



Universidad Nacional Mayor de San Marcos

Universidad del Perú. Decana de América

Facultad de Ciencias Matemáticas

Escuela Profesional de Estadística

**Aplicación del análisis de conglomerados para la
segmentación de grupos en el fútbol base para una
entidad deportiva en el año 2022**

TRABAJO DE SUFICIENCIA PROFESIONAL

Para optar el Título Profesional de Licenciado en Estadística

AUTOR

Jorge Luis BALTA PAU

ASESOR

Dra. Ofelia ROQUE PAREDES

Lima, Perú

2023



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

Referencia bibliográfica

Balta, J. (2023). *Aplicación del análisis de conglomerados para la segmentación de grupos en el fútbol base para una entidad deportiva en el año 2022*. [Trabajo de Suficiencia Profesional de pregrado, Universidad Nacional Mayor de San Marcos, Facultad de Ciencias Matemáticas, Escuela Profesional de Estadística]. Repositorio institucional Cybertesis UNMSM.

Metadatos complementarios

Datos de autor	
Nombres y apellidos	Jorge Luis Balta Pau
Tipo de documento de identidad	DNI
Número de documento de identidad	70671670
URL de ORCID	https://orcid.org/0009-0001-3318-0410
Datos de asesor	
Nombres y apellidos	Ofelia Roque Paredes
Tipo de documento de identidad	DNI
Número de documento de identidad	06243124
URL de ORCID	https://orcid.org/0000-0001-8280-021X
Datos del jurado	
Presidente del jurado	
Nombres y apellidos	Zoraida Judith Huamán Gutiérrez
Tipo de documento	DNI
Número de documento de identidad	09890094
Miembro del jurado 1	
Nombres y apellidos	Hugo Marino Rodríguez Orellana
Tipo de documento	DNI
Número de documento de identidad	40162362
Datos de investigación	
Línea de investigación	A.3.2.1 Análisis Multivariante

Grupo de investigación	No aplica.
Agencia de financiamiento	Sin financiamiento.
Ubicación geográfica de la investigación	Universidad Nacional Mayor de San Marcos País: Perú Departamento: Lima Provincia: Lima Distrito: Lima Coordenadas geográficas Latitud: -12.058333 Longitud: -77.083333
Año o rango de años en que se realizó la investigación	Año 2022
URL de disciplinas OCDE	Matemáticas aplicadas https://purl.org/pe-repo/ocde/ford#1.01.02



UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS

Universidad del Perú. Decana de América
FACULTAD DE CIENCIAS MATEMÁTICAS
ESCUELA PROFESIONAL DE ESTADÍSTICA

**ACTA DE SUSTENTACIÓN DEL TRABAJO DE SUFICIENCIA PROFESIONAL
PARA LA OBTENCIÓN DEL TÍTULO PROFESIONAL DE LICENCIADO EN
ESTADÍSTICA
(PROGRAMA DE TITULACIÓN PROFESIONAL 2023)**


En la UNMSM – Ciudad Universitaria – Facultad de Ciencias Matemáticas, siendo las *16...* horas del sábado 21 de octubre del 2023, se reunieron los docentes designados como Miembros del Jurado Evaluador (PROGRAMA DE TITULACIÓN PROFESIONAL 2023): Dra. Zoraida Judith Huamán Gutiérrez (PRESIDENTE), Mg. Hugo Marino Rodríguez Orellana (MIEMBRO) y la Dra. Ofelia Roque Paredes (MIEMBRO ASESOR), para la sustentación del Trabajo de Suficiencia Profesional titulado: “**APLICACIÓN DEL ANÁLISIS DE CONGLOMERADOS PARA LA SEGMENTACIÓN DE GRUPOS EN EL FÚTBOL BASE PARA UNA ENTIDAD DEPORTIVA EN EL AÑO 2022**”, presentado por el señor **Bachiller JORGE LUIS BALTA PAU**, para optar el Título Profesional de Licenciado en Estadística.


Luego de la exposición del Trabajo de Suficiencia Profesional, la Presidente invitó al expositor a dar respuesta a las preguntas formuladas.

Realizada la evaluación correspondiente por los Miembros del Jurado Evaluador, el expositor mereció la aprobación *Sobresaliente*, con un calificativo promedio de *Diecisiete (17)*

A continuación, los Miembros del Jurado Evaluador dan manifiesto que el participante **Bachiller JORGE LUIS BALTA PAU**, en vista de haber aprobado la sustentación de su Trabajo de Suficiencia Profesional, será propuesto para que se le otorgue el Título Profesional de Licenciado en Estadística.

Siendo las *16:30* horas se levantó la sesión firmando para constancia la presente Acta.


Dra. Zoraida Judith Huamán Gutiérrez
PRESIDENTE


Mg. Hugo Marino Rodríguez Orellana
MIEMBRO

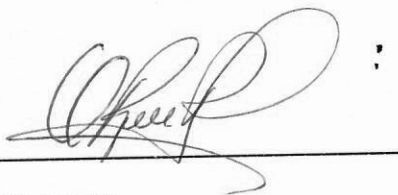

Dra. Ofelia Roque Paredes
MIEMBRO ASESOR

CERTIFICADO DE SIMILITUD

Yo, Ofelia Roque Paredes en mi condición de asesora acreditada con Resolución Decanal N° 001606-2023-D-FCM/UNMSM del Trabajo de Suficiencia Profesional, cuyo título es "APLICACIÓN DEL ANÁLISIS DE CONGLOMERADOS PARA LA SEGMENTACIÓN DE GRUPOS EN EL FÚTBOL BASE PARA UNA ENTIDAD DEPORTIVA EN EL AÑO 2022", presentado por el bachiller JORGE LUIS BALTA PAU, para optar el título de Licenciado en Estadística.

Certifico que se ha cumplido con lo establecido en la Directiva de Originalidad y de Similitud de Trabajos Académicos, de Investigación y Producción Intelectual. Según la revisión, análisis y evaluación mediante el software de similitud textual, el documento evaluado cuenta con el porcentaje de **13%** de similitud, nivel **PERMITIDO** para continuar con los trámites correspondientes y para su **publicación en el repositorio institucional**.

Se emite el presente certificado en cumplimiento de lo establecido en las normas vigentes, como uno de los requisitos para la obtención del título correspondiente.



DNI: 06243124

Dra. Ofelia Roque Paredes



Huella Digital

RESUMEN

En el presente trabajo se utiliza la técnica estadística del análisis de conglomerados para agrupar jugadores de categoría o edad formativa, que participan en un torneo de fútbol. En los últimos años se ha visto la utilidad que tiene la estadística para diferentes áreas de investigación, esto incluye la aplicación en los deportes, ya sea aplicada en niños, jóvenes y adultos que se dedican al fútbol de manera amateur y profesional, siendo en esta ocasión aplicada en el fútbol base.

Detallaremos el uso de la estadística a través del análisis de conglomerados jerárquico a través del método de Ward y del análisis de conglomerados no jerárquico K-medias, aplicado en el software R Studio, en jóvenes de las categorías 12 a 16 años pertenecientes al Centro Formativo Benjamins FC. Los cuales participaron en el torneo Federación Regional 2022, quedando en el puesto de Sub-campeón. Mostrando así, la importancia de la técnica utilizada para observar patrones que ayuden a mejorar el rendimiento de los jóvenes en futuros partidos.

ABSTRACT

In the present work, the statistical technique of cluster analysis is used to group players of educational category or age, who participate in a soccer tournament. In recent years, the usefulness of statistics for different research areas has been seen, this includes the application in sports, whether it is applied to children, youth and adults who are engaged in amateur and professional soccer, being in this occasion applied in grassroots football We will detail the use of statistics through the hierarchical cluster analysis through the Ward method and the non-hierarchical K-means cluster analysis, applied in the R Studio software, in youths from categories 12 to 16 years belonging to the Benjamins FC Training Center. Which participated in the 2022 Regional Federation tournament, remaining in the position of Runner-up Thus showing the importance of the technique used to observe patterns that help improve the performance of young people in future matches.

*Dedicado a mis padres, mi soporte. A mis abuelas y a la
universidad por los grandes momentos dentro y fuera de
sus aulas.*

AGRADECIMIENTO

Al Centro de Formación Benjamins FC, por su confianza en darme un espacio donde aplicar mis conocimientos y permitirme colaborar en el trabajo del fútbol de menores, de esa manera contribuir con el desarrollo de nuestra sociedad.

INDICE

I.	INTRODUCCIÓN	7
II.	DESCRIPCIÓN DE LA ACTIVIDAD	8
2.1	Reseña breve de la empresa	8
2.2	Organigrama	9
2.3	Problemática	9
2.4	Objetivo principal	10
2.5	Objetivos específicos	10
2.6	Breve descripción de la metodología	10
2.7	Alcances y límites de la investigación	11
2.8	Justificación	11
2.8.1	Justificación social	12
2.8.2	Justificación metodológica	12
III.	MARCO TEÓRICO	13
3.1	Análisis de Conglomerados	13
3.2	Análisis de conglomerados jerárquico	14
3.3	Método de Ward	14
3.4	Análisis de conglomerados no jerárquico	16
3.5	Método jerárquico de K-medias	16
3.6	Análisis exploratorio de datos	17
3.6.1	Método del codo	17
3.6.2	Coeficiente de la silueta	17
3.6.3	Criterio de C-H	
3.7	Fútbol	18
3.8	Fútbol base	18
3.9	Escuela de fútbol o academia	18
3.10	Copa federación regional	19
3.11	Base de datos	20
3.12	Descripción de la base de datos	20
3.13	Mapa conceptual	22
3.14	Antecedentes Internacionales	26
IV.	METODOLOGÍA	28
4.1	Resultados	36

V.	CONCLUSIONES	40
VI.	RECOMENDACIONES	41
VII.	BIBLIOGRAFÍA	42
VIII.	ANEXOS	
8.1	Anexo 1	
8.2	Anexo 2	
8.3	Anexo 3	
8.4	Anexo 4	

I. INTRODUCCIÓN

En el ámbito futbolístico es de conocimiento que, para lograr gran nivel de competencia, la formación de atletas inicia en el fútbol base, siendo esta conocida como divisiones menores o canteras. Es la etapa formativa y la más importante para todos los que dedican parte de su vida a jugar fútbol. Es por ello que existen ligas en países europeos (Inglaterra, Alemania, Italia, etc.) y a nivel sudamericano (Brasil, Argentina, etc.) que demuestran tener grandes resultados, con una cantidad de futbolistas de alto nivel que sobrepasa el nuestro; siendo para nosotros una limitante al momento de competir con selecciones que si trabajan en la formación de sus divisiones menores. A través de este TSP detallo como es que la estadística con la aplicación del análisis de conglomerados, hoy en día sería muy útil para colaborar en la formación de futuros futbolistas profesionales.

Desarrollamos nuestra aplicación en el Centro de Formación Benjamins FC, academia que viene formando divisiones menores desde el 2018, a la cual llamaremos Academia BFC.

En el año 2022 la Academia BFC participó en el torneo Copa Federación Regional, a nivel de todo Lima Metropolitana, quedando en segundo lugar y obteniendo el ascenso a Copa Federación Elite 2023. El TSP contiene:

En primer lugar, hablaremos sobre la empresa en la que me desempeño, en el segundo capítulo se presenta la descripción de las actividades que se realizan y desarrollaremos la problemática principal en la formación desde el fútbol base.

En el tercer capítulo será detallado el marco teórico, los conceptos que se utilizan en este análisis, los antecedentes nacionales e internacionales y la aplicación de la técnica estadística.

En el cuarto capítulo veremos la metodología utilizada en el programa R Studio, para la obtención de datos y los resultados de dicha aplicación. En el quinto capítulo tendremos las conclusiones de los resultados.

En la última parte tenemos las recomendaciones, la bibliografía que nos ha ayudado en la elaboración de este TSP y los anexos correspondientes.

II. DESCRIPCIÓN DE LA ACTIVIDAD

2.1 Reseña Breve de la Empresa

El CF Benjamins FC, a la cual llamamos Academia BFC, es un proyecto dirigido a todos los niños que deseen mejorar su rendimiento deportivo, así como sus habilidades y manejo de los gestos técnicos de este deporte.

Es una academia en donde la formación de los niños es guiada por auténticos especialistas en la dirección y el control del entrenamiento, a través de una metodología adecuada a cada una de las edades.

Los valores en el programa deportivo de Academia BFC son la base en nuestra metodología, que se propone para la práctica son: Respeto, juego limpio (fair play), responsabilidad, deseo de superación, espíritu de lucha, solidaridad y puntualidad.

Misión

Incentivar, educar y formar integralmente a la niñez y juventud a través de la práctica del fútbol mediante una enseñanza especializada en entrenamiento para el fútbol base, con metodología apropiada para cada edad, a cargo de una calificada plana de formadores cuyo objetivo es prepararlos adecuadamente para el futuro; brindándole las herramientas necesarias para desarrollar sus habilidades, destrezas, competencias y valores.

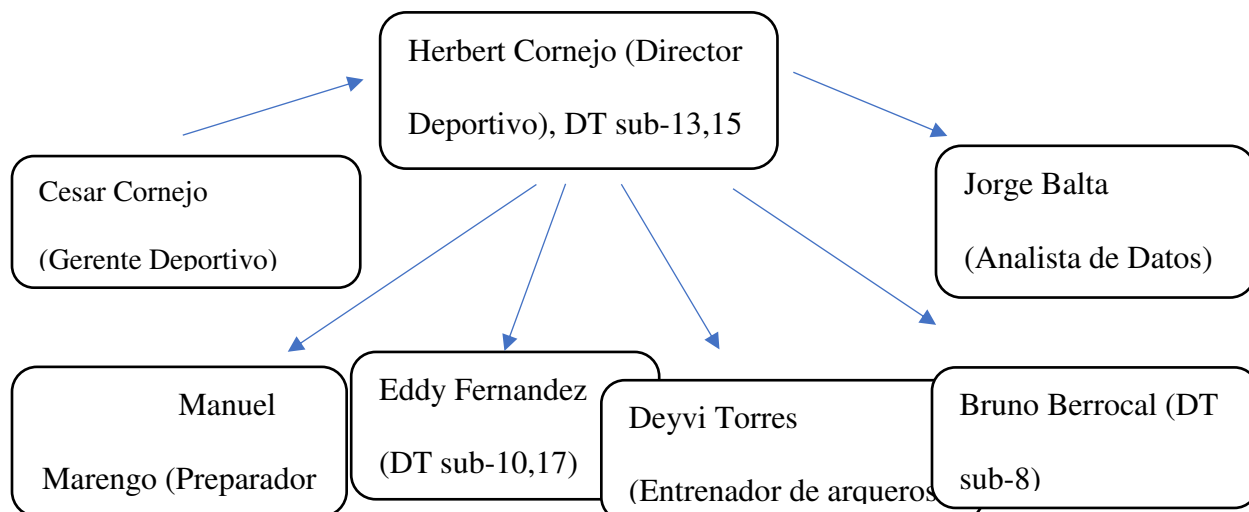
Visión

Educar a través del deporte. Aspiramos a constituirnos en un centro modelo de formación orientado al fútbol base que tenga trascendencia y que nos permita aportar con el desarrollo social, educativo y deportivo de la comunidad por el bien del juego y de la sociedad.

2.2 Organigrama

Figura 1

Organigrama de la Academia BFC



Nota. La figura muestra el personal de labores que conforma la Academia Benjamins FC.

2.3 Problemática

El momento futbolístico en el año del torneo fue muy favorable para la Academia BFC, los resultados fueron positivos para el equipo y con ello consiguieron jugar la final del torneo, actualmente la Academia BFC viene participando en juegos amistosos y campeonatos relámpagos, en un nivel superior al que se jugaron en el torneo del año 2022; razones por las cuales se está buscando factores que influyan en la preparación técnica y física requerida para la

competencia de este año. Es donde puede intermediar el análisis de conglomerados para formar grupos con características parecidas y potenciar el rendimiento de cada jugador.

2.4 Objetivo principal

Segmentar a los jóvenes que pertenecen al equipo competitivo de la Academia BFC, en el torneo copa federación 2022 utilizando en análisis de conglomerados jerárquico y no jerárquico.

2.5 Objetivos específicos

- Determinar las variables adecuadas que permiten realizar la segmentación.
- Determinar el método jerárquico de Ward
- Determinar el método no jerárquico K- medias

2.6 Breve descripción de la metodología

Tenemos un total de 29 jóvenes que pertenecen a los equipos competitivos de la Academia BFC, para obtener los datos de ellos en partidos tenemos nuestro método, el cual inicia en la obtención de datos, a través de videos filmados en los partidos. Como segundo paso, descargamos todos los eventos ocurridos en una base de datos del software SPSS, en este software realizamos los análisis cuantitativos respectivos para luego ser entregados y explicados a los cuerpos técnicos y ellos tratar de reflejar los resultados obtenidos, aplicándolos en los entrenamientos para potenciar e incrementar su interacción en los futuros partidos.

3 Alcances y límites de la investigación

En el proceso de recolectar información mediante videos e imágenes de los partidos, existen algunas complicaciones en temas de ubicación, ya que algunos campos deportivos que visitamos no cuentan con una plataforma o tribunas donde se pueda obtener las imágenes de un buen ángulo, es necesario el uso de una plataforma que eleve la cámara a una altura la cual distinga a los jugadores para obtener los datos precisos a analizar.

La tecnología es esencial para un trabajo mas automatizado, para esto se necesitan herramientas como los softwares de video análisis, chalecos con dispositivos GPS, el cual nos ayuda a medir su rendimiento físico, dispositivos electrónicos con memoria de video amplia, como laptop o Tablet.

4. Justificación

Como aficionado del futbol y estadístico profesional, trato de contribuir mediante este TSP con una idea para los entrenadores de fútbol de menores o profesionales y así se apoyen en la estadística, a través del análisis de conglomerados, a tomar decisiones que ayuden a contribuir en el desarrollo profesional de todos los jóvenes dedicados a este deporte.

4.1 Justificación social

El futbol peruano no es competitivo a nivel internacional a comparación con la mayoría de países de nuestro continente, y eso también ocurre a nivel de futbol de menores, por ello con este trabajo trato de aportar una idea, en base a estadística y al análisis de conglomerados, la cual podría ser utilizada por diferentes entrenadores de fútbol, para formar grupos de jugadores y lograr repotenciar su rendimiento a nivel competitivo profesional.

4.2 Justificación metodológica

La manera efectiva de obtener datos es a través de las grabaciones en cada partido, luego de acabado el campeonato realizamos dichos análisis en el software R Studio. Dichos resultados servirán para optimizar el trabajo en los entrenamientos y lograr mejores resultados a futuro.

III. MARCO TEÓRICO

A continuación, presentaremos las definiciones de los términos estadísticos y deportivos utilizados en el trabajo:

3.1 Análisis de Conglomerados

El análisis de conglomerados implica agrupar un conjunto de objetos en grupos significativos y útiles. Es una noción utilizada para indicar que tipo de agrupaciones estamos tratando de identificar. Los métodos de análisis de conglomerados clasifican grupos de casos o elementos, en base a criterios cualitativos o cuantitativos. A veces, en lugar de los casos, se forman grupos con las variables. Existe en análisis de conglomerados jerárquicos y no jerárquicos. La agrupación no se refiere a algoritmos específicos, sino que es un proceso para crear grupos en la medida de similitud (Singh, 2017).

Según Pérez Márquez (2015), el análisis de conglomerados “constituye uno de los procedimientos estadísticos más utilizados hoy en día para la segmentación, su finalidad es revelar concentraciones de los datos (casos o variables) para su agrupamiento eficiente en conglomerados según su similitud. Cuando se tiene una muestra de individuos, de cada uno de los cuales se dispone una serie de observaciones, el análisis de conglomerados sirve para clasificarlos en grupos más homogéneos posibles en base a las variables observadas

En este trabajo, la aplicación que tiene este análisis en el fútbol base son principalmente: la segmentación de jugadores, cuando los jugadores se agrupan en base a su semejanza de acuerdo a su estilo de juego respecto a las variables seleccionadas y también el comportamiento del jugador, cuando se quiere identificar jugadores homogéneos.

3.1.1 Análisis de conglomerados jerárquico

En este análisis, se intenta identificar grupos relativamente homogéneos basándonos en características seleccionadas, no tenemos una idea “a priori” de cuantos grupos podemos formar con los datos obtenidos. Entre los diferentes métodos jerárquicos para la formación de grupos, nosotros utilizaremos el mas común, que es el método entre grupos.

Permite aglomerar tanto casos como variables y elegir entre una gran variedad de métodos de aglomeración y medidas de distancia.

3.1.2 Método de Ward:

También llamado eslabonamiento de Ward, es un método de conglomerados jerárquico cuyo algoritmo es también conocido como el de mínima varianza ya que el objetivo es unir aquellos grupos que resulten en el menor incremento en la varianza de los grupos.

El uso del método de Ward está bastante difundido porque suele discriminar mejor entre grupos en relación a otros métodos jerárquicos (Arroyo-Borja, 2018).

Este método fue difundido por Ward (1963), quien argumentó que los conglomerados debían constituirse, de tal manera que, al fundirse dos elementos, la pérdida de información resultante de la fusión sea mínima. En este contexto, la cantidad de información se cuantifica como la suma de distancias al cuadrado de cada elemento respecto al centroide del conglomerado al que pertenece. Para ello se comienza calculando, en cada conglomerado, el vector de medias de todas las variables, es decir, el centroide. Luego se calculan las distancias euclídeas al cuadrado entre cada elemento y centroides de todos los conglomerados. Por último, se suman las distancias correspondientes a todos los elementos.

En cada paso se unen aquellos conglomerados (o elementos) que dan lugar a un menor incremento de la SCE, es decir, de la suma de cuadrados de las distancias intra-conglomerado.

3.1.3 Análisis de conglomerados no jerárquico

Dentro de este método, tenemos una idea “a priori” de cuantos grupos podemos formar, y las observaciones son asignadas a esos grupos, de tal manera que se maximiza la homogeneidad de los sujetos asignados a un mismo grupo. Según Perez (2015), el análisis conglomerado no jerárquico precisa que el investigador fije de antemano el número de clusters en que quiere agrupar sus datos. En algunas ocasiones, nos basamos en la cantidad de grupos que resultaron en el análisis de conglomerados jerárquicos para iniciar el análisis con este método. Existen diferentes tipos de métodos no jerárquicos, el que se utiliza en este trabajo es el método de K-medias.

3.1.4 Método no jerárquico k-medias

Es el algoritmo de agrupamiento más usado, para utilizarlo debemos especificar antes el número de grupos que deseamos encontrar, k . Las distintas etapas de este método son:

1. Elección de los k centroides de manera aleatoria.
2. Asignación de cada muestra al centroide que se encuentra más cerca de ésta.
3. Actualización de los k centroides al valor medio de entre todos los puntos que conforman el cluster. Se repite los dos últimos pasos hasta que no haya diferencia con la iteración anterior (Junquera, 2021-22).

Principalmente, como se dice, es necesario especificar la cantidad de grupos que deseamos encontrar, para esto estudiaremos tres métodos efectivos para estimar el valor óptimo de conglomerados “ k ”: El método del codo, el coeficiente de la silueta y el Criterio C-H.

3.2 ANÁLISIS EXPLORATORIO DE DATOS:

Para saber que cantidad de grupos es necesario para nuestro estudio, necesitamos corroborarlo basándonos en métodos estadísticos que nos ayudan a demostrar científicamente lo que buscamos, para esto utilizaremos los 3 métodos nombrados anteriormente.

3.2.1 Método del Codo

El método del codo usa los valores de la inercia una vez se ha aplicado k-medias con un número “n” de conglomerados. La inercia es la suma de las distancias al cuadrado de cada punto del conglomerado a su centroide.

La grafica se asemeja a un brazo, entonces el punto donde notemos una atenuación (el codo del brazo) en la inercia, representará el número óptimo de conglomerados (Pineda,2022).

3.2.2 Coeficiente de Silueta

En el análisis exploratorio de datos que se efectúa para saber la cantidad de grupos optimas, Este análisis es utilizado como una herramienta gráfica para representar una medida del grado de agrupación de los ejemplos en los conglomerados (Raschka, 2023).

3.3 Fútbol

En la actualidad, es uno de los deportes más practicados en casi todo el mundo a nivel amateur y profesional. En el cual se requieren de un manejo administrativo y futbolístico, disciplinas que son necesarias para el sostenimiento, donde este último presenta varias categorías según la edad de los participantes, entre ellas las divisiones menores conocidas como futbol base.

3.3.1 Fútbol Base

Es la etapa formativa de este deporte, humana, física y técnica, conocida como las divisiones menores o canteras, infantiles hasta juveniles de 18 años. Donde existen a nivel

mundial, diferentes academias formativas, clubes y profesionales particulares, que comparten y aplican su conocimiento para la formación adecuada de los que participan.

3.3.2 Escuela de Fútbol o Academia

Una escuela de futbol es una especial organización de la participación de los jóvenes deportistas, que trata con metodologías, el logro de los objetivos físico-técnicos y de las restantes áreas formativas, pero respetando y favoreciendo el natural proceso evolutivo de esos jóvenes. Existen academias con fútbol de carácter recreativo, donde la participación de los niños es ligada al entretenimiento y las de carácter competitivo, donde la progresión deportiva sirve de apoyo a la formación integral y favoreciendo a la salud del joven futbolista, como de aquellos que tienen un objetivo mas intenso en las mejoras técnicas y tácticas y de formación profesional (Wanceulen,2017).

3.3.3 Copa Federación Regional

Torneo en el cual participan todos los clubes y academias que desean inscribirse, esto es a nivel de todo el país, dirigida por la comisión de menores de la Liga Departamental de Futbol de Lima, ente afiliado a la Federación Peruana de Futbol (FPF). En dicho año la Copa se llamo “Copa Centenario FPF” y dentro de sus disposiciones generales tiene artículos, entre los cuales se encuentran:

Artículo 1.- Los Campeonatos de Menores Masculino son organizados, coordinados, supervisados, deportiva y administrativamente, por la Federación Peruana de Fútbol, a través de las Ligas Departamentales de Fútbol.

Artículo 2.- Los Campeonatos de Fútbol de Menores Masculino se desarrollarán en todas las Regiones del país, bajo la normativa vigente estipulada por las autoridades estatales competentes en cada región, respecto a salud pública y bioseguridad.

Entre otros artículos (anexo B y C), son parte del reglamento o bases para disputar el torneo, dirigidos para los equipos, los jugadores y árbitros.

3.4 Análisis de la base de datos

Este bloque consta de dos partes que se corresponden a la descripción de la base de datos y al tratamiento de la misma para su posterior análisis.

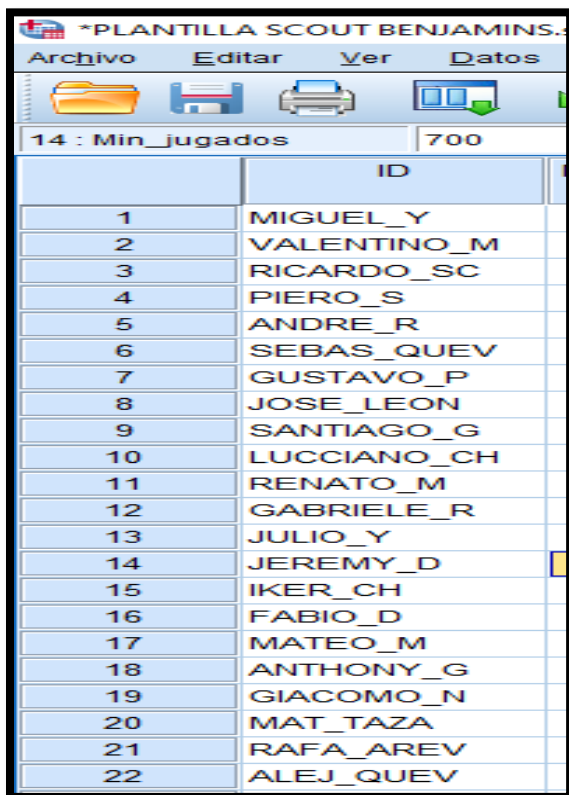
3.4.1 Descripción de la base de datos

La base de datos consta de 29 jóvenes y 39 variables explicativas, de las cuales el nombre del jugador es la única variable cualitativa. El resto son variables cuantitativas y recogen la información sobre los eventos que ocurren en el juego, por ejemplo: número de pases efectivos, duelos ganados, número de goles, quites de pelota, etc.

En la siguiente figura, están abreviados los datos de los jóvenes pertenecientes a los equipos competitivos de la Academia BFC.

Figura 2

Lista de los jóvenes que pertenecen al equipo competitivo de Academia BFC.



	ID
1	MIGUEL_Y
2	VALENTINO_M
3	RICARDO_SC
4	PIERO_S
5	ANDRE_R
6	SEBAS_QUEV
7	GUSTAVO_P
8	JOSE_LEON
9	SANTIAGO_G
10	LUCCIANO_CH
11	RENATO_M
12	GABRIELE_R
13	JULIO_Y
14	JEREMY_D
15	IKER_CH
16	FABIO_D
17	MATEO_M
18	ANTHONY_G
19	GIACOMO_N
20	MAT_TAZA
21	RAFA_AREV
22	ALEJ_QUEV

Nota. Relación de jóvenes que jugaron el torneo Copa Federación Regional, para la Academia BFC. (Anexo A)

Figura 3

Imagen referencial de la base de datos

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	ID	Cadena	15	0		Ninguno	Ninguno	11	Izquierda	Nominal	Entrada
2	Min_jugados	Numérico	8	0		Ninguno	Ninguno	8	Derecha	Escala	Entrada
3	Min_posibles	Numérico	8	0		Ninguno	Ninguno	8	Derecha	Escala	Entrada
4	Part_jugados	Numérico	8	0		Ninguno	Ninguno	8	Derecha	Escala	Entrada
5	Part_posibles	Numérico	8	0		Ninguno	Ninguno	8	Derecha	Escala	Entrada
6	Part_titular	Numérico	8	0		Ninguno	Ninguno	8	Derecha	Escala	Entrada
7	Goles	Numérico	8	0		Ninguno	Ninguno	8	Derecha	Escala	Entrada
8	PROM_GOLXPART	Numérico	8	3		Ninguno	Ninguno	8	Derecha	Escala	Entrada
9	Gol_penal	Numérico	8	0		Ninguno	Ninguno	8	Derecha	Escala	Entrada
10	Gol_pieDer	Numérico	8	0		Ninguno	Ninguno	8	Derecha	Escala	Entrada
11	Gol_pielzq	Numérico	8	0		Ninguno	Ninguno	8	Derecha	Escala	Entrada
12	Asistencias	Numérico	8	0		Ninguno	Ninguno	8	Derecha	Escala	Entrada
13	Toque_balon	Numérico	8	0		Ninguno	Ninguno	8	Derecha	Escala	Entrada
14	Pases	Numérico	8	0		Ninguno	Ninguno	8	Derecha	Escala	Entrada
15	Pases_claves	Numérico	8	0		Ninguno	Ninguno	8	Derecha	Escala	Entrada
16	Pases_precisos	Numérico	8	0		Ninguno	Ninguno	8	Derecha	Escala	Entrada
17	Pases_propia_cancha	Numérico	8	0		Ninguno	Ninguno	8	Derecha	Escala	Entrada
18	Pases_largos_precisos	Numérico	8	0		Ninguno	Ninguno	8	Derecha	Escala	Entrada
19	Pases_medDist_precisos	Numérico	8	0		Ninguno	Ninguno	8	Derecha	Escala	Entrada
20	Pases_errados	Numérico	8	0		Ninguno	Ninguno	8	Derecha	Escala	Entrada
21	Despejes	Numérico	8	0		Ninguno	Ninguno	8	Derecha	Escala	Entrada
22	Tiros_a_puerta	Numérico	8	0		Ninguno	Ninguno	8	Derecha	Escala	Entrada
23	Quites_pelota	Numérico	8	0		Ninguno	Ninguno	8	Derecha	Escala	Entrada
24	Tackles_partido	Numérico	8	0		Ninguno	Ninguno	8	Derecha	Escala	Entrada

Nota. La base cuenta con 29 jóvenes registrados y 39 variables explicativas

3.5 Mapa conceptual:

Son 39 variables explicativas, ellas son elegidas en base a los eventos ocurridos dentro de los partidos, variables las cuales describen las cualidades técnicas, físicas y tácticas del jugador. Cada variable presente en la base de datos lleva a resaltar las características similares de cada jugador al momento de competir en el campeonato y eso colabora de manera que el cuerpo técnico les encuentre la posición idónea dentro del campo.

Distribuimos las variables en 3 niveles, los cuales se definen como defensivos, ofensivos y creación de juego. Estas variables son registradas de manera cuantitativa en el SPSS para el análisis y posteriormente la presentación de informes al cuerpo técnico (tenemos la definición de cada variable en el Anexo D).

A continuación, se presentan diferentes investigaciones desde el campo internacional relacionados con el problema de investigación

3.3.1 Antecedentes internacionales

Malagon, Selma en el 2019 en su investigación titulada *Machine Learning en el Futbol* dijo, "En el método k-means es el analista quien debe escoger el valor de k" (número de grupos), en esta investigación, Malagon busca analizar el rendimiento en el terreno de juego de jugadores profesionales de las grandes ligas europeas de fútbol durante la temporada 2017-18 a partir de estrategias y técnicas de Machine Learning y Big Data como el análisis de componentes principales, análisis clúster, gráficos de contribuciones y gráficos de radar, entre otros métodos.

Garrido Tapias en las conclusiones de su trabajo de fin de grado año 2021-2022, llamado "Uso de técnicas de clustering para encontrar perfiles de jugadores en una competición de fútbol profesional" de la Universidad de Valladolid dice, "la influencia que tiene el parámetro k en el algoritmo k-medias, ya que cuanto menor era el valor de éste los cluster obtenidos se acercaban más hacia las posiciones más generales en el mundo del fútbol que son: portero, defensa, mediocentro y delantero. En cambio, cuanto más grande ha sido k nos ha permitido ver como se dividían grupos que parecían fijos en un primer momento. Por tanto, respecto a este parámetro podemos concluir que cuanto menor sea k mayor será el número de habilidades o aspectos del juego en los que destacarán los jugadores que formen los clusters. Mientras que cuanto mayor sea el número de grupos preseleccionado lograremos clusters más específicos, alejándonos de la idea de jugadores "multiposición".

Existe un estudio en la plataforma campus bigdata, el cual titula *Big Data y en Análisis de Conglomerados: Descubriendo los Diferentes Perfiles de los Mediocentro – Informe*, en el cual agrupa jugadores mediocampistas respecto a sus características, defensivos, ofensivos y creadores de juego. Dicho informe dice que “el centro del campo es un lugar donde se abarcan casi todos los parámetros que medimos los jugadores donde muchos de ellos no tienen un perfil específico “. Utilizan el análisis de conglomerados no jerárquico, fijando ellos la cantidad de grupos previamente a formar, usando la librería “NbClust” en el R Studio.

3.3.2 Antecedentes nacionales

No contamos con investigaciones sobre análisis de conglomerados realizadas en el Perú.

IV. METODOLOGÍA

Todo inicia en el traslado al campo de juego o estadio, el cual puede ser en cualquier distrito de Lima Metropolitana. Nos ubicamos en la parte más alta a la que podamos tener acceso e instalamos nuestros equipos de filmación. El siguiente paso es iniciar la grabación del partido de las dos categorías y de manera telefónica estar en comunicación con el cuerpo técnico, indicándole las incidencias del partido desde un mejor ángulo. Como tercer paso, recopilamos los datos a través de videos filmados en los partidos amistosos y competitivos, los cuales son descargados a una hoja de datos en SPSS, posteriormente es cargada en el software R Studio para los análisis requeridos.

Para esta aplicación, utilizamos el análisis de conglomerados jerárquico (entre grupos) y no jerárquico (método de K-medias), en el software R Studio.

Iniciamos con la descargar de la base de datos que tenemos en SPSS al programa R studio, luego aplicaremos códigos cuya función va desde mostrar la estructura de nuestra base de datos, seguidamente del análisis exploratorio de datos el cual nos mostrara la cantidad óptima de conglomerados a formar y por ultimo la distribución de los jugadores a través de los métodos conglomerados jerárquico de Ward y del no jerárquico K-medias.

Figura 4

Inicio del análisis en R, cargamos la base de datos

```
#####
# Cargamos el conjunto de datos
library(haven)
datos <- read_spss("PLANTILLA SCOUT BENJAMINS.sav")
```

Nota. La figura muestra el código que carga una base de datos de SPSS en R.

Figura 5

Código para obtener la estructura de la base de datos.

```
> str(datos)
'data.frame': 29 obs. of 39 variables:
 $ ID : chr "MIGUEL_Y" "VALENTINO_M" "RICARDO_SC" "PIERO_S" ...
 ..- attr(*, "format.spss")= chr "A15"
 $ Min_jugados : num 880 560 880 880 880 880 880 880 575 800 825 ...
 ..- attr(*, "format.spss")= chr "F8.0"
 $ Min_posibles : num 880 880 880 880 880 880 880 880 880 880 ...
 ..- attr(*, "format.spss")= chr "F8.0"
 $ Part_jugados : num 11 10 11 11 11 11 11 10 11 11 ...
 ..- attr(*, "format.spss")= chr "F8.0"
 $ Part_posibles : num 11 11 11 11 11 11 11 11 11 11 ...
 ..- attr(*, "format.spss")= chr "F8.0"
 $ Part_titular : num 11 9 11 11 11 11 11 9 11 11 ...
 ..- attr(*, "format.spss")= chr "F8.0"
 $ Goles : num 0 0 1 1 2 2 1 3 5 7 ...
 ..- attr(*, "format.spss")= chr "F8.0"
 $ PROM_GOLXPART : num 0 0 0.0909 0.0909 0.1818 ...
```

Nota. Resultados de la estructura de la data.

Figura 6

Código que verifica presencia de datos perdidos

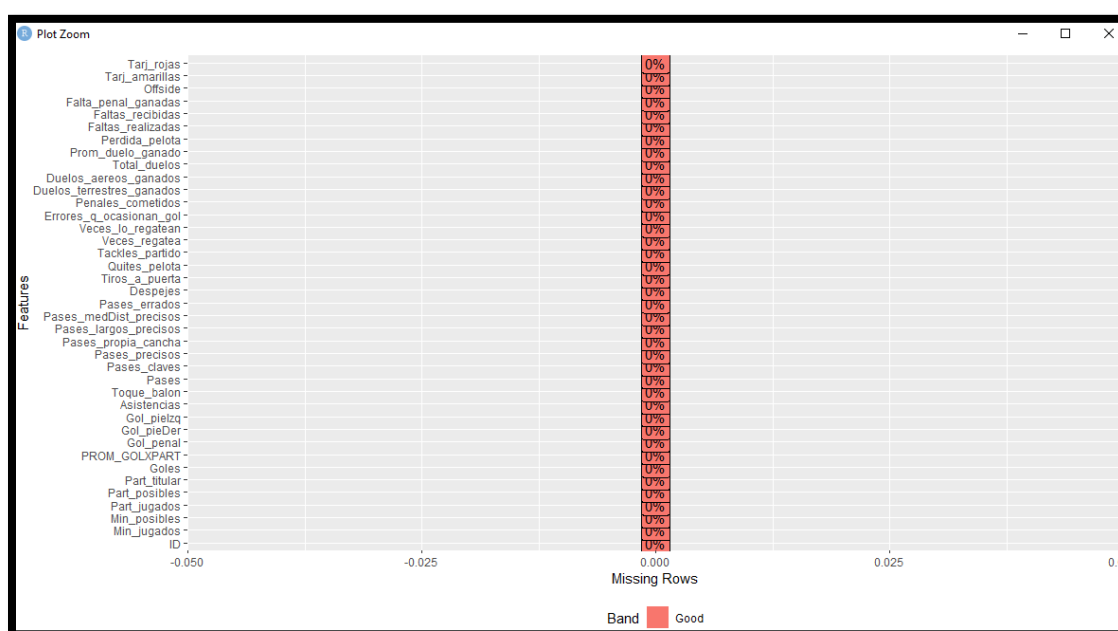
```
# Cambiamos a un data frame
datos <- as.data.frame(datos)
head(datos)

# verificación de datos perdidos
library(DataExplorer)
plot_missing(datos)
```


Es necesario hacer una revisión de datos perdidos o nulos, porque estos pueden traer problemas en el análisis.

Figura 7

Gráfico de datos perdidos o nulos.

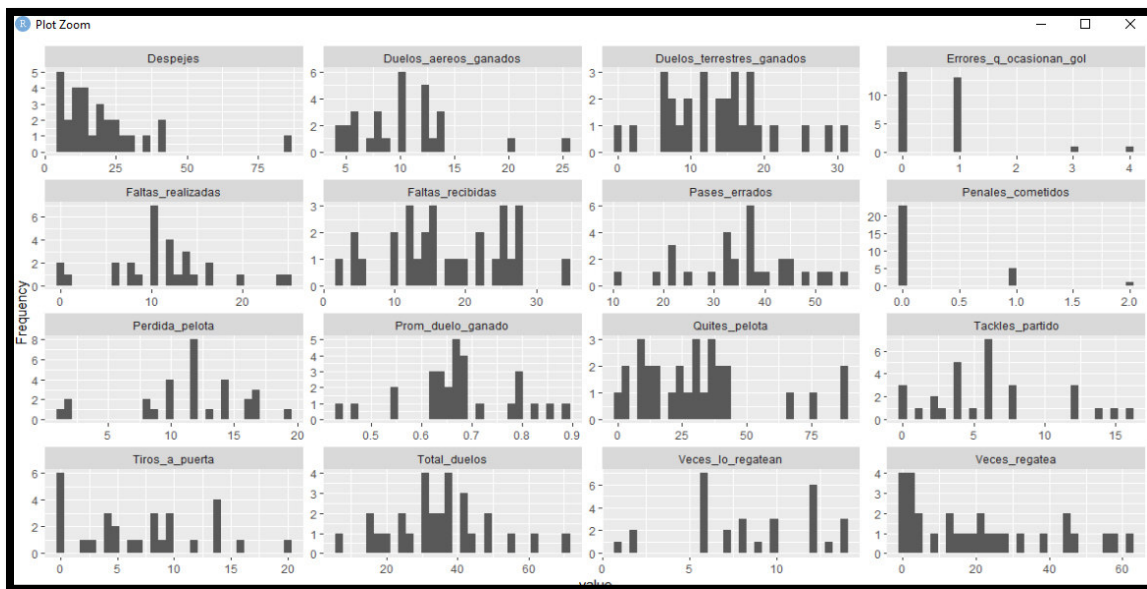


Nota. El gráfico referencia que no hay presencia de datos perdidos.

En la siguiente figura mostraremos histogramas de las variables numéricas, esto sirve para observar el comportamiento de cada variable.

Figura 8

Histograma de las variables numéricas



Luego de observar la estructura de la base de datos, tenemos algunas observaciones en las cuales hay variables que dentro de sus datos todas son iguales, por ejemplo, en tarjetas rojas todos tienen 0; otra manera de ver los datos repetitivos es en la cantidad de Minutos Posibles, donde todos los jugadores tienen la misma cantidad de jugar por igual.

Dado estos resultados en dichas variables, procedemos a realizar un nuevo análisis exploratorio, pero esta vez, descartando dichas variables que pueden influenciar en los resultados.

Figura 9

Análisis Exploratorio de Datos

```
> # .....
> datos1<-as.data.frame(scale(datos[,c(2,4,6:38)]))
> head(datos1)
  Min_jugados Part_jugados Part_titular      Goles PROM_GOLXPART Gol_penal
1  0.9697501  0.4955355  0.5823592 -1.0691418  -1.07940601 -0.4211174
2 -0.7776831 -0.1887754 -0.6239563 -1.0691418  -1.07940601 -0.4211174
3  0.9697501  0.4955355  0.5823592 -0.4728896  -0.49564105 -0.4211174
4  0.9697501  0.4955355  0.5823592 -0.4728896  -0.49564105 -0.4211174
5  0.9697501  0.4955355  0.5823592  0.1233625   0.08812391 -0.4211174
6  0.9697501  0.4955355  0.5823592  0.1233625   0.08812391 -0.4211174
  Gol_pieDer Gol_pieIzq Asistencias Toque_balon      Pases Pases_claves
1 -0.90033447 -0.4282161  -0.1225867  0.1353038 -0.1513705  -0.5523754
2 -0.90033447 -0.4282161  -1.0113405  0.6837932 -0.2650944  -0.8727531
3 -0.90033447 -0.4282161  -1.0113405  0.2196867 -1.0611619  -0.8727531
```

Nota. En este AED se excluye a las variables que no contribuyen en el análisis por ser repetitivas.

El siguiente paso dentro del AED, es verificar la cantidad de grupos a formar, esto lo haremos mediante el método del codo, el método de la silueta y el criterio de C-H.

Figura 10

Pasos para observar cantidad óptima de grupos, método del codo.

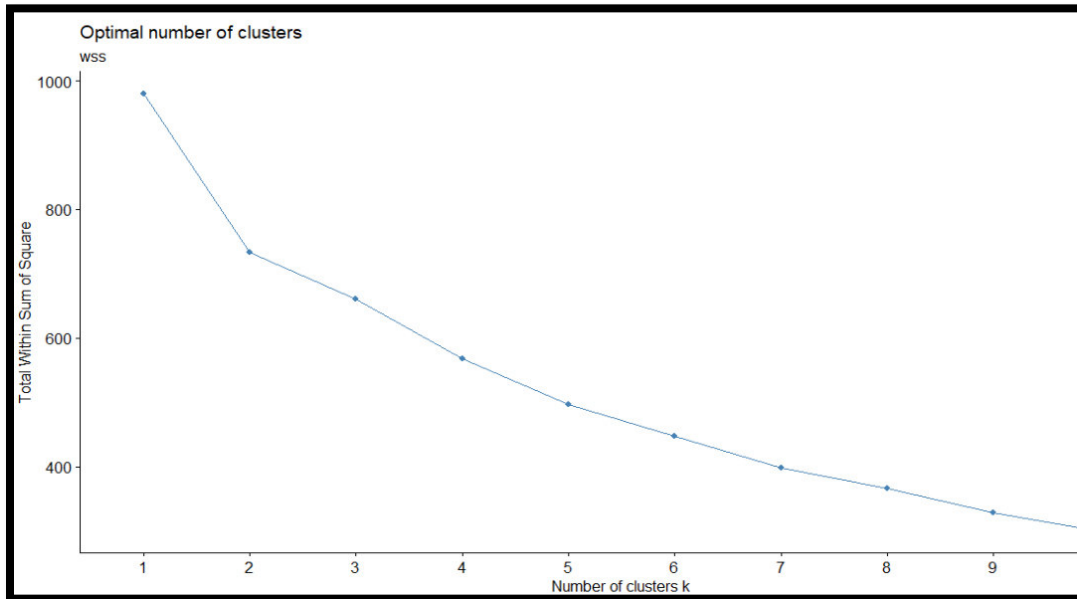
```
# Paso 3: Evaluar la cantidad optima de grupos
# usando el paquete factoextra
# Método 1: Método del codo
library(factoextra)
set.seed(123)
fviz_nbclust(datos1, kmeans, method = "wss") +
  labs(subtitle = "wss")
```

Nota. Código del método del codo

Dentro de los resultados, tenemos primero al método del codo el cual se refleja en la siguiente figura.

Figura 11

Gráfico del método del codo.



Nota. El método del codo muestra que la cantidad de grupos óptima es 2.

De esta manera, el método del codo nos dice que el número óptimo de conglomerados es 2, sin embargo, existen otras maneras de corroborar estos resultados y por eso mostramos cuánto se forma con el método de la silueta y criterio C-H.

Figura 12

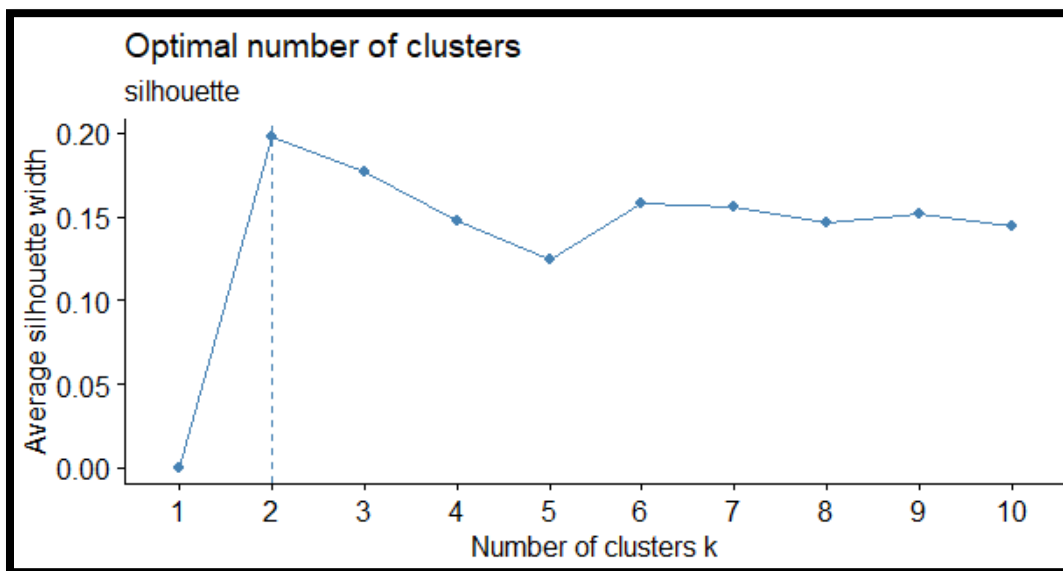
Método de la silueta.

```
# Método 2: silueta
fviz_nbclust(datos1, kmeans, method = "silhouette") +
  labs(subtitle = "silhouette")
```

Nota. Es un método confiable para la cantidad de grupos al igual que el del codo.

Figura 13

Gráfico del método de la silueta.



Nota. La marca nos indica la cantidad de conglomerados óptimo con este método.

De la misma manera, demostraremos cuantos grupos formamos bajo el criterio de C-H.

Figura 14

Código en R para el criterio C-H

```
> # Método 3: Criterio de C-H
> library(fpc)
> kmeansruns(datos1,criterion="ch")
K-means clustering with 2 clusters of sizes 17, 12
```

Nota. Método confiable para obtener la cantidad de grupos.

El criterio de C-H, el cual es un método más, al igual que el método del codo y de la silueta, que sirve de análisis para detectar la cantidad óptima de grupos a formar, nos da como resultados 2 conglomerados, conformado el primero por 17 jugadores y el segundo por 12 jugadores.

Figura 15

Aplicación del Método no jerárquico K-medias.

```
> agrup_kmeans <- kmeans(datos1,
+                          centers=2,      # Número de cluster
+                          nstart = 100,   # Número de puntos iniciales
+                          iter.max = 1000)
> agrup_kmeans
K-means clustering with 2 clusters of sizes 17, 12
```

Nota. Dentro del método K-medias, nos arroja al igual que el método anterior, que los 2 grupos tiene distribuidos 17 y 12 jugadores respectivamente.

El siguiente paso en este análisis corresponde a formar los 2 grupos con los jugadores con dichos códigos. Bajo el método K-medias obtenemos lo siguiente.

Figura 16

Código para generar el gráfico del método K-medias.

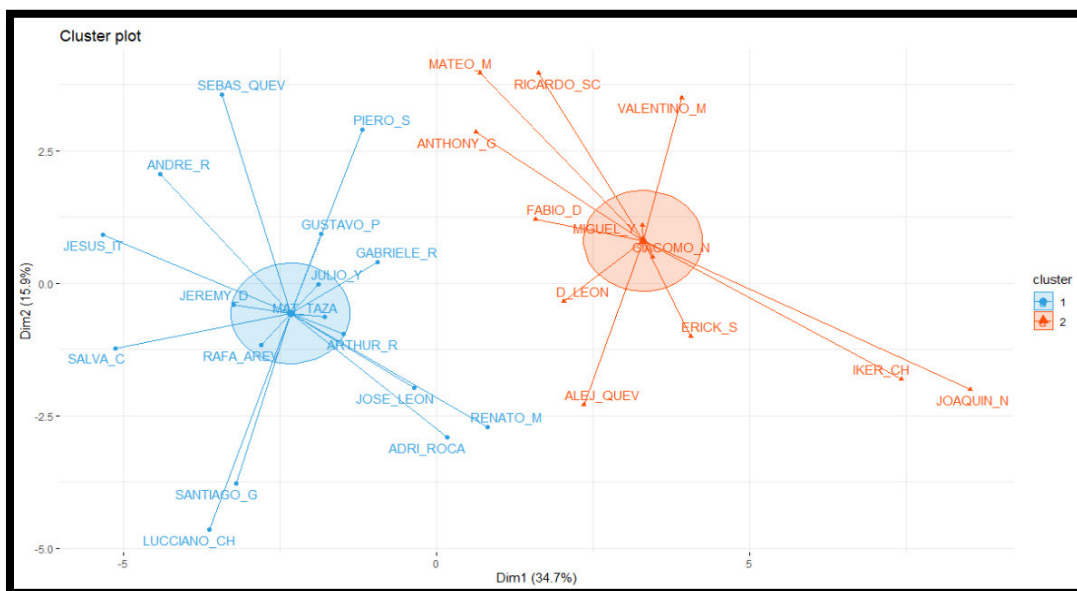
```
km.res<-kmeans(scale(datos1), 2, nstart = 25)

fviz_cluster(km.res,data = datos1,stand = TRUE,
              palette = c("#2E9FDF", "#FC4E07"),
              ellipse.type = "euclid", # Elipse de concentración
              star.plot = TRUE, # agrega segmentos a los centroides
              repel = TRUE, # impresión de etiquetas
              ggtheme = theme_minimal()
)
```

Nota. Formaremos los 2 conglomerados, esta vez con los jugadores distribuidos.

Figura 17

Gráfico de los 2 conglomerados.



Nota. Jugadores clasificados por características similares de juego.

4.1 RESULTADOS

4.1.1 INTERPRETACIÓN:

En la Figura 14, en primer lugar se forman los 2 grupos los cuales fueron corroborados por 3 métodos de análisis exploratorio de datos (Codo, Silueta, Criterio C-H), donde el primero conglomerado de color azul es quien tiene a la mayoría de jugadores y a raíz de la similitud que presentan entre ellos, por ejemplo a los jugadores que se encuentran más cerca al centroide de su conglomerado donde observamos a RAFA_AREV, MAT_TAZA, JEREMY_D, JULIO_Y, ARTHUR_R; ellos tienen un papel de creadores de juego, todos ellos juegan como volantes del

equipo y la gran parte del partido son ellos los que presentan mayor toque de balón, en comparación a sus compañeros. Un poco más lejos y asignados hacia la parte superior podemos observar a GABRIELE_R, JESUS_IT, ANDRE_R, GUSTAVO_P, SEBAS_QUEV; jugadores encargados del mediocampo, cuya función principal es defender, por ellos también circula el balón con frecuencia y tienen como característica similar la recuperación del balón. Por otro lado, vemos a SANTIAGO_G, LUCCIANO_CH, JOSE_LEON, ADRI_ROCA, RENATO_M; ellos son los delanteros en sus categorías, tienen características similares con respecto a jugadas de ataque, tienen como características similares la cantidad de goles, entre otras.

En el otro conglomerado, observamos que presenta la menor parte de jugadores y con características similares, dentro de este conglomerado observamos a los porteros y a los defensas y solo un delantero que es ALEJ_QUEV, es uno de los jugadores con menos participación en los partidos.

4.2 Método Jerárquico De Ward

Ahora veremos el análisis conglomerado jerárquico, aquí de igual manera se formaran los 2 grupos bajo el método de Ward.

Para esto tenemos un código en R Studio, el cual se muestra en la siguiente figura

Figura 18

Cluster Jerárquico: Método Jerárquico de Ward.


```
# CLUSTER JERARQUICO
#-----
a=hclust(dist(scale(datos1)),method="ward.D")
plot(a)

(b=cutree(a, h=20))
plot(a)
```

Nota. Código para el método de Ward en el R studio.

Tenemos como resultados a que conglomerados son distribuidos cada jugador, ellos son clasificados por números que indican a que grupo van, dadas sus características de juego.

Figura 19

Distribución de cada jugador por conglomerado

MIGUEL_Y	VALENTINO_M	RICARDO_SC	PIERO_S	ANDRE_R	SEBAS_QUEV	GUSTAVO_P
1	1	1	2	2	2	2
JOSE_LEON	SANTIAGO_G	LUCCIANO_CH	RENATO_M	GABRIELE_R	JULIO_Y	JEREMY_D
1	2	2	1	1	1	2
IKER_CH	FABIO_D	MATEO_M	ANTHONY_G	GIACOMO_N	MAT_TAZA	RAFA_AREV
1	1	1	1	1	2	2
ALEJ_QUEV	ADRI_ROCA	JESUS_IT	ARTHUR_R	JOAQUIN_N	ERICK_S	D_LEON
1	1	2	1	1	1	1
SALVA_C						
2						

Nota. En el 1 están los jugadores con mejores indicadores a nivel defensivo y en el 2, los que tienen mejores indicadores de creación de juego y ofensivo.

Lo siguiente es observar los resultados en el grafico representativo del método jerárquico, el dendograma.

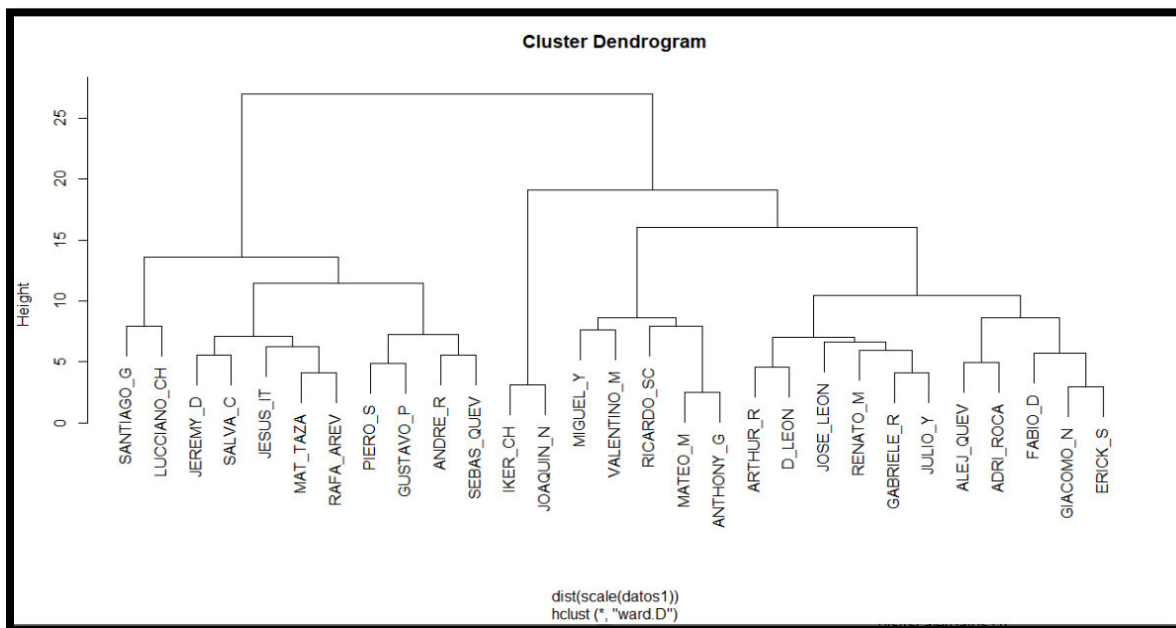
Figura 20*Dendograma*

Figura 19: Dendograma del método jerárquico.

El método jerárquico de Ward nos muestra la distribución en 2 conglomerados, en el conglomerado de la izquierda se observa a volantes de estilo defensivo como SEBAS_QUEV, ANDRE_R, GUSTAVO_P, PIERO_S, cuya característica similar es la recuperación de balón, también a los volantes de creación de juego como RAFA_AREV, MAT_TAZA, y a los atacantes SANTIAGO_G, LUCCIANO_CH.

En el conglomerado de la derecha tenemos a la mayoría de jugadores, entre los cuales se encuentran los porteros IKER_CH, JOAQUIN_N, MIGUEL_Y, también se encuentran todos los defensas por las bandas como VALENTINO_M, FABIO_D, ERICK_S y los defensas centrales MATEO_M, ANTHONY_G; a su vez también están volantes de creación y de marca como

ARTHUR_R, GABRIELE_R, JULIO_Y, los cuales también tienen características similares como la cantidad de toques de balón, pases efectivos, etc.

Por otra parte, dentro de este conglomerado se encuentran delanteros como ADRI_ROCA, ALEJ_QUEV.

V. CONCLUSIONES

Los análisis de conglomerados no jerárquico K-medias y jerárquico Ward son útiles para la aplicación en el fútbol, de la misma manera para cualquier otro deporte.

Es necesario eliminar variables que son repetitivas para que de esa manera los conglomerados formados sean homogéneos.

Al observar los resultados de ambos métodos, resulta que el método no jerárquico de k-medias es más preciso que el método jerárquico de Ward.

La etapa formativa (división de menores o canteras) del fútbol base es la más importante para el desarrollo de los que dedican tiempo de su vida a practicarlo, el uso de la estadística y la tecnología vienen siendo utilizadas profundamente en este deporte.

Las variables elegidas nos permiten observar las características similares en defensiva, ofensiva y creación de juego de cada niño y joven que pertenece al equipo competitivo; con esto decimos que la distribución de los jugadores en 2 conglomerados nos permite interpretar las características similares de juego que presentan los jóvenes por naturaleza para ser utilizado en formaciones, entrenamientos, didácticas y dinámicas que se realizan dentro de la Academia BFC.

A su vez, aunque sabemos que ambos métodos son útiles para este tipo de estudios, podemos decir que el método de K-medias es más preciso y óptimo para este tipo de análisis, ya que muestra un resultado cuantitativo más próximo a los objetivos de la Academia BFC, y es corroborado en el análisis de los partidos amistosos y de competencia.

VI. RECOMENDACIONES

Existen maneras de obtener datos para un análisis cuantitativo en el fútbol, las variables marcan parametros tanto defensivos, ofensivos y creacion de juego que son, los eventos que ocurren en los partidos de futbol. Gracias a los analisis estadisticos previos o durante el partido que se puede realizar, un equipo de futbol puede tomar mejores decisiones al momento de competir. Entonces se recomienda lo siguiente:

A la Academia BFC, adquirir mejores herramientas para futuros analisis respectivos, estas serian una mejor cámara filmadora, unos chalecos GPS que sirven para medir el rendimiento fisico del jugador y a traves de esto ver su evolucion constantemente como tambien poder evitar lesiones, tambien una tablet o un dispositivo electrónico donde almacenar estos datos recogidos a traves de las filmaciones y de los GPS.

A los entrenadores de futbol formativo, que no sean esquivos a los resultados estadisticos, que vean estos resultados como apoyo para resultados positivos futuros.

A la escuela de Estadística de la FCM-UNMSM, implementar temas estadísticos relacionados al deporte, ahora con la tecnología que tenemos al alcance, deberiamos capacitar a los estudiantes de pre-grado en dicha área.

VII. BIBLIOGRAFÍA

Garrido, T. M. (2021-22). *Uso de técnicas de clustering para encontrar perfiles de jugadores en una competición de fútbol profesional*. Valladolid.

Singh, K. R. (2017). *Machine Learning Using R*. New Delhi: Apress.

Malagón Selma, MDP (2019). *Machine Learning en el mundo del fútbol*.

Federación Peruana de Fútbol - FPF (2022). *Reglamento Campeonato Federación Regional, Copa Centenario FPF 2022*.

Marqués, M. P. (2015). *Minería de datos: a través de ejemplos*. Colombia: España.

Sistema de Análisis Estadístico con SPSS (2006). (n.d.). (n.p.): IICA Biblioteca Venezuela.

Wanceulen Ferrer, A. (2017). *Las Escuelas de Fútbol: Objetivos, Contenidos, Metodología y Evaluación*. España: Wanceulen Editorial S.L.

Arroyo López, P. E., Borja Medina, J. C. (2018). *Análisis multivariante para la inteligencia de mercados*. México: Editorial Digital del Tecnológico de Monterrey.

Pineda Pertuz, C. (2022). *Aprendizaje automático y profundo en Python: Una mirada hacia la inteligencia artificial*. Colombia: Ediciones de la U.

Raschka, S. (2023). *Machine Learning con PyTorch y Scikit-Learn*. España: Marcombo.

https://www.campusbigdata.com/difusion/Informe_Conglomerados_Mediocentros.pdf

VIII. ANEXOS

ANEXO A: Listado de jóvenes participantes del torneo federación regional 2022 de la Academia BFC.

Lista de datos etiquetados:

Sub 15:

MIGUEL_Y: Miguel Yllaconza

VALENTINO_M: Valentino Molina

RICARDO_SC: Ricardo Scaglioni

PIERO_S: Piero Sanchez

ANDRE_R: Andre Ramos

SEBAS_QUEV: Sebastian Quevedo

GUSTAVO_P: Gustavo Perales

JOSE_LEON: Jose Leon

SANTIAGO_G: Santiago Galvez

LUCCIANO_CH: Lucciano Chale

RENATO_M: Renato Morales

GABRIELE_R: Gabriele Romani

JULIO_Y: Julio Yataco

JEREMY_D: Jeremy Durant

Sub 13:

IKER_CH: Iker Chileno

FABIO_D: Fabio Davila

MATEO_M: Mateo Mallma

ANTHONY_G: Anthony Gil

GIACOMO_N: Giacomo Novoa

MAT_TAZA: Mathias Taza

RAFA_AREV: Rafael Arevalo

ALEJ_QUEV: Alejandro Quevedo

ADRI_ROCA: Adriano Roca

JESUS_IT: Jesus Itache

ARTHUR_R: Arthur Romero

JOAQUIN_N: Joaquin Noriega

ERICK_S: Erick Santamaria

D_LEON: Diego Leon

SALVA_C: Salvador Carreño

ANEXO B: DISPOSICIONES GENERALES DE LA COMISION DE MENORES DE LA LIGA DEPARTAMENTAL DE FÚTBOL

Artículo 1.- Los Campeonatos de Menores Masculino son organizados, coordinados, supervisados, deportiva y administrativamente, por la Federación Peruana de Fútbol, a través de las Ligas Departamentales de Fútbol.

Artículo 2.- Los Campeonatos de Fútbol de Menores Masculino se desarrollarán en todas las Regiones del país, bajo la normativa vigente estipulada por las autoridades estatales competentes en cada región, respecto a salud pública y bioseguridad.

Artículo 3.- El presente campeonato se regirá de acuerdo con el Reglamento Único de Menores, el Texto Único de Bases y Reglamentos del Sistema Nacional de Campeonatos de la FPF.

Artículo 4.- En el Campeonato Federación, será de aplicación el Reglamento Único de Justicia de la FPF.

ANEXO C: DISPOSICIONES PARA EL JUGADOR

Artículo 14.- Para participar en el campeonato Federación Regional, los jugadores deberán contar con los siguientes requisitos:

Documento de Identidad y Ficha de Afiliación:

- DNI.

- Ficha de afiliación otorgada por las Ligas Departamentales y/o Federación Peruana de Fútbol.

b) Fotografía:

- Dos (02) fotografías tamaño carné formato JPG, en posición de frente, con fondo blanco, cabeza descubierta y obligatoriamente a colores. Las fotografías deben ser actuales.

c) Exámenes Médicos:

- Ficha de Afiliación de Asegurado (Sis, Essalud, Privado)

- Certificado médico individual.

- Carnet de vacunación (dosis completa según disposición del gobierno)

d) Autorización del Padre de Familia y/o Apoderado:

- Formulario de participación.

- Copia del DNI del padre y/o apoderado.

Artículo 15.- Un jugador no podrá inscribirse y/o actuar por Clubes distintos en un mismo Campeonato Copa Federación y que correspondan al mismo año; salvo los Jugadores con transferencia o Cesión Temporal en la 1era. Etapa (apertura) o 2da. Etapa (clausura).

Artículo 16.- Es requisito indispensable para participar la presentación el Carnet de Jugador y su DNI original, si por alguna circunstancia no posee el DNI original, podrá hacerlo con el Certificado Original de identidad expedido por la RENIEC con la debida imagen del deportista.

ANEXO D:**Variables nivel defensivo**

Despejes: Número de toques de balón que son lanzados a los laterales o tiros de esquina.

Quites_pelota: Número de veces en las que el jugador recupera la posesión del balón del equipo contrario.

Veces_lo_regatean: Número de veces en las que logran pasar con el balón a un jugador del equipo.

Errores_q_ocasionan_gol: Número de veces en las que el error del jugador ocasiona que al equipo les anoten gol.

Penales_cometidos: Número de faltas dentro del área ocasionada por el jugador durante el torneo.

Duelos_terrestres_ganados: Número de veces en las que el jugador gana la posesión del balón en una jugada dividida con los pies o partes de la pierna.

Duelos_aereos_ganados: Número de veces en las que el jugador gana la posesión del balón en una jugada dividida con la cabeza.

Total_duelos: Total de duelos terrestres y duelos aéreos, ganados y perdidos.

Prom_duelo_ganado: Promedio de duelos terrestres y aéreos ganados durante el torneo.

Tackles_partido: Número de contacto cuerpo-cuerpo de jugadores de dos equipos.

Variables creación de juego

Toque_balon. Número total de veces en las que el jugador tuvo contacto con el balón durante el torneo.

Pases. Número de pases de balón o acción colectiva (que permite relacionar al equipo) de cada jugador durante el torneo.

Pases_claves. Número de pases de cada jugador durante el torneo, que pueden ocasionar una jugada de peligro a favor del equipo.

Pases_precisos. Número de pases de cada jugador durante el torneo, que fueron dirigidos correctamente al compañero de equipo.

Pases_propia_cancha. Número de pases de cada jugador durante el torneo, en cancha propia.

Pases_largos_precisos. Número de pases de cada jugador durante el torneo, lanzados a más de 30 metros de distancia.

Pases_medDist_precisos. Numero de pases de cada jugador durante el torneo, a una distancia no menor a 15 ni mayor a 30 metros.

Pases_errados. Numero de pases de cada jugador durante el torneo, que no fueron dirigidos correctamente al compañero de equipo.

Veces_regatea. Número de veces en las que el jugador logra pasar con el balón a uno o más jugadores del equipo contrario.

Faltas_recibidas. Número de fouls recibidos.

Variables nivel ofensivo

Goles. Número de anotaciones de cada jugador

PROM_GOLXPART. Promedio de goles por partido durante el torneo.

Gol_Penal. Número de anotaciones a través de un tiro penal de cada jugador.

Gol_pieDer. Número de anotaciones con el pie derecho de cada jugador.

Gol_pieIz. Número de anotaciones con el pie izquierdo de cada jugador.

Asistencias. También llamado pase gol, es el pase que le permite anotar el gol al equipo.

Tiros_a_puerta. Numero de tiros al arco contrario efectuado por cada jugador durante el torneo.

A parte de las variables que definen las características similares de cada jugador, tenemos algunas que son importantes para medir el rendimiento físico, tales como

ID. Nombre abreviado del jugador

Min_jugados. Número total de minutos jugados durante el torneo.

Min_posibles. Número total de minutos posibles que se puede jugar en el torneo.

Part_jugados. Número de partidos de futbol en los que participo el jugador

Part_titular. Número de partidos de futbol en los que el jugador figuró como titular.

Perdida_pelota. Número de veces en las que el jugador perdió la posesión del balón con el equipo contrario.

Faltas_realizadas. también llamados foul, es la cantidad de fouls ocasionados por cada jugador.

Falta_penal_ganadas. Número de fouls en las que el jugador ocasiona un penal a favor del equipo.

Offside. Número de fuera de lugar.

Tarj_amarillas. Número total de tarjetas amarillas en el torneo.

Tarj_rojas. Número total de tarjetas rojas en el torneo.

Tabla de figuras

Figura 1 Organigrama de la Academia BFC	11
Figura 2 Lista de los jóvenes que pertenecen al equipo competitivo de Academia BFC. ¡Error! Marcador no definido.	
Figura 3 Imagen referencial de la base de datos.....	19
Figura 4 Inicio del análisis en R, cargamos la base de datos.....	25
Figura 5 Código para obtener la estructura de la base de datos.....	25
Figura 6 Código para verificar si hay datos perdidos	25
Figura 7 Gráfico de datos perdidos o nulos	26
Figura 8 Histogramas de variables numéricas	¡Error! Marcador no definido.
Figura 9 Análisis Exploratorio de Datos, sin las variables con datos repetitivos	¡Error! Marcador no definido.
Figura 10 Código para método del codo.....	21
Figura 11 Gráfico representativo del método del codo.....	29
Figura 12 Código para coeficiente de Silueta	29
Figura 13 Gráfico representativo del coeficiente de Silueta ¡Error! Marcador no definido.	
Figura 14 Criterio C-H.....	¡Error! Marcador no definido.
Figura 15 Aplicación del Método no jerárquico K-medias	31
Figura 16 Código para el método de K-medias	31
Figura 17 Gráfico de conglomerados no jerárquico K-medias.....	32
Figura 18 Método de Ward.....	34

Figura 19 Distribución de jugadores por conglomerado 34

Figura 20 Dendograma del método jerárquico 35