

Inteligencia Artificial I

PREDICCIÓN DEL VALOR DE LOS JUGADORES DE LA RFEF

David Anaya - 2211714
Yiber Romero - 2221835
Cesar Vanegas - 2220040

Introducción

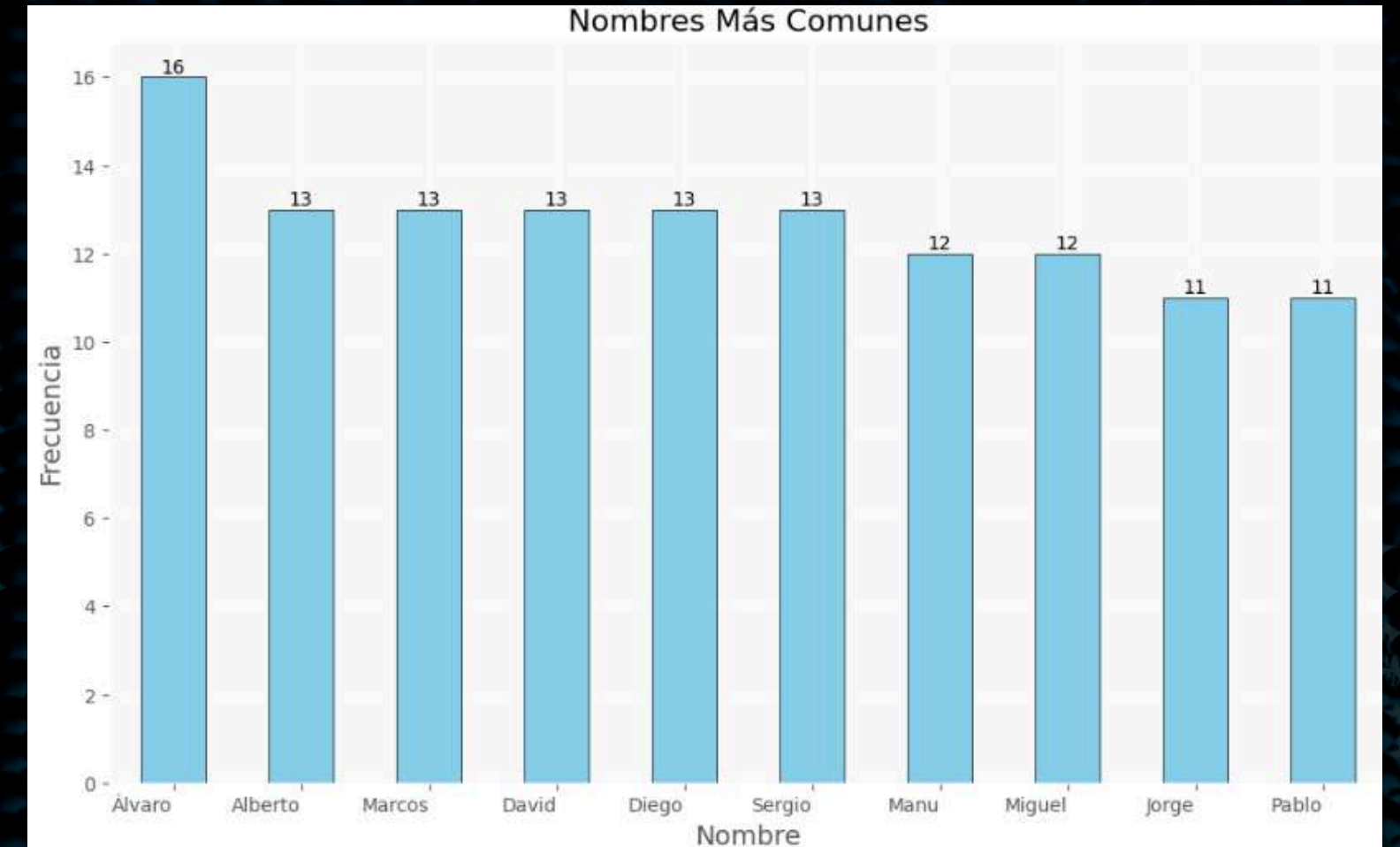
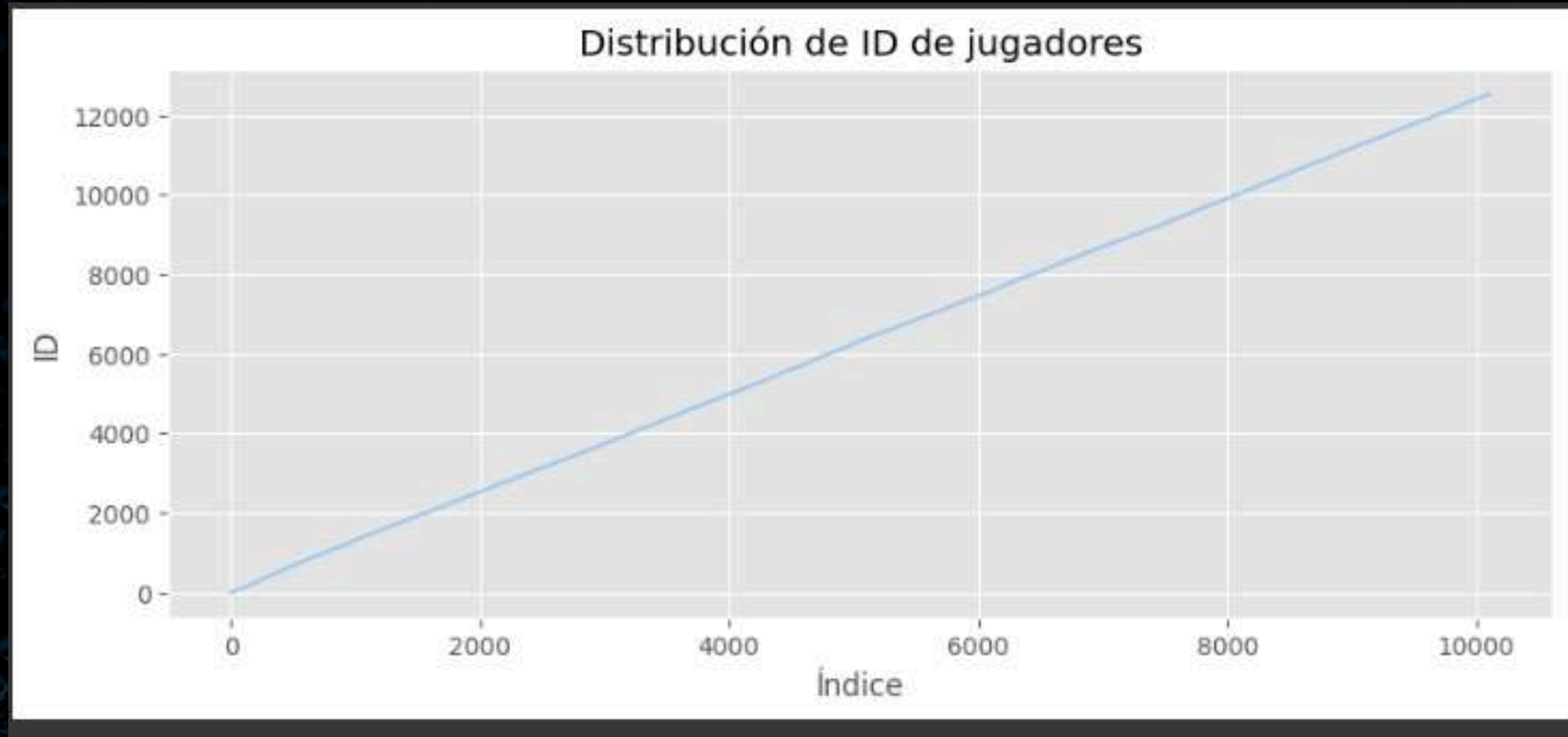


El dataset utilizado es sobre la Real Federación Española de Fútbol (RFEF) en la temporada de 2023-2024 gracias al esfuerzo de Isabelo Castillo. Dicho conjunto de datos recopila información sobre el desempeño de los jugadores de las categorías regionales de España, abarcando aspectos como los partidos, titularidades, goles, tarjetas que ha recibido, posición en el campo, entre otros.

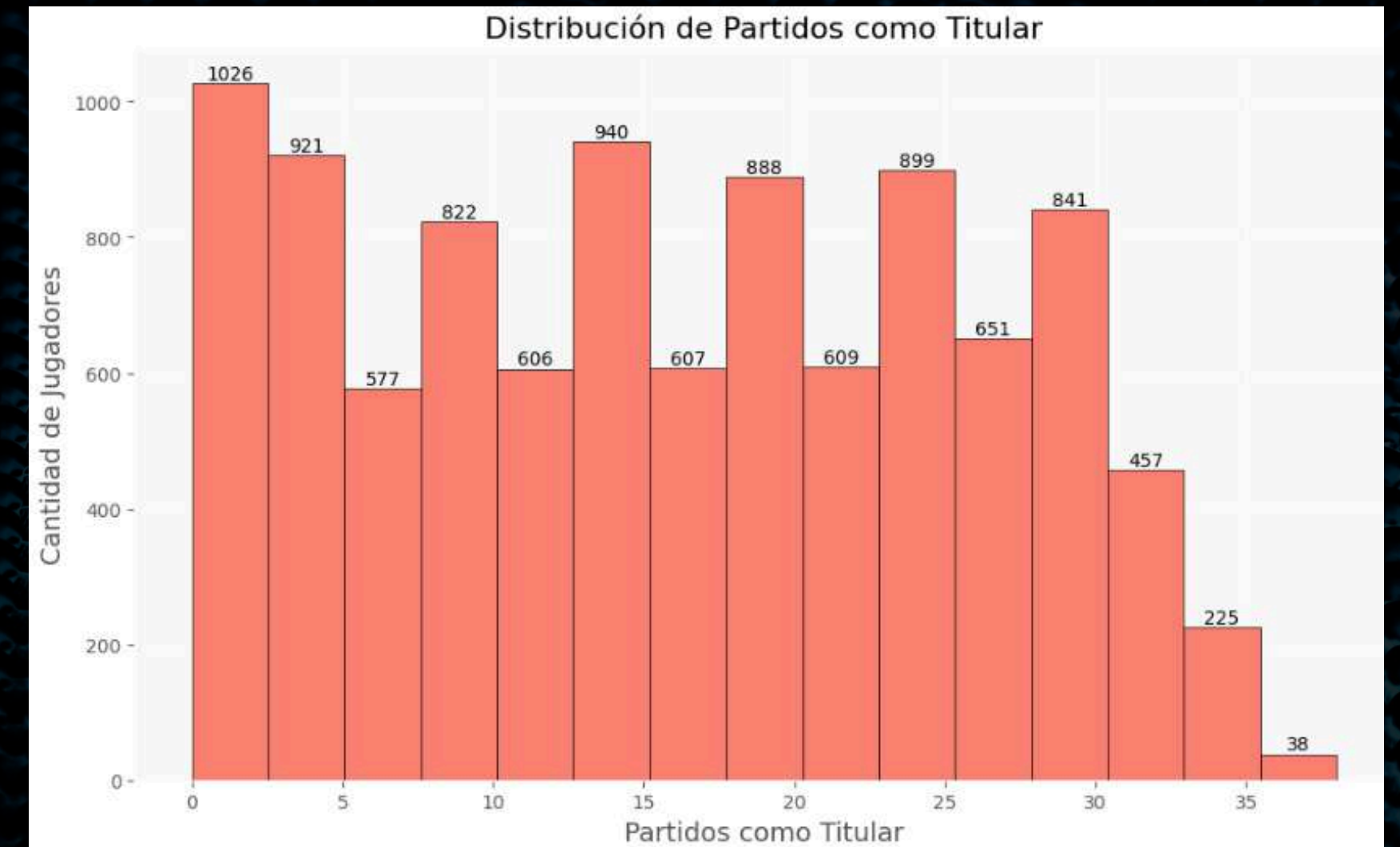
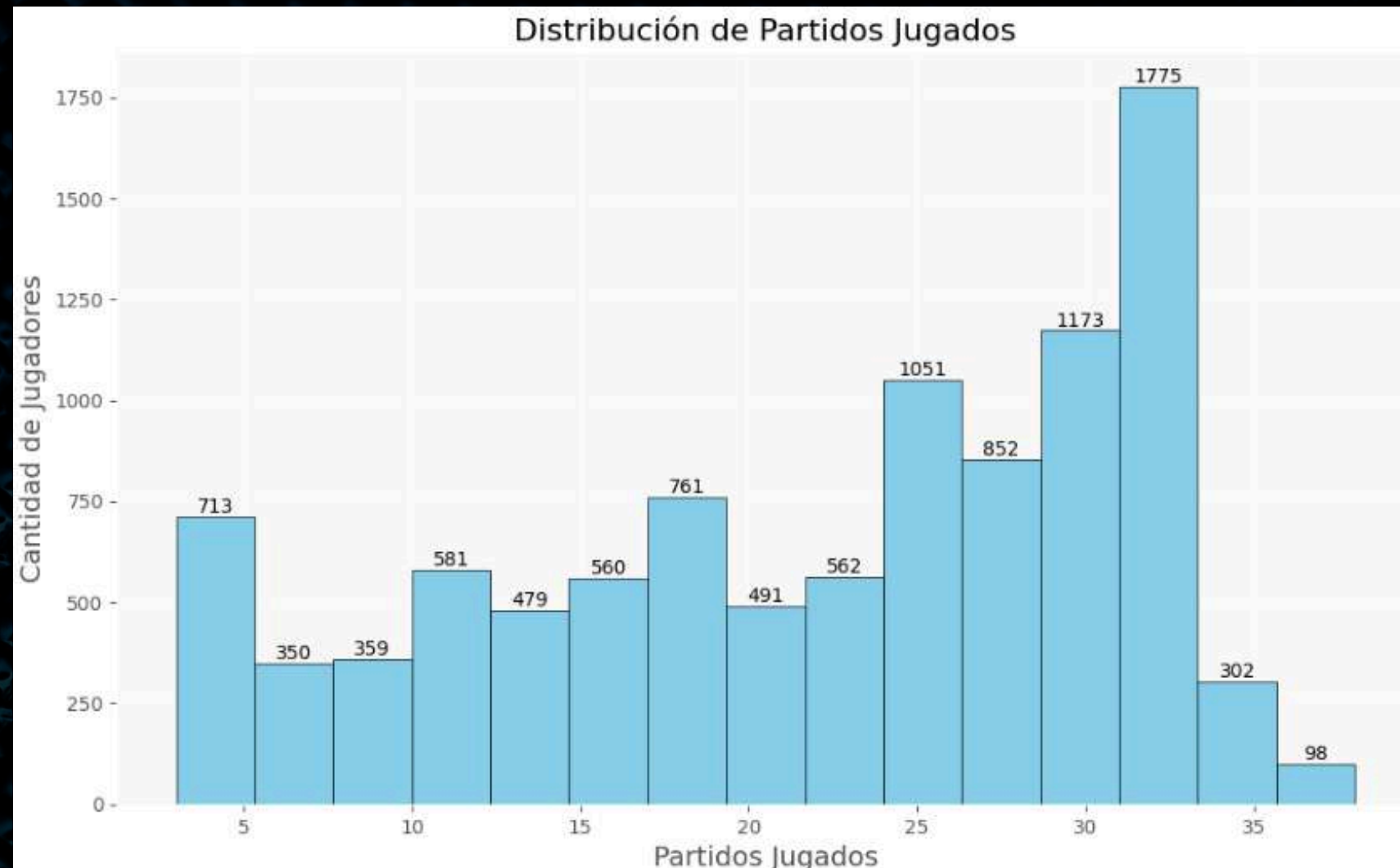
Analisis del Dataset

----- Display data -----													
	id	nombre	partidos_jugados	partidos_titular	goles	asistencias	tarjeta_amarilla	edad	altura	valor	rating	posicion	
0	0	Alberto Varo	34	34	21	1	2	31	191.000000	174640.0	59	Portero	
1	2	Dani Parra	5	4	3	0	1	24	188.000000	91220.0	44	Portero	
2	4	Joan Oriol	37	36	1	3	9	37	175.000000	119000.0	62	Lateral Izquierdo	
3	5	P. Trigueros	37	37	5	1	6	31	188.000000	217500.0	57	Defensa Central	
4	6	Nacho González	31	31	2	0	8	29	185.000000	340340.0	61	Defensa Central	
5	7	Pol Domingo	29	28	0	0	5	24	180.000000	392580.0	57	Lateral Derecho	
6	8	Unai Dufur	18	7	0	0	0	25	187.000000	310340.0	54	Defensa Central	

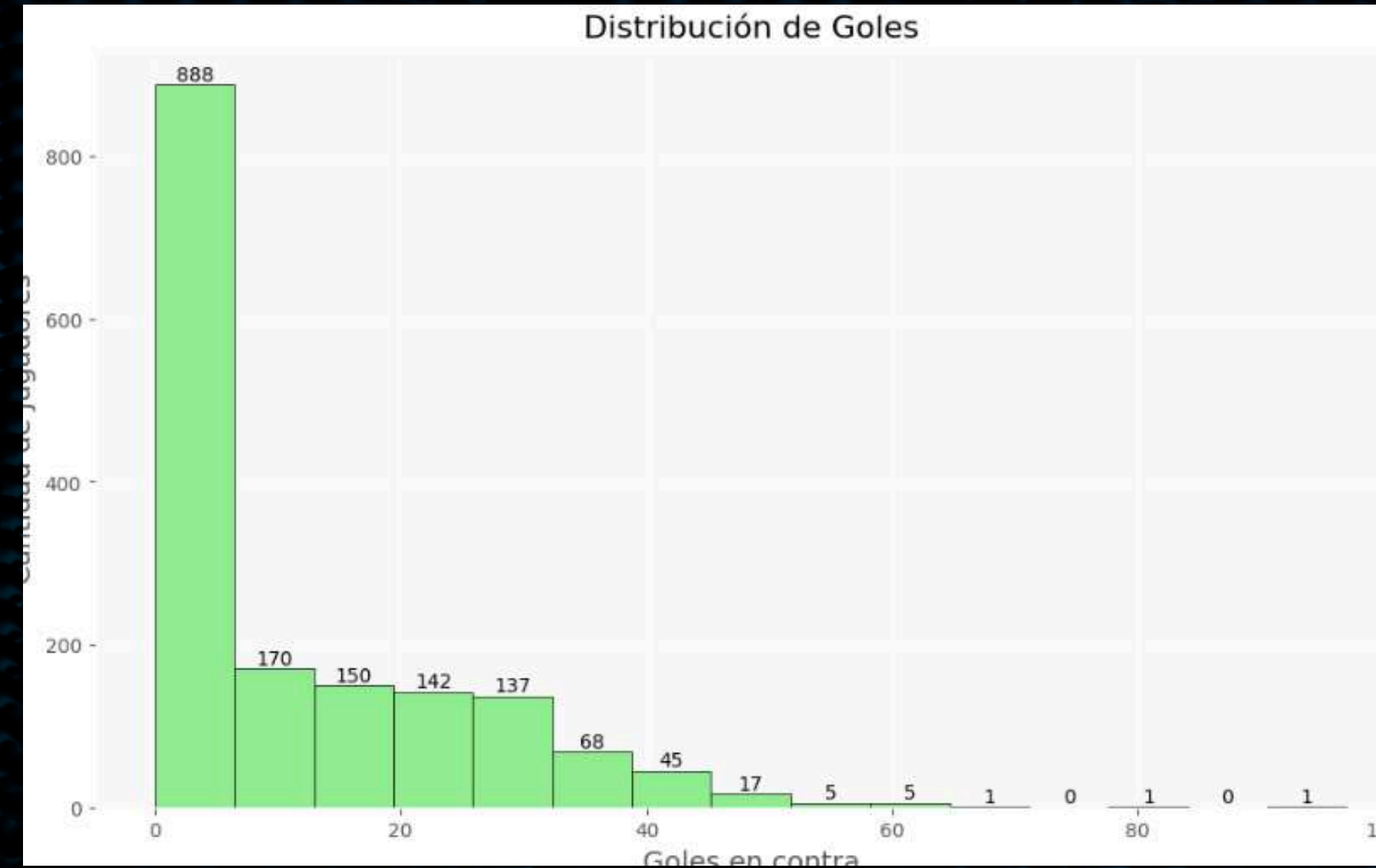
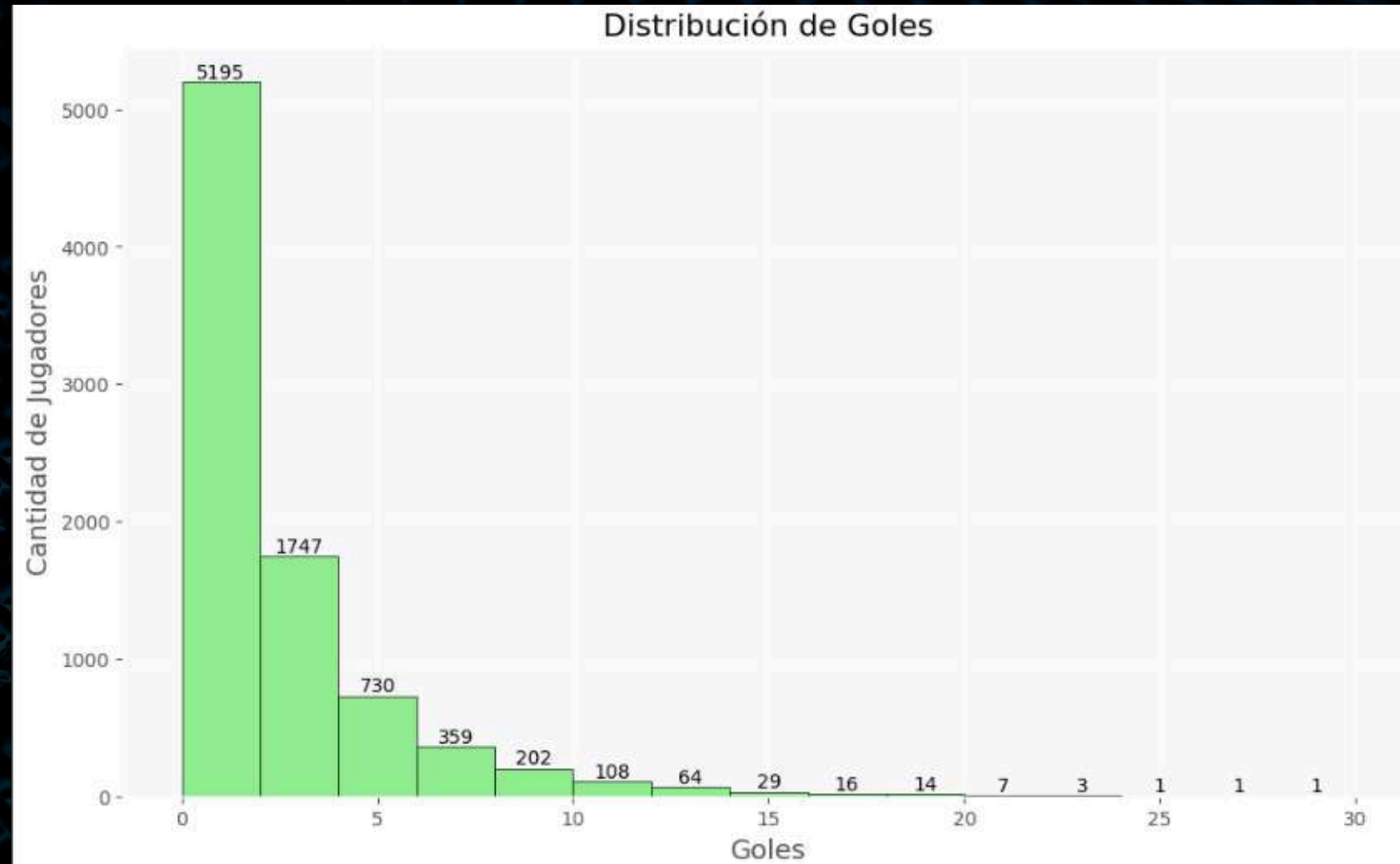
Analisis del Dataset



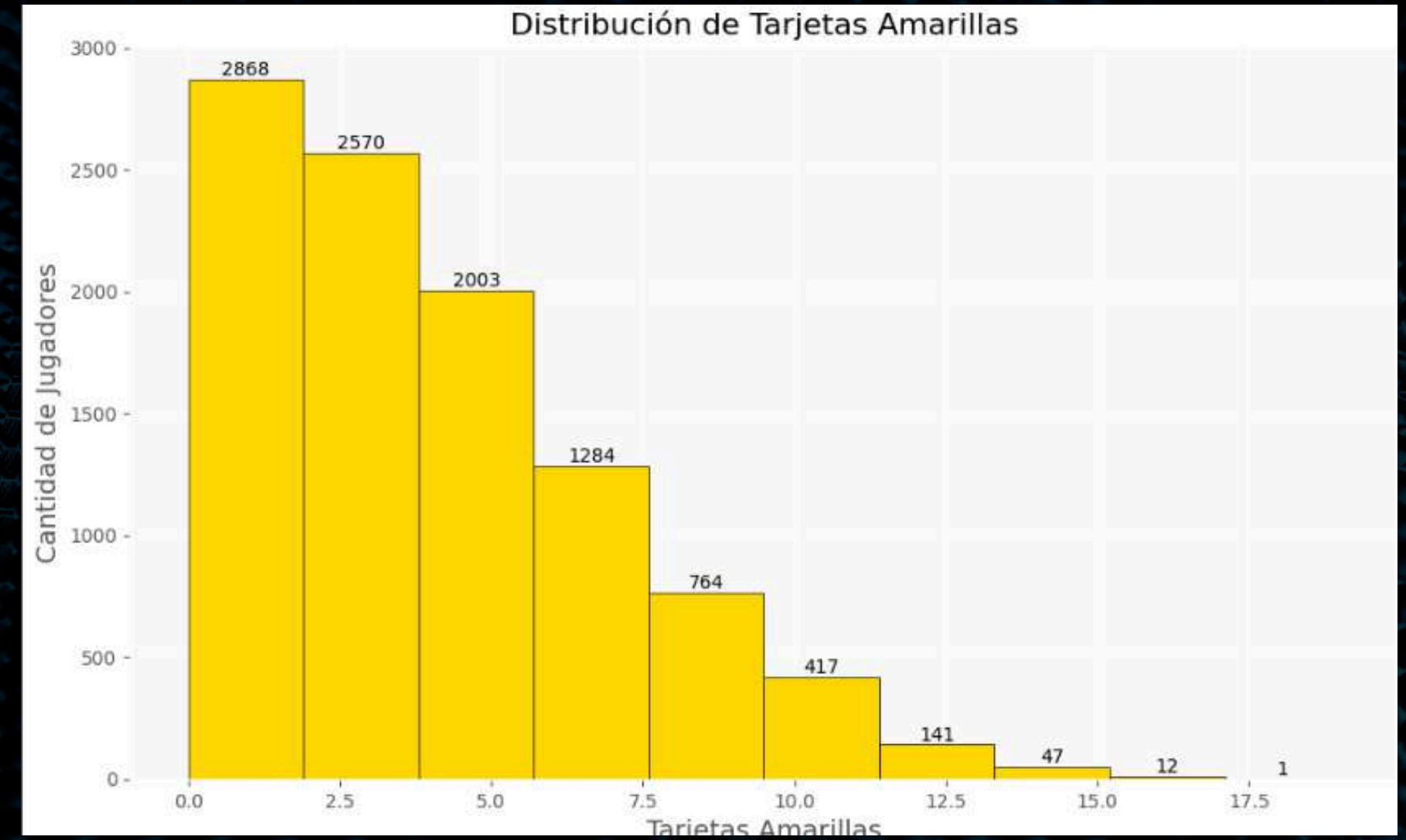
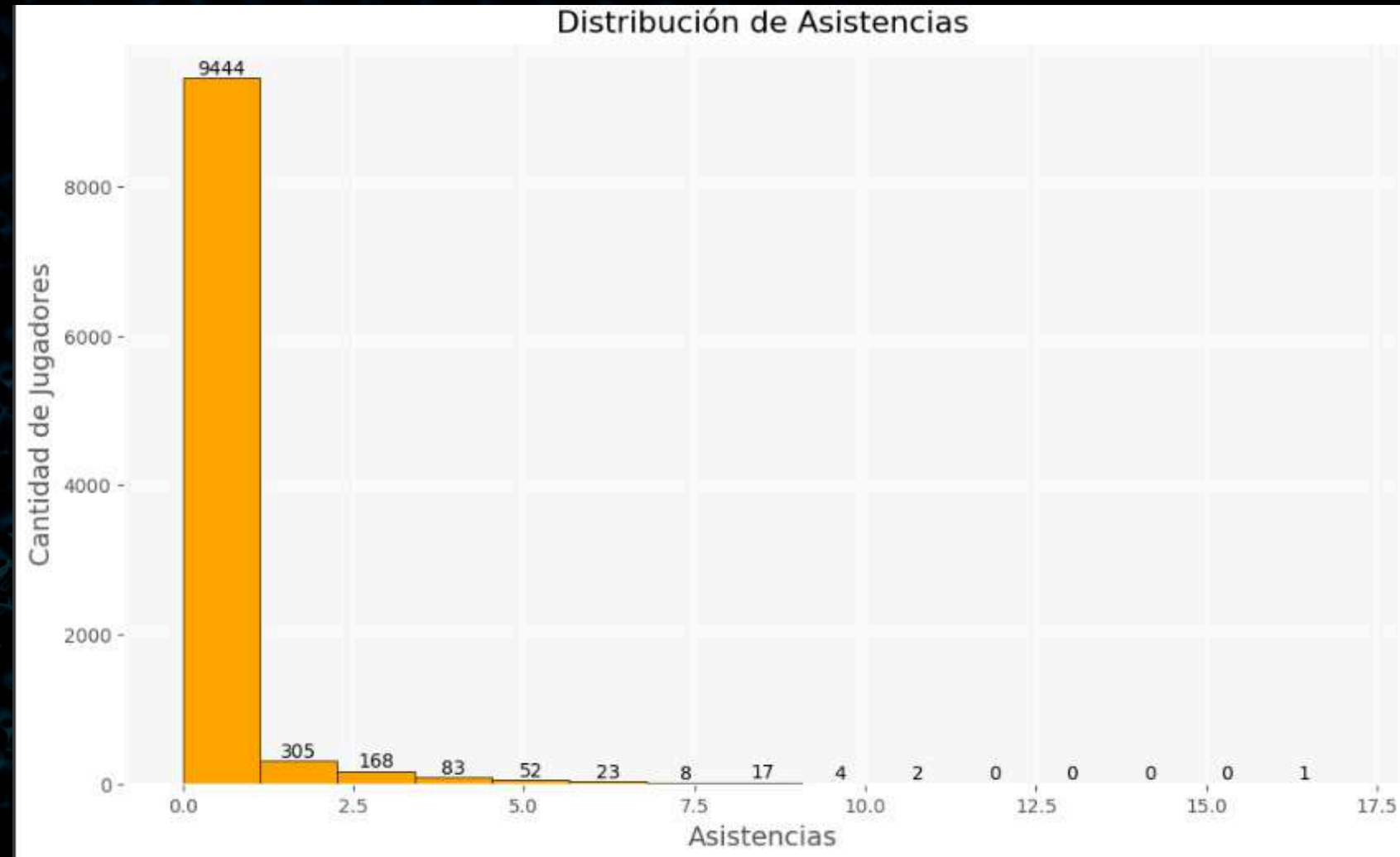
Analisis del Dataset



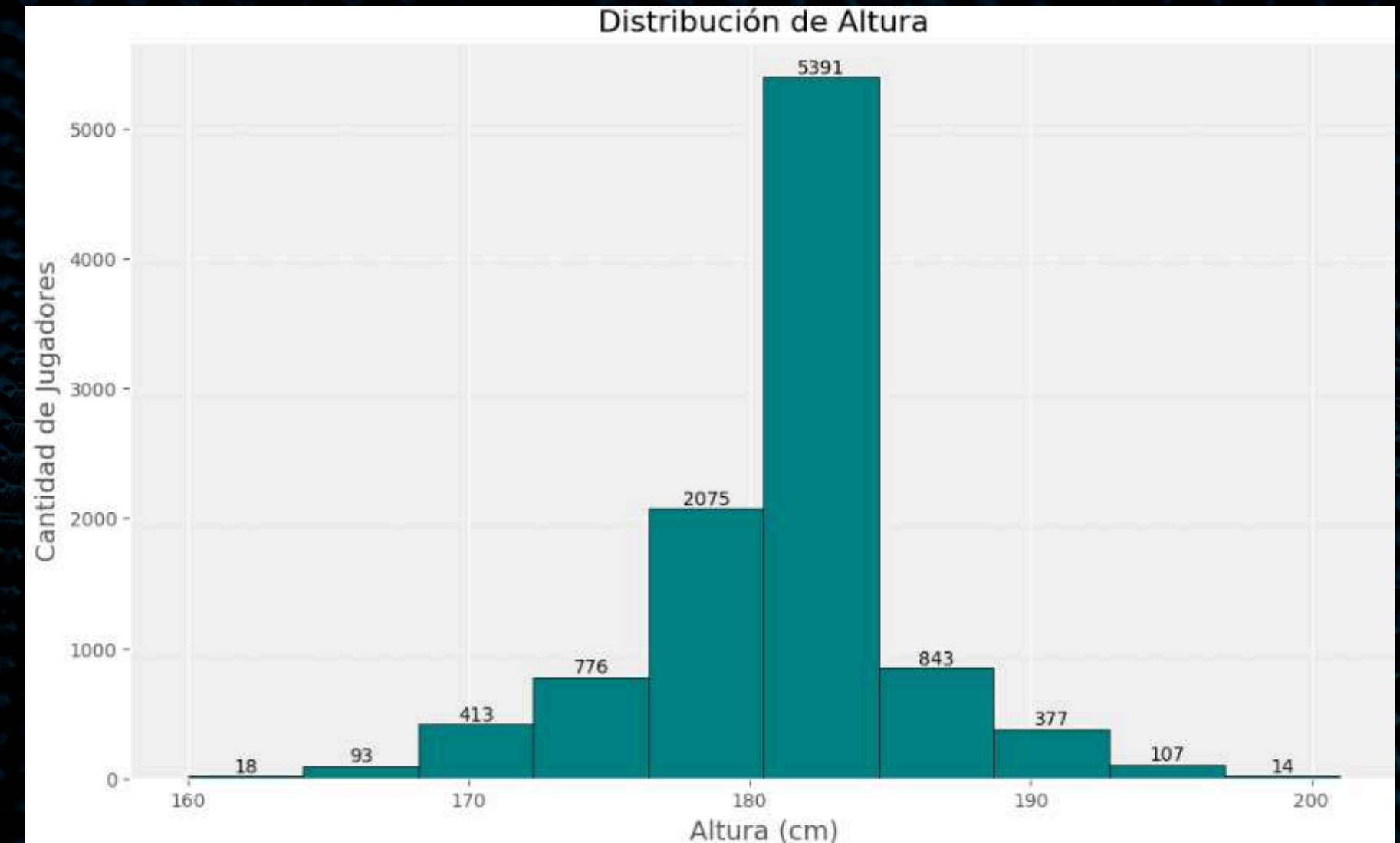
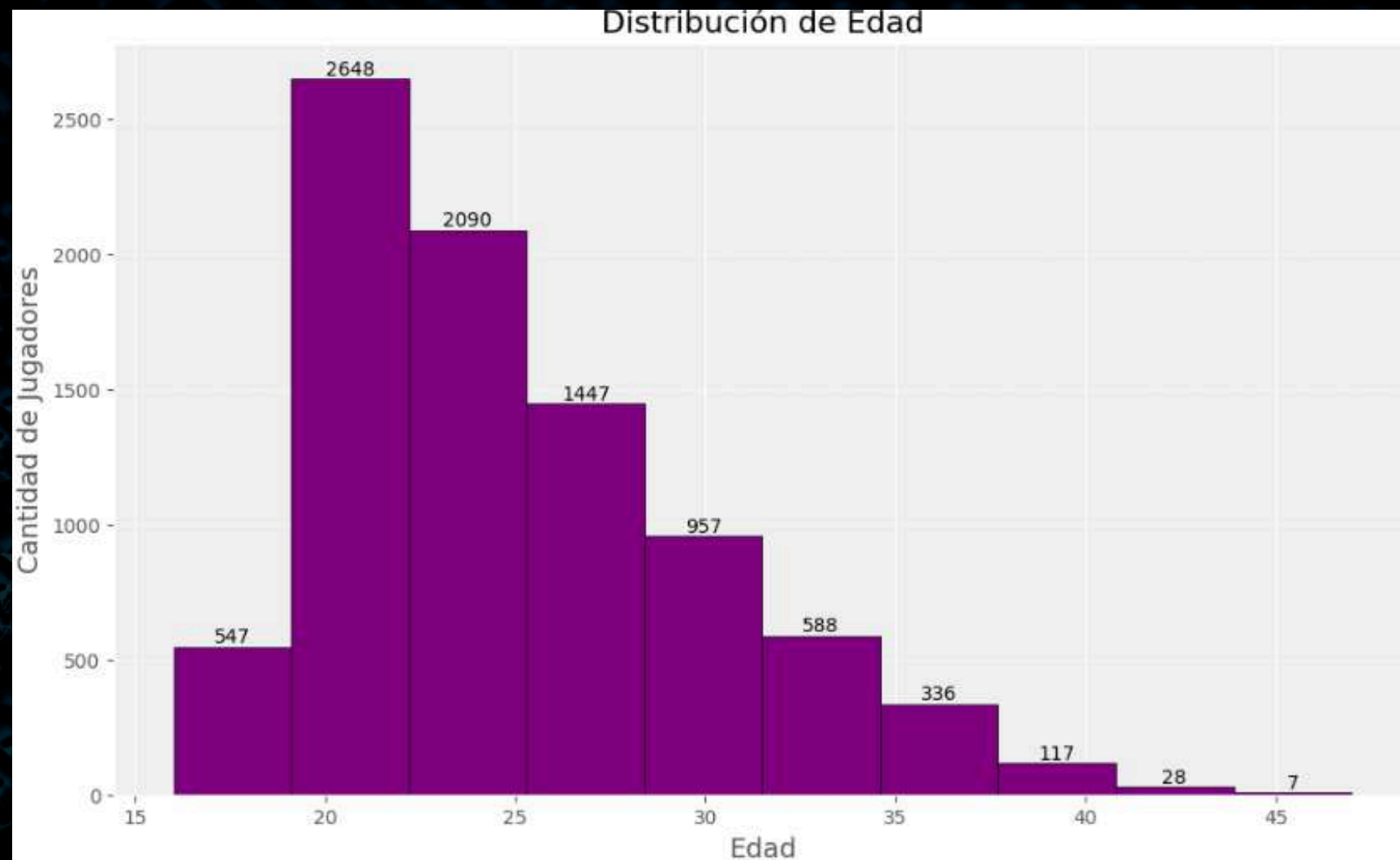
Analisis del Dataset



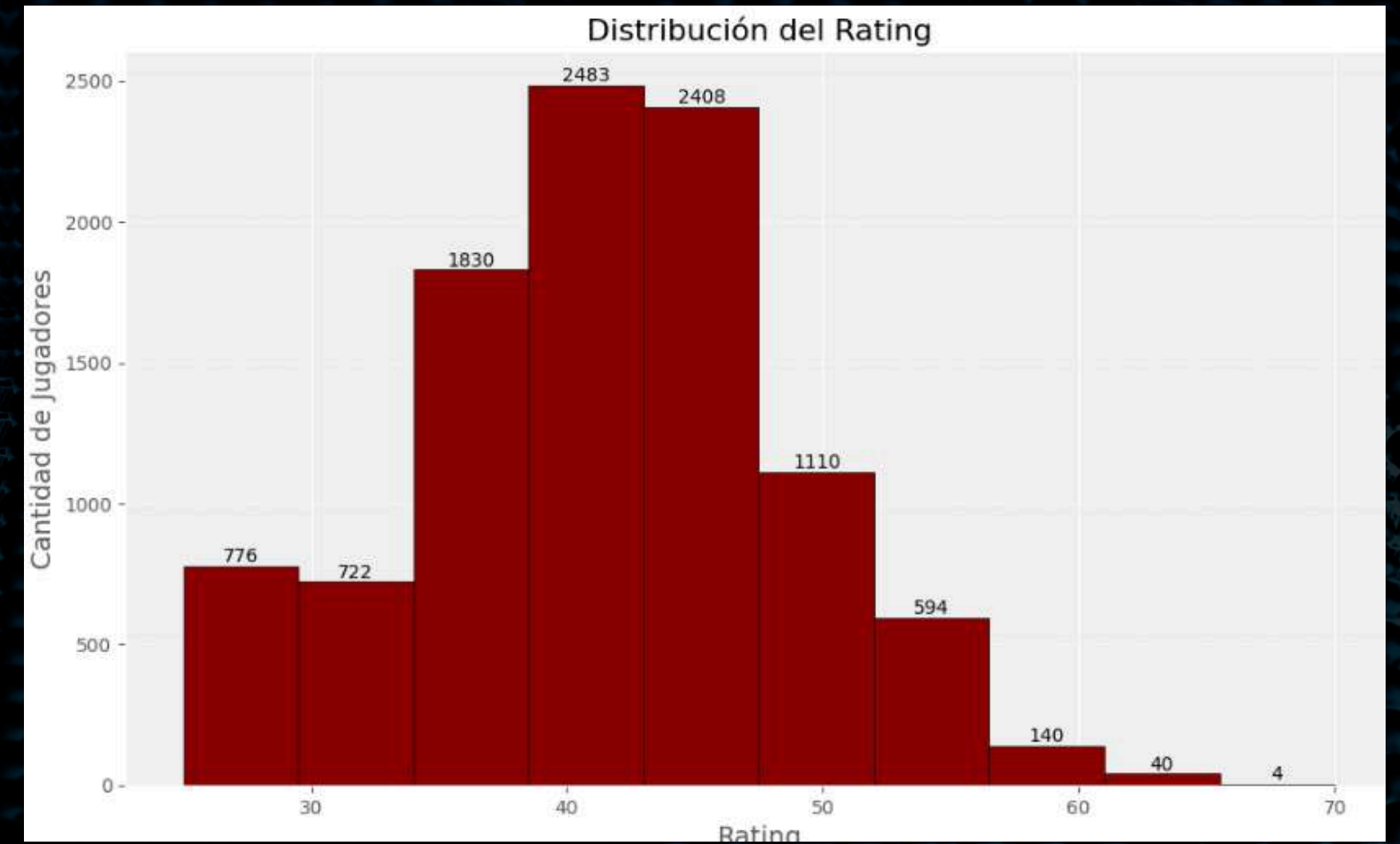
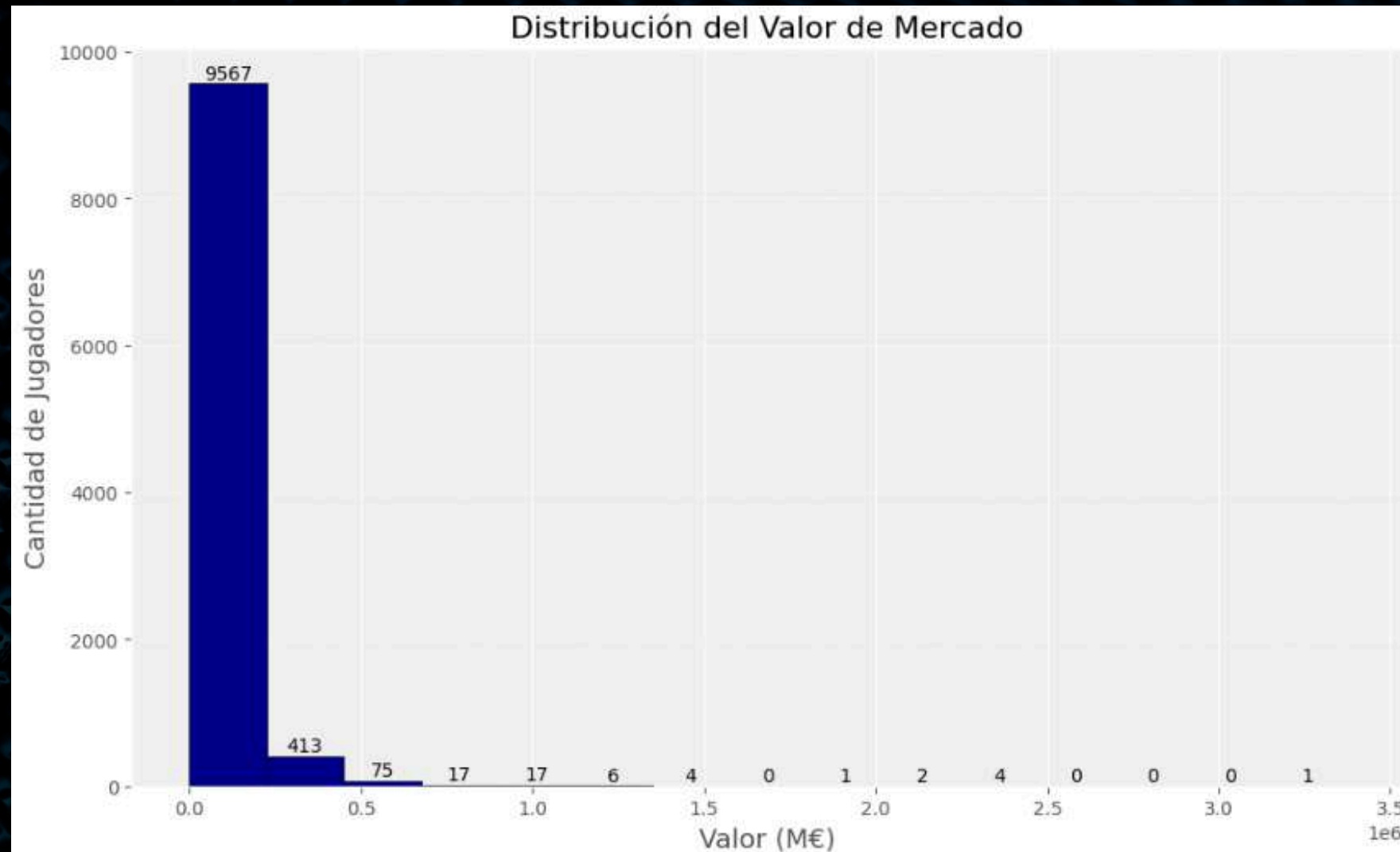
Analisis del Dataset



Analisis del Dataset

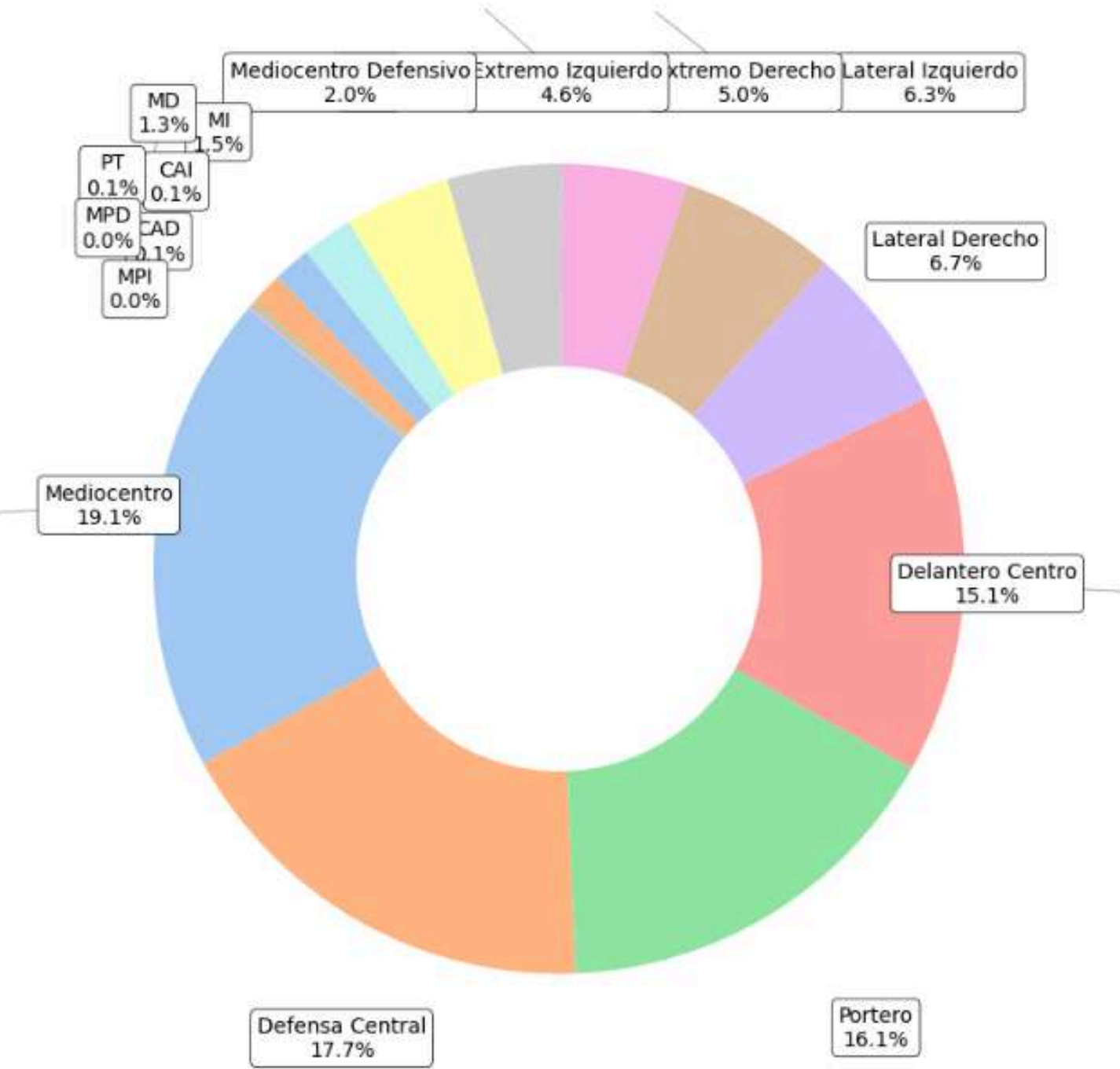


Analisis del Dataset



Analisis del Dataset

Distribución de Posiciones (etiquetas externas + tabla de leyenda)



Posición	Cantidad	Porcentaje
Mediocentro	1929	19.1%
Defensa Central	1788	17.7%
Portero	1630	16.1%
Delantero Centro	1530	15.1%
Lateral Derecho	673	6.7%
Lateral Izquierdo	636	6.3%
Extremo Derecho	503	5.0%
Extremo Izquierdo	464	4.6%
MP	426	4.2%
Mediocentro Defensivo	207	2.0%
MI	152	1.5%
MD	136	1.3%
CAI	14	0.1%
CAD	10	0.1%
PT	6	0.1%
MPI	2	0.0%
MPD	1	0.0%

Análisis de regresión para los jugadores de la RFEF

Este modelo permite predecir el valor de mercado de jugadores de fútbol en función de variables como goles, edad, altura, partidos jugados, asistencias y tarjetas amarillas.

Utilizando árboles de decisión, evaluamos diferentes niveles de complejidad para encontrar el equilibrio ideal entre precisión y sobreajuste.

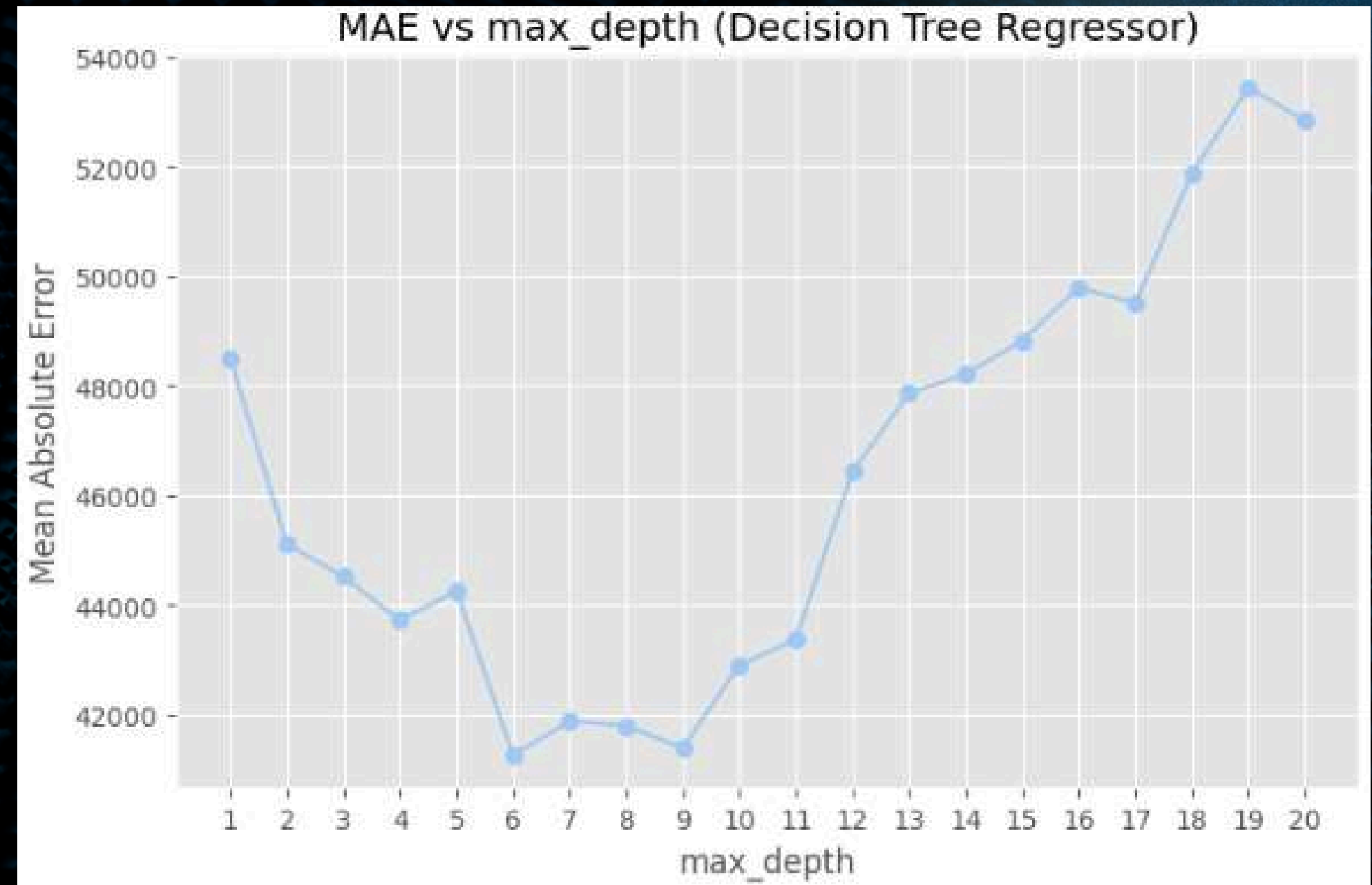
Mediante la métrica de error absoluto medio (MAE), identificamos qué tan buenas son las predicciones del modelo. Además, la visualización del MAE frente a la profundidad del árbol nos ayuda a seleccionar el mejor modelo posible.



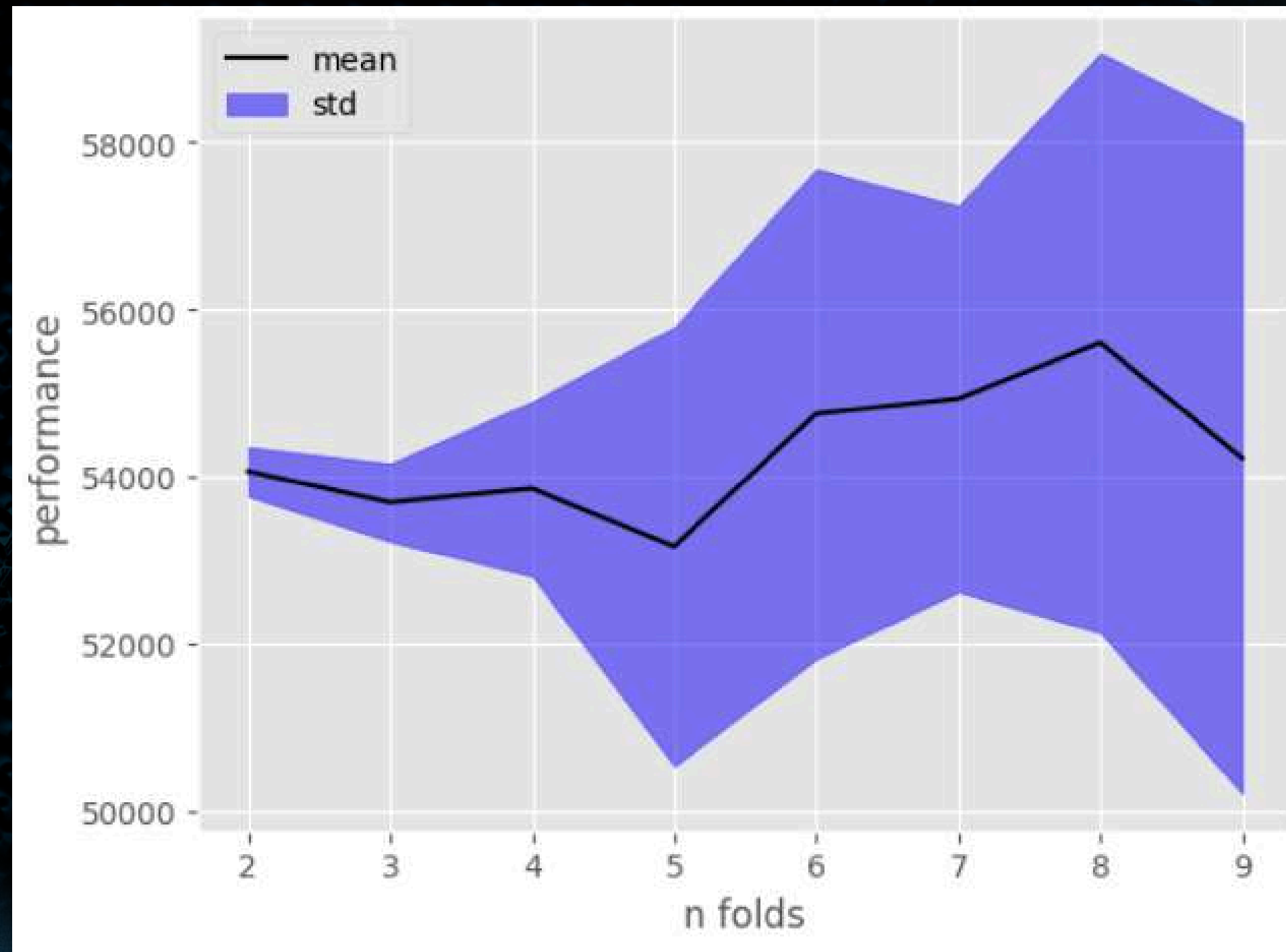


```
MAE for max_depth 1: 48500.15916793943
Relative MAE for max_depth 1: 1.07
MAE for max_depth 2: 45109.12234659425
Relative MAE for max_depth 2: 0.99
MAE for max_depth 3: 44515.35610367874
Relative MAE for max_depth 3: 0.98
MAE for max_depth 4: 43716.73747354037
Relative MAE for max_depth 4: 0.96
MAE for max_depth 5: 44250.00035937916
Relative MAE for max_depth 5: 0.97
MAE for max_depth 6: 41258.61768905903
Relative MAE for max_depth 6: 0.91
MAE for max_depth 7: 41878.7208796056
Relative MAE for max_depth 7: 0.92
MAE for max_depth 8: 41794.45640923411
Relative MAE for max_depth 8: 0.92
MAE for max_depth 9: 41385.470943065484
Relative MAE for max_depth 9: 0.91
MAE for max_depth 10: 42895.79162914715
Relative MAE for max_depth 10: 0.94
MAE for max_depth 11: 43370.75091638737
Relative MAE for max_depth 11: 0.95
MAE for max_depth 12: 46433.529664969494
Relative MAE for max_depth 12: 1.02
MAE for max_depth 13: 47856.35253479052
Relative MAE for max_depth 13: 1.05
MAE for max_depth 14: 48206.42767896451
Relative MAE for max_depth 14: 1.06
MAE for max_depth 15: 48814.358284555085
Relative MAE for max_depth 15: 1.07
MAE for max_depth 16: 49791.66507888634
Relative MAE for max_depth 16: 1.1
MAE for max_depth 17: 49493.45994641521
Relative MAE for max_depth 17: 1.09
MAE for max_depth 18: 51869.77218965484
Relative MAE for max_depth 18: 1.14
MAE for max_depth 19: 53435.71432125372
Relative MAE for max_depth 19: 1.18
MAE for max_depth 20: 52833.46453385828
Relative MAE for max_depth 20: 1.16
Best Relative MAE: 0.91
```

MAE vs Max Depth for DT



Best n-fold for CV



Random Forest

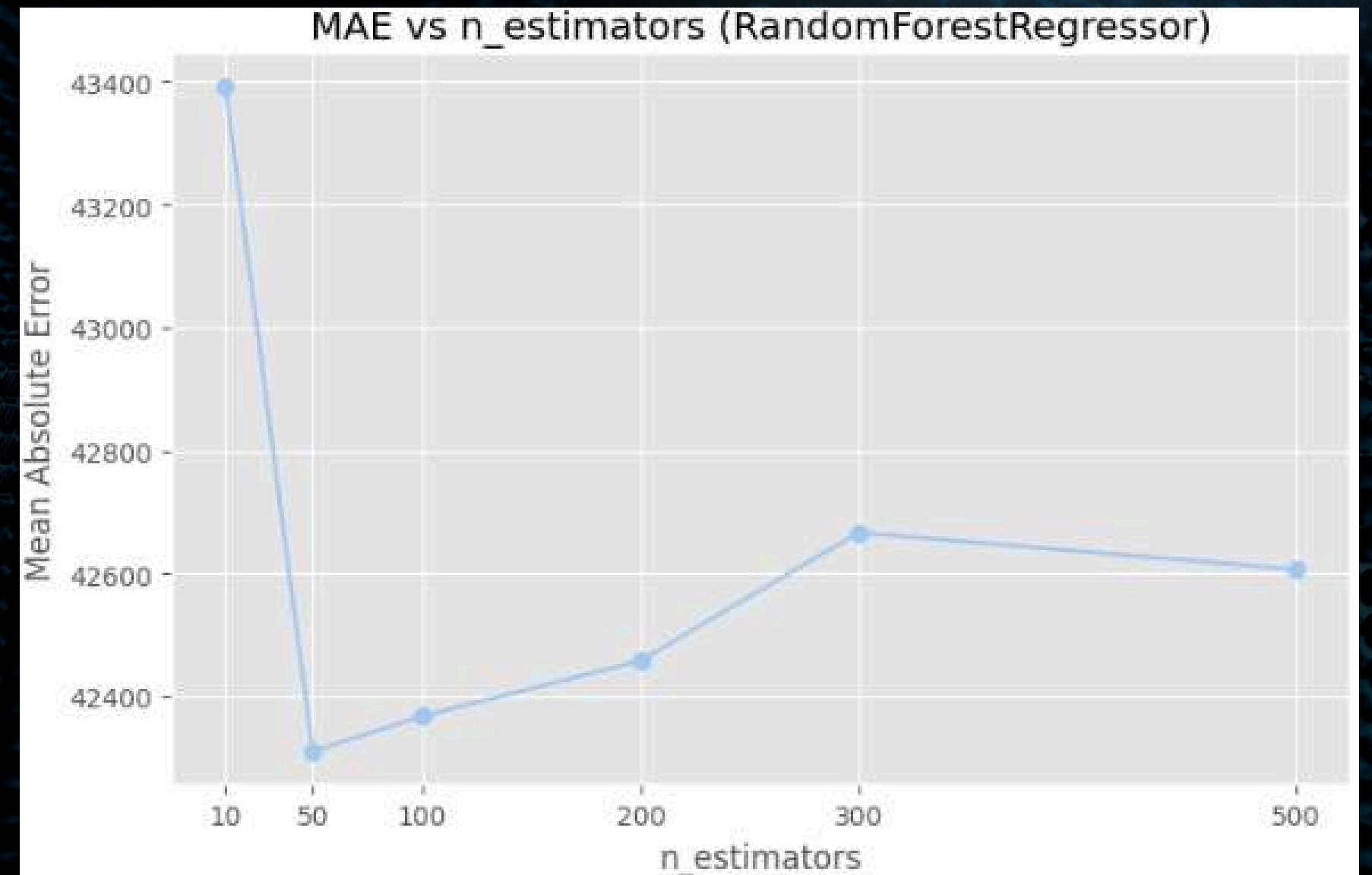
Este modelo utiliza una técnica llamada bosque aleatorio para predecir el valor de mercado de un jugador de fútbol, usando datos como su edad, altura, partidos jugados, asistencias y tarjetas amarillas.

En lugar de un solo árbol de decisión, el modelo combina muchos árboles diferentes y toma el promedio de todos ellos. Esto hace que las predicciones sean más estables, precisas y confiables. Probamos diferentes cantidades de árboles para ver cuántos se necesitan para obtener buenos resultados sin perder eficiencia.



MAE vs n_estimators for RF

```
MAE for n_estimator 10: 43392.27235989012  
Relative MAE for n_estimator 10: 0.95  
MAE for n_estimator 50: 42310.31821520568  
Relative MAE for n_estimator 50: 0.93  
MAE for n_estimator 100: 42367.49445193635  
Relative MAE for n_estimator 100: 0.93  
MAE for n_estimator 200: 42457.919983990345  
Relative MAE for n_estimator 200: 0.93  
MAE for n_estimator 300: 42666.3679737579  
Relative MAE for n_estimator 300: 0.94  
MAE for n_estimator 500: 42606.458878810314  
Relative MAE for n_estimator 500: 0.94  
Best Relative MAE: 0.93
```

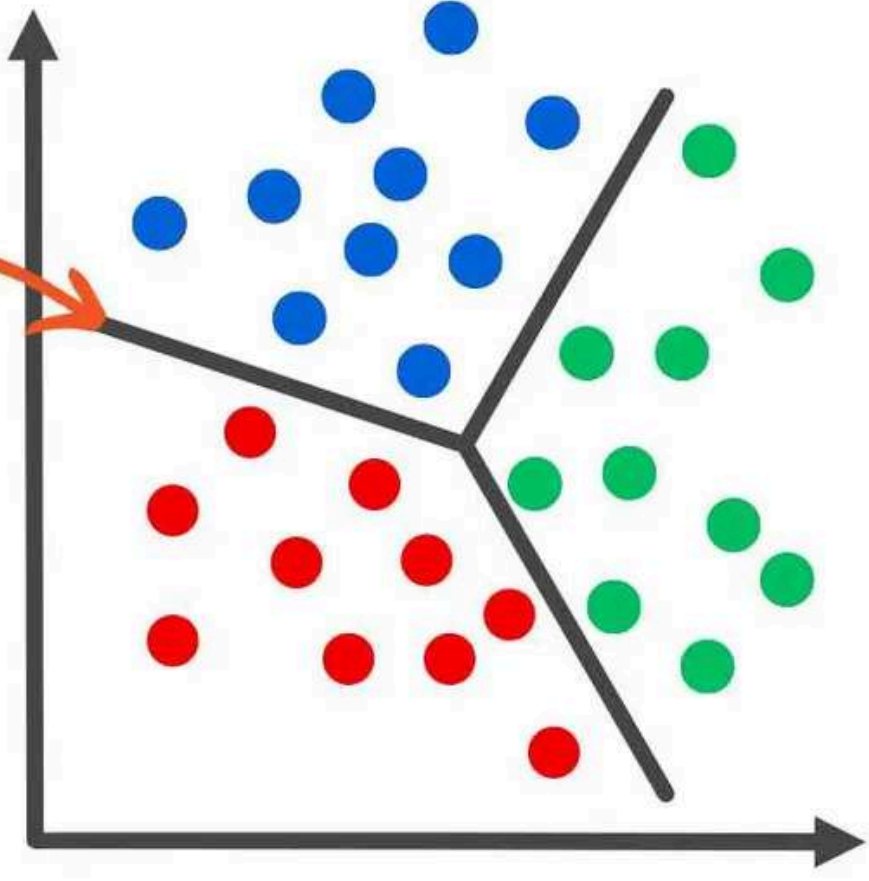


SVM (Support Vector Machine)

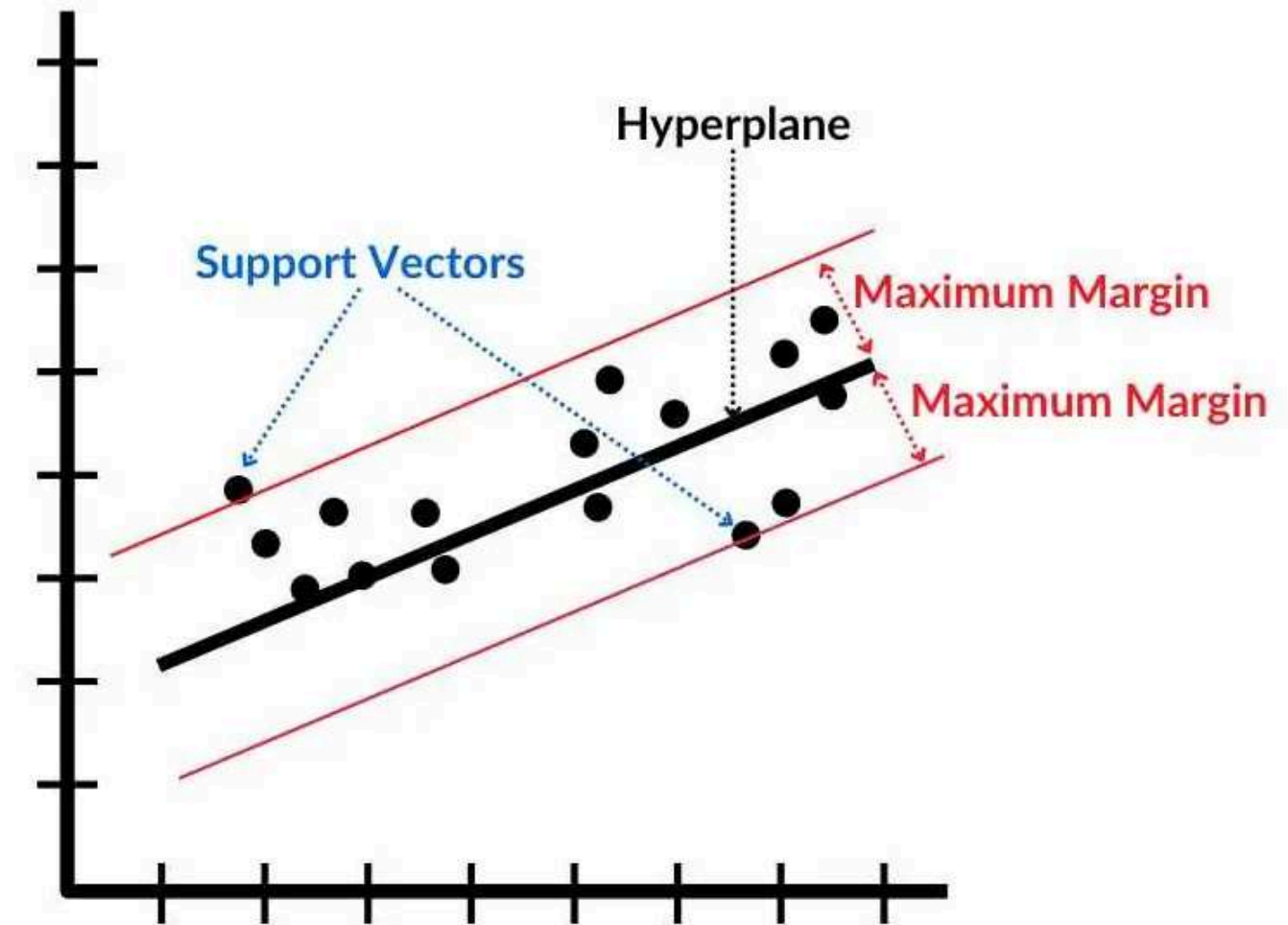
SVM (Máquina de Vectores de Soporte) es un algoritmo de aprendizaje automático supervisado que se usa principalmente para clasificación, busca encontrar una línea (o plano o superficie) óptima que separe datos

Support Vector
Machines (SVM)

Hyperplanes that Best
Separates Different Classes



Support Vector Regression (SVR)

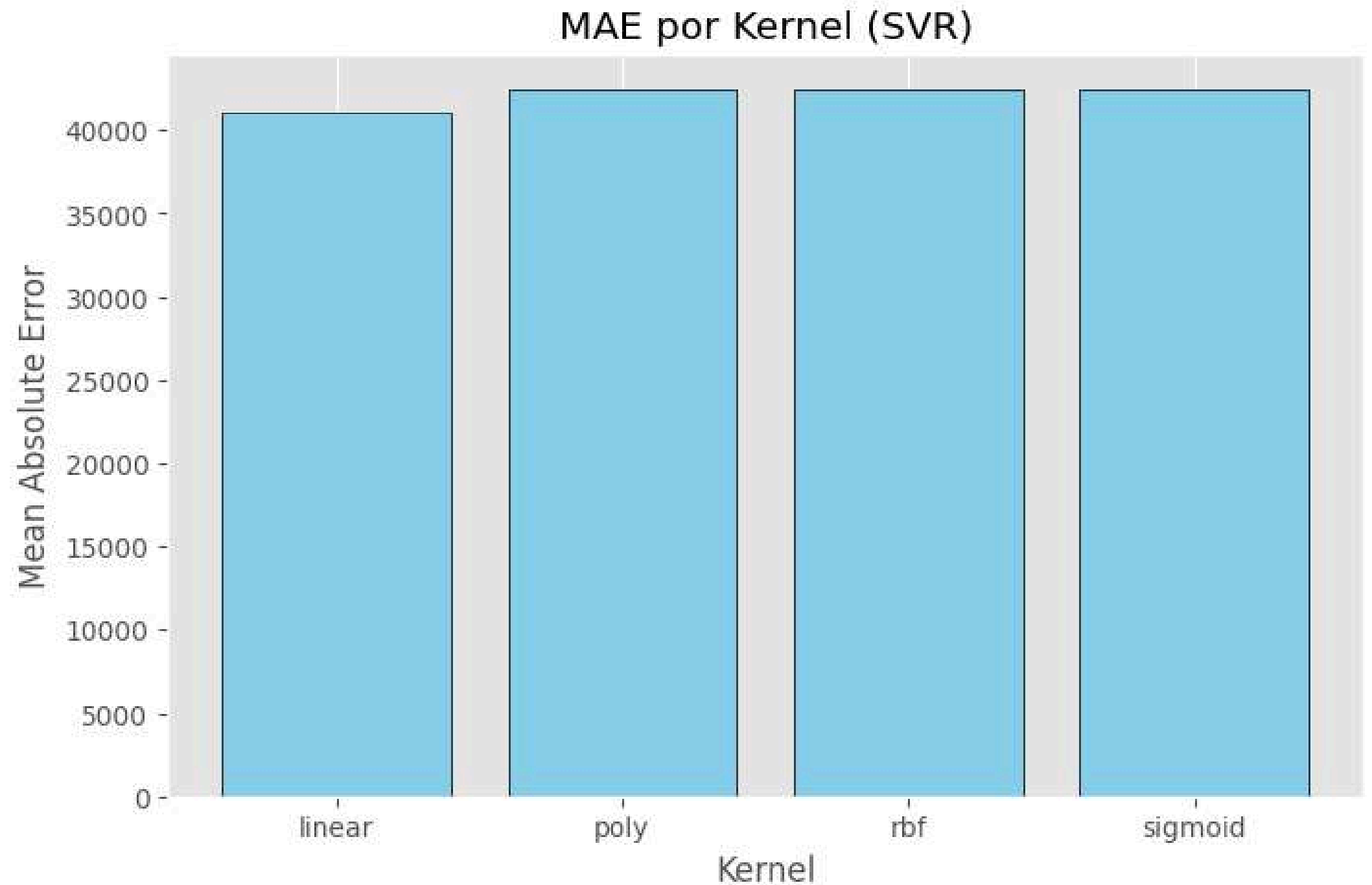


SVR (Support Vector Regression)

es una técnica de aprendizaje automático utilizada para resolver problemas de regresión. Similar a las SVM para clasificación. En SVR, se permite un pequeño margen de error (llamado "epsilon") y el modelo intenta que la mayoría de los puntos estén dentro de ese margen, sin penalizar demasiado.

MAE for kernel linear: 41046.73673334646
Relative MAE for kernel linear: 0.9
MAE for kernel poly: 42415.34798170879
Relative MAE for kernel poly: 0.93
MAE for kernel rbf: 42441.725852636235
Relative MAE for kernel rbf: 0.93
MAE for kernel sigmoid: 42452.61727977079
Relative MAE for kernel sigmoid: 0.93
Best Relative MAE: 0.9

MAE vs kernels for SVM



¿Por qué esto es útil?



Análisis de rendimiento económico: Clubes pueden usar este tipo de modelos para ver si un jugador está sobrevalorado o infravalorado respecto a sus estadísticas reales.

**Toma de decisiones en scouting o fichajes
Permite estimar el valor de un jugador sin necesidad de saberlo de antemano, solo conociendo su desempeño en el campo y algunos datos físicos.**



**Comparación de modelos de predicción:
Al usar varios kernels (linear, polinómico, RBF, etc.), evaluamos qué modelo se ajusta mejor a los datos reales del fútbol español.**



Gracias!