

# **ESCUELA POLITÉCNICA NACIONAL**

**FACULTAD DE INGENIERIA DE SISTEMAS**

**DESARROLLO DE MODELOS DE APRENDIZAJE AUTOMÁTICO  
APLICADOS A LA PATOLOGÍA DE ALZHEIMER**

**MODELO DE APRENDIZAJE AUTOMÁTICO NO-SUPERVISADO  
PARA EL ANÁLISIS DE PACIENTES CON DETERIORO  
COGNITIVO RELACIONADO CON LA PATOLOGÍA DE  
ALZHEIMER**

**TRABAJO DE INTEGRACIÓN CURRICULAR PRESENTADO COMO  
REQUISITO PARA LA OBTENCIÓN DEL TÍTULO DE INGENIERO EN  
CIENCIAS DE LA COMPUTACIÓN**

**ALBARRACÍN VINUEZA KLEBER ANDRÉS**

**[kleber.albarracin@epn.edu.ec](mailto:kleber.albarracin@epn.edu.ec)**

**DIRECTOR: Dr. AGUIAR PONTES JOSAFÁ DE JESÚS**

**[josafa.aguiar@epn.edu.ec](mailto:josafa.aguiar@epn.edu.ec)**

**DMQ, Septiembre 2024**

## **CERTIFICACIONES**

Yo, Kleber Andrés Albarracín Vinueza declaro que el trabajo de integración curricular aquí descrito es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

---

**Kleber Andres Albarracín Vinueza**

Certifico que el presente trabajo de integración curricular fue desarrollado por Kleber Andres Albarracín Vinueza, bajo mi supervisión.

**Dr. Josafá de Jesús Aguiar Pontes**  
**DIRECTOR**

## **DECLARACIÓN DE AUTORÍA**

A través de la presente declaración, afirmamos que el trabajo de integración curricular aquí descrito, así como el (los) producto(s) resultante(s) del mismo, son públicos y estarán a disposición de la comunidad a través del repositorio institucional de la Escuela Politécnica Nacional; sin embargo, la titularidad de los derechos patrimoniales nos corresponde a los autores que hemos contribuido en el desarrollo del presente trabajo; observando para el efecto las disposiciones establecidas por el órgano competente en propiedad intelectual, la normativa interna y demás normas.

Kleber Andrés Albarracín Vinueza

Dr. Josafá de Jesús Aguiar Pontes

## DEDICATORIA

Con profundo agradecimiento, quiero dedicar este trabajo a quienes han sido un pilar fundamental en mi vida durante toda mi carrera universitaria. En primer lugar, a la Madre Dolorosa, cuya imagen me ha acompañado en los momentos más difíciles. Su presencia ha sido una fuente de fortaleza espiritual, dándome el coraje necesario para superar los desafíos académicos y personales. En los momentos de duda y dificultad, su amor incondicional me ha recordado la importancia de la fe, la paciencia y la perseverancia. A ella, a quien dirijo mis oraciones, dedico cada paso que he dado hacia la culminación de este logro, que no habría sido posible sin su intercesión y protección.

A mis padres, quienes me han brindado su apoyo incondicional desde el primer día. Con amor y sacrificio, han estado a mi lado en cada etapa de este largo camino, dándome las herramientas necesarias no solo para alcanzar mis metas académicas, sino también para crecer como persona. Su fe en mí, su paciencia y sus palabras de aliento en los momentos más difíciles me impulsaron a no rendirme. Agradezco cada consejo, cada conversación y cada sacrificio que hicieron para que pudiera llegar a este momento. Esta tesis es tanto suya como mía, pues sin ellos, este sueño no habría sido posible.

A mi tía, cuya presencia y apoyo constante han sido fundamentales para mí. Ella ha sido una segunda madre, ofreciéndome su sabiduría y amor en los momentos en que más lo necesitaba. Su ejemplo de fortaleza y dedicación me ha inspirado a seguir adelante, incluso cuando el camino parecía imposible de recorrer. Sus palabras siempre llenas de cariño me han dado consuelo en los momentos de mayor incertidumbre, y por ello, le dedico también este logro.

Finalmente, a mi pareja, mi compañera en este viaje, quien ha sido una fuente inagotable de amor y comprensión. Gracias por estar a mi lado en cada triunfo y cada tropiezo, por tu paciencia en los días más agotadores y por tu constante fe en mí. Tu apoyo ha sido un faro de luz en los momentos más oscuros, recordándome siempre que el esfuerzo y la dedicación valen la pena. Esta tesis es también un reflejo de lo que hemos construido juntos, de la fortaleza que hemos desarrollado como equipo y del amor que hemos cultivado a lo largo de este camino.

A todos ustedes, les dedico este trabajo con la más profunda gratitud, pues han sido mi inspiración y mi fortaleza. Esta tesis no solo marca el fin de una etapa académica, sino también el inicio de una nueva, que emprendo con la certeza de que seguiré contando con su amor y apoyo incondicional.

## AGRADECIMIENTO

Quiero expresar mi más sincero agradecimiento a todas las personas que, de una forma u otra, han sido parte fundamental de este camino académico. En primer lugar, mi más profunda gratitud a la Madre Dolorosa, cuya devoción y consuelo me han acompañado a lo largo de estos años. En los momentos de mayor dificultad, su presencia me ha dado fuerzas para seguir adelante, recordándome siempre la importancia de la fe y la esperanza. A ella le debo no solo la fortaleza espiritual para enfrentar los desafíos, sino también la serenidad en los momentos de incertidumbre. Su guía ha sido una luz constante que me ha inspirado a superar cada obstáculo.

A mis padres, quienes, con su amor, sacrificio y dedicación, han sido el pilar sobre el cual he construido este sueño. Ellos han estado siempre a mi lado, brindándome su apoyo incondicional y alentándome a dar lo mejor de mí en cada paso de este recorrido. Sus consejos, paciencia y ejemplo me han impulsado a perseverar incluso cuando parecía que las dificultades eran insuperables. Su fe en mí me ha motivado a no rendirme y a seguir adelante con determinación. Todo lo que he logrado hasta ahora es también reflejo de su esfuerzo y confianza en mí, por lo que les estaré eternamente agradecido.

A mi tía, quien ha sido una presencia invaluable en mi vida, siempre brindándome su apoyo y cariño. Su amor constante y su preocupación genuina han sido un consuelo en los momentos más difíciles. A lo largo de este camino, ella ha sido una fuente de sabiduría y aliento, ofreciéndome palabras de ánimo cuando más las necesitaba. Su ejemplo de fortaleza y su compromiso con mi bienestar me han inspirado profundamente, y este logro es también resultado de todo lo que me ha enseñado.

A mi pareja, mi compañera y apoyo incondicional, gracias por estar a mi lado en cada paso de este proceso. Tu amor, paciencia y comprensión han sido fundamentales para que yo pudiera mantenerme enfocado y motivado, incluso en los momentos más agotadores. Has sido mi refugio en los días difíciles y mi motivación constante para seguir adelante. Sin tu apoyo inquebrantable y tu confianza en mí, este logro habría sido mucho más difícil de alcanzar. Este trabajo es también tuyo, y no puedo agradecerte lo suficiente por todo lo que has hecho por mí a lo largo de este camino.

A todos aquellos que de alguna manera han sido parte de esta etapa de mi vida, ya sea con su apoyo, consejos o simplemente estando presentes, les agradezco de corazón.

# ÍNDICE DE CONTENIDO

CERTIFICACIONES.....	I
DECLARACIÓN DE AUTORÍA.....	II
DEDICATORIA.....	III
AGRADECIMIENTO.....	IV
ÍNDICE DE CONTENIDO.....	V
RESUMEN .....	VII
ABSTRACT .....	VIII
1 DESCRIPCIÓN DEL COMPONENTE DESARROLLADO.....	1
1.1 Objetivo general.....	1
1.2 Objetivos específicos .....	2
1.3 Alcance .....	2
1.4 Descripción del problema.....	2
2 MARCO TEÓRICO .....	8
2.1 Análisis de Componentes Principales (PCA).....	8
2.2 K-Means clustering.....	11
2.3 Comparación de PCA y K-Means.....	14
2.4 Ambiente de trabajo y lenguaje de programación .....	14
3 METODOLOGÍA.....	16
3.1 Importación de librerías.....	16
3.2 Descripción del dataset:.....	19
3.3 Preprocesamiento de datos:.....	20
3.4 Análisis de correlaciones, selección y reducción de variables. ....	21
3.5 Escalado de variables .....	26
3.6 Desarrollo del modelo de aprendizaje no supervisado.....	28
3.7 Evaluación del modelo .....	33
4 RESULTADOS, CONCLUSIONES Y RECOMENDACIONES.....	37
4.1 Resultados .....	37
4.2 Conclusiones.....	51
4.3 Recomendaciones.....	52
5 REFERENCIAS BIBLIOGRÁFICAS.....	54
6 ANEXOS.....	58

ANEXO I.....	59
ANEXO II.....	62
ANEXO III.....	63
ANEXO IV .....	63

## RESUMEN

Para este trabajo de integración curricular se ha escogido como tema el progreso de los pacientes de Alzheimer, para el desarrollo de un modelo de aprendizaje no supervisado, mismo que se realizó a partir de un conjunto de datos.

El desarrollo del modelo, se trabajó en el ambiente de trabajo de Google Colab, mismo que evidenció la necesidad de utilizar un tratamiento de valores faltantes sobre los datos, para poder realizar la posterior ejecución de los algoritmos correspondientes.

Luego se realizó la correlación entre las variables, para poder eliminar las que causan ruido dentro del análisis. A partir de esto se escalaron los datos, para realizar el análisis de componentes principales (PCA) y reducir la dimensionalidad del conjunto de datos y así poder observar la separabilidad de los datos.

La aplicación de la clusterización permitió asignar a los pacientes de Alzheimer al grupo de estado al que pertenecen, esto es al de demencia, no demencia y en transición o convertida.

Los algoritmos de aprendizaje no supervisado no son susceptibles de medición en porcentaje de certeza, los resultados obtenidos a partir de este trabajo, pueden tomarse como base para algoritmos de aprendizaje supervisado para permitir la predicción del avance o no de la enfermedad hacia una demencia diagnosticada.

**PALABRAS CLAVE:** Alzheimer, conjunto de datos, análisis de componentes principales, dimensionalidad, separabilidad, grupos.

## **ABSTRACT**

The progress of Alzheimer's patients has been chosen as the topic for this paper to curricular integration, in order to define the development of an unsupervised learning model, which was carried out from a set of data.

The development of the model was worked in the Google Colab work environment, which showed the need to use a treatment of missing values on the data, in order to carry out the subsequent execution of the corresponding algorithms.

Then the optimization was carried out between the variables, in order to eliminate those that cause noise within the analysis. From this, the data were scaled to perform principal component analysis (PCA) and reduce the dimensionality of the data set and thus be able to observe the separability of the data.

The application of clustering allowed Alzheimer's patients to be assigned to the status group to which they belong, that is, dementia, non-dementia, and transition or conversion.

Unsupervised learning algorithms are not susceptible to measurement in percentage of certainty, the results obtained from this work can be taken as a basis for supervised learning algorithms to allow prediction of the progression or not of the disease towards a diagnosed dementia.

**KEYWORDS:** Alzheimer, set of data, principal component analysis, dimensionality, separability, cluster.

# 1 DESCRIPCIÓN DEL COMPONENTE DESARROLLADO

Al ser el Alzheimer una de las dolencias más comunes en la actualidad entre las personas de la tercera edad, se ha escogido esta como tema para el desarrollo de un modelo de aprendizaje no supervisado, mismo que se realizó a partir de un conjunto de datos referenciales obtenidos de Kaggle, en donde se encontraban almacenados, con el fin de mostrar estadísticas al respecto.

Para el desarrollo del modelo, se ingresó en el ambiente de trabajo de Google Colab, el que se trabaja en la nube, en el navegador web, dentro del cual se dio un primer análisis de los datos, mismo que evidenció la necesidad de utilizar un tratamiento de valores faltantes sobre los datos, para poder realizar la posterior ejecución de los algoritmos correspondientes.

A partir de ello se realizó la correlación entre las variables, para determinar las que tienen una baja correlación con la variable de trabajo y que, por lo tanto, pueden generar ruido dentro del análisis y que por ello debían ser eliminadas.

Posterior a esto se escalaron o estandarizaron los datos, esto para poder realizar el Análisis de Componentes Principales (PCA), lo cual permitió reducir la dimensionalidad del conjunto de datos con el fin de poder observar la separabilidad de los datos.

Ya con la dimensionalidad reducida, se aplicó un método de clusterización (K-Means), con el que se busca asignar a los pacientes de Alzheimer al grupo de estado al que pertenecen, esto es al de demencia (confirmada y diagnosticada), no demencia (que no han mostrado síntomas o signos de la enfermedad) y en transición, es decir, los pacientes que al inicio del estudio no mostraron síntomas sensibles de la enfermedad, pero que, conforme fue avanzando el estudio, fueron deteriorando su condición y por lo tanto, mostrando los primeros signos de demencia.

A partir del PCA se pudo ver una clara separabilidad de los pacientes en cada una de las fases, solo con un pequeño solapamiento de la etapa de transición temprana y tardía, esto por la misma naturaleza de la enfermedad.

A pesar de que los algoritmos de aprendizaje no supervisado no son susceptibles de medición en porcentaje de certeza, los resultados obtenidos a partir de este trabajo, pueden tomarse como base para algoritmos de aprendizaje supervisado para permitir la predicción del avance o no de la enfermedad hacia una demencia diagnosticada.

## 1.1 Objetivo general

Desarrollar un modelo de aprendizaje automático no supervisado que permita analizar las características de los pacientes que presentan Alzheimer e identificar aquellas que se considera relevantes.

## **1.2 Objetivos específicos**

- Aplicar técnicas de preprocesamiento de datos, como normalización y manejo de valores faltantes, para asegurar la coherencia y la calidad de los datos utilizados en el modelo.
- Desarrollar un modelo de aprendizaje no supervisado que pueda identificar patrones correlacionados con personas sanas y patrones correlacionados con personas enfermas.
- Seleccionar y probar algoritmos de aprendizaje no supervisado, ajustando sus parámetros hasta lograr resultados convenientes.

## **1.3 Alcance**

Identificación de variables relevantes, En esta actividad, se llevará a cabo una revisión de la literatura para identificar las variables clínicas más relevantes para el Alzheimer y sus etapas de deterioro cognitivo.

Preparación de los datos: En esta actividad, se trabajará con el conjunto de datos Alzheimer.csv, para llevar a cabo este proyecto de titulación. Así también, se realizarán las tareas de limpieza y preprocesamiento de datos que se estimen necesarios.

Utilización de herramientas de visualización: Se integrarán herramientas de visualización que permitan la interpretación de los resultados, facilitando la comprensión de los patrones identificados.

Codificación: Se implementará un modelo de aprendizaje no supervisado para la identificación de características relevantes, y un modelo de aprendizaje no supervisado para la clusterización, utilizando herramientas y bibliotecas de programación de Python.

Pruebas, Evaluación, Conducción de pruebas: Se realizarán pruebas de los modelos.

Análisis de Resultados y Conclusiones: Se llevará a cabo un análisis detallado de los resultados obtenidos, identificando patrones relevantes y conclusiones que puedan ser de utilidad para profesionales de la salud en el manejo del Alzheimer.

## **1.4 Descripción del problema.**

Desde siempre, los seres humanos, se han visto colocados en situaciones de difícil resolución, situaciones que, en siglos pasados y de manera muy frecuente los llevaban a una muerte temprana y súbita, enfermedades, guerras, la misma vida cotidiana era dura y difícil de afrontar.

Con el paso del tiempo, gracias al avance de la tecnología los problemas humanos han cambiado, las guerras se resuelven en su mayoría a nivel diplomático, muchas enfermedades antes mortales hoy son fácilmente tratables y curables, la vida diaria se ha facilitado a través de modernos medios de transporte, máquinas e incluso recursos para pasar los momentos de ocio. Todo esto ha contribuido a alargar el tiempo de vida humana, lo cual ha traído nuevos problemas, esto es que la población envejecida presenta diferentes tipos de enfermedades y síndromes relacionados con la pérdida de las capacidades físicas y mentales propias de la juventud.

Durante la última etapa de la vida humana, pueden presentarse varios de estos síndromes relacionados con la disminución de la capacidad mental, una de ellas es el Alzheimer.

Si bien este no se desarrolla exclusivamente en pacientes mayores de 80 años, si se presenta en un mayor porcentaje dentro de ese rango de edad y se presenta desde aproximadamente los 60 años.

A continuación, se expondrán algunos hechos sobre este síndrome y sus características relevantes.

#### 1.4.1 Del Alzheimer

El Alzheimer es una patología neurodegenerativa progresiva, con una variabilidad relativa, es decir, los síntomas y signos son diferentes para cada paciente, así como el daño neuropatológico dentro del cerebro (ver Anexo I, Tabla 1).

El Alzheimer no tiene cura hasta el presente, por lo que es de gran importancia su diagnóstico temprano, lo que, si bien no permite revertir sus efectos, si ayuda sobre todo a los familiares de los pacientes a prepararse de mejor manera para afrontar la situación y a organizarse para el cuidado del paciente.

##### 1.4.1.1 Demencia

“La demencia es una disminución de la capacidad intelectual del paciente al compararla con el nivel previo de esta función. Esta disminución también se encuentra acompañada de cambios psicológicos y del comportamiento, lo que conlleva la alteración de las capacidades sociales y laborales del paciente” (Casanova, 1999).

La demencia no se refiere de una manera precisa a una enfermedad, sino, a un síndrome, ya que es un conjunto de síntomas y signos (ver Anexo I, Tabla 2). Este síndrome se evidencia por la pérdida progresiva de la capacidad del individuo de poder cuidar de sí mismo, de interactuar con el medio y con otros individuos volviéndolo finalmente dependiente de sus cuidadores, mismos que pueden ser sus familiares o si su nivel socioeconómico lo permite, personal de cuidado especializado.

La demencia puede tener diferentes causas, cada una de ellas recibe un nombre en particular:

- “Alzheimer”.
- “Demencia vascular”.
- “Demencia con cuerpos de Lewy”.
- “Enfermedad de Parkinson con demencia”.
- “Demencia del lóbulo frontal (enfermedad de Pick)”.
- “Hidrocefalia normotensiva”.
- “Demencias degenerativas”.
- “Combinaciones de síndromes y síndromes de solapamiento”.
- “Enfermedades por priones”.
- “Atrofias focales progresivas”.
- “Demencias metabólicas y tóxicas”.
- “Infecciones”.
- “Enfermedades neurológicas que se pueden presentar como una demencia”.
- “Demencias misceláneas infrecuentes”.
- “Demencias pediátricas que se pueden presentar en la edad adulta”.
- “Demencias reversibles” (Casanova, 1999).

Dentro de estos síndromes, el Alzheimer es el que cuenta con un mayor porcentaje de incidencia dentro de la población de la tercera edad, situándose entre el 50 % al 60 % de las demencias que se presentan (Casanova, 1999).

#### *1.4.1.2 Alzheimer*

El Alzheimer es un tipo de demencia que pueden afectar a las personas mayores de 65 años. De los estudios clínicos realizados, existen los siguientes factores de riesgo para la aparición del Alzheimer:

- Edad

- Apo E4 (presencia de por lo menos una copia del alelo E4 del gen de la Apo-E (alopipoproteína E) en el cromosoma 19).
- Historia familiar de demencia.
- Trauma craneal.
- Nivel educativo bajo.
- Síndrome de Down (pueden desarrollar Alzheimer a edades tempranas, 50 años o antes).
- Hipertensión arterial sistólica.

Si bien estos se consideran factores de riesgo, estos pueden o no presentarse en un paciente con Alzheimer, es decir, no son decisivos. La edad no es un factor de peso para la aparición del Alzheimer, no todos los pacientes de la tercera edad lo padecen, sin embargo, si la longevidad del paciente aumenta, también aumenta la probabilidad de desarrollarlo

#### *1.4.1.3 Tipos de Alzheimer*

El Alzheimer de acuerdo a la edad en la que aparece, se puede clasificar como:

- De inicio precoz, si aparece antes de los 65 años.
- De inicio tardío, si aparece después de los 65 años-

Dentro de este a su vez se tiene dos subtipos:

- Familiar, si dentro de la historia familiar del paciente existen casos de demencia, principalmente diagnosticada como Alzheimer-
- Esporádica, si no se tienen casos de demencia dentro de la historia familiar del paciente.

#### *1.4.1.4 Fases del Alzheimer*

De acuerdo con la etapa en la que se encuentra, el Alzheimer se puede dividir en tres fases:

- Inicial, en la cual los síntomas y signos son leves y el paciente puede realizar las tareas cotidianas y mantiene autonomía en su comportamiento, y solo necesita de ayuda en las tareas más complejas.

- Intermedia o moderada, en la cual los síntomas y signos han aumentado de manera moderada, haciendo que la dependencia del paciente hacia su cuidador aumente también.
- Terminal o severa, en el que los síntomas y signos se encuentran ya tan avanzados que la dependencia del paciente hacia su cuidador es completa.

#### *1.4.1.5 Diagnóstico*

Ya que no todas las demencias son Alzheimer, se debe primero reconocer el tipo de demencia que padece el paciente, esto se realiza mediante una evaluación clínica completa del paciente. Para esto se realiza de manera inicial una valoración en la que los familiares deberán aportar con todos los detalles tanto de su comportamiento como de sus antecedentes familiares y médicos, además, se le realizará exámenes generales, neurológicos y neuropatológicos.

Dentro de la exploración neuropsicológica se encuentra el MMSE (Minimal State Examination), mismo que consta de una serie de preguntas simples destinadas a determinar el grado de orientación, memoria, atención, cálculo, lenguaje, escritura y dibujo del paciente (ver Anexo I, Figura 1).

En cuanto a los exámenes neuropatológicos, se realizan tomografías, y estudios con resonancia magnética, en las que se puede notar una conformación llamada ovillo neurofibrilar, que se halla en varias partes del cerebro, sustituyendo a las neuronas, esto se observa dentro del cerebro de los pacientes de la tercera edad de manera común, pero en un mayor porcentaje en los pacientes con Alzheimer.

#### *1.4.1.6 Pronóstico*

Dado que el Alzheimer no tiene cura y es degenerativo, no se tiene un buen pronóstico, ya que el paciente perderá de manera progresiva su independencia en las tareas diarias y luego incluso en el control de sus funciones corporales. Al presentarse el Alzheimer en los años dorados de los pacientes, esta termina al terminar la vida del paciente.

El Alzheimer no es la causa directa de la muerte del paciente, ya que no produce daño en los órganos vitales como corazón o sistema respiratorio, pero puede ser causa de accidentes graves si el paciente sale sin sus cuidadores y se enfrenta a los peligros del exterior sin compañía, se podría decir que, el Alzheimer es el acompañante que causa el deterioro de las capacidades mentales de su portador quitándole también su independencia e identidad.

#### *1.4.1.7 Tratamiento y prevención*

Al no existir una cura, el tratamiento es solo paliativo, es decir, busca el bienestar del paciente y el apoyo a los familiares y cuidadores por medio de ayuda psicológica y psiquiátrica.

Es importante entonces tratar de prevenir que se presente el síndrome, o que se retrase el máximo tiempo posible su aparición, para esto, se debe administrar al paciente vitaminas como: complejo B, E, C, ácido fólico y en general antioxidantes, mantener una buena alimentación, una vida sana y exenta de vicios.

## 2 MARCO TEÓRICO

Este marco teórico tiene como objetivo principal introducir los algoritmos, Análisis de Componentes Principales (PCA) y K-Means en el análisis de datos de pacientes con Alzheimer, para ello se proporciona una descripción general de cada algoritmo, incluyendo su historia, fundamentos matemáticos y aplicaciones específicas en el análisis de datos de esta patología.

Además, se incluye una breve descripción del lenguaje de programación empleado y el ambiente de desarrollo utilizado.

### 2.1 Análisis de Componentes Principales (PCA)

#### 2.1.1 Historia y origen del PCA

El análisis de componentes principales (PCA), en estadística, es un método para sintetizar un conjunto de datos en nuevas variables no correlacionadas (componentes). El PCA fue creado por Karl Pearson en 1901, más tarde fue formalizado y nombrado por Harold Hotelling en 1933 como *Análisis de Componentes Principales* o por sus siglas en inglés PCA (Principal Analysis Component) (Hotelling, 1933).

El PCA es un método estadístico que sirve para simplificar la complejidad de espacios lineales o conjuntos de datos con muchas dimensiones, conservando su información, es decir, existe una muestra con  $n$  individuos cada uno con  $k$  variables  $(X_1, X_2, \dots, X_k)$  teniendo un espacio lineal de  $k$  dimensiones (Hotelling, 1933).

El PCA permite disminuir la dimensionalidad de un conjunto de datos al identificar un número menor de factores (llamados componentes principales) que explican la mayor parte de la variabilidad de los datos originales (Hotelling, 1933).

Inicialmente, se tienen  $k$  variables que representan las características de cada individuo. Al aplicar PCA, se encuentran  $z$  componentes principales, donde  $z$  es menor que  $k$ . Cada uno de los componentes principales es una combinación lineal de las variables originales, y en conjunto, estas nuevas variables capturan la información esencial de los datos (Rodrigo, 2022).

En consecuencia, para caracterizar a cada individuo, solo se necesitan  $z$  valores de los componentes principales, en lugar de los  $k$  valores originales. Esto se entiende como una reducción significativa de la dimensionalidad de los datos, sin que esto signifique perder información relevante (Rodrigo, 2022).

Por tanto, el método PCA permite “concentrar” la información de múltiples variables en unos pocos componentes (Rodrigo, 2022)

### 2.1.2 Fundamentos matemáticos del PCA.

El PCA es una técnica matemática que transforma un conjunto de datos originales en un nuevo sistema de coordenadas, en el que, las variables que capturan la mayor cantidad de datos se representan en los primeros ejes, los datos deben ser estandarizados previamente (esto quiere decir que tengan como media 0 y desviación estándar 1) (Jolliffe, 1986).

Para lograr esta transformación lineal, se calcula una matriz cuadrada ( $k \times k$ ) nombrada como la matriz de covarianza o matriz de correlación del conjunto de datos original. Esta matriz refleja las relaciones entre ellas de todas las variables convirtiéndola en una gran herramienta para resumir un gran conjunto de datos (Jolliffe, 1986).

Se calcula la relación lineal entre dos variables. La matriz se construye a partir del cálculo del coeficiente de correlación para cada par de variables, e insertándolo en la celda correspondiente, al estar los datos estandarizados desde el principio se puede considerar que la matriz de covarianza y la matriz de correlación de Pearson son iguales esto se demuestra en el Anexo 2 tabla (1), este proceso se lleva a cabo con la ecuación 1 (Jolliffe, 1986).

$$Corr(X_{est}, Y_{est}) = \frac{1}{(n-1)} \sum_{i=1}^n (X_{est,i})(Y_{est,i})$$

**Ecuación 1.** Fórmula para calcular los elementos de la matriz de correlación de Pearson

Disponible en (Jolliffe, 1986)

La matriz de covarianza/correlación ( $Corr(X)$ ), presenta los vectores propios y valores propios. Los vectores propios (Eigenvectors) son no nulos, se calculan mediante la multiplicación entre la matriz de correlación y un vector ( $V_i$ ) dando como resultante, un vector del cual se puede decir que es la multiplicación de un múltiplo ( $\lambda_i$ ) por el vector original. A todo vector propio le corresponde su valor propio y viceversa, esto se puede entender de mejor manera con un ejemplo práctico como se muestra en la ecuación 2 (Profesores del departamento de matemática aplicada de la Escuela Técnica Superior de Ingeniería, 2011).

$$Corr(X) \cdot V_i = \lambda_i \cdot V_i$$

Donde  $i$  varía desde 1 hasta  $k$

## **Ecuación 2.** Cálculo de los vectores y valores propios

Disponible en (Jolliffe, 1986)

Los vectores propios solo existen en matrices cuadradas, son ortogonales(perpendiculares) entre ellos, independientemente de las dimensiones que tengan (Profesores del departamento de matemática aplicada de la Escuela Técnica Superior de Ingeniería, 2011).

Se construye la matriz característica a partir de la matriz de correlaciones y restando un valor  $\lambda$  por la matriz identidad, luego calcular el determinante obteniendo el polinomio característico, encontrando las soluciones de  $\lambda$  para esta ecuación que son los valores propios y con la ecuación 2 se obtiene los vectores propios (Profesores del departamento de matemática aplicada de la Escuela Técnica Superior de Ingeniería, 2011).

En el método PCA, a cada uno de los componentes le corresponde un vector y un valor propios, y el orden del componente se establece de manera decreciente utilizando los valores propios ( $\lambda_i$ ), por tanto, el primer componente tiene el vector y valor propio asociado más alto, es decir selecciona los vectores propios(componentes) con mayores valores propios (variabilidad explicada) (Jolliffe, 1986).

El cálculo de los componentes principales tiene como fin obtener cada componente principal  $Z_j$ , se obtiene por combinación lineal de las variables originales, es decir, que son nuevas variables obtenidas al combinar de una determinada forma las variables originales.

Se tiene un grupo de variables ( $X_1, X_2, \dots, X_k$ ) que tienen mayor varianza, se realiza la combinación lineal normalizada como se muestra en las ecuaciones 3 y 4 (Mangas, 2012).

$$\sum_{j=1}^k \phi_{j1}^2 = 1$$

### **Ecuación 3.** Definición de combinación lineal normalizada

$$Z_1 = \phi_{11}X_1 + \phi_{21}X_2 + \dots + \phi_{k1}X_k$$

### **Ecuación 4,** Combinación lineal normalizada

Donde:

$\phi_{11}, \dots, \phi_{k1}$  se puede interpretar como la importancia de cada variable en cada componente, este se denomina Loadings.

Luego del cálculo del primer componente ( $Z_1$ ) se calcula el segundo componente ( $Z_2$ ) repitiendo el mismo proceso, pero con la condición de que la combinación lineal no puede estar correlacionada con el primer componente, es decir que ( $Z_1$ ) y ( $Z_2$ ) sean ortogonales(perpendiculares). Este proceso se repite de forma iterativa hasta calcular todos los componentes principales o hasta que se decida detener el proceso (Mangas, 2012).

### 2.1.3 Aplicaciones y ventajas del PCA

Existen varias aplicaciones del PCA, por ejemplo:

Finanzas, predecir los precios de las acciones en un tiempo futuro, basándose en su comportamiento histórico ha sido un tema de investigación durante mucho tiempo, el PCA es una herramienta útil para este propósito, ya que permite simplificar y analizar conjuntos de datos complejos relacionados con los precios de las acciones. El PCA funciona reduciendo la dimensionalidad de los datos, facilitando el análisis y la identificación de patrones y tendencias que pueden ayudar a predecir los precios futuros (Datacamp, 2024).

Procesamiento de imágenes, el PCA se utiliza ampliamente para comprimir imágenes manteniendo la calidad, es decir al reducir la dimensionalidad de los datos, se conservan detalles esenciales de la imagen mientras su tamaño es reducido siendo principalmente útil para transmitir compartir y almacenar de manera eficiente (Datacamp, 2024).

El PCA también se emplea en tareas de reconocimiento de imágenes como la clasificación de objetos, la detección de rostros y la identificación de patrones ya que, al analizar los componentes principales de una imagen, se puede extraer características distintas que permiten discriminar las imágenes (Datacamp, 2024).

Sanidad, en este campo, el PCA se utiliza para el análisis de imágenes médicas como, resonancias magnéticas (MRI) y tomografías computarizadas (TC), reduciendo la dimensionalidad de las imágenes, facilitando la visualización e interpretación por parte de los profesionales médicos, el PCA puede ayudar a detectar y diagnosticar enfermedades a partir de imágenes médicas (Datacamp, 2024).

Seguridad, los sistemas biométricos, como el recogimiento de huellas dactilares y reconocimiento facial, utilizan PCA para extraer características relevantes de los datos biométricos, analizando los componentes principales de una huella dactilar o rostro, identificando patrones únicos que permiten la identificación y verificación de individuos (Datacamp, 2024).

## 2.2 K-Means clustering

### 2.2.1 Historia y origen del K-Means

El algoritmo K-Means (K-Medias) fue propuesto como una idea y presentado por Hugo Steinhaus en 1957, en 1965 E.W. Forgy publicó un método similar, más tarde en el año 1967 el término K-Means “K-Medias” fue adoptado por James MacQueen, publicándose en 1982. (Hartian, 1975).

K-Means es un algoritmo de agrupamiento que tiene por objetivo dividir un conjunto de datos en K grupos, cada observación se asigna al grupo cuyo promedio, que de aquí en adelante se denominara centroide, es más cercano, la cantidad de grupos a encontrar es el parámetro K que debe fijarse al principio, es ampliamente utilizado en minería de datos para descubrir patrones y estructuras ocultas en grandes conjuntos de datos (Hartigan, 1975).

El método de clustering da inicio con K centroides que son tomados aleatoriamente de filas de la data, y establece cada observación al centroide más cercano, luego de ser asignados, los centroides se mueven a la ubicación promedio de todos los datos designados a ellos, y se vuelven a reasignar los puntos de acuerdo a las nuevas posiciones (Hartigan, 1975).

### 2.2.2 Fundamentos matemáticos del K-Means

El objetivo principal del algoritmo es agrupar las observaciones de un conjunto de datos dentro un mismo grupo con características similares entre sí, para lograr esto, K-Means utiliza la distancia euclídea para cuantificar la similitud (cerca del centroide del grupo) o no semejanza (lejos del centroide del grupo) entre las observaciones (Jurjo, 2022).

Una forma de evaluar que tan bien representan a los miembros de un grupo es mediante los centroides, la suma de errores al cuadrado que calcula la suma de las distancias al cuadrado de cada observación al centroide de su grupo, si se tiene un valor bajo, el centroide representa bien a su grupo, mientras que, con un valor alto, la representación no es buena (Jurjo, 2022).

El K-Means se considera un proceso iterativo, es decir, que repite el proceso hasta llegar a una solución, por cada iteración, trata de minimizar el valor de la suma de los errores al cuadrado, reasignando las observaciones a los grupos de manera que, cada observación quede en un grupo cuyo centroide sea el más cercano, la medida es entonces la sumatoria de las distancias al cuadrado de cada observación al centroide de su grupo como se demuestra en la ecuación 5 (Jurjo, 2022).

$$\min_S E(\mu_i) = \min_S \sum_{i=1}^k \sum_{x_j \in S_i} \|x_j - \mu_i\|^2$$

### **Ecuación 5.** Cálculo de las distancias (observación-centroide)

Disponible en (Jurjo, 2022)

Donde:

$\min_S$  las asignaciones, S, que minimiza la fórmula.

k cantidad de grupos

$x_j \in S_i$  por cada punto asignado al grupo i

$\mu_i$  centroide del grupo i

Luego, para el refinamiento iterativo el algoritmo asigna cada observación al grupo con el centroide más cercano, como se muestra en la ecuación 6.

Sea  $m_1^{(1)}, m_2^{(1)}, \dots, m_k^{(1)}$  conjunto inicial de centroides.

$$S_i^{(t)} = \{x_p: \|x_p - m_i^t\| \leq \|x_p - m_j^t\| \forall 1 \leq j \leq k\}$$

### **Ecuación 6.** Asignación de observación al centroide

Disponible en (Jurjo, 2022)

Luego, se calculan los nuevos centroides, como el centroide de las observaciones del grupo, como se muestra en la ecuación 7.

$$C_k = \frac{1}{|C_k|} \sum_{x \in C_k} x$$

### **Ecuación 7.** Actualización de centroides.

disponible en (Jurjo, 2022)

Se considera que el algoritmo ha convergido, cuando las asignaciones de los centroides no cambian significativamente.

#### 2.2.3 Aplicaciones y Ventajas del K-Means

El algoritmo K-Means es dada sus ventajas, uno de los métodos de clustering de uso generalizado y más populares en el aprendizaje automático y en la analítica de datos y tiene varias aplicaciones.

Segmentación de clientes, en marketing se usa para segmentar clientes en grupos basados en sus comportamientos y características, así se identifican grupos de consumidores, permitiendo la creación de estrategias de marketing más personalizadas y efectivas (Congacha Griogia, 2020).

Detección de fraude, dentro de la banca y finanzas para la identificación de patrones anómalos en transacciones financieras que podrían indicar fraude (Congacha Griogia, 2020).

Análisis de redes sociales, para identificar grupos de usuarios con intereses o comportamientos similares en redes sociales (Congacha Griogia, 2020).

El K-Means es un método simple, fácil de entender y de implementar, siendo accesible para una amplia gama de usuarios, además, es computacionalmente eficiente y puede procesar grandes conjuntos de datos de manera rápida, demostrando ser efectivo para encontrar grupos con centros definidos (Congacha Griogia, 2020).

## **2.3 Comparación de PCA y K-Means.**

### **2.3.1 Diferencias clave entre PCA y K-Means.**

Mientras que el PCA se centra en reducir la dimensionalidad de un conjunto de datos, el algoritmo K-means se emplea para agrupar los datos en clusters o grupos homogéneos.

El PCA transforma los datos originales en un nuevo espacio de menor dimensión, conservando la mayor parte de la información relevante. Esto permite visualizar y analizar los datos de manera más sencilla, identificando los patrones y tendencias subyacentes.

Por otro lado, K-means agrupa los datos en clusters basándose en la proximidad entre ellos. El objetivo es encontrar grupos de puntos que sean lo más similares posible dentro de cada grupo y lo más diferentes posible entre grupos.

En resumen, el PCA se utiliza para simplificar la representación de datos complejos, mientras que K-means se utiliza para descubrir estructuras y relaciones dentro de los datos.

### **2.3.2 Combinación de PCA y K-Means en el análisis de Alzheimer.**

La combinación de PCA y K-Means proporciona herramientas de gran utilidad para el análisis de datos complejos en el contexto del Alzheimer, al reducir la dimensionalidad y agrupar a los pacientes en clusters significativos, se pueden identificar patrones y características relevantes que son importantes para la comprensión e identificación de esta enfermedad.

## **2.4 Ambiente de trabajo y lenguaje de programación**

#### 2.4.1 Google colab

Google Colab (Colaboraty) es una herramienta de Google, la cual fue diseñada para escribir y ejecutar código Python directamente en el navegador, siendo especialmente útil para proyectos de análisis de datos y aprendizaje automático.

Colab es una plataforma que se encuentra en la nube, es decir, que con ella se puede acceder a los notebooks desde cualquier dispositivo, en cualquier lugar que cuente con conexión a internet, viene preconfigurado con muchas bibliotecas de Python para la ciencia de datos, aprendizaje automático y procesamiento de imágenes como TensorFlow, PyTorch, Pandas, NumPy, y Scikit-Learn entre otras, la plataforma de Colab ofrece acceso gratuito a recursos computacionales potentes como GPU y TPU (unidades de procesamiento tensorial), que permiten entrenar modelos de aprendiza automático eficientemente (Google, 2024).

#### 2.4.2 Python

Es un lenguaje de programación de alto nivel, muy utilizado y de propósito general en la ciencia de datos y el aprendizaje automático ya que combina simplicidad y funcionalidad (Python, 2024).

Python tiene una sintaxis clara y concisa, lo que lo hace fácil de leer y escribir, permitiendo a los desarrolladores enfocarse en resolver problemas gracias a una amplia colección de bibliotecas estándar que facilitan diversas tareas, desde la manipulación de datos hasta la construcción de modelos de aprendizaje no supervisado en lugar de preocuparse por la complejidad del lenguaje (Python , 2024).

## 3 METODOLOGÍA

### 3.1 Importación de librerías

La primera sección del cuaderno implica la importación de librerías necesarias para la manipulación de datos, preprocesamiento, desarrollo de modelos, y visualización de resultados, a continuación, se dará una explicación más detallada de su función para este proyecto.

#### 3.1.1 Pandas

Esta librería se especializa en el manejo y análisis de estructura de datos permitiendo leer y escribir fácilmente en ficheros en formato CSV de Excel, accede a los datos mediante nombres de filas y columnas, para este proyecto se utilizó el DataFrame que es una estructura de dos dimensiones (tablas) (Alberca A. S., 2020).

Pandas cuenta con una serie de funciones que permiten hacer un resumen descriptivo de los atributos, ver las características, añadir y eliminar columnas de un DataFrame (Alberca A. S., 2020).

Valores faltantes, dentro de esta librería existe una serie de funciones que nos permite encontrar, sumar y rellenar valores faltantes o nulos dentro de un DataFrame, para lo que se utilizan los métodos "isnull().sum()" que devuelve la suma de valores nulos dentro del DataFrame (Burrueco i. D., 2022) (pandas.pydata.org, 2024).

Luego se aplicó el método "fillna" que permite sustituir los valores inexistentes de un DataFrame por un valor dado de acuerdo con ciertos criterios, y se pueden sustituir por un valor en concreto que para este caso es la media aritmética debido a que los lugares donde faltaban datos eran columnas numéricas, la media aritmética se utiliza para reducir el sesgo en los resultados del análisis (Burrueco I. D., 2022).

La función "get dummies" se aplicó para la codificación denominada "one hot encoding" la serie de pandas para codificar una columna categórica que da como resultado un Dataframe que incluye tanto los índices de la fila como de la columna, esto quiere decir que se agrega un registro diferente a cada columna y los registros pasan a ser sustituidos por valores binarios (0,1), 1 en caso de ser positivo, y 0 en el caso contrario (Burrueco D. , <https://interactivechaos.com>, 2022).

El método "corr()" en un DataFrame es utilizado para calcular la correlación entre las columnas numéricas, este método nos permite especificar el algoritmo de cálculo, en este

caso se tiene por defecto(Pearson). “El índice de correlación de Pearson es una medida de relación lineal entre dos variables y se basa en la covarianza entre ellas” (Torres, 2020).

“La correlación es una medida estadística que indica la relación lineal entre dos variables, el valor oscila entre -1 y 1 donde 1 es una correlación positiva perfecta y -1 es una correlación negativa perfecta” (Torres, 2020).

### 3.1.2 NumPy

Es una librería especializada en cálculo numérico y el análisis de datos, proporciona una herramienta para trabajar con arreglos multidimensionales y funciones matemáticas de alto nivel (Alberca D. A., <https://aprendeconalf.es>, 2021).

### 3.1.3 Matplotlib.pyplot

Es una colección de funciones que trabaja como Matlab, esta función mantiene varios estados a través de la llamada de la función que realiza un seguimiento de elementos como la figura en la que se está trabajando, es decir, el área de trazado (Matplotlib, 2021).

Esto permite la creación de histogramas, varios tipos de diagramas y mapas de color, cuenta con una manera de personalizar los colores, marcadores de puntos, estilo de líneas hasta títulos del gráfico y los ejes (Alberca D. A., <https://aprendeconalf.es>, 2020).

### 3.1.4 Seaborn

“Es una herramienta de visualización de datos, permite transformar datos brutos en gráficos y diagramas que facilitan la lectura y comprensión de los datos, está basada en matplotlib y se integra con las estructuras de pandas, es una herramienta ideal para la visualización estadística que en este proyecto se utilizó para crear un mapa de calor y así visualizar las correlaciones de las variables” (Data Science, 2023).

### 3.1.5 Sklearn.model\_selection

Proporciona herramientas para dividir un Dataset en dos bloques de datos destinados a ser de entrenamiento y validación del modelo, también se utiliza para la validación cruzada y la selección de parámetros, es útil sobre todo para realizar la división de arreglos o matrices en subconjuntos aleatorios de entrenamiento (train) y ensayo (test) (David Cournapeau, <https://scikit-learn.org>, 2021).

### 3.1.6 Sklearn.preprocessing:

Esta librería incluye varias herramientas como: clases y funciones para el preprocesamiento de los datos y transformadores para cambiar características sin

procesar, así como escalado, normalización y codificación de variables categóricas (David Cournapeau, <https://scikit-learn.org>, 2021), para este proyecto se utilizaron las siguientes:

#### 3.1.6.1 *StandardScaler*:

Esta clase estandariza los datos, este tipo de escalado se denomina normalización estándar o escalado estándar o normalización z, estandariza los datos al eliminar la media y escalando los datos de manera que su varianza sea igual a 1 (Burrueco I. D., 2022).

#### 3.1.6.2 *LabelEncoder*:

“Codifica etiquetas de una columna categórica en valores numéricos entre 0 y el número de clases menos 1, una vez instanciado el método “fit:transform” primero crea el mapeado entre las etiquetas y los números para luego transformar las etiquetas que se incluyan como argumentos en los números correspondientes” (Burrueco D. , <https://interactivechaos.com>, 2022).

#### 3.1.7 *Sklearn.decomposition*:

Es una librería que utiliza algoritmos de descomposición matricial como el algoritmo PCA esta librería se puede considerar para técnicas de reducción de dimensionalidad (David Cournapeau, <https://scikit-learn.org>, 2021)

#### 3.1.8 *Sklearn.cluster*:

Esta librería se utiliza para algoritmos de agrupamiento entre mínimo dos variables, y da como resultado una matriz de etiquetas correspondientes a cada cluster, para la clase etiquetas y después del entrenamiento, los datos se pueden encontrar en el atributo “labels\_” (David Cournapeau, <https://scikit-learn.org>, 2021).

Esta librería cuenta con varios métodos de los cuales el que se utilizó fue el método K-Means que agrupa observaciones que son similares para descubrir los patrones que a simple vista se desconocen, utiliza la regla de codo que nos representa el número de centroides. “El objetivo del K-Means es minimizar la suma de las distancias entre los puntos y el centroide al que pertenece” (Sanz, 2015).

#### 3.1.9 *Sklearn.metrics*:

Es un módulo que proviene de “Scikit-learn” que proporciona una variedad de funciones para medir el rendimiento de modelos de aprendizaje automático, estas métricas se pueden utilizar para evaluar los modelos de aprendizaje no supervisado (David Cournapeau, <https://scikit-learn.org/stable/about.html>, 2021).

## 3.2 Descripción del dataset:

Esta sección describe cómo se carga el conjunto de datos `alzheimer.csv` y se realiza un análisis inicial para entender su estructura y contenido.

El conjunto de datos de nombre "alzheimer.csv" fue adquirido dentro de la plataforma web "Kaggle" disponible en "*Alzheimer Features*" (Dincer, 2021).

En esta parte se cargan el conjunto de datos y se realiza un análisis inicial para comprender su estructura y contenido.

Dentro de la data o conjunto de datos encontramos una serie de características del Alzheimer las cuales se las denominará de aquí en adelante como variables y variable de trabajo, las cuales se mencionan a continuación junto con una pequeña descripción de cada una.

### Variables

- M/F: Género del sujeto (M para masculino, F para femenino).
- Age: Edad del sujeto en el momento del estudio.
- Education: Nivel educativo del sujeto medido en años de educación.
- SES (Socioeconomic Status): Estatus socioeconómico del sujeto, medido en una escala cualitativa.
- MMSE (Mini-Mental State Examination): Puntuación en una prueba estandarizada que evalúa el estado mental del sujeto. Se utiliza comúnmente para diagnosticar demencia.
- CDR (Clinical Dementia Rating): Escala que mide la severidad de la demencia en el sujeto, desde 0 (sin demencia) hasta 3 (demencia severa).
- eTIV (Estimated Total Intracranial Volume): Volumen total intracraneal estimado, medido en milímetros cúbicos.
- nWBV (Normalized Whole Brain Volume): Volumen total del cerebro normalizado, expresado como un porcentaje del volumen intracraneal total.
- ASF (Atlas Scaling Factor): Factor de escala utilizado para normalizar los volúmenes cerebrales según un atlas médico de referencia.

### Variable de trabajo

Group: Indica la clasificación del sujeto en uno de los tres grupos:

- Demented: Personas diagnosticadas con demencia.
- No Demented: Personas diagnosticadas sin demencia.
- Converted: Personas que inicialmente se diagnosticó que no tenían demencia, pero que desarrollaron la enfermedad durante el seguimiento.

Se utiliza la función “read\_csv” de la librería pandas para leer el archivo alzheimer.csv y cargarlo en un DataFrame, que es una estructura de datos tabular que facilita la manipulación y análisis de los datos

Se emplea el método “info()” de la librería pandas para obtener una descripción rápida del conjunto de datos, esta descripción incluye el número de filas(entradas) y de columnas(variables), tipo de datos de cada columna(numéricos, categóricos), y el número de valores no nulos de cada columna, con esta información se puede identificar la presencia de valores faltantes o valores nulos para comprender de manera general el conjunto de datos.

Existen variables numéricas y variables categóricas (contienen texto), de las 10 columnas de las que consta el conjunto de datos, dos de ellas son categóricas “M/F” y la variable de trabajo “Group”. Para lo cual primero se realizará un tratamiento de datos para su correcto manejo.

Al realizar el primer análisis se puede notar que hay presencia de valores faltantes para lo que se va a requerir un tratamiento de datos para su análisis posterior.

### **3.3 Preprocesamiento de datos:**

El preprocesamiento de datos es una etapa importante en cualquier proyecto que requiera análisis de datos, ya que asegura la calidad y coherencia de los datos antes de aplicarlos en modelos de aprendizaje automático, en esta sección se describen las técnicas de preprocesamiento aplicadas al conjunto de datos “alzheimer.csv”.

#### **3.3.1 Manejo de valores faltantes:**

El primer paso para el preprocesamiento de datos es identificar las columnas que contienen valores faltantes y la cantidad de estos valores, la columna “SES” cuenta con 19 valores faltantes, la columna “MMSE” tiene 2 valores faltantes, la decisión de imputar o eliminar dichos valores puede depender de varios factores como el porcentaje de valores faltantes.

Cuando existe 5 % o menos de valores faltantes generalmente, se recomienda asignar estos valores faltantes con la media, mediana o moda, como para nuestro conjunto de datos el porcentaje de valores nulos no excede el 5 % se aplicó el comando para ingresar en las celdas vacías la media aritmética (GeeksforGeeks, 2024).

### 3.3.2 Codificación de variables categóricas:

Las variables categóricas son las que permiten clasificar una serie de datos mediante la utilización de valores fijos asociados a una cualidad o categoría, su entrada suele ser texto, para poder manejarlas se recurre a convertirlas a forma numérica para ser utilizadas posteriormente.

El método que se utilizó fue el “one hot encoding”, que convierte variables categóricas en una serie de columnas binarias que corresponde a 1 cuando en esa celda estuvo dicho valor categórico y en caso de 0 es cuando no se tuvo ese valor en esa celda. Se aplicó este método a la columna del género “M/F” que contiene dos tipos de valores categóricos M para masculino y F para femenino fue reemplaza por las columnas binarias “M/F\_F” y “M/F\_M”.

### 3.3.3 Revisión y limpieza final:

Se realizó una inspección final del DataFrame para asegurar que no existan errores y que todas las columnas estén en el formato adecuado. El método anteriormente nombrado elimina la columna original, pero se realizó una inspección para corroborar ese hecho, el DataFrame ha sido completamente procesado y está listo para aplicarle los modelos de aprendizaje automático.

## **3.4 Análisis de correlaciones, selección y reducción de variables.**

Para el primer análisis se realizó una copia del Dataframe para en ella realizar la conversión de la variable de trabajo de categórica a numérica con el método “LabelEncoder” esto ayuda a visualizar de mejor manera las correlaciones entre las variables con respecto a la variable de trabajo y a no afectar al DataFrame original.

### 3.4.1 Matriz de correlación:

Es una tabla que muestra las correlaciones entre todas las variables del DataFrame, cada celda en la matriz contiene el coeficiente de correlación de dos variables específicas, el “coeficiente de correlación de Pearson” es la medida más utilizada en este análisis, que

cuantifica la dirección y la fuerza de la relación lineal entre dos variables, su valor oscila entre -1 y 1, donde:

1, indica que existe una correlación positiva para los valores cercanos a 1 pero en el caso de 1 se conoce como una correlación positiva perfecta.

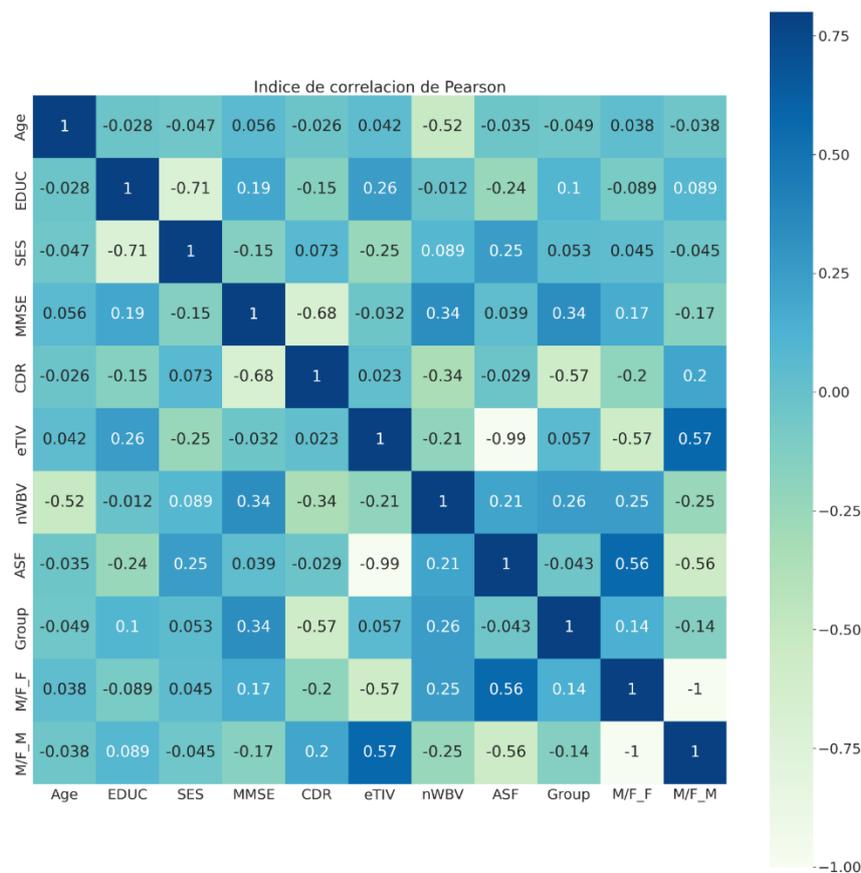
0, al estar cercanos al 0, esto indica que la correlación es baja entre esas variables y que en algunos casos puede generar ruido lo que llevaría a la mala interpretación del modelo o puede conducir a un problema de sobreajuste por lo cual si el valor está oscilando entre -0,1 y 0,1 es recomendable eliminar esas variables.

-1, indica que hay una correlación negativa en valores cercanos al -1, pero, en el caso de ser -1, la correlación es negativa perfecta.

La matriz de correlación es una tabla que muestra los coeficientes de correlación que se encuentra entre todas las combinaciones de variables del conjunto de datos.

#### 3.4.2 Visualización de la matriz de correlación

Para facilitar la interpretación de la matriz de correlación, se utiliza un mapa de calor (heatmap), esto permite visualizar los coeficientes de correlación con colores donde los colores más intensos representan correlaciones más fuertes sean estas positivas o negativas, con esta visualización se busca identificar de manera eficiente las relaciones más fuertes entre las variables. Esto se puede ver de mejor manera en la Figura 1.



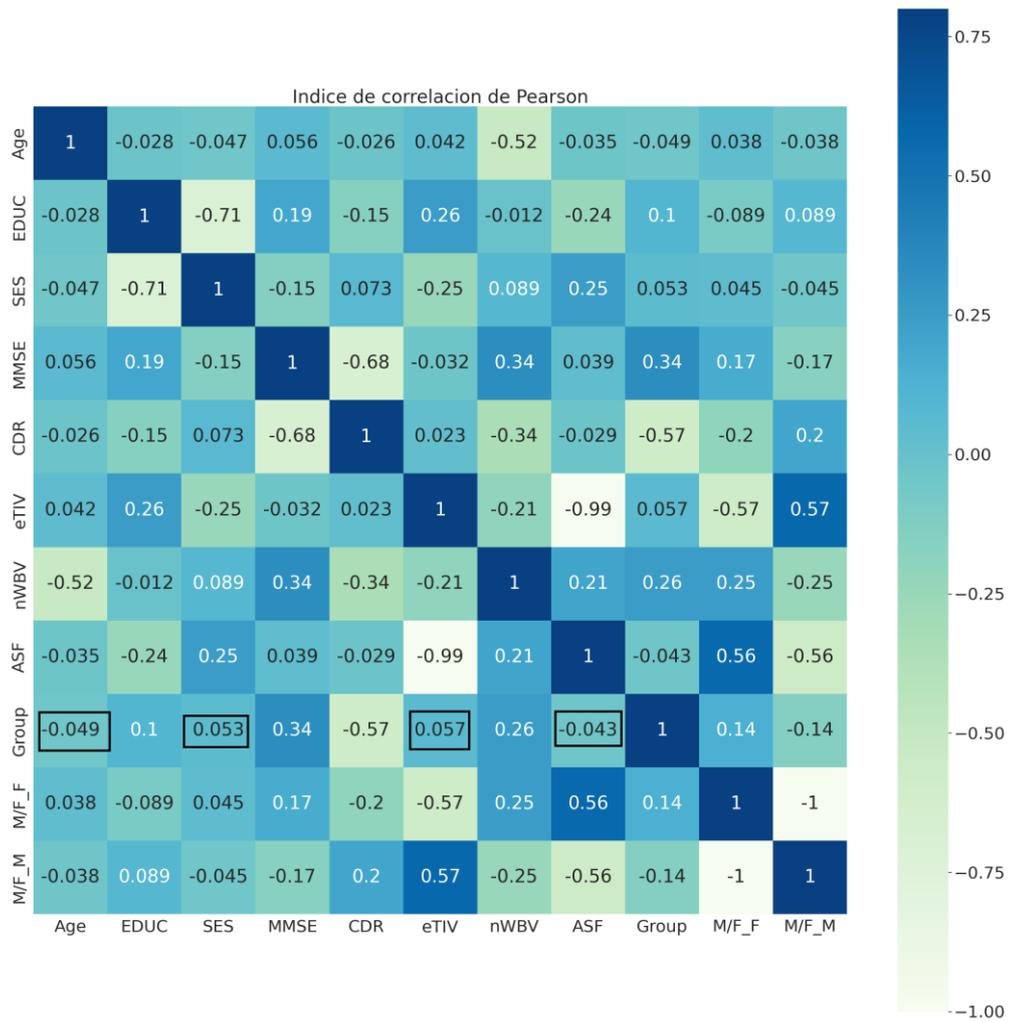
**Figura 1.** Índice de correlación de Pearson. Una representación visual de la matriz de correlación que facilita la identificación de correlaciones fuertes y débiles entre las variables.

(Autoría propia)

### 3.4.3 Análisis de correlaciones relevantes.

En la identificación de correlaciones fuertes las variables cercanas a 1 o -1 se consideran altamente correlacionadas, estas variables pueden estar correlacionadas de manera positiva o negativa para determinar el tipo de correlación se realizó el análisis para la variable "Group".

Después del primer análisis se pueden encontrar correlaciones que están por debajo del  $\pm 0,1$  como se muestra en la Figura 2, por esto se tomó la decisión de eliminarlas del Dataframe original, las variables eliminadas son "Age", "SES", "eTIV", "ASF", debido a que en el gráfico de correlación de Pearson se muestra que dichas variables no son influyentes dentro de la data existe la posibilidad de que generen ruido.



**Figura 2.** Índice de correlación de Pearson, variables a ser eliminadas por su baja correlación

(Autoría propia)

Una vez eliminadas las variables de correlación baja, la matriz queda como se muestra en la Figura 3.

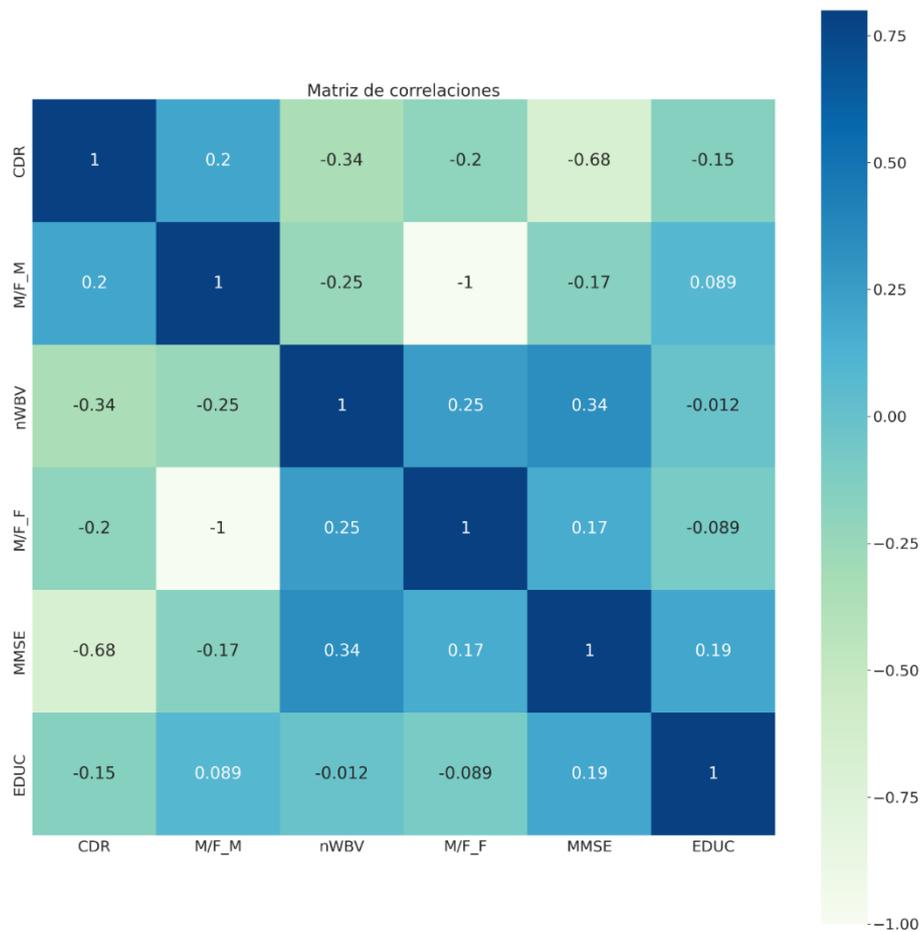


**Figura 3.** Índice de correlación de Pearson final

(Autoría propia)

#### 3.4.4 Selección de variables:

En esta sección se seleccionan y reducen las variables para mantener solo las variables de trabajo, para esto se verifica el DataFrame para luego excluir las variables categóricas del DataFrame que en este caso es solo una y es la variable de trabajo, a continuación, se muestra la matriz de correlaciones con las variables escogidas en la Figura 4.



**Figura 4.** Matriz de correlaciones  
(Autoría propia)

### 3.5 Escalado de variables

El escalado de variables es un paso importante para el preprocesamiento de datos, especialmente cuando se utilizan algoritmos de aprendizaje automático que son sensibles a la escala de las características asegurando que todas las variables contribuyan de manera equitativa al modelo y ayuden a mejorar su desempeño.

#### 3.5.1 Escalado de variables

Se realiza este proceso debido a que las variables con mayores rangos pueden dominar a aquellas con rangos menores, lo que puede sesgar el modelo, por ello el escalado asegura que todas las variables tengan una influencia equitativa al momento de pasar por los algoritmos, este proceso se realiza, principalmente, cuando hay algoritmos como el PCA o los basados en distancia como el K-Means (Burrueco D. , <https://interactivechaos.com>, 2022).

### 3.5.2 Métodos comunes de escalado

En este caso existen dos métodos de escalado, los más utilizados y son la estandarización y normalización.

“Estandarización, es una técnica utilizada en el análisis de datos y la estadística, transforma las características para que tengan una media de 0 y una desviación estándar de 1” (Urrego, 2023).

Normalización, es utilizada en el análisis estadístico de datos, escala las características para que estén en un rango específico usualmente entre 0 y 1 (Urrego, 2023).

### 3.5.3 Estandarización de variables

Para este proceso se utilizó una de las clases de la librería “sklearn.preprocessing” el cual se utiliza para estandarizar las características eliminando la media y escalando a la varianza unitaria.

Para un mejor entendimiento, la fórmula matemática detrás del algoritmo se presenta en la ecuación 8.

$$z = \frac{x - \mu}{\sigma}$$

#### **Ecuación 8. Estandarización Z**

Disponible en: (Universidad de Mucia, 2023)

Donde:

x: valor original

$\mu$ : media de la característica

$\sigma$ : desviación estándar de la característica

Este proceso da como resultado un array, donde cada característica tiene una media de 0 y una desviación estándar de 1.

El proceso de escalado de variables asegura que todas las características contribuyan de manera equitativa con el modelo y mejore su eficiencia y la precisión del aprendizaje automático, el método empleado es ampliamente utilizado para transformar las características a una escala con media 0 y desviación estándar de 1 (Universidad de Mucia, 2023).

## 3.6 Desarrollo del modelo de aprendizaje no supervisado

El desarrollo de modelos de aprendizaje no supervisado es una etapa importante para identificar patrones y relaciones en los datos, sin la necesidad de etiquetas predefinidas, en este ambiente, se han implementado dos técnicas principales: el Análisis de Componentes Principales (PCA) y el algoritmo K-Means. A continuación, se describen estas dos técnicas.

### 3.6.1 Implementación de PCA

“Una de las principales dificultades en la estadística multivariada es la visualización de datos, pudiendo realizarse gráficas para dos o hasta tres variables, es decir, con una vista bi o tridimensional, pero cuando existen más de tres variables ya no es posible visualizar sus relaciones, sin embargo, en los conjuntos de datos con más de tres variables, suelen existir grupos de variables que relacionan la misma información, esta es una de las razones por las que es posible que exista más de una variable midiendo el mismo principio motor que rige el comportamiento de los datos, se puede sacar partido de esta redundancia simplificando y sustituyendo un grupo de variables por una nueva variable única” (Mathworks, 2024).

“El PCA tiene como propósito la reducción de la dimensionalidad que transforma un conjunto de variables posiblemente correlacionadas en un conjunto más pequeño de variables no correlacionadas llamadas componentes principales, esto ayuda a simplificar el modelo, reducir el ruido y mejorar la interpretabilidad de los datos” (Burrueco D. , <https://interactivechaos.com>, 2022).

La forma en que el PCA funciona es identificar las direcciones (componentes principales) en las que los datos varían más y proyecta los datos en estas nuevas dimensiones (Burrueco D. , <https://interactivechaos.com>, 2022).

“El primer componente principal es el eje inicial en el espacio, al proyectar las observaciones sobre ese eje los resultados forman una nueva variable, la varianza es la máxima entre todas las opciones del primer eje” (Mathworks, 2024).

“El segundo componente principal es el otro eje en el espacio, perpendicular al primero, proyectar las observaciones de este eje genera otra nueva variable, su varianza es máxima entre todas las posibles opciones del segundo eje” (Mathworks, 2024).

Al aplicar el algoritmo PCA sobre nuestro conjunto de datos escalados se obtienen los vectores propios y valores propios en forma de un array como se muestra en las Figuras 5 y 6.

Lo vectores propios nos indican como se relacionan los componentes principales con las variables originales, mientras que los valores propios nos dicen cuanta información se halla presente del conjunto de datos original, es decir, cuanta información es capaz de capturar cada componente principal.

```
array([2.47217286e+00, 1.60398331e+00, 9.30025817e-01, 6.90811832e-01,
       3.19135214e-01, 3.14627103e-32])
```

**Figura 5.** Valores propios del PCA

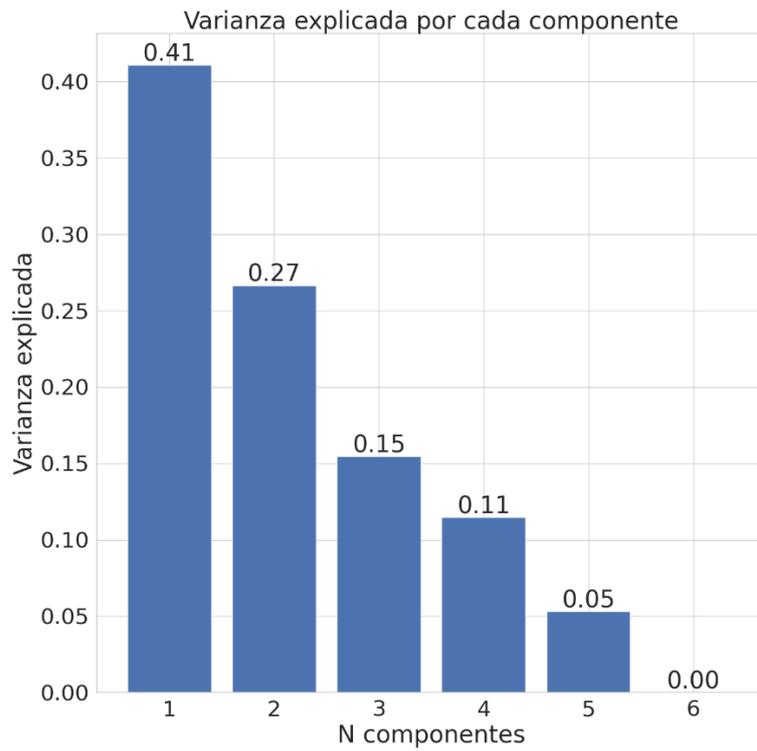
Autoría propia

```
array([[ 4.17197061e-01, -3.08921679e-02, -5.14971918e-01,
        5.14971918e-01, -3.66675539e-01, -4.00186436e-01],
       [ 4.46580121e-01, -3.97831561e-01,  4.37612139e-01,
       -4.37612139e-01, -1.59961531e-01, -4.83425056e-01],
       [ 1.03886751e-01,  8.33791264e-01,  2.05111335e-01,
       -2.05111335e-01, -4.52877708e-01, -6.89937156e-02],
       [ 3.51702114e-01,  3.79218071e-01, -3.23437081e-02,
        3.23437081e-02,  7.96685643e-01, -3.09352601e-01],
       [-7.01445631e-01,  4.19702875e-02, -1.35190648e-02,
        1.35190648e-02,  1.24548982e-02, -7.11120149e-01],
       [-0.00000000e+00, -3.35658909e-15, -7.07106781e-01,
       -7.07106781e-01,  3.94169733e-15, -6.31804910e-16]])
```

**Figura 6.** Vectores propios del PCA

Autoría propia

La varianza explicada y la varianza explicada acumulada, son vectores de mucha utilidad cuando se debe decidir el número de componentes principales que se van a utilizar, en este primer análisis se calculan todos los componentes principales del conjunto de datos original estandarizado por lo que, se está almacenando toda la información de la data original, el sumario de la proporción de varianza explicada acumulada de todas las componentes es siempre 1, en la Figura 7 se puede observar la varianza explicada y en la Figura 8 la varianza explicada acumulada



**Figura 7.** Varianza explicada por cada componente  
Autoría propia.

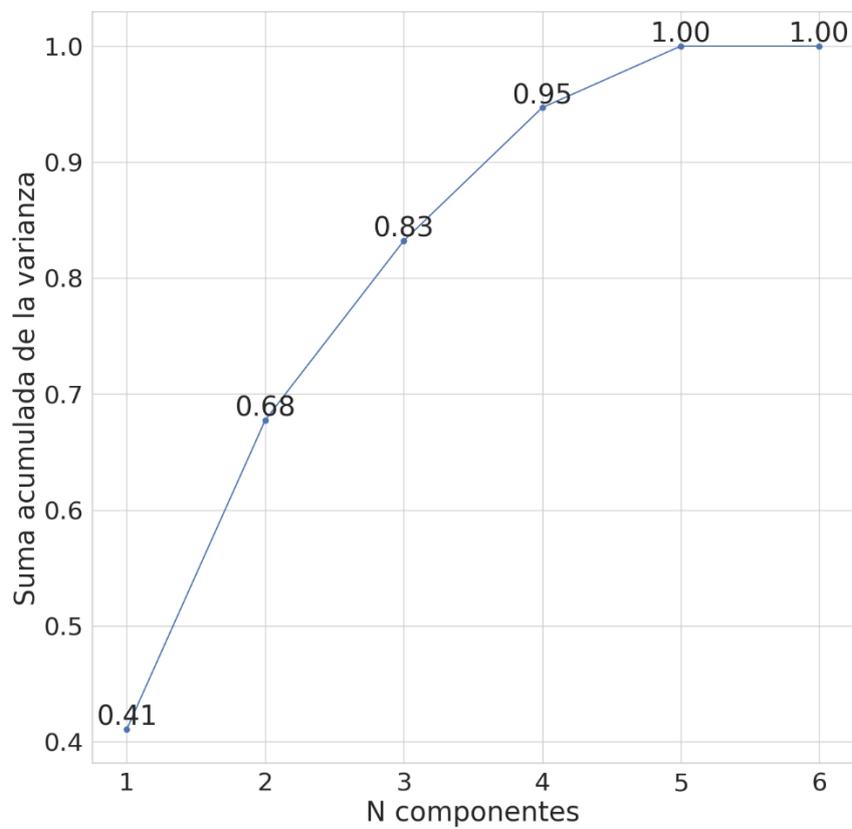


Figura 8. Grafica de la suma de la varianza explicada acumulada vs. El número de componentes  
Autoría propia

Para este análisis se recurrió a la varianza explicada acumulada que nos muestra que el primer componente se encuentra con un 0,41, es decir, 41% de la varianza observada en los datos y la segunda con un 0,27, es decir, 27%, el tercer componente tiene un 0,15, es decir, 15%, el cuarto componente tiene 0,05, es decir, 5%, el sexto componente tiene un valor muy pequeño que se puede considerar como 0.

Para seleccionar el número de componentes es recomendado, que el porcentaje de la varianza acumulada sea superior al 80%, esto se cumpliría con 3 componentes, pero en este caso al estar tratando con datos médicos se consideró adecuado un valor superior al 90% por lo que se seleccionó 4 componentes de los cuales sus vectores y valores propios se muestran en la Figura 9.

```
[181] pca_v2.explained_variance_#contienes los valores propios
↳ array([2.47217286, 1.60398331, 0.93002582, 0.69081183])

[179] pca_v2.components_#contiene los vectores propios
↳ array([[ 0.41719706, -0.03089217, -0.51497192,  0.51497192, -0.36667554,
          -0.40018644],
         [ 0.44658012, -0.39783156,  0.43761214, -0.43761214, -0.15996153,
          -0.48342506],
         [ 0.10388675,  0.83379126,  0.20511133, -0.20511133, -0.45287771,
          -0.06899372],
         [ 0.35170211,  0.37921807, -0.03234371,  0.03234371,  0.79668564,
          -0.3093526 ]])
```

Figura 9. Vectores y valores propios  
Autoría propia

Luego, se obtienen nuevas variables (componentes principales) que contienen la mayor parte de la variabilidad de los datos originales, cada componente principal de los datos originales tiene su porcentaje de variabilidad y un gráfico de dispersión que muestra los datos proyectados en los dos primeros componentes principales facilitando la identificación de patrones y agrupaciones.

Una vez obtenidos estos nuevos datos se crea un nuevo DataFrame que los contendrá, cabe destacar que el PCA solo reduce dimensionalidad a nivel de variables (columnas) las filas mantienen su misma dimensión, esto se puede entender de mejor manera en el anexo (3) tabla (1).

### 3.6.2 Implementación de K-means

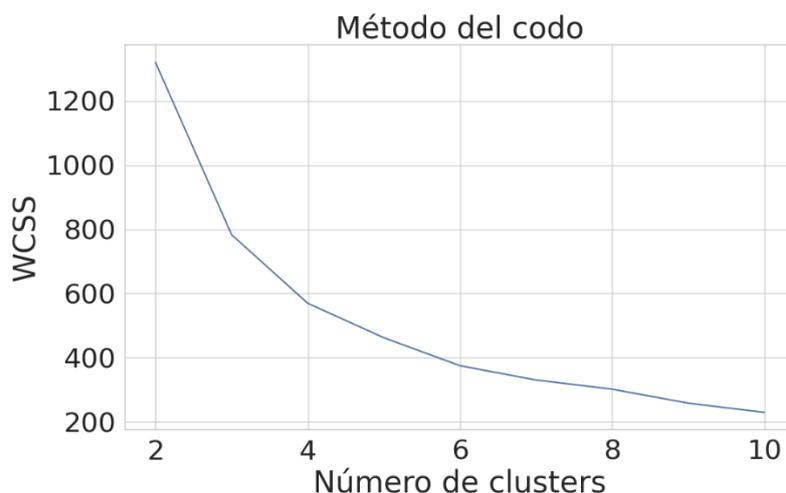
“La agrupación en clusters K-means es uno de los algoritmos de aprendizaje automático no supervisado más simple y popular, este algoritmo hace inferencias a partir de un conjunto de datos utilizando únicamente vectores de entrada sin hacer referencia a resultados conocidos, es decir, el objetivo de K-Means es agrupar puntos de datos similares y descubrir patrones subyacentes. Para lograr este objetivo, K-Means busca un número fijo (k) de grupos (es una colección de puntos de datos agregados debido a ciertas similitudes) en un conjunto de datos” (Education Ecosystem (LEDU), 2018).

“Una vez definido el número objetivo k, este es el número de centroides (ubicación imaginaria o real que representa el centro del grupo) que se necesita en el conjunto de datos, es decir, el algoritmo K-Means identifica el número de centroides y luego asigna cada punto de datos al grupo más cercano manteniendo los centroides lo más pequeños posible, recalcula los centroides de los clusters y repite estos pasos hasta que las asignaciones de los clusters no cambien” (David Cournapeau, <https://scikit-learn.org>, 2021).

Se aplicó el K-Means a los datos ya transformados previamente por el algoritmo PCA, para este caso se seleccionó los tres primeros componentes principales para realizar este estudio.

Para identificar el número objetivo k se empleó el método del codo, este varía el número de grupos (k) de 2 a 10, para cada valor de K se calcula la suma de cuadrados dentro del grupo o WCSS (Within-Cluster Sum of Square), también conocido como inercia (Inertia), es la suma de la distancia al cuadrado entre cada punto y el centroide de su cluster asignado, recorre en iteración un rango de posibles valores de k, para cada k, realiza una agrupación de k-medias y calcula el WCSS (Saji, 2024).

Cuando se traza el WCSS con el valor de k el gráfico parece un codo, a medida que aumenta el número de clusters, el valor de WCSS comienza a disminuir. El valor de WCSS es mayor cuando k=2 como se muestra en la Figura 10, el gráfico cambiará rápidamente en un punto, creando una forma de codo hasta que a partir de un punto, la gráfica se mueve casi paralela al eje x (número de clusters). El valor de K correspondiente a este punto es el valor óptimo de k, es decir, el valor óptimo de clusters (Saji, 2024).

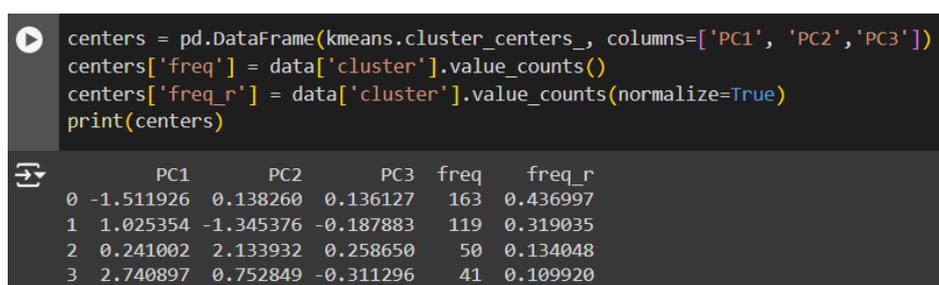


**Figura 10.** Método del codo

Autoría propia

Analizando la gráfica con lo anteriormente expuesto, se puede notar que 4 clústers sería el número óptimo para agrupar los datos basados en la similitud de las características, esto se puede visualizar en el espacio de los componentes principales.

Luego, se realizó el cálculo de la frecuencia que es el número de puntos de datos asignados a cada clúster y la frecuencia relativa que es la proporción de puntos de datos asignados a cada clúster para los centroides dentro del espacio de los componentes principales, como se muestra en la Figura 11.



**Figura 11.** Gráfico de frecuencias y frecuencias relativas para cada cluster

Autoría propia

### 3.7 Evaluación del modelo

La evaluación del modelo se considera como una fase importante para este proyecto ya que determina el rendimiento y la efectividad, para un modelo de aprendizaje no supervisado, la evaluación se realiza utilizando diversas métricas y técnicas de visualización para interpretar los resultados. A continuación, se describe las métricas que se van a emplear.

### 3.7.1 Métricas de evaluación

Las métricas de evaluación para un modelo de clustering como K-Means a diferencia de los modelos supervisados donde existen métricas como la precisión y el error, no se realiza de forma tan directa, debido a que no hay una respuesta única y correcta para la agrupación de datos. Sin embargo, existen métricas que ayudan a comprender la calidad y la estructura de los grupos que se ha creado.

A continuación, se presentan las métricas empleadas para la evaluación.

#### 3.7.1.1 WCSS:

“Es la suma de las distancias cuadráticas entre los puntos de datos y el centroide de su cluster correspondiente, es decir, representa que tan ajustados están los puntos a sus centroides, cuando el valor del WCSS(inercia) es más bajo, esto es un indicador de que los puntos están más cerca de sus centroides, lo que sugiere un buen ajuste, sin embargo, este valor disminuye a medida que se incrementa el número de clusters, por lo que es recomendable utilizarlo en combinación con el método del codo para determinar el número óptimo de clusters” (Strickland, 2015).

Para realizar el cálculo el algoritmo se utiliza una fórmula matemática que se presenta en la ecuación 9

$$WCSS = \sum_{i=1}^k \sum_{x \in S_i} \|x - \mu_i\|^2$$

**Ecuación 9.** Formula del WCSS

Disponible en (López, 2024)

Donde:

k es el número de clusters.

$S_i$  representa el conjunto de puntos en el cluster i.

x es un punto dentro del cluster  $S_i$ .

$\mu_i$  es el centroide del cluster i, es decir, el punto que representa el centro del cluster.

$\|x - \mu_i\|^2$  es la distancia al cuadrado entre el punto x y el centroide  $\mu_i$ , lo que da una idea de cuán lejos está cada punto de su centro.

Para nuestro caso, el valor de inercia es de 569.695 unidades, lo que nos indica que los datos tienen una gran dispersión, es decir existen puntos que están más alejados de sus centroides.

### 3.7.1.2 Coeficiente de Silueta (Silhouette Score):

Es una métrica para evaluar la calidad de la agrupación obtenida con el algoritmo de clustering en este caso K-Means. Su función es identificar cual es el número óptimo de agrupamientos. El coeficiente de Silueta es un valor comprendido entre -1 y 1 donde -1 indica que los datos están mal asignados a su cluster, 0 indica que los puntos de datos están exactamente al límite entre otros clusters y 1 indica que los puntos de datos están bien asignados a su cluster y está lejos de los puntos de otro cluster, para una observación  $i$  se denota como  $s(i)$  y se define con la Ecuación 10

$$s(i) = \frac{b - a}{\max(a, b)}$$

**Ecuación 10.** Ecuación del coeficiente de Silueta

Disponible en: (Banchemo, 2015)

Donde:

$a$  es el promedio de las distancias de la observación  $i$  con las demás observaciones del cluster al que pertenece  $i$

$b$  es la distancia mínima a otro cluster, que no es el mismo en el que está la observación  $i$ , ese cluster es la segunda mejor opción para  $i$  y se designa como vecindad  $i$ .

El valor de  $s(i)$  puede ser obtenido combinando los valores de  $a$  y  $b$  como se muestra en la Ecuación 11:

$$s(i) = \begin{cases} 1 - \frac{a}{b}, & \text{si } a < b \\ 0, & \text{si } a = b \\ \frac{b}{a} - 1, & \text{si } a > b \end{cases}$$

**Ecuación 11.** Fórmula del coeficiente de silueta

Para este proyecto el promedio del coeficiente de silueta fue de 0,36 lo que se puede considerar como un valor aceptable.

### 3.7.2 Distancia entre Centroides:

Esta métrica mide la distancia entre los centroides de los clusters, también se puede definir la distancia entre el centroide del cluster  $i$  y el centroide del cluster  $j$ , es decir la distancia euclidiana entre dos puntos en un espacio  $n$ -dimensional, se puede observar en la ecuación

12

$$d(C_i, C_j) = \sqrt{\sum_{m=1}^k (C_{im} - C_{jm})^2}$$

**Ecuación 12.** Distancia entre centroides

Disponible en (Moix Bergadà, 2013)

Donde

$(C_i$  y  $C_j)$  son los centroides de los clusters  $i$ -ésimo y  $j$ -ésimo.

$(C_{ik}$  y  $C_{jk})$  son coordenadas de los centroides  $(C_i$  y  $C_j)$  en la  $k$ -ésima dimensión.

$m$  varia entre 1 hasta  $k$ .

Luego, la matriz de distancias se puede observar en la tabla 1

Tabla 1. Matriz de distancias entre centroides

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Cluster 1	0	3,121530	3,327233	5,259697
Cluster 2	3,121530	0	3,730976	3,370589
Cluster 3	3,327233	3,730976	0	3,079237
Cluster 4	5,259697	3,370589	3,079237	0

En la tabla se puede observar que existen distancias variables entre los centroides, pero no son consideradas pequeñas, por lo que, se puede decir que existe una mejor separabilidad entre los clusters.

## 4 RESULTADOS, CONCLUSIONES Y RECOMENDACIONES

### 4.1 Resultados

En esta sección, se presenta los resultados obtenidos después de realizar el análisis, evaluación de los modelos de aprendizaje no supervisado aplicado a el conjunto de datos del Alzheimer.cvs.

#### 4.1.1 Resultados del Análisis de Componentes Principales (PCA)

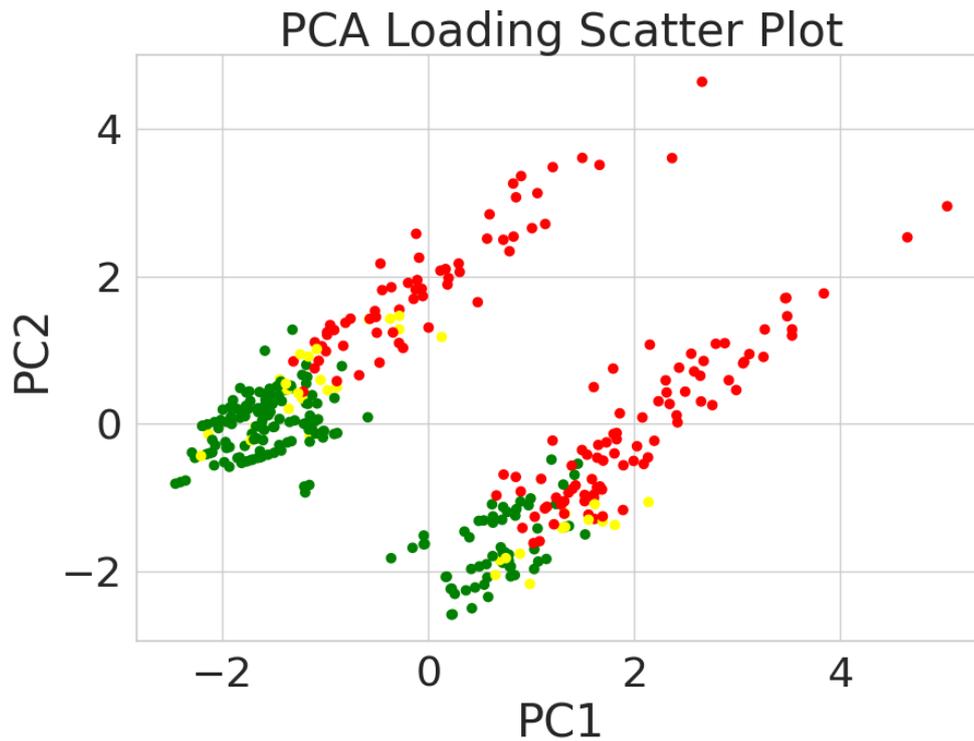
El PCA se empleó para reducir la dimensionalidad de los datos y visualizar las principales fuentes de variación, al terminar de ejecutar el modelo de aprendizaje no supervisado se obtuvo tanto la varianza explicada como la varianza explicada acumulada por los primeros 4 componentes principales como se muestra en la Figura 12

```
Varianza explicada por las primeras componentes principales:  
[0.41092417 0.26661385 0.15458874 0.11482663]  
Varianza explicada acumulada: 0.9469533961357217
```

**Figura 12.** Valores de la varianza explicada y varianza explicada acumulada.

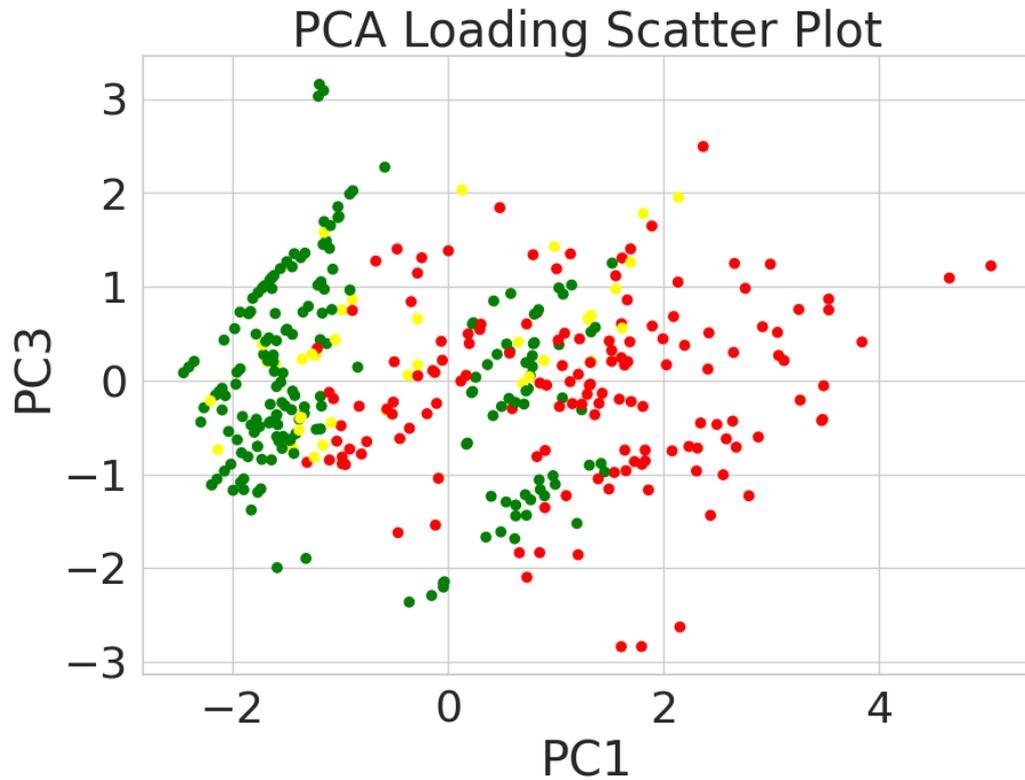
Autoría propia

Luego, se proyecta los datos de los primeros componentes principales obtenidos del PCA y colorea los puntos según sus asignaciones de la variable de trabajo "Group", esto se visualiza de mejor manera en las Figuras 13, 14, 15, 16, 17,18.



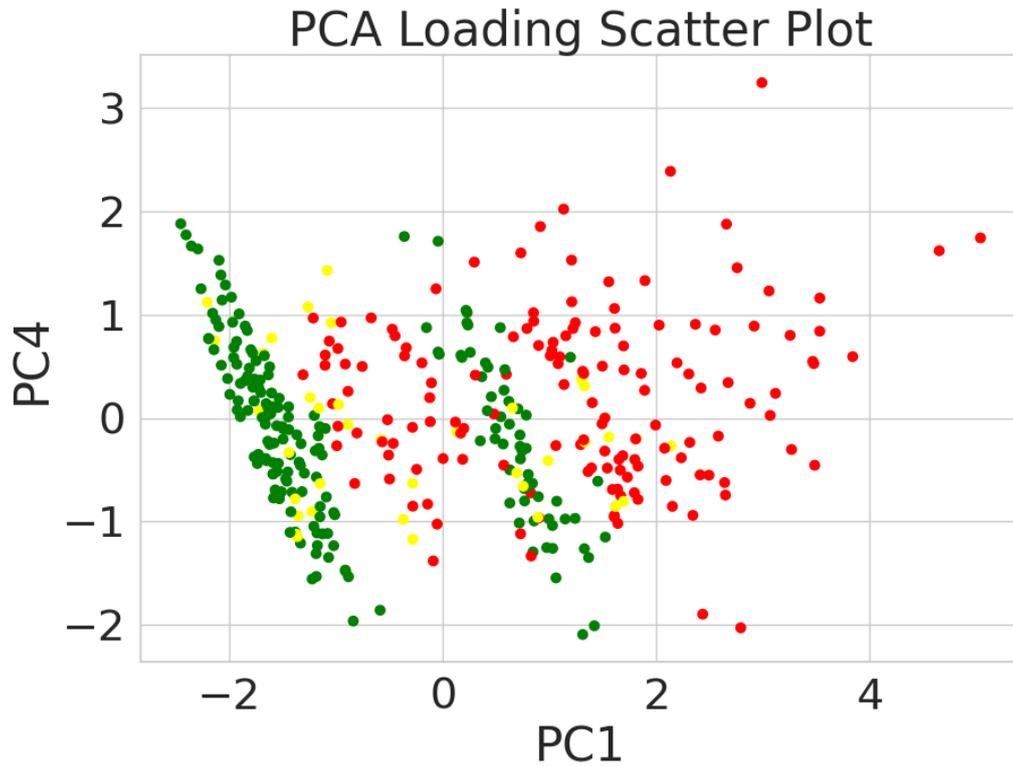
**Figura 13.** Gráfico del componente “PC1” vs “PC2”

El gráfico muestra la relación que existe entre el primer componente principal y el segundo componente principal, se puede observar que los puntos rojos(demente) tienden a agruparse en una región distinta de los grupos verdes (no demente) y los puntos amarillos (“converted”) se encuentran entre los dos grupos principales, esto sugiere que los primeros componentes principales son efectivos para distinguir entre los 3 estados de los individuos.



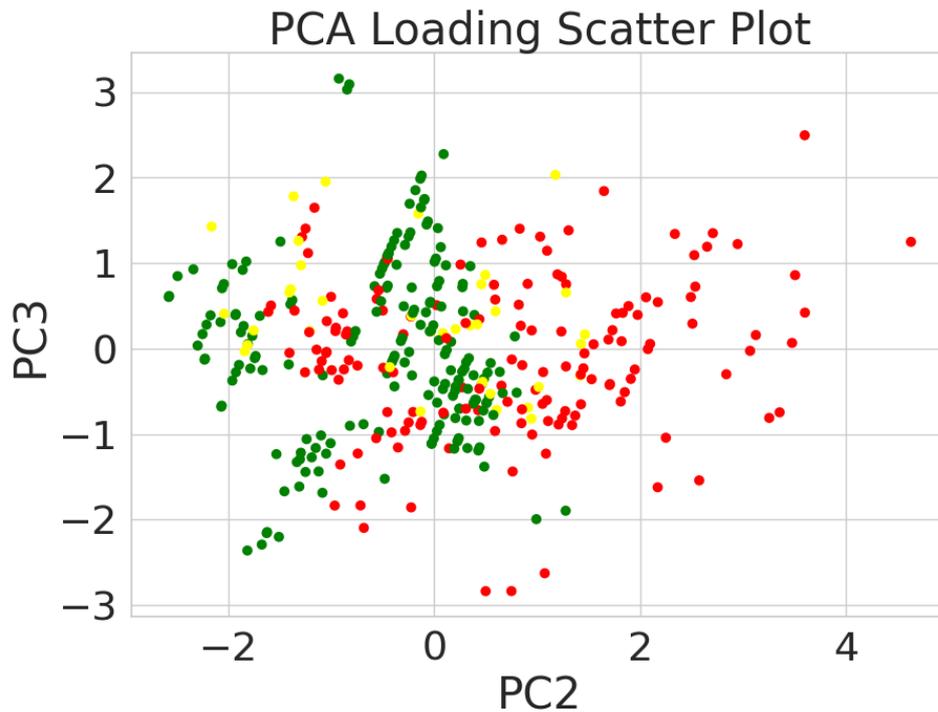
**Figura 14.** Gráfico del componente “PC1” vs “PC3”

El gráfico nos muestra que se compara el primer componente principal con el tercer componente principal, aunque sigue existiendo separación entre los grupos, la dispersión es mayor en comparación a la Figura 13. Los puntos verdes y rojos están más mezclados, los amarillos continúan apareciendo en posiciones intermedias, lo que indica que el tercer componente principal aporta información adicional, pero no tan clara como el segundo componente.



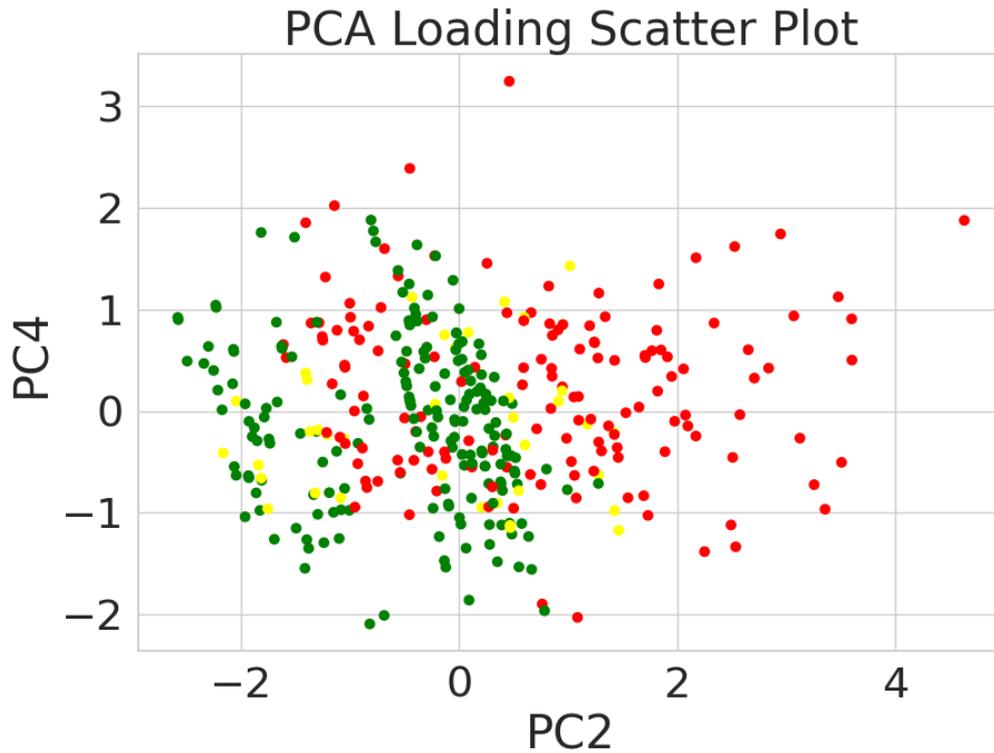
**Figura 15.** Gráfico del componente "PC1" vs "PC4"

En este gráfico se muestra el primer componente principal frente al cuarto componente principal, aquí la separación entre grupos es menos clara que en los anteriores. Los puntos de diferentes colores se mezclan más, esto significa que el cuarto componente principal no es tan efectivo para distinguir entre los estados del individuo.



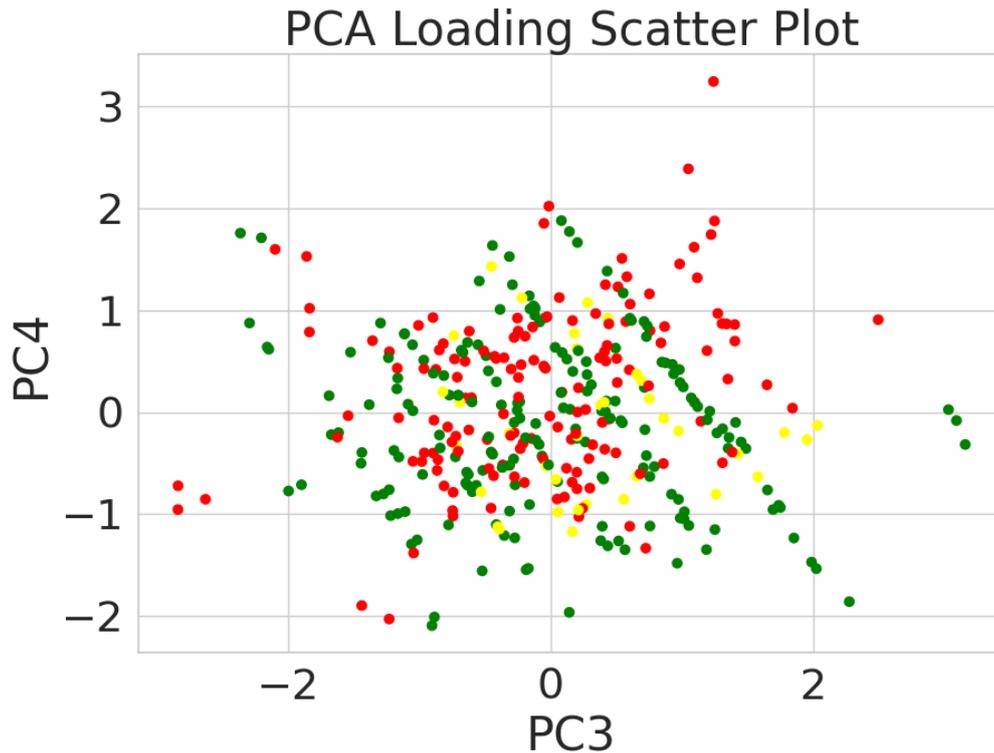
**Figura 16.** Gráfico del componente “PC2” vs “PC3”

Este grafico presenta el segundo componente principal frente al tercer componente principal mostrando que la separación entre los grupos está siendo moderada, se observa que los puntos rojos y verdes están más mezclados en comparación con el “Gráfico de PC1 vs PC2” (Figura 13), los puntos amarillos se encuentran en posiciones intermedias, pero aún se pueden apreciar ciertas agrupaciones.



**Figura 17.** Gráfico del componente "PC2" vs "PC4"

Este gráfico muestra que al comparar el segundo componente principal con el cuarto componente principal existe una separación menos clara, los puntos de los 3 colores están bastante mezclados, esto indica que el cuarto componente principal tiene una menor capacidad discriminativa en comparación con la segunda.



**Figura 18.** Gráfico del componente “PC3” vs “PC4”

Finalmente, el gráfico muestra al tercer componente principal frente al cuarto componente principal. La separación entre los grupos es menos clara que en todos los gráficos anteriores, los puntos están muy mezclados, lo que sugiere que estos componentes principales no son efectivos para discriminar entre los estados de los individuos por si solos.

Después de este análisis se puede afirmar que los dos primeros componentes principales son los más efectivos para mostrar una separabilidad entre los estados de los individuos.

#### 4.1.2 Resultados del Clustering K-Means:

El algoritmo de K-Means se utilizó para agrupar los datos en clusters, por lo que se generaron varios clusters, cada uno representando a un grupo de individuos con características similares.

Al calcular la frecuencia y la frecuencia relativa se obtuvieron cuantos datos están en cada cluster y el porcentaje de total de los datos como se muestra en la tabla 2.

**Tabla 2.** Tabla de frecuencias

Frecuencia	Frecuencia Relativa
163	43.70%
119	31.90%
50	13.40%
41	10.99%

La distribución de los datos en los clusters y la posición de los centroides van a proporcionar información sobre la estructura de los datos, esto se puede visualizar en las Figuras 19, 20, 21, 22, 23 que continúan en la siguiente hoja.

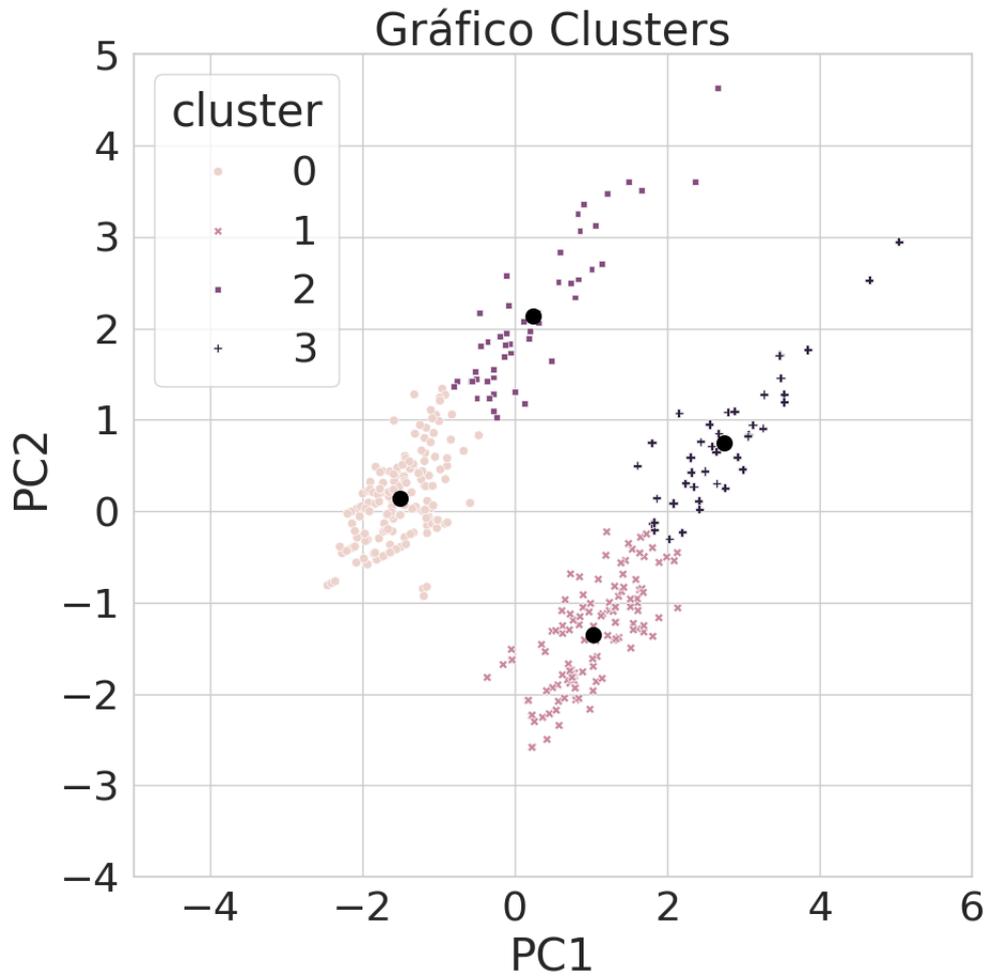


Figura 19. Gráfico cluster de los componentes principales “PC1” vs “PC2”

El gráfico muestra los puntos que se agrupan en cuatro clusters distintos, representados por diferentes formas y colores, los centroides de los clusters están marcados y son claramente visibles, la separación entre los clusters sugiere que el K-Means ha sido efectivo al identificar agrupaciones naturales en los datos.

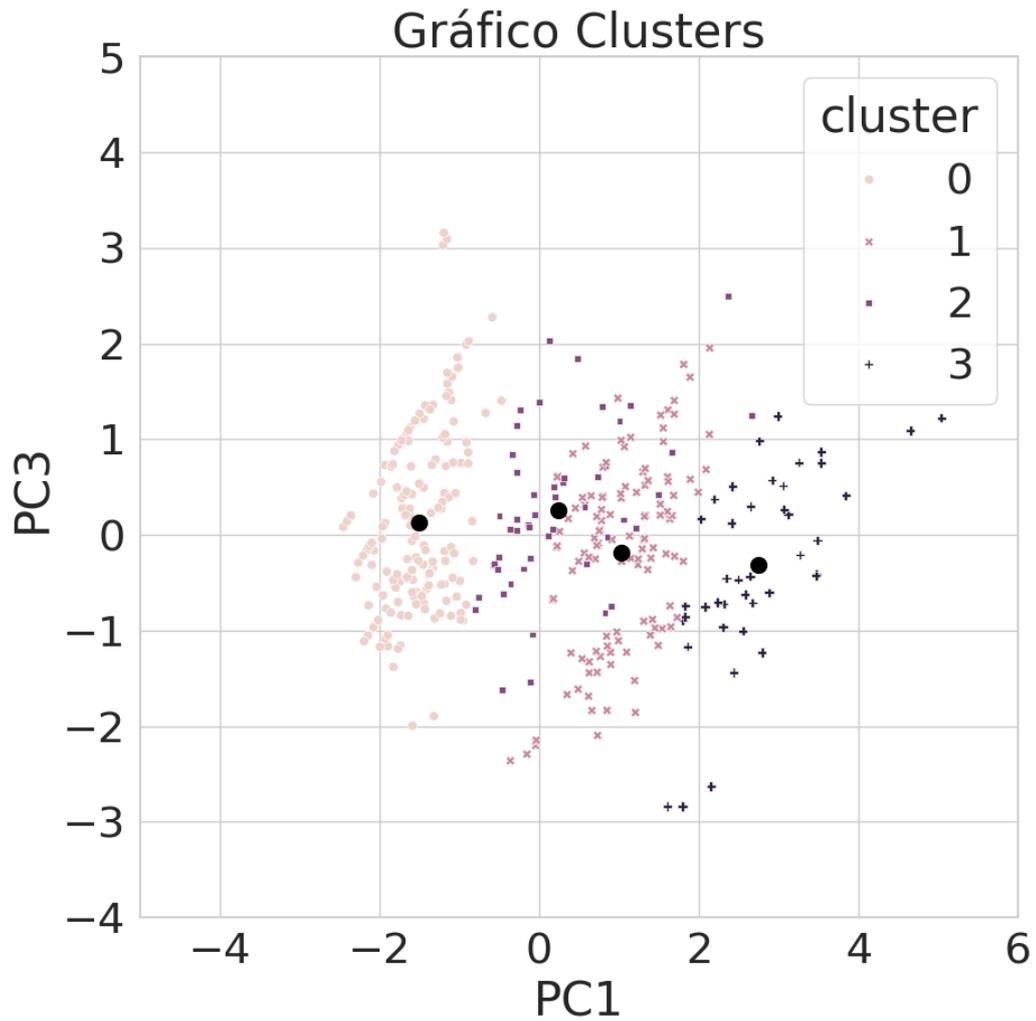


Figura 19. Gráfico cluster de los componentes principales "PC1" vs "PC3"

En el gráfico se puede observar una mayor mezcla entre los clusters en comparación al gráfico anterior, pero aun así se tiene una diferenciación visible entre los grupos, los centroides son visibles y las distancias entre ellos indican que los clusters están separados.

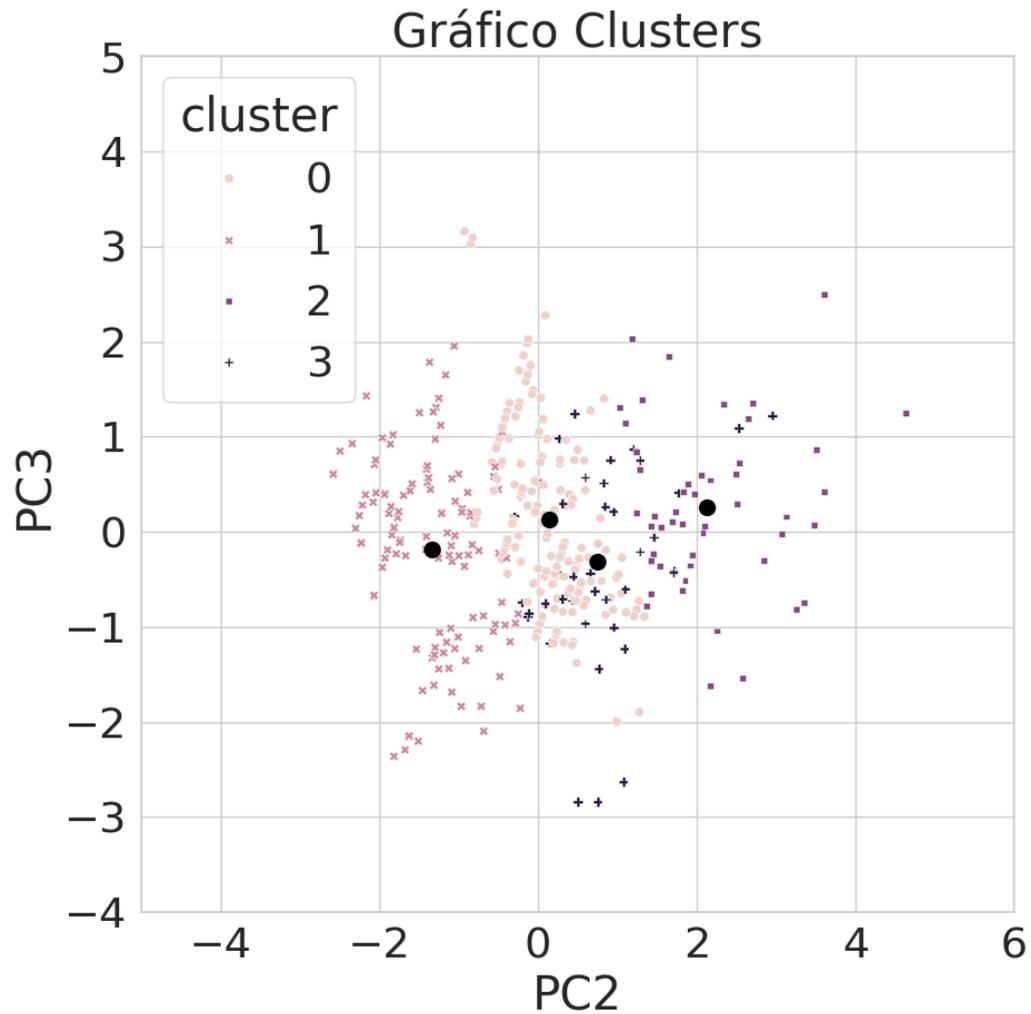


Figura 19. Gráfico cluster de los componentes principales “PC2” vs “PC3”

Este gráfico muestra que la separación entre clusters es moderada, cierta superposición entre los puntos de los clusters y los centroides es visible y las distancias entre ellos proporcionan información sobre la estructura de los datos en este espacio reducido.

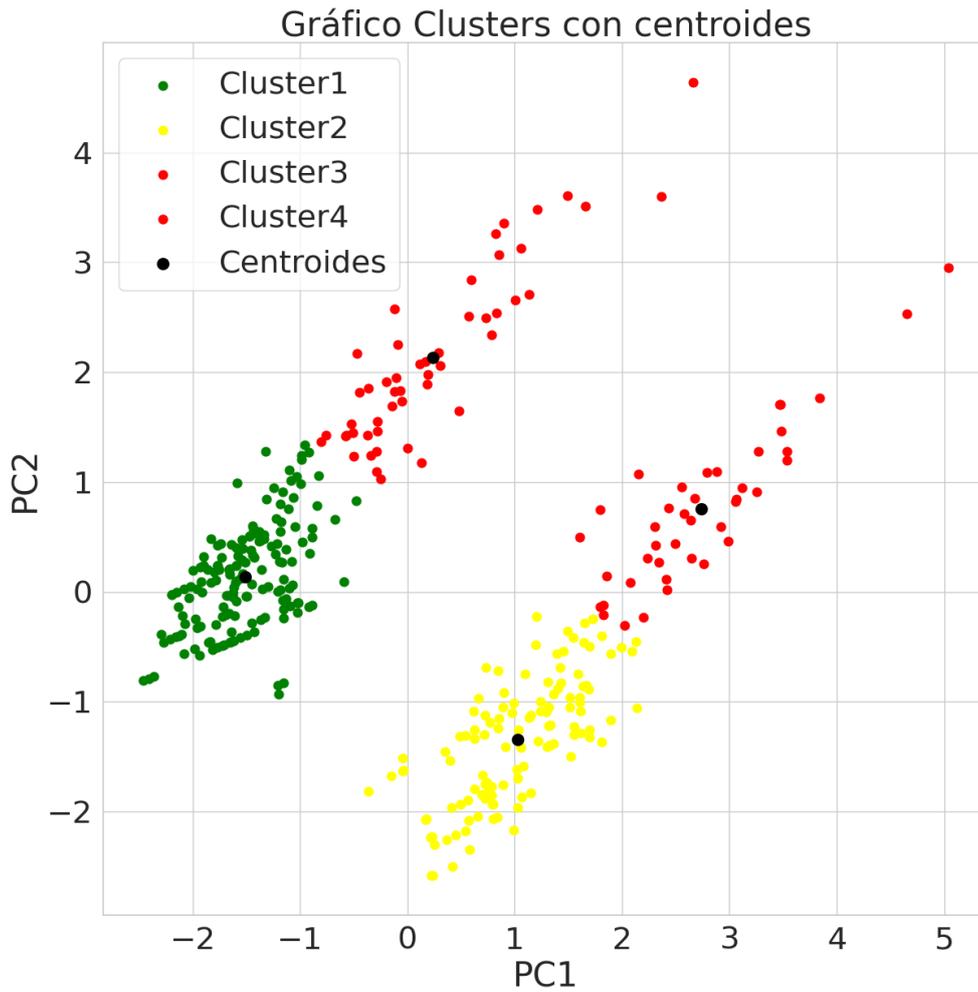


Figura 19. Gráfico cluster con centroides de los componentes principales “PC1” vs “PC2”

Este gráfico se representa como el más aceptado para la comparación, muestra una clara separación entre los clusters, con los centroides bien definidos y las distancias entre ellos bien marcadas.

### 4.1.3 Comparación de Gráficos PCA y K-Means

**Tabla 3.** Comparación de Gráficos PCA y K-Means

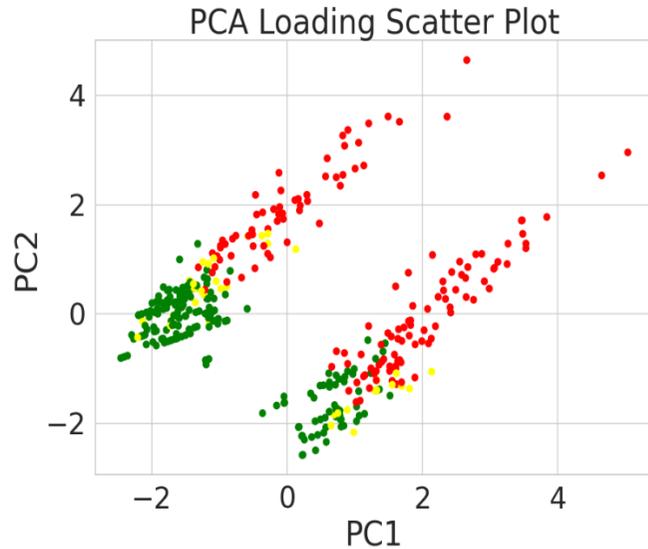


Figura 20. Gráfico de los 2 primeros componentes principales

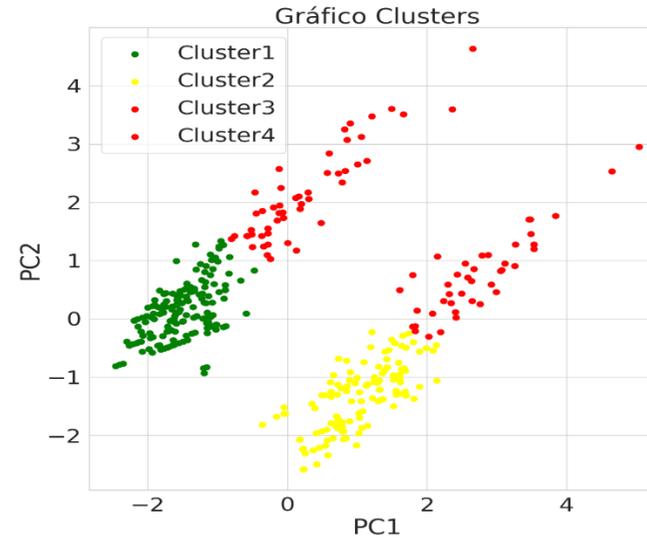


Figura 21 Gráfico de cluster de los componentes principales

El gráfico del PCA (Figura 20) seleccionado para la comparación muestra los dos primeros componentes principales que mejor separan los estados de los individuos, observando una separación clara entre los puntos rojos, verdes y un poco entre amarillos. En el gráfico del K-Means (Figura 21) seleccionado por tener más aceptación mostrando una separación clara entre los clusters, existiendo una separación entre los mismos que indica que el algoritmo ha sido capaz de identificar agrupaciones naturales en los datos.

Al comparar los dos gráficos, notamos que el PCA ha sido efectivo a la hora de reducir la dimensionalidad y capturar la variabilidad en los datos, permitiendo una clara visualización de los diferentes estados, por otro lado, el clustering K-Means ha sido efectivo agrupando los datos en clusters distintos permitiendo identificar patrones naturales en los datos.

La combinación de PCA y K-Means resulto ser un acierto ya que el PCA al reducir la dimensionalidad facilita la visualización, mientras que K-Means identifica y agrupa patrones similares.

En el gráfico de PCA, se observa que existen dos grupos de individuos del mismo color, además de algunos individuos en transición que se encuentran mezclados. Sin embargo, al aplicar el clustering K-Means, esta mezcla se resuelve, permitiendo identificar claramente los grupos en los que se encuentran los individuos. Los puntos verdes representan a los individuos sin demencia, los puntos amarillos representan a los individuos en transición (de no demente a demente), y los puntos rojos representan a los individuos con demencia.

De manera visual, se nota que los individuos de color rojo (demente) se distribuyen en diferentes clusters. Al mirar el cluster que contiene a los individuos de color verde (sin demencia), junto con el cluster que contiene a algunos individuos de color rojo (demente), se sugiere una transición acelerada de un estado a otro. En contraste, otro grupo de clusters, que contiene puntos amarillos y rojos, sugiere una transición más lenta de no demente a demente. Estadísticamente, se muestra claramente que el algoritmo K-Means está tendiendo a entremezclar a los pacientes no dementes con los que están en transición, esto debido a que sus datos están solapados.

## **4.2 Conclusiones**

Aplicar técnicas de preprocesamiento de datos como el manejo de valores faltantes y la estandarización ha sido de gran ayuda para asegurar la coherencia y calidad de los datos, permitiendo que los modelos de aprendizaje no supervisado operaran sobre un conjunto limpio y homogéneo de datos, mejorando la precisión y fiabilidad de los resultados obtenidos.

El desarrollo de un modelo de aprendizaje no supervisado utilizando específicamente los modelos PCA y K-Means, permitió identificar patrones significativos en los datos. Estos patrones distinguen claramente entre individuos no dementes, aquellos en transición y aquellos con demencia, proporcionando así una herramienta valiosa para la identificación temprana y el tratamiento de la enfermedad.

La selección y pruebas de algoritmos de aprendizaje no supervisado, como PCA y K-Means demostró ser adecuado para el análisis de los datos de pacientes con Alzheimer. La implementación de estos algoritmos y la optimización de sus parámetros han llevado a resultados convenientes y útiles, destacando la importancia de una adecuada elección y ajuste de modelos en análisis de datos.

La utilización de herramientas de visualización resultó ser fundamental para la interpretación de los datos del clustering a través de los gráficos generados permitieron una comprensión más clara de los patrones identificados, facilitando de esta manera la comunicación de hallazgos a los profesionales de salud y apoyando la toma de decisiones basadas en datos.

La fase de pruebas y evaluación de los modelos implementados fue relevante para garantizar la robustez y precisión de los resultados, a través de distintas métricas, se confirmó que los modelos desarrollados pueden identificar características y agrupar efectivamente a los pacientes según la progresión de su estado, que puede ser demente o no demente, proporcionando una base para su aplicación en entornos clínicos.

Se analizaron los resultados obtenidos mostrando patrones relevantes que pueden ser de gran utilidad para los profesionales de la salud en el manejo del Alzheimer. Las conclusiones obtenidas indican que el modelo de aprendizaje no supervisado desarrollado es efectivo en la identificación de los estados de salud, sino que, también puede ofrecer una percepción para el desarrollo de estrategia de tratamiento y gestión de la enfermedad.

### **4.3 Recomendaciones**

Es fundamental seguir optimizando las técnicas de preprocesamiento de datos, como el manejo de valores faltantes y la estandarización por lo que se recomienda explorar métodos adicionales de asignación de valores a los datos faltantes y técnicas avanzadas de escalado para asegurar la máxima coherencia y calidad de los datos utilizados en los modelos.

Aunque el PCA y K-Means han mostrado buenos resultados, es recomendable investigar y probar otros algoritmos de aprendizaje no supervisado, como el Análisis de Componentes Independientes (ICA) y Clustering jerárquico, ofreciendo perspectivas adicionales y la posibilidad de mejorar la identificación en los datos.

Ampliar el conjunto de datos con variables clínicas adicionales relevantes para la patología. Puede mejorar los resultados del modelo por lo que se recomienda una revisión continua

de la literatura para identificar nuevas variables que podrían ser relevantes para el análisis y el modelado de los datos.

## 5 REFERENCIAS BIBLIOGRÁFICAS

- Alberca, A. S. (2020). <https://aprendeconalf.es>. Obtenido de <https://aprendeconalf.es>: <https://aprendeconalf.es/docencia/python/manual/pandas/>
- Alberca, D. A. (2020). <https://aprendeconalf.es>. Obtenido de <https://aprendeconalf.es>: <https://aprendeconalf.es/docencia/python/manual/matplotlib/>
- Alberca, D. A. (2021). <https://aprendeconalf.es>. Obtenido de <https://aprendeconalf.es>: <https://aprendeconalf.es/docencia/python/manual/numpy/>
- Bancho, S. (Octubre de 2015). <https://www.labredes.unlu.edu.ar>. Obtenido de <https://www.labredes.unlu.edu.ar>: <https://www.labredes.unlu.edu.ar/sites/www.labredes.unlu.edu.ar/files/site/data/bdm/coeficiente-silueta.pdf>
- Burrueco, D. (2022). <https://interactivechaos.com>. Obtenido de <https://interactivechaos.com>: <https://interactivechaos.com/es/python/function/labelencoder>
- Burrueco, D. (2022). <https://interactivechaos.com>. Obtenido de <https://interactivechaos.com>: <https://interactivechaos.com/es/manual/tutorial-de-machine-learning/la-funcion-getdummies>
- Burrueco, D. (2022). <https://interactivechaos.com>. Obtenido de <https://interactivechaos.com>: <https://interactivechaos.com/es/manual/tutorial-de-machine-learning/escalado-de-datos>
- Burrueco, D. (2022). <https://interactivechaos.com>. Obtenido de <https://interactivechaos.com>: <https://interactivechaos.com/es/wiki/analisis-de-componentes-principales-pca>
- Burrueco, i. D. (2022). <https://interactivechaos.com>. Obtenido de <https://interactivechaos.com>: <https://interactivechaos.com/es/manual/tutorial-de-pandas/la-funcion-isnull>
- Burrueco, I. D. (2022). <https://interactivechaos.com>. Obtenido de <https://interactivechaos.com>: <https://interactivechaos.com/es/manual/tutorial-de-pandas/el-metodo-fillna>
- Burrueco, I. D. (2022). <https://interactivechaos.com>. Obtenido de <https://interactivechaos.com>: <https://interactivechaos.com/es/manual/tutorial-de-machine-learning/standardscaler>
- Carvajal, L. (2006). *Metodología de la Investigación Científica. Curso general y aplicado* (28 ed.). Santiago de Cali: U.S.C.
- Casanova, J. P. (1999). *Enfermedad de Alzheimer. Del diagnóstico a la terapia: conceptos y hechos*. Barcelona: Fundación "la Caixa".
- Congacha Griogia, V. C. (26 de enero de 2020). <https://rpubs.com>. Obtenido de <https://rpubs.com>: <https://rpubs.com/Giorgia96/kmeans>
- Data Science. (26 de enero de 2023). <https://datascientest.com>. Obtenido de <https://datascientest.com>: <https://datascientest.com/es/seaborn-la-herramienta-de-data-visualization-python>
- Datacamp. (marzo de 2024). <https://www.datacamp.com>. Obtenido de <https://www.datacamp.com>: <https://www.datacamp.com/es/tutorial/pca-analysis-r>

- David Cournapeau, M. B. (2021). <https://scikit-learn.org>. Obtenido de [https://scikit-learn.org:https://scikit-learn.org/stable/modules/generated/sklearn.model\\_selection.train\\_test\\_split.html](https://scikit-learn.org:https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html)
- David Cournapeau, M. B. (2021). <https://scikit-learn.org>. Obtenido de <https://scikit-learn.org:https://scikit-learn.org/stable/modules/preprocessing.html>
- David Cournapeau, M. B. (2021). <https://scikit-learn.org>. Obtenido de <https://scikit-learn.org:https://scikit-learn.org/stable/api/sklearn.decomposition.html>
- David Cournapeau, M. B. (2021). <https://scikit-learn.org>. Obtenido de <https://scikit-learn.org:https://scikit-learn.org/stable/modules/clustering.html>
- David Cournapeau, M. B. (2021). <https://scikit-learn.org>. Obtenido de <https://scikit-learn.org:https://scikit-learn.org/stable/modules/clustering.html#k-means>
- David Cournapeau, M. B. (2021). <https://scikit-learn.org/stable/about.html>. Obtenido de <https://scikit-learn.org:https://scikit-learn.org/stable/about.html:https://scikit-learn.org/stable/api/sklearn.metrics.html#module-sklearn.metrics>
- Dincer, B. (2021). <https://www.kaggle.com>. Obtenido de <https://www.kaggle.com:https://www.kaggle.com/datasets/brsdincer/alzheimer-features/data>
- DINCER, B. (2021). <https://www.kaggle.com>. Obtenido de <https://www.kaggle.com:https://www.kaggle.com/datasets/brsdincer/alzheimer-features/data>
- Education Ecosystem (LEDU). (12 de septiembre de 2018). <https://towardsdatascience.com>. Obtenido de <https://towardsdatascience.com:https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1>
- GeeksforGeeks. (28 de febrero de 2024). <https://www.geeksforgeeks.org>. Obtenido de <https://www.geeksforgeeks.org:https://www.codificandobits.com/blog/manejo-datos-faltantes/>
- Google. (2024). <https://colab.research.google.com>. Obtenido de <https://colab.research.google.com:https://colab.research.google.com/?hl=es#scrollTo=OwuxHmxllTwN>
- Google. (s.f.). <https://www.google.com>. Obtenido de <https://www.google.com:intl/es-419/about/datacenters/>
- Hartian, J. A. (1975). *Clustering algortihms*. New York, Wiley: Printdisabled.
- Hartigan, J. A. (1975). *Clustering algorithms*. New York: Wiley: printdisabled.
- Hotelling, H. (1933). *Analysis of complex of statistical variables into principal components*. Journal of Educational Psychology,.
- Jolliffe, I. (1986). *Principal Component Analysis, Second Edition*. New York, Berlin: Springer.
- Jurjo, C. P. (2022). *El Método K medias*. Valladolid, España: Universidad de Valladolid.

López, O. G. (2 de febrero de 2024). <https://medium.com/about>. Obtenido de <https://medium.com/about>: <https://medium.com/@oriolgilabertlopez/entendiendo-el-within-cluster-sum-of-squares-wcss-14935cb64672>

Mangas, A. S. (25 de abril de 2012). Análisis de componentes principales. *PROYECTO FIN DE CARRERA*, pág. 130.

Mathworks. (2024). <https://la.mathworks.com/>. Obtenido de <https://la.mathworks.com/>: <https://la.mathworks.com/help/stats/principal-component-analysis-pca.html>

Matplotlib. (Marzo de 2021). <https://matplotlib.org>. Obtenido de <https://matplotlib.org>: <https://matplotlib.org/stable/tutorials/pyplot.html>

Moix Bergadà, M. (2013). *Cálculo de la matriz de distancias entre centroides con TRANSCAD*. Obtenido de <https://upcommons.upc.edu>: <https://upcommons.upc.edu/handle/2117/22751>

pandas.pydata.org. (2024). <https://pandas.pydata.org>. Obtenido de <https://pandas.pydata.org>: <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.sum.html>

PREISENDORFER, R. W. (1988). National Oceanic and Atmospheric Administration, Pacific Marine Environmental Laboratory, Seattle, Washington, U.S.A. En R. W. PREISENDORFER, *Principal Component Analysis in Meteorology and Oceanography* (pág. 419). Amsterdam - Oxford - New York - Tokyo: ELSEVIER.

Profesores del departamento de matemática aplicada de la Escuela Técnica Superior de Ingeniería. (2011). *ÁLGEBRA LINEAL, notas de clase*. Sevilla, España.

Python . (2024 de junio de 2024). <https://packaging.python.org>. Obtenido de <https://packaging.python.org>: <https://packaging.python.org/en/latest/>

Python. (2024). <https://wiki.python.org>. Obtenido de <https://wiki.python.org>: <https://wiki.python.org/moin/BeginnersGuide>

Rodrigo, J. A. (2022). <https://cienciadedatos.net>. Obtenido de <https://cienciadedatos.net>: [https://cienciadedatos.net/documentos/35\\_principal\\_component\\_analysis](https://cienciadedatos.net/documentos/35_principal_component_analysis)

Saji, B. (01 de julio de 2024). <https://www.analyticsvidhya.com/about>. Obtenido de <https://www.analyticsvidhya.com/about>: <https://www.analyticsvidhya.com/blog/2021/01/in-depth-intuition-of-k-means-clustering-algorithm-in-machine-learning/#:~:text=K%20Means%20Clustering%20Using%20the%20Elbow%20Method,-ln%20the%20Elbow&text=For%20each%20value%20of%20K,plot%20looks%20like%20an>  
%

Sanz, F. (2015). <https://www.themachinelearners.com>. Obtenido de <https://www.themachinelearners.com>: <https://www.themachinelearners.com/k-means/>

scikit-learn. (s.f.). <https://scikit-learn.org/stable/about.html>. Obtenido de <https://scikit-learn.org/stable/about.html>: <https://scikit-learn.org/stable/api/sklearn.metrics.html#module-sklearn.metrics>

Strickland, P. J. (26 de mayo de 2015). <https://bicorner.com>. Obtenido de <https://bicorner.com>:  
<https://bicorner.com/2015/05/26/k-means-clustering-using-r/>

Torres, J. (2020). *CodeToDevs*. Obtenido de CodeToDevs:  
<https://www.codetodevs.com/correlacion-variables-pandas-corr/>

Universidad de Murcia. (2023). <https://www.um.es>. Obtenido de <https://www.um.es>:  
[https://www.um.es/web/universidad/buscador?q=tec\\_resi&ie=UTF-8&cx=009167550808675375826%3Aqh8wfgo2pbk](https://www.um.es/web/universidad/buscador?q=tec_resi&ie=UTF-8&cx=009167550808675375826%3Aqh8wfgo2pbk)

Urrego, N. (11 de julio de 2023). <https://medium.com>. Obtenido de <https://medium.com>:  
<https://nicolasurrego.medium.com/transformando-datos-en-oro-c%C3%B3mo-la-estandarizaci%C3%B3n-y-normalizaci%C3%B3n-mejoran-tus-resultados-fbe0840d2b94#:~:text=La%20estandarizaci%C3%B3n%20Z%2Dscore%20es,unidades%20de%20medida%20y%20escalas.>

## **6 ANEXOS**

ANEXO I. Datos médicos relevantes.

ANEXO II. Demostración de Formulas

ANEXO III Link al repositorio que contiene el código

ANEXO IV. Datos extensos obtenidos del PCA

## ANEXO I

### 1. Criterios diagnósticos de demencia

Tabla 1. Criterios diagnósticos de demencia

Desarrollo de múltiples trastornos cognitivos	Alteración significativa del funcionamiento familiar, social o laboral.	Inicio gradual y disminución cognitiva continuada respecto al nivel previo de funcionamiento.
Alteración de la memoria, incapacidad para fijar nuevos recuerdos y/o para evocar recuerdo antiguos	Alteración de uno o más de los siguientes aspectos	
	<ul style="list-style-type: none"><li>• Afasia: alteración del lenguaje (comprender, expresarse).</li><li>• Apraxia: alteración de la gestualidad (manipular objetos).</li><li>• Agnosia: fallos en el reconocimiento de objetos.</li><li>• Alteración de la función ejecutiva: * planificar, secuenciar, abstraer</li></ul>	

Criterios diagnósticos de demencia simplificados y adaptados a partir del Manual diagnóstico y estadístico de la Asociación Americana de Psiquiatría (DSM-IV. American Psychiatric Association, 1994) (Casanova, 1999).

## 2. Definición operativa de la enfermedad de Alzheimer (JPC)

Tabla 2. Definición operativa de la enfermedad de Alzheimer (JPC)

Enfermedad neurodegenerativa progresiva	Clínica característica:	Neuropatología característica:	Variabilidad
	<p>Alteración cognitiva: alteración progresiva de memoria y de capacidades intelectivas en general</p> <p>Alteración funcional: pérdida progresiva de autonomía personal</p>	<p>Atrofia cerebral</p> <p>Pérdida de neuronas y espinas neuronales</p> <p>Placas seniles</p> <p>Ovillos neurofibrilares</p> <p>Otras alteraciones: degeneración granulo vacuolar, angiopatía amiloide</p>	<p>Edad de inicio variable: presenil o senil</p> <p>Evolución (lenta, rápida)</p> <p>Perfil neuropsicológico (con predominios variables de los trastornos en la memoria, el lenguaje, el reconocimiento, la escritura...)</p> <p>Alteraciones psicológicas y del comportamiento (variables en síntomas como la ansiedad, la depresión, las alucinaciones, la agresividad...)</p> <p>En los hallazgos neuropatológicos (de autopsia)</p>

Disponible en (Casanova, 1999)

**Orientación**

1. ¿En qué año estamos?  
¿En qué estación estamos?  
¿En qué mes estamos?  
¿Qué día del mes es?  
¿Qué día de la semana es? (5 puntos)
2. ¿En qué país estamos?  
¿En qué región estamos?  
¿En qué ciudad estamos?  
¿Cómo se llama este hospital?  
¿En qué planta nos encontramos? (5 puntos)

**Memoria inmediata**

3. Repetir tres nombres («árbol», «puente», «farol») Presentarlos de nuevo hasta que aprenda los tres nombres y anotar el número de ensayos. (3 puntos)

**Atención y cálculo**

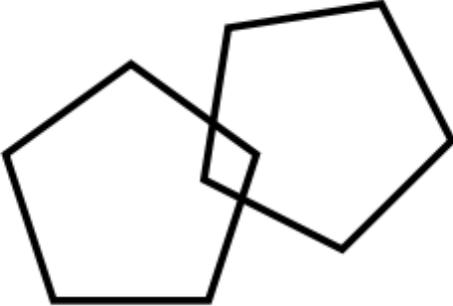
4. Restar 7 partiendo de 100, cinco veces consecutivas  
Alternativa: deletrear «mundo» a la inversa (5 puntos)

**Recuerdo diferido**

5. Repetir los tres nombres aprendidos antes (3 puntos)

**Lenguaje y construcción**

6. Nombrar un lápiz y un reloj mostrados (2 puntos)
7. Repetir la frase: «Ni síes, ni noes, ni peros» (1 punto)
8. Realizar correctamente las tres órdenes siguientes:  
«Coja este papel con la mano derecha, dóblelo por la mitad y póngalo en el suelo» (3 puntos)
9. Leer y ejecutar la frase  
«Cierre los ojos» (1 punto)
10. Escribir una frase con sujeto y predicado (1 punto)
11. Copiar el dibujo de dos pentágonos (1 punto)



**Figura 1** test introducido por Folstein y colaboradores

Para la evaluación rápida y sencilla de los trastornos cognitivos. Cuanto menor es la puntuación (más errores), peor es el estado de las capacidades mentales superiores (Casanova, 1999).

## ANEXO II

**Tabla 1.** Demostración simple de Igualdad

Sea  $X_{est}, Y_{est}$ ; variables estandarizadas. Sea  $n$ ; número de variables, tenemos que:

$$Cov(X_{est}, Y_{est}) = \frac{1}{(n-1)} \sum_{i=1}^n ((X_{est,i} - \overline{X_{est,i}})(Y_{est,i} - \overline{Y_{est,i}}))$$

Dado que la media de variables estandarizadas es 0  $\overline{X_{est}} = 0$  y  $\overline{Y_{est}} = 0$ .

$$Cov(X_{est}, Y_{est}) = \frac{1}{(n-1)} \sum_{i=1}^n (X_{est,i})(Y_{est,i})$$

Luego, la correlación de Pearson entre variables estandarizadas  $X_{est}, Y_{est}$  es

$$Corr(X_{est}, Y_{est}) = \frac{Cov(X_{est}, Y_{est})}{\sigma_{X_{est}} \sigma_{Y_{est}}}$$

Dado que  $\sigma_{X_{est}} = 1$  y  $\sigma_{Y_{est}} = 1$  las variables estandarizadas tienen una desviación estándar igual a 1, la formula se simplifica quedando de la siguiente manera

$$Corr(X_{est}, Y_{est}) = \frac{Cov(X_{est}, Y_{est})}{1} \equiv Cov(X_{est}, Y_{est})$$

Por tanto, al tener variables estandarizadas la matriz de correlación de Pearson y la matriz de covarianza son iguales.

Por lo que se puede decir que

$$Corr(X_{est}, Y_{est}) = \frac{1}{(n-1)} \sum_{i=1}^n (X_{est,i})(Y_{est,i})$$

### ANEXO III

<https://github.com/AndresA-epn/Trabajo-de-integraci-n-curricular/tree/main>

### ANEXO IV

**Tabla 1.** Datos obtenidos del PCA

PC1	PC2	PC3	PC4
1.23958908	1.08441182	0.31127171	0.97006092
1.06062942	1.41505403	0.18451717	1.54528647
1.85963193	0.14411745	-1.1661857	0.43264402
1.54159106	0.41557961	0.97935176	0.48310009
2.31478039	0.42686559	0.71996456	-0.2349171
1.13194487	0.06217831	1.48888518	0.35626444
1.10195379	0.03509571	1.40999821	0.10007487
0.99380991	1.00821121	1.10600566	0.97470391
1.4522757	0.53873373	0.97373569	0.61039422
0.84484133	1.24157941	1.05813808	1.29373989
1.15251463	1.12053705	-0.2456907	0.7968604
2.02690829	0.30123088	0.16843074	0.90017657
2.19499377	-0.2279039	0.37603162	0.53497281
1.18845846	0.28001039	0.43673721	1.31078369
1.12880523	0.39022446	0.39448569	1.11904184
0.73115476	-0.6849405	2.09624313	1.59910829
1.20673585	0.22489773	1.85664318	1.52985662
1.31128492	0.84770244	0.87197336	0.42056512
1.03410932	1.05281025	0.64556492	0.13969034
1.59524643	0.08087144	0.42549126	0.1117092
1.49637262	0.03773792	0.54760943	0.10311655

-1.298625	0.04852912	0.79184576	0.53276803
-	-	-	-
1.33503488	0.48367848	0.35140099	1.20961513
-	-	-	-
1.38447179	0.46211172	0.41246008	1.10220226
-	-	-	-
1.17650785	0.63702607	0.27153434	1.23269902
2.42040194	0.02190969	0.50876491	0.29155199
3.53340849	1.28207214	0.75388029	1.16183556
-	-	-	-
2.43591073	0.76236066	1.43677741	1.89675997
-	-	-	-
2.79251431	1.08630924	1.22819685	2.02885805
-	-	-	-
0.48922455	1.31252611	1.61328589	0.19972992
-	-	-	-
1.19586166	0.48221296	1.52228775	0.5894273
-	-	-	-
0.35047228	-1.457247	1.66872891	0.21961118
-	-	-	-
0.61841837	1.08778462	1.68557711	0.1622712
-	-	-	-
1.35654393	0.20668341	0.22913632	0.94557993
-	-	-	-
1.22767904	0.34709094	0.27236752	-0.9042161
-	-	-	-
0.28374478	1.28092454	0.65644926	0.63064065
-	-	-	-
0.98654383	2.16601862	1.42936055	0.41106101
-	-	-	-
1.81150059	1.36827921	1.78242291	0.20033197
-	-	-	-
2.13844203	1.05727068	1.95436803	0.26798232
-	-	-	-
1.61441084	1.28587833	1.30714175	0.87059411
-	-	-	-
1.55541549	1.22742448	1.11834843	1.32012685
-	-	-	-
2.26791727	0.45851332	0.28938985	1.25188694
-	-	-	-
2.06984071	0.28791232	0.16067593	1.14287276
-	-	-	-
0.72847597	2.49525854	0.60395911	1.11910949
-	-	-	-
0.82734978	2.53839206	0.72607728	1.33393523
-	-	-	-
1.65068119	0.28379878	0.96054419	0.39877111

1.82865405	-	-	-
-	0.20615845	0.74073149	0.78545744
1.89828666	0.32215138	1.16299695	0.33552889
-	0.19468391	-1.1695927	0.22971734
1.99748941	0.22487737	1.08410999	0.07933932
-	0.23781743	-	-
1.92827774	0.23781743	1.04747454	0.0148916
-1.8986156	0.55940028	0.582022	1.33056117
1.89430753	0.25559254	0.98393162	1.45535991
2.75881381	0.53350562	-	-
-1.413805	0.53350562	0.56461794	0.71711724
-	0.39741144	-	-
-1.5327825	0.39741144	0.59563733	0.77996365
-	-	-	-
2.45751179	0.80883537	0.08354659	1.8811259
-	-	-	-
2.40807489	0.78726861	0.14460567	1.77371303
-	0.47073842	-	-
1.36469703	0.47073842	0.38803644	1.14516741
-	0.46642507	-	-
1.37458441	0.46642507	0.40024826	1.12368483
-	1.42613878	0.05710438	-
0.37132586	1.42613878	0.05710438	0.97900483
-	0.65354016	-	-
2.6413929	0.65354016	0.43249703	0.62320475
-	0.27113773	-	-
2.34378466	0.27113773	0.45228428	0.94063941
-	0.11322939	0.97502301	-
1.15031428	0.11322939	0.97502301	0.85500664
-	0.00301532	1.01727452	-
-1.2099675	0.00301532	1.01727452	1.04674849
-	0.01595538	1.05390997	-
1.18030536	0.01595538	1.05390997	1.11119621
-	0.06340225	1.18823996	-
1.07154417	0.06340225	1.18823996	1.34750453
-	0.38447139	-	-
1.56244465	0.38447139	0.63227278	0.71551592
-	0.80138063	-	-
1.18573736	0.80138063	0.51479099	0.56994185
-	0.02599291	-	-
-2.0774556	0.02599291	0.96480676	0.51211296
-	0.05187302	-	-
2.01813131	0.05187302	0.89153586	0.38321751
-1.9192575	0.09500654	-0.7694177	0.16839177

-	-	-	-
1.85960428	0.20522061	0.81166921	0.36013362
-	-	-	-
1.49248309	0.35281241	1.15593326	0.05504992
-	-	-	-
1.39295141	0.56461384	1.04700659	0.48149877
-	-	-	-
1.79932084	0.13476454	0.89288936	0.40037243
-	-	-	-
1.82898298	0.12182449	0.85625391	0.46482015
-	-	-	-
0.67249228	0.66030105	1.2745879	0.96995714
-	-	-	-
0.47441572	0.83090204	1.40330182	0.86094295
-	-	-	-
1.79496525	0.75036116	2.83550644	0.72146271
-	-	-	-
1.60644714	0.49973955	2.83648613	0.95456839
-	-	-	-
2.15156884	1.07430974	2.62692588	0.85356078
-	-	-	-
1.60447594	0.08348313	0.18223461	0.77446636
-	-	-	-
1.68390392	0.03535764	0.2000625	0.62568966
-	-	-	-
1.04713634	0.59609413	0.43662275	0.92395031
-	-	-	-
1.78397381	-0.2970093	0.49594722	0.63041313
-	-	-	-
1.60600095	0.21936897	0.71575992	0.2437268
-	-	-	-
0.95402347	1.34031894	0.89443763	0.92974164
-	-	-	-
0.75627585	1.42658597	-0.6502013	0.50009015
3.1182281	0.94760993	0.21363512	0.23984714
3.84198052	1.76848837	0.41196323	0.59544301
3.06846226	0.84170922	0.26809845	0.02662272
-	-	-	-
1.61708441	0.44220046	1.121551	0.0551071
-	-	-	-
-1.44867	0.28453952	1.21362946	0.01054063
-0.064787	1.83197181	0.4189328	1.25173882
0.29447605	2.17398181	0.54460702	1.51023695
-	-	-	-
-1.3480619	0.02696236	0.73078668	0.42535516
-	-	-	-
1.16943117	0.27327061	0.71955454	-0.1707669
3.0578274	0.82202938	0.51428034	1.23108723
2.91874619	0.59297453	0.57435974	0.89056868

3.25557501	0.90829641	0.75851667	0.80143575
4.65364296	2.52848653	1.09411382	1.62004035
5.04023762	2.94970913	1.22380742	1.74413185
-	-	-	-
2.15915608	0.41106645	0.15505987	1.01557863
-	-	-	-
2.12949393	0.39812639	0.11842442	0.9511309
-	-	-	-
2.09983179	0.38518634	0.08178897	0.88668318
2.30555088	0.59122016	-0.9632212	0.42784007
3.47722384	1.7085948	-0.4137901	0.52795361
-	-	-	-
1.20582119	0.84762664	3.03132725	0.02615141
-	-	-	-
1.15638428	0.82605988	3.09238634	0.08126146
-	-	-	-
1.19626274	0.92764725	3.15906149	0.31596846
-	-	-	-
0.57300812	2.07881104	0.31150331	0.27013706
-	-	-	-
0.80008894	-2.0639379	0.7078975	0.54459944
-	-	-	-
-1.713895	0.13263166	0.27894946	0.36950009
-	-	-	-
1.44660678	0.0681628	0.49314609	0.11010788
-	-	-	-
1.15204857	0.15513897	1.57998396	0.63393659
-	-	-	-
0.12904345	1.17835046	2.03270022	0.12885678
-	-	-	-
0.36195736	1.81765855	2.36102165	1.75830648
-	-	-	-
-0.0449033	1.51096337	2.20128835	1.7121387
-	-	-	-
-1.5437463	0.29339764	0.23393302	0.10386357
-	-	-	-
1.59384107	0.10316297	0.06394727	0.42999815
-	-	-	-
1.73053012	0.31114441	0.83987366	0.35031216
-	-	-	-
1.64121476	0.43429853	0.84548972	0.22301803
-	-	-	-
1.57233203	0.38015804	0.64448459	0.69403335
-	-	-	-
1.36122605	0.93025622	0.36315867	-0.5166183
-	-	-	-
1.80747393	0.39881955	0.27571659	0.20078495
-	-	-	-
1.58896475	0.74671517	0.19780931	0.69008021

0.89771522	-	-	-
0.91446562	0.91446562	1.35511694	0.70316496
1.09579178	-	-	-
0.74386462	0.74386462	1.22640303	0.59415077
0.36504418	-	-	-
2.25372538	2.25372538	0.17057757	0.40063382
0.45403061	-	-	-
2.21490522	2.21490522	0.28048392	0.20729065
0.54301704	-	-	-
2.17608505	2.17608505	0.39039027	0.01394748
1.09272428	-	-	-
0.12925886	0.12925886	1.65325486	0.76283204
1.02384155	-	-	-
0.18339935	0.18339935	1.85425999	1.23384735
1.02351262	-	-	-
-0.0990654	-0.0990654	1.73873757	0.91321005
0.99488873	-	-	-
0.98572969	0.98572969	0.48119523	0.26687725
0.50942026	-	-	-
1.45008581	1.45008581	0.22938347	-0.3576115
0.8242483	-	-	-
3.25754477	3.25754477	0.81000761	0.72272195
0.09035275	-	-	-
2.25231727	2.25231727	1.04193149	1.38138241
0.90101683	-	-	-
3.35832413	3.35832413	0.74492916	0.96454146
1.6089172	-	-	-
-1.0037619	-1.0037619	0.60706264	1.06160285
1.08091081	-	-	-
-1.5877668	-1.5877668	0.50483236	0.52703388
0.62731894	-	-	-
1.33647314	1.33647314	1.32679805	0.82112326
0.62764788	-	-	-
1.25213919	1.25213919	1.44232046	0.50048596
0.76607121	-	-	-
1.19175226	1.19175226	1.27135503	-0.801242
0.8949361	-	-	-
1.05134473	1.05134473	1.22812383	0.75987817
1.18672417	-	-	-
0.54837876	0.54837876	0.16822374	1.53185374
-0.8403369	-	-	-
0.78368003	0.78368003	0.14366742	1.96310654
1.01362523	-	-	-
0.09475204	0.09475204	1.75094939	0.93469263
-1.1523775	-	-	-
0.23947293	0.23947293	1.69550637	0.95457389
0.7818392	-	-	-
1.76971251	1.76971251	0.14994856	0.02910527
0.72218597	-	-	-
1.87992658	1.87992658	0.19220008	0.16263658

0.78151026	- 1.85404647	- 0.26547098	- 0.29153202
0.80095609	- 1.92975372	- 0.40541703	- 0.65513447
0.79106871	- 1.93406707	- 0.39320521	- 0.63365189
0.19569184	- 1.97754686	- 0.39513998	- 0.09919384
0.18547553	- 1.88889955	- 0.49845058	- 0.39834856
- 1.95934523	- 0.02790291	- -0.6313069	- 0.68549439
-1.7513813	- 0.20281726	- 0.49038117	- 0.55499763
-1.8008182	- 0.1812505	- 0.55144025	- 0.6624105
- 1.83080928	- 0.08397648	- 0.47255329	- 0.40622093
- 1.78137238	- 0.10554324	- -0.4114942	- 0.29880805
- 0.46574429	- 2.17274386	- 1.62150294	- -0.2444073
- 0.11869915	- 2.57671305	- -1.5406566	- 0.03438551
- 1.49843584	- 0.39044024	- 1.26809279	- 0.20268379
- 1.33002143	- -0.2327793	- 1.36017126	- 0.24725026
- 1.42922417	- 0.36024678	- 1.35357551	- 0.35306181
- 0.70360729	- 1.67003514	- 0.25022644	- 0.09043084
- 0.74282788	- 1.73711569	- 0.08585676	- 0.31613675
- 1.32651229	- 1.39829398	- 0.519118	- 1.26297133
- 1.36606182	- 1.38104057	- 0.56796527	- 1.34890163
- 1.02923299	- 1.69636245	- 0.38380833	- -1.2597687
- 0.44878023	- 1.81330176	- 0.61820223	- 0.79604225
- 1.20869014	- 3.47936648	- 0.06751823	- 1.12605884
- 1.49542418	- 3.60445368	- 0.42166091	- 0.50306418
- 1.69817454	- 0.49611665	- 0.22308851	- 0.46669569
- 1.29114724	- 1.09463387	- 0.14616091	- 0.25570525

1.4005663	-	-	-
-	0.87851908	0.24287576	0.14926102
0.14325823	1.69401379	0.10693133	0.83182944
-0.0542718	1.73283395	0.21683768	1.02517261
-	0.99260202	1.99435254	-0.772571
1.58830928	1.27773044	1.89567833	0.71132591
-	-	-	-
1.32069212	0.05645413	0.54204338	1.2885273
-	-	-	-
2.03898249	0.00038055	0.38328976	1.00925384
-	0.47928626	1.65023129	1.84368215
1.91044654	1.65023129	1.84368215	0.04058891
0.47928626	3.60223244	2.49632321	0.9083771
2.36748371	-	-	-
3.2682731	1.28067858	0.20814858	0.30346152
-	-	-	-
3.48612441	1.45990627	0.05501104	0.45544085
-	-	-	-
1.64013594	0.45678005	0.74171118	1.01856313
-	-	-	-
2.49508377	0.43823338	0.46753578	0.55164451
-	-	-	-
0.2170624	-2.2340917	0.12812209	1.04350973
-	-	-	-
0.22694978	2.22977835	0.11591027	1.02202715
-	-	-	-
0.25628299	2.30117225	0.03624759	0.63694214
-	-	-	-
0.72685063	1.12467171	1.43572471	0.39467441
-	-	-	-
0.85505764	1.15293209	1.16144868	0.99458517
-	-	-	-
-1.4431382	0.60489953	0.71677581	0.33203222
-	-	-	-
0.57863193	1.41989235	0.31486618	0.20723348
-	-	-	-
1.93174631	0.31185936	0.12581191	0.52147942
-	-	-	-
1.96140846	0.32479941	0.08917646	0.58592714
-	-	-	-
0.83963847	-2.0466845	0.75674477	0.63052974
-	-	-	-
1.06737717	1.86314345	0.92209414	0.80399165
-	-	-	-
1.14647621	1.82863663	1.01978867	0.97585224

-	-	-	-
1.58275762	0.32599446	0.46973835	0.24137845
-	-	-	-
1.33524416	0.51816221	0.27996535	0.45780551
-	-	-	-
0.56452609	1.89908998	0.19014704	0.25008771
-	-	-	-
0.76227371	1.81282295	0.05408929	-0.6797392
-	-	-	-
0.66095629	0.96813584	-1.8351586	0.78757442
-	-	-	-
0.8494744	0.71751423	1.83417892	1.0206801
-	-	-	-
0.91480288	1.27323839	0.73006795	0.52317404
-	-	-	-
-0.3597938	1.85212192	0.50829588	0.60269908
-	-	-	-
1.82820785	0.48652903	1.37999471	0.07461585
-	-	-	-
1.73955036	0.44101524	1.15456595	0.43936461
-	-	-	-
-1.7692125	0.42807518	-1.1912014	0.37491689
-	-	-	-
-1.9334806	0.58022772	0.73077286	0.74254947
-	-	-	-
1.76539512	0.50690074	0.93837375	0.37734571
-	-	-	-
1.81483203	-0.5284675	0.87731466	0.48475858
-	-	-	-
1.63753451	0.85933259	0.16573025	0.68695805
-	-	-	-
1.66719665	0.84639254	0.2023657	0.75140578
-	-	-	-
1.60754343	0.95660661	0.24461722	0.94314763
-	-	-	-
1.72411131	0.22127897	0.38226006	0.07034536
-	-	-	-
0.88959611	0.49643984	0.86305665	0.06104547
-	-	-	-
0.97858254	0.45761967	0.7531503	0.1322977
-	-	-	-
1.99392882	0.50090027	0.44574646	0.06724651
3.53307955	1.19773818	0.8694027	0.84119827
-	-	-	-
0.97370621	1.10117187	1.01490688	1.25237606
-	-	-	-
0.71630537	1.29765298	-1.2168917	1.01446643
-	-	-	-
0.59456605	2.84011909	0.29896039	0.42361963

1.05960189	3.12718059	0.15947258	0.26542406
1.13199239	1.14453024	0.01172063	2.02280749
0.91381214	-1.4080919	0.04933576	1.85414953
1.41983444	-0.6885529	0.88337803	2.00898962
1.31074431	0.82033373	0.90218559	-2.0933186
2.55372221	0.95205583	1.00449304	0.85268761
2.88000578	1.09439644	0.60150309	0.14376266
-0.2820105	1.5492929	0.05148831	-0.8517107
0.57060681	2.51057887	0.29304763	-0.454751
-0.1233638	1.82145818	0.08726819	0.19765232
0.50072896	1.23588102	0.20083124	0.58919634
0.56874454	1.4242057	0.30265437	0.22871605
0.16689684	2.09879099	0.05602406	0.14528114
0.98401455	1.24304492	0.81555066	0.67355206
1.10299205	1.10695074	0.84657005	0.61070566
0.23477394	2.57816732	0.61678491	0.90097733
0.22488656	2.58248067	0.60457309	0.92245991
0.42263418	2.49621364	0.84880943	0.49280843
0.58116121	2.34286605	0.92867607	0.46972454
2.13764703	0.13407138	0.73559719	0.75154343
1.08429371	1.015877	0.44823029	1.43008515
1.53472606	0.16352681	0.08075927	0.19291602
1.61415404	0.04468604	0.09858715	0.04413932
1.62437036	0.04396127	0.20189775	-0.2550154
0.30424376	2.05947747	0.60090561	0.41630747
2.6587928	4.63587594	1.24988997	1.87760115
1.43304155	0.57472914	0.77599964	1.10532441

-	-	-	-
1.22540655	0.66530953	0.51955149	1.55645847
-	-	-	-
1.69444917	0.20833891	0.41889551	0.00589764
-	-	-	-
1.66478703	0.19539886	0.45553096	0.05855008
-	-	-	-
1.03353713	1.25663124	0.27671008	0.734014
-	-	-	-
1.24183	0.99738293	0.25130677	0.92415453
-	-	-	-
-1.4632419	0.51193886	0.62567702	0.60970437
-	-	-	-
1.54234095	0.47743205	0.72337156	0.43784377
-	-	-	-
1.47312928	0.50762551	0.63788884	0.58822179
-	-	-	-
0.65210717	2.04430422	0.40919784	0.09827647
-	-	-	-
1.29876213	-1.4085391	0.65796991	0.37505454
-	-	-	-
1.32842427	1.39559905	0.69460536	0.31060682
-	-	-	-
1.55583403	1.29639195	0.97547714	0.18349238
-	-	-	-
1.69392843	1.32033899	1.26196499	0.80488572
-	-	-	-
1.62457963	0.00947753	0.2733334	0.49679422
-	-	-	-
-1.50626	0.04205127	0.53539761	0.08163397
-	-	-	-
0.91508036	0.13595249	1.98858997	1.47015567
-	-	-	-
0.88541822	0.12301243	2.02522542	1.53460339
-	-	-	-
0.58846785	0.09072208	2.2760575	1.85844332
-	-	-	-
1.16297421	0.91240271	0.68879612	0.09832651
-	-	-	-
1.24440378	0.94416844	0.81910672	0.20022819
-	-	-	-
1.38414285	0.54644568	0.52798249	0.78156496
-	-	-	-
0.28233943	1.46495894	0.16701073	-1.172348
-	-	-	-
0.98380527	1.20856119	0.88698631	0.07825755
-	-	-	-
0.80550348	1.37053548	0.78269603	0.14430659
-	-	-	-
1.02749871	1.96473081	0.98876929	1.03869865

1.52285456	- 1.49606134	1.25279287	- 1.15091547
1.21999201	- 1.35871196	0.44475297	0.86755243
1.02158652	- 1.61364691	0.43156147	0.65592932
-1.9709669	- 0.24477881	- 0.03855777	0.92804701
- 1.67401654	- 0.03104429	0.21227431	0.60420709
- 1.21555075	0.43843319	0.34454429	0.96851678
- 0.88926718	0.58077379	0.74753423	0.25959183
- 0.28515013	1.09689013	1.1458878	-0.0889333
- 0.24592955	1.02980958	1.31025748	-0.4955009
0.00191285	1.30631129	1.38450806	- 0.39129066
- 2.35863798	- 0.76570185	0.20566475	1.66630016
- 2.08146238	- 0.56059404	0.4320732	1.38542538
- 2.29725047	- 0.38711942	- 0.44154772	1.63697196
- 2.09917392	- 0.21651842	-0.3128338	1.52795777
- 1.70607084	- 0.48102063	1.01164465	0.24845027
- 1.65663393	- 0.45945387	1.07270373	0.1410374
- 1.36957095	- 0.25003271	1.31132399	- 0.16131996
0.69306204	- 1.84301641	- 0.03139343	- 0.52936118
1.32982962	- 1.21156464	0.20516683	- 0.23110053
- 2.20859298	- 0.43263321	- 0.21611895	1.1229915
- 1.26498766	0.41686643	0.2834852	1.07592965
1.1355613	2.70928463	1.3515666	0.32549274
1.00702535	2.65321106	1.19281298	0.60476621
- 1.77061785	- 0.24404078	- 0.70176286	0.16679046
- 1.77061785	- 0.24404078	- 0.70176286	0.16679046

-	-	-	-
1.59231606	0.40601507	0.59747258	0.10074142
-1.6621856	0.20715369	0.45191047	0.39015516
-	-	-	-
1.45422166	0.38206803	0.31098473	0.52065192
-	-	-	-
1.51387488	0.27185396	0.26873322	0.71239377
-	-	-	-
0.82680326	1.05905667	0.27359435	0.63208101
-	-	-	-
0.11745994	2.07722423	0.00503503	0.03786827
-	-	-	-
1.98258857	0.51746052	0.55419137	1.17059964
-	-	-	-
1.85405262	0.46138695	0.71294498	0.89132617
-	-	-	-
1.83427785	0.45276024	0.73736861	0.84836103
-	-	-	-
1.63653024	0.36649321	0.98160495	0.41870954
-	-	-	-
1.16160701	0.07511836	1.45224973	0.29181672
-	-	-	-
-1.7258456	0.48964733	0.98722101	0.29141541
-	-	-	-
1.63685917	0.45082717	1.09712736	0.09807225
-	-	-	-
0.85175761	3.07108393	0.02553993	0.93751962
-	-	-	-
1.66285179	3.50911274	0.86030662	0.50341417
-	-	-	-
1.08023547	0.27760704	0.75802525	1.11591968
-	-	-	-
-0.91215	0.35093402	0.96562613	1.48112345
-	-	-	-
1.10505527	0.75424842	0.12608668	0.51113841
-	-	-	-
1.06517681	0.85583579	0.19276183	0.74584541
-	-	-	-
1.72978024	0.24929196	0.86284966	-0.5706317
-	-	-	-
2.0791559	0.08840469	0.74938725	0.29065099
-	-	-	-
0.62417931	1.78887591	0.23239856	0.05834586
-	-	-	-
0.49531442	1.92928345	0.27562976	0.09970969
-	-	-	-
0.72305312	-1.7457424	0.11028039	-0.2731716
-	-	-	-
2.41159098	0.11729678	0.12263698	0.54931008

2.64921706	0.30515119	0.30019816	0.74425457
2.13351665	-0.4531849	1.05059296	2.38855754
2.99112394	0.45988994	1.24186201	3.24607237
2.67437237	0.85320955	0.70981275	0.34421834
3.46733646	1.70428145	0.42600192	0.54943618
-	-	-	-
1.55776012	0.41632035	1.19482189	0.07378835
-	-	-	-
1.63685917	0.45082717	1.09712736	0.09807225
0.16903085	2.07162406	0.67861971	0.60921525
0.17891823	-2.0673107	0.66640789	0.58773268
0.41621537	1.96379026	0.37332429	0.0721509
1.31994224	1.21587799	0.19295501	0.20961795
1.51801879	-1.045277	0.32166893	0.31863214
-	-	-	-
0.15378415	1.67722794	2.29153156	0.8760001
-	-	-	-
0.04502296	1.62978107	2.15720157	0.63969178
-	-	-	-
0.03513558	1.62546772	2.14498976	0.61820921
1.43001917	0.83109529	0.13480467	0.83662292
1.31071273	1.05152343	0.05030164	0.45313922
1.32060011	1.04721007	0.03808982	0.43165664
0.75238633	-1.8171363	0.04187747	0.65825662
0.89080966	1.75674937	0.2128429	0.95901266
1.61656367	1.08647744	0.55930951	0.85409518
0.53791397	1.30632584	1.29383311	0.87583914
0.39883277	1.53538068	1.23375371	0.53532059
-	-	-	-
1.65283643	0.16161682	0.25274059	0.01953459
-	-	-	-
1.44487249	0.33653116	0.11181486	0.11096217

-	-	-	-
1.55429155	0.12041637	0.01510001	0.51592845
-	-	-	-
1.69350989	1.25137151	1.40483628	0.69873352
-	-	-	-
1.8912575	1.16510448	1.64907261	0.26908203
-	-	-	-
-0.5189787	1.53010642	-0.3571177	0.01549163
-	-	-	-
0.19673025	1.91133942	0.35340406	0.53477647
-	-	-	-
0.10774382	1.95015959	0.24349771	0.34143331
-	-	-	-
2.23535241	0.30802482	0.70213667	-0.3836938
-	-	-	-
2.58239755	0.71199401	0.62129034	-0.173672
-	-	-	-
1.65229822	0.21146704	0.43969865	0.41163773
-	-	-	-
1.59297393	0.23734715	0.36642775	0.54053318
-	-	-	-
1.59297393	0.23734715	0.36642775	0.54053318
-	-	-	-
1.42488846	0.31067413	0.15882687	0.90573694
-	-	-	-
0.34381655	1.23967794	0.84157207	0.68123673
0.78830689	2.33979918	1.34215591	0.86728057
-	-	-	-
1.51834773	0.96094304	0.20614651	0.00200516
-	-	-	-
1.68643321	0.88761606	0.41374739	-0.3631986
-	-	-	-
2.0924737	0.54210071	0.68338704	0.60270955
-	-	-	-
2.19610417	0.02576731	1.11134856	0.76990385
-	-	-	-
2.14666727	0.00420055	1.05028948	0.66249098
-	-	-	-
2.19610417	0.02576731	1.11134856	0.76990385