

CIENCIA
UNEMI

**Ciencia, tecnología y
medio ambiente**

Ciencia de datos aplicada al manejo de residuos sólidos en Los Ríos, Ecuador

María Mercedes Alarcón-Bermúdez¹; Orlando Erazo²;
Génesis Acosta Gonzalez³; María Fernanda Alarcón⁴

Resumen

La gestión de residuos sólidos es un desafío para las ciudades y los países por la urbanización y por el aumento de su población. Esto lleva a necesidad de buscar opciones que faciliten su análisis, como la Ciencia de Datos. Por ello, este estudio aplicó técnicas como el *Clustering* para analizar la generación y manejo de residuos sólidos no peligrosos, tomando como caso los cantones de la provincia de Los Ríos en Ecuador. Se identificaron 16 variables influyentes mediante análisis de Chi-Cuadrado y ANOVA, y se aplicaron algoritmos de *Clustering* (*K-Means*, *DBSCAN* y *Clustering Jerárquico*) para agrupar cantones con características similares. Los resultados muestran que la mayoría de los cantones no clasifican adecuadamente los residuos y los eliminan junto con los residuos domésticos. Sin embargo, algunos cantones implementan prácticas eficientes como el compostaje y el reciclaje. Así, se recomienda la implementación de programas educativos y estrategias específicas para cada grupo de cantones, con base en los patrones identificados.

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

Data science applied to solid waste management in Los Ríos-Ecuador

Abstract

Solid waste management is a challenge for cities and countries due to urbanization and population growth. This fact creates the need to explore options that facilitate analysis, such as Data Science. Therefore, this study applied techniques like Clustering to analyze the generation and management of non-hazardous solid waste, using the cantons of the Los Ríos province in Ecuador as a case study. Sixteen influential variables were identified through Chi-Square and ANOVA analysis, and Clustering algorithms (*K-Means*, *DBSCAN*, and *Hierarchical Clustering*) were applied to group cantons with similar characteristics. The results show that most cantons do not properly classify waste and dispose of it along with household waste. However, some cantons implement efficient practices such as composting and recycling. Therefore, the implementation of educational programs and specific strategies for each group of cantons is recommended, based on the identified patterns.

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

Recibido: 24 de febrero de 2025

Aceptado: 19 de agosto de 2025

¹ Universidad Técnica Estatal de Quevedo, malarconb@uteq.edu.ec, <https://orcid.org/0009-0004-7592-6198>

² Universidad Técnica Estatal de Quevedo, oerazo@uteq.edu.ec, <https://orcid.org/0000-0001-5642-9920>

³ Universidad Técnica Estatal de Quevedo, gacostag2@uteq.edu.ec, <https://orcid.org/0000-0003-2492-6142>

⁴ Empresa D-LASER, maria.alarcon2015@uteq.edu.ec, <https://orcid.org/0009-0004-1144-1124>

I. INTRODUCCIÓN

En la actualidad, la gestión adecuada de los residuos sólidos representa uno de los desafíos más críticos a nivel global, con implicaciones significativas para el medio ambiente, la salud pública y la sostenibilidad urbana (Organización de las Naciones Unidas, 2018). El crecimiento acelerado de la población, junto con el desarrollo económico, ha incrementado sustancialmente la generación de residuos, ejerciendo una presión sin precedentes sobre los ecosistemas naturales (Sáez et al., 2014).

En Ecuador, esta situación no es ajena, ya que la gestión de residuos sólidos se ha convertido en un desafío ambiental prioritario, especialmente en regiones con alta densidad poblacional y crecimiento urbano acelerado (Ministerio del Ambiente, Agua y Transición Ecológica, 2020). Este es el caso también de la provincia de Los Ríos, donde el aumento de la población y la expansión de las áreas urbanas han agravado la problemática de la gestión de residuos sólidos. A pesar de los esfuerzos institucionales, persisten prácticas inadecuadas como la quema de residuos, la disposición en vertederos improvisados y el uso limitado de sistemas de recolección formal (Ministerio del Ambiente, 2019).

Según datos del Censo de Población y Vivienda (Instituto Nacional de Estadísticas y Censo, 2022), estas prácticas aún se presentan en la provincia, lo que refleja la urgencia de implementar estrategias más eficientes y sostenibles. Además, estudios globales indican que, en 2016, se generaron aproximadamente 2010 millones de toneladas de residuos sólidos municipales, cifra que podría alcanzar los 3400 millones de toneladas para 2050 si no se adoptan medidas correctivas (Kaza et al., 2018).

En este contexto, el presente trabajo se centra en analizar el comportamiento de los residuos sólidos no peligrosos en la provincia de Los Ríos, utilizando técnicas de Ciencia de Datos. El objetivo principal es explorar patrones de generación y manejo de residuos, con el fin de proponer soluciones innovadoras que contribuyan a mejorar la gestión local y avanzar hacia un modelo más sostenible. Cabe notar que este estudio se enmarca en los Objetivos de Desarrollo Sostenible

(ODS), específicamente en el ODS 11: Ciudades y Comunidades Sostenibles, que promueve la gestión eficiente de residuos como un pilar fundamental para la sostenibilidad urbana (Organización de las Naciones Unidas, 2022).

TRABAJO RELACIONADO

El uso de técnicas de ciencia de datos en la gestión de residuos sólidos ha sido ampliamente explorado en la literatura, demostrando su efectividad para identificar patrones y mejorar estrategias de manejo. Gondo (2019) empleó el Análisis de Clúster Jerárquico (HCA) para evaluar los riesgos ambientales y de salud asociados con prácticas deficientes de gestión de residuos en 26 ciudades y pueblos de Etiopía. Su estudio reveló que factores como la infraestructura, la planificación territorial y las condiciones socioeconómicas influyen significativamente en estos riesgos, proporcionando un marco analítico útil para abordar la gestión de residuos desde una perspectiva integral.

Por otro lado, Izquierdo-Horna *et al.* (2021) aplicaron el algoritmo K-Means para analizar la generación de residuos sólidos en la comunidad del distrito de Puente Piedra, en Perú. A través de un enfoque basado en el análisis sociodemográfico y sociocultural, segmentaron la población y caracterizaron su comportamiento en relación con la generación y acumulación de residuos. De manera similar, Dunkel *et al.* (2022) utilizaron datos socioeconómicos abiertos combinados con modelos de Clustering para clasificar países según sus características de generación de residuos, identificando los factores que más influyen en la producción de desechos a nivel nacional. Estos enfoques resaltan la utilidad del aprendizaje automático para la categorización de datos y el desarrollo de estrategias de gestión adaptadas a distintos contextos.

Además, estudios como el de Warintarawej y Nillaor (2023) han demostrado la eficacia del análisis de datos en la optimización de estrategias sostenibles. En su investigación, implementaron técnicas de *Clustering* mixto con el coeficiente de Gower para evaluar la gestión de residuos en una comunidad de bajos recursos en Tailandia, relacionando variables socioeconómicas con

prácticas de reutilización, reducción y reciclaje (3R). Su propuesta derivó en el desarrollo de guías prácticas para la gestión municipal de residuos, evidenciando el impacto positivo de la participación comunitaria en la sostenibilidad de estos procesos. En la misma línea, Izquierdo-Horna *et al.* (2022) destacaron la importancia de incluir variables socioculturales, ambientales y económicas en los modelos de análisis de residuos, aunque señalaron la limitación en la disponibilidad de estos datos, lo que representa un reto metodológico para la construcción de modelos predictivos más precisos.

Finalmente, la revisión sistemática realizada por Assef *et al.* (2022) identificó diversas aplicaciones de técnicas de Clustering en la gestión de residuos, resaltando el predominio del algoritmo K-Means y su eficacia en el diseño de estrategias de recolección y clasificación de desechos. No obstante, los autores subrayan la necesidad de incorporar restricciones más realistas en los modelos para mejorar su aplicabilidad en contextos urbanos complejos.

A partir de estos antecedentes, el presente estudio se enfoca en la aplicación de técnicas de Clustering para analizar la generación de residuos sólidos no peligrosos en la provincia de Los Ríos, Ecuador. Este enfoque permitirá identificar patrones clave y factores determinantes en la gestión de residuos, proporcionando información útil para la formulación de estrategias más efectivas y sostenibles en la región.

II. MATERIALES Y MÉTODOS

El estudio se basó en un enfoque analítico para la revisión y procesamiento de una base de datos disponible, siguiendo un procedimiento estructurado en varias etapas: (1) exploración inicial de los datos, (2) limpieza y preparación de los datos, (3) análisis exploratorio y (4) aplicación de técnicas de Ciencia de Datos. Como técnica principal, se empleó el *Clustering*, un método de aprendizaje no supervisado que permite agrupar observaciones con características similares. Para la implementación de estas técnicas, se seleccionó *Python* como lenguaje de programación debido a su flexibilidad y capacidad para manejar grandes volúmenes de datos complejos. Además, se utilizaron bibliotecas especializadas como *Pandas*

para la manipulación de datos y *Scikit-Learn* para la aplicación de algoritmos de aprendizaje automático.

La base de datos utilizada corresponde al período 2015-2019 y fue obtenida de los datos abiertos proporcionados por el Instituto Nacional de Estadísticas y Censos (INEC) de Ecuador (Instituto Nacional de Estadística y Censos, 2024). Dado que el estudio se basó en datos ya existentes, no requirió la realización de encuestas o entrevistas. Se obtuvo un conjunto de datos limitado únicamente a los trece cantones de la provincia de Los Ríos. Esta base de datos inicial estuvo compuesta por 5570 observaciones, correspondientes a información sobre el manejo de residuos sólidos no peligrosos en la provincia, y contenía 33 variables (27 cualitativas y 6 cuantitativas).

En la etapa de limpieza, se identificaron y eliminaron cinco variables que no eran relevantes para el análisis (Tabla 1). Además, se realizaron imputaciones para variables con valores faltantes utilizando medidas de tendencia central: la mediana para las variables "TotalPilas", "TotalFocos" y "FocosAhorradores", y la media para el coeficiente de silueta.

Tabla 1. 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

Para la selección de las variables más influyentes, se evaluó la independencia entre variables categóricas y numéricas. En el caso de las variables categóricas, se aplicó la prueba de Chi-Cuadrado, mientras que, para las variables numéricas en relación con las categóricas, se utilizó el análisis de varianza (ANOVA) (Whatley, 2022). Este último permitió determinar si las medias de las variables numéricas diferían significativamente entre las categorías.

En la etapa de *Clustering*, se prepararon los datos para aplicar técnicas de aprendizaje no

supervisado. Para ello, se realizó una codificación *One-Hot* en las variables categóricas con más de dos categorías, lo que mejoró el rendimiento computacional. Posteriormente, se aplicó el Análisis de Componentes Principales (PCA) para reducir la dimensionalidad de los datos y resaltar las variables más relevantes en la generación y manejo de residuos sólidos no peligrosos (Libório et al., 2022). Esta reducción facilitó la aplicación de tres algoritmos de *Clustering*: K-Means, DBSCAN y *Clustering Jerárquico*, permitiendo agrupar los

cantones de manera eficiente y precisa según sus características similares.

III. RESULTADOS

En la Figura 1 se presentan los valores de Chi-Cuadrado para las combinaciones de variables categóricas que mostraron un alto grado de independencia. Se observa que los valores de Chi-Cuadrado son inferiores a 15, lo que indica una baja dependencia entre las variables analizadas.

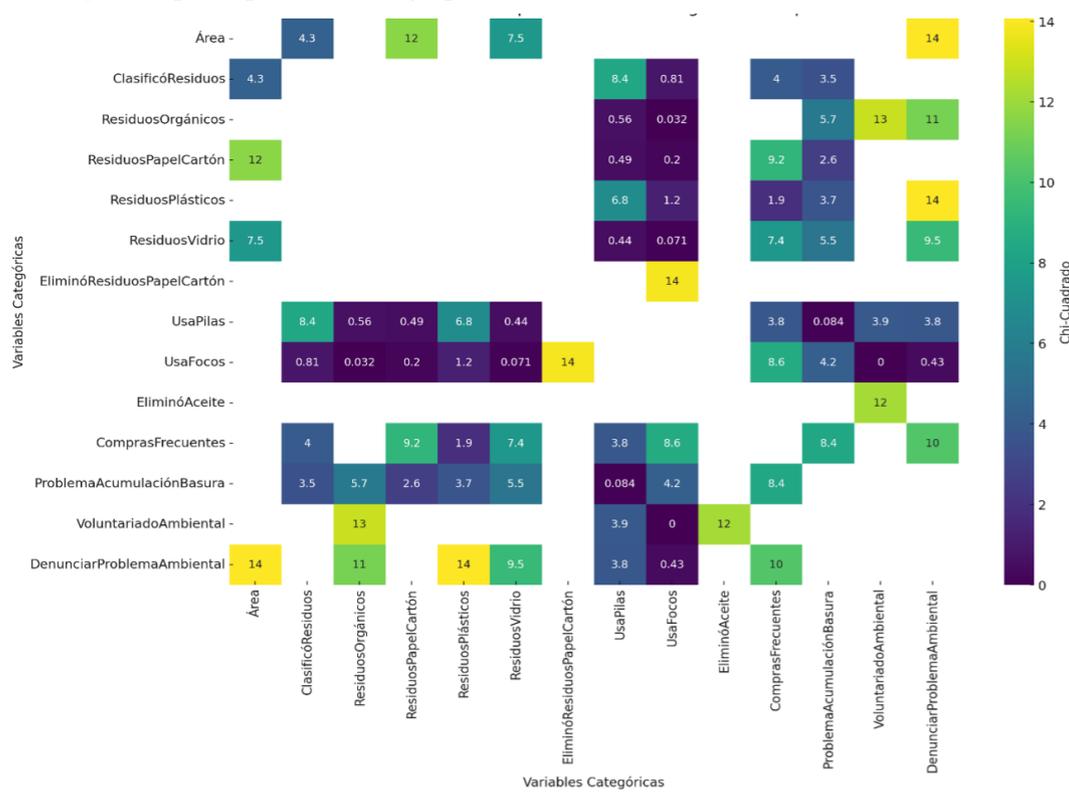


Figura 1. Valores de Chi-Cuadrado para variables categóricas.

Por otro lado, en la Tabla 2 se muestran las combinaciones de variables numéricas y categóricas que presentaron un valor p mayor a 0,05, lo que sugiere que no existen diferencias

significativas entre las medias de las variables numéricas en las distintas categorías. Esto confirma la independencia entre estas variables.

Tabla 2. Comparación de variables numéricas y categóricas

Variable numérica	Variable categórica	Valor p
Total Pilas	Residuos Orgánicos	0,201421
	Residuos Papel Cartón	0,059691
	Residuos Vidrio	0,836748
	Compras Frecuentes	0,160619

Pilas Recargables	Residuos Vidrio	0,836748
	Voluntariado Ambiental	0,092553
	Compras Frecuentes	0,178841
Total Focos	Clasificó Residuos	0,499194
	Residuos Orgánicos	0,567124
	Compras Frecuentes	0,623096
	Problema Acumulación Residuos	0,087894
Focos Ahorradores	Clasificó Residuos	0,499194
	Residuos Orgánicos	0,567127
	Usa Focos	0,474980
	Compras Frecuentes	0,215426
Focos LED	Clasificó Residuos	0,561850
	Residuos Vidrio	0,180486
	Usa Focos	0,158551
	Compras Frecuentes	0,437515

Los resultados del análisis de varianza (ANOVA) indican que las combinaciones de variables listadas en la Tabla 2 no presentan diferencias significativas en las medias de las variables numéricas entre las categorías correspondientes. Por ejemplo, el valor p de 0,201421 para las variables "TotalPilas" y "ResiduosOrgánicos" sugiere que la cantidad total de pilas no varía significativamente según la

clasificación de residuos orgánicos, lo que indica que estas variables son independientes.

Tras aplicar las pruebas de Chi-Cuadrado y ANOVA, se identificaron las variables más relevantes para el análisis de *Clustering*. En la Tabla 3 se presentan las 16 variables seleccionadas, que incluyen tanto variables numéricas como categóricas.

Tabla 3. Variables útiles para el análisis de clusters

VARIABLES NUMÉRICAS	VARIABLES CATEGÓRICAS
Total Pilas	Clasificó Residuos
Pilas Recargables	Residuos Orgánicos
Total Focos	Residuos Papel Cartón
Focos Ahorradores	Residuos Vidrio
Focos LED	Eliminó Residuos Papel Cartón
	Usa Pilas
	Usa Focos
	Compras Frecuentes
	Problema Acumulación Residuos
	Voluntariado Ambiental
	Denunciar Problema Ambiental

Para aplicar los algoritmos de *Clustering*, se realizó una codificación *One-Hot* en las variables categóricas, transformándolas en columnas binarias (1 y 0) que representan la presencia o ausencia de cada categoría.

Aplicación de algoritmos de *Clustering*

En *K-Means*, para determinar el número óptimo de *clusters*, se aplicó el método del codo, dando como resultado cuatro *clusters*. Las principales características de cada uno se presentan en la Tabla 4, habiendo sido identificadas mediante los valores medios de las variables más relevantes.

Tabla 4. Clusters y características de agrupación por K-Means

Clusters	Ciudades	Características	Valor medio
Cluster 0	Puebloviejo	Eliminó residuos plásticos	2,561
		Eliminó residuos papel/cartón	2,512
		Eliminó aceite	4,366
		Eliminó focos	2,134
Cluster 1	Quevedo, Buena Fe, Valencia, Montalvo, Babahoyo, Ventanas y Urdaneta	Residuos hora	1,765
		Clasificó residuos	0,234
		Voluntariado ambiental	0,049
		Denunciar problemas ambientales	0,039
Cluster 2	Mocache, Baba y Palenque	Eliminó residuos orgánicos	2,156
		Residuos orgánicos	0,278
		Compras frecuentes	0,178
Cluster 3	Quinsaloma y Vinces	Eliminó residuos vidrio	2,018
		Focos LED	0,174
		Eliminó focos	1,807
		Problemas de acumulación de residuos	0,165

Para DBSCAN, se utilizó el método de la rodilla para determinar el parámetro "eps" óptimo, que resultó ser 0,5. La visualización de

los clusters obtenidos se presenta en la Figura 2. Se identificaron algunos cantones como ruido (marcados con -1), lo que indica que no pertenecen a clusters bien definidos.

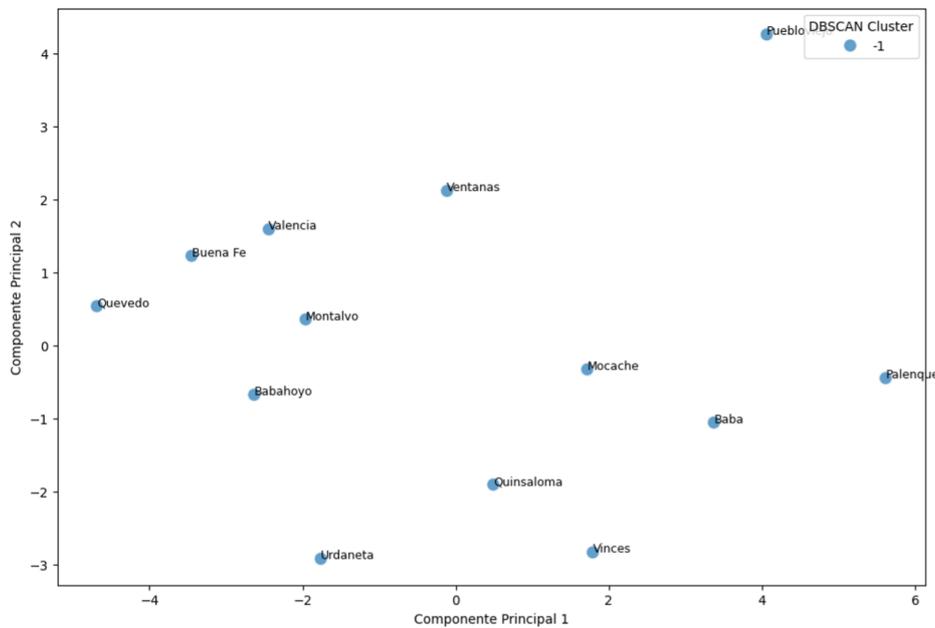


Figura 2. Visualización de clusters DBSCAN.

El Clustering Jerárquico también identificó cuatro clusters, basándose en un dendrograma,

siendo los cantones agrupados y sus características detalladas en la Tabla 5.

Tabla 5. Clusters y características de agrupación para Clustering Jerárquico

Clusters	Ciudades	Características	Valor medio
Cluster 1	Quevedo, Buena Fe, Valencia, Montalvo y Babahoyo	Clasificó residuos	0,239
		Residuos plásticos	0,293
		Voluntariado ambiental	0,054
		Denuncias voluntariado ambiental	0,046
Cluster 2	Puebloviejo y Palenque	Residuos orgánicos	0,342
		Eliminó residuos plásticos	2,564
		Eliminó residuos papel/ cartón	2,408
		Eliminó aceites	4,638
Cluster 3	Ventanas, Mocache y Baba	Clasificó residuos	0,261
		Eliminó aceites	3,763
		Eliminó focos	1,939
Cluster 4	Urdaneta, Quinsaloma y Vinces	Eliminó residuos vidrio	1,457
		Focos LED	0,148
		Problemas acumulación de residuos	0,208

Estrategias de gestión de residuos identificadas

En la Tabla 6 se presentan las estrategias predominantes en cada *cluster* formado por *K-Means*. Se observa que, aunque algunos

cantones aplican prácticas como el compostaje y el reciclaje, la mayoría aún depende de métodos menos sostenibles, como la disposición de residuos con el resto de la basura.

Tabla 6. Estrategias de gestión de residuos identificadas por K-means

Variable	Cluster 0	Cluster 1	Cluster 2	Cluster 3
Residuo Hora	Si	Si	Si	Si
Clasificó Residuos	Si	No	No	Si
Compras Frecuentes	Bolsa plástica (desechable)	Bolsa plástica (desechable)	Bolsa plástica (desechable)	Bolsa plástica (desechable)
Denuncia Problema Ambiental	No	No	No	No
Eliminó Aceite	Con residuos generales	Con residuos generales	Con residuos generales	Con residuos generales
Eliminó Focos	Con residuos generales	Con residuos generales	Con residuos generales	Con residuos generales
Eliminó Residuos Orgánicos	Composta/abono animal	Con residuos generales	Con residuos generales	Donó o vendió
Eliminó Residuos Papel Cartón	Con residuos generales	Con residuos generales	Con residuos generales	Donó o vendió
Eliminó Residuos Plástico	Donó o vendió	Con residuos generales	Con residuos generales	Donó o vendió
Eliminó Residuos Vidrio	Con residuos generales	Con residuos generales	Con residuos generales	Donó o vendió
Problemas Acumulación Residuos	No	No	No	No
Residuos Orgánicos	No	No	No	Si
Residuos Papel Cartón	No	No	No	Si

Residuos Plásticos	Si	No	No	Si
Residuos Vidrio	No	No	No	Si
Usa Focos	Si	Si	Si	Si
Usa Pilas	Si	Si	Si	Si
Voluntariado Ambiental	No	No	No	No

En la Tabla 7 se detallan las estrategias predominantes identificadas mediante Clustering Jerárquico. Este método resalta la necesidad de mejorar prácticas como la denuncia de problemas ambientales y el voluntariado.

Tabla 7. Estrategias identificadas por Clustering Jerárquico

Variable	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Residuo Hora	Si	Si	Si	Si
Clasificó Residuos	No	No	No	Si
Compras Frecuentes	Bolsa plástica (desechable)	Bolsa plástica (desechable)	Bolsa plástica (desechable)	Bolsa plástica (desechable)
Denuncia Problema Ambiental	No	Si	No	No
Eliminó Aceite	Con residuos generales	Con residuos generales	Con residuos generales	Con residuos generales
Eliminó Focos	Con residuos generales	Con residuos generales	Con residuos generales	Con residuos generales
Eliminó Residuos Orgánicos	Con residuos generales	Con residuos generales	Composta/ abono animal	Composta/ abono animal
Eliminó Residuos Papel Cartón	Con residuos generales	Con residuos generales	Donó o vendió	Con residuos generales
Eliminó Residuos Plástico	Con residuos generales	Con residuos generales	Donó o vendió	Donó o vendió
Eliminó Residuos Vidrio	Con residuos generales	Con residuos generales	Donó o vendió	Con residuos generales
Problemas Acumulación Residuos	No	No	No	No
Residuos Orgánicos	No	No	Si	No
Residuos Papel Cartón	No	No	Si	No
Residuos Plásticos	No	No	Si	Si
Residuos Vidrio	No	No	Si	No
Usa Focos	Si	Si	Si	Si
Usa Pilas	Si	Si	Si	Si
Voluntariado Ambiental	No	Si	No	No

IV. DISCUSIÓN

La identificación de variables clave en la generación y manejo de residuos sólidos no peligrosos en los cantones de la provincia de Los Ríos reveló que factores como la clasificación de residuos, el uso de bolsas plásticas y las técnicas de eliminación son cruciales, coincidiendo con estudios previos (Cheng y Hu, 2010; Ferronato y Torretta, 2019). Para garantizar la calidad del análisis, se aplicaron métodos de imputación como la mediana, respaldados por Li *et al.* (2024) y Templ (2023), y se evaluó la independencia entre

variables mediante pruebas de Chi-Cuadrado, siguiendo a Watson (2014) y Caruso *et al.* (2019), lo que permitió una agrupación precisa y menos sesgada de los cantones.

El uso de técnicas de clasificación, como K-Means y *Clustering* jerárquico, permitió identificar patrones y características relevantes en la producción y manejo de residuos sólidos no peligrosos en los cantones de la provincia de Los Ríos. Sin embargo, uno de los principales desafíos en este tipo de análisis es el alto costo computacional asociado.

Para abordar este problema, se aplicó la reducción de dimensionalidad mediante el Análisis de Componentes Principales (PCA). Como señalan Boutsidis et al. (2014), el PCA no solo simplifica los datos al transformar las variables originales en un número menor de componentes principales, sino que también mejora la eficiencia de los algoritmos de Clustering al eliminar el ruido y la redundancia en los datos.

En cuanto a los algoritmos de *Clustering*, K-Means demostró ser una herramienta efectiva para la agrupación de cantones según sus características similares. Festa *et al.* (2023) destacan que K-Means, combinado con PCA, mejora significativamente la capacidad de agrupación y la identificación de patrones.

En este estudio, el método del codo permitió determinar que el número óptimo de clusters era cuatro, lo que facilitó una segmentación adecuada de los cantones. Por otro lado, el algoritmo DBSCAN no logró formar clusters bien definidos, ya que todos los cantones fueron clasificados como ruido (representados con -1). Este resultado sugiere que las variables analizadas no presentaban una densidad suficiente para formar grupos claros con este método.

Ambos métodos de *Clustering* identificaron cuatro grupos y resaltaron la necesidad de mejorar la clasificación de residuos y la participación en prácticas sostenibles. Aunque existen similitudes, también hay diferencias significativas. Ambos métodos destacan la falta de participación en voluntariado ambiental y la necesidad de mejorar la clasificación de residuos. Sin embargo, K-Means identificó mejor las prácticas específicas como el compostaje y el reciclaje, mientras que *Clustering* Jerárquico resaltó áreas de mejora en denuncias y voluntariado ambiental.

Por otro lado, el análisis de las estrategias de gestión de residuos en los diferentes clusters reveló patrones claros y áreas de mejora significativas. En general, se observó, con base en los datos obtenidos, que la mayoría de los cantones no clasifican sus residuos y que la eliminación se realiza principalmente junto con el resto de los desechos domésticos. Estos hallazgos coinciden con estudios previos que identifican la falta de clasificación de residuos como un problema común

en países en desarrollo (Ferronato y Torretta, 2019). Además, Singh y Ordoñez (2016) resaltan la importancia de la recuperación de recursos y el reciclaje en el marco de la economía circular, lo cual es directamente aplicable a los resultados obtenidos en este estudio.

Para abordar estas deficiencias, se recomienda la implementación de programas de reciclaje y la promoción de prácticas sostenibles adaptadas a las características específicas de cada grupo de cantones. Como señalan Cheng y Hu (2010), las estrategias de gestión de residuos deben ser individualizadas para maximizar su eficacia y sostenibilidad. En este sentido, la identificación de características específicas en la generación y manejo de residuos, facilitada por algoritmos de *Clustering* como K-Means, permite una segmentación precisa y la implementación de acciones específicas para cada grupo.

V. CONCLUSIONES

Este trabajo, centrado en el análisis del comportamiento de los residuos sólidos no peligrosos en los cantones de la provincia de Los Ríos, desde un enfoque de Ciencia de Datos, identificó 16 variables clave que influyen en la generación de residuos sólidos no peligrosos, utilizando análisis de Chi cuadrado y ANOVA para garantizar la estabilidad y precisión en los modelos de Clustering. La aplicación de técnicas de agrupamiento, como K-Means y Clustering Jerárquico, facilitó la segmentación de los cantones en cuatro clusters distintos, cada uno con características particulares en la generación y gestión de residuos. La reducción de dimensionalidad mediante PCA mejoró la visualización de los grupos, evidenciando patrones en los datos, aunque el algoritmo DBSCAN no logró una segmentación óptima debido a la presencia de ruido en los datos. Asimismo, el análisis de estrategias de gestión reveló deficiencias en la clasificación de residuos en la mayoría de los cantones, aunque algunos clusters presentaron prácticas sostenibles como el compostaje y el reciclaje. La comparación entre K-Means y Clustering Jerárquico mostró que, si bien ambos métodos son complementarios, K-Means resulta más preciso para identificar prácticas

específicas, mientras que el Clustering Jerárquico permite resaltar oportunidades de mejora en la gestión ambiental. Estos hallazgos proporcionan información relevante para la optimización de estrategias de manejo de residuos en la región, contribuyendo a una gestión más eficiente y sostenible.

Por otro lado, este estudio proporciona una base para la implementación de estrategias de gestión de residuos más eficientes y sostenibles en la provincia de Los Ríos, Ecuador. La identificación de *clusters* con características similares y la detección de áreas para aplicar estrategias ambientales deben contribuir a la formulación de acciones específicas y adaptadas a las necesidades de cada cantón, contribuyendo así a la sostenibilidad ambiental y al cumplimiento de los Objetivos de Desarrollo Sostenible.

Es necesario involucrar a la comunidad mediante talleres, colaboración en escuelas, campañas de concientización o incluso incentivos propuestos por los municipios de cada cantón, como parte de inculcar en la sociedad la educación ambiental necesaria que permitirá mejorar las prácticas de gestión de residuos, reduciendo la generación y promoviendo mejores prácticas de residuos. También se debe considerar la colaboración entre municipalidades, debido que tienen una experiencia adquirida en la gestión de residuos, conocimiento que pueden servir a otros cantones, para implementar las mismas prácticas en mejora del medio ambiente y los ciudadanos.

VI. BIBLIOGRAFÍA

Assef, F., Steiner, M. T., y Lima, E. P. (2022). A review of Clustering techniques for waste management. *Heliyon*, 8(1). <https://doi.org/10.1016/j.heliyon.2022.e08784>

Boutsidis, C., Zouzias, A., Mahoney, M. W., y Drineas, P. (2014). Randomized Dimensionality Reduction for k-means Clustering. *IEEE Trans. Inf. Theory*, 61(2), 1-27. <https://doi.org/10.48550/arXiv.1110.2897>

Caruso, G., Gattone, S. A., Balzanella, A., y Di Battista, T. (2019). Cluster Analysis: An Application to a Real Mixed-Type Data Set. En C. Flaut, Š. Hořková-Mayerová, y D. Flaut, *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., y Hu, Y. (2010). Municipal solid waste (MSW) as a renewable source of energy: Current and future practices in China. *Bioresour. Technol.*, 101(11), 3816-3824. <https://doi.org/10.1016/j.biortech.2010.01.040>

Dunkel, J., Dominguez, D., Borzdynski, Ó. G., y Sánchez, Á. (2022). Solid Waste Analysis Using Open-Access Socio-Economic Data. *Sustainability*, 14(3), 1233. <https://doi.org/10.3390/su14031233>

Ferronato, N., y Torretta, V. (2019). Waste Mismanagement in Developing Countries: A Review of Global Issues. *International Journal of Environmental Research and Public Health*, 16(6), 1060. <https://doi.org/10.3390/ijerph16061060>

Festa, D., Novellino, A., Hussain, E., Bateson, L., Casagli, N., Confuorto, P., . . . 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. *Jàmba : Journal of Disaster Risk Studies*, 11(2), 716. <https://doi.org/10.4102/jamba.v11i2.716>

Instituto Nacional de Estadística y Censos. (2024). *Censos- Instituto Nacional de Estadística y Censos*. Instituto Nacional de Estadística y Censos: <https://www.ecuadorencifras.gob.ec/informacion-de-anos-antiores-hogares/>

Instituto Nacional de Estadísticas y Censo. (2022). *Boletín Técnico Nacional*. Censo Ecuador cuenta conmigo 2022: https://www.censoecuador.gob.ec/public/Boletin_Nacional.htm

Izquierdo-Horna, L., Kahhat, R., y Vázquez-Rowe, I. (2022). Reviewing the influence of sociocultural, en-

- vironmental and economic variables to forecast municipal solid waste (MSW) generation. *Sustainable Production and Consumption*, 33, 809-819. <https://doi.org/10.1016/j.spc.2022.08.008>
- Izquierdo-Horna, L., Zevallos, J., Damazo, M., y 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*, 6(5), 875-881. <https://doi.org/10.18280/ijstdp.160508>
- Kaza, S., Yao, L., Bhada-Tata, P., y Van Woerden, F. (2018). *What a Waste 2.0: A Global Snapshot of Solid Waste Management to 2050*. World Bank Group. <https://doi.org/https://doi.org/10.1596/978-1-4648-1329-0>
- Li, J., Guo, S., Ma, R., He, J., Zhang, X., Rui, D., . . . y 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., Machado-Coelho, T. M., Laudares, S., y 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>
- Ministerio del Ambiente. (2019). *Dirección Provincial del Ambiente de Los Rios Informe de Rendición de cuenta 2019*. https://www.ambiente.gob.ec/wp-content/uploads/downloads/2020/04/Los-R%C3%ADos_Fase-2.pdf
- Ministerio del Ambiente, Agua y Transición Ecológica. (2020). *Ecuador impulsa la gestión adecuada de residuos orgánicos en las ciudades*. Ministerio del Ambiente, Agua y Transición Ecológica: <https://www.ambiente.gob.ec/ecuador-impulsa-la-gestion-adeuada-de-residuos-organicos-en-las-ciudades/>
- Organización de las Naciones Unidas. (2018). *Cómo la basura afecta al desarrollo de América Latina*. Naciones Unidas: <https://news.un.org/es/story/2018/10/1443562>
- Organización de las Naciones Unidas. (2022). *Objetivos de desarrollo sostenible*. 11 Ciudades y comunidades sostenibles: <https://www.un.org/sustainabledevelopment/es/cities/>
- Sáez, A., Urdaneta, G., y Joheni, A. (2014). Manejo de residuos sólidos en América Latina y el Caribe. *Omnia*, 20(3), 121-135. <https://www.redalyc.org/articulo.oa?id=73737091009>
- Singh, J., y 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), 2729. <https://doi.org/10.3390/math11122729>
- Warintarawej, P., y 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. Michalos, *Encyclopedia of Quality of Life and Well-Being Research* (pp. 601-604). Springer. 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, *Introduction to Quantitative Analysis for International Educators* (pp. 57-74). Springer. https://doi.org/10.1007/978-3-030-93831-4_5