

EDA

LOS MEJORES JUGADORES DE
LA LIGA 2020/2021
A DÍA 31/12/2020

ALEJANDRO CÁRABE ARRANZ



LaLiga



ÍNDICE

1. *Introducción.* ---- p.2
2. *Hipótesis.* ----- p.2
3. *Variables.* ----- p.3
4. *Conclusiones* --- p.5
5. *Recursos utilizados* --- p.8

1. Introducción

He realizado el trabajo sobre este año debido a que, a pesar de ser una liga atípica marcada por el coronavirus, el Atlético de Madrid consiguió salir campeón realizando una primera vuelta casi perfecta.

Para entender las diferentes hipótesis que se establecerán en este trabajo, es de vital importancia conocer cómo iba la clasificación de este torneo a día 31 de diciembre de 2020.

Para confeccionar la plantilla, se va a establecer una alineación con formación 4-4-2.

Clasificación

Pos	Equipo	PJ	PG	PE	PP	DG	PT
1	Atlético de Madrid	14	11	2	1	-22	35
2	Real Madrid	16	10	3	3	-13	33
3	Real Sociedad	17	8	5	4	-14	29
4	Sevilla FC	14	8	2	4	-7	26
5	Villarreal	16	6	8	2	-4	26
6	Barcelona	15	7	4	4	-14	25
7	Granada CF	15	7	3	5	-4	24
8	Celta de Vigo	16	6	5	5	0	23
9	Cádiz	16	5	4	7	-9	19
10	Real Betis	16	6	1	9	-11	19
11	Athletic Club	16	5	3	8	-1	18
12	Levante	15	4	6	5	-1	18
13	Alavés	16	4	6	6	-4	18
14	Getafe	15	4	5	6	-4	17
15	Elche	14	3	7	4	-4	16
16	Eibar	16	3	7	6	-4	16
17	Valencia CF	16	3	6	7	-2	15
18	Real Valladolid	16	3	6	7	-9	15
19	Osasuna	15	3	4	8	-10	13
20	Huesca	16	1	9	6	-11	12

2. Hipótesis

La principal pregunta que nos realizamos para la elaboración de este trabajo es la siguiente: Cuáles fueron los mejores futbolistas, por posición, en la primera mitad de La Liga 2020.

A partir de aquí se establecen distintas hipótesis a las cuales se tratará de dar respuesta mediante un estudio estadístico profundo.

Hipótesis principal: Los mejores jugadores pertenecen a alguno de los 4 equipos que encabezaban la tabla a día 31/12/2020.

Hipótesis secundarias:

- El número de pases acertados por un centrocampista es directamente proporcional al número de regates realizados por el mismo.
- El número de pases acertados por los centrocampistas de un equipo es directamente proporcional a la posición en la tabla del equipo.
- El número de despejes realizados por los defensas de un equipo es directamente proporcional a los goles encajados.

- Pregunta: ¿Qué defensas sobresalen a nivel estadístico en alguna de las variables de estudio?

3. Variables

Se van a estudiar las estadísticas de los jugadores en sus distintas posiciones, estableciendo para cada posición las más relevantes.

Para los **porteros** se ha tenido en cuenta:

- Únicamente los guardametas que superen los 900 minutos jugados de los 1440 estudiados.
- Por otra parte, como la base de datos viene segmentada por partido, solo se han tenido en cuenta los partidos donde el jugador haya estado presente un tiempo igual o superior a 45 minutos.
 1. El número de goles recibidos medio por partido, con valor negativo y ponderado entre el total de los valores.
 2. El porcentaje de paradas medio por partido entre los tiros a puerta recibidos, ponderado entre el total de los valores, con valor positivo. Este valor se ha extraído debido a que no todos los tiros a puerta son goles y/o paradas, así como no todos los goles son debidos a tiros a puerta.
 3. El número de paradas medio por partido, con valor positivo y ponderado entre el total de los valores.

Para los **defensas** se ha tenido en cuenta:

- Como la base de datos viene segmentada por partido, solo se han tenido en cuenta los partidos donde el jugador haya estado presente un tiempo igual o superior a 45 minutos.
- Se han considerado únicamente los defensas que superen los 600 minutos jugados de los 1440 estudiados.
 1. El número de recuperaciones de balón medio por partido, con valor positivo y ponderado entre el total de los valores.
 2. El número de intercepciones de balón medio por partido, con valor positivo y ponderado entre el total de los valores.
 3. El número de disparos bloqueados medio por partido, con valor positivo y ponderado entre el total de los valores.
 4. El número de despejes medio por partido, con valor positivo y ponderado entre el total de los valores.
 5. El número de goles recibidos medio por partido, con valor negativo y ponderado entre el total de los valores.

Para los **laterales** se ha tenido en cuenta:

- Como la base de datos viene segmentada por partido, solo se han tenido en cuenta los partidos donde el jugador haya estado presente un tiempo igual o superior a 45 minutos. Por otra parte, como se ha estudiado el porcentaje de pases medios

acertados por partido, así como el porcentaje de regates acertados, solo se consideran los jugadores que han entrado en contacto con el balón 400 veces en el total de minutos estudiado.

- Se han considerado únicamente los defensas que superen los 720 minutos jugados de los 1440 estudiados, debido a que suelen ser más volátiles debido a la distancia recorrida por partido.
 1. El número de recuperaciones de balón medio por partido, con valor positivo y ponderado entre el total de los valores.
 2. El porcentaje de pases acertados medio por partido, con valor positivo y ponderado entre el total de los valores.
 3. El número total de asistencias, que obtiene un valor de 0.01 por asistencia creada.
 4. El porcentaje de regates acertados medio por partido
 5. El número de goles recibidos medio por partido, con valor negativo y ponderado entre el total de los valores.

Para los **centrocampistas** se ha tenido en cuenta:

- Como la base de datos viene segmentada por partido, solo se han tenido en cuenta los partidos donde el jugador haya estado presente un tiempo igual o superior a 45 minutos. Por otra parte, como se ha estudiado el porcentaje de pases medios acertados por partido, así como el porcentaje de regates acertados, solo se consideran los jugadores que han entrado en contacto con el balón 400 veces en el total de minutos estudiado.
- Se han considerado únicamente los centrocampistas que superen los 520 minutos jugados de los 1440 estudiados, debido a que suelen ser más volátiles debido a la distancia recorrida por partido.
 1. El número de recuperaciones de balón medio por partido, con valor positivo y ponderado entre el total de los valores.
 2. El porcentaje de pases acertados medio por partido, con valor positivo y ponderado entre el total de los valores.
 3. El número total de asistencias, que obtiene un valor de 0.01 por asistencia creada.
 4. El porcentaje de regates acertados medio por partido
 5. El número de goles, que obtiene un valor de 0.01 por gol.

Para las **bandas** se ha tenido en cuenta:

- Como la base de datos viene segmentada por partido, solo se han tenido en cuenta los partidos donde el jugador haya estado presente un tiempo igual o superior a 45 minutos. Por otra parte, como se ha estudiado el porcentaje de pases medios acertados por partido, así como el porcentaje de regates acertados, solo se consideran los jugadores que han entrado en contacto con el balón 350 veces en el total de minutos estudiado.

- Se han considerado únicamente los centrocampistas que superen los 500 minutos jugados de los 1440 estudiados, debido a que suelen ser más volátiles y propensos a rotar.
 1. El número de recuperaciones de balón medio por partido, con valor positivo y ponderado entre el total de los valores.
 2. El porcentaje de pases acertados medio por partido, con valor positivo y ponderado entre el total de los valores.
 3. El número total de asistencias, que obtiene un valor de 0.01 por asistencia creada.
 4. El porcentaje de regates acertados medio por partido
 5. El número de goles, que obtiene un valor de 0.01 por gol.

Para los **delanteros** se ha tenido en cuenta:

- Se han considerado únicamente los centrocampistas que superen los 500 minutos jugados de los 1440 estudiados, debido a que suelen ser más volátiles debido a la distancia recorrida por partido.
- Por otra parte, como se ha estudiado el porcentaje de pases medios acertados por partido, así como el porcentaje de regates acertados, solo se consideran los jugadores que han entrado en contacto con el balón 400 veces en el total de minutos estudiado.
 1. El número de recuperaciones de balón medio por partido, con valor positivo y ponderado entre el total de los valores.
 2. El porcentaje de pases acertados medio por partido, con valor positivo y ponderado entre el total de los valores.
 3. El número total de asistencias, que obtiene un valor de 0.01 por asistencia creada.
 4. El porcentaje de regates acertados medio por partido
 5. El número de goles, que obtiene un valor de 0.01 por gol.

4. Conclusiones:

El equipo final es el siguiente:

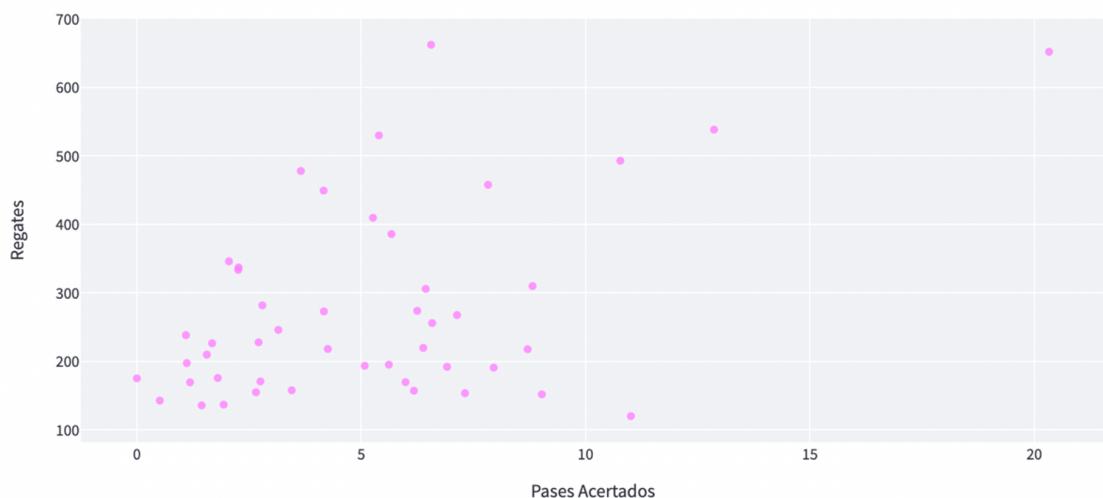
- Porteros: Jan Oblak (Atlético de Madrid)
- Defensas: Stefan Savić (Atlético de Madrid), Jules Koundé (Sevilla FC)
- Laterales: Javi Galán (SD Huesca), Jesús Navas (Sevilla FC)
- Centrocampistas: Carlos Soler (Valencia) , Luka Modrić (Real Madrid)
- Bandas: Mikel Oyarzabal (Real Sociedad), Marcos Llorente (Atlético de Madrid)
- Delanteros: Iago Aspas (Celta), Karim Benzema (Real Madrid)

Hay 3 jugadores en la plantilla que no corresponden a los 4 primeros de la tabla. Por tanto, rechazamos la hipótesis nula que establecía que al menos el 80% de los jugadores alineados pertenecen a los primeros clasificados en La Liga 2020/2021 a día 31 de diciembre de 2020.

Para la hipótesis: **El número de pases acertados por un futbolista es directamente proporcional al número de regates realizados por el mismo.**

- Establecemos un Scatterplot y un Bubble Chart con los regates acertados situados sobre el eje “x” y los pases acertados sobre el eje “y”.

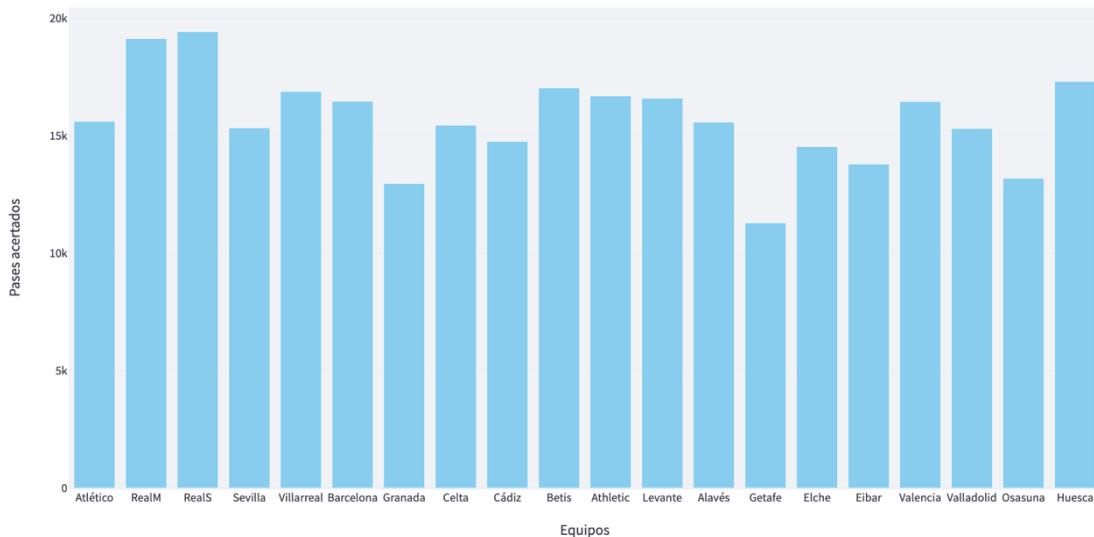
Correlación Pases acertados / Regates Acertados



- Como no se puede apreciar visualmente con claridad si existe una relación, se procede a realizar el coeficiente de correlación entre las variables.
- El coeficiente de correlación es de 0.47 positivo. Al ser menor que 0.5, no podemos afirmar que exista una correlación representativa entre las variables.

Para la hipótesis: **El número de pases acertados es directamente proporcional a la posición en la tabla de cada equipo.**

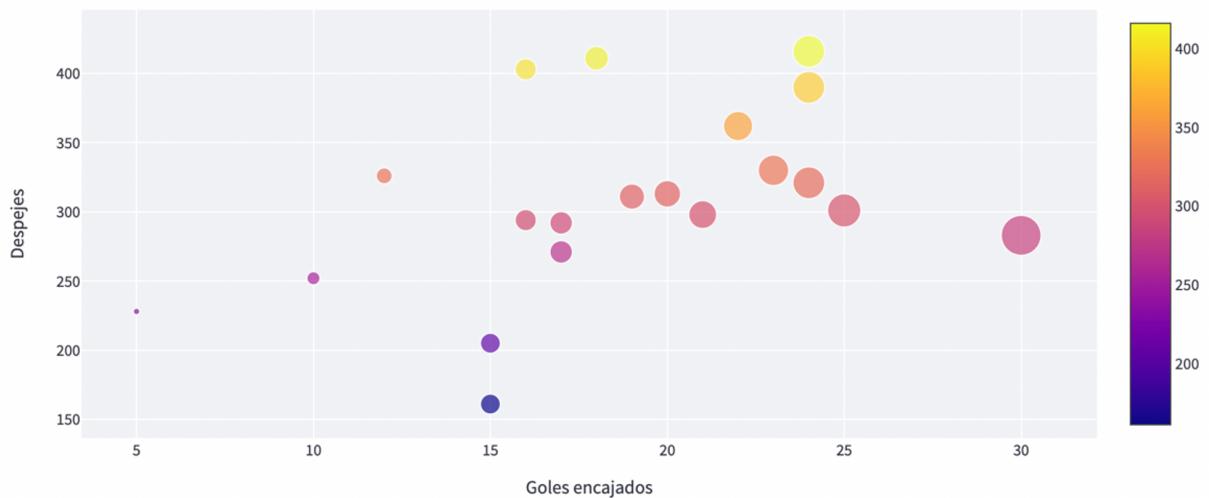
- Establecemos un gráfico de barras con el total de pases acertados por los centrocampistas de cada equipo. En el eje x situamos a los equipos.



- Para ver con mayor claridad que equipos realizaron más pases acertados, dibujamos un pie chart donde se ve que equipo a realizado el mayor porcentaje de pases acertados.
- Como no se puede apreciar visualmente con claridad si existe una relación, se procede a realizar el coeficiente de correlación entre las variables.
- El coeficiente de correlación es de 0.47 positivo. Al ser menor que 0.5, no podemos afirmar que exista una correlación representativa entre las variables.

Para la hipótesis: **El número de despejes realizados por equipo es directamente proporcional a los goles encajados.**

- Establecemos un Bubble Chart con los goles encajados situados en el eje x y el número total de despejes en el eje y.



- Como no se puede apreciar visualmente con claridad si existe una relación, se procede a realizar el coeficiente de correlación entre las variables.
- El coeficiente de correlación es de 0.47 positivo. Al ser menor que 0.5, no podemos afirmar que exista una correlación representativa entre las variables.

Para la pregunta: **¿Qué defensas sobresalen a nivel estadístico en alguna de las variables de estudio?**

Para una mejor visualización, se realiza un boxplot donde podremos observar los outliers de cada posición. Destacan:

- En recuperaciones de balón, sobresalen David García del Osasuna, el cual fue uno de los centrales de la temporada y recuperó una media de 15 balones por partido, y Djené Dakonam, perteneciente al Getafe FC, con una estadística similar.
- En disparos bloqueados, destaca principalmente Rubén Duarte, lateral izquierdo del Deportivo Alavés, el cual bloqueó una media de 3 disparos por partido. Otros jugadores destacados en este ámbito fueron Pau Torres, defensa del Villarreal FC que posteriormente sería convocado por la Selección Española de Fútbol. Por último, destaca la escasa cantidad de disparos bloquea Yeray, el defensa central habitualmente titular del Athletic Club.

5. Recursos utilizados

Obtención de datasets: *Kaggle*.

Lenguaje de programación: *Python*.

Interfaz web de código: *Jupyter Notebook*.

Editor de código: *Visual Studio Code*.

Desarrollo de la web: *Streamlit*.

Librerías utilizadas:

- *Numpy*
- *Pandas*
- *Plotly*
- *Streamlit*

