

# Agrupamiento de Hogares Ecuatorianos según sus Prácticas y Conciencia Ambiental usando K-means

Víctor E. Herrera G., Pablo A. Solano C. - Facultad de Ingeniería, Universidad de Cuenca

**Resumen**—El presente documento describe el procedimiento realizado para la ejecución de un análisis por agrupamiento de los hogares ecuatorianos según sus prácticas y conciencia ambiental. Para esto, se utiliza el dataset resultante de la Encuesta de Seguimiento al Plan Nacional de Desarrollo (EPSPND) realizada por el INEC en el año 2019. Luego de una selección de variables de interés y un tratamiento del dataset se utiliza el algoritmo K-means, apoyado de técnicas como el Método del Codo para agrupar a los hogares ecuatorianos encuestados. Se discuten los resultados con un enfoque que busca clasificar a los hogares ecuatorianos según su nivel de conciencia ambiental y buenas prácticas ambientales. Por último, se discute como los grupos se distribuyen en distintas ciudades del Ecuador.

**Index Terms**—Prácticas Ambientales, Conciencia Ambiental, Hogar Ecuatoriano, Clustering, K-means

## I. INTRODUCCIÓN

La preservación del medio ambiente es una de las principales problemáticas a las que se enfrenta el ser humano hoy en día. Debido a esto es importante que el país cuente con información ambiental coherente y consistente para realizar un seguimiento efectivo sobre el estado y uso de los recursos naturales e informar a la ciudadanía respecto a las principales prácticas para proteger el medio ambiente.

Existen muchas maneras de obtener información relevante respecto de las prácticas y conciencia ambiental de las personas, una de ellas es a través de encuestas. Estos instrumentos ayudan a recolectar distintos tipos de datos que pueden ser sometidos a un análisis estadístico que permite identificar las prácticas y concepciones ambientales más comunes entre las personas.

A nivel mundial, muchos tipos de estudios se han llevado a cabo con este propósito. Sin embargo, la gran parte de estos estudios están enfocados en un proceso descriptivo que busca determinar que es lo hacen o piensan la mayor parte de las personas respecto del medio ambiente, sin considerar la heterogeneidad del grupo al que estudian [1].

La heterogeneidad en las prácticas y actitudes de las personas frente al medio ambiente es un aspecto que puede estar relacionado a distintos factores. Bajo este preámbulo, el análisis a través de agrupamiento o clusterización es una herramienta que puede ser aplicado para analizar la heterogeneidad en las actitudes y conciencia ambiental de las personas.

Existen distintos enfoques a la hora de aplicar un análisis de agrupamiento para determinar grupos de personas con distintas prácticas y conciencia ambiental. Estos enfoques dependen de la información que se utilice para realizar el análisis de agrupamiento. Esta información puede considerar aspectos como patrones de reciclaje [2] o incluso relaciones entre

la conciencia ambiental y el nivel de educación y estrato económico [3].

En el caso de este estudio, el enfoque del análisis de agrupamiento que se pretende realizar depende de la información disponible en Ecuador respecto de las prácticas y conciencia ambiental de los ecuatorianos. Para esto, se realizó una búsqueda exhaustiva a través de distintas fuentes, con el objetivo de encontrar información relevante de las prácticas y conciencia ambiental de los hogares ecuatorianos.

En este contexto, el INEC ejecuta anualmente el Módulo de Información Ambiental en Hogares [4]. Hasta el año 2017 la información se levantaba como parte del módulo de la Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU), y a partir de 2019, se lo realiza a través de la Encuesta de Seguimiento al Plan Nacional de Desarrollo (EPSPND).

Los resultados de esta encuesta incluyen un dataset que contiene información de hábitos y prácticas de 11,528 hogares ecuatorianos, con 109 variables que constituyen las respuestas a preguntas respecto a temas ambientales como clasificación de residuos, disposición de residuos peligrosos y especiales, prácticas de uso de agua, prácticas de uso energía, pautas de consumo responsable, transporte y movilidad, conciencia y responsabilidad ambiental.

El objetivo de este estudio es agrupar los hogares con hábitos y prácticas ambientales similares con el objetivo de que los grupos generados permitan clasificar a los hogares ecuatorianos según su nivel de conciencia ambiental y buenas prácticas ambientales. Esto permite describir la heterogeneidad de los hogares ecuatorianos respecto a las prácticas ambientales y conciencia ambiental de los mismos.

## II. MATERIALES Y MÉTODOS

### II-A. Contexto de la Encuesta

El objetivo inicial de la recolección de datos por medio de una encuesta de las prácticas, hábitos y conciencia ambiental de los hogares Ecuatorianos fue para cumplir con el Módulo de Información Ambiental en Hogares que desarrolla anualmente el INEC. Este estudio se realiza con la intención de que el país cuente con información ambiental, coherente y consistente para realizar un seguimiento efectivo sobre el estado y uso de los recursos naturales, además de informar a la ciudadanía respecto a las principales prácticas para proteger el medio ambiente [4].

El INEC tiene como objetivo principal determinar los hábitos y prácticas de los hogares y las personas respecto a temas

ambientales como: clasificación y eliminación de residuos, prácticas de ahorro de agua y energía, transporte y movilidad, pautas de consumo, conciencia y responsabilidad ambiental [4].

El INEC evaluó varios factores que se deben tomar en cuenta como introducción a la información tratada a lo largo de este documento. Estos factores se pueden dividir en los siguientes dominios:

1. Clasificación de residuos habituales
  - Pregunta sí o no a clasificación de residuos orgánicos.
  - Pregunta sí o no a clasificación de residuos inorgánicos como: - Papel, cartón, plástico, vidrio, metal, Tetrapak.
  - Pregunta de opción múltiple de como se eliminó los residuos dentro del hogar. Siete opciones de respuesta para 6 tipos de residuos.
2. Disposición de residuos peligrosos y especiales
  - Pregunta si y no sobre si se utiliza pilas en el hogar. Si se responde si se indica cuantas pilas se tiene en el hogar y cuantas de estas son recargables.
  - Pregunta sí y no sobre si se utiliza focos en el hogar. Si se responde sí se indica cuantos focos se tiene en el hogar, cuantos de estos son ahorradores, y cuantos de estos son focos LED.
  - Pregunta de opción múltiple de como se eliminó los residuos especiales como pilas, focos o aceites dentro del hogar. Siete opciones de respuesta para 6 tipos de residuos.
3. Prácticas de uso de agua
  - Preguntas de tres opciones (sí, no y no aplica) sobre el uso de agua en el hogar como si se reusa el agua para varias actividades, si se utiliza balde en lugar de manguera, si se cierran las llaves mientras se realiza actividades de aseo, etc. Ocho preguntas en total.
4. Prácticas de uso de energía
  - Preguntas de tres opciones (sí, no y no aplica) sobre el uso de energía en hogares como si se desconectan aparatos electrónicos cuando no se usan, si apaga los focos al salir de una habitación, si se hace uso de eventos naturales para evitar el gasto de energía, etc. Siete preguntas en total.
5. Pautas de consumo responsable
  - Pregunta de tres opciones donde se da paso a responder sobre que contenedor que se uso para realizar compras el mes pasado.
  - Pregunta de opción múltiple donde se pregunta acerca que factor es el que mas le importa al momento de adquirir un nuevo producto electrodoméstico, de alimentación, etc.

## 6. Transporte y movilidad

- Pregunta acerca la posesión de al menos un vehículo motorizado.
- Si se dispone de un vehículo motorizado se responde acerca que combustible se usa con mayor frecuencia.

## 7. Conciencia y responsabilidad ambiental - Preguntas

- de si o no acerca problemas ambientales que afectan el barrio, entre los posibles problemas están: contaminación visual, agua contaminada, ruidos excesivos, acumulación de basura, etc.
- Pregunta acerca la preocupación que se tiene por la situación ambiental en su barrio, se da una escala de preocupación de cinco ítems ordenados de menor a mayor.
- Preguntas de si o no acerca parte actividades que algún miembro familiar ha formado, entre las posibles actividades están: colaborar con tiempo y dinero en defensa del ambiente, participar en voluntariados ambientales, manifestarse contra situaciones perjudiciales para el ambiente, denunciar algún problema ambiental que se haya identificado.
- Preguntas a calificar a partir que tan de acuerdo se esta en una debida afirmación sobre la responsabilidad de varios agentes en el cuidado del medio ambiente. Se da una escala de 5 ítems donde 1 es totalmente en desacuerdo y 5 es totalmente de acuerdo, además se ubica 99 cuando no responde.

Un diccionario de variables detallado se puede encontrar en las fuente del INEC [4]. Las preguntas sobre transporte, pautas de consumo responsable y desechos eléctricos y electrónicos no se consideraron en este estudio, la razón de esto se explica en la siguiente sección.

### II-B. Selección de Variables a Utilizar

Estudios que consideran aspectos socio-económicos como el de Philippsen, Soares y Santana [3], concluyen que la conciencia ambiental puede estar relacionada a factores como el nivel de educación y los ingresos de una persona. Para poder seguir un enfoque similar al de este estudio es necesario contar con información tanto de aspectos socio-económicos así como de conciencia ambiental.

Por otro lado, estudios como el de Elgaaeid [2] realizan un análisis por agrupamiento considerando aspectos como patrones de reciclaje asociados únicamente a las prácticas y conciencia ambiental. Este estudio no considera variables socio-económicas ni ninguna otra variable que puede estar asociada a las mismas.

El dataset proporcionado por el INEC corresponde a una encuesta multipropósito que contiene información sobre prácticas ambientales, conciencia ambiental, pautas de consumo responsable y transporte. Estos dos últimos aspectos, pautas de consumo y transporte, están relacionados con aspectos

socio-económicos de los hogares ecuatorianos. Sin embargo, el dataset en cuestión no cuenta con información socio-económica adicional de los hogares encuestados.

Respecto de las pautas de consumo por ejemplo, un análisis rápido demostró que la mayor parte de los hogares solo consideran el precio a la hora de comprar algún producto. Sin variables socio-económicas asociadas a esto no se puede encontrar mayores diferencias entre estos hogares del porque preferir el precio sobre otros aspectos según su situación socio-económica.

Algo parecido sucede con el transporte, la mayor parte de hogares encuestados no cuenta con automóvil propio, los hogares que si cuentan con uno contestan preguntas sobre consumo de combustibles, aunque estos son muy pocos en comparación a los que no. El problema es como diferenciar entre los hogares que no cuentan con un automóvil, no hay información adicional de como se transportan, simplemente se sabe que no utilizan automóvil.

Todo lo descrito anteriormente hace que el enfoque adecuado para nuestro estudio sea similar al de Elgaaeid [2], donde no se tomen en cuenta variables asociadas a aspectos socio-económicos y solo se consideren aspectos ligados a las prácticas y conciencia ambiental, descartando las variables relacionadas a pautas de consumo y transporte.

Por último, respecto de las variables de disposición de residuos eléctricos y electrónicos, un análisis previo demostró que la mayor parte de la población no genera este tipo de residuos. La cantidad de hogares que lo hacen es prácticamente despreciable. Por ende, tampoco se consideran las variables concernientes a este hecho en el agrupamiento, debido a que estas variables no aportan la heterogeneidad necesaria para ello.

De esta manera las 60 variables seleccionadas tienen que ver únicamente con aspectos relacionados a prácticas y conciencia ambiental de los siguientes tópicos:

- Clasificación de residuos (papel, plástico, vidrio, metal, etc.)
- Disposición de residuos peligrosos y especiales (pilas, focos, desechos farmacéuticos, etc.)
- Prácticas de uso de agua
- Prácticas de uso energía
- Conciencia y responsabilidad ambiental

## II-C. Tratamiento del Dataset

Las variables seleccionadas del dataset constituyen las respuestas a varios tipos de preguntas. Se pretende codificar numéricamente las variables no numéricas para facilitar el proceso de agrupamiento. La idea es codificar numéricamente todas las variables de manera que un valor mayor es un indicador de una menor conciencia ambiental. Para esto, respuestas afirmativas, respuestas que indiquen estar a favor de algo o que exhiban una buena práctica ambiental deben ser codificadas con valores numéricos menores.

El primer problema surgió con las preguntas de Sí o No, en las que se pueden dar respuestas de No Aplica, o con las preguntas con una escala de Likert en las que se pueden dar

respuestas de No Sabe / No Responde (NS/NR). Si bien estas no son respuestas faltantes, son respuestas que no dicen nada. Para tratar con este tipo de respuestas se siguió un criterio de neutralidad, en donde se asigna una valor intermedio para estas respuestas [5].

Por otro lado, las opciones de las preguntas de opción múltiple de residuos y las escalas de Likert pueden ser consideradas ordinales, por lo que una codificación de número entero es suficiente [6]. La codificación numérica de todas estas variables se puede observar en el Cuadro I.

Cuadro I  
CODIFICACIÓN DE VARIABLES

Tipo Pregunta	Respuesta	Codificación	
Sí / No	Sí	=	0
	No Aplica	=	0.5
	No	=	1
Opción Residuos	No generó estos residuos	=	0
	Preparó Compost Abono	=	1
	Envío a centro de acopio	=	2
	Regaló, vendió	=	3
	Almacenó en hogar	=	4
	Depositó en basura común	=	5
	Quemó, enterró, botó	=	6
Escala Likert	Totalmente De Acuerdo	=	0
	...	...	...
	Neutral o NS/NR	=	2
	...	...	...
	Totalmente En Desacuerdo	=	4

Por otro lado, se tienen variables numéricas como las cantidades de uso de pilas, pilas recargables, focos, focos ahorradores y focos ahorradores LED. Existen valores faltantes cuando se responde de manera negativa al uso de estos artefactos. Estos se rellenan con cero, el no usar algo es igual a usar una cantidad nula de ese algo.

Además, las cantidades de pilas recargables, focos ahorradores y focos ahorradores LED son transformadas a las diferencias entre pilas y pilas recargables, focos y focos ahorradores, focos ahorradores y focos ahorradores LED respectivamente. Esto se hace para cumplir con el criterio de que un valor mayor es un indicador de una menor conciencia ambiental.

Por último, todas las variables ya transformadas y codificadas se normalizan en valores entre 0 y 1. Esto para facilitar

el procesamiento y la interpretación de los datos respecto a diferentes escalas.

#### II-D. Algoritmos y Técnicas Utilizadas

Para el proceso de clusterización de la información recolectada y tratada, se hará uso del algoritmo K-means con su respectivo análisis previo por medio de dos técnicas diferentes para determina el numero de clústers a utilizar [7].

*II-D1. K-means:* El algoritmo K-means es uno de los algoritmos mas simples y poderosos para realizar el proceso de clusterización.

Este realiza la clusterización basándose en un numero k de centroides definidos previamente a su ejecución. Cada centroide definirá un cluster, por lo que la posición inicial de estos es crucial para el resultado final, dos configuraciones de centroides diferentes dará resultados diferentes.

El proceso del algoritmo consiste en tomar cada uno de los puntos dados por el dataset y asociarlo con el centroide mas cercano. Cuando todos los puntos se han asociado la primera iteración del algoritmo se ha completado. En este punto se deben recalculan nuevos centroides dado los del paso previo y se repite el proceso explicado con estos nuevos centroides. Por cada iteración del algoritmo se vera como los centroides cambian de paso a paso, el algoritmo acaba cuando no halla cambio en la nueva posición de los centroides.

*II-D2. Método del Codo:* Tal vez el método mas conocido para determinar el numero de clústers en un dataset es el método del codo. Este es un método visual, el cual consiste de calcular la clusterización de la información para  $k = 1, 2, 3, 4, 5, \dots$ . Al calcular los clústers, se usa el costo que se realiza para el entrenamiento y se gráfica punto por punto, en alguno de estos el costo bajara dramáticamente y después de esto se encontrara con un plateau en el cual no hay un cambio importante en el costo. EL valor k en el cual el costo descendió dramáticamente es la cantidad de clústers ideal al cual clústerizar el dataset [7].

*II-D3. Silhouette:* Otro análisis para determinar el numero óptimo de clústers en un dataset es el método de la silueta o "silhouette method". Este análisis se puede utilizar para estudiar la distancia de separación entre los clústers resultantes. El coeficientes de silueta muestra un resumen de la medida de qué tan cerca está cada punto en un grupo de puntos en los grupos vecinos y, por lo tanto, proporciona una forma de evaluar parámetros como el número de grupos visualmente. Esta medida tiene un rango de  $[-1, 1]$ . Los coeficientes de silueta cercanos a +1 indican que la muestra está lejos de los clústers vecinos. Un valor de 0 indica que la muestra está en o muy cerca del límite de decisión entre dos clústers vecinos y los valores negativos indican que esas muestras podrían haber sido asignadas al clúster incorrecto.

#### II-E. Herramientas Utilizadas

Para el tratamiento, clusterización y análisis de datos se hacen uso de las siguientes librerías de Python 3:

- scikit-learn
- pandas

- matplotlib
- plotly

Además, se hace uso de Weka para ciertas tareas de pre-procesamiento de datos, como llenado de valores vacíos, eliminación de atributos y normalización.

### III. PROCEDIMIENTO

Se procedió a realizar un proceso de agrupamiento considerando todas las variables seleccionadas. Para esto primero se aplicó el Método del Codo usando scikit-learn para poder determinar el número de grupos óptimo [8]. Adicionalmente, también se aplica el método Silhouette para tener una segunda opinión respecto del número de grupos óptimo. Sin embargo, este método no tuvo muy buenos resultados como se podrá ver en la sección de Resultados, por lo que se utiliza el criterio del codo.

Luego de haber obtenido el número de grupos óptimo, se procede a realizar el agrupamiento por K-means con todas las variables seleccionadas. Los resultados se interpretan usando usando gráficos de tipo Line Polar [9], que permiten ver los valores promedio de cada variable para todos los clústers. Cada variable se muestra en un radio de una circunferencia en donde la magnitud de la misma asciende desde el centro del círculo hacia afuera.

Para cada clúster la unión de todas las variables forma un polígono que describe las magnitudes de las variables en conjunto a través del área del mismo. El área de este polígono no es el único indicador de las características del clúster. Por ejemplo, un clúster con un área menor a otro podría indicar que de manera general y conjunta las magnitudes de las variables del mismo son menores, pero sin embargo el clúster de área mayor podría tener variables menores en magnitud al clúster de área menor. Además, las distintas variables pueden tener distintos pesos o relevancias. Este tipo de casos específicos se abordan en la sección de Discusión.

Sin embargo, recordando que las variables fueron codificadas de manera que un valor mayor es un indicador de una menor conciencia ambiental, el área del polígono de un clúster podría ser capaz de indicar de manera general el nivel de conciencia ambiental de sus miembros. Mientras más pequeña sea el área del polígono del clúster mayor será el nivel de conciencia ambiental de manera general de sus miembros.

El gráfico Line Polar resultante de usar todas las variables seleccionadas permite observar diferencias claras entre algunas variables, sin embargo las diferencias entre clústers respecto de otras variables no son tan visibles. Esto podría deberse a que son muy pocas o pequeñas. No obstante, es de interés del estudio encontrar la mayor cantidad de diferencias entre los hogares ecuatorianos en cada una de las secciones de la encuesta. Por esta razón, se procedió a realizar un procedimiento de agrupamiento de los hogares por cada sección de la encuesta INEC utilizada para el estudio.

De esta manera, se obtiene un agrupamiento de todos los hogares encuestados por cada una de las siguientes secciones: clasificación de residuos, disposición de residuos especiales,

prácticas de uso de agua, prácticas de uso de energía, conciencia y responsabilidad ambiental. Para cada una de estas se aplica el Método del Codo y Silhouette para determinar el número óptimo de clústers. Sin embargo, cuando ninguno de los métodos es lo suficientemente preciso por lo que se trata de mantener un número de clústers uniforme para todas las secciones.

Por último, se comparan el proceso de agrupamiento usando todas las variables al mismo tiempo con el agrupamiento por sección y se usan los clústeres obtenidos en el agrupamiento con todas las variables para describir de manera general la situación respecto a los niveles de conciencia ambiental en las distintas áreas (urbana o rural) o cantones encuestados [10].

#### IV. RESULTADOS Y DISCUSIÓN

##### IV-A. Agrupamiento con Todas las Variables

El resultado de aplicar el Método del Codo para el dataset con todas la variables seleccionadas para este estudio se puede ver en la Figura 1.

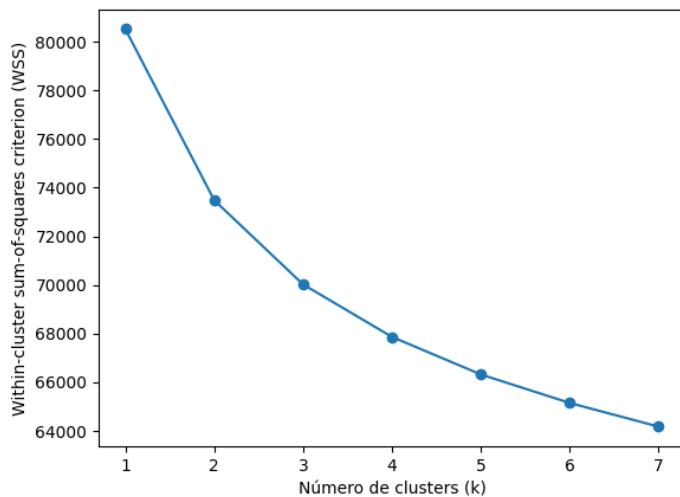


Figura 1. Método del Codo - Agrupamiento con Todas las Variables

Se puede ver que el codo se encuentra en  $k = 3$ , lo que significa que la clusterización del dataset con todas la variables se debería realizar con 3 clústers. Adicionalmente, para tener más seguridad en la elección del número de clústers, se aplica el Método de la Silueta, los s-cores para  $k = 1, 2, 3, 4, 5, 6, 7$  se pueden ver en el Cuadro II.

Se puede ver que los s-cores del método de la silueta para  $k$ , desde  $k = 2$  a  $k = 7$ , son muy cercanos a cero. Esto podría indicar que la mayor parte de las instancias del dataset están muy cerca del límite de decisión entre dos clústers vecinos. A pesar de que los mejores resultados se obtienen para  $k = 2$  y  $k = 3$ , se concluye que el método de la silueta no es adecuado para determinar el numero de clústers de este dataset. Por lo que la elección de  $k$  se basa en el método del codo.

Cuadro II  
COEFICIENTES DEL MÉTODO SILUETA

k = número de clústers	S-core
k = 2	0.09656924976049075
k = 3	0.07512973746434341
k = 4	0.0621648317580256
k = 5	0.05586786366197981
k = 6	0.04919352308044649
k = 7	0.050173998244743244

Se realiza el proceso de agrupamiento con todas las variables y con un número de clústers igual a tres. El gráfico Line Polar que ilustra los resultados este proceso se puede ver en el la Figura 2.

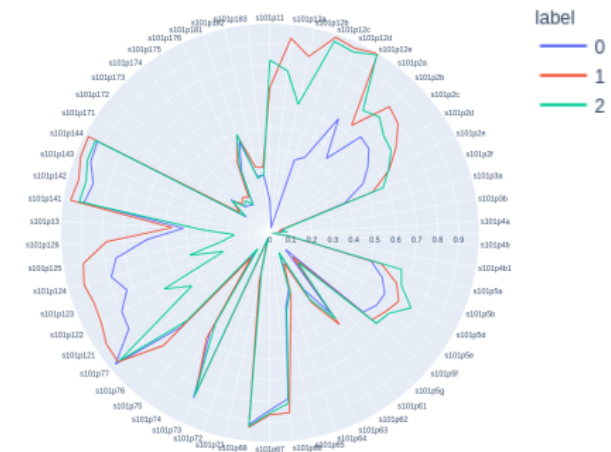


Figura 2. Line Polar - Agrupamiento con Todas las Variables

Esta figura permite ver que las mayores diferencias entre los clústers están presentes en las variables concernientes a clasificación de residuos y con la conciencia ambiental respecto de la percepción de problemas ambientales en los barrios de los hogares. No obstante, también es importante saber en que variables se parecen los clústers y como lo hacen.

Comenzando con la similitudes, parece que la mayor parte de los hogares usan pilas y focos de manera moderada. Además, pareciese que la gran parte de los hogares tienen buenos hábitos de ahorro de agua, a excepción de cuestiones como disponer de economizadores de chorro, inodoros de doble descarga o colocar una botella de agua u otro objeto en el inodoro. Adicionalmente, pareciese que la gran parte de los hogares tiene buenos hábitos de ahorro de energía, a excepción de cuestiones como evitar introducir alimentos calientes en el refrigerado o planchar la mayor cantidad de ropa de una sola vez. Por último, la mayor parte de hogares parece tener una buena conciencia ambiental.

Por otro lado, comenzando con las diferencias entre los clústers, en el Clúster 0 se encuentran la gran mayoría de los hogares que sí clasifican la mayor parte de residuos orgánicos e inorgánicos como papel, cartón, plástico, vidrio, metal y tetrapak. El Clúster 2 contiene una buena parte de los hogares que no clasifican desechos como vidrio o metal, pero tienen una clasificación moderada del resto de residuos. El Clúster 1 contiene a la mayor parte de hogares que no clasifican casi ningún tipo de residuo, a excepción de residuos orgánicos.

Respecto de la eliminación de residuos habituales, en el Clúster 0 se encuentran la mayor parte de hogares que realizan actividades como preparar ábono con desechos orgánicos y que no generan muchos de los tipos de desechos inorgánicos o que de hacerlo los envían a un centro de acopio o los depositan en un contenedor especial. En el Clúster 2 se encuentran la mayor parte de las personas que regalan o conservan algunos tipos de desechos inorgánicos, además de una parte que deposita desechos en la basura común. En el Clúster 1 la mayor parte de las personas desechan todo tipo de residuos en la basura común, y otra parte bota, arroja o quema los desechos.

Por otra parte, en el Clúster 2 se encuentran la mayor parte de hogares que perciben problemas ambientales como contaminación visual, agua contaminada, ruidos excesivos, acumulación de basura, contaminación del aire o presencia de animales y sus excrementos en sus barrios. Esta no es una característica común en los hogares del Clúster 0 o del Clúster 1. Además, existen diferencias muy ligeras en variables concernientes a disposición de residuos especiales. Sin embargo, este gráfico podría indicar que el resto de las variables son muy similares para todos los clústers. Debido a esto en la siguiente sección se presentan y se discuten los resultados obtenidos con el agrupamiento por secciones que permite observar diferencias no visibles de manera tan clara con este procedimiento.

Por último, siguiendo el criterio del área del polígono se puede ver que de manera general el Clúster 0 podría representar la mejor conciencia ambiental, seguido del Clúster 2, y por último el Clúster 1. En la Figura 3 se puede observar la distribución porcentual de los hogares encuestados en los clústers descritos anteriormente.

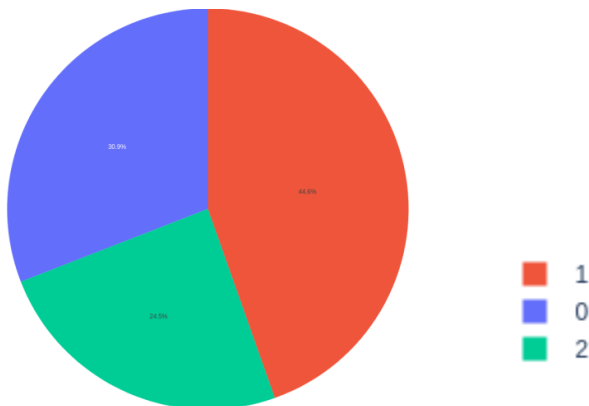


Figura 3. Distribución Porcentual - Agrupamiento General

Se puede ver que el 30.9 % clasifica en el Clúster 0, esto quiere decir que tiene buenas prácticas en la clasificación de desechos y residuos habituales. Además, se puede ver que el 44.6 % de los hogares encuestados clasifica en Clúster 1, es decir que podrían mejorar mucho sus hábitos de clasificación y eliminación de residuos habituales. Por último, el 24.5 % de los hogares clasifican en el Clúster 2, esto quiere decir que existen problemas ambientales en sus barrios.

#### IV-B. Agrupamiento por Sección de Preguntas

Debido a que el análisis previo no permite diferenciar a los hogares en base a ciertos aspectos como disposición de residuos especiales, prácticas de uso de agua, prácticas de uso de energía o ciertos aspectos de conciencia y responsabilidad ambiental se procede a realizar un análisis de agrupamiento por cada una de las secciones de la encuesta.

El proceso que se sigue para cada una de las secciones es el mismo realizado para el análisis que considera todas las variables. La diferencia es que se consideran únicamente las variables de cada sección. La Figura 4 permite ver el resultado del gráfico Line Polar para el agrupamiento para la sección clasificación de residuos. Únicamente se incluye el gráfico de esta sección, pero los resultados para cada una de las secciones de la encuesta se presentan y discuten en las siguientes subsecciones del documento.

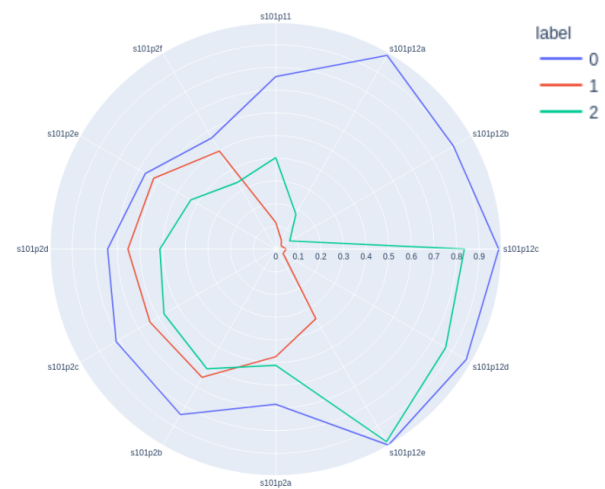


Figura 4. Line Polar - Agrupamiento con Preguntas de Residuos

#### IV-C. Agrupamiento por Clasificación de Residuos

A partir del criterio del área del polígono aplicada sobre la clusterización representada por medio del gráfico line polar, se puede ver una clara diferencia entre los tres clústers donde el primero tiene una gran tendencia a caer en lo que clasificamos como malas prácticas hacia el medio ambiente en lo que se trata la clasificación de residuos pues en ninguna pregunta los hogares que forman parte de este clúster caen debajo del 0.5 en la escala de magnitud establecida. Este cluster, reflejado en el agrupamiento general se podría asociar como el clúster 1 pues se puede ver como no clasifica ningún tipo de residuos (orgánico e inorgánico), además de esto se ve como para

la eliminación de residuos este clúster posee prácticas muy alejadas de lo ideal, es decir la mayoría de hogares agrupados por este clúster tienen una tendencia a aplicar prácticas como depositar todo tipo de desecho en basura común, quemarla o enterrarla.

Por otro lado, el segundo clúster ejerce malas prácticas para la clasificación de los tipos de residuos de plástico, vidrio, metal y tetra pak, mientras que para los tipos de residuos como papel, cartón y plástico tiene lo que se consideraría como ideales. Sumado a esto se ve que las prácticas de eliminación de residuos se consideran tolerables pues existe una tendencia normal a realizar prácticas como regalar vender o almacenar en su hogar. Por todo esto, este clúster se puede ver reflejado en el agrupamiento general como el clúster 2.

El último cluster, que se ve reflejado por el clúster 0 en el agrupamiento general pues posee buenas prácticas al momento de clasificar residuos ya sean orgánicos e inorgánicos, mientras que para la eliminación de desechos como vidrio, metal y tetra pak los hogares clasificados por este clúster tienen prácticas un poco similares al primer clúster explicado. Este cluster, hablando solamente de las prácticas de eliminación de residuos se podría ver como un punto medio entre el primer y segundo clúster explicado.

#### *IV-D. Agrupamiento por Disposición de Residuos Especiales y Peligrosos*

Lo primero a notar en la clusterización de los hogares en este tipo de preguntas es que uniformemente los hogares tienen un coeficiente muy bajo en la escala de magnitud establecida, esto se debe interpretar a que el uso de focos y pilas es moderado debido a que una mala práctica con respecto a estos elementos no se considera poseer un gran número de estos, sino que, la diferencia entre por ejemplo, los focos y los focos ahorradores que se posee sea grande, debido al consumo de energía que focos normales representa. Esta tendencia es claramente evidenciada en el agrupamiento general.

Dentro de las cualidades propias de cada clúster, podemos destacar que los hogares que caen en el primero, con respecto al cómo se elimina elementos como pilas, desechos farmacéuticos, aceites y grasas, focos ahorradores, etc, tiene una gran tendencia a realizar prácticas como quemar, enterrar y botar siendo esta las actividades principales invariantes para este cluster, por esta razón se asocia la mayoría de este clúster con el clúster 2 del agrupamiento general, donde el porcentaje de hogares ecuatorianos que caen en este clúster es de 53.2 %. El siguiente clúster es muy similar al anterior, pero la variante de este sería que además de quemar, enterrar y botar este grupo también toma parte en practicas de eliminación como el depósito de basura común, que se podría catalogar como una actividad menos negativa contra el medio ambiente. Además, y tal vez la más grande variante con el clúster anterior es que los hogares de este clúster prácticamente no genera residuos de cartuchos de impresora y toners. Por estas razones podemos asociar la mayoría de este clúster con el clúster 1 del agrupamiento general, donde caen un total del 29.3 % de los hogares ecuatorianos.

El último clúster y que a simple vista se puede catalogar como el clúster que aplica prácticas más conscientes con el medio ambiente abarca un total de 17.4 % de hogares ecuatorianos, donde se aplican buenas prácticas de eliminación de residuos especiales excepto por residuos de grasa y aceite donde se emplean técnicas como regalar y vender o almacenar en su hogar, por lo que la mayoría de hogares de este clúster se asocia con el clúster 0 del agrupamiento general.

#### *IV-E. Agrupamiento por Prácticas de Uso de Agua*

Acerca la clusterización del uso agua por parte de los hogares ecuatorianos se puede decir que sobre las preguntas sobre el tiempo que toman para ducharse, sí revisan regularmente las tuberías, sobre si usan economizadores de chorro, sobre el inodoro doble descarga o sobre sí colocan una botella de agua o otro objeto dentro del inodoro, todos los hogares independientes de clúster se comportan de la misma manera, por lo que podemos concluir que sí se duchan en menos de 10 minutos, sí revisan las tuberías, no disponen de economizadores, no disponen de inodoros doble descarga y no colocan objetos en el tanque del inodoro.

En donde se existen variaciones entre clústers son en las preguntas sobre el reuso de agua y sí se usa un balde en lugar de manguera.

El primer clúster en donde caen el 35.8 % de hogares ecuatorianos, es el que tiene peores prácticas en el uso de agua, es decir para las preguntas que muestran diferencias con respecto a los otros clústers estos hogares no la realizan. Lo que significa que este clúster en su mayor parte se puede asociar con el clúster 1.

El segundo clúster el cual tiene la menor cantidad de hogares ecuatorianos con el 25.4 %, no reusa el agua mientras que si utiliza balde en lugar de manguera para ciertas actividades, la mayor parte de este clúster se la identifica como el clúster 2. Por parte del último clúster a analizar, se puede ver que si reusa el agua al igual que utiliza un balde en lugar de manguera, en este caen el 38.8 % de hogares ecuatorianos y se asocia en su mayoría con el clúster 0 del agrupamiento general.

#### *IV-F. Agrupamiento por Prácticas de Uso de Energía*

La discusión sobre los clústers de uso de energía merece un análisis por cada pregunta pues establecer puntos comunes o diferentes entre cada uno resulta complicado. Acerca si los hogares desconectan aparatos electrónicos cuando no los usan se puede decir que todos los clústers agrupan a hogares que sí realizan esta práctica. Acerca si los hogares apagan focos al salir de una habitación, todos los hogares si realizan esta práctica. Acerca si los hogares introducen alimentos calientes al refrigerador, se puede decir que el primer clúster si lo hace, el segundo los hogares en su mayoría no lo hacen y en el tercero prácticamente todos los hogares que caen dentro de este no lo hacen. Acerca si los hogares abren las cortinas para aprovechar la luz del sol, todos los clústers consisten de hogares que en su mayoría si lo hace. Acerca si los hogares



utilizan electrodomésticos ahorradores de energía, el primer clúster y el último consisten de hogares que no lo hacen, mientras que en el segundo clúster est[an ]la mayor[i]a de hogares que si lo hacen. Y por último, acerca de los hogares que disponen de paneles solares, todos los clústers consisten de hogares que no cumplen con esta práctica. Por todo esto se puede concluir que el primer clúster que contiene el 39.3 % de todos los hogares ecuatorianos, se asocia en su mayoría con el clúster 1. Mientras que el segundo clúster se lo puede asociar en su mayoría con el clúster 0, conteniendo el 33.9 % total de los hogares ecuatorianos. Y finalmente el último clúster se asocia en su mayoría con el clúster 2, con un total de 26.8 % de hogares ecuatorianos.

#### IV-G. Agrupamiento por Conciencia y Responsabilidad Ambiental

Por parte de la clusterización de las preguntas de conciencia ambiental se puede decir que todos los clústers tienen una tendencia similar para preguntas de participación en actividades pro-ambientales donde se ve que ninguno toman parte en este tipo de actividades, por otro lado en las preguntas sobre responsabilidad ambiental donde la tendencia se revierte y aunque variante la mayoría de hogares en todos los clústers optan que la responsabilidad es dependiente de cada sujeto explicado en cada pregunta. En la pregunta sobre si proteger el medio ambiente encarece la vida, existe en todos los clústers tanto hogares que estan de acuerdo con esta afirmación como que no, mientras que en el resto de estas preguntas sobre si el ambiente genera ahorros para su hogar y si proteger el ambiente requiere tiempo y esfuerzo se está en su mayoría de acuerdo.

El primer clúster donde se agrupan 52.1% de hogares, respecto a las preguntas sobre tipos de problemas ambientales en el barrio, se cuenta con hogares que en su mayoría no tienen, pudiendo así asociarlo en mayor parte con el clúster 1. El segundo clúster donde se encuentran 26.9 % de hogares, se ve que hay una tendencia a que no se encuentran problemas ambientales en el barrio para la mayoría de hogares en este clúster excepto por el problema ambiental de la presencia de animales callejero y sus excrementos, asociándolo así con el clúster 2.

El último clúster entra en contraste evidente con los otros dos pues para todas la preguntas sobre problemas ambientales en el barrio se encuentra que si existen todos los problemas planteados en el cuestionario para la mayoría de hogares en el cluster, siendo este el 21 % de todos los hogares encuestados, y que se lo puede asociar en su mayoría con el clúster 0.

#### IV-H. Análisis Final - Agrupamiento Total y Por Sección

Los análisis descritos anteriormente nos han permitido llegar a conocer las características principales de cada uno de los clústers que se describen en este documento, sus similitudes y principalmente sus diferencias. A continuación se describen cada uno de los mismos:

- **Clúster 0:** La principal característica diferenciadora de los hogares de este clúster es que clasifican y desechan la

mayoría de residuos habituales, especiales y peligrosos con buenas prácticas amigables con el ambiente. Además, a diferencia de los hogares en otros clústers tienen algunas mejores prácticas en el uso de agua.

- **Clúster 1:** La principal característica diferenciadora de los hogares de este clúster es que NO clasifican y desechan muchos de los residuos habituales, especiales y peligrosos con buenas prácticas amigables con el ambiente. Además, irónicamente son los que menos problemas ambientales reportan en sus barrios.
- **Clúster 2:** La principal característica diferenciadora de los hogares de este clúster es que tienen varios problemas ambientales en sus barrios. Además, los hogares de estos clústers desechan muchos de los residuos especiales y peligrosos de manera irresponsable.

#### IV-I. Grupos por Área

Es importante poder conocer cuál es la situación de los clústeres dependiendo del área en la que se ubican los hogares. En la Figura 5 se puede observar la distribución de los clústers entre las dos diferentes áreas (urbana y rural) de los hogares ecuatorianos encuestados.

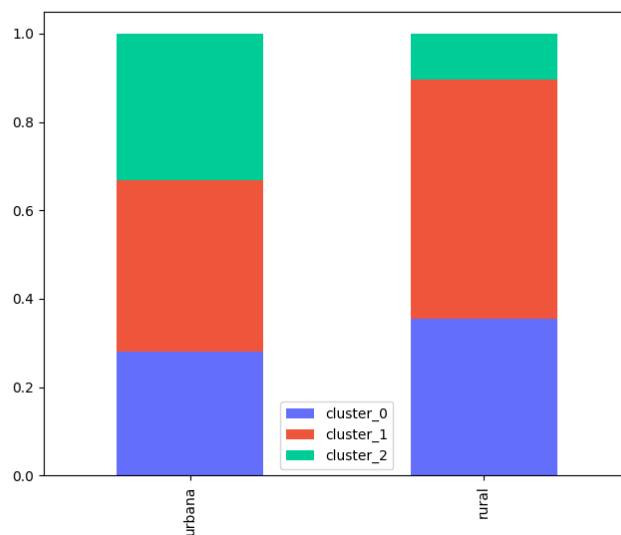


Figura 5. Clústers por Área

Se puede observar que en el área urbana los hogares clasificados en el Clúster 2 representan aproximadamente la tercera parte de los hogares urbanos. Algo similar ocurre con el Clúster 0. Mientras que los hogares clasificados en el Clúster 1 son aproximadamente el 40 % de los hogares urbanos.

En el área rural los hogares en clasificados en el Clúster 2 son una muy pequeña parte de menos del 10 %. Los hogares clasificados en Clúster 1 representan más de la mitad de los hogares rurales. Por último, el Clúster 0 contiene poco menos de la cuarta parte de los hogares rurales.



#### IV-J. Grupos por Cantón

Por otra parte, es importante poder conocer cuál es la situación de los clústeres dependiendo de los cantones distribuidos en las distintas regiones del país. En la Figura 6 se puede observar la distribución de los clústeres entre 17 cantones que constituyen los cantones con más hogares encuestados de cada región.

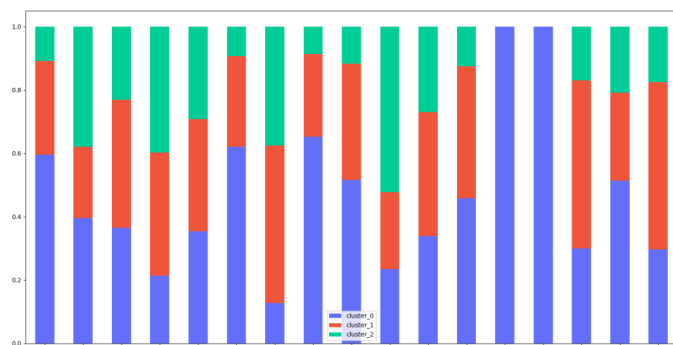


Figura 6. Clústeres por Cantón

Lo más destacable de este gráfico es que los dos cantones de la región insular tienen todos sus hogares agrupados en el Clúster 0. Por otro lado, se puede apreciar que grandes urbes como Quito, Guayaquil o Manta tienen una gran cantidad de hogares en el Clúster 2 y una pequeña cantidad de hogares en el Clúster 0.

Por otra parte, al sur del país se encuentran ciudades como Cuenca, Macas y Loja que cuentan con las menores cantidades de hogares en el Clúster 2 y las mayores cantidades en el Clúster 0. Por último, no parece haber diferencias claras entre una región u otra (exceptuando la región insular), pero las ciudades de la costa son las que tienen las mayores cantidades de hogares en el Clúster 1.

#### V. CONCLUSIONES

Este estudio ha permitido agrupar en tres grupos diferentes a los hogares ecuatorianos según su nivel de conciencia y prácticas ambientales. El primero de estos grupos, describe a los hogares con mejores prácticas de clasificación y desecho de todo tipo de residuos, además de una mejor cultura con respecto al uso del agua. Esto podría indicar que este grupo es el que mejores prácticas y conciencia ambiental tiene.

Por otro lado, el segundo grupo describe a los hogares con prácticas no amigables con el ambiente en la clasificación y desecho de todo tipo de residuos. Sin embargo, este grupo es el que menos problemas ambientales reporta en sus barrios.

Además, el tercer grupo describe a los hogares que tienen la mayor parte de los problemas ambientales en sus barrios. En este grupo es donde no hay una buena cultura al momento de clasificar y desechar residuos especiales y peligrosos, lo que podría explicar los problemas que estos hogares se encuentran en su barrio.

Respecto a como se distribuyen estos grupos en las áreas urbana y rural se puede ver que hay una cantidad parecida de hogares con buenas prácticas ambientales en ambos. Por otra parte, el área urbana es la que más problemas ambientales presenta, lo cual contrasta con el área rural que es la que tiene menos hogares con problemas ambientales, a pesar de que una gran parte de los hogares en el área rural tienen malas prácticas ambientales. Esto podría indicar que la mayor parte de los hogares pertenecientes al segundo grupo no tienen problemas ambientales a pesar de sus malas prácticas solo porque se encuentran en el área rural.

Por otra parte, con relación a la distribución de los grupos en los distintos cantones en las cuatro regiones del Ecuador se puede concluir que la región con mejores prácticas ambientales es la región insular, esto podría deberse a la cultura de preservación ambiental existente en las islas Galápagos.

Con respecto al resto de cantones, podemos decir que las grandes urbes como Quito, Guayaquil o Manta son las que más problemas ambientales tienen, lo que podría ser causado por el gran número de hogares con malas prácticas ambientales.

Una excepción a esta regla es el cantón Cuenca que junto a otras ciudades del sur del país como Loja y Macas poseen una gran cantidad de hogares con buenas prácticas ambientales, lo cual además se evidencia con su reducida cantidad de hogares con problemas ambientales.

Contemplando únicamente el Ecuador continental no parece que ningún grupo pueda asociarse a una región específica. Sin embargo, la mayor parte de las ciudades de la costa son las que tienen una gran cantidad de hogares con malas prácticas ambientales.

Por último, se puede concluir que los ecuatorianos deben mejorar sus prácticas relacionadas con la clasificación y desecho de residuos habituales, especiales y peligrosos, este estudio demuestra que este es el problema más recurrente en muchos hogares. Para esto tanto los ciudadanos como autoridades deben tomar medidas respecto a este problema con mayor prioridad frente a otros aspectos ambientales.

#### REFERENCIAS

- [1] T. K. Yoon y S. Ahn, "Clustering Koreans' Environmental Awareness and Attitudes into Seven Groups: Environmentalists, Dissatisfieds, Inactivators, Bystanders, Honeybees, Optimists, and Moderates", *Sustainability*, vol. 12, n° 20, p. 8370, oct. 2020, doi: 10.3390/su12208370.
- [2] L. Elgaaid, "Exploring the role of anticipated guilt on pro-environmental behavior – a suggested typology of residents in France based on their recycling patterns", *Journal of Consumer Marketing*, vol. 29, n° 5, pp. 369–377, jul. 2012, doi: 10.1108/07363761211247488.
- [3] J. Strieder Philippsen, F. H. Soares Angeoletto, y R. G. Santana, "El nivel de educación y los ingresos son importantes para la buena conciencia ambiental: un estudio de caso desde el sur de Brasil", *Ecol. Austral*, vol. 27, n° 1, pp. 039–044, feb. 2017, doi: 10.25260/EA.17.27.1.0.300.
- [4] I. N. de E. y Censos, "Hogares", Instituto Nacional de Estadística y Censos. <https://www.ecuadorencifras.gob.ec/hogares/> (accedido feb. 22, 2021).
- [5] G. Papadourakis et al., "Clustering analysis on questionnaire data for program accreditation", en 2008 19th EAEEIE Annual Conference, Tallinn, jun. 2008, pp. 76–81, doi: 10.1109/EAEEIE.2008.4610162.
- [6] D. Schuurman, K. De Moor, L. De Marez, y J. Van Looy, "Fanboys, competitors, escapists and time-killers: a typology based on gamers' motivations for playing video games", en Proceedings of the 3rd international conference on Digital Interactive Media in Entertainment and Arts - DIMEA '08, Athens, Greece, 2008, p. 46, doi: 10.1145/1413634.1413647.

- [7] T. Kodinariya y P. Makwana, "Review on Determining of Cluster in K-means Clustering", International Journal of Advance Research in Computer Science and Management Studies, vol. 1, pp. 90–95, ene. 2013.
- [8] "Practicing Clustering Techniques on Survey Dataset — by Tama Handika — Medium". <https://medium.com/@blazetamareborn/practicing-clustering-techniques-on-survey-dataset-f7d7a322e6ff> (accedido feb. 22, 2021).
- [9] M. Letelier, "Clustering With More Than Two Features? Try This To Explain Your Findings", Medium, jul. 23, 2020. <https://towardsdatascience.com/clustering-with-more-than-two-features-try-this-to-explain-your-findings-b053007d680a> (accedido feb. 22, 2021).
- [10] "Ecuador - Encuesta Anual de Edificaciones 2012, Permisos de Construcción - variable - V122". <https://anda.inec.gob.ec/anda/index.php/catalog/503/datafile/F2/V122> (accedido feb. 22, 2021).