

**Diseño de un tablero de control en Power BI para la caracterización de casos de Dengue en la región Caribe de Colombia (2010-2024) mediante analítica de datos y un modelo de clustering K-means**

Raúl Hernán Pérez Ávila

Asesor

Isaac Esteban Camargo Freile

Universidad Nacional Abierta y a Distancia UNAD

Escuela de Ciencias Sociales Artes y Humanidades ECSAH

Especialización en Ciencias de Datos y Analítica

2025

### **Dedicatoria**

A mi esposa y mis hijos, quienes me apoyaron con su amor, fuerza y fe; sacrificando tiempo en familia para dedicar a la construcción de este proyecto.

A mis padres, por sus oraciones y amor incondicional, por enseñarme el valor de los pequeños éxitos.

### **Agradecimientos**

Agradezco al Espíritu Santo de Dios, quien me acompañó, me guió y me dio la sabiduría para cumplir esta nueva meta.

## Resumen

El Dengue es más que una estadística en los municipios y departamentos de la región Caribe de Colombia; es una realidad que afecta a miles de familias cada año, desafiando constantemente a nuestro sistema de salud pública. Detrás de cada caso reportado hay una historia humana, una comunidad en riesgo y una necesidad urgente de respuesta más ágiles y focalizadas. Conscientes de esta realidad, este proyecto se propone transformar la manera en que se observan y se entienden estos datos, convirtiéndolos en una herramienta para la acción empática y efectiva.

Para lograrlo, se diseñará un tablero de control interactivo en Power BI que permita explorar y comprender la dinámica de los casos de Dengue ocurridos entre 2010 y 2024 en los municipios de los departamentos de la región Caribe de Colombia. A través de la analítica de datos, no solo visualizaremos patrones históricos, geográficos y demográficos, sino que implementaremos un modelo de clustering no supervisado (K-means). Esta técnica nos permitirá ir un paso más allá, identificando y agrupando municipios con perfiles de riesgo similares, revelando focos de transmisión y características comunes que a simple vista son invisibles.

El resultado final será una herramienta intuitiva y accesible, pensada para que los tomadores de decisiones en salud pública puedan navegar la información, generar conocimiento y, lo más importante, orientar estrategias de prevención y control que protejan de manera más eficientes a las comunidades más vulnerables de nuestra región.

***Palabras clave:*** Dengue, Analítica de Datos, Power BI, Salud Pública, Clustering.

## Abstract

In the Caribbean region of Colombia, Dengue is more than a statistic; it is a reality that affects thousands of families each year, constantly challenging our public health system. Behind every reported case, there is a human story, a community at risk, and an urgent need for more agile and focused responses. Aware of this reality, this project aims to transform how this data is observed and understood, turning it into a tool for empathetic and effective action.

To achieve this, an interactive dashboard will be designed in Power BI to explore and comprehend the dynamics of Dengue cases reported between 2010 and 2024. Through data analytics, we will not only visualize historical, geographical, and demographic patterns but also implement an unsupervised clustering model (K-means). This technique will allow us to go a step further by identifying and grouping municipalities with similar risk profiles, revealing transmission hotspots and common characteristics that are otherwise invisible.

The final result will be an intuitive and accessible tool, designed for public health decision-makers to navigate the information, generate knowledge, and, most importantly, guide prevention and control strategies that more efficiently protect the most vulnerable communities in our region.

**Keywords:** Dengue, Data Analytics, Power BI, Public Health, Clustering.

## Tabla de Contenido

Introducción .....	13
Descripción del Problema .....	15
Planteamiento del Problema .....	16
Pregunta Principal de Investigación.....	16
Justificación .....	18
Objetivos .....	21
Objetivo General .....	21
Objetivos Específicos.....	21
Estado del Arte.....	22
Estudio sobre la Caracterización del Dengue en Colombia y la Región.....	22
Aplicación de Machine Learning en la Vigilancia Epidemiológica .....	23
Inteligencia de Negocios y Visualización en la Salud Publica .....	23
Identificación de Brechas y Aportes al Proyecto .....	24
Marco Teórico.....	25
Dengue .....	25
Vigilancia Epidemiológica.....	25
Inteligencia de Negocios (Business Intelligence - BI).....	26
Análisis Exploratorio de Datos (EDA) .....	26
Modelos Analíticos y Machine Learning.....	26
Clustering no Supervisado: El Algoritmo K-means.....	27
Python como Herramienta para el Análisis de Datos.....	27
Marco Conceptual.....	28

Capa 1 El Fundamento de los Datos – El Microdato como Unidad de Análisis .....	28
Capa 2 Modelo Analítico – Del Dato al Patrón Oculto .....	28
Capa 3 La interfaz de Conocimiento – La Democratización de la Decisión .....	29
Marco Normativo.....	31
Derecho a la Salud y Deberes del Estado.....	31
Ley de Protección de Datos Personales (Habeas Data) .....	31
Marco Regulatorio de la Vigilancia en Salud Pública .....	32
Normativas Éticas en Investigación en Salud .....	33
Metodología .....	34
Comprensión del Negocio (o del Problema).....	34
Procedimiento de Selección de Fuentes .....	34
Comprensión de los Datos .....	35
Preparación de los Datos.....	36
Modelado .....	37
Construcción de Variables.....	38
Procesamiento .....	39
Evaluación.....	39
Despliegue (Implementación) .....	39
Tipo de Estudio .....	39
Recolección de Datos .....	40
Resultados .....	43
Caracterización Epidemiológica .....	43
Análisis Temporal (¿Cuándo Ocurre?).....	44

Estacionalidad (Canal Endémico).....	46
Análisis Espacial (¿Dónde Ocurre?) .....	47
Análisis Demográfico y de Gravedad (¿Quién y Como?).....	49
Segmentación de Riesgo Municipal.....	55
Determinación del Número Optimo de Clústeres .....	55
Perfilamiento y Caracterización de los Clústeres.....	56
Distribución Geoespacial del Riesgo .....	58
Implementación Tecnológica y Automatización del Modelo .....	59
Conclusiones.....	61
La Dinámica y la Caracterización Epidemiológica.....	61
Segmentación y el Modelo Analítico.....	61
Nodo de Alta Transmisión (Clúster 0) .....	62
Foco de Alerta por Gravedad (Clúster 3) .....	62
Municipios de Riesgo Pediátricos (Clúster 2).....	62
Zona de Riesgo Bajo/Disperso (Clúster 1).....	62
Arquitectura Tecnológica e Integración.....	62
Calidad de los Datos.....	63
Recomendaciones .....	64
Intervención Intersectorial en Entornos Escolares.....	64
Auditorías Clínicas Focalizadas en el Clúster de Gravedad .....	64
Atención Diferenciada por Vulnerabilidades Biológicas.....	65
Integración de Variables Climáticas para la Predicción .....	65
Adopción de Inteligencia Artificial para la Vigilancia en la Salud .....	66

Referencias Bibliográficas .....	67
Apéndices.....	71

**Lista de Tablas**

**Tabla 1** *Variables Seleccionadas para el Análisis*..... 37

**Tabla 2** *Tabla de Perfilamiento Clústeres (Promedios)* ..... 57

## Lista de Figuras

<b>Figura 1</b> <i>Diagrama de Flujo del Proceso de Selección de Literatura (Metodología PRISMA)..</i>	35
<b>Figura 2</b> <i>Proceso de Datos en la Vigilancia Epidemiológica.....</i>	42
<b>Figura 3</b> <i>Panorama General y Geográfico (Dengue).....</i>	44
<b>Figura 4</b> <i>Tendencia de Casos de Dengue en la Región Caribe (2010-2024) .....</i>	44
<b>Figura 5</b> <i>Estacionalidad del Dengue: Promedio de Casos por Semana Epidemiológica para el Año 2024 .....</i>	46
<b>Figura 6</b> <i>Casos Totales de Dengue por Departamento (2010-2024).....</i>	48
<b>Figura 7</b> <i>Top 10 Municipios con Mayor Número de Casos (2010-2024).....</i>	49
<b>Figura 8</b> <i>Perfil Demográfico y de Gravedad .....</i>	50
<b>Figura 9</b> <i>Distribución de Casos por Sexo (2010-2024).....</i>	51
<b>Figura 10</b> <i>Distribución de Casos por Área de Residencia .....</i>	51
<b>Figura 11</b> <i>Pirámide Poblacional de Casos de Dengue (2010-2024).....</i>	52
<b>Figura 12</b> <i>Indicadores de Gravedad (KPIs) .....</i>	53
<b>Figura 13</b> <i>Determinación de K Mediante el Método del Codo.....</i>	55
<b>Figura 14</b> <i>Coeficiente de Silueta Según el Número de Clústeres .....</i>	56
<b>Figura 15</b> <i>Distribución Espacial de los Clústeres de Riesgo de Dengue en la Región Caribe (K=4). .....</i>	59

**Lista de Apéndices**

<b>Apéndice A</b> <i>Cronograma de Actividades</i> .....	<b>71</b>
--	-----------

## Introducción

El dengue ha dejado de ser una enfermedad lejana, confinada a rincones tropicales, para convertirse en una preocupación de salud pública global. Su avance, silencioso pero implacable, es impulsado por fenómenos como el cambio climático, la urbanización acelerada y la creciente movilidad de las personas. La Organización Mundial de la Salud (OMS) ha alertado sobre su dramática expansión, reportando que la incidencia se ha multiplicado por ocho en las últimas dos décadas, poniendo a casi la mitad de la población mundial en zona de riesgo (WHO, 2025). Este llamado global no es solo un aviso estadístico, sino un recordatorio de la vulnerabilidad compartida y de la necesidad de estrategias más inteligentes y colaborativas para enfrentar al mosquito *Aedes Aegypti*, su principal vector.

En el continente americano, la situación refleja esta tendencia global con una urgencia particular. La Organización Panamericana de Salud (OPS) describe un escenario de brotes cada vez más frecuente e intenso, que ponen a prueba la resiliencia de los sistemas de salud desde el Cono Sur hasta el Caribe (PAHO/WHO, 2024). Las Américas se han convertido en un epicentro de la enfermedad, no por falta de esfuerzo, sino por la complejidad de factores sociales y ambientales que facilitan la propagación del virus. La lucha contra el Dengue en nuestra región es una batalla diaria que se libra en los hogares, barrios y centros de salud, donde la prevención y la atención oportuna son las principales herramientas para evitar desenlaces fatales.

Colombia, por su ubicación geográfica y diversidad climática, es un territorio donde el Dengue se manifiesta como una enfermedad endémica y de un desafío prioritario. La responsabilidad de monitorear cada caso recae en el Sistema Nacional de Vigilancia de Salud Pública (Sivigila), cuya información es gestionada por el Instituto Nacional de Salud (INS). Los datos del Sivigila no son simples registros; son el testimonio de como la enfermedad impacta al

país, revelando ciclos epidémicos con picos preocupantes, como los ocurridos en 2010 y 2019 (Instituto Nacional de Salud, 2024). Estos informes nos demuestran que, si bien contamos con un sistema de vigilancia robusto, el volumen y la complejidad de los datos requieren de nuevas capacidades analíticas para pasar de un enfoque reactivo a uno predictivo.

Es en la región Caribe de Colombia donde este desafío adquiere un rostro particularmente humano y tangible. Sus condiciones climáticas, sumadas al factor socioeconómico, crean un entorno propicio para la transmisión sostenida de Dengue, afectando de manera desproporcionada a sus comunidades (Villar et al., 2015a). Para las familias de esta región, el Dengue trasciende las cifras epidemiológicas para convertirse en una carga socioeconómica tangible, manifestada en índices de ausentismo escolar, pérdida de productividad laboral y una constante presión sobre la capacidad hospitalaria. Es aquí donde nuestro proyecto encuentra su propósito fundamental: tomar los datos, propiedad del Instituto Nacional de la Salud, y transformarlos en conocimiento accionable. Buscamos trascender el análisis tradicional para construir una herramienta que permita a los líderes de salud pública ver patrones, anticipar brotes y, en última instancia, proteger con mayor eficiencia a la gente del Caribe colombiano.

## **Descripción del Problema**

La región Caribe de Colombia. Un territorio de gran riqueza cultural y natural enfrenta de manera persistente un desafío de salud pública que afecta profundamente a sus comunidades: el Dengue. A pesar de los continuos esfuerzos de las autoridades sanitarias, la enfermedad mantiene un comportamiento endemoepidémico, lo que significa que siempre está presente, pero con brotes cíclicos que saturan el servicio de salud y genera una considerable carga social y económica.

La vigilancia de la enfermedad se centraliza en los datos recopilados por el Sistema Nacional de Vigilancia en la Salud Pública (Sivigila). Esta base de datos, propiedad del Instituto Nacional de Salud, es un repositorio masivo y detallado que registra más de catorce años de historia de la enfermedad en el país. Contiene variables demográficas, geográficas y clínicas de cada caso, representando una oportunidad sin precedentes para un análisis profundo.

Sin embargo, la practicas actual de la vigilancia epidemiológica a menudo se ve limitada por un enfoque retrospectivo. Los informes y boletines, aunque informativos, tienden a presentar los datos de forma agregada y estática. Esta aproximación dificulta la identificación en tipo real de patrones complejos, la detención de conglomerados de casos a nivel local y la comprensión de las dinámicas de transmisión que subyacen a los brotes. En esencia, existe una brecha significativa entre la enorme cantidad de datos disponibles y la capacidad para transformarlos en conocimiento estratégico y oportuno que guie intervenciones preventivas eficaces.

## **Planteamiento del Problema**

El problema central que aborda este proyecto es la limitada capacidad de los métodos de análisis epidemiológicos tradicionales para explotar la riqueza de los microdatos del Sivigila, lo que resulta en una vigilancia de la salud pública predominantemente reactiva y con un alcance estratégicamente restringido.

Actualmente, las entidades territoriales de la salud en la región Caribe carecen de herramientas interactivas y dinámicas que les permitan explorar de manera ágil el comportamiento histórico y actual del Dengue. La falta de un sistema de visualización georreferenciado y la ausencia de modelos analíticos avanzados impiden responder a preguntas cruciales de forma rápida: ¿Dónde están surgiendo los focos de transmisión en este preciso momento? ¿Qué municipios comparten perfiles de riesgos similares, aunque no sean geográficamente contiguos? ¿Existen patrones demográficos o temporales que se repiten antes de cada gran brote?

Estas limitaciones trascienden lo meramente técnico y genera repercusiones humanas directas. La incapacidad para anticipar brotes o focalizar recursos de manera óptima debilita las campañas de prevención, aumentando el riesgo de que los casos escalen a niveles epidémicos. En ese contexto, la literatura científica reconoce que, en la era de la Big Data, el verdadero desafío en la salud pública ya no es recolectar datos, sino desarrollar la capacidad analítica para interpretarlos y actuar sobre ellos de manera proactiva (Gopalakrishnan et al., 2022).

### **Pregunta Principal de Investigación**

¿De qué manera el diseño y la implementación de una herramienta de inteligencia de negocios en Powe BI, que integra un modelo de machine learning no supervisado (k-means), puede fortalecer la capacidad de análisis y la vigilancia estratégica del Dengue para la toma de

decisiones en salud pública en la región Caribe de Colombia, utilizando los datos del Sivigila del periodo 2010-2024?

## Justificación

A nivel mundial, el Dengue se ha consolidado como la enfermedad viral transmitida por mosquito de más rápida propagación, representando una amenaza creciente para la salud pública global (WHO, 2025). Organizaciones internacionales advierten que cerca de la mitad de la población del planeta vive actualmente en áreas de riesgo, con una incidencia que ha aumentado que ha aumentado en las últimas décadas debido a factores como el cambio climático y la urbanización no planificada. En el contexto de América, esta situación es particularmente crítica, donde los brotes cíclicos saturan los servicios sanitarios y exponen la vulnerabilidad de las economías emergentes (OPS, 2024). Colombia por sus características geográficas, no es ajena a este escenario, manteniéndose como un territorio endémico donde la enfermedad persiste y evoluciona cada año.

Específicamente en la región Caribe de Colombia, la temporada de lluvias trae consigo no solo un cambio en el clima, sino también la sombra de una amenaza recurrente: El Dengue. Esta enfermedad transmitida, por un mosquito que convive en nuestros hogares y comunidades, representa una carga inmensa que va mucho más allá de las cifras de salud. El Dengue tiene un profundo impacto económico y social, medido en días de trabajo perdidos, costo de hospitalización que agobian a las familias y una presión constante sobre un sistema de salud que lucha por responder con recursos limitados (Laserna et al., 2018). La magnitud de este problema no es una abstracción; es una realidad que exige herramientas más sofisticadas y humanas para su comprensión y control.

En Colombia, contamos con una fuente de datos invaluable: el Sistema Nacional de Vigilancia en Salud Pública (Sivigila), administrado por el Instituto Nacional de Salud. Este sistema recopila meticulosamente cada caso sospechoso y confirmado en el país. Sin embargo,

aquí surge el nudo de nuestra historia: poseer los datos no es sinónimo de tener el conocimiento. La información del Sivigila, a menudo presenta en extensas tablas y reportes estáticos, es como un libro escrito en idiomas que pocos pueden interpretar con la agilidad que una epidemia demanda. Los análisis tradicionales suelen ofrecer una vista panorámica, agregando casos por departamentos o municipios, pero fallan en revelar los micro-focos de transmisión, esos conglomerados de barrios o veredas donde la intervención sería más efectiva. Esta brecha entre la recolección de datos y la toma de decisiones estratégicas ha sido reconocida como un desafío clave en la salud pública moderna (Taipalus et al., 2023).

Aquí es donde nuestro proyecto propone un desenlace innovador. La creación de un tablero de control en Power BI no es simplemente un ejercicio de visualización; es una respuesta directa a la necesidad de democratizar y dinamizar la información epidemiológica. A diferencia de los informes estáticos, un tablero interactivo permite a los epidemiólogos y gestores de salud interactuar dinámicamente con los datos: filtrando por semana, explorar un municipio específico, o cruzar con variables demográficas con un solo clic. Estudios previos han demostrado cómo las herramientas de Business Intelligence, como Power BI, pueden revolucionar el análisis en la salud al transformar datos complejos en conocimiento intuitivo y accionable (Prakash Mahto, 2025). Lo novedoso de nuestra propuesta es que los caos no se mostraran como cifras agrupadas, sino como polígonos georreferenciados en un mapa, permitiendo una identificación visual inmediata de los puntos calientes.

Pero la verdadera vanguardia de este trabajo reside en la integración de un componente de ciencias de datos a través de Python. No nos conformaremos con visualizar lo que es evidente; buscaremos descubrir lo que esta oculto. Mediante la implementación de un modelo de machine Learning no supervisado, específicamente el algoritmo K-means, daremos un salto

cualitativo. Este modelo analizará el comportamiento del Dengue a lo largo de catorce años y agrupará los municipios no por su cercanía geográfica, sino por la similitud de sus patrones epidémicos. La aplicación de clustering para identificar zonas de riesgo de Dengue es una técnica avanzada que permite optimizar la asignación de recursos y diseñar campañas de prevención mucho más focalizadas (Sembiring et al., 2021a).

En resumen, este proyecto se justifica por su capacidad de cerrar la brecha entre la abundancia de datos y la escasez de conocimiento estratégico. Al combinar la potencia visual de Power BI con la profundidad analítica de un modelo de K-means, entregaremos una herramienta que no solo reporta el pasado, sino que ilumina el camino para un futuro donde la vigilancia del Dengue en la región Caribe sea más proactiva, precisa y, sobre todo, más humana.

## Objetivos

### Objetivo General

Desarrollar una herramienta de inteligencia de negocios en Power BI que integre un modelo de segmentación no supervisado, basado en la información de la serie histórica de los microdatos del Sivigila correspondiente al periodo 2010-2024, para la vigilancia epidemiológica y la toma de decisiones estratégicas frente al Dengue en la región Caribe de Colombia.

### Objetivos Específicos

Caracterizar el comportamiento epidemiológico del Dengue en los departamentos y municipios de la región Caribe de Colombia, mediante el análisis de sus patrones de distribución geográfica, temporal y demográfica a partir de los microdatos del Sivigila del periodo 2010-2024.

Implementar un modelo de clustering no supervisado (K-means) para segmentar los municipios de la región en grupos con perfiles de riesgo epidemiológico similares, integrando la potencia estadística de Python dentro de un entorno de inteligencia de negocios de Microsoft Power BI.

Diseñar un tablero de control interactivo en Power BI que visualice de manera efectiva los resultados del análisis descriptivo y la segmentación del modelo k-means, para facilitar la exploración de los datos y soportar la toma de decisiones en salud pública.

## **Estado del Arte**

Para contextualizar la relevancia y originalidad de este proyecto, es fundamental realizar una revisión de la literatura científica existente. Este análisis se ha estructurado en tres áreas convergentes: los estudios epidemiológicos sobre el Dengue en nuestro contexto, la aplicación de técnicas de Machine Learning para el análisis de enfermedades infecciosas y el uso emergentes de herramientas de inteligencia de negocios para la visualización de datos en salud.

### **Estudio sobre la Caracterización del Dengue en Colombia y la Región**

La investigación sobre el Dengue en Colombia no es un campo nuevo. Durante años, la academia y las instituciones de salud han dedicado esfuerzos significativos a entender su comportamiento. Un trabajo fundamental en esta línea es la revisión sistemática de Villar et al. (2015b), que describe las tendencias epidemiológicas del Dengue en Colombia entre 2000 y 2011. Este estudio fue clave para caracterizar la enfermedad a nivel nacional, identificando los ciclos epidémicos y la distribución geográfica de los casos, confirmado a la región Caribe como una de las zonas de más alta endemicidad. De manera similar, estudios más localizados como el de Alvis-Guzman et al. (2015) en el departamento de Córdoba, han proporcionado una visión detallada de la carga de la enfermedad a nivel subregional, analizando incidencias y tasas de mortalidad.

Estos trabajos, si bien son pilares en la epidemiología del Dengue en el país, comparten una característica común: se basan en métodos estadísticos descriptivos tradicionales. Aunque robustos y necesarios, sus resultados suelen presentarse en tablas y gráficos estadísticos, ofreciendo una “fotografía” del pasado en lugar de una herramienta dinámica para el explorar el presente.

## **Aplicación de Machine Learning en la Vigilancia Epidemiológica**

Con el avance de las ciencias de datos, ha surgido una nueva ola de investigaciones que aplican técnicas de Machine Learning para ir más allá del análisis descriptivo. A nivel internacional, es creciente la evidencia sobre el uso de algoritmo de clustering, como el K-means, para la identificación de “puntos calientes” o zonas de alta transmisión de enfermedades vectoriales. Por ejemplo, el trabajo de Sembiring et al. (2021b) demuestra como el algoritmo K-means puede ser utilizado eficazmente para mapear la propagación del Dengue, segmentando áreas geográficas en clústeres de riesgos (bajo, medio y alto). Este tipo de análisis permite una focalización de recursos que el análisis tradicional no facilita.

En el contexto latinoamericano, si bien la adopción de estas técnicas es más incipiente, ya existen investigaciones que exploran el uso de modelos predictivos y de segmentación para enfermedades como el Zika, Chikungunya y el propio Dengue. Sin embargo, una revisión de la literatura revela que la aplicación específica de clustering no supervisado sobre la serie histórica de microdatos del Sivigila para la región Caribe colombiana es un área escasamente explorada.

## **Inteligencia de Negocios y Visualización en la Salud Pública**

El tercer pilar nuestro estado del arte se centra en cómo se comunican los hallazgos. Un análisis, por muy avanzado que sea, tiene un impacto limitado sino se presenta de una manera que sea comprensible y útil para los tomadores de decisiones. Aquí es donde la inteligencia de negocios (BI) cobra relevancia. Artículos como el de Prakash Mahto (2025) argumentan que herramientas como Power BI están “revolucionando” el sector salud al permitir la transformación de datos crudos en conocimiento accionable a través de dashboards interactivos.

En Colombia, varias entidades del sector salud han comenzado a adoptar estas herramientas para el monitoreo de indicadores de gestión. No obstante, su aplicación para la

vigilancia epidemiológica de enfermedades infecciosas a nivel regional, de una manera que integre análisis descriptivo profundos con los resultados de modelos de machine Learning, es aún un campo emergente. La mayoría de las visualizaciones disponibles públicamente se limitan a gráficos de barras o mapas de coropletas que totalizan casos, sin permitir la exploración dinámica o el análisis de clústeres.

### **Identificación de Brechas y Aportes al Proyecto**

La revisión de la literatura nos permite identificar una brecha clara y una oportunidad significativa, por un lado, tenemos estudios epidemiológicos sólidos sobre el Dengue en la región Caribe, pero con un enfoque analítico tradicional. Por otro lado, existen metodologías de Machine Learning válidas para la segmentación de riesgos, pero raramente aplicadas a nuestra fuente de datos específicas. Y finalmente, disponemos de potentes herramientas de visualización que aún no han sido explotadas en todo su potencial para la vigilancia del Dengue en nuestro contexto.

Este proyecto, por lo tanto, no busca aplicar lo ya echo, sino integrar estos tres pilares, el aporte novedoso de nuestro trabajo reside en ser el primer estudio, hasta donde sabemos que utiliza la serie histórica completa de microdatos de SiviGila (2010-2024) para la región Caribe, aplicando un modelo de clustering K-means para descubrir patrones de riesgo no evidentes y, crucialmente, consolidado todos estos hallazgos en una herramienta de inteligencia de negocios interactiva y funcional en Power BI. De esta manera, no solo generamos nuevo conocimiento, sino que lo entregamos en un formato diseñado para la acción.

## **Marco Teórico**

En el desarrollo de este proyecto convergen conceptos del dominio de la salud pública y de la ciencia de datos. Por consiguiente, es indispensable establecer una base conceptual sólida que defina con precisión cada uno de los términos claves que sustentan esta investigación. A continuación, se detallan los pilares teóricos sobre los cuales se construye este análisis.

### **Dengue**

El Dengue es una enfermedad viral, sistémica y dinámica, transmitida principalmente por la picadura del mosquito hembra de la especie *Aedes Aegypti*. La Organización Panamericana de la Salud (OPS) la describe como una infección febril aguda que, aunque en algunos casos cursa de forma leve, puede evolucionar a formas graves con manifestaciones hemorrágicas, shock y, en últimas instancias, la muerte (OPS, 2024). Su complejidad radica en la existencia de cuatro serotipos virales distintos (DENV-1, DENV-2, DENV-3 y DENV-4). Una infección con serotipo proporciona inmunidad de por vida contra ese serotipo en particular, pero una infección posterior con un serotipo diferente es un factor de riesgo conocido para desarrollar Dengue grave. Este fenómeno inmunológico convierte el Dengue en un desafío recurrente para la salud pública, especialmente en regiones tropicales y subtropicales como la nuestra.

### **Vigilancia Epidemiológica**

Para combatir amenazas como el Dengue, la salud pública se apoya en una herramienta fundamental: la vigilancia epidemiológica. El Ministerio de Salud y Protección Social de Colombia, en línea con los Centros de Control y la Prevención de Enfermedades (CDC), la define como el proceso sistemático y constante de recolección, análisis, interpretación y divulgación de datos específicos relacionados con la salud (Ministerio de Salud y Protección Social de Colombia, 2022). El propósito de la vigilancia no es meramente acumulativo; su valor

reside en su capacidad para “utilización en la planificación, ejecución y evaluación de la práctica en la salud pública”. En otras palabras, la información (análisis) y generar una respuesta (acción).

### **Inteligencia de Negocios (Business Intelligence - BI)**

Tradicionalmente asociado al mundo empresarial, el concepto de Inteligencia de Negocios (BI) ha demostrado ser extraordinariamente valioso para el sector salud. Se puede definir como el conjunto de procesos, tecnologías y herramientas que convierten datos crudos en información significativa y conocimiento accionable para facilitar la toma de decisiones más informadas (Microsoft, 2022). En el contexto de la salud pública, una herramienta de BI como Power BI trasciende los reportes estáticos, permitiendo a los epidemiológicos interactuar con los datos, explorar hipótesis en tiempo real y comunicar hallazgos de manera visual e intuitiva.

### **Análisis Exploratorio de Datos (EDA)**

Antes de aplicar modelos complejos, es fundamental “escuchar lo que los datos tienen que decir. Esta es la filosofía del Análisis Exploratorio de Datos (EDA), un enfoque acuñado por el matemático John Tukey. El propósito principal del EDA es utilizar técnicas estadísticas y de visualización para resumir las principales características de un conjunto de datos, identificar errores, detectar patrones y relaciones, formular hipótesis (IBM, n.d.). En nuestro proyecto el EDA será el primer paso crucial para comprender la distribución de los casos de Dengue, identificar valores atípicos y preparar los datos para el modelo posterior.

### **Modelos Analíticos y Machine Learning**

Un modelo analítico es una representación matemática simplificada de un proceso del mundo real, diseñada para ayudarnos a entender ese proceso y hacer predicciones. Dentro del vasto campo de los modelos analíticos, se encuentra el Machine Learning o Aprendizaje

Automático. Esta es una rama de la inteligencia artificial en la que los algoritmos no se propagan con reglas explícitas, sino que “aprenden” patrones directamente de los datos (AWS, n.d.).

El Machine Learning se divide principalmente en dos categorías: supervisado (cuando los datos están etiquetados con la respuesta correcta) y no supervisados (cuando los datos no están etiquetados). Nuestro proyecto se enfoca en este último.

### **Clustering no Supervisado: El Algoritmo K-means**

El clustering es una técnica de aprendizaje no supervisado cuyo objetivo es encontrar una estructura o patrones en una colección de datos sin etiquetar. El algoritmo K-means es uno de los métodos de clustering más populares y efectivos. Su lógica es agrupar los datos en un número predefinido de clústeres (K), donde cada punto de datos pertenece al clúster cuyo centro (centroide) está más cerca (Bagnato, 2018). El algoritmo funciona de manera iterativa, reajustando los centroides en cada paso hasta que las asignaciones de los clústeres se estabilizan. Para nuestro estudio, K-means es la técnica para segmentar municipios con comportamientos similares del Dengue, revelando agrupaciones que no son evidentes a simple vista.

### **Python como Herramienta para el Análisis de Datos**

La elección de Python como lenguaje de programación para el componente de Machine Learning no es casual. Python se ha consolidado como el estándar de facto en la comunidad de ciencias de datos debido a su sintaxis sencilla y legible, su vasta comunidad de soporte y, sobre todo, su robusto ecosistema de librerías especializadas (Canales Luna, 2024). Bibliotecas como Pandas (para la manipulación de datos), Scikit-learn (para la implementación de modelos de Machine Learning como el K-means) y Matplotlib/Seaborn (para la visualización) proporcionan un entorno de trabajo potente y eficiente que permite pasar de la teoría a la aplicación de manera práctica y fluida.

## **Marco Conceptual**

El presente proyecto se fundamenta en un modelo conceptual diseñado para transformar la vigilancia epidemiológica del Dengue, pasando de un paradigma reactivo basado en reportes de casos a un enfoque proactivo de vigilancia estratégica. Este modelo concibe la solución no como un conjunto de herramientas aisladas, sino como un sistema integrado de tres capas interdependientes que, en conjunto, convierten el dato crudo en conocimiento accionable para la toma de decisiones en salud pública.

A continuación, se describe cada una de las capas que componen este marco conceptual.

### **Capa 1 El Fundamento de los Datos – El Microdato como Unidad de Análisis**

La base de todo el sistema es el repositorio de datos del Sivigila. Conceptualmente, este no se considera simplemente una base de datos, sino la memoria epidemiológica del Caribe y del país. Cada fila o registro, referido aquí como un microdato, trasciende su naturaleza numérica para representar la historia de una persona afectada por el Dengue. Este enfoque humanista es fundamental, pues orienta el análisis a entender las realidades detrás de las cifras. La riqueza de este fundamento radica en su granularidad, permitiendo un análisis detallado a nivel de municipios, semana epidemiológica y perfil demográfico, lo cual es esencial para una vigilancia de precisión.

### **Capa 2 Modelo Analítico – Del Dato al Patrón Oculto**

Esta capa constituye el núcleo metodológico del proyecto y representa el puente entre el dato en su estado bruto y el descubrimiento de patrones significativos. El motor analítico se concibe como un proceso secuencial y sinérgico:

1. Análisis Descriptivo y Exploratorio (EDA): funciona como la primera fase del diagnóstico, donde se “interrogan” a los datos para comprender su estructura, calidad y las

tendencias más evidentes. Es el punto de partida para formular hipótesis sobre comportamientos de la enfermedad.

Modelo de Segmentación (K-means): Es el componente de descubrimiento del motor. Mientras el análisis descriptivo muestra “que paso”, el clustering no supervisado busca responder “donde están ocurriendo patrones similares”. Al aplicar K-means, se trasciende el análisis geográfico tradicional para generar una taxonomía de riesgo epidemiológico, agrupando municipios no por su frontera, sino por la similitud de su “huella” epidémica a lo largo del tiempo. Este tipo de segmentación es clave para la optimizar intervenciones, tal como sugieren estudios sobre la aplicación de ciencias de datos en la gestión pública (Vallejos et al., 2024).

El motor analítico, que combina Python y sus librerías, es el encargado de procesar la memoria epidemiológica para extraer de ellas no solo resúmenes, sino también inteligencia predictiva y de segmentación.

### **Capa 3 La interfaz de Conocimiento – La Democratización de la Decisión**

Un análisis, por más profundo que sea, pierde su valor si no es accesible para quienes deben tomar decisiones. Esta tercera etapa conceptualiza la herramienta de Power BI no como un simple visualizador, sino como un vehículo para democratización del conocimiento. Su función es traducir los complejos hallazgos del motor analítico a un lenguaje visual, interactivo e intuitivo.

A través de mapas de georreferenciación, gráficos dinámicos y filtros interactivos, esta interfaz permite que un epidemiólogo o un secretario de salud pueda explorar los datos sin necesidad de ser experto en programación. Se convierte en un “laboratorio virtual” donde se pueden validar hipótesis y monitorear los clústeres de riesgo. La implementación de soluciones de inteligencia de negocio es un paso fundamental hacia una gobernanza basada en la evidencia,

donde las decisiones estratégicas se sustentan en un análisis riguroso y accesible de los datos (Gan, 2025).

En conclusión, este marco conceptual integra el microdato del Sivigila, un motor analítico avanzado y una interfaz de conocimiento interactiva para materializar el concepto de vigilancia estratégica. Este modelo no solo busca describir el pasado, sino generar una comprensión profunda de los patrones de riesgo para informar un futuro con intervenciones más inteligentes y eficaces contra el Dengue en la región Caribe.

## **Marco Normativo**

El desarrollo de un proyecto que involucra el análisis de datos de salud pública, incluso cuando estos son anónimos, deben estar enmarcados y respaldado por un conjunto de normas que garantizan tanto el derecho fundamental a la salud de la población como la protección de la información personal. Este capítulo describe el andamiaje legal y ético sobre el cual se sustenta esta investigación.

### **Derecho a la Salud y Deberes del Estado**

La base de toda acción en salud pública en Colombia reside en la Constitución Política de 1991. En su artículo 49, se consagra la salud como un derecho fundamental y un servicio público a cargo del Estado. Este mandato constitucional implica que el Estado no solo debe garantizar la atención, sino también “organizar, dirigir y reglamentar la prestación del servicio de salud a los habitantes conforme a los principios de eficiencia, universalidad y solidaridad” (Asamblea Nacional Constituyente, 1991). La vigencia de epidemiológica, como la que realiza el Sivigila, es una manifestación directa de este deber del Estado, pues es la herramienta principal para monitorear el estado de salud de la población y dirigir las acciones de prevención y control.

### **Ley de Protección de Datos Personales (Habeas Data)**

El manejo de información es el aspecto más sensible de nuestro proyecto. La norma rectora en esta materia es la Ley Estatutaria 1581 de 2012, por lo cual se dictan disposiciones generales para la protección de datos personales. Esta ley establece un marco de principios para el tratamiento de datos, incluyen la legalidad, finalidad, libertad, veracidad, transparencia, seguridad y confidencialidad (Congreso de Colombia, 2012).

Es crucial destacar que la ley, en su artículo 3, define los datos sensibles como “aquellos que afectan la intimidad del Titular o cuyo uso indebido puede generar su discriminación”,

mencionado explícitamente los datos relativos de la salud. Nuestro proyecto maneja datos de salud; sin embargo, opera en total conformidad con la ley por dos razones fundamentales:

1. Finalidad: El tratamiento de los datos tiene una finalidad de interés público, específicamente la investigación científica y la vigilancia en la salud pública, propósitos contemplados y permitidos por la normativa.

2. Anonimización: El proyecto utiliza una base de datos secundaria proporcionada por el Instituto Nacional de Salud, en la cual la información ha sido previamente anonimizada. Esto significa que los datos que identifican de manera directa e inequívoca a una persona han sido disociados, garantizando la protección de la identidad y la intimidad de los individuos.

### **Marco Regulatorio de la Vigilancia en Salud Pública**

El sistema que nos provee los datos, Sivigila, no es una entidad aislada, sino que está regulado por un marco normativo específico. El Decreto 780 de 2016, conocido como Decreto Único Reglamentado del Sector Salud y Protección Social, compila y reorganiza todas las normas del sector. En su libro 2, parte 8, título 8, capítulo 1, este decreto define y estructura el Sistema de Vigilancia en Salud Pública (Ministerio de Salud y Protección Social, 2016).

Este decreto establece las responsabilidades de los diferentes actores del sistema de salud (municipios, departamentos, EPS, IPS y el Instituto Nacional de Salud) en la recolección, notificación y análisis de los datos de los Eventos de Interés en Salud Pública, como el Dengue. Por lo tanto, el uso de los datos del Sivigila para este proyecto nos es solo legítimo, sino que se alinea con el propósito mismo para el cual es sistema fue creado: generar conocimiento útil para la toma de decisiones que protejan la salud colectiva.

## **Normativas Éticas en Investigación en Salud**

Finalmente, toda investigación que involucra información de seres humanos debe adherirse a principios éticos. En Colombia, la norma de referencia es la Resolución 8430 de 1993, por la cual se establecen las normas científicas, técnicas y administrativas para la investigación en salud. En su artículo 11, esta resolución clasifica las investigaciones según su nivel de riesgo (Ministerio De Salud, 1993).

De acuerdo con esta clasificación, presente proyecto como una “investigación sin riesgo”. Esta categoría aplica a “estudios que emplean técnicas y métodos de investigación documental retrospectivos y aquellos en los que no se realiza ninguna intervención o modificación intencionada de las variables biológicas, filosóficas, psicológicas o sociales de los individuos que participan en el estudio”. Al ser análisis retrospectivo de una base de datos secundaria y anonimizada, nuestro proyecto no expone a ninguna persona a riesgo físico o de privacidad, cumpliendo así con los más altos estándares éticos.

## **Metodología**

El proyecto se desarrollará bajo el método cuantitativo y analítico, adoptando como marco de trabajo la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining). Esta metodología es un estándar de facto en la industria de la ciencia de datos y se caracteriza por su enfoque estructurado e interactivo, lo que hace ideal para un proyecto aplicado como este. Las fases de CRISP-DM se adaptarán a nuestro contexto de la siguiente manera:

### **Comprensión del Negocio (o del Problema)**

Esta fase inicial sienta las bases teóricas y operativas del proyecto. Además de la identificación de las necesidades, la justificación y la definición de objetivos, los cuales se desarrollaron ampliamente en los capítulos introductorios del documento. Adicional, se llevó a cabo una revisión sistemática de la literatura para fundamentar la estrategia analítica.

El objetivo de la revisión fue identificar antecedentes sobre el uso de técnicas de minería de datos y aprendizaje autónomo aplicada a la vigilancia epidemiológicas del Dengue, así como validar las variables clínicas y demográficas críticas para la caracterización del evento en el contexto colombiano.

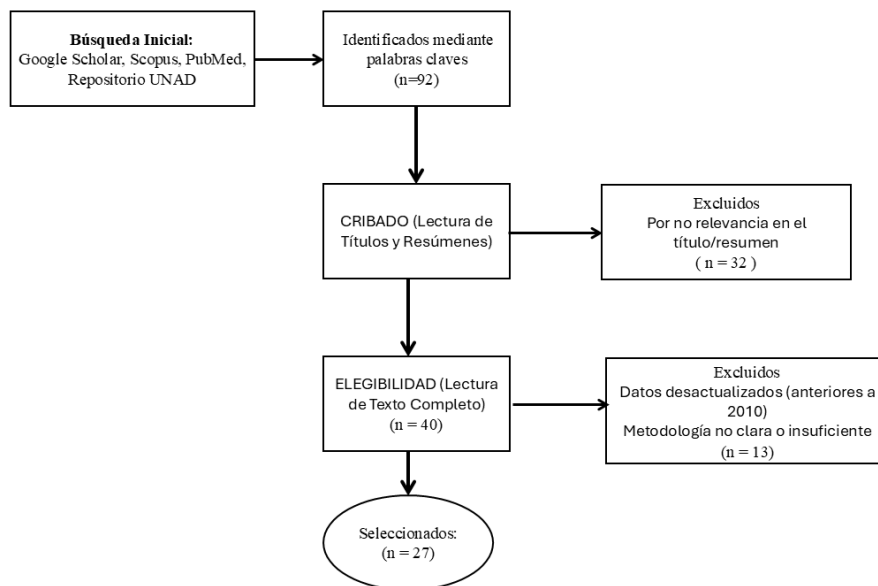
### ***Procedimiento de Selección de Fuentes***

Para garantizar la validez teórica, se siguió un flujo e búsqueda y selección estructurado bajo los lineamientos PRISMA. La búsqueda se realizó en bases de datos académicas (Google Scholar, Scopus, PubMed y Repositorio Institucional UNAD) utilizando ecuaciones de búsqueda que combinan términos como "Dengue Colombia", "Clustering Epidemiology", "Machine Learning Public Health", "Vigilancia en Salud Pública".

El Proceso de cribado y elegibilidad, que partió de 92 registros iniciales hasta consolidar las 27 referencias finales que sustentan este trabajo, se detallan en la Figura 1.

**Figura 1**

*Diagrama de Flujo del Proceso de Selección de Literatura (Metodología PRISMA)*



## Comprensión de los Datos

En esta fase se realizó la recolección inicial de los datos y se procedió con una exploración para familiarizarse con la información, identificar problemas de calidad y descubrir primeros insights.

- Fuente de Datos: La fuente primaria son los microdatos anonimizados del SIVIGILA (código 210), proporcionado por el INS, para el periodo 2010-2024.
- Consolidación Inicial: Se unificaron los archivos anuales, los cuales se nombran “Datos\_año\_210.xlsx”, en un único conjunto de datos que se llamara “Dengue\_Consolidado\_2010\_2024.csv”.

- **Filtrado Geográfico:** Se filtro el dataset maestro para incluir únicamente los registros de la Región Caribe, definidos por los códigos DANE departamentales 08, 13, 20, 23, 44, 47, 70 y 88.

### **Preparación de los Datos**

Es una de las fases más intensivas y el objetivo es construir el conjunto de datos final que será utilizado para el modelado y la visualización. Se realizaron las siguientes tareas:

- **Selección de Variables:** De las 75 variables originales, se seleccionó un subconjunto relevante para los objetivos del estudio (ver Tabla 1)
- **Limpieza de Datos:** Se estandarizaron valores categóricos, se manejan fechas convirtiéndolas a formato datetime y se gestionaron valores nulos.
- **Construcción de Variables:** Se crearon nuevas variables necesarias para el análisis, destacando la unificación de la edad (variable EDAD\_ANOS) a partir de la columna EDAD y UNI\_MED.

**Tabla 1***Variables Seleccionadas para el Análisis*

Variable	Descripción	Fase de Uso
FEC_NOT, INI_SIN	Fechas de notificación e inicio de síntomas	Caracterización, Dashboard
ANO, SEMANA	Variables de tiempo para análisis de tendencia	Caracterización, Dashboard
COD_DPTO_R, COD_MUN_R	Códigos DANE de residencia	Caracterización, Clustering, Dashboard
Departamento_residencia	Nombre del Departamento	Caracterización, Dashboard
Municipio_residencia	Nombre del Municipio	Caracterización, Dashboard
AREA	Área de residencia (Urbano/Rural)	Caracterización, Dashboard
EDAD_ANOS	Edad de la paciente unificada en años	Caracterización, Clustering, Dashboard
SEXO	Sexo del paciente	Caracterización, Dashboard
PAC_HOS	Indicador de hospitalización	Caracterización, Clustering, Dashboard
CON_FIN	Condición final (Vivo/Muerto)	Caracterización, Clustering, Dashboard
ESTADO_CASO_AGRUPADO	Estado final del caso (Confirmado, Probable)	Caracterización, Dashboard

**Modelado**

En esta fase se aplicó el algoritmo de aprendizaje no supervisado K-means con el objetivo de segmentar el comportamiento epidemiológico en la Región Caribe. Se definió una

granularidad espaciotemporal de Municipio-Año, lo que permite evaluar la evolución de riesgo de cada territorio a lo largo del tiempo, facilitando su visualización dinámica.

### ***Construcción de Variables***

Para caracterizar multidimensionalmente el evento, se construyen siete variables sintéticas a partir del dataset consolidado, agrupando tres dimensiones de análisis.

#### **Dimensión de Magnitud y Severidad.**

- Volumen de casos: Cantidad total de notificaciones en el municipio para el año específico.
- Proporción de Hospitalización: Porcentaje de caos que requieren manejos intrahospitalarios (indicador de gravedad clínica).
- Tasa de Letalidad: Porcentaje de fallecimientos sobre el total de casos (indicador de impacto mortal)

#### **Dimensión Demográfico y Geográfico.**

- Proporción de Casos Pediátricos (<15 años): Identificar municipios donde la carga de enfermedad afecta principalmente a la población infantil vulnerable.
- Proporción de Sexo Femenino: Variable de control para identificar patrones diferenciales de exposición.
- Proporción de Ruralidad: Porcentaje de casos reportados en centros poblados y rurales dispersos, diferenciando el perfil epidemiológico urbano vs rural.

#### **Dimensión de Vigilancia.**

- Índice de Confirmación: Proporción de casos confirmados por laboratorios o nexos epidemiológico frente a casos sospechosos/probables. Esto actúa como un proxy de la calidad de la vigilancia y la capacidad diagnóstica local.

### ***Procesamiento***

Se realizó una estandarización de los datos utilizando StandardScaler (media 0, desviación 1) para garantizar que variables con diferentes escalas contribuyan equitativamente al cálculo de las distancias del algoritmo.

### ***Evaluación***

Una vez generado el modelo de segmentación, se procedió a evaluar su capacidad técnica y su pertinencia epidemiológica.

**Evaluación Técnica.** Se validó la consistencia interna de los clústeres mediante el análisis de la varianza intra-clase y el Coeficiente de Silueta, confirmando que la configuración de  $K=4$  maximiza la distribución entre los grupos formados.

**Evaluación de Negocio.** Se interpretaron los centroides de cada clúster para asignarles una etiqueta clínica y social. Se verificó que los grupos resultantes tuvieran sentido desde la perspectiva de la salud pública, diferenciando claramente perfiles de magnitud (volúmenes de casos) de perfiles de severidad (letalidad), lo cual valida la utilidad del modelo para la toma de decisiones diferenciadas.

### ***Despliegue (Implementación)***

Los resultados tanto del análisis descriptivo (fase 2) como del modelado (fase 4) serán integrados y visualizados en la herramienta de inteligencia de negocios. Se diseñará y construirá el tablero de control final en Power BI, que concluye el entregable principal del proyecto.

### **Tipo de Estudio**

El presente trabajo se clasifica como un proyecto aplicado con un diseño de investigación cuantitativo, descriptivo y retrospectivo.

- **Proyecto Aplicado:** Porque su objetivo general no es solo generar conocimiento teórico, sino desarrollar una solución tecnológica tangible y funcional (tablero de Power BI) que responde a una necesidad específica del sector salud.
- **Cuantitativo:** Porque se basa en análisis numéricos y estadísticos de un gran volumen de datos para identificar patrones y relaciones.
- **Descriptivo:** En línea con el primer objetivo específico, el estudio busca caracterizar y describir en detalle el comportamiento del Dengue en la región Caribe a lo largo del tiempo, espacio y población.
- **Retrospectivo:** Porque utiliza como fuente de información datos históricos ya existentes, correspondientes al periodo 2010-2024, sin intervenir en las variables ni hacer seguimiento a los sujetos en el futuro.

### **Recolección de Datos**

La recolección de la información para este proyecto se basa en una fuente de datos secundaria.

- **Fuente de Información:** La base de datos oficial de caso de Dengue notificados al Sistema Nacional de Vigilancia en Salud Pública (Sivigila).
- **Entidad Custodia:** El Instituto Nacional de Salud (INS) de Colombia.
- **Proceso de Obtención:** Los datos serán obtenidos a través de las fuentes disponibles del portal web del SIVIGILA (Sivigila, 2025), en la sesión de búsqueda de microdatos. La información se puede consultar y extraer año por año, lo que representa un proceso de extracción y consolidación manuales o automatizada por medio de un script de Python. La entidad proporciona una base de datos anonimizada, en cumplimiento con la Ley

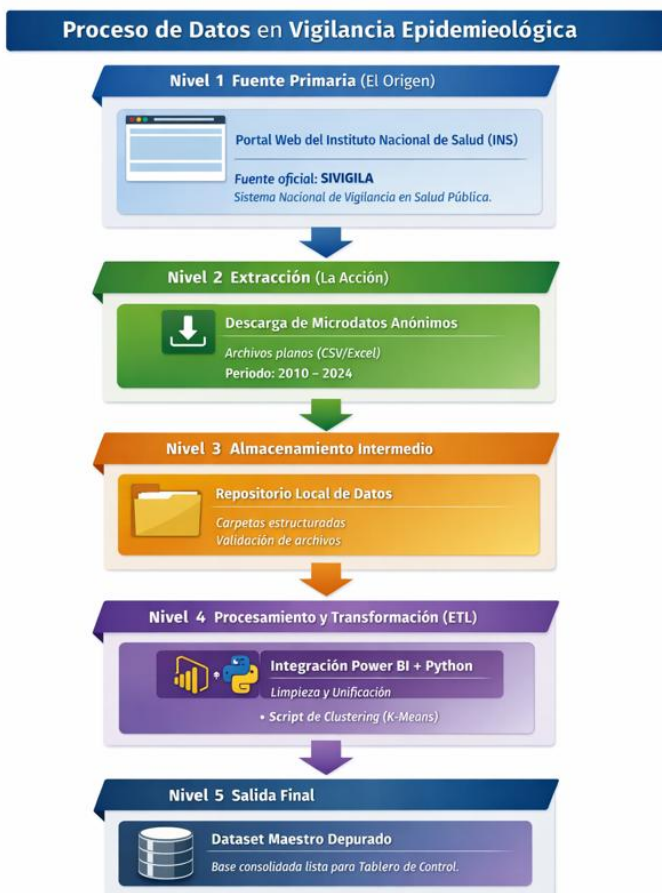
1581 de 2012 de Protección de Datos Personales, garantizando la confidencialidad de la identidad de los pacientes.

- Universo de Datos: La base de datos consolidada incluirá todos los registros de casos de Dengue notificados en los departamentos que componen la región Caribe de Colombia (Atlántico, Bolívar, Cesar, Córdoba, La Guajira, Magdalena y Sucre) durante el periodo comprometido entre la semana epidemiológica 1 de 2010 y la semana epidemiológica 52 de 2024.

- Variables de Interés: Aunque el dataset completo contiene numerosas variables, el análisis se centrará en:

- Variables Geográficas: Departamentos y municipios de ocurrencia del caso.
- Variables Temporales: Año, semana epidemiológica y fecha de inicio de síntomas.
- Variables Demográficas: Edad y sexo del paciente.
- Variables Clínicas: Clasificación final del caso (Dengue, Dengue Grave) y si requirió hospitalización.

Finalmente, se puede observar en la Figura 2, el procedimiento de recolección y flujo de los datos utilizados en el proyecto.

**Figura 2***Proceso de Datos en la Vigilancia Epidemiológica*

*Nota. Tomada de Imagen generada mediante ChatGPT/DALL·E, OpenAI (2025).*

## Resultados

Este capítulo presenta los hallazgos principales del proyecto, derivado de la aplicación de la metodología CRISP-DM. Los resultados se han estructurado en función de los objetivos específicos planteados, comenzando con la caracterización descriptiva del Dengue y avanzando hacia análisis de patrones mediante modelos.

Los análisis presentados a continuación -temporales, espaciales, demográficos y de gravedad- constituyen la base analítica sobre la cual se diseña y construye el tablero de control en Power BI. Cada sección documenta los descubrimientos claves obtenidos a partir del conjunto de datos limpios (2010-2024), proporcionando una radiografía completa del comportamiento de la enfermedad en la Región Caribe.

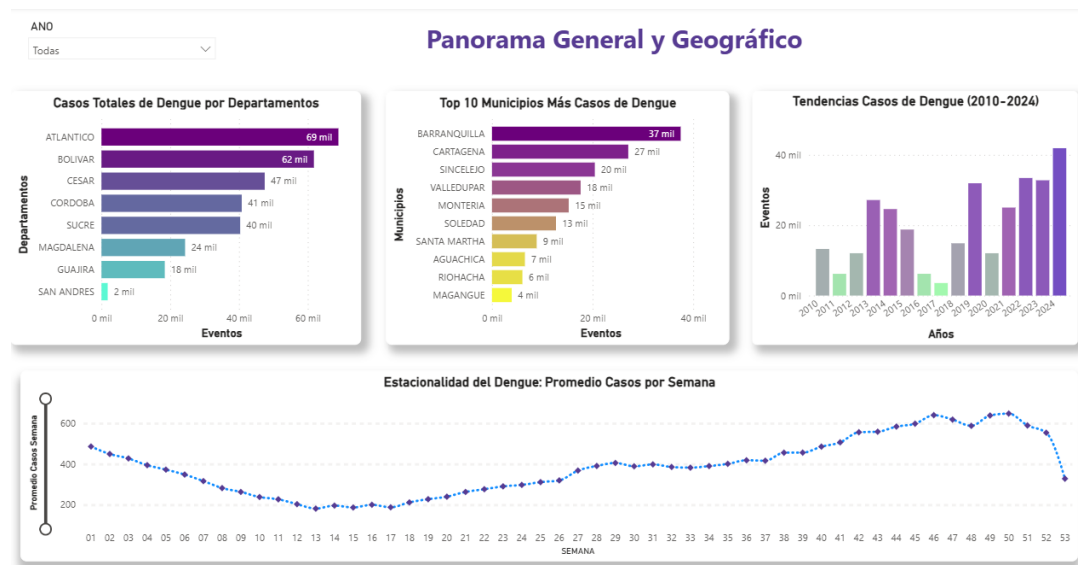
### Caracterización Epidemiológica

El primer paso para el diseño del tablero de control consiste en un análisis descriptivo de la data histórica. Esta caracterización se divide en análisis temporal (¿cuándo?), espacial (¿dónde?), demográfico (¿quiénes?) y de gravedad (¿cómo?).

En la primera hoja del tablero se inicia graficando los resultados de los análisis temporales (¿Cuándo?) y los análisis espaciales (¿Dónde?), se diseñan graficas que muestran los comportamientos por años y por semanas epidemiológicas, al igual que se graficas que muestran los casos y comportamientos en los diferentes departamentos y municipios de la Región Caribe, como se evidencia en la **Figura 3**.

**Figura 3**

*Panorama General y Geográfico (Dengue)*

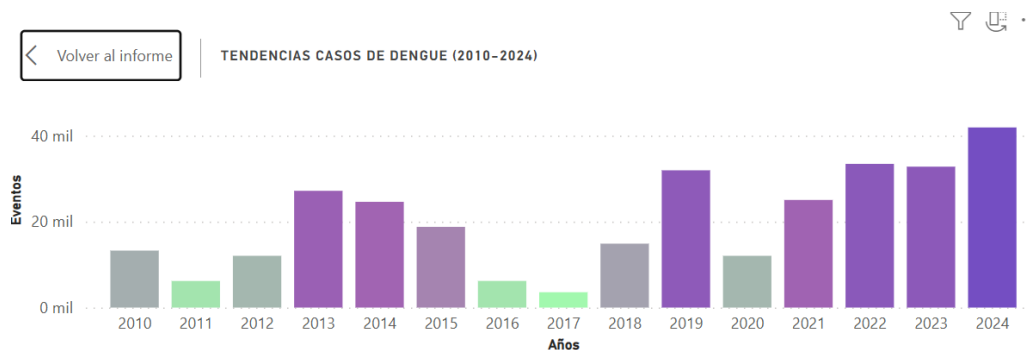


**Análisis Temporal (¿Cuándo Ocurre?)**

El comportamiento del Dengue en la Región Caribe durante el periodo 2010-2024 revela un patrón cíclico y una alta sensibilidad a fenómenos climáticos y eventos de salud pública, como se evidencia en la **Figura 4**.

**Figura 4**

*Tendencia de Casos de Dengue en la Región Caribe (2010-2024)*



## Compartimientos Endemo-Epidémicos (Picos y Valles)

1. Años de Alta Incidencia (Picos): El gráfico muestra un comportamiento endemo-epidémico. Los principales picos epidémicos se identifican claramente en los años 2013, 2014, 2019 y los brotes históricos de 2023-2024.

- *Asociados con el Fenómeno de El Niño:* Estos picos epidémicos tienen una correlación directa y documentada con los ciclos del Fenómeno de El Niño. Este fenómeno climático provoca un aumento en las temperaturas y escasez de lluvias en la Región Caribe. Paradójicamente, la falta de lluvias obliga a la población a almacenar agua en tanques y albercas, creando criaderos artificiales ideales para el mosquito *Aedes aegypti*, lo que dispara la transmisión. Los años 2013, 2014, 2019 y 2023-2024 fueron años con presencia notable de El Niño, lo que explica científicamente los brotes (Ministerio de Salud, 2024).

2. Años de Baja Incidencia (Valles): Se observan valles de baja transmisión, destacando dos periodos claves:

- *Periodo 2016 – 2018:* El año 2016 no fue un pico, sino el inicio de un periodo de baja incidencia que continuó en el 2017 y 2018. La investigación epidemiológica sugiere dos factores claves para esto:

- *Epidemia del Zika (2015-2016):* La emergencia del virus del Zika, también transmitido por el *Aedes Aegypti*, generó una alerta sanitaria masiva. Esto pudo causar un sesgo de clasificación y, a su vez, las intensas campañas de control vectorial (fumigaciones, eliminación de criaderos) contra el Zika impactaron también en la población de mosquitos transmisores de Dengue reduciendo los casos.

- *Periodo Inter-epidémico*: Estos valles son esperados después de grandes epidemias como la del 2014, ya que la población adquiere inmunidad temporal a los serotipos circulantes.

- *Periodo 2020 – 2021 (La Caída de la Pandemia)*: Esta caída no se debe a una ausencia de Dengue. Este hallazgo es un artefacto directo de la pandemia del COVID-19:

- *Sub-registro masivo*: El Sistema de Vigilancia de Salud Pública (SIVIGILA) colapso operativamente, priorizando el reporte de COVID-19 sobre todos los demás eventos.

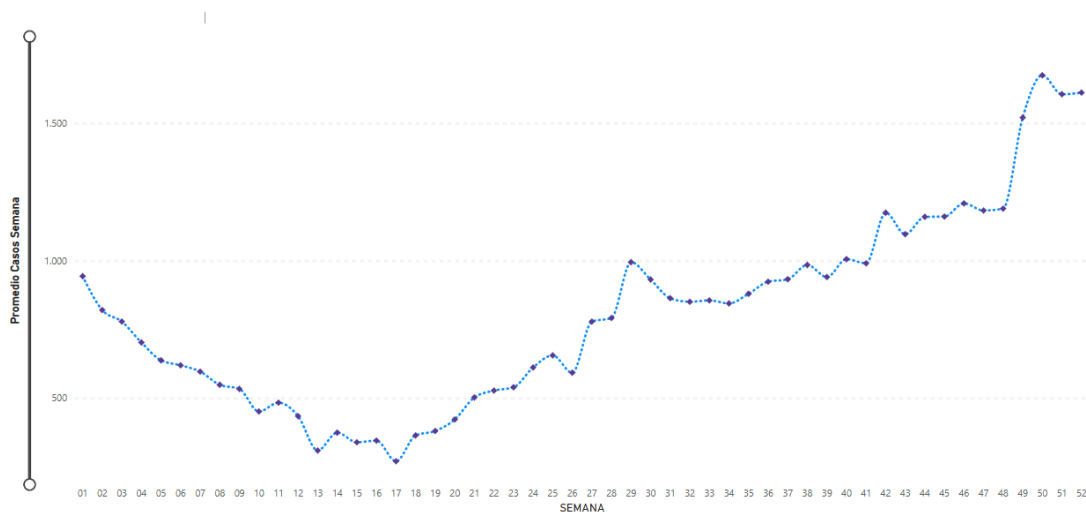
- *Cambio del Comportamiento*: Las cuarentenas y el miedo al contagio hicieron que la población con síndromes febriles (como el Dengue) no asistieran a los centros de salud, quedando fuera del registro.

El dramático repunte en el 2023-2024, tras el fin de la emergencia del COVID-19 y la llegada de un nuevo Fenómeno de El Niño, confirma que el virus seguirá circulando y encontró una población altamente susceptible, generando el pico más alto de la década.

### Estacionalidad (Canal Endémico)

#### Figura 5

*Estacionalidad del Dengue: Promedio de Casos por Semana Epidemiológica para el Año 2024*

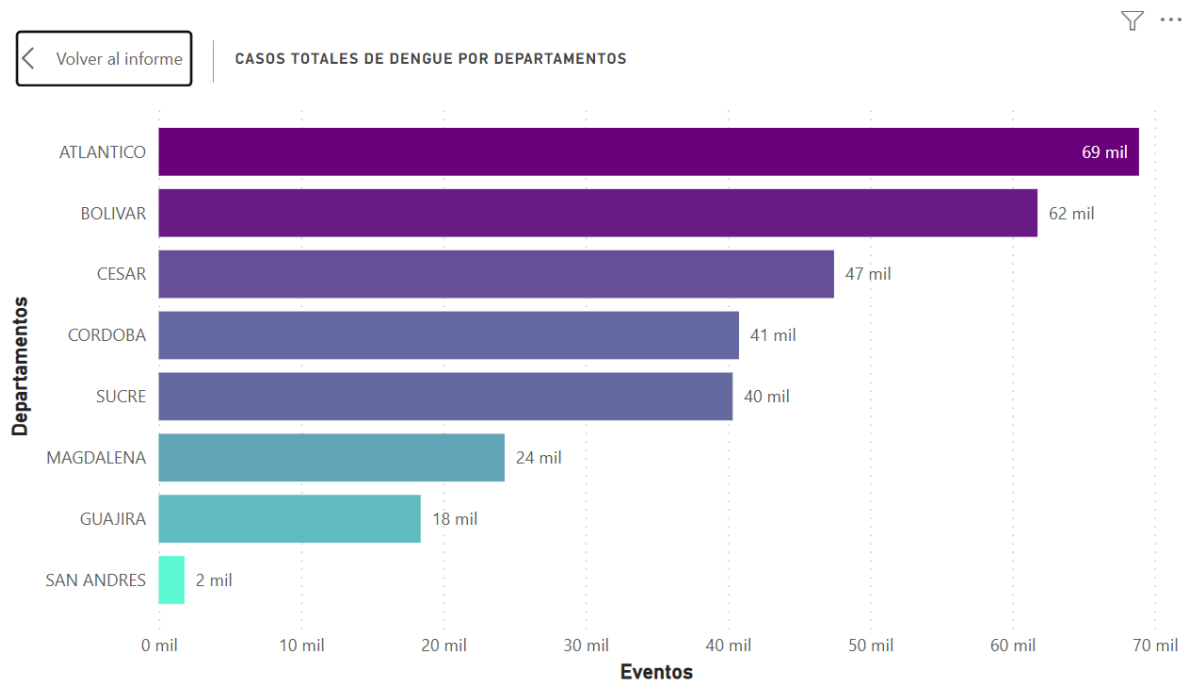


El análisis de la estacionalidad semanal promedio del año 2024, el cual es el año con mayor cantidad de casos reportados (Ver Figura 5) muestra un descenso de la incidencia desde la semana 1 hasta la semana 13, seguido por niveles bajos hasta la semana 18. A continuación ocurre un aumento pronunciado entre las semanas 18 y 30, con una meseta entre la semana 30 y 38. Se observan picos agudos en las semanas 46 y 50 siendo los picos más altos en esta semana, tras lo cual la incidencia tiende a disminuir. Este patrón sugiere la presencia de dos ventanas temporales de mayor transmisión potencialmente asociadas con variaciones de climáticas, cambios en la dinámica de vectores o en la circulación de serotipos.

### ***Análisis Espacial (¿Dónde Ocurre?)***

El Dengue es una enfermedad focalizada. El análisis espacial revela que la carga de la enfermedad no se distribuye de manera homogénea.

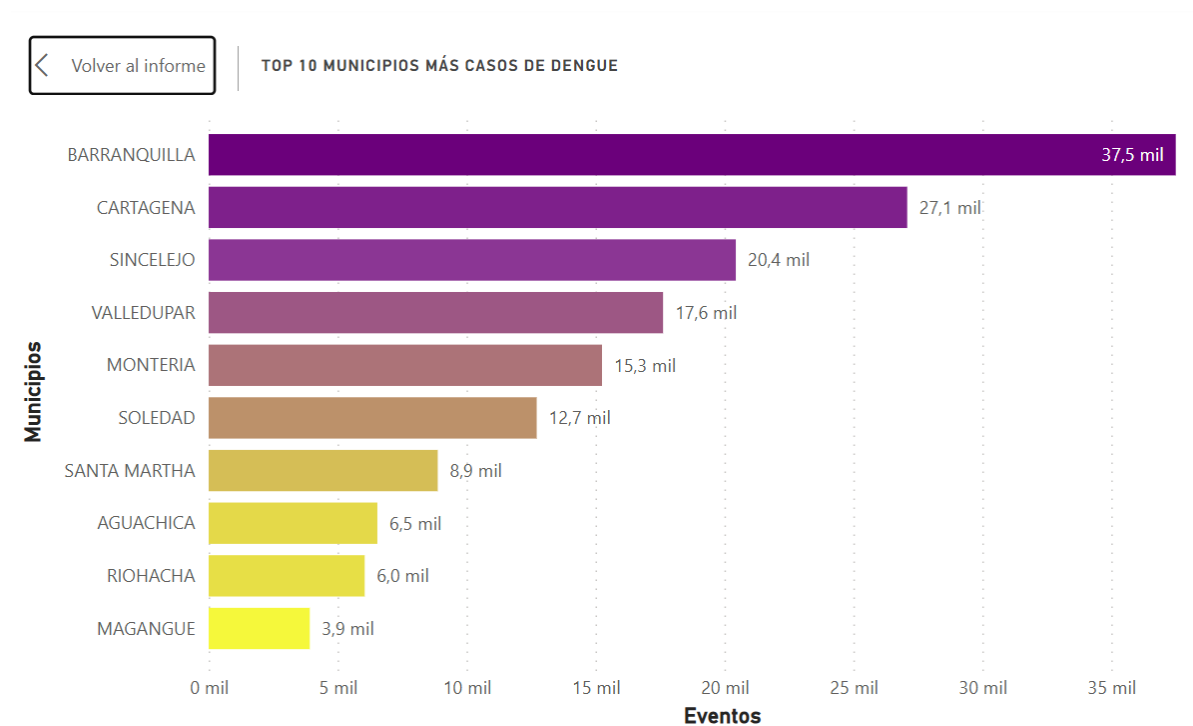
**Distribución por Departamento.** Al ver los casos del periodo 2010-2024, se observa una alta concentración de los departamentos con mayor densidad de poblacional y principales centros urbanos (Ver Figura 6). Atlántico y Bolívar lideran la casuística regional, seguidos por Cesar, Córdoba y Sucre.

**Figura 6***Casos Totales de Dengue por Departamento (2010-2024)*

**Identificación de “Hotspots” Municipales.** El análisis a nivel de departamental es insuficiente, ya que el Dengue es una enfermedad intrínsecamente urbana. El Análisis del Top 10 de Municipios (Ver Figura 7) confirma esta hipótesis de manera contundente.

## Figura 7

### Top 10 Municipios con Mayor Número de Casos (2010-2024)

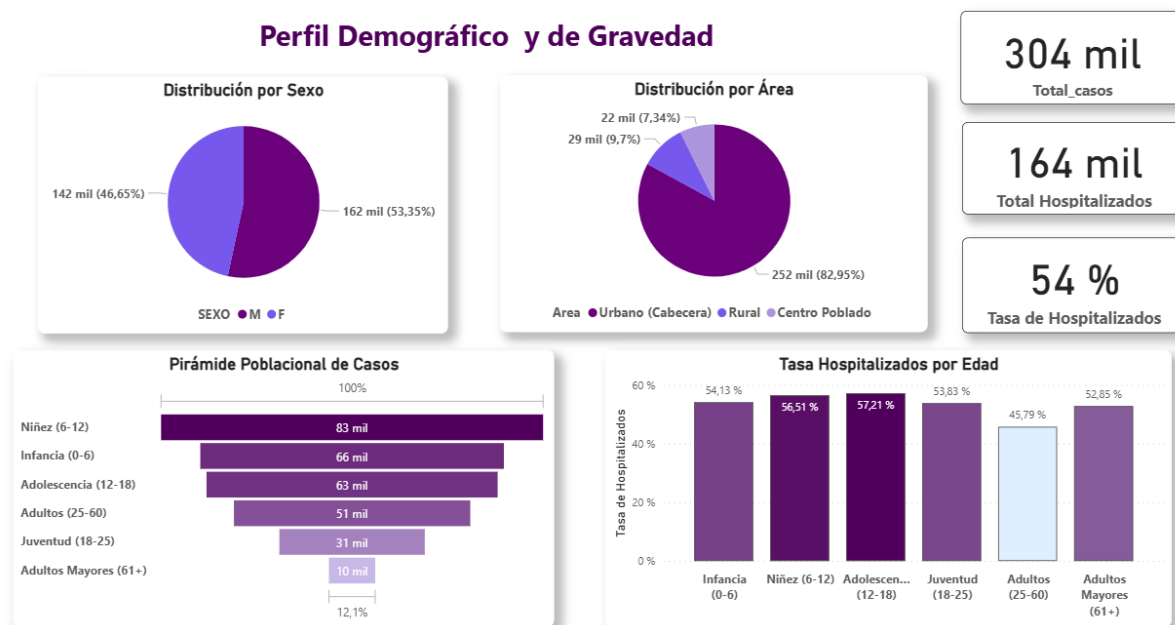


Se observa que las capitales departamentales (Barranquilla, Cartagena, Sincelejo, Valledupar, Montería) acumulan la gran mayoría de los casos, Barranquilla y Cartagena se consolidan como los principales hotspots de las enfermedades en la región.

Este hallazgo justifica la necesidad del Objetivo Especifico 2 (Clustering), ya que es crucial entender los patrones a nivel municipal para que las inversiones de salud pública (fumigaciones, control de criaderos) sean costo-efectivas y se focalicen donde realmente se necesitan.

### ***Análisis Demográfico y de Gravedad (¿Quién y Como?)***

En esta sección nos enfocamos en el perfil de los pacientes y la severidad de la enfermedad en la Región Caribe durante el periodo 2010-2024 (Ver Figura 8)

**Figura 8***Perfil Demográfico y de Gravedad*

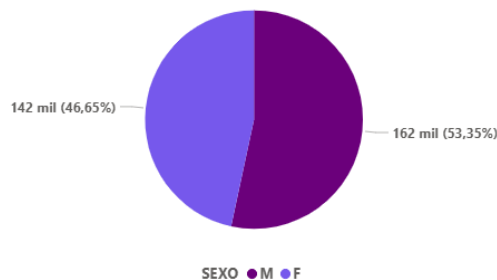
El análisis del perfil demográfico revela tres hallazgos claves:

1. *Distribución por Sexo*: El Dengue en la Region Caribe no presente una diferencia significativa por sexo, como observamos en la Figura 9, la distribución de los casos equitativa, afectando a hombres 53% y mujeres en un 47% aproximadamente en proporciones con una diferencia de 6 puntos.

## Figura 9

### *Distribución de Casos por Sexo (2010-2024)*

DISTRIBUCIÓN POR SEXO

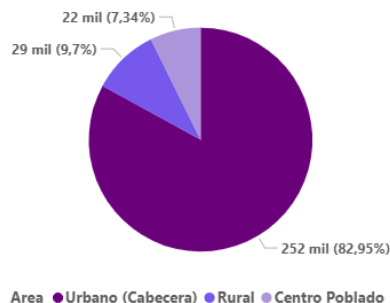


2. *Carácter Urbano de la Enfermedad:* En contraste con el sexo, el análisis por área de residencia (Ver Figura 10) es contundente. El Dengue es un fenómeno abrumadoramente urbano: más del 82% de los casos reportados provienen del sector urbano. Esto reafirma que el *Aedes aegypti* es un vector domestico adaptados a los entornos urbanos, donde la densidad poblacional y las soluciones de almacenamiento de aguas (albercas, tanques) facilitan su reproducción.

## Figura 10

### *Distribución de Casos por Área de Residencia*

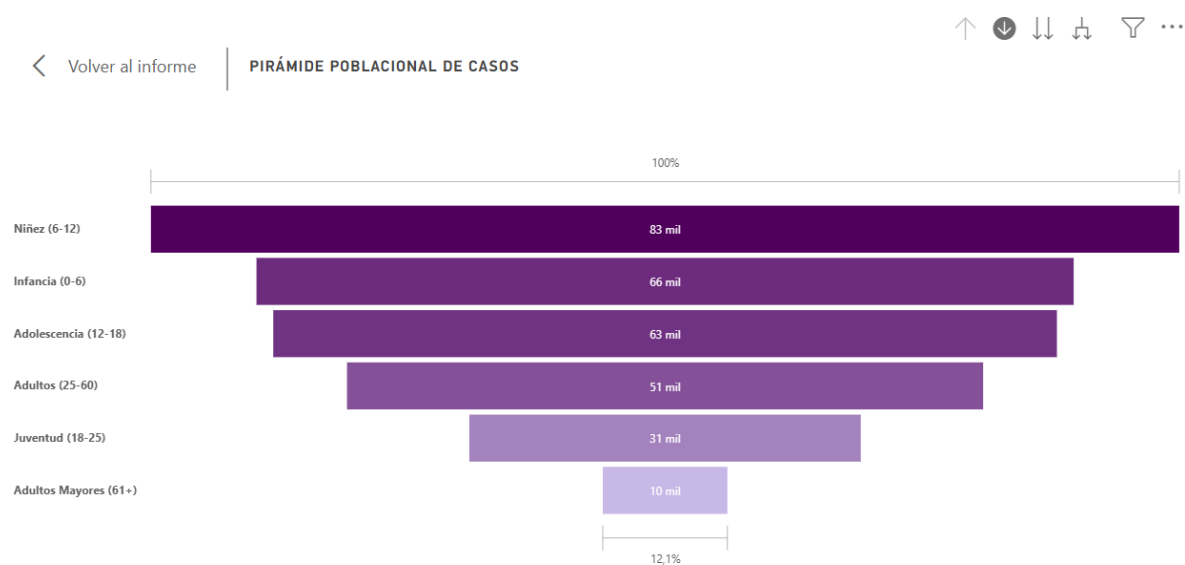
DISTRIBUCIÓN POR ÁREA



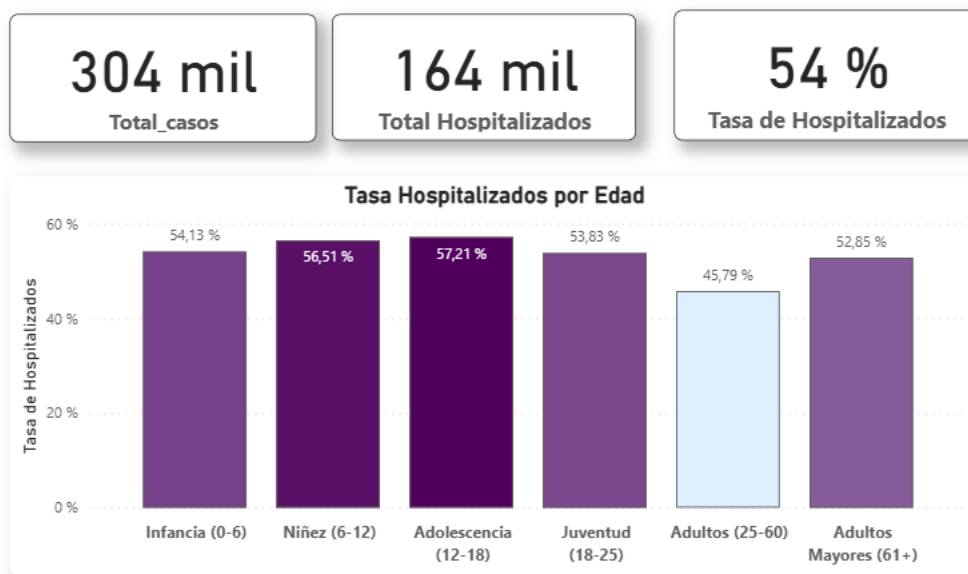
3. *Grupo de Edad más Afectado:* El análisis de la pirámide poblacional (Ver Figura 11) es uno de los hallazgos más relevantes. Si bien el Dengue afecta a todos los grupos de edad, la mayor carga de enfermedad se concentra en la población de infancia, niñez y adolescencia. Los grupos de edades entre los 0 y 18 años son los que presentan el mayor número de casos, adicional, también observamos que en los casos del grupo de edad de los adolescentes, el sexo masculino presenta más casos que el sexo contrario. Todos estos patrones antes vistos tienen un impacto socioeconómico directo, afectando a la población en edad escolar e inicio de la etapa productiva laboralmente.

### Figura 11

#### *Pirámide Poblacional de Casos de Dengue (2010-2024)*



**Análisis de gravedad:** El análisis de severidad (Ver Figura 12) nos permite cuantificar el impacto clínico del Dengue en la región.

**Figura 12***Indicadores de Gravedad (KPIs)*

Para evaluar la carga real que el evento impone sobre el sistema de salud, se analizó la proporción de casos que requirieron manejo intrahospitalario durante el periodo del estudio. Se registraron un total de 163.876 pacientes hospitalizados, lo que representa el 54% de los casos reportados, exigiendo una demanda significativa de camas y recursos asistenciales.

Sin embargo, al desagregar este indicador por curso de la vida (Ver Figura 12), se rompe la hipótesis de un riesgo homogéneo y se revela un patrón de vulnerabilidad focalizado en la población en edad escolar y adolescente:

- *Máxima Severidad en Escolares y Adolescentes:* Contrario a lo esperado en otras patologías, los grupos de 12 a 18 años (57,21%) y 6 a 12 años (65.51%) presenta la tasa de hospitalización mas alta de toda la serie, superando significativamente la tasa general de toda la regional. Esto sugiere que la respuesta inmunológica en estos grupos etarios, posiblemente

asociadas a segundas infecciones, tienden a desencadenar con mayor frecuencia cuadros que requieren vigilancia clínica estricta.

- *Alta vulnerabilidad en Primera Infancia:* El grupo de 0 a 6 años mantiene una tasa elevada del 54.13% alineada con la tasa regional general. Aunque no es el pico mas alto, representa un desafío crítico dado que el manejo de líquidos en pacientes pediátricos de corta edad requiere una precisión clínica mayor para evitar complicaciones.
- *Resiliencia en la Adulthood:* El grupo de 25 a 60 años registran la tasa más baja de hospitalización (45,79%). Este descenso notable de casi 12 puntos porcentuales por debajo de los adolescentes, indica que la población adulta económicamente activa tiende a desarrollar cuadros menos severos o a ser manejada de forma ambulatoria con mayor éxito.
- *Repunte en Adultos Mayores:* Finalmente, se observa un nuevo incremento en la tasa de hospitalización para los mayores de 60 años (52,85%). Si bien no alcanza los niveles de la población adolescente, este repunte alerta sobre la necesidad de internación preventiva en este grupo debido a la presencia de comorbilidades que pueden descompensarse por la infección viral.

En general, el perfil de gravedad del Dengue en la Región Caribe es predominantemente pediátrico y adolescente. La estrategia de preparación hospitalaria debe priorizar la disponibilidad de camas y unidades de rehidratación para menores de 18 años, quienes constituyen el grueso de la demanda de alta complejidad, mientras que el manejo del adulto joven ofrece una ventana de oportunidad para el tratamiento ambulatorio seguro.

## Segmentación de Riesgo Municipal

Para la implementación del modelo de clustering no supervisado K-means, se realizó un proceso previo de validación para determinar el número óptimo de grupos (K) que mejor segmentan los municipios de la Región Caribe según su perfil de riesgo epidemiológico.

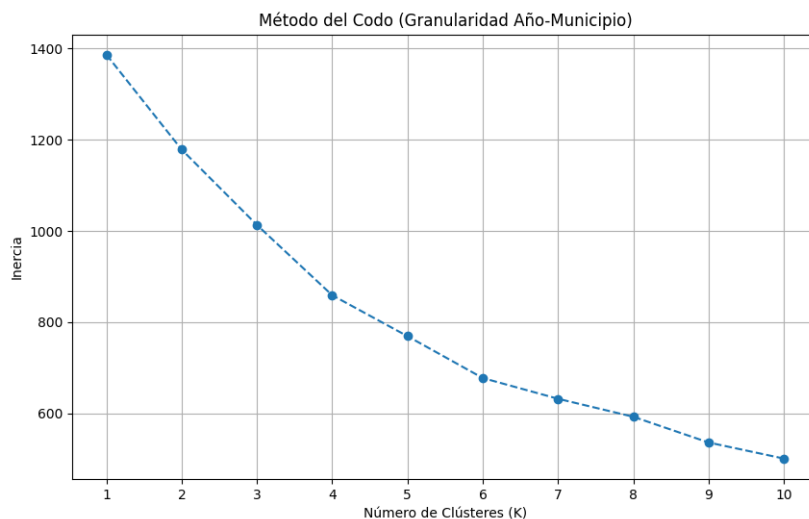
### *Determinación del Número Óptimo de Clústeres*

La selección del parámetro K no se realiza de manera arbitraria, sino que se fundamentó en la convergencia de dos métodos estadísticos de validación: el método del Codo (Elbow Method) y el análisis de Coeficiente de Silueta.

Método del Codo en primer lugar, se calculó la inercia (suma de las distancias al cuadrado dentro de los clústeres) para un rango de 1 a 10 grupos como se observa en la Figura 13, la curva presenta un punto de inflexión o codo marcado en K=4. Esto indica que, a partir de cinco grupos, la ganancia en la compactación de los clústeres disminuye marginalmente, sugiriendo que 4 es un número eficiente para la segmentación.

### **Figura 13**

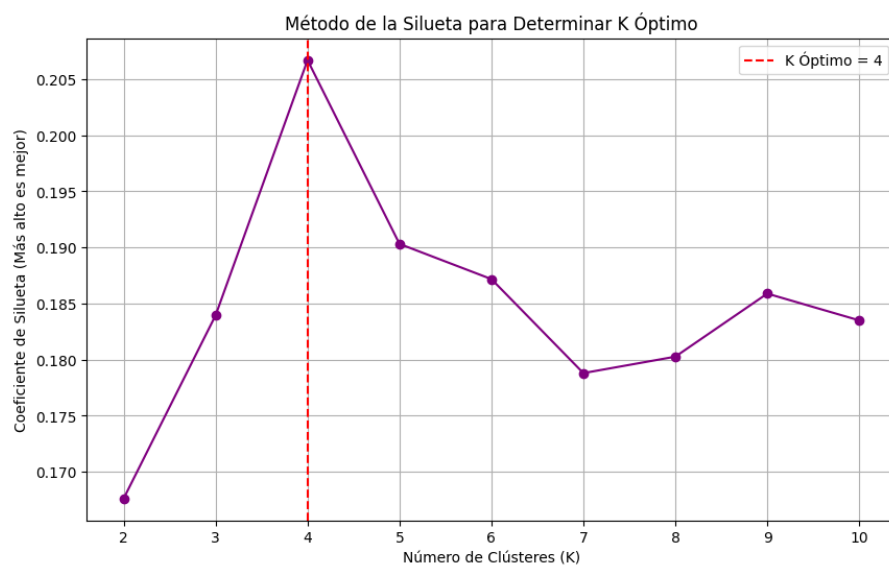
#### *Determinación de K Mediante el Método del Codo*



Método de la Silueta: para corroborar este hallazgo, se aplicó el método de la Silueta, el cual evalúa la cantidad de agrupamiento midiendo qué tan similar es un objeto a su propio clúster (cohesión) en comparación con otros clústeres (separación). Los resultados, presentados en la figura 14, muestran que el coeficiente de la silueta alcanza su valor máximo (0.2067) cuando  $K=4$ .

**Figura 14**

*Coefficiente de Silueta Según el Número de Clústeres*



Decisión del Modelado: Dado que ambos métodos convergen en  $K=4$ , se procedió a entrenar el algoritmo k-means final configurado con 4 clústeres.

### ***Perfilamiento y Caracterización de los Clústeres***

La aplicación del algoritmo K-means permitió identificar cuatro perfiles epidemiológicos distintos en la Región Caribe. La Tabla 2 resume las características promedio de cada grupo, revelando dinámicas de riesgo heterogéneas.

**Tabla 2***Tabla de Perfilamiento Clústeres (Promedios)*

Clúster	Total Casos	Prop. Hospitaliz	Tasa. Letalidad	Prop. Menor	Prop. Mujer	Prop. Ruralidad	Prop. Confi	Cant. Municip
0	21769	47.91%	0.0069%	55.69%	46.83%	6.19%	69.78%	6
1	652	58.11%	0.0005%	45.93%	43.94%	45.96%	74.19%	77
2	1048	53.79%	0.0005%	63.41%	47.52%	21.16%	69.37%	110
3	1308	63.28%	0.1039%	57.85%	46.41%	15.90%	73.40%	5

*Nota.* Resultados de los promedios de los clústeres generados

A partir de estos indicadores, se definió las siguientes tipologías:

**Clúster 0 Centros Urbanos de Alta Transmisión (N=6 Municipios).**

- *Características:* Este grupo está conformado por solo 6 municipios, los cuales son las grande capitales y distritos. Se caracteriza por una magnitud masiva de casos, con un promedio histórico mayor a 21 mil casos, lo que lo hace muy superior al resto.

- *Perfil:* Posee la menor ruralidad (6%), confirmando que son núcleos netamente urbanos. Aunque tiene la mayor carga de enfermedad, su tasa de letalidad es baja, lo que sugiere una mejor capacidad de respuesta en sus redes hospitalarias.

- *Estrategia Sugerida:* Control vectorial intensivo y masivo debido a la alta densidad del vector.

**Clúster 3 Municipios de Alta Severidad (N=5 Municipios).**

- *Características:* Este es el grupo de mayor riesgo clínico, aunque su volumen de caso es moderado con promedio de 1.308, presenta la tasa de hospitalización con la mayor proporción (63%).

- *Perfil:* Son municipios donde, proporcionalmente se hospitaliza más gente por Dengue. Esto alerta sobre posibles fallas en la detención temprana o en el manejo clínico de urgencias.

- *Estrategia Sugerida:* Auditoria urgente de la calidad de la atención clínica y reentrenamiento médico.

### **Clúster 2 Municipios de Riesgo Pediátrico Predominante (N=110 Municipios).**

- *Características:* Es el grupo más numeroso, aglomera la mayoría de los municipios.

- *Perfil:* Su característica distintiva es que presenta la mayor proporción de casos en menores de 15 años (63%). El Dengue en estos territorios es fundamentalmente una enfermedad infantil. Tiene una ruralidad media (21%).

- *Estrategia Sugerida:* Enfoque en entornos escolares y protección familiar.

### **Clúster 1 Municipios de Riesgo Rural Disperso (N=77 Municipios).**

- *Características:* Municipios con menor carga de casos (promedio 652).

- *Perfil:* Se distingue por ser lo más rurales (46% de ruralidad) y tener la menor afectación en niños (45%). Su comportamiento sugiere brotes más esporádicos y dispersos.

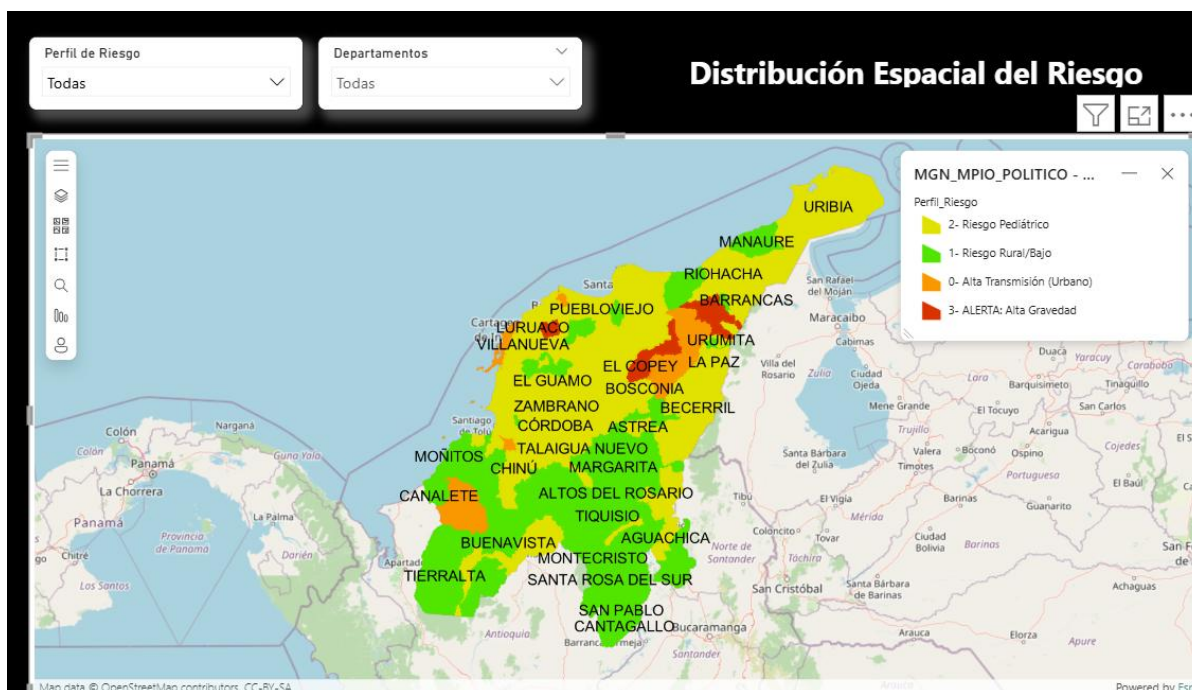
- *Estrategia Sugerida:* Vigilancia comunitaria y educación en zonas veredales.

### ***Distribución Geoespacial del Riesgo***

La asignación de los clústeres a nivel municipal permitió generar una cartografía del riesgo epidemiológico para la Región Caribe. La Figura 15 muestra la distribución espacial de los cuatro perfiles identificados, facilitando la detección rápida de patrones territoriales.

**Figura 15**

*Distribución Espacial de los Clústeres de Riesgo de Dengue en la Región Caribe (K=4).*



Como se evidencia en la visualización, los municipios de “Alta Transmisión” (Clúster 0) corresponde a los nodos urbanos principales (Barranquilla, Cartagena, Montería, Sincelejo y Valledupar), Actualmente como epicentro de la epidemia. Por su parte, los municipios de “Alerta por Letalidad” (Clúster 3) se presentan como focos específicos que, aunque dispersos, requieren una intervención focalizada inmediata debido a la severidad de los desenlaces clínicos reportados.

### **Implementación Tecnológica y Automatización del Modelo**

Uno de los aportes más relevantes de este proyecto fue la implementación de una arquitectura de análisis híbrida, integrando la potencia estadística de Python dentro de un entorno de inteligencia de negocios de Microsoft Power BI.

A diferencia de los enfoques tradicionales donde el modelo se realizaba de forma estática y externa (en archivos aislados), este desarrollo utilizó la capacidad de scripting de Python embebida con Power Query. Esto permitió operacionalizar el algoritmo K-means directamente en el flujo de transformación de datos (ETL).

Esta integración ofrece tres ventajas técnicas y operativas claves para la vigilancia de la salud pública:

1. *Reproducibilidad Automática:* El modelo de clustering no es estático; al actualizar la base de datos con nuevos reportes del SIVIGILA, el script de Python se ejecuta automáticamente, recalculando los centroides y reasignando los municipios a su clúster correspondiente sin intervención manual.
2. *Capacidad Analítica Avanzada:* Se superan las limitaciones nativas de las herramientas de BI convencionales (limitadas a estadísticas descriptivas), incorporando librerías de Machine Learning (*Scikit-learn*) para el descubrimiento de patrones complejos (aprendizaje no supervisado) dentro del mismo tablero de control.
3. *Visualización Integrada:* Los resultados del modelo matemático, se integran nativamente con las capacidades gráficas de Power BI, permitiendo a los tomadores de decisiones interactuar con resultados de ciencias de datos a través de mapas y gráficos dinámicos, democratizando el acceso a la información avanzada.

## Conclusiones

La ejecución del presente trabajo permitió diseñar e implementar una solución de inteligencia de negocios y analítica avanzada, transformando 15 años de registros históricos en conocimiento estratégico para la vigilancia epidemiológica del Dengue en la Region Caribe de Colombia. A partir del procesamiento integral de los datos (2010-2014) y la integración de modelos de aprendizaje no supervisado, se presentan las siguientes conclusiones estructurales:

### **La Dinámica y la Caracterización Epidemiológica**

El análisis descriptivo evidencio que el Dengue en la región no es un evento estático, sino que obedece a una dinámica endemoepidémica cíclica y estacional. Temporalmente, se identificó un comportamiento bimodal con dos periodos de transmisión intensificada al año, un pico menor hacia las semanas epidemiológicas 12-17 y un pico mayor sostenido entre la entre las semanas 30 y 46, coincidiendo con los regímenes de lluvias posequías. Históricamente, los brotes epidémicos más severos (2010, 2013, 2014, 2019, 2023 y 2024) mostrando una correlación directa con la ocurrencia del Fenómeno de El Niño, validando la hipótesis que la variabilidad climática actúa como un catalizador crítico para la proliferación del vector.

Demográficamente, el estudio revelo una realidad compleja. Aunque la enfermedad es predominantemente urbana (88% de los casos en cabeceras municipales), la carga de morbilidad recae de manera masiva sobre la población infantil y adolescentes (0 a 18 años). Los datos son contundentes: el grupo de 6 a 12 años presenta la mayor incidencia, seguido por la primera infancia (0-5 años) y los adolescentes (12-18 años).

### **Segmentación y el Modelo Analítico**

La aplicación del algoritmo K-means (K=4) logro segmentar con éxito la heterogeneidad territorial de la región, superando las limitaciones de la clasificación binaria tradicional.

Mediante la desnormalización de los centroides, fue posible interpretar epidemiológicamente los cuatro perfiles de riesgo identificados, concluyendo que:

### ***Nodo de Alta Transmisión (Clúster 0)***

Conformado por las grandes capitales, este grupo actúa como el motor de la endemia debido a su volumen masivo de casos, aunque mantiene una letalidad controlada, sugiriendo una mejor capacidad de respuesta hospitalaria.

### ***Foco de Alerta por Gravedad (Clúster 3)***

Este es el hallazgo más crítico del modelo. Identifica un grupo de municipios que, pese a tener una transmisión moderada, presenta fallas graves en los desenlaces clínicos, con tasas de letalidad y hospitalización altas. Estos territorios representan puntos ciegos en la calidad de la atención que requieren auditoria inmediata.

### ***Municipios de Riesgo Pediátricos (Clúster 2)***

Agrupar la mayoría de los territorios, caracterizándose porque la epidemia es sostenida fundamentalmente por la población menor de 15 años, lo que apunta a la necesidad de intervenciones centradas en el entorno escolar y familiar.

### ***Zona de Riesgo Bajo/Disperso (Clúster 1)***

Municipios con transmisión esporádica y mayor componente rural, donde el riesgo es menor pero latente.

## **Arquitectura Tecnológica e Integración**

Se demostró la necesidad y eficacia de implementar una arquitectura híbrida que integra Python dentro del flujo de datos de Power BI. Se concluye que el uso exclusivo de herramientas de ETL convencionales, en este caso Power Query, hubiesen sido insuficientes para los objetivos analíticos del proyecto, dada la ausencia de motores estadísticos nativos para la ejecución de

algoritmos de agrupamientos como k-means. La incorporación de script de Python permitió embeber el modelo de Machine Learning directamente en la canalización de datos, logrando que la clasificación de riesgos de los municipios no sea estática, sino que se recalcula directamente con cada actualización del sistema de vigilancia, visualizándose los resultados integrados en el mapa del tablero de control.

### **Calidad de los Datos**

La fragmentación anual y la inconsistencia en los registros históricos del SIVIGILA se identificaron con la principal barrera técnica para el análisis longitudinal en la salud pública. El proceso de consolidación y limpieza de 15 años de datos fue un prerrequisito indispensable, sin el cual la aplicación de cualquier modelo predictivo o descriptivo carecería de validez estadística.

## **Recomendaciones**

Con base a la experiencia técnica de los resultados obtenidos durante la ejecución del proyecto, se plantean las siguientes recomendaciones dirigidas a entidades de la salud pública y a la academia o futuros investigadores:

### **Intervención Intersectorial en Entornos Escolares**

El análisis demográfico y el predominio de “Clúster de Riesgo Pediátrico” (Clúster 2), confirma que la transmisión se concentra en menores de 15 años. Estudios epidemiológicos (Villar et al., 2015a) han señalado a las escuelas como centro de amplificación de la transmisión debido a las actividades diurnas del vector y a la aglomeración de susceptibles.

Se recomienda, trascender el control vectorial domiciliario y articular acciones con las Secretarías de Educación. Adicional, implementar programas de escuelas limpias y libres de criaderos en los municipios de este clúster, focalizando fumigaciones en las instituciones educativas antes de los picos estacionales detectados (semana 30 en adelante).

### **Auditorías Clínicas Focalizadas en el Clúster de Gravedad**

El modelo identificó un grupo de municipios donde la hospitalización supera significativamente el promedio regional, pese a no tener la mayor carga de casos. La literatura de la Organización Panamericana de la Salud (OPS) establece que la mortalidad por Dengue es evitable en más del 98% de los casos si el manejo clínico es oportuno y adecuado.

Se insta a las Secretarías de Salud Departamentales a priorizar estos municipios para realizar auditorías de calidad de los servicios de urgencias. Se debe implementar un reentrenamiento obligatorio del personal médico en las Guías de Atención Clínica Integral del Paciente con Dengue de la OPS, enfocándose en la identificación temprana de signos de alarma y

el manejo de líquidos, ya que la alta letalidad es un indicador proxy de fallas en la calidad de la atención.

### **Atención Diferenciada por Vulnerabilidades Biológicas**

Se evidencio una disociación entre la morbilidad (niños y jóvenes en etapa escolar) y la mortalidad (en los extremos de la vida). El pico de la letalidad en menores de 5 años y mayores de 60 años es consistente con la evidencia sobre la fragilidad capilar en infantes y la inmunosenescencia o comorbilidades en adultos mayores como factores de mal pronóstico.

Se recomienda, implementar una ruta de atención prioritaria para estos grupos etarios desde el primer contacto con el sistema de salud, garantizando seguimientos estrictos independientemente de los signos de alarmas iniciales, para mitigar el riesgo de desenlaces fatales.

### **Integración de Variables Climáticas para la Predicción**

La correlación temporal identificada entre los picos epidémicos del proyecto y el Fenómeno de El Niño se alinean consistentemente con estudios regionales (INS, OPS) que asocian el aumento de temperaturas y las sequías con el incremento de criaderos de criaderos por almacenamiento de agua.

Se recomienda, Evolucionar el tablero de control actual hacia un sistema de alerta temprana, integrando bases de datos meteorológicas del IDEAM al modelo de datos en Power BI, permitiendo predecir brotes con semanas de antelación, tal como sugiere la literatura sobre modelos clima-salud para optimizar la asignación de recursos preventivos.

### **Adopción de Inteligencia Artificial para la Vigilancia en la Salud**

La investigación reciente en la salud pública digital (Casco et al., 2024) respalda el uso de técnicas de minería de datos (Data mining) y agrupamientos (Clustering) para optimizar la asignación de recursos limitados.

Se recomienda, institucionalizar el uso del tablero de control desarrollado, aprovechando la integración de Python y Power BI. Adicional, se sugiere a las entidades territoriales utilizar la clasificación automática del modelo para emitir las alertas epidemiológicas semanales, reduciendo el tiempo de análisis manual y permitiendo una respuesta reactiva basada en datos.

### Referencias Bibliográficas

- Alvis-Guzman, N., Rodríguez-Barreto, H., & Velilla, S. M. (2015). Dengue en un área de Colombia ian Caribbean. *Colombia Medica*, 46(1), 3–7.  
<https://doi.org/10.25100/cm.v46i1.1487>
- Asamblea Nacional Constituyente. (1991). *Leyes desde 1992 - Vigencia expresa y control de constitucionalidad [CONSTITUCION\_POLITICA\_1991]*.  
[http://www.secretariassenado.gov.co/senado/basedoc/constitucion\\_politica\\_1991.html](http://www.secretariassenado.gov.co/senado/basedoc/constitucion_politica_1991.html)
- AWS. (n.d.). *¿Qué es el machine learning? - Explicación de la tecnología ML - AWS*. Retrieved October 12, 2025, from <https://aws.amazon.com/es/what-is/machine-learning/>
- Bagnato, J. I. (2018, March 12). *K-Means con Python paso a paso | Aprende Machine Learning*.  
<https://www.aprendemachinelearning.com/k-means-en-python-paso-a-paso/>
- Canales Luna, J. (2024, February 20). *Python vs. R para ciencia de datos: ¿cuál deberías aprender? | DataCamp*. <https://www.datacamp.com/es/blog/python-vs-r-for-data-science-whats-the-difference>
- Casco, M. E., Fogliato, N. M., & Roatta, S. E. (2024). Minería de datos espaciales como herramienta de mejora en la asignación de recursos en Salud Pública: Un enfoque basado en evidencias. *XXX Congreso Argentino de Ciencias de La Computación*, 639–648.  
<http://sedici.unlp.edu.ar/handle/10915/176529>
- Congreso de Colombia. (2012). *Ley\_1581\_de\_2012*.
- Gan, M. (2025, March 2). *Cómo la visualización de datos está transformando la toma de decisiones en salud. - Soy Marta Gan*. <https://soymartagan.com/visualizacion-de-datos-en-salud/>

- Gopalakrishnan, U., Thiagarajan, K., Felicita, A. S., Gosh, P., Alshehri, A., Awadh, W., Alzahrani, K. J., Alzahrani, F. M., Alsharif, K. F., Halawani, I. F., Alshammeri, S., Alamoudi, A., Albar, D. H., Baeshen, H. A., & Patil, S. (2022). In-Vitro Assessment of the Corrosion Potential of an Oral Strain of Sulfate-Reducing Bacteria on Metallic Orthodontic Materials. *International Journal of Environmental Research and Public Health* 2022, Vol. 19, Page 15312, 19(22), 15312. <https://doi.org/10.3390/IJERPH192215312>
- IBM. (n.d.). *¿Qué es el análisis exploratorio de datos?* | IBM. Retrieved October 12, 2025, from <https://www.ibm.com/mx-es/think/topics/exploratory-data-analysis>
- Instituto Nacional de Salud. (2024). *Protocolo de vigilancia en salud pública. Dengue*. <https://doi.org/10.33610/JQVP8800>
- Laserna, A., Barahona-Correa, J., Baquero, L., Castañeda-Cardona, C., & Rosselli, D. (2018). Economic impact of dengue fever in Latin America and the Caribbean: a systematic review. *Rev Panam Salud Publica*;42, Sept. 2018, 42. <https://doi.org/10.26633/RPSP.2018.111>
- Microsoft. (2022). *Qué es Inteligencia empresarial* | Microsoft Power BI. <https://www.microsoft.com/es-es/power-platform/products/power-bi/topics/business-intelligence/what-is-business-intelligence>
- Ministerio De Salud. (1993). *Resolución 8430 de 1993*.
- Ministerio de Salud. (2024). *El Ministerio de Salud llama a intensificar las medidas de prevención y control del dengue ante el impacto del fenómeno de El Niño y la Niña*. <https://www.minsalud.gov.co/Paginas/intensificar-las-medidas-de-prevencion-y-control-del-dengue.aspx>

Ministerio de Salud y Protección Social. (2016). *Por medio del cual se expide el Decreto Único Reglamentario del Sector Salud y Protección Social EL PRESIDENTE DE LA REPÚBLICA DE COLOMBIA.*

Ministerio de Salud y Protección Social de Colombia. (2022). *Páginas - Vigilancia en salud pública.* <https://www.minsalud.gov.co/salud/publica/epidemiologia/Paginas/vigilancia-salud-publica.aspx>

OPS. (2024). *Dengue: Síntomas, Prevención y Tratamientos - OPS/OMS.*  
<https://www.paho.org/es/temas/dengue>

PAHO/WHO. (2024). *Epidemiological Update Increase in dengue cases in the Region of the Americas.* <https://www3.paho.org/data/index.php/en/mnu-topics/indicadores-dengue-en.html>

Prakash Mahto, M. (2025). International Journal of Research Publication and Reviews Power BI: Revolutionizing Healthcare Through Actionable Insights and Data Visualization. In *International Journal of Research Publication and Reviews* (Issue 6). [www.ijrpr.com](http://www.ijrpr.com)

Sembiring, M. A., Agus, T. A., Fitri, M., Sibuea, L., Tinggi, S., Informatika, M., & Royal, D. K. (2021a). PENERAPAN METODE ALGORITMA K-MEANS CLUSTERING UNTUK PEMETAAN PENYEBARAN PENYAKIT DEMAM BERDARAH DENGUE (DBD). *JOURNAL OF SCIENCE AND SOCIAL RESEARCH*, 4(3), 336–341.  
<https://doi.org/10.54314/JSSR.V4I3.712>

Sembiring, M. A., Agus, T. A., Fitri, M., Sibuea, L., Tinggi, S., Informatika, M., & Royal, D. K. (2021b). PENERAPAN METODE ALGORITMA K-MEANS CLUSTERING UNTUK PEMETAAN PENYEBARAN PENYAKIT DEMAM BERDARAH DENGUE (DBD).

*JOURNAL OF SCIENCE AND SOCIAL RESEARCH*, 4(3), 336–341.

<https://doi.org/10.54314/JSSR.V4I3.712>

Sivigila. (2025). *PortalSivigila2019 Buscador*.

<https://portalsivigila.ins.gov.co/Paginas/Buscador.aspx#>

Taipalus, T., Isomöttönen, V., Erkkilä, H., & Äyrämö, S. (2023). Data Analytics in Healthcare:

A Tertiary Study. *SN Computer Science*, 4(1), 1–14. [https://doi.org/10.1007/S42979-022-](https://doi.org/10.1007/S42979-022-01507-0/TABLES/4)

[01507-0/TABLES/4](https://doi.org/10.1007/S42979-022-01507-0/TABLES/4)

Vallejos, L., Mariño, S. I., & Britos, P. V. (2024). Minería de datos espaciales como herramienta

de mejora en la asignación de recursos en Salud Pública: Un enfoque basado en evidencias.

*XXX Congreso Argentino de Ciencias de La Computación*, 639–648.

<http://sedici.unlp.edu.ar/handle/10915/176529>

Villar, L. A., Rojas, D. P., Besada-Lombana, S., & Sarti, E. (2015a). Epidemiological Trends of

Dengue Disease in Colombia (2000-2011): A Systematic Review. *PLOS Neglected Tropical*

*Diseases*, 9(3), e0003499. <https://doi.org/10.1371/JOURNAL.PNTD.0003499>

Villar, L. A., Rojas, D. P., Besada-Lombana, S., & Sarti, E. (2015b). Epidemiological Trends of

Dengue Disease in Colombia (2000-2011): A Systematic Review. *PLOS Neglected Tropical*

*Diseases*, 9(3), e0003499. <https://doi.org/10.1371/JOURNAL.PNTD.0003499>

WHO. (2025, August 21). *Dengue*. [https://www.who.int/news-room/fact-sheets/detail/dengue-](https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue)

[and-severe-dengue](https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue)

## Apéndices

### Apéndice A

#### *Cronograma de Actividades*

<b>Tiempo</b>	<b>2025</b>							
<b>Actividad</b>	<b>Mes 1</b>	<b>Mes 2</b>	<b>Mes 3</b>	<b>Mes 4</b>	<b>Mes 5</b>	<b>Mes 6</b>	<b>Mes 7</b>	<b>Mes 8</b>
Identificación del problema								
Análisis de fuentes y bases de datos								
Diseño de propuesta de trabajo de grado								
Extracción, limpieza Procesamiento de datos								
Análisis exploratorio de datos (EDA)								
Aplicación de Modelos analíticos								
Construcción de Tablero de visualización								