

Proyecto ETL - Football Player Stats
Juan Camilo Vargas - 2216273
Sebastian Diaz Noguera - 2216299

Análisis

Este conjunto de datos contiene estadísticas de jugadores de fútbol de los años 2021 y 2022. Hay cerca de 2900 filas, que representan a cada jugador, y 143 columnas, que representan diferentes estadísticas, incluyendo nombre, edad, nacionalidad, posición, equipo, goles, asistencias, partidos jugados, tarjetas amarillas, rojas entre otros. Este conjunto de datos se puede utilizar para una gran variedad de propósitos, incluyendo análisis de rendimiento, identificación de talentos y desarrollo de estrategias. Por ejemplo, un equipo de fútbol podría utilizar el conjunto de datos para analizar el rendimiento de sus jugadores e identificar áreas en las que pueden mejorar. Finalmente, el conjunto de datos se puede utilizar para desarrollar estrategias de juego que ayuden a un equipo a ganar partidos.

Nuestro segundo conjunto de datos posee información básica sobre los jugadores, pero adicional a esto posee el dato del precio de los jugadores, el cual nos va a permitir de acuerdo a las estadísticas calcular el precio del jugador mediante una regresión lineal.

En nuestra búsqueda de un tema de interés, nos llamó bastante la atención estos dos datasets relacionados a las diferentes estadísticas que nos brinda el fútbol, que en pocas ocasiones son tenidos en cuenta para contratar jugadores. Por otro lado, aunque ya se han realizado aplicaciones sobre el tema, creemos que esta puede ser una buena oportunidad para desarrollar las habilidades necesarias para el manejo de datos complejos.

El objetivo de este proyecto es la de crear una herramienta que le sirva a los equipos de fútbol en la creación de la plantilla ideal acorde a las estadísticas que busque de cada jugador en una posición específica, teniendo en cuenta los recursos económicos del equipo, reduciendo los tiempos de búsqueda, y adecuando el presupuesto de equipo de acuerdo a jugador deseado, aumentando la eficiencia. Para abordar esta situación, estamos considerando proporcionar estadísticas detalladas de los jugadores. Creemos que brindar información estadística objetiva puede ofrecer una visión más completa y precisa de las habilidades y el rendimiento de un jugador. Al contar con datos concretos sobre su historial y estadísticas de juego, los equipos tendrán una base más sólida para tomar decisiones informadas en el proceso de selección y contratación de jugadores y así facilitarle a los equipos la evaluación de talentos, ayudando así a mitigar la influencia de representantes que

solo destacan lo positivo. Al proporcionar datos cuantificables y relevantes, buscamos contribuir a una toma de decisiones más equitativa y beneficios para todos los involucrados en el mundo del fútbol.

Siguiendo la línea de pensamiento y análisis basados en las estadísticas de los jugadores de fútbol, se identificó la oportunidad de llevar a cabo una regresión lineal múltiple como un enfoque clave para el proyecto. Esta técnica de análisis estadístico, ampliamente utilizada en el campo de la ciencia de datos y el análisis de datos deportivos, tiene el potencial de proporcionar insights valiosos para el cálculo del precio de los jugadores. Dado que el precio de un jugador es un factor crucial en las decisiones de contratación de los equipos, encontrar una manera de estimar este valor faltante es de suma importancia. La regresión lineal múltiple permite explorar las relaciones complejas entre las diversas estadísticas de los jugadores y su precio, lo que puede proporcionar una estimación precisa y fundamentada

1.2021-2022 Football Player Stats:

Este primer dataset contiene información sobre las estadísticas de más de 2900 jugadores de fútbol, el cual posee un enorme potencial analítico ya que se podría crear modelos los cuales facilitan la selección de un jugador de acuerdo a su desempeño en diferentes áreas o posiciones del equipo.

Columnas relevantes

Las columnas seleccionadas como 'relevantes' son las que poseen información de gran utilidad para la realización de este proyecto, con el fin de dar un análisis estadístico de cada jugador de forma precisa.

- **Player:** Nombre del jugador, esta columna es esencial para la identificación del jugador y será usada para la creación de la dimensión que contenga dicho nombre junto con un ID.
- **Nation:** Nacionalidad del jugador, que nos servirá para obtener más información del jugador, también se creará una dimensión y de esta manera dividirlos por selección en la que juegan.
- **Pos:** Posición del jugador, esta columna nos permitirá conocer la posición del jugador en el campo de juego tales como: GK, DF, MF, FW.
- **Squad:** Equipo en donde juega, es importante para poder comparar a los jugadores de forma justa, ya que las estadísticas de un jugador pueden verse afectadas por el equipo en el que juega.
- **Comp:** La columna de liga del dataset nos indica la competición en la que está jugando el jugador. Esto es importante porque las diferentes ligas tienen diferentes niveles de competencia y diferentes estilos de juego.
- **Age:** Edad del jugador, es importante realizar un análisis del rendimiento de los jugadores de fútbol porque es un indicador de madurez y experiencia de un jugador. Los jugadores más jóvenes suelen ser más rápidos y ágiles,

mientras que los jugadores más experimentados suelen ser más inteligentes y hábiles tácticamente.

- **MP** : Partidos jugados, es un dato que nos permite saber si un jugador ha sido tenido en cuenta en su equipo, lo que nos refleja la regularidad con la disputado partidos.
- **Starts**: Partidos en los que fueron titulares, este datos nos dice cuántas veces el jugador empezó como parte del once inicial del partido.
- **Min**: Minutos jugados, este datos es importante ya que nos permite medir el periodo de tiempo en el que el jugador ha estado en el campo y que a su vez nos demuestra también el rendimiento que puede llegar a tener.
- **Goals**: Promedio de gol, este dato es la división de goles hechos por el jugador con los partidos disputados, es un dato muy relevante a la hora de medir el rendimiento.
- **Shots**: Promedio total de tiros (No incluye tiros penales), es la división de los tiros que no van con dirección al arco por cada partido jugado, al igual que la anterior variable nos permite conocer el desempeño del jugador a la hora de realizar tiros.
- **SoT**: Promedio tiros a puerta (No incluye tiros penales) es la división de los tiros al arco (dentro los tres palos) por cada partido jugado, es una variable muy importante para medir la efectividad del jugador.
- **SoT%**: Porcentaje de tiros a puerta (No incluye tiros penales) esta variable es para tener una medida más exacta a la anterior, ya que es un dato más fácil de comprender para los usuarios de la herramienta.
- **ShoDist**: Distancia promedio, en yardas, desde x punto a la portería de todos los tiros realizados (No incluye penales) es una variable que ayudará a identificar de manera rápida qué capacidad de tiro tiene un jugador ya que puede ser de larga o medio distancia.
- **ShopK** : Penaltis hechos (Promedio), es la división de penaltis hechos por el jugador con los penaltis intentados.
- **PKatt** : Penaltis intentados (Promedio) nos puede permitir indicar de una manera más clara cuantos penaltis intentados tiene cada jugador.
- **PasTotCmp** : Pases completados (Promedio) es la división de los pases completados con los pases intentados, es una variable muy significativa debido a la frecuencia con la que ocurre la acción en un partido, lo que conlleva a ser una cualidad muy relevante para los directivos.
- **PasAtt** : Pases intentados, permite ver de manera comparativa los pasos que completo a los que intento.
- **PasTotCmp%** : Porcentaje de finalización del pase, esta variable es para tener una medida más exacta a la anterior, ya que es un dato más fácil de comprender para los usuarios de la herramienta.
- **PasShoCmp%** : Porcentaje de pases completados (Pases entre 5 y 15 yardas),ya que nos sirve para el análisis de rendimiento de los jugadores, especialmente en las posiciones relacionadas la distribución del balón, como centrocampistas y mediocampistas.

- **PasMedCmp%** : Porcentaje de pases completados (Pases entre 15 y 30 yardas), esta métrica se enfoca en la habilidad y precisión de un jugador al ejecutar pases de alcance intermedio, lo cual tiene un impacto considerable en la efectividad de generar juego
- **PasLonCmp%** : Porcentaje de pases completados (Pases de más de 30 yardas), es la capacidad de un jugador para ejecutar pases de larga distancia.
- **Asistencias** : Asistencias (promedio) Es el resultado de dividir las asistencias (Pase que permite a otro jugador marcar gol) dadas por un jugador por el número de partidos disputados, La cantidad de asistencias realizadas por un jugador refleja su destacado desempeño en el campo. Esta métrica se convierte en un factor crucial a tener en cuenta por los directivos.
- **CrsPA** : Centros completos al área, nos parece un dato considerable que refleja la capacidad de un jugador para generar oportunidades de gol y peligro para el equipo contrario, demuestra su impacto en el campo de manera significativa.
- **PasPress**: Pases realizados bajo la presión del oponente, este factor evidencia la resiliencia del jugador a la hora de estar en situaciones de estrés.
- **Sw** : Pases que recorren más de 40 yardas del ancho de la cancha (cambió de frente) refleja la visión de juego que puede tener un jugador, una cualidad diferenciadora de un jugador.
- **PaswLeft**: Pases intentados con el pie izquierdo, este dato nos permite conocer el pie dominante del jugador que puede denotar un factor importante en la eficacia del jugador.
- **PaswRight**: Pases intentados con el pie derecho, este dato nos permite conocer el pie dominante del jugador que puede denotar un factor importante en la eficacia del jugador.
- **SCA** : Acciones de creación de tiro, esta métrica se centra en las habilidades y la capacidad de un jugador para crear situaciones que culminen en intentos de tiro al arco, principalmente son ejecutadas por los mediocampistas (MF) y delanteros (FW).
- **TkIDef3rd** : Tackles en 1/3 defensivo, es la habilidad y efectividad de un jugador para ejecutar entradas y recuperar el balón en la parte del campo cercana a su propia portería, métrica que sirve para identificar a un buen defensa (DF)
- **TkIMid3rd** : Tackles en medio 1/3 es la habilidad y efectividad de un jugador para ejecutar entradas y recuperar el balón en la parte del medio campo, métrica que sirve para identificar a un mediocampista defensivo (MF)
- **TkIAtt3rd** : Tackles en ataque 1/3 es la habilidad y efectividad de un jugador para ejecutar entradas y recuperar el balón en la parte del medio campo, métrica que sirve para identificar a un delantero con rasgos defensivos (FW)
- **Int** : Intercepciones (balones interceptados) es una variable muy importante principalmente para los jugadores de la zona defensiva.

- **Tkl+Int** : representa la suma de dos métricas esenciales en la zona defensiva: los tacleadas exitosas y las intercepciones, nos ayuda a reflejar el rendimiento de los jugadores en la zona defensiva.
- **DriSucc**: Regates completados con éxito, es una cualidad muy importante ya que nos permite saber la destreza que tienen un jugador para desequilibrar a sus oponentes y generar peligro al equipo rival.
- **DriAtt** : Regates intentados
- **DriSucc%** : Porcentaje de regates completados con éxito, es una variable que nos permite exponer de mejor manera a los usuarios la habilidad de que tiene cada jugador de poder driblar a un oponente.
- **CrdY** : Tarjetas amarillas, nos permite analizar el rendimiento de los jugadores ya que este es un indicador de la disciplina de un jugador.
- **2CrdY** : Segunda tarjeta amarilla, nos muestra la reincidencia del jugador a la hora de cometer faltas, lo que puede afectar tanto el desempeño del jugador como del equipo.
- **CrdR** : Tarjetas rojas, es el número de veces en el que el jugador ha sido expulsado del partido por su reincidencia al cometer faltas o ya puede ser por agresiones u otro tipo de comportamiento, un claro factor a considerar dentro del análisis de desempeño.
- **Fls** : Faltas cometidas, se centra en la cantidad de infracciones que un jugador comete durante un partido lo que ofrece una perspectiva sobre su agresividad en el juego y su capacidad para detener las acciones del equipo contrario.
- **Fld** : Faltas recibidas, reflejar el nivel de atención que el jugador recibe por parte del equipo contrario y su capacidad para atraer la marca y crear oportunidades para su equipo.
- **PKcon** : Penales concedidos, es una variable que nos permite identificar con que frecuencia los jugadores conceden penales en contra
- **OG** : Goles en propia puerta, metrica que aunque a simple vista no parece ser significativa, puede expresar mucho acerca del rendimiento del jugador.
- **AerWon** : Aéreas ganadas, es relevante para influir en el juego en zonas clave del campo y para establecer el control del balón en situaciones donde la disputa por el balón se da en el aire.
- **AerLost** : Aéreas pérdidas, proporciona información sobre la capacidad de un jugador para competir, su fortaleza física y su capacidad para mantener la posesión o evitar que el oponente tome el control del balón después de un duelo en el aire
- **AerWon%** : Porcentaje de Aéreas ganadas, al igual que algunas variables vistas anteriormente, esta métrica nos sirve para proporcionar le una medida más comprensible para el público.
- **PPA** : Pases completos al área de 18 yardas (Área pequeña del equipo rival), número de veces que el jugador centró el balón al área de 18 yardas. Otra variable que nos permite evaluar la capacidad de generar peligro al equipo contrario.

TkIDri : Número de dribladores abordados, esto puede indicar su agilidad, capacidad de anticipación y habilidad para mantener la posesión o recuperar el balón. Además, esta estadística puede revelar la influencia del jugador en la contención defensiva y su capacidad para frustrar las jugadas ofensivas del oponente

- **TkIDri%** : Porcentaje de dribladores abordados para simplificar la métrica anterior.
- **TouDefPen** : Toques en área defensiva, determina la habilidad para iniciar la transición desde la defensa al ataque y su contribución en la salida controlada del balón desde situaciones potencialmente peligrosas.
- **ShoFK** : Remates de tiros libres (Promedio)

Columnas a considerar

Las columnas de esta sección, son columnas a considerar, es decir, columnas que podrían usarse en un futuro para un análisis más profundo pero que para el objetivo de este proyecto no se necesitan de momento.

- **Press**: Número de veces que se aplica presión al jugador contrario que recibe, lleva o suelta el balón (promedio)
- **G/Sh** : Goles por tiro (Promedio)
- **PasFK** : Pases intentados de tiros libres
- **CK**: Saques de esquina
- **PasInt** : Interceptado (Pase interceptado)
- **GCA** : Acciones creadoras de objetivos
- **TouDef3rd** : Toques en 1/3 defensivo
- **TouMid3rd**: toques en medio 1/3
- **TouAtt3rd** : Toques en ataque 1/3
- **CarTotDist**: Distancia total, en yardas, que un jugador movió la pelota mientras la controlaba con los pies, en cualquier dirección
- **CarPrgDist**: Distancia total, en yardas, que un jugador movió el balón mientras lo controlaba con los pies hacia la portería contraria.
- **Acarreos**: Número de veces que el jugador controló el balón con los pies
- **CarDis** : Número de veces que un jugador pierde el control del balón tras ser placado por un jugador contrario
- **Rec**: número de veces que un jugador recibió un pase con éxito
- **Rec%** : Porcentaje de tiempo que un jugador recibió con éxito un pase
- **Crs** : Cruces
- **BlkShSv** : Número de veces que bloquea un tiro que estaba en el blanco, interponiéndose en su camino.
- **Blocks**: Número de veces que bloquea la pelota parándose en su camino (anticipo), se enfoca en la cantidad de veces que un arquero interrumpe el

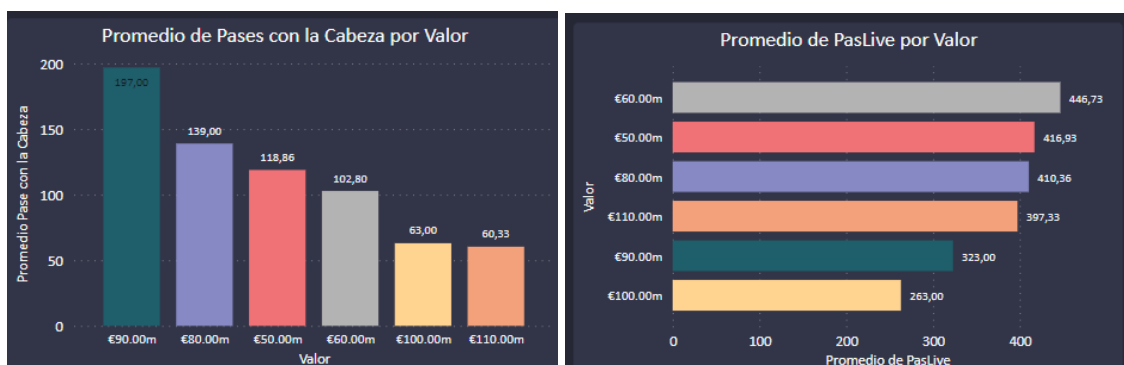
trayecto de un disparo o un pase del equipo contrario, con el objetivo de impedir que el balón llegue a la portería

Columnas eliminadas

La mayoría de las columnas eliminadas hasta el momento son debido que estas no representan información importante o aplicable para el análisis de los datasets las cuales pueden tener columnas repetidas o valores incoherentes con las demás columnas.

Columnas eliminada de pases

Siguiendo nuestra métricas de elección consideramos que estas variables acerca de esta acción de juego se eliminaron ya que cuenta con información excesivamente detallada acerca de esta, por lo que no genera de alguna manera significación a la hora ya bien sea cuando un director técnico está buscando la elección de su plantilla ideal ya que no son cualidades que diferencian a un jugador de otro, y de igual de manera al momento de calcular el valor mediante la regresión lineal que queremos llevar a cabo, para respaldar esta decisión se crearon gráficas en la herramienta Power bi que reflejaban lo mencionado.



En estas 2 gráficas se puede observar como por ejemplo “ Pases con la cabeza” no afecta de gran manera el valor del mercado del jugador , ya que se puede ver como los mejores pagados (100.00m,110.00m) cuentan casi con la mitad de promedio de pases que los menos pagado en esta gráfica.

Lo mismo se puede observar en la segunda gráfica de Pases con el “balón en movimiento” donde se observa que los menos pagados son los que mejor promedio tienen.

También se eliminaron algunas columnas que contaban con información repetidas de las seleccionadas anteriormente como, por ejemplo :


```
p = players_df[['PasAtt', 'PasTotAtt']]
a = p.head(10)
print(a)
```

	PasAtt	PasTotAtt
0	45.00	45.00
1	47.00	47.00
2	61.00	61.00
3	49.80	49.80
4	12.00	12.00
5	17.20	17.20
6	9.09	9.09
7	20.20	20.20
8	35.50	35.50
9	71.10	71.10

```
p = players_df[['PasCmp', 'PasTotCmp']]
a = p.head(10)
print(a)
```

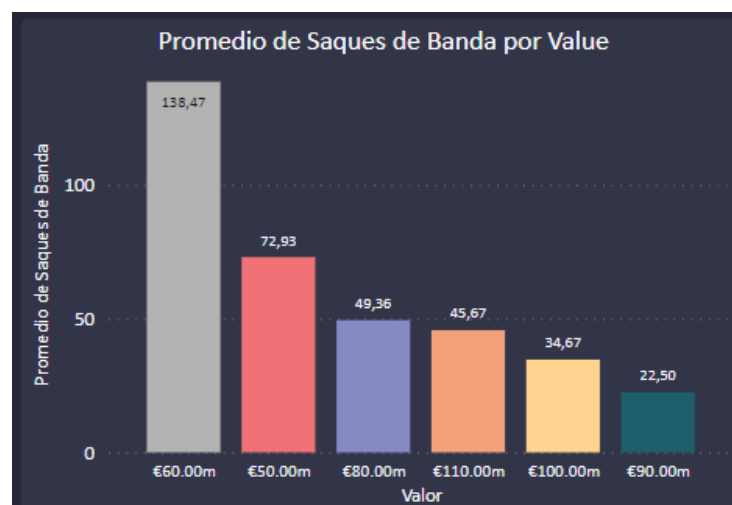
	PasCmp	PasTotCmp
0	34.00	34.00
1	38.70	38.70
2	55.90	55.90
3	40.70	40.70
4	4.00	4.00
5	11.10	11.10
6	7.27	7.27
7	14.60	14.60
8	31.30	31.30
9	64.30	64.30

- **PasTotAtt** : Pases intentados (parecen promedios) (Columna Repetida)
- **PasCmp** : Pases completados (Columna Repetida)
- **GcaPassLive** : Pases completos de balón vivo que conducen a un gol (Columna Repetida)
- **GcaPassDead**: Pases a balón parado completados que conducen a un gol (Columna Repetida)
- **ScaPassLive**: Pases completos de balón vivo que conducen a un intento de tiro
- **ScaPassDead**: Pases a balón parado completados que conducen a un intento de tiro
- **TB** :Pase completo enviado entre defensores traseros al espacio abierto
- **PasLive** : Pases de balón en vivo
- **PasDead**: Pases a balón parado
- **PasHigh**: Pases que están por encima del nivel de los hombros en la altura máxima
- **PaswHead** : Pases intentados usando la cabeza
- **PasLow**: pases que dejan el suelo, pero se mantienen por debajo del nivel de los hombros
- **PasTotDist**: distancia total, en yardas, que los pases completos han recorrido en cualquier dirección
- **PasTotPrgDist** : Distancia total, en yardas, que han recorrido los pases completos hacia la meta del oponente
- **PasShoCmp** : Pases completados (Pases entre 5 y 15 yardas)
- **PasShoAtt** : Pases intentados (Pases entre 5 y 15 yardas)
- **PasMedCmp** : Pases completados (Pases entre 15 y 30 yardas)
- **PasMedAtt** : Pases intentados (Pases entre 15 y 30 yardas)
- **PasLonCmp** : Pases completados (Pases de más de 30 yardas)
- **PasLonAtt** : Pases intentados (Pases de más de 30 yardas)
- **PasAss**: pases que conducen directamente a un tiro (tiros asistidos)
- **Pas3rd**: Pases completos que entran en el 1/3 del campo más cercano a la portería

- **PasProg**: Pases completos que mueven el balón hacia la meta del oponente al menos 10 yardas desde su punto más lejano en los últimos seis pases, o cualquier pase completo al área de penalti.
- **PasGround** : Pases de tierra
- **PaswOther** : Pases intentados usando partes del cuerpo que no sean la cabeza o los pies del jugador
- **PasOut**: fuera de los límites
- **RecProg**: Pases completos que mueven el balón hacia la meta del oponente al menos 10 yardas desde su punto más lejano en los últimos seis pases, o cualquier pase completo al área penal

Columnas de reglas o sucesos eliminadas

Estas columnas se eliminaron ya que estas reglas o estas acciones no muestran las capacidades de un jugador, por lo que no se utilizan como una métrica o un factor para determinar el valor del mercado. De igual manera se utilizó power bi para respaldar lo mencionado.



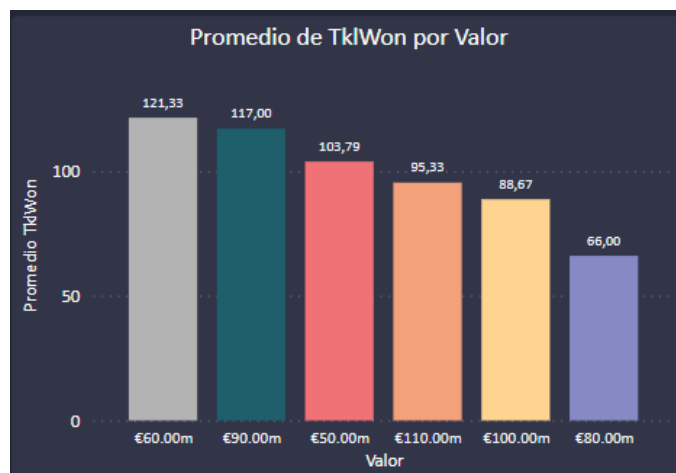
Como se puede observar en la gráfica el promedio de los saques de banda no afectó de ninguna manera el valor del mercado de cada jugador, ya que como lo mencionamos anteriormente no es un factor diferenciador, es por esta razón que no van a ser tenidas en cuentas estas columnas:

- **CkIn** : Tiros de esquina hacia adentro
- **CkOut** : Saques de esquina desviados
- **CkStr** : Saques de esquina directos
- **TI** : Saques de banda realizados
- **PasOff** : Fuera de juego

- **PasOut**: fuera de los límites
- **Clr** : Espacios libres

Columnas de información del equipo eliminadas

Teniendo en claro nuestro objetivo de proporcionar el valor de los jugadores al tener información de estadísticas en general de los equipos de cada jugador, estas columnas no generan alguna influencia en el valor de mercado es por esto que se tomó la decisión de eliminarlas, para respaldar lo mencionado tenemos esta gráfica.



Esta gráfica tiene el promedio de las disputas del balón ganadas por el equipo total, y de igual manera que en las anteriores gráficas esta estadística no representa significancia a la hora de calcular el valor de cada jugador. Teniendo en cuenta esto se eliminaron las siguientes columnas:

- **TklWon** : Tackles en los que el equipo del tackleador ganó la posesión del balón, fue eliminada ya que se centra específicamente en el equipo del jugador, no en el jugador como tal.
- **PresSucc** : Número de veces que el equipo ganó la posesión dentro de los cinco segundos de aplicar presión
- **Press%** : Porcentaje de tiempo que el equipo obtuvo la posesión dentro de los cinco segundos de aplicar presión

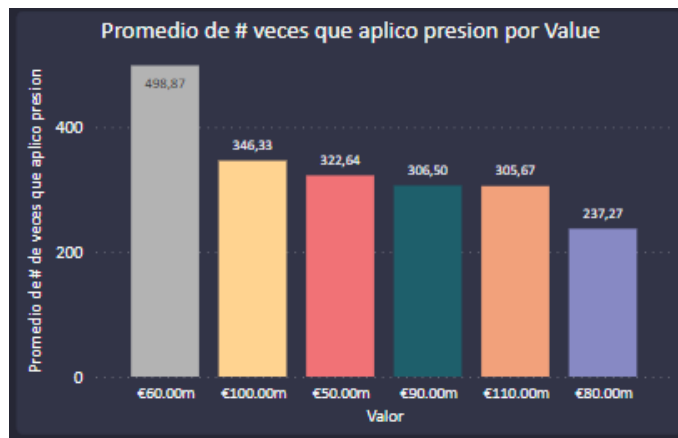
Columnas Redundantes y No Relevantes

Las columnas individuales eliminadas, se retiraron aquellas columnas que no ofrecían contribuciones sustanciales al análisis global o cuyos valores se podían deducir directamente de otras variables disponibles en la base de datos. Un ejemplo claro de esto se encuentra en las columnas "**Born**" y "**G/Sot**", estas columnas se consideran redundantes, ya que la información sobre la fecha de nacimiento de un jugador (Born) y la relación entre goles y tiros a puerta (G/Sot) se pueden obtener de manera más eficiente a partir de otros datos existentes en el conjunto

- **Born**: Año de nacimiento,
- **G/SoT**: Promedio goles por tiro a puerta (No incluye penales)
- **Recov** : Número de balones sueltos recuperados,
- **PKwon** : Penaltis ganados
- **90s** : minutos jugados divididos por 90
- **Err** : Errores que conducen al tiro de un oponente
- **TkIW** : Tackles en los que el equipo del tackleador ganó la posesión del balón(repetida)
- **CPA** : Acarreo al área de 18 yardas
- **CarProg** : Acarreos que mueven el balón hacia la meta del oponente por lo menos 5 yardas, o cualquier acarreo dentro del área penal.
- **GcaDrib**: regates exitosos que conducen a un gol, esta métrica fue descartada debido a que ya hay otras que ofrecen el mismo tipo de información
- **ScaDrib**: regates exitosos que conducen a un intento de tiro, columna irrelevante ya que existe una que ya proporciona información
- **BlkPass**: número de veces que bloquea un pase al pararse en su camino

Columnas eliminadas por criterio

Las columnas eliminadas por criterio se refieren a aquellas características o variables que fueron retiradas de un conjunto de datos columnas que no aportan información significativa al análisis o que pueden ser redundantes. El objetivo principal de eliminar estas columnas es simplificar el conjunto de datos, reducir la complejidad y mejorar la calidad de los datos que se utilizarán en análisis posteriores.



- **PasBlocks**: Bloqueado por el oponente que estaba parado en el camino
- **ScaSh**: tiros que conducen a otro intento de tiro
- **ScaFld** : Faltas cometidas que conducen a un intento de tiro
- **ScaDef**: acciones defensivas que conducen a un intento de tiro
- **GcaSh** : Disparos que conducen a otro tiro de gol
- **GcaFld** : Faltas cometidas que conducen a gol
- **GcaDef** : Acciones defensivas que conducen a un gol
- **TkIDriAtt** : Número de veces que se regatea más el número de tacleadas
- **TkIDriPast** : Número de veces que un jugador contrario regatea
- **Tkl** : Número de jugadores placados tacleados
- **TouAttPen** : Toques en el área de penalti de ataque
- **TouLive**: toques de balón en vivo. No incluye tiros de esquina, tiros libres, saques de banda, saques de banda, saques de meta ni tiros penales.
- **DriPast** : Número de jugadores pasados driblando
- **Car3rd** : Acarreos que entran en el 1/3 del campo más cercano a la portería
- **CarMis** : Número de veces que un jugador falló al intentar hacerse con el control de un balón
- **RecTarg**: Número de veces que un jugador fue el objetivo de un intento de pase
- **DriMegs**: número de veces que un jugador regatea el balón a través de las piernas de un jugador contrario
- **Toques**: Número de veces que un jugador tocó el balón. Nota: recibir un pase, luego driblar y luego enviar un pase cuenta como un toque
- **PresDef3rd** : Número de veces que aplica presión al jugador contrario que está recibiendo, llevando o soltando el balón, en el 1/3 defensivo
- **PresMid3rd** : Número de veces que aplica presión al jugador contrario que está recibiendo, llevando o soltando el balón, en el medio 1/3
- **PresAtt3rd** : Número de veces que aplica presión al jugador contrario que está recibiendo, llevando o soltando el balón, en el 1/3 de ataque

2.Transfermarkt player datas

El segundo conjunto de datos nos brinda datos sobre una amplia cantidad de jugadores, superando los 10,000 registros. Este conjunto de datos incluye información detallada sobre cada jugador, como su nombre, edad, equipo, nacionalidad, posición y un atributo especialmente crucial: su valor. Este valor se convierte en un elemento esencial para llevar a cabo el cálculo del precio de aquellos jugadores cuyos valores están ausentes en la base de datos.

Columnas relevantes

- **_id**: Identificador
- **Player_Name**: Nombre del jugador, esta columna es esencial para la identificación del jugador
- **Value**: Precio del jugador, esta columna es esencial para la estimación del precio de los jugadores que no poseen dicho dato.

Columnas Eliminadas

Estas columnas fueron eliminadas debido a que el dataset anterior ya posee dicha información.

Age: Edad del jugador

Team: Equipo en el que juega

Nationality: Nacionalidad del jugador

Position: Posicion del jugador

Clasificación de las columnas por posición

Para facilitar la búsqueda y selección de jugadores con las características deseadas en una posición específica, hemos implementado una cuidadosa clasificación de las columnas en nuestra base de datos. Esta clasificación se ha desarrollado considerando las diversas posiciones que existen en el deporte en cuestión y la importancia de brindar a nuestros usuarios la capacidad de identificar rápidamente a los jugadores que mejor se adapten a sus necesidades.

Por ello decidimos clasificar las columnas de la siguiente manera:

ARQUEROS (GK) :

Player Nation, Pos, Squad, Comp, Age, MP, Starts, Min, PasTotCmp, PasAtt, PasTotCmp%, PasMedCmp%, PasLonCmp%, PaswLeft, PaswRight, TouDefPen, CrdY, CrdR , 2CrdY

DEFENSAS (DF) :

Player Nation, Pos, Squad, Comp, Age, MP, Starts, Min, PasTotCmp, PasAtt, PasTotCmp%, PasMedCmp%, PasLonCmp%, PaswLeft, PaswRight, TouDefPen,

CrdY, CrdR , 2CrdY, (TklDef3rd + TklDri), TklDri%, Int , Tkl+Int, Assists, Goals. Fls ,Fld, AerWon,AerLost , AerWon% OG. PKcon, BlkSh .

MEDIOCAMPISTAS (MF) :

Player Nation, Pos, Squad, Comp, Age, MP, Starts, Min, PasTotCmp, PasAtt, PasTotCmp%, PasMedCmp%, PasLonCmp%, PaswLeft, PaswRight, CrdY, CrdR , 2CrdY, Assists, Goals. Fls ,Fld, AerWon,AerLost , AerWon% , (TklDef3rd + TklMid3rd) Assists, Goals. Fls ,Fld, AerWon,AerLost , AerWon%, DriSucc , DriAtt , DriSucc%, Shots, SoT , SoT% , SCA. ShoDist, ShoFK

DELANTEROS (MF) :

Player Nation, Pos, Squad, Comp, Age, MP, Starts, Min, PasTotCmp, PasAtt, PasTotCmp%, PasMedCmp%, PasLonCmp%, PaswLeft, PaswRight, CrdY, CrdR , 2CrdY, Assists, Goals. Fls ,Fld, AerWon,AerLost , AerWon% , (TklMid3rd + TklAtt3rd) Assists, Goals. Fls ,Fld, AerWon,AerLost , AerWon%, DriSucc , DriAtt , DriSucc%, Shots, SoT , SoT% , SCA. ShoDist, ShoCK

Referencias

- *2021-2022 football player stats.* (2022, 29 mayo). Kaggle.
<https://www.kaggle.com/datasets/vivovinco/20212022-football-player-stats>
- *Transfermarkt Player datas , over 10.000 player.* (2023, 14 abril). Kaggle.
<https://www.kaggle.com/datasets/merfarukgnaydn/transfermarkt-player-data-over-10000-player>