

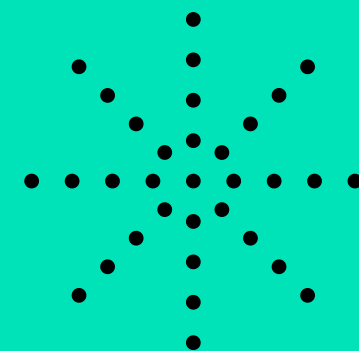
PROYECTO DE ESTADISTICA II

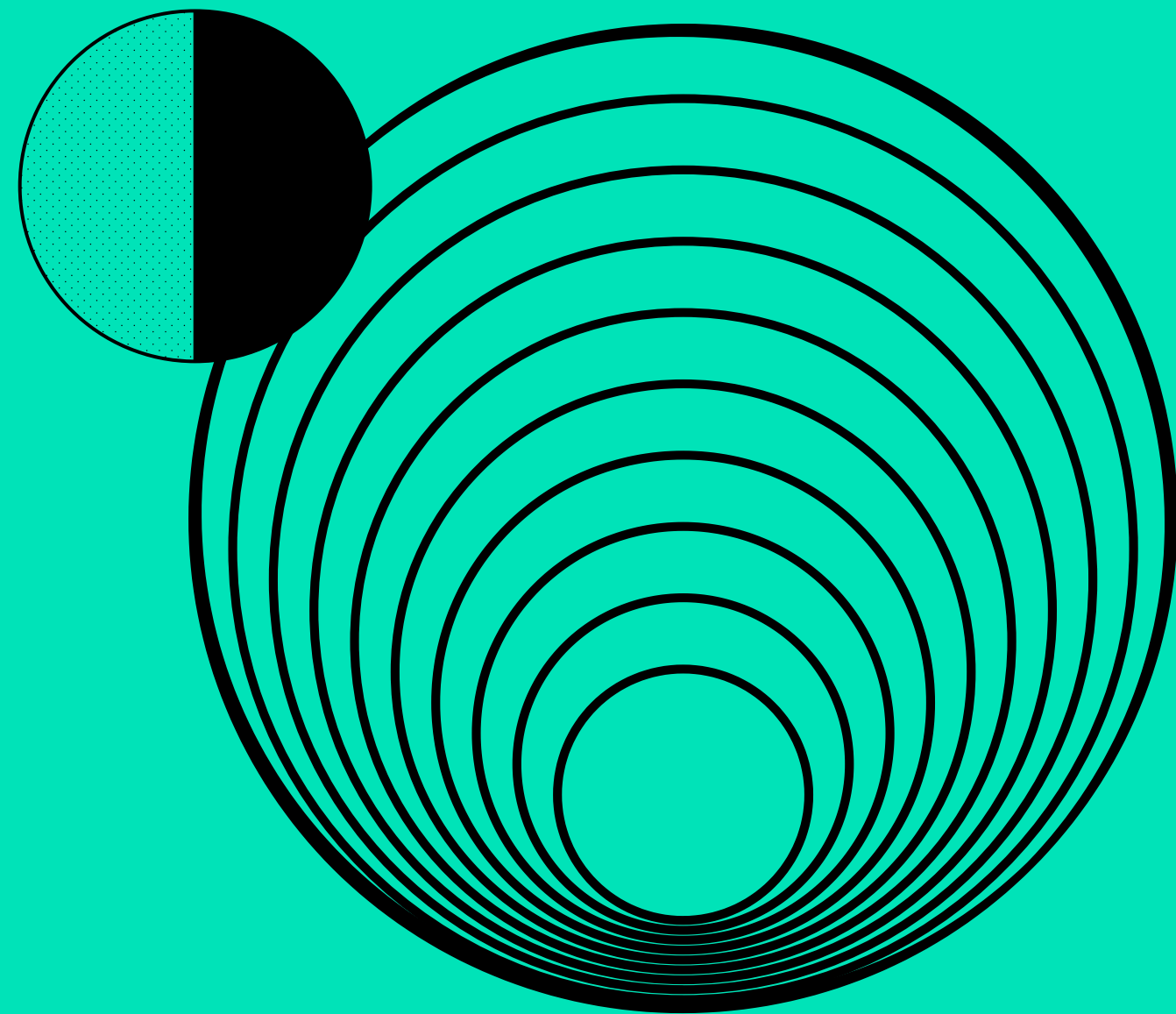
ANÁLISIS DEL RENDIMIENTO DE JUGADORES DE FÚTBOL DURANTE LA TEMPORADA 2022-2023

Por:

Leider Esteban

Nestor Hernandez



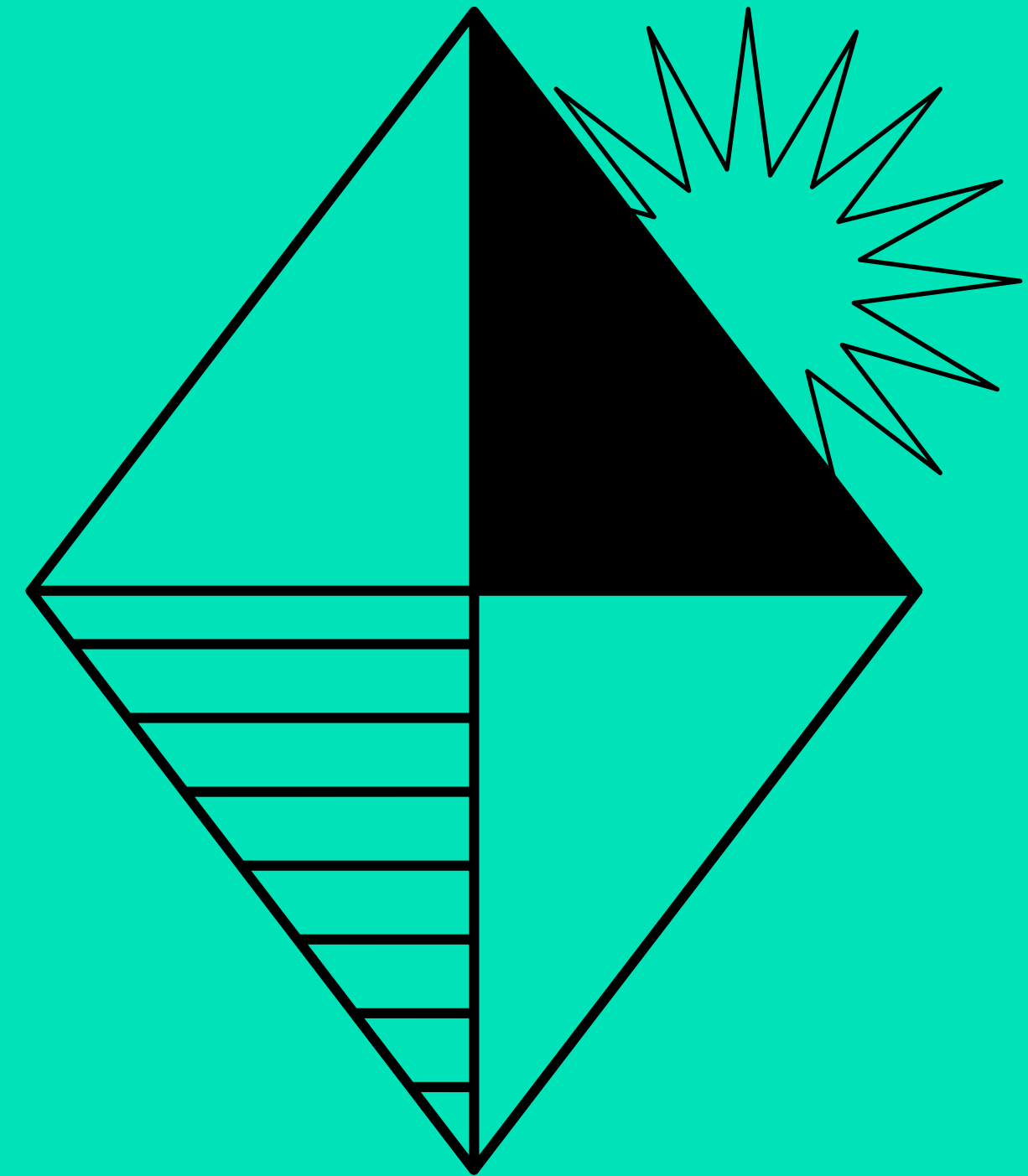


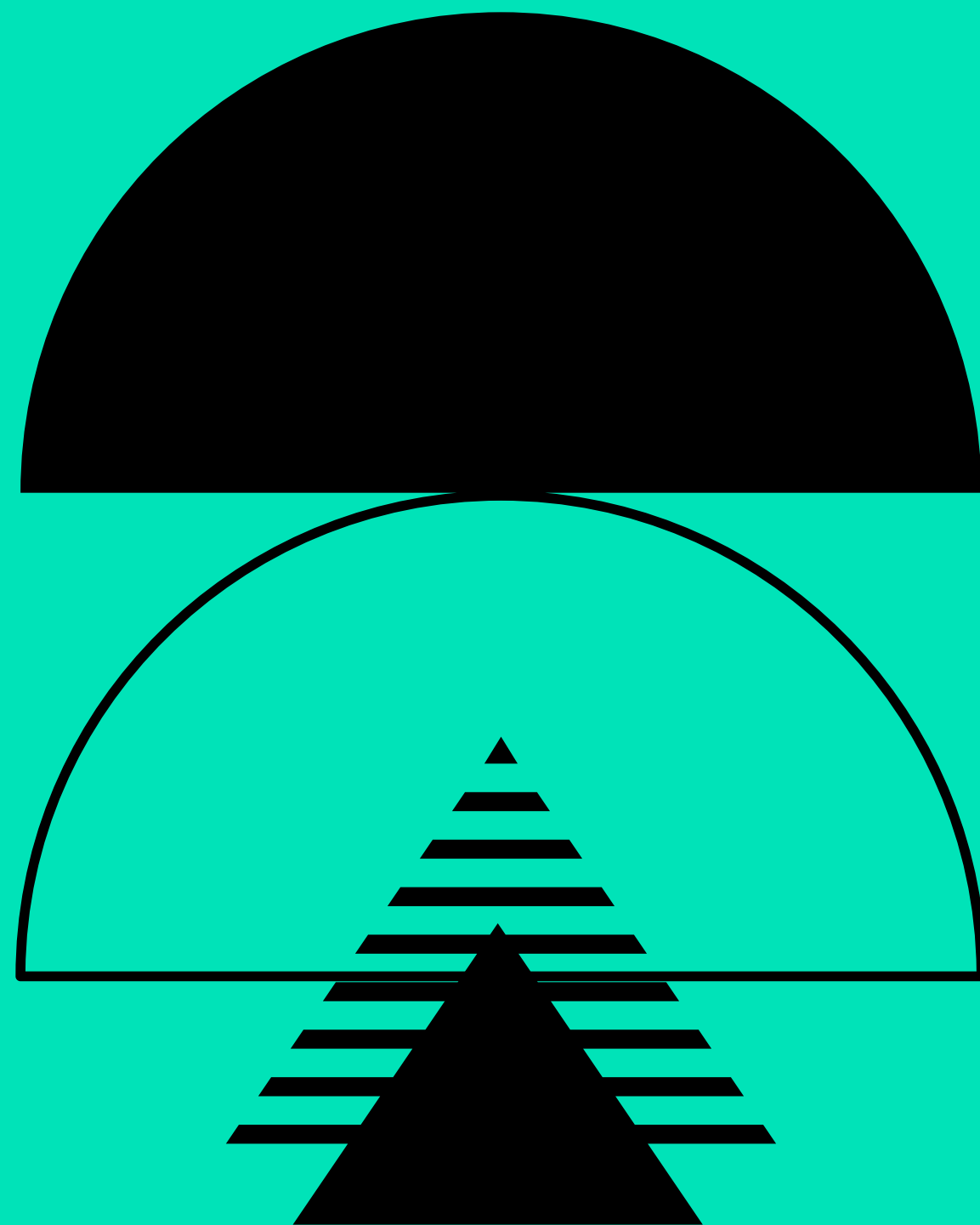
Contenidos.

1. Introducción.
2. Limpieza y organización
3. del dataset.
4. Datos generales.
5. Puntos clave.
6. Estadísticas.
7. Cifras obtenidas.
8. Análisis detallado.
9. Los mercados.
- Conclusiones.

Introducción.

El presente proyecto de Estadística 2 tiene como objetivo analizar un conjunto de datos correspondiente al de futbolistas en la temporada 2022-2023, incluyendo goles, asistencias, minutos jugados y tarjetas. Busca identificar patrones, comparar jugadores y posiciones, y aplicar técnicas estadísticas como medidas de tendencia central, dispersión y regresión. Este estudio conecta la estadística con un contexto práctico y relevante como el fútbol.

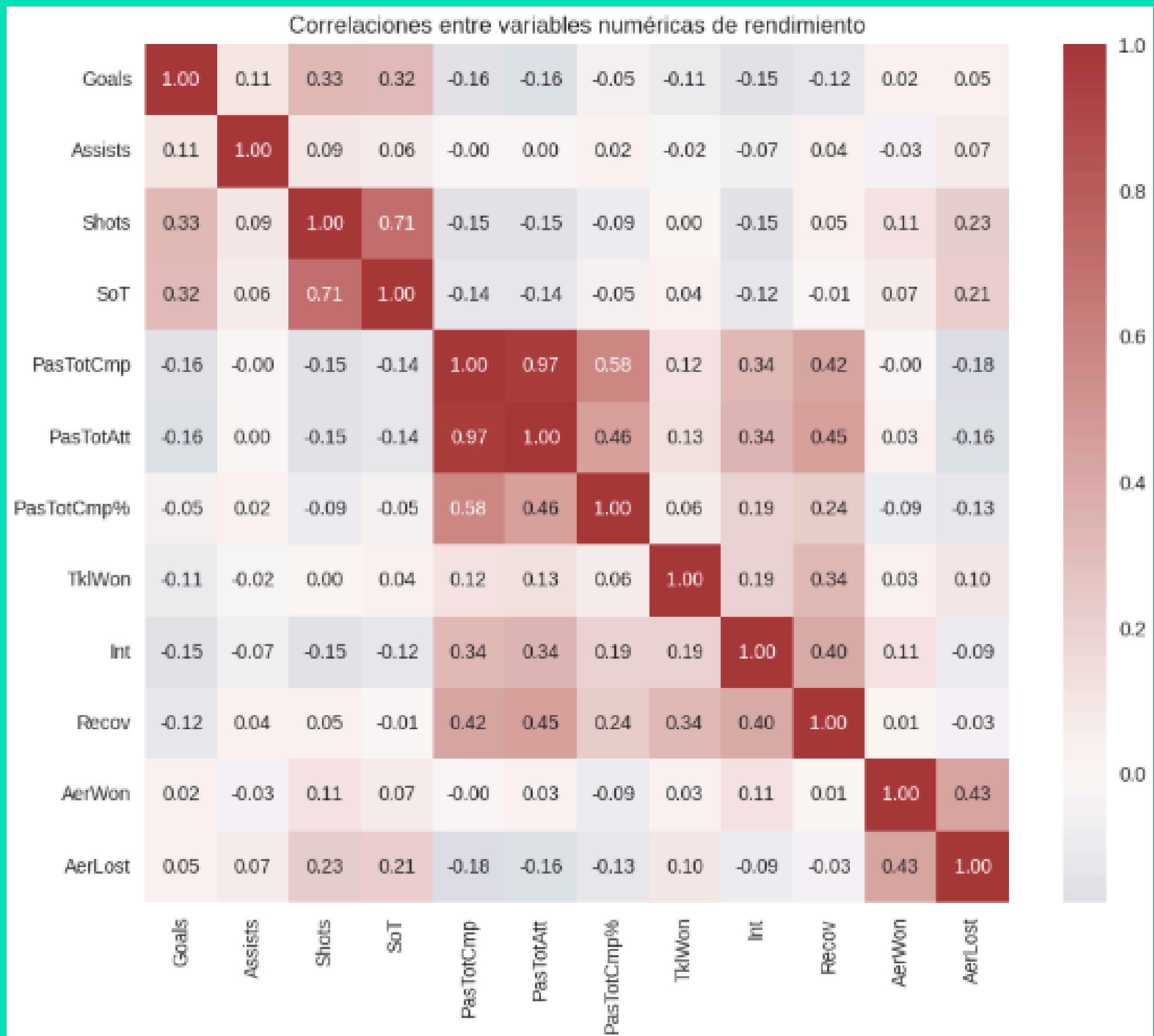




Limpieza y organización del dataset

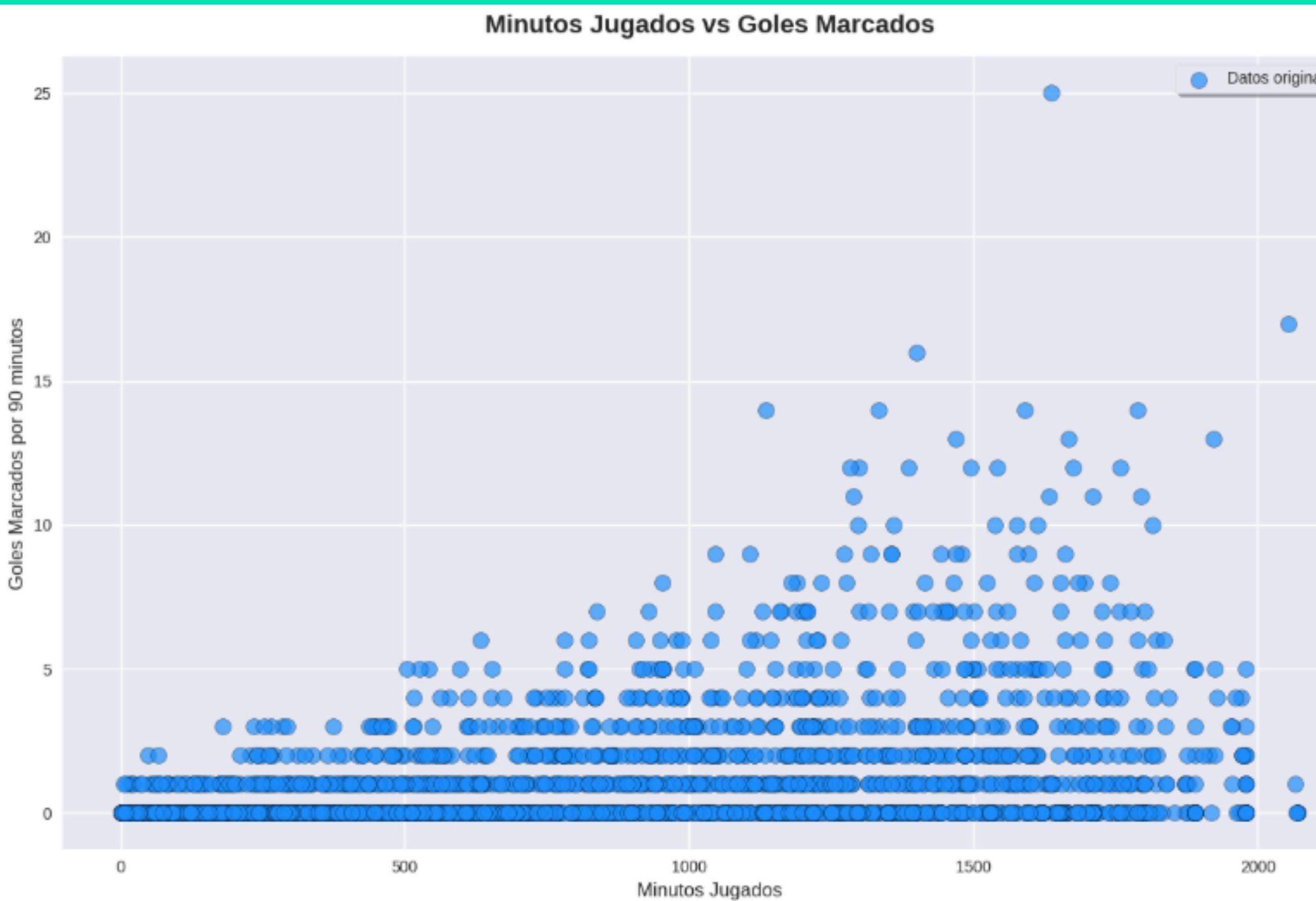
```
#@title **Limpieza y organización del dataset**  
# Ver dimensiones y tipos de datos  
df_info = df.info()  
# Verificar valores nulos por columna  
missing_values = df.isnull().sum().sort_values(ascending=False)  
# Contar duplicados  
duplicates_count = df.duplicated().sum()  
# Mostrar resultados clave  
df.shape, missing_values.head(10), duplicates_count
```

Matriz de correlación



Se seleccionan variables clave del rendimiento futbolístico (goles, asistencias, disparos, pases, etc.) para calcular una matriz de correlación, visualizada con un mapa de calor que muestra relaciones entre métricas.

Grafica De Dispersion



Análisis de la Gráfica de Dispersión:

- Alta concentración de jugadores con pocos o ningún gol por 90 minutos, incluso con muchos minutos jugados.

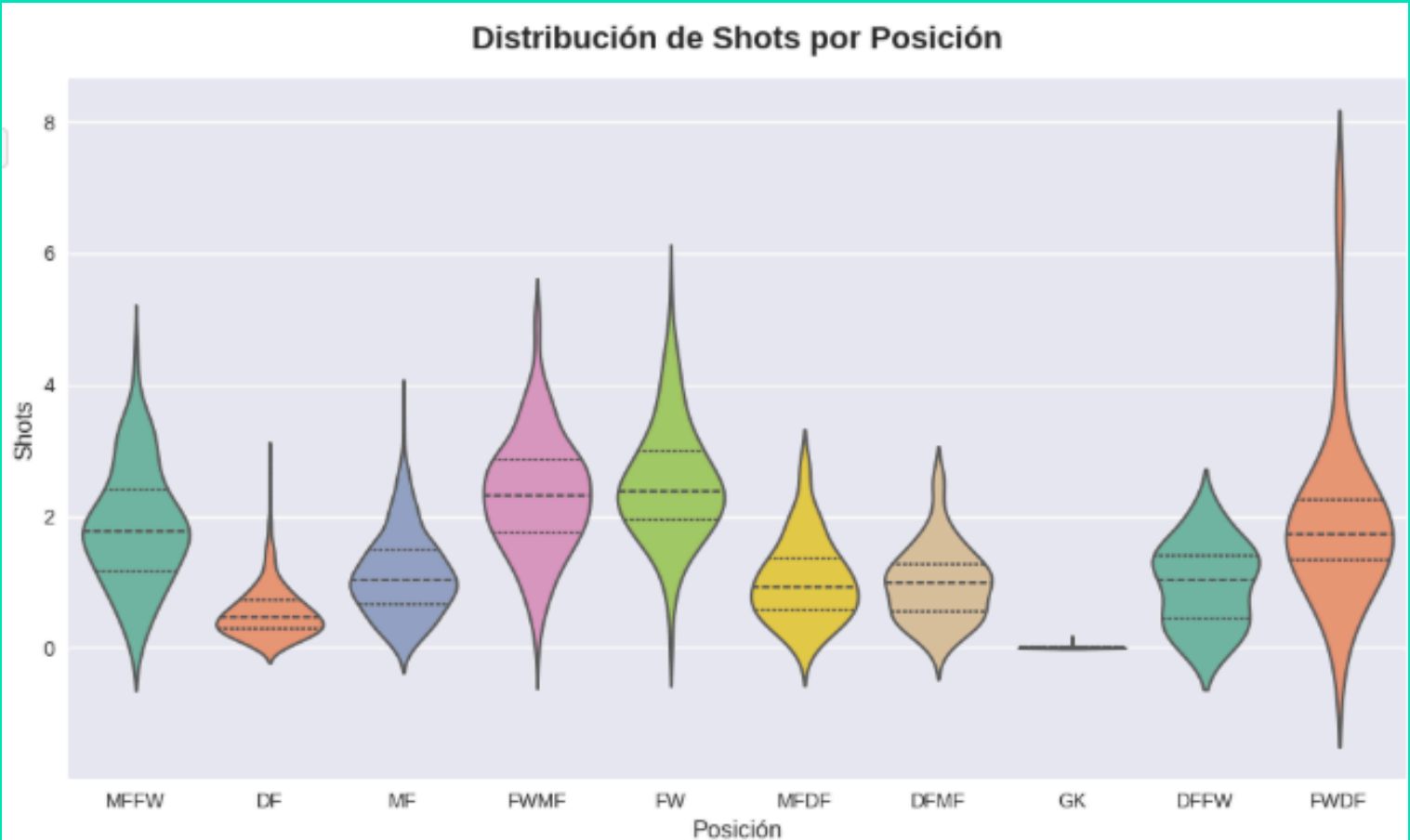
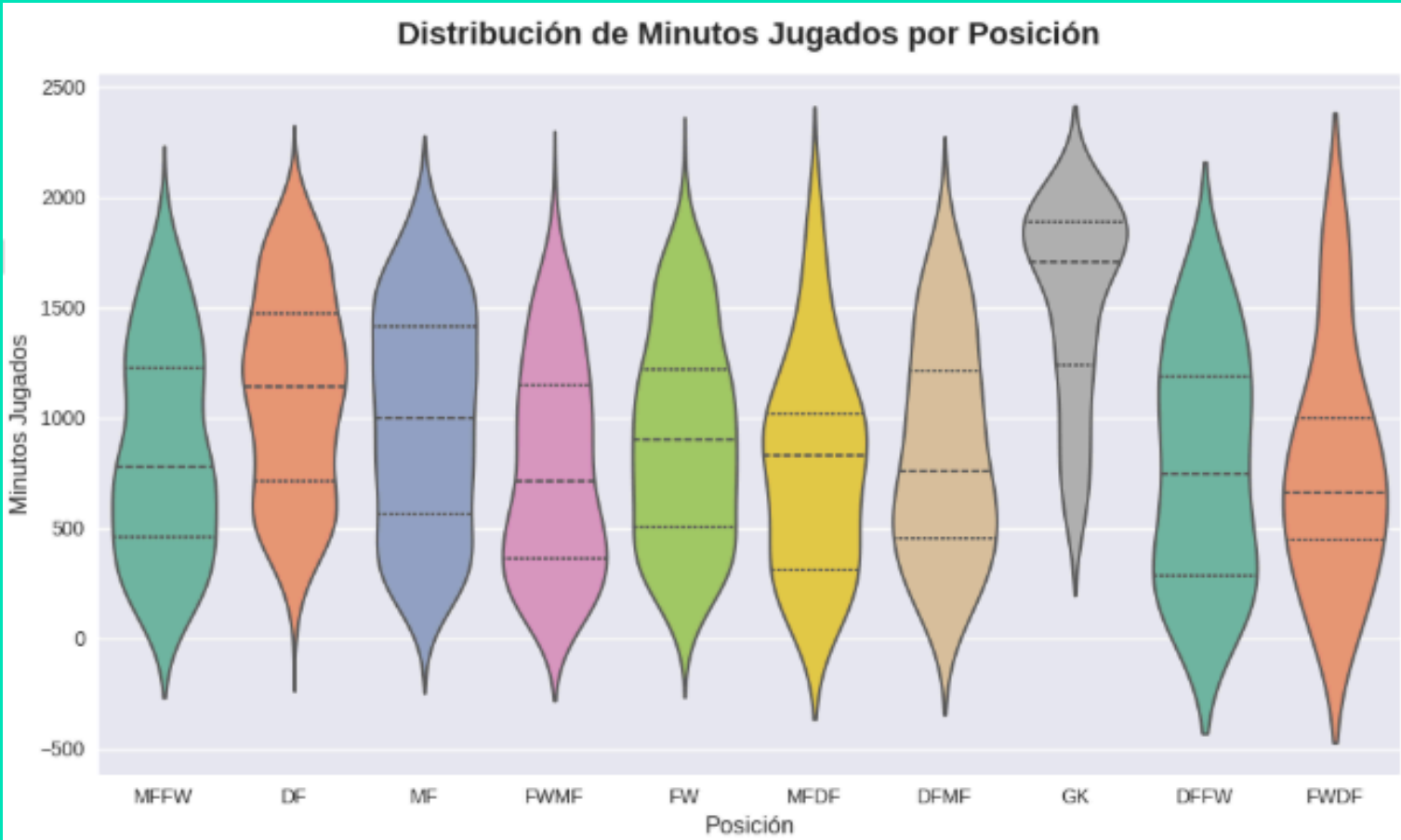
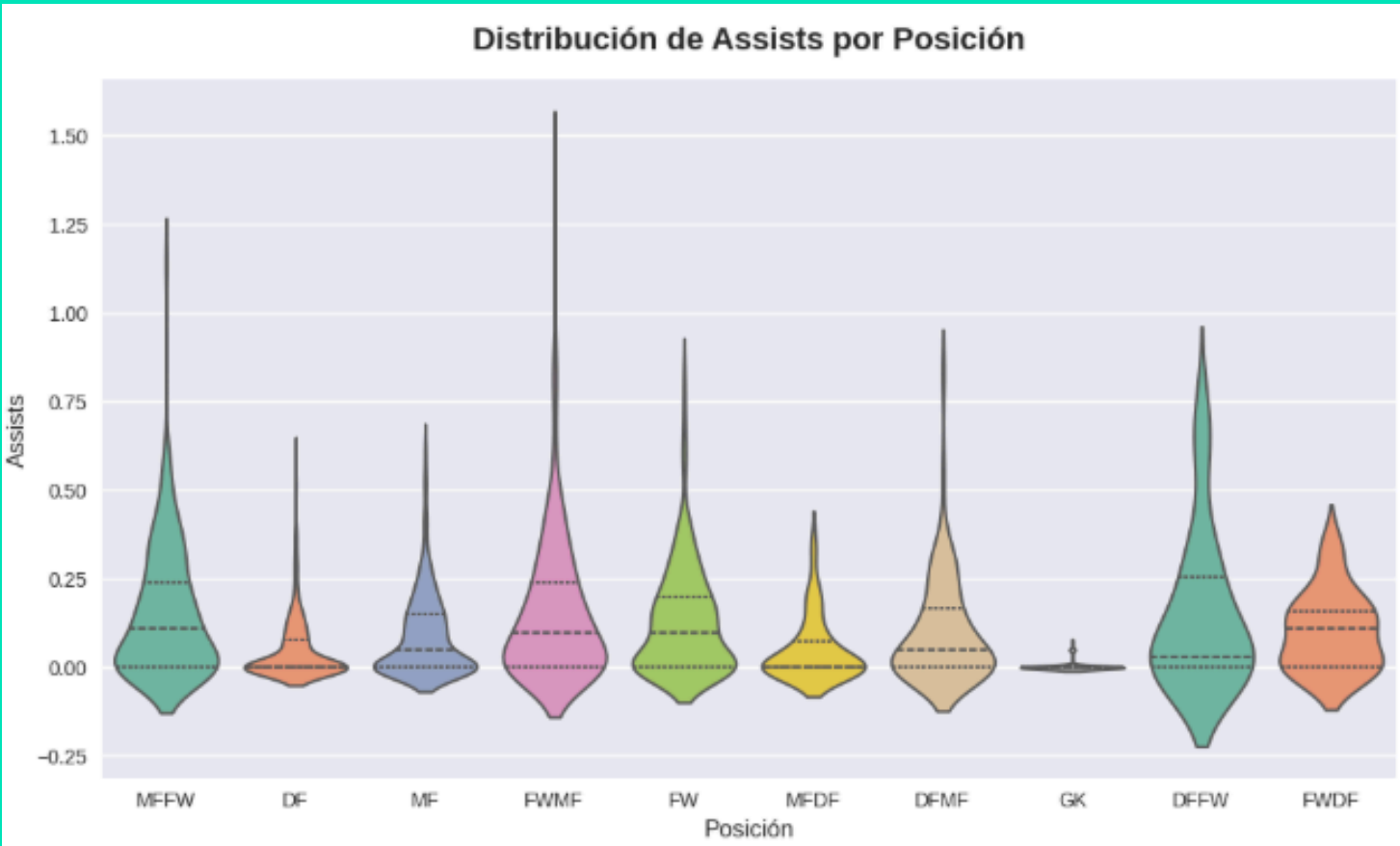
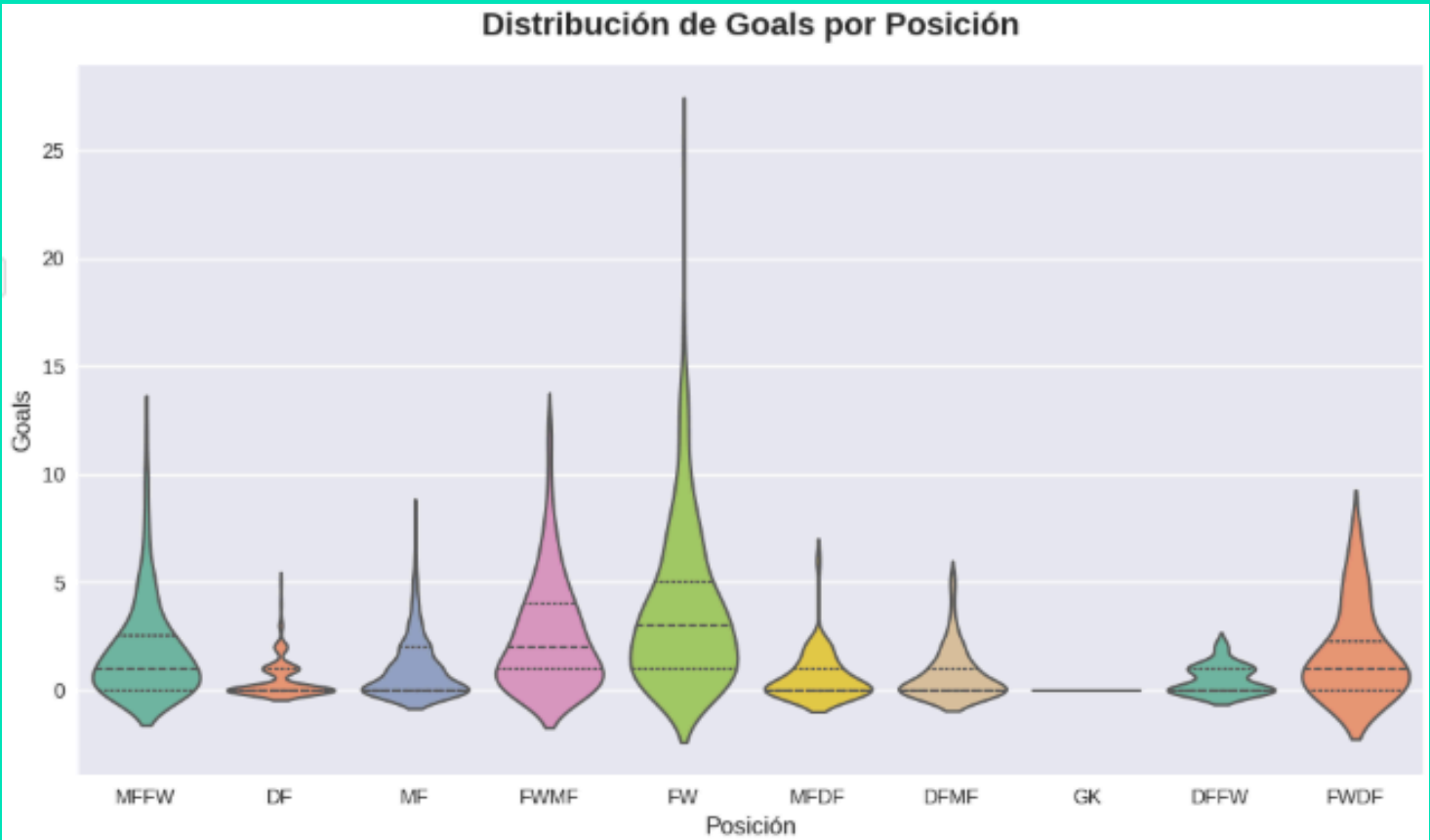
- Algunos jugadores con pocos minutos muestran tasas de goles muy altas, distorsionando el análisis.

- Mayor estabilidad en tasas de gol entre jugadores con más minutos en cancha.

- No hay correlación lineal clara entre minutos jugados y goles por 90 minutos.

- Presencia de outliers con valores extremos (más de 10 goles/90 min) y poca participación.

Graficas de Violin

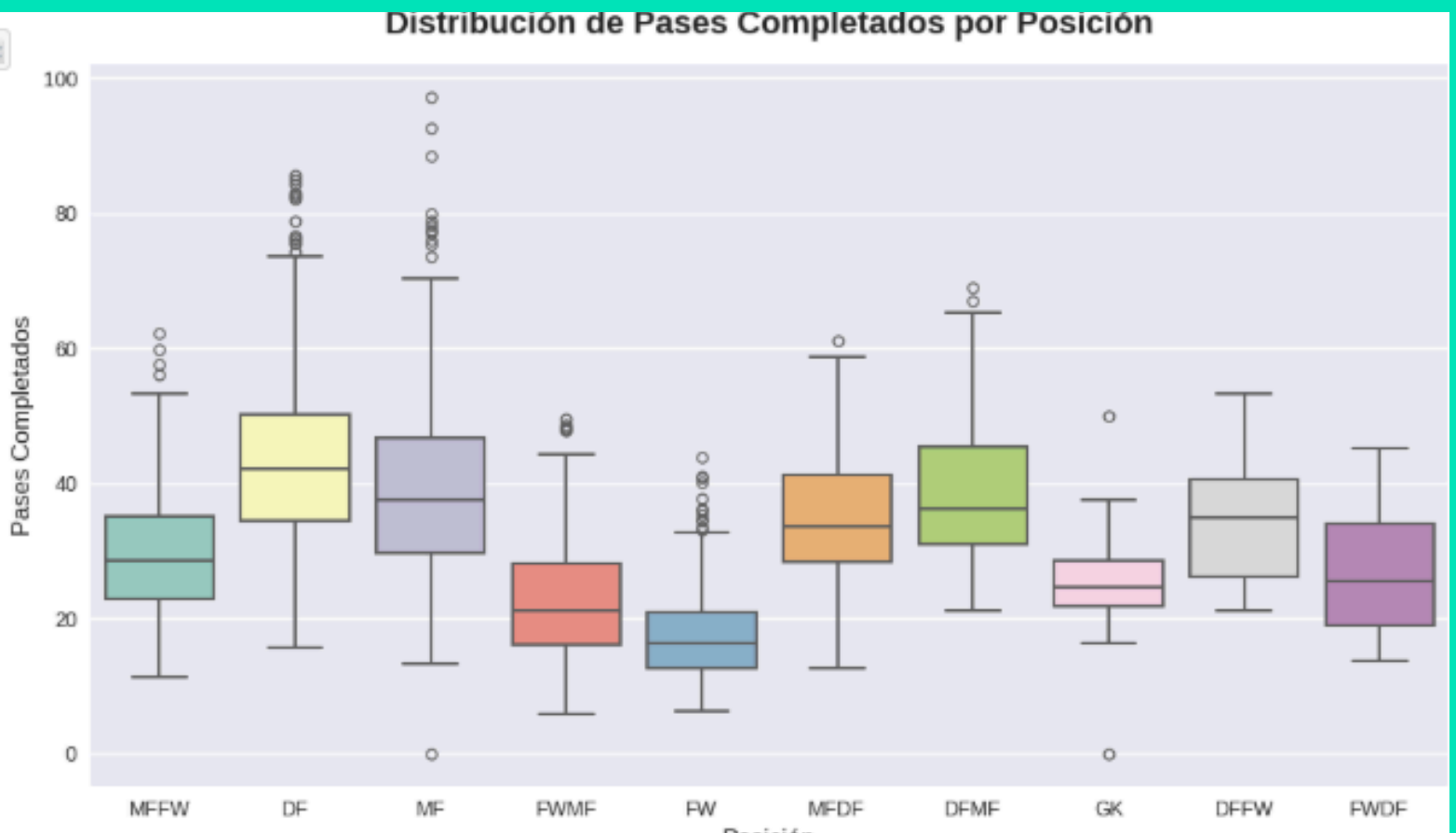
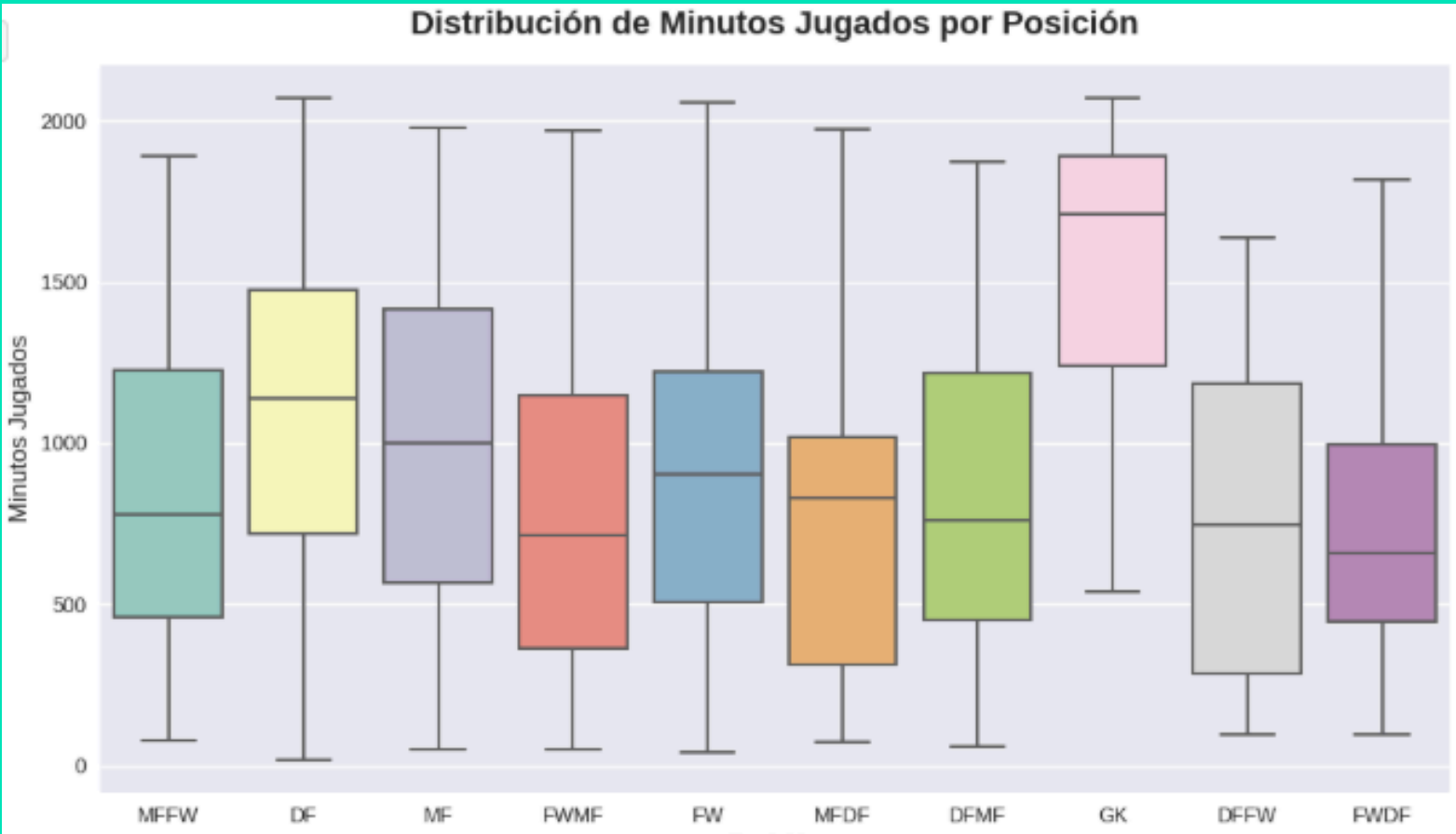
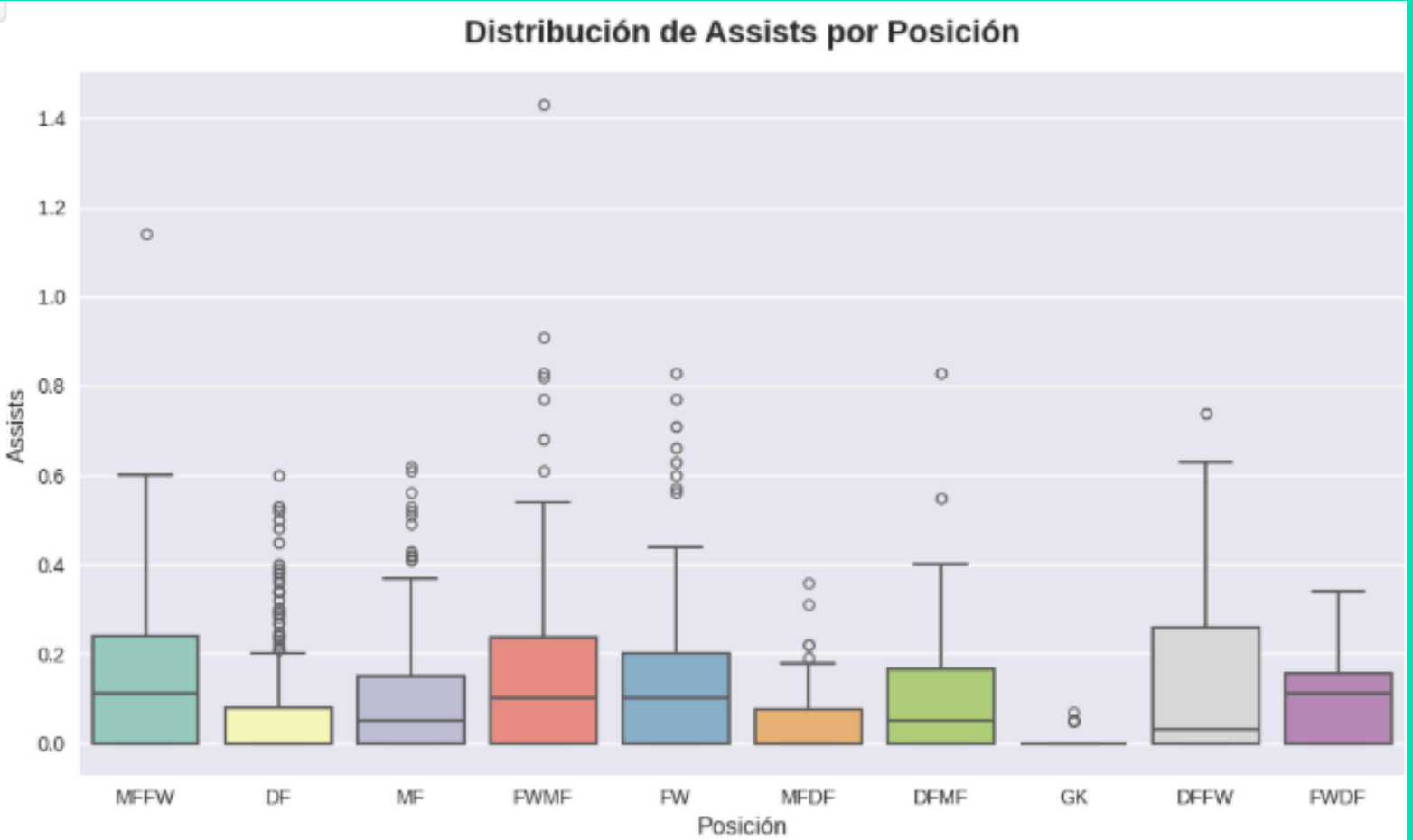
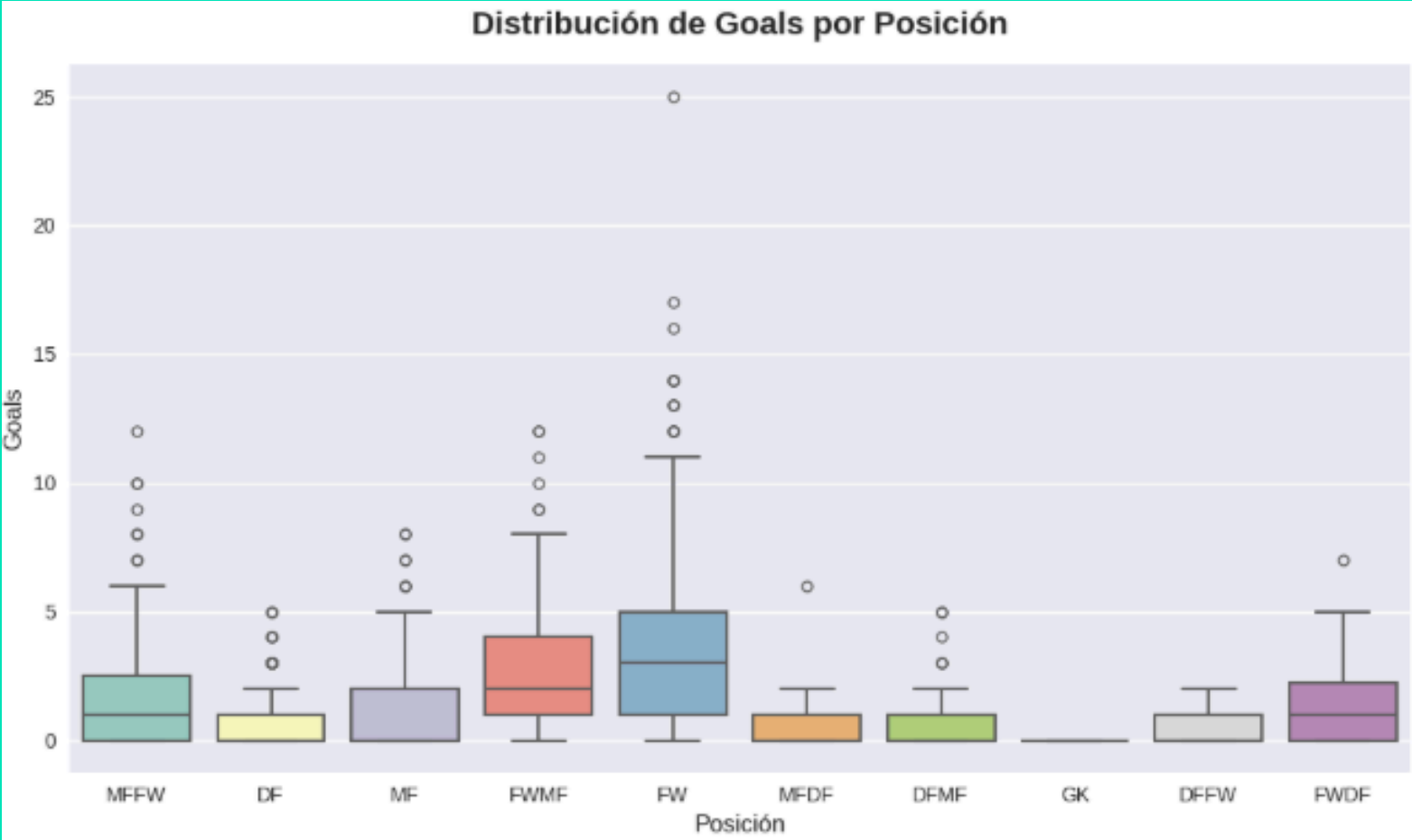


Graficas de Violin

Los gráficos de violín muestran la distribución estadística por posición en el campo.

- **Goles:** Los delanteros (FW) y mediocampistas ofensivos (MFFW) destacan, mientras que defensores (DF) y arqueros (GK) tienen contribuciones mínimas.
- **Asistencias:** Más homogéneas, con mayor concentración en MF y MFFW, reflejando su rol ofensivo. GK y DF apenas participan.
- **Minutos Jugados:** Distribución similar en todas las posiciones, con DF y GK mostrando menor dispersión por su participación constante.
- **Pases Completados:** MF lidera, consistente con su función de distribución; GK y FW tienen menor aporte.
- **Disparos:** FW domina, seguido por MFFW; otras posiciones, especialmente GK, son menos activas.
- **Tackles Ganados:** DF y DFMF sobresalen, reflejando su rol defensivo; GK y posiciones ofensivas tienen poca actividad.

Graficas de cajas

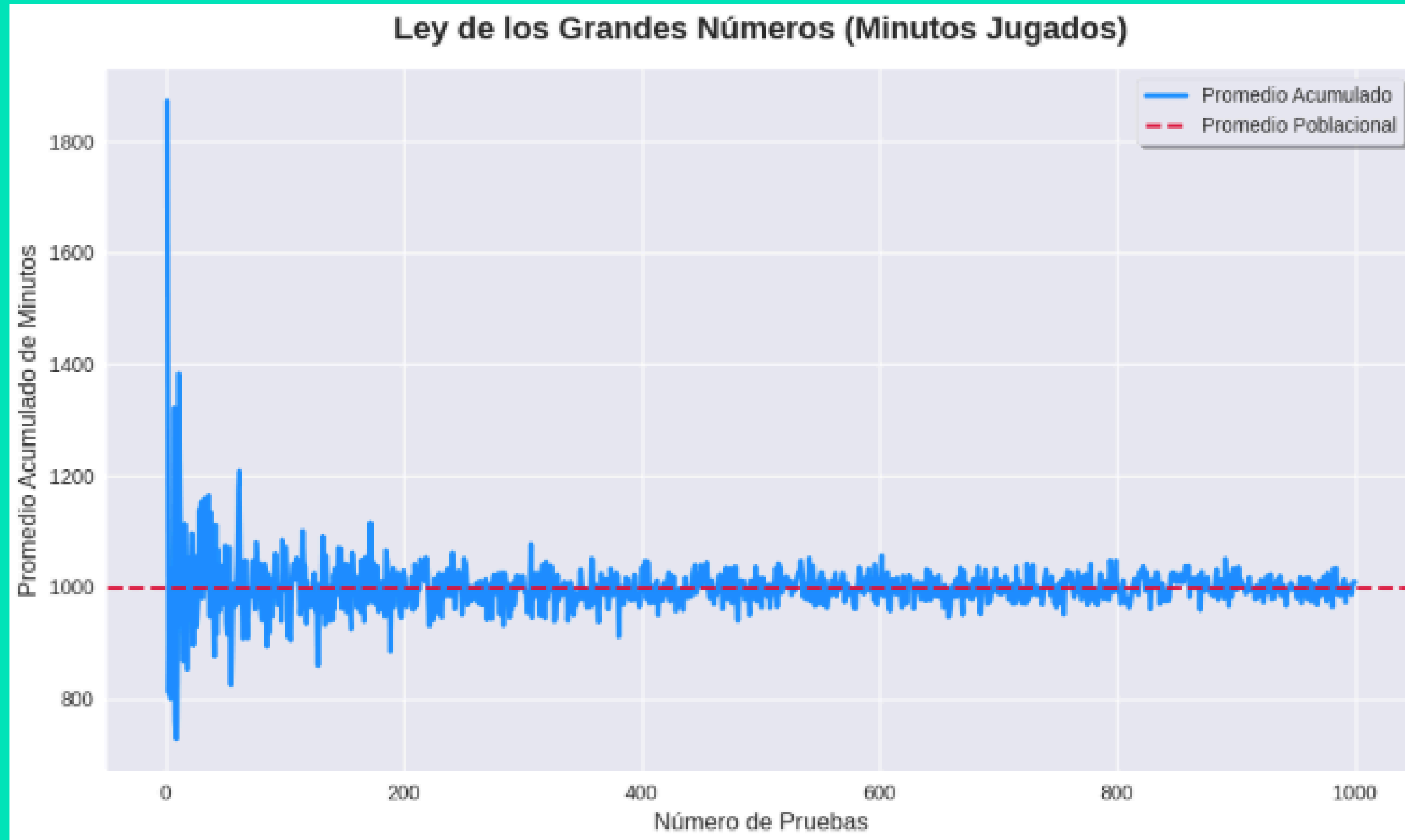


Graficas de Violin

Los gráficos muestran la distribución de métricas futbolísticas según la posición del jugador.

- **Goles:** Los delanteros (FW, FWDF, FWMF) tienen mayores medianas, mientras que defensas y porteros apenas anotan.
- **Asistencias:** Las posiciones ofensivas (MF, FWMF, FW) destacan, aunque con alta dispersión; GK y DF tienen valores mínimos.
- **Minutos Jugados:** Distribución equilibrada, con porteros (GK) mostrando mayor constancia.
- **Pases Completados:** Los defensas (DF) sobresalen, mientras que FWDF y GK tienen menos.
- **Tiros:** Delanteros (FW, FWMF) lideran, con defensas mostrando medianas cercanas a cero.
- **Tackles Ganados:** Defensas (DF, DFMF, DFFW) y mediocampistas defensivos (MFDF) tienen los mayores valores, confirmando su rol en recuperación.

Grafica de la Ley de los grandes Numeros

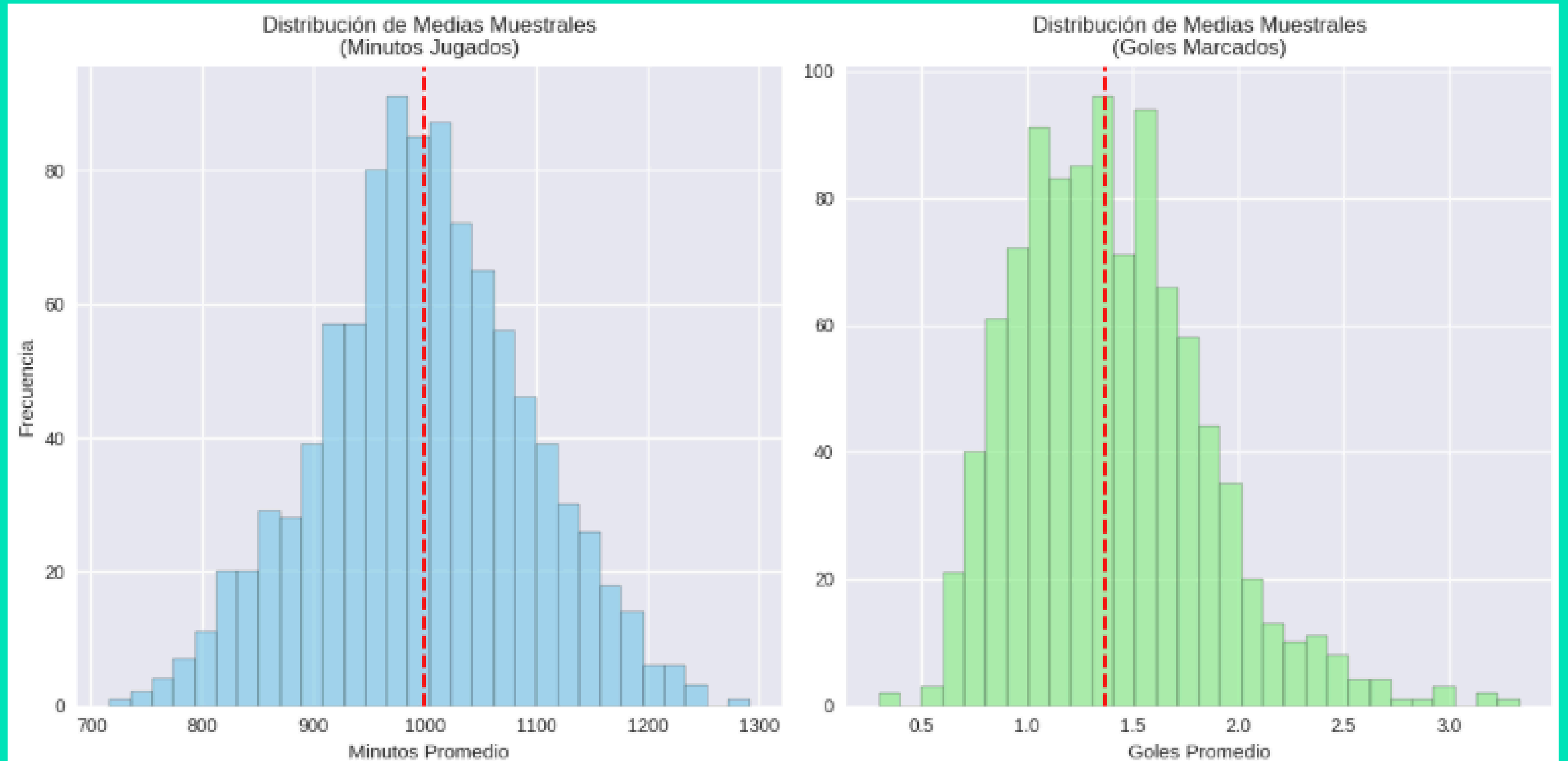


Ley de los grandes Numeros

Observaciones Clave:

- Alta variabilidad inicial debido a pocas observaciones.
- Convergencia del promedio acumulado al aumentar pruebas.
- Aproximación al promedio poblacional, validando la Ley de los Grandes Números.
- Reducción de oscilaciones con más datos, mejorando la consistencia.
- Evidencia de estabilidad estadística con un mayor tamaño de muestra.

Teorema del Límite Central

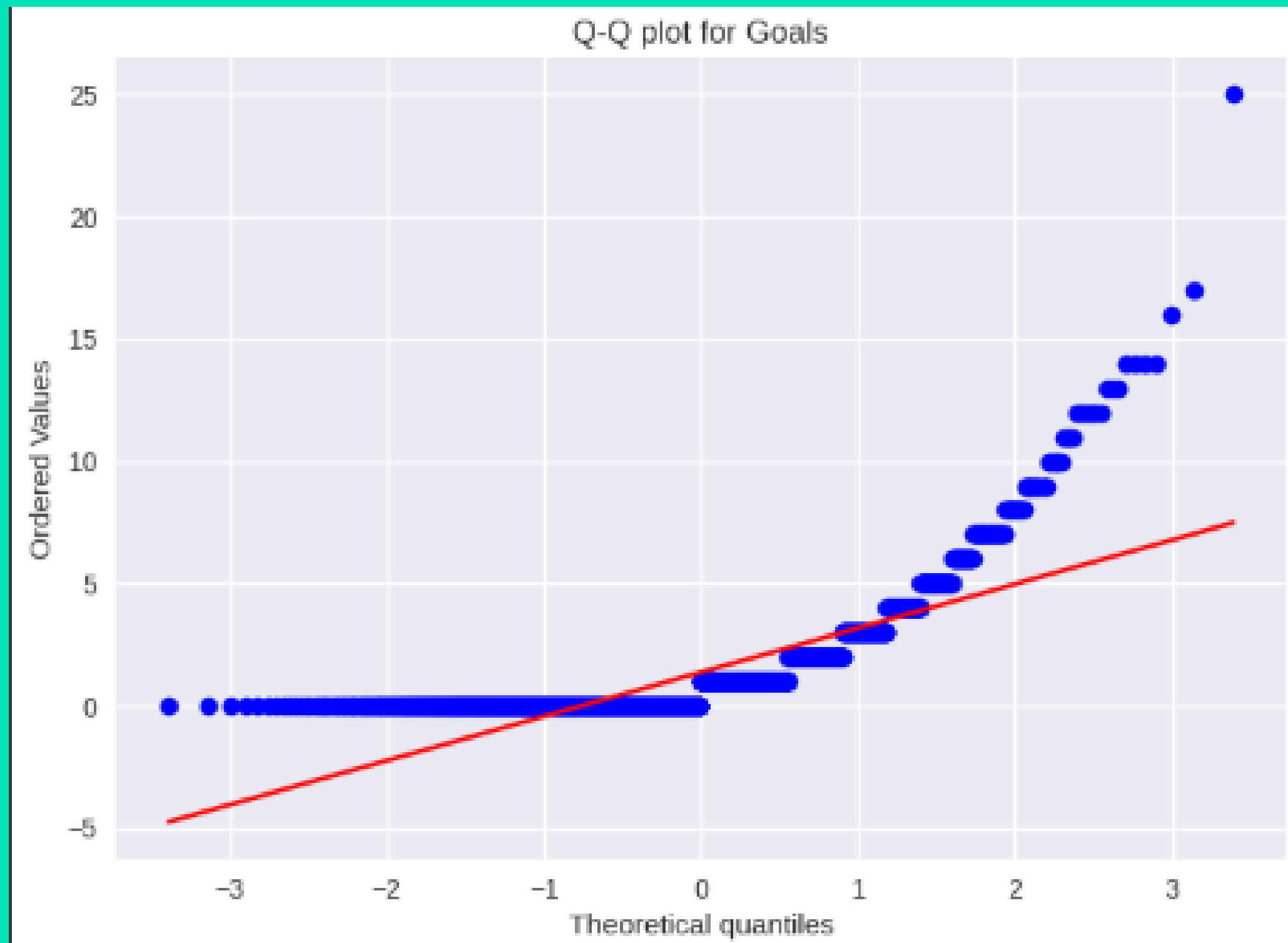


Teorema del Límite Central

Observaciones Clave:

1. Ambas distribuciones tienen forma de campana, respaldando el Teorema Central del Límite.
2. Las medias están centradas en los valores poblacionales, mostrando muestras representativas.
3. Menor dispersión en minutos jugados, con una distribución más simétrica y concentrada.
4. Mayor asimetría en goles marcados, con sesgo a la derecha por jugadores con muchos goles.
5. Consistencia muestral: los promedios de las muestras se agrupan cerca del promedio poblacional.

gráficos Q-Q

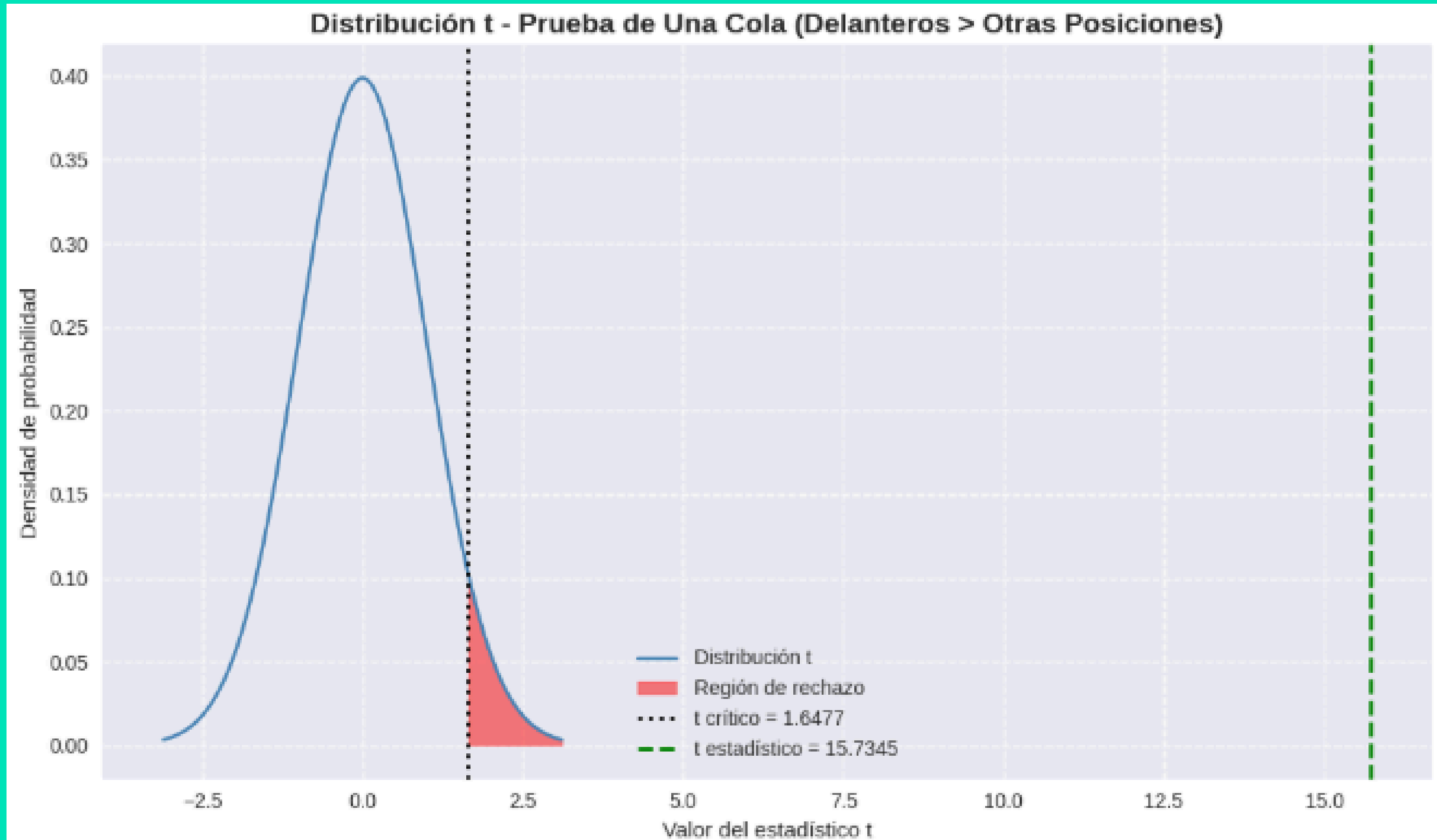


gráficos Q-Q

Las observaciones en datos de conteo, comunes en deportes como el fútbol donde los eventos son raros o irregulares, pueden analizarse mejor utilizando distribuciones como la de Poisson o la binomial negativa, más adecuadas para modelar eventos infrecuentes y variables discretas.

El análisis de residuos y la detección de jugadores con rendimientos excepcionales (outliers) aportan información clave para estrategias de juego y evaluaciones de desempeño, permitiendo ajustar expectativas y mejorar predicciones sobre el rendimiento de jugadores y equipos en futuros eventos deportivos.

PRUEBA DE HIPOTESIS DE UNA MUESTRA



PRUEBA DE HIPOTESIS DE UNA MUESTRA

Objetivo: Determinar si la probabilidad media de marcar goles de los delanteros es mayor que la de jugadores de otras posiciones.

Contexto: Dado el rol de los delanteros y el análisis exploratorio, se plantea que tienen una mayor producción goleadora.

Hipótesis:

- $H_0: \mu_{\text{delanteros}} \leq \mu_{\text{otras}}$
- $H_1: \mu_{\text{delanteros}} > \mu_{\text{otras}}$

Prueba: t de Student para muestras independientes (Welch), unilateral, con $\alpha = 0.05$.

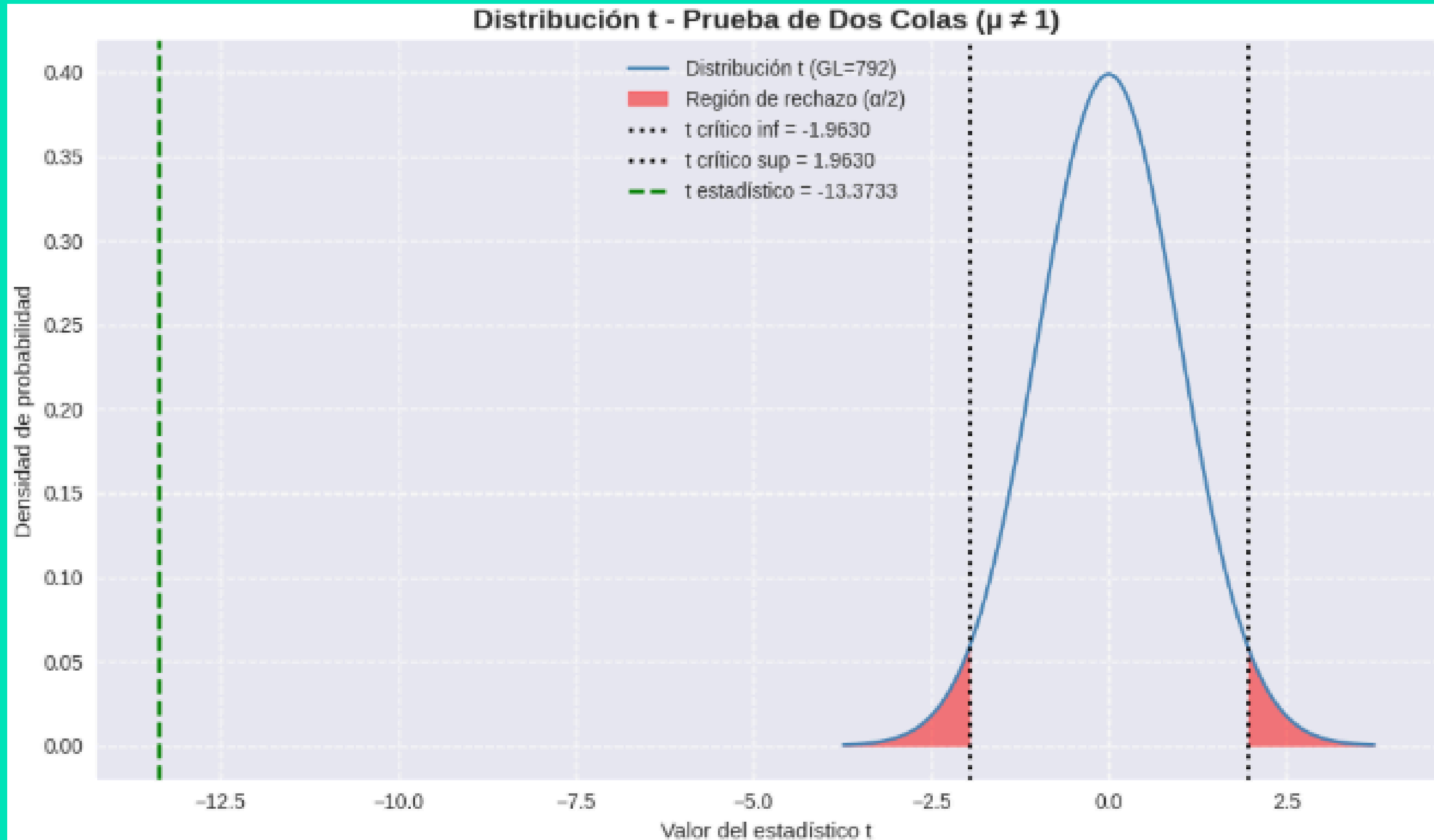
Resultados:

- Media muestral (Delanteros): 2.7441
- Media muestral (Otras Posiciones): 0.6447
- t-estadístico: 17.2187
- Grados de libertad: 772.43
- Valor crítico t: 1.6468
- p-valor: 0.0000
- IC unilateral (95%): Límite superior = 2.3002

Conclusión:

Como el p-valor (0.0000) < α (0.05) y el t-estadístico (17.2187) > valor crítico t (1.6468), se rechaza H_0 . Existe evidencia estadística significativa para afirmar que la probabilidad media de marcar goles de los delanteros es mayor que la de jugadores en otras posiciones.

PRUEBA DE HIPOTESIS DE DOS MUESTRAS



PRUEBA DE HIPOTESIS DE DOS MUESTRAS

Objetivo: Determinar si la media de goles anotados por defensores difiere significativamente de 1 gol.

Contexto: En el fútbol, los defensores tienen un rol menos ofensivo. Es importante evaluar si su media goleadora difiere de 1 gol.

Prueba: Se realizó una prueba t de Student para una muestra, de dos colas, con $\alpha = 0.05$.

$H_0: \mu_{\text{defensores}} = 1$

$H_1: \mu_{\text{defensores}} \neq 1$

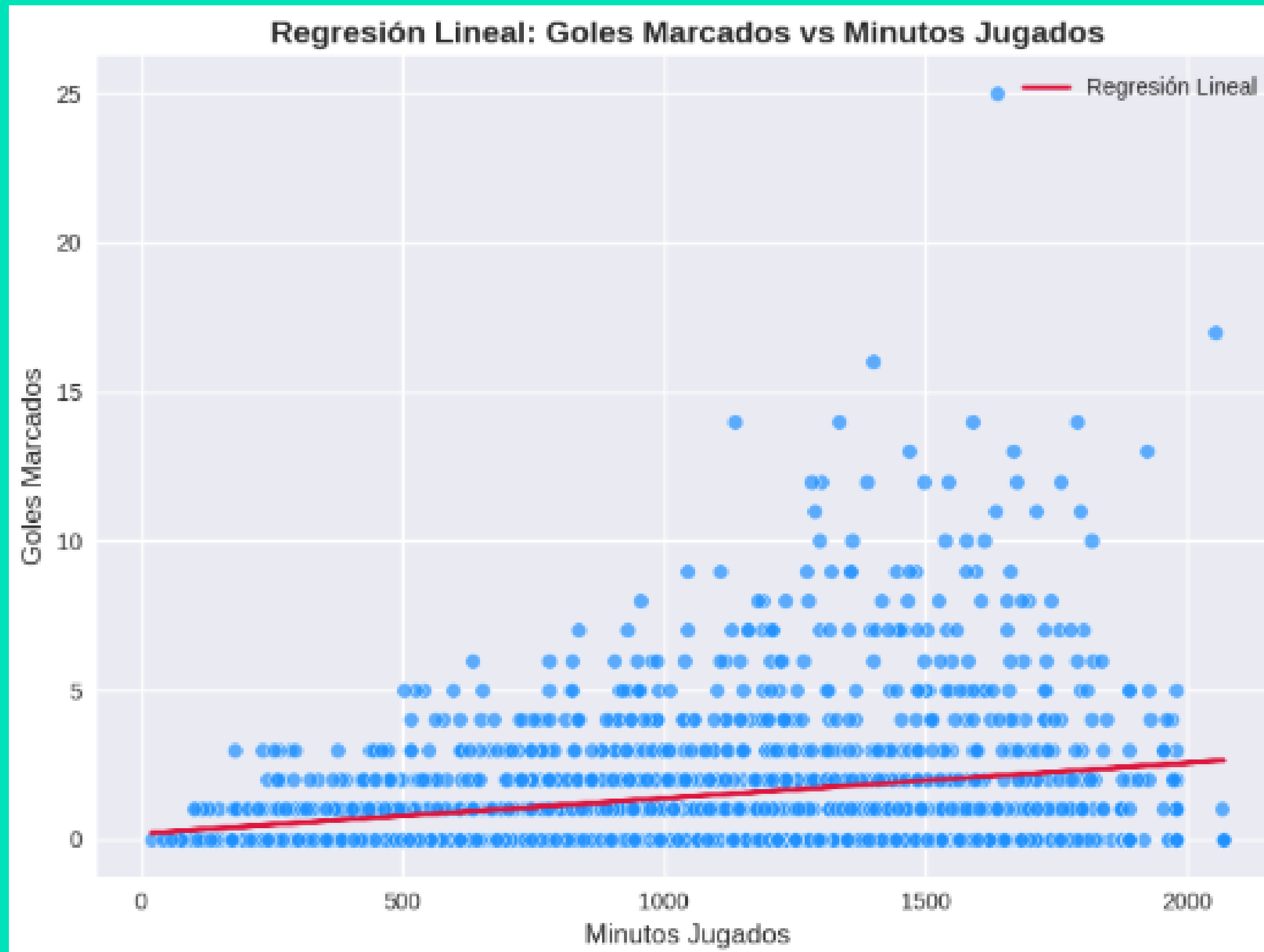
Resultados:

- Valor p = 0.0000 < 0.05
- Estadístico t observado = 13.3733 > t crítico = 1.9630

Conclusión: Se rechaza H_0 , indicando que la media de goles de los defensores es significativamente diferente de 1.

Relevancia: Esto refuerza la idea de que los defensores suelen marcar menos de un gol por temporada, destacando su función principal en la defensa.

Regresión Lineal Simple



Regresión Lineal Simple

El gráfico de dispersión muestra la relación entre minutos jugados y goles marcados, con una línea de regresión que indica una leve tendencia positiva: más minutos tienden a asociarse con más goles, aunque la relación es débil. Hay alta dispersión, con muchos jugadores que, pese a superar los 1000 minutos, no marcan más de 5 goles. También se observan valores atípicos, como jugadores con más de 10 goles, probablemente delanteros destacados. En conclusión, aunque jugar más minutos ofrece más oportunidades, el rendimiento goleador depende de múltiples factores adicionales, y la regresión lineal solo identifica una tendencia general.

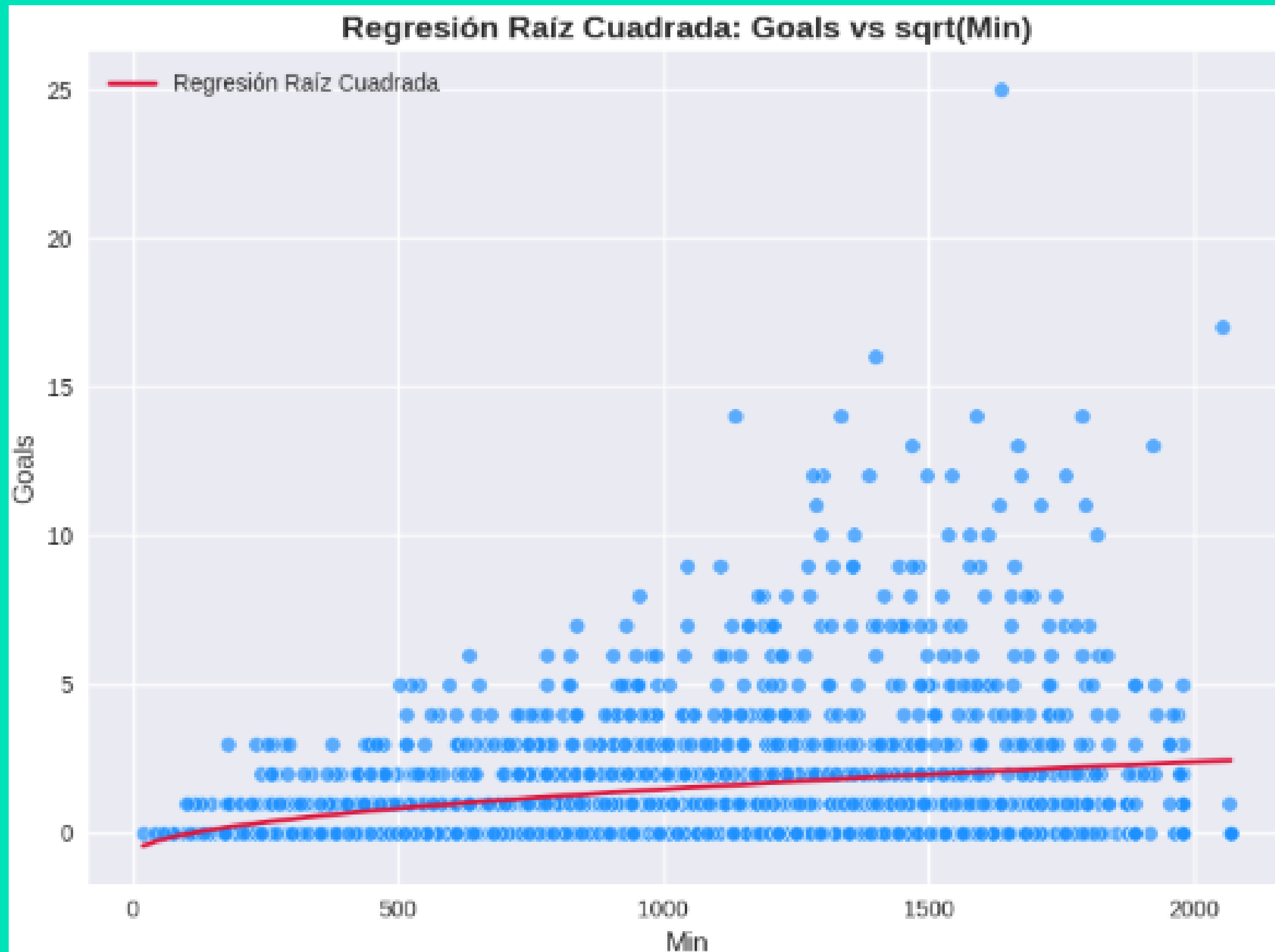
Regresión Con Transformación Logarítmica



Regresión Con Transformación Logarítmica

El gráfico ilustra la relación entre goles marcados y minutos jugados usando una función logarítmica, mostrando que la regresión logarítmica captura mejor el aumento inicial de goles con pocos minutos. A partir de los 1000 minutos, la curva se estabiliza, indicando un rendimiento decreciente. Comparado con la regresión lineal, el modelo logarítmico es más realista al reflejar un crecimiento moderado en contextos deportivos. En conclusión, la regresión logarítmica es ideal para modelar la relación entre minutos y goles, especialmente útil para jugadores con menos tiempo de juego.

Regresión Con Transformación de Raíz Cuadrada



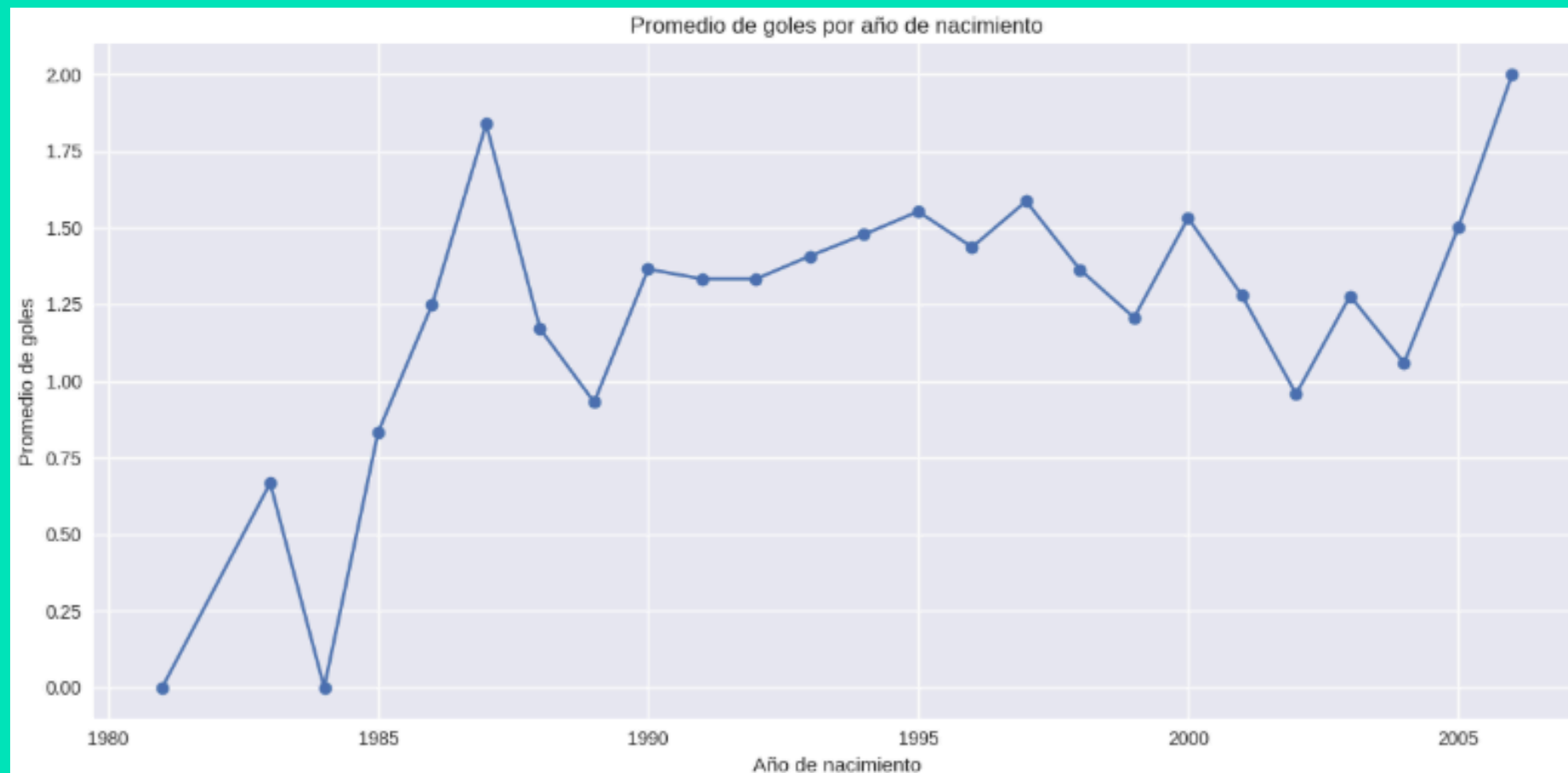
Regresión Con Transformación de Raíz Cuadrada

Análisis de Regresión Raíz Cuadrada: Goles vs $\sqrt{\text{Minutos}}$

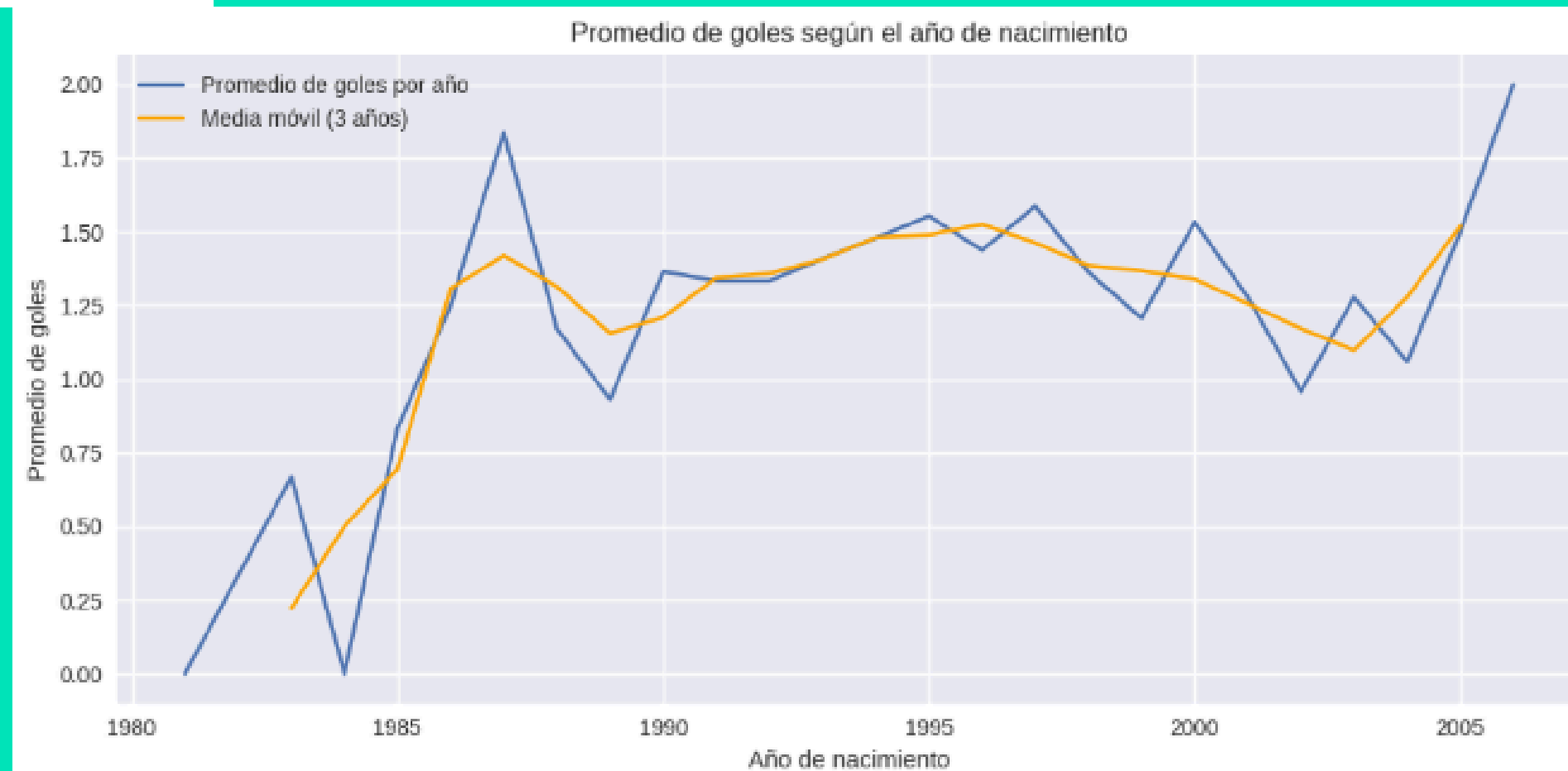
Este gráfico muestra la relación entre los goles y la raíz cuadrada de los minutos jugados, junto con una curva ajustada.

Observaciones Clave:

- **Crecimiento Suave:** La raíz cuadrada permite un aumento progresivo de los goles con los minutos, sin aplanarse bruscamente.
- **Modelo Intermedio:** Ofrece un equilibrio entre crecimiento lineal y logarítmico, útil para tasas decrecientes menos pronunciadas.
- **Minutos Altos:** Captura cierta progresión en jugadores con más tiempo, aunque no ajusta completamente la dispersión en los extremos.



Analisis de Serie de Tiempo



Analisis de Serie de Tiempo

Crecimiento desde los años 80: Los jugadores nacidos a principios de los 80 tienen menos goles en promedio, con gran variabilidad entre 1980 y 1986.

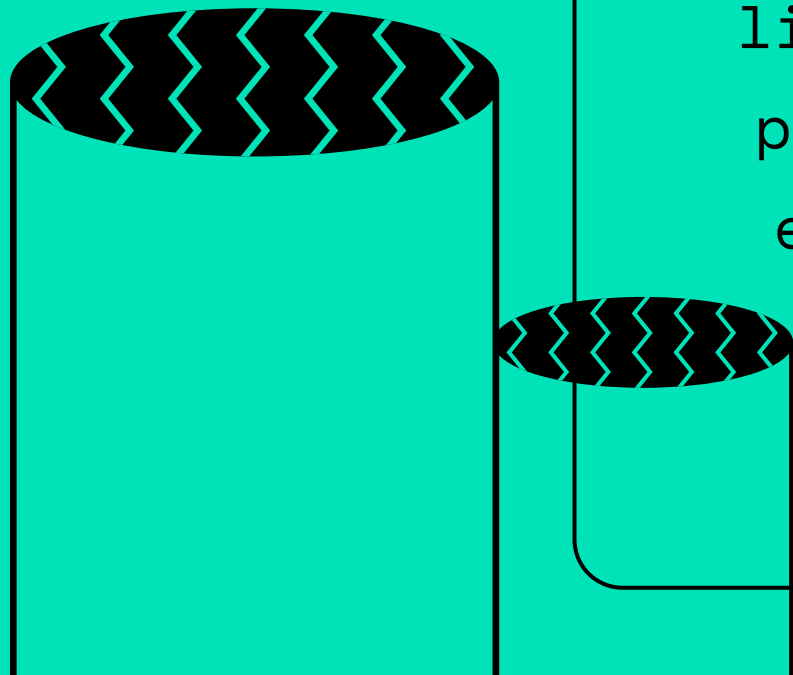
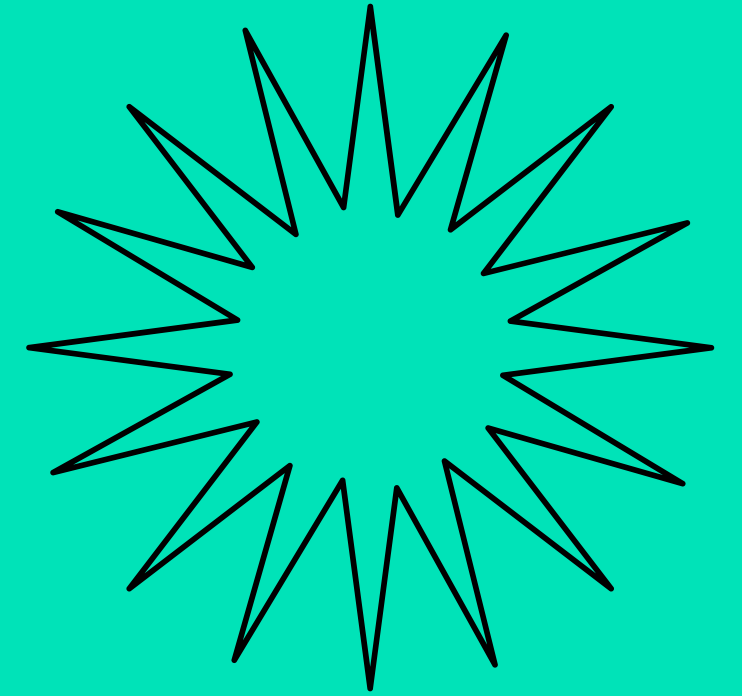
Pico en 1987: Este año destaca con un promedio de casi 1.9 goles por jugador, posiblemente por una generación más ofensiva o con más tiempo de juego.

Estabilización en los 90s: Los nacidos en los 90s muestran promedios entre 1.3 y 1.6 goles, con menor variación, indicando un rendimiento más homogéneo.

Alza después del 2000: Desde 2004-2005, el promedio sube a 2 goles por jugador, reflejando jugadores jóvenes más talentosos o con mayor protagonismo ofensivo.

Conclusiones.

Este análisis estadístico confirma que el rendimiento de los jugadores de fútbol es multifacético y fuertemente influenciado por su posición en el campo. Si bien los principios estadísticos como la Ley de los Grandes Números y el Teorema del Límite Central se validan en este conjunto de datos, la complejidad de métricas como el gol requiere un análisis cuidadoso y a menudo no lineal. La diferenciación del rendimiento por posición, validada por pruebas estadísticas, es fundamental para una comprensión precisa del deporte.





GRACIAS