

# Feedback Matemáticas y Estadística para la IA

Unai Garrido Ramírez

21 de enero de 2025

## Resumen

En este trabajo, se exponen de manera fundamentada, las diferentes conclusiones y resultados relacionados al ejercicio de Feedback de la asignatura Matemáticas y Estadística para la IA del Máster Universitario de IA impartido por la Universidad Alfonso X el Sabio. En este trabajo se encuentra un análisis de datos exhaustivo mediante las técnicas aprendidas a lo largo del transcurso de la asignatura. Todas estas técnicas están aplicadas sobre un conjunto de datos futbolísticos, que contiene información de todo tipo.

## 1. Exploración Inicial y Limpieza de Datos

Durante este primer proceso de Exploratory Data Analysis (EDA), se han cubierto diferentes tareas con objetivo de dar claridad, analizar superficialmente y entender el conjunto de datos objetivo. Entre ellas se encuentran el manejo de valores faltantes, transformaciones en las variables y el análisis de la consistencia, como se expone a continuación.

### 1.1. Identificación, manejo de valores faltantes y análisis de la estructura y consistencia de los datos.

Para comenzar con el EDA, primeramente se llevó a cabo la identificación de los datos y la estructura del dataset. Mediante el uso de funciones de la librería Pandas y Numpy se pudo identificar con que tipo de variables estamos trabajando.

Como se puede observar en el dataframe trabajamos con bastantes variables categóricas, aunque también hay numéricas y booleanas. Seguimos con la identificación de ciertas inconsistencias del dataframe como valores nulos o valores duplicados haciendo uso de diferentes comandos como `Results.isnull().sum()`, `Results.duplicated().sum()`, entre otros.

Obteniendo como resultado que no hay celdas nulas ni duplicadas. De forma preventiva se realizó una modificación de consistencia en el dataframe para que todos los datos siguieran el mismo formato.

Listing 1: Estandarización de nombres y formatos de fecha

```
Results['tournament'] = Results['tournament'].str.strip()
Results['country'] = Results['country'].str.title()
Results['city'] = Results['city'].str.title()
Results['home_team']= Results['home_team'].str.title()
Results['away_team']= Results['away_team'].str.title()
Results['date'] = pd.to_datetime(Results['date'])
```

Para llevar a cabo el manejo de datos faltantes se ha hecho uso de la imputación por moda así como la interpolación lineal para las variables numéricas. Quedando así perfectamente preparado para los siguientes pasos del EDA y posteriores análisis.

θ									
date	object	home_team	object	away_team	object	home_score	int64	away_score	int64
tournament	object	0	1872-11-30	Scotland	England	0	0	Friendly	Glasgow
city	object	1	1873-03-08	England	Scotland	4	2	Friendly	London
country	object	2	1874-03-07	Scotland	England	2	1	Friendly	Glasgow
neutral	bool	3	1875-03-06	England	Scotland	2	2	Friendly	London
		4	1876-03-04	Scotland	England	3	0	Friendly	Glasgow
								Scotland	False

((a)) Tipos de datos

((b)) Estructura del Dataframe

## 1.2. Transformaciones realizadas a las variables.

Después de las anteriores transformaciones mencionadas y llevadas a cabo para otorgar consistencia, la columna 'neutral' (True/False) se convirtió a tipo entero (1/0) usando astype(int).

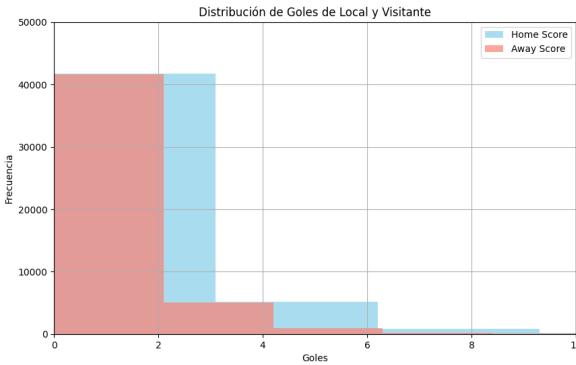
Así podemos utilizar esta variable en modelos que requieren entradas numéricas, como la matriz de correlación, entre otros algoritmos posteriormente utilizados.

Con el objetivo de brindar un conjunto de datos limpio, informativo y adecuado para la implementación de modelos predictivos consistentes, también se han definido nuevas variables, que aportan información relevante y tienen un formato más adecuado para la posterior implementación de los modelos.

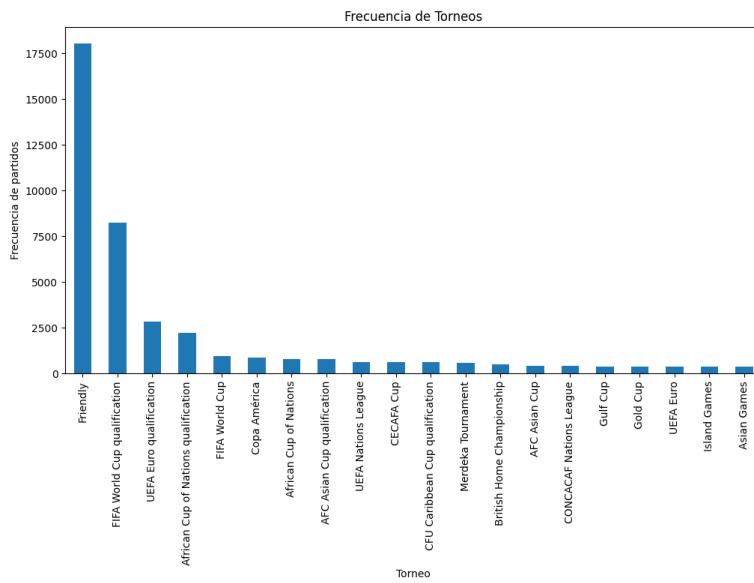
- **Diferencia de goles ('Goal diff')**: Esta variable captura la magnitud de la victoria o derrota, proporcionando información más allá del simple resultado (ganar, empatar, perder). Puede ser útil para analizar el rendimiento del equipo o para predecir resultados futuros.
- **Década del partido ('Decade')**: Permite agrupar y analizar las tendencias a lo largo del tiempo, observando cómo ha evolucionado el fútbol en diferentes décadas.
- **Resultado del partido ('Win')**: Simplifica el análisis del resultado, especialmente para modelos que se basen en algún tipo de clasificación.
- **Año del partido ('Year')**: Similar a nuestra columna Decade pero aportando más especificidad, útil en los casos en los que se quiera hacer un estudio de cierto año en específico.
- **Si el partido fue amistoso ('is friendly')**: Permite agrupar partidos en función de su tipo, ya que los partidos amistosos pueden tener características y dinámicas diferentes a los partidos de competición oficial.

### 1.3. Visualizaciones y representaciones de los datos.

Para concluir con el Análisis Exploratorio de Datos, se han buscado visualizaciones fáciles de interpretar y que representen los patrones iniciales de nuestro conjunto de datos. Tenemos las siguientes:



((a)) Número de goles por partido



((b)) Torneos

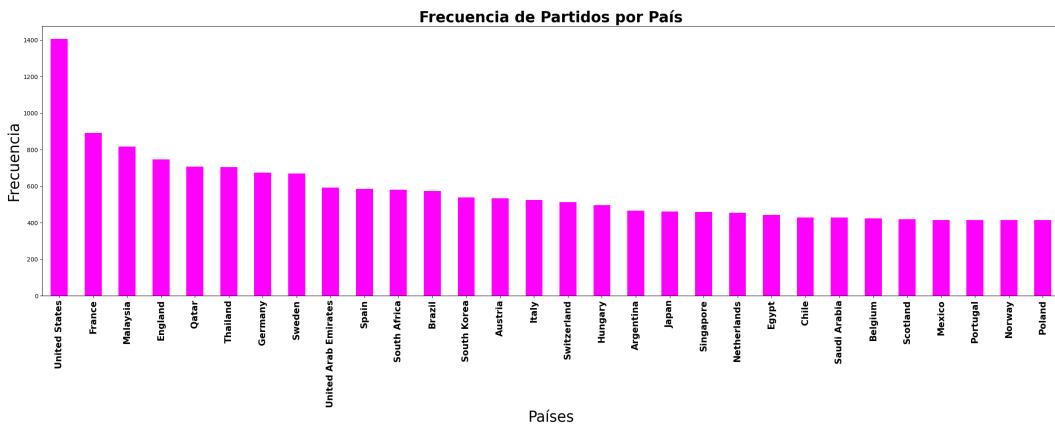


Figura 3: Frecuencia de partidos por País

## 2. Análisis Descriptivo

### 2.1. Estadísticas generales del conjunto de datos.

Para hallar las estadísticas de nuestro conjunto de datos usamos el comando `Describe` de pandas. Lógicamente este comando solo se puede ejecutar con variables numéricas, entonces suprimimos las columnas no numéricas y nos quedamos con (date, home score, away score, neutral, goladiff, decade y Win).

Aun así si analizamos detenidamente nuestras columnas numéricas algunas de ellas no tienen sentido incluirlas en el análisis ya que corresponden a fechas o son variables booleanas a las cuales no tiene sentido aplicar funciones estadísticas como la media o la mediana, contando con eso, nos queda la siguiente tabla de estadísticas descriptivas:

Estadística	home_score	away_score	Goal_diff
count	47917	47917	47917
mean	1.75787	1.18175	0.576121
std	1.77254	1.39957	2.41132
min	0	0	-21
25%	1	0	-1
50%	1	1	0
75%	2	2	2
max	31	21	31

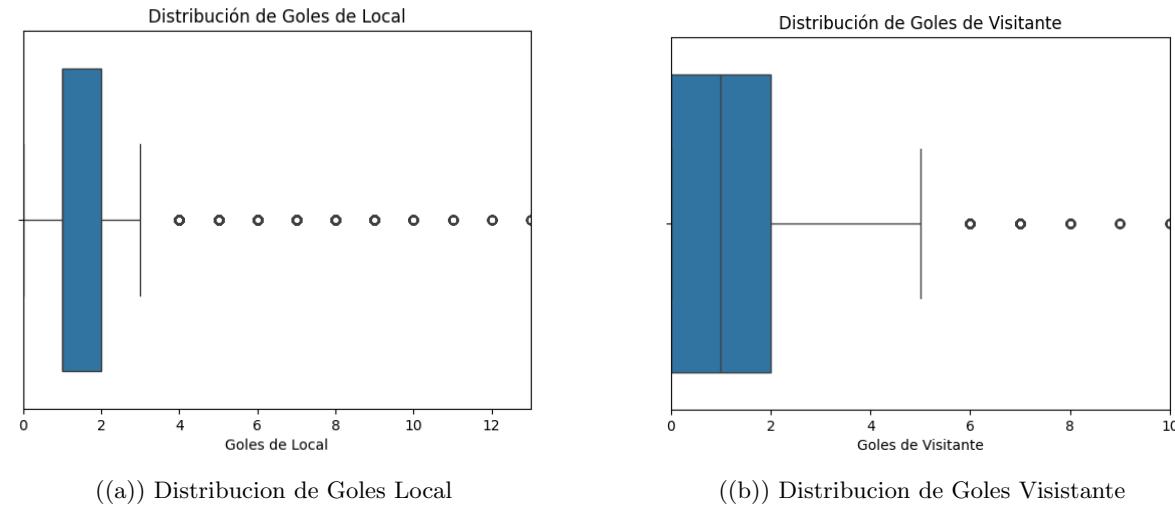
Figura 4: Funciones Estadísticas básicas

En cuanto a mis variables categóricas, tenemos varios resultados interesantes los cuales sacamos de la simple observación de nuestros datos sin analizar profundamente nuestras visualizaciones.

- Kuala Lumpur es la ciudad en la que más partidos se han jugado.
- Uruguay es el equipo que más ha jugado como visitante.
- Brazil es el equipo que más ha jugado como Local.
- EL torneo 'FIFA World Cup Qualification' es en el que más partidos se han jugado, excluyendo los partidos amistosos.

## 2.2. Visualizaciones relevantes e Identificación de patrones o tendencias.

El objetivo de este apartado del análisis descriptivo realizado es resumir y describir las características principales de nuestro conjunto de datos a través de diversas visualizaciones y su estudio. En este caso hacemos uso de diferentes tipos de visualizaciones como por ejemplo, histogramas, box-plots, heatmaps, gráficos de barras, entre otros que se comentarán para remarcar las características propias de nuestro conjunto de datos.



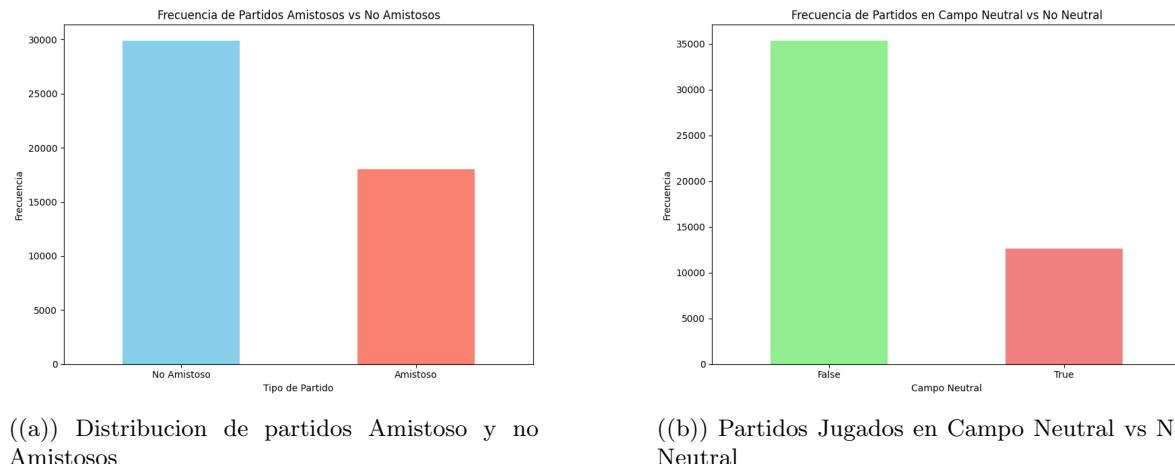
((a)) Distribucion de Goles Local

((b)) Distribucion de Goles Visistante

Si nos fijamos en nuestras representaciones de cajas, en el caso Local la mayoría de los partidos están comprendidos entre 1 y 2 goles.

Estudiando el caso visitante tenemos una representación parecida pero en este caso hay prácticamente la misma probabilidad de que haya un partido con 0 o 1 goles que uno con 1 o 2.

Ambas representaciones tienen una desviación estándar que llega hasta los 5 goles. También presentan outliers de prácticamente todos los valores, es decir, ha habido partidos en los que se han metido muchos más goles de lo habitual.



((a)) Distribucion de partidos Amistoso y no Amistosos

((b)) Partidos Jugados en Campo Neutral vs No Neutral

Podemos ver en estas distribuciones que la mayoría de partidos pertenecen a torneos oficiales y que se juegan normalmente en campos no neutrales.

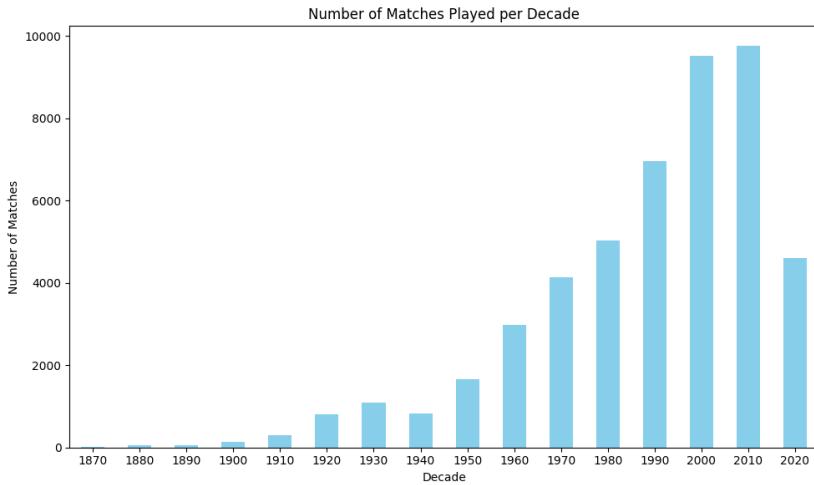


Figura 7: Número de partidos por año

Se observa un crecimiento exponencial, especialmente a partir de la segunda mitad del siglo XX. Esto refleja el desarrollo y la popularización del fútbol a nivel mundial, con la creación de más torneos y la participación de más países.

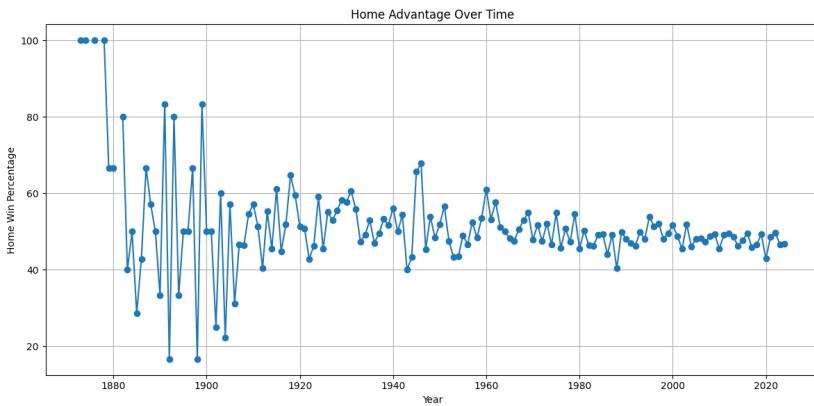


Figura 8: Porcentaje de Victorias en casa por año

La ventaja de localía parece haber disminuido a lo largo del tiempo. En las primeras décadas, el porcentaje de victorias locales era notablemente mayor, pero ha ido decreciendo gradualmente.

Podrían existir períodos donde la ventaja local fue más pronunciada y otros donde se redujo, posiblemente debido a factores como la profesionalización o cambios en las reglas del juego.

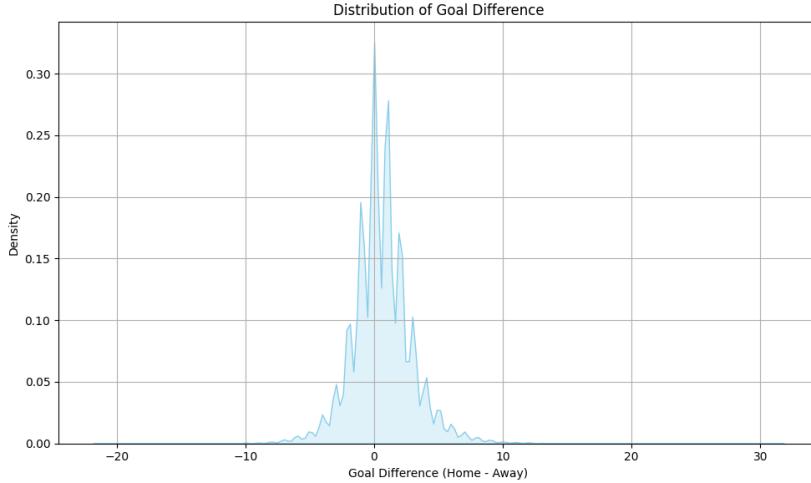


Figura 9: Densidad de diferencia de goles

La distribución es ligeramente asimétrica hacia la derecha, lo que indica que, en general, hay más partidos donde el equipo local gana por una diferencia de goles pequeña o empata.

Por la forma de la distribución vemos que la mayoría de partidos acaban sin diferencia de goles o con una diferencia muy ajustada de aproximadamente un gol.

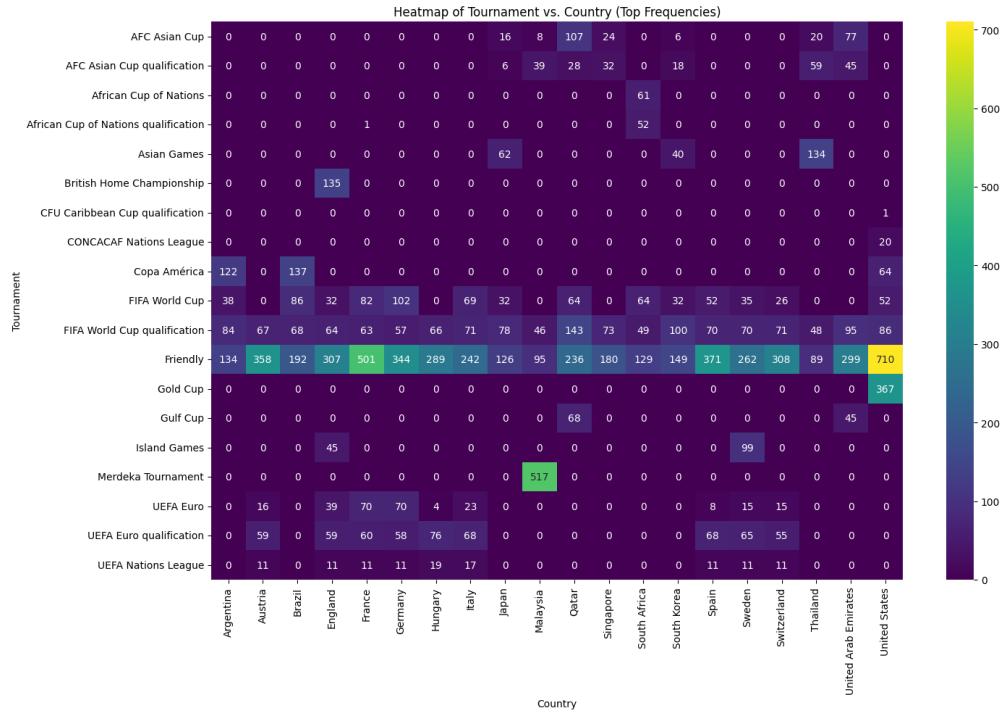


Figura 10: Heatmap del Número de partidos albergados por cada País y Torneo

Se puede observar que hay muchos torneos que son específicos de ciertas regiones o países, mientras que hay otros que son más globales.

Los Friendly Games y los clasificatorios de la Copa Mundial de la FIFA se juegan en una amplia gama de países, indicando su carácter internacional. En contraste, algunos torneos regionales o continentales, como la Copa América o la Copa Africana de Naciones, tienen una mayor concentración en países específicos de esas regiones.

USA es el país que alberga la mayor cantidad de partidos de un torneo o tipo, en este caso, amistosos.

### 3. Reducción de dimensionalidad. Análisis de componentes principales.

Nuestro objetivo con el PCA es analizar si las variables que tenemos pueden llegar a ser redundantes y hasta dónde se puede reducir la dimensionalidad de nuestro problema.

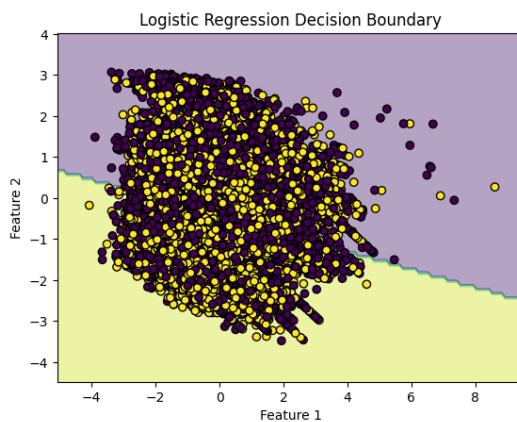
Para ello escogí las variables que deseo para mi modelo y las estandarizo, para luego crear variables derivadas que recojan su comportamiento. En nuestro caso, las variables son:  
tournament encoded, city encoded, is friendly encoded, home team avg goals scored, away team avg goals scored, Year.

No añadimos ninguna otra ya que o son redundantes, o nos aparece el famoso "Data Leakage", lo cual, para la construcción de un modelo predictivo, no es lo óptimo.

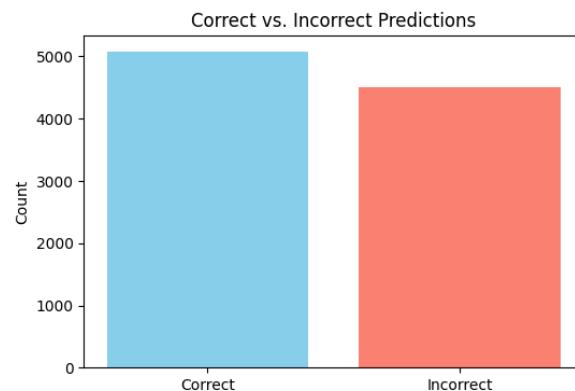
Dividimos el modelo en 2 componentes principales, las cuales nos dieron unas varianzas de [0,2532 y 0,180].

Construimos un modelo de regresión logística para predecir el resultado de un partido en función de las componentes principales, cuya accuracy nos otorga un 0,53.

Analizando el resultado y viendo que no se ha usado ninguna variable que influya en el resultado del partido directamente tenemos que nuestro modelo se comporta de la siguiente manera y clasifica de cada 100 partidos 53 de manera correcta.



((a)) Clasificación de las victorias con respecto a las componentes principales



((b)) Distribución de partidos acertados

### 4. Resultados de Análisis Bayesiano

Se ha utilizado, en este caso, el Teorema de Bayes para calcular las probabilidades condicionales, sujetas al resultado de partidos de fútbol en diferentes situaciones.

Este Teorema nos permite actualizar nuestras creencias sobre un evento (la probabilidad a priori) a medida que obtenemos nueva evidencia (los datos). La fórmula general del Teorema de Bayes es:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1)$$

En el análisis de nuestro caso de uso diferenciamos cada parte del teorema de las siguientes formas:

- **Probabilidad de que un equipo gane jugando como local:**

- Se definió el evento A como el equipo gana.
- Se definió el evento B como el equipo juega como local.

- Se utilizó la información del conjunto de datos para estimar las probabilidades a priori y la verosimilitud.
- Se aplicó el Teorema de Bayes para calcular la probabilidad posterior  $P(A|B)$ .

• **Probabilidad de empate en un campo neutral:**

- Se definió el evento A como el partido termina en empate.
- Se definió el evento B como el partido se juega en un campo neutral.
- Se utilizó la información del conjunto de datos para estimar las probabilidades a priori y la verosimilitud.
- Se aplicó el Teorema de Bayes para calcular la probabilidad posterior  $P(A|B)$ .

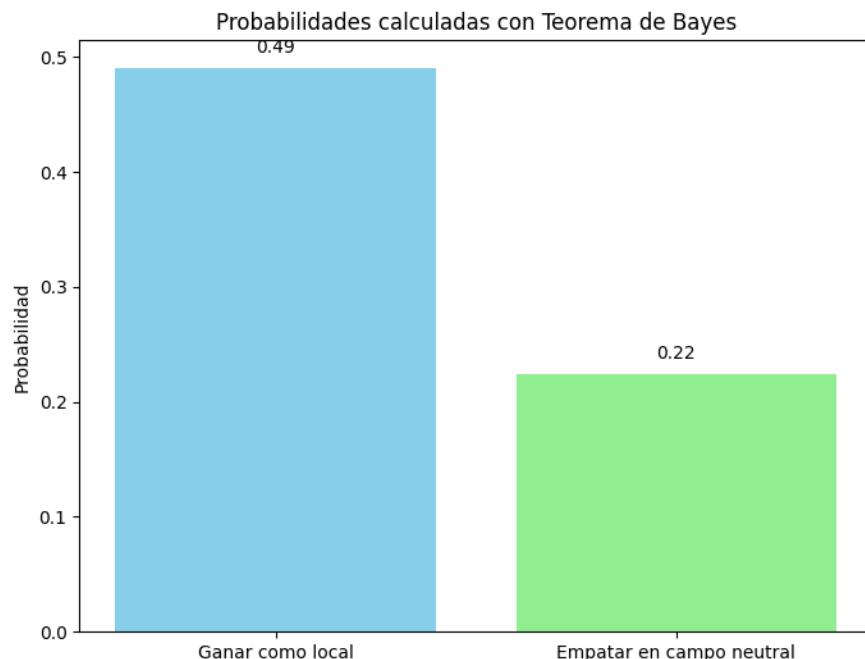


Figura 12: Probabilidades a posteriori aplicando el Teorema de bayes de mis dos supuestos

#### 4.1. Caso Aplicado: Probabilidad de que un partido sea emocionante (Muchos Goles)

Para este apartado extra hemos definido los eventos siguientes. **Evento A:** El partido tiene más de X goles (donde X es el umbral que define un partido "emocionante", en este caso 3), y **Evento B:** El partido se juega entre ciertos equipos (Equipo1 vs Equipo2) en este caso (Argentina vs Uruguay).

##### Probabilidad a priori $P(A)$

Se trata de la probabilidad de que exista un partido emocionante, es decir en este caso, más de 3 goles. La calculamos con el registro histórico de partidos con más de 3 goles.

##### Probabilidad a priori $P(B)$

Si B es "partido entre Equipo1 y Equipo2",  $P(B)$  sería la frecuencia de partidos entre esos equipos en el registro global.

##### La verosimilitud $P(B|A)$

Se trata en nuestro caso de estudio de la probabilidad de que sea un partido Argentina Uruguay debido a que este tuvo más de 3 goles. Es decir, la división del número de partidos Argentina Vs Uruguay con

más de 3 goles entre los partidos totales con más de 3 goles.

Tras la aplicación del Teorema de Bayes tenemos que la probabilidad a posteriori o  $P(A/B)$  en este caso es de 0.00112

## 5. Modelos Predictivos

A la hora de realizar un modelo predictivo hay que tener cuidado con el problema de Data Leakage considerando varias cosas previamente.

Si calculamos 'Goal diff' directamente como la diferencia entre los goles del equipo local (home score) y los goles del equipo visitante (away score), estaremos introduciendo esa fuga de datos, ya que queremos predecir el resultado de un partido y la diferencia de goles solo está disponible después de que el partido ha terminado.

Introduciendo esta variable la clasificación del modelo se regiría con una simple regla tal que:

- Si 'Goaldiff' es positivo, el equipo local gana.
- Si 'Goaldiff' es negativo, el equipo visitante gana.
- Si 'Goaldiff' es 0, es un empate.

Esta regla conllevaría a un modelo sobreajustado y con una accuracy irreal de valor 1.

Es por eso que se proponen dos tipos de modelo, un Naive Bayes Categórico el cual simplifica mucho el problema ya que supone la asunción de independencia ingenua que hace que las características sean independientes entre sí. El otro, sin embargo, es un modelo en las que las variables goaldiff, homescore y awayscore no están presentes ya que queremos evitar ese escape de datos. Primeramente visualizamos como de relacionadas están nuestras variables principales:

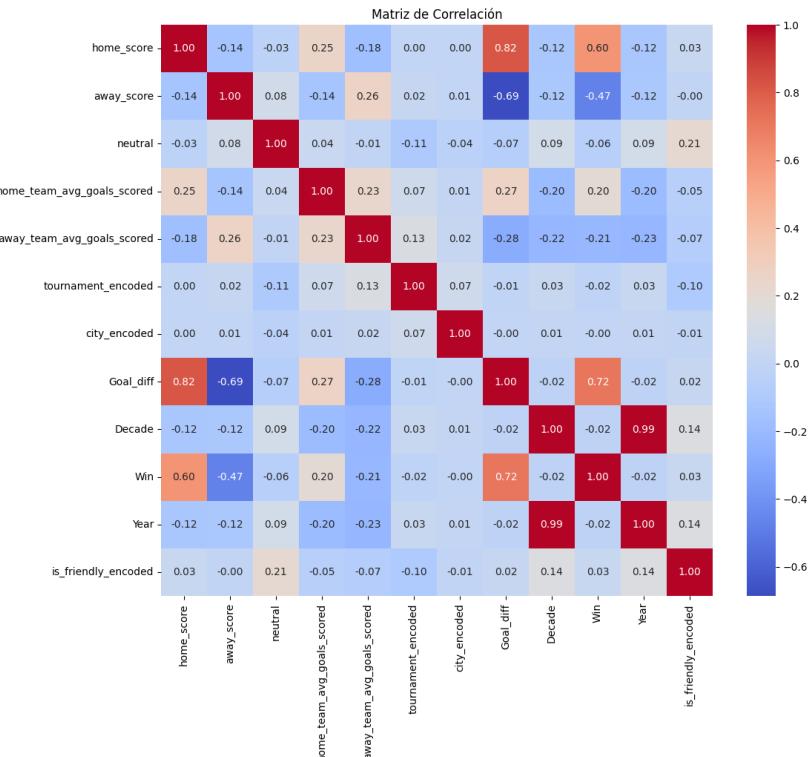


Figura 13: Matriz de Correlación

## 5.1. Algoritmos implementados (CategoricalNB, RandomForest y Regresión logística)

### 5.1.1. Clasificación Bayesiana

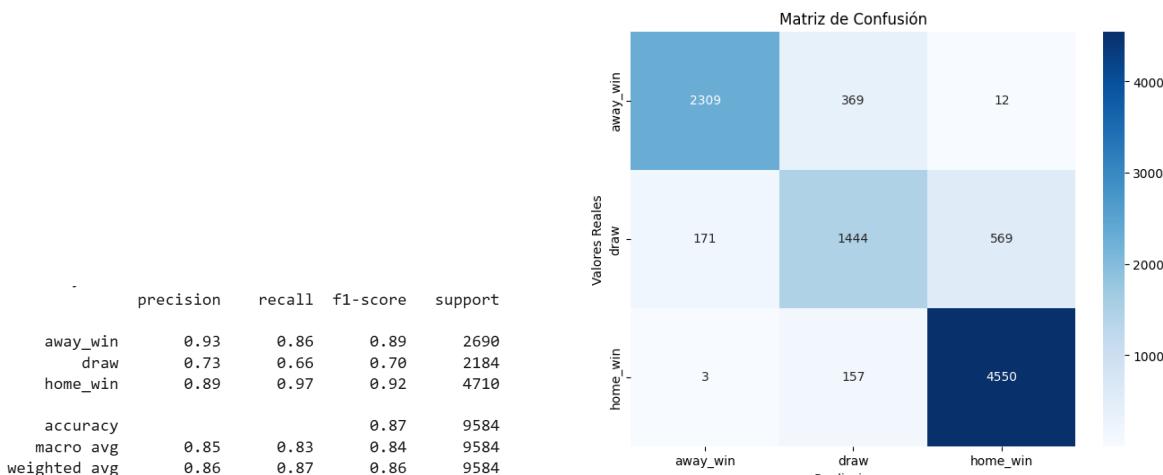
Al hacer uso de este algoritmo podemos incluir Goaldiff debido a la manera que tiene de tratar los datos.

En el caso del categoriclaNB, trata todos los datos como variables categóricas, discretizando las variables contínuas, además de esto hace uso de la suposición de independencia ingenua, la cual establece que la probabilidad de que se de una característica es independiente de las demás.

Este tratamiento se traduce en el hecho de que el modelo está tratando Goaldiff como una indicación de la probabilidad de observar diferentes resultados, sin asumir una relación causal directa.

El procedimiento de implementación de este algoritmo se compone de las siguientes partes:

- Selección de Características, y codificación One-Hot, en este caso fueron: 'tournamentencoded', 'cityencoded', 'isfriendlyencoded', 'homescorecat', 'awayscorecat'.
- División de datos en entrenamiento y prueba.
- Ajuste del modelo con fit().
- Realización de predicciones con predict() y métricas de evaluación.



((a)) Accuracy, F1 Score, precisión y recall.

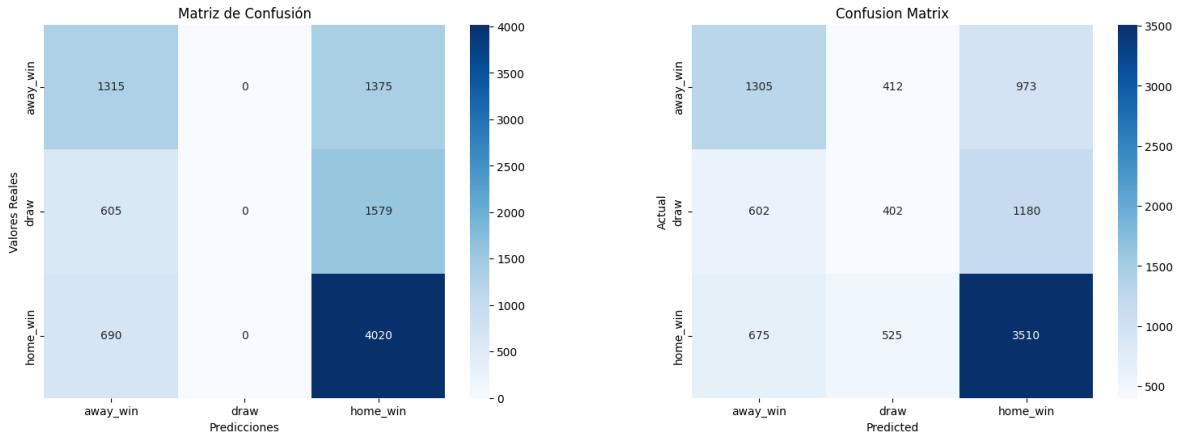
((b)) Matriz de Confusión Categorical Naive Bayes

### 5.1.2. Random Forest y Regresión logística

La clave en la elección de que variables incluir en cada modelo reside en como estos manejan los datos. En el caso de random forest o regresión logística los métodos buscan patrones en los datos para establecer relaciones directas entre las variables predictoras y las que queremos predecir, en nuestro caso el resultado de un partido.

Si incluimos 'Goaldiff' como una característica, estos modelos aprenderán a utilizarla como un predictor muy fuerte, ya que está directamente relacionada con el resultado del partido. Esto podría llevar a un sobreajuste (overfitting), donde el modelo se ajusta demasiado a los datos de entrenamiento y no generaliza bien a nuevos datos. Además, introduce el problema de data leakage, ya que 'Goaldiff' u otras variables como 'HomeScore' o 'AwayScore', no estarían disponibles en el momento de hacer predicciones sobre partidos futuros.

Para incluir más información en los modelos se ha utilizado dos nuevas variables derivadas llamadas 'home team win rate' y 'away team win rate'.



((a)) Matriz de Confusión Regresión Logística

((b)) Matriz de Confusión Random Forest

## 5.2. Comparación de métodos y conclusión

Las diferencias en los rendimientos de los 3 métodos residen en diferentes factores, como son:

- Complejidad de las relaciones entre variables:** Los datos de fútbol a menudo contienen relaciones complejas y no lineales entre las variables. Random Forest es capaz de modelar estas relaciones de manera más efectiva, lo que explica su rendimiento.
- Independencia de características:** Naive Bayes Categórico asume que las características son independientes, lo cual no siempre es cierto en el fútbol. Esto puede afectar su precisión en comparación con modelos que no hacen esta asunción, como Random Forest y Regresión Logística.
- Escalado de características:** Regresión Logística se beneficia del escalado de características, mientras que Random Forest es menos sensible a este preprocesamiento. Naive Bayes Categórico no requiere escalado de características, ya que trabaja con variables categóricas.

Como conclusión obtenemos que al tener un problema tan complejo, si usamos métodos que modelen las relaciones entre variables de manera más efectiva su precisión bajará. En caso de utilizar modelos más simples podremos conseguir un mayor rendimiento pero a cambio de una simplificación del problema.

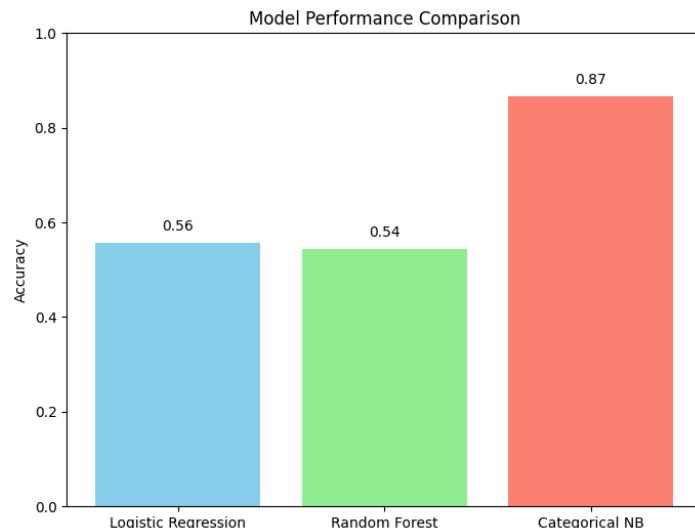


Figura 16: Comparación entre métodos

## 6. Grafos

Al visualizar los grafos debemos tener en cuenta varios factores, entre los que destacan la claridad de la representación y la capacidad de diferenciar tendencias, es por eso que en esta sección se abordan dos tipos distintos de enfoque, cada uno con sus ventajas y desventajas.

### 6.1. Grafo General

Cuando visualizamos un grafo para todo nuestro conjunto de datos, el objetivo principal es intentar identificar ciertos patrones y tendencias en cuanto a los equipos en general y los partidos que han jugado. Más abajo se enumeran las ventajas y algunas de las desventajas de este modelo y se comentarán que conclusiones podemos sacar de sus métricas y visualización.

- **Visión General:** Permite observar las relaciones entre todos los equipos y torneos a la vez, brindando una perspectiva global de las interacciones.
  - **Identificación de patrones generales:** Facilita la detección de patrones o tendencias generales en el conjunto de datos, como la existencia de grupos de equipos con un alto nivel de interacción.
  - **Comparación global:** Posibilita la comparación del rendimiento o las características de los equipos en todos los torneos de forma simultánea.

En cuanto a las desventajas de este enfoque podemos enumerar varias, como la complejidad de la visualización, la superposición de información y la no posibilidad de identificación de patrones específicos.

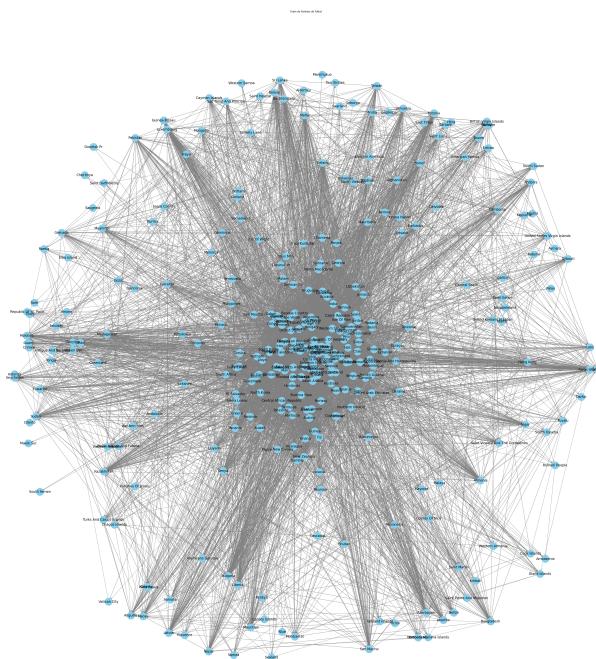


Figura 17: Grafo General de todos los enfrentamientos

Las métricas de nuestro grafo son las siguientes:  
Número de nodos: 336 ; Número de aristas: 11951 ; Grado promedio: 71.13690476190476 Densidad: 0.10617448471926084 ; Clustering coefficient promedio: 0.5421983046742967.

De estos valores podemos destacar varias conclusiones, como por ejemplo que es una red muy dispersa ya que tenemos un gran número de conexiones pero la baja densidad (0.11), nos indica que muchos equipos no han jugado entre sí.

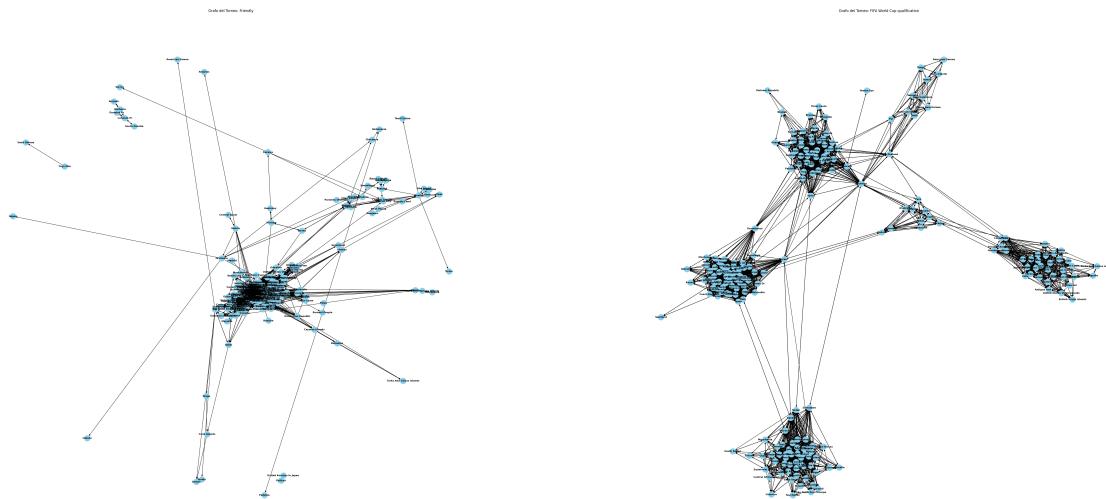
El alto clustering coefficient promedio (0.54) revela la existencia de grupos o clusters dentro de la red. Estos clusters representan ligas o regiones geográficas donde los equipos juegan entre sí con mayor frecuencia.

Este clustering coefficient nos indica que podríamos hacer un análisis más exhaustivo de ciertas regiones o torneos, como se realiza en el siguiente punto. Al observar el grado promedio podemos ver que cada equipo ha jugado en términos generales con 71 equipos diferentes.

## 6.2. Grafos Alternativos Por Torneo

- **Enfoque específico:** Permite centrarse en un torneo en particular, facilitando el análisis de las relaciones entre los equipos que participan en él.
- **Mayor Claridad:** Al visualizar solo la información de un torneo, se reduce la complejidad del grafo, mejorando la claridad y la interpretación.
- **Identificación de patrones específicos:** Facilita la detección de patrones o tendencias específicas de un torneo, como la existencia de equipos dominantes o la formación de grupos dentro del torneo.

Al observar las diferentes ventajas de este tipo de representación, parece ser en un principio una mejor manera de sacar conclusiones de nuestro conjunto de datos, pero este tipo de aproximación también tiene desventajas como pueden ser la pérdida de visión global, el mayor tiempo de computación debido a la generación por separado de los grafos y la mayor dificultad de comparación entre torneos.



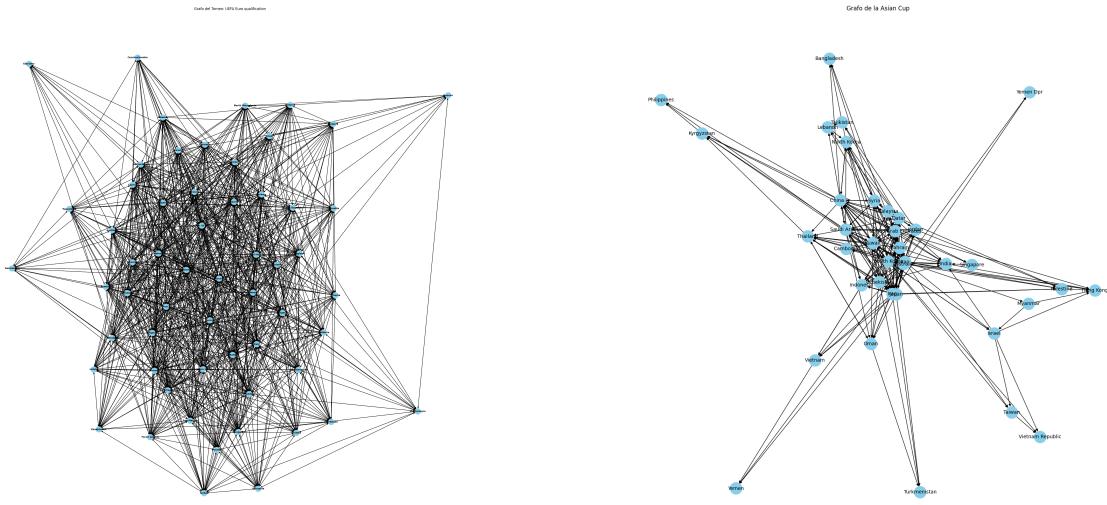
((a)) Grafo Friendly Games

((b)) Grafo Clasificatorio de la copa del mundo

Se puede observar fácilmente, en el caso de la world cup qualification, existen varios grupos que se dividen por continentes o regiones para luego enfrentarse entre si. Este tipo de comportamiento era indistinguible en el grafo general. En cuanto a los friendly games, vemos que hay un conjunto de equipos central que se enfrentan entre ellos con bastante frecuencia, sin embargo hay clusters de menor tamaño e independientes que solo han llevado a cabo enfrentamientos independientes.

Si nos fijamos en las métricas y ordenamos los equipos por importancia tenemos:

- Friendly Games: 'Saudi Arabia', 'China Pr', 'South Korea', 'Egypt', 'Tunisia'. Estos equipos tienen mayor grado de centralidad, es decir son los que, con más frecuencia participan en partidos.
- WC Qualification: 'Turkey', 'Finland', 'Sweden', 'Republic Of Ireland', 'Australia'



((a)) Grafo Copa de Europa

((b)) Grafo de la Copa Asiática

■ **Eurocopa (UEFA European Championship):**

- **Centralidad de Grado:** Si observamos, equipos como Alemania, España, Francia, Italia o Inglaterra tienen una alta centralidad de grado, esto confirma su rol como selecciones importantes en la historia de la Eurocopa, ya que han jugado más partidos en el torneo.
- **Centralidad de Intermediación:** Equipos con alta centralidad de intermediación podrían ser, por ejemplo, aquellos que se encuentran en zonas geográficas intermedias de Europa o que han participado en eliminatorias contra equipos de diferentes regiones. Esto indica su rol de conectores.<sup>en</sup> la red de la Eurocopa.
- **Centralidad de Vector Propio:** Equipos como los mencionados anteriormente (Alemania, España, etc.) tienen también alta centralidad de vector propio, esto refuerza su influencia en el torneo, ya que están conectados a otros equipos importantes. Podrían ser considerados como los "núcleos" de la red de la Eurocopa.
- **Componentes conectados:** Existen bastantes componentes conectados. Esto es un indicador de la poca cohesión y la integración de las selecciones europeas en el torneo.

■ **Copa Asiática (AFC Asian Cup):**

- **Centralidad de Grado:** Si observas que equipos como Japón, Corea del Sur, Australia, Irán o Arabia Saudita tienen alta centralidad de grado, esto confirma su participación frecuente y su importancia en la Copa Asiática.
- **Centralidad de Intermediación:** En la Copa Asiática, los equipos con alta centralidad de intermediación podrían ser aquellos que conectan diferentes regiones geográficas dentro de Asia. Por ejemplo, equipos de Asia Central que han jugado contra equipos del este, oeste y sur de Asia. Esto refleja la diversidad geográfica del torneo.
- **Componentes conectados:** Observamos más de un componente conectado en la Copa Asiática, esto podría reflejar la existencia de subgrupos o clusters de equipos que han interactuado más entre sí. Esto podría deberse a factores geográficos, políticos o de desarrollo futbolístico en las diferentes regiones de Asia.

La Eurocopa parece tener una red de equipos más cohesionada e integrada, con un único componente conectado y una mayor concentración de la influencia en torno a las selecciones tradicionalmente fuertes.

La Copa Asiática podría presentar una red más fragmentada, con la posibilidad de múltiples componentes conectados y una influencia más distribuida entre diferentes equipos de distintas regiones de Asia. Esto refleja la mayor diversidad geográfica y futbolística del continente asiático.

Como conclusión podemos decir que la representación de los grafos por separado proporciona un enfoque más claro, contextual y relevante para el análisis de datos de fútbol. Permite identificar patrones específicos de cada torneo, realizar comparaciones entre torneos y simplificar el proceso de análisis. Aunque un grafo conjunto podría mostrar todas las relaciones a la vez, la complejidad resultante podría dificultar la obtención de insights valiosos.