



Plataforma de gestión de datos para el
monitoreo de cultura ciudadana en el área
metropolitana de Bucaramanga durante el período
2026-2027

Modalidad: Desarrollo Tecnológico

:

Jair Francisco Alvarado Bautista

1005257674

UNIDADES TECNOLÓGICAS DE SANTANDER
Ciencias Naturales e Ingenierías
Tecnología en Desarrollo de Sistemas Informáticos
Bucaramanga
30 de Junio del 2026



Plataforma de gestión de datos para el
monitoreo de cultura ciudadana en el área
metropolitana de Bucaramanga durante el período
2026-2027

Modalidad: Desarrollo Tecnológico

Jair Francisco Alvarado Bautista

1005257674

**Trabajo de Grado para optar al título de
Tecnólogo en Desarrollo de Sistemas Informáticos**

DIRECTOR

Genner Andrés Carrillo Rueda

**UNIDADES TECNOLÓGICAS DE SANTANDER
Ciencias Naturales e Ingenierías
Tecnología en Desarrollo de Sistemas Informáticos
Bucaramanga
30 de Junio del 2026**

Nota de Aceptación

Este informe final de trabajo de grado,
en modalidad Desarrollo Tecnológico, fue APROBADO
En cumplimiento de uno de los requisitos exigidos por las
Unidades Tecnológicas de Santander para optar el
Título de Tecnólogo en Desarrollo de Sistemas Informáticos,
Según acta No. 21 del 2026 - 09 – 15,
del Comité de Trabajo de Grado.

Hugo Pedroza
Evaluador

Genner Andrés Carrillo Rueda
Director

DEDICATORIA

A Dios, por darme la fortaleza y la sabiduría para llegar hasta aquí, y por poner en mi camino a cada una de las personas que hicieron posible este logro. A mis padres, Stella Bautista y Porfirio Alvarado, por creer en mí incluso en los momentos en que yo dudé; su esfuerzo silencioso sostiene cada página de este trabajo. A mi mejor amigo, el geólogo José Alberto Guerrero, por sus palabras justas en el momento justo, que me devolvieron las ganas de seguir cuando más las necesité. A mi novia, Sara Sofía Gómez Cáceres, por su amor, su comprensión infinita y por no dejarme rendir, gracias por acompañarme y orientarme en esta etapa académica. A quienes hicieron parte de mi formación, en especial a la agencia Y2D, por acompañar mi crecimiento. Y con el corazón partido, pero lleno de gratitud, a mi hermano Freddy Rentería Reyes (Q.E.P.D.), quien me mostró este maravilloso mundo de la tecnología, sin imaginar que esas palabras se convertirían en el motor de todo lo que hoy soy. No estás aquí para verlo con tus propios ojos, pero sé que cada línea de este trabajo lleva un poco de ti, y que estarías orgulloso. Este logro es tan mío como tuyo. Te sigo llevando conmigo, hermano, hoy y siempre.

AGRADECIMIENTOS

El autor expresa su agradecimiento a **Genner Andrés Carrillo Rueda**, docente de las Unidades Tecnológicas de Santander y director de este trabajo de grado, por su orientación y acompañamiento a lo largo del desarrollo del proyecto.

A las **Unidades Tecnológicas de Santander**, por la formación académica recibida durante el desarrollo del programa de Tecnología en Desarrollo de Sistemas Informáticos, y por brindar el marco institucional dentro del cual fue posible desarrollar este trabajo.

A la **Secretaría del Interior de la Alcaldía de Bucaramanga**, por poner a disposición del público, a través del portal de Datos Abiertos, el conjunto de datos delictivos del municipio que constituyó una de las fuentes principales de información de este trabajo.

A todas las personas que, de manera voluntaria y anónima, respondieron la encuesta de percepción ciudadana aplicada como parte de la prueba piloto de este proyecto, cuya participación permitió validar la operatividad de la plataforma desarrollada.

TABLA DE CONTENIDO

RESUMEN EJECUTIVO.....	10
INTRODUCCIÓN.....	12
1. DESCRIPCIÓN DEL TRABAJO DE INVESTIGACIÓN	15
1.1. PLANTEAMIENTO DEL PROBLEMA	15
1.2. JUSTIFICACIÓN.....	17
1.3. OBJETIVOS	19
1.3.1. OBJETIVO GENERAL	19
1.3.2. OBJETIVOS ESPECÍFICOS.....	19
1.4. ESTADO DEL ARTE.....	21
2. MARCO REFERENCIAL	25
2.2 MARCO CONCEPTUAL	28
2.3 MARCO LEGAL	29
3. DISEÑO DE LA INVESTIGACIÓN.....	31
3.1 TIPO DE INVESTIGACIÓN	31
3.2 ENFOQUE DE LA INVESTIGACIÓN	32
3.4 MÉTODO DE INVESTIGACIÓN	33
3.4.1 TÉCNICAS E INSTRUMENTOS DE RECOLECCIÓN DE INFORMACIÓN	33
3.5. FASES DEL DESARROLLO METODOLÓGICO	35
3.6. CONSIDERACIONES ÉTICAS.....	36
4. DESARROLLO DEL TRABAJO DE GRADO.....	38
4.1 FASE DE DIAGNÓSTICO.....	38
4.1.1 IDENTIFICACIÓN Y CARACTERIZACIÓN DE FUENTES DE DATOS	38
4.1.2 DEFINICIÓN DE INDICADORES	39
4.2 FASE DE DISEÑO	40
4.2.1 ARQUITECTURA GENERAL	40
4.2.2 MODELO DE DATOS.....	41
4.3 FASE DE IMPLEMENTACIÓN.....	42
4.3.1 EXTRACCIÓN, TRANSFORMACIÓN Y CARGA DE DATOS (ETL).....	42
4.3.2 DESARROLLO DEL BACKEND.....	43

4.3.3	DESARROLLO DEL FRONTEND	45
4.3.4	DESPLIEGUE	46
4.4	FASE DE VALIDACIÓN	47
5.	<u>RESULTADOS</u>	<u>52</u>
5.1.	RESULTADOS DEL DIAGNÓSTICO E INTEGRACIÓN DE FUENTES DE DATOS	52
5.2.	RESULTADOS DEL DESARROLLO DE LA PLATAFORMA	53
5.3.	RESULTADOS DE LA PRUEBA PILOTO: ENCUESTA DE PERCEPCIÓN CIUDADANA	53
5.4.	RESULTADOS DEL ANÁLISIS DE SENTIMIENTO EN REDES SOCIALES	55
5.5.	ANÁLISIS INTEGRADO DE RESULTADOS	56
5.6.	SÍNTESIS DE CUMPLIMIENTO DE LOS OBJETIVOS ESPECÍFICOS	57
5.7.	VALORACIÓN DEL ÉXITO DEL TRABAJO REALIZADO	57
6.	<u>CONCLUSIONES</u>	<u>59</u>
7.	<u>RECOMENDACIONES</u>	<u>63</u>
8.	<u>REFERENCIAS BIBLIOGRAFICAS</u>	<u>66</u>
9.	<u>APÉNDICES</u>	<u>70</u>
10.	<u>ANEXOS.....</u>	<u>85</u>

LISTA DE FIGURAS

No se encuentran elementos de tabla de ilustraciones.

LISTA DE TABLAS

Tabla 1. <i>Comparación de plataformas e iniciativas revisadas en el estado del arte</i>	24
Tabla 2. <i>Matriz de pruebas de los endpoints de la API</i>	48
Tabla 3. <i>Distribución de respuestas de encuesta y promedios de percepción por comuna</i>	54
Tabla 4. <i>Tecnologías empleadas</i>	86
Tabla 5. <i>Referencia de la API — Módulo de Delitos</i>	88
Tabla 6. <i>Referencia de la API — Módulos de Encuestas, Redes Sociales y Chat</i>	88

RESUMEN EJECUTIVO

El presente trabajo abordó la dificultad que enfrentan las autoridades de Bucaramanga para monitorear la cultura ciudadana debido a la desarticulación entre las fuentes de información disponibles. Se planteó como objetivo general desarrollar una plataforma web funcional que integrara datos de encuestas, fuentes institucionales y redes sociales, permitiendo el análisis y la visualización de indicadores de cultura ciudadana como herramienta de apoyo a la gestión pública.

Para alcanzar este objetivo se adoptó un enfoque mixto, desarrollado en cuatro fases: diagnóstico de fuentes de datos e indicadores, diseño de una arquitectura de tres capas con un modelo de datos dimensional, implementación de un backend con procesos de extracción, transformación y carga, y validación mediante verificación técnica y una prueba piloto de recolección de datos reales.

Como resultado, se integraron en un modelo de datos único los registros oficiales de delitos del municipio, respuestas de una encuesta pública de percepción ciudadana recolectadas en distintas comunas del área metropolitana, y comentarios de redes sociales analizados mediante inteligencia artificial. La totalidad de los módulos de consulta fue verificada mediante sumas de control frente a los datos de origen, y la plataforma se desplegó públicamente, incorporando además un agente conversacional capaz de responder preguntas en lenguaje natural con base en los datos reales del sistema.

Se concluyó que la integración de fuentes heterogéneas resulta viable cuando el modelo de datos se diseña para acomodar la naturaleza propia de cada fuente, y que la percepción ciudadana de seguridad no siempre guarda relación proporcional con la frecuencia oficial de delitos, lo que respalda mantener ambas fuentes integradas en lugar de privilegiar una como indicador único. Asimismo, se concluyó que la representatividad territorial de un instrumento de participación ciudadana depende de contar con canales de difusión institucionales, más allá de la disponibilidad tecnológica de la plataforma.

PALABRAS CLAVE: cultura ciudadana, integración de datos, percepción ciudadana, inteligencia artificial, ciudades inteligentes.

INTRODUCCIÓN

Las ciudades intermedias latinoamericanas han incorporado en los últimos años tecnologías de información y comunicación para mejorar la gestión de sus servicios urbanos, siguiendo el paradigma conocido como *ciudad inteligente*, definido por Townsend (2013) como un modelo de desarrollo urbano que combina infraestructura tecnológica y analítica de datos para optimizar la eficiencia de los servicios y la calidad de vida de los habitantes. Esta línea de pensamiento se ha consolidado con el tiempo: estudios posteriores como el de Hashem et al. (2015) documentaron cómo la analítica de grandes volúmenes de datos permite identificar patrones urbanos que resultan invisibles con métodos tradicionales de recolección, y trabajos más recientes, como el de Koirala (2025), muestran que la integración de sensores, GPS y cámaras continúa siendo un campo activo de desarrollo, aplicado hoy principalmente a la optimización del tráfico en tiempo real. Esta literatura, tanto la fundacional como la más reciente, evidencia que el aprovechamiento de datos urbanos heterogéneos es un campo maduro y todavía en expansión.

Sin embargo, esa misma literatura revela también una limitación persistente. Kitchin (2014) advierte que buena parte de las plataformas que se presentan como innovadoras replican el mismo enfoque: capturar variables tangibles y fácilmente cuantificables tráfico, infraestructura, seguridad, dejando de lado fenómenos sociales más complejos y menos susceptibles de una medición directa, como la cultura ciudadana, definida por Mockus (2001) como el conjunto de comportamientos y normas compartidas que sostienen la convivencia y el respeto por lo público. Esta brecha se confirma en

ejemplos actuales de aplicación práctica: plataformas como SeeClickFix, integrada hoy a CivicPlus, permiten a los ciudadanos reportar daños puntuales en el mobiliario urbano (Southern Maryland Chronicle, 2025), pero no están diseñadas para observar comportamientos sociales continuos. A nivel regional, Valenzuela (2015) señala además que la brecha digital condiciona de forma desigual la adopción de estas tecnologías en América Latina, lo que hace que soluciones diseñadas para otros contextos no sean directamente trasladables a una ciudad como Bucaramanga.

Es precisamente en ese punto donde se ubica el aporte de este trabajo. Pese a la amplia disponibilidad de literatura y de plataformas sobre ciudades inteligentes, ninguna de las revisadas integra de manera simultánea encuestas, información institucional de movilidad y percepción ciudadana en redes sociales con el propósito específico de monitorear la cultura ciudadana en un contexto como el del área metropolitana de Bucaramanga. El proyecto no pretende refutar ni sustituir la literatura existente, sino ocupar un vacío que esa misma literatura permite identificar: la ausencia de herramientas que traten la cultura ciudadana con el mismo rigor analítico que hoy se aplica al tráfico o a la infraestructura.

El método empleado para dar solución al problema planteado es de tipo computacional: consiste en el desarrollo de una plataforma web que integra y procesa datos provenientes de fuentes heterogéneas mediante técnicas de analítica de datos y procesamiento de lenguaje natural. Este desarrollo se apoya en una fase previa de recolección de información de enfoque mixto, que combina datos cuantitativos de

encuestas y de fuentes institucionales de datos abiertos con datos cualitativos de naturaleza textual provenientes de redes sociales. El detalle técnico de esta solución computacional, incluida su arquitectura y su implementación, se desarrolla en el capítulo de Desarrollo del trabajo de grado de este documento, mientras que el diseño metodológico de la fase de recolección de datos se presenta en el capítulo de Diseño de la Investigación.

El presente informe se estructura de la siguiente manera: el Capítulo 1 presenta el planteamiento del problema, la justificación y los objetivos que orientan el trabajo, junto con una revisión del estado del arte sobre plataformas similares desarrolladas a nivel nacional e internacional. El Capítulo 2 desarrolla el marco referencial teórico, conceptual y legal que sustenta las decisiones tomadas a lo largo del proyecto. El Capítulo 3 describe el diseño metodológico de la investigación. El Capítulo 4 documenta el desarrollo técnico de la plataforma en sus cuatro fases. El Capítulo 5 presenta los resultados obtenidos, incluyendo los de la prueba piloto de validación. Los Capítulos 6 y 7 presentan, respectivamente, las conclusiones y las recomendaciones derivadas del trabajo.

1. DESCRIPCIÓN DEL TRABAJO DE INVESTIGACIÓN

1.1. PLANTEAMIENTO DEL PROBLEMA

La gestión urbana de ciudades intermedias como Bucaramanga enfrenta una limitación estructural para comprender la cultura ciudadana: la información con la que cuentan las autoridades locales proviene de fuentes que operan de manera aislada. Por un lado, las encuestas de percepción se aplican de forma esporádica y generan datos que quedan desactualizados en pocos meses; por otro, los reportes manuales de las entidades de control dependen de la disponibilidad de personal en campo y cubren solo una fracción del territorio. A esto se suma una tercera fuente, apenas explorada de forma sistemática en el contexto local: las opiniones que los ciudadanos expresan de manera espontánea en redes sociales sobre el uso del espacio público, la movilidad o la convivencia.

El problema no es la ausencia de datos, sino su desarticulación. Los sensores IoT que podrían capturar comportamientos objetivos flujo peatonal, ocupación de andenes, niveles de ruido no están conectados con los sistemas donde se almacenan los resultados de encuestas, y ninguno de los dos se cruza con el análisis de lo que se dice en redes sociales (Hashem et al., 2015). Como consecuencia directa, cuando una autoridad municipal necesita decidir dónde priorizar una campaña de cultura ciudadana o evaluar si una intervención tuvo efecto, debe apoyarse en información parcial, recolectada en momentos distintos y sin posibilidad real de correlación entre variables. El efecto de esta desarticulación es una toma de decisiones reactiva, que responde a quejas puntuales en lugar de anticiparse a partir de tendencias identificables.

De este análisis surge la pregunta que orienta el presente trabajo: ¿en qué medida la integración de datos de encuestas, sensores IoT y redes sociales en una plataforma web permite identificar correlaciones entre el uso del espacio público y la percepción ciudadana en el área metropolitana de Bucaramanga durante el período 2026-2027?

1.2. JUSTIFICACIÓN

Este trabajo se justifica, en primer lugar, por una necesidad práctica: las decisiones sobre convivencia y cultura ciudadana en Bucaramanga suelen tomarse hoy con información incompleta o desactualizada, lo que reduce su efectividad. Contar con una plataforma capaz de cruzar distintas fuentes de datos no resuelve por sí sola los problemas de convivencia, pero sí amplía el margen de evidencia disponible para que las autoridades diseñen intervenciones más focalizadas por ejemplo, saber con mayor precisión qué parques requieren atención prioritaria o si una campaña educativa tuvo impacto medible en el comportamiento ciudadano.

En el plano tecnológico, el proyecto implica resolver retos concretos de interoperabilidad entre fuentes heterogéneas datos estructurados de encuestas, series de tiempo de sensores y texto no estructurado de redes sociales, lo cual constituye en sí mismo un aporte técnico replicable para el monitoreo de otros fenómenos urbanos, no solo de cultura ciudadana. En el plano social, al facilitar el acceso a información agregada sobre el estado de la convivencia, el proyecto favorece la transparencia y ofrece a la ciudadanía un insumo adicional para participar en la discusión pública sobre su entorno. En cuanto a las implicaciones económicas, la plataforma permitiría optimizar la asignación de recursos públicos al dirigir las campañas y acciones municipales hacia los puntos críticos identificados con evidencia, reduciendo el gasto en intervenciones de baja efectividad; del mismo modo, un eventual impacto ambiental indirecto se deriva de una mejor gestión del espacio público, que incide en la ocupación ordenada de andenes y zonas verdes, contribuyendo a un entorno urbano más sostenible.

El proyecto es viable en el marco de un trabajo de pregrado tecnológico, dado que se cuenta con acceso a conjuntos de datos abiertos de la Alcaldía de Bucaramanga y del Área Metropolitana, así como con herramientas de código abierto para el procesamiento de texto (NLP) y la visualización de datos (Power BI, Python o similares). El alcance se ha delimitado a un piloto en tres comunas de la ciudad, lo que garantiza su ejecución en el tiempo disponible (2026-2027) sin requerir infraestructura computacional avanzada más allá de la disponible en los laboratorios de la institución. Adicionalmente, la disponibilidad de datos públicos de redes sociales y sensores IoT de bajo costo hace factible la implementación sin depender de presupuestos externos.

Para las Unidades Tecnológicas de Santander, el trabajo se alinea con las líneas de investigación institucionales orientadas a la aplicación de tecnologías de la información en la solución de problemáticas del entorno, y evidencia cómo un proyecto de grado de pregrado tecnológico puede convertirse en un insumo con utilidad real para la gestión pública local. Para el estudiante, este proyecto representa una oportunidad formativa integral: implica desarrollar competencias en análisis de datos heterogéneos, minería de textos y visualización de información, habilidades altamente demandadas en el sector tecnológico y en el ámbito de las ciudades inteligentes. La experiencia de trabajar con datos reales y con una problemática concreta del entorno fortalece su capacidad de resolver problemas complejos y su preparación para enfrentar los desafíos de la cuarta revolución industrial en contextos de gestión pública, aportando así no solo a su formación académica sino también a su proyección profesional.

1.3. OBJETIVOS

1.3.1. OBJETIVO GENERAL

Desarrollar una plataforma web funcional que integre datos de encuestas, sensores IoT y redes sociales, permitiendo el análisis y la visualización de indicadores de cultura ciudadana como herramienta de apoyo para la gestión pública.

1.3.2. OBJETIVOS ESPECÍFICOS

- Diagnosticar los requerimientos de información y los indicadores clave de cultura ciudadana relevantes para el entorno local mediante la revisión de literatura y consultas con actores de interés.
- Diseñar la arquitectura de software de la plataforma, definiendo sus módulos, la estructura de la base de datos y las tecnologías a utilizar para garantizar su funcionalidad y escalabilidad.
- Implementar los componentes funcionales de la plataforma, programando el backend para el procesamiento de datos y el frontend para la visualización interactiva de la información.
- Validar la operatividad de la plataforma a través de una prueba piloto controlada, verificando su capacidad para la consulta y análisis de los indicadores definidos.

Adicionalmente, se consultó a tres actores de interés vinculados a distintas comunas del área metropolitana (Comuna 3, Comuna 4 y Comuna 11), cuyas actas se documentan en el Apéndice D. Las consultas confirmaron la pertinencia de los indicadores definidos a partir de la revisión de literatura: los tres consultados asociaron la cultura ciudadana con comportamientos de cuidado del espacio público y respeto por la tranquilidad de los vecinos, coincidiendo con las dimensiones de convivencia y cuidado de lo público descritas por Mockus (2001), y priorizaron la percepción de seguridad y el uso adecuado del espacio público como los temas más relevantes para su sector. Como hallazgo adicional, uno de los consultados, presidente de una Junta de Acción Comunal, señaló interés específico en conocer los horarios en que se concentran los delitos indicadores que la plataforma ya incorpora a través del módulo de análisis por rango horario, y otro consultado destacó como valioso que la ciudadanía no pueda alterar los datos de la plataforma, reforzando la pertinencia del diseño de solo lectura pública adoptado en este trabajo.

1.4. ESTADO DEL ARTE

La literatura sobre monitoreo urbano mediante tecnología se ha desarrollado principalmente alrededor del concepto de *Smart City*, entendido como un modelo de gestión que combina tecnologías de la información, sensórica y analítica de datos para mejorar la eficiencia de los servicios urbanos (Townsend, 2013). Dentro de ese campo más amplio, existen dos tradiciones que resultan relevantes para este trabajo: las plataformas de reporte ciudadano y los estudios de *big data* urbano.

En Estados Unidos, aplicaciones como SeeClickFix hoy integrada a CivicPlus permiten a los residentes reportar problemas no urgentes, como baches o daños en el mobiliario público, y a partir de esos reportes las administraciones locales priorizan la asignación de recursos (Southern Maryland Chronicle, 2025). Se trata de un modelo probado, pero limitado a problemas tangibles y fácilmente cuantificables; no está diseñado para capturar comportamientos o percepciones, que es precisamente el terreno de la cultura ciudadana.

Un segundo grupo de trabajos se ocupa de la analítica de datos urbanos a gran escala. Koirala (2025) muestra cómo la integración de sensores, GPS y cámaras permite optimizar la gestión del tráfico mediante procesamiento en tiempo real, evidenciando que el valor de estas plataformas está menos en la cantidad de datos recolectados que en la capacidad de cruzarlos entre sí. En América Latina, este tipo de iniciativas avanza de forma desigual: mientras ciudades como Medellín avanzan en la consolidación de ecosistemas de datos orientados a la movilidad dentro de sus planes de ciencia, tecnología e innovación, en Bucaramanga los esfuerzos de digitalización de servicios públicos siguen siendo puntuales y no integrados entre sí. Valenzuela (2015) atribuye parte de este rezago regional a la brecha digital, que condiciona tanto la disponibilidad de infraestructura como los patrones de uso de redes sociales que alimentarían un sistema de este tipo.

Vale la pena introducir aquí una salvedad que atraviesa buena parte de la literatura sobre *big data* urbano: la tendencia a presentar los datos generados por sensores como evidencia objetiva y neutral de lo que ocurre en una ciudad. Kitchin (2014) cuestiona esa idea y advierte que los datos urbanos no están libres de sesgos, pues dependen de dónde se instalan los sensores, qué variables se decide medir y qué interpretación se hace de ellas; el riesgo, señala el autor, es que la aparente objetividad de los números oculte decisiones de diseño que sí son políticas.

Esta advertencia es pertinente para el presente proyecto: un indicador como "respeto a las cebras" medido por una cámara en una única intersección no representa el comportamiento de toda la ciudad, y su interpretación debe hacerse con esa limitación en mente, más aún cuando se combine con datos de percepción provenientes de redes sociales, que tampoco son representativos del conjunto de la ciudadanía sino de quienes participan activamente en esas plataformas.

De la revisión anterior se identifica un vacío específico: las plataformas existentes, tanto internacionales como locales, se concentran en problemas tangibles tráfico, infraestructura, seguridad y prácticamente ninguna integra de manera simultánea sensores IoT, encuestas estructuradas y análisis de redes sociales con el propósito de monitorear un fenómeno tan multidimensional como la cultura ciudadana, entendida en los términos planteados por Mockus (2001) como el conjunto de comportamientos y normas compartidas que sostienen la convivencia. Este trabajo busca ocupar ese espacio, con la salvedad metodológica señalada por Kitchin (2014) como criterio de diseño y no solo como advertencia teórica.

Tabla 1. Comparación de plataformas e iniciativas revisadas en el estado del arte

Plataforma o iniciativa	Ubicación	Fuentes de datos integradas	Enfoque principal	Limitación identificada respecto a cultura ciudadana
SeeClickFix / CivicPlus (Southern Maryland Chronicle, 2025)	Estados Unidos	Reportes ciudadanos individuales sobre infraestructura	Gestión y priorización de mantenimiento urbano	No captura percepción ciudadana ni comportamientos; se limita a reportes puntuales de daños o fallas
Analítica de tráfico urbano con sensores, GPS y cámaras (Koirala, 2025)	Contexto internacional (sin ciudad específica)	Sensores IoT de tráfico en tiempo real	Optimización de la gestión vial	Se concentra en un único fenómeno tangible (tráfico); no incorpora percepción ciudadana ni redes sociales
Ecosistemas de datos de movilidad en ciudades colombianas intermedias y capitales	Colombia (referencia general)	Datos institucionales de movilidad y transporte público	Planificación y monitoreo del transporte	No integra encuestas de percepción ni análisis de redes sociales; enfoque exclusivamente operativo
Plataforma desarrollada en este trabajo de grado	Bucaramanga, Colombia	Datos institucionales de delitos, encuesta de percepción ciudadana y comentarios de redes sociales analizados con inteligencia artificial	Monitoreo integral de cultura ciudadana mediante integración de fuentes heterogéneas	Cobertura geográfica aún parcial de la encuesta piloto (8 de 23 comunas), reconocida explícitamente en el Capítulo 5

Nota. Elaborada por el autor a partir de las fuentes citadas en el apartado 1.4 (Estado del arte) de este documento

2. MARCO REFERENCIAL

2.1 Marco teórico

2.1.1. Ciudades inteligentes y gobernanza de datos

El concepto de ciudad inteligente describe un modelo de desarrollo urbano que combina infraestructura tecnológica, sensórica y analítica de datos con el propósito de mejorar la eficiencia de los servicios, la sostenibilidad ambiental y la calidad de vida de sus habitantes (Townsend, 2013). Este trabajo se inscribe en dicho paradigma, en la medida en que busca aplicar tecnología para comprender mejor una dimensión de la vida urbana la cultura ciudadana que suele quedar fuera de los sistemas convencionales de gestión de la ciudad, más centrados en infraestructura física que en comportamientos sociales.

Sin embargo, adoptar este paradigma exige también asumir una postura crítica frente a él. Kitchin (2014) advierte que buena parte del discurso sobre ciudades inteligentes presenta los datos generados por sensores como evidencia neutral y objetiva de lo que ocurre en una ciudad, cuando en realidad esos datos dependen de decisiones de diseño qué se mide, dónde se ubican los sensores, qué variables se descartan que no son neutrales sino técnicas y, en el fondo, también políticas. Esta advertencia resulta directamente aplicable al presente proyecto: un indicador construido a partir de una cámara instalada en una única intersección, o a partir de publicaciones en redes sociales de un segmento particular de la población, no puede leerse como un retrato objetivo del comportamiento ciudadano en su conjunto. Por esta razón, la plataforma que se propone debe entenderse como una herramienta de apoyo a la gestión pública y no como un sistema de medición exacta e infalible de la cultura ciudadana.

Vinculado a lo anterior, la gobernanza de datos define las políticas, roles y procesos necesarios para tratar la información como un activo estratégico. En este proyecto adquiere especial relevancia porque la plataforma procesará información generada por y sobre personas: garantizar que los datos se recolecten, almacenen y utilicen de forma ética y segura no es un requisito accesorio, sino una condición para que la herramienta genere confianza tanto en las autoridades como en la ciudadanía que eventualmente participe en ella.

2.1.2 Cultura ciudadana

El concepto de cultura ciudadana, desarrollado y aplicado en Bogotá durante la administración de Antanas Mockus, se define como el conjunto de comportamientos, valores y normas compartidas que favorecen la convivencia, el respeto por lo público y el cumplimiento voluntario de las reglas (Mockus, 2001). Su valor conceptual para este trabajo no está solo en la definición, sino en el énfasis que pone en la autorregulación: la cultura ciudadana no se mide únicamente verificando si existe una norma, sino observando si las personas la cumplen incluso cuando nadie las vigila. Esta distinción es la que orienta el tipo de indicadores que la plataforma buscará capturar, por ejemplo, el respeto a las cebras peatonales o el uso correcto de las canecas de basura, que son comportamientos observables y no simples declaraciones de intención recogidas en una encuesta.

2.1.3 Internet de las Cosas y sensores urbanos

El Internet de las Cosas (IoT) corresponde a la red de dispositivos físicos equipados con sensores y capacidad de conexión que les permite capturar y transmitir datos del entorno en tiempo real. En el contexto de una ciudad, estos sensores funcionan como una fuente de información que no depende de la percepción subjetiva de quien responde una encuesta, sino

de un registro directo del comportamiento observado, por ejemplo, mediante visión por computador para contar vehículos que respetan un semáforo, o sensores acústicos para medir contaminación auditiva en un parque. Para este proyecto, el IoT cumple la función de contrastar lo que las personas dicen (a través de encuestas o redes sociales) con lo que efectivamente ocurre en el espacio público.

2.1.4 Analítica de Big Data y visualización de datos

La plataforma propuesta integrará información proveniente de fuentes distintas y en formatos heterogéneos estructurados, semiestructurados y no estructurados, lo que técnicamente se conoce como *big data*. La analítica de *big data* aporta las técnicas necesarias para procesar ese volumen de información y detectar patrones o correlaciones que no serían visibles analizando cada fuente por separado (Hashem et al., 2015). La visualización de datos, por su parte, es la disciplina encargada de traducir esos resultados en representaciones gráficas comprensibles: mapas de calor, gráficos de tendencia, tableros interactivos (dashboards), sin las cuales el análisis técnico no tendría ninguna utilidad práctica para quien debe tomar una decisión de gestión pública.

2.1.5 Procesamiento de Lenguaje Natural para análisis de redes sociales

El Procesamiento de Lenguaje Natural (PLN) es la rama de la inteligencia artificial que permite a un sistema interpretar texto en lenguaje humano. En este proyecto, el PLN es la tecnología que hace posible aprovechar una de las fuentes más ricas y también más complejas de información: lo que los ciudadanos publican espontáneamente en redes sociales. Mediante técnicas como el análisis de sentimientos o el modelado de tópicos, es posible identificar si las conversaciones sobre un espacio público específico son mayoritariamente positivas o

negativas, y qué temas de convivencia concentran la atención en un sector determinado. Esta fuente, sin embargo, comparte la misma limitación señalada por Kitchin (2014): quienes publican en redes sociales no son una muestra representativa de toda la ciudadanía, por lo que sus resultados deben interpretarse como una señal complementaria y no como una medición poblacional.

2.2 Marco conceptual

Para facilitar la comprensión del documento, se precisan a continuación algunos conceptos empleados de forma recurrente:

- **Cultura ciudadana:** conjunto de comportamientos, valores y normas compartidas que favorecen la convivencia, el respeto por lo público y el cumplimiento voluntario de las reglas (Mockus, 2001).
- **Indicador de cultura ciudadana:** variable observable y medible que representa un comportamiento asociado a la convivencia o al respeto por lo público (por ejemplo, el porcentaje de peatones que cruzan por la cebra en un punto determinado).
- **Dato estructurado:** información organizada en un formato predefinido, como los registros de una base de datos relacional (por ejemplo, las respuestas cerradas de una encuesta).
- **Dato no estructurado:** información que no sigue un formato fijo, como el texto libre de una publicación en redes sociales.
- **Interoperabilidad:** capacidad de distintos sistemas o fuentes de datos para intercambiar información y ser interpretados de manera conjunta.

- **Dashboard (tablero de control):** panel visual interactivo que agrupa distintos indicadores para facilitar su consulta y análisis.
- **Anonimización:** proceso mediante el cual se elimina o transforma la información que permitiría identificar a una persona natural a partir de un conjunto de datos.
- **Internet de las Cosas (IoT):** red de dispositivos físicos equipados con sensores y capacidad de conexión que les permite capturar y transmitir datos del entorno en tiempo real.
- **Procesamiento de Lenguaje Natural (PLN):** rama de la inteligencia artificial que permite a un sistema interpretar, analizar y generar texto en lenguaje humano.
- **Big data:** conjunto de técnicas y herramientas para procesar grandes volúmenes de datos heterogéneos que no pueden ser analizados con métodos tradicionales.

2.3 Marco legal

El desarrollo de la plataforma implica el tratamiento de datos que provienen, directa o indirectamente, de personas naturales, por lo cual su diseño debe ajustarse al marco normativo colombiano vigente en materia de protección de datos y gestión de información pública.

En primer lugar, el derecho al habeas data está consagrado en el artículo 15 de la Constitución Política de Colombia (1991), que reconoce el derecho de toda persona a conocer, actualizar y rectificar la información que se haya recogido sobre ella en bases de datos. Este derecho fue desarrollado por la Ley Estatutaria 1581 de 2012 (Congreso de la República de Colombia, 2012), que establece los principios generales para el tratamiento de datos personales en Colombia autorización previa y expresa del titular, finalidad específica del

tratamiento, seguridad de la información y designa a la Superintendencia de Industria y Comercio como autoridad de vigilancia. Esta ley, reglamentada parcialmente por el Decreto 1377 de 2013 (Presidencia de la República de Colombia, 2013), resulta directamente aplicable al proyecto en la medida en que cualquier encuesta o mecanismo de recolección de datos personales que alimente la plataforma deberá contar con el consentimiento informado del titular, tal como ya se contempló en las consideraciones éticas de la propuesta de trabajo de grado. Cabe anotar que, al momento de elaboración de este documento, cursa en el Congreso de la República un proyecto de reforma a la Ley 1581 de 2012 orientado a actualizar el régimen de protección de datos personales frente a los retos de la inteligencia artificial y la economía digital; dado que se trata de una iniciativa aún no aprobada, el proyecto se rige por la normativa actualmente vigente, sin perjuicio de que su diseño busque anticiparse a principios como la transparencia algorítmica que dicha reforma plantea.

En segundo lugar, en tanto la plataforma está orientada a apoyar la gestión pública y eventualmente a poner información a disposición de la ciudadanía, resulta pertinente la Ley 1712 de 2014 (Congreso de la República de Colombia, 2014), que regula el derecho de acceso a la información pública y establece el principio de divulgación proactiva por parte de las entidades del Estado. Aunque el proyecto no es desarrollado directamente por una entidad pública, su propósito de apoyo a la gestión municipal lo alinea con este principio de transparencia.

En el ámbito de política pública, el documento CONPES 3920 de 2018 (Departamento Nacional de Planeación [DNP], 2018) Política Nacional de Explotación de Datos (Big Data) constituye el referente nacional en la materia: reconoce los datos como un activo estratégico

para la generación de valor social y económico, y plantea la necesidad de superar la desarticulación entre fuentes de información como una de las principales barreras para su aprovechamiento en Colombia. Este diagnóstico coincide, a escala local, con el problema que motiva el presente trabajo de grado, y ofrece un respaldo de política pública a la pertinencia de desarrollar soluciones que integren fuentes de datos dispersas.

Finalmente, si bien el proyecto no involucra obras civiles ni intervenciones físicas que generen afectación directa a recursos naturales, se reconoce la importancia de la normativa ambiental vigente en Colombia, en particular la Ley 99 de 1993 (Congreso de la República de Colombia, 1993) que creó el Sistema Nacional Ambiental (SINA) y el Decreto 1076 de 2015 (Presidencia de la República de Colombia, 2015) (Decreto Único Reglamentario del Sector Ambiente y Desarrollo Sostenible), en lo relacionado con el ordenamiento territorial y el uso del espacio público. La plataforma, al facilitar una gestión más eficiente del espacio urbano, contribuye indirectamente a reducir impactos ambientales asociados a la ocupación desordenada del territorio.

3. DISEÑO DE LA INVESTIGACIÓN

3.1 Tipo de investigación

El presente trabajo se clasifica como una investigación de tipo aplicado, en tanto no busca generar conocimiento teórico nuevo sobre la cultura ciudadana, sino utilizar el conocimiento ya existente sobre este fenómeno para desarrollar una herramienta tecnológica concreta que apoye su monitoreo. Dentro de esta categoría, el estudio tiene un alcance descriptivo: pretende identificar y caracterizar los indicadores de cultura ciudadana relevantes

para el área metropolitana de Bucaramanga a partir de la integración de distintas fuentes de datos, sin que ello implique establecer relaciones causales entre las variables observadas. Un alcance descriptivo es apropiado en esta etapa porque el problema identificado en el Capítulo 1 no es la ausencia de teoría sobre cultura ciudadana, sino la falta de una herramienta que permita observarla de manera integrada; explicar por qué ocurren esos comportamientos quedaría para una fase posterior, una vez la plataforma esté operando y se disponga de series de datos más amplias.

3.2 Enfoque de la investigación

El proyecto adopta un enfoque mixto, que combina componentes cuantitativos y cualitativos en función de la naturaleza de cada fuente de datos (Hernández-Sampieri et al., 2014).

El componente cuantitativo se apoya en los datos numéricos provenientes de encuestas estructuradas y de sensores IoT, que permiten cuantificar comportamientos observables, por ejemplo, el porcentaje de peatones que respeta un cruce señalizado. El componente cualitativo se apoya en el análisis de contenido generado espontáneamente por los ciudadanos en redes sociales, cuya naturaleza textual y no estructurada requiere técnicas distintas de interpretación, como el análisis de sentimientos. La combinación de ambos enfoques no es una decisión metodológica arbitraria: responde directamente al problema planteado, que identifica la desarticulación entre fuentes estructuradas y no estructuradas como la causa principal de la limitación actual en el monitoreo de cultura ciudadana

3.4 Método de investigación

Se emplea un método de observación, complementado con procedimientos de análisis documental en la fase de diagnóstico inicial. La observación se aplica de dos maneras distintas: de forma directa, a través de los sensores IoT que registran comportamientos en el espacio público sin intervención del observador, y de forma indirecta, a través de la revisión de lo que los ciudadanos expresan por su propia iniciativa en redes sociales.

El análisis documental, por su parte, se emplea en la fase de diagnóstico para identificar, a partir de literatura académica y experiencias previas revisadas en el estado del arte, qué indicadores de cultura ciudadana resultan pertinentes y viables de capturar tecnológicamente en el contexto de Bucaramanga. Adicionalmente, se emplea un método inductivo en el proceso de análisis de datos: a partir de la observación de comportamientos específicos registrados por las distintas fuentes, se buscarán patrones y correlaciones que permitan construir indicadores agregados de cultura ciudadana, sin partir de hipótesis previas sobre el comportamiento esperado.

3.4.1 Técnicas e instrumentos de recolección de información

Dado que la plataforma integra tres fuentes de datos de naturaleza distinta, se emplean técnicas diferenciadas para cada una.

En primer lugar, las encuestas estructuradas serán aplicadas a una muestra de ciudadanos del área metropolitana de Bucaramanga, con preguntas cerradas orientadas a medir percepción sobre convivencia y uso del espacio público. El instrumento se diseñará y

validará durante la fase de diagnóstico (ver Plan de actividades, Fase 1) y se incluirá como apéndice del informe final una vez definido.

En segundo lugar, en lugar de instrumentar sensores propios, la plataforma consumirá conjuntos de datos ya publicados por entidades del área metropolitana, principalmente a través del portal de Datos Abiertos del Área Metropolitana de Bucaramanga (AMB), integrado con el Portal de Datos Abiertos del Gobierno Nacional, y del portal de Datos Abiertos de la Dirección de Tránsito de Bucaramanga (DTB). Estas fuentes publican series relacionadas con movilidad (parque automotor, transporte público, validaciones de usuarios) y calidad ambiental (calidad del aire, datos hídricos), que se emplearán como proxy de comportamiento urbano mientras no exista un conjunto de datos específico de cultura ciudadana. Esta decisión reduce el alcance técnico del proyecto respecto de instrumentar hardware propio, pero aumenta su viabilidad dentro del tiempo disponible para el trabajo de grado, y se ajusta al marco legal de datos abiertos (Ley 1712 de 2014) descrito en el Capítulo 2.

En tercer lugar, se realizarán consultas con actores de interés mediante entrevistas o reuniones exploratorias con actores institucionales o comunitarios que permitan validar la pertinencia de los indicadores propuestos antes de iniciar el desarrollo técnico.

Finalmente, se empleará la extracción de datos de redes sociales mediante la recolección de publicaciones públicas relacionadas con espacios y temas de convivencia en Bucaramanga, sujeta a los términos de uso de cada plataforma y a los principios de tratamiento de datos personales descritos en el marco legal.

3.5. Fases del desarrollo metodológico

El desarrollo del trabajo se organiza en cuatro fases, que corresponden directamente a los objetivos específicos planteados y al plan de actividades definido en la propuesta de trabajo de grado.

La Fase 1, denominada Diagnóstico, se asocia al Objetivo específico 1 e incluye la revisión de literatura, la consulta con actores de interés, y la identificación y caracterización de fuentes de datos disponibles. El producto esperado de esta fase es un documento con los indicadores de cultura ciudadana definidos y las fuentes de datos caracterizadas.

La Fase 2, denominada Diseño, se asocia al Objetivo específico 2 y comprende la definición de la arquitectura de software, la estructura de la base de datos y las tecnologías a emplear para garantizar la interoperabilidad entre las tres fuentes de datos identificadas. El producto esperado es el modelo de integración y la arquitectura de la plataforma documentada.

La Fase 3, denominada Implementación, se asocia al Objetivo específico 3 y consiste en el desarrollo del backend para el procesamiento y correlación de los datos, y del frontend para su visualización interactiva. El producto esperado es una plataforma web funcional con módulo de análisis y visualización.

La Fase 4, denominada Validación, se asocia al Objetivo específico 4 y consiste en la ejecución de una prueba piloto controlada mediante el despliegue público de la plataforma y la difusión del instrumento de encuesta a través de canales institucionales y comunitarios, sin restringir la participación a un número predefinido de comunas. Esta decisión metodológica responde a la naturaleza misma del producto desarrollado: al tratarse de una plataforma web

de acceso público, limitar la prueba piloto a tres comunas habría subutilizado su capacidad real de alcance, mientras que una convocatoria abierta permite valorar la operatividad del sistema en condiciones más cercanas a las de su uso final. El producto esperado es un informe de validación con los resultados de la prueba piloto, la cobertura geográfica efectivamente lograda y las recomendaciones derivadas de su análisis.

3.6. Consideraciones éticas

Dado que el proyecto involucra el tratamiento de datos provenientes de personas naturales, se adoptan las siguientes consideraciones éticas, alineadas con el marco legal descrito en el Capítulo 2.

En primer lugar, se informó a los participantes, mediante un texto introductorio visible en el formulario, que sus respuestas serían utilizadas exclusivamente con fines académicos y de análisis agregado, y que su participación era voluntaria y anónima. No obstante, se reconoce como limitación metodológica que el formulario no incluyó un mecanismo formal de consentimiento informado (por ejemplo, una casilla de aceptación explícita o un documento firmado), a diferencia de lo inicialmente previsto. Esta limitación se documenta de manera transparente y se recomienda su corrección en fases posteriores del proyecto, dado que la naturaleza anónima de los datos recolectados sin captura de nombres, identificaciones ni otros datos personales mitiga, aunque no elimina, el riesgo asociado a esta omisión

En segundo lugar, toda la información recolectada será anonimizada, eliminando cualquier dato que permita identificar a una persona natural (nombres, cédulas, direcciones, etc.).

En tercer lugar, en lo relacionado con el tratamiento de datos de redes sociales, solo se utilizarán publicaciones de acceso público, respetando los términos de uso de cada plataforma. No se almacenará información que permita identificar a los usuarios.

Finalmente, los datos recolectados serán almacenados en servidores seguros con acceso restringido al equipo investigador, y serán destruidos una vez finalizado el proyecto.

4. DESARROLLO DEL TRABAJO DE GRADO

El desarrollo del trabajo de grado se organizó siguiendo las cuatro fases metodológicas definidas en el Capítulo 3: diagnóstico, diseño, implementación y validación. A continuación, se documenta el trabajo realizado en cada una, con énfasis en las decisiones técnicas tomadas y su justificación.

4.1 Fase de diagnóstico

4.1.1 *Identificación y caracterización de fuentes de datos*

Conforme a lo planteado en la propuesta de trabajo de grado, se preveía integrar tres fuentes de datos: encuestas de percepción ciudadana, sensores IoT y redes sociales. Durante esta fase se llevó a cabo un diagnóstico de viabilidad real de cada fuente, que condujo a dos ajustes justificados respecto del planteamiento inicial.

En primer lugar, se determinó que instrumentar sensores IoT propios excedía el alcance temporal y presupuestal de un trabajo de grado de pregrado tecnológico. Como alternativa, se identificó y evaluó el portal de Datos Abiertos del municipio de Bucaramanga (datos.gov.co), encontrándose disponible el conjunto de datos "Información delictiva del municipio de Bucaramanga", a cargo de la Secretaría del Interior, con registros desde 2016 hasta la fecha de consulta y actualización periódica. Este conjunto de datos se adoptó como sustituto de los sensores IoT: al igual que un sensor, constituye un registro objetivo de comportamientos observables en el espacio público a diferencia de una encuesta, no depende de la percepción subjetiva de quien reporta, y además guarda una relación conceptual más directa con las

dimensiones de cultura ciudadana descritas por Mockus (2001), en particular la de convivencia y respeto por la vida, que la información de movilidad considerada inicialmente.

En segundo lugar, se evaluó el acceso a APIs de redes sociales convencionales (X/Twitter, Facebook/Instagram). Se encontró que, a la fecha de esta consulta, X eliminó por completo su nivel de acceso gratuito para nuevos desarrolladores, y que el acceso a contenido público de Meta (Facebook/Instagram) requiere un proceso de revisión de aplicación y verificación empresarial no compatible con los tiempos de un trabajo de grado, sin garantía de aprobación. Como alternativa viable, gratuita y de acceso inmediato, se seleccionó la API de datos de YouTube (YouTube Data API v3), utilizando como fuente los comentarios públicos de videos del canal del periódico Vanguardia relacionados con seguridad y convivencia en Bucaramanga. Si bien YouTube no es una red social en el sentido estrictamente sociotécnico del término, cumple la misma función que se buscaba en el planteamiento original: capturar percepción ciudadana espontánea y no estructurada, complementaria a la de las otras dos fuentes.

4.1.2 Definición de indicadores

A partir de la revisión de literatura documentada en el Capítulo 2 y de la estructura de las fuentes de datos identificadas, se definieron los siguientes indicadores de cultura ciudadana a monitorear:

- Frecuencia y distribución geográfica de delitos, como indicador objetivo de la dimensión de convivencia y respeto por la vida.

- Distribución temporal (horaria, diaria, mensual) de los delitos, como insumo para la focalización de intervenciones.
- Percepción ciudadana de seguridad, diferenciada entre horario diurno y nocturno, como indicador subjetivo complementario al anterior.
- Percepción ciudadana de uso adecuado del espacio público.
- Nivel de participación comunitaria autorreportada.
- Sentimiento predominante en las conversaciones ciudadanas espontáneas sobre seguridad y convivencia.

4.2 Fase de diseño

4.2.1 *Arquitectura general*

Se diseñó una arquitectura de tres capas: una capa de datos (base de datos relacional PostgreSQL), una capa de backend (API REST desarrollada en Python con el framework FastAPI (Ramírez, 2018), utilizando SQLAlchemy (SQLAlchemy Contributors, 2006) como capa de mapeo objeto-relacional para la comunicación con la base de datos) y una capa de presentación (dashboard web desarrollado en HTML, CSS y JavaScript). A esta arquitectura se sumó un componente transversal de inteligencia artificial, materializado en un agente conversacional capaz de consultar los propios endpoints del sistema para responder preguntas en lenguaje natural con datos verificables, evitando así que el modelo de lenguaje genere cifras no respaldadas por los datos reales de la plataforma.

Esta arquitectura se seleccionó por ser la más adecuada para separar responsabilidades: la base de datos gestiona la persistencia, el backend gestiona la lógica de negocio y la agregación de información, y el frontend gestiona exclusivamente la presentación

visual, sin lógica de negocio propia. Esta separación facilita que, en trabajos futuros, cualquiera de las capas pueda reemplazarse o extenderse sin afectar a las demás.

4.2.2 Modelo de datos

Para la base de datos se adoptó un modelo dimensional, también conocido como esquema estrella, compuesto por una tabla de hechos y ocho tablas de dimensión. Esta decisión de diseño no es una preferencia estética: los esquemas dimensionales son el estándar reconocido en el campo de la analítica de datos precisamente porque optimizan el tipo de consultas que la plataforma necesita resolver agrupaciones y cruces por comuna, por tiempo, por tipo de delito, frente a un modelo completamente normalizado, que resulta más eficiente para transacciones individuales, pero más costoso computacionalmente para consultas analíticas agregadas (Hashem et al., 2015).

La tabla de hechos, denominada hecho delictivo, almacena un registro por cada caso reportado, con referencias a ocho tablas de dimensión: tipología del delito, delito específico, tipo de sitio, arma o medio empleado, móvil de la víctima y del agresor, comuna, barrio y rango de edad (curso de vida). A esta estructura inicial se sumaron dos tablas adicionales para las otras fuentes de datos: encuesta respuesta, que almacena cada respuesta del formulario de percepción ciudadana, y red_social_comentario, que almacena cada comentario extraído de YouTube junto con su clasificación de sentimiento.

Cabe destacar una decisión de diseño orientada a la protección de datos personales: en la tabla red_social_comentario no se almacena el nombre de usuario ni ningún otro dato que permita identificar al autor del comentario, únicamente el contenido textual público y el

resultado de su análisis. Esta decisión aplica el principio de minimización de datos, coherente con el marco legal descrito en el Capítulo 2, y responde a una consideración señalada por Kitchen (2014): la disponibilidad técnica de un dato no implica la obligación ni la conveniencia de recolectarlo en su totalidad.

4.3 Fase de implementación

4.3.1 *Extracción, transformación y carga de datos (ETL)*

Para poblar la base de datos a partir del conjunto de datos de delitos, se desarrolló un proceso de extracción, transformación y carga (ETL) mediante un script en Python que emplea la librería pandas. Este proceso realizó, entre otras, las siguientes tareas de limpieza y transformación:

- Conversión de tipos de datos (fechas, horas) desde el formato de texto original a tipos de fecha y hora nativos.
- Corrección de valores atípicos: se detectó un registro con un valor de edad negativo (-2), evidentemente un error de captura del dato original, el cual fue convertido a valor nulo en lugar de omitirse o corregirse arbitrariamente, dejando constancia explícita del ajuste realizado.
- Construcción de las ocho tablas de dimensión a partir de los valores únicos presentes en el conjunto de datos original, y posterior vinculación de cada uno de los 130.207 registros de la tabla de hechos con su dimensión correspondiente mediante llaves foráneas.
- Verificación de integridad post-carga, confirmando que ningún registro quedó sin vincular a sus dimensiones obligatorias.

Durante este proceso se identificaron y documentaron limitaciones inherentes a los datos de origen, consistentes con las dimensiones de calidad descritas por Vetrò et al. (2016) para datos abiertos gubernamentales, entre ellas la redundancia de la columna "localidad" que resultó ser la concatenación de otras dos columnas ya existentes y una inconsistencia en la nomenclatura de la comuna 7, referenciada indistintamente como "La Ciudadela" y "Ciudadela Real de Minas" en distintos periodos del conjunto de datos, la cual se documenta como limitación de los datos de origen sin alterar la información entregada por la entidad responsable.

Para la fuente de redes sociales, se desarrolló un segundo script que, mediante la API de YouTube, identifica videos relevantes del canal de Vanguardia a partir de un conjunto de palabras clave asociadas a seguridad y convivencia, extrae los comentarios públicos de dichos videos, y aplica un modelo de procesamiento de lenguaje natural preentrenado para clasificar el sentimiento de cada comentario en positivo, negativo o neutral, junto con un porcentaje de confianza de la clasificación.

4.3.2 Desarrollo del backend

El backend se desarrolló como una API REST con FastAPI, organizada en cuatro módulos funcionales: delitos, encuestas, redes sociales y un agente conversacional. En conjunto, estos módulos exponen dieciocho endpoints que permiten consultar los datos agregados desde distintas perspectivas (por comuna, por tipología, por tiempo, por rango horario, entre otras), registrar nuevas respuestas de encuesta, y generar respuestas e informes en lenguaje natural mediante inteligencia artificial.

Un aspecto que merece mención particular es la implementación de un mecanismo de protección de datos sensibles a nivel de código, no solo como enunciado de política. Al identificarse, durante la fase de diagnóstico, la presencia de categorías de delito relacionadas con menores de edad y delitos sexuales clasificadas como datos sensibles bajo la Ley 1581 de 2012, se implementó una regla que oculta automáticamente el conteo exacto de casos cuando una consulta desagregada por barrio, para una categoría sensible, arroja un número de casos por debajo de un umbral definido. Esta medida busca reducir el riesgo de que el cruce de una categoría sensible con una ubicación muy específica permita identificar indirectamente a una posible víctima, aplicando el principio de minimización de datos de forma operativa y verificable, y no meramente declarativa.

El agente conversacional se implementó mediante la técnica de llamado a funciones ("tool calling"), en la cual el modelo de lenguaje Gemini, de Google (Google LLC, 2024a), tiene acceso a un conjunto de funciones que corresponden a los propios endpoints de consulta de la plataforma. De esta manera, cuando un usuario formula una pregunta en lenguaje natural, el modelo determina qué función consultar, obtiene el dato real correspondiente, y redacta una respuesta en lenguaje natural fundamentada en ese dato, en lugar de generar una respuesta basada únicamente en el conocimiento con el que fue entrenado. Este mismo mecanismo se reutilizó para implementar la generación automática de un informe ejecutivo, estructurado en secciones de resumen general, hallazgos y recomendaciones, exportable en formato PDF.

Respecto de la operación del agente, se documentan las siguientes características técnicas y de gobernanza: el modelo empleado es gemini-3.6-flash, de Google, en su versión disponible a través de la API de Gemini al momento del desarrollo. Cada consulta al agente se

procesa de forma independiente, sin almacenamiento de conversaciones previas ni memoria entre solicitudes distintas; el sistema no conserva un historial persistente de las interacciones de los usuarios. Ante un error de comunicación con la API externa (por ejemplo, por indisponibilidad del servicio o superación de un límite de uso), el sistema no implementa actualmente un mecanismo de reintento automático ni mensajes de error diferenciados por tipo de falla; el frontend captura la excepción y muestra al usuario un mensaje genérico de error de conexión. La herramienta se opera bajo el nivel gratuito de la API de Gemini, con los límites de uso diario que dicho nivel establece. Es responsabilidad de trabajos futuros incorporar límites explícitos de tokens por respuesta, un registro de auditoría de las consultas realizadas, y pruebas sistemáticas frente a consultas adversariales orientadas a extraer del agente conteos protegidos por el mecanismo de datos sensibles descrito en este mismo apartado. Se advierte, además, que las respuestas del agente relativas a percepción ciudadana o sentimiento en redes sociales heredan las limitaciones de representatividad de sus fuentes de datos, descritas en el Capítulo 5, y no deben interpretarse como evidencia estadísticamente representativa de la ciudadanía.

4.3.3 Desarrollo del frontend

El panel de visualización (dashboard) se desarrolló como una aplicación web de una sola página, sin frameworks de JavaScript adicionales, empleando la librería Chart.js (Chart.js Contributors, 2013) para la generación de gráficas interactivas. El panel se organiza en cuatro secciones temáticas panorama general de delitos, análisis temporal y de perfil de los casos, percepción ciudadana y redes sociales, y consulta al agente conversacional, e incluye filtros interactivos que permiten segmentar la información por año.

Cada gráfica del panel se validó individualmente contra los datos crudos antes de incorporarse a la versión final del sistema, siguiendo un procedimiento de verificación cruzada: para cada endpoint nuevo, se sumó manualmente el total de casos devuelto y se comparó contra el total general de 130.207 registros, confirmando en todos los casos una coincidencia exacta que descarta la duplicación o pérdida de información durante el proceso de agregación.

Adicionalmente, se implementó un formulario público independiente para la recolección de respuestas de la encuesta de percepción ciudadana, con validación de los campos requeridos y confirmación visual del registro exitoso de cada respuesta.

4.3.4 Despliegue

Con el propósito de que la plataforma fuera accesible públicamente requisito explícito del producto "plataforma web funcional y desplegada" definido en la propuesta de trabajo de grado, se adoptó una arquitectura de despliegue con servicios separados para el backend y la base de datos. Esta decisión se tomó tras identificar que los servicios de base de datos gratuitos ofrecidos por algunas plataformas de alojamiento web caducan tras un periodo determinado (usualmente 30 días), lo cual resultaba incompatible con la necesidad de conservar los datos durante todo el periodo de desarrollo y validación del proyecto.

En consecuencia, la base de datos se alojó en el servicio Neon (Neon, Inc., 2021), que ofrece PostgreSQL gestionado. Con un nivel gratuito sin fecha de expiración mientras que el backend se desplegó en el servicio Render (Render Services, Inc., 2019), configurado para desplegarse automáticamente ante cada actualización del repositorio de código fuente alojado en GitHub. La plataforma resultante es accesible públicamente en la dirección <https://cultura->

ciudadana-bucaramanga.onrender.com, sin restricciones de registro o autenticación para su consulta.

El código fuente completo del proyecto se encuentra disponible públicamente en el repositorio <https://github.com/JairAlvaradoB/cultura-ciudadana-bucaramanga>, bajo la rama principal (main). La versión específica evaluada en este documento corresponde al commit c7523ee4829aa6c7aa7b7676912c05f16616fb66. El repositorio incluye una nota de licenciamiento (LICENSE.md) que reserva los derechos del trabajo mientras se define, en conjunto con la institución, la política formal de propiedad intelectual aplicable a este trabajo de grado.

Se verificó la disponibilidad y correcto funcionamiento de la plataforma pública, incluyendo el panel de visualización, el formulario de encuesta, el asistente conversacional y la descarga del informe en PDF, el 10 de septiembre de 2026, confirmando su operatividad más allá del entorno de desarrollo local

4.4 Fase de validación

La validación de la operatividad de la plataforma, correspondiente al Objetivo específico 4, se llevó a cabo mediante dos procedimientos complementarios, enfocados principalmente en la característica de adecuación funcional descrita por la norma ISO/IEC 25010 (International Organization for Standardization [ISO], 2011) para la evaluación de calidad de software.

En primer lugar, se realizó una verificación técnica exhaustiva de cada uno de los dieciocho endpoints de la API, contrastando los resultados devueltos contra los datos de origen mediante sumas de control manuales, tal como se describió en el apartado 4.3.3.

Esta verificación permitió identificar y corregir, antes de su publicación, errores de referencias entre módulos que de otro modo habrían generado resultados incorrectos sin manifestarse como fallas visibles del sistema.

En segundo lugar, se realizó una prueba piloto de recolección de datos reales mediante la difusión del formulario de encuesta de percepción ciudadana mencionado en el apartado 4.3.3, ya desplegado en su dirección pública, a personas externas al equipo de desarrollo del proyecto.

Tabla 2. Matriz de pruebas de los endpoints de la API

N.º	Endpoint	Entrada / parámetros de prueba	Resultado esperado	Resultado obtenido	Tipo
1	GET /delitos/resumen-por-comuna	Sin parámetros (consulta general)	Lista de 21 comunas/corregimientos con total de casos, suma igual a 130.207	Devolvió 21 registros; suma verificada manualmente = 130.207. Coincidencia exacta	Positiva
2	GET /delitos/resumen-por-tipologia	Sin parámetros	Lista de tipologías con total de casos, verificable contra el dashboard	5 tipologías devueltas; total coincide con KPI del panel (130.207)	Positiva
3	GET /delitos/resumen-temporal	Sin parámetros	Serie de casos por año y mes, 2016-2026	Serie completa devuelta; permitió identificar caída real en 2020 (pandemia) y año 2026 incompleto	Positiva
4	GET /delitos/resumen-por-barrio	tipologia = "SEXUALES" (categoría sensible)	Barrios con menos de 5 casos de esa categoría deben devolver total_casos = null y dato_protegido = true	Barrios como "1 DE MAYO" y "20 DE JULIO" devueltos como null/protegido; barrio "12 DE OCTUBRE" con 14 casos (por encima del umbral) devuelto sin protección, verificado descargando el JSON completo	Positiva (prueba de seguridad/límite)

N.º	Endpoint	Entrada / parámetros de prueba	Resultado esperado	Resultado obtenido	Tipo
5	GET /delitos/resumen- por-rango-horario	Sin parámetros	24 franjas horarias con total de casos, suma = 130.207	24 registros devueltos; suma verificada = 130.207, coincidencia exacta con el resultado del caso 1	Positiva
6	GET /delitos/rango- horario-por-tipologia	Sin parámetros	Cruce de rango horario y tipología; la suma de una franja específica debe coincidir con el caso 5	Franja "0:00-0:59": suma de las 5 tipologías = 5.571, idéntico al valor devuelto en el caso 5 para esa misma franja	Positiva (validación cruzada)
7	GET /delitos/tendencia- mensual-por-tipologia	Sin parámetros	Serie mensual por tipología, consistente con el total general	Datos devueltos correctamente; utilizados en el gráfico de tendencia mensual del panel sin discrepancias observadas	Positiva
8	GET /delitos/resumen- por-sexo	Sin parámetros	Total de casos por sexo de la víctima	3 categorías devueltas (masculino, femenino, no disponible); consumido correctamente por el panel	Positiva
9	GET /delitos/resumen- por-movil-victima	Sin parámetros	14 categorías de móvil, suma = 130.207	14 registros devueltos, suma verificada = 130.207. Reveló que el 90 % de las víctimas estaban "a pie"	Positiva
10	GET /delitos/piramide- edad-sexo	Sin parámetros	37 combinaciones de rango de edad y sexo, suma = 130.207	37 registros devueltos; suma verificada manualmente = 130.207, coincidencia exacta	Positiva
11	GET /encuestas/opciones- comuna	Sin parámetros	Lista de 23 comunas/corregimientos para el formulario	23 registros devueltos con id, nombre y número; usados para poblar el formulario público	Positiva

N.º	Endpoint	Entrada / parámetros de prueba	Resultado esperado	Resultado obtenido	Tipo
12	GET /encuestas/opciones- rango-edad	Sin parámetros	19 rangos de edad para el formulario	19 registros devueltos correctamente	Positiva
13	POST /encuestas/	Payload válido con comuna_id, curso_vida_id, sexo y escalas 1-5	Código 201, registro creado con id y fecha_respuesta	Código 201 confirmado en pruebas por /docs y desde el formulario público en producción	Positiva
14	POST /encuestas/ (validación de escalas)	Envío de un valor fuera del rango 1-5 en un campo de escala	El código valida internamente el rango (función validar_escala) y debería responder 422	No se ejecutó una solicitud real con datos inválidos durante las pruebas documentadas en este trabajo; la validación existe en el código fuente pero queda pendiente como prueba explícita a ejecutar en trabajos futuros	Pendiente de ejecución
15	GET /encuestas/resumen- por-comuna	Sin parámetros	Promedios de percepción por comuna con al menos 1 respuesta registrada	Devolvió 8 comunas con 31 respuestas en total; promedios ponderados verificados manualmente (ver Capítulo 5)	Positiva
16	GET /redes- sociales/resumen- sentimiento	Sin parámetros	Total de comentarios por categoría de sentimiento, suma = 137	3 categorías devueltas (neutral, negativo, positivo), suma = 137. Verificado también por consulta directa a la base de datos (psql)	Positiva
17	GET /redes- sociales/comentarios- recientes	limite = 15	15 comentarios con su clasificación de sentimiento y porcentaje de confianza	15 comentarios devueltos, con texto, sentimiento y confianza coherentes con una revisión manual de una muestra	Positiva
18	POST /chat/	mensaje = "¿Qué comuna tiene más casos de delitos?"	Respuesta en lenguaje natural con la comuna correcta y cifra exacta, obtenida mediante	Respondió correctamente "Oriental, 15.822 casos", cifra idéntica	Positiva

N.º	Endpoint	Entrada / parámetros de prueba	Resultado esperado	Resultado obtenido	Tipo
			consulta real a la base de datos (no generada por el modelo)	a la mostrada en el panel de visualización	
19	GET /chat/informe	Sin parámetros	Documento PDF descargado con seis secciones (resumen, tipología, percepción, sentimiento, hallazgos, recomendaciones)	PDF generado y descargado correctamente en un tiempo de 20 a 40 segundos, con las seis secciones presentes	Positiva
20	Manejo de errores no controlado (evidencia de desarrollo)	Falta de importación del modelo Movil en el módulo de rutas	El sistema debe fallar de forma visible (no silenciosa) ante un error de programación	Error 500 devuelto correctamente y visible en el registro del servidor; corregido en el mismo ciclo de desarrollo mediante la importación faltante	Negativa (detectada y corregida)

Nota. Elaborada por el autor a partir de las pruebas realizadas mediante la documentación interactiva de la API (/docs) y mediante verificación cruzada de sumas de control contra los datos de origen. El caso 14 se documenta explícitamente como pendiente de ejecución, en cumplimiento del principio de no reportar como ejecutadas pruebas que no se realizaron durante el desarrollo de este trabajo.

5. RESULTADOS

En este capítulo se presentan los resultados obtenidos durante el desarrollo del trabajo de grado, organizados según los cuatro objetivos específicos planteados. Cada apartado incluye tanto la evidencia cuantitativa disponible como un análisis cualitativo de su alcance y sus limitaciones.

5.1. Resultados del diagnóstico e integración de fuentes de datos

Como resultado de la fase de diagnóstico, se logró la integración efectiva de tres fuentes de datos heterogéneas en un modelo de datos único: 130.207 registros de delitos provenientes del portal de Datos Abiertos de Bucaramanga, distribuidos en 21 comunas y corregimientos con datos disponibles; 31 respuestas de encuesta de percepción ciudadana, recolectadas en 8 comunas reales; y 137 comentarios de YouTube analizados con inteligencia artificial. La integración de estas tres fuentes en un modelo dimensional común permite consultar simultáneamente el dato oficial, la percepción ciudadana y la opinión pública espontánea sobre un mismo fenómeno, lo cual constituye un resultado directamente vinculado a la pregunta de investigación planteada en el Capítulo 1. Esta afirmación se circunscribe a las plataformas revisadas en el estado del arte (Capítulo 1, apartado 1.4) y no constituye una revisión sistemática exhaustiva del conjunto de soluciones existentes a nivel local, nacional o internacional.

Un resultado adicional de esta fase, de carácter metodológico, fue la identificación de limitaciones reales en los datos de origen que no habrían sido evidentes sin un proceso de verificación exhaustivo: un registro con edad negativa, una columna redundante, y una inconsistencia en la nomenclatura de la comuna 7 a lo largo del tiempo. Estos hallazgos se

documentan como parte de los resultados del proceso, en tanto evidencian que la calidad de los datos abiertos institucionales, aunque generalmente alta, no está exenta de imperfecciones que deben tratarse explícitamente en cualquier análisis derivado de ellos.

5.2. Resultados del desarrollo de la plataforma

La plataforma desarrollada expone dieciocho endpoints de consulta distribuidos en cuatro módulos (delitos, encuestas, redes sociales y agente conversacional), y un panel de visualización con siete gráficas interactivas verificadas. La totalidad de los endpoints fue sometida a una prueba de suma de control, comparando el total de casos devuelto por cada consulta agregada contra el total general de 130.207 registros; en todos los casos evaluados se obtuvo una coincidencia exacta, lo cual constituye evidencia directa de la integridad del proceso de agregación de datos implementado.

La plataforma se encuentra desplegada y accesible públicamente en <https://cultura-ciudadana-bucaramanga.onrender.com>, cumpliendo con el producto "plataforma web funcional y desplegada" definido en la propuesta de trabajo de grado. Como resultado adicional no contemplado originalmente en el alcance mínimo del proyecto, se implementó un agente conversacional basado en inteligencia artificial capaz de responder preguntas en lenguaje natural consultando los datos reales de la plataforma, así como un generador automático de informes ejecutivos exportables en formato PDF.

5.3. Resultados de la prueba piloto: encuesta de percepción ciudadana

Durante el periodo de recolección se obtuvieron 31 respuestas válidas, distribuidas en 8 comunas del área metropolitana de Bucaramanga, según se detalla en la Tabla 2.

Tabla 3. *Distribución de respuestas de encuesta y promedios de percepción por comuna*

Comuna	Numero de Respuestas	Seguridad de día (1-5)	Seguridad de noche (1-5)	Mal uso del espacio público (1-5)
García Rovira	11	3,91	3,00	4,27
Occidental	8	4,25	3,75	3,12
San Francisco	3	3,67	2,00	3,67
Oriental	3	4,33	4,33	4,00
Cabecera del Llano	2	4,00	2,50	3,50
Provenza	1	5,00	4,00	4,00
Mutis	1	4,00	3,00	3,00
Lagos del Cacique	1	5,00	5,00	5,00

Nota. Elaborada por el autor a partir de los datos registrados en la plataforma. Escala: 1 = percepción más desfavorable, 5 = percepción más favorable (para seguridad); en mal uso del espacio público, 5 indica mayor frecuencia percibida de uso inadecuado.

El promedio ponderado general, calculado sobre el total de 30 respuestas correspondientes a comunas identificadas (excluyendo una respuesta registrada como "no disponible"), fue de 4,10 sobre 5 para la percepción de seguridad diurna, 3,30 sobre 5 para la percepción de seguridad nocturna, y 3,80 sobre 5 para la percepción de mal uso del espacio público. La diferencia de 0,80 puntos entre la percepción de seguridad diurna y nocturna es consistente con el hallazgo obtenido en el análisis de los datos oficiales de delitos (Capítulo 5, apartado 5.4), en el cual las franjas horarias de mayor concentración de casos no corresponden a la madrugada sino a la media mañana y al inicio de la noche, lo que sugiere que la percepción ciudadana de mayor riesgo nocturno responde a factores adicionales a la sola frecuencia de casos, como la menor visibilidad o la menor afluencia de personas en el espacio público durante esas horas.

Se identifica como limitación relevante que la encuesta constituye una muestra piloto de tipo no probabilístico, recolectada por conveniencia a través de canales informales de difusión,

y no un muestreo aleatorio o estratificado representativo de la población del área metropolitana de Bucaramanga. La cobertura lograda (8 de 23 comunas y corregimientos) no permite generalizar los resultados de percepción al conjunto del municipio, y cinco de las ocho comunas con respuesta cuentan con menos de cinco registros, umbral por debajo del cual la propia plataforma advierte que no se trata de una muestra representativa. En consecuencia, los resultados presentados en este apartado deben interpretarse como una validación inicial de la operatividad de la plataforma, y no como una medición representativa de la percepción ciudadana en Bucaramanga. Esta limitación se retoma en el Capítulo 7 como recomendación para la continuidad del proyecto.

5.4. Resultados del análisis de sentimiento en redes sociales

Del total de 137 comentarios analizados, extraídos de videos de un único canal de YouTube (el periódico Vanguardia) relacionados con seguridad y convivencia en Bucaramanga, 53 (38,7 %) fueron clasificados como neutrales, 49 (35,8 %) como negativos y 35 (25,5 %) como positivos, mediante el modelo preentrenado robertuito-sentiment-analysis (versión disponible en el repositorio Hugging Face al momento de la consulta), documentado en el Apéndice B. Estos resultados deben entenderse como un análisis exploratorio de sentimiento y no como una medición definitiva o representativa de la opinión ciudadana, por al menos tres razones: primero, el modelo no fue validado contra una muestra de comentarios clasificados manualmente por un evaluador humano, por lo que no se dispone de métricas de exactitud, precisión o sensibilidad para este dominio específico; segundo, la muestra proviene de un único canal y de un criterio de ordenamiento por relevancia definido por la propia plataforma de YouTube, no de un muestreo aleatorio de la conversación pública; y tercero, quienes comentan en videos de noticias no son necesariamente representativos del conjunto de la ciudadanía. La

proporción combinada de comentarios neutrales y negativos (74,5 %) es notablemente superior a la de comentarios positivos, lo cual es consistente con la naturaleza de las noticias de origen relacionadas mayoritariamente con hechos delictivos, aunque no se explica únicamente por un rechazo emocional: una revisión cualitativa de los comentarios clasificados como neutrales muestra que una proporción de ellos corresponde a reacciones informativas o descriptivas más que a expresiones de carga emocional, lo que constituye una observación adicional sobre los límites de la clasificación automática de sentimiento aplicada a este tipo de contenido.

5.5. Análisis integrado de resultados

Al cruzar los resultados de las tres fuentes de datos para las comunas con información disponible en las tres, se observa un patrón parcialmente consistente entre percepción y realidad, con al menos una divergencia relevante que merece destacarse. San Francisco, la segunda comuna con mayor número de casos de delitos registrados en el periodo 2016-2026 (15.232 casos), presenta también una de las percepciones de seguridad nocturna más bajas de la muestra (2,00 sobre 5), lo cual es consistente con lo esperado. Sin embargo, Oriental, la comuna con mayor número de casos registrados en todo el periodo (15.822 casos), presenta entre sus respondientes una percepción de seguridad nocturna alta (4,33 sobre 5). Esta divergencia, aunque basada en una muestra reducida de tres respuestas para dicha comuna, ilustra el tipo de hallazgo que la integración de fuentes hace posible identificar y que ninguna fuente por separado habría revelado: la frecuencia oficial de delitos no siempre se traduce en una percepción ciudadana proporcional de inseguridad, lo cual es coherente con la advertencia de Kitchin (2014) sobre los límites de interpretar los datos urbanos como un reflejo directo y completo de la realidad social que pretenden representar.

5.6. Síntesis de cumplimiento de los objetivos específicos

El Objetivo específico 1 (diagnóstico de requerimientos e indicadores) se cumplió mediante la identificación y caracterización documentada de las tres fuentes de datos y la definición de seis indicadores de cultura ciudadana, descritos en el apartado 4.1.

El Objetivo específico 2 (diseño de la arquitectura) se cumplió mediante el diseño e implementación de un modelo dimensional de datos y una arquitectura de tres capas, verificable en el código fuente del proyecto y documentada en el apartado 4.2.

El Objetivo específico 3 (implementación de los componentes funcionales) se cumplió mediante el desarrollo y despliegue público de dieciocho endpoints y un panel de visualización interactivo, con evidencia de verificación técnica documentada en el apartado 4.3.

El Objetivo específico 4 (validación mediante prueba piloto) se cumplió mediante la verificación técnica de los dieciocho endpoints y la recolección de 31 respuestas reales de encuesta en 8 comunas, si bien con la limitación de cobertura geográfica señalada en el apartado 5.3, que se retoma como recomendación para trabajos futuros.

5.7. Valoración del éxito del trabajo realizado

Con el propósito de valorar de manera explícita el éxito del trabajo desarrollado, y no únicamente su cumplimiento formal frente a los objetivos planteados, se establecieron a priori tres criterios de éxito, evaluados a continuación de forma cuantitativa.

El primer criterio, la integridad de la integración de datos, se cumplió en su totalidad: los 130.207 registros de delitos se integraron sin pérdida ni duplicación, condición verificada

mediante sumas de control en los dieciocho endpoints de la plataforma, todos los cuales arrojaron una coincidencia exacta frente al total de referencia.

El segundo criterio, la disponibilidad pública y estable de la plataforma, se cumplió: el sistema se encuentra desplegado y accesible sin restricciones desde la fase de implementación del proyecto, con una arquitectura que garantiza la persistencia permanente de los datos, a diferencia de alternativas de alojamiento gratuito que habrían comprometido la continuidad del proyecto tras treinta días.

El tercer criterio, la obtención de una muestra mínima viable de percepción ciudadana (definida internamente en un rango de 25 a 30 respuestas), se cumplió y se superó levemente, con 31 respuestas recolectadas en 8 comunas distintas del área metropolitana.

En términos cualitativos, el principal indicador de éxito del trabajo es la posibilidad, demostrada en el apartado 5.5, de identificar un hallazgo la divergencia entre la frecuencia oficial de delitos y la percepción ciudadana de seguridad en la comuna Oriental que ninguna de las tres fuentes de datos habría permitido evidenciar por sí sola. Este resultado constituye evidencia directa de que el propósito central del trabajo, formulado en la pregunta de investigación del Capítulo 1, se alcanzó de manera efectiva, aun cuando el tamaño de la muestra de percepción ciudadana recomienda tratar este hallazgo específico con cautela y profundizarlo en fases posteriores del proyecto.

6. CONCLUSIONES

Antes de presentar las conclusiones específicas, conviene precisar el alcance real de los hallazgos derivados de la encuesta de percepción ciudadana y del análisis de sentimiento en redes sociales. Dado que ambas fuentes se basan en muestras no probabilísticas una encuesta piloto de alcance geográfico parcial y un conjunto de comentarios de un único canal de YouTube, la plataforma desarrollada no debe presentarse como un instrumento de medición representativa de la totalidad de la ciudadanía del área metropolitana de Bucaramanga en su estado actual. Las conclusiones que siguen se formulan teniendo en cuenta esta limitación, y se refieren a la viabilidad técnica y metodológica demostrada por el proyecto, no a la generalización estadística de los datos de percepción recolectados hasta el momento.

La integración de fuentes de datos de naturaleza distinta un registro administrativo estructurado, una encuesta de percepción y contenido textual espontáneo generado en redes sociales resultó viable gracias a que el diseño de la base de datos no intentó forzar una estructura homogénea única, sino que empleó un modelo dimensional capaz de acomodar cada fuente según su propia naturaleza y vincularlas posteriormente a través de dimensiones comunes, como la comuna y el tiempo. Esto permite concluir que la dificultad habitual de integrar big data urbano heterogéneo no reside tanto en la tecnología disponible como en la decisión de diseño de datos que se adopte desde el principio; un modelo mal escogido habría obligado a sacrificar información propia de cada fuente para lograr una compatibilidad artificial entre ellas.

Durante el desarrollo se presentaron dos limitaciones externas que obligaron a modificar el planteamiento original de la propuesta de trabajo de grado: la inviabilidad de instrumentar sensores IoT propios dentro del alcance de un trabajo de grado de pregrado, y la restricción reciente de acceso gratuito a las interfaces de programación de las principales redes sociales. En ambos casos, la solución adoptada no consistió en sustituir la fuente por cualquier alternativa disponible, sino en identificar una fuente distinta que preservara la misma función conceptual que la fuente original cumplía dentro del diseño de la investigación un registro objetivo de comportamiento urbano en el primer caso, y percepción ciudadana espontánea en el segundo. Esto permite concluir que la flexibilidad metodológica frente a restricciones externas no debe entenderse como una renuncia a los objetivos planteados, sino como una condición necesaria para sostenerlos, siempre que el reemplazo se justifique en función del papel que la fuente original cumplía y no únicamente en su disponibilidad.

El proceso de limpieza de los datos abiertos institucionales reveló que, aun tratándose de una fuente oficial y de calidad general alta, dicha información no está exenta de inconsistencias como un valor de edad inválido, una columna redundante o una nomenclatura territorial cambiante en el tiempo, las cuales solo se hicieron evidentes mediante una revisión activa y no automática de los datos. De ello se concluye que la disponibilidad pública de un conjunto de datos no garantiza por sí sola su calidad, y que cualquier trabajo que pretenda construir conocimiento a partir de datos abiertos debe incorporar una fase explícita de verificación, en lugar de asumir que el origen institucional de la información exime de dicha revisión.

El diseño de un mecanismo de protección de datos sensibles a nivel de código, y no solo como enunciado de política, condujo a concluir que las consideraciones éticas de un proyecto que trata información relacionada con personas no deben limitarse a una declaración de principios en el documento escrito, sino traducirse en decisiones técnicas verificables dentro del propio sistema. La alternativa de omitir esta protección habría sido técnicamente más simple, pero habría trasladado el riesgo de identificación indirecta de víctimas al usuario final de la plataforma, lo cual resulta inaceptable para una herramienta orientada a la gestión pública.

El hallazgo de una divergencia entre la frecuencia oficial de delitos y la percepción ciudadana de seguridad en al menos una de las comunas analizadas permite concluir que estas dos variables, aunque relacionadas, no deben tratarse como equivalentes ni como sustitutas la una de la otra en el diseño de indicadores de cultura ciudadana. La razón de esta divergencia probablemente no obedece a un error de medición en ninguna de las dos fuentes, sino a que la percepción ciudadana se construye a partir de la experiencia personal, la exposición mediática y la interacción social cotidiana, factores que no necesariamente guardan una relación proporcional con la frecuencia estadística de un fenómeno. Esta conclusión refuerza la pertinencia de mantener ambas fuentes integradas en la plataforma, en lugar de privilegiar una sobre la otra como indicador único de cultura ciudadana.

La cobertura geográfica desigual alcanzada en la recolección de respuestas de encuesta, concentrada en las comunas donde el autor y su red de contactos inmediatos tienen mayor presencia, evidencia una limitación propia de los canales informales de difusión empleados durante la prueba piloto. De ello se concluye que la representatividad territorial de un instrumento de participación ciudadana depende en gran medida de contar con canales de difusión institucionales, y no únicamente de que la plataforma tecnológica esté disponible y en funcionamiento; una plataforma correctamente desplegada no garantiza, por sí sola, una muestra geográficamente equilibrada si no se acompaña de una estrategia de difusión con alcance comparable en las distintas zonas del municipio.

Finalmente, la implementación de un agente conversacional que consulta los propios datos de la plataforma en lugar de generar respuestas basadas únicamente en el conocimiento previo del modelo de lenguaje permite concluir que la utilidad de la inteligencia artificial en un sistema orientado a la gestión pública depende menos de la sofisticación del modelo empleado que de su anclaje verificable a datos reales; un agente sin este anclaje habría podido generar respuestas plausibles pero no necesariamente ciertas, lo cual resulta especialmente riesgoso en un dominio, como el de la seguridad y la convivencia ciudadana, donde las decisiones informadas por datos erróneos pueden tener consecuencias sobre la asignación de recursos públicos.

7. RECOMENDACIONES

Se recomienda ampliar la estrategia de difusión del instrumento de encuesta de percepción ciudadana más allá de los canales informales empleados durante la prueba piloto de este trabajo de grado, vinculando canales institucionales como la Alcaldía de Bucaramanga, las Juntas de Acción Comunal de cada comuna, o la propia comunidad académica de las Unidades Tecnológicas de Santander. Esta ampliación permitiría alcanzar una cobertura geográfica más equilibrada entre las veintitrés comunas y corregimientos del área metropolitana, condición necesaria para que los indicadores de percepción ciudadana adquieran representatividad estadística suficiente para fundamentar decisiones de gestión pública.

Se recomienda automatizar la actualización periódica de los datos que alimentan la plataforma, en particular el conjunto de datos de delitos del portal de Datos Abiertos y los comentarios extraídos de YouTube, mediante tareas programadas que ejecuten los procesos de extracción, transformación y carga sin intervención manual. Actualmente estos procesos se ejecutan de forma manual, lo cual es adecuado para la fase de desarrollo y validación del proyecto, pero limita la vigencia de la información en un escenario de operación continua.

Se recomienda diversificar las fuentes de percepción ciudadana en redes sociales más allá de un único canal de noticias, incorporando otros medios de comunicación locales o, si las condiciones de acceso y presupuesto lo permiten en el futuro, retomar la

evaluación de las interfaces de programación de aplicaciones de plataformas como X o Meta, cuyo acceso gratuito no estuvo disponible durante el desarrollo de este trabajo. Ampliar la diversidad de fuentes reduciría el riesgo de que la percepción capturada esté condicionada por la línea editorial o el enfoque particular de un solo medio.

Se recomienda que, una vez alcanzada una muestra de encuesta de mayor tamaño y cobertura, se complementen los análisis descriptivos presentados en este trabajo con técnicas de estadística inferencial que permitan establecer si las diferencias de percepción observadas entre comunas son estadísticamente significativas, y no únicamente diferencias descriptivas entre promedios calculados sobre muestras pequeñas.

Se recomienda profundizar cualitativamente en las divergencias identificadas entre la frecuencia oficial de delitos y la percepción ciudadana de seguridad mediante técnicas complementarias, como grupos focales o entrevistas semiestructuradas en las comunas donde dicha divergencia resulte más marcada, con el fin de comprender los factores que la explican y que un análisis exclusivamente cuantitativo no permite capturar en profundidad.

Se recomienda a las Unidades Tecnológicas de Santander y a los grupos de investigación institucionales considerar la apropiación del modelo metodológico desarrollado en este trabajo la integración de datos administrativos, de percepción ciudadana y de redes sociales mediante un modelo dimensional común como base replicable para el monitoreo de otros indicadores urbanos de interés institucional, distintos

a la cultura ciudadana, dado que la arquitectura desarrollada no depende conceptualmente del dominio específico aplicado en este trabajo.

Finalmente, se recomienda dar seguimiento a la evolución del marco normativo colombiano en materia de protección de datos personales, en particular al proyecto de reforma a la Ley 1581 de 2012 que se encontraba en trámite en el Congreso de la República al momento de la elaboración de este documento, e incorporar oportunamente los ajustes que dicha reforma pudiera introducir en los mecanismos de tratamiento y protección de datos sensibles implementados en la plataforma.

8. REFERENCIAS BIBLIOGRAFICAS

Chart.js Contributors. (2013). *Chart.js*. <https://www.chartjs.org>

Congreso de la República de Colombia. (1993). *Ley 99 de 1993: Por la cual se crea el Ministerio del Medio Ambiente, se reordena el sector público encargado de la gestión y conservación del medio ambiente y los recursos naturales renovables, se organiza el Sistema Nacional Ambiental (SINA) y se dictan otras disposiciones*.

Congreso de la República de Colombia. (2012). *Ley Estatutaria 1581 de 2012: Por la cual se dictan disposiciones generales para la protección de datos personales*.

Congreso de la República de Colombia. (2014). *Ley 1712 de 2014: Por medio de la cual se crea la Ley de Transparencia y del Derecho de Acceso a la Información Pública Nacional*.

Constitución Política de Colombia. (1991). *Artículo 15 [Título II, Capítulo 1]*.

Departamento Nacional de Planeación. (2018). *Documento CONPES 3920: Política nacional de explotación de datos (Big Data)*.

<https://colaboracion.dnp.gov.co/CDT/Conpes/Econ%C3%B3micos/3920.pdf>

Google LLC. (2024a). *Gemini API*. <https://ai.google.dev/gemini-api>

Google LLC. (2024b). *YouTube Data API*. <https://developers.google.com/youtube/v3>

Hashem, I. A. T., Chang, V., Anuar, N. B., Adewole, K., Yaqoob, I., Gani, A., Ahmed, E., & Chiroma, H. (2015). The role of Big Data in smart city. *International Journal of Information Management*, 36(5), 748-758.
<https://doi.org/10.1016/j.ijinfomgt.2016.05.002>

Hernández-Sampieri, R., Fernández-Collado, C., & Baptista-Lucio, P. (2014). *Metodología de la investigación* (6.^a ed.). McGraw-Hill.

International Organization for Standardization. (2011). *ISO/IEC 25010:2011—Systems and software engineering—Systems and software quality requirements and evaluation (SQuaRE)—System and software quality models*.
<https://www.iso.org/standard/35733.html>

Kitchin, R. (2014). The real-time city? Big data and smart urbanism. *GeoJournal*, 79(1), 1-14. <https://doi.org/10.1007/s10708-013-9516-8>

Koirala, L. (2025). Big data analytics for smart cities: Optimizing urban traffic management using real-time data processing. *Journal of Computer Science and Technology Application*. <https://journal.corisinta.org/corisinta/article/view/56>

Mockus, A. (2001). *Cultura ciudadana, programa contra la violencia en Bogotá*. Alcaldía Mayor de Bogotá.

Neon, Inc. (2021). Neon. <https://neon.tech>

Pérez, J. M., Furman, D. A., Alonso Alemany, L., & Luque, F. M. (2022). RoBERTuito: A pre-trained language model for social media text in Spanish. *Proceedings of the Thirteenth Language Resources and Evaluation Conference (LREC 2022)*.
<https://aclanthology.org/2022.lrec-1.785/>

Presidencia de la República de Colombia. (2013). *Decreto 1377 de 2013: Por el cual se reglamenta parcialmente la Ley 1581 de 2012*.

Presidencia de la República de Colombia. (2015). *Decreto 1076 de 2015: Por medio del cual se expide el Decreto Único Reglamentario del Sector Ambiente y Desarrollo Sostenible*.

Ramírez, S. (2018). FastAPI. <https://fastapi.tiangolo.com>

Render Services, Inc. (2019). Render. <https://render.com>

Southern Maryland Chronicle. (2025, 9 de octubre). *Charles debuts mobile reporting system*. <https://southernmarylandchronicle.com/2025/10/09/charles-debuts-mobile-reporting-system/>

SQLAlchemy Contributors. (2006). *SQLAlchemy*. <https://www.sqlalchemy.org>

Townsend, A. M. (2013). *Smart cities: Big data, civic hackers, and the quest for a new utopia*. W. W. Norton & Company.

Valenzuela, S. (2015). *Understanding social media use in Latin America*.
https://www.researchgate.net/publication/281791236_Understanding_Social_Media_Use_in_Latin_America

Vetrò, A., Canova, L., Torchiano, M., Minotas, C. O., Iemma, R., & Morando, F. (2016). Open data quality measurement framework: Definition and application to Open Government Data. *Government Information Quarterly*, 33(2), 325-337.
<https://doi.org/10.1016/j.giq.2016.02.001>

9. APÉNDICES

APÉNDICE A. INSTRUMENTO DE ENCUESTA DE PERCEPCIÓN CIUDADANA

Encuesta de Percepción Ciudadana — Área Metropolitana de Bucaramanga}

Instrucciones para el respondiente: Sus respuestas son anónimas y se usan únicamente con fines académicos y de análisis agregado.

1. ¿En qué comuna vives? Selección única entre las 23 comunas y corregimientos del área metropolitana de Bucaramanga (lista desplegable poblada dinámicamente desde la base de datos de la plataforma).
2. Rango de edad Selección única entre los siguientes rangos: 0-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85 o más.
3. Sexo Selección única: Masculino / Femenino / Prefiero no decir.
4. ¿Qué tan seguro te sientes caminando en tu barrio de día? Escala de opción única de 1 a 5, donde 1 = nada seguro y 5 = muy seguro.
5. ¿Qué tan seguro te sientes caminando en tu barrio de noche? Escala de opción única de 1 a 5, donde 1 = nada seguro y 5 = muy seguro.
6. ¿Con qué frecuencia observas mal uso del espacio público (basura, invasión de andenes, etc.)? Escala de opción única de 1 a 5, donde 1 = nunca y 5 = siempre.
7. ¿Has participado en alguna actividad comunitaria en el último año? Selección única: Sí / No.

8. Comentario adicional (opcional) Campo de texto libre: "¿Algo más que quieras contarnos sobre tu barrio?"

Nota. Instrumento de elaboración propia, aplicado en formato electrónico a través de la plataforma web desarrollada en este trabajo de grado, disponible en <https://cultura-ciudadana-bucaramanga.onrender.com/dashboard/encuesta.html>. El diseño de las preguntas 4, 5 y 6 se fundamentó en las dimensiones de convivencia y cuidado de lo público descritas por Mockus (2001), citado en el Capítulo 2 de este documento.

APÉNDICE B. PROTOCOLO DE EXTRACCIÓN DE COMENTARIOS DE REDES SOCIALES

Fuente de datos: canal de YouTube del periódico Vanguardia, playlist "Noticias de Bucaramanga" (identificador de playlist: PL0z3JydD83hZGbAA1SQsPEjjurPhR26pK).

Criterio de selección de videos: se revisaron los videos de la playlist en orden cronológico, filtrando por la presencia de al menos una de las siguientes palabras clave en el título del video: seguridad, inseguridad, hurto, robo, convivencia, espacio público, andén, policía, delito, violencia, atraco, homicidio, ciudadano.

Parámetros de extracción:

- Máximo de videos relevantes procesados por ejecución: 15.
- Máximo de comentarios extraídos por video: 30.
- Criterio de ordenamiento de comentarios solicitado a la API: relevancia.

- Comentarios excluidos del análisis: aquellos vacíos o con una extensión superior a 500 caracteres, por exceder el límite de procesamiento del modelo de clasificación empleado.

Herramienta de extracción: YouTube Data API v3 (Google LLC, 2024b), consultada mediante la librería google-api-python-client.

Modelo de clasificación de sentimiento: robertuito-sentiment-analysis (Pérez et al., 2022), modelo preentrenado en español disponible en el repositorio Hugging Face, aplicado mediante la librería transformers. El modelo clasifica cada comentario en una de tres categorías (positivo, negativo o neutral) y asigna un porcentaje de confianza a la clasificación.

Dato no recolectado, por decisión de diseño: nombre de usuario o cualquier identificador del autor del comentario, en aplicación del principio de minimización de datos descrito en el Capítulo 2 de este documento.

Nota. Protocolo de elaboración propia, implementado como parte del proceso de extracción, transformación y carga (ETL) descrito en el apartado 4.3.1 de este documento.

APÉNDICE C. POLÍTICA DE TRATAMIENTO DE DATOS PERSONALES

Plataforma de gestión de datos para el monitoreo de cultura ciudadana en el área metropolitana de Bucaramanga

1. Responsable del tratamiento

Jair Francisco Alvarado Bautista, estudiante del programa de Tecnología en Desarrollo de Sistemas Informáticos de las Unidades Tecnológicas de Santander (UTS), identificado con cédula de ciudadanía n.º 1005257674, en calidad de autor y responsable del tratamiento de los datos recolectados a través de la plataforma desarrollada en el marco del presente trabajo de grado. Correo de contacto: jfranciscoalvarado@uts.edu.co
Director del trabajo de grado: Genner Andrés Carrillo Rueda, docente de las Unidades Tecnológicas de Santander.

2. Datos objeto de tratamiento

La plataforma recolecta los siguientes datos, exclusivamente a través del formulario público de encuesta de percepción ciudadana:

- Comuna o corregimiento de residencia.
- Rango de edad.
- Sexo.
- Respuestas de percepción sobre seguridad y uso del espacio público (escalas numéricas de 1 a 5).

- Indicación de participación en actividades comunitarias (sí/no).
- Comentario adicional de texto libre (opcional).

No se recolecta nombre, número de identificación, dirección exacta, número de teléfono, correo electrónico ni ningún otro dato que permita identificar directamente a la persona que responde. En consecuencia, los datos tratados no constituyen, en sentido estricto, datos personales identificables bajo los términos generales de la Ley 1581 de 2012, si bien se atienden en esta política los principios de dicha norma como buena práctica de gobernanza de datos.

Adicionalmente, la plataforma procesa comentarios públicos publicados por terceros en la plataforma YouTube, obtenidos a través de su interfaz de programación de aplicaciones (API) oficial. De estos comentarios se conserva únicamente el contenido textual público y el resultado de su análisis de sentimiento; no se almacena el nombre de usuario, identificador de cuenta ni ninguna otra información que permita identificar al autor del comentario.

3. Finalidad del tratamiento

Los datos recolectados se utilizan exclusivamente para: a) El desarrollo del trabajo de grado y su correspondiente sustentación académica ante las Unidades Tecnológicas de Santander. b) La generación de indicadores agregados y estadísticos de percepción ciudadana, presentados siempre de forma consolidada y nunca de manera individualizada. c) La visualización pública, en el panel de la plataforma, de dichos indicadores agregados.

No se utilizarán los datos recolectados con fines comerciales, publicitarios, ni se transferirán ni venderán a terceros.

4. Principio de minimización de datos

En cumplimiento del principio de minimización, la plataforma se diseñó para recolectar únicamente los datos estrictamente necesarios para los fines descritos, evitando la captura de información adicional que no sea indispensable para el análisis de percepción ciudadana.

5. Almacenamiento y seguridad

Los datos se almacenan en una base de datos gestionada por el proveedor de infraestructura en la nube Neon (PostgreSQL como servicio), con acceso restringido mediante credenciales que solo conoce el responsable del tratamiento. La conexión entre la plataforma y la base de datos se realiza mediante protocolos cifrados (SSL/TLS).

6. Derechos de los titulares

Cualquier persona que haya participado en la encuesta, aunque los datos recolectados sean anónimos, puede ejercer los siguientes derechos, en la medida en que sean aplicables dada la naturaleza anónima de la información: a) Conocer, actualizar y rectificar la información general recolectada por la plataforma. b) Solicitar información sobre el uso que se ha dado a los datos. c) Presentar quejas ante la autoridad competente (Superintendencia de Industria y Comercio) por infracciones a la normativa de protección de datos. d) Revocar la autorización y/o solicitar la eliminación de los datos,

cuando dicha solicitud sea técnicamente posible dada la naturaleza anónima de la información recolectada.

Estos derechos pueden ejercerse escribiendo al correo de contacto señalado en el numeral 1 de este documento.

7. Vigencia y limitación reconocida

Esta política entra en vigencia a partir de su publicación en este documento. Se reconoce como limitación que, al momento de la recolección de las primeras respuestas de la encuesta, esta política formal aún no había sido redactada ni publicada; dicha recolección se realizó bajo un aviso general de anonimato y uso académico visible en el formulario, pero sin el nivel de detalle que aquí se presenta. Se recomienda que, en fases posteriores de este proyecto, la política de tratamiento de datos esté disponible y visible desde el primer momento de recolección de cualquier dato nuevo.

Fecha de elaboración: septiembre de 2026.

APÉNDICE D. ACTAS DE CONSULTA A ACTORES DE INTERÉS

Acta 1

Persona consultada: Sara Sofía Gómez Cáceres

Rol o vínculo con la comunidad: Ciudadana

Comuna: Comuna 11

Fecha: 10 de septiembre de 2026

Duración: 10 minutos

Medio: Presencial

P1. Desde su experiencia en su comuna o barrio, ¿qué comportamientos o situaciones asocia usted con una "buena" o "mala" cultura ciudadana?

Buena cultura ciudadana es cuando los vecinos botan la basura el día que corresponde; mala cultura ciudadana es cuando sacan a sus mascotas y no recogen sus heces.

P2. ¿Cuáles temas considera más relevantes para entender la seguridad y la convivencia en su sector?

La percepción de seguridad de la gente.

P3. ¿Cree que una herramienta que integre datos de delitos, encuestas de percepción y redes sociales sería útil para su comunidad o las autoridades?

Sí, ya que ayuda a fortalecer la seguridad en el sector.

P4. ¿Hay algo que esperaba ver en una plataforma así, que no se le haya mencionado?

Ver las estadísticas de la peligrosidad de cada sector.

P5. ¿Tiene algún comentario adicional sobre cultura ciudadana en Bucaramanga?

No, considera que el tema está bien abordado.

Consentimiento: la persona consultada autorizó el uso de esta información de forma anónima y agregada para este trabajo de grado.

Acta 2

Persona consultada: Rafael González

Rol o vínculo con la comunidad: Presidente de la Junta de Acción Comunal (JAC)

Comuna: Comuna 4

Fecha: 10 de septiembre de 2026

Duración: 10 minutos

Medio: Presencial

P1. Desde su experiencia en su comuna o barrio, ¿qué comportamientos o situaciones asocia usted con una "buena" o "mala" cultura ciudadana?

Buena cultura ciudadana es que la comunidad sea solidaria y vele por el bienestar de todos; mala cultura ciudadana son los vecinos que ponen música en las noches afectando la tranquilidad de los demás.

P2. ¿Cuáles temas considera más relevantes para entender la seguridad y la convivencia en su sector?

La frecuencia de delitos y el uso adecuado del espacio público.

P3. ¿Cree que una herramienta que integre datos de delitos, encuestas de percepción y redes sociales sería útil para su comunidad o las autoridades?

Sí, sería muy útil para llevar un control y seguimiento de los delitos en el barrio.

P4. ¿Hay algo que esperaba ver en una plataforma así, que no se le haya mencionado?

Saber en qué horario se manifiestan más los robos.

P5. ¿Tiene algún comentario adicional sobre cultura ciudadana en Bucaramanga?

Que la ciudadanía no tenga acceso a editar los datos de la plataforma, ya que eso conserva que sean verídicos. *Consentimiento: la persona consultada autorizó el uso de esta información de forma anónima y agregada para este trabajo de grado.*

Acta 3

Persona consultada: José Alberto Guerrero

Rol o vínculo con la comunidad: Ciudadano

Comuna: Comuna 3

Fecha: 10 de septiembre de 2026

Duración: 10 minutos

Medio: Presencial

P1. Desde su experiencia en su comuna o barrio, ¿qué comportamientos o situaciones asocia usted con una "buena" o "mala" cultura ciudadana?

Los vecinos que cuidan las zonas públicas hacen parte de una buena cultura ciudadana; los que fuman en espacios públicos y son groseros hacen parte de una mala cultura ciudadana.

P2. ¿Cuáles temas considera más relevantes para entender la seguridad y la convivencia en su sector?

El uso adecuado del espacio público y la percepción de seguridad de la gente.

P3. ¿Cree que una herramienta que integre datos de delitos, encuestas de percepción y redes sociales sería útil para su comunidad o las autoridades?

Sí, para que las autoridades conozcan los focos de delitos.

P4. ¿Hay algo que esperaríamos ver en una plataforma así, que no se le haya mencionado?

No, considera que la plataforma está muy completa.

P5. ¿Tiene algún comentario adicional sobre cultura ciudadana en

Bucaramanga?

No tiene comentarios adicionales. *Consentimiento: la persona consultada autorizó el uso de esta información de forma anónima y agregada para este trabajo de grado.*

APÉNDICE E. SOLICITUD DE AUTORIZACIÓN Y RECONOCIMIENTO INSTITUCIONAL

A continuación, se presenta la comunicación formal enviada a la Secretaría del Interior de la Alcaldía de Bucaramanga, solicitando reconocimiento institucional del proyecto y la eventual expedición de una constancia de conformidad. La solicitud fue radicada exitosamente ante la Alcaldía de Bucaramanga el 10 de septiembre de 2026, bajo el número de radicado 1-WEB-202609-00177115, mediante el cual es posible consultar el estado de la solicitud. A la fecha de elaboración de este documento, se encuentra pendiente de respuesta por parte de la entidad.

Bucaramanga, 10 de septiembre de 2026

Señores

SECRETARÍA DEL INTERIOR

Alcaldía de Bucaramanga

Ciudad

Asunto: Solicitud de reconocimiento institucional Trabajo de grado "Plataforma de gestión de datos para el monitoreo de cultura ciudadana en el área metropolitana de Bucaramanga"

Respetados señores:

Mi nombre es Jair Francisco Alvarado Bautista, identificado con cédula de ciudadanía n.º 1005257674, estudiante del programa de Tecnología en Desarrollo de Sistemas Informáticos de las Unidades Tecnológicas de Santander (UTS). Me dirijo a ustedes en el marco de mi trabajo de grado, desarrollado bajo la dirección del docente Genner Andrés Carrillo Rueda.

El proyecto consiste en una plataforma web que integra tres fuentes de datos el conjunto de datos delictivos de Bucaramanga publicado por esa Secretaría a través del portal de Datos Abiertos de Colombia (datos.gov.co), una encuesta de percepción ciudadana, y un análisis de comentarios públicos en redes sociales con el propósito de apoyar el monitoreo de indicadores de cultura ciudadana en el municipio. La plataforma se encuentra desplegada públicamente y disponible para su consulta en: <https://cultura-ciudadana-bucaramanga.onrender.com>

Dado que el proyecto emplea de manera central el conjunto de datos publicado por esa Secretaría, y que uno de los objetivos del trabajo es que la herramienta desarrollada resulte de utilidad para la gestión pública municipal, solicito respetuosamente:

1. Un espacio breve para presentar la plataforma ante el área que la Secretaría considere pertinente.

2. La expedición de una constancia o acta de reconocimiento en la que se certifique que la Secretaría tuvo conocimiento del proyecto y, de considerarlo pertinente, su valoración sobre la utilidad de la herramienta presentada.

Quedo atento a cualquier información adicional que se requiera para dar trámite a esta solicitud, y agradezco de antemano la atención prestada.

Cordialmente,

Jair Francisco Alvarado Bautista

Estudiante — Tecnología en Desarrollo de Sistemas Informáticos

Unidades Tecnológicas de Santander

Correo: jfranciscoalvarado@uts.edu.co

Teléfono: 3136291492

Con copia a:

Genner Andrés Carrillo Rueda — Director del trabajo de grado

Unidades Tecnológicas de Santander

10. ANEXOS

Plataforma de gestión de datos para el monitoreo de cultura ciudadana en el área metropolitana de Bucaramanga

Este manual describe el funcionamiento técnico y el modo de uso de la plataforma desarrollada en este trabajo de grado. Se divide en dos partes: la primera dirigida a quien necesite instalar, mantener o extender el sistema (manual técnico); la segunda dirigida a cualquier persona que use la plataforma para consultar información sobre cultura ciudadana en el municipio (manual de usuario). La plataforma es de acceso público en <https://cultura-ciudadana-bucaramanga.onrender.com/dashboard/>

Arquitectura general

La plataforma sigue una arquitectura de tres capas: capa de datos (PostgreSQL, alojada en Neon), capa de backend (API REST en FastAPI/Python, alojada en Render) y capa de presentación (dashboard web en HTML, CSS y JavaScript, servido como archivos estáticos desde el mismo backend). Un componente adicional, el agente conversacional, utiliza el modelo Gemini mediante llamado a funciones para consultar los propios endpoints de la plataforma y responder preguntas en lenguaje natural con datos verificables.

Tabla 4. *Tecnologías empleadas*

Categoría	Tecnología	Función en el proyecto
Backend	Python 3.13 + FastAPI	Framework para la construcción de la API REST
Base de datos	PostgreSQL (Neon)	Almacenamiento relacional en la nube, nivel gratuito permanente
ORM	SQLAlchemy 2.0	Mapeo objeto-relacional entre Python y PostgreSQL
Hosting backend	Render	Despliegue del servicio web, nivel gratuito
Frontend	HTML5, CSS3, JavaScript	Dashboard interactivo, sin frameworks adicionales
Visualización	Chart.js 4.4	Renderizado de gráficas interactivas
PLN / Sentimiento	transformers + robertuito-sentiment-analysis	Clasificación de sentimiento en comentarios
Datos de redes sociales	YouTube Data API v3	Extracción de comentarios públicos
Agente conversacional	Google Gemini API	Generación de respuestas e informes en lenguaje natural
Control de versiones	Git + GitHub	Historial de cambios y despliegue continuo

Modelo de datos

La base de datos se diseñó bajo un modelo dimensional (esquema estrella), compuesto por una tabla de hechos (hecho_delictivo) y ocho tablas de dimensión: tipología, delito, clase de sitio, arma o medio, móvil, comuna, barrio y curso de vida. Dos tablas adicionales, encuesta_respuesta y red_social_comentario, almacenan las otras dos fuentes de datos de la plataforma.

Instalación en entorno local

Requisitos previos: Python 3.13, PostgreSQL 17, Git, y llaves de acceso a la API de YouTube y a la API de Gemini.

- Clonar el repositorio del proyecto desde GitHub.
- Crear y activar un entorno virtual de Python.
- Instalar las dependencias listadas en requirements.txt.
- Crear un archivo .env con las variables DATABASE_URL, YOUTUBE_API_KEY y GEMINI_API_KEY.
- Levantar el servidor con el comando: `python -m uvicorn main:app --reload`
- Acceder al panel en la dirección `http://127.0.0.1:8000/dashboard/`

Despliegue en producción

El sistema está desplegado con servicios separados para garantizar la persistencia de los datos: la base de datos se aloja en Neon (PostgreSQL gestionado, nivel gratuito permanente), y el backend se despliega en Render, configurado para actualizarse automáticamente ante cada cambio subido al repositorio de GitHub. Limitación conocida: el nivel gratuito de Render pone el servicio en reposo tras 15 minutos de inactividad; la primera solicitud posterior puede tardar entre 30 y 60 segundos en responder.

Tabla 5. Referencia de la API — Módulo de Delitos

Método	Endpoint	Descripción
GET	/delitos/resumen-por-comuna	Total, de casos agrupados por comuna, con filtro opcional por año
GET	/delitos/resumen-por-tipologia	Total de casos agrupados por tipología de delito
GET	/delitos/resumen-temporal	Total de casos por año y mes
GET	/delitos/resumen-por-barrio	Total de casos por barrio, con protección de datos sensibles
GET	/delitos/resumen-por-rango-horario	Total de casos por franja horaria del día
GET	/delitos/rango-horario-por-tipologia	Cruce de rango horario y tipología de delito
GET	/delitos/tendencia-mensual-por-tipologia	Serie temporal mensual desglosada por tipología
GET	/delitos/resumen-por-sexo	Total de casos por sexo de la víctima
GET	/delitos/resumen-por-movil-victima	Total de casos por medio de desplazamiento de la víctima
GET	/delitos/piramide-edad-sexo	Total de casos por rango de edad y sexo de la víctima

Tabla 6. Referencia de la API — Módulos de Encuestas, Redes Sociales y Chat

Método	Endpoint	Descripción
GET	/encuestas/opciones-comuna	Lista de comunas para el formulario
GET	/encuestas/opciones-rango-edad	Lista de rangos de edad para el formulario
POST	/encuestas/	Registra una nueva respuesta de encuesta
GET	/encuestas/resumen-por-comuna	Promedios de percepción ciudadana por comuna
GET	/redes-sociales/resumen-sentimiento	Total de comentarios por categoría de sentimiento
GET	/redes-sociales/comentarios-recientes	Muestra de comentarios analizados recientemente
POST	/chat/	Envía un mensaje al asistente y recibe una respuesta basada en datos reales
GET	/chat/informe	Genera un informe ejecutivo completo con inteligencia artificial

La documentación interactiva completa de todos los endpoints, generada automáticamente por FastAPI, está disponible en <https://cultura-ciudadana-bucaramanga.onrender.com/docs>

Consideraciones de seguridad y protección de datos

En cumplimiento de la Ley 1581 de 2012, la plataforma no almacena el nombre de usuario de quienes comentan en YouTube, ni ningún dato que permita identificar directamente a una persona en las respuestas de encuesta. Adicionalmente, cuando se consulta una categoría de delito sensible (relacionada con menores de edad o delitos sexuales) desagregada por barrio, los conteos inferiores a un umbral definido se ocultan automáticamente para reducir el riesgo de identificación indirecta de víctimas.

Acceso a la plataforma

La plataforma es de acceso público y no requiere registro. Se ingresa desde cualquier navegador a la dirección <https://cultura-ciudadana-bucaramanga.onrender.com/dashboard/>

El panel principal (dashboard)

El panel muestra cuatro indicadores generales (total de casos, respuestas de encuesta, comentarios analizados y comunas monitoreadas), un filtro por año, y las gráficas organizadas en las secciones: Delitos registrados; Cuándo y cómo ocurren los delitos; Percepción ciudadana y redes sociales; y Comentarios recientes analizados. Todas las gráficas son interactivas.

Cómo responder la encuesta de percepción ciudadana

- Ingresar a la dirección .../dashboard/encuesta.html
- Seleccionar la comuna de residencia y el rango de edad.
- Seleccionar el sexo.
- Calificar, de 1 a 5, la percepción de seguridad de día y de noche.
- Calificar, de 1 a 5, la frecuencia percibida de mal uso del espacio público.
- Indicar si ha participado en actividades comunitarias en el último año.
- Escribir, opcionalmente, un comentario adicional.
- Presionar "Enviar respuesta".

Cómo usar el asistente de la plataforma

En el panel del asistente, identificado con un ícono circular, se puede escribir cualquier pregunta sobre delitos, percepción ciudadana o redes sociales en Bucaramanga. El asistente consulta los datos reales de la plataforma antes de responder, en vez de generar una respuesta basada únicamente en su conocimiento previo.

Cómo descargar el informe ejecutivo con inteligencia artificial

Al presionar el botón "Descargar informe (PDF)" en la parte superior del panel, el sistema genera, mediante inteligencia artificial, un informe ejecutivo con seis secciones (resumen general, tipología predominante, percepción ciudadana, análisis de sentimiento, hallazgos clave y recomendaciones), y lo descarga automáticamente en formato PDF. El proceso tarda entre 20 y 40 segundos.

F-DC-125

**INFORME FINAL DE TRABAJO DE GRADO EN MODALIDAD DE PROYECTO
DE INVESTIGACIÓN, DESARROLLO TECNOLÓGICO, MONOGRAFÍA,
EMPRENDIMIENTO Y SEMINARIO**

VERSIÓN: 2.0

ELABORADO POR:
Docencia

REVISADO POR:
Sistema Integrado de Gestión

APROBADO POR: Líder del Sistema Integrado de Gestión
FECHA APROBACIÓN: Octubre de 2023