



ANÁLISIS PREDICTIVO DE LA UEFA CHAMPIONS LEAGUE 2023

TFG BUSINESS ANALYTICS

ÍÑIGO OLALQUIAGA MUÑOZ DE DIOS
E-3 ANALYTICS
Universidad Pontificia de Comillas - ICADE

“Para ganar una Champions hay que tener un buen equipo, un gran entrenador y ganar al Real Madrid”

Franz Beckenbauer (2017)

Índice

1. Introducción	2
1.1 Objetivo del trabajo.....	2
1.2 Historia de la competición.....	2
2. Descripción de la base de datos utilizada	4
3. Participantes: equipos y ligas	10
3.1 Sistema de clasificación	10
3.2 Equipos participantes: datos históricos generales.	10
4. Análisis predictivo y estadístico de cada equipo	20
4.1 Metodología	20
4.2 Análisis predictivo.....	20
5. Análisis predictivo y estadístico de jugadores	47
5.1 Máximo goleador de la competición	47
5.2 Mejor portero de la competición	53
5.3 Mejor defensa de la competición.....	56
6. Conclusiones	60
7. Anexos	62
8. Bibliografía	74

1. Introducción

1.1 Objetivo del trabajo

El objetivo del trabajo es llevar a cabo un análisis predictivo de la competición de fútbol más importante de Europa, la UEFA Champions League. A lo largo de la carrera de Business Analytics, he descubierto que todo lo aprendido no es solo una ciencia teórica, sino que tiene aplicación práctica. Hoy en día, son cada vez más los equipos de fútbol que utilizan la inteligencia artificial y el *Big Data* para tomar decisiones como a que jugador fichar o cómo afrontar un partido.

Además, quiero poner en práctica dos herramientas utilizadas a lo largo de mi grado en ICADE: Python y PowerBi. Estas herramientas no solo son utilizadas en la facultad, sino que son utilizadas por las grandes empresas del mundo a nivel global. Mediante la utilización de ambas, conseguiré predecir y plasmar los datos obtenidos en mi investigación para llegar, así, a una conclusión fundamentada. No obstante, el fútbol, como cualquier otro deporte, es difícil de predecir con exactitud. Quiero aclarar que el estudio que voy a llevar a cabo va a estar basado en pura estadística y probabilidad y la historia nos ha demostrado que no siempre aciertan.

Le daré a los datos una forma, un nombre y un fin, para así sacar conclusiones aproximadas y cercanas a una hipotética realidad. Analizaré a los equipos y a sus jugadores a lo largo de la competición. Realizaré predicciones como qué equipo será el ganador, que jugador será el máximo goleador o que portero será el que menos goles reciba.

En conclusión, intentaré, con los medios de los que dispongo, determinar cuál va a ser el mejor equipo de Europa a nivel de clubes.

1.2 Historia de la competición

La Champions League, conocida anteriormente como la Copa de Europa, nace en el año 1955. En junio de este mismo año, la Unión de Federaciones Europeas de Fútbol (en adelante, UEFA), aprobó organizar una competición entre clubes europeos llamada Copa de Clubes Campeones Europeos. Impulsada por el periódico francés *L'Equipe* y apoyada por Santiago Bernabéu, presidente del Real Madrid en aquel año, se inauguró la competición como un espejo del Campeonato Sudamericano de Campeones. Gabriel Hanot, exfutbolista y periodista francés, fue el ideólogo de lo que conocemos hoy como la UEFA Champions League.

No obstante, esta idea no fue aprobada desde el primer momento. Tras ser rechazada por la UEFA, fueron necesarias largas conversaciones con la Fédération Internationale de Football Association (en adelante, FIFA) y con varios clubes, muchos de ellos en gestación, para que la idea se hiciera realidad. Desde entonces, se han ido introduciendo cambios y la competición ha evolucionado hasta lo que reconocemos hoy como la mejor competición de fútbol europeo a nivel de clubes.

En primera instancia, fueron dieciocho equipos los que apoyaron la iniciativa en el año 1955, representando a dieciocho países diferentes. Estos países son: Suiza, España, Portugal, Yugoslavia, Austria, Países Bajos, Italia, el Protectorado de Sarre, Dinamarca, Francia, Hungría, Bélgica, Suecia, Polonia, Alemania y Escocia. A pesar de la inclusión de grandes países en la competición, grandes clubes históricos se quedaron fuera de la misma por la falta de consolidación de la UEFA. Entre las grandes ausencias destacan las de la Unión Soviética, Checoslovaquia e Inglaterra. Finalmente, por motivos de organización en el seno de la UEFA, fueron dieciséis los participantes de la primera edición de la Copa de Europa en formato de eliminación directa.

Desde 1955 en adelante, el *trono* de la competición ha ido cambiando de manos. El Real Madrid, club que más veces ha ganado la competición, tuvo un inicio apabullante: ganó las cinco primeras ediciones liderado por un jugador histórico, Alfredo Di Stefano. Tras un inicio fulgurante, serían Benfica y los clubes italianos (Milán e Inter de Milán), los encargados de recoger la batuta. Cabe destacar las épocas doradas del Ajax de Ámsterdam, campeones de tres ediciones seguidas entre el 1971 y el 1973, y del Bayern de Múnich, vencedor por partida triple entre los años 1974 y 1976. Desde ese año, comenzaría a florecer con fuerza el fútbol inglés, haciéndose con siete de las diez ediciones siguientes. El Liverpool sería el actor principal en esta década.

En 1992, tras una variedad interesante de campeones durante los años 80, la competición cambia de formato y de nombre a lo que hoy conocemos como la UEFA Champions League o Liga de Campeones de la UEFA. Se eliminan los cuartos de final y las semifinales y se añade la fase de grupos. Esta sería una pequeña liguilla, dividida en dos grupos, en la que los primeros de cada grupo se enfrentarían en la final. En la edición siguiente esto se modificó y se volvieron a introducir los cuartos de final y las semifinales, previas a la gran final. Progresivamente, se iban añadiendo más participantes a la competición hasta llegar, hoy, a treinta y dos.

Es en la temporada 1996-1997, cuando aflora la polémica Ley Bosman. La Ley Bosman permitía a los jugadores extracomunitarios, es decir, de fuera del continente europeo, jugar sin ninguna restricción. Anteriormente, los clubes tenían un número de plazas limitadas para estos jugadores y no podían rebasarlo. Por ende, los clubes con mayor poder adquisitivo se lanzaron a por las estrellas del panorama internacional. Los clubes de menor entidad y con menor poder financiero se vieron relegados a un segundo plano, con unas posibilidades mínimas de hacer frente a los grandes *clubes-empresa*.

Desde entonces, solo equipos de las cuatro grandes ligas (inglesa, italiana, española y alemana) han conseguido proclamarse campeones de la competición, con excepción del equipo portugués Oporto FC en 2004. Chelsea, Liverpool, Bayern, Barcelona o Real Madrid han sido los más laureados durante las últimas dos décadas, destacando a este último, campeón de cinco de las últimas diez ediciones.

2. Descripción de la base de datos utilizada

Las bases de datos utilizadas en mi trabajo han sido extraídas de la página web: <https://fixturedownload.com/sport/football>, o recabadas de la página de la <https://www.uefa.com/> e incluidas en un csv de elaboración propia.

Todos los archivos han sido extraídos en formato .csv para el mejor manejo posterior con la herramienta Python.

En primer lugar, las bases de datos utilizadas para los clubes son:

1. **champions_league_2020_2022:** esta base de datos contiene información relativa a las estadísticas de los clubes durante la Champions League o Liga de Campeones de la UEFA durante las temporadas 2020-2021 y 2021-2022. Esta base de datos está compuesta de 9 variables y de 2192 observaciones. Las variables que utilizaré en mi trabajo serán:

	Nombre de la variable	Descripción	Tipo de variable
1	Match Number	Número de orden del partido en la competición	Numérica
2	Round Number	Número de ronda en la que se disputa ese partido	Numérica
3	Date	Fecha del partido	Numérica
4	Location	Estadio en el que se disputa el partido	Categórica
5	Home Team	Equipo local	Categórica
6	Away Team	Equipo visitante	Categórica
7	Group	Grupo de la competición al que pertenecen ambos equipos	Categórica
8	Home Goals	Goles marcados por el equipo local en ese partido	Numérica

9	Away Goals	Goles marcados por el equipo visitante en ese partido	Numérica
---	------------	---	----------

Tabla 1.1: Variables Clubes I

2. **clean_champions_league_fixtures:** esta base de datos contiene información relativa a los cruces que se van a dar durante la Champions League o Liga de Campeones de la UEFA de la temporada 2022-2023, es decir, la que voy a predecir. Esta base de datos está compuesta de 3 variables y de 375 observaciones. Las variables que utilizaré en mi trabajo serán las siguientes:

	Nombre de la variable	Descripción	Tipo de variable
1	Home	Equipo local	Categórica
2	Away	Equipo visitante	Categórica
3	Year	Año en el que se disputa el partido	Numérica

Tabla 1.2: Variables Clubes II

3. **leagues_2021_2022:** esta base de datos contiene información relativa a las estadísticas de los clubes de las cinco grandes ligas, por partido, durante la temporada pasada (por ello, 2021 y 2022). Esta base de datos está compuesta de 8 variables y de 5.616 observaciones. Las variables que utilizaré en mi trabajo serán las siguientes:

	Nombre de la variable	Descripción	Tipo de variable
1	Match Number	Número de orden del partido en la competición	Numérica
2	Round Number	Número de ronda en la que se disputa ese partido	Numérica

3	Date	Fecha del partido	Numérica
4	Location	Estadio en el que se disputa el partido	Catógica
5	Home Team	Equipo local	Catógica
6	Away Team	Equipo visitante	Catógica
7	Home Goals	Goles marcados por el equipo local en ese partido	Numérica
8	Away Goals	Goles marcados por el equipo visitante en ese partido	Numérica

Tabla 1.3: Variables Clubes III

4. **groups_champions:** esta base de datos contiene información relativa a los grupos de la competición de la edición de este año, 2022-2023. Esta base de datos está compuesta de 10 variables y de 320 observaciones. Las variables que utilizaré en mi trabajo serán las siguientes:

	Nombre de la variable	Descripción	Tipo de variable
1	Pos	Posición en el grupo	Numérica
2	Team	Equipo	Catógica
3	Pld	Partidos jugados	Numérica
4	W	Partidos ganados	Numérica
5	D	Partidos empatados	Numérica
6	L	Partidos perdidos	Numérica
7	GF	Goles marcados por el equipo en la fase de grupos	Numérica
8	GA	Goles recibidos por el equipo en la fase de grupos	Numérica
9	GD	Diferencia de goles entre GF y GA	Numérica
10	Pts	Puntos obtenidos	Numérica

Tabla 1.4: Variables Clubes IV

5. **nondata_leagues_2021_2022**: esta base de datos contiene información relativa a los equipos que no participaron en ninguna de las ediciones pasadas de la Champions League (2020-2021 y 2021-2022) ni aparecen en la base de datos *leagues_2021_2022*. Esta base de datos está compuesta de 4 variables y de 52 observaciones. Las variables que utilizaré en mi trabajo serán las siguientes:

	Nombre de la variable	Descripción	Tipo de variable
1	Team	Equipo	Categórica
2	Goals Scored	Goles anotados	Numérica
3	Goals Conceded	Goles encajados	Numérica
4	TS	<i>Team strength</i>	Numérica

Tabla 1.4: Variables Clubes V

En segundo lugar, las bases de datos utilizadas para los jugadores son:

1. **Player Stats – CL 2021 – 2022**: esta base de datos contiene información relativa a las estadísticas de los jugadores durante la Champions League o Liga de Campeones de la UEFA de la temporada pasada (por ello, 2021 y 2022). Esta base de datos está compuesta de 52 variables y de 1.020 observaciones. Las variables que utilizaré en mi trabajo serán:

	Nombre de la variable	Descripción	Tipo de variable
1	Player	Nombre del jugador participante en la competición.	Categórica
2	Minutes Played	Minutos jugados por el jugador	Numérica
3	Matches Played	Partidos jugados por el jugador	Numérica
4	Goals	Goles	Numérica
5	Assists	Asistencias	Numérica
6	Distance Covered	Distancia cubierta	Numérica
7	Total Goals	Goles a favor	Numérica
8	Right foot	Goles con la pierna derecha	Numérica
9	Left foot	Goles con la pierna izquierda	Numérica
10	Head	Goles con la cabeza	Numérica

11	Other	Goles con una parte distinta al pie/cabeza	Numérica
12	Inside area	Goles dentro del área	Numérica
13	Outside area	Goles fuera del área	Numérica
14	Penalties scored	Goles de penalti	Numérica
15	Matches Played	Partidos jugados por el jugador	Numérica
16	Total Attempts	Número de disparos totales	Numérica
17	On target	Número de disparos a portería	Numérica
18	Off target	Número de disparos fuera de portería	Numérica
19	Blocked	Número de disparos bloqueados por el portero contrario	Numérica
20	Matches Played	Partidos jugados por el jugador	Numérica
21	Passing Accuracy (%)	Porcentaje de acierto en el pase	Numérica
22	Passes Attempted	Pases intentados	Numérica
23	Passes Completed	Pases completados	Numérica
24	Crossing Accuracy (%)	Porcentaje de acierto en los centros	Numérica
25	Crosses attempted	Centros intentados	Numérica
26	Crosses completed	Centros completados	Numérica
27	Free-kicks taken	Faltas lanzadas	Numérica
28	Matches Played	Partidos jugados por el jugador	Numérica
29	Attacks	Ataques del jugador	Numérica
30	Corners taken	Corners sacados	Numérica
31	Offsides	Fueras de juego	Numérica
32	Dribbles	Regates	Numérica
33	Matches Played	Partidos jugados por el jugador	Numérica
34	Balls Recovered	Balones recuperados	Numérica
35	Tackles	Entradas	Numérica
36	Tackles won	Entradas con éxito	Numérica
37	Tackles lost	Entradas perdidas	Numérica
38	Clearences attempted	Despejes de balón intentados	Numérica
39	Matches Played	Partidos jugados por el jugador	Numérica
40	Saves	Paradas del portero	Numérica
41	Goals Conceded	Goles encajados	Numérica
42	Own goals conceded	Goles en propia portería concedidos	Numérica
43	Saves from penalties	Paradas del portero a penaltis	Numérica

44	Clean sheets	Portería a cero	Numérica
45	Punches made	Despejes con los puños del portero	Numérica
46	Matches Played	Partidos jugados por el jugador	Numérica
47	Fouls committed	Faltas cometidas	Numérica
48	Fouls suffered	Faltas sufridas	Numérica
49	Yellow cards	Tarjetas amarillas	Numérica
50	Red cards	Tarjetas rojas	Numérica
51	Minutes Played	Minutos jugados por el jugador	Numérica
52	Matches Played	Partidos jugados por el jugador	Numérica

Tabla 1.5: Variables Jugadores I (Las variables que aparecen en negrita son aquellas utilizadas para hacer el análisis)

- 2. Player Stats por partido - CL 2021 - 2022:** esta base de datos contiene información relativa a las estadísticas de los jugadores, por partido, durante la Champions League o Liga de Campeones de la UEFA de la temporada pasada (por ello, 2021 y 2022). Esta base de datos está compuesta de 51 variables y de 1.020 observaciones. Las variables que utilizaré en mi trabajo serán las mismas que en el cuadro anterior pero divididas entre el número de partidos de juego cada equipo (*Matches Played*).

3. Participantes: equipos y ligas

3.1 Sistema de clasificación

Hoy en día, la Champions League goza de un total de 32 equipos participantes. Este año, en concreto, estos 32 participantes provienen de quince ligas distintas de Europa: inglesa (Premier League), española (Liga Santander), italiana (Serie A), alemana (Bundesliga), francesa (Ligue 1), holandesa (Eredivisie), portuguesa (Primeira Liga), escocesa (Scottish Premiership), belga, austriaca (Bundesliga austriaca), checa, croata, ucraniana, danesa e israelí.

Estos 32 participantes son escogidos según la posición que ocupen a la hora de la consecución de sus ligas domésticas, es decir, las ligas de su país, tal y como indica el artículo 3.01 de la Regulación de la UEFA Champions League¹. No obstante, no todas las ligas aportan los mismos participantes ni tienen acceso a las mismas plazas que otras. La elección acerca de qué ligas aportan más participantes, o menos, ha sido desarrollada en base a la historia e importancia de los clubes en el continente europeo desde su nacimiento.

Las ligas más importantes gozan de una posición mayor en el ranking, medido por un coeficiente. Este coeficiente mide varios parámetros y coloca las ligas en orden descendente. Por ello, ateniéndonos al ranking, gozan de 4 participantes con clasificación directa las siguientes ligas: inglesa, española, italiana y alemana. Serán Portugal y Francia las que aporten dos participantes directos más a la competición. Finalmente, aportando un solo participante directo, tenemos a los equipos vencedores de la liga en Holanda, Bélgica (tras la eliminación de Rusia por motivos bélicos), Austria y Escocia.

La suma de los participantes que clasifican directamente es de veinticuatro equipos. Por ende, son ocho los equipos que quedarían por clasificar para completar el cuadro. Los terceros de las ligas que aportan dos participantes directos, los segundos de aquellas que aportan únicamente un participante y los ganadores de aquellas ligas que no gozan de participación directa, participarán en unas eliminatorias denominadas *play-offs* para determinar las últimas ocho plazas.

3.2 Equipos participantes: datos históricos generales.

En esta edición de la Champions League 2022/2023, son las grandes ligas las que vuelven a mandar en el número de participantes. Los participantes se dividen en países de la siguiente forma:

¹ <https://documents.uefa.com/r/Regulations-of-the-UEFA-Champions-League-2022/23/Article-3-Entries-for-the-competition-Online>

- I. **Liga inglesa:** Manchester City, Liverpool, Chelsea y Tottenham Hotspur.
- II. **Liga española:** Real Madrid, FC Barcelona, Atlético de Madrid y Sevilla.
- III. **Liga italiana:** AC Milán, Inter de Milán, Juventus de Turín y Nápoles.
- IV. **Liga alemana:** Bayern de Múnich, Borussia Dortmund, Leipzig, Bayer Leverkusen y Eintracht Frankfurt².
- V. **Liga francesa:** Paris Saint-Germain y Marseille.
- VI. **Liga portuguesa:** Benfica, Oporto y Sporting de Portugal.
- VII. **Liga holandesa:** Ajax de Ámsterdam.
- VIII. **Liga escocesa:** Celtic y Rangers.
- IX. **Liga belga:** Club Brujas.
- X. **Liga checa:** Viktoria Plzen.
- XI. **Liga austriaca:** Salzburgo.
- XII. **Liga croata:** Dinamo de Zagreb.
- XIII. **Liga ucraniana:** Shakhtar Donetsk.
- XIV. **Liga danesa:** Copenhagen.
- XV. **Liga israelí:** Maccabi Haifa.

2022/23 UEFA Champions League

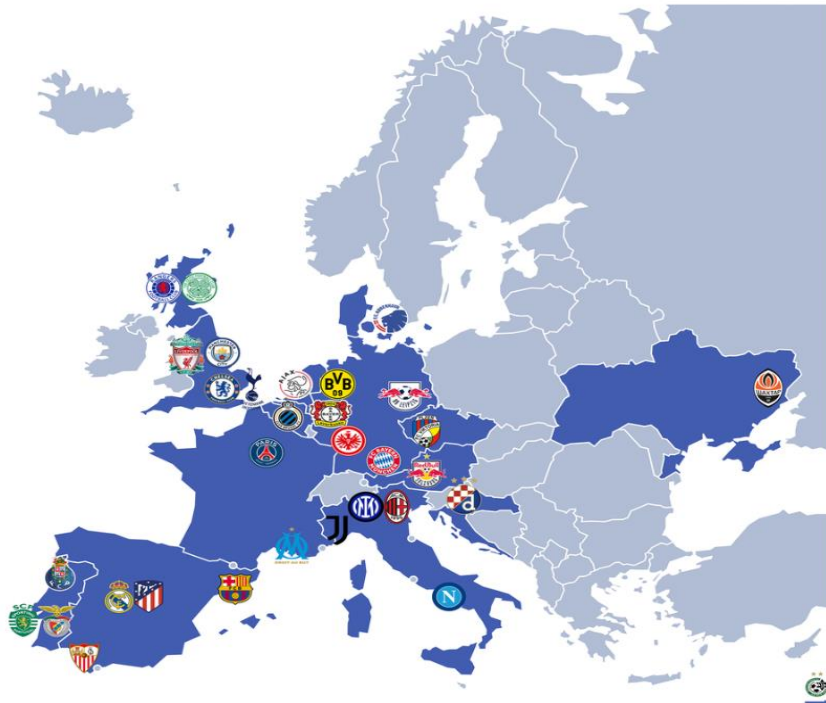


Figura 1: Mapa de equipos por país.

² Es el quinto clasificado de Alemania debido a su victoria en la UEFA Europa League, la segunda competición europea más importante a nivel de clubes.

I. Liga Inglesa.

Manchester City

El club inglés fue el primer clasificado de la liga inglesa (Premier League) en la temporada pasada, la temporada 2021 - 2022. Ha participado en trece ediciones de la UEFA Champions League y se encuentra en la treintava posición en el ranking histórico de la competición. Nunca ha conseguido el título de la UEFA Champions League.

Su mejor participación fue alcanzar la final en la edición de la UEFA Champions League 2020 – 2021, en la que perdió con el Chelsea³.

Liverpool

El club inglés es el más laureado de su país en la competición, con seis campeonatos en sus vitrinas. Ha participado en veintisiete ediciones de la UEFA Champions League y se encuentra en la sexta posición en el ranking histórico de la competición.

Fue subcampeón de la edición pasada de la UEFA Champions League, perdiendo la final contra el Real Madrid. Su último trofeo fue en la edición 2018-2019⁴.

Chelsea

El club inglés ha conseguido ganar la competición en dos ocasiones. Ha participado en 19 ediciones de la UEFA Champions League y se encuentra en la doceava posición en el ranking histórico de la competición.

Fue campeón de la edición de la UEFA Champions League 2020 – 2021, en la que derrotó al Manchester City⁵.

Tottenham Hotspur

El club inglés nunca ha ganado una UEFA Champions League. Ha participado en siete ediciones de la misma y se encuentra en la posición número sesenta y cuatro del ranking histórico de la competición.

La final de la temporada 2018-2019 fue su mejor resultado en la competición, en la que perdió contra el Liverpool⁶.

³ Ver Anexo 1

⁴ Ver Anexo 2

⁵ Ver Anexo 3

⁶ Ver Anexo 4

II. Liga Española

Real Madrid

El club español es el más laureado de todos los tiempos en la competición. Ha conseguido un total de catorce títulos y suma cincuenta y tres participaciones en la UEFA Champions League. Además, ocupa el primer lugar del ranking histórico de la competición.

El último título logrado fue conseguido en la edición pasada, la UEFA Champions League 2021 – 2022, en la que derrotaron al Liverpool⁷.

Barcelona

El club español es el segundo más laureado del fútbol español. Ha conseguido un total de cinco títulos de la UEFA Champions League. Ha participado en treinta y tres ediciones de la UEFA Champions League y se encuentra en la tercera posición en el ranking histórico de la competición.

Su último título fue conseguido en la edición de la UEFA Champions League 2014 – 2015, en la que derrotaron a la Juventus en la final⁸.

Atlético de Madrid

El club español suma un total de dieciocho participaciones en la competición y se encuentra en la dieciseisava posición del ranking histórico de la competición.

Su mejor participación fue el alcance de las dos finales en las temporadas 2013 – 2014 y 2015 – 2016⁹.

Sevilla

El club español suma un total de diez participaciones en la competición y se encuentra en la posición número cincuenta y ocho del ranking histórico de la competición.

Su mejor participación en la competición fue la clasificación a cuartos de final, en la edición de la UEFA Champions League 2017 – 2018.

⁷ Ver Anexo 5

⁸ Ver Anexo 6

⁹ Ver Anexo 7

III. Liga Italiana

AC Milan

El club italiano fue el primer clasificado de la liga italiana en la temporada 2021-2022. Es el segundo club con más títulos en la competición, con 7. Ha participado en treinta ediciones de la UEFA Champions League y ocupa la séptima posición en el ranking histórico de la competición. Su último título fue conseguido en la edición de la UEFA Champions League 2006 – 2007, en la que vencieron al Liverpool en la final¹⁰.

Inter de Milán

El club italiano ha ganado la competición en tres ocasiones. Ha participado en veinticuatro ediciones de la competición y ocupa la decimoquinta posición en el ranking histórico de la competición.

Su último título fue conseguido en la edición de la UEFA Champions League 2009 – 2010, en la que vencieron al Bayern de Múnich en la final¹¹.

Juventus

El club italiano tiene dos títulos en la competición europea y ha llegado a dos finales en los últimos diez años. Ha participado en treinta y siete ediciones de la competición y ocupa la quinta posición en el ranking histórico de la misma.

Su último título fue conseguido en la edición de la UEFA Champions League 1995-1996, en la que vencieron al Ajax de Ámsterdam¹².

Nápoles

El club italiano suma un total de diez participaciones en la competición y se encuentra en la posición número sesenta y tres del ranking histórico de la competición.

Su mejor participación fue el alcance de las semifinales¹³.

¹⁰ Ver Anexo 9

¹¹ Ver Anexo 10

¹² Ver Anexo 11

¹³ Ver Anexo 12

IV. Liga Alemana

Bayern de Múnich

El club alemán ha obtenido seis títulos a lo largo de la historia, además de sumar una innumerable cantidad de finales. Ha participado en treinta y nueve ediciones de la UEFA Champions League y ocupa la segunda posición en el ranking histórico de la competición.

Su último título fue conseguido en la edición de la UEFA Champions League 2019 – 2020, en la que se impusieron al Paris Saint – Germain en la final¹⁴.

Borussia Dortmund

El club alemán suma un título en la UEFA Champions League. Cuenta con veintiún participaciones en la competición y ocupa el lugar número diecisiete en el ranking histórico de la misma.

El título fue logrado en la edición de la UEFA Champions League 1996 – 1997, en la que se impusieron a la Juventus¹⁵.

Leipzig

El club alemán es el más joven de la competición. Creado en 2009, guarda como mejor participación las semifinales alcanzadas en la temporada 2019-2020. Ha participado en cinco ediciones de la UEFA Champions League y ocupa la posición número noventa y cinco en el ranking histórico de la competición.

Creado en 2009, guarda como mejor participación las semifinales alcanzadas en la temporada 2019-2020¹⁶.

Bayer Leverkusen

El club alemán logró clasificarse como cuarto en su liga doméstica. Ha participado en trece ediciones de la UEFA Champions League y ocupa la posición número treinta y siete del ranking histórico de la competición.

Su mejor actuación data en la temporada 2001-2002, en la que fue derrotado en la final por el Real Madrid¹⁷.

¹⁴ Ver Anexo 13

¹⁵ Ver Anexo 14

¹⁶ Ver Anexo 15

¹⁷ Ver Anexo 16

Eintracht Frankfurt

El club alemán se proclamó campeón de la edición pasada de la UEFA Europa League (el segundo campeonato europeo más importante a nivel de clubes). Por ello, se clasificó como el quinto equipo alemán a la máxima competición europea. Ha participado en dos ediciones de la UEFA Champions League y ocupa el puesto ciento cincuenta y cinco en el ranking histórico de la competición.

Su mejor participación fue conseguida en la primera edición de la UEFA Champions League, en la que perdió la final con el Real Madrid¹⁸.

V. Liga Francesa

Paris Saint-Germain

El club francés fue el campeón de su liga doméstica la temporada 2021-2022. Ha participado en dieciséis ediciones de la UEFA Champions League y ocupa el veintavo puesto del ranking histórico de la competición.

Su mejor participación fue el alcance de la gran final de la UEFA Champions League en la temporada 2019-2020, en la que perdió con el Bayern de Múnich¹⁹.

Marseille

El club francés tiene un título de la competición en sus vitrinas. Ha participado en dieciséis ocasiones en la competición y ocupa la posición número treinta y ocho del ranking histórico de la competición.

El título fue conseguido en la edición de la UEFA Champions League 1992-1993, en la que derrotaron al Milán²⁰.

¹⁸ Ver Anexo 17

¹⁹ Ver Anexo 18

²⁰ Ver Anexo 19

VI. Liga Portuguesa

Benfica

El club portugués tiene dos títulos en sus vitrinas. Ha participado en cuarenta y dos ediciones de la UEFA Champions League y ocupa la octava posición en el ranking histórico de la competición.

Su último título fue conseguido en la edición de la UEFA Champions League 1961 – 1962²¹.

Oporto

El club portugués suma dos títulos en la UEFA Champions League. Ha participado en treinta y siete ediciones de la UEFA Champions League y ocupa el noveno puesto en el ranking histórico de la competición.

Su último título fue conseguido en la edición de la UEFA Champions League 2003 – 2004, en la que derrotó al Mónaco en la final²².

Sporting de Portugal

El club portugués suma veintitrés participaciones en la competición. Se clasificó para la edición de esta UEFA Champions League 2022-2023 vía *play-offs*. Ocupa el puesto número cincuenta y tres del ranking histórico de la competición²³.

VII. Liga Holandesa

Ajax de Ámsterdam

El club holandés ha sido campeón de la UEFA Champions League en 4 ocasiones, consiguiendo tres títulos de manera consecutiva. Ha participado en treinta y nueve ediciones de la UEFA Champions League y ocupa el décimo puesto del ranking histórico de la competición.

Su último título fue conseguido en la edición de la UEFA Champions League 1994 – 1995, en la que derrotaron al Milán²⁴.

²¹ Ver Anexo 20

²² Ver Anexo 21

²³ Ver Anexo 22

²⁴ Ver Anexo 23

VIII. Liga Escocesa

Celtic

El club escocés se proclamó campeón de la competición en una ocasión. Desde entonces, sus participaciones no han sido destacadas. Ha participado en treinta y siete ediciones de la UEFA Champions League y ocupa el catorceavo lugar en el ranking histórico de la competición.

El título obtenido fue conseguido en la edición de la UEFA Champions League 1966 – 1967, en la que derrotaron al Inter de Milán²⁵.

Rangers

El club escocés es, tras el Celtic, el equipo más laureado de Escocia. Ha participado en la competición en un total de treinta y dos ediciones y ocupa el puesto número veintidós en el ranking histórico de la competición²⁶.

IX. Liga Belga

Club Brujas

El club belga es el más laureado de la historia en su país. Ha participado en veintitrés ediciones de la UEFA Champions League y ocupa el puesto número treinta y seis del ranking histórico de la competición.

Su mejor participación ha sido la clasificación a los octavos de final de la competición²⁷.

X. Liga Checa

Viktoria Plzen

El club checo ha participado nueve veces en la competición y, actualmente, ocupa el puesto número setenta y ocho del ranking histórico de la competición.

Su mejor participación ha sido la clasificación a la fase de grupos²⁸.

²⁵ Ver Anexo 24

²⁶ Ver Anexo 25

²⁷ Ver Anexo 26

²⁸ Ver Anexo 27

XI. Liga Austriaca

Salzburgo

El club austriaco ha participado dieciocho veces en la competición y, actualmente, ocupa el puesto número cuarenta y cinco del ranking histórico de la competición²⁹.

XII. Liga Croata

Dinamo de Zagreb

El club croata ha participado en veinticuatro ocasiones en la competición, logrando meterse entre los treinta primeros del ranking histórico de la competición, en la posición número veinticinco³⁰.

XIII. Liga Ucraniana

Shakhtar Donetsk

El club ucraniano ha participado en veintitrés ocasiones en la competición, logrando colocarse en la posición número veintisiete del ranking histórico de la competición³¹.

XIV. Liga Danesa

Copenhagen

El club danés ha participado en dieciséis ediciones de la competición y ocupa el puesto número cuarenta y tres del ranking histórico de la competición³².

XV. Liga Israelí

Maccabi Haifa

El club israelí ha participado en diez ocasiones en la máxima competición europea de clubes de la historia y ocupa el puesto número ochenta y cinco del ranking histórico de la competición³³.

²⁹ Ver Anexo 28

³⁰ Ver Anexo 29

³¹ Ver Anexo 30

³² Ver Anexo 31

³³ Ver Anexo 32

4. Análisis predictivo y estadístico de cada equipo

4.1 Metodología

Para llevar a cabo el análisis predictivo de la UEFA Champions League 2022-2023, he utilizado dos aplicaciones fundamentales para el cálculo y la visualización de los datos: Python y PowerBi. He utilizado las dos aplicaciones a lo largo de la carrera, por lo que estoy familiarizado con las mismas. En primer lugar, con Python, he llevado a cabo el análisis predictivo en sí mismo. Esto es, el análisis de probabilidades, de estadísticas y de datos puros. La plataforma permite calcular medias, medianas, percentiles y todo tipo de variables estadísticas necesarias para realizar una predicción. Mediante la utilización de las fórmulas aprendidas en el grado de Business Analytics, he podido darles valor a los datos contenidos en el Excel. Los ficheros Excel han sido extraídos de una página web o son de elaboración propia, tras haber recabado datos de infinitas fuentes tanto de la UEFA Champions League como de las ligas domésticas de cada país en el que haya un equipo participante.

En segundo lugar, he utilizado PowerBi para la visualización de los datos y de las predicciones hechas. De nada sirve predecir y averiguar estadísticas, si estas no son plasmadas en gráficos, tablas o mapas que nos permitan apreciarlas. PowerBi es una herramienta de Microsoft 365 que permite la creación de distintos tipos de gráficos, así como la fácil interpretación de los mismos.

Para la generación de la base de datos, he utilizado las páginas web oficiales, tanto de las ligas de las que son miembros los equipos participantes, como la de la UEFA (encargada de la Champions League). En ella he incluido datos tanto de los jugadores, a nivel individual, como de los equipos, a nivel colectivo. Para la generación de la misma he utilizado Excel, de Microsoft 365, ya que Python y PowerBi funcionan excelentemente con un archivo de estas características. Las bases de datos que han sido extraídas de una web, las he sacado de la web *Fixture Download*, una página web que contiene una gran cantidad de datos de competiciones deportivas, en especial de fútbol.

4.2 Análisis predictivo

La UEFA Champions League es una competición que consta de dos partes: fase de grupos y fase eliminatoria. En la fase de grupos, se dividen los 32 equipos en 8 grupos de 4 equipos. Todos los equipos del mismo grupo se enfrentan dos veces: ida (en la ciudad del equipo en cuestión) y vuelta (en la ciudad del equipo al que se enfrentan). Por ello, cada equipo juega un total de 6 partidos en esta primera fase. Los dos primeros clasificados de un grupo se clasificarán para la siguiente fase eliminatoria. En cambio, los dos últimos clasificados del grupo quedarán eliminados de la competición. Los grupos de este año son los siguientes:



Figura 2.1: Grupos de la competición (UEFA Champions League)

A) FASE DE GRUPOS

En primer lugar, he simulado los 12 partidos que se juegan en un grupo de la UEFA Champions League para determinar quiénes son el primer y el segundo clasificado, es decir, los que se clasifican para la siguiente ronda. Esta siguiente ronda es la de octavos de final, a la que solo acceden la mitad de los equipos (16). Como en la mayoría de las competiciones futbolísticas, el ganador de un partido obtiene 3 puntos, aquel que empate obtiene 1 punto y el que pierde el partido obtendrá 0 puntos.

Para ello, he empezado mi código en Python importando las bibliotecas que vaya a necesitar. Estas bibliotecas me han dado acceso a las funciones y herramientas que he utilizado posteriormente. Concretamente, he importado las bibliotecas *pandas*, *pickle* y *poisson*, de la biblioteca *spicy.stats*.

Importando *pandas*, puedo estructurar mis datos y permitir que Python lea el *.csv* que voy a utilizar. *Pandas* proporciona estructuras de datos eficientes y flexibles para trabajar con datos tabulares, como hojas de cálculo y bases de datos. Algunas de las funcionalidades que ofrece *Pandas* son:

- La capacidad de leer y escribir datos en diferentes formatos de archivo, como CSV, Excel, SQL, etc.
- La capacidad de manipular y transformar datos, incluyendo la selección, filtrado, agregación y combinación de datos.

La principal utilidad de la biblioteca *pickle* es que permite a los programadores de Python guardar objetos complejos, como diccionarios, listas y clases, en un archivo y luego recuperarlos más tarde en su forma original.

Por último, con *poisson*, importo la función estadística para calcular las probabilidades de los equipos para conseguir los 3 puntos. Esta función permite calcular la probabilidad de que ocurra un número específico de eventos en un intervalo de tiempo o espacio, dada una tasa promedio conocida de ocurrencia de eventos.³⁴

Tras incorporar las bibliotecas, he cargado los archivos *.csv* con los que he trabajado. Estos archivos, descritos anteriormente en el apartado “*Descripción de las bases de datos utilizadas*”, han sido extraídos de una web con información relevante de la competición.

Mediante el uso de *pandas*, he cargado los archivos **champions_league_2020_2022**, **groups_champions** y **clean_champions_league_fixtures**. Estos archivos contienen información de los goles marcados y recibidos durante las ediciones pasadas de la competición, los grupos de esta y los cruces que hay en una edición de la UEFA Champions League.³⁵

Una vez cargados los archivos, he utilizado el csv **champions_league_2020_2022** para iniciar el análisis con esos datos. Esta base de datos contiene aquellos relativos a la competición durante las temporadas 2020-2021 y 2021-2022. Estos datos han sido los primeros que he utilizado para realizar la simulación. Más adelante, en cambio, he utilizado los datos de las ligas para ver como variaría, haciendo una media de ambos conjuntos de datos.

El *team strength* ha sido la variable que he utilizado para determinar porque un equipo es más fuerte que otro, es decir, porque un equipo tiene más probabilidades de ganar a su contrincante. Este dato lo he extraído haciendo la diferencia entre las medias de goles anotados y de goles concedidos. Para ello, no he utilizado todas las variables del csv **champions_league_2020_2022**, solo “*Home Team*”, “*Away Team*”, “*Home Goals*” y “*Away Goals*”. He dividido los datos según sean del equipo local o del equipo visitante. Para ello, dentro de la base de datos he creado dos conjuntos que contengan solo estos datos denominados *df_home* y *df_away*.³⁶

³⁴ Ver Anexo 33

³⁵ Ver Anexo 34

³⁶ Ver Anexo 35

Tras definir *df_home* y *df_away*, he renombrado las columnas “Home Goals” y “Away Goals” de *df_home* como “GoalsScored” y “GoalsConceded” y las columnas “Home Goals” y “Away Goals” de *df_away* como “GoalsConceded” y “GoalsScored”.³⁷

Por último, he sacado el *team strength* de los equipos realizando las medias de “GoalsScored” y “GoalsConceded”, ordenándolas por equipo³⁸ (*Tstrength en el siguiente gráfico*):

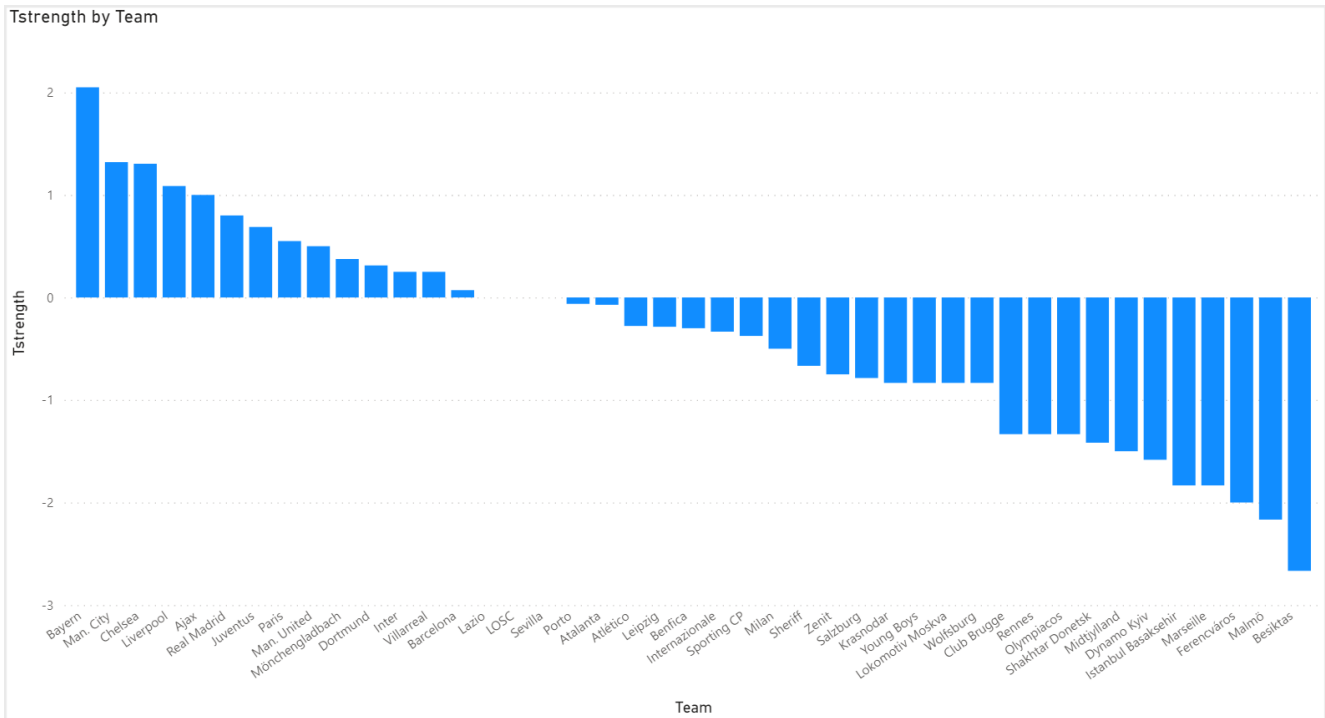


Figura 2.2: Team Strength Champions League 2020-2022 (Elaborado con PowerBi).

En el gráfico aparecen muchos de los equipos que participan en la competición, pero hay muchos otros que no participan en esta edición. He hecho un filtrado de aquellos que sí que participan. Además, muchos de los que participan en esta edición, no aparecen en la base de datos anterior. Por ello, de esta base de datos he utilizado solo los equipos que sí aparecen y de los que si se pueden sacar conclusiones. Respecto a los equipos **que aparecen y no participan** no he tenido nada en cuenta.

Una vez cargados los documentos y habiendo seleccionado las variables que me han interesado para el cálculo del *team strength*, he elaborado la función con la que voy a simular los partidos de la competición.

³⁷ Ver Anexo 36

³⁸ Ver Anexo 37

Predict points

Para el cálculo de los puntos de cada equipo en la fase de grupos, he elaborado una función denominada **predict_points**³⁹. Como indica su nombre, me ha permitido calcular los puntos de un equipo por partido. La función **predict_points** coge los datos de las columnas extraídas de la base de datos utilizada para calcular la probabilidad que tiene un equipo de ganar a otro. La probabilidad se calcula con *poisson*. La distribución de Poisson es una distribución de probabilidad que se utiliza para modelar la ocurrencia de eventos raros en un intervalo de tiempo o espacio. Esta distribución describe la probabilidad de que un número específico de eventos ocurra en un intervalo dado, dada una tasa promedio conocida de ocurrencia de eventos. La distribución de Poisson se utiliza comúnmente en estadísticas, física, ingeniería, y otras áreas donde es importante modelar eventos raros.

La función *predict_points* tiene una entrada de dos datos: equipo local (home) y equipo visitante (away). Si alguno de los equipos que se escriben en la entrada de datos no aparece en la misma, la función devolverá dos ceros (0,0). En cambio, si ambos equipos aparecen en la base de datos, la función funcionará.

Primero, verifica si los nombres de los equipos están en el DataFrame *df_team_strength*, que suponemos contiene las estadísticas de fortaleza de los equipos en términos de goles marcados y recibidos. Si ambos equipos están en el DataFrame, se calcula la tasa de goles esperados para cada equipo, utilizando las estadísticas de goles marcados y recibidos del otro equipo.

A continuación, utiliza un bucle *for* para calcular la probabilidad de que cada equipo marque cierto número de goles en el partido. La probabilidad se calcula utilizando la función *poisson.pmf* () de la biblioteca *scipy.stats*.

Finalmente, se calcula el puntaje esperado para cada equipo en el partido sumando la probabilidad de ganar (3 puntos) y la probabilidad de empate (1 punto) ponderadas por la probabilidad de que ocurra cada resultado. La función devuelve dos valores entre 0 y 3, que hacen referencia a cómo se repartirían los tres puntos del partido según el *team strength* de cada uno. Por ejemplo, en el partido Real Madrid – Leipzig⁴⁰.

³⁹ Ver Anexo 38

⁴⁰ Ver Anexo 39

Esto quiere decir, que el Real Madrid se llevaría 2,3034428455066207 puntos de los 3 que hay en juego, mientras que el Leipzig tan solo se llevaría 0,56624904295391. Esto indica que el Real Madrid tiene una probabilidad de ganar cercana al 79% ($2,3034/3$) mientras que el Leipzig tan solo del 21 % ($0,5662/3$).

Leipzig and Real Madrid

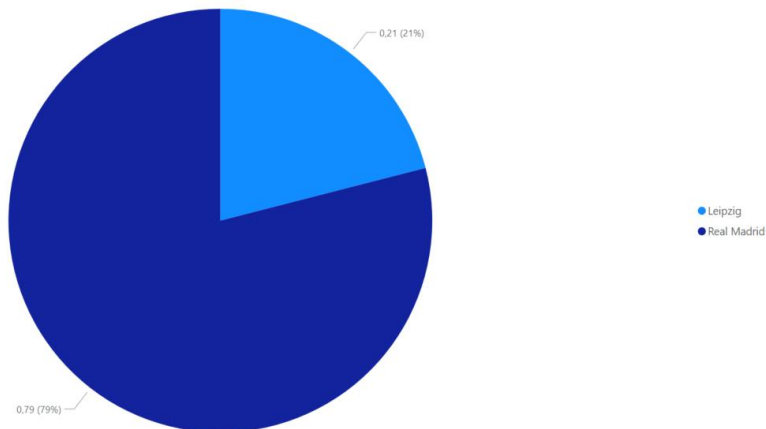


Figura 2.3: Probabilidades Real Madrid – Leipzig I (Elaborado con PowerBi)

La función *predict_points* la he aplicado a todos los partidos de los grupos que incluían a los equipos que aparecían en la base de datos **champions_league_2020_2022**. Es decir, los cruces de los distintos grupos en los cuales ambos equipos estaban incluidos en la base de datos mencionada, por ende, que habían jugado al menos una de las dos ediciones anteriores de la Champions League, son los que he simulado. He podido hacer este ejercicio con 22 de los 32 equipos participantes.⁴¹

⁴¹ Ver Anexos 40 - 47.

Predict_points2

No obstante, las probabilidades obtenidas con la función *predict_points* solo son aplicables a aquello que han realizado los equipos en las ediciones de la Champions League 2020-2021 y 2021-2022, sin tener en cuenta sus actuaciones en sus respectivas ligas. Es importante ver el desempeño de los equipos no solo en la competición, sino también cómo lo hacen a lo largo del año en su país. Para ello, he ejecutado una segunda función **predict_points2**, con la que he simulado como serían los cruces entre los equipos teniendo en cuenta los datos de sus ligas. El problema al utilizar el dataset **leagues_2021_2022** es que solo incluye datos de los equipos de las cinco grandes ligas. Estas ligas son: la Liga Inglesa, la Liga Española, la Liga Italiana, la Liga Alemana y la Liga Francesa. Al ser las grandes ligas, muchos de los equipos que figuran en esta base de datos ya salen en **champions_league_2020_2022**.

Para estos equipos, tengo dos probabilidades. La primera, está basada en los datos contenidos en la base de datos **champions_league_2020_2022**, es decir, la obtenida con *predict_points*. La segunda, está basada en los contenidos de la base de datos **leagues_2021_2022**, la obtenida con la función *predict_points2*. En cambio, para los equipos que no figuran en la base de datos **champions_league_2020_2022**, la probabilidad obtenida con la función **predict_points2** es la única que he analizado. Para aquellos equipos para los que he podido calcular los puntos mediante la función **predict_points** y **predict_points2** calcularé la probabilidad sobre un máximo de seis puntos. En cambio, para aquellos cuyos datos solo aparecen para una de las dos funciones, solo utilizaré la probabilidad sobre un máximo de tres puntos (**predict_points** o **predict_points2**, solamente).

Como con la función *predict_points*, he importado el csv **leagues_2021_2022**.⁴² y he creado *df_team_strength2* para recoger el *team strength* de los equipos de esta base de datos.

Con la misma estructura que la función *predict_points*, he ejecutado el mismo código con el nombre de *predict_points2*⁴³ y tomando como valores los datos relativos a la base de datos **leagues_2021_2022**. Igual que con la función anterior, la nueva función creada devuelve dos valores que hacen referencia al reparto de 3 puntos posibles entre los dos equipos. Si estos dos equipos no pertenecen a la base de datos a la que se refiere esta nueva función, la misma devolverá dos ceros, como ocurría con *predict_points*.

⁴² Ver Anexo 48

⁴³ Ver Anexo 49

Al coger otros datos, han cambiado las probabilidades y el Leipzig, ahora, tiene más probabilidades de ganar ante el Real Madrid⁴⁴. Por ello, he sumado las probabilidades obtenidas con las dos funciones (**predict_points** y **predict_points2**). De esta manera, las probabilidades del Real Madrid de ganar el partido son del 69,25 % ($1,73 + 2,30/6$) y las del Leipzig son del 30,75 % ($1,08 + 0,56/6$).

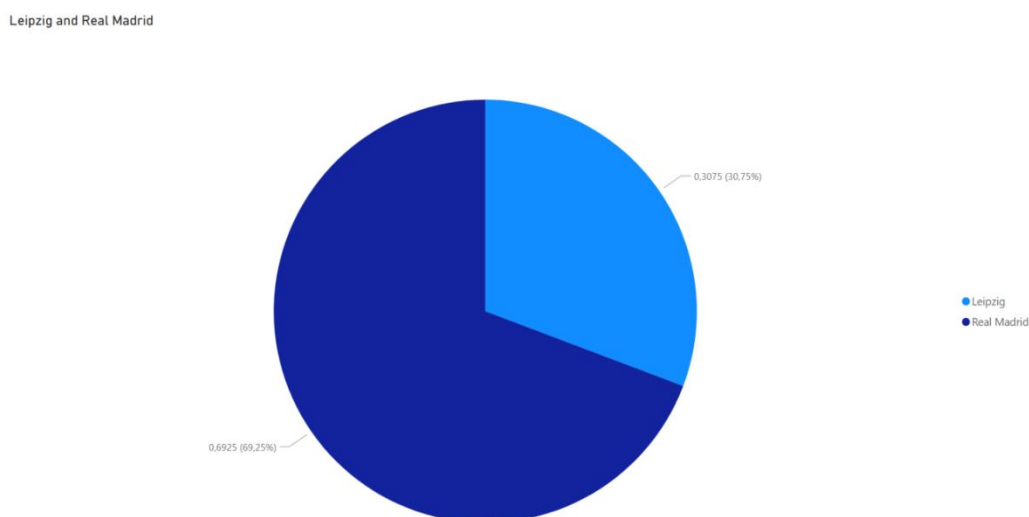


Figura 2.4: Probabilidades Real Madrid – Leipzig II (Elaborado con PowerBi)

Mediante esta función, he obtenido resultados para 19 de los 32 equipos participantes⁴⁵.

Extra Teams

Los equipos que no participaron en ninguna de las ediciones pasadas de la UEFA Champions League (ediciones 2021-2022 y 2020-2021) y que tampoco pertenecen a las grandes ligas, no aparecen en ninguna de las bases de datos de las que tiran las funciones **predict_points** y **predict_points2**. Por ello, he calculado su *team strength* y lo he comparado de manera simple con aquel de los grandes equipos, sin necesidad de crear una nueva función *predict_points3*. Además, coincide que los equipos que aparecen en este epígrafe son aquellos con menos nivel o con menos participaciones en la competición.

⁴⁴ Ver Anexo 50

⁴⁵ Ver Anexo 51 - 52

Tras simular todos los partidos de la fase de grupos, he obtenido los siguientes resultados por grupo. He incluido un 10% de probabilidad añadida a aquel equipo que juega en su estadio, mientras que se le resta a aquel que está jugando fuera de su estadio:

En el grupo A:



Liverpool - Ajax



IDA

VUELTA

Liverpool: 66% (+10% por jugar en su estadio)

Ajax: 54%

Ajax: 34% (-10% por jugar fuera su estadio)

Liverpool: 46%



Liverpool - Nápoles



IDA

VUELTA

Liverpool: 71%

Napoli: 49%

Napoli: 29%

Liverpool: 51%

Para los enfrentamientos de los equipos con el Rangers, he hecho una comparación directa de su *Team Strength*, ya que, al no haber participado en las ediciones anteriores ni pertenecer a una de las 5 grandes ligas, no ha sido posible su simulación. Además, he hecho lo mismo para el Ajax – Napoli, ya que cada uno de ellos está dentro de una función *predict_points* diferente y no son comparables. La comparación arroja el siguiente resultado:

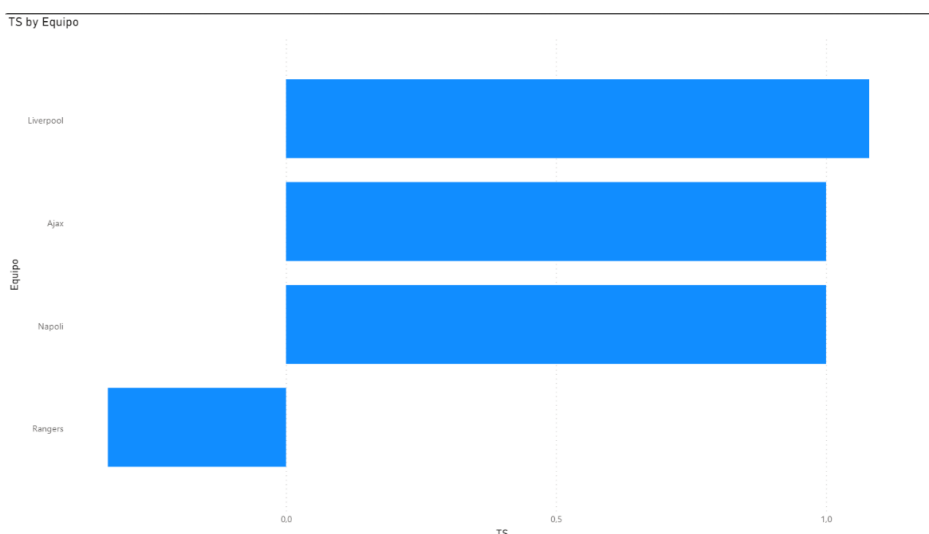


Figura 2.5: Team Strength de los equipos del Grupo A (Elaborado con PowerBi)

En el grupo B:



IDA

Oporto: 66,5%

Atlético de Madrid: 33,5%

Oporto – Atlético de Madrid



VUELTA

Atlético de Madrid: 53,5%

Oporto: 56,5%



IDA

Oporto: 86%

Brujas: 14%

Oporto – Brujas



VUELTA

Brujas: 34%

Oporto: 66%



IDA

Atlético de Madrid: 77%

Brujas: 23%

Atlético de Madrid – Brujas



VUELTA

Brujas: 43%

Atlético de Madrid: 57%



IDA

Atlético de Madrid: 58,5%

Bayer Leverkusen: 41,5%

Atlético de Madrid – Bayer Leverkusen



VUELTA

Bayer Leverkusen: 61,5%

Atlético de Madrid: 38,5%

Para los enfrentamientos del Bayer Leverkusen con el Brujas y con el Oporto, he hecho una comparación directa de su *Team Strength*, ya que no ha sido posible su simulación porque cada uno de ellos está dentro de una función *predict_points* y no son comparables. La comparación arroja el siguiente resultado:

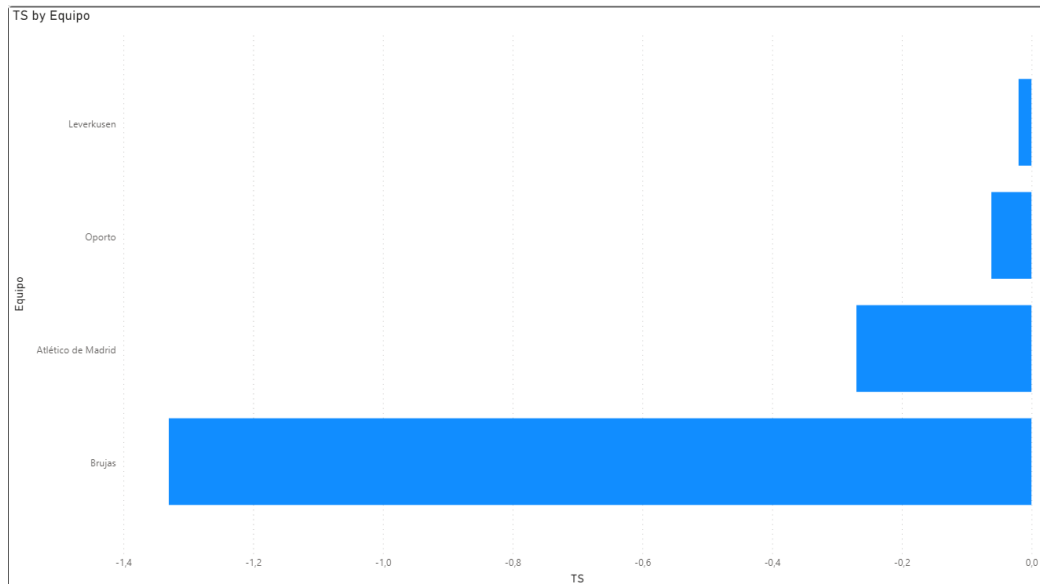


Figura 2.6: Team Strength de los equipos del Grupo B (Elaborado con PowerBi)

En el grupo C:



IDA

Bayern de Múnich: 87%

Barcelona: 13%

Bayern de Múnich – Barcelona



VUELTA

Barcelona: 33%

Bayern de Múnich: 67%



IDA

Bayern de Múnich: 76,5%

Inter de Milán: 23,5%

Bayern de Múnich – Inter de Milán



VUELTA

Inter de Milán: 43,5%

Bayern de Múnich: 56,5%

Barcelona – Inter de Milán



IDA

Barcelona: 51%

Inter de Milán: 49%



VUELTA

Inter de Milán: 69%

Barcelona: 31%

Para los enfrentamientos de los equipos con el Viktoria Plzen, he hecho una comparación directa de su *Team Strength*, ya que, al no haber participado en las ediciones anteriores ni pertenecer a una de las 5 grandes ligas, no ha sido posible su simulación. La comparación arroja el siguiente resultado:

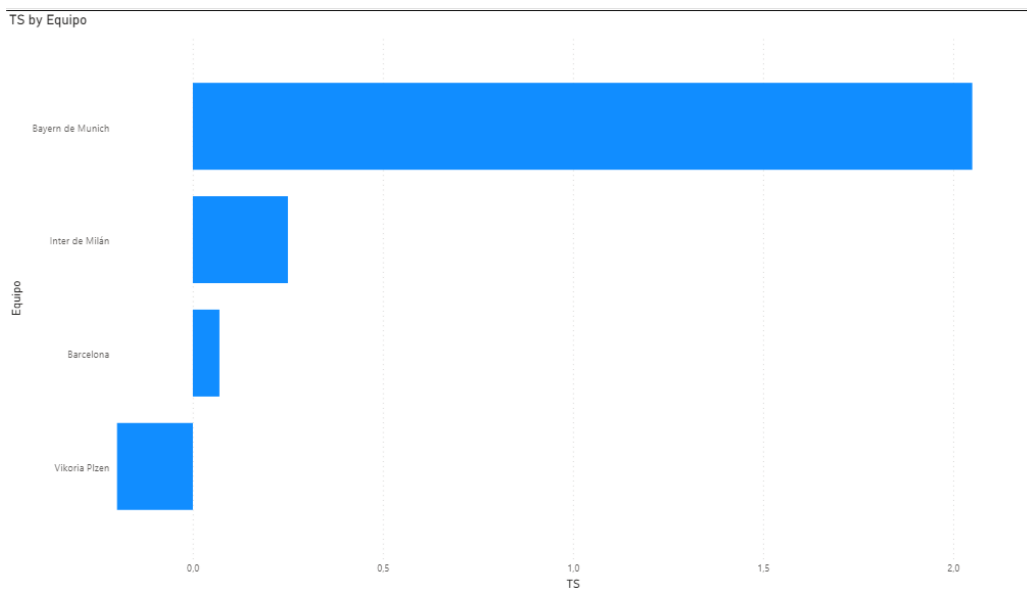


Figura 2.7: Team Strength de los equipos del Grupo C (Elaborado con PowerBi)

En el grupo D:



IDA

Marseille: 15%

Sporting CP: 85%

Marseille – Sporting CP



VUELTA

Sporting CP: 99%

Marseille: 1%



IDA

Eintracht Frankfurt: 40,5%

Tottenham Hotspur: 59,5%

Eintracht Frankfurt – Tottenham



VUELTA

Tottenham Hotspur: 79,5%

Eintracht Frankfurt: 20,5%



IDA

Eintracht Frankfurt: 39,5%

Marseille: 60,5%

Eintracht Frankfurt – Marseille



VUELTA

Marseille: 80,5%

Eintracht Frankfurt: 19,5%



IDA

Tottenham Hotspur: 58%

Marseille: 42%

Tottenham – Marseille



VUELTA

Marseille: 62%

Tottenham Hotspur: 38%

Para los enfrentamientos Sporting CP-Tottenham Hotspur y Sporting CP-Eintracht he hecho una comparación directa de su *Team Strength*, ya que cada uno de ellos está dentro de una función predict_points y no son comparables. La comparación arroja el siguiente resultado:

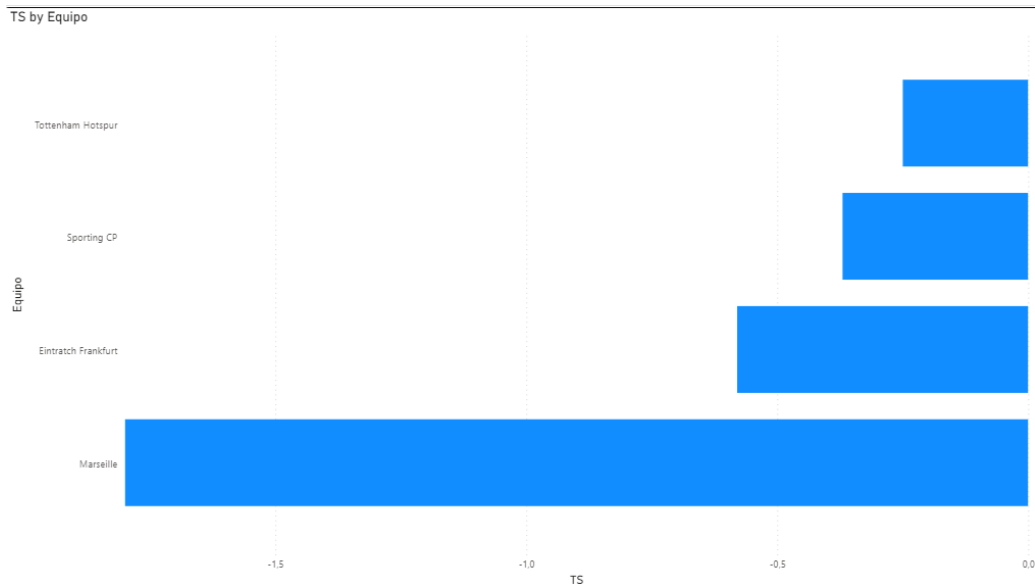


Figura 2.8: Team Strength de los equipos del Grupo D (Elaborado con PowerBi)

En el grupo E:



IDA

Milán: 39,5%

Chelsea: 60,5%

Milán – Chelsea



VUELTA

Chelsea: 80,5%

Milán: 19,5%



IDA

Milán: 61,5%

Salzburgo: 39,5%

Milán – Salzburgo



VUELTA

Salzburgo: 58,5%

Milán: 41,5%



Chelsea – Salzburgo



IDA

Chelsea: 95%

Salzburgo: 5%

VUELTA

Salzburgo: 16,5%

Chelsea: 83,5%

Para los enfrentamientos de los equipos con el Dinamo Zagreb, he hecho una comparación directa de su *Team Strength*, ya que, al no haber participado en las ediciones anteriores ni pertenecer a una de las 5 grandes ligas, no ha sido posible su simulación. La comparación arroja el siguiente resultado:

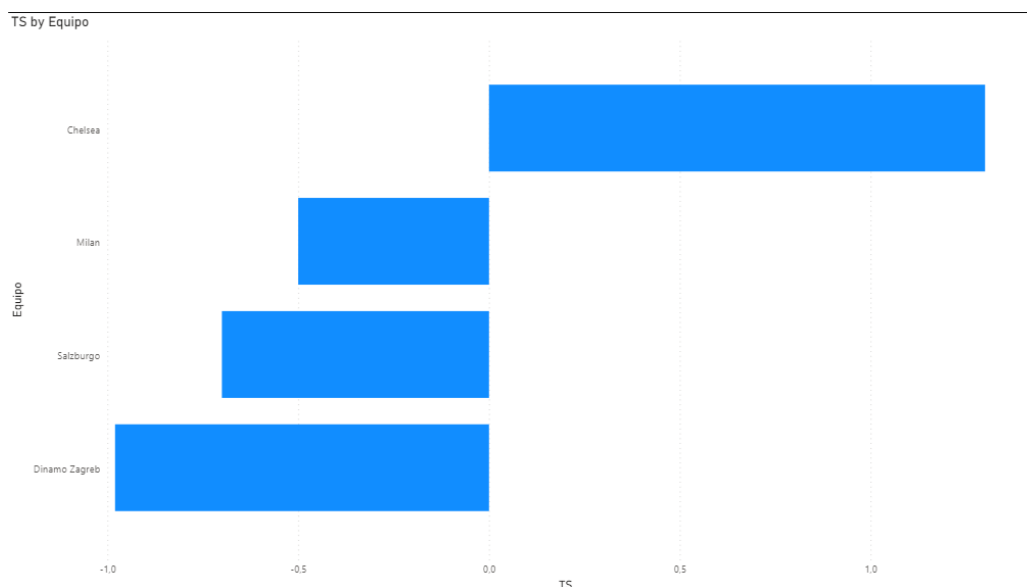


Figura 2.9: Team Strength de los equipos del Grupo E (Elaborado con PowerBi)

En el grupo F:



Real Madrid – Shakhtar Donetsk



IDA

Real Madrid: 95%

Shakhtar Donetsk: 5%

VUELTA

Shakhtar Donetsk: 16,5%

Real Madrid: 83,5%



Real Madrid – Leipzig



IDA

Real Madrid: 80%

Leipzig: 20%

VUELTA

Leipzig: 40%

Real Madrid: 60%



Shakhtar Donetsk – Leipzig



IDA

Shakhtar Donetsk: 23,5%

Leipzig: 76,5%

VUELTA

Leipzig: 96,5%

Shakhtar Donetsk: 3,5%

Para los enfrentamientos de los equipos con el Celtic, he hecho una comparación directa de su *Team Strength*, ya que, al no haber participado en las ediciones anteriores ni pertenecer a una de las 5 grandes ligas, no ha sido posible su simulación. La comparación arroja el siguiente resultado:

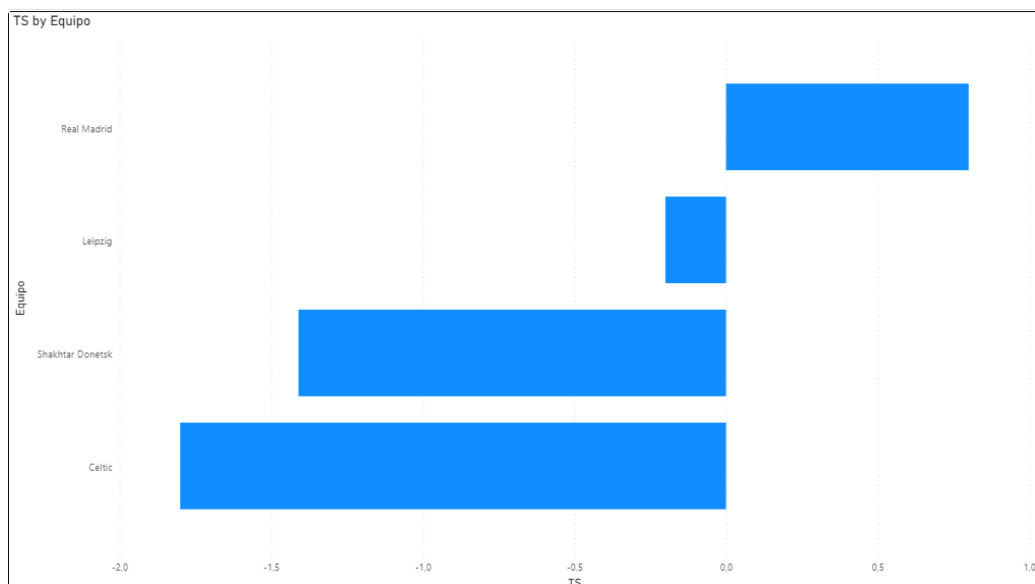


Figura 2.10: Team Strength de los equipos del Grupo F (Elaborado con PowerBi)

En el grupo G:



IDA

Man. City: 86,5%

Sevilla: 13,5%

Man. City – Sevilla



VUELTA

Sevilla: 33,5%

Man. City: 66,5%



IDA

Man. City: 90,5%

Borussia Dortmund: 9,5%

Man. City – Borussia Dortmund



VUELTA

Borussia Dortmund: 29,5%

Man. City: 70,5%



IDA

Sevilla: 58,5%

Borussia Dortmund: 41,5%

Sevilla – Borussia Dortmund



VUELTA

Borussia Dortmund: 61,5%

Sevilla: 38,5%

Para los enfrentamientos de los equipos con el Copenhague, he hecho una comparación directa de su *Team Strength*, ya que, al no haber participado en las ediciones anteriores ni pertenecer a una de las 5 grandes ligas, no ha sido posible su simulación. La comparación arroja el siguiente resultado:

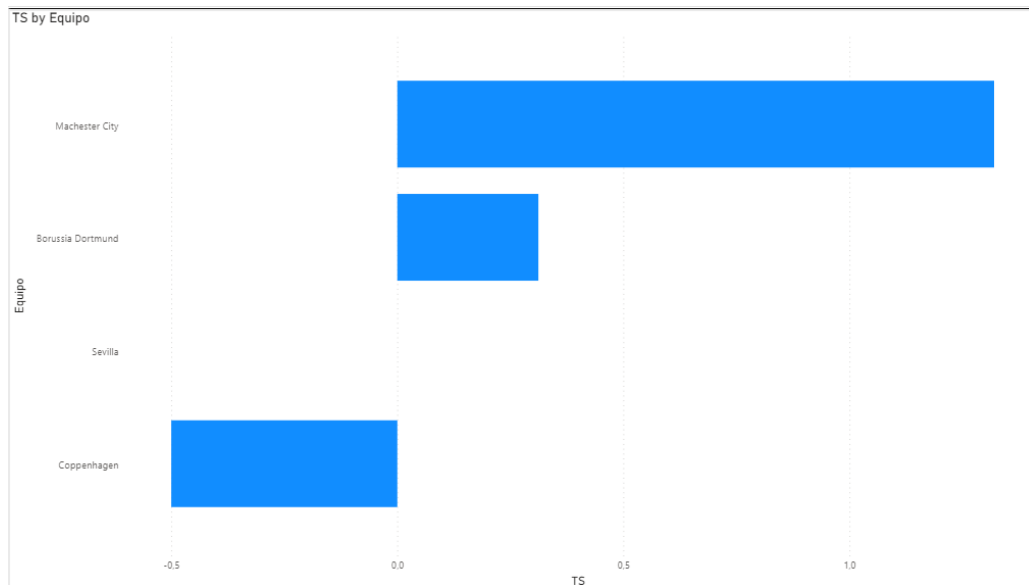


Figura 2.11: Team Strength de los equipos del Grupo G (Elaborado con PowerBi)

En el grupo H:



IDA

Paris Saint-Germain: 67%

Juventus: 33%

Paris Saint- Germain – Juventus



VUELTA

Juventus: 53%

Paris Saint-Germain: 47%



Paris Saint- Germain – Benfica



IDA

Paris Saint-Germain: 82,5%

Benfica: 17,5%

VUELTA

Benfica: 37,5%

Paris Saint-Germain: 62,5%



Juventus- Benfica



IDA

Juventus: 85%

Benfica: 15%

VUELTA

Benfica: 35%

Juventus: 65%

Para los enfrentamientos de los equipos con el Maccabi Haifa, he hecho una comparación directa de su *Team Strength*, ya que, al no haber participado en las ediciones anteriores ni pertenecer a una de las 5 grandes ligas, no ha sido posible su simulación. La comparación arroja el siguiente resultado:

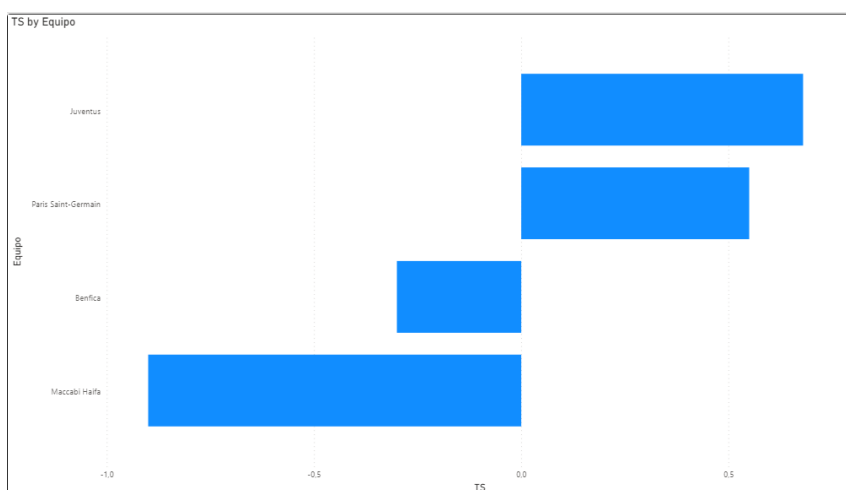


Figura 2.12: Team Strength de los equipos del Grupo H (Elaborado con PowerBi)

Resultado final fase de grupos (clasificados a la siguiente ronda)

Al final, tras la simulación de todos los partidos, he obtenido a los clasificados por grupo y sus respectivos puntos, teniendo en cuenta los siguientes criterios:

- Si un equipo tiene más de un 60% de probabilidad de ganar un partido, se llevará los **3 puntos** relativos a ese partido.
- Si los porcentajes de un partido están entre el 40% y el 60%, se llevará **1 punto** cada equipo en ese partido.
- Si un equipo tiene menos de 40% de probabilidad, se llevará **0 puntos**.

Esto determina no solo el número de puntos que consigue un equipo sino también el número de victorias, derrotas y empates. Tras realizar el cálculo de los mismos, una vez simulados todos los partidos, incluyendo el dato de *team_strength* utilizado para medir a los equipos que no pertenecían a ninguna de las funciones, la clasificación por grupos ha quedado de la siguiente manera:

GRUPO A					
Posición	Equipo	Puntos	V	D	E
1	Liverpool	14	4	0	2
2	Ajax	10	3	2	1
3	Napoli	10	3	2	1
4	Rangers	0	0	6	0

GRUPO B					
Posición	Equipo	Puntos	V	D	E
1	Oporto	10	3	2	1
2	Bayer Leverkusen	8	2	2	2
3	Atlético	6	1	2	3
4	Brujas	1	0	5	1

GRUPO C					
Posición	Equipo	Puntos	V	D	E
1	Bayern de Munich	16	5	0	1
2	Inter de Milán	11	3	1	2
3	Barcelona	7	2	3	1
4	Viktoria Plzen	0	0	6	0

GRUPO D					
Posición	Equipo	Puntos	V	D	E
1	Sporting CP	13	4	1	1
2	Tottenham	11	3	1	2
3	Marseille	6	2	4	0
4	Eintracht Frankfurt	2	0	4	2

GRUPO E					
Posición	Equipo	Puntos	V	D	E
1	Chelsea	18	6	0	0
2	Milan	10	3	2	1
3	Salzburgo	7	2	3	1
4	Dinamo Zagreb	0	0	6	0

GRUPO F					
Posición	Equipo	Puntos	V	D	E
1	Real Madrid	18	6	0	0
2	Leipzig	12	4	2	0
3	Shakhtar Donetsk	6	2	4	0
4	Celtic	0	0	6	0

GRUPO G					
Posición	Equipo	Puntos	V	D	E
1	Man. City	18	6	0	0
2	Borussia Dortmund	10	3	2	1
3	Sevilla	7	2	3	1
4	Copenhagen	0	0	6	0

GRUPO H					
Posición	Equipo	Puntos	V	D	E
1	Paris Saint-Germain	16	5	0	1
2	Juventus	13	4	1	1
3	Benfica	6	2	4	0
4	Maccabi Haifa	0	0	6	0

*Entre el Ajax y el Napoli, que empatan a puntos, he supuesto que el segundo clasificado es el Ajax por la posición en el ranking histórico.

B) OCTAVOS DE FINAL

En la Champions League, al llegar a la fase de octavos de final, no hay un cuadro predeterminado de cruces en los octavos de final. Esto quiere decir que los equipos, una vez clasificados, no sabrán a que equipo se enfrentan hasta que se realiza el sorteo.

En el sorteo de octavos de final hay dos bombos o urnas. En la primera urna, se colocan las papeletas de los primeros clasificados de cada grupo. En la segunda urna, los segundos clasificados de los mismos. Al azar, se cogerán dos bolas, una de cada urna, para determinar los emparejamientos de esta ronda. No obstante, hay dos restricciones a este sorteo:

1. No se pueden enfrentar el primer clasificado y el segundo de un mismo grupo en los octavos de final
2. No podrán enfrentarse equipos de la misma nacionalidad, es decir, de la misma liga de un país. Por ejemplo, el enfrentamiento Tottenham (2º) vs. Chelsea (1º) no podría darse.

Por ello, he simulado un sorteo virtual en internet para obtener los cruces de esta ronda, teniendo en cuenta las restricciones, con el siguiente resultado:

OCTAVOS DE FINAL			
Chelsea	vs.	Bayer Leverkusen	
Paris Saint- Germain	vs.	Tottenham Hotspur	
Real Madrid	vs.	Juventus	
Manchester City	vs.	Inter de Milán	
Sporting CP	vs.	Milan	
Bayern de Munich	vs.	Ajax	
Liverpool	vs.	Leipzig	
Oporto	vs.	Borussia Dortmund	

En la columna de la izquierda, aparecen los primeros clasificados de sus grupos. En la columna de la derecha, los segundos clasificados de esta fase inicial.

Como en la fase de grupos, he simulado los partidos para ver que probabilidades tendrían de clasificarse los equipos, esta vez, siendo solo posible el empate en la ronda de ida⁴⁶.

⁴⁶ Ver Anexo 53



IDA

Bayer Leverkusen: 43%

Chelsea: 57%

Bayer Leverkusen – Chelsea



VUELTA

Chelsea: 77%

Bayer Leverkusen: 23%

CLASIFICADO: Chelsea



IDA

Tottenham Hotspur: 41,5%

PSG: 58,5%

Tottenham – Paris Saint Germain



VUELTA

PSG: 78,5%

Tottenham Hotspur: 21,5%

CLASIFICADO: Paris Saint-Germain



IDA

Juventus: 51,5%

Real Madrid: 48,5%

Juventus – Real Madrid



VUELTA

Real Madrid: 68,5%

Juventus: 31,5%

CLASIFICADO: Real Madrid



IDA

Inter de Milán: 44%

Manchester City: 56%

Man. City – Inter de Milán



VUELTA

Manchester City: 76%

Inter de Milán: 24%

CLASIFICADO: Manchester City



IDA

Milán: 51,5%

Sporting CP: 48,5%

Milán – Sporting CP



VUELTA

Sporting CP: 68,5%

Milán: 31,5%

CLASIFICADO: Sporting CP



IDA

Ajax: 40%

Bayern de Múnich: 60%

Ajax – Bayern de Múnich



VUELTA

Bayern de Múnich: 80%

Ajax: 20%

CLASIFICADO: Bayern de Múnich



IDA

Leipzig: 32,5%

Liverpool: 67,5%

Leipzig – Liverpool



VUELTA

Liverpool: 87,5%

Leipzig: 12,5%

CLASIFICADO: Liverpool



IDA

Borussia Dortmund: 68,5%

Oporto: 31,5%

Borussia Dortmund – Oporto



VUELTA

Oporto: 51,5%

Borussia Dortmund: 48,5%

CLASIFICADO: Borussia Dortmund

C) CUARTOS DE FINAL

Tras la ronda de Octavos de Final, los clasificados se ordenan en una llave, a sorteo, que posteriormente determine como serán los cruces hasta la final. En este caso, ya no hay ningún tipo de restricción a los cruces que se puedan dar. Todos los equipos se pueden enfrentar entre ellos. He simulado los partidos de nuevo⁴⁷.

Para obtener la llave de cruces, he simulado, de nuevo, un sorteo ficticio. La llave obtenida ha sido la siguiente:



Figura 2.13: Cuadro Sorteo Cuartos de Final



Sporting CP – Liverpool



IDA

Sporting CP: 24%

Liverpool: 76%

VUELTA

Liverpool: 96%

Sporting CP: 4%

CLASIFICADO: Liverpool



Man. City – Chelsea



IDA

VUELTA

⁴⁷ Ver Anexo 54

Manchester City: 63,5%

Chelsea: 56,5%

Chelsea: 36,5%

Manchester City: 43,5%

CLASIFICADO: Manchester City



IDA

Borussia D. – Bayern de Múnich



VUELTA

Borussia Dortmund: 30%

Bayern Múnich: 90%

Bayern Munich: 70%

Borussia Dortmund: 10%

CLASIFICADO: Bayern de Munich



IDA

Real Madrid- Paris Saint- Germain



VUELTA

Real Madrid: 62%

PSG: 42%

PSG: 38%

Real Madrid: 58%

CLASIFICADO: Real Madrid

D) SEMIFINALES

Tras la ronda de Cuartos de Final, he simulado esta ronda igual que la anterior⁴⁸. Los clasificados a esta ronda son los siguientes:

- Liverpool vs. Manchester City
- Bayern vs. Real Madrid

Los resultados obtenidos son los siguientes:

⁴⁸ Ver Anexo 55



Man. City- Liverpool



IDA

Manchester City: 64%

Liverpool: 36%

VUELTA

Liverpool: 56%

Manchester City: 44%

CLASIFICADO: Manchester City



Real Madrid- Bayern de Múnich



IDA

Real Madrid: 48%

Bayern de Múnich: 52%

VUELTA

Bayern de Múnich: 72%

Real Madrid: 28%

CLASIFICADO: Bayern de Munich

E) FINAL

La gran final de la Champions League 2022/2023 sería Bayern de Múnich vs. Manchester City. Para ello, he simulado el partido final⁴⁹, teniendo en cuenta el porcentaje obtenido en Python y sin tener en cuenta el 10% adicional por jugar de local en su estadio.



Final

Bayern de Múnich: 50,5%

Manchester City: 49,5%



⁴⁹ Ver anexo 56

La diferencia entre ambos es prácticamente mínima y sería muy difícil acertar el resultado o sacar una conclusión de un resultado tan ajustado. Por ello, debido a la igualdad, tomaré como referencia el ránking histórico de la competición para determinar como campeón al **Bayern de Múnich**. Este ocupa el segundo lugar del mencionado ránking, mientras el Manchester City ocupa el puesto número 30.

5. Análisis predictivo y estadístico de jugadores

Después de predecir el campeón de la competición, a nivel de equipo, he hecho un análisis más concreto por jugador. Para ello, he utilizado los datos recogidos de la página web de la UEFA e insertados en las bases de datos generadas por mí. Las bases de datos utilizadas en este apartado son **Player Stats CL – 2021-2022** y **Player Stats por partido CL – 2021-2022**.

En este análisis de jugadores he utilizado tanto Python como PowerBi para realizar un análisis predictivo y de probabilidad de ciertas facetas del juego. Más concretamente, en este apartado, voy a predecir:

- El máximo goleador de la competición (El jugador que más goles ha anotado)
- El mejor portero de la competición (El jugador que menos goles ha recibido)
- El mejor defensa de la competición (El jugador que se prevé que tenga las mejores estadísticas defensivas)

5.1 Máximo goleador de la competición

El máximo goleador de una competición de fútbol es aquel que más goles es capaz de transformar en la portería del equipo rival. De la página web de la UEFA, he extraído los datos de los veinte mejores delanteros, es decir, de los veinte jugadores que más goles han transformado. Las estadísticas por partido de estos jugadores se ven muy bien reflejadas en los siguientes gráficos:

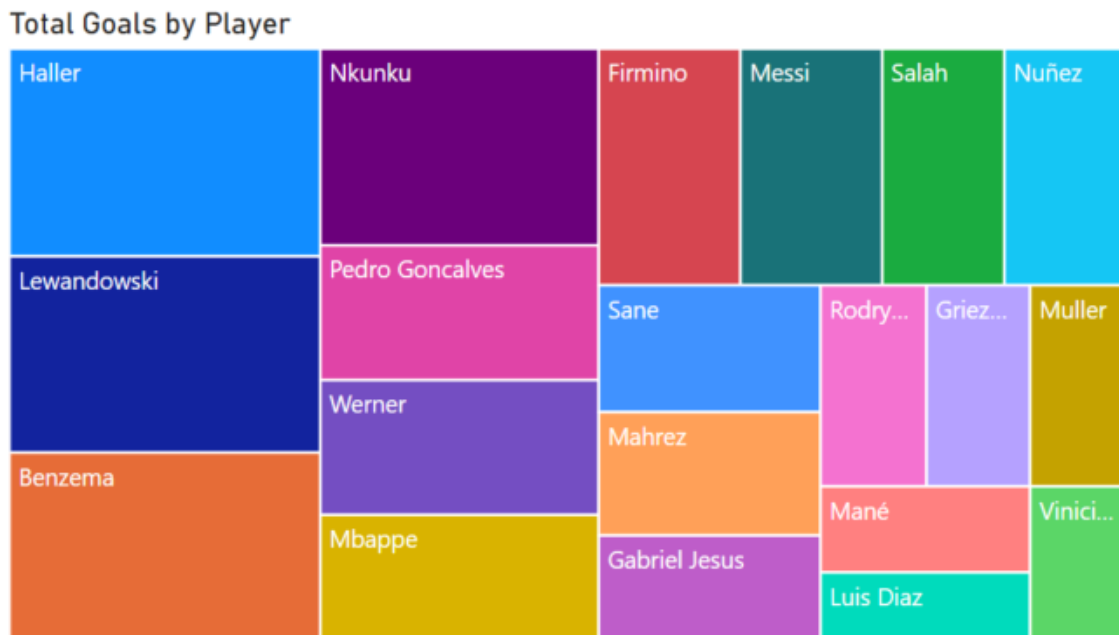


Figura 3.1: Goles por partido anotados por los máximos goleadores (Elaborado con PowerBi)

El gráfico anterior enseña cómo se han distribuido los goleadores si se tomase como total la suma de todos sus goles por partido. Haller se coloca en la cabeza, seguido por Lewandoswki y Nkunku.

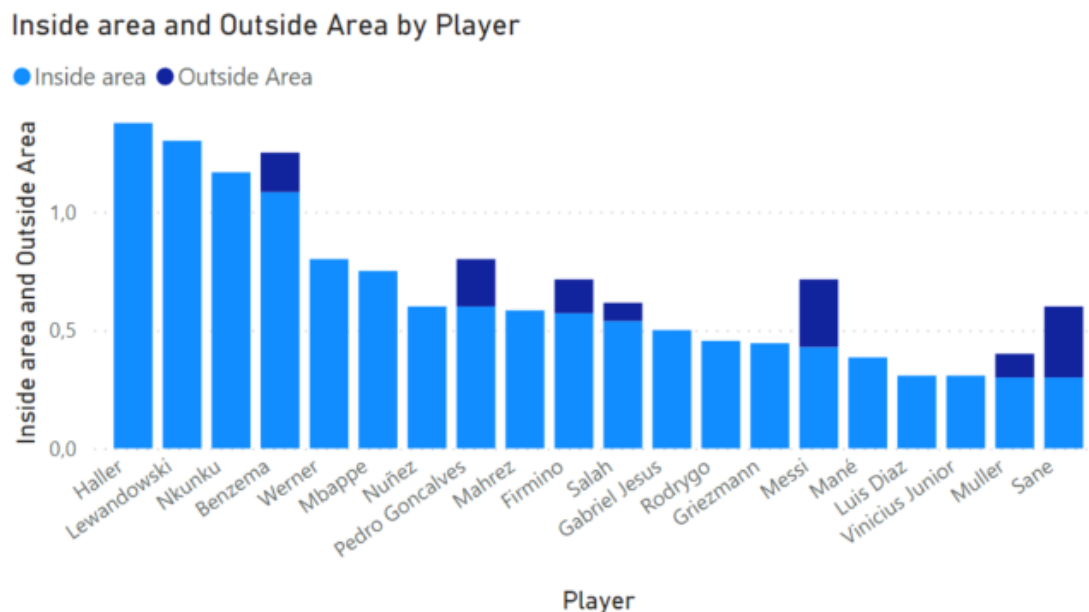


Figura 3.2: Goles desde dentro y fuera del área por partido (Elaborado con PowerBi)

El gráfico anterior muestra que cantidad de goles anotados por partido por los jugadores analizados han sido desde dentro del área, es decir, desde la zona donde el portero puede coger el balón con las manos, y qué goles han sido anotados desde fuera de la misma. En este caso, es interesante que ninguno de los tres máximos anotadores por partido ha conseguido anotar desde fuera del área.

Al ser jugadores, los cuales la mayoría llegaron a pasar la ronda de fase de grupos, para realizar la predicción he decidido utilizar las cifras en valor absoluto (no por partido) y he estimado los goles que meterán a nivel global. Estas cifras son más reales, ya que las cifras de goles por partidos están situadas entre 1 y 0 por lo que no se pueden sacar grandes conclusiones.

De la edición pasada de la Champions League, he extraído las siguientes estadísticas goleadoras de los veinte máximos anotadores de goles. Estos datos han sido extraídos de la base de datos *Player Stats – CL 2021-2022.csv*:

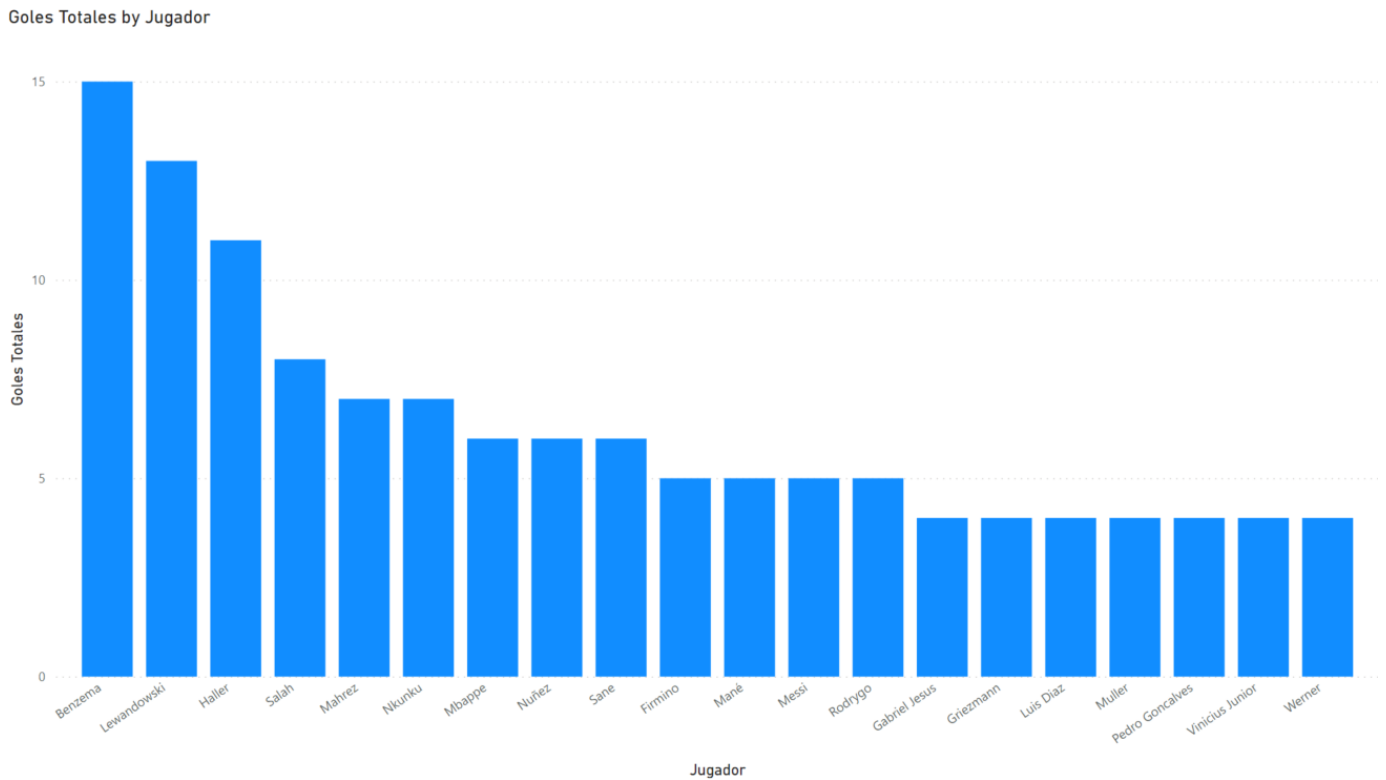


Figura 3.3: Máximos goleadores la Champions League 2021-2022. (Elaborado con PowerBi)

Los **goles** anotados que aparecen en la gráfica anterior, junto con los **tiros intentados** (disparos) y los **tiros a puerta** (disparos hacia la portería) han sido las variables que he seleccionado para la predicción. Entre las distintas variables hay una gran correlación y esto ayuda a una mayor precisión en el cálculo de los futuros goles⁵⁰.

En primer lugar, he importado la librería *NumPy*. La librería *NumPy* (Numerical Python) de Python es una biblioteca de funciones matemáticas de alto rendimiento que se utiliza para realizar operaciones numéricas y matriciales en Python. NumPy se utiliza en la ciencia de datos, la ingeniería, la física y otras disciplinas que involucran cálculos numéricos y análisis de datos. Esta librería proporciona una estructura de datos de matriz multidimensional (llamada "ndarray") que permite el almacenamiento y la manipulación de grandes conjuntos de datos de una manera eficiente.

⁵⁰ Ver Anexo 57

Además, he importado *LinearRegression()* de la biblioteca *sklearn.linear_model* para llevar a cabo la predicción con el método de regresión lineal múltiple.

En segundo lugar, he introducido los datos de goles anotados (*goals_scored*), tiros intentados (*shots_attempted*) y de tiros a puerta (*shots_target*) de cada jugador. Los valores han sido extraídos de la base de datos **Player Stats – CL 2021- 2022**. Los he introducido manualmente en una lista de Python, ya que he tenido un error en mi ordenador.

En tercer lugar, he creado una lista en Python con los datos de los veinte jugadores analizados. Para cada jugador he introducido tres inputs relativos a las tres variables con las que he trabajado en este apartado. He creado una tupla que contiene las listas con los datos para los jugadores. Para cada jugador he creado una lista en la que los datos de las tres variables están separados por comas.

En cuarto lugar, he dividido las variables en variables dependientes e independientes. El número de goles anotados (*goals_scored*) es nuestra variable dependiente, mientras que las variables tiros intentados (*shots_attempted*) y tiros a puerta (*shots_target*) son las variables independientes. Las variables independientes son aquellas que se pueden manipular y modificar y cuyo impacto se ve reflejado en la variable dependiente. Las variables independientes también se conocen como variables explicativas. Por otro lado, las variables dependientes son aquellas que no existen si no hay al menos una variable independiente que le dé un valor. En este caso, *goals_scored* (goles anotados) es la variable dependiente de las variables independientes *shots_attempted* y *shots_target*.

En quinto lugar, he creado el modelo de Regresión Lineal llamando a la función *LinearRegression()*. En el modelo creado he introducido las variables definidas en el párrafo anterior. Un modelo de regresión lineal múltiple es un modelo estadístico que busca establecer una relación lineal entre una variable dependiente y dos o más variables independientes. En otras palabras, trata de predecir el valor de una variable dependiente a partir de dos o más variables independientes que se consideran relevantes.

El modelo se puede expresar matemáticamente como:

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_px_p + \varepsilon$$

Donde:

y es la variable dependiente

x₁, x₂, ..., x_p son las variables independientes

β₀, β₁, β₂, ..., β_p son los coeficientes del modelo

ε es el término de error o residuo, que representa la diferencia entre el valor observado y el valor predicho por el modelo.

El objetivo del modelo es encontrar los valores de los coeficientes $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ que minimizan la suma de los cuadrados de los residuos. Para ello se utilizan técnicas de regresión múltiple, como el método de mínimos cuadrados o el método de máxima verosimilitud.

Si solo hubiese utilizado una, el modelo sería simple.

En sexto lugar, he introducido los datos relativos al jugador en cuestión, e introducidos previamente, en el modelo de predicción.

Por último, he utilizado la función *print* para que la función devuelva una frase en la que se indique que el jugador ha marcado el número determinado de goles predicho anteriormente.

He repetido este ejercicio con los 20 jugadores, obteniendo los siguientes resultados:

Jugador	Total Goles Año 2021-2022	Predicción Goles
Benzema	15	13,7
Lewandowski	13	12,66
Haller	11	10,33
Salah	8	12,4
Nkunku	7	10,33
Mahrez	7	4,13
Mbappe	6	7,23
Núñez	6	4,13
Sane	6	8,01
Firmino	5	2,84
Mané	5	3,61
Rodrygo	5	5,68
Messi	5	5,42
Gabriel Jesus	4	5,42
Werner	4	2,83
Pedro Goncalves	4	5,68
Vinicius Junior	4	5,17
Luis Diaz	4	6,46
Muller	4	4,13
Griezmann	4	5,17

Figura 3.4: Comparativa de goles anotados la edición pasada vs. la predicción estimada para esta edición de la Champions League.

Al haber cogido los datos del año pasado y haber hecho una estimación basada en los mismos, los 3 máximos goleadores no han cambiado: Benzema, Lewandowski y Haller. No obstante, es interesante analizar como bajará o subirá el número total de goles que anotan. Por ejemplo, mientras jugadores como Núñez o Mahrez bajarán sus cifras goleadoras, otros jugadores como Salah o Nkunku subirán las suyas notablemente. Por lo general, se mantienen cerca de los goles que anotaron el año pasado ya que la correlación de goles anotados con tiros y tiros a puerta es muy fuerte. Por ello, los jugadores citados que han subido sus estadísticas son aquellos que

tuvieron un número alto de disparos y de disparos a puerta y para los que se prevé una mayor cantidad de goles:

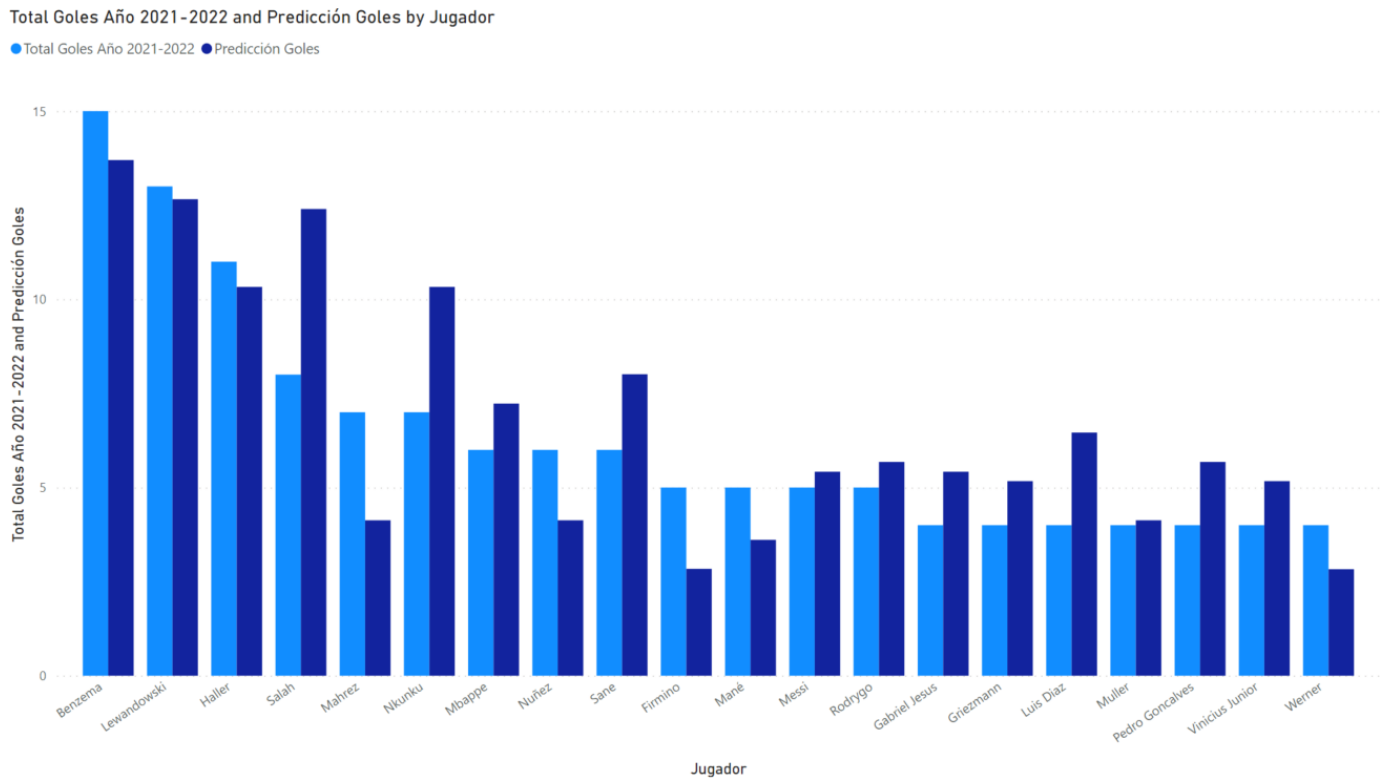


Figura 3.5: Comparativa de goles anotados la edición pasada vs. la predicción estimada para esta edición (Elaborado con PowerBi)

El jugador con más goles anotados de la competición es **Karim Benzema**.

5.2 Mejor portero de la competición

La predicción para determinar el número de paradas que realizarán los veinte porteros analizados es muy similar a la realizada con los delanteros. No obstante, en vez de tomar los valores absolutos de la competición para las variables utilizadas, he tomado las cifras por partido. Como con el análisis de máximo goleador, también voy a coger a los veinte mejores porteros de la edición pasada que participen también este año. Las gráficas siguientes muestran los datos de los porteros la edición pasada:

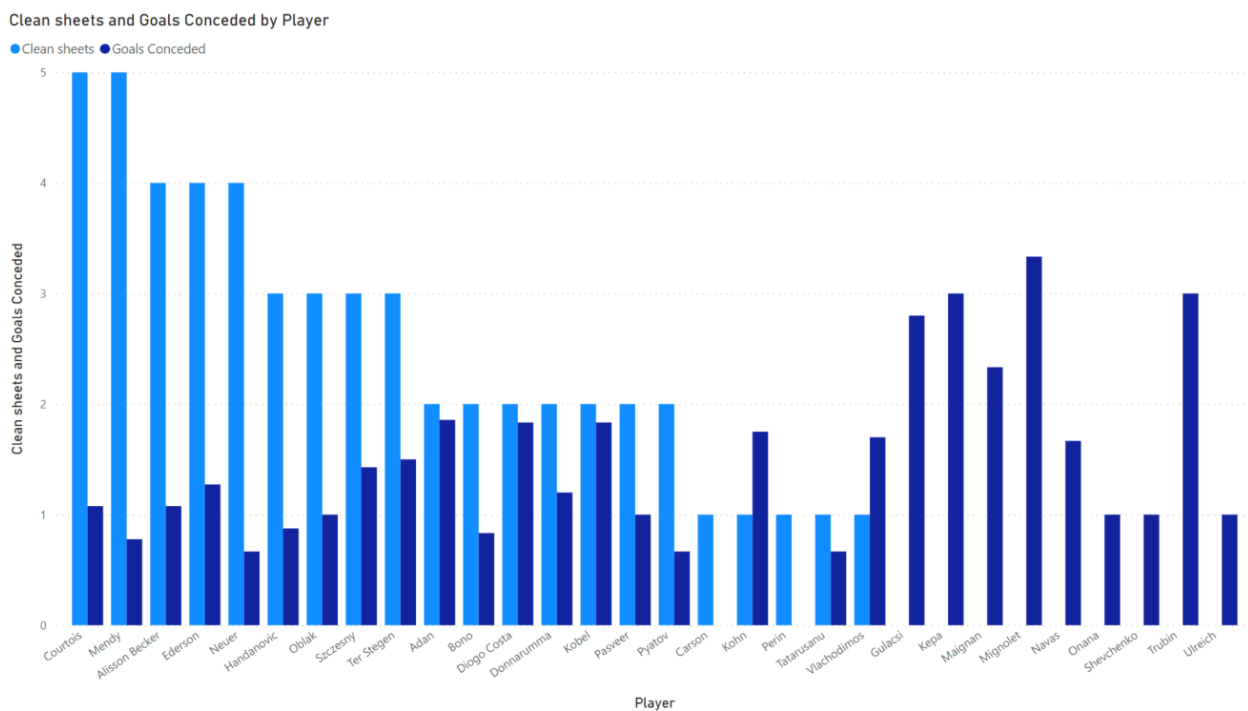


Figura 3.6: Porterías a cero (*clean sheets*) y goles encajados por partido (*goals conceded*) por los porteros la edición pasada de la competición (Elaborado con PowerBi)

La gráfica anterior indica que hay una relación directa entre no encajar un gol, es decir, dejar la portería a cero y los goles encajados. Los porteros con más goles encajados no han sido capaces de mantener su portería a cero ni un partido.

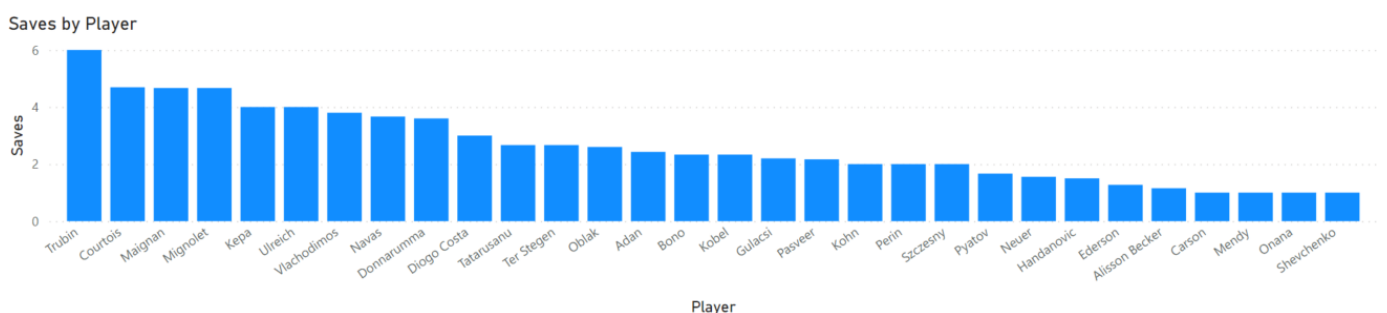


Figura 3.7: Paradas (*Saves*) por partido de cada portero (Elaborado con PowerBi)

El gráfico anterior muestra el número de paradas por partido realizadas por los porteros. Este demuestra que realizar muchas paradas no está relacionado con conceder menos goles, pero si tiene cierta correlación. En el caso del portero Trubin, es el que más paradas hace, pero de los que más goles reciben por partido. En cambio, Courtois, segundo en la lista de más paradas, es de los porteros que menos goles reciben por partido.

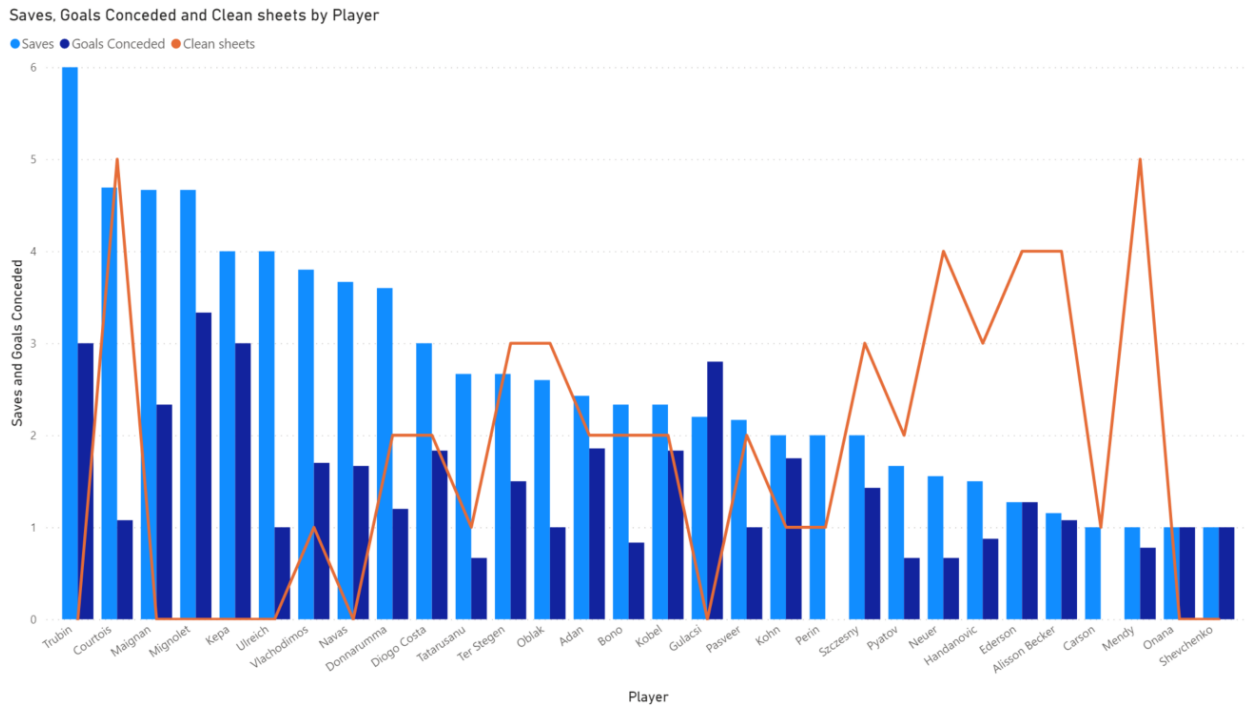


Figura 3.8: Paradas por partido (*Saves*), Goles concedidos por partido (*Goals Conceded*) y Porterías a cero (*clean sheets*) de los porteros en la edición pasada de la competición (Elaborado con PowerBi)

En esta última gráfica, el mantener la portería a cero está altamente relacionado con los goles concedidos, como aparece en la Figura 3.6. No obstante, el realizar más paradas no es en ningún caso indicativo de mantener la portería a cero. De hecho, salvo el caso de Courtois, la línea de portería a cero (*clean sheets*) y la de paradas realizadas (*saves*) están correlacionadas contrariamente.

Entrando en la predicción y el análisis, para el análisis del mejor portero voy a analizar las variables:

- Goles encajados (*Goals Conceded*)
- Paradas (*Saves*)
- Partidos con la portería a cero (*Clean sheets*)

El código que utilizaré es igual que el utilizado en el apartado de máximo goleador, pero basándome en las tres variables indicadas anteriormente. No obstante, en el caso de los porteros, voy a utilizar las variables por partido, en vez de en toda la competición, ya que la diferencia de goles concedidos y partidos jugados es notable⁵¹.

Al realizar el mismo ejercicio con el resto de porteros, he tenido en cuenta, únicamente, a los 20 mejores porteros de los 30 que hay en mi muestra. El resultado que he obtenido se ve reflejado en la siguiente gráfica:

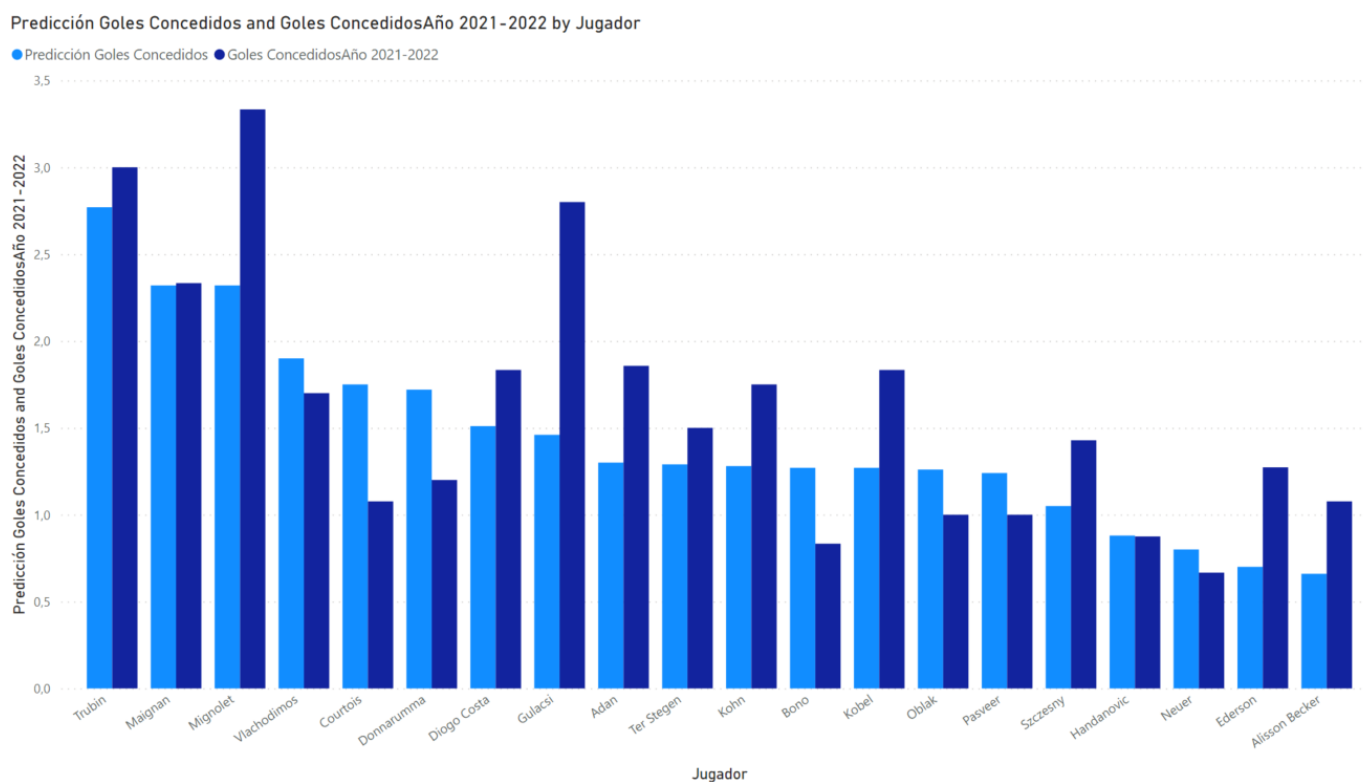


Figura 3.9: Comparativa predicción de goles con los goles encajados la edición pasada de la competición (Elaborado con PowerBi)

Las barras azules claras representan el número de goles que se ha predicho que van a encajar los porteros por partido en esta edición, mientras que las barras azules oscuras indican el número de goles que les encajaron la temporada pasada por partido.

⁵¹ Ver anexo 58.

Aunque la predicción realizada sea por partido, es una realidad que es más fácil recibir menos goles en menos partidos. Por ello, he filtrado por segunda vez los resultados de la gráfica para obtener el resultado de aquellos porteros que han jugado más de 7 partidos, es decir, que han pasado de la fase de grupos de la competición. El resultado es el siguiente:

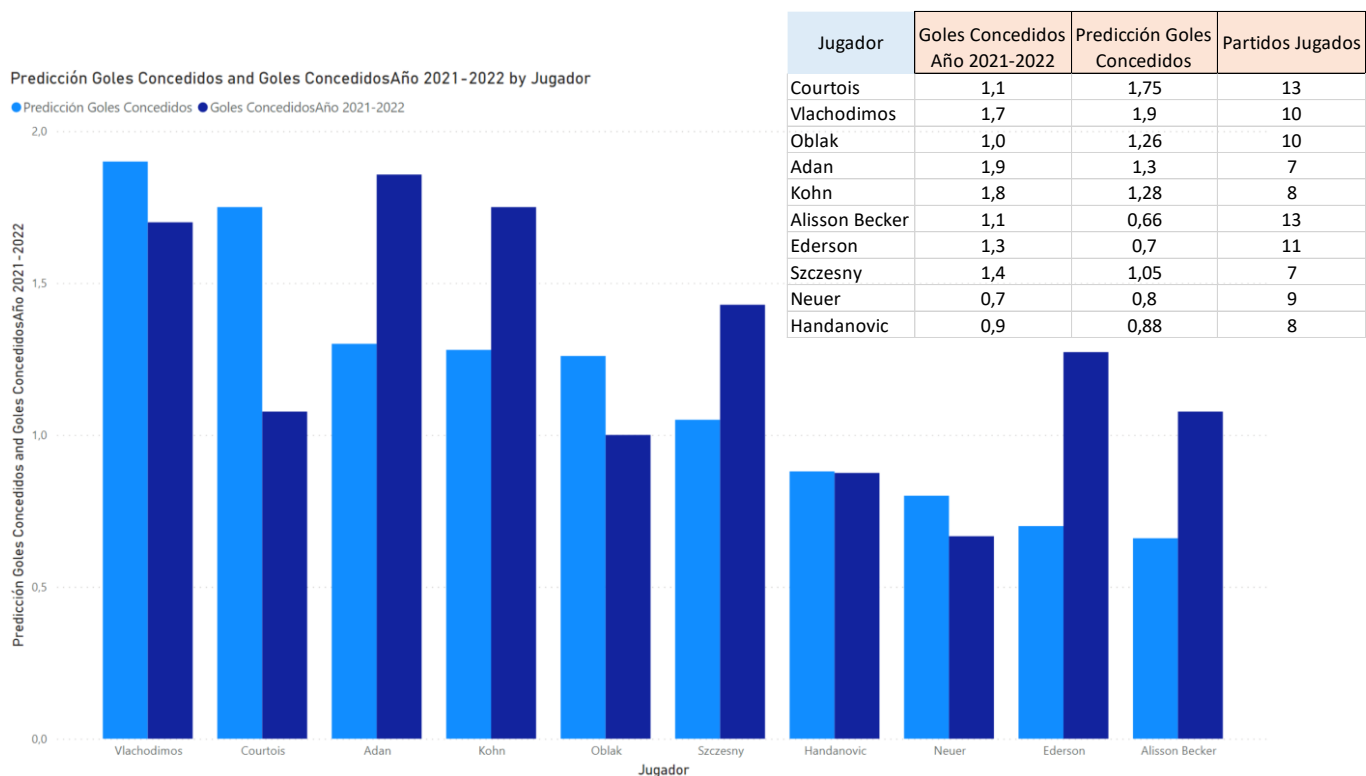


Figura 3.10: Comparativa de la predicción de goles concedidos esta edición vs. los goles concedidos en la edición pasada por los diez mejores porteros de la competición (Elaborado con PowerBi)

El resultado para cada portero es muy distinto, yendo desde casi dos goles por partido a un mínimo de 0,66 goles por partido. Muchos de los porteros ven minorada su estadística, mientras muchos otros ven como su estadística va al alza.

El portero con menos goles concedidos de la competición es **Alisson Becker**.

5.3 Mejor defensa de la competición

Respecto a los defensas, he extraído los datos de los diez defensas con mejores estadísticas de la competición. En concreto, he centrado el foco en la variable balones recuperados. Con balón recuperado, me refiero a robar o tomar el balón cuando la tiene el equipo rival. He filtrado a la mitad el número de defensas a analizar, ya que el resto estaban muy lejos en estadísticas.

Las estadísticas de balones recuperados en la edición pasada fueron las siguientes:

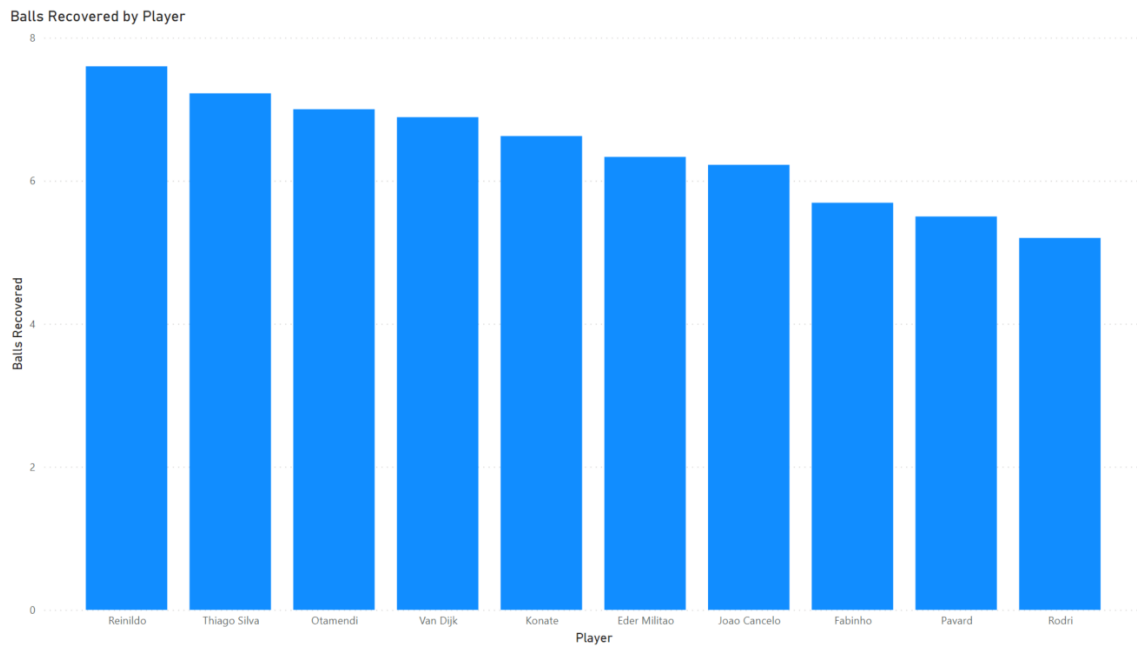


Figura 3.11: Balones recuperados por partido (*Balls Recovered*) por los diez mejores defensas de la edición pasada (Elaborado con PowerBi)

El jugador con más balones recuperados fue Reinildo, seguido por Thiago Silva y Otamendi. En cambio, Rodri, Pavard y Fabinho fueron los jugadores con menos recuperaciones de balón.

Respecto a las entradas realizadas y entradas con éxito, he filtrado el resultado es el siguiente:

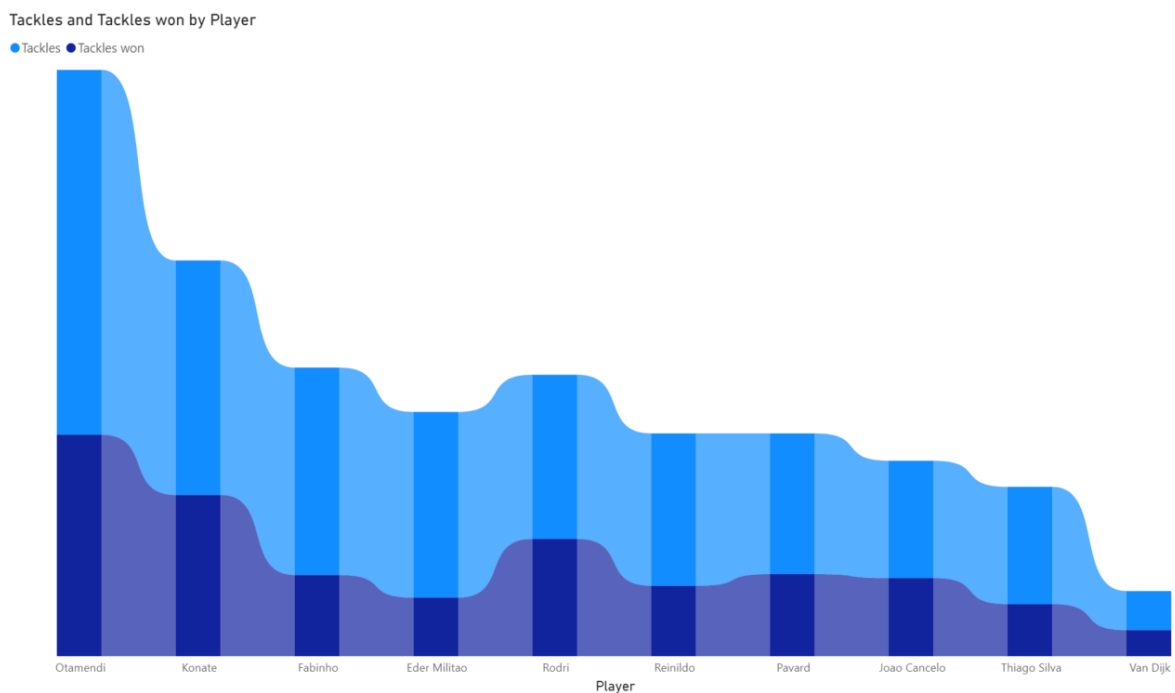


Figura 3.12: Entradas (*tackles*) y Entradas con éxito (*tackles won*) por partido por los diez mejores defensas de la edición pasada (Elaborado con PowerBi)

Para continuar con el análisis por posición, he predicho quién será el mejor defensa de la competición con el mismo modelo utilizado para el máximo goleador y para el mejor portero.

Para los defensas he tenido en cuenta las siguientes variables:

- Balones recuperados (*Balls Recovered*)
- Entradas realizadas (*Tackles*)
- Entradas realizadas con éxito (*Tackles won*)

En cambio, yendo a la variable a calcular, el número de balones recuperados por los diez defensas analizados en la edición pasada fue el siguiente:

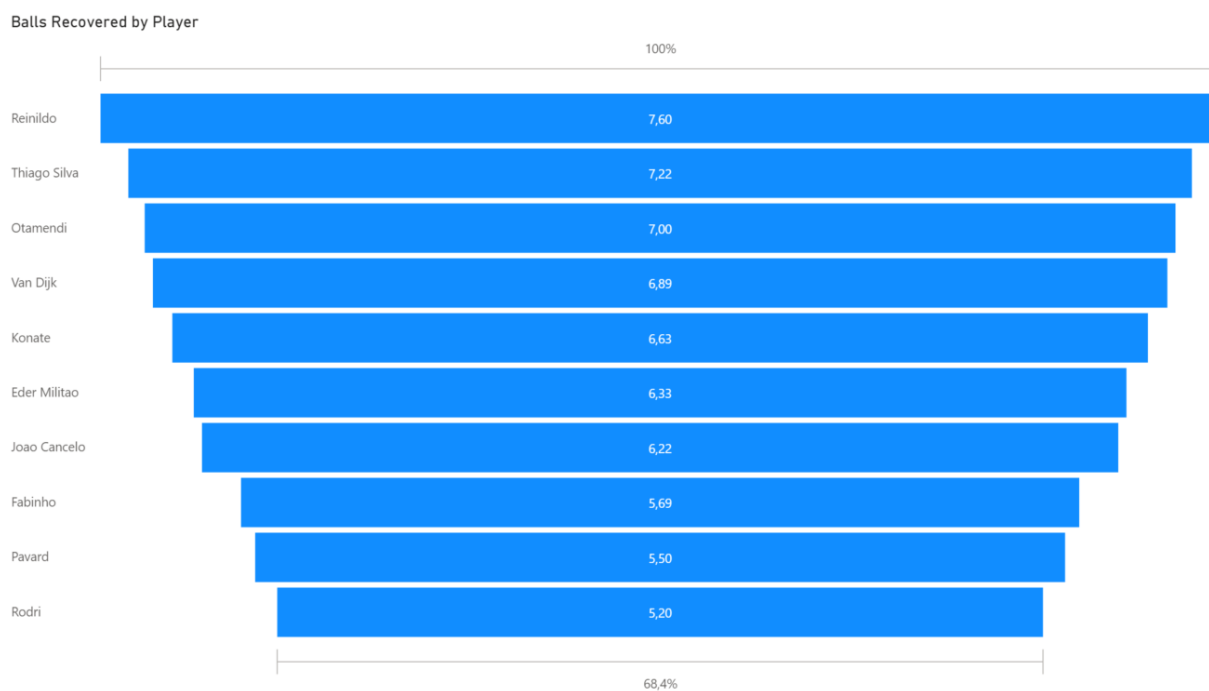


Figura 3.13: Balones recuperados (*Balls Recovered*) por partido por los diez mejores defensas de la edición pasada (Elaborado con PowerBi)

El código que he utilizado en este caso es el mismo que para los dos apartados anteriores⁵². El resultado obtenido en la predicción es el siguiente:

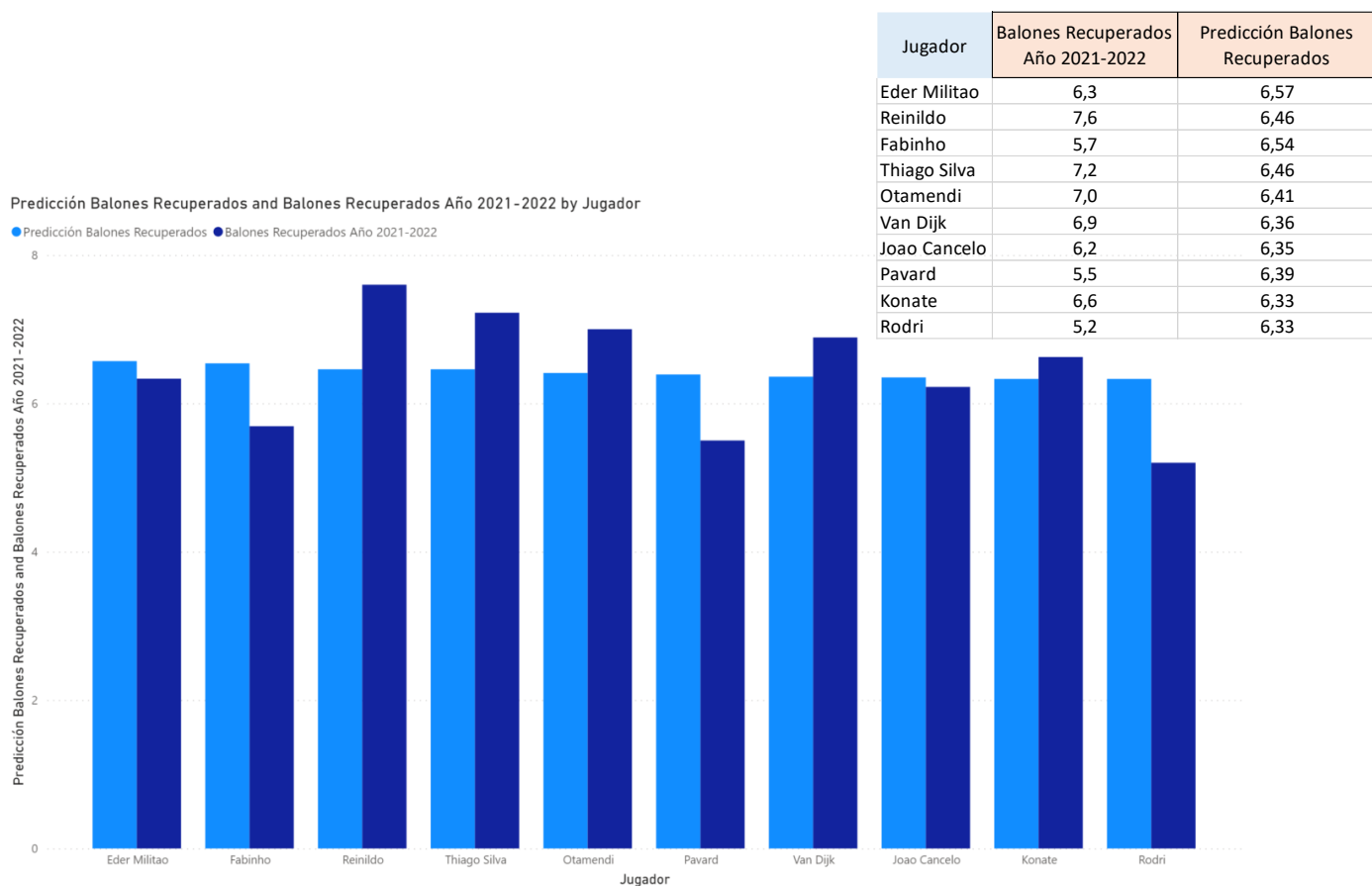


Figura 3.14: Comparativa de la predicción de balones recuperados vs. balones recuperados en la edición pasada por los diez mejores defensas de la competición (Elaborado con PowerBi)

El resultado es muy ajustado entre los diez defensas, ya que las variables independientes, entradas por partido y entradas con éxito por partido, eran muy similares para todos los jugadores analizados.

El jugador con más balones recuperados de la competición será **Eder Militao**.

⁵² Ver anexo 59.

6. Conclusiones

En el mundo del fútbol, la predicción del futuro siempre ha sido un tema muy debatido. Sin embargo, con la creciente cantidad de datos disponibles y de las herramientas de análisis adecuadas, se ha vuelto más fácil hacer predicciones precisas. Concretamente, en este trabajo, he utilizado Python y PowerBi como herramientas principales para conseguir los objetivos: predecir quién será el campeón, el máximo goleador, el mejor portero y el mejor defensa de la UEFA Champions League.

El análisis de datos es una disciplina que tiene como objetivo extraer información útil y conocimiento de conjuntos de datos grandes y complejos. En este proyecto, he utilizado Python para recopilar y procesar los datos de las últimas temporadas de la Champions League. Por otro lado, he utilizado PowerBi para visualizar y analizar los datos de manera más efectiva.

En primer lugar, he analizado la eficacia de cada equipo en las últimas temporadas. Para ello, se han considerado factores como el número de goles marcados y el número de goles encajados. Después de analizar los datos, he concluido que el equipo que tiene más probabilidades de ganar la Champions League en la próxima temporada es el Bayern de Múnich. El club alemán se impondría por poco al Manchester City en una hipotética final.

En segundo lugar, he analizado la eficacia de los jugadores, en particular y como si su actuación fuese ajena a la de su equipo, en la última edición de la UEFA Champions League. Respecto al máximo goleador de la competición, he predicho que será Karim Benzema, delantero del Real Madrid. Benzema es uno de los jugadores más talentosos de la historia del fútbol y ha demostrado ser un goleador implacable en la Champions League en las últimas temporadas, siendo el máximo goleador de la competición en la edición pasada y consiguiendo el trofeo a mejor futbolista del mundo, el Balón de Oro.

Respecto al mejor portero, el análisis también ha revelado que el mejor portero de la competición será Alisson Becker, portero del Liverpool de Inglaterra. Alisson Becker ha demostrado una consistencia impresionante en las últimas temporadas y fue uno de los principales protagonistas de la edición de la UEFA Champions League en la que el Liverpool se proclamó campeón en la temporada 2018 – 2019.

Finalmente, en cuanto al mejor defensa, he concluido que Eder Militao, del Real Madrid, es el jugador más destacado. Eder Militao es conocido por su físico imponente y sus habilidades defensivas excepcionales, lo que lo convierte en uno de los defensores más temidos en el fútbol mundial. Se consagró como un jugador imprescindible en el Real Madrid tras conseguir la edición pasada de la Champions League.

Es importante tener en cuenta que, aunque los datos y el análisis son importantes para hacer predicciones precisas, el fútbol es un deporte muy impredecible y cualquier cosa puede suceder en un partido. Además, las lesiones y otros factores imprevistos pueden cambiar el rumbo de la competición en cualquier momento. Este tipo de variables no han sido contempladas en el análisis ya que serían innumerables las variables a tener en cuenta.

Por ello, las predicciones realizadas en este trabajo no son más que hipótesis que podrían suceder basadas en un modelo matemático. Estas hipótesis tienen altas probabilidades de suceder, pero como todo lo que sucederá en el futuro, no sabemos si finalmente sucederán o no.

7. Anexos

- Anexo 1

Pos	Club	Country	Part	Titles	Pld	W	D	L	F	A	Pts	GD
30	Manchester City FC	ENG	13	0	112	62	21	29	227	128	145	99

- Anexo 2

Pos	Club	Country	Part	Titles	Pld	W	D	L	F	A	Pts	GD
6	Liverpool FC	ENG	27	6	246	142	50	54	470	222	334	248

- Anexo 3

Pos	Club	Country	Part	Titles	Pld	W	D	L	F	A	Pts	GD
12	Chelsea FC	ENG	19	2	197	103	53	41	340	176	259	

- Anexo 4

Pos	Club	Country	Part	Titles	Pld	W	D	L	F	A	Pts	GD
64	Tottenham Hotspur	ENG	7	0	61	28	12	21	116	89	68	27

- Anexo 5

Pos	Club	Country	Part	Titles	Pld	W	D	L	F	A	Pts	GD
1	Real Madrid CF	ESP	53	14	470	281	80	109	1036	514	642	522

- Anexo 6

Pos	Club	Country	Part	Titles	Pld	W	D	L	F	A	Pts	GD
3	FC Barcelona	ESP	33	5	339	197	76	66	667	343	470	324

- Anexo 7

Pos	Club	Country	Part	Titles	Pld	W	D	L	F	A	Pts	GD
16	Club Atlético de Madrid	ESP	18	0	160	76	42	42	226	152	194	74

- Anexo 8

Pos	Club	Country	Part	Titles	Pld	W	D	L	F	A	Pts	GD
58	Sevilla FC	ESP	10	0	72	30	19	23	110	103	79	7

- Anexo 9

Pos	Club	Country	Part	Titles	Pld	W	D	L	F	A	Pts	GD
7	AC Milan	ITA	30	7	261	129	66	66	434	247	324	187

- Anexo 10

Pos	Club	Country	Part	Titles	Pld	W	D	L	F	A	Pts	GD
15	FC Internazionale Milano	ITA	24	3	198	94	52	52	281	200	240	81

- Anexo 11

Pos	Club	Country	Part	Titles	Pld	W	D	L	F	A	Pts	GD
5	Juventus	ITA	37	2	301	153	70	78	479	301	376	178

- Anexo 12

Pos	Club	Country	Part	Titles	Pld	W	D	L	F	A	Pts	GD
63	SSC Napoli	ITA	10	0	58	27	15	16	100	71	69	29

- Anexo 13

Pos	Club	Country	Part	Titles	Pld	W	D	L	F	A	Pts	GD
2	FC Bayern München	GER	39	6	378	227	75	76	800	369	529	431

- Anexo 14

Pos	Club	Country	Part	Titles	Pld	W	D	L	F	A	Pts	GD
17	Borussia Dortmund	GER	21	1	168	80	33	55	286	212	193	74

- Anexo 15

Pos	Club	Country	Part	Titles	Pld	W	D	L	F	A	Pts	GD
95	RB Leipzig	GER	5	0	36	18	4	14	65	62	40	3

- Anexo 16

Pos	Club	Country	Part	Titles	Pld	W	D	L	F	A	Pts	GD
37	Bayer 04 Leverkusen	GER	13	0	117	43	26	48	171	180	112	-9

- Anexo 17

Pos	Club	Country	Part	Titles	Pld	W	D	L	F	A	Pts	GD
155	Eintracht Frankfurt	GER	2	0	13	7	3	3	30	23	17	7

- Anexo 18

Pos	Club	Country	Part	Titles	Pld	W	D	L	F	A	Pts	GD
20	Paris Saint-Germain	FRA	16	0	141	77	27	37	283	163	181	120

- Anexo 19

Pos	Club	Country	Part	Titles	Pld	W	D	L	F	A	Pts	GD
38	Olympique de Marseille	FRA	16	1	114	46	19	49	161	139	111	22

- Anexo 20

Pos	Club	Country	Part	Titles	Pld	W	D	L	F	A	Pts	GD
8	SL Benfica	POR	42	2	283	128	66	89	465	328	322	137

- Anexo 21

Pos	Club	Country	Part	Titles	Pld	W	D	L	F	A	Pts	GD
9	FC Porto	POR	37	2	267	121	60	86	395	303	302	92

- Anexo 22

Pos	Club	Country	Part	Titles	Pld	W	D	L	F	A	Pts	GD
53	Sporting Clube de Portugal	POR	23	0	106	31	22	53	140	178	84	-38

- Anexo 23

Pos	Club	Country	Part	Titles	Pld	W	D	L	F	A	Pts	GD
10	AFC Ajax	NED	39	4	247	112	64	71	396	282	288	114

- Anexo 24

Pos	Club	Country	Part	Titles	Pld	W	D	L	F	A	Pts	GD
14	Celtic FC	SCO	37	1	222	101	39	82	337	270	241	67

- Anexo 25

Pos	Club	Country	Part	Titles	Pld	W	D	L	F	A	Pts	GD
22	Rangers FC	SCO	32	0	173	64	41	68	242	248	169	-6

- Anexo 26

Pos	Club	Country	Part	Titles	Pld	W	D	L	F	A	Pts	GD
36	Club Brugge	BEL	23	0	125	43	30	52	156	179	116	-23

- Anexo 27

Pos	Club	Country	Part	Titles	Pld	W	D	L	F	A	Pts	GD
78	FC Viktoria Plzeň	CZE	9	0	53	22	9	22	82	105	53	-23

- Anexo 28

Pos	Club	Country	Part	Titles	Pld	W	D	L	F	A	Pts	GD
45	FC Salzburg	AUT	18	0	86	35	22	29	125	115	92	10

- Anexo 29

Pos	Club	Country	Part	Titles	Pld	W	D	L	F	A	Pts	GD
25	GNK Dinamo Zagreb	CRO	24	0	156	65	32	59	230	217	162	13

- Anexo 30

Pos	Club	Country	Part	Titles	Pld	W	D	L	F	A	Pts	GD
27	FC Shakhtar Donetsk	UKR	23	0	152	58	36	58	214	231	152	-17

- Anexo 31

Pos	Club	Country	Part	Titles	Pld	W	D	L	F	A	Pts	GD
43	F.C. Copenhagen	DEN	16	0	94	38	23	33	126	113	99	13

- Anexo 32

Pos	Club	Country	Part	Titles	Pld	W	D	L	F	A	Pts	GD
85	Maccabi Haifa FC	ISR	10	0	52	18	9	25	88	88	45	0

Código

Equipos:

- Anexo 33

```
import pandas as pd
import pickle
from scipy.stats import poisson
```

- Anexo 34

```
df_table=pd.read_csv('C:/Users/inigo/Downloads/groups_champions.csv',sep=',')
df_cl=pd.read_csv('C:/Users/inigo/Downloads/champions_league_2020_2022.csv',sep=',')
df_fixtures=pd.read_csv('C:/Users/inigo/Downloads/clean_champions_league_fixtures.csv',sep=',')
```

- Anexo 35

```
df_home=df_c1[['Home Team','Home Goals',' Away Goals']]
df_away=df_c1[['Away Team','Home Goals',' Away Goals']]
```

- Anexo 36

```
df_home=df_home.rename(columns={'Home Team': 'Team','Home Goals': 'GoalsScored', ' Away Goals': 'GoalsConceded'})
df_away=df_away.rename(columns={'Away Team': 'Team','Home Goals': 'GoalsConceded', ' Away Goals': 'GoalsScored'})
```

- Anexo 37

```
df_team_strength=pd.concat([df_home,df_away],ignore_index=True).groupby('Team').mean()
df_team_strength
```

- Anexo 38

```
def predict_points(home,away):
    if home in df_team_strength.index and away in df_team_strength.index:
        lamb_home=df_team_strength.at[home,'GoalsScored'] * df_team_strength.at[away,'GoalsConceded']
        lamb_away=df_team_strength.at[away,'GoalsScored'] * df_team_strength.at[home,'GoalsConceded']
        prob_home,prob_away,prob_draw=0,0,0
        for x in range(0,11):
            for y in range(0,11):
                p=poisson.pmf(x,lamb_home) * poisson.pmf(y,lamb_away)
                if x==y:
                    prob_draw += p
                elif x>y:
                    prob_home += p
                else:
                    prob_away += p

        points_home= 3 * prob_home + prob_draw
        points_away= 3 * prob_away + prob_draw
        return (points_home,points_away)
    else:
        return(0,0)
```

- Anexo 39

```
predict_points('Real Madrid','Leipzig')
(2.3034428455066207, 0.56624904295391)
```

- Anexo 40

Grupo A

```
In [17]: predict_points('Ajax','Liverpool')
Out[17]: (1.2237270864900394, 1.5682720925638023)
```

```
In [18]: predict_points('Liverpool','Ajax')
Out[18]: (1.5682720925638023, 1.2237270864900398)
```

- Anexo 41

Grupo B

```
In [19]: M predict_points('Porto','Atlético')
Out[19]: (1.562006632831657, 1.170827742269591)

In [20]: M predict_points('Porto','Club Brugge')
Out[20]: (2.201385441973947, 0.6424796570679254)

In [21]: M predict_points('Atlético','Club Brugge')
Out[21]: (1.9102427163597737, 0.8884037895697865)
```

- Anexo 42

Grupo C

```
In [22]: M predict_points('Bayern','Barcelona')
Out[22]: (2.608495616917506, 0.29334083273531136)

In [23]: M predict_points('Bayern','Inter')
Out[23]: (2.31128656217372, 0.5255773943960045)

In [24]: M predict_points('Inter','Barcelona')
Out[24]: (1.5627737863845748, 1.1859659550948782)
```

- Anexo 43

Grupo D

```
In [25]: M predict_points('Marseille','Sporting CP')
Out[25]: (0.16446532350065435, 2.761767651081061)
```

- Anexo 44

Grupo E

```
In [26]: M predict_points('Milan','Chelsea')
Out[26]: (0.2713096274810588, 2.6146777405695043)

In [27]: M predict_points('Milan','Salzburg')
Out[27]: (1.4408965153826756, 1.3614876344809563)

In [28]: M predict_points('Chelsea','Salzburg')
Out[28]: (2.7722702163006416, 0.155109662226995)
```

- Anexo 45

Grupo F

```
In [29]: M predict_points('Real Madrid','Shakhtar Donetsk')
Out[29]: (2.786237206209474, 0.14490183697090503)

In [30]: M predict_points('Real Madrid','Leipzig')
Out[30]: (2.3034428455066207, 0.56624904295391)

In [31]: M predict_points('Leipzig','Shakhtar Donetsk')
Out[31]: (2.543495275359623, 0.3483697695654691)
```

- Anexo 46

Grupo G

```
In [32]: M predict_points('Man. City','Sevilla')
Out[32]: (2.3340952610451593, 0.5142394948297218)

In [33]: M predict_points('Man. City','Dortmund')
Out[33]: (2.271951585881175, 0.5844581597742338)

In [34]: M predict_points('Dortmund','Sevilla')
Out[34]: (1.62621007539121, 1.1769005433198632)
```

- Anexo 47

Grupo H

```
In [35]: M predict_points('Paris','Juventus')
Out[35]: (1.2515835640739037, 1.55406370156956)

In [36]: M predict_points('Paris','Benfica')
Out[36]: (2.0812673461992457, 0.7604888905278865)

In [37]: M predict_points('Juventus','Benfica')
Out[37]: (2.1841447470859516, 0.6618384599635807)
```

- Anexo 48

```
df_leagues=pd.read_csv('C:/Users/inigo/Downloads/leagues_2021_2022.csv',sep=';')
print(df_leagues)
```

- Anexo 49

```
def predict_points2(home,away):
    if home in df_team_strength2.index and away in df_team_strength2.index:
        lamb_home=df_team_strength2.at[home,'GoalsScored'] * df_team_strength2.at[away,'GoalsConceded']
        lamb_away=df_team_strength2.at[away,'GoalsScored'] * df_team_strength2.at[home,'GoalsConceded']
        prob_home,prob_away,prob_draw=0,0,0
        for x in range(0,11):
            for y in range(0,11):
                p=poisson.pmf(x,lamb_home) * poisson.pmf(y,lamb_away)
                if x==y:
                    prob_draw += p
                elif x>y:
                    prob_home += p
                else:
                    prob_away += p

        points_home= 3 * prob_home + prob_draw
        points_away= 3 * prob_away + prob_draw
        return (points_home,points_away)
    else:
        return(0,0)
```

- Anexo 50

```
predict_points2('Real Madrid', 'RB Leipzig')  
(1.7321207260607943, 1.0739459056491834)
```

- Anexo 51

Grupo A

```
In [45]: M predict_points2('Liverpool', 'Napoli')  
Out[45]: (1.717113079223523, 1.0691508153356741)
```

Grupo B

```
In [46]: M predict_points2('Atlético de Madrid', 'Bayer 04 Leverkusen')  
Out[46]: (1.3566725276661964, 1.4640594159503568)
```

Grupo C

```
In [47]: M predict_points2('FC Bayern', 'FC Barcelona')  
Out[47]: (1.8589203253234337, 0.9668791461598575)
```

```
In [48]: M predict_points2('FC Bayern', 'Inter')  
Out[48]: (1.4861899503574296, 1.3269571955754107)
```

```
In [49]: M predict_points2('FC Barcelona', 'Inter')  
Out[49]: (1.0724496694300794, 1.7196668623133913)
```

Grupo D

```
In [50]: M predict_points2('Eintracht Frankfurt', 'Spurs')  
Out[50]: (0.8147565814476668, 2.0065669356685936)
```

```
In [51]: M predict_points2('Eintracht Frankfurt', 'Olympique de Marseille')  
Out[51]: (0.7902072032747466, 2.023291139296107)
```

```
In [52]: M predict_points2('Spurs', 'Olympique de Marseille')  
Out[52]: (1.3339052554791944, 1.4450846170883063)
```

Grupo E

```
In [53]: M predict_points2('Milan', 'Chelsea')  
Out[53]: (1.3340966578607498, 1.421698199635595)
```

- Anexo 52

Grupo F

```
In [54]: M predict_points2('Real Madrid', 'RB Leipzig')
Out[54]: (1.7321207260607943, 1.0739459056491834)
```

Grupo G

```
In [55]: M predict_points2('Man City', 'Borussia Dortmund')
Out[55]: (2.4475309440965867, 0.433573952612429)
```

```
In [56]: M predict_points2('Man City', 'Sevilla FC')
Out[56]: (2.0744375239096544, 0.7261092452442461)
```

```
In [57]: M predict_points2('Sevilla FC', 'Borussia Dortmund')
Out[57]: (1.5518482366403419, 1.2587360892864392)
```

Grupo H

```
In [58]: M predict_points2('Juventus', 'Paris Saint-Germain')
Out[58]: (0.8281688120191738, 1.9736894388354642)
```

- Anexo 53

Octavos de Final

```
In [59]: M predict_points2('Chelsea', 'Bayer 04 Leverkusen')
Out[59]: (1.9256019272775273, 0.9058582338784997)
```

```
In [60]: M predict_points2('Paris Saint-Germain', 'Spurs')
Out[60]: (1.9654880337374419, 0.8535747115692904)
```

```
In [61]: M predict_points('Real Madrid', 'Juventus')
Out[61]: (1.4727871470335379, 1.325649706458055)
```

```
In [62]: M predict_points2('Real Madrid', 'Juventus')
Out[62]: (1.839211478230238, 0.9429026680444114)
```

```
In [63]: M predict_points('Man. City', 'Inter')
Out[63]: (2.006412009555609, 0.7727445700021497)
```

```
In [64]: M predict_points2('Man City', 'Inter')
Out[64]: (1.7143813406354749, 1.078739658811259)
```

```
In [65]: M predict_points('Sporting CP', 'Milan')
Out[65]: (1.6727181686123287, 1.1426789435127533)
```

```
In [66]: M predict_points('Bayern', 'Ajax')
Out[66]: (2.096646866976798, 0.7448781085827246)
```

```
In [67]: M predict_points('Liverpool', 'Leipzig')
Out[67]: (2.5131191848527275, 0.37655400500959946)
```

```
In [68]: M predict_points2('Liverpool', 'RB Leipzig')
Out[68]: (2.036985155877384, 0.7924857716716431)
```

```
In [69]: M predict_points('Porto', 'Dortmund')
Out[69]: (1.1230028331080362, 1.676331002659862)
```

- Anexo 54

Cuartos de Final

```
In [70]: M predict_points('Sporting CP', 'Liverpool')
Out[70]: (0.3523886577042903, 2.5406473312944424)

In [71]: M predict_points('Man. City', 'Chelsea')
Out[71]: (1.1837040574815925, 1.572480249763801)

In [72]: M predict_points2('Man City', 'Chelsea')
Out[72]: (1.7705891187528038, 1.0191808224941679)

In [73]: M predict_points('Dortmund', 'Bayern')
Out[73]: (0.3217561631085966, 2.5742108306574827)

In [74]: M predict_points2('FC Bayern', 'Borussia Dortmund')
Out[74]: (2.283894583067695, 0.5651678183910827)

In [75]: M predict_points('Paris', 'Real Madrid')
Out[75]: (1.1758914673837082, 1.6345761069732185)

In [76]: M predict_points2('Paris Saint-Germain', 'Real Madrid')
Out[76]: (1.5240641918853133, 1.2688912430522967)
```

- Anexo 55

Semifinales

```
In [80]: M predict_points('Man. City', 'Liverpool')
Out[80]: (1.533474119419077, 1.245931077273442)

In [82]: M predict_points2('Man City', 'Liverpool')
Out[82]: (1.4905283533608369, 1.2976654913633563)

In [83]: M predict_points('Real Madrid', 'Bayern')
Out[83]: (0.6332907977213117, 2.219376453824724)

In [84]: M predict_points2('Real Madrid', 'FC Bayern')
Out[84]: (1.2532546448109059, 1.560681393692978)
```

- Anexo 56

Final

```
In [85]: M predict_points('Man. City', 'Bayern')
Out[85]: (1.0706927692901849, 1.7366649027822065)

In [86]: M predict_points2('Man City', 'FC Bayern')
Out[86]: (1.6967316834809285, 1.122208564466197)
```


Jugadores:

- Anexo 57

Benzema

```
In [30]: import numpy as np
        from sklearn.linear_model import LinearRegression

In [181]: goals_scored=15
         shots_attempted=45
         shots_target=23

In [183]: data=np.array ([[15,45,23],[13,42,22],[11,37,18],[8,33,13],[7,31,19],[7,29,11],[6,7,28],[6,27,16],[6,27,8],[5,26,11],[5,26,16],
        <----->

In [184]: X=data[:,1:]
         y=data[:,0]

In [185]: model=LinearRegression()
         model.fit(X,y)

Out[185]: LinearRegression()

In [186]: X_new=np.array([[shots_attempted,shots_target]])
         prediction=model.predict(X_new)

In [187]: print(f"The predicted number of goals for the striker is:{prediction[0]:.2f}")

The predicted number of goals for the striker is:13.70
```

- Anexo 58

Courtois

```
In [215]: saves=4.7
         goals_conceded=1.1
         clean_sheets=5

In [216]: datos=np.array ([[1.1,4.7,5],[1.7,3.8,1],[3.3,4.7,0],[1.2,6,3],[1.8,3,2],[3,6,0],[1.2,3.6,2],[1.9,2.4,2],[1.8,2,1],[1.5,2.7,2],
        <----->

In [217]: z=datos[:,1:]
         a=datos[:,0]

In [218]: modelo=LinearRegression()
         modelo.fit(z,a)

Out[218]: LinearRegression()

In [221]: z_new=np.array([[saves,clean_sheets]])
         predic=modelo.predict(z_new)

In [222]: print(f"The predicted number of goals conceded for the goalkeeper is:{predic[0]:.2f}")

The predicted number of goals conceded for the goalkeeper is:1.75
```

- Anexo 59

Reinildo

```
In [397]: M balls_recovered1=7.6
          tackles1=1.3
          tackles_won1=0.6

In [398]: M dat1=np.array ([[6.3,1.6,0.5],[7.6,1.3,0.6],[5.7,1.8,0.7],[7.2,1,0.4],[7,3.1,1.9],[6.9,0.3,0.2],[6.2,1,0.7],[5.5,1.2,0.7],[6.
          <
          >

In [399]: M h1=dat1[:,1:]
          i1=dat1[:,0]

In [400]: M mol=LinearRegression()
          mol.fit(h1,i1)

Out[400]: LinearRegression()

In [401]: M h_new1=np.array([[tackles1,tackles_won1]])
          pred1=mol.predict(h_new1)

In [402]: M print(f"The predicted number of balls recovered by the defender is:{pred1[0]:.2f}")
          The predicted number of balls recovered by the defender is:6.46
```

8. Bibliografía

- *Download Football/Soccer fixtures, schedules and results as CSV, XLSX, ICS and JSON / Fixture Download.* (s. f.). <https://fixturedownload.com/sport/football>
- UEFA.com. (s. f.). *Season 2021/22 Club stats | UEFA Champions League 2021/22.* Recuperado 26 de octubre de 2022, de <https://www.uefa.com/uefachampionsleague/history/seasons/2022/statistics/clubs/>
- UEFA.com. (s. f.-b). *Season 2021/22 Player stats | UEFA Champions League 2021/22.* Recuperado 26 de octubre de 2022, de <https://www.uefa.com/uefachampionsleague/history/seasons/2022/statistics/players/>
- Diario AS. (s. f.). *Ranking equipos LaLiga Santander 2021/2022.* AS.com. Recuperado 26 de octubre de 2022, de https://resultados.as.com/resultados/futbol/primera/2021_2022/ranking/
- Diario AS. (s. f.-b). *Rankings jugadores LaLiga Santander 2021/2022.* AS.com. Recuperado 26 de octubre de 2022, de https://resultados.as.com/resultados/futbol/primera/2021_2022/ranking/jugadores/
- *Índice de Deportistas y Equipos que empiezan por p | AS.com.* (2022, 26 octubre). <https://resultados.as.com/resultados/ficha/listado/>
- *Premier League First Team Club Statistics, Team & Player Stats.* (s. f.). Recuperado 26 de octubre de 2022, de <https://www.premierleague.com/stats>
- *Ligue 1 Uber Eats - Official statistics - - Gãñnãñral.* (s. f.). Ligue1. Recuperado 26 de octubre de 2022, de <https://www.ligue1.com/stats>
- *Teams Statistics | Lega Serie A.* (s. f.). Recuperado 26 de octubre de 2022, de <https://www.legaseriea.it/en/serie-a/teams-statistics>

- *Player Statistics / Lega Serie A*. (s. f.). Recuperado 26 de octubre de 2022, de <https://www.legaseriea.it/en/serie-a/statistics>
- *Serie A 2021/22* :: *ceroacero.es*. (s. f.). www.ceroacero.es. Recuperado 26 de octubre de 2022, de <https://www.ceroacero.es/edition.php?id=156515>
- Diario AS. (s. f.-b). *Ranking equipos Serie A 2021/2022*. AS.com. Recuperado 26 de octubre de 2022, de https://resultados.as.com/resultados/futbol/italia/2021_2022/ranking/
- Zeileis, A. (2022, 14 noviembre). *Machine learning of a 2022 FIFA World Cup multiverse*. Achim Zeileis. <https://www.zeileis.org/news/fifa2022/>
- *A Mathematician's Guide to the World Cup* | *Mathematical Institute*. (s. f.). <https://www.maths.ox.ac.uk/node/61756>
- Icaro. (2021, 15 diciembre). Creating an ML model for predicting English Premiership results (or using Machine Learning to figure out if Arsenal will return to the Champions League in 2020–2021). *Medium*. <https://medium.com/analytics-vidhya/creating-an-ml-model-for-predicting-english-premiership-results-or-using-machine-learning-to-875b369c1b48>
- *Club Soccer Predictions*. (2023, 20 abril). FiveThirtyEight. <https://projects.fivethirtyeight.com/soccer-predictions/>
- Sblendorio, D. (2021). How to predict NFL Winners with Python. *ActiveState*. <https://www.activestate.com/blog/how-to-predict-nfl-winners-with-python/>