



Universidad del Desarrollo
Facultad de Ingeniería

DINÁMICAS DE BALOMPIÉ EN SUDAMÉRICA; UN ANÁLISIS TÁCTICO Y
TÉCNICO MEDIANTE DATA SCIENCE
Predicción, Clasificación y Clustering

POR: ALEJANDRO EDUARDO HENRÍQUEZ ARDILES

GABRIEL ROBERTO OCARANZA HENRÍQUEZ

Capstone presentado a la Facultad de Ingeniería de la Universidad del
Desarrollo para optar al grado académico de Magíster en Data Science

PROFESOR GUÍA:

Dra. LORETO BRAVO

Noviembre 2023

SANTIAGO

TABLA DE CONTENIDO

RESUMEN.....	1
1. INTRODUCCIÓN	1
2. TRABAJO RELACIONADO	3
3. HIPÓTESIS Y OBJETIVOS.....	7
3.1. HIPÓTESIS.....	7
3.2. OBJETIVO GENERAL	8
3.3. OBJETIVOS ESPECÍFICOS.....	8
4. DATOS Y METODOLOGÍA.....	10
4.1. DATOS	10
4.2.1 <i>Limpieza y Preparación de Datos</i>	14
4.2.2 <i>Análisis Exploratorio de Datos</i>	15
4.2. METODOLOGÍA.....	34
4.2.1 <i>Support Vector Machines (SVM)</i>	35
4.2.2 <i>Regresión Logística</i>	37
4.2.3 <i>Random Forest</i>	38
4.2.4 <i>PCA (Análisis de Componentes Principales)</i>	39
4.2.5 <i>Distancia Euclidiana</i>	41
4.2.6 <i>Similitud del Coseno</i>	42
4.2.7 <i>Evaluación de Modelos</i>	43
5. RESULTADOS	44
5.1. MÉTRICAS DE LOS MODELOS	44
5.2. ANÁLISIS DE LOS PESOS DE LAS VARIABLES POR PAÍS.....	48
<i>Argentina</i>	48

<i>Bolivia</i>	50
<i>Brasil</i>	53
<i>Chile</i>	55
<i>Colombia</i>	58
<i>Ecuador</i>	60
<i>Paraguay</i>	62
<i>Peru</i>	65
<i>Uruguay</i>	67
<i>Venezuela</i>	69
5.3. SIMILARIDAD DE LIGAS	74
5.4. CLUSTERIZACION POR EQUIPOS	81
6. CONCLUSIONES	100
7. TRABAJOS FUTURO	102
8. LIMITACIONES	103
BIBLIOGRAFÍA	105

Resumen

El estudio sobre equipos de CONMEBOL, utilizando datos de la plataforma Wyscout, incluyó cuatro equipos populares por liga. Se descubrió una alta similaridad en estilos de juego entre ligas, lo que sugirió la necesidad de reorganizar la agrupación de equipos basada en tácticas más que en nacionalidad. Se identificaron cuatro clusters distintos, reflejando desde estilos de posesión hasta enfoques más directos. Sin embargo, la versatilidad táctica de los equipos no correlacionó directamente con el éxito en rankings como el de la FIFA. El uso del Modelo de Regresión Logística permitió una interpretación profunda de los estilos de juego, evidenciando la compleja red de tácticas que trascienden fronteras nacionales y resaltando la diversidad estratégica en el fútbol sudamericano.

1. Introducción

El fútbol sudamericano, reconocido por su fervor y dinamismo, ha sido históricamente un ámbito de estudio continuo debido a la rica diversidad táctica presente en sus ligas. Las naciones que integran la Conmebol, con sus distintos enfoques y estilos de juego, ofrecen un escenario propicio para investigar posibles variaciones en la dinámica futbolística de cada país.

El propósito primordial de esta investigación es explorar y determinar las diferencias tácticas y técnicas en el juego de las distintas ligas de fútbol que conforman la Conmebol. Las implicancias de identificar dichas diferencias trascienden el ámbito académico, pues

un entendimiento profundo de estos aspectos puede ser determinante para la adaptación y optimización de estrategias en el contexto competitivo.

Con la incursión de técnicas avanzadas de data science en el ámbito deportivo, se ha abierto una ventana de oportunidad para analizar el fútbol desde una perspectiva cuantitativa y cualitativa sin precedentes. Este estudio, a través de la aplicación de métodos analíticos avanzados, aspira a desarrollar un modelo que asigne valores específicos a las distintas variables del juego, permitiendo una comparación objetiva entre las ligas.

Mediante esta investigación, se delinearán la metodología adoptada, los instrumentos de recolección de datos empleados y el proceso analítico seguido. Finalmente, se presentarán los hallazgos que, se espera, ofrezcan una comprensión detallada y sistemática de la naturaleza del fútbol en el ámbito de la Conmebol.

2. Trabajo Relacionado

En los últimos años, hemos sido testigos de un progreso significativo en la recopilación y análisis de datos en partidos de fútbol. Esta evolución ha desencadenado un creciente interés por parte de la comunidad científica y la industria deportiva en la evaluación avanzada del rendimiento de los jugadores de fútbol. Un claro ejemplo de este avance es el estudio "PlayeRank", liderado por Luca Pappalardo y su equipo. PlayeRank se destaca como un marco de trabajo innovador, diseñado para realizar una evaluación multidimensional y consciente del rol de los jugadores, aprovechando el análisis de grandes volúmenes de datos generados durante los partidos, conocidos como "soccer-logs". Este enfoque representa un cambio paradigmático en cómo se mide y comprende el rendimiento en el fútbol, brindando una perspectiva más rica y detallada que los métodos tradicionales. El marco de PlayeRank se desarrolló en colaboración con Wyscout, una empresa líder en el ámbito del scouting en el fútbol. Utiliza un enfoque de aprendizaje automático para evaluar el desempeño de los jugadores, basado en un conjunto de datos masivo que incluye millones de eventos de partidos que abarcan cuatro temporadas de 18 competiciones de fútbol destacadas. PlayeRank implementa un sistema de tres fases: aprendizaje, calificación y clasificación. Durante la fase de aprendizaje, se determinan los pesos de las características del rendimiento de los jugadores, utilizando un enfoque supervisado que vincula las características del equipo con el resultado de los partidos. En la fase de calificación, se evalúa el rendimiento de cada jugador en un partido mediante el producto escalar entre los pesos de las características y los valores de estas características

observadas en ese partido. Finalmente, en la fase de clasificación, se construyen clasificaciones basadas en roles para los jugadores, considerando sus calificaciones de rendimiento y los roles asignados durante las fases anteriores. Un aspecto innovador de PlayeRank es su enfoque en la detección de roles, identificando los roles de los jugadores en el campo mediante algoritmos como k-means y considerando la posición promedio de los jugadores durante los partidos. Esto permite un análisis más detallado y contextualizado del rendimiento, superando las limitaciones de los enfoques unidimensionales anteriores que no tenían en cuenta la especificidad de los roles de los jugadores en el campo. En términos de resultados experimentales, PlayeRank ha demostrado ser robusto y efectivo, superando a los algoritmos existentes en términos de acuerdo con las clasificaciones de jugadores proporcionadas por expertos en fútbol profesional. Esta investigación destaca la creciente importancia de los análisis basados en datos en el deporte y demuestra cómo las tecnologías avanzadas pueden proporcionar insights valiosos sobre el rendimiento de los jugadores más allá de las métricas tradicionales. El enfoque multidimensional de PlayeRank y su capacidad para integrar y analizar una amplia gama de datos de rendimiento lo distingue como un desarrollo significativo en la analítica del fútbol. Este tipo de herramientas ofrece nuevas oportunidades para la evaluación y selección de talentos, así como para la mejora del rendimiento de los jugadores y equipos en el fútbol moderno.

Otro análisis crucial que se revisó abarcó el estudio realizado sobre la Liga Belga, el cual se destaca por su significativa contribución en la implementación de técnicas de aprendizaje automático y análisis de datos avanzados. Este estudio representa un hito en

la predicción y comprensión de los resultados de los partidos de fútbol. El enfoque principal de este estudio fue la identificación de variables críticas con un alto poder predictivo sobre los desenlaces de los partidos. Se utilizó un modelo de aprendizaje automático que integró una extensa gama de variables, sumando hasta 100 diferentes factores. La recopilación de estos datos se extendió a lo largo de tres temporadas de la Liga Belga, apoyándose en el sofisticado sistema de seguimiento óptico SportVU. La metodología del estudio resaltó por la inclusión de técnicas avanzadas como el Factor de Inflación de Varianza y BorutaShap para la efectiva selección de características, además del uso del Boosting Extremo de Gradientes en la predicción de los resultados de los partidos. Este estudio logró una notable precisión en sus predicciones, clasificando acertadamente la mayoría de los partidos analizados. Tal hallazgo enfatiza el inmenso potencial de incluir un espectro amplio de variables en la predicción y evaluación de resultados, proveyendo así herramientas valiosas para el análisis táctico y estratégico en el fútbol. Por lo tanto, este estudio se erige como un ejemplo destacado de cómo el análisis avanzado de datos está revolucionando nuestra percepción y entendimiento del fútbol, ofreciendo perspectivas más profundas sobre el deporte y abriendo nuevas posibilidades para el análisis del rendimiento y la toma de decisiones estratégicas en este ámbito.

Otro estudio analizado es "Predicting Match Outcomes in Association Football Using Team Ratings and Player Ratings" realizado por Halvard Arntzen y Lars Magnus Hvattum, que se enfoca en comparar el rendimiento de las calificaciones de equipos y las calificaciones individuales de los jugadores en la predicción de resultados de partidos de

fútbol. El sistema de clasificación Elo, bien conocido, se utiliza para calcular las calificaciones de los equipos, mientras que una variante de las calificaciones plus-minus se usa para calificar a los jugadores individuales.

El estudio hace la Comparación de Modelos Estadísticos para Pronósticos de Partidos, donde se comparan dos tipos de modelos estadísticos para la previsión de resultados de partidos, un modelo de regresión logística ordenada (OLR), conocido por generar pronósticos pre-partido razonables, y un nuevo modelo de riesgo competitivo basado en análisis de supervivencia, que también puede usarse para predicciones durante el juego.

Uso de Calificaciones de Jugadores y Equipos para Predecir Resultados: Se compara el uso de calificaciones individuales de jugadores y calificaciones directas de equipos como base para pronosticar resultados de partidos.

Los datos utilizados en los experimentos reportados incluyen partidos de las cuatro divisiones del sistema de ligas inglés y la Copa de la Liga inglesa, abarcando diez temporadas desde 2009/2010 hasta 2018/2019, resultando en un total de 21,129 partidos.

En el modelo OLR, se consideraron dos covariables: la diferencia en las calificaciones Elo de los equipos (x_{Elo}) y la diferencia promedio en las calificaciones individuales de los jugadores de los equipos (x_{PM}). El estudio probó tres combinaciones de covariables usando los modelos OLR y de riesgo competitivo: utilizando solo las calificaciones Elo, solo las calificaciones plus-minus, o ambas simultáneamente. Para comparar las seis variantes resultantes, se utilizaron la pérdida de información (LI) y la pérdida cuadrática (L2) como métricas de evaluación. Este informe es un ejemplo importante de cómo se pueden aplicar modelos estadísticos avanzados y análisis de datos en el contexto del fútbol

para mejorar la precisión de las predicciones de resultados de partidos, utilizando tanto calificaciones de equipos como evaluaciones detalladas del desempeño individual de los jugadores.

3. Hipótesis y Objetivos

3.1. Hipótesis

La hipótesis central de este estudio se basa en la premisa de que existen variaciones distintivas en las características de juego entre las diez ligas que componen la Confederación Sudamericana de Fútbol (Conmebol). Se espera que estas diferencias se reflejen en una variedad de aspectos fundamentales del juego, como la cantidad promedio de goles por partido, el control del balón, la precisión en los pases, la eficacia defensiva, entre otras métricas clave que necesitan ser claramente definidas.

En caso de que no se identifiquen diferencias significativas entre las ligas, el estudio se orientará hacia la búsqueda de nuevas agrupaciones entre los equipos. El objetivo será definir patrones de juego similares entre los equipos, lo que podría revelar tendencias y estilos de juego compartidos más allá de las divisiones geográficas y ligas individuales. Este enfoque permitirá una comprensión más profunda de las estrategias y tácticas predominantes en el fútbol sudamericano, contribuyendo a una visión más matizada de la diversidad y riqueza táctica presente en esta región.

3.2. Objetivo General

Las características de juego serán definidas utilizando modelos de machine learning como modelos lineales o similares, que analizarán los datos recopilados de las distintas ligas. Estos modelos permitirán calcular los coeficientes de cada variable, lo que revelará la contribución relativa de cada aspecto del juego en la diferencia observada entre las ligas. El análisis de estas diferencias proporcionará una comprensión más profunda de las particularidades y variaciones en el estilo de juego en las distintas ligas de la región sudamericana.

3.3. Objetivos Específicos

- Construir modelos de clasificación que permita predecir el resultado de un partido en base a la información obtenida a través de la plataforma Wyscout, para cada una de las ligas que componen la Conmebol.
- Identificar a través del modelo, cuáles son las características de juego de cada liga que compone Conmebol a través del peso de cada variable.
- Realizar comparaciones entre las distintas ligas y determinar si existen diferencias en el estilo de juego de cada liga.
- Explorar y definir agrupaciones alternativas de equipos dentro de la Conmebol: Desarrollar un análisis de agrupación para identificar patrones y similitudes en el estilo de juego entre diferentes equipos, independientemente de su liga de origen. Este objetivo se enfocará en descubrir agrupaciones transversales que trasciendan

las divisiones tradicionales por ligas, revelando afinidades y características comunes en el juego que podrían estar presentes en equipos de distintas ligas de la Conmebol. Este análisis ayudará a entender si las estrategias y tácticas empleadas por los equipos tienen correlaciones que no se limitan a las fronteras de sus respectivas ligas nacionales.

4. Datos y Metodología

4.1. Datos

La fuente de información para el desarrollo de este trabajo se extrajo de la plataforma Wyscout, considerando la información de resultados de partidos desde el año 2015 a 2022 de cada liga perteneciente a Conmebol, de esta forma se obtuvo el siguiente número de partidos por cada liga seleccionada:

País	Equipos	Partidos
Argentina	Boca Junior, Independiente, Racing Club, River Plate	1445
Bolivia	Bolivar, Oriente Petrolero, The Strongest, Wilstermann.	1352
Brasil	Flamengo, Palmeiras, Santos, Sao Paulo.	2431
Chile	Colo Colo, Unión Española, Universidad Católica, Universidad de Chile.	1295
Colombia	América de Cali, Atlético Nacional, Medellín, Millonarios.	1601
Ecuador	Barcelona, Delfín, El Nacional, Emelec, LDU Quito.	1548
Paraguay	Cerro Porteño, Libertad, Nacional Asunción, Olimpia.	1694

Peru	Alianza Lima, Cienciano, Melgar, Sporting Cristal, Universitario.	1561
Uruguay	Danubio, Defensor Sporting, Nacional, Peñarol.	1332
Venezuela	Caracas, Deportivo la Guaira, Deportivo la Táchira, Metropolitano.	969

Tabla 1: Datos empleados para análisis

La selección de equipos se desarrolló en base a la popularidad de cada uno de ellos en su respectivo país, por motivos de acceso a la plataforma Wyscout se estimó emplear solo esta cantidad de equipos, entendiendo que un mejor análisis hubiera resultado de la inclusión de todos los equipos pertenecientes a cada liga.

Las variables extraídas para cada liga fueron las siguientes:

Información Básica: 'Fecha', 'Partido', 'Competición', 'Duración', 'Equipo', 'Seleccionar esquema'.

Rendimiento del Equipo: 'Goles', 'xG (Goles Esperados)', 'Tiros / a la portería'.

Pases y Posesión: 'Pases / logrados', 'Posesión del balón, %'.

Balones Perdidos y Recuperados: 'Balones perdidos / bajos / medios / altos', 'Balones recuperados /bajos / medios / altos'.

Duelos y Tiros: 'Duelos / ganados', 'Tiros de fuera del área / a la portería'.

Tipos de Ataque: 'Ataques posicionales / con remate', 'Contraataques / con remate', 'Jugadas a balón parado / con remate', 'Córneres / con remate', 'Tiros libres / con remate', 'Penaltis / marcados'.

Precisión y Tipos de Pases: 'Centros / precisos', 'Pases cruzados en profundidad completados', 'Pases en profundidad completados', 'Pases hacia adelante / logrados', 'Pases hacia atrás / logrados', 'Pases laterales / logrados', 'Pases largos / logrados', 'Pases en el último tercio / logrados', 'Pases progresivos / precisos'.

Duelos y Jugadas Defensivas: 'Duelos ofensivos / ganados', 'Duelos defensivos / ganados', 'Duelos aéreos / ganados', 'Entradas a ras de suelo / logradas', 'Interceptaciones', 'Despejes'.

Faltas y Tarjetas: 'Faltas', 'Tarjetas amarillas', 'Tarjetas rojas'.

Otros Detalles Técnicos: 'Fuera de juego', 'Goles recibidos', 'Tiros en contra / a la portería', 'Desmarques / logrados', 'Saques laterales / logrados', 'Saques de meta'.

Métricas Analíticas: 'Intensidad de paso', 'Promedio pases por posesión del balón', 'Lanzamiento largo %', 'Distancia media de tiro', 'Longitud media pases', 'PPDA (Presión Defensiva Pasiva)'.

Dado que la hipótesis presentada como principal objetivo es la comparación de las características de juego de las distintas ligas se procedió a realizar el siguiente análisis para cada una de las ligas Conmebol:

1. Limpieza y Preparación de Datos:

1. Eliminación de registros duplicados.

2. Tratamiento de valores faltantes.
3. Segmentación o categorización, si es necesario.

2. Análisis Exploratorio de Datos:

1. Estadísticas descriptivas (media, mediana, moda, desviación estándar, etc.).
2. Visualización de datos (histogramas, gráficos de dispersión, box plots, etc.).

3. Selección y Transformación de Características:

1. Ingeniería de características.
2. Normalización o escalado de características, si es necesario.

4.2.1 Limpieza y Preparación de Datos

Conforme a lo previamente establecido, el proceso de limpieza y preparación de datos se implementó meticulosamente en cada uno de los diez conjuntos de datos extraídos de la plataforma Wyscout. El protocolo de trabajo estructurado para esta depuración abarcó procedimientos tales como: la detección de entradas vacías o missing data, la supresión de columnas no pertinentes, la identificación de columnas categóricas univariantes, el reconocimiento de columnas numéricas con valores constantes, el rastreo de registros duplicados y la localización de outliers o valores atípicos.

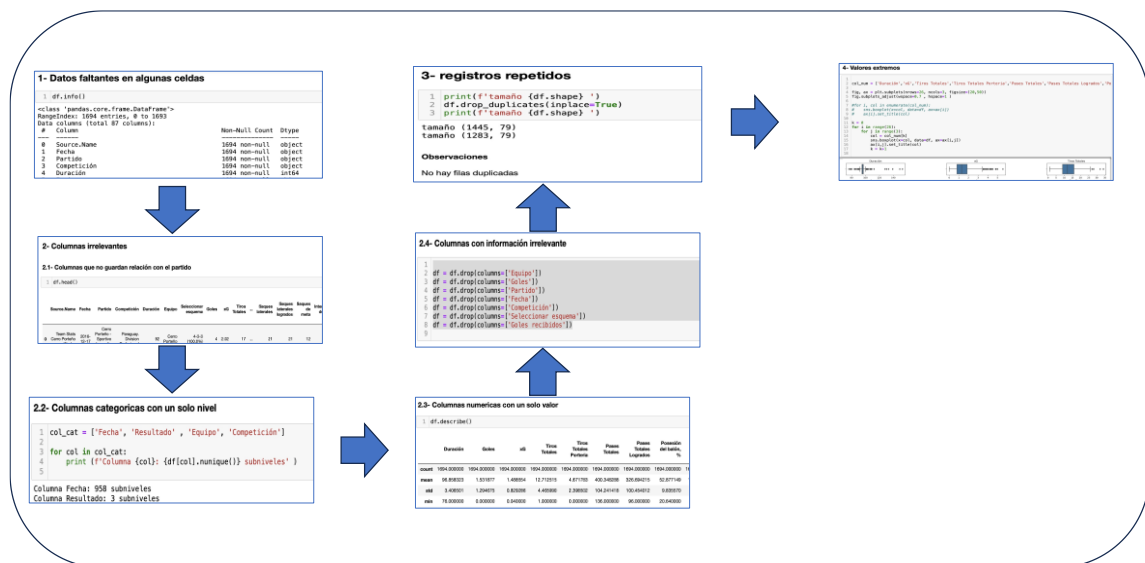


Ilustración 1: Proceso de Limpieza y Preparación de Datos

Tras ejecutar el procedimiento de limpieza, se ha validado que el conjunto de datos está bastante limpio, es decir, no se han identificado registros que necesiten ser eliminados, ni se han detectado valores nulos o espacios en blanco. No obstante, el análisis preliminar de la información revela aproximadamente 109 atributos que caracterizan cada encuentro deportivo. Al profundizar en esta data, es evidente que algunos de estos atributos son

derivados mediante operaciones de cálculo porcentual basadas en los datos originales. Estos atributos derivados introducen redundancia y, por lo tanto, se ha decidido prescindir de dichas columnas para optimizar el análisis. Adicionalmente, se ha incorporado un atributo denominado 'Resultado', el cual clasifica cada partido según tres categorías: triunfo, empate o derrota. Como consecuencia de estas intervenciones, el número total de atributos se ha recalibrado a 81.

```

3- Tipos de Datos
In [6]: df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1445 entries, 0 to 1444
Data columns (total 81 columns):
#   Column                                     Non-Null Count  Dtype
---  ---
0   Duración                                   1445 non-null   int64
1   Goles                                     1445 non-null   int64
2   xG                                         1445 non-null   float64
3   Tiros Totales                             1445 non-null   int64
4   Tiros Totales Portería                    1445 non-null   int64
5   Pases Totales                             1445 non-null   int64
6   Pases Totales Logrados                    1445 non-null   int64
7   Posesión del balón, %                     1445 non-null   float64
8   Balones perdidos                          1445 non-null   int64
9   Balones Perdidos bajos                    1445 non-null   int64
10  Balones perdidos medios                   1445 non-null   int64
11  Balones perdidos altos                    1445 non-null   int64
12  Balones recuperados /bajos / medios / altos 1445 non-null   int64
13  Balones recuperados bajos                 1445 non-null   int64
14  Balones recuperados medios                1445 non-null   int64
15  Balones recuperados altos                 1445 non-null   int64
16  Duelos Totales                            1445 non-null   int64
17  Duelos Ganados                            1445 non-null   int64
18  Tiros de Fuera del área                   1445 non-null   int64
19  Tiros de fuera del área a la portería     1445 non-null   int64
20  Ataques posicionales                      1445 non-null   int64
21  Ataques posicionales con remate           1445 non-null   int64
22  Contraataques                             1445 non-null   int64
23  Contraataques con remate                  1445 non-null   int64
24  Jugadas a balón parado                    1445 non-null   int64
25  Jugadas a balón parado con remate         1445 non-null   int64
26  Córneres                                  1445 non-null   int64
27  Córneres con remate                       1445 non-null   int64
28  Tiros libres                              1445 non-null   int64
29  Tiros libres con remate                   1445 non-null   int64
30  Penaltis                                  1445 non-null   int64
31  Penaltis marcados                         1445 non-null   int64
32  Centros                                   1445 non-null   int64
33  Centros precisos                          1445 non-null   int64
34  Pases cruzados en profundidad completados 1445 non-null   int64

```

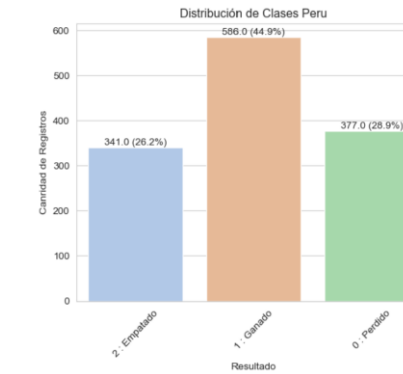
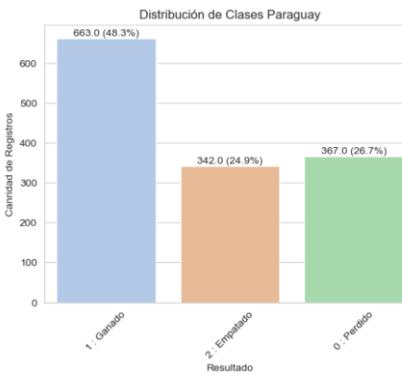
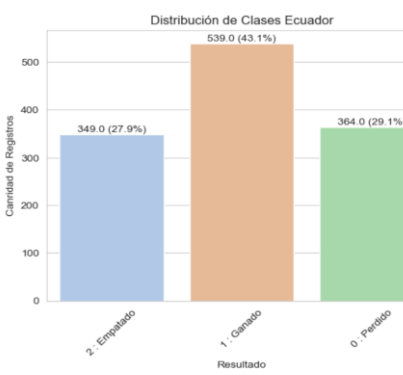
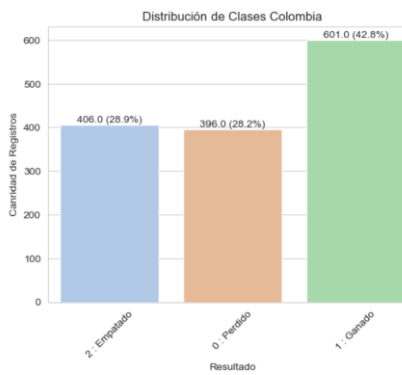
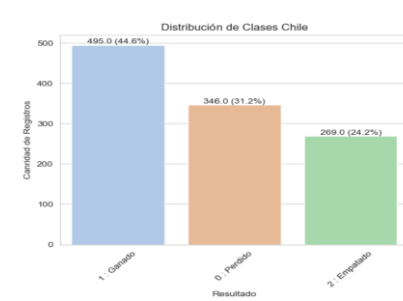
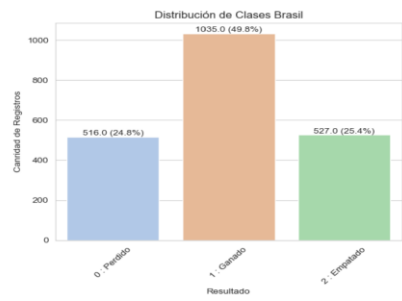
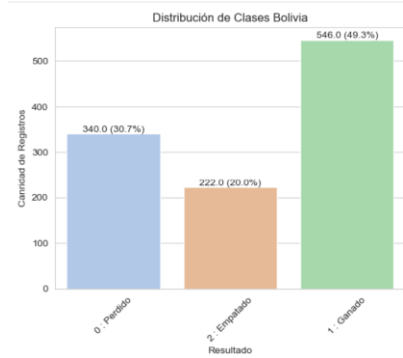
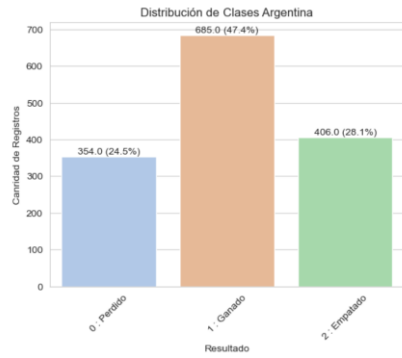
Ilustración 2: Características de la Data

Se destaca que los atributos son del tipo int64 y float64.

4.2.2 Análisis Exploratorio de Datos

Tomando en consideración el rendimiento óptimo esperado de los algoritmos predictivos a ser implementados para las distintas ligas de la Conmebol, se considera fundamental, como etapa inicial de análisis, evaluar la distribución balanceada de las clases: triunfos, empates y derrotas. Este equilibrio es crucial para evitar sesgos en el entrenamiento y

evaluación de los modelos. A continuación, se presentan visualizaciones gráficas detalladas que representan la distribución de estas clases para cada liga.



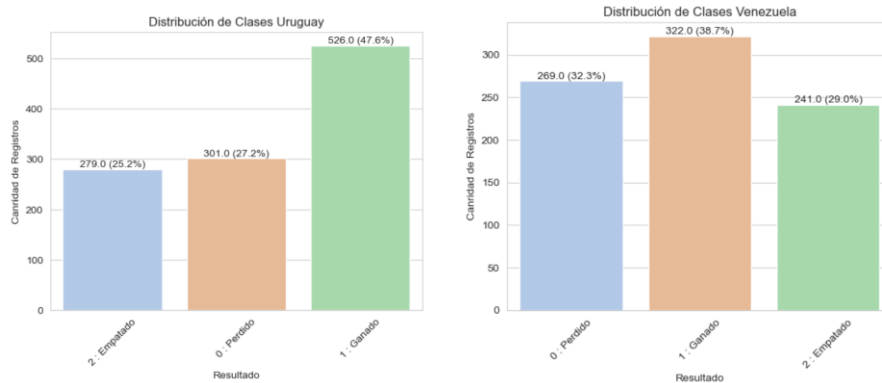


Ilustración 3: Distribución de Clases por País

Tras la inspección de las visualizaciones, se identifica un desequilibrio significativo en la distribución de las categorías. Notablemente, la cantidad de partidos ganados supera con creces a las otras categorías en todas las ligas analizadas. Con el fin de rectificar esta disparidad y garantizar un análisis robusto, se genera re-muestreo. Se desarrolla un algoritmo que identifique la categoría con la menor frecuencia y, posteriormente, se procede con un muestreo aleatorio estratificado para los partidos ganados y perdidos. De este modo, se busca establecer una homogeneidad entre las tres categorías del juego.

Adicionalmente se realiza una serie de histogramas para todos los atributos comprendidos en los datos analizados, dada la enorme cantidad de atributos presentes se presentan algunas interpretaciones de los datos para algunas variables contenidas en cada país.

4.2.2.1 Argentina

Duración: La mayoría de los partidos tienen una duración cercana a los 90 minutos, lo cual es esperado. Sin embargo, hay algunos partidos con duraciones superiores, lo que podría deberse a tiempos extra, lesiones prolongadas, o pausas significativas.

Goles: La mayoría de los partidos resultan en 0 o 1 gol. Sin embargo, hay una cantidad considerable de partidos donde se marcaron hasta 3 goles. Es menos común ver partidos con 4 o más goles.

xG (Expected Goals): Similar a la distribución de goles, la mayoría de los valores de xG están cerca de 0 o 1, indicando que en la mayoría de los partidos se esperaba que se marcaran pocos goles. No obstante, hay ocasiones donde xG es mayor, lo que sugiere que hubo oportunidades claras para marcar que no se concretaron.

Tiros Totales: La mayoría de los equipos tienen entre 0 y 10 tiros en total durante un partido. Sin embargo, hay equipos que realizan hasta 20 tiros o más en un solo partido, lo que indica un enfoque ofensivo agresivo.

Posesión del balón, %: La distribución de la posesión del balón está centrada alrededor del 50%, lo que indica un equilibrio en la posesión del balón entre los equipos. Sin embargo, hay equipos que dominan la posesión, llegando hasta el 75% o más.

Balones perdidos: La mayoría de los equipos pierden el balón entre 0 y 10 veces por partido. Sin embargo, hay ocasiones en las que un equipo puede perder el balón más de 20 veces, lo que sugiere un juego menos controlado o una alta presión del equipo contrario.

Duelos Totales: La distribución muestra que la mayoría de los equipos tienen entre 10 y 20 duelos totales en un partido. Hay equipos que alcanzan más de 30 duelos, lo que indica un juego físico y confrontacional.

Faltas: La mayoría de los equipos cometen entre 0 y 5 faltas por partido. No obstante, hay equipos que pueden cometer más de 10 faltas en un solo partido, lo que sugiere un juego más agresivo o tácticas defensivas más robustas.

Tiros Totales Portería: La mayoría de los equipos realizan entre 0 y 5 tiros directos a la portería durante un partido. Hay equipos que logran más de 10 tiros a portería, lo que indica una ofensiva con precisión y eficacia.

Pases Totales: Los equipos tienden a realizar entre 100 y 300 pases en un partido. Sin embargo, hay equipos que superan los 400 pases, lo que sugiere una estrategia de posesión y control del juego.

Pases Totales Logrados: Similar a los pases totales, la mayoría de los equipos logran completar entre 100 y 300 pases. Equipos con una alta cantidad de pases logrados probablemente tienen una buena precisión en el pase y mantienen el balón con eficacia.

Balones Perdidos bajos: La distribución muestra que la mayoría de los equipos pierden el balón en zonas bajas (probablemente cerca de su propia área) entre 0 y 10 veces. Estos balones perdidos pueden resultar en contraataques peligrosos por el equipo contrario.

Balones perdidos medios: La mayoría de los equipos pierden el balón en zonas medias entre 0 y 10 veces. La zona media es crucial para controlar el ritmo del juego, por lo que perder el balón aquí puede cambiar la dinámica del partido.

Balones perdidos altos: Los equipos tienden a perder el balón en zonas altas (cerca del área contraria) con más frecuencia, indicando la presión alta del equipo defensor o los intentos de penetrar defensas sólidas.

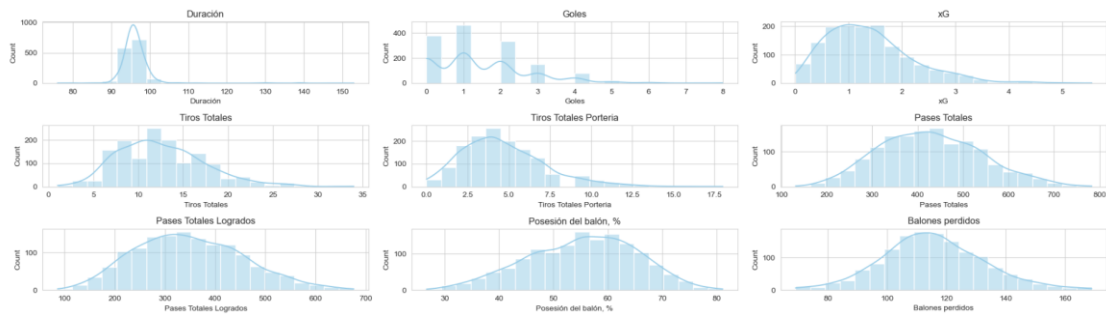


Ilustración 4: Distribución Argentina

4.2.2.2 Bolivia

Duración: La mayoría de los partidos en Bolivia duran alrededor de 90 minutos, lo cual es el estándar. Sin embargo, hay una notable cantidad de partidos que superan esta duración, lo que podría indicar tiempos extra, pausas prolongadas o lesiones.

Goles: En Bolivia, la mayoría de los partidos resultan en 0, 1 o 2 goles. Es menos frecuente encontrar partidos con 3 o más goles.

xG (Expected Goals): La métrica xG sugiere que en la mayoría de los partidos se esperaba que se marcaran 0, 1 o 2 goles. Hay menos ocasiones donde el xG es mayor a 2, lo que indica partidos donde se esperaba una mayor cantidad de goles.

Tiros Totales: La mayoría de los equipos en Bolivia realizan entre 0 y 15 tiros en total por partido. Es raro ver equipos que superen los 20 tiros en un juego.

Posesión del balón, %: La posesión del balón tiende a centrarse alrededor del 50%, lo que indica un equilibrio general entre los equipos. Sin embargo, hay equipos que logran dominar la posesión, alcanzando porcentajes superiores al 70%.

Balones perdidos: La mayoría de los equipos en Bolivia pierden el balón entre 0 y 15 veces en un partido. Las ocasiones en las que un equipo pierde el balón más de 20 veces son menos comunes.

Duelos Totales: En Bolivia, la mayoría de los equipos tienen entre 10 y 25 duelos totales en un partido. Es menos común ver equipos con más de 30 duelos en un partido.

Faltas: La mayoría de los equipos en Bolivia cometen entre 0 y 7 faltas por partido. Equipos que cometen más de 10 faltas son menos frecuentes.

Tiros Totales Portería: La mayoría de los equipos en Bolivia realizan entre 0 y 5 tiros directos a portería por partido. Sin embargo, hay equipos que logran más de 10 tiros directos a la portería, lo que sugiere una precisión ofensiva notable.

Pases Totales: La mayoría de los equipos realizan entre 100 y 300 pases en un partido. Hay equipos que superan los 400 pases, lo que indica una estrategia centrada en la posesión y control del juego.

Pases Totales Logrados: La mayoría de los equipos logran completar entre 100 y 300 pases por partido. Es raro ver equipos que logren completar más de 400 pases con éxito.

Balones Perdidos bajos: La mayoría de los equipos pierden el balón en zonas bajas (probablemente cerca de su propia área) entre 0 y 10 veces por partido.

Balones perdidos medios: La mayoría de los equipos pierden el balón en zonas medias entre 0 y 10 veces. Es menos frecuente ver equipos que pierden el balón más de 15 veces en esta zona.

Balones perdidos altos: En Bolivia, es común que los equipos pierdan el balón en zonas altas (cerca del área contraria) entre 0 y 15 veces por partido.

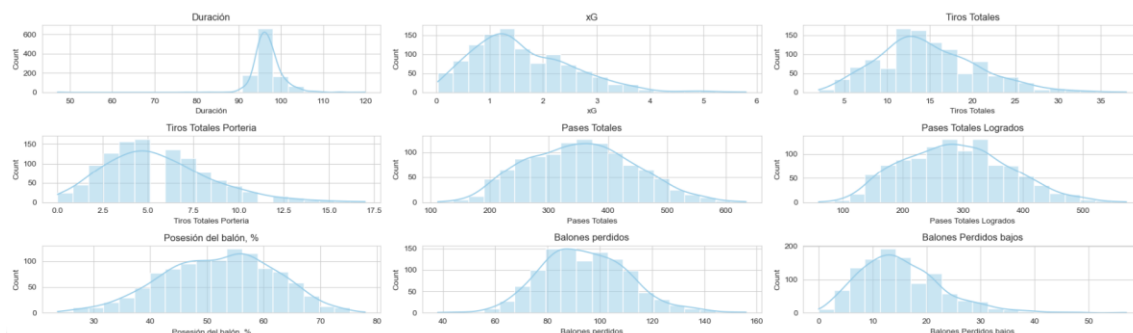


Ilustración 5: Distribución Bolivia

4.2.2.3 Brasil

Duración: La mayoría de los partidos en Brasil duran alrededor de 90 minutos. Sin embargo, se observa un rango considerable de partidos que superan esta duración, posiblemente debido a tiempos extra, pausas o lesiones.

Goles: La distribución de goles muestra que la mayoría de los partidos en Brasil resultan en 0, 1 o 2 goles. Es menos común tener partidos con 3 o más goles.

xG (Expected Goals): La métrica xG sugiere que en la mayoría de los partidos se esperaba que se marcaran 0, 1 o 2 goles. Hay ocasiones donde el xG es mayor a 2, lo que indica partidos donde se esperaban más goles.

Tiros Totales: Los equipos en Brasil tienden a realizar entre 0 y 15 tiros en total por partido. Es menos común observar equipos que superen los 20 tiros en un juego.

Posesión del balón, %: La posesión del balón en Brasil muestra una distribución centrada alrededor del 50%. Sin embargo, hay equipos que dominan la posesión, alcanzando porcentajes superiores al 70%.

Balones perdidos: En Brasil, la mayoría de los equipos pierden el balón entre 0 y 15 veces por partido. Es menos común observar equipos que pierden el balón más de 20 veces en un juego.

Duelos Totales: La mayoría de los equipos en Brasil tienen entre 10 y 25 duelos totales por partido. Es raro observar equipos con más de 30 duelos en un partido.

Faltas: La mayoría de los equipos en Brasil cometen entre 0 y 8 faltas por partido. Equipos que cometen más de 10 faltas son menos comunes.

Tiros Totales Portería: La mayoría de los equipos en Brasil realizan entre 0 y 5 tiros directos a portería por partido. Sin embargo, hay equipos que logran más de 10 tiros directos a la portería.

Pases Totales: Los equipos en Brasil tienden a realizar entre 100 y 350 pases por partido. Algunos equipos superan los 400 pases, lo que indica una estrategia de control del juego centrada en la posesión.

Pases Totales Logrados: Similar a los pases totales, la mayoría de los equipos logran completar entre 100 y 350 pases por partido.

Balones Perdidos bajos: La mayoría de los equipos pierden el balón en zonas bajas entre 0 y 10 veces por partido.

Balones perdidos medios: En Brasil, es común que los equipos pierdan el balón en zonas medias entre 0 y 10 veces por partido.

Balones perdidos altos: La mayoría de los equipos pierden el balón en zonas altas entre 0 y 15 veces por partido.

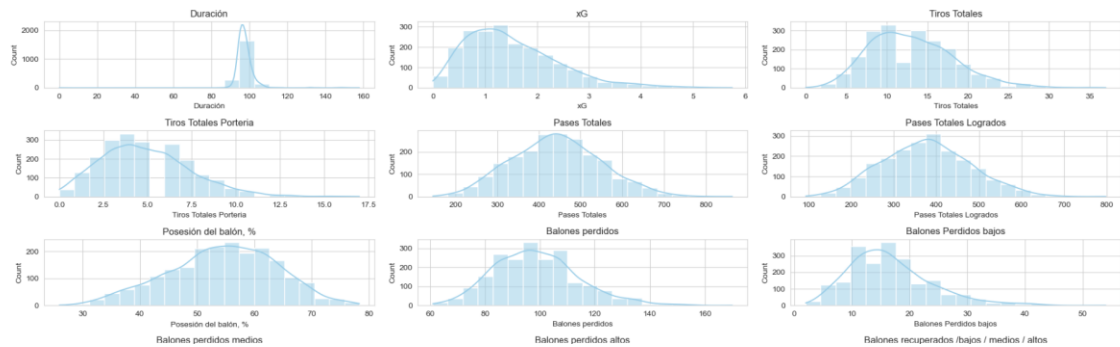


Ilustración 6: Distribución Brasil

4.2.2.4 Chile

Goles: La distribución de goles muestra que la mayoría de los partidos en Chile resultan en 1.00 goles. Es menos común tener partidos con 3 o más goles.

xG: La métrica xG sugiere que en la mayoría de los partidos en Chile se esperaba que se marcaran 1.38 goles. Hay ocasiones donde el xG es mayor a 2, lo que indica partidos donde se esperaban más goles.

Tiros Totales: Los equipos en Chile tienden a realizar entre 0 y 32 tiros en total por partido. Es menos común observar equipos que superen los 20 tiros en un juego.

Posesión del balón, %: La posesión del balón en Chile muestra una distribución centrada alrededor del 52%. Sin embargo, hay equipos que dominan la posesión, alcanzando porcentajes superiores al 70%.

Duelos Totales: La mayoría de los equipos en Chile tienen entre 132 y 608 duelos totales por partido. Es raro observar equipos con más de 30 duelos en un partido.

Faltas: La mayoría de los equipos en Chile cometen entre 4 y 31 faltas por partido. Equipos que cometen más de 10 faltas son menos comunes.

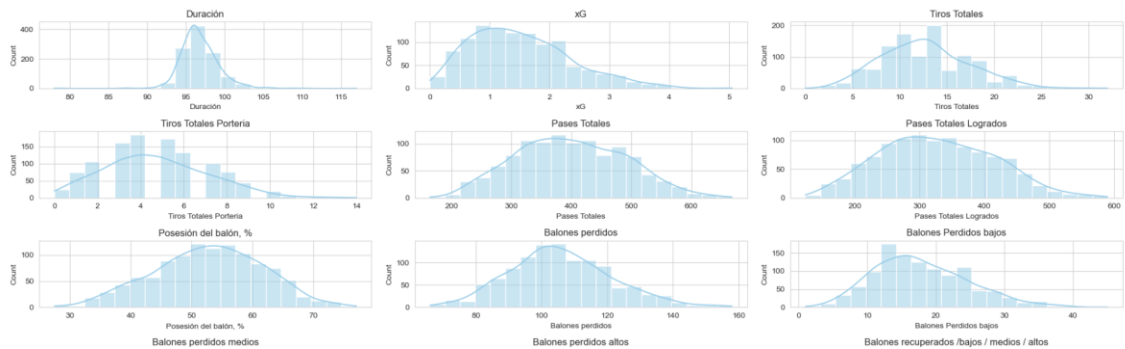


Ilustración 7: Distribución Chile

4.2.2.5 Colombia

Goles: La distribución de goles muestra que la mayoría de los partidos en Colombia resultan en 1.00 goles. Es menos común tener partidos con 3 o más goles.

xG: La métrica xG sugiere que en la mayoría de los partidos en Colombia se esperaba que se marcaran 1.25 goles. Hay ocasiones donde el xG es mayor a 2, lo que indica partidos donde se esperaban más goles.

Tiros Totales: Los equipos en Colombia tienden a realizar entre 1 y 31 tiros en total por partido. Es menos común observar equipos que superen los 20 tiros en un juego.

Posesión del balón, %: La posesión del balón en Colombia muestra una distribución centrada alrededor del 52.65%. Sin embargo, hay equipos que dominan la posesión, alcanzando porcentajes superiores al 70%.

Duelos Totales: La mayoría de los equipos en Colombia tienen entre 112 y 383 duelos totales por partido. Es raro observar equipos con más de 30 duelos en un partido.

Faltas: La mayoría de los equipos en Colombia cometen entre 3.00 y 31.00 faltas por partido. Equipos que cometen más de 10 faltas son menos comunes.

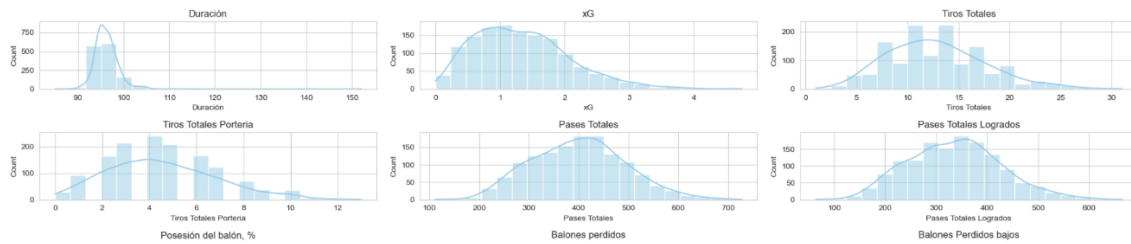


Ilustración 8: Distribución Colombia

4.2.2.6 Ecuador

Goles: La distribución de goles muestra que la mayoría de los partidos en Ecuador resultan en 1.00 goles. Es menos común tener partidos con 3 o más goles.

xG: La métrica xG sugiere que en la mayoría de los partidos en Ecuador se esperaba que se marcaran 1.29 goles. Hay ocasiones donde el xG es mayor a 2, lo que indica partidos donde se esperaban más goles.

Tiros Totales: Los equipos en Ecuador tienden a realizar entre 1 y 33 tiros en total por partido. Es menos común observar equipos que superen los 20 tiros en un juego.

Poseción del balón, %: La posesión del balón en Ecuador muestra una distribución centrada alrededor del 51.22%. Sin embargo, hay equipos que dominan la posesión, alcanzando porcentajes superiores al 70%.

Duelos Totales: La mayoría de los equipos en Ecuador tienen entre 116.00 y 321.00 duelos totales por partido. Es raro observar equipos con más de 30 duelos en un partido.

Faltas: La mayoría de los equipos en Ecuador cometen entre 4.00 y 39.00 faltas por partido. Equipos que cometen más de 10 faltas son menos comunes.

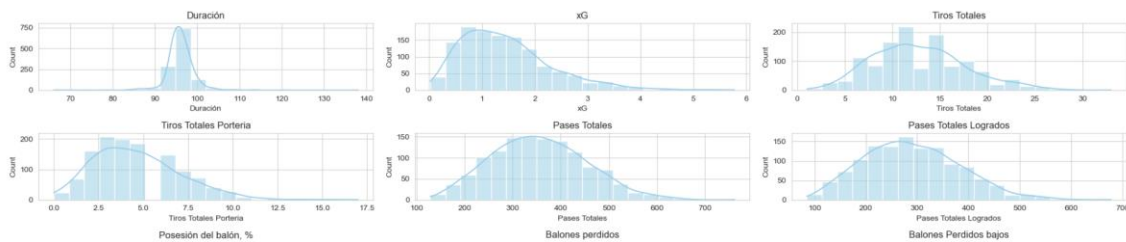


Ilustración 9: Distribución Ecuador

4.2.2.7 Paraguay

Goles: La distribución de goles muestra que la mayoría de los partidos en Paraguay resultan en 1.00 goles. Es menos común tener partidos con 3 o más goles.

xG: La métrica xG sugiere que en la mayoría de los partidos en Paraguay se esperaba que se marcaran 1.36 goles. Hay ocasiones donde el xG es mayor a 2, lo que indica partidos donde se esperaban más goles.

Tiros Totales: Los equipos en Paraguay tienden a realizar entre 1.00 y 33.00 tiros en total por partido. Es menos común observar equipos que superen los 20 tiros en un juego.

Poseción del balón, %: La posesión del balón en Paraguay muestra una distribución centrada alrededor del 52.88%. Sin embargo, hay equipos que dominan la posesión, alcanzando porcentajes superiores al 70%.

Duelos Totales: La mayoría de los equipos en Paraguay tienen entre 126.00 y 374.00 duelos totales por partido. Es raro observar equipos con más de 30 duelos en un partido.

Faltas: La mayoría de los equipos en Paraguay cometen entre 1.00 y 28.00 faltas por partido. Equipos que cometen más de 10 faltas son menos comunes.

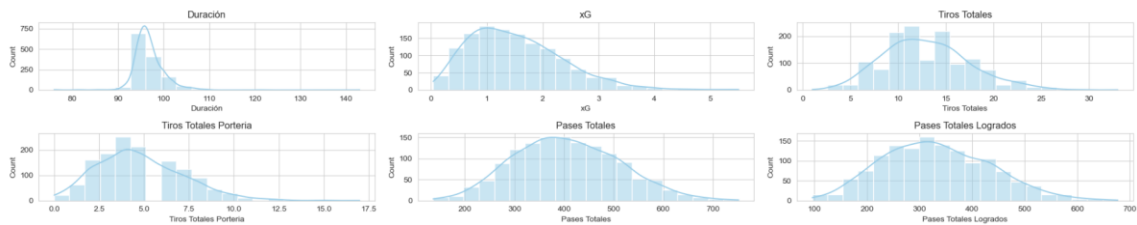


Ilustración 10: Distribución Paraguay

4.2.2.8 Peru

Goles: La distribución de goles muestra que la mayoría de los partidos en Peru resultan en 1.00 goles. Es menos común tener partidos con 3 o más goles.

xG: La métrica xG sugiere que en la mayoría de los partidos en Peru se esperaba que se marcaran 1.35 goles. Hay ocasiones donde el xG es mayor a 2, lo que indica partidos donde se esperaban más goles.

Tiros Totales: Los equipos en Peru tienden a realizar entre 1.00 y 37.00 tiros en total por partido. Es menos común observar equipos que superen los 20 tiros en un juego.

Poseción del balón, %: La posesión del balón en Peru muestra una distribución centrada alrededor del 52.35%. Sin embargo, hay equipos que dominan la posesión, alcanzando porcentajes superiores al 70%.

Duelos Totales: La mayoría de los equipos en Peru tienen entre 84.00 y 365.00 duelos totales por partido. Es raro observar equipos con más de 30 duelos en un partido.

Faltas: La mayoría de los equipos en Peru cometen entre 3.00 y 31.00 faltas por partido. Equipos que cometen más de 10 faltas son menos comunes.

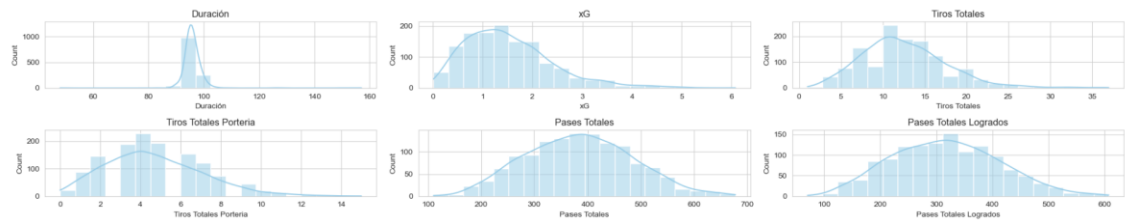


Ilustración 11: Distribución Perú

4.2.2.9 Uruguay

Goles: La distribución de goles muestra que la mayoría de los partidos en Uruguay resultan en 1.00 goles. Es menos común tener partidos con 3 o más goles.

xG: La métrica xG sugiere que en la mayoría de los partidos en Uruguay se esperaba que se marcaran 1.37 goles. Hay ocasiones donde el xG es mayor a 2, lo que indica partidos donde se esperaban más goles.

Tiros Totales: Los equipos en Uruguay tienden a realizar entre 2.00 y 29.00 tiros en total por partido. Es menos común observar equipos que superen los 20 tiros en un juego.

Poseción del balón, %: La posesión del balón en Uruguay muestra una distribución centrada alrededor del 51.87%. Sin embargo, hay equipos que dominan la posesión, alcanzando porcentajes superiores al 70%.

Duelos Totales: La mayoría de los equipos en Uruguay tienen entre 135.00 y 355.00 duelos totales por partido. Es raro observar equipos con más de 30 duelos en un partido.

Faltas: La mayoría de los equipos en Uruguay cometen entre 4.00 y 35.00 faltas por partido. Equipos que cometen más de 10 faltas son menos comunes.

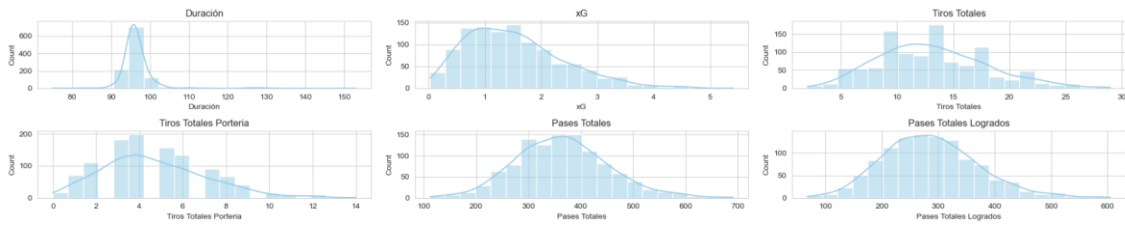


Ilustración 12: Distribución Uruguay

4.2.2.10 Venezuela

Goles: La distribución de goles muestra que la mayoría de los partidos en Venezuela resultan en 1.00 goles. Es menos común tener partidos con 3 o más goles.

xG: La métrica xG sugiere que en la mayoría de los partidos en Venezuela se esperaba que se marcaran 1.21 goles. Hay ocasiones donde el xG es mayor a 2, lo que indica partidos donde se esperaban más goles.

Tiros Totales: Los equipos en Venezuela tienden a realizar entre 0.00 y 32.00 tiros en total por partido. Es menos común observar equipos que superen los 20 tiros en un juego.

Poseción del balón, %: La posesión del balón en Venezuela muestra una distribución centrada alrededor del 49.92%. Sin embargo, hay equipos que dominan la posesión, alcanzando porcentajes superiores al 70%.

Duelos Totales: La mayoría de los equipos en Venezuela tienen entre 130.00 y 341.00 duelos totales por partido. Es raro observar equipos con más de 30 duelos en un partido.

Faltas: La mayoría de los equipos en Venezuela cometen entre 5.00 y 35.00 faltas por partido. Equipos que cometen más de 10 faltas son menos comunes.

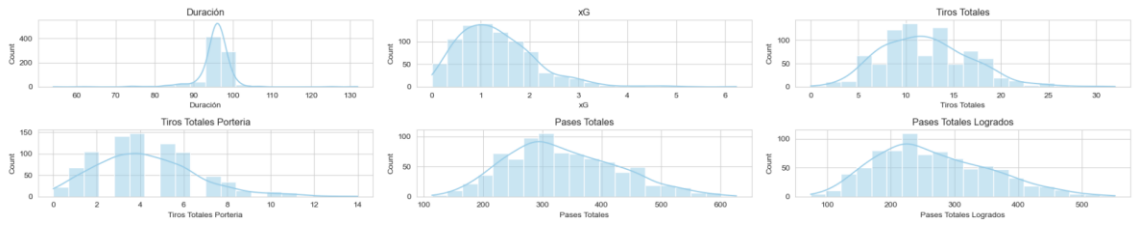


Ilustración 13: Distribución Venezuela

4.2.2.10 Tabla Comparativa

Se escogen algunas variables por país y se realizan comparaciones respecto a la distribución de datos para Duelos Totales, Goles y xG.

País	Duelos Totales	Goles	xG
Argentina	Argentina tiene un promedio de 245.23 duelos totales por partido, con una mediana de 243.00, una desviación estándar de 36.90, un mínimo de 125.00 y un máximo de 369.00. Este es uno de los promedios más altos, superando a países como Bolivia y Brasil y Chile y Colombia y Ecuador y Paraguay y Peru y Uruguay y Venezuela.	Argentina tiene un promedio de 1.46 goles por partido, con una mediana de 1.00, una desviación estándar de 1.33, un mínimo de 0.00 y un máximo de 8.00. El promedio es mayor que Colombia y Uruguay y Venezuela pero menor que Bolivia y Brasil y Chile y Ecuador y Paraguay y Peru.	Argentina tiene un promedio de 1.38 xg por partido, con una mediana de 1.27, una desviación estándar de 0.82, un mínimo de 0.01 y un máximo de 5.55. El promedio es mayor que Colombia y Venezuela pero menor que Bolivia y Brasil y Chile y Ecuador y Paraguay y Peru y Uruguay.
Bolivia	Bolivia tiene un promedio de 197.58 duelos totales por partido, con una mediana de 196.00, una desviación estándar de 35.85, un mínimo de 47.00 y un máximo de 334.00. Este es uno de los promedios más bajos, siendo superado por países como Argentina y Brasil y Chile y Colombia y Ecuador y Paraguay y Peru y Uruguay y Venezuela.	Bolivia tiene un promedio de 1.73 goles por partido, con una mediana de 1.00, una desviación estándar de 1.52, un mínimo de 0.00 y un máximo de 8.00. Este es uno de los promedios más altos, superando a países como Argentina y Brasil y Chile y Colombia y Ecuador y Paraguay y Peru y Uruguay y Venezuela.	Bolivia tiene un promedio de 1.58 xg por partido, con una mediana de 1.40, una desviación estándar de 0.92, un mínimo de 0.03 y un máximo de 5.80. Este es uno de los promedios más altos, superando a países como Argentina y Brasil y Chile y Colombia y Ecuador y Paraguay y Peru y Uruguay y Venezuela.
Brasil	Brasil tiene un promedio de 215.32 duelos totales por partido, con una mediana de 213.00, una desviación estándar de 33.59, un mínimo de 116.00 y un máximo de 411.00. El promedio es mayor que Bolivia y Colombia y Ecuador pero menor que Argentina y Chile y Paraguay y Peru y Uruguay y Venezuela.	Brasil tiene un promedio de 1.51 goles por partido, con una mediana de 1.00, una desviación estándar de 1.28, un mínimo de 0.00 y un máximo de 9.00. El promedio es mayor que Argentina y Chile y Colombia y Ecuador y Peru y Uruguay y Venezuela pero menor que Bolivia y Paraguay.	Brasil tiene un promedio de 1.48 xg por partido, con una mediana de 1.32, una desviación estándar de 0.88, un mínimo de 0.00 y un máximo de 5.75. El promedio es mayor que Argentina y Colombia y Ecuador y Venezuela pero menor que Bolivia y Chile y Paraguay y Peru y Uruguay.
Chile	Chile tiene un promedio de 228.86 duelos totales por partido, con una mediana de 228.00, una desviación estándar de 36.22, un mínimo de 132.00 y un máximo de 608.00. El promedio es mayor que Bolivia y Brasil y Colombia y Ecuador y Paraguay y Peru y Venezuela pero menor que Argentina y Uruguay.	Chile tiene un promedio de 1.48 goles por partido, con una mediana de 1.00, una desviación estándar de 1.22, un mínimo de 0.00 y un máximo de 8.00. El promedio es mayor que Argentina y Colombia y Ecuador y Peru y Uruguay y Venezuela pero menor que Bolivia y Brasil y Paraguay.	Chile tiene un promedio de 1.50 xg por partido, con una mediana de 1.38, una desviación estándar de 0.84, un mínimo de 0.00 y un máximo de 5.05. El promedio es mayor que Argentina y Brasil y Colombia y Ecuador y Paraguay y Peru y Uruguay y Venezuela pero menor que Bolivia.
Colombia	Colombia tiene un promedio de 209.95 duelos totales por partido, con una mediana de 209.00, una desviación estándar de 31.53, un mínimo de 112.00 y un máximo de 383.00. El promedio es mayor que Bolivia pero menor que Argentina y Brasil y Chile y Ecuador y Paraguay y Peru y Uruguay y Venezuela.	Colombia tiene un promedio de 1.34 goles por partido, con una mediana de 1.00, una desviación estándar de 1.14, un mínimo de 0.00 y un máximo de 7.00. El promedio es mayor que Venezuela pero menor que Argentina y Bolivia y Brasil y Chile y Ecuador y Paraguay y Peru y Uruguay.	Colombia tiene un promedio de 1.33 xg por partido, con una mediana de 1.25, una desviación estándar de 0.74, un mínimo de 0.00 y un máximo de 4.74. El promedio es mayor que Venezuela pero menor que Argentina y Bolivia y Brasil y Chile y Ecuador y Paraguay y Peru y Uruguay.
Ecuador	Ecuador tiene un promedio de 212.95 duelos totales por partido, con una mediana de 213.00, una desviación estándar de 36.16, un mínimo de 116.00 y un máximo de 321.00. El promedio es mayor que Bolivia y Colombia pero menor que Argentina y Brasil y Chile y Paraguay y Peru y Uruguay y Venezuela.	Ecuador tiene un promedio de 1.46 goles por partido, con una mediana de 1.00, una desviación estándar de 1.23, un mínimo de 0.00 y un máximo de 7.00. El promedio es mayor que Argentina y Colombia y Uruguay y Venezuela pero menor que Bolivia y Brasil y Chile y Paraguay y Peru.	Ecuador tiene un promedio de 1.44 xg por partido, con una mediana de 1.29, una desviación estándar de 0.87, un mínimo de 0.02 y un máximo de 5.78. El promedio es mayor que Argentina y Colombia y Venezuela pero menor que Bolivia y Brasil y Chile y Paraguay y Peru y Uruguay.
Paraguay	Paraguay tiene un promedio de 225.03 duelos totales por partido, con una mediana de 224.00, una desviación estándar de 34.96, un mínimo de 126.00 y un máximo de 374.00. El promedio es mayor que Bolivia y Brasil y Colombia y Ecuador y Peru y Venezuela pero menor que Argentina y Chile y Uruguay.	Paraguay tiene un promedio de 1.53 goles por partido, con una mediana de 1.00, una desviación estándar de 1.29, un mínimo de 0.00 y un máximo de 7.00. El promedio es mayor que Argentina y Brasil y Chile y Colombia y Ecuador y Peru y Uruguay y Venezuela pero menor que Bolivia.	Paraguay tiene un promedio de 1.49 xg por partido, con una mediana de 1.36, una desviación estándar de 0.83, un mínimo de 0.04 y un máximo de 5.51. El promedio es mayor que Argentina y Brasil y Colombia y Ecuador y Venezuela pero menor que Bolivia y Chile y Peru y Uruguay.
Peru	Peru tiene un promedio de 217.55 duelos totales por partido, con una mediana de 218.00, una desviación estándar de 38.77, un mínimo de 84.00 y un máximo de 365.00. El promedio es mayor que Bolivia y Brasil y Colombia y Ecuador y Venezuela pero menor que Argentina y Chile y Paraguay y Uruguay.	Peru tiene un promedio de 1.48 goles por partido, con una mediana de 1.00, una desviación estándar de 1.27, un mínimo de 0.00 y un máximo de 8.00. El promedio es mayor que Argentina y Colombia y Ecuador y Uruguay y Venezuela pero menor que Bolivia y Brasil y Chile y Paraguay.	Peru tiene un promedio de 1.49 xg por partido, con una mediana de 1.35, una desviación estándar de 0.88, un mínimo de 0.01 y un máximo de 6.08. El promedio es mayor que Argentina y Brasil y Colombia y Ecuador y Paraguay y Venezuela pero menor que Bolivia y Chile y Uruguay.
Uruguay	Uruguay tiene un promedio de 243.13 duelos totales por partido, con una mediana de 240.00, una desviación estándar de 35.38, un mínimo de 135.00 y un máximo de 355.00. El promedio es mayor que Bolivia y Brasil y Chile y Colombia y Ecuador y Paraguay y Peru y Venezuela pero menor que Argentina.	Uruguay tiene un promedio de 1.39 goles por partido, con una mediana de 1.00, una desviación estándar de 1.18, un mínimo de 0.00 y un máximo de 6.00. El promedio es mayor que Colombia y Venezuela pero menor que Argentina y Bolivia y Brasil y Chile y Ecuador y Paraguay y Peru.	Uruguay tiene un promedio de 1.50 xg por partido, con una mediana de 1.37, una desviación estándar de 0.87, un mínimo de 0.03 y un máximo de 5.41. El promedio es mayor que Argentina y Brasil y Colombia y Ecuador y Paraguay y Peru y Venezuela pero menor que Bolivia y Chile.
Venezuela	Venezuela tiene un promedio de 215.52 duelos totales por partido, con una mediana de 213.00, una desviación estándar de 31.41, un mínimo de 130.00 y un máximo de 341.00. El promedio es mayor que Bolivia y Brasil y Colombia y Ecuador pero menor que Argentina y Chile y Paraguay y Peru y Uruguay.	Venezuela tiene un promedio de 1.26 goles por partido, con una mediana de 1.00, una desviación estándar de 1.16, un mínimo de 0.00 y un máximo de 7.00. Este es uno de los promedios más bajos, siendo superado por países como Argentina y Bolivia y Brasil y Chile y Colombia y Ecuador y Paraguay y Peru y Uruguay.	Venezuela tiene un promedio de 1.32 xg por partido, con una mediana de 1.21, una desviación estándar de 0.79, un mínimo de 0.00 y un máximo de 6.23. Este es uno de los promedios más bajos, siendo superado por países como Argentina y Bolivia y Brasil y Chile y Colombia y Ecuador y Paraguay y Peru y Uruguay.

Tabla 2: Tabla Comparativa

4.2. Metodología

El objetivo de este trabajo es determinar las características de juego de cada una de las ligas que componen la Conmebol y determinar si existen similitudes en cuanto al juego que desarrollan, para determinar las características de juego se opta por emplear modelos de clasificación que permitan determinar el peso de cada variable y de esta forma conformar vectores de peso por país y luego comparar estos vectores con técnicas como PCA, igualdad coseno, distancia euclidiana, con las herramientas mencionadas se determinará un criterio que nos permita indicar la similitud entre las distintas ligas que componen Conmebol. La evaluación de los modelos se realizará empleando métricas que contribuyan a entender la calidad de los modelos desarrollados.

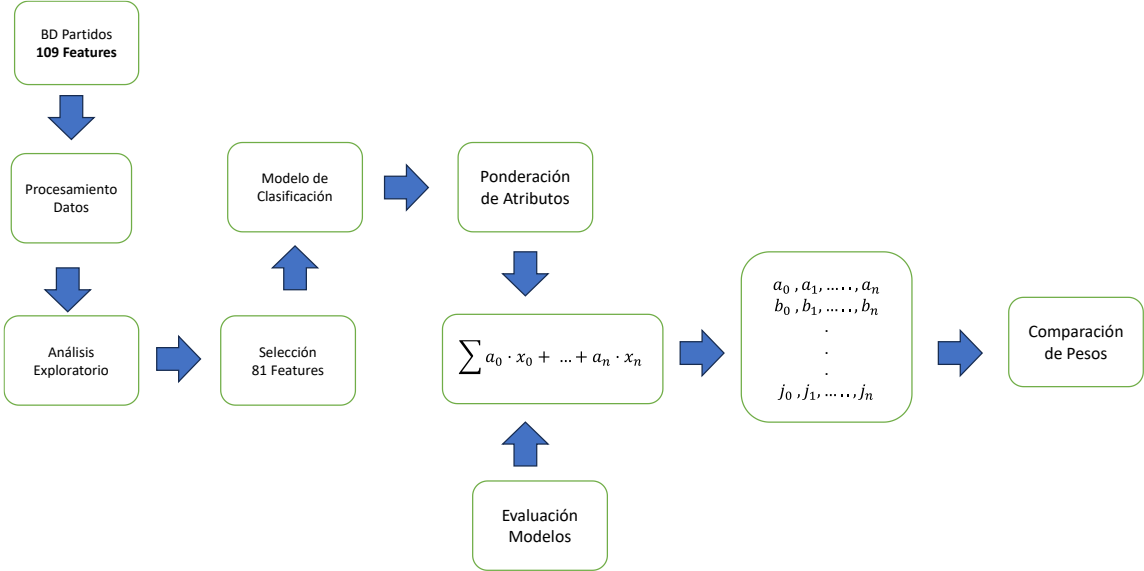


Ilustración 14: Metodología Empleada para ligas Nacionales

En caso de que no se detecten diferencias significativas entre las distintas ligas, se procederá a agrupar por equipos. Esto se hará con el objetivo de identificar las variaciones sutiles entre ellos, lo cual ayudará a definir y distinguir diversos estilos de juego.

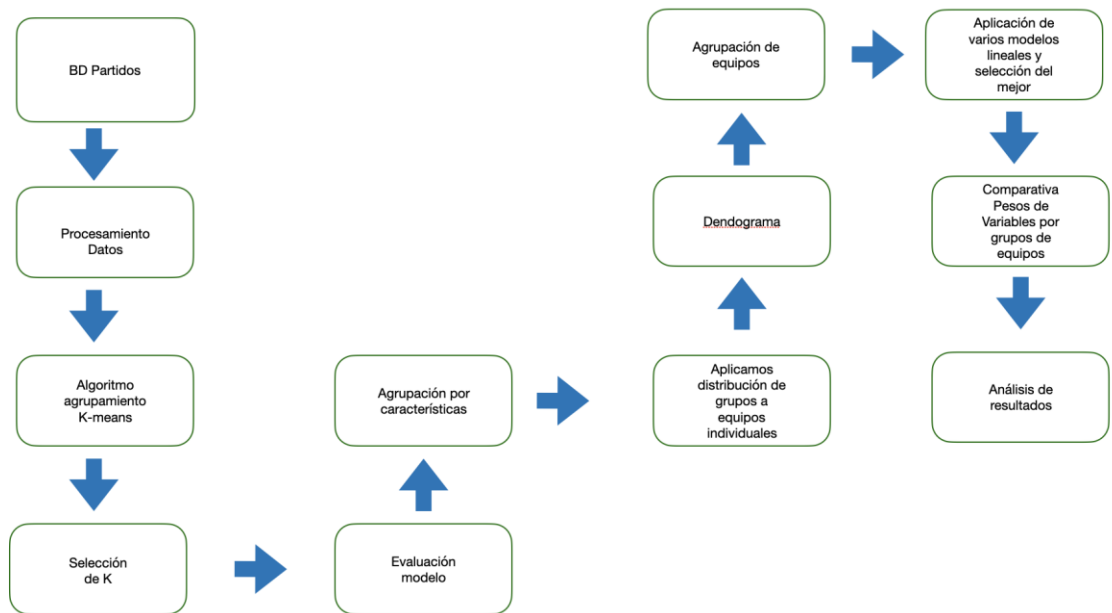


Ilustración 15: Metodología Empleada en clustering

4.2.1 Support Vector Machines (SVM)

Support Vector Machines, comúnmente conocidas como SVM, son un conjunto de algoritmos de aprendizaje supervisado utilizados para clasificación y regresión. Originadas de la teoría de aprendizaje estadístico, las SVM han demostrado ser eficientes en espacios de alta dimensionalidad y situaciones donde la división no es linealmente separable.

El núcleo de las SVM es encontrar un hiperplano que mejor separe las clases en el espacio de características. Matemáticamente, un hiperplano en un espacio (p) -dimensional es una subvariedad plana de dimensión $(p-1)$. El hiperplano se define generalmente por la ecuación:

$$\mathbf{w} \cdot \mathbf{x} + \mathbf{b} = 0$$

Donde (w) es el vector normal al hiperplano, (x) son los datos y (b) es el término de sesgo. Las SVM buscan maximizar el margen entre las clases. El margen es básicamente la distancia entre el hiperplano y el punto más cercano de cada clase. Para maximizar este margen, la SVM intenta minimizar $(\|w\|)$, sujeto a las restricciones de clasificación.

$$\text{Margen} = 2 / \|w\|$$

La belleza de las SVM radica en el uso de funciones kernel, que permiten operar en un espacio de características de alta dimensión sin calcular explícitamente las coordenadas de los datos en ese espacio, sino simplemente calculando el producto interno entre las imágenes de los datos. Este mecanismo es lo que permite a las SVM realizar clasificación no lineal.

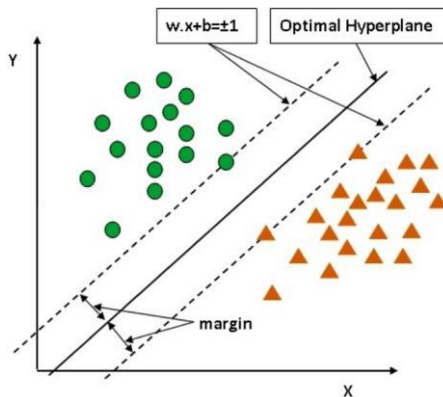


Ilustración 16: SVM

4.2.2 Regresión Logística

La regresión logística es un método estadístico para analizar conjuntos de datos en los que hay una o más variables independientes que determinan un resultado. El resultado se mide con una variable dicotómica (en la que solo hay dos resultados posibles). Es utilizada para predecir un resultado binario (1 / 0, Sí / No, Verdadero / Falso) dada una serie de variables independientes.

La regresión logística predice la probabilidad de ocurrencia de un evento binario utilizando una función logística.

$$P(Y=1|X) = e^{(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)} / (1 + e^{(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)})$$

Aquí, $P(Y=1|X)$ es la probabilidad de que el evento ocurra, e es la base del logaritmo natural, β_0 es el término de intersección, y $(\beta_1, \dots, \beta_p)$ son los coeficientes de las variables independientes (X_1, \dots, X_p) . Estos coeficientes indican la importancia de cada variable independiente en la predicción del resultado.

La regresión logística estima las probabilidades utilizando una función logística, que es una curva en forma de S. Por lo tanto, a medida que la suma ponderada de las entradas (o la puntuación) aumenta, la probabilidad resultante de la salida se acerca a 1, y a medida que disminuye, se acerca a 0.

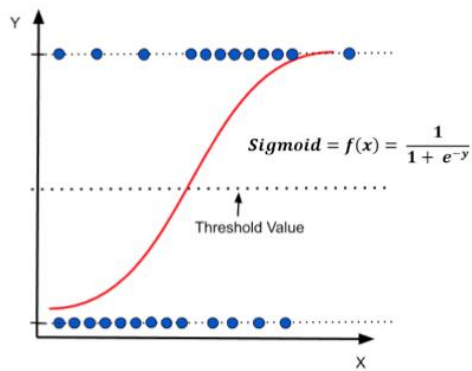


Ilustración 17: Regresión logística

4.2.3 Random Forest

Random Forest es un algoritmo de ensamblado que combina las decisiones de múltiples árboles de decisión para mejorar la precisión y evitar el sobreajuste. La idea principal detrás de Random Forest es combinar múltiples árboles de decisión en la determinación del resultado final en lugar de depender de árboles individuales.

Cada árbol se construye utilizando un subconjunto aleatorio de los datos y características. Esto introduce variabilidad en la construcción de cada árbol y hace que el modelo final sea más robusto. Cuando se realiza una predicción, la clase que obtenga la mayoría de votos de todos los árboles es la elegida.

Matemáticamente, si denotamos la función de predicción del k -ésimo árbol como $f_k(x)$, y tenemos un total de M árboles, la predicción final de Random Forest para una entrada x se calcula como:

$$\mathbf{RF}(x) = \text{mode}(f_1(x), f_2(x), \dots, f_M(x))$$

Aquí, la función 'mode' devuelve la clase que aparece con más frecuencia entre las predicciones de los árboles individuales.

Uno de los aspectos más importantes de Random Forest es su capacidad para calcular la importancia de las características. La importancia se mide observando cuánto disminuye la precisión del modelo al excluir cada característica.

Random Forest Classifier

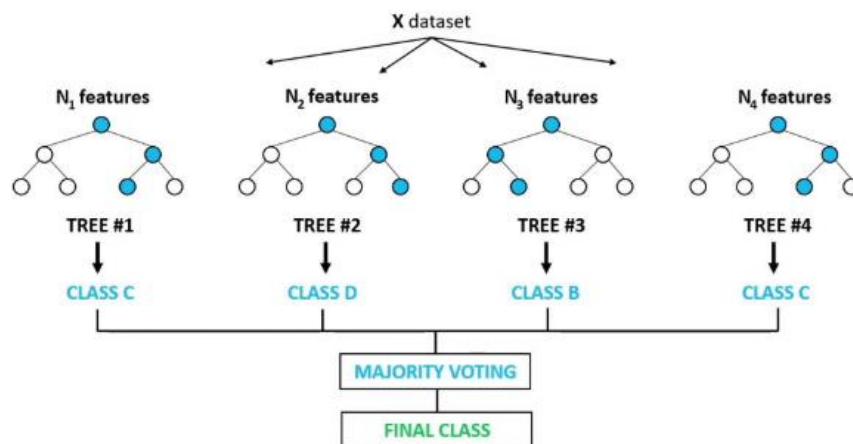


Ilustración 18: Random Forest

4.2.4 PCA (Análisis de Componentes Principales)

El Análisis de Componentes Principales, conocido como PCA por sus siglas en inglés, es una técnica de reducción de dimensionalidad que se utiliza frecuentemente en estadísticas

y aprendizaje automático. La idea principal detrás de PCA es transformar las variables originales en un nuevo conjunto de variables, los componentes principales, que son ortogonales (no correlacionados), y que reflejan la máxima varianza en los datos.

Matemáticamente, PCA busca la proyección que maximiza la varianza de los datos. Si (X) es la matriz de datos centrada (media 0), entonces la descomposición de valores singulares (SVD) de (X) se da como:

$$X = U \Sigma V^T$$

Aquí, (U) y (V) son matrices ortogonales y (Σ) es una matriz diagonal de valores singulares. Los componentes principales son las columnas de (V) . Estos componentes principales representan las direcciones de máxima varianza.

Una vez calculados los componentes principales, es posible representar los datos originales en este nuevo espacio de características reducido, lo que puede ser beneficioso tanto para la visualización como para la eficiencia computacional en algoritmos de aprendizaje automático.

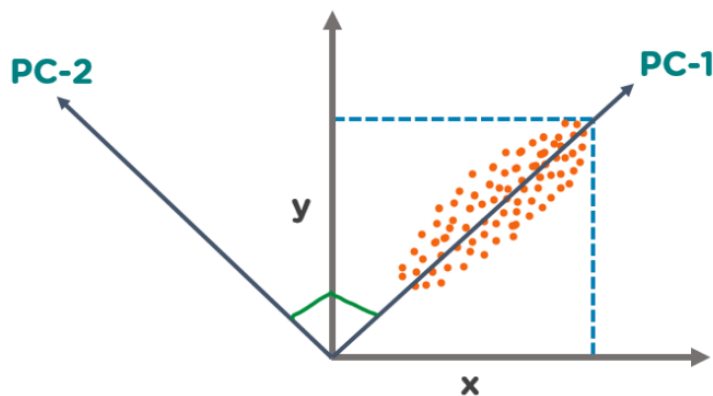


Ilustración 19: Análisis PCA

4.2.5 Distancia Euclidiana

La distancia euclidiana es una métrica que mide la distancia 'lineal' entre dos puntos en el espacio euclidiano. Es una de las métricas más utilizadas para medir similitudes y diferencias entre vectores en muchos campos, incluido el aprendizaje automático.

Matemáticamente, la distancia euclidiana entre dos puntos (A) y (B) en un espacio (N) -dimensional se define como la longitud del segmento que conecta los dos puntos.

Se calcula utilizando la fórmula:

$$d(A, B) = \sqrt{\sum_{i=1}^N (A_i - B_i)^2}$$

Aquí, (A_i) y (B_i) son las coordenadas del i -ésimo componente de los puntos (A) y (B) , respectivamente. La suma se extiende sobre todas las dimensiones del espacio.

La distancia euclidiana es intuitiva y tiene propiedades deseables, como ser siempre no negativa y ser cero solo cuando los dos puntos son idénticos.

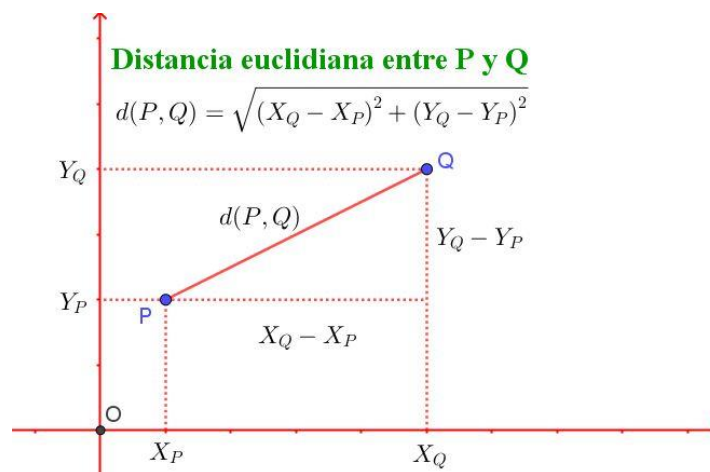


Ilustración 20: Distancia Euclidiana

4.2.6 Similitud del Coseno

La similitud del coseno es una métrica utilizada para determinar cuán similares son dos vectores. Esta técnica es ampliamente usada en sistemas de recomendación y otros campos donde se necesita medir la similitud entre dos conjuntos de datos.

Matemáticamente, la similitud del coseno entre dos vectores (A) y (B) se calcula como el coseno del ángulo entre ellos. La fórmula es la siguiente:

$$\text{similitud}(A, B) = (A \cdot B) / (\|A\| \|B\|)$$

Donde $(A \cdot B)$ es el producto punto de los vectores y $(\|A\|)$ y $(\|B\|)$ son las magnitudes (o normas) de los vectores (A) y (B) , respectivamente.

La similitud del coseno resultante estará en el rango $[-1, 1]$. Un valor de 1 indica que los vectores son idénticos, un valor de -1 indica que son totalmente disimilares y un valor de 0 indica ortogonalidad (no hay similitud). En la mayoría de los casos prácticos, los valores estarán en el rango $[0, 1]$, especialmente cuando los vectores representan cuentas o medidas no negativas.

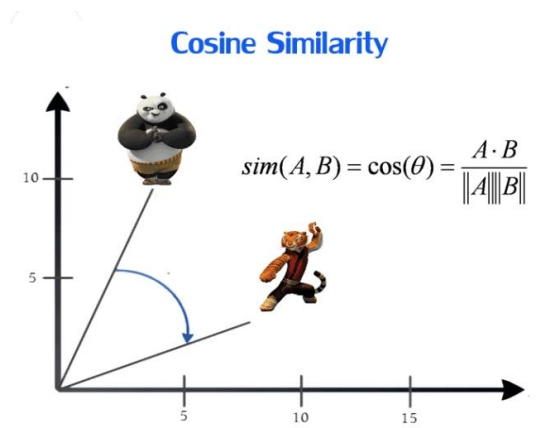


Ilustración 21: Similitud Coseno

4.2.7 Evaluación de Modelos

Para validar el rendimiento de los modelos propuestos, se tomarán en cuenta las siguientes métricas:

- Exactitud o Accuracy: Es la proporción de predicciones correctas en relación con el total de predicciones hechas. Se calcula como:

$$Accuracy = \frac{Numero\ Predicciones\ Correctas}{Numero\ Predicciones\ Totales}$$

No obstante, su uso no es recomendado en conjuntos de datos desbalanceados, pues un modelo podría tener un alto rendimiento prediciendo correctamente una clase dominante, pero fallar en predecir una clase minoritaria. Por ello, se determinará posteriormente si esta métrica es adecuada para el estudio.

- Precisión o Precision: Representa la exactitud de las predicciones positivas del modelo. Es la proporción de verdaderos positivos frente al total de casos positivos (verdaderos positivos y falsos positivos). Se calcula con la fórmula:

$$Precision = \frac{Verdaderos\ Positivos}{Verdaderos\ Positivos + Falsos\ Positivos}$$

Indica cómo el modelo distingue correctamente los casos positivos de todos los que predice como positivos.

- Exhaustividad o Recall: Esta métrica cuantifica cuántos de los casos positivos reales fueron identificados correctamente por el modelo. Se define como la proporción de

verdaderos positivos respecto al total de casos positivos reales (verdaderos positivos y falsos negativos). Su fórmula es:

$$Recall = \frac{Verdaderos\ Positivos}{Verdaderos\ Positivos + Falsos\ Negativos}$$

Muestra la habilidad del modelo para detectar y clasificar correctamente todos los casos positivos presentes en el conjunto de datos.

Cada métrica proporciona una visión distinta del rendimiento del modelo, y dependiendo de la naturaleza del problema y los objetivos del estudio, algunas pueden ser más relevantes que otras.

5. Resultados

5.1. Métricas de los Modelos

Tras la limpieza de los datos, se desarrollaron diversos modelos analíticos con el objetivo de predecir victorias o derrotas en las distintas ligas de la Conmebol. Los modelos implementados en cada país incluyen el Support Vector Machine (SVM), la Regresión Logística, Random Forest y XGBoost. Estos modelos no solo tuvieron en cuenta la normalización de los datos, sino también el equilibrio entre partidos ganados y perdidos en cada liga, asegurando así una representación correcta. Además, se aplicaron técnicas de muestreo aleatorio para recopilar los datos, minimizando cualquier sesgo potencial y

garantizando la fiabilidad de los resultados obtenidos.

Pais	Model	Accuracy	Precision	Recall	F1 Score
Argentina	Logistic Regression	0.888	0.890	0.888	0.888
Argentina	XGBoost	0.872	0.872	0.872	0.871
Argentina	Linear SVM	0.866	0.866	0.866	0.866
Argentina	Random Forest	0.844	0.846	0.844	0.844
Bolivia	Linear SVM	0.917	0.917	0.917	0.917
Bolivia	Logistic Regression	0.912	0.913	0.912	0.912
Bolivia	XGBoost	0.892	0.894	0.892	0.892
Bolivia	Random Forest	0.882	0.889	0.882	0.883
Brasil	Logistic Regression	0.923	0.923	0.923	0.923
Brasil	Linear SVM	0.916	0.917	0.916	0.916
Brasil	XGBoost	0.903	0.904	0.903	0.903
Brasil	Random Forest	0.900	0.900	0.900	0.900
Chile	XGBoost	0.918	0.919	0.918	0.918
Chile	Linear SVM	0.909	0.909	0.909	0.909
Chile	Logistic Regression	0.904	0.904	0.904	0.904
Chile	Random Forest	0.875	0.877	0.875	0.875
Colombia	Random Forest	0.912	0.914	0.912	0.912
Colombia	Logistic Regression	0.895	0.896	0.895	0.895
Colombia	Linear SVM	0.887	0.888	0.887	0.887
Colombia	XGBoost	0.878	0.878	0.878	0.878
Ecuador	Logistic Regression	0.932	0.932	0.932	0.932
Ecuador	Linear SVM	0.918	0.918	0.918	0.918
Ecuador	Random Forest	0.890	0.894	0.890	0.891
Ecuador	XGBoost	0.890	0.892	0.890	0.891
Paraguay	XGBoost	0.698	0.702	0.698	0.700
Paraguay	Linear SVM	0.688	0.692	0.688	0.689
Paraguay	Logistic Regression	0.682	0.682	0.682	0.679
Paraguay	Random Forest	0.672	0.674	0.672	0.673
Peru	XGBoost	0.938	0.940	0.938	0.938
Peru	Logistic Regression	0.934	0.934	0.934	0.934
Peru	Linear SVM	0.921	0.925	0.921	0.921
Peru	Random Forest	0.907	0.912	0.907	0.908
Uruguay	Random Forest	0.901	0.901	0.901	0.901
Uruguay	Logistic Regression	0.895	0.897	0.895	0.895
Uruguay	Linear SVM	0.890	0.892	0.890	0.890
Uruguay	XGBoost	0.884	0.886	0.884	0.884
Venezuela	Random Forest	0.932	0.932	0.932	0.932
Venezuela	XGBoost	0.889	0.889	0.889	0.889
Venezuela	Logistic Regression	0.877	0.879	0.877	0.876
Venezuela	Linear SVM	0.870	0.873	0.870	0.870

Tabla 3: Comparación de Modelos

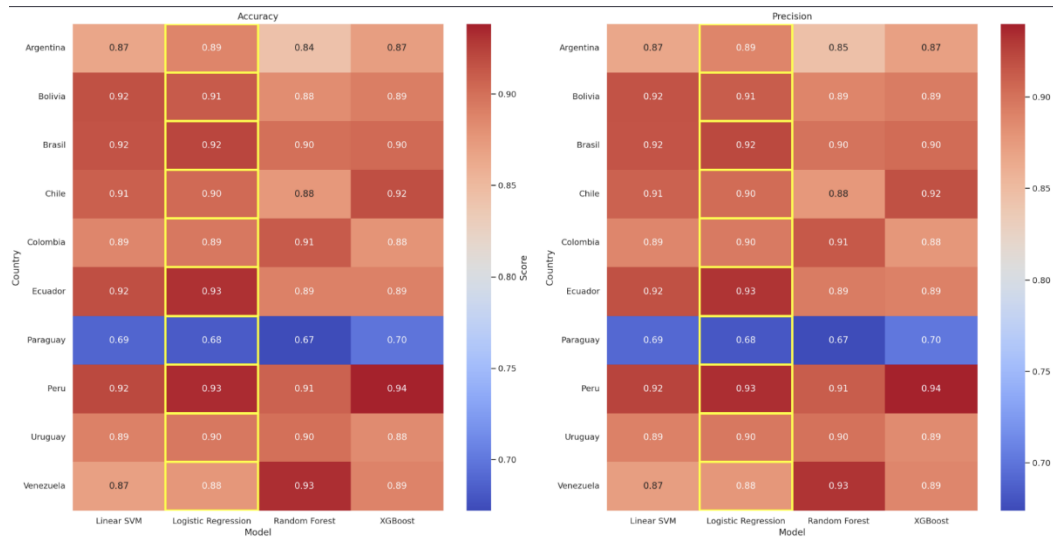


Ilustración 22: Accuracy y Precision

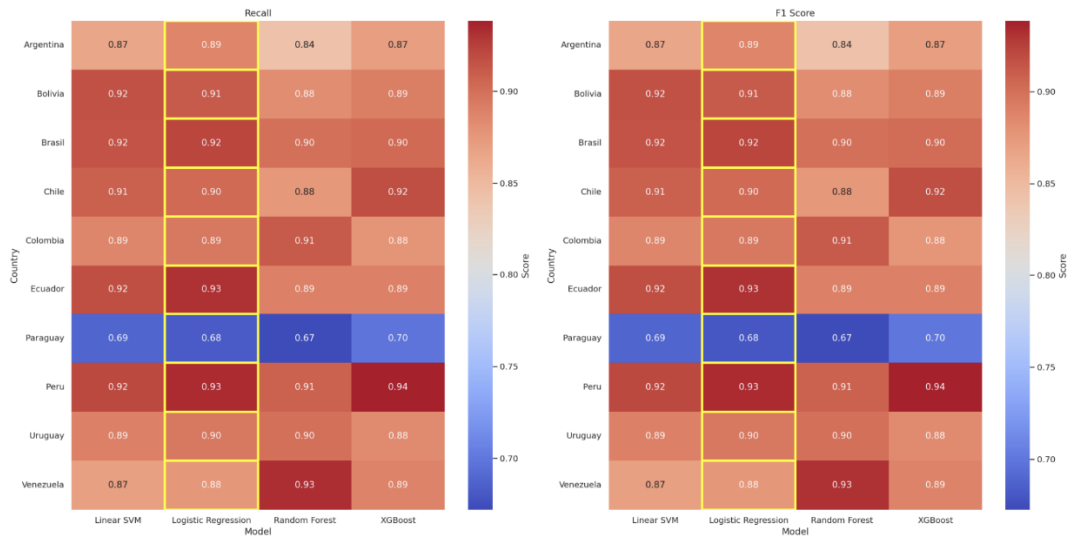


Ilustración 23: Recall y F1 Score

El modelo de Regresión Logística fue seleccionado debido a su buen rendimiento en términos de métricas de evaluación en la mayoría de los países. Una de las principales ventajas de la Regresión Logística es su capacidad para identificar claramente el peso o importancia de cada variable en la predicción. Esto permitió discernir cuáles características tienen el mayor impacto positivo o negativo en el modelo de cada país.

5.2. Análisis de los pesos de las variables por País

A continuación, presentamos un resumen detallado que destaca las cinco variables con mayor y menor influencia en el modelo, proporcionando así una visión integral de los factores clave que afectan los resultados de las predicciones.

Argentina

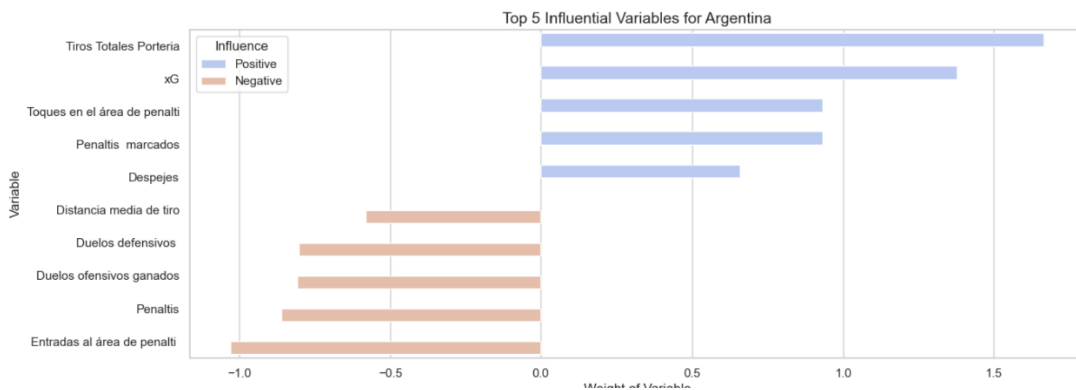


Ilustración 24: Top 5 Argentina

Análisis de las Variables Positivas:

Tiros Totales a Portería: Indica un equipo ofensivo que crea muchas oportunidades de gol. Un alto número de tiros a portería sugiere una estrategia de ataque directo y presión constante en el área rival.

xG (Expected Goals): Representa la calidad de las oportunidades de gol creadas. Un alto xG sugiere que el equipo no solo ataca frecuentemente, sino que también lo hace con eficacia, creando oportunidades claras de gol.

Toques en el Área de Penalti: Refleja un equipo que logra penetrar las defensas rivales y mantener la posesión en zonas peligrosas. Esto puede ser indicativo de un buen juego de pases y movimiento sin balón.

Penalti Marcado: Un equipo con muchos penaltis marcados puede indicar una habilidad para provocar faltas en el área rival, lo que a menudo resulta de una agresiva ofensiva en el área.

Despeje: La habilidad para despejar balones sugiere una defensa sólida y efectiva en situaciones de presión, lo que es crucial para mantener un resultado favorable.

Análisis de las Variables Negativas:

Duelos Defensivos Ganados: Si esta variable afecta negativamente, puede indicar debilidades en la defensa, especialmente en situaciones de uno contra uno.

Penaltis: Los penaltis concedidos pueden ser un signo de indisciplina o tácticas defensivas arriesgadas.

Entradas al Área de Penalti: Si afecta negativamente, podría indicar problemas en controlar las incursiones del equipo rival en áreas peligrosas.

Goles Recibidos: Naturalmente, recibir muchos goles es un indicativo de problemas defensivos, ya sea en la línea defensiva o en el portero.

Estilo de Juego General:

Con estas variables, el estilo de juego parece ser muy ofensivo y basado en la creación constante de oportunidades de gol. Este equipo probablemente se destaque en el ataque, con jugadores habilidosos en el área rival y una táctica que favorece el juego en el tercio ofensivo del campo. Sin embargo, hay una preocupación notable en la defensa, especialmente en la efectividad de los duelos defensivos y en la tendencia a conceder penaltis y goles. Esto sugiere un equipo que podría adoptar un enfoque de "mejor ataque que defensa", confiando en superar a sus oponentes en lugar de mantener una sólida línea defensiva.

Este análisis resalta la importancia del equilibrio entre ataque y defensa. Mientras que un enfoque ofensivo puede resultar en muchos goles marcados, la vulnerabilidad defensiva podría ser costosa, especialmente contra equipos que son efectivos en el contraataque o en situaciones de juego estático.

Bolivia

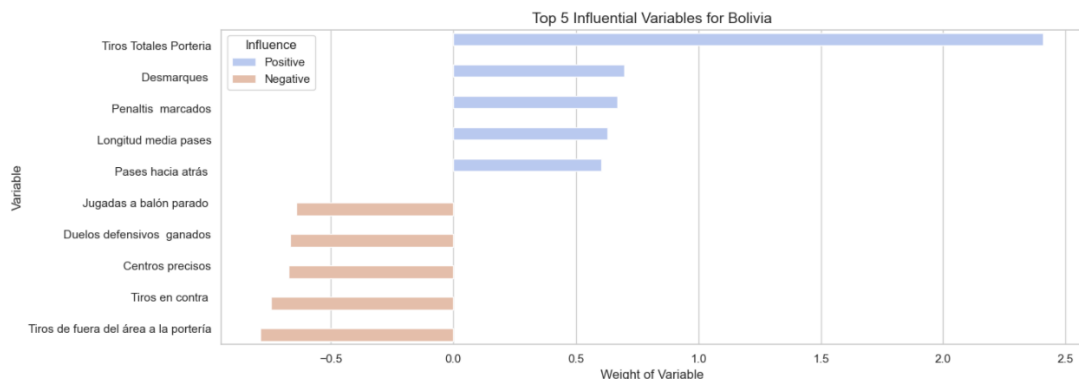


Ilustración 25: Top 5 Bolivia

Análisis de las Variables Positivas:

Tiros Totales a Portería: Indica un enfoque ofensivo, con un equipo que genera muchas oportunidades de gol. La capacidad para rematar a portería frecuentemente sugiere una presión ofensiva constante.

Desmarques: Refleja la movilidad y la inteligencia táctica de los jugadores, especialmente en el ataque. Un alto número de desmarques exitosos sugiere un juego dinámico y la capacidad de encontrar espacios en la defensa rival.

Penaltis Marcados: Implica eficacia en situaciones de penalti, lo cual puede ser un indicador de habilidad para forzar faltas en el área o de contar con buenos ejecutores de penaltis.

Longitud Media de Pases: Una mayor longitud media en los pases puede indicar un estilo de juego que favorece los pases largos, posiblemente buscando cambiar rápidamente el juego o encontrar a jugadores desmarcados.

Pases hacia Atrás: Un alto número de pases hacia atrás sugiere un estilo de juego que prioriza la posesión y la reconstrucción cuidadosa de las jugadas desde la defensa o el mediocampo.

Análisis de las Variables Negativas:

Duelos Defensivos Ganados: Si esta métrica es baja, indica debilidades en la defensa, especialmente en situaciones de uno contra uno.

Centros Precisos: La falta de precisión en los centros puede apuntar a una debilidad en la creación de oportunidades de gol desde las bandas.

Tiros en Contra: Un alto número de tiros en contra sugiere que el equipo permite a los oponentes crear oportunidades de gol, posiblemente debido a debilidades defensivas o a una táctica de alto riesgo.

Tiros Fuera del Área a la Portería: La capacidad del equipo rival para realizar tiros desde fuera del área indica posiblemente una falta de presión defensiva en la media cancha.

Goles Recibidos: Como en el análisis anterior, un alto número de goles recibidos es claramente indicativo de problemas defensivos.

Estilo de Juego General:

Este conjunto de variables sugiere un equipo con un enfoque ofensivo fuerte, que favorece el juego directo y los pases largos, junto con una movilidad táctica en el ataque. La habilidad para ejecutar desmarques y generar oportunidades a través de penaltis marcados indica un equipo agresivo en el área rival.

En contraste, hay claras debilidades defensivas. La falta de efectividad en duelos defensivos y la tendencia a recibir tiros y goles sugiere que mientras el equipo es fuerte en el ataque, puede ser vulnerable a contraataques y a equipos que sepan aprovechar espacios en la media y defensa. Además, la falta de precisión en los centros puede limitar las opciones ofensivas por las bandas.

En resumen, este equipo parece tener un estilo de juego enfocado en un ataque potente y directo, posiblemente sacrificando el equilibrio defensivo. La efectividad en el ataque

puede compensar las debilidades defensivas, pero contra equipos tácticamente astutos y defensivamente sólidos, estos puntos débiles podrían ser explotados.

Brasil

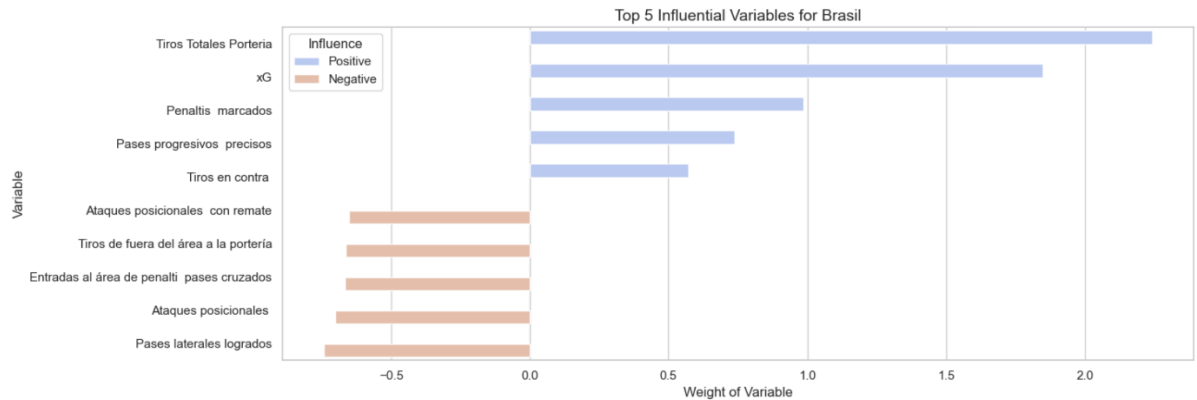


Ilustración 26: Top 5 Brasil

Análisis de las Variables Positivas:

Tiros Totales a Portería: Un alto número sugiere un equipo muy ofensivo que crea muchas oportunidades de gol, indicando una presión constante en el área rival.

xG (Expected Goals): Un alto xG implica que el equipo no solo crea muchas oportunidades, sino que también las crea de alta calidad, aumentando la probabilidad de marcar goles.

Penaltis Marcados: Indica una capacidad para forzar faltas en el área rival y contar con buenos ejecutores de penaltis.

Pases Progresivos Precisos: Refleja un equipo que es eficiente en avanzar el balón hacia adelante, indicando un juego de posesión con intención ofensiva.

Tiros en Contra: Si esta variable es positiva, sugiere que el equipo puede estar jugando con una táctica de alto riesgo, permitiendo tiros en contra pero confiando en su capacidad ofensiva para compensar.

Análisis de las Variables Negativas:

Tiros de Fuera del Área a la Portería: La ineficacia en esta área puede indicar una falta de amenaza desde larga distancia o una preferencia por construir jugadas más cercanas al área.

Entradas al Área de Penalti Pases Cruzados: Una baja eficacia aquí sugiere dificultades en crear oportunidades a través de pases cruzados en el área, posiblemente debido a la defensa rival o a una falta de precisión en los centros.

Ataques Posicionales: Si esto afecta negativamente, podría indicar que el equipo tiene problemas para mantener ataques sostenidos o para penetrar defensas organizadas.

Pases Laterales Logrados: Una baja en esta métrica puede reflejar una falta de amplitud en el juego o una preferencia por el juego vertical directo en lugar de la distribución lateral del balón.

Goles Recibidos: Como antes, muchos goles recibidos son indicativos de problemas defensivos.

Estilo de Juego General:

Este equipo parece centrarse en un juego ofensivo potente, creando oportunidades de gol de alta calidad y avanzando el balón de manera efectiva hacia el área rival. La habilidad para convertir penaltis y crear tiros a portería sugiere un ataque formidable.

Sin embargo, hay debilidades claras en otros aspectos del juego, especialmente en la generación de oportunidades a través de pases cruzados y ataques posicionales sostenidos. La falta de amenaza desde fuera del área y la limitada utilización de pases laterales sugieren una tendencia a centrarse en el juego directo y vertical. Además, la defensa parece ser un punto débil, dada la alta cantidad de goles recibidos.

En resumen, este equipo se caracteriza por su fortaleza en ataque directo y la creación de oportunidades de gol, pero necesita mejorar en la distribución del juego y fortalecer su línea defensiva para ser más equilibrado y menos vulnerable a los contraataques.

Chile

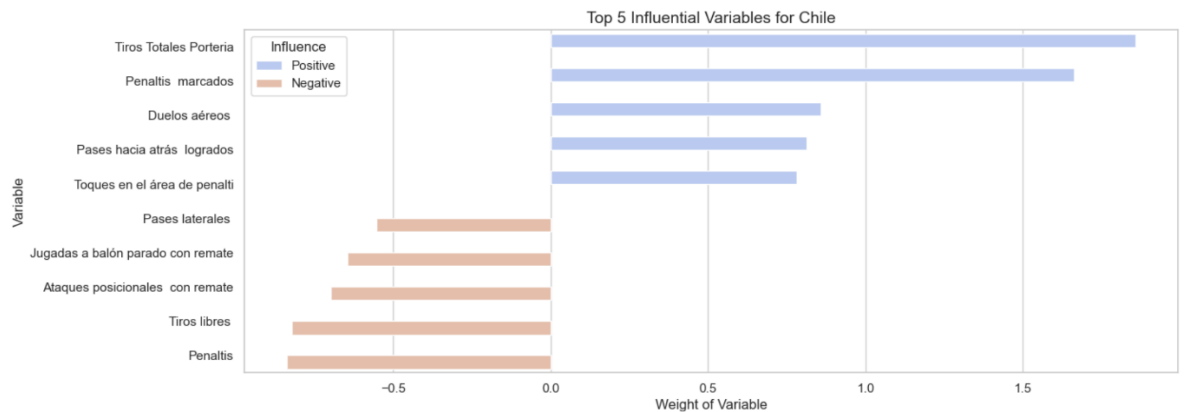


Ilustración 27: Top 5 Chile

Análisis de las Variables Positivas:

Tiros Totales a Portería: Esta métrica alta indica un equipo que genera muchas oportunidades de gol, destacando su eficacia ofensiva y la capacidad de presionar constantemente en el área rival.

xG (Expected Goals): Un alto xG sugiere que las oportunidades de gol creadas son de alta calidad, lo que aumenta la probabilidad de convertir estas oportunidades en goles.

Intensidad de Paso: Esta variable sugiere un ritmo de juego rápido, con una transición ágil del balón, lo que puede ser indicativo de un estilo de juego dinámico y proactivo.

Despejes: Un alto número de despejes refleja una defensa sólida y reactiva, capaz de lidiar con amenazas en su propia área.

Pases Cruzados en Profundidad Completados: La eficiencia en este tipo de pases indica habilidad para crear oportunidades de gol a través de jugadas elaboradas y pases precisos en profundidad.

Análisis de las Variables Negativas:

Tiros Libres: Si esta variable afecta negativamente, sugiere una falta de eficacia en situaciones de balón parado, lo cual puede ser una desventaja en juegos ajustados.

Ataques Posicionales con Remate: Una influencia negativa aquí puede indicar dificultades en finalizar jugadas ofensivas organizadas con un remate efectivo.

Entradas al Área de Penalti Pases Cruzados: Si esta métrica es negativa, sugiere problemas en la creación de oportunidades a través de pases cruzados, lo que puede limitar la variedad ofensiva del equipo.

Tiros en Contra a la Portería: Un alto número indica que el equipo permite a los oponentes crear oportunidades de gol claras, posiblemente debido a debilidades defensivas o a una táctica de alto riesgo.

Goles Recibidos: Como en análisis anteriores, muchos goles recibidos son claramente indicativos de problemas defensivos.

Estilo de Juego General:

Este conjunto de variables sugiere un equipo con un enfoque ofensivo fuerte, capaz de crear y aprovechar oportunidades de gol de alta calidad, especialmente a través de un juego rápido y pases en profundidad. La habilidad para realizar despejes efectivos indica una defensa sólida en situaciones críticas.

Sin embargo, hay claras áreas de mejora. La falta de eficacia en tiros libres y la dificultad en finalizar ataques posicionales con remates exitosos pueden limitar las oportunidades de gol. Además, la vulnerabilidad a los tiros en contra a la portería y la cantidad de goles recibidos sugieren una defensa que puede ser superada, especialmente en situaciones de contraataque.

En resumen, el equipo muestra un estilo de juego dinámico y ofensivo, con énfasis en la creación de oportunidades a través de pases en profundidad y un juego rápido. Sin embargo, debe mejorar en situaciones de balón parado y fortalecer su defensa para evitar ser vulnerable a los ataques de los oponentes.

Colombia

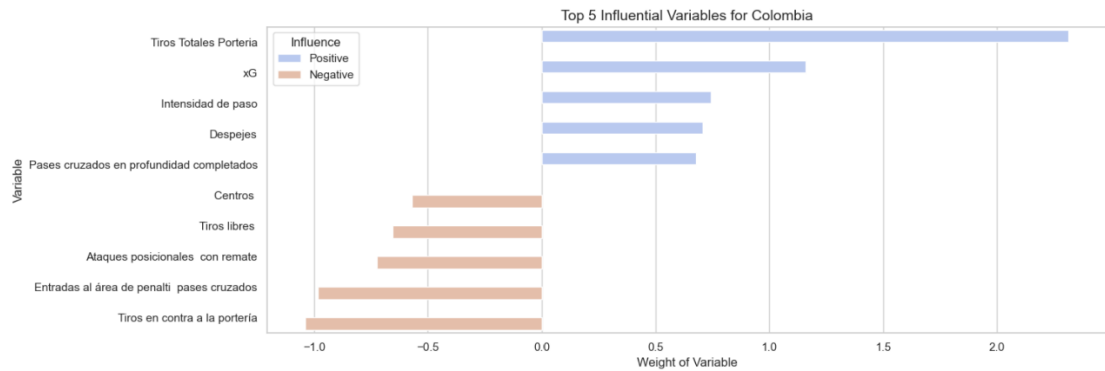


Ilustración 28: Top 5 Colombia

Análisis de las Variables Positivas:

Tiros Totales a Portería: Una alta cantidad de tiros a portería indica un equipo ofensivo, con una fuerte presencia en el área rival y una tendencia a crear muchas oportunidades de gol.

xG (Expected Goals): Un xG elevado sugiere que el equipo crea oportunidades de gol de alta calidad, lo que implica una mayor probabilidad de convertir estas oportunidades en goles.

Intensidad de Paso: Esta métrica refleja un juego rápido y dinámico, con transiciones ágiles y un enfoque en el movimiento rápido del balón, lo que puede indicar un estilo de juego proactivo y enfocado en la ofensiva.

Despejes: Un alto número de despejes sugiere una defensa sólida, capaz de lidiar eficazmente con las amenazas en su propia área.

Pases Cruzados en Profundidad Completados: La eficacia en estos pases indica una habilidad para crear oportunidades de gol mediante jugadas elaboradas y pases precisos, especialmente en situaciones de juego abierto.

Análisis de las Variables Negativas:

Tiros Libres: Una baja eficacia en tiros libres puede ser una desventaja, especialmente en partidos ajustados donde estas situaciones pueden ser cruciales.

Ataques Posicionales con Remate: Si esta métrica afecta negativamente, sugiere problemas en la finalización efectiva de jugadas ofensivas organizadas.

Entradas al Área de Penalti Pases Cruzados: Una baja efectividad aquí indica dificultades en crear oportunidades a través de pases cruzados dentro del área, lo que puede limitar la variedad ofensiva.

Tiros en Contra a la Portería: Un alto número sugiere que el equipo permite oportunidades claras de gol al rival, posiblemente debido a debilidades defensivas o a tácticas de alto riesgo.

Goles Recibidos: Una alta cantidad de goles recibidos es un claro indicativo de problemas en la defensa.

Estilo de Juego General:

Este equipo parece tener un enfoque ofensivo fuerte, destacándose en la creación y aprovechamiento de oportunidades de gol de calidad. La intensidad en el paso del juego y la habilidad para ejecutar pases en profundidad apuntan a un estilo de juego dinámico y ofensivo.

No obstante, hay áreas claras de mejora. La falta de eficacia en tiros libres y en la finalización de ataques posicionales sugiere que podrían estar perdiendo oportunidades cruciales de gol. Además, la vulnerabilidad a los tiros en contra y la cantidad de goles

recibidos indican una defensa que puede ser propensa a errores o superada por tácticas de contraataque.

En resumen, el equipo muestra un estilo de juego dinámico, centrado en la ofensiva con un enfoque en la creación de oportunidades a través de pases en profundidad y un juego rápido. Sin embargo, para ser más equilibrado y efectivo, necesita mejorar en situaciones de balón parado y fortalecer su defensa.

Ecuador

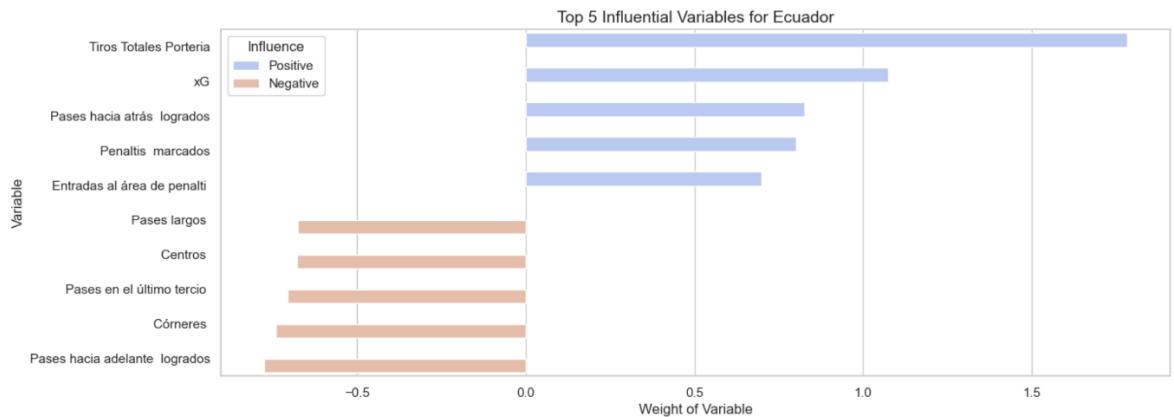


Ilustración 29: Top 5 Ecuador

Análisis de las Variables Positivas:

Tiros Totales a Portería: Esta métrica alta sugiere un equipo muy ofensivo que crea muchas oportunidades de gol, mostrando una presión constante en el área rival.

xG (Expected Goals): Un alto xG indica que el equipo no solo crea oportunidades de gol, sino que estas son de alta calidad, lo que aumenta la probabilidad de convertir.

Pases hacia Atrás Logrados: Este aspecto sugiere un estilo de juego que prioriza la posesión y la construcción cuidadosa del juego, posiblemente para organizar ataques de forma metódica.

Penaltis Marcados: Implica una habilidad para provocar faltas en el área rival y tener efectividad en la ejecución de penaltis.

Entradas al Área de Penalti: Indica un equipo que logra penetrar en las defensas rivales y crear oportunidades de gol desde posiciones peligrosas.

Análisis de las Variables Negativas:

Centros: Una baja efectividad en los centros puede indicar una debilidad en la creación de oportunidades de gol desde las bandas.

Pases en el Último Tercio: Si esta variable es negativa, sugiere dificultades en mantener la posesión y crear oportunidades en la zona más crítica del campo.

Córneres: Una baja eficacia en córneres podría indicar problemas en aprovechar situaciones de balón parado.

Pases hacia Adelante Logrados: Si esta métrica es baja, sugiere una falta de eficacia en avanzar el balón hacia posiciones ofensivas.

Goles Recibidos: Como en análisis anteriores, una alta cantidad de goles recibidos es indicativo de problemas defensivos.

Estilo de Juego General:

Este equipo parece centrarse en un estilo de juego ofensivo, destacando en la creación de oportunidades de gol, especialmente a través de jugadas dentro del área rival y ejecución

de penaltis. La tendencia a realizar pases hacia atrás exitosos sugiere un enfoque en la posesión y en la construcción cuidadosa de ataques.

Sin embargo, el equipo muestra debilidades en la ejecución de centros, mantenimiento de la posesión en el último tercio y en la eficacia de pases hacia adelante, lo que podría limitar su versatilidad ofensiva. Además, la defensa parece ser un punto débil, dado el alto número de goles recibidos.

En conclusión, este equipo parece adoptar un enfoque ofensivo, privilegiando la creación de oportunidades de gol a través de la posesión y penetración en el área rival, pero necesita mejorar en la explotación de situaciones de balón parado y fortalecer su defensa para evitar ser vulnerable a contraataques o equipos que presionan alto.

Paraguay

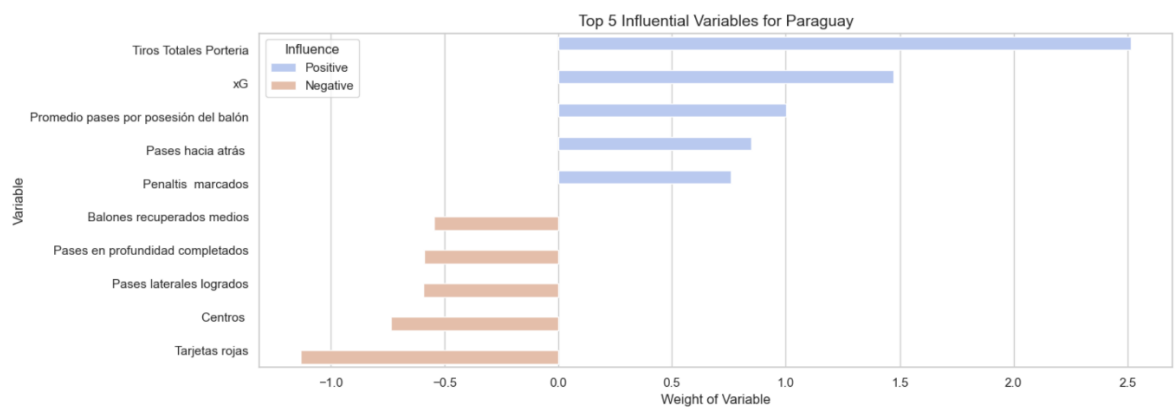


Ilustración 30: Top 5 Paraguay

Análisis de las Variables Positivas:

Tiros Totales a Portería: Un alto número indica un equipo con un enfoque ofensivo fuerte, capaz de crear y ejecutar muchas oportunidades de gol.

xG (Expected Goals): Un xG elevado sugiere que el equipo no solo crea oportunidades de gol, sino que lo hace con alta calidad, incrementando las probabilidades de marcar.

Promedio de Pases por Posesión del Balón: Esto implica un juego basado en la posesión, donde el equipo mantiene el control del balón y construye sus ataques de manera metódica.

Pases hacia Atrás: La tendencia a realizar pases hacia atrás sugiere un enfoque en mantener la posesión y posiblemente reorganizar el juego para crear oportunidades de ataque.

Penaltis Marcados: La habilidad para marcar penaltis indica una ventaja en situaciones de juego detenido y puede reflejar la habilidad de provocar faltas en el área rival.

Análisis de las Variables Negativas:

Pases en Profundidad Completados: Una baja eficacia en esta área puede indicar dificultades en penetrar las defensas rivales con pases directos.

Pases Laterales Logrados: Si esta variable es negativa, sugiere una falta de efectividad en la distribución del juego a través de pases laterales, lo que podría limitar la amplitud en el ataque.

Centros: Una baja eficacia en los centros puede indicar problemas en la creación de oportunidades de gol desde las bandas.

Tarjetas Rojas: Un alto número de tarjetas rojas puede ser indicativo de indisciplina o tácticas defensivas agresivas que resultan en expulsiones.

Goles Recibidos: Como en análisis anteriores, una alta cantidad de goles en contra sugiere debilidades defensivas.

Estilo de Juego General:

Este equipo parece centrarse en un estilo de juego ofensivo, con énfasis en la creación de oportunidades de gol a través de un juego controlado y posesión del balón. La tendencia a realizar pases hacia atrás y un alto promedio de pases por posesión del balón indica un enfoque en la construcción cuidadosa de las jugadas y en mantener el control del juego.

Sin embargo, hay claras debilidades en otros aspectos del juego, como la dificultad en ejecutar pases en profundidad y en la efectividad de los centros, lo que puede limitar la variedad ofensiva. Además, la indisciplina indicada por las tarjetas rojas y las deficiencias defensivas, como se refleja en los goles recibidos, sugieren áreas que necesitan mejoras significativas.

En resumen, el equipo muestra un enfoque ofensivo basado en la posesión y la creación metódica de oportunidades, pero necesita mejorar en la explotación de situaciones de balón parado y fortalecer su defensa para ser más equilibrado y efectivo globalmente.

Peru

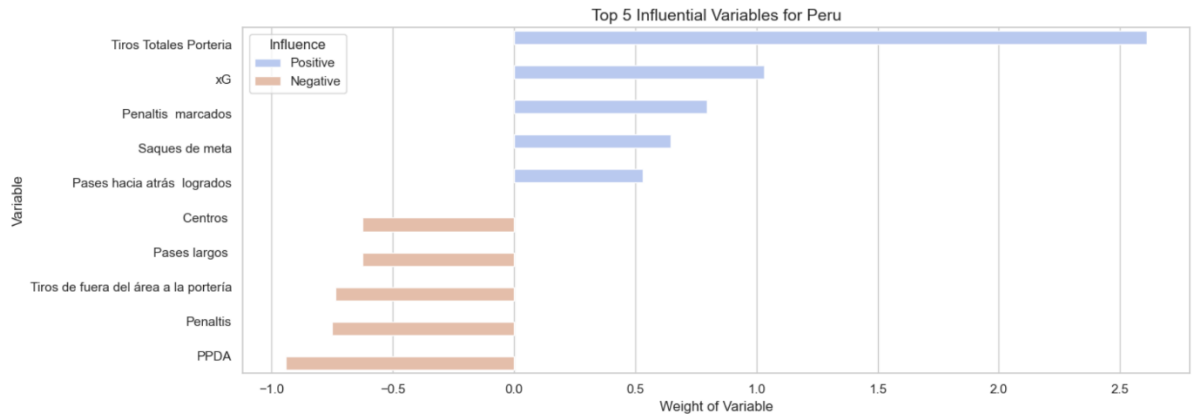


Ilustración 31: Top 5 Peru

Análisis de las Variables Positivas:

Tiros Totales a Portería: Una alta cantidad indica un equipo ofensivamente fuerte, capaz de crear muchas oportunidades de gol y ejercer presión en el área rival.

xG (Expected Goals): Un alto xG sugiere que las oportunidades de gol creadas son de alta calidad, lo que aumenta la probabilidad de marcar.

Penaltis Marcados: La eficacia en la ejecución de penaltis puede reflejar habilidad para forzar faltas en el área rival y contar con buenos ejecutores.

Saques de Meta: Una influencia positiva de los saques de meta puede indicar una defensa sólida que logra repeler los ataques y reiniciar el juego de forma efectiva.

Pases hacia Atrás Logrados: Esto sugiere un estilo de juego que prioriza la posesión y una reconstrucción cuidadosa del juego desde atrás, posiblemente para organizar ataques de manera metódica.

Análisis de las Variables Negativas:

Pases Largos: La ineficacia en pases largos puede indicar una limitación en cambiar rápidamente el enfoque del juego o en encontrar jugadores en posiciones avanzadas.

Tiros de Fuera del Área a la Portería: Una baja eficacia aquí sugiere una falta de amenaza desde distancias más largas.

Penaltis (Concedidos): Si esta variable es negativa, indica una tendencia a cometer faltas en el área propia, lo que resulta en oportunidades de gol para el equipo contrario.

PPDA (Pases Permitidos por Acción Defensiva): Un bajo PPDA sugiere que el equipo no ejerce una alta presión defensiva, permitiendo al equipo rival pasar el balón con más facilidad.

Goles Recibidos: Una alta cantidad de goles en contra indica debilidades en la defensa.

Estilo de Juego General:

Este equipo parece tener un fuerte enfoque ofensivo, destacándose en la creación y aprovechamiento de oportunidades de gol, especialmente a través de acciones dentro del área rival. La tendencia a realizar pases hacia atrás exitosos y los eficientes saques de meta sugieren un enfoque en mantener la posesión y organizar el juego desde la defensa.

Sin embargo, el equipo muestra debilidades significativas en defensa, como lo indican los altos goles recibidos y la tendencia a conceder penaltis. La ineficacia en los pases largos y la falta de amenaza en tiros desde fuera del área podrían limitar su capacidad para variar el ataque y sorprender al equipo contrario. Además, un bajo PPDA sugiere una estrategia defensiva menos agresiva, lo que podría permitir a los oponentes construir sus ataques con mayor facilidad.

En conclusión, el equipo muestra un estilo de juego ofensivo y orientado a la posesión, pero necesita mejorar en la variedad de su ataque y fortalecer significativamente su defensa para evitar ser vulnerable, especialmente contra equipos que son efectivos en el contraataque o en situaciones de juego estático.

Uruguay

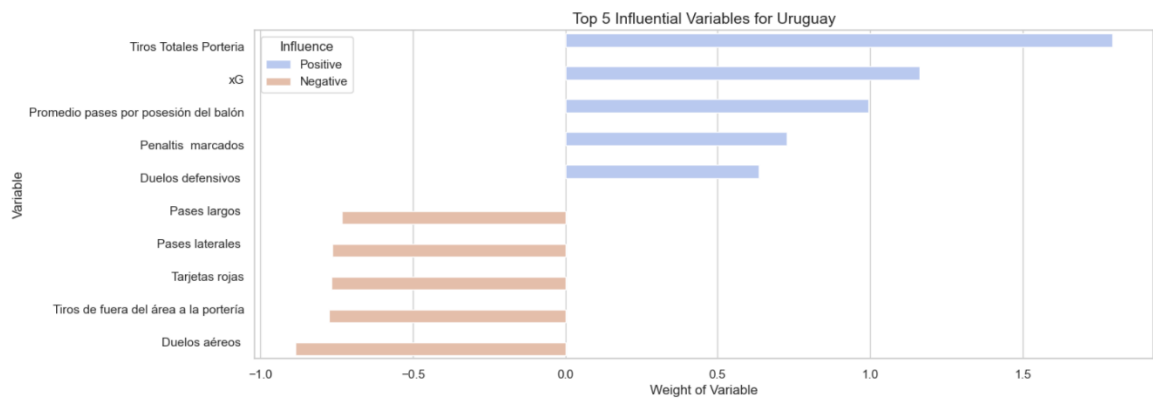


Ilustración 32: Top 5 Uruguay

Análisis de las Variables Positivas:

Tiros Totales a Portería: Un alto número indica una fuerte presencia ofensiva, con muchas oportunidades de gol creadas y una presión constante en el área rival.

xG (Expected Goals): Un alto xG sugiere que el equipo no solo crea oportunidades de gol, sino que estas son de alta calidad, aumentando la probabilidad de marcar.

Promedio de Pases por Posesión del Balón: Esto implica un juego basado en la posesión, donde el equipo controla el balón y construye sus ataques de manera metódica.

Penaltis Marcados: La habilidad para marcar penaltis indica efectividad en situaciones de juego detenido y puede reflejar la habilidad de provocar faltas en el área rival.

Duelos Defensivos: Una alta eficacia en duelos defensivos sugiere una defensa sólida, capaz de recuperar el balón y detener ataques rivales.

Análisis de las Variables Negativas:

Pases Laterales: Si esta variable es negativa, puede indicar una falta de amplitud en el juego o una preferencia por el juego vertical en lugar de la distribución lateral del balón.

Tarjetas Rojas: Un alto número de tarjetas rojas puede ser indicativo de indisciplina o tácticas defensivas agresivas que resultan en expulsiones.

Tiros de Fuera del Área a la Portería: Una baja eficacia en esta área sugiere una falta de amenaza desde distancias más largas.

Duelos Aéreos: Si esta variable es negativa, indica debilidades en el juego aéreo, lo que puede ser un problema en situaciones defensivas y ofensivas.

Goles Recibidos: Como en análisis anteriores, una alta cantidad de goles en contra sugiere debilidades en la defensa.

Estilo de Juego General:

Este equipo parece tener un enfoque ofensivo fuerte, destacando en la creación de oportunidades de gol, especialmente a través de un juego controlado y posesión del balón.

La habilidad para ejecutar penaltis y una alta eficacia en duelos defensivos sugiere un enfoque equilibrado entre ataque y defensa.

Sin embargo, hay claras debilidades en otros aspectos del juego, como la limitada efectividad en pases laterales, que podría restringir la amplitud en el ataque, y en el juego aéreo. La indisciplina indicada por las tarjetas rojas y los altos goles recibidos sugieren áreas que necesitan mejoras significativas, especialmente en la defensa y en evitar situaciones que conduzcan a expulsiones.

En resumen, el equipo muestra un estilo de juego ofensivo y orientado a la posesión, con una defensa efectiva en duelos en el suelo, pero necesita mejorar en el juego aéreo y en la disciplina para ser más equilibrado y efectivo en diferentes aspectos del juego.

Venezuela

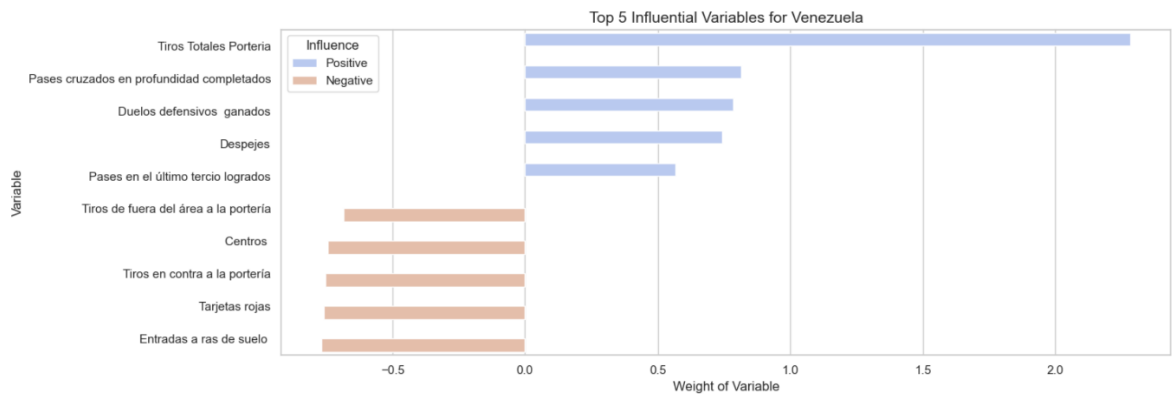


Ilustración 33: Top 5 Venezuela

Análisis de las Variables Positivas:

Tiros Totales a Portería: Un alto número indica una fuerte capacidad ofensiva, con muchas oportunidades de gol creadas y una presión constante en el área rival.

Pases Cruzados en Profundidad Completados: Esta eficacia sugiere habilidad en crear oportunidades de gol mediante jugadas elaboradas y pases precisos en profundidad.

Duelos Defensivos Ganados: Un alto número indica una defensa sólida y efectiva en situaciones de uno contra uno, crucial para recuperar el balón y detener ataques rivales.

Despejes: Una alta cantidad de despejes refleja una defensa capaz de lidiar con situaciones de presión y eliminar amenazas en su área.

Pases en el Último Tercio Logrados: Indica eficacia en mantener la posesión y crear oportunidades en la zona más crítica del campo.

Análisis de las Variables Negativas:

Centros: Una baja eficacia en los centros puede indicar problemas en la creación de oportunidades de gol desde las bandas.

Tiros en Contra a la Portería: Un alto número sugiere que el equipo permite a los oponentes crear oportunidades de gol claras, posiblemente debido a debilidades defensivas o a una táctica de alto riesgo.

Tarjetas Rojas: Un alto número de tarjetas rojas puede ser indicativo de indisciplina o tácticas defensivas agresivas que resultan en expulsiones.

Entradas a Ras de Suelo: Si esta métrica es baja, sugiere una debilidad en recuperar el balón en situaciones de juego cercano.

Goles Recibidos: Una alta cantidad de goles en contra indica debilidades en la defensa.

Estilo de Juego General:

Este equipo muestra un fuerte enfoque ofensivo, destacando en la creación de oportunidades de gol y manteniendo una presión constante en el área rival. La habilidad

para ejecutar pases en profundidad y mantener la posesión en el último tercio del campo sugiere un estilo de juego elaborado y táctico.

Sin embargo, hay claras debilidades defensivas. La alta cantidad de goles recibidos y los tiros en contra indican una defensa que puede ser superada con facilidad, especialmente en situaciones de contraataque. La falta de efectividad en los centros y las tarjetas rojas sugieren áreas que necesitan mejora, como la disciplina y la capacidad para crear oportunidades a partir de jugadas por las bandas.

En conclusión, mientras el equipo es fuerte ofensivamente y en duelos defensivos directos, necesita mejorar en aspectos de juego aéreo, disciplina y en la prevención de situaciones que conducen a oportunidades claras para el rival. La mejora en estas áreas podría llevar a un estilo de juego más equilibrado y efectivo.

Al describir las variables mas y menos influyentes que se extarjeron de los modelos de regresion logística es posible darse cuenta que gran parte de las variables se repiten para ambos casos de esta forma se generan los siguientes wordclouds y gráficos de barras que permiten ver la frecuencia de ocurrencia de cada una de ellas:

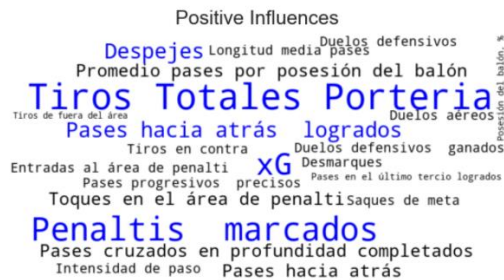


Ilustración 34: Variables Positivas

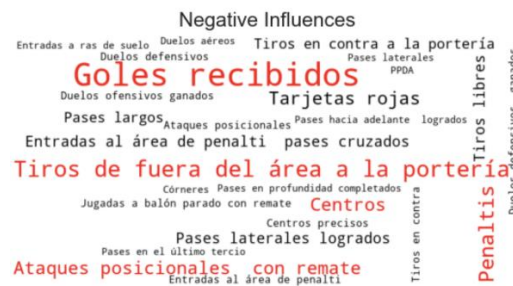


Ilustración 35: Variables Negativas

Del análisis se observa que ciertas variables juegan un papel crucial en determinar la victoria o la derrota de un equipo. De esta manera, 'Tiros Totales a Portería', 'Penaltis Marcados', 'Expectativa de Gol (xG)' y 'Pases hacia atrás exitosos' emergen como factores con una influencia positiva significativa. Estas variables no solo son consistentes en su importancia a través de los 10 modelos analizados, sino que también enfatizan la eficacia ofensiva y la precisión en el juego.

Por otro lado, en el espectro negativo, variables como 'Goles Recibidos', 'Tiros Desviados de la Portería', 'Penaltis Concedidos', 'Centros al Área' y 'Ataques Posicionales con Remate' resaltan por su repetición en los modelos, sugiriendo que estos aspectos pueden ser indicativos de vulnerabilidades defensivas o falta de precisión en el ataque. Gráficamente, y tras un meticuloso ordenamiento de los datos, la información se presenta

de la siguiente manera, permitiendo una visualización clara de cómo estos factores influyen en el rendimiento del equipo en el campo.

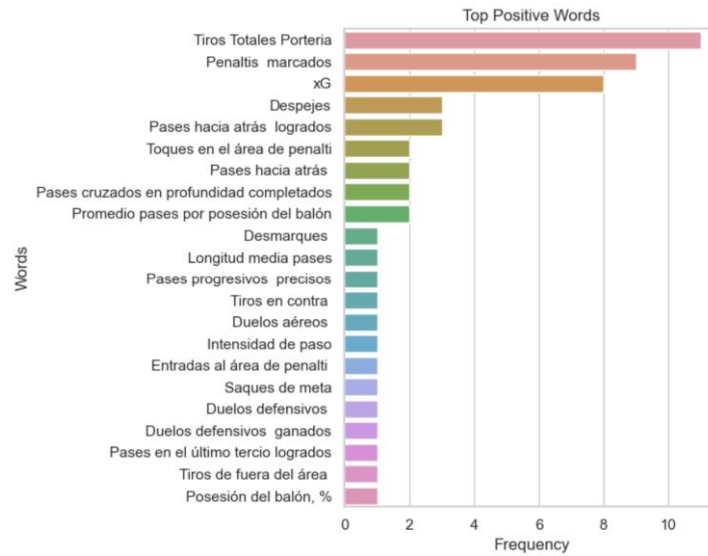


Ilustración 36: Variables Positivas



Ilustración 37: Variables Negativas

5.3. Similaridad de Ligas

Uno de los aspectos cruciales de este análisis es determinar si existen diferencias entre las distintas ligas de fútbol, lo cual nos permitirá definir si hay estilos de juego únicos asociados a cada país. Para lograr este objetivo, hemos elegido la similitud de coseno como nuestra principal técnica de comparación. Esta metodología se destaca por su eficacia en la medición de similitudes en espacios de alta dimensionalidad, como es el caso de nuestras variables asociadas a los estilos de juego. A diferencia de la distancia euclidiana, que mide la magnitud entre dos puntos en un espacio, la similitud de coseno se enfoca en la orientación de los vectores, independientemente de su magnitud. Esto es particularmente útil en nuestro contexto, ya que nos interesa más la dirección (o patrón) de los estilos de juego que la magnitud de las características individuales. Al comparar las ligas de fútbol utilizando la similitud de coseno, podemos capturar más efectivamente las similitudes en los estilos de juego, incluso cuando las magnitudes absolutas de las variables (como frecuencia de pases, agresividad defensiva, etc.) difieren. Este enfoque nos permite ir más allá de las diferencias superficiales y adentrarnos en la estructura subyacente de los estilos de juego, ofreciendo una perspectiva más profunda sobre cómo los equipos en diferentes ligas abordan estratégicamente el juego de fútbol.

Inicialmente se hace PCA sobre todas las ligas encontrándose lo siguiente:

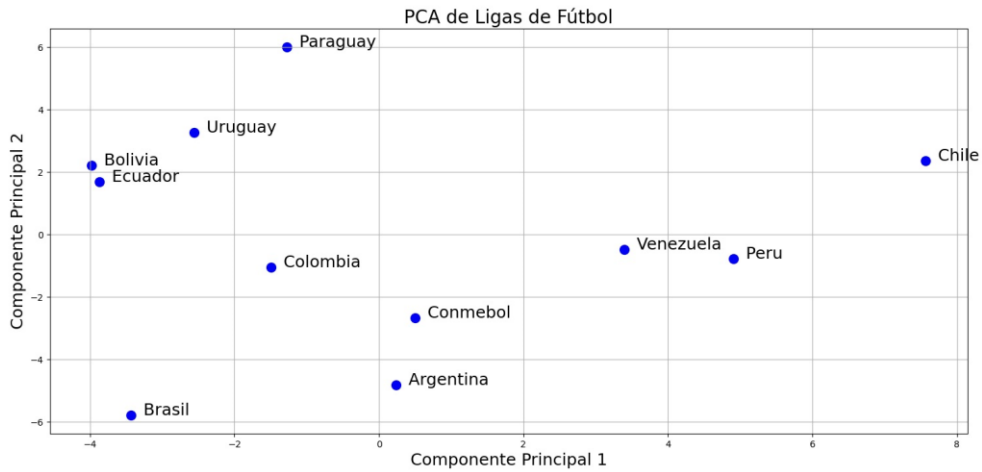


Ilustración 38: PCA de Ligas

Se observan patrones de agrupación interesantes, tales como la proximidad entre Venezuela y Perú, así como la agrupación de Bolivia, Ecuador y Uruguay. Adicionalmente, países como Chile, Argentina, Brasil, Colombia y Paraguay se destacan por su ubicación más aislada respecto a estos grupos. Es crucial subrayar que la técnica utilizada permite una eficiente reducción de la dimensionalidad de los datos. Cabe resaltar la sensibilidad del Análisis de Componentes Principales (PCA) a la escala y variabilidad de los datos, lo que implica que las características con mayor varianza influirán significativamente en los resultados. A continuación, se define la similitud coseno en las 10 ligas, y los valores obtenidos son los siguientes:

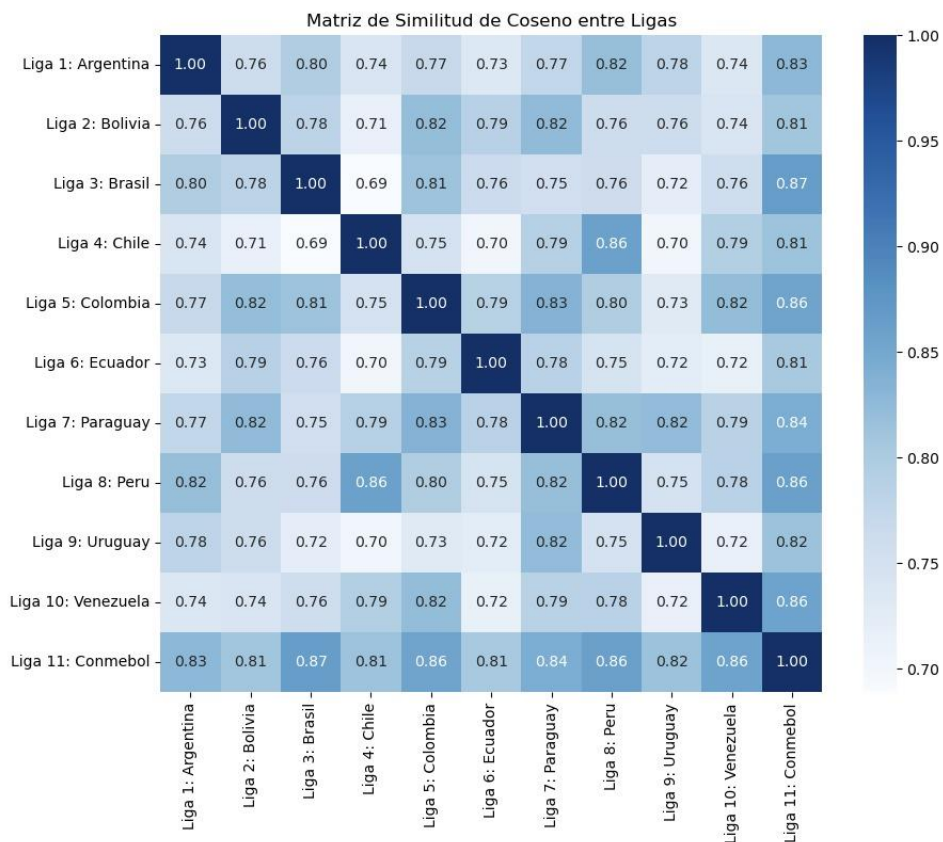


Ilustración 39: Similitud Coseno Ligas

La matriz proporciona una serie de valores que facilitan la interpretación de altos y bajos niveles de similitud, lo que nos ofrece indicativos clave sobre las semejanzas en los estilos de juego entre las distintas ligas. Al observar la matriz se observan que los valores de similitud oscilan entre 0.69 y 0.86, estos valores se consideran altos lo que revela que existe una alta similitud entre las ligas, a continuación, se presenta una tabla por país que muestra el promedio de similitud, los valores máximo-mínimo y los países correspondientes a esta categoría.

País	Promedio	Mínimo	Máximo	Desviación Estándar	Mínimo Similitud	Máximo Similitud
Argentina	0.77	0.73	0.82	0.03	Ecuador	Perú
Bolivia	0.77	0.71	0.82	0.04	Chile	Colombia
Brasil	0.76	0.69	0.81	0.04	Chile	Colombia
Chile	0.75	0.69	0.86	0.06	Brasil	Perú
Colombia	0.79	0.73	0.83	0.03	Uruguay	Paraguay
Ecuador	0.75	0.70	0.79	0.03	Chile	Bolivia
Paraguay	0.80	0.75	0.83	0.03	Brasil	Colombia
Perú	0.79	0.75	0.86	0.04	Ecuador	Chile
Uruguay	0.74	0.70	0.82	0.04	Chile	Paraguay
Venezuela	0.76	0.72	0.82	0.04	Ecuador	Colombia

Tabla 4: Resumen de Similitud Coseno

La mayor similitud se da entre Chile y Perú en un valor de 0.86 y adicionalmente Chile, presenta la más baja similitud con Brasil con un valor de 0.69, luego existe un grupo importante de valores de similitud que oscila entre 0.70 a 0.75 (Brasil- Uruguay-Chile-Ecuador-Argentina- Colombia Venezuela-Bolivia) y finalmente existe otro grupo donde los valores de similitud oscilan entre 0.80 a 0.85(Uruguay-Brasil-Peru-Paraguay-Bolivia-Colombia-Argentina Venezuela). Como una forma de visualizar la data y comprender la similitud entre ligas se procedió a generar una red con los valores de similaridad, encontrándose los siguientes resultados:

Red de Similitud entre Ligas de Fútbol Sudamericanas con Ajustes en Rangos y Colores

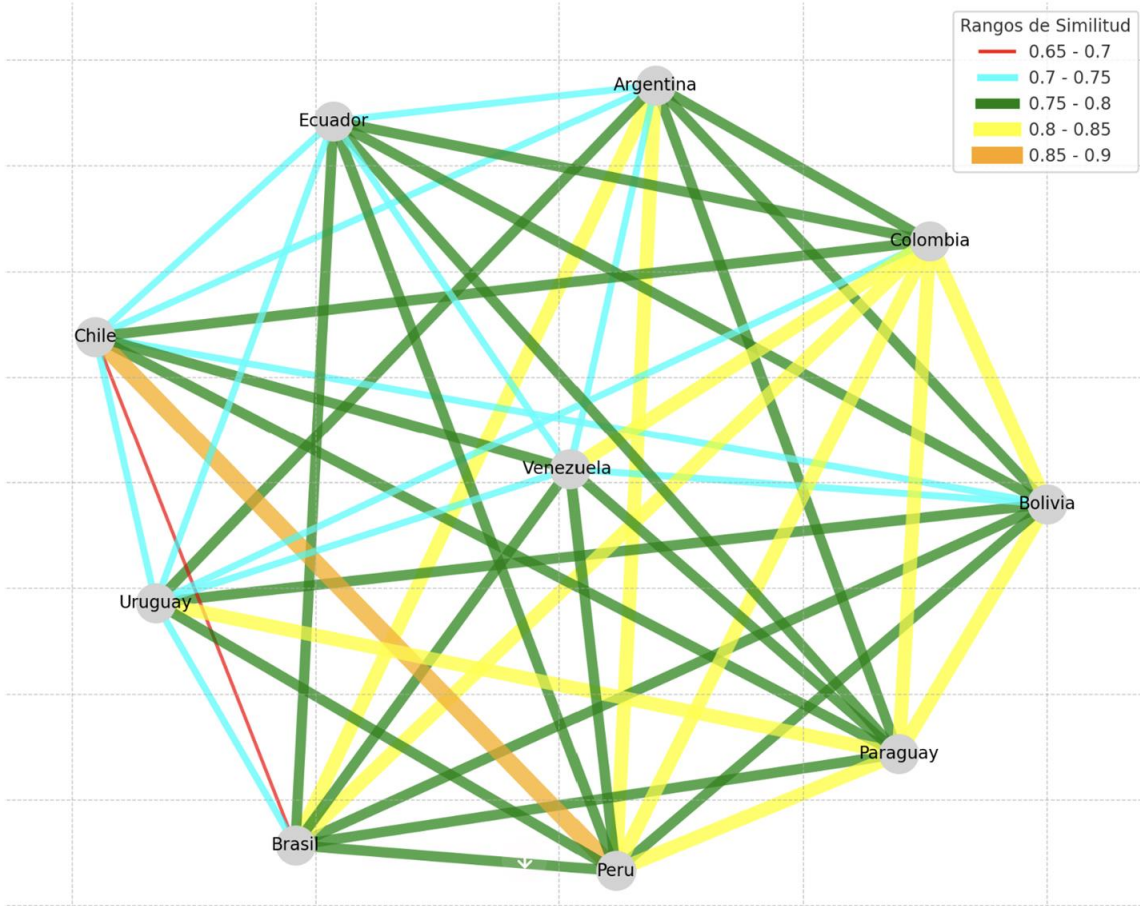


Ilustración 40: Red de Similitud Coseno

Se observan como los nodos están agrupados y los rangos de valores antes mencionados, donde destaca Chile por poseer el mayor y menor valor de similitud, al analizar los datos de similitud se procede a calcular el grado de cada Nodo el cual representa la similitud de cada liga respecto al resto de las ligas sudamericanas.

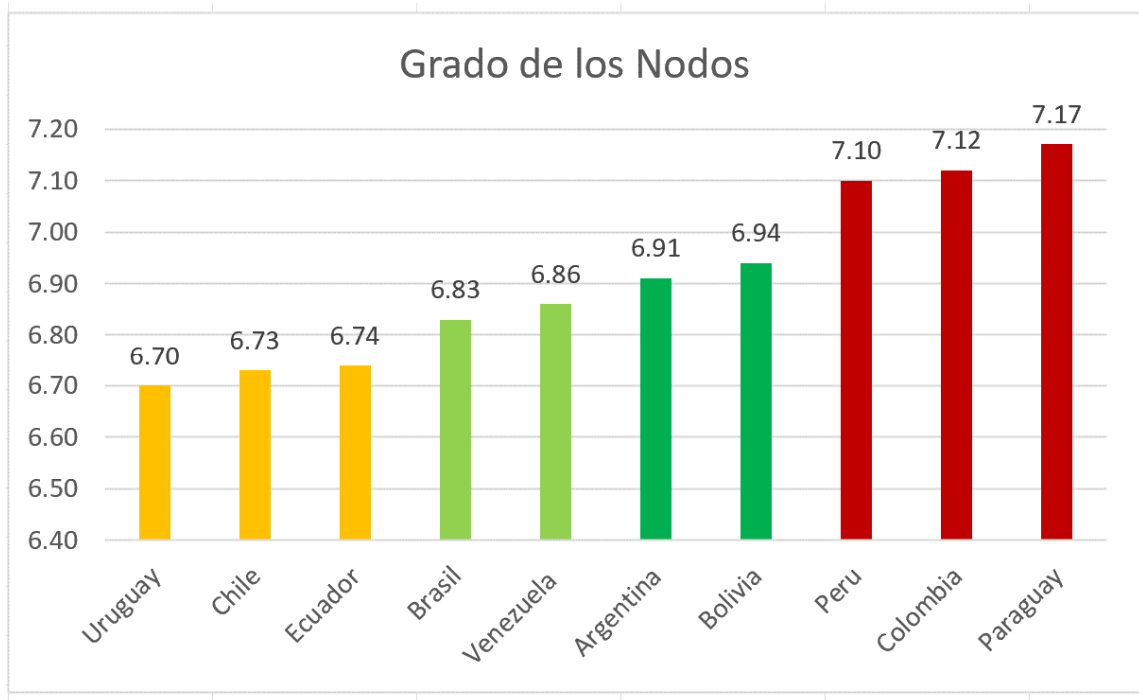


Ilustración 41: Resumen de Grado de Nodo

Al organizar la información y comparar cada país respecto a todos los demás se pudo identificar tres grupos:

Uruguay-Chile-Ecuador: Son los equipos que menos se parecen en estilo de juego, al resto de los equipos sudamericanos. Dado su bajo valor de grado de nodo.

Brasil-Venezuela-Argentina-Bolivia: Poseen un valor de grado de nodo intermedio, por lo tanto son medianamente parecidos en su estilo de juego respecto al resto de los equipos sudamericanos.

Peru-Bolivia-Paraguay: Poseen el más alto valor de grado de nodo, por lo tanto, poseen un alto parecido en su estilo de juego respecto al resto de los equipos sudamericanos.

Al profundizar en el análisis de los datos recopilados, se evidencia que los valores de similitud coseno entre las ligas son significativamente altos, oscilando entre 0.69 y 0.86. Esta medida de similitud, basada en el método de similitud coseno, refleja un alto grado de coincidencia en las características de los equipos analizados, independientemente de sus ligas nacionales. En el campo de la ciencia de datos, valores de similitud coseno en este rango indican una fuerte correlación en los patrones de juego y estrategias entre los equipos, sugiriendo que las diferencias entre las ligas pueden no ser tan pronunciadas como se había hipotetizado inicialmente.

Dado este hallazgo, surge la necesidad de reevaluar la agrupación original basada en ligas nacionales. La alta similaridad coseno observada conduce a la consideración de una nueva estratificación de los datos, donde la agrupación de equipos podría basarse en criterios alternativos, posiblemente trascendiendo las fronteras nacionales. Esta reorganización de los grupos se justifica por la aplicación técnica del análisis de similitud coseno, que mide la magnitud de la similaridad en un espacio multidimensional, donde cada dimensión representa una variable del modelo de regresión logística utilizado.

Este enfoque contrasta con la hipótesis inicial que postulaba que la nacionalidad era el principal factor distintivo en los estilos de juego. Sin embargo, el análisis detallado mediante la similitud coseno revela que, aunque existen diferencias en los patrones de juego, estas no son tan marcadas como se esperaba. Por lo tanto, se procede a un reprocesamiento y reagrupación de los datos, teniendo en cuenta estas nuevas

perspectivas. Este paso es fundamental para garantizar que el análisis subsiguiente esté alineado con las verdaderas similitudes y diferencias que emergen de los datos, conforme a los estándares metodológicos en ciencia de datos y análisis estadístico avanzado.

5.4. Clusterización por Equipos

Utilizamos los datos de 15225 partidos obtenidos de la plataforma Wyscout para realizar un análisis de clúster. Se emplea la información preprocesada del proceso anterior. Este proceso nos ayudará a identificar características distintivas y variados estilos de juego. Comenzamos determinando el número óptimo de clústeres mediante el método del codo, lo que nos permite crear una gráfica más definida y significativa.

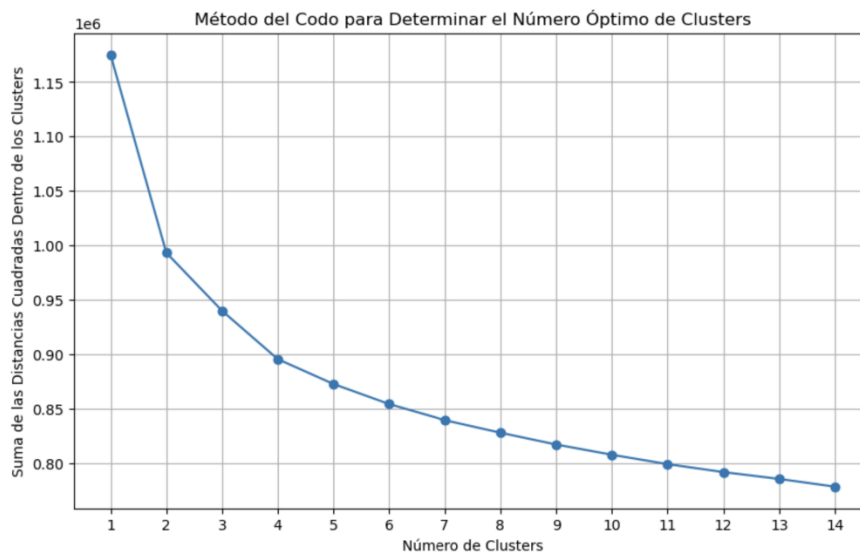


Ilustración 42: Método del Codo

Del análisis se determina que a través del método del codo el número óptimo de clústeres es de alrededor de 4. Con esta información se procede a graficar el número de clúster definido:

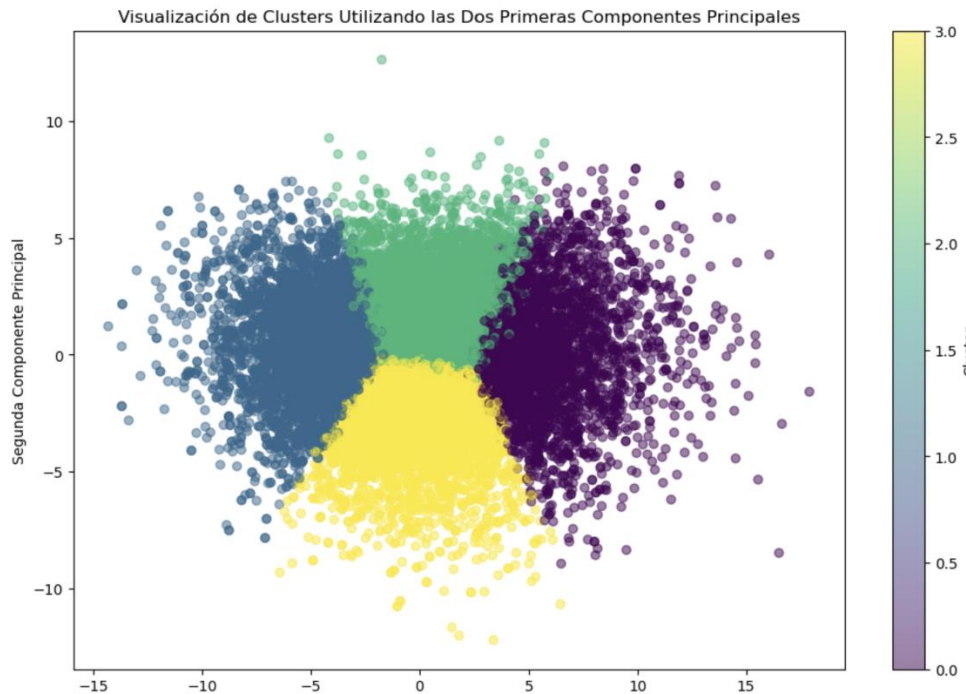


Ilustración 43: Clusterización de Equipos

Tras definir los clusters, se procede a un análisis detallado de los equipos de fútbol, utilizando métricas ampliamente reconocidas en el análisis deportivo para evaluar y comparar su rendimiento. Para este propósito, se han seleccionado variables específicas que son fundamentales para entender distintos aspectos del juego. Estas incluyen la efectividad ofensiva, el estilo de juego, la táctica defensiva y la agresividad.

Entre las estadísticas elegidas para comparar los clusters, se encuentran indicadores clave de la ofensiva como la posesión del balón, los goles marcados, los xG (goles esperados) y los tiros totales, que juntos aportan una visión integral de la eficiencia en ataque. En contraste, para analizar la estrategia defensiva, se consideran métricas como las faltas cometidas, el PPDA (pases permitidos por acción defensiva) y el porcentaje de

lanzamientos largos, que ofrecen insights valiosos sobre el enfoque defensivo y la aproximación general al juego.

A continuación, se detallan las variables empleadas en el análisis:

- **Posesión del balón (%):** Indica qué tan a menudo un equipo controla el balón durante un partido. Una alta posesión puede sugerir un estilo de juego basado en el control y la construcción metódica de jugadas, mientras que una posesión más baja puede indicar un enfoque en el contraataque o un estilo de juego más directo.
- **Goles:** Es la métrica más directa del éxito ofensivo de un equipo. Comparar los promedios de goles por cluster ayuda a entender qué tan efectivos son los diferentes estilos de juego en términos de marcar.
- **xG (Goles Esperados):** Proporciona una medida de la calidad de las oportunidades de gol creadas. Un alto xG sugiere que un equipo crea muchas oportunidades claras de gol, independientemente de si estas se convierten en goles reales.
- **Tiros Totales:** Indica la frecuencia con la que un equipo intenta marcar. Un mayor número de tiros puede reflejar un enfoque más ofensivo o agresivo.
- **Pases Totales:** Ayuda a entender el estilo de juego en términos de construcción de jugadas. Un alto número de pases suele asociarse con un juego de posesión y control, mientras que un número más bajo puede indicar un juego más directo.
- **Faltas:** Da una idea del nivel de agresividad o intensidad defensiva de un equipo. Más faltas pueden sugerir un juego más físico o una táctica de presión alta.
- **PPDA (Presión Defensiva Pasiva):** Mide la intensidad de la presión defensiva de un equipo. Un PPDA más bajo indica una presión más alta y agresiva para recuperar el

balón, mientras que un PPDA más alto sugiere un enfoque más pasivo o posicional en la defensa.

- Porcentaje de Lanzamiento Largo: Esta estadística refleja la propensión de un equipo a utilizar pases largos. Un porcentaje más alto puede indicar un estilo de juego más directo, buscando avanzar rápidamente hacia la portería contraria, mientras que un porcentaje más bajo sugiere un enfoque en el juego corto y la retención del balón.

Al graficar estas variables podemos observar algunas diferencias por cada Cluster:

Comparación de Características Clave por Cluster

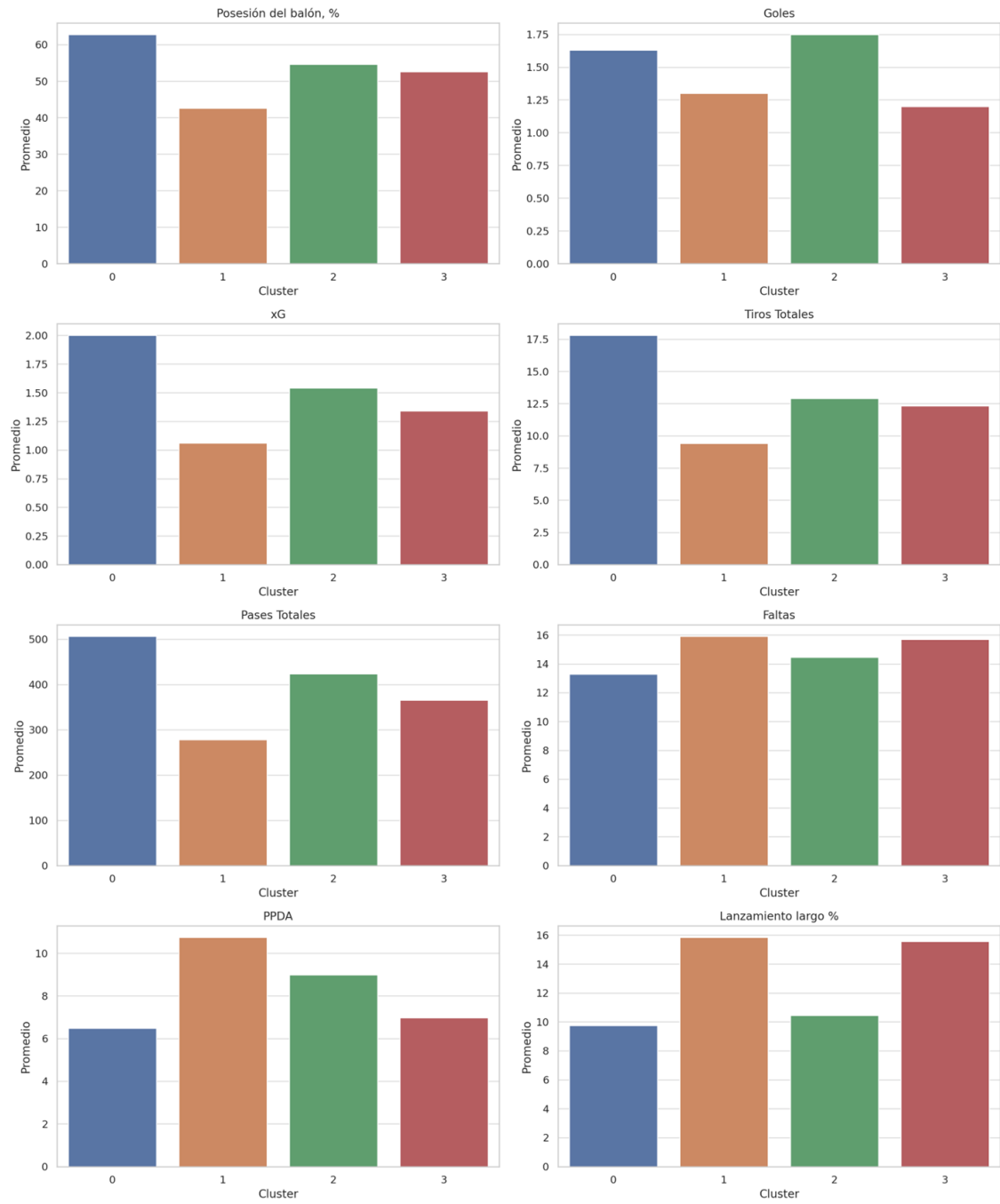


Ilustración 44: Comparación de Características por Cluster

CLUSTER 0

Estilo de Juego: Este cluster se caracteriza por un juego de posesión dominante (62.81%) con un alto número de pases totales (506.44) y una buena precisión de pase (430.95 pases totales logrados).

Ofensiva: Exhibe un alto volumen de tiros (17.80 tiros totales) con una cantidad significativa de ellos en portería (6.22). Los goles promedio (1.63) y los goles esperados (xG de 2.00) sugieren un enfoque ofensivo eficiente.

Defensiva y Transición: Menos faltas (13.28) y un bajo PPDA (6.49) indican una presión efectiva y recuperación rápida del balón. Los saques de meta son relativamente bajos (6.39), lo que puede indicar menos situaciones defensivas.

Estilo de Pase: Preferencia por pases más cortos y precisos, con una distancia media de tiro de 19.20 y una longitud media de pases de 20.40.

CLUSTER 1

Estilo de Juego: Juego más directo con menor posesión del balón (42.56%) y menor cantidad de pases (277.79 pases totales).

Ofensiva: Menor actividad ofensiva, con menos tiros (9.42 tiros totales) y menor precisión de tiro (3.57 tiros totales a portería). Los goles (1.30) y el xG (1.06) están en el extremo inferior.

Defensiva y Transición: Mayor número de faltas (15.91) y PPDA más alto (10.75) sugieren un enfoque más físico y menos presión en la recuperación del balón. Los saques de meta son más altos (11.08), posiblemente indicando más situaciones defensivas.

Estilo de Pase: Mayor preferencia por lanzamientos largos (15.85%) y pases más largos en promedio (21.15).

CLUSTER 2

Estilo de Juego: Juego equilibrado con buena posesión del balón (54.66%) y un número decente de pases (423.92).

Ofensiva: Actividad ofensiva moderada con 12.90 tiros totales y 4.97 en portería. Los goles (1.75) y xG (1.54) son decentes, indicando eficacia en la finalización.

Defensiva y Transición: Faltas moderadas (14.45) y PPDA (8.99) reflejan un balance entre juego físico y táctico.

Estilo de Pase: Lanzamiento largo moderado (10.46%) con una distancia media de tiro de 19.74 y longitud media de pases de 20.18.

CLUSTER 3

Estilo de Juego: Similar al Cluster 2 en términos de posesión (52.58%) pero con menos pases (365.37).

Ofensiva: Menos eficacia ofensiva (1.20 goles, 1.34 xG) con un número moderado de tiros (12.34).

Defensiva y Transición: Faltas y PPDA ligeramente altos (15.70 y 6.98 respectivamente) sugieren un estilo de juego más disruptivo.

Estilo de Pase: Mayor preferencia por lanzamientos largos (15.56%) y pases más largos en promedio (21.29).

En función a la información anterior, cada cluster podría definirse de la siguiente forma;

Cluster 0: Estilo de juego de posesión dominante, ofensiva eficiente, presión efectiva y recuperación rápida del balón, preferencia por pases cortos y precisos.

Cluster 1: Juego más directo, menor actividad ofensiva, enfoque físico y menos presión en la recuperación del balón, preferencia por lanzamientos largos y pases más largos.

Cluster 2: Juego equilibrado, actividad ofensiva moderada, balance entre juego físico y táctico, lanzamiento largo y pases moderados.

Cluster 3: Menos eficacia ofensiva, estilo de juego disruptivo, mayor preferencia por lanzamientos largos y pases más largos.

El análisis de los datos muestra que cada equipo tiene la capacidad de migrar entre los cuatro clústeres definidos. Para facilitar la interpretación de cómo los equipos cambian su estilo de juego y cómo se agrupan según esta variabilidad, se generó un dendrograma utilizando la similitud coseno:

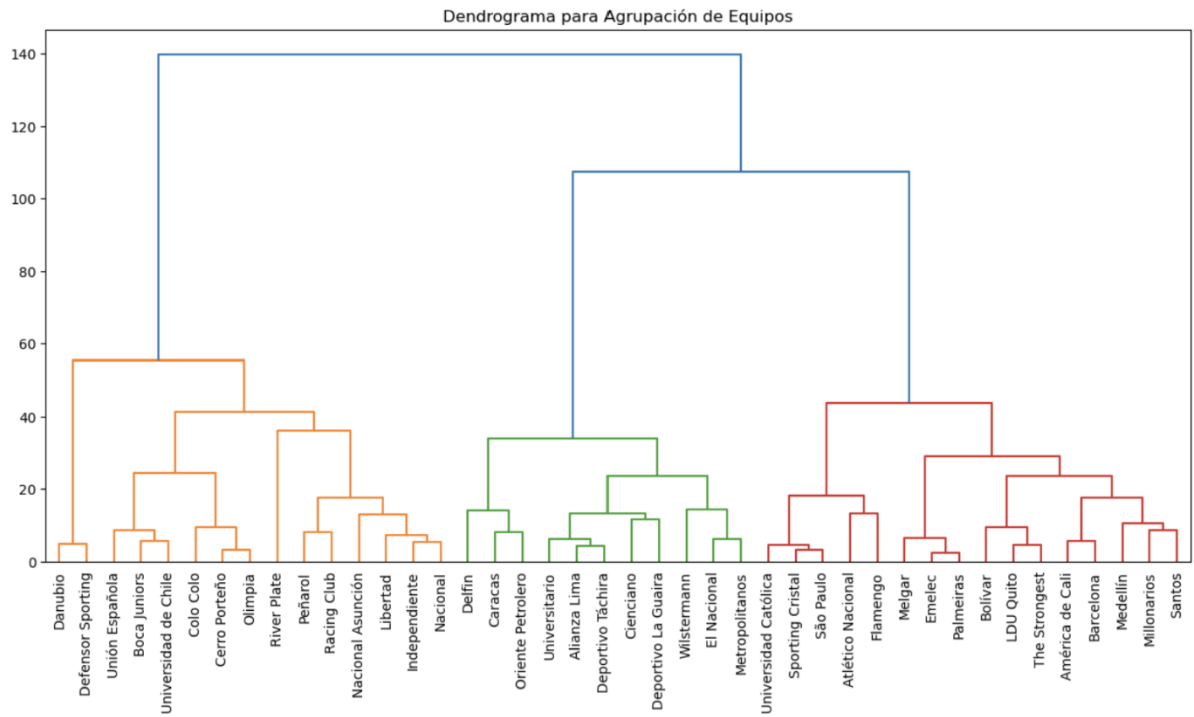


Ilustración 45: Dendograma Jerarquizado

Se observan varias agrupaciones en base al dendograma, , de esta forma se procede a graficar la distribución porcentual de los equipos en clusters.

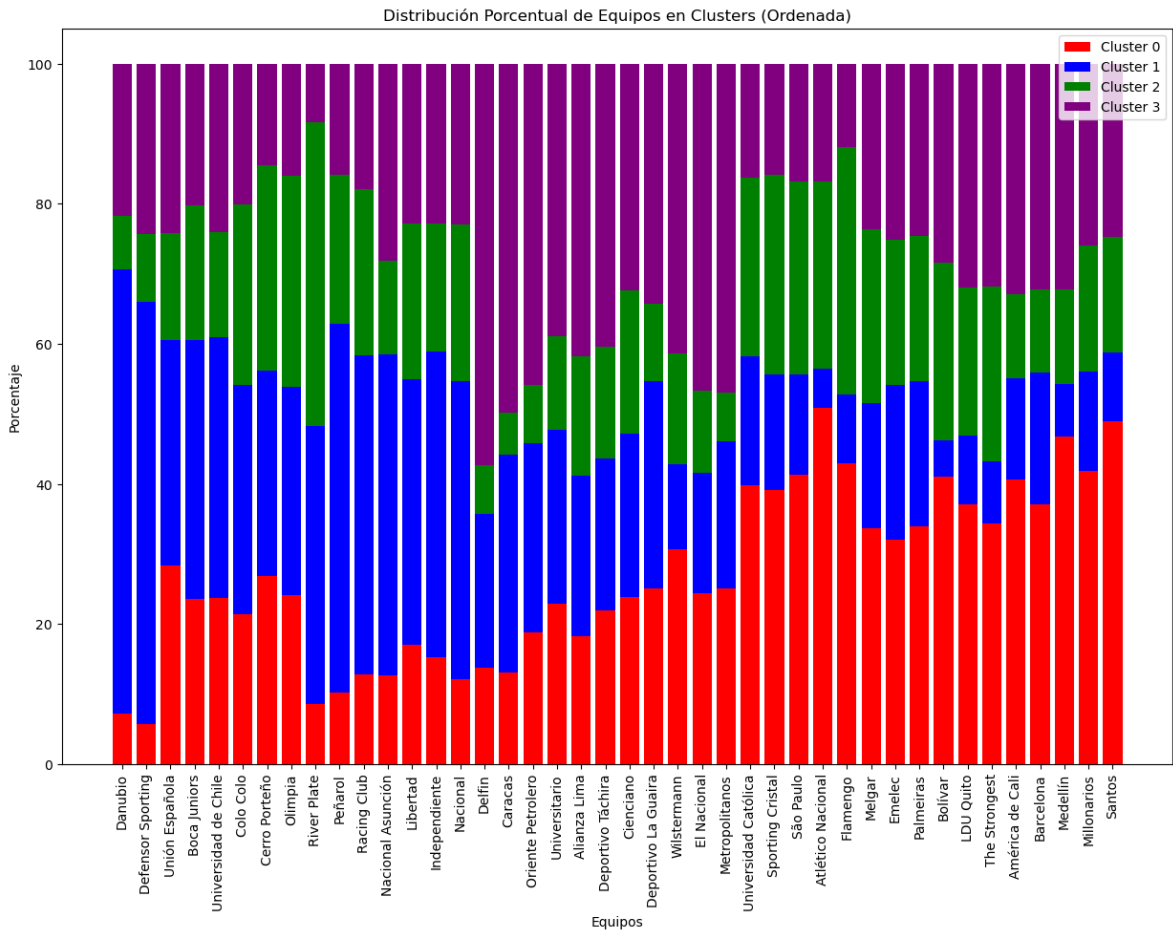


Ilustración 46: Distribución Porcentual por Equipos Ordenada

Al observar como los equipos se distribuyen en los distintos clusters, surge la duda de como medir la versatilidad de un equipo para moverse de un estilo a otro de esta forma se emplea el Índice de Shannon, también conocido como la entropía de Shannon, es una medida comúnmente utilizada en teoría de la información y estadística para cuantificar la diversidad o incertidumbre de un conjunto de datos. En el contexto del fútbol, este índice puede aplicarse para evaluar cómo un equipo se distribuye entre diferentes estilos de juego, representados aquí como cuatro clusters. Cada cluster representa un estilo de juego distinto, y la presencia de un equipo en un cluster específico indica su afinidad o tendencia

hacia ese estilo en particular. La distribución de un equipo a través de estos clusters puede variar: algunos equipos pueden mostrar una fuerte preferencia por un único estilo, mientras que otros pueden distribuirse más uniformemente entre varios estilos.

El Índice de Shannon se calcula utilizando las proporciones de la presencia de un equipo en cada cluster. Se determina sumando el producto de la proporción de la presencia del equipo en cada cluster y el logaritmo natural de esta proporción, y luego tomando el negativo de esta suma. Un valor más alto de la entropía de Shannon indica una distribución más uniforme entre los clusters, lo que sugiere una mayor versatilidad o flexibilidad táctica del equipo.

Al aplicar este índice a varios equipos, se puede generar un ranking basado en la capacidad de cada equipo para distribuirse entre los diferentes estilos de juego. Un equipo con un índice de Shannon alto se clasificará más arriba en el ranking, indicando que no muestra una preferencia fuerte por un único estilo de juego y se adapta a diferentes tácticas. Por otro lado, un índice bajo sugiere una especialización o consistencia en un estilo particular. Este ranking proporciona una perspectiva valiosa para entrenadores, analistas y aficionados, ya que revela no solo la versatilidad táctica de los equipos, sino también su capacidad para adaptarse a diferentes estrategias de juego. En resumen, el Índice de Shannon, aplicado en este contexto, es una herramienta útil para evaluar y comparar la adaptabilidad táctica de los equipos en el fútbol moderno. De esta forma el ranking generado fue el siguiente:

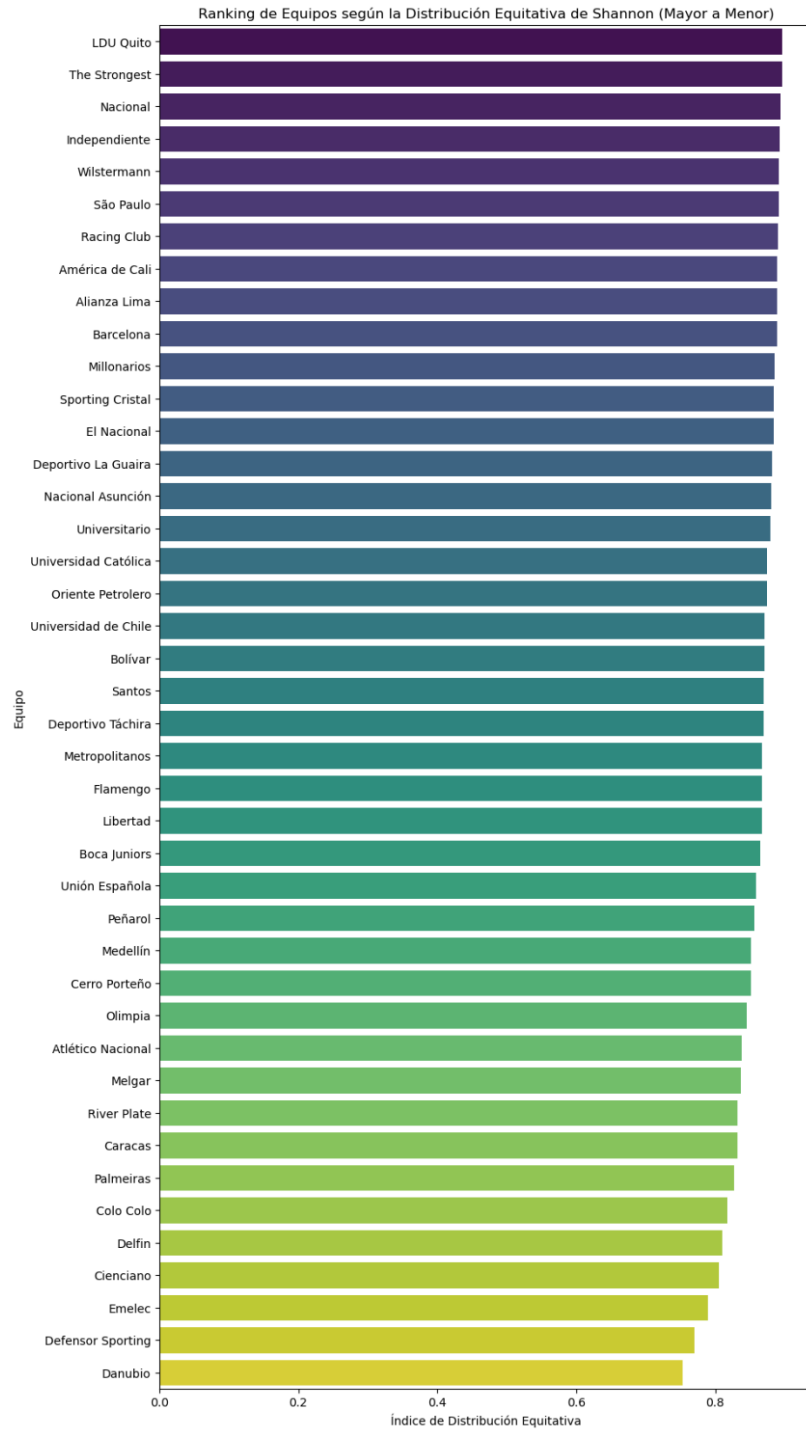
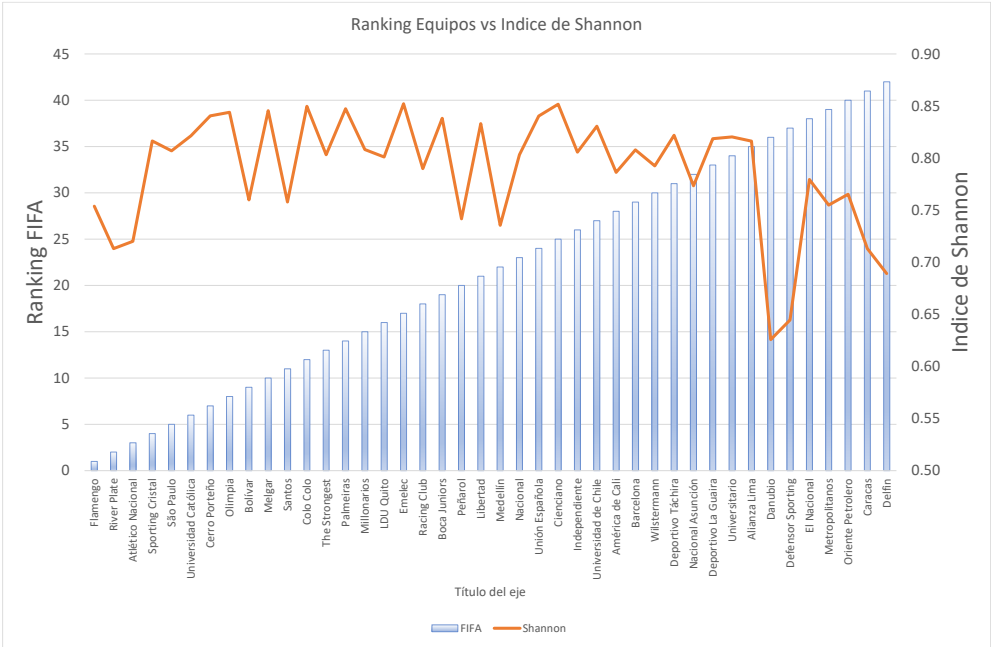


Ilustración 47: Ranking Índice de Shannon

La gráfica revela claramente que equipos como LDU Quito, The Strongest, Nacional, Independiente y Wilstermann exhiben una notable versatilidad para adaptarse a diversos estilos de juego. Esta capacidad para moverse entre diferentes tácticas los sitúa en una posición privilegiada en términos de flexibilidad estratégica. Por otro lado, equipos como Danubio, Defensor Sporting, Emelec y Cienciano muestran los índices más bajos, lo cual sugiere una menor versatilidad. Estos equipos parecen tener un estilo de juego más definido y consistente, lo que podría implicar una menor adaptabilidad a diferentes enfoques tácticos. Esta diferencia en los índices subraya la diversidad de estrategias adoptadas por los distintos equipos y su capacidad para ajustarse a variados estilos de juego en el fútbol moderno.

No necesariamente la versatilidad de juego implica más éxito cuando comparamos con el Ranking FIFA de Diciembre 2022.



Tras la creación del dendrograma, se formaron agrupaciones entre los diferentes equipos.

Los grupos conformados en base al dendrograma fueron los siguientes:

Grupo	Equipos
Danubio	Danubio, Defensor Sporting
Union Española	Unión Española, Boca Juniors, Universidad de Chile, Colo Colo, Cerro Porteño, Olimpia
River Plate	River Plate, Peñarol, Racing Club, Nacional Asunción, Libertad, Independiente, Nacional
Delfin	Delfin , Caracas, Oriente Petrolero
Universitario	Universitario, Alianza Lima, Deportivo Tachira, CIenciano, Deportivo Guaira, WIlsterman, El Nacional, Metropolitano
Universidad Católica	Universidad Católica, Sporting Cristal, Sao Paulo, Atletico Nacional, Flamengo
Melgar	Melgar, Emelec, Palmeiras, Bolivar, LDU Quito, Strongest, America de Cali, Barcelona, Medellin, Millonario, Santos

Tabla 5: Agrupacion de Equipos

A continuación, se implementaron modelos de regresión logística con el objetivo de identificar la importancia de cada variable para los nuevos grupos conformados. Esto nos permitió construir vectores que facilitan la determinación de la similitud entre estos grupos recién formados. Las métricas utilizadas para evaluar cada uno de los modelos fueron las siguientes:

Grupo	Model	Accuracy	Precision	Recall	F1 Score
Danubio	Logistic Regression	0.865	0.888	0.865	0.865
Union_Espanola	Logistic Regression	0.935	0.936	0.935	0.935
River_Plate	Logistic Regression	0.897	0.899	0.897	0.898
Delfin	Logistic Regression	0.880	0.881	0.880	0.880
Universitario	Logistic Regression	0.926	0.926	0.926	0.926
Universidad_Catolica	Logistic Regression	0.926	0.926	0.926	0.926
Melgar	Logistic Regression	0.925	0.925	0.925	0.925

Tabla 6: Resultado del Modelamiento

Las métricas obtenidas para cada uno de los grupos de los modelos resulto ser bastante aceptable con valores por sobre 0.87 en términos de accuracy, Danubio resultó con un valor menor respecto a la tendencia de los otros grupos, no obstante para la realización del análisis se definió como un valor aceptable. De esta forma se emplearon los pesos de las variables obtenidos de loa regresión Logistica y se procedió a generar la matriz de similitud con el objeto de comparar la similitud de los nuevos grupos generados.

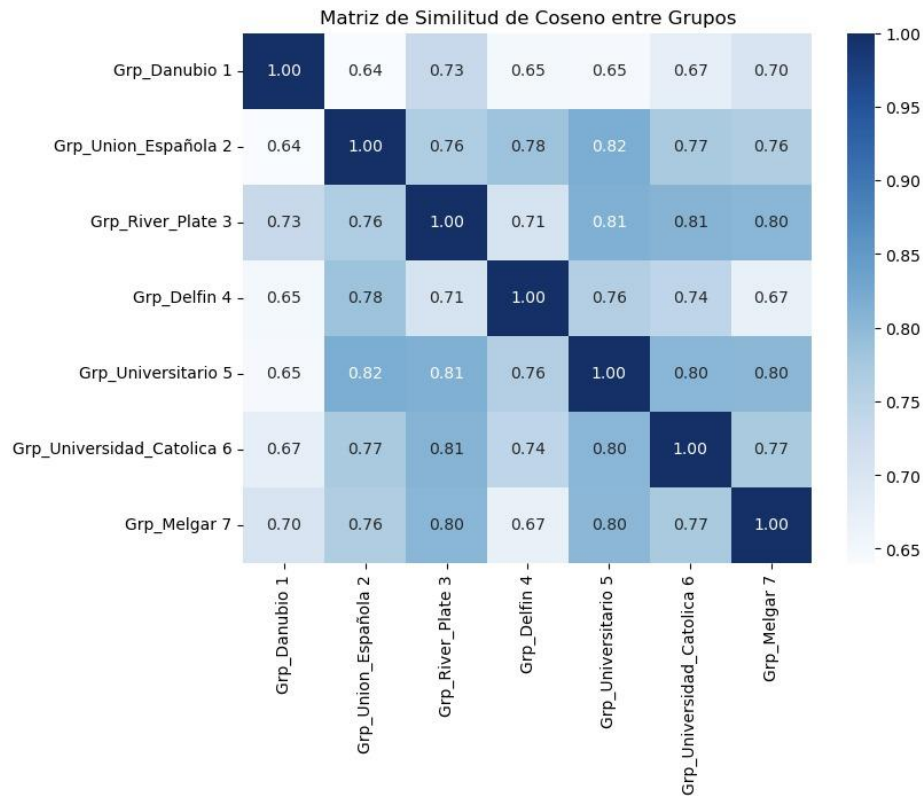


Ilustración 48: Resultado de Matriz de Similaridad

La matriz de similitud proporciona información esencial sobre la similitud entre los grupos recientemente conformados. Para una comprensión más clara y concisa, esta matriz se ha sintetizado en la tabla siguiente:

Grupo	Media	Desviación Estándar	Mínimo	Máximo	Mínimo	Máximo2
Grp_Danubio 1	0.673	0.035	0.640	0.730	Grp_Union_Española 2	Grp_River_Plate 3
Grp_Union_Española 2	0.755	0.061	0.640	0.820	Grp_Danubio 1	Grp_Universitario 5
Grp_River_Plate 3	0.770	0.043	0.710	0.810	Grp_Danubio 1	Grp_Universidad_Catolica 6
Grp_Delfin 4	0.718	0.051	0.650	0.780	Grp_Danubio 1	Grp_Union_Española 2
Grp_Universitario 5	0.773	0.064	0.650	0.820	Grp_Danubio 1	Grp_Union_Española 2
Grp_Universidad_Catolica 6	0.760	0.051	0.670	0.810	Grp_Danubio 1	Grp_River_Plate 3
Grp Melgar 7	0.750	0.054	0.670	0.800	Grp_Universitario 5	Grp_Delfin 4

Tabla 7: Resumen Matriz de Similaridad

Rango de Similitud: Los valores de similitud coseno varían entre 0 y 1, donde valores cercanos a 1 indican una alta similitud. En tu conjunto de datos, los valores oscilan entre

0.64 y 0.82. Esto sugiere que, aunque hay diferencias entre los grupos, todos comparten cierto nivel de similitud en las características analizadas.

Grupo con Menor Similitud (Grp_Danubio 1): Este grupo tiene el valor mínimo de similitud (0.64), lo que indica que es el que presenta mayores diferencias respecto a los demás grupos en el conjunto de datos. Podría ser interesante investigar cuáles características específicas hacen que este grupo sea tan distinto.

Grupo con Mayor Similitud (Grp_Union_Española 2): Por otro lado, este grupo muestra el valor máximo de similitud (0.82), lo que sugiere que tiene muchas características en común con los otros grupos. Esto podría indicar un estilo de juego o características de equipo similares a los demás.

Desviación Estándar: Los valores de desviación estándar son relativamente bajos en todos los grupos, lo que indica que la similitud de coseno entre los diferentes grupos no varía drásticamente. Esto sugiere una cierta uniformidad en las características o el desempeño de los equipos dentro de cada grupo.

Análisis por Grupo: Cada grupo tiene su propia media de similitud, lo que refleja cómo se compara en promedio con los otros grupos. Grupos con una media más alta tienen, en general, más características en común con otros grupos. Esto podría ser útil para análisis de emparejamientos o competiciones, donde se buscan equipos con estilos similares o diferenciados.

En función a la información anterior se procede a generar una red con la información extraída de esta forma la representación grafica queda de la siguiente forma:

Red de Similitud Coseno entre Grupos de Equipos de Fútbol (Color y Grosor por Valor de Similitud)

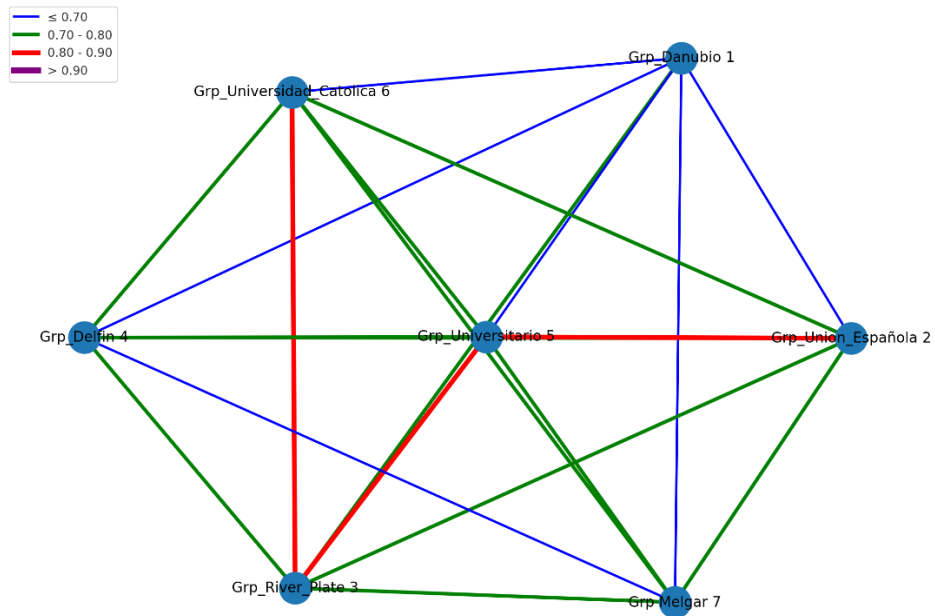


Ilustración 49: Red con Nuevas Agrupaciones

De la red se destacan 4 grandes grupos asociados en base a su similitud destacando la similitud entre el Grp_Universidad Católica, Grp_River_Plate, Grp_Universitario y Grp_Union_Española y a su vez la conformación de una red de baja similitud conformada por Grp_Melgar, Grp_Delfin, Grp_Universidad_Catolica, Grp_Danubio, Grp_Union_Española.

Se procede a calcular el grado de nodo de los distintos Grupos conformados, determinándose que estos valores son menores respecto a la asociación inicial supuesta en base a la nacionalidad de los equipos.

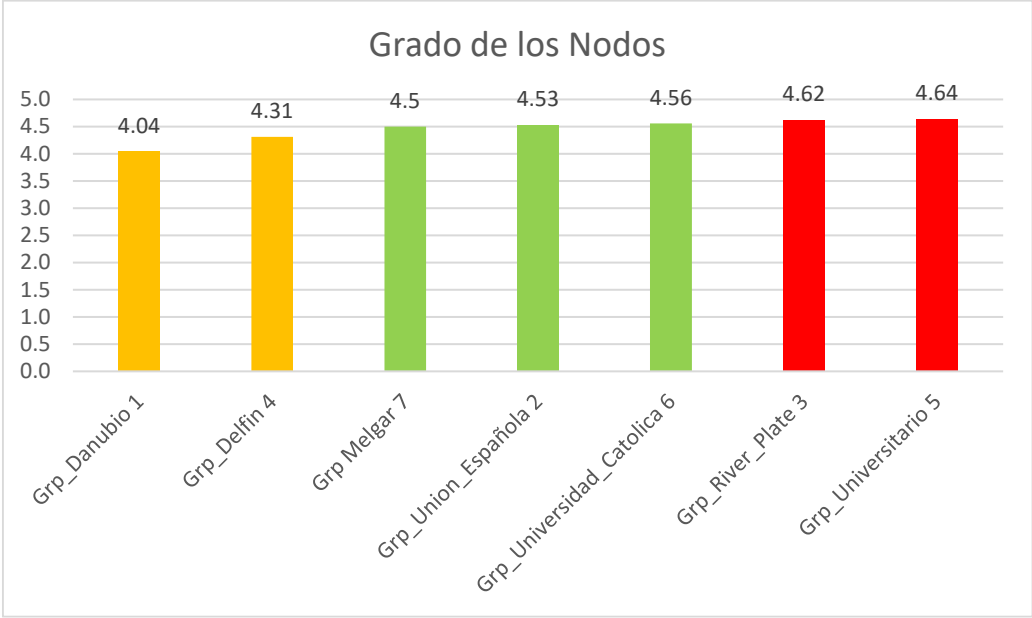


Ilustración 50: Grado de Nodo nueva Agrupación

Se aprecia que los valores de Grado varían entre 4.04 y 4.64, lo que sugiere que la metodología aplicada recientemente ha sido efectiva en resaltar y diferenciar las características de juego dentro de esta nueva agrupación de manera más precisa.

6. Conclusiones

Las Principales conclusiones del estudio realizado se pueden dividir en los siguientes puntos:

Similaridad de Ligas y Reagrupación de Equipos:

A pesar de las diferencias percibidas entre las ligas de CONMEBOL, el estudio reveló una alta similitud en los estilos de juego, con valores de similitud coseno entre 0.69 y 0.86. Esta observación sugiere que las diferencias entre las ligas no son tan pronunciadas como se pensaba, lo que llevó a la necesidad de reorganizar la agrupación de equipos más allá de las fronteras nacionales, basada en similitudes tácticas más que en la nacionalidad.

Clusterización por Equipos y Versatilidad Táctica:

Se identificaron cuatro clusters distintos, reflejando una variedad de estilos de juego, desde un juego de posesión dominante hasta un estilo más directo y físico. La capacidad de los equipos para moverse entre estos clusters indica una notable versatilidad y adaptabilidad en sus enfoques tácticos. Sin embargo, esta versatilidad táctica, medida a través del Índice de Shannon, no mostró una correlación directa con el éxito en rankings como el de la FIFA. Equipos con alta versatilidad táctica no necesariamente ocuparon

posiciones altas en el ranking, lo que sugiere que la flexibilidad táctica es solo uno de varios factores que influyen en el éxito en el fútbol.

Nuevas Agrupaciones y Representación de Redes:

La creación de un dendrograma permitió formar nuevas agrupaciones de equipos basadas en la similitud coseno. Estas nuevas agrupaciones proporcionaron una visión más detallada de la similitud entre los estilos de juego de los equipos. La representación gráfica de estas agrupaciones mediante redes y el cálculo del grado de nodo para cada grupo revelaron que las nuevas agrupaciones ofrecían una distinción más precisa de las características de juego, destacando la utilidad de este enfoque en resaltar diferencias y similitudes tácticas más sutiles entre los equipos.

Uso del Modelo de Regresión Logística en el Análisis:

El Modelo de Regresión Logística desempeñó un papel fundamental en este estudio. La interpretación de los pesos de las variables en el modelo proporcionó insights valiosos sobre los factores más influyentes en los estilos de juego de los equipos. Esta metodología permitió identificar las características específicas que diferencian un estilo de juego de otro, ofreciendo una comprensión más profunda y matizada de los enfoques tácticos en el fútbol sudamericano.

Conclusión General del Estudio:

El estudio concluyó que existe una gran diversidad táctica dentro del fútbol sudamericano, y que la agrupación basada en la nacionalidad no refleja necesariamente una uniformidad en el enfoque táctico. La clasificación de los estilos de juego en cuatro categorías principales proporciona un marco para evaluar y comparar tácticas en el fútbol contemporáneo, destacando la riqueza y profundidad estratégica del deporte. En resumen, el análisis detallado de los equipos de CONMEBOL utilizando técnicas de machine learning reveló una compleja red de similitudes y diferencias tácticas que trascienden las fronteras nacionales. La capacidad de adaptarse a diferentes estilos de juego, aunque valiosa, es solo un aspecto de lo que contribuye al éxito en el fútbol, y la representación mediante redes y el análisis de grado de nodo, junto con el uso del Modelo de Regresión Logística, proporcionan herramientas adicionales para comprender la dinámica táctica en el fútbol sudamericano.

7. Trabajos Futuro

Continuando con el estudio sobre estilos de juego en las ligas de CONMEBOL utilizando data science, existen varias direcciones interesantes para futuros trabajos. Un área prometedora es el desarrollo de modelos predictivos avanzados que integren los estilos de juego identificados para predecir resultados de partidos, considerando factores como la condición de local o visitante y la influencia de jugadores clave. Además, sería valioso investigar cómo evolucionan estos estilos de juego a lo largo del tiempo y cómo esto afecta el rendimiento y las tácticas. Ampliar el análisis para incluir una comparación

internacional con otras confederaciones también sería enriquecedor. Esto permitiría explorar cómo los estilos de juego de CONMEBOL se comparan y contrastan con los de otras regiones, como UEFA o AFC, y entender las tendencias tácticas a nivel global. Otro aspecto importante es examinar el impacto de factores externos, como cambios en la gestión de los equipos, en los estilos de juego. Asimismo, la implementación de técnicas de machine learning más sofisticadas, como el deep learning, podría proporcionar una comprensión más detallada de las tácticas de juego. Un enfoque detallado en el análisis de jugadores individuales y cómo sus estilos influyen en el juego general del equipo también podría revelar dinámicas interesantes. Además, aplicar ciencia de datos en la gestión deportiva podría mejorar significativamente aspectos como la planificación de entrenamientos y estrategias de reclutamiento. Finalmente, estudiar la eficiencia del juego, como la relación entre la posesión del balón y los goles marcados, y cómo esto se relaciona con el éxito del equipo, podría ofrecer insights valiosos para la formulación de estrategias en el campo. Estos enfoques no solo proporcionarían una comprensión más profunda del fútbol sudamericano, sino que también avanzarían en el campo emergente de la ciencia de datos aplicada al deporte.

8. Limitaciones

Las limitaciones principales de este estudio están relacionadas con la cantidad de datos disponibles por liga y el acceso limitado a la plataforma Wyscout, debido a su alto costo de suscripción. Además, la selección de equipos para el análisis, que se basó inicialmente

en la popularidad, podría introducir un sesgo en el estudio. Otro aspecto a considerar es el marco temporal empleado para cada equipo, estimado en una ventana de cuatro años. Aunque esta aproximación es válida, realizar una clusterización anual podría ofrecer una perspectiva más detallada sobre cómo varían los estilos de juego de los equipos a lo largo del tiempo.

Bibliografía

1. Pappalardo, L., Cintia, P., Ferragina, P., Massucco, E., Pedreschi, D., & Giannotti, F. (2018). PlayeRank: Multi-dimensional and role-aware rating of soccer player performance. ArXiv
2. Herberger, T., & Litke, C. (2021). The Impact of Big Data and Sports Analytics on Professional Football: A Systematic Literature Review. Springer Proceedings in Business and Economics, 147 - 171.
3. González Ramos, P., Martín Agüero, O., Montero Quesada, J., & Rice Nelson, D. (2021). Guía para el control observacional de la táctica grupal ofensiva en el fútbol. Revista de Ciencia y Tecnología en la Cultura Física, 162 - 176.
4. Albarrán Jardón, E. (2020). Inteligencia deportiva: tecnología aplicada al deporte. Pensamiento libre. Estado de Meximo: Pensamiento Libre. Dufour, M., Phillips, J., &
5. Ernwein, V. (2017). What makes the difference? Analysis of the 2014 World. Journal of Human Sport and Exercise, 616-629. Gudmundsson, J., & Horton, M.

(2018). Spatio-Temporal Analysis of Team Sports . ACM Computing Surveys (CSUR), 1-34. Brooks,

6. J., Kerr, M., & Guttag, J. (2016). Developing a Data-Driven Player Ranking in Soccer Using Predictive Model Weights. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 49-55
7. Oliver Shulte and Zeyu Zhao. 2017. Apples-to-Apples: Clustering and Ranking NHLPlayers Using Location Information and Scoring Impact. In MIT Sloan Sports Analytics Conference. Hynes Convention Center, Boston, MA, US
8. Hosmer, David W.; Lemeshow, Stanley (2000). Applied Logistic Regression (2nd ed.). Wiley. ISBN 978-0-471-35632-5.