# DETERMINACIÓN Y POST-PROCESAMIENTO DE PERFILES CON TÉCNICAS NO SUPERVISADAS DE MINERÍA DE DATOS COMO SOPORTE A LA TOMA DE DECISIONES EN LA ESTRATEGIA UNIDOS PARA LA SUPERACIÓN DE LA POBREZA EN COLOMBIA

**Andrés Eugenio Silva Monsalve**

**Escuela Colombiana de Ingeniería Julio Garavito**
**Decanatura de Ingeniería de Sistemas**
**Maestría Gestión de Información**
**Bogotá D.C., Colombia**
**2018**

# DETERMINACIÓN Y POST-PROCESAMIENTO DE PERFILES CON TÉCNICAS NO SUPERVISADAS DE MINERÍA DE DATOS COMO SOPORTE A LA TOMA DE DECISIONES EN LA ESTRATEGIA UNIDOS PARA LA SUPERACIÓN DE LA POBREZA EN COLOMBIA

**Andrés Eugenio Silva Monsalve**

Trabajo de investigación para optar al título de
Magíster en Gestión de Información

Director
PhD. Dante Conti

Co director
PhD. Victoria Eugenia Ospina Becerra

**Escuela Colombiana de Ingeniería Julio Garavito**
**Decanatura de Ingeniería de Sistemas**
**Maestría en Gestión de Información**
**Bogotá D.C., Colombia**
**2018**

Publicado en 2018 por la Escuela Colombiana de Ingeniería Julio Garavito. Avenida 13 No 205-59 Bogotá. Colombia
TEL: +57 – 1 668 36 00

## Agradecimientos

Sin ninguna duda soy lo que soy gracias a los esfuerzos de mi madre. Cada peldaño que alcanzo es para honrar su dedicación por mí.

**Resumen**

En la búsqueda de garantizar que los hogares más pobres y vulnerables puedan superar sus condiciones de vida, la Estrategia Red UNIDOS, consolida y gestiona información que permite conocer la situación de los hogares y personas, para la toma de decisiones relacionadas con la planeación y ejecución del proceso de acompañamiento. Este proyecto determina el valor agregado y potencial de la información mediante modelado basado en datos, específicamente clustering. Al respecto, se crea una vista minable donde se aplica clusterización jerárquica (método Ward, con distancia coseno) y se decide cortar los dendogramas en tres (3) y seis (6) grupos respectivamente, que posteriormente fueros postprocesados y validados para describir en detalle los perfiles y su interpretación usando los centroides de cada grupo. Finalmente, los perfiles hallados y plenamente descritos sirven como base para dar soporte a la toma de decisiones al proceso de acompañamiento de la Estrategia Red UNIDOS y son resumidos a su vez en dos (2) dashboards de resultados.

**Abstract**

In the search to guarantee that the poorest and most vulnerable households can overcome their living conditions, the UNIDOS Network Strategy consolidates and manages information that allows knowing the situation of households and people, for making decisions related to planning and execution of the accompaniment process. This project determines the aggregate and potential value of the information through modeling based on data, specifically clustering. In this regard, a minable view is created where hierarchical clustering is applied (Ward method, with cosine distance) and it is decided to cut the dendograms into three (3) and six (6) groups respectively, which were subsequently post-processed and validated to describe in detail the profiles and their interpretation using the centroids of each group. Finally, the profiles found and fully described serve as a basis to support the decision-making

process to accompany the UNIDOS Network Strategy and are summarized in two (2) results dashboards.

# Tabla de contenido

# Lista de Figuras

# Lista de Tablas

# 1 Introducción

## 1.1 Problemática (Justificación)

Colombia calcula la pobreza a partir de los ingresos monetarios de los hogares y a través del Índice de Pobreza Multidimensional (IPM), el cual, evalúa privaciones en la educación de la niñez y juventud, el trabajo, la salud, el acceso a servicios públicos y las condiciones de vivienda. La pobreza en 2015 se calcula en 10,3 por ciento de la población y el gobierno a través de diferentes programas brinda ofertas para suplir las necesidades básicas de los hogares en condición vulnerable y de esta forma disminuir las cifras (Administrativo, Estadística, & Nacional, 2014).

A través de la Ley 1785 de 2016, se establece la Red para la Superación de la Pobreza Extrema Red UNIDOS y se dictan otras disposiciones, dicha Red estará conformada por las entidades del Estado que presten servicios sociales dirigidos a la población en pobreza extrema, alcaldías y gobernaciones, el sector privado y organizaciones de la sociedad civil, que busca asegurar que los hogares más pobres y vulnerables puedan superar las condiciones que los mantienen en pobreza y consoliden sus capacidades para el desarrollo y el ejercicio de sus derechos (Congreso de la República de Colombia, 2016).

Para la implementación de la Estrategia Red UNIDOS, se consolida y se gestiona la información a través del Sistema de Información Misional, el cual, provee datos específicos, que permite conocer la situación de las comunidades, hogares y personas focalizadas, para la toma de decisiones relacionadas con la planeación, implementación y ejecución de la citada Estrategia.

Sin embargo, la gestión de información y la madurez para tomar decisiones dependen no solo del uso de tecnologías de punta, sino de la tendencia, filosofía o

cultura que la organización debe cultivar al interior y exterior para generar apropiación y valor cuantitativo y cualitativo sobre cada uno de los datos que hay en el ambiente empresarial y social, además de la pertinencia en los datos internos y externos que responden preguntas y generan valor para la articulación sector social.

Para Davenport, Acles y Prusak (1999), la información se ha convertido en la "moneda" clave de la organización, se ha vuelto demasiado valiosa para la mayoría de los gerentes como para regalar. Para que las organizaciones basadas en la información tengan éxito, las empresas deben aprovechar el poder de la política, es decir, permitir que las personas negocien el uso y la definición de la información, del mismo modo que negociamos el intercambio de otras monedas.

Es así como hoy se observan políticas, directrices, planes de gobierno o estratégicos, planes de acción, indicadores, estrategias de ventas y mapas de riesgos, con mejoras importantes no solo con antecedentes y experiencia del negocio sino con otras variables que permiten ser más asertivos en su funcionamiento de los procesos del negocio. Lo anterior con el fin de generar valor y sostenibilidad en el sector social y la población atendida.

Bajo las premisas anteriores, esta propuesta de investigación se focaliza en determinar valor agregado y potencial de la información mediante modelado y análisis de datos, de tal manera que se facilite el procesamiento e interpretación de ésta y así, dar soporte optimizado al proceso de toma de decisiones asociado a la evaluación y clasificación de perfiles en la estrategia UNIDOS para la superación de la pobreza en Colombia.

## Objetivos y Pregunta de Investigación

### 1.1.1 Objetivo general

Diseñar un modelo de datos basado en técnicas de aprendizaje no supervisado para detectar y etiquetar perfiles como soporte a la toma de decisiones en la Estrategia Unidos para la superación de la pobreza en Colombia.

### 1.1.2 Objetivos específicos

- Caracterizar las fuentes de información de la gestión del acompañamiento de la Estrategia Unidos.

- Identificar agentes y variables de la Estrategia Unidos para la toma de decisiones.

- Proponer métodos de análisis para la gestión del acompañamiento de la Estrategia Unidos.

- Representar un prototipo de modelado de datos y tableros de control que responda a la gestión del acompañamiento realizado a los hogares por la Estrategia Unidos basado en técnicas no supervisadas de aprendizaje de máquinas (clusterización).

- Efectuar el post-procesamiento de la información (interpretación de perfiles obtenidos con las técnicas no supervisadas) para dar mayor soporte y robustez a los tableros de visualización de resultados.

### 1.1.3 Pregunta

Por lo ya expuesto, este proyecto busca entonces responder a la pregunta ¿cómo usar técnicas de aprendizaje no supervisado (clusterización y/o mapas de autoorganización) y las herramientas de post-procesamiento y visualización para la

toma de decisiones en la Estrategia Unidos? y las siguientes preguntas específicas: ¿establecer los beneficios del perfilamiento de hogares a través de clústeres de información para mejorar la prestación del servicio a la población vulnerable en Colombia?, ¿establecer los beneficios de la visualización de información respecto al acompañamiento de los hogares en situación de vulnerabilidad?.

La respuesta a las preguntas formuladas contribuye a potenciar la gestión y las capacidades de decisión que realiza Prosperidad Social en el acompañamiento ofrecido por la Estrategia Unidos a los hogares pobres de Colombia.

## 1.2 Alcance y Limitaciones

1. La investigación explora el cálculo de veintiséis (26) logros familiares para los hogares acompañados por la Estrategia Unidos de Prosperidad Social en los departamentos de Caldas, Quindío y Risaralda de Colombia.

2. La investigación explora la información recolectada por Prosperidad Social durante los años 2016 y 2017.

3. La investigación se acota en revisar los estados de cumplimiento de logros Alcanzado y Por alcanzar que identifica las fortalezas y debilidades del hogar al momento de su caracterización.

4. La investigación aborda variables cuantificadas de la conformación de cada uno de los hogares objeto de estudio.

5. Dado que las variables objeto de análisis (veintiséis logros familiares) son producto de la recolección del formulario de caracterización desde la ubicación geográfica del hogar, la investigación no validará la veracidad del cálculo, las variables recolectadas en el formulario y las condiciones de vida de los hogares.

6. Los resultados arrojados por la investigación son admitidos para la toma de decisiones en los departamentos acotados. Sin embargo, la toma de decisiones en otra ubicación geográfica de Colombia es válida si se realiza el ejercicio de etiquetado como en los hogares analizados.

7. Para el uso y tratamiento de información es fundamental contar con el aval del dueño de los datos. En este caso, Prosperidad Social a través de la Dirección de Acompañamiento Familiar y Comunitario encargada del diseño e implementación de la Estrategia Unidos en Colombia.

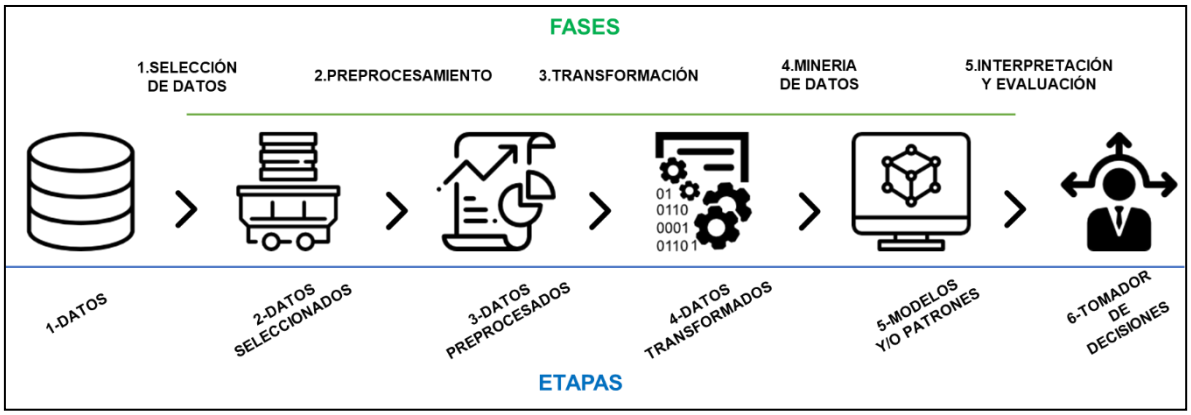
## 1.3 Metodología

El proyecto de investigación se basa en la metodología Knowledge Discovery in Databases (KDD) introducida por Fayyad en 1996 para resolver el problema.

### 1.3.1 Marco de Referencia KDD

Para Fayyad (1996). El KDD puede definirse como un proceso que busca a partir de datos y métodos de minería de datos, identificar patrones válidos, novedosos, potencialmente útiles y en última instancia comprensibles para describir conocimiento. En la figura 1-1 se especifican las etapas y fases del proceso KDD que permiten obtener conclusiones a partir de la correlación de información.

Figura 1-1 Etapas y Fases KDD




Fuente: Construcción propia

### 1.3.2 Fuentes de Información y Muestra

Este apartado de la investigación incluye el preprocesamiento de los datos mediante informes exploratorios y de estadística descriptiva que tiene como fuente de información las bases de datos de los hogares acompañados en Colombia que se encuentran en el sistema de información misional de la Estrategia Unidos de Prosperidad Social, caracterizadas por contar con un cogestor social, el cual realiza el acompañamiento. Estos hogares cuentan con información de ubicación actualizada en los últimos dos años, además de contar con el cálculo de los veintiséis (26) indicadores de pobreza denominados logros con detalle familiar e individual. La diversidad en información respecto al avance del acompañamiento y las situaciones en la condición de vida de cada hogar será útil para relacionar las variables objeto indispensable en la toma de decisiones.

Para la selección de la muestra se tomará como referente los hogares con acompañamiento familiar de la Estrategia Unidos en los departamentos de Caldas, Quindío y Risaralda que se encuentren registrados en el sistema de información

misional. Con base en los criterios anteriores el tamaño aproximado para el procesamiento de datos corresponde a 25.000 registros.

### 1.3.3 Variables a Evaluar

Con base en la fundamentación teórica antes presentada las siguientes son las variables para evaluar en este estudio:

- Estados de logros familiares de los hogares acompañados por la Estrategia Unidos durante septiembre de 2016 y abril de 2017.
- Distribución del acompañamiento en el territorio nacional colombiano, específicamente los departamentos de Caldas, Quindío y Risaralda.
- Distribución del hogar por sexo, edad, discapacidad.

### 1.3.4 Técnicas para la recolección de información

La técnica para la obtención de la información en este proyecto de investigación será a través de análisis de contenido de listados estructurados de las encuestas que previamente la Estrategia Unidos ha recolectado a corte abril de 2017, donde se relaciona la información correspondiente al levantamiento de información que realiza el cogestor social en el proceso de Ubicación y Caracterización de los hogares y personas acompañadas por la Estrategia Unidos y los indicadores de logros familiares para la gestión del acompañamiento.

### 1.3.5 Procedimiento para el trabajo de campo

El desarrollo del trabajo de campo para este estudio se realizará de la siguiente forma:

- Obtener el total de hogares acompañados a corte abril de 2017, en formato plano con variables de identificación de hogares y personas anonimizadas.

- Establecer contacto con los tomadores de decisiones y a través de la observación y experiencia, identificar la zona geográfica del territorio nacional que sea representativa para la operación y tenga diferentes modelos de acompañamiento (urbano y rural) además de condiciones complejas para gestión del acompañamiento de la Estrategia Unidos.

- Diseñar y ajustar el modelo de datos para la correlación de información de los datos previamente extraídos de la fuente de información.

- Aplicación de herramientas técnicas para realizar el análisis descriptivo del conglomerado de información.

- Diseñar las bases de datos minables de la bodega de datos.

- Aplicación de clusterización jerárquica, técnicas no supervisadas y etiquetado a la base de datos minable, para la construcción de clústeres de información.

- Interpretación de etiquetados, construcción de perfiles y herramienta de visualización para la toma de decisiones.

### 1.3.6 Aspectos éticos del estudio

El tratamiento de las personas participantes del estudio (director y codirector) y el manejo de la información estarán enmarcados dentro de los cánones de la ética investigativa y gozarán de la confidencialidad y el respeto que ello merece. Será obligación inherente a los investigadores guardar prudencia y acatamiento a la normatividad ética investigativa.

Al respecto, el Departamento Administrativo para la Prosperidad Social a través de respuesta escrita el 12 de octubre de 2017 con número de radicado 20172201296511, autoriza el uso de datos de la Estrategia Unidos y se firma entre las partes el Acuerdo de Confidencialidad de información con terceros.

### 1.3.7 Presentación del documento

El documento se presenta en seis capítulos, de los cuales el capítulo 2 corresponde a la fundamentación teórica en la cual se detalla aspectos técnicos de pobreza, las iniciativas de medición de pobreza y las metodologías más utilizadas en el mundo con énfasis en Colombia. Adicional, detalla conceptos técnicos que se utilizarán en la investigación como minería de datos, clusterización, métodos jerárquicos y técnicas no supervisadas. El capítulo 3, corresponde al análisis descriptivo, en el cual, se verifica el estado actual de los hogares, su composición demográfica, la distribución de los hogares en los veintiséis indicadores de pobreza, la propuesta de agrupamientos, la propuesta de variables score por dimensiones de pobreza y se construye el primer dashboard para visualizar e interactuar con la información. El capítulo 4, corresponde al análisis exploratorio a través de clusterización jerárquica (método Ward, con distancia coseno), se describen e interpretan los perfiles de cada grupo identificado y se construye el segundo dashboard con el agrupamiento sugerido para la interpretación. En el capítulo 5 se puntualizan los principales resultados de acuerdo con las fases detalladas en la metodología KDD. Y finalmente en el capítulo 6, se realizan las conclusiones, recomendaciones y se describen los trabajos futuros.

# 2 Fundamentación teórica

Este capítulo se estructura en tres pilares para el desarrollo de la investigación. El primero (numeral 2.1) relacionado con los procesos de negocio a intervenir, en donde se identifica y entrega conceptos para comprender la pobreza, las metodologías para realizar mediciones y el entendimiento de los procesos internos de la Estrategia Red UNIDOS para la superación de la pobreza en Colombia. El segundo (numeral 2.2) relacionado con prácticas y conjuntos de herramientas tecnológicas de minería de datos para la intervención y búsqueda de perfiles objeto de investigación. Y el tercero (numeral 2.3) relacionado con antecedentes o intervenciones anteriores con técnicas de minería de datos a problemas de pobreza y es quien otorga importancia al ejercicio propuesto.

## 2.1 Pobreza

De acuerdo con Anand & Sen (2000), el primer requisito para conceptuar la pobreza es tener un criterio que permita definir quién debe estar en el centro de nuestro interés. Especificar algunas “normas de consumo” o una “línea de pobreza” puede abrir parte de la tarea, los pobres son aquellos cuyos niveles de consumo caen por debajo de estas normas, o cuyos ingresos están por debajo de esa línea. La pobreza es un mundo complejo y complicado que requiere un análisis claro para descubrir todas sus dimensiones.

### 2.1.1 Iniciativas de medición de Pobreza en el mundo

Para Alkire & Foster (2007), actualmente existen tres enfoques principales para identificar a los pobres en un contexto multidimensional. El primer enfoque es el ‘unidimensional’, a través del cual se combinan los distintos indicadores de

bienestar en una sola variable agregada y una persona es identificada como pobre cuando la variable cae debajo de una determinada línea de corte.

Al respecto, Sen (1976), define los criterios para distinguir a las personas pobres de las no pobres, y la 'agregación', mediante la cual se reúnen los datos sobre las personas pobres para crear un indicador general de pobreza.

Para Alkire et al (2007), este método de identificación toma en cuenta las privaciones dimensionales, pero sólo en tanto que afectan al indicador agregado. Existe un margen mínimo para evaluar las privaciones dimensionales en sí mismas, lo cual a menudo es visto como una característica esencial para un enfoque multidimensional.

Para Alkire et al (2007), el segundo es el enfoque de 'unión', que considera a una persona que sufre privaciones en una sola dimensión como pobre en el sentido multidimensional. Generalmente se reconoce que este enfoque es excesivamente inclusivo y puede llegar a generar estimaciones exageradas de la pobreza. El tercer enfoque principal es el método de la 'intersección', que exige que una persona sufra privaciones en todas las dimensiones para ser identificada como pobre.

### 2.1.2 Iniciativas de medición de Pobreza en Colombia

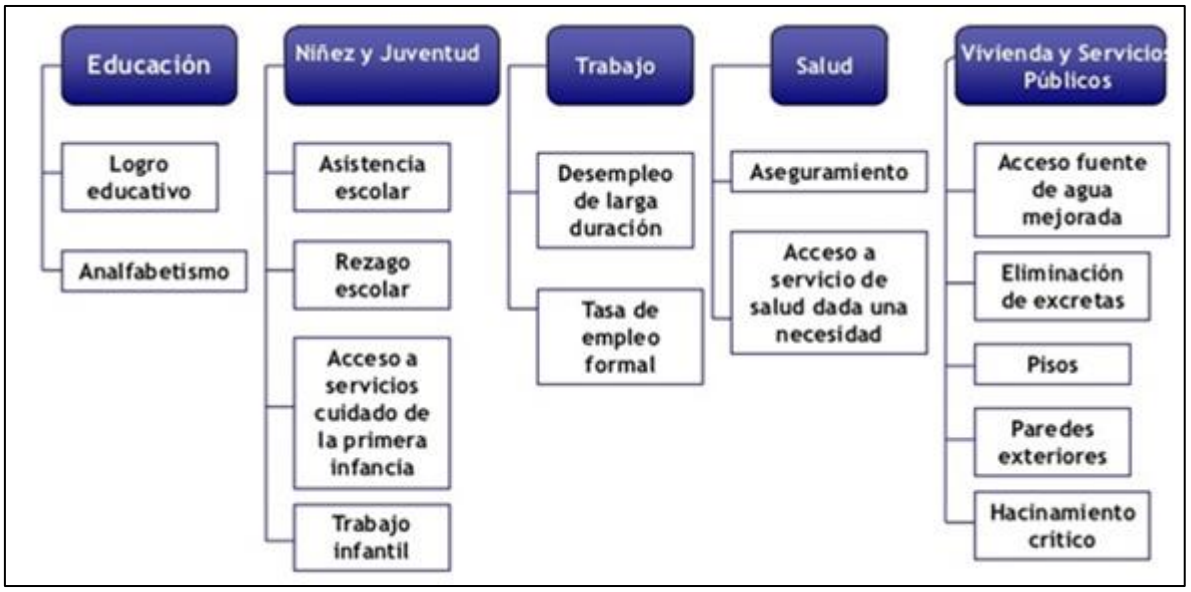
Para Angulo, Díaz & Pardo (2011), el Departamento Nacional de Planeación permite utilizar la metodología Alkire y Foster (2007) como instrumento de seguimiento y monitoreo de las acciones de política pública para la reducción de la pobreza. La nitidez de la noción multidimensional de la pobreza que expresa el indicador se transmite a la discusión multisectorial del diseño y la planeación de estrategias para la reducción de la pobreza. Es posible generar alarmas a partir de

la identificación de las dimensiones y variables con mayor privación, las más rezagadas y las más críticas en la población pobre.

Para Prosperidad Social, (2015) Colombia cuenta con dos (2) metodologías oficiales de medición de pobreza: Pobreza Monetaria y el Índice de Pobreza Multidimensional –IPM Colombia– (CONPES Social 150, 2012). Cada una de estas metodologías aborda aspectos distintos pero complementarios del fenómeno de la pobreza y son los indicadores principales para el seguimiento a las políticas sociales incluidas en el último Plan Nacional de Desarrollo.

Para Prosperidad Social, (2015) en la guía del acompañamiento familiar, en el país también se ha reconocido la pobreza como un fenómeno multidimensional, que aborda no solo aspectos monetarios, sino que tiene en cuenta la situación de los hogares en 15 indicadores de bienestar relacionados con: i) educación; ii) niñez y la juventud; iii) trabajo; iv) salud; y v) servicios públicos domiciliarios. La pobreza multidimensional es entendida como una situación en la cual un hogar tiene alrededor de cinco (5) carencias en quince (15) indicadores seleccionados (ver Figura 2-1). En el PND 2014-2018 se estableció como meta la reducción de la pobreza multidimensional de 24,8% en 2013 a 17,8% de 2018. Estos dos enfoques en la conceptualización de la pobreza implican reconocer tanto la perspectiva monetaria como multidimensional en el proceso de superación de la pobreza.

Figura 2-1 Variables del Índice de Pobreza Multidimensional para Colombia




Fuente: CONPES 150 de 2012

Al respecto, el Departamento Administrativo Nacional de Estadística en Colombia (2018), La medición de la pobreza se hace tradicionalmente de forma directa e indirecta, siguiendo la clasificación de Amartya Sen (1981). El método directo evalúa los resultados de satisfacción (o no privación) que tiene un individuo respecto a ciertas características que se consideran vitales como salud, educación, empleo, entre otras. En Colombia se realiza la medición directa por medio del Índice de Pobreza Multidimensional (IPM). Por otra parte, el método indirecto busca evaluar la capacidad adquisitiva de los hogares respecto a una canasta, para esto observa su ingreso, el cual es un medio y no un fin para lograr la satisfacción.

Para el Consejo Nacional de Política Económica y Social de Colombia (2012), Las personas se clasifican como pobres si su ingreso promedio al mes es inferior al valor de la línea de pobreza. De forma equivalente, una persona se identifica como pobre extremo si su ingreso promedio al mes es inferior al valor de la línea de indigencia. La incidencia de la pobreza (pobreza extrema) es el porcentaje de

personas identificadas como pobres (pobres extremas). Ambas incidencias tienen como denominador la población total del país.

### 2.1.3 Metodología Estrategia Unidos

Para Prosperidad Social et al, (2015), la Estrategia de Acompañamiento Familiar y Comunitario, Estrategia UNIDOS, es el pilar de un conjunto de acciones del Gobierno Nacional que apuntan al cumplimiento de la meta de una Colombia sin pobreza extrema. Esta estrategia de intervención se sustenta en la articulación del trabajo de entidades públicas y privadas, así como la unión de esfuerzos de los distintos niveles del Estado, encaminados a la generación de sinergias en torno a la población que más necesita del apoyo del Estado.

Para Prosperidad Social, (2015), para garantizar el Acompañamiento Familiar y Comunitario, y el acceso a los servicios Sociales del Estado dirigido a los hogares en pobreza extrema previstos por la Ley 1785 de 2016, Prosperidad Social implementa, como parte de su oferta social, la Estrategia UNIDOS, brinda Un Acompañamiento Familiar a los hogares en pobreza extrema a fin de lograr que cada uno de ellos reconozca sus fortalezas y potencialidades, consolide sus vínculos familiares, sus redes de interacción social, adquiera o afiance habilidades sociales y acceda a la oferta de bienes y servicios institucionales, para superar su situación de pobreza. Este acompañamiento tiene metodologías y tiempos definidos orientados a que los beneficiarios alcancen condiciones mínimas de vida.

Al respecto Prosperidad Social et al, (2015), el rol principal del Acompañamiento Familiar lo cumplen los Cogestores Sociales, quienes deben implementar la metodología que se ha diseñado para apoyar a los hogares en la identificación de los Logros Familiares por alcanzar y en las actividades que les permitan gestionarlos en el período de intervención de la Estrategia UNIDOS. El rol del Cogestor es motivar los hogares para que movilicen sus propios recursos, sus

conocimientos, sus habilidades y estimular en el hogar la toma de decisiones para la consecución de los Logros Familiares establecidos en el Plan del Hogar.

## 2.2 Minería de datos

Como concepto para Microsoft (2018), la minería de datos es el proceso de detectar la información procesable de los conjuntos grandes de datos. Utiliza el análisis matemático para deducir los patrones y tendencias que existen en los datos. Normalmente, estos patrones no se pueden detectar mediante la exploración tradicional de los datos porque las relaciones son demasiado complejas o porque hay demasiados datos.

### 2.2.1 Técnicas no supervisadas

De acuerdo con Moreno & Lopez (2018), las técnicas de minería de datos no supervisadas, también conocidas con el nombre de técnicas de descubrimiento del conocimiento, se utilizan para la detección de patrones ocultos en bases de datos de gran tamaño. Dichos patrones representan por sí mismos información útil que puede ser utilizada directamente en la toma de decisiones. Trabajos recientes muestran que los algoritmos no supervisados pueden utilizarse con éxito para resolver problemas de clasificación.

Al respecto Amat (2017), las particiones se establecen de forma que, las observaciones que están dentro de un mismo grupo son similares entre ellas y distintas a las observaciones de otros grupos. Se trata de un método no supervisado, ya que el proceso ignora la variable respuesta que indica a que grupo pertenece realmente cada observación (si es que existe tal variable).

Para Padmanabhan, & Tuzhilin (1999), el objetivo de esas propuestas es la obtención de un número reducido de reglas con valores altos del factor de soporte

y de confianza. Además de esos factores, para determinar el interés de una regla se pueden usar medidas subjetivas como la incertidumbre y la accionabilidad

Las técnicas no supervisadas se pueden diferenciar en tres grupos de métodos. Para Amat et al, (2017), existen métodos de particionamiento de clustering, métodos de agrupación jerárquica, y métodos que combinan o modifican los anteriores.

Al respecto, para Foster & Fawcett (2013). los métodos jerárquicos se dividen en:

> "***Los métodos jerárquicos aglomerativos*** *comienzan con tantos clusters como objetos tengamos que clasificar y en cada paso se recalculan las distancias entre los grupos existentes y se unen los dos grupos más similares o menos disimilares. El algoritmo acaba con un clúster conteniendo todos los elementos.*
>
> ***Los métodos jerárquicos divisivos*** *comienzan con un clúster que engloba a todos los elementos y en cada paso se divide el grupo más heterogéneo. El algoritmo acaba con tantos clusters (de un elemento cada uno) como objetos se hayan clasificado.*"

Algunos de los métodos jerárquicos aglomerativos para Moreno (2018) son método del enlace simple, método de enlace completo, método del promedio entre grupos, método del centroide, método de la mediana y método de Ward.

### 2.2.2 Modelo de datos y Clusterización

Para Han, Kamber & Pei. (2012), las técnicas de agrupamiento consideran las tuplas de datos como objetos. Particionan los objetos en grupos, o clusters, para que los objetos dentro de un clúster sean "similares" entre sí y "disímiles" a objetos en otros

grupos. La similitud se define comúnmente en términos de cómo "Cerrar" los objetos están en el espacio, en función de una función de distancia. La "calidad" de un clúster puede representarse por su diámetro, la distancia máxima entre dos objetos cualquiera en el grupo.

Para De La Fuente et al (2011), el Análisis Cluster, conocido como Análisis de Conglomerados, es una técnica estadística multivariante que busca agrupar elementos (o variables) tratando de lograr la máxima homogeneidad en cada grupo y la mayor diferencia entre los grupos.

Según Amat et al (2015):

> *"El clustering basado en modelos considera que las observaciones proceden de una distribución que es a su vez una combinación de dos o más componentes (clusters), cada uno con una distribución propia. En principio, cada cluster puede estar descrito por cualquier función de densidad, pero normalmente se asume que siguen una distribución multivariante normal. Para estimar los parámetros que definen la función de distribución de cada clúster (media y matriz de covarianza si se asume que son de tipo normal) se recurre al algoritmo de Expectation-Maximization (EM). Este resuelve distintos modelos en los que el volumen, forma y orientación de las distribuciones pueden considerarse iguales para todos los clústeres o distintas para cada uno*."

Para Hahn & Packowski (2015), la minería de datos y el modelado construyen la base metodológica para casos de uso reactivo de sentido y respuesta. Ayudan a descubrir conocimientos y patrones que se pueden traducir en reglas comerciales para (Semiautomáticas) a los eventos empresariales predefinidos. El modelado de datos implica el amplio conjunto de métodos estadísticos multivariados.

Para Foster et al (2013), los métodos jerárquicos tienen por objetivo agrupar clústeres para formar un nuevo o bien separar alguno ya existente para dar origen a otros dos, de tal forma que, si sucesivamente se va efectuando este proceso de aglomeración o división, se minimice alguna distancia o bien se maximice alguna medida de similitud.

Para Han, et al. (2012), los métodos de transformación de datos, en este paso de preprocesamiento, los datos se transforman o consolidan para que el proceso de minería resultante pueda ser más eficiente, y los patrones encontrados pueden ser más fáciles de entender.

Para Lizma & Boccado (2014), el método de coeficiente de correlación de Spearman, muestra una asociación entre variables. Permite obtener un coeficiente de asociación ente variables que no se comportan normalmente, entre variables ordinales. Se calcula con base a una serie de rangos asignados. los valores van de 1 a 1, siendo 0 el valor que indica no correlación, y los signos indican correlación directa e inversa.

Para Han, et al. (2012), el método de Ward es un procedimiento jerárquico en el cual, en cada etapa, se unen los dos clústeres para los cuales se tenga el menor incremento en el valor total de la suma de los cuadrados de las diferencias, dentro de cada clister, de cada individuo al centroide del clúster.

Al respecto Foster et al. (2013), el objetivo del método de Ward es encontrar en cada etapa aquellos dos clústeres cuya unión proporcione el menor incremento en la suma total de errores.

Para Santi, Aloise, & Blanchard (2016), los algoritmos de clústeres determinan grupos de objetos de una manera en que los objetos en el mismo grupo, llamados clústeres, son más similares entre sí que a los de otros grupos.). La

agrupación es omnipresente, con aplicaciones en las ciencias naturales, psicología, medicina, ingeniería, economía, comercialización y otros campos.

Para Han, et al. (2012), los atributos nominales tienen un finito (pero posiblemente grande) número de valores distintos, sin ordenar entre los valores. La definición manual de jerarquías conceptuales puede ser una tarea tediosa y que requiere mucho tiempo para un usuario o un experto de dominio. Afortunadamente, muchas jerarquías están implícitas en el esquema de base de datos y se puede definir automáticamente en el nivel de definición de esquema. Las jerarquías conceptuales se pueden usar para transformar los datos en múltiples niveles de granularidad.

A este respecto Cleofas-Sánchez, Sánchez, García, & Valdovinos (2016), desde un punto de vista práctico, la clasificación se refiere a la asignación- de un conjunto finito de muestras a clases predefinidas basadas en un número de variables o atributos observados. El diseño efectivo de los clasificadores son una tarea compleja donde la calidad subyacente de los datos se vuelve crítico para lograr aún más clasificaciones precisas.

Para Benati, Puerto, & Rodríguez-Chía (2017), el problema de la agrupación consiste en descubrir o detectar cómo se divide una población en dos o más subgrupos, cada uno de los subgrupos especificados por distintas características. El resultado típico de un algoritmo de agrupamiento es la asignación de observaciones a los grupos.

A este respecto, Santi et al, (2016), el modelo tiene como objetivo dividir objetos en grupos de forma tal que la suma de las distancias de cada objeto al ejemplar central de su clúster (es decir, mediana) es mínima. N objetos dados a ser agrupados y un número conocido de clústeres.

### 2.2.3 Distancia por similitud

Todos los métodos de clusterización no supervisados tienen en común que para llevar a cabo agrupaciones necesita definir y cuantificar la similitud. Al respecto, para De la Fuente (2011) poder unir variables o individuos es necesario tener algunas medidas numéricas que caractericen las relaciones entre las variables o los individuos. Cada medida refleja asociación en un sentido particular y es necesario elegir una medida apropiada para el problema concreto que se esté tratando. La medida de asociación puede ser una distancia o una similaridad.

Respecto a medidas de distancia, para Amat et al (2017), el término distancia se emplea dentro del contexto del clustering como cuantificación de la similitud o diferencia entre observaciones. Si se representan las observaciones en un espacio p dimensional, siendo p el número de variables asociadas a cada observación, cuando más se asemejen dos observaciones más próximas estarán, de ahí que se emplee el término distancia. Algunas de las más utilizadas son:

La distancia Coseno, para Amat et al, (2017), el coseno del ángulo que forman dos vectores puede interpretarse como una medida de similitud de sus orientaciones, independientemente de sus magnitudes. Si dos vectores tienen exactamente la misma orientación (el ángulo que forman es 0º) su coseno toma el valor de 1, si son perpendiculares (forman un ángulo de 90º) su coseno es 0 y si tienen orientaciones opuestas (ángulo de 180º) su coseno es de -1. La fórmula es

$$\cos(\alpha) = \frac{X \times Y}{\|X\| \, \|Y\|}$$

La distancia euclídea, para Amat et al, (2017), entre dos puntos p y q se define como la longitud del segmento que une ambos puntos. En coordenadas cartesianas, la distancia euclídea se calcula empleando el teorema de Pitágoras, en

el que cada punto está definido por las coordenadas (x, y), la distancia euclídea entre p y q viene dada por la ecuación

$$d_{euc}(p,q) = \sqrt{\left(x_p - x_q\right)^2 + \left(y_p - y_q\right)^2}$$

La distancia Manhattan, para Amat et al, (2017), define la distancia entre dos puntos p y q como el sumatorio de las diferencias absolutas entre cada dimensión. Esta medida se ve menos afectada por outliers (es más robusta) que la distancia euclídea debido a que no eleva al cuadrado las diferencias.

$$d_{man}(p,q) = \sum_{i=1}^{n} |(p_i - q_i)|$$

### 2.2.4 Herramientas para la toma de decisiones

De acuerdo con Foreman (2013) y Antoniadis, Tsiakiris, & Tsopogloy (2015), las organizaciones llevan a cabo negocios y operaciones, mejorando la transparencia financiera, la comercialización y los servicios al cliente, la cadena de suministro y la gestión de operaciones, gestión de recursos humanos, integrando todos los recursos e información en una sola plataforma.

Para Silahtaroğlu & Alayoglu (2016), las herramientas y los conceptos relacionados con los sistemas de información (IS) tienen la máxima importancia cuando se trata de hacer planes estratégicos y tomar decisiones en la vida empresarial. Hoy en día, las empresas invierten en datos para extraer información y utilizarlo para fines operativos, tácticos y estratégicos.

Así, para Iglesias, Tiemblo, Ledezma, & Sanchis (2016), la sociedad moderna genera enormes cantidades de información cada día, especialmente en formato

digital, que obstruyen el almacenamiento y procesamiento y análisis posterior. Los grandes datos se pueden definir como una escala de conjunto de datos que va más allá de la herramienta de gestión de base de datos, aun cuando existen capacidades de recopilación, almacenamiento, administración y análisis de datos.

Así, para Tello et al. (2016), el reto se perfila en la necesidad de lograr una mayor eficiencia y eficacia en sus procesos de producción, principalmente si estos se sustentan en los activos intangibles que son generados a partir de una estrategia de desarrollo basada en el conocimiento.

De acuerdo con Sensuse et al. (2015), la gestión del conocimiento de procesos se define como actividades de identificación, creación, captura, intercambio y transferencia de conocimiento para mejorar el desempeño organizacional. Estas actividades son prácticas tanto horizontales como Vertical en la organización.

A este respecto Tello et al. (2016), en las organizaciones se presenta una situación que versa en el manejo de la información, que cada vez es más numerosa y difícil de categorizar. La competitividad de las empresas considera importante que estas aprendan y que con el tiempo puedan replicar el conocimiento que se concentra en ellas a partir de los diferentes agentes involucrados en su operación, pudiendo ser empleados, directivos, clientes, etc.

Para Silahtaroğlu & Alayoglu (2016),

> *"Las principales gerencias de las empresas confían en tableros, gráficos, tablas, números y estadísticas cuando toman sus decisiones finales. Todos estos valores, simplemente hablando, provienen de sistemas de información como sistemas de información de gestión (MIS), decisión sistemas de apoyo (DSS) y sistemas de información ejecutiva (EIS). Estos sistemas procesan datos para generar o extraer información y, finalmente, presentarlo a la alta dirección en*

*forma de tablas, cuadros o cuadros de mando. Ellos sirven en diferentes niveles de toma de decisiones. Estos niveles son operativos, tácticos y estratégicos"* (p. 4).

A este respecto Gröger, Hillmann, Hahn, Mitschang, & Westkämper (2013), los paneles de control se basan principalmente en estadísticas y la presentación de informes con los servicios básicos de alerta.

Para Wang (2016), las percepciones de los usuarios sobre los atributos de diseño se convierten en grados cuantitativos de satisfacción del cliente para implicaciones gerenciales.

Para Larson & Chang (2016) y Foreman (2013), la visualización, si bien no es un concepto nuevo, es un componente clave De análisis rápido. La visualización como parte de un servicio permite los usuarios a comprender rápidamente conjuntos de datos complejos creados a partir análisis o modelos analíticos.

Al respecto Hahn & Packowski (2015), análisis de negocios y conceptos relacionados que describen el análisis de datos empresariales con fines de toma de decisiones han recibido atención tanto en las empresas e instituciones académicas. La alta gerencia generalmente ven el análisis de negocios como un factor diferenciador y una ventaja competitiva y, por lo tanto, están cada vez más interesados en este valor potencial.

Según Holsapple et al. (2014), las razones para análisis de negocios:

- Lograr una ventaja competitiva.
- El apoyo de los objetivos estratégicos y tácticos de la organización.
- Mejor rendimiento de la organización.
- Mejores resultados de decisiones y procesos para la toma.
- La producción de conocimientos.

- La obtención de valor a partir de los datos

## 2.3 Minería de datos para problemas de pobreza

Para Bulos, Delfino & Rivera (2014), la extracción de patrones ocultos de información de grandes bases de datos, más allá de regresión, que permitirá la generación de una predicción sobre la dirección y el alcance del cambio en el estado de una muestra. Después de determinar estos patrones ocultos, la aplicación de la gamificación puede ser utilizado como un mecanismo de cambio de comportamiento particularmente para las personas que están predispuestas participar en juegos sobre problemas basados en la realidad.

Para Pluliková (2015), la pobreza representa un tema importante en la práctica del desarrollo. En este caso se intentan para entender, medir y predecir este fenómeno a través de modelos que efectivamente predigan los niveles de pobreza. Para lograr esta tarea, también se presenta una forma de medir la pobreza a través de un índice de pobreza. Los datos que utilizamos provienen de Indonesia y, por lo tanto, los resultados reflejan la condición en este país.

Para Tarozzi & Deaton (2008), en los últimos años han visto un uso generalizado de mapas de pobreza de áreas pequeñas basados en datos censales enriquecidos por relaciones estimadas a partir de encuestas de hogares que predicen variables no cubiertas por el censo. Estos métodos son usados para realizar estimaciones de pobreza y desigualdad para áreas tan pequeñas como 20.000 hogares. El estudio argumenta para unir de manera útil los datos de encuestas y censos que requiere un grado de homogeneidad espacial para el cual el método no proporciona ninguna base, y que es improbable que esté satisfecho en la práctica. Se utiliza datos del censo de 2000 de México para construir "encuestas de hogares" sintéticas y para simular el proceso de mapeo de la pobreza.

Para Coromaldy & Zoli (2007), la medición de pobreza convencional se basa en los ingresos de los hogares, y muestran que la pobreza y la desigualdad han aumentado en Italia a través de los últimos quince años. El principal inconveniente de este método es que lo hace no incluir otras variables no monetarias relevantes para definir las necesidades de los hogares. En la actualidad el mundo acordó que la pobreza debería conceptualizarse como un fenómeno multidimensional, más relacionado con el nivel de vida de la persona o el hogar que a la simple incapacidad de satisfacer la subsistencia básica necesariamente. Se propone medir la pobreza en Italia complementando la información de ingresos con indicadores no monetarios. Con este fin, se realiza un análisis multidimensional de la pobreza utilizando una muestra representativa basada en la primera ola (2004) del componente italiano de las estadísticas europeas sobre ingresos y condiciones de vida.

Al respecto Shirvanian & Bakhshoodeh (2012), en este estudio, la pobreza rural en Irán se investiga aplicando un enfoque multidimensional, la técnica de minería de reglas de asociación y las pruebas de Levine, F y Tukey a los datos de hogares de 2008. Los resultados indican que la pobreza en sus multidimensiones es un problema epidémico en el Irán rural. Los resultados también muestran que hay 11 patrones de pobreza en las áreas rurales, que incluyen cuatro patrones principales con una cobertura del 99.62% y siete subpatrones con una cobertura de casi el 0.38%. En estos patrones, la vivienda y la educación del hogar son las dimensiones más importantes de la pobreza y la pobreza de ingresos es la dimensión menos importante.

# 3 Análisis Descriptivo

Este capítulo presenta una visión global del conjunto de datos de los hogares pobres que son objeto del análisis, en el cual, se detalla la distribución de los integrantes que conforman los hogares por quinquenio y sexo. Se destaca la distribución porcentual por estado de cada uno de los logros que calcula la Estrategia Red UNIDOS previa caracterización de los hogares pobres. Es importante acotar que la línea de investigación es la búsqueda de agrupamientos de acuerdo con los comportamientos de la información. Para el caso descriptivo, este agrupamiento se realiza desde la perspectiva del negocio y se concluye el capítulo con una propuesta de medición de hogares de acuerdo con la dimensión de logros y la interpretación de los grupos producto del ejercicio.

Para iniciar el ejercicio se toma como fuente de información la base de datos que comprende una población de noventa y ocho mil novecientos cincuenta y un (98.951) personas agrupadas en veinticinco mil cuatrocientos sesenta y cuatro (25.464) hogares, acompañados durante los años 2016 y 2017 en los departamentos de Caldas, Quindío y Risaralda (eje cafetero). A continuación, se detalla el análisis a través de tres agrupamientos objetivo, así:

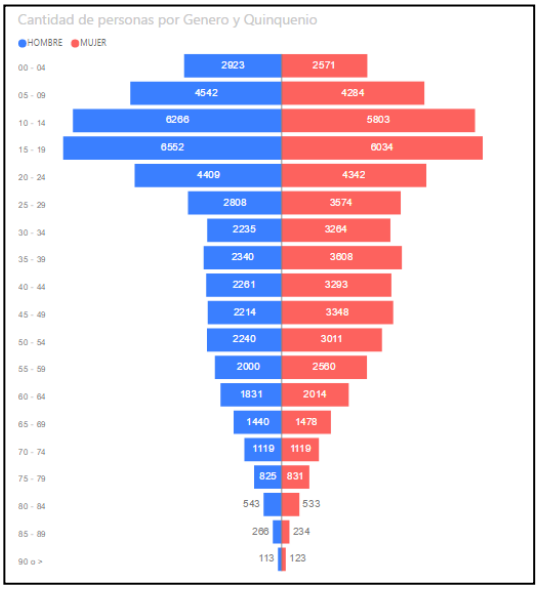
1. Pirámide poblacional por edad y sexo de las personas acompañadas.
2. Indicadores familiares de los hogares acompañados.
3. Indicadores individuales de las personas acompañadas.

## 3.1 Pirámide poblacional analizada.

Población muy homogénea por sexo, y con altas concentraciones en los grupos de adolescencia y adulto joven al desagregarse en grupos quinquenales. El comportamiento en los grupos quinquenales en la población pobre analizada, tiene

una alta similitud con la pirámide poblacional del censo 2015 realizado por el Departamento Administrativo Nacional de Estadística (DANE), para cada uno de los departamentos, en donde se visualiza también altas concentraciones de personas en la población adolescente y joven.

Figura 3-1 Cantidad de personas por género y quinquenio




Fuente: Construcción propia

La Estrategia Unidos establece para un hogar identificado como pobre, veintiséis (26) indicadores o escenarios que deben predominar para mejorar su condición, distribuidos en cinco (5) dimensiones.

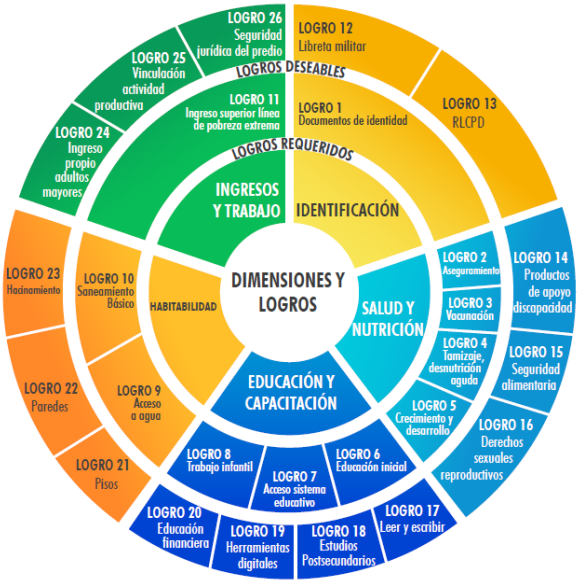
A través de algoritmos por cada indicador y con la información previamente recolectada en el formulario de caracterización, se calcula y asigna un valor a cada indicador (Alcanzado, Por Alcanzar o No Aplica), el cual, permite una acción o gestión particular por parte de la Estrategia o sus aliados y permite cuantificar el avance del hogar en términos de superación de pobreza.

## 3.2 Estudio de los Indicadores Familiares.

La Estrategia Unidos establece para un hogar identificado como pobre, veintiséis (26) indicadores o escenarios que deben predominar para mejorar su condición, distribuidos en cinco (5) dimensiones.

A través de algoritmos por cada indicador y con la información previamente recolectada en el formulario de caracterización, se calcula y asigna un valor a cada indicador (Alcanzado, Por Alcanzar o No Aplica), el cual, permite una acción o gestión particular por parte de la Estrategia o sus aliados y permite cuantificar el avance del hogar en términos de superación de pobreza.

Figura 3-2 Distribución dimensión y logros




Fuente: Dirección de Acompañamiento Familiar y Comunitario. Prosperidad Social

Para el análisis de correlación (método Spearman) de los indicadores se homogeniza las variables tomando únicamente la frecuencia del estado Por Alcanzar (déficit del hogar) por municipio. Al respecto, se evidencian varias

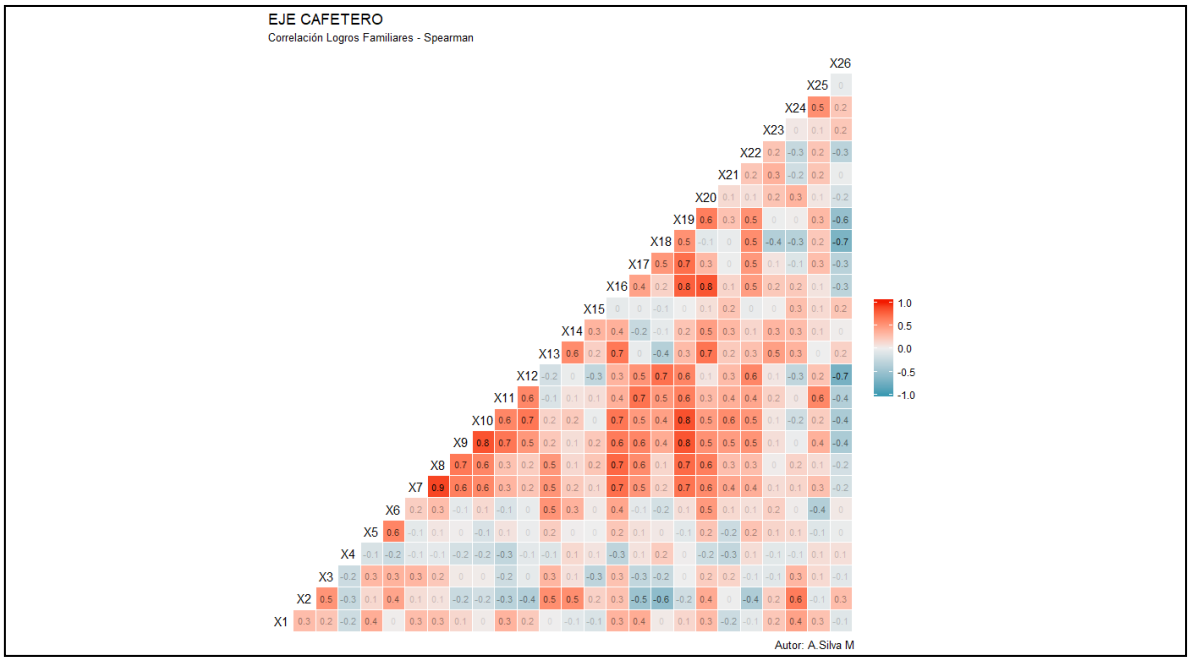
relaciones positivas y se agrupan los indicadores con mayor peso en tres (3) grupos. **A** (Indicador 7 y 8), **B** (Indicador 9, 10 y 19), **C** (Indicador 16, 19 y 20). En consecuencia, los grupos A y B tienen indicadores que son requeridos para el proceso de superación de pobreza. En el caso **A-** "**Educación infantil"**, un menor de edad sin acceso al sistema educativo es más propenso a ejercer trabajo infantil. En el caso **B- "Acceso a servicios",** un hogar con acceso a servicios básicos como agua y saneamiento básico tiene mayor probabilidad de acceder a herramientas digitales. Y en el caso **C- "Responsabilidad",** en donde una persona con educación en derechos sexuales y reproductivos también tienen educación financiera y utilización de herramientas digitales. Es importante indicar que también se visualizan relaciones con peso negativo que para el caso no son clasificadas dado que no representan valor en el negocio. Estas relaciones negativas son (Indicadores 18 y 26), (Indicadores 12 y 26), (Indicadores 19 y 26), (Indicadores 2 y 18).

Figura 3-3 Mapa de Calor - Correlación logros por Spearman

| | LOGRO_01 | LOGRO_02 | LOGRO_03 | LOGRO_04 | LOGRO_05 | LOGRO_06 | LOGRO_07 | LOGRO_08 | LOGRO_09 | LOGRO_10 | LOGRO_11 | LOGRO_12 | LOGRO_13 | LOGRO_14 | LOGRO_15 | LOGRO_16 | LOGRO_17 | LOGRO_18 | LOGRO_19 | LOGRO_20 | LOGRO_21 | LOGRO_22 | LOGRO_23 | LOGRO_24 | LOGRO_25 | LOGRO_26 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| LOGRO_01 | 1,000 | 0,288 | 0,242 | -0,208 | 0,374 | 0,014 | 0,281 | 0,274 | 0,132 | 0,023 | 0,339 | 0,211 | 0,005 | -0,074 | -0,133 | 0,337 | 0,381 | 0,032 | 0,137 | 0,296 | -0,219 | -0,058 | 0,226 | 0,433 | 0,256 | -0,091 |
| LOGRO_02 | 0,288 | 1,000 | 0,512 | -0,257 | 0,129 | 0,421 | 0,098 | 0,072 | -0,230 | -0,184 | -0,328 | -0,370 | 0,546 | 0,546 | 0,167 | 0,261 | -0,505 | -0,586 | -0,195 | 0,391 | -0,009 | -0,395 | 0,184 | 0,637 | -0,074 | 0,283 |
| LOGRO_03 | 0,242 | 0,512 | 1,000 | -0,217 | 0,310 | 0,342 | 0,321 | 0,219 | 0,032 | 0,019 | -0,209 | -0,049 | 0,325 | 0,084 | -0,302 | 0,256 | -0,279 | -0,246 | 0,002 | 0,193 | 0,153 | -0,065 | -0,081 | 0,323 | 0,118 | -0,077 |
| LOGRO_04 | -0,208 | -0,257 | -0,217 | 1,000 | -0,141 | -0,219 | -0,070 | -0,054 | -0,160 | -0,185 | -0,283 | -0,072 | -0,109 | 0,124 | 0,061 | -0,300 | 0,072 | 0,197 | 0,009 | -0,204 | -0,277 | 0,098 | -0,075 | -0,068 | 0,097 | 0,087 |
| LOGRO_05 | 0,374 | 0,129 | 0,310 | -0,141 | 1,000 | 0,592 | -0,061 | 0,050 | 0,010 | -0,138 | 0,067 | -0,018 | 0,188 | 0,010 | 0,001 | 0,228 | 0,084 | 0,038 | -0,073 | 0,241 | -0,217 | 0,250 | 0,146 | 0,117 | -0,098 | -0,026 |
| LOGRO_06 | 0,014 | 0,421 | 0,342 | -0,219 | 0,592 | 1,000 | 0,182 | 0,282 | -0,056 | 0,119 | -0,123 | -0,037 | 0,516 | 0,344 | 0,028 | 0,412 | -0,065 | -0,167 | 0,091 | 0,484 | 0,125 | 0,086 | 0,172 | 0,016 | -0,367 | 0,035 |
| LOGRO_07 | 0,281 | 0,098 | 0,321 | -0,070 | -0,061 | 0,182 | 1,000 | 0,886 | 0,619 | 0,572 | 0,326 | 0,240 | 0,511 | 0,209 | 0,126 | 0,705 | 0,504 | 0,161 | 0,681 | 0,563 | 0,411 | 0,365 | 0,105 | 0,146 | 0,265 | -0,156 |
| LOGRO_08 | 0,274 | 0,072 | 0,219 | -0,054 | 0,050 | 0,282 | 0,886 | 1,000 | 0,654 | 0,586 | 0,254 | 0,193 | 0,496 | 0,126 | 0,233 | 0,740 | 0,560 | 0,140 | 0,730 | 0,626 | 0,275 | 0,281 | 0,018 | 0,156 | 0,104 | -0,170 |
| LOGRO_09 | 0,132 | -0,230 | 0,032 | -0,160 | 0,010 | -0,056 | 0,619 | 0,654 | 1,000 | 0,830 | 0,667 | 0,491 | 0,218 | 0,095 | 0,219 | 0,644 | 0,589 | 0,381 | 0,823 | 0,526 | 0,467 | 0,514 | 0,093 | -0,019 | 0,353 | -0,413 |
| LOGRO_10 | 0,023 | -0,184 | 0,019 | -0,185 | -0,138 | 0,119 | 0,572 | 0,586 | 0,830 | 1,000 | 0,588 | 0,663 | 0,156 | 0,211 | -0,028 | 0,668 | 0,481 | 0,447 | 0,839 | 0,467 | 0,554 | 0,507 | 0,053 | -0,195 | 0,191 | -0,415 |
| LOGRO_11 | 0,339 | -0,328 | -0,209 | -0,283 | 0,067 | -0,123 | 0,326 | 0,254 | 0,667 | 0,588 | 1,000 | 0,579 | -0,125 | 0,070 | 0,096 | 0,354 | 0,686 | 0,486 | 0,618 | 0,300 | 0,363 | 0,414 | 0,167 | 0,047 | 0,582 | -0,367 |
| LOGRO_12 | 0,211 | -0,370 | -0,049 | -0,072 | -0,018 | -0,037 | 0,240 | 0,193 | 0,491 | 0,663 | 0,579 | 1,000 | -0,170 | -0,026 | -0,281 | 0,333 | 0,516 | 0,691 | 0,593 | 0,061 | 0,293 | 0,584 | 0,051 | -0,279 | 0,249 | -0,719 |
| LOGRO_13 | 0,005 | 0,546 | 0,325 | -0,109 | 0,188 | 0,516 | 0,511 | 0,496 | 0,218 | 0,156 | -0,125 | -0,170 | 1,000 | 0,556 | 0,177 | 0,658 | -0,035 | -0,356 | 0,296 | 0,677 | 0,207 | 0,265 | 0,453 | 0,337 | 0,002 | 0,165 |
| LOGRO_14 | -0,074 | 0,546 | 0,084 | 0,124 | 0,010 | 0,344 | 0,209 | 0,126 | 0,095 | 0,211 | 0,070 | -0,026 | 0,556 | 1,000 | 0,335 | 0,356 | -0,242 | -0,051 | 0,244 | 0,454 | 0,312 | 0,137 | 0,326 | 0,288 | 0,109 | 0,011 |
| LOGRO_15 | -0,133 | 0,167 | -0,302 | 0,061 | 0,001 | 0,028 | 0,126 | 0,233 | 0,219 | -0,028 | 0,096 | -0,281 | 0,177 | 0,335 | 1,000 | -0,047 | -0,032 | -0,070 | -0,002 | 0,089 | 0,244 | -0,032 | 0,007 | 0,272 | 0,118 | 0,234 |
| LOGRO_16 | 0,337 | 0,261 | 0,256 | -0,300 | 0,228 | 0,412 | 0,705 | 0,740 | 0,644 | 0,668 | 0,354 | 0,333 | 0,658 | 0,356 | -0,047 | 1,000 | 0,435 | 0,163 | 0,772 | 0,823 | 0,146 | 0,463 | 0,216 | 0,214 | 0,067 | -0,309 |
| LOGRO_17 | 0,381 | -0,505 | -0,279 | 0,072 | 0,084 | -0,065 | 0,504 | 0,560 | 0,589 | 0,481 | 0,686 | 0,516 | -0,035 | -0,242 | -0,032 | 0,435 | 1,000 | 0,496 | 0,705 | 0,309 | -0,014 | 0,488 | 0,063 | -0,111 | 0,279 | -0,330 |
| LOGRO_18 | 0,032 | -0,586 | -0,246 | 0,197 | 0,038 | -0,167 | 0,161 | 0,140 | 0,381 | 0,447 | 0,486 | 0,691 | -0,356 | -0,051 | -0,070 | 0,163 | 0,496 | 1,000 | 0,530 | -0,074 | -0,033 | 0,549 | -0,365 | -0,346 | 0,219 | -0,689 |
| LOGRO_19 | 0,137 | -0,195 | 0,002 | 0,009 | -0,073 | 0,091 | 0,681 | 0,730 | 0,823 | 0,839 | 0,618 | 0,593 | 0,296 | 0,244 | -0,002 | 0,772 | 0,705 | 0,530 | 1,000 | 0,630 | 0,258 | 0,532 | -0,012 | 0,012 | 0,307 | -0,550 |
| LOGRO_20 | 0,296 | 0,391 | 0,193 | -0,204 | 0,241 | 0,484 | 0,563 | 0,626 | 0,526 | 0,467 | 0,300 | 0,061 | 0,677 | 0,454 | 0,089 | 0,823 | 0,309 | -0,074 | 0,630 | 1,000 | 0,116 | 0,086 | 0,242 | 0,333 | 0,053 | -0,153 |
| LOGRO_21 | -0,219 | -0,009 | 0,153 | -0,277 | -0,217 | 0,125 | 0,411 | 0,275 | 0,467 | 0,554 | 0,363 | 0,293 | 0,207 | 0,312 | 0,244 | 0,146 | -0,014 | -0,033 | 0,258 | 0,116 | 1,000 | 0,230 | 0,344 | -0,193 | 0,239 | -0,026 |
| LOGRO_22 | -0,058 | -0,395 | -0,065 | 0,098 | 0,250 | 0,086 | 0,365 | 0,281 | 0,514 | 0,507 | 0,414 | 0,584 | 0,265 | 0,137 | -0,032 | 0,463 | 0,488 | 0,549 | 0,532 | 0,086 | 0,230 | 1,000 | 0,242 | -0,256 | 0,235 | -0,318 |
| LOGRO_23 | 0,226 | 0,184 | -0,081 | -0,075 | 0,146 | 0,172 | 0,105 | 0,018 | 0,093 | 0,053 | 0,167 | 0,051 | 0,453 | 0,326 | 0,007 | 0,216 | 0,063 | -0,365 | -0,012 | 0,242 | 0,344 | 0,242 | 1,000 | 0,039 | 0,093 | 0,225 |
| LOGRO_24 | 0,433 | 0,637 | 0,323 | -0,068 | 0,117 | 0,016 | 0,146 | 0,156 | -0,019 | -0,195 | 0,047 | -0,279 | 0,337 | 0,288 | 0,272 | 0,214 | -0,111 | -0,346 | 0,012 | 0,333 | -0,193 | -0,256 | 0,039 | 1,000 | 0,539 | 0,225 |
| LOGRO_25 | 0,256 | -0,074 | 0,118 | 0,097 | -0,098 | -0,367 | 0,265 | 0,104 | 0,353 | 0,191 | 0,582 | 0,249 | 0,002 | 0,109 | 0,118 | 0,067 | 0,279 | 0,219 | 0,307 | 0,053 | 0,239 | 0,235 | 0,093 | 0,539 | 1,000 | -0,035 |
| LOGRO_26 | -0,091 | 0,283 | -0,077 | 0,087 | -0,026 | 0,035 | -0,156 | -0,170 | -0,413 | -0,415 | -0,367 | -0,719 | 0,165 | 0,011 | 0,234 | -0,309 | -0,330 | -0,689 | -0,550 | -0,153 | -0,026 | -0,318 | 0,225 | 0,225 | -0,035 | 1,000 |

Fuente: Construcción propia

Figura 3-4 Pirámide Correlación logros familiares - Spearman




Fuente: Construcción propia

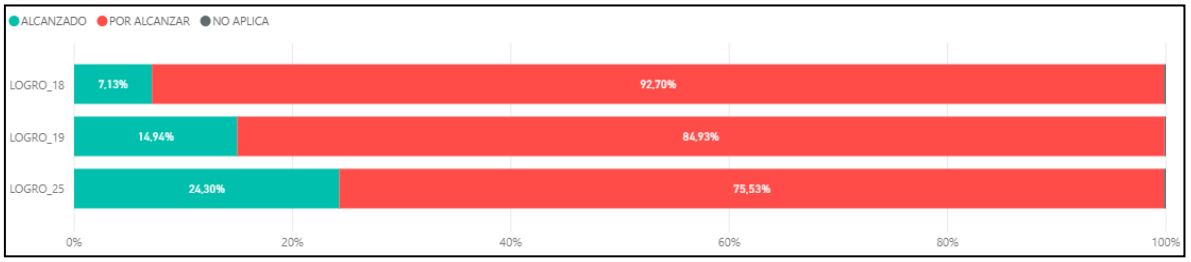
Un segundo ejercicio con los indicadores familiares consiste en agruparlos de acuerdo con la similitud en los estados adquiridos para la cobertura municipal analizada. De este modo se identifican los grupos **D-Debilidades**, **E-Baja Aplicabilidad** y **F-Fortalezas**.

Figura 3-5 Distribución en Porcentajes de hogares por Logro y Estado


Fuente: Construcción propia

**Grupo Debilidades**. Indicadores (18, 19 y 25), relacionado con estudios postsecundarios, herramientas digitales y vinculación laboral. En este grupo se asocian aquellos indicadores de difícil cumplimiento a corto plazo para un hogar pobre. Se visualiza con claridad una estrecha relación entre una vinculación laboral formal y el desarrollo de habilidades para el trabajo.

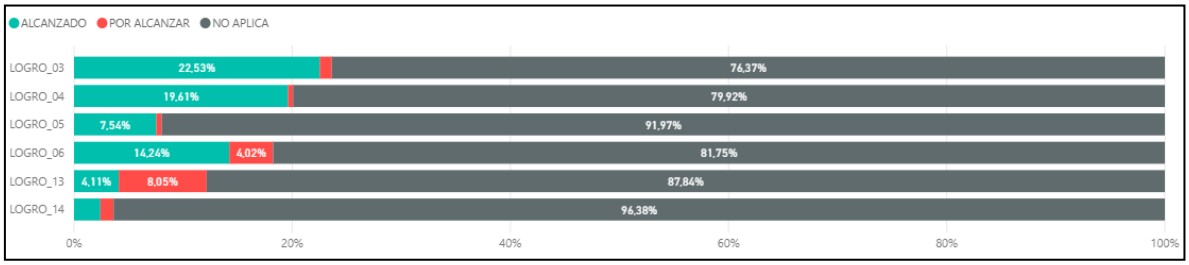
Figura 3-6 Grupo Debilidades




Fuente: Construcción propia

**Grupo de baja aplicabilidad**. Indicadores (3, 4, 5, 6, 13 y 14), relacionado con primera infancia y discapacidad. En este grupo se asocian aquellos indicadores que por lo regular No Aplican para el hogar dado que no cuenta con integrantes en situación de discapacidad o en grupo etario de primera infancia.

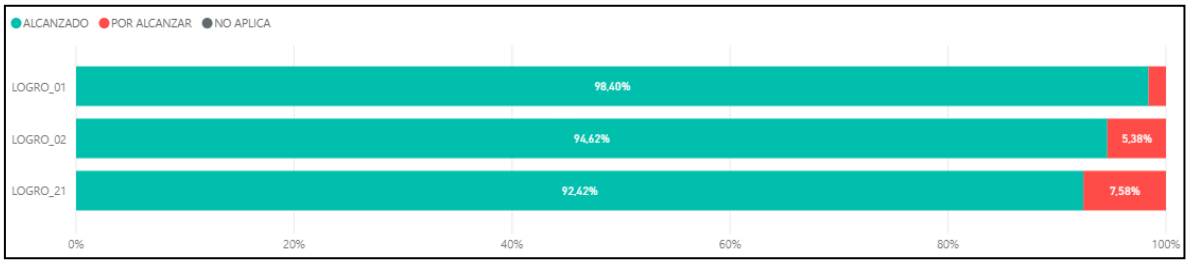
Figura 3-7 Grupo Baja Aplicabilidad




Fuente: Construcción propia

**Grupo Fortalezas.** Indicadores (1, 2 y 21) relacionado con identificación, aseguramiento de salud y pisos de la vivienda. En este grupo se asocian aquellos indicadores de fácil cumplimiento.

Figura 3-8 Grupo Fortalezas



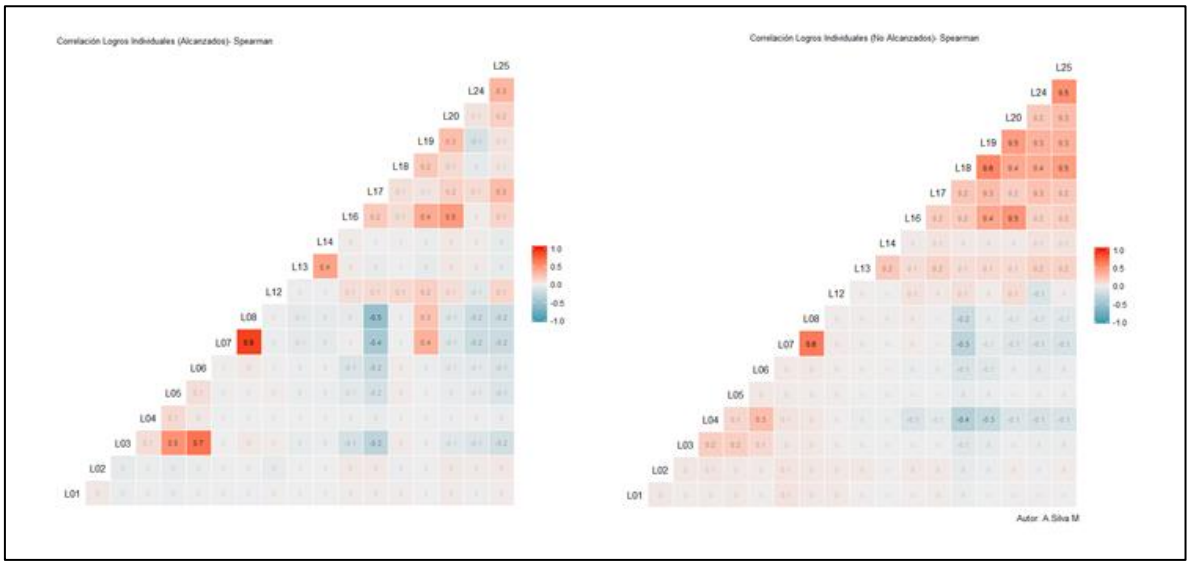

Fuente: Construcción propia

## 3.3 Estudio de los Indicadores individuales.

Las características de dieciocho (18) indicadores, la Estrategia calcula inicialmente la calificación a nivel de la persona. A partir de esta información se ponderan para conocer la calificación familiar entre todos los integrantes del hogar.

A partir del detalle de la calificación individual se realiza el ejercicio de correlación (método Spearman) y se confirma la relación del grupo **A-** "**Educación infantil"** y **D- "Debilidades"**, previamente identificados con los ejercicios anteriores. Sin embargo, al verificar las relaciones negativas con mayor peso, se identifica un nuevo grupo **G** (Indicador 8 y 17). En este caso **G- "Educación y Trabajo"** hay una relación inversamente proporcional entre saber leer y escribir con el trabajo infantil.

Figura 3-9 Correlación Logros Individuales Estado por Spearman




Fuente: Construcción propia

## 3.4 Hallazgos del análisis descriptivo

Se concluye que, si bien los indicadores se encuentran agrupados por dimensiones que distan de su objeto, si se evidencia interacciones al compararlos, encontrando siete (7) grupos focales en donde los indicadores apoyan la lectura o comportamiento del conjunto. Los grupos focales son:

A. Educación Infantil
B. Acceso a servicios
C. Responsabilidad

D. Debilidades
E. Baja aplicabilidad
F. Fortalezas
G. Educación y Trabajo

El comportamiento de la muestra de indicadores familiares e individuales es reflejo de los porcentajes a nivel nacional y municipal. Se resalta que la comparación de indicadores familiares e individuales es igual, dado que uno se construye a partir de la ponderación del otro. Al respecto el enfoque del proyecto será únicamente a partir de los indicadores familiares.

Se evidencia que por dimensión no hay un dato que permita el análisis por este agrupamiento y que refleje o resuma el estado interno de sus indicadores. Producto del análisis se validan los pesos por dimensión a partir del Índice de Pobreza Multidimensional (IMP) y se distribuye el sub-peso de los indicadores por cada dimensión asignando el 80% a los requeridos y el 20% para los deseados, cuando el estado del indicador es Alcanzado o No aplica. Se detalla a través de la siguiente tabla:

Tabla 3-1 Distribución de pesos por Dimensión y Logro

| DIMENSION | PESO | INDICADOR | DISTRIBUCIÓN PESO | REQUERIDO | DESCRIPCION INDICADOR |
|---|---|---|---|---|---|
| 1-Identificacion | 0,2 | 1 | 0,8 | SI | Documento de Identidad |
| | | 12 | 0,1 | | Libreta Militar |
| | | 13 | 0,1 | | Registro para la Localización y Caracterización de Personas con Discapacidad – RLCPD |
| 2-Salud y Nutrición | 0,2 | 2 | 0,2 | SI | Aseguramiento Seguridad Social |
| | | 3 | 0,2 | SI | Vacunación |
| | | 4 | 0,2 | SI | Tamizaje, desnutrición aguda |

| DIMENSION | PESO | INDICADOR | DISTRIBUCIÓN PESO | REQUERIDO | DESCRIPCION INDICADOR |
|---|---|---|---|---|---|
| | | 5 | 0,2 | SI | Crecimiento y desarrollo |
| | | 14 | 0,066666667 | | productos de apoyo discapacidad |
| | | 15 | 0,066666667 | | Seguridad alimentaria |
| | | 16 | 0,066666667 | | Derechos sexuales y reproductivos |
| 3-Educacion y Capacitación | 0,2 | 6 | 0,266666667 | SI | Educación inicial |
| | | 7 | 0,266666667 | SI | Acceso sistema educativo |
| | | 8 | 0,266666667 | SI | Trabajo infantil |
| | | 17 | 0,05 | | Leer y escribir |
| | | 18 | 0,05 | | Estudios Postsecundarios |
| | | 19 | 0,05 | | Herramientas digitales |
| | | 20 | 0,05 | | Educación financiera |
| 4-Habitabilidad | 0,2 | 9 | 0,4 | SI | Acceso a agua |
| | | 10 | 0,4 | SI | Saneamiento básico |
| | | 21 | 0,066666667 | | Pisos |
| | | 22 | 0,066666667 | | Paredes |
| | | 23 | 0,066666667 | | Hacinamiento |
| 5-Ingresos y Trabajo | 0,2 | 11 | 0,8 | SI | Ingreso superior línea de pobreza |
| | | 24 | 0,066666667 | | Ingreso propio adultos mayores |
| | | 25 | 0,066666667 | | Vinculación actividad productiva |
| | | 26 | 0,066666667 | | Seguridad jurídica del predio |

Para conceptualizar el ejercicio de análisis se construye herramienta en PowerBI en la dirección https://goo.gl/wQPnoN

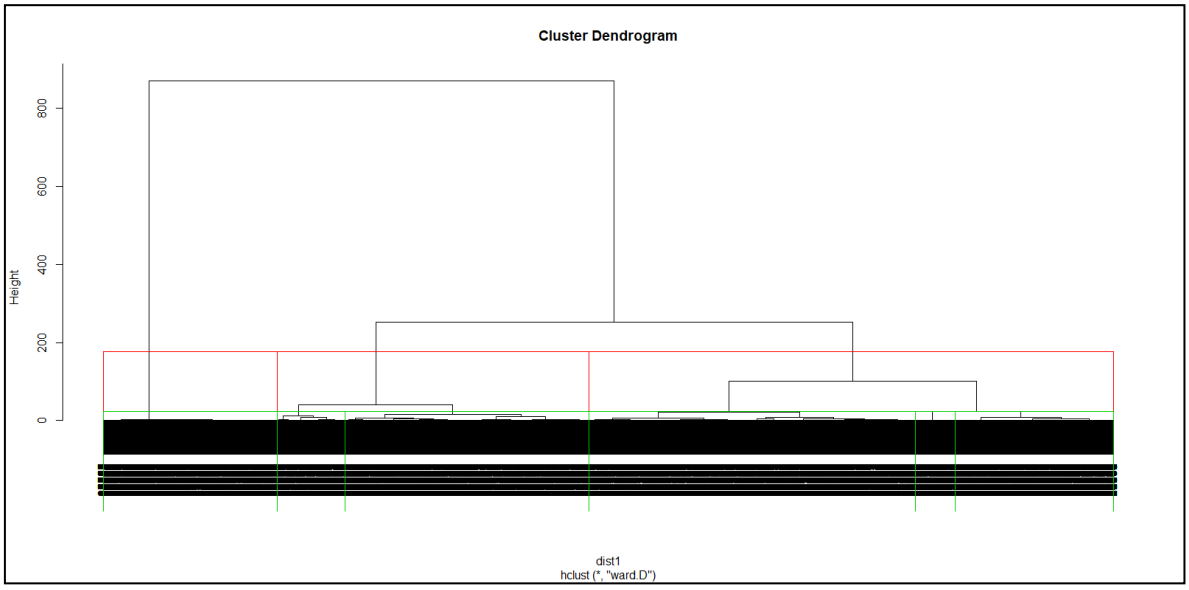
# 4 Análisis Exploratorio

Este capítulo presenta los resultados del postprocesamiento de hogares pobres que son objeto del análisis a través de las herramientas de clusterización jerárquica (método Ward, con distancia coseno) y se decide cortar los dendogramas en tres (3) y seis (6) grupos respectivamente, que posteriormente fueros interpretados y validados para describir en detalle los perfiles usando los centroides de cada grupo y las variables principales.

## 4.1 Aplicación de técnicas no supervisadas

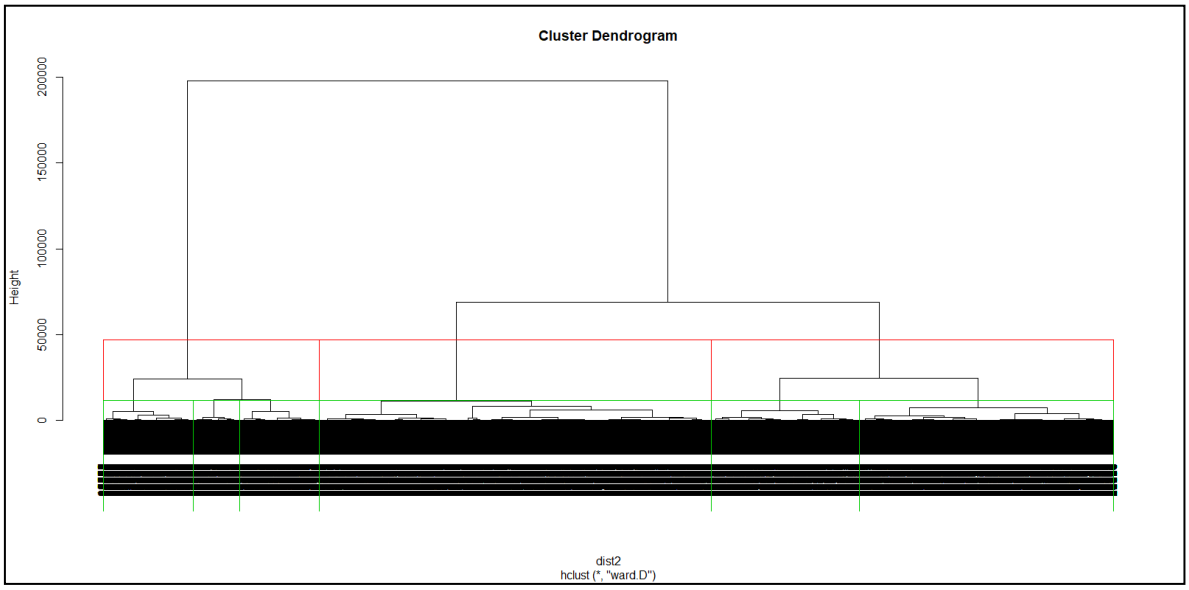
Se realiza la clusterización usando el método WARD a razón que garantiza la mínima varianza con clústeres más compactos que permiten ser más fáciles de interpretar. Y a través de los algoritmos de distancia por similitud de coseno, euclídea y manhattan, se forman las etiquetas o clasificadores realizando cortes para formar tres (3) y seis (6) grupos respectivamente. Estos cortes (k=3 y k=6) se grafican (ver figuras 4-1, 4-2 y 4-3) a través de dendogramas para visualizar el ejercicio.

*Figura 4-1Dendograma técnica coseno metodo ward*



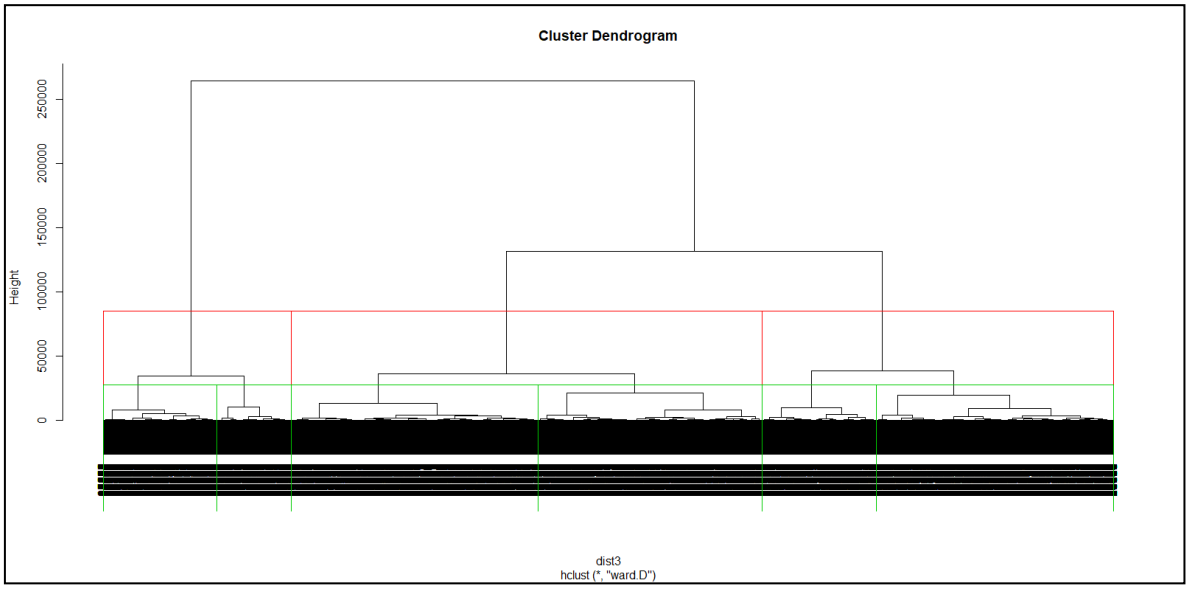

Fuente: Construcción propia

*Figura 4-2 Dendograma técnica euclidean método Ward*




Fuente: Construcción propia

*Figura 4-3 Genograma técnica manhattan método Ward*




Fuente: Construcción propia

Finalizada la diagramación por dendogramas y para continuar con el análisis exploratorio, se limita a la similitud coseno porque es muy útil en datos con cambios de signo como se puede observar en los valores que toman las variables de los logros que para el caso es 1, 0, -1. Es importante resaltar que la similitud por coseno mide esa direccionalidad en los datos, además de ofrecer un ángulo que puede diferenciar los valores tomados. Otras distancias por similitud como manhattan o euclidean anulan ese sentido porque al elevarse al cuadrado dan positivas.

Respecto al etiquetado se realiza en cada uno de los hogares objeto del análisis y por cada etiqueta creada en la fuente de información a partir del análisis exploratorio de las similitudes mencionados con los cortes en k=3 y k=6, se efectúa el cálculo de la media, mediana, desviación estándar, máximo y mínimo.

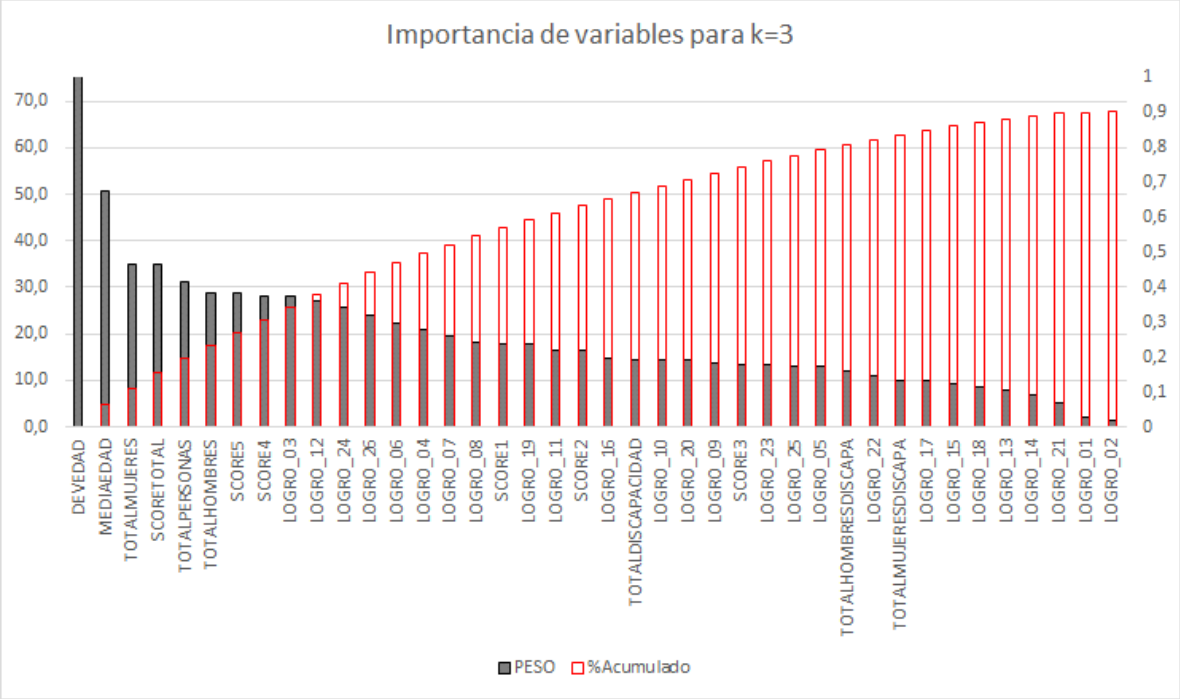
Este resultado consolidado permite que a través de Random Forest verificar la integridad, acertando en un 96.5% (tan solo 3.5% de error) en la clasificación de los hogares para el etiquetado por tan solo tres grupos, sin embargo, al aumentar a

una clasificación por seis grupos se disminuye a un 92% de asertividad (con un error del 8%).

## 4.2 Priorización de variables

El Random Forest permite identificar la importancia de las variables que definen la variedad y tipología de la muestra cuando el clasificador es para tres grupos, como se muestra en la figura 4-4 y figura 4-5 para los agrupamientos por k=3 y k=6.

*Figura 4-4 Importancia de variables para clasificación por tres grupos*

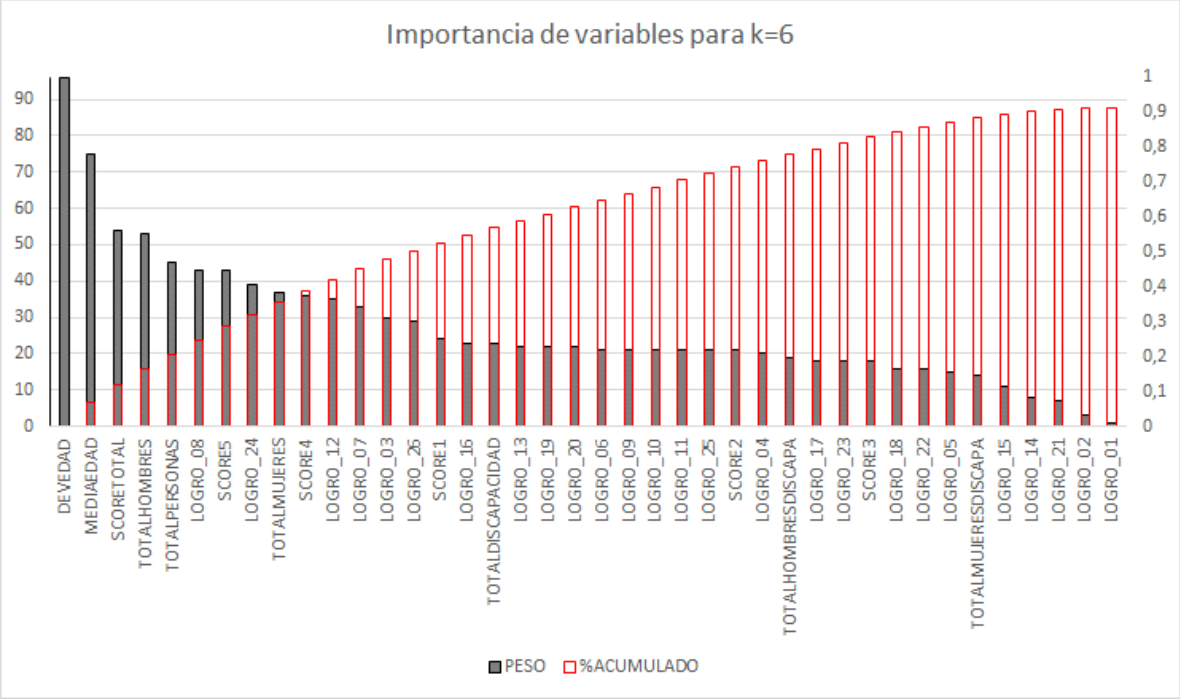



Fuente: Construcción propia

Para el etiquetado por tres grupos, el diagrama de Pareto de la figura 4-4 muestra las variables a tener en cuenta para el análisis de acuerdo con aquellas que presentan mayor peso respecto al porcentaje acumulado, para el análisis llama la atención el Score5, Score4 (hace referencia a las dimensiones Ingresos y Trabajo

y Habitabilidad respectivamente) y la variable LOGRO_03 que es parte integral del cálculo del Score2 (dimensión Salud y Nutrición), sin embargo, para tener mayor detalle se tomarán las variables siguientes LOGRO_12 y LOGRO_06 que hace referencia respectivamente a la construcción del Score1 y Score3 (dimensiones de Identificación y Educación y Capacitación) y de esta forma se complementa el universo de las dimensiones mencionadas en la sección 3.4.

*Figura 4-5 Importancia de variables para clasificación por seis grupos*




Fuente: Construcción propia

Para el etiquetado por seis grupos, el diagrama de Pareto de la figura 4-5 muestra las variables a tener en cuenta para el análisis de acuerdo con aquellas que presentan mayor peso respecto al porcentaje acumulado, para el análisis llama la atención el Score Total que agrupa todas las dimensiones, Score5 (hace referencia a las dimensiones Ingresos y Trabajo), la variable LOGRO_08 que es parte integral del cálculo del Score3 (dimensión educación y capacitación), sin

embargo, para tener mayor detalle se tomarán las variables siguientes Score4 (dimensión de habitabilidad), LOGRO_12 que es parte integral del cálculo del Score1 (dimensión identificación), LOGRO_03 que es parte integral del cálculo del Score2 (dimensión salud y nutrición) y de esta forma se complementa el universo de las dimensiones mencionadas en la sección 3.4.

## 4.3 Estadística descriptiva

Teniendo en cuenta las variables más representativas con sus valores estadísticos que se visualizan en la figura 4-6 y 4-7, se extrae y analiza cada uno de los tres o seis grupos resultado de la clasificación que se obtuvo a través de la distancia por similitud por coseno y se define el perfil verificando los centroides de los cálculos mediana, media, desviación estándar, máximo y mínimo para traducirlos a términos del negocio.

*Figura 4-6 Similitud Coseno - Variables Representativas para K=3*

| | COSINE K=3 & ETIQUETA=1 | | | | | COSINE K=3 & ETIQUETA=2 | | | | | COSINE K=3 & ETIQUETA=3 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| VARIABLES | MEDIANA | MEDIA | DES.ESTANDAR | MAXIMO | MINIMO | MEDIANA | MEDIA | DES.ESTANDAR | MAXIMO | MINIMO | MEDIANA | MEDIA | DES.ESTANDAR | MAXIMO | MINIMO |
| DEVEDAD | 16,39 | 16,98168024 | 5,084641574 | 39,5 | 3,39 | 14,82 | 15,55886479 | 4,912864109 | 38 | 2,5 | 0 | 1,250477496 | 2,037938998 | 10 | 0 |
| MEDIAEDAD | 23 | 23,688111 | 6,625591946 | 52,6 | 7,2 | 33 | 35,11960396 | 12,14741849 | 83,5 | 11 | 63 | 61,29750514 | 14,72469235 | 101 | 4 |
| TOTALMUJERES | 3 | 3,079557026 | 1,463029515 | 13 | 0 | 2 | 1,869322047 | 0,92193632 | 7 | 0 | 1 | 0,707790724 | 0,526983139 | 4 | 0 |
| SCORETOTAL | 0,75 | 0,746108282 | 0,133228017 | 1 | 0,26 | 0,796666667 | 0,810262767 | 0,124734473 | 1 | 0,28 | 0,843333333 | 0,835172493 | 0,118692545 | 1 | 0,38 |
| TOTALPERSONAS | 5 | 5,588594705 | 2,124523778 | 23 | 2 | 4 | 3,686947321 | 1,271090729 | 12 | 2 | 1 | 1,431345671 | 0,567157868 | 5 | 1 |
| TOTALHOMBRES | 2 | 2,509037678 | 1,49437129 | 15 | 0 | 2 | 1,817625274 | 1,100296455 | 8 | 0 | 1 | 0,723554946 | 0,524426536 | 4 | 0 |
| SCORE5 | 0,133333333 | 0,362822471 | 0,379370333 | 1 | 0 | 0,2 | 0,521306024 | 0,413875554 | 1 | 0 | 0,866666667 | 0,593557231 | 0,416981746 | 1 | 0 |
| SCORE4 | 0,933333333 | 0,820807875 | 0,281943771 | 1 | 0 | 1 | 0,862031089 | 0,255542758 | 1 | 0 | 1 | 0,803868708 | 0,306565804 | 1 | 0 |
| LOGRO_03 | 1 | 0,494271894 | 0,848688825 | 1 | -1 | -1 | -0,998941879 | 0,04431829 | 1 | -1 | -1 | -0,999543066 | 0,030230249 | 1 | -1 |
| LOGRO_12 | 0 | -0,157204684 | 0,747314497 | 1 | -1 | 0 | -0,132038395 | 0,757842789 | 1 | -1 | -1 | -0,791181174 | 0,542764353 | 1 | -1 |
| LOGRO_06 | 0 | 0,051171079 | 0,931479951 | 1 | -1 | -1 | -0,999017459 | 0,041683422 | 1 | -1 | -1 | -0,999543066 | 0,030230249 | 1 | -1 |

Fuente: Construcción propia

*Figura 4-7 Similitud Coseno - Variables Representativas para K=6*

| | COSINE K=6 & ETIQUETA=1 | | | | | COSINE K=6 & ETIQUETA=2 | | | | | COSINE K=6 & ETIQUETA=3 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| VARIABLES | MEDIANA | MEDIA | DES.ESTANDAR | MAXIMO | MINIMO | MEDIANA | MEDIA | DES.ESTANDAR | MAXIMO | MINIMO | MEDIANA | MEDIA | DES.ESTANDAR | MAXIMO | MINIMO |
| DEVEDAD | 17,55 | 17,91676208 | 5,100021005 | 39,5 | 4,99 | 15,4 | 16,20290938 | 5,260890405 | 38 | 5,25 | 12,03 | 12,10259663 | 2,905006078 | 21,5 | 3,7 |
| MEDIAEDAD | 24,2 | 24,75120995 | 6,739781096 | 52,6 | 7,67 | 28,6 | 29,883026 | 8,302173391 | 68,33 | 11 | 60 | 59,84091179 | 9,356757996 | 83,5 | 30 |
| SCORETOTAL | 0,763333333 | 0,761910338 | 0,128587551 | 1 | 0,266666667 | 0,793333333 | 0,80441448 | 0,127358104 | 1 | 0,286666667 | 0,793333333 | 0,81011893 | 0,118723958 | 1 | 0,386666667 |
| TOTALHOMBRES | 2 | 2,1019678 | 1,171880068 | 7 | 0 | 2 | 1,930150632 | 1,181954207 | 8 | 0 | 1 | 1,342913776 | 0,726040464 | 4 | 0 |
| TOTALPERSONAS | 5 | 4,980484632 | 1,527649003 | 13 | 2 | 4 | 3,993683188 | 1,301460892 | 12 | 2 | 2 | 2,591674926 | 0,723589261 | 7 | 2 |
| SCORE5 | 0,133333333 | 0,389960427 | 0,391053857 | 1 | 0 | 0,2 | 0,501522514 | 0,413986044 | 1 | 0 | 0,8 | 0,495341923 | 0,412787894 | 1 | 0 |
| LOGRO_08 | 1 | 0,848755895 | 0,362826812 | 1 | -1 | 1 | 0,689868805 | 0,653507964 | 1 | -1 | -1 | -1 | 0 | -1 | -1 |
| LOGRO_24 | -1 | -0,635550496 | 0,631048309 | 1 | -1 | -1 | -0,679664723 | 0,606756983 | 1 | -1 | 0 | 0,019821606 | 0,559552301 | 1 | -1 |
| TOTALMUJERES | 3 | 2,878516832 | 1,232044236 | 9 | 0 | 2 | 2,063532556 | 0,947906432 | 7 | 0 | 1 | 1,24876115 | 0,549676087 | 3 | 0 |
| SCORE4 | 0,933333333 | 0,838456117 | 0,267458354 | 1 | 0 | 1 | 0,860908649 | 0,255070838 | 1 | 0 | 1 | 0,838916419 | 0,277707717 | 1 | 0 |
| LOGRO_12 | 0 | -0,216945845 | 0,766978101 | 1 | -1 | 0 | -0,192662779 | 0,760176721 | 1 | -1 | -1 | -0,354806739 | 0,768281934 | 1 | -1 |
| LOGRO_03 | 1 | 0,515368353 | 0,838038972 | 1 | -1 | -1 | -0,99829932 | 0,056177373 | 1 | -1 | -1 | -1 | 0 | -1 | -1 |

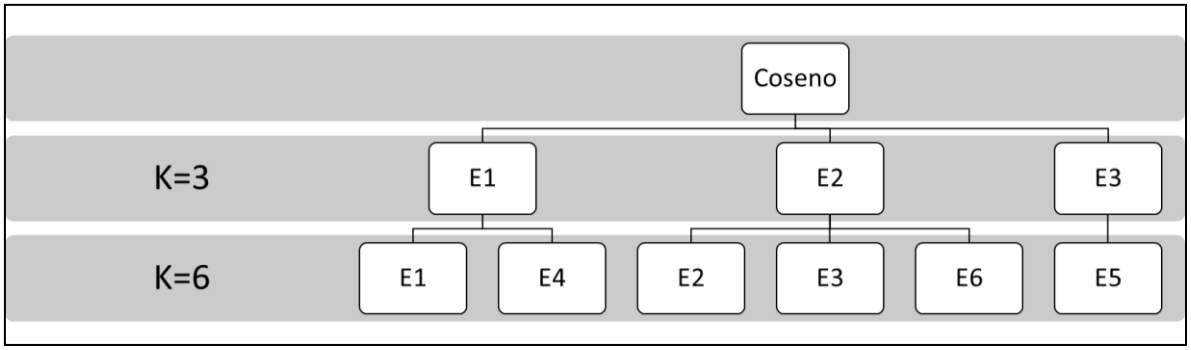
| | COSINE K=6 & ETIQUETA=4 | | | | | COSINE K=6 & ETIQUETA=5 | | | | | COSINE K=6 & ETIQUETA=6 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| VARIABLES | MEDIANA | MEDIA | DES.ESTANDAR | MAXIMO | MINIMO | MEDIANA | MEDIA | DES.ESTANDAR | MAXIMO | MINIMO | MEDIANA | MEDIA | DES.ESTANDAR | MAXIMO | MINIMO |
| DEVEDAD | 13,2 | 13,61330404 | 3,286350681 | 26,26 | 3,39 | 0 | 1,250477496 | 2,037938998 | 10 | 0 | 14,82 | 15,10412782 | 4,089550396 | 30,92 | 2,5 |
| MEDIAEDAD | 19,63 | 19,85858817 | 4,435857917 | 38,82 | 7,2 | 63 | 61,29750514 | 14,72469235 | 101 | 4 | 38,67 | 39,67191228 | 9,651354529 | 78,33 | 15,5 |
| SCORETOTAL | 0,696666667 | 0,689185706 | 0,134150561 | 0,99 | 0,26 | 0,843333333 | 0,835172493 | 0,118692545 | 1 | 0,38 | 0,83 | 0,822365079 | 0,119809956 | 1 | 0,28 |
| TOTALHOMBRES | 4 | 3,975395431 | 1,607603309 | 15 | 0 | 1 | 0,723554946 | 0,524426536 | 4 | 0 | 2 | 1,705513784 | 0,950647112 | 6 | 0 |
| TOTALPERSONAS | 7 | 7,779144698 | 2,497564764 | 23 | 3 | 1 | 1,431345671 | 0,567157868 | 5 | 1 | 3 | 3,331077694 | 1,052119632 | 10 | 2 |
| SCORE5 | 0,133333333 | 0,265065417 | 0,31520049 | 1 | 0 | 0,866666667 | 0,593557231 | 0,416981746 | 1 | 0 | 0,866666667 | 0,568688388 | 0,410113953 | 1 | 0 |
| LOGRO_08 | 1 | 0,731693029 | 0,448467109 | 1 | -1 | -1 | -0,995202193 | 0,089303682 | 1 | -1 | -1 | -0,792982456 | 0,58153321 | 1 | -1 |
| LOGRO_24 | -1 | -0,867018161 | 0,43902076 | 1 | -1 | 0 | -0,219785241 | 0,740041533 | 1 | -1 | -1 | -0,555889724 | 0,670817888 | 1 | -1 |
| TOTALMUJERES | 4 | 3,803749268 | 1,927352326 | 13 | 0 | 1 | 0,707790724 | 0,526983139 | 4 | 0 | 2 | 1,62556391 | 0,80866957 | 6 | 0 |
| SCORE4 | 0,933333333 | 0,757234915 | 0,321017431 | 1 | 0 | 1 | 0,803868708 | 0,306565804 | 1 | 0 | 1 | 0,870192147 | 0,250281793 | 1 | 0 |
| LOGRO_12 | 0 | 0,057996485 | 0,626336015 | 1 | -1 | -1 | -0,791181174 | 0,542764353 | 1 | -1 | 0 | 0,049373434 | 0,714297884 | 1 | -1 |
| LOGRO_03 | 1 | 0,41827768 | 0,882070242 | 1 | -1 | -1 | -0,999543066 | 0,030230249 | 1 | -1 | -1 | -1 | 0 | -1 | -1 |

Fuente: Construcción propia

## 4.4 Interpretación de Perfiles

Para facilitar el procesamiento y toma de decisiones del negocio, en la figura 4-8 se detalla las etiquetas objeto de interpretación y cómo éstas se relacionan cuando el agrupamiento pasa de tres a seis.

*Figura 4-8 Etiquetas objeto de interpretación*




Fuente: Construcción propia

### 4.4.1 Interpretación para tres grupos

*Tabla 4-1 Interpretación clasificación por tres grupos similitud Coseno*

| K=3 | Etiqueta 1 | Etiqueta 2 | Etiqueta 3 |
|---|---|---|---|
| Nombre propuesto para el grupo | **PROYECTO FORMACIÓN** | **PROYECTOS PRODUCTIVOS** | **PROYECTO SALUD Y NUTRICIÓN** |
| Porcentaje de marcación | 30.85% | 51.96% | 17.19% |
| Promedio de edad en el hogar | 23 años (+/- 5) | 33 años | 63 años |
| Promedio de integrantes en el hogar | 5 (+/- 2) | 4 (+/- 1) | 1 |
| Características del grupo | Hogar conformado mayoritariamente por mujeres jóvenes. Cuentan con condiciones habitacionales apropiadas para garantizar su seguridad y salubridad. Pero con una notable ausencia para desarrollar su potencial productivo y capacidades que permita generar oportunidades en el mercado laboral. | Hogar conformado equitativamente por género. Con tendencia a la ausencia de primera infancia y niños. Tienen condiciones habitacionales apropiadas para garantizar su seguridad y salubridad. Pero con una notable ausencia para desarrollar su potencial productivo y capacidades que permita generar oportunidades en el mercado. | Hogar con tendencia unifamiliar de adultos mayores. Cuentan con condiciones apropiadas en los escenarios habitacionales y de ingreso. Pero con una notable ausencia en condiciones de salud y nutrición. |
| Posible sugerencia | Acercar ofertas de estudio como media técnica, tecnológica y/o universitaria. | Acercar ofertas de estudio como tecnológica y/o universitaria. Adicional de cursos complementarios para incentivar el emprendimiento e incubación de empresa. | Oferta de promoción y prevención en salud y nutrición |

## 4.4.2 Interpretación para seis grupos

*Tabla 4-2 Interpretación clasificación por seis grupos similitud Coseno*

| K=6 | Nombre propuesto para el grupo | Porcentaje de marcación | Promedio de edad en el hogar | Promedio de integrantes en el hogar | Características del grupo | Posible sugerencia |
|---|---|---|---|---|---|---|
| Etiqueta 1 | PROYECTO FORMACIÓN | 24,15% | 24 años (+/- 6) | 5 | Hogar conformado mayoritariamente por mujeres jóvenes. Cuentan con condiciones habitacionales apropiadas para garantizar su seguridad y salubridad. Pero con una notable ausencia para resolver la situación militar y desarrollar sus habilidades y capacidades que permita generar oportunidades en el mercado laboral. | Acercar ofertas de estudio como media técnica, tecnológica y/o universitaria. |
| Etiqueta 4 | | 6,70% | 19 años (+/- 4) | 7 | Hogar conformado equitativamente por género. Cuentan con condiciones habitacionales apropiadas para garantizar su seguridad y salubridad. Pero con una notable ausencia para resolver la situación militar y desarrollar sus habilidades y capacidades que permita generar oportunidades en el mercado laboral. | |
| Etiqueta 2 | PROYECTOS PRODUCTIVOS | 32,33% | 28 años (+/- 8) | 4 | Hogar conformado equitativamente por género. Con tendencia a la ausencia de primera infancia y niños. Tienen condiciones habitacionales apropiadas para garantizar su seguridad y salubridad. Pero con una notable ausencia para desarrollar su potencial productivo y capacidades que permita generar oportunidades en el mercado. | Acercar ofertas de estudio como tecnológica y/o universitaria. Adicional de cursos complementarios para incentivar el emprendimiento e incubación de empresa. |

| K=6 | Nombre propuesto para el grupo | Porcentaje de marcación | Promedio de edad en el hogar | Promedio de integrantes en el hogar | Características del grupo | Posible sugerencia |
|---|---|---|---|---|---|---|
| Etiqueta 3 | | 3,96% | 60 años (+/- 9) | 2 | Hogar conformado equitativamente por género. Con tendencia a la ausencia de primera infancia y niños. Tienen condiciones habitacionales apropiadas para garantizar su seguridad y salubridad. Pero con una notable ausencia para desarrollar su potencial productivo y capacidades que permita generar oportunidades en el mercado. | |
| Etiqueta 6 | | 15,67% | 39 años (+/- 9) | 3 | Hogar conformado equitativamente por género. Con tendencia a la ausencia de primera infancia y niños. Tienen condiciones habitacionales apropiadas para garantizar su seguridad y salubridad. Pero con una notable ausencia para desarrollar su potencial productivo y capacidades que permita generar oportunidades en el mercado. | |
| Etiqueta 5 | PROYECTO SALUD Y NUTRICIÓN | 17,19% | 63 años (+/- 14) | 1 | Hogar con tendencia unifamiliar de adultos mayores. Cuentan con condiciones apropiadas en los escenarios habitacionales y de ingreso. Pero con una notable ausencia en condiciones de salud y nutrición | Oferta de promoción y prevención en salud y nutrición |

Para conceptualizar el ejercicio con los resultados del postprocesamiento y tomar decisiones a través de la lectura desde un visualizador que brinde mayor eficacia y eficiencia en la interpretación de los perfiles, se construye herramienta en PowerBI en la dirección https://goo.gl/wQPnoN en la página 2.

# 5 Principales Resultados

De acuerdo con la metodología KDD los principales resultados por cada fase son:

1. **Selección de datos**. La elección de una zona territorial de Colombia que cuenta con gran cantidad de dinámicas en términos de pobreza y el acompañamiento familiar que realiza la Estrategia Unidos. Además de la identificación de variables por hogar y persona para el desarrollo de la investigación.

2. **Preprocesamiento**. La construcción de una herramienta de visualización y posterior la identificación de siete (7) grupos focales a partir de un análisis descriptivo para entender el comportamiento de los hogares respecto a los logros previamente calculados. Esta identificación aporta en la preparación y limpieza de los datos para las fases posteriores de la investigación.

3. **Transformación**. La construcción de un dato que permita el análisis para el agrupamiento por dimensión y que refleje o resuma el estado interno de sus logros. Producto del análisis se validan los pesos por dimensión a partir del Índice de Pobreza Multidimensional (IMP) y se distribuye el sub-peso de los indicadores por cada dimensión asignando el 80% a los requeridos y el 20% para los deseados, cuando el estado del indicador es Alcanzado o No aplica.

4. **Minería de datos**. La practicidad en la aplicación de técnicas no supervisadas a la fuente de información para la construcción del etiquetado. La definición por juicio de expertos de un único algoritmo de distancia por similitud.

5. **Interpretación y Evaluación**. La identificación en términos del negocio para tres perfiles de hogares pobres colombianos, el cual se desagrega y detalla a seis perfiles, que en conjunto conforman un portafolio de interpretación y que sirve como herramienta para la toma decisiones en procesos de planeación e implementación de Estrategia UNIDOS o intervención de entidades aliadas en apoyo a la superación de la pobreza en Colombia. Este ejercicio se detalla con la construcción de una herramienta de visualización.

# 6 Conclusiones y Recomendaciones

En esta investigación se ha explotado la información recolectada por la Estrategia Unidos de los hogares pobres focalizados por Prosperidad Social en los departamentos de Caldas, Quindío y Risaralda. El conjunto de datos analizados adopta la metodología del Índice de Pobreza Multidimensional a través del estado de los veintiséis logros familiares que se distribuyen en cinco dimensiones

El análisis descriptivo a través de la correlación de variables permitió conocer la composición demográfica y la situación de los hogares, agrupados en siete subconjuntos con características comunes en el estado de logros (Educación Infantil, Acceso a servicios, Responsabilidad, Debilidades, Baja aplicabilidad, Fortalezas, Educación y Trabajo). Los subconjuntos detallados corresponden al primer objeto de intervención social por parte del estado colombiano a través de entregas de beneficios tangibles e intangibles para el mejoramiento y permanencia de las condiciones.

Para comparar los resultados proporcionados por el análisis descriptivo, se asumió el cálculo del score del hogar a través de una distribución equitativa por dimensión, pero internamente en cada dimensión con sub-pesos diferenciados por la importancia del logro para la superación de la pobreza de acuerdo con la metodología de la Estrategia Unidos.

Las variables existentes junto con las nuevas variables score hicieron parte del análisis exploratorio a través de la identificación de etiquetas por medio del cálculo de distancias por similitud (técnicas no supervisadas). El ejercicio permite conocer tres diferentes perfiles de hogares que se pueden detallar hasta seis tipos, además de las variables que tienen mayor importancia para la creación del

agrupamiento, de las cuales se destacan el tipo score que tienen un análisis interno más amplio dada su constitución.

Clasificar un hogar por alguno de los tres perfiles tiene un error de tan solo 3.5% y normalmente sucede en la intersección del promedio de edad del hogar en dos de los perfiles identificados.

Los perfiles identificados con previa interpretación del negocio y visualización, permite tomar decisiones de tipo estratégico, táctico y operativo respecto al acompañamiento e intervención posterior de un hogar.

Tanto para el análisis descriptivo como para el exploratorio/resultados del postprocesamiento del clustering, se construye una representación gráfica que comprende las variables del negocio (ubicación geográfica, estado de logros y ciclo vital de hogares) que se complementa con las variables producto de la investigación (cálculos score y agrupamientos sugeridos con cortes en tres y seis). Esta visualización permite al usuario explorar la información de forma amigable, comprender el comportamiento y dinámicas de los hogares a nivel departamental y municipal, conocer la distribución por sexo y ciclo vital, responder preguntas de la operación del acompañamiento familiar que brinda Prosperidad Social a través de la Estrategia Red UNIDOS en los procesos de caracterización, plan de trabajo familiar, gestión oferta orientado a la superación de pobreza y la consecución de los logros que los hogares tengan por alcanzar o construcción de temáticas para fortalecer los logros que los hogares tengan cumplidos. Finalmente tomar decisiones a través de la lectura del visualizador con mayor eficacia y eficiencia a través de los perfiles interpretados.

En consecuencia, esta nueva experiencia de intervención mejora la prestación del servicio a la población vulnerable, permitiendo estandarizar a través

de la identificación, planeación, ordenamiento, priorización y garantía del uso adecuado de los recursos del estado colombiano.

Dado que el análisis propuesto sugiere agrupamientos por hogares, se necesitaban investigaciones para identificar y detallar mejor a las personas acompañadas en cada uno de los perfiles construidos. En especial cuando el 82.81 % de los hogares (dos de los tres perfiles), tienen dificultades comunes en la dimensión de ingresos y trabajo.

Finalmente se recomienda complementar el ejercicio de investigación integrando información de las demás entidades públicas, privadas, ONG, cooperación internacional, etc. Que se movilizan en torno a entregar oferta oportuna previa identificación de las necesidades de los hogares por la Estrategia Unidos.

# Referencias


Anand, S., & Sen, A. (2000). Human Development and Economic Sustainability. World Development 2029-2049. https://doi.org/10.1016/S0305-750X(00)00071-1

Angulo, R., Diaz, Y., & Pardo, R. (2011). Indice de Pobreza Multidimensional para Colombia (IPM-Colombia) 1997-2010. Departamento Nacional de Planeación. https://colaboracion.dnp.gov.co/cdt/estudios%20econmicos/382.pdf

Amat, J. (2017). Clustering y heatmaps: aprendizaje no supervisado. https://rpubs.com/Joaquin_AR/310338

Antoniadis, I., Tsiakiris, T., & Tsopogloy, S. (2015). Business Intelligence During Times of Crisis: Adoption and Usage of ERP Systems by SMEs. Procedia - Social and Behavioral Sciences, 175, 299–307. https://doi.org/10.1016/j.sbspro.2015.01.1204

Aufaure, M. A., & Chiky, R. (2014). From business intelligence to semantic data stream management. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 8823, 85–93. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-84910630749&partnerID=tZOtx3y1

Benati, S., Puerto, J., & Rodríguez-Chía, A. M. (2017). Clustering data that are graph connected. European Journal of Operational Research, 261(1), 43–53. https://doi.org/10.1016/j.ejor.2017.02.009

Bulos, R., & Delfino, N. (2014). Exploring data mining and gamification as tools for poverty analysis and policy formulation: a methodological framework. https://pdfs.semanticscholar.org/758a/bb0c8dd7d33d77fe82cb1fdc0fdabf36887d.pdf

Cleofas-Sánchez, L., Sánchez, J. S., García, V., & Valdovinos, R. M. (2016). Associative learning on imbalanced environments: An empirical study. Expert Systems with Applications, 54, 387–397. https://doi.org/10.1016/j.eswa.2015.10.001

Congreso de la República de Colombia. (2016). Ley 1785 de 21 de junio de 2016 “Por medio de la cual se establece la Red para la Superación de la Pobreza Extrema Red Unidos y se dictan otras disipaciones,” 1–8. Retrieved from http://es.presidencia.gov.co/normativa/normativa/LEY 1785 DEL 21 DE JUNIO DE 2016.pdf

Consejo Nacional de Política Económica y Social de Colombia. (2012). Conpes Social 150. Departamento Nacional de Planeación. https://colaboracion.dnp.gov.co/CDT/Conpes/Social/150.pdf

Coromaldi, M., & Zoli, M. (2007). A multidimensional poverty analysis. evidence from italian data. https://art.torvergata.it/retrieve/handle/2108/43383/50721/257%20coromaldi_zoli_9ott%20CEIS.pdf

Côrte-Real, N., Ruivo, P., & Oliveira, T. (2014). The Diffusion Stages of Business Intelligence & Analytics (BI&A): A Systematic Mapping Study. Procedia Technology, 16, 172–179. https://doi.org/10.1016/j.protcy.2014.10.080

Gröger, C., Hillmann, M., Hahn, F., Mitschang, B., & Westkämper, E. (2013). The operational process dashboard for manufacturing. Procedia CIRP, 7, 205–210. https://doi.org/10.1016/j.procir.2013.05.035

Davenport, T, Eccles, R, & Prusak, L (1999). Chapter 2 - Information Politics, In Knowledge and Special Libraries, 29-48. https://doi.org/10.1016/B978-0-7506-7084-5.50004-X

Departamento Administrativo Nacional de Estadística en Colombia (2018). Pobreza Monetaria y Multidimensional en Colombia. Boletín Técnico. https://www.dane.gov.co/files/investigaciones/...vida/pobreza/bol_pobreza_17.pdf

Departamento para la Prosperidad Social (2017). Guía del Acompañamiento Familiar.GGA2. Version 2. Grupo de Diseño Metodológico y Formación.

Departamento para la Prosperidad Social (2017). Manual operative de la estrategia de acompañamiento familiar y comunitario. MGA1 Version 5. Grupo de Diseño Metodológico y Formación.

Earley, S., Henderson, D., & Data Management Association. (2017). DAMA-DMBOK: Data management body of knowledge.

Fayyad, U., Stolorz, P. (1997) Data mining and KDD: Promise and challenges, Future Generation Computer Systems, 99-115. https://doi.org/10.1016/S0167-739X(97)00015-0

Foreman, J. (2013). Data Smart: Using Data Science to Transform Information into Insight.

Foster, P., Fawcett, T. (2013). Data Science for Business: What you need to know about data mining and data-analytic thinking

De la Fuente, S. (2011). Análisis de conglomerados. Universidad Autonoma de Madrid. http://www.fuenterrebollo.com/Economicas/ECONOMETRIA/SEGMENTACION/CONGLOMERADOS/conglomerados.pdf

Jiawei Han, Micheline Kamber & Jian Pei. (2012). 3 - Data Preprocessing, In the Morgan Kaufmann Series in Data Management Systems, Morgan Kaufmann, Boston, Pages 83-124, Data Mining (Third Edition), ISBN 9780123814791, https://doi.org/10.1016/B978-0-12-381479-1.00003-4.

Hahn, G. J., & Packowski, J. (2015). A perspective on applications of in-memory analytics in supply chain management. Decision Support Systems, 76, 45–52. https://doi.org/10.1016/j.dss.2015.01.003

Holsapple, C., Lee-post, A., & Pakath, R. (2014). A unified foundation for business analytics. Decision Support Systems, 64, 130–141. https://doi.org/10.1016/j.dss.2014.05.013

Iglesias, J. A., Tiemblo, A., Ledezma, A., & Sanchis, A. (2016). Web news mining in an evolving framework. Information Fusion, 28(January 2014), 90–98. https://doi.org/10.1016/j.inffus.2015.07.004

Larson, D., & Chang, V. (2016). International Journal of Information Management A review and future direction of agile, business intelligence, analytics and data science. International Journal of Information Management, 36(5), 700–710. https://doi.org/10.1016/j.ijinfomgt.2016.04.013

Lin, W. C., Tsai, C. F., Hu, Y. H., & Jhang, J. S. (2017). Clustering-based undersampling in class-imbalanced data. Information Sciences, 409–410, 17–26. https://doi.org/10.1016/j.ins.2017.05.008

Liu, B., Hsu, W., Chen, S., Ma, Y. (2000). Analyzing the subjective Interestingness of Association Rules. IEEE Intelligent Systems, september/October (2000) 47-55.)

Lizma, P. & Boccado, G. (2014). Guía de Asociación entre variables (Pearson y Spearman en SPSS). Universidad de Chile. https://www.u-cursos.cl/facso/2014/2/SO01007/1/material_docente/bajar?id_material=994690.

Microsoft. (2018). Conceptos de minería de datos. https://docs.microsoft.com/es-es/sql/analysis-services/data-mining/data-mining-concepts?view=sql-analysis-services-2017

Moreno, M., & López, V. (2018). Uso de técnicas no supervisadas en la construcción de modelos de clasificación en ingenieria de software. www.lsi.us.es/redmidas/Capitulos/LMD14.pdf

Nguyen, D., Nguyen, L. T. T., Vo, B., & Pedrycz, W. (2016). Efficient mining of class association rules with the itemset constraint. Knowledge-Based Systems, 103, 73–88. https://doi.org/10.1016/j.knosys.2016.03.025

Ochin, Kumar, S., & Joshi, N. (2016). Rule Power Factor: A New Interest Measure in Associative Classification. Procedia Computer Science, 93(September), 12–18. https://doi.org/10.1016/j.procs.2016.07.175

Olson, D. L., Delen, D., & Meng, Y. (2012). Comparative analysis of data mining methods for bankruptcy prediction. Decision Support Systems, 52(2), 464–473. https://doi.org/10.1016/j.dss.2011.10.007

Padmanabhan, B., & Tuzhilin, A. (1999) Knowledge refinement based on the discovery of unexpected patterns in data mining. Decision Support Systems 27 (1999) 303– 318.)

Pluliková, N. (2015). Poverty analysis using machine learning methods. http://www.iam.fmph.uniba.sk/institute/stehlikova/BC/2016-plulikova.pdf

Rouhani, S., Ghazanfari, M., & Jafari, M. (2012). Evaluation model of business intelligence for enterprise systems using fuzzy TOPSIS. Expert Systems with Applications, 39(3), 3764–3771. https://doi.org/10.1016/j.eswa.2011.09.074

Santi, É., Aloise, D., & Blanchard, S. J. (2016). A model for clustering data from heterogeneous dissimilarities. European Journal of Operational Research, 253(3), 659–672. https://doi.org/10.1016/j.ejor.2016.03.033

Sensuse, D. I., Cahyaningsih, E., & Wibowo, W. C. (2015). Identifying Knowledge Management Process of Indonesian Government Human Capital Management Using Analytical Hierarchy Process and Pearson Correlation Analysis.

Procedia Computer Science, 72(81), 233–243. https://doi.org/10.1016/j.procs.2015.12.136

Shirvanian, A., & Bakhshoodeh, M. (2012). Investigating poverty in rural Iran: The multidimensional poverty approach. http://dx.doi.org/10.4236/as.2012.35077

Silahtaroğlu, G., & Alayoglu, N. (2016). Using or Not Using Business Intelligence and Big Data for Strategic Management: An Empirical Study Based on Interviews with Executives in Various Sectors. Procedia - Social and Behavioral Sciences, 235(October), 208–215. https://doi.org/10.1016/j.sbspro.2016.11.016

Tarozzi, A., & Deaton, A. (2008). Using census and survey data to estimate poverty and inequality for small areas. https://www.princeton.edu/rpds/papers/WP_246.pdf

Tello, E. A., Alberto, J. M., & Velasco, P. (2016). Inteligencia de negocios: estrategia para el desarrollo de competitividad en empresas de base tecnológica Business intelligence: Strategy for competitiveness development in technology-based firms. Contaduría Y Administración, 61(1), 127–158. https://doi.org/10.1016/j.cya.2015.09.006

Trieu, V.-H. (2016). Getting value from Business Intelligence systems: A review and research agenda. Decision Support Systems, 93, 111–124. https://doi.org/10.1016/j.dss.2016.09.019

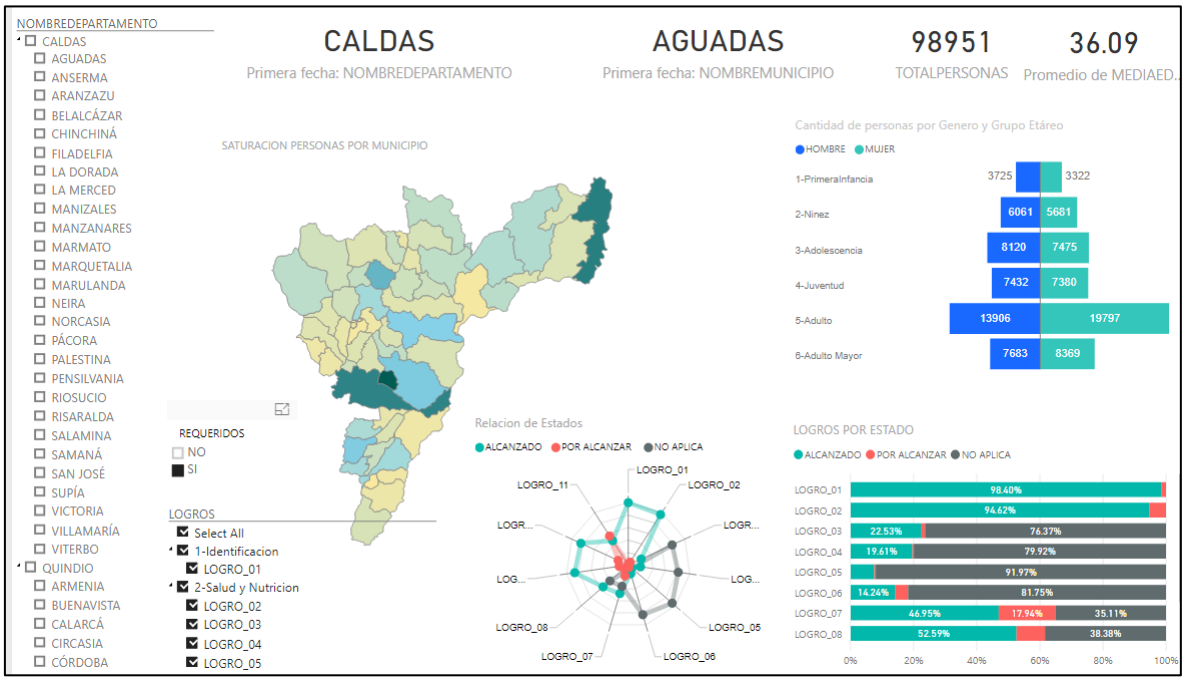
Wang, C. H. (2016). A novel approach to conduct the importance-satisfaction analysis for acquiring typical user groups in business-intelligence systems. Computers in Human Behavior, 54, 673–681. https://doi.org/10.1016/j.chb.2015.08.014

# Abreviaciones

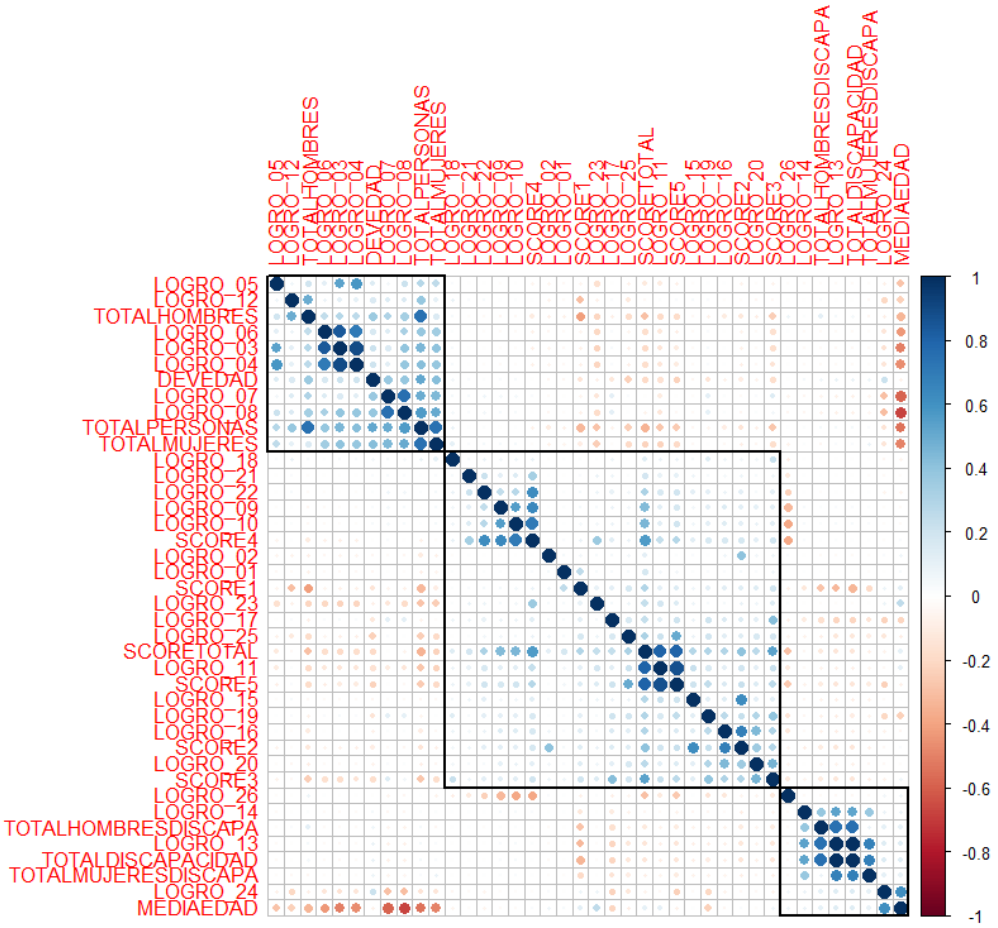
| | |
|---|---|
| KDD | Descubrimiento de conocimiento en bases de datos (Knowledge discovery in databases) |
| IPM | Índice de Pobreza Multidimensional |

# Anexos

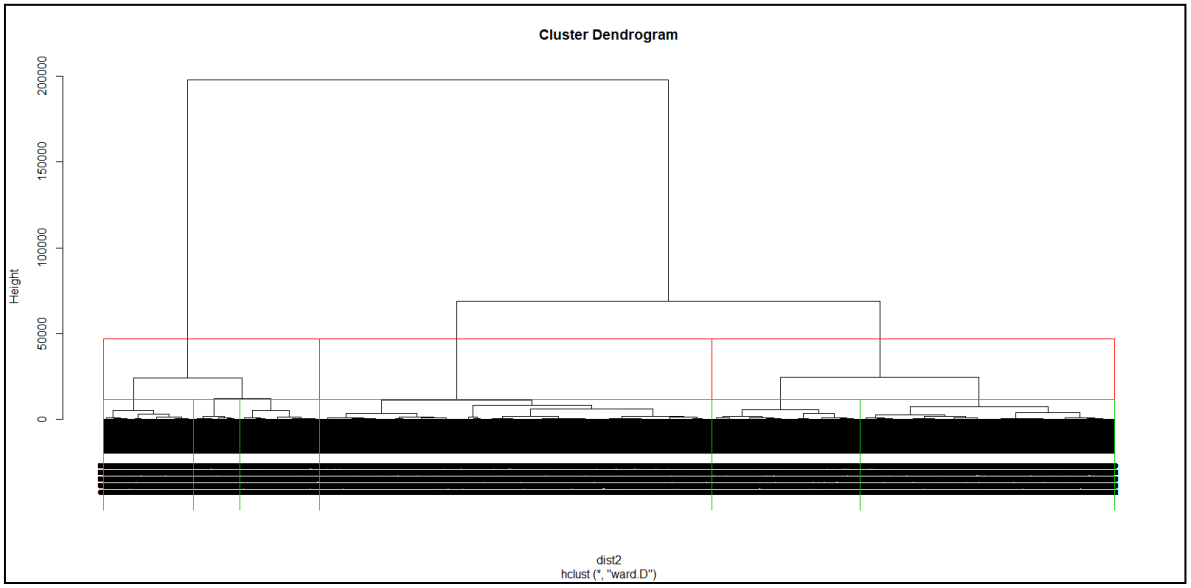
1. Imagen con la representación gráfica en PowerBI del análisis exploratorio propuesto.



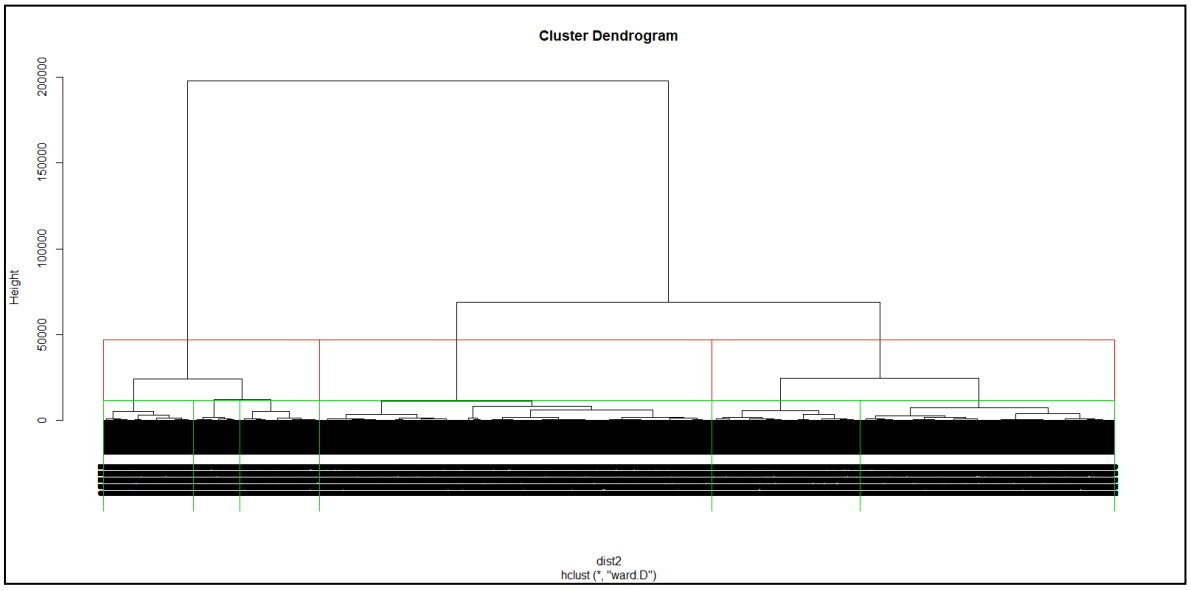
2. Matriz de correlación de variables con agrupamiento en k=3



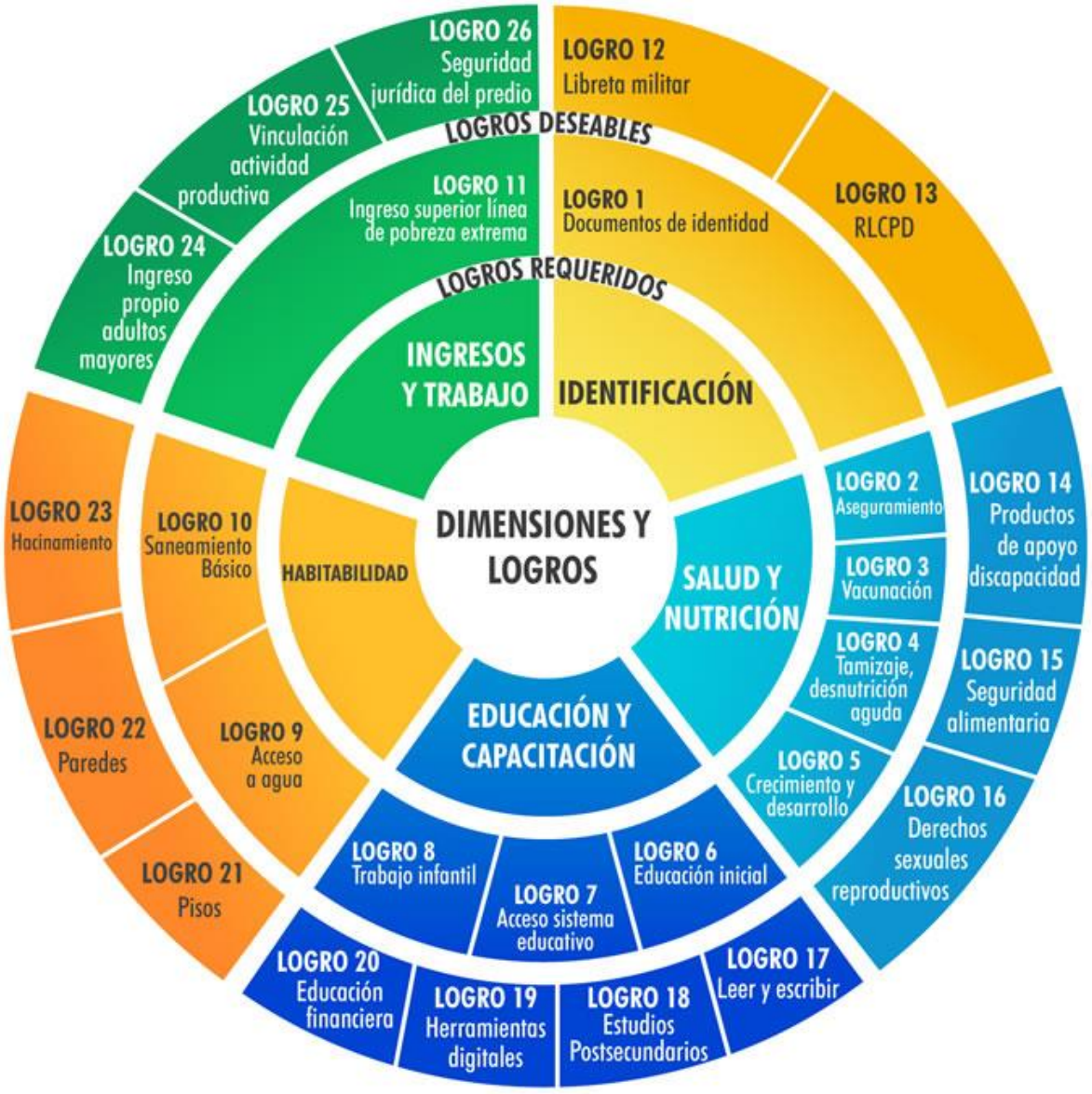
3. Dendograma método Ward por Euclidean




4. Dendograma método Ward por Manhattan



5. Esquema de Logros




Tomado de Dirección de Acompañamiento Familiar y Comunitario
Dirección:
http://www.prosperidadsocial.gov.co/ent/gen/prg/Paginas/Acompa%C3%B1amiento-Familiar-y-Comunitario.aspx

6.