

# UCUENCA

**Universidad de Cuenca**

Facultad de Ciencias Químicas

Carrera de Ingeniería Industrial

**Sistema de Clasificación de Inventarios basado en Algoritmos de Machine Learning**


Trabajo de titulación previo a la obtención del título de Ingeniero Industrial

**Autor:**

Carlos Johao Romero Bustamante

**Director:**

Juan Carlos Llivisaca Villazhañay

ORCID:  0000-0003-2154-3277

**Cuenca, Ecuador**

2023-08-25

## Resumen

La gestión efectiva de inventarios es esencial para optimizar el control, almacenamiento y distribución de productos dentro de un sistema. En este estudio, se utilizó un enfoque basado en análisis estadístico y algoritmos de machine learning para determinar la clasificación óptima de ítems en el sistema de inventario de repuestos del sector automotriz. Para esto se examinó una base de datos que contenía las ventas de repuestos de una empresa automotriz a lo largo de un año. Mediante la aplicación de los algoritmos K-medias, Clustering Large Applications (CLARA) y Divisive Analysis (DIANA), se identificó una clasificación óptima distribuida en tres clústeres. Además, se realizó un análisis comparativo con la clasificación ABC para definir las características de cada agrupación. Los resultados demostraron que el algoritmo CLARA mejora la gestión de inventarios, permitiendo optimizar los espacios de almacenamiento, aumentar la eficiencia operativa, reducir costos, mejorar el servicio al cliente y tomar decisiones informadas. Se puede mencionar que, algunos productos destacados en las agrupaciones resultantes fueron el 2452084002, 5810159A00 y 3910045800 de las agrupaciones 1, 2 y 3 respectivamente; estos productos son relevantes debido a su total de ventas en cada agrupación relacionando su cantidad, costo y precio de venta. Este estudio contribuye al campo de la gestión de inventarios al demostrar cómo el uso de algoritmos de machine learning mediante análisis estadístico puede optimizar la clasificación de artículos en el inventario, siendo relevante en la toma de decisiones estratégicas mediante una distribución más precisa y adaptada a las necesidades de la empresa.

*Palabras clave: clara, diana, k-medias, agrupamiento, inventario*



El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Cuenca ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por la propiedad intelectual y los derechos de autor.

Repositorio Institucional: <https://dspace.ucuenca.edu.ec/>

### Abstract

Effective inventory management is essential to optimize the control, storage and distribution of products within a system. In this study, an approach based on statistical analysis and machine learning algorithms was used to determine the optimal classification of items in an automotive parts inventory system. For this purpose, a database containing the spare parts sales of an automotive company over the course of a year was examined. By applying the K-means, Clustering Large Applications (CLARA) and Divisive Analysis (DIANA) algorithms, an optimal classification distributed in three clusters was identified. In addition, a comparative analysis with the ABC classification was performed to define the characteristics of each cluster. The results showed that the CLARA algorithm improves inventory management, allowing to optimize storage space, increase operational efficiency, reduce costs, improve customer service and make informed decisions. It can be mentioned that, some outstanding products in the resulting clusters were 2452084002, 5810159A00 and 3910045800 from clusters 1, 2 and 3 respectively; these products are relevant due to their total sales in each cluster relating their quantity, cost and sales price. This study contributes to the field of inventory management by demonstrating how the use of machine learning algorithms through statistical analysis can optimize the classification of items in the inventory, being relevant in strategic decision making through a more accurate distribution adapted to the needs of the company.

*Keywords: clara, diana, k-means, clustering, inventory.*



The content of this work corresponds to the right of expression of the authors and does not compromise the institutional thinking of the University of Cuenca, nor does it release its responsibility before third parties. The authors assume responsibility for the intellectual property and copyrights.

Institutional Repository: <https://dspace.ucuenca.edu.ec/>

---

## Índice de contenido

1. Introducción .....	9
1.1. Objetivo general: .....	13
1.1.1. Objetivos específicos: .....	13
1.2. Hipótesis: .....	13
2. Marco teórico .....	13
2.1. Inventarios de repuestos .....	14
2.2. Importancia de los Inventarios .....	14
2.3. Características de inventarios .....	16
2.4. Técnicas de machine learning en la segmentación de inventarios .....	17
2.4.1. Algoritmos de clustering .....	18
2.5. Clustering por particiones .....	18
2.5.1. K-medias .....	18
2.5.2. Clustering Large Applications (CLARA) .....	19
2.6. Clustering jerárquico .....	20
2.6.1. Divisive Analysis (DIANA) .....	21
2.7. Segmentación y beneficios en diferentes campos .....	21
2.8. Número de clústeres y métodos de obtención .....	23
3. Metodología .....	24
3.1. Enfoque de la investigación: .....	25
3.2. Fase 1: Obtención, comprensión y preparación de la base de datos .....	25
3.3. Fase 2: Obtención del óptimo número de clústeres. ....	27
3.4. Fase 3: Aplicación de los algoritmos K-medias, CLARA y DIANA. ....	28
3.5. Fase 4: Interpretación de los resultados y clasificación de repuestos .....	28
4. Resultados .....	29
4.1. Obtención del número de clústeres .....	30
4.2. Ejecución de los algoritmos K-medias, CLARA y DIANA con 4 clústeres .....	33
4.3. Análisis de los resultados obtenidos utilizando 4 clústeres .....	37
4.4. Elección óptima y definitiva del número de clústeres .....	39
4.5. Ejecución de los algoritmos K-medias, CLARA y DIANA con 3 clústeres .....	40
4.6. Análisis comparativo utilizando el Sistema de clasificación ABC .....	44
4.7. Selección del algoritmo óptimo para la clasificación adecuada de ítems .....	49
5. Discusión .....	51
6. Conclusiones .....	55
Referencias .....	56
Anexos A .....	61
Anexos B .....	62

**Índice de figuras**

Figura 1 Unidades de vehículos vendidas anualmente durante el periodo 2000 - 2022	10
Figura 2 Evolución de la importación de autopartes durante el periodo 2016 - 2022	11
Figura 3 Metodología del proyecto de investigación	25
Figura 4 Obtención del número de clústeres para el algoritmo K-medias	31
Figura 5 Obtención del número de clústeres para el algoritmo CLARA	32
Figura 6 Agrupación de datos mediante el algoritmo K-medias (4 clústeres)	34
Figura 7 Agrupación de datos mediante el algoritmo CLARA (4 clústeres)	35
Figura 8 Agrupación de datos mediante el algoritmo DIANA (4 clústeres)	36
Figura 9 Agrupación de datos mediante el algoritmo K-medias (3 clústeres)	41
Figura 10 Agrupación de datos mediante el algoritmo CLARA (3 clústeres)	42
Figura 11 Agrupación de datos mediante el algoritmo DIANA (3 clústeres)	43
Figura 12 Diagrama de Pareto - Sistema de clasificación ABC	45
Figura 13 Distribución de datos mediante el algoritmo CLARA	47
Figura 14 Distribución de datos mediante el algoritmo DIANA	47
Figura 15 Distribución de datos mediante el algoritmo K-medias	48

**Índice de tablas**

Tabla 1 Distribución de datos en los clústeres asignados a cada algoritmo .....	37
Tabla 2 Ítems asignados a clústeres como único elemento .....	38
Tabla 3 Obtención del número de clústeres para el algoritmo K-medias y CLARA.....	39
Tabla 4 Resultados obtenidos mediante Clasificación ABC.....	46
Tabla 5 Distribución de datos en los clústeres de cada algoritmo y segmento ABC .....	49
Tabla 6 Distribución de datos en los clústeres del algoritmo CLARA según las marcas de vehículos .....	50
Tabla 7 Ítems más destacados en cada agrupación y para cada marca de repuestos .....	53

### Dedicatoria

A mi madre, Marianela Bustamante. Tu presencia constante a lo largo de mi vida ha sido mi mayor fortaleza y motivación para alcanzar mis metas. Tu humildad y sabias enseñanzas me han convertido en la persona que soy hoy. Eres la mejor maestra que he tenido, guiándome con tu amor y ejemplo. Te amo profundamente, mamá, y te agradezco por todo lo que has hecho por mí.

A mi padre, Juan Carlos Romero. Tu inquebrantable apoyo y enseñanzas han sido fundamentales en mi camino. Gracias por mostrarme el valor de los sacrificios y la importancia de la dedicación, guiándome por el camino correcto y asegurándote de que no renuncie a mis sueños. Este logro también es tuyo.

A mis hermanos, Juan Romero y Ana Romero. Ustedes no solo son mis hermanos, sino también mis mejores amigos, aquellos que han estado a mi lado en cada paso del camino. Juntos hemos enfrentado los momentos más difíciles con valentía y siempre hemos encontrado motivos para sonreír. Su compañía y apoyo incondicional han sido mi mayor fortaleza. Me siento profundamente orgulloso de ustedes y agradezco infinitamente el amor y la complicidad que compartimos. Sigamos juntos, motivándonos mutuamente en este hermoso viaje llamado vida.

A Xiomara Aguirre, tu compañía y apoyo incondicional han sido un regalo invaluable en mi vida y en la de mi familia. Eres una persona excepcional, siempre mostrando tu bondad y generosidad hacia nosotros. Cada momento compartido ha sido una fuente constante de inspiración y motivación. Gracias por ser ese ejemplo vivo de una buena persona. Tu presencia en mi vida ha dejado una huella imborrable.

## Agradecimientos

Quiero agradecer en primer lugar a Dios y a la Virgen María por haber derramado innumerables bendiciones sobre mi amada familia y sobre mí a lo largo de esta etapa repleta de momentos de alegría, tristeza, sacrificio, desafíos y con la profunda satisfacción de haberlo concluido de manera exitosa porque de la mano de Dios todo es posible.

Agradezco a todas las personas que contribuyeron para la finalización de este proceso académico. Quiero destacar de manera especial a mi tutor, Ing. Juan Carlos Llivisaca, por su constante apoyo y guía invaluable que hizo posible la realización de este trabajo de titulación, demostrando durante esta etapa ser más que un profesional, un amigo.

Quiero expresar mi más sincero agradecimiento a la Universidad de Cuenca por haber sido mi hogar académico durante estos años de formación. Agradezco profundamente a esta prestigiosa institución por brindarme la oportunidad de adquirir conocimientos, desarrollar habilidades y forjar un camino hacia el éxito profesional.

A mis amigos, quiero transmitirles mi gratitud por su apoyo constante a lo largo de esta travesía. Sus palabras de aliento y los momentos felices que compartimos han sido un impulso importante en mi camino hacia el éxito. Sin duda, su compañía ha sido fundamental y valiosa para mi desarrollo personal y académico.

Una vez más, agradezco a cada persona que formó parte de este logro, en especial a mi familia, su contribución ha dejado una huella indeleble en mi vida.



## 1. Introducción

El análisis de segmentación mediante algoritmos de machine learning hace referencia a la búsqueda específica de patrones o tendencias que permitan la creación de modelos en minería de datos con el fin de obtener datos procesables y estadísticas detalladas (Servente, 2002).

Según Torres & Cárdenas (2021), los algoritmos usados para la clasificación están diseñados para dividir un grupo de información en conjuntos de datos, cuyas características presentan homogeneidad y estos puedan diferenciarse de otros. Métodos muy utilizados como técnica de agrupación son los denominados algoritmos de clúster y algoritmos jerárquicos ya que, estos permiten una mayor apreciación de los arreglos en el conjunto de datos a través de gráficas de dispersión y dendrogramas. En la actualidad, es común que las empresas y organizaciones gestionen grandes volúmenes de datos, aprovechando el avance tecnológico actual para llevar a cabo investigaciones y aplicar diversos métodos analíticos sobre estos. Los resultados obtenidos de estas investigaciones les permiten a las empresas desarrollar estrategias que impulsen mejoras en sus procesos y, por consiguiente, fomenten el crecimiento productivo en la industria.

Por otro lado, la utilización de datos de ventas para segmentar productos es un elemento crucial para la formulación de estrategias y la toma de decisiones que impulsan la aplicación de modelos relevantes para mejorar las actividades comerciales de un negocio. Con el transcurso del tiempo y considerando el impacto de la globalización, la implementación de estrategias de agrupación de artículos con el objetivo de que una empresa pueda llevar a cabo iniciativas en diferentes áreas, optimizar recursos y generar mayores ingresos no es algo novedoso. En las organizaciones, una función crucial se encuentra en la gestión de inventarios o almacenes, donde se pueden almacenar volúmenes significativos de artículos. Este aspecto es estratégico para el éxito de un negocio, ya que su gestión adecuada puede tener un impacto positivo o negativo en el mismo.

Se conoce que en Ecuador, uno de los sectores industriales que estima un ligero crecimiento para el año 2023 es el sector automotor. Según el Boletín Sector Automotriz en Cifras de la Asociación de Empresas Automotrices del Ecuador (AEADE) se proyecta un crecimiento de entre 2 y 3% en ventas en comparación con el año 2022, el cual según gremios se considera un año de recuperación económica y comercial (Zúñiga, 2023).

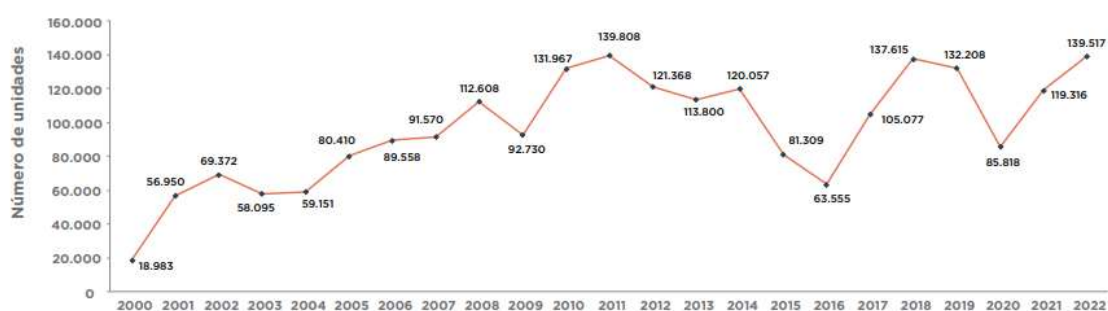
La actividad productiva del sector automotor en Ecuador está centrada en la venta de vehículos, mantenimiento, reparación de vehículos, venta de autopartes y mantenimiento de motocicletas; si bien, los mayores ingresos se obtienen con la venta de vehículos es importante tener en cuenta que la disposición de autopartes representa un factor estratégico con el que se pueden desarrollar técnicas que, con la ayuda de la tecnología permitan establecer mejoras en el área que requiera de estos ítems. Por lo expuesto, es importante conocer que el sector automotriz es un pilar fundamental de la economía nacional y un sector con crecimiento constante.

Durante el año 2022, el sector automotriz ha desempeñado un papel positivo en el impulso del empleo en la economía nacional. Según el Registro Estadístico de Empleo en la Seguridad Social (REESS) elaborado por Hernández (2022), se registraron un total de 533601 puestos de trabajo en el sector automotor hasta el mes de diciembre. Esto representa un incremento significativo en la generación de empleo en comparación con el REESS realizado por Zambrano & Vera (2022) hasta marzo del mismo año, que reportaba 517871 plazas de empleo.

Según el anuario elaborado por la AEADE (2022), se presenta otra evidencia que respalda el notable crecimiento mostrado por el sector automotriz. En dicho informe se reporta un total de 139517 vehículos vendidos, abarcando tanto vehículos livianos como pesados, tal como se muestra en la Figura 1. Estos datos constituyen un indicador claro del impulso y la demanda en el mercado automotriz.

### Figura 1

*Unidades de vehículos vendidas anualmente durante el periodo 2000 - 2022*



*Nota:* Esta figura muestra la cantidad de vehículos vendidos anualmente desde el año 2000 hasta el 2022. Se incluyen tanto los vehículos livianos como los pesados. Adaptado de Anuario 2022, de: AEADE, 2022.

Además, el anuario también revela un aumento en la importación de autopartes durante el mismo año, alcanzando un valor de USD 475 millones FOB (Free on board). Este valor supera los registros de años anteriores, lo que indica un crecimiento significativo en la importación de componentes automotrices, como se ilustra en la Figura 2.

**Figura 2**

*Evolución de la importación de autopartes durante el periodo 2016 - 2022*



*Nota:* Esta figura muestra la evolución de la importación de autopartes en millones FOB desde 2016 hasta 2022, revelando un promedio anual de 371 millones FOB durante este período. Adaptado de Anuario 2022, de: AEADE, 2022.

Se conoce que los algoritmos utilizados para segmentar conjuntos de datos tienen como objetivo agrupar cada uno de los elementos en clústeres o agrupaciones que exhiben características distintivas. Los algoritmos de agrupación, conocidos como algoritmos de conglomerados, se dividen en jerárquicos y no jerárquicos (particiones). Lo común de los dos es que se puede realizar particiones, jerarquización, densidades, entre otras actividades, y permiten a los usuarios formar grupos de manera interpretativa y de fácil comprensión. La agrupación centrada en particiones comienza con un valor conocido de  $k$  clústeres a los cuales, las observaciones se distribuirán de manera iterativa minimizando la distancia de cada observación en relación con el centroide; por otro lado, la agrupación jerárquica aglomera (en caso de que los clústeres vayan fusionándose) o divide (en caso de que los clústeres se separen) un conjunto de datos en forma de dendrograma, con el objetivo de formar agrupaciones que contengan datos con características similares.

En esta investigación se emplearon algoritmos de agrupación por particiones y jerárquica, como K-medias, Clustering Large Applications (CLARA) y Divisive Analysis (DIANA). Las razones para elegir estos algoritmos son las siguientes: entre los métodos de agrupación

por particiones, el algoritmo K-medias es ampliamente utilizado debido a su simplicidad y rapidez en la obtención de resultados, aunque puede presentar dificultades al tratar con grandes volúmenes de datos. Por esta razón, se utilizó también el algoritmo CLARA, que es más robusto y supera las limitaciones del K-medias al trabajar con grandes cantidades de datos. En cambio, para poder comparar los métodos de particiones y jerárquicos, se toma el algoritmo DIANA, que permite la formación de clústeres independientes a partir de un único clúster, con sus respectivas observaciones. Estos tres algoritmos han sido ampliamente utilizados en numerosos estudios, incluyendo aquellos llevados a cabo por Rifa et al. (2020), Nietto & Nicoletti (2017) y Popat & Emmanuel (2014).

Por otro lado, el conjunto de datos con el que se realizó la ejecución de los algoritmos fue obtenido de una base de datos que representa el movimiento de venta de autopartes o repuestos en una empresa del sector automotriz dentro de la ciudad de Cuenca. Este caso de estudio es representativo a una empresa del sector automotriz que se encuentra consolidada muchos años en el mercado nacional dedicándose a la comercialización de vehículos y la venta de repuestos.

En el caso de estudio, el objetivo es aplicar mediante algoritmos una segmentación de los productos que permita clasificar los mismos de manera óptima dentro del sistema de inventarios que tiene la empresa. Este aspecto adquiere una gran relevancia para la empresa en sus operaciones de venta, ya que abordar eficientemente los problemas de abastecimiento en la bodega puede resultar en un aumento de las ventas. Al contar con los ítems necesarios y satisfacer las expectativas del cliente al momento de realizar una compra, se fomenta la fidelidad y se genera un mayor nivel de satisfacción. Esto contribuye a cubrir las necesidades del cliente de manera efectiva y cumplir con sus expectativas, fortaleciendo así la relación con el cliente.

Durante el desarrollo de esta investigación, se llevó a cabo una exhaustiva revisión de diversas perspectivas teóricas relacionadas con la clasificación de inventarios utilizando algoritmos de machine learning. La sección metodológica detalla las cuatro fases necesarias para el desarrollo de esta investigación, con un enfoque en el uso de la base de datos y la ejecución de los algoritmos en la misma, lo que permitió obtener resultados precisos para una correcta clasificación de repuestos. Asimismo, se identificaron y discutieron posibles limitaciones de este estudio y se propusieron recomendaciones para futuras investigaciones. Esta investigación busca contribuir al avance del conocimiento en la clasificación de inventarios, con un enfoque específico en el uso de algoritmos de machine learning. La intención es aportar nueva evidencia y ofrecer una base sólida para

investigaciones futuras en este campo. A lo largo de este trabajo de investigación, se utilizó un enfoque riguroso y analítico para abordar los objetivos planteados y alcanzar conclusiones fundamentadas.

### **1.1. Objetivo general:**

Determinar la clasificación adecuada de ítems dentro del sistema de inventario de repuestos, mediante el análisis estadístico basado en algoritmos de machine learning para mejorar la clasificación de artículos en las empresas del sector automotriz.

#### **1.1.1. Objetivos específicos:**

- Definir las estrategias de clustering y variables que se tomarán en cuenta para la preparación del conjunto de datos y su correspondiente segmentación.
- Analizar los algoritmos de agrupamiento K-medias, CLARA y DIANA en la segmentación de datos junto con la correcta elección del número de clústeres.
- Proponer un sistema de clasificación de inventarios para los repuestos categorizados mediante los algoritmos de segmentación.

### **1.2. Hipótesis:**

La clasificación de artículos basada en algoritmos de segmentación permite una adecuada distribución de repuestos dentro del inventario, puesto que ofrecen información más acertada para la gestión de los mismos.

## **2. Marco teórico**

La industria automotriz representa un porcentaje importante en cuanto a la participación económica de un país ya que, esta origina altos ingresos debido a las actividades que se realizan directa e indirectamente en la misma. Por ende, un punto importante dentro de las empresas que pertenecen al sector automotriz son sus inventarios; es por eso, que según Angulo-Rivera (2019), mantener un monitoreo y gestión adecuada evitará la generación de problemas relacionados con la desorganización y carencia de control, lo que permitirá un adecuado uso de los recursos que guíen a las empresas a conseguir beneficios.

En la industria automotriz existen empresas dedicadas a la comercialización de repuestos en las que según Tovar Giraldo & Fuentes Romero (2019), al momento de generar una mayor cantidad de ganancias debido al crecimiento de ventas, estas deben diseñar mejoras para la optimización de sus procesos de inventarios. Estos procesos ayudarán a mantener

una cantidad óptima de ítems en stock, que eviten un incremento de costos innecesarios y la insatisfacción de clientes.

Según el estudio de García et al. (2021), los factores a tener en cuenta para una correcta gestión de inventarios son: comprobación de entrada de ítems, ubicación de ítems, despacho de ítems, sistema de actualización de ítems, codificación de ítems, toma de inventario físico y reporte de riesgos internos. Todos estos factores están ligados a una correcta clasificación de ítems dentro del sistema de inventario que permita un mejor manejo de los mismos al momento de la toma de decisiones.

### **2.1. Inventarios de repuestos**

En el estudio de Humberto (2017), se indica que el inventario centrado en la venta de artículos comerciales o de uso como repuestos, significa un activo muy importante al momento de realizar estudios sobre el mismo, ya que representa un punto fuerte involucrándose directamente en la generación de ganancias para la empresa. Al momento de contar con elevadas cantidades de estos, las empresas corren el riesgo de quebrar, provocando una liquidación de la empresa; en cambio, cuando se cuenta con un déficit de estos, afectan a la imagen de la empresa y rompe lazos comerciales.

Estos tipos de inventarios almacenan ciertas cantidades de artículos que pueden servir para evitar paros en la producción al momento de hacer remplazos de piezas en maquinaria o equipos. También, es posible que el inventario contenga repuestos para comercializar, como es el caso de los repuestos de vehículos, maquinaria, equipos, entre otros. Es por eso que resulta necesario hacer estudios puntuales sobre cómo realizar un buen manejo de estos tipos de inventario con el objetivo de dotar a las empresas que hagan uso de estos, información que les permita mejorar la planificación y organización de sus negocios.

### **2.2. Importancia de los Inventarios**

En el ámbito empresarial, el manejo de inventarios constituye una base comercial fundamental para cada empresa. Este proceso permite a las organizaciones mantener un control preciso de los bienes almacenados y su disponibilidad. Los inventarios en las empresas cumplen con un papel muy importante ya que, como se indica en el estudio de Marqués et al. (2017), se pueden destacar las siguientes ventajas:

- Conocer de manera confiable la situación económica de la empresa al final de un periodo.

- Capacidad de realizar pronósticos o predicciones que permitan elaborar cronogramas de producción y planificar actividades relacionadas al uso de ítems dentro de la bodega de la empresa.
- Establecer reservas de ítems de acuerdo a las fluctuaciones de la demanda de mercado.

Los sistemas de inventario tienen como objetivo principal dotar a la empresa u organización de los ítems necesarios para realizar las actividades de forma continua y con un desenvolvimiento constante, de manera que sirva para hacer frente a la demanda de clientes durante periodos de tiempo distintos y consecuentemente, aporte positivamente al crecimiento económico de la empresa.

Es posible evidenciar la importancia del inventario en el estudio de Lima Prudente (2020), en el que se indica que en las empresas dedicadas al sector comercial, los sistemas de inventario son fundamentales para que dentro de las mismas, se cumplan principios de organización, distribución y aprovisionamiento de ítems de manera correcta ya que, de ocurrir lo contrario, las posibilidades de generar grandes pérdidas económicas son bastante altas y estas pueden llevar a una empresa al cierre de la misma o a perder oportunidades dentro del mercado.

Teniendo en cuenta la importancia del inventario dentro de las empresas, es necesario resaltar lo fundamental que es mantener un control eficiente de este junto a una clasificación apropiada de ítems, ya que este genera problemas complejos al ser manejado de manera inadecuada. Un manejo inadecuado presenta un problema para las empresas generándose desorden, planificaciones ineficientes, distribución inadecuada de ítems, clientes insatisfechos, entre otros.

Según el estudio realizado por Holguín et al. (2004), uno de los desafíos más frecuentes que enfrentan los directivos en empresas industriales o comerciales al administrar el inventario es la presencia de excesos y faltantes, comúnmente conocido como desbalanceo de inventarios según el artículo de Gutiérrez & Vidal (2008). Estos estudios resaltan la importancia de abordar de manera efectiva este problema, ya que puede tener un impacto significativo en la gestión operativa y financiera de una organización.

En el estudio llevado a cabo por Villegas (2021), se establece que el exceso de inventario para un producto se define como la presencia de una cantidad significativamente mayor en comparación con su consumo durante un periodo determinado. Por otro lado, la faltante de inventarios se define como la incapacidad de satisfacer la demanda de un producto debido



a su escasez. Por tanto, una gestión adecuada del sistema de inventarios se vuelve fundamental para abordar de manera temprana este desbalance y evitar la devaluación de la empresa.

Los costos asociados a la gestión de inventarios desempeñan un papel crucial en la eficiente administración de los mismos, como se señala en el estudio realizado por Pinzón et al. (2010). Estos costos abarcan la adquisición, almacenamiento, agotamiento y el costo total del inventario. Por lo tanto, resulta fundamental implementar estrategias efectivas de gestión de inventario con el objetivo de mantener un equilibrio entre la disponibilidad y la demanda de los productos, optimizando así los recursos y garantizando la satisfacción del cliente. Al tener en cuenta estos costos y aplicar estrategias adecuadas, las empresas pueden lograr una gestión eficiente de sus inventarios, reduciendo los costos asociados y maximizando el rendimiento operativo y financiero.

### **2.3. Características de inventarios**

Al tener en cuenta que, los inventarios toman en consideración varias cantidades de ítems diferentes que son pieza fundamental en las empresas, tanto para la venta de los mismos como para la producción de otros artículos; dependiendo del propósito de los ítems. Es por eso que, según en el estudio de Durán (2012), existen diferentes tipos de inventarios ya que, estos dependen de la naturaleza de la empresa y el punto de vista que esta tenga planteado. En su artículo se señala una clasificación completa sobre los tipos de inventarios según el criterio al que se haga referencia, siendo los más importantes detallados a continuación:

#### *Primer criterio: Funcional*

Dentro de este primer criterio, se considera la función o actividades a las que la empresa se dedica como por ejemplo, compañías dedicadas a la elaboración de productos, comercialización de artículos o prestación de servicios. Estos comprenden los materiales que actúan como punto inicial de algún proceso (materia prima), productos que se encuentran en elaboración (productos en proceso) y artículos listos para la venta (productos terminados).

#### *Segundo criterio: Existencias almacenadas*

Dentro de este segundo criterio, se considera el motivo por el cual la empresa está almacenando ciertas cantidades de ítems. Estos comprenden la materia prima o artículos



terminados que están listos en caso de que existan fluctuaciones de crecimiento en la demanda del mercado (precautelativos), materia prima o artículos terminados que se encuentran activos para ser utilizados o vendidos con el fin de generar ganancias (operativo) y materia prima o artículos terminados que se encuentran retenidos en caso de que existan variaciones en los precios y se pueda sacar provecho de eso al momento de su venta (especulativo).

#### *Tercer criterio: Duración*

Dentro de este tercer criterio, se considera el tiempo de durabilidad que tienen ciertos ítems dentro del inventario. Estos comprenden la materia prima o artículos terminados tanto con fecha de caducidad (perecederos) como con durabilidad bastante amplia (no perecederos).

#### *Cuarto criterio: Origen*

Dentro de este cuarto criterio, se considera el punto de partida o precedencia de los ítems dentro del inventario. Estos comprenden la materia prima o artículos terminados que son de origen extranjero y hayan sido ingresados al país (importados) y a la materia prima o artículos terminados que son de origen propio tanto en su elaboración como en su comercialización (nacionales).

## **2.4. Técnicas de machine learning en la segmentación de inventarios**

Para que una empresa pueda ofrecer una estrategia diferenciada dentro del mercado es necesario que exista la integración de nuevos métodos y aplicaciones que permitan una relación coordinada entre la gestión del inventario, operaciones y cadena de suministro. Es por ello que, en el estudio realizado por García-Barrios et al. (2021), se destaca el uso de machine learning como herramienta de categorización de productos en stock y resalta su importancia como guía/apoyo en la toma de decisiones de gestión en inventarios con el fin de desarrollar un método eficiente y flexible.

Mediante técnicas de machine learning es posible realizar la segmentación de un conjunto de datos para la toma de decisiones. Técnicas muy utilizadas son el aprendizaje supervisado y el no supervisado que, según Sandoval Serrano (2018), estos se diferencian en que el primero utiliza conjunto de datos etiquetados ideal para realizar predicciones y el segundo utiliza conjuntos de datos no etiquetados, ideal para la agrupación de datos según la similitud en sus características.

### 2.4.1. Algoritmos de clustering

El uso de algoritmos de machine learning ha sido ampliamente utilizado para la segmentación de patrones, agrupación de artículos, clasificación de clientes, entre otros. Aunque existan varios estudios relacionados con metodologías de machine learning, según Moharana & Sarmah (2018), la literatura existente revela que la aplicación de la minería de datos tiene estudios bastante limitados para la gestión de inventarios y más aún cuando se habla de inventarios de repuestos. Es por eso que, es necesario conocer como son aplicados los algoritmos de machine learning en la clasificación de ítems y como se comprende el análisis de sus resultados mediante el tipo de variables que se tengan presentes en el estudio.

## 2.5. Clustering por particiones

Según Leiva-Valdebenito & Torres-Avilés (2010), los algoritmos de partición determinan agrupaciones a los conjuntos de datos sin tener en cuenta niveles jerárquicos, donde el número de clústeres preespecificados pueden ser asignados de manera aleatoria o mediante métodos de obtención, para luego de aplicado un algoritmo de segmentación, designar observaciones de datos a cada agrupación de modo que satisface los criterios del estudio.

Gupta & Panda (2019), en su estudio describen que estos tipos de algoritmos tienen un principio de agrupación basado en la división de conjuntos de datos según un punto central. Este punto central puede referirse a la media, mediana u otros aspectos que permitirán el ajuste de datos en distintos clústeres o agrupaciones. Sin embargo, cuentan con el inconveniente de que necesita de manera previa, el número de agrupaciones que se tienen que realizar. Los algoritmos de clustering más conocidos que emplean este principio son K-medias, K-medoides, Partitioning around Medoids (PAM) y Clustering Large Applications (CLARA).

### 2.5.1. K-medias

Según Ramírez Mendoza (2022), el algoritmo de segmentación K-medias tiene el objetivo de asociar mediante grupos a un conjunto de datos, de los cuales se dividen según la similitud que existen entre ellos. Este método de segmentación mueve los centros de cada clúster o agrupación hacia las posiciones medias de los datos constituyentes, reasignando una y otra vez a cada grupo hasta que ya no existan variaciones significativas.

Para el correcto funcionamiento del algoritmo K-medias necesita obligatoriamente un valor de entrada referente al número de clústeres, asociando centroides en los cuales los datos se asignarán dependiendo de la cercanía que tengan con estos. Este algoritmo es uno de los más utilizados en los estudios relacionados con la minería de datos y la clasificación de ítems, ya que este es un método rápido y sencillo al momento de ser aplicado a un conjunto de datos.

A pesar que, este método de clustering es uno de los algoritmos más comunes al momento de realizar una segmentación de datos, este resulta ser muy sensible al ser ejecutado cuando en el conjunto de datos existe la presencia de outliers o datos atípicos. Para resolver dicha cuestión, la mejor opción es disminuir el peso de los datos atípicos mediante técnicas robustas. Es muy común que, cuando se trabaja con conjuntos de datos que contienen miles de observaciones, los datos atípicos se encuentren presentes con normalidad. La presencia de datos atípicos, multivariados y la necesidad de un algoritmo que pueda aplicarse a grandes cantidades de datos, da como solución el uso de algoritmos de clustering como CLARA (Clustering Large Applications) y DIANA (Divisive Analysis).

### **2.5.2. Clustering Large Applications (CLARA)**

En el estudio de Leiva-Valdebenito & Torres-Avilés (2010), se puede considerar que el algoritmo de segmentación Clustering Large Applications (CLARA) tiene el objetivo de hacer una división de muestras tomadas de un conjunto de datos aplicando el algoritmo PAM, encontrando los K-medoides de todas las observaciones existentes en una de esas muestras aleatorias y agrupando todo el set de datos.

Este algoritmo es una extensión del algoritmo K-medoides según lo establece Kaufman & Rousseeuw (1990), teniendo en cuenta la ventaja de ser capaz de manejar grandes conjuntos de datos y por ende ser menos sensible a los datos atípicos, haciéndolo un método más robusto. Este método selecciona como agrupaciones finales los medoides que obtienen una cantidad mínima en la suma total de sus distancias intra-cluster.

Debido a que lo más común en estudios relacionados con la minería de datos es trabajar con conjuntos de datos de gran volumen, CLARA es una mejor opción que K-medias. Esto se ha evidenciado en varios casos como el de Gupta & Panda (2019), en el cual se utiliza el algoritmo CLARA para hacer comparaciones con K-medias, siendo aplicados simultáneamente a un mismo conjunto de datos. Esta aplicación sencilla afirma que de los dos algoritmos, el método de clustering CLARA es el que ofrece los mejores resultados al momento de realizar una segmentación de un gran volumen de observaciones.

Teniendo en cuenta que el algoritmo CLARA es un método de clasificación muy efectivo, este también puede ser complementado con otro tipo de metodologías, sirviendo como base de la aplicación de otras clases de segmentación. El estudio realizado por Abera & Khedkar (2020) aplica lo mencionado anteriormente, ejecutando el método CLARA a una base de datos de electrodomésticos con el fin de conocer una clasificación de los mismos según el consumo eléctrico. A partir de eso, se usan metodologías como Support Vector Machine (SVM) y Artificial Neural Network (ANN) para efectuar otro tipo de clasificación más precisa que permita obtener predicciones sobre el consumo de energía eléctrica que tienen los clientes en ciertos periodos específicos.

Si bien K-medias y CLARA son algoritmos que clasifican el número de observaciones de un conjunto de datos de manera factible, la visualización de esta clasificación es muy importante al momento de realizar análisis. Estos algoritmos son visualizados normalmente mediante gráficas de dispersión, las cuales indican de manera sencilla la distribución de cada clasificación junto con los datos que le corresponden. Sin embargo, existen otros métodos de análisis en los cuales se puede visualizar la segmentación de los datos de manera jerárquica, permitiendo realizar otro tipo de análisis al momento de ejecutar una clasificación por grupos de forma multivariada.

## **2.6. Clustering jerárquico**

Los algoritmos de jerarquización determinan agrupaciones de modo aglomerativo y divisivo, resultando un número creciente de categorías anidadas; es decir, no es necesario contar de manera previa con un número específico de clústeres. Estos algoritmos están representados gráficamente mediante dendrogramas, los cuales resumen la segmentación de datos en un análisis de agrupaciones, cuyas líneas verticales hacen referencia a los grupos de cada etapa.

Estos tipos de algoritmos tienen un principio de agrupación que da inicio a una parte única de la cual se van dividiendo varias clases de manera sucesiva. Uno de los algoritmos de clustering jerárquico más conocidos y utilizados es el Divisive Analysis (DIANA), en el cual al ser aplicado, divide el conjunto de datos en varias agrupaciones de las cuales es posible tomar directamente el número de clústeres que se desean utilizar o tomar en cuenta según un estudio determinado.

### 2.6.1. Divisive Analysis (DIANA)

Muñoz Quintero (2019), indica en su estudio que el algoritmo de segmentación Divisive Analysis (DIANA) tiene como objetivo realizar una segmentación de datos de manera jerárquica, entrando en el grupo de modo divisivo, ya que el primer agrupamiento engloba a los datos de todo el conjunto de observaciones. Luego de esta primera agrupación, comienza la división de los datos en dos agrupaciones siguientes y así sucesivamente hasta quedar divididos de tal manera que solo contengan un solo dato.

Es importante tener en cuenta que este método, al ser jerárquico e inverso, posibilita hacer la elección del número de agrupaciones que se deseen tener ya que este se va dividiendo de forma consecutiva. Además, este tipo de algoritmos son representados mediante una división en forma de árbol inverso denominado dendrograma, en el que se pueden observar cómo se van dividiendo todas las observaciones de un conjunto de datos y permite realizar otros tipos de análisis.

El método DIANA es un algoritmo de segmentación que agrupa de manera jerárquica al conjunto de datos según el análisis que se requiera realizar. Este ha sido muy utilizado para realizar segmentaciones que requieran un análisis posterior de toma de decisiones, o comparativas con datos previamente establecidos. Algo similar se puede evidenciar en el caso de Patnaik et al. (2016), en el que se utiliza el método DIANA para clasificar una gran cantidad de datos de velocidades recogidos por un receptor GPS, en distintas clases según el rango de velocidad al que pertenezca. Gracias a dicho estudio se pueden clasificar redes de carreteras urbanas en un número de clases de calles y velocidades medias de las mismas para una posterior comparativa con datos establecidos en el Highway Capacity Manual del año 2000.

### 2.7. Segmentación y beneficios en diferentes campos

El uso de algoritmos de segmentación se ha visto reflejado sobre todo, en investigaciones en la que se centra la elaboración de estrategias de marketing haciendo una clasificación de clientes. En artículos académicos, como es el caso de Anitha & Patil (2019), es posible mediante un análisis de segmentación de clientes que da información sobre el patrón de compra frecuente del cliente, compra reciente y el beneficio obtenido (RFM) junto con la aplicación del algoritmo de segmentación K-medias, conocer el comportamiento de compra de los consumidores de una empresa del sector minorista para su clasificación en distintos grupos. El objetivo de dicha investigación, es conocer en qué área específica se han realizado la máxima cantidad de ventas, sirviendo de apoyo a la empresa para el diseño de

estrategias de ventas y marketing, permitiendo el aumento de ventas y mejoras en la retención de clientes.

La segmentación de clientes basada en algoritmos de machine learning ha servido de estudio para no solamente ser realizada mediante uno o dos tipos de algoritmos, sino del uso de varios métodos, los cuales, luego de un análisis sea posible escoger el más adecuado para el estudio. En los artículos de conferencia, como es el caso de Parikh & Abdelfattah (2020) y Saraf et al. (2022), se tiene como objetivo conocer el comportamiento de compra de los consumidores en las empresas de tipo físico y online mediante varios métodos y algoritmos de segmentación. Dentro de su estudio se realiza la aplicación de métodos de clustering como Mean-Shift, DBSCAN, Clustering aglomerativo y K-medias, de los cuales K-medias fue el tomado en cuenta al momento de segmentar a los clientes debido a su menor tiempo de ejecución y a su agrupación efectiva.

Estudios realizados para la segmentación de productos se han visto reflejados en varios artículos de investigación, como es el caso de Lolli et al. (2019), Vesga Acevedo (2020) y Gustriansyah et al. (2020), en los cuales se destaca la clasificación de productos tanto para inventarios de e-commerce como para inventarios físicos. En dichos estudios, se aplica la metodología del algoritmo K-medias para clasificar productos en distintas categorías con el fin de conseguir información de los ítems que permita resaltar el impacto económico que tienen los inventarios y mejorar la precisión del proceso de gestión de existencias de los mismos. Es importante tener en cuenta que, una correcta clasificación de inventarios necesita de una adecuada gestión y control de los mismos, la cual puede estar sujeta a una política de reordenación específica en función de su prioridad.

Como se ha podido evidenciar en varios estudios, se puede realizar una ejecución de una gran cantidad de algoritmos de clustering y de estos, elegir el más conveniente para el caso que se esté analizando. Además, estos algoritmos también se complementan cuando están relacionados de forma similar, como es el caso del artículo realizado por Datta & Datta (2006), el que se realiza un análisis de datos de alta dimensión sobre expresión genética dividida en dos medidas de validación: medición de consistencia estadística y congruencia funcional biológica. En este estudio se evalúan 6 tipos de algoritmos: Unweighted Pair Group Method using Arithmetic averages (UPGMA), K-medias, DIANA, Fuzzy Analysis Clustering Algorithm (FANNY), Model-Based y Self-Organizing Maps (SOM), de los cuales no existe ninguno que de manera simultánea pueda realizar una segmentación de datos óptima que abarque los distintos tipos de análisis requeridos y pueda generar conclusiones para el estudio.

Sin embargo, existen algoritmos que muestran un comportamiento similar para un determinado tipo de análisis dentro del estudio realizado. Esto se puede observar en los resultados de las gráficas, en los que se existe relación de los algoritmos de clustering K-medias y DIANA para el análisis de datos sobre esporulación de levadura, los cuales a la vez ofrecen los mejores resultados con respecto a la medida de validación. Esto demuestra que los algoritmos se pueden relacionar entre sí, coincidiendo en varios factores de manera que ofrezcan los mejores resultados según el tipo de análisis requerido y el conjunto de datos con los que se trabaje.

Otro estudio similar ha sido realizado por Muñoz Quintero (2019), en este se puede evidenciar otro tipo de relación de algoritmos con el objetivo de caracterizar a los usuarios del transporte público mediante análisis cluster, utilizando métodos como: CLARA, DIANA y Agglomerative Nesting (AGNES). En este estudio se resalta una relación congruente entre los tres algoritmos evidenciando 2 grupos de usuarios con perfiles similares, en los cuales dependiendo del análisis que se requiera. Estos estudios dan a conocer que gracias al empleo de varios tipos de algoritmos de segmentación se pueden establecer mejores análisis al momento de hacer un estudio de casos, especialmente si se trata de un análisis para conjuntos de datos multivariados.

## **2.8. Número de clústeres y métodos de obtención**

En los algoritmos de particiones es muy común determinar el número de particiones, a esto se les conoce como clústeres. Los clústeres ( $k$ ) hacen referencia a grupos en los cuales se pueden introducir ciertos elementos con características similares con el objetivo de determinar o establecer patrones, asociar artículos o seccionar clientes. Es una técnica muy utilizada cuando existe la necesidad de crear agrupaciones para un análisis posterior, siendo utilizados al momento de aplicar algoritmos de segmentación que hacen uso de estas unidades de asignación como por ejemplo K-medias y CLARA.

Según Gupta & Panda (2019), dado que el algoritmo K-medias es ampliamente utilizado para temas de segmentación es necesario tener en cuenta que para su ejecución es obligatorio determinar la cantidad de clústeres que se utilizarán en el modelo. El número de clústeres es importante ya que, indicará la cantidad de grupos en los que se clasificarán los ítems. En el artículo de Gustriansyah et al. (2020), se realiza la obtención del número de clústeres mediante varios métodos como método del codo, índice de la silueta, índice de Calinski-Harabasz, índice de Davies-Bouldin, índice Ratkowski, índice de Hubert, índice de Ball-Hall e índice de Krzanowski-Lai.



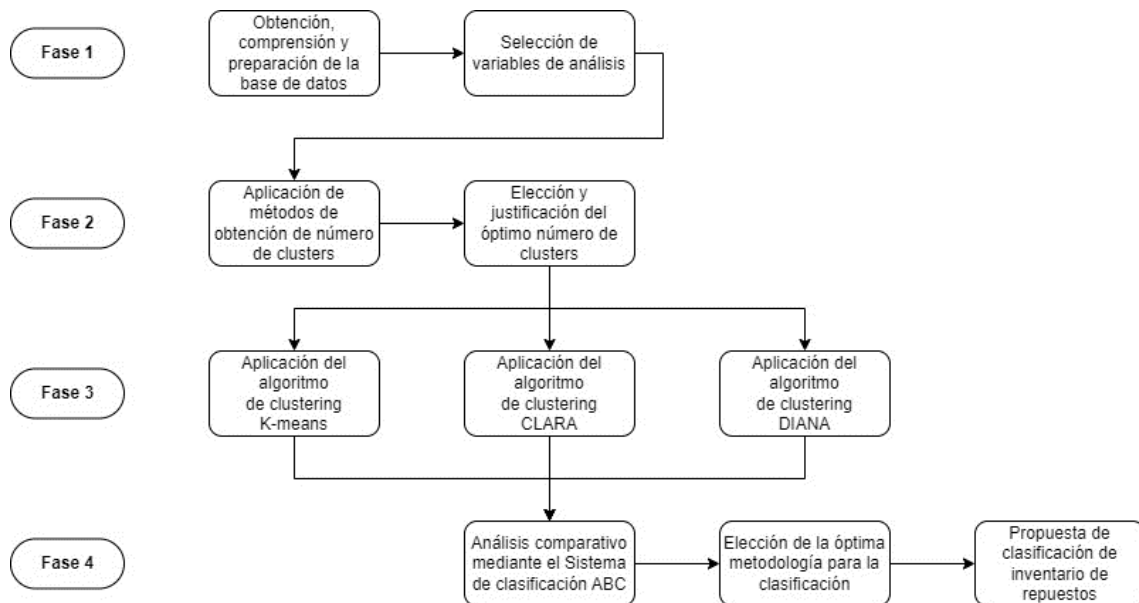
Para la elección del número óptimo de los clústeres al momento de realizar la comparación de distintos métodos, dicha elección puede realizarse a través de una evaluación de la varianza de la relación entre la distancia intra-clúster y la distancia inter-cluster. Por otro lado, en los métodos de agrupación jerárquicas se puede determinar el número de clústeres considerando la altura a la cual se corta el dendrograma, de manera que se pueda tener un número determinado de agrupaciones.

### 3. Metodología

El desarrollo de este trabajo se llevó a cabo a través de varias etapas, en las cuales se establecieron parámetros específicos con el propósito de obtener resultados que faciliten la distribución adecuada de repuestos dentro del inventario utilizando algoritmos de machine learning. Estos resultados proporcionan información precisa y relevante para la gestión de inventarios en la empresa objeto de estudio. En la Figura 3 se muestra el diagrama que ilustra las etapas de la metodología, así como los parámetros asociados a cada una de ellas.

El análisis estadístico realizado mediante la aplicación de algoritmos de machine learning se realizó sobre un conjunto de datos provenientes de la empresa caso de estudio, perteneciente al sector automotriz de la ciudad de Cuenca. Estos datos representan las ventas de repuestos durante un período específico. Es importante resaltar que, con el fin de preservar la confidencialidad de la información sensible de la empresa objeto de estudio, se utilizarán otros tipos de nombres en las variables de análisis. Esta medida tiene el objetivo de preservar la identidad empresarial y evitar riesgos de divulgación no autorizada de información.



**Figura 3***Metodología del proyecto de investigación*

*Nota:* Esta figura muestra la metodología diseñada para el proyecto de investigación distribuida en cuatro fases.

### 3.1. Enfoque de la investigación:

La investigación que se presenta se enfocó en el uso de herramientas estadísticas cuantitativas y algoritmos de machine learning para analizar un conjunto de datos y así poder segmentarlos y clasificarlos de manera objetiva, evitando cualquier tipo de subjetividad o interpretación. Este enfoque se basó en la medición numérica y el análisis de datos, utilizando herramientas especializadas para obtener resultados precisos y confiables. El objetivo de la investigación fue evaluar el comportamiento de los datos y a través del análisis, obtener una clasificación objetiva y precisa, lo que permitió una mejor comprensión de la información y la toma de decisiones fundamentadas en los resultados obtenidos.

### 3.2. Fase 1: Obtención, comprensión y preparación de la base de datos.

La base de datos corresponde al volumen de ventas de repuestos realizados en una empresa del sector automotriz desde el 5 de enero hasta el 28 de diciembre de 2022. Este conjunto de datos incluye información cuantitativa y cualitativa, la cual se utilizó para aplicar los algoritmos de clasificación. Sin embargo, es importante que se realice una depuración de la base de datos antes de su análisis, para asegurarse de que no haya información incorrecta o inconsistente que pueda afectar los resultados y dificultar su interpretación posterior.

Las variables de las cuales está compuesta la base de datos corresponden a: sucursal, número de factura, fecha, documento del cliente, cliente, vendedor, forma de pago, marca, tipo de cliente, familia, picking, número de parte, descripción, línea de repuesto, categoría del repuesto, cantidad, costo unitario, costo total, precio unitario, precio total, descuento, precio neto, total y número de crédito.

El proceso de depuración de la base de datos se llevó a cabo con un enfoque riguroso que incluyó dos pasos principales. En primer lugar, se realizó una selección minuciosa de las variables que se consideraban relevantes para el análisis. Esto implicó la revisión exhaustiva de todas las variables disponibles en la base de datos y la eliminación de aquellas que no eran necesarias o que no aportaban información significativa para los objetivos de análisis del estudio.

La selección de variables relevantes dentro del conjunto de datos requiere un cuidadoso análisis y consideración, teniendo en cuenta tanto el objetivo de la investigación como la calidad de los datos de origen. Es crucial tener en cuenta que, según estudios como Morales & Morales (2019), Guerrero (2017) y Sánchez (2021), la ejecución de algoritmos como K-medias y CLARA se basa en datos cuantitativos exclusivamente.

Para garantizar la correcta ejecución de los algoritmos mencionados en el caso de estudio, se han seleccionado las siguientes variables para la interpretación de los datos: número de parte (código de identificación de cada ítem), cantidad de repuesto vendido, costo total del repuesto, precio neto del repuesto y precio total de venta del repuesto incluido el IVA. Estas variables proporcionarán la información necesaria para llevar a cabo un análisis adecuado utilizando los algoritmos mencionados, permitiendo obtener conclusiones relevantes para el caso de estudio.

En segundo lugar, después de seleccionar las variables relevantes, se llevó a cabo una exhaustiva revisión de los datos asociados a estas variables para identificar y corregir posibles errores. El proceso de depuración de los datos comenzó con una revisión gramatical del número de parte de cada ítem, siguiendo los criterios establecidos por Uribe & Ramírez (2009). Estos criterios incluyeron verificar el orden de las palabras, el uso de mayúsculas y minúsculas, la presencia de espacios en blanco, palabras faltantes, errores ortográficos, errores tipográficos, palabras truncadas, prefijos, sufijos y sinónimos.

Una vez aplicados estos criterios, se procedió a corregir manualmente la identificación de cada repuesto para evitar datos duplicados y recopilar los valores correspondientes. En esta etapa, se sumaron los valores cuantitativos de cada variable en cada ítem, de manera que

cada número de parte contenía los valores totales en cada variable. Esto permitió consolidar varios registros en un solo ítem, que incluía la información completa.

La depuración de datos se llevó a cabo para garantizar la consistencia, validez e integridad de los datos utilizados en el análisis. Durante este proceso, se verificó la calidad de los datos y se detectaron posibles valores atípicos o faltantes en las variables seleccionadas para el análisis. Estos valores fueron corregidos o eliminados según corresponda, con el objetivo de asegurar la confiabilidad de los datos utilizados en el análisis.

El estudio realizado por Uribe (2010) señala que los datos faltantes son un desafío importante al analizar conjuntos de datos, ya que la ausencia de información puede afectar la integridad y la precisión del análisis. Por otro lado, los valores atípicos son observaciones que se desvían significativamente del comportamiento esperado de los datos y pueden influir en los resultados del análisis estadístico. Estos datos atípicos y faltantes no fueron tomados en consideración por falta de información relacionada con las variables elegidas; de esta manera solo se tienen en cuenta el resto de datos que representan el comportamiento real de ventas de la empresa.

### **3.3. Fase 2: Obtención del óptimo número de clústeres.**

En esta etapa, se empleó el software estadístico RStudio (versión 4.2.1) para ejecutar los distintos métodos estadísticos utilizados en la obtención de los clústeres. Para garantizar una programación adecuada de las líneas de código en el software, se tomaron en cuenta las recomendaciones establecidas por Wickham et al. (2023) en relación a la ejecución de líneas de código en el campo de la ciencia de datos.

Para determinar el número óptimo de clústeres que se tomarán en cuenta para la ejecución de los algoritmos. Este número de clústeres fue obtenido mediante la aplicación de diversos métodos de obtención; siguiendo las recomendaciones establecidas en el artículo de Gustriansyah et al. (2020), que incluyen el método del codo, silueta media y estadística de la brecha. La determinación del número óptimo de clústeres permitió clasificar las observaciones en grupos en función de la similitud de sus características.

El número de clústeres que se obtiene en un análisis de clustering depende en gran medida del método que se utilice. Cada método mencionado tiene una forma diferente de agrupar los datos, lo que puede dar lugar a resultados muy distintos. Por lo tanto, es importante comparar los resultados obtenidos por diferentes métodos para determinar la cantidad óptima de clústeres. Una vez que se ha determinado la cantidad óptima de clústeres, se

utilizó esta información para realizar la restricción de agrupamiento en todos los algoritmos de clustering que se utilizaron. Esto significa que, tanto si se utilizan algoritmos basados en particiones como si se utilizan algoritmos jerárquicos, se utiliza la misma cantidad de clústeres para agrupar los datos. Esto simplifica el análisis y hace que los resultados sean más comparables entre los diferentes algoritmos a probar.

#### **3.4. Fase 3: Aplicación de los algoritmos K-medias, CLARA y DIANA.**

En la tercera fase, se utilizaron los algoritmos K-medias, CLARA y DIANA, haciendo uso del programa estadístico RStudio (versión 4.2.1), con el fin de explorar la agrupación del conjunto de datos de repuestos y generar la línea de código que permita ejecutar la segmentación del mismo teniendo en cuenta las recomendaciones de Wickham et al. (2023). La cantidad de clústeres utilizada para la agrupación se basó en los resultados obtenidos en la Fase 2 de la metodología.

De acuerdo con Katsov (2017), las técnicas de agrupación pueden considerarse como métodos de ingeniería de características, ya que permiten agrupar datos en función de sus características y, de esta manera, realizar análisis más precisos al establecer modelos de segmentación. Por esta razón, el uso de algoritmos como K-medias, CLARA y DIANA al realizar agrupaciones contribuye a obtener el mejor rendimiento posible de estos métodos de machine learning.

La agrupación de los datos depende en gran medida del número de clústeres seleccionados, ya que cada algoritmo de clustering creó un número específico de agrupaciones. La visualización de la agrupación de los datos se realizó mediante diversas técnicas gráficas según el algoritmo que se utilizó; por ejemplo con los algoritmos K-medias y CLARA se mostraron gráficos de dispersión y con el algoritmo DIANA, gráficos de jerarquización. La visualización gráfica resultó ser una herramienta muy útil para representar la segmentación de datos en grupos, lo que facilita la identificación de segmentos estadísticamente significativos en términos de similitud o diferencia del conjunto de datos.

#### **3.5. Fase 4: Interpretación de los resultados y clasificación de repuestos.**

Una vez que los datos han sido asignados a sus respectivos clústeres, se extrajeron estos datos para poder conocer específicamente mediante dicha asignación, la identificación de los repuestos que se encuentran en cada clúster y su correspondiente agrupación en el inventario. Al conocer la información específica de cada dato y su respectiva asignación a

un grupo determinado, fue posible llevar a cabo una clasificación adecuada de los repuestos dentro del inventario.

Al interpretar adecuadamente los resultados obtenidos a través del análisis comparativo con el sistema de clasificación ABC, se logró seleccionar el algoritmo de agrupación más apropiado. Esta elección se basó en el índice más alto obtenido por dicho algoritmo en comparación con los demás, siguiendo la recomendación del estudio realizado por Chavez Valderrama & Salinas Flores (2021). Según este estudio, un índice más alto en la distribución de datos indica que el algoritmo seleccionado genera clústeres de mejor calidad en comparación con los otros algoritmos evaluados.

En esta fase final del proceso de clasificación de repuestos, se llevó a cabo un análisis para lograr una interpretación apropiada de los datos y una clasificación efectiva de los repuestos para el inventario de la empresa. En este análisis, se determinaron las agrupaciones finales y se identificaron los conjuntos de datos correspondientes. A partir de ahí, se procedió a proponer una clasificación que sea aplicable dentro del inventario. Esta clasificación es capaz de asegurar una correcta distribución de los repuestos dentro de la bodega y permite una señalización efectiva para abastecer ciertos productos debido a su rotación.

La importancia de esta fase es evidenciada en el estudio de Durán (2012), radica en que la clasificación adecuada de los repuestos permite una gestión eficiente del inventario, reducción de costos y un mejor servicio al cliente. Además, una clasificación inadecuada puede provocar problemas como la falta de stock de ciertos repuestos o el exceso de inventario de otros, lo que puede llevar a una pérdida financiera. Por lo tanto, esta etapa es esencial en el proceso de clasificación de repuestos y debe ser abordada con la debida atención y rigor para lograr resultados exitosos y beneficiosos para la empresa.

#### **4. Resultados**

La base de datos utilizada en el estudio inicialmente constaba de 2773 registros de ventas, los cuales fueron depurados mediante un proceso que unificó la cantidad de productos vendidos, costos y valor en ventas totales para cada ítem. Además, se corrigieron los errores gramaticales relacionados con los números de parte. Para garantizar la integridad de los datos, se decidió no considerar las ventas que incluían notas de crédito en el análisis. Esto se debió a que dichas ventas podrían estar asociadas con devoluciones o cancelaciones, lo que podría causar anomalías en los datos. Este proceso de depuración tenía como objetivo recopilar información precisa y evitar confusiones al interpretar los datos

en las segmentaciones realizadas por los algoritmos. Como resultado, la base de datos depurada cuenta con 1602 registros.

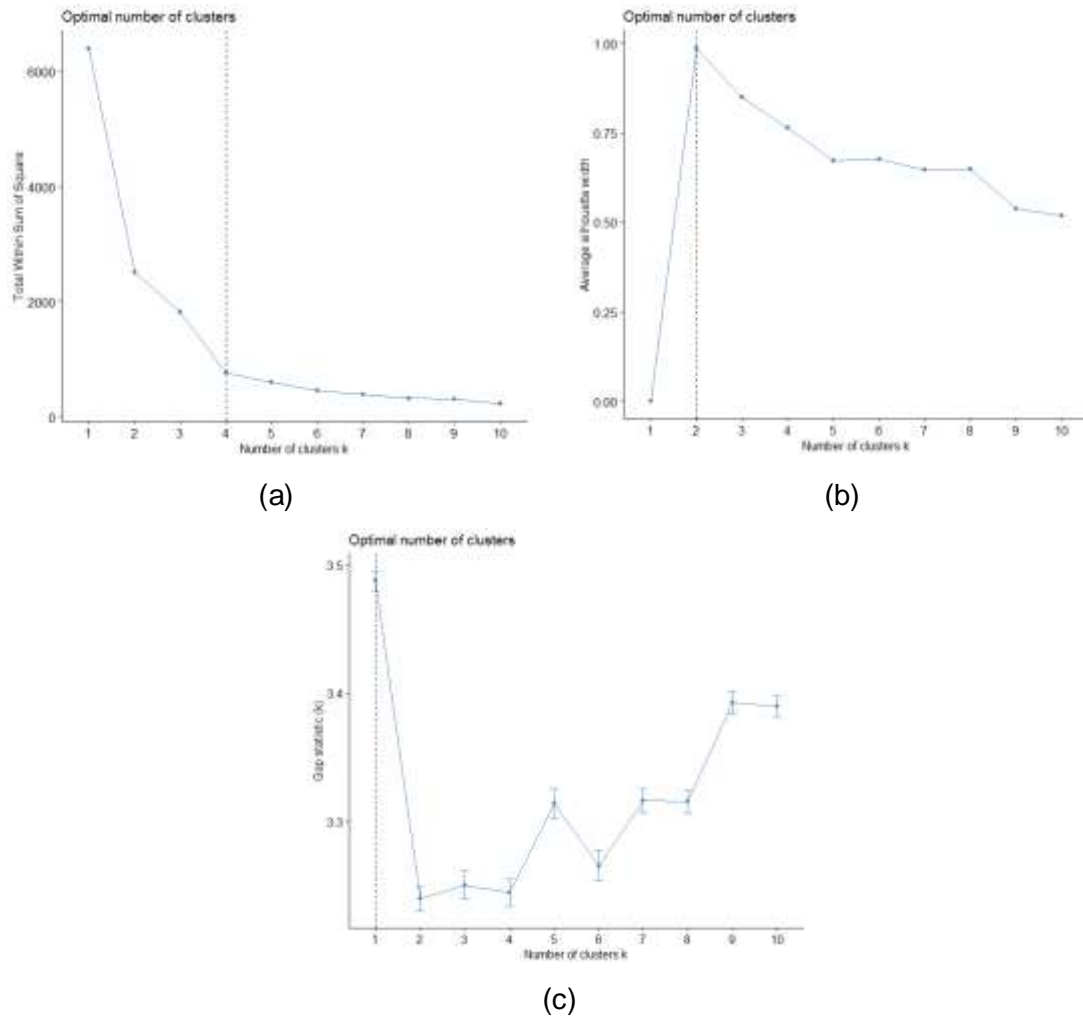
#### **4.1. Obtención del número de clústeres**

Se ha aplicado cada fase de la metodología y se han conseguido resultados relevantes al momento de la aplicación de los algoritmos K-medias, CLARA y DIANA al conjunto de datos; eligiendo un número de clústeres de forma específica. Esta elección ha sido relevante para comprender el comportamiento de agrupación de los datos y facilitar la toma de decisiones cuando sea necesario. A través de pruebas realizadas utilizando el método del codo, silueta y estadístico de la brecha (gap), se ha observado que el número óptimo de clústeres más recomendado es 4 debido a que representa la mayor cantidad de agrupaciones obtenidas para los algoritmos K-medias y CLARA.

Esta cantidad de clústeres representa el número de agrupaciones factibles realizadas por los métodos mencionados teniendo en cuenta que, realizar agrupaciones de 1 o 2 clústeres no resulta viable para la empresa debido a su sistema de manejo de inventarios. Al elegir 4 clústeres, podemos obtener una comprensión más completa del comportamiento de los datos en relación a los clústeres a los que pertenecen. En las Figuras 4 y 5 se muestra la obtención del número de clústeres mediante los métodos mencionados anteriormente y se puede notar que no existe una pérdida significativa de información.

**Figura 4**

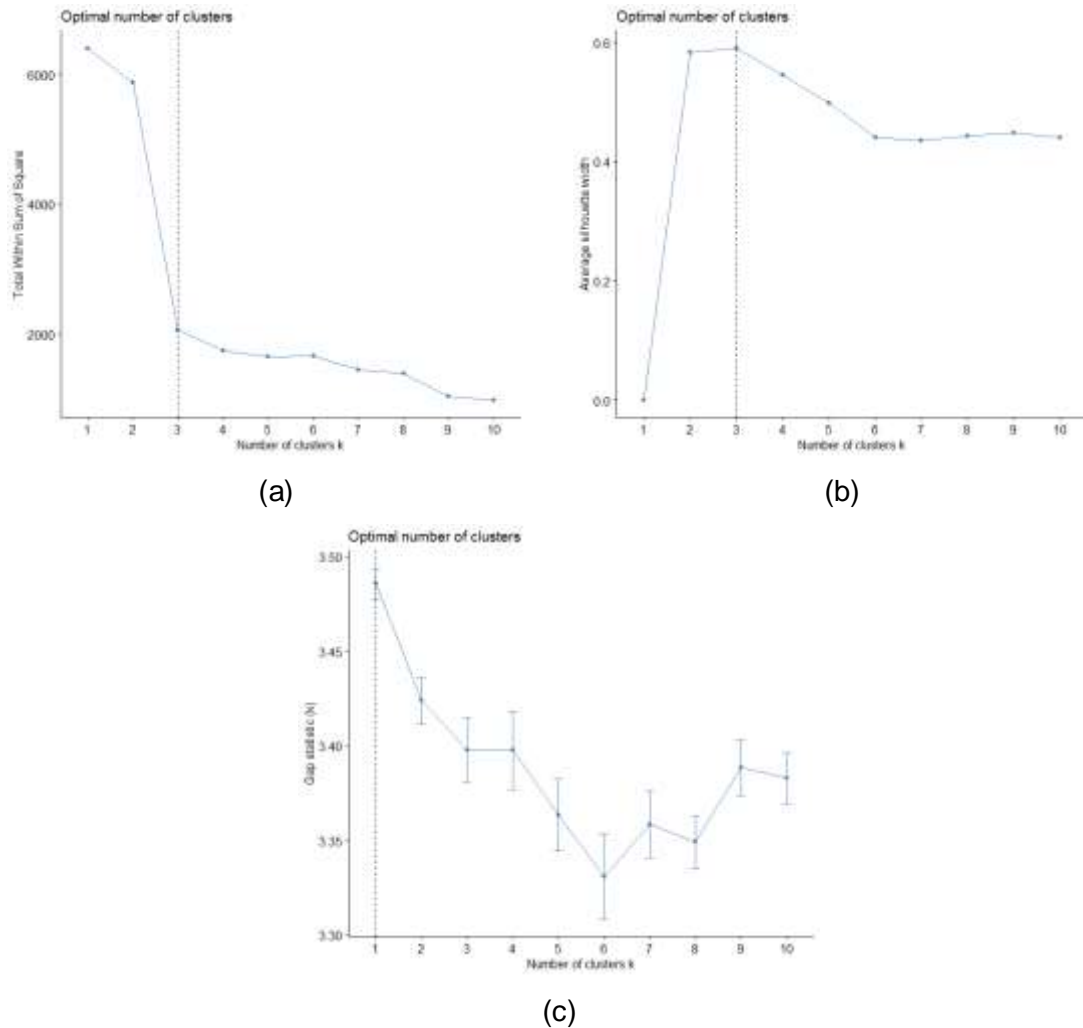
*Obtención del número de clústeres para el algoritmo K-medias*



*Nota:* Esta figura muestra la obtención del número de clústeres mediante: (a) método del codo, (b) método de la silueta y (c) método del estadístico gap; tomando en cuenta al método de agrupamiento K-medias.

**Figura 5**

*Obtención del número de clústeres para el algoritmo CLARA*



*Nota:* Esta figura muestra la obtención del número de clústeres mediante: (a) método del codo, (b) método de la silueta y (c) método del estadístico gap; tomando en cuenta al método de agrupamiento CLARA.

Para determinar el número óptimo de clústeres utilizando el algoritmo DIANA, es necesario ejecutar el algoritmo en sí, ya que no es posible obtener el número de agrupaciones óptimas utilizando un método de obtención aplicado a los algoritmos de clustering K-medias y CLARA. El algoritmo DIANA realiza una división de los datos y la elección de la cantidad de grupos queda a criterio del investigador que realiza el estudio. En este sentido, se puede realizar un corte en una sección específica del dendrograma que indique el número de agrupaciones más adecuado para el análisis.

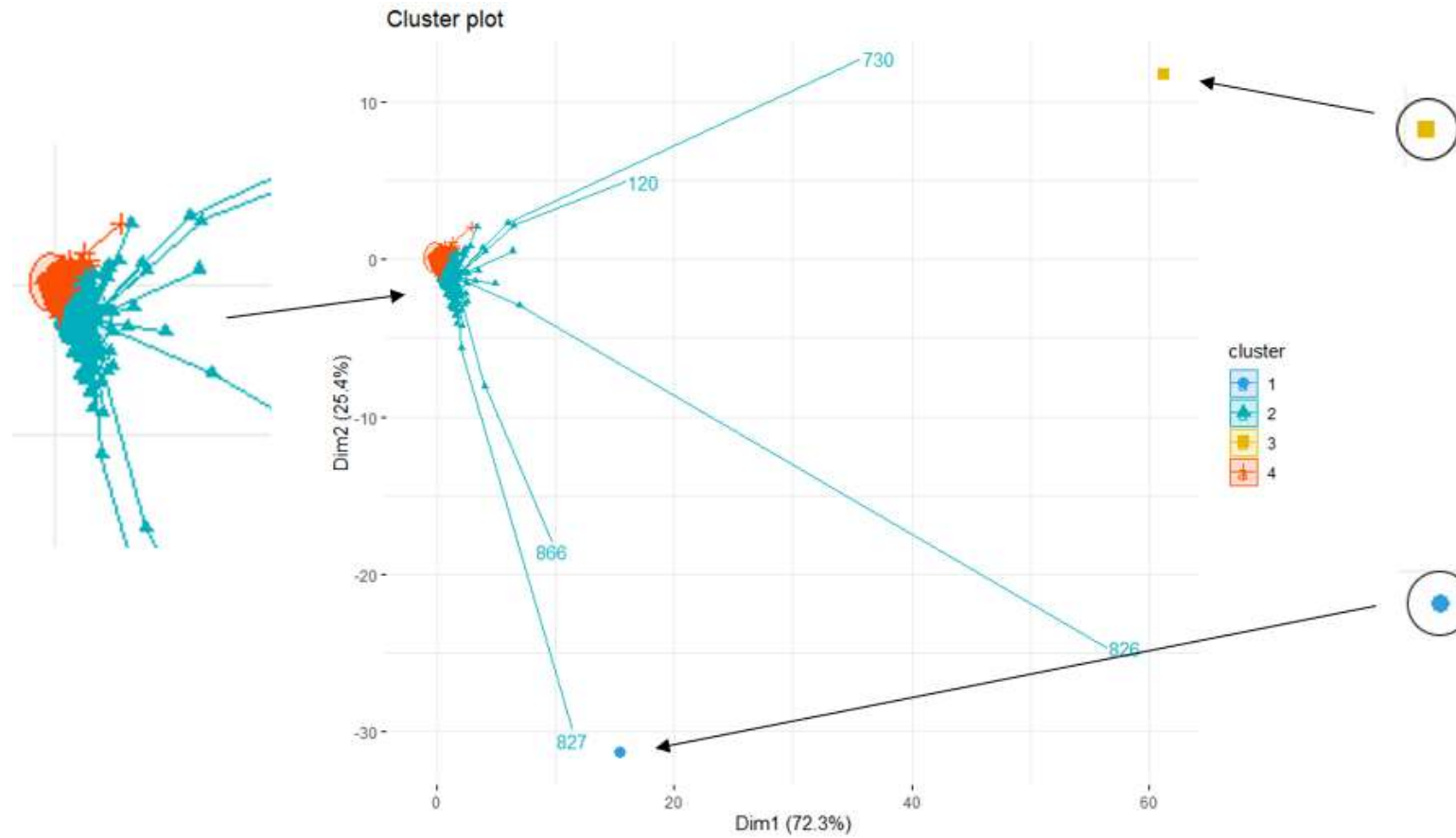


#### **4.2. Ejecución de los algoritmos K-medias, CLARA y DIANA con 4 clústeres**

De esta manera, se puede obtener una configuración óptima de clústeres para el caso de estudio en cuestión. Por lo tanto, se tomará el mismo número de clústeres en la aplicación de todos los algoritmos de particiones. Las gráficas para K-medias, CLARA y DIANA se muestran en las Figuras 6, 7 y 8 respectivamente.

Figura 6

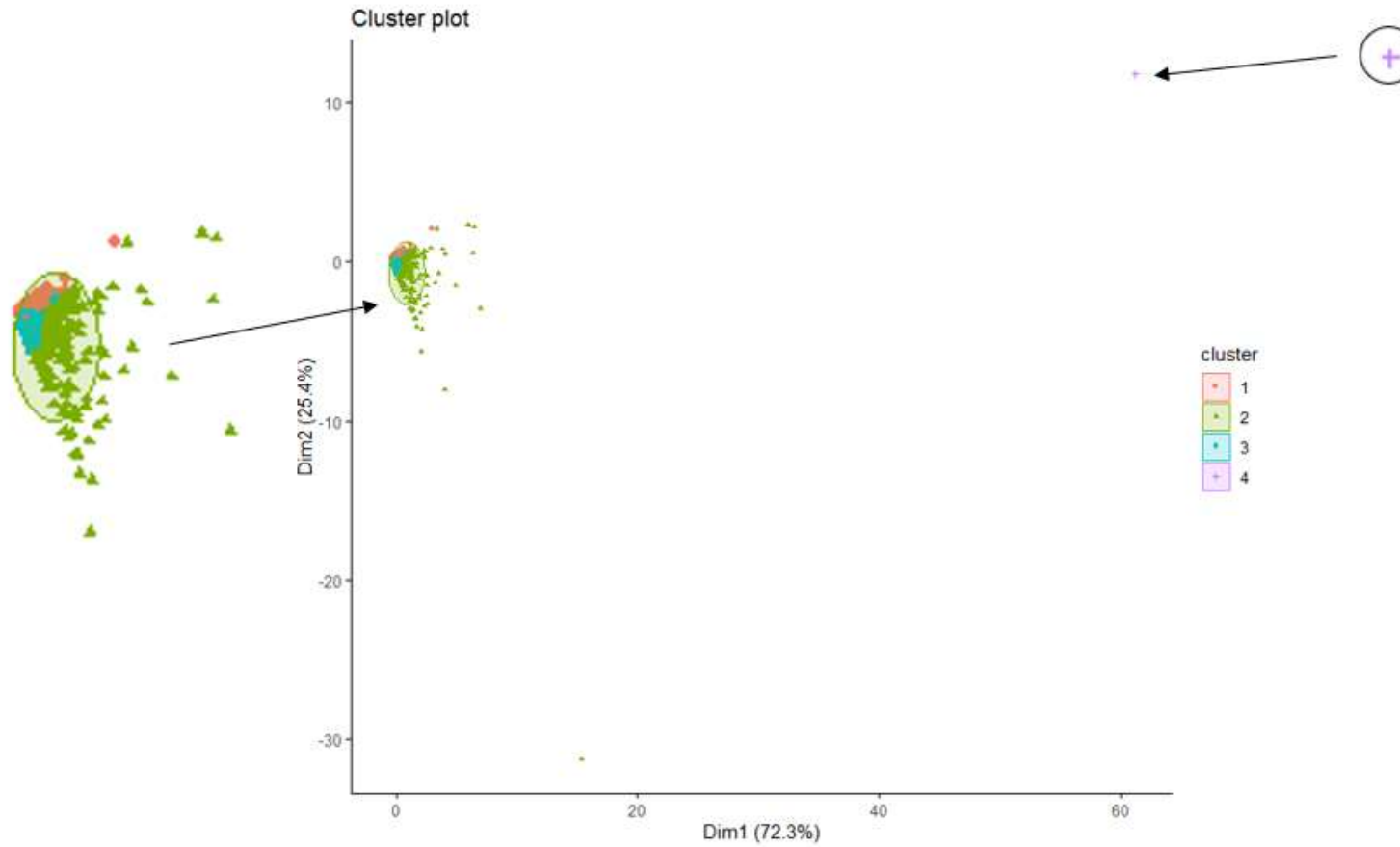
Agrupación de datos mediante el algoritmo K-medias (4 clústeres)



Nota: Esta figura muestra la agrupación de los datos mediante el algoritmo K-medias teniendo en cuenta 4 clústeres.

**Figura 7**

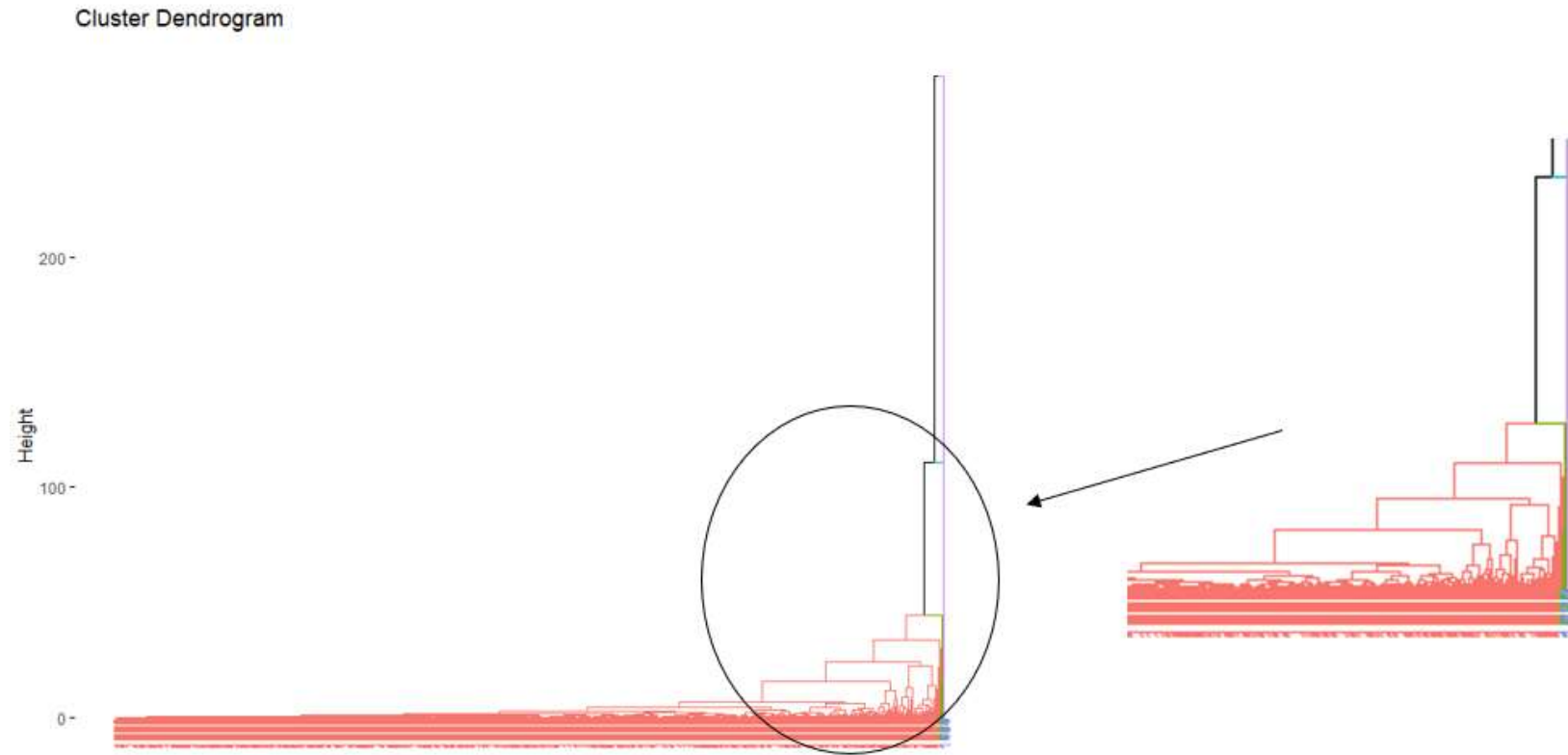
*Agrupación de datos mediante el algoritmo CLARA (4 clústeres)*



*Nota:* Esta figura muestra la agrupación de los datos mediante el algoritmo CLARA teniendo en cuenta 4 clústeres.

**Figura 8**

*Agrupación de datos mediante el algoritmo DIANA (4 clústeres)*



*Nota:* Esta figura muestra la agrupación de los datos mediante el algoritmo DIANA teniendo en cuenta 4 clústeres.

### 4.3. Análisis de los resultados obtenidos utilizando 4 clústeres

Al analizar las gráficas mostradas en las Figuras 6, 7 y 8 se evidencia la presencia de puntos de datos que se encuentran significativamente distantes de las agrupaciones más comunes, exhibiendo diferencias considerables en términos de distancia de los puntos al centro. Esta observación resulta de gran interés en el proceso de agrupación ya que mediante la comparación de los diferentes clústeres de cada algoritmo, se muestran estos datos atípicos. Sin embargo, estos datos fueron tomados en cuenta junto con los demás debido a que representan el comportamiento de consumo real en la empresa. Por lo tanto, mediante la exportación y visualización de los datos, es posible identificar y examinar estos ítems con el fin de comprender las causas que originan este comportamiento en los datos, lo cual tiene relevancia para los resultados obtenidos.

En cuanto a la comparación de los diferentes algoritmos, se puede realizar por el número de elementos distribuidos en cada clúster. Para realizar esto, mediante una tabla dinámica se llevó a cabo un conteo de los datos en función de los clústeres a los cuales están asignados.

Al examinar la Tabla 1, se observa la presencia de clústeres que contienen únicamente un solo elemento en su agrupación. Este comportamiento es relevante para la toma de decisiones, por lo que se procedió a identificar los datos responsables de esta situación con el fin de realizar análisis posteriores. A continuación, se detallan los ítems asignados a cada algoritmo:

**Tabla 1**

*Distribución de datos en los clústeres asignados a cada algoritmo*

Algoritmo	Clúster				Total general
	1	2	3	4	
CLARA	1084	196	321	1	1602
DIANA	1596	4	1	1	1602
K-medias	1	129	1	1471	1602

- Con el algoritmo CLARA, se ha asignado un solo dato al clúster 4.
- Con el algoritmo DIANA, tanto el clúster 3 como el clúster 4 tienen asignado un solo dato en cada uno.

- Con el algoritmo K-medias, tanto el clúster 1 como el clúster 3 han asignado un solo dato en cada uno.

Estos hallazgos proporcionan información valiosa sobre la distribución de los datos y podrían influir en la interpretación de los resultados obtenidos mediante los algoritmos utilizados. Los ítems asignados como único dato dentro de un clúster se presentan de forma recopilada en la Tabla 2.

**Tabla 2**

*Ítems asignados a clústeres como único elemento*

Número	Clúster	Número de parte	Algoritmo
1494	4	KIT-EMBRA-003	CLARA
1494	3	KIT-EMBRA-003	DIANA
1589	4	TOLVA	DIANA
1589	1	TOLVA	K-medias
1494	3	KIT-EMBRA-003	K-medias

Al examinar la Tabla 2, se ha llevado a cabo un análisis exhaustivo para determinar el motivo por el cual ciertos datos han sido asignados a un único clúster. A partir de dicho análisis, se han obtenido los siguientes resultados:

- El ítem denominado "KIT-EMBRA-003" ha registrado el mayor número de ventas y, por ende, ha generado los mayores ingresos (63.120,40 USD) en comparación con los demás ítems de otros clústeres, mostrando una diferencia significativa. Esta disparidad en términos de ingresos es la razón por la cual este ítem ha sido separado y asignado a un único clúster en los tres algoritmos aplicados al conjunto de datos.
- El ítem "TOLVA" se destaca como el segundo repuesto que ha generado mayores ingresos (15.000,00 USD), además de ser el que presenta el costo unitario más elevado (13.392,86 USD), mostrando una diferencia significativa con respecto a los demás ítems de otros clústeres. Por esta razón, este ítem se encuentra asignado a un único clúster en dos de los tres algoritmos aplicados al conjunto de datos.

Estos hallazgos son fundamentales para comprender el comportamiento de los datos y su asignación a clústeres específicos, proporcionando una visión más profunda sobre la importancia de ciertos ítems, en términos de ventas y generación de ingresos.

#### 4.4. Elección óptima y definitiva del número de clústeres

Debido a este análisis en los datos asignados se ha tomado la decisión de excluir los ítems KIT-EMBRA-003 y TOLVA del conjunto de datos ya que, su notoriedad de diferenciación es bastante obvia al ser los dos repuestos con mayores ingresos en las ventas. Por lo tanto, estos dos ítems se han tomado en cuenta en clústeres únicos. Por lo tanto, los resultados de agrupación que se han realizado a continuación se han modificado sin tomar en cuenta los productos descritos y se consideró realizar una nueva clasificación del inventario.

Con el fin de obtener una distribución óptima de los ítems y facilitar la elección del algoritmo apropiado para la clasificación del inventario, se procedió a realizar un nuevo análisis de los datos. Esta nueva fase del análisis permite obtener una mejor distribución de los ítems restantes y, en consecuencia, determinar el algoritmo más adecuado para la clasificación precisa dentro del inventario.

Previo a la ejecución de los algoritmos sobre el conjunto de datos, se llevó a cabo un procedimiento para determinar el número de clústeres apropiado según el algoritmo de clasificación. En este sentido, se han aplicado los métodos del codo (wss), la silueta (silhouette) y el estadístico gap (stat\_gap) para los algoritmos K-medias y CLARA. Los resultados se pueden visualizar en la Tabla 3 para el algoritmo K-medias y CLARA respectivamente.

**Tabla 3**

*Obtención del número de clústeres para el algoritmo K-medias y CLARA*

Algoritmo	Método	Número de clústeres
K-medias	Codo	6
	Silueta	2
	Estadístico gap	2
CLARA	Codo	6
	Silueta	2
	Estadístico gap	2

*Nota:* Esta tabla muestra la obtención del número de clústeres mediante: método del codo, método de la silueta y método del estadístico gap tomando en cuenta los métodos de agrupamiento K-medias y CLARA con la decisión de no tomar en consideración los datos mencionados en la Tabla 2.

- Mediante el método del codo, que busca optimizar la suma de los cuadrados de las distancias (wss), se obtuvo un resultado de 6 clústeres.
- Mediante el método de la silueta, que selecciona el número de clústeres en base al valor máximo del ancho medio entre grupos, se obtuvo un resultado de 2 clústeres.
- Mediante el método del estadístico de gap, que compara la varianza intra-clúster en diferentes valores de clústeres con un valor esperado basado en una distribución uniforme de referencia, se obtuvo también un resultado de 2 clústeres.

Debido a que la mayoría de los inventarios requieren una clasificación eficiente para lograr una distribución rápida y efectiva, muchos sistemas de clasificación se basan en el popular sistema de clasificación ABC. Este sistema identifica los productos que generan los mejores resultados económicos. Para optimizar la gestión de la clasificación de inventarios, se utiliza una clasificación en tres segmentos (A, B y C), lo que implica la elección óptima de 3 grupos de datos que presenten características similares.

Aunque el número de clústeres puede variar según el algoritmo de clasificación utilizado, en nuestro caso, la elección de 3 clústeres es la más adecuada para satisfacer los intereses recomendados por la empresa del estudio. Por lo tanto, se evaluó la clasificación de los datos bajo la condición de 3 segmentos de clasificación, con esto se puede tener una clasificación más acertada a lo que la empresa de estudio realiza actualmente.

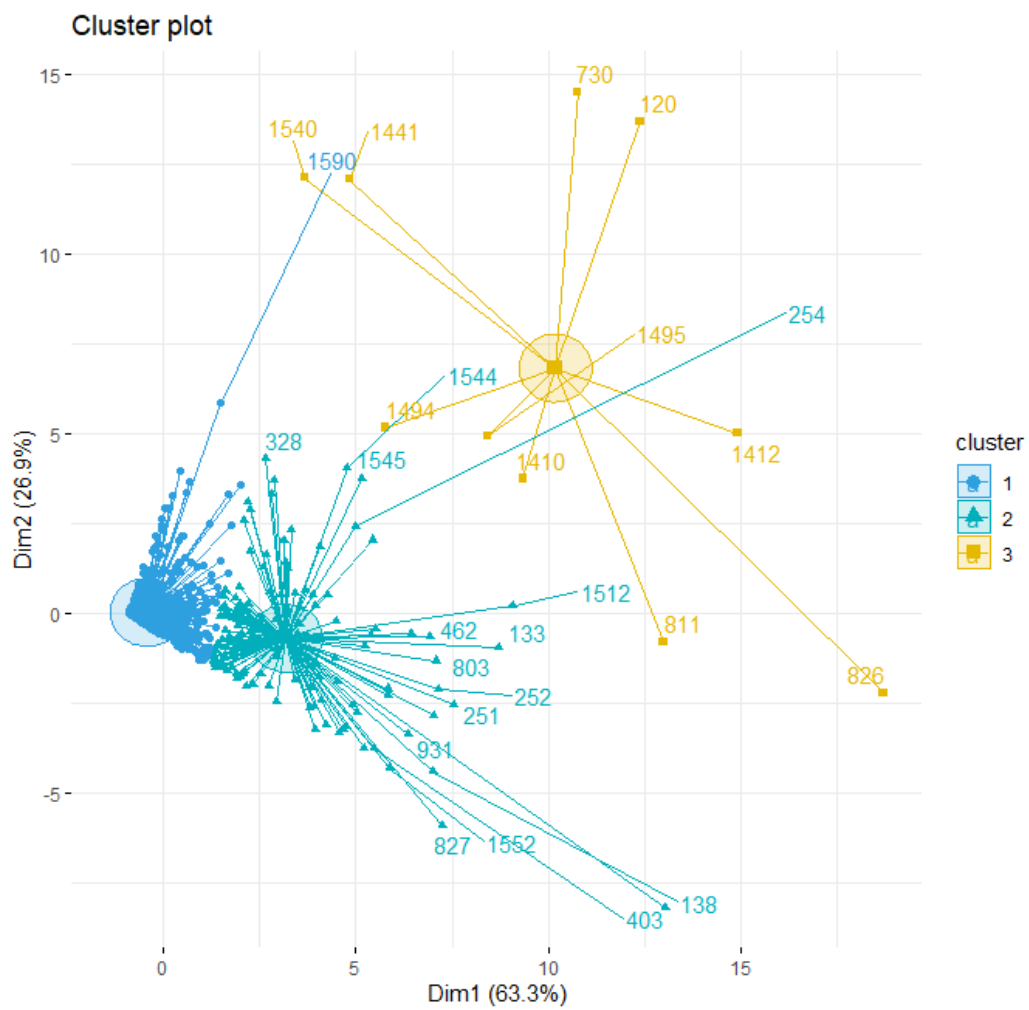
#### **4.5. Ejecución de los algoritmos K-medias, CLARA y DIANA con 3 clústeres**

La agrupación del conjunto de datos mediante los tres algoritmos y con el número predefinido de clústeres, ha arrojado los siguientes resultados:



**Figura 9**

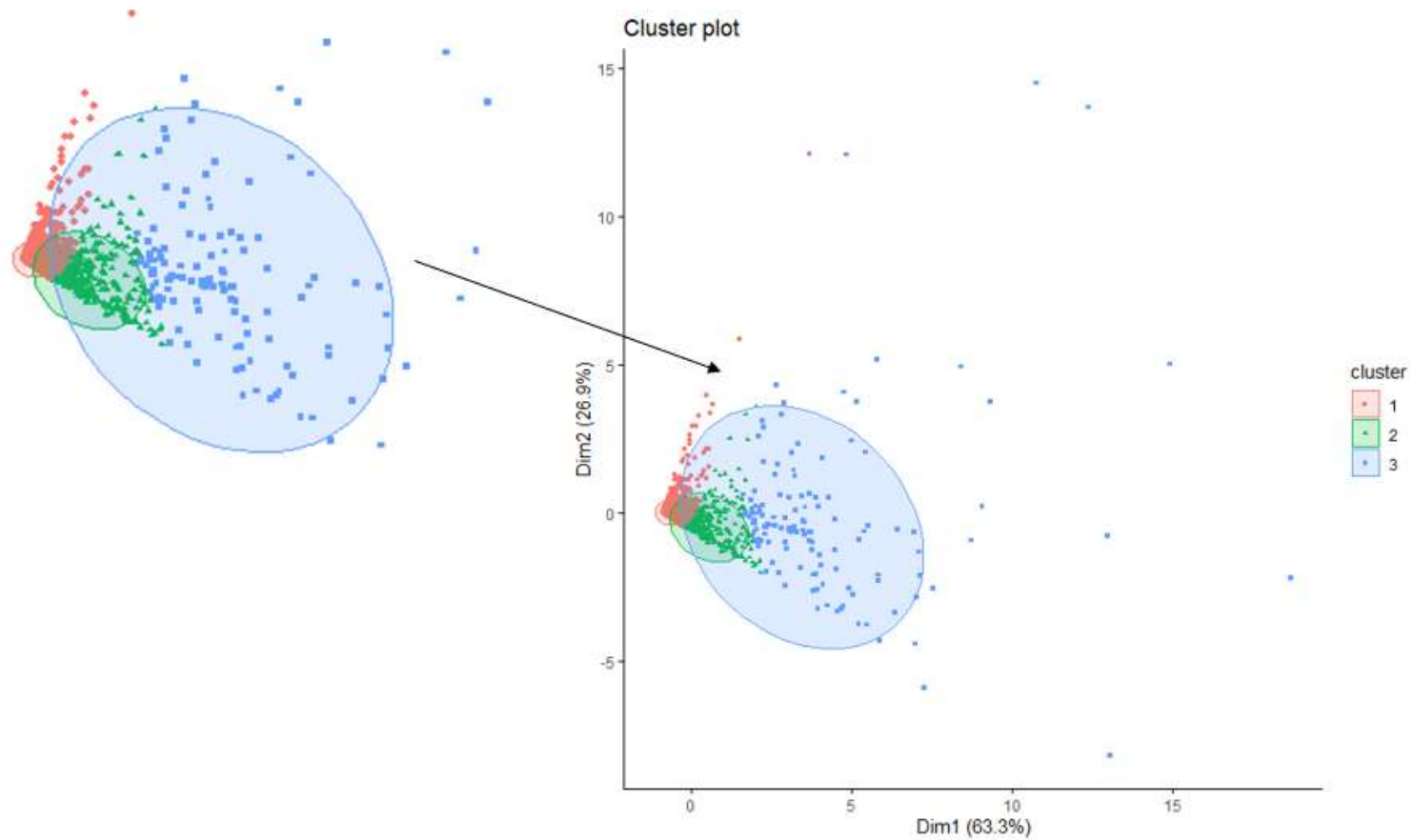
*Agrupación de datos mediante el algoritmo K-medias (3 clústeres)*



*Nota:* Esta figura muestra la agrupación de los datos mediante el algoritmo K-medias teniendo en cuenta 3 clústeres.

**Figura 10**

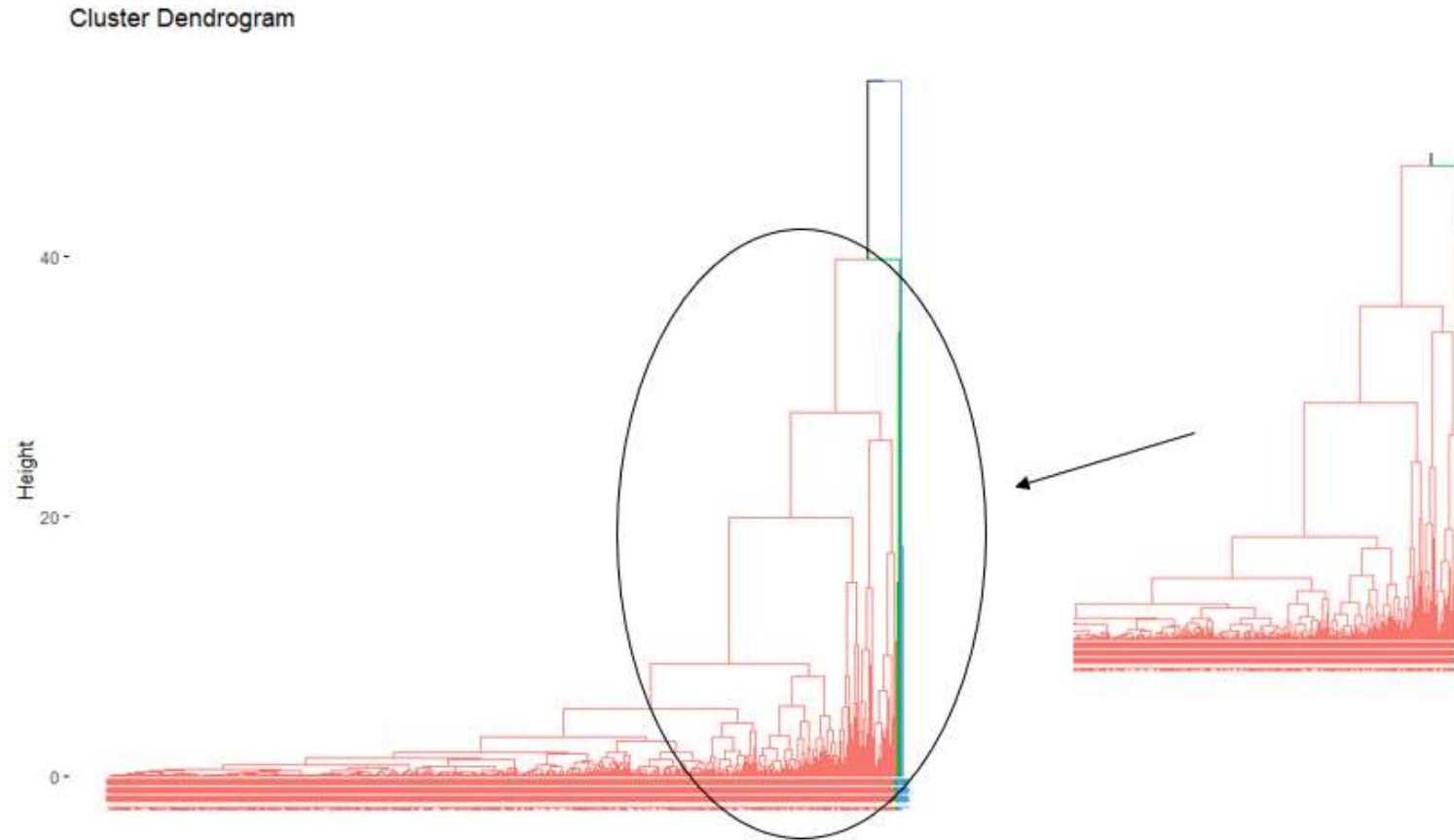
*Agrupación de datos mediante el algoritmo CLARA (3 clústeres)*



*Nota:* Esta figura muestra la agrupación de los datos mediante el algoritmo CLARA teniendo en cuenta 3 clústeres.

**Figura 11**

*Agrupación de datos mediante el algoritmo DIANA (3 clústeres)*



*Nota:* Esta figura muestra la agrupación de los datos mediante el algoritmo DIANA teniendo en cuenta 3 clústeres.

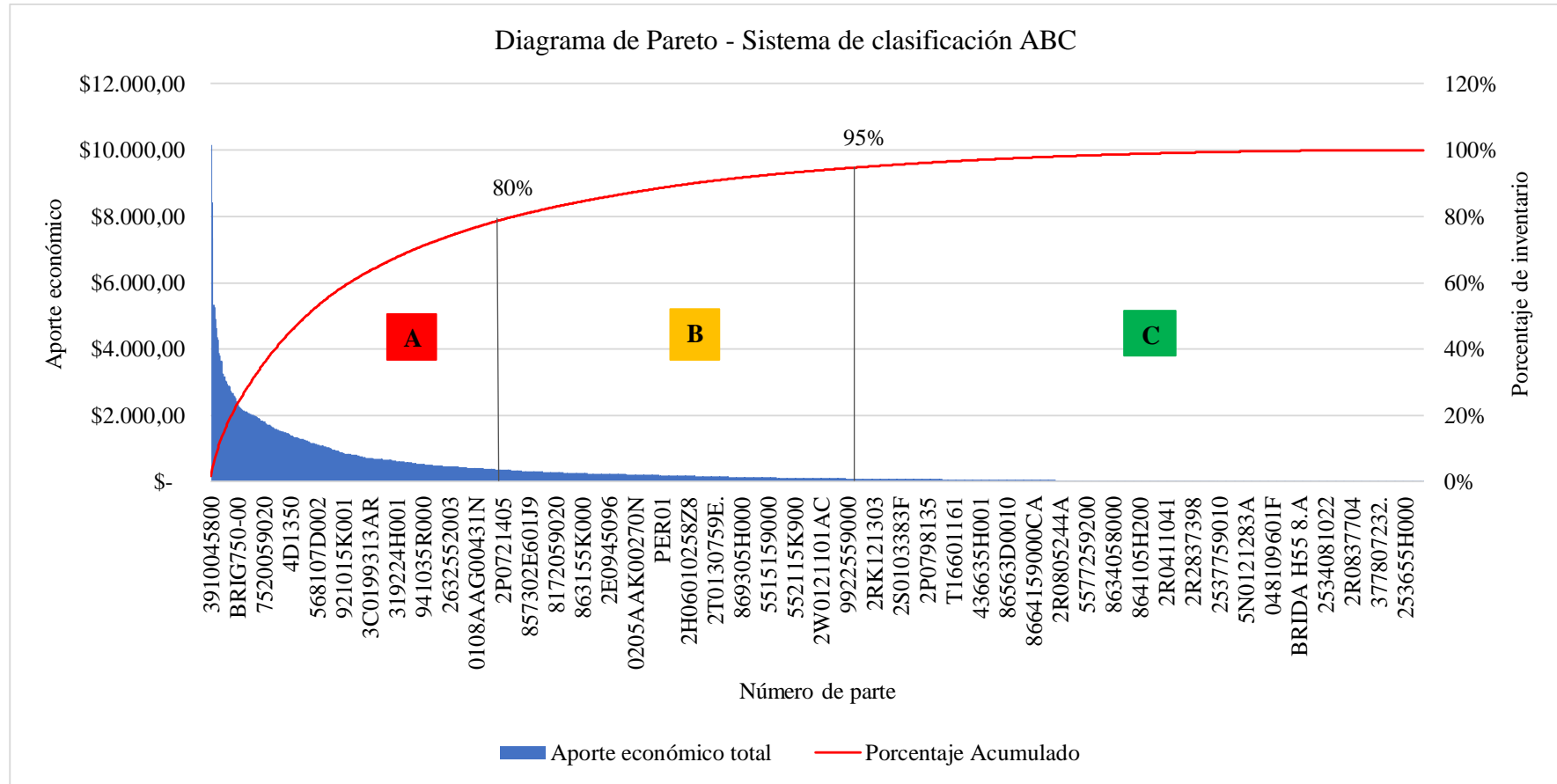
Como se puede observar en las Figuras 9 y 10, las agrupaciones son más apreciables en comparación con las Figuras 6 y 7 en las gráficas de dispersión. Esto se debe a la decisión de excluir dos ítems que causaban una segmentación inadecuada de los datos. Es importante destacar que aunque estos ítems resultaron en una segmentación inapropiada, no se consideran datos atípicos, ya que tienen una mayor relevancia en comparación con los demás ítems. Sin embargo, se determinó que su inclusión afectaba negativamente la calidad de la segmentación, por lo que se tomaron medidas para evitar este efecto indeseado.

#### **4.6. Análisis comparativo utilizando el Sistema de clasificación ABC**

Luego de obtener la clasificación mediante clústeres, se procedió a contrastarla con el Sistema de Clasificación ABC. Para realizar este análisis, se utilizó el Diagrama de Pareto, como se muestra en la Figura 12. A través de este análisis, se identificaron 400 ítems que se encuentran distribuidos en el segmento de datos tipo A (aporte económico alto), 466 en el segmento tipo B (aporte económico medio) y 734 ítems en el segmento tipo C (aporte económico bajo).

Figura 12

Diagrama de Pareto - Sistema de clasificación ABC



El diagrama de Pareto, representado en la Figura 12, es utilizado para identificar las secciones que representan las categorías ABC dentro del sistema de inventario. La Tabla 4 muestra la cantidad de repuestos correspondiente a cada segmento, según el análisis de clasificación ABC.

**Tabla 4**

*Resultados obtenidos mediante Clasificación ABC*

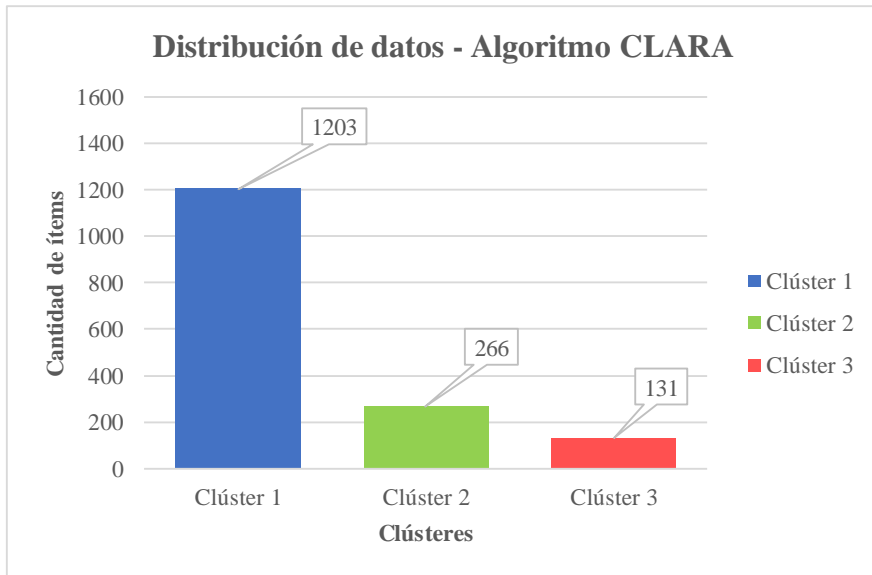
Porcentaje Pareto	Zona	Cantidad	Cantidad acumulada	Porcentaje Cantidad Acumulada	Porcentaje Inventario	Porcentaje Inventario Acumulado
0 - 80	A	400	400	25%	79,96%	79,96%
80 -95	B	466	866	54%	15,04%	95,00%
95 -100	C	734	1600	100%	5,00%	100,00%

*Nota:* muestra los resultados obtenidos mediante la clasificación ABC, donde se presentan la cantidad de ítems distribuidos en cada zona y sus respectivos porcentajes de inventario.

Mediante la exportación de datos se pudo obtener la cantidad de repuestos distribuidos en cada uno de los clústeres según el algoritmo utilizado para la segmentación de los ítems. Las Figuras 13, 14 y 15 presentan de manera clara los datos recopilados en los tres clústeres elegidos para los tres algoritmos de segmentación.

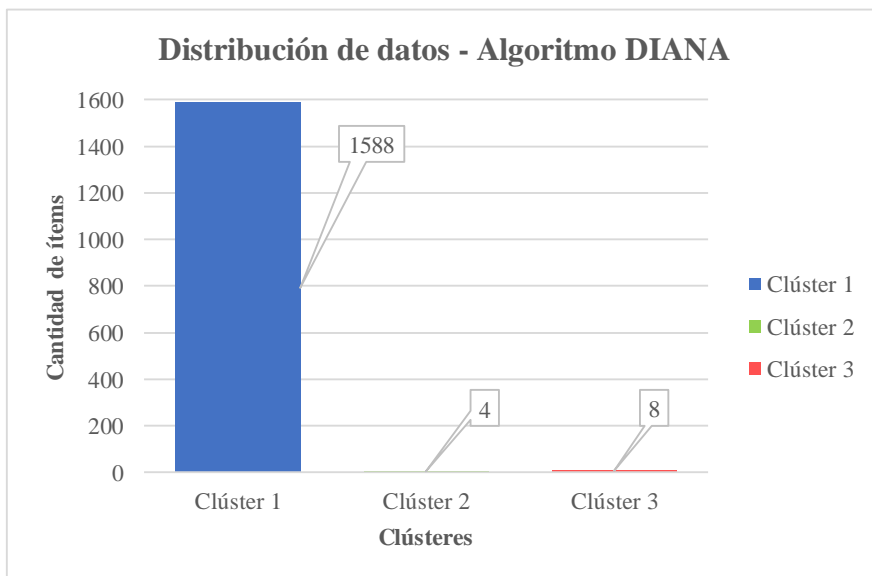
**Figura 13**

*Distribución de datos mediante el algoritmo CLARA*



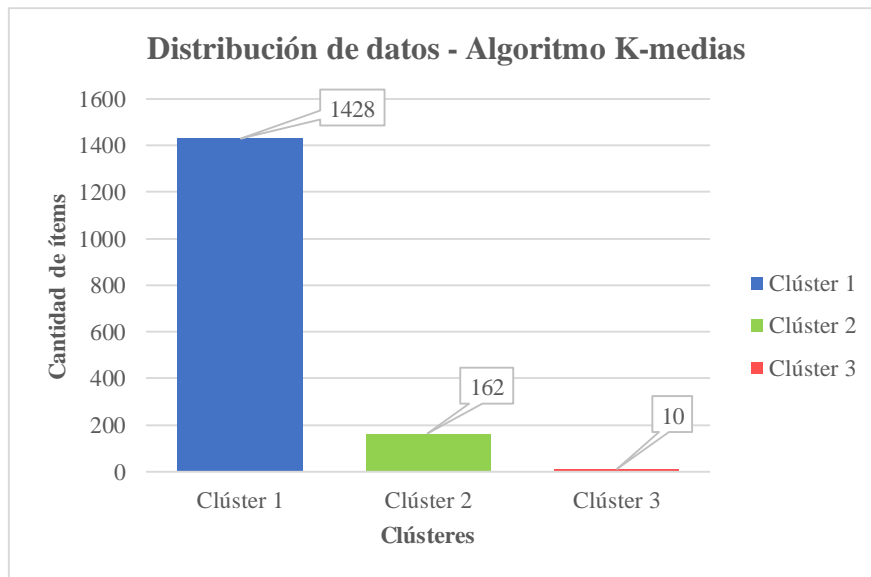
**Figura 14**

*Distribución de datos mediante el algoritmo DIANA*



**Figura 15**

*Distribución de datos mediante el algoritmo K-medias*



Al realizar una comparación entre la distribución de los ítems del inventario obtenida mediante los algoritmos y el sistema de clasificación ABC, se puede observar la relación entre los datos. La Tabla 5 muestra la distribución de los datos, en donde se destacan las siguientes observaciones:

- La mayoría de los ítems se encuentran en el clúster 1, incluyendo todos los datos clasificados en C para los algoritmos CLARA, DIANA y K-medias.
- El algoritmo CLARA cuenta con la mayoría de los ítems clasificados en B
- Los algoritmos DIANA y K-medias tienen todos los ítems en la categoría B. Sin embargo, la distribución de los ítems clasificados en A varía para los tres algoritmos.



**Tabla 5***Distribución de datos en los clústeres de cada algoritmo y segmento ABC*

<b>Algoritmo y ABC</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>Total general</b>
<b>CLARA</b>	<b>1203</b>	<b>266</b>	<b>131</b>	<b>1600</b>
A	60	209	131	400
B	409	57		466
C	734			734
<b>DIANA</b>	<b>1588</b>	<b>4</b>	<b>8</b>	<b>1600</b>
A	388	4	8	400
B	466			466
C	734			734
<b>K-medias</b>	<b>1428</b>	<b>162</b>	<b>10</b>	<b>1600</b>
A	228	162	10	400
B	466			466
C	734			734

*Nota:* Esta tabla muestra la distribución de los datos de cada clúster en los segmentos de la clasificación ABC según el algoritmo empleado.

#### **4.7. Selección del algoritmo óptimo para la clasificación adecuada de ítems**

Los resultados obtenidos demuestran que la distribución más acertada se logra con el algoritmo CLARA. Dado que la empresa tiene intereses específicos, es importante evitar una categoría con muy pocos datos. Además, la elección de la agrupación se basa en el estudio realizado por Amat (2017), donde se explica que el algoritmo CLARA es preferible en términos de robustez para la inicialización de centroides. CLARA utiliza muestras aleatorias de un tamaño determinado en las que se ejecuta el algoritmo PAM (K-medoides) para obtener una agrupación óptima.

Este proceso se repite hasta que los clústeres finales tengan una menor suma total de distancias. Esta metodología es menos afectada por la presencia de puntos de datos que difieren considerablemente del patrón general de los demás; esto aumenta la probabilidad de obtener una solución robusta y reduce el sesgo de muestreo en comparación con DIANA o K-medias.

Es necesario tener en cuenta lo mencionado en el artículo de Saraf et al. (2022), donde se destaca la elección de algoritmos de clustering no jerárquico debido a su cálculo más rápido

y con clústeres más compactos. La segmentación realizada por el algoritmo CLARA se ajusta a las condiciones mencionadas y a recomendaciones basadas en estudios previos, por lo tanto es la opción más adecuada para la empresa en comparación con las segmentaciones realizadas por los algoritmos DIANA y K-medias

La propuesta de clasificación de inventarios para el caso de estudio de la empresa del sector automotriz es como ya se había mencionado, la realizada por el algoritmo CLARA, de la que podemos destacar la cantidad de ítems pertenecientes a cada uno de los clústeres asignados. Estos clústeres que, comparados con la clasificación ABC hacen posible relacionar las características de aporte económico con las características correspondientes al comportamiento de compra de los ítems.

En la Tabla 6 se muestra cómo se distribuyeron los datos en cada clúster según la marca de los ítems. En estas agrupaciones se observó que la cantidad de datos ubicados en cada uno de los clústeres es directamente proporcional a la cantidad de datos de cada una de las marcas; es decir que, para el clúster 1 habrá una mayor cantidad de datos para cada marca en relación con el total de datos pertenecientes a la misma.

**Tabla 6**

*Distribución de datos en los clústeres del algoritmo CLARA según las marcas de vehículos*

<b>Algoritmo y Marcas</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>Total general</b>
<b>CLARA</b>	<b>1203</b>	<b>266</b>	<b>131</b>	<b>1600</b>
Marca 1	644	135	87	866
Marca 2	97	10	3	110
Marca 3	10	6	2	18
Marca 4	1			1
Marca 5	451	115	39	605

*Nota:* Esta tabla muestra la distribución de los datos de cada clúster en los segmentos de la clasificación ABC según el algoritmo empleado.

## 5. Discusión

La elección para la clasificación del inventario está basada en el algoritmo CLARA debido a los intereses particulares de la empresa por tener una agrupación más distribuida de los repuestos; los cuales con su correcta aplicación ayudaría a conseguir buenos resultados. Además, esta idea de clasificación mediante CLARA puede ser respaldada por el estudio realizado por Gupta & Panda (2019), en el cual se concluye, a partir de una base de datos establecida, que el algoritmo CLARA ofrece mejores resultados en la agrupación mediante clústeres en comparación con los algoritmos K-medias y K-medoides ya que considera datos de comportamiento de sistemas. Esto es especialmente beneficioso al gestionar grandes conjuntos de datos, lo cual es común en el ámbito de la gestión de inventarios.

Existen investigaciones, como el estudio realizado por Lolli et al. (2019), que proponen el uso de técnicas de agrupación supervisada, donde se simulan exhaustivamente escenarios a nivel de un solo artículo, extrayendo una muestra de la población de elementos y determinando las mejores políticas de reordenamiento para cada artículo en esa muestra. Sin embargo, a diferencia con esta investigación, se ha decidido considerar toda la población de artículos vendidos al utilizar técnicas de agrupación no supervisada aplicando algoritmos. En este enfoque, cada tipo de artículo se trata de forma independiente y se clasifica en grupos según las variables asignadas, teniendo en cuenta los intereses específicos de la organización.

De esta manera, al considerar todo el conjunto de datos, se obtiene una visión completa del comportamiento de ventas de cada ítem, lo que permite aplicar estrategias de reordenamiento en intervalos de tiempo específicos. Además, esta metodología ofrece la flexibilidad de introducir o dar de baja ciertos ítems según sea necesario, asegurando que la gestión de inventario esté constantemente ajustada a las necesidades de la organización.

Como se había mencionado previamente, la empresa analizada en el caso de estudio tiene implementado una clasificación ABC en su sistema de inventarios; esto resulta relevante al contar con ítems representativos como KIT-EMBRA-003 y TOLVA debido a su alto aporte económico representado en ventas, ya que estos modifican el proceso de segmentación al usar dicho sistema. Sin embargo, el uso del sistema tradicional puede llevar a descuidar otros ítems igualmente importantes para las ventas. Por lo tanto, es imperativo contar con un nuevo sistema de clasificación que garantice que la importancia de estos ítems no sea pasada por alto.

Teniendo en cuenta la distribución de repuestos en cada una de las agrupaciones correspondientes y el análisis mediante la clasificación ABC teniendo en cuenta los segmentos A (aporte económico alto), B (aporte económico medio) y C (aporte económico bajo) se resaltan los siguientes grupos:

Al considerar la distribución de repuestos en las respectivas agrupaciones y el análisis basado en la clasificación ABC, que divide los segmentos que se manejan en el caso de estudio como A (alto aporte económico), B (aporte económico medio) y C (aporte económico bajo), se destacan los siguientes grupos:

- Agrupación del clúster 1: Agrupación formada por 1203 ítems de repuestos, en los que mediante el análisis de clasificación ABC enfocado en el aporte económico, este grupo contiene una mezcla de todos los ítems correspondientes al segmento C y la mayoría del segmento B. En esta agrupación de datos existen la mayoría de los ítems de cada una de las distintas marcas de repuestos vendidos por la empresa.
- Agrupación del clúster 2: Agrupación formada por 266 ítems de repuestos, en los que mediante el análisis de clasificación ABC enfocado en el aporte económico, en este grupo existe una mezcla se una mayoría de ítems del segmento A y unos cuantos más del segmento B. En esta agrupación de datos existe una cantidad considerable de ítems vendidos por la empresa para cada una de las distintas marcas de repuestos en relación con el total de aquellos ítems de cada marca.
- Agrupación del clúster 3: Agrupación formada por 131 ítems de repuestos, en los que mediante el análisis de clasificación ABC enfocado en el aporte económico, todos estos forman parte de forma directa del segmento A, correspondiente al grupo de mayores resultados económicos. En esta agrupación existe la menor cantidad de ítems vendidos en la empresa por cada una de las distintas marcas.

De estas agrupaciones destacan ciertos datos para cada marca, los cuales representan a ítems específicos y se muestran en la Tabla 7.

**Tabla 7**

*Ítems más destacados en cada agrupación y para cada marca de repuestos*

<b>Marca</b>	<b>Número de parte</b>	<b>Total</b>
<b>Agrupación 1</b>		
Marca 1	2452084002	613,07 USD
Marca 2	1001CAA14881N	314,24 USD
Marca 3	81437220070 L	477,84 USD
Marca 4	A6111800009	15,01 USD
Marca 5	07W105401A	497,28 USD
<b>Agrupación 2</b>		
Marca 1	5810159A00	1.305,32 USD
Marca 2	0102EAZ00440NA	731,98 USD
Marca 3	81437026079	703,16 USD
Marca 5	2H6853601A DPJ	1.173,68 USD
<b>Agrupación 3</b>		
Marca 1	3910045800	10.147,74 USD
Marca 2	0305BM0061N	2.012,26 USD
Marca 3	2V5201512	4.274,26 USD
Marca 5	07W115436C	5.327,17 USD

*Nota:* Esta tabla muestra los ítems más destacados de cada una de las agrupaciones enfocado en términos de venta totales junto a sus valores.

En relación con los resultados obtenidos, la clasificación de repuestos para la empresa del sector automotriz depende en gran medida de cómo se asignan los recursos para desarrollar el sistema de inventarios. Es importante tener en cuenta que este estudio se ve limitado por factores externos como la demanda de los clientes, estrategias de marketing y otros aspectos que influyen en el desempeño del inventario. Sin embargo, se garantiza que mantener un sistema de inventarios organizado facilitará la disposición de repuestos en ubicaciones establecidas, lo cual mejorará la gestión del inventario.

En cuanto a la elección del número óptimo de agrupaciones dentro del inventario, el artículo de Gustriansyah et al. (2020) propone seleccionar dicho número mediante ocho índices de validez y un análisis comparativo estadístico. Sin embargo, en contraste con dicho enfoque, este estudio se enfoca en determinar las agrupaciones según los intereses específicos de la empresa, en lugar de depender únicamente de pruebas estadísticas para obtener

agrupaciones ideales. Esto se debe a que cada organización cuenta con condiciones particulares que influyen en las decisiones de distribución de inventarios. Sin embargo, los resultados de agrupación obtenidos a través de los índices del codo, la silueta y el estadístico gap fueron de 6, 2 y 2 clústeres respectivamente. Estas métricas indicaron los grupos que podrían considerarse estadísticamente ideales. No obstante, es importante tener en cuenta que estas métricas dependen del algoritmo utilizado y pueden no ajustarse a los intereses específicos de la empresa.

El artículo de Anitha & Patil (2019) aborda el uso del machine learning y algoritmos aplicados a bases de datos de ventas para obtener información valiosa sobre el comportamiento de compra de los consumidores. En relación con la presente investigación, se destaca que el comportamiento de compra de los consumidores está estrechamente relacionado con la distribución de los ítems dentro del inventario, ya que es necesario que los repuestos estén disponibles para su adquisición al momento de realizar una venta. Por lo tanto, la aplicación de técnicas de machine learning mediante algoritmos de segmentación permite una distribución más eficiente de los ítems dentro del inventario de una empresa del sector automotriz que se dedique a la venta de repuestos y que esto permita conocer el comportamiento de rotación de los mismos.

La comparación entre la agrupación realizada mediante algoritmos de machine learning y la clasificación tradicional más popular, el método ABC, ha sido objeto de estudio en artículos como el realizado por Khanorkar y Kane (2023). En dicho estudio se analizan el algoritmo K-medias, la clasificación ABC y las técnicas de toma de decisiones multicriterio para agrupar elementos en un inventario. Los resultados resaltan el rendimiento sobresaliente del algoritmo de clasificación en comparación con las otras dos técnicas mencionadas. En esta investigación, se comparte la misma perspectiva, dado que el uso de algoritmos de clasificación facilita el manejo de grandes conjuntos de datos y evita la necesidad de realizar cálculos específicos en las técnicas de toma de decisiones multicriterio. Por lo tanto, se ha decidido centrarse exclusivamente en la aplicación de algoritmos de segmentación de machine learning y realizar una comparación con la clasificación ABC en relación a ciertos parámetros.

Por lo expuesto en párrafos anteriores, el enfoque de utilizar algoritmos de machine learning para la clasificación de inventarios permite aprovechar la capacidad de estos algoritmos para procesar eficientemente grandes volúmenes de datos. Además, el uso de algoritmos de machine learning brinda la oportunidad de lograr una clasificación más precisa y automatizada, lo que a su vez facilita la toma de decisiones relacionadas con las políticas

de almacenamiento, reabastecimiento y control de inventarios, tal como lo indica el artículo de investigación de Raja et al. (2016). Esta perspectiva se relaciona directamente con la presente investigación, ya que los ítems se agruparon en función de sus características específicas, formando grupos con políticas de inventario propias.

## 6. Conclusiones

El objetivo de este estudio de investigación consistió en determinar la clasificación óptima de los ítems dentro del sistema de inventario de repuestos utilizando algoritmos de machine learning, con el propósito de mejorar la clasificación de artículos en empresas del sector automotriz. Para lograr esto, se utilizó una base de datos que contenía información de ventas para cada ítem, la cual fue preparada previamente para la ejecución de los algoritmos K-medias, CLARA y DIANA. Obtenidos los resultados independientemente de cada algoritmo, se procedió a comparar los mismos y encontrar una segmentación adecuada en tres grupos o clústeres considerados ideales por la empresa que proporcionó los datos.

Para el sistema de clasificación, se optó por plantear la segmentación utilizando el algoritmo CLARA. Sin embargo, se decidió excluir los ítems KIT-EMBRA-003 y TOLVA de esta segmentación debido a su notable relevancia dentro del conjunto de datos. Estos ítems serán tratados de manera individual y separada para garantizar un análisis más detallado y preciso de los datos restantes presentes en el sistema.

Para comprender y justificar la agrupación obtenida, se realizó un análisis comparativo con el sistema de clasificación ABC utilizado por la empresa del caso de estudio. Esto permitió examinar el comportamiento de los datos en relación con dicha agrupación. Además, se tuvieron en cuenta los objetivos de la empresa de lograr una distribución más equilibrada de los ítems en el inventario.

La segmentación llevada a cabo mediante el algoritmo CLARA demostró ser adecuada y respondió de manera satisfactoria a los requerimientos establecidos. Esto validó el modelo de clasificación propuesto para el sistema de inventarios, el cual logra una distribución más equilibrada de los ítems, en línea con los intereses de la empresa. Esta conclusión se refuerza al contrastarla con los resultados obtenidos mediante la clasificación ABC, lo que resulta atractivo para la empresa caso de estudio.

El análisis realizado ha permitido identificar la cantidad de ítems por marca y su distribución en los tres grupos distintos. Estos grupos están compuestos por una mezcla de ítems de

todas las marcas y se han distribuido siguiendo los criterios mencionados en la metodología. Es importante destacar que la clasificación de inventarios se basa en criterios relacionados con las ventas de repuestos, los cuales indican que la agrupación de cada ítem no se basa específicamente en la marca, sino en su rotación. Esto evita adquirir cantidades similares de repuestos para cada marca cuando los movimientos no respaldan tales decisiones. Las agrupaciones obtenidas permiten una distribución adecuada de los repuestos en el inventario, ya que se basan en información real de la empresa, resultando ideal para la gestión del inventario.

Los resultados revelan que los clústeres obtenidos presentan una combinación de ítems de diversas marcas, lo cual proporciona una distribución más equilibrada en función del volumen de ventas de cada ítem. Esta estrategia evita la clasificación del inventario únicamente por marcas y ayuda a prevenir problemas de exceso o escasez en productos específicos. Una clasificación efectiva respaldada por un sólido sistema de gestión contribuye a evitar pérdidas económicas significativas.

Si bien este estudio se centró específicamente en una empresa del sector automotriz, es importante tener en cuenta la variabilidad de los resultados al aplicar los algoritmos de machine learning en los inventarios de otras industrias. Esta variabilidad puede ser influenciada por el tipo de empresa, la disponibilidad de datos y los criterios de gestión utilizados en los sistemas de inventario. Se espera que este estudio sirva como punto de partida para futuras investigaciones relacionadas con la gestión de inventarios, brindando un respaldo valioso para la toma de decisiones en empresas que requieran de análisis similares en sus sistemas de inventario.

### Referencias

- Abera, F. Z., & Khedkar, V. (2020). Machine Learning Approach Electric Appliance Consumption and Peak Demand Forecasting of Residential Customers Using Smart Meter Data. *Wireless Personal Communications*, 111(1), 65-82. <https://doi.org/10.1007/s11277-019-06845-6>
- AEADE. (2022). *Anuario AEADE 2022*. Obtenido de Asociación de empresas automotrices del Ecuador: <https://www.aeade.net/anuario/>
- Amat, J. (21 de Septiembre de 2017). *RPubs*. Obtenido de [https://rpubs.com/Joaquin\\_AR/310338](https://rpubs.com/Joaquin_AR/310338)
- Angulo-Rivera, R. J. (2019). Control interno y gestión de inventarios de la empresa constructora Peter Contratistas S.R. Ltda. *Gaceta Científica*, 5(2), Art. 2. <https://doi.org/10.46794/gacien.5.2.696>



- Anitha, P., & Patil, M. M. (2019). RFM model for customer purchase behavior using K-medias algorithm. *Journal of King Saud University - Computer and Information Sciences*, 34(5), 1785-1792. <https://doi.org/10.1016/j.jksuci.2019.12.011>
- Chavez Valderrama, L. A. W., & Salinas Flores, J. W. (2021). *Aplicación del algoritmo K-Medoid para la segmentación de los alumnos ingresantes de una universidad*. <http://dspace.esPOCH.edu.ec/handle/123456789/14598>
- Datta, S., & Datta, S. (2006). Evaluation of clustering algorithms for gene expression data. *BMC Bioinformatics*, 7(4), S17. <https://doi.org/10.1186/1471-2105-7-S4-S17>
- Durán, Y. (2012). Administración del inventario: Elemento clave para la optimización de las utilidades en las empresas. *Visión Gerencial*, 1, 55-78.
- García, M. L., Dávila, P. F. P., Abanto, S. E. S., & Cruz, A. S. V. de la. (2021). Gestión de inventarios y la rentabilidad de una empresa del sector automotriz. *Sapienza: International Journal of Interdisciplinary Studies*, 2(4), Art. 4. <https://doi.org/10.51798/sijis.v2i4.157>
- García-Barrios, D., Palomino, K., García-Solano, E., & Cuello-Quiroz, A. (2021). A Machine Learning based Method for Managing Multiple Impulse Purchase Products: An Inventory Management Approach. *Journal of Engineering Science and Technology Review*, 14(1), 25-37. Scopus. <https://doi.org/10.25103/jestr.141.02>
- Guerrero, I. (Mayo de 2017). *Algoritmos de clustering y búsqueda de asociaciones*. Obtenido de Amazonaws: [http://rstudio-pubs-static.s3.amazonaws.com/285458\\_4ccd5a5e29774f80a4aa9aa21aef74e6.html](http://rstudio-pubs-static.s3.amazonaws.com/285458_4ccd5a5e29774f80a4aa9aa21aef74e6.html)
- Gupta, T., & Panda, S. P. (2019). A Comparison of K-medias Clustering Algorithm and CLARA Clustering Algorithm on Iris Dataset. *International Journal of Engineering & Technology*, 7(4), Art. 4. <https://doi.org/10.14419/ijet.v7i4.21472>
- Gustriansyah, R., Suhandi, N., & Antony, F. (2020). Clustering optimization in RFM analysis Based on K-medias. *Indonesian Journal of Electrical Engineering and Computer Science*, 18, 470. <https://doi.org/10.11591/ijeecs.v18.i1.pp470-477>
- Gutiérrez, V., & Vidal, C. J. (2008). Modelos de Gestión de Inventarios en Cadenas de Abastecimiento: Revisión de la Literatura. *Revista Facultad de Ingeniería Universidad de Antioquia*, 43, 134-149.
- Holguín, C. J. V., Ortega, J. C. L., & Rengifo, F. C. (2004). Aplicación de los Modelos de Inventarios en una Cadena de Abastecimiento de Productos de Consumo Masivo con una Bodega y N Puntos de Venta. *INGENIERÍA Y COMPETITIVIDAD*, 6(1), Article 1. <https://doi.org/10.25100/iyc.v6i1.2287>
- Humberto, G. S. (2017). *Inventarios manejo y control*. Ecoe Ediciones.
- Katsov, I. (2017). *Introduction to algorithmic marketing: Artificial intelligence for marketing operations*. Ilia Katcov.

- Kaufman, L., & Rousseeuw, P. J. (1990). *Finding Groups in Data: An Introduction to Clúster Analysis*. John Wiley & Sons.
- Khanorkar, Y., & Kane, P. V. (2023). Selective inventory classification using ABC classification, multi-criteria decision making techniques, and machine learning techniques. *Materials Today: Proceedings*, 72, 1270-1274. <https://doi.org/10.1016/j.matpr.2022.09.298>
- Leiva-Valdebenito, S. A., & Torres-Avilés, F. J. (2010). Una revisión de los algoritmos de partición más comunes en el análisis de conglomerados: Un estudio comparativo. *Revista Colombiana de Estadística*, 33(2), 321-339.
- Lima Prudente, E. K. (2020). *Importancia del control de inventario en las empresas comerciales*. [BachelorThesis, La Libertad: Universidad Estatal Península de Santa Elena, 2020]. <https://repositorio.upse.edu.ec/handle/46000/5512>
- Lolli, F., Balugani, E., Ishizaka, A., Gamberini, R., Rimini, B., & Regattieri, A. (2019). Machine learning for multi-criteria inventory classification applied to intermittent demand. *Production Planning & Control*, 30(1), 76-89. <https://doi.org/10.1080/09537287.2018.1525506>
- Marqués, A. O., Domínguez, S. P. P., Durán, J. I. T., & Gómez, A. R. (2017). Nivel de importancia del control interno de los inventarios dentro del marco conceptual de una empresa. *Liderazgo Estratégico*, 7(1), Art. 1.
- Moharana, U. C., & Sarmah, S. P. (2018). Joint replenishment of associated spare parts using clustering approach. *The International Journal of Advanced Manufacturing Technology*, 94(5), 2535-2549. <https://doi.org/10.1007/s00170-017-0909-6>
- Morales, V., & Morales, B. (2019). *Una técnica de agrupación robusta para un enfoque Big Data: CLARABD para tipos de datos mixtos*. <http://dspace.epoch.edu.ec/handle/123456789/11225>
- Muñoz Quintero, C. V. (2019). *Caracterización del transporte usado por los que se movilizan hacia el aeropuerto el Dorado. Un caso de estudio*. <https://repository.usta.edu.co/handle/11634/15465>
- Nietto, P. R., & Nicoletti, M. D. C. (2017). Case studies in divisive hierarchical clustering. *International Journal of Innovative Computing and Applications*, 8(2), 102-112. <https://doi.org/10.1504/IJICA.2017.084893>
- Parikh, Y., & Abdelfattah, E. (2020). Clustering Algorithms and RFM Analysis Performed on Retail Transactions. *2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, 0506-0511. <https://doi.org/10.1109/UEMCON51285.2020.9298123>
- Patnaik, A. K., Bhuyan, P. K., & Krishna Rao, K. V. (2016). Divisive Analysis (DIANA) of hierarchical clustering and GPS data for level of service criteria of urban streets.

- Alexandria Engineering Journal*, 55(1), 407-418.  
<https://doi.org/10.1016/j.aej.2015.11.003>
- Pinzón, I., Pérez Ortega, G., & Arango, M. D. (2010). Mejoramiento en la gestión de inventarios. Propuesta metodológica. *Revista Universidad EAFIT*, 46(160), Article 160.
- Raja, A. M. L., Ai, T. J., & Astanti, R. D. (2016). A Clustering Classification of Spare Parts for Improving Inventory Policies. *IOP Conference Series: Materials Science and Engineering*, 114(1), 012075. <https://doi.org/10.1088/1757-899X/114/1/012075>
- Ramírez Mendoza, D. Y. (2022). *Métodos de machine learning con algoritmos de clúster no supervisados, una alternativa de segmentación de las pymes colombianas para plantear estrategias de acuerdo con sus condiciones económicas* [MasterThesis, Universidad EAFIT]. <http://repository.eafit.edu.co/handle/10784/31580>
- Rifa, I. H., Pratiwi, H., & Respatiwan, R. (2020). Clustering of earthquake risk in Indonesia using K-medoids and K-means Algorithms. *MEDIA STATISTIKA*, 13(2), 194-205. <https://doi.org/10.14710/medstat.13.2.194-205>
- Sánchez, R. (2021). Clasificación no supervisada de imágenes médicas y minería de datos. Algoritmo S3 vs K-medias. *Revista Cubana de Investigaciones Biomédicas*, 40. [http://scielo.sld.cu/scielo.php?script=sci\\_abstract&pid=S0864-03002021000200006&lng=es&nrm=iso&tlng=en](http://scielo.sld.cu/scielo.php?script=sci_abstract&pid=S0864-03002021000200006&lng=es&nrm=iso&tlng=en)
- Sandoval Serrano, L. J. (2018). *Algoritmos de aprendizaje automático para análisis y predicción de datos*. <http://redicces.org.sv/jspui/handle/10972/3626>
- Saraf, E., Pradhan, S., Joshi, S., & Sountharajan, S. (2022). Behavioral Segmentation with Product Estimation using K-medias Clustering and Seasonal ARIMA. *2022 6th International Conference on Trends in Electronics and Informatics (ICOEI)*, 1641-1648. <https://doi.org/10.1109/ICOEI53556.2022.9776834>
- Servente, M. (2002). *Algoritmos TDIDT aplicados a la Minería de datos inteligente*. Buenos Aires: Universidad de Buenos Aires.
- Torres, J. I. S., & Cárdenas, E. G. (2021). Análisis y aplicación de algoritmos de minería de datos. *Revista Perspectivas*, 6(21), Article 21. <https://doi.org/10.26620/uniminuto.perspectivas.6.21.2021.71-88>
- Tovar Giraldo, J. M., & Fuentes Romero, B. C. (2019). Diseño de un sistema de gestión de inventario para minimizar costos en una empresa comercializadora de repuestos automotriz. *Universidad San Ignacio de Loyola*. <https://renati.sunedu.gob.pe/handle/sunedu/3064965>
- Uribe, I. A. (2010). *Guía metodológica para la selección de técnicas de depuración de datos*. <https://repositorio.unal.edu.co/handle/unal/69915>

- Uribe, I. A., & Ramírez, C. J. (2009). Hacia una Metodología para la selección de técnicas de depuración de datos (junio 2009). *Avances en Sistemas e Informática*, 6(1), Article 1.
- Vesga Acevedo, D. N. (2020). Modelo de planeación de inventarios para E-commerce, utilizando herramientas de inteligencia artificial para hacer pronósticos de demanda y clasificación de inventarios. *instname:Universidad de los Andes*. <https://repositorio.uniandes.edu.co/handle/1992/48641>
- Villegas, L. (2021). *Abastecimiento múltiple para la administración de exceso de inventarios: Caso de estudio en compañía de manufactura* [Masters, Universidad Autónoma de Nuevo León]. <http://eprints.uanl.mx/22264/>
- Wickham, H., Çetinkaya-Rundel, M., & Grolemund, G. (2023). *R for Data Science*. O'Reilly Media, Inc.
- Zambrano, C., & Vera, D. (Marzo de 2022). *Registro Estadístico de Empleo en la Seguridad Social - REESS*. Obtenido de Instituto Nacional de Estadísticas y Censos - INEC: [https://www.ecuadorencifras.gob.ec/documentos/web-inec/Estadisticas\\_Economicas/Estadistica\\_empleo\\_seguridad\\_social/1\\_2022\\_REES\\_S\\_Bolet%C3%ADn.pdf](https://www.ecuadorencifras.gob.ec/documentos/web-inec/Estadisticas_Economicas/Estadistica_empleo_seguridad_social/1_2022_REES_S_Bolet%C3%ADn.pdf)
- Zúñiga, C. (10 de Enero de 2023). *El universo*. Obtenido de <https://www.eluniverso.com/noticias/economia/un-ligero-crecimiento-en-ventas-preve-el-sector-automotriz-de-ecuador-en-este-2023-pese-a-mas-desafios-nota/>

## Anexos

## Anexo A

Ítems más representativos: Número de parte y referencia

Número de parte	Referencia
2452084002	Balancín escape
1001CAA14881N	Filtro de combustible
81437220070 L	Buje
A6111800009	Filtro de aceite
07W105401A	Biela
5810159A00	Pastillas de freno de disco delantero
0102EAZ00440NA	Capó
81437026079	Amortiguador delantero
2H6853601A DPJ	Emblema
3910045800	Unidad de control eléctrico
0305BM0061N	Bomba de inyección de combustible
2V5201512	Filtro de combustible
07W115436C	Cartucho de filtro

## Anexo B

Línea de código en RStudio (versión 4.2.1) aplicada al caso de estudio

```

1
2 library(dplyr)
3 library(kableExtra)
4 library(readxl)
5
6 # Carga de la base
7 BD_R <- read_excel("D:/10. Décimo Ciclo/UIC/BD_R.xlsx")
8
9 df <- BD_R
10 head(df) %>%
11   kable(caption = "Base de Datos: Ventas de repuestos de una Empresa del Sector Automotriz " ,
12         align = "c",
13         digits = 2) %>%
14   kable_classic_2(html_font = "sans-serif",
15                  lightable_options = c("hover", "striped")) %>%
16   row_spec(0,
17           bold = T,
18           color = "white",
19           background = "#219B6D")
20
21
22 # Remover registros con algún valor faltante
23 df <- na.omit(df)
24

```

```

25 # Normalización de los datos
26 df <- scale(df[,c("sumacant", "CostoUnit", "SumaPTotal", "SumaTotal")])
27 head(df) %>%
28   kable(caption = "Base de Datos Normalizada " ,
29         align = "c",
30         digits = 6) %>%
31   kable_classic_2(html_font = "sans-serif",
32                  lightable_options = c("hover", "striped")) %>%
33   row_spec(0,
34            bold = T,
35            color = "white",
36            background = "#219B6D")
37
38 #### MÉTODO 1: K-MEANS CLUSTERING
39 #### ESTIMACIÓN ÓPTIMA DEL NÚMERO DE CONGLOMERADOS
40 #### Determinamos el número de grupos o conglomerados (k)
41 library(factoextra)
42
43 #Determinación del número de grupos o conglomerados (k)
44 library(factoextra)
45 #Obtención del número de clústeres
46 #Método del codo
47 fviz_nbclust(df, kmeans, method = "wss")+
48   geom_vline(xintercept = 4, linetype = 2)
49
50 #Método de la silueta
51 fviz_nbclust(df, kmeans, method = "silhouette")+
52   geom_vline(xintercept = 2, linetype = 2)
53
54 #Método de la estadística gap
55 fviz_nbclust(df, kmeans, method = "gap_stat")+
56   geom_vline(xintercept = 1, linetype = 2)
57
58
59 #Método K-medias
60 set.seed(123)
61 km.res <- kmeans(df, 3, nstart = 25)
62 print(km.res)
63
64 aggregate(df, by = list(cluster = km.res$cluster), mean) %>%
65   kable(caption = "Media de Cada variable por Conglomerado" ,
66         align = "c",
67         digits = 6) %>%
68   kable_classic_2(html_font = "sans-serif",
69                  lightable_options = c("hover", "striped")) %>%
70   row_spec(0,
71            bold = T,
72            color = "white",
73            background = "#219B6D")
74
75 dd <- cbind(df, cluster = km.res$cluster)
76 head(dd) %>%
77   kable(caption = "Puntuaciones" ,
78         align = "c",
79         digits = 6) %>%
80   kable_classic_2(html_font = "sans-serif",
81                  lightable_options = c("hover", "striped")) %>%
82   row_spec(0,
83            bold = T,
84            color = "white",
85            background = "#219B6D")
86
87
88 #Visualización de datos
89 library(factoextra)
90 fviz_cluster(
91   km.res,
92   data = df,
93   palette = c("#2E9FDF", "#00AFBB", "#E7B800", "#FC4E07"),
94   ellipse.type = "euclid",
95
96   star.plot = TRUE,
97
98   repel = TRUE,
99
100  ggtheme = theme_minimal()
101 )
102
103 #Exportación de datos
104 bd.kmeans=as.data.frame(km.res$cluster)
105 bd.kmeans1=as.data.frame(km.res$size)
106 bd.kmeans.unido=data.frame(bd.kmeans,BD_RT)
107 write.csv(bd.kmeans.unido,"mydatakmeansT.csv")

```

```

132 #Método CLARA
133
134 set.seed(1234)
135
136 # Generación de los datos
137 df_CLARA <- as.data.frame(df)
138
139 colnames(df_CLARA) <- c("x", "y", "a", "b")
140 rownames(df_CLARA) <- paste0("S", 1:nrow(df_CLARA))
141
142 head(df_CLARA, nrow = 10) %>%
143   kable(caption = "Conjunto de datos aleatorios" ,
144         align = "c",
145         digits = 6) %>%
146   kable_classic_2(html_font = "sans-serif",
147                  lighttable_options = c("hover", "striped")) %>%
148   row_spec(0,
149           bold = T,
150           color = "white",
151           background = "#219B6D")
152
153 library(cluster)
154 library(factoextra)
155 #Obtención del número de clústeres
156 fviz_nbclust(df_CLARA, clara, method = "wss")+
157   geom_vline(xintercept = 3, linetype = 2)
158
159 fviz_nbclust(df_CLARA, clara, method = "silhouette")+
160   geom_vline(xintercept = 3, linetype = 2)
161
162 fviz_nbclust(df_CLARA, clara, method = "gap_stat")+
163   geom_vline(xintercept = 1, linetype = 2)
164
165
166 #Ejecución del metodo CLARA
167 clara.res <- clara(df_CLARA, 3, samples = 50, pamLike = TRUE)
168
169 print(clara.res)
170
171 dd_CLARA <- cbind(df_CLARA, cluster = clara.res$cluster)
172 head(dd_CLARA, n = 3) %>%
173   kable(caption = "Puntuaciones" ,
174         align = "c",
175         digits = 6) %>%
176   kable_classic_2(html_font = "sans-serif",
177                  lighttable_options = c("hover", "striped")) %>%
178   row_spec(0,
179           bold = T,
180           color = "white",
181           background = "#219B6D")
182
183 # Medooides
184 clara.res$medoids %>%
185   kable(caption = "Medooides" ,
186         align = "c",
187         digits = 6) %>%
188   kable_classic_2(html_font = "sans-serif",
189                  lighttable_options = c("hover", "striped")) %>%
190   row_spec(0,
191           bold = T,
192           color = "white",
193           background = "#219B6D")
194
195 # Clustering
196 head(clara.res$clustering, 10)
197
198 ## visualización de datos
199 library(factoextra)
200 fviz_cluster(
201   clara.res,
202   #palette = c("#00AFBB", "#FC4E07"),
203   # color palette
204   ellipse.type = "t",
205   # Concentration ellipse
206   geom = "point",
207   pointsize = 1,
208   ggtheme = theme_classic()
209 )
210
211 #Exportación de datos
212 idcluster=as.data.frame(clara.res$clustering)
213 bdcluster=data.frame(idcluster,BD_RT)
214 write.csv(bdcluster,"mydataclaraT.csv")

```



```
217 #Método DIANA
218
219 library(cluster)
220 library(factoextra)
221 # Divisive ANalysis Clustering
222 res.diana <- diana(x = df, # data matrix
223                  stand = TRUE, # standardize the data
224                  metric = "euclidean" # metric for distance matrix
225 )
226
227 print(res.diana)
228
229 #visualización de datos
230 fviz_dend(res.diana, cex = 0.6, k = 3)
231
232 #Corte
233 grp <- cutree(res.diana, k = 3)
234
235 datosgrupo<-cbind(df,grupo=grp)
236
237 #Exportación de datos
238 table(grp)
239 tabla.diana = data.frame(res.diana$order, grp)
240 bdclusterdiana=data.frame(tabla.diana,BD_RT)
241 write.csv(bdclusterdiana,"mydatadianaT.csv")
```