



Pontificia Universidad
JAVERIANA
Cali

**ANÁLISIS DE CONGLOMERADOS EN DESIGUALDADES Y ENFERMEDADES
CARDIO VASCULARES (ECV): UNA BASE PARA LA ASIGNACIÓN DE
RECURSOS EN POLÍTICAS PÚBLICAS EN SALUD**

*Juan Sebastián Meza Buchely
Karen Marcela Rosero Ramos*

*Proyecto Aplicado para optar al título de
Magister en Ciencia de Datos*

Director:
Juan Camilo Paz Roa, Ph.D.

FACULTAD DE INGENIERÍA Y CIENCIAS
MAESTRÍA EN CIENCIA DE DATOS
SANTIAGO DE CALI, DICIEMBRE DE 2024

TABLA DE CONTENIDO

	Pág.
INTRODUCCIÓN.....	7
1. DEFINICIÓN DEL PROBLEMA.....	8
a) PLANTEAMIENTO DEL PROBLEMA	9
b) FORMULACIÓN DEL PROBLEMA.....	11
2. OBJETIVOS DEL PROYECTO	12
a) OBJETIVO GENERAL	12
b) OBJETIVOS ESPECÍFICOS	12
3. MARCO DE REFERENCIA.....	13
3.1. MARCO TEÓRICO.....	13
3.1.1. Determinantes socioeconómicos y salud	13
3.1.2. Desigualdad socioeconómica	14
3.1.3. Enfermedades Cardiovasculares ECV	15
3.1.4. Aprendizaje automático	17
3.1.5. Agrupación de K-means	18
3.1.6. Agrupación geoespacial.....	19
3.2. ANTECEDENTES	20
3.2.1. Antecedentes Internacionales.....	20
3.2.2. Antecedentes Nacionales	22
4. METODOLOGÍA.....	24
4.1. Tipo de investigación	24
4.2. Enfoque metodológico	24
4.3. Procedimiento metodológico.....	24
5. CARACTERÍSTICAS SOCIOECONÓMICAS Y MORTALIDAD DE ECV EN LA REGIÓN PACÍFICA COLOMBIANA EN EL PERIODO 2016-2020.....	28
5.1. Análisis descriptivo de los datos	35
5.1.1. Medidas de tendencia central	35
5.2. Análisis estadístico de variables.....	38
5.3. Diagramas de correlación.....	51
5.4. Preparación y manipulación previa de datos	55
6. CONGLOMERADOS ESPACIALES DE CONDICIONES SOCIODEMOGRAFICAS Y MORTALIDAD POR ECV	57
6.1. Análisis de incidencia de variable	57
6.2 Resultados de clústeres por medio de K-means con ACP	62
6.3. Resultados de clústeres por medio de K-means con T-SNE.....	64
6.4. Resultados de clústeres por medio de Kmeans++ con T-SNE.....	67
6.5. Resultados de clústeres por medio de Gaussian Mixture con T-SNE.....	68

6.6.	Resultados de clústeres por medio de OPTICS con T-SNE	69
6.7.	Resultados de clústeres por medio de K-means (sin reducción de dimensionalidad)	71
6.8.	Evaluación de modelos de clusterización	72
7.	VALIDACIÓN DE RESULTADOS	76
7.1.	Investigaciones previas sobre ECV y determinantes sociales de la salud	76
7.2.	Validación de resultados con expertos en políticas públicas	82
7.	DESPLIEGUE DEL TABLERO DE INFORMACIÓN	94
8.	CONCLUSIONES	98
9.	RECOMENDACIONES	100
	REFERENCIAS	101
	ANEXOS	108

ÍNDICE DE FIGURAS

Figura 1 Balance de datos antes y después de imputación	36
Figura 2 Población.....	38
Figura 3 Población por grupo etario	39
Figura 4 Necesidades Básicas Insatisfechas	40
Figura 5 Cobertura en educación.....	41
Figura 6 Graduados de estudios universitarios.....	41
Figura 7 coberturas de Servicios Públicos.....	42
Figura 8 Valor Agregado	43
Figura 9 IRCA	45
Figura 10 Población formal en el mercado laboral.....	46
Figura 11 Inversión en deporte y gasto público en salud	48
Figura 12 Nacimientos y Defunciones Totales.....	49
Figura 13 Enfermedades cardiovasculares	50
Figura 14 Diagrama de calor de Coberturas en Educación	52
Figura 15 Diagramas de calor de Población.....	53
Figura 16 Correlación de coberturas	54
Figura 17 Correlación variables de inversión y población formal	55
Figura 18 Dispersión de datos agrupados por municipales y departamentos de la zona pacifica colombiana en el periodo 2016-2020.	59
Figura 19 Comportamiento de la tasa por cada 100 mil habitantes de ECV por departamentos de la zona pacífica, 2016-202	60
Figura 20 Varianza de las variables, resultados ACP.....	60
Figura 21 Representación en el plano cartesianos de los dos primeros componentes ACP.	61
Figura 22 Resultados del número de clústeres –método del codo.....	62
Figura 23 Resultados del número de clústeres –método de la silueta	62
Figura 24 Resultados del número de municipios por cada clúster	63
Figura 25 Resultados de la reducción de dimensionalidad por t -SNE	64
Figura 26 Resultados de clusterización mediante algoritmo de K-means.....	65
Figura 27 Resultados de clústeres desagregados por tasa de ECV.....	66
Figura 28 Resultados de clústeres obtenidos mediante el método K-means++	67
Figura 29 Resultados de clústeres obtenidos mediante el método Gaussian Mixture	69
Figura 30 Resultados de clústeres obtenidos mediante el método OPTICS	70
Figura 31 Visualización pestaña tabulación del tablero de información	94
Figura 32 Visualización pestaña de modelación de los datos método K-means	95
Figura 33 Visualización pestaña de modelación de los datos gráfico de radar	96
Figura 34 Visualización de datos de resumen por clúster obtenido	96
Figura 35 Visualización mapa de coropletas	97

ÍNDICE DE TABLAS

Tabla 1 Clasificación 6/67 de enfermedades del sistema circulatorios propuesto por la OPS	15
Tabla 2 Categorías y subcategorías de variables	29
Tabla 3 Estadísticas descriptivas de la base de datos	36
Tabla 4 Incidencia variables independientes vs tasa ECV	57
Tabla 5 Asignaciones de clústeres mediante el método K-means++	68
Tabla 6 Asignaciones de clústeres mediante el método Gaussian Mixture.....	69
Tabla 7 Asignaciones de clústeres mediante el método OPTICS	70
Tabla 8 Resultados clusterización datos originales	75
Tabla 9 Evaluación de modelos mediante métricas de separación y dispersión.....	77

ÍNDICE DE ANEXOS

Anexo A <i>Guion de preguntas para validación de resultados con expertos</i>	108
Anexo B <i>Pantallazos del código de salida de la programación</i>	111

INTRODUCCIÓN

El presente proyecto aplicado tuvo como objetivo identificar la incidencia de las condiciones socioeconómicas en la mortalidad por Enfermedades Cardiovasculares (en adelante ECV) para fortalecer la focalización de recursos de las políticas públicas de salud en la región pacífica colombiana, teniendo como base los resultados obtenidos en el periodo 2016-2020 mediante un análisis de conglomerados geoespacial.

El propósito de la investigación surgió al reconocer las dificultades que presentan las administraciones locales para la focalización de recursos en las políticas públicas de promoción y prevención en salud debido a la ausencia de información actual, verificada y objetiva para la toma de decisiones [1, 2]. En ese orden de ideas, los procesos de focalización de recursos públicos para las políticas públicas en salud se han fundamentado en el análisis de variables como estrato, nivel de ingreso o sexo; sin embargo, dichos mecanismos no aseguran la identificación de los grupos poblacionales más vulnerables [1]. A raíz de ello, la implementación de los programas en salud ha presentado dificultades para asegurar la calidad, eficacia y pertinencia de sus estrategias aplicadas, lo cual afecta, no solo las condiciones de bienestar de la población, sino también la gobernabilidad en el territorio nacional [3]. A partir de ello, el presente estudio utilizó como base los datos de mortalidad por ECV y condiciones socioeconómicas obtenidas en el periodo 2016-2020 para identificar tendencias y conglomerados que permitan a las entidades públicas mejorar el proceso de toma de decisiones, y por ende, de la focalización de los recursos de las políticas públicas en salud.

En ese orden de ideas, el estudio analizó algunas condiciones socioeconómicas de los municipios de la zona pacífica colombiana (integrada por Chocó, Cauca, Valle del Cauca y Nariño) a través de un estudio de conglomerados. La metodología propuesta comenzó con un análisis descriptivo de las variables de interés, posteriormente, se aplicaron técnicas de reducción de dimensionalidad que se justificaron al comprender que estas técnicas permiten eliminar características excesivas y no necesarias en los modelos de aprendizaje y permite la resolución de problemas complejos con modelos simple. Luego de aplicar las técnicas de reducción de dimensionalidad, se procedió a realizar los conglomerados mediante diferentes técnicas de clusterización, incluyendo el método K-means para validar los resultados encontrados. La unidad de análisis del proceso estadístico son los municipios de la zona pacífica colombiana.

El desarrollo del proyecto aplicado, se encuentra ordenada en 10 capítulos de la siguiente manera: en los capítulos 1 y 2 se encuentran la descripción del problema de estudio y los objetivos; posteriormente, en el capítulo 3 se desarrolla el marco teórico y antecedentes del estudio. Seguidamente, se encuentra la metodología de trabajo implementada. En los capítulos 5 a 8 se evidencian los hallazgos de cada uno de los objetivos específicos del estudio. Finalmente, en el capítulo 9 y 10 se mencionan las conclusiones y las áreas de trabajo futuro. De forma general, se puede mencionar que la realización del proyecto aplicado aporta al conocimiento en el área de salud y políticas públicas de salud debido a que los análisis de determinantes socioeconómicos en salud han presentado mayor énfasis en las Enfermedades Diarreicas Agudas (EDA) e Infecciones Respiratorias Agudas (IRA).

1. DEFINICIÓN DEL PROBLEMA

El presente estudio se centra en fortalecer los procesos de asignación de recursos durante la ejecución de las políticas públicas en el sector de salud de Colombia. Cabe reconocer que las políticas públicas se han constituido como una herramienta de planificación ampliamente utilizada por el Estado colombiano para el direccionamiento de recursos y estrategias encaminadas a la resolución de una problemática social. A pesar de la creciente importancia de las políticas públicas en la administración estatal, existe una preocupación sobre las estrategias que se han implementado para la focalización de individuos, familias, o colectivos que cumplan con los requisitos para acceder a los servicios y bienes que se ofrecen a través de las entidades estatales [1].

Generalmente, los procesos de focalización de los recursos, estrategias, actores y programas en las políticas públicas se han realizado mediante criterios como estrato, nivel de ingresos, sexo, edad, entre otros. Pero dichos mecanismos, no siempre logran identificar a los grupos sociales más vulnerables o que requieren la atención y prestación de los bienes y servicios del Estado [1].

En palabras de Arroyave [2], los constantes retos en los procesos de focalización de las políticas públicas han desdibujado su importancia y relevancia como instrumentos de la gestión estatal. Debido a ello, las nuevas tendencias en los procesos de focalización de políticas públicas han generado que las entidades gubernamentales presenten una mayor preocupación por el quién y el dónde se realiza la intervención del Estado [1]. En efecto, las debilidades en las metodologías para la identificación de los beneficiarios de los diferentes programas del Estado han generado fallas en los procesos diagnósticos, con lo cual se resta la efectividad e impacto que pueden tener las políticas públicas en la atención de un problema social [3].

Los inadecuados procesos de focalización de los recursos públicos en las políticas públicas generan no solo problemas de efectividad en el manejo de las finanzas estatales, sino que al mismo tiempo provocan pérdidas de tipo administrativo, problemas en la gobernabilidad, generación de efectos adversos en la resolución de la problemática, estigmas y hasta el rechazo de la población en relación con la implementación de una política pública [1, 2]. Efectivamente, las debilidades en los procesos de focalización se pueden reforzar unos a otros, dificultando aún más el proceso de identificación de beneficiarios [4], generando nuevos retos para la identificación de mecanismos de focalización más efectivos.

Las dificultades en los procesos de focalización de los recursos públicos se explican por una diversidad de factores, sin embargo, una de las razones más frecuentes es la ausencia de datos e información de calidad que facilite la toma de decisiones por parte de la administración pública [5]. La ausencia de datos y modelos estadísticos fiables generan que las entidades tengan prácticas de auto-reporte, identificación errónea de beneficiarios, inadecuada asignación de recursos en territorios y priorización de problemas públicos que no corresponden a la realidad de las comunidades [1, 5].

Frente a este contexto, la administración pública se ha preocupado por mejorar los procesos de focalización a través de diferentes herramientas y estudios que permitan comprender el

estado actual de un problema público [4]. La necesidad de fortalecer las estrategias de focalización de los recursos públicos es un reto en todas las áreas del sector público, sin embargo, existen algunos sectores, como el de salud, que presentan mayores dificultades debido a la gran cantidad de recursos que poseen, la población que atienden, la diversidad de información que se produce diariamente, la ausencia de metodologías de procesamiento de datos, y en algunos casos, la ausencia de profesionales que puedan procesar los datos para generar información valiosa que pueda ayudar a la toma de decisiones por parte de las instancias decisorias de política pública [4, 5].

Al considerar las problemáticas actuales en los procesos de focalización de políticas públicas, especialmente en el sector salud, el presente proyecto aplicado pretende ofrecer información valiosa que facilite la toma de decisiones en instituciones y entidades de salud que tienen la responsabilidad de ejecutar y monitorear políticas públicas. Para ello, el presente estudio se concentrará en analizar las desigualdades sociales y su impacto en el desarrollo de aparición de ECV.

a) PLANTEAMIENTO DEL PROBLEMA

Las ECV se constituye como uno de los grupos de mortalidad con mayor prevalencia en el mundo y en América Latina [6], en efecto, las estadísticas reportadas por la Organización Panamericana de la Salud (OPS) establecen que el incremento de enfermedades no transmisibles implica un reto en las sociedades y en los estados nacionales para diseñar estrategias, programas y políticas públicas encaminadas a la prevención y promoción de la salud [5, 6].

De acuerdo con estudios realizados por la OPS y la OMS, las ECV son desde hace 20 años el principal grupo de enfermedades no transmisibles en el mundo. Según sus estimaciones, este grupo de patologías seguirá constituyéndose como uno de los principales retos en salud pública debido a las condiciones y hábitos de vida que incrementan el riesgo de padecer dichas enfermedades [6]. Esta realidad, también se logra evidenciar en Colombia, ya que según las estadísticas vitales proporcionadas por el DANE (2022) se ha visto un incremento y alta incidencia de las muertes relacionadas con la ECV, situación que se prevé que continúe en los siguientes años [7]. En efecto, las ECV son las enfermedades más mortíferas del mundo, cobrando la vida de cerca de 17.3 millones de personas cada año, cifra que se incrementaría a 23.6 millones para el 2030 [6]. Para Colombia, en el año 2021 se registraron un total de 51.988 muertes relacionadas con la ECV, un incremento del 12% en relación con el año 2020, presentando un aumento en la incidencia de casos en mujeres [7].

El crecimiento de las enfermedades no transmisibles y las ECV en el mundo, ha generado que organizaciones como la OPS recomienden a los gobiernos nacionales el fortalecimiento de las políticas públicas de promoción y prevención, realizando especial énfasis en atender los factores sociales y económicos que pueden estar vinculados al desarrollo y aparición de dichas enfermedades [6]. La necesidad de fortalecer las políticas de promoción y prevención, así como su focalización se centra en los altos costos que pueden estar asociados al tratamiento y diagnóstico del ECV, por ejemplo, un estudio realizado en 2017 demostró que

Colombia era el tercer país con los mayores gastos en el tratamiento de enfermedades cardíacas con 6.4 billones de pesos anuales, solo por detrás de Brasil y México; no obstante, las medidas adoptadas por el Estado no han permitido mejorar las desigualdades regionales que existen en el territorio nacional [8].

Considerando la incidencia de las enfermedades cardiovasculares en el país, y su potencial incremento en los siguientes años, se hace evidente analizar los diferentes factores que pueden influir en el desarrollo de dichas enfermedades. Por ello, si bien es cierto, la aparición y desarrollo de ECV pueden estar relacionadas con una multiplicidad de factores, el presente estudio lo realizará considerando las condiciones socioeconómicas de los individuos. Esta posición ha sido expuesta previamente por autores como Whitehead y Dahlgren [9] y Marmot [10] quienes argumentan que las diferencias sistemáticas en salud están correlacionadas con los factores socioeconómicos, debido a que la magnitud y causas de las enfermedades varían en función de las condiciones de los individuos.

En consecuencia, al mejorar la comprensión de los factores socioeconómicos relacionados con la mortalidad de las ECV, el Estado colombiano puede fortalecer los procesos de focalización, tanto en materia económica como técnica, para fortalecer los programas de prevención y promoción en salud.

Bajo dicho enfoque, el análisis de conglomerados permitirá identificar en qué municipios prevalecen condiciones socioeconómicas que presentan una mayor correlación con la mortalidad de las ECV para que las entidades nacionales, departamentales y municipales puedan mejorar sus intervenciones en los territorios y focalizar sus recursos de acuerdo con las necesidades de cada uno. Cabe reconocer que, en Colombia, todavía existe un reto latente para disminuir las desigualdades en la atención en salud por regiones. Frente a ello, las estadísticas nacionales han evidenciado la existencia de brechas significativas en los territorios, siendo especialmente afectados departamentos de la costa pacífica, atlántica y la Orinoquía [7]. Debido a ello, los retos a nivel nacional invitan al estudio de conglomerados como una herramienta que ofrece información, tanto cualitativa como cuantitativa, que pueden coadyuvar en la toma de decisiones de agentes claves en políticas públicas para fortalecer las Rutas Integrales de Atención en Salud (RIAS) que están enfocadas en la promoción y prevención de la salud [11, 12, 13].

Esta realidad se ha evidenciado en diferentes instituciones y entidades municipales donde se ha establecido la importancia de mejorar los procesos de focalización de recursos de las políticas públicas con el objetivo de fortalecer el impacto y resultados que llegue a tener los programas gubernamentales. Dicha realidad, se hace evidente en las diferentes políticas públicas del Estado colombiano, sin embargo, en el presente estudio se pretende identificar cómo inciden los determinantes socioeconómicos sobre la mortalidad en las enfermedades cardiovasculares mediante un análisis de conglomerados geoespacial, de modo que esto posibilite focalizar estrategias de intervención gubernamental que permita a las entidades encargadas fortalecer los procesos de planificación y al mismo tiempo el manejo o ejecución de los recursos.

Así, la realización del proyecto aplicado beneficia a las secretarías de salud municipales y departamentales quienes tendrán un insumo para la toma de decisiones basado en datos para

la ejecución, planificación y evaluación de políticas públicas de promoción y prevención en la salud. Al mismo tiempo, contribuye a entidades territoriales de salud para la construcción de los informes epidemiológicos de salud (ASIS) porque el estudio aborda los determinantes sociales de la salud como un factor explicativo para la adquisición de ECV. Desde el área de la ciencia de datos, también aporta a la articulación de esta rama del conocimiento a la administración pública en relación al tratamiento de datos y a la generación de estadísticas con enfoque territorial y poblacional para los planes de gestión, estrategias de intervención y planes de acción de los comités interinstitucionales para la puesta en marcha de acciones y recursos públicos.

Para realizar dicho objetivo, se ha propuesto analizar algunas condiciones socioeconómicas y de mortalidad de ECV en la región pacífica durante el periodo 2016-2020 para identificar qué factores tienen una mayor relación con la mortalidad de ECV y a partir de ello, ayudar a los gobiernos departamentales y municipales en la focalización de las políticas públicas de salud en el largo plazo.

b) FORMULACIÓN DEL PROBLEMA

Al considerar la descripción realizada previamente, la pregunta problema del presente estudio es la siguiente: ¿Cómo fortalecer la focalización de recursos en las políticas públicas de Salud a través de un análisis de conglomerados que identifique las condiciones socioeconómicas y su incidencia sobre la mortalidad de ECV en la región pacífica colombiana?

Para la resolución de la pregunta problema, se establecieron las siguientes preguntas de sistematización:

- ¿Cuáles son las principales características sociodemográficas en la región pacífica colombiana durante el periodo 2016-2020?
- ¿Cómo se relacionan los determinantes socioeconómicos y la incidencia de mortalidad de ECV en los diferentes municipios de la región pacífica colombiana durante 2016-2020?
- ¿Cómo validar los resultados obtenidos en los conglomerados espaciales?
- ¿Cuál es la mejor estrategia para la difusión de los resultados del estudio para facilitar la toma de decisiones en materia de política pública en salud?

2. OBJETIVOS DEL PROYECTO

a) OBJETIVO GENERAL

Identificar la incidencia de las condiciones socioeconómicas en la mortalidad por ECV para fortalecer la focalización de recursos de las políticas públicas de salud en la región pacífica colombiana teniendo como base información correspondiente al periodo 2016-2020 mediante un análisis de conglomerados geoespacial.

b) OBJETIVOS ESPECÍFICOS

- Recopilar las características socioeconómicas de los municipios de la región pacífica colombiana en el periodo 2016-2020 mediante revisión de fuentes secundarias que consoliden la base de datos de entrenamiento del modelo de conglomerados.
- Determinar los conglomerados de los municipios de la región pacífica colombiana mediante una comparación de la aplicación de diferentes métodos de aprendizaje automatizado a la base de datos de entrenamiento que permita encontrar las relaciones entre los determinantes socioeconómicos y la incidencia de mortalidad de la ECV.
- Validar los resultados obtenidos en los conglomerados espaciales
- Desplegar un tablero con la información obtenida del análisis de conglomerados que permita facilitar la toma de decisiones en políticas públicas de salud.

3. MARCO DE REFERENCIA

En el presente capítulo, se realiza una revisión teórica sobre los principales determinantes socioeconómicos de la salud, la desigualdad económica y la concepción de las enfermedades cardiovasculares. De igual manera, el capítulo realiza una revisión sintética de investigaciones previas que han abordado la problemática analizada en esta investigación con el objetivo de identificar sus principales aportes al presente estudio.

3.1. MARCO TEÓRICO

3.1.1. *Determinantes socioeconómicos y salud*

La relación entre el estatus socioeconómico (SES) y la salud se ha observado desde tiempos inmemoriales. La bibliografía publicada sobre las disparidades sociales en salud se remonta al siglo XIX. Ya en 1840, Edwin Chadwick informó de una diferencia en la edad de fallecimiento entre las distintas clases sociales [14, 15]. El estudio de Whitehall en la década de 1960 demostró no sólo una asociación entre el grado de empleo y la mortalidad cardiovascular, sino también un gradiente definitivo entre categorías. El personal administrativo presentaba la menor mortalidad, mientras que los oficinistas y otros trabajadores de clase baja tenían la mortalidad más alta [16].

De acuerdo con Marmota, et al [17] el interés por los resultados sanitarios fue progresando lentamente a lo largo de los años, pasando de la mortalidad a la morbilidad y, finalmente, al nivel de los factores de riesgo. Las investigaciones posteriores se centraron en los mecanismos causales que vinculan uno o más componentes del SES con estos resultados sanitarios. A partir de los datos disponibles, Solar e Irwin desarrollaron un marco para comprender cómo se vinculan los determinantes sociales con los resultados sanitarios. El objetivo de toda esta gama de investigaciones era desarrollar intervenciones que redujeran las disparidades socioeconómicas en materia de salud. Según Irwin y Solar [18] sigue siendo un hecho que el estatus socioeconómico fue y será una exposición común estudiada en la investigación de los resultados sanitarios. A menos que se vincule causalmente con el resultado sanitario y se delinee el mecanismo causal, sería de poca utilidad en el desarrollo de intervenciones basadas en la población.

El estatus socioeconómico se refiere a los niveles absolutos o relativos de recursos económicos, poder y prestigios estrechamente asociados con la riqueza de un individuo, comunidad o país. El estatus socioeconómico es un constructo multidimensional que comprende múltiples factores, como ingresos, educación, situación laboral y otros factores. El nivel socioeconómico bajo se asocia con un mayor riesgo de desarrollar y morir por enfermedad cardiovascular (ECV) [19]. Según Havranek, et al [20] específicamente, la Asociación Estadounidense del Corazón señala que el nivel de ingresos, el nivel educativo y la situación laboral a nivel individual y de vecindario se asocian consistentemente con las enfermedades cardiovasculares en los países de altos ingresos. Los factores socioeconómicos pueden afectar el estado de salud directamente a nivel individual y también pueden influir en características más amplias a nivel del hogar, el vecindario o la comunidad, que luego pueden afectar la salud [21].

En esta línea de argumentos Schultz, et al [21] afirman que los factores socioeconómicos afectan la capacidad de una persona para participar en actividades de salud, pagar atención médica y vivienda y controlar el estrés. Por ejemplo, el empleo proporciona ingresos, lo que permite el acceso a vivienda, educación, cuidado infantil, alimentación, atención médica y otras necesidades. A nivel comunitario, los barrios de bajos ingresos tienen menos probabilidades de tener acceso a atención médica de alta calidad. Los factores socioeconómicos también pueden interactuar o confundir las relaciones entre otras variables y la salud. Por ejemplo, los efectos combinados del nivel socioeconómico y la raza/etnicidad o el sexo pueden influir en la salud de manera diferente entre diferentes grupos.

3.1.2. Desigualdad socioeconómica

Las diferencias socioeconómicas en salud se han descrito desde los siglos XVI y XVII [22], pero sólo recientemente reducirlas ha sido fundamental para las políticas de salud pública en muchos países del mundo [23]. De esta forma se menciona que la diferenciación social hace alusión a la existencia de cualidades y roles diversos, de esta forma todos los grupos y sociedades se encuentran diferenciados de alguna manera. En esta medida se afirma que la desigualdad social es la jerarquización de esas diferencias y por ende es la condición por la que los individuos tienen un acceso inequitativo a recursos valorados. Cabe destacar que la misma tiende a incrementar con el tamaño de los grupos. Con base en lo anterior, la fuente asegura que todas las sociedades cuentan con un determinado nivel de desigualdad en un momento dado.

De acuerdo con la ACNUR [24] la desigualdad no es única ni afecta a un solo aspecto de la vida de los individuos, sino que se puede hablar de diferentes tipos: desigualdad social, hace alusión al trato diferente que recibe una persona a consecuencia de su posición social, su situación económica, la religión que profesa, su género, la cultura de la que proviene o sus preferencias sexuales, entre otros elementos; desigualdad económica, hace alusión a la distribución de la riqueza entre los individuos, de esta forma, las diferencias de ingresos entre personas implica un problema de acceso a bienes y servicios, lo que se ve reflejado en quienes tienen menos ingresos; desigualdad educativa, que hace referencia a un elemento que tiene como base la desigualdad social y económica, debido a que supone que los individuos no cuentan con las mismas oportunidades al momento de acceder a una formación; desigualdad de género, como un elemento que se genera cuando la persona no cuenta con acceso a las oportunidades de igual manera que otra persona de otro sexo; finalmente, la desigualdad legal, que se genera cuando las leyes de los tribunales favorecen más a unos que a otros.

Desde esta perspectiva, la CEPAL [25] plantea que la desigualdad corresponde a una característica histórica y estructural propia de las sociedades latinoamericanas y caribeñas, la cual se ha prolongado incluso en etapas de crecimiento económico. De esta manera, se resalta que pese a los avances orientados al desarrollo social y a la promoción de políticas activas de carácter redistributivo e incluyente, aún persisten elevados niveles de desigualdad que conspiran en contra del desarrollo y constituyen una importante barrera para la erradicación de la pobreza. En efecto, se establece que la desigualdad dificulta que los individuos asciendan socialmente y consigan mayores niveles de bienestar.

Durante las últimas tres décadas, los estudios epidemiológicos han confirmado la existencia de desigualdades socioeconómicas en una variedad de resultados de salud, incluida la mortalidad prematura, las enfermedades cardiovasculares, la obesidad, la diabetes, la mala salud autodeclarada y los cánceres relacionados con el tabaquismo, y han explorado posibles mecanismos que vinculan una posición socioeconómica más baja a una salud más deficiente. De acuerdo con Lawlor [26] existe evidencia de que una peor posición socioeconómica se asocia con una peor morbilidad, mortalidad y salud autoinformada en las personas mayores.

3.1.3. Enfermedades Cardiovasculares ECV

De acuerdo con la Organización Mundial de la Salud [27] las Enfermedades Cardiovasculares son la principal causa de muerte a nivel mundial, pues anualmente fallecen más personas por ECV que por cualquier otra enfermedad. La definición que se plantea señala que las ECV constituyen un grupo de trastornos del corazón y de los vasos sanguíneos, donde se incluyen enfermedades como: la enfermedad coronaria, que hace referencia a los vasos sanguíneos involucrados en el musculo cardiaco; la enfermedad cerebrovascular, que involucra los vasos sanguíneos que irrigan el cerebro; la enfermedad arterial periférica, como un padecimiento asociado a los vasos sanguíneos que irrigan los brazos y las piernas; la enfermedad cardiaca reumática, como el daño al musculo cardiaco y a las válvulas cardiacas, derivado de una fiebre reumática y bacterias estreptocócicas; cardiopatía congénita, asociada a los defectos congénitos que impactan negativamente en el desarrollo y funcionamiento del corazón, lo que es consecuencia de malformaciones en la estructura del órgano desde el nacimiento; trombosis venosa profunda y embolia pulmonar, constituida por coágulos de sangre en las venas de las piernas que se pueden desprender y moverse hasta el corazón o pulmones.

Como se menciona previamente, las enfermedades cardiovasculares incluyen una diversidad de patologías, por ello, entidades internacionales como la OPS ha reconocido que un reto para los procesos analíticos y de investigación en materia de salud es la diversidad de patologías. A raíz de ello, desde el año 1987 la entidad con varios países de las Américas realizó proyectos de investigación sobre perfiles de salud/análisis de mortalidad para lograr una clasificación de enfermedades según signos, síntomas y estados epidemiológicos de la enfermedad. A partir de ello, surge la lista de clasificación de enfermedades 6/67 de la OPS, la cual se utiliza en las estadísticas vitales del DANE, y también se utilizó, para la recopilación de los datos para el presente trabajo. En resumen, la lista 6/67 agrupa a las enfermedades del sistema circulatorio en el grupo 3, el cual se presenta a continuación.

Tabla 1

Clasificación 6/67 de enfermedades del sistema circulatorios propuesto por la OPS

Código		Enfermedades del Sistema Circulatorio
3	3.01	Fiebre reumática aguda y enfermedades cardiacas reumáticas crónicas
	3.02	Enfermedades hipertensivas

3.03	Enfermedades isquémicas del corazón
3.04	Enfermedad cardiopulmonar, de la circulación pulmonar y otras formas de enfermedades del corazón
3.05	Paro cardíaco
3.06	Insuficiencia cardíaca
3.07	Enfermedades cerebrovasculares
3.08	Aterosclerosis
3.09	Otras enfermedades del sistema circulatorio

Fuente: Organización Mundial de la Salud [27]

La sintomatología de cada enfermedad depende de su patología. Por ejemplo, según la OPS, los accidentes cerebrovasculares generalmente se califican como eventos agudos y son ocasionados principalmente por un bloqueo que imposibilita que la sangre fluya de manera adecuada hacia el corazón o el cerebro. En esta medida, la causa más frecuente es la acumulación de depósitos de grasa encontrados en las paredes internas de los vasos sanguíneos que irrigan el corazón o el cerebro. Pese a las diferentes causales de las enfermedades cardiovasculares, es preciso señalar que existen factores de riesgo que comúnmente conducen al desarrollo y prevalencia de este tipo de padecimientos.

Los factores de riesgo conductuales más importantes de enfermedades cardíacas y accidentes cerebrovasculares son una dieta poco saludable, la inactividad física, el consumo de tabaco y el consumo nocivo de alcohol. Los efectos de los factores de riesgo conductuales pueden manifestarse en individuos como presión arterial elevada, glucosa en sangre elevada, lípidos en sangre elevados y sobrepeso y obesidad. Estos “factores de riesgo intermedios” se pueden medir en los centros de atención primaria e indican un mayor riesgo de ataque cardíaco, accidente cerebrovascular, insuficiencia cardíaca y otras complicaciones [27].

Uno de los factores que también se relacionan con el desarrollo de este tipo de enfermedades son las desigualdades sociales que a su vez son consecuencia de las condiciones socioeconómicas de los individuos. De esta manera, Pérez [28] afirma que las condiciones socioeconómicas influyen en el desarrollo y la prevalencia de los factores de riesgo cardiovascular, los cuales pueden variar entre países y con el paso del tiempo, pero que se manifiestan en mayor medida en los adultos mayores.

Al respecto, Davari [29] afirma que el nivel socioeconómico tiene un efecto medible y significativo en la salud cardiovascular. Los factores de riesgo biológico, conductual y psicosocial que prevalecen en personas desfavorecidas acentúan el vínculo entre el nivel socioeconómico y las enfermedades cardiovasculares. Los factores socioeconómicos, como el empleo y los ingresos, también afectan la tasa de mortalidad a través del impacto en los factores de riesgo relacionados con el estilo de vida antes y después del ataque cardíaco. Después del ataque cardíaco, los grupos de nivel socioeconómico más bajo enfrentarían consecuencias de salud más graves a largo plazo. Se han registrado desigualdades socioeconómicas en casi todos los países occidentales en la prevalencia y ocurrencia de muertes por ECV. La inequidad socioeconómica es uno de los principales desafíos en materia de ECV.

3.1.4. Aprendizaje automático

El aprendizaje automático (ML), ha crecido rápidamente en los últimos años en el contexto del análisis de datos y la informática que normalmente permite que las aplicaciones funcionen de manera inteligente [30]. El aprendizaje automático generalmente proporciona a los sistemas la capacidad de aprender y mejorar a partir de la experiencia automáticamente sin estar programados específicamente y generalmente se lo conoce como las últimas tecnologías más populares en la cuarta revolución industrial (Industria 4.0) [31].

De acuerdo con Mohammed, Badruddin y Bashier [32] la Industria 4.0 suele ser la automatización continua de las prácticas industriales y de fabricación convencionales, incluido el procesamiento exploratorio de datos, utilizando nuevas tecnologías inteligentes como la automatización del aprendizaje automático. Por lo tanto, para analizar inteligentemente estos datos y desarrollar las correspondientes aplicaciones del mundo real, los algoritmos de aprendizaje automático son la clave.

De acuerdo con Sarker [33] los algoritmos de aprendizaje automático se dividen principalmente en cuatro categorías: aprendizaje supervisado, aprendizaje no supervisado, aprendizaje semisupervisado y aprendizaje por refuerzo. El aprendizaje supervisado suele ser la tarea de aprendizaje automático para aprender una función que asigna una entrada a una salida en función de pares de entrada-salida de muestra, por ende, emplea datos de entrenamiento etiquetados e incluye un enfoque basado en tareas; el aprendizaje no supervisado analiza conjuntos de datos sin etiquetar sin necesidad de interferencia humana, es decir, radica en un proceso basado en datos; el aprendizaje semisupervisado, el cual se define como una hibridación de los métodos supervisados y no supervisados, ya que opera tanto con datos etiquetados como sin etiquetar; el aprendizaje por refuerzo, que se fundamenta en un tipo de algoritmo automático que permite a los agentes de software y a las máquinas evaluar automáticamente el comportamiento óptimo e un contexto o entorno particular para mejorar su eficiencia.

Según la fuente mencionada, la eficacia y eficiencia de una solución de aprendizaje automático dependen de la naturaleza y las características de los datos y del rendimiento de los algoritmos de aprendizaje. En el área de los algoritmos de aprendizaje automático, existen técnicas de análisis de clasificación, regresión, agrupación de datos, ingeniería de características y reducción de dimensionalidad, aprendizaje de reglas de asociación o aprendizaje por refuerzo para construir eficazmente sistemas basados en datos.

En esta medida, [34] afirma que dentro del aprendizaje automático, se emplean comúnmente dos enfoques básicos: el aprendizaje supervisado y el aprendizaje no supervisado. La principal diferencia es que uno utiliza datos etiquetados para ayudar a predecir resultados, mientras que el otro no. Sin embargo, existen algunos matices entre los dos enfoques y áreas clave en las que uno supera al otro.

Por otro lado, [35] menciona que el aprendizaje supervisado es un enfoque de aprendizaje automático que se define por el uso de conjuntos de datos etiquetados. Estos conjuntos de datos están diseñados para entrenar o "supervisar" algoritmos para clasificar datos o predecir

resultados con precisión. Utilizando entradas y salidas etiquetadas, el modelo puede medir su precisión y aprender con el tiempo.

Así pues, el autor mencionado señala que el aprendizaje supervisado se puede dividir en dos tipos de problemas cuando se extraen datos, tales como:

- **Los problemas de clasificación** utilizan un algoritmo para asignar con precisión datos de prueba en categorías específicas y se pueden utilizar algoritmos de aprendizaje supervisado para clasificar el spam en una carpeta separada de su bandeja de entrada. Los clasificadores lineales, las máquinas de vectores de soporte, los árboles de decisión y los bosques aleatorios son tipos comunes de algoritmos de clasificación.
- **La regresión** es otro tipo de método de aprendizaje supervisado que utiliza un algoritmo para comprender la relación entre variables dependientes e independientes. Los modelos de regresión son útiles para predecir valores numéricos basados en diferentes puntos de datos, como las proyecciones de ingresos por ventas para una empresa determinada. Algunos algoritmos de regresión populares son la regresión lineal, la regresión logística y la regresión polinómica.

Al respecto, Sarker [33] afirma que la eficacia y eficiencia de una solución de aprendizaje automático dependen de la naturaleza y las características de los datos y del rendimiento de los algoritmos de aprendizaje. En el área de los algoritmos de aprendizaje automático, existen técnicas de análisis de clasificación, regresión, agrupación de datos, ingeniería de características y reducción de dimensionalidad, aprendizaje de reglas de asociación o aprendizaje por refuerzo para construir eficazmente sistemas basados en datos.

3.1.5. Agrupación de K-means

De acuerdo con Oyana [36] la agrupación delinea la operación de objetos dentro de un conjunto de datos que tienen cualidades similares en grupos homogéneos. Permite el descubrimiento de similitudes y diferencias entre patrones con el fin de derivar conclusiones útiles sobre ellos. Determinar la estructura o los patrones dentro de los datos es un componente importante en la clasificación y visualización, lo que permite la extracción geoespacial de conjuntos de datos de gran volumen. Si bien se han desarrollado muchas técnicas de agrupación a lo largo de los años, la técnica de agrupación es una técnica considerada como común y flexible. La función principal del algoritmo k -means es dividir los datos en k subgrupos disjuntos y luego la calidad de estos grupos se mide mediante diferentes métodos de validación.

La agrupación de K-means es según Kim, Kim y Yi [37] un algoritmo de aprendizaje automático no supervisado que agrupa datos en k grupos. Más específicamente, el algoritmo primero selecciona aleatoriamente k muestras del conjunto de datos para inicializar los k centroides del grupo. Luego, cada punto de datos se asigna al grupo correspondiente al centroide más cercano medido por la distancia euclidiana. Luego, las ubicaciones de los centroides se actualizan para que correspondan a la media de todos los puntos de datos en el

grupo correspondiente. El proceso de asignar puntos de datos al centroide del grupo más cercano y ajustar las ubicaciones de los centroides se repite hasta que los centroides del grupo sean estables. Por lo tanto, la agrupación k-means intenta identificar agrupaciones naturales en los datos.

A grandes rasgos Sharma [38] señala que K-Means tiene como objetivo dividir las observaciones en un número predefinido de grupos (k) en los que cada punto pertenece al grupo con la media más cercana. Comienza seleccionando aleatoriamente k centroides y asignando los puntos al grupo más cercano, luego actualiza cada centroide con la media de todos los puntos del grupo. El método de agrupamiento de k-medias considera la correlación espacial entre las muestras que pertenecen al mismo grupo y las relaciones estadísticas entre los puntos de observación dentro del conjunto de datos.

3.1.6. Agrupación geoespacial

La agrupación geoespacial es según Wang y Wang [39] el método de agrupar un conjunto de objetos espaciales en grupos llamados "clústeres". Los objetos dentro de un grupo muestran un alto grado de similitud, mientras que los grupos son lo más diferentes posible. Uno de los propósitos de la agrupación es hacer una generalización y revelar una relación entre atributos espaciales y no espaciales. Otro aspecto importante de los datos geoespaciales es su dimensión temporal, que refleja los cambios y la dinámica de los fenómenos a lo largo del tiempo. La agrupación temporal tiene como objetivo agrupar los datos geoespaciales en función de sus patrones temporales, como tendencias, ciclos o anomalías.

Con base en lo anterior, Zhong, Li y Zhu [40] plantean que la agrupación temporal puede ayudar a descubrir regularidades temporales, anomalías y evolución de los grupos, así como a comprender las causas y consecuencias de los cambios. Algunos de los enfoques recientes en la agrupación temporal incluyen la agrupación de series temporales, la agrupación de trayectoria y la agrupación espacio-temporal. La agrupación de series temporales utiliza varias medidas de similitud, funciones de distancia o métodos de extracción de características para comparar y agrupar datos de series temporales, como imágenes de satélite o lecturas de sensores. La agrupación de trayectorias utiliza varias técnicas, como la segmentación, la alineación o la compresión, para comparar y agrupar datos de trayectoria, como pistas GPS o movimientos de animales. La agrupación espacio-temporal combina las dimensiones espaciales y temporales de los datos, y utiliza varios modelos, como cadenas de Markov, modelos ocultos de Markov o factorización tensorial, para capturar las dependencias espacio-temporales y las variaciones de los clústeres.

Por otra parte, Pedersen, et al [41] sugiere que la agrupación espectral-espacial puede mejorar la precisión, robustez e interpretabilidad de los grupos, así como reducir los efectos del ruido y los valores atípicos. Algunos de los enfoques recientes en la agrupación espectral-espacial incluyen métodos basados en gráficos, métodos de aprendizaje profundo y métodos basados en superpíxeles. Los métodos basados en gráficos utilizan gráficos para representar los datos y sus relaciones, y aplican algoritmos de corte de gráficos o agrupamiento de gráficos para particionar el gráfico. Los métodos de aprendizaje profundo utilizan redes neuronales para aprender características y representaciones de alto nivel de los datos y, a continuación, aplican algoritmos de agrupación en clústeres en las características aprendidas. Los métodos

basados en superpíxeles utilizan superpíxeles, que son pequeñas regiones de píxeles con características similares, como unidades básicas para la agrupación en clústeres, y luego los fusionan o dividen de acuerdo con algunos criterios.

3.2. ANTECEDENTES

3.2.1. Antecedentes Internacionales

Para llevar a cabo el presente proyecto es importante identificar estudios previos que han analizado la temática y la han abordado desde diferentes regiones encontrando elementos importantes que pueden ser claves para el estudio en curso. De esta manera es preciso hacer referencia a la investigación elaborada por Aungkulanon, et al [42] donde se aborda el tema de diferenciales de mortalidad y privación socioeconómica a nivel de área en Tailandia: resultados del análisis de componentes principales y del análisis de conglomerados. De este modo el artículo plantea como objetivo examinar los patrones de mortalidad en diferentes estratos socioeconómicos en Tailandia. Para la realización del estudio, se aplicó un análisis transversal considerando los datos obtenidos en el Censo de Población y Vivienda de 2010 sobre la condición socioeconómica a nivel de área y se comparó con los indicadores de mortalidad de 2010 de la base de datos del registro civil a nivel de superdistritos. Se utilizó el análisis de componentes principales para construir un índice de privación socioeconómica y el análisis de conglomerados K-mean (método de agrupamiento) para agrupar el estatus socioeconómico y la mortalidad por causa específica. El estudio reconoce la importancia de aplicar el método de ACP para la construcción de índices poblacionales al reducir la dimensionalidad de los datos y elimina la colinealidad de los datos, sin embargo también se reconoce algunas debilidades en relación a la interpretación del índice y a su sensibilidad a los datos ruidosos.

Los resultados del estudio analizado permitieron evidenciar un exceso de mortalidad por todas las enfermedades, excepto el cáncer colorrectal, entre los superdistritos con un nivel socioeconómico bajo. La agrupación espacial fue evidente en la distribución del estatus socioeconómico y las tasas de mortalidad. El análisis de conglomerados reveló que los superdistritos predominantemente urbanos tendían a tener una tasa de mortalidad estandarizada por todas las causas baja, pero una tasa de mortalidad específica por cáncer colorrectal alta. Las muertes por cáncer de hígado, diabetes y enfermedades renales eran frecuentes en los superdistritos de nivel socioeconómico bajo, que albergaban a un tercio de la población total tailandesa. Desde esta perspectiva, el principal aporte del presente estudio radica en las conclusiones derivadas del mismo, donde se logró demostrar que las zonas socialmente desfavorecidas presentan un exceso de muertes globales y por causas específicas. Las poblaciones que viven en zonas más prósperas, a pesar de la baja mortalidad general, siguen teniendo muchas muertes evitables, como el cáncer colorrectal. Estos hallazgos justifican futuros estudios epidemiológicos que investiguen las diversas causas del exceso de muertes en zonas no desfavorecidas y la aplicación de políticas para reducir la brecha de mortalidad entre zonas ricas y pobres.

En esta medida, se encuentra el estudio de Muthalaly, et al [43] donde se definen las relaciones entre diferentes dominios de la salud socioeconómica, los factores de riesgo

cardiovascular tradicionales y los eventos cardiovasculares. La investigación aborda un estudio transversal de las áreas de gobierno local de Victoria, Australia. Se utilizaron datos de una encuesta de salud de la población combinados con datos de eventos cardiovasculares derivados de datos hospitalarios y gubernamentales. A partir de 22 variables se generaron cuatro dominios socioeconómicos: nivel educativo, bienestar económico, lejanía y salud psicosocial. El resultado primario fue un compuesto de IAMCEST, IAMCEST, insuficiencia cardiaca y muertes cardiovasculares por cada 10.000 personas. Se utilizaron la regresión lineal y el análisis de conglomerados para evaluar las relaciones entre los factores de riesgo y los eventos.

Los resultados del estudio permitieron conocer que todos los dominios socioeconómicos estaban asociados con la carga de factores de riesgo tradicionales, incluida la hipertensión, el tabaquismo, la mala alimentación, la diabetes y la obesidad. Por otro lado, se encontró que el bienestar financiero, el nivel educativo y la lejanía estaban correlacionados con los eventos cardiovasculares desde un análisis univariado. En cuanto a los resultados de ajuste multivariado por edad y sexo se pudo establecer que el bienestar financiero, el bienestar psicosocial y la lejanía se vincularon con eventos cardiovasculares, en tanto que el nivel educativo no. En el análisis de los factores de riesgo tradicionales, se pudo establecer que el bienestar financiero y la lejanía permanecieron correlacionados con los eventos cardiovasculares. En conclusión, el estudio representa un importante aporte en lo que concierne a sus demostraciones, ya que se comprobó una fuerte asociación entre los determinantes socioeconómicos del bienestar financiero, los niveles de educación, el bienestar psicosocial, la lejanía y las tasas de eventos cardiovasculares a nivel del área del gobierno local. Finalmente, el estudio también identificó agrupaciones relacionadas con factores de salud socioeconómicos deficientes entre áreas de gobierno local que sugieren áreas de alto riesgo y que pueden proporcionar un plan para implementar intervenciones políticas interdisciplinarias.

Finalmente, se puede mencionar el estudio realizado por Augusto y Lanza [49] quienes realizaron un análisis espacial de la mortalidad cardiovascular y sus factores asociados en el mundo. Para el desarrollo del trabajo se utilizaron como variables la mortalidad por enfermedades cardiovasculares, los grupos poblacionales por curso de vida, el PIB per Cápita y la escolaridad promedio en cada uno de los países. La técnica para el análisis de los datos fue la regresión por el método de cuadrados ordinarios y estadística espacial (Spatial Lag Model) para la generación de un mapa de calor en cada uno de las naciones. El estudio también aplica métodos de reducción de dimensionalidad como el ACP y el t-SNE debido a que permite reducir los datos y mejorar el rendimiento de algoritmos especialmente cuando se trabajan con bases de datos extensas. Al realizar los estudios de reducción de dimensionalidad se argumenta que dichas técnicas se deben a analizar a cada caso concreto porque la eliminación de datos redundantes no siempre implica que el funcionamiento del modelo siempre sea más consistente.

Al aplicar las técnicas de regresión y análisis de los datos, el estudio encontró que la mortalidad por ECV ha presentado un incremento en los países en desarrollo, mientras que en los países desarrollados se ha generado un estancamiento en la disminución de estas patologías. En concreto, el estudio evidenció que variables como el PIB per Cápita, el nivel de urbanización de los países y el nivel escolar de los individuos se constituyen como las tres

variables más importantes para la explicación de la mortalidad por ECV. En efecto, el estudio establece que la escolaridad fue la variable estadísticamente más significativa en sus modelos estadísticos, lo cual evidencia la necesidad de fortalecer los procesos de educación y sensibilización en las políticas de salud alrededor del mundo.

3.2.2. Antecedentes Nacionales

El estudio elaborado por Parra [11] plantea como propósito general establecer las desigualdades sociales en la mortalidad por enfermedades cardiovasculares en los municipios de los Departamentos del triángulo del café, para lo que se busca describir el comportamiento de las tasas de mortalidad por enfermedades isquémicas cardíacas, hipertensivas y accidentes cerebrovasculares, empleando una metodología adecuada para medir las desigualdades sociales en la mortalidad a causa de las mencionadas enfermedades, según indicadores económicos en los municipios de los Departamentos de Caldas, Quindío, y Risaralda. De esta manera, para analizar la mortalidad y efectuar el cálculo de las tasas y Necesidades Básicas Insatisfechas (NBI) fue necesario recurrir a la información del Departamento Nacional de Estadística; el Producto Interno Bruto (PIB) y la implementación de la técnica multivariada de análisis de conglomerados, para agrupar los municipios en grupos, teniendo en cuenta similitudes en sus características.

Los resultados de la investigación evidenciaron la identificación de tres clases: en primer lugar, se encontraba la clase uno, en la que incursionaban aquellos municipios que tenían un PIB per cápita más alto, el NBI más bajo, con un promedio elevado de tasa de mortalidad por accidente cerebrovascular, un promedio de tasa de mortalidad por hipertensión más bajo; en segundo lugar, la clase dos, donde se presentó un PIB per cápita más bajo y el promedio de la tasa de mortalidad por isquemia cardíaca más alto; finalmente, la clase tres, donde se encontraban los municipios con un NBI más elevado y con ello arrojaron un resultado promedio más alto en la mortalidad por hipertensión e isquemia cardíaca. Desde lo anterior, el aporte del mencionado estudio es crucial en el proyecto en curso, pues demuestra una relación entre los determinantes socioeconómicos con la prevalencia de enfermedades cardiovasculares y desde sus conclusiones se logró verificar y plantear que la conformación de los conglomerados permite establecer un vínculo entre un NBI alto con tasas de mortalidad por hipertensión e isquemia cardíaca, en tanto que se encontró una relación entre un PIB per cápita alto y NBI bajo con la tasa de mortalidad por accidente cerebrovascular.

El estudio de Pérez y Achcar [44] está muy relacionado con el proyecto en curso, debido a que presenta un propósito similar en la región Pacífico, donde se analizan de manera espacial y temporalmente la mortalidad por enfermedades cardiovasculares en la región Pacífico de Colombia para el periodo de 2002-2015 y su asociación con algunos indicadores socioeconómicos municipales usando técnicas estadísticas de análisis espacial. En este sentido, el estudio incluyó una metodología cuantitativa con un tipo de investigación descriptivo-ecológico en los 177 municipios de la región Pacífico, utilizando los datos de mortalidad por ECV, bajo códigos 100-199 de la Clasificación Internacional de Enfermedades (CIE-10), y siete indicadores socioeconómicos municipales. Con base en ello también se incluyeron cálculos sobre las tasas de mortalidad crudas y estandarizadas, las

cuales estuvieron asignadas según sexo, para las ECV y sus principales causas, empleando una modelación de conteo de defunciones utilizando modelos jerárquicos Bayesianos.

El aporte importante del estudio se encuentra en los principales hallazgos, donde fue posible demostrar que durante el periodo 2002-2015 se presentaron tasas estandarizadas de mortalidad por enfermedades cardiovasculares, evidenciando una tendencia descendente en hombres (129,0 a 119,3) y en mujeres (129,0 a 110,0), siendo las enfermedades isquémicas del corazón, seguidas de las cerebrovasculares, las principales causas de muerte en la región. De la misma manera, se logró demostrar que en general, el riesgo de mortalidad por ECV fue mayor en aquellos municipios más desfavorecidos en lo que corresponde a los aspectos económico y social.

Finalmente, se puede mencionar el trabajo realizado por Pico et al [50] que tuvo como objetivo analizar espacialmente el riesgo cardiovascular de la población adulta mayor del municipio de Cali. Para ello, se utilizó una metodología cuantitativa de corte transversal con una muestra de 4092 adultos mayores de 55 años donde se analizaron variables sociodemográficas con SPSS versión 24 y se utilizó el software GeoCODE para estandarizar las direcciones.

El estudio analizó una serie de variables donde se encuentran el sexo, los rangos de edad, el régimen de salud de los adultos mayores, el estrato y condiciones de vida como el consumo de tabaco, consumo de frutas y verduras y actividad física. De acuerdo con los resultados del estudio, se evidenció que de la muestra el 84.1% presentaba un bajo riesgo cardiovascular; el 13.1% riesgo medio y el 2.7% alto. Igualmente, el riesgo cardiovascular era más alto en hombres en comparación de las mujeres, más aún, se reconoció que la población analizada presentaba hábitos de vida saludable. Al ejecutar el análisis espacial, el estudio encontró que el riesgo cardiovascular alto se presentó en las comunas 2, 6, 7, 13, 16 y 19 de Cali.

Los resultados de este estudio también aportan a la presente investigación porque realiza una aproximación al estudio espacial de determinantes socioeconómicos en la región pacífica colombiana. De igual manera, se analizan algunas variables de importancia para el estudio como la afiliación al SGSSS, los hábitos de vida saludable y el estrato. Igualmente, ofrece una visión de las condiciones socioeconómicas en la ciudad de Cali, el cual se incluye también en la presente investigación.

4. METODOLOGÍA

En la presente sección se establecen el tipo de estudio, enfoque y procedimientos a utilizar para el desarrollo de la investigación. De igual manera, se realiza una breve descripción de los pasos a ejecutar por cada objetivo específico con la finalidad de delimitar una ruta de acción.

4.1. Tipo de investigación

El desarrollo del presente trabajo es de tipo teórico-práctico porque en un primer momento, la fase de recolección de información y datos se concentró en bases secundarias y de libre acceso, a partir de ello, se procede a realizar asociaciones entre los factores socioeconómicos y la prevalencia de mortalidad de ECV. Luego de realizar la primera fase de análisis, los investigadores desplegaron un tablero en línea con el objetivo de que las entidades e instituciones interesadas en los datos arrojados por el estudio puedan visualizar los datos y generar una aplicabilidad a la información obtenida por el presente estudio.

4.2. Enfoque metodológico

Considerando la naturaleza del estudio, se propone el uso de la metodología Cross-Industry Standard Process for Data Mining (CRISP-DM). De acuerdo con Wirth [45] esta metodología incluye seis etapas que determinan los procesos a seguir dentro de un proyecto de análisis de datos; las fases del método CRISP-DM incluyen: comprensión del fenómeno, comprensión de los datos, preparación de los datos, modelados, evaluación y despliegue.

De igual manera, el presente estudio utilizó técnicas de datos espaciales que centran en la representación de puntos, líneas o áreas geográficas en el geoespacio para determinar que fenómenos se presentan en dichas zonas. En consecuencia, este tipo de técnicas se centran en analizar patrones de puntos, procesos estacionarios, representación espectral y simulaciones geoestadísticas gaussianas [2].

4.3. Procedimiento metodológico

Para la realización del presente estudio, en una primera etapa, fue necesario comprender y entender cómo las características socioeconómicas pueden influir en la prevalencia de mortalidad por ECV, este procedimiento requiere de una comprensión cualitativa y cuantitativa, que permite determinar estadísticas descriptivas para analizar las variables de interés y encontrar en qué regiones de la zona pacífica colombiana se evidencian mayores factores de riesgo para la prevalencia de mortalidad de las ECV.

Comprender la naturaleza de los datos, permitió alcanzar el primer objetivo específico, el cual también incluye la recopilación, sistematización y consolidación de la base de datos de entrenamiento del modelo de conglomerados. Las actividades para alcanzar este objetivo incluyeron:

- (i) Consolidar los datos del modelo mediante una búsqueda minuciosa en repositorios institucionales como DANE, Ministerio de Educación, Ministerio de Salud, informes de epidemiología ASIS y Datos Abiertos Colombia, entre otros que tienen información que aporta al desarrollo de la presente investigación. La información que se recolectó mediante esta base de datos, incluye tanto los componentes sociodemográficos, así como los índices de mortalidad que son calculados para cada municipio tanto en los informes ASIS como en las estadísticas vitales del DANE.
- (ii) Todos los datos recolectados en la fase de búsqueda requirieron un posterior análisis estadístico exploratorio y descriptivo, comprensión de las estadísticas y tasas, y limpieza de los datos para consolidar la base de datos que será el insumo principal para la realización de los modelos de conglomerados en la región pacífica.
- (iii) Con la ayuda de las estadísticas descriptivas, los investigadores pudieron plantear algunas hipótesis iniciales que permitan establecer las relaciones entre las variables analizadas y el espacio donde se presentan dichas correlaciones.

Al finalizar el proceso de consolidación de las bases de datos, se procedió a realizar los modelos de conglomerados. Frente al análisis de conglomerados se debe reconocer que los procedimientos de autocorrelación espacial permiten medir la relación entre la misma variable y sus vecindades o contigüidades en múltiples espacios [2]. En palabras de Getis [34], la autocorrelación espacial “se enfoca en mostrar las relaciones dentro de una variable a lo largo del espacio, en lugar de mostrar las relaciones entre variables” (p. 12). En ese orden de ideas, la autocorrelación espacial es un indicador de estadística descriptiva que permite medir las formas y manera cómo se distribuyen los fenómenos en un espacio geográfico [46], por ello, la autocorrelación permite medir el grado en qué una variable geográfica está correlacionada con ella misma en dos puntos o zonas diferentes del área de estudio, para medir la similitud en un área concreta. Así mismo, este tipo de análisis permitirá indagar puntos de autocorrelación positiva o negativa [47].

Al desarrollar los conglomerados, se da alcance al segundo objetivo específico. Para ello, las acciones que se deben desarrollar para dicho proceso son las siguientes:

- (i) En primer lugar, se calculó un indicador poblacional y de mortalidad con el objetivo de evitar que las estadísticas de correlación espacial estén sesgadas a los municipios donde se presentan una mayor población. Por ello, al transformar la tasa total de mortalidad en cada región mediante una tasa se pudo evitar que los análisis se sesgaran.
- (ii) Posteriormente, se aplicó una técnica de reducción de dimensionalidad a través del análisis componentes principales (PCA, por sus siglas en inglés) para crear un índice sociodemográfico que incluyó a las variables objeto de estudio. Tal y como se establece en los antecedentes, los métodos de reducción de dimensionalidad son técnicas ampliamente utilizadas para simplificar los datos al eliminar las redundancias y conservando la mayor parte de la variabilidad original, así mismo, permite eliminar problemas de colinealidad mejorando su análisis, interpretación y generan características compactas que pueden ser útiles para crear nuevas variables que representen mejor los datos originales. Si bien es cierto, la

utilización de métodos de reducción de dimensionalidad son métodos que permiten mejorar la consistencia de los datos, se debe analizar cada caso en concreto porque dichas técnicas pueden ser sensibles a los datos ruidosos, no todos capturan las relaciones no lineales y pueden depender de la varianza. Al considerar las ventajas y desventajas de los métodos de reducción de dimensionalidad y analizar los resultados de diferentes técnicas de reducción de dimensionalidad (ACP y T-SNE) se decide optar por el método T-SNE porque el resultado con el PCA porque este método utiliza transformaciones lineales, sin embargo, la naturaleza de los datos evidenció que el dataframe no era lineal y no capturaba las estructuras locales o agrupaciones específicas. En contraste, el método T-SNE no es lineal, permite visualizar estructuras intrínsecas de los datos aunque es un método más exigente en términos computacionales.

- (iii) Al finalizar el procedimiento de reducción de la dimensionalidad de los datos, se procedió a realizar el análisis por conglomerados. Para dicha tarea, se ha propuesto utilizar el método de K-means que permitirá realizar una serie de agrupaciones mediante unos algoritmos que se repetirán para todos los valores de k . No obstante, el estudio también analizará otras técnicas de clusterización como el K-means ++, el Gaussian Mixture o el Optics. La decisión final sobre el método a escoger dependerá de los resultados y el proceso de validación tanto cualitativo como cuantitativa de los clústeres.
- (iv) Para la realización de los clústeres, se analizaron como unidades de medida a los diferentes municipios que integran la zona pacífica colombiana. Para ello, el presente trabajo adopta las recomendaciones de Aungkulanon y otros [42] “quienes comenzaron el proceso de clusterización con la identificación de “súper distritos”, en el presente caso, ese procedimiento se suplió con el cálculo de las tasas de mortalidad y poblacionales para evitar las varianzas en las estadísticas entre municipios con una alta población y municipios con pocos habitantes.
- (v) Luego del análisis inicial, se procedió a realizar los clústeres de mortalidad y condiciones socioeconómicas mediante la agrupación de municipios en la región pacífica colombiana. Para la identificación de los n clústeres en cada región agrupada se tendrá en cuenta el valor resultante de R^2 , posterior a ello, se procedió a realizar su respectivo análisis.
- (vi) Luego de la realización de la clusterización mediante la técnica de K-means, se procedió a realizar la validación de dichos resultados con la utilización de otras técnicas para identificar si existieron cambios abruptos en las diferentes técnicas y se utilizó métricas de evaluación de consistencia de clusterización. Luego de este procedimiento, se encontró una consistencia en los diferentes métodos aplicados y se decide adoptar los resultados obtenidos por el método del K-means. Finalizado este procedimiento, se realizó su interpretación y posteriores recomendaciones y sugerencias de acuerdo a los hallazgos evidenciados.

Al momento de obtener los modelos previos, el procedimiento analítico de los clústeres requiere validar y evaluar si los modelos obtenidos son consistentes. Para ello, se requiere realizar un análisis que permita identificar si el número de clústeres es el adecuado, si las tendencias en las agrupaciones son las correctas y verificar la generalidad del modelo. Para ello, se puede implementar diferentes técnicas tanto cualitativas como cuantitativas para analizar si los resultados del modelo ofrecen resultados fiables y que presenten características

de consistencia, validez interna y externa. El proceso de validación es un paso central antes de publicar y diseñar el tablero de información que podrá ser accesible al público en general. Todo el procedimiento de validación de los resultados de los conglomerados permite dar cumplimiento al tercer objetivo específico del trabajo. El proceso de validación propuesta en el estudio fue a través de expertos en políticas públicas de salud quienes tienen experiencia y amplio conocimiento en el tema. El total de expertos consultados para el proceso de validación fue de 6 individuos en los departamentos de Nariño y Valle del Cauca.

Cuando se finalizó el proceso de validación, los investigadores procedieron a desplegar el tablero con la información resultante del estudio mediante una herramienta digital (el link del tablero se presenta en el capítulo 8). La publicidad de la información con esta herramienta se puede publicar de forma gratuita, sin necesidad de adquirir la licencia de uso, a través de una publicación en la web y de open source para la comunidad en general, y de forma concreta para las entidades municipales y regionales encargadas de adoptar decisiones de política pública.

Al validar los principales modelos, se pudo identificar las principales características sociodemográficas asociadas a la prevalencia de mortalidad por ECV y a partir de ello, se obtuvo conclusiones para realizar una serie de recomendaciones sobre política pública que permita ajustar las estrategias, programas y planes que estén implementando las entidades estatales para la promoción y prevención de la salud en la región pacífica colombiana.

5. CARACTERÍSTICAS SOCIOECONÓMICAS Y MORTALIDAD DE ECV EN LA REGIÓN PACÍFICA COLOMBIANA EN EL PERIODO 2016-2020

En el presente capítulo, se sintetiza los principales resultados relacionados con el proceso de recopilación, sistematización y análisis inicial de la base de datos construida para la investigación. Para la construcción de la base de datos se utilizaron datos secundarios que provienen de las siguientes fuentes:

- **DANE:** Los datos recopilados del Departamento Administrativo Nacional de Estadísticas (DANE) incluyeron: los datos poblacionales, el NBI, Valor Agregado por Municipio y las estadísticas de mortalidad por ECV. En relación con los datos demográficos, la clasificación de los grupos etarios se realizó de acuerdo con los rangos de edad establecidos en las políticas públicas de salud de la siguiente manera:
 - a) **Primera infancia:** Se refiere al grupo de edad constituido entre los 0 a 5 años de edad.
 - b) **Infancia:** Hace referencia a la población contenida entre los 6 y 12 años.
 - c) **Adolescencia y juventud:** Integra a los grupos poblacionales entre los 13 y 29 años.
 - d) **Adultez:** Se relaciona con los grupos poblacionales contenidos entre los 30 a 59 años de edad.
 - e) **Adulto mayor:** Es el grupo poblacional contenido entre los 60 años y más.
- **Datos Abiertos de Colombia:** En este repositorio se buscaron los datos relacionados con la cobertura educativa de los municipios de la zona pacífica. Se decidió usar la cobertura neta porque en esta variable no se cuentan a las estudiantes que se encuentran en extraedad. En relación con la cobertura neta se desagregó la variable en cobertura neta total, primaria, secundaria y media. En el mismo repositorio se buscó los datos relacionados con el porcentaje de cobertura del Sistema General de Seguridad Social en Salud – SGSSS.
- **Repositorio Terridata -DNP:** En este repositorio se buscaron y sistematizaron los datos relacionados con la cobertura de los servicios públicos (electricidad, acueducto y alcantarillado). De igual manera, se obtuvieron los datos del gasto público en salud, deporte y recreación. Así mismo, en este repositorio se obtuvieron los datos del porcentaje de población formal.
- **Reportes ASIS del Instituto Nacional de Salud:** En este repositorio se buscaron y obtuvieron los datos del Índice de Riesgo de Calidad del Agua para consumo humano (IRCA) que evalúa los rangos óptimos que debería tener el recurso hídrico para evitar enfermedades relacionadas con su consumo.
- **Observatorio de Educación de la Pontificia Universidad Javeriana:** De este repositorio se lograron obtener los datos del número de graduados de programas universitarios (pregrado).

Luego de realizar la descripción de las principales fuentes de información obtenidas en el estudio, se explican la utilidad de cada una de las variables utilizadas y su aporte para el desarrollo del estudio.

Tabla 2

Categorías y subcategorías de variables

Construcción de la base de datos			
Categorías de variables	Subcategorías	Estudios previos	Aportes trabajo de grado
Variables demográficas	Nacimientos	N/A	Estas variables no han sido analizadas en estudios previos, sin embargo, estas variables sociodemográficas podrían aportar un elemento de análisis al estudio considerando el comportamiento de las tasas totales de natalidad y defunción.
	Defunciones		
	Población por Sexo	Pico y Otros [50] Vergara y Otros [51]	La literatura ha demostrado que el sexo es una variable importante al analizar el riesgo de que los hombres o mujeres tengan para el desarrollo de una enfermedad. Empíricamente, el sexo masculino ha sido asociado a un mayor riesgo de enfermedades cardiovasculares
	Población total	Gómez y Otros [52]	La población total podría ayudar a entender la prevalencia de las ECV en cada una de las regiones analizadas, al tiempo, que permitiría identificar si unos grupos poblacionales son más propensos a presentar diferentes patologías por ECV
	Población por área	Aungkulanon y Otros [42] Augusto y Lanza [49] De Andrade y Otros [53] Hernández y Otros [54]	La distribución geográfica permite identificar las condiciones de vida de los individuos que viven en zonas urbanas o rurales y su riesgo para el patologías por ECV (por ejemplo mayor servicios hospitalarios en la zona urbana o estilos de vida más saludables en la zona rural)

	Población por grupo etario	Parra y Otros [11] Briceño y otros [55] Pico y Otros [50] Vergara y Otros [51] Aungkulanon y Otros [42] Augusto y Lanza [49] De Andrade y Otros [53]	El curso de vida es otro de los factores más determinantes para explicar el desarrollo de patologías asociadas a las ECV. Los estudios empíricos demuestran que a mayor edad, existe una mayor posibilidad de desarrollar enfermedades cardiovasculares.
Variables económicas	Valor agregado	N/A	El valor agregado se toma como una variable proxy del PIB de cada municipio, al considerar que en la mayoría de los municipios de la región pacífica colombiana no hay registro del PIB.
	Gasto público en salud	N/A	Esta variable se analiza en el estudio porque se considera un elemento valioso para el desarrollo del presente estudio porque estudios teóricos han supuesto que un mayor gasto público en políticas de promoción, prevención y sensibilización podría reducir los riesgos de desarrollar una ECV al ofrecer mayores programas y estrategias de medicina preventiva
	NBI	Parra y Otros [11]	El NBI es un indicador de las necesidades básicas insatisfechas de la población colombiana que incluye una multiplicidad de variables analizadas y que no se limitan a la generación de ingresos. Esta variable es importante al considerar que las zonas con mayores NBI tendrían, hipotéticamente, más probabilidades de presentar inadecuadas condiciones de salud.

	PBI Per cápita	Parra y Otros [11] Vergara y Otros [51] Augusto y Lanza [49] Ford y Highfield [56]	Esta es una variable que mide el nivel de ingresos en una región por persona. Para el presente estudio no se tendrá en cuenta debido a la ausencia de datos disponibles.
	Estrato	Pico y Otros [50]	El estrato es otra medida socioeconómica de las personas. En este estudio no se tendrá en cuenta al considerar que ya se tiene un indicador como el valor agregado y el NBI que son variables económicas con mayor capacidad explicativa que el estrato.
	% población en trabajo formal	Vergara y Otros [51] Aungkulanon y Otros [42] Ford y Highfield [56]	El porcentaje de población con trabajo formal es otro de los indicadores que podría tener un alto poder explicativo en el desarrollo de ECV, debido a que las personas con mejores oportunidades laborales presentan un mejor nivel educativo, mayores oportunidades financieras para cubrir gastos de medicina prepagada y aseguramiento al SGSSS
VARIABLES SOCIALES	Cobertura educativa neta primaria	Vergara y Otros [51] Aungkulanon y Otros [42] Augusto y Lanza [49] De Andrade y Otros [53]	Los niveles de formación educativa se consideran como una de las variables sociales con mayor relevancia para la explicación de una multiplicidad de fenómenos económicos. En el área de la salud, se ha considerado que las personas con mayores niveles educativos pueden presentar menores riesgos para el desarrollo de ECV debido a: (a) mayor conocimiento en salud; (b) mayor sensibilización sobre la importancia de hábitos
	Cobertura educativa neta secundaria	Vergara y Otros [51] Aungkulanon y Otros [42] Augusto y Lanza [49] De Andrade y Otros [53]	

Cobertura educativa neta media	Vergara y Otros [51] Aungkulanon y Otros [42] Augusto y Lanza [49] De Andrade y Otros [53]	saludables; (c) más oportunidades para ingresar a esquemas de protección a salud. Si bien es cierto, en muchos estudios no siempre se ha analizado esta variable, el presente estudio si la considera al reconocer las brechas del capital humano al interior de la región pacífica. A raíz de ello, es importante analizar si esta variable puede presentar un alto nivel explicativo para la mortalidad de enfermedades cardiovasculares.
Cobertura educativa neta total	Vergara y Otros [51] Aungkulanon y Otros [42] Augusto y Lanza [49] De Andrade y Otros [53]	
# graduados universitarios	Vergara y Otros [51] Aungkulanon y Otros [42] Augusto y Lanza [49] De Andrade y Otros [53]	
IRCA	N/A	El IRCA no se ha considerado en ninguno de los estudios analizados, sin embargo, la calidad del agua es una variable muy importante porque el consumo del recurso hídrico es un factor fundamental para la preservación de un buen estado de salud. Si bien es cierto, el nivel de calidad del agua ha sido asociado a enfermedades gastrointestinales, en los últimos años, también se ha ido asociando a las ECV al considerar los efectos nocivos del agua contaminada en la preparación de alimentos para la salud humana.

Cobertura servicios públicos	N/A	Esta variable no ha sido analizada en los estudios encontrados previamente. Sin embargo, también se considera una variable de interés porque la cobertura a los mismos es considerada como determinantes intermedios de la salud según la OMS.
Cobertura SGSSS	Pico y Otros [50] Briceño y Otros [55]	La cobertura al SGSSS y el tipo de cobertura es otra variable de interés que se considera en el estudio. Este atributo no se ha considerado en muchos estudios, pero el porcentaje y tipo de cobertura puede tener una influencia en los programas de prevención y promoción en la salud.
Hábitos alimentarios saludables	Pico y Otros [50] Briceño y Otros [55] Vergara y Otros [51] Augusto y Lanza [49]	La alimentación es una variable de importancia para la explicación de diversas enfermedades. Sin embargo, el presente estudio no abordó esta variable debido a la inexistencia de datos agregados disponibles por municipios.
Actividad Física	Pico y Otros [50] Vergara y Otros [51] Augusto y Lanza [49]	La actividad física se constituye como una variable explicativa para el desarrollo de ECV. En el presente estudio, se analizó dicha variable a través de las inversiones de los municipios en deporte y creación, al considerar que no existen datos a nivel municipal
VARIABLES DE VIVIENDA (conexión a internet, calidad de vivienda, déficit de vivienda)	Aungkulanon y Otros [42]	el presente estudio analizó estas variables a través de las inversiones en vivienda por cada municipio. Adicional a ello, las condiciones de vivienda también se desagregaron de acuerdo a los niveles de cobertura de servicios públicos básicos en cada municipio.

VARIABLES DE MORTALIDAD Y MORTALIDAD POR ECV

Antecedentes familiares	Vergara y Otros [51]	Constituyen un elemento importante para el análisis de enfermedades debido a que la genética desempeña un papel fundamental en la posibilidad de desarrollar una enfermedad o no. Sin embargo, esta variable no se analizó en el estudio debido a que esta información es confidencial a cada paciente.
Enfermedades hipertensivas	Parra y Otros [11] Gómez y Otros [52] De Andrade y Otros [53]	Como se logra observar, el tipo de mortalidad más analizada a nivel nacional e internacional son las enfermedades cardiovasculares porque se constituyen como uno de los grupos de enfermedades más frecuentes en pacientes alrededor del mundo. Sin embargo, en el estudio se propone un análisis que incluye a los 6 principales grupos de ECV de acuerdo a la clasificación 6/67 de la OPS que agrega signos, síntomas y estados morbosos de acuerdo a la patología de cada enfermedad.
Enfermedades isquémicas de corazón	Parra y Otros [11] Gómez y Otros [52] De Andrade y Otros [53]	
Enfermedades cardiovasculares	Gómez y Otros [52] Augusto y Lanza [49] Pico y Otros [50] De Andrade y otros [53] Ford y Highfield [56]	
Insuficiencia cardíaca	Gómez y Otros [52] De Andrade y Otros [53]	
Enfermedades cerebrovasculares	Parra y Otros [11] Hernández y Otros [54] Gómez y Otros [52]	
Otras del sistema circulatorio	N/A	
Mortalidad	Aungkulanon y Otros [42]	

Fuente: Realización propia

5.1. Análisis descriptivo de los datos

El análisis de variables relacionadas con el tema de análisis de conglomerados en desigualdades y enfermedades cardiovasculares (ECV) desempeña un papel fundamental en la formulación y ejecución de políticas públicas de salud. Dado que las enfermedades cardiovasculares representan una carga significativa para los sistemas de salud en todo el mundo y están estrechamente vinculadas a factores socioeconómicos y de estilo de vida, por ello, entender las disparidades en su incidencia y prevalencia es esencial para una asignación efectiva de recursos. El análisis de conglomerados permite identificar patrones y agrupaciones de regiones o comunidades que comparten características similares en términos de desigualdades socioeconómicas y de salud, lo que proporciona información valiosa para la elaboración de estrategias de intervención específicas y adaptadas a las necesidades locales.

Al examinar variables relacionadas con la desigualdad y las enfermedades cardiovasculares a través del análisis de conglomerados, los responsables de la formulación de políticas pueden identificar áreas geográficas o poblaciones con mayor riesgo o vulnerabilidad, lo que les permite dirigir recursos y servicios de manera más efectiva y equitativa. Además, este enfoque puede ayudar a destacar las disparidades en el acceso a la atención médica y los determinantes sociales de la salud, lo que subraya la importancia de abordar las causas subyacentes de las desigualdades en salud. En última instancia, al comprender mejor la relación entre desigualdades y enfermedades cardiovasculares a través del análisis de conglomerados, las políticas públicas pueden diseñarse de manera más precisa y estratégica para promover la equidad en salud y mejorar los resultados cardiovasculares a nivel poblacional. En esta medida, es importante conocer las variables desde una perspectiva descriptiva.

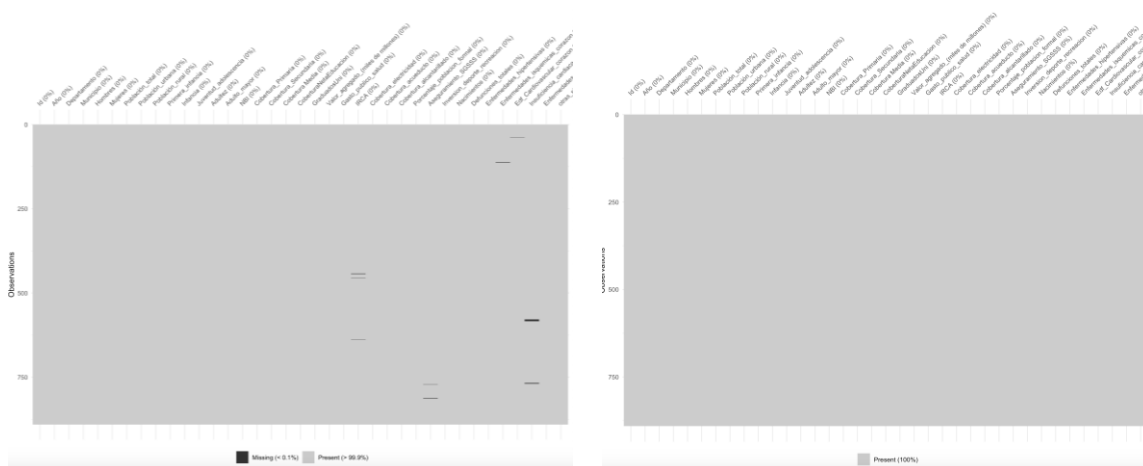
5.1.1. Medidas de tendencia central

Para iniciar el proceso de exploración inicial de los datos, se comenzó identificando el porcentaje de datos ausentes de la base de datos. A partir de ello, se logró identificar que la base de datos presentaba un 99.9% de datos presentes y solo el 0.01% estaban ausentes (ver Figura 1). De forma concreta, los datos ausentes se encontraron en las variables de IRCA, aseguramiento al SGSSS, mortalidad por enfermedades isquémicas, enfermedades cardiovasculares e insuficiencia cardiaca.

Sin embargo, al considerar que para la realización de los conglomerados se requiere que la base de datos tenga el 100% de los atributos, se procedió a realizar el proceso de imputación de los datos en las variables mencionadas previamente. Para el proceso de imputación es importante reconocer que existen diversos métodos como los probabilísticos (media o promedio) o a través de métodos de regresión (lineal, estocástica, LOC o múltiple). Sin embargo, al considerar la naturaleza de los datos y el porcentaje de balance del dataframe, se procedió a realizar un proceso de imputación de datos mediante la media de las variables.

Figura 1

Balace de datos antes y después de imputación



Fuente: Realización propia

Luego de finalizar el proceso de imputación de los datos, se procedió a identificar algunas estadísticas descriptivas de las variables, previo a este proceso se habían transformado las variables a través de un logaritmo. Los resultados de este proceso se resumen a continuación:

Tabla 3

Estadísticas descriptivas de la base de datos

Variable	Mínimo/ Máximo	Media	Promedio	Cuartiles (1 y 3)	Desviación
Hombres	7.48 13.86	9.27	9.08	8.50 9.89	1.02
Mujeres	7.26 13.99	9.15	8.99	8.46 9.64	0.96
Población total	7.83 14.62	9.71	9.55	9.10 10.07	0.96
Población Urbana	6.30 14.60	8.53	8.12	7.66 9.21	1.34
Población Rural	7.42 12.05	9.30	9.23	8.71 9.78	0.82
Primera infancia	5.98 12.18	7.59	7.46	6.84 8.14	0.98
Infancia	5.99 12.22	7.64	8.33	7.82 9.05	0.95
Juventud y adolescencia	6.85 13.20	8.52	8.33	7.82 9.05	0.95
Adulthood	6.77 13.78	8.84	8.68	8.15 9.38	0.99

Adulto mayor	5.28 12.77	7.79	7.67	7.17 8.26	0.97
NBI	0.08 86.52	26.15	18.02	11.07 32.51	2.15
Cobertura primaria	2.98 9.63	6.36	4.69	4.32 8.91	2.30
Cobertura secundaria	1.95 9.61	6.19	4.61	4.18 8.76	2.30
Cobertura media	0.35 9.04	5.51	4.13	3.56 8.11	2.36
Cobertura neta	10.58 172.76	79.37	79.17	70.32 89.80	1.88
Graduados universitarios	0 10.22	1.23	0.69	0 1.38	2.05
Valor agregado	2.62 10.68	5.09	4.86	4.16 5.65	1.29
Gasto público salud	0 26.25	21.18	21.98	19.45 22.73	2.48
IRCA	0 100	27.64	19.52	3.97 48.45	2.63
Cobertura electricidad	8.82 100	89.00	96.99	84.52 100	1.69
Cobertura acueducto	0 100	59.54	61.31	38.07 85.49	3.01
Cobertura alcantarillado	0 100	42.55	40.16	17.60 64.81	2.89
Población formal	0.46 3.76	1.66	1.59	1.31 1.89	0.57
SGSSS	0 185.31	81.79	83.48	73.57 92.67	1.71
Inversión deporte y recreación	0 26.02	19.37	19.23	18.71 20.09	1.80
Nacimientos	1.17 10.18	5.06	4.87	4.26 5.76	1.15
Defunciones totales	0.69 9.79	4.18	4.04	3.47 4.72	1.18
E. hipertensivas	0 5.77	1.33	1.38	0.69 1.94	1.04
E. isquémicas	0 7.94	2.08	2.07	1.09 2.99	1.36
E. cardiovascular	0 6.29	0.60	0.50	0 1.09	0.82
Insuficiencia cardíaca	0 6.29	0.60	0.50	0 1.09	0.82
E. cerebrovascular	0 6.98	1.59	1.60	0.69 2.19	1.13

Otras sistema circulatorio	0 5.11	0.41	0.38	0 0.69	0.67
---------------------------------------	-----------	------	------	-----------	------

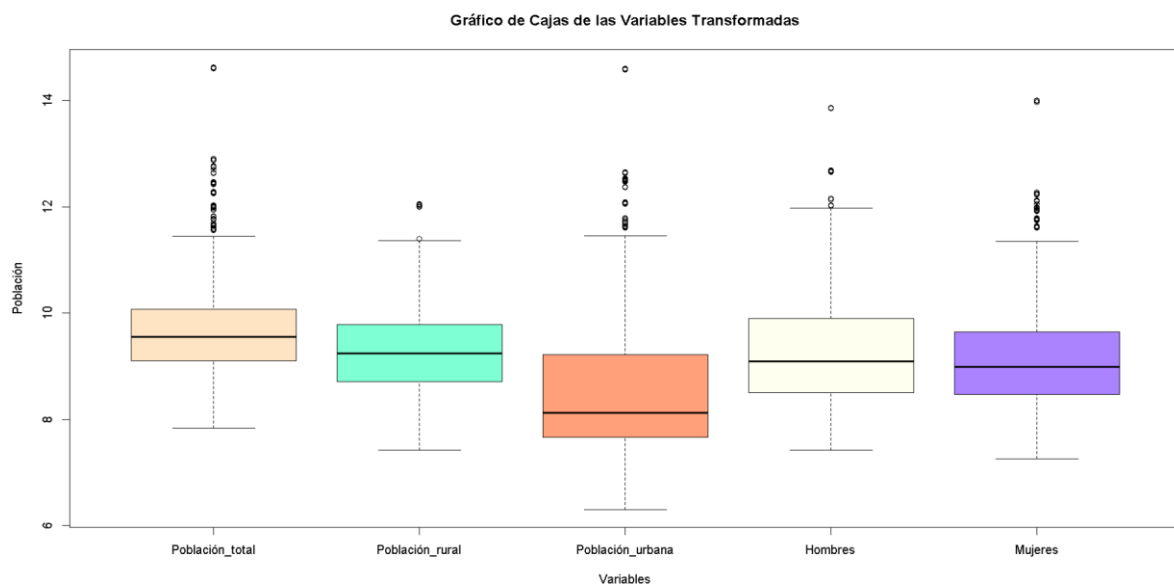
Fuente: Realización propia

5.2. Análisis estadístico de variables

Luego de realizar el análisis y exploración inicial de los datos, se procedió a realizar un análisis estadístico de los datos contenidos en el dataframe. Este proceso incluyó el análisis de las cajas de bigotes y las correlaciones de las variables.

Figura 2

Gráfico de cajas variable población



Fuente: Realización propia

El gráfico de cajas de las variables transformadas muestra la distribución de cinco categorías poblacionales: población total, población rural, población urbana, hombres y mujeres. Cada caja representa la dispersión de los datos en estas categorías, permitiendo identificar la mediana, los cuartiles y la extensión de los valores atípicos.

Para la población total, la mediana se sitúa alrededor de 10, con una variabilidad que se extiende desde aproximadamente 8 hasta 11.5. Los bigotes alcanzan hasta cerca de 12, pero hay varios valores atípicos por encima de este límite, lo que sugiere que existen regiones o periodos con una población total significativamente mayor. La población rural tiene una mediana similar de 10, pero una dispersión más concentrada entre aproximadamente 9 y 11, con valores atípicos más dispersos que alcanzan hasta 12.

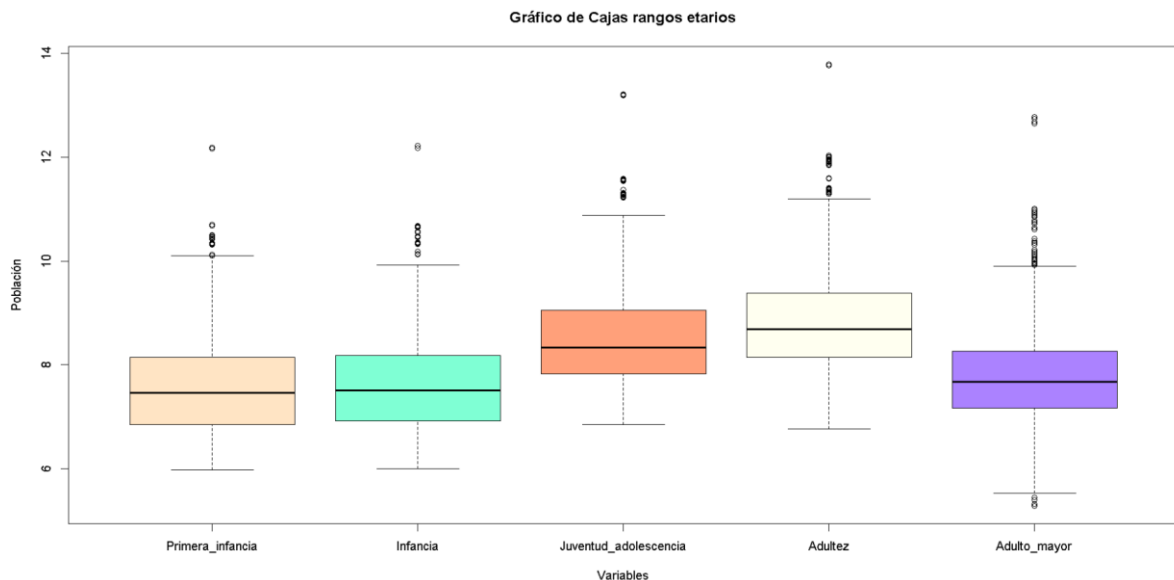
La población urbana muestra una mediana cercana a 9, con una mayor variabilidad que se extiende desde aproximadamente 8 hasta 10.5. Los bigotes y los valores atípicos alcanzan hasta cerca de 12, indicando una mayor heterogeneidad en la población urbana. En cuanto a

la distribución por género, tanto hombres como mujeres presentan medianas similares alrededor de 9. La dispersión para hombres se extiende desde aproximadamente 8 hasta 11, con varios valores atípicos hasta cerca de 13. Para las mujeres, la dispersión se extiende desde aproximadamente 8 hasta 11, con algunos valores atípicos por encima de este límite.

Este gráfico de cajas permite visualizar las diferencias en la distribución de la población en distintas categorías, mostrando tanto la centralidad de los datos como la variabilidad y la presencia de valores atípicos. Esto puede ser útil para identificar patrones y tendencias en la distribución poblacional, así como para detectar áreas con características demográficas inusuales.

Figura 3

Gráfico de cajas población por grupo etario



Fuente: Realización propia

El gráfico de cajas de los rangos etarios presenta una comparación entre cinco grupos de edad: primera infancia, infancia, juventud/adolescencia, adultez y adulto mayor. Cada caja representa la distribución de la población en estos rangos etarios, proporcionando una visión clara de la mediana, los cuartiles y la dispersión de los datos.

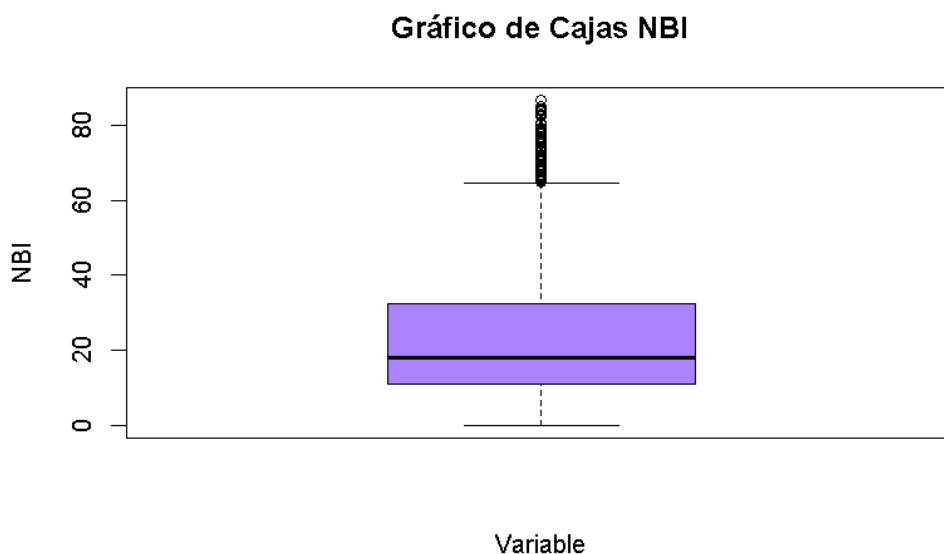
Para el grupo de primera infancia, la mediana se sitúa alrededor de 8, con una variabilidad que se extiende desde aproximadamente 7 hasta 9.5. Los bigotes muestran la extensión de la población de este grupo, con algunos valores atípicos por encima de 10, lo que indica una menor cantidad de áreas o periodos con una población más alta en este rango de edad. En comparación, el grupo de infancia muestra una mediana similar de 8, pero con una dispersión más uniforme que se extiende desde aproximadamente 7 hasta 9.5, y algunos valores atípicos que alcanzan hasta 12.

En el grupo de juventud y adolescencia, la mediana también está alrededor de 8, pero con una dispersión ligeramente mayor en comparación con la infancia, extendiéndose desde aproximadamente 7 hasta 10. Los valores atípicos alcanzan hasta 12, sugiriendo que hay áreas con una mayor población juvenil/adolescente en comparación con la primera infancia y la infancia.

Para los grupos de adultez y adulto mayor, la mediana está alrededor de 9 y 8, respectivamente. La caja para el grupo de adultez muestra una mayor variabilidad, con una dispersión que se extiende desde aproximadamente 7.5 hasta 10.5, y varios valores atípicos que alcanzan hasta 12. El grupo de adulto mayor tiene una dispersión similar a la de la primera infancia, con una mediana de 8 y algunos valores atípicos por encima de 10, lo que sugiere una distribución más homogénea de la población en este rango de edad. La presencia de valores atípicos en todos los grupos indica que existen áreas con una población significativamente diferente al resto, destacando la diversidad demográfica en los diferentes rangos etarios.

Figura 4

Gráfico de cajas variable Necesidades Básicas Insatisfechas (NBI)



Fuente: Realización propia

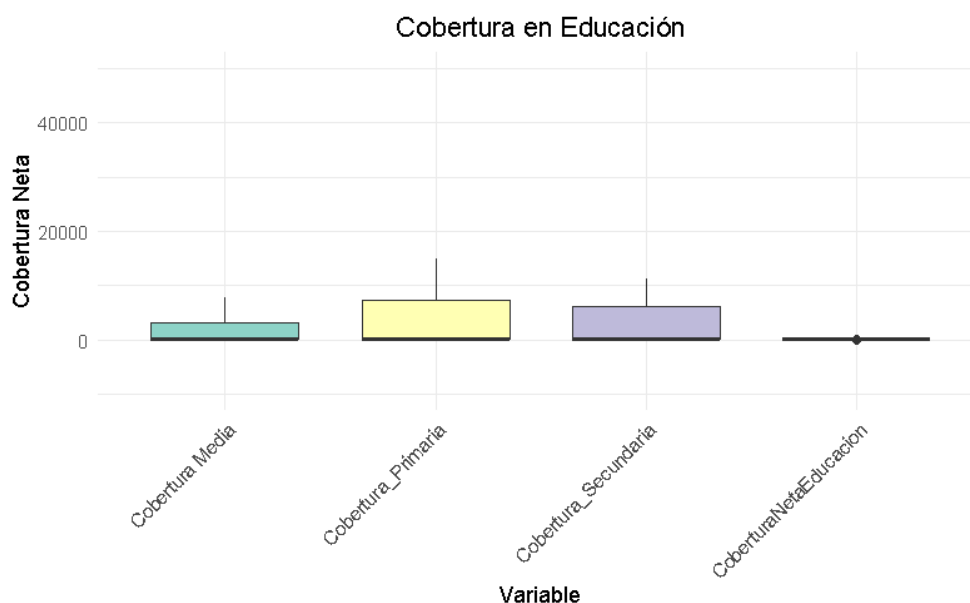
El gráfico de cajas sobre el índice de necesidades básicas insatisfechas (NBI) muestra que la mediana de NBI está alrededor del 20%, lo que indica que la mitad de las observaciones tienen menos de este porcentaje de NBI. El rango intercuartílico (IQR) sugiere una dispersión moderada, con la mayoría de los datos situándose entre aproximadamente el 10% y el 30%. Los bigotes se extienden desde alrededor del 0% hasta cerca del 50%, lo que muestra la variabilidad en el índice de NBI en diferentes áreas o periodos. La presencia de múltiples valores atípicos por encima del 50% indica que existen áreas con niveles significativamente

más altos de NBI, alcanzando hasta el 80%, lo cual resalta las disparidades en términos de necesidades básicas satisfechas.

En lo relacionado con la cobertura educativa, se evidencia que la única variable que muestra datos atípicos es la cobertura neta en educación media, así como también se muestra que esta representa una cobertura menor en comparación a la cobertura en secundaria y primaria, donde se encuentra que existe un mejor indicador para la cobertura primaria.

Figura 5

Gráfico de cajas cobertura en educación



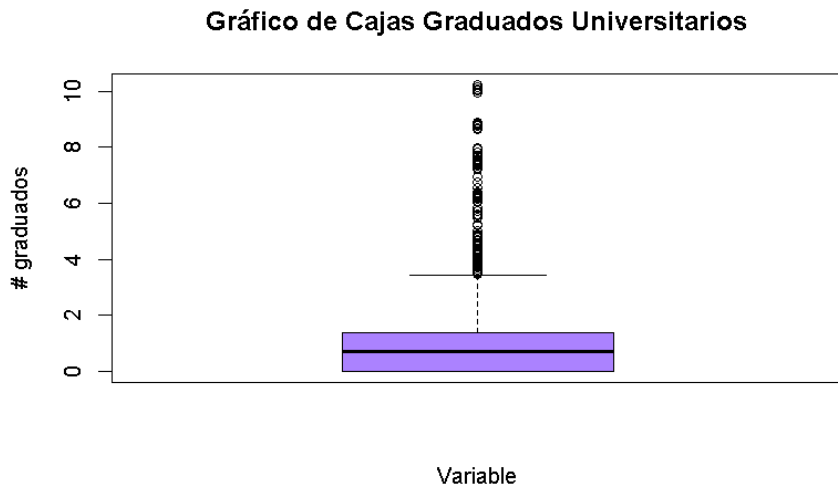
Fuente: Realización propia

Por otro lado, se puede mencionar que las cuatro variables señalan una mediana orientada hacia el cuartil 25, lo que es señal de sesgos y asimetrías en las mismas, sin embargo, un elemento de importancia es que la cobertura primaria presenta una cantidad de datos agrupados mayor en comparación a las demás coberturas, no obstante, dicha diferencia no es tan significativa.

El gráfico de cajas sobre graduados universitarios muestra la distribución del número de graduados en diferentes áreas o periodos. La mediana del número de graduados se encuentra cerca de 1, lo que indica que, en promedio, la mayoría de las áreas tienen alrededor de un graduado universitario. El rango intercuartílico (IQR) es bastante estrecho, reflejando una baja variabilidad entre el primer y tercer cuartil. Sin embargo, la presencia de numerosos valores atípicos por encima de la mediana indica que hay ciertas áreas con un número significativamente mayor de graduados (ver Figura 6).

Figura 6

Gráfico de cajas graduados de estudios universitarios

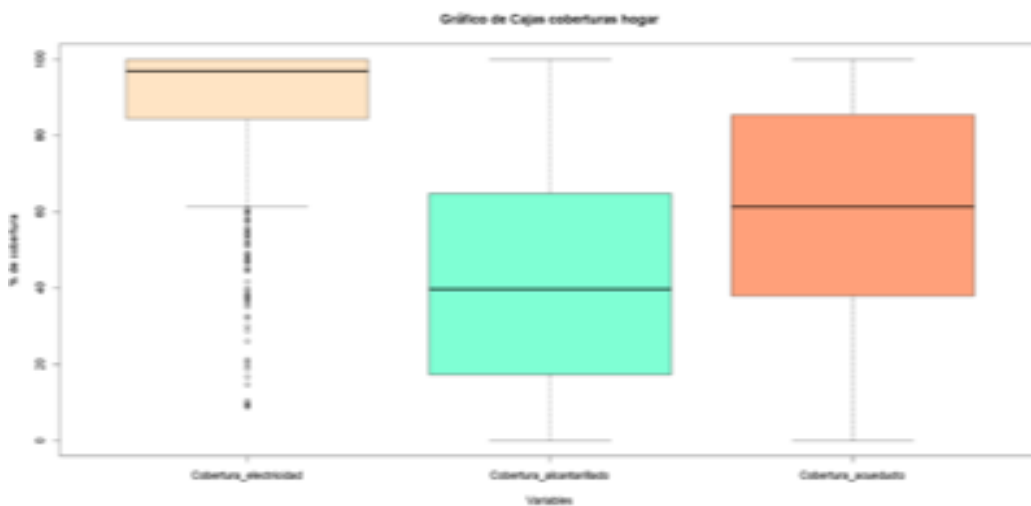


Fuente: Realización propia

Los bigotes del gráfico se extienden desde 0 hasta cerca de 3 graduados, lo que sugiere que la mayoría de los datos se encuentran dentro de este rango. Sin embargo, la concentración de valores atípicos en la parte superior del gráfico, algunos alcanzando hasta 10 graduados, destaca que, aunque la mayoría de las áreas tienen un número bajo de graduados, hay excepciones notables con una mayor cantidad. Esto podría indicar disparidades significativas en la educación superior entre diferentes regiones o periodos, sugiriendo que algunos lugares tienen un acceso y una tasa de finalización de estudios universitarios mucho más alta que otros.

Figura 7

Gráfico de cajas coberturas de Servicios Públicos



Fuente: Realización propia

El gráfico de cajas sobre coberturas del hogar presenta la distribución porcentual de la cobertura de electricidad, acueducto y alcantarillado en diferentes áreas. En cuanto a la cobertura de electricidad, la mediana se sitúa cerca del 100%, lo que indica que la mayoría de las áreas tienen una cobertura eléctrica completa o casi completa. Sin embargo, se observan numerosos valores atípicos hacia la parte inferior del gráfico, lo que sugiere que hay áreas con coberturas significativamente más bajas. Los bigotes se extienden desde cerca del 60% hasta el 100%, reflejando una amplia variabilidad en la cobertura eléctrica entre las diferentes áreas.

La cobertura del acueducto tiene una mediana alrededor del 75%, lo que indica una menor proporción de áreas con cobertura completa de acueducto en comparación con la electricidad. El rango intercuartílico (IQR) para la cobertura del acueducto es amplio, extendiéndose desde aproximadamente el 50% hasta el 90%, lo que señala una considerable variabilidad en la provisión de servicios de acueducto. Los bigotes del gráfico, que van desde el 0% hasta casi el 100%, junto con la presencia de algunos valores atípicos en ambos extremos, sugieren que mientras algunas áreas tienen una cobertura deficiente, otras disfrutan de una cobertura casi total.

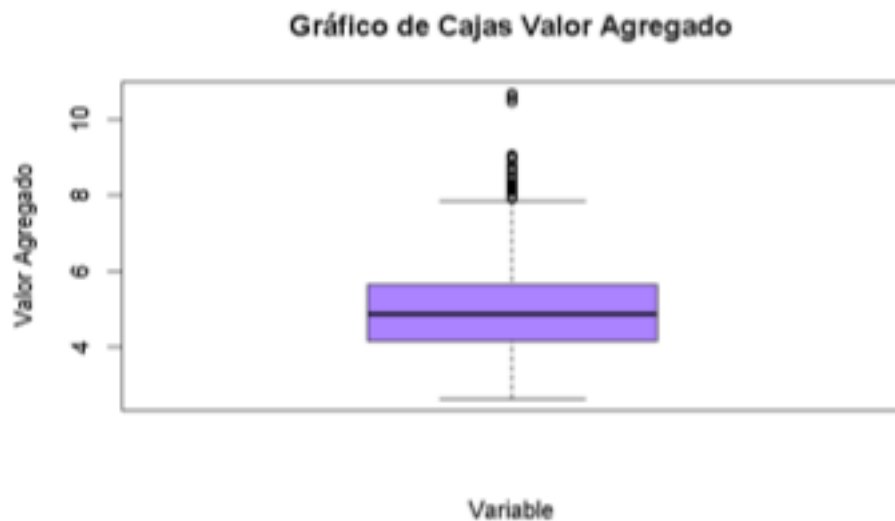
La cobertura de alcantarillado, representada por la caja verde, muestra una mediana alrededor del 60%, lo que indica que menos de la mitad de las áreas tienen una cobertura completa de alcantarillado. El IQR de esta variable es amplio, abarcando desde aproximadamente el 40% hasta el 80%, lo que revela una gran dispersión en la disponibilidad de servicios de alcantarillado. Los bigotes, que se extienden desde cerca del 10% hasta el 90%, y la presencia de valores atípicos, destacan aún más la variabilidad y los desafíos en la implementación de sistemas de alcantarillado en ciertas áreas.

En resumen, el gráfico revela una disparidad en la cobertura de servicios básicos entre electricidad, acueducto y alcantarillado en las áreas estudiadas. La cobertura eléctrica es en general alta, con algunas excepciones notables, mientras que la cobertura de acueducto y alcantarillado muestra una mayor variabilidad y áreas con servicios insuficientes. Esta información puede ser crucial para planificar mejoras en la infraestructura y garantizar una distribución más equitativa de estos servicios esenciales.

Sobre el gráfico de cajas del valor agregado muestra la distribución del valor agregado en diferentes áreas. La mediana del valor agregado se encuentra cerca de 6, lo que indica que la mitad de las áreas analizadas tienen un valor agregado inferior a este punto. El rango intercuartílico (IQR), que se extiende aproximadamente de 5 a 7, sugiere que la mayoría de las áreas tienen un valor agregado dentro de este rango, mostrando una variabilidad moderada en la cantidad de valor agregado producido (ver Figura 8).

Figura 8

Gráfico de Cajas Valor Agregado

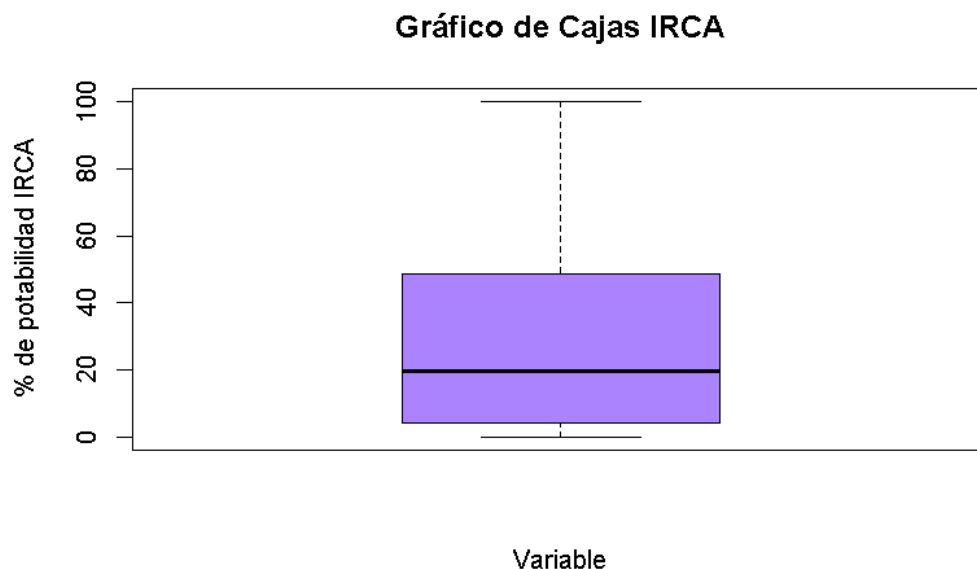


Fuente: Realización propia

Los bigotes del gráfico se extienden desde cerca de 4 hasta 8, lo que refleja una dispersión considerable en los datos de valor agregado. Además, se observan varios valores atípicos por encima de 8, indicando que algunas áreas tienen un valor agregado significativamente mayor que la mayoría. Estos valores atípicos pueden señalar áreas con una alta eficiencia o productividad en comparación con otras. En resumen, el análisis del gráfico de cajas del valor agregado revela una distribución moderadamente concentrada alrededor de la mediana con algunas áreas sobresalientes que podrían estar impulsando el crecimiento económico en la región.

Figura 9

Gráfico de cajas IRCA



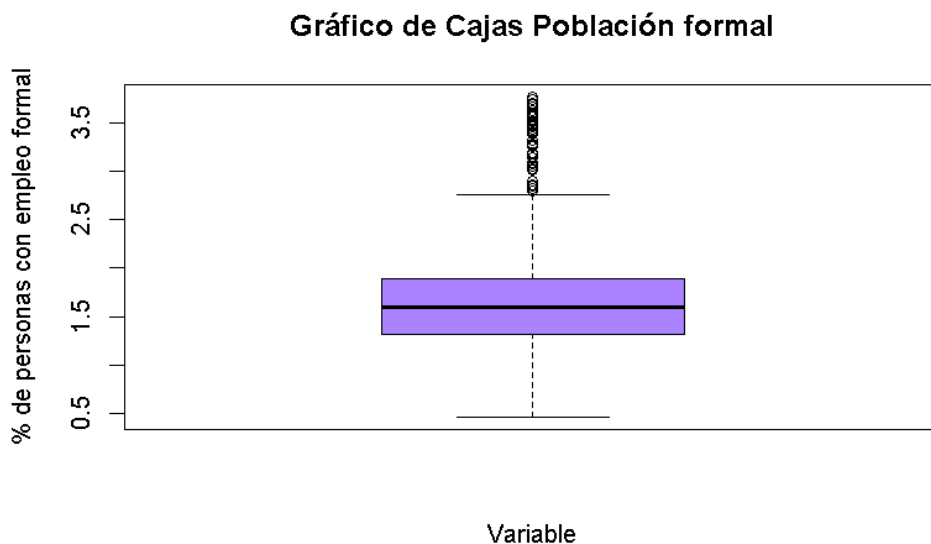
Fuente: Realización propia

El gráfico de cajas del Índice de Riesgo de Calidad del Agua (IRCA) muestra la distribución porcentual de la potabilidad del agua. La mediana del porcentaje de potabilidad se encuentra alrededor del 20%, indicando que la mitad de las áreas analizadas tiene una potabilidad del agua inferior a este valor. El rango intercuartílico (IQR) se extiende aproximadamente desde el 5% hasta el 50%, lo que sugiere que el 50% central de los datos se concentra en este intervalo. Esto indica una variabilidad moderada en los niveles de potabilidad del agua en las áreas estudiadas.

Los bigotes del gráfico se extienden desde cerca de 0% hasta 100%, mostrando una considerable dispersión en los datos de potabilidad del agua. Esto refleja que existen áreas con niveles de potabilidad muy bajos, así como áreas con niveles muy altos, posiblemente indicando diferencias significativas en la calidad del agua entre las diversas regiones. La ausencia de valores atípicos sugiere que la mayoría de las áreas se encuentran dentro de este amplio rango de variabilidad. En conclusión, el análisis del IRCA revela una distribución amplia en la calidad del agua potable, con una mediana relativamente baja, sugiriendo que muchas áreas podrían estar enfrentando desafíos significativos en términos de calidad del agua.

Figura 10

Gráfico de cajas Población formal en el mercado laboral



Fuente: Realización propia

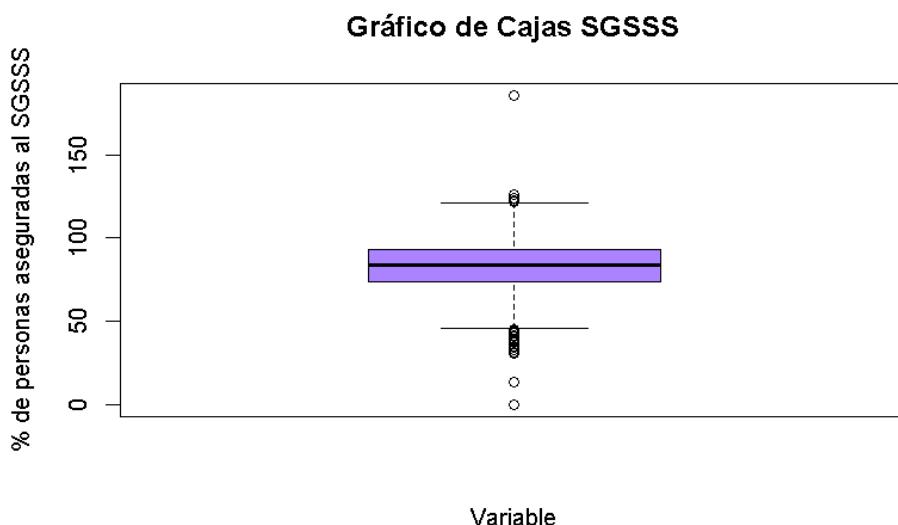
El gráfico de cajas de la población con empleo formal muestra que la mediana del porcentaje de personas con empleo formal se encuentra alrededor del 1.5%. Esto indica que, en la mayoría de las áreas analizadas, la mitad de los datos de empleo formal está por debajo de este valor, lo que sugiere una baja prevalencia de empleo formal en las áreas estudiadas. El rango intercuartílico (IQR), que se extiende aproximadamente desde el 1% hasta el 2%, revela una concentración de los datos en este rango, indicando una dispersión relativamente baja en los niveles de empleo formal en la mayoría de las áreas.

Los valores atípicos observados, principalmente por encima del 2.5%, sugieren la existencia de áreas con una proporción significativamente mayor de empleo formal en comparación con la mayoría. Los bigotes del gráfico, que se extienden desde 0.5% hasta 2.5%, muestran una variabilidad considerable en los datos de empleo formal. Esto refleja que, aunque la mayoría de las áreas tienen una baja tasa de empleo formal, existen excepciones notables que podrían requerir una atención especial para entender las causas subyacentes de estas variaciones. El análisis sugiere que el empleo formal es limitado en la región estudiada, con algunas áreas que presentan tasas significativamente más altas que el promedio.

El gráfico de cajas del porcentaje de personas aseguradas al Sistema General de Seguridad Social en Salud (SGSSS) muestra que la mediana está alrededor del 75%, indicando que la mitad de los datos se encuentra por debajo de este valor. Esto sugiere que, en la mayoría de las áreas analizadas, al menos el 75% de la población está asegurada. El rango intercuartílico (IQR), que se extiende aproximadamente desde el 60% hasta el 90%, revela una concentración significativa de los datos de aseguramiento en este intervalo, sugiriendo una distribución relativamente uniforme en el rango intermedio.

Figura 11

Gráfico de cajas Seguridad Social en Salud



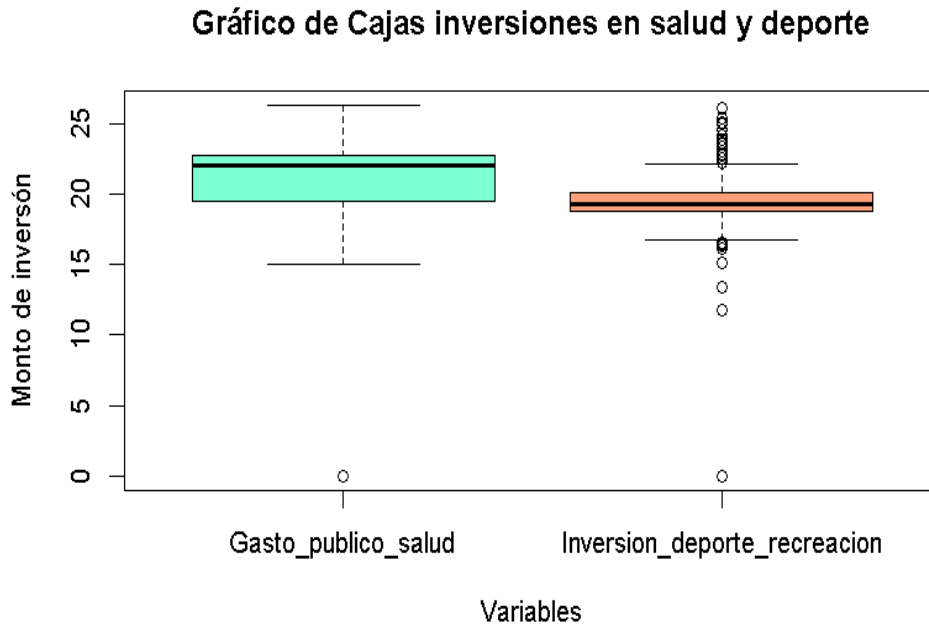
Fuente: Realización propia

Los valores atípicos observados, tanto por encima del 100% como por debajo del 50%, sugieren casos excepcionales de alta y baja cobertura de aseguramiento, respectivamente. Los bigotes del gráfico, que se extienden aproximadamente desde el 50% hasta el 100%, muestran una variabilidad considerable en los datos, reflejando la diversidad en los niveles de aseguramiento entre las diferentes áreas analizadas. En resumen, mientras que la mayoría de las áreas tienen un nivel de aseguramiento alrededor del 75%, existen diferencias notables que podrían requerir una atención especial para mejorar la equidad en la cobertura del sistema de salud.

La figura 12 hace alusión al gráfico de cajas de las variables gasto público en salud y la inversión en deporte y recreación, donde se proporciona una visión clara sobre cómo se distribuyen los recursos en estos sectores. La mediana del gasto público en salud es de aproximadamente 20 unidades monetarias, lo que indica que la mitad de los datos de inversión en salud se encuentra por debajo de este valor. El rango intercuartílico (IQR), que se extiende de 15 a 25 unidades, sugiere que la mayor parte de las inversiones se concentran en este rango. Los bigotes del gráfico, que van de 10 a 27 unidades, muestran una variabilidad moderada en la inversión total, con un valor atípico por debajo del límite inferior, sugiriendo un caso excepcional de baja inversión.

Figura 12

Gráfico de Cajas inversión en deporte y gasto público en salud



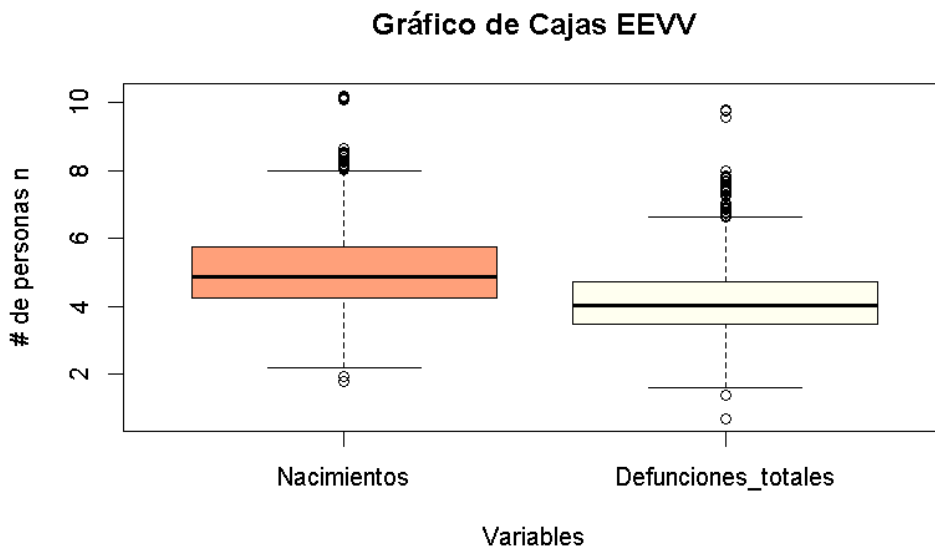
Fuente: Realización propia

Por otro lado, la segunda variable mencionada presenta una mediana ligeramente menor, cercana a 18 unidades monetarias. El IQR para esta categoría es más estrecho, de 15 a 20 unidades, indicando una menor dispersión central en comparación con el gasto en salud. Sin embargo, esta categoría muestra una mayor cantidad de valores atípicos, tanto por encima como por debajo de los límites de los bigotes, que se extienden de 10 a 22 unidades. Esto sugiere que, aunque la mayoría de las inversiones en deporte y recreación se concentran en un rango más estrecho, hay una variabilidad considerable con casos de inversión significativamente diferentes del promedio.

En contraste, el análisis estadístico revela que el gasto público en salud tiene una mayor mediana y una distribución más uniforme, indicando una asignación de recursos más consistente y predecible en el sector salud. En cuanto a la inversión en deporte y recreación existe una mayor dispersión y una cantidad significativa de valores atípicos, reflejando una asignación de recursos más variable y posiblemente menos controlada. Estos hallazgos subrayan la necesidad de una revisión detallada y posiblemente una reestructuración de la distribución de los recursos para asegurar una inversión más equilibrada y eficiente en ambos sectores, especialmente en deporte y recreación donde la variabilidad es más pronunciada.

Figura 13

Gráfica de cajas Nacimientos y Defunciones Totales



Fuente: Realización propia

El gráfico de cajas presentado muestra la distribución de dos variables: Nacimientos y Defunciones totales. El gráfico permite visualizar la dispersión y la simetría de los datos, así como para identificar posibles valores atípicos.

En esta medida, se encuentra que respecto a la variable Nacimientos existen diversos puntos de interés, entre lo que se considera la mediana, que hace alusión a línea horizontal dentro de la caja representa la mediana, que parece estar alrededor de 6. Esto indica que el número medio de nacimientos es aproximadamente 6 personas.

Por otro lado, se encuentra el Rango Intercuartílico (IQR), donde se perciba que la caja abarca el rango entre el primer cuartil (Q1) y el tercer cuartil (Q3), que parece estar entre 4 y 7. Esto sugiere que el 50% central de los datos de nacimientos se encuentra en este rango. En este sentido, también se encuentran los Valores Atípicos, en lo que se observan varios puntos fuera de los bigotes, tanto por encima como por debajo, lo que indica la presencia de valores atípicos. Estos valores atípicos representan nacimientos significativamente mayores o menores que la mayoría de los datos.

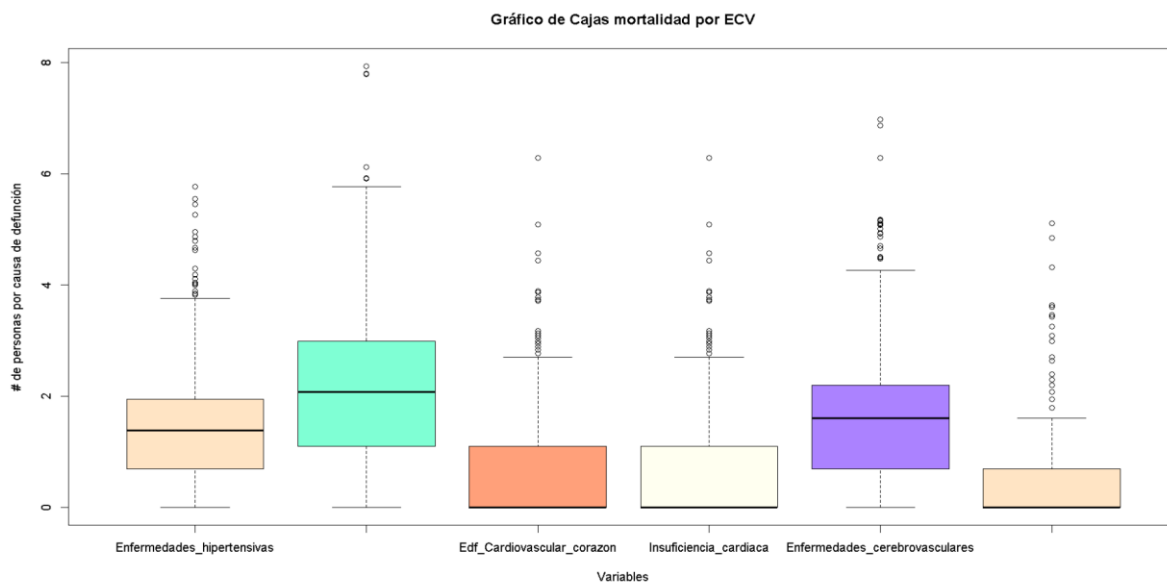
En cuanto a las Defunciones Totales, es posible evidenciar resultados respecto a la mediana, la cual se encuentra alrededor de 5, indicando que el número medio de defunciones totales es aproximadamente 5 personas. Igualmente, el Rango Intercuartílico (IQR) señala que la caja para las defunciones totales se extiende desde aproximadamente 4 hasta 6, indicando que el 50% central de los datos se encuentra en este rango. Cabe destacar que en esta variable también se identifican valores atípicos por encima y por debajo de los bigotes, lo que indica la presencia de defunciones significativamente mayores o menores. Finalmente, los bigotes

abarcan un rango mayor que en el caso de los nacimientos, desde aproximadamente 2 hasta 8, mostrando una mayor variabilidad en los datos de defunciones totales.

Desde un análisis comparativo, es preciso establecer que la mediana de los nacimientos es ligeramente superior a la de las defunciones totales, sugiriendo que, en promedio, hay más nacimientos que defunciones. En cuanto a la variabilidad (medida por el rango intercuartílico) es similar para ambas variables, aunque las defunciones totales parecen tener un rango ligeramente mayor, lo que indica una dispersión más amplia y en ambos conjuntos de datos presentan varios valores atípicos, lo que sugiere la presencia de eventos inusuales que podrían requerir una investigación adicional para comprender mejor sus causas.

Figura 14

Gráfico de cajas Enfermedades cardiovasculares



Fuente: Realización propia

El gráfico de cajas de mortalidad por enfermedades cardiovasculares (ECV) presenta un análisis comparativo entre varias causas de muerte: Enfermedades hipertensivas, Enfermedades cardiovasculares corazón, Insuficiencia cardiaca y Enfermedades cerebrovasculares y otras enfermedades. Este análisis es fundamental para entender la distribución y centralización de las defunciones en cada categoría y resaltar las diferencias y similitudes entre ellas.

En primer lugar, al observar la variable Enfermedades hipertensivas, se aprecia una mediana de aproximadamente 3 personas, indicando que la mitad de las defunciones están por debajo de este valor. El rango intercuartílico (IQR) se extiende desde 1 hasta 5, lo que sugiere una concentración central de los datos en este intervalo. Los valores atípicos por encima del límite superior indican la presencia de casos excepcionales de mortalidad elevada. Los bigotes muestran una variabilidad total excluyendo los atípicos, extendiéndose desde 0 hasta alrededor de 6.

En cuanto a las enfermedades isquémicas del corazón se muestra una mediana cercana a las 6 personas, lo cual es notablemente más alto en comparación con las otras causas de ECV analizadas. El rango intercuartílico (IQR) se extiende aproximadamente desde 4 hasta 8, indicando que el 50% central de las defunciones por estas enfermedades se encuentra en este intervalo. Esto sugiere una concentración significativa de casos de mortalidad en estos rangos. Además, se observan varios valores atípicos por encima y por debajo del rango, lo que indica la presencia de casos excepcionales de mortalidad significativamente mayores o menores que la media. Los bigotes se extienden desde aproximadamente 2 hasta 10, mostrando una considerable dispersión y variabilidad en los datos de mortalidad por enfermedades isquémicas del corazón. Esto refleja una alta incidencia y variabilidad en las defunciones causadas por estas enfermedades, destacando la importancia de enfocarse en su prevención y tratamiento.

Por otra parte, las enfermedades cardiovasculares del corazón señalan una mediana cercana a las 3 personas, lo que indica que la mitad de los datos de mortalidad se encuentran por debajo de este valor. El rango intercuartílico (IQR) se extiende aproximadamente desde 2 hasta 4, sugiriendo que el 50% central de las defunciones por estas enfermedades se concentra en este intervalo. Los valores atípicos son evidentes por encima y por debajo de los bigotes, lo que sugiere la presencia de casos excepcionales de mortalidad que se desvían significativamente de la media. Los bigotes, que se extienden desde aproximadamente 1 hasta 5, muestran una variabilidad moderada en los datos. Esta dispersión y presencia de valores atípicos indican que, aunque la mayoría de los casos se agrupan en torno a la mediana, existen eventos de mortalidad significativamente altos y bajos que podrían requerir una investigación más profunda para entender sus causas y determinar intervenciones específicas.

La variable Insuficiencia cardiaca presenta una mediana más baja, alrededor de 2 personas, indicando una menor mortalidad en comparación con otras causas de ECV. El IQR se extiende desde 1 hasta 4, sugiriendo una menor dispersión central. Aunque también se observan valores atípicos por encima del límite superior, los bigotes muestran una variabilidad total menor, extendiéndose desde 0 hasta aproximadamente 5, lo que indica una menor dispersión en los datos de mortalidad.

Por último, las Enfermedades cerebrovasculares muestra una mediana similar a la de "Enfermedades hipertensivas", alrededor de 3 personas. El IQR se extiende desde 1 hasta 5, indicando una dispersión comparable. Los valores atípicos presentes sugieren variaciones significativas en la mortalidad, mientras que los bigotes, que se extienden desde 0 hasta aproximadamente 6, reflejan una variabilidad total similar a la de las enfermedades hipertensivas.

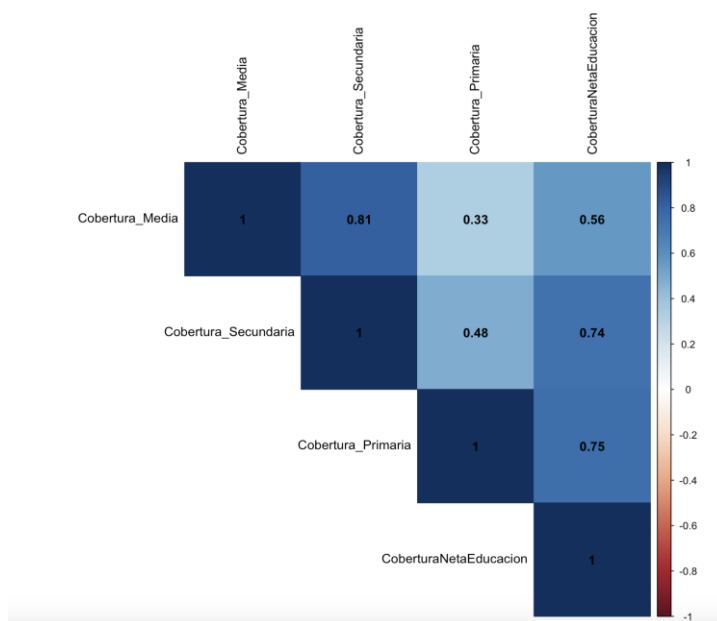
5.3. Diagramas de correlación

En un tercer momento, se realizó un análisis de las correlaciones de las variables para identificar el grado como se relacionan los principales grupos de variables entre sí. Este análisis permite identificar de forma clara cuál es el comportamiento de los atributos y su relación con el objetivo de identificar problemas de multicolinealidad que podría afectar la modelación y generación de los conglomerados. De igual manera, este análisis inicial permite

obtener un marco de referencia inicial sobre el comportamiento de los datos y establece un criterio de decisión al momento de realizar los clústeres y modelación final.

Figura 15

Diagrama de calor de Coberturas en Educación

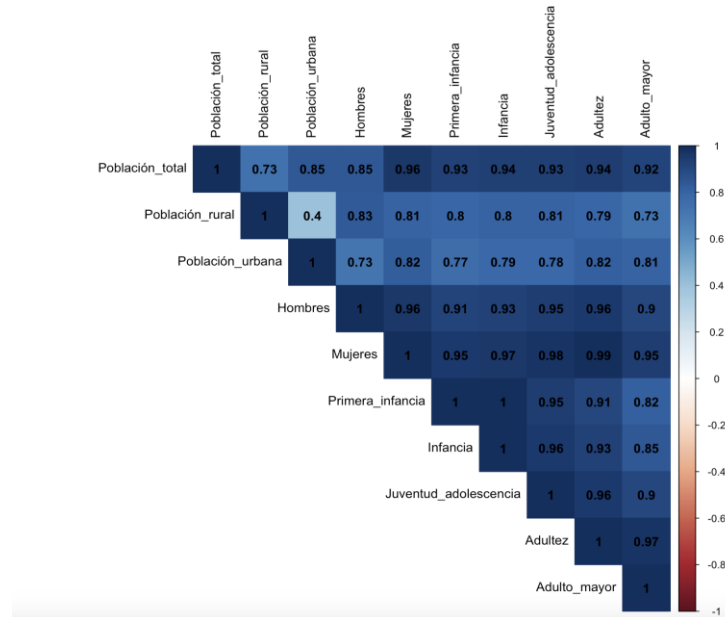


Fuente: Realización propia

Como se puede observar en la anterior gráfica, se observa una correlación positiva fuerte entre los diferentes niveles de cobertura (primaria, secundaria y media). Sin embargo, al analizar las coberturas parciales con la neta se evidencia que la correlación entre las variables es débil. Los resultados obtenidos mediante esta gráfica permiten identificar que introducir todas las variables al modelo final de conglomerados puede generar un problema de multicolinealidad, este proceso se validará en el siguiente capítulo.

Figura 16

Diagramas de calor de Población

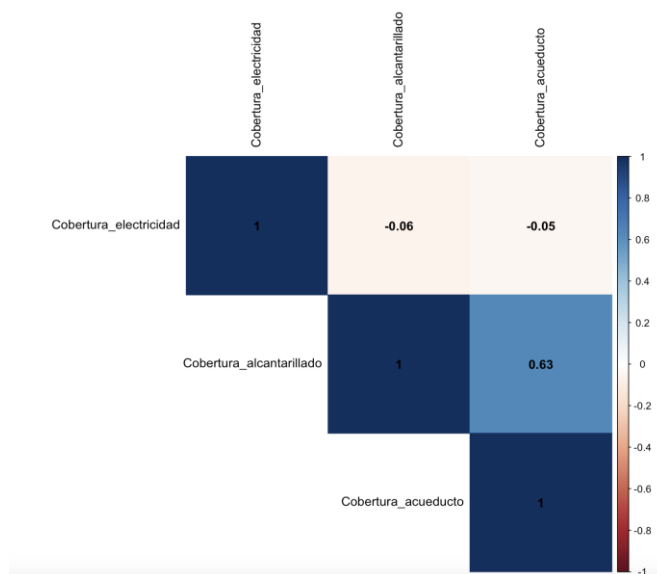


Fuente: Realización propia

En relación con las variables demográficas poblacionales se evidencia que existe una correlación fuerte en la mayoría de los atributos. Este comportamiento implica que al momento de realizar el modelado es importante considerar qué variables tienen mayor impacto en la explicación de la mortalidad de ECV y así evitar problemas de multicolinealidad que pueda afectar la estabilidad del modelo.

Figura 17

Correlación de coberturas a servicios públicos

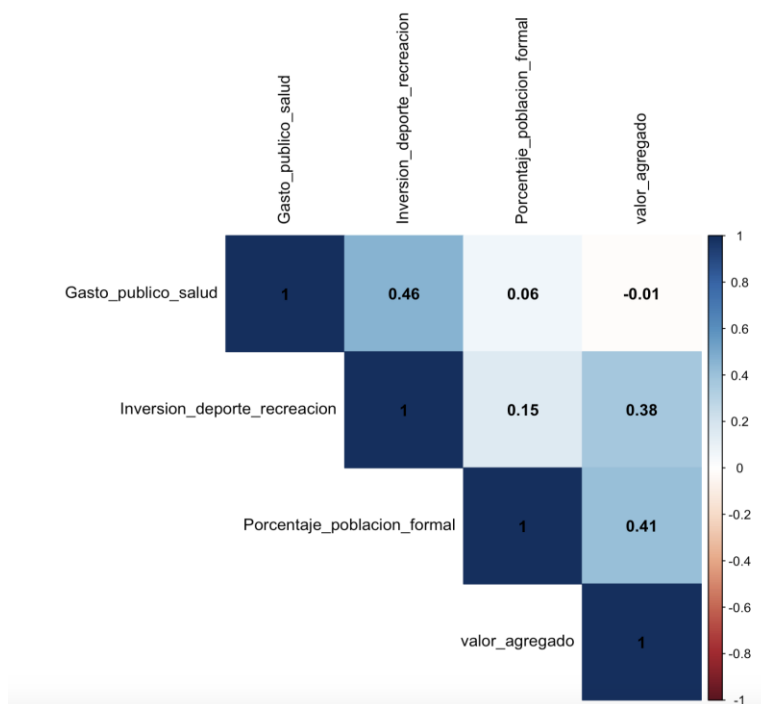


Fuente: Realización propia

En la anterior figura, se observa que entre las variables de cobertura de alcantarillado, electricidad y acueducto no se observa una correlación muy fuerte entre ellas. La única excepción de una correlación moderadamente fuerte se da entre la cobertura de alcantarillado y acueducto, lo cual guarda relación, al considerar que estos servicios públicos se instalan casi de forma simultánea.

Figura 18

Correlación variables de inversión y población formal



Fuente: Realización propia

En relación con las variables de valor agregado, inversiones y porcentaje de población formal en el mercado laboral se evidencia que existe nuevamente una baja correlación, especialmente entre variables como valor agregado, inversión en deporte y recreación y gasto público en salud. Se evidencia una correlación moderada entre valor agregado y población formal lo cual es lógico al reconocer que estas dos variables están íntimamente relacionadas con la producción nacional.

Finalmente, es importante mencionar que los análisis de correlación se realizaron a todas las variables objeto de estudio, sin embargo, en el documento se presentaron algunas de las más importantes según los resultados obtenidos.

5.4. Preparación y manipulación previa de datos

Luego de realizar el análisis exploratorio de los datos, se procedió con la manipulación y transformación de algunas variables del dataframe. En un primer momento, se procedió a realizar el cálculo de unos índices poblacionales en las variables demográficas con el objetivo de evitar que los clústeres presenten algunos de los siguientes errores:

- *Sesgo en el análisis:* Los municipios con mayor cantidad de población pueden dominar el análisis de los clústeres, ya que, los cálculos pueden estar dominados por

regiones con mayor cantidad de habitantes. Lo anterior implica que los clústeres no reflejen patrones espaciales verdaderos, sino simplemente la concentración de la población.

- *Sobreestimación de poblaciones altas:* Al no transformar las variables, los municipios con poblaciones más grandes generan correlaciones artificiales entre las variables debido al tamaño de la población y no las relaciones subyacentes entre los atributos a analizar.
- *Desbalance de los clústeres:* Al trabajar con las variables poblacionales sin transformación se puede inflar artificialmente la homogeneidad dentro de los clústeres y exagerar las diferencias entre ellos [57].

Al considerar los problemas que puede generar el uso de variables poblacionales sin la transformación adecuada, se procedió a generar unos índices poblacionales. Estos se aplicaron tanto a las variables demográficas (población por sexo, edad, zona de residencia), así como a los datos de mortalidad por ECV. El cálculo del índice se realizó considerando una tasa por 100.000 habitantes, y se aplicó la siguiente fórmula.

$$\text{Índice poblacional} = \left(\frac{\text{Número de casos o eventos}}{\text{Población total}} \right) \times 100.000$$

De igual manera, el procesamiento previo de los datos incluyó la estandarización de los datos que tiene como finalidad: (a) igualar la influencia de las variables para evitar sesgos en variables que tienen mayores escalas y pueden dominar el proceso de agrupamiento de los clústeres; (b) ayudar en la convergencia de los algoritmos al considerar que método K-means basa sus cálculos en las distancias (euclidiana, Manhattan) y (c) reducción de la varianza para facilitar la detección de patrones y estructuras subyacentes en los datos [57]. El proceso para la estandarización se realizó a través de la normalización Z-score que permite convertir los datos de tal manera para que tengan una media de 0 y una desviación estándar de 1.

6. CONGLOMERADOS ESPACIALES DE CONDICIONES SOCIODEMOGRAFICAS Y MORTALIDAD POR ECV

En el presente capítulo se presenta el modelo de estadística espacial que permite identificar la relación existente entre las condiciones sociodemográficas y la mortalidad por ECV en la región pacífica colombiana, teniendo como punto de partida los datos en el periodo 2016-2020. Al finalizar el capítulo, se podrán obtener conclusiones importantes que aporten a las entidades públicas a tomar decisiones considerando datos de periodos anteriores para que se fortalezca la focalización de recursos en políticas públicas de salud, y específicamente en programas de promoción y prevención en salud.

6.1. Análisis de incidencia de variable

En un primer momento, el análisis de la estadística espacial comienza realizando un análisis de incidencia de las variables explicativas sobre la variable dependiente para identificar qué atributos tienen un mayor peso explicativo en relación con la mortalidad por ECV. Este paso es el inicial debido a que se requiere seleccionar las variables que van a hacer parte del modelado final.

Teniendo en cuenta lo expresado previamente, se procedió a analizar la incidencia de cada una de las variables de la mortalidad por ECV en relación con las condiciones socioeconómicas identificadas previamente. Este procedimiento se sintetiza en la siguiente tabla.

Tabla 4

Incidencia variables independientes vs tasa ECV

Variable	Valor t	Probabilidad	Variable	Valor t	Probabilidad
IRCA	-0.514	0.607487	Mujeres	-3.654e+10	<2e-16***
Cobertura electricidad	2.763	0.005854**	Hombres	-3.663e+10	<2e-16***
Cobertura acueducto	1.978	0.0483*	P. urbana	2.570e+10	<2e-16***
Cobertura alcantarillado	2.854	0.00441**	P. rural	-6.104e+09	<2e-16***
SGSSS	3.590	0.000348***	Adulto mayor	-3.463e+10	<2e-16***
NBI	-3.462	0.000562***	Primera infancia	-0.283	0.777
Cobertura primaria	1.81	0.0707	Infancia	-1.245	0.213
Cobertura secundaria	2.318	0.0207*	Juventud y adolescencia	-0.129	0.897
Cobertura media	3.214	0.00136**	Adultez	9.548	<2e-16***

Cobertura neta	2.477	0.013450*	Valor agregado	8.014	3.49e-15***
Universitarios	35.972	< 2e-16***	Gasto público en salud	5.360	1.06e-07***
Población	3.675e+10	<2e-16***	Población formal	7.973	4.78e-15***
Inversión en deporte y recreación		-2.122		0.0341*	

Nota. * significancia al 95%, ** significancia al 99%, *** significancia al 99.9%,

La implementación de los procesos de correlación entre las variables de interés y la tasa por ECV permitió identificar qué atributos se deberían escoger o seleccionar para continuar con el proceso de clústeres. A partir de ello, se realizó una serie de selecciones y análisis previo donde se decidió escoger como variables para la clusterización la tasa por ECV, la tasa de defunción, el NBI, el valor agregado, la cobertura neta en educación, el gasto público en salud, el IRCA y la inversión de los municipios en programas de deporte y recreación.

Teniendo en cuenta los datos presentados anteriormente, se puede evidenciar que la variable de mortalidad de las enfermedades hipertensivas presenta relación estadísticamente significativa con la mayoría de los atributos recopilados en el estudio. Para continuar con el análisis inicial del modelo de clusterización se realizó una visualización inicial de los datos agrupándolos por los diferentes municipios de cada uno de los departamentos de la región pacífica colombiana. Este proceso permitió identificar algunos patrones iniciales en cada una de las zonas de interés como por ejemplo mayores tasas de NBI en los municipios del departamento de Chocó que en del Valle del Cauca, mayores tasas de cobertura neta y producto económico en los municipios del Valle del Cauca en relación con las regiones de los otros departamentos de la zona pacífica colombiana (ver figura 19).

Por otro lado, antes de calcular los componentes principales de las variables, se procedió a identificar el comportamiento de la tasa de mortalidad por ECV. En ese orden de ideas, el estudio identificó que en cada uno de los departamentos se presenta comportamientos similares, sin embargo, sí se evidencia un incremento de las muertes por ECV desde el año 2020 (ver figura 20). Este comportamiento se puede explicar por el fenómeno ocurrido por la COVID-19 donde se evidencia que durante dicho periodo se presentaron mayores casos de mortalidad por enfermedades cardiovasculares debido a una multiplicidad de fenómenos entre los cuales se encuentran la preexistencia de una enfermedad por ECV y el incremento en la posibilidad de presentar alteraciones en la coagulación que aumentaba el número de infartos por miocardio [58] e incremento de accidente cardiovascular, especialmente en personas con mayor marginación social y económica [58, 59].

Figura 19

Dispersión de datos agrupados por municipios y departamentos de la zona pacifica colombiana en el periodo 2016-2020.

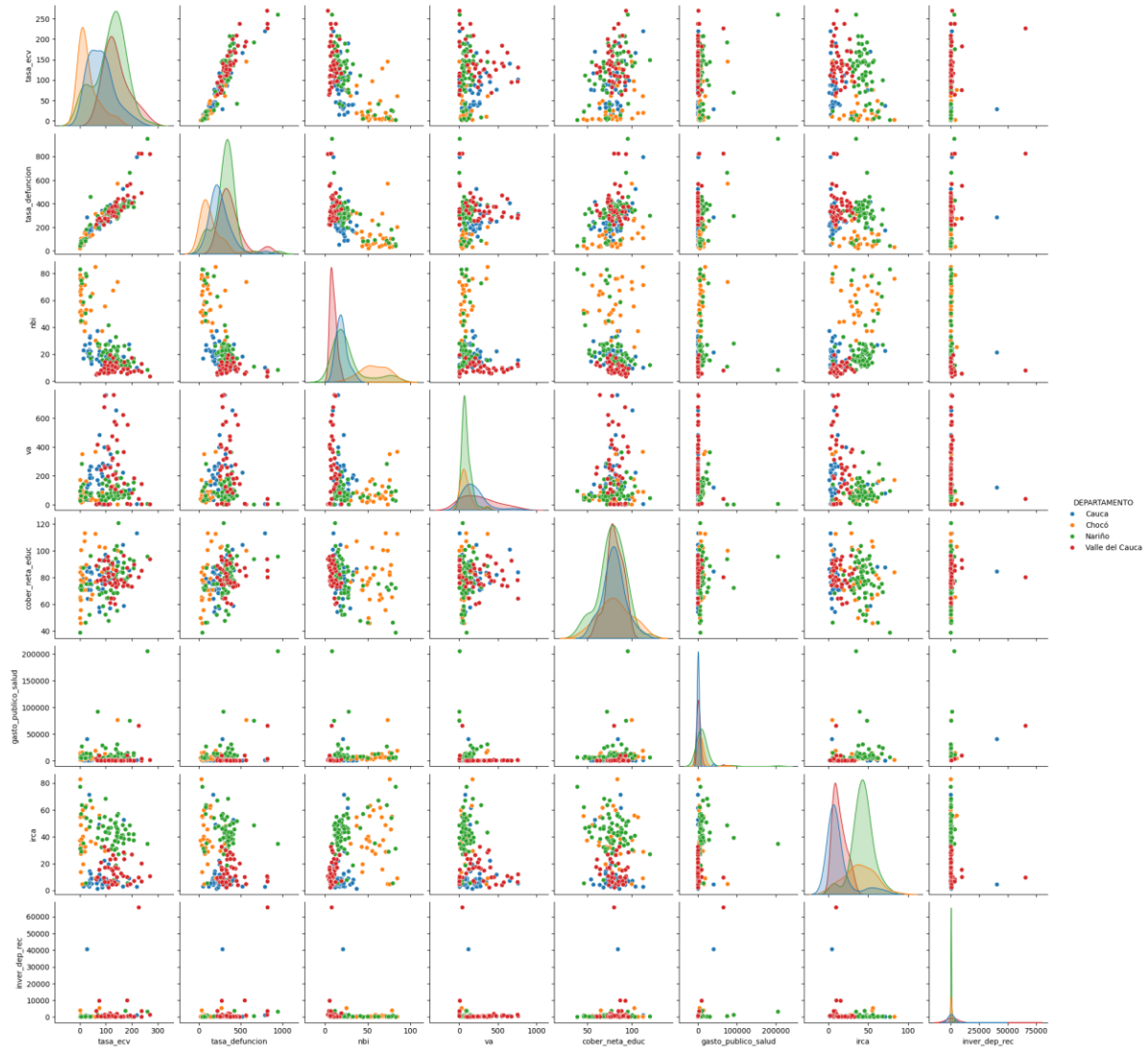
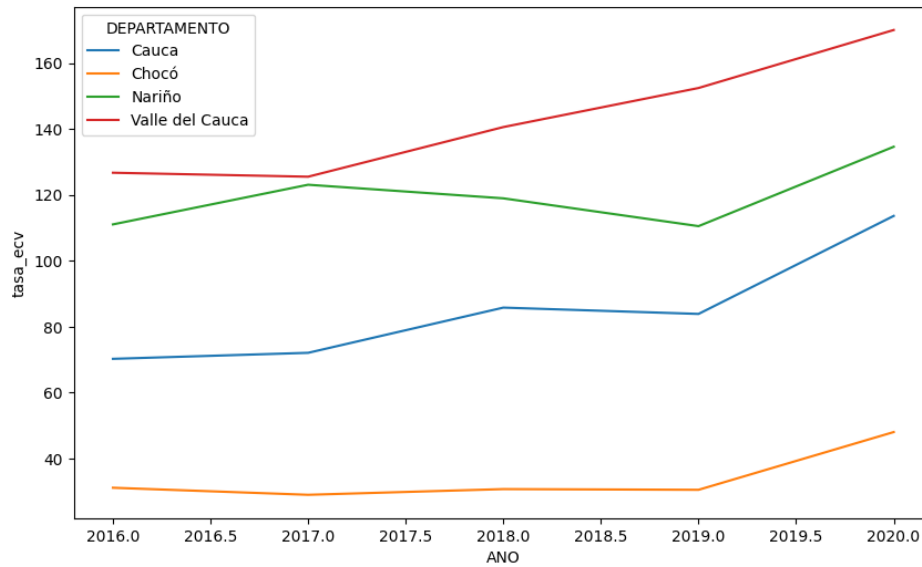


Figura 20

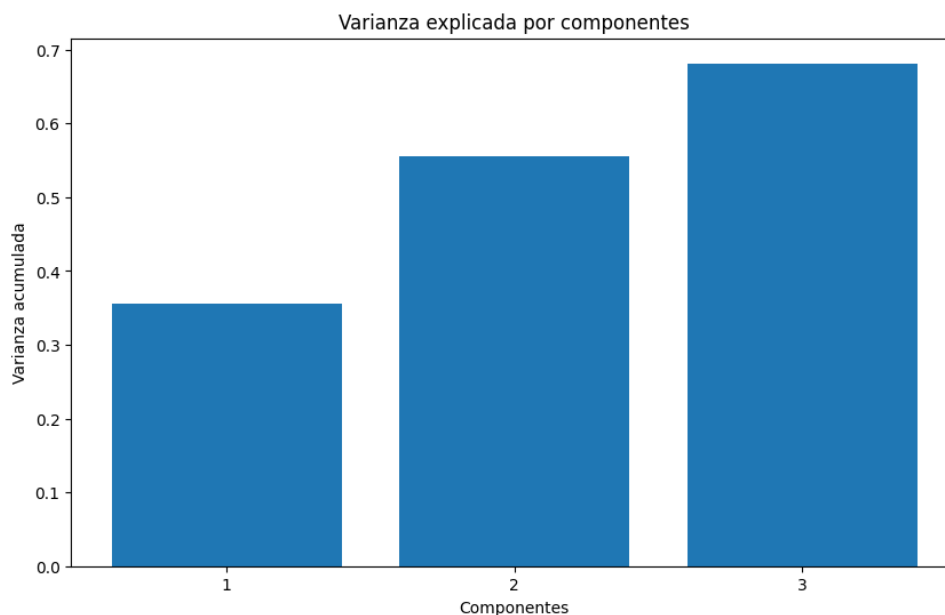
Comportamiento de la tasa por cada 100 mil habitantes de ECV por departamentos de la zona pacifica, 2016-2020



Al finalizar, el proceso de análisis inicial se continuó con el cálculo de los componentes principales (ACP) con las variables mencionadas previamente. Los resultados del ACP evidencia que solo 3 componentes explican casi un 70% de la varianza total, es decir, que existe una baja variabilidad en algunos atributos.

Figura 21

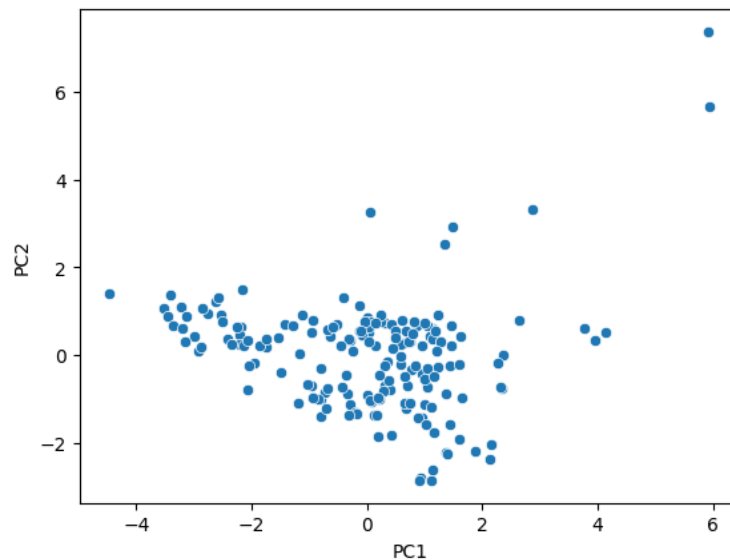
Varianza de las variables, resultados ACP



Con los resultados obtenidos en el modelo ACP, se escogió los dos primeros componentes para analizar la distribución de los datos dentro de un plano. Al realizar dicho proceso, se evidenció que existió algunos datos que se alejaban de la distribución de la mayoría de los datos (observaciones atípicas). Estos resultados se deben considerar debido a que puedan afectar la consistencia y comportamiento de los clústeres.

Figura 22

Representación en el plano cartesianos de los dos primeros componentes ACP



Debido a los resultados presentados previamente, se ejecutó iterativamente varios modelos para dejar de forma aislada aquellos municipios que tuvieran un comportamiento atípico y no afectaran los resultados del clúster. Posterior a este procedimiento, se procedió a realizar la evaluación de los modelos de clusterización: k-means con ACP y k-means con T-SNE. Posteriormente, se desarrolla un proceso de análisis donde se comparan los resultados obtenidos mediante el algoritmo de K-means con T-SNE con un método alternativo para validar las evidencias arrojadas por el algoritmo. Para el desarrollo del proceso de validación con diferentes técnicas de modelación se parte nuevamente de la representación bidimensional obtenida del algoritmo t-SNE, dado que representa de mejor forma las relaciones en un plano cartesiano a comparación del análisis PCA. Dicho lo anterior se iteraron entre 3 algoritmos adicionales (Kmeans++, Optics y Gassian Mixture) la idea principal con esto es poder medir la consistencia de los resultados usando diferentes algoritmos.

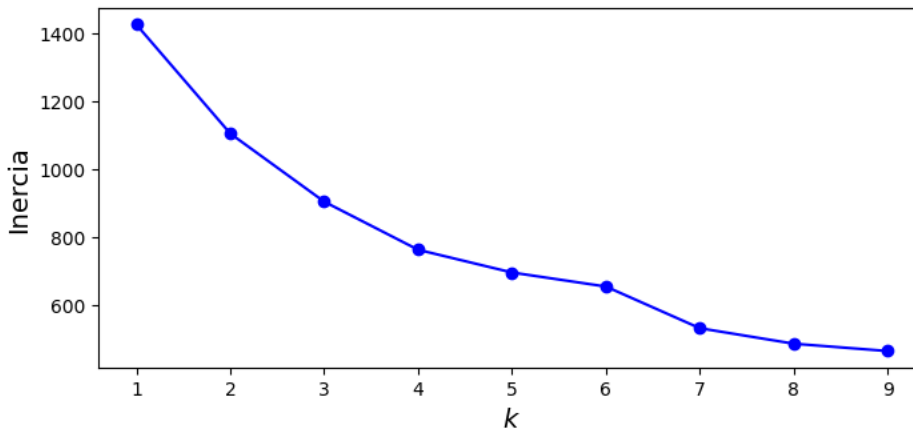
Finalmente, también se comparó con el K-Means sin reducción de dimensionalidad. La aplicación de estos seis modelos y su correspondiente optimización de hiperparámetros se describe en las secciones 6.2 a 6.9. Finalmente, la evaluación respecto a las métricas es presentada en la sección 6.8 y la sección 6.9 presenta la interpretación de los clústeres considerando la metodología que obtuvo los criterios estadísticamente más adecuados para el análisis.

6.2 Resultados de clústeres por medio de K-means con ACP

Para comenzar los procesos de cálculo de los clústeres de K-means con ACP, en un primer momento se aplicó las métricas de la silueta y de inercia con el objetivo de determinar el número óptimo de clústeres en función de los datos obtenidos en el análisis de ACP. Al aplicar el método del codo, no se logró encontrar una diferencia significativa en la disminución de las distancias al incluir un clúster adicional, por lo que el resultado por este método no es concluyente. Los resultados obtenidos mediante este método también permiten inferir que al considerar la nube de puntos mostrada en la figura 23 no es fácil separar los datos de forma lineal.

Figura 23

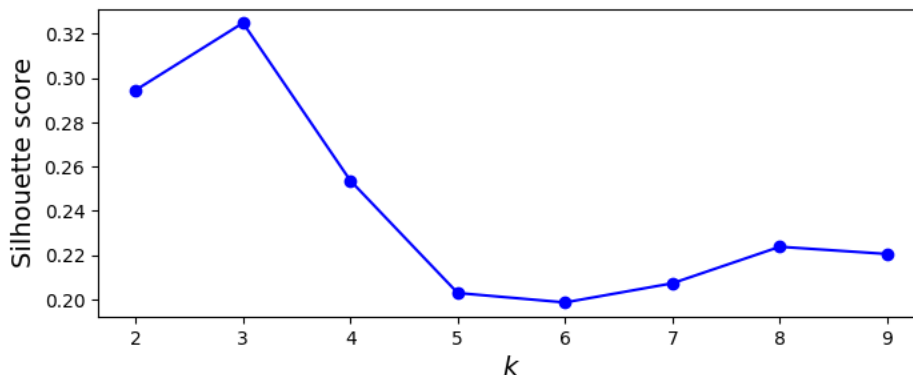
Resultados del número de clústeres –método del codo



Al evidenciar las limitaciones obtenidas en el método del codo, se procedió a aplicar la técnica de la silueta. En esta métrica si se evidenció de forma clara que el número óptimo de clústeres debería ser 3.

Figura 24

Resultados del número de clústeres –método de la silueta



Teniendo en cuenta los resultados obtenidos previamente, se procedió a realizar la primera agrupación mediante clústeres. Al realizar el primer procedimiento, se encontró que la

clusterización presentaba un comportamiento atípico por la alta diferencia que se presentaba en la cantidad de municipios que constituían cada clúster (ver figura 25a). Al observar dichos resultados, fue necesario realizar varias iteraciones para dejar por un lado los municipios de la zona pacífica colombiana que presentaba comportamientos muy diferentes al resto del conjunto de datos.

Al realizar dicho proceso, se llegó a una composición de dos clústeres. Frente a ello, se debe resaltar el clúster -1 que contiene todos los municipios que no pudieron ser categorizados correctamente al considerar las imitaciones dadas en las relaciones lineales obtenidas en los dos componentes principales (ver figura 25b). A pesar de que el segundo modelo presenta un comportamiento más estable, no es recomendable utilizar los resultados obtenidos en el clúster -1 porque implica que en este se agrupa todos los municipios que no pudieron ser agrupados adecuadamente. Lo anterior implica, que en dicha clusterización existe una variabilidad intrínseca elevada dentro del grupo, lo cual va en contra de la lógica de los clústeres y la metodología del K-means.

Figura 25

Resultados del número de municipios por cada clúster

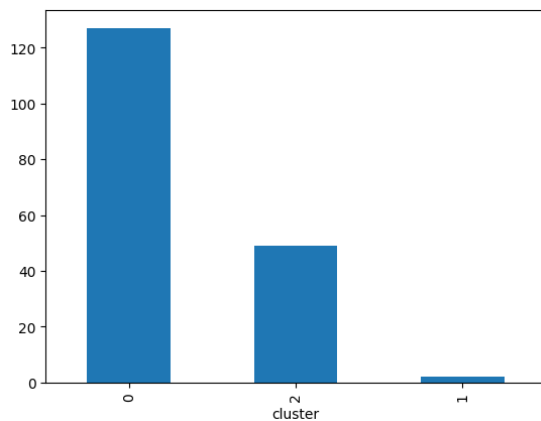


Figura 25a ACP inicial

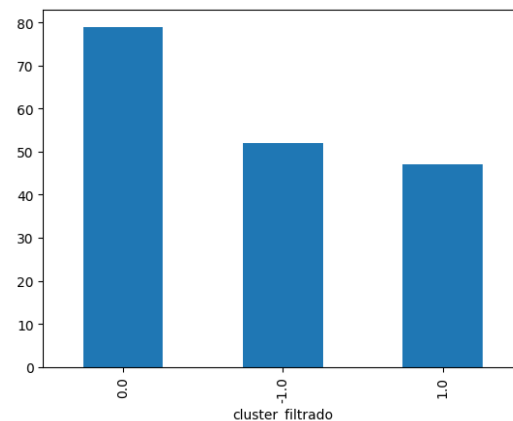


Figura 25b ACP con iteraciones

Al observar los resultados obtenidos previamente, se decide adoptar un modelo de reducción de dimensiones más potente, el T-SNE, que no depende de la linealidad de los datos, sino que se basa en la distribución de los mismos a lo largo de todas las características en un espacio multidimensional. La idea de esta reducción es similar al análisis de ACP, con la diferencia, de que a través del método T-SNE se mantiene las relaciones entre municipios similares y se separa a aquellos que presentan un comportamiento diferente.

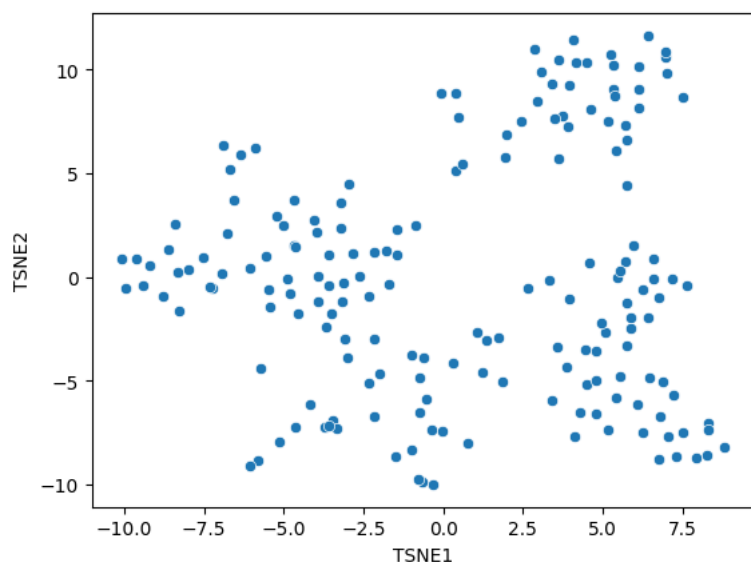
El método t-SNE es una técnica de reducción de dimensionalidad usada para transformar datos complejos en una representación que mantenga la estructura de proximidad entre puntos. En otras palabras, intenta conservar las relaciones de cercanía entre los puntos de datos originales cuando se proyectan en un espacio de menor dimensión. Una de las principales utilidades de este método es calcular la probabilidad de que dos puntos en el espacio de alta dimensión sean similares (vecinos), basándose en la distancia entre ellos mediante el cálculo de la distribución gaussiana. A raíz de ello, el método t-SNE minimiza

la diferencia entre las distribuciones utilizando una función de costo que mide la divergencia entre las probabilidades del espacio original y el espacio reducido.

Al aplicar este método de reducción de la dimensionalidad, se logró obtener una mejor comprensión de los resultados porque se evidencia una separación de los datos en grupos o clúster mejor definidos (ver figura 26).

Figura 26

Resultados de la reducción de dimensionalidad por t-SNE



Considerando los resultados obtenidos previamente se procedió a realizar la clusterización a través del algoritmo de K-means.

6.3. Resultados de clústeres por medio de K-means con T-SNE

En un segundo momento, se procedió a realizar una nueva modelación a través del método de K-means mediante la técnica de reducción por T-SNE debido a que la reducción de la dimensionalidad mediante ACP indicó que uno de los clústeres presentaba municipios que no pudieron ser agrupados adecuadamente. A raíz de ello, se procedió a calcular el algoritmo K-means con la reducción de dimensionalidad mediante T-SNE, para corroborar los resultados obtenidos a través de la reducción de la dimensionalidad, también se procedió a inicializar de forma aleatoria el hiperparametro K del modelo con el objetivo de verificar si el número de clústeres identificados mediante las técnicas de reducción de dimensionalidad era acertados. Por ello, al dar inicio a los centroides mediante el algoritmo de K-means ++ se encontró que el número adecuado de clústeres eran 4.

Antes de realizar el proceso de clusterización, se agregó una variable de ubicación geográfica denominada “norte”, el cual determina si el municipio se encuentra al norte del país (departamentos de Chocó o Valle del Cauca) o no. Con esas visualizaciones se podrá

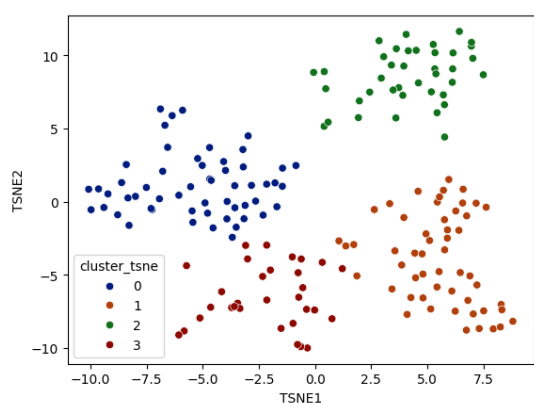
caracterizar a cada uno de los 178 municipios y de esa forma determinar un perfil que pueda ser usado posteriormente para mejorar los procesos de focalización de recursos en las políticas públicas de salud.

Al aplicar el algoritmo de K-means con los dos componentes principales de método t-SNE se obtuvieron los resultados presentados en la figura 27. Así, en la figura 27a se observan los cuatro clúster claramente identificados en una distribución bidimensional; mientras que en la figura 27b se puede determinar la posición de cada municipio en los clústeres identificados. Para facilitar la visualización y comprensión de los datos obtenidos, cada departamento fue representado por un color específico, y cada punto, determinar un municipio de dicha región.

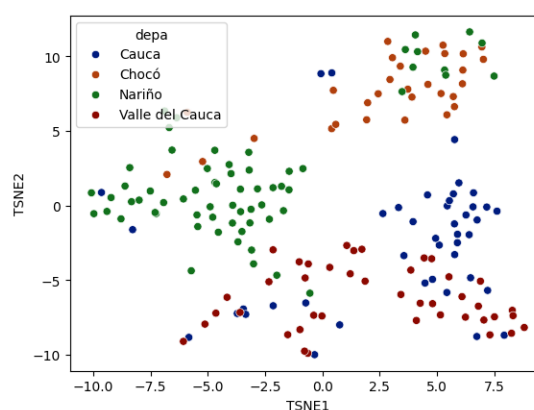
Observando la figura 27b, se evidencian que la mayoría de los municipios de cada departamento se concentran en algún clúster. Sin embargo, existen otras regiones donde se representa una mezcla entre diferentes municipios de la zona pacífica colombiana.

Figura 27

Resultados de clusterización mediante algoritmo de K-means.



27^a Distribución bidimensional por clúster

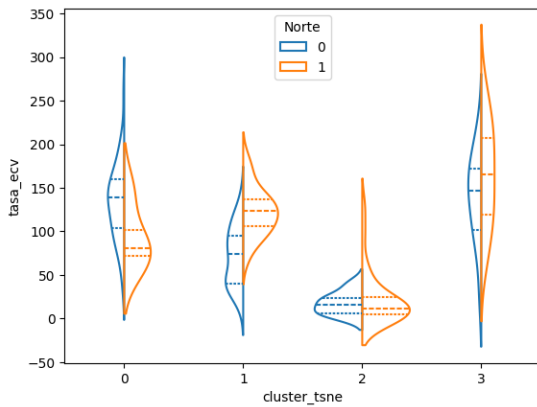


27^b Distribución bidimensional por departamento

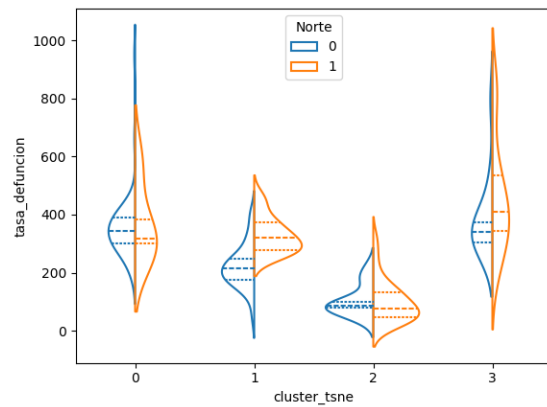
Al agregar la variable norte, se pudo desagregar los diferentes resultados de los clústeres considerando la tasa por ECV y los atributos analizados en el proceso de clusterización. Los resultados de la desagregación se observan en las figuras 28.

Figura 28

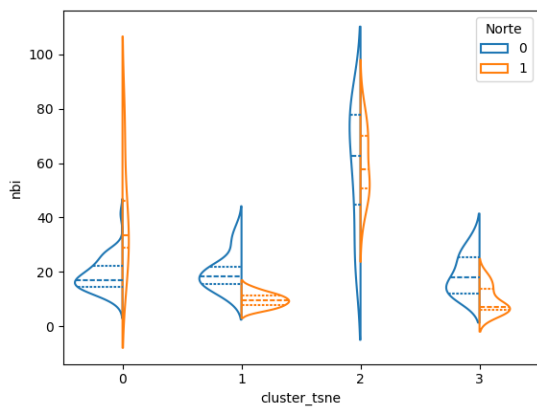
Resultados de clústeres desagregados por tasa de ECV



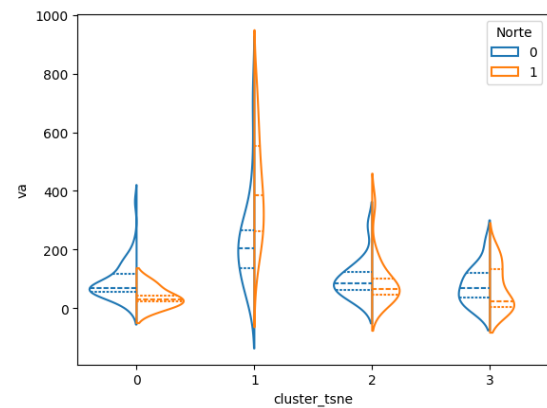
28a Distribución de la Tasa ECV por Clústeres y Ubicación Geográfica (Norte/Sur)



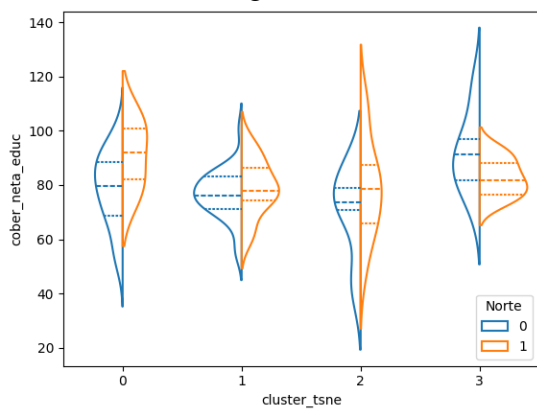
28b Distribución de la Tasa de Defunción por Clústeres y Ubicación Geográfica (Norte/Sur)



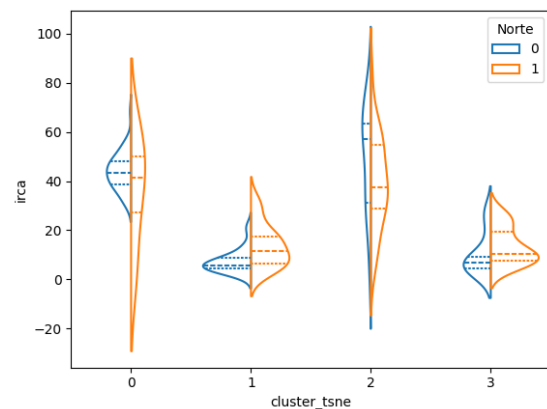
28c Distribución de las Necesidades Básicas Insatisfechas (NBI) por Clústeres y Ubicación Geográfica (Norte/Sur)



28d Distribución del Valor Agregado (VA) por Clústeres y Ubicación Geográfica (Norte/Sur)



28e Distribución de la Cobertura Neta Educativa por Clústeres y Ubicación Geográfica (Norte/Sur)



28f Distribución del IRCA por Clústeres y Ubicación Geográfica (Norte/Sur)

6.4. Resultados de clústeres por medio de Kmeans++ con T-SNE

Es una variación del algoritmo tradicional de K-means que tiene como objetivo mejorar la forma como se inicializan los centroides, esta variación incrementa el nivel de ajuste porque la inicialización de los centroides es crucial porque afecta la convergencia y calidad de las agrupaciones finales del algoritmo. En consecuencia, la optimización en la inicialización del centroide genera una menor distorsión y mejora la calidad del clustering.

En esta variación se modificó la forma en la que se inicializan los centroides, reduciendo la aleatoriedad del ejercicio. Frente a los resultados de esta primera variación, es importante señalar que los hallazgos fueron idénticos al obtenido mediante el método K-means. Estos resultados se evidencian en la siguiente figura y tabla.

Figura 29

Resultados de clústeres obtenidos mediante el método K-means++

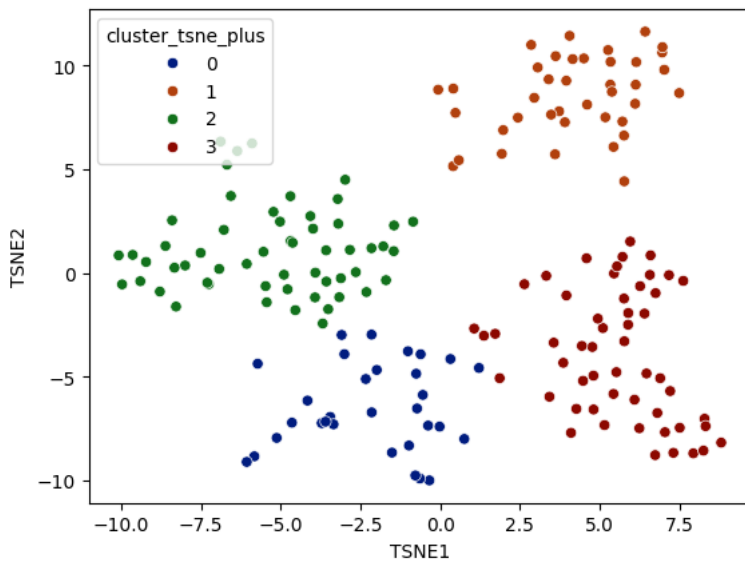


Tabla 5

Asignaciones de clústeres mediante el método K-means++

cluster_tsne_plus	0	1	2	3
cluster_tsne				
0	0	0	55	0
1	0	0	0	53
2	0	39	0	0
3	31	0	0	0

En conclusión, revisando los resultados obtenidos mediante la modelación de K-means++, se evidencia que no hay variación entre la asignación de los clústeres, cada grupo está compuesto por las mismas observaciones en cada grupo. En resumen, se puede afirmar que con esta primera validación se obtienen resultados que validan los hallazgos expuestos en el desarrollo del capítulo.

6.5. Resultados de clústeres por medio de Gaussian Mixture con T-SNE

El modelo de Mezcla Gaussiana es un enfoque estadístico que permite modelar la distribución de los datos en un espacio de características diversas, es un modelo que se basa en una distribución de datos gaussiana. A diferencia del algoritmo de K-means donde se asume que los clústeres son esféricos y tienen tamaños similares, este modelo permite que los clústeres tengan distintas formas, tamaños y orientaciones, permitiendo que se más potente y flexible en algunos casos. Este modelado presenta un mejor ajuste en datos que tienen baja probabilidad de pertenecer a uno de los clústeres, por ello, es de gran ayuda para la detección de anomalías.

Para la realización de este procedimiento, se obtuvo que las variaciones fueron muy pequeñas, solamente cambiaron 7 observaciones de 178, lo que sigue manteniendo una consistencia fuerte a las asignaciones realizadas. Estas variaciones están explicadas principalmente por la ubicación en la frontera entre clústeres, dicho de otra forma, en la representación bidimensional de los datos hay algunos municipios que tienen características que son difíciles de categorizar a un grupo con respecto a las características de los demás municipios.

En ese orden de ideas, se puede mencionar que, hasta el momento, los clústeres obtenidos en la primera modelación siguen siendo consistentes entre 3 algoritmos. Los resultados obtenidos de este proceso de modelación se encuentran en la siguiente figura y tabla.

Figura 30

Resultados de clústeres obtenidos mediante el método Gaussian Mixture

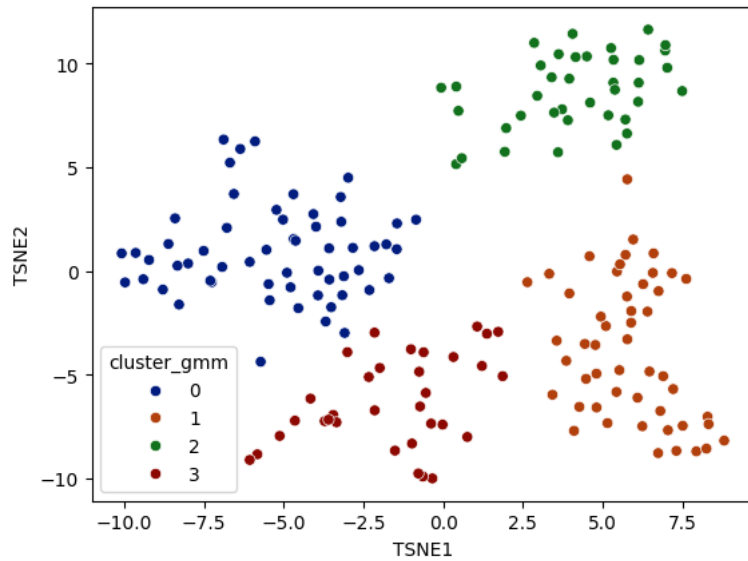


Tabla 6

Asignaciones de clústeres mediante el método Gaussian Mixture

cluster_gmm	0	1	2	3
cluster_tsne				
0	55	0	0	0
1	0	49	0	4
2	0	1	38	0
3	2	0	0	29

En consecuencia, se puede afirmar que hasta el momento los resultados obtenidos por el algoritmo K-means siguen siendo consistentes.

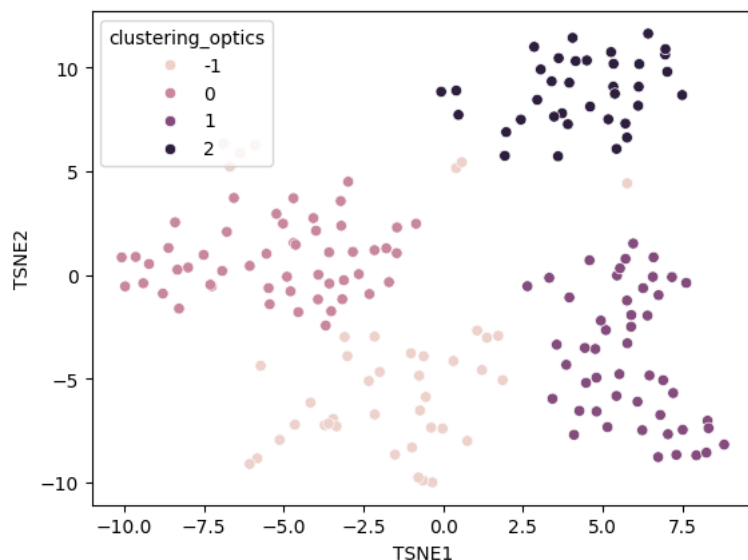
6.6. Resultados de clústeres por medio de OPTICS con T-SNE

OPTICS (Ordering Points To Identify the Clustering Structure), estrechamente relacionado con DBSCAN, encuentra núcleos de muestra de alta densidad y expande los clústeres a partir de ellos. Este tipo de algoritmo está diseñado para identificar la estructura de los datos espaciales, incluso cuando tienen densidades variables, por ello, este modelo se basa en la densidad de los datos. Este modelo tiene como ventaja el manejo de densidades variables porque puede tener un mejor manejo de ellas y descubrir las estructuras de clustering complejas. De igual manera, este tipo de algoritmo puede revelar una estructura jerárquica en los datos y así se puede mostrar algunos subclústeres dentro de los clústeres más grandes.

Un aspecto relevante de este algoritmo es que encuentra aquellas observaciones que tienen una menor densidad, para estos casos estas observaciones son agregadas a un clúster -1. Los resultados de este algoritmo se presentan a continuación.

Figura 31

Resultados de clústeres obtenidos mediante el método OPTICS



Según los resultados presentados previamente, se identificó una consistencia con los resultados obtenidos con K-means. Para este caso en concreto, el clúster 3 de K-means es el clúster -1 de OPTICS, a su vez esto indica que esas observaciones quedaron agregadas a un clúster de baja densidad.

Para los otros clústeres hay solo 11 observaciones que se habían agregado a otros grupos, en línea con lo mencionado anteriormente representan observaciones difíciles de clasificar dada su lejanía con otros datos, este algoritmo puede fácilmente agregar en el clúster -1 aquellos municipios con características particulares que vale la pena analizar a mayor profundidad de forma separada. De forma general, el número de asignaciones en cada clúster se encuentran resumidas en la tabla No 7.

Tabla 7

Asignaciones de clústeres mediante el método OPTICS

clustering_optics	-1	0	1	2
cluster_tsne				
0	4	51	0	0
1	4	0	49	0
2	3	0	0	36
3	31	0	0	0

6.7. Resultados de clústeres por medio de K-means (sin reducción de dimensionalidad)

Con el objetivo de verificar los resultados de las modelaciones previas, se procedió a realizar la clusterización sin considerar técnicas de reducción de la dimensionalidad, es decir, sin realizar ningún tratamiento previo a los datos. El objetivo de realizar este último modelo se justifica al considerar que se pretende evaluar el comportamiento de los datos originales en el proceso de clusterización y evidenciar si existe un ajuste razonable o consistente con los datos sin tratamiento, o en efecto, se requiere realizar un tratamiento previo al dataset.

A partir de ello, al aplicar el algoritmo de clusterización a los datos originales, se obtuvieron los siguientes resultados.

Tabla 8

Resultados clusterización datos originales

No. Clúster	Valor obtenido
0	83
3	53
2	40
1	2

De acuerdo con los resultados obtenidos previamente, se logra identificar que al realizar la clusterización de los datos originales, sin ningún tipo de transformación, se evidencia que la inicialización también genera 4 grupos diferenciales, sin embargo, el clúster número 1 contiene solamente dos observaciones. Este resultado puede indicar que al calcular los clústeres sin ninguna transformación se presentan datos atípicos que pueden afectar la consistencia del modelo.

6.8. Evaluación de modelos de clusterización

Al desarrollar todas las opciones de modelación de clusterización, se presentan algunas hipótesis sobre el funcionamiento de los modelos. Por ejemplo, una primera hipótesis es que la consistencia del modelo con los datos originales no es la más óptima, sin embargo, dicha afirmación requiere de la verificación de estadísticos que evidencien efectivamente si el modelo presenta algún tipo de problema. En relación con los otros modelos, se evidencia que no existen diferencias significativas, no obstante, dicha hipótesis también debe ser verificada mediante pruebas de validación estadísticas.

En ese orden de ideas, las 6 metodologías presentadas evidencian resultados similares en relación al número de clústeres y a la composición del número de municipios de cada uno de ellos. Sin embargo, la valoración final de cada uno de los modelos se realiza mediante los estadísticos de Bunn y Davis Boulding que se presentan en el siguiente apartado.

Para la evaluación de los modelos presentados en el capítulo, se tuvieron en cuenta dos criterios para su evaluación que tiene como objetivo identificar si los clústeres están bien separados y bien agrupados. Estos indicadores fueron:

- Índice de Dunn: Es un índice que permite medir si los clústeres están bien separados y compactos; el objetivo es que el índice de Dunn sea lo más alto posible porque indica que los clústeres están bien separados y compactados.
- Índice de Davis Boulding: El índice de DBI tiene como objetivo medir la dispersión dentro de los clústeres y la separación entre ellos. En este escenario, el objetivo es obtener un menor valor lo cual indica que un número bajo identifica clústeres bien separados y compactos.

Frente a los indicadores expuestos previamente, se realizó el cálculo para cada uno de las modelaciones obteniendo los siguientes resultados (ver anexo A).

Tabla 9

Evaluación de modelos mediante métricas de separación y dispersión

<i>Métrica/Modelo</i>	<i>Índice Davies-Boulding</i>	<i>Índice de Dunn</i>
<i>K-Means</i>	0.65	0.08
<i>OPTICS</i>	2.46	0.05
<i>Gaussian Mixture</i>	0.69	0.07
<i>K-means (sin tratamiento)</i>	1.18	0.05

Considerando los resultados obtenidos a través de la evaluación mediante las métricas de Dunn y Davis Boulding, se reconoce que el modelo con el mejor comportamiento es el obtenido con el algoritmo KMeans. De acuerdo con los resultados de la tabla No. 9 se reconoce que el modelo de K-means presenta mejores condiciones en término de separación de clústeres, dispersión dentro de los clústeres y son los más compactos.

A raíz de ello, se selecciona el modelo de K-means debido a que la evaluación de los criterios de consistencia, dispersión, separación y compactos. Al realizar la selección de la primera clusterización, se procede a realizar la validación cualitativa de los mismos a través de dos mecanismos: (a) la validación de estudios previos y (b) la validación con expertos para indagar con ellos el nivel de certeza, afinidad y congruencia con las realidades evidenciadas en los territorios.

6.9. Interpretación de los clústeres obtenidos con K-means con T-SNE (otra subsección)

Considerando la evaluación de los métodos de clusterización, se logró evidenciar que el mejor algoritmo para los datos analizados en el presente proyecto fue el método de K-means utilizando la técnica de reducción de dimensionalidad mediante T-SNE.

Al realizar la clusterización mediante K-means con T-SNE se logró obtener las siguientes interpretaciones para cada uno de los clústeres identificados. Frente a la interpretación de los clústeres es importante reconocer que una de las limitaciones del estudio estuvo relacionada con la inclusión del tiempo en algunas variables como hábitos de vida saludables en deporte y recreación porque dichos procesos pueden presentar variaciones en periodos de tiempo muy largo, sin embargo, la presencia de datos fue limitada. No obstante, cabe reconocer que al ser un estudio de tipo exploratorio para la región pacífica este proyecto puede dar paso a estudios posteriores que consideran un marco temporal más amplio y se comparen con los resultados obtenidos aquí.

Clúster 3: “Municipios Resilientes con Desafíos de Salud”

Los municipios pertenecientes a este clúster se caracterizan por presentar altos niveles promedio de mortalidad por enfermedades cardiovasculares (ECV), a pesar de tener bajos niveles de Necesidades Básicas Insatisfechas (NBI). Esto sugiere que, aunque la mayoría de la población cuenta con acceso a servicios básicos y una calidad de vida relativamente buena, los factores de riesgo para ECV siguen siendo prevalentes. Estos municipios también muestran bajos niveles de Valor Agregado, lo que indica una economía menos desarrollada o diversificada, lo cual podría limitar las oportunidades económicas y aumentar el estrés, un factor contribuyente a las ECV.

A pesar de estos desafíos, se observa un alto gasto en salud y una considerable inversión en recreación y deporte, lo que sugiere un esfuerzo significativo por parte de las autoridades locales para mitigar los riesgos de salud mediante la promoción de un estilo de vida más activo y la mejora en la infraestructura de salud. Sin embargo, la persistencia de altas tasas de ECV indica la necesidad de reforzar estas intervenciones o explorar nuevas estrategias de salud pública para reducir la incidencia de estas enfermedades. Frente al análisis concreto de este clúster se debe reconocer que una limitación en relación a la interpretación de las inversiones en recreación y deporte porque: (a) la adopción de hábitos saludables puede requerir un periodo de intervención muy alto, situación que no pudo ser considerada en el estudio debido a la ausencia de datos a muy largo plazo, por lo tanto, su impacto no puede ser considerado. No obstante, este estudio puede ser una primera aproximación para consolidar bases de datos en los municipios de la región pacífica colombiana y evaluar su relación con la aparición de ECV; y (b) porque no toda la inversión realizada en recreación

y deporte se dedica a programas o estrategias de intervención, sino también a mantenimiento de infraestructura.

Los municipios que se integran en este grupo pueden ser considerados como un claro ejemplo para una mejor focalización de los recursos públicos, porque a pesar de las altas inversiones en salud, recreación y deporte siguen presentando altas tasas de mortalidad por ECV. Lo anterior, implica que desde las entidades públicas se debe reconsiderar y analizar porque las inversiones ejecutadas no tienen el impacto esperado, para ello, se recomienda a las entidades evaluar los programas ejecutados y a partir de ello, establecer nuevas estrategias y planes de intervención que respondan a las realidades del territorio.

Clúster 2: "Municipios Vulnerables con Riesgo Latente"

Los municipios que conforman este clúster se distinguen por tener bajos niveles por enfermedades cardiovasculares (ECV) y una baja mortalidad poblacional general. Sin embargo, estos resultados aparentemente positivos contrastan con su realidad socioeconómica, marcada por altos niveles de Necesidades Básicas Insatisfechas (NBI), lo que indica una población con acceso limitado a servicios esenciales.

Además, estos municipios presentan baja cobertura educativa y el índice de Riesgo de Calidad del Agua (IRCA) más alto entre todos los clústeres, lo que sugiere serios desafíos en términos de calidad del agua y posibles riesgos para la salud pública. La inversión en deporte y recreación es baja, lo que podría limitar las oportunidades para fomentar un estilo de vida saludable. Aunque las tasas de ECV son actualmente bajas, la combinación de altos niveles de NBI, baja educación, y malas condiciones de infraestructura sugiere un riesgo latente que podría agravar las condiciones de salud en el futuro si no se toman medidas adecuadas. Esto también se puede deber a la falta de información y acceso a servicios de salud que limitan la información disponible.

Los municipios que se encuentran en este clúster requieren de un proceso de fortalecimiento en inversión en algunas áreas del desarrollo humano tales como la educación y el acceso al recurso hídrico en condiciones de calidad. Estos factores se pueden abordar desde la articulación de políticas públicas en salud (mejoramiento del IRCA e inversión en deporte y recreación) y de políticas públicas en educación (cobertura neta) para disminuir las condiciones de riesgo a las cuales se pueden exponer los habitantes de este municipio a factores que puedan incrementar la mortalidad por ECV en el mediano y largo plazo.

Clúster 1: "Municipios de Contrastes Socioeconómicos"

Este grupo de municipios presenta una situación paradójica, caracterizada por una tasa de enfermedades cardiovasculares (ECV) de 94 por cada 100,000 habitantes, que ocupa el tercer lugar más alto entre los clústeres, y una tasa de defunción general de 269 por cada 100,000 habitantes, también en tercer lugar.

A pesar de estos indicadores preocupantes de salud, los municipios en este clúster se destacan por tener el Valor Agregado más alto (314), lo que sugiere una economía relativamente robusta en comparación con los otros clústeres. Sin embargo, este desarrollo económico no se refleja adecuadamente en el bienestar social, ya que la cobertura educativa es la segunda más baja, y el gasto en salud y recreación es el más bajo de los cuatro clústeres, lo que puede

estar contribuyendo a los altos niveles de mortalidad. En términos de calidad del agua, estos municipios tienen el índice de Riesgo de Calidad del Agua (IRCA) más bajo, lo que indica que la infraestructura es favorable. Este clúster representa una dicotomía entre prosperidad económica y deficiencias significativas en la inversión en salud y educación, lo que podría estar limitando el potencial para mejorar los indicadores de salud pública.

Al igual que el clúster 3, este tipo grupo de municipios puede ser calificado como un clúster de especial atención en los procesos de focalización de recursos en materia de políticas públicas. En ese orden de ideas, se puede argumentar que este grupo de municipios evidencia las desigualdades latentes en las regiones del país porque el crecimiento económico no se refleja en las condiciones de bienestar y desarrollo humano; por ello, es necesario que desde las entidades públicas se adopten medidas que permitan generar una mejor redistribución de la riqueza, apostando por inversiones en salud, recreación y deporte para disminuir las brechas existentes entre los diferentes municipios de la zona pacífica colombiana.

Clúster 0: "Municipios con Alta Carga Sanitaria y Desigualdades Latentes"

Los municipios de este clúster se enfrentan a serios desafíos en salud pública, evidenciados por la segunda tasa más alta de enfermedades cardiovasculares (ECV), con 129.75 casos por cada 100,000 habitantes, y una tasa de defunción de 355.16 por cada 100,000 habitantes, también la segunda más elevada de todos los clústeres.

A pesar de estos problemas de salud, los niveles de Necesidades Básicas Insatisfechas (NBI) son relativamente moderados en comparación con otros grupos (20.3%), lo que sugiere una población que no enfrenta carencias extremas. Sin embargo, el Valor Agregado es medio (85.5 miles de millones), lo que indica una economía menos desarrollada, y la cobertura educativa, aunque no la más baja, es limitada (79.5%), lo que puede influir negativamente en la capacidad de la población para acceder a mejores oportunidades. Notablemente, estos municipios registran el mayor gasto en salud de todos los clústeres, lo que refleja un esfuerzo significativo para enfrentar los problemas de salud, aunque no parece ser suficiente para reducir la alta mortalidad. Además, presentan el IRCA más alto, lo que indica problemas graves en la calidad del agua, un factor que podría estar exacerbando los problemas de salud. El gasto en recreación es similar al de otros clústeres, pero dada la carga sanitaria, podría ser necesario reorientar recursos adicionales hacia estrategias preventivas y educativas.

Nuevamente, los resultados evidenciados en este clúster reconocen la importancia de fortalecer los procesos de focalización de recursos públicos en las políticas públicas porque a pesar de tener el mayor gasto en salud pública tiene una alta tasa de mortalidad por ECV. Por ello, desde la institucionalidad se debe identificar y analizar los principales determinantes de riesgo de enfermedades por ECV en los municipios y empezar a intervenir dichos factores. En efecto, los resultados de este clúster reconocen que la importancia de realizar procesos continuos de seguimiento y evaluación a las políticas públicas porque los impactos esperados de las inversiones en salud no presentan el comportamiento deseado.

7. VALIDACIÓN DE RESULTADOS

En el presente capítulo se realiza una validación cualitativa de los resultados obtenidos en el capítulo 6. El objetivo de la validación cualitativa es indagar entre expertos en políticas públicas acerca de la veracidad de la información obtenida en los clústeres de acuerdo con las realidades que se presentan en los territorios de la zona pacífica colombiana.

Para el proceso de validación cualitativa con expertos, se presentó los resultados obtenidos mediante el algoritmo de K-means y la reducción de dimensionalidad mediante T-SNE. Se decide presentar este modelo a los expertos, debido a que la valoración de los resultados, evidenciaron que dicho modelo era el que presentaba los resultados más consistentes en relación con las seis metodologías de clusterización que se utilizaron.

El proceso de validación cualitativa permite enriquecer el trabajo de grado porque permite articular la ciencia de datos con la experticia de expertos en políticas públicas quienes pueden utilizar los datos encontrados como argumentos que puedan justificar el cambio o transformación de metodologías en la asignación de recursos públicos, diseño de programas y estrategias de políticas públicas y aumentar la efectividad, incidencia e impacto de las políticas públicas en el sector salud.

7.1. Investigaciones previas sobre ECV y determinantes sociales de la salud

En relación con los resultados encontrados en el presente estudio se debe resaltar que los hallazgos evidencian dos tendencias claras. Por un lado, los municipios con mayores necesidades o desigualdades socioeconómicas suelen tener tasas de mortalidad por ECV elevadas, reconociendo que los determinantes sociales y económicos se consolidan como un factor explicativo para promover políticas de promoción y prevención en salud. Por otro lado, algunos municipios con altos niveles de desarrollo socioeconómico también tienen tasas elevadas de mortalidad por ECV, este comportamiento, permite inferir la importancia de indagar los factores que explican dicha dinámica para realizar ajustes y mejorar la focalización de los programas de salud pública.

Frente a los resultados presentados aquí, es importante reconocer algunas limitaciones del estudio en relación a la temporalidad de los datos. En primer lugar, es importante reconocer que algunas acciones de política pública pueden implicar un espacio de ejecución temporal mayor para medir impactos positivos, específicamente al analizar de hábitos, tendencias y tratamientos médicos, sin embargo, al momento de realizar el estudio las bases de datos sobre estas temáticas siguen siendo de difícil hallazgo. No obstante, estos estudios pueden generar un mayor interés para fortalecer las bases de datos en dicha materia. Al mismo tiempo, se debe reconocer que el estudio no cuenta con datos relacionados con hábitos de vida saludables debido a la ausencia de los mismos a nivel municipal (solo se encuentran a nivel departamental), lo cual se considera como un factor esencial para la explicación y desarrollo de muchas enfermedades.

A pesar de las limitaciones del estudio, se debe reconocer que los resultados aquí expuestos se pueden constituir como un primer acercamiento a la comprensión de los determinantes

socioeconómicos en materia de mortalidad por ECV en la región pacífica colombiana. A partir de ello, se puede fomentar otras investigaciones que aborden periodos temporales más amplios y apoyar en los procesos de toma de decisiones en los agentes institucionales.

Bajo dichas consideraciones, el estudio indagó otros estudios similares para identificar sus principales hallazgos y su relación con los resultados de la presente investigación. En primer lugar, se puede mencionar el estudio realizado por Pérez y Achar [60] donde se realizó un análisis de correlación de indicadores socioeconómicos con la mortalidad por ECV en la región pacífica colombiana durante el periodo 2002-2015. A través de un proceso de modelación jerárquica Bayesiana, el autor evidenció que la mortalidad por ECV tendía a presentar un incremento en municipios pequeños, con mayor índice de NBI, con un alto indicador en la pobreza multidimensional y en las regiones con mayor déficit de vivienda. Así mismo, el estudio resaltó que otras variables como la cobertura en salud, el índice de desarrollo y la tasa de alfabetismo era más frecuentes en municipios con tasas de mortalidad por ECV más elevadas, aunque en estas tres variables no obtuvieron una relación estadísticamente significativa.

En el mismo orden de ideas, Pérez y Achar [60] argumentan que a nivel geográfico las áreas con mayor mortalidad se ubicaron en los departamentos de Chocó, Nariño y sur del Cauca quienes se caracterizan por tener un desarrollo menor en relación con el departamento del Valle. Paradójicamente, el estudio encontró que en grandes centros poblados la mortalidad era relativamente alta, aunque dicho comportamiento puede estar explicado por la cantidad de individuos que habitan en las zonas urbanas.

Los resultados del estudio expuesto previamente, validan los resultados encontrados en la presente investigación donde la clusterización evidenció que los municipios con alta tasas de necesidades básicas insatisfechas tienden a presentar un mayor indicador de mortalidad. Similarmente, algunos municipios con un nivel de desarrollo socioeconómico alto también presentan altas tasas de mortalidad por ECV, lo cual implica, que los procesos de prevención y promoción en la salud identifiquen qué otros factores pueden estar explicando dicho comportamiento y adoptar acciones específicas para disminuir la mortalidad por estas patologías.

En relación con el comportamiento particular de áreas de alta desarrollo y una alta mortalidad por ECV, algunos estudios internacionales han evidenciado que el aumento en las condiciones socioeconómicas puede incrementar el riesgo por ECV, debido a que una mayor prosperidad en zonas previamente empobrecidas puede generalizar comportamientos poco saludables que son inicialmente adoptados por un alto número de población [61, 62]. En efecto, algunos estudios identificados en Colombia evidencian que los centros urbanos más grandes tienen estilos de vida que se caracterizan por un mayor consumo de hidratos de carbono, azúcares refinados, sedentarismo y tabaquismo que pueden incrementar el riesgo a padecer ECV [63].

Si bien es cierto, un cuerpo de la literatura establece que un alto nivel de desarrollo puede estar relacionado con una mayor probabilidad de mortalidad por ECV, existen otros estudios donde se contempla que algunas tasas de mortalidad se pueden ver afectadas porque los pacientes con cuadros clínicos más severos se remiten a otros centros con mayores recursos

[64]. Esta explicación se puede aplicar al contexto colombiano donde los centros urbanos y capitales de departamentos tienen la infraestructura de salud más amplia y con mejor capacidad tecnológica para atender una gran diversidad de patologías y enfermedades. Frente a este fenómeno, el informe de la RAP Pacífico en 2020 detalló que las capitales de los departamentos del pacífico colombiano aglomeran más del 50% de centros médicos, IPS y hospitales de alta complejidad [65].

La concentración de la infraestructura médica de alta complejidad en las capitales de los departamentos o en centros de desarrollo intermedio de cada región puede explicar en parte el por qué algunos municipios con un alto nivel de desarrollo económico presentan un alta de mortalidad por ECV al considerar que algunos casos se pueden trasladar hasta centros de atención de mayor complejidad.

Por otro lado, otra investigación que ha obtenido resultados similares a los del presente estudio es el realizado por Romero [66] quien analiza las causas socioeconómicas de la mortalidad en la región pacífica colombiana. Los resultados del estudio evidencian que las tasas de mortalidad de esta región del país son superiores al nivel nacional, a raíz de ello, el principal reto de las políticas en salud pública es apostar por modelos de intervención independientemente del nivel de ingreso por habitante y desarrollar una estrategia que permita mejorar la calidad en el agua para consumo humano y acceso igualitario a los tratamientos médicos.

A pesar de que el estudio de Romero [66] no analiza específicamente la mortalidad por ECV sí reconoce que existe una correlación estadística entre el menor nivel de desarrollo económico y PIB per cápita bajo en la mortalidad por enfermedades no transmisibles (donde se clasifican las ECV), lo cual invita a la necesidad de fortalecer esquemas de prevención en la región y mejorar las estrategias que promueven el bienestar de la población que pueden generar un incentivo a la disminución de la tasa de mortalidad de algunas enfermedades como las relacionadas con el sistema circulatorio.

Los resultados aquí mencionados también se pueden comparar con otros estudios de mortalidad por ECV y condiciones sociodemográficas que se han realizado en otras regiones del país. Por ejemplo, la investigación realizada por Parra y otros [11] analizó algunas causas de mortalidad por ECV, como las enfermedades cerebrovasculares y las enfermedades hipertensivas, donde encontraron que los municipios más desarrollados del eje cafetero colombiano tuvieron la mortalidad más alta por enfermedades cerebrovasculares, pero la mortalidad más baja por enfermedades hipertensivas.

En contraste, el clúster compuesto por los municipios con las condiciones socioeconómicas menos favorecidas (PIB per cápita más bajo y NBI más alto) presentaron las tasas de mortalidad por enfermedad isquémica e hipertensivas más altas de la región cafetera. De acuerdo a los resultados obtenidos por las investigadoras se reconoce que las condiciones socioeconómicas pueden influir en el incremento de mortalidad de algunas patologías cardíacas (como las enfermedades isquémicas e hipertensivas); mientras que para otras enfermedades la relación no es tan clara (enfermedades cerebrovasculares).

Un resultado importante de la investigación es la evidencia que los municipios con mayor NBI se encuentran en cercanía al departamento del Chocó y Oriente de Caldas lo cual evidencia anillos de desigualdad en municipios fronterizos. Más aún, en estos municipios el NBI no ha tenido un avance desde el 2005 y dichas entidades territoriales hacen parte del clúster con la mortalidad más alta de enfermedades isquémicas e hipertensivas [11]. Uno de los aspectos más preocupantes del estudio es que se evidencia que los avances en el NBI por municipio en los últimos años han sido mínimos, situación que también es argumentado por Aponte y otros [67]. El estancamiento del NBI es un elemento preocupante al considerar que esta variable se considera como uno de los elementos más explicativos al momento de analizar la mortalidad por ECV, en efecto, los estudios mencionados previamente sostienen que el país todavía presenta retos muy importantes de desigualdad y brechas económicas al interior del territorio nacional, siendo algunos de los municipios de la costa pacífica y atlántica colombiana los que tienen el NBI más alto [67].

A pesar de las limitaciones estadísticas encontradas en dicho estudio, las autoras reconocen que las evidencias estadísticas se constituyen en una base fuerte para sugerir que las desigualdades socioeconómicas en la mortalidad tienen relación con las condiciones regionales y del territorio, lo cual es fundamental para la formulación de políticas públicas y apoyar el proceso de toma de decisiones en agentes e instituciones con un enfoque multinivel e intersectorial.

Otro estudio que se puede mencionar en relación con las desigualdades sociales en salud y la mortalidad es el realizado por Asprilla y Montenegro [68] donde se identificó que los departamentos del pacífico que tuvieron el peor desempeño del NBI presentaron las mayores tasas de mortalidad en patologías como la mortalidad materna, infantil y la esperanza de vida. Efectivamente, el estudio encontró que de los departamentos del Chocó y el Cauca tuvieron el peor desempeño tanto en el NBI como los resultados en salud; en contraste, el Valle del Cauca tuvo el mejor desempeño en dichos indicadores, pero fue el territorio más heterogéneo en algunas causas de mortalidad.

Igualmente, se puede mencionar el estudio realizado por Palacios [69] quien analizó las condiciones socioeconómicas y su posible relación con las tasas de mortalidad en el departamento del Chocó. En su estudio se encontró que los municipios más alejados de la capital del departamento fueron los que presentaron las tasas de mortalidad más alta del departamento (Unguía y alto Baudó). En general, las características socioeconómicas de estos municipios se caracterizan por tener un bajo desarrollo en la infraestructura médica y terrestre y un mayor indicador de necesidades básicas insatisfechas; de forma concreta, el estudio encontró que estudios con mayor nivel de analfabetismo y NBI se presentaron tasas de mortalidad más alta en las enfermedades hipertensivas, isquémica del corazón y enfermedades cerebro vasculares.

De forma general, los estudios aquí analizados establecen que las condiciones socioeconómicas suelen tener un alto nivel explicativo al momento de analizar la mortalidad por ECV. En efecto, una de las variables analizadas en cada uno de los estudios es el NBI donde se encuentra que los municipios de la zona pacífica donde hay un alto indicador se puede presentar una mayor mortalidad por ECV; sin embargo, también se debe resaltar que algunas regiones o municipios que presentan un alto nivel de desarrollo también se

caracterizan por tener una mortalidad elevada. En estos municipios, la literatura evidencia que el factor explicativo no se genera por razones netamente económicas, sino también por factores como hábitos saludables, incluyendo alimentación y sedentarismo, y en alguna medida porque los centros poblacionales más grandes presentan una mayor concentración de personas con estas enfermedades por poseer centros clínicos de mayor complejidad.

Los resultados aquí evidenciados resaltan la importancia de contar con datos que permitan mejorar la comprensión del fenómeno sobre mortalidad por ECV y ajustar las decisiones adoptadas en políticas públicas de salud. Los resultados presentados en el estudio pueden aportar en el desarrollo de programas de prevención y promoción en salud, sin embargo, es importante reconocer que los hallazgos aquí presentados también presentan algunas limitaciones relacionadas con la capacidad de contar con una serie de datos más amplia para identificar el impacto de las políticas públicas en las condiciones de salud, o la ausencia de datos relacionados con los hábitos de vida saludables como alimentación, deporte, descanso, salud mental, entre otros que no se encuentran desagregados por municipios.

Por otro lado, es importante destacar los desiertos alimentarios como una problemática que si bien no se percibe con claridad también puede ser común en áreas urbanas marginadas y comunidades rurales de bajos ingresos, tanto en países desarrollados como en vías de desarrollo, dicha problemática son los desiertos alimentarios que según Beaulac, Kristjansson y Cummins [70] son áreas caracterizadas por la falta de acceso físico, económico y cultural a alimentos saludables y asequibles. Estas regiones, que suelen ubicarse en comunidades de bajos ingresos, reflejan profundas desigualdades socioeconómicas. La ausencia de supermercados cercanos y las limitaciones económicas obligan a los residentes a depender de alimentos ultraprocesados, lo que compromete su salud y calidad de vida. Este fenómeno no solo afecta a las zonas rurales, sino también a comunidades urbanas marginadas, donde las opciones alimenticias nutritivas están significativamente restringidas.

Desde una perspectiva de salud pública, los desiertos alimentarios son un determinante clave en la prevalencia de enfermedades cardiovasculares, especialmente entre poblaciones vulnerables. Una dieta deficiente, alta en grasas saturadas, sodio y azúcar, y baja en frutas y verduras, contribuye directamente a factores de riesgo como la hipertensión, el colesterol elevado y la obesidad. Esto demuestra cómo las desigualdades socioeconómicas pueden perpetuar disparidades en la salud cardiovascular, resaltando la necesidad de intervenciones que aborden tanto la accesibilidad a alimentos saludables como las condiciones estructurales que mantienen estos desiertos alimentarios.

El artículo de Beaulac, Kristjansson y Cummins [70] aborda el fenómeno de los desiertos alimentarios en los Estados Unidos, caracterizados por la falta de acceso a alimentos saludables en comunidades de bajos ingresos. Esta problemática se debe principalmente a la ausencia de supermercados cercanos, el costo elevado de los alimentos nutritivos y la falta de transporte. Estas barreras estructurales limitan el acceso a frutas, verduras y otros productos esenciales para una dieta equilibrada, lo que afecta de manera desproporcionada a comunidades rurales y urbanas marginadas.

Una de las principales consecuencias de los desiertos alimentarios es el incremento en la prevalencia de enfermedades crónicas relacionadas con la dieta. Los residentes de estas áreas

suelen depender de alimentos ultraprocesados, ricos en grasas saturadas, azúcar y sodio, que son más accesibles y económicos. Esto contribuye a tasas más altas de obesidad, diabetes tipo 2 y enfermedades cardiovasculares. Además, la deficiencia en el consumo de nutrientes esenciales, como fibra, vitaminas y minerales, exacerba el riesgo de hipertensión arterial y niveles elevados de colesterol, aumentando la carga de enfermedades en estas comunidades.

El impacto de la falta de acceso a alimentos saludables no se limita a la salud física, sino que también afecta la salud mental. El estrés asociado con la inseguridad alimentaria, combinado con la preocupación por las enfermedades crónicas, puede desencadenar trastornos como la ansiedad y la depresión. Estos problemas son especialmente prevalentes en familias que enfrentan dificultades para proporcionar una dieta balanceada a sus hijos, lo que perpetúa un ciclo de vulnerabilidad y desigualdad intergeneracional.

El artículo sugiere que la solución a esta problemática requiere un enfoque multifacético. Es esencial no solo mejorar la disponibilidad de alimentos saludables mediante la construcción de supermercados en áreas desfavorecidas, sino también implementar políticas públicas que ofrezcan subsidios para alimentos nutritivos, programas educativos sobre nutrición y acceso a transporte público. Estas medidas podrían reducir la incidencia de enfermedades crónicas, mejorar la calidad de vida y promover la equidad en salud en estas comunidades.

De la misma manera se encuentra el estudio de Bloque y Subramanian [71], donde se afirma que los desiertos alimentarios son áreas con acceso limitado a alimentos saludables, son un factor clave en el desencadenamiento de enfermedades cardiovasculares (ECV), particularmente en comunidades de bajos ingresos. La falta de acceso a alimentos frescos y nutritivos fuerza a los residentes a depender de opciones más económicas y accesibles, como alimentos ultraprocesados, ricos en grasas, azúcares y sodio. Estos patrones alimentarios están estrechamente asociados con el desarrollo de enfermedades como la hipertensión, la diabetes tipo 2 y, en última instancia, las ECV. La mala alimentación, unida a la falta de ejercicio y otras condiciones de vida adversas, crea un caldo de cultivo perfecto para la aparición de enfermedades crónicas.

El impacto de los desiertos alimentarios no puede analizarse sin considerar las desigualdades sociales que influyen directamente en la salud. Las comunidades marginadas enfrentan barreras adicionales que complican aún más su acceso a alimentos saludables, como el bajo nivel educativo sobre nutrición, la escasez de tiempo para preparar comidas saludables y el estrés constante derivado de la inseguridad económica. Estos factores sociales y económicos profundizan las desigualdades en salud, ya que las personas, en su mayoría, recurren a soluciones rápidas y económicas, como las comidas procesadas, que son perjudiciales para la salud cardiovascular. La evidencia muestra que la falta de conocimiento sobre una alimentación balanceada y el estrés crónico juegan un papel crucial en la adopción de hábitos poco saludables que aumentan el riesgo de ECV.

Además, aunque el aumento de supermercados en áreas de bajos recursos podría parecer una solución, los estudios sugieren que simplemente mejorar el acceso a alimentos saludables no resuelve el problema de fondo. Las políticas públicas deben ir más allá de la construcción de establecimientos comerciales y centrarse en cambios estructurales que aborden las causas subyacentes de la mala alimentación. Programas de educación nutricional y estrategias que

incentiven el consumo de alimentos saludables, como el uso de beneficios en programas de subsidios alimentarios, son esenciales para cambiar los patrones alimenticios. Estos programas no solo deben centrarse en el acceso a alimentos, sino también en empoderar a las comunidades con los conocimientos necesarios para tomar decisiones informadas que favorezcan la salud cardiovascular.

Por último, la combinación de desigualdades sociales y la falta de acceso a alimentos saludables perpetúa un ciclo de mala alimentación y enfermedades crónicas. Para combatir el desencadenamiento de ECV en estas poblaciones, es crucial adoptar un enfoque integral que considere tanto las soluciones inmediatas, como el acceso a alimentos nutritivos, como las intervenciones a largo plazo en educación y políticas públicas. La integración de estos enfoques, con énfasis en la equidad en salud, puede ser fundamental para reducir las disparidades en salud y prevenir enfermedades cardiovasculares, mejorando así la calidad de vida de las comunidades más vulnerables.

Desde una perspectiva general se puede afirmar que, a pesar de las limitaciones encontradas en el estudio, se debe mencionar que los resultados aquí presentados implican un primer acercamiento a estudios sobre desigualdad social y mortalidad por ECV en la región pacífica colombiana. A raíz de ello, es importante que el país continúe investigando sobre estas problemáticas, situación que adquiere especial atención si se consideran las tendencias de crecimiento de prevalencia de las enfermedades por ECV a nivel global y el envejecimiento de la población colombiana. En consecuencia, el análisis y comprensión del fenómeno hará posible que las entidades nacionales, departamentales y locales puedan desarrollar estrategias de multinivel e intersectoriales con un enfoque territorial que permitan tener un mayor alcance, efectividad, impactos y sostenibilidad en los recursos públicos implementados en las políticas de salud.

7.2. Validación de resultados con expertos en políticas públicas

El proceso de validación de los resultados con expertos en políticas públicas se realizó con 6 profesionales del área de la salud y economía que han estudiado temáticas relacionadas con: (a) el ciclo de vida de las políticas públicas: diseño, implementación, monitoreo, seguimiento y evaluación; (b) profesionales de la salud que han investigado sobre problemáticas de salud y su articulación con las políticas públicas; y (c) un profesional que se ha desempeñado en cargos directivos en materia de salud pública.

Del total de los 6 expertos consultados, 2 de ellos pertenecen a la academia y han realizado sus investigaciones sobre el ciclo de vida de la política pública; 3 de ellos son investigadores o funcionarios públicos en instituciones privadas y públicas donde han analizado el impacto de las políticas públicas para la prevención y promoción de la salud; y uno de ellos se ha desempeñado en cargos directivos y secretarías de salud. Del total de los expertos consultados, el 66.66% han trabajado en políticas públicas en los departamentos de Nariño, Cauca y Putumayo y el 33.33% han desarrollado su vida profesional en investigaciones en el departamento del Valle del Cauca.

El análisis de validación de expertos permitió que se desarrollara una encuesta que abarcara elementos administrativos relacionados con las políticas públicas y con los resultados obtenidos en el presente estudio. Sus aportes, basados en su experiencia y trayectoria, abordan diversos desafíos relacionados con las desigualdades en salud y las ECV. Estas reflexiones enfatizan la importancia de la asignación eficiente de recursos, la necesidad de fortalecer la coordinación interinstitucional y la implementación de políticas públicas basadas en la prevención, la equidad y la sostenibilidad, con el objetivo de mejorar los indicadores de salud en los territorios más vulnerables y garantizar una señal de impacto.

Se consultó a los entrevistados sobre cómo se determina la distribución de recursos en salud pública a los municipios más vulnerables o con mayores necesidades, y cuáles son los factores clave considerados en este proceso. En sus respuestas, explicaron que este reparto se fundamenta en un enfoque basado en la equidad y la priorización de necesidades, utilizando indicadores clave que reflejan las condiciones socioeconómicas y de salud de la población. Mencionaron que factores como la prevalencia de enfermedades, los índices de pobreza, la densidad poblacional y la infraestructura sanitaria existente son determinantes esenciales para identificar las áreas que requieren mayor atención.

La toma de decisiones en este proceso también depende de la recopilación y análisis de datos confiables que permitan establecer un panorama claro de las necesidades específicas de cada región. Herramientas como estudios epidemiológicos, estadísticas demográficas e informes de accesibilidad a los servicios de salud son esenciales para orientar las políticas públicas.

Si bien es cierto, el análisis contextual forma parte de los procesos de asignación de recursos, los expertos determinan que dicho proceso no es altamente flexible porque depende del Sistema General de Participaciones (SGP) y de la categoría de los departamentos y municipios. Generalmente, los municipios con mayores vulnerabilidades tienen menores recursos de transferencia del SGP y menor capacidad de inversión en áreas claves de salud a partir de los recursos propios de cada entidad territorial; de ahí, que se explique porque algunos municipios, especialmente de la costa pacífica, tengan un alto NBI en materia de salud, una baja infraestructura hospitalaria y grandes retos en atención sanitaria.

En dicho escenario, los expertos consultados sostienen que una de las debilidades en el diseño de las políticas públicas se centra en el proceso de asignación y determinación de recursos para la financiación de acciones y estrategias contenidas en los documentos de política. Por ello, desde la academia se ha recomendado que el proceso de diseño y formulación abarque un estudio presupuestal y se alineen a estrategias preexistentes. De hecho, los estudios de seguimiento y monitoreo han descubierto que muchas políticas públicas tienen un bajo impacto porque existe replicación de acciones con estrategias existentes, metas ambiciosas con nulos recursos, ausencia en el diseño de indicadores claros y medibles, baja preparación de los profesionales a cargo de su implementación y, en algunos casos, desarticulación en los comités intersectoriales de política pública.

Por consiguiente, los expertos resaltan la importancia de integrar enfoques interdisciplinarios y colaborativos en la asignación de recursos. Esto implica que las decisiones no solo deben basarse en criterios técnicos, sino también en políticas públicas sólidas que incorporan aspectos sociales, culturales y económicos. Al combinar estas perspectivas, se puede lograr

una gestión de recursos más inclusiva y sostenible, reduciendo las brechas en el acceso a servicios de salud y fortaleciendo los sistemas de atención en los municipios más vulnerables.

En relación con los retos para abordar las desigualdades socioeconómicas para analizar la prevalencia de las ECV en la región pacífica colombiana se destacó que enfrentar estas desigualdades implica superar diversos desafíos, como garantizar un acceso equitativo a servicios de salud preventivos y de tratamiento. Señalaron que las condiciones económicas adversas, la falta de infraestructura sanitaria en regiones vulnerables y las barreras culturales limitan la implementación efectiva de políticas públicas, perpetuando las desigualdades en los resultados de salud. Además, explicaron que estos factores dificultan tanto la adopción de hábitos saludables como el acceso oportuno a intervenciones médicas, lo que contribuye a una mayor incidencia de estas enfermedades en comunidades desfavorecidas.

Otro desafío significativo es la necesidad de adaptar las políticas públicas al contexto particular de cada región. Las estrategias diseñadas de manera centralizada muchas veces no consideran las dinámicas locales, lo que afecta su relevancia y eficacia. La falta de recursos humanos capacitados y de financiamiento sostenible también limita la capacidad de los sistemas de salud para responder a las necesidades específicas de las comunidades más afectadas por las enfermedades cardiovasculares. Además, la ausencia de datos actualizados y desagregados dificulta la identificación precisa de las brechas y la planificación de intervenciones efectivas.

Así, los expertos reconocen que los procesos de asignación inicial para las políticas públicas también deben ir acompañados del monitoreo y seguimiento en la fase de implementación de las políticas. Frente a ello, reconocen que las limitaciones que han tenido las políticas públicas frente a la disminución de las desigualdades se ha generado por una baja gestión del conocimiento en el ciclo de la política pública, por ello, estudios geospaciales como el presente pueden aportar en los procesos de planificación estratégica porque la institucionalidad adoptaría decisiones basados en evidencia. Por ejemplo, los expertos reconocen que en algunas ciudades como Bogotá ya se está realizando un seguimiento de políticas públicas a través de estadísticas espaciales que han colaborado en gran medida a fortalecer la toma de decisiones en políticas públicas de salud considerando las particularidades de las localidades y ajustar sus estrategias de intervención según factores de riesgos encontrados en cada comuna.

En consecuencia, reconocen que la presencia de datos actualizados y desagregados es fundamental para asegurar el éxito de las intervenciones en políticas públicas de salud. Desafortunadamente, la mayoría de las entidades municipales no tienen la capacidad financiera, técnica y humana para realizar un proceso de seguimiento y planificación a profundidad, sin embargo, si se recomienda que en materia de abordaje de políticas en salud se realice un monitoreo más estricto al reconocer que los determinantes en salud no solo son clínicos sino también sociales.

Al aterrizar los retos y manejo de desigualdades como factor explicativo en las ECV, los expertos consultados argumentan que dicho proceso requiere un enfoque basado en datos y estrategias adaptadas a las realidades específicas de las comunidades afectadas. Señalaron que uno de los mayores aprendizajes ha sido comprender que las desigualdades en salud no

solo están relacionadas con factores clínicos, sino también con determinantes sociales como la pobreza, el nivel educativo y el acceso limitado a servicios de salud. Así mismo, reconocen las limitaciones de estudios estadísticos por regiones y con datos desagregados en el área de ECV, por ejemplo, la mayor parte de la investigación se han centrado en determinar la relación existente en barreras de acceso a programas de prevención y promoción, atención oportuna y NBI con enfermedades como el EDA (Enfermedad Diarreica Aguda) o el IRA (Infección Respiratoria Aguda).

Así, los expertos reconocen que es indispensable incrementar los estudios sobre las ECV porque las investigaciones sobre este grupo de patologías no siempre han logrado encontrar correlaciones positivas porque este tipo de enfermedades tienen diferentes ventanas de oportunidad en su estado morbo y la caracterización de la población incluyen más variables que solamente las condiciones económicas, sino que también se valoran enfermedades familiares, hábitos de vida saludable y cuadros genéticos. Así, en dicho contexto, es esencial incrementar el conocimiento sobre cómo las diferentes condiciones socioeconómicas de las regiones, en este caso de la zona pacífica, pueden influir en las condiciones de salud de los individuos. Esta necesidad se agudiza al reconocer, que el tratamiento de la ECV requiere combatir la desinformación y las barreras culturales que impiden a muchas personas buscar atención médica temprana. Estos desafíos exigen un equilibrio entre medidas centralizadas y acciones locales que respondan a las particularidades de cada región.

Los expertos coinciden en que el manejo eficaz de estas desigualdades depende en gran medida de la capacidad de los sistemas de salud para trabajar de manera integrada y colaborativa. Esto incluye alianzas intersectoriales y la participación activa de las comunidades en el diseño e implementación de las intervenciones. La experiencia muestra que una aproximación inclusiva, basada en la escucha activa y la cooperación entre actores claves es esencial para reducir las brechas en el acceso a la atención cardiovascular y mejorar los resultados de salud a largo plazo.

En el análisis de los "Municipios Resilientes con Desafíos de Salud", identificados como aquellos donde a pesar de una alta inversión en salud persisten elevados niveles de mortalidad por ECV, los expertos sostuvieron que esta paradoja refleja una desconexión entre la inversión financiera y la efectividad en la implementación de políticas y programas. Explicaron que, en muchas ocasiones, los recursos destinados a la salud no se traducen en intervenciones adaptadas al contexto local, lo que limita significativamente su impacto en la mejora de los indicadores de salud.

Otro factor clave es la persistencia de determinantes sociales de la salud que no se abordan de manera integral. Aspectos como la desigualdad económica, los bajos niveles educativos, las condiciones laborales precarias y el acceso limitado a una alimentación saludable continúan afectando a las comunidades, contrarrestando los beneficios de las inversiones en infraestructura o servicios médicos. Además, la fragmentación de los sistemas de salud y la falta de coordinación intersectorial dificultan la implementación de estrategias preventivas eficaces y sostenibles para reducir la incidencia de ECV.

Los expertos enfatizan que abordar este desafío requiere un cambio en el enfoque de las políticas de salud, priorizando no solo el gasto, sino también la calidad, pertinencia y

sostenibilidad de las intervenciones. Es fundamental fortalecer las capacidades locales para identificar y resolver problemas específicos, promover la educación en salud y fomentar la participación comunitaria en la planificación y evaluación de programas. Solo a través de un enfoque más holístico e inclusivo se puede superar esta paradoja y mejorar de manera significativa la salud cardiovascular en regiones resilientes, pero aún vulnerables.

En relación con el clúster "Municipios Vulnerables con Riesgo Latente", caracterizados por bajos niveles de mortalidad por ECV pero enfrentando grandes desafíos socioeconómicos, los expertos consultados sostuvieron que los resultados son consistentes con investigaciones previas porque la mayor parte de los municipios clasificados en dicha categoría se encuentran en el departamento del Chocó y la costa pacífica. Este resultado es consistente con la realidad de estos municipios que se caracterizan por la ausencia de centros de salud y hospitales de alta complejidad que puedan atender algunos cuadros clínicos graves, mayores dificultades para acceder a programas de prevención y promoción y un alto nivel de NBI. A pesar de las debilidades en materia económica, este tipo de municipios se caracterizan por tener unos hábitos alimentarios y de actividad física más positivos que en las ciudades, situación que puede llegar a explicar la baja tasa de mortalidad por ECV, ya que, este tipo de enfermedades no solo se explicación por condiciones sociales y económicos, sino también por hábitos alimenticios y de ejercicio.

En el mismo clúster, los expertos sostienen que los resultados del estudio reafirman los planes de acción nacionales donde se reconoce la necesidad de una intervención temprana en municipios del Departamento del Chocó y la costa pacífica colombiana para evitar que las dificultades actuales se traduzcan en problemas de salud más graves a largo plazo, incluidos incrementos significativos en las ECV. Destacaron que factores como la pobreza, el limitado acceso a servicios de salud, la falta de infraestructura sanitaria y la deficiente educación en salud son elementos que, si no se abordan oportunamente, pueden actuar como catalizadores de problemas crónicos, especialmente frente a cambios demográficos o de estilo de vida que puedan surgir en estas comunidades.

Para estos municipios, los expertos consideran prioritario implementar políticas públicas que se enfoquen en la prevención y promoción de la salud. Iniciativas como la educación en hábitos saludables, la mejora de la accesibilidad a alimentos nutritivos y la creación de entornos que fomenten la actividad física son fundamentales para mitigar los riesgos a largo plazo. También es clave fortalecer los sistemas de atención primaria de salud, asegurando que las comunidades tengan acceso a servicios preventivos, detección temprana y control de factores de riesgo como la hipertensión y la diabetes, que están estrechamente relacionados con las ECV.

Igualmente, para los municipios agregados en este clúster es esencial adoptar un enfoque multisectorial que aborde las causas subyacentes de las desigualdades socioeconómicas. Esto incluye programas de desarrollo económico, mejora de la infraestructura básica y promoción de la equidad educativa, así como el fortalecimiento de programas de promoción y prevención para evitar que las transiciones demográficas y nuevos estilos de vida puedan agudizar la morbilidad de ECV.

En relación con el clúster "Municipios de Contrastes Socioeconómicos" donde la prosperidad económica contrasta con una baja inversión en salud y educación, y qué tipo de intervenciones podrían equilibrar esta dicotomía. Los expertos sostuvieron que este fenómeno refleja un patrón recurrente en ciertas regiones, donde la prosperidad económica no se traduce en una inversión proporcional en servicios básicos. Explicaron que esta paradoja está asociada a una falta de planificación estratégica en el uso de los recursos y a la priorización de sectores económicos sobre las necesidades sociales. También destacaron que la baja inversión en salud y educación en estos municipios impacta directamente en la calidad de vida de la población, perpetuando desigualdades estructurales.

Para equilibrar esta dicotomía, los expertos consideran fundamental promover políticas públicas que fortalezcan la gobernanza local y aseguren que los ingresos generados se reinviertan en infraestructura y servicios esenciales. Esto incluye el desarrollo de sistemas de salud accesibles y de calidad, así como la implementación de programas educativos que potencializan el capital humano en estas comunidades. También resulta clave fomentar la transparencia en la administración de recursos, garantizando que las decisiones sobre inversión respondan a las necesidades reales de la población.

Para estos municipios, los expertos sostienen que adquiere mayor importancia los procesos de monitoreo y seguimiento de las políticas públicas porque se evidencia una debilidad en la focalización de los recursos públicos. En efecto, los profesionales argumentan que desde los nuevos lineamientos establecidos por el Departamento Nacional de Planeación se ha reconocido la debilidad en el manejo presupuestal de algunas políticas públicas en todo el país, situación que se puede corroborar en este estudio, al reconocer que no siempre la prosperidad económica significa que las entidades municipales están adoptando estrategias de alto impacto para las comunidades.

Así mismo, los expertos han reconocido que la ausencia de los procesos de monitoreo, seguimiento y evaluación de impacto de las políticas públicas ha generado que desde los procesos de planificación se inviertan en ejes y líneas de acción que no corresponden a la realidad del territorio. En muchos otros casos, la asignación de recursos no depende, exclusivamente, de la entidad territorial sino de los recursos del SGP generando que algunas líneas de acción queden absolutas o sin acción. Más aún, cabe reconocer que los problemas de gobernabilidad, gestión financiera y corrupción también afecta la asignación de recursos, lo cual puede explicar, que en algunos municipios se evidencie prosperidad económica pero baja inversión en salud.

Además, se recomienda implementar enfoques intersectoriales que integren salud, educación y desarrollo económico. Iniciativas como alianzas público-privadas, incentivos para proyectos de impacto social y mecanismos de participación comunitaria pueden contribuir a cerrar las brechas entre prosperidad económica y bienestar social. En última instancia, este enfoque equilibrado no solo mejora las condiciones de vida, sino que también asegura un desarrollo sostenible.

Igualmente, se propone los procesos de veeduría ciudadana y estrategias de políticas de salud colectivas donde la comunidad adquiere un papel más activa en la gestión de recursos, estrategias y programa. Como ejemplo específico, los expertos resaltaron la política pública

de salud colectiva del Municipio de Pasto que se constituyó en un referente de gestión en salud a nivel nacional y latinoamericano por abordar y conceptualizar el concepto de salud desde una concepción multidisciplinaria, participación ciudadana y gestión inter e intrainstitucional de programas de salud. La apuesta de dicha política, que entró en vigencia en el año 2023, es la disminución de brechas de desigualdad en salud, disminución de indicadores de mortalidad (incluyendo las ECV), gestión del conocimiento, transparencia y participación y movilización ciudadana como factores transversales para el mejoramiento en las condiciones de salud.

En relación con los municipios catalogados como de "Alta Carga Sanitaria y Desigualdades Latentes", los entrevistados señalaron que la persistencia de altos índices de enfermedades cardiovasculares (ECV) y mortalidad, a pesar de una considerable inversión en salud, se debe a múltiples factores estructurales y sistémicos. Destacaron que uno de los principales problemas es la ineficiencia en la asignación y gestión de los recursos, los cuales, en muchos casos, no están dirigidos a intervenciones específicas y contextualizadas para las necesidades locales. Además, mencionaron que la fragmentación de los sistemas de salud, junto con la falta de integración entre niveles de atención y sectores sociales, limita significativamente el impacto de las inversiones realizadas.

Otro factor crítico es que las desigualdades sociales y económicas subyacentes no suelen abordarse de manera integral, perpetuando las condiciones que generan las altas tasas de ECV y mortalidad. Aspectos como la falta de educación en salud, la precariedad laboral y el acceso desigual a recursos básicos como agua potable y alimentos saludables son determinantes sociales que no pueden ser resueltos únicamente mediante inversión en servicios médicos. Esto evidencia la necesidad de políticas más inclusivas y transversales que consideren la salud como un resultado de múltiples factores interrelacionados.

Desde la perspectiva de los expertos, superar estas limitaciones requiere un cambio en la estrategia de inversión hacia enfoques más preventivos y comunitarios. La promoción de estilos de vida saludables, el fortalecimiento de la atención primaria y la participación activa de la comunidad en el diseño e implementación de políticas son clave para cerrar las brechas existentes. Además, es esencial mejorar los sistemas de monitoreo y evaluación para garantizar que las inversiones tengan un impacto real en la reducción de las desigualdades, promoviendo así resultados más equitativos y sostenibles en salud.

Aunado a lo anterior, se indagó a los expertos que, de acuerdo a los resultados del estudio, cómo se podría impactar positivamente en la reducción de mortalidad por ECV. En esta medida, mencionaron que, los municipios clasificados en este clúster deben priorizar estrategias integrales que combinen prevención, atención médica accesible y abordaje de los determinantes sociales de la salud. Los expertos coinciden en que es esencial fortalecer los programas de promoción de estilos de vida saludables, enfocados en reducir factores de riesgo como el tabaquismo, la mala alimentación y el sedentarismo. Estas iniciativas deben complementarse con campañas de educación que sensibilicen a las comunidades sobre la importancia de la prevención y el autocuidado en salud.

Además, se recomienda priorizar el fortalecimiento de la atención primaria de salud, asegurando que los servicios médicos estén disponibles y sean de calidad en las regiones más vulnerables. Esto incluye la capacitación de personal sanitario, la mejora de la infraestructura y el acceso a tecnologías médicas que permitan un diagnóstico temprano y el manejo adecuado de las ECV. También es fundamental garantizar la equidad en la asignación de recursos, de manera que las comunidades con mayores necesidades puedan acceder a la intervención.

Las políticas públicas deben integrar enfoques intersectoriales que aborden las desigualdades socioeconómicas subyacentes. Programas que promueven la seguridad alimentaria, el acceso a empleo digno y la mejora de las condiciones de vida son fundamentales para atacar las causas estructurales que contribuyen a la alta carga sanitaria. Los expertos destacan que solo a través de un enfoque holístico y participativo se puede lograr un impacto significativo en la reducción de las ECV y mejorar la calidad de vida en estos municipios.

Desde un ámbito más clínico, las recomendaciones que ofrecieron los expertos para fortalecer las políticas públicas de salud para reducción de la mortalidad por ECV se centró en que las estrategias de intervención en Colombia no se deben centrar en la fase crónica de la enfermedad, sino en su fase aguda. De forma general, sostienen que los lineamientos de política pública en salud están enfocados en programas durante la fase crónica de la enfermedad (por ejemplo, el programa crónicos de hipertensos o ACV). Sin embargo, en la fase crónica de la enfermedad, la posibilidad terapéutica es baja mientras que en la fase aguda los procesos de prevención permiten hacer un seguimiento al individuo para evitar que la patología se agrave e incrementar favorablemente las condiciones de vida de los individuos al abordar la enfermedad en una fase que no genera tantas consecuencias en salud para el individuo. Más aún, al hablar de focalización el diseño de programas en sus fases agudas permite ahorrar una gran cantidad de recursos económicos porque su costo es mucho más barato que los tratamientos en la fase crónica.

Teniendo en cuenta la trayectoria de los entrevistados, se preguntó la opinión que tenían respecto a los cambios en los resultados de salud cardiovascular dado la implementación de políticas públicas específicas y los factores que han facilitado o dificultado estos cambios. En este sentido, mencionaron que, a lo largo de los años, se han observado cambios significativos en los resultados de salud cardiovascular en ciertas regiones como resultado de la implementación de políticas públicas específicas, aunque los avances han sido desiguales. Entre los factores que han facilitado estos cambios destacan la promoción de programas de prevención que fomentan estilos de vida saludables, el fortalecimiento de la atención primaria y la introducción de tecnologías médicas para el diagnóstico y manejo temprano de las enfermedades cardiovasculares (ECV). Estas iniciativas han demostrado ser particularmente efectivas cuando se combinan con campañas educativas y el apoyo de comunidades bien informadas.

Sin embargo, también persisten desafíos que han dificultado un impacto más amplio y sostenido. La falta de recursos económicos en algunas regiones, junto con la fragmentación de los sistemas de salud, limita la continuidad y alcance de las intervenciones. Además, las desigualdades sociales y económicas, como el acceso desigual a alimentos saludables y la

atención médica, siguen siendo barreras estructurales que afectan los resultados de las políticas públicas. La resistencia al cambio cultural y la desinformación en torno a la prevención de ECV también han obstaculizado los avances en algunas comunidades.

Los expertos subrayan que los cambios sostenibles requieren un enfoque integral que combine la implementación técnica de políticas con la construcción de capacidades locales, el fortalecimiento de alianzas intersectoriales y el monitoreo constante de los resultados. También destacan la importancia de involucrar a las comunidades en el diseño y ejecución de estas políticas, asegurando que sean culturalmente relevantes y ajustadas a las realidades locales. Este enfoque, centrado en la participación y la equidad, es clave para lograr mejoras significativas y duraderas en la salud cardiovascular.

En concreto, los expertos sostuvieron que los cambios estructurales también varían de acuerdo al territorio. Por ejemplo, en municipios con un bajo desarrollo económico se encuentran limitantes de infraestructura, acceso a servicios sanitarios y servicios básicos, alimentación y actitudes comportamentales que frenan el posible impacto de las intervenciones de política pública. En contraste, en municipio con mayor nivel de desarrollo, las limitantes principales están asociadas a comportamientos, creencias y alto nivel de sedentarismo y malas prácticas alimenticias de los individuos más que en el acceso a los servicios sanitarios.

Dichos determinantes sociales, según los expertos, permiten explicar porque las ciudades principales de los cuatro departamentos analizados presentan una alta tasa de mortalidad por ECV a pesar de su desarrollo socioeconómico. Así mismo, otros expertos también consideran que otra potencial razón para el alto número de mortalidad por ECV en dichos territorios se puede generar por fallas en los procesos de recolección de estadísticas en Colombia, al considerar que en muchas ocasiones el registro del deceso de un paciente se realiza por lugar de ocurrencia y no de procedencia. Esta debilidad es una situación que se debe transformar de forma estructural para que los datos y análisis estadísticos estén más ajustados a las realidades de los territorios y planificar sus acciones con estrategias más focalizadas.

De la misma manera, se interrogó a los entrevistados sobre la opinión que tenían respecto al rol de la coordinación interinstitucional (salud, educación y desarrollo social) en la reducción de las desigualdades en enfermedades cardiovasculares. En esta medida, coinciden en que la coordinación interinstitucional es fundamental en el área de salud, ya que, permite abordar de manera integral los determinantes sociales y estructurales que inciden en la prevalencia de estas enfermedades. La colaboración entre los sectores de salud, educación y desarrollo social es clave para diseñar e implementar estrategias efectivas que combinen la promoción de estilos de vida saludables, la prevención de factores de riesgo y la mejora de las condiciones de vida en comunidades vulnerables.

Uno de los mayores retos en esta coordinación es la falta de comunicación y alineación de objetivos entre las instituciones involucradas. Esto se traduce en programas fragmentados que no logran el impacto esperado en las comunidades más afectadas. Para mejorar esta colaboración, es necesario establecer mecanismos formales de planificación conjunta, compartir información y datos relevantes, y fomentar la participación activa de todos los actores clave, incluidos los gobiernos locales, organizaciones comunitarias y el sector

privado. Además, se requiere la asignación de recursos específicos para garantizar la ejecución efectiva de las políticas intersectoriales.

Así mismo, los expertos reconocen que desde el área de salud también se debe fortalecer algunas estrategias y procedimientos que permitan fortalecer el perfil epidemiológico del individuo. Dichas acciones pueden incluir estrategias de estudios genéticos, antecedentes familiares, perfil fisiopatológico y antecedentes de salud. Los estudios más recientes sobre las ECV han demostrado que es posible predecir la posibilidad de desarrollar alguna patología de este grupo de enfermedades a través de una multiplicidad de variables del perfil epidemiológico, sin embargo, en muchos territorios del país no es posible contar con la información que permita realizarlo y trazar una ruta de atención integral a los pacientes. La recolección de estos datos, es según criterio de los expertos, un elemento fundamental para próximas políticas públicas en salud.

Desde la perspectiva de los expertos, fortalecer la colaboración y articulación institucional implica también la implementación de políticas públicas basadas en la evidencia, que integren indicadores compartidos de éxito y permitan medir el impacto colectivo de las intervenciones. La capacitación continua de los actores institucionales en enfoques colaborativos y la promoción de una visión compartida sobre la equidad en salud son elementos esenciales para maximizar los beneficios de la coordinación interinstitucional y reducir de manera significativa las desigualdades en ECV.

Los expertos dieron su opinión respecto a las estrategias o innovaciones que consideran cruciales para reducir las desigualdades en salud, prestando especial atención a las enfermedades cardiovasculares, donde mencionaron que se requiere la implementación de estrategias e innovaciones que combinen enfoques preventivos, tecnológicos y comunitarios. Los expertos coinciden en que una de las claves es fortalecer la atención primaria mediante la expansión de programas de promoción de la salud y prevención de factores de riesgo, como el tabaquismo, la obesidad y la inactividad física. Estas intervenciones deben ser accesibles, culturalmente relevantes y centradas en las necesidades específicas de las comunidades vulnerables.

La incorporación de tecnologías innovadoras también es crucial para cerrar las brechas en salud. Herramientas como la telemedicina, el análisis de big data y las aplicaciones móviles para el monitoreo de la salud permiten mejorar el acceso a servicios médicos, especialmente en áreas rurales o con limitaciones de infraestructura. Estas tecnologías facilitan la detección temprana de factores de riesgo y la gestión personalizada de enfermedades, optimizando los recursos disponibles y ampliando la cobertura de las intervenciones.

Los expertos subrayan la importancia de fomentar la participación comunitaria y la coordinación intersectorial como pilares fundamentales para abordar las desigualdades de manera estructural. Estrategias como la educación en salud, la promoción de entornos saludables y el fortalecimiento de alianzas entre sectores como educación, desarrollo social y transporte son esenciales para atacar las causas subyacentes de las ECV. Este enfoque integral, que combina innovación, prevención y participación activa, no solo mejora los resultados de salud, sino que también promueve la equidad y la sostenibilidad en las comunidades más afectadas.

Al considerar tanto los resultados obtenidos y la validación de los resultados, se exponen las siguientes recomendaciones para cada uno de los clústeres obtenidos.

1. Clúster 3: Municipios resilientes con desafíos de salud: Para los municipios agrupados en esta categoría la principal recomendación se concentra en fomentar hábitos de vida saludable en la población y establecer criterios de medición para medir el impacto de dichos hábitos en la promoción, prevención y mantenimiento de la salud. Así mismo, en este grupo de municipios, se recomienda la revisión de los planes de inversión y planes de acción para determinar los rubros a los cuales se destinan los recursos en salud, recreación y deporte porque son uno del clúster con mayor inversión en dichos rubros.

En materia de condiciones socioeconómicas, se reconoce la importancia de diversificar la economía porque los antecedentes investigativos y la opinión de los expertos reconocen que un bajo desarrollo económico, limita el acceso de las personas a los programas de promoción y prevención.
2. Clúster 2: Municipios vulnerables con riesgo latente: La principal recomendación en los municipios contenidos en este grupo se encaminan al fortalecimiento de programas de atención social debido a los bajos niveles de NBI. Si bien es cierto, en este clúster los índices de mortalidad por ECV son bajos, existe un riesgo latente de la población a ser expuesta a condiciones de sanidad, hábitos alimenticios y acceso a servicios públicos que pueden provocar crisis en salud pública.

Frente a los municipios de este clúster es esencial aumentar la presencia del Estado a través de programas de intervención y estrategias de salud pública. En efecto, el resultado del proyecto aplicado reconoce que la mayoría de los municipios de este clúster están localizados en los departamentos del Chocó y la costa pacífica colombiana que se han caracterizado por la ausencia de programas de prevención en salud, baja cobertura educativa y un alto riesgo de agua no potable (IRCA). Estas áreas son de especial atención para disminuir posibles riesgos en materia de salud pública.
3. Clúster 1: Municipios de contrastes socioeconómicos: En este clúster una de las principales recomendaciones está enfocada a analizar y determinar el proceso mediante el cual se focaliza los recursos en salud, educación, deporte y recreación porque a pesar de tener el valor agregado más alto se evidencia bajas inversiones en dichos sectores. Para ello, es fundamental reconocer cómo se está realizando la focalización de recursos en estas áreas y determinar si es más pertinente destinar algunos recursos adicionales a programas de salud y educación, y menos a programas de saneamiento básico al considerar que el IRCA es el más bajo de todos los clústeres.

De forma general, se recomienda abordar las variables mencionadas previamente con el objetivo de continuar con la tendencia positiva de una baja mortalidad por ECV, dado que una mayor inversión en temas de salud, hábitos saludables y educación pueden disminuir los factores de riesgo a morbilidad y mortalidad por ECV como han expuestos los expertos y los estudios previos.

4. Clúster 0: Municipios con alta carga sanitaria y desigualdades latentes: La principal recomendación de estos municipios se enfoca en fortalecer la focalización de los programas sociales debido a que se evidencia un desarrollo económico moderado pero un comportamiento negativo en indicadores sociales como inversión en recreación, deporte; el mayor nivel de IRCA y un NBI moderado. Este clúster evidencia que los principales retos están en mejorar la focalización de servicios y evaluar el impacto de los programas porque tiene la segunda tasa de mortalidad más alta por ECV, un bajo nivel de NBI y el gasto público en salud más alto lo que puede entenderse como inadecuados procesos de focalización en procesos de promoción y prevención en salud, es decir, no se pueden estar priorizando otros factores de riesgos latentes en dichas regiones y que pueden estar asociados a las enfermedades ECV. En estos clúster los programas de promoción y prevención deberían invertir parte de los recursos de salud en matizar factores de riesgo en la población para comprender factores genéticos, ambientales, sociales y económicos que pueden estar relacionados para la mortalidad por ECV y atender dichos factores de forma prioritaria porque se puede evidenciar que los recursos en salud no están generando un impacto positivo afectando los criterios de eficiencia, efectividad y economía en políticas públicas.

7. DESPLIEGUE DEL TABLERO DE INFORMACIÓN

El proceso de la creación del tablero de información se realizó mediante la plataforma Reflex App que permite su creación de forma gratuita, y posteriormente, brinda la posibilidad de migrar los datos a aplicaciones web con mayor complejidad o de mayor capacidad de datos.

La plataforma para el despliegue de los datos se encuentra localizada en la siguiente página web: <https://enfermedades-cardiovasculares.reflex.run/> donde se encuentra la información y su índice de navegación principal tiene 4 opciones adicionales que son: información técnica, general, tabulación y modelación. En la pestaña de información general, se explica el procedimiento y metodología de análisis de los datos donde se explica al público cómo se realizó el proceso, y un resumen de los 4 clústeres encontrados en el desarrollo del trabajo.

En la página tabulación, se encuentra la tabulación de todos los datos utilizados para el proyecto (ver figura 32). Posteriormente, en la segunda página ya se encuentra la modelación de los clústeres obtenidos en el capítulo 6.

Figura 32

Visualización pestaña tabulación del tablero de información



Departamento	Municipio	Año	Enfermedades CV
CHOCÓ	QUIBDÓ	2016	179
CHOCÓ	ACANDÍ	2016	5
CHOCÓ	ALTO BAUDO	2016	4
CHOCÓ	ATRATO	2016	6
CHOCÓ	BAGADÓ	2016	5
CHOCÓ	BAHÍA SOLANO	2016	6
CHOCÓ	BAJO BAUDÓ	2016	8
CHOCÓ	BOJAYA	2016	2
CHOCÓ	EL CANTÓN DEL SAN PABLO	2016	2
CHOCÓ	CARMEN DEL DARIEN	2016	3

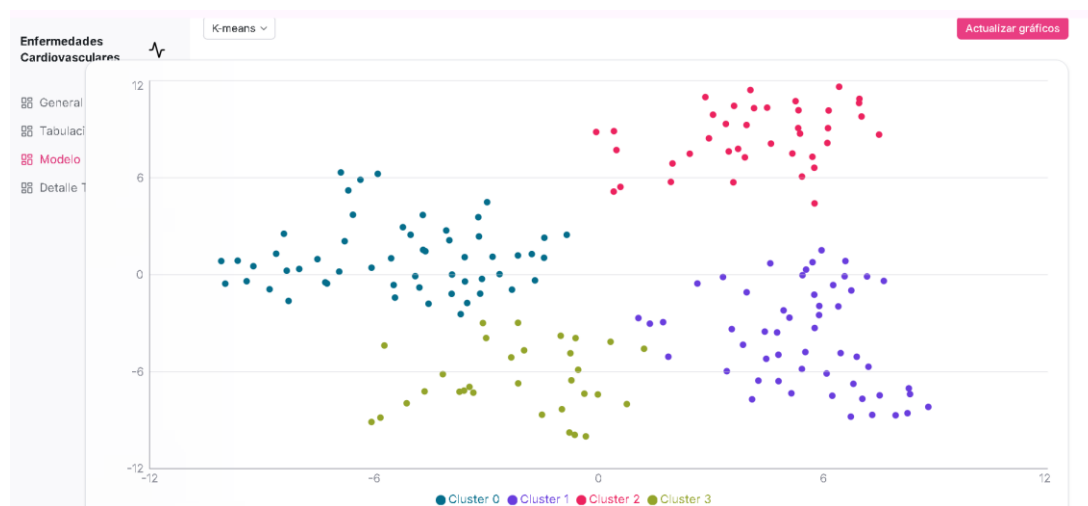
En ese orden de ideas, en la pestaña denominada “Modelo” se puede encontrar los resultados del clúster obtenido en el estudio. A raíz de ello se puede visualizar los resultados de los diferentes clústeres obtenidos en el estudio, al mismo tiempo, se logra visualizar las características en materias de NBI, defunciones, cobertura neta en educación, IRCA y gasto público en salud de cada uno de los clústeres. Adicional a ello, en esta pestaña también se pueden seleccionar y visualizar los resultados obtenidos con las diferentes técnicas de modelación.

La siguiente imagen obtenida del tablero muestra una visualización de datos usando el método de K-means clustering, que agrupa a los datos en diferentes clústeres basados en sus características. En este caso, el gráfico está representando los determinantes socioeconómicos en las enfermedades cardiovasculares y segmenta a la población en

diferentes grupos o clústeres en función de sus similitudes. El gráfico indica cuatro clústeres identificados por colores, cada punto de un clúster es la representación de una observación con características similares respecto a los determinantes socioeconómicos que influyen en las enfermedades cardiovasculares (ver figura 33). En la opción de visualización de los clústeres, el tablero de información también brinda la opción de escoger los diferentes métodos de clusterización utilizados en el estudio.

Figura 33

Visualización pestaña de modelación de los datos método K-means



Por otro lado, se encuentra un gráfico de radar que representa los determinantes socioeconómicos para cada uno de los clústeres identificados en el análisis previo de las enfermedades cardiovasculares (ECV). Cada gráfico muestra la distribución de varias variables en función de su importancia o valor en cada clúster.

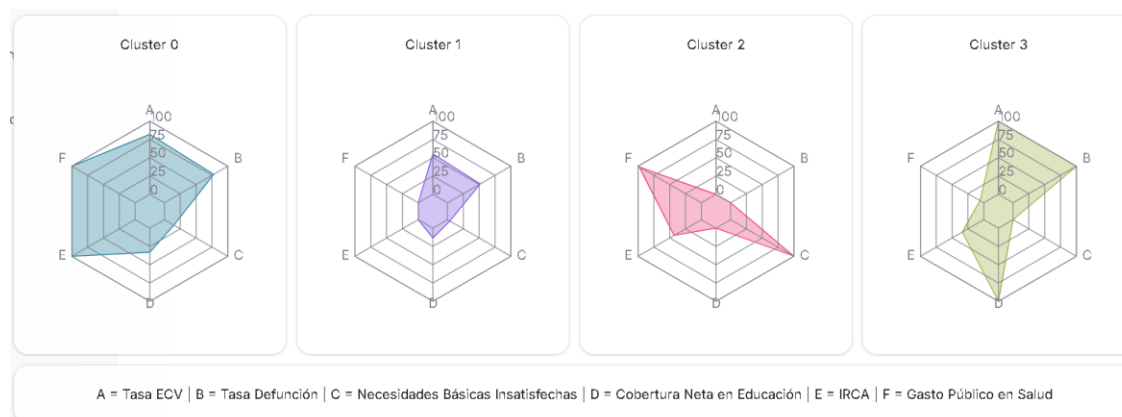
En los gráficos de rada se muestra los valores promedio de estos determinantes para un clúster específico, por ejemplo, el clúster 0 (azul) evidencia una tasa de ECV y tasa de defunción (A y B) más altas, pero un gasto público en salud (F) relativamente elevado. Esto podría sugerir que este grupo enfrenta un alto riesgo de enfermedades cardiovasculares, pero con intervenciones públicas en curso. En cuanto al clúster 1 (morado) se encuentra que presenta bajos niveles de tasa de ECV y defunción, y parece tener mejores condiciones en términos de educación (D) y gasto en salud (F), lo que podría indicar un grupo de bajo riesgo cardiovascular. El clúster 2 (rojo) indica tener as necesidades básicas insatisfechas (C) más altas y baja cobertura en educación (D), lo que podría indicar vulnerabilidad socioeconómica y un riesgo más elevado de sufrir enfermedades cardiovasculares. Finalmente, el clúster 3 (verde) muestra una tasa de defunción (B) relativamente alta y un bajo gasto en salud (F), lo que podría indicar una deficiencia en las políticas de salud pública, a pesar de tener un bajo IRCA (E), que refleja una mejor calidad del agua.

La comparación entre los clústeres ayuda a entender qué áreas necesitan mayor atención en términos de políticas públicas, como el acceso a la salud, la educación, o la mejora en las condiciones de vida (como el acceso al agua potable o la cobertura de necesidades básicas).

Esto proporciona una herramienta valiosa para diseñar intervenciones específicas para cada grupo.

Figura 34

Visualización pestaña de modelación de los datos gráfico de radar



En cuanto al resumen por clúster se encuentra la siguiente tabla, donde se evidencia que el clúster 0 tiene la tasa de ECV más baja (129), pero una tasa de defunción relativamente alta (355), lo que sugiere que otros factores podrían estar afectando la mortalidad. También tiene el menor porcentaje de NBI (10.4%), lo cual es positivo, pero su cobertura educativa (66.4%) y el IRCA (4.5) indican áreas de mejora.

Por otro lado, el clúster 3 tiene la tasa de ECV más alta (153) y la mayor tasa de defunción (427), lo que podría ser resultado de su alto porcentaje de NBI (32.5%) y la menor cobertura educativa (44.2%). El IRCA de 10.4 también refleja un entorno más desafiante. El gasto público en salud es crucial. Los clústeres con mejor cobertura y menores tasas de NBI tienden a tener mejores resultados de salud, lo que subraya la importancia de políticas sociales y económicas sólidas.

Figura 35

Visualización de datos de resumen por clúster obtenido

Cluster	Tasa ECV (A)	Tasa Defunción (B)	NBI (C)	Cobertura Neta en Educación (D)	IRCA (E)	Gasto Público en Salud (F)
0	129	355	20	79	43	16233
1	94	269	15	77	10	464
2	20	101	59	76	42	7247
3	153	427	13	86	12	5807

Finalmente, en la misma pestaña se encuentra el mapa de coropletas donde se logra observar cómo se distribuye la tasa de ECV y las condiciones socioeconómicas para cada uno de los municipios de la zona pacífica colombiana. Cada uno de los clústeres se encuentra representado por un color específico, lo que tiene un significado importante en su interpretación. Adicional a ello, en el mapa de la región pacífica se encuentra una breve interpretación del significado de cada clúster.

Figura 36

Visualización mapa de coropletas



El tablero proporciona una herramienta dinámica de análisis que no solo ilustra la situación actual, sino que también permite monitorear el efecto de futuras intervenciones. La capacidad de desglosar los determinantes clave y observar su interacción dentro de cada clúster facilita la toma de decisiones basadas en datos y evidencia. Esto es crucial para garantizar que los recursos se asignen de manera eficiente, dirigiendo mayores esfuerzos hacia aquellos grupos con mayores carencias en educación, acceso a servicios de salud o calidad de vida, contribuyendo así a reducir las desigualdades socioeconómicas y los riesgos asociados a las ECV en la población.

8. CONCLUSIONES

El estudio tuvo como objetivo identificar la incidencia de las condiciones socioeconómicas en la mortalidad por ECV en la región pacífica colombiana para ayudar en la focalización de los recursos en políticas públicas de salud a través de un análisis geoestadístico. Bajo dicho contexto, el estudio realizó en primer momento, un análisis descriptivo de las estadísticas recopiladas para posteriormente identificar los factores con mayor capacidad explicativa en la mortalidad por ECV en la región pacífica colombiana.

A raíz de ello, se encontró que las variables con mayor poder explicativo fueron el NBI, la cobertura neta en educación, el IRCA, el valor agregado y el gasto público en salud. Con estas variables de explicativas se procedió a realizar la clusterización de los datos encontrando que los municipios de la región pacífica colombiana se podrían clasificar en cuatro grandes clústeres. El método utilizado para obtener estos resultados fue el de K-means, sin embargo, para verificar los datos encontrados también se utilizaron otras técnicas como el K-means ++, OPTICS y GMM. Luego de realizar la evaluación de los modelos mediante métricas estadísticas para identificar la consistencia, dispersión y separación de los clústeres, se encontró que el mejor modelo fue el de K-means con la reducción de dimensionalidad de T-SNE porque fue el modelo que obtuvo las mejores métricas de Dunn (que evalúa si los clústeres están bien separados y compactados) y la métrica de Davis Boulding que permite evaluar su la dispersión dentro de los clústeres es óptima.

Frente a estos resultados, la clusterización evidenció 4 grupos claramente definidos que se clasificaron así: (a) clúster 0: municipios con alta carga y desigualdades latentes; (b) clúster 1: Municipios de contrastes socioeconómicos; (c) clúster 2: Municipios vulnerables con riesgos latentes; y (d) clúster 3: Municipios resilientes con desafíos en salud. Cada grupo se caracterizó por tener unas condiciones socioeconómicas particulares y una tasa de mortalidad por ECV.

Para realizar la validación de los resultados obtenidos, se consultó a expertos en el sector salud y de políticas públicas acerca de los hallazgos de la investigación. De forma general, los expertos consultados resaltaron que los resultados evidencian la radiografía en salud que tiene el país porque se evidencia que los determinantes socioeconómicos son una variable de interés al momento de diseñar, planificar, ejecutar y evaluar las políticas públicas que se implementan desde el sector público. En efecto, los expertos reconocen la importancia de realizar estudios relacionados con la temática porque dicha información es esencial para facilitar el proceso de toma de decisiones en política social y ayudar a focalizar acciones, estrategias y presupuestos de acuerdo a las necesidades del territorio.

Así mismo, reconocieron que el aporte del proyecto aplicado al área de salud y políticas públicas es importante porque la mayoría de los estudios que analizan las condiciones socioeconómicas en salud se han centrado en enfermedades como el EDA y el IRA, siendo las ECV un área que se puede profundizar considerando las estadísticas de crecimiento de su mortalidad en largo plazo y una población mayor creciente. Igualmente, los expertos consideran que los estudios de este tipo también ayudan a fortalecer los programas de promoción y prevención en salud porque evidencia la importancia de abordar las

enfermedades en su fase aguda y no crónica, facilitando la focalización y eficacia de los recursos públicos en salud.

Así, el análisis geoespacial realizado en la investigación permite identificar las características de los diferentes municipios de la región pacífica colombiana y su tasa de mortalidad por ECV con el objetivo de determinar acciones de política pública que puedan ayudar en la disminución de la mortalidad por este grupo de enfermedades. A juicio de los expertos, los resultados del estudio apoyan los procesos de seguimiento y monitoreo de las políticas públicas y aportan al fortalecimiento institucional y articulación entre entidades territoriales para incrementar los impactos positivos de las estrategias ejecutadas por el gobierno.

Al considerar las recomendaciones de expertos y la revisión de literatura se establecieron una serie de recomendaciones concretas a cada clúster de municipio con el objetivo de fortalecer programas de intervención social, mejorar las metodologías de focalización en políticas públicas, evaluar e identificar los impactos de las estrategias y políticas públicas existentes en los territorios y abordar políticas de prevención y mitigación de riesgo para los clústeres que obtuvieron resultados positivos en la mortalidad por ECV. Por ello, el estudio ofrece recomendaciones concretas que pueden ayudar a la toma de decisiones a secretarías de salud, instituciones públicas y decisores de política pública para la planificación, ejecución, monitoreo y seguimiento de políticas sociales en la región pacífica colombiana.

El estudio de investigación culminó con el despliegue de un tablero de acceso público que tiene como objetivo ser difundido a entidades municipales para que sirva como insumo para la toma de decisiones en materia de política pública en salud.

A pesar de los resultados obtenidos en el estudio, es importante resaltar que su ejecución también incluyó algunas limitaciones al proceso analítico. En primer lugar, se debe reconocer que el periodo de análisis temporal (2016-2020) fue corto, lo cual no puede afectar el análisis de impacto de algunas políticas públicas que se diseñan con periodos de 10 años o más. Sin embargo, al ser un estudio exploratorio se espera que esta investigación se constituya como una base y soporte para próximos estudios. De igual manera, es importante destacar que algunas variables de interés para las enfermedades ECV no pudieron ser analizadas debido a la ausencia de datos a nivel municipal, dichas variables fueron hábitos alimenticios y actividad física que solo se encuentra desagregada para las principales áreas metropolitanas del país.

9. RECOMENDACIONES

La principal recomendación del estudio se fundamenta en la utilización de técnicas de analítica de datos para apoyar los procesos de seguimiento, monitoreo y evaluación de políticas públicas tanto en el área de salud como de otros ámbitos. Esta recomendación surge al reconocer que el acceso a datos estadísticos contextualizados y actualizados pueden apoyar los procesos de toma de decisión por parte de las entidades públicas. El acceso a información actual permitirá que los procesos de seguimiento y monitoreo de políticas públicas sea más eficiente y apoye a la planificación y al curso de acción de estrategias y planes que se desarrollen desde el gobierno. Esta recomendación se dirige principalmente a las entidades departamentales (gobernaciones) debido a que el procesamiento de datos no siempre es factible en todos los municipios, sin embargo, bajo los principios de complementariedad de la acción pública, las entidades gubernamentales pueden apoyar en la planificación de políticas públicas de entidades territoriales que así lo requieran.

En segundo lugar, se recomienda incentivar las investigaciones sobre salud y política pública al evidenciar que este tipo de estudios no es muy frecuente en el país, sin embargo, puede llegar a tener un impacto positivo en la gestión pública de recursos y programas ejecutados por el gobierno nacional. En ese orden de ideas, apoyar el desarrollo de este tipo de investigaciones desde la academia, el sector público o privado permite contar con información más reciente para fortalecer la institucionalidad de las políticas públicas y apoyar en los esfuerzos de focalización, disminución de brechas de desigualdad e impacto en las condiciones de salud y bienestar de los individuos.

Finalmente, también se recomienda el fortalecimiento de la analítica de datos en las instituciones públicas debido a que la comprensión de datos estadísticos es una herramienta poderosa para guiar y facilitar los procesos de toma de decisiones. Hoy en día, los profesionales de ciencia de datos pueden aportar desde su qué-hacer profesional a la comprensión de una diversidad de fenómenos mediante técnicas analíticas que apoyen el diseño, ejecución y evaluación de las políticas sociales del Estado.

REFERENCIAS

- [1] S. Pérez. *Focalización espacial en políticas públicas: una aproximación para mejorar la ruta integral de atención en salud para la población materno perinatal mediante patrones geográficos en Colombia*. Bogotá: Universidad de Los Andes, 2023.
- [2] S. Arroyave, “Las políticas públicas en Colombia. Insuficiencias y desafíos”, *FORUM*, No. 1, 95-111, Julio 2011.
- [3] C. Hooton. *Reframing spatial policy through targeting diagnostic tools: potential and deprivation*. Londres: Cambridge Journal of Regions, Economy and Society, 2021
- [4] A. Sen, “La economía política de la focalización” *Comercio Exterior* Vol. 53, No. 6, 555-562, 2003
- [5] L. Reina, “Focalización y movilidad social: una revisión del caso colombiano”, *Panorama Económico*, Vol. 29, No. 9, 145-159, Abril 2021.
- [6] Organización Panamericana de la Salud. (2022, mayo 13). Día mundial del corazón: enfermedades cardiovasculares causan 1.9 millones de muerte al año en las Américas. [En Línea]. Disponible: https://www3.paho.org/hq/index.php?option=com_content&view=article&id=7252:2012-world-heart-day-cardiovascular-diseases-cause-1-9-million-deaths-year-americas&Itemid=0&lang=es#gsc.tab=0
- [7] Innpulsa Colombia. (2022, septiembre 29). Enfermedades cardiovasculares, un gran desafío para la salud pública en Colombia. [En Línea]. Disponible: [https://www.innpulsacolombia.com/cemprende/noticias/enfermedades-cardiovasculares-un-gran-desafio-para-la-salud-publica-en-colombia#:~:text=En%20Colombia%2C%20durante%202021%2C%20se,de%20Estad%20C3%ADsticas%20\(DANE\)%204.](https://www.innpulsacolombia.com/cemprende/noticias/enfermedades-cardiovasculares-un-gran-desafio-para-la-salud-publica-en-colombia#:~:text=En%20Colombia%2C%20durante%202021%2C%20se,de%20Estad%20C3%ADsticas%20(DANE)%204.)
- [8] Deloitte Access Economics. (2017, marzo 3). Economía colombiana perdió 6.4 billones por enfermedades cardíacas. [En Línea]. Disponible: <https://pcnpost.com/economia-colombiana-perdio-64-billones-enfermedades-cardiacas/>
- [9] M. Whitehead, G. Dahlgren. *Conceptos y principios de la lucha contra las desigualdades sociales en salud: Desarrollando el máximo potencial de salud para toda la población*. Liverpool: Universidad de Liverpool – OMS, 2010.
- [10] M. Marmot. “Economics and social determinants of disease”, *World Health Organization*, Vol. 79, No. 10, 988-989, 2001.
- [11] J. Parra., D. Cardona, Cerezo. M. “Análisis de conglomerados para el estudio de las desigualdades sociales por enfermedades cardiovasculares”, *Rev. Salud Pública*, Vol. 19, No. 4: 475-483, 2017.

[12] A. Hernández, D. Díaz, D. Espinoza, S. Vilcarromero. “Análisis espacial de la mortalidad distrital por enfermedades cardiovasculares en las provincias de Lima y Callao”, *Rev Peru Med Exp Salud Pública*, Vol, 33, No. 1: 185-196, 2016.

[13] D. Camelo. *Modelo de conglomerado para el mapa de datos epidemiológicos*, Granada: Universidad de Granada, 2021.

[14] P. Stack. *Métodos de Aprendizaje Supervisado y no Supervisado para la Estimación de Microestructura Cerebral en Datos de DWMR*, Guanajuato: Centro de Investigación en Matemáticas, A. C., 2021.

[15] S Macintyre. “The Black Report and beyond: what are the issues?” *Soc Sci Med*. Vol, 44, No. 44: 723-45, 1997.

[16] M. Gibbons. “A Historical Overview of Health Disparities and the Potential of eHealth Solutions” *Journal of Medical Internet Research*. Vol, 7. No. 5, 2005.

[17] M marmota, G Rosa, M Shipley, P Hamilton. “Employment grade and coronary heart disease in British civil servants”. *Journal Epidemiol Community Health*, Vol. 32. No. 4: 244-249. 1978.

[18] World Healt Organization. “A *conceptual framework for action on the social determinants of health*”. Geneva. World Health Organization 2010.

[19] Braveman PA, Cubbin C, Egerter S, Chideya S, Marchi KS, Metzler M, et al. Socioeconomic status in health research: One size does not fit all. *JAMA*. 2006;294(22):2879–88.

[20] Havranek EP, Mujahid MS, Barr DA, Blair IV, Cohen MS, Cruz-Flores S, et al. Social determinants of risk and outcomes for cardiovascular disease: A scientific statement from the American Heart Association. *Circulation*. 2015;132(9):873–98. doi:10.1161/CIR.0000000000000228.

[21] Schultz WM, Kelli MH, Lisko JC, Varghese T, Shen J, Sandesara P, et al. Socioeconomic status and cardiovascular outcomes: Challenges and interventions. *Circulation*. 2018;137(20):2166–178. doi:10.1161/CIRCULATIONAHA.117.029652.

[22] Whitehead M. Life and death over the millennium. In: Drever F, Whitehead M, eds. *Health inequalities* London: Stationery Office, 1997:7-28.

[23] J. Mackenbach. Social inequality and death as illustrated in late-medieval death dances. *Am J Public Health* 1995;85:1285-92.

[24] ACNUR (8 de agosto 2018). ¿Qué es desigualdad, qué tipos existen y qué consecuencias tiene? [En línea] Disponible: https://eacnur.org/es/blog/que-es-desigualdad-que-tipos-existen-y-que-consecuencias-tiene-tc_alt45664n_o_pstn_o_pst

[25] CEPAL. La matriz de la desigualdad social en América Latina. Santiago: Comisión Económica para América Latina y el Caribe. 2016.

[26] D. Lawlor. “Socioeconomic inequalities in health”. *BMJ*. Vol, 12. No. 7601: 963-964. doi: 10.1136/bmj.39203.661829.BE.

[27] Organización Mundial de la Salud. (11 de junio 2021) “Enfermedades cardiovasculares (ECV)” [En Línea]. Disponible: [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds))

[28] B. Pérez, E. García, A. Graciani, P. Guallar, E. López, L. Muñoz, F. Rodríguez, “*Desigualdades sociales en los factores de riesgo cardiovascular de los adultos mayores de España: estudio ENRICA-Seniors*”. Vol, 70. No. 3. 2017.

[29] M Davari, M Reza, E. Khorasani. “Socioeconomic status, cardiac risk factors, and cardiovascular disease: A novel approach to determination of this association”. *ARYA Atheroscler*. Vol. 15. No. 6: 260-266. 2019.

[30] Sarker, Iqbal. Context-aware rule learning from smartphone data: survey, challenges and future directions. 95, 2019, *Journal of Big Data*.

[31] Sarker, Iqbal. Cybersecurity data science: an overview from machine learning perspective. y otros. 41, 2020, *Journal of Big Data*.

[32] Mohammed, Mohssen, Badruddin, Muhammad y Bashier, Eihab. *Machine Learning Algorithms and Applications*. s.l. : CRC Press , 2016.

[33] Sarker, Iqbal. *Machine Learning: Algorithms, Real-World Applications and Research Directions*. 160, 2021, *SN Computer Science*.

[34] Y. Tishan, «Understanding the Difference Between Supervised and Unsupervised Learning Techniques,» University of Plymouth, 2023.

[35] B. Wah, *Supervised and Unsupervised Learning for Data Science*, Springer, 2020.

[36] T. Oyana, «A New-Fangled FES-k-Means Clustering Algorithm for Disease Discovery and Visual Analytics,» *EURASIP J Bioinform Syst Biol.*, vol. 1, 2010.

[37] B. Kim, J. Kim y G. Yi, «Analysis of Clustering Evaluation Considering Features of Item Response Data Using Data Mining Technique for Setting Cut-Off Scores,» *Symmetry*, vol. 9, n° 5, 2017.

[38] P. Sharma, «The Ultimate Guide to K-Means Clustering: Definition, Methods and Applications,» 2023. [En línea]. Available: <https://www.analyticsvidhya.com/blog/2019/08/comprehensive-guide-k-means-clustering/>.

[39] X. Wang y J. Wang, «Using Clustering Methods in Geospatial Information Systems,» *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 64, n° 3, 2009.

[40] Y. Zhong, J. Li y S. Zhu, «Clustering Geospatial Data for Multiple Reference Points,» *IEEE Access PP*, vol. 99, 2019.

[41] K. Pedersen, R. Jensen, L. Hall, M. Cutler, M. Transtrum, K. Gee y S. V. Lympany, «K-Means Clustering of 51 Geospatial Layers Identified for Use in Continental-Scale Modeling of Outdoor Acoustic Environments,» *Appl. Sci.*, vol. 13, n° 14, 2023.

[42] S. Aungkulanon, V. Tangcharoensathien, K. Shibuya, K. Bundhamcharoen, V. Chongsuvivatwong. “Area-level socioeconomic deprivation and mortality differentials in Thailand: results from principal component analysis and cluster análisis”, *International Journal for Equity in Health*, vol, 16, No, 117, 2017.

[43] R. Muthalaly, A. Nelson, A. Baradi, H. Mehta, M Wilson, A. Nasis. “Socioeconomic determinants of health, traditional risk factors and cardiovascular outcomes in Australia”, *International Journal of Cardiology Cardiovascular Risk and Prevention*, vol, 17, No. 200184, 2023.

[44] M Pérez, J. Achcar. “Desigualdades socioeconómicas en la mortalidad por enfermedades cardiovasculares: Región Pacifico de Colombia, 2002-2015”. *Ciência & Saúde Coletiva*, Vol, 26, No. 3: 5201-5214, 2021.

[45] Wirth, Rüdiger Jochen Hipp: *Crisp-dm: Towards a standard process model for data mining CRISP-DM: Towards a standard process model for data mining. Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, 1, 29–39. Manchester, 2000.

[46] A. Getis. *Reflections on Spatial Autocorrelation., Regional Science and Urban Economics.* 2007

[47] W. Siabato. “La autocorrelación espacial y el desarrollo de la geografía cuantitativa., Cuadernos de Geografía” *Revista Colombiana de Geografía*, Vol. 28, 2019.

[48] V. Olaya. *Sistemas de Información Geográfica., Sistemas de Información Geográfica.* 2020.

[49] E Augusto, B. Lanza. “Spatial análisis of cardiovascular mortality and associated factors around the world”. *BMC Public Health*, Vol, 22, No. 15: 1556, 2022.

[50] S Pico, M Hernández, L Muñoz. “Descripción especial del riesgo cardiovascular en población adulta mayor: caso de Cali-Colombia”. *Nutrición Clínica y Dietética Hospitalaria*, Vol, 42, No. 2: 133-141, 2022.

[51] N Vergara, D Correa, C Moreno, J Mercado, G Basto, J Ocampo. “Determinación del riesgo cardiovascular: un estudio de caso”. *Salud Jalisco*, Vol, 7: 16-23, 2020.

[52] D Gómez, M Prieto, A Mellado, A Moreno. “Análisis espacial de la mortalidad por enfermedades cardiovasculares en la ciudad de Madrid, España”. *Revista Española de Salud Pública*, Vol, 89, No. 1: 35-44, 2015.

[53] L De Andrade, V Zanini, A Batilana, E Carvahlo. “Regional disparities in mortality after ischemic heart disease in a Brazilian state from 2005 to 2010”. *Revista Española de Salud Pública*, Vol, 89, No. 19, 2013.

[54] A Hernández, G Bendezu, D Díaz, M Santero, N Mickas, D Azanedo, D Antiporta. “Spatial analysis of childhood obesity and overweight in Peru”. *RevPeru Med Exp Salud Pública*, Vol, 33: 489-97, 2016.

[55] G Briceño, M Fernández, J Céspedes. “Prevalencia elevada de factores de riesgo cardiovascular en una población pediátrica”. *Biomédica*, Vol, 35, No. 2: 219-226, 2015.

[56] M Ford, L Highfield. “Exploring the spatial association between social deprivation and cardiovascular disease mortality at the neighborhood level”. *PloSOne*, Vol, 11: 229-246, 2016.

[57] J. Jamiolkwski, A. Genowska y A. Pajak. “Is area-level socioeconomic deprivation associated with mortality due to circulatory system diseases in Poland”. *BMC Public Health*, vol. 23, nº 7, pp. <https://doi.org/10.1186/s12889-022-14914-y>, 2023.

[58] O. Medina. “Impacto de la pandemia de COVID-19 en las tendencias de mortalidad por enfermedades cardiovasculares en México, 2000-2022”. *Medicina de Familia SEMERGEN*, vol. 50, n° 3, pp.24-36. DOI: 10.1016/j.semerg.2023.102170, 2024.

[59] J. Fernández, X. Merchán, G. Persi, J. Seguí, V. Aldino y otros. “Efectos de la pandemia por COVID-19 con relación al ACV isquémico ¿La pandemia realmente lo cambió todo? Un estudio comparativo pre-post COVID-19 con revisión de literatura”. *Neurología Argentina*, vol. 14, n° 4, pp.244-250. doi: 10.1016/j.neuarg.2022.08.004, 2022.

[60] M. Pérez, J. Achar. “Desigualdades socioeconómicas en la mortalidad por enfermedades cardiovasculares: Región Pacífico de Colombia, 2002-2015”. *Ciencia & Salud Colectiva*, vol. 26, n° 5, pp.5201-5214, DOI: 10.1590/1413-812320212611.3.02562020, 2021.

[61] X. Tang, D. Laskowitz, L. He, T. Ostbye, Y. Cao, y otros, “Neighborhood socioeconomic status and the prevalence of stroke and coronary heart disease in rural China: A population-based study”. *Int J Stroke*. vol. 10 n° 3, pp.388-395. DOI: 10.1111/ij.s.12343, 2015.

[62] L. Morgenstern, B. Kissela, “Stroke disparities: Large global problem that must be addressed”. *Stroke Lippincott Williams and Wilkins*, vol.2, pp. 3560-3563. DOI: 10.1161/STROKEAHA.115.009533, 2015.

[63] D. Guerrero, G. Pestaña, B. Díaz, R. Vargas, N. Alvis. “Mortalidad de las enfermedades cardiovasculares en Colombia: 30 años de observación”. *Acta Neurológica Colombiana*, vol. 37 n° 4, pp.23-35, <https://doi.org/10.22379/24224022386>, 2021.

[64] P. Rodríguez. “Estrategias para la prevención y control de las enfermedades cerebrovasculares”, *Revista Cubana de Neurología y Neurocirugía*, vol. 2 n° 1, pp.61-86, 2012.

[65] Rap Pacífico. “Reflexiones regionales: sobre la salud en el pacífico colombiano”. 2020, disponible en: <https://rap-pacifico.gov.co/wp-content/uploads/2022/04/Inclusividad-en-la-salud-una-respuesta-integral-para-el-territorio.pdf>

[66] J. Romero. “Aspectos socioeconómicos de la mortalidad en el pacífico colombiano”. *Documentos de Trabajo sobre Economía Regional*, vol. Junio n° 240, pp.1-60, 2016.

[67] C. Aponte, E. Romero, L. Santa. “Análisis de datos espaciales del Índice de Necesidades Básicas Insatisfechas en la región andina”. *Perspectiva Geográfica*. vol. 20 n° 2, pp.391-418, ISSN: 0123-3769, 2015.

[68] K. Asprilla, G. Montenegro. “Desigualdades sociales en salud: análisis de la region pacífica colombiana”. *Salud UIS*. vol. 54, pp. e22048. doi: <https://doi.org/10.18273/saluduis.54.e:22048>, 2022.

[69] H. Palacios. “*Desigualdades sociales en la mortalidad por enfermedades transmisibles, tumores malignos, enfermedades cardiovasculares, lesiones de causa externa y diabetes mellitus y su relación con las características socioeconómicas en el departamento del Chocó 2009-2013*”. Editorial Universidad Autónoma de Manizales, 2019.

[70] J. Beaulac, E. Kristjansson, and S. Cummins, "Una revisión sistemática de los desiertos alimentarios, 1966-2007," **Prev. Chronic Dis.**, vol. 6, no. 3, p. A105, 2009. [Online]. Available: http://www.cdc.gov/pcd/issues/2009/jul/08_0163.htm. [Accessed: Nov. 22, 2024].

[71] J. P. Block and S. V. Subramanian, "Moving Beyond 'Food Deserts': Reorienting United States Policies to Reduce Disparities in Diet Quality," **PLoS Med.**, vol. 12, no. 12, p. e1001914, Dec. 2015, doi: 10.1371/journal.pmed.1001914.

ANEXOS

Anexo A

Métricas de evaluación de los modelos de clusterización

a. Evaluación modelo K-means

```
print(f"Davis Bouldin: {davies_bouldin_index(X_tsne[['TSNE1', 'TSNE2']], kmeans_plus.labels_, centroids=kmeans_plus.cluster_centers_)")  
print(f"Dunn Index: {dunn_index(X_tsne[['TSNE1', 'TSNE2']], kmeans_plus.labels_)}")
```

✓ 0.0s

Python

Davis Bouldin: 0.6511975922270532
Dunn Index: 0.0778880437704063

b. Evaluación modelo OPTICS

```
print(f"Davis Bouldin: {davies_bouldin_index(X_tsne[['TSNE1', 'TSNE2']], optics_clustering.labels_, centroids=np.array(centroids_optics))")  
print(f"Dunn Index: {dunn_index(X_tsne[['TSNE1', 'TSNE2']], optics_clustering.labels_)}")
```

✓ 0.0s

Python

Davis Bouldin: 2.4551692138490506
Dunn Index: 0.04507031920928816

c. Evaluación modelo Gaussian Mixture

```
print(f"Davis Bouldin: {davies_bouldin_index(X_tsne[['TSNE1', 'TSNE2']], gmm.predict(X_tsne[['TSNE1', 'TSNE2']]), centroids=gmm.means_)")  
print(f"Dunn Index: {dunn_index(X_tsne[['TSNE1', 'TSNE2']], gmm.predict(X_tsne[['TSNE1', 'TSNE2']]))}")
```

✓ 0.0s

Python

Davis Bouldin: 0.6899541953812869
Dunn Index: 0.06998410332752678

d. Evaluación modelo Kmeans (sin tratamiento previo)

```
print(f"Davis Bouldin: {davies_bouldin_index(X_scaled, kmeans_direct.labels_, centroids=kmeans_direct.cluster_centers_)")  
print(f"Dunn Index: {dunn_index(X_scaled, kmeans_direct.labels_)}")
```

✓ 0.0s

Davis Bouldin: 1.1865116313274215
Dunn Index: 0.057304448750272394

Anexo B

Guion de preguntas para validación de resultados con expertos

Objetivo: El presente guion de preguntas tiene como objetivo realizar una validación de los resultados obtenidos en el estudio titulado “**ANÁLISIS DE CONGLOMERADOS EN DESIGUALDADES Y ENFERMEDADES CARDIO VASCULARES (ECV): UNA BASE PARA LA ASIGNACIÓN DE RECURSOS EN POLÍTICAS PÚBLICAS EN SALUD**”. Las respuestas consignadas por usted serán tratadas de forma confidencial y tienen fines meramente académicos.

Guion de preguntas

1. En el contexto de la salud pública, ¿cómo se determina la distribución de recursos a los municipios más vulnerables o con mayores necesidades? ¿Cuáles son los factores clave que se consideran en este proceso?
2. A partir de su experiencia en el diseño y ejecución de políticas públicas, ¿cuáles considera que son los mayores retos al abordar las desigualdades socioeconómicas que inciden en la prevalencia de enfermedades cardiovasculares en distintas regiones? ¿Cómo afectan estas desigualdades la efectividad de las intervenciones en salud?
3. ¿Cómo ha sido su experiencia en la identificación y manejo de desigualdades en salud, especialmente en relación con las enfermedades cardiovasculares (ECV)?
4. En el análisis de conglomerados, uno de los clústeres describe “Municipios Resilientes con Desafíos de Salud”, donde a pesar de una alta inversión en salud, se observan elevados niveles de ECV. En su experiencia, ¿ha observado una situación similar en algunas regiones del país? ¿A qué factores podría atribuir esto?
5. Entre algunos de los resultados obtenidos en nuestro estudio se puede observar la existencia de “Municipios Vulnerables con Riesgo Latente”, es decir, que muestra bajos niveles de ECV en la actualidad, pero que al mismo tiempo enfrentan grandes desafíos socioeconómicos. ¿Cree que estos desafíos podrían traducirse en mayores problemas de salud en el futuro? ¿Qué políticas públicas cree que deberían priorizarse en estos municipios?
6. El clúster “Municipios de Contrastes Socioeconómicos” refleja una paradoja entre prosperidad económica y baja inversión en salud y educación. ¿Ha notado un patrón similar en su trabajo? ¿Qué tipo de intervenciones podrían equilibrar esta dicotomía?
7. Finalmente, los “Municipios con Alta Carga Sanitaria y Desigualdades Latentes” tienen altos índices de ECV y mortalidad, pero también una considerable inversión en salud. Desde su perspectiva, ¿por qué cree que esta inversión no ha logrado reducir las desigualdades en estos casos?
8. Considerando los municipios identificados como "Municipios con Alta Carga Sanitaria y Desigualdades Latentes", ¿qué tipo de políticas públicas cree que se deben priorizar para lograr un impacto significativo en la reducción de las enfermedades cardiovasculares?
9. En su trayectoria, ¿ha visto cambios significativos en los resultados de salud cardiovascular debido a la implementación de políticas públicas específicas? ¿Cuáles han sido los factores que han facilitado o dificultado estos cambios?

10. ¿Cómo se asegura la sostenibilidad de las políticas implementadas para combatir las desigualdades en salud y enfermedades cardiovasculares? ¿Qué estrategias considera clave para que las políticas a largo plazo se mantengan eficaces?

11. Desde su perspectiva, ¿cuál es el rol de la coordinación interinstitucional (salud, educación, desarrollo social) en la reducción de las desigualdades en enfermedades cardiovasculares? ¿Cómo se puede mejorar esta colaboración en políticas públicas?

12. ¿Qué estrategias o innovaciones considera cruciales para reducir las desigualdades en salud, especialmente en lo que respecta a las enfermedades cardiovasculares?

Anexo C

Pantallazos del código de salida de la programación

```
# dataset original entregado
data_cleaned = pd.read_excel("data_cleaned.xlsx",
                             usecols=['id_dep', 'id_mun', 'año', 'nbi', 'va', 'gasto_publico_salud', 'irca', 'inver_dep_rec',
                             dtype={'id_dep':str, 'id_mun':str})
data_cleaned.rename(columns={'id_dep': 'COD_DPTO', 'id_mun': 'COD_MUNIC', 'año': 'ANO'}, inplace=True)
data_cleaned['COD_MUNIC'] = data_cleaned['COD_MUNIC'].apply(lambda x: str(x).zfill(3))
data_cleaned['COD_MUNIC'] = data_cleaned['COD_DPTO'] + data_cleaned['COD_MUNIC']

# validaciones poblacion
poblacion18 = pd.read_excel("data/poblacion2018.xlsx", skiprows=10, header=1)
poblacion18 = poblacion18[poblacion18['ÁREA GEOGRÁFICA'] == 'Total'][['COD_DPTO', 'DEPARTAMENTO', 'COD_DPTO-MPIO', 'MUNICIPIO',
poblacion18.columns = ['COD_DPTO', 'DEPARTAMENTO', 'COD_MUNIC', 'MUNICIPIO', 'ANO', 'Poblacion']

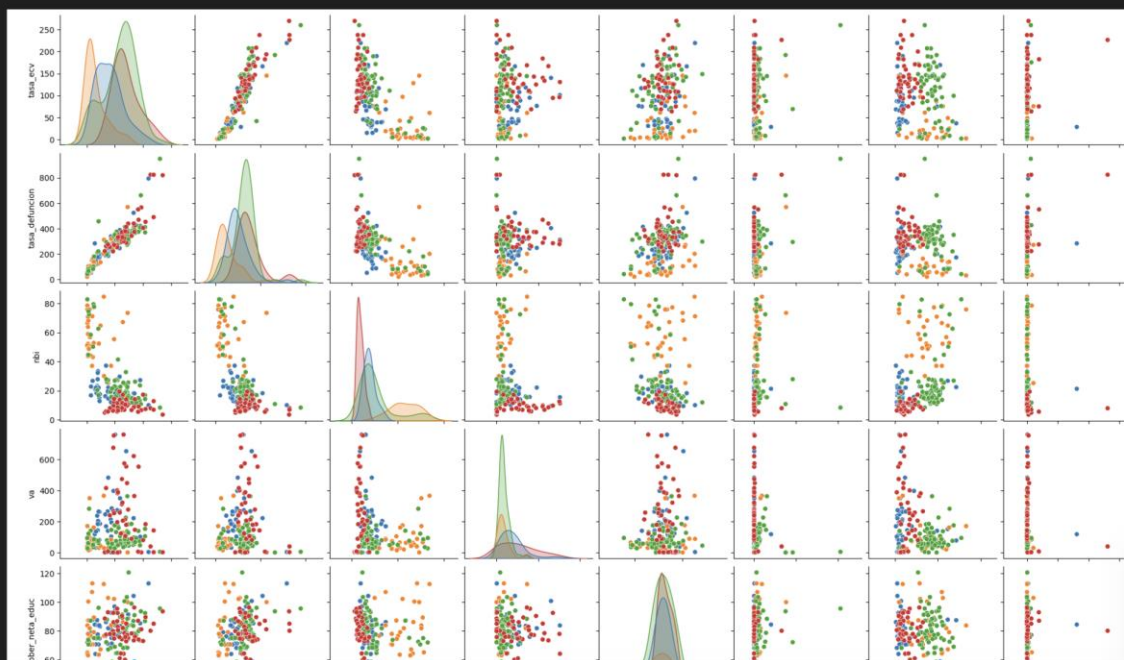
poblacion19 = pd.read_excel("data/poblacion2019.xlsx", skiprows=10, header=1)
poblacion19 = poblacion19[poblacion19['ÁREA GEOGRÁFICA'] == 'Total'][['DP', 'DPNOM', 'MPIO', 'DPMP', 'AÑO', 'Población']]
poblacion19.columns = ['COD_DPTO', 'DEPARTAMENTO', 'COD_MUNIC', 'MUNICIPIO', 'ANO', 'Poblacion']
poblacion19 = poblacion19[(poblacion19['ANO'] >= 2015) & (poblacion19['ANO'] <= 2017)]

poblacion = pd.concat([poblacion18, poblacion19], axis=0)
poblacion.sort_values(['COD_DPTO', 'COD_MUNIC', 'ANO'], inplace=True)
poblacion.drop_duplicates(inplace=True)
poblacion['COD_DPTO'] = poblacion['COD_DPTO'].astype(str)
poblacion['COD_MUNIC'] = poblacion['COD_MUNIC'].astype(str)
```

Python

```
# sns.scatterplot(data=input_modelo, x='tasa_ecv', y='tasa_defuncion')
sns.pairplot(data=input_modelo, hue='DEPARTAMENTO')
```

<seaborn.axisgrid.PairGrid at 0x30ff32e10>



```
sns.scatterplot(X_tsne, x='TSNE1', y='TSNE2', hue='cluster_tsne_plus', palette='dark')
```

<Axes: xlabel='TSNE1', ylabel='TSNE2'>

