



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA



Grado en Ciencia de Datos

## PROYECTO FÚTBOL

Análisis estadístico e inferencial a partir de varias fuentes de recogida  
de datos del mundo del fútbol

**AUTORES:** Josep Carles Vega  
Joan Castillo Esteve  
Didac Cerdà Falcó  
Dani Goig Martínez  
Vicente Torres Martínez

**AÑO ACADÉMICO:** 2019/2020

## ÍNDICE

- I. [INTRODUCCIÓN/ALCANCE DEL PROYECTO](#)
  - II. [OBJETIVOS](#)
  - III. [III. CONJUNTO DE DATOS/FUENTES DE DATOS](#)
  - IV. [INTEGRACIÓN Y TRANSFORMACIÓN DE LOS DATOS](#)
  - V. [HERRAMIENTAS Y TÉCNICAS UTILIZADAS](#)
  - VI. [RESULTADOS/CONCLUSIONES](#)
  - VII. [LECCIONES APRENDIDAS PARA FUTUROS PROYECTOS DE CIENCIA DE DATOS](#)
  - VIII. [BIBLIOGRAFÍA](#)
  - IX. [ANEXOS](#)
- 

### [I. INTRODUCCIÓN/ALCANCE DEL PROYECTO](#)

El proyecto que estamos realizando pretende hacer un análisis a fondo de 5 temporadas de la Liga Santander, estudiando los jugadores, los resultados de los partidos disputados, profundizando más en los Clásicos, ya que estos son los partidos más mediáticos a nivel mundial.

La principal restricción que encontramos en este proyecto, es la tremenda dificultad de encontrar datos relacionados con el mundo del fútbol. En este aspecto, llegamos incluso a contactar con profesionales en busca de datos y todos ellos nos respondieron que el negocio de los datos relacionados con el fútbol va en aumento y que habría que pagar una suma mínimo de 10.000 euros para hacernos con una buena base de datos acorde con nuestros objetivos. Además, en la misma línea de este problema, para solucionarlo, íbamos a asistir a una charla impartida en la Universidad Politécnica de Valencia relacionada con el mundo del fútbol y de la que esperábamos sacar ideas y datos. Sin embargo, debido a la pandemia mundial causada por el virus del Covid-19 y que nos obligó a trabajar desde casa, no pudimos asistir. Así como restricciones menores, tuvimos que ponernos de acuerdo mediante reuniones realizadas a través de skype para valorar subjetivamente a los jugadores y formar una matriz de comparación pareada.

---

### **II. OBJETIVOS**

Tras debatir entre los miembros del grupo qué objetivos serían los más interesantes para realizar, optamos por escoger varios objetivos para las distintas bases de datos que teníamos.

Para la base de datos de resultados, hemos decidido hacer un análisis global de los equipos visitantes y locales durante las temporadas 2014-2015, 2015-2016, 2016-2017, 2017-2018 y 2018-2019. En dicho análisis comparamos los enfrentamientos directos entre dos equipos concretos y observamos el rendimiento que han tenido estos equipos como local y como visitante.

Otra base que tenemos es la base de los Clásicos. Para esta base hemos marcado varios objetivos.

En primer lugar compara las distintas posiciones (portero, defensa, medio-centro, delantero...) de cada equipo durante los 10 clásicos.

También hemos realizado análisis para ver cuáles han sido los mejores jugadores durante estos 10 partidos y realizar un 11 ideal, es decir mejor equipo formado por los once mejores jugadores, ya sean del Real Madrid o del F.C.Barcelona.

Finalmente, a raíz del resultado anterior, descubrir cuál de los dos equipos, Madrid o Barça, ha sido mejor en los últimos 5 años.

La última base que tenemos es la de jugadores, la cual es la más grande de todas. Al ser la más grande es la base con la que más podemos trabajar por lo que hemos marcado bastantes objetivos. Estos son:

La comparación de los jugadores y estudiar los mejores jugadores por posición. Al realizar estos dos objetivos decidimos introducir un nuevo objetivo, el de determinar si un jugador tiene el nivel adecuado para su equipo o si debería, por un lado, estar en un equipo de un nivel superior o por otro lado, en un equipo con un nivel inferior.

El análisis del rendimiento de los jugadores nos lleva a pensar en estudiar la construcción de un equipo adecuado a sus objetivos o necesidades. Junto al objetivo anterior, podemos ver que fichajes serían adecuados para cada equipo.

Finalmente estudiar las características de los jugadores y compararlos con otros jugadores que juegan en la misma posición y en el mismo equipo para conocer cuál de los dos sería más adecuado que jugara en el equipo en cuestión.

---

### III. CONJUNTO DE DATOS/FUENTES DE DATOS

Hemos utilizado varias bases de datos para realizar este proyecto, estas bases las hemos obtenido de distintas fuentes de datos. Las bases son las siguientes:

- **Base de resultados:** La base de datos resultante tiene un total de 1900 filas o resultados diferentes, 380 de cada temporada. Tiene un total de 26 columnas. Las variables de temporada y fecha sirven para saber cuando se ha disputado el partido. Tiene 2 variables para referirse tanto al equipo local como al equipo visitante. Las variables restantes hacen referencia a diferentes estadísticas (Tiros, Goles, Tarjetas, Faltas, Corners etc) para cada uno de los equipos.  
Estos datos han sido descargados de la página web Datahub.
- **Base de los Clásicos:** El resultado de esta integración es un dataframe de 360 casos y 33 variables. Cada caso, pertenece a un jugador y en las variables, encontramos las estadísticas de cada jugador en cada clásico de los recogidos, la temporada a la que pertenece cada partido y el estadio donde se ha jugado. Algunas de las variables a destacar son nota, goles, asistencias, tiros, pérdidas, tarjetas amarillas y tarjetas rojas entre otras.  
Los datos usados para crear esta base de datos han sido extraídos de la página web Whoscored copiandolos a mano en un documento excel.
- **Bases de jugadores:** Estas bases están separadas por temporadas, ya que era contraproducente juntar todas las temporadas en una misma base ya que para los objetivos marcados queríamos estudiar las temporadas por separado. Estas bases las hemos extraído mediante webscrapping de la página web FBref.  
Como finalmente los objetivos fijados estaban más relacionados con la evolución de los equipos decidimos centrarnos en la última temporada (2018-2019). Para la base ya obtenida añadimos varias tablas de la misma página, estas tablas las obtuvimos copiando un archivo csv que proporciona la página web. Finalmente hemos obtenido una base de datos con 564 casos, es decir, todos los jugadores que jugaron en dicha temporada, y 106 variables. Entre estas variables destacan los goles, asistencias, tarjetas, pases completados, despejes, intercepciones, entre muchas otras que hemos utilizado en nuestro análisis.  
Como ya hemos dicho, esta bases las hemos obtenido mediante webscrapping y documentos csv que proporcionaba la página web.

## IV. INTEGRACIÓN Y TRANSFORMACIÓN DE LOS DATOS.

### ***Base Resultados***

La base de Resultados está formada por cinco bases de datos que tuvimos que juntar, una para cada temporada desde 2014 hasta 2019.

Para esta base lo único que tuvimos que hacer fue cambiar el nombre de las variables para más comodidad, ya que todas las variables eran útiles y no había datos faltantes.

### ***Base Clásicos***

Como la base de Clásicos la obtuvimos de forma manual, no tuvimos que eliminar variables innecesarias o tener en cuenta valores nulos, ya que no los había. Lo que sí que tuvimos que hacer fue crear variables nuevas necesarias para el análisis. Estas son: Goles recibidos, Posición (Portero, Defensa, Mediocentro, Delantero), Goles Por Partido y Asistencias Por Partido.

### ***Base Jugadores***

Lo primero que tuvimos que hacer para la base de jugadores fue juntar varios conjuntos de datos que obtuvimos de internet. Seguidamente eliminamos variables que eran inútiles o estaban repetidas y borramos filas que no eran jugadores. También cambiamos el nombre a las variables, para así poder realizar los análisis de manera más cómoda.

Finalmente, para la base de jugadores creamos varias variables nuevas, como pueden ser, Temporada, Equipo y Posición. También creamos; 5 variables de Valores obtenidos mediante un proceso analítico jerárquico y el nivel del equipo según la posición que quedaron en la clasificación.

---

## V. HERRAMIENTAS Y TÉCNICAS UTILIZADAS.

A lo largo del proyecto hemos utilizado principalmente el lenguaje de programación R, pero también hemos utilizado excel.

En R hemos utilizado muchas librerías como *gridExtra*, *corrplot*, *gplots*, *FactoMineR*, *factoextra*, *dplyr*, *ggplot2*, *RColorBrewer*, *ggthemes*, *ggrepel*, *reshape2*, *data.table*, *devtools*, *VGAM*, *SportsAnalytics270*, *magrittr*, *png*, *ggsignif*, *data.table*, *formattable*, *rvest*, *plyr*, *dplyr*.

Hemos hecho varios análisis PCA para conocer mejor los datos de las bases de datos.

Proceso analítico jerárquico para sacar los valores de los jugadores.

---

## VI. RESULTADOS/CONCLUSIONES

Los resultados/conclusiones de nuestro proyecto los podemos dividir en función de la base de datos:

## 1. Base resultados

Los siguientes gráficos muestran las probabilidades que tiene cada equipo de ganar, perder o empatar.

### TOP 3 La Liga

	Gana Local	Empate	Gana Visitante
Barça-Madrid	0.550	0.165	0.285
Madrid-Barça	0.485	0.176	0.339
Barça-At.Madrid	0.714	0.144	0.142
At.Madrid-Barça	0.303	0.194	0.503
At.Madrid-Madrid	0.313	0.197	0.490
Madrid-At.Madrid	0.666	0.163	0.171

El Atlético de Madrid tiene una baja probabilidad de ganar en todos los encuentros. Esto se puede deber a que estas probabilidades se obtienen por los goles marcados de cada equipo, por lo que el At.Madrid , al ser el equipo que menos goles marca de los tres, tenga poca probabilidad de ganar.

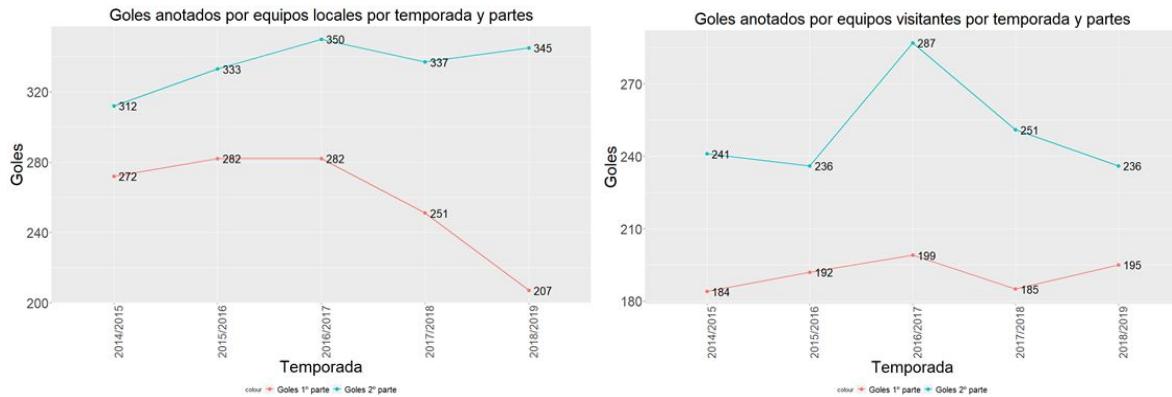
En los Clásicos, la probabilidad de ganar es mayor para el equipo local, siendo mayor para el Barcelona.

### Partidos Importantes Valencia CF

	Gana_Valencia	Empate	Pierde_Valencia
Barça-Valencia/Valencia-Barça	0.199	0.167	0.634
Getafe-Valencia/Valencia-Getafe	0.480	0.253	0.267
At.Madrid-Valencia/Valencia-At.Madrid	0.357	0.231	0.412
Real Madrid-Valencia/Valencia-Real Madrid	0.218	0.179	0.604
Levante-Valencia/Valencia-Levante	0.460	0.250	0.290
Sevilla-Valencia/Valencia-Sevilla	0.372	0.231	0.397

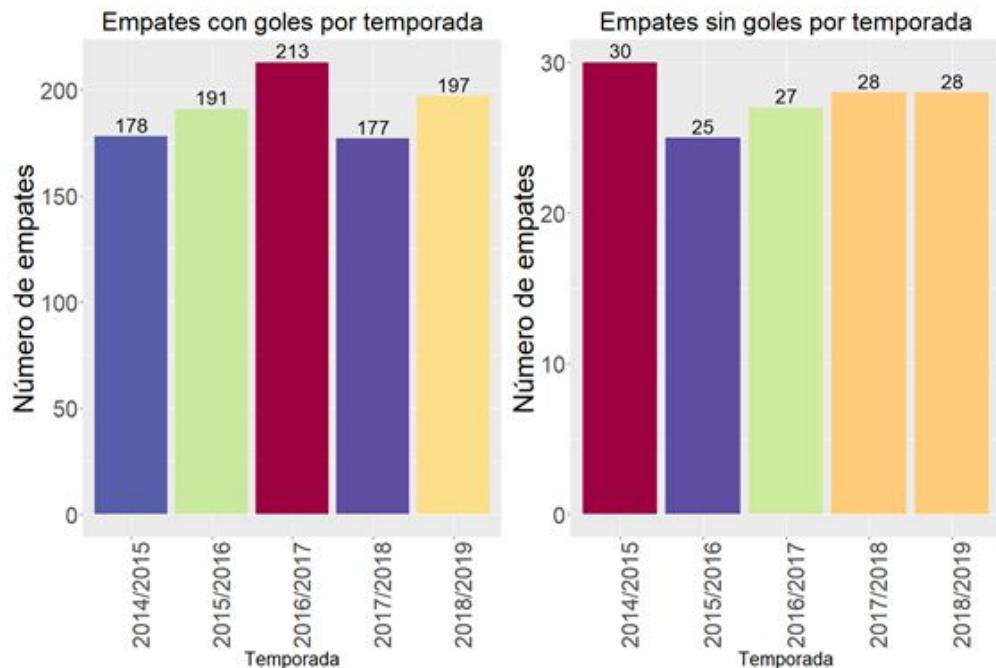
Los partidos entre valencia y sevilla son los más igualados y el equipo que más probabilidades tiene de ganarle al Valencia es el Barcelona.

Los siguientes gráficos mostrarán la evolución de los goles en las diferentes temporadas.



A partir de los siguientes gráficos, obtuvimos el sorprendente resultado de que se marcan más goles en la segunda temporada y, además, se marcan muchos más goles como local que como visitante.

Gráfico de empates con goles/sin goles

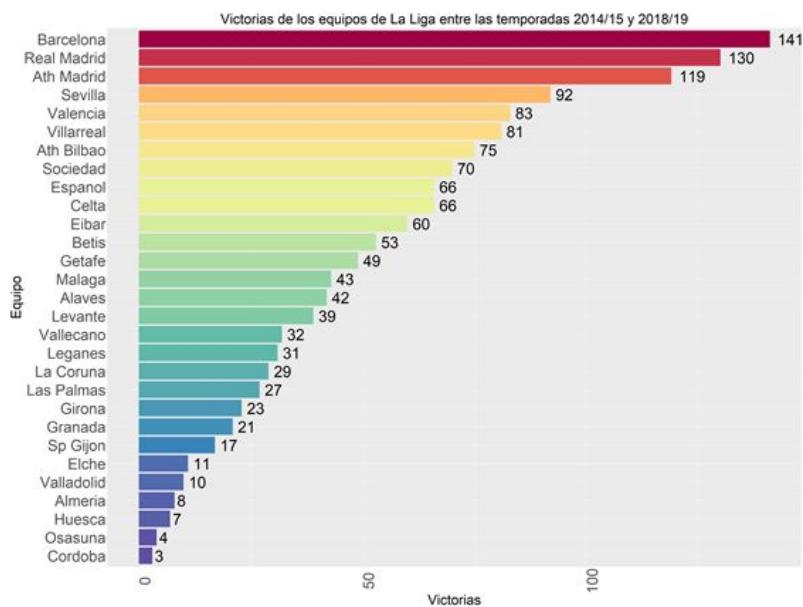


A la vista de los gráficos, observamos que es mejor quedarse en los partidos al final, ya que en todas las temporadas los empates con goles superan a los empates sin goles. Además, la diferencia entre la temporada con más empates con goles, que es la temporada 2016/2017, respecto de la temporada con menos empates sin goles, que es la temporada 2015/2016 es de casi 190 partidos y en todos los casos, la diferencia entre lo anteriormente comentado, supone 150 partidos, hecho que resalta la gran capacidad ofensiva de los equipos españoles, ya que frente al total, los partidos sin goles suponen una cantidad minúscula.

- Victorias totales de los equipos en las últimas 5 temporadas

El Barcelona es el equipo que más victorias ha conseguido en estos últimos 5 años, seguido del Real Madrid y el Atlético de Madrid con una diferencia de 11 victorias entre cada uno de ellos. El cuarto equipo con más victorias es el Sevilla, que tiene 92 victorias. El Valencia y el Villareal son los siguientes y tienen un número de

victorias muy similar (83-81). Los equipos que menos victorias han conseguido, y ya hemos visto antes son el Elche(11), Valladolid(10), Almeria (8), Huesca (7), Osasuna (4) o el Córdoba (3), el equipo con menos victorias.



A continuación, pondremos las conclusiones del análisis de victorias como local y visitante de cada equipo, las derrotas y las derrotas de los equipos como local y visitante. Gráficos de estas conclusiones están en [Anexos Proyecto II](#).

#### - Conclusiones Victorias como local y visitante

El Barcelona es el equipo más sólido tanto en su estadio como fuera de él, siendo el mejor local y el mejor visitante de las últimas 5 temporadas en La Liga. El Real Madrid y el Atlético de Madrid son, aparte del Barcelona, los mejores locales de La Liga con prácticamente las mismas victorias en su estadio (71-70). Esta diferencia se acentúa en los partidos como visitante donde el Real Madrid ha ganado 10 partidos más estos últimos 5 años. El Sevilla ha conseguido 64 victorias como local, pero solo 28 como visitante. Podemos decir que el Sevilla es un equipo muy sólido en su estadio y más débil como visitante. Equipos como el Valencia, Villareal o Athletic de Bilbao son los siguientes equipos en ambas tablas, siendo el número de victorias en casa mucho mayor que el de fuera. El Córdoba, que sólo ha disputado 1 de esas últimas 5 temporadas sólo ha logrado una victoria en casa, y dos fuera de su estadio. Otros equipos como Osasuna (2), Almeria(3), Huesca(5) o Elche(6) son también muy débiles en casa. Repiten también como visitantes el Osasuna (2), Huesca(2), Almeria(5), Elche(5) y el Valladolid (5), que junto al Sp Gijon (5), Las Palmas(6) o el Granada(7) son los equipos que menos partidos ganan como visitantes.

#### - Conclusiones derrotas en general

El Eibar es el equipo que más derrotas ha sufrido en estos últimos 5 años con un total de 80 partidos. Muy cerca se encuentran el Celta, Malaga, Real Sociedad, Español, Levante o Deportivo con más de 68 partidos perdidos. El Real Madrid ha perdido más partidos que el Atlético de Madrid (31-29 partidos) a pesar de lograr un campeonato de Liga en estos últimos 5 años. El equipo que menos partidos ha perdido es el Barcelona. Solo ha perdido 17 partidos en estas últimas 5 temporadas. Equipos como Osasuna, Córdoba, Almeria, Huesca, Elche o Valladolid han perdido pocos partidos porque han disputado 1 o 2 temporadas de las 5 posibles.

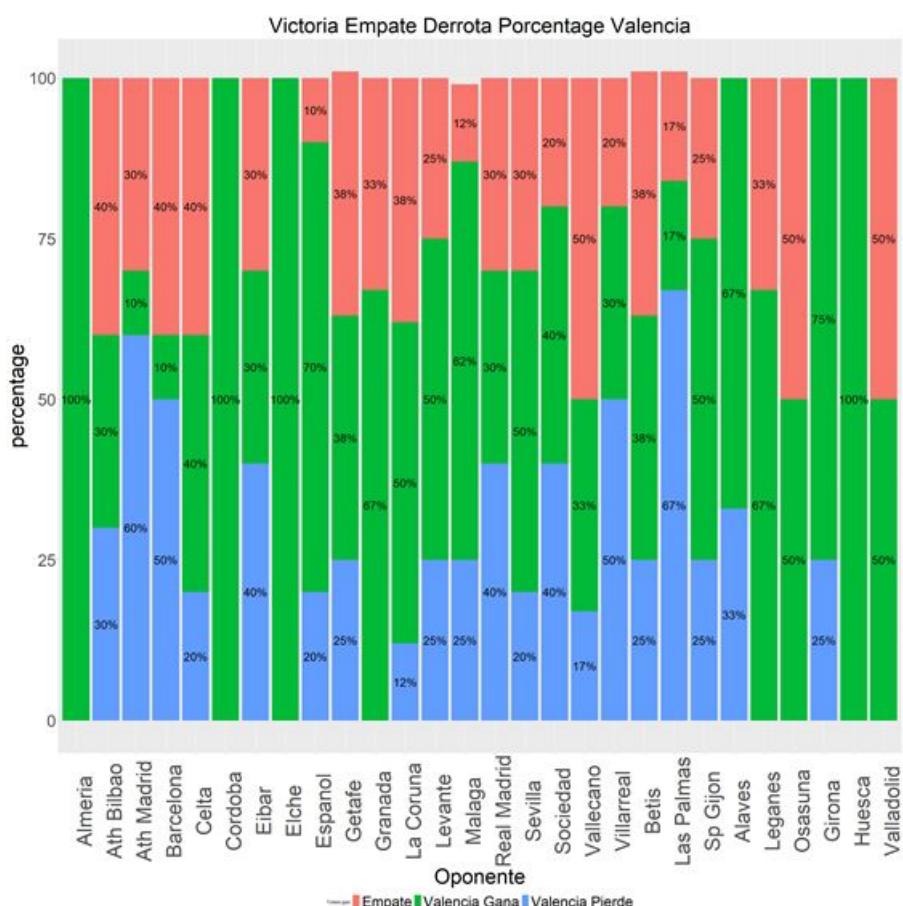
#### - Conclusiones Derrotas como local y visitante

El equipo que más partidos ha perdido en casa estas últimas 5 temporadas es el Eibar. Equipos como Deportivo, Málaga, Real Sociedad, Celta, Levante, Español, Getafe tienen también bastantes derrotas como local (todos más de 25). El equipo que menos partidos ha perdido es el Barcelona. ¡Solo ha perdido 6 partidos en las últimas 5 temporadas como local!. El Atlético de Madrid (8) y el Real Madrid (12) han perdido también pocos partidos en su estadio. Equipos como Huesca(8), Almería(9), Valladolid(9), Elche (10), Osasuna(10) o Córdoba (10) han perdido un número de partidos similar que los 3 equipos de antes, pero esto se debe a que solo han disputado 1 o 2 de las últimas 5 temporadas. De ahí el número tan bajo de derrotas.

Ocurre lo mismo con estos equipos cuando son visitantes. El Eibar ha perdido muchos más partidos que los equipos anteriores porque ha partido en la mayoría de las temporadas o en todas. El Celta es el equipo que más partidos pierde como visitante (49 partidos). Otros equipos que también han perdido muchos partidos como visitante son el Español, Eibar, Athletic de Bilbao, Real Sociedad, Málaga, Levante o Sevilla. De los equipos que han disputado todas las temporadas el Barcelona es el equipo que menos partidos ha perdido como visitante (11 partidos). El Real Madrid ha perdido 19 partidos como visitante y el Atlético de Madrid 21. Podemos concluir, por tanto, que el Atlético de Madrid es mejor local y peor visitante que el Real Madrid. También podemos ver que equipos como el Athletic de Bilbao o el Sevilla son muy buenos como locales pero son bastante débiles como visitantes.

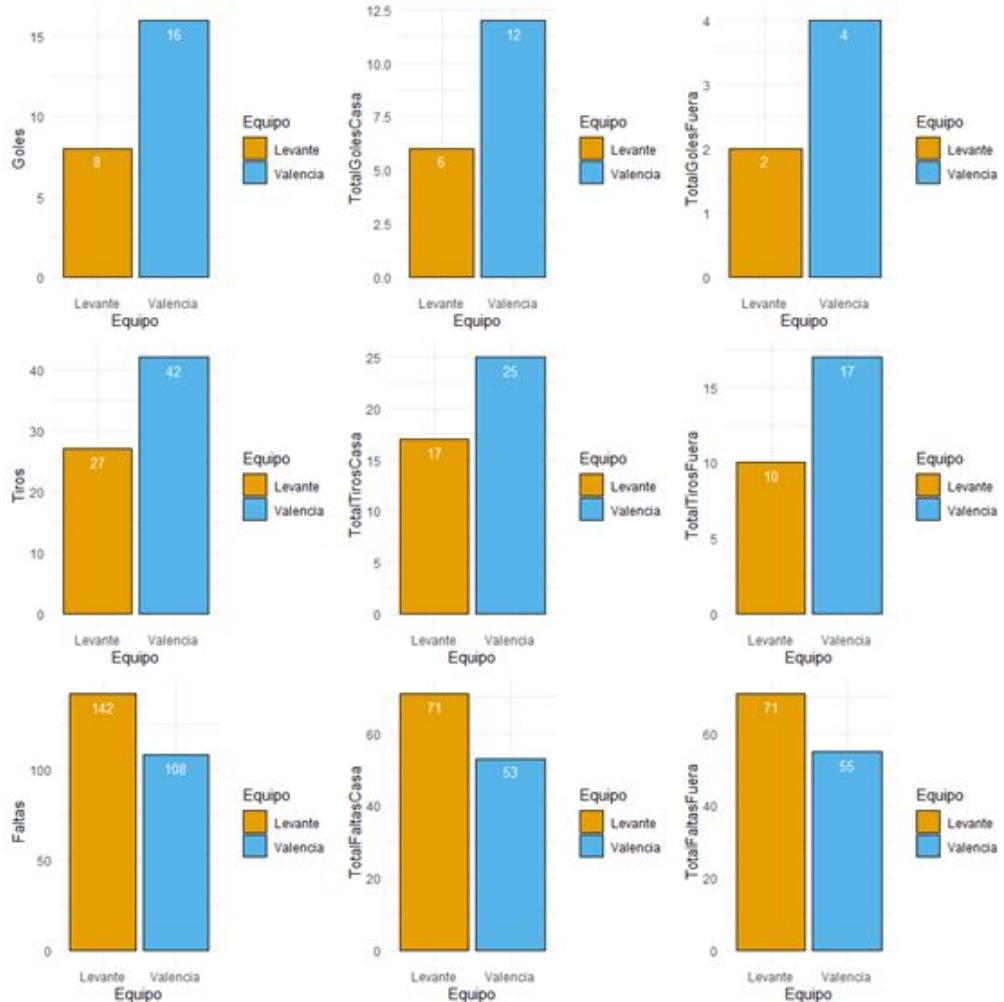
### Análisis Valencia CF

El gráfico siguiente muestra el porcentaje de victorias, empates y derrotas del Valencia frente a otros equipos.



El Valencia ha ganado todos los partidos jugados contra el Almería, el Córdoba, el Elche y Huesca. Por otro lado, tenemos que los equipos contra los que más ha perdido o les cuesta ganar son el Atlético de Madrid, el Barcelona, el Villarreal, el Madrid y Las Palmas.

A continuación, mostraremos unas gráficas comparando los tiros, los goles y las faltas de los enfrentamientos directos entre el Levante y el Valencia.



En la primera fila de los gráficos, tenemos las gráficas relacionadas con los goles. A la vista de las tres gráficas, observamos como el Valencia ha marcado el doble de goles que el Levante en general, aunque esto no quiere decir que en todos los partidos haya sido así.

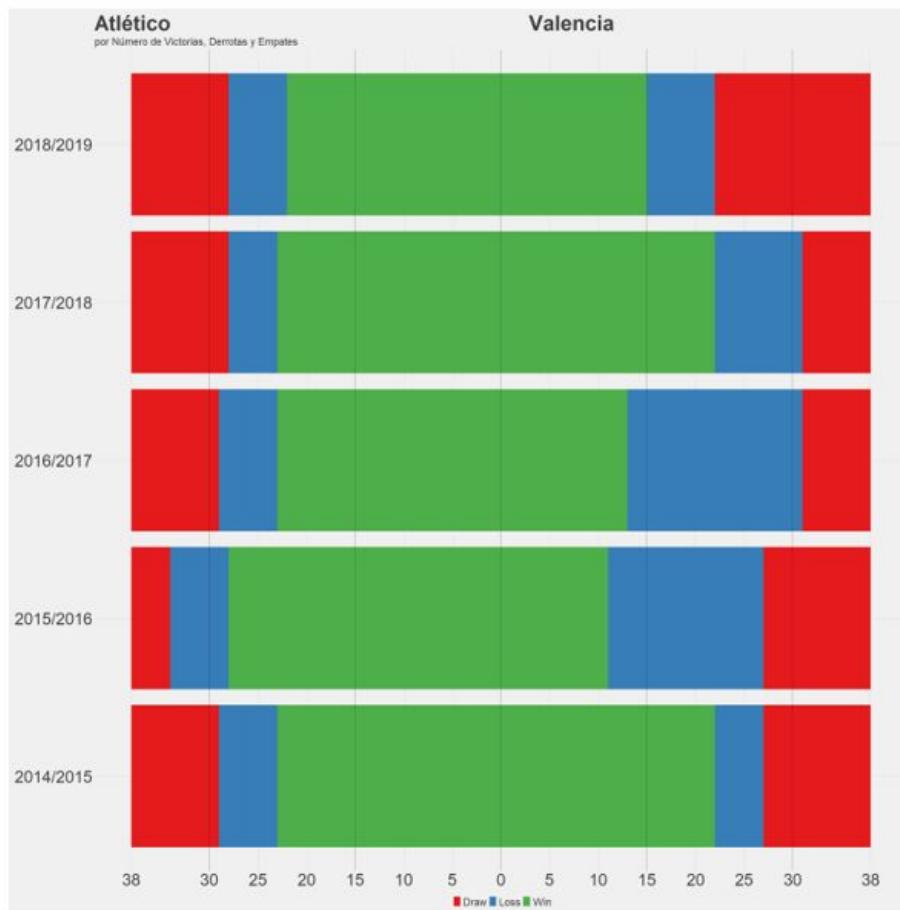
En la segunda fila tenemos las gráficas relacionadas con los tiros. A la vista de las gráficas,, obtenemos que el Valencia le ha disparado entre los tres palos al Levante en 42 ocasiones, habiendo marcado 16 goles en total, lo que supone que para marcarle un gol al Levante, el Valencia necesita tirarle 2.62 tiros a portería. Por otra parte, el Levante le ha tirado a portería al Valencia en 27 ocasiones, habiendo marcado 8 goles en total, lo que supone que para marcarle un gol al Valencia, el Levante necesita tirarle 3.5 tiros a portería. Concluimos entonces que los delanteros del Valencia han sido más efectivos o las actuaciones de los porteros del Levante han sido más flojas.

En la última fila analizamos un aspecto importante en los derbis. Este aspecto es el de las faltas, donde hay numerosos casos que confirman que, dada la rivalidad existente entre equipos de una misma ciudad, se

producen más faltas. A la vista de los gráficos, vemos que las faltas, que en principio es una estadística negativa, es decir, cuantas menos faltas comete un equipo es porque tiene más el balón y no le hace falta hacer faltas para robarlo, el Levante supera al Valencia tanto en faltas totales, como faltas hechas como local y faltas hechas como visitante.

Como conclusión, confirmamos que el Valencia ha sido mejor que el Levante porque lo supera en las estadísticas más importantes.

Por último, vamos a comparar el Valencia con el Atlético de Madrid, un equipo de categoría similar a la del Valencia.

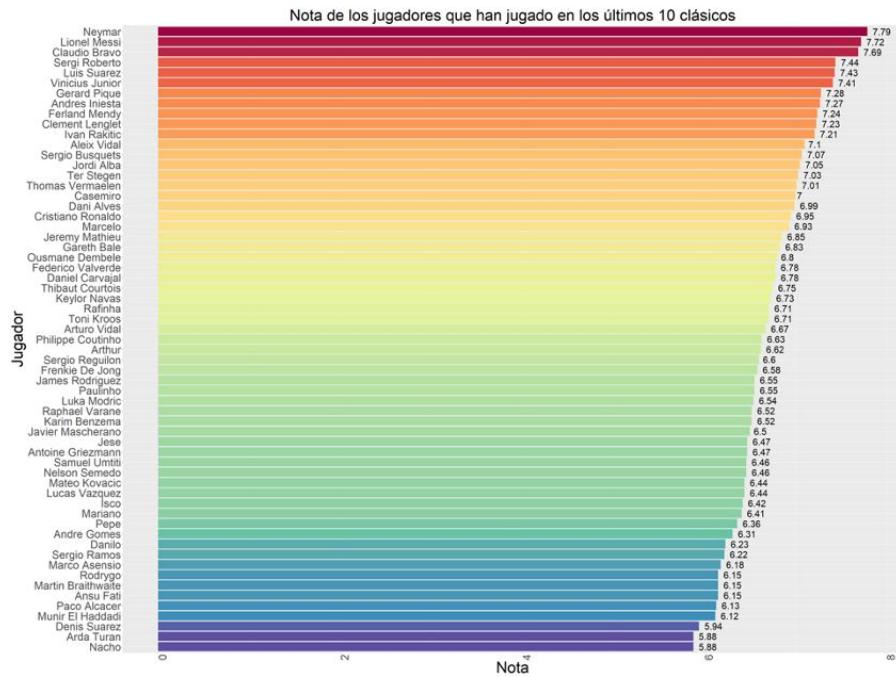


En este gráfico, la franja verde, que pertenece a las victorias, y parte desde el 0 en el medio hacia la parte del Atlético de Madrid, es mayor que la del Valencia en todas las temporadas menos en la temporada 2017/2018, que es igual. También, para la franja azul, que pertenece a las derrotas, tenemos que es menor para el Atlético en todas las temporadas menos en la temporada 2014/2015. Sin embargo, como el Atlético ganó un partido más que el Valencia, quedó un punto por delante en la clasificación.

Por lo tanto, podemos decir que el Atlético de Madrid ha sido mejor equipo que el Valencia en las últimas temporadas.

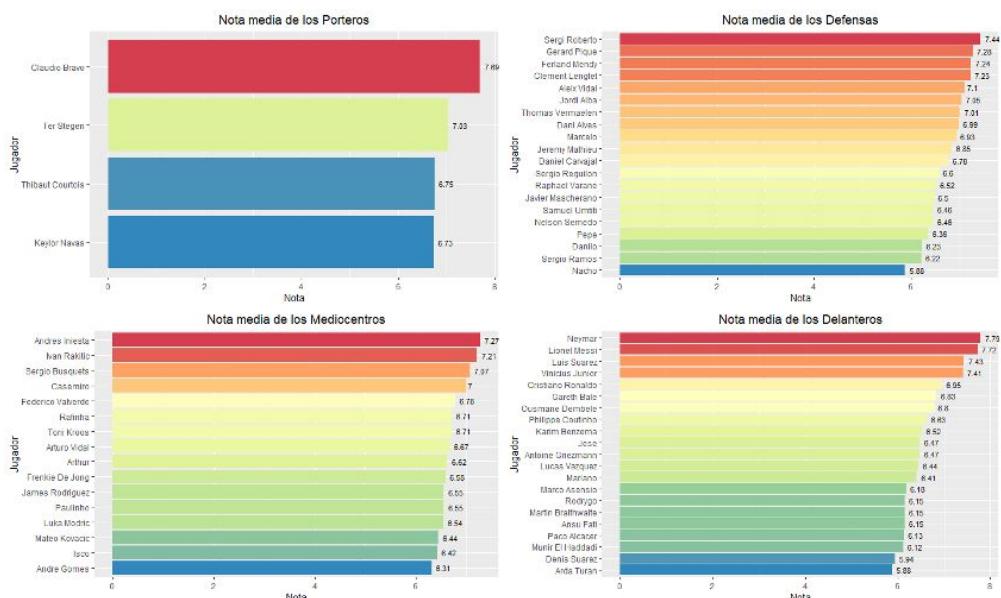
## 2. Base Clásicos.

En primer lugar veremos los mejores jugadores de los Clásicos ligueros de los últimos 5 años según su nota media.



Las notas son más o menos similares, varían como mucho 2 puntos entre los mejores jugadores y los peores. El top 3 está formado por Neymar, Leo Messi y Claudio Bravo. Los peores jugadores son Denis Suárez, Arda Turan y Nacho. Llama la atención que Cristiano Ronaldo aparezca en el puesto 19, cuando es considerado uno de los mejores jugadores del mundo. Se observa en general una mayor nota en jugadores del Barcelona que trataremos de analizar más adelante.

Ahora vamos a desglosar las Notas por Posición.



El mejor de cada posición es un jugador del FC Barcelona.

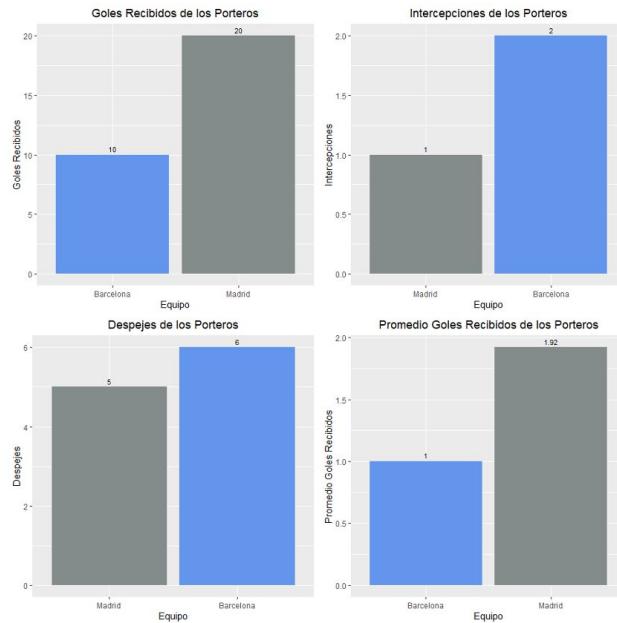
- El mejor Portero es Claudio Bravo.
- El mejor Defensa es Sergi Roberto.
- El mejor Centrocampista es Andres Iniesta.

- El mejor Delantero es Neymar.

En general se observa una clara superioridad de los jugadores del FC Barcelona en términos de calificación respecto a los jugadores del Real Madrid. Esto, evidencia una superioridad del Barça en los últimos 10 clásicos ligueros.

Las gráficas que se muestran a continuación tratan una comparativa de las estadísticas entre ambos equipos.

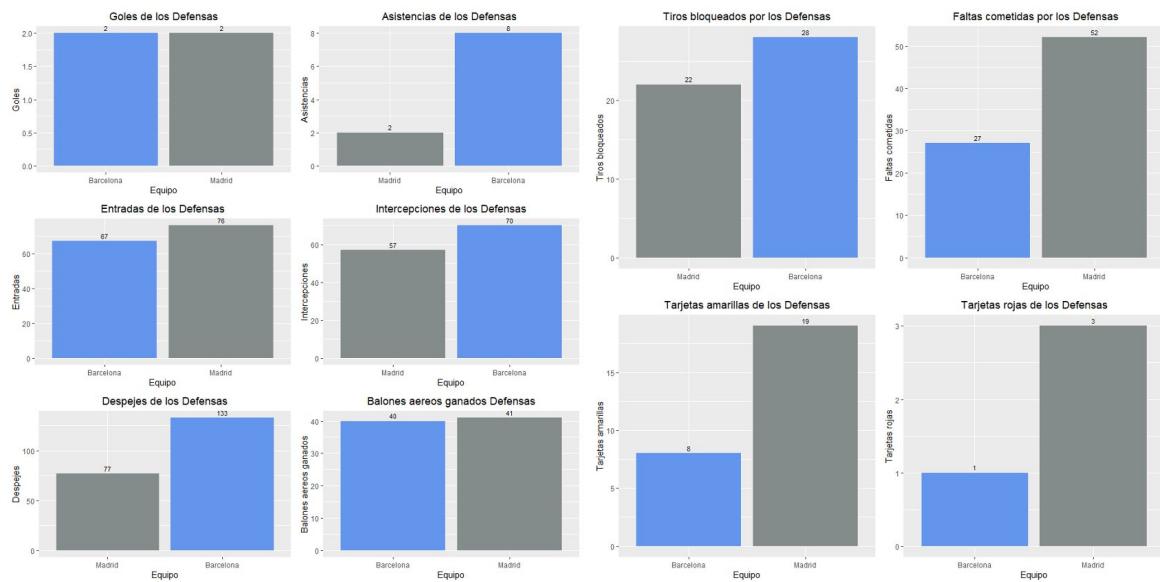
### PORTEROS



Podemos ver como los porteros del Barcelona han tenido mejores actuaciones. Esto se puede deber tanto a mejores estadísticas propias como a las del resto de sus compañeros. Es más sencillo obtener una mejor calificación si tus compañeros también la obtienen.

Si un equipo gana con comodidad, todos los jugadores sacan buenas notas porque generan ocasiones y evitan las del rival. El portero, al recibir menos goles, obtiene por tanto una mejor nota.

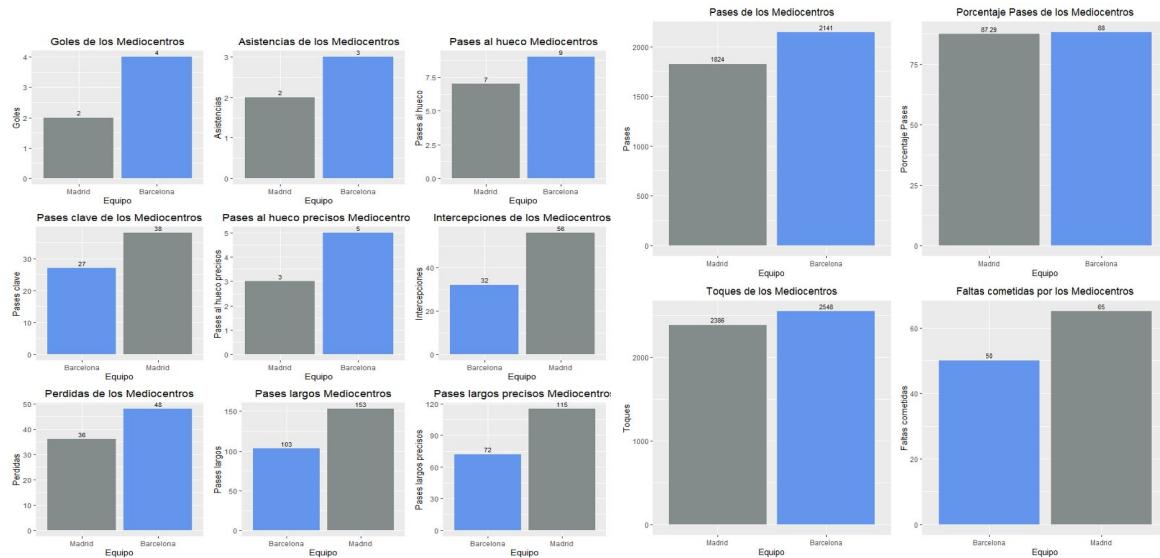
## DEFENSAS



Cuando el Barcelona es superior en estadísticas, lo es muy superior al Real Madrid. Por ejemplo en despejes y asistencias, estadísticas muy importantes en un defensa. En cambio, cuando es superior el Madrid es por la mínima. Esto ocurre en Balones aéreos ganados o entradas.

En aspectos negativos como son las Faltas Cometidas o las Tarjetas recibidas, el Real Madrid supera al FC Barcelona.

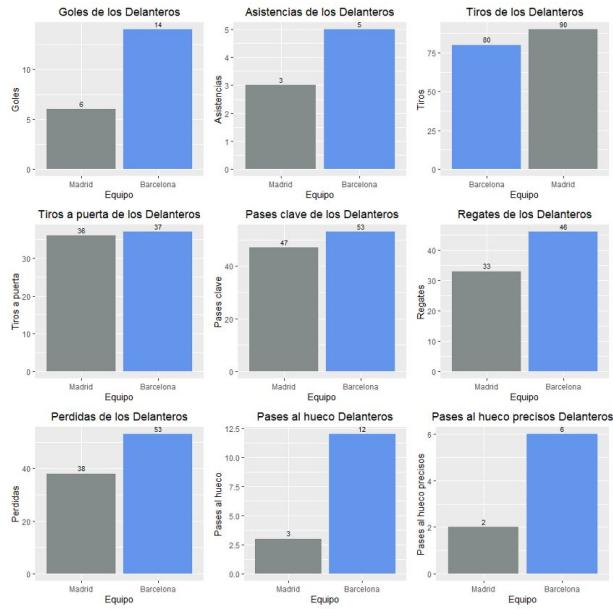
## MEDIOCENTROS



En general los 3 jugadores titulares del Real Madrid, Toni Kroos, Casemiro y Luka Modric, hacen que el Madrid supere al Barcelona en las estadísticas de mediocampistas.

La superioridad vista por los jugadores del Barcelona frente al Real Madrid se revierte en el caso de los Mediocentros.

## DELANTEROS



La victoria en esta posición vuelve a ser del Barcelona. Esto recalca la superioridad del Barcelona en los Clásicos analizados. El Barcelona anota casi el doble de goles que el Real Madrid a pesar de disparar menos. Destacar por tanto la gran efectividad de los delanteros del FC Barcelona.

### **Mejor y peor alineación de los Clásicos.**

Teniendo en cuenta las calificaciones medias y la posición de cada jugador y empleando una formación de 4-3-3, una de las más habituales en el fútbol, formamos el mejor 11 (en la parte izquierda) y el peor 11 (en la parte derecha)

Los dos campos de la parte inferior, muestran el mejor/peor 11 teniendo en cuenta las mejores/peores actuaciones de los jugadores y no su cómputo global. Es decir, tenemos en cuenta la mejor/peor calificación del jugador y no su calificación media.

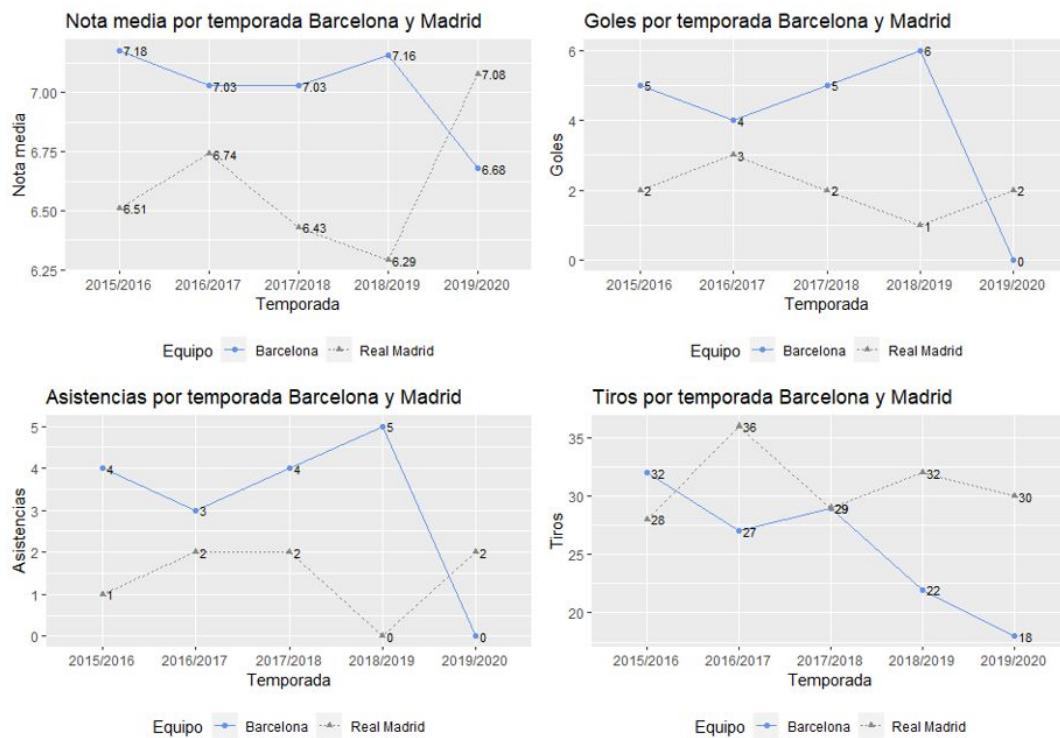
En el mejor equipo aparecen una gran mayoría de jugadores del FC Barcelona. En el peor equipo aparecen muchos jugadores del Madrid. Los jugadores del Barcelona en el peor equipo son suplentes.

Esto demuestra la superioridad del FC Barcelona en los clásicos ligeros de las últimas 5 temporadas.



Finalmente, ¿ha sido mejor el FC Barcelona o el Real Madrid?

Las variables que más influyen en el resultado de un partido son Goles, Asistencia, Tiros y la Nota Media de los jugadores de cada equipo.



El FC Barcelona ha anotado más goles y asistencias que el Real Madrid entre la temporada 2015/2016 y la 2018/2019. En la última temporada, la 2019/2020, el Real Madrid parece haber dado la vuelta a la situación ya que consiguió anotar 2 goles, dejando su portería a 0 en ambos partidos.

La Nota Media puede ser también un buen indicador de cómo funciona el equipo. En las 4 primeras temporadas son los jugadores del FC Barcelona quienes sacaron mejores notas. Sin embargo, en la última temporada (que es la actual), como ya hemos dicho antes el Real Madrid parece haber reaccionado al dominio azulgrana. En esta temporada los jugadores del Real Madrid han sacado una mejor nota que los del FC Barcelona.

Como conclusión, podemos decir que el FC Barcelona ha sido claramente mejor equipo que el Real Madrid en cuanto a enfrentamientos directos se refiere en estos últimos 5 años. Sin embargo, el Real Madrid parece haber cambiado esta situación de superioridad del FC Barcelona en la última temporada, que veremos si se prolonga o no en el tiempo.

### 3. Base Jugadores

El primer estudio de esta base de datos consistirá en la visualización de los mejores y peores jugadores de la liga, de forma que según el valor de los jugadores que hemos creados a partir de sus estadísticas y características mediante el proceso analítico jerárquico.

**Mejor Equipo de La Liga**



**Peor Equipo de La Liga**



Del **mejor once** todos los jugadores menos Diego López, Damián Suárez y Ever Banega pertenecen a equipos que tienen el nivel más alto, por lo que es de esperar que esos jugadores estuvieran en el mejor once de la

liga. Sin embargo, **Diego López, Damián Suárez y Ever Banega**, podrían considerarse las revelaciones de la liga ya que son jugadores que no pertenecen a los equipos del nivel más alto y se han ganado un puesto entre los mejores de la liga.

#### JUGADORES REVELACIÓN DE LA TEMPORADA 2018/2019 :

- **DIEGO LÓPEZ - ESPANYOL**
- **DAMIÁN SUÁREZ - GETAFE**
- **ÉVER BANEGA - SEVILLA**

Por otro lado del **peor once** no se puede sacar muchas conclusiones, todos los jugadores son menores de 24 años menos Juanjo Camacho que tienen 37 años y es su última temporada antes de retirarse. Por tanto los peores jugadores de la liga son jugadores jóvenes que están empezando (apenas suman 10 partidos entre los 11 jugadores) o jugadores a punto de retirarse.

El segundo estudio de esta base de datos se centrará en un equipo de fútbol en concreto, el Valencia CF. Sobre este analizaremos a los distintos jugadores por posiciones del Valencia:

- ¿En qué posiciones es más débil y más fuerte?
- ¿Cuáles se podrían mejorar con jugadores de niveles inferiores?
- Comparar el Valencia con el resto de equipos de su nivel por posición.

Mejor Equipo del Valencia



Peor Equipo del Valencia



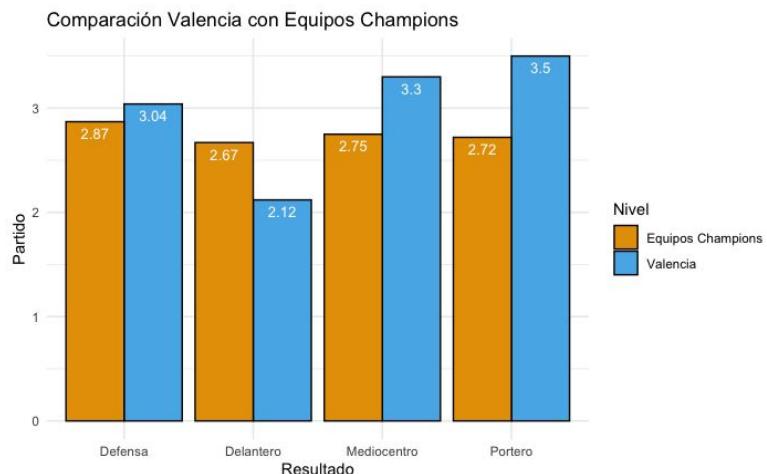
Muchos de los jugadores del equipo de los peores tienen un buen valor pero están en este peor once porque sus compañeros tienen mayor posición, si esto sucede consideraremos que el Valencia es muy fuerte en esa posición.

La media del valor de los porteros de la liga es de 3.04, la media de los porteros de los equipos nivel Champions es 3.20 y el mejor portero del Valencia Neto tiene un valor muy superior ambas de 4.93, por lo que el Valencia es fuerte en portería.

La media del valor de los defensas de la liga es de 2.8, la media de los defensas de los equipos nivel Champions es 2.95 y los mejores defensas del Valencia tienen una media de 4.17 y su mejor defensa José Luis Gayá tiene un valor de 4.95 por lo que el Valencia también es fuerte en esta posición.

La media de los mediocentros de los equipos nivel Champions es 3.2 y los mejores mediocentros del Valencia tienen una media de 3.89 y el mediocentro que más valor tiene del Valencia es el mejor mediocentro de la liga, pero los otros mediocentros.

La media del valor de los delanteros de la liga es de 2 la media de los delanteros de los equipos nivel Champions es 2.8 y los delanteros del Valencia tienen una media de 2.12 y su mejor delantero es Kevin Gameiro con 3.29. Por esto no podemos considerar que sea fuerte en esta posición, tiene buenos titulares pero los suplentes no dan la talla así que vamos a buscar un jugador que haya mostrado un valor al nivel de los titulares del Valencia Gameiro, Rodrigo y Santi Mina, y pueda sustituir al más bajo de ellos Santi Mina con un valor de 2.79.



En este gráfico de barras comparamos por posiciones el Valor medio del resto los equipos del nivel champions league con la media del Valencia CF

Ahora según esto dicho anteriormente, vamos a buscar a nuestro nuevo delantero en equipos de nivel Uefa Europa League y Posición de mitad de tabla, hasta dos niveles por debajo del nivel del Valencia CF.

Los mejores delanteros de los equipos nivel Europa League son:

1. **Jorge Molina** → 3.41
2. **Borja Iglesias** → 3.35
3. **Wissam Ben Yedder** → 3.32

Los mejores delanteros de los equipos nivel Mitad de tabla son:

1. **Mikel Oyarzabal** → 3.40
2. **Iñaki Williams** → 3.27
3. **Iker Muniain** → 2.94

Vamos a valorar uno de cada nivel ya que presentan valores similares y no hay ninguno que se vaya del resto. Según nuestro criterio creemos que es más interesante determinar el valor monetario de **Borja Iglesias** y **Mikel Oyarzabal**.

El valor monetario de Borja Iglesias y Mikel Oyarzabal se obtendrá utilizando el Método del proceso analítico jerárquico, aquí mostraremos el valor final el proceso de este método al completo se encontrará en el anexo **VALOR BORJA IGLESIAS / VALOR MIKEL OYARZABAL**

**BORJA IGLESIAS = 22 Millones de EUROS**

**MIKEL OYARZABAL = 33 Millones de EUROS**

Con esto proporcionamos el valor monetario de 2 jugadores que han demostrado un nivel propio de un equipo del nivel del Valencia CF por lo que podrían reemplazar la posición más débil del equipo, la posición de SANTI MINA, esta información se la pasamos a los directivos del Valencia CF y ya queda en sus manos la decisión de fichar uno a otro, o pedir que valoremos a otros jugadores.

Sobre los valores monetarios obtenidos por este grupo, se ha valorado a BORJA IGLESIAS con 22 millones en la temporada 2018/2019, en el verano de 2019 el Betis pagó 25 millones por la propiedad de la ficha federativa de Borja Iglesias. Por lo que teniendo en cuenta que solo tenemos datos de la liga y no de la copa que también participó Borja Iglesias, consideramos que hemos hecho una buena valoración.

Otro estudio de esta base de datos, consistirá en tras la visualización de distintos equipos de distintos niveles, estudiaremos los mejores once y mejores jugadores que puedan de cada nivel de equipo. Y veremos la posibilidad de que alguno de los jugadores pueda jugar en un nivel superior si es de los que más destaca en su nivel y tiene el nivel suficiente para un nivel superior.

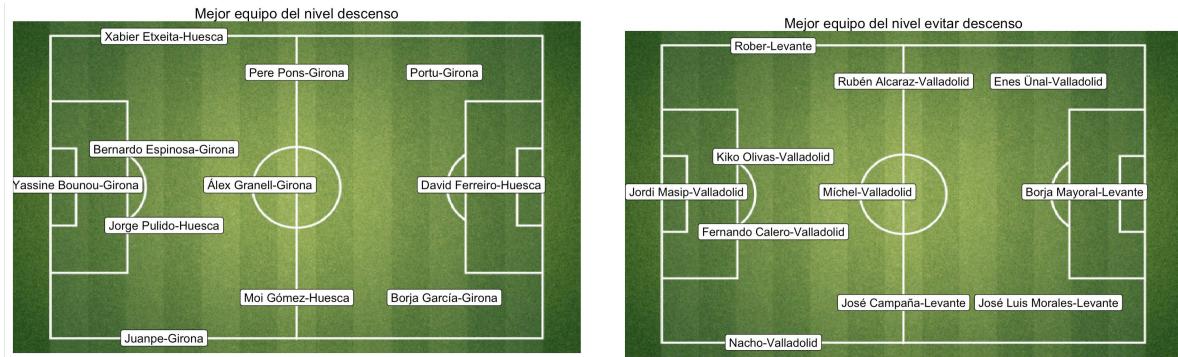
**Nota:** un jugador sólo será considerado para cambiar a equipos como máximo dos niveles por encima del nivel de su equipo. Esto es porque cada jugador compite con otros compañeros de equipo por jugar y en equipos de nivel superior la competencia es más dura que en niveles bajos.

Comenzamos con los jugadores que pertenecen a equipos de **nivel de DESCENSO de división**, ya que son los jugadores que para equipos que se mantienen en la división son más accesibles y más rentables. Vamos a considerar los jugadores que estén mínimo 1.5 por encima de la media.

De entre los jugadores del HUESCA, RAYO VALLECANO y GIRONA los jugadores que según su valor más vale la pena observar son los que se ven en el gráfico de la izquierda

#### **NIVEL EVITAR EL DESCENSO.**

De entre los jugadores del LEVANTE, CELTA DE VIGO y VALLADOLID los jugadores que según su valor más vale la pena observar son los que se ven en el gráfico de la derecha

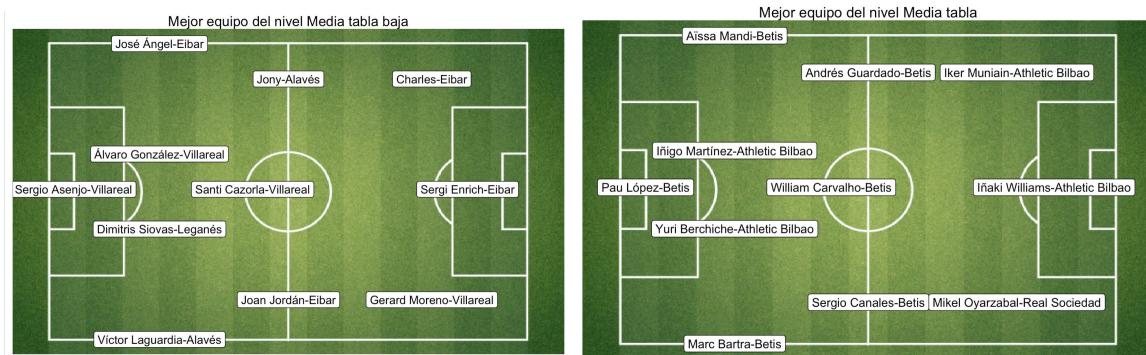


### NIVEL MITAD DE TABLA BAJA.

De entre los jugadores del ALAVÉS, EIBAR y VILLARREAL los jugadores que según su valor más vale la pena observar son los que se ven en el gráfico de la izquierda.

### NIVEL MITAD TABLA.

De entre los jugadores del ATHLETIC BILBAO, LEGANÉS y BETIS los jugadores que según su valor más vale la pena observar son los que se ven en el gráfico de la derecha.



### NIVEL UEFA EUROPA LEAGUE.

De entre los jugadores del ESPANYOL, SEVILLA, GETAFE los jugadores que según su valor más vale la pena observar son los que se ven en el gráfico de la izquierda

### NIVEL CHAMPIONS LEAGUE

De entre los jugadores del ATLÉTICO MADRID, REAL MADRID, VALENCIA CF, BARCELONA los jugadores que según su valor más vale la pena observar son los que se ven en el gráfico de la derecha.



## VII. LECCIONES APRENDIDAS PARA MIS FUTUROS PROYECTOS DE CIENCIA DE DATOS.

Mientras hacíamos el trabajo nos han surgido varias dudas que hemos ido resolviendo. Esto nos puede servir para futuros proyectos. Algunas lecciones que hemos aprendido son:

Proponer objetivos que no sean muy específicos y ir especificando a medida que vamos progresando en los análisis. Durante este proyecto hemos tenido que cambiar varios objetivos porque era imposible realizarlos con nuestros datos, o bien porque hemos propuesto otros que nos parecían más interesantes.

También hemos aprendido bastantes cosas de programación en R, ya que antes de este trabajo no habíamos profundizado mucho en la programación en R, ya que utilizábamos siempre python. Entre estas podemos destacar a graficar imágenes sacadas de internet, juntar bases de datos, y muchos aspectos básicos de R que antes no conocíamos.

---

## VIII. BIBLIOGRAFÍA

### I. Páginas Web de dónde hemos sacado las Bases de Datos

- Base Resultados: [Datahub](#)
- Base Clásicos: [WhoScored](#)
- Base Jugadores: [FBref](#)

### II. Noticias relacionadas con este proyecto

- [Big Data en el Getafe](#)

El equipo de Bordalás trabajó con una empresa israelí de 'big data' para prevenir lesiones y mejorar el rendimiento físico de su plantilla. El Getafe ha cerrado la campaña 2018/2019 como el equipo de Primera con menos lesiones musculares, tan sólo ocho. El Atlético o el Real Madrid rondaron las 30.

- [Real Madrid La crisis de Gareth Bale, explicada con el Big Data](#)

En las últimas dos temporadas el proyecto del Real Madrid ha pasado por altibajos especialmente en el ataque. Tras la salida de Cristiano Ronaldo el equipo ha perdido fiabilidad cara a puerta y la producción goleadora ha caído en picado desde entonces. Gareth Bale parecía uno de los nombres propios con más capacidad para maquillar la marcha del portugués pero su influencia ha seguido la dinámica descendente de los últimos cursos. La consultora [Driblab](#) muestra de manera clara el mal momento de Gareth Bale. Si analizamos las cifras de las últimas temporadas la caída del astro galés ha sido total.

- [Cómo se aplica Big Data en el Real Madrid](#)

El Real Madrid es uno de los mejores equipos, y económicamente entre los más potentes, por lo que tiene la exigencia necesaria de hacer uso de tecnologías aplicadas al fútbol para mejorar su rendimiento y ayudarle a ganar partidos. » El Big Data no te gana partidos, pero ayuda». Y ya desde la época de Carlo Ancelotti tenían implantados sistemas de tratamiento de datos, principalmente físicos, para controlar el nivel actual del jugador y compararlo con datos históricos para determinar la situación física del mismo.

- [Soccer Analytics usando R](#)

Breve artículo donde se explica la aplicación de R para el análisis futbolístico.

---

## IX. ANEXOS

Para el apartado de Anexos hemos creado una carpeta virtual a la que se accede con el siguiente enlace:  
[Anexos Proyecto II](#)

La carpeta consta de los siguientes carpetas:

- Fichas
    - Fichas configuración - fuentes de datos
    - Fichas configuración - integración datos - plantilla
  - Clásicos
  - Jugadores
  - Resultados
- En las 3 carpetas tenemos los Script en R y PDF del análisis en R y las Bases de datos en CSV
- Proceso analítico jerárquico
    - Matrices de comparación pareada
    - Modelo de estructura jerárquica
    - Explicación del valor para la base de Jugadores
    - Archivos Excel con el Valor Monetario de varios Jugadores necesarios para el análisis.
-