



PREDICCIÓN VALORES DE TRANSFERENCIA DE JUGADORES DE FÚTBOL

Proyecto final Coding Dojo - Data Science

Matías Vidal Céspedes



ANTECEDENTES DEL CONJUNTO DE DATOS

- El conjunto de datos se trabajará para crear un modelo predictivo para los valores de transferencia de jugadores de fútbol.
- Contiene:
 - información básica de los jugadores de fútbol como edad, altura, posición de juego.
 - estadísticas profesionales como goles, asistencias (temporadas 2021-2022 y 2022-2023).
 - lesiones y premios totales individuales y de equipo en su carrera.

ANTECEDENTES DEL CONJUNTO DE DATOS

11 LIGAS EUROPEAS

4 LIGAS AMERICANAS

1 LIGA AFRICANA

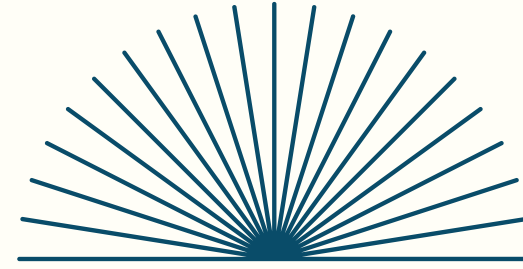
4 LIGAS ASIATICAS



DATAFRAME

El dataframe consta de 10754 filas y 22 columnas.

Nombre columna	Descripción
player	The link to the Transfermarkt page contains player data. Please add a prefix "www.transfermarkt.com/" to further navigating.
team	Name of the team that player played for at the time data was collected.
name	Name of player.
position	Position that player played most.
height	Height of player, cm unit.
age	Age of player.
appearance	The number of times a player appear on field.
goals	Goals scored
asissts	Assits to goals.
yellow cards	Yellow cards.
second yellow	Number of time player get 2 yellow card in 1 match.
red cards	Red cards.
goals conceded	Goal conceded (only Goalkeeper).
clean sheets	Clean sheets (only Goalkeeper).
minutes played	Minutes played.
days_injured	Days injured.
games_injured	Games injured.
award	Total award in whole career.
current_value	Valuated price, Euro unit.
highest_value	Highest valuated price in the past, Euro unit.
position_encode	Goalkeeper – 1, Defender – 2, midfield – 3, Attack - 4
winger	1: winger, 0: not a winger



ANÁLISIS EXPLORATORIO DE DATOS



LIMPIEZA DE LOS DATOS

Duplicados

No existen valores duplicados.

Valores faltantes

No existen valores faltantes.

#	Column	Non-Null Count	Dtype
0	player	10754 non-null	object
1	team	10754 non-null	object
2	name	10754 non-null	object
3	position	10754 non-null	object
4	height	10754 non-null	float64
5	age	10754 non-null	float64
6	appearance	10754 non-null	int64
7	goals	10754 non-null	float64
8	assists	10754 non-null	float64
9	yellow cards	10754 non-null	float64
10	second yellow cards	10754 non-null	float64
11	red cards	10754 non-null	float64
12	goals conceded	10754 non-null	float64
13	clean sheets	10754 non-null	float64
14	minutes played	10754 non-null	int64
15	days_injured	10754 non-null	int64
16	games_injured	10754 non-null	int64
17	award	10754 non-null	int64
18	current_value	10754 non-null	int64
19	highest_value	10754 non-null	int64
20	position_encoded	10754 non-null	int64
21	winger	10754 non-null	int64

Inconsistencias

Se revisan las 22 columnas de forma individual y no se identifican inconsistencias en sus valores.

LIMPIEZA DE LOS DATOS

Columnas innecesarias

Se identificaron que para efectos de la predicción del valor de transferencia de un futbolista no es necesario contar con las siguientes columnas.

PLAYER

TEAM

NAME

POSITION



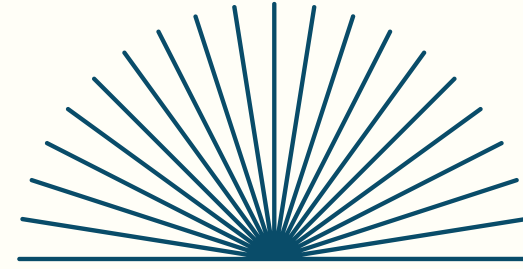
ANÁLISIS EXPLORATORIO DE DATOS

**10754 FILAS
22 COLUMNAS**



**10754 FILAS
18 COLUMNAS**

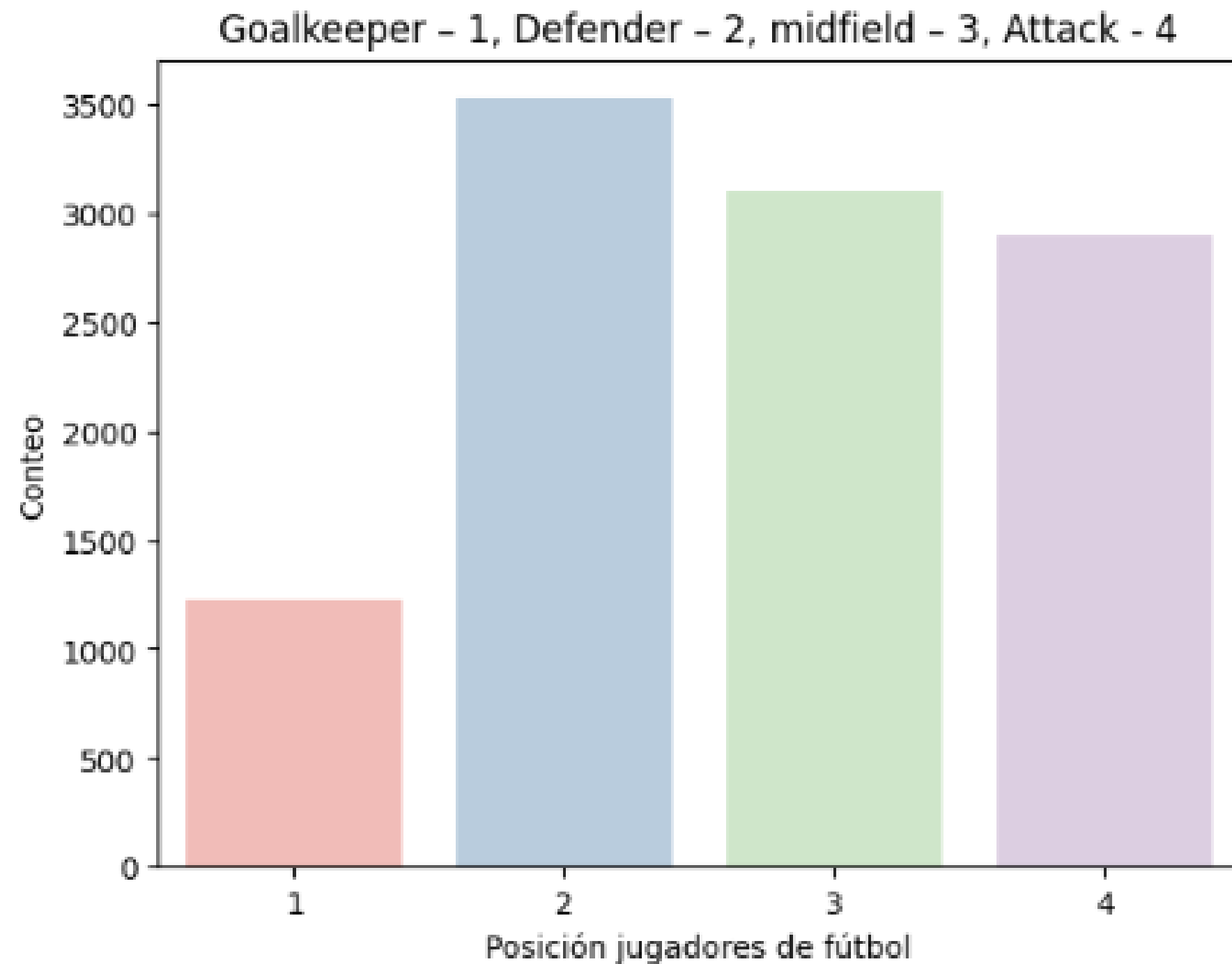




EXPLORACIÓN DE LOS DATOS

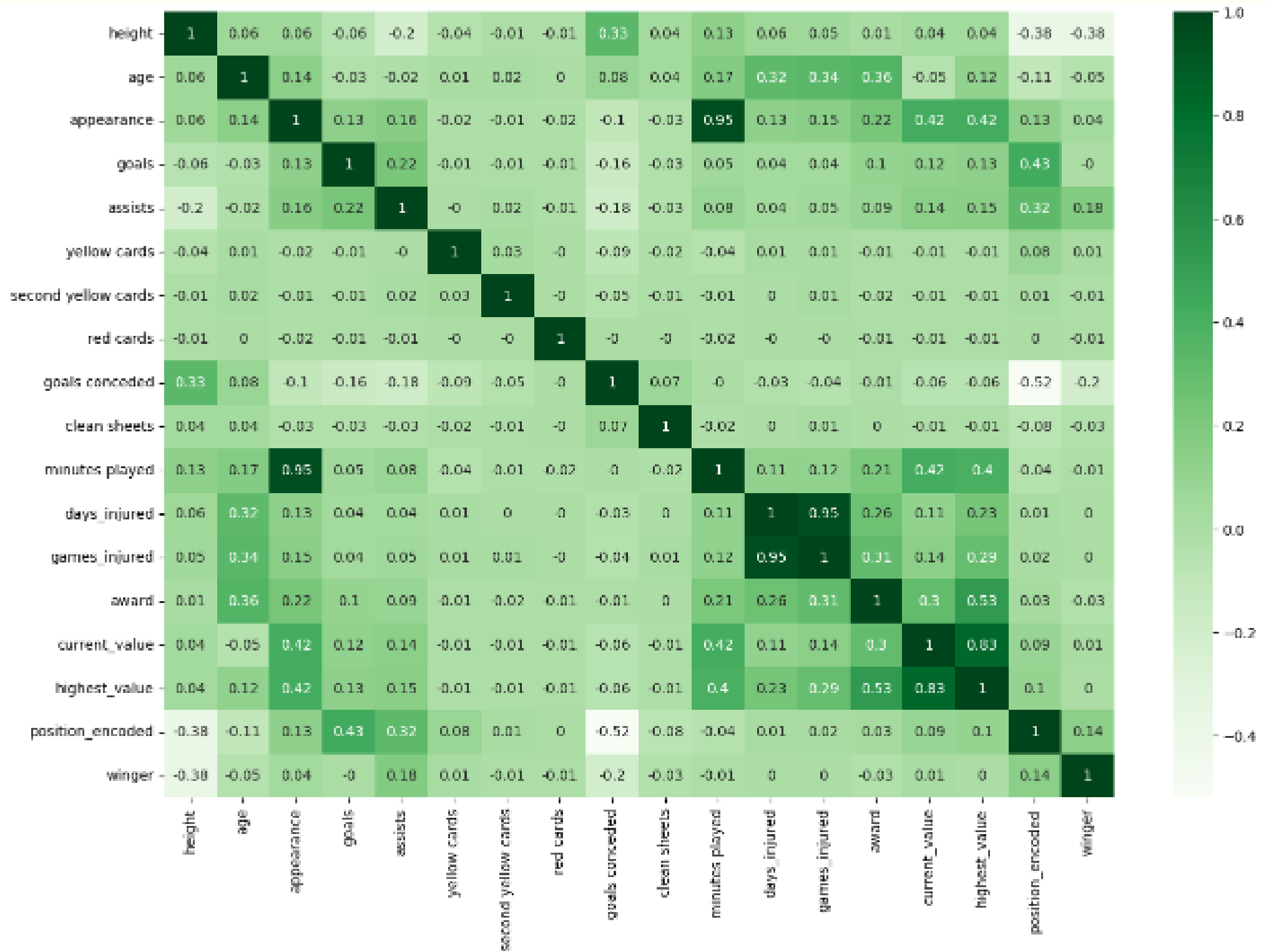


VISUALIZACIÓN POSICIONES DE LOS JUGADORES DE FÚTBOL



Se revela disparidad en la distribución de jugadores según sus posiciones. En particular, se observa que la categoría de defensas abarca un mayor número de jugadores en comparación con el grupo minoritario de arqueros.

VISUALIZACIÓN DE LA CORRELACIÓN DE LAS VARIABLES



Alta correlación:

- "highest_value" tiene una fuerte correlación positiva de aproximadamente 0.835. Esto sugiere que los valores más altos en la columna "highest_value" tienden a estar asociados con valores más altos en "current_value".

Moderada correlación:

- "minutes played", "appearance", "award" muestran correlaciones moderadamente positivas con "current_value", alrededor de 0.421, 0.420 y 0.300 respectivamente. Esto sugiere que jugadores con más minutos jugados, más apariciones o más premios tienden a tener un mayor valor actual.

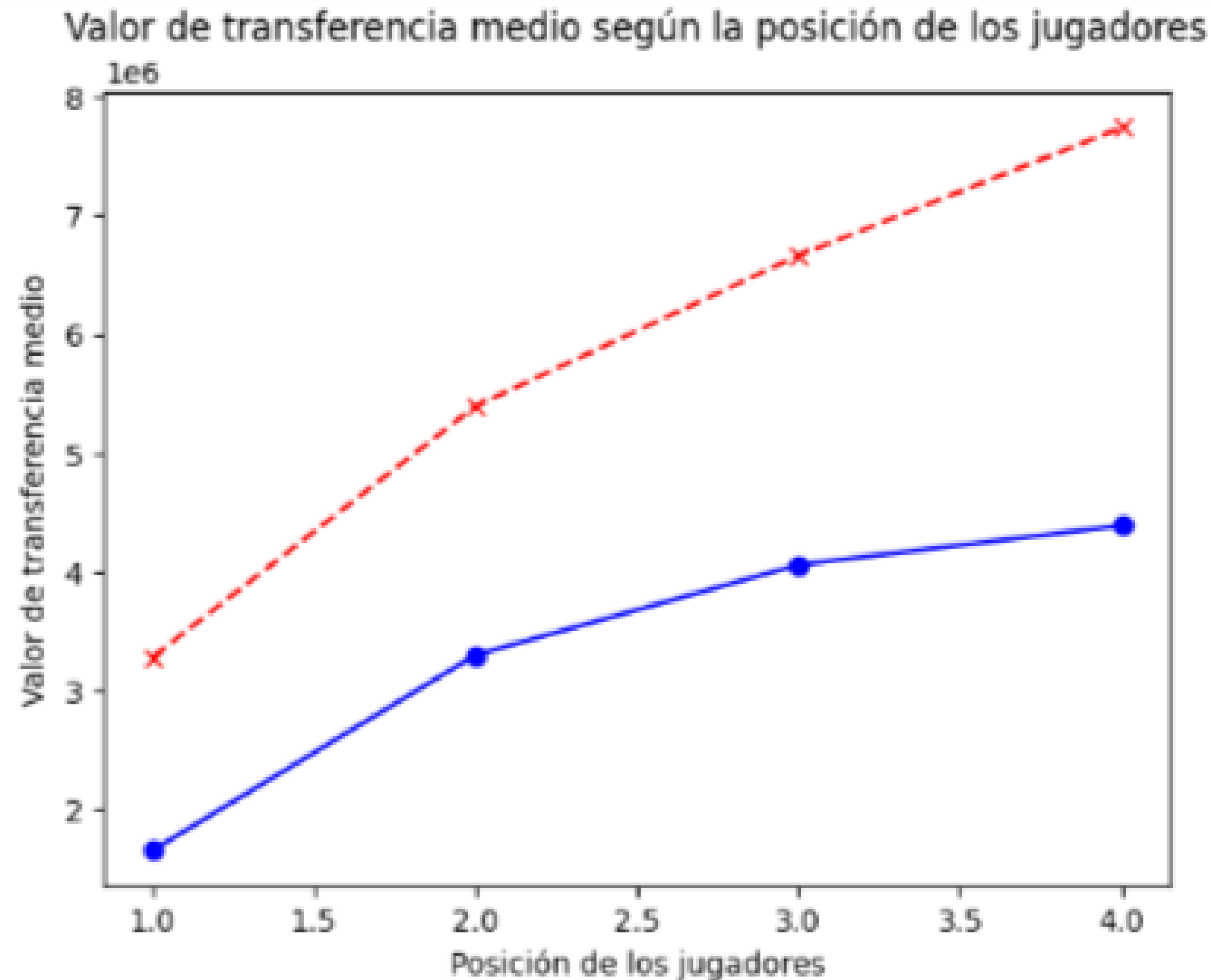
Débil correlación:

- "assists", "games_injured", "goals" y "days_injured" tienen correlaciones débiles con "current_value". Esto indica que hay alguna relación positiva, pero la fuerza de la relación es limitada.

Baja o nula correlación:

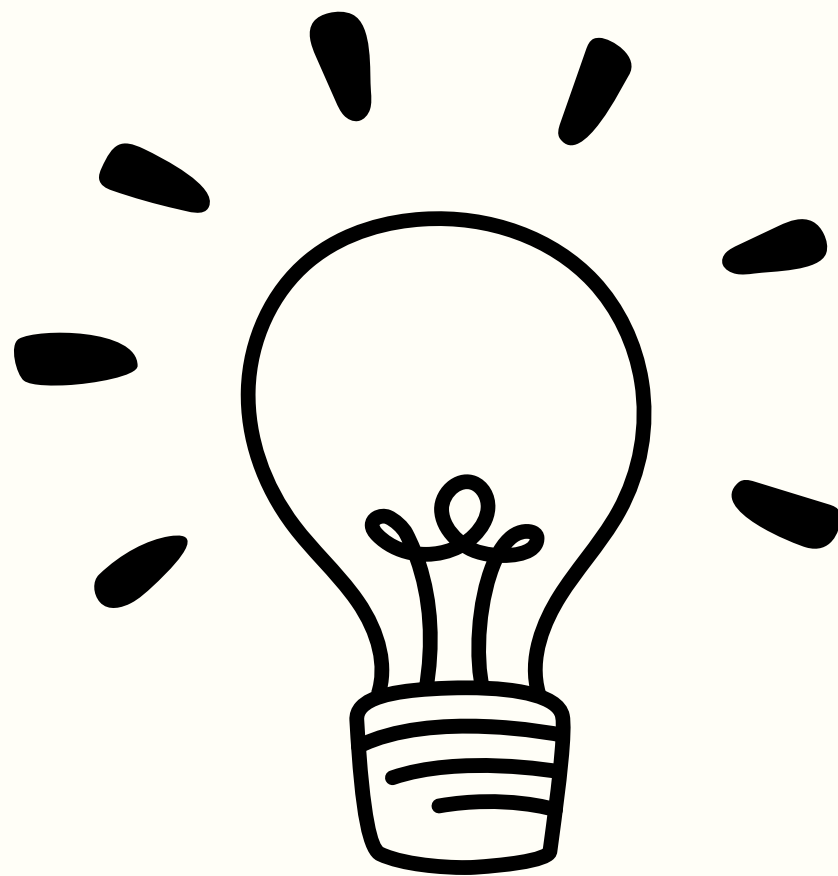
- "position_encoded", "goals conceded", "age", "height", "second yellow cards", "yellow cards", "clean sheets", "red cards" y "winger" tienen correlaciones bajas o nulas con "current_value".

VISUALIZACIÓN MEDIA DEL VALOR DE TRANSFERENCIA Y POSICIÓN DE LOS JUGADORES



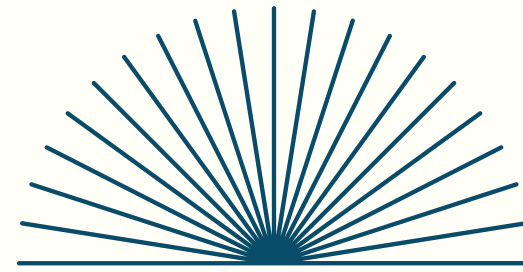
En base a estos datos, se puede extraer que:

- Existen diferencias entre el valor de transferencia de los jugadores de fútbol según su posición en el campo. Los atacantes (valor 4) presentan los valores más altos, tanto en el promedio actual como en el promedio más alto. Por otro lado, los arqueros (valor 1) tienden a tener los precios más bajos en ambas categorías.
- Esta diferencia está asociada a un aumento progresivo en los precios promedio desde las posiciones de arquero hasta las posiciones de ataque. Esta tendencia sugiere una correlación entre la posición de juego y el valor de mercado de los jugadores.



DESAFÍOS

- Incorporar más ligas de fútbol.
- Incorporar más estadísticas profesionales:
 - Convocatorias a la selección nacional.
 - Partidos de copas internacionales (Champions League, Europa League, entre otros).
- Incorporar más temporadas de fútbol.
- Incorporar valores de ventas anteriores.



Muchas gracias