



UNIVERSIDAD TÉCNICA ESTATAL DE QUEVEDO
FACULTAD DE POSGRADO
MAESTRÍA EN CIENCIA DE DATOS

Proyecto de investigación previa la
obtención del Grado Académico
de Magíster en Ciencia de Datos.

TEMA

“Análisis del comportamiento de los residuos sólidos no peligrosos en los
cantones de la provincia de Los Ríos: Un enfoque de Ciencia de Datos”

AUTORA

Ing. María Mercedes Alarcón Bermúdez

DIRECTOR

Ing. Orlando Erazo Moreta, PhD.

QUEVEDO – ECUADOR

2023 – 2024

CERTIFICACIÓN DE CULMINACIÓN DEL PROYECTO DE INVESTIGACIÓN



El suscrito, Ing. Orlando Erazo Moreta, PhD., docente de la Universidad Técnica Estatal de Quevedo, certifica que la Ing. María Mercedes Alarcón Bermúdez, realizó el Proyecto de Investigación de grado titulado **“Análisis del comportamiento de los residuos sólidos no peligrosos en los cantones de la provincia de Los Ríos: Un enfoque de Ciencia de Datos”**, previo a la obtención del título de Magister en Ciencia de Datos, bajo mi dirección, habiendo cumplido con las disposiciones reglamentarias establecidas para el efecto.

Quevedo, 04 de agosto de 2024



Firmado electrónicamente por:
ORLANDO RAMIRO
ERAZO MORETA

Ing. Orlando Erazo Moreta, PhD.

DIRECTOR DEL PROYECTO DE INVESTIGACIÓN

DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS



Yo, María Mercedes Alarcón Bermúdez, declaro que la investigación aquí descrita es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

La Universidad Técnica Estatal de Quevedo, puede hacer uso de los derechos correspondientes a este documento, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normativa institucional vigente.

Ing. María Mercedes Alarcón Bermúdez

C.C. #0941620106

AGRADECIMIENTO

Agradezco a Dios y a todas las personas que han contribuido al desarrollo de este trabajo. Expreso mi más sincero agradecimiento a mi tutor, Orlando Erazo, por su orientación, paciencia y apoyo durante todo el proceso. Su experiencia y conocimiento han sido esenciales para la realización de este trabajo.

A la unidad de posgrado y los docentes de la Maestría, cuyas enseñanzas y dedicación fueron fuente de inspiración para mi desarrollo académico, compartiendo sus conocimientos con nosotros.

A mi esposo Jorge Luis Casanova, a su familia y a mi familia, gracias por su amor, comprensión y por ser ese pilar de apoyo en cada momento. Su apoyo incondicional ha sido mi fuente constante de fortaleza y ánimo.

Finalmente, mi gratitud se extiende a todos los que han hecho posible la culminación y elaboración de este proyecto de investigación.

Mercedes Alarcón Bermúdez.

DEDICATORIA

Dedico este trabajo a mi hija Elizabeth, que es mi principal fuente de inspiración, esfuerzo y amor. A mi esposo Jorge, por brindarme ese apoyo incondicional en cada momento, brindarme su amor y motivarme a seguir adelante.

También se la dedico a toda mi familia, quienes de una u otra forma me han apoyado en esta etapa de mi vida, brindándome su apoyo moral en cada paso.

Con un corazón lleno de gratitud, agradezco a Dios por tenerlos en mi vida. ¡Gracias por ser esa fuente de inspiración y amor!

Mercedes Alarcón Bermúdez

RESUMEN

La gestión eficiente de los residuos sólidos es un desafío global que afecta la calidad de vida de las comunidades y del medio ambiente. El manejo inadecuado de los residuos sólidos ha generado preocupaciones debido al constante crecimiento de la población y la provincia de Los Ríos no se encuentra exenta de esta problemática. La Ciencia de Datos se ha hecho presente en diferentes áreas de trabajo, incluyendo la ambiental. Por ello, este proyecto tiene como objetivo principal analizar el comportamiento de los residuos, aplicando técnicas de Ciencia de Datos. Para identificar las principales variables influyentes en la generación y manejo de residuos, se realizaron análisis de Chi-Cuadrado y ANOVA, asegurando la independencia de las variables seleccionadas. Posteriormente, para agrupar los cantones según características similares, se aplicaron técnicas de *clustering* como K-means, DBSCAN y *clustering* jerárquico, donde se utilizó el Análisis de Componentes Principales (PCA) para reducir la dimensionalidad de los datos. En los resultados se identificaron 16 variables influyentes relacionadas con la generación y forma de eliminación de los residuos sólidos no peligrosos en los hogares. Al aplicar las técnicas de *clustering*, se identificaron cuatro grupos con características similares en eliminación y manejo de residuos. Las estrategias identificadas en cada cluster revelaron que ciertos cantones participan en voluntariado ambiental y que, en su mayoría, los residuos son depositados con los demás residuos del hogar. Estos resultados demuestran cómo la Ciencia de Datos puede contribuir en el beneficio ambiental.

Palabras clave: Gestión sostenible, residuos no peligrosos, segmentación, minería de datos, aprendizaje automático.

ABSTRACT

Efficient solid waste management is a global challenge that affects the quality of life of communities and the environment. The inadequate handling of solid waste has raised concerns due to the constant population growth, and the province of Los Ríos is not the exception to this issue. Data Science has become crucial in various fields, including environmental work. Therefore, this project aims to analyze waste behavior by applying Data Science techniques. To identify the main variables influencing waste generation and management, Chi-Square and ANOVA analyses were conducted, ensuring the independence of the selected variables. Subsequently, to group the cantons according to similar characteristics, clustering techniques such as K-means, DBSCAN, and hierarchical clustering were applied, using Principal Component Analysis (PCA) to reduce data dimensionality. The results identified 16 influential variables related to the generation and disposal methods of non-hazardous solid waste in households. By applying clustering techniques, four groups with similar characteristics in waste disposal and management were identified. The strategies identified in each cluster revealed that certain cantons actively participate in environmental volunteering and that, for the most part, waste is disposed of with other household waste. These results demonstrate how Data Science can significantly contribute to environmental benefits.

Keywords: Sustainable management, non-hazardous waste, segmentation, data mining, machine learning.

ÍNDICE

CERTIFICACIÓN DE CULMINACIÓN DEL PROYECTO DE INVESTIGACIÓN	2
DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS	3
AGRADECIMIENTO	4
DEDICATORIA	5
RESUMEN	6
ABSTRACT	7
INTRODUCCIÓN	13
CAPÍTULO I	15
MARCO CONTEXTUAL DE LA INVESTIGACIÓN	15
1.1. SITUACIÓN ACTUAL DE LA PROBLEMÁTICA	16
1.2. PROBLEMA DE INVESTIGACIÓN	17
1.2.1. Problemas derivados	17
1.3. DELIMITACIÓN DEL PROBLEMA	17
1.4. OBJETIVOS	18
1.4.1. Objetivo general	18
1.4.2. Objetivos específicos	18
1.5. JUSTIFICACIÓN	19
CAPÍTULO II	21
MARCO TEÓRICO	21
2.1. FUNDAMENTACIÓN CONCEPTUAL	22

2.1.1. Residuo	22
2.1.2. Residuos sólidos	22
2.1.3. Estimación de generación de residuos	22
2.1.4. Comportamiento de los residuos sólidos	23
2.1.5. Clasificación de residuos sólidos no peligrosos	23
2.1.6. Técnicas de <i>clustering</i>	25
2.2. FUNDAMENTACIÓN TEÓRICA.....	26
2.3. FUNDAMENTACIÓN LEGAL	29
CAPÍTULO III	31
METODOLOGÍA DE LA INVESTIGACIÓN	31
3.1. TIPO DE INVESTIGACIÓN	32
3.1.1. Investigación correlacional	32
3.2. MÉTODOS UTILIZADOS EN LA INVESTIGACIÓN	32
3.2.1. Método analítico	32
3.2.2. Procedimiento lógico.....	32
3.3. CONSTRUCCIÓN METODOLÓGICA DEL OBJETO DE INVESTIGACIÓN	34
3.3.1. Técnicas de investigación	34
3.3.2. Instrumentos de investigación	34
3.4. ELABORACIÓN DEL MARCO TEÓRICO	35
3.5 RECOLECCIÓN DE LA INFORMACIÓN	35
3.6 PROCESAMIENTO Y ANÁLISIS.....	36

3.6.1. Identificación de variables influyentes	37
3.6.2. Agrupación por características similares	40
3.6.3. Comportamientos identificados	43
CAPÍTULO IV	44
RESULTADOS Y DISCUSIÓN	44
4.1. PRINCIPALES VARIABLES QUE INFLUYEN EN LA GENERACIÓN DE RESIDUOS SÓLIDOS NO PELIGROSOS EN LOS CANTONES DE LA PROVINCIA DE LOS RÍOS	45
4.1.1. Preparación y limpieza de los datos	45
4.1.2. Análisis de Chi-Cuadrado para variables categóricas	52
4.1.3. Análisis de ANOVA para las variables numéricas y categóricas ($p > 0,05$).....	54
4.1.4. Discusión	56
4.2. TÉCNICAS DE CLUSTERING PARA AGRUPAR CANTONES CON CARACTERÍSTICAS SIMILARES EN TÉRMINOS DE GENERACIÓN Y MANEJO DE LOS RESIDUOS SÓLIDOS NO PELIGROSOS.....	58
4.2.1. Aplicación de técnicas de <i>clustering</i>	60
4.2.2. Discusión	68
4.3. ESTRATEGIAS DE GESTIÓN DE RESIDUOS NO PELIGROSOS ENCONTRADAS EN LOS DIFERENTES <i>CLUSTERS</i> DE CANTONES	69
4.3.1. Estrategias o prácticas identificadas con K-means	69
4.3.2. Estrategias o prácticas identificadas por Clustering Jerárquico	72
4.3.3. Discusión	75

CAPÍTULO IV.....	77
CONCLUSIONES Y RECOMENDACIONES	77
5.1. CONCLUSIONES.....	78
5.2. RECOMENDACIONES	79
REFERENCIAS BIBLIOGRÁFICAS.....	80
ANEXOS.....	87

ÍNDICE DE ILUSTRACIONES

Ilustración 1 Procedimiento propuesto para la ejecución del estudio.....	33
Ilustración 2 Procesos para cumplir con los objetivos.....	36
Ilustración 3 Diagrama de Caja de “TotalPila” antes y después de la imputación.	48
Ilustración 4 Diagrama de Caja de la variable "TotalFocos" antes y después de la imputación.....	49
Ilustración 5 Diagrama de Caja de la variable "FocosAhorradores" antes y después de la imputación.....	50
Ilustración 6 Comparación del Coeficiente de Silueta por Método de Imputación.....	51
Ilustración 7 Valores de Chi-Cuadrado para variables categóricas.....	54
Ilustración 8 One-Hot encoding en variables categóricas.....	58
Ilustración 9 Agrupación de valores agregados por cantón de la provincia.....	59
Ilustración 10 Varianza explicada por componentes principales	60
Ilustración 11 Método del codo para determinar número óptimo de clusters en K-means.	61
Ilustración 12 Representación de los clusters formados por K-means	62

Ilustración 13 Método de Rodilla para DBSCAN	64
Ilustración 14 Visualización de clusters DBSCAN.....	65
Ilustración 15 Dendograma del clustering jerárquico	66
Ilustración 16 Visualización de clusters jerárquico.....	66

ÍNDICE DE TABLAS

Tabla 1 Procesos para el análisis de componentes principales	42
Tabla 2 Nuevo nombre de las variables	45
Tabla 3 Variables eliminadas.....	47
Tabla 4 Resultados de la prueba de chi-cuadrado.....	53
Tabla 5 Comparación de variables numéricas y categóricas.....	55
Tabla 6 Variables útiles para el análisis de clusters.....	56
Tabla 7 Clusters formados por K-means.....	61
Tabla 8 Características de agrupación por K-means.....	63
Tabla 9 Cantones asignados a cada clusters por clusters jerárquico	66
Tabla 10 Características de agrupación para clustering jerárquico.....	68
Tabla 11 Estrategias de gestión de residuos identificadas por K-means.....	70
Tabla 12 Estrategias identificadas por clustering jerárquico.....	72

INTRODUCCIÓN

En la actualidad, la gestión adecuada de los residuos sólidos es un desafío crucial, no solo a nivel local, sino también a escala mundial. El constante aumento de la población y el crecimiento económico han llevado a un incremento en la generación de residuos, lo que ha ejercido una presión significativa sobre el medio ambiente y la calidad de vida de las personas (Sáez et al., 2014). Este problema, que afecta a numerosas zonas en América Latina y el Caribe, no es ajeno a Ecuador, donde la gestión de residuos sólidos se ha convertido en un desafío medioambiental apremiante (Ministerio del Ambiente, s. f.).

En particular, en la provincia de Los Ríos, Ecuador, el aumento constante de la población y el crecimiento de las áreas urbanas han agravado la problemática de la gestión de residuos sólidos. A pesar de los esfuerzos locales, existe una preocupación importante respecto al manejo inadecuado y la disposición final de estos residuos (Goswami et al., 2024). Según datos recopilados a través del Censo de Población y Vivienda (2022), la eliminación de los residuos en esta provincia se lleva a cabo de varias maneras, incluyendo el uso de carros recolectores, la quema de residuos y la disposición en vertederos improvisados. Estos datos subrayan la necesidad de abordar este desafío de manera más efectiva (Dirección Provincial del Ambiente de Los Ríos, 2019).

Dado este escenario, este proyecto se centró en emplear técnicas de Ciencia de Datos para explorar el comportamiento de los residuos sólidos no peligrosos en la provincia de Los Ríos. El objetivo principal es analizar la generación y manejo de estos residuos, aplicando técnicas de Ciencia de Datos. Se detectaron patrones que inciden en la gestión de residuos de los cantones de la provincia de Los Ríos a nivel hogar, proporcionando así descubrimientos que podrían facilitar la toma de decisiones más informada y efectiva. Este enfoque, centrado en la Ciencia de Datos, promete aportar contribuciones

significativas para el entendimiento y mejora de las estrategias de la gestión de residuos sólidos en la provincia de Los Ríos, representando un aporte a las prácticas sostenibles en esta área.

El proyecto tiene objetivos específicos claramente definidos. Primero, se identificaron las principales variables que influyen en la generación de residuos sólidos no peligrosos en los distintos cantones de la provincia. Luego, se aplicaron técnicas de *clustering* para agrupar cantones con características similares en términos de generación y manejo de residuos. Finalmente, se evaluaron las estrategias de gestión de residuos no peligrosos en los diferentes *clusters*. Así, este proyecto aspira a identificar soluciones prácticas y sostenibles basadas en datos para el manejo de residuos, contribuyendo significativamente al bienestar de la comunidad y a la protección del medio ambiente en la provincia de Los Ríos.

El presente trabajo se encuentra estructurado de la siguiente manera. El **capítulo I, “Marco contextual de la investigación”**, especifica aspectos del proyecto como la contextualización de la problemática, el problema general y problemas derivados de la investigación, objetivos y justificación. El **capítulo II, Marco teórico de la investigación**, presenta los conceptos, trabajos relacionados y el marco legal del trabajo. El **capítulo III, “Metodología de la investigación”**, define la metodología aplicada para el cumplimiento de cada uno de los objetivos. En el **capítulo IV, “Resultados y discusión”**, se detallan los hallazgos del análisis realizado, indicando el cumplimiento de cada uno de los objetivos propuestos. Finalmente, el **capítulo V, “Conclusiones y recomendaciones”**, contiene las conclusiones generales de los hallazgos del trabajo y las propuestas de la autora.

CAPÍTULO I

MARCO CONTEXTUAL DE LA INVESTIGACIÓN

1.1. SITUACIÓN ACTUAL DE LA PROBLEMÁTICA

La recolección de residuos sólidos en un área urbana es un desafío y una tarea complicada, debido a que estos tienen múltiples formas de generarse y cada actividad humana es considerada un punto de generación, ya sean propiedades privadas, espacios públicos o incluso áreas deshabilitadas. Los residuos producidos por las personas siempre han existido. Sin embargo, su aparición como un problema ambiental es reciente. Los residuos sólidos se originan debido a la expansión humana y sus ocupaciones diarias (Alcocer, Cevallos, Knudsen - 2019).

El aumento constante de residuos sólidos representa una problemática ambiental en constante crecimiento. Cada día, las personas generan una gran cantidad de residuos debido a las actividades diarias, que incluyen el trabajo, los estudios o las actividades en el hogar. Esto se ha convertido en un problema ambiental en constante aumento que se agrava a medida que crece la sociedad. Por ello, la gestión adecuada de residuos sólidos se torna esencial para mitigar los impactos negativos asociados a su acumulación y disposición inadecuada.

Ante esta situación, la realización del presente proyecto representa un desafío donde se deben aplicar técnicas avanzadas de Ciencia de Datos. Aunque con el uso y aplicación de estas técnicas se pretende revelar factores cruciales en la generación y manejo de residuos, su aplicación exitosa enfrenta obstáculos significativos, siendo uno de los más importantes la necesidad de datos confiables y precisos que presenten la situación real en cada cantón. Intentar segmentar los cantones con base en sus características comunes en la generación y manejo de residuos no solo representa un reto técnico desde el punto de análisis de datos, sino que también exige una necesidad de comprensión de las estrategias de gestión. Con este enfoque se pretende sintetizar la técnica analítica con conocimientos

prácticos existentes, y producir un análisis pertinente que pueda ser transformado en estrategias efectivas para la gestión de residuos.

1.2. PROBLEMA DE INVESTIGACIÓN

¿Cómo se pueden entender los patrones y factores que influyen en la generación y manejo de residuos sólidos no peligrosos en los cantones de la Provincia de Los Ríos?

1.2.1. Problemas derivados

¿Cuáles son las principales variables que influyen en la generación de residuos sólidos no peligrosos en los cantones de la Provincia de Los Ríos y cómo se relacionan?

¿Cómo se pueden agrupar los cantones con características similares en generación y manejo de residuos sólidos no peligrosos, con base en los datos y estadísticas, para facilitar la identificación de estrategias comunes?

¿Qué estrategias de manejo de residuos sólidos no peligrosos son actualmente implementadas en los diferentes cantones de la Provincia de Los Ríos, según datos disponibles?

1.3. DELIMITACIÓN DEL PROBLEMA

CAMPO: Tecnología de la información y la comunicación

ÁREA: Tecnología de la información y la comunicación

LÍNEA: Ciencia de Datos.

LUGAR: Provincia de Los Ríos

TIEMPO: octubre 2023 – junio 2024.

1.4. OBJETIVOS

1.4.1. Objetivo general

Analizar el comportamiento de los residuos sólidos no peligrosos en los cantones de la provincia de Los Ríos utilizando técnicas de Ciencia de Datos.

1.4.2. Objetivos específicos

1. Identificar las principales variables que influyen en la generación de residuos sólidos no peligrosos en los cantones de la provincia de Los Ríos, para aplicar técnicas de Ciencia de Datos y análisis del comportamiento de los residuos.
2. Aplicar técnicas de *clustering* para agrupar cantones con características similares en términos de generación y manejo de los residuos sólidos no peligrosos, facilitando la identificación de patrones y estrategias comunes.
3. Examinar las estrategias de gestión de residuos no peligrosos encontradas en los diferentes *clusters* de cantones, identificando prácticas eficientes y oportunidades de mejora para una gestión sostenible de residuos.

1.5. JUSTIFICACIÓN

La rápida urbanización, el aumento de la población y el progreso económico han generado una mayor cantidad de residuos. Según estudios estadísticos realizados por Kaza et al. (2018), en el año 2016, a nivel mundial se generaron 2010 millones de toneladas de residuos sólidos municipales, esperando que aumenten a 3400 millones de toneladas para el 2050. Además, en este mismo reporte se menciona que el 33 % de los residuos sólidos producidos se manejan de manera insegura y se eliminan en vertederos ilegales o no monitoreados. Esto hace que los residuos sólidos sean un tema de especial interés.

La disposición de residuos sólidos no peligrosos es un desafío crítico en el contexto global debido a sus impactos ambientales, económicos y sociales. La provincia de Los Ríos no se encuentra libre de este problema, debido a que la población aumenta constantemente y en consecuencia también aumenta la producción de residuos sólidos. La acumulación de residuos sólidos puede provocar problemas de salud como enfermedades transmitidas por vectores y contaminación del aire, lo que destaca la importancia de abordar eficazmente este problema (Rondón et al., 2016).

El propósito de este trabajo es realizar un análisis del comportamiento de los residuos sólidos no peligrosos en la provincia de Los Ríos, aplicando técnicas avanzadas de ciencias de datos, como el *clustering*. Se persigue emplear técnicas analíticas que no solo ayuden a comprender mejor las tendencias actuales en la generación de residuos, sino que, al proporcionar un análisis profundo, este proyecto podría contribuir significativamente en la toma de decisiones basadas en datos, permitiendo a las personas responsables implementar estrategias más efectivas y adaptadas a las necesidades específicas de los distintos cantones de la provincia.

Este trabajo, centrado en el análisis de residuos sólidos no peligrosos en la provincia de Los Ríos, se considera en el marco de los Objetivos de Desarrollo Sostenible (ODS). Específicamente, contribuye al ODS 11: Ciudades y Comunidades Sostenibles, al abordar la gestión eficiente de residuos sólidos para mejorar la sostenibilidad urbana (Organización de las Naciones Unidas, 2022). Además, se alinea con el Plan Nacional de Desarrollo de Ecuador, orientado hacia la promoción de prácticas sostenibles y la mejora de la calidad de vida de sus habitantes (Secretaría Nacional de Planificación, 2021). Al aplicar técnicas de Ciencia de Datos, este proyecto no solo busca optimizar la gestión de residuos, sino también fomentar un desarrollo más sostenible y responsable en la provincia, en línea con los compromisos nacionales e internacionales de sostenibilidad.

CAPÍTULO II

MARCO TEÓRICO

2.1. FUNDAMENTACIÓN CONCEPTUAL

2.1.1. Residuo

En el manual de aprovechamiento elaborado por el Ministerio del Ambiente y Agua del Ecuador (2020), se define a los residuos como todo objeto, sustancia, material o elemento sólido que no tiene características peligrosas resultantes de diferentes actividades humanas. Además, el documento del mismo ministerio menciona que un residuo puede ser transformado o reutilizado para crear un nuevo producto que podría tener un valor económico agregado.

2.1.2. Residuos sólidos

Los residuos sólidos son definidos generalmente como todo material sólido o semi sólido que es descartado o que ya no es considerado útil. Según la Agencia de Protección Ambiental de Estados Unidos (EPA) (2020), estos residuos incluyen varios tipos de materiales que son desechados como residuos domésticos, comerciales, construcción e industriales no peligrosos. EPA establece que una gestión correcta de los residuos sólidos podría ser crucial para la salud humana y el medio ambiente, promoviendo prácticas como la reducción de la generación de residuos, el reciclaje e incluso la recuperación de energía.

2.1.3. Estimación de generación de residuos

Hoornweg & Bhada-Tata (2012) señalan que la estimación de generación de residuos implica el proceso de calcular la cantidad de residuos que se espera generar en una determinada área y período de tiempo. Esto se logra mediante el análisis de datos históricos, considerando factores demográficos, tendencias de consumo y otros parámetros relevantes.

2.1.4. Comportamiento de los residuos sólidos

Se entiende por comportamiento al conjunto de patrones y dinámicas asociadas a la generación y manejo de los residuos sólidos dentro de una zona o área específica. En este proyecto se entiende por “comportamiento” a toda actividad que se realiza con los residuos sólidos no peligrosos, desde la generación hasta la forma de reutilización, manejo o eliminación. Un análisis de los comportamientos de los residuos sólidos podría incluir aspectos como la generación, que es la cantidad y tipo de residuo producidos por hogares, industrias u otras actividades económicas; la recolección y transporte, que hace referencia a los métodos en que los residuos son recolectados; tratamiento, reciclaje y, en algunos casos, la disposición final.

2.1.5. Clasificación de residuos sólidos no peligrosos

La clasificación de los residuos sólidos no peligrosos es esencial para su manejo adecuado y para minimizar su impacto ambiental. Estos residuos son aquellos que no representan un riesgo significativo para la salud humana o el medio ambiente. Aquí se presenta una clasificación detallada de los residuos sólidos no peligrosos:

2.1.5.1. Residuos orgánicos

Los residuos orgánicos son materiales de origen biológico que pueden descomponerse naturalmente. Estos incluyen restos de alimentos, poda de jardines, residuos agrícolas y otros desechos de origen vegetal o animal. Los residuos orgánicos son ideales para procesos de compostaje, lo que permite convertirlos en abono natural, enriqueciendo el suelo y reduciendo la necesidad de fertilizantes químicos (Anderson, 2023).

2.1.5.2. Residuos inorgánicos

Los residuos inorgánicos no tienen origen biológico y no se descomponen fácilmente.

Estos se dividen en varias categorías:

- **Plásticos:** Incluyen bolsas, botellas, envases y otros productos plásticos. El reciclaje de plásticos es crucial para reducir la contaminación y la dependencia de recursos fósiles (Altikat et al., 2022).
- **Metales:** Como latas de aluminio, chatarra de hierro y otros metales. Los metales pueden ser fundidos y reutilizados repetidamente, lo que hace su reciclaje muy eficiente en términos de conservación de recursos y energía (Yang et al., 2021).
- **Vidrio:** Botellas y frascos. El vidrio puede ser reciclado infinitamente sin perder su calidad, lo que lo convierte en un material ideal para el reciclaje (Yang et al., 2021)
- **Papel y cartón:** Periódicos, revistas, cajas y otros productos de papel. El reciclaje de papel ayuda a conservar los bosques y reducir la contaminación del agua y del aire (Kawai & Tasaki, 2016).

2.1.5.3. Residuos comerciales

Estos son generados en establecimientos comerciales, tales como tiendas, restaurantes, hoteles y oficinas. Incluyen una variedad de materiales reciclables como papel, cartón, plásticos y metales. La gestión adecuada de estos residuos puede mejorar la sostenibilidad de las operaciones comerciales y reducir los costos de eliminación de residuos (Altikat et al., 2022) .

2.1.5.4. Residuos industriales no peligrosos

Estos provienen de procesos industriales y manufactureros y no presentan características peligrosas. Incluyen residuos como el metal de desecho, residuos plásticos, madera, y productos de papel y cartón. La gestión eficiente de estos residuos implica su reciclaje y reutilización en procesos industriales, minimizando el impacto ambiental (Li et al., 2017)

2.1.5.5. Residuos de Construcción y Demolición

Generados en actividades de construcción, remodelación y demolición, estos residuos incluyen materiales como concreto, ladrillos, madera y metal. Su reciclaje y reutilización en nuevas construcciones reduce la demanda de materiales vírgenes y disminuye los residuos enviados a los vertederos (Luciani et al., 2015).

Importancia de la Clasificación y Manejo Adecuado

La clasificación adecuada de los residuos sólidos no peligrosos es fundamental para su manejo eficiente. Facilita el reciclaje y la reutilización, reduce la cantidad de residuos enviados a los vertederos y ayuda a conservar los recursos naturales. Además, promueve prácticas sostenibles y contribuye a la economía circular, donde los materiales se mantienen en uso durante más tiempo, reduciendo así la necesidad de nuevos recursos (Kawai & Tasaki, 2016).

2.1.6. Técnicas de *clustering*

Riquelme et al. (2006) se refieren al *clustering* como el proceso mediante el cual se agrupan diferentes registros y observaciones en categorías que tienen similitud entre sí, pero distintas a los registros de otro *cluster*. Además, Riquelme menciona que esta técnica se encuentra presente en diferentes áreas de estudio, dando como ejemplo el de las ciencias ambientales, donde es utilizado para identificar o descubrir patrones con el fin de mejorar su gestión o control.

Aragones y Graells (2019) mencionan que *clustering* es una técnica que pertenece al tipo de aprendizaje no supervisado. Esta técnica se basa en diferentes enfoques y algoritmos. Algunos algoritmos trabajan con supuestos iniciales, como un número predefinido de grupos, mientras que otros determinan el número óptimo de grupos basándose en la estructura de los datos.

Rodríguez et al. (2019) mencionan diferentes métodos dentro del clustering:

Métodos jerárquicos. Son aquellos que incluyen enfoques aglomerativos y divisivos, utilizan la distancia como criterio principal para agrupar o separar puntos. Este método de agrupación se basa en similitudes o distancias de similitud entre las observaciones.

Métodos de partición. Un método muy común de partición es el k-means, donde se forman grupos alrededor de centroides que representan el promedio de los puntos dentro de cada clúster.

Método basado en densidad. Entre los métodos basados en densidad se destaca DBSCAN, donde se agrupan puntos en zonas de alta densidad, sin necesidad de especificar parámetros de antemano.

2.2. FUNDAMENTACIÓN TEÓRICA

El estudio realizado por Gondo (2019) se enfoca en evaluar los riesgos para la salud humana y la vida acuática asociados con prácticas deficientes de gestión de residuos sólidos municipales en Etiopía, empleando el Análisis de Clúster Jerárquico (HCA). Analizando 26 ciudades y pueblos, el estudio revela que las amenazas para la salud están relacionadas con factores económicos, infraestructura de gestión de residuos y políticas de planificación del uso del suelo. Este enfoque metodológico proporciona una comprensión detallada de cómo las características geográficas y socioeconómicas

influyen en los riesgos para la salud, ofreciendo así un marco valioso para abordar la gestión de residuos de manera más efectiva.

El artículo presentado por Izquierdo-Horna et al. (2021) presenta un análisis exploratorio de la generación de residuos sólidos en la comunidad del distrito de Puente Piedra, en Perú. Este estudio utiliza el algoritmo de clasificación no supervisada K-Means y un conjunto de indicadores sociales para segmentar la población y comprender su comportamiento en relación con la generación y acumulación de residuos sólidos municipales. Los resultados del estudio revelan perfiles demográficos predominantes y patrones de comportamiento relacionados con la generación de residuos, proporcionando información para la gestión efectiva de los residuos sólidos en la región. Este enfoque metodológico combina análisis socioeconómico, sociodemográfico y sociocultural con técnicas avanzadas de aprendizaje automático, ofreciendo una visión integral y práctica para abordar los desafíos de la gestión de residuos.

Assef et al. (2022) en cambio realizaron una revisión sistemática sobre el uso de técnicas de análisis de clúster en la gestión de residuos, destacando nueve aplicaciones distintas, con especial énfasis en la recolección de residuos como el área más estudiada. El trabajo reveló que el algoritmo K-means es la técnica más popular en este campo y sugiere la necesidad de incorporar restricciones realistas para hacer las metodologías más aplicables prácticamente.

Izquierdo-Horna et al. (2022) destacan en su artículo la importancia de incluir variables socioculturales, ambientales y económicas en el proceso de análisis de residuos sólidos. Sin embargo, señalan que la disponibilidad de datos relacionados con estas variables es limitada, llegando a la conclusión de que el uso de variables específicas de estudios de

casos a través de datos primarios podría potenciar el desempeño predictivo de los modelos, lo que les proporcionaría una herramienta eficaz para la gestión de los residuos.

Warintarawej y Nillaor (2023) implementan técnicas de clustering mixto para la gestión sostenible de residuos en una comunidad de bajo presupuesto de Tailandia. Utilizaron el coeficiente de Gower para analizar variables socioeconómicas con variables numéricas relacionadas al comportamiento de las 3R: reutilizar, reducir y reciclar. Ellos tuvieron como resultado el desarrollo de guías prácticas para la gestión municipal de residuos sólidos, dejando como evidencia la eficacia de la participación comunitaria y el análisis de datos en la creación de estrategias sostenibles para la gestión de residuos.

Dunkel et al. (2022) analizan cómo los datos socioeconómicos abiertos pueden utilizarse para el análisis de residuos sólidos, empleando tanto modelos de regresión como técnicas de *clustering*. Utilizaron el algoritmo K-means para agrupar países según características similares de residuos, identificando factores socioeconómicos y demográficos que más influyen en la producción de residuos. Este enfoque facilita la comprensión entre países y el desarrollo de estrategias de gestión de residuos más efectivas a nivel nacional.

De acuerdo con los trabajos mencionados, se puede verificar que existen varias investigaciones donde se aplican técnicas de Ciencia de Datos para analizar datos relacionados a los residuos sólidos. Estas investigaciones han demostrado la eficacia de diversas técnicas de *clustering* y análisis de datos para identificar patrones y mejorar las estrategias de gestión de residuos. Sin embargo, es necesario seguir aplicando y desarrollando nuevas técnicas para abordar de manera efectiva posibles problemáticas relacionadas al manejo de los residuos sólidos no peligrosos. En el presente trabajo se aplicaron técnicas de *clustering* para analizar la generación de residuos sólidos no

peligrosos en la provincia de Los Ríos. Este enfoque proporcionará una comprensión de la situación de la provincia en relación con la gestión de residuos, lo que podría servir como base para futuros trabajos o el desarrollo de estrategias más efectivas y sostenibles. La aplicación de estas técnicas permitirá identificar patrones clave y factores influyentes, contribuyendo así a mejorar las prácticas de gestión de residuos y promover un manejo más eficiente y sostenible de los residuos sólidos en la comunidad.

2.3. FUNDAMENTACIÓN LEGAL

SECCIÓN I GESTIÓN INTEGRAL DE RESIDUOS Y/O DESECHOS SÓLIDOS NO PELIGROSOS

Art. 55 De la gestión integral de residuos y/o desechos sólidos no peligrosos.- La gestión integral constituye el conjunto de acciones y disposiciones regulatorias, operativas, económicas, financieras, administrativas, educativas, de planificación, monitoreo y evaluación, que tienen la finalidad de dar a los residuos sólidos no peligrosos el destino más adecuado desde el punto de vista técnico, ambiental y socio-económico, de acuerdo con sus características, volumen, procedencia, costos de tratamiento, posibilidades de recuperación y aprovechamiento, comercialización o finalmente su disposición final.

Está dirigida a la implementación de las fases de manejo de los residuos sólidos que son la minimización de su generación, separación en la fuente, almacenamiento, recolección, transporte, acopio y/o transferencia, tratamiento, aprovechamiento y disposición final.

Una gestión apropiada de residuos contribuye a la disminución de los impactos ambientales asociados a cada una de las etapas de manejo de éstos. Art. 56 Normas técnicas. - La Autoridad Ambiental Nacional establecerá la norma técnica para la gestión integral de residuos y/o desechos sólidos no peligrosos, en todas sus fases.

Art. 57 Responsabilidades de los Gobiernos Autónomos Descentralizados Municipales.- Garantizarán el manejo integral de residuos y/o desechos sólidos generados en el área de su competencia, ya sea por administración o mediante contratos con empresas públicas o privadas; promoviendo la minimización en la generación de residuos y/o desechos sólidos, la separación en la fuente, procedimientos adecuados para barrido y recolección, transporte, almacenamiento temporal de ser el caso, acopio y/o transferencia; fomentar su aprovechamiento, dar adecuado tratamiento y correcta disposición final de los desechos que no pueden ingresar nuevamente a un ciclo de vida productivo; además dar seguimiento para que los residuos peligrosos y/o especiales sean dispuestos, luego de su tratamiento, bajo parámetros que garanticen la sanidad y preservación del ambiente.

Art. 60 Del Generador. - Todo generador de residuos y/o desechos sólidos no peligrosos debe:

Tener la responsabilidad de su manejo hasta el momento en que son entregados al servicio de recolección y depositados en sitios autorizados que determine la autoridad competente.

Tomar medidas con el fin de reducir, minimizar y/o eliminar su generación en la fuente, mediante la optimización de los procesos generadores de residuos.

Realizar separación y clasificación en la fuente conforme lo establecido en las normas específicas.

Art. 61 De las prohibiciones. - No depositar sustancias líquidas, pastosas o viscosas, excretas, ni desechos peligrosos o de manejo especial, en los recipientes destinados para la recolección de residuos sólidos no peligrosos.

CAPÍTULO III

METODOLOGÍA DE LA INVESTIGACIÓN

3.1. TIPO DE INVESTIGACIÓN

3.1.1. Investigación correlacional

Al aplicar técnicas de *clustering* y segmentación en la investigación, es necesario aplicar este tipo de investigación para realizar un análisis más profundo y explorar cómo diferentes combinaciones de variables interactúan e influyen en la generación y manejo de los residuos, descubriendo patrones subyacentes que podrían no ser identificados solo en un análisis descriptivo o aplicado.

3.2. MÉTODOS UTILIZADOS EN LA INVESTIGACIÓN

3.2.1. Método analítico

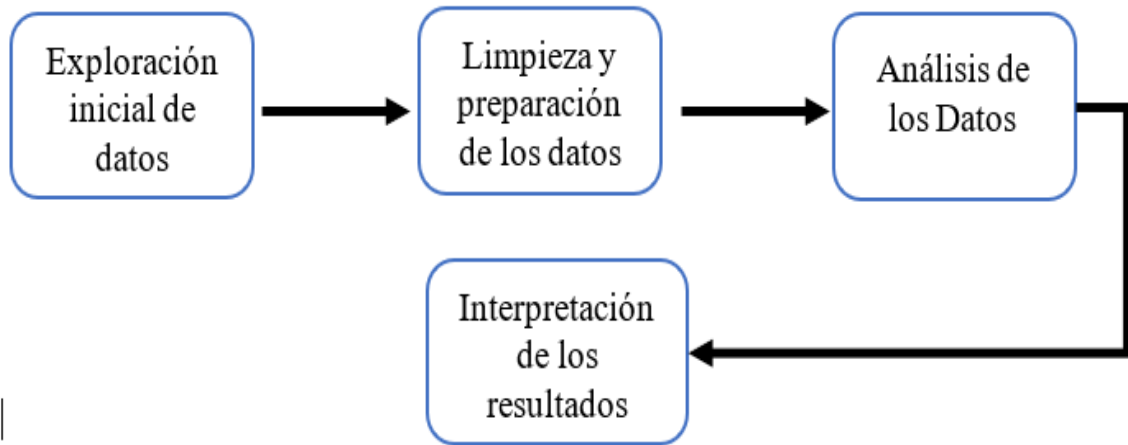
El método analítico fue utilizado en el desarrollo del proyecto para la revisión de la base de datos disponible, que contiene información del manejo de los residuos sólidos en los diferentes cantones de la provincia de interés. Esto permitió realizar la limpieza y validación de los datos, representando un paso importante en el desarrollo de la investigación. Además, facilitó la aplicación de técnicas de Ciencia de Datos para identificar patrones, tendencias y anomalías en los datos.

3.2.2. Procedimiento lógico

En la **Ilustración 1** se observa el procedimiento lógico de la investigación que implica una secuencia ordenada de pasos que incluye la exploración inicial de los datos, limpieza y preparación de estos, análisis de los datos, interpretación y presentación de los resultados obtenidos.

Ilustración 1

Procedimiento propuesto para la ejecución del estudio.



Nota: Elaborado por la autora.

Exploración inicial de datos. En esta etapa se realizó una revisión de la base de datos existente que contiene información de los residuos sólidos en los cantones de la provincia de Los Ríos. Se identificó también la estructura de los datos, las variables disponibles y los metadatos relevantes, permitiendo verificar la integridad de los datos e identificar los problemas como valores faltantes.

Limpieza y preparación de datos. Esta etapa de la investigación permitió corregir errores de entradas en los datos, como inconsistencias o valores incorrectos e imputación de valores faltantes de manera adecuada, aplicando técnicas basadas en modelos, además de identificar y manejar los valores atípicos que podrían afectar el análisis en las otras etapas y conducir el estudio a conclusiones incorrectas.

Análisis de datos. Se realizó un análisis estadístico inicial para resumir los datos, incluyendo medidas de tendencia central y medidas de dispersión. Además, se utilizaron herramientas para la visualización de los resultados obtenidos y exploración de los patrones visuales. También, se aplicaron técnicas de Ciencia de Datos más avanzadas,

como *clustering*, para agrupar los cantones según sus características en la generación y manejo de los residuos, para de esta forma descubrir patrones y relaciones entre los datos.

3.3. CONSTRUCCIÓN METODOLÓGICA DEL OBJETO DE INVESTIGACIÓN

3.3.1. Técnicas de investigación

En este trabajo se utilizaron técnicas de análisis de ciencias de datos para explorar el comportamiento de los residuos sólidos no peligrosos. Como técnica de investigación de Ciencia de Datos, se aplicó *clustering* para crear diferentes grupos con características similares, permitiendo identificar diferentes grupos de prácticas en relación con el manejo de los residuos sólidos.

3.3.2. Instrumentos de investigación

En esta investigación, los instrumentos se centraron en la manipulación y el análisis de datos provenientes de la base de datos de gestión de residuos sólidos en la provincia de Los Ríos. A diferencia de los instrumentos tradicionales como cuestionarios o encuestas, en este caso, se trabajó con datos ya recopilados.

Software de Ciencia de Datos. Para cumplir con los objetivos establecidos se seleccionó *Python* como el entorno de programación debido a su flexibilidad y capacidad para manejar datos complejos. Este lenguaje de programación incluye las bibliotecas como *Pandas* y *Scikit-Learn*, que permiten la manipulación y el análisis de datos. *Pandas*, gracias a su eficiente estructura de datos *Dataframe*, permitió manejar tablas de datos grandes y complejas. Esta biblioteca ofreció funciones eficientes para la limpieza, filtrado, agrupación y transformación de datos, facilitando que el proceso de la preparación de los datos sea menos propenso a errores. Además, *Scikit-Learn* fue crucial

para acceder a los algoritmos de aprendizaje que incluyen los métodos de *K-means*, análisis de componentes principales (PCA), *clustering* jerárquico y DBSCAN.

3.4. ELABORACIÓN DEL MARCO TEÓRICO

Para la elaboración del marco teórico, se identificaron los conceptos clave relacionados con la gestión de residuos sólidos, permitiendo tener una mejor visión de lo que es la gestión de residuos y su generación por cantón. Además, se realizó una revisión de la literatura ya existente en diferentes fuentes académicas relacionadas con la gestión de residuos sólidos, la Ciencia de Datos y la aplicación de técnicas de *clustering*. Esta revisión incluyó la consulta de bases de datos académicas como, *Springer*, revistas especializadas como *PLOS ONE*, revisión de literaturas académicas del motor de búsqueda *Google Scholar*, entre otras. Se emplearon palabras claves específicas para filtrar artículos más relevantes. También se consideraron trabajos comparativos y análisis de diversas técnicas de *clustering*, incluyendo métodos jerárquicos, de partición y K-Means, proporcionando una idea para la aplicación de estas técnicas en el análisis de los datos de los residuos sólidos.

3.5 RECOLECCIÓN DE LA INFORMACIÓN

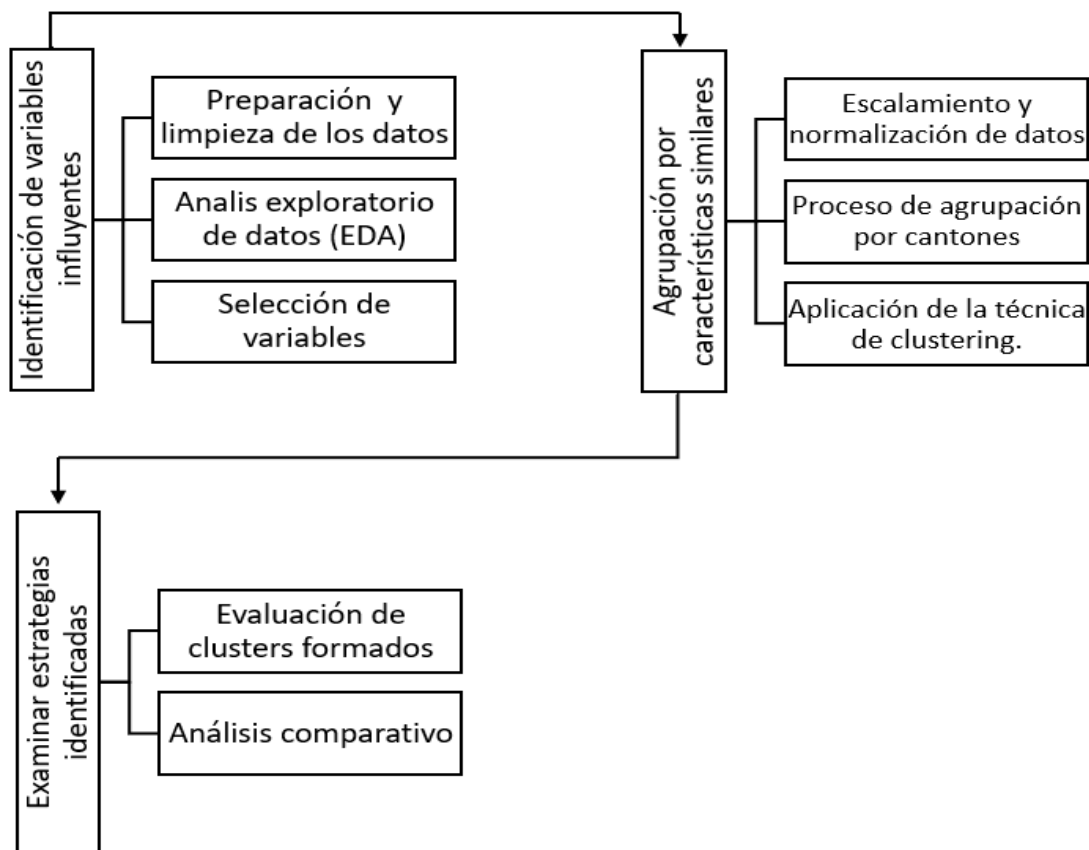
Dado que la investigación se basó en una base de datos existente, no se realizaron encuestas ni entrevistas. La información primaria necesaria se obtuvo de la base de datos que describe cómo se manejan los residuos sólidos en los hogares de los cantones de la provincia de Los Ríos. La base de datos está construida para el periodo 2015 a 2019, información que fue obtenida de los datos abiertos del Instituto Nacional de Estadísticas y Censos INEC.

3.6 PROCESAMIENTO Y ANÁLISIS

En la **Ilustración 2** se observa el flujo lógico y estructurado de las tareas y subtarefas que se realizaron en el procesamiento de los datos. Se encuentra dividido en las fases principales que apoyaron el cumplimiento de los objetivos específicos y, por consiguiente, el objetivo general.

Ilustración 2

Procesos para cumplir con los objetivos



Nota: Elaborado por la autora.

Revisión independiente y unificación de bases de datos

Antes de empezar a trabajar en cada objetivo de la investigación, fue necesario realizar una revisión preliminar de los datos obtenidos del Instituto Nacional de Estadísticas y Censos INEC (2023) para entender la información de cada base de datos disponible acerca del manejo de los residuos sólidos en diferentes cantones de la provincia de Los Ríos. Se

llevó a cabo una revisión preliminar para conocer su estructura y su diccionario de variables. Esta fase incluyó la unificación de formatos y la unificación y creación de una sola base de datos que incluya solo las variables que se relacionan a los residuos sólidos. Esta integración de los datos fue necesaria para que la información que se encontraba en múltiples conjuntos de datos quede en una única base unificada. Esto permitió asegurar un mejor análisis, lo que aumentó la calidad de los hallazgos de las siguientes etapas del proyecto.

3.6.1. Identificación de variables influyentes

3.6.1.1. Preparación y limpieza de los datos

Previo a la identificación de las principales variables influyentes en el manejo de los residuos sólidos, se realizó el proceso de carga y procesamiento de datos, eliminación o imputación de datos faltantes y la transformación de los datos. A continuación, se detallan las etapas realizadas en la preparación y limpieza de datos.

- **Carga y procesamiento de datos**

Se cargó al entorno de trabajo la base de datos ya unificada en un documento de MS Excel. Se realizó una inspección preliminar que permitió entender la estructura de los datos y tener una mejor visión del conjunto de datos, haciendo posible una mejor comprensión de las características presentes. Se pudo visualizar el número de filas, columnas y el tipo de datos en cada variable. Con esto se logró tener una mejor idea de la cantidad de información a utilizar.

Luego, se realizó el renombramiento de las variables. Esta etapa fue necesaria porque los nombres originales de las variables eran muy largos y complicados de manejar. Al simplificar y abreviar los nombres de las variables, se facilitó el uso y manipulación

durante el análisis de datos. De esta forma, se puede mantener una consistencia y mejor manejo dentro del análisis.

- **Eliminación e imputación de registros incompletos**

Para determinar las variables útiles y aplicar *clustering* es necesario que no existan registros incompletos o con un número alto de valores faltantes o perdidos. La revisión detallada de los datos reveló que existían diferentes registros con información faltante. Debido a que se trabajó con una base de datos ya existente, fue necesario revisar las columnas que presentaban más del 50 % de registros faltantes y proceder a su eliminación.

Para las variables que presentaron menos del 50 % de datos perdidos, se evaluaron diferentes estrategias de imputación. Para los datos numéricos, se evaluaron los métodos de la media, mediana y moda. Se utilizó el diagrama de caja de bigotes para comparar la distribución de cada variable antes y después de aplicar cada método de imputación. Al realizar este análisis de comparación visual de los efectos de cada método de imputación en la distribución de las variables numéricas, facilitó la toma de decisión informada sobre el tratamiento de los datos faltantes. Esta comparación permitió evaluar cómo cada método afecta a la distribución de los datos, ayudando a seleccionar el método que mejor preserva la estructura original de los datos y minimizar sesgos. Según Li et al. (2024), las métricas de la *Silhouette Score* ayudan a seleccionar la mejor estrategia de imputación. Basándonos en este estudio, se utilizó esta métrica para determinar la estrategia más adecuada que debía aplicarse a los datos numéricos de la base de datos que contiene información del manejo de los residuos sólidos no peligrosos de los cantones de la provincia de Los Ríos.

La imputación de datos faltantes en las variables categóricas fue realizada mediante la moda. Khan & Hoque (2020) destacan que la imputación de datos categóricos con el

método de la moda ayuda a mantener la distribución original de las categorías, minimizando la alteración de la estructura de los datos originales y reduciendo el sesgo que podría introducirse con métodos más complejos.

3.6.1.2. Análisis de Chi-Cuadrado y ANOVA

Para la elección de las variables más influyentes se determinó la independencia entre variables categóricas y numéricas. Primero se realizó la división entre variables categóricas y numéricas. Luego, para evaluar la independencia de las variables categóricas, se utilizó la prueba de Chi-Cuadrado. Para las variables numéricas en relación con las categóricas, se aplicó la prueba de análisis de varianza (ANOVA), evaluando si las medias de las variables numéricas difieren significativamente entre las categóricas (Whatley, 2022).

Pasos para el análisis de Chi-Cuadrado para variables categóricas

Se calcularon los valores de Chi-Cuadrado para todas las combinaciones de las variables categóricas, con la finalidad de evaluar la independencia de variables y obtener finalmente las más influyentes para la creación de los *clusters*.

Para el análisis de Chi-Cuadrado entre todas las combinaciones de variables categóricas e identificar su independencia se realizó el siguiente procedimiento:

1. Se crearon tablas de contingencia para cada par de variables categóricas, donde se identificaron las frecuencias de las combinaciones de categorías.
2. Se calculó el valor de Chi-Cuadrado, aplicando la prueba sobre cada tabla de contingencia para obtener el valor de Chi-Cuadrado, indicando el grado de asociación entre las dos variables.

3. Se construyó la matriz de correlación en la cual cada celda contiene el valor de Chi cuadrado entre las dos variables categóricas comparadas.

Análisis de ANOVA para variables numéricas y categóricas

Para la evaluación de independencia entre las variables numéricas y categóricas del conjunto de datos se utilizó el análisis de varianza ANOVA. A continuación, se detallan los pasos realizados:

1. Para el análisis de ANOVA, primero se separaron las variables numéricas y categóricas del conjunto de datos.
2. Para cada combinación de variables numéricas y categóricas se realizó la prueba de ANOVA, comparando las medias de la variable numérica entre distintas categorías de la variable categórica. Para la prueba de ANOVA se creó la siguiente hipótesis:
 - Hipótesis nula: Las medias de la variable numérica son iguales entre categorías.
 - Hipótesis alternativa: Al menos una de las medias es diferente.
3. Se calculó el valor p para cada prueba ANOVA, donde un valor alto de p , o sea mayor a 0,05, indica que no existe diferencia significativa entre medias, sugiriendo independencia entre variables.

Se filtraron las variables independientes. Se crearon las tablas de resultados donde se indican las variables numéricas independientes y las variables categóricas independientes, obteniendo de esta forma las variables influyentes que fueron utilizadas en el análisis posterior.

3.6.2. Agrupación por características similares

En esta etapa del proyecto se prepararon los datos para aplicar la técnica de *machine learning* de aprendizaje no supervisado. La técnica utilizada es la de *clustering*. Esto

consistió en elegir el algoritmo que mejor se ajuste al conjunto de datos. Los procesos realizados se encuentran detallados a continuación.

3.6.2.1. Escalamiento y normalización de datos

Antes de aplicar los algoritmos de *clustering*, se realizó la codificación *One-Hot* en cada una de las variables categóricas con más de dos categorías. Con esto se mejoró el consumo computacional. Esta codificación creó una columna para cada categoría de la variable original. Se aplicó el escalamiento de datos a las variables numéricas seleccionadas para garantizar la integridad y precisión del análisis. Este paso fue necesario para que todas las variables se encuentren en una escala comparable, permitiendo que el análisis se realizara de forma equitativa; además, mejoró la comparabilidad de las variables y facilitó la convergencia de los algoritmos de *clustering*, reduciendo el sesgo que podría surgir de las diferencias de las magnitudes de los datos. Con la normalización de los datos se logró una representación más precisa y coherente de las variables, lo que permitió interpretar correctamente los patrones y factores influyentes en la generación y manejo de residuos sólidos no peligrosos, proporcionando una base para los siguientes procesos a realizar en el trabajo (Han et al., 2011).

3.6.2.2. Proceso de agrupación por cantones

Para el proceso de agrupación de cantones según características similares, se utilizó el Análisis de Componentes Principales (PCA). Este método permitió reducir la dimensionalidad de los datos y resaltar las principales variables que influyen en la generación y manejo de residuos sólidos no peligrosos (Libório et al., 2022). En la **Tabla I** se muestran los pasos realizados para aplicar PCA a los datos, detallando cada etapa del procedimiento.

Tabla 1*Procesos para el análisis de componentes principales*

Proceso	Descripción
Preprocesamiento de los datos	Los datos fueron normalizados para garantizar que todas las variables tengan una escala comparable.
Cálculo de la matriz de covarianza	Se calculó la matriz de covarianza para comprender las relaciones entre las variables.
Selección de componentes principales	Se seleccionaron los primeros componentes principales que impliquen más del 90 % de la varianza de los datos.
Transformación de los datos	Los datos originales fueron transformados en el nuevo espacio de componentes principales.

Nota: Elaborado por la autora.

La aplicación de PCA facilitó la aplicación de los algoritmos de *clustering*, permitiendo agrupar los cantones de manera más eficiente y precisa en función de sus características similares.

3.6.2.3. Aplicación de técnicas de clustering

Se aplicaron tres algoritmos de *clustering*: K-Means, DBSCAN y *Clustering* Jerárquico. A continuación, se detallan los procesos realizados en cada algoritmo para obtener las agrupaciones de cantones según sus características similares.

K-means clustering

Primero se determinó el número óptimo de clusters, para este paso se aplicó el método del codo (*elbow method*) (Onumanyi et al., 2022). Luego de conocer el número óptimo de clusters para el algoritmo, se empleó K-means al conjunto de datos escalados para agrupar los cantones con características similares identificadas por el algoritmo. Por último, se realizó la visualización de los resultados de los clusters, graficándolos en las dos principales componentes, donde se identificó a qué clusters pertenece cada cantón.

DBSCAN clustering

Para agrupar los cantones por sus características similares, primero se realizó la determinación de los parámetros de ϵ (eps) y la cantidad mínima de muestra para el algoritmo DBSCAN. Luego, se ejecutó el algoritmo DBSCAN para de esta forma identificar clusters de forma arbitraria y manejar los puntos de ruido. Por último, se visualizaron los clusters resultantes en sus dos componentes principales.

Clustering jerárquico.

Primero se utilizó el método de enlace de Ward para construir un dendograma y determinar la estructura jerárquica de los cantones. Luego de tener el dendograma, se determinó el número de clusters a forma y a continuación aplicar el método de *Clustering* jerárquico. Por último, se visualizaron los clusters resultantes en sus dos componentes principales.

3.6.3. Comportamientos identificados

Se finalizó con el detalle del análisis de los *clusterings* identificados en la etapa anterior. Este paso fue importante para comprender cuáles son las similitudes que se presentan en la generación y manejo de residuos sólidos no peligrosos. Consistió en realizar el análisis detallado de los *clusterings*, con la presentación de los hallazgos en los datos y grupos.

3.6.3.1. Evaluación de clustering formados

Se revisaron los clusters formados por los diferentes algoritmos de *clustering* para identificar los grupos de cantones con características similares, con el objetivo de observar las separaciones de los clusters formados identificando si son representativas las características observadas.

CAPÍTULO IV

RESULTADOS Y DISCUSIÓN

4.1. PRINCIPALES VARIABLES QUE INFLUYEN EN LA GENERACIÓN DE RESIDUOS SÓLIDOS NO PELIGROSOS EN LOS CANTONES DE LA PROVINCIA DE LOS RÍOS

4.1.1. Preparación y limpieza de los datos

- **Carga y procesamiento de datos**

En la inspección preliminar de la base de datos se notó que estaba constituida de 5570 observaciones que corresponden a datos referentes al manejo de residuos sólidos no peligrosos de los diferentes cantones de la provincia de Los Ríos. Además, estaba compuesta por 33 columnas que corresponden a 27 variables cualitativas y 6 cuantitativas. La información que constituye la base de datos corresponde a los años 2015 hasta el 2019.

En la **Tabla 2** se presentan las variables originales con los nuevos nombres asignados luego del proceso de renombramiento de variables. Este proceso ayudó en la consistencia de los datos y permitió su uso en el estudio.

Tabla 2

Nuevo nombre de las variables

Nuevo nombre de la variable	Nombre antes de renombramiento
Año	AÑO
Área	AREA
Ciudad	CIUDAD
Cantón	CANTON
ResiduoHora	Durante los ÚLTIMOS 12 MESES, en su hogar, sacó la RESIDUO a las horas previstas de recolección
ClasificóResiduos	Durante los ÚLTIMOS 12 MESES, clasificó los RESIDUOS en su hogar
ResiduosOrgánicos	CLASIFICÓ durante los ÚLTIMOS 12 MESES, residuos: Orgánicos

ResiduosPapelCartón	CLASIFICÓ durante los ÚLTIMOS 12 MESES, residuos: Inorgánicos - Papel, cartón
ResiduosPlásticos	CLASIFICÓ durante los ÚLTIMOS 12 MESES, residuos: Inorgánicos - Plástico
ResiduosVidrio	CLASIFICÓ durante los ÚLTIMOS 12 MESES, residuos: Inorgánicos - Vidrio
EliminóResiduosOrgánicos	PRINCIPALMENTE, en su hogar durante los ÚLTIMOS 12 MESES, cómo eliminó los Residuos Orgánicos
EliminóResiduosPlásticos	PRINCIPALMENTE, en su hogar durante los ÚLTIMOS 12 MESES, cómo eliminó los Residuos de Plástico
EliminóResiduosVidrio	PRINCIPALMENTE, en su hogar durante los ÚLTIMOS 12 MESES, cómo eliminó los Residuos de Vidrio
EliminóResiduosPapelCartón	PRINCIPALMENTE, en su hogar durante los ÚLTIMOS 12 MESES, cómo eliminó los Residuos de Papel, cartón
RazónNoClasifica	Cuál es la PRINCIPAL RAZÓN por la que usted NO CLASIFICA:
UsaPilas	En su hogar Utiliza PILAS (control remoto, relojes, calefón, juguetes, etc.) que estén en funcionamiento
TotalPilas	Cuántas pilas tiene en total el hogar
PilasRecargables	De las pilas registradas cuántas son recargables
UsaFocos	En su hogar, utiliza FOCOS
TotalFocos	Cuántos focos tiene en total el hogar
FocosAhorradores	Cuántos focos ahorradores
FocosLed	De los focos ahorradores cuántos focos son LED
EliminóPilas	PRINCIPALMENTE, en su hogar durante los ÚLTIMOS 12 MESES, cómo eliminó: Las pilas
EliminóFarmacéuticos	PRINCIPALMENTE, en su hogar durante los ÚLTIMOS 12 MESES, cómo eliminó: Los desechos farmacéuticos (medicamentos)
EliminóElectrónicos	PRINCIPALMENTE, en su hogar durante los ÚLTIMOS 12 MESES, cómo eliminó: Desechos electrónicos / eléctricos
EliminóAceites	PRINCIPALMENTE, en su hogar durante los ÚLTIMOS 12 MESES, cómo eliminó: Aceite y/o grasas (cocina)
EliminóFocos	PRINCIPALMENTE, en su hogar durante los ÚLTIMOS 12 MESES, cómo eliminó: Focos ahorradores

ComprasFrecuentes	Durante el MES PASADO, al momento de hacer sus compras, en su hogar utilizó con mayor frecuencia
ProblemaAcumulaciónResiduo	Dentro de los siguientes PROBLEMAS AMBIENTALES, cuáles son los que afectan a su barrio - Acumulación de Residuo
VoluntariadoAmbiental	Durante los ÚLTIMOS 12 MESES, Ud. o algún miembro participó en voluntariados ambientales
DenunciarProblemaAmbiental	Durante los ÚLTIMOS 12 MESES, Ud. o algún miembro participó en - Denunciar personalmente un problema ambiental
FactorExpansión	Factor de expansión

Nota: Elaborado por la autora.

- **Eliminación e imputación de registros incompletos**

En la **Tabla 3** se presentan las variables que fueron eliminadas de la base de datos debido a que presentan un porcentaje mayor al 50 % de datos faltantes. Adicionalmente, se eliminaron otras variables que no son consideradas esenciales para el análisis.

Tabla 3
Variables eliminadas

Nombre de la variable	Motivo de eliminación
Ciudad	El estudio se va a realizar a nivel de cantón, no de ciudad.
FactorExpansión	No es necesario extrapolar datos.
Año	No se realizarán series temporales
RazónNoClasifica	64,51 % de datos faltantes
UtilizóOtro	99 % de datos faltantes

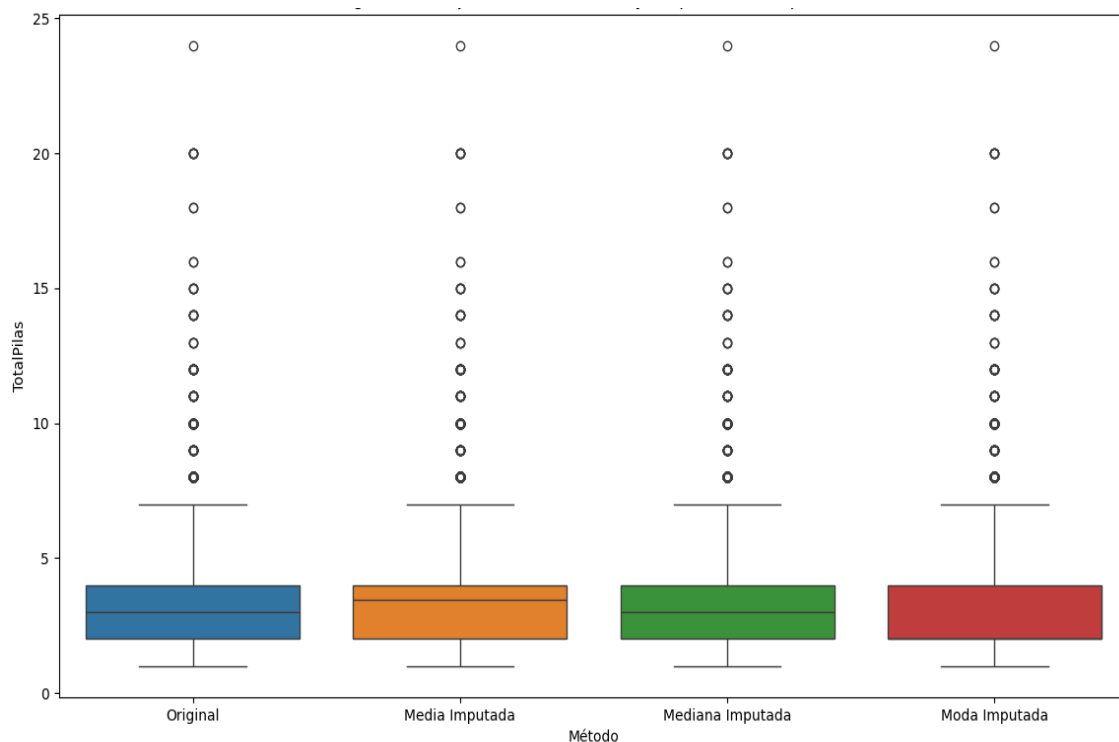
Nota: Elaborado por la autora.

En la **Ilustración 3** se presenta el diagrama de caja para comparar la distribución de la variable “TotalPilas” antes y después de la imputación de datos, aplicando los métodos

de la media, mediana y moda. Esto permitió entender qué método es el más apropiado para la imputación.

Ilustración 3

Diagrama de caja de “TotalPila” antes y después de la imputación.



Nota: Elaborado por la autora.

La distribución original de la variable “TotalPilas” muestra una cantidad considerable de valores atípicos que se encuentran dispersos hasta un valor cercano a 25 pilas, indicando una alta variabilidad en los datos. La línea dentro de la caja representa a la mediana de los datos originales, ubicada aproximadamente en 3 pilas por hogar. Los bigotes de la caja muestran que el 50 % central de los datos se encuentra entre 2 y 7 pilas.

Al imputar los valores faltantes con la media de la variable, se observa que la distribución permanece similar a la original. Sin embargo, hay un ligero cambio en la línea de la mediana, lo que indica que existe una pequeña reducción en la variabilidad de los datos.

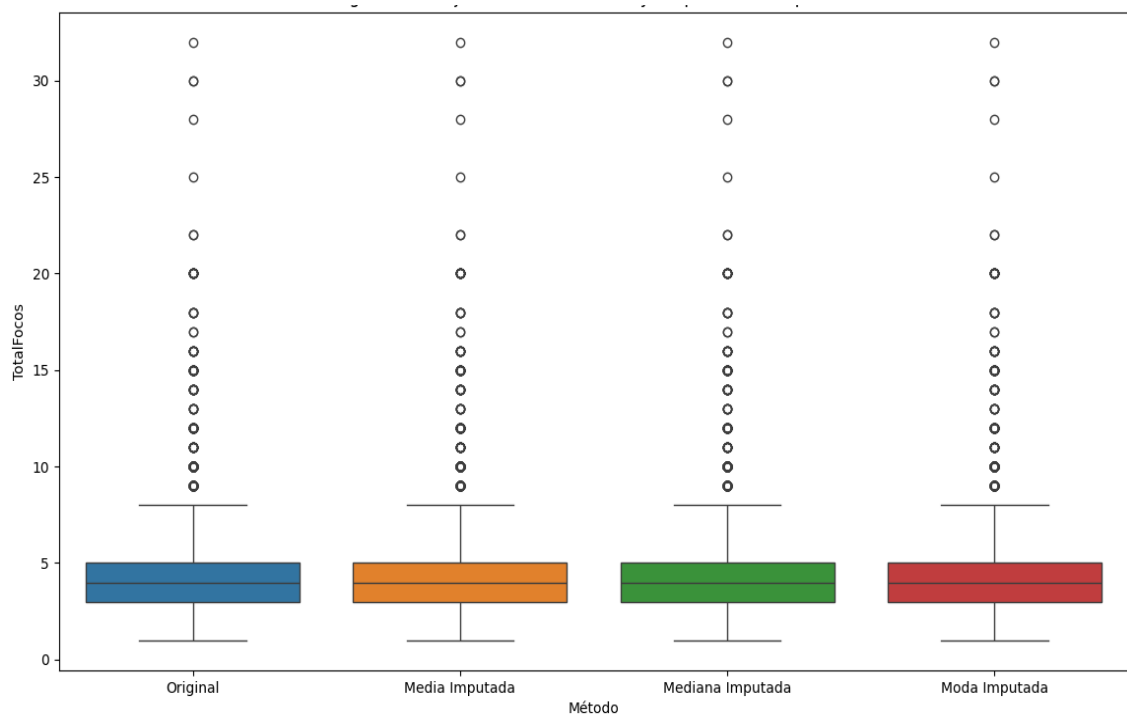
Cuando los valores faltantes son reemplazados por la mediana, la distribución resultante es muy similar a la original, conservando el valor de la mediana y sus cuartiles. A pesar de que la cantidad de valores atípicos es igual a la distribución original, la mediana tiende a ser menos influenciada por estos valores extremos, lo que ayuda a mantener una distribución central más estable.

En el caso de la imputación por moda, se observa una variabilidad en la línea de la mediana en comparación con la distribución original. Esto sugiere que este método puede introducir cambios en los datos, alterando la distribución respecto a los valores originales.

En la **Ilustración 4** se muestra la distribución de la variable “TotalFocos”, comparando los cambios que presenta después de aplicar los métodos de imputación por media, mediana y moda.

Ilustración 4

Diagrama de Caja de la variable "TotalFocos" antes y después de la imputación.



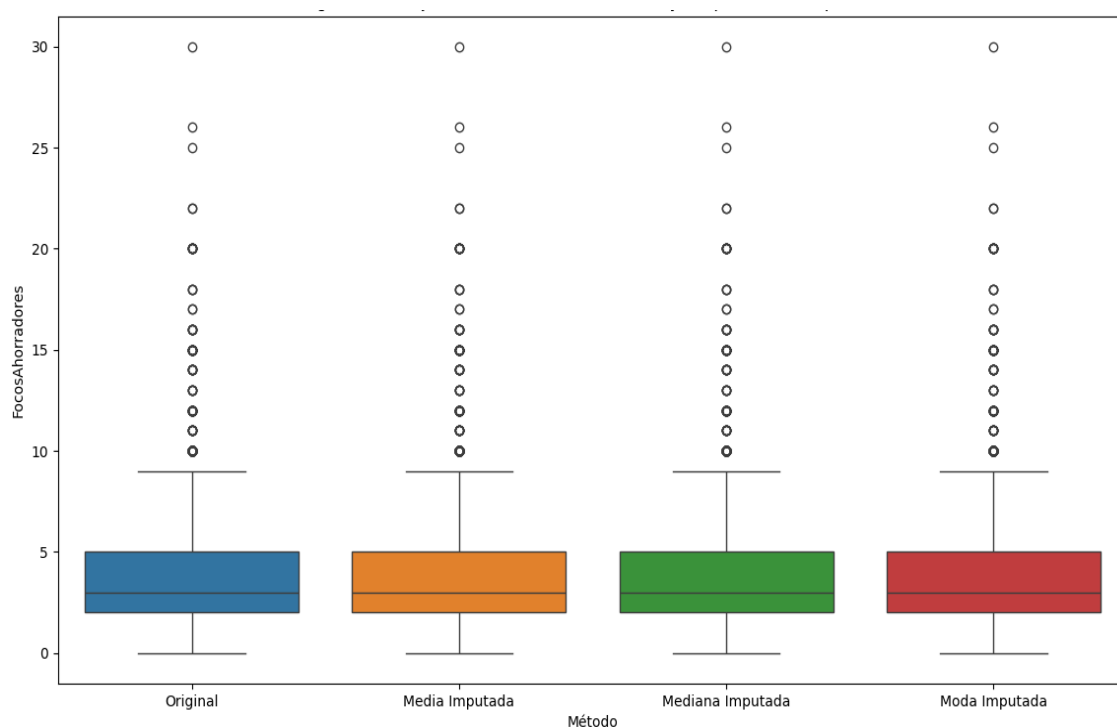
Nota: Elaborado por la autora.

La **Ilustración 4** muestra que los métodos de imputación por la media, mediana y moda tienen un impacto mínimo en la distribución de la variable “TotalFocos”, donde el valor de su mediana es aproximadamente 5, sus cuartiles se mantienen entre 3 y 7 focos y los valores atípicos están entre 10 a 30 focos por hogar.

A continuación, se muestra la **Ilustración 5** que compara la distribución de la variable “FocosAhorradores” después de aplicar los métodos de imputación por media, mediana y moda. El diagrama de caja facilita la comparación visual de los efectos de cada método de imputación en comparación con la distribución original de la variable.

Ilustración 5

Diagrama de Caja de la variable "FocosAhorradores" antes y después de la imputación.



Nota: Elaborado por la autora.

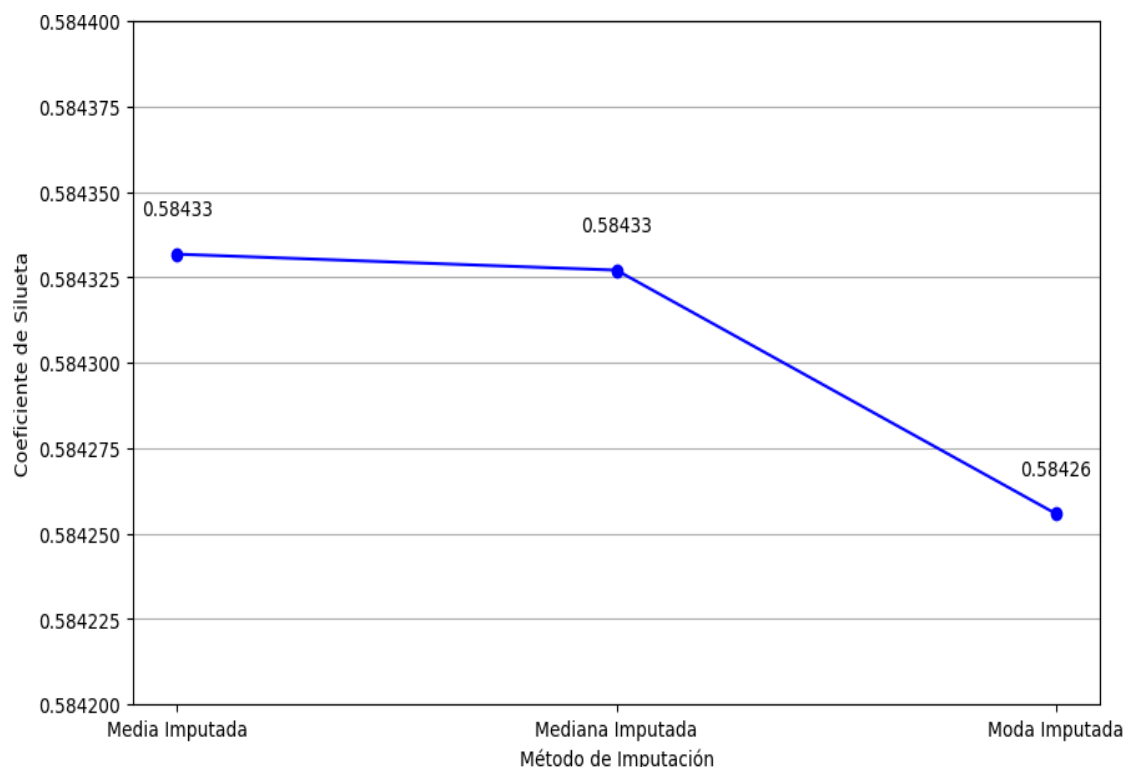
Se observa que al aplicar los tres métodos de imputación a la variable “FocosAhorradores” su distribución original no se ve significativamente afectada. En cada uno de los métodos, el valor de tendencia central se mantiene aproximadamente en

5 focos ahorradores por hogar. La variabilidad de los datos se concentra entre 3 a 7 focos ahorradores por hogar, lo que sugiere una consistencia en la distribución a pesar de la imputación. Los valores atípicos se encuentran en un rango de 20 a 30 focos ahorradores; también se mantienen similares en todos los métodos de imputación, indicando que ninguno de los métodos altera de manera significativa la presencia de estos valores extremos.

En la **Ilustración 6** se muestra el análisis comparativo del coeficiente de silueta para los tres métodos de imputación: media, mediana y moda, aplicados en las variables numéricas. Se observa que la moda presenta una variabilidad en su resultado en comparación con la imputación por media y mediana. Mientras más alto es el valor del coeficiente de silueta, mejor es el método de imputación.

Ilustración 6

Comparación del Coeficiente de Silueta por Método de Imputación



Nota: Elaborado por la autora.

El coeficiente de silueta para la imputación por media es de 0,58433, representando a un valor de silueta alto, lo que indica que existe una buena cohesión de los datos. De manera similar, la imputación por mediana muestra un valor de silueta alto, siendo un valor de 0,58433, lo que indica que este método también es válido para realzar la imputación de datos en las variables numéricas de la base de datos.

Por otro lado, la imputación por moda, con un coeficiente de silueta ligeramente más bajo 0,58426, sugiere una menor cohesión de los datos. Debido a este menor rendimiento, se recomienda descartar el uso del método de imputación por moda, decidiendo por la media o la mediana.

De acuerdo con los resultados presentados, y basándose en el trabajo de Maheswari et al. (2020), se eligió la mediana como método de imputación para los datos numéricos. A pesar de que las diferencias entre los métodos de media y mediana son mínimas, la mediana se considera el método que mejor preserva las características originales de los datos. La presencia de datos atípicos justifica la elección de este método, ya que no se ve afectado por estos valores, resultando menores sesgos en análisis posteriores. Esta elección asegura que la imputación por mediana refleja de manera más precisa la tendencia central de los datos sin ser influenciada por valores extremos.

4.1.2. Análisis de Chi-Cuadrado para variables categóricas

En la **Tabla 4** se muestran los resultados de las pruebas de independencia chi-cuadrado, ilustrando la relación entre diversas combinaciones de variables categóricas.

Tabla 4*Resultados de la prueba de chi-cuadrado*

Variable 1	Variables 2	Valor de Chi-Cuadrado
Área	ClasificóResiduos	4,31
Área	ResiduosPapelCartón	11,60
Área	ResiduosVidrio	7,52
ClasificóResiduos	UsaPilas	8,38
ClasificóResiduos	UsaFocos	0,81
ClasificóResiduos	ComprasFrecuentes	3,98
ClasificóResiduos	ProblemaAcumulaciónResiduo	3,49
ResiduosOrgánicos	UsaPilas	0,56
ResiduosOrgánicos	UsaFocos	0,03
ResiduosOrgánicos	VoluntariadoAmbiental	12,88
ResiduosOrgánicos	DenunciarProblemaAmbiental	11,16
ResiduosPapelCartón	UsaPilas	0,49
ResiduosPapelCartón	UsaFocos	0,20
ResiduosPapelCartón	ComprasFrecuentes	9,20
ResiduosPapelCartón	ProblemaAcumulaciónResiduo	2,61
ResiduosPlásticos	UsaPilas	6,77
ResiduosPlásticos	UsaFocos	1,24
ResiduosPlásticos	ComprasFrecuentes	1,92
ResiduosPlásticos	ProblemaAcumulaciónResiduo	3,73
ResiduosPlásticos	DenunciarProblemaAmbiental	14,07
Área	DenunciarProblemaAmbiental	14,02

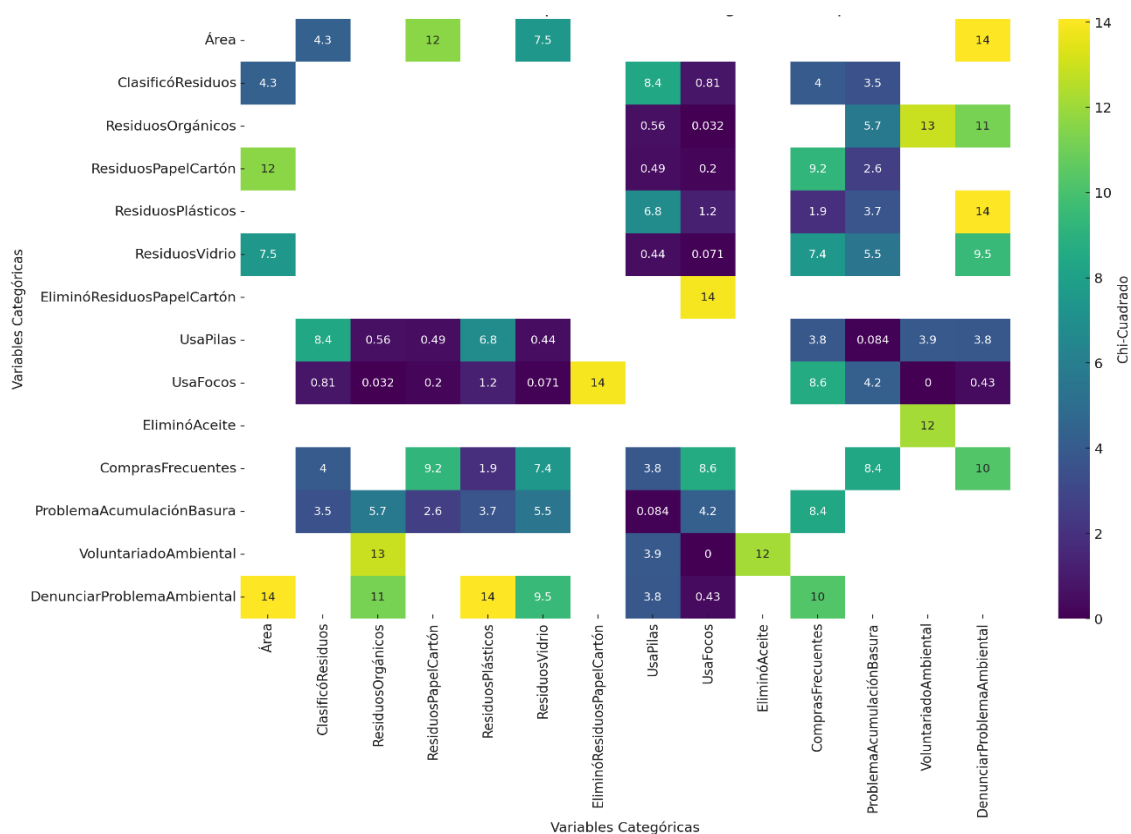
Nota: Elaborado por la autora.

Se identificaron las variables categóricas independientes, evaluando los valores de Chi-Cuadrado que cumplen con la regla del valor umbral menores a 15. Un valor de Chi Cuadrado menor a 15 generalmente indica una asociación débil entre las variables. En la tabla se observa que los valores de Chi-Cuadrado entre las comparaciones de variables son menores a 15, indicando que existe independencia entre las variables categóricas.

En la **Ilustración 7** se observan los valores de Chi-Cuadrado para las combinaciones de las variables categóricas que presentaron un alto grado de independencia, donde el valor de Chi-Cuadrado en la relación de variables es menor a 15; mientras el valor de Chi cuadrado es más bajo, más independiente es la variable.

Ilustración 7

Valores de Chi-Cuadrado para variables categóricas.



Nota: Elaborado por la autora.

4.1.3. Análisis de ANOVA para las variables numéricas y categóricas ($p > 0,05$)

En la **Tabla 5** se muestran las combinaciones de las variables numéricas y categóricas que mostraron un valor p mayor que 0,05, indicando que no existe una diferencia significativa entre las medias; por lo tanto, son variables independientes.

Tabla 5*Comparación de variables numéricas y categóricas.*

Variable numérica	Variable categórica	Valor p
TotalPilas	ResiduosOrgánicos	0,201421
TotalPilas	ResiduosPapelCartón	0,059691
TotalPilas	ResiduosVidrio	0,836748
TotalPilas	ComprasFrecuentes	0,160619
PilasRecargables	ResiduosVidrio	0,836748
PilasRecargables	VoluntariadoAmbiental	0,092553
PilasRecargables	ComprasFrecuentes	0,178841
TotalFocos	ClasificóResiduos	0,499194
TotalFocos	ResiduosOrgánicos	0,567124
TotalFocos	ComprasFrecuentes	0,623096
TotalFocos	ProblemaAcumulaciónResiduo	0,087894
FocosAhorradores	ClasificóResiduos	0,499194
FocosAhorradores	ResiduosOrgánicos	0,567127
FocosAhorradores	UsaFocos	0,474980
FocosAhorradores	ComprasFrecuentes	0,215426
FocosLED	ClasificóResiduos	0,561850
FocosLED	ResiduosVidrio	0,180486
FocosLED	UsaFocos	0,158551
FocosLED	ComprasFrecuentes	0,437515

Nota: Elaborado por la autora.

Los resultados de ANOVA muestran que las combinaciones de las variables listadas en la tabla no presentan diferencias significativas en las medias de las variables numéricas entre las categorías correspondientes. Si se observa el valor de p para las variables “TotalPilas” y “ResiduosOrgánicos”, se nota que la cantidad total de pilas no varía significativamente según la clasificación de los residuos orgánicos, sugiriendo que la cantidad de pilas utilizadas no depende de si los residuos orgánicos son clasificados o no.

Luego de aplicar estos dos métodos para la evaluación de independencia de las variables presentes en la base de datos, se obtuvieron las variables que se consideran más importantes para trabajar las agrupaciones de los cantones de acuerdo con sus características de manejo de residuos sólidos no peligrosos. En la **Tabla 6** se observan las variables que fueron utilizadas en la aplicación de los diferentes grupos de algoritmos de *clustering*, representando una cantidad total de dieciséis variables.

Tabla 6
Variables útiles para el análisis de clusters

Variables numéricas	Variables categóricas
TotalPilas	ClasificóResiduos
PilasRecargables	ResiduosOrgánicos
TotalFocos	ResiduosPapelCartón
FocosAhorradores	ResiduosVidrio
FocosLED	EliminóResiduosPapelCartón
	UsaPilas
	UsaFocos
	ComprasFrecuentes
	ProblemaAcumulaciónResiduo
	VoluntariadoAmbiental
	DenunciarProblemaAmbiental

Nota: Elaborado por la autora.

4.1.4. Discusión

La identificación de las principales variables que influyen en la generación y manejo de los residuos sólidos no peligrosos es esencial para tener una mejor visión de las características de las prácticas de los diferentes cantones de la provincia de Los Ríos. Con el análisis preliminar realizado en el trabajo, se puede observar que los factores como la clasificación de los residuos, el uso de bolsas plásticas y las técnicas de eliminación, juegan un papel importante en la generación de residuos.

Los procesos realizados en cada uno de los pasos para identificar las principales variables influyentes en el manejo de los residuos sólidos fueron cruciales para el análisis posterior de los datos. La eliminación de registros incompletos e imputación de datos faltantes representó un paso importante para garantizar la integridad del análisis. Esto va acorde a lo señalado por Jäger et al. (2021), quienes resaltan la importancia de mantener la calidad de los datos, debido a que en casi todo análisis se van a encontrar datos faltantes que son necesarios tratar para mantener la calidad necesaria de los datos para evitar sesgos en los resultados.

Además, Templ (2023) indica que es importante probar los diferentes métodos de imputación que existen y decidir por el que no representa un gran impacto de valores extremos, permitiendo mantener la consistencia de los datos. Además, Li et al. (2024) evaluaron la efectividad de diferentes métodos de imputación y encontraron que métodos sencillos como la imputación por la mediana pueden ofrecer un buen rendimiento para los datos faltantes.

Así, en este trabajo se realizaron pruebas de imputación con la media, mediana y la moda y de esta forma se determinó el método de imputación de datos que fue más efectivo para tratar los valores faltantes en el conjunto de datos, lo que se realizó de acuerdo con el tipo de variables presentes, numéricas y categóricas.

Además, fue necesario realizar pruebas de independencias de variables. La independencia entre variables reduce la multicolinealidad, mejorando la estabilidad y la interpretación de los *clusterings*. Esta idea se basa en el trabajo de Caruso et al. (2019) que indica que las variables independientes permiten que los clusters reflejen mejor las características subyacentes de los datos.

Asimismo, la selección de variables independientes asegura que la información redundante sesgue los resultados de *clustering*. Watson (2014) menciona que las pruebas de independencia, como Chi-Cuadrado para las variables categóricas, son fundamentales para identificar asociaciones significativas y evitar redundancia en el análisis. En este trabajo, se aplicaron estas técnicas para identificar las variables con mayor independencia entre ellas. Al seleccionar las variables con mayor independencia, permite optimizar la formación de los clusters, permitiendo que el análisis de *clustering* tenga una agrupación de cantones más precisa y con menos sesgos posibles.

4.2. TÉCNICAS DE CLUSTERING PARA AGRUPAR CANTONES CON CARACTERÍSTICAS SIMILARES EN TÉRMINOS DE GENERACIÓN Y MANEJO DE LOS RESIDUOS SÓLIDOS NO PELIGROSOS

La codificación *One-Hot* creó una columna para cada categoría de la variable original. En la **Ilustración 8** se puede observar cómo se van asignando los valores de 1 y 0, donde uno representa la presencia de una característica y 0 la ausencia de ella. Esta transformación de las variables categóricas a otro formato fue necesaria para aplicar los algoritmos de *clustering*.

Ilustración 8

One-Hot encoding en variables categóricas

Nombres de los cantones	1	1	Babahoyo	0	0	0	0	0	1	1	1	...	0	0	0	0	0	0	0	0	0
	2	1	Babahoyo	0	0	0	0	0	0	1	0	...	0	0	0	0	0	0	0	0	0
	3	1	Babahoyo	0	0	0	0	0	1	1	0	...	0	0	0	0	0	0	0	0	0
	4	1	Babahoyo	0	0	0	0	0	1	1	0	...	0	0	0	0	0	0	0	0	0
	5	1	Babahoyo	0	0	0	0	0	1	1	0	...	0	0	0	0	0	0	0	0	0

Asignación de valores según la codificación One - Hot

Nota: Elaborado por la autora.

En la **Ilustración 9** se observan los valores agregados por cantón, mostrando las proporciones y promedios de las variables seleccionadas. Estos valores permiten una comprensión más profunda de las características y comportamientos de los residuos sólidos no peligrosos en cada cantón, facilitando así la identificación de áreas que requieren atención particular y la implementación de estrategias de gestión más efectivas. La visualización de los datos agregados resalta las diferencias y similitudes entre cantones,

Ilustración 9

Agrupación de valores agregados por cantón de la provincia

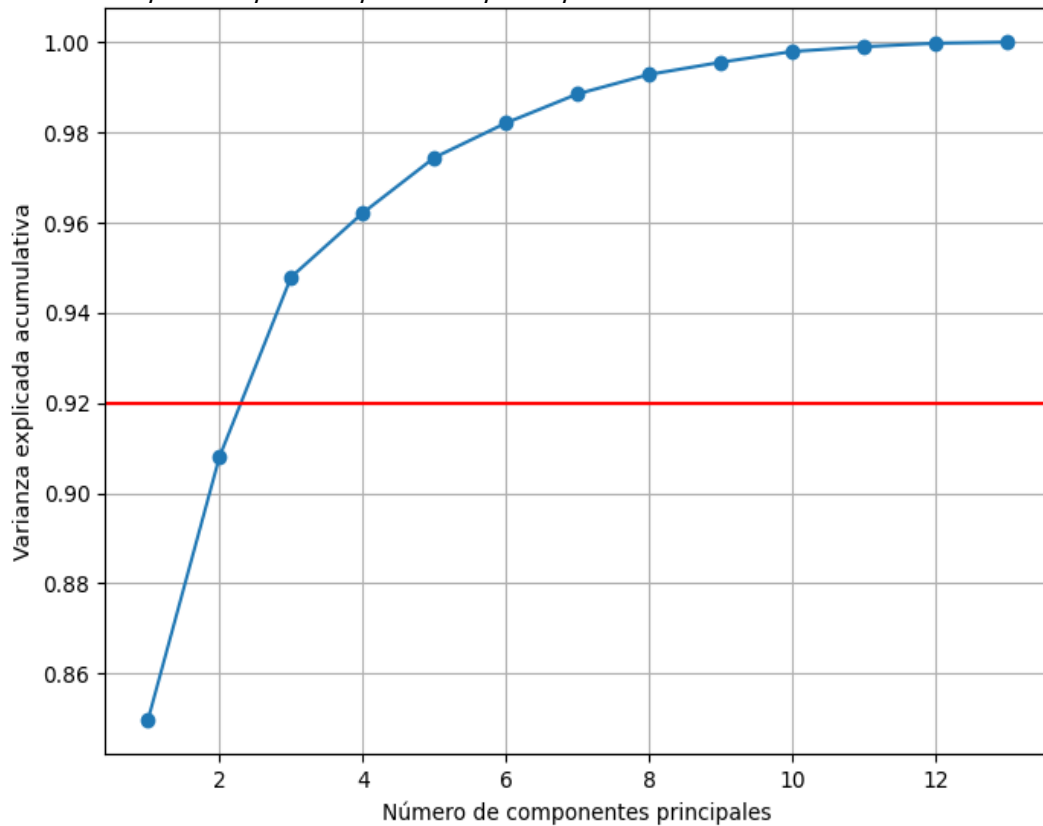
Cantones de la provincia de Los Ríos	Baba	0.335329	0.224551	0.320359	0.071856	0.209581	0.050898	0.724551	0.985030	0.104790	0.014970
	Babahoyo	0.653571	0.205357	0.147321	0.141071	0.291071	0.076786	0.833036	0.987500	0.232143	0.041071
	Buena Fe	0.830472	0.263948	0.289700	0.214592	0.302575	0.098712	0.841202	0.995708	0.182403	0.034335
	Mocache	0.284038	0.295775	0.248826	0.112676	0.225352	0.077465	0.840376	0.988263	0.159624	0.016432
	Montalvo	0.765625	0.242188	0.250000	0.140625	0.296875	0.062500	0.757812	1.000000	0.343750	0.046875
	Palenque	0.258427	0.117978	0.264045	0.129213	0.191011	0.106742	0.842697	0.988764	0.089888	0.022472
	Puebloviejo	0.213415	0.158537	0.420732	0.237805	0.274390	0.219512	0.823171	0.981707	0.170732	0.042683
	Quevedo	0.942584	0.228708	0.157895	0.187560	0.291866	0.086124	0.852632	0.997129	0.297608	0.052632
	Quinsaloma	0.198864	0.113636	0.238636	0.142045	0.204545	0.085227	0.823864	0.965909	0.130682	0.022727
	Urdaneta	0.805668	0.182186	0.145749	0.113360	0.246964	0.076923	0.744939	0.971660	0.295547	0.044534
	Valencia	0.707989	0.256198	0.297521	0.181818	0.283747	0.101928	0.815427	0.991736	0.143251	0.096419
	Ventanas	0.531323	0.262181	0.324826	0.206497	0.310905	0.106729	0.791183	0.993039	0.180974	0.025522
	Vinces	0.404472	0.126016	0.199187	0.130081	0.160569	0.093496	0.770325	0.981707	0.199187	0.014228
	Valor promedio de cada variable										

Nota: Elaborado por la autora.

Por otro lado, en la **Ilustración 10** se muestra la varianza explicada por componentes principales. Aquí se observa que más del 90 % de la varianza se encuentra explicada por las dos primeras componentes principales. Este resultado justifica la reducción de dimensionalidad a dos componentes, facilitando la visualización del análisis de los datos sin perder información significativa.

Ilustración 10

Varianza explicada por componentes principales



Nota: Elaborado por la autora.

El resultado de aplicar PCA a los datos fue una reducción de la tabla a dos componentes principales para cada cantón. Esta reducción de dimensionalidad facilitó la visualización de los clusters y la identificación de patrones en los datos.

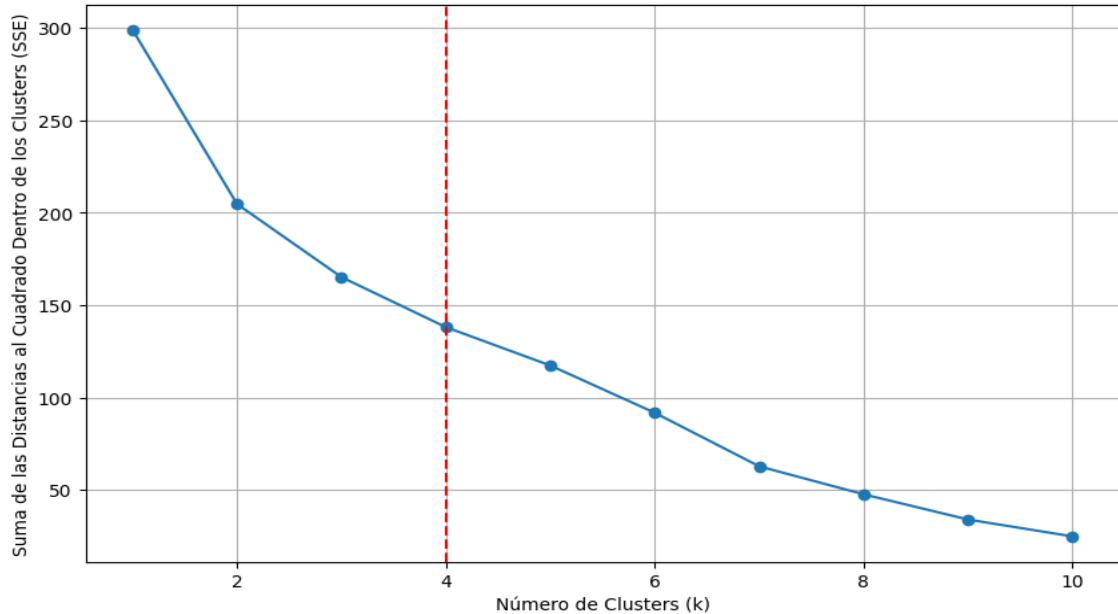
4.2.1. Aplicación de técnicas de *clustering*

Algoritmo K-Means

En la **Ilustración 11** se presenta el método del codo, permitiendo identificar el número óptimo de clusters. En el eje X se representa el número de clusters (k) y en el eje Y se muestra la suma de las distancias al cuadrado dentro de los clusters (SSE), permitiendo visualizar que en el valor k el codo comienza a disminuir de manera menos pronunciada.

Ilustración 11

Método del codo para determinar número óptimo de clusters en K-means.



Nota: Elaborado por la autora.

Como resultado del análisis del método del codo se obtuvo que el valor óptimo de clusters para este conjunto de datos es cuatro. Esto quiere decir que, al agrupar los cantones en cuatro clusters, se obtiene una segmentación adecuada que minimiza la variación dentro de cada clusters sin perder la eficiencia de la segmentación.

Luego de determinar el número óptimo de clusters se aplicó el algoritmo de K-means a los datos. En la **Tabla 7** se pueden observar los cantones asignados a cada cluster formado por el algoritmo de K-means.

Tabla 7

Clusters formados por K-means

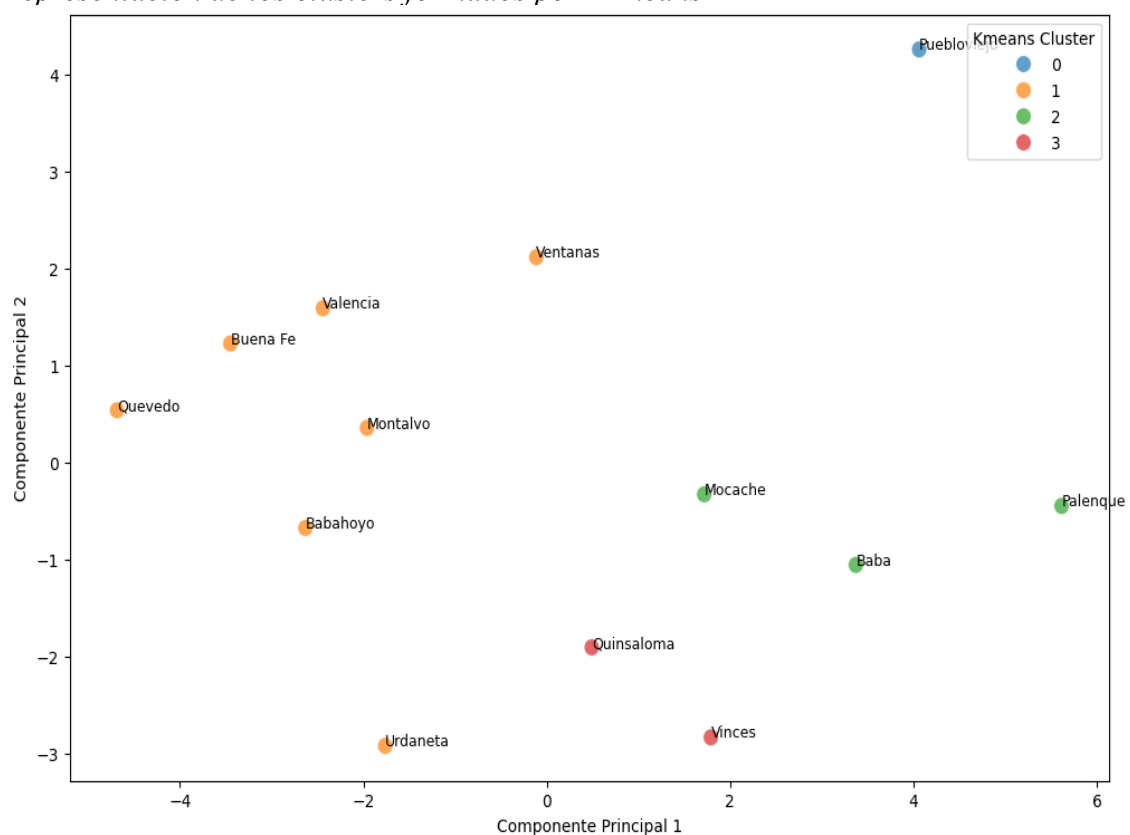
Número de clusters	Cantones
Kmeans Cluster 0	Puebloviejo
Kmeans Cluster 1	Quevedo, Buena Fe, Valencia, Montalvo, Babahoyo, Ventanas y Urdaneta
Kmeans Cluster 2	Mocache, Baba y Palenque
Kmeans Cluster 3	Quinsaloma y Vines

Nota: Elaborado por la autora.

En la **Ilustración 12** se muestra la visualización de los clusters obtenidos mediante el algoritmo K-means utilizando PCA para reducir la dimensionalidad de los datos a dos componentes principales. Cada punto en la ilustración representa un cantón, coloreado según el cluster al que pertenece. Se observa la distribución de los cantones en el espacio bidimensional definido por las componentes principales. Al visualizar los clusters dentro de las componentes facilita una mejor comprensión de la formación de clusters.

Ilustración 12

Representación de los clusters formados por K-means



Nota: Elaborado por la autora.

Características de agrupación para los clusters creados por K-means

La agrupación de cantones por clusters se realizó de acuerdo con características similares. Para identificar estas características se consideraron las variables que representaban un valor medio más alto en comparación con otros clusters. En la **Tabla 8** se detallan las variables con características similares en cada agrupación.

Tabla 8
Características de agrupación por K-means

Clusters	Característica	Valor medio
Cluster 0	Eliminó residuos plásticos	2,561
	Eliminó residuos papel/cartón	2,512
	Eliminó aceite	4,366
	Eliminó focos	2,134
Cluster 1	Residuo hora	1,765
	Clasificó residuos	0,234
	Voluntariado ambiental	0,049
	Denunciar problemas ambientales	0,039
Cluster 2	Eliminó residuos orgánicos	2,156
	Residuos orgánicos	0,278
	Compras frecuentes	0,178
Cluster 3	Eliminó residuos vidrio	2,018
	Focos LED	0,174
	Eliminó focos	1,807
	Problemas de acumulación de residuo	0,165

Nota: Elaborado por la autora.

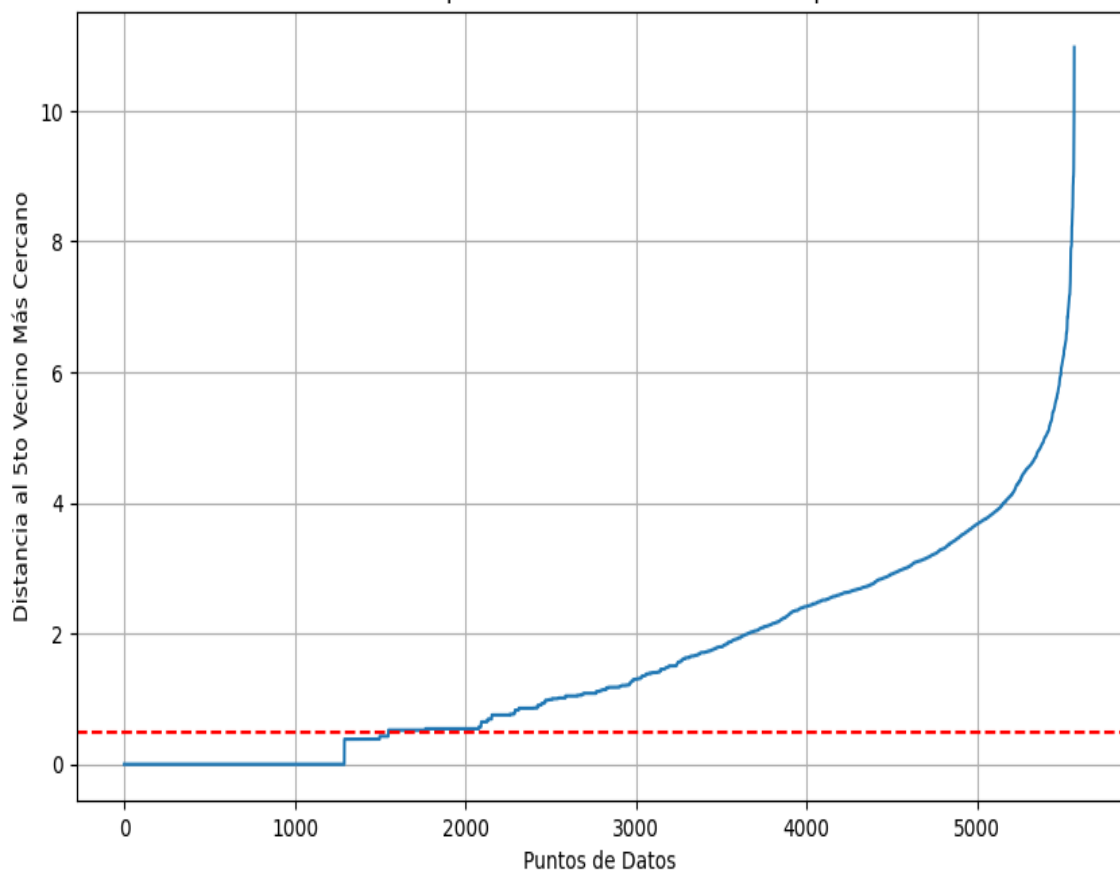
Al aplicar el algoritmo de agrupación K-Means, se segmentaron los cantones de acuerdo con los valores medios más similares entre ellos, considerando que estas variables representaban estrategias muy similares aplicadas en cada cantón. De esta manera, en un análisis posterior se identificaron estrategias aplicadas en cada agrupación.

DBSCAN

En la **Ilustración 13** se presenta el método de rodilla, utilizado para determinar el parámetro “eps” en el algoritmo DBCSCAN. En el eje X se muestra los puntos de los datos y en el eje Y la distancia al quinto vecino más cercano. Cuando la pendiente de la curva cambia de manera más pronunciada, se considera un valor “eps” óptimo.

Ilustración 13

Método de Rodilla para DBSCAN



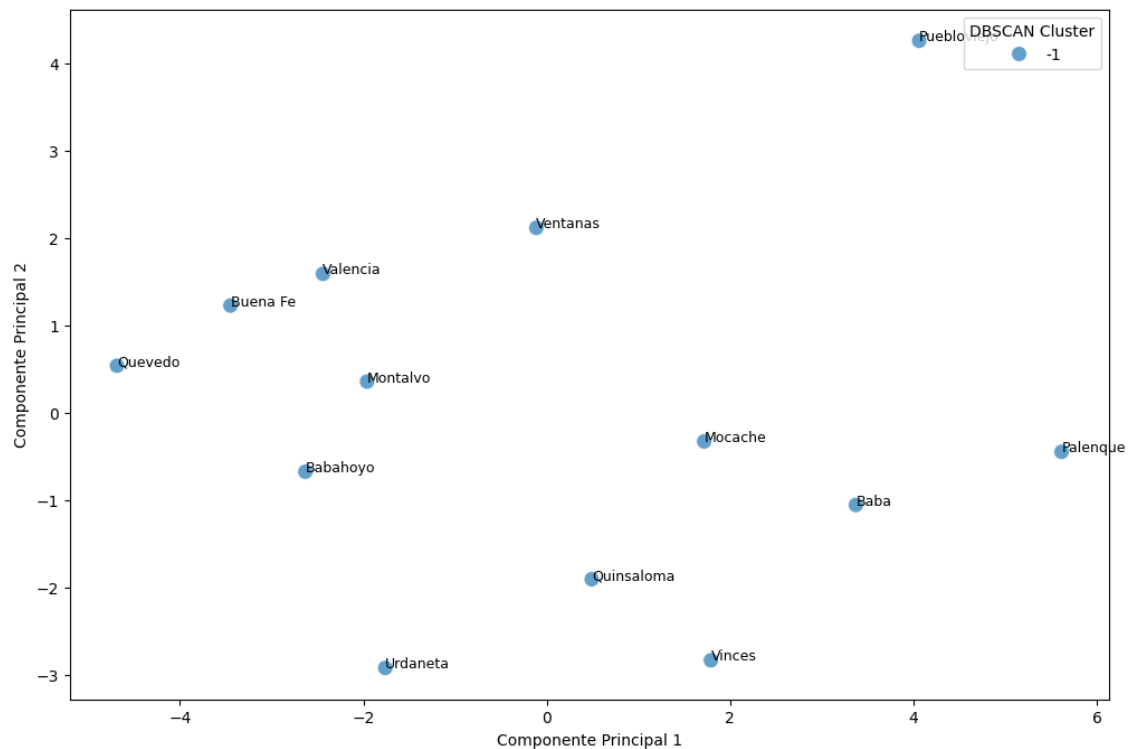
Nota: Elaborado por la autora.

El análisis del método de rodilla sugiere que el parámetro “eps” óptimo para aplicar el algoritmo DBSCAN es aproximadamente 0,5. Este valor será utilizado para ejecutar el algoritmo y obtener las agrupaciones de los cantones de la provincia.

En la **Ilustración 14** se muestra la visualización de los clusters obtenidos después de aplicar el algoritmo DBSCAN al conjunto de datos. Se utilizó PCA para reducir la dimensionalidad de los datos a dos componentes principales. Cada punto en la ilustración representa un cantón.

Ilustración 14

Visualización de clusters DBSCAN



Nota: Elaborado por la autora.

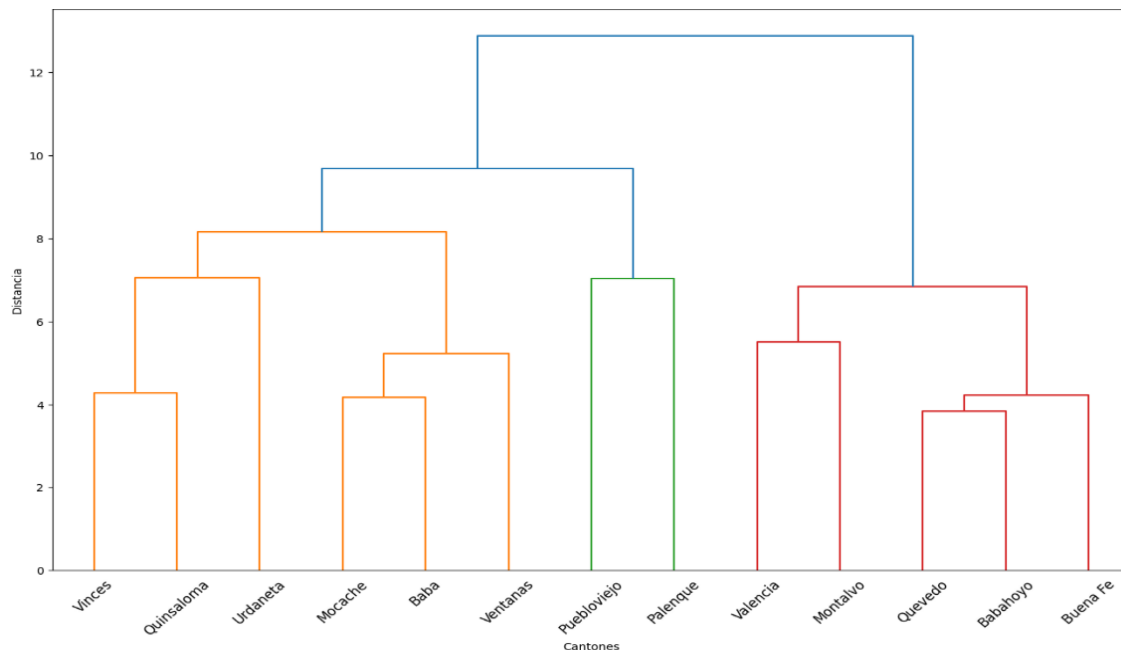
El análisis mediante DBSCAN identifica a los cantones como ruido; estos se reconocen porque se encuentran marcados con -1, valor que indica que los cantones no pertenecen a clusters bien definidos. Estos resultados son útiles cuando se desea detectar áreas con características únicas o menos comunes en la generación y manejo de residuos.

Clustering Jerárquico

En la **Ilustración 15** se muestra el dendograma resultante del *clustering* jerárquico aplicado a los datos escalados de los cantones. Se visualiza la manera en que los cantones se agrupan en función de sus similitudes. Los cantones que se encuentran más cerca en el dendograma son similares entre sí. Además, este dendograma indica cómo los cantones se encuentran jerárquicamente agrupados en clusters, proporcionando una representación visual de las relaciones y similitudes entre ellos.

Ilustración 15

Dendograma del clustering jerárquico



Nota: Elaborado por la autora.

En la **Nota:** Elaborado por la autora.

Ilustración 16 se muestra la visualización de los clusters obtenidos mediante el algoritmo de *clustering* jerárquico utilizando PCA para reducir la dimensionalidad de los datos a dos componentes principales. Cada punto de la ilustración representa a un cantón de la provincia de Los Ríos. En la **Tabla 9** se detalla el nombre del cluster formado con este algoritmo y los cantones que pertenecen a cada grupo.

Tabla 9

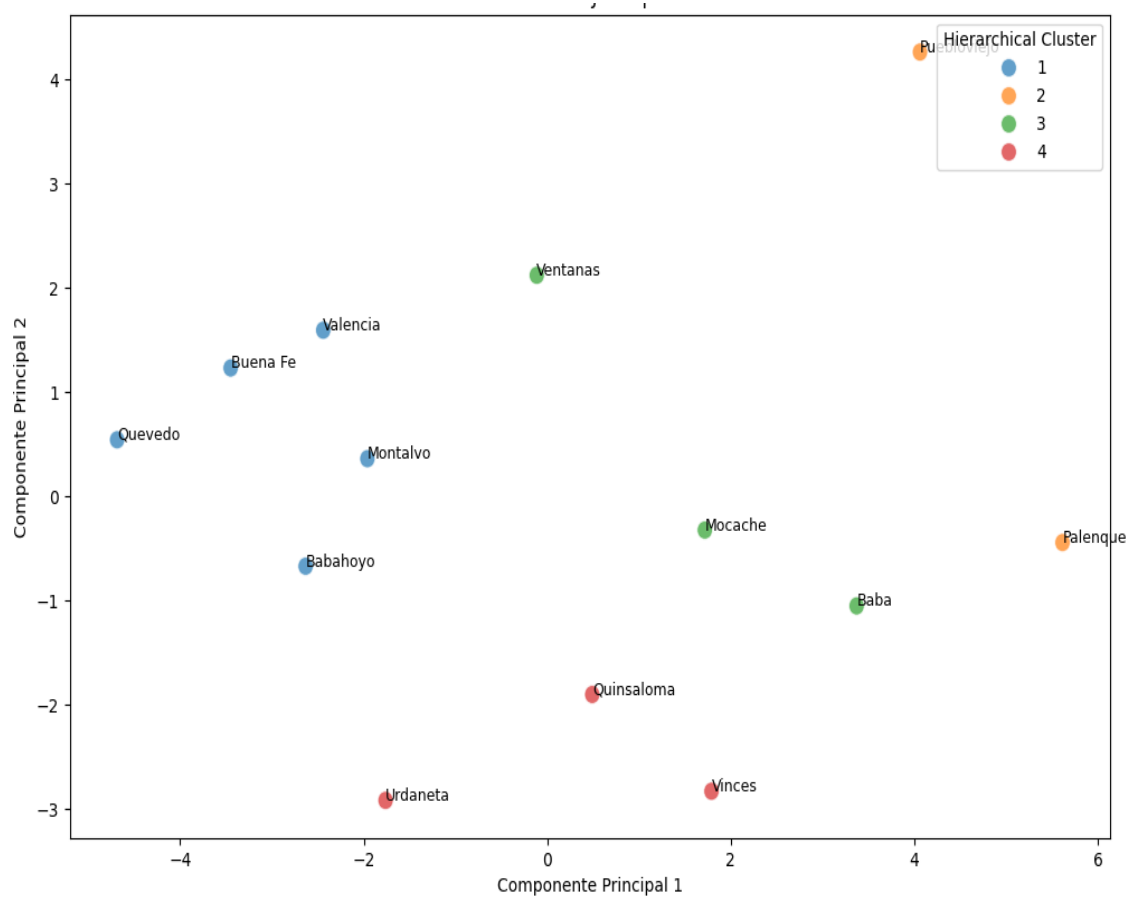
Cantones asignados a cada clusters por clusters jerárquico

Clusters	Cantones
Cluster 1	Quevedo, Buena Fe, Valencia, Montalvo y Babahoyo
Cluster2	Pueblviejo y Palenque
Cluster 3	Ventanas, Mocache y Baba
Cluster 4	Urdaneta, Quinsaloma y Vincés

Nota: Elaborado por la autora.

Ilustración 16

Visualización de clusters jerárquico



Nota: Elaborado por la autora.

Al visualizar la distribución de los cantones en el espacio bidimensional de las componentes principales, se pueden identificar claramente los clusters formados mediante el algoritmo de *clustering* jerárquico. A continuación,

Características sobresalientes de agrupación para Clustering Jerárquico

La **Tabla 10** detalla las variables con características similares en cada agrupación de cantones generadas al aplicar el algoritmo de *Clustering* Jerárquico al conjunto de datos. Además, se visualiza el valor medio de cada variable en comparación a los demás clusters.

Tabla 10*Características de agrupación para clustering jerárquico*

Clusters	Característica	Valor medio
Cluster 1	Clasificó residuos	0,239
	Residuos plásticos	0,293
	Voluntariado ambiental	0,054
	Denuncias voluntariado ambiental	0,046
Cluster 2	Residuos orgánicos	0,342
	Eliminó residuos plásticos	2,564
	Eliminó residuos papel/cartón	2,408
	Eliminó aceites	4,638
Cluster 3	Clasificó residuos	0,261
	Eliminó aceites	3,763
	Eliminó focos	1,939
Cluster 4	Eliminó residuos vidrio	1,457
	Focos LED	0,148
	Problemas acumulación de residuo	0,208

Nota: Elaborado por la autora.

Las características seleccionadas como sobresalientes en cada clúster son aquellas que tienen valores más significativos en comparación con otros clusters. Estas características de agrupación permiten identificar patrones predominantes y prácticas de manejo de residuos sólidos no peligrosos en cada grupo de cantón.

4.2.2. Discusión

El uso de técnicas de clasificación como K-Means y *clustering* jerárquico permitió identificar características importantes en la producción y manejo de residuos sólidos no peligrosos entre los cantones de la provincia de Los Ríos.

Uno de los principales desafíos en el análisis de datos es el costo computacional que puede representar. Para mejorar esta parte del análisis se aplicó la reducción de

dimensionalidad mediante PCA (Análisis de Componentes Principales). Boutsidis et al. (2014) indican que aplicar PCA ayuda a mantener la mayor parte de variabilidad de los datos simplificándolo y transformando las variables originales en un número menor de componentes principales. Esto no solo mejora el costo computacional, sino que además mejora la eficiencia de los algoritmos de *clustering* y elimina el ruido y la redundancia presente en los datos.

Festa et al. (2023) en cambio hacen referencia al uso del algoritmo de K-Means como el algoritmo de agrupación con mejores resultados, resaltando su capacidad de agrupación y su facilidad de adaptación con PCA. Al aplicar el análisis de sus componentes principales mejora su capacidad de agrupación y la identificación de características.

Por último, el algoritmo DBSCAN no logró formar clusters por características similares. Esto se debe a que al momento de agrupación todos los cantones fueron representados con -1, lo que significa que existe ruido dentro de esas variables.

4.3. ESTRATEGIAS DE GESTIÓN DE RESIDUOS NO PELIGROSOS ENCONTRADAS EN LOS DIFERENTES *CLUSTERS* DE CANTONES

Para el análisis de identificación de estrategias predominantes en el manejo de residuos sólidos no peligrosos en los cantones de la provincia de Los Ríos se utilizaron los resultados obtenidos con los métodos de *clustering*: K-means y Clustering Jerárquico para agrupar los cantones en clusters y determinar las estrategias predominantes para cada variable en cada cluster.

4.3.1. Estrategias o prácticas identificadas con K-means

En la **Tabla II** se muestran las estrategias predominantes de cada variable, de acuerdo con las agrupaciones identificadas por el método de agrupación K-means.

Tabla 11*Estrategias de gestión de residuos identificadas por K-means*

Variable	Cluster 0	Cluster 1	Cluster 2	Cluster 3
ResiduoHora	Si	Si	Si	Si
ClasificóResiduos	Si	No	No	Si
Compras Frecuentes	Bolsa de plástico (desechables)	Bolsa de plástico (desechables)	Bolsa de plástico (desechables)	Bolsa de plástico (desechables)
DenunciaProblemaAmbienta	No	No	No	No
EliminóAceite	Depositó con el resto de los residuos	Depositó con el resto de los residuos	Depositó con el resto de los residuos	Depositó con el resto de los residuos
EliminóFocos	Depositó con el resto de los residuos	Depositó con el resto de los residuos	Depositó con el resto de los residuos	Depositó con el resto de los residuos
EliminóResiduosOrgánicos	Composta abono/alimento de animales	Depositó con el resto de los residuos	Depositó con el resto de los residuos	Regaló (escuela, vecina), vendió
EliminóResiduosPapelCartón	Depositó con el resto de los residuos	Depositó con el resto de los residuos	Depositó con el resto de los residuos	Regaló (escuela, vecina), vendió
EliminóResiduosPlástico	Regaló (escuela, vecina), vendió	Depositó con el resto de los residuos	Depositó con el resto de los residuos	Regaló (escuela, vecina), vendió
EliminóResiduosVidrio	Depositó con el resto de los residuos	Depositó con el resto de los residuos	Depositó con el resto de los residuos	Regaló (escuela, vecina), vendió
ProblemasAcumulaciónResiduos	No	No	No	No
ResiduosOrganicos	No	No	No	Si
ResiduosPapelCartón	No	No	No	Si

ResiduosPlásticos	Si	No	No	Si
ResiduosVidrio	No	No	No	Si
UsaFocos	Si	Si	Si	Si
UsaPilas	Si	Si	Si	Si
VoluntariadoAmbien tal	No	No	No	No

Nota: Elaborado por la autora.

Cluster 0: Pueblviejo

El clúster 0 se destaca por sus prácticas positivas en el manejo de residuos. Los hogares de este cantón realizan estrategias como la clasificación de residuos y la eliminación de residuos orgánicos, papel/cartón y plástico mediante el compostaje y reciclaje. Sin embargo, existen áreas de mejora como la participación en voluntariado ambiental y denuncia de problemas ambientales.

Cluster 1: Quevedo, Buena Fe, Valencia, Montalvo, Babahoyo, Ventanas y Urdaneta

En el cluster 1, se observa una baja clasificación de residuos y participación en programas de voluntariado ambiental. La eliminación de los residuos se realiza principalmente depositándolos junto con el resto de los residuos doméstica, lo que sugiere oportunidades significativas de mejora en prácticas de reciclaje y compostaje.

Cluster 2: Mocache, Baba y Palenque

Los cantones del cluster 2 muestran un nivel bajo de clasificación de residuos y una eliminación inadecuada de residuos plásticos y papel/cartón siendo su mayor práctica es depositarlos con la mayoría de los residuos del hogar. Además, presenta un bajo nivel de participación en prácticas sostenibles y voluntariado ambiental.

Cluster 3: Quinsaloma y Vinces

El cluster 3 presenta prácticas de clasificación de residuos y la eliminación de residuos como el plástico, vidrio, papel/cartón; lo realizan mediante reciclaje o donación. Sin embargo, áreas como la participación en programas de voluntariado ambiental y la denuncia de problemas ambientales requieren mejoras.

4.3.2. Estrategias o prácticas identificadas por Clustering Jerárquico

En la **Tabla 12** se muestran las estrategias predominantes de cada variable, de acuerdo con las agrupaciones identificadas por el método de *Clustering Jerárquico*.

Tabla 12

Estrategias identificadas por clustering jerárquico

Variable	Cluster 1	Cluster 2	Cluster 3	Cluster 4
ResiduoHora	Si	Si	Si	Si
ClasificóResiduos	No	No	No	Si
ComprasFrecuentes	Bolsa de plástico (desechables)	Bolsa de plástico (desechables)	Bolsa de plástico (desechables)	Bolsa de plástico (desechables)
DenunciaProblema Ambiental	No	Si	No	No
EliminóAceite	Depositó con el resto de los residuos	Depositó con el resto de los residuos	Depositó con el resto de los residuos	Depositó con el resto de los residuos
EliminóFocos	Depositó con el resto de los residuos	Depositó con el resto de los residuos	Depositó con el resto de los residuos	Depositó con el resto de los residuos
EliminóResiduosOrgánicos	Depositó con el resto de los residuos	Depositó con el resto de los residuos	Composta abono/alimento de animales	Composta abono/alimento de animales
EliminóResiduosPapelCartón	Depositó con el resto de los residuos	Depositó con el resto de los residuos	Regaló (escuela, vecina), vendió	Depositó con el resto de los residuos

EliminóResiduosPlástico	Deposité con el resto de los residuos	Deposité con el resto de los residuos	Regaló (escuela, vecina), vendió	Regaló (escuela, vecina), vendió
EliminóResiduosVidrio	Deposité con el resto de los residuos	Deposité con el resto de los residuos	Regaló (escuela, vecina), vendió	Deposité con el resto de los residuos
ProblemasAcumulaciónResiduos	No	No	No	No
ResiduosOrganicos	No	No	Si	No
ResiduosPapelCartón	No	No	Si	No
ResiduosPlásticos	No	No	Si	Si
ResiduosVidrio	No	No	Si	No
UsaFocos	Si	Si	Si	Si
UsaPilas	Si	Si	Si	Si
VoluntariadoAmbienta	No	Si	No	No

Nota: Elaborado por la autora.

Cluster 1: Quevedo, Buena Fe, Valencia, Montalvo y Babahoyo

El cluster 1 presenta una baja participación en estrategias como la clasificación de residuos y programas de voluntariado ambiental. La eliminación de sus residuos orgánicos y plásticos son realizados de manera inadecuada, ya que son depositados junto a los residuos que genera el hogar.

Cluster 2: Pueblviejo y Palenque

En el cluster 2 se observa una mejor participación en la denuncia de problemas ambientales. Sin embargo, el manejo de residuos aún requiere mejoras en relación con la clasificación y eliminación adecuada de los residuos.

Cluster 3: Ventanas, Mocache y Baba

El cluster 3 se destaca por la clasificación de residuos y prácticas eficientes en la eliminación de residuos orgánicos y plásticos mediante compostaje y reciclaje. Además, muestra un mayor manejo en la eliminación de residuos comparado con los otros clusters.

Cluster 4: Urdaneta, Quinsaloma y Vines

Los cantones del cluster 4 presentan una combinación de prácticas eficientes en la eliminación de residuos plásticos y papel/cartón mediante reciclaje y donación. Sin embargo, necesita mejoras en la participación en prácticas sostenibles y voluntariado ambiental.

A continuación, se muestran las similitudes y las diferencias que existen al aplicar los métodos de agrupación K-means y *Clustering* Jerárquico. Aunque ambos métodos lograron identificar cuatro grupos diferentes, se observó que las características de las agrupaciones varían entre los dos algoritmos.

Similitudes

- Los dos métodos muestran que existe la necesidad de mejorar la clasificación de residuos y la participación en prácticas sostenibles como el voluntariado ambiental.
- Ambos métodos identifican prácticas como el compostaje y reciclaje entre algunos cantones. Estas estrategias de manejo de residuos son utilizadas principalmente para la eliminación de plásticos y papel/cartón.

Diferencias

- El algoritmo de K-mean ayuda a identificar de una mejor manera los cantones que presentan una mejor aplicación de prácticas específicas como la eliminación de residuos orgánicos y plásticos mediante compostaje y reciclaje.

- *Clustering* Jerárquico destaca las áreas de mejora en términos de estrategias ambientales como la denuncia de problemas y el voluntariado ambientales.

4.3.3. Discusión

El análisis de las estrategias de gestión de residuos en varios grupos reveló patrones y áreas de mejora evidentes. En general, se encontró que la mayoría de los cantones no utilizan la clasificación de residuos y que la eliminación se realiza principalmente con el resto de los residuos del hogar.

Estos resultados destacan la importancia de adoptar estrategias de manejo de residuos que se adapten a las características particulares de cada grupo. Todos los grupos reciben recomendaciones importantes, incluida la implementación de programas de reciclaje, y la promoción de prácticas sostenibles.

Cheng y Hu (2010) señalan que las estrategias de gestión de residuos deben adaptarse a las particularidades de cada área para aumentar la eficiencia y la sostenibilidad. Es fundamental implementar un enfoque individualizado para lograr una gestión de residuos más eficiente y sostenible en la Provincia de Los Ríos. Esto permite adaptar las estrategias de gestión a las particularidades de cada cantón, lo que maximiza la eficacia de las acciones que se implementen. La identificación de características específicas en la generación y el manejo de residuos se facilita mediante algoritmos de agrupamiento como K-Means, lo que permite una segmentación precisa para cada grupo de cantones.

La clasificación de residuos y la participación en prácticas sostenibles son áreas críticas que necesitan mejoras en la mayoría de los clusters. Existen clusters que destacan por sus prácticas eficientes en la eliminación de residuos, pero aún tienen áreas significativas de mejora. Basándose en estudios previos, como el de Ferronato y Torretta (2019), también

se identifica que la falta de clasificación de residuos es un problema común en países en desarrollo, lo que corrobora los resultados aquí obtenidos. Además, Singh y Ordoñez (2016) destacan la importancia de la recuperación de recursos y el reciclaje en la economía circular, lo cual es directamente aplicable a las observaciones descritas sobre la clasificación de residuos.

CAPÍTULO IV

CONCLUSIONES Y RECOMENDACIONES

5.1. CONCLUSIONES

Con el análisis de Chi cuadrado y ANOVA se identificaron 16 variables que influyen en la generación de los residuos sólidos no peligrosos en los cantones de la provincia de Los Ríos, garantizando una mayor estabilidad y precisión en el análisis de *clustering*. Se demostró que la imputación de datos por la mediana es un método eficaz para la imputación de datos numéricos debido a su capacidad de preservar la estructura original de los datos y minimizar sesgos.

La aplicación de técnicas de *clustering*, como K-Means y *Clustering* Jerárquico, permitieron agrupar los cantones con características similares en términos de generación y manejo de residuos sólidos no peligrosos. Se lograron identificar cuatro clusters diferentes luego de la aplicación de estos métodos de agrupación, cada uno con características y prácticas de manejo específicas. La reducción de dimensionalidad mediante PCA facilitó la visualización de los clusters y la identificación de patrones en los datos. Sin embargo, el algoritmo DBSCAN no logró formar clusters bien definidos debido a la presencia de ruido en los datos.

El análisis de las estrategias de gestión de residuos sólidos no peligrosos en los diferentes clusters de cantones reveló patrones claros y áreas de mejora. La mayoría de los cantones no implementa adecuadamente la clasificación de residuos y depositarlos con todos los residuos del hogar. Sin embargo, se identificaron prácticas eficientes en algunos clusters, como el compostaje y reciclaje de residuos orgánicos y plásticos. La comparación entre K-means y *clustering* jerárquico mostró que, aunque ambos métodos tienen similitudes, el K-means es más eficaz para identificar prácticas específicas, mientras que el *clustering* jerárquico resalta áreas de mejora en estrategias ambientales.

5.2. RECOMENDACIONES

La mayoría de los cantones no clasifica adecuadamente los residuos sólidos no peligrosos, lo que subraya la necesidad de implementar programas de educación y sensibilización. Estos programas deben centrarse en la importancia de la clasificación de residuos y las técnicas de eliminación adecuadas. Involucrar a las comunidades mediante talleres, campañas de concienciación y colaboraciones con escuelas locales es esencial, porque la educación ambiental mejora significativamente las prácticas de gestión de residuos, reduciendo la generación de residuos y promoviendo mejores prácticas.

La aplicación de técnicas de *clustering*, como K-Means y *clustering* jerárquico ha demostrado ser eficaz para identificar patrones y estrategias de manejo de residuos en los cantones. Se recomienda que las autoridades locales utilicen estos métodos para diseñar planes de gestión específicos para cada clusters, asegurando que las estrategias se adapten a las características y necesidades particulares de cada grupo. Es fundamental que la recolección de datos sea lo más consistente y real posible para garantizar la precisión y validez de los análisis.

Los resultados mostraron que algunos cantones realizan prácticas como el compostaje y reciclaje de residuos orgánicos y plásticos. Se recomienda extender estas prácticas a otros cantones, promoviendo el intercambio de conocimientos y experiencias entre los clusters. Así puedan mejorar significativamente las actitudes y comportamientos hacia la gestión sostenible de residuos.

REFERENCIAS BIBLIOGRÁFICAS

- Agencia de Protección Ambiental de los Estados Unidos, & Oficina de Conservación y Recuperación de Recursos. (2020). *Mejores prácticas para la gestión de los residuos sólidos: Una Guía para los responsables de la toma de decisiones en los países en vías de desarrollo* (p. 169). https://www.epa.gov/sites/default/files/2021-02/documents/swm_guide-spanish-reducedfilesize_pubnumber_october.pdf
- Altikat, A., Gulbe, A., & Altikat, S. (2022). Intelligent solid waste classification using deep convolutional neural networks. *International Journal of Environmental Science and Technology*, 19(3), 1285-1292. <https://doi.org/10.1007/s13762-021-03179-4>
- Anderson, L. (2023, octubre 17). *Residuos sólidos: Clasificación y definición - eco360*. <https://eco360.info/medio-ambiente-y-conservacion/residuos-solidos-clasificacion-y-definicion/>
- Aragones, X. F., & Graells, J. A. (2019). *Técnicas de clustering* (1.^a ed.). Oberta UOC Publishing, SL. https://openaccess.uoc.edu/bitstream/10609/147174/10/AnaliticaDeDatos_Modulo5_TecnicasDeClustering.pdf
- Assef, F. M., Steiner, M. T. A., & Lima, E. P. de. (2022). A review of clustering techniques for waste management. *Heliyon*, 8(1), e08784. <https://doi.org/10.1016/j.heliyon.2022.e08784>
- Boutsidis, C., Zouzias, A., Mahoney, M. W., & Drineas, P. (2014). *Randomized Dimensionality Reduction for k-means Clustering* (arXiv:1110.2897). arXiv. <https://doi.org/10.48550/arXiv.1110.2897>

- Caruso, G., Gattone, S. A., Balzanella, A., & Di Battista, T. (2019). Cluster Analysis: An Application to a Real Mixed-Type Data Set. En C. Flaut, Š. Hošková-Mayerová, & D. Flaut (Eds.), *Models and Theories in Social Systems* (pp. 525-533). Springer International Publishing. https://doi.org/10.1007/978-3-030-00084-4_27
- Cheng, H., & Hu, Y. (2010). Municipal solid waste (MSW) as a renewable source of energy: Current and future practices in China. *Bioresource Technology*, *101*(11), 3816-3824. <https://doi.org/10.1016/j.biortech.2010.01.040>
- Dirección Provincial del Ambiente de Los Rios. (2019). *Rendición de cuenta 2019*. https://www.ambiente.gob.ec/wp-content/uploads/downloads/2020/04/Los-R%C3%ADos_Fase-2.pdf
- Dunkel, J., Dominguez, D., Borzdynski, Ó. G., & Sánchez, Á. (2022). Solid Waste Analysis Using Open-Access Socio-Economic Data. *Sustainability*, *14*(3), Article 3. <https://doi.org/10.3390/su14031233>
- Ferronato, N., & Torretta, V. (2019). Waste Mismanagement in Developing Countries: A Review of Global Issues. *International Journal of Environmental Research and Public Health*, *16*(6), Article 6. <https://doi.org/10.3390/ijerph16061060>
- Festa, D., Novellino, A., Hussain, E., Bateson, L., Casagli, N., Confuorto, P., Del Soldato, M., & Raspini, F. (2023). Unsupervised detection of InSAR time series patterns based on PCA and K-means clustering. *International Journal of Applied Earth Observation and Geoinformation*, *118*, 103276. <https://doi.org/10.1016/j.jag.2023.103276>
- Gondo, T. (2019). A hierarchical cluster-based segmentation analysis of potential solid waste management health hazards in urban Ethiopia. *Jàmbá : Journal of Disaster Risk Studies*, *11*(2), 716. <https://doi.org/10.4102/jamba.v11i2.716>

- Goswami, M., Maurya, P., Nautiyal, S., Banerjee, S., Gupta, A. K., & Premkumar, A. (2024). Status of household solid waste in Bengaluru and its periphery: Synergies and disjunctions between waste management practices and circular economy. *Journal of Material Cycles and Waste Management*, 26(1), 295-312. <https://doi.org/10.1007/s10163-023-01826-1>
- Han, J., Kamber, M., & Pei, J. (2011). *Data-Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf* (Tercera). ELSEVIER. <https://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf>
- Hoornweg, D., & Bhada-Tata, P. (2012). What a waste: A global review of solid waste management. *Urban Dev Ser Knowl Pap*, 15, 87-88.
- Instituto Nacional de Estadísticas y Censo. (2022). *Instituto Nacional de Estadística y Censos*. <https://www.ecuadorencifras.gob.ec/institucional/home/>
- Instituto Nacional de Estadísticas y Censo. (2023). *Hogares*. Instituto Nacional de Estadísticas y Censo. <https://www.ecuadorencifras.gob.ec/hogares/>
- Izquierdo-Horna, L., Kahhat, R., & Vázquez-Rowe, I. (2022). Reviewing the influence of sociocultural, environmental and economic variables to forecast municipal solid waste (MSW) generation. *Sustainable Production and Consumption*, 33. <https://doi.org/10.1016/j.spc.2022.08.008>
- Izquierdo-Horna, L., Zevallos, J., Damazo, M., & Yanayaco, D. (2021). Exploratory Data Analysis of Community Behavior Towards the Generation of Solid Waste Using

- K-Means and Social Indicators. *International Journal of Sustainable Development and Planning*, 16, 875-881. <https://doi.org/10.18280/ijstdp.160508>
- Jäger, S., Allhorn, A., & Bießmann, F. (2021). A Benchmark for Data Imputation Methods. *Frontiers in Big Data*, 4. <https://doi.org/10.3389/fdata.2021.693674>
- Kawai, K., & Tasaki, T. (2016). Revisiting estimates of municipal solid waste generation per capita and their reliability. *Journal of Material Cycles and Waste Management*, 18(1), 1-13. <https://doi.org/10.1007/s10163-015-0355-1>
- Kaza, S., Yao, L., Bhada-Tata, P., Van Woerden, F., Ionkova, K., Morton, J., Poveda, R. A., Sarraf, M., Malkawi, F., Harinath, A. S., Banna, F., An, G., Imoto, H., & Levine, D. (2018). *What a waste 2.0: A global snapshot of solid waste management to 2050*. World Bank Group. <https://doi.org/10.1596/9781464813290>
- Khan, S. I., & Hoque, A. S. M. L. (2020). SICE: An improved missing data imputation technique. *Journal of Big Data*, 7(1), 37. <https://doi.org/10.1186/s40537-020-00313-w>
- Li, J., Gao, K., & Xu, Z. (2017). Charge-decay electrostatic separation for removing Polyvinyl chloride from mixed plastic wastes. *Journal of Cleaner Production*, 157, 148-154. <https://doi.org/10.1016/j.jclepro.2017.03.233>
- Li, J., Guo, S., Ma, R., He, J., Zhang, X., Rui, D., Ding, Y., Li, Y., Jian, L., Cheng, J., & Guo, H. (2024). Comparison of the effects of imputation methods for missing data in predictive modelling of cohort study datasets. *BMC Medical Research Methodology*, 24(1), 41. <https://doi.org/10.1186/s12874-024-02173-x>
- Libório, M. P., da Silva Martinuci, O., Machado, A. M. C., Machado-Coelho, T. M., Laudaes, S., & Bernardes, P. (2022). Principal component analysis applied to

- multidimensional social indicators longitudinal studies: Limitations and possibilities. *GeoJournal*, 87(3), 1453-1468. <https://doi.org/10.1007/s10708-020-10322-0>
- Luciani, V., Bonifazi, G., Rem, P., & Serranti, S. (2015). Upgrading of PVC rich wastes by magnetic density separation and hyperspectral imaging quality control. *Waste Management*, 45, 118-125. <https://doi.org/10.1016/j.wasman.2014.10.015>
- Maheswari, K., Packia Amutha Priya, P., Ramkumar, S., & Arun, M. (2020). Missing Data Handling by Mean Imputation Method and Statistical Analysis of Classification Algorithm. En A. Haldorai, A. Ramu, S. Mohanram, & C. C. Onn (Eds.), *EAI International Conference on Big Data Innovation for Sustainable Cognitive Computing* (pp. 137-149). Springer International Publishing. https://doi.org/10.1007/978-3-030-19562-5_14
- Ministerio del Ambiente. (s. f.). *Ministerio del Ambiente, Agua y Transición Ecológica*. Recuperado 23 de febrero de 2024, de <https://www.ambiente.gob.ec/perfil-ministro/>
- Ministerio del Ambiente y Agua del Ecuador. (2020). *Manual de aprovechamiento de residuos orgánicos municipales* (1.^a ed.). <https://www.ambiente.gob.ec/wp-content/uploads/downloads/2020/07/MANUAL-DE-APROVECHAMIENTO-DE-RESIDUOS-ORGANICOS-MUNICIPAL.pdf>
- Onumanyi, A. J., Molokomme, D. N., Isaac, S. J., & Abu-Mahfouz, A. M. (2022). AutoElbow: An Automatic Elbow Detection Method for Estimating the Number of Clusters in a Dataset. *Applied Sciences*, 12(15), Article 15. <https://doi.org/10.3390/app12157515>

- Organización de las Naciones Unidas. (2022). *Ciudades y comunidades sostenibles*. Desarrollo Sostenible. <https://www.un.org/sustainabledevelopment/es/cities/>
- Quinteros—2019—DE LOS RESIDUOS SÓLIDOS URBANOS EN EL CANTÓN DE QU.pdf. (s. f.). Recuperado 21 de mayo de 2024, de <http://scielo.sld.cu/pdf/rus/v11n5/2218-3620-rus-11-05-362.pdf>
- Riquelme, J. C., Ruiz, R., & Gilbert, K. (2006). *Minería de Datos: Conceptos y Tendencias*. 10(29).
- Rodriguez, M. Z., Comin, C. H., Casanova, D., Bruno, O. M., Amancio, D. R., Costa, L. da F., & Rodrigues, F. A. (2019). Clustering algorithms: A comparative approach. *PLOS ONE*, 14(1), e0210236. <https://doi.org/10.1371/journal.pone.0210236>
- Rondón, E., Szantó, M., Francisco, J., Contreras, E., & Gálvez, A. (2016). Guía general para la gestión de residuos sólidos domiciliarios. *Naciones Unidas*, 211.
- Sáez, A., Urdaneta, G., & Joheni, A. (2014). Manejo de residuos sólidos en América Latina y el Caribe. *Omnia*, 20(3), 16.
- Secretaría Nacional de Planificación. (2021). *Plan de Creación de Oportunidades 2021 2025 Aprobado.pdf*. 122.
- Singh, J., & Ordoñez, I. (2016). Resource recovery from post-consumer waste: Important lessons for the upcoming circular economy. *Journal of Cleaner Production*, 134, 342-353. <https://doi.org/10.1016/j.jclepro.2015.12.020>
- Templ, M. (2023). Enhancing Precision in Large-Scale Data Analysis: An Innovative Robust Imputation Algorithm for Managing Outliers and Missing Values. *Mathematics*, 11(12), Article 12. <https://doi.org/10.3390/math11122729>

- Warintarawej, P., & Nillaor, P. (2023). Implementing data analysis based on the mixed clustering technique for sustainable participatory waste management in a low-budget area. *foresight*, 25(1), 108-125. <https://doi.org/10.1108/FS-09-2021-0179>
- Watson, K. B. (2014). Categorical Data Analysis. En A. C. Michalos (Ed.), *Encyclopedia of Quality of Life and Well-Being Research* (pp. 601-604). Springer Netherlands. https://doi.org/10.1007/978-94-007-0753-5_291
- Whatley, M. (2022). One-Way ANOVA and the Chi-Square Test of Independence. En M. Whatley (Ed.), *Introduction to Quantitative Analysis for International Educators* (pp. 57-74). Springer International Publishing. https://doi.org/10.1007/978-3-030-93831-4_5
- Yang, T., Xu, J., Zhao, Y., Gong, T., Zhao, R., Sun, M., & Xi, B. (2021). Classification technology of domestic waste from 2000 to 2019: A bibliometrics-based review. *Environmental Science and Pollution Research*, 28(21), 26313-26324. <https://doi.org/10.1007/s11356-021-12816-x>

ANEXOS.

Anexo 1. Certificado antiplagio



Yo, Ing. Orlando Erazo Moreta, PhD., en calidad de director del Proyecto de Investigación titulado **“Análisis del comportamiento de los residuos sólidos no peligrosos en los cantones de la provincia de Los Ríos: Un enfoque de Ciencia de Datos”**, me permito manifestar a usted y por medio del Consejo Directivo de Facultad lo siguiente:

Que la estudiante María Mercedes Alarcón Bermúdez, egresada en la Facultad de Posgrado, ha cumplido con las correcciones pertinentes e ingresado su Proyecto de Investigación al sistema COMPILATIO, por lo que tengo a bien certificar la siguiente información sobre el informe del sistema antiplagio con un porcentaje del 3 %.



Ing. Orlando Erazo Moreta, PhD.
DIRECTOR DEL PROYECTO DE INVESTIGACIÓN