

Análisis del Impacto del xG *Delta* y *PPDA* en el Rendimiento de Equipos de Fútbol: Visualización y Evaluación Comparativa

Sergio Soler Rocha

Universidad Nacional de Educación a Distancia

Resumen

Este estudio examina la influencia de las métricas avanzadas xG *Delta* y *PPDA* en el rendimiento de los equipos de fútbol, con un enfoque en su relación con la clasificación final en la liga. A lo largo del análisis, se realizaron múltiples visualizaciones que permitieron identificar tendencias significativas entre estas métricas y el éxito de los equipos.

El xG *Delta*, que mide la diferencia entre los goles esperados y los goles reales, mostró una correlación positiva con las posiciones altas en la tabla, indicando que los equipos con un xG *Delta* más alto tienden a rendir mejor. Por otro lado, el *PPDA*, que cuantifica la intensidad de la presión defensiva, presentó una correlación negativa, sugiriendo que un *PPDA* bajo, asociado a una mayor presión defensiva, se relaciona con un mejor rendimiento.

Además, se desarrolló un enfoque visual que agrupa los datos por posiciones en la liga, promediando las métricas para reducir la variabilidad y destacar las tendencias generales. Este método permitió crear visualizaciones más claras y comprensibles, incluso para personas no especializadas en el análisis de datos deportivos.

Las conclusiones del estudio subrayan la importancia de ambas métricas en la evaluación del rendimiento de los equipos y sugieren que su combinación ofrece una perspectiva más completa que el análisis de cada métrica por separado. Finalmente, se proponen posibles mejoras y extensiones del trabajo, incluyendo la incorporación de más ligas y temporadas, así como la integración de otras métricas avanzadas para enriquecer el análisis.

1. Introducción

En el análisis del fútbol moderno, las métricas avanzadas han ganado una importancia creciente, permitiendo a analistas y entrenadores obtener una visión más profunda y precisa del rendimiento de los equipos y jugadores. Entre estas métricas, el **xG Delta** (Expected Goals Delta) y el **PPDA** (Passes Per Defensive Action) destacan por su capacidad de medir aspectos críticos del juego que no siempre son evidentes en las estadísticas tradicionales.

El **xG Delta** es una métrica que compara la cantidad de goles que un equipo realmente anota con la cantidad de goles esperados (xG), basados en la calidad de las oportunidades de gol generadas. Un xG Delta positivo indica que un equipo ha sido más eficiente en la conversión de sus oportunidades en goles, mientras que un xG Delta negativo sugiere lo contrario.

Por otro lado, el **PPDA** mide la intensidad defensiva de un equipo, calculando el número de pases permitidos al oponente antes de realizar una acción defensiva como una entrada o intercepción. Un valor de *PPDA* bajo indica una alta presión sobre el equipo contrario, lo que puede ser un indicativo de un estilo de juego agresivo y proactivo.

Este trabajo tiene como objetivo estudiar la relevancia de estas dos métricas, xG Delta y *PPDA*, en relación con el rendimiento general de los equipos en diversas ligas europeas. A través de un análisis exhaustivo y visualizaciones claras, se busca demostrar cómo estas métricas pueden influir en la clasificación final de los equipos en sus respectivas ligas y cómo su comprensión puede ofrecer ventajas competitivas.

El objetivo final de este estudio es ofrecer una visualización que evidencie de manera contundente la importancia del xG Delta y el *PPDA*, subrayando su papel en la evaluación del rendimiento de los equipos. Este enfoque permitirá no solo apreciar la eficacia y la intensidad defensiva de los equipos, sino también comprender mejor los factores que contribuyen a su éxito o fracaso a lo largo de una temporada.

1.1. PPDA: Passes Per Defensive Action

El **PPDA** (Passes Per Defensive Action) es una métrica avanzada utilizada para evaluar la intensidad defensiva de un equipo. Se define como el número de pases que un equipo contrario completa antes de que el equipo analizado realice una acción defensiva significativa, como una entrada, intercepción, o cualquier otra acción que interrumpa la posesión del oponente.

$$\text{PPDA} = \frac{\text{Número de pases completados por el oponente en su propia mitad}}{\text{Número de acciones defensivas en la mitad del oponente}} \quad (1)$$

Un valor bajo de PPDA indica que un equipo ejerce una presión alta y constante, intentando recuperar la posesión lo más cerca posible del área contraria. Por el contrario, un valor alto sugiere que el equipo permite más pases al oponente antes de intentar recuperar el balón, lo que puede reflejar un enfoque defensivo más pasivo.

El cálculo del PPDA se realiza contando el número de pases completados por el equipo contrario en su propia mitad del campo, y dividiendo ese número por las acciones defensivas realizadas en esa misma área. Esta métrica se utiliza para analizar el estilo de juego de un equipo y su eficacia en la presión alta, siendo un indicador clave de la intensidad defensiva.

1.2. xG: Expected Goals

El **xG** (Expected Goals) es una métrica estadística utilizada para evaluar la probabilidad de que un disparo termine en gol, basándose en las características de la oportunidad de gol. Esta métrica se ha convertido en una herramienta crucial para analizar el rendimiento de un equipo o jugador más allá de los goles marcados.

El xG de un disparo se calcula utilizando un modelo que tiene en cuenta varios factores, tales como:

- **Posición del disparo:** Cuanto más cerca y centrado esté el disparo con respecto a la portería, mayor será el valor de xG.
- **Tipo de disparo:** Disparos de pie, cabezazos, o tiros de volea tienen diferentes probabilidades de éxito.
- **Situación del juego:** Si el disparo proviene de una jugada en movimiento, un balón parado, un contraataque, etc.
- **Presión defensiva:** La proximidad y la cantidad de defensores que rodean al jugador que dispara.
- **Ángulo de disparo:** El ángulo desde el cual se realiza el disparo también influye en la probabilidad de marcar.
- **Tipo de asistencia:** Si el disparo fue asistido por un pase corto, un centro, un pase largo, etc.

Cada uno de estos factores se considera en un modelo estadístico que asigna un valor xG a cada disparo, que representa la probabilidad de que el disparo se convierta en gol. Por ejemplo, un disparo desde muy cerca de la portería podría tener un xG de 0.75, lo que indica que, de cada 10 disparos similares, se esperaría que 7.5 terminen en gol.

1.3. xG Delta: Expected Goals Delta

El **xG Delta** es una métrica que compara los goles esperados (**xG**) con los goles realmente marcados por un equipo. La métrica xG evalúa la calidad de las oportunidades de gol que un equipo genera, asignando una probabilidad de gol a cada tiro en función de varios factores, como la posición del disparo, el tipo de asistencia, la presión defensiva, entre otros.

$$\text{xG Delta} = \text{Goles Marcados} - \text{xG} \quad (2)$$

Un **xG Delta** positivo indica que un equipo ha marcado más goles de los que se esperaría según la calidad de sus oportunidades, lo cual puede deberse a una finalización excepcional o a una cierta

dosis de suerte. Por otro lado, un **xG Delta** negativo sugiere que un equipo ha sido ineficiente en la conversión de sus oportunidades en goles.

El cálculo del xG Delta se realiza restando el valor total de xG de un equipo (suma de xG de cada oportunidad de gol) del número total de goles realmente marcados en un periodo de tiempo, como una temporada o un partido. Esta métrica es esencial para entender si un equipo está obteniendo resultados por encima o por debajo de sus expectativas estadísticas, y puede ser utilizada para evaluar la sostenibilidad de su rendimiento a lo largo del tiempo.

1.4. Descripción del Dataset

El dataset utilizado en este estudio fue obtenido de [Kaggle](#) y contiene información detallada sobre equipos, partidos, jugadores y estadísticas avanzadas en varias ligas de fútbol. Este dataset es fundamental para analizar y comprender las métricas **xG Delta** y **PPDA**, así como su impacto en el rendimiento de los equipos a lo largo de las temporadas. A continuación, se describen los principales archivos y sus contenidos:

- **teams.csv**: Este archivo contiene la información básica sobre los equipos participantes. Incluye:
 - *teamID*: Identificador único del equipo.
 - *name*: Nombre del equipo.
- **teamstats.csv**: Este archivo almacena estadísticas avanzadas de los equipos para cada partido, incluyendo:
 - *gameID*: Identificador único del partido.
 - *teamID*: Identificador del equipo.
 - *season*: Temporada en la que se jugó el partido.
 - *date*: Fecha del partido.
 - *location*: Indica si el equipo jugó como local (*h*) o visitante (*a*).
 - *goals*: Número de goles marcados por el equipo en el partido.
 - *xGoals*: Goles esperados basados en la calidad y cantidad de las oportunidades.
 - *shots*: Total de disparos realizados.
 - *shotsOnTarget*: Total de disparos a puerta.
 - *deep*: Número de veces que el equipo llegó a las zonas más avanzadas del campo.
 - *ppda*: *Passes per Defensive Action*, métrica que mide la intensidad defensiva del equipo.
 - *fouls*: Número de faltas cometidas.
 - *corners*: Número de saques de esquina obtenidos.
 - *yellowCards*: Número de tarjetas amarillas recibidas.
 - *redCards*: Número de tarjetas rojas recibidas.
 - *result*: Resultado del partido desde la perspectiva del equipo.
- **shots.csv**: Este archivo contiene datos detallados sobre los disparos realizados en los partidos, incluyendo:
 - *gameID*: Identificador del partido.
 - *shooterID*: Identificador del jugador que realizó el disparo.
 - *assisterID*: Identificador del jugador que asistió el disparo (si aplica).
 - *minute*: Minuto del partido en que se realizó el disparo.
 - *situation*: Situación del disparo (juego abierto, balón parado, etc.).
 - *lastAction*: Última acción antes del disparo (pase, regate, etc.).
 - *shotType*: Tipo de disparo (tiro, cabezazo, etc.).
 - *shotResult*: Resultado del disparo (gol, atajado, fuera, etc.).

- *xGoal*: Probabilidad de que el disparo se convirtiera en gol.
 - *positionX*: Coordenada X de la posición del disparo en el campo.
 - *positionY*: Coordenada Y de la posición del disparo en el campo.
- **players.csv**: Este archivo incluye información sobre los jugadores, con los siguientes campos:
 - *playerID*: Identificador único del jugador.
 - *name*: Nombre del jugador.
 - **leagues.csv**: Este archivo detalla la información sobre las ligas de fútbol, con las siguientes columnas:
 - *leagueID*: Identificador único de la liga.
 - *name*: Nombre de la liga.
 - *understatNotation*: Notación utilizada por Understat para identificar la liga.
 - **games.csv**: Este archivo contiene datos sobre los partidos jugados, incluyendo:
 - *gameID*: Identificador único del partido.
 - *leagueID*: Identificador de la liga en la que se jugó el partido.
 - *season*: Temporada en la que se jugó el partido.
 - *date*: Fecha del partido.
 - *homeTeamID*: Identificador del equipo local.
 - *awayTeamID*: Identificador del equipo visitante.
 - *homeGoals*: Goles marcados por el equipo local.
 - *awayGoals*: Goles marcados por el equipo visitante.
 - *homeProbability*: Probabilidad de victoria del equipo local.
 - *drawProbability*: Probabilidad de empate.
 - *awayProbability*: Probabilidad de victoria del equipo visitante.
 - *homeGoalsHalfTime*: Goles marcados por el equipo local en la primera mitad.
 - *awayGoalsHalfTime*: Goles marcados por el equipo visitante en la primera mitad.

Además, este archivo incluye varias columnas adicionales relacionadas con probabilidades y cuotas de apuestas proporcionadas por diferentes casas de apuestas.

- **appearances.csv**: Este archivo recoge los datos de las apariciones de los jugadores en los partidos, con los siguientes campos:
 - *gameID*: Identificador del partido.
 - *playerID*: Identificador del jugador.
 - *goals*: Goles marcados por el jugador.
 - *ownGoals*: Goles en propia puerta.
 - *shots*: Disparos realizados por el jugador.
 - *xGoals*: Goles esperados para el jugador.
 - *xGoalsChain*: Contribución del jugador al xG en toda la secuencia de juego.
 - *xGoalsBuildup*: Contribución del jugador al xG en la construcción de la jugada.
 - *assists*: Asistencias del jugador.
 - *keyPasses*: Pases clave realizados.
 - *xAssists*: Asistencias esperadas.
 - *position*: Posición del jugador en el campo.
 - *positionOrder*: Orden de la posición del jugador.
 - *yellowCard*: Indica si el jugador recibió tarjeta amarilla.

- *redCard*: Indica si el jugador recibió tarjeta roja.
- *time*: Minutos jugados.
- *substituteIn*: Minuto en que el jugador ingresó como sustituto.
- *substituteOut*: Minuto en que el jugador fue sustituido.
- *leagueID*: Identificador de la liga en la que se jugó el partido.

Este dataset es una rica fuente de información para el análisis avanzado de datos en el fútbol, permitiendo la exploración y evaluación de diversas métricas clave como el **xG Delta** y el **PP-DA**, que son fundamentales para comprender el rendimiento de los equipos a lo largo de varias temporadas.

2. Preprocesado

2.1. Tratamiento de Valores Nulos

Durante el preprocesado de los datos, se identificaron varios casos de valores nulos en diferentes columnas de las tablas. A continuación, se detalla cómo se manejaron estos valores nulos:

2.1.1. Tabla *shots* Columna *assisterID*

La columna *assisterID* en la tabla *shots* contiene 84344 valores nulos, lo que indica que el tiro no tuvo ningún asistente. Para manejar estos valores nulos, se rellenaron con el valor -1, indicando explícitamente la ausencia de un asistente.

```
# Cargar datos (especificando el encoding si es necesario)
shots = pd.read_csv('shots.csv', encoding='latin1')

# Rellenar valores nulos en assisterID con -1 para indicar que no hubo asistente
shots['assisterID'].fillna(-1, inplace=True)

# Verificar que los valores nulos han sido tratados
print(shots['assisterID'].isnull().sum())
```

2.1.2. Tabla *shots* Columna *lastAction*

La columna *lastAction* en la tabla *shots* contiene 36896 valores nulos. Los valores posibles para esta columna son: *Pass*, *Cross*, *None*, *Aerial* y *Standard*. Los valores nulos indican que no se registró ninguna acción específica antes del tiro. Estos valores nulos se rellenaron con *None*.

```
# Cargar datos (especificando el encoding si es necesario)
shots = pd.read_csv('shots.csv', encoding='latin1')

# Rellenar valores nulos en lastAction con 'None' para indicar que no hubo una
acción específica
shots['lastAction'].fillna('None', inplace=True)

# Verificar que los valores nulos han sido tratados
print(shots['lastAction'].isnull().sum())
```

2.1.3. Tabla *teamstats* columna *yellowCards*

En la tabla *teamstats*, la columna *yellowCards* contiene un valor nulo. Para manejar este valor nulo, se decidió rellenarlo con la mediana de la columna, ya que la mediana es una medida robusta y no se ve afectada por valores atípicos.

```
# Cargar datos (especificando el encoding si es necesario)
teamstats = pd.read_csv('teamstats.csv', encoding='latin1')

# Calcular la mediana de la columna yellowCards
median_yellow_cards = teamstats['yellowCards'].median()

# Rellenar el valor nulo con la mediana
teamstats['yellowCards'].fillna(median_yellow_cards, inplace=True)

# Verificar que los valores nulos han sido tratados
print(teamstats['yellowCards'].isnull().sum())
```

2.1.4. Tabla *games* Columnas Cuotas de Apuestas

En la tabla *games*, varias columnas representan las cuotas de apuestas de diferentes casas de apuestas para los resultados de los partidos. Las columnas con valores nulos y el número de nulos en cada una son las siguientes:

- **Bet365:** *B365H* (5), *B365D* (5), *B365A* (5)
- **Bwin:** *BWH* (3), *BWD* (3), *BWA* (3)
- **Interwetten:** *IWH* (18), *IWD* (18), *IWA* (18)
- **Pinnacle Sports:** *PSH* (20), *PSD* (20), *PSA* (20)
- **William Hill:** *WHH* (6), *WHD* (6), *WHA* (6)
- **BetVictor:** *VCH* (4), *VCD* (4), *VCA* (4)
- **Pinnacle Sports (Cierre):** *PSCH* (2), *PSCD* (2), *PSCA* (2)

Para manejar estos valores nulos, se decidió eliminar dichas columnas, ya que, serán irrelevantes para el propósito del estudio.

```
# Eliminar columnas con valores nulos en el DataFrame 'games'
games = games.dropna(axis=1)

# Verificar las columnas restantes
print(games.columns)
```

2.2. Transformación de Variables

En esta sección se realizaron diversas transformaciones de variables para facilitar el análisis de los datos. A continuación, se detalla la transformación aplicada a la tabla *shots*.

2.2.1. Tabla *shots* Columna *xg_delta*

Para la tabla *shots*, se calculó una nueva columna llamada *xg_delta*, que representa la diferencia entre los goles esperados (*xGoal*) y el resultado real del tiro.

El siguiente código muestra cómo se implementó esta transformación:

```
# Calcular el xG Delta en el DataFrame shots
shots['xg_delta'] = shots.apply(
    lambda row: 1 - row['xGoal'] if row['shotResult'] == 'Goal' else -row['xGoal'],
    axis=1
)
```

Se realizó una búsqueda de datos incoherentes y duplicados, entre otros. Sin embargo, se constató que las tablas estaban bastante limpias en ese aspecto, por lo que se decidió pasar al siguiente apartado.

3. Procesado y Análisis

3.1. Rendimiento en términos absolutos a lo largo de las temporadas

Aunque inicialmente no estaba previsto dentro de los objetivos marcados para este estudio y visualización, finalmente decidí incluir este apartado porque, al explorar diversas visualizaciones, me di cuenta de la importancia de analizar las métricas mencionadas en el apartado de objetivos. Fue precisamente al experimentar con estas visualizaciones que surgió la idea central de este estudio. En otras palabras, este proceso de exploración visual no solo complementó el análisis, sino que fue el catalizador que dio origen a la propia estructura y enfoque del estudio.

3.1.1. Análisis de los goles por temporada

Por ser el gol el elemento central y la culminación de todo esfuerzo en el fútbol, decidí iniciar este estudio con un análisis simple de los goles a lo largo de las temporadas. Comprender la distribución y evolución de los goles es fundamental para desentrañar las dinámicas del juego y evaluar el rendimiento de los equipos.

Inicialmente, se calculó el promedio de goles por partido para cada temporada. Esta métrica fue seleccionada porque permite comparar de manera equitativa temporadas con diferentes números de partidos. Posteriormente, se decidió utilizar un gráfico de barras para visualizar estos datos, dado que este formato facilita la comparación directa entre las temporadas y destaca las variaciones en el promedio de goles de una manera clara y comprensible.

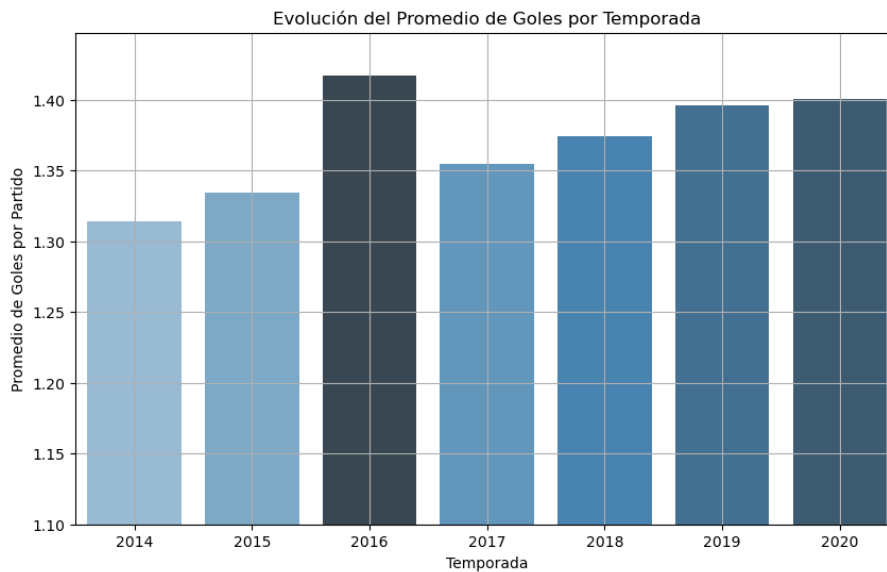


Figura 1: Evolución del Promedio de Goles por Temporada

Se observa una tendencia ascendente en general; sin embargo, destaca la temporada 2016 como la que tuvo más goles por partido.

3.1.2. Promedio de Goles con y sin Asistencia por Temporada

La asistencia es una métrica fundamental en el análisis del rendimiento de los jugadores y equipos, ya que mide la participación directa en la creación de goles.

Para visualizar el promedio de goles con y sin asistencia por temporada, se realizaron las siguientes transformaciones sobre los datos originales. Primero, se calcularon los promedios de goles asistidos y no asistidos por temporada. Posteriormente, ambos conjuntos de datos fueron combinados en un único DataFrame que permite la comparación directa entre estas dos categorías.

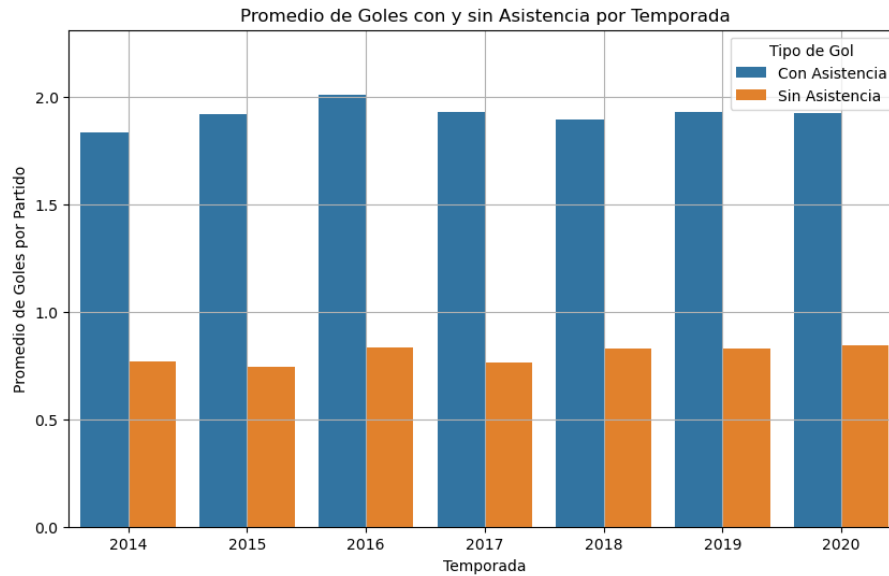


Figura 2: Promedio de Goles con y sin Asistencia por Temporada

3.1.3. Análisis de Empates

El empate es probablemente el resultado menos preferido por los aficionados, ya que deja la sensación de una resolución incompleta. No obstante, el análisis de los empates es importante para entender las dinámicas del juego y las tácticas utilizadas por los equipos.

Probabilidad de Empate según el Número Total de Goles en un Partido Para analizar cómo la probabilidad de empate varía en función del número total de goles anotados en un partido, se realizaron las siguientes transformaciones: primero, se calculó el número total de goles por partido sumando los goles anotados por ambos equipos. Luego, se identificaron los partidos que terminaron en empate y se calculó la probabilidad de empate para cada número total de goles. Finalmente, se utilizó un gráfico de barras para visualizar estos datos.

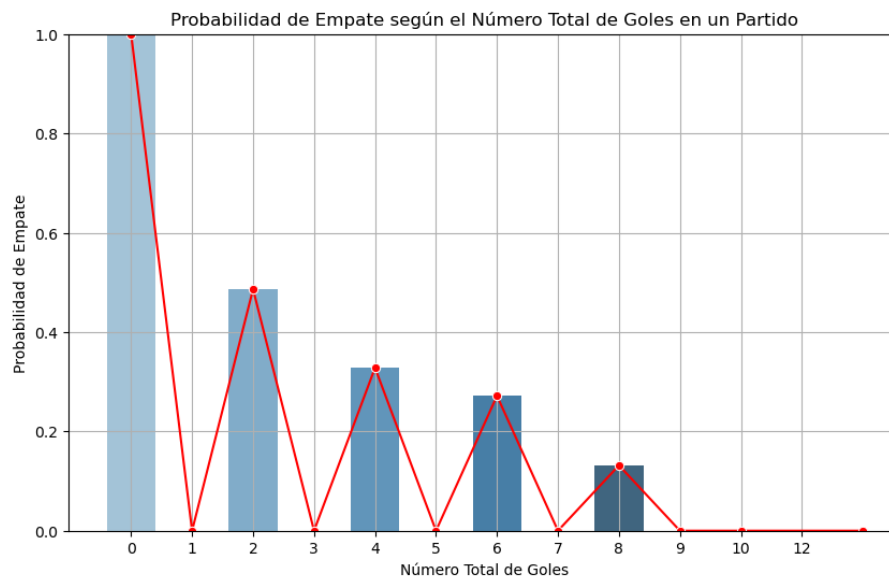


Figura 3: Probabilidad de Empate según el Número Total de Goles en un Partido

Porcentaje de Empates por Temporada El análisis del porcentaje de empates por temporada requirió calcular el número total de partidos que terminaron en empate cada temporada y dividirlo

por el número total de partidos jugados en esa misma temporada, obteniendo así el porcentaje de empates.

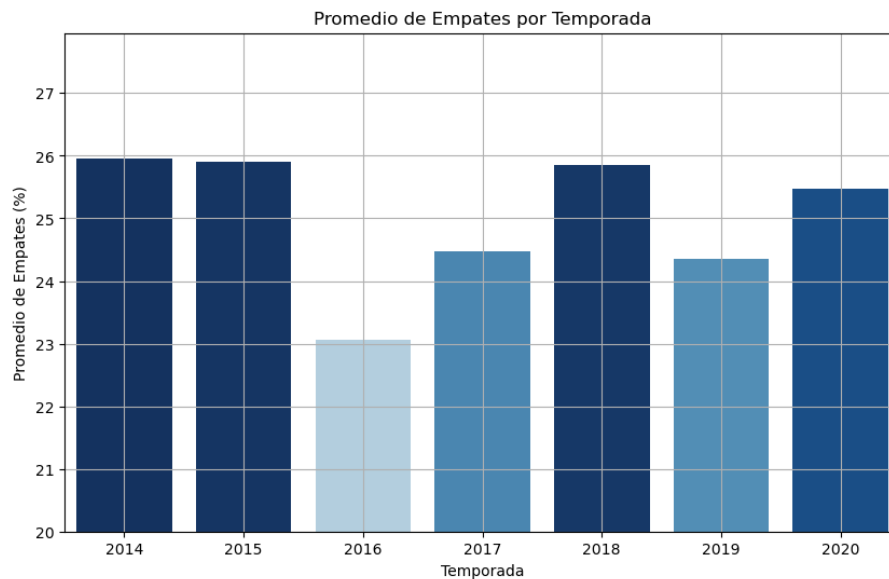


Figura 4: Porcentaje de Empates por Temporada

Después de visualizar las gráficas anteriores, llegué a la conclusión de que, en términos generales, la temporada 2016 fue un buen año en cuanto a la cantidad de goles y la baja cantidad de empates en comparación con otras temporadas. Esto me motivó a analizar más métricas a lo largo de las temporadas, como el promedio de faltas, corners, tarjetas, entre otras. Sin embargo, no he incluido estos análisis en el informe, ya que se desvían de los objetivos principales y podrían diluir el enfoque central del estudio.

3.1.4. Análisis de xGoals (xG Delta)

Para calcular el xG Delta por temporada, primero se obtuvo el promedio de goles y xGoals por temporada. Luego, se calculó la diferencia entre estos dos valores, resultando en el xG Delta. Para visualizar esta diferencia, se utilizó un gráfico de líneas que permite observar cómo ha evolucionado esta métrica a lo largo de las temporadas. Se añadió una línea de referencia en 0 para resaltar las temporadas en las que el equipo estuvo por encima o por debajo de las expectativas de goles según los xGoals.

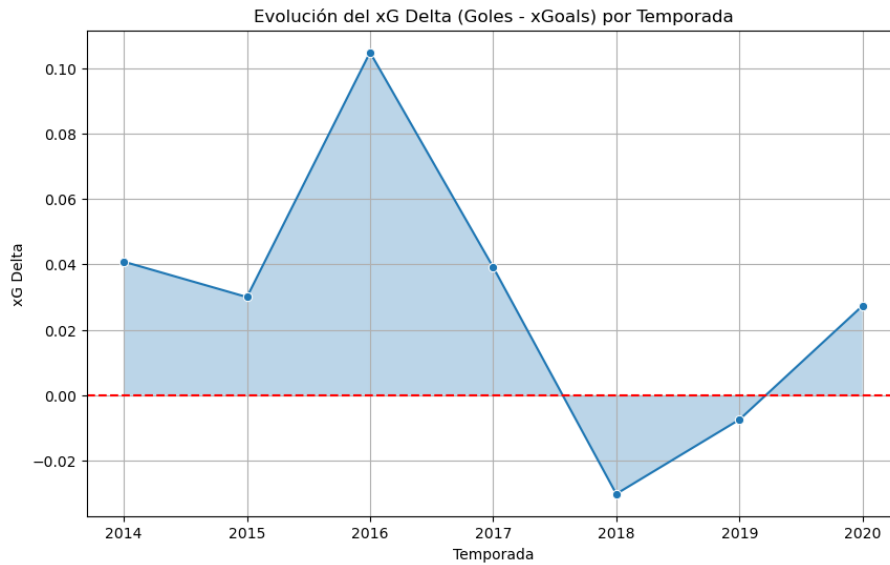


Figura 5: Evolución del xG Delta (Goles - xGoals) por Temporada

Fue con esta gráfica (y con la siguiente) donde surgió el objetivo del presente trabajo. Anteriormente, había concluido que la temporada 2016 fue destacada en comparación con las otras analizadas, tanto en términos de goles como de bajos índices de empates. Entonces, al observar que en la temporada 2016 se registra el pico más alto del xG Delta, me surgió la pregunta: ¿Cómo afecta esta métrica, que no es muy conocida entre el gran público, al rendimiento de los clubes en una temporada? Al investigar, encontré artículos que explicaban en qué consiste la métrica, pero no un análisis detallado y visualmente claro que respondiera a mi inquietud. Por ello, decidí enfocarme en esta tarea.

3.1.5. Análisis de PPDA (Passes per Defensive Action) por temporada

Para calcular el PPDA por temporada, primero se obtuvo el promedio de PPDA por temporada para cada equipo. Luego, se representó este valor en un gráfico de líneas, lo que permite observar cómo ha evolucionado esta métrica a lo largo de las temporadas.

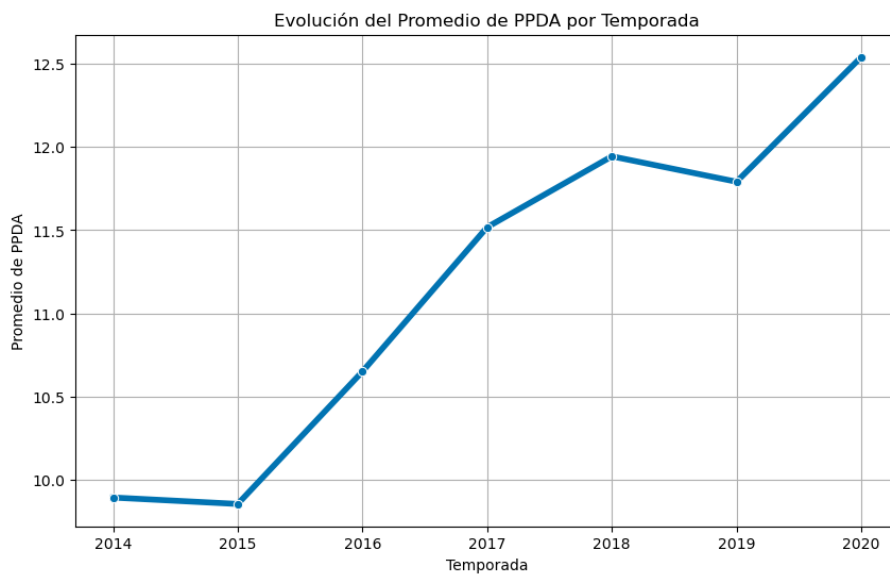


Figura 6: Evolución del PPDA por Temporada

En este caso, al observar la clara tendencia ascendente del promedio de PPDA por temporada, me surgió la misma pregunta que anteriormente: ¿Existe alguna relación entre el rendimiento de

los equipos en una temporada y la métrica PPDA? ¿Hay también una relación con la métrica xG Delta? ¿Cómo pueden ambas influir en el rendimiento?

3.2. Comprendiendo el xG Delta con Visualizaciones

El análisis y la comprensión del xG Delta se benefician significativamente del uso de visualizaciones. Estas permiten interpretar mejor cómo los equipos y jugadores se desempeñan en relación con las expectativas de goles, proporcionando una visión clara de las áreas en las que se superan o no las expectativas.

3.2.1. Visualización del xG Delta en un Campo de Fútbol

Para entender mejor el concepto del xG Delta, se visualizó esta métrica en el contexto espacial de un campo de fútbol. Se seleccionó un partido aleatorio y se representaron los disparos realizados por el equipo local en ese juego. Los disparos se posicionaron en el campo de acuerdo con las coordenadas (X, Y) y se colorearon en función de su xG Delta, lo que permite identificar visualmente las oportunidades de gol que superaron o no las expectativas.

Para ello, se realizó un merge entre los datos de disparos (shots) y los datos de partidos (games) basándose en el identificador del partido (gameID). Posteriormente, se seleccionó un partido aleatorio y se filtraron los disparos del equipo local en ese partido específico. Finalmente, se generó la visualización de los disparos en el campo de fútbol.

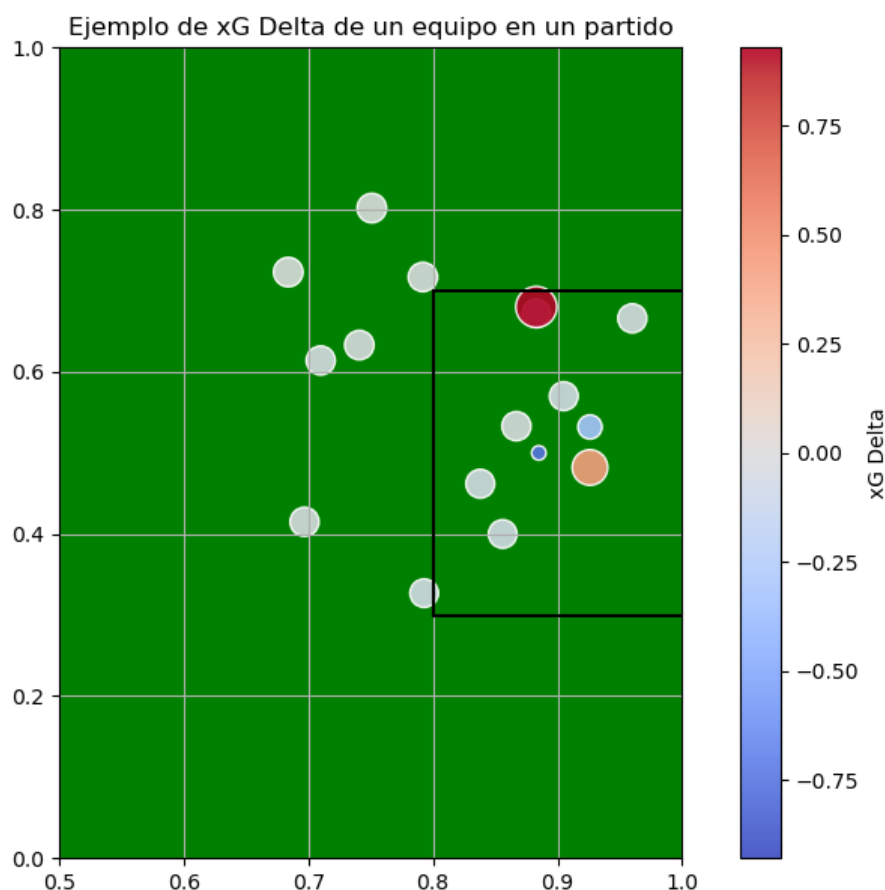


Figura 7: Visualización del xG Delta de un equipo local en un partido aleatorio, mostrando la ubicación y el resultado de los disparos en el campo de fútbol.

En la imagen se pueden observar los diferentes disparos a puerta realizados por el equipo local. Se destacan dos goles, siendo uno de ellos especialmente interesante: el disparo que aparece en rojo intenso, lo que indica que, a pesar de no ser una situación favorable, se logró convertir en gol,

resultando en un xG Delta alto. El otro gol, con un xG Delta mucho más bajo, sugiere que era una oportunidad bastante favorable para el atacante, lo cual es intuitivo dada su cercanía a la portería.

Además, se observan dos ocasiones claras que no se concretaron en gol, reflejadas por sus xG Delta negativos, lo que indica que, a pesar de ser buenas oportunidades, no fueron materializadas. El resto de los disparos muestran xG Deltas negativos pero cercanos a cero, lo que sugiere que eran situaciones en las que no era muy probable que el disparo se convirtiera en gol.

3.2.2. Comprobación de la Eficiencia del Modelo de xG

Para evaluar la precisión del modelo que genera los datos de xG, se calculó el xG Delta, que mide la diferencia entre la probabilidad de gol asignada por el modelo y el resultado real (gol o no gol).

Con el fin de visualizar cómo se comporta el modelo a lo largo del campo, se creó un heatmap que muestra el promedio de xG Delta en diferentes zonas del terreno de juego. Se dividió el campo en una cuadrícula (en este caso, de 20x20 celdas), y se calculó el promedio de xG Delta para los disparos realizados dentro de un radio específico en cada celda. Este análisis permite identificar áreas del campo donde el modelo de xG tiende a ser más optimista o pesimista en sus predicciones.

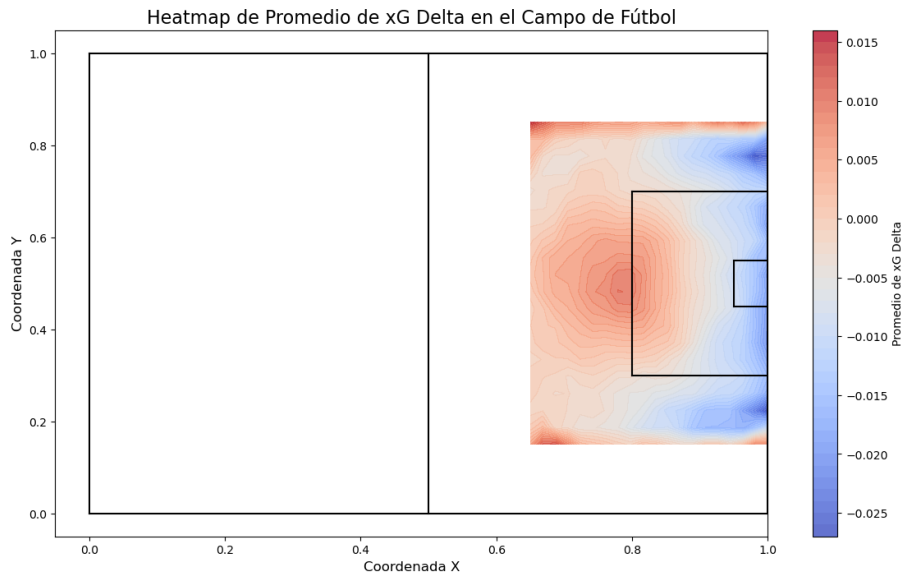


Figura 8: Heatmap de Promedio de xG Delta en el Campo de Fútbol

En el heatmap, se observan zonas donde el xG Delta tiende a ser ligeramente positivo (máximo alrededor de 0.015) y otras donde es negativo (mínimo alrededor de -0.025). Esto indica que, en promedio, el modelo es bastante preciso, con pequeñas variaciones que son esperables en cualquier modelo predictivo. Las zonas con valores más positivos pueden corresponder a situaciones donde el modelo subestima la probabilidad de gol, mientras que las zonas con valores negativos indican que el modelo sobreestima ligeramente dicha probabilidad. Estos resultados sugieren que el modelo de xG es confiable, con una precisión aceptable en la mayoría de las situaciones.

3.3. Entendiendo el Impacto del xG Delta y PPDA en el Rendimiento de los Equipos

3.3.1. Evolución del Promedio de xG Delta por Temporada para los 5 Mejores Equipos

En este análisis, se calculó el promedio total del xG Delta de los equipos para identificar a los cinco equipos con el mejor rendimiento en esta métrica. Posteriormente, se examinó la evolución de su xG Delta a lo largo de las temporadas y se comparó con sus posiciones en la clasificación de la liga.

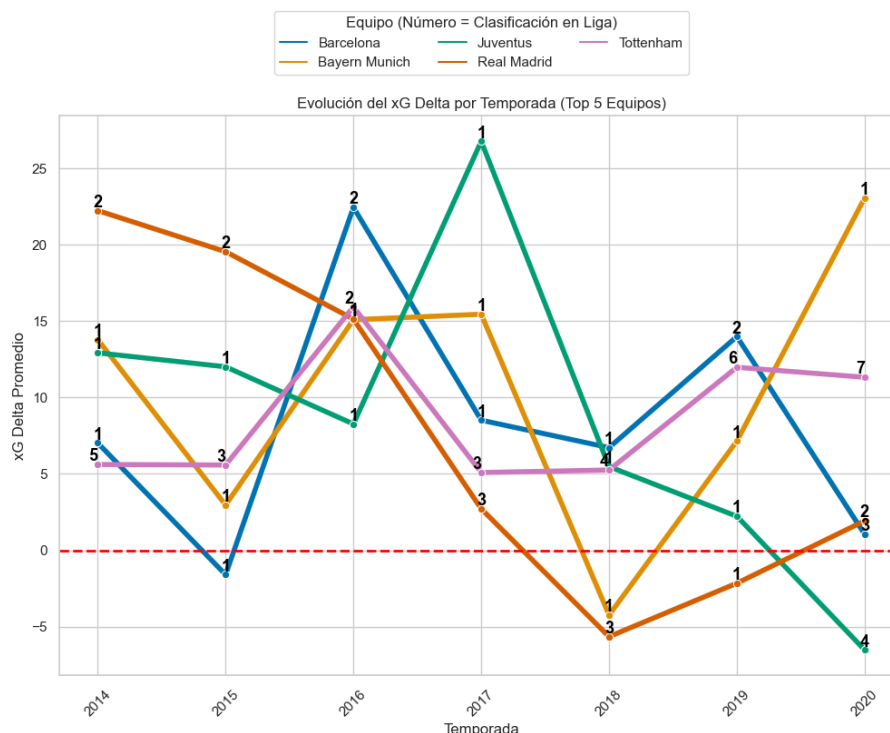


Figura 9: Evolución del Promedio de xG Delta por Temporada para los 5 Mejores Equipos

En la Figura 9, observamos que los clubes analizados son algunos de los más importantes de Europa: *Barcelona*, *Real Madrid*, *Bayern Munich*, *Juventus* y, en menor medida pero también relevante, *Tottenham*. En general, estos equipos muestran posiciones altas en la clasificación de sus respectivas ligas, y en la mayoría de las temporadas presentan un xG Delta positivo.

Sin embargo, es interesante notar que, en algunas temporadas, ciertos equipos lograron la primera posición en la liga a pesar de tener un xG Delta negativo. Esto sugiere que, aunque el xG Delta es una métrica importante, existen otros factores que también influyen significativamente en el rendimiento de los equipos. El análisis revela que, aunque los equipos con un alto xG Delta tienden a tener buenos resultados, esta métrica puede presentar dispersiones o varianzas que no siempre se correlacionan de manera lineal con la posición final en la clasificación.

Además, se realizó un análisis similar para los cinco equipos con el xG Delta más bajo. En su mayoría, estos equipos obtuvieron posiciones bajas en la tabla, con varios de ellos descendiendo a la categoría inferior. No se representaron gráficamente, ya que, al haber descensos no se puede apreciar evolución temporal.

3.3.2. Evolución del Promedio de PPDA por Temporada para los 5 Equipos con el PPDA más Bajo

En esta sección se presenta la evolución del promedio de PPDA (Passes per Defensive Action) por temporada para los cinco equipos que registraron los valores más bajos en esta métrica. El gráfico nos permite observar cómo estos equipos han desempeñado en términos defensivos a lo largo de varias temporadas.

El análisis revela la presencia de equipos de élite en Europa, como el Bayern Munich, el Barcelona y el Paris Saint Germain, que lograron posiciones destacadas en sus respectivas ligas. Sin embargo, el caso del Rayo Vallecano es particularmente interesante, ya que el equipo aparece con dos descensos y una clasificación en la media tabla. Por otro lado, el Leeds United muestra un rendimiento notable al alcanzar la novena posición en la última temporada analizada.

Estos equipos, a pesar de tener un PPDA significativamente por debajo del promedio general, han logrado diferentes grados de éxito. Este comportamiento sugiere que, aunque el PPDA es una métrica importante en la evaluación del rendimiento defensivo de un equipo, no es un factor determinante por sí solo.

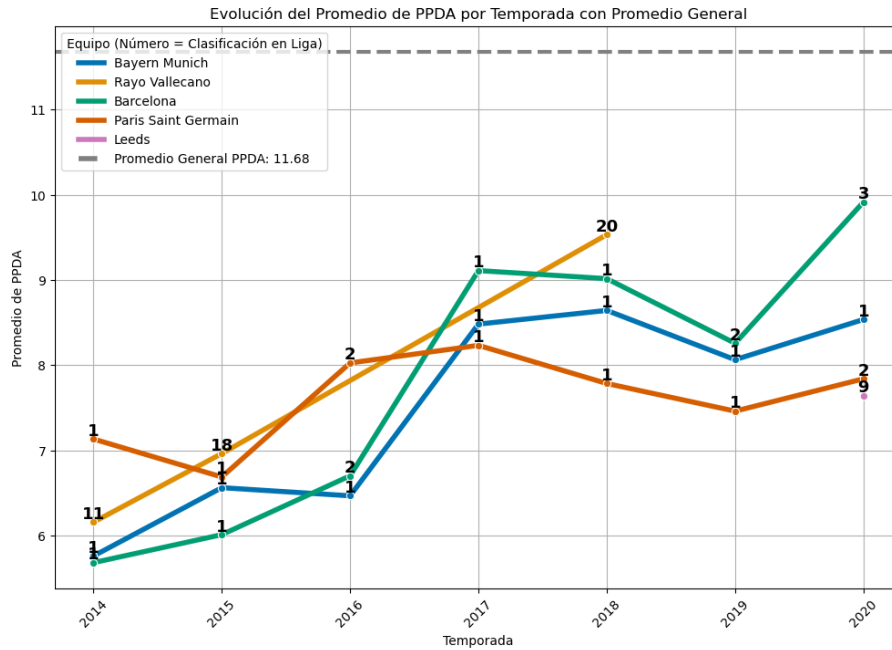


Figura 10: Evolución del Promedio de PPDA por Temporada para los Equipos con el PPDA más Bajo

También se analizaron los cinco equipos con el promedio de PPDA más alto. Todos ellos experimentaron descensos en sus respectivas temporadas. Esto sugiere de manera intuitiva que un promedio muy alto de PPDA tiende a estar asociado con malos resultados, mientras que un promedio bajo generalmente se correlaciona con un mejor desempeño, aunque con un grado considerable de variabilidad y dispersión.

3.4. Relación entre los puntos obtenidos con el xG Delta y el PPDA promedio de los equipos

La puntuación obtenida en una liga es una métrica directa y tangible del rendimiento de un equipo a lo largo de una temporada. Refleja no solo la consistencia y capacidad del equipo para ganar partidos, sino también su eficiencia en traducir oportunidades en resultados. Por lo tanto, la puntuación es una medida integral que permite evaluar el éxito competitivo de un equipo.

3.4.1. Relación entre Puntos Obtenidos y xG Delta Promedio

Para analizar la relación entre los puntos obtenidos y el xG Delta promedio de los equipos, se recopilamos datos de todas las ligas y temporadas del dataset. Primero, se calcularon los puntos totales obtenidos por cada equipo en una temporada, así como su xG Delta promedio. Luego, se combinó esta información para evaluar cómo el xG Delta influye en el rendimiento, medido en puntos obtenidos.

A continuación, se realizó una representación gráfica de esta relación, acompañada de una línea de regresión para visualizar la tendencia. Aunque se observa una considerable dispersión en los datos, la tendencia ascendente sugiere que existe una correlación entre un mayor xG Delta y la obtención de más puntos.

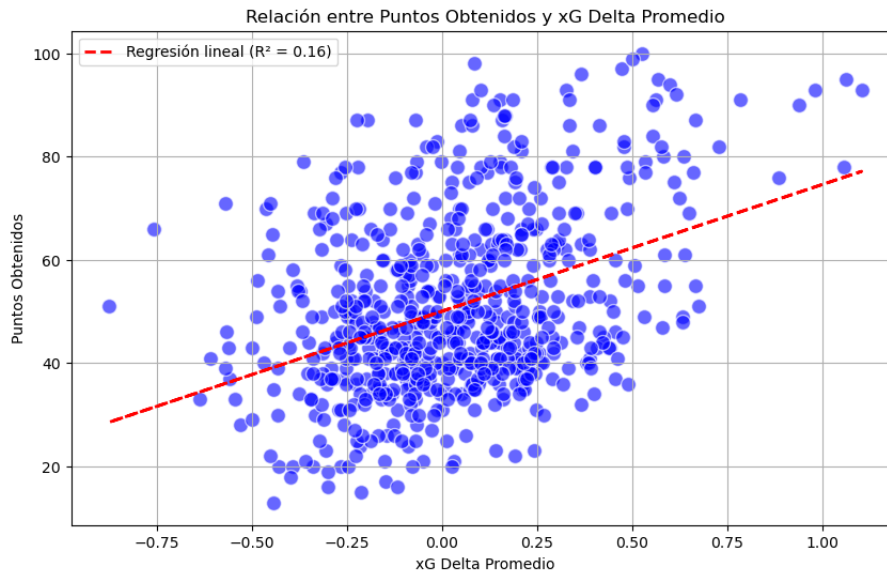


Figura 11: Relación entre Puntos Obtenidos y xG Delta Promedio

El coeficiente de determinación ($R^2 = 0,16$) indica que, aunque la relación es significativa, hay una alta variabilidad en los datos, lo que sugiere que el xG Delta es un factor relevante, pero no el único determinante del rendimiento de un equipo.

3.4.2. Relación entre Puntos Obtenidos y PPDA Promedio

El modus operandi para analizar la relación entre los puntos obtenidos y el PPDA promedio sigue el mismo enfoque que en el caso del xG Delta. Se recopilieron los datos de todas las ligas y temporadas del dataset, calculando tanto los puntos totales obtenidos por cada equipo en una temporada como su PPDA promedio. Esta información se combinó para evaluar cómo el PPDA, que mide la intensidad defensiva de un equipo, influye en su rendimiento en términos de puntos obtenidos.

La representación gráfica muestra una clara tendencia descendente: a medida que aumenta el PPDA promedio, existe una tendencia a que los equipos obtengan menos puntos. Esto se alinea con nuestros pronósticos, indicando que un mayor PPDA, que refleja una menor presión defensiva, suele estar asociado con un menor rendimiento en la liga.

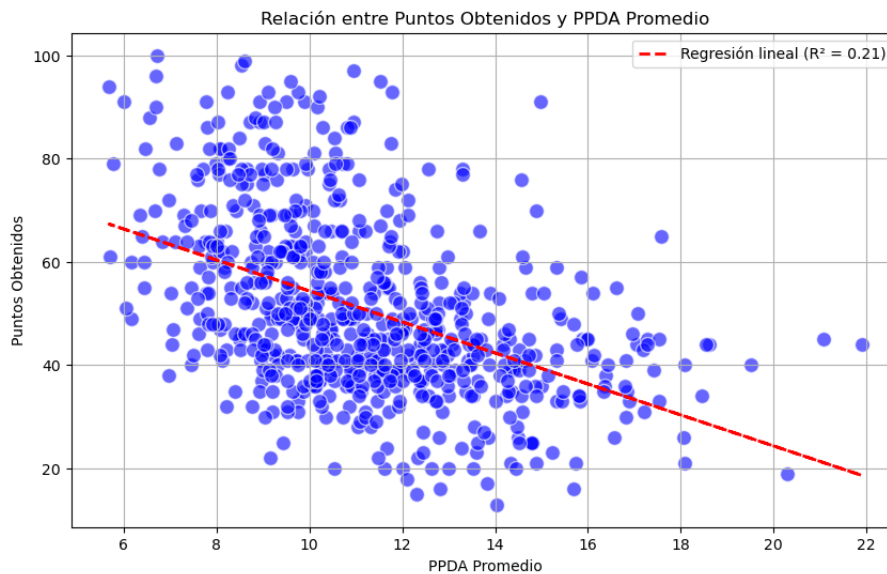


Figura 12: Relación entre Puntos Obtenidos y PPDA Promedio

El coeficiente de determinación ($R^2 = 0,21$) sugiere que la relación entre el PPDA y los puntos obtenidos es más significativa que la observada con el xG Delta. Sin embargo, al igual que antes, la dispersión y la varianza en los datos son considerables, lo que indica que, aunque el PPDA es un factor relevante, no es el único que determina el rendimiento de un equipo.

3.4.3. Relación entre xG Delta, PPDA y Puntuación

En esta sección, se explora la relación entre las tres variables clave: xG Delta, PPDA y puntuación. Aprovechando los cálculos realizados en apartados anteriores, se llevó a cabo una regresión múltiple que examina cómo estas métricas, en conjunto, influyen en el rendimiento de los equipos en términos de puntos obtenidos.

Para la representación, se optó por una visualización en dos dimensiones donde se graficaron los puntos obtenidos en el eje X y el xG Delta en el eje Y. La tercera dimensión, el PPDA, se representó mediante burbujas de tamaño variable según su valor. Además, se utilizó una escala de color con la paleta *bwr*: las burbujas más grandes y de color rojizo corresponden a valores altos de PPDA, mientras que las burbujas más pequeñas y azuladas indican valores bajos.

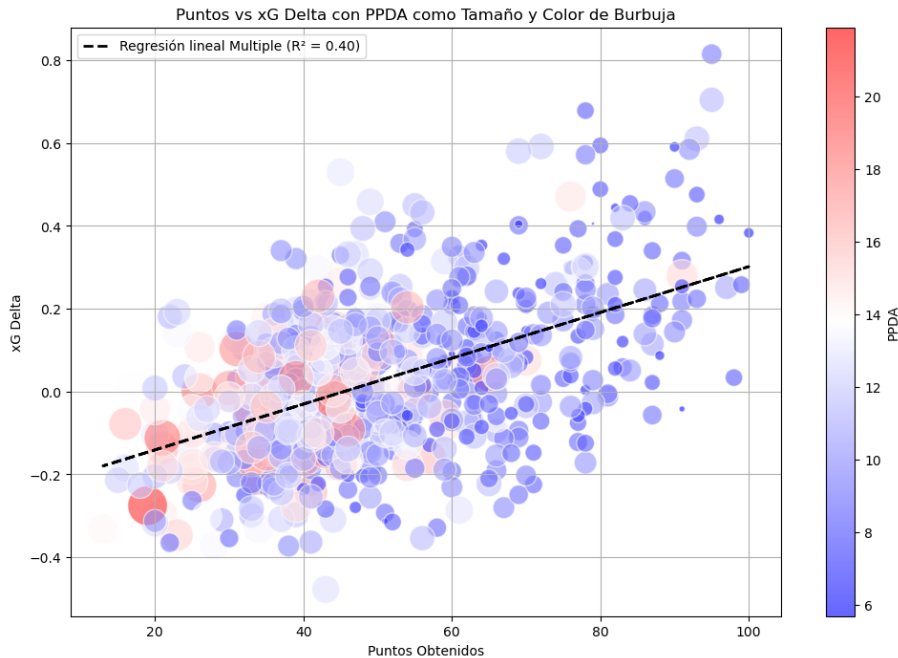


Figura 13: Relación entre Puntos Obtenidos, xG Delta y PPDA Promedio

El ajuste realizado muestra una mejora significativa en el coeficiente de determinación ($R^2 = 0,4$), lo que sugiere que considerar conjuntamente las métricas xG Delta y PPDA proporciona una comprensión más profunda de cómo estas influyen en el rendimiento de un equipo. Este análisis multidimensional permite capturar matices que podrían pasar desapercibidos al evaluar cada métrica por separado, destacando la importancia de un enfoque integral en el estudio del rendimiento deportivo.

4. Visualización

En esta sección, se utilizará un método de agrupación y promediación para reducir la varianza y dispersión en los datos, con el fin de obtener una visualización más clara y accesible de la relación entre las métricas xG Delta, PPDA, y la posición en la clasificación de los equipos.

4.1. Método de Agrupación y Promediación

El enfoque descrito consiste en agrupar a los equipos de diferentes ligas y temporadas según su posición final en la clasificación, es decir, se crean estratos donde se agrupan todos los equipos

que terminaron en la misma posición (primeros, segundos, etc.). Una vez agrupados, se calcula el promedio de las métricas de xG Delta y PPDA para cada uno de estos grupos. Este método permite disminuir la varianza y la dispersión de los datos, facilitando una interpretación más directa y comprensible.

4.2. Objetivo de la Visualización

El objetivo de este enfoque es generar tres visualizaciones que se basen en las posiciones en la liga como métrica principal, en lugar de analizar cada equipo individualmente. Este método permitirá una representación visual más clara y menos ruidosa de cómo las métricas xG Delta y PPDA influyen en el rendimiento de los equipos. Estas visualizaciones serán comparadas con las anteriores, mostrando de manera más sencilla y didáctica cómo las posiciones en la liga se relacionan con el rendimiento según las métricas estudiadas.

Al agrupar por posiciones y promediar las métricas, la visualización resultante ofrece una herramienta poderosa para explicar, incluso a personas no especializadas, las conclusiones extraídas de los análisis anteriores. Específicamente, se mostrará cómo la combinación de xG Delta y PPDA se correlaciona con el éxito o fracaso de los equipos, permitiendo así una comprensión más profunda y accesible de su impacto en el rendimiento en la liga.

4.3. Análisis Comparativo: Posición vs. xG Delta Promedio y Posición vs. PPDA Promedio

En este apartado, se analizan conjuntamente las visualizaciones resultantes de las gráficas de *Posición vs. xG Delta Promedio* y *Posición vs. PPDA Promedio*. Estas visualizaciones han sido diseñadas con un enfoque estilístico y cromático específico para facilitar la interpretación de los datos y destacar las tendencias clave que emergen al aplicar la técnica de promediación por posiciones en la liga.

4.3.1. Justificación del Estilo y Colores Elegidos

El estilo de las gráficas sigue un enfoque minimalista, con el fin de centrar la atención del observador en las tendencias clave que se desean destacar. Los colores han sido seleccionados estratégicamente para facilitar la diferenciación entre las dos métricas analizadas. Este esquema de color también ayuda a mantener una coherencia visual entre las dos gráficas, facilitando así la comparación directa entre las tendencias observadas.

4.3.2. Análisis de las Gráficas

Las gráficas muestran ahora con mayor claridad la tendencia general de cómo las métricas xG Delta y PPDA se relacionan con las posiciones en la clasificación, gracias a la técnica de promediación por posiciones discutida anteriormente.

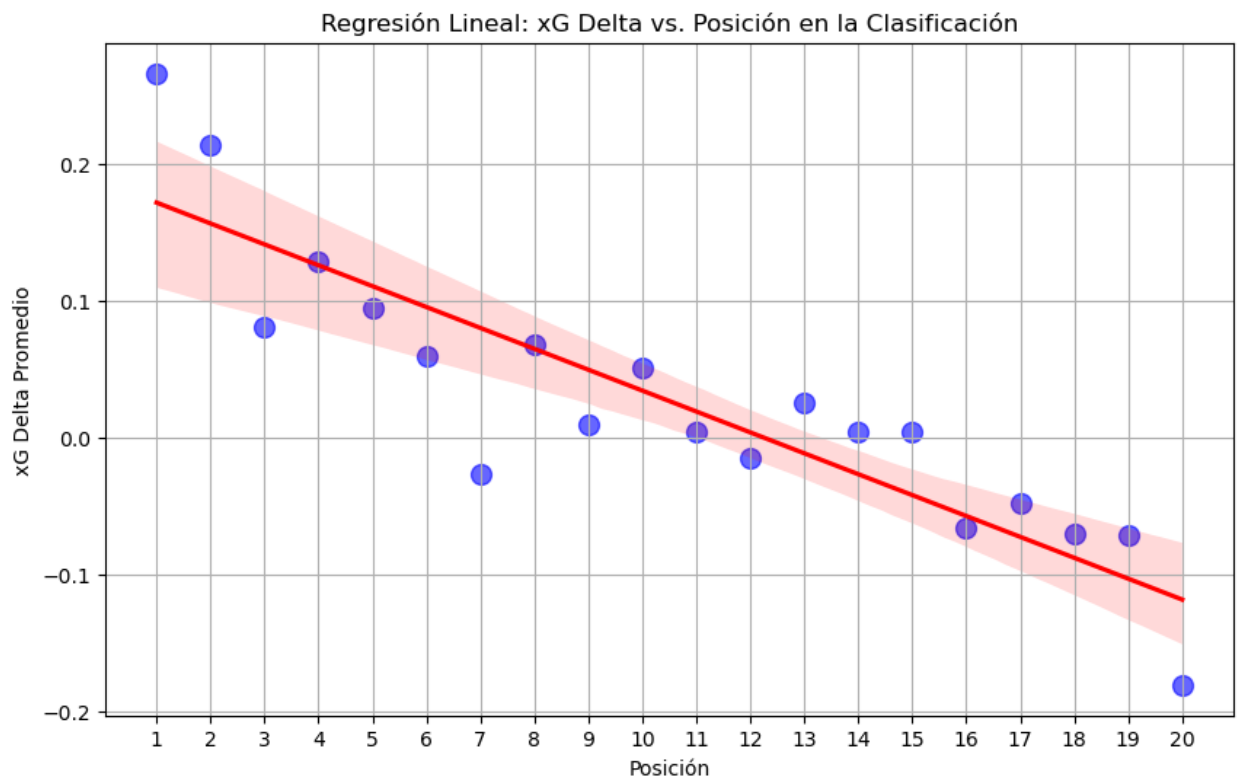


Figura 14: Relación entre Posición en la Clasificación con xG Delta Promedio

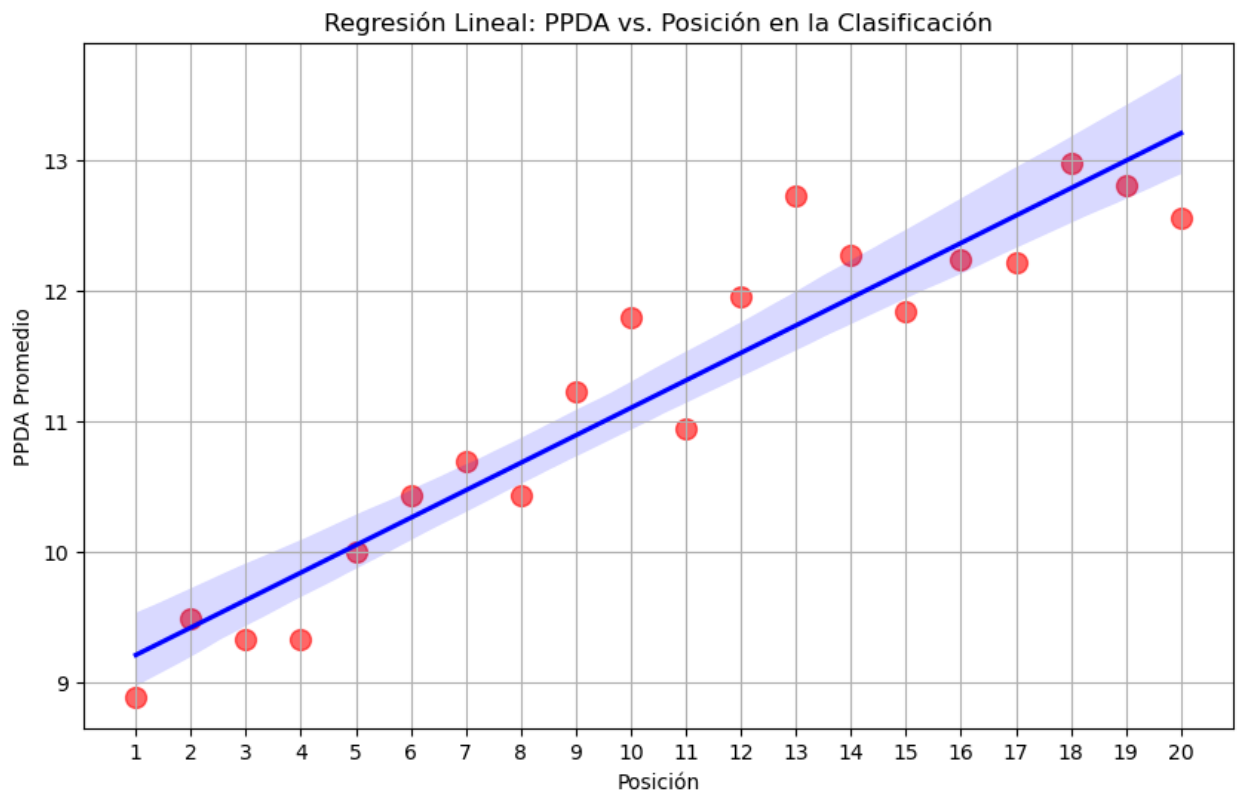


Figura 15: Relación entre Posición en la Clasificación con PPDA Promedio

- **Regresión Lineal: xG Delta vs. Posición en la Clasificación.** En esta gráfica, se

observa una tendencia descendente con un coeficiente de determinación (R^2) de 0.79, lo que indica una fuerte relación entre un mayor xG Delta y mejores posiciones en la tabla. Este resultado sugiere que equipos con un alto xG Delta suelen terminar en posiciones superiores, aunque es importante recordar que se trata de una media, y la variabilidad individual puede ser significativa.

- **Regresión Lineal: PPDA vs. Posición en la Clasificación.** Por otro lado, la gráfica de PPDA muestra una tendencia ascendente con un R^2 de 0.90, lo que refleja una relación igualmente fuerte entre un bajo PPDA y mejores posiciones en la liga. Esto confirma que equipos que ejercen una presión defensiva más alta (reflejada en un PPDA bajo) tienden a tener un mejor rendimiento en la clasificación.

Ambas gráficas, al analizarse conjuntamente, permiten a cualquier persona, independientemente de su nivel de especialización, comprender la importancia de estas métricas en el rendimiento de los equipos. No obstante, es crucial tener en cuenta que estos resultados representan promedios, y la variabilidad observada en estudios anteriores sigue siendo un factor importante a considerar.

4.3.3. Representación Conjunta de Posición vs xG Delta vs PPDA

En este subapartado se presenta una visualización que integra las tres variables clave estudiadas: la posición en la clasificación, el xG Delta y el PPDA. La representación se estructura como un gráfico de burbujas, donde:

- El eje X muestra las posiciones en la clasificación de los equipos.
- El eje Y representa el xG Delta promedio obtenido por los equipos.
- El tamaño y el color de las burbujas varían en función del PPDA promedio, donde burbujas más grandes y de color rojizo indican valores más altos de PPDA, y burbujas más pequeñas y de color azulado indican valores más bajos.

Este tipo de visualización tiene varias ventajas importantes:

1. **Claridad Visual:** Al combinar tres variables en un único gráfico, se proporciona una representación visual clara y comprensible de cómo las métricas xG Delta y PPDA se relacionan con el rendimiento del equipo en términos de su posición en la clasificación.
2. **Análisis Multidimensional:** El uso de tamaño y color de las burbujas permite captar la influencia simultánea de múltiples factores, lo que facilita la identificación de patrones y relaciones complejas entre las variables.
3. **Intuición y Comprensión:** Esta representación intuitiva permite que cualquier persona, incluso sin conocimientos avanzados en métricas futbolísticas, comprenda la importancia de mejorar tanto el xG Delta como el PPDA para alcanzar un mejor rendimiento en la clasificación.
4. **Ajuste Lineal:** Además, se incluye una línea de regresión que refuerza la relación entre las posiciones en la clasificación y el xG Delta, mostrando una tendencia lineal significativa.

El análisis de esta representación sugiere una fuerte correlación entre estas métricas, con un coeficiente de determinación R^2 de 0.93, indicando que la combinación de un alto xG Delta y un bajo PPDA está asociada con mejores posiciones en la clasificación. Esta visualización resalta la importancia de trabajar en ambas métricas para mejorar el rendimiento a lo largo de varias temporadas, aunque se debe tener en cuenta la variabilidad inherente en los datos.

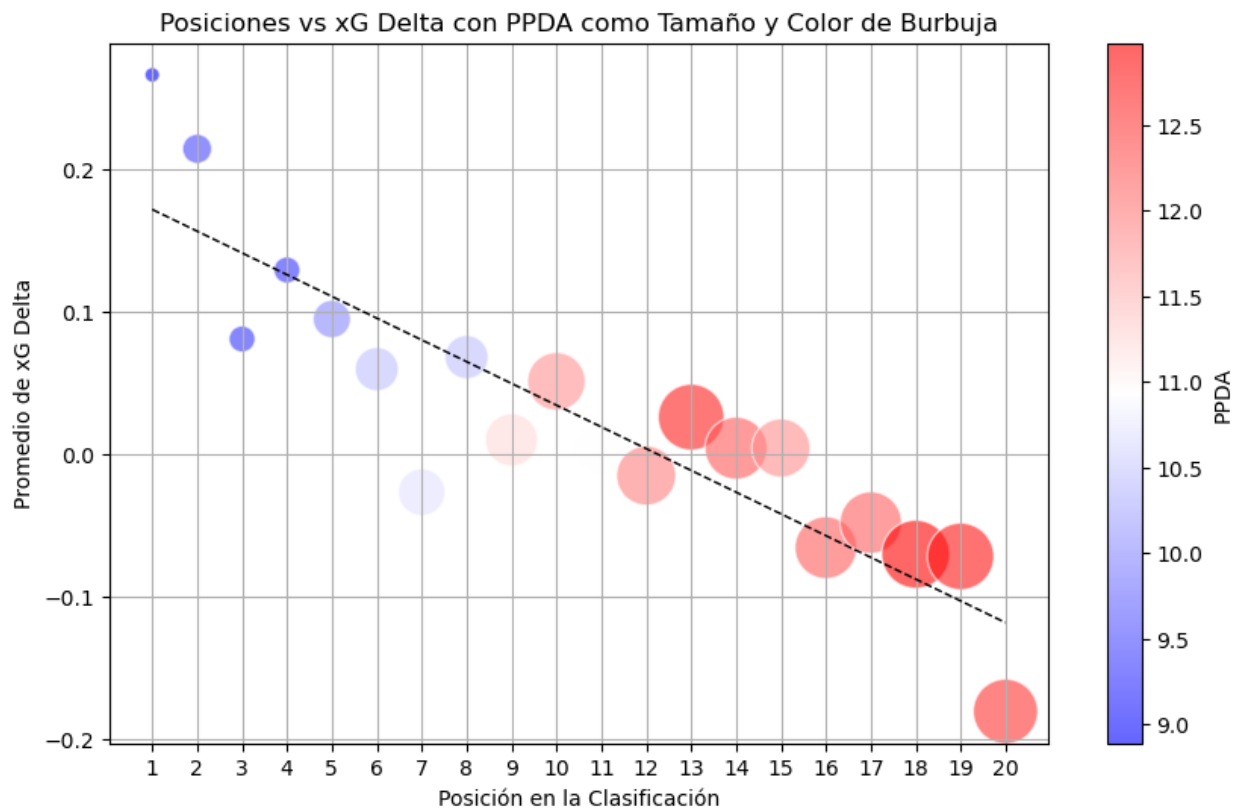


Figura 16: Posiciones vs xG Delta con PPDA como Tamaño y Color de Burbuja

5. Discusión, Conclusiones y Posibles Mejoras

5.1. Discusión

El presente estudio ha explorado la relevancia de las métricas *xG Delta* y *PPDA* en el rendimiento de los equipos de fútbol, analizando cómo estas influyen en su clasificación final en la liga. A través de una serie de visualizaciones, se ha demostrado que ambas métricas tienen una relación significativa con el éxito o el fracaso de los equipos. La combinación de estas dos métricas proporciona una visión más completa del rendimiento, permitiendo un análisis más profundo de las estrategias y enfoques que podrían adoptar los equipos.

En particular, la técnica de agrupación por posiciones ha sido clave para reducir la varianza y el ruido en los datos, lo que ha permitido obtener visualizaciones más claras y comprensibles. Al agrupar los datos según las posiciones finales en la liga y calcular los promedios de *xG Delta* y *PPDA*, hemos podido observar tendencias más nítidas que destacan la importancia de estas métricas. Sin embargo, es importante señalar que, aunque se observa una clara tendencia, la variabilidad y la dispersión dentro de los datos originales indican que otros factores también juegan un papel crucial en el rendimiento global de un equipo.

5.2. Conclusiones

A lo largo del análisis se ha demostrado que:

- El *xG Delta* tiene una correlación positiva significativa con el rendimiento del equipo, reflejado en la posición en la liga. Un *xG Delta* alto generalmente está asociado con mejores posiciones en la tabla.
- El *PPDA*, por otro lado, muestra una correlación negativa con el rendimiento, lo que significa que equipos con un *PPDA* bajo (indicativo de una presión defensiva alta) tienden a rendir mejor en la liga.

- La combinación de *xG Delta* y *PPDA* en un análisis multivariable mejora la capacidad predictiva sobre el rendimiento del equipo, tal como lo indica el aumento en el coeficiente de determinación R^2 en la regresión múltiple.

Estos resultados subrayan la importancia de considerar múltiples métricas avanzadas para evaluar el rendimiento de los equipos de fútbol. El *xG Delta* y el *PPDA* juntos ofrecen una perspectiva más completa que cualquiera de las métricas por sí sola.

5.3. Posibles Mejoras

Aunque el análisis presentado proporciona valiosas ideas, existen varias áreas donde el estudio podría mejorar y ampliarse:

- **Incorporación de más temporadas y ligas:** Expandir el análisis para incluir más temporadas y ligas adicionales podría ayudar a generalizar los resultados y verificar si las tendencias observadas son consistentes en diferentes contextos.
- **Análisis de factores adicionales:** Incluir otras métricas avanzadas como *xG Chain*, *xG Buildup* y métricas defensivas específicas podría enriquecer el análisis y proporcionar una visión aún más completa del rendimiento de los equipos.
- **Análisis de contexto:** Integrar análisis cualitativos que consideren cambios en la dirección técnica, lesiones clave, y otras circunstancias contextuales podría explicar algunas de las variabilidades observadas en los datos. .

En resumen, el estudio destaca la importancia de las métricas *xG Delta* y *PPDA* en la evaluación del rendimiento de los equipos de fútbol, y ofrece un punto de partida para futuros análisis y desarrollos en este ámbito.

6. Referencias

1. **The PFSA.** The Power of Analytics in Football. Retrieved from <https://thepfesa.co.uk>
2. **Frontiers in Sports.** Data-Driven Visual Performance Analysis in Football. Retrieved from <https://www.frontiersin.org/articles/10.3389/fspor.2021.682707/full>
3. **Understat.** Advanced Football Analytics. Retrieved from <https://understat.com/>
4. **Kaggle Dataset.** Football Database by Technika148. Retrieved from <https://www.kaggle.com/datasets/technika148/football-database/data?select=teamstats.csv>
5. ¿Para qué sirve el indicador xG (Goles esperados) y cómo se calcula? <https://objetivoanalista.com/xg-goles-esperados/>
6. ¿Qué son los PPDA (Passes allowed Per Defensive Action)? [https://vicdatadeporte.wordpress.com/2021/01/29/que-son-los-ppda-passes-allowed-per-defensive-action/#:~:text=Los%20PPDA%20\(siglas%20en%20ingl%C3%A9s,de%20uno%20o%20varios%20partidos.](https://vicdatadeporte.wordpress.com/2021/01/29/que-son-los-ppda-passes-allowed-per-defensive-action/#:~:text=Los%20PPDA%20(siglas%20en%20ingl%C3%A9s,de%20uno%20o%20varios%20partidos.)
7. **Apuntes de la asignatura**