



Análisis de sentimientos en reseñas hoteleras de Buenaventura mediante minería de texto

Andrés Felipe Duque Caicedo

Trabajo de grado presentado para optar al título de Ingeniero en Analítica de Datos

Director: Andrés Alberto Osorio Londoño, Doctor (PhD) en Ingeniería

Codirector: Edgar Rafael Jiménez López, Magíster (MSc) en Gestión y Desarrollo De Proyectos
De Software

Universidad de Manizales
Facultad de Ciencias e Ingeniería
Ingeniería en Analítica de Datos
Manizales, Caldas, Colombia
2026

Cita	(Duque Caicedo, 2026)
Referencia	Duque Caicedo, A. F. (2026). <i>Análisis de sentimientos en reseñas hoteleras de Buenaventura mediante minería de texto</i> [Trabajo de grado profesional]. Universidad de Manizales. RIDUM: Repositorio Institucional Universidad de Manizales.
Estilo APA 7 (2020)	



Grupo de Investigación y Desarrollo en Informática y Telecomunicaciones

Declaración de inteligencia artificial: el o los autores de este trabajo de grado declaran que han utilizado herramientas de inteligencia artificial (IA), tales como [ChatGPT, Perplexity], de manera ética y responsable, tal como se establece en el Acuerdo UManizales 002 (julio 26 de 2023) sobre propiedad intelectual e IA. Estas herramientas son empleadas como apoyo en la redacción, revisión gramatical y generación de ideas, pero en ningún caso sustituyen el análisis crítico, la argumentación académica ni la originalidad del trabajo. Asimismo, cualquier contenido generado con asistencia de IA está citado y referenciado adecuadamente, garantizando la integridad académica y el cumplimiento de los principios éticos de la investigación.

Biblioteca y Centro de Recursos: biblioteca.umanizales.edu.co

Repositorio Institucional: ridum.umanizales.edu.co

Universidad de Manizales: umanizales.edu.co

Revistas: revistasum.umanizales.edu.co

Fondo Editorial: editorialum.umanizales.edu.co

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Manizales ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

Dedicatoria

A Dios

Agradecimientos

Familia, Amigos y Compañeros

Tabla de contenido

Resumen.....	10
Abstract.....	11
Introducción	12
1 Planteamiento del Problema de Investigación y Justificación.....	14
1.1 Descripción del Área Problemática	14
1.2 Formulación del Problema.....	18
1.3 Justificación	18
2 Objetivos.....	20
2.1 Objetivo General.....	20
2.2 Objetivos Específicos.....	20
3 Antecedentes	21
4 Referente Normativo y Legal.....	24
5 Referente Teórico.....	26
5.1 Fundamentos Técnicos del Análisis de Sentimientos	26
5.2 Fundamentos conceptuales de la percepción turística	32
6 Metodología	34
6.1 Enfoque Metodológico.....	35
6.2 Tipo de Estudio	36
6.3 Procedimiento	36
7 Resultados	49
7.1 Caracterización General del Corpus y del Modelo de Análisis de Sentimientos	49
7.2 Indicadores Clave de Reputación Digital por Hotel	52
7.3 Evolución Anual de los Sentimientos por Hotel.....	57
7.4 Resultados Temáticos de las Reseñas Negativas	62

7.5 Resultados Temáticos de las Reseñas Positivas.....	69
7.6 Síntesis de Resultados por Objetivo Específico.....	77
8 Conclusiones	79
8.1 Conclusión General del Estudio.....	79
8.2 Conclusiones Frente a los Objetivos Específicos	81
8.3 Aportes y Limitaciones del Estudio	83
9 Recomendaciones	85
9.1 Recomendaciones Generales para los Hoteles Analizados.....	85
9.2 Recomendaciones Específicas por Hotel	87
9.3 Recomendaciones para Actores del Destino y Futuras Investigaciones	90
Referencias.....	92
Anexo A. Perspectiva de referencia y crecimiento potencial atribuible a una mayor presencia en línea para Colombia	96
Anexo B. Oportunidad para un mayor gasto en turismo en Colombia (2021–2025)	97
Anexo C. Evidencia del repositorio del proyecto en GitHub	98
Anexo D. Evidencia de la Página Web Desarrollada	99

Lista de tablas

Tabla 1. Columnas incluidas en el análisis y justificación técnica	42
Tabla 2. Columnas excluidas del análisis y justificación	44
Tabla 3. Reseñas finales por hotel después del ETL	45
Tabla 4. Campos registrados en Supabase para control de versiones del ETL.....	46
Tabla 5. Distribución de reseñas textuales por hotel y periodo de análisis (2020–2025).....	50
Tabla 6. Resultados de la verificación manual externa (3%).....	51
Tabla 7. Ejemplos ilustrativos de la verificación manual externa	51
Tabla 8. Temáticas negativas — Hotel Cosmos Pacífico	62
Tabla 9. Temáticas negativas — Hotel Cordillera.....	64
Tabla 10. Temáticas negativas — Steven Buenaventura.....	66
Tabla 11. Temáticas negativas — Hotel Torre Mar	67
Tabla 12. Temáticas negativas — Hotel Magüipí	68
Tabla 13. Temáticas positivas — Hotel Cosmos Pacífico.....	70
Tabla 14. Temáticas positivas — Hotel Cordillera.....	71
Tabla 15. Temáticas positivas — Steven Buenaventura.....	73
Tabla 16. Temáticas positivas — Hotel Torre Mar	74
Tabla 17. Temáticas positivas — Hotel Magüipí	75

Lista de figuras

Figura 1. Fuentes de información utilizadas por los visitantes recibidos en el Valle del Cauca en el pre-viaje. Periodo 2017–2018	15
Figura 2. Políticas clave para fomentar el desarrollo sostenible en Buenaventura en Buenaventura	16
Figura 3. Problemáticas percibidas en Buenaventura por empresarios afiliados a la CCBUN (2024)	17
Figura 4. Esquema del análisis de sentimientos mediante Azure Cognitive Services – Text Analytics	27
Figura 5. Configuración inicial del scraper con las URL de los hoteles seleccionado	39
Figura 6. Parámetro de límite de reseñas por establecimiento en Apify	39
Figura 7. Configuración del criterio de ordenamiento de reseñas (más recientes primero)	40
Figura 8. Configuración predeterminada del parámetro temporal en Apify	40
Figura 9. Configuración por defecto sin filtro de idioma	40
Figura 10. Selección del origen de las reseñas desde Google	41
Figura 11. Ejecución del scraper y resultados generales de la extracción	41
Figura 12. Archivos originales descargados desde Apify almacenados en la capa Bronze	42
Figura 13. Flujo general del proceso ETL implementado (Bronze → Silver → Gold)	47
Figura 14. Panel de indicadores de reputación digital – Hotel Cordillera	53
Figura 15. Panel de indicadores de reputación digital – Cosmos Pacífico	54
Figura 16. Panel de indicadores de reputación digital – Hotel Magüipí	54
Figura 17. Panel de indicadores de reputación digital – Hotel Steven Buenaventura	55
Figura 18. Panel de indicadores de reputación digital – Hotel Torre Mar	56
Figura 19. Evolución anual de los sentimientos en las reseñas del Hotel Cosmos Pacífico	57
Figura 20. Evolución anual de los sentimientos en las reseñas del Hotel Cordillera	58
Figura 21. Evolución anual de los sentimientos en las reseñas del Hotel Steven Buenaventura	59
Figura 22. Evolución anual de los sentimientos en las reseñas del Hotel Torre mar	60

Figura 23. Evolución anual de los sentimientos en las reseñas del Hotel Magüipi	61
---	----

Siglas, acrónimos y abreviaturas

F1-Score	Métrica que combina precisión y recall mediante su media armónica
NLP	Procesamiento de Lenguaje Natural (Natural Language Processing)
UGC	Contenido generado por el usuario (User-Generated Content)
CRISP-DM	Proceso estándar de la industria para minería de datos
SCRUM	Metodología ágil basada en sprints
API	Interfaz de programación de aplicaciones (Application Programming Interface)
KPI	Key performance indicator / Indicador clave de desempeño
LLM	Modelo de lenguaje de gran escala (Large Language Model)
BERT	Representaciones codificadoras bidireccionales de transformadores (Bidirectional Encoder Representations from Transformers)
ABSA	Análisis de sentimientos basado en aspectos (Aspect-Based Sentiment Analysis)
Precisión / Precision	Precisión; proporción de predicciones positivas correctas sobre el total de casos clasificados como positivos
Azure	Plataforma en la nube utilizada para almacenamiento, procesamiento de datos y consumo de servicios cognitivos de IA.
Recall	Exhaustividad o sensibilidad; proporción de casos positivos reales que el modelo logra identificar correctamente
Azure Cognitive Services – Text Analytics	Servicio de procesamiento de lenguaje natural utilizado para clasificar sentimientos, extraer frases clave y generar información accionable.
VADER	Modelo basado en reglas para análisis de sentimientos (Valence Aware Dictionary and SEntiment Reasoner)

Resumen

El análisis de sentimientos aplicado al turismo se ha consolidado como una herramienta fundamental para comprender las percepciones de los visitantes y orientar la toma de decisiones en destinos emergentes. En Buenaventura, ciudad con alto potencial turístico pero marcada por desafíos estructurales como la seguridad y la informalidad empresarial, no existían estudios que sistematizaran de manera rigurosa las opiniones disponibles en plataformas digitales.

Este proyecto examinó 3.727 reseñas publicadas en Google entre 2020 y 2025 sobre los cinco hoteles más comentados de la ciudad, mediante técnicas de minería de texto, análisis automatizado de sentimientos y visualización de datos. El proceso siguió el modelo CRISP-DM, articulando un scraping automatizado con Apify, un flujo de procesamiento en Azure basado en la arquitectura Bronze – Silver – Gold y un enriquecimiento semántico generado a través de Azure Cognitive Services – Text Analytics.

El estudio clasificó la polaridad emocional de las reseñas, identificó aspectos temáticos y reconoció tendencias narrativas asociadas a la experiencia turística. La validación manual de una muestra del 3% arrojó una coherencia interpretativa del 90,2%. Los resultados fueron integrados en dashboards interactivos en Power BI, conformando un sistema visual capaz de revelar patrones de satisfacción, alertas temáticas y brechas en la respuesta hotelera, con valor estratégico para turistas, empresarios y entidades locales.

Palabras clave: análisis de sentimientos, minería de texto, percepción turística, reseñas en línea, reputación digital, hoteles, Buenaventura

Abstract

Sentiment analysis applied to tourism has become a fundamental tool for understanding visitors' perceptions and guiding decision-making in emerging destinations. In Buenaventura, a city with significant tourism potential but affected by structural challenges such as security conditions and business informality, no studies had rigorously systematized the opinions available on digital platforms.

This project examined 3,727 reviews posted on Google between 2020 and 2025 concerning the five most frequently reviewed hotels in the city, using text mining techniques, automated sentiment analysis, and data visualization. The process followed the CRISP-DM model, integrating automated scraping through Apify, a processing workflow in Azure based on the Bronze–Silver–Gold architecture, and semantic enrichment generated through Azure Cognitive Services – Text Analytics.

The study enabled the classification of emotions, the identification of thematic aspects, and the detection of narrative trends associated with the tourism experience. Manual validation of a 3% sample yielded an interpretative coherence of 90.2%. The findings were incorporated into interactive dashboards in Power BI, creating a visual system capable of revealing satisfaction patterns, thematic alerts, and gaps in hotel responses, offering strategic value for tourists, business owners, and local institutions.

Keywords: sentiment analysis, text mining, tourist perception, online reviews, digital reputation, hotels; Buenaventura

Introducción

El crecimiento de las plataformas digitales ha transformado la forma en que los viajeros evalúan, comparan y seleccionan servicios turísticos. En el sector hotelero, las reseñas en línea constituyen una fuente relevante para aproximarse a la experiencia de los huéspedes, dado que reúnen opiniones espontáneas sobre los servicios recibidos. Este tipo de contenido generado por usuarios influye en la imagen de los destinos y en los procesos de decisión turística, como lo señalan Gretzel y Yoo (2008), al destacar el papel de las opiniones en línea en la planificación de viajes.

En este escenario, el análisis de sentimientos ha adquirido relevancia dentro de la investigación turística, porque facilita la interpretación sistemática de grandes volúmenes de opiniones digitales. Manosso y Domareski (2021) evidencian que esta técnica se ha consolidado como una línea de estudio útil para comprender percepciones, valoraciones y experiencias de los visitantes. De manera complementaria, la minería de texto permite extraer conocimiento a partir de datos no estructurados, como comentarios, reseñas y documentos escritos (Feldman & Sanger, 2007).

A partir de este marco, el presente trabajo de grado analiza reseñas hoteleras publicadas en Google sobre establecimientos ubicados en Buenaventura, con el propósito de identificar patrones de percepción turística mediante técnicas de minería de texto. Para ello, se examina la distribución de sentimientos positivos, neutrales y negativos, así como los aspectos de la experiencia hotelera asociados a la satisfacción o insatisfacción de los usuarios.

Para el desarrollo del proyecto se adoptó el modelo CRISP-DM, propuesto por Chapman et al. (2000), debido a su utilidad para estructurar proyectos de minería de datos en fases ordenadas. El proceso integró extracción automatizada de reseñas mediante Apify, organización de datos bajo una arquitectura Bronze – Silver – Gold, enriquecimiento semántico con Azure Cognitive Services – Text Analytics y visualización de resultados en Power BI. Además, se realizó una validación manual externa sobre una muestra del corpus para contrastar la coherencia interpretativa de la clasificación automática.

El documento se estructura en varios capítulos. En primer lugar, se presenta el planteamiento del problema, su formulación y la justificación del estudio. Luego se exponen los objetivos, antecedentes, referentes normativos y fundamentos teóricos. Posteriormente, se describe

la metodología empleada, se presentan los resultados obtenidos, se desarrollan las conclusiones y se proponen recomendaciones orientadas a los hoteles analizados, a los actores del destino y a futuras investigaciones.

1 Planteamiento del Problema de Investigación y Justificación

Este capítulo presenta el problema central que dio origen al desarrollo del proyecto de grado, así como los argumentos que justifican su relevancia académica y social. Se expone la descripción del área problemática, la formulación del problema y las razones que sustentaron la pertinencia de la investigación. El propósito fue explicar con claridad qué se estudió, por qué fue necesario hacerlo y cuál fue su valor para Buenaventura y su sector turístico.

1.1 Descripción del Área Problemática

Buenaventura, municipio portuario del Valle del Cauca, se proyecta como un destino turístico estratégico en Colombia. Solo entre enero y febrero de 2025 recibió más de 31.000 visitantes, consolidándose como un motor importante de la economía regional (Cámara de Comercio de Buenaventura, 2025). Sus atractivos naturales y culturales incluyen el avistamiento de ballenas jorobadas entre julio y octubre, especialmente en la zona del Parque Nacional Natural Uramba Bahía Málaga, así como playas reconocidas como Juanchaco, Ladrilleros y La Barra, y celebraciones tradicionales como los carnavales de San Buenaventura en junio y julio (Colombia Travel, s.f.; Comfenalco Valle, s.f.).

En el ámbito académico, las revisiones sistemáticas sobre el uso del análisis de sentimientos en turismo confirmaron un creciente interés internacional en esta metodología. Destacaron China (35 publicaciones) y Estados Unidos (24) como los países con mayor producción científica en el área. Estos estudios evidenciaron el valor de las opiniones de los usuarios en línea para tomar decisiones estratégicas en el sector turístico (Manosso & Domareski, 2021).

No obstante, este potencial enfrentó obstáculos significativos. La violencia y el estigma asociados al conflicto urbano y a la cobertura mediática contribuyeron a consolidar una imagen negativa de la ciudad, afectando la confianza de los visitantes y el desarrollo turístico (El País, 2024; CWMas, 2025). A ello se sumó la dificultad para acceder a información clara y organizada: aunque miles de turistas comparten sus experiencias en plataformas como Google Reviews, estas reseñas permanecen dispersas y poco estructuradas. Este fenómeno fue identificado por la Organización Mundial del Turismo (2021), al advertir que la sobrecarga de información digital sin clasificación temática representa una barrera para la toma de decisiones informadas. En el caso de Buenaventura, el Hotel Cordillera acumuló más de 2.900 opiniones, mientras que otros como el Cosmos Pacífico, Maguipi, Torre Mar y Steven Buenaventura superaron fácilmente los 600

comentarios cada uno. Esta saturación informativa hizo inviable que un visitante promedio leyera todos los contenidos y formara un criterio confiable, lo cual reforzó la necesidad de contar con herramientas automatizadas que sintetizaran y clasificaran dicha información de forma comprensible.

En línea con esta problemática, se revisó la Encuesta Ritmo Empresarial (Cámara de Comercio de Buenaventura, 2025), la cual resaltó el turismo como sector clave para el desarrollo económico del distrito. Dicho documento evidenció la necesidad de mejorar la visibilidad digital de la oferta hotelera y turística, ya que muchos visitantes tomaban decisiones con base en percepciones poco estructuradas. Esta situación se reforzó con los hallazgos del Sistema de Información Turística del Valle del Cauca – SITUR, que mostraron que la mayoría de los visitantes realizó búsquedas de información antes de su viaje, lo que subrayó la importancia de contar con contenidos digitales claros y actualizados. En efecto, el 89,09% de los encuestados manifestó haber consultado fuentes previas a su visita, como se observa en la Figura 1.

Figura 1.

Fuentes de información utilizadas por los visitantes recibidos en el Valle del Cauca en el pre-viaje. Periodo 2017–2018

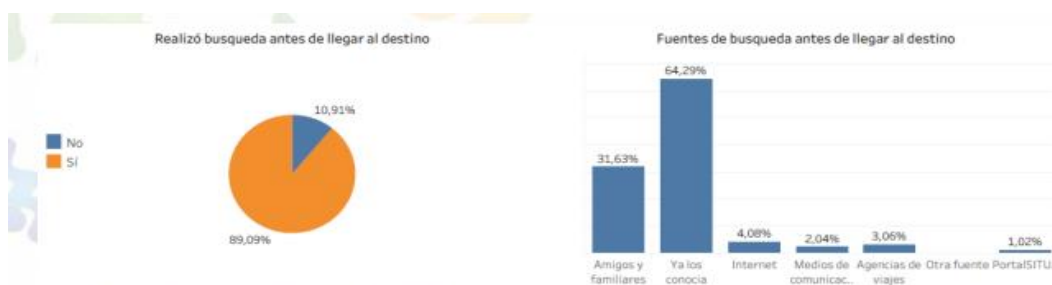
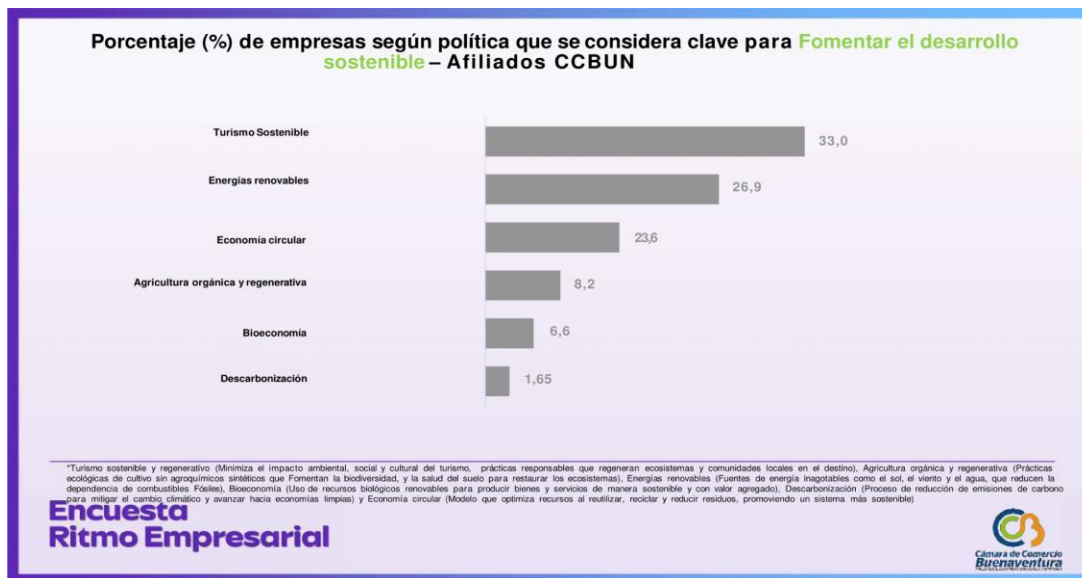


Figura N°A27. Principales fuentes de información utilizada por los visitantes recibidos por el Valle del Cauca en el pre-viaje. Periodo: 2017-2018

Nota. Tomado de <https://tinyurl.com/situr> (SITUR Valle del Cauca & Cotelco, 2018).

Comprender la percepción turística resultó fundamental para fortalecer la confianza de los visitantes y promover el desarrollo económico local. En este sentido, el turismo sostenible se destacó como la política más priorizada para alcanzar dicho desarrollo en Buenaventura, representando el 33% de las prioridades de las empresas locales (Cámara de Comercio de Buenaventura, 2024), como se muestra en la Figura 2.

Figura 2.*Políticas clave para fomentar el desarrollo sostenible en Buenaventura*

Nota. Tomado de Cámara de Comercio de Buenaventura (2024).

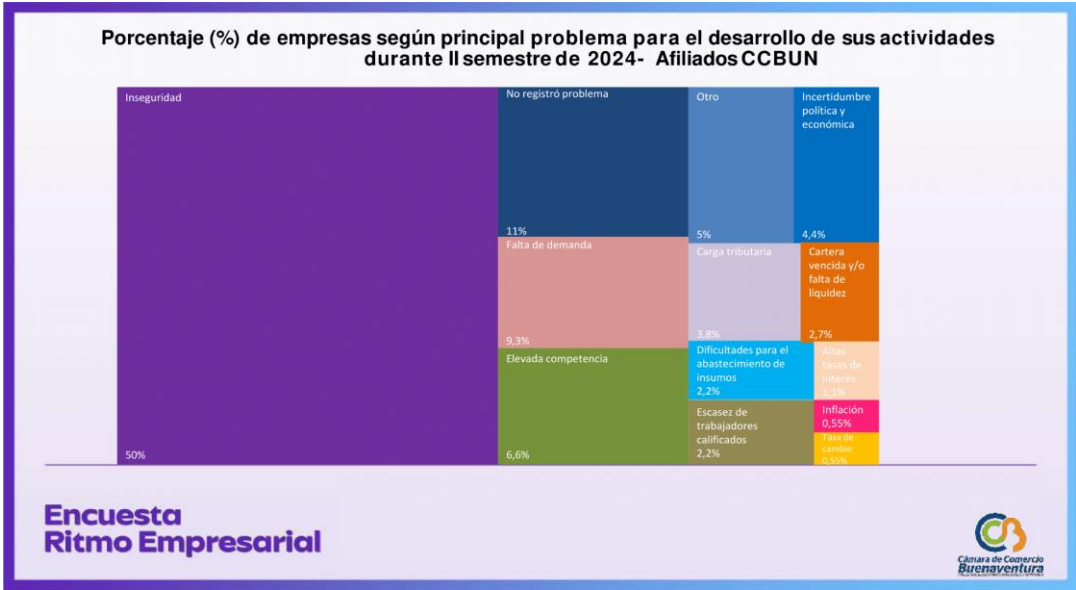
A nivel internacional, se revisaron estudios que aplican análisis de sentimientos en reseñas turísticas para identificar patrones de percepción del servicio. Por ejemplo, el estudio de Oliveira et al. (2022) analizó 8.596 reseñas de hoteles en la región del Algarve (Portugal), demostrando cómo los sentimientos variaban según la procedencia de los visitantes y aportando insumos útiles para segmentar clientes y mejorar la gestión hotelera.

En Colombia, el estudio de Mendoza-Moreno et al. (2021) constituyó uno de los pocos antecedentes nacionales que aplicaron minería de opinión en el sector turístico. Esta investigación desarrolló una ontología para la trazabilidad del turista e incorporó técnicas de análisis de sentimientos (NLP y visualización georreferenciada) para calcular un KPI de satisfacción con base en opiniones de visitantes a diferentes destinos turísticos del país. Este trabajo validó la aplicabilidad de este tipo de metodologías en contextos nacionales y respaldó la capacidad técnica disponible.

Sin embargo, no se encontraron investigaciones que sistematizaran las percepciones de los visitantes en el Valle del Cauca, y menos aún en Buenaventura. Esta brecha informativa limitó a los turistas, al carecer de información estructurada para decidir, y también a empresarios y entidades locales, que no contaban con retroalimentación analítica para mejorar la oferta y posicionar la ciudad como destino turístico.

Por otra parte, la Encuesta Ritmo Empresarial (Cámara de Comercio de Buenaventura, 2024) también reveló que la inseguridad fue percibida como el principal obstáculo para el desarrollo empresarial en la ciudad, seguida por la falta de demanda y la alta competencia, tal como se observa en la Figura 3.

Figura 3.
Problemáticas percibidas en Buenaventura por empresarios afiliados a la CCBUN (2024)



Nota. Tomado de Cámara de Comercio de Buenaventura (2024).

Finalmente, este estudio exploró la relación entre dos dimensiones clave: por un lado, los sentimientos expresados por los usuarios en las reseñas publicadas en Google Reviews; y por otro, la percepción turística general del destino Buenaventura. El análisis de estas opiniones identificó patrones de percepción del servicio hotelero y aportó elementos para comprender su relación con la imagen proyectada de la ciudad. Esta relación adquirió relevancia a la luz de estudios recientes en Colombia, los cuales demostraron que una mayor visibilidad en línea se traduce en mayores ingresos, empleos y desarrollo para el sector turístico. Por ejemplo, un estudio conjunto de Google y Oxford Economics proyectó que la digitalización del turismo podría generar 26.000 nuevos empleos, 600.000 noches de estadía adicionales y hasta US\$800 millones en gasto turístico acumulado para 2025 (AmCham Colombia, 2022). Además, entre 2012 y 2019, el uso creciente de canales digitales contribuyó con un aumento del 7% en la economía del turismo del país (AmCham Colombia, 2021). En este marco, se concluye que proyectos como el presente aportan

a la sistematización de la percepción turística en plataformas digitales, facilitando la toma de decisiones basada en datos confiables.

1.2 Formulación del Problema

¿Cuáles son los patrones de percepción turística de los sentimientos expresados en las reseñas de Google sobre los cinco hoteles más reseñados de Buenaventura entre el 1.º de enero de 2020 y el 21 de agosto de 2025?

1.3 Justificación

Este proyecto surgió del interés por aportar al desarrollo de Buenaventura, ciudad natal del investigador, aprovechando el gran potencial turístico de la región. Representó una oportunidad valiosa para aplicar los conocimientos adquiridos en la carrera de Ingeniería en Analítica de Datos a un problema real, cercano y con impacto social. La motivación también radicó en fomentar la adopción de la transformación digital en el sector turístico local, donde todavía prevalecían prácticas tradicionales como la confianza en recomendaciones orales, frecuentemente subjetivas y difíciles de verificar.

Desde una perspectiva práctica, la utilidad del proyecto se manifestó en su capacidad para generar información organizada, clara y confiable, dirigida a múltiples actores. Estos resultados podrán beneficiar a distintos grupos: por un lado, los turistas contarán con información comprensible para tomar decisiones informadas al momento de elegir un alojamiento; por otro, los empresarios hoteleros accederán a insumos analíticos útiles para identificar fortalezas y debilidades en sus servicios, impulsando la mejora continua. A su vez, instituciones como la Cámara de Comercio de Buenaventura y Cotelco dispondrán de evidencia empírica relevante para orientar sus estrategias de promoción y posicionamiento.

Finalmente, este estudio abordó una temática novedosa dentro del contexto local. Hasta la fecha, no se había identificado una investigación que aplicara análisis de sentimientos a gran escala sobre reseñas hoteleras en Buenaventura mediante técnicas de minería de texto y NLP. El valor agregado del proyecto no radicó en el desarrollo de software especializado, sino en la transformación de datos dispersos en conocimiento útil a través de metodologías accesibles, confiables y replicables. Esta aproximación generó informes y visualizaciones que reflejaron de

forma sistemática la percepción de los visitantes y aportaron a la consolidación de Buenaventura como un destino turístico competitivo.

2 Objetivos

Este capítulo presenta los objetivos que guiaron el desarrollo del proyecto. Fueron formulados a partir de la problemática identificada, asegurando coherencia entre el planteamiento investigativo y la metodología aplicada. El objetivo general resume el propósito central del estudio, mientras que los objetivos específicos descomponen dicho propósito en acciones concretas que orientaron la recolección, procesamiento y análisis de datos.

2.1 Objetivo General

Identificar patrones de percepción turística en Buenaventura mediante el análisis automatizado de sentimientos en reseñas de Google sobre los cinco hoteles más comentados entre el 1.º de enero 2020 y el 21 de agosto de 2025.

2.2 Objetivos Específicos

- Determinar de forma comparativa la distribución de sentimientos (positivos, neutrales y negativos) en las reseñas de los cinco hoteles más comentados de Buenaventura
- Establecer relaciones entre la polaridad emocional de las reseñas y la percepción turística del destino.
- Proponer recomendaciones basadas en datos que orienten la mejora de la gestión hotelera y la experiencia del huésped en Buenaventura.

3 Antecedentes

El análisis de sentimientos aplicado al turismo ha despertado un creciente interés académico a nivel internacional, impulsado por el auge de plataformas digitales y la relevancia de las opiniones en línea como insumo para la toma de decisiones (Manosso & Domareski, 2021). En Europa, diversos estudios han demostrado la utilidad de automatizar la clasificación de reseñas para segmentar clientes y optimizar la gestión hotelera mediante minería de texto y aprendizaje automático (Oliveira et al., 2022; Bagherzadeh et al., 2021; Ahani et al., 2019). Este desarrollo ha ido acompañado de propuestas orientadas a mejorar la transparencia de los algoritmos, como el uso de técnicas interpretables tipo LIME, que buscan explicar el funcionamiento de modelos complejos de NLP o de tipo “caja negra” (Atrey, 2023).

En el plano metodológico, distintos trabajos académicos han explorado soluciones que integran scraping, minería de datos, visualizaciones y marcos de referencia consolidados como CRISP-DM. Este último se ha consolidado como estándar en proyectos de minería de datos, pues estructura el proceso en fases replicables y ha sido utilizado en investigaciones en turismo y sectores afines. Por ejemplo, Barrera y Barrera (2022) aplicaron un proceso CRISP-DM para clasificar reseñas gastronómicas, mientras que Giehl (2018) y Valle (2021) lo adaptaron a la recolección automatizada de datos y validación de modelos. Estos trabajos demostraron que marcos estructurados permiten organizar de manera rigurosa los proyectos basados en datos.

Asimismo, se desarrollaron estrategias de clasificación temática que permitieron diferenciar aspectos dentro de las reseñas, como calidad del servicio, infraestructura o limpieza, aportando un nivel de análisis más granular útil para investigaciones posteriores (Coello, 2023). En esta misma línea, Rashidi (2024) subrayó la importancia de plantear proyectos realistas, fundamentados en tecnologías accesibles y con decisiones técnicas justificadas, lo que refuerza la pertinencia de emplear metodologías disponibles para organizar y representar información turística.

Desde el enfoque sectorial, la literatura reciente ha mostrado cómo el análisis de sentimientos contribuye a identificar factores clave de satisfacción del cliente en servicios turísticos. Guadarrama-Gómez et al. (2024) analizaron más de 7.500 reseñas de establecimientos de alimentos en Huatulco (México), demostrando que la mayoría de opiniones tenían carga positiva y que los aspectos negativos eran puntuales, lo que facilitó mejoras estratégicas en los servicios. En hotelería, Araya-Pizarro y Verelst (2024) identificaron, mediante minería de textos y análisis de sentimientos aplicados a reseñas de Booking.com en hoteles de Santiago de Chile, que variables

como infraestructura, ubicación, comida, personal y relación precio-calidad son los principales determinantes de experiencias positivas. Estos hallazgos reafirman el potencial del análisis de sentimientos para guiar decisiones basadas en evidencia.

En cuanto al análisis a escala de destinos turísticos, la minería de opiniones se ha utilizado para evaluar percepciones agregadas. Herrera y Salgado (2022) emplearon técnicas de big data y análisis de sentimientos para calificar servicios turísticos en Quito, mientras que Guerrero Rodríguez et al. (2018) evidenciaron en Guanajuato valoraciones similares entre turistas nacionales y extranjeros frente a los principales atractivos. Estos hallazgos ayudan a identificar elementos comunes de satisfacción y apoyan tanto el diseño de políticas públicas como el fortalecimiento del posicionamiento de marca de un destino.

En términos de innovación, destacan experiencias como el foro Rest-Mex (2021), donde se entrenaron modelos avanzados de NLP sobre más de 7.400 reseñas en español, logrando mejoras significativas en la clasificación de polaridad mediante técnicas de deep learning (Álvarez-Carmona et al., 2022). De igual forma, la investigación de Yannacopoulou y Kallinikos (2025) en Kastoria (Grecia) combinaron análisis de sentimientos/percepción (vía texto y fotos de usuarios) con análisis de contenido visual y categorización espacial de atractivos naturaleza, paisaje, monumentos, etc. para ofrecer una visión integrada de la imagen del destino, incluyendo dimensiones perceptivas, espaciales y afectivas.

También se ha estudiado el efecto de la interacción entre los establecimientos y los usuarios. Un estudio de TripAdvisor e Ipsos MORI (2019) reveló que el 77% de los viajeros estaría más dispuesto a reservar si el propietario del hotel responde de forma personalizada a las reseñas, y que el 89% mejora su percepción del negocio cuando recibe una respuesta cuidada, incluso ante críticas negativas. De manera similar, investigaciones del Cornell Center for Hospitality Research han encontrado que responder reseñas especialmente las negativas se asocia con mejoras en la reputación y visibilidad del hotel, aunque con rendimientos decrecientes más allá del 40% de respuestas (Anderson & Han, 2016). Estos hallazgos refuerzan el valor estratégico de gestionar activamente la reputación digital mediante respuestas oportunas.

Por otro lado, la percepción de seguridad ha sido identificada como un factor determinante en la elección de destino, incluso por encima de atributos como ubicación o calidad del servicio. Chan y Lam (2013) demostraron que los sistemas de seguridad y protección son elementos estratégicos de decisión, especialmente en contextos donde la percepción de riesgo es alta. Este hallazgo fue

reforzado por estudios posteriores que vinculan directamente la sensación de seguridad con la intención de regresar a un alojamiento (Rittichainuwat, 2011), destacando que sin una percepción de seguridad básica, otros atributos valorados pueden quedar neutralizados.

Finalmente, resulta importante considerar las plataformas digitales utilizadas como fuente de información para el análisis de sentimientos. Diversas plataformas concentran el contenido generado por los viajeros, siendo TripAdvisor, Booking y Google Reviews algunas de las más utilizadas. TripAdvisor fomenta comentarios más extensos debido a sus requisitos mínimos de publicación y permite segmentar las opiniones por tipo de viajero, fecha o nivel de valoración, lo que la convierte en una fuente útil para estudios cualitativos detallados. En cambio, Google Reviews sobresale por su volumen masivo de datos, facilitado por su integración con Google Maps y Android, aunque muchas reseñas carecen de contenido textual profundo. Un estudio comparativo reciente en destinos enológicos de La Rioja (España), que recopiló 26.078 reseñas de Google y 9.998 de TripAdvisor, evidenció que Google ha superado a TripAdvisor en cantidad y actualidad de opiniones en ciertos contextos, posicionándose como una fuente más eficaz para el monitoreo continuo de percepciones turísticas (Piñero & Mellinas, 2025). Estas características fundamentan su elección como insumo principal en la presente investigación, complementando la rigurosidad metodológica con la amplitud y representatividad de los datos analizados.

4 Referente Normativo y Legal

Este capítulo presenta la normativa y los instrumentos jurídicos que enmarcaron el desarrollo del proyecto, organizados jerárquicamente según la pirámide de Hans Kelsen. En la cúspide se encuentra la Constitución Política de Colombia (1991), cuyo artículo 52 reconoce el turismo como un “derecho social y económico” y promueve su desarrollo en armonía con el entorno natural y cultural. En desarrollo de este mandato, se expidió la Ley 300 de 1996 (Ley General de Turismo), que define el turismo como un sector de libre iniciativa privada e introduce la obligatoriedad del Registro Nacional de Turismo para los prestadores de servicios. Posteriormente, la Ley 1558 de 2012 fortaleció los principios de calidad, descentralización y sostenibilidad, mientras que la Ley 2068 de 2020 actualizó la normativa incorporando estándares vinculados con la sostenibilidad ambiental y la protección del turista.

En el ámbito reglamentario, el Decreto 1074 de 2015 (Decreto Único Reglamentario del sector Comercio, Industria y Turismo) estableció normas operativas para los prestadores, incluyendo la inscripción anual y estándares de servicio, y el Decreto 1379 de 2021 reglamentó el guionaje turístico, definiendo competencias y requisitos para los guías registrados. A estas disposiciones se suma la Ley 1581 de 2012 sobre protección de datos personales, que regula el tratamiento de la información e incluye excepciones para el uso de datos públicos con fines científicos. En el marco de este proyecto, las reseñas utilizadas provinieron de acceso público y fueron debidamente anonimizadas, omitiendo cualquier dato que permitiera la identificación de los usuarios, en cumplimiento estricto de esta normativa.

A nivel local, el Acuerdo Municipal 03 de 2001 adoptó el Plan de Ordenamiento Territorial de Buenaventura, en el que se reconoce al turismo como motor económico y se promueve su articulación con el desarrollo urbano y portuario. Además, en junio de 2025, la Alcaldía Distrital y la Gobernación del Valle del Cauca presentaron un prototipo de plataforma web interactiva orientada a promover el turismo local, integrando hoteles, restaurantes y empresarios mediante soluciones digitales y sistemas de incentivos (Alcaldía Distrital de Buenaventura, 2025). Asimismo, el Congreso de Turismo Responsable 2025, titulado “Trenzando Turismo Responsable y Sostenible: Buenaventura, un sueño en común”, reflejó el interés institucional por fortalecer la oferta turística mediante iniciativas sostenibles y apoyadas en plataformas digitales (Soy de Buenaventura, 2025).

En conjunto, este marco normativo legitimó el análisis de la percepción turística como una actividad que contribuyó al desarrollo social y económico de Buenaventura. La normativa sobre protección de datos personales habilitó el uso académico de reseñas públicas, siempre que se garantizara la privacidad, lo cual fue cumplido rigurosamente. Asimismo, las leyes sectoriales del turismo establecieron estándares de calidad cuya percepción fue analizada desde la experiencia de los visitantes. Finalmente, los acuerdos y proyectos municipales ofrecieron respaldo institucional a la digitalización del turismo, reforzando la oportunidad de esta investigación.

5 Referente Teórico

Este capítulo integró los fundamentos conceptuales que sustentaron el desarrollo del proyecto, organizados en dos ejes principales: (1) fundamentos técnicos del análisis de sentimientos y (2) fundamentos conceptuales de la percepción turística. Esta estructura articuló los aportes metodológicos y teóricos necesarios para abordar el problema desde una perspectiva analítica, contextual e interpretativa.

5.1 Fundamentos Técnicos del Análisis de Sentimientos

Análisis de Sentimientos

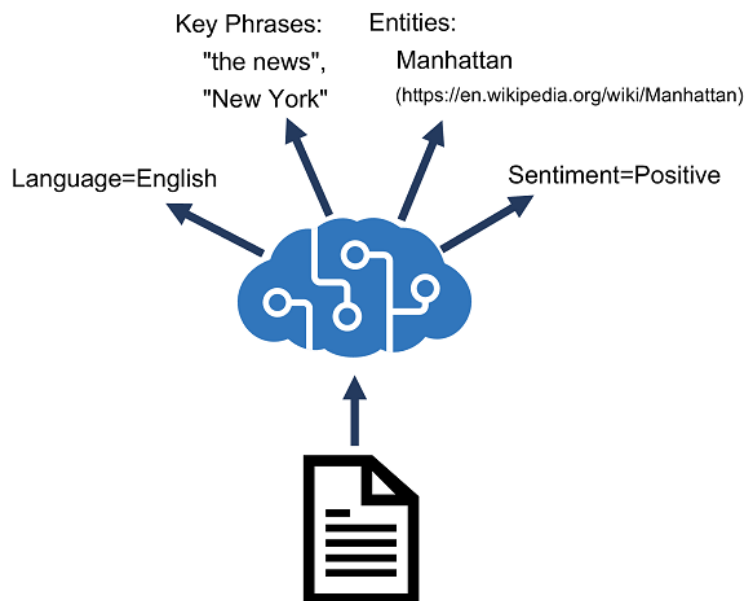
El análisis de sentimientos constituyó la técnica central de este proyecto, al identificar la carga emocional de los textos escritos por usuarios y clasificarla en categorías como positiva, negativa o neutra (Liu, B, 2012). Su aplicación a las reseñas hoteleras de Buenaventura ofreció una vía práctica para comprender la percepción turística desde el comportamiento digital.

Procesamiento de Lenguaje Natural (NLP)

Esta rama de la inteligencia artificial fue una de las tecnologías clave empleadas en este proyecto, al orientar el procesamiento computacional del lenguaje humano (Zhang et al., 2018). Incluye operaciones como tokenización, lematización, identificación de idioma, extracción de frases clave, reconocimiento de entidades y clasificación de sentimientos. Para la implementación del proyecto, se utilizó Azure Cognitive Services – Text Analytics, servicio preentrenado que procesó automáticamente los textos recopilados sin necesidad de construir modelos desde cero, como se muestra en la Figura 4.

Figura 4.

Esquema del análisis de sentimientos mediante Azure Cognitive Services – Text Analytics



Nota. Tomado de *Analyze Text with AI Language* [Gráfico], por Microsoft, s.f., en *Microsoft Learn*. <https://tinyurl.com/2yn6rwk6>

Este servicio operó mediante modelos preentrenados de NLP, capaces de determinar la polaridad emocional de un texto (positiva, neutra o negativa) y devolver un puntaje de confianza para cada categoría. La clasificación se fundamentó en el análisis contextual y semántico de las frases, facilitando la sistematización de percepciones subjetivas expresadas por los usuarios en grandes volúmenes de texto.

Enfoques Modernos para el Análisis Automatizado de Sentimientos

El análisis automatizado de sentimientos ha evolucionado desde enfoques léxicos y modelos supervisados clásicos hacia arquitecturas de aprendizaje profundo, modelos transformer y, más recientemente, modelos de lenguaje de gran escala. De acuerdo con Mao et al. (2024), esta evolución refleja el interés por mejorar la capacidad de los sistemas para interpretar textos no estructurados, capturar relaciones semánticas y adaptarse a dominios específicos. En los enfoques

léxicos, como VADER, la clasificación se apoya en diccionarios de palabras y reglas predefinidas de polaridad, lo que permite implementaciones simples, rápidas e interpretables (Hutto & Gilbert, 2014). Sin embargo, este tipo de métodos puede presentar limitaciones frente a textos en español, expresiones locales, negaciones, ironía o reseñas mixtas, donde una misma opinión combina elementos positivos y negativos.

Por otra parte, los enfoques de machine learning clásico abordan el análisis de sentimientos como una tarea supervisada de clasificación, utilizando algoritmos como Naive Bayes, Maximum Entropy o Support Vector Machines a partir de características extraídas del texto (Pang et al., 2002). Estos métodos pueden adaptarse a dominios específicos, pero requieren un corpus previamente etiquetado para entrenamiento y evaluación. En el caso de estudios exploratorios sobre reseñas reales, como el presente, esta condición puede representar una restricción metodológica cuando no se dispone inicialmente de una base de datos manualmente clasificada como referencia.

El desarrollo de la arquitectura transformer marcó un cambio relevante en el NLP, al introducir mecanismos de atención capaces de analizar relaciones entre palabras sin depender estrictamente del orden secuencial tradicional (Vaswani et al., 2017). A partir de esta línea, modelos preentrenados como BERT permitieron representar el significado de las palabras considerando simultáneamente su contexto anterior y posterior, lo cual favorece la interpretación de frases ambiguas o semánticamente complejas (Devlin et al., 2019). Más recientemente, los modelos de lenguaje de gran escala han ampliado las posibilidades del análisis de sentimientos al permitir tareas más flexibles, como clasificación contextual, explicación de opiniones y análisis de sentimientos basado en aspectos; no obstante, también introducen retos asociados con costo computacional, reproducibilidad, control de salidas y auditabilidad metodológica (Zhang et al., 2024).

En este contexto, el presente proyecto no tuvo como propósito entrenar ni comparar experimentalmente modelos léxicos, modelos clásicos de machine learning, BERT o modelos de lenguaje de gran escala (LLM), sino aplicar un servicio preentrenado y administrado de análisis de sentimientos sobre un corpus real de reseñas hoteleras. Esta decisión respondió a la naturaleza exploratorio-descriptiva del estudio y a las características de los datos disponibles, dado que las reseñas analizadas correspondían al periodo 2020–2025 y no existía inicialmente un corpus etiquetado de referencia que permitiera entrenar y evaluar un modelo supervisado propio. En este

sentido, enfoques como machine learning clásico o ajuste fino de modelos basados en BERT habrían requerido una fase previa de anotación manual, definición de criterios de etiquetado, entrenamiento, ajuste de hiperparámetros y validación con métricas supervisadas, actividades que excedían el alcance del proyecto.

Frente a estas alternativas, Azure Cognitive Services – Text Analytics ofreció una solución pertinente por su disponibilidad como servicio preentrenado, su compatibilidad con textos en español y su capacidad para clasificar polaridad, generar puntajes de confianza y extraer elementos semánticos relevantes sin desarrollar un algoritmo desde cero. Esta elección también respondió a criterios de viabilidad técnica, temporal y económica, dado que permitió aprovechar los créditos académicos disponibles en la plataforma Azure y concentrar el esfuerzo investigativo en la recolección, depuración, estructuración, validación e interpretación de las reseñas hoteleras. Además, al no tratarse de un sistema en tiempo real ni de un modelo que requiriera aprendizaje continuo, el uso de un servicio preentrenado resultó coherente con el objetivo de analizar retrospectivamente un corpus delimitado.

No obstante, este enfoque presenta limitaciones que deben ser reconocidas. Al tratarse de un servicio propietario, no es posible inspeccionar completamente la arquitectura interna del modelo, sus pesos, reglas de decisión ni criterios exactos de clasificación. Por tanto, aunque sus salidas permiten aproximarse de manera sistemática a la percepción expresada por los usuarios, no deben interpretarse como una medición absoluta ni plenamente explicable de la experiencia del huésped. Asimismo, como ocurre en tareas de NLP, pueden presentarse errores ante reseñas muy cortas, ambiguas, irónicas o mixtas, especialmente cuando coexisten valoraciones positivas y negativas dentro de un mismo comentario. Adicionalmente, al no tratarse de un modelo supervisado entrenado por el investigador ni partir de un corpus previamente etiquetado como verdad de referencia, no resultaba metodológicamente adecuado calcular métricas como precisión, recall o F1-score, pues estas requieren contrastar las predicciones del modelo con etiquetas reales previamente definidas y validadas. Por esta razón, los resultados fueron interpretados de forma agregada y complementados con una validación manual externa orientada a contrastar la coherencia interpretativa de las etiquetas generadas.

Análisis de Sentimientos Basado en Aspectos (ABSA)

Esta aproximación representa una forma más granular que el análisis de sentimientos tradicional, debido a que no se limita a clasificar un texto completo como positivo, negativo o neutro, sino que busca identificar los aspectos específicos sobre los cuales se expresa una opinión y la polaridad asociada a cada uno de ellos. En el contexto hotelero, esta perspectiva resulta especialmente pertinente, ya que una misma reseña puede contener valoraciones diferenciadas sobre dimensiones como habitaciones, limpieza, atención, restaurante, ubicación, infraestructura o relación calidad-precio. Por ejemplo, un usuario puede valorar positivamente la ubicación del hotel, pero expresar inconformidad frente al estado de la habitación o la atención recibida; en estos casos, la polaridad global de la reseña no representa completamente la complejidad de la experiencia turística.

Estudios recientes han resaltado la utilidad de ABSA para generar conocimiento más detallado sobre la experiencia del cliente en contextos turísticos y hoteleros. Água et al. (2025) señalan que el análisis basado en aspectos, apoyado en modelos de lenguaje de gran escala, permite obtener perspectivas más precisas sobre la experiencia del cliente al desagregar las opiniones en atributos específicos del servicio. Esta perspectiva resulta coherente con el presente proyecto, dado que el análisis de las reseñas no se orientó únicamente a determinar la polaridad general, sino también a identificar temas recurrentes asociados con satisfacción o insatisfacción, tales como limpieza, habitaciones, atención, alimentación, infraestructura, entorno y respuesta del establecimiento.

En este estudio no se entrenó un modelo ABSA propio; en su lugar, se utilizó Azure Cognitive Services – Text Analytics como un servicio preentrenado que ya incorporaba capacidades predeterminadas para el análisis de sentimientos y la extracción de elementos semánticos relevantes en textos no estructurados. Esta funcionalidad permitió identificar frases de opinión, aspectos mencionados y temas recurrentes asociados a las opiniones expresadas por los usuarios, lo cual facilitó una lectura más detallada de los factores que incidían en la satisfacción o insatisfacción hotelera. Por tanto, el proyecto se posiciona como una aproximación aplicada al paradigma ABSA, en la medida en que combina la polaridad general de las reseñas con la identificación y agrupación temática de aspectos de la experiencia hotelera.

Esta decisión permitió pasar de una lectura agregada del sentimiento a una interpretación más específica de los factores que explican la percepción de los huéspedes. No obstante, los

aspectos identificados deben entenderse como categorías analíticas construidas para interpretar el corpus, no como el resultado de un modelo ABSA entrenado y validado específicamente para el dominio hotelero de Buenaventura. En consecuencia, el valor del enfoque radica en su utilidad aplicada para organizar y explicar patrones temáticos dentro de las reseñas, manteniendo claridad sobre sus límites metodológicos.

Minería de Texto

La minería de texto consiste en extraer conocimiento útil a partir de datos no estructurados como reseñas, comentarios o publicaciones digitales (Feldman & Sanger, 2007). En el contexto de este trabajo, la técnica organizó y convirtió información dispersa en insumos analíticos confiables, útiles para sistematizar la percepción de los visitantes.

Metodología CRISP-DM

Para estructurar el proceso analítico, se adoptó el modelo CRISP-DM (Cross-Industry Standard Process for Data Mining), el cual guió las fases de comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación y despliegue (Chapman et al., 2000). En este estudio, cada fase fue adaptada al contexto turístico: la comprensión del negocio permitió delimitar los factores que afectaban la percepción de Buenaventura; la comprensión de los datos facilitó el reconocimiento de la estructura, calidad y utilidad del corpus obtenido; la preparación de los datos implicó la recolección, limpieza, deduplicación, normalización de formatos y depuración de reseñas de Google Reviews; el modelado se centró en la aplicación del análisis automatizado de sentimientos; y la evaluación consistió en interpretar los resultados frente a la reputación digital y competitividad turística del destino.

Desde una perspectiva teórica, CRISP-DM no solo funcionó como una secuencia operativa de fases, sino como un marco de trazabilidad y validación del conocimiento generado a partir de datos. En la fase de comprensión del negocio, el modelo permitió vincular el análisis técnico con una problemática turística concreta de Buenaventura, apoyada en la revisión del contexto local y en el conocimiento situado del investigador sobre el territorio, sin sustituir la validación empírica del proceso. En la fase de comprensión de los datos, el análisis exploratorio permitió identificar características relevantes del corpus extraído mediante Apify, tales como reseñas muy cortas, columnas redundantes, campos con baja utilidad analítica y variables necesarias para segmentar la

información por hotel, fecha, idioma o respuesta del establecimiento. Posteriormente, la fase de preparación de los datos permitió transformar esa comprensión inicial en decisiones metodológicas documentadas, como la depuración de columnas innecesarias, el control de duplicados, la normalización de campos y la selección de reseñas con contenido textual suficiente para el análisis semántico.

En la fase de modelado, el uso de Azure Cognitive Services – Text Analytics permitió aplicar análisis automatizado de sentimientos y extracción de elementos semánticos sobre el corpus depurado, sin que el proyecto dependiera del entrenamiento de un modelo supervisado propio. La fase de evaluación aportó un mecanismo de contraste mediante validación manual externa, orientado a revisar la coherencia interpretativa entre las etiquetas automáticas y la lectura humana de una muestra de reseñas. Finalmente, la fase de despliegue permitió representar los resultados en Power BI mediante indicadores, gráficos y tablas temáticas útiles para la interpretación turística. En conjunto, CRISP-DM fortaleció la coherencia entre problema, datos, transformación, análisis, validación y visualización, reduciendo decisiones arbitrarias y fortaleciendo la reproducibilidad del proceso analítico.

5.2 Fundamentos conceptuales de la percepción turística

Percepción Turística

La percepción turística fue entendida como el conjunto de imágenes, opiniones y emociones que los visitantes construyen acerca de un destino, a partir de sus experiencias personales, las reseñas de otros usuarios y factores externos como los medios de comunicación (Echtner & Ritchie, 1991). En el caso específico de Buenaventura, donde confluyen tanto atractivos naturales como problemáticas sociales, el análisis de esta percepción identificó fortalezas del destino y áreas susceptibles de mejora.

Contenido Generado por Usuarios (UGC)

El contenido generado por usuarios (User-Generated Content, UGC) incluyó opiniones, reseñas y calificaciones publicadas voluntariamente en plataformas digitales. Este tipo de contenido se ha convertido en una de las fuentes más influyentes para el turismo, ya que los viajeros tienden a confiar más en la experiencia de otros turistas que en mensajes publicitarios

tradicionales (Gretzel & Yoo, 2008). En esta investigación, las reseñas de Google Reviews constituyeron el UGC principal; su análisis reflejó, de forma sistemática, la percepción turística del destino Buenaventura.

Online Travel Reviews (OTR)

Dentro del universo del contenido generado por usuarios, las reseñas turísticas en línea u *Online Travel Reviews* (OTR) representan una categoría particularmente influyente en los procesos de toma de decisiones de los viajeros. Estas reseñas, publicadas de forma espontánea por huéspedes en plataformas digitales, no solo reflejan experiencias personales, sino que también moldean la imagen y reputación de los destinos y establecimientos turísticos (Filieri et al., 2015). Las OTR combinan componentes cognitivos (por ejemplo, comentarios sobre limpieza, infraestructura o servicio) con componentes afectivos (emociones, juicios, satisfacción), lo que las convierte en insumos valiosos para el análisis de percepciones. En este proyecto, las OTR recopiladas desde Google Reviews permitieron construir una representación detallada de la experiencia turística en Buenaventura, al condensar un volumen significativo de valoraciones reales, actualizadas y referidas a distintos aspectos de la estancia.

6 Metodología

La metodología aplicada orientó el cumplimiento sistemático de los objetivos específicos, además de la respuesta a la pregunta de investigación relacionada con la percepción turística de los hoteles más representativos de Buenaventura. Para ello, se desarrolló un proceso metodológico estructurado que combinó técnicas cualitativas y cuantitativas, integrando minería de texto, análisis automatizado de sentimientos, visualización analítica y evaluación interpretativa de los resultados.

Todas las actividades se realizaron bajo los lineamientos del modelo CRISP-DM, el cual guio de forma ordenada las fases de comprensión del problema, comprensión de los datos, preparación, modelado, evaluación y despliegue. Este marco metodológico resultó adecuado, pues organizó las tareas de análisis, además de asegurar coherencia con los objetivos planteados.

Como parte de la preparación de los datos, se implementó un proceso formal de extracción, transformación y carga (ETL), que garantizó limpieza, trazabilidad, reproducibilidad y control de versiones. Las reseñas fueron extraídas mediante Apify – Google Maps Reviews Scraper, transformadas mediante procedimientos de normalización, deduplicación y depuración, y posteriormente almacenadas en Azure Blob Storage y Supabase, organizadas en capas Bronze, Silver y Gold.

Una vez consolidado y depurado el conjunto de datos, se aplicó análisis automatizado de sentimientos mediante Azure Cognitive Services – Text Analytics, herramienta utilizada para clasificar la polaridad emocional de los textos, además de extraer aspectos temáticos mencionados por los visitantes. Los resultados del modelo fueron estructurados mediante vistas analíticas en SQL, diseñadas en Supabase, las cuales facilitaron la agregación por hotel, año, aspecto y polaridad, así como el cálculo de métricas descriptivas que responden directamente a los objetivos del estudio.

Posteriormente, los hallazgos fueron representados en paneles interactivos utilizando Power BI, donde se visualizaron comparaciones de sentimientos entre hoteles, tendencias anuales, aspectos críticos mencionados, distribución porcentual de emociones y métricas de gestión como la tasa de respuesta del establecimiento. Estas visualizaciones permitieron interpretar los patrones identificados y fundamentar recomendaciones orientadas a fortalecer la calidad del servicio hotelero en Buenaventura.

La metodología en su conjunto articuló la minería de texto, el análisis estadístico y la representación visual, ofreciendo un proceso riguroso, reproducible y alineado con los objetivos propuestos.

6.1 Enfoque Metodológico

El enfoque metodológico adoptado en este trabajo fue mixto secuencial exploratorio (QUAL → QUAN), lo que implicó una integración ordenada de técnicas cualitativas y cuantitativas para analizar la percepción turística expresada en las reseñas digitales de los hoteles más representativos de Buenaventura. Esta secuencia inició con una exploración interpretativa del contenido textual, seguida de una sistematización cuantitativa orientada a la agregación de resultados, además de la generación de indicadores.

En la fase cualitativa, se examinó el contenido de las reseñas para identificar patrones narrativos, temas recurrentes y aspectos del servicio mencionados por los visitantes. Esta revisión preliminar reconoció dimensiones clave como limpieza, atención, infraestructura, alimentación, habitaciones y experiencia del personal; estas orientaron las decisiones analíticas posteriores, además de facilitar la comprensión contextual antes del modelado automatizado.

En la fase cuantitativa, las descripciones identificadas se transformaron en métricas mediante el uso de Azure Cognitive Services – Text Analytics, herramienta que clasificó la polaridad emocional de los textos y extrajo aspectos temáticos asociados a valoraciones positivas o negativas. Estos resultados fueron estructurados mediante vistas analíticas en SQL; así, la información se organizó según hotel, año, polaridad y aspecto. A partir de esta estructura se construyeron indicadores sobre distribución de sentimientos, tendencias temporales y patrones lingüísticos relevantes.

La articulación de ambas aproximaciones integró los hallazgos cualitativos con su sistematización cuantitativa; con ello, fortaleció la interpretación de los resultados, además de mantener coherencia con los objetivos del estudio. En conjunto, el enfoque metodológico combinó lectura contextual, procesamiento automatizado y análisis estadístico para ofrecer una visión integral del fenómeno turístico desde los datos generados por los propios usuarios.

6.2 Tipo de Estudio

El presente trabajo se desarrolló bajo un diseño exploratorio–descriptivo, acorde con la naturaleza del problema y con el tipo de datos disponibles. Fue exploratorio en la medida en que abordó un fenómeno poco estudiado en el contexto local la percepción turística expresada en reseñas digitales de los hoteles más representativos de Buenaventura permitiendo identificar patrones emergentes, tendencias temporales y aspectos recurrentes que no habían sido previamente sistematizados.

Asimismo, fue descriptivo, dado que se orientó a caracterizar de manera detallada la distribución de sentimientos, los patrones lingüísticos y las variaciones de polaridad emocional entre hoteles y a lo largo del tiempo. Para ello se emplearon métricas cuantitativas como porcentajes, conteos, proporciones anuales y análisis comparativos entre establecimientos. Estas métricas representaron con claridad el comportamiento de los sentimientos positivos, negativos y neutrales en el periodo analizado.

El análisis incorporó además un componente interpretativo, mediante el cual se identificaron posibles hipótesis derivadas de los patrones encontrados, por ejemplo, incrementos inusuales de sentimiento negativo en ciertos periodos o recurrencia de aspectos criticados en varios años consecutivos con el propósito de orientar recomendaciones estratégicas para la gestión hotelera. Estas interpretaciones se basan en la evidencia empírica y buscan apoyar la toma de decisiones, sin constituir explicaciones causales estrictas.

Este enfoque exploratorio–descriptivo integró la identificación de patrones con la representación sistemática de los resultados; así, proporcionó una comprensión amplia de las percepciones turísticas expresadas en Google Reviews, además de insumos útiles para su análisis estratégico.

6.3 Procedimiento

El desarrollo del proyecto implicó una serie de actividades técnicas y analíticas realizadas de forma ordenada y progresiva. Este apartado documenta dichas acciones, agrupadas en fases, con el fin de mostrar cómo se obtuvo, procesó, enriqueció y analizó la información para responder a los objetivos planteados.

El proceso de extracción, transformación y carga (ETL) constituyó la base técnica del proyecto, pues obtuvo, depuró y estructuró los datos textuales que posteriormente fueron

analizados mediante técnicas de minería de texto y análisis automatizado de sentimientos. Este proceso se diseñó bajo una arquitectura de tres capas (Bronze, Silver y Gold), ampliamente utilizada en ingeniería de datos por su capacidad para garantizar trazabilidad, auditabilidad, control de versiones y reproducibilidad del flujo analítico.

La organización metodológica del pipeline se fundamentó en el modelo CRISP-DM (Cross-Industry Standard Process for Data Mining), debido a su enfoque iterativo y comprobado para proyectos de analítica. Cada una de sus fases se adaptó explícitamente al contexto de los hoteles más representativos de Buenaventura. En la fase de comprensión del negocio se identificó la necesidad estratégica de caracterizar la percepción turística sobre estos establecimientos, entendiendo la reputación digital como un factor clave para la competitividad del sector hotelero local. A partir de ello se definieron los objetivos del estudio, su alcance, los actores involucrados y los requerimientos analíticos.

En la fase de comprensión de los datos se exploró la estructura, formato, idioma, volumen y metadatos asociados a las reseñas de Google Reviews obtenidas mediante Apify. Esta inspección inicial identificó elementos relevantes, tales como reseñas sin contenido textual, campos redundantes, diversidad de idiomas y variables con baja variabilidad; estos incidieron en las decisiones de limpieza, filtrado y selección de columnas.

En la fase de preparación de los datos se desarrolló e implementó el proceso ETL, que incluyó la extracción automatizada de las reseñas, su depuración y estandarización, la normalización de tipos de datos, la eliminación de duplicados mediante el identificador de la reseña y la creación de variables derivadas, como el año y mes de publicación, la longitud del texto o la oportunidad de respuesta del establecimiento. El almacenamiento se organizó en capas Bronze, Silver y Gold para mantener una separación clara entre datos crudos, datos limpios y datos enriquecidos semánticamente.

La fase de modelado correspondió al análisis automatizado de sentimientos mediante Azure Cognitive Services – Text Analytics, el cual clasificó la polaridad emocional de cada reseña (positiva, neutra o negativa) y extrajo aspectos temáticos (aspects) asociados a valoraciones positivas o negativas. Estas salidas se integraron con las variables numéricas y categóricas seleccionadas durante la preparación de datos y se consolidaron en una tabla analítica central, utilizada posteriormente para la construcción de vistas en SQL y paneles en Power BI.

En la fase de evaluación se verificó la coherencia y utilidad interpretativa de los resultados, combinando la revisión descriptiva de los indicadores generados con la comprobación manual de una muestra de reseñas seleccionadas. Esta revisión contrastó la clasificación automática del sentimiento con el contenido textual y la calificación en estrellas, con el fin de verificar la consistencia semántica, además de la pertinencia contextual para el análisis turístico en Buenaventura. Los valores obtenidos, así como los ejemplos representativos de coincidencias, divergencias y casos ambiguos, se presentan en el capítulo de Resultados, dado que esta sección se limita a describir el procedimiento de validación aplicado.

Finalmente, la fase de despliegue consistió en la integración de los datos enriquecidos en Supabase, la construcción de vistas analíticas en SQL para facilitar la segmentación y comparación entre hoteles y periodos, y el diseño de dashboards interactivos en Power BI. Estos paneles permitieron visualizar la distribución de sentimientos por hotel, las tendencias temporales, los aspectos más criticados o valorados y métricas de gestión como el porcentaje de reseñas respondidas por cada establecimiento. De esta manera, el procedimiento implementado articuló la ruta completa desde la obtención de las reseñas hasta la presentación de resultados interpretables para los actores del sector.

Como evidencia complementaria del proceso técnico desarrollado, se incluye el repositorio del proyecto, en el cual se documenta la estructura general, los archivos, scripts y componentes implementados durante el flujo analítico descrito en esta metodología (véase Anexo C).

Además, el estudio adoptó un criterio temporal explícito. Aunque el scraper extrajo inicialmente todas las reseñas disponibles en los perfiles de Google Maps de cada hotel, el análisis se restringió posteriormente al periodo comprendido entre el 1.º de enero de 2020 y el 21 de agosto de 2025, con el propósito de capturar percepciones contemporáneas y variaciones recientes en la reputación de los establecimientos.

Extracción

La etapa de extracción se llevó a cabo a partir de fuentes públicas de acceso abierto, específicamente las reseñas publicadas por usuarios en Google Reviews. Para su obtención se utilizó la herramienta Apify – Google Maps Reviews Scraper para automatizar la descarga de reseñas textuales, junto con los metadatos asociados, tales como fecha de publicación, idioma, calificación numérica, información básica del autor y respuesta del establecimiento.

Para iniciar el proceso se configuraron en el scraper las direcciones URL correspondientes a los perfiles oficiales en Google Maps de los cinco hoteles seleccionados: Hotel Cordillera, Cosmos Pacífico, Magüipí, Torre Mar y Steven Buenaventura. El criterio de selección se basó en el volumen de reseñas disponibles, de manera que el análisis se apoyara en un conjunto de datos suficientemente amplio y estable para obtener resultados comparativos entre establecimientos.

Figura 5.

Configuración inicial del scraper con las URL de los hoteles seleccionado

Provide either Google place URLs. You will get all the reviews that are connected with the selected Google Place.

Google Maps place URLs (optional) ?

https://maps.app.goo.gl/sjsdc63TA9xMnxfZ9	Advanced	×
https://maps.app.goo.gl/5di9S1YzdqCK66p67	Advanced	×
https://maps.app.goo.gl/P8T6LT7fcPAEt54m7	Advanced	×
https://maps.app.goo.gl/w7di32u8hfDoeABK6	Advanced	×
https://maps.app.goo.gl/TwxuxMuHGXFpbsKD8	Advanced	×

Posteriormente se estableció un límite máximo de reseñas por hotel con el fin de evitar truncamientos indeseados y asegurar una recolección completa dentro de las capacidades técnicas del sistema. Se priorizó la extracción de reseñas más recientes con el propósito de obtener información pertinente al periodo de interés.

Figura 6.

Parámetro de límite de reseñas por establecimiento en Apify

Number of reviews (optional) ?

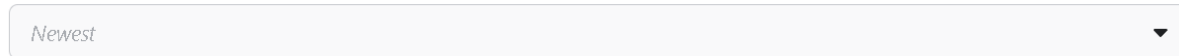
10000 reviews per place

Con el fin de preservar la secuencia temporal de las percepciones, se configuró el scraper para ordenar las reseñas desde las más recientes hasta las más antiguas. Esta configuración facilitó la reconstrucción posterior de patrones y tendencias a lo largo de los años considerados.

Figura 7.

Configuración del criterio de ordenamiento de reseñas (más recientes primero)

Sort reviews by (optional) ⓘ

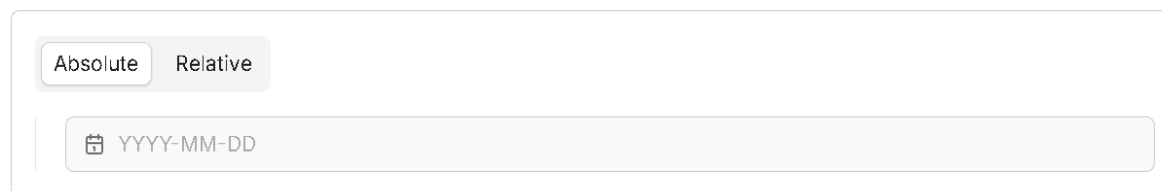
A screenshot of a web interface showing a dropdown menu for sorting reviews. The text 'Sort reviews by (optional) ⓘ' is above the dropdown. The dropdown itself is light gray with the word 'Newest' in a darker gray font and a small downward arrow on the right side.

La herramienta no ofrecía una opción para fijar simultáneamente una fecha inicial y una fecha final en la interfaz de extracción; por tal razón, el scraper recuperó todas las reseñas disponibles para cada hotel. El rango temporal utilizado en el análisis (2020–2025) se definió posteriormente en la fase de transformación, mediante filtros aplicados sobre la fecha de publicación.

Figura 8.

Configuración predeterminada del parámetro temporal en Apify

Only scrape reviews newer than [date] (optional) ⓘ

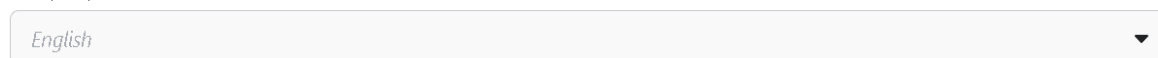
A screenshot of the Apify configuration interface. It shows two tabs: 'Absolute' (selected) and 'Relative'. Below the tabs is a date input field with a calendar icon and the placeholder text 'YYYY-MM-DD'.

Respecto al idioma, se mantuvo la configuración predeterminada de Apify, sin aplicar filtros restrictivos, de modo que se conservaron reseñas tanto de visitantes nacionales como extranjeros. Esta decisión conservó la riqueza original del corpus; al mismo tiempo, facilitó el análisis de textos en distintos idiomas mediante variables específicas.

Figura 9.

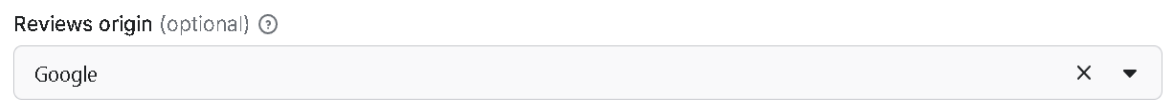
Configuración por defecto sin filtro de idioma

Language (optional) ⓘ

A screenshot of a web interface showing a dropdown menu for language selection. The text 'Language (optional) ⓘ' is above the dropdown. The dropdown itself is light gray with the word 'English' in a darker gray font and a small downward arrow on the right side.

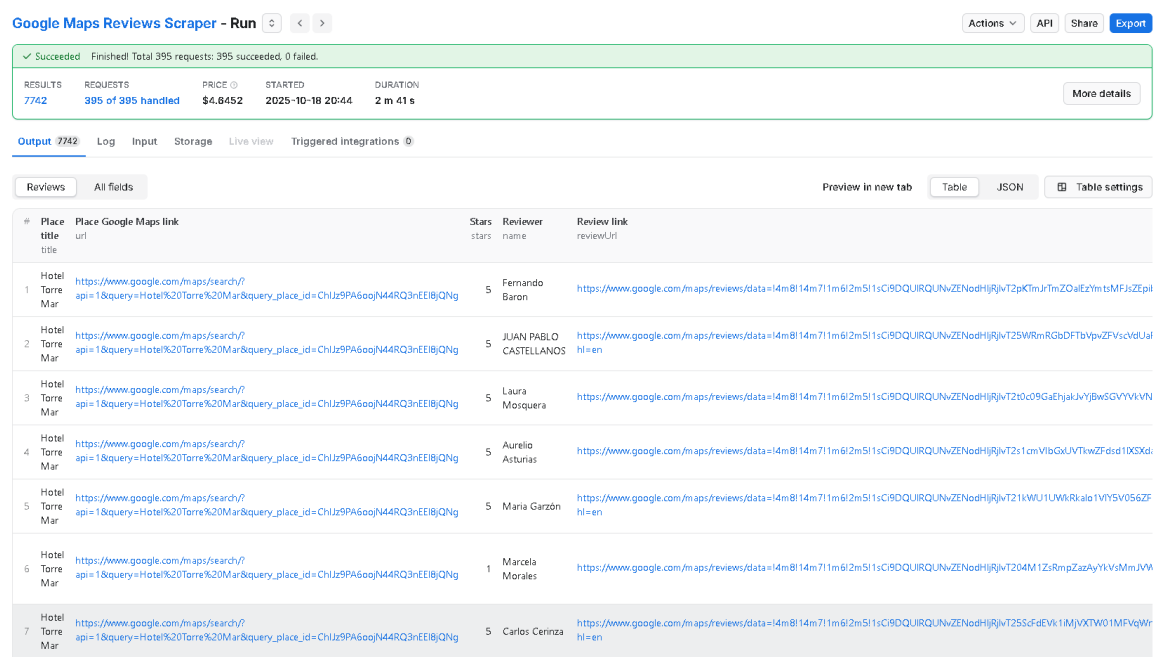
En cuanto al origen de los comentarios, se especificó que únicamente se consideraran reseñas provenientes de Google, dado que el estudio se centró exclusivamente en dicha plataforma y no contempló la integración con otras fuentes como TripAdvisor u otras.

Figura 10.
Selección del origen de las reseñas desde Google



Una vez definidos los parámetros, se ejecutó el scraper para cada hotel. El proceso finalizó satisfactoriamente, obteniendo un total de 7.742 reseñas distribuidas entre los cinco establecimientos, lo que proporcionó una base suficiente para la posterior depuración y análisis.

Figura 11.
Ejecución del scraper y resultados generales de la extracción

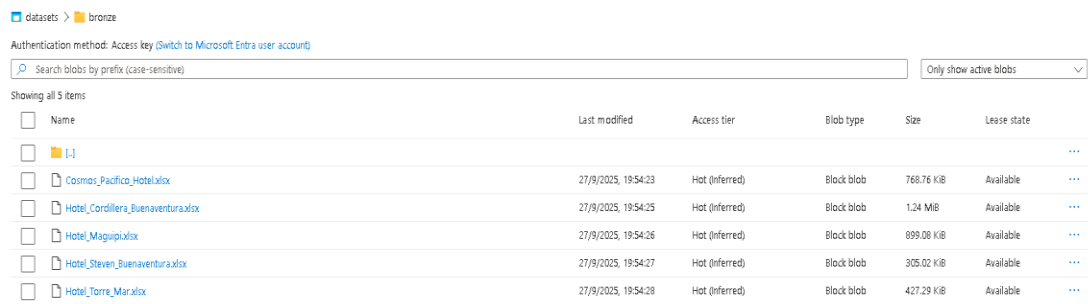


Transformación

La segunda fase correspondió a la transformación de los datos, en la cual se aplicaron procedimientos de limpieza, estandarización, validación y modelado estructural con el propósito de convertir los archivos crudos provenientes de la capa Bronze en datos consistentes, homogéneos y analíticamente utilizables. Este proceso garantizó la calidad y la trazabilidad necesarias para los análisis estadísticos y descriptivos desarrollados posteriormente.

En la capa Bronze se almacenaron los archivos originales descargados desde Apify, principalmente en formato Excel (.xlsx), sin modificaciones. Esta capa funcionó como respaldo fiel del estado inicial de la información.

Figura 12.
Archivos originales descargados desde Apify almacenados en la capa Bronze



Showing all 5 items						
☐ Name	Last modified	Access tier	Blob type	Size	Lease state	
☐ [.]						...
☐ Cosmos_Pacifico_Hotel.xlsx	27/9/2025, 19:54:23	Hot (Inferred)	Block blob	766.76 KiB	Available	...
☐ Hotel_Cordillera_Buenaventura.xlsx	27/9/2025, 19:54:25	Hot (Inferred)	Block blob	1.24 MiB	Available	...
☐ Hotel_Magupi.xlsx	27/9/2025, 19:54:26	Hot (Inferred)	Block blob	899.08 KiB	Available	...
☐ Hotel_Steven_Buenaventura.xlsx	27/9/2025, 19:54:27	Hot (Inferred)	Block blob	305.02 KiB	Available	...
☐ Hotel_Torre_Mar.xlsx	27/9/2025, 19:54:28	Hot (Inferred)	Block blob	427.29 KiB	Available	...

A continuación, se realizó la selección de columnas relevantes para el análisis. Se conservaron únicamente las variables necesarias para el análisis de sentimientos, la caracterización de los hoteles, la descripción básica de los revisores, la medición de la respuesta del establecimiento y la trazabilidad temporal y técnica del proceso. En la Tabla 1 se presentan las principales columnas incluidas y su justificación técnica.

Tabla 1.
Columnas incluidas en el análisis y justificación técnica

Columna	Justificación técnica
reviewId	Identificador único de cada reseña; permite controlar duplicados y asegurar unicidad.
placeId	Identificador estable del establecimiento; segmenta correctamente por hotel.
title	Complementa el texto y enriquece el contexto semántico.
text	Insumo principal para el análisis de sentimientos.
publishedAtDate	Permite análisis temporales y construcción de tendencias.
stars	Indicador cuantitativo que facilita comparaciones y resúmenes agregados.
textTranslated	Facilita el análisis uniforme cuando la reseña no está en español.

Columna	Justificación técnica
originalLanguage	Permite segmentar por idioma y examinar diversidad lingüística.
reviewOrigin	Identifica la fuente de la reseña (Google).
totalScore	Métrica global del hotel útil para contextualizar la reputación.
reviewsCount	Indica el volumen global de reseñas del establecimiento.
hotelStars	Contextualiza las expectativas asociadas al nivel del hotel.
price	Útil para contrastar percepción de servicio y rango de precios.
isLocalGuide	Señal de experiencia o credibilidad del revisor.
reviewerNumberOfReviews	Indica actividad previa del revisor en la plataforma.
likesCount	Mide la recepción social de la reseña.
responseFromOwnerText	Aporta información sobre la gestión de la respuesta del hotel.
responseFromOwnerDate	Permite medir la oportunidad temporal de la respuesta.
reviewContext/...	Subcategorías temáticas que describen el contexto de la experiencia.
reviewDetailedRating/...	Subpuntuaciones específicas del servicio (por ejemplo, limpieza o ubicación).
scrapedAt	Marca de auditoría que respalda la trazabilidad técnica del proceso.
categoryName	Verificación de pertenencia a la categoría “Hotel/Lodging”.

Nota. Las columnas incluidas se seleccionaron según su utilidad para el análisis de sentimientos, la exploración descriptiva y la segmentación comparativa entre hoteles.

Una vez seleccionadas y depuradas las variables relevantes, los datos se almacenaron en la capa Silver utilizando formatos optimizados para consulta analítica. En esta etapa se realizaron transformaciones clave orientadas a asegurar consistencia y calidad. Se normalizaron los tipos de datos, unificando formatos de fecha, estandarizando variables numéricas y homogenizando campos booleanos. Se aplicó limpieza textual sobre las reseñas, eliminando enlaces, caracteres residuales y símbolos no lingüísticos, con el propósito de mejorar la calidad del texto sin alterar su significado.

Asimismo, se llevó a cabo un proceso de deduplicación, para el cual se utilizó el identificador reviewId como clave única, conservando una sola versión de cada reseña y evitando

sesgos producidos por registros repetidos. Se crearon variables derivadas como `year_month` (año y mes de publicación), `review_length` (longitud del texto) y `response_delay_days` (diferencia entre la fecha de la reseña y la fecha de respuesta del establecimiento), que facilitaron el análisis temporal y la interpretación de la gestión de los hoteles.

La identificación de reseñas excesivamente cortas resultó particularmente relevante para detectar textos con baja densidad informativa. La literatura ha señalado que los textos muy cortos tienden a reducir la precisión y la estabilidad de los modelos de análisis de sentimientos (McCartney, Hensman & Longo, 2017), lo cual refuerza la importancia de considerar la longitud del texto en la interpretación de los resultados.

En paralelo, se eliminaron variables redundantes, poco informativas o asociadas a posibles riesgos de privacidad, siguiendo criterios de pertinencia metodológica, economía computacional y claridad analítica. La Tabla 2 resume las principales columnas excluidas y el motivo de su eliminación.

Tabla 2.

Columnas excluidas del análisis y justificación

Columna	Motivo de exclusión
address, street, neighborhood, city, state, postalCode, countryCode	Todos los hoteles pertenecen a Buenaventura; no aportan variabilidad útil para el análisis.
fid, cid	Identificadores redundantes; placeId fue suficiente como referencia estable.
temporarilyClosed, permanentlyClosed	Todos los valores se encontraban en “false”; sin utilidad analítica en el contexto del estudio.
imageUrl, reviewImageUrls/..., reviewerPhotoUrl	Información multimedia no utilizada en el análisis textual y con riesgos de privacidad.
reviewerUrl, reviewUrl, url	Campos de navegación irrelevantes para la minería de texto.

Columna	Motivo de exclusión
rating	Duplicaba la información contenida en stars; se mantuvo únicamente esta última.
searchString	Solo almacenaba la cadena de búsqueda utilizada por el scraper.
visitedIn	Campo vacío para todos los registros.
translatedLanguage	Redundante frente a originalLanguage.
categories/...	Baja variabilidad; la mayoría de registros se clasificaban como “Hotel/Lodging”.

Nota. La exclusión se basó en criterios de redundancia, baja variabilidad, privacidad de los usuarios y pertinencia respecto a los objetivos del estudio.

Reseñas Finales Después del ETL

Tras los procesos de depuración, eliminación de duplicados y filtrado temporal, el conjunto de datos analítico quedó conformado por 3.727 reseñas, distribuidas entre los cinco hoteles incluidos en el estudio. Esta reducción respecto al total extraído obedeció a la eliminación de registros duplicados, reseñas fuera del rango temporal definido y casos sin texto útil para el análisis.

Tabla 3.
Reseñas finales por hotel después del ETL

<i>Hotel</i>	<i>Reseñas finales</i>
Torre Mar	575
Cosmos	788
Magüipí	765
Steven	406
Cordillera	1.193
<i>Total</i>	<i>3727</i>

Nota. Estas cifras corresponden al corpus definitivo utilizado para el análisis, una vez aplicados los filtros y transformaciones del proceso ETL.

Carga y trazabilidad

La fase de carga consistió en almacenar los datos transformados dentro del entorno de Azure Blob Storage, organizado en las tres capas del flujo ETL (Bronze, Silver y Gold), y en habilitar su explotación analítica a través de Supabase. La capa Bronze conservó los archivos originales; la capa Silver almacenó los datos limpios y estandarizados; y la capa Gold reunió los datos enriquecidos con información semántica procedente de Azure Cognitive Services – Text Analytics. Con el fin de asegurar la trazabilidad del proceso, se registró información básica de cada ejecución del pipeline, incluyendo el hotel procesado, la cantidad de reseñas importadas, el rango temporal aplicado, la fecha de ejecución y la ubicación del archivo generado en la nube. Esta información se gestionó en tablas auxiliares de control alojadas en Supabase, las cuales funcionaron como índices de verificación y facilitaron la auditoría del flujo de datos.

Tabla 4.

Campos registrados en Supabase para control de versiones del ETL

<i>Campo</i>	<i>Descripción</i>
run_id	Identificador único de cada ejecución del pipeline.
hotel_id	Hotel asociado al proceso de carga.
layer	Capa generada o actualizada (Bronze, Silver o Gold).
rows	Número de registros procesados en la ejecución.
from_date	Fecha inicial del rango temporal procesado.
to_date	Fecha final del rango temporal procesado.
storage_path	Ruta del archivo almacenado en Azure Blob Storage.

Nota. Esta estructura documentó el flujo del ETL, con soporte para control de versiones y reconstrucción histórica del dataset en caso de ser necesario.

Además, se diseñaron vistas SQL en Supabase que permitieron consultar de manera directa los datos contenidos en la capa Gold, sin necesidad de acceder manualmente a los archivos físicos en Blob Storage. Estas vistas fueron fundamentales para el análisis posterior, especialmente para la obtención de métricas temporales, la segmentación por hotel y la construcción de indicadores que se integraron en los dashboards de Power BI.

Figura 13.*Flujo general del proceso ETL implementado (Bronze → Silver → Gold)*

Nota. Adaptado de Medallion Architecture: Data Lakehouse, Data Mesh and Data Quality [Gráfico], por N. Emilio, s. f., en Bismart Blog. <https://blog.bismart.com/en/medallion-architecture>

Esta representación gráfica resume la lógica estructural del pipeline desarrollado en este proyecto, validando que la organización de datos en capas responde a un diseño secuencial, reproducible y orientado al análisis.

Enriquecimiento semántico y validación (Capa Gold)

En la capa Gold se aplicaron las capacidades de Azure Cognitive Services – Text Analytics para generar el enriquecimiento semántico de las reseñas. Para cada texto se obtuvo la clasificación de la polaridad del sentimiento (positiva, neutra o negativa) y se extrajeron aspectos temáticos (aspects) asociados a valoraciones positivas o negativas, tales como limpieza, atención, habitaciones, alimentación o infraestructura. Estos resultados se incorporaron al dataset previamente depurado, dando lugar a una tabla analítica que combinó información numérica, textual y semántica. La elección de Azure Cognitive Services – Text Analytics fue seleccionado por ofrecer compatibilidad nativa con el idioma español, alta capacidad para procesar textos cortos y no estructurados, operación sin requerir entrenamiento supervisado, interpretación transparente basada en probabilidades y una integración directa con pipelines en la nube. Estas características redujeron significativamente los tiempos de implementación y evitaron la necesidad de desarrollar modelos propios desde cero. En consecuencia, el esfuerzo metodológico pudo centrarse en la interpretación de resultados y en la construcción de visualizaciones útiles para la toma de decisiones.

La validación de los resultados se apoyó en dos estrategias complementarias. Por un lado, se revisaron las distribuciones agregadas de sentimientos y aspectos en las vistas SQL y dashboards de Power BI, con el fin de detectar patrones anómalos o inconsistencias evidentes. Por otro lado, se realizó una revisión manual de una muestra de reseñas, seleccionadas de manera intencional entre los distintos hoteles y tipos de polaridad, para contrastar la clasificación automática con la lectura humana. Esta verificación respaldó la coherencia general del modelo en el dominio específico de Buenaventura, además de aportar confianza adicional sobre la utilidad interpretativa de los resultados. El producto final de esta fase consistió en un conjunto de datos limpio, estructurado y semánticamente enriquecido, alojado en la capa Gold y expuesto mediante vistas analíticas en Supabase, que sirvió de base para la elaboración de los paneles de Power BI y para el cumplimiento de los objetivos específicos del proyecto.

7 Resultados

En este capítulo se presentan los principales hallazgos obtenidos a partir del análisis de las reseñas textuales publicadas en Google sobre cinco hoteles ubicados en Buenaventura (Hotel Cordillera, Hotel Cosmos Pacífico, Hotel Magüipí, Hotel Torre Mar y Hotel Steven Buenaventura) durante el periodo 2020–2025.

Los resultados se organizan en seis secciones. En primer lugar, se caracteriza el corpus de reseñas y el proceso de verificación manual del modelo de análisis de sentimientos (Sección 7.1). En segundo lugar, se presentan los indicadores clave de reputación digital por hotel (Sección 7.2). Posteriormente, se analiza la evolución anual de los sentimientos por establecimiento (Sección 7.3). Luego, se examinan las temáticas más frecuentes en las reseñas negativas y positivas (Secciones 7.4 y 7.5). Finalmente, se sintetizan los hallazgos en función de los objetivos específicos del estudio (Sección 7.6), que servirá de base para las conclusiones y recomendaciones de los capítulos siguientes.

De esta manera, el presente capítulo constituye el puente entre la descripción metodológica del proyecto y la discusión de hallazgos en clave gerencial y estratégica para el destino turístico.

7.1 Caracterización General del Corpus y del Modelo de Análisis de Sentimientos

El corpus analizado está conformado por las reseñas textuales publicadas en Google por usuarios que se hospedaron o interactuaron con cinco hoteles ubicados en Buenaventura (Hotel Cordillera, Hotel Cosmos Pacífico, Hotel Magüipí, Hotel Torre Mar y Hotel Steven Buenaventura) durante el periodo 2020–2025. Para efectos del presente estudio se consideraron únicamente las reseñas que contenían texto escrito, descartando las valoraciones numéricas sin descripción, con el fin de garantizar la disponibilidad de información lingüística suficiente para aplicar técnicas de análisis de sentimientos y de extracción de aspectos. En términos generales, el conjunto de datos recoge opiniones emitidas a lo largo de seis años calendario, con una distribución heterogénea entre hoteles y entre periodos. Algunos establecimientos concentran un volumen alto de reseñas, mientras que otros aportan un número menor de observaciones, lo que implica que su peso relativo en el análisis agregado es diferente. Esta variación en el número de reseñas por hotel y por año se tuvo en cuenta al momento de interpretar los resultados.

Con el fin de evitar confusiones entre el total de valoraciones (que incluye también calificaciones numéricas sin texto) y el subconjunto de reseñas efectivamente analizadas mediante

técnicas de minería de texto, se construyó una tabla específica con la distribución de reseñas textuales por hotel y por periodo de análisis. En la Tabla 5 se presenta esta distribución, que constituye la base sobre la cual se calcula la muestra del 3 % utilizada en la verificación manual descrita más adelante.

Tabla 5.

Distribución de reseñas textuales por hotel y periodo de análisis (2020–2025)

Hotel	2020	2021	2022	2023	2024	2025	Total
Hotel Cordillera	81	148	116	84	60	30	519
Hotel Cosmos Pacífico	61	98	82	68	40	17	366
Hotel Magüipí	74	103	99	77	44	21	417
Hotel Steven Buenaventura	16	22	18	15	69	31	171
Hotel Torre Mar	8	71	40	30	69	19	237
Total	240	442	354	274	282	118	1710

Para la clasificación automática de los sentimientos se utilizó Azure Cognitive Services – Text Analytics, configurado en idioma español. El modelo asignó a cada reseña una categoría de polaridad (positiva, negativa o neutra) y una probabilidad asociada; así, las opiniones se agruparon según su carga emocional predominante. Adicionalmente, el servicio identificó frases de opinión asociadas a aspectos específicos (por ejemplo, habitaciones, servicio, restaurante, ubicación), lo cual facilitó la construcción de tablas de frecuencias temáticas por hotel y por año, que se presentan en las secciones posteriores. Con el propósito de verificar la confiabilidad de la clasificación automática, se realizó una verificación manual externa sobre una muestra aleatoria equivalente al 3% del total de reseñas textuales reportado en la Tabla 6 (51 reseñas). Cada reseña de la muestra fue leída y clasificada manualmente como positiva, negativa o neutra, considerando el balance narrativo y el predominio emocional del texto. Los resultados de este ejercicio muestran que en 46 de las 51 reseñas hubo coincidencia clara entre la etiqueta humana y la etiqueta generada por Azure. En 4 reseñas adicionales se identificaron opiniones ambiguas o mixtas: aunque el contenido combinaba elementos positivos y negativos, tanto la evaluación humana como el modelo

coincidieron en la categoría de sentimiento asignada, por lo que se consideran coincidencias en reseñas ambiguas. Finalmente, se registró una única divergencia, en la cual la etiqueta asignada por el modelo no correspondía al sentido predominante identificado en la lectura manual. La Tabla 6 resume estos resultados.

Tabla 6.

Resultados de la verificación manual externa (3%)

Resultado	Cantidad	Porcentaje
Coincidencias entre etiqueta automática y manual	46	90,20%
Divergencias	1	1,96%
Casos ambiguos	4	7,84%
Total	51	100%

Finalmente, con el fin de ilustrar de forma cualitativa el procedimiento de verificación, se construyó una tabla adicional en la que se presentan algunos ejemplos de reseñas en las que se comparan la etiqueta asignada por el modelo y la etiqueta determinada manualmente, incluyendo un caso de plena coincidencia, un caso de opinión mixta con acuerdo en la etiqueta final y el único caso de divergencia identificado.

Tabla 7.

Ejemplos ilustrativos de la verificación manual externa

Reseña completa	Etiqueta Azure	Etiqueta humana	Observación
“Control de tv malo, aire acondicionado feito, le faltaba la manija de la ducha, H502. La ubicación es muy buena.”	Positiva	Negativa	Divergencia real. El modelo privilegia la frase final (“muy buena”), pero el contenido central es negativo y la experiencia predominante es insatisfactoria.

Reseña completa	Etiqueta Azure	Etiqueta humana	Observación
“Excelente ubicación al Puerto de Buenaventura, habitaciones cómodas, único comentario negativo es que las tarjetas de las habitaciones en ocasiones no funcionaban por lo que tocaba bajar hasta la recepción para que las activaran de nuevo.”	Positiva	Positiva Ambigua	Caso ambiguo genuino. La reseña mezcla elogios fuertes con un problema operativo molesto. La valoración global sigue siendo positiva, pero la reseña presenta un balance emocional mixto.
“Excelente servicio y buena atención, mejor hotel de la ciudad.”	Positiva	Positiva	Coincidencia clara sin ambigüedad. Reseña 100% positiva, sin elementos neutrales ni negativos.

A partir de este ejercicio se observa que el modelo de análisis de sentimientos empleado ofrece un nivel de concordancia suficiente para utilizar sus etiquetas como insumo válido en el análisis agregado de la reputación digital de los hoteles estudiados. En las secciones siguientes, dichas etiquetas se integran con indicadores cuantitativos (KPIs) y con la clasificación temática de las reseñas para describir la evolución de la percepción de los usuarios en el periodo 2020–2025.

7.2 Indicadores Clave de Reputación Digital por Hotel

Con el fin de sintetizar la reputación digital de cada establecimiento, se construyó un conjunto de indicadores clave de desempeño (KPIs) a partir de las reseñas analizadas. Estos KPIs permiten

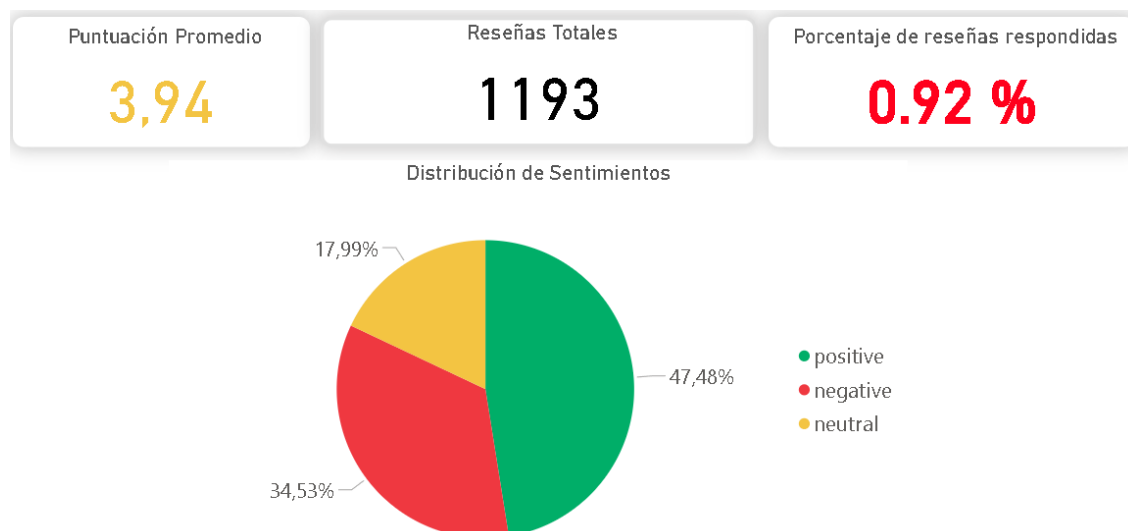
observar, de manera integrada, el volumen de opiniones disponibles, la calificación promedio reportada en Google, el balance emocional de las reseñas y el nivel de interacción de los hoteles con los comentarios de sus huéspedes. Para cada hotel se diseñó un panel de indicadores en Power BI que incluye, al menos, los siguientes elementos:

- Puntuación promedio del hotel en Google
- Número total de reseñas textuales y no textuales utilizadas en el análisis
- Porcentaje de reseñas respondidas por el establecimiento, como aproximación a la gestión activa de la reputación en línea.
- Distribución porcentual de sentimientos (positivos, negativos y neutros) representada mediante un gráfico circular.

En las Figuras 14–18 se presentan los paneles de KPIs para los cinco hoteles analizados. Cada panel resume, en un solo vistazo, el volumen de información disponible y el tono general de las opiniones emitidas por los usuarios, así como el grado de respuesta de los administradores a dichas reseñas.

Figura 14.

Panel de indicadores de reputación digital – Hotel Cordillera

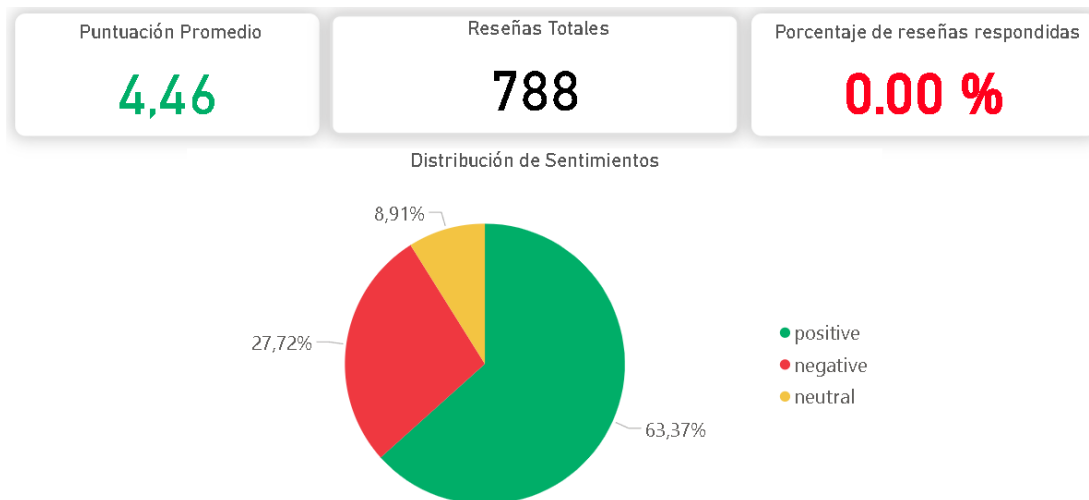


En el caso del Hotel Cordillera, el panel de indicadores muestra un volumen elevado de reseñas en el periodo 2020–2025, con un total de 1.193 reseñas analizadas y una puntuación promedio de 3,94. La distribución de sentimientos evidencia un predominio de reseñas positivas (47,48 %), aunque también se registra una proporción apreciable de opiniones negativas (34,53 %)

y neutras (17,99 %). El indicador de reseñas respondidas corresponde al 0,92 %, lo que muestra que la mayoría de los comentarios permaneció sin respuesta dentro del periodo analizado.

Figura 15.

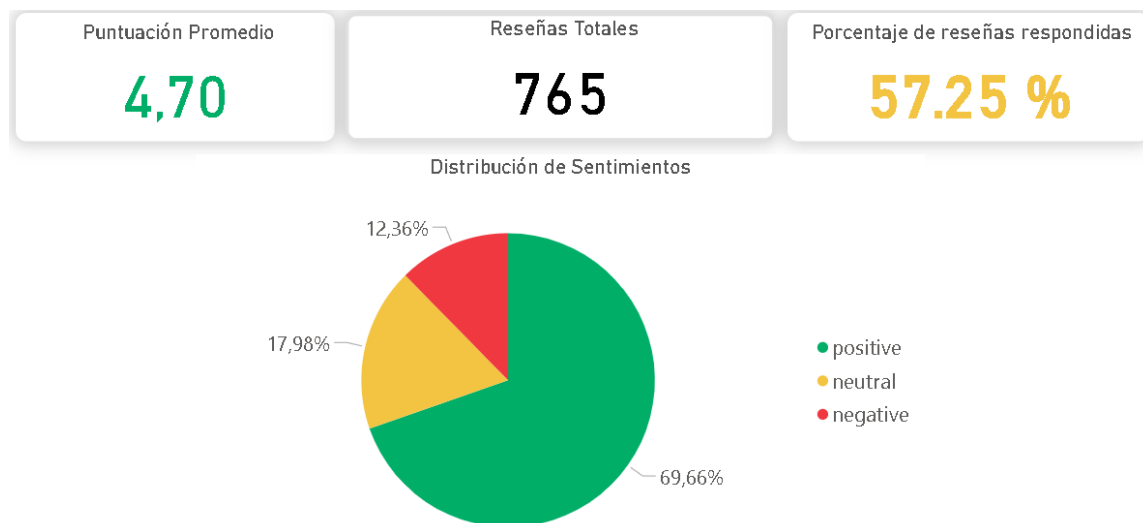
Panel de indicadores de reputación digital – Cosmos Pacífico



El Hotel Cosmos Pacífico presenta un total de 788 reseñas analizadas y una puntuación promedio de 4,46. En la distribución de sentimientos se observa un predominio de reseñas positivas (63,37 %), acompañado de opiniones negativas (27,72 %) y neutras (8,91 %). El porcentaje de reseñas respondidas corresponde al 0 %, lo que evidencia ausencia de interacción visible del establecimiento frente al conjunto de comentarios analizados.

Figura 16.

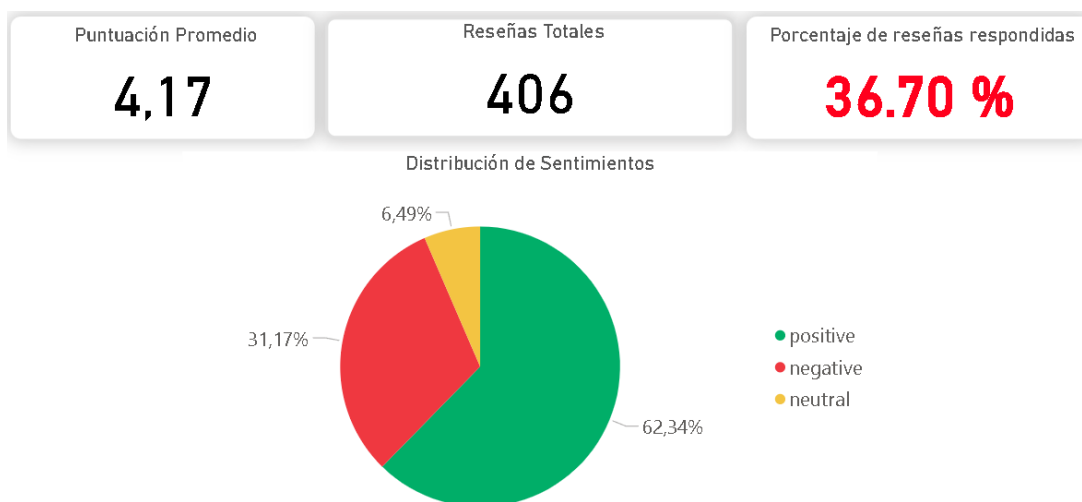
Panel de indicadores de reputación digital – Hotel Magüipí



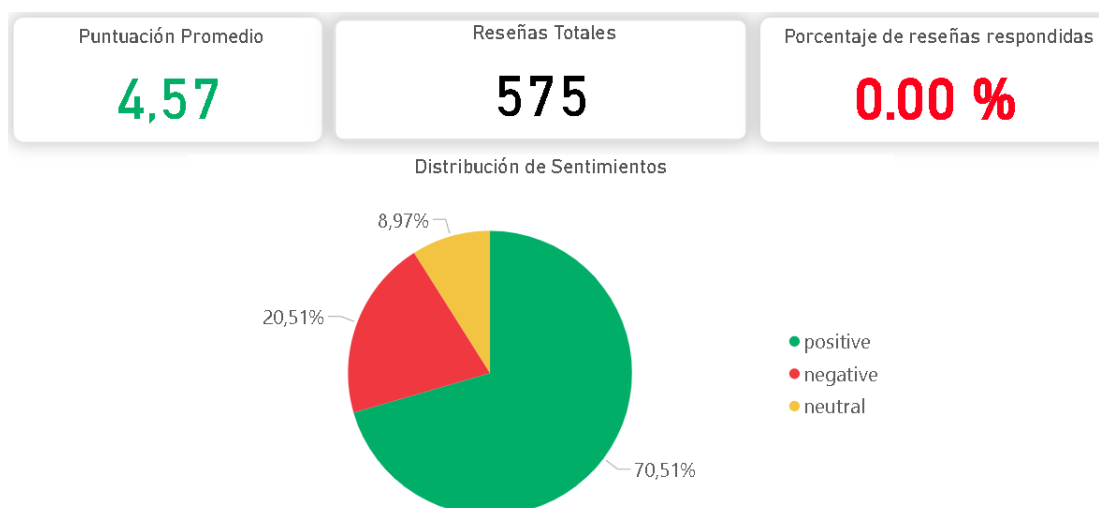
En el Hotel Magüipí, el panel de KPIs evidencia una puntuación promedio de 4,70 sobre un total de 765 reseñas analizadas. La distribución de sentimientos muestra una proporción elevada de reseñas positivas (69,66 %), frente a porcentajes menores de opiniones neutras (17,98 %) y negativas (12,36 %). Además, el porcentaje de reseñas respondidas alcanza el 57,25 %, valor comparativamente superior al de los demás establecimientos del estudio.

Figura 17.

Panel de indicadores de reputación digital – Hotel Steven Buenaventura



El Hotel Steven Buenaventura presenta un volumen menor de reseñas en comparación con los demás hoteles, con 406 reseñas analizadas y una puntuación promedio de 4,17. La distribución de sentimientos se inclina hacia opiniones positivas (62,34 %), seguida por reseñas negativas (31,17 %) y neutras (6,49 %). El porcentaje de reseñas respondidas corresponde al 6,70 %, lo que refleja una baja interacción visible entre el establecimiento y los usuarios.

Figura 18.*Panel de indicadores de reputación digital – Hotel Torre Mar*

Por su parte, el Hotel Torre Mar cuenta con 575 reseñas analizadas y una puntuación promedio de 4,57. La distribución global de sentimientos muestra una proporción mayoritaria de reseñas positivas (70,51 %), junto con opiniones negativas (20,51 %) y neutras (8,97 %). En este establecimiento, el indicador de reseñas respondidas se mantiene en 0 %, lo que significa que, para el periodo analizado, no se registraron respuestas del hotel a los comentarios publicados.

A partir de estos paneles se observan diferencias relevantes entre los hoteles. En términos de volumen, el Hotel Cordillera concentra el mayor número de reseñas analizadas (1.193), mientras que el Hotel Steven Buenaventura registra el menor volumen (406). En cuanto a la puntuación promedio, los valores oscilan entre 3,94 en Hotel Cordillera y 4,70 en Hotel Magüipí. La distribución de sentimientos también evidencia contrastes entre establecimientos: Hotel Torre Mar presenta la mayor proporción de reseñas positivas (70,51 %), mientras que Hotel Cordillera registra la mayor proporción de reseñas negativas (34,53 %). Finalmente, el porcentaje de reseñas respondidas varía entre 0 % en Cosmos Pacífico y Torre Mar, y 57,25 % en Hotel Magüipí, lo que muestra diferencias importantes en la gestión activa de la reputación digital. Estas diferencias generales sirven como punto de partida para el análisis temporal y temático desarrollado en las secciones siguientes.

Estos resultados se articulan con la literatura sobre reseñas turísticas en línea y reputación digital hotelera, en la medida en que las opiniones publicadas por los usuarios no solo reflejan experiencias individuales, sino que también inciden en la imagen percibida de los establecimientos, así como en los procesos de decisión de otros viajeros. Gretzel y Yoo (2008)

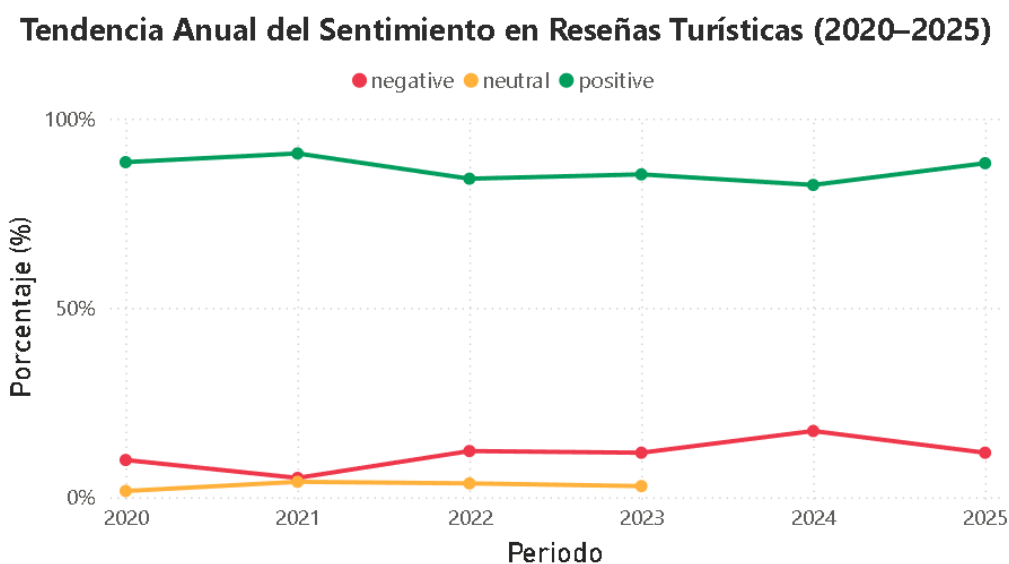
señalan que las reseñas en línea cumplen un papel relevante en la planificación de viajes; por su parte, Filieri et al. (2015) destacan que este tipo de opiniones integra valoraciones cognitivas y afectivas sobre atributos concretos del servicio. En este sentido, indicadores como la calificación promedio, la distribución de sentimientos, además del porcentaje de reseñas respondidas, permiten aproximarse a la reputación digital de cada hotel desde una perspectiva cuantitativa e interpretativa. Asimismo, Anderson y Han (2016), junto con TripAdvisor e Ipsos MORI (2019), respaldan la importancia de la respuesta del establecimiento como práctica asociada a la gestión activa de la reputación en línea.

7.3 Evolución Anual de los Sentimientos por Hotel

Además de los indicadores globales de reputación presentados en la sección anterior, se analizó la evolución temporal de los sentimientos expresados en las reseñas textuales de cada hotel. Para ello se calcularon, por año y por establecimiento, los porcentajes de reseñas negativas, neutras y positivas respecto del total de comentarios textuales publicados entre 2020 y 2025. Estos resultados se representaron mediante gráficos de líneas, donde cada serie corresponde a una categoría de sentimiento (positivo, negativo, neutro).

Figura 19.

Evolución anual de los sentimientos en las reseñas del Hotel Cosmos Pacífico

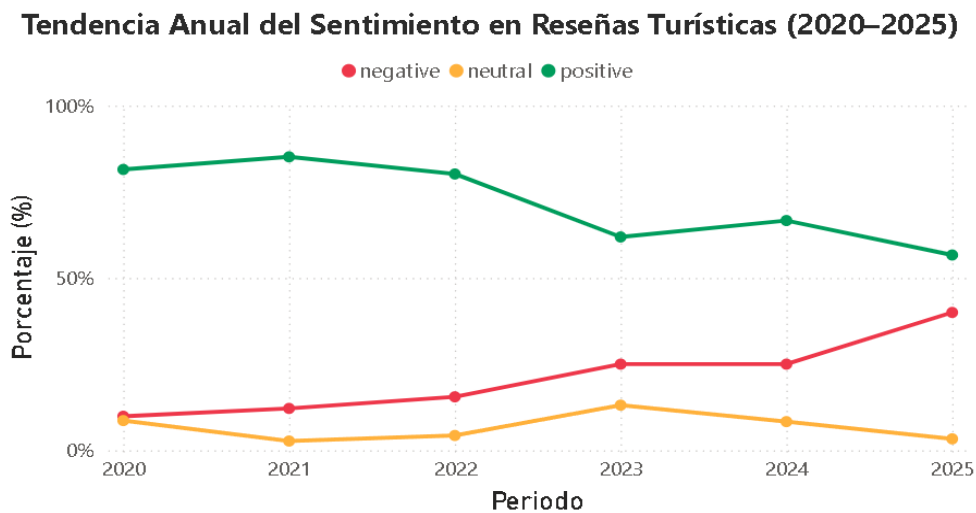


En el caso del Hotel Cosmos Pacífico, la evolución anual muestra un predominio consistente de reseñas positivas a lo largo de todo el periodo. Como se observa en la Figura 19,

los porcentajes de opiniones negativas se mantienen en niveles relativamente bajos, con variaciones moderadas entre años: en algunos periodos se alcanza uno de los valores más altos de reseñas positivas y uno de los porcentajes más bajos de reseñas negativas, mientras que en otros (por ejemplo, 2024) se registra un aumento puntual de las opiniones negativas que no altera de forma drástica el balance general. En términos agregados, Cosmos Pacífico presenta una mayoría clara de reseñas positivas y una fracción reducida de reseñas negativas sobre el total de reseñas textuales analizadas.

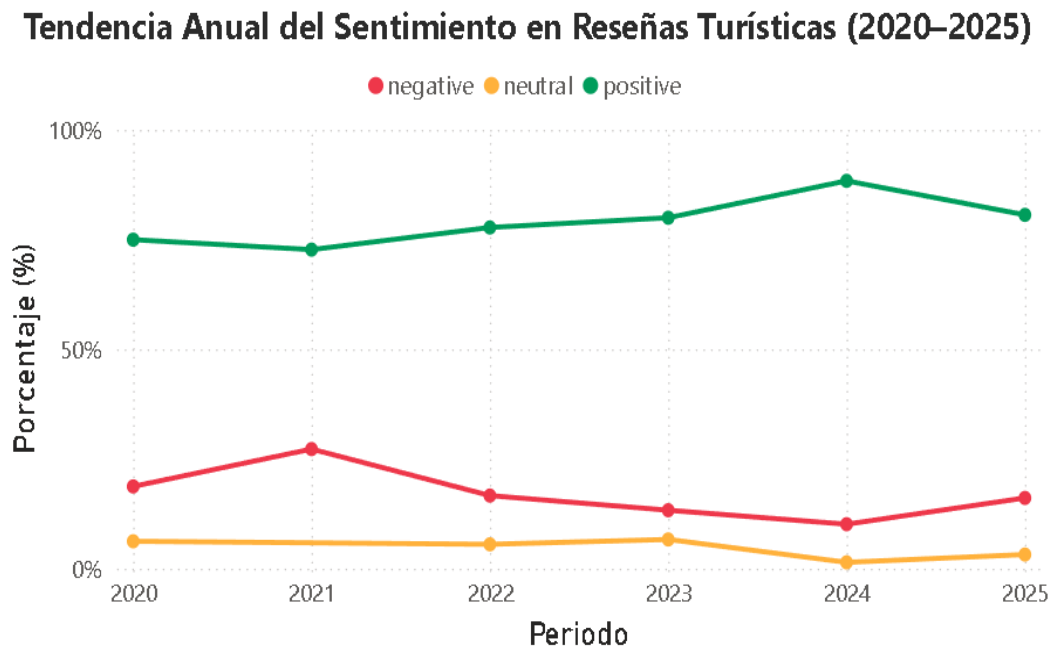
Figura 20.

Evolución anual de los sentimientos en las reseñas del Hotel Cordillera



En el Hotel Cordillera, la serie temporal de sentimientos presenta un comportamiento distinto al de Cosmos. Como se aprecia en la Figura 20, durante los primeros años del periodo (2020–2022) predominan claramente las reseñas positivas, con porcentajes elevados y un nivel relativamente bajo de reseñas negativas. Sin embargo, a partir de 2023 se observa un cambio de tendencia: la proporción de reseñas negativas aumenta de manera sostenida, mientras que el porcentaje de reseñas positivas disminuye. Esta dinámica se acentúa en 2024 y 2025, años en los que las reseñas negativas alcanzan el 25 % y el 40 %, respectivamente, mientras que las reseñas positivas descienden al 66,67 % y al 56,67 %. Paralelamente, el número total de reseñas anuales tiende a reducirse en los años en que se concentra una mayor proporción de opiniones negativas, lo que refuerza la idea de un deterioro en la percepción del hotel durante el tramo final del periodo analizado.

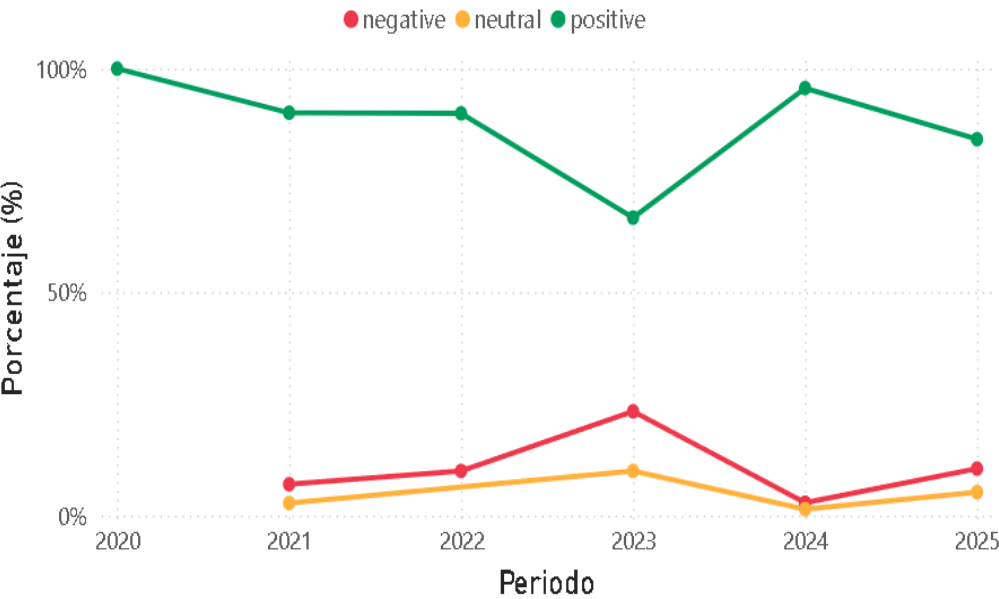
Figura 21.
Evolución anual de los sentimientos en las reseñas del Hotel Steven Buenaventura



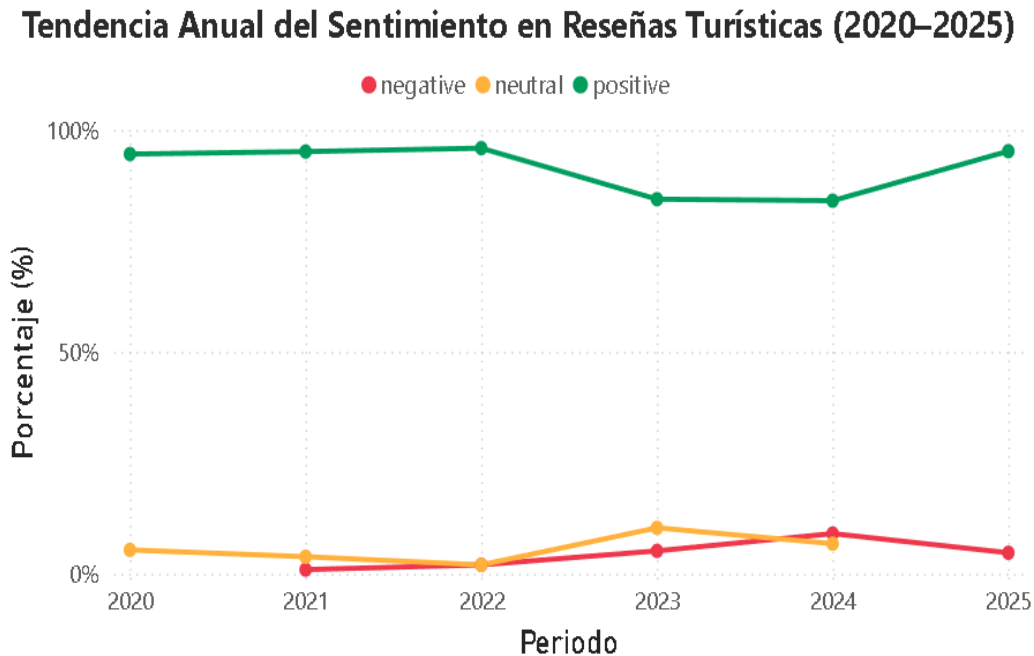
En el Hotel Steven Buenaventura, las tendencias de sentimiento a lo largo del tiempo muestran un patrón de mejora progresiva. Como se observa en la Figura 21, en los primeros años (2020–2022) los porcentajes de reseñas negativas son comparativamente más altos y la proporción de reseñas positivas es más moderada. No obstante, a partir de 2023 se aprecia una disminución gradual en el porcentaje de opiniones negativas y un incremento en la proporción de reseñas positivas. En 2024 se alcanza uno de los valores más altos de positividad del periodo, acompañado además por un aumento en el número de reseñas, lo que otorga mayor robustez a esta tendencia. En conjunto, la serie sugiere un cambio en la percepción de los huéspedes hacia una valoración más favorable en los años recientes.

Figura 22.
Evolución anual de los sentimientos en las reseñas del Hotel Torre mar

Tendencia Anual del Sentimiento en Reseñas Turísticas (2020–2025)



En el Hotel Torre Mar, la evolución temporal de los sentimientos refleja un comportamiento más irregular. Tal como se muestra en la Figura 22, entre 2020 y 2022 los porcentajes de reseñas positivas se mantienen en niveles relativamente altos, mientras que las reseñas negativas presentan valores moderados. Sin embargo, en 2023 se registra un incremento marcado en la proporción de reseñas negativas, que alcanzan el 23,33 %, acompañado de una reducción de las reseñas positivas hasta el 66,67 %. Este comportamiento convierte a 2023 en un punto crítico dentro de la serie, especialmente si se compara con 2024, año en el que las opiniones positivas aumentan al 95,65 % y las negativas disminuyen al 2,90 %. Posteriormente, en 2025 la serie tiende hacia valores más cercanos al promedio general del periodo.

Figura 23.*Evolución anual de los sentimientos en las reseñas del Hotel Magüipi*

En el Hotel Magüipí, las series temporales de sentimientos presentan una trayectoria especialmente favorable. Como se observa en la Figura 23, durante todo el periodo 2020–2025 las reseñas positivas constituyen un porcentaje muy elevado del total, mientras que las reseñas negativas se mantienen en niveles bajos. Esta estabilidad se evidencia, por ejemplo, en 2022, año en el que las reseñas positivas alcanzan el 95,92 % y las negativas apenas el 2,04 %. Incluso en 2024, cuando se observa una ligera reducción de la positividad, las opiniones favorables se mantienen en 84,09 %, frente a un 9,09 % de reseñas negativas. En términos generales, la evolución temporal confirma que Magüipí conserva una reputación predominantemente positiva a lo largo de los años analizados.

Esta variación temporal es coherente con la literatura sobre análisis de sentimientos en turismo, la cual señala que las reseñas en línea permiten observar cambios en la percepción de los visitantes y detectar patrones asociados a la experiencia turística en distintos periodos. Manosso y Domareski (2021) destacan que el análisis de sentimientos se ha consolidado como una herramienta útil para estudiar percepciones turísticas a partir de opiniones digitales, mientras que Oliveira et al. (2022) evidencian que las reseñas hoteleras pueden utilizarse para segmentar comportamientos y reconocer diferencias en la valoración de los establecimientos. En este sentido, las tendencias anuales identificadas en los hoteles analizados no deben entenderse como datos

aislados, sino como señales de evolución reputacional que permiten reconocer estabilidad, deterioro o recuperación en la percepción de los huéspedes.

7.4 Resultados Temáticos de las Reseñas Negativas

Además del balance global de sentimientos, se realizó un análisis temático de las reseñas clasificadas como negativas, con el fin de identificar qué aspectos específicos de la experiencia turística generan mayor insatisfacción en los huéspedes. Para ello, las frases de opinión extraídas por Azure Cognitive Services – Text Analytics fueron agrupadas y reclasificadas en categorías semánticamente coherentes (por ejemplo, habitaciones, servicio, comida/restaurante, instalaciones, ubicación, entorno, higiene, entre otras), construyendo tablas de frecuencia por hotel y por año.

En las tablas de esta sección se presentan, para cada establecimiento, las temáticas negativas más mencionadas durante el periodo 2020–2025, así como su distribución temporal. Estas tablas permiten observar tanto los focos de inconformidad más recurrentes como la aparición de problemas puntuales en determinados años. A continuación, se describen de manera sintética los patrones más relevantes identificados en las reseñas negativas de cada hotel.

En el caso del Hotel Cosmos Pacífico, las temáticas negativas se organizan en torno a aspectos de servicio, gastronomía e infraestructura. En la Tabla 8 se presentan las frecuencias anuales de las principales categorías negativas asociadas a este establecimiento.

Tabla 8.
Temáticas negativas — Hotel Cosmos Pacífico

Temática	2020	2021	2022	2023	2024	2025	Total
Servicio	3	2	4	0	1	0	10
Trato / atención del personal	1	0	2	2	0	1	6
Recepción	0	0	1	0	1	1	3
Mantenimiento	2	0	4	0	1	0	7
Restaurante / alimentos	3	4	3	1	2	0	13
Publicidad engañosa	1	0	0	0	0	0	1

Temática	2020	2021	2022	2023	2024	2025	Total
Parqueadero	0	0	1	1	2	0	4
Aire acondicionado	1	0	2	0	0	0	3
Desayuno	2	1	2	2	1	0	8
Calidad de alimentos	0	2	0	0	0	0	2
Personal del hotel	0	1	1	1	0	0	3
Sábanas / lencería	0	1	1	0	0	1	3
Baños	0	0	1	1	1	0	3
Habitaciones	0	1	3	2	3	1	10
Áreas comunes	0	0	2	0	0	0	2
Piscina	0	0	1	0	1	1	3
Instalaciones / infraestructura	0	0	3	1	3	1	8
Precios / precio	0	0	1	2	1	0	4
Aseo / limpieza	0	0	1	1	0	1	3
Wifi / internet	0	0	1	1	0	0	2
Dispositivos (teléfono/tarjeta)	0	0	0	1	0	1	2
Camas / confort	0	0	1	0	0	1	2
Manejo de vajilla	0	2	0	0	0	0	2
Consideración con huéspedes	0	0	0	1	0	0	1
Organización	0	0	0	0	1	0	1
Experiencia negativa	0	0	0	1	0	0	1
Lugar / entorno	0	0	0	1	0	0	1

Temática	2020	2021	2022	2023	2024	2025	Total
Total por año	11	11	32	22	20	12	108

A partir de esta tabla se observa que una parte importante de las reseñas negativas hace referencia al servicio y la atención del personal, incluyendo comentarios sobre la rapidez en la respuesta a solicitudes, la gestión de situaciones específicas y, en algunos casos, la percepción del trato recibido. De igual forma, aparecen de manera recurrente inconformidades relacionadas con la comida y el restaurante, ya sea por la percepción de que la oferta gastronómica no cumple las expectativas en sabor, variedad o relación calidad–precio. En menor medida, pero de forma constante, se registran menciones a problemas de ruido en determinadas zonas del hotel y a detalles de aseo o mantenimiento en habitaciones y áreas comunes, que los huéspedes destacan como elementos que afectan la comodidad de la estadía.

En el Hotel Cordillera, las reseñas negativas se concentran de forma muy marcada en el estado físico y sanitario de las instalaciones. En la Tabla 9 se presenta la distribución de las temáticas negativas reportadas para este hotel durante el periodo de análisis.

Tabla 9.
Temáticas negativas — Hotel Cordillera

Temática	2020	2021	2022	2023	2024	2025	Total
Habitaciones	4	8	7	9	3	2	33
Hotel	2	1	4	3	7	3	20
Atención	1	2	5	4	1	1	14
Aseo	1	1	0	4	3	1	10
Servicio	2	0	1	3	2	1	9
Personal / empleados	1	2	2	3	1	0	9
Baños	0	0	2	3	1	0	6
Lugar / sitio	0	1	0	3	2	0	6

Temática	2020	2021	2022	2023	2024	2025	Total
Camas / colchones	0	1	1	1	1	3	7
Instalaciones	0	0	0	3	1	0	4
Experiencia	0	1	1	1	1	1	5
Teléfono / teléfonos	0	2	1	1	0	0	4
Pintura	0	1	1	1	0	0	3
Desayuno	0	0	0	2	0	1	3
Olores / olor	1	1	1	0	0	0	3
Aire acondicionado	0	1	0	0	1	1	3
Paredes	0	0	0	0	0	1	1
Sábanas	0	0	0	0	0	1	1
TV	0	0	0	0	0	1	1
Citófonos	0	0	1	0	0	0	1
Facturación electrónica	0	0	0	1	0	0	1
Total por año	12	23	32	46	24	38	175

La información de la tabla muestra que la categoría de habitaciones es la fuente principal de inconformidad, con comentarios que mencionan incomodidad en las camas, presencia de humedad, problemas de ventilación, olores desagradables, plagas y fallas en el funcionamiento de elementos básicos como duchas, llaves, aire acondicionado o cerraduras. Estas observaciones se repiten a lo largo de varios años, lo que indica que se trata de un foco recurrente de insatisfacción. A ello se suman reseñas que señalan deficiencias en el aseo y la limpieza tanto de las habitaciones como de las zonas comunes, así como críticas dirigidas al servicio y la atención en recepción u otros puntos de contacto. Algunas opiniones negativas mencionan además ruido proveniente del entorno o de otras áreas del hotel y episodios puntuales relacionados con la seguridad percibida.

En conjunto, las temáticas negativas de Cordillera se orientan principalmente al estado de las instalaciones y las condiciones de higiene.

En el Hotel Steven Buenaventura, el número total de reseñas negativas es menor que en otros establecimientos del estudio, pero también se identifican patrones específicos. La Tabla 10 presenta las temáticas negativas más frecuentes y su distribución por año para este hotel.

Tabla 10.
Temáticas negativas — Steven Buenaventura

Temática	2020	2021	2022	2023	2024	2025	Total
Habitaciones	2	2	0	1	3	1	9
Camas / Colchones	0	1	0	1	0	0	2
Baño / Plomería	0	1	0	1	0	0	2
Aseo / Limpieza	1	0	0	1	0	0	2
Infraestructura / Mant.	1	0	2	1	0	0	4
Infraestructura general	0	1	1	0	1	0	3
Servicio / Atención	3	3	2	1	5	2	16
Internet / Conectividad	1	0	0	0	1	0	2
Comunicación interna	1	0	0	0	0	0	1
Ubicación / Entorno	0	0	1	0	2	0	3
Total Por Año	8	9	6	5	9	8	45

A partir de la tabla se observa que las críticas se concentran en aspectos de confort y equipamiento de las habitaciones, tales como el tamaño percibido, la comodidad del mobiliario o el estado de algunos elementos del baño. También se registran menciones a ruido y a la ubicación, especialmente cuando el entorno inmediato del hotel se percibe como poco favorable para el descanso debido al tráfico o a la actividad nocturna. En menor medida se encuentran reseñas que cuestionan aspectos de servicio y coordinación, como tiempos de respuesta o gestión de reservas,

aunque estas categorías no alcanzan la misma frecuencia que en otros hoteles. En este caso, las inconformidades se concentran más en detalles puntuales de confort y entorno que en fallas generalizadas en la atención.

En el Hotel Torre Mar, las reseñas negativas presentan un patrón más diverso, en el que confluyen aspectos de servicio, instalaciones y logística. En la Tabla 11 se resumen las principales temáticas negativas identificadas para este establecimiento.

Tabla 11.
Temáticas negativas — Hotel Torre Mar

Temática	2021	2022	2023	2024	2025	Total
Alimentos / Bebidas	3	4	3	0	1	11
Equipos / Tecnología	0	0	3	2	0	5
Habitación / Confort	0	0	1	2	0	3
Parqueadero / Acceso	1	0	0	0	0	1
Precio	1	0	0	0	0	1
Restaurante - Mesa	1	1	2	0	0	4
Servicio / Atención	2	5	2	0	0	9
Otros	0	1	0	0	0	1
Total por año	7	10	10	4	1	32

La tabla evidencia que las quejas sobre servicio y atención ocupan un lugar relevante, con comentarios que mencionan demoras en la respuesta a solicitudes, dificultades en la coordinación del check-in y el manejo de reservas, o percepciones de poca eficiencia en momentos de alta demanda. A ello se suman reseñas que señalan problemas de mantenimiento y condiciones de las instalaciones, incluyendo desgaste visible, fallas en el funcionamiento de ciertos equipos y condiciones de aseo mejorables en algunas habitaciones o zonas comunes. En determinados años se observa un aumento en la frecuencia de reseñas negativas, asociado a la acumulación de

comentarios críticos en torno a estas categorías, lo que se articula con los patrones de sentimiento explicados en otros apartados del capítulo.

Finalmente, en el Hotel Magüipí, el volumen de reseñas negativas es relativamente reducido en comparación con los demás hoteles, pero también se identifican algunas temáticas recurrentes. En la Tabla 12 se presentan las categorías negativas registradas para este establecimiento y su distribución temporal.

Tabla 12.
Temáticas negativas — Hotel Magüipí

Temática	2020	2021	2022	2023	2024	2025	Total
Baño / Plagas	0	0	0	0	0	1	1
Camas / Colchones	1	0	0	0	0	0	1
Comida / Restaurante	0	1	4	2	1	1	9
Cultura / Playas / Entorno	0	0	1	2	1	0	4
Habitaciones	2	0	0	0	1	0	3
Instalaciones / Mantenimiento	0	0	0	0	1	0	1
Internet / Señal	0	3	0	0	1	0	4
Música / Ambiente	0	2	0	0	0	0	2
Planes / Actividades	0	0	0	0	0	1	1
Servicio / Atención	0	6	1	3	4	0	14
Total por año	3	12	6	7	9	3	40

La lectura de la tabla muestra que una parte de las opiniones negativas se relaciona con servicio y atención, especialmente en situaciones puntuales en las que los huéspedes perciben demoras, falta de coordinación o dificultades para resolver solicitudes. Otra fracción de las reseñas hace referencia a comida y restaurante, mencionando expectativas no cumplidas respecto a la variedad, la disponibilidad de ciertos platos o la organización en determinados horarios. También

aparecen menciones aisladas a aspectos de infraestructura y mantenimiento puntual (como el funcionamiento de neveras o las condiciones de los baños), a la señal de internet y a cuestiones de cultura o limpieza del entorno de playas, donde se alude tanto al hotel como al comportamiento de otros visitantes. En conjunto, las reseñas negativas de Magüipí se distribuyen entre un conjunto acotado de temas y no alcanzan, en frecuencia, los niveles observados en otros hoteles del estudio.

Como balance de las reseñas negativas, los focos de inconformidad varían entre establecimientos: en algunos casos se concentran en habitaciones y limpieza; en otros, en servicio, atención, experiencia gastronómica, logística o entorno. Estos patrones serán contrastados con las temáticas positivas para obtener una lectura más equilibrada de las fortalezas y debilidades percibidas en cada hotel.

Estos hallazgos son consistentes con estudios previos sobre experiencia hotelera y análisis de reseñas, en los cuales la satisfacción o insatisfacción del huésped suele asociarse con atributos concretos del servicio, tales como infraestructura, limpieza, atención, comodidad, ubicación, además de la relación calidad–precio. En esta línea, Araya-Pizarro y Verelst (2024) identificaron que variables como infraestructura, ubicación y relación precio–calidad inciden en la experiencia hotelera; por su parte, Vargas-Calderón et al. (2021) evidenciaron la utilidad de las reseñas de clientes para evaluar la calidad del servicio en el sector hotelero. Desde esta perspectiva, las temáticas negativas identificadas en los hoteles de Buenaventura no representan únicamente opiniones aisladas, sino señales operativas sobre dimensiones específicas de la experiencia que pueden afectar la reputación digital, así como la percepción turística del destino.

7.5 Resultados Temáticos de las Reseñas Positivas

De manera complementaria al análisis de las reseñas negativas, se realizó una clasificación temática de las opiniones etiquetadas como positivas, con el propósito de identificar qué aspectos de la experiencia turística son valorados de forma recurrente por los huéspedes. Al igual que en el caso anterior, las frases de opinión extraídas por Azure Cognitive Services – Text Analytics fueron agrupadas y reclasificadas en categorías semánticas como atención y servicio, habitaciones, comida y restaurante, instalaciones, ubicación, entorno, vistas, precio, entre otras, construyendo tablas de frecuencia por hotel y por año. Las tablas de esta sección presentan, para cada establecimiento, las temáticas positivas más mencionadas durante el periodo 2020–2025 y su distribución temporal. Estas tablas permiten reconocer tanto las fortalezas percibidas por los

usuarios como aquellos elementos que se convierten en atributos diferenciadores de cada hotel. A continuación, se describen los patrones más relevantes observados en las reseñas positivas.

En el caso del Hotel Cosmos Pacífico, las temáticas positivas se organizan principalmente en torno a la calidad del servicio, la percepción general del hotel y la ubicación. En la Tabla 13 se presentan las frecuencias anuales de las principales categorías positivas asociadas a este establecimiento.

Tabla 13.
Temáticas positivas — Hotel Cosmos Pacífico

Tema	2020	2021	2022	2023	2024	2025	Total
Atención y trato del personal	22	23	24	26	22	15	132
Percepción general del hotel	20	14	14	12	13	4	77
Ubicación y entorno	9	10	9	7	8	6	49
Vistas y paisaje	8	7	7	7	9	6	44
Habitaciones y confort	6	15	19	4	9	0	53
Restaurante y calidad de la comida	8	13	6	8	3	2	40
Desayuno	4	9	7	7	5	4	36
Piscina y zonas húmedas	6	5	4	7	4	2	28
Servicio general del hotel	3	3	4	3	3	7	23
Instalaciones generales	1	3	6	2	6	3	21
Experiencia global de la estadía	2	2	3	3	2	3	15
Ambiente y atmósfera	0	1	4	3	3	0	11
Precio y relación calidad-precio	0	1	1	2	2	3	9
Conectividad e internet	0	0	2	2	1	1	6
Seguridad	0	1	3	1	1	0	6

Tema	2020	2021	2022	2023	2024	2025	Total
Perfil huésped (familias/vacaciones)	1	0	2	1	0	0	4
Accesibilidad	0	1	1	0	1	0	3
Servicios adicionales (spa/eco)	0	0	1	0	1	0	2
Otros aspectos positivos	1	2	3	1	0	1	8
Total por año	78	91	111	96	94	63	533

Los datos muestran que la atención y el trato del personal constituyen una de las categorías con mayor número de menciones, lo que indica que los huéspedes valoran de forma reiterada la amabilidad, disposición y profesionalismo del equipo de trabajo. También se destaca la percepción general del hotel, con comentarios que lo describen como un lugar agradable, cómodo y recomendable. De igual forma, la ubicación y el entorno aparecen como fortalezas, resaltando la cercanía a puntos estratégicos de la ciudad y las vistas desde determinadas áreas. En un segundo nivel de frecuencia se encuentran aspectos como las habitaciones y el confort, las instalaciones y el paisaje, que complementan la imagen positiva del establecimiento a lo largo del periodo analizado.

En el Hotel Cordillera, las reseñas positivas muestran un perfil claramente orientado al servicio, la ubicación y la relación calidad–precio. En la Tabla 14 se presenta la distribución de las temáticas positivas para este hotel, luego de un proceso de reclasificación que agrupó las opiniones en categorías específicas y evitó el uso de etiquetas ambiguas como “otros”.

Tabla 14.
Temáticas positivas — Hotel Cordillera

Tema	2020	2021	2022	2023	2024	2025	Total
Aseo / Limpieza	2	1	0	1	0	0	4
Atención / Servicio	29	64	48	23	24	7	195
Comida / Restaurante	4	10	3	10	3	0	30

Experiencia	6	8	5	7	5	3	34
Habitaciones	7	6	14	9	7	3	46
Instalaciones	8	9	7	4	5	2	35
Precio	6	13	7	6	6	4	42
Tecnología	0	2	2	0	3	0	7
Ubicación	18	24	21	20	13	5	101
Vista / Entorno	7	18	14	10	5	5	59
Total	87	162	121	90	69	29	558

Las frecuencias señalan que la categoría de atención y servicio concentra el mayor número de menciones positivas, con huéspedes que resaltan de forma consistente la amabilidad del personal, el buen trato y la disposición para resolver solicitudes. La ubicación constituye otra fortaleza relevante, asociada a la cercanía al muelle turístico y a otros puntos de interés de Buenaventura. Asimismo, la temática de habitaciones aparece con una frecuencia significativa, destacando comentarios sobre comodidad, tamaño adecuado y condiciones favorables para el descanso, especialmente en ciertos periodos. Otras categorías relevantes son precio, donde se resalta la percepción de tarifas cómodas o justas, y vista/entorno, con reseñas que valoran positivamente el paisaje observado desde el hotel. En conjunto, las reseñas positivas de Cordillera se concentran en la combinación de buen servicio, ubicación estratégica y una relación costo–beneficio percibido como favorable.

En el Hotel Steven Buenaventura, el volumen total de reseñas positivas es menor que en otros hoteles, pero la tabla de temáticas permite identificar con claridad qué aspectos son más valorados por los huéspedes. En la Tabla 15 se presentan las categorías positivas y su distribución por año para este establecimiento.

Tabla 15.*Temáticas positivas — Steven Buenaventura*

Temática	2020	2021	2022	2023	2024	2025	Total
Aseo / Limpieza	1	0	0	0	1	2	4
Baño	0	1	0	0	0	0	1
Camas / Colchones	0	1	0	0	0	0	1
Comida / Desayuno	1	0	1	0	2	0	4
Equipos / Internet	0	0	0	0	0	2	2
Experiencia general	0	0	1	0	1	0	2
Habitaciones	2	2	1	3	11	4	23
Instalaciones / Confort	3	5	4	2	18	9	41
Precio / Relación valor	2	0	4	2	3	5	16
Servicio / Atención	4	5	8	7	46	12	82
Ubicación / Entorno	3	6	5	3	17	8	42
Total por año	16	20	24	17	99	42	218

En este caso, las menciones se concentran en atención y servicio, donde los usuarios resaltan el buen trato, la cordialidad y la cercanía del personal. También aparecen referencias a la experiencia general de la estadía, con comentarios que describen el hotel como un lugar agradable para pasar la noche o para estancias cortas. En menor medida se registran menciones positivas sobre habitaciones y limpieza, así como valoraciones favorables de la ubicación cuando esta facilita el acceso a determinados puntos de la ciudad. Aunque las frecuencias son más bajas que en otros establecimientos, las reseñas positivas construyen una imagen de hotel funcional y acogedor, especialmente apreciado por su trato humano.

En el Hotel Torre Mar, las reseñas positivas están fuertemente asociadas a la experiencia en el restaurante, la calidad del servicio y la ubicación. En la Tabla 16 se resumen las temáticas positivas con mayor frecuencia en este hotel durante el periodo 2020–2025.

Tabla 16.*Temáticas positivas — Hotel Torre Mar*

Temática	2020	2021	2022	2023	2024	2025	Total
Comida / Restaurante	8	77	19	30	53	4	191
Experiencia general / Estadía	1	3	2	1	5	1	13
Habitaciones	0	6	7	1	0	0	14
Instalaciones / Ambiente	1	7	3	3	4	3	21
Precio / Valor	0	2	2	0	0	0	4
Servicio / Atención	3	21	16	4	22	1	67
Ubicación / Entorno	0	12	12	3	16	2	45
Vista / Paisaje	0	1	7	1	6	1	16
Total por año	13	129	68	43	106	12	371

La tabla permite apreciar que la categoría de comida y restaurante concentra el mayor número de menciones positivas, convirtiéndose en una de las principales fortalezas del establecimiento. Los huéspedes destacan tanto la calidad de los platos como la experiencia de consumo en espacios como el mirador o la terraza. La temática de servicio y atención también presenta una frecuencia elevada, con comentarios que reconocen la amabilidad y disposición del personal en diversas áreas del hotel. Adicionalmente, se registran numerosas menciones a la ubicación y el entorno, que resaltan las vistas al mar, la cercanía al puerto y la conveniencia del emplazamiento. Aspectos como las instalaciones y el ambiente y las habitaciones aparecen igualmente en el conjunto de temáticas positivas, contribuyendo a una percepción general favorable de la estadía, pese a la coexistencia de reseñas negativas analizadas en la sección anterior.

Finalmente, en el Hotel Magüipí, las reseñas positivas presentan un patrón especialmente robusto en torno a la experiencia global, las instalaciones y el entorno natural. En la Tabla 17 se recogen las principales categorías positivas y su frecuencia para este establecimiento.

Tabla 17.*Temáticas positivas — Hotel Magüipi*

Temática	2020	2021	2022	2023	2024	2025	Total
Actividades / Planes	2	5	7	4	1	0	20
Ambiente / Tranquilidad	1	1	2	1	1	0	6
Aseo / Limpieza	0	0	0	0	1	1	2
Atención / Servicio	19	42	31	26	12	7	137
Bebidas / Bar / Tienda	0	2	1	1	1	0	5
Comida / Restaurante	19	21	14	14	12	7	87
Experiencia General / Estadía	17	31	22	24	18	5	117
Habitaciones / Cabañas	1	7	3	6	1	0	18
Instalaciones / Infraestructura	37	40	54	37	19	14	201
Logística / Organización	0	0	0	1	0	1	2
Paisajes / Vistas / Entorno Natural	4	15	16	13	4	9	61
Personal / Anfitriones	7	11	7	9	5	1	40
Piscinas / Zonas húmedas	0	1	3	2	0	0	6
Playas / Mar / Charcos	3	10	6	9	1	2	31
Precio / Relación calidad- precio	0	2	1	5	1	1	10
Spa / Bienestar	0	0	1	3	0	0	4
Tecnología / Internet / Señal	0	0	0	2	0	0	2
Ubicación / Acceso	0	0	0	1	0	0	1

Temática	2020	2021	2022	2023	2024	2025	Total
Total por año	111	188	168	158	77	48	750

La información de la tabla muestra que las categorías de instalaciones/infraestructura y experiencia general/estadía concentran un número muy elevado de menciones, lo que indica que los huéspedes valoran de forma reiterada tanto las condiciones físicas del lugar (piscinas, zonas comunes, cabañas, áreas de descanso) como la vivencia integral del viaje. La atención y el servicio también son destacados con frecuencia, reforzados por comentarios sobre la calidez y amabilidad del personal. Otra temática central es la de paisajes, vistas y entorno natural, donde se resaltan las playas, el mar, la vegetación y las actividades asociadas al turismo de naturaleza (paseos, charcos, avistamiento de ballenas, entre otros). La categoría de comida y restaurante presenta igualmente un volumen importante de menciones positivas, que subrayan la calidad de la gastronomía ofrecida. En conjunto, las reseñas positivas de Magüipí configuran un perfil en el que la infraestructura, el servicio, la naturaleza y la experiencia vivida se articulan como elementos clave de satisfacción.

En síntesis, el análisis temático de las reseñas positivas muestra que, aunque existen patrones compartidos como la importancia de la atención del personal y la ubicación, cada hotel presenta también fortalezas particulares: en algunos casos destaca el restaurante, en otros el entorno natural o las instalaciones, y en otros la relación calidad–precio o la experiencia general de la estadía. Esta lectura complementa el panorama ofrecido por las temáticas negativas y proporciona insumos para la interpretación posterior de los resultados en clave comparativa, que se abordará en los capítulos de conclusiones y recomendaciones.

Estos resultados dialogan con la literatura sobre contenido generado por usuarios y percepción turística, en la medida en que las reseñas positivas permiten identificar los atributos que los huéspedes asocian con experiencias satisfactorias, además de contribuir a la construcción de una imagen favorable del establecimiento. Gretzel y Yoo (2008) señalan que las opiniones en línea influyen en la planificación de viajes; por su parte, Filieri et al. (2015) destacan que las reseñas turísticas integran componentes cognitivos y afectivos de la experiencia, como la valoración del servicio, la infraestructura, la ubicación o las emociones asociadas a la estadía. En el caso de los hoteles analizados, la recurrencia de menciones positivas sobre atención, entorno

natural, gastronomía, instalaciones y experiencia general confirma que la reputación digital se construye a partir de múltiples dimensiones de la experiencia turística, no únicamente desde la calificación numérica asignada por el usuario.

7.6 Síntesis de Resultados por Objetivo Específico

Esta sección presenta una síntesis de los principales hallazgos del capítulo de resultados en relación con los objetivos específicos planteados en el estudio, con el fin de establecer un puente entre la descripción detallada de los análisis y las conclusiones que se desarrollarán en el capítulo siguiente. En conjunto, los indicadores globales de reputación, la evolución temporal de los sentimientos y el análisis temático de las reseñas permiten evaluar en qué medida se cumplieron los objetivos propuestos y qué percepción ofrecen las opiniones de los huéspedes sobre los hoteles analizados en Buenaventura.

En relación con el primer objetivo específico, orientado a determinar de forma comparativa la distribución de sentimientos en las reseñas de los cinco hoteles más comentados de Buenaventura, los resultados de las secciones 8.2 y 8.3 muestran que el objetivo fue cumplido. Los indicadores globales permitieron comparar la reputación digital de los establecimientos, mientras que las series temporales evidenciaron variaciones en la estabilidad de la percepción durante el periodo 2020–2025. En conjunto, estos hallazgos muestran una reputación hotelera mayoritariamente positiva, aunque con trayectorias diferenciadas entre establecimientos.

En cuanto al segundo objetivo específico, centrado en establecer relaciones entre la polaridad emocional de las reseñas y la percepción turística del destino, los resultados de las secciones 8.4 y 8.5 evidencian que la percepción de los huéspedes no depende únicamente de una calificación global, sino de atributos concretos de la experiencia hotelera. Las tablas temáticas permitieron distinguir focos recurrentes de insatisfacción y fortalezas percibidas, vinculadas tanto a dimensiones funcionales del servicio como a elementos emocionales de la estadía. De esta forma, la polaridad de las reseñas se relaciona con aspectos específicos que contribuyen a la imagen turística proyectada por Buenaventura.

En lo que respecta al tercer objetivo específico, orientado a proponer recomendaciones basadas en datos para mejorar la gestión hotelera y la experiencia del huésped en Buenaventura, los resultados del capítulo aportan la base empírica necesaria para su formulación. Si bien las recomendaciones se desarrollan en el capítulo siguiente, los hallazgos aquí presentados permiten

orientar acciones de mejora diferenciadas por hotel y por dimensión de la experiencia. En este sentido, el objetivo se aborda desde la generación de evidencia analítica y se concreta posteriormente en las recomendaciones del estudio.

En síntesis, los resultados del capítulo responden a los tres objetivos específicos del estudio: caracterizan la distribución comparativa de sentimientos, relacionan la polaridad emocional con atributos concretos de la experiencia turística y sustentan la formulación de recomendaciones basadas en datos. Esta lectura permite pasar de la descripción analítica hacia una interpretación estratégica de la reputación digital hotelera en Buenaventura.

Desde una lectura integrada de destino, los resultados evidencian que la percepción turística de Buenaventura no puede reducirse a una valoración homogénea de su oferta hotelera, sino que se configura a partir de experiencias diferenciadas entre establecimientos. En conjunto, los cinco hoteles analizados muestran una reputación predominantemente positiva, asociada principalmente con la atención del personal, la ubicación, el entorno natural y la experiencia gastronómica; de igual forma, se destacan las instalaciones y las condiciones físicas como atributos relevantes en la satisfacción de los huéspedes. No obstante, también se identifican focos de insatisfacción recurrentes relacionados con habitaciones, mantenimiento, limpieza, servicio, conectividad, alimentación y gestión de la respuesta a las reseñas. Esta combinación de fortalezas y debilidades permite interpretar la percepción turística de Buenaventura como un fenómeno mixto: el destino posee atributos valorados por los visitantes y capacidad para generar experiencias satisfactorias; sin embargo, requiere fortalecer aspectos operativos y de reputación digital para consolidar una imagen turística más confiable, competitiva y sostenible.

Como producto aplicado de los resultados obtenidos, se desarrolló una página web orientada a facilitar la consulta y visualización de la información analítica del proyecto. Este recurso permite presentar de manera accesible los hallazgos derivados del análisis de sentimientos, así como la información asociada a los hoteles analizados en Buenaventura (véase Anexo D).

8 Conclusiones

Este capítulo presenta las conclusiones del estudio con base en los hallazgos obtenidos a través del análisis de sentimientos aplicado a las reseñas de Google sobre los cinco hoteles más comentados de Buenaventura entre 2020 y 2025. A diferencia del apartado de resultados centrado en la exposición detallada de cifras, temáticas y patrones narrativos, aquí el foco está en interpretar su significado desde una perspectiva integral, conectando lo observado con los objetivos planteados en la investigación.

La sección se organiza en tres bloques. Primero, se formula una conclusión general que integra los principales hallazgos cuantitativos y cualitativos. Luego, se desarrollan conclusiones específicas alineadas con cada uno de los objetivos particulares, mostrando en qué medida se cumplieron y qué revelan sobre la reputación digital y la experiencia del huésped en Buenaventura. Finalmente, se destacan los aportes del estudio en términos metodológicos, académicos y prácticos, junto con sus principales limitaciones.

Este proceso se apoyó en el análisis automatizado de contenido generado por usuarios (User Generated Content o UGC), que ha demostrado ser una fuente valiosa para captar percepciones reales de los turistas en línea (Cambria, 2011). El uso de minería de opiniones y NLP transformó grandes volúmenes de texto no estructurado en evidencias concretas para la toma de decisiones en el ámbito turístico (Zhang et al., 2018).

8.1 Conclusión General del Estudio

De manera global, el análisis de las reseñas de Google correspondientes a los cinco hoteles estudiados en Buenaventura (Cordillera, Cosmos Pacífico, Magüipí, Torre Mar y Steven Buenaventura) entre el 1.º de enero de 2020 y el 21 de agosto de 2025 permite concluir que la reputación digital del conjunto de establecimientos es mayoritariamente positiva, pero está lejos de ser homogénea. Todos los hoteles registran un predominio de opiniones favorables y calificaciones en rango satisfactorio; sin embargo, la intensidad de esa positividad, la estabilidad en el tiempo y el peso de las reseñas negativas varían de forma importante entre unos y otros. Mientras que hoteles como Magüipí y, en buena medida, Cosmos Pacífico sostienen una imagen muy favorable y relativamente estable, otros como Torre Mar y Cordillera enfrentan períodos concretos de mayor tensión reputacional, con incrementos notorios en la proporción de reseñas

negativas y señales claras de deterioro en la percepción de los huéspedes, especialmente en los últimos años del periodo analizado.

En respuesta a la pregunta de investigación: “¿Cuáles son los patrones de percepción turística de los sentimientos expresados en las reseñas de Google sobre los cinco hoteles más reseñados de Buenaventura entre el 1.º de enero de 2020 y el 21 de agosto de 2025?”, los resultados permiten concluir que la percepción turística identificada en las reseñas hoteleras es predominantemente positiva, aunque diferenciada entre establecimientos. Los patrones hallados muestran variaciones por hotel, cambios temporales en la reputación digital y relaciones claras entre la polaridad emocional de las reseñas y aspectos concretos de la experiencia del huésped, como atención, ubicación, habitaciones, limpieza, instalaciones, alimentación, entorno natural y gestión de la respuesta hotelera.

El estudio también evidencia que la reputación digital de estos hoteles no se limita a un porcentaje de reseñas positivas o a una puntuación promedio; se construye a partir de un conjunto de dimensiones específicas de la experiencia turística. Las reseñas favorables reiteran la importancia de la atención del personal, la ubicación, el entorno natural, las instalaciones, la gastronomía y otros elementos que actúan como ejes centrales de satisfacción. Por ejemplo, en Magüipí se destaca la combinación entre paisaje, infraestructura y experiencia integral; en Torre Mar sobresale el componente gastronómico; mientras que en Cosmos Pacífico y Cordillera la ubicación y la percepción general del hotel funcionan como diferenciales. En sentido inverso, las reseñas negativas concentran su atención en problemas de habitaciones, aseo, mantenimiento, dificultades logísticas (check-in, tiempos de respuesta, coordinación interna), así como en aspectos de ruido y condiciones del entorno inmediato que inciden en la calidad del descanso y la sensación de confort.

Desde la perspectiva del destino, las opiniones analizadas sugieren que la experiencia turística en Buenaventura se percibe como una combinación de fortalezas asociadas al territorio (naturaleza, paisaje, actividades en el mar y la selva) y desafíos en la calidad y consistencia del servicio hotelero. Esta percepción compuesta refuerza hallazgos de investigaciones previas que muestran cómo las emociones expresadas en línea forman parte de la construcción de imagen del destino (Echtner & Ritchie, 1991), en especial cuando el contenido proviene de usuarios y no de fuentes institucionales (Cambria, 2011). Las reseñas no solo describen a cada hotel de manera

aislada, sino que, en conjunto, configuran un relato más amplio sobre Buenaventura como lugar de visita.

En este sentido, el análisis de sentimientos aplicado a reseñas en línea, enmarcado dentro de la categoría de Online Travel Reviews (OTR) y contenido generado por el usuario (UGC), se confirma como una herramienta útil para sistematizar percepciones en tiempo casi real, tal como lo han señalado autores como Zhang et al. (2018) y Feldman & Sanger (2007). La aplicación del enfoque CRISP-DM estructuró el proceso analítico de forma replicable y validada, en línea con los antecedentes de Giehl (2018) y Barrera & Barrera (2022), quienes demostraron la pertinencia de esta metodología para transformar datos textuales en hallazgos estratégicos en contextos turísticos.

En síntesis, el estudio permite concluir que la reputación digital de los hoteles analizados es, en general, favorable, pero atraviesa por trayectorias diferenciadas: algunos establecimientos consolidan una imagen robusta y coherente con las expectativas de sus huéspedes, mientras que otros muestran señales de desgaste o de variabilidad en la calidad percibida durante el periodo analizado. Este panorama abre la puerta a la formulación de recomendaciones específicas y basadas en datos, orientadas a reforzar las fortalezas identificadas, abordar las debilidades más recurrentes y aprovechar la información disponible en las reseñas en línea como insumo estratégico para la gestión hotelera y el fortalecimiento de la experiencia turística en Buenaventura.

8.2 Conclusiones Frente a los Objetivos Específicos

En relación con el primer objetivo específico, orientado a determinar de forma comparativa la distribución de sentimientos (positivos, neutrales y negativos) en las reseñas de los cinco hoteles más comentados de Buenaventura, el análisis mostró que, si bien la reputación digital global es positiva, esta se manifiesta de manera heterogénea entre hoteles y variable a lo largo del tiempo. Las trayectorias reputacionales no son estáticas: algunos hoteles mantienen niveles altos y estables de satisfacción (como Magüipí), mientras que otros experimentan fluctuaciones, descensos o fases de recuperación. Este comportamiento confirma hallazgos previos en análisis de sentimientos aplicados a turismo (Guadarrama-Gómez & Rodríguez-Rodríguez, 2024; Araya-Pizarro & Verelst, 2024), donde se reconoce que la percepción digital responde tanto a factores estructurales como coyunturales. Además, respalda teóricamente el enfoque de polaridad emocional como dimensión clave del análisis automático de reseñas (Zhang, Wang & Liu, 2018).

Con respecto al segundo objetivo específico, enfocado en establecer relaciones entre la polaridad emocional de las reseñas y la percepción turística del destino, se concluye que las dimensiones que configuran la satisfacción o insatisfacción del huésped no se distribuyen al azar, sino que siguen patrones temáticos recurrentes. Las reseñas negativas apuntan con regularidad a deficiencias en aseo, mantenimiento, ruidos, tiempos de respuesta o problemas logísticos, tal como lo han reportado estudios previos que aplicaron análisis temático en hoteles urbanos (Guerrero Rodríguez et al., 2018). En contraste, los comentarios positivos suelen destacar la atención del personal, la ubicación y el entorno natural, lo cual confirma que la experiencia del huésped es multidimensional y está mediada por factores tangibles e intangibles.

En términos de contenido, estas observaciones refuerzan lo planteado por Echtner & Ritchie (1991) sobre la construcción de imagen de un destino a partir de elementos funcionales y psicológicos, así como lo señalado por Cambria (2011) sobre el papel de las emociones codificadas en lenguaje natural como indicador de actitud hacia una experiencia. Además, este hallazgo se conecta con estudios que han utilizado minería de opiniones para mapear percepciones agregadas sobre destinos turísticos (Herrera & Salgado, 2022), donde se reconoce que las reseñas actúan como relato colectivo y espejo de los valores atribuidos al territorio.

En cuanto al tercer objetivo específico, relacionado con proponer recomendaciones basadas en datos que orienten la mejora de la gestión hotelera y la experiencia del huésped en Buenaventura, se concluye que los resultados obtenidos, tanto cuantitativos como cualitativos, permiten identificar ejes de acción estratégicos para cada establecimiento. El procesamiento automatizado de texto mediante técnicas de NLP, complementado con dashboards en Power BI, brindó una base empírica sólida para señalar fortalezas diferenciales, debilidades persistentes y áreas críticas de intervención. Esta capacidad de extracción de conocimiento útil desde datos no estructurados valida los aportes metodológicos del modelo CRISP-DM (Chapman et al., 2000) y del uso de herramientas como Apify y Azure en flujos de análisis semántico (Giehl, 2018; Valle, 2021).

Finalmente, el análisis también reafirma la pertinencia de emplear Google Reviews como fuente principal de estudio. Tal como se argumentó en los antecedentes, su masividad y actualización constante lo posicionan como una herramienta clave para monitorear la percepción turística en tiempo casi real (Piñero & Mellinas, 2025), aunque su naturaleza breve y no filtrada exige una sistematización rigurosa para garantizar hallazgos interpretables.

8.3 Aportes y limitaciones del Estudio

Desde un enfoque académico, este trabajo aporta evidencia empírica sobre el uso de técnicas de minería de texto y análisis automatizado de sentimientos en el sector hotelero de un destino emergente como Buenaventura, para el cual no existían antecedentes documentados de sistematización masiva de reseñas digitales. Al combinar análisis cuantitativo y cualitativo, el estudio permite caracterizar patrones de percepción a lo largo del tiempo y entre distintos establecimientos, con base en un corpus de 3.727 reseñas publicadas en Google Reviews entre enero de 2020 y agosto de 2025. Este tipo de aproximación refuerza lo planteado por Manosso y Domareski (2021) respecto a la utilidad del análisis de sentimientos como herramienta de monitoreo reputacional, y valida lo propuesto por Pang (2006) y Cambria (2011) sobre la capacidad del lenguaje natural para codificar estados emocionales que reflejan la experiencia del usuario.

En el plano metodológico, el estudio demuestra la viabilidad de estructurar un flujo de procesamiento de datos a partir de herramientas accesibles y replicables. La implementación del modelo CRISP-DM (Chapman et al., 2000) organizó de manera sistemática las fases de recolección automatizada (scraping con Apify), almacenamiento, transformación, análisis semántico y visualización. La tasa de coincidencia del 90,2% obtenida entre etiquetas automáticas y validación manual respalda la consistencia interpretativa del proceso y refleja que, incluso con modelos de análisis no personalizados, es posible lograr resultados confiables cuando se establecen filtros y ajustes en las fases de enriquecimiento. Esto coincide con lo señalado por Rashidi (2024) en cuanto a la importancia de diseñar proyectos realistas y técnicamente justificados en contextos de recursos limitados.

En cuanto al valor aplicado, los hallazgos generados se presentan en un formato visual y segmentado, mediante dashboards construidos en Power BI, que facilita la toma de decisiones por parte de actores turísticos locales. Al identificar dimensiones críticas como limpieza, atención, infraestructura, entorno o logística, el estudio convierte las reseñas en insumos estratégicos para mejorar la gestión hotelera y elevar la calidad de la experiencia turística. Este tipo de aplicaciones respalda lo señalado por Gretzel & Yoo (2008) sobre el potencial transformador de las online travel reviews (OTR) como herramienta de retroalimentación y toma de decisiones.

No obstante, el estudio presenta varias limitaciones que deben ser reconocidas. La primera es la fuente única de datos: al basarse exclusivamente en Google Reviews, se excluyen otras

plataformas relevantes como TripAdvisor o Booking, que podrían ofrecer percepciones complementarias o segmentadas según tipo de viajero. Aunque Google fue seleccionada por su cobertura y volumen, como se argumentó en los antecedentes (Piñero & Mellinas, 2025), esta elección introduce un sesgo posible en el perfil del usuario cuyas opiniones son analizadas. La segunda limitación está relacionada con el periodo temporal: si bien la muestra cubre más de cinco años, el corte en agosto de 2025 deja por fuera eventos o tendencias del último cuatrimestre del año, lo cual puede limitar inferencias longitudinales.

También existen restricciones operativas asociadas al desarrollo del proyecto. El trabajo fue diseñado y ejecutado por una sola persona en el marco de un semestre académico, lo que implicó acotar la profundidad de ciertos análisis y asumir criterios unipersonales en la limpieza, clasificación y validación de datos. A ello se suma una limitación de tipo financiero: el uso de servicios en la nube se restringió a los recursos disponibles en la cuenta educativa, lo cual obligó a reducir la frecuencia del scraping y a limitar la integración de modelos avanzados. Tal como han señalado estudios previos (Giehl, 2018; Valle, 2021), la infraestructura tecnológica puede condicionar el alcance de los proyectos basados en datos, especialmente en contextos donde no se cuenta con créditos extensivos o servidores dedicados.

En resumen, aunque el estudio enfrenta restricciones de fuente, cobertura temporal y recursos operativos, su aporte metodológico, empírico y aplicado resulta relevante como línea base para futuras investigaciones en turismo digital en Colombia. Más allá de describir un fenómeno, ofrece herramientas y aprendizajes útiles para replicar, escalar y profundizar el análisis de percepciones turísticas en destinos similares, donde la reputación digital aún no ha sido sistematizada con rigurosidad.

9 Recomendaciones

En este capítulo se presentan las recomendaciones derivadas de los resultados y conclusiones del estudio, con el propósito de orientar acciones de mejora en la gestión hotelera y en la experiencia del huésped en Buenaventura. A diferencia del capítulo de resultados, centrado en describir lo que muestran los datos, y del capítulo de conclusiones, enfocado en interpretar esos hallazgos, aquí el énfasis se sitúa en proponer líneas de acción concretas, sustentadas en la evidencia generada a partir del análisis de las reseñas digitales. Las recomendaciones se organizan en tres niveles. En primer lugar, se formulan recomendaciones generales aplicables al conjunto de hoteles analizados, que abordan aspectos transversales como la gestión de reseñas en línea, el seguimiento sistemático de indicadores y el uso de herramientas de analítica como apoyo a la toma de decisiones. En segundo lugar, se proponen recomendaciones específicas para cada hotel, en función de los patrones de sentimiento y de las temáticas de satisfacción e insatisfacción identificadas en el estudio, de modo que cada establecimiento cuente con insumos diferenciados para priorizar acciones de mejora. Finalmente, se presentan recomendaciones dirigidas a actores del destino y a futuras investigaciones, orientadas a aprovechar los resultados del trabajo en clave de política turística, articulación sectorial y continuidad académica de la línea de análisis de reputación digital en Buenaventura.

9.1 Recomendaciones Generales para los Hoteles Analizados

Con base en los resultados obtenidos y en las conclusiones formuladas, es posible plantear un conjunto de recomendaciones generales que son aplicables, en mayor o menor medida, a todos los hoteles analizados. Estas recomendaciones no sustituyen las acciones específicas que cada establecimiento debería emprender según su propio perfil, pero sí ofrecen una hoja de ruta común para fortalecer la gestión de la reputación digital y la experiencia del huésped en Buenaventura.

La gestión activa de reseñas no es solo intuitiva: está respaldada por evidencia empírica. Un estudio global de TripAdvisor en asociación con Ipsos MORI, realizado a más de 23 000 viajeros, reveló que el 63 % de los encuestados afirmó que estaría más dispuesto a reservar un hotel o restaurante si el propietario respondía a la mayoría de las reseñas. Además, cuando las respuestas del propietario son personalizadas, el 77 % manifestó que su probabilidad de reservar aumentaba. Por otra parte, el 89 % señaló que una respuesta cuidadosa a una reseña negativa mejoraba su impresión sobre el establecimiento. Estos hallazgos evidencian que no basta con

acumular reseñas positivas: la interacción con los comentarios especialmente mediante respuestas personalizadas y atentas influye significativamente en la percepción del servicio, en la intención de reserva y en la reputación digital del establecimiento (TripAdvisor & Ipsos MORI, 2019).

En primer lugar, se recomienda institucionalizar el seguimiento sistemático de las reseñas en línea como parte de la gestión cotidiana de cada hotel. Esto implica no limitarse a observar la calificación promedio en estrellas, sino monitorear de manera periódica indicadores como la distribución de sentimientos (porcentaje de reseñas positivas, negativas y neutras), la evolución de estos indicadores en el tiempo y las temáticas más mencionadas en los comentarios. El uso de paneles de control y dashboards, como los construidos en este estudio, puede incorporarse como herramienta de apoyo para reuniones de seguimiento, comités de calidad o espacios de planeación interna.

En segundo lugar, se recomienda fortalecer la gestión activa de las reseñas, en particular mediante la respuesta oportuna y respetuosa a los comentarios publicados por los huéspedes. Responder reseñas negativas de forma empática, ofreciendo explicaciones cuando corresponda e indicando las acciones que se tomarán, contribuye a mitigar el impacto de las experiencias insatisfactorias y envía un mensaje claro de compromiso con la mejora. De igual forma, agradecer las reseñas positivas y destacar los aspectos que el huésped valoró refuerza la relación con el cliente y ayuda a consolidar una comunidad de usuarios que se siente escuchada. Para ello, se sugiere definir lineamientos internos sobre tiempos de respuesta, tono comunicativo y responsables de esta tarea.

En tercer lugar, los resultados muestran que muchos focos de insatisfacción se concentran en dimensiones operativas básicas, por lo que se recomienda priorizar acciones de mejora en habitaciones, aseo, mantenimiento y logística del servicio. Aunque las particularidades varían entre hoteles, en términos generales es conveniente revisar de forma sistemática el estado de las habitaciones (mobiliario, baños, ventilación, control de plagas, olores), los protocolos de limpieza, el funcionamiento de equipos (aires acondicionados, duchas, cerraduras, tarjetas de acceso) y los procesos de check-in, manejo de reservas y tiempos de respuesta. La información extraída de las reseñas puede utilizarse como insumo para elaborar listas de verificación, planes de mantenimiento preventivo y ajustes en los procedimientos de atención.

En cuarto lugar, las reseñas positivas confirman el peso central de la atención y trato del personal en la percepción de los huéspedes. Por ello, se recomienda invertir de manera sostenida

en formación y acompañamiento del talento humano, no solo en aspectos técnicos, sino también en habilidades blandas relacionadas con la comunicación, la gestión de conflictos, la hospitalidad y la empatía. La sistematización de buenas prácticas identificadas en algunos hoteles del estudio puede servir como referencia para otros establecimientos que busquen mejorar su cultura de servicio.

Finalmente, se recomienda aprovechar la información disponible en las reseñas para diseñar estrategias de posicionamiento y mejora continua. Las fortalezas identificadas como la gastronomía, el entorno natural, la ubicación, la relación calidad-precio o determinadas instalaciones pueden convertirse en elementos explícitos de la propuesta de valor de cada hotel, mientras que las debilidades recurrentes deben ser abordadas con planes de acción concretos y medibles. La combinación de análisis cuantitativo (indicadores y tendencias) con análisis cualitativo (lectura de comentarios y temáticas) permite pasar de una gestión intuitiva de la reputación a una gestión basada en datos, más alineada con las exigencias del turismo contemporáneo y con las expectativas de los visitantes que consultan reseñas antes de tomar decisiones de viaje.

9.2 Recomendaciones Específicas por Hotel

A partir de los patrones de sentimientos, las temáticas de satisfacción e insatisfacción y la evolución temporal de las reseñas, es posible formular recomendaciones diferenciadas para cada uno de los hoteles analizados. Estas propuestas buscan orientar acciones concretas de mejora, partiendo de las particularidades de cada establecimiento.

Hotel Cosmos Pacífico

En el caso de Cosmos Pacífico, las reseñas muestran una reputación globalmente favorable, con fortalezas claras en atención del personal, percepción general del hotel y ubicación. Se recomienda consolidar estas fortalezas convirtiéndolas en parte explícita de la propuesta de valor del establecimiento (por ejemplo, resaltando en la comunicación la cercanía a puntos estratégicos y la calidad del servicio). Al mismo tiempo, las reseñas negativas señalan focos específicos relacionados con ruido en ciertas áreas y aspectos puntuales de comida/restaurante. En consecuencia, se sugiere: (i) revisar horarios, fuentes de ruido y posibles medidas de mitigación (distribución de habitaciones, información anticipada a los huéspedes, ajustes en actividades

nocturnas) y (ii) utilizar los comentarios sobre gastronomía como insumo para ajustar la oferta (variedad, tiempos de servicio, relación calidad–precio), priorizando los aspectos que más se repiten en las reseñas críticas.

Hotel Cordillera

Para el Hotel Cordillera, los resultados evidencian un volumen alto de reseñas y un deterioro reputacional en los últimos años, asociado sobre todo a problemas de habitaciones, aseo, mantenimiento y, en menor medida, a aspectos de servicio y ruido. Se recomienda priorizar un plan integral de mejora en infraestructura y condiciones sanitarias, que incluya: diagnóstico detallado del estado de habitaciones (humedades, ventilación, olores, plagas), revisión y reparación de equipos (duchas, aire acondicionado, cerraduras, agua caliente) y fortalecimiento de los protocolos de limpieza. A la par, las reseñas positivas muestran fortalezas importantes en atención del personal, ubicación y relación calidad–precio. Por ello, además de corregir los problemas físicos, se sugiere reforzar estas fortalezas en la comunicación y en la experiencia del huésped, presentando las mejoras en infraestructura como una respuesta directa a las inconformidades evidenciadas en las reseñas. Resulta clave, en este hotel, activar una política de respuesta sistemática a las reseñas negativas, explicando ajustes realizados y mostrando disposición a corregir fallas recurrentes.

Hotel Steven Buenaventura

En Steven Buenaventura, el número de reseñas es menor, pero la evolución reciente muestra una tendencia de mejora en la percepción. Las reseñas positivas destacan la atención y trato del personal y la experiencia general, mientras que las críticas se concentran en confort de las habitaciones, ruido y entorno inmediato, así como en algunos aspectos de logística. Se recomienda capitalizar la tendencia positiva reciente poniendo en valor la calidad del servicio y, simultáneamente, implementar acciones específicas para mejorar el confort: revisión de mobiliario, ropa de cama, condiciones de ventilación y aislamientos básicos frente al ruido, en la medida en que la infraestructura lo permita. También es aconsejable revisar los procesos de check-in, manejo de reservas y tiempos de respuesta a solicitudes sencillas (toallas, llaves, cambios de habitación), de modo que la percepción positiva que ya existe sobre el personal se vea acompañada por una sensación de mayor fluidez operativa. Dado el menor volumen de reseñas, cada comentario

tiene un peso relativo alto; por ello, se sugiere llevar un seguimiento cercano de las nuevas opiniones y responderlas de forma constante.

Hotel Torre Mar

En el Hotel Torre Mar, los resultados combinan una buena valoración de la gastronomía, la ubicación y ciertas experiencias (como el mirador/terraza) con la presencia de un año crítico en términos de aumento de reseñas negativas, asociado a problemas de servicio, mantenimiento e infraestructura. Se recomienda, en primer lugar, proteger y potenciar el componente gastronómico como sello distintivo del hotel, asegurando consistencia en la calidad de los platos, tiempos de servicio y atención en el restaurante, especialmente en períodos de alta demanda. En segundo lugar, se sugiere implementar un plan de mantenimiento preventivo y correctivo de habitaciones y zonas comunes, orientado a reducir las inconformidades por equipos que no funcionan, desgaste visible o condiciones de aseo mejorables. Dado que el porcentaje de reseñas respondidas es nulo, una recomendación prioritaria es comenzar a responder las reseñas de forma sistemática, en particular aquellas que documentan experiencias negativas argumentadas, de modo que se muestre al huésped que sus comentarios son leídos y generan acciones concretas.

Hotel Magüipí

En Magüipí, la reputación digital es especialmente favorable, con una proporción muy alta de reseñas positivas asociadas a instalaciones, entorno natural, experiencia general, atención del personal y gastronomía, y un volumen reducido de reseñas negativas. En este caso, la principal recomendación es sistematizar y mantener las buenas prácticas que han permitido construir esta imagen, evitando que el crecimiento de la demanda o cambios operativos comprometan la calidad percibida. En términos específicos, se sugiere: (i) seguir cuidando la experiencia integral (actividades, contacto con la naturaleza, comodidad de las zonas comunes y cabañas) como elemento central de diferenciación; (ii) utilizar las reseñas positivas como insumo para campañas de comunicación y posicionamiento del hotel y del destino; y (iii) atender de manera puntual las reseñas negativas que mencionan aspectos de logística, mantenimiento puntual o señal de internet, para evitar que pequeños problemas se conviertan en focos de insatisfacción. Dado que Magüipí presenta también un porcentaje relativamente alto de reseñas respondidas, se recomienda mantener

y formalizar esta práctica como parte de su estrategia de gestión de reputación en línea, reforzando la imagen de cercanía y cuidado hacia el huésped.

9.3 Recomendaciones para Actores del Destino y Futuras Investigaciones

A partir de los resultados obtenidos en este estudio, se identifican diversas líneas de trabajo que podrían ser abordadas en investigaciones futuras o en desarrollos posteriores orientados a fortalecer la analítica turística en Buenaventura. Estas proyecciones buscan ampliar el alcance, la profundidad y la aplicabilidad del análisis realizado, atendiendo tanto a limitaciones metodológicas como a nuevas oportunidades detectadas durante el proceso.

En primer lugar, se propone ampliar la cobertura de fuentes de información, incorporando reseñas de plataformas especializadas como Booking, TripAdvisor o Airbnb. Esta triangulación permitiría contrastar si los patrones de percepción observados en Google Reviews se mantienen o varían según el canal, el tipo de usuario o la lógica de funcionamiento de cada plataforma. Como lo sugiere Piñero y Mellinas (2025), la diversificación de fuentes puede mejorar la representatividad y riqueza descriptiva del análisis.

En segundo lugar, se recomienda extender la serie temporal más allá del año 2025 y establecer esquemas de captura más frecuentes y automatizados. Un sistema de monitoreo casi en tiempo real como el planteado en estudios previos sobre destinos con alta variabilidad reputacional (Gretzel & Yoo, 2008) permitiría observar la evolución dinámica de la percepción del huésped y ajustar decisiones en función de cambios inmediatos en el comportamiento del consumidor.

En tercer lugar, se sugiere integrar nuevas variables externas que complementen el análisis de sentimientos con indicadores cuantitativos, tales como tarifas promedio, niveles de ocupación, procedencia geográfica de los visitantes o estadísticas oficiales de turismo. Esta triangulación de fuentes aportaría una comprensión más robusta sobre cómo la percepción expresada en las reseñas se relaciona con el desempeño económico, la competitividad del destino y los factores de contexto.

Finalmente, futuras investigaciones podrían profundizar en el uso de técnicas avanzadas de análisis de texto, como el modelado de tópicos (topic modeling), análisis de emociones más allá de la polaridad tradicional (positiva, neutra, negativa) o segmentaciones basadas en idioma, tipo de huésped o dispositivo de origen. Estas técnicas permitirían matizar la interpretación de las experiencias y capturar sutilezas emocionales y temáticas que pueden ser críticas para la gestión turística. Asimismo, el fortalecimiento de modelos explicativos e interpretables como los

propuestos por Atrey (2023) y Cambria (2011) abre la posibilidad de desarrollar herramientas más comprensibles y accionables para los actores locales.

En conjunto, estas líneas futuras buscan consolidar un enfoque de análisis turístico que sea continuo, multisource, contextualizado y útil para la toma de decisiones públicas y privadas en destinos como Buenaventura, donde la voz del visitante, sistematizada a través de tecnologías del lenguaje, puede convertirse en un insumo estratégico de primer orden

.

Referencias

- Água, M., António, N., Carrasco, P., & Rassal, C. (2025). Large language models powered aspect-based sentiment analysis for enhanced customer insights. *Tourism & Management Studies*, 21(1), 1–19. <https://doi.org/10.18089/tms.20250101>
- Ahani, A., Nilashi, M., Ibrahim, O., Sanzogni, L., & Weaven, S. (2019). Market segmentation and travel choice prediction in Spa hotels through TripAdvisor's online reviews. *International Journal of Hospitality Management*, 80, 52–77. <https://doi.org/10.1016/j.ijhm.2019.01.003>
- Alcaldía Distrital de Buenaventura. (2025). Alcaldía Distrital y Gobernación del Valle desarrollan proyecto tecnológico para fomentar el turismo en Buenaventura. Alcaldía Distrital de Buenaventura. <https://acortar.link/n8Cc8G>
- Álvarez-Carmona, M. Á., Aranda, R., Guerrero-Rodríguez, R., Rodríguez-González, A. Y., & López-Monroy, A. P. (2022). A combination of sentiment analysis systems for the study of online travel reviews: Many heads are better than one. *Computación y Sistemas*, 26(2). <https://doi.org/10.13053/cys-26-2-4055>
- AmCham Colombia. (2021, noviembre 24). La digitalización de la industria turística en Colombia será indispensable para acelerar su recuperación: Google. <https://acortar.link/uhwIjU>
- AmCham Colombia. (2022, junio 14). Recuperación del turismo en Colombia necesita apalancarse en herramientas digitales. <https://acortar.link/Yf4qTl>
- Anderson, C. K., & Han, S. (2016). Hotel performance impact of socially engaging with consumers. *Cornell Hospitality Report*, 16(10). <https://hdl.handle.net/1813/71227>
- Araya-Pizarro, S., & Verelst, N. (2024). Más allá de las reseñas: Desvelando la experiencia hotelera con minería de texto. *Revista Interamericana de Ambiente y Turismo*, 20(2), 126–139. <https://doi.org/10.4067/S0718-235X2024000200126>
- Atrey, P. (2023). Studying product reviews using sentiment analysis based on interpretable machine learning [Tesis de maestría, University of Maryland]. <https://doi.org/10.13016/6cvw-dlwd>
- Bagherzadeh, S., Shokouhyar, S., Jahani, H., & Sigala, M. (2021). A generalizable sentiment analysis method for creating a hotel dictionary: using big data on TripAdvisor hotel reviews. *Journal of Hospitality and Tourism Technology*, 12(2), 1–30. <https://acortar.link/HEHt8H>
- Barrera Barrera, J. C., & Barrera Barrera, K. M. (2022). Desarrollo de una aplicación híbrida inteligente y un modelo de análisis de datos en fases para resumir y representar los comentarios de clientes sobre productos y servicios de bares y restaurantes mediante Django, IONIC, Angular, Deep Learning y redes sociales [Trabajo de grado, Universidad Politécnica Salesiana]. Repositorio Institucional de la Universidad Politécnica Salesiana. <https://dspace.ups.edu.ec/handle/123456789/23420>
- Cámara de Comercio de Buenaventura. (2025). Encuesta Ritmo Empresarial Buenaventura 2025. <https://tinyurl.com/26rcjw2h>

- Cambria, E. (2011). Application of Common Sense Computing for the Development of a Novel Knowledge-Based Opinion Mining Engine [Tesis doctoral, University of Stirling]. <https://dspace.stir.ac.uk/bitstream/1893/6497/1/thesis.pdf>
- Chan, E. S. W., & Lam, D. (2013). Hotel safety and security systems: Bridging the gap between managers and guests. *International Journal of Hospitality Management*, 32(1), 202–216. <https://doi.org/10.1016/j.ijhm.2012.05.010>
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide. CRISP-DM Consortium. <https://tinyurl.com/yedy8ps3>
- Coello Vallecillo, R. E. (2023). Análisis de sentimiento basado en los aspectos de las reseñas de una cadena de restaurantes en Honduras [Trabajo de grado, Universidad Tecnológica Centroamericana]. <https://repositorio.unitec.edu/xmlui/handle/123456789/12901>
- Colombia Travel. (s.f.). Buenaventura. <https://colombia.travel/es/buenaventura>
- Comfenalco Valle. (s.f.). La Bocana – Destinos y atractivos turísticos. <https://www.comfenalcovalle.com.co/politica-de-turismo-sostenible/la-bocana-destinos-y-atractivos-turisticos/>
- CWMas. (2025, abril 5). Turismo de Buenaventura en riesgo por ola de violencia. <https://tinyurl.com/2bxs2pd4>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171–4186. <https://doi.org/10.18653/v1/N19-1423>
- Echtner, C. M., & Ritchie, J. R. B. (1991). The meaning and measurement of destination image. *Journal of Tourism Studies*, 2(2), 2–12. <https://tinyurl.com/msdxc92d>
- El País. (2024, enero 9). Los indicadores que encendieron otra alarma sobre la economía de Buenaventura. <https://www.elpais.com.co/valle/los-indicadores-que-encendieron-otra-alarma-sobre-la-economia-de-buenaventura-que-esta-pasando-0901.html>
- Feldman, R., & Sanger, J. (2007). *The text mining handbook: Advanced approaches in analyzing unstructured data*. Cambridge University Press. <https://tinyurl.com/3766mynp>
- Filieri, R., Alguezaui, S., & McLeay, F. (2015). Why do travelers trust TripAdvisor? Antecedents of trust towards consumer-generated media and its influence on recommendation adoption and word of mouth. *Tourism Management*, 51, 174–185. <https://acortar.link/WDrPPH>
- Giehl, P. M. (2018). Automated Sentiment Analysis of the Swiss Media Landscape [Tesis de maestría, Université de Fribourg]. <http://tiny.cc/j464101>
- Gretzel, U., & Yoo, K.-H. (2008). Use and impact of online travel reviews. En P. O'Connor, W. Höpken, & U. Gretzel (Eds.), *Information and Communication Technologies in Tourism 2008* (pp. 35–46). Springer. <https://tinyurl.com/2xtrqlwc>

- Guadarrama-Gómez, H., & Rodríguez-Rodríguez, A. (2024). Análisis de sentimientos en reseñas de establecimientos de alimentos y bebidas de Bahías de Huatulco. <https://tinyurl.com/5n7rpk89>
- Guerrero Rodríguez, R., Pérez Sánchez, M., & Alvarado Sizzo, I. (2018). ¿Qué opinan nuestros turistas? Un análisis temático en TripAdvisor sobre atractivos turísticos de la ciudad de Guanajuato. *Dimensiones Turísticas*, 2(3), 91–110. <https://doi.org/10.47557/VTGU4902>
- Herrera Herrera, N. I., & Salgado Reyes, N. E. (2022). Evaluación de sitios turísticos mediante análisis de sentimientos de comentarios emitidos por usuarios en redes sociales. *Revista Tecnológica - ESPOL*, 34(2), 124–138. <https://doi.org/10.37815/rte.v34n2.921>
- Hutto, C. J., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1), 216–225. <https://doi.org/10.1609/icwsm.v8i1.14550>
- Liu, B. (2012). Sentiment analysis and opinion mining. University of Illinois at Chicago. <https://tinyurl.com/y82cjygl>
- Manosso, F. C., & Domareski Ruiz, T. C. (2021). Using sentiment analysis in tourism research: A systematic, bibliometric, and integrative review. *Journal of Tourism, Heritage & Services Marketing*, 7(2), 16–27. <https://doi.org/10.5281/zenodo.5548426>
- Mao, Y., Liu, Q., & Zhang, Y. (2024). Sentiment analysis methods, applications, and challenges: A systematic literature review. *Journal of King Saud University - Computer and Information Sciences*, 36(4), 102048. <https://doi.org/10.1016/j.jksuci.2024.102048>
- McCartney, A., Hensman, S., & Longo, L. (2017, December 7-8). “How short is a piece of string?”: The impact of text length and text augmentation on short-text classification accuracy (pp. 102-114). En *Proceedings of the 25th Irish Conference on Artificial Intelligence and Cognitive Science (AICS 2017)*, Dublín, Irlanda. <https://doi.org/10.21427/d7151m>
- Mendoza-Moreno, J. F., Santamaria-Granados, L., Fraga Vázquez, A., & Ramirez-Gonzalez, G. (2021). OntoTouTra: Tourist traceability ontology based on big data analytics. *Applied Sciences*, 11(22), 11061. <https://doi.org/10.3390/app112211061>
- Oliveira, A. S., Renda, A. I., Correia, M. B., & Antonio, N. (2022). Hotel customer segmentation and sentiment analysis through online reviews: an analysis of selected European markets. *Tourism and Management Studies*, 18(1), 1–12. <https://doi.org/10.18089/tms.2022.180103>
- Organización Mundial del Turismo. (2021). Entender y gestionar los datos del turismo en la era digital. <https://www.unwto.org/es/publicaciones>
- Pang, B. (2006). Automatic analysis of document sentiment [Tesis doctoral, Cornell University]. <https://www.cs.cornell.edu/home/llee/extra/pang-thesis.pdf>
- Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)*, 79–86. <https://doi.org/10.3115/1118693.1118704>

- Piñero, M. S., & Mellinas Cánovas, J. P. (2025). Comparativa entre Google Reviews y TripAdvisor en destinos turísticos especializados: El caso de La Rioja. En E. González Fandos & Á. Sánchez Hernández (Eds.), *Tendencias en vino y agroalimentación en el siglo XXI: retos y oportunidades* (pp. 175–183). Universidad de La Rioja. <https://dialnet.unirioja.es/servlet/articulo?codigo=10413530>
- Rashidi, S. (2024). *Your AI survival guide: scraped knees, bruised elbows, and lessons learned from real-world AI deployments* (1st ed.). John Wiley & Sons, Inc.
- Rittichainuwat, B. N., & Chakraborty, G. (2012). Perceptions of importance and what safety is enough. *Journal of Business Research*, 65(1), 42–50. <https://doi.org/10.1016/j.jbusres.2011.07.013>
- Soy de Buenaventura. (2025). Congreso de Turismo Responsable 2025: Trenzando turismo responsable y sostenible: Buenaventura, un sueño en común. Soy de Buenaventura. <http://tiny.cc/dk54101>
- Tourism Economics. (2021). *Driving the Tourism Recovery in Colombia: Google Colombia Final Report*. Google. <https://tinyurl.com/drivingtourism>
- TripAdvisor & Ipsos MORI. (2019, diciembre 12). TripAdvisor study reveals 77% of travelers more likely to book when business owners respond to reviews. TripAdvisor MediaRoom. <https://tinyurl.com/mr3fm9d5>
- Valle Gutiérrez, G. (2021). Aplicación de técnicas de web scraping y procesamiento del lenguaje natural para la extracción y evaluación de información de una página web de empleo [Trabajo Fin de Máster, Universidad Pontificia Comillas]. <https://acortar.link/qVZtRe>
- Van der Ende, D. (2020). *Scraping Scratch: A Dataset of Comments and their Sentiment* [Tesis de maestría, Leiden University]. <https://theses.liacs.nl/1883>
- Vargas-Calderón, V., Moros Ochoa, A., Castro Nieto, G. Y., & Camargo, J. E. (2021). Machine learning for assessing quality of service in the hospitality sector based on customer reviews. *Information Technology & Tourism*, 23(3), 351–379. <https://doi.org/10.1007/s40558-021-00207-4>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. En I. Guyon et al. (Eds.), *Advances in Neural Information Processing Systems*, 30. Curran Associates. <https://papers.neurips.cc/paper/7181-attention-is-all-you-need>
- Yannacopoulou, A., & Kallinikos, K. (2025). Measuring destination image using AI and big data: Kastoria's image on TripAdvisor. *Societies*, 15(1), 5. <https://doi.org/10.3390/soc15010005>
- Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. *arXiv*. <https://arxiv.org/abs/1801.07883>
- Zhang, W., Deng, Y., Liu, B., Pan, S. J., & Bing, L. (2024). Sentiment analysis in the era of large language models: A reality check. *Findings of the Association for Computational Linguistics: NAACL 2024*, 3881–3906. <https://doi.org/10.18653/v1/2024.findings-naacl.246>

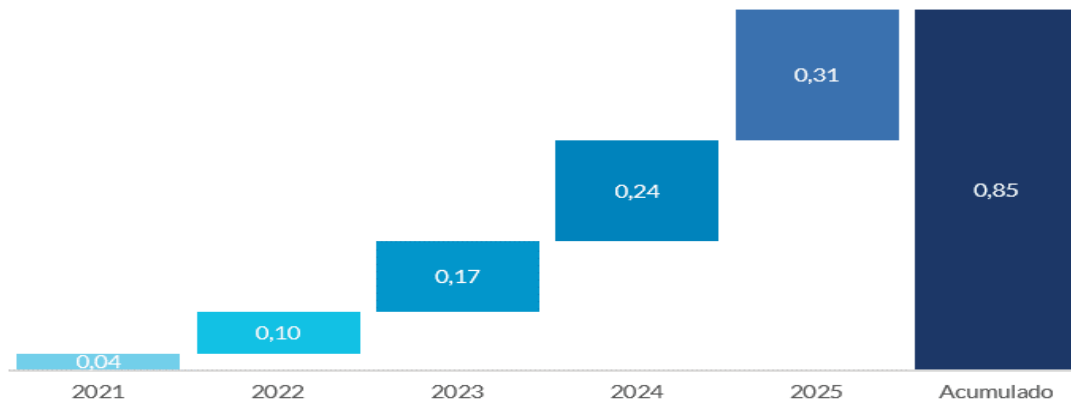
Anexo A. Perspectiva de referencia y crecimiento potencial atribuible a una mayor presencia en línea para Colombia

		2021	2022	2023	2024	2025
Noches (en millones)	De referencia	20,8	26,1	29,8	32,2	34,3
	Perspectiva de oportunidad	20,9	26,3	30,1	32,6	34,8
	Aumento potencial	0,1	0,2	0,3	0,4	0,6
Gastos (miles de millones de USD)	De referencia	6,6	8,7	10,2	11,2	12,2
	Perspectiva de oportunidad	6,7	8,8	10,4	11,4	12,5
	Aumento potencial	0,0	0,1	0,2	0,2	0,3
PIB (miles de millones de USD)	De referencia	9,3	12,5	14,7	16,2	17,7
	Perspectiva de oportunidad	9,3	12,7	15,0	16,5	18,1
	Aumento potencial	0,1	0,1	0,2	0,3	0,4
Empleos (miles)	De referencia	662	842	944	996	1041
	Perspectiva de oportunidad	666	852	959	1017	1067
	Aumento potencial	4	10	15	21	26

Nota. Fuente <https://tinyurl.com/drivingtourism> (Tourism Economics, 2021).

Anexo A. Oportunidad para un mayor gasto en turismo en Colombia (2021–2025)

En millones de USD nominales



Fuente: Tourism Economics

Nota. Fuente <https://tinyurl.com/drivingtourism> (Tourism Economics, 2021).

Anexo C. Evidencia del repositorio del proyecto en GitHub

El repositorio de GitHub constituye una evidencia técnica del desarrollo realizado, al permitir consultar los archivos, scripts, estructura del proyecto y componentes implementados durante el proceso analítico.

Enlace: <https://github.com/Pipe199x/Buenaventura-Hotels>

Anexo D. Evidencia de la Página Web Desarrollada

La página web desarrollada funciona como un producto aplicado del proyecto, al presentar de manera accesible los resultados derivados del análisis de sentimientos y la visualización de información asociada a los hoteles analizados en Buenaventura.

Enlace: <https://buenaventuradatos.com>