

Premier League

Analítica Avanzada Para Tu Equipo

Paula Sofía Torres

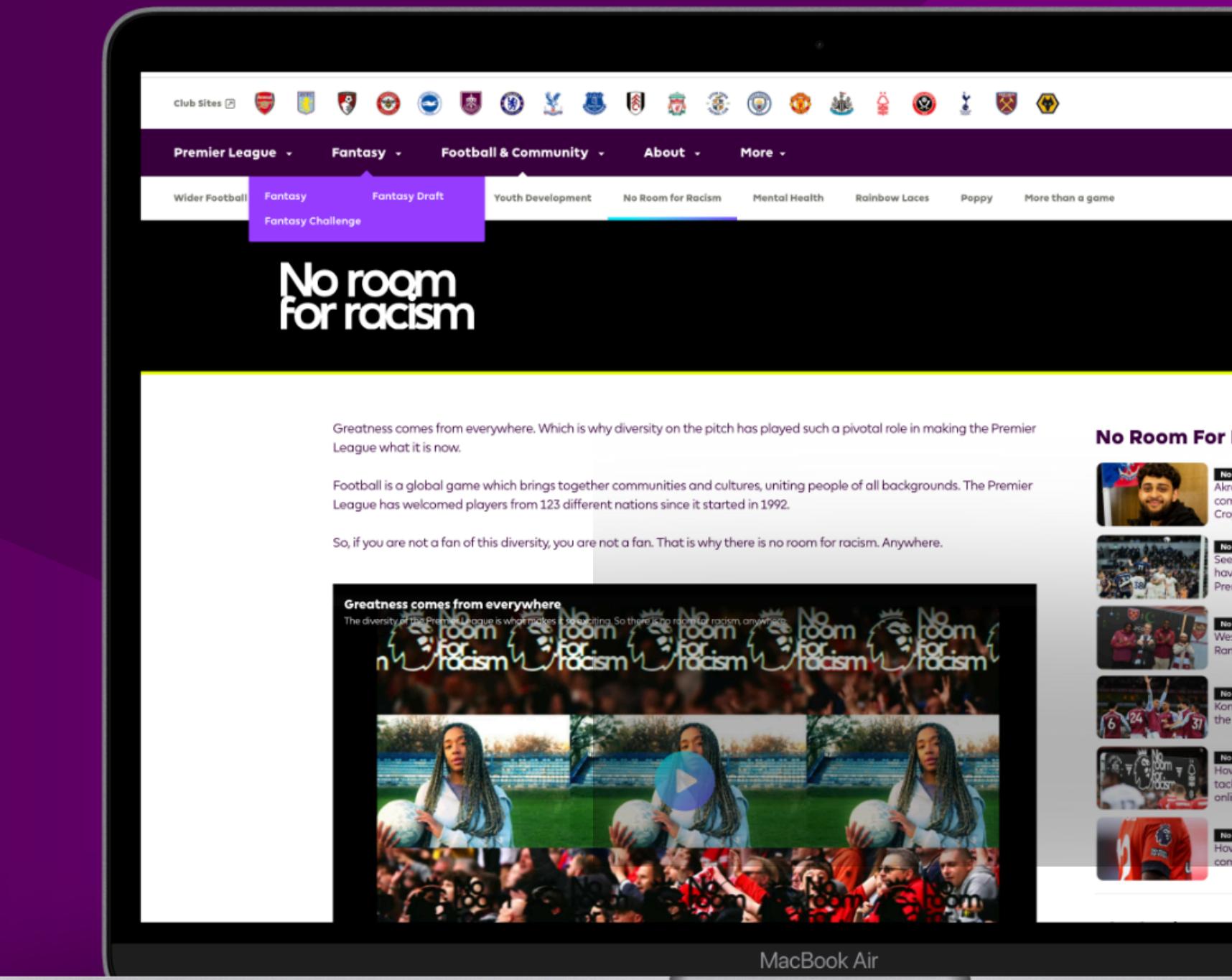


Fuentes de Datos

Transfermarkt: Se uso para estimar el precio promedio a lo largo de la temporada de los jugadores.

Estadísticas proporcionadas por la **Premier League**:

- Características de **rendimiento de los equipos** para cada partido a lo largo de la temporada.
- Características de **rendimiento de los jugadores** en los partidos a lo largo de la temporada.
- **Resultados de los todos los partidos** que se jugaron en el calendario oficial de la temporada.



Se recopilaron y estructuraron datos que abarca múltiples aspectos del rendimiento y características de los jugadores y equipos en la liga.

Incluye información detallada sobre más de 40 variables para cada partido y jugador, desde estadísticas de juego como goles, pases y tarjetas, hasta datos personales de los jugadores y características de los equipos.



Por jugador

```
root
|-- game_id: integer (nullable = true)
|-- team_id: integer (nullable = true)
|-- team_name: string (nullable = true)
|-- date: date (nullable = true)
|-- player_name: string (nullable = true)
|-- player_id: integer (nullable = true)
|-- player_position_info: string (nullable = true)
|-- player_rating: float (nullable = true)
|-- good_high_claim: integer (nullable = true)
|-- touches: integer (nullable = true)
|-- saves: integer (nullable = true)
|-- total_pass: integer (nullable = true)
|-- formation_place: integer (nullable = true)
|-- accurate_pass: integer (nullable = true)
|-- aerial_won: integer (nullable = true)
|-- aerial_lost: integer (nullable = true)
|-- fouls: integer (nullable = true)
|-- total_scoring_att: integer (nullable = true)
|-- total_tackle: integer (nullable = true)
|-- won_contest: integer (nullable = true)
|-- penalty_conceded: integer (nullable = true)
|-- blocked_scoring_att: integer (nullable = true)
|-- man_of_the_match: integer (nullable = true)
|-- goals: integer (nullable = true)
|-- att_pen_goal: integer (nullable = true)
|-- goal_assist: integer (nullable = true)
|-- six_yard_block: integer (nullable = true)
|-- yellow_card: integer (nullable = true)
|-- error_lead_to_goal: integer (nullable = true)
|-- post_scoring_att: integer (nullable = true)
|-- second_yellow: integer (nullable = true)
|-- red_card: integer (nullable = true)
|-- clearance_off_line: integer (nullable = true)
|-- own_goals: integer (nullable = true)
|-- last_man_tackle: integer (nullable = true)
|-- att_pen_target: integer (nullable = true)
|-- penalty_save: integer (nullable = true)
|-- att_pen_miss: integer (nullable = true)
|-- att_pen_post: integer (nullable = true)
```

Por equipo

```
root
|-- game_id: integer (nullable = true)
|-- team_id: integer (nullable = true)
|-- team_name: string (nullable = true)
|-- team_rating: float (nullable = true)
|-- date: string (nullable = true)
|-- won_corners: integer (nullable = true)
|-- won_contest: integer (nullable = true)
|-- total_tackle: integer (nullable = true)
|-- aerial_lost: integer (nullable = true)
|-- possession_percentage: float (nullable = true)
|-- accurate_pass: integer (nullable = true)
|-- total_pass: integer (nullable = true)
|-- shot_off_target: integer (nullable = true)
|-- total_offside: integer (nullable = true)
|-- blocked_scoring_att: integer (nullable = true)
|-- ontarget_scoring_att: integer (nullable = true)
|-- aerial_won: integer (nullable = true)
|-- goals: integer (nullable = true)
|-- att_pen_goal: integer (nullable = true)
|-- post_scoring_att: integer (nullable = true)
|-- Salvadas Portero: integer (nullable = true)
|-- Total Shots: integer (nullable = true)
|-- Fouls: integer (nullable = true)
|-- Yellow Cards: integer (nullable = true)
|-- Red Cards: integer (nullable = true)
|-- Resultado: string (nullable = true)
```



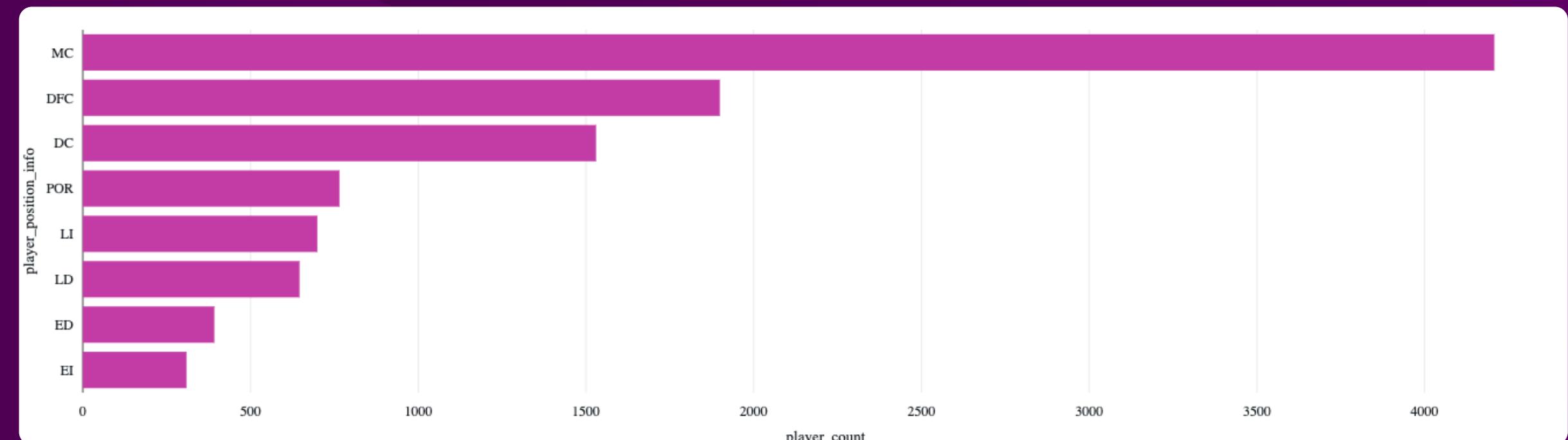
Sobre los jugadores



team_name	player_name	avg_player_ratin
Liverpool FC	Philippe Coutinho	8.03
Manchester City FC	Sergio Agüero	7.81
Arsenal FC	Alex Oxlade Chamberlain	7.81
Manchester City FC	Kevin De Bruyne	7.80
Liverpool FC	Mohamed Salah	7.69

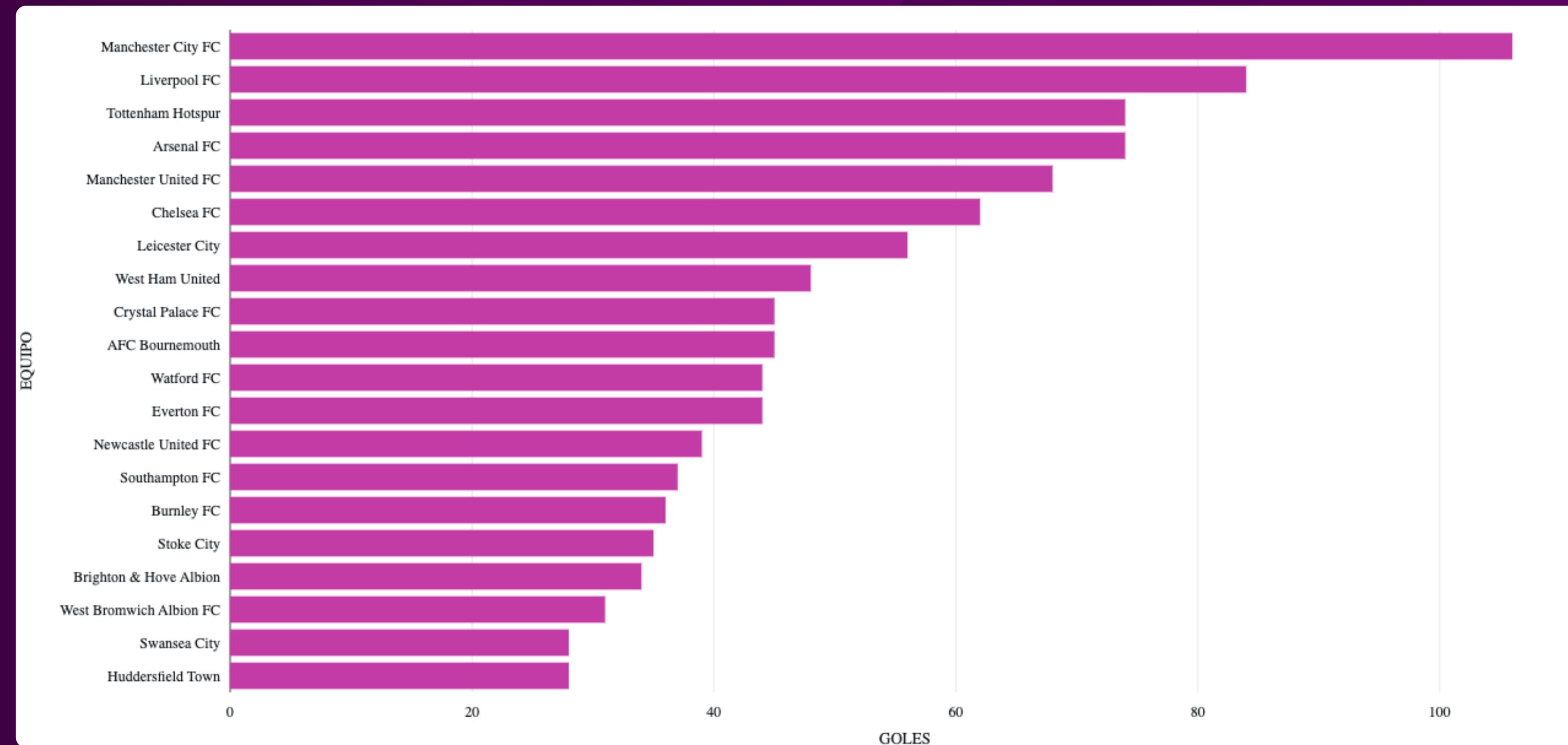
player_name	total_foul	total_yellow_card	total_red_card
Luka Milivojevic	69	8	0
Jordan Ayew	63	2	1
Glenn Murray	60	9	0
Wilfred Ndidi	60	6	2
Erik Pieters	58	5	0
Granit Xhaka	57	10	0
Roberto Firmino	57	1	0

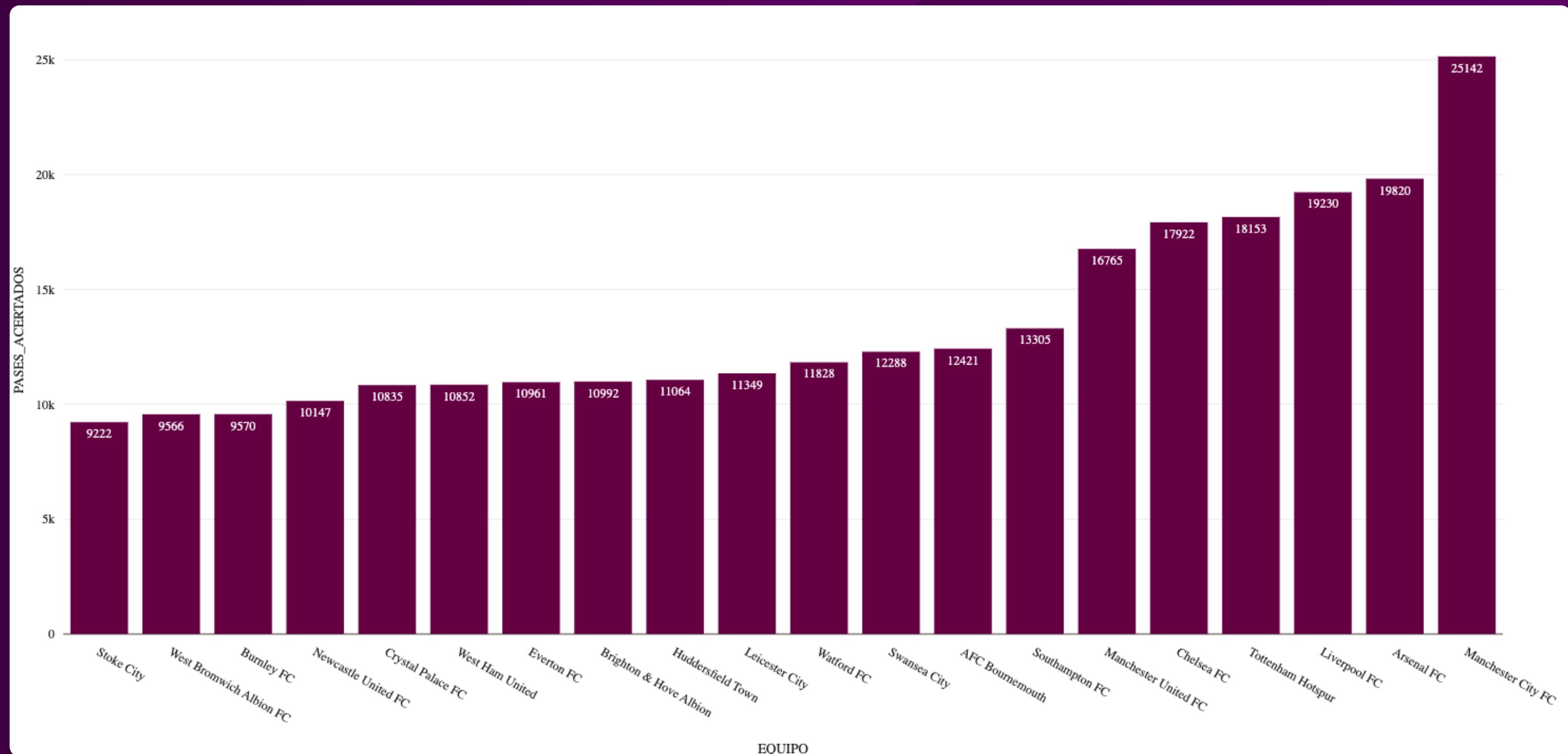
player_name	total_goal
Mohamed Salah	32
Harry Kane	30
Sergio Agüero	21
Jamie Vardy	20
Raheem Sterling	18
Romelu Lukaku	16
Roberto Firmino	15

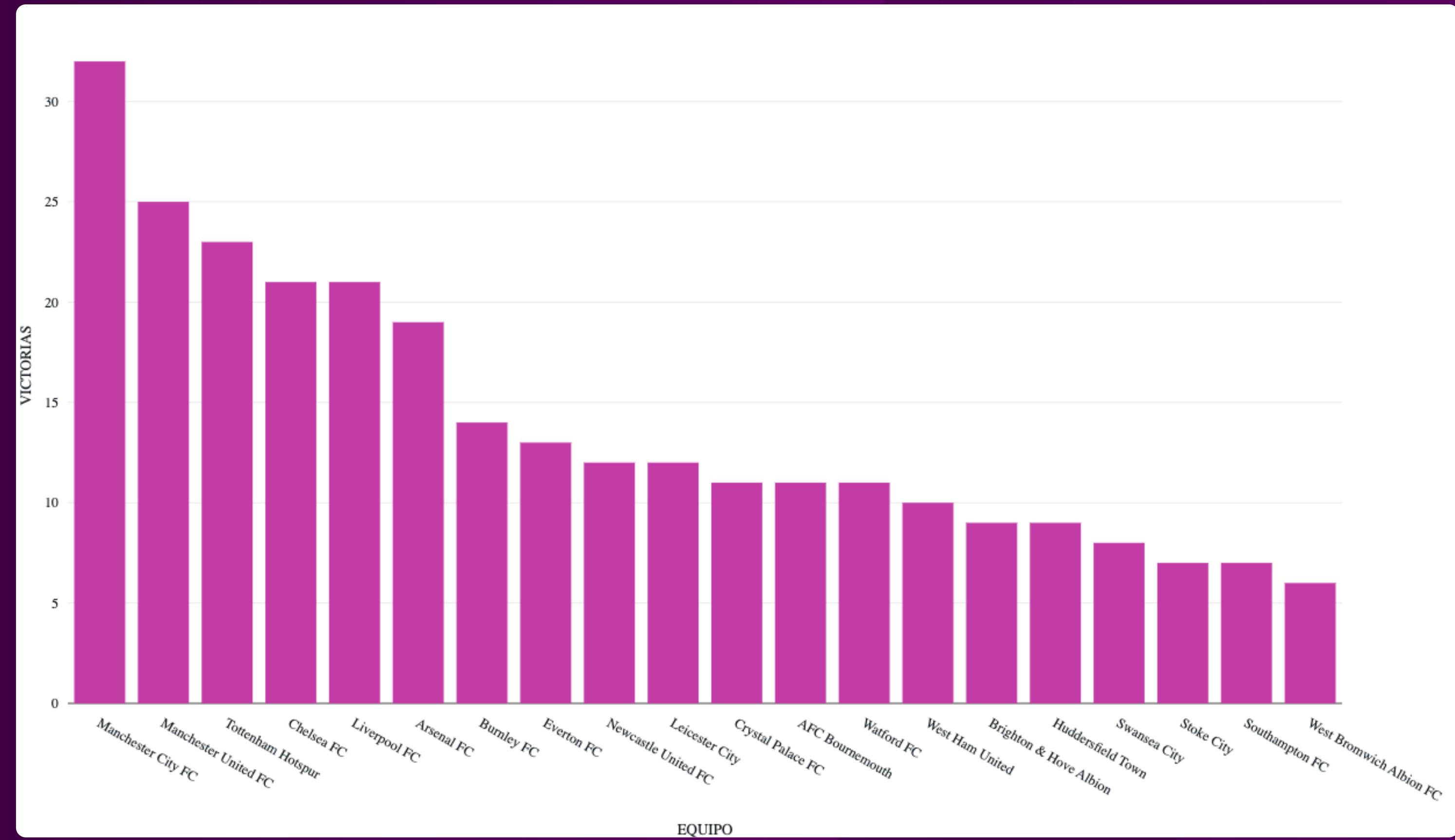


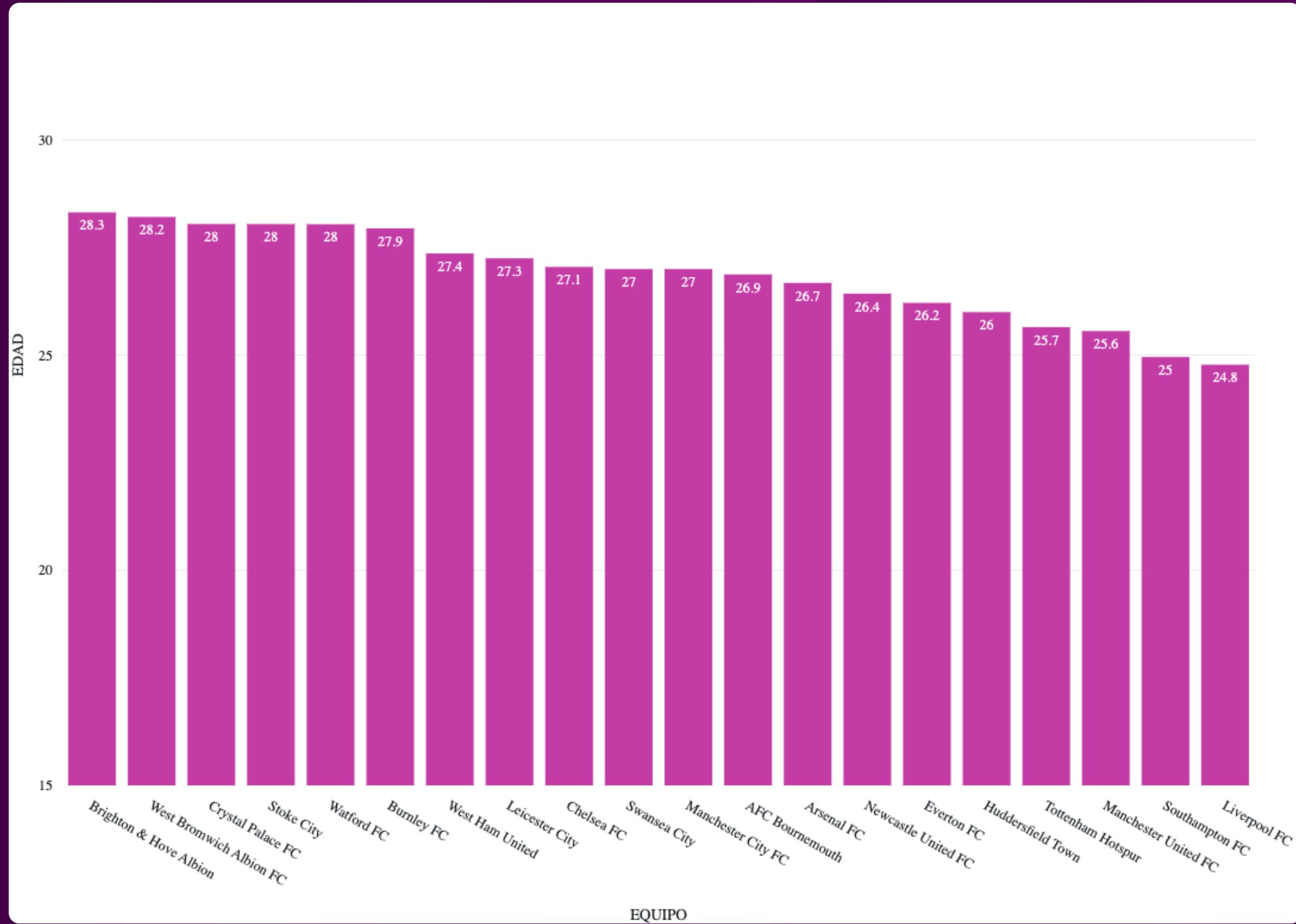
Sobre los equipos











**¿Quiénes son los 11 jugadores más
valiosos e imprescindibles para
conformar un equipo de élite en la liga
inglesa?**



Performance Score

Posición	Fórmula de Rendimiento
Delantero Centro (DC)	$\text{performance} = (\text{goals} \times 2.38 + \text{goal_assist} \times 0.95 + \text{total_scoring_att} \times 1.19 + \text{att_pen_goal} \times 1.43 + \text{won_contest} \times 0.48 - \text{fouls} \times 0.24 + \text{man_of_the_match} \times 3.33)$
Defensa Central (DFC)	$\text{performance} = (\text{total_tackle} \times 1.5 + \text{aerial_won} \times 2.0 + \text{clearance_off_line} \times 1.0 + \text{blocked_scoring_att} \times 1.7 + \text{man_of_the_match} \times 2.0 + \text{last_man_tackle} \times 1.5 + \text{accurate_pass} \times 0.3)$
Lateral Izquierdo/Derecho (LI/LD)	$\text{performance} = (\text{total_tackle} \times 1.04 + \text{aerial_won} \times 1.04 + \text{won_contest} \times 2.08 - \text{fouls} \times 0.63 + \text{accurate_pass} \times 1.46 + \text{man_of_the_match} \times 3.75)$
Portero (POR)	$\text{performance} = (\text{saves} \times 3 + \text{penalty_save} \times 1.5 + \text{good_high_claim} \times 1.5 - \text{error_lead_to_goal} \times 3.5 + \text{accurate_pass} \times 0.5)$
Medio Centro (MC)	$\text{performance} = (\text{accurate_pass} \times 1.0 + \text{total_tackle} \times 1.2 + \text{won_contest} \times 1 + \text{goal_assist} \times 2.0 + \text{goals} \times 2.2 + \text{man_of_the_match} \times 2 + \text{total_pass} \times 0.7 + \text{aerial_won} \times 1.3)$
Extremo Izquierdo/Derecho (EI/ED)	$\text{performance} = (\text{goals} \times 2.5 + \text{goal_assist} \times 1.25 + \text{total_scoring_att} \times 1.04 + \text{accurate_pass} \times 0.65 + \text{won_contest} \times 1.65 + \text{man_of_the_match} \times 2.92)$



Dashboard de resultados



Modelos



Regresión Logística



Propósito

- **Actualizar las probabilidades de ganar o perder en tiempo real:** Proporcionando una base de información que permite al equipo **tomar decisiones tácticas inmediatas**, como ajustar la defensa o intensificar el ataque durante el juego.
- **Evaluación del rendimiento del entrenador:** Sirviendo como **indicador de las decisiones tomadas** y su efectividad, ayudando a guiar y evaluar el desempeño del entrenador.

probability	Resultado	prediction
[0.9999979968363805, 2.003163619535009E-6]	0	0.0
[0.00108296177394623, 0.9989170382260538]	1	1.0
[0.9977277725633367, 0.00227227436663332]	0	0.0
[0.9199420408560445, 0.08005795914395553]	0	0.0
[0.21908395002952935, 0.7809160499704706]	1	1.0
[0.06934926365252528, 0.9306507363474747]	1	1.0
[0.9978776463827057, 0.0021223536172942925]	0	0.0
[0.07929259781218825, 0.9207074021878118]	1	1.0
[0.08324242751024516, 0.9167575724897549]	1	1.0
[0.9999983292221541, 1.6707778458524558E-6]	0	0.0



Elección Variables

- **Variable Objetivo:** La variable objetivo es **Resultado** ya que es binaria y de ella podemos tener las probabilidades de saber si el equipo es más propenso a perder o ganar en el momento.
- **Variables Explicativas:** Algunas de las variables que se incluyen para explicar el modelo son el **rating del equipo**, el número de **tiros de esquina**, de **duelos ganados**, de **tarjetas**, de **faltas**, de **opciones de marcar**, de **tiros a puerta**, de **salvadas del portero**, etc.

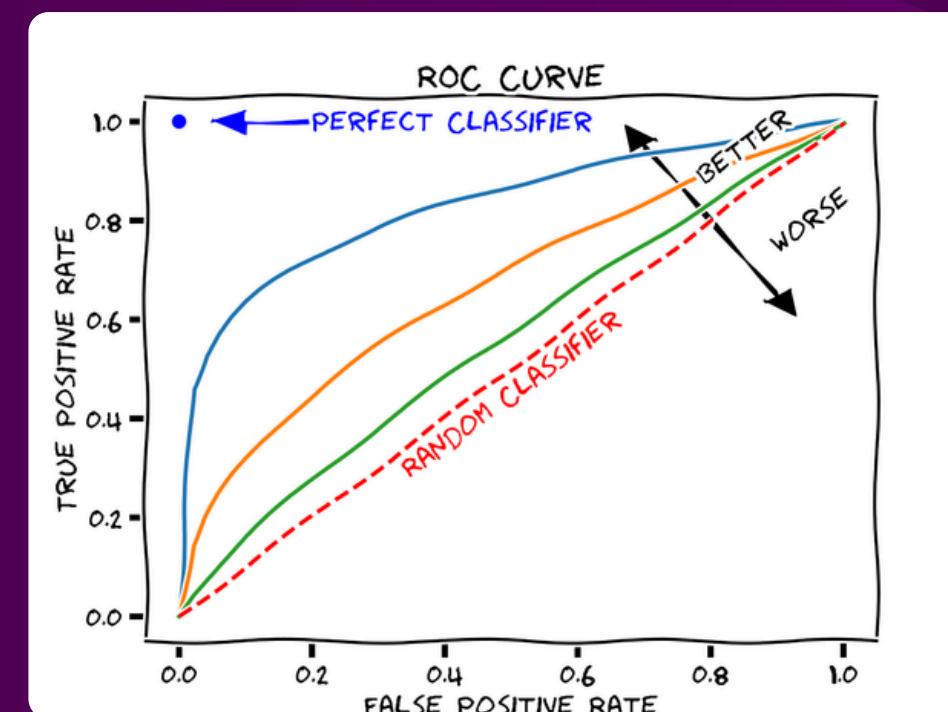


Para analizar el **rendimiento** de los **diferentes modelos de regresión logística**, se utilizaron diversas configuraciones de regularización **Regularización De Ridge** que es más severa castigando los coeficientes y un número **máximo de iteraciones**.

Las métricas utilizadas fueron **AUC** y **Accuracy**.

Accuracy		Max Iterations		
		1	2	5
Regularización Ridge	0.001	0.781395	0.865116	0.92093
	0.5	0.753488	0.767442	0.767442
	1	0.725581	0.744186	0.739535

AUC		Max Iterations		
		1	2	5
Regularización Ridge	0.001	0.875912	0.943758	0.964814
	0.5	0.875912	0.923451	0.935055
	1	0.875912	0.909695	0.915965



Análisis Serie De Tiempo



Propósito

Se busca analizar el puntaje que indica el rendimiento de los equipos a lo largo de la temporada para predecir cuál será el rendimiento en los siguientes partidos.

Esto le permitiría a los inversionistas:

- Los inversionistas pueden **comparar** el rendimiento proyectado de varios equipos y decidir cuál ofrece la mejor oportunidad de retorno sobre la inversión.
- Con predicciones precisas del rendimiento futuro, los inversionistas pueden valorar mejor los **riesgos** asociados con sus inversiones.
- Al modelar el rendimiento de los equipos a lo largo de la temporada Esta información es valiosa para tomar decisiones informadas sobre **en qué equipos invertir y cuándo hacerlo.**



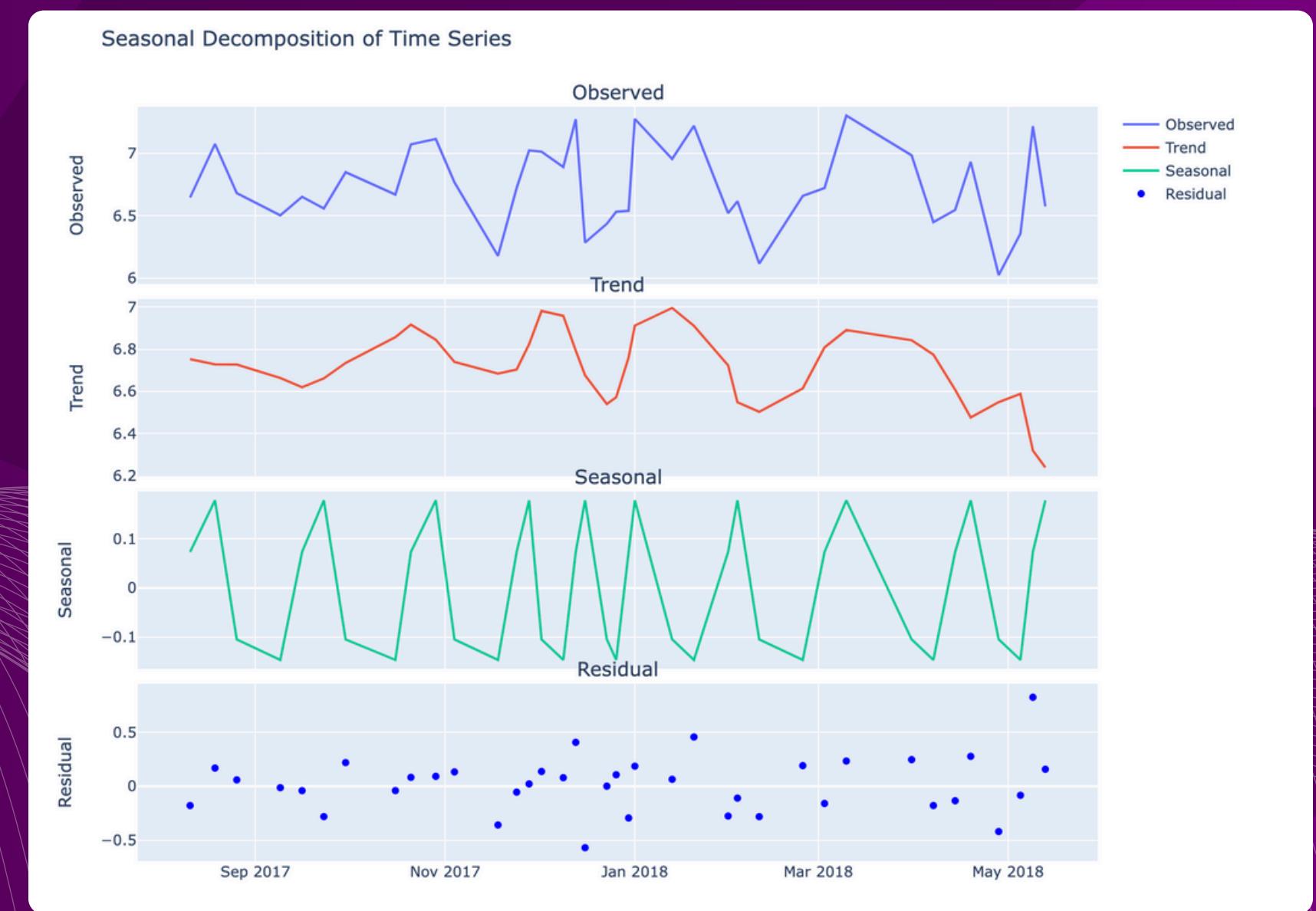
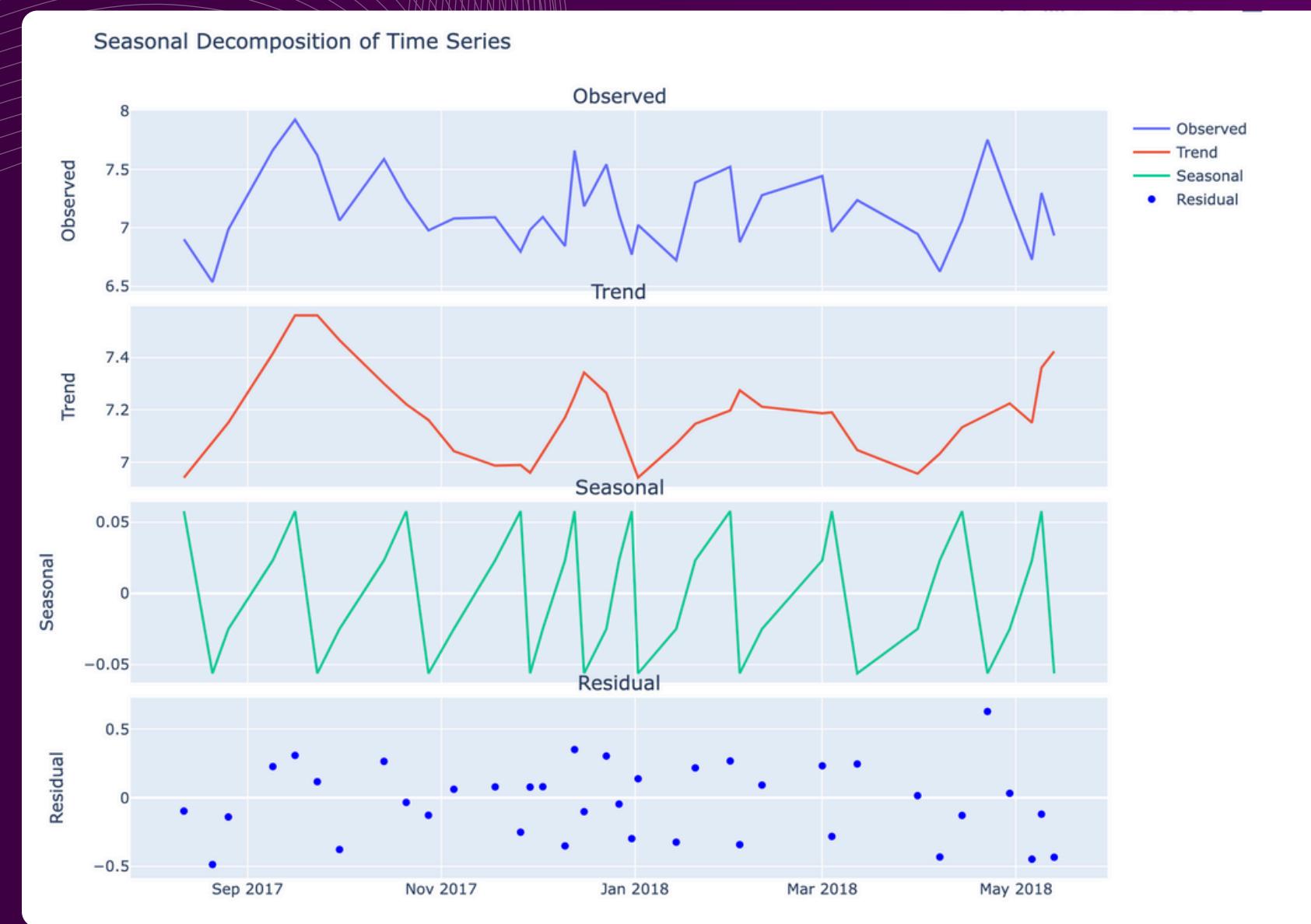
1	Man City
2	Man United
3	Tottenham ▲
4	Liverpool
5	Chelsea
6	Arsenal
7	Burnley
8	Everton
9	Leicester



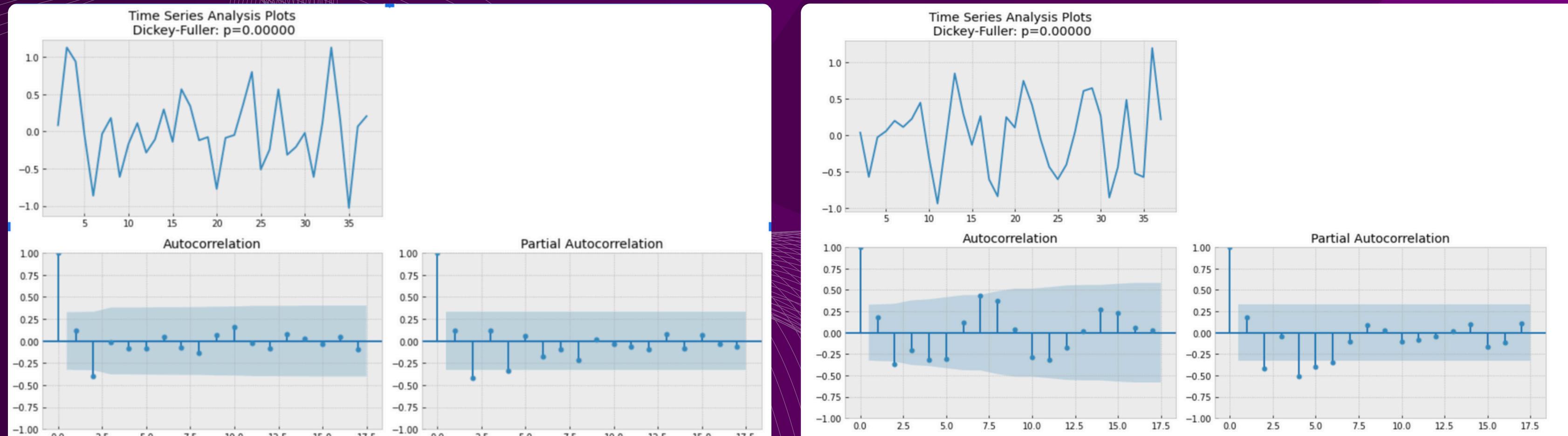
Descomposición de las Series De Tiempo



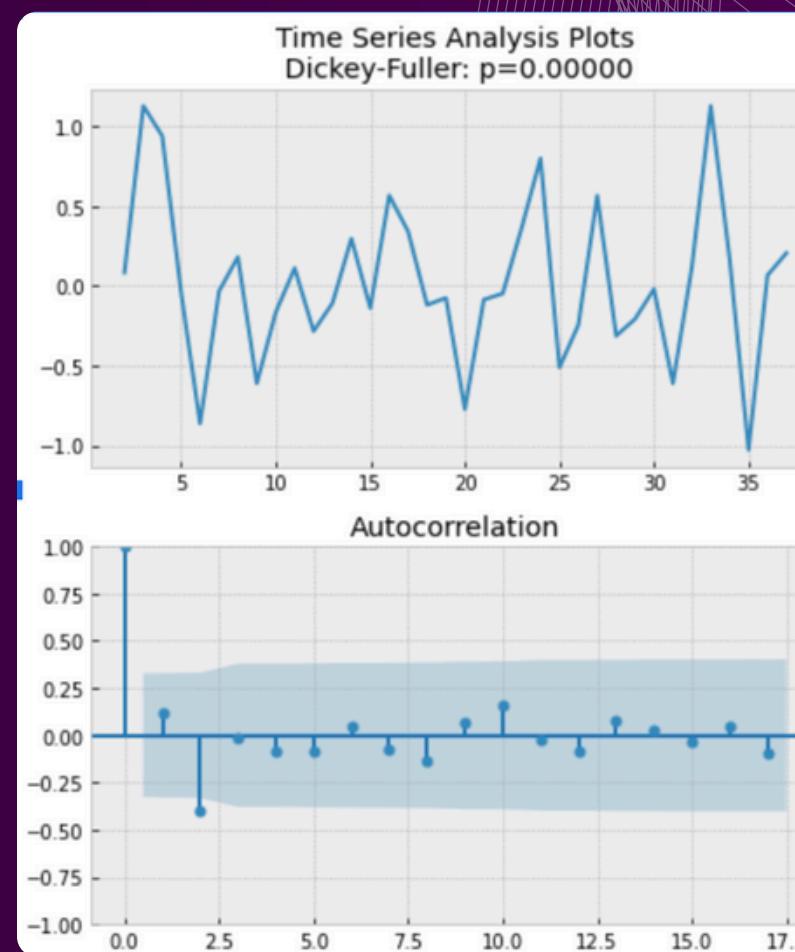
```
decomposition = seasonal_decompose(man_city_data['team_rating'], model='additive', period=4, extrapolate_trend='freq')
```



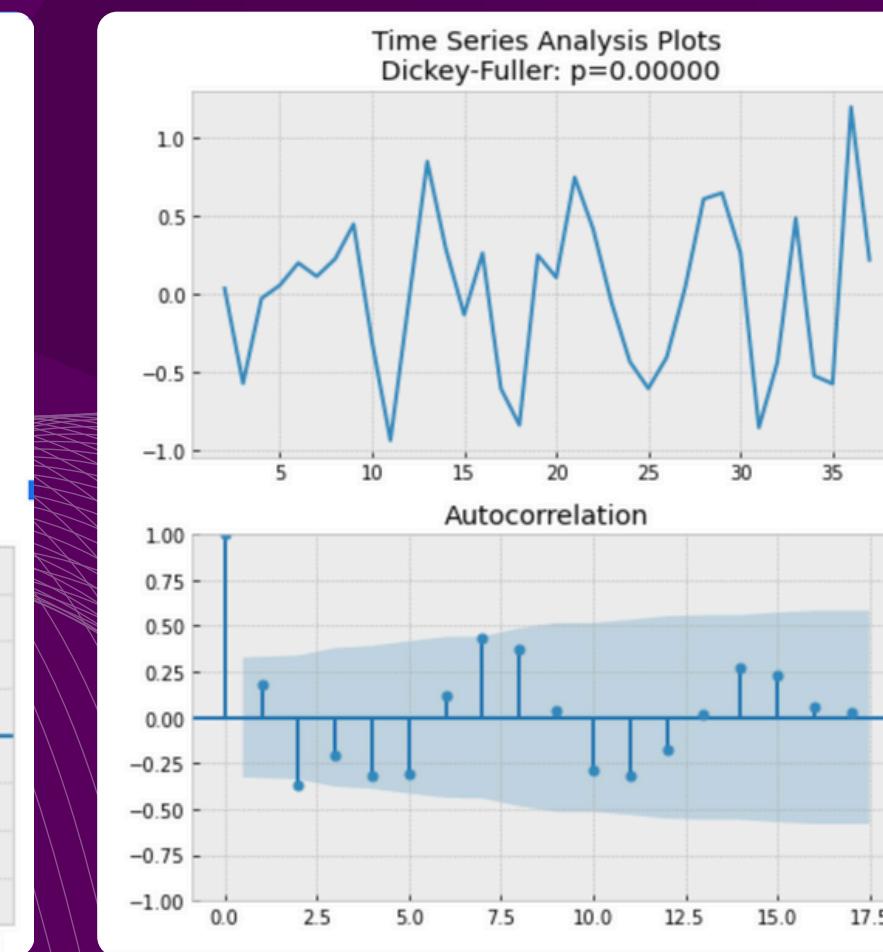
Selección de parámetros y prueba de estacionariedad



Selección de parámetros y prueba de estacionariedad



```
# Definir valores iniciales
ps = range(2, 5)
d = 1
qs = range(0, 2)
Ps = range(0, 2)
D = 0
Qs = range(0, 2)
s = 4
```



```
# Definir valores iniciales
ps = range(2, 6)
d = 1
qs = range(0, 2)
Ps = range(0, 2)
D = 0
Qs = range(0, 2)
s = 4
```



```
model=sm.tsa.statespace.SARIMAX(man_city_pd['team_rating'], order=(param[0], d, param[1]),  
                                 seasonal_order=(param[2], D, param[3], s)).fit(disp=-1)
```

```
# Definir valores iniciales  
ps = range(2, 5)  
d = 1  
qs = range(0, 2)  
Ps = range(0, 2)  
D = 0  
Qs = range(0, 2)  
s = 4
```



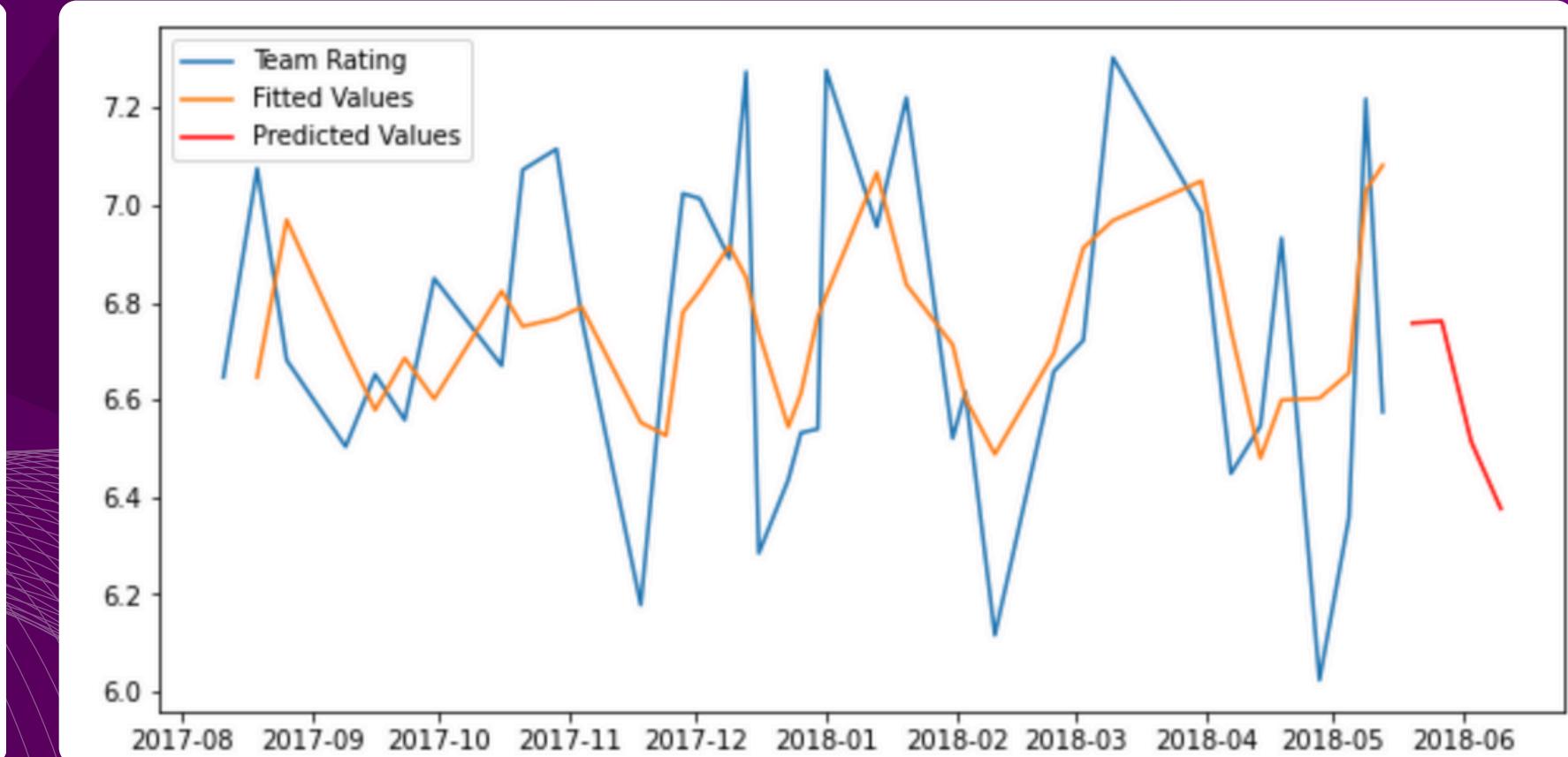
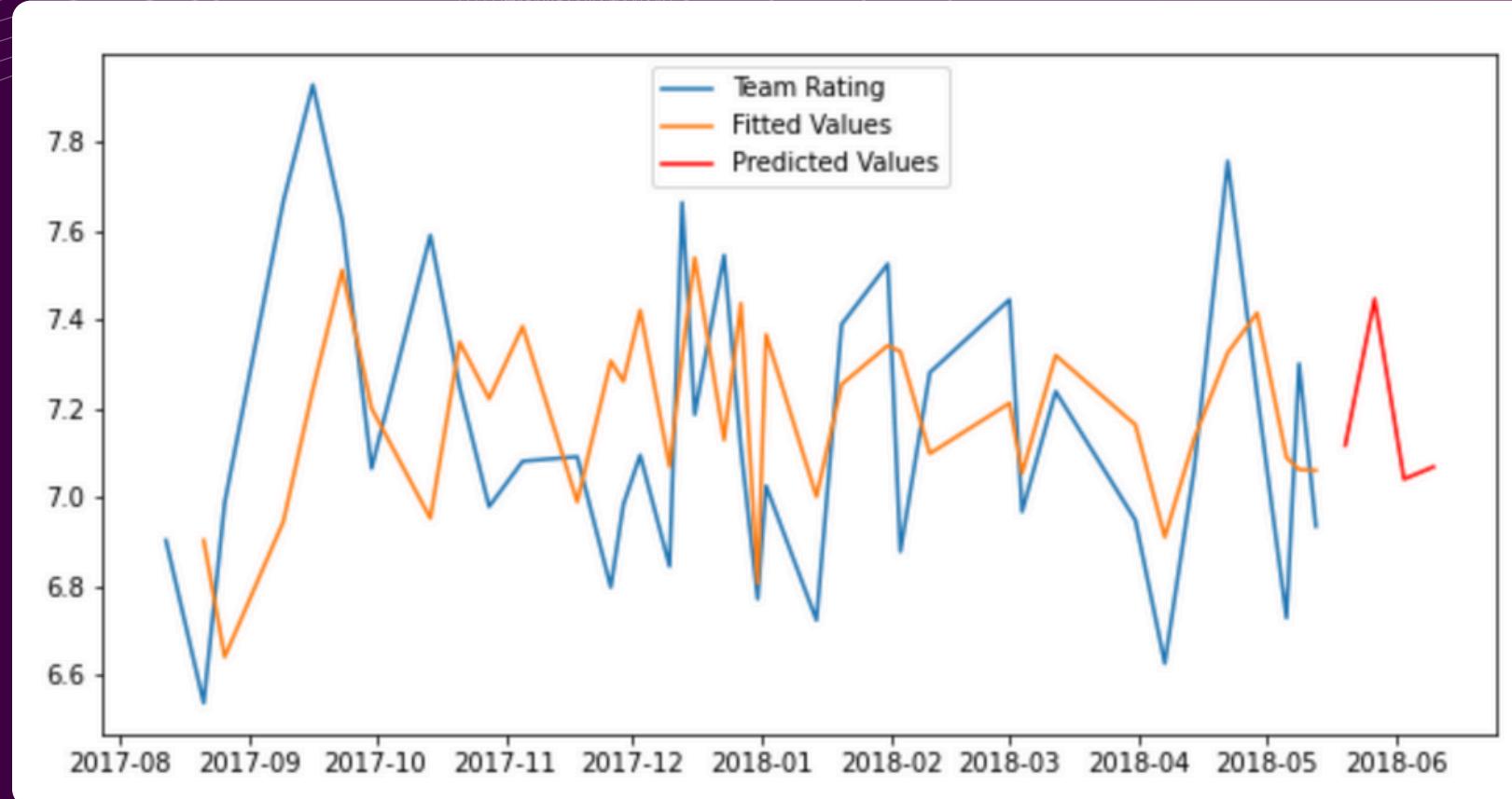
	parameters	aic
1	▶ [2, 1, 0, 1]	29.283520206079313
2	▶ [3, 0, 0, 1]	30.255124022894183
3	▶ [2, 1, 1, 1]	31.11029018696127
4	▶ [3, 1, 0, 1]	31.175119679902565
5	▶ [3, 0, 1, 1]	31.61022674188798

```
# Definir valores iniciales  
ps = range(2, 6)  
d = 1  
qs = range(0, 2)  
Ps = range(0, 2)  
D = 0  
Qs = range(0, 2)  
S = 4
```

	parameters	aic
1	▶ [5, 1, 0, 1]	24.290933099299806
2	▶ [5, 1, 0, 0]	24.76078996093778
3	▶ [4, 1, 0, 1]	27.498257933972894
4	▶ [3, 0, 1, 0]	29.969521794608998
5	▶ [3, 1, 1, 0]	30.127833578862184
6	▶ [3, 0, 1, 1]	31.162825952657056



Predicciones



```
mean_absolute_percentage_error: 0.06447429897212295  
median_absolute_error: 0.27775327996369237
```

```
mean_absolute_percentage_error: 0.05944849757592226  
median_absolute_error: 0.24106909049886438
```



Decisión de elección de modelos





🚀 Plan Estratégico Nueva York: Innovación y Seguridad



🎯 Objetivo General

Desarrollar estrategias específicas para la ciudad de Nueva York para reducir arrestos y abordar la pobreza, utilizando análisis de datos de 2016 y 2024.

📌 Objetivos Específicos

1. Utilizar K-Means Clustering para identificar patrones en los arrestos.
2. Analizar la relación entre educación, pobreza, etnia, género y tasas de arresto.
3. Desarrollar un modelo de clasificación para determinar el empleo basado en datos de pobreza.

Análisis de Publicidad y Ventas

Este repositorio contiene un análisis detallado sobre el impacto de los presupuestos de publicidad en las ventas, así como los modelos de regresión lineal desarrollados para predecir las ventas en función de los presupuestos de publicidad en diferentes medios.

Propósito del Análisis

El propósito de este análisis es examinar la influencia de los presupuestos de publicidad en radio y televisión sobre las ventas totales. Se pretende comparar la eficacia de estos dos medios de publicidad para determinar cuál tiene una mayor relación lineal con el aumento de ventas.

🏆 Oscars Awards Analysis

🎬 Descripción del Proyecto

Este proyecto 📈 se centra en el análisis de los premios Oscar, con un enfoque particular en la categoría de Mejor Película Internacional 🌎. Utilizamos técnicas de procesamiento de lenguaje natural 🗣️ para extraer insights de las sinopsis de las películas y explorar diversas temáticas y tendencias a lo largo de los años.

📁 Estructura del Repositorio

El repositorio se organiza de la siguiente manera:

- `main.ipynb` : Notebook principal que contiene el flujo de trabajo completo del análisis.
- `oscars_s.csv` : Dataset original de los premios Oscar.
- `oscars_sfinales.csv` : Dataset procesado y limpio tras la limpieza de datos.

Stock Trend Analysis for Ecopetrol📈

Team Members

- Santiago Moreno
- Juan David Lopez
- Paula Sofía Torres
- Mauricio Rodriguez
- Sofia Catalina Galindo

From Pontificia Universidad Javeriana

Objective

