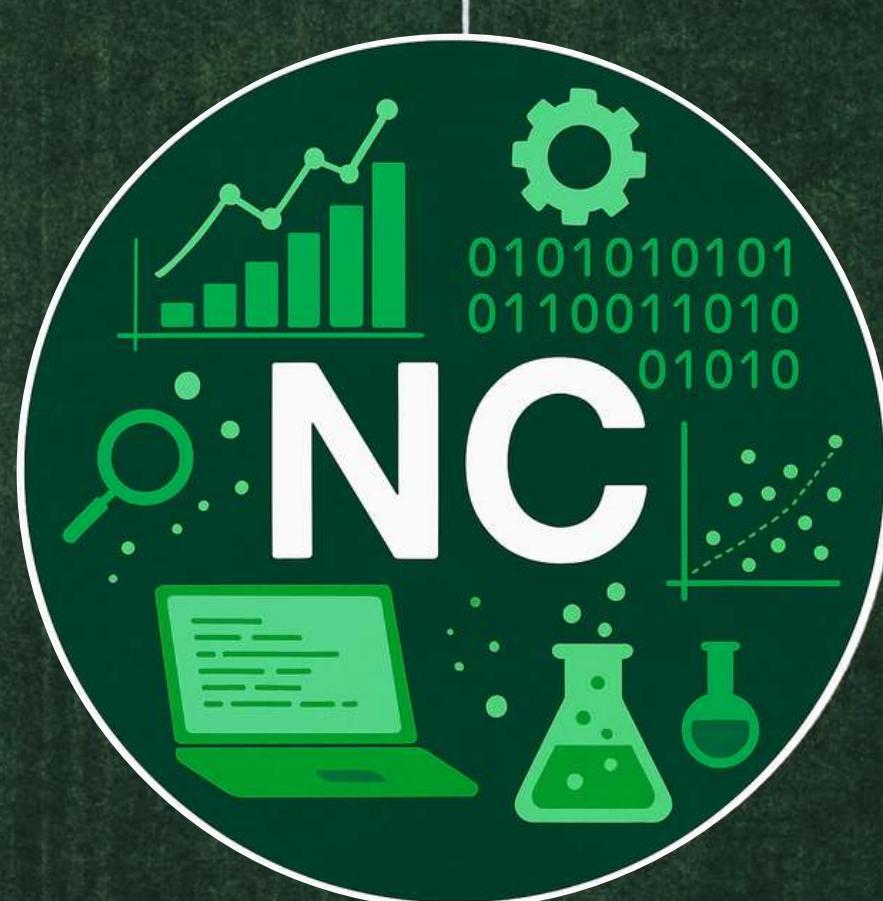


ANÁLISIS FUTBOLÍSTICO DATOS Y ESTADÍSTICAS

DATA SCIENCE II
CODERHOUSE



AUTOR:
NAHUEL CUELLO

El fútbol explicado con datos...



INTRODUCCIÓN AL PROYECTO

Análisis Futbolístico



El fútbol, más allá de la pasión, es un juego de patrones y decisiones. Este proyecto busca traducir esas dinámicas en datos concretos, analizando más de 1800 partidos y comparando comportamientos entre ligas locales y la Copa del Mundo 2022.



MOTIVACIÓN Y PROPOSITO



- Comprender cómo los datos reflejan el rendimiento real dentro del campo de juego.
- Identificar los factores que influyen directamente en la victoria o la derrota.
- Evidenciar que el fútbol mantiene patrones estadísticos consistentes en distintos contextos competitivos.
- Transformar información numérica en conclusiones tácticas y estratégicas aplicables.
- Promover una mirada analítica del fútbol que complemente la interpretación subjetiva del juego.





DESCRIPCIÓN DE LOS DATOS

DATASETS DEL PROYECTO:

- soccer-matches.csv
- statsbombpy

soccer-matches.csv:

El conjunto de datos contiene registros de 1845 partidos de fútbol correspondientes a diversas ligas de la temporada 2021/2022.

Librería statsbombpy (API):

Para la integración de datos se utilizó la librería statsbombpy, con el fin de obtener todos los datos oficiales del Mundial de Fútbol 2022.

METODOLOGÍA DE ANÁLISIS

1. Limpieza y unificación de datos (CSV + API).

2. Formulación de preguntas clave.

3. Exploración estadística y visualización (EDA).

4. Obtención de insights e interpretación.

ANÁLISIS

El análisis se desarrolló a partir de la integración de datos de distintas fuentes, la formulación de preguntas clave y la exploración visual de variables.



PASO A PASO DEL ANÁLISIS



Limpieza de Datos

- Eliminación de valores nulos y duplicados.
- Normalización de nombres de equipos y ligas.
- Creación de nuevas variables (porcentaje de pases, duelos ganados, etc.).
- Unificación de formato entre datasets locales y API.

Preguntas clave

- Se definieron interrogantes para guiar el enfoque del análisis.
- Se generaron gráficos específicos para responder cada pregunta.
- Se obtuvieron conclusiones claras basadas en evidencia estadística.

Exploración de Datos (EDA)

- Análisis descriptivo general del rendimiento.
- Visualización de patrones por variable (pases, goles, faltas, duelos).
- Estudio de correlaciones para identificar relaciones fuertes.

Obtención de insights

- Interpretación de los resultados visuales y estadísticos.
- Comparación entre ligas locales y datos del Mundial.
- Identificación de relaciones consistentes entre variables clave.
- Redacción de conclusiones claras.



PREGUNTAS CLAVE



¿Influye jugar de local en las probabilidades de ganar?



¿Qué formaciones son más efectivas para ganar partidos?



¿Mayor posesión, implica mayor probabilidad de ganar?



¿Cuánto más disparos al arco, mayor probabilidad de victoria?



¿La precisión de pases, está relacionada con el resultado?



¿Los regates, influyen en el resultado final?



¿Ganar los duelos aéreos, puede darte una victoria?



¿Qué impacto tienen las faltas en el desarrollo del partido?

USO DE LAS PREGUNTAS CLAVE



Las preguntas clave fueron el punto de partida del análisis. Sirvieron para definir el enfoque del estudio y orientar la exploración de los datos hacia aspectos realmente relevantes del juego.



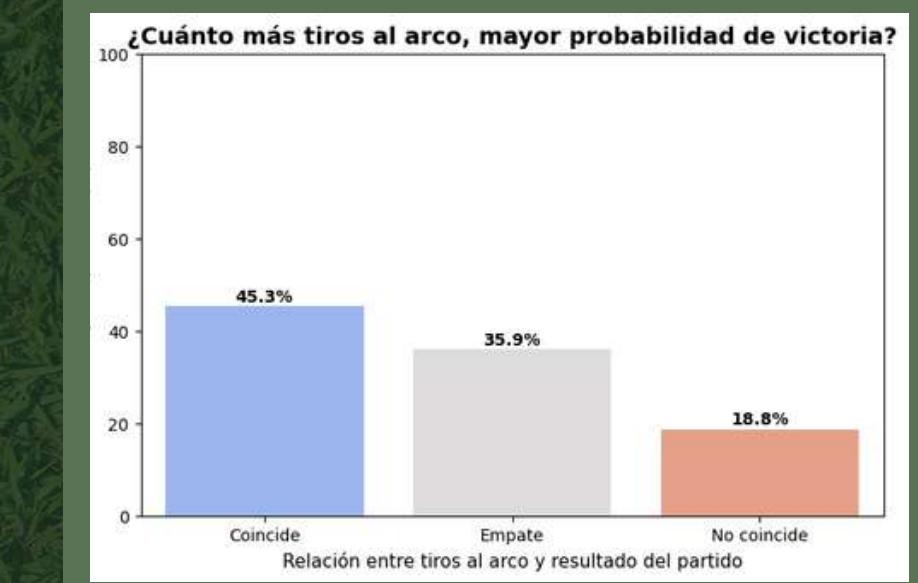
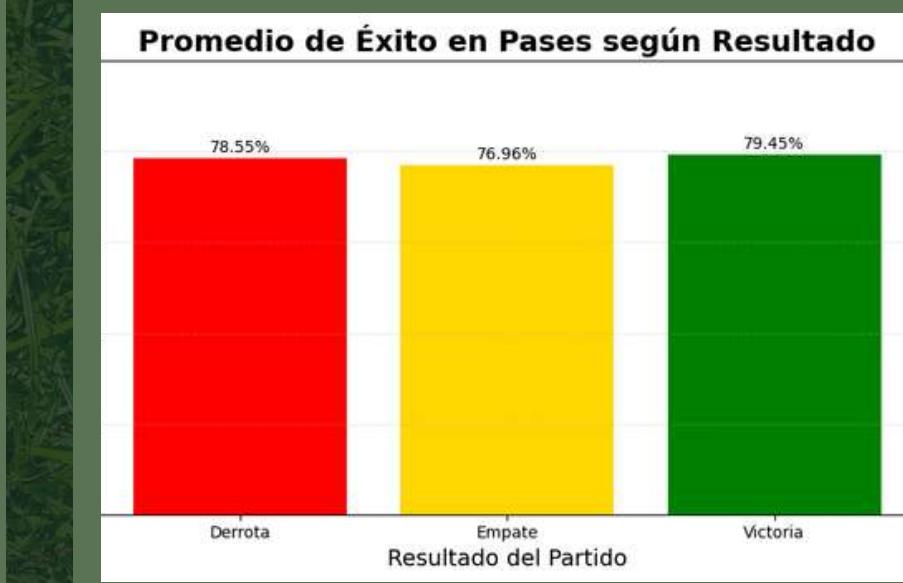
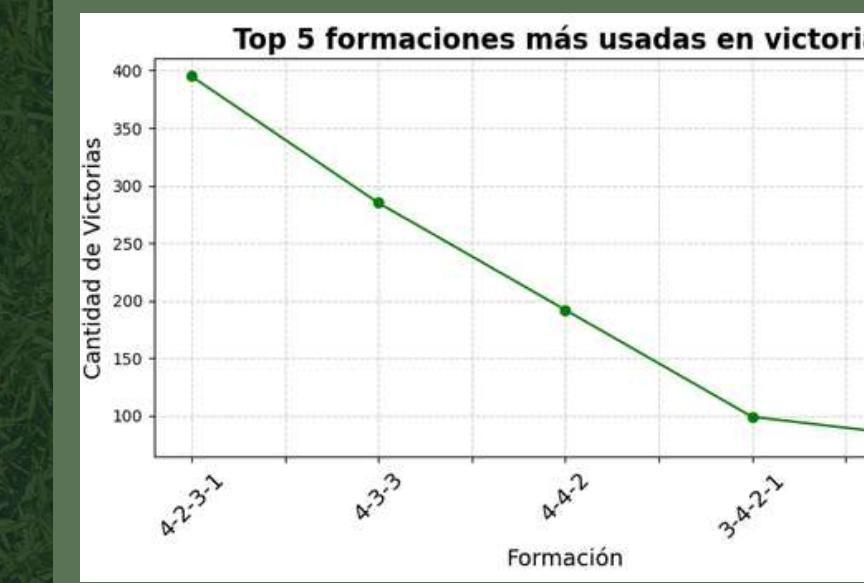
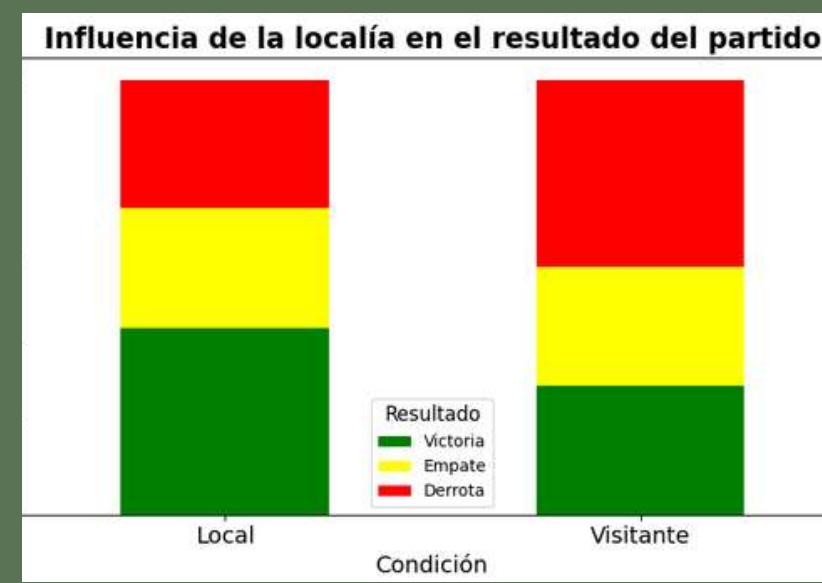
A través de ellas se establecieron los objetivos concretos que guiaron cada etapa del proceso: desde la limpieza y selección de variables hasta la interpretación de los resultados.



En lugar de analizar todos los datos de forma aislada, las preguntas permitieron concentrar el análisis en relaciones específicas, como el impacto de los pases, los duelos o las faltas en el resultado final. De este modo, cada visualización y cada métrica respondieron a una pregunta clara, transformando el análisis en insights útiles y aplicables al fútbol real.



RESPUESTAS GRÁFICAS A LAS PREGUNTAS CLAVE



Gráficos

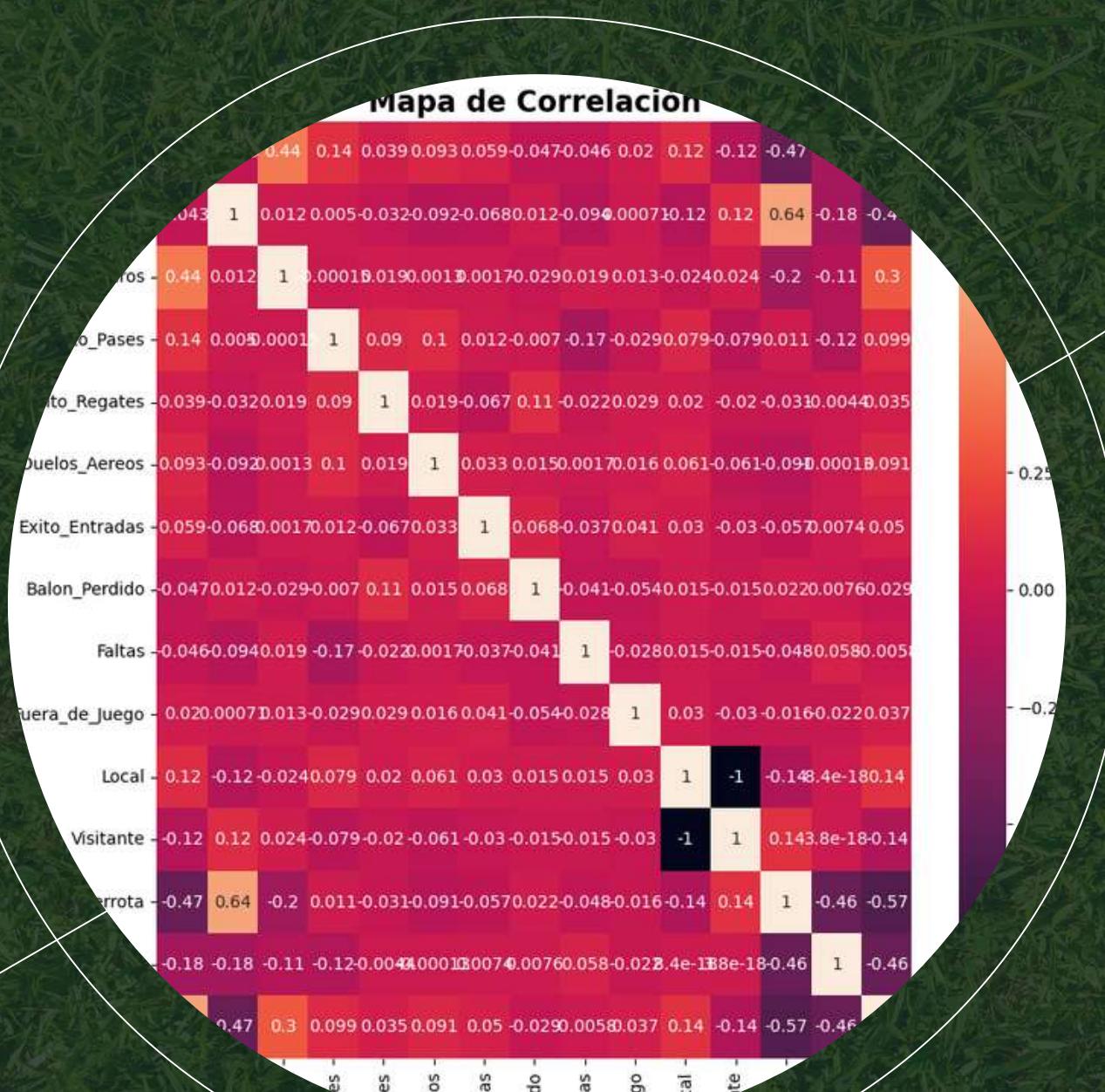


Utilizando las preguntas clave como punto de partida, se generó una serie de gráficos que permitieron visualizar de manera clara y concreta las respuestas a cada una de ellas. A partir de estas visualizaciones, fue posible obtener conclusiones precisas que ayudaron a responder las principales incógnitas del proyecto.

ANÁLISIS DE VARIABLES



Se realizó la codificación y correlación de variables para identificar relaciones significativas entre los distintos indicadores de rendimiento. Este proceso permitió transformar los datos en información interpretable y detectar patrones clave dentro del juego.



MACHINE LEARNING



Objetivo:

- ⚽ Predecir el resultado de los partidos (victoria, empate o derrota) a partir de variables estadísticas del juego.

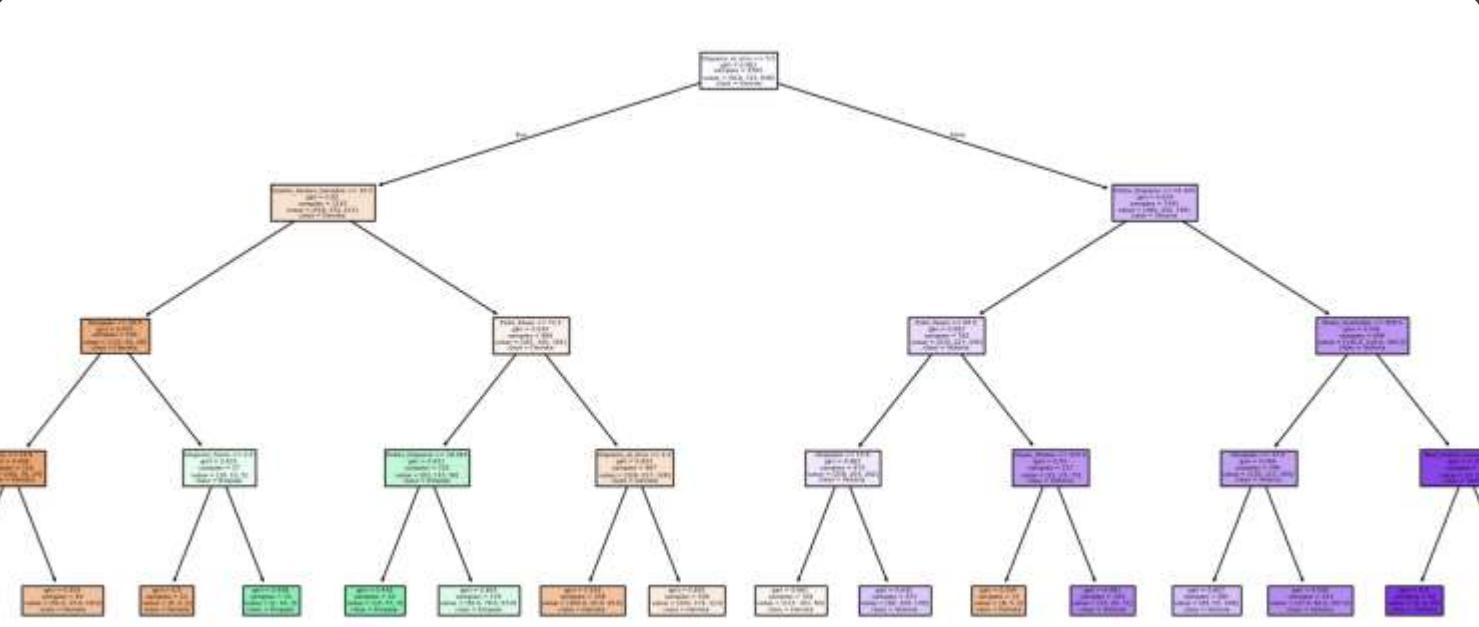
Procedimiento:

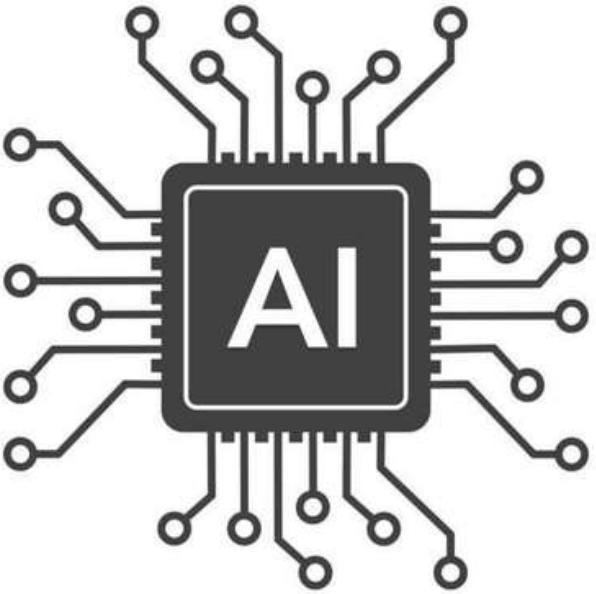
- ⚽ Se aplicaron modelos de machine learning usando datos históricos de los partidos, se entrenaron y evaluaron los modelos para medir su precisión.

Algoritmos:

- ⚽ Decision Tree Classifier.

- ⚽ Random Forest Classifier.





MACHINE LEARNING(API)



Objetivo:

- ⚽ Predecir el ganador del partido (Local, Empate o Visitante).

Procedimiento:

- ⚽ Entrenamos varios modelos de Machine Learning usando los datos del Mundial 2022 para predecir el ganador de cada partido.

👤 Cuello Nahuel.
👤 Data Science II.
👤 CoderHouse.



Algoritmos:

- ⚽ Decision Tree Classifier.
- ⚽ Random Forest Classifier.
- ⚽ XGBoost Classifier.

Hiperparámetros:

Elegimos el **Decision Tree** por su mejor equilibrio entre rendimiento y simplicidad. Finalmente lo optimizamos ajustando sus hiperparámetros para mejorar su desempeño.

CONCLUSIÓN FINAL

El análisis confirma que ganar no es cuestión de azar, sino de ejecutar bien ciertos aspectos del juego, minimizar errores y mantener el control del balón. A través del estudio de variables como la posesión, la precisión en los pases, los duelos ganados y la efectividad en jugadas clave, se evidencia que los equipos que dominan estos factores tienden a obtener mejores resultados. Esto refuerza la importancia de una estrategia bien planificada y del trabajo colectivo dentro del campo.



[Ver proyecto en GitHub](#)



GRACIAS
POR SU
REVISIÓN



Cuello Nahuel.



Data Science II.



CoderHouse.