

# Análisis histórico y estadístico de la English Premier League



Autor: Gastón Padro

Institución: CODERHOUSE

Curso: Data Analytics

Fecha de presentación: 08/06/2025

## Contenido

Breve descripción.....	3
Hipótesis .....	3
Tabla de versionado.....	3
Objetivo General.....	4
Herramientas tecnológicas implementadas .....	4
Temáticas y alcance .....	4
Tipos de procesos .....	4
Exploración primaria.....	5
Usuario final y nivel de aplicación del análisis: .....	5
Data set seleccionado .....	5
Diagrama de entidad relación:.....	6
Listado de tablas y descripción: .....	6
Descripción y definición de llaves primarias (PK) y foráneas (FK): .....	7
Listado de columnas de cada tabla e información: .....	7
Transformación de datos .....	10
Medidas calculadas.....	10
1. Máximo puntos temporada: .....	10
2. Promedio puntos por Equipo en Temporada: .....	10
3. Equipo y Diferencia de Goles: .....	10
4. Diferencia de Goles: .....	11
5. Equipo Campeón Visual:.....	11
6. Equipo con más puntos en una temporada: .....	11
7. Equipo con más puntos históricos.....	11
8. Equipo con más temporadas jugadas.....	12
9. Mejor diferencia de goles en una temporada.....	12
10. Promedio Puntos por Temporada.....	12
11. Puntos del Campeón por Temporada .....	13
12. Puntos del Campeón por Temporada .....	13
13. Puntos del Campeón por Temporada .....	13
Diagrama de Entidad Relación de Power Bi.....	13
Imágenes de páginas del tablero .....	14
Portada: .....	14

Glosario:.....	14
Análisis por temporada:.....	15
Estadísticas por equipo:.....	15
Records históricos:.....	16
Futuras líneas:.....	16

## Breve descripción

El proyecto tiene como objetivo realizar un análisis exhaustivo sobre datos históricos de la Premier League, la máxima categoría del fútbol inglés, desde la temporada 1888-1889 hasta la temporada 2023-2024. El análisis se enfocará en obtener insights que permitan evaluar el desempeño histórico de equipos y temporadas específicas, facilitando la visualización de patrones, tendencias y estadísticas relevantes para la toma de decisiones, predicción de resultados y evaluación comparativa del rendimiento de equipos.

## Hipótesis

1. Existe una correlación entre el desempeño histórico (cantidad de partidos ganados, goles anotados y recibidos, rendimiento como local y visitante) y la probabilidad de un equipo de lograr una mejor posición final en la tabla de la temporada actual. En resumen: Los equipos con un rendimiento histórico superior presentan mayores probabilidades de mantenerse competitivos a lo largo del tiempo.
2. El análisis estadístico histórico permite predecir patrones relevantes en relación con la influencia de la localía, la eficacia ofensiva y defensiva, y cómo estas variables afectan los resultados finales de una temporada. En resumen: Factores históricos estadísticos pueden predecir la influencia del desempeño de local y visitante en los resultados obtenidos.

## Tabla de versionado

Versión	Fecha	Delta
<b>V1</b>	28/03/2025	Primera versión donde se establecen las hipótesis del análisis
<b>V2</b>	11/04/2025	Se realizó la conexión de datos y se generó el diseño, las medidas y los gráficos principales del trabajo.
<b>V3</b>	08/06/2025	Se generaron las páginas faltantes y se finalizó el tablero.

## Objetivo General

Utilizando la metodología SMART (Específico, Medible, Alcanzable, Relevante y con Marco temporal definido), el objetivo general es desarrollar un análisis descriptivo y predictivo basado en los datos históricos de la Premier League, permitiendo identificar patrones claros y predecibles que ayuden a la toma de decisiones estratégicas relacionadas con el rendimiento de equipos, influencia de estadios y tendencias históricas relevantes.

## Herramientas tecnológicas implementadas

Para realizar este trabajo se utilizaron las siguientes herramientas:

- Microsoft Power Bi
- Microsoft Excel
- Microsoft Word
- Adobe PDF

## Temáticas y alcance

El análisis abarcará datos históricos completos de partidos y equipos de la Premier League desde la temporada inicial 1888-1889 hasta la última temporada completa disponible (2023-2024).

Se analizarán variables como:

- Partidos ganados, empatados y perdidos.
- Goles marcados y recibidos.
- Rendimiento como equipo local y visitante.
- Estadísticas avanzadas (penales, tarjetas, faltas).
- Evolución del rendimiento histórico y patrones recurrentes.

Esto permitirá determinar tendencias claras que puedan utilizarse para predecir resultados futuros o realizar análisis comparativos entre temporadas o equipos.

## Tipos de procesos

El análisis implicará:

- Proceso Descriptivo: Identificar patrones históricos, tendencias y estadísticas relevantes.

- Proceso Diagnóstico: Evaluar por qué ciertos equipos o temporadas han tenido resultados específicos en función de variables históricas disponibles.
- Proceso Predictivo: Utilizar estadísticas históricas para predecir comportamientos futuros sobre la base de variables identificadas como significativas.
- Proceso Prescriptivo: Recomendar estrategias específicas basadas en el análisis realizado, especialmente útiles para equipos y analistas deportivos.

## Exploración primaria

Tras una primera exploración, se observa que:

- Equipos históricamente exitosos tienden a mantener un rendimiento consistente en temporadas recientes.
- Existe una clara influencia del desempeño como local en los resultados generales de cada equipo.
- Las estadísticas históricas permiten identificar patrones claros en el rendimiento ofensivo y defensivo, que pueden usarse como indicadores predictivos.
- No se detectan errores significativos, aunque es recomendable una segunda revisión para asegurar calidad en todos los registros.

## Usuario final y nivel de aplicación del análisis:

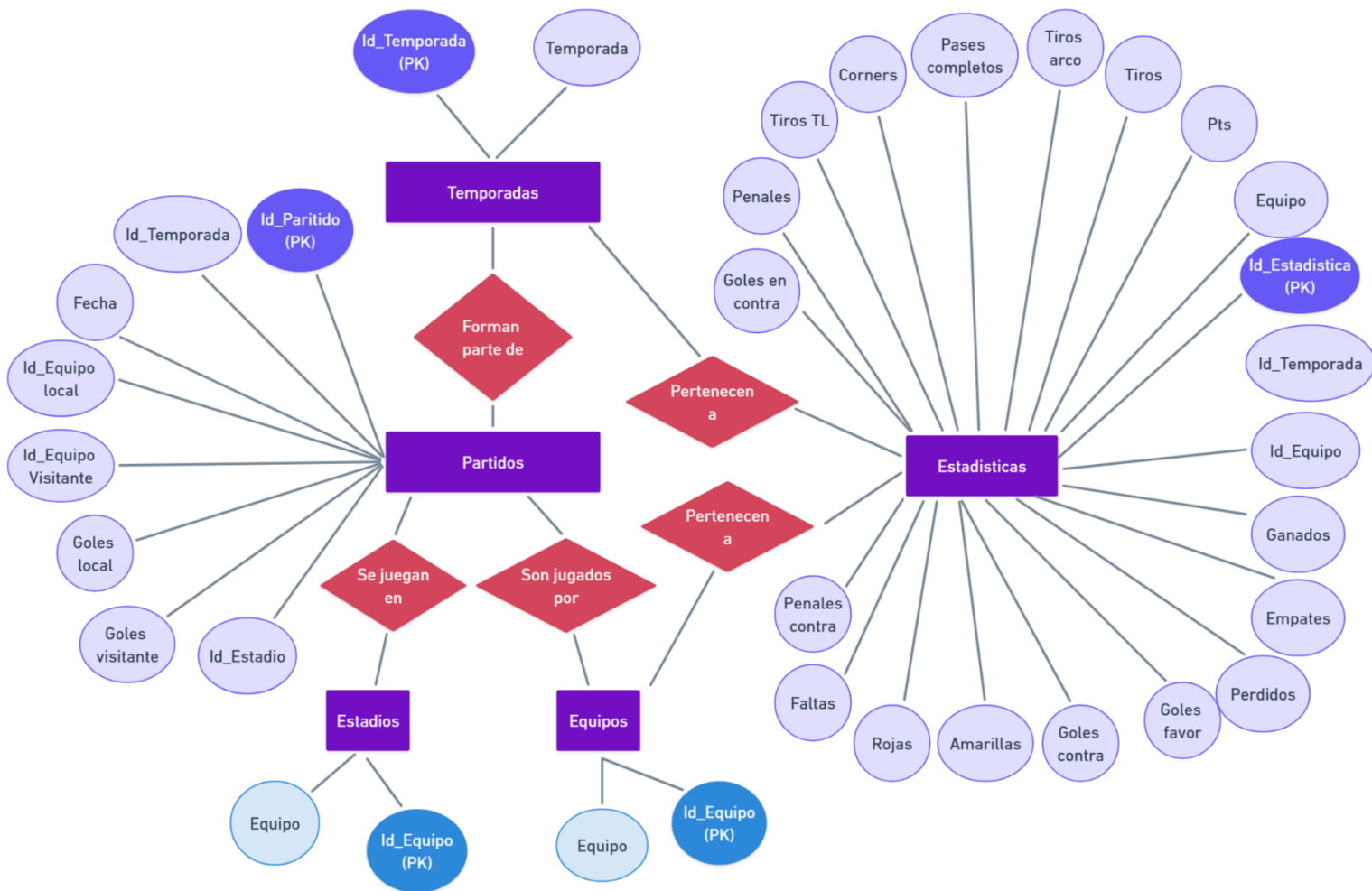
El usuario final del proyecto será principalmente:

- **Nivel Estratégico:** Directivos o cuerpo técnico de clubes deportivos que deseen utilizar datos históricos para decidir estratégicamente sobre contrataciones, tácticas o identificar rivales fuertes o débiles históricamente.
- **Nivel Táctico:** Analistas deportivos o periodistas que busquen fundamentos históricos y estadísticos claros para comentar sobre rendimientos y resultados en sus medios.
- **Nivel Operativo:** Apostadores deportivos o aficionados que buscan mejorar su precisión en pronósticos deportivos basándose en patrones históricos detectados.

## Data set seleccionado

La base de datos seleccionada contiene información histórica de la Premier League desde la temporada 1888-1889 hasta 2023-2024.

## Diagrama de entidad relación:



Made with  Whimsical

## Listado de tablas y descripción:

- Temporadas: Información general sobre cada temporada (campeón, temporadas específicas).
- Partidos: Datos de cada encuentro histórico jugado (fecha, equipos, goles, estadio).
- Equipos: Información básica y asignación de IDs para cada equipo participante.
- Estadísticas Temporada: Estadísticas detalladas de cada equipo durante cada temporada.
- Estadios: Identificación y asignación de IDs únicos para cada estadio utilizado.

## Descripción y definición de llaves primarias (PK) y foráneas (FK):

### Tabla: Temporadas

- **PK:** id\_Temporada (Identificador único para cada temporada).

### Tabla: Equipos

- **PK:** id\_Equipo (Identificador único para cada equipo).

### Tabla: Estadios

- **PK:** Id\_Estadio (Identificador único para cada estadio).

### Tabla: Partidos

- **PK:** id\_partido (Identificador único para cada partido).
- **FK:** id\_Temporada (Se relaciona con tabla Temporadas).
- **FK:** id\_Equipo\_local y id\_Equipo\_visitante (Se relacionan con tabla Equipos).
- **FK:** Id\_Estadio (Se relaciona con tabla Estadios).

### Tabla: Estadísticas\_Temporada

- **PK:** Id\_Estadistica (combinación de id\_Temporada e id\_Equipo).
- **FK:** id\_Temporada (Se relaciona con tabla Temporadas).
- **FK:** id\_Equipo (Se relaciona con tabla Equipos).

## Listado de columnas de cada tabla e información:

### Tabla: Temporadas

Tipo de clave	Campo	Tipo de campo	Detalle
PK	id_Temporada	INT	Identificador único para cada temporada
	Temporada	Texto	Nombre o período temporal (Ej: 2023/2024)
	Campeón	Texto	Equipo campeón de la temporada

**Tabla: Equipos**

Tipo de clave	Campo	Tipo de campo	Detalle
PK	id_Equipo	INT	Identificador único para cada equipo
	Equipo	Texto	Nombre del equipo

**Tabla: Estadios**

Tipo de clave	Campo	Tipo de campo	Detalle
PK	id_Estadio	INT	Identificador único del estadio
	Estadio	Texto	Nombre del estadio

**Tabla: Partidos**

Tipo de clave	Campo	Tipo de campo	Detalle
PK	Id_Partido	Texto	Identificador único del partido (id Temporada + equipo local + visitante)
FK	Id_Temporada	INT	Temporada en la que se jugó el partido (Tabla Temporadas)
	Fecha	Fecha	Fecha en la que se jugó el partido
FK	Id_Equipo local	INT	Identificador del equipo local (Tabla Equipos)
	Goles local	INT	Goles marcados por el equipo local
	Goles visitante	INT	Goles marcados por el equipo visitante
FK	Id_Equipo Visitante	INT	Identificador del equipo visitante (Tabla Equipos)
FK	Id_Estadio	INT	Identificador del estadio (Tabla Estadios)



**Tabla: Estadísticas**

Tipo de clave	Campo	Tipo de campo	Detalle
PK	Id_Estadistica	Texto	Identificador único (id Temporada + id Equipo)
	Equipo	Texto	Temporada relacionada a las estadísticas (Tabla Temporadas)
FK	Id_Temporada	INT	Equipo relacionado a las estadísticas (Tabla Equipos)
	Ganados	INT	Partidos ganados en la temporada
	Empate	INT	Partidos empatados en la temporada
	Perdidos	INT	Partidos perdidos en la temporada
	Goles favor	INT	Goles marcados en total por el equipo
	Goles contra	INT	Goles recibidos en total por el equipo
	Pts	INT	Puntos totales obtenidos en la temporada
	Tiros	INT	Tiros totales realizados en la temporada
	Tiros Arco	INT	Tiros al arco realizados en la temporada
	Tiros TL	INT	Tiros de tiros libres realizados
	Penales	INT	Penales a favor
	Pases Completos	INT	Cantidad de pases completos
	Corners	INT	Cantidad de corners a favor
	Amarillas	INT	Tarjetas amarillas recibidas
	Rojas	INT	Tarjetas rojas recibidas
	Faltas	INT	Faltas cometidas

	Penales contra	INT	Penales en contra
	Goles en contra	INT	Autogoles realizados

## Transformación de datos

Se verificó el formato de datos de todas las tablas para asegurarme que sean los correctos. Además, se creó una tabla “Calendario” que cubre los años de todos los partidos jugados en las distintas temporadas y se agregó una columna con datos numéricos para el año, el mes y el día con formato texto.

## Medidas calculadas

Se creó una tabla llamada “Medidas”, donde se agruparán todas las medidas calculadas.

### 1. Máximo puntos temporada:

Máximo Puntos Temporada =

```
CALCULATE(
    MAX(Estadisticas[Pts]),
    REMOVEFILTERS(Estadisticas[Equipo]))
```

### 2. Promedio puntos por Equipo en Temporada:

Promedio Puntos por Equipo en Temporada =

```
AVERAGEX(
    VALUES(Estadisticas[Equipo]),
    CALCULATE(SUM(Estadisticas[Pts])))
```

### 3. Equipo y Diferencia de Goles:

Equipo y Diferencia de Goles =

```
VAR TablaFiltrada =
    ADDCOLUMNS(
        Estadisticas,
        "Dif", Estadisticas[Goles Favor] - Estadisticas[Goles contra]
    )
VAR MaxDif = MAXX(TablaFiltrada, [Dif])
VAR EquipoGanador =
    CALCULATE(
        FIRSTNONBLANK(Estadisticas[Equipo], 1),
        (Estadisticas[Goles Favor] - Estadisticas[Goles contra]) = MaxDif
    )
RETURN
    EquipoGanador & " – " & MaxDif & " goles"
```

#### 4. Diferencia de Goles:

Diferencia de Goles =

`SUM(Estadisticas[Goles Favor]) - SUM(Estadisticas[Goles contra])`

#### 5. Equipo Campeón Visual:

Equipo Campeón Visual =

```
IF(
    HASONEVALUE(Temporadas[Temporada]),
    SELECTEDVALUE(Temporadas[Campeon]),
    "Seleccioná una temporada"
)
```

#### 6. Equipo con más puntos en una temporada:

Equipo con más puntos en una temporada =

```
VAR TablaMax =
TOPN(
    1,
    ADDCOLUMNS(
        Estadisticas,
        "TemporadaNombre", RELATED(Temporadas[Temporada])
    ),
    Estadisticas[Pts],
    DESC
)
RETURN
MAXX(
    TablaMax,
    Estadisticas[Equipo] & " – " &
    Estadisticas[Pts] & " pts (" &
    [TemporadaNombre] & ")"
)
```

#### 7. Equipo con más puntos históricos

Equipo con más puntos históricos =

```
VAR TablaPuntos =
SUMMARIZE(
    Estadisticas,
    Estadisticas[Equipo],
    "TotalPts", SUM(Estadisticas[Pts])
)
VAR TopEquipo =
TOPN(1, TablaPuntos, [TotalPts], DESC)
RETURN
```

```
MAXX(
    TopEquipo,
    Estadisticas[Equipo] & " – " & FORMAT([TotalPts], "#,0") & " pts"
)
```

## 8. Equipo con más temporadas jugadas

Equipo con más temporadas jugadas =

```
VAR TablaTemporadas =
    SUMMARIZE(
        Estadisticas,
        Estadisticas[Equipo],
        "TotalTemporadas", DISTINCTCOUNT(Estadisticas[Id_Temporada])
    )
VAR TopEquipo =
    TOPN(1, TablaTemporadas, [TotalTemporadas], DESC)
RETURN
    MAXX(
        TopEquipo,
        Estadisticas[Equipo] & " – " & [TotalTemporadas] & " temporadas"
    )
```

## 9. Mejor diferencia de goles en una temporada

Mejor diferencia de goles en una temporada =

```
VAR TablaDif =
    ADDCOLUMNS(
        Estadisticas,
        "DifGoles", Estadisticas[Goles Favor] - Estadisticas[Goles contra],
        "Temporada", RELATED(Temporadas[Temporada])
    )
VAR TopDif =
    TOPN(1, TablaDif, [DifGoles], DESC)
RETURN
    MAXX(
        TopDif,
        Estadisticas[Equipo] & " – +" & [DifGoles] & " goles (" & [Temporada] & ")"
    )
```

## 10. Promedio Puntos por Temporada

Promedio Puntos por Temporada =

```
DIVIDE(
    [Puntos totales históricos],
    [Temporadas Jugadas]
)
```

### 11. Puntos del Campeón por Temporada

Puntos del Campeón por Temporada =

```
CALCULATE(
    MAX(Estadisticas[Pts]),
    VALUES(Estadisticas[Id_Temporada])
)
```

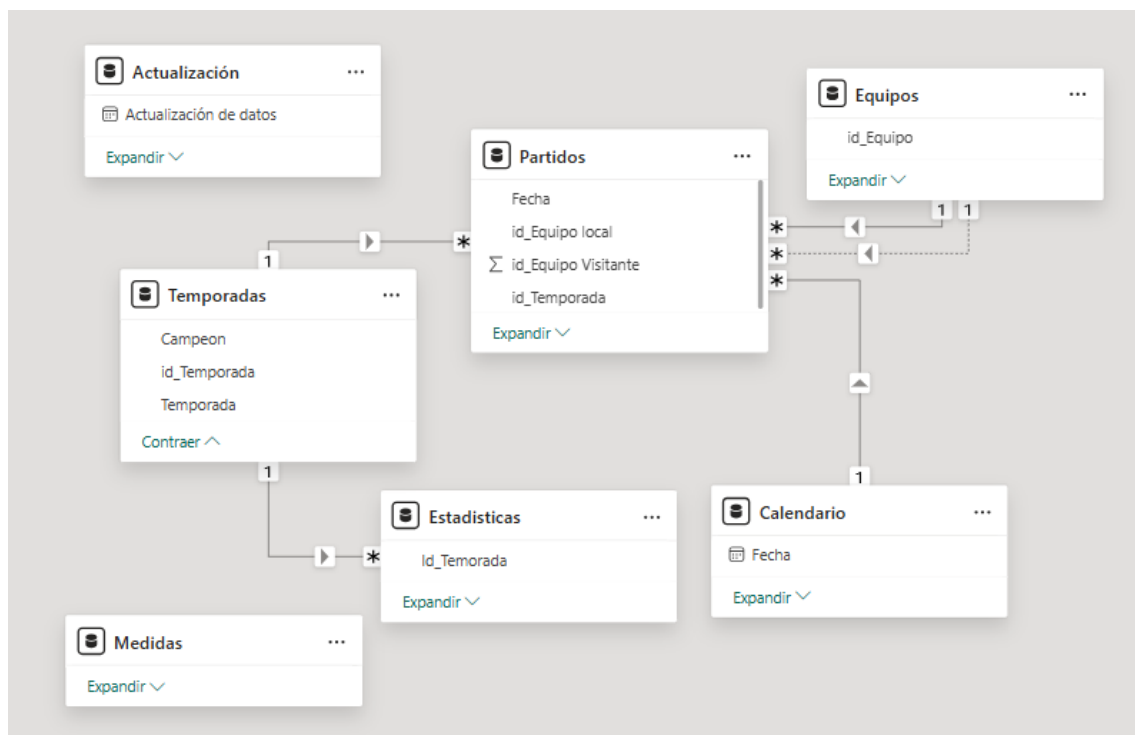
### 12. Puntos del Campeón por Temporada

Puntos totales históricos = SUM(Estadisticas[Pts])

### 13. Puntos del Campeón por Temporada

Temporadas jugadas = DISTINCTCOUNT(Estadisticas[Id\_Temporada])

## Diagrama de Entidad Relación de Power Bi



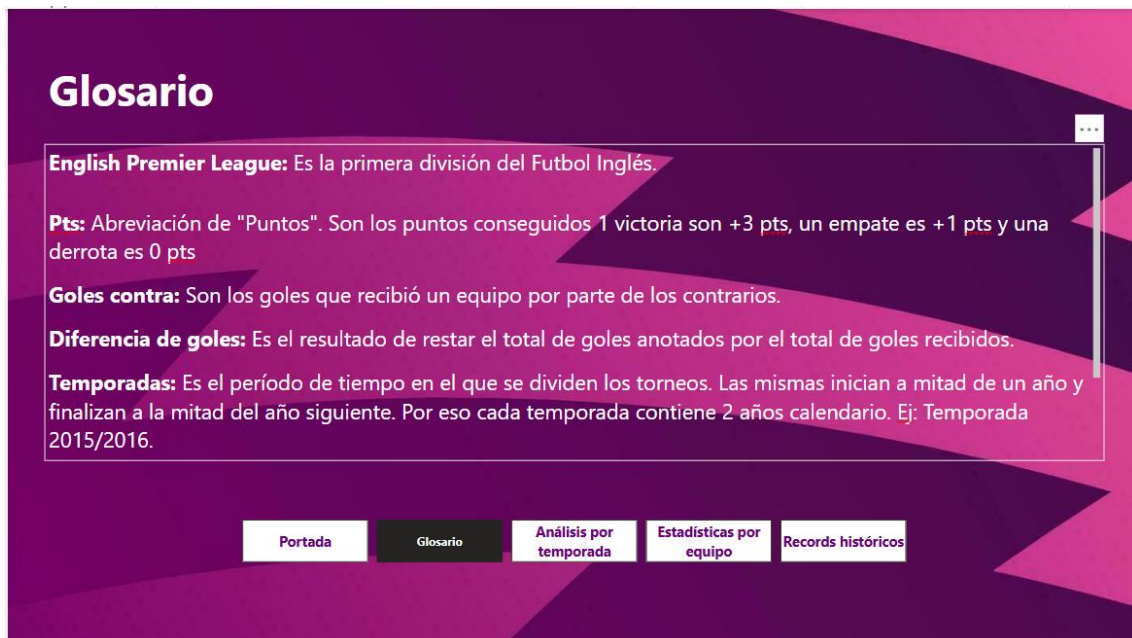
## Imágenes de páginas del tablero

Portada:



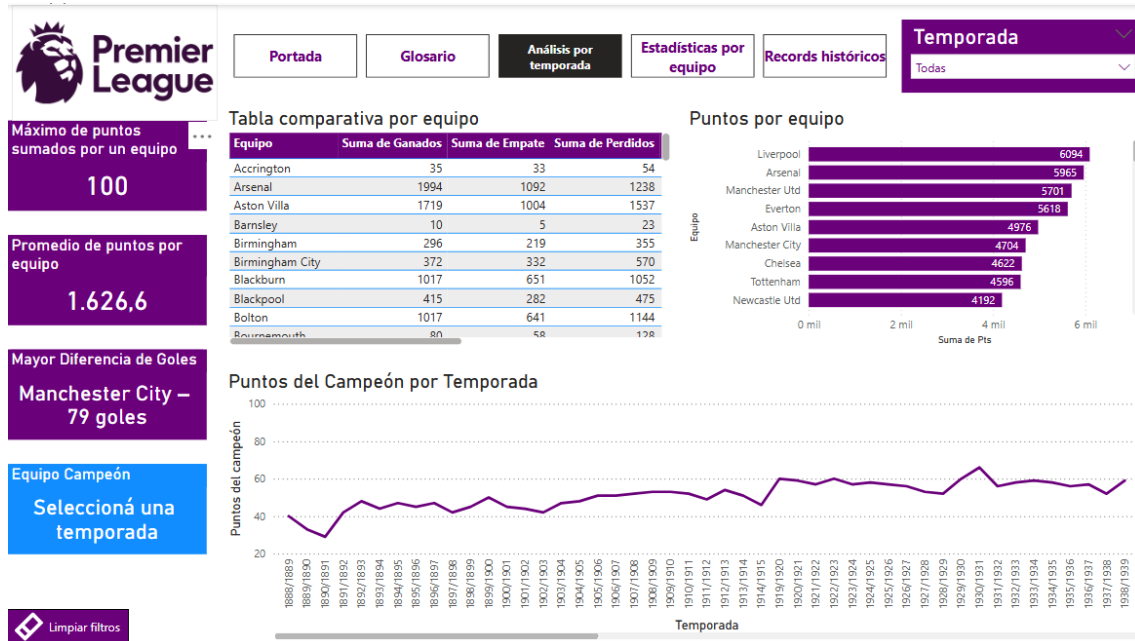
La portada cuenta de 3 accesos a distintas páginas y con un acceso a un Glosario con la explicación de algunos términos que sirven para comprender el análisis. A su vez la portada cuenta con la fecha de la última actualización del tablero.

Glosario:



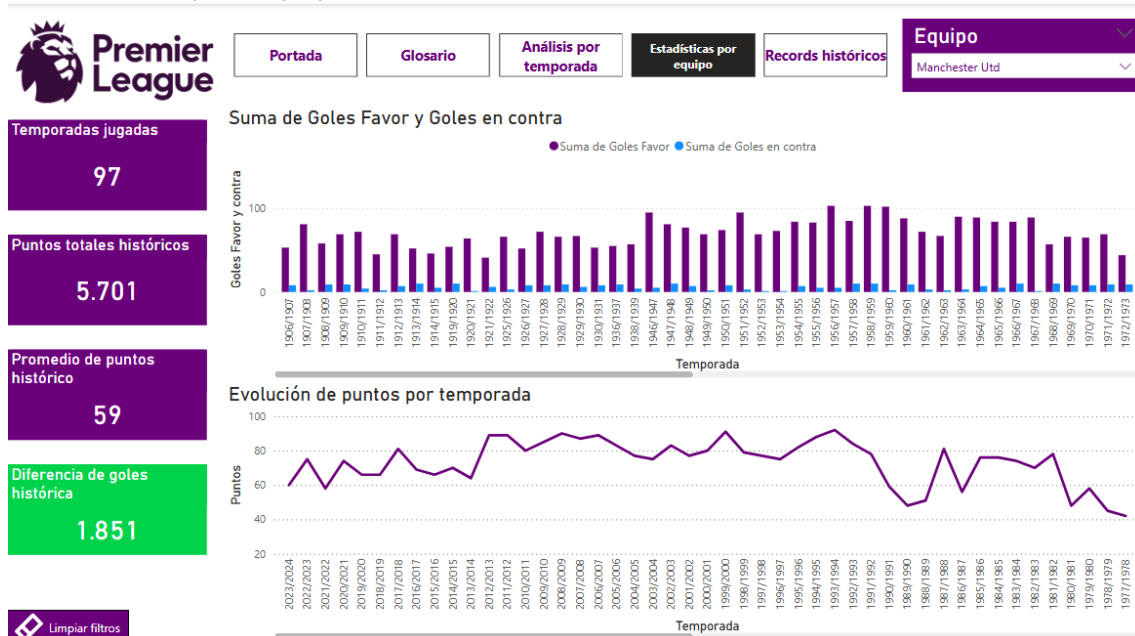
El Glosario nos da información sobre determinados términos técnicos utilizados en el proyecto.

## Análisis por temporada:



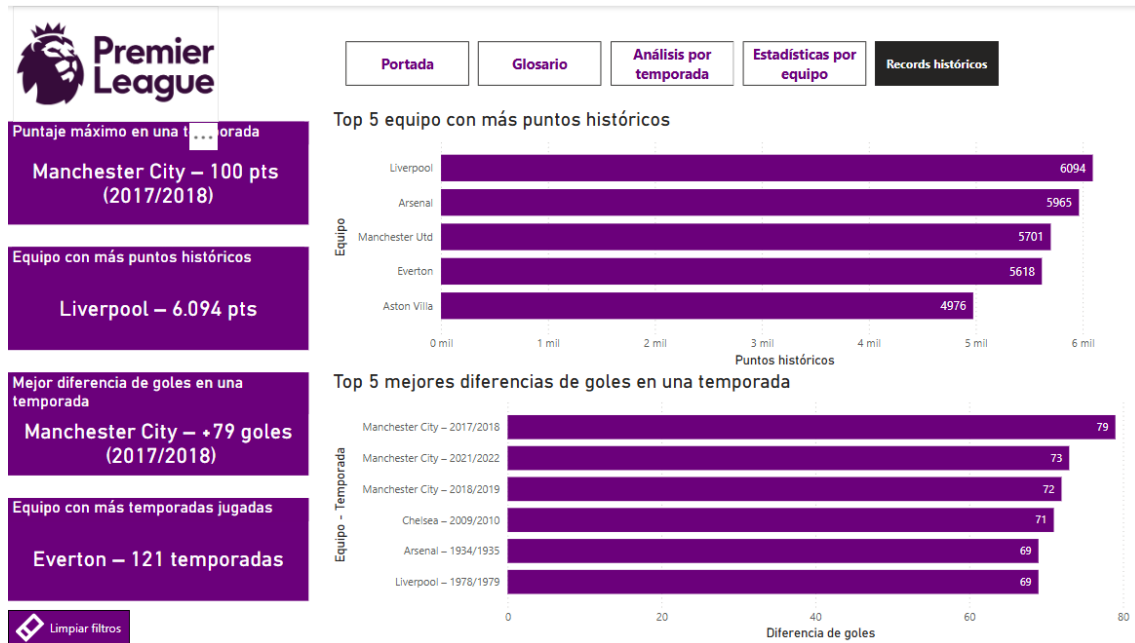
El primer análisis brinda información general sobre distintas estadísticas de la Liga Inglesa filtradas por temporada. En el mismo se pueden ver el máximo de puntos sumados, el promedio de puntos por equipo, la mayor diferencia de goles, entre otras métricas generales.

## Estadísticas por equipo:



El segundo análisis brinda información general sobre distintas estadísticas de la Liga Inglesa filtradas por equipo. En el mismo se puede ver información por cada equipo respecto a las temporadas jugadas, la suma histórica de puntos obtenidos, el promedio histórico de puntos, la diferencia de goles a favor o en contra histórica, entre otros.

## Records históricos:



Por último, el tercer análisis muestra distintos records de estadísticas logradas junto al equipo que las consiguió. Además, se muestra el top 5 de equipos líderes en cuanto a puntos históricos y en cuanto a mejores diferencias de goles.

## Futuras líneas:

A continuación, se presentan algunas posibles iniciativas para complementar y mejorar este proyecto:

- Incorporación de estadísticas de jugadores**  
Ampliar la base de datos con información individual de futbolistas (goles, asistencias, tarjetas, minutos jugados) permitiría realizar análisis más detallados de rendimiento, identificar a los máximos goleadores de cada temporada o construir rankings por posición.
- Análisis por técnico o entrenador**  
Relacionar temporadas con entrenadores permitiría evaluar su impacto en el rendimiento de los equipos, visualizando patrones como eficacia por ciclo o evolución de puntos con diferentes DT.
- Visualización geográfica de los equipos**  
Añadir coordenadas o ubicaciones de los clubes posibilitaría representar datos en mapas, destacando regiones dominantes en distintas épocas o el crecimiento de ciertas zonas.