



**FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA**

SEGMENTACIÓN DEL SECTOR PÚBLICO PERUANO A TRAVÉS DE  
CLUSTERING BASADO EN DATOS DE CONTRATACIÓN ELECTRÓNICA

**Línea de investigación:**

**Estadística y bioestadística**

Trabajo de suficiencia profesional para optar el título profesional de  
Licenciada en Estadística

**Autora:**

Timoteo Jacinto, Gianina Rossmery

**Asesor:**

Ruiz Arias, Raúl Alberto

ORCID: 0000-0002-8877-6158

**Jurado:**

Rivas Arguelles, José Walter

Estrada Cantero, Jeanette Nazaria

Bustillos Borja, Rubén Hernán

**Lima - Perú**

**2024**



# SEGMENTACIÓN DEL SECTOR PÚBLICO PERUANO A TRAVÉS DE CLUSTERING BASADO EN DATOS DE CONTRATACIÓN ELECTRÓNICA

## INFORME DE ORIGINALIDAD

15%

INDICE DE SIMILITUD

14%

FUENTES DE INTERNET

4%

PUBLICACIONES

5%

TRABAJOS DEL ESTUDIANTE

## FUENTES PRIMARIAS

1	<a href="https://cdn.www.gob.pe">cdn.www.gob.pe</a> Fuente de Internet	4%
2	<a href="https://saeusceprod01.blob.core.windows.net">saeusceprod01.blob.core.windows.net</a> Fuente de Internet	2%
3	<a href="https://hdl.handle.net">hdl.handle.net</a> Fuente de Internet	2%
4	<a href="https://wb2server.congreso.gob.pe">wb2server.congreso.gob.pe</a> Fuente de Internet	<1%
5	<a href="https://www.perucompras.gob.pe">www.perucompras.gob.pe</a> Fuente de Internet	<1%
6	<a href="https://docplayer.es">docplayer.es</a> Fuente de Internet	<1%
7	<a href="https://www.unfv.edu.pe">www.unfv.edu.pe</a> Fuente de Internet	<1%
8	<a href="https://pdfcookie.com">pdfcookie.com</a> Fuente de Internet	<1%



Universidad Nacional  
**Federico Villarreal**

**VRIN** | VICERRECTORADO  
DE INVESTIGACIÓN

FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA

**SEGMENTACIÓN DEL SECTOR PÚBLICO PERUANO A TRAVÉS DE  
CLUSTERING BASADO EN DATOS DE CONTRATACIÓN ELECTRÓNICA**

**Línea de Investigación:**

Estadística y Bioestadística

Modalidad de Experiencia Profesional para optar el Título Profesional de Licenciada en  
Estadística

**Autora:**

Timoteo Jacinto, Gianina Rossmery

**Asesor:**

Ruiz Arias, Raúl Alberto

ORCID: 0000-0002-8877-6158

**Jurado:**

Rivas Arguelles, José Walter

Estrada Cantero, Jeanette Nazaria

Bustillos Borja, Rubén Hernán

Lima – Perú

2024

### **Dedicatoria**

A mis padres, cuyo sacrificio y dedicación han sido mi ejemplo para seguir. A mis hermanas, por su compañía y sabios consejos a lo largo de este camino. Gracias por ser mi mayor apoyo.

## ÍNDICE

RESUMEN .....	7
ABSTRACT .....	8
I. INTRODUCCIÓN .....	9
1.1. Trayectoria de la Autora.....	10
1.2. Descripción de la empresa.....	11
1.2.1 Misión .....	12
1.2.2 Visión.....	12
1.3. Organigrama de la Empresa .....	12
1.4. Áreas y funciones desempeñadas .....	14
II. DESCRIPCIÓN DE UNA ACTIVIDAD ESPECÍFICA .....	16
2.1 Planteamiento del problema .....	16
2.1.1 Determinación del problema.....	16
2.1.2 Problema principal .....	17
2.1.3 Problemas secundarios.....	17
2.1.4 Objetivo principal .....	17
2.1.5 Objetivos secundarios .....	17
2.1.6 Justificación de la Investigación .....	18
2.1.7 Alcances y limitaciones .....	18
2.2 Marco Teórico .....	19
2.2.1 Antecedentes Bibliográficos .....	19
2.2.2 Bases Teóricas .....	21
2.2.3 Definición de términos básicos.....	32
2.3 Propuesta de solución.....	34
2.3.1 Metodología de la solución.....	34

2.3.2 Desarrollo de la solución. ....	34
III. APORTES MÁS DESTACABLES A LA EMPRESA .....	58
IV. CONCLUSIONES .....	59
V. RECOMENDACIONES .....	61
VI. REFERENCIAS.....	62
VII. ANEXOS.....	67

## ÍNDICE DE TABLAS

Tabla 1. Objetivos Estratégicos Institucionales de PERÚ COMPRAS.....	11
Tabla 2. Análisis del Uso de Catálogos Electrónicos de bienes en el Sector Público Peruano (2023) .....	36
Tabla 3. Estadísticos descriptivos de las variables .....	37
Tabla 4. Clúster Jerárquico Aglomerativo: AGNES.....	41
Tabla 5. Criterios en enlazamiento (Linkage).....	41
Tabla 6. Análisis de la Matriz de Correlación Cofenética .....	42
Tabla 7. Evaluación del método de Clustering .....	46
Tabla 8. Validación interna de los Clústeres .....	50
Tabla 9. Análisis de la calidad de los clústeres utilizando múltiples Índices .....	51
Tabla 10. Clasificación de Entidades Públicas .....	53

## ÍNDICE DE FIGURAS

Figura 1. Estructura Orgánica de la Central de Compras Públicas.....	14
Figura 2. Clasificación de los métodos de agrupamiento .....	24
Figura 3. Método de Enlace Simple.....	25
Figura 4. Método de Enlace Completo .....	26
Figura 5. Método de Enlace Promedio .....	26
Figura 6. Método de Centroid Linkage.....	27
Figura 7. Método de Ward's Linkage.....	28
Figura 8. Índice de Silueta .....	31
Figura 9. Distribución de histogramas y diagrama de caja.....	38
Figura 10. Matriz de distancias euclidianas.....	40
Figura 11. Análisis de Clustering K-means con $k=3$ .....	43
Figura 12. Análisis del algoritmo PAM para Clúster con $k=3$ .....	44
Figura 13. Método del codo .....	45
Figura 14. Criterio de la silueta.....	45
Figura 15. Índice NbClust del clúster jerárquico Ward .....	47
Figura 16. Dendrograma obtenido mediante el método jerárquico aglomerativo de Ward, con $k=3$ .....	48
Figura 17. Análisis de Clustering Jerárquico de Sectores Públicos con Visualización PCA .....	49
Figura 18. Análisis de Clustering Jerárquico .....	54
Figura 19. Análisis de clústeres mediante mapas de calor en contrataciones públicas electrónicas .....	56



## RESUMEN

**Objetivo:** Desarrollar una metodología robusta para segmentar los sectores de contratación pública peruana según sus patrones de compra de bienes en los Catálogos Electrónicos de Acuerdos Marco. Se analizaron datos de contrataciones realizadas entre enero y diciembre de 2023, considerando variables como el volumen y frecuencia de contratación, el número de proveedores, y el número de bienes, productos y catálogos. **La población y muestra:** Está conformada por los 32 sectores que realizaron contrataciones electrónicas a través de los Catálogos Electrónicos de Acuerdos Marco. **Metodología:** El estudio fue descriptivo transversal de corte mixto. Para identificar los patrones de compra de los sectores, se utilizaron técnicas de clustering jerárquico (Ward.D, Ward.D2, Complete y McQuitty) y particional (K-means, PAM). El análisis se realizó utilizando el software estadístico R Studio. **Resultados:** El análisis de clustering reveló que los métodos jerárquicos (Ward.D, Ward.D2, Complete y McQuitty) proporcionan una segmentación más robusta y consistente, con una notable superioridad del método de Ward. Aunque K-means y PAM con  $k=3$  explican una alta proporción de la varianza total (98.3%), el análisis de silueta mostró cierta ambigüedad en la clasificación de algunos casos, especialmente en la frontera entre los clústeres 2 y 3. **Conclusiones:** El método jerárquico de Ward fue el más efectivo, identificando tres clústeres: 1) "Núcleo de Gobierno", abarcando el 43.7% (14 sectores); 2) "Control, Justicia, Administración Pública y Desarrollo Social", con un 50.0% (16 sectores); y 3) "Gobiernos Subnacionales", representando el 6.3% restante (2 sectores).

*Palabras clave:* Métodos jerárquicos, métodos de partición, Ward K-means, PAM, contrataciones públicas

## ABSTRACT

**Objective:** Develop a robust methodology to segment the sectors of Peruvian public procurement according to their purchasing patterns of goods in the Electronic Catalogs of Framework Agreements. Data from procurements made between January and December 2023 were analyzed, considering variables such as the volume and frequency of procurement, the number of suppliers, and the number of goods, products, and catalogs. **Population and Sample:** Composed of the 32 sectors that made electronic procurements through the Electronic Catalogs of Framework Agreements. **Methodology:** The study was a mixed cross-sectional descriptive analysis. To identify the purchasing patterns of the sectors, hierarchical clustering techniques (Ward.D, Ward.D2, Complete, and McQuitty) and partitional clustering techniques (K-means, PAM) were used. The analysis was performed using the statistical software R Studio. **Results:** The clustering analysis revealed that the hierarchical methods (Ward.D, Ward.D2, Complete, and McQuitty) provide a more robust and consistent segmentation, with a notable superiority of the Ward method. Although K-means and PAM with  $k=3$  explain a high proportion of the total variance (98.3%), the silhouette analysis showed some ambiguity in the classification of certain cases, especially on the border between clusters 2 and 3. **Conclusions:** The hierarchical Ward method was the most effective, identifying three clusters: 1) "Government Core," covering 43.7% (14 sectors); 2) "Control, Justice, Public Administration, and Social Development," with 50.0% (16 sectors); and 3) "Subnational Governments," representing the remaining 6.3% (2 sectors).

*Keywords:* Hierarchical methods, Partitioning methods, Ward, K-means, PAM, public procurement

## I. INTRODUCCIÓN

En el contexto actual de la gestión pública en Perú, la optimización de los procesos de contratación es esencial para garantizar tanto la eficiencia como la transparencia en el uso de los recursos públicos. Los Catálogos Electrónicos de Acuerdos Marco han emergido como una herramienta clave para agilizar las adquisiciones; no obstante, la ausencia de una segmentación detallada de los sectores públicos limita significativamente su potencial. Esto puede derivar en la adquisición de bienes inapropiados, costos elevados, calidad insuficiente y retrasos en los procedimientos.

Este estudio aborda esta problemática mediante un análisis basado en técnicas de Clustering, con el objetivo de agrupar los sectores públicos peruanos según sus contrataciones realizadas entre enero y diciembre de 2023. Para ello, se aplicaron métodos de agrupamiento tanto jerárquicos como particionales, incluyendo Ward, Single, Complete, Average, McQuitty, Median, Centroid, K-means y PAM. La segmentación se realizó considerando múltiples variables, tales como el volumen y frecuencia de contratación, el número de proveedores, y la diversidad de bienes, productos y catálogos empleados.

La identificación de estos grupos permite diseñar estrategias de compra más eficientes y adaptadas a las necesidades de cada sector, mejorando los catálogos electrónicos. Este estudio contribuyó a la literatura existente al proponer un modelo de segmentación que puede servir como referencia para futuras investigaciones en el ámbito de las contrataciones públicas.

Esta investigación se estructura en seis capítulos interconectados. El primer capítulo contextualiza la investigación, describiendo la institución de estudio y el problema abordado. El segundo capítulo desarrolla el marco teórico, revisando antecedentes y fundamentos conceptuales. El tercer capítulo detalla el diseño metodológico, incluyendo la población, la muestra y los instrumentos de recolección de datos. Las conclusiones del estudio se presentan

en el cuarto capítulo, seguidas de recomendaciones en el quinto capítulo. Finalmente, el sexto capítulo ofrece una lista de referencias bibliográficas que sustentan el trabajo realizado.

### **1.1. Trayectoria de la Autora**

Bachiller en la carrera profesional de Estadística y egresada de la Maestría en Ingeniería Industrial de la Universidad Nacional Mayor de San Marcos, con formación especializada en diversas áreas, entre las que destacan: “Proyectos de Inversión Pública”, “Business Analytics: Gestión de la Predictibilidad”, “Machine Learning con R”, “Analytics y Minería de Datos”, “Big Data para el Desarrollo Sostenible”, “Sistema Nacional de Abastecimiento”, “Contrataciones del Estado”, “Proyectos de Inversión Pública”, “Business Intelligence & Analytics”, “Muestreo” y “Finanzas”, entre otros programas de capacitación.

Desde 2007, he desempeñado diversas funciones en el ámbito de la investigación y el análisis de datos. En el Instituto Nacional de Estadística e Informática (INEI), durante más de 8 años, participé en proyectos relacionados con el análisis de datos, muestreo y elaboración de indicadores en áreas como economía, salud, educación y seguridad ciudadana. Posteriormente, entre 2014 y 2016, trabajé en el sector privado como Analista Business Intelligence en Mapcity y Jaque Mate Decisiones Inteligentes, donde me dediqué al análisis de datos, proyección de niveles socioeconómicos y desarrollo de modelos predictivos. Desde 2017 hasta la actualidad, en la Central de Compras Públicas, he desempeñado el rol de Asistente de Estudios de Mercado, realizando análisis de datos, diseño de encuestas, muestreo y modelado predictivo para optimizar las contrataciones estatales.

En las investigaciones realizadas se emplearon diversas fuentes de datos, tanto internas de la institución como provenientes de registros administrativos de otras entidades y empresas privadas. Se desarrollaron modelos estadísticos y se optimizaron los procesos mediante la creación de códigos, scripts, do-files y sintaxis en software especializados como SPSS, Power BI, SQL, Stata, R, Python, entre otros.

## 1.2. Descripción de la empresa

La Central de Compras Públicas - PERÚ COMPRAS fue creada el 3 de junio de 2008 mediante el Decreto Legislativo N.º 1018 y comenzó a operar formalmente el 18 de marzo de 2016. Es un organismo público adscrito al Ministerio de Economía y Finanzas, con autonomía técnica, funcional y administrativa. Su principal objetivo es optimizar las contrataciones públicas a nivel nacional mediante sistemas y procedimientos eficientes y dinámicos, apoyados por personal altamente especializado y el uso de tecnologías de la información, así como el aprovechamiento de las economías de escala. PERÚ COMPRAS forma parte del Sistema Nacional de Abastecimiento.

El propósito de PERÚ COMPRAS se enfoca en optimizar las adquisiciones públicas a nivel nacional mediante la implementación de estrategias, herramientas y procedimientos especializados en contratación. A mediano plazo, la institución establece tres Objetivos Estratégicos Institucionales (OEI) alineados con su misión, los cuales contribuyen al logro de los objetivos sectoriales definidos en el Plan Estratégico Sectorial Multianual del Sector Economía y Finanzas.

### Tabla 1.

#### Objetivos Estratégicos Institucionales de PERÚ COMPRAS

OEI	Objetivo
OEI 1	Optimizar las contrataciones públicas electrónicas a nivel nacional.
OEI 2	Fortalecer la gestión institucional con integridad y transparencia.
OEI 3	Fortalecer la Gestión de Riesgos de Desastres.

*Nota:* Resolución Jefatural N° 000028-2022-PERÚ COMPRAS-JEFATURA.

Resolución Gerencial N° 000014-2023-PERÚ COMPRAS-GG.

### **1.2.1 Misión**

PERÚ COMPRAS desarrolla mecanismos y estrategias de compra eficientes, transparentes, ágiles e innovadoras para el Estado, con el objetivo de contribuir al bienestar y la competitividad del país, conforme a su Ley de Creación y normativa complementaria.

### **1.2.2 Visión**

PERÚ COMPRAS, en concordancia con la visión del Sector Economía y Finanzas definida en la Resolución Ministerial N° 411-2016-EF/41, tiene como propósito promover un crecimiento económico sostenible, mejorar la calidad de vida de los peruanos y garantizar una política fiscal responsable y transparente, todo ello dentro de un contexto de estabilidad macroeconómica.

## **1.3. Organigrama de la Empresa**

El organigrama de PERÚ COMPRAS está estructurado en órganos de línea y oficinas, cada uno con roles claramente definidos que permiten el cumplimiento de sus responsabilidades en la gestión de las adquisiciones públicas. A continuación, se detallan los principales órganos de línea, los cuales asumen funciones operativas y técnicas esenciales dentro del sistema nacional de compras públicas:

- **Dirección de Acuerdos Marco:** Esta dirección lidera el diseño, organización y ejecución de los Catálogos Electrónicos de Acuerdos Marco (CEAM). Sus actividades comprenden la selección de proveedores, la implementación y administración de los catálogos, así como la renovación de estos instrumentos esenciales para las contrataciones.
- **Dirección de Compras Corporativas:** Encargada de gestionar las compras corporativas asignadas a PERÚ COMPRAS, esta dirección agrupa la demanda, ofrece asistencia técnica en sus áreas de competencia y propone mejoras para optimizar las contrataciones institucionales. Asimismo, aborda tareas específicas que le son asignadas para fortalecer la eficiencia del sistema.

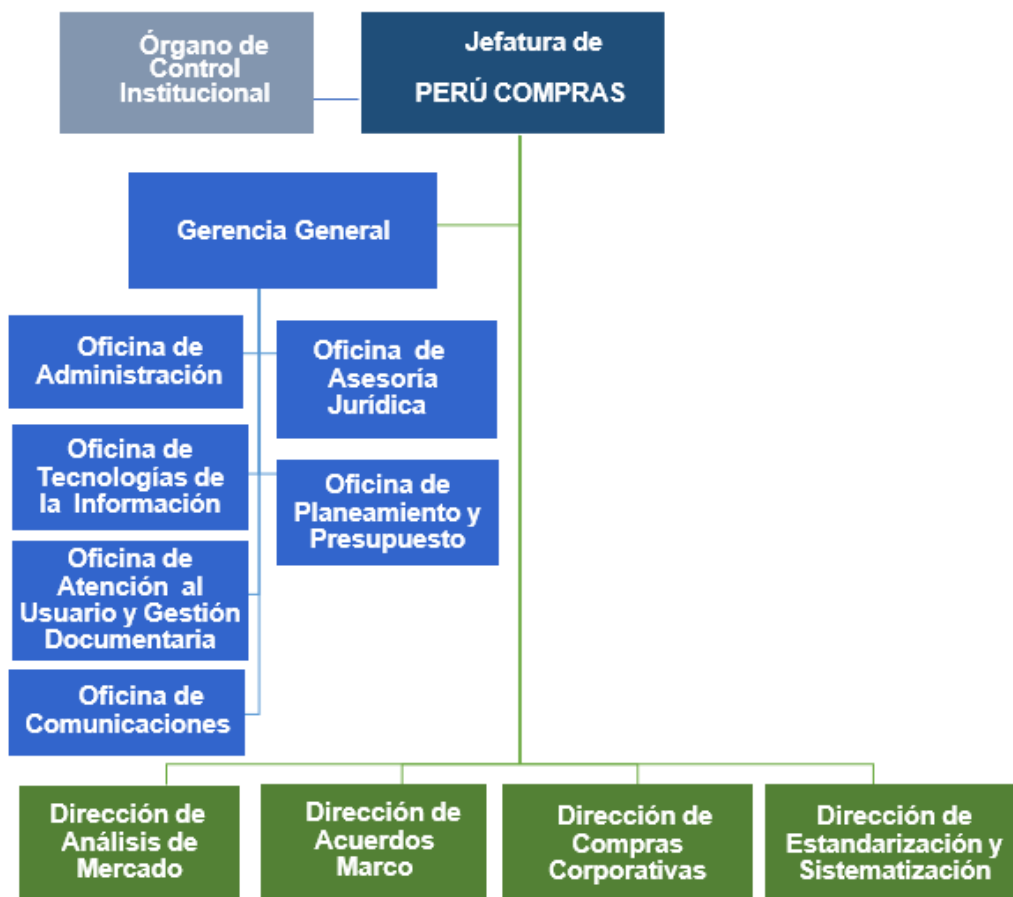
- **Dirección de Estandarización y Sistematización:** Su responsabilidad principal es gestionar el Listado de Bienes y Servicios Comunes de Subasta Inversa, además de estructurar las Fichas Producto necesarias para los catálogos electrónicos. Promueve la estandarización de los requerimientos del Estado a través de procesos de homologación de bienes y servicios, contribuyendo a la uniformidad en las adquisiciones.
- **Dirección de Análisis de Mercado:** Esta dirección genera información estratégica para optimizar las compras públicas, con un enfoque en estudios y análisis del mercado tanto nacional como internacional. Sus investigaciones abarcan:
  - El análisis del abastecimiento público.
  - Estudios de satisfacción relacionados con los procesos de Subasta Inversa Electrónica, Acuerdos Marco y Homologación.
  - Factores que inciden en la declaratoria de desierto en ciertos rubros.
  - La relación entre la estandarización de productos y la eficiencia en las compras públicas.
  - Análisis de costos específicos, como los asociados a servicios de boletos aéreos.

Los resultados de estos estudios están disponibles en el observatorio público PERÚ

OBSERVA: <https://www.perucompras.gob.pe/observatorio/que-es.php>.

**Figura 1.**

*Estructura Orgánica de la Central de Compras Públicas*



*Nota:* Adaptado de la estructura organizativa de PERÚ COMPRAS, 2019, según el Reglamento de Organización y Funciones.

#### 1.4. Áreas y funciones desempeñadas

En su rol como asistente de estudios en la Dirección de Análisis de Mercado (DAMER), la autora desempeñó las siguientes funciones principales:

- **Diseño y ejecución de encuestas de satisfacción:** Participó en el desarrollo y aplicación de encuestas dirigidas a usuarios, con el propósito de recolectar datos relevantes para respaldar decisiones estratégicas en las diversas áreas de la organización. Sus responsabilidades incluyeron la formulación de preguntas, la interpretación de los resultados obtenidos y la



redacción de informes detallados que ofrecieron recomendaciones útiles para mejorar los servicios y procesos internos.

- **Elaboró estudios y documentos técnicos:** Contribuyó a la creación de informes y análisis técnicos destinados a impulsar buenas prácticas en las adquisiciones públicas. Estos documentos proporcionaron pautas claras y sugerencias prácticas, apoyando la gestión eficiente de recursos públicos y promoviendo la transparencia en las contrataciones.
- **Desarrolló de modelos estadísticos:** Empleó herramientas especializadas, como SPSS y R, para procesar y analizar grandes conjuntos de datos, obteniendo indicadores clave que mejoraron los procesos de compra pública. Los modelos creados permitieron identificar tendencias y patrones esenciales, fortaleciendo la toma de decisiones informada y optimizando la gestión de adquisiciones.
- **Análisis de datos:** Realizó evaluaciones exhaustivas para detectar patrones y tendencias significativas en las contrataciones públicas. Esta labor generó información valiosa para mejorar los procedimientos de adquisición, facilitando la creación de reportes periódicos que sirvieron como insumo para decisiones estratégicas de las entidades públicas, contribuyendo a una administración de recursos más efectiva y orientada a resultados.

## II. DESCRIPCIÓN DE UNA ACTIVIDAD ESPECÍFICA

### 2.1 Planteamiento del problema

#### 2.1.1 *Determinación del problema*

Desde su fundación en marzo de 2016, la Central de Compras Públicas – PERÚ COMPRAS ha registrado avances destacados en la mejora de sus procesos de adquisición. En 2017, la entidad gestionaba 12 Catálogos Electrónicos, pero este número se incrementó significativamente, alcanzando un total de 45 al cierre de 2023. Este crecimiento exponencial ha generado una mayor demanda operativa, dificultando en algunos casos la determinación óptima de las necesidades de las entidades públicas. Por ello, se hace evidente la relevancia de analizar y comprender los patrones de compra de manera sistemática, lo que permitirá implementar estrategias de contratación más efectivas y adaptadas a las particularidades de cada sector público.

La segmentación es una herramienta clave para formular estrategias más efectivas (Kotler y Armstrong, 2012; Kotler et al., 2016), dado que facilita la adaptación de los productos y servicios a las características específicas de cada grupo de usuarios. Este enfoque no solo maximiza el impacto, sino que también mejora la eficiencia en las contrataciones públicas, permitiendo un uso óptimo de los recursos.

La segmentación de los sectores públicos representa una oportunidad para optimizar los procesos de contratación estatal y maximizar la eficiencia en el uso de los catálogos electrónicos. Sin embargo, para su adecuada implementación, es necesario contar con un enfoque metodológico sólido respaldado por herramientas analíticas avanzadas. Al definir criterios de segmentación específicos y relevantes mediante técnicas analíticas apropiadas, es posible obtener un entendimiento profundo de las necesidades y patrones de compra de cada sector, lo que permite diseñar estrategias de contratación más adaptadas y efectivas.

### **2.1.2 Problema principal**

- ¿Cómo se puede desarrollar una metodología robusta para segmentar los sectores de contratación pública en función de sus patrones de compra de bienes en los Catálogos Electrónicos de Acuerdos Marco?

### **2.1.3 Problemas secundarios**

- ¿Qué técnicas de Clustering serán más efectivas para obtener resultados robustos y significativos al analizar variables como volumen y frecuencia de contratación, número de proveedores, y número de bienes, productos y catálogos utilizados, basándose en los datos de contrataciones realizadas a través de los Catálogos Electrónicos de Acuerdos Marco?
- ¿Cuál es el número óptimo de segmentos que maximiza la homogeneidad dentro de cada grupo y minimiza la varianza entre ellos?
- ¿Cómo se pueden identificar segmentos homogéneos utilizando los criterios de agrupamiento establecidos?
- ¿Qué patrones de compra caracterizan a cada segmento, en términos de las variables utilizadas?

### **2.1.4 Objetivo principal**

- Desarrollar una metodología robusta para segmentar los sectores de contratación públicas según sus patrones de compra de bienes en los Catálogos Electrónicos de Acuerdos Marco

### **2.1.5 Objetivos secundarios**

- Seleccionar la técnica de Clustering más apropiadas para segmentar el sector público peruano en función de variables como volumen y frecuencia de contratación, número de proveedores, y diversidad de bienes, productos y catálogos utilizados, basándose

en los datos de contrataciones realizadas a través de los Catálogos Electrónicos de Acuerdos Marco.

- Identificar el número de segmentos que mejor representa la heterogeneidad de los datos de los Catálogos Electrónicos de Acuerdos Marco.
- Identificar segmentos homogéneos utilizando los criterios de agrupamiento establecidos.
- Caracterizar los segmentos en términos de las variables utilizadas

### **2.1.6 Justificación de la Investigación**

**2.1.6.1 Justificación Teórica.** La identificación de patrones de compra específicos en los diferentes sectores de la contratación pública permite desarrollar un marco teórico y metodológico que contribuya al avance del conocimiento en el campo de las compras públicas.

**2.1.6.2 Justificación Práctica.** Desde una perspectiva práctica, al agrupar a los diferentes sectores según sus características y necesidades específicas, es posible diseñar estrategias de adquisición más personalizadas y eficientes.

**2.1.6.3 Justificación Metodológica.** La metodología implementada facilita la identificación de patrones de compra, proporcionando una base sólida no solo para optimizar las estrategias de contratación, sino también para impulsar futuras investigaciones en este ámbito, contribuyendo a una gestión más eficiente y alineada con las necesidades de cada sector.

### **2.1.7 Alcances y limitaciones**

Este estudio tiene como objetivo analizar los patrones de compra en diversos sectores públicos, con el fin de identificar características específicas y diseñar estrategias de mejora personalizadas. A través del análisis de datos de contratación, se pretende caracterizar el comportamiento de compra comunes.

Una de las limitaciones más significativas de este estudio es la ausencia de un historial de contratación suficientemente extenso en ciertos sectores, lo cual dificultó llevar a cabo un análisis completo a nivel de las entidades públicas. Además, aspectos importantes como la ubicación geográfica o el tipo específico de bienes contratados no pudieron ser analizados en profundidad. Esta restricción afecta la capacidad para entender de manera integral las variaciones en los patrones de compra entre diferentes entidades y sectores.

Para investigaciones futuras, se recomienda abordar estos aspectos cuando se disponga de datos más completos y detallados. Al incluir estas variables, será posible llevar a cabo un análisis más profundo que refleje no solo las diferencias entre las entidades públicas, ubicación geográfica y tipo de contratación, sino también contribuir a una optimización más eficaz de los recursos públicos a través de un enfoque más segmentado.

## **2.2 Marco Teórico**

### **2.2.1 Antecedentes Bibliográficos**

Correa et al. (2024) proponen un enfoque integral para la segmentación de clientes utilizando R, cubriendo tanto métodos básicos como avanzados, con énfasis en el Clustering. En su obra *Segmentación Efectiva de Clientes utilizando R: Técnicas para Administración y Marketing Avanzado*, muestran cómo identificar patrones de compra y comportamientos clave de los consumidores a través del análisis de datos. El libro<sup>1</sup> resalta la importancia de una gestión adecuada de bases de datos, alineándose con los objetivos de este estudio.

En el estudio de Pastrán y Gongora (2021), se analizan distintos algoritmos de clusterización para datos no supervisados, comparando técnicas de clasificación dentro del proceso KDD (Knowledge Discovery in Databases). Los métodos se agrupan en tres categorías: partición, densidad y difusos. El análisis muestra que el algoritmo K-medoides es

---

1 [https://www.researchgate.net/publication/377189890\\_Segmentacion\\_Efectiva\\_de\\_Clientes\\_utilizando\\_R\\_Tecnicas\\_para\\_Administracion\\_y\\_Marketing\\_Avanzado](https://www.researchgate.net/publication/377189890_Segmentacion_Efectiva_de_Clientes_utilizando_R_Tecnicas_para_Administracion_y_Marketing_Avanzado)

más eficiente para segmentar en grupos pequeños, mientras que K-means ofrece mejores resultados para un mayor número de grupos, aunque se vuelve costoso computacionalmente con grandes volúmenes de datos o un número de grupos superior a 10.

Ventura y Fernández (2020) realizaron un análisis de agrupamiento (clustering) basado en la expresión génica en pacientes con cáncer de mama. Los subtipos moleculares resultantes fueron correlacionados con variables clínicas de interés. Aunque se exploró la aplicación de K-means, un análisis comparativo exhaustivo con métodos jerárquicos no fue llevado a cabo.

Lira (2018) empleó el algoritmo K-means para analizar patrones de consumo de agua potable en Andahuaylas, segmentando a los usuarios en tres categorías: bajo, medio y alto consumo. A través de este enfoque, la autora demostró cómo la minería de datos puede ser útil para entender el comportamiento de los consumidores y mejorar la gestión del agua mediante la implementación de estrategias más eficaces.

Tripathi et al. (2018) exploraron la importancia de la segmentación de clientes en CRM y compararon los algoritmos de K-means y clustering jerárquico. Ambos métodos identificaron cinco segmentos de clientes ( $K=5$ ), pero el Clustering jerárquico, al no requerir definir a priori el número de grupos, demostró ser más flexible y proporcionó resultados más robustos, especialmente cuando se trabajaron con datos aleatorios.

En el estudio de Ochoa et al. (2017), se realizó un análisis exhaustivo de diversas técnicas de minería de datos no supervisadas para segmentar estudiantes. El estudio utilizó los algoritmos K-means y PAM dentro del Clustering particional, así como Clustering jerárquico aglomerativo, para identificar patrones en el comportamiento estudiantil. Los resultados mostraron que K-means generó grupos de mayor calidad, destacándose con mayor similitud dentro de la agrupación y mayor separación entre los grupos formados, validado mediante el coeficiente de silueta. Este enfoque permitió personalizar intervenciones educativas y mejorar

el rendimiento académico, destacando la relevancia de los métodos avanzados de análisis en el ámbito educativo.

### **2.2.2 Bases Teóricas**

**2.2.2.1 Análisis de clúster.** Autores como Ward (1963), MacQueen (1967), Kaufman y Rousseeuw (1990) y más recientemente Jain (2010) han realizado contribuciones fundamentales al campo del análisis de clúster. Estos métodos comparten el objetivo común de identificar grupos homogéneos de objetos, aunque difieren en sus detalles específicos. Los pasos generales involucrados en este proceso son:

1. Selección de las variables
2. cálculo de la matriz de distancias o similitud
3. Selección del algoritmo de Clustering
4. Determinación del número óptimo de clústeres
5. Evaluación de los resultados

**2.2.2.2 Medidas de distancia o similitud.** La elección de una métrica de distancia o similitud es crucial en el análisis de clúster. Esta decisión determina cómo se calculan las distancias entre los objetos de datos y, por consiguiente, la forma en que se agrupan.

Diversos autores, como Cuadras (1989), Legendre y Legendre (1979), Gower y Legendre (1986), Legendre, Dallot y Legendre (1985), Hair et al. (2010) y Everitt et al. (2011), propusieron criterios y metodologías para la selección de la métrica más apropiada en cada caso, considerando características como el tipo de variables (cuantitativas, categóricas o mixtas), la distribución de los datos y la naturaleza de la relación entre las observaciones. Entre las medidas de distancia más comúnmente empleadas para datos cuantitativos, se encuentran:

- Distancia Euclídea: es adecuada para datos continuos y sin outliers. Si expresión es:

$$distancia(x, y) = \left\{ \sum_i (x_i - y_i)^2 \right\}^{\frac{1}{2}}$$

- Distancia Manhattan, según Han y Kamber (2001) enfatizan la importancia de seleccionar la distancia adecuada según el problema en cuestión. Por ejemplo, la distancia de Manhattan puede ser más robusta en presencia de ruido (outliers) y útil para datos categóricos ordinales.

$$distancia(x, y) = \sum_i |x_i - y_i|$$

- Distancia de Minkowski, según Han y Kamber (2001) en su obra sobre minería de datos, destacan la importancia de la distancia de Minkowski como una herramienta fundamental para analizar y comparar datos. Su versatilidad y su capacidad para capturar diferentes tipos de similitud la convierten en una medida muy utilizada en diversos campos, donde  $q \geq 1$ .

$$distancia(x, y) = \left( \sum_i |x_i - y_i|^q \right)^{\frac{1}{q}}$$

- La distancia de Mahalanobis, propuesta por Mahalanobis (1936), es una medida estadística que supera a la distancia euclidiana al considerar las correlaciones entre variables, lo que la hace más robusta para el análisis de datos multivariados.

$$distancia(x, y) = \left( (x_i - y_i)' s^{-1} (x_i - y_i) \right)^{\frac{1}{2}}$$

La distancia de Bray-Curtis, propuesta por Bray y Curtis (1957), es una medida de disimilitud ecológica que cuantifica la diferencia relativa en la abundancia de especies entre dos sitios, siendo especialmente útil en comunidades con variada riqueza de especies.

$$BC_{IJ} = 1 - \frac{2C_{IJ}}{s_i + s_j}$$



Donde  $C_{IJ}$  es la suma de las abundancias menores para cada especie común en ambos sitios, y  $S_i$  y  $(S_j)$  son los totales de especímenes en los sitios (i) y (j) respectivamente.

- La métrica de Canberra, propuesta por Lance y Williams (1966), es útil para datos con ceros o valores pequeños, ya que pondera las diferencias de manera distinta a la distancia euclidiana.

$$distancia(x, y) = \sum_i \frac{|x_i - y_i|}{|x_i| + |y_i|}$$

**2.2.2.3 Método de agrupamiento.** También conocido como clustering, tiene como objetivo agrupar observaciones similares entre sí, formando clústeres o conglomerados. Existen diversos métodos de agrupamiento, cada uno con sus propias características y aplicaciones, tales como:

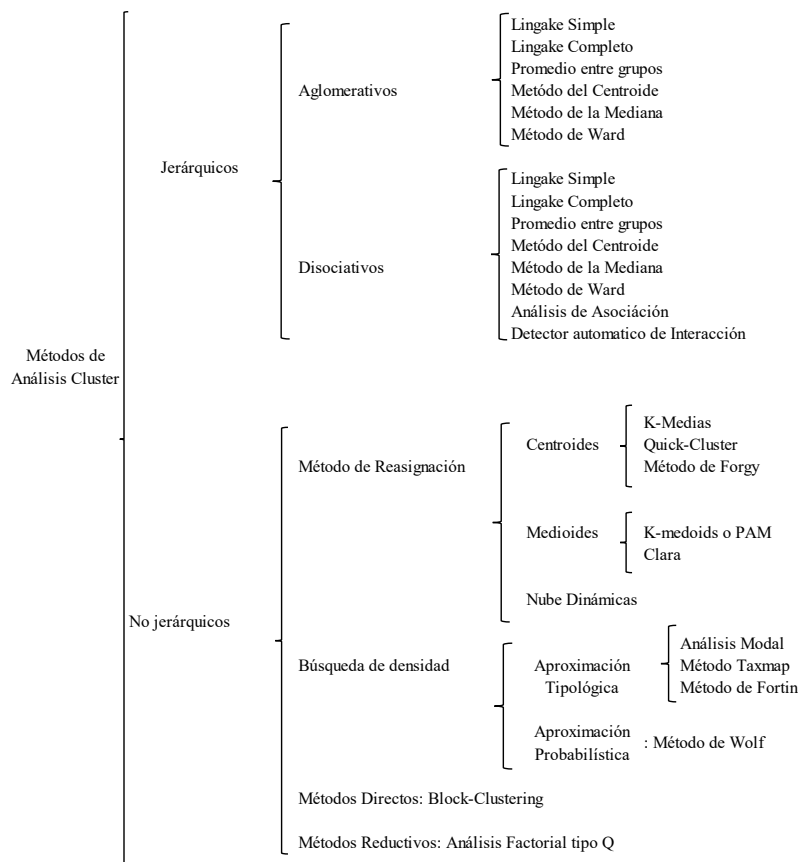
A. **Métodos Jerárquicos:** Estos métodos organizan los datos de manera jerárquica según su cercanía o similitud, formando clústeres que se representan en un árbol o Dendrograma. Se pueden dividir en dos tipos:

- **Aglomerativos:** Comienzan con cada dato como un clúster individual y van fusionándolos de acuerdo con su similitud.
- **Divisivos:** Comienzan con un solo clúster que contiene todos los datos y lo dividen en subclústeres.

B. **Métodos de partición.** Se emplean cuando se cuenta con datos heterogéneos y se desea segmentarlos en una cantidad de grupos que ha sido determinada previamente.

C. **Métodos basados en densidades.** Estas técnicas evalúan la concentración de los datos para identificar los grupos, como en el caso del DBSCAN

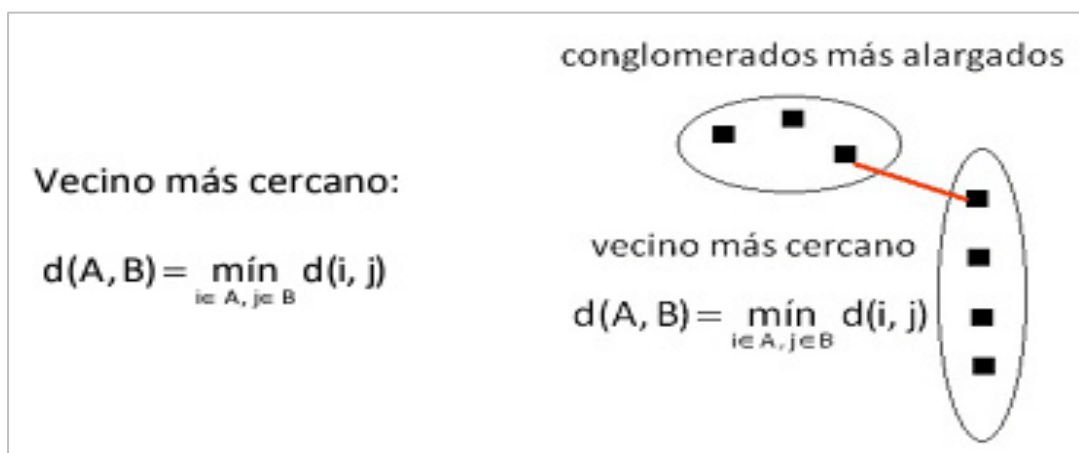
D. **Métodos basados en distribuciones.** En este enfoque, se plantea un modelo como hipótesis y se busca ajustar los datos a dicho modelo, como ocurre con los modelos de mezcla gaussiana, que se ajustan mediante el algoritmo de maximización de expectativas.

**Figura 2.***Clasificación de los métodos de agrupamiento*

Nota: Adaptado de gráfico métodos de análisis clúster (p. 7), por De la Fuente Crespo, 2000, Universidad autónoma de Madrid.

**2.2.2.4 Clustering jerárquico.** Agrupa datos en una jerarquía de clústeres. En este tipo de análisis es fundamental el método de linkage, que define cómo se calcula la distancia entre dos clústeres y, en consecuencia, cómo se fusionan o dividen en cada paso del algoritmo. Existen diferentes tipos de linkage, tales como:

- **Método Single-Linkage** conocido como el vecino más próximo, establece la distancia entre dos clústeres como la mínima distancia que existe entre cualquier par de puntos, cada uno perteneciente a un clúster diferente. En el caso de que se utilicen medidas de similitudes, se toma en cuenta la mayor similitud entre los puntos involucrados (Sharma, 1996).

**Figura 3.***Método de Enlace Simple*

*Nota:* Adaptado de gráfico distancias entre conglomerados (p.8), por De la Fuente Crespo, 2000, Universidad autónoma de Madrid.

Sea  $x_i \in C_i$  y  $x_j \in C_j$ , donde  $|C_i| = n_i$  y  $|C_j| = n_j$ . La distancia entre  $C_i$  y  $C_j$  luego de  $k$  iteraciones es:

a) En el caso de emplear **distancias**:

$$d(C_i, C_j) = \text{Min}_{(i_1, j_1)=1, 2, \dots, (n-k) | i_1 \neq j_1} \{d(C_{i_1}, C_{j_1})\}$$

$$= \text{Min}_{(i_1, j_1)=1, 2, \dots, (n-k) | i_1 \neq j_1} \left\{ \text{Min}_{x_l \in C_{i_1}, x_m \in C_{j_1}} d(x_l, x_m) \right\}$$

$$(l = 1, 2, \dots, n_{i_1})(m = 1, 2, \dots, n_{j_1})$$

b) En el caso de emplear similaridades:

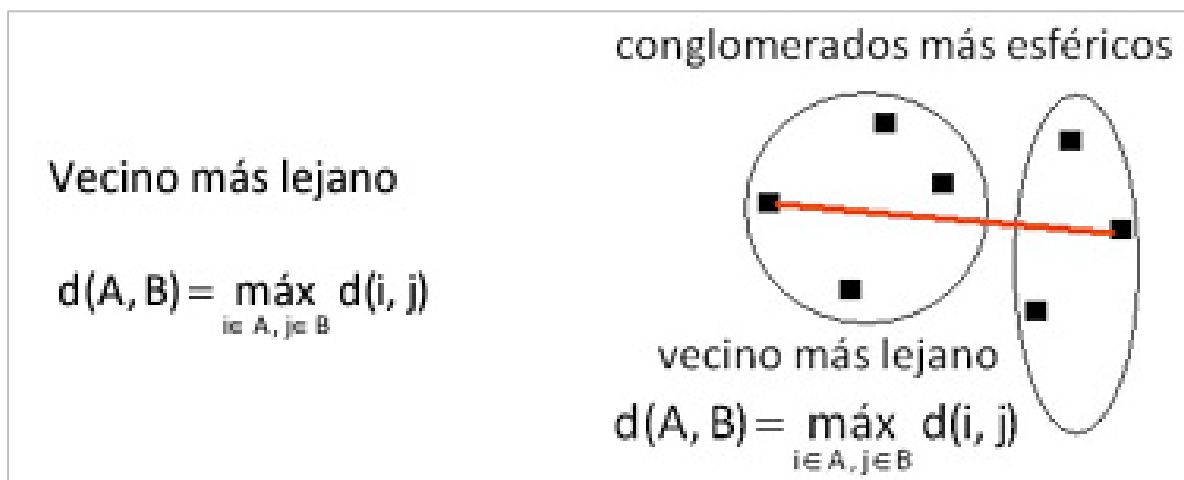
$$s(C_i, C_j) = \text{Max}_{(i_1, j_1)=1, 2, \dots, (n-k) | i_1 \neq j_1} \{s(C_{i_1}, C_{j_1})\} = \text{Max}_{(i_1, j_1)=1, 2, \dots, (n-k) | i_1 \neq j_1}$$

$$\left\{ \text{Max}_{x_l \in C_{i_1}, x_m \in C_{j_1}} s(x_l, x_m) \right\} (l = 1, 2, \dots, n_{i_1})(m = 1, 2, \dots, n_{j_1})$$

o **Método Complete linkage (vecino más lejano)**: Define la distancia entre dos clústeres como la distancia máxima entre cualquier par de puntos, uno en cada clúster (King, 1967).

Figura 4.

Método de Enlace Completo

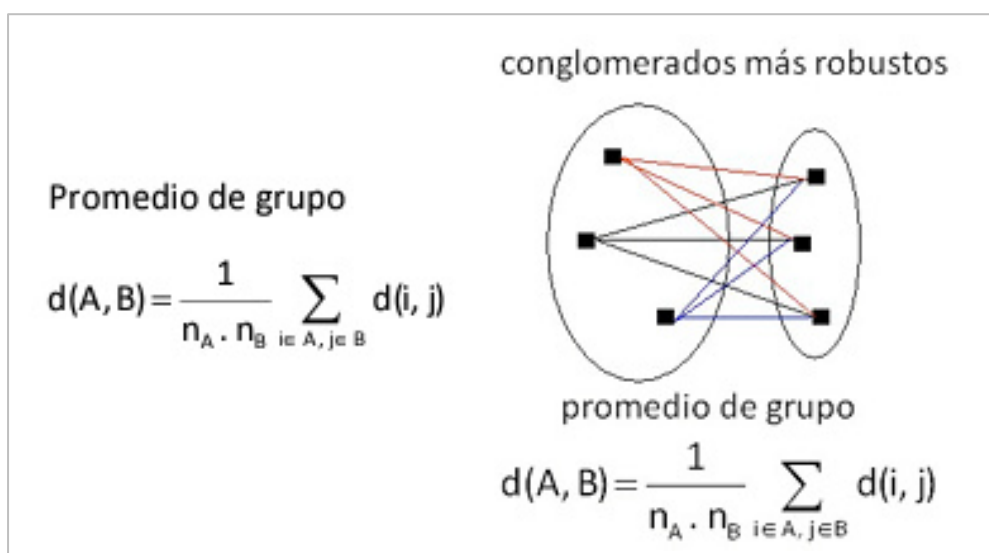


*Nota:* Adaptado de gráfico distancias entre conglomerados (p.8), por De la Fuente Crespo, 2000, Universidad autónoma de Madrid.

- **Método Average-Linkage (de la media):** Utilizan la distancia media entre los elementos de dos clústeres como criterio de enlace, como propusieron Ward (1963) y Murtagh (1984).

Figura 5.

Método de Enlace Promedio

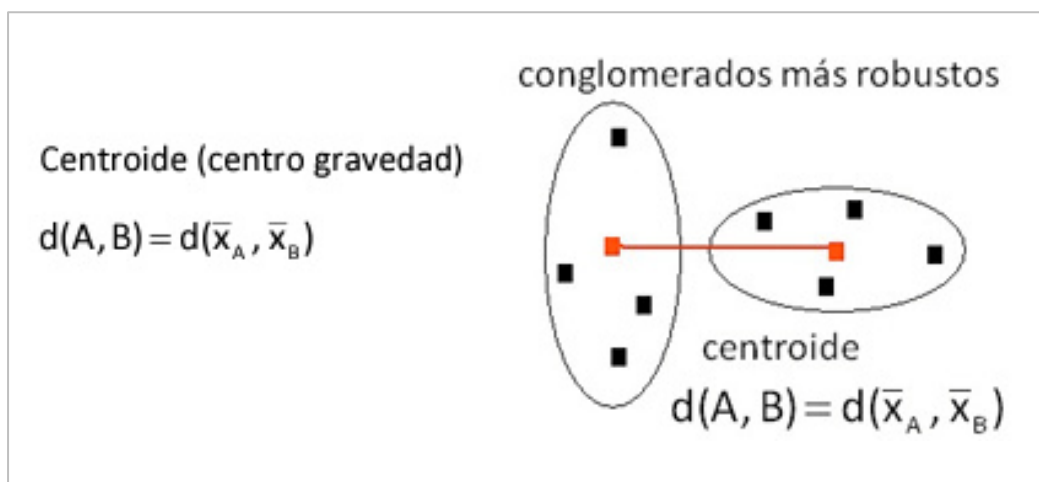


*Nota:* Adaptado de gráfico distancias entre conglomerados (p.8), por De la Fuente Crespo, 2000, Universidad autónoma de Madrid.

- **Centroid linkage:** En este método, la similitud entre dos clústeres se determina calculando la distancia entre sus respectivos centroides.

**Figura 6.**

*Método de Centroid Linkage*



*Nota:* Adaptado de gráfico distancias entre conglomerados (p.8), por De la

Fuente Crespo, 2000, Universidad autónoma de Madrid.

Supongamos que queremos medir la distancia entre el clúster  $C_i$  (con  $n_i$  elementos, formado por  $C_{i1}$  y  $C_{i2}$  con  $n_{i1}$  y  $n_{i2}$  elementos, respectivamente) y el clúster  $C_j$  (con  $n_j$  elementos). Sean  $m_i, m_{i1}, m_{i2}$  y  $m_j$  los centroides (vectores  $n$ -dimensionales) de estos clústeres (Gallardo, 1994).

$$m_l^{i1} = \frac{n_{i1}m_l^{i1} + n_{i2}m_l^{i2}}{n_{i1} + n_{i2}} \quad (l = 1, 2, \dots, n)$$

$$d^2(C_i, C_j) = \sum_{l=1}^n (m_l^j - m_l^i)^2$$

$$d^2(C_i, C_j) = \frac{n_{i1}}{n_{i1} + n_{i2}} d^2(C_{i1}, C_j) + \frac{n_{i2}}{n_{i1} + n_{i2}} d^2(C_{i2}, C_j) - \frac{n_{i1}n_{i2}}{n_{i1} + n_{i2}} d^2(C_{i1}, C_{i2})$$

Al asumir que ambos subgrupos de  $C_i$  tienen el mismo tamaño ( $n_{i1} = n_{i2}$ ), el método posiciona el centroide del clúster  $C_i$  en el punto medio entre  $C_{i1}$  y  $C_{i2}$ . En consecuencia, el centroide del clúster resultante de la unión de  $(C_i, C_j)$  se ubica en el centroide del triángulo formado por  $C_{i1}$ ,  $C_{i2}$  y  $C_j$  (Gallardo, 1994).

• **Para calcular distancias:**

$$d(C_i, C_j) = \frac{1}{2} [d(C_{i1}, C_j) + d(C_{i2}, C_j)] + \frac{1}{4} [d(C_{i1}, C_{i2})]$$

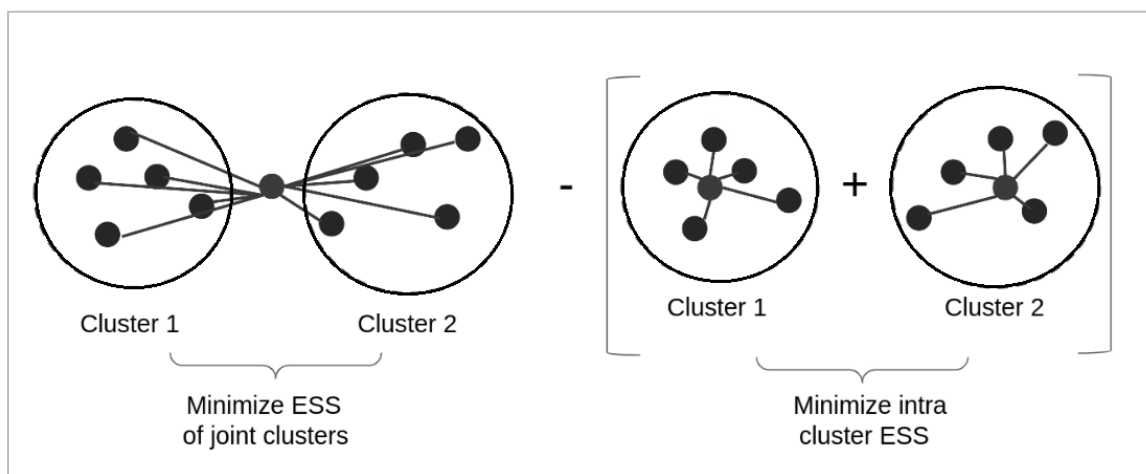
• **Para calcular similitudes:**      -                                      -

$$s(C_i, C_j) = \frac{1}{2} [s(C_{i1}, C_j) + s(C_{i2}, C_j)] + \frac{1}{4} [1 - s(C_{i1}, C_{i2})]$$

○ **Ward's linkage:** Su objetivo es reducir la pérdida de información durante el proceso de fusión de clústeres. A través del criterio de Suma de Errores Cuadráticos (ESS) busca minimizar la suma de los cuadrados de las desviaciones de cada punto con respecto al centroide de su clúster. Un valor más alto en este criterio indica una mayor dispersión de los datos dentro de cada clúster, lo que se traduce en una mayor pérdida de información.

**Figura 7.**

*Método de Ward's Linkage*



**2.2.2.5 Métodos de partición.** Dada su amplia aplicación en diversas áreas, se han seleccionado dos algoritmos de Clustering representativos para este estudio.

- El método K-means, propuesto por MacQueen et al. (1967), es un algoritmo de clasificación no supervisada que se utiliza para agrupar objetos en k grupos según sus características. Este proceso de agrupamiento se lleva a cabo minimizando la suma de las distancias entre cada objeto y el centroide correspondiente de su grupo o clúster, habitualmente utilizando la distancia cuadrática
- El Método K-medoides. Es una alternativa a K-means que, en lugar de calcular centroides, selecciona como representantes de los clústeres puntos de datos reales llamados medoides. Esta característica lo hace más robusto ante valores atípicos y adecuado para datos categóricos. Al elegir medoides, K-Medoids evita las distorsiones que pueden surgir en K-means cuando se calculan promedios de datos heterogéneos. (Rousseeuw y Kaufman, 1987).

**2.2.2.6 Técnicas gráficas.** Son herramientas visuales esenciales para explorar y evaluar la calidad de las agrupaciones obtenidas mediante algoritmos de clustering. Entre las técnicas más comunes se encuentran las siguientes:

- Dendrograma: Se trata de una representación visual del proceso de agrupamiento que ilustra la forma en que los clústeres se fusionan en etapas sucesivas.
- Gráfico de codo: Esta herramienta se emplea para determinar el número ideal de clústeres en el análisis de k-medias (k-means). Representa la suma de las distancias cuadradas dentro de los clústeres (inercia) en relación con la cantidad de clústeres.
- Matriz de disimilitud: Representa las distancias o diferencias entre pares de objetos en una matriz visual. Puede ser utilizada para observar patrones en las distancias y evaluar la formación de clústeres.

- Mapa de calor: Representa visualmente las disimilitudes entre casos en una matriz donde los colores indican la magnitud de las distancias. Puede facilitar la identificación de patrones y la determinación de la cantidad de clústeres necesarios.
- El índice NbClust es una herramienta estadística creada para asistir a los investigadores en la identificación del número óptimo de clústeres durante un análisis de agrupamiento. Este índice integra múltiples índices de validación, tanto internos como externos, con el fin de ofrecer una evaluación más sólida y confiable del número de grupos que se encuentran en un conjunto de datos.

**2.2.2.7 Calidad de los clústeres.** Para asegurar la solidez y fiabilidad de los resultados, se emplean principalmente:

A. *Índices de Validación Interna:* Se utiliza exclusivamente la información disponible en los datos mediante varios índices de validación interna. Estos índices miden tanto la cohesión interna de los clústeres como la separación entre ellos. Algunos de los índices más comúnmente empleados son:

- Índice de silueta. Según Banchemo (2015), mide qué tan bien asignado está un objeto a su propio clúster en comparación con otros. Donde el coeficiente de Silueta para una observación  $i$  se denota como  $s(i)$  y se define como:

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

Donde:

$a$ : es el promedio de las distancias de la observación  $i$  con las demás observaciones del clúster.

$b$  es la distancia mínima a otro clúster que no es el mismo en el que está la observación

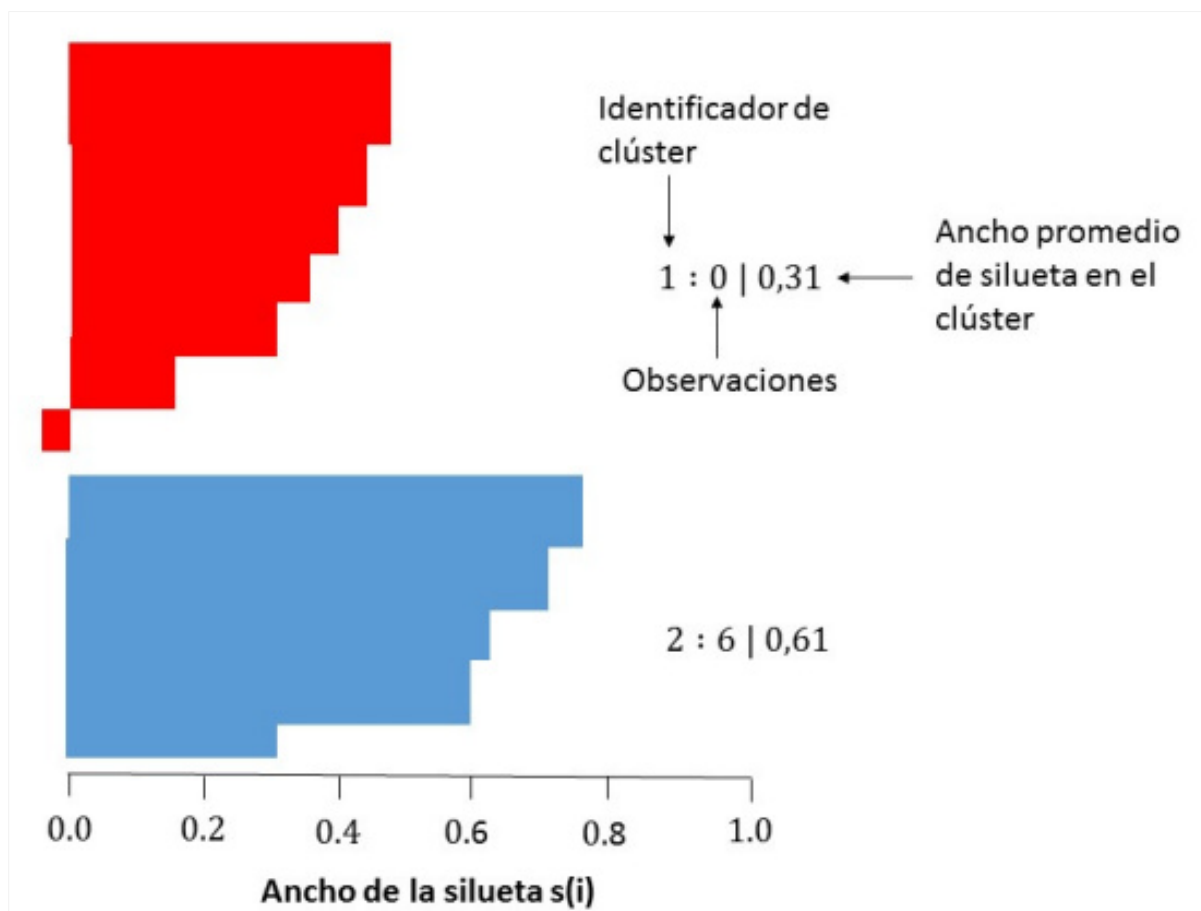


Resumiendo:

- $s(i) \approx 1$ , la observación  $i$  está bien asignada a su clúster.
- $s(i) \approx 0$ , la observación  $i$  está entre dos clústeres.
- $s(i) \approx -1$ , la observación  $i$  está mal asignada a su clúster

**Figura 8.**

*Índice de Silueta*



*Nota:* Tomado de gráfico Índice Silueta (p.44), por Pastrán L. y Gongora S, 2021, Universidad Tecnológica de Pereira

- Suma de Cuadrados Intra-clúster (SSE): Cuantifica la variabilidad dentro de cada clúster.

- El Criterio de Calinski-Harabasz analiza la proporción entre la dispersión entre los clústeres y la dispersión interna a cada uno de ellos.
- El Índice de Davies-Bouldin, el Índice de Dunn y el Coeficiente de Cophenetic ofrecen medidas adicionales para evaluar tanto la separación como la cohesión de los clústeres
- Los índices Average Proximity to Nearest Neighbor (APN), Average Distance (AD), Average Distance between Means (ADM), Figure of Merit (FOM) y Connectivity proporcionan métricas complementarias para analizar la calidad de los clústeres desde diversas perspectivas.

**B. Validación Externa de los Clúster:** Se evaluó mediante índices de validación externa, los cuales comparan la agrupación resultante con información externa conocida. Los índices utilizados incluyen:

- Pureza: Proporciona una medida sencilla de la homogeneidad de los clústeres.
- ARI, Fowlkes-Mallows y AMI: Estos índices cuantifican la similitud entre la agrupación obtenida y una agrupación de referencia, teniendo en cuenta el azar. El ARI, por ejemplo, es ampliamente utilizado en la literatura.
- Evaluación Visual: La inspección visual de los clústeres, a través de representaciones gráficas, complementa los índices numéricos y permite identificar patrones y estructuras no evidentes en los datos.

### **2.2.3 Definición de términos básicos**

De acuerdo con la Directiva N.º 006-2021-PERÚ COMPRAS (Resolución Jefatural N.º 139-2021-PERÚ COMPRAS), que establece los lineamientos para la implementación y operación del Catálogo Electrónico de Acuerdos Marco, se utilizan las siguientes definiciones:

- **Los Acuerdo Marco.** Refiérase al acuerdo establecido entre PERÚ COMPRAS y el proveedor adjudicatario, en el cual se establecen las obligaciones y derechos de las partes durante la vigencia del mismo.
- **Catálogos Electrónicos:** Se refiere a la herramienta implementada por PERÚ COMPRAS, para canalizar las contrataciones de bienes y/o servicios.
- **Contratación a través de los Catálogos Electrónicos de Acuerdos Marco.** Se refiere al método especial de contratación a través del cual las entidades contratan, mediante Reglas previamente definidas por PERÚ COMPRAS, los bienes y/o servicios que se encuentran disponibles en los Catálogos Electrónicos de Acuerdos Marco, siempre que estos permitan la atención del requerimiento y se cuente con la disponibilidad de recursos.
- **Costo total del bien o servicio.** Es el valor o monto resultante que contemple todos los conceptos y aspectos referidos a las características y condiciones del bien o servicio, tales como: (i) precio; (ii) el rendimiento y vida útil; (iii) las condiciones comerciales, como los plazos de garantía, los plazos de entrega, entre otros; (iv) el costo por flete; y, (v) otros aplicables de acuerdo a las condiciones del mercado en los que se comercien los bienes o servicios contenidos en los Catálogos Electrónicos de Acuerdos Marco. (Perú Compras, 2021).
- **Fichas-producto.** Identificación inequívoca de un único bien respecto de sus características definidas previamente.
- **Proveedor adjudicatario:** Es aquel que luego del resultado de la evaluación de propuestas de un procedimiento de implementación o extensión de la vigencia de los Catálogos Electrónicos de Acuerdos Marco, obtiene al menos una oferta adjudicada
- **Rubro.** Refiérase a las actividades de un mismo tipo o relacionados a la misma especialidad

- **Producto.** Identificación de lo que requiere una Entidad para el desarrollo de sus actividades y el cumplimiento de sus funciones y fines, el cual podrá contener una o varias Fichas-producto públicas y tomar decisiones informadas para mejorar su eficiencia y eficacia.

## 2.3 Propuesta de solución

### 2.3.1 Metodología de la solución

**2.3.1.1 Tipo de investigación.** Esta investigación tiene un enfoque aplicado y utiliza un método cuantitativo. Se emplearán técnicas de Clustering como K-means, PAM y métodos jerárquicos para analizar los datos de las contrataciones y segmentar los sectores públicos peruanos según sus patrones de adquisición en los catálogos electrónicos de bienes.

**2.3.1.2 Diseño de investigación.** Se eligió un enfoque de investigación no experimental y de carácter transversal para analizar las contrataciones electrónicas efectuadas mediante los Catálogos Electrónicos de Acuerdos Marco. La naturaleza estática de los datos recopilados en un solo momento temporal facilitó la obtención de una visión general de la situación en ese momento específico, es decir, en 2023.

**2.3.1.3 Población y muestra.** La población de estudio está conformada por todos los 32 sectores que realizaron contrataciones electrónicas a través de los Catálogos Electrónicos de Acuerdos Marco durante el periodo de enero a diciembre de 2023. Dada la naturaleza exhaustiva del estudio, se ha realizado un censo de la población, incluyendo todos los datos disponibles de las contrataciones.

### 2.3.2 Desarrollo de la solución.

Este estudio se realizó en el entorno de programación R, aprovechando sus funcionalidades para desarrollar modelos de clasificación, producir gráficos de alta calidad y ajustar parámetros con el fin de mejorar los resultados.

**2.3.2.1 Construcción de las bases de datos y variables.** Para este estudio, se utilizó un conjunto de datos relacionado con las contrataciones realizadas a través de los Catálogos Electrónicos de Acuerdos Marco en Perú durante el año 2023. La información, disponible en el Anexo 1 y proporcionada mediante el enlace correspondiente, proviene del reporte N°129-2024-PERÚ COMPRAS-DAMER. Se llevará a cabo un análisis de clústeres basado en seis criterios:

- Frecuencia de contratación: es el número total de órdenes de compra formalizadas en el año 2023
- Volumen de contratación: es el monto total contratado en soles durante el año 2023
- Diversidad de catálogos: es el número de catálogos electrónicos distintos utilizados.
- Diversidad de bienes: es el número de categorías distintas contratadas.
- Diversidad de productos: es el número de Ficha-producto distintas contratadas.
- Número de proveedores: Número de proveedores distintos que negociaron.

En la Tabla 2 se presenta el detallado de las contrataciones por sector, considerando las variables mencionadas anteriormente.

**Tabla 2.***Análisis del Uso de Catálogos Electrónicos de bienes en el Sector Público Peruano (2023)*

Sector	Sector Público	Diversidad de catálogos	Diversidad de bienes	Diversidad de productos	Número de proveedores	Frecuencia de contratación	Volumen de contratación (Mill.)
PCM	PRESIDENCIA CONSEJO DE MINISTROS	27	151	1,909	516	1,669	56.95
Mincul	CULTURA	27	130	1,016	332	767	12.92
PJ	PODER JUDICIAL	23	131	2,691	753	5,292	160.08
Minam	AMBIENTAL	22	105	766	268	540	18.26
MINJUSDH	JUSTICIA	24	133	2,079	564	1,830	30.15
Mininter	INTERIOR	27	156	3,002	661	2,797	80.09
RREE	RELACIONES EXTERIORES	15	68	259	82	141	3.24
MEF	ECONOMÍA Y FINANZAS	29	141	1,455	449	1,268	46.45
Minedu	EDUCACIÓN	33	207	7,257	1,215	9,662	176.47
Minsa	SALUD	26	161	2,778	601	3,154	67.37
MTPE	TRABAJO Y PROMOCIÓN DEL EMPLEO	27	130	1,651	563	2,010	41.77
Midagri	AGRICULTURA	29	151	2,751	712	2,637	52.36
Minem	ENERGÍA Y MINAS	27	136	1,074	390	897	15.45
CGR	CONTRALORÍA GENERAL	20	89	451	131	214	36.37
DP	DEFENSORÍA DEL PUEBLO	10	32	78	38	55	0.64
JNJ	JUNTA NACIONAL DE JUSTICIA	10	42	85	44	53	0.99
MPFN	MINISTERIO PÚBLICO	21	105	913	305	774	51.23
TC	TRIBUNAL CONSTITUCIONAL	10	22	50	39	55	0.53
Mindef	DEFENSA	31	184	4,756	680	5,271	46.51
FMP	FUERO MILITAR POLICIAL	13	74	143	86	103	1.28
Congreso	CONGRESO DE LA REPÚBLICA	17	78	179	62	82	4.86
JNE	JURADO NACIONAL DE ELECCIONES	9	19	59	30	47	0.35
ONPE	OFICINA NACIONAL DE PROCESOS ELECTORALES	16	71	359	91	143	0.51
Reniec	REGISTRO NACIONAL DE IDENTIFICACIÓN Y ESTADC	16	63	253	80	136	11.71
Mincetur	COMERCIO EXTERIOR Y TURISMO	21	71	257	150	237	5.98
MTC	TRANSPORTES Y COMUNICACIONES	29	152	1,400	430	1,100	50.5
MVCS	VIVIENDA CONSTRUCCIÓN Y SANEAMIENTO	29	144	1,660	540	1,532	25.42
PRODUCE	PRODUCCIÓN	20	97	475	209	363	7.72
MIMP	MINISTERIO DE LA MUJER Y POBLACIONES VULNERA	22	110	867	260	593	13.71
Midis	DESARROLLO E INCLUSION SOCIAL	17	98	609	204	388	4.67
GL	GOBIERNO LOCAL	37	239	18,277	1,720	35,407	401.57
GR	GOBIERNOS REGIONALES	37	237	16,224	1,711	37,480	459.88

**2.3.2.2 Análisis exploratorio de datos.** En esta parte se presenta un análisis descriptivo de las variables del estudio, incluyendo medidas de tendencia central y dispersión para identificar patrones y posibles anomalías.

**Tabla 3.**

*Estadísticos descriptivos de las variables*

	Diversidad de catálogos	Diversidad de bienes	Diversidad de productos	Número de proveedores	Frecuencia de contratación	Volumen de contratación (Mill.)
<b>Min.</b>	9	19	50	30	47	0.4
<b>1st Qu.</b>	17	73	259	90	143	4.8
<b>Median</b>	23	120	965	319	771	21.8
<b>Mean</b>	23	116	2,368	435	3,647	58.9
<b>3rd Qu.</b>	28	151	2,232	573	2,167	51.5
<b>Moda</b>	27	71 <sup>a</sup>	50 <sup>a</sup>	30 <sup>a</sup>	55	0.35 <sup>a</sup>
<b>Max.</b>	37	239	18,277	1,720	37,480	459.9
<b>Desviación</b>	7.7	56.1	4,203.3	435.6	8,844.8	106.4
<b>Asimetría</b>	-0.1	0.3	2.8	1.6	3.2	2.7
<b>Curtosis</b>	-0.8	-0.3	7.9	2.5	9.7	7.4

a. Existe n múltiples modos. Se muestra el valor más pequeño

La Tabla 3 presenta una considerable variabilidad en la cantidad de catálogos electrónicos que utilizan los distintos sectores, abarcando desde 7 hasta 37 catálogos, con un promedio de 25. El número más común es de 27 catálogos, lo que sugiere que más de la mitad de los sectores están involucrados con más de la mitad de los catálogos electrónicos disponibles. Este hallazgo indica una notable adopción y uso de estos recursos por parte de los sectores analizados.

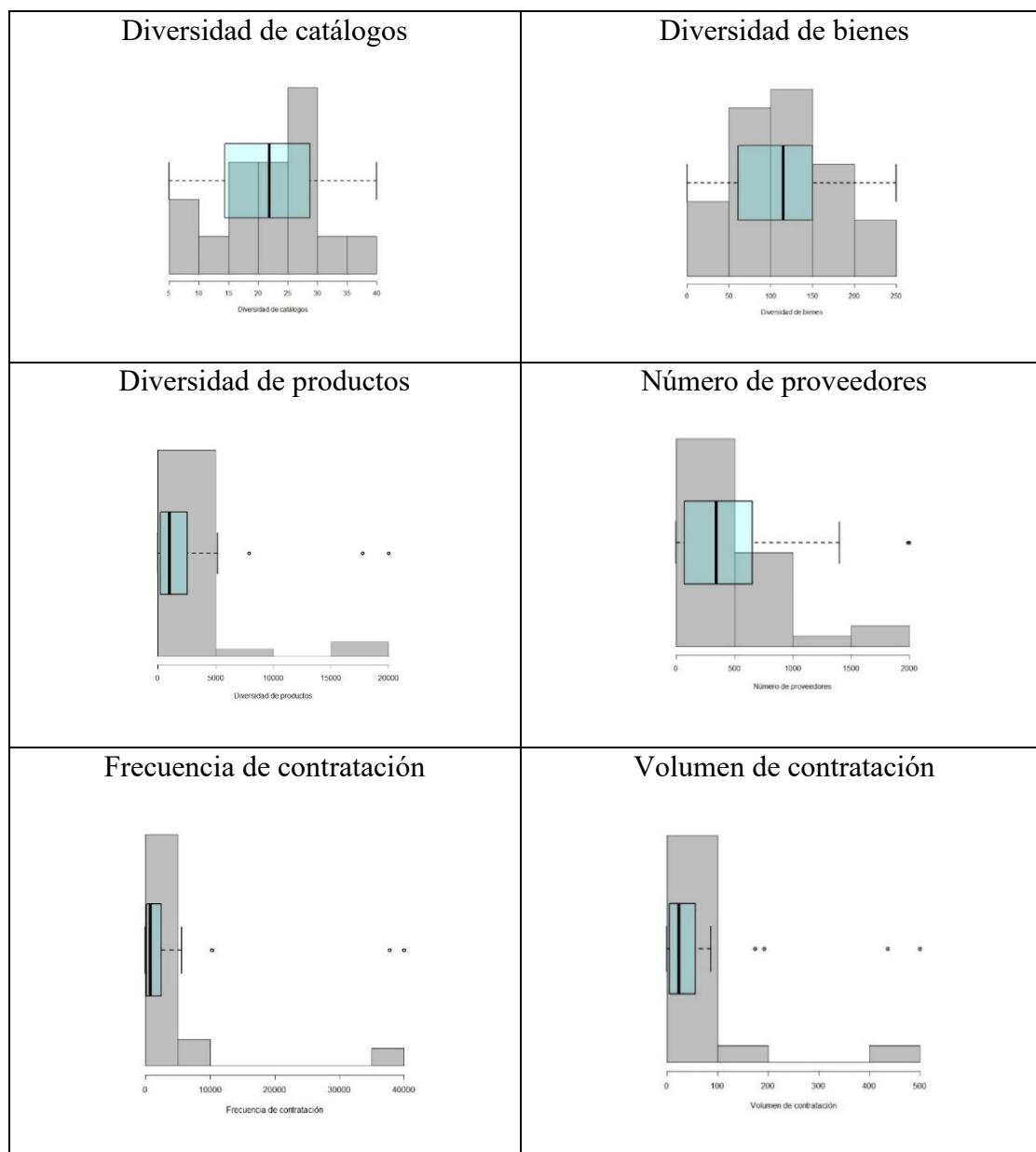
En cuanto al análisis de las órdenes de compra por sector, se observa una gran heterogeneidad, con un mínimo de 47 órdenes y un máximo de 37,480 órdenes. La frecuencia más alta fue de 55 órdenes, lo que señala una tendencia hacia volúmenes de compra moderados. Sin embargo, el percentil 75, situado en 2,167 órdenes, sugiere que una proporción significativa

de sectores lleva a cabo volúmenes de compra notablemente mayores. Esta variabilidad podría estar vinculada al tamaño del sector, su actividad económica y otros factores.

El volumen promedio de contratación de los contratos adjudicados a través de los Catálogos Electrónicos de Acuerdo Marco en 2023 fue de S/ 58.9 millones, con un máximo de S/ 459.9 millones. Esta amplia dispersión señala una gran heterogeneidad en los patrones de compra de los diversos sectores.

**Figura 9.**

*Distribución de histogramas y diagrama de caja*





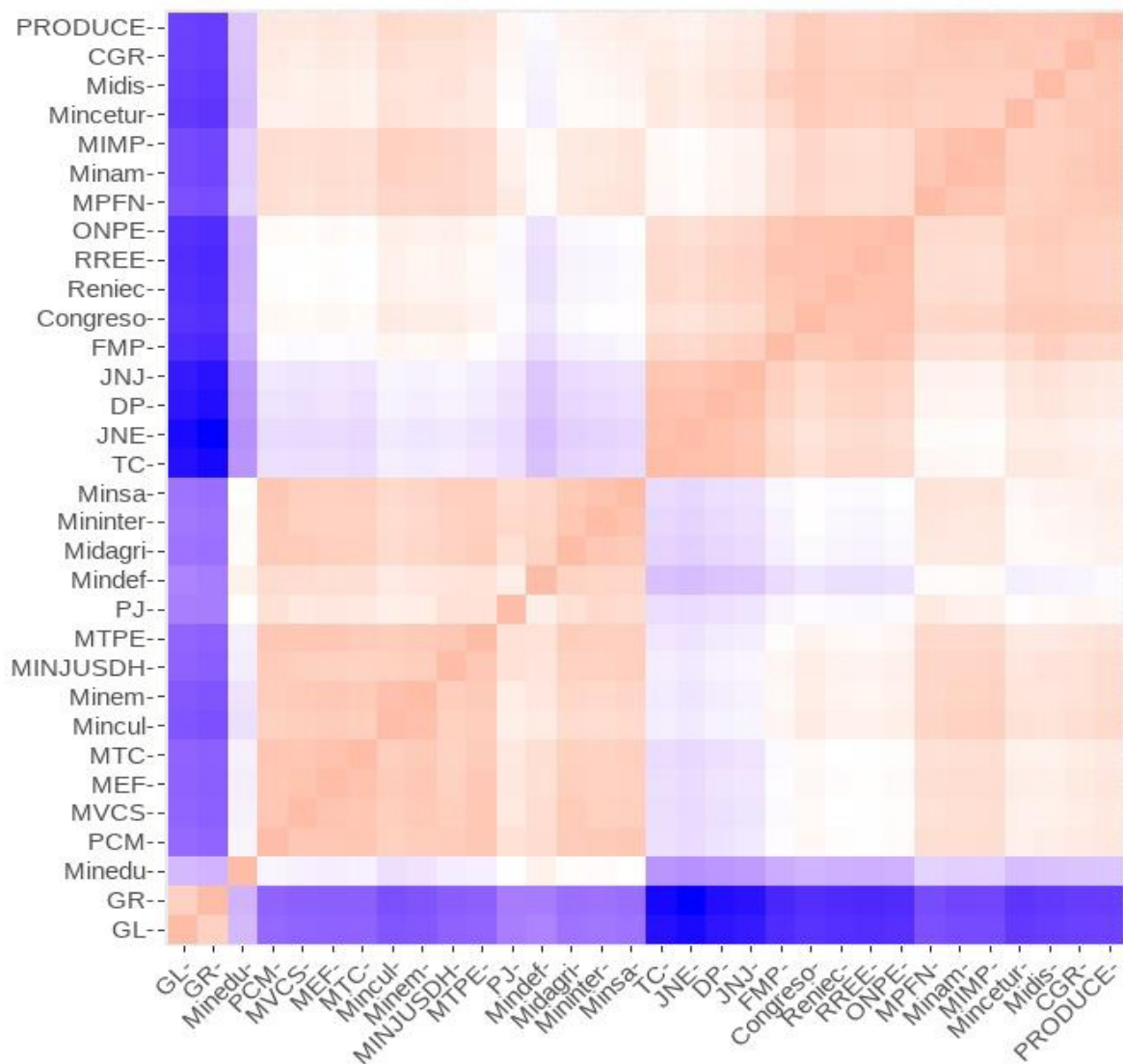
Los datos analizados revelan una variabilidad significativa entre los diferentes sectores, evidenciando una considerable heterogeneidad en las variables estudiadas. Aunque la mayoría de los valores se distribuyen de manera esperada, se detectaron valores atípicos que exceden el rango intercuartílico en algunas variables. Estas anomalías podrían estar influenciadas por factores sectoriales específicos; para más detalles, consulte el anexo C.

**2.3.2.3 Estandarización de los datos.** Dado que las variables se medían en diferentes escalas, se optó por estandarizar las variables a puntuación z, eliminando así los efectos de diferentes escalas y garantizando una contribución igualitaria de cada variable.

**2.3.2.4 Análisis Clúster.** Se calculó la matriz de distancias de los sectores. Tal como se observa en la figura 10, los sectores con alta correlación (representados en tonos rojizos) tienden a agruparse, mientras que aquellos con baja correlación (tonos azulados) se separan en grupos distintos. Los sectores atípicos, con una correlación muy baja con el resto, se identifican por los intensos tonos azules en las celdas correspondientes.

**Figura 10.**

*Matriz de distancias euclidianas*



En cuanto al estadístico de Hopkins, el valor obtenido es 0.9999 indica que los datos tienen una estructura de clustering muy fuerte. Este resultado es una excelente indicación de que los datos están altamente agrupados y es adecuado proceder con técnicas de clustering, ya que probablemente se obtendrán buenos resultados.

En la Tabla 4, se presenta el análisis comparativo entre diversas técnicas de Clustering jerárquico y particional, como Ward, Single, Complete y Average. Los resultados indican que

el enlace 'Ward.D' del Clustering jerárquico ofrece la solución más favorable. Esto sugiere que estas técnicas logran formar grupos que son más homogéneos y claramente definidos.

**Tabla 4.**

*Clúster Jerárquico Aglomerativo: AGNES*

Average	Single	Complete	Ward
0.94	0.92	0.95	0.97

En la Tabla 5, los resultados indican que el método Average presenta la mayor correlación (0.95), sugiriendo que mantiene mejor las relaciones originales entre los datos. A fin de validar esta afirmación y explorar posibles sesgos.

**Tabla 5.**

*Criterios en enlazamiento (Linkage)*

Ward.D	Ward.D2	Single	Complete	Average	Centroid	Mcquitty	Median
0.54	0.82	0.93	0.83	0.95	0.94	0.86	0.88

Además, se analizó la matriz de correlación de cofenética con el fin de evaluar la correspondencia entre la matriz de distancias originales (euclídea) y la matriz de uniones generada por las técnicas de clustering jerárquico, como Ward, Single, Complete, Average, McQuitty, Median y Centroid. De la Tabla 6, se obtuvo que los métodos Ward.D, Ward.D2, Complete y McQuitty presentan una alta correlación, sugiriendo una fuerte tendencia hacia la formación de grupos compactos y jerárquicos, priorizando la minimización de la varianza intra-clúster.

Por el contrario, los métodos Single y Complete exhiben una baja similitud, lo que sugiere que podrían ser más adecuados para identificar estructuras más dispersas o enfoques analíticos distintos. Los métodos Average y Median se sitúan en una posición intermedia, mostrando una correlación moderada con respecto a los demás, lo que los convierte en opciones más versátiles.

La visual de los Dendrogramas (Anexos E-H) confirma la diversidad de resultados, lo que indica la presencia de subestructuras más complejas en los datos y la necesidad de explorar diferentes enfoques de agrupamiento.

**Tabla 6.**

*Análisis de la Matriz de Correlación Cofenética*

	<b>Ward.D</b>	<b>Ward.D2</b>	<b>Single</b>	<b>Complete</b>	<b>Average</b>	<b>Centroid</b>	<b>Mcquitty</b>	<b>Median</b>
<b>Ward.D</b>	1.00	0.89	0.35	0.88	0.53	0.44	0.84	0.80
<b>Ward.D2</b>	0.89	1.00	0.72	1.00	0.85	0.79	0.99	0.98
<b>Single</b>	0.35	0.72	1.00	0.74	0.98	0.99	0.79	0.83
<b>Complete</b>	0.88	1.00	0.74	1.00	0.86	0.81	1.00	0.99
<b>Average</b>	0.53	0.85	0.98	0.86	1.00	0.99	0.90	0.92
<b>Centroid</b>	0.44	0.79	0.99	0.81	0.99	1.00	0.85	0.88
<b>Mcquitty</b>	0.84	0.99	0.79	1.00	0.90	0.85	1.00	1.00
<b>Median</b>	0.80	0.98	0.83	0.99	0.92	0.88	1.00	1.00

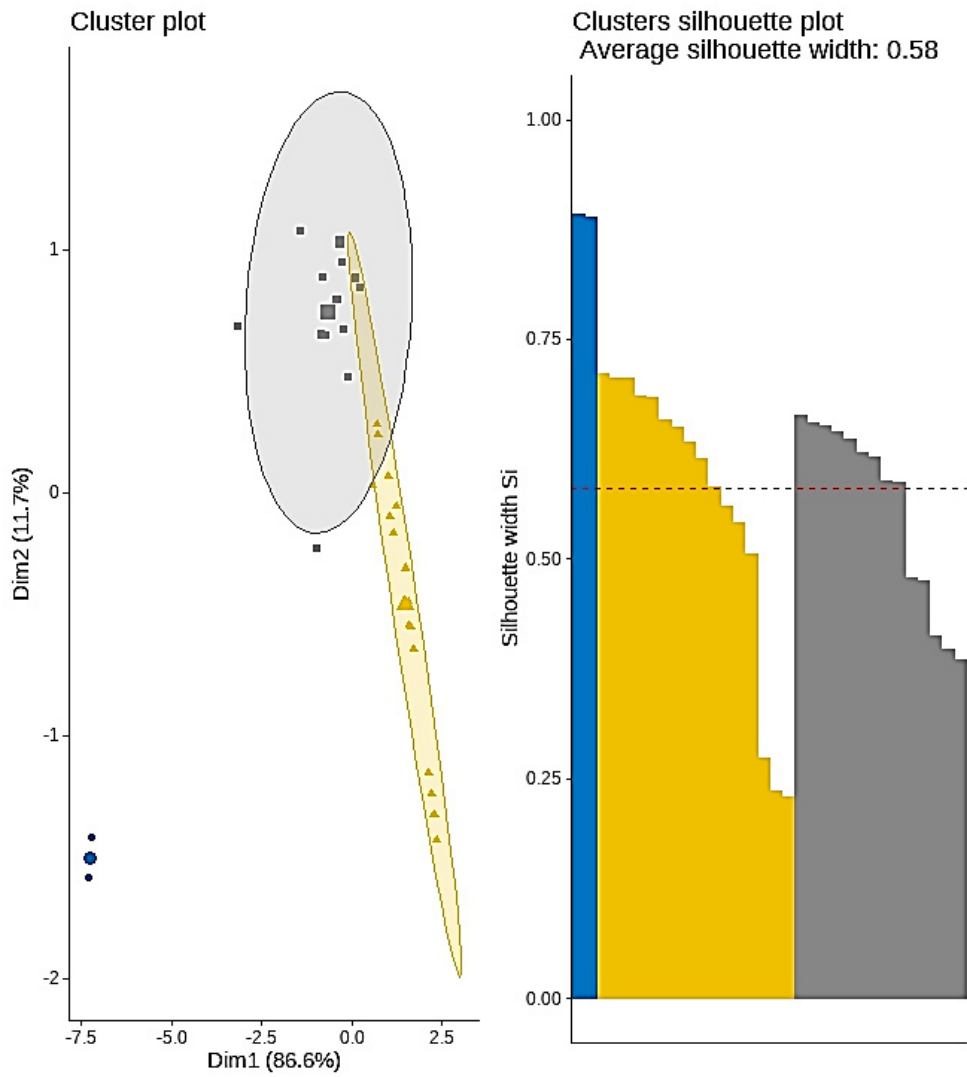
Nota: La matriz muestra la similitud entre métodos de clustering jerárquico. Valores cercanos a 1 indican Dendrograma muy similares, mientras que valores cercanos a 0 indican Dendrograma distintos.

Los resultados del análisis de Clustering no jerárquico, tanto para K-means como para PAM con  $k=3$ , son prometedores. Estos algoritmos explican aproximadamente el 98.3% de la varianza total, lo que indica que los tres clústeres identificados capturan la mayor parte de la estructura subyacente en los datos. La visualización en la Figura 11 corrobora esta interpretación, mostrando una clara separación entre los grupos. No obstante, un análisis más detallado utilizando el coeficiente de silueta muestra cierta ambigüedad en la asignación de algunos puntos. Aunque el valor promedio de 0.58 indica una buena clasificación en general, cerca del 15.0% de los casos presentan valores cercanos a cero, especialmente en la frontera

entre los clústeres 2 y 3, lo que sugiere que estos puntos podrían pertenecer a más de un grupo. Esta superposición podría estar relacionada con la naturaleza intrínseca de los datos y la elección del número de clústeres.

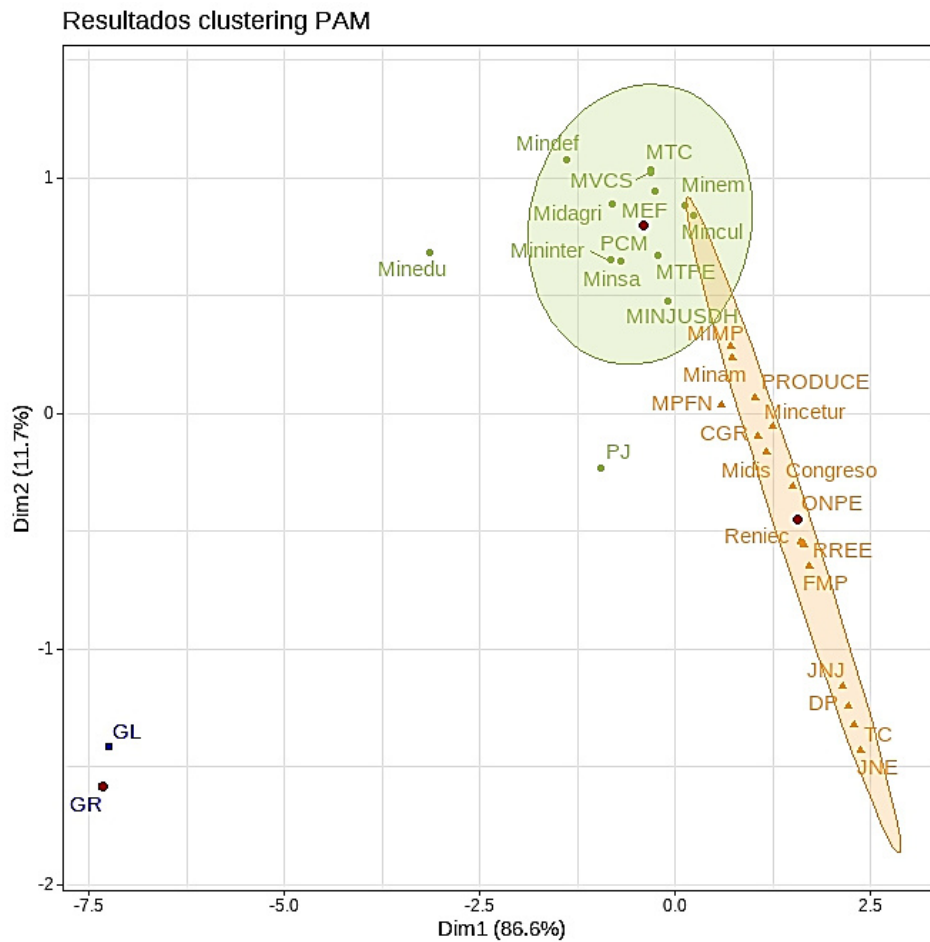
**Figura 11.**

*Análisis de Clustering K-means con  $k=3$*



**Figura 12.**

*Análisis del algoritmo PAM para Clúster con  $k=3$*

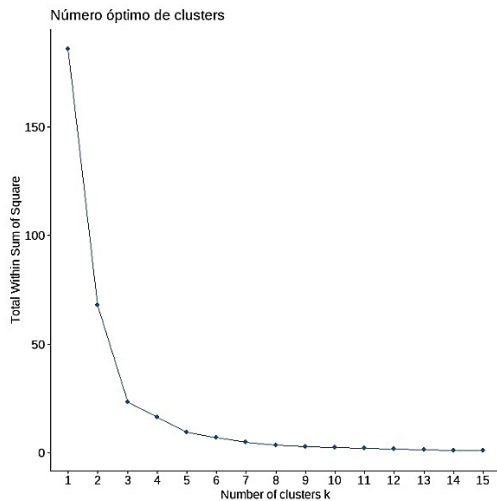


Tras evaluar diferentes métodos, se optó por el método de Ward debido a su capacidad para minimizar la varianza intra-clúster, lo que garantiza grupos más homogéneos y significativos. Esta elección se justifica por la naturaleza de los datos, que presentan variables numéricas continuas, y por el objetivo del estudio, que busca identificar segmentos de sectores públicos con características similares. Además, la comparación con otros métodos de enlace, como Complete y McQuitty, confirmó la robustez y fiabilidad de los resultados obtenidos con el método de Ward.

**2.3.2.5 Evaluar la cantidad óptima de grupos.** Se utilizó una variedad de métodos para determinar el número óptimo de grupos, entre ellos el método del codo, el índice de silueta, el criterio GAP, el índice NbClust y el análisis visual.

**Figura 13.**

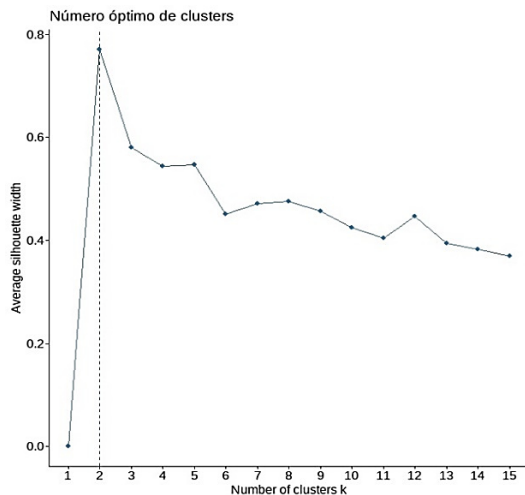
*Método del codo*



De la figura 13, el valor óptimo de k se selecciona en el punto donde la disminución de WCSS se vuelve menos pronunciada a partir de k=3. Otra alternativa corresponde al método Silhouette, este método sugiere una solución de K=2.

**Figura 14.**

*Criterio de la silueta*



Mientras que, la Tabla 7 ofrecen un panorama complejo, al evaluar diversos métodos de clustering y métricas de evaluación se obtuvo resultados variados. Este, sugiere que el método jerárquico con K=2 presenta puntajes competitivos en algunas métricas.

**Tabla 7.**

*Evaluación del método de Clustering*

---

Clustering Methods:  
 hierarchical kmeans diana fanny pam clara agnes

Cluster sizes:  
 2 3 4 5

Validation Measures:

		2	3	4	5
hierarchical	Connectivity	3.8579	6.7869	8.6087	14.1103
	Dunn	0.7869	0.4297	0.3150	0.3705
	Silhouette	0.7701	0.5289	0.5592	0.5471
kmeans	Connectivity	3.8579	5.6798	8.6087	14.1103
	Dunn	0.7869	0.2174	0.3150	0.3705
	Silhouette	0.7701	0.5803	0.5592	0.5471
diana	Connectivity	3.8579	5.6798	8.6087	14.1103
	Dunn	0.7869	0.2174	0.3150	0.3705
	Silhouette	0.7701	0.5803	0.5592	0.5471
fanny	Connectivity	1.8218	5.6798	14.2960	22.7837
	Dunn	0.0953	0.2174	0.0575	0.1279
	Silhouette	0.4505	0.5803	0.4673	0.4188
pam	Connectivity	3.8579	5.6798	13.2357	17.6028
	Dunn	0.7869	0.2174	0.0809	0.1418
	Silhouette	0.7701	0.5803	0.5091	0.4991
clara	Connectivity	3.8579	5.6798	12.9456	17.6028
	Dunn	0.7869	0.2174	0.1418	0.1418
	Silhouette	0.7701	0.5803	0.4606	0.4991
agnes	Connectivity	3.8579	6.7869	8.6087	14.1103
	Dunn	0.7869	0.4297	0.3150	0.3705
	Silhouette	0.7701	0.5289	0.5592	0.5471

Optimal Scores:

	Score	Method	Clusters
Connectivity	1.8218	fanny	2
Dunn	0.7869	hierarchical	2
Silhouette	0.7701	hierarchical	2

---



Mientras que, con el NbClust, se evaluó una amplia gama de índices, tales como el índice de silueta y el índice de Calinski-Harabasz, este indicó la presencia de tres clústeres.

**Figura 15.**

*Índice NbClust del clúster jerárquico Ward*

```
*****  
* Among all indices:  
* 1 proposed 2 as the best number of clusters  
* 14 proposed 3 as the best number of clusters  
* 2 proposed 4 as the best number of clusters  
* 1 proposed 5 as the best number of clusters  
* 1 proposed 7 as the best number of clusters  
* 1 proposed 8 as the best number of clusters  
* 4 proposed 9 as the best number of clusters  
* 3 proposed 10 as the best number of clusters
```

\*\*\*\*\* Conclusion \*\*\*\*\*

```
* According to the majority rule, the best number of clusters is 3
```

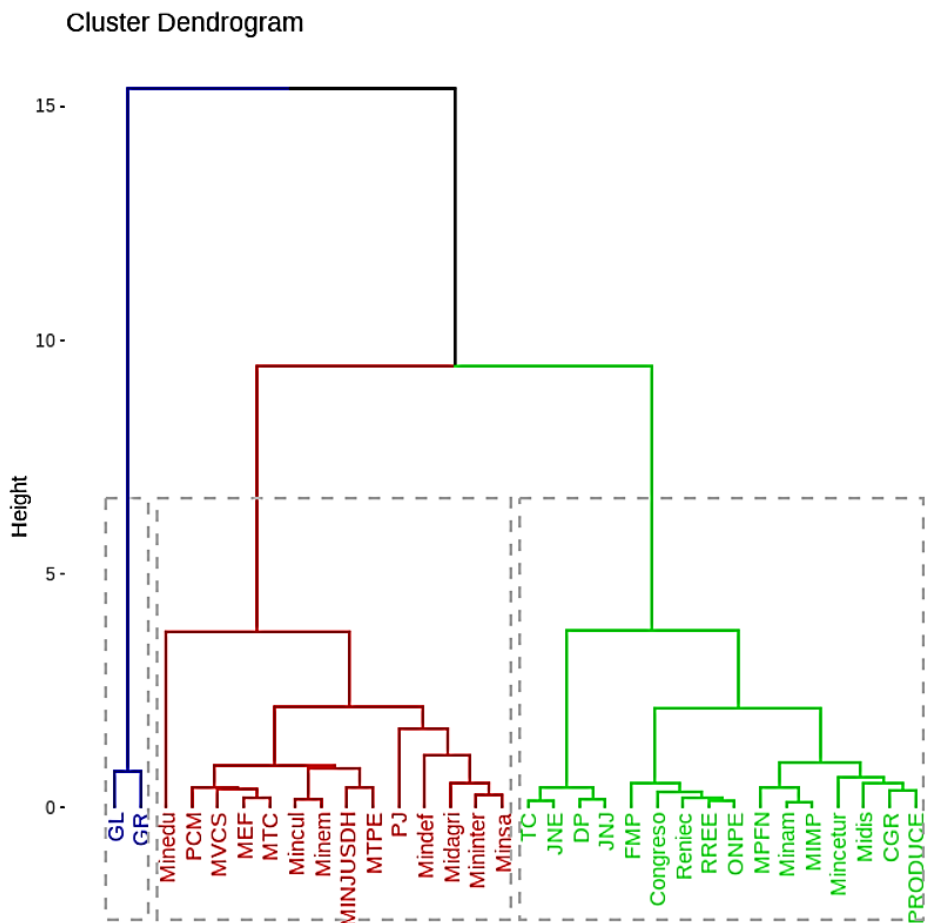
```
*****
```

Considerando conjuntamente los resultados del índice de validación NbClust, el análisis visual de los dendrogramas y los objetivos de nuestra investigación se optó por el método de Ward. Esta decisión se fundamenta en la capacidad de este método para minimizar la varianza intra-cluster, lo cual se alinea perfectamente con la naturaleza de nuestros datos y con nuestra intención de formar grupos altamente homogéneos.

**2.3.2.6 Obtención de los clústeres.** Al aplicar el método de Ward, se identificaron tres segmentos de sectores con perfiles de contratación diferenciados. Según la Figura 16, el clúster más grande (rojo) agrupa principalmente a ministerios (16 sectores, 50.0% del total). El clúster verde (14 sectores, 43.7% del total) incluye a entidades como el Congreso y el Tribunal Constitucional]. Por último, el clúster azul, conformado por solo dos sectores (GL y GR), representando el 6.3% restante (2 sectores). Basándonos en la Figura 17, se observa que los datos presentan una cierta heterogeneidad, ya que los tres clústeres se encuentran bastante separados entre sí.

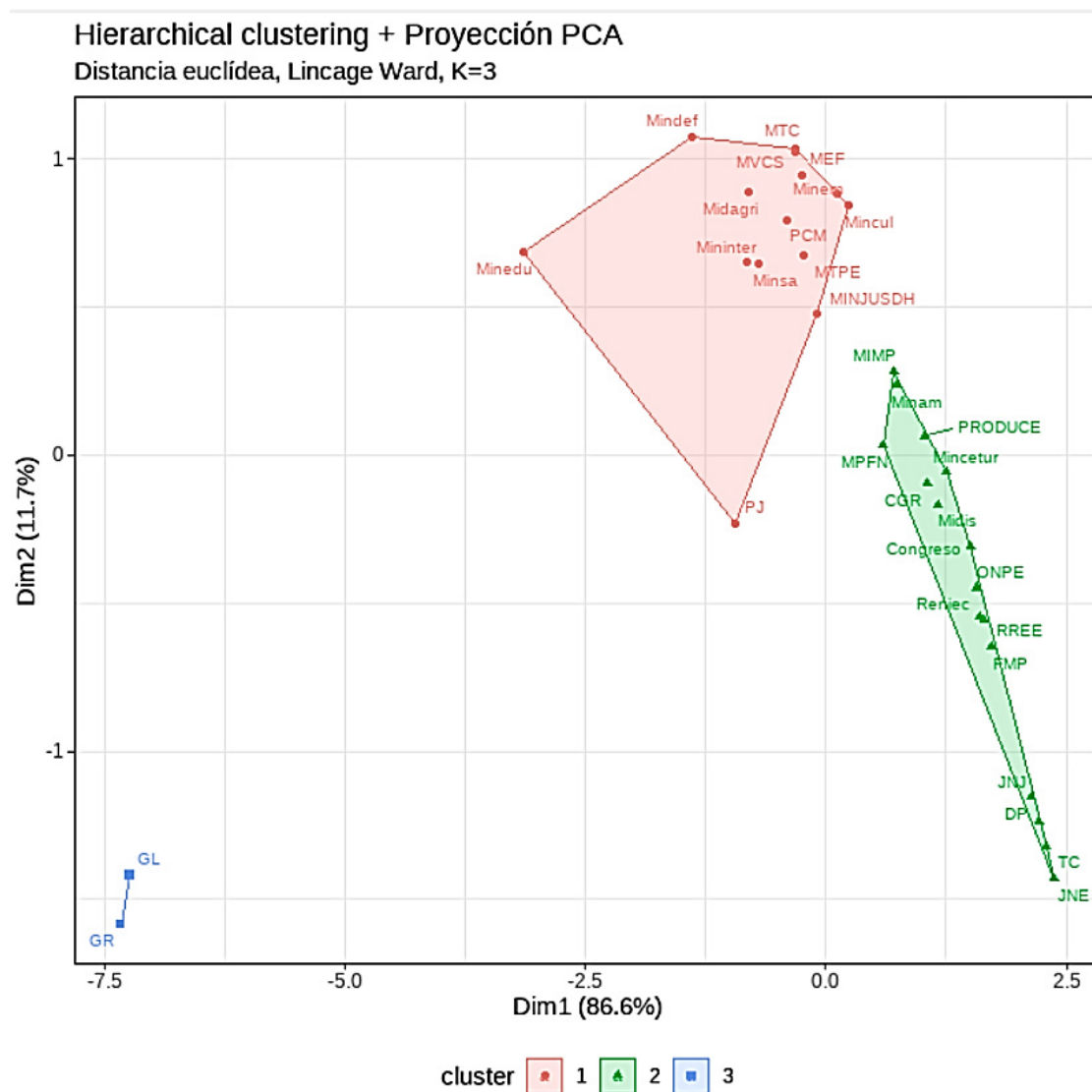
**Figura 16.**

*Dendrograma obtenido mediante el método jerárquico aglomerativo de Ward, con  $k=3$*



**Figura 17.**

*Análisis de Clustering Jerárquico de Sectores Públicos con Visualización PCA*



**Validación de los clústeres.** Se utilizó el paquete cIValid para comparar simultáneamente múltiples algoritmos. De la Tabla 8, se obtuvo que el método jerárquico con 2 clústeres exhibe la menor conectividad entre ellos, respaldado principalmente por los índices de silueta y Dunn, *sin embargo*, los índices APN, AD, ADM y FOM no brindan una evidencia concluyente. Por el contrario, sugiere que la segmentación obtenida no depende de un algoritmo específico, sino que refleja una partición genuina de los datos. La notable cohesión interna de los clústeres y la adecuada separación

entre ellos, respaldadas por los valores obtenidos de los índices Silhouette, Davies-Bouldin y Calinski-Harabasz, refuerzan esta interpretación, tabla 9.

Tabla 8.

Validación interna de los Clústeres

---

Clustering Methods:  
hierarchical kmeans pam

Cluster sizes:  
2 3 4 5 6

Validation Measures:

		2	3	4	5	6
hierarchical	APN	0.0101	0.0000	0.0000	0.0192	0.0503
	AD	1.7099	1.4926	0.8266	0.6838	0.6154
	ADM	0.0694	0.0000	0.0000	0.1237	0.1357
	FOM	0.6050	0.4911	0.2848	0.2833	0.2651
	Connectivity	3.8579	6.7869	8.6087	14.1103	16.9393
	Dunn	0.7869	0.4297	0.3150	0.3705	0.4035
	Silhouette	0.7701	0.5289	0.5592	0.5471	0.5049
kmeans	APN	0.0302	0.0000	0.0700	0.0843	0.1007
	AD	1.7696	0.9656	0.8902	0.7163	0.6323
	ADM	0.2083	0.0000	0.1632	0.2101	0.1881
	FOM	0.6291	0.3519	0.3048	0.2970	0.2627
	Connectivity	3.8579	5.6798	8.6087	14.1103	16.9393
	Dunn	0.7869	0.2174	0.3150	0.3705	0.4035
	Silhouette	0.7701	0.5803	0.5592	0.5471	0.5049
pam	APN	0.0778	0.0098	0.0559	0.0497	0.0595
	AD	1.7969	0.9750	0.7966	0.6998	0.5373
	ADM	0.3044	0.0247	0.0945	0.1258	0.0805
	FOM	0.6811	0.3533	0.3233	0.3123	0.2257
	Connectivity	3.8579	5.6798	13.2357	17.6028	20.5317
	Dunn	0.7869	0.2174	0.0809	0.1418	0.2631
	Silhouette	0.7701	0.5803	0.5091	0.4991	0.5207

Optimal Scores:

	Score	Method	Clusters
APN	0.0000	hierarchical	3
AD	0.5373	pam	6
ADM	0.0000	hierarchical	3
<del>FOM</del>	<del>0.2257</del>	<del>pam</del>	<del>6</del>
Connectivity	3.8579	hierarchical	2
Dunn	0.7869	hierarchical	2
Silhouette	0.7701	hierarchical	2

---

**Tabla 9.***Análisis de la calidad de los clústeres utilizando múltiples Índices*

Clúster	Algoritmos	Índice Silhouette	Índice Davies-Bouldin	Índice Calinski-Harabasz
<b>k=2</b>	Hierarchical	0.77	0.24	0.79
	kmeans	0.77	0.24	0.79
	PAM	0.77	0.24	0.79
<b>k=3</b>	Hierarchical	0.52	0.54	0.22
	kmeans	0.58	0.54	0.22
	PAM	0.58	0.54	0.22
<b>k=4</b>	Hierarchical	0.56	0.51	0.20
	kmeans	0.56	0.51	0.20
	PAM	0.58	0.56	0.08

Nota: Se presenta los resultados de aplicar tres algoritmos de clustering (jerárquico, k-means y PAM), variando el número de clústeres (k=2, 3 y 4).

**2.3.2.7 Caracterización de los clústeres.** Se identificó características particulares y tendencias que distinguen a cada clúster, el resultado obtenido es el siguiente:

- El Clúster 1, denominado 'Núcleo de Gobierno', agrupa a los ministerios y organismos con mayor poder de decisión en la gestión pública. A pesar de su rol central en la definición y ejecución de políticas, el análisis del Clúster 1, revela un perfil de actividad que podría considerarse intermedio en comparación con otros clústeres. Estos sectores presentan una diversidad promedio de 28 catálogos, 151 bienes contratados y 2,534 fichas de productos, lo que indica una actividad moderada. Además, negocian con un promedio de 600 proveedores y realizan 2,849 transacciones anuales, generando un volumen de contratación promedio de S/. 62 millones.
- El Clúster 2, denominado 'Control, justicia, administración pública, y desarrollo social', agrupa a los ministerios y organismos encargados de funciones altamente especializadas, lo

que refleja una menor diversificación en sus catálogos y productos contratados. Con un promedio de 16 catálogos, 86 bienes y 363 productos contratados, este clúster muestra una actividad más focalizada. Asimismo, el número reducido de proveedores (130) y transacciones (245 anuales) con un volumen promedio de S/. 10 millones, sugiere operaciones más específicas y con un alcance más limitado, lo cual es coherente con su rol institucional.

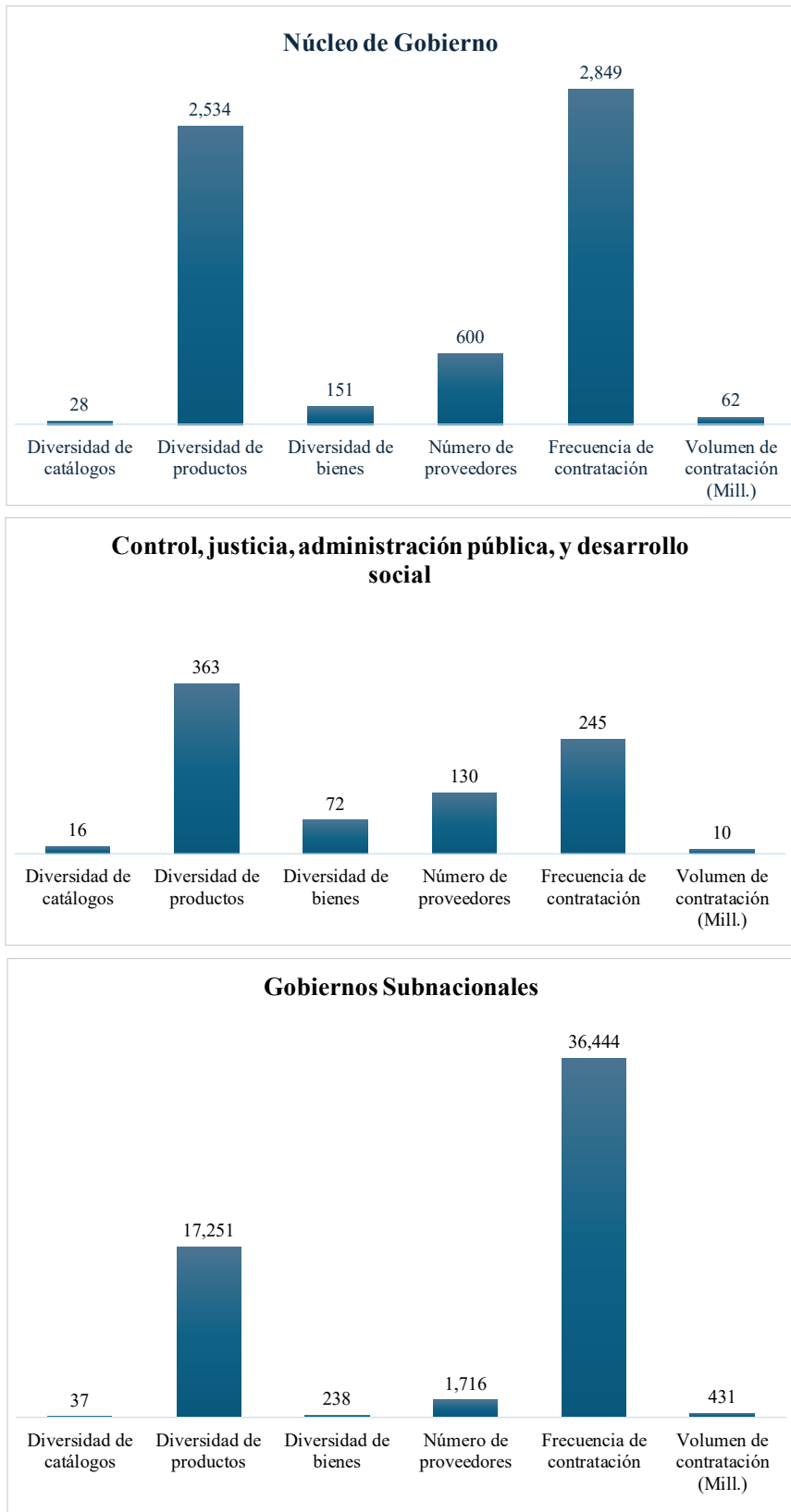
- El Clúster 3, identificado como "Gobiernos Subnacionales," agrupa diversas entidades públicas tanto regionales como locales, caracterizadas por una notable diversidad de necesidades y un volumen de contratación significativamente mayor en comparación con los demás clústeres. Esta diversidad se manifiesta en un promedio de 37 catálogos utilizados, 238 bienes y 17,251 productos contratados, lo que refleja una amplia gama de productos adquiridos. Asimismo, el alto número de proveedores (1,715) y transacciones (36,444 anuales), junto con un volumen promedio de S/. 431 millones, destaca su rol clave en la provisión de servicios públicos a nivel local.

**Tabla 10.***Clasificación de Entidades Públicas*

<b>Clúster</b>	<b>Sector Público</b>
<b>1: Núcleo de Gobierno</b>	Agricultura
	Cultura
	Defensa
	Economía y Finanzas
	Educación
	Energía y Minas
	Interior
	Justicia
	Poder Judicial
	Presidencia Consejo De Ministros
	Salud
	Trabajo y Promoción del Empleo
	Transportes y Comunicaciones
Vivienda Construcción y Saneamiento	
<b>2: Control, justicia, administración pública, y desarrollo social</b>	Ambiental
	Comercio Exterior y Turismo
	Congreso de la República
	Contraloría General
	Defensoría del Pueblo
	Desarrollo E Inclusion Social
	Fuero Militar Policial
	Junta Nacional de Justicia
	Jurado Nacional de Elecciones
	Ministerio de la Mujer y Poblaciones Vulnerables
	Ministerio Público
	Oficina Nacional de Procesos Electorales
	Producción
	Registro Nacional de Identificación y Estado Civil
	Relaciones Exteriores
Tribunal Constitucional	
<b>3: Gobiernos Subnacionales</b>	Gobierno Local
	Gobiernos Regionales

**Figura 18.**

*Análisis de Clustering Jerárquico*





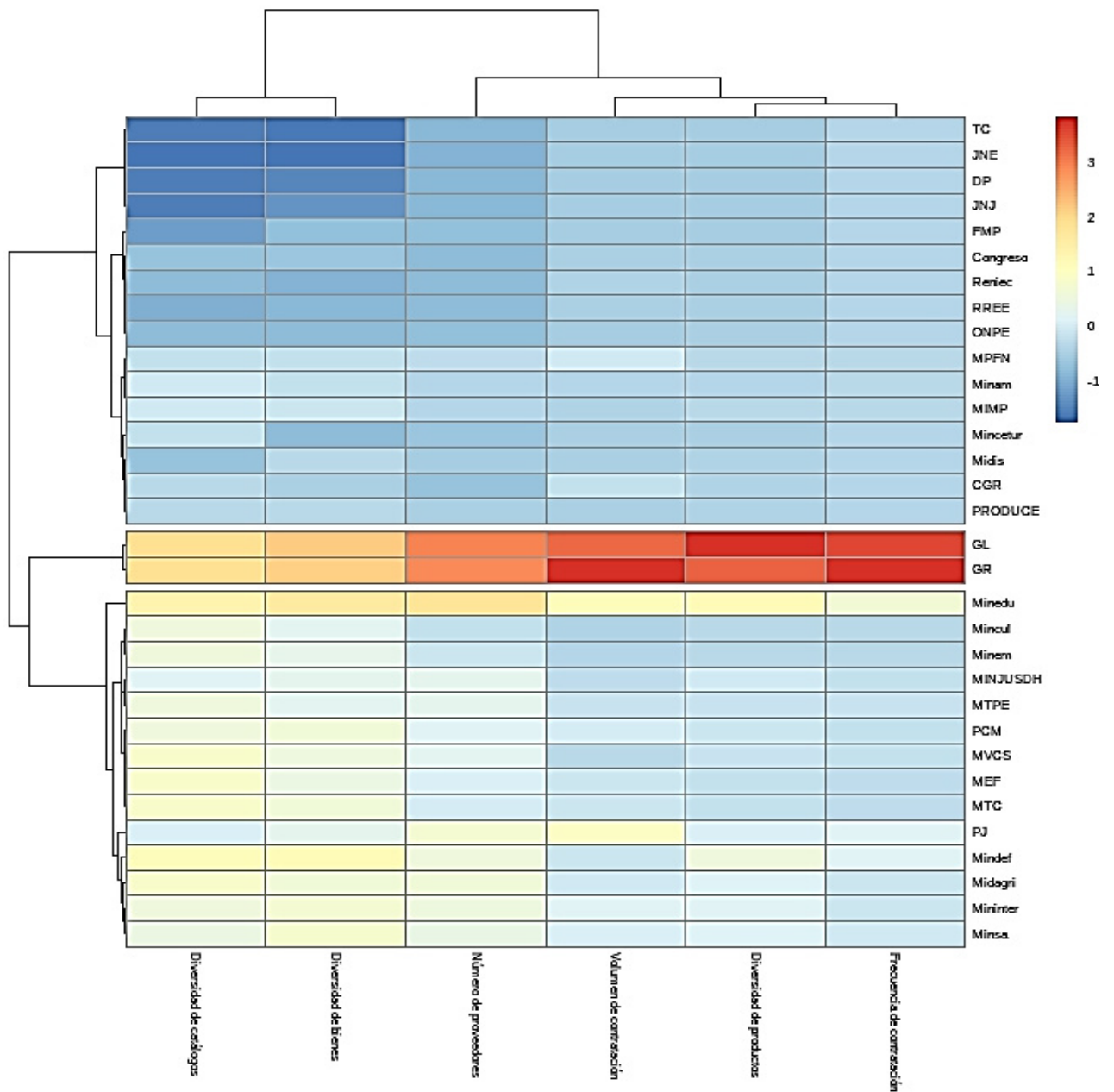
El mapa de calor (Figura 19) confirma visualmente esta segmentación de cada sector. Los colores más cálidos (rojos) indican valores altos en las variables analizadas, mientras que los colores más fríos (azules) representan valores bajos.

- Clúster 1: Núcleo de Gobierno
  - o Frecuencia de Contratación: Alta frecuencia de contratación, lo que indica un mayor ritmo de adquisición de bienes
  - o Volumen de Contratación: Alto volumen de contratación, reflejando grandes presupuestos y necesidades extensas.
  - o Diversidad de Catálogos: Amplia variedad de catálogos utilizados, abarcando múltiples categorías de bienes y productos
  - o Proveedores: Gran número de proveedores, incluyendo tanto Lima como otras regiones
- Clúster 2: Control, justicia, administración pública, y desarrollo social
  - o Frecuencia de Contratación: Moderada frecuencia de compras, enfocada en necesidades específicas y especializadas.
  - o Volumen de Contratación: Volumen moderado, con compras significativas, pero menos frecuentes
  - o Diversidad de Catálogos: Uso de catálogos especializados, centrados en bienes específicos para sus funciones.
  - o Proveedores: Menor número de proveedores
- Clúster 3: Gobiernos Subnacionales
  - o Frecuencia de Contratación: Variable, dependiendo de las necesidades locales.
  - o Volumen de Contratación: Volumen bajo a moderado, reflejando presupuestos más limitados.

- o Diversidad de Catálogos: Uso de catálogos variados, pero menos extensos que el Núcleo de Gobierno.
- o Proveedores: Menor número de proveedores, con una tendencia a utilizar proveedores locales.

**Figura 19.**

*Análisis de clústeres mediante mapas de calor en contrataciones públicas electrónicas*



Para optimizar la gestión de compras, se propone adaptar las políticas y herramientas de PERU COMPRAS a las características específicas de cada clúster. Por ejemplo, para los sectores con alto volumen de compras (como el “Núcleo de Gobierno”), podrían implementarse mecanismos de compra centralizada, mientras que para los sectores con bajo volumen (como algunos de los “Gobiernos Subnacionales”), se podrían simplificar los procesos de contratación.

### III. APORTES MÁS DESTACABLES A LA EMPRESA

En mi rol como miembro del equipo de la Central de Compras Públicas, mis principales contribuciones fueron:

- Participé en el análisis de datos del proyecto estratégico de Perú Compras, enfocado en mapear necesidades de abastecimiento y definir estrategias diferenciadas para mejorar adquisiciones públicas, optimizando procesos, reduciendo riesgos y generando ahorros.
- Mediante el análisis de datos históricos de contrataciones estatales, identifiqué patrones y tendencias en el comportamiento de compra. Con base en estos hallazgos, propuse criterios objetivos para optimizar los procesos de contratación electrónica, tales como montos mínimos de contratación, tiempos máximos de inactividad de productos y límites de precios unitarios. Estas recomendaciones, respaldadas por técnicas estadísticas, permitieron una toma de decisiones más estratégica y eficiente, alineada con los objetivos de la empresa y mejorando la satisfacción del cliente
- Colaboré en el desarrollo de un módulo de encuestas automatizadas para medir la satisfacción de usuarios de catálogos electrónicos.
- Realicé un aporte integral en el proceso estadístico, abarcando desde la recopilación y limpieza de datos hasta el análisis, interpretación y presentación de resultados en diversos estudios
- Colaboré en los proyectos destacados de Buenas Prácticas en Gestión Pública, como el Cotizador Electrónico y PERÚ OBSERVA, los cuales fueron premiados por Ciudadanos al Día (CAD). Además, participé en el proyecto El Camino Estratégico para Compras Públicas Eficientes: Perú Compras, reconocido por la Universidad del Pacífico.

## IV. CONCLUSIONES

Este estudio se basa en datos de un período determinado, por lo que los resultados obtenidos pueden estar sujetos a variaciones debido a cambios en el contexto económico y político. A continuación, se presentan los hallazgos más significativos:

- Para desarrollar una metodología robusta para segmentar los sectores de contratación pública, se siguen estos pasos clave: definir los objetivos del análisis, recopilar y preparar los datos, realizar un análisis exploratorio para comprender las variables, aplicar técnicas de segmentación como el Clustering, y finalmente, validar e interpretar los resultados para perfilar los clústeres.
- Al evaluar diferentes métodos Clustering, se optó por el método de Ward debido a su capacidad para minimizar la varianza intra-clúster, lo que garantiza grupos más homogéneos y significativos. Esta elección se justifica por la naturaleza de los datos, que presentan variables numéricas continuas, y por el objetivo del estudio, que busca identificar segmentos de sectores públicos con características similares para optimizar los procesos de contratación. Además, la comparación con otros métodos de enlace, como Complete y McQuitty, confirmó la robustez y fiabilidad de los resultados obtenidos con el método de Ward.
- Tras aplicar varios métodos de análisis, como el método del codo, el índice de silueta, el criterio GAP y el índice NbClust, se concluyó que la solución óptima consiste en tres clústeres. Este enfoque permitió identificar perfiles claramente diferenciados entre los grupos, basados en las variables evaluadas.
- Respecto a la caracterización de los clústeres obtenidos, se obtuvo:
  - Clúster 1, denominado " Núcleo de Gobierno" revela un perfil de actividad que podría considerarse intermedio en comparación con otros clústeres. Estos sectores presentan

una diversidad promedio de 28 catálogos, 151 bienes contratados y 2,534 fichas de productos, lo que indica una actividad moderada. Además, negocian con un promedio de 600 proveedores y realizan 2,849 transacciones anuales, generando un volumen de contratación promedio de S/. 62 millones.

- El Clúster 2, denominado "Control, justicia, administración pública, y desarrollo social", refleja una menor diversificación en sus catálogos y productos contratados. Con un promedio de 16 catálogos, 86 bienes contratados y 363 productos, este clúster muestra una actividad más focalizada. Asimismo, el número reducido de proveedores (130) y transacciones (245 anuales) con un volumen promedio de S/. 10 millones, sugiere operaciones más específicas y con un alcance más limitado, lo cual es coherente con su rol institucional.
- El Clúster 3, denominado "Gobiernos Subnacionales", agrupa una amplia variedad de entidades públicas regionales y locales, caracterizadas por una alta diversidad de necesidades y un volumen de contratación significativamente superior al resto de los clústeres. Esta diversidad se refleja en un promedio de 37 catálogos, 238 bienes contratados y 17,251 fichas de productos, lo que indica una amplia gama de producto adquiridos. Asimismo, el alto número de proveedores (1,715) y transacciones (36,444 anuales), junto con un volumen promedio de S/. 431 millones, destaca su rol clave en la provisión de servicios públicos a nivel local.

## V. RECOMENDACIONES

Para asegurar una gestión eficaz y adaptada a los cambios en las dinámicas de compra, se recomienda:

- Actualizar la segmentación de los clústeres debido a que los sectores no son estáticos y pueden evolucionar con el tiempo por cambios en las dinámicas de compra, la incorporación de nuevos productos, variaciones en el mercado o en las necesidades institucionales.
- Implementar estrategias personalizadas para cada uno de los tres clústeres identificados. Esta segmentación permitirá adaptar las acciones a las características específicas de cada grupo, considerando la dinámica de los mercados y el entorno legal.
- Monitorizar los patrones de compras de los sectores para adaptar las estrategias de manera proactiva, asegurando que las políticas estén alineadas con las condiciones actuales y futuras, contribuyendo a una gestión más eficiente.
- Ampliar el alcance del estudio para incluir todas las entidades públicas que realizan adquisiciones de bienes mediante catálogos electrónicos durante un periodo de tiempo más extenso, con el fin de obtener una visión más detallada de los patrones de contratación. Además, sería útil incorporar información más precisa sobre los bienes adquiridos para identificar características específicas por sector o entidad.

## VI. REFERENCIAS

- Bray R. y Curtis J. (1957). *An ordination of the upland forest communities of southern Wisconsin*. Ecological Monographs, 27(4), pp. 325–349.  
<https://doi.org/10.2307/1942268>
- Correa M., Almeida B., Espinoza M., Cabezas A. y Castillo K. (2024). *Segmentación Efectiva de Clientes utilizando R: Técnicas para Administración y Marketing Avanzado*.  
[https://www.researchgate.net/publication/377189890\\_Segmentacion\\_Efectiva\\_de\\_Clientes\\_utilizando\\_R\\_Tecnicas\\_para\\_Administracion\\_y\\_Marketing\\_Avanzado](https://www.researchgate.net/publication/377189890_Segmentacion_Efectiva_de_Clientes_utilizando_R_Tecnicas_para_Administracion_y_Marketing_Avanzado)
- Cuadras, C. (1989). *Distancias Estadísticas*. Estadística Española, 30, pp. 295–378.  
<https://doi.org/10.1016/B978-0-444-88029-1.50047-2>
- De la Fuente Crespo, J. (2000). *Análisis Conglomerados Clúster*. Universidad autónoma de Madrid, pp. 3-8. [https://www.estadistica.net/Master-Econometria/Analisis\\_Cluster.pdf](https://www.estadistica.net/Master-Econometria/Analisis_Cluster.pdf)
- Everitt, B., Landau, S., Leese, M. y Stahl, D. (2011). *Cluster Analysis*. Reino Unido. (5ª ed.). Wiley.
- Gallardo, J. A., Gutiérrez, R., González, A. y Torres, F. (1994). *Técnica de análisis de datos multivariable*. Tratamiento computacional. Universidad de Granada.
- Gower, J. (1985). *Measures of similarity, dissimilarity and distance*. Encyclopedia of Statistics (5ª ed). Wiley.
- Gower, J., y Legendre, P. (1986). *Metric and Euclidean properties of dissimilarity coefficients*. Journal of Classification 3, pp. 5–48. <https://doi.org/10.1007/BF01896809>



Hair, J., Black, W., Babin, B. y Anderson, R. (2010). *Análisis multivariante*. (5ª ed). Pearson Educación.

Han, J. y Kamber, M. (2001). *Data Mining: Concepts and Techniques*. (3ª ed). Morgan Kaufmann Publishers.

Jain, A. (2010). *Data Clustering: 50 Years Beyond K-means*. Pattern Recognition Letters, 31(8), (pp. 651-666). <https://doi.org/10.1016/j.patrec.2009.09.011>.

Kaufman, L. y Rousseeuw, P. (1990). *Finding Groups in Data*. Nueva York: Wiley & Sons Inc. <https://doi.org/10.1002/9780470316801>

King, B. (1967). *Step-wise clustering procedure*. *Journal of the American Statistical Association*. Wiley, 69, pp. 86-101. DOI:[10.1080/01621459.1967.10482890](https://doi.org/10.1080/01621459.1967.10482890)

Kotler, P., y Keller, K. L. (2012). *Dirección de marketing*. México D.F.: (12ª ed). Pearson Educación.  
[https://books.google.com.pe/books?hl=es&lr=&id=CoHT8SmJVDQC&oi=fnd&pg=PP33&dq=Kotler,+P.,+%26+Keller,+K.+L.+\(2012\).+Direcci%C3%B3n+de+marketing.+M%C3%A9xico+D.F.:+Pearson.&ots=17uKXzCQXj&sig=ViLhJccGTex2PLI6WGVEtxReiUU#v=onepage&q&f=false](https://books.google.com.pe/books?hl=es&lr=&id=CoHT8SmJVDQC&oi=fnd&pg=PP33&dq=Kotler,+P.,+%26+Keller,+K.+L.+(2012).+Direcci%C3%B3n+de+marketing.+M%C3%A9xico+D.F.:+Pearson.&ots=17uKXzCQXj&sig=ViLhJccGTex2PLI6WGVEtxReiUU#v=onepage&q&f=false)

Kotler, P., Keller, K. L., y Pineda Ayala, L. E. (2016). *Dirección de marketing*. México, D.F.: (15ª ed). Pearson Educación.

Lance, G. y Williams, W. (1966). *Una teoría general de las estrategias de ordenamiento clasificatorio*. 1. Sistemas jerárquicos. *Revista informática*,9(4), (pp. 373-380). <https://api.semanticscholar.org/CorpusID:51476597>

- Legendre, P., Dallot, S., y Legendre, L. (1985). *Succession of species within a community: Chronological clustering, with applications to marine and freshwater zooplankton*. American Naturalist, The University of Chicago Press, 125(2), (pp. 257-288).  
<http://www.jstor.org/stable/2461635>
- Legendre, P. y Legendre, L. (1979). *Ecología digital* (Vols. 1 y 2). Masson, París. (197 pp., 247 pp.)
- Lira A. (2018). *Determinación de patrones de comportamiento de consumo de agua potable con algoritmos de clusterización en la provincia de Andahuaylas* [Tesis de pregrado, Universidad Nacional José María Arguedas]. Repositorio Institucional UNAJMA.  
[https://repositorio.unajma.edu.pe/bitstream/handle/20.500.14168/440/Anai\\_Luz\\_Amparo\\_Tesis\\_Bachiller\\_2018.pdf?sequence=1&isAllowed=y](https://repositorio.unajma.edu.pe/bitstream/handle/20.500.14168/440/Anai_Luz_Amparo_Tesis_Bachiller_2018.pdf?sequence=1&isAllowed=y)
- MacQueen, J. (1967). *Some methods for classification and analysis of multivariate observations*. In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability (Vol. 1, pp. 281–297). University of California Press, Berkeley. <http://projecteuclid.org/euclid.bsm/1200512992>
- Mahalanobis, P. (1936). *On the generalized distance in statistics*. Proceedings of the National Institute of Sciences of India, 2(1), (pp. 49-55).
- Murtagh, F. (1983). *A survey of recent advances in hierarchical clustering algorithms which use cluster centers*. Journal of Computational. The Computer Journal, 26(4), (pp. 354–359). DOI:[10.1093/comjnl/26.4.354](https://doi.org/10.1093/comjnl/26.4.354)
- Ochoa, L., Paredes, K. y Tejada, J. (2017). *Estudio comparativo de técnicas no supervisadas de minería de datos para segmentación de alumnos* (pp. 115-120). En 15° Multi-Conferencia Internacional de LACCEI en Ingeniería, Educación y Tecnología:

"Alianza Global para el Desarrollo y la Educación en Ingeniería".

DOI:[10.18687/LACCEI2017.1.1.115](https://doi.org/10.18687/LACCEI2017.1.1.115)

Pastrán L.y Gongora S. (2021). *Algoritmo de selección y validación del método de clusterización óptimo para datos no supervisados* [Tesis de maestría, Universidad Tecnológica de Pereira]. Repositorio Institucional UTP. <https://hdl.handle.net/11059/13672>

Pearson, K. (1926). *On the coefficient of racial likeness*. Biometrika, (18), (pp. 337–343). <https://doi.org/10.1093/biomet/18.1-2.105>

Perú Compras. (2021). *Resolución Jefatural N.º 139-2021-PERÚ COMPRAS*. (09 de julio de 2021). Central de Compras Públicas ([https://cdn.www.gob.pe/uploads/document/file/2010369/RJ\\_139\\_2021.pdf?v=1728581172](https://cdn.www.gob.pe/uploads/document/file/2010369/RJ_139_2021.pdf?v=1728581172))

Reglamento de Organización y Funciones (ROF) de la Central de Compras Públicas (Perú 31 de diciembre de 2018). *Compras*. Central de Compras Públicas [https://cdn.www.gob.pe/uploads/document/file/534208/NUEVO\\_ROF\\_Y\\_ANEXOS.pdf?v=1668554716](https://cdn.www.gob.pe/uploads/document/file/534208/NUEVO_ROF_Y_ANEXOS.pdf?v=1668554716)

Reglamento de la Ley N° 30225, *Ley de Contrataciones del Estado* (31 de diciembre de 2018). Decreto Supremo N° 344-2018. Diario Oficial El Peruano.

Rousseeuw, P y Kaufman, L. (1987). *Clustering by means of medoids*. Dodge, (pp. 405-416).

Salazar, B., Espinoza, M., Cabezas, A. y Castillo, K. (2024). *Segmentación Efectiva de Clientes utilizando R: técnicas para Administración y Marketing Avanzado*. DOI:[10.62131/978-9942-7173-2-0](https://doi.org/10.62131/978-9942-7173-2-0)

Sharma, S. (1996). *Applied multivariate techniques*. New York: John Wiley & Sons. 6(8), pp.185-236. [https://es.scribd.com/document/387117643/Applied-Multivariate-](https://es.scribd.com/document/387117643/Applied-Multivariate-Techniques-Sharma-pdf)

[Techniques-Sharma-pdf](https://es.scribd.com/document/387117643/Applied-Multivariate-Techniques-Sharma-pdf)

Tripathi, S., Bhardwaj, A., y E, P. (2018). *Enfoques de la agrupación en clústeres en la segmentación de clientes*. Revista Internacional de Ingeniería y Tecnología, 7(3.12), (pp. 802-807).DOI:[10.14419/ijet.v7i3.12.16505](https://doi.org/10.14419/ijet.v7i3.12.16505)

Ventura, C., y Fernández, D. (2020). *Evaluación de análisis de clustering jerárquico en datos moleculares de alta dimensión* [Tesis de maestría, Universidad Oberta de Catalunya].

Repositorio

Institucional

UOC.

<https://openaccess.uoc.edu/bitstream/10609/120648/6/milumbrerasTFM0620memoria.pdf>

Ward, J. H. (1963). *Hierarchical grouping to optimize an objective function*. Journal of the American Statistical Association, 58(301), (pp .236–244).

<https://iv.cns.iu.edu/sw/data/ward.pdf>

## VII. ANEXOS

### Anexo A: Información solicitada sobre Boletines Estadísticos de Acuerdos Marco

	<b>PERÚ</b>	Ministerio de Economía y Finanzas	Central de Compras Públicas - PERÚ COMPRAS	Dirección de Análisis De Mercado
---	-------------	-----------------------------------	--	----------------------------------

*"Decenio de la Igualdad de Oportunidades para Mujeres y Hombres"*  
*"Año del Bicentenario, de la consolidación de nuestra Independencia, y de la conmemoración de las heroicas batallas de Junín y Ayacucho"*

San Isidro, 07 de Junio del 2024

**MEMORANDO N° 000129-2024-PERÚ COMPRAS-DAMER**

Firmado digitalmente por SEMORILE CHAU Manuel Felipe FAU 20605027010 acb  
Cargo: Director De La Dirección De Análisis De Mercado  
Motivo: Soy el autor del documento  
Fecha: 07.06.2024 21:01:29 -05:00

Para : **PATRICIA HERMINIA ESPEJO URIOSTE**  
FUNCIONARIO RESPONSABLE DE ENTREGAR INFORMACIÓN PÚBLICA  
ACCESO A LA INFORMACIÓN PÚBLICA

De : **MANUEL FELIPE SEMORILE CHAU**  
DIRECTOR DE LA DIRECCIÓN DE ANÁLISIS DE MERCADO  
DIRECCIÓN DE ANÁLISIS DE MERCADO

Asunto : Acceso a la Información Pública.

Referencia : a) Memorando N° 000109-2024-PERÚ COMPRAS-FREI(3Jun2024)  
b) Correo: Transparencia – PERÚ COMPRAS(31Mayo2024)

Es grato dirigirme a usted para saludarla y en atención al documento y correo de referencia b), en el que se tramita el pedido de información de la Srta. Gianina Rossmery Timoteo Jacinto, a través del cual solicita los archivos que sirvieron como base por la Dirección de Análisis de Mercado para la elaboración del Boletín Estadístico de Acuerdos Marco para los meses de (enero a diciembre) para los años 2022 y 2023; asimismo, del Boletín N°4 del año 2024.

Al respecto, se adjunta la información solicitada en el siguiente enlace: [Gianina Timoteo-Abril2024](#)

Boletín Estadístico de Acuerdos Marco N°12-2022, N°12-2023 y N°04-2024 (enero-abril)

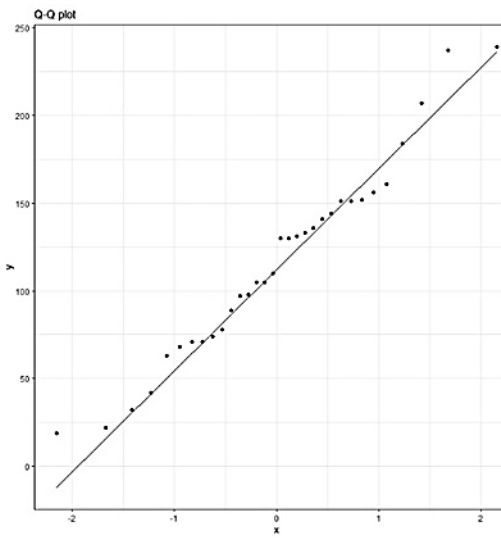
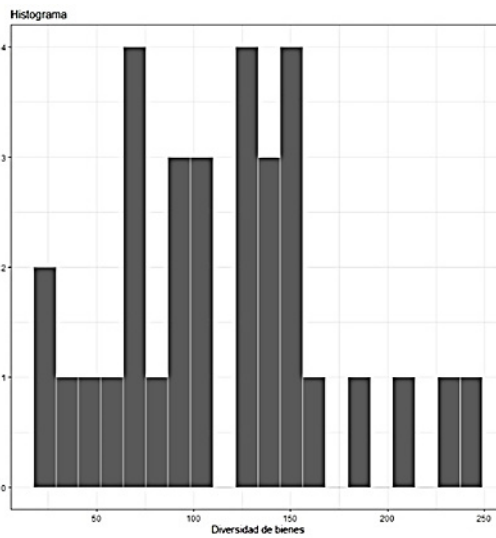
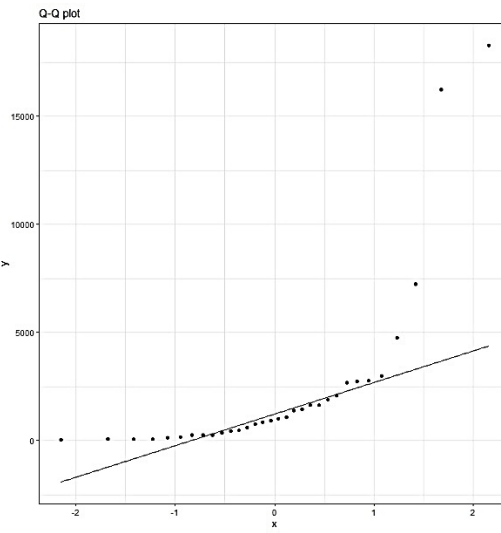
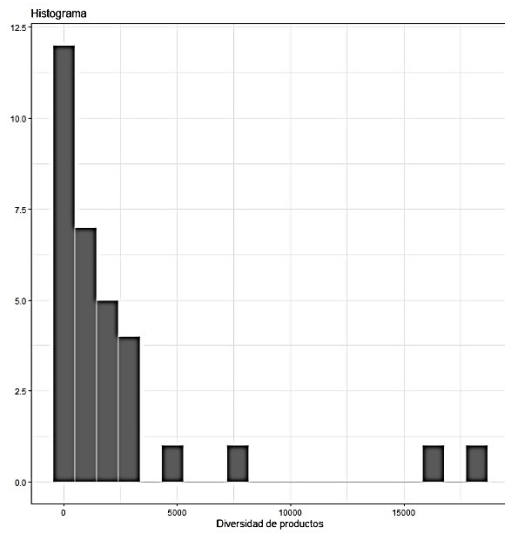
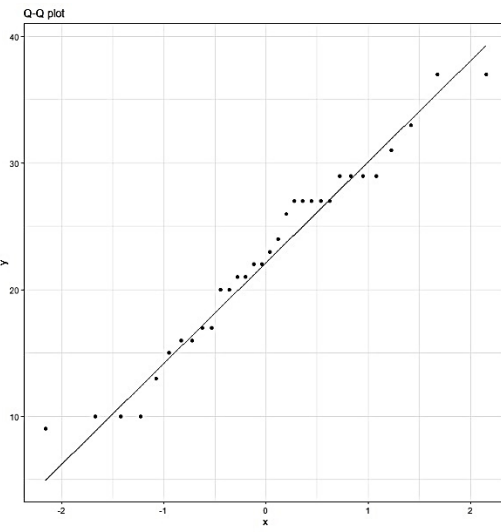
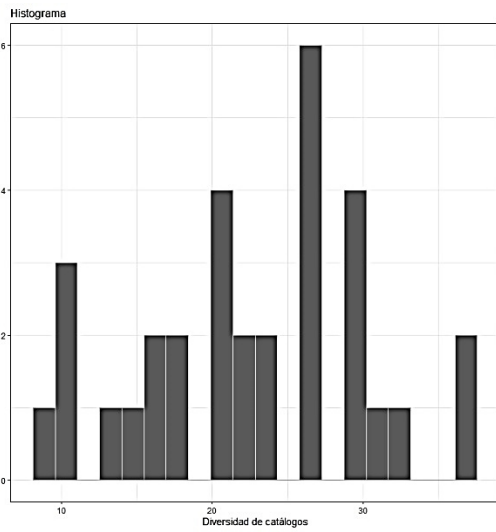
- Reporte de datos Boletín CEAM (enero-diciembre 2022).xlsx.
- Reporte de datos Boletín CEAM (ene-dic) 2023.xlsx.
- Reporte de datos Boletín CEAM (abril) 2024.

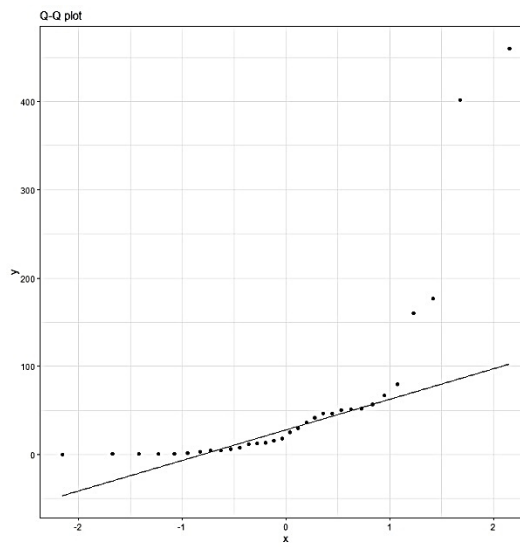
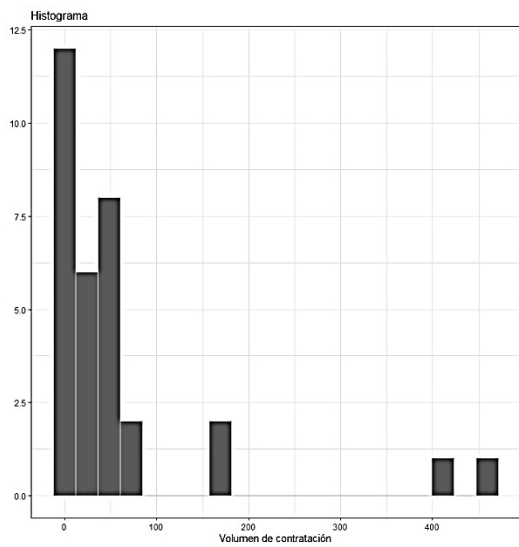
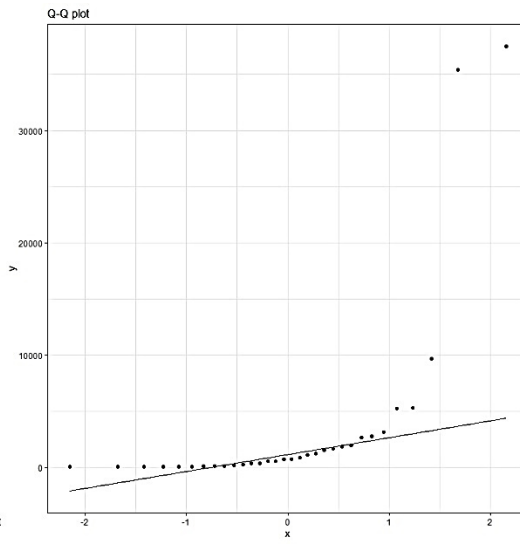
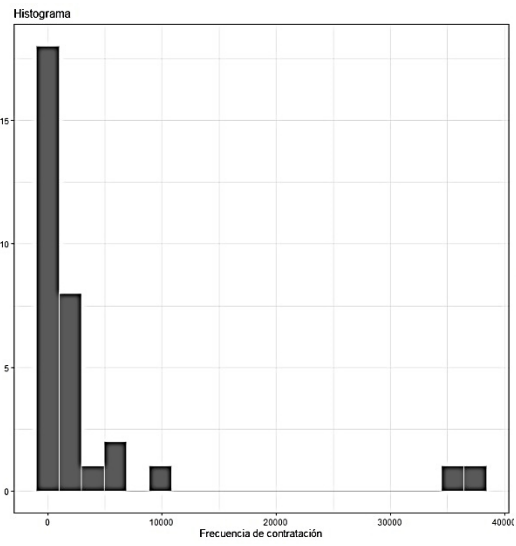
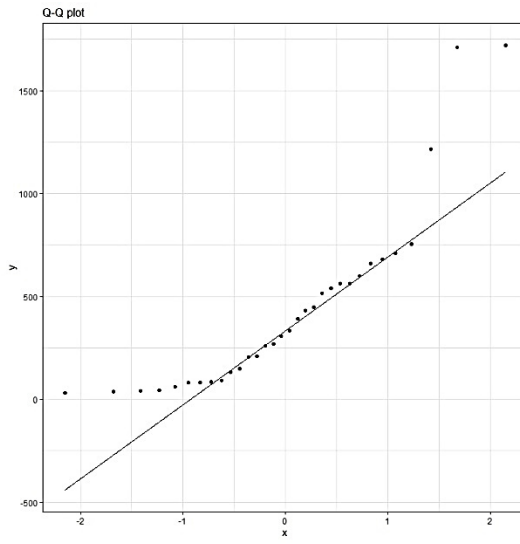
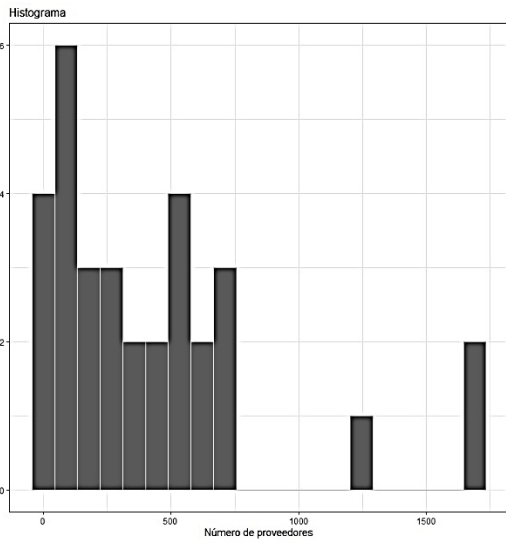
Atentamente,

Documento firmado digitalmente  
**MANUEL FELIPE SEMORILE CHAU**  
DIRECTOR DE LA DIRECCIÓN DE ANÁLISIS DE MERCADO  
Central de Compras Públicas PERÚ COMPRAS

MSC/lvd

## Anexo B: Distribución de las variables mediante histogramas y gráficos Q-Q

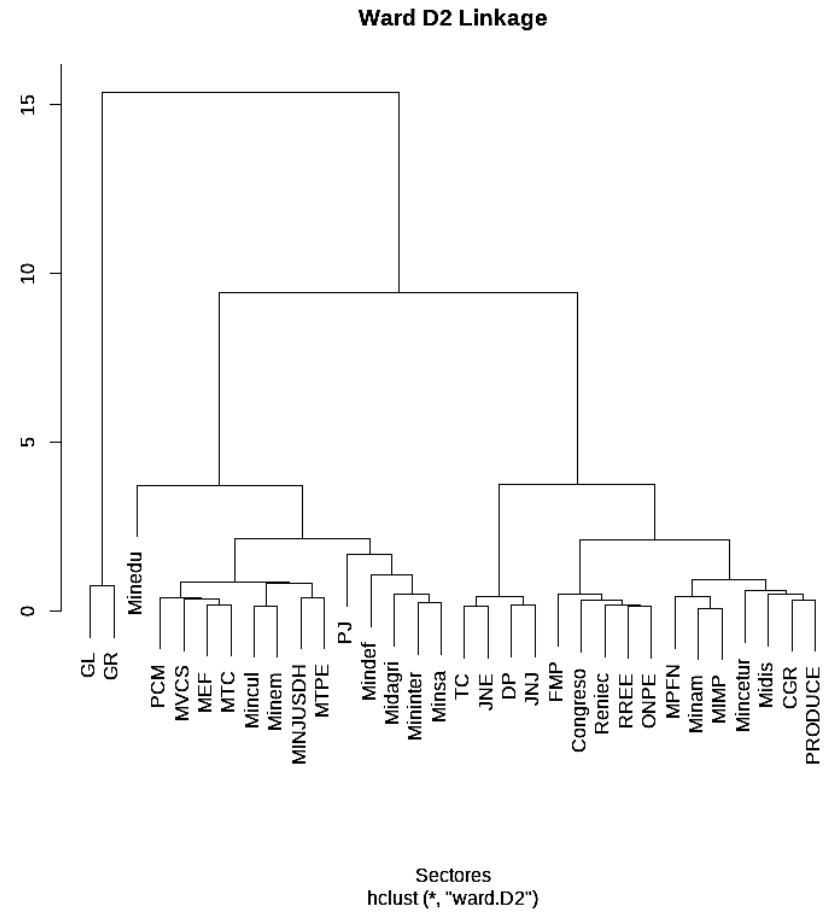
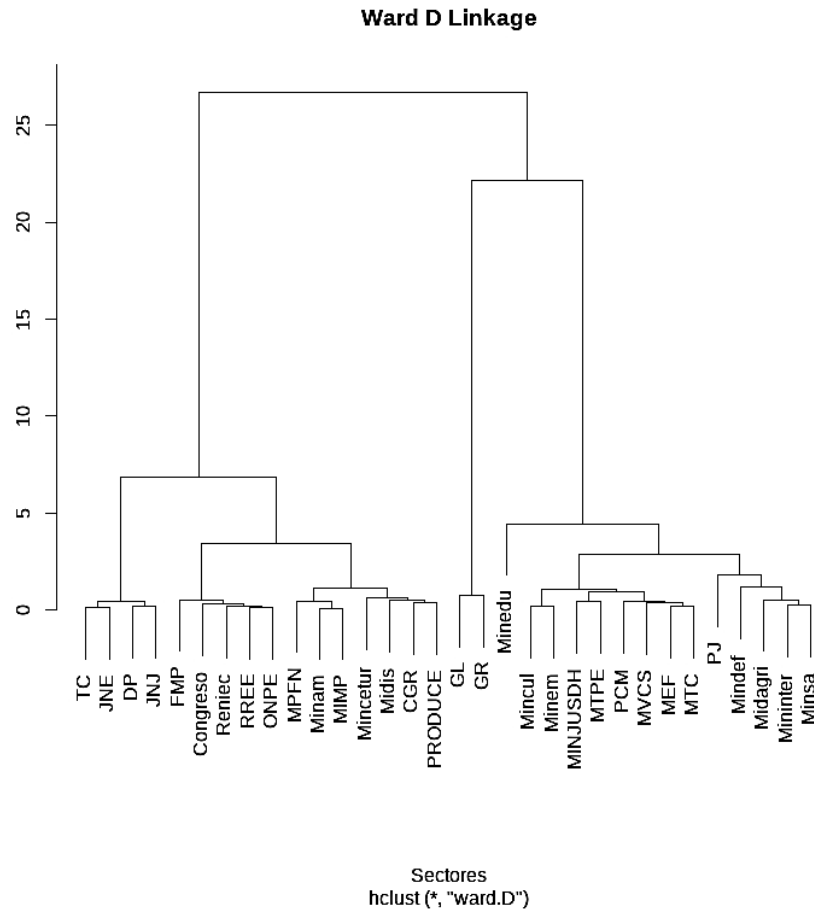




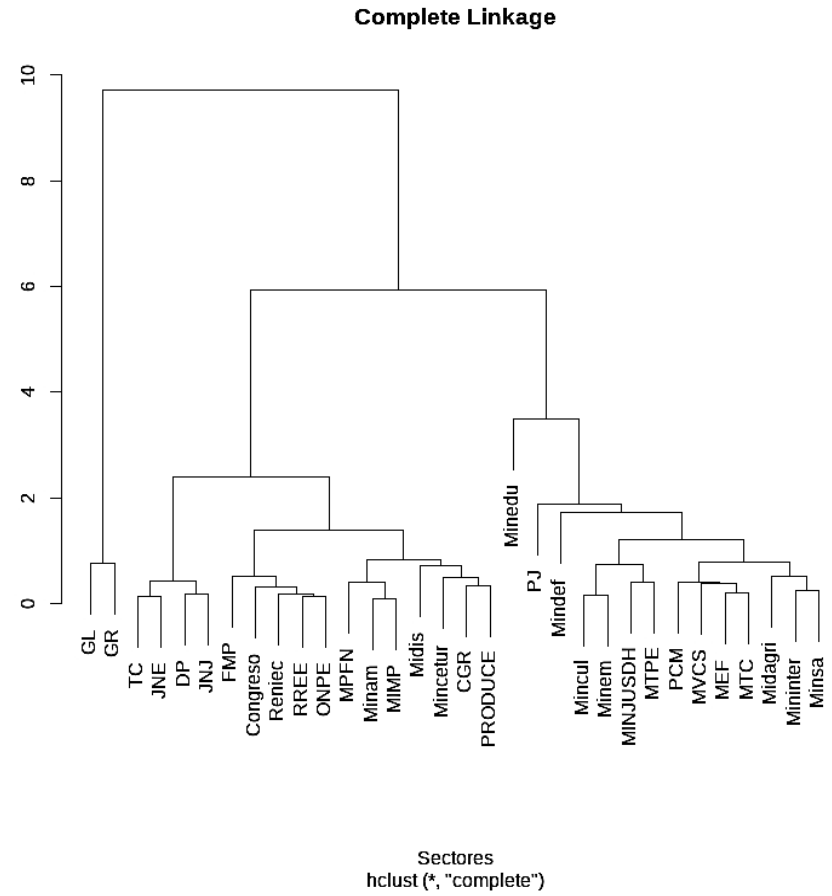
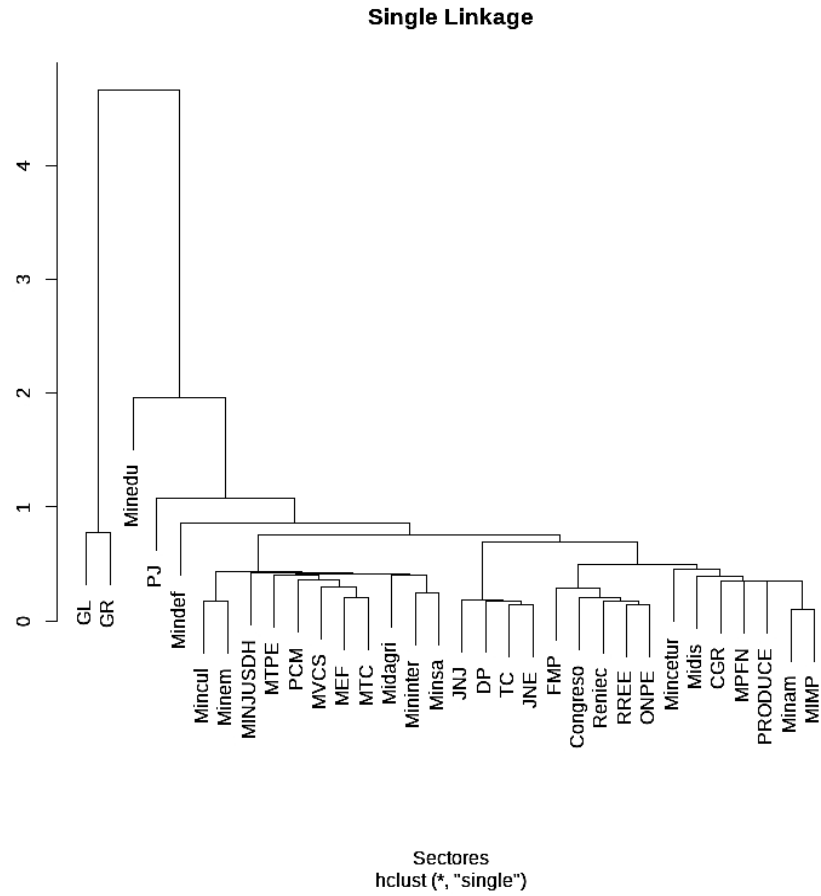




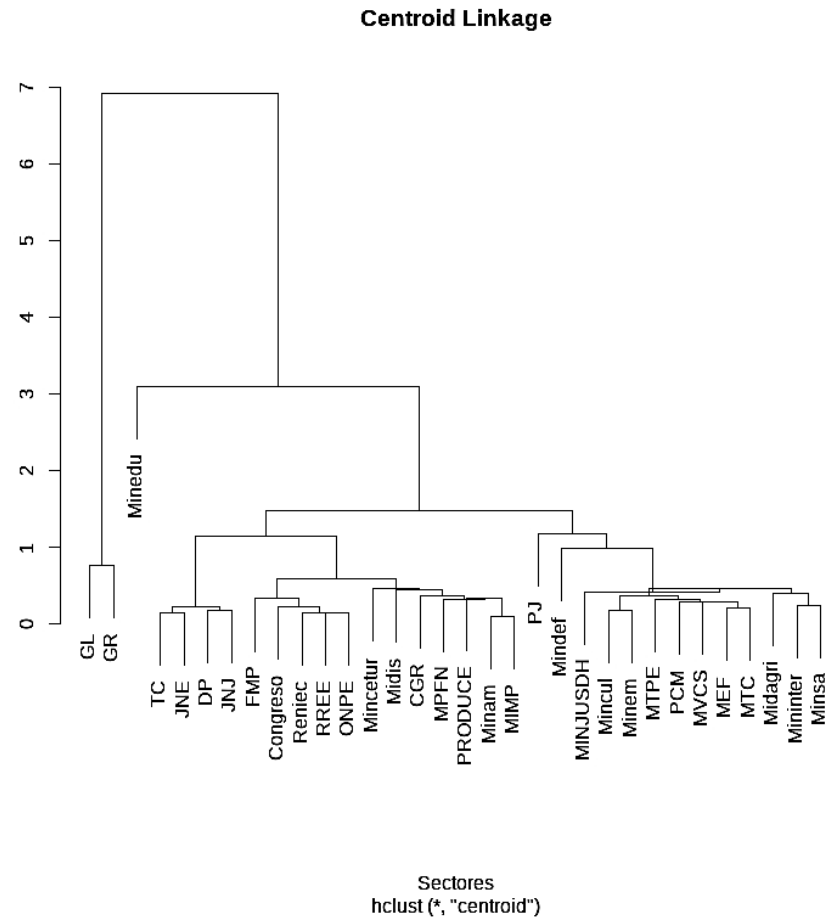
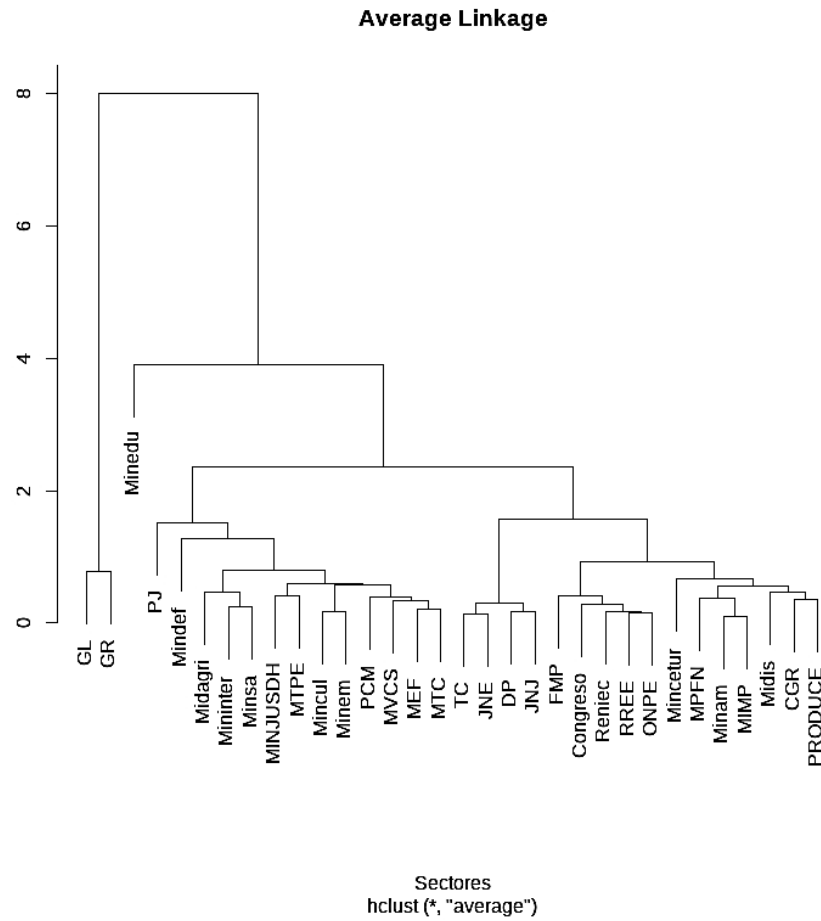
## Anexo D: Clustering Jerárquico de Sectores Públicos: Comparación entre Ward D y Ward D2



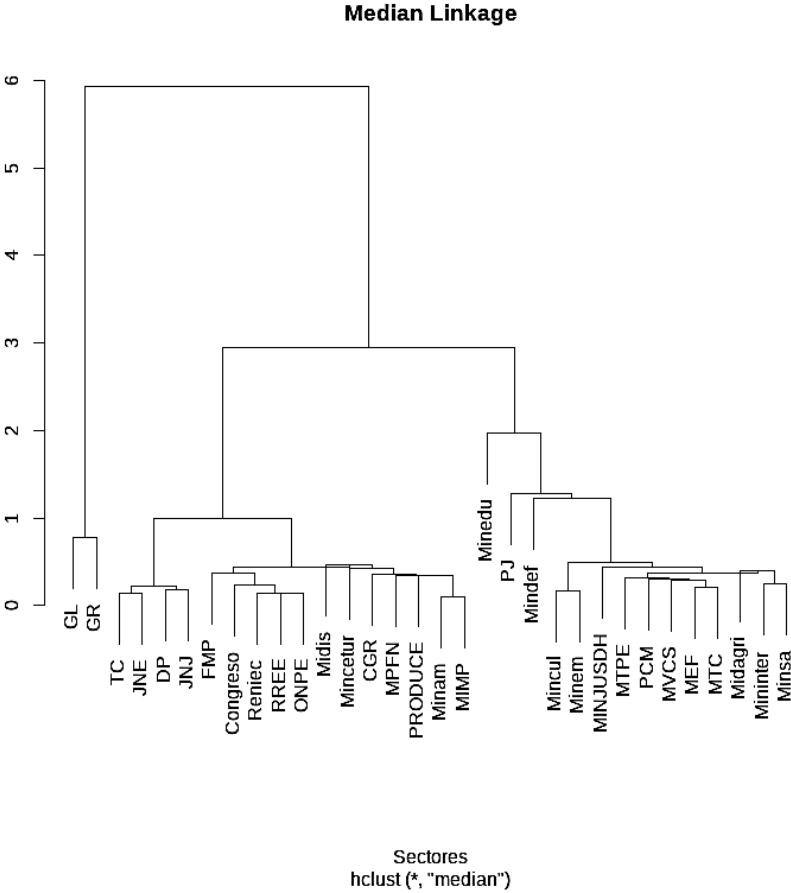
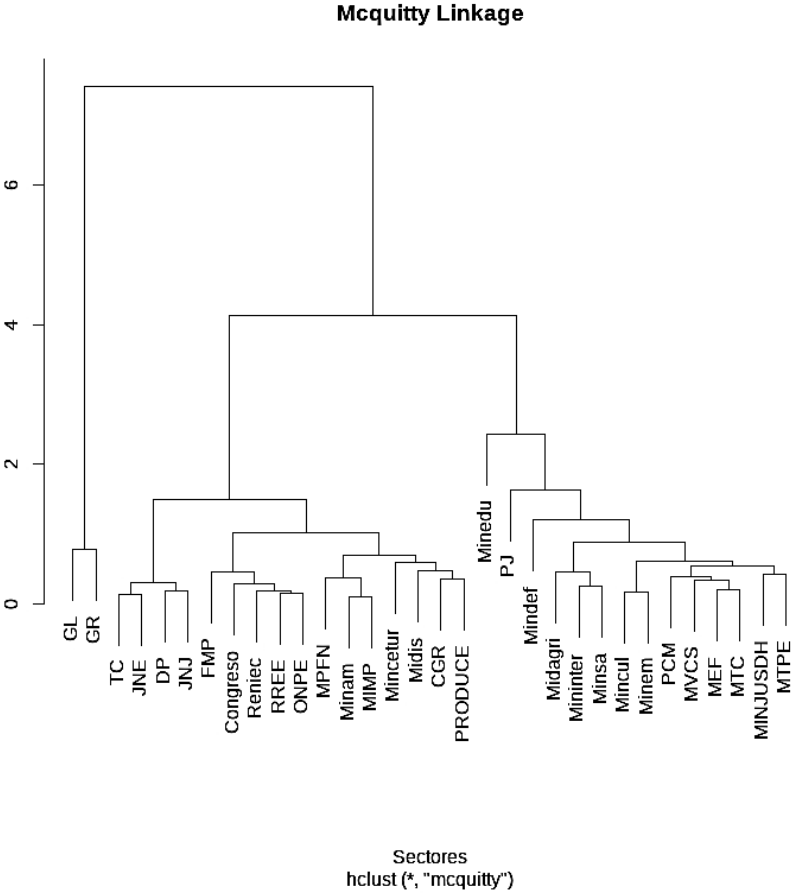
## Anexo E: Clustering Jerárquico de Sectores Públicos: Comparación entre Enlace Simple y Completo



## Anexo F: Clustering Jerárquico de Sectores Públicos: Comparación entre Enlace Promedio vs. Centroidal



### Anexo G: Clustering Jerárquico de Sectores Públicos: Comparación entre Enlace McQuitty y Media



## Anexo H: Clasificación del Sector Público

Sector	Sector Público	Diversidad de catálogos	Diversidad de bienes	Diversidad de productos	Número de proveedores	Frecuencia de contratación	Volumen de contratación (Mill.)	Clúster
PCM	PRESIDENCIA CONSEJO DE MINISTROS	27	151	1,909	516	1,669	56.95	1
Mincul	CULTURA	27	130	1,016	332	767	12.92	1
PJ	PODER JUDICIAL	23	131	2,691	753	5,292	160.08	1
Minam	AMBIENTAL	22	105	766	268	540	18.26	2
MINJUSDH	JUSTICIA	24	133	2,079	564	1,830	30.15	1
Mininter	INTERIOR	27	156	3,002	661	2,797	80.09	1
RREE	RELACIONES EXTERIORES	15	68	259	82	141	3.24	2
MEF	ECONOMÍA Y FINANZAS	29	141	1,455	449	1,268	46.45	1
Minedu	EDUCACIÓN	33	207	7,257	1,215	9,662	176.47	1
Minsa	SALUD	26	161	2,778	601	3,154	67.37	1
MTPE	TRABAJO Y PROMOCIÓN DEL EMPLEO	27	130	1,651	563	2,010	41.77	1
Midagri	AGRICULTURA	29	151	2,751	712	2,637	52.36	1
Minem	ENERGÍA Y MINAS	27	136	1,074	390	897	15.45	1
CGR	CONTRALORÍA GENERAL	20	89	451	131	214	36.37	2
DP	DEFENSORÍA DEL PUEBLO	10	32	78	38	55	0.64	2
JNJ	JUNTA NACIONAL DE JUSTICIA	10	42	85	44	53	0.99	2
MPFN	MINISTERIO PÚBLICO	21	105	913	305	774	51.23	2
TC	TRIBUNAL CONSTITUCIONAL	10	22	50	39	55	0.53	2
Mindef	DEFENSA	31	184	4,756	680	5,271	46.51	1
FMP	FUERO MILITAR POLICIAL	13	74	143	86	103	1.28	2
Congreso	CONGRESO DE LA REPÚBLICA	17	78	179	62	82	4.86	2
JNE	JURADO NACIONAL DE ELECCIONES	9	19	59	30	47	0.35	2
ONPE	OFICINA NACIONAL DE PROCESOS ELECTORALES	16	71	359	91	143	0.51	2
Reniec	REGISTRO NACIONAL DE IDENTIFICACIÓN Y ESTADO CIVIL	16	63	253	80	136	11.71	2
Mincetur	COMERCIO EXTERIOR Y TURISMO	21	71	257	150	237	5.98	2
MTC	TRANSPORTES Y COMUNICACIONES	29	152	1,400	430	1,100	50.5	1
MVCS	VIVIENDA CONSTRUCCIÓN Y SANEAMIENTO	29	144	1,660	540	1,532	25.42	1
PRODUCE	PRODUCCIÓN	20	97	475	209	363	7.72	2
MIMP	MINISTERIO DE LA MUJER Y POBLACIONES VULNERABLES	22	110	867	260	593	13.71	2
Midis	DESARROLLO E INCLUSION SOCIAL	17	98	609	204	388	4.67	2
GL	GOBIERNO LOCAL	37	239	18,277	1,720	35,407	401.57	3
GR	GOBIERNOS REGIONALES	37	237	16,224	1,711	37,480	459.88	3