



PREDICCIÓN VALORES DE TRANSFERENCIA DE JUGADORES DE FÚTBOL

Proyecto final Coding Dojo - Data Science

Matías Vidal Céspedes



PROBLEMÁTICA

**FACTORES
PERSONALES**



**FACTORES
MERCADO**



INCERTIDUMBRE

**FACTORES
EXTERNOS**

**FACTORES
EQUIPO**

ANTECEDENTES DEL CONJUNTO DE DATOS



**Información
básica
de los jugadores
de fútbol**



**Estadísticas
profesionales
2021-2023**



**Premios
individuales y
por equipo**

ANTECEDENTES DEL CONJUNTO DE DATOS

11 LIGAS EUROPEAS

4 LIGAS AMERICANAS

1 LIGA AFRICANA

4 LIGAS ASIATICAS

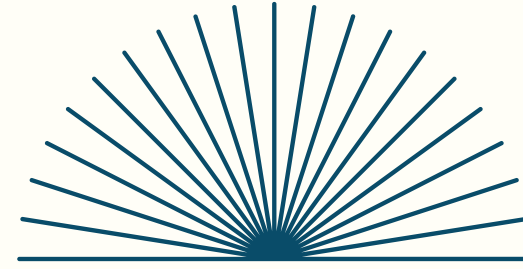


DATAFRAME

10.754 filas

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

Valores faltantes

Inconsistencias

LIMPIEZA DE LOS DATOS



Columnas innecesarias

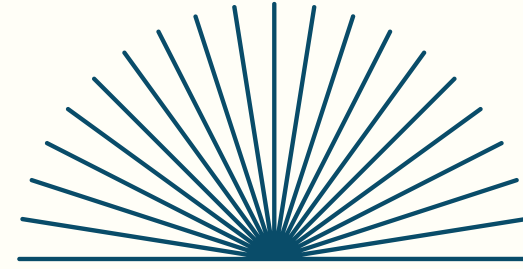
Se identificaron que para efectos de la predicción del valor de transferencia de un futbolista no es necesario contar con las siguientes columnas.

PLAYER

TEAM

NAME

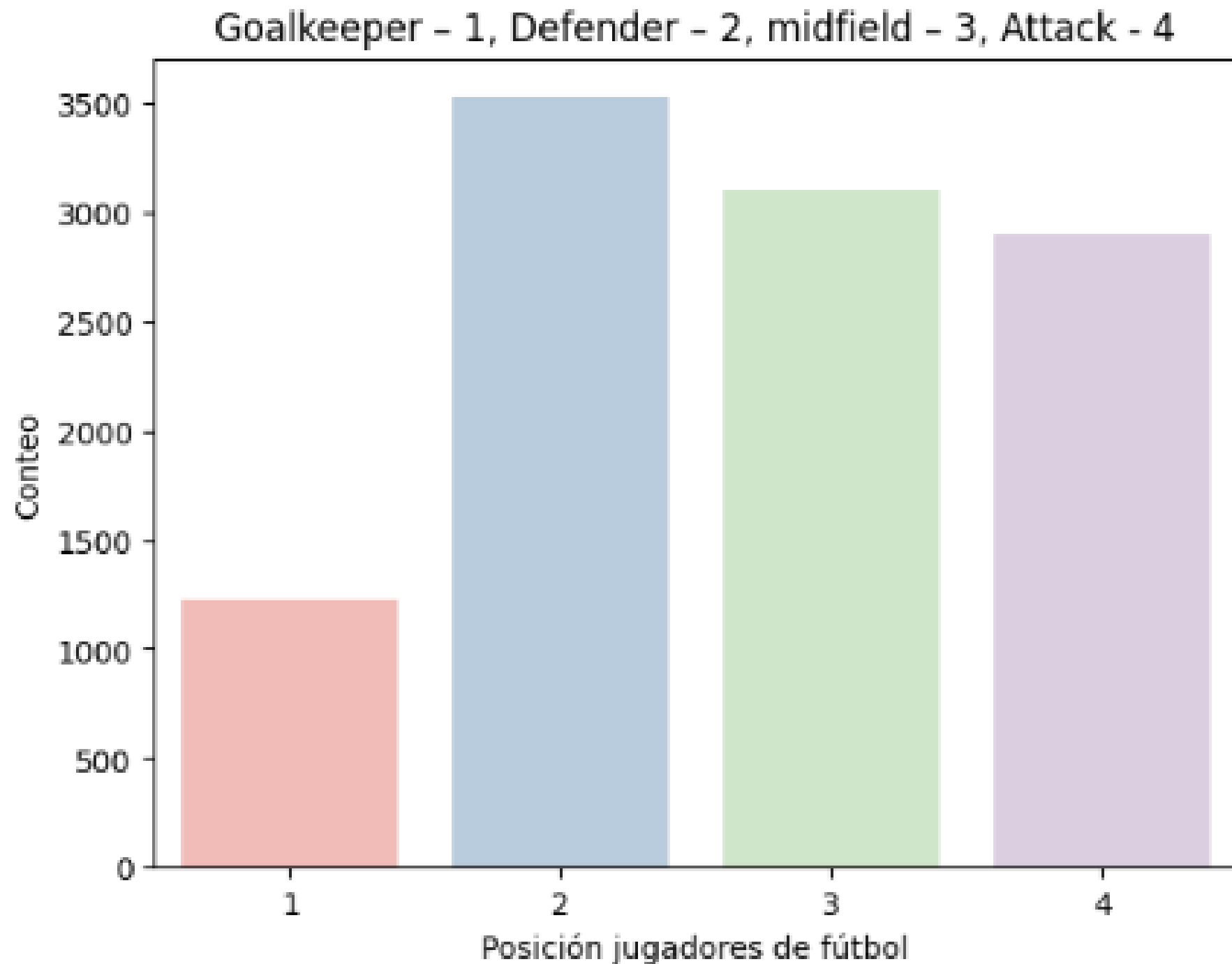
POSITION



EXPLORACIÓN DE LOS DATOS



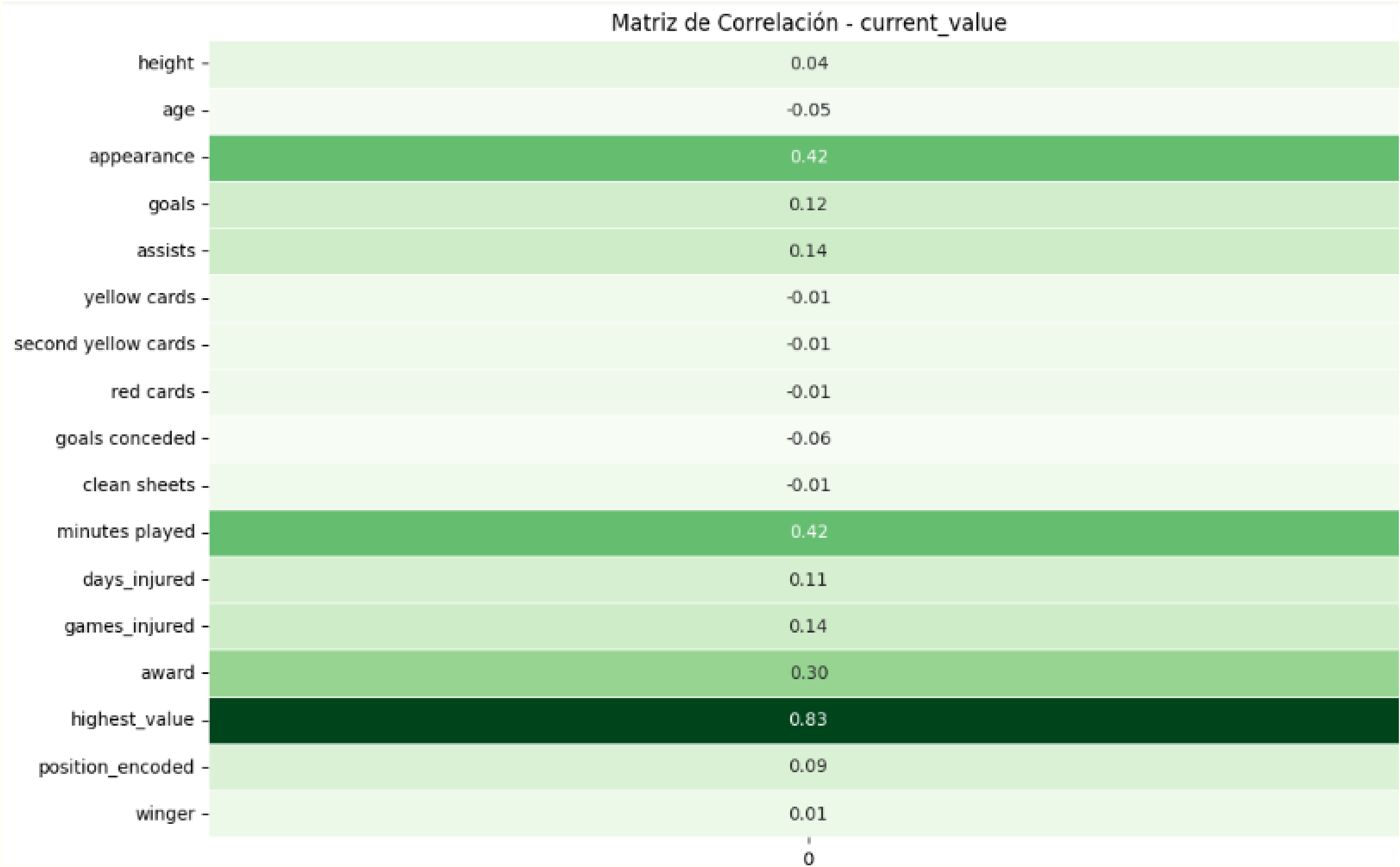
VISUALIZACIÓN POSICIONES DE LOS JUGADORES DE FÚTBOL



Disparidad en la distribución

- Arqueros 11,4%
- Defensas 32,8%
- Mediocampistas 28,8%
- Delanteros 27,0%

VISUALIZACIÓN DE LA CORRELACIÓN DE LAS VARIABLES



Alta correlación:

- "highest_value"

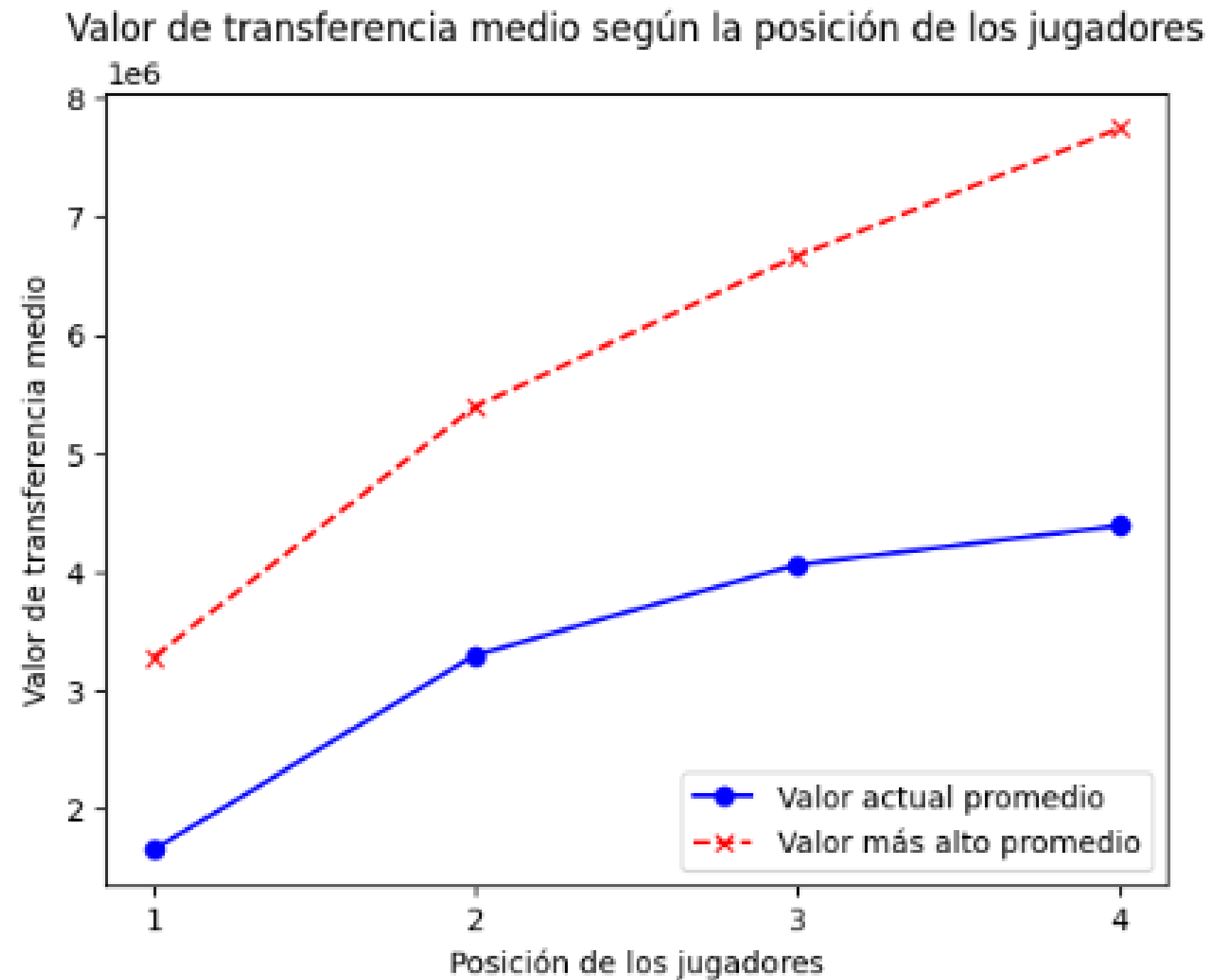
Moderada correlación:

- "minutes played"
- "appearance"
- "award"

Débil correlación:

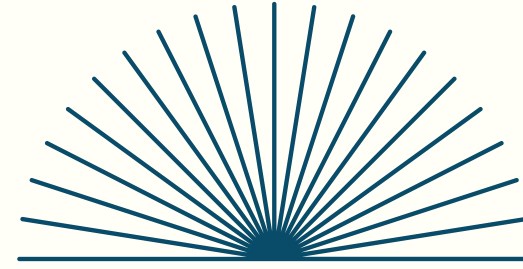
- "assists"
- "games_injured"
- "goals"
- "days_injured"

VISUALIZACIÓN MEDIA DEL VALOR DE TRANSFERENCIA Y POSICIÓN DE LOS JUGADORES



Diferencia valor de transferencia según posición

Relación directa entre la posición del jugador y el valor promedio de transferencia.

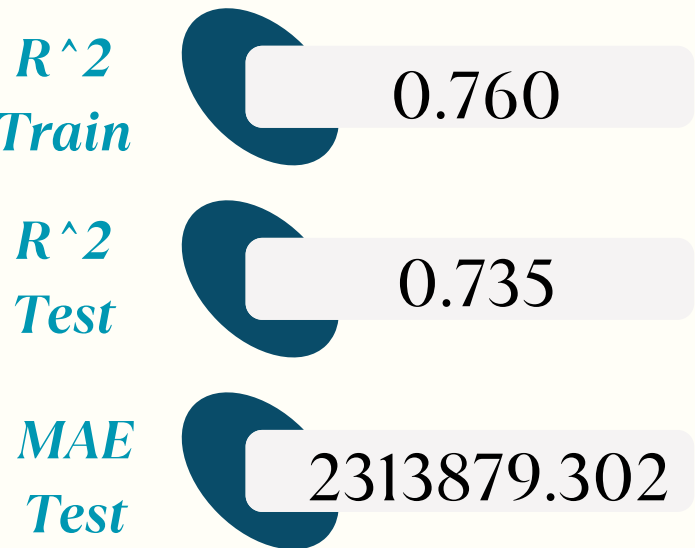


MODELOS DE PREDICCIÓN

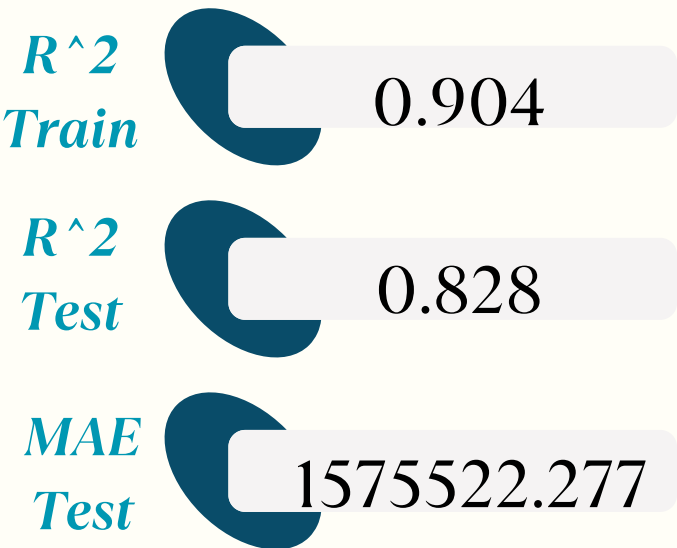


MODELOS DE REGRESIÓN - RESULTADOS

Regresión lineal



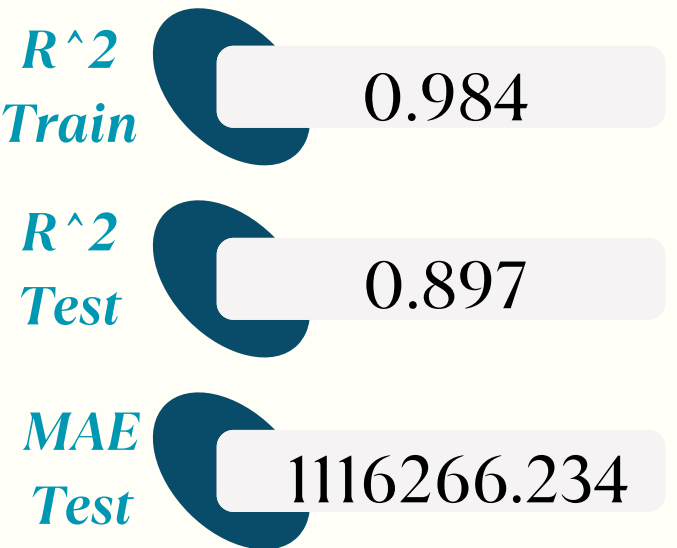
Decisión Tree



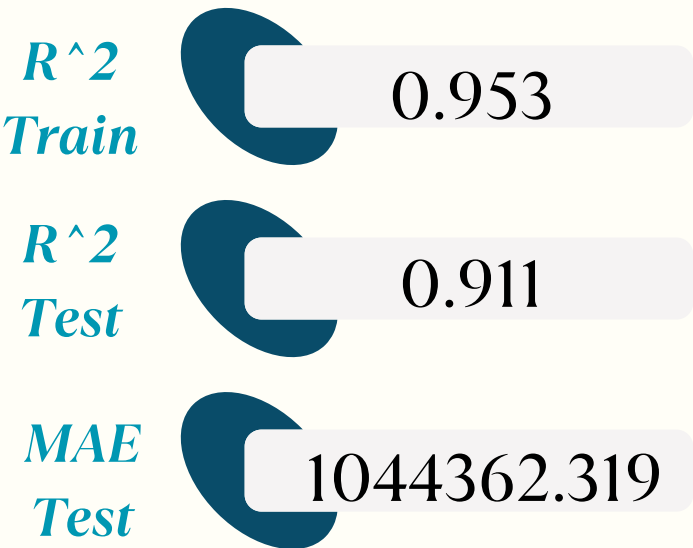
Bagged Tree



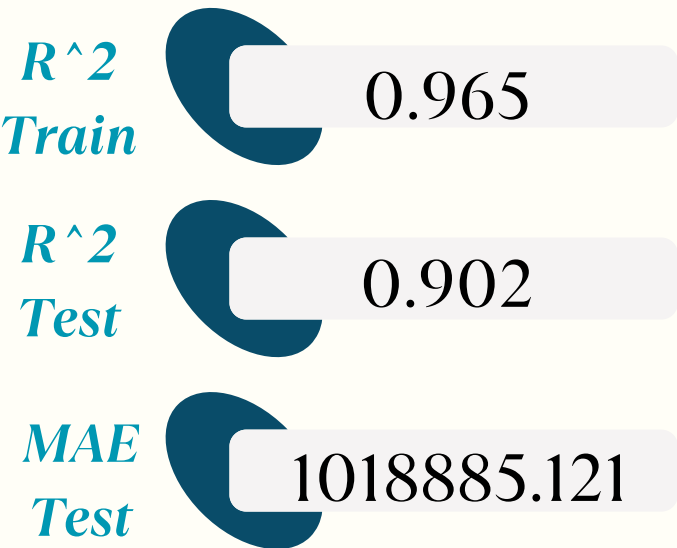
Random Forest



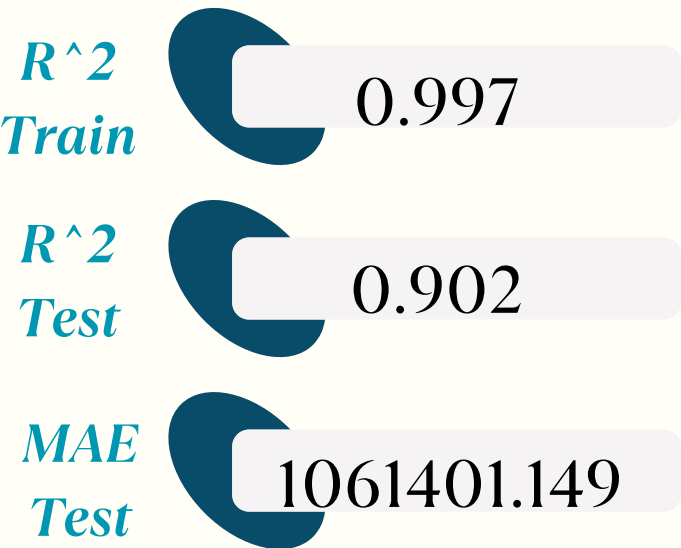
Gradient Boosting

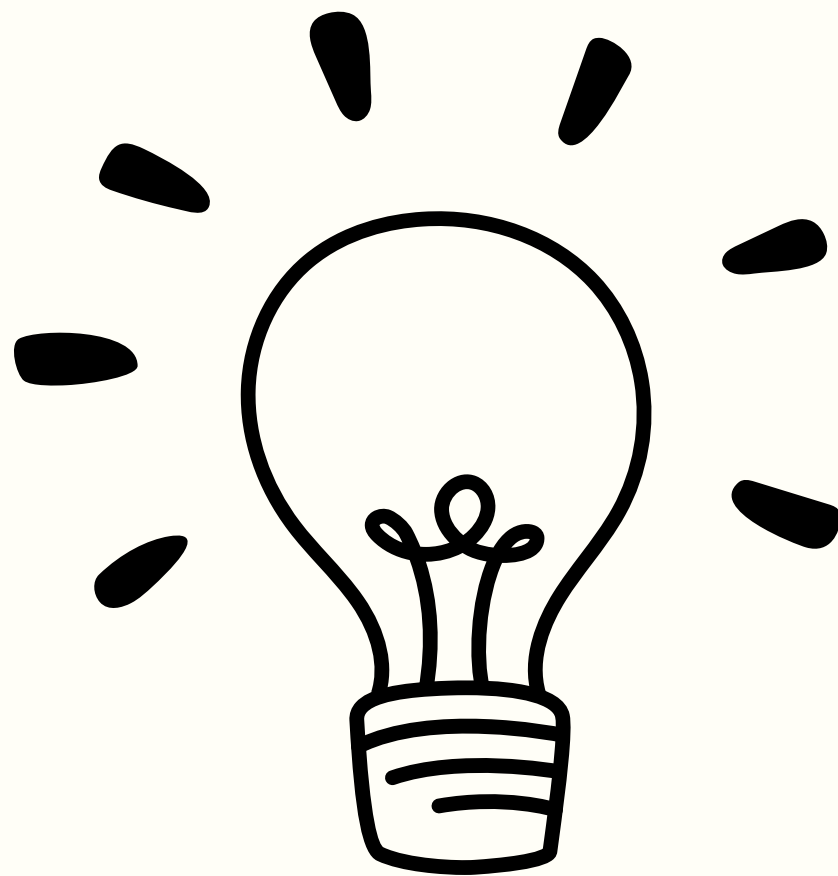


LightGBM



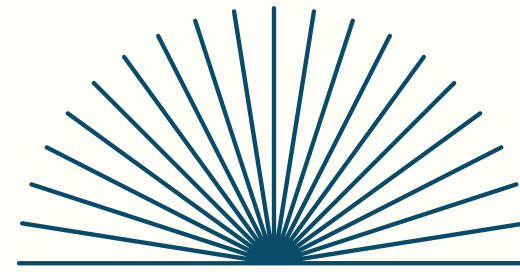
XGBoost





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.
- Asignar puntaje a los equipos de fútbol (Ranking FIFA).



Muchas gracias