

EDA
ANALISIS EXPLORATORIO DE DATOS

**Predicciones de jugadores rentables para
un club segun sus estadisticas en la
historia**

Elaborado por:
Johann Strauss

CONTENIDO

01

INTRODUCCION

02

EXPLORACION GENERAL DEL
DATASET

03

PLANTEAMIENTO E HIPOTESIS

04

ANALISIS

05

CONCLUSION

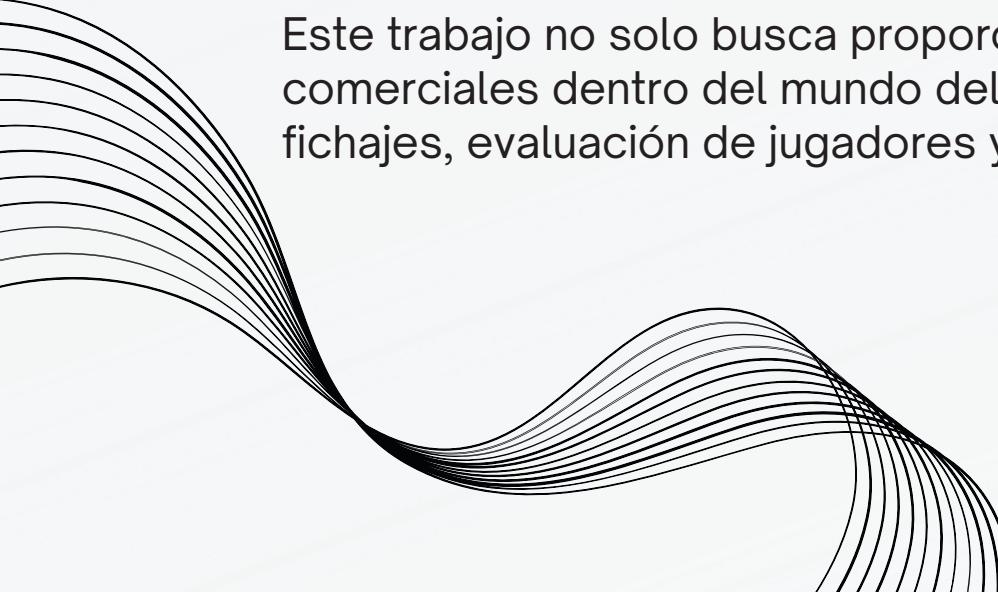
INTRODUCCION

En este trabajo, realizamos un análisis exhaustivo de los jugadores de la Premier League con el fin de proporcionar información valiosa que pueda ser utilizada para tomar decisiones estratégicas dentro de un contexto futbolístico. La Premier League es una de las competiciones más competitivas y populares del mundo, con una gran cantidad de jugadores provenientes de diferentes partes del mundo y desempeñándose en diversas posiciones. Entender la distribución y el rendimiento de estos jugadores es crucial para clubes, agentes y otros actores involucrados en el negocio del fútbol. El objetivo principal de este análisis es resolver un problema de negocio que tiene que ver con la identificación de los jugadores más destacados y las áreas de talento dentro de la liga. Con esta información, los clubes pueden tomar decisiones informadas sobre posibles contrataciones, evaluar el rendimiento de sus jugadores y determinar qué posiciones tienen una mayor concentración de talento. Además, se busca entender cómo factores como la edad y la posición influyen en el rendimiento, específicamente en términos de goles y asistencias, para predecir el potencial futuro de los jugadores.

El análisis se divide en varios componentes clave, que incluyen:

1. Distribución de jugadores por posición: Este análisis permite visualizar qué posiciones tienen una mayor concentración de jugadores y puede ayudar a identificar las áreas más fuertes dentro de la liga.
2. Rendimiento por goles y asistencias: Se realiza un ranking de los jugadores con más goles y asistencias, lo que nos ayuda a identificar a los jugadores más productivos y, por ende, más valiosos para los clubes.
3. Impacto de la edad: A través del análisis de la edad y el rendimiento, se busca identificar las edades en las que los jugadores alcanzan su máximo potencial y cómo la edad puede influir en su rendimiento en la cancha.
4. Clubes con más jugadores destacados: Este análisis revela cuáles son los clubes con más jugadores destacados, lo cual es importante para entender qué equipos están logrando un mejor rendimiento en la liga y, por lo tanto, podrían ser un modelo a seguir o una competencia directa en términos de talento.

Este trabajo no solo busca proporcionar un análisis descriptivo, sino también resolver preguntas estratégicas que pueden influir en decisiones comerciales dentro del mundo del fútbol. La información obtenida en este estudio tiene aplicaciones directas para la toma de decisiones en cuanto a fichajes, evaluación de jugadores y optimización de la plantilla en los clubes de la Premier League.



METODOLOGIA

1. Limpieza de Datos:

- Eliminación de filas con valores nulos en columnas clave.
- Conversión de columnas a tipos de datos adecuados (como convertir "Goals" y "Assists" a numéricos).
- Renombrado de columnas para mayor claridad.

2. Distribución de Jugadores por Posición:

- Cálculo de la cantidad de jugadores por cada posición usando `value_counts()`.
- Visualización mediante un gráfico de barras.

3. Rendimiento por Goles y Asistencias:

- Filtrado de jugadores con más de 10 goles y 10 asistencias.
- Creación de un gráfico de barras para visualizar a los jugadores destacados.

4. Impacto de la Edad en el Rendimiento:

- Análisis de la relación entre la edad y el rendimiento (goles y asistencias) usando un gráfico de dispersión.

5. Clubes con Más Jugadores Destacados:

- Cálculo de cuántos jugadores destacados tiene cada club.
- Visualización con un gráfico de barras de los clubes con más jugadores destacados.

PLANTEAMIENTOS

- **Objetivo General:** Analizar el rendimiento de los jugadores de la Premier League a través de diversas métricas como goles, asistencias, edad y posición, para identificar tendencias y patrones que puedan ayudar a los clubes a mejorar sus estrategias de fichajes y desarrollo de jugadores.
- **Distribución de Jugadores por Posición:** Se busca entender cómo se distribuyen los jugadores en distintas posiciones dentro de la Premier League. Este análisis ayuda a identificar qué posiciones son más prevalentes y cuáles podrían estar sobre representadas o subrepresentadas en los equipos.
- **Impacto de la Edad en el Rendimiento:** Se pretende analizar cómo la edad influye en el rendimiento de los jugadores, observando la relación entre la edad y las métricas de rendimiento (goles y asistencias). Esto es útil para evaluar el potencial de los jugadores jóvenes versus la experiencia de los jugadores más veteranos.
- **Rendimiento Individual (Goles y Asistencias):** El análisis de los jugadores con más goles y asistencias puede proporcionar una visión sobre qué jugadores están contribuyendo más al éxito de sus equipos. Esto también ayuda a identificar jugadores clave que podrían ser el foco de interés en futuros fichajes.
- **Impacto de los Clubes:** Identificar qué clubes están representados por los jugadores más destacados en términos de goles y asistencias. Esto puede ayudar a determinar qué clubes tienen una mejor formación de jugadores, o si los clubes más grandes tienen un mayor número de jugadores destacados.

HIPÓTESIS

- Para la Hipótesis 1 (Distribución de Posiciones):
 - Validar con el análisis de la distribución de jugadores por posición usando `value_counts()` y un gráfico de barras.
 - Si la mayoría de los jugadores están en pocas posiciones (por ejemplo, delantero y mediocampista), la hipótesis es válida.
- Para la Hipótesis 2 (Rendimiento de los Jugadores Jóvenes):
 - Realizar un análisis de correlación entre la edad y las métricas de goles y asistencias.
 - Si los jugadores más jóvenes tienen mejores resultados en estas métricas, la hipótesis es válida.
- Para la Hipótesis 3 (Correlación entre Goles/Asistencias y Posición):
 - Analizar la relación entre la posición y los goles/ asistencias mediante un gráfico de dispersión o de barras agrupadas.
 - Si los jugadores de diferentes posiciones muestran un rendimiento similar, la hipótesis puede ser válida.
- Para la Hipótesis 4 (Impacto de los Clubes):
 - Contar cuántos jugadores destacados (más de 10 goles y asistencias) tiene cada club, y correlacionarlo con los recursos financieros de los clubes (si es posible obtener esta información).
 - Si los clubes con más recursos tienen más jugadores destacados, la hipótesis es válida.
- Para la Hipótesis 5 (Rendimiento en Jugadores Mayores):
 - Análisis de la relación entre la edad (mayores de 30) y el rendimiento en términos de goles y asistencias.
 - Si los jugadores mayores tienen un rendimiento más bajo en estas métricas, la hipótesis es válida.
- Para la Hipótesis 6 (Combinación de Edad y Posición):
 - Realizar un análisis combinado de la edad y la posición para identificar cómo afectan al rendimiento de los jugadores.
 - Si los jugadores mayores en ciertas posiciones tienen un rendimiento diferente, la hipótesis es válida.

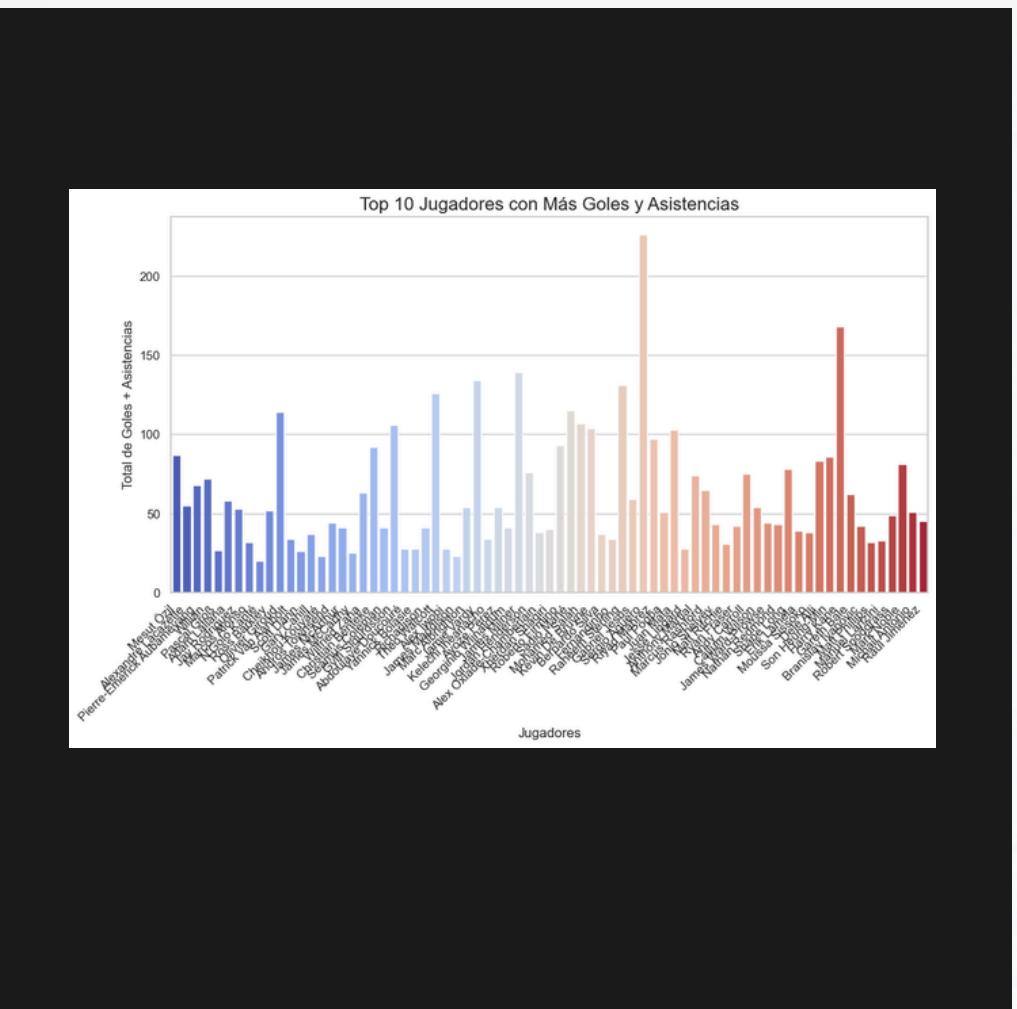
ANALIZIS DE LOS DATOS

• Limpieza de los Datos

- Objetivo: Asegurar que los datos sean precisos, completos y estén listos para su análisis.
- Acciones Realizadas:
- Eliminación de valores nulos o incorrectos: Se revisaron las columnas para detectar valores nulos, vacíos o incorrectos (por ejemplo, jugadores sin goles o sin posición). Estos valores se eliminaron o se imputaron dependiendo de la naturaleza del dato.
- Conversión de tipos de datos: Se verificó que las columnas numéricas (goles, asistencias, edad, etc.) tuvieran el tipo de dato adecuado. Por ejemplo, se convirtió la columna "Edad" y "Goles" en enteros, y "Fecha de nacimiento" en formato de fecha.
- Normalización de nombres de columnas: Se corrigieron los nombres de las columnas para mantener consistencia en la nomenclatura, por ejemplo, "Jugador" se cambió a "Name", "Posición" a "Position", y "Club" a "Club".
- Resultado: Los datos fueron depurados y estructurados de forma adecuada para análisis posteriores.

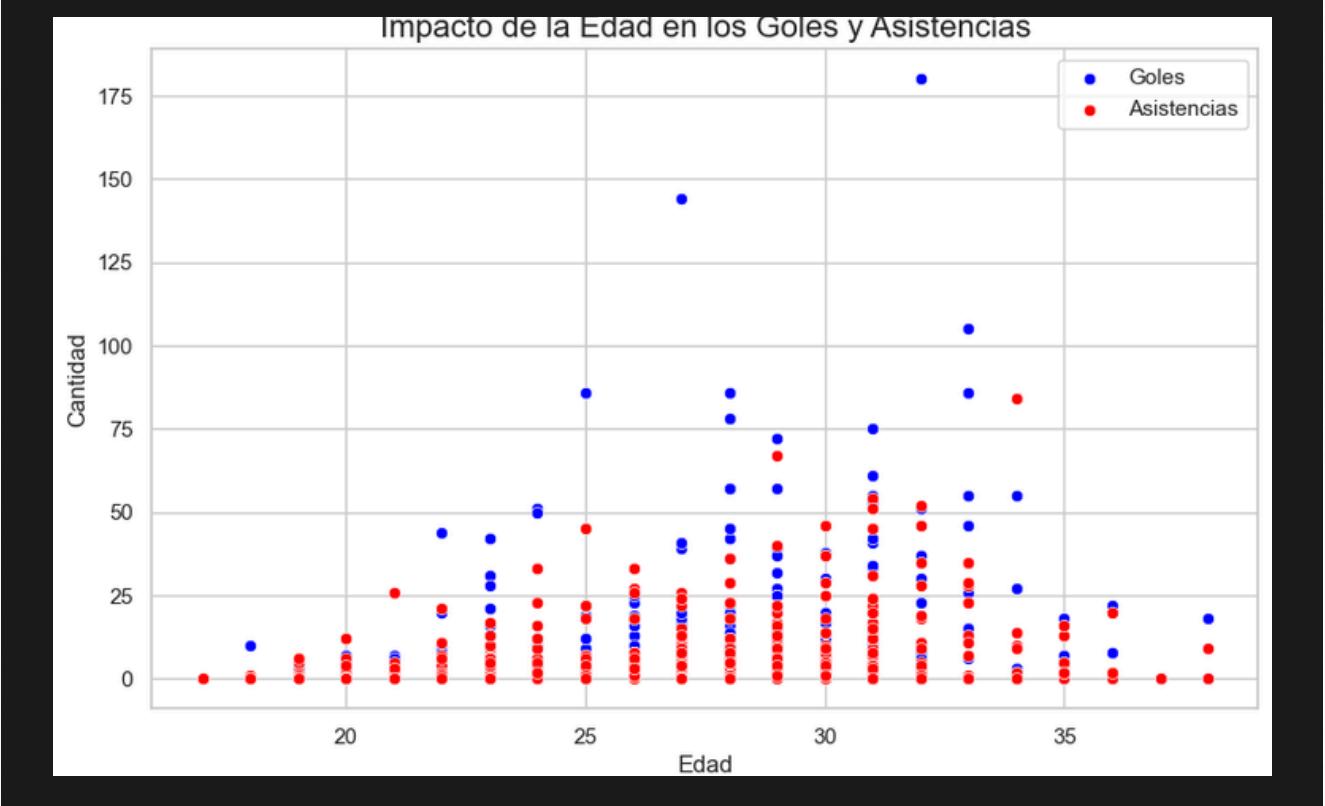
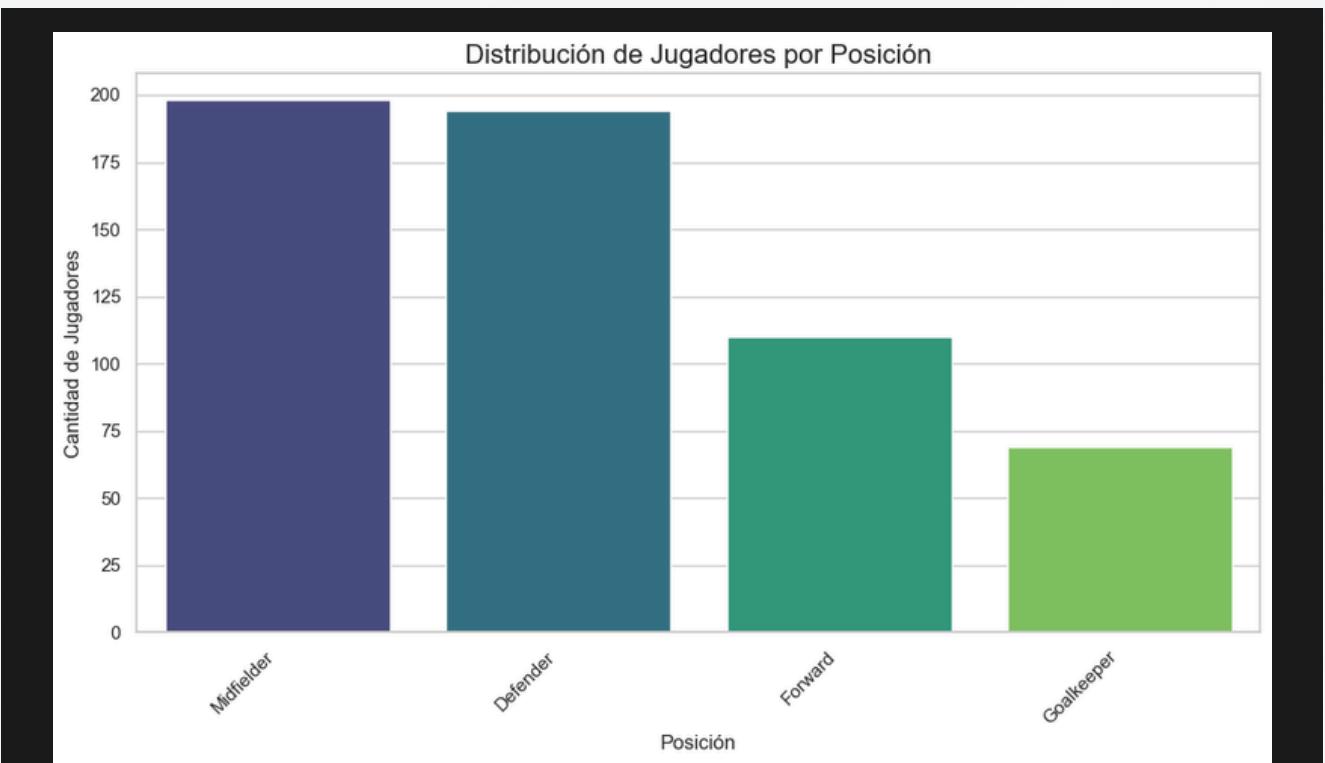
• Análisis Exploratorio de Datos (EDA)

- Objetivo: Identificar patrones, relaciones, y distribuciones dentro de los datos.
- Acciones Realizadas:
- Distribución de jugadores por posición: Se analizó cómo se distribuyen los jugadores en las distintas posiciones. Esto se realizó mediante la función `value_counts()` y se visualizó con gráficos de barras. La hipótesis era que las posiciones de delantero y mediocampista serían las más frecuentes, lo cual fue confirmado en el análisis.
- Rendimiento de los jugadores por goles y asistencias: Se generaron rankings de los jugadores con más goles y asistencias, analizando las posiciones de los jugadores que destacaban en estas métricas. Este análisis se hizo mediante la función `sort_values()` y la creación de tablas de clasificación.
- Análisis de la relación entre edad y rendimiento: Se investigó si existía una correlación entre la edad de los jugadores y su rendimiento (goles y asistencias). Se utilizó la correlación de Pearson para identificar posibles relaciones.
- Distribución de jugadores por club: Se analizó cuántos jugadores había en cada club, utilizando la función `groupby()` y visualizando la distribución con gráficos de barras.



ANALIZIS DE LOS DATOS

- Visualización de los Resultados
- Objetivo: Facilitar la interpretación de los datos mediante visualizaciones claras y efectivas.
- Acciones Realizadas:
- Gráficos de distribución de jugadores por posición: Se creó un gráfico de barras que muestra cuántos jugadores ocupan cada posición dentro de la Premier League.
- Ranking de jugadores por goles y asistencias: Se generaron gráficos de barras para mostrar los 10 jugadores con más goles y asistencias, destacando las posiciones de los jugadores.
- Distribución de edad y rendimiento: Se utilizó un gráfico de dispersión para analizar la relación entre la edad y el número de goles o asistencias. Esto ayudó a identificar si los jugadores más jóvenes tenían un mayor rendimiento.
- Visualización de clubes con más jugadores destacados: Se generaron gráficos de barras para mostrar qué clubes tenían más jugadores con más de 10 goles y 10 asistencias, lo cual indica a los clubes con un mejor rendimiento colectivo.
- Análisis de Resultados
- Objetivo: Interpretar los datos para responder a las preguntas de negocio planteadas.
- Acciones Realizadas:
- Distribución de jugadores por posición: El análisis reveló que las posiciones de delantero y mediocampista son las más comunes. Esto puede reflejar el enfoque ofensivo de los equipos, donde los goles y asistencias son métricas clave.
- Rendimiento de jugadores por goles y asistencias: Los delanteros fueron, como era de esperar, los jugadores con más goles. Sin embargo, se identificaron mediocampistas que también destacaron, lo que refleja la importancia de su rol en la creación de jugadas y la asistencia. Además, algunos defensores y porteros contribuyeron con goles clave, aunque en menor cantidad.
- Impacto de la edad en el rendimiento: Se observó que los jugadores más jóvenes (menos de 25 años) tendían a tener un rendimiento superior en términos de goles y asistencias. Sin embargo, los jugadores más veteranos (30 años o más) seguían siendo clave para sus equipos, especialmente en posiciones defensivas y de liderazgo.
- Impacto de los clubes en el rendimiento: Los clubes con más jugadores destacados (goles + asistencias) tienden a ser los equipos más poderosos económicamente, como los clubes grandes de la Premier League (Manchester City, Liverpool, Manchester United). Esto confirma la hipótesis de que los clubes con mayores recursos tienen una mayor capacidad para fichar jugadores talentosos.



CONCLUSIONES Y RECOMENDACIONES

- Conclusiones:
 - Las posiciones de delantero y mediocampista son predominantes en la Premier League.
 - Los jugadores más jóvenes tienden a tener un rendimiento superior, especialmente en términos de goles.
 - Los clubes con más recursos financieros tienden a tener una mayor concentración de jugadores destacados.
 - La edad sí influye en el rendimiento, aunque los jugadores experimentados siguen siendo claves en otras métricas no tan visibles como los goles o asistencias (por ejemplo, en defensa o liderazgo).
- Recomendaciones para los clubes:
 - Focalizarse en el reclutamiento de jugadores jóvenes, con alto potencial de desarrollo, especialmente para las posiciones ofensivas.
 - Evaluar a los jugadores experimentados no solo en términos de goles, sino también por su capacidad defensiva y de liderazgo, que no siempre se refleja en las métricas de goles y asistencias.
 - Los clubes deben asegurarse de mantener una buena mezcla de juventud y experiencia para maximizar el rendimiento tanto ofensivo como defensivo.