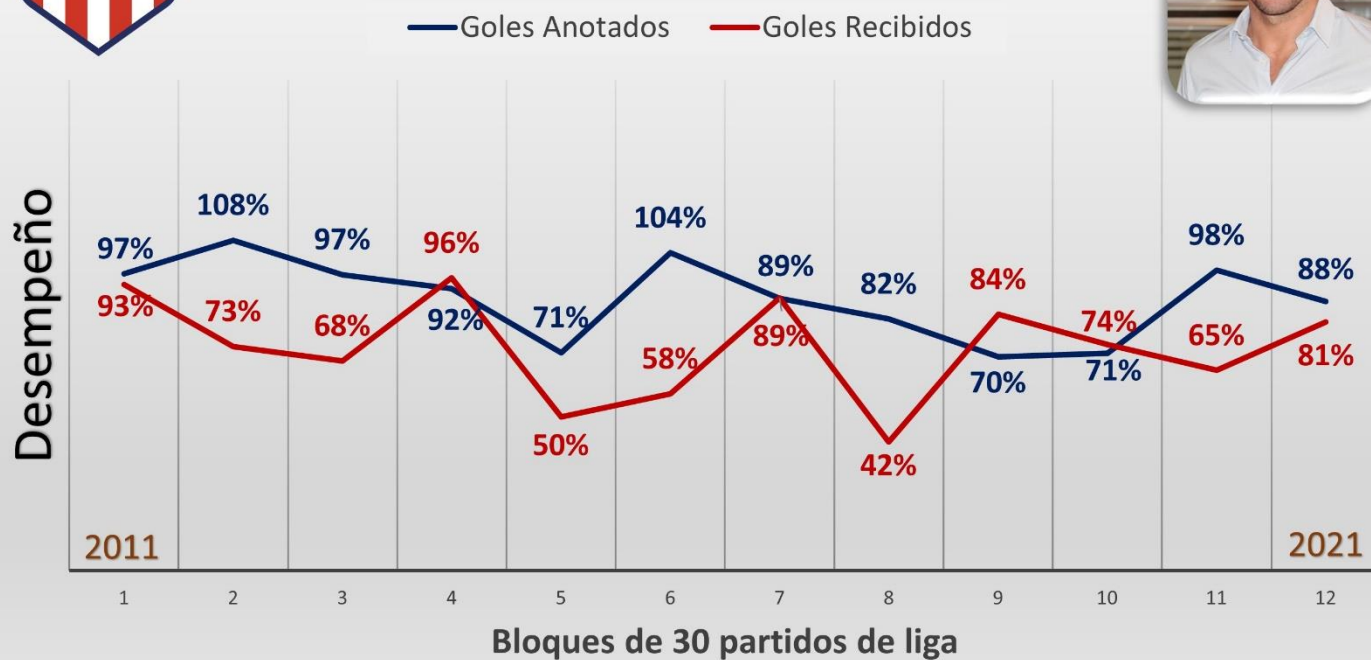
Por otro lado, ¿qué papel juegan los DTs y sus cuerpos técnicos en todo esto? Una pista: los buenos DTs tienen claro el *qué* deben conseguir (es decir, 100% de puntos esperados), pero sí varían en el *cómo* se llega al 100% de los puntos. Veamos un adelanto, en la siguiente gráfica se muestra el desempeño en puntos (es decir, puntos obtenidos entre puntos esperados) del Atlético de Madrid de Diego Simeone:



Diego Simeone ha obtenido en promedio un 106% de los puntos que se esperaban del Atlético de Madrid, pero eso no es sorprendente en sí mismo, porque resulta que la práctica totalidad de los DTs que dirigen, al menos, 60 partidos, se mantienen cercanos al 100%. **Lo que diferencia a los buenos DTs es el cómo consiguen el 100% de sus puntos.** Para dar una idea de lo anterior, veamos la siguiente gráfica que muestra el performance de goles anotados y goles recibidos del equipo dirigido por el Cholo:



Desempeño en Goles Diego Simeone

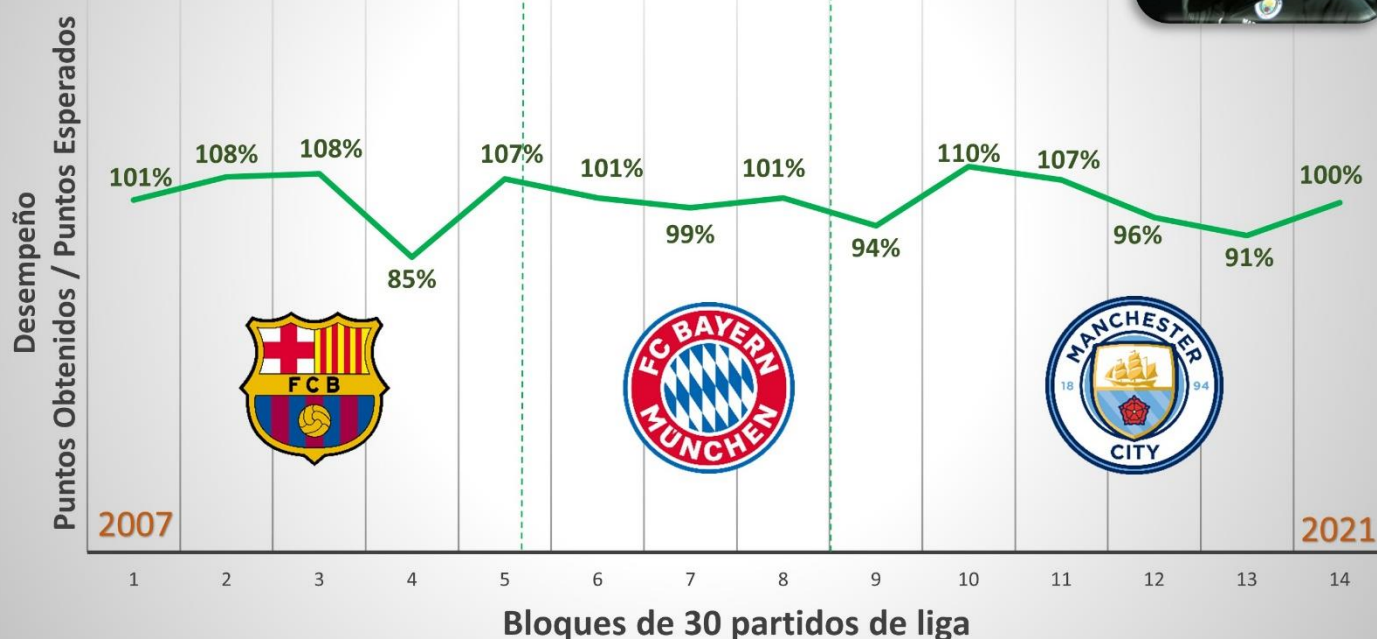


Diego Simeone es un DT que, en el largo plazo, en el entorno en que ha dirigido, tiende a obtener 106% de los Puntos Esperados, anotando 89% de los goles y recibiendo 73% de los goles esperados.

¿Cómo será el desempeño de Josep Guardiola? Demos una mirada:

Desempeño en Puntos

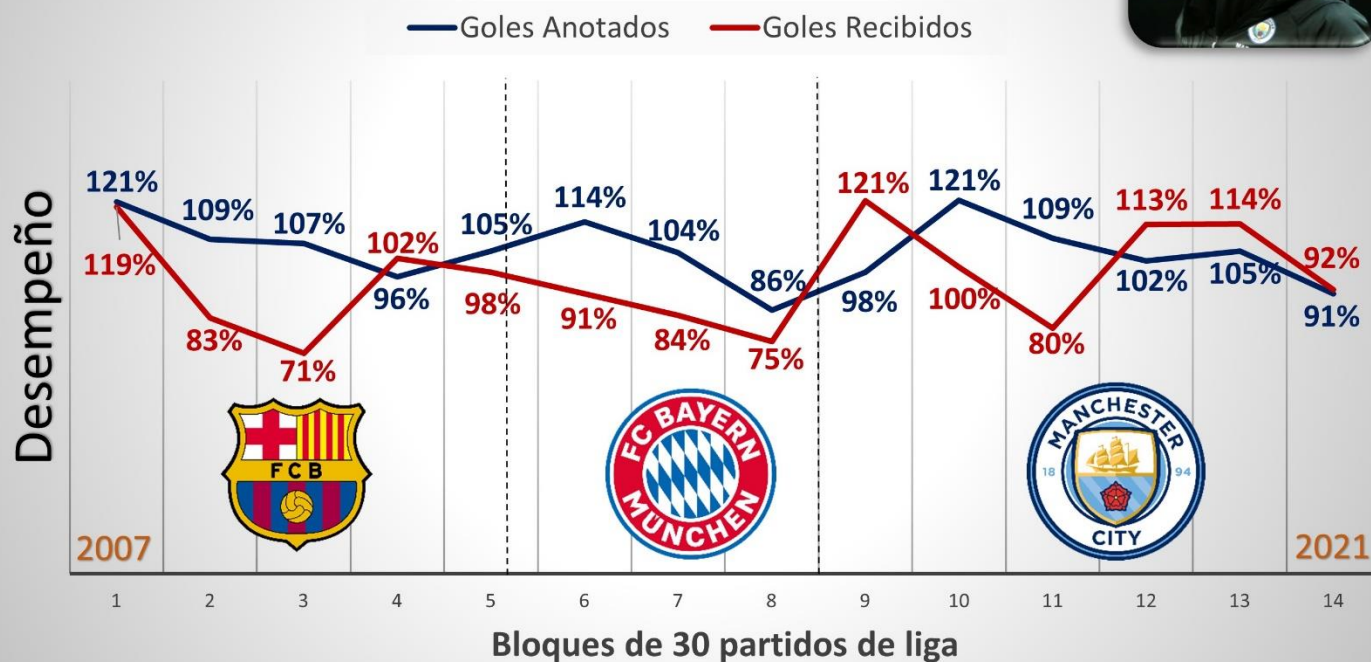
Josep Guardiola



En lo que respecta al desempeño en puntos Guardiola, en general, ha cumplido con las expectativas puestas sobre él en los torneos de liga al obtener un desempeño promedio del 101% a lo largo de su carrera. Pero ¿cómo le habrá ido a Pep en cuanto a goles anotados y recibidos? Veamos:

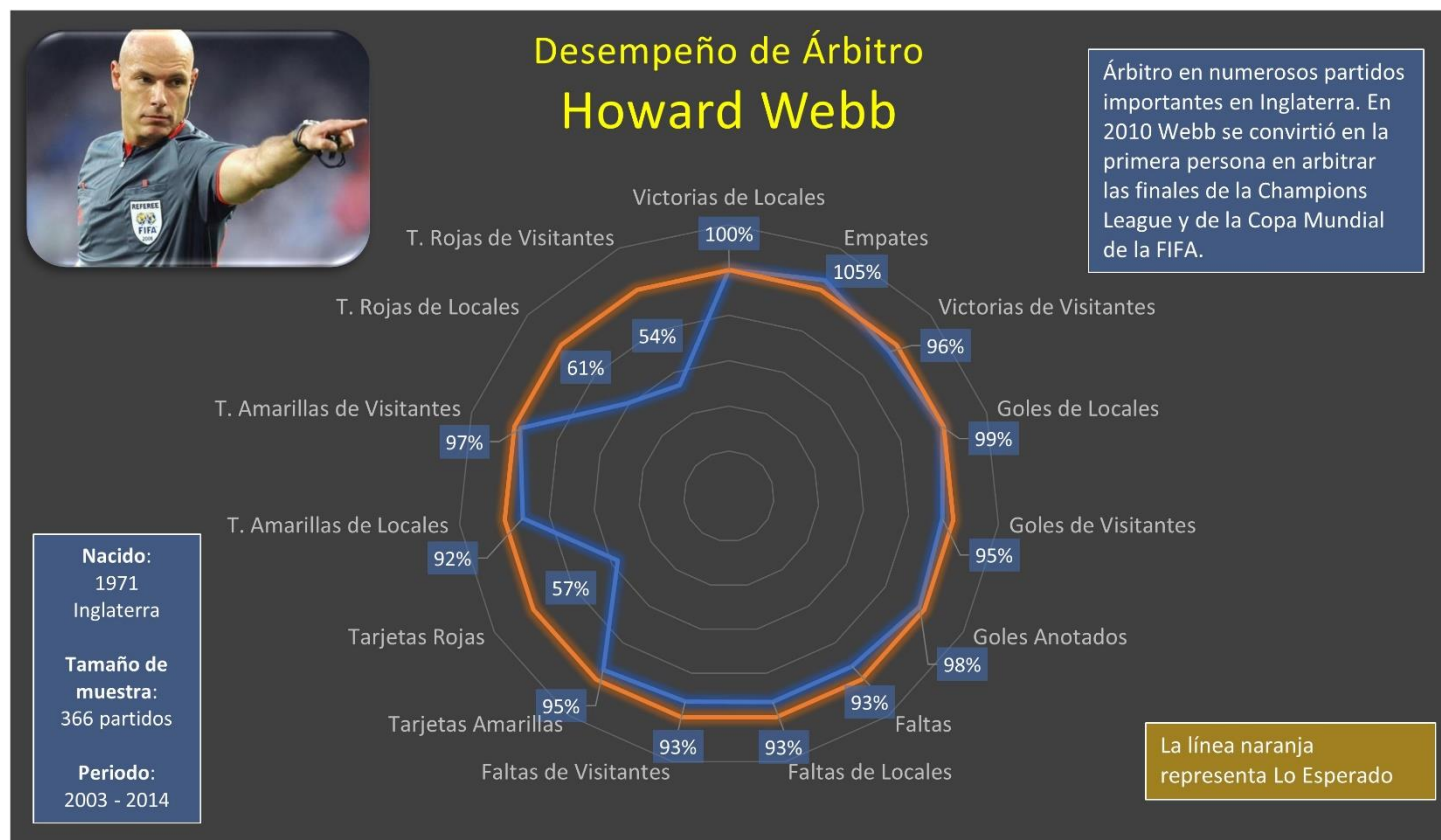
Desempeño en Goles

Josep Guardiola

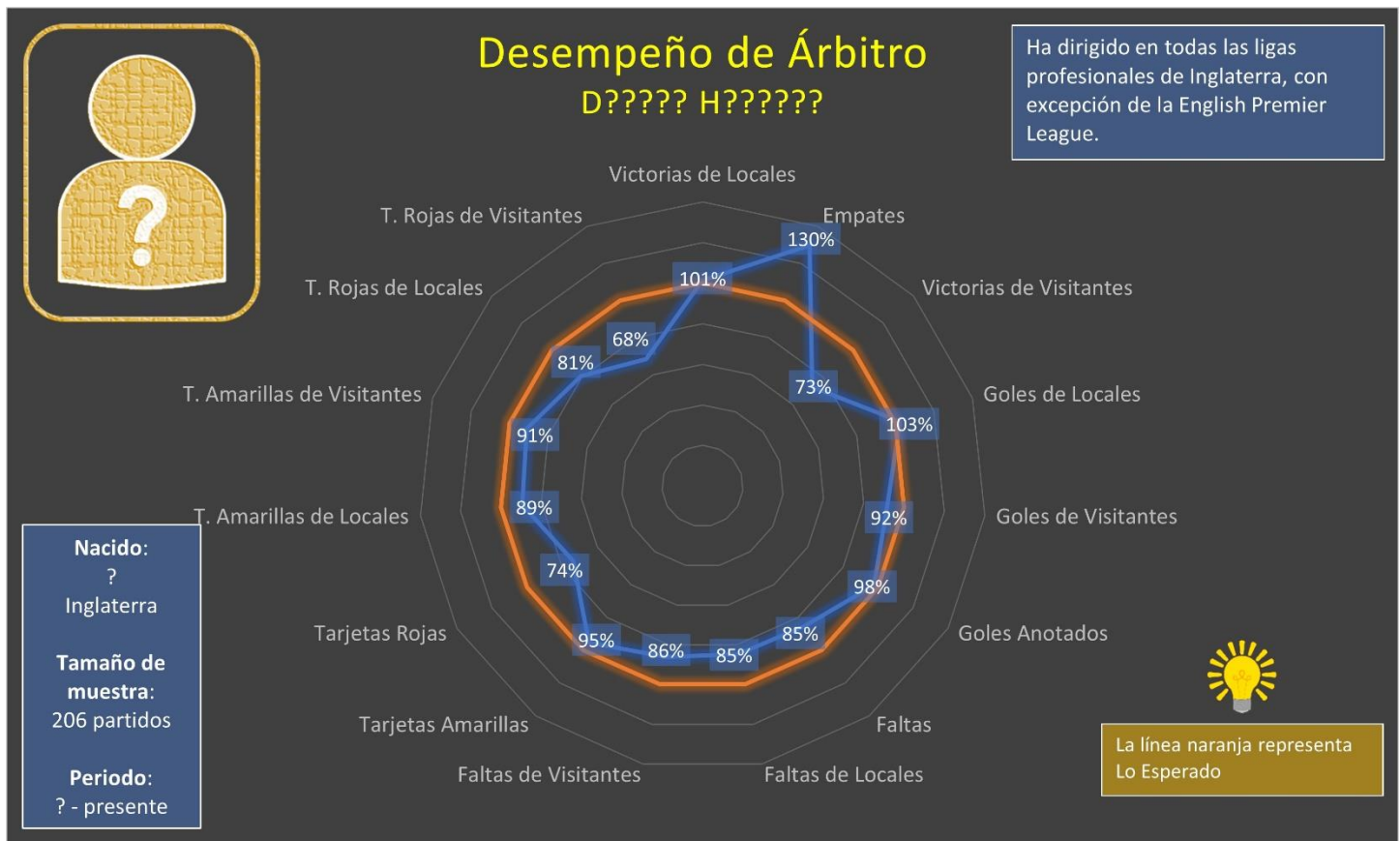


A lo largo de su carrera Pep Guardiola, a lo largo de su carrera, el desempeño en goles anotados es de 105%, mientras que su desempeño en goles recibidos es de 95%.

¿Y los árbitros? ¿se podrá saber en qué medida y en qué aspectos los silbantes tienen injerencia en los partidos? ¿cómo reconocer a los mejores árbitros? Una pista: los mejores son los que pasan más desapercibidos (menor diferencia sobre lo esperado). Pero veamos un ejemplo, con el caso de un histórico árbitro inglés, Howard Webb:



Tal parece que Howard tiene una afición por guardarse tarjetas rojas, aunque no así (tanto) con las amarillas. La mayoría de los árbitros, suelen pasar desapercibidos, es decir, que no se alejan mucho de Lo Esperado, pero aún así hay excepciones notables como la siguiente:



Este árbitro desconocido (hasta ahora), tiene más de una cosa curiosa, pero resalta una: **130% de Empates**. Dicho de otra forma, de entre los 206 partidos dirigidos por este árbitro, se esperaba que hubiese 55 empates, pero resultó que hubo 71. Curioso, ¿no?

¿Se podrá usar esta métrica para construir una estrategia de apuestas? Una pista: sí.

Mucho que explorar en futuros capítulos.

Ing. Jaime G. Meléndez

Twitter: [@JamesGMelendez](https://twitter.com/JamesGMelendez)

Web: <https://academiathi.com/>

LinkedIn: [jaimemelendezthi](https://www.linkedin.com/in/jaimemelendezthi)

Github: <https://github.com/castorx9000>

Síguenos en Enrachados Deportes

Podcast: [Enrachados jueves 21:00 horas](#)

Twitter e Instagram: [@Enrachados_deportes](#)