

UNIVERSIDAD MIGUEL HERNÁNDEZ DE
ELCHE

Facultad de Ciencias Sociales y Jurídicas de Elche



Optimización de equipos deportivos

GRADO EN ESTADÍSTICA EMPRESARIAL

Curso Académico 2022/2023

ESTUDIANTE **Alexandre Buforn Úbeda**

TUTOR **José Luís Sainz-Pardo Auñon**

Trabajo Fin de Grado

Elche, junio de 2023

Este concepto de TFG ha surgido desde la pasión más profunda del estudiante por el fútbol y la implementación de todos los conocimientos aprendidos durante el grado para ayudar a cualquier equipo deportivo.





Índice

1. Resumen.....	4
2. Introducción.....	5
2.1. Presentación.....	5
2.2. Motivos.....	8
3. Marcos teórico.....	9
4. Objetivos.....	10
5. Metodología.....	11
5.1. Recopilación de información.....	11
6. Resultados.....	15
6.1. Preparación de Datos.....	15
6.2. Estadística Descriptiva.....	16
6.3. Notas jugadores sin algoritmo.....	25
6.4. Notas jugadores con algoritmo.....	28
6.5. Modelo.....	35
7. Conclusiones.....	42
8. Bibliografía.....	44
9. Anexos.....	48
9.1. Código R Estadística Descriptiva.....	50
9.2. Código R Obtención Notas.....	59
9.3. Código R Obtención Notas Regresión.....	60
9.4. Código R Modelo	65

1. Resumen

La mala gestión de un club puede llevar a la mayor catástrofe deportiva y económica de una entidad. Deportivamente hablando, la peor consecuencia es la pérdida de puestos en la tabla y no poder participar en competiciones europeas, incluso, pelear hasta la última jornada por no descender de categoría. En cambio, en el ámbito económico, el club puede verse cerca de la ruina más absoluta hasta el punto de tener que vender a sus mejores jugadores para poder afrontar todas las deudas y salarios.

La elaboración de un modelo estadístico nos va a permitir ayudar a cualquier entidad deportiva, en este caso, al Valencia CF. Esto supone la mejora de la plantilla a través de fichajes con unas características determinadas por un precio mínimo y la venta o fin de cesión de varios jugadores que no han cumplido con las expectativas esperadas.

Palabras clave: regresión lineal múltiple, modelo estadístico, fútbol, notas, algoritmo.

Abstract

The mismanagement of a club can lead to the biggest sporting and economic catastrophe of an entity. Sportingly speaking, the worst consequence is the loss of places in the table and not being able to participate in European competitions, even fighting until the last day for not being relegated. On the other hand, in the economic field, the club can be seen close to absolute ruin to the point of having to sell its best players to be able to face all the debts and salaries of the players.

The development of a statistical model will allow us to help any sports entity, in this case Valencia CF. This supposes the improvement of the squad through signings with certain characteristics for a minimum price and the sale or end of assignment of several players who have not met the expected expectations.

Keywords: multiple linear regression, statistical model, soccer, grades, algorithm.

2. Introducción

2.1. Presentación

El fútbol¹ que conocemos hoy en día no tiene nada en común con el que se practicaba en sus inicios. Sus orígenes se remontan a los siglos III y II a.C. en la antigua China, donde se practicaba un deporte parecido al de patear una pelota. Durante el paso del tiempo, esta actividad fue avanzando en diferentes lugares del mundo, como por ejemplo el Lejano Oriente y el Mediterráneo.

Aunque existen varias hipótesis que indican que el fútbol lo inventaron los *guaraníes* (pueblo de guerreros paraguayos) en el siglo XV, existen más evidencias de que lo inventaron los ingleses. Hasta el año 1863, patear una pelota con los pies con el objetivo de meterla en la red contraria no se llamaría fútbol o balompié.

La *FA (Football Association)* fue la primera asociación de fútbol que se creó en todo el mundo, en concreto, se fundó en Inglaterra. Los integrantes de esta asociación (12 equipos) se reunieron en la *Taberna Freemason's* y acordaron la creación del primer reglamento oficial. Esta normativa estaría compuesta por 14 reglas y la llamarían *Cambridge*².

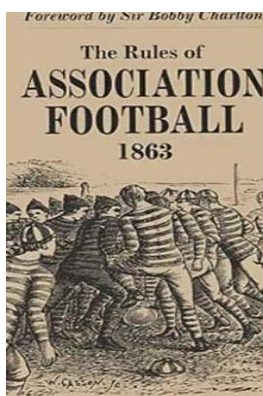


Figura 1: primera normativa del fútbol profesional. Fuente: LA RAZÓN.

1. Puede ver en el [siguiente enlace](#) los descubrimientos de National Geographic.
2. Puede ver en el [siguiente enlace](#) el reglamento Cambridge.

A partir de la creación de este reglamento, la normativa se ha ido actualizando a los nuevos tiempos y actualmente, el documento³ que se utiliza es el actualizado en 2016 con la implementación del VAR (*Video Assistant Referee*).

A partir de esta organización, apareció la que hoy en día es la más importante de todas, la FIFA (*Fédération Internationale de Football Association*).



Figura 2: Logo FIFA. Fuente: FIFA.

Esta asociación se creó el año 1904. Principalmente, se encarga de gobernar las federaciones de fútbol a nivel global y de organizar los mayores acontecimientos futbolísticos como por ejemplo La Copa Mundial de Fútbol.

Durante el paso de los años, la expansión de este deporte ha ido creciendo de manera exponencial hasta convertirse en el deporte rey. Las zonas más populares son Europa y América del Sur gracias a sus torneos continentales tanto a nivel de clubes como de selecciones. Los más conocidos son: la *Champions League*, *Eurocopa*, *Conmebol* y la *Copa América*.

Durante los últimos años, este deporte se ha ido haciendo más popular en lugares como Estados Unidos, China, Japón, Catar y Arabia Saudí. Esto es debido a la aparición de empresarios multimillonarios y jeques que están convenciendo a los mejores jugadores actuales con salarios desorbitados para que acaben sus carreras en sus equipos. Los casos más destacados a nivel europeo son el Paris Saint Germain y el Manchester City, los cuales han ingresado centenares de millones de euros para poder fichar a los mejores jugadores del mundo y conseguir los trofeos más significativos.

3. [Aquí](#) podrá encontrar el documento en formato PDF de la normativa actual.

Posición	Dueño	Equipos	Patrimonio en dólares (2021)
1	Mohamed Bin Salmán	Newcastle United	435 billones
2	Sheikh Mansour	Manchester City, Melbourne City, New York City	31.59 billones
3	Dietrich Mateschitz	Red Bull Salzburg, New York Red Bulls, RB Leipzig	21.38 billones
4	Andrea Agnelli	Juventus	14.98 billones
5	Roman Abramovich	Chelsea	13.8 billones
6	Philip Anschutz	LA Galaxy	11.03 billones
7	Stan Kroenke	Arsenal, Colorado Rapids	9.26 billones
8	Nasser Al-Khelaifi	PSG	8.85 billones
9	Zhang Jindong	Inter Milan	8.44 billones
10	Guo Guangchang	Wolverhampton	7.08 billones

Cuadro 1: dueños de clubes de fútbol más ricos del mundo. Fuente: GOAL.

En el ámbito nacional, la llegada de estas personalidades adineradas al *Campeonato Nacional de Liga de Primera División* no es nada nuevo, ya que, en el año 2010, el emir catari *Abdullah ben Nasser Al Thani* se hizo con la totalidad del *Málaga CF* por una cantidad de 36 millones de euros. En su momento, el club andaluz pagó la deuda pendiente de 14 millones de euros y llegó a puestos europeos, realizando la mejor temporada de todos sus tiempos. Hoy en día, este equipo se encuentra en Segunda División y con una deuda mayor a la anterior, por lo que está a punto de desaparecer, a no ser que otro empresario compre el club.

2.2. Motivos

La aparición del COVID-19⁴ en diciembre de 2019 hizo que gran parte de los países de todo el mundo cerraran sus fronteras e impusieran normas de convivencia para poder luchar contra esta epidemia y poder acabar cuanto antes con ella. Los meses de confinamiento no solo afectaron a la clase media, sino que también golpearon a la economía de todas las grandes y pequeñas empresas de nuestro país, obligando a algunas a tener que cerrar.

Los equipos de fútbol profesionales de nuestro país también se vieron afectados por esta situación tan inesperada. Desde el punto de vista contable, ninguna previsión a los siguientes años había calculado una situación que desestabilizara tanto las cuentas de cualquier entidad. Esto provocó tantas pérdidas que obligaron a algunos presidentes prescindir de sus jugadores máspreciados o vender el club a cualquier empresa o jeque multimillonario.

El caso que más impactó en el mundo del balompié fue el del *FC Barcelona* y *Lionel Messi*. Tras haber acabado la campaña 2020/2021 con unas **pérdidas de más de 1200 millones de euros**, el club tuvo que tomar la drástica decisión de desprenderse del jugador que más le había dado al club tanto a nivel futbolístico como económico.

Actualmente, los únicos equipos españoles que han podido aguantar la presión de todos los “petrodólares” y de los empresarios asiáticos y norteamericanos son el *FC Barcelona*, *Real Madrid CF*, *Athletic Club De Bilbao* y *CA Osasuna*. Estas entidades pertenecen a sus socios, es decir, que estos son los que deciden a través de unas elecciones y lo transmiten a la Junta Directiva del club correspondiente. El resto de los integrantes de la Liga Española son sociedades anónimas.

Como dato curioso, los últimos equipos mencionados son los actuales participantes en la competición de la Copa de su Majestad el Rey del año 2023. Este hecho es bueno para el fútbol porque demuestra que no siempre tiene que ganar el dinero.

4. En este [enlace](#) podemos obtener más información sobre el COVID

3. Marco teórico

Las nuevas tecnologías no solamente han mejorado y modernizado nuestras vidas, sino que también han cambiado el trabajo de muchos sectores. En nuestro caso en concreto, vamos a hablar de cómo se introdujo la estadística en el deporte y en el fútbol.

La primera vez que se usó el **Big Data** en el deporte fue en los *Juegos Olímpicos de Londres 2012*, pero donde se le dio más visibilidad y se aplicó con más eficacia fue en los siguientes Juegos, los de *Río 2016*. Principalmente se utilizó para poder predecir las medallas que se iban a repartir según el país y la modalidad.

En cuanto al balompié, **la recopilación de datos empezó en los años 50 de forma manual**⁵, pero cuando ha dado el gran salto ha sido en las últimas temporadas. El principal culpable ha sido el *Brentford FC*⁶, actual equipo de la *Premier League*. Esto le ha permitido al humilde club poder crecer a niveles tan grandes que han podido pelear contra los más grandes y formar un equipo competitivo.

A nivel nacional, el director deportivo del *Sevilla FC*, *Monchi*, es considerado uno de los mejores en su puesto, ya que consigue el fichaje de jugadores poco conocidos y los vende a precios de estrellas. Los casos más conocidos han sido el de Dani Alves, Sergio Ramos y, el más reciente, el de Jules Koundé.

Las claves para poder fichar bien y barato son las siguientes: **tener en la base de datos la mayor cantidad de jugadores posibles, consultar con el entrenador la posición que más necesita, añadir las características que más demandan para así poder decidir entre una menor cantidad de jugadores**, y, por último, **una vez tenemos la respuesta del algoritmo, buscar el mejor precio según el mercado** (Monchi, 2020)⁷.

5. En este [enlace](#) podemos ver un artículo donde lo explica de una forma más detallada.

6. En este [enlace](#) podemos ver la historia del club inglés y el uso de la estadística.

7. En este [enlace](#) podemos ver una parte de la entrevista que le realizan al director hispalense.

4. Objetivos

La finalidad principal de este trabajo es poder ayudar al club valenciano que más veces ha llegado a finales de competiciones europeas y que más títulos nacionales tiene en sus arcas, el *Valencia CF*⁸.

En el año 2014, la entidad valenciana estuvo a punto de desaparecer porque el banco español *Bankia* ya no iba a refinanciar al club y estaba dispuesto a venderlo. En ese mismo momento, entró en escena el multimillonario empresario singapurense *Peter Lim* dispuesto a salvar al club de la quiebra total. La situación actual del club es parecida a la de la entidad malaguista. A día de hoy, el club valenciano se ha quedado a un solo punto de descender a la segunda división y con unas pérdidas de 45.8 millones de euros, incluso sin tener dinero para pagar las nóminas de sus jugadores.

El descenso a una categoría inferior implicaría la venta de jugadores muy importantes, ya que los ingresos que pueden recibir no tienen ni punto de comparación con los que perciben los clubes de primera división.

El trabajo que vamos a realizar será ayudar al club utilizando el Big Data. Esto significa que vamos a poder recopilar toda la información de los jugadores del Valencia CF y del resto de equipos españoles de primera y segunda división española. Esta información la vamos a emplear para saber cómo podemos mejorar el equipo a través de ventas y compras que aumenten el rendimiento general en los partidos para poder obtener mejores resultados.

Para poder saber que jugador es más efectivo que otro, sacaremos las notas de las 3 últimas temporadas y haremos una media ponderada entre ellas. La obtención de estas calificaciones las vamos a conseguir mediante varias aplicaciones especializadas en fútbol y a través de una regresión lineal múltiple que explicaremos más adelante. Estos programas deportivos tienen un algoritmo secreto que no permiten dar a conocer como han obtenido estas calificaciones, por lo que vamos a desarrollar por nuestra cuenta un método que nos permita acercarnos lo máximo posible.

8. En este [enlace](#) podemos ver el ejercicio contable de la temporada 20/21 y 21/22 en las páginas 9, 10, 11 y 12 del documento.

5. Metodología

El software que hemos utilizado ha sido la *versión 2022.07.2+576 de Rstudio*⁹, un espacio integrado en el lenguaje de R. De software libre y código abierto.

5.1. Recopilación de información

Los datos que vamos a utilizar han sido obtenidos de diferentes fuentes de información deportiva especializadas en el mundo del fútbol. Las páginas que hemos empleado han sido [Transfermarkt](#)¹⁰, [Sofascore](#)¹¹, [SalarySport](#)¹² y [Onefootball](#)¹³.

La primera de ellas la hemos utilizado para poder obtener toda la información relacionada con el valor de mercado actual de los jugadores seleccionados, el número de tarjetas amarillas y rojas que han recibido durante toda su carrera, el porcentaje de minutos que han jugado con sus equipos durante las últimas 3 temporadas y los minutos disputados con sus respectivas selecciones nacionales hasta el día de hoy.

La siguiente fuente la hemos utilizado para obtener información sobre el porcentaje de goles, pases, defensa y paradas de la actual temporada, obtener la nota media calculada en cada uno de los últimos 3 años, obtener el precio de su último fichaje y saber si es un jugador que en el último año ha tendido a lesionarse.

La siguiente fuente la hemos utilizado para obtener información sobre el salario que perciben actualmente cada uno de los jugadores seleccionados en nuestra base de datos.

9. En este [enlace](#) podemos ver toda la información relacionada con Rstudio
10. En este [enlace](#) podemos ver toda la información relacionada con Transfermarkt.
11. En este [enlace](#) podemos ver toda la información relacionada con Sofascore.
12. En este [enlace](#) podemos ver toda la información relacionada con SalarySport.
13. En este [enlace](#) podemos ver toda la información relacionada con OneFootball.

Por último, la última fuente la hemos utilizado para obtener información sobre el nombre de los jugadores, el equipo al que pertenecen, la posición que ocupan en el campo, la edad y el tipo de formación que más han jugado esta temporada en sus entidades.

La base de datos que hemos rellenado manualmente a través de diferentes fuentes de información contiene las siguientes variables:

- **Jugadores:** contiene el nombre de todos los futbolistas que vamos a analizar.
- **Equipo:** contiene el nombre de las 5 entidades que vamos a analizar. Estas son el **FC Barcelona**, el **Valencia CF**, el **Levante UD**, el **Sevilla FC** y la **Real Sociedad**.
- **Posición:** se trata de una variable categórica que expresa el lugar que ocupa cada jugador en el campo: **1** si el jugador es portero, **2** si es defensa, **3** si es centrocampista y **4** si es delantero o atacante.
- **Edad:** es una variable numérica que indica los años que tiene cada jugador a 5 de junio de 2023.
- **Tipo_ formación:** se trata de una variable categórica que indica el sistema de juego que ha empleado cada jugador esta temporada. Para los porteros, asignamos un **0** porque en todas las formaciones solamente hay un portero sobre el campo. Para los defensas, asignamos un **1** si han jugado con 4 defensas o un **2** si han jugado con 5 defensas en el once titular. Para los centrocampistas, asignamos un **3** si han jugado con 3 centrocampistas o un **4** si han jugado con 4 centrocampistas en el once titular. Para los delanteros: asignamos un **5** si han jugado con 3 atacantes, un **6** si han jugado con 2 atacantes o un **7** si han jugado con 1 solo atacante en el once titular.
- **T_amarillas:** se trata de una variable numérica que indica la cantidad de tarjetas amarillas que ha recibido el jugador durante toda su carrera en todas las competiciones hasta el día 20 de febrero de 2023.
- **T_rojas:** se trata de una variable numérica que indica la cantidad de tarjetas rojas que ha recibido el jugador durante toda su carrera en todas las competiciones hasta el día 20 de febrero de 2023.

- **Goles:** se trata de una variable numérica que indica el porcentaje de goles respecto a los disparos realizados por el jugador en cada partido de la actual temporada 2022/2023 hasta la jornada 22. La variable original que hemos utilizado se llama *goals conversion (%)*.
- **Pases:** se trata de una variable numérica que indica el porcentaje de precisión de pases respecto a todos los que ha realizado durante cada partido de la actual temporada 2022/2023 hasta la jornada 22. La variable original que hemos utilizado se llama *accurate per games (%)*.
- **Defensa:** se trata de una variable numérica que indica el porcentaje de acciones defensivas realizadas con éxito respecto a todas las que ha realizado durante cada partido de la actual temporada 2022/2023 hasta la jornada 22. La variable original que hemos utilizado se llama *total duels won (%)*.
- **Paradas:** se trata de una variable numérica que indica el porcentaje de paradas realizadas respecto a todos los disparos recibidos durante cada partido de la actual temporada 2022/2023 hasta la jornada 22. La variable original que hemos utilizado se llama *saves per game (%)*.
- **Nota_20_21:** se trata de una variable numérica que indica las calificaciones obtenidas por cada jugador en la temporada 2020/2021. Hay que tener en cuenta que algunos jugadores no jugaban para el club actual o aún no habían debutado en el primer equipo.
- **Nota_21_22:** se trata de una variable numérica que indica las calificaciones obtenidas por cada jugador en la temporada 2021/2022. Hay que tener en cuenta que algunos jugadores no jugaban para el club actual o aún no habían debutado en el primer equipo.
- **Nota_22_23:** se trata de una variable numérica que indica las calificaciones obtenidas por cada jugador en la actual temporada 2022/2023 hasta la jornada 22.
- **Min_20_21:** se trata de una variable numérica que indica el porcentaje de minutos jugados respecto a los posibles en cada liga correspondiente en la temporada 2020/2021. Hay que tener en cuenta que varios jugadores no jugaron en la liga española y la cantidad de partidos disponibles no es

el mismo¹⁴. Los valores los hemos calculado haciendo la división entre los minutos jugados y los posibles.

- **Min_21_22:** se trata de una variable numérica que indica el porcentaje de minutos jugados respecto a los posibles en cada liga correspondiente en la temporada 2021/2022. Hay que tener en cuenta que varios jugadores no jugaron en la liga española y la cantidad de partidos disponibles no es el mismo. Los valores los hemos calculado haciendo la división entre los minutos jugados y los posibles.
- **Min_22_23:** se trata de una variable numérica que indica el porcentaje de minutos jugados respecto a los posibles en la liga española en la temporada 2022/2023. Los valores los hemos calculado haciendo la división entre los minutos jugados y los posibles.
- **Selección:** se trata de una variable binaria. Le asignamos el valor **1** si el jugador ha ido alguna vez convocado con la selección absoluta, y **0** en caso contrario.
- **Min_selección:** se trata de una variable numérica que indica la cantidad de minutos disputados de cada jugador en la selección absoluta de su país.
- **Precio_actual:** se trata de una variable numérica que indica el valor de mercado correspondiente a cada jugador expresado en millones de euros.
- **Precio_fichaje:** se trata de una variable numérica que indica el precio que pagó el club por el fichaje del jugador expresado en millones de euros.
- **Lesiones:** se trata de una variable binaria. Le asignamos el valor **1** si el jugador tiende a lesionarse (2 o más veces en los últimos 12 meses), y **0** en caso contrario.
- **Salario:** se trata de una variable numérica que indica la cantidad de dinero que percibe cada jugador al año expresado en millones de euros.

14. En este [enlace](#) podemos ver la cantidad de minutos totales que se juega en las ligas más importantes.

6. Resultados

6.1. Preparación de Datos

Primero de todo, vamos a cargar los datos recogidos en la base de datos que hemos explicado en el punto anterior. Vemos que tenemos una cantidad de 117 jugadores. Podemos observar que se dividen en 5 grandes grupos o equipos: el 18.8% de los futbolistas pertenecen al FC Barcelona, el 19.65% pertenecen al Valencia CF, el 19.65% pertenecen al Levante UD, el 23.93% pertenecen al Sevilla FC y el 21.36% pertenecen a la Real Sociedad.

Antes de empezar con todos los análisis, vamos a limpiar y procesar los datos para comprobar que todas las variables están como queremos nosotros. Podemos observar que *precio_fichaje* no es numérica porque algunos valores son "cedido". Por tanto, lo que vamos a hacer es sustituirlo por el valor 0.1, ya que el precio mínimo que paga un club por la cesión es 100000 euros entre el traspaso y el salario compartido. Una vez sustituido, transformamos la variable en numérica para poder seguir con el proyecto.

A	precio_fichaje	Min. : 0.000
B	precio_fichaje	1st Qu.: 0.000
C	precio_fichaje	Median : 0.100
D	precio_fichaje	Mean : 6.815
E	precio_fichaje	3rd Qu.: 6.500
F	precio_fichaje	Max. :125.000

Cuadro 2: summary de una variable numérica. Fuente: elaboración propia.

A	jugadores	Length:117
B	jugadores	Class :character
C	jugadores	Mode :character

Cuadro 3: summary de una variable carácter. Fuente: elaboración propia.

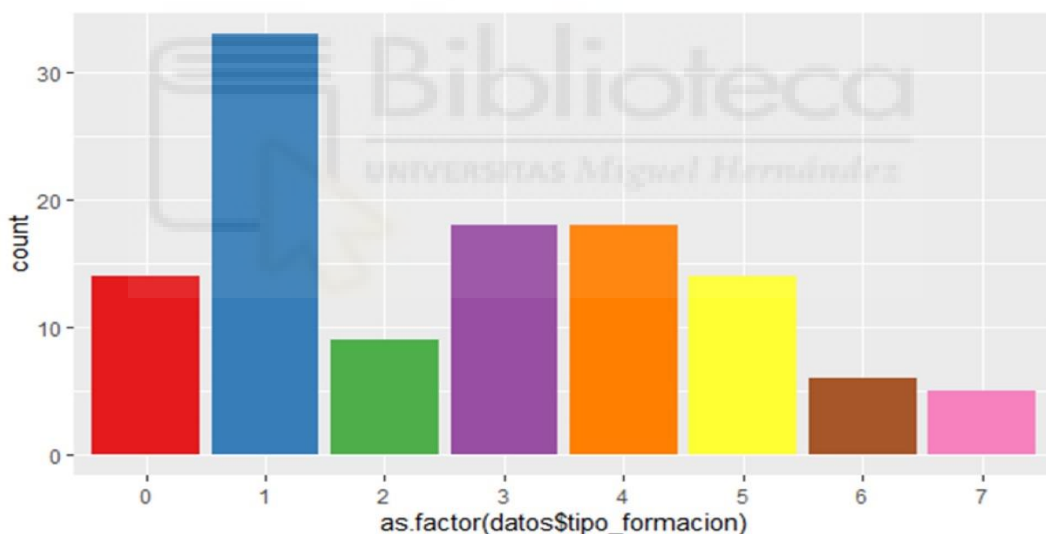
Una vez hemos limpiado, preparado y procesado los datos, vamos a realizar la estadística descriptiva.

6.2. Estadística Descriptiva

Tras haber hecho las modificaciones necesarias en el apartado anterior, procedemos a efectuar un análisis de los datos que tenemos. Lo necesario para llevar a cabo un buen descriptivo es lo siguiente: un análisis univariante, bivariante y multivariante.

Primero de todo vamos a empezar con el **análisis univariante**. Aquí solamente vamos a estudiar una sola variable de todas las que tenemos en nuestra base de datos. En este caso, analizaremos la variable *tipo_formacion*.

La información que queremos sacar es el sistema de juego que más se usa entre estos jugadores y la menos utilizada para jugar los partidos. Para poder verlo gráficamente, vamos a utilizar un *barplot* o un diagrama de barras.



```

0          1          2          3          4          5
0.11965812 0.28205128 0.07692308 0.15384615 0.15384615 0.11965812
6          7
0.05128205 0.04273504

```

Gráfico 1: diagrama de barras sobre el uso de cada tipo de formación. Fuente: elaboración propia.

También hemos hecho una tabla que nos muestre los porcentajes exactos para poder hacer nuestro análisis más exacto.

Podemos observar como la formación más utilizada en la línea defensiva es la formada por 4 jugadores con un 28.2% respecto a la formada por 3 con un 7.69%. Para la línea de creación del juego, la de los centrocampistas, el uso de ambas formaciones es exactamente la misma con un 15.38%. Por tanto, se usarán tanto 3 como 4 jugadores. Por último, el área atacante utiliza bastante más la línea de 3 con un 11.97% que la de 2 o 1 delantero (ambas cerca del 5%).

En resumen, el sistema de juego que más emplean en sus equipos estos jugadores es la formada por un portero, 4 defensores, 3 mediocentros y 3 delanteros. La formación 4-3-3 es de las más populares en el fútbol español, ya que se caracteriza por el dominio de la posesión del balón y no tanto por los contraataques característicos del fútbol inglés. La segunda que más se emplea es la 4-4-2, la cual es una conversión del sistema anterior, ya que uno de los dos delanteros siempre será referencia y uno de los mediocentros será más defensivo que el otro. Por último, el sistema menos popular y utilizado es el 5-4-1, ya que solamente lo usan los equipos que quieren salir rápido al contraataque y no tener ningún tipo de protagonismo en el partido.

Ahora vamos a realizar un **análisis bivariante**, el cual trata de estudiar dos variables de la base de datos y como se relacionan entre ellas.

En este caso, analizaremos la relación que hay entre la posición que ocupa cada jugador en el campo y el salario que percibe anualmente. Por tanto, las variables que vamos a estudiar son *salario* y *posición*. Para poder verlo, vamos a realizar un histograma.

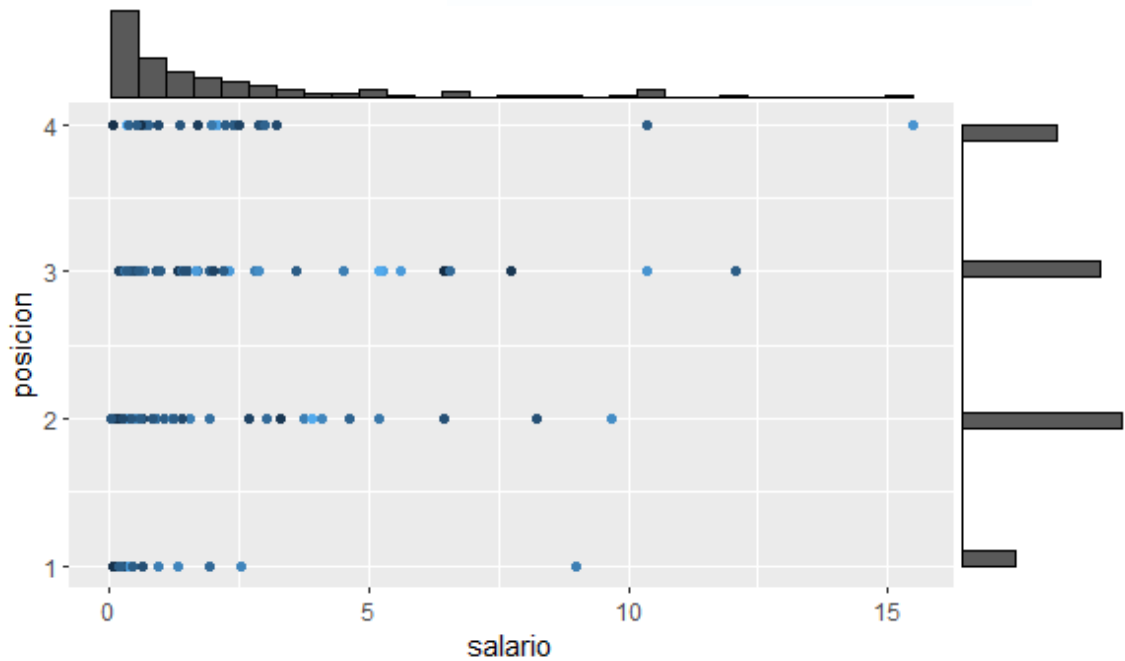


Gráfico 2: histograma sobre el salario de los jugadores en función de la posición del campo.
Fuente: elaboración propia.

Como hemos visto en la imagen, el jugador que más dinero ingresa cada año ocupa la posición de delantero con un salario de más de 15 millones de euros anuales. En este caso, si vamos a nuestros datos, podemos observar que es el jugador polaco del FC Barcelona *Robert Lewandowski*.

A parte de este dato, los segundos que más ingresan son los centrocampistas con más de 10 millones de euros por temporada. Luego ya observamos que los mejores defensas y el mejor portero perciben menos de 10 millones de euros. Este orden jerárquico no solamente ocurre en España, sino que también a nivel mundial, ya que la importancia de los goles es bastante más elevada que la de hacer pases, defender o realizar paradas defensivas. Esto también lo podemos ver en la entrega de premios, donde los galardonados son normalmente atacantes.

Por tanto, en líneas generales, los jugadores que ocupan la posición de atacantes son los que más perciben, aunque, a excepción de los dos mejores, los otros cobran bastante menos que los mediocentros y los defensas. La información que podemos concluir es la siguiente: el mejor jugador del FC Barcelona es el delantero y, por tanto, cobra más; pero los otros clubes tienen a sus jugadores más determinantes en la línea defensiva y del centro del campo.

También vamos a analizar la relación que existe entre la edad de cada futbolista con las tarjetas que ha recibido durante toda su carrera. Para ello, vamos a realizar un *boxplot* o diagrama de cajas.

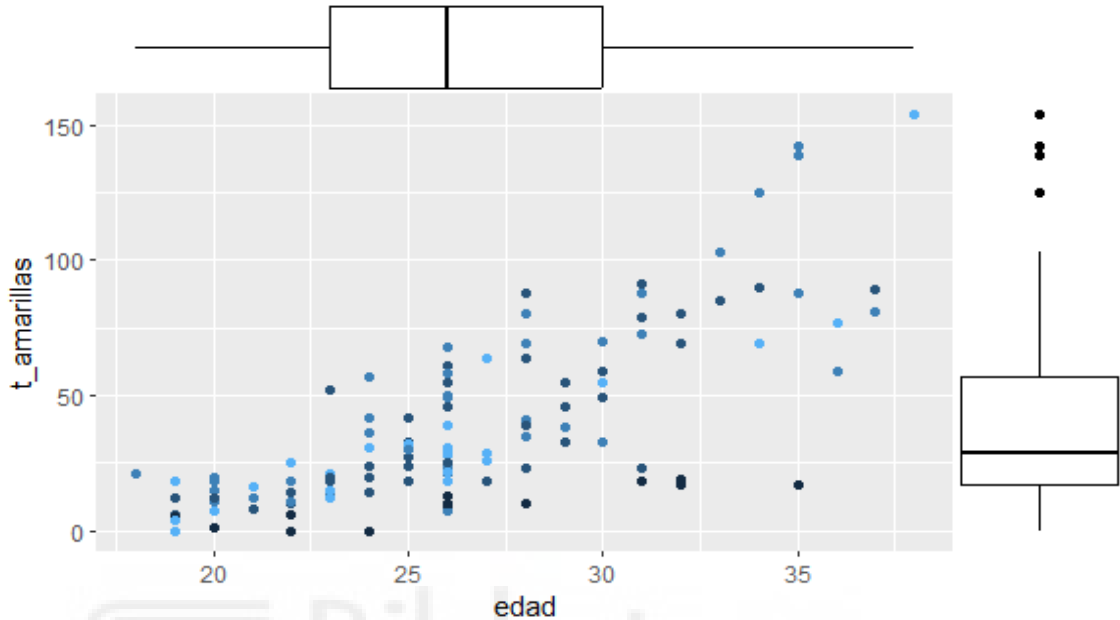
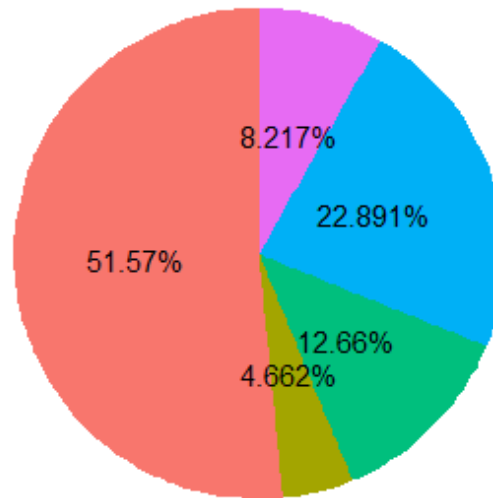


Gráfico 3: diagrama de cajas sobre las tarjetas amarillas obtenidas según la edad que tienen.
Fuente: elaboración propia.

Como era de esperar, los jugadores que más años de carrera tienen a sus espaldas son los que más han sido sancionados con tarjetas amarillas. Podemos observar que hay 4 casos atípicos, pero, si nos fijamos bien, estos pertenecen a futbolistas cerca de los 35 años, por lo que no es nada extraño de ver.

Más allá de esta excepción, la gran mayoría de jugadores tienen menos de 100 tarjetas amarillas. Esto se debe a varias razones. Una de ellas es la posición del campo, ya que un defensa siempre va a ser más apercibido que otra línea del campo y los menos sancionados serán los porteros, esto se debe a su poca intervención en el juego. Otra razón es por la edad de los jugadores como ya hemos dicho anteriormente. Cuanto más joven sea, menos tarjetas habrá percibido como norma general.

Además de los dos análisis anteriores, también vamos a ver qué equipo es el que más salario le ingresa a sus jugadores. Para esto, vamos a realizar un *piechart* o gráfico circular.



grupo ■ FC Barcelona ■ Levante UD ■ Real Sociedad ■ Sevilla FC ■ Valencia CF

Gráfico 4: gráfico circular sobre el porcentaje de salario que gasta cada equipo.
Fuente: elaboración propia.

Podemos observar como el club que más gastos tiene en cuanto a salarios es el FC Barcelona con un 51.57% respecto a los demás. Esto es lógico porque la entidad catalana es una de las mejores del mundo, por lo tanto, necesita los mejores jugadores para poder competir por todas las competiciones tanto nacionales como continentales para poder mantener su nivel.

Además del FC Barcelona, el club que más paga a sus jugadores es el Sevilla FC con un 22.89%. El club hispalense también trata de pelear por las máximas competiciones, pero siempre a menor exigencia que los culers. Los demás clubes que les siguen son la Real Sociedad y el Valencia CF, los cuales no tienen una participación tan habitual en competiciones europeas. El Levante UD es el que menos porcentaje de salario tiene con un 4.66% porque actualmente está jugando en la segunda división española y no puede mantener los jugadores y los salarios que se podía permitir en la máxima división.

Para finalizar con el análisis bivariante, vamos a calcular el coeficiente de correlación que existe entre cada una de las variables de nuestra base de datos.

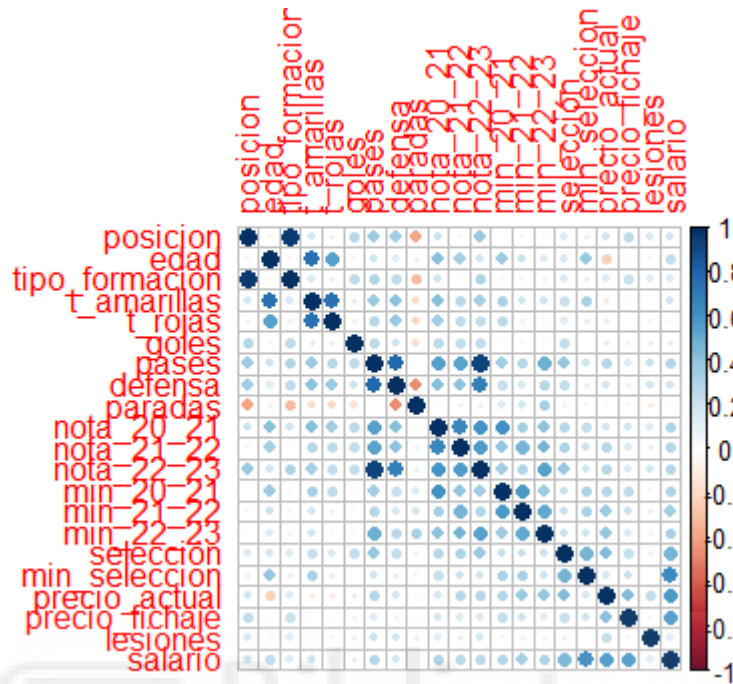


Gráfico 5: coeficientes de correlación entre las variables. Fuente: elaboración propia.

Como podemos observar, las variables que más relación directa tienen entre sí son *posicion* con *tipo_formacion*, *pases* con *nota_22_23*, *edad* con *t_amarillas*, *t_rojas* con *t_amarillas*, *defensa* con *pases* y *defensa* con *nota_22_23*. Por otro lado, existe relación indirecta entre *posicion* y *paradas*, y entre *defensa* y *paradas*.

Por último, vamos a realizar un **análisis multivariante**, el cual trata de estudiar más de dos variables de la base de datos.

En este caso, vamos a estudiar la relación entre la variable *jugadores*, *equipo*, *min_20_21*, *min_21_22* y *min_22_23*. Es decir, nuestro objetivo es saber que futbolistas han tenido más minutos cada temporada en cada uno de sus equipos. Por ejemplo, vamos a observar gráficamente a través de un *circular barplot* o diagrama de barras circular los minutos que han jugado los valencianistas cada una de las 3 últimas temporadas. Hay que tener en cuenta que varios de ellos no jugaron anteriormente en el equipo che y esta es su primera temporada.

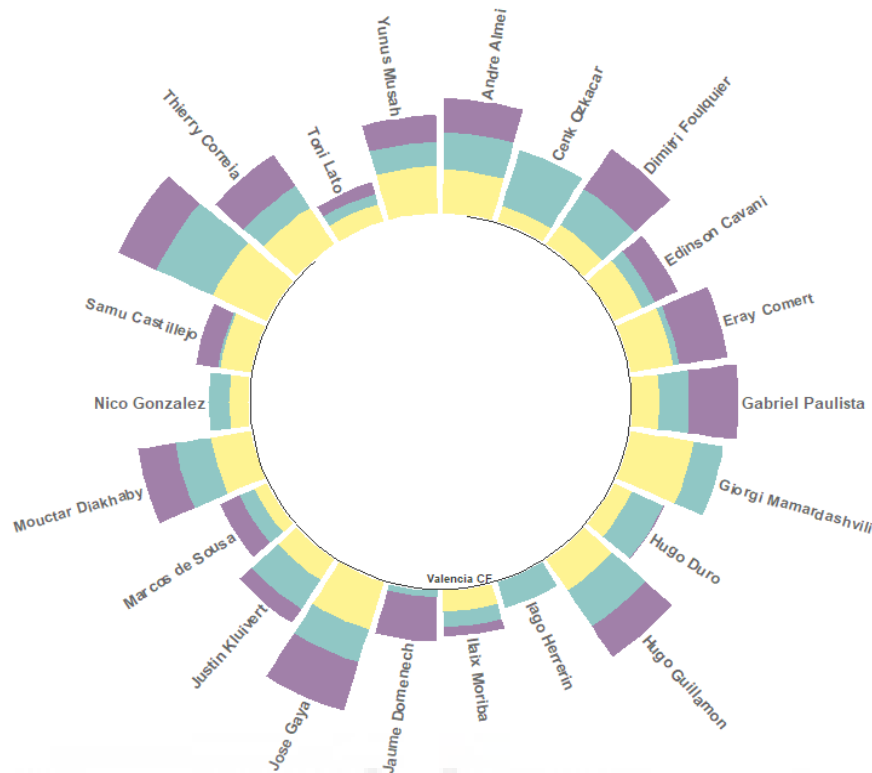


Gráfico 6: diagrama de barras circular sobre los minutos de las 3 temporadas de los jugadores del Valencia CF. Fuente: elaboración propia.

Como podemos observar, tenemos 3 colores que representan cada una de las temporadas. El **color morado** pertenece a la **campana 2020/2021**. El **color azul claro** pertenece a la **campana 2021/2022**. El **color amarillo** pertenece a la actual **temporada 2022/2023 hasta la jornada 22**.

Empezando por la portería, el guardameta que más ha jugado esta temporada ha sido *Mamardashvili* con la totalidad de los partidos. En el año anterior, los 3 jugadores compartieron titularidad acumulando cada uno de ellos varios minutos. Sin embargo, en la temporada 2020/2021, solamente *Jaume Domenech* fue el dueño de la portería con la totalidad de los minutos. Esto puede deberse a que sus otros dos compañeros no estaban en el club valencianista.

Siguiendo con la línea defensiva, los 4 jugadores con más minutos acumulados en la actual temporada han sido el capitán *José Gayá*, *Mouctar Diakhaby*, *Thierry Correia* y *Eray Comert*. Estos 4 futbolistas han tenido más o menos el mismo nivel de minutos durante las temporadas anteriores, aunque los únicos futbolistas que se pueden acercar han sido *Gabriel Paulista* y *Hugo Guillamón* que ha estado las 3 temporadas en el Valencia CF.

En la línea de creación del juego, los 3 jugadores que más protagonismo han tenido durante esta campaña han sido *Yunus Musah*, *André Almeida* y *Samu Castillejo*. Los dos últimos es el primer año que juegan en el club. Por último, los atacantes que más han jugado han sido *Cavani*, *Samuel Lino* y *Justin Kluivert*. Estos 3 delanteros también es el primer año que se encuentran en el equipo.

Por tanto, así quedaría el 11 titular con los jugadores que más minutos han acumulado durante el curso actual.



Figura 3: 11 titular del Valencia CF según los minutos jugados. Fuente: FUTBIN.

Para el club valenciano esta ha sido una de las peores temporadas tanto a nivel de títulos como financiero, ya que sus mejores jugadores han tenido que abandonar el club para poder cuadrar las cuentas y no hundir al club en la miseria. Es por eso por lo que han tenido que conseguir varias cesiones de jugadores jóvenes y agentes libres disponibles en el mercado.

Por otro lado, también hemos comparado las variables *equipo*, *precio_actual*, *precio_fichaje*, *salario* y *edad*. Lo que vamos a hacer es compararlas por sus valores medios.

equipo	precio_actual	precio_fichaje	salario	edad
<chr>	<dbl>	<dbl>	<dbl>	<dbl>
1 FC Barcelona	34.6	21.0	6.40	25.5
2 Levante UD	1.75	0.548	0.579	27.2
3 Real Sociedad	15.6	3.69	1.44	26.0
4 Sevilla FC	8.91	6.40	2.32	27.8
5 Valencia CF	11.1	2.82	1.02	25.8

Cuadro 4: valores medios del precio actual, precio fichaje, salario y edad en función del equipo.
Fuente: elaboración propia.

Como podemos observar, el club que tiene la plantilla más cara es el FC Barcelona con 34.6 millones de euros de media, con una media de fichajes de 21 millones y un salario medio de 6.4 millones de euros. Es bastante más elevado que los demás, ya que es un club con bastante más exigencias deportivas que los demás. En cambio, es el club más joven de las 5 que hemos analizado con una mediana de 25.5 años, por lo que podemos intuir que el valor de *precio_actual* subirá con el paso de los años junto a los salarios.

En cambio, el Levante UD es el más barato, el que menos coste en fichajes ha gastado y el que menos salario gasta en sus jugadores. Esto tiene sentido, ya que es un club que se encuentra en Segunda División, y el dinero que perciben por victorias y por objetivos es bastante más bajo que un club que esté en Primera División. A esto le sumamos que han tenido que acudir al mercado de agentes libres, donde la edad media de los jugadores tiende a ser mayor.

Por último, vemos que hay un caso bastante curioso, ya que, si miramos la actual clasificación, el Sevilla FC está bastante más por delante que el club valencianista, teniendo estos un valor de plantilla medio mayor que los hispalenses. Esto nos demuestra, que en el mundo del fútbol no todo es siempre tener a los mejores jugadores donde el más rico gana y el más pobre pierde.

6.3. Notas jugadores sin algoritmo

Después de haber hecho un análisis estadístico de algunas variables y haber sacado la información necesaria para poder seguir con el proyecto, vamos a calcular una valoración que englobe el rendimiento de las 3 temporadas.

Vamos a darle unas ponderaciones para las habilidades de cada jugador (*paradas, defensa, pases, goles*). Cada una de ellas será diferente dependiendo de la posición que ocupen en el campo. Le vamos a dar importancia también de forma positiva a las convocatorias con la selección absoluta y de forma negativa a las sanciones de tarjetas amarillas y rojas y a las lesiones. También lo multiplicaremos por 2 para poder valorar en una escala del 0 al 10.

Para los **porteros**, le he dado un 100% de importancia a la variable paradas porque es la labor principal de un guardameta, y un 50% a pases para que también sea importante la salida de balón desde atrás. Las otras variables de defensa y goles no les he dado protagonismo. La fórmula para sacar sus notas sería la siguiente:

$$\begin{aligned} & \left(pases * \min_{22,23} * 0.5 + paradas * \min_{22,23} * 1 + \frac{nota_{20,21} * \min_{20,21}}{10} \right. \\ & \quad + \frac{nota_{21,22} * \min_{21,22}}{10} + \frac{nota_{22,23} * \min_{22,23}}{10} + seleccion \\ & \quad + \left(\frac{\min_{seleccion}}{\max[\min_{seleccion_1}]} \right) - \left(\frac{t_{amarillas}}{\max[t_{amarillas_1}]} \right) - \left(\frac{t_{rojas}}{\max[t_{rojas_1}]} \right) \\ & \quad \left. + \left(\frac{precio_{actual}}{\max[precio_{actual_1}]} \right) - lesiones \right) * 2 \end{aligned}$$

Para los **defensores**, le he dado un 100% de importancia a la variable *defensa* porque es su labor principal, un 67% a *pases* para que también sea importante la salida de balón desde atrás y, por último, un 33% a *goles*. La otra variable de *paradas* no le he dado protagonismo. La fórmula para sacar sus notas sería la siguiente:

$$\begin{aligned} & \left(\textit{pases} * \textit{min}_{22,23} * 0.67 + \textit{goles} * \textit{min}_{22,23} * 0.33 + \textit{defensa} * \textit{min}_{22,23} * 1 \right. \\ & + \frac{\textit{nota}_{20,21} * \textit{min}_{20,21}}{10} + \frac{\textit{nota}_{21,22} * \textit{min}_{21,22}}{10} + \frac{\textit{nota}_{22,23} * \textit{min}_{22,23}}{10} \\ & + \textit{seleccion} + \left(\frac{\textit{min}_{\textit{seleccion}}}{\max[\textit{min}_{\textit{seleccion}_2}]} \right) - \left(\frac{\textit{t}_{\textit{amarillas}}}{\max[\textit{t}_{\textit{amarillas}_2}]} \right) \\ & \left. - \left(\frac{\textit{t}_{\textit{rojas}}}{\max[\textit{t}_{\textit{rojas}_2}]} \right) + \left(\frac{\textit{precio}_{\textit{actual}}}{\max[\textit{precio}_{\textit{actual}_2}]} \right) - \textit{lesiones} \right) * 2 \end{aligned}$$

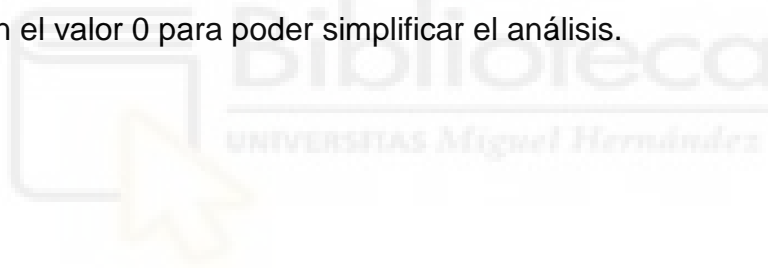
Para los **centrocampistas**, le he dado un 100% de importancia a la variable *pases* porque es su labor principal, un 67% a *goles y defensa* para que también lleguen a portería con efectividad y defiendan las jugadas rivales. La otra variable de *paradas* no le he dado protagonismo. La fórmula para sacar sus notas sería la siguiente:

$$\begin{aligned} & \left(\textit{pases} * \textit{min}_{22,23} * 1 + \textit{goles} * \textit{min}_{22,23} * 0.67 + \textit{defensa} * \textit{min}_{22,23} * 0.67 \right. \\ & + \frac{\textit{nota}_{20,21} * \textit{min}_{20,21}}{10} + \frac{\textit{nota}_{21,22} * \textit{min}_{21,22}}{10} + \frac{\textit{nota}_{22,23} * \textit{min}_{22,23}}{10} \\ & + \textit{seleccion} + \left(\frac{\textit{min}_{\textit{seleccion}}}{\max[\textit{min}_{\textit{seleccion}_3}]} \right) - \left(\frac{\textit{t}_{\textit{amarillas}}}{\max[\textit{t}_{\textit{amarillas}_3}]} \right) \\ & \left. - \left(\frac{\textit{t}_{\textit{rojas}}}{\max[\textit{t}_{\textit{rojas}_3}]} \right) + \left(\frac{\textit{precio}_{\textit{actual}}}{\max[\textit{precio}_{\textit{actual}_3}]} \right) - \textit{lesiones} \right) * 2 \end{aligned}$$

Para los **atacantes**, le he dado un 100% de importancia a la variable *goles* porque es su labor principal, un 67% a *pases* para que también sea importante la transición entre ellos para dar el último pase y, finalmente, un 33% a *defensa*. La otra variable de ***paradas*** no le he dado protagonismo. La fórmula para sacar sus notas sería la siguiente:

$$\begin{aligned} & \left(\textit{pases} * \textit{min}_{22,23} * 0.67 + \textit{goles} * \textit{min}_{22,23} * 1 + \textit{defensa} * \textit{min}_{22,23} * 0.33 \right. \\ & + \frac{\textit{nota}_{20,21} * \textit{min}_{20,21}}{10} + \frac{\textit{nota}_{21,22} * \textit{min}_{21,22}}{10} + \frac{\textit{nota}_{22,23} * \textit{min}_{22,23}}{10} \\ & + \textit{seleccion} + \left(\frac{\textit{min}_{\textit{seleccion}}}{\max[\textit{min}_{\textit{seleccion}_4}]} \right) - \left(\frac{\textit{t}_{\textit{amarillas}}}{\max[\textit{t}_{\textit{amarillas}_4}]} \right) \\ & - \left(\frac{\textit{t}_{\textit{rojas}}}{\max[\textit{t}_{\textit{rojas}_4}]} \right) + \left(\frac{\textit{precio}_{\textit{actual}}}{\max[\textit{precio}_{\textit{actual}_4}]} \right) - \textit{lesiones} \left. \right) * 2 \end{aligned}$$

Por último, los jugadores que tenían valoraciones negativas las hemos convertido en el valor 0 para poder simplificar el análisis.



6.4. Notas jugadores con algoritmo

La valoración de los jugadores que hemos sacado anteriormente ha sido creada e inventada totalmente por mí. Existen 3 variables que hemos extraído de la página web *Sofascore* que expresaban la valoración media de todos los futbolistas en cada temporada que habían jugado. Estas son: *nota_20_21*, *nota_21_22* y *nota_22_23*.

Después de haber investigado un poco por los foros y los blogs de los seguidores de esta aplicación, en todos ellos indicaban que el algoritmo que usan en este portal es secreto, ya que no quieren que se desvelen como lo usan. Por tanto, nosotros vamos a emplear la estadística para ver el margen de error que tenemos en nuestros resultados.

Concretamente, vamos a emplear **regresión lineal múltiple**.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots \beta_n X_n + \varepsilon,$$

Figura 4: ecuación general regresión lineal múltiple. Fuente: Econometría UMH del Elche.

En esta imagen podemos observar cómo consideramos el modelo de regresión lineal múltiple, donde las X son las variables independientes, Y es la variable dependiente y ε es el error aleatorio. Esto nos permitirá saber que variables son las que se utilizaran en el modelo y la ponderación o importancia que tienen cada una de ellas. Cabe recordar que las variables que indican las habilidades de los jugadores pertenecen a la actual temporada.

Para empezar, vamos a analizar la **temporada 2020/2021**.

```
Step: AIC=144.39
nota_20_21 ~ edad + goles + pases + min_20_21

Coefficients:
(Intercept)      edad      goles      pases      min_20_21
-1.3211         0.1044         3.3619         3.2308         3.5454
```

Cuadro 5: modelo de regresión lineal múltiple temporada 2020/2021. Fuente: elaboración propia.

Como podemos observar, el modelo que se utiliza para sacar las notas de la campaña 2020/2021 tiene un AIC=144.39 y es el siguiente:

$$nota_{20,21} = -1.3211 + 0.1044 * edad + 3.3619 * goles + 3.2308 * pases + 3.5454 * min_{20,21}$$

Una vez tenemos el modelo y hemos sacado las notas pertenecientes a esa temporada, vamos a ver el error que hay en cada una de las valoraciones de cada jugador entre el Big Data y Sofascore. Para poder verlo, hemos calculado la diferencia en valor absoluto entre los valores calculados recientemente y las notas que nos proporcionaba nuestra base de datos. También lo hemos representado en un histograma y un gráfico circular para verlo gráficamente.

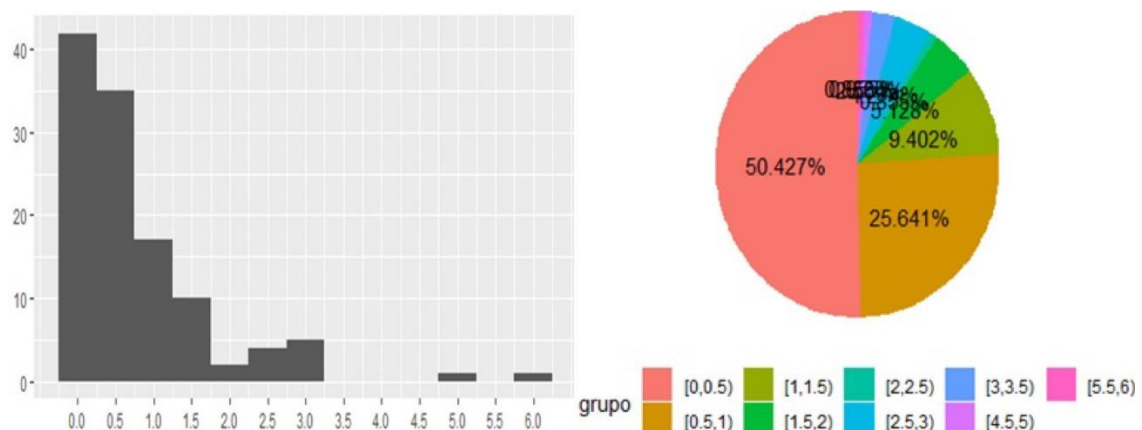


Gráfico 7: representación gráfica del porcentaje de margen de error temporada 2020/2021. Fuente: elaboración propia.

Como se aprecia en ambas imágenes, el margen de error más frecuente se encuentra entre un 0 y 0.5 con un 50.42%, lo cual nos indica que más de la

mitad de las valoraciones que hemos obtenido con el modelo se acercan al algoritmo secreto. A partir de ahí, tenemos que un cuarto de las notas se equivoca hasta por un punto, lo cual no es una barbaridad, pero ya no es tan exacto como lo anterior. Más allá, nos encontramos con menos frecuencia errores mayores que 1 hasta tener dos casos donde la diferencia entre una valoración y otra es de 5 y 6 puntos. Esto es algo que nos podíamos esperar, ya que estamos implementando en el modelo variables que no pertenecen a esta temporada.

Ahora seguiremos con la **temporada 2021/2022**.

```
Step: AIC=83.86
nota_21_22 ~ edad + tipo_formacion + t_amarillas + pases + min_20_21 +
min_21_22 + nota_20_21 + precio_actual + salario

Coefficients:
(Intercept)          edad  tipo_formacion  t_amarillas
-0.80354          0.13014         -0.11744         -0.01095
pases          min_20_21  min_21_22          nota_20_21
 2.02501         -1.67718          2.16870          0.42890
precio_actual          salario
 0.02402         -0.09108
```

Cuadro 6: modelo de regresión lineal múltiple temporada 2021/2022. Fuente: elaboración propia.

Como podemos ver, el modelo resultante utilizado para sacar las calificaciones de la campaña 2021/2022 tiene un AIC=83.86 y es el siguiente:

$$\begin{aligned}
 nota_{21,22} = & -0.80354 + 0.13014 * edad - 0.11744 * tipo_{formacion} - 0.01095 \\
 & * t_{amarillas} + 2.02501 * pases - 1.67718 * min_{20,21} + 2.16870 \\
 & * min_{21,22} + 0.42890 * nota_{20,21} + 0.02402 * precio_{actual} - 0.09108 \\
 & * salario
 \end{aligned}$$

Una vez tenemos el modelo y hemos sacado las notas pertenecientes a esa temporada, podemos observar como la nota del año anterior influye sobre la actual temporada de forma positiva. Ahora vamos a ver el error que hay en cada una de las valoraciones de cada jugador entre el Big Data y Sofascore. Para poder verlo, hemos calculado la diferencia en valor absoluto entre los valores calculados recientemente y las notas que nos proporcionaba nuestra base de datos, tal y como lo habíamos hecho en la temporada anterior. También lo hemos representado en un histograma y un gráfico circular para verlo gráficamente.

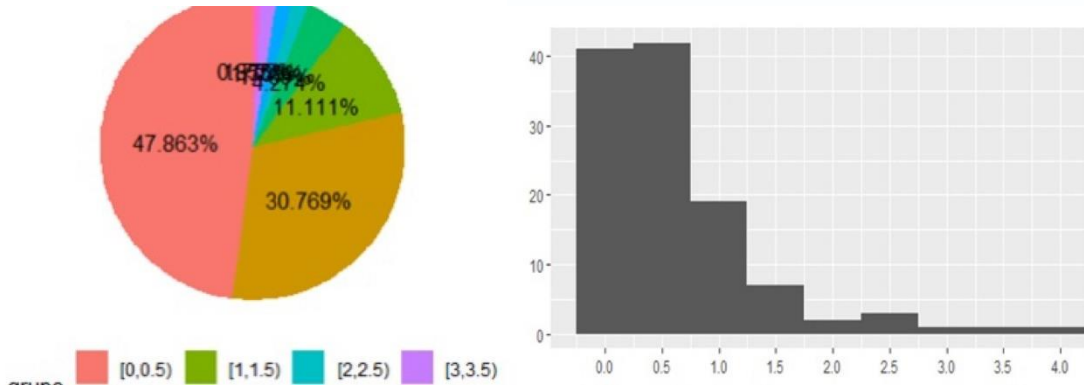


Gráfico 8: representación gráfica del porcentaje de margen de error temporada 2021/2022.
Fuente: elaboración propia.

Como se aprecia en ambas imágenes, el margen de error más frecuente se encuentra entre un 0 y 0.5 con un 47.863%, lo cual nos indica que un poco menos de la mitad de las valoraciones que hemos obtenido con el modelo se acercan al algoritmo secreto. A partir de ahí, tenemos que una tercera parte de las notas se equivoca hasta por un punto, lo cual no es muy grande, pero ya no es tan exacto como lo anterior. Más allá, nos encontramos con menos frecuencia errores mayores que 1 hasta tener un caso donde la diferencia entre una valoración y otra es de 4 puntos. Esto es algo que nos podíamos esperar, ya que estamos implementando en el modelo variables que no pertenecen a esta temporada. Cabe destacar que el máximo margen de error es más pequeño esta temporada que la anterior, esto se puede deber a que hemos implementado en el modelo la nota de la campaña 2020/2021.

Por último, vamos a analizar la **campeña 2022/2023 hasta la jornada 22.**

```
Step: AIC=-86.85
nota_22_23 ~ posicion + tipo_formacion + pases + paradas + min_22_23 +
            nota_20_21 + salario

Coefficients:
(Intercept)      posicion  tipo_formacion      pases
      -0.66872         0.92729         -0.29602         6.07088
      paradas      min_22_23      nota_20_21      salario
       1.89286         0.63320         0.09103        -0.04016
```

Cuadro 7: modelo de regresión lineal múltiple temporada 2022/2023. Fuente: elaboración propia.

Como podemos ver, el modelo resultante utilizado para sacar las calificaciones de la campaña 2022/2023 tiene un AIC=-86.85 (es el más bajo de los 3 calculados anteriormente) y es el siguiente:

$$\begin{aligned}
 nota_{22,23} = & -0.66872 + 0.92729 * posicion - 0.29602 * tipo_{formacion} + 6.07088 \\
 & * pases + 0.63320 * min_{22,23} + 0.09103 * nota_{20,21} - 0.04016 \\
 & * salario
 \end{aligned}$$

Una vez tenemos el mejor modelo y hemos sacado las notas pertenecientes a esa temporada, podemos observar como la nota de la temporada 2020/2021 y los minutos disputados en este mismo año influye de forma positiva. Ahora vamos a ver el error que hay en cada una de las valoraciones de cada jugador entre el Big Data y Sofascore. Para poder verlo, hemos calculado la diferencia en valor absoluto entre los valores calculados recientemente y las notas que nos proporcionaba nuestra base de datos, tal y como lo habíamos hecho en las anteriores temporadas. También lo hemos representado en un histograma y un gráfico circular para verlo gráficamente.

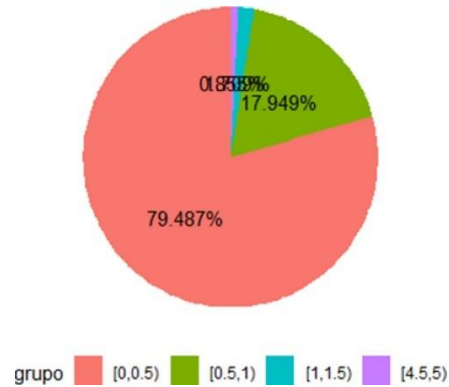
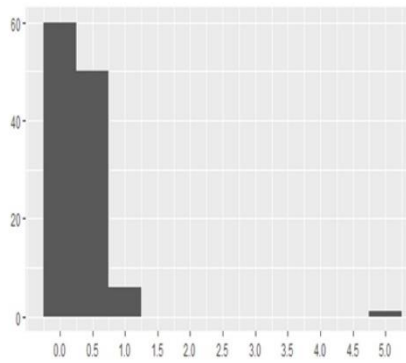


Gráfico 9: representación gráfica del porcentaje de margen de error temporada 2022/2023.
Fuente: elaboración propia.

Como se aprecia en ambas imágenes, el margen de error más frecuente se encuentra entre un 0 y 0.5 con un 79.487%, lo cual nos indica que solo una cuarta parte de las valoraciones que hemos obtenido con el modelo no se acercan al algoritmo secreto. A partir de ahí, tenemos que una quinta parte de las notas se equivoca hasta por un punto y medio, lo cual no es un problema, pero ya no es tan exacto como lo anterior. Más allá, nos encontramos con un solo caso excepcional que se equivoca por 5 puntos. La reducción del margen de error no es casualidad, ya que las variables que expresan las habilidades de los jugadores pertenecen a la misma temporada que estamos analizando.

Para concluir, vamos a observar el contraste entre los 3 histogramas juntos para saber qué año se acerca más a la realidad y si nos hemos podido acercar un poco a la gran incógnita que guarda esta página deportiva.

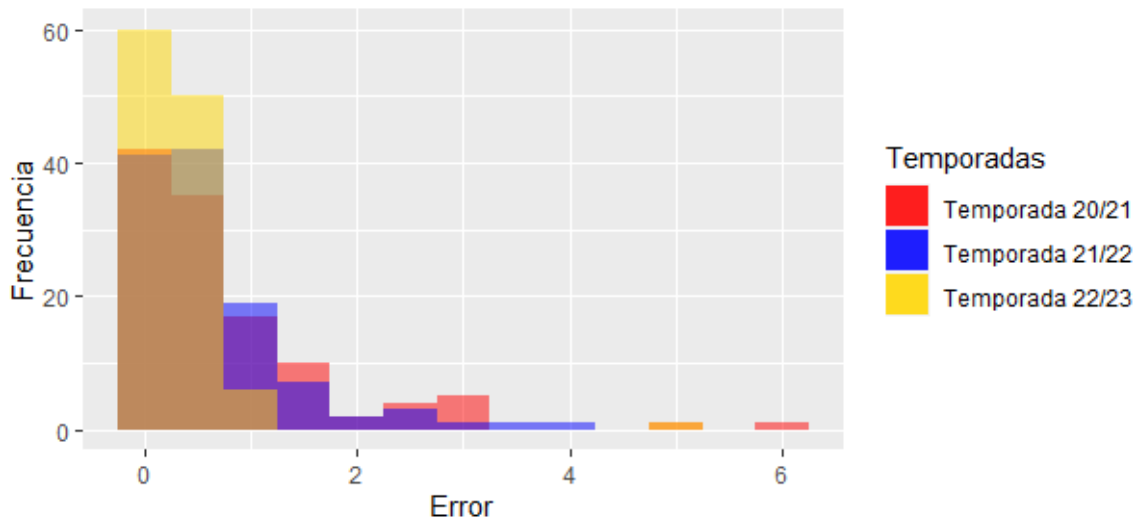


Gráfico 10: comparación de las 3 temporadas diferenciadas por colores. Fuente: elaboración propia.

El intervalo de 0 hasta 2 es el que se repite con más frecuencia, donde tenemos hasta un máximo de 60 casos en la actual temporada y unos 40 en las otras dos anteriores. Luego ya tenemos varios casos que tienen un error bastante más elevado de 2 y confirman que el modelo utilizado no se acerca demasiado para los casos de las dos últimas temporadas.

Por tanto, podemos concluir que la regresión que hemos realizado en la última campaña ha sido la más correcta y, por tanto, aunque no sea 100% efectivo, podríamos hacernos una idea de cómo los trabajadores especializados en Big Data de Sofascore han podido calcular las notas medias de cada temporada.

6.5. Modelo

Después de haber visto cómo podemos valorar las actuaciones de cada jugador por temporada, vamos al principal objetivo de nuestro proyecto. A día de hoy, el Valencia CF se encuentra en una situación muy delicada tanto a nivel económico como futbolístico. El club se encuentra muy cerca de los puestos de descenso, y aunque no parece que vayan a descender a la segunda categoría, una entidad de este nivel no puede repetir esta terrible temporada.

El uso de la estadística y el Big Data ha mejorado a varios equipos como ya dijimos anteriormente. Por tanto, nosotros vamos a emplear varias técnicas para poder buscar en nuestra base de datos jugadores que puedan mejorar a la actual plantilla por un precio asequible para la situación financiera del club.

Vamos a recordar los actuales integrantes de esta plantilla posición por posición, indicando también los jugadores cedidos:

- **PORTEROS:** *Iago Herrerín, Jaume Domenech y Giorgi Mamardashvili.*
- **DEFENSAS:** como defensa central tenemos a *Gabriel Paulista, Hugo Guillamón, Mouctar Diakhaby, Cenk Ozkacar* (cedido) y *Eray Comert*; como lateral izquierdo nos encontramos a *Jose Luís Gayá* y *Toni Lato*; por último, en el lateral derecho tenemos a *Thierry Correia* y *Dimitri Foulquier.*
- **CENTROCAMPISTAS:** como pivote defensivo tenemos a *Ilaix Moriba* (cedido) y *Nico González* (cedido); como interior tenemos a *Yunus Musah, Samu Castillejo* y *Andre Almeida.*
- **ATACANTES:** como extremos tenemos a *Justin Kluivert* (cedido) y *Samuel Lino* (cedido); como delantero centro tenemos a Edinson Cavani, *Hugo Duro* y *Marcos de Sousa.*

Aunque la demarcación principal es la que hemos indicado anteriormente, también pueden ocupar otras posiciones en el campo de juego. Por ejemplo, el defensor *Hugo Guillamón* también puede jugar como mediocentro defensivo. Todos los centrocampistas valencianistas pueden jugar tanto como interiores como pivotes más atrasados. Por último, el español *Samu Castillejo* también puede jugar como extremo. Todo esto dependerá de los jugadores sancionados y lesionados que tenga el club.

Cabe destacar que este año, la entidad valencianista ha cedido a un jugador que le puede dar bastante juego en el centro del campo, más concretamente como interior. Este futbolista es el serbio de 25 años *Uros Racic*.

Una vez sabemos toda la información correspondiente a la plantilla del Valencia CF en la temporada 2022/2023, vamos a realizar un modelo capaz de ajustarse al presupuesto disponible para la siguiente temporada y que pueda mejorar la posición que vaya a ocupar.

Tras la nefasta gestión de la presidenta singapurense *Lay Hoon Chan* a cargo del club valenciano, los principales accionistas del club han decidido venderlo a otra entidad. El principal objetivo es salvar la terrible situación económica del club y ayudarlo a salir de este bache tan grande y poder volver a puestos europeos.

Para ello, vamos a realizar un modelo estadístico que nos permita obtener los mejores jugadores por un precio asequible y que puedan mejorar las actuales posiciones. Los conjuntos que vamos a utilizar para realizar el modelo son los siguientes:

- **P:** jugadores que ocupan la posición de portero.
- **D:** jugadores que ocupan la posición de defensor.
- **C:** jugadores que ocupan la posición de centrocampista.
- **A:** jugadores que ocupan la posición de atacantes.

Los parámetros que hemos creado para poner en uso el modelo son los siguientes:

- **LT ψ y:** jugadores que tienen una edad inferior a ψ años.
- **MT ψ y:** jugadores que tienen una edad superior a ψ años.
- **MT ψ p:** jugadores que tienen un porcentaje en pases superior a $\psi\%$.
- **MT ψ d:** jugadores que tienen un porcentaje en defensa superior a $\psi\%$.
- **MT ψ g:** jugadores que tienen un porcentaje en goles superior a $\psi\%$.
- **MT ψ n:** jugadores que tienen una nota superior a ψ .
- **MT ψ min23:** jugadores que han disputado un porcentaje de minutos superior a $\psi\%$ en la actual temporada 2022/2023.

- **MT ψ nota:** jugadores que tienen una nota media de las tres temporadas superior a ψ . Esta calificación la hemos sacado de la siguiente forma:
$$\frac{nota_{20,21} + nota_{21,22} + nota_{22,23}}{3}$$
- **MT ψ min:** jugadores que han jugado un porcentaje de media de minutos entre las 3 temporadas de $\psi\%$. Esta calificación la hemos sacado de la siguiente forma:
$$\frac{min_{20,21} + min_{21,22} + min_{22,23}}{3}$$
- **LT ψ pa:** jugadores que tienen un precio de mercado actual inferior a ψ millones de euros.
- **MT ψ pa:** jugadores que tienen un precio de mercado actual superior a ψ millones de euros.

El símbolo ψ indica que podemos insertar el valor que queramos en cada momento de nuestras restricciones. Las abreviaturas **MT** y **LT** vienen del inglés “More Than” (mayor estricto que) y “Lower Than” (menor estricto que), respectivamente.

La única variable que vamos a utilizar es la siguiente:

- **X_i**: es una variable binaria que toma el 1 si ficho al jugador i, y 0 en caso contrario.

Por tanto, el modelo final con la función objetivo y sus respectivas restricciones quedaría así:

F.Obj: Min *precio_{actual}*

s.a.

$$\sum_{\{i \in LT28y \cap MT0n \cap P\}} x_i \geq 1$$

$$\sum_{\{i \in LT32y \cap MT71p \cap MT46d \cap MT25min23 \cap MT5nota \cap MT5pa \cap LT40pa \cap MT4n \cap D\}} x_i \geq 1$$

$$\sum_{\{i \in LT30y \cap MT20y \cap MT81p \cap MT50d \cap MT29min23 \cap MT4nota \cap MT35min \cap MT5pa \cap LT50pa \cap MT0n \cap C\}} x_i \geq 1$$

$$\sum_{\{i \in LT31y \cap MT22y \cap MT50p \cap MT3g \cap MT21min23 \cap MT65nota \cap MT40min \cap MT20pa \cap MT0n \cap A\}} x_i \geq 1$$

Primero de todo, la función objetivo que hemos planteado ha sido minimizar la variable *precio_actual*. Para esto, hemos creado un vector con todos los precios de mercado actuales de todos los jugadores de nuestra base de datos. Nuestro objetivo será escoger los valores más pequeños en función de la posición que queramos completar y sus respectivas restricciones.

Para los **porteros**, hemos visto que lo que más nos conviene es un guardameta que tenga menos de 28 años, y su nota sea mayor que 0. Los únicos que cumplen con estos requisitos han sido *Iñaki Peña*, *Arnau Tenas* y *Dani Cárdenas*. Desde el club nos han transmitido que quieren un portero de garantías y con experiencia, pero que no tenga el rol de titular. Por tanto, vamos a elegir al actual guardameta del Levante UD con un valor de traspaso de 1500000€: *Dani Cárdenas*. Para no tener un exceso de jugadores en esta posición, hemos decidido vender al tercer portero del club che por un valor de 500000€: *Iago Herrérin*. Lo vamos a vender a su anterior equipo: Athletic de Bilbao. Este movimiento nos genera un gasto de solamente 1 millón de euros.



Figura 5: comparación entre Dani Cárdenas y Iago Herrérin. Fuente: FUTBIN.

Siguiendo con los **defensas**, nos comunican que buscan un perfil con menos de 32 años, más de un 71% en pases completados con acierto, más de un 46% en acciones defensivas exitosas, que haya jugado más de un 25% de los minutos posibles esta temporada, que tenga una nota media de más de un 5, que tenga un precio de mercado entre 5 y 40 millones de euros y, por último, una nota mayor que 4. Después de analizar el mercado, los únicos defensores que cumplen estos requisitos son *Andreas Christensen*, *Eric García* y *Robin Le Normand*. Finalmente, hemos decidido que vamos a elegir al defensor español del FC Barcelona con un precio de traspaso de 18000000€: *Eric García*. A parte, vamos a vender a *Dimitri Foulquier* al Granada CF para rejuvenecer la plantilla

por un valor de 3000000€ y también activamos la opción de compra de *Cenk Ozkacar* por 1800000€. Por tanto, la nueva línea defensiva tendría un coste total de 16.8 millones de euros.

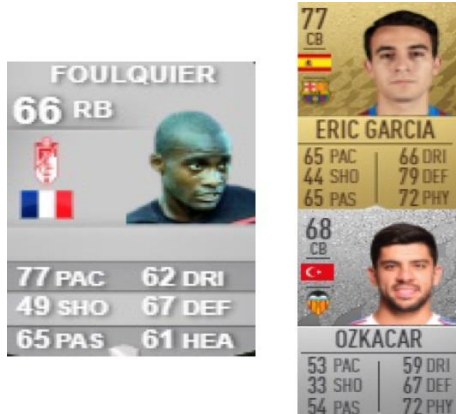


Figura 6: comparación entre Dimitri Foulquier, Eric García y Cenk Ozkacar. Fuente: FUTBIN.

Avanzando a la zona del **centro del campo**, las características que más nos piden son las siguientes: debe tener entre 20 y 30 años, más de un 81% de pases completados con éxito, más de un 50% en acciones defensivas realizadas con acierto, más de un 29% de minutos posibles en esta temporada, más de un 4 de nota media, más de un 35% de minutos jugados entre las tres últimas campañas, un precio de mercado entre 5 y 50 millones de euros y, por último, una nota mayor que 0. A partir de estos datos, los mediocentros que nos aparecen son *Joan Jordán*, *Martín Zubimendi* e *Igor Zubeldía*. Finalmente, se ha decidido fichar al español hispalense por una cantidad de 14000000€: *Joan Jordán*. Además, vamos a activar la opción de compra de *Ilaix Moriba* por un valor de 7000000€ y devolvemos a *Nico González* a su anterior club (FC Barcelona), ya que no ha cumplido con las expectativas. Por tanto, la línea de creación de juego nos va a costar 21 millones de euros.



Figura 7: comparación entre Nico González, Ilaix Moriba y Joan Jordán. Fuente: FUTBIN.

Por último, vemos que, en la **línea más ofensiva** del equipo, nos piden que tenga entre 22 y 31 años, más de un 50% de pases completados con éxito, más de un 3% de disparos que han terminado en gol, más de un 21% de minutos posibles en esta temporada, más de un 6.5 de nota media, más de un 40% de minutos jugados entre las últimas tres campañas, un precio de mercado mayor de 20 millones de euros y, por último, una nota mayor que 0. Tras haber analizado nuestra base de datos, los atacantes que vemos disponibles son *Ousmane Dembele*, *Raphinha* y *Ferran Torres*. Hemos decidido que vamos a darle una segunda oportunidad al extremo español por un valor de traspaso de 35000000€: *Ferran Torres*. También vamos a vender a *Edinson Cavani* al Napoli por 5000000€ y ejerceremos la opción de compra de *Samuel Lino* y *Justin Kluivert* por un valor de 16000000€ y 14000000€, respectivamente. Por tanto, la línea de finalización nos costará un total de 60 millones de euros.



Figura 8: comparación entre Edinson Cavani, Ferran Torres, Justin Kluivert y Samuel Lino. Fuente: FUTBIN.

Para concluir, la revolución del equipo valencianista para que vuelva a los puestos donde se merece tendrá un coste de 98.8 millones de euros, los cuales se irán amortizando a medida que la entidad gane más títulos, se clasifique para puestos europeos y encuentre más formas de financiación, como, por ejemplo, marcas y patrocinadores.

La **plantilla final**, con el once titular y los suplentes, quedaría de la siguiente forma:

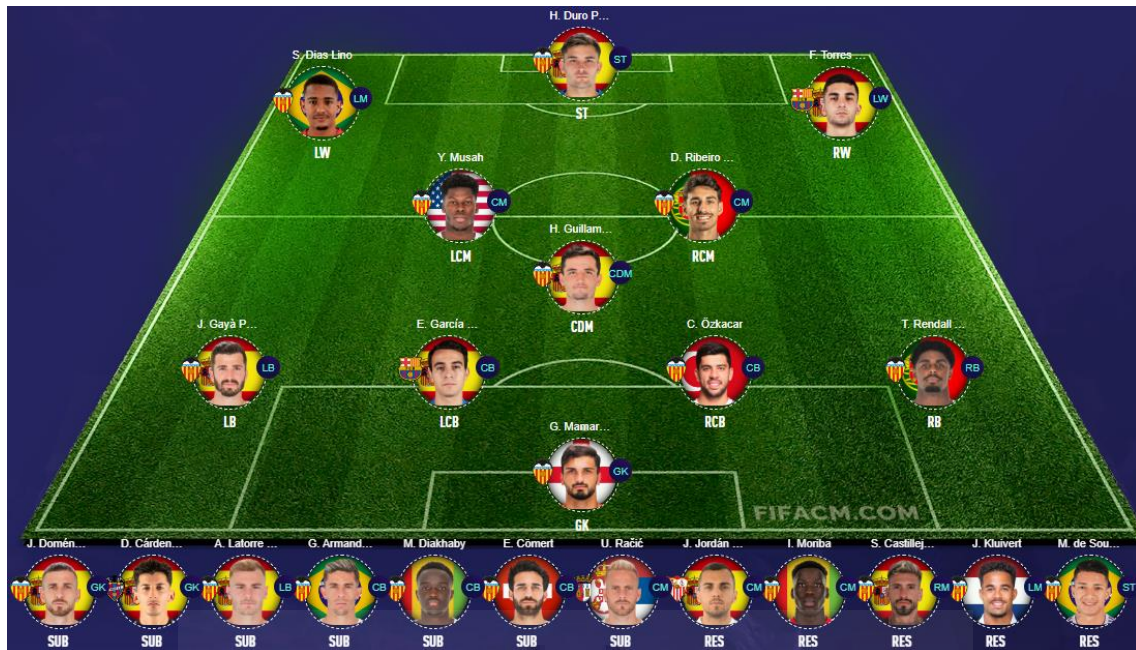


Figura 9: equipo Valencia CF temporada 2023/2024. Fuente: FIFACM.

En el caso de haber querido formar un nuevo equipo basándonos en la valoración de los jugadores, la función objetivo hubiera estado formada por los elementos de la variable *notas*.

Por tanto, a partir de nuestro modelo, podemos hacer cualquier equipo a partir de las características que nos pidan.

7. Conclusiones

La realización de este trabajo final nos ha permitido obtener mucha información tanto a nivel económico como deportivo del Valencia CF.

Por una parte, en el ámbito deportivo, hemos visto como la plantilla estaba mal estructurada porque tenía más defensas que centrocampistas y delanteros. También hemos concluido que teníamos una plantilla con varios jugadores con una edad bastante avanzada y necesitábamos darle un aire nuevo al club.

Por otra parte, a nivel económico, hemos visto como, con una inversión de 100 millones de euros, podemos mejorar bastante los números del equipo. Aunque es bastante dinero, si lo comparamos con el actual mercado de fichajes, donde jugadores de 18 años llegan a valer entre 40 y 50 millones, podríamos considerarlo un gran triunfo empresarial.

También hemos observado la gran diferencia entre tener un presidente que solamente busca el beneficio propio sin importar las consecuencias ocasionadas para el club y tener un club donde las decisiones se toman a votación entre los socios. Esto lo hemos podido concluir gracias a las ventas más significativas de jugadores como *Carlos Soler* o *Gonçalo Guedes*.

Más allá de los resultados, cabe destacar la elaboración de un modelo estadístico mediante la regresión lineal múltiple, para poder llegar a conocer todas aquellas valoraciones que tanto nos quieren ocultar las páginas especializadas en datos futbolísticos. Hemos observado que las notas que hemos obtenido en la última temporada, por lo general, solamente se han distanciado en 1 solo punto sobre la realidad. Esto nos hace intuir que el algoritmo realizado es bastante parecido al de estas aplicaciones.

Para concluir, tenemos que destacar la importancia de tener una responsabilidad tan grande como la de gestionar un club entero. Entre estas obligaciones, está la adaptación a las nuevas tecnologías y técnicas que puedan mejorar tanto la plantilla como el juego dentro del campo. Todo esto se lo debemos al desarrollo del Big Data dentro del ámbito deportivo, o más concretamente, en el mundo del fútbol. Esto podrá ayudar en nuestro caso al



club che a poder mejorar su plantilla sin tener que gastar más dinero de lo necesario y así poder evitar las deudas tanto con otros clubes como con tus propios futbolistas. Esto también ha hecho que clubes menos conocidos han ido escalando puestos y subiendo de categoría hasta llegar a poder competir en competiciones europeas, como es el caso del club inglés Brighton & Hove Albion FC.



8. Bibliografía

- Guterman, T. (s. f.). *El análisis de la eficacia técnica del jugador de fútbol en competición*. <https://www.efdeportes.com/efd82/futbol.htm>
- Transfermarkt. (s. f.). *Fichajes de fútbol, rumores, valores de mercado y noticias*. <https://www.transfermarkt.es/>
- <https://www.sofascore.com/>
- *Salary Sport | Comprehensive Salary Database for Sports Clubs & Players*. (s. f.). <https://salarysport.com/>
- *inicio | OneFootball*. (s. f.). OneFootball. <https://onefootball.com/es/inicio>
- Comunate. (2020, 7 octubre). SofaScore da detalles sobre su sistema de puntuación. *Comunate.com*. <https://www.comunate.com/noticias/6595/sofascore-da-detalles-sobre-su-sistema-de-puntuacion>
- <https://www.sofascore.com/news/sofascore-clasificacion-de-estadisticas/#top>
- Gallo, J. (2021). Tutorial: cómo analizar las notas de SofaScore. *ComunioMagazine*. <https://magazine.comunio.es/tutorial-analisis-de-las-notas-de-sofascore/#:~:text=El%20algoritmo%20de%20SofaScore%20es,su%20acierto%20o%20desacierto%20estad%C3%ADstico>.
- Holtz, Y. (s. f.). *The R Graph Gallery – Help and inspiration for R charts*. The R Graph Gallery. <https://r-graph-gallery.com/>
- BBC News Mundo. (2014, 29 agosto). ¿Quién inventó el fútbol: los ingleses o los guaraníes? *BBC News Mundo*. https://www.bbc.com/mundo/noticias/2014/08/140827_deportes_futbol_guarani_vs

- Pérez, M. (2023, 10 febrero). ¿En qué año se inventó el fútbol? *sport*.
<https://www.sport.es/es/noticias/futbol/en-que-ano-invento-se-el-futbol-82793812>
- García, A. (2023, 31 enero). El Valencia no tiene ni para pagar a los jugadores: otra vez pagarés. *Superdeporte*. <https://www.superdeporte.es/valencia-cf/2023/01/31/valencia-pagar-jugadores-vez-pagares-82271658.html>
- Goal. (2021, 8 octubre). ¿Quiénes son los dueños más ricos de los equipos de fútbol? Ranking 2021 actualizado. *Goal.com España*.
<https://www.goal.com/es/noticias/quienes-duenos-mas-ricos-equipos-futbol/kw5gne4dp4sa18h5myr09kwey>
- Collection, K. G. N. G. I. (2023). Aunque el deporte que conocemos hoy lo formalizaron los británicos, sus raíces se remontan a la China antigua o Mesoamérica. *National Geographic*.
<https://www.nationalgeographic.es/historia/donde-surgio-el-futbol-esto-dicen-los-arqueologos>
- Trujillo, I. (2021, 19 mayo). El día que nació el primer reglamento de fútbol: 14 normas, sin córners y sin árbitro. *La Razón*.
<https://www.larazon.es/deportes/20201208/j7b6z5vwfvhxldu6734brjyj6e.html>
- <https://digitalhub.fifa.com/m/938d26afa7ec425/original/fzqqbeaxkffqqfgo83k3-pdf.pdf>
- http://seguro.valenciacf.com/bd/archivos/GRUPO_ccaa_2021-2022_con_EINF.pdf
- Goal. (2021b, octubre 8). ¿Quiénes son los dueños más ricos de los equipos de fútbol? Ranking 2021 actualizado. *Goal.com España*.
<https://www.goal.com/es/noticias/quienes-duenos-mas-ricos-equipos-futbol/kw5gne4dp4sa18h5myr09kwey>

- colaboradores de Wikipedia. (2023). Historia del fútbol. *Wikipedia, la enciclopedia libre*.
https://es.wikipedia.org/wiki/Historia_del_f%C3%BAtbol#Juegos_antiguos
- colaboradores de Wikipedia. (2023a). COVID-19. *Wikipedia, la enciclopedia libre*. <https://es.wikipedia.org/wiki/COVID-19#Historia>
- **GarciaSergio**. (1657205651000). *Deportes & Big Data: los puntos que se ganan fuera de los estadios*. LinkedIn.com.
<https://www.linkedin.com/pulse/deportes-big-data-los-puntos-que-se-ganan-fuera-de/>
- BBC News Mundo. (2014a, abril 6). El uso del «Big Data»: ¿una revolución en ciernes para el fútbol? *BBC News Mundo*.
https://www.bbc.com/mundo/noticias/2014/04/140330_tecnologia_geeks_futbol_finde_aa
- Casañ, J. (2021, junio 1). *Brentford FC, el club del “Big Data” y el “Moneyball” llega a la Premier*. Golsmedia; Golsmedia Sports SL.
<https://golsmedia.com/viral/2021/06/01/brentford-fc-club-big-data-moneyball-llega-premier/>
- Contreras, E. (2020, 9 febrero). Así ficha Monchi: departamento I+D, data y algoritmo entre 18.000 futbolistas. *Marca.com*.
<https://www.marca.com/primera-plana/2020/02/08/5e3709e822601d8a5b8b458e.html>
- @SUPERMONTORO. (2014, 25 diciembre). Así llegó Peter Lim al Valencia. *Superdeporte*. <https://www.superdeporte.es/valencia-cf/2014/12/25/llego-peter-lim-valencia-53573488.html>



- colaboradores de Wikipedia. (2023a). RStudio. *Wikipedia, la enciclopedia libre*.

<https://es.wikipedia.org/wiki/RStudio>



9. Anexos

El siguiente [enlace](#) contiene la base de datos utilizada “Base de datos TFG.xlsx” que contiene toda la información de los jugadores. Para poder llevar a cabo nuestro proyecto, Rstudio nos ha permitido utilizar las siguientes librerías:

- **Readxl:** la utilizamos para leer datos de hojas de cálculo de Excel, podemos importar datos de archivos “.xls” y “.xlsx”.
- **Dplyr:** es una de las librerías más populares en R para manipular y transformar datos. Podemos realizar una variedad de tareas comunes de manipulación de datos, como filtrar, seleccionar, ordenar y agrupar datos de manera rápida y sencilla.
- **DT:** es una herramienta muy útil para la visualización interactiva de datos, proporciona varias funciones para crear y personalizar tablas, es una librería muy útil para la visualización interactiva de datos.
- **Ggplot2:** es una de las herramientas más populares para la visualización de datos, es una librería de gráficos basada en la gramática de los gráficos, lo que significa que utiliza una sintaxis coherente y lógica para crear gráficos sofisticados y personalizados, puedes crear gráficos de alta calidad de una manera fácil e intuitiva.
- **GgExtra:** es una librería de visualización de datos útil para aquellos que buscan personalizar aún más sus gráficos y añadir características avanzadas a sus visualizaciones, es una extensión de la librería “ggplot2” que hemos comentado anteriormente.
- **Corrplot:** se utiliza para visualizar matrices de correlación de datos en forma de gráficos, es útil para crear visualizaciones de matrices de correlación que muestran las relaciones entre las variables en un conjunto de datos.
- **Tidyyverse:** es un conjunto de paquetes que proporcionan herramientas para el procesamiento y análisis de datos.
- **Viridis:** proporciona paletas de colores para la visualización de datos en gráficos y mapas, estas paletas de colores se han diseñado

especialmente para ser perceptualmente uniformes y evitar la distorsión en la percepción de los datos.

- **GridExtra:** proporciona herramientas para la construcción de gráficos, proporciona varias funciones útiles para combinar gráficos y personalizar el diseño de los gráficos.
- **LpSolve:** sirve para poder resolver problemas de programación lineal, que consisten en maximizar o minimizar una función lineal sujeta a un conjunto de restricciones lineales y poder encontrar la solución óptima.

Una vez hemos cargado las librerías necesarias, vamos a mostrar el código utilizado para realizar nuestro trabajo final.



9.1. Código R Estadística Descriptiva

```
#####  
#####LIBRERIAS#####  
#####  
  
library(readxl)  
library(dplyr)  
library(ggplot2)  
library(ggExtra)  
library(corrplot)  
library(tidyverse)  
library(viridis)  
library(DT)  
library(gridExtra)  
library(lpSolve)  
  
#####CARGAMOS EL CONJUNTO DE DATOS  
#####  
  
datos<- read_excel("C:/Users/admin/Desktop/Base de datos TFG.xlsx")  
View(datos)  
  
#####EXPLORAMOS, LIMPIAMOS Y PROCESAMOS LOS  
#####DATOS#####  
  
head(datos)  
summary(datos)  
str(datos)  
nrow(datos)  
ncol(datos)
```

```
datos$precio_fichaje <- ifelse(datos$precio_fichaje == "cedido", 0.1,
datos$precio_fichaje)
head(datos)
summary(datos)
str(datos)
nrow(datos)
ncol(datos)
datos <- datos %>% mutate(precio_fichaje = as.numeric(precio_fichaje))
datos=na.omit(datos)
complete.cases(datos)
is.na(datos)
View(datos)
datosdf=as.data.frame(datos)
datatable(head(datosdf), rownames = head(LETTERS))
datatable(summary(datosdf), rownames = head(LETTERS))

#####
##### ESTADÍSTICA DESCRIPTIVA
#####
#####

#####
##### ANALISIS UNIVARIANTE #####
#####

### BARPLOT
databp <- data.frame(datos$tipo_formacion)
bptform=ggplot(databp, aes(x=as.factor(datos$tipo_formacion),
fill=as.factor(datos$tipo_formacion) )) +
  geom_bar( ) +
  scale_fill_brewer(palette = "Set1") +
  theme(legend.position="none")
bptform

tabla<-prop.table(table(datos$tipo_formacion));table

#####
##### ANALISIS BIVARIANTE #####
#####

data=as.data.frame(datos)
data

### HISTOGRAMA
p <- ggplot(data, aes(x=salario, y=posicion, color=edad, size=)) +
  geom_point() +
  theme(legend.position="none")

phist <- ggMarginal(p, type="histogram")
phist
```

```

### BOXPLOT
p <- ggplot(data, aes(x=edad, y=t_amarillas, color=posicion, size=)) +
  geom_point() +
  theme(legend.position="none")

pbp <- ggMarginal(p, type="boxplot")
pbp

### PIECHART
datosB=datos[datos$equipo == "FC Barcelona",]
datosV=datos[datos$equipo == "Valencia CF",]
datosL=datos[datos$equipo == "Levante UD",]
datosS=datos[datos$equipo == "Sevilla FC",]
datosRS=datos[datos$equipo == "Real Sociedad",]

meanB=((sum(datosB$salario))/sum(datos$salario))*100
meanV=((sum(datosV$salario))/sum(datos$salario))*100
meanL=((sum(datosL$salario))/sum(datos$salario))*100
meanS=((sum(datosS$salario))/sum(datos$salario))*100
meanRS=((sum(datosRS$salario))/sum(datos$salario))*100

grafico=c(meanB,meanV,meanL,meanS,meanRS)
grafico

df <- data.frame(grupo = c("FC Barcelona", "Valencia CF", "Levante UD", "Sevilla
FC", "Real Sociedad"), porcentaje = c(grafico[1], grafico[2], grafico[3], grafico[4],
grafico[5]))

piesalario=ggplot(df, aes(x = "", y = porcentaje, fill = grupo)) +
  geom_bar(width = 1, stat = "identity") +
  coord_polar("y", start = 0) +
  theme_void() +
  theme(legend.position = "bottom") +
  geom_text(aes(label = paste0(round(porcentaje,digits=3), "%")), position =
position_stack(vjust = 0.5), size = 4)
piesalario

### COEFICIENTE DE CORRELACION

cor(datos[,3:23])
corrplot(cor(datos[,3:23]))
correlaciones=corrplot(cor(datos[,3:23]), method = "number",
  title = "Correlaciones",
  tl.pos = "n", mar = c(2, 1, 3, 1))
correlaciones

```

```
#####
##### ANALISIS MULTIVARIANTE #####
#####

##### FC BARCELONA #####
data <- data.frame(
  individual=datos$jugadores[datos$equipo=="FC Barcelona"],
  group=datos$equipo[datos$equipo=="FC Barcelona"],
  value1=datos$min_20_21[datos$equipo=="FC Barcelona"],
  value2=datos$min_21_22[datos$equipo=="FC Barcelona"],
  value3=datos$min_22_23[datos$equipo=="FC Barcelona"]
)
data <- data %>% gather(key = "observation", value="value", -c(1,2))
empty_bar <- 2
nObsType <- nlevels(as.factor(data$observation))
to_add <- data.frame( matrix(NA, empty_bar*nlevels(data$group)*nObsType,
ncol(data)) )
colnames(to_add) <- colnames(data)
to_add$group <- rep(levels(data$group), each=empty_bar*nObsType )
data <- rbind(data, to_add)
data <- data %>% arrange(group, individual)
data$id <- rep( seq(1, nrow(data)/nObsType) , each=nObsType)
label_data <- data %>% group_by(id, individual) %>% summarize(tot=sum(value))
number_of_bar <- nrow(label_data)
angle <- 90 - 360 * (label_data$id-0.5) /number_of_bar
label_data$hjust <- ifelse( angle < -90, 1, 0)
label_data$angle <- ifelse(angle < -90, angle+180, angle)
base_data <- data %>%
  group_by(group) %>%
  summarize(start=min(id), end=max(id) - empty_bar) %>%
  rowwise() %>%
  mutate(title=mean(c(start, end)))
grid_data <- base_data
grid_data$end <- grid_data$end[ c( nrow(grid_data), 1:nrow(grid_data)-1)] + 1
grid_data$start <- grid_data$start - 1
grid_data <- grid_data[-1,]
pbarcelona <- ggplot(data) +
  geom_bar(aes(x=as.factor(id), y=value, fill=observation), stat="identity", alpha=0.5)
+
  scale_fill_viridis(discrete=TRUE) +

  ylim(-3,max(label_data$tot, na.rm=T)) +
  theme_minimal() +
  theme(
    legend.position = "none",
    axis.text = element_blank(),
    axis.title = element_blank(),
    panel.grid = element_blank(),
    plot.margin = unit(rep(-1,4), "cm")
  ) +
  coord_polar() +

  geom_text(data=label_data, aes(x=id, y=tot+0.05, label=individual, hjust=hjust),
  color="black", fontface="bold", alpha=0.6, size=3.5, angle= label_data$angle,
  inherit.aes = FALSE ) +
```

```

geom_segment(data=base_data, aes(x = start, y = -0.025, xend = end, yend = -
0.025), colour = "black", alpha=0.8, size=0.6 , inherit.aes = FALSE ) +
  geom_text(data=base_data, aes(x = title, y = -0.09, label=group), hjust=c(1),
colour = "black", alpha=0.8, size=2.5, fontface="bold", inherit.aes = FALSE)

pbarcelona

##### VALENCIA CF #####

data <- data.frame(
  individual=datos$jugadores[datos$equipo=="Valencia CF"],
  group=datos$equipo[datos$equipo=="Valencia CF"],
  value1=datos$min_20_21[datos$equipo=="Valencia CF"],
  value2=datos$min_21_22[datos$equipo=="Valencia CF"],
  value3=datos$min_22_23[datos$equipo=="Valencia CF"]
)

data <- data %>% gather(key = "observation", value="value", -c(1,2))

empty_bar <- 2
nObsType <- nlevels(as.factor(data$observation))
to_add <- data.frame( matrix(NA, empty_bar*nlevels(data$group)*nObsType,
ncol(data)) )
colnames(to_add) <- colnames(data)
to_add$group <- rep(levels(data$group), each=empty_bar*nObsType )
data <- rbind(data, to_add)
data <- data %>% arrange(group, individual)
data$id <- rep( seq(1, nrow(data)/nObsType) , each=nObsType)

label_data <- data %>% group_by(id, individual) %>% summarize(tot=sum(value))
number_of_bar <- nrow(label_data)
angle <- 90 - 360 * (label_data$id-0.5) /number_of_bar
label_data$hjust <- ifelse( angle < -90, 1, 0)
label_data$angle <- ifelse(angle < -90, angle+180, angle)

base_data <- data %>%
  group_by(group) %>%
  summarize(start=min(id), end=max(id) - empty_bar) %>%
  rowwise() %>%
  mutate(title=mean(c(start, end)))

grid_data <- base_data
grid_data$end <- grid_data$end[ c( nrow(grid_data), 1:nrow(grid_data)-1)] + 1
grid_data$start <- grid_data$start - 1
grid_data <- grid_data[-1,]

pvalencia <- ggplot(data) +

  geom_bar(aes(x=as.factor(id), y=value, fill=observation), stat="identity", alpha=0.5)
+
  scale_fill_viridis(discrete=TRUE) +

```



```
ylim(-3,max(label_data$tot, na.rm=T)) +
  theme_minimal() +
  theme(
    legend.position = "none",
    axis.text = element_blank(),
    axis.title = element_blank(),
    panel.grid = element_blank(),
    plot.margin = unit(rep(-1,4), "cm")
  ) +
  coord_polar() +

  geom_text(data=label_data, aes(x=id, y=tot+0.05, label=individual, hjust=hjust),
    color="black", fontface="bold",alpha=0.6, size=3.5, angle= label_data$angle,
    inherit.aes = FALSE ) +

  geom_segment(data=base_data, aes(x = start, y = -0.025, xend = end, yend = -
0.025), colour = "black", alpha=0.8, size=0.6 , inherit.aes = FALSE ) +
  geom_text(data=base_data, aes(x = title, y = -0.09, label=group), hjust=c(1),
    colour = "black", alpha=0.8, size=2.5, fontface="bold", inherit.aes = FALSE)

pvalencia

##### LEVANTE UD #####

data <- data.frame(
  individual=datos$jugadores[datos$equipo=="Levante UD"],
  group=datos$equipo[datos$equipo=="Levante UD"],
  value1=datos$min_20_21[datos$equipo=="Levante UD"],
  value2=datos$min_21_22[datos$equipo=="Levante UD"],
  value3=datos$min_22_23[datos$equipo=="Levante UD"]
)

data <- data %>% gather(key = "observation", value="value", -c(1,2))

empty_bar <- 2
nObsType <- nlevels(as.factor(data$observation))
to_add <- data.frame( matrix(NA, empty_bar*nlevels(data$group)*nObsType,
ncol(data)) )
colnames(to_add) <- colnames(data)
to_add$group <- rep(levels(data$group), each=empty_bar*nObsType )
data <- rbind(data, to_add)
data <- data %>% arrange(group, individual)
data$id <- rep( seq(1, nrow(data)/nObsType) , each=nObsType)

label_data <- data %>% group_by(id, individual) %>% summarize(tot=sum(value))
number_of_bar <- nrow(label_data)
angle <- 90 - 360 * (label_data$id-0.5) /number_of_bar
label_data$hjust <- ifelse( angle < -90, 1, 0)
label_data$angle <- ifelse(angle < -90, angle+180, angle)

base_data <- data %>%
  group_by(group) %>%
  summarize(start=min(id), end=max(id) - empty_bar) %>%
  rowwise() %>%
  mutate(title=mean(c(start, end)))
```

```

grid_data <- base_data
grid_data$end <- grid_data$end[ c( nrow(grid_data), 1:nrow(grid_data)-1)] + 1
grid_data$start <- grid_data$start - 1
grid_data <- grid_data[-1,]

plevante <- ggplot(data) +

  geom_bar(aes(x=as.factor(id), y=value, fill=observation), stat="identity", alpha=0.5)
+
  scale_fill_viridis(discrete=TRUE) +

ylim(-3,max(label_data$tot, na.rm=T)) +
  theme_minimal() +
  theme(
    legend.position = "none",
    axis.text = element_blank(),
    axis.title = element_blank(),
    panel.grid = element_blank(),
    plot.margin = unit(rep(-1,4), "cm")
  ) +
  coord_polar() +

  geom_text(data=label_data, aes(x=id, y=tot+0.05, label=individual, hjust=hjust),
  color="black", fontface="bold", alpha=0.6, size=3.5, angle= label_data$angle,
  inherit.aes = FALSE ) +

  geom_segment(data=base_data, aes(x = start, y = -0.025, xend = end, yend = -
0.025), colour = "black", alpha=0.8, size=0.6 , inherit.aes = FALSE ) +
  geom_text(data=base_data, aes(x = title, y = -0.09, label=group), hjust=c(1),
  colour = "black", alpha=0.8, size=2.5, fontface="bold", inherit.aes = FALSE)

plevante

##### SEVILLA FC #####

data <- data.frame(
  individual=datos$jugadores[datos$equipo=="Sevilla FC"],
  group=datos$equipo[datos$equipo=="Sevilla FC"],
  value1=datos$min_20_21[datos$equipo=="Sevilla FC"],
  value2=datos$min_21_22[datos$equipo=="Sevilla FC"],
  value3=datos$min_22_23[datos$equipo=="Sevilla FC"]
)

data <- data %>% gather(key = "observation", value="value", -c(1,2))

empty_bar <- 2
nObsType <- nlevels(as.factor(data$observation))
to_add <- data.frame( matrix(NA, empty_bar*nlevels(data$group)*nObsType,
ncol(data)) )
colnames(to_add) <- colnames(data)
to_add$group <- rep(levels(data$group), each=empty_bar*nObsType )
data <- rbind(data, to_add)
data <- data %>% arrange(group, individual)
data$id <- rep( seq(1, nrow(data)/nObsType) , each=nObsType)

```

```

label_data <- data %>% group_by(id, individual) %>% summarize(tot=sum(value))
number_of_bar <- nrow(label_data)
angle <- 90 - 360 * (label_data$id-0.5) /number_of_bar
label_data$hjust <- ifelse( angle < -90, 1, 0)
label_data$angle <- ifelse(angle < -90, angle+180, angle)

base_data <- data %>%
  group_by(group) %>%
  summarize(start=min(id), end=max(id) - empty_bar) %>%
  rowwise() %>%
  mutate(title=mean(c(start, end)))

grid_data <- base_data
grid_data$end <- grid_data$end[ c( nrow(grid_data), 1:nrow(grid_data)-1)] + 1
grid_data$start <- grid_data$start - 1
grid_data <- grid_data[-1,]

psevilla <- ggplot(data) +

  geom_bar(aes(x=as.factor(id), y=value, fill=observation), stat="identity", alpha=0.5)
+
  scale_fill_viridis(discrete=TRUE) +

  ylim(-3,max(label_data$tot, na.rm=T)) +
  theme_minimal() +
  theme(
    legend.position = "none",
    axis.text = element_blank(),
    axis.title = element_blank(),
    panel.grid = element_blank(),
    plot.margin = unit(rep(-1,4), "cm")
  ) +
  coord_polar() +

  geom_text(data=label_data, aes(x=id, y=tot+0.05, label=individual, hjust=hjust),
  color="black", fontface="bold", alpha=0.6, size=3.5, angle= label_data$angle,
  inherit.aes = FALSE ) +

  geom_segment(data=base_data, aes(x = start, y = -0.025, xend = end, yend = -
0.025), colour = "black", alpha=0.8, size=0.6 , inherit.aes = FALSE ) +
  geom_text(data=base_data, aes(x = title, y = -0.09, label=group), hjust=c(1),
  colour = "black", alpha=0.8, size=2.5, fontface="bold", inherit.aes = FALSE)

psevilla

##### REAL SOCIEDAD #####

data <- data.frame(
  individual=datos$jugadores[datos$equipo=="Real Sociedad"],
  group=datos$equipo[datos$equipo=="Real Sociedad"],
  value1=datos$min_20_21[datos$equipo=="Real Sociedad"],
  value2=datos$min_21_22[datos$equipo=="Real Sociedad"],
  value3=datos$min_22_23[datos$equipo=="Real Sociedad"]
)

```

```

data <- data %>% gather(key = "observation", value="value", -c(1,2))

empty_bar <- 2
nObsType <- nlevels(as.factor(data$observation))
to_add <- data.frame( matrix(NA, empty_bar*nlevels(data$group)*nObsType,
ncol(data)) )
colnames(to_add) <- colnames(data)
to_add$group <- rep(levels(data$group), each=empty_bar*nObsType )
data <- rbind(data, to_add)
data <- data %>% arrange(group, individual)
data$id <- rep( seq(1, nrow(data)/nObsType) , each=nObsType)

label_data <- data %>% group_by(id, individual) %>% summarize(tot=sum(value))
number_of_bar <- nrow(label_data)
angle <- 90 - 360 * (label_data$id-0.5) /number_of_bar
label_data$hjust <- ifelse( angle < -90, 1, 0)
label_data$angle <- ifelse(angle < -90, angle+180, angle)

base_data <- data %>%
  group_by(group) %>%
  summarize(start=min(id), end=max(id) - empty_bar) %>%
  rowwise() %>%
  mutate(title=mean(c(start, end)))

grid_data <- base_data
grid_data$end <- grid_data$end[ c( nrow(grid_data), 1:nrow(grid_data)-1)] + 1
grid_data$start <- grid_data$start - 1
grid_data <- grid_data[-1,]

prs <- ggplot(data) +

  geom_bar(aes(x=as.factor(id), y=value, fill=observation), stat="identity", alpha=0.5)
+
  scale_fill_viridis(discrete=TRUE) +

ylim(-3,max(label_data$tot, na.rm=T)) +
theme_minimal() +
theme(
  legend.position = "none",
  axis.text = element_blank(),
  axis.title = element_blank(),
  panel.grid = element_blank(),
  plot.margin = unit(rep(-1,4), "cm")
) +
coord_polar() +

  geom_text(data=label_data, aes(x=id, y=tot+0.05, label=individual, hjust=hjust),
color="black", fontface="bold",alpha=0.6, size=3.5, angle= label_data$angle,
inherit.aes = FALSE ) +

  geom_segment(data=base_data, aes(x = start, y = -0.025, xend = end, yend = -
0.025), colour = "black", alpha=0.8, size=0.6 , inherit.aes = FALSE ) +
  geom_text(data=base_data, aes(x = title, y = -0.09, label=group), hjust=c(1),
colour = "black", alpha=0.8, size=2.5, fontface="bold", inherit.aes = FALSE)
prs

```

```
ggsave(pbarcelona, file="barcelona.png", width=10, height=10)  
ggsave(plevante, file="levante.png", width=10, height=10)  
ggsave(pvalencia, file="valencia.png", width=10, height=10)  
ggsave(psevilla, file="sevilla.png", width=10, height=10)  
ggsave(prs, file="real sociedad.png", width=10, height=10)
```

```
datos %>%  
  select(equipo, precio_actual, precio_fichaje, salario, edad) %>%  
  group_by(equipo) %>%  
  summarise_all(.funs = "mean")
```



9.2. Código R Obtención Notas

```
#####
#####
##### SACAMOS LAS NOTAS JUGADORES
#####
#####

### PORTEROS

datos$notas <- ifelse(datos$posicion == 1,
((datos$pases*datos$min_22_23*0.5)+(datos$paradas*datos$min_22_23*1)+((dato
s$nota_20_21*datos$min_20_21)/10)+((datos$nota_21_22*datos$min_21_22)/10)+
((datos$nota_22_23*datos$min_22_23)/10)+datos$selección+((datos$min_seleccio
n)/max(datos$min_seleccion[datos$posicion==1]))-
((datos$t_amarillas)/max(datos$t_amarillas[datos$posicion==1]))-
((datos$t_rojas)/max(datos$t_rojas[datos$posicion==1]))+((datos$precio_actual)/m
ax(datos$precio_actual[datos$posicion==1]))-datos$lesiones)*2,

### DEFENSAS

      ifelse(datos$posicion == 2,
((datos$goles*datos$min_22_23*0.33)+(datos$pases*datos$min_22_23*0.67)+(dat
os$defensa*datos$min_22_23*1)+((datos$nota_20_21*datos$min_20_21)/10)+((dat
os$nota_21_22*datos$min_21_22)/10)+((datos$nota_22_23*datos$min_22_23)/10
)+datos$selección+((datos$min_seleccion)/max(datos$min_seleccion[datos$posicio
n==2]))-((datos$t_amarillas)/max(datos$t_amarillas[datos$posicion==2]))-
((datos$t_rojas)/max(datos$t_rojas[datos$posicion==2]))+((datos$precio_actual)/m
ax(datos$precio_actual[datos$posicion==2]))-datos$lesiones)*2,

### CENTROCAMPISTAS

      ifelse(datos$posicion == 3,
((datos$goles*datos$min_22_23*0.67)+(datos$pases*datos$min_22_23*1)+(datos$
defensa*datos$min_22_23*0.67)+((datos$nota_20_21*datos$min_20_21)/10)+((dat
os$nota_21_22*datos$min_21_22)/10)+((datos$nota_22_23*datos$min_22_23)/10)
+datos$selección+((datos$min_seleccion)/max(datos$min_seleccion[datos$posicio
n==3]))-((datos$t_amarillas)/max(datos$t_amarillas[datos$posicion==3]))-
((datos$t_rojas)/max(datos$t_rojas[datos$posicion==3]))+((datos$precio_actual)/m
ax(datos$precio_actual[datos$posicion==3]))-datos$lesiones)*2,

### DELANTEROS

((datos$goles*datos$min_22_23*1)+(datos$pases*datos$min_22_23*0.67)+(datos$
defensa*datos$min_22_23*0.33)+((datos$nota_20_21*datos$min_20_21)/10)+((dat
os$nota_21_22*datos$min_21_22)/10)+((datos$nota_22_23*datos$min_22_23)/10)
+datos$selección+((datos$min_seleccion)/max(datos$min_seleccion[datos$posicio
n==4]))-((datos$t_amarillas)/max(datos$t_amarillas[datos$posicion==4]))-
((datos$t_rojas)/max(datos$t_rojas[datos$posicion==4]))+((datos$precio_actual)/m
ax(datos$precio_actual[datos$posicion==4]))-datos$lesiones)*2))

datos$notas=ifelse(datos$notas<0,0,datos$notas)
View(datos)

datajn <- data.frame(datos$jugadores, datos$notas)
View(datajn)
```

9.3. Código R Obtención Notas Regresión

```
#####  
#####  
##### NOTAS JUGADORES ALGORITMO REGRESION  
#####  
#####  
#####  
  
# TEMPORADA 20/21  
  
modelo_20_21 <- lm(nota_20_21 ~  
posicion+edad+tipo_formacion+t_amarillas+t_rojas+  
goles+pases+defensa+paradas+min_20_21+  
selección+min_seleccion+precio_actual+  
precio_fichaje+lesiones+salario, data = datos)  
summary(modelo_20_21)  
  
step(object = modelo_20_21, direction = "both", trace = 1)  
  
modelo_20_21 <- (lm(formula = nota_20_21 ~ edad + goles + pases + min_20_21,  
data = datos))  
summary(modelo_20_21)  
  
datos$pred_20_21=-  
1.3211+0.1044*datos$edad+3.3619*datos$goles+3.2308*datos$pases+  
3.5454*datos$min_20_21  
  
dif2021=abs(datos$pred_20_21-datos$nota_20_21)  
  
pred2021 <- data.frame(datos$jugadores,datos$equipo,datos$nota_20_21,  
datos$pred_20_21, dif2021)  
  
pred2021$datos.pred_20_21=ifelse(pred2021$datos.nota_20_21==0,0,pred2021$d  
atos.pred_20_21)  
  
pred2021$dif2021=abs(pred2021$datos.nota_20_21-pred2021$datos.pred_20_21)  
View(pred2021)  
  
confint(lm(formula = nota_20_21 ~ edad + goles + pases + min_20_21, data =  
datos))  
  
data2021=data.frame(pred2021$dif2021)  
  
h2021=ggplot(data2021, aes(x=pred2021$dif2021)) +  
geom_histogram(binwidth=0.5) +  
scale_x_continuous(breaks=seq(0, 6, by=0.5)) +  
labs(x = "", y = "", title = "")  
h2021  
x_cut <- cut(data2021$pred2021.dif2021, breaks = seq(0, 6, by = 0.5), right =  
FALSE)  
  
x_count <- table(x_cut)  
  
x_prop <- prop.table(x_count) * 100
```



```

tabla2021 <- data.frame(Intervalo=as.character(x_count),
Frecuencia=as.numeric(x_count), Porcentaje=x_prop)
tabla2021 <- subset(tabla2021, Frecuencia > 0)
tabla2021

grafico2021=c(tabla2021$Porcentaje.Freq)
grafico2021

df2021 <- data.frame(grupo = tabla2021$Porcentaje.x_cut, porcentaje =
tabla2021$Porcentaje.Freq)

pie2021=ggplot(df2021, aes(x = "", y = porcentaje, fill = grupo)) +
  geom_bar(width = 1, stat = "identity") +
  coord_polar("y", start = 0) +
  theme_void() +
  theme(legend.position = "bottom") +
  geom_text(aes(label = paste0(round(porcentaje,digits=3), "%")), position =
position_stack(vjust = 0.5), size = 4)
pie2021

# TEMPORADA 21/22

modelo_21_22 <- lm(nota_21_22 ~
posicion+edad+tipo_formacion+t_amarillas+t_rojas+
goles+pases+defensa+paradas+min_20_21+min_21_22+nota_20_21+
selección+min_seleccion+precio_actual+
precio_fichaje+lesiones+salario, data = datos)
summary(modelo_21_22)

step(object = modelo_21_22, direction = "both", trace = 1)

modelo_21_22 <- (lm(formula = nota_21_22 ~ edad + tipo_formacion + t_amarillas
+
      pases + min_20_21 + min_21_22 + nota_20_21 + precio_actual +
      salario, data = datos))
summary(modelo_21_22)

datos$pred_21_22=-0.80354+0.13014*datos$edad+(-
0.11744)*datos$tipo_formacion+
(-0.01095)*datos$t_amarillas+2.02501*datos$pases+
(-1.67718)*datos$min_20_21+2.16870*datos$min_21_22+
0.42890*datos$nota_20_21+0.02402*datos$precio_actual+
(-0.09108)*datos$salario

dif2122=abs(datos$pred_21_22-datos$nota_21_22)

pred2122 <- data.frame(datos$jugadores,datos$equipo,datos$nota_21_22,
datos$pred_21_22, dif2122)

pred2122$datos.pred_21_22=ifelse(pred2122$datos.nota_21_22==0,0,pred2122$d
atos.pred_21_22)

pred2122$dif2122=abs(pred2122$datos.nota_21_22-pred2122$datos.pred_21_22)
View(pred2122)

```

```

confint(lm(formula = nota_21_22 ~ edad + tipo_formacion + t_amarillas +
           pases + min_20_21 + min_21_22 + nota_20_21 + precio_actual +
           salario, data = datos))

data2122=data.frame(pred2122$dif2122)

h2122=ggplot(data2122, aes(x=pred2122$dif2122)) +
  geom_histogram(binwidth=0.5) +
  scale_x_continuous(breaks=seq(0, 4, by=0.5)) +
  labs(x = "", y = "", title = "")
h2122

x_cut <- cut(data2122$pred2122.dif2122, breaks = seq(0, 4, by = 0.5), right =
FALSE)
x_count <- table(x_cut)

x_prop <- prop.table(x_count) * 100

tabla2122 <- data.frame(Intervalo=as.character(x_count),
Frecuencia=as.numeric(x_count), Porcentaje=x_prop)
tabla2122 <- subset(tabla2122, Frecuencia > 0)
tabla2122

grafico2122=c(tabla2122$Porcentaje.Freq)
grafico2122

df2122 <- data.frame(grupo = tabla2122$Porcentaje.x_cut, porcentaje =
tabla2122$Porcentaje.Freq)

pie2122=ggplot(df2122, aes(x = "", y = porcentaje, fill = grupo)) +
  geom_bar(width = 1, stat = "identity") +
  coord_polar("y", start = 0) +
  theme_void() +
  theme(legend.position = "bottom") +
  geom_text(aes(label = paste0(round(porcentaje,digits=3), "%")), position =
position_stack(vjust = 0.5), size = 4)
pie2122

# TEMPORADA 22/23 hasta jornada 22

modelo_22_23 <- lm(nota_22_23 ~
posicion+edad+tipo_formacion+t_amarillas+t_rojas+
goles+pases+defensa+paradas+min_20_21+min_21_22+min_22_23+
nota_20_21+nota_21_22+
selección+min_seleccion+precio_actual+
precio_fichaje+lesiones+salario, data = datos)
summary(modelo_22_23)

step(object = modelo_22_23, direction = "both", trace = 1)

modelo_22_23 <- (lm(formula = nota_22_23 ~ posicion + tipo_formacion + pases +
paradas + min_22_23 + nota_20_21 + salario, data = datos))
summary(modelo_22_23)

```

```

datos$pred_22_23=-0.66872+0.92729*datos$posicion+(-
0.29602)*datos$tipo_formacion+
(6.07088)*datos$pases+1.89286*datos$paradas+
(0.63320)*datos$min_22_23+0.09103*datos$nota_20_21+
(-0.04016)*datos$salario

dif2223=abs(datos$pred_22_23-datos$nota_22_23)

pred2223 <- data.frame(datos$jugadores,datos$equipo,datos$nota_22_23,
datos$pred_22_23, dif2223)

pred2223$datos.pred_22_23=ifelse(pred2223$datos.nota_22_23==0,0,pred2223$d
atos.pred_22_23)

pred2223$dif2223=abs(pred2223$datos.nota_22_23-pred2223$datos.pred_22_23)
View(pred2223)

confint(lm(formula = nota_22_23 ~ posicion + tipo_formacion + pases +
paradas + min_22_23 + nota_20_21 + salario, data = datos))

data2223=data.frame(pred2223$dif2223)

h2223=ggplot(data2223, aes(x=pred2223$dif2223)) +
geom_histogram(binwidth=0.5) +
scale_x_continuous(breaks=seq(0, 5, by=0.5)) +
labs(x = "", y = "", title = "")
h2223

x_cut <- cut(data2223$pred2223.dif2223, breaks = seq(0, 5, by = 0.5), right =
FALSE)

x_count <- table(x_cut)

x_prop <- prop.table(x_count) * 100

tabla2223 <- data.frame(Intervalo=as.character(x_count),
Frecuencia=as.numeric(x_count), Porcentaje=x_prop)
tabla2223 <- subset(tabla2223, Frecuencia > 0)
tabla2223

grafico2223=c(tabla2223$Porcentaje.Freq)
grafico2223

df2223 <- data.frame(grupo = tabla2223$Porcentaje.x_cut, porcentaje =
tabla2223$Porcentaje.Freq)

pie2223=ggplot(df2223, aes(x = "", y = porcentaje, fill = grupo)) +
geom_bar(width = 1, stat = "identity") +
coord_polar("y", start = 0) +
theme_void() +
theme(legend.position = "bottom") +
geom_text(aes(label = paste0(round(porcentaje,digits=3), "%")), position =
position_stack(vjust = 0.5), size = 4)
pie2223

```

```
grid.arrange(h2021, h2122, h2223, ncol=3,  
             top="Histogramas de tres años",  
             left="Frecuencia")
```

```
ggplot(data2021, aes(x=pred2021$dif2021)) +  
  geom_histogram(aes(fill="Temporada 20/21"), binwidth=0.5, alpha=0.5) +  
  geom_histogram(data=data2122, aes(x=pred2122$dif2122, fill="Temporada  
21/22"), binwidth=0.5, alpha=0.5) +  
  geom_histogram(data=data2223, aes(x=pred2223$dif2223, fill="Temporada  
22/23"), binwidth=0.5, alpha=0.5) +  
  scale_fill_manual(name="Temporadas", values=c("Temporada 20/21"="red",  
"Temporada 21/22"="blue", "Temporada 22/23"="gold")) +  
  labs(x="Error", y="Frecuencia", title="")
```



9.4. Código R Modelo

```
fObj=c(30,4,2,60,30,9,5,6,60,
      18,15,5,100,35,50,90,3,60,
      45,50,35,50,0.5,1.5,25,14,2.5,
      4,25,8,40,1.8,3,4,25,7,
      8,14,14,5,14,16,7,4,1.5,
      0.8,0.15,1,1.2,1.5,4,0.5,0.7,2,
      0.1,3,2,1,8,0.5,0.6,4,4,
      1,0.1,0.8,5,15,0.1,12,14,5,
      9,3,14,10,12,0.15,18,4,7,
      14,4,14,10,4,9,4,14,12,
      15,12,0.4,25,0.3,3,14,6,3,
      10,6,35,0.5,40,2,15,50,12,
      5,4,30,6,4,60,15,10,18)

m<-length(datos$jugadores)
bin<-seq(1,m)

####   PORTEROS   #####

#P=datos$jugadores[datos$posicion==1]
#LT28y=datos$jugadores[datos$edad<28]
#MT0n=datos$jugadores[datos$notas>0]

datosP <- data.frame(
  jugadores = datos$jugadores,
  posicion = datos$posicion,
  edad = datos$edad,
  notas = datos$notas,
  equipoV=datos$equipo
)
xP <- ifelse(datosP$equipoV!='Valencia CF' & datosP$posicion == 1 &
datosP$edad < 28 & datosP$notas > 0, 1, 0)
print(xP)
which(xP==1)

#x2+x3+x45>=1

# Construimos la matriz de restricciones
A<-matrix(0,nrow=1, ncol=m)
A[1, ]<-xP
A

# Vector de terminos independientes
b<-1
b
signos<-c('>=')
signos

solucion<-lp('min', fObj, A, signos, b, binary.vec=bin)
x<-solucion$solution
x
v<-solucion$objval
v
```

```
##### DEFENSAS #####

#D=datos$jugadores[datos$posicion==2]
#LT32y=datos$jugadores[datos$edad<32]
#MT71p=datos$jugadores[datos$pases>0.71]
#MT46d=datos$jugadores[datos$defensa>0.46]
#MT25min23=datos$jugadores[datos$min_22_23>0.25]
#MT5nota=datos$jugadores[((datos$nota_20_21+datos$nota_21_22+datos$nota_22_23)/3)>5]
#MT0min=datos$jugadores[((datos$min_20_21+datos$min_21_22+datos$min_22_23)/3)>0]
#MT5pa=datos$jugadores[datos$precio_actual>5]
#LT40pa=datos$jugadores[datos$precio_actual<40]
#MT4n=datos$jugadores[datos$notas>4]

datosD <- data.frame(
  jugadores = datos$jugadores,
  posicion = datos$posicion,
  edad = datos$edad,
  pases = datos$pases,
  defensa = datos$defensa,
  min23 = datos$min_22_23,
  medianota = (datos$nota_20_21+datos$nota_21_22+datos$nota_22_23)/3,
  pa = datos$precio_actual,
  notas = datos$notas,
  equipoV = datos$equipo
)

xD <- ifelse(datosD$equipoV!='Valencia CF' & datosD$posicion == 2 &
datosD$edad < 32 & datosD$pases>0.71 &
datosD$defensa>0.46 & datosD$min23>0.25 & datosD$medianota>5 &
datosD$pa>5 & datosD$pa<40 & datosD$notas > 4, 1, 0)

print(xD)
which(xD==1)

#x5+x10+x102>=1

# Construimos la matriz de restricciones
A<-matrix(0,nrow=1, ncol=m)
A[1, ]<-xD
A

# Vector de terminos independientes
b<-1
b
signos<-c('>=')
signos

solucion<-lp('min', fObj, A, signos, b, binary.vec=bin)
x<-solucion$solution
x
v<-solucion$objval
v
```

```
##### MEDIOCENTROS #####

#C=datos$jugadores[datos$posicion==3]
#LT30y=datos$jugadores[datos$edad<30]
#MT20y=datos$jugadores[datos$edad>20]
#MT81p=datos$jugadores[datos$pases>0.81]
#MT50d=datos$jugadores[datos$defensa>0.5]
#MT29min23=datos$jugadores[datos$min_22_23>0.29]
#MT4nota=datos$jugadores[((datos$nota_20_21+datos$nota_21_22+datos$nota_22_23)/3)>4]
#MT35min=datos$jugadores[((datos$min_20_21+datos$min_21_22+datos$min_22_23)/3)>0.35]
#MT5pa=datos$jugadores[datos$precio_actual>5]
#LT50pa=datos$jugadores[datos$precio_actual<50]
#MT0n=datos$jugadores[datos$notas>0]

datosC <- data.frame(
  jugadores = datos$jugadores,
  posicion = datos$posicion,
  edad = datos$edad,
  pases = datos$pases,
  defensa = datos$defensa,
  min23 = datos$min_22_23,
  medianota = (datos$nota_20_21+datos$nota_21_22+datos$nota_22_23)/3,
  mediamin = (datos$min_20_21+datos$min_21_22+datos$min_22_23)/3,
  pa = datos$precio_actual,
  notas = datos$notas,
  equipoV = datos$equipo
)

xC <- ifelse(datos$equipo!='Valencia CF' & datosC$posicion == 3 & datosC$edad <
30 & datosC$edad>20 & datosC$pases>0.81 &
  datosC$defensa>0.5 & datosC$min23>0.29 & datosC$medianota>4 &
  datosC$mediamin>0.35 & datosC$pa>5 & datosC$pa<50 & datosC$notas
> 0, 1, 0)
print(xC)
which(xC==1)
#x82+x104+x106>=1
# Construimos la matriz de restricciones

A<-matrix(0,nrow=1, ncol=m)
A[1, ]<-xC
A

# Vector de terminos independientes
b<-1
b
signos<-c('>=')
signos

solucion<-lp('min', fObj, A, signos, b, binary.vec=bin)
x<-solucion$solution
x
v<-solucion$objval
v
```



```
##### ATACANTES #####

#A=datos$jugadores[datos$posicion==4]
#LT31y=datos$jugadores[datos$edad<31]
#MT22y=datos$jugadores[datos$edad>22]
#MT50p=datos$jugadores[datos$pases>0.5]
#MT3g=datos$jugadores[datos$goles>0.03]
#MT21min23=datos$jugadores[datos$min_22_23>0.21]
#MT65nota=datos$jugadores[((datos$nota_20_21+datos$nota_21_22+datos$nota_
22_23)/3)>6.5]
#MT40min=datos$jugadores[((datos$min_20_21+datos$min_21_22+datos$min_22
_23)/3)>0.4]
#MT20pa=datos$jugadores[datos$precio_actual>20]
#MT0n=datos$jugadores[datos$notas>0]

datosA <- data.frame(
  jugadores = datos$jugadores,
  posicion = datos$posicion,
  edad = datos$edad,
  pases = datos$pases,
  goles = datos$goles,
  min23 = datos$min_22_23,
  medianota = (datos$nota_20_21+datos$nota_21_22+datos$nota_22_23)/3,
  mediamin = (datos$min_20_21+datos$min_21_22+datos$min_22_23)/3,
  pa = datos$precio_actual,
  notas = datos$notas,
  equipoV = datos$equipo
)
xA <- ifelse(datosA$equipoV!='Valencia CF' & datosA$posicion == 4 &
datosA$edad < 31 & datosA$edad>22 & datosA$pases>0.5 &
  datosA$goles>0.03 & datosA$min23>0.21 & datosA$medianota>6.5 &
  datosA$mediamin>0.4 & datosA$pa>20 & datosA$notas > 0, 1, 0)
print(xA)
which(xA==1)

#x18+x21+x22>=1

# Construimos la matriz de restricciones
A<-matrix(0,nrow=1, ncol=m)
A[1, ]<-xA
A

# Vector de terminos independientes
b<-1
b
signos<-c('>=')
signos

solucion<-lp('min', fObj, A, signos, b, binary.vec=bin)
x<-solucion$solution
x
v<-solucion$objval
v
```