



**Propuesta de un modelo de inteligencia de negocios para la toma de decisiones estratégica en la gestión de un destino turístico inteligente.**

JUAN SEBASTIÁN BLANDÓN GÓMEZ

Universidad Ean

Facultad de Ingeniería

Maestría en Inteligencia de Negocios

Bogotá D.C., Colombia

19/11/2025

**Propuesta de un modelo de inteligencia de negocios para la toma de decisiones estratégica en la gestión de un destino turístico inteligente.**

**JUAN SEBISTIÁN BLANDÓN GÓMEZ**

Trabajo de grado presentado como requisito para optar al título de:

**Magister en Inteligencia de negocios**

Director (a):

FAJARDO MORENO WILLIAM STIVE

Modalidad:

**Monografía**

Universidad Ean

Facultad de Ingeniería

Maestría en Inteligencia de Negocios

Bogotá D.C., Colombia

19/11/2025

Nota de aceptación:

---

---

---

---

---

---

Firma del jurado

---

Firma del jurado

---

Firma del director del trabajo de grado

Bogotá, 19/11/2025

A mi mamá, cuyo amor ha sido mi refugio,  
y a papá, quienes me enseñaron a  
encontrar mi fuerza y mi propósito en  
Dios.

"Porque yo sé los planes que tengo para  
ustedes —afirma el Señor—, planes de  
bienestar y no de calamidad, a fin de  
darles un futuro y una esperanza."

Jeremías 29:11

## Resumen

El turismo constituye uno de los sectores económicos más dinámicos a nivel global y el principal motor del Archipiélago de San Andrés, cuya creciente dependencia de esta actividad plantea desafíos en torno a la gestión del destino. La ausencia de sistemas integrados de información limita la capacidad de las instituciones locales para planificar con base en datos, lo que evidencia la necesidad de modelos innovadores basados en datos abiertos y colaborativos que fortalezcan la toma de decisiones.

El propósito de esta investigación fue proponer un modelo de inteligencia de negocios para apoyar la gestión estratégica de un destino turístico inteligente en San Andrés Isla. En función de este objetivo, la primera parte del documento presenta la fundamentación teórica sobre ambos ejes temáticos; la segunda describe la metodología de investigación, el enfoque mixto y las fases del desarrollo; la tercera parte expone el diseño e implementación del pipeline tecnológico; y la última parte aborda los resultados, conclusiones y líneas futuras de investigación.

Los resultados evidenciaron la aplicabilidad de diversos KPI como el Score de Reputación Digital, el Índice de Sentimiento Turístico y el Net Promoter Score; confirmaron percepciones mayoritariamente positivas sobre el destino y mostraron la relevancia de un amplio número de alojamientos pequeños dentro de la oferta. Se concluye que el modelo es viable y replicable, ofreciendo indicadores estratégicos que fortalecen la gestión turística y abren líneas futuras orientadas a integrar métricas sociales, ambientales y de gobernanza.

**Palabras clave:** Smart Tourism Destination, Business Intelligence, Sentiment Analysis, San Andrés Island, Data Analysis.

### **Abstract**

Tourism is one of the most dynamic economic sectors worldwide and the primary driver of the Archipelago of San Andrés, where increasing reliance on this activity poses significant challenges for destination management. The absence of integrated information systems limits the capacity of local institutions to plan based on data, highlighting the need for innovative models grounded in open and collaborative data that strengthen decision-making processes.

The purpose of this research was to propose a business intelligence model to support the strategic management of a smart tourism destination in San Andrés Island. In line with this objective, the first part of the document presents the theoretical foundation on both thematic axes; the second part describes the research methodology, including the mixed approach and development phases; the third part outlines the design and implementation of the technological pipeline; and the final part discusses the results, conclusions, and future research directions.

The findings demonstrate the applicability of several key performance indicators (KPIs), including the Digital Reputation Score, the Tourism Sentiment Index, and the Net Promoter Score. Results also confirm predominantly positive perceptions of the destination and highlight the central role of small-scale accommodation within the local tourism offer. The study concludes that the proposed model is both viable and replicable, providing strategic indicators that enhance destination management. It further identifies opportunities for future research, particularly the integration of social, environmental, and governance metrics into the broader framework of smart tourism destinations.

**Keywords:** Destino turístico inteligente, Inteligencia de negocios, Análisis de sentimiento, San Andrés Isla, Análisis de datos.

## Contenido

	<b>Pág.</b>
<b>1. Introducción .....</b>	<b>12</b>
<b>2. Objetivos .....</b>	<b>17</b>
2.1 <i>Objetivo general</i> .....	17
2.2 <i>Objetivos específicos</i> .....	17
<b>3. Justificación .....</b>	<b>18</b>
<b>4. Marco Teórico .....</b>	<b>20</b>
4.1 <i>Destino Turístico Inteligente (Smart Tourism Destination)</i> .....	20
4.2 <i>Inteligencia de Negocios y Big Data</i> .....	22
4.3 <i>Big Data</i> .....	27
4.4 <i>Técnicas de análisis avanzado para el análisis de datos en turismo</i> .....	30
<b>5 Hipótesis.....</b>	<b>34</b>
<b>6 Metodología .....</b>	<b>35</b>
6.1 <i>Tipo de investigación</i> .....	35
6.2 <i>Fases de investigación</i> .....	36
6.3 <i>Desafíos metodológicos y gestión de limitaciones</i> .....	39
6.4 <i>Población y muestra</i> .....	40
6.5 <i>Técnicas para el análisis de datos y herramientas</i> .....	40

**7 Revisión sistemática de literatura sobre Inteligencia de negocios y destinos turísticos inteligente.**

**42**

7.1	<i>Planificación</i> .....	43
7.2	<i>Búsqueda de literatura científica y resultados</i> .....	44
7.3	<i>Selección de artículos relevantes</i> .....	48
7.4	<i>Evaluación de calidad</i> .....	50
7.5	<i>Resultados</i> .....	50
7.5.1	<i>Metodologías predominantes</i> .....	51
7.5.2	<i>Variables frecuentes</i> .....	53
7.5.3	<i>Síntesis de resultados</i> .....	55
7.5.4	<i>Hallazgos relevantes</i> .....	56
7.5.5	<i>KPIs identificados</i> .....	57
7.5.6	<i>Tecnologías aplicadas</i> .....	58

**8 Desarrollo del modelo de inteligencia de negocios para la gestión turística inteligente en San**

**Andrés Isla..... 62**

8.1	<i>Enfoque metodológico propuesto</i> .....	63
8.1.1	<i>Variables del modelo</i> .....	64
8.1.2	<i>KPIs del modelo</i> .....	64
8.1.3	<i>Tecnologías utilizadas en el desarrollo</i> .....	69
8.2	<i>Diseño de la arquitectura del modelo BI</i> .....	71
8.2.1	<i>Componentes del modelo</i> .....	71

Propuesta de un modelo de inteligencia de negocios para la toma de decisiones estratégica en la gestión de un destino turístico inteligente.	10
8.2.1.1 <i>Capa de extracción de datos</i> .....	71
8.2.1.2 <i>Capa de procesamiento e integración (ETL): Flujo con Airflow</i> .....	75
8.2.1.2.1 <i>Zona Raw</i> .....	76
8.2.1.2.2 <i>Zona Staging</i> .....	76
8.2.1.2.3 <i>Zona Data WareHouse</i> .....	77
8.2.1.3 <i>Capa semántica y de análisis</i> .....	80
8.2.1.3.1 <i>Análisis de Clustering</i> .....	81
8.2.1.3.2 <i>Procesamiento de lenguaje natural NLP y análisis de sentimiento</i> .....	82
8.2.1.4 <i>Capa de visualización</i> .....	84
8.2.1.4.1 <i>Diagrama general del modelo</i> .....	85
8.3 <i>Implementación y resultados del modelo</i> .....	87
8.3.1 <i>Implementación operativa del modelo en Power BI</i> .....	87
8.3.2 <i>Resultados</i> .....	89
8.3.2.1 <i>Alojamientos</i> .....	89
8.3.2.2 <i>Atracciones turísticas</i> .....	92
8.3.2.3 <i>Satisfacción del Usuario</i> .....	93
8.3.2.4 <i>Percepción del Destino</i> .....	96
8.2.1 <i>El dashboard como herramienta para la inteligencia estratégica del destino</i> .....	98
8.2.2 <i>Contribución del modelo a la gestión de un destino turístico inteligente</i> .....	99
8.2.2.1 <i>Gobernanza y Gestión eficiente</i> .....	100
8.2.2.2 <i>Innovación tecnológica</i> .....	100

Propuesta de un modelo de inteligencia de negocios para la toma de decisiones estratégica en la gestión de un destino turístico inteligente.	11
8.2.2.3 <i>Sostenibilidad y planificación territorial</i> .....	101
8.2.2.4 <i>Orientación al visitante y experiencia del turista</i> .....	101
<b>9    Validación del modelo mediante el instrumento V de Aiken</b> .....	<b>103</b>
<b>10    Discusión</b> .....	<b>106</b>
10.1 <i>Aportes del modelo a la literatura</i> .....	111
10.2 <i>Limitaciones</i> .....	113
<b>11    Conclusiones y Trabajo Futuro</b> .....	<b>116</b>
<b>Referencias</b> .....	<b>120</b>

## Lista de Figuras

	<b>Pág.</b>
Figura 1 - Distribución de publicaciones por año .....	45
Figura 2 - Concurrencia de Palabras Clave de investigación .....	47
Figura 3 - Flujo principal ScraperApp .....	73
Figura 4 - Pipeline Modelo BI .....	75
Figura 5 - Modelo de entidad relación .....	79
Figura 6 - Diagrama General del Modelo .....	86
Figura 7 - Dashboard general .....	88
Figura 8 - Dashboard Alojamientos Por Zona Geográfica .....	90
Figura 9 - Dashboard Alojamientos por Tipo Alojamiento .....	92
Figura 10 - Dashboard Atracciones Turísticas .....	93
Figura 11 - Dashboard Satisfacción Usuario .....	95
Figura 12 - Dashboard Percepción Destino .....	97

### Lista de Tablas

	<b>Pág.</b>
Tabla 1 Oferta de servicios turísticos.....	65
Tabla 2 Percepción del destino turístico:.....	67
Tabla 3 - Síntesis por serie de variables V de Aiken .....	104

## 1. Introducción

El turismo global ha enfrentado numerosos retos y transformaciones en los últimos años, especialmente a raíz de la pandemia de Covid-19, que alteró significativamente la movilidad y, por consiguiente, en la demanda de destinos turísticos a nivel mundial. A principios de 2020, el sector turístico experimentó una caída abrupta debido a las restricciones de movilidad y medidas de confinamiento impuestas por diversos países. Sin embargo, recientemente el sector turístico ha mostrado una notable recuperación. Según el reporte de World Tourism Organization, 2024, el turismo internacional ha tenido una recuperación del 88% comparado con los niveles previos a la pandemia, con 1,286 millones de turistas internacionales registrados en 2023, teniendo un aumento con respecto al año anterior por encima del 34% (2024, p. 1). En el caso de Colombia, el mismo informe señala un crecimiento de más del 33% en llegadas internacionales en comparación con 2019 (World Tourism Organization, 2024).

Para impulsar esta recuperación económica, el nuevo turismo pospandemia, ha requerido la implementación de estrategias de transformación digital y sostenibles, enfocadas en la promoción de prácticas responsables que responde a las nuevas expectativas y comportamientos de los viajeros (Rodríguez Jiménez & Martínez Martínez, 2022, p. 135) No obstante, como se indica en Tauber & Bausch, estos cambios hacia prácticas más sostenibles no siempre se sostienen en el largo plazo, debido a las fuertes presiones económicas de los agentes participantes del mercado turístico, cuyo enfoque sigue priorizando la maximización del uso de recursos existentes por encima del desarrollo de modelos sostenibles (Tauber & Bausch, 2022, p. 21).

En este contexto, la gestión de destinos turísticos inteligentes y sostenibles cobra una alta prioridad, ya que permite rediseñar destinos bajo un enfoque centrado en la creación y generación de valor más allá del volumen masivo de visitantes (Rodríguez Jiménez & Cristóbal Martínez Martínez, n.d., p. 129). Esto pone de manifiesto el debate entre los modelos turísticos masivos y los enfoques alternativos más sostenibles, que convergen dentro del “concepto paraguas” de turismo inteligente y responsable. Por lo tanto, resulta necesario ampliar la discusión sobre cómo una gestión estratégica basada en datos puede contribuir a un desarrollo turístico más sostenible y resiliente en un escenario pospandemia, involucrando a empresas e instituciones del sector para garantizar la generación de valor compartido y bienestar para todos los actores intervinientes, tal como lo argumenta (Rodríguez et al., 2020) citado en (Rodríguez Jiménez & Cristóbal Martínez Martínez, n.d., p. 76).

El Archipiélago de San Andrés es uno de los departamentos con mayor dependencia económica del turismo en Colombia, con una participación significativa en el PIB regional y en la generación de empleo local. Se estima que más del 50% de la población ocupada trabaja en actividades relacionadas directamente al turismo (James Cruz, 2013). Este protagonismo económico ha venido acompañado de una expansión sostenida del sector turístico, alcanzando en el 2022 más de 1,3 millones de turistas, algo más de 3.800 visitantes diarios (James et al., 2024). Pasando de 780 empresas de alojamiento y comidas en el año 2016, a un total de 1326 establecimientos de alojamiento y comidas en el 2024, con una tasa de crecimiento promedio anual del 7%. (Cámara de Comercio de San Andrés, 2024)

A pesar del crecimiento continuo y el potencial de la industria, el modelo actual de desarrollo turístico masivo del departamento enfrenta importantes desafíos estructurales. La ausencia de un enfoque integral y sostenible ha dado lugar a problemáticas críticas, tales como la alta concentración de turistas, falta de equidad en la distribución de la riqueza generada y un impacto negativo sobre la calidad de vida de la comunidad residente. Además, el deterioro ambiental derivado del turismo masivo representa una amenaza crítica para los recursos naturales del archipiélago y expone la necesidad urgente de políticas y nuevas prácticas responsables en la gestión del destino.

El plan de desarrollo territorial 2024-2027 de la Gobernación de San Andrés no cuenta con un sistema estructurado e integral de monitoreo que permita evaluar de manera sistemática de indicadores clave de gestión turística inteligente (Gobernación Departamento Archipiélago de San Andrés, 2024). Esta falta de datos oficiales y confiables limita la capacidad de las autoridades locales para implementar políticas basadas en datos, afectando la competitividad del destino en un entorno turístico cada vez más dinámico y exigente.

Ante estos desafíos, la aplicación de herramientas de Inteligencia de Negocios (BI) surge como una solución clave para transformar la gestión turística. Estas herramientas permiten establecer sistemas de información que integran datos históricos y proyectan escenarios futuros, facilitando así una toma de decisiones informada y basada en evidencia. (Rodríguez Jiménez & Cristóbal Martínez Martínez, 2022). El uso de BI, ofrece la posibilidad de recolectar datos de plataformas colaborativas, estos datos pueden proporcionar información actualizada y confiable sobre las experiencias y

expectativas de los usuarios, contribuyendo a la creación de indicadores claves de desempeño (KPI) alineados con las necesidades reales del destino.

En este sentido, las herramientas de Inteligencia de Negocios (BI) emergen como un recurso clave para la recolección, análisis y proyección de datos, permitiendo a las autoridades y organizaciones del sector turístico tomar decisiones más informadas y responsables. A partir de esta necesidad, surge la siguiente pregunta orientadora de la investigación: ¿Cómo puede el uso de herramientas de Inteligencia de Negocios (BI), contribuir a la generación de un modelo que permita fundamentar la toma de decisiones informadas y promueva la gestión inteligente del destino turístico en el Archipiélago de San Andrés?

El alcance de este estudio abarca desde la identificación de fuentes de información hasta la validación de un modelo de inteligencia de negocios que permita fundamentar la toma de decisiones informadas para la gestión turística inteligente del Archipiélago de San Andrés. No se contempla la implementación operativa del modelo dentro de las instituciones locales, sino que se presenta como una propuesta conceptual para futuras aplicaciones.

Las principales limitaciones del estudio se centran en la dependencia de datos externos de plataformas abiertas y colaborativas, cuya disponibilidad y calidad puede variar. Asimismo, se considera posibles restricciones legales y éticas en el acceso y uso de estos datos. El estudio no plantea fuentes de información internas de actores del sector turístico, y se resalta el entorno altamente dinámico y cambiante que pueden

afectar las preferencias de los viajeros a corto, mediano y largo plazo, influyendo en la relevancia de los datos recopilados.

Finalmente, el documento está estructurado en cuatro partes. En la primera parte se presenta la fundamentación teórica, que incluye la revisión sistemática de literatura y el marco conceptual sobre destinos turísticos inteligentes e inteligencia de negocios. La segunda parte describe la metodología de investigación, detallando el enfoque, las fases de desarrollo, las técnicas de análisis de datos y las herramientas tecnológicas empleadas. La tercera parte aborda los resultados obtenidos, exponiendo el diseño del modelo de inteligencia de negocios, los hallazgos derivados del análisis de los datos y la validación realizada. Por último, la cuarta parte presenta las conclusiones y recomendaciones, donde se responden las hipótesis planteadas, se discuten los aportes a la literatura y se proponen líneas de trabajo futuro orientadas a fortalecer la gestión de destinos turísticos inteligentes.

## **2. Objetivos**

### **2.1 Objetivo general**

Proponer un modelo de inteligencia de negocios para el Archipiélago de San Andrés que permita la toma de decisiones estratégicas para la gestión de un destino turístico inteligente.

### **2.2 Objetivos específicos**

- Analizar en la literatura científica las variables y metodologías empleadas en modelos previos de inteligencia de negocios aplicadas al sector turismo y destinos turísticos inteligentes.
- Desarrollar un modelo de inteligencia de negocios para la toma de decisiones estratégicas en la gestión de un destino turístico inteligente.
- Validar que el modelo de inteligencia de negocios permita la toma de decisiones informadas a los principales actores del sector y promueva la gestión de un destino turístico inteligente.

### **3. Justificación**

La presente investigación es de alta relevancia debido a la creciente urgencia de desarrollar un destino turístico inteligente, que logre un proceso de toma de decisiones del destino y su desarrollo turístico basado en datos, especialmente en un destino altamente dependiente del sector; como el Archipiélago de San Andrés, donde la creciente presión sobre los recursos naturales ha llevado a afrontar retos críticos relacionados con la capacidad de carga turística, la falta de control en la expansión hotelera y el impacto negativo en la comunidad local y el medio ambiente.

Actualmente, la gestión turística del Archipiélago enfrenta limitaciones derivadas de la falta de datos públicos y de calidad, dificultando la implementación de estrategias sostenibles, lo que perpetúa problemas como la saturación de la infraestructura, la insatisfacción de los turistas y el deterioro del entorno natural. En este contexto, la adopción de herramientas de BI se presenta como una solución clave, permitiendo consolidar información de múltiples fuentes para optimizar la toma de decisiones basadas en datos y mejorar la sostenibilidad del destino.

El Auge de la revolución digital y el acceso masivo a la información han permitido un entorno donde los usuarios puedan mantenerse siempre conectados y comunicados (Cruz-Chóez, 2022). Como lo indica Bustamante (2019), la web 2.0 ha transformado a los usuarios no solo en consumidores de información, sino también en co-creadores y distribuidores de contenido. Esta clase de contenidos creados por los usuarios (User-generated Content, UGC) es cada vez más percibido como confiables, creíbles,

relevantes, actualizados y atractivos; influyendo significativamente en las decisiones de otros usuarios a lo hora de reservar o alquilar un bien y/o servicios en plataformas digitales (Bustamante, 2019, p. 4).

El modelo propuesto en esta investigación busca integrar datos colaborativos con herramientas de análisis avanzado, como clustering, minería de texto y análisis de sentimientos, para generar KPI's que respalden la toma de decisiones informadas. Esta propuesta también contribuirá significativamente a la literatura académica y la práctica profesional, proporcionando un modelo replicable de gestión turística basado en inteligencia de negocios, que podría ser aplicado a otros destinos con características similares. El enfoque propuesto se alinea con las tendencias globales del turismo inteligente y sostenibles, fortaleciendo la competitividad del Departamento, permitiendo a los actores del sector aplicar estrategias comerciales más eficientes y alineadas con las necesidades del destino. Al igual, que permita contribuir a una distribución más equitativa de los beneficios que genera esta actividad económica, facilitando la formulación de políticas públicas que promuevas un desarrollo sostenible a largo plazo.

En conclusión, este estudio es pertinente y necesario, ya que responde a una problemática crítica para el desarrollo sostenible de San Andrés, proponiendo una solución innovadora que busca optimizar la gestión del turismo en San Andrés, alineándose con los principios de responsabilidad que demanda el sector, basándose en un modelo de inteligencia de negocios que permitirá obtener resultados claves para el destino turístico, el sector de servicios de alojamiento y aproximarnos a los patrones de comportamiento de los turistas que visitan el destino.

#### **4. Marco Teórico**

##### **4.1 Destino Turístico Inteligente (Smart Tourism Destination)**

El concepto de “Destino turístico inteligente” es un término aún emergente en la literatura, vinculado directamente con el desarrollo del concepto de “Ciudades inteligentes” (Gretzel, Sigala, et al., 2015, p. 186). Según Del Chiappa & Baggio, 2015, aún persisten debates sobre su definición y conceptualización, pero en esencia, estos destinos se configuran como un sistema en red que integran los actores claves del sector turísticos mediante una infraestructura tecnológica orientada a la cooperación, intercambio de conocimiento y la innovación (Del Chiappa & Baggio, 2015; Marine-Roig & Anton Clavé, 2015).

Por su parte, el concepto de ciudades inteligentes se fundamenta en cómo las tecnologías emergentes pueden mejorar y potenciar la generación de valor, especialmente para los habitantes de las ciudades (Zeng et al., 2020), definiéndose como espacios que generan nuevas formas de entender los problemas urbanos, coordinando nuevas tecnologías digitales que permiten el diseño de modelos y métodos para utilizar datos en la resolución de desafíos críticos relacionados con movilidad, energía y gobernanza urbana, al tiempo que identifican riesgos e incertidumbres (Del Chiappa & Baggio, 2015).

De esta forma se puede establecer que las ciudades inteligentes se centran en optimizar procesos enfocados en los ciudadanos, mientras que los destinos turísticos inteligentes integran digitalmente la vida de los habitantes y los ecosistemas de servicios

turísticos, con el propósito de mejorar la calidad de las experiencias de los turistas a través del uso de las Tecnologías de la Información y la Comunicación (TIC's). Esta integración se aborda desde dos perspectivas complementarias: la demanda, que se enfoca en las necesidades y expectativas de los turistas, y la oferta, que prioriza la capacidad del destino para ofrecer servicios eficientes y adecuados. (Marine-Roig & Anton Clavé, 2015c; Zeng et al., 2020)

Autores como Jankova et al. definen el concepto de destino turístico inteligente como “un ecosistema complejo y dinámico en el que los componentes físicos y digitales están estructurados y estrechamente relacionados y evolucionan como un todo.” (2023, p. 508). En esencia, un destino turístico inteligente se fundamenta en las interconexiones que se dan entre el destino turístico y sus múltiples grupos de interés, mediante el uso de plataformas que fomentan el flujo de comunicación y aumento de conocimiento (Del Vecchio & Passiante, 2017, p. 164). Estas características fortalecen la capacidad del destino para aprender, adaptarse y evolucionar, posibilitando un diseño y una gestión fundamentada en sistemas de información (Del Chiappa & Baggio, 2015, p. 146; Marine-Roig & Anton Clavé, 2015c).

Desde esta perspectiva, los destinos turísticos inteligentes, tal como las ciudades inteligentes, se fundamentan en el uso avanzado de aplicaciones e infraestructuras TIC's (Gretzel, Werthner, et al., 2015; Marine-Roig & Anton Clavé, 2015c). Tecnologías como el Internet de las Cosas (IoT), el cloud computing, servicios de internet para usuario final, aplicaciones de Big Data, aplicaciones móviles y el Internet 4.0 G son elementos claves para los objetivos del turismo inteligente (Gretzel, Sigala, et al., 2015, p. 182; Jankova et al., 2023, p. 508; Wang et al., 2013, p. 60).

De esta forma, la construcción de un destino turístico inteligente implica la conformación de una infraestructura física con un correspondiente nivel de progreso tecnológico (Jankova et al., 2023, p. 508), por lo cual, las colaboraciones publico privadas son determinantes para la conformación de este ecosistema interconectado donde los elementos físicos y digitales influyen directamente en la mejora de los procesos de toma de decisiones, que así mismo, implica la proyección de un crecimiento económico y un desarrollo sostenible (Del Chiappa & Baggio, 2015; Díaz-González et al., 2022; Marine-Roig & Anton Clavé, 2015; Zeng et al., 2020).

Esta visión holística del sector turístico, orientada por datos, implica en la generación de políticas públicas capaces de construir mejores condiciones de vida sobre sus habitantes, derivando así un nuevo modelo de planificación del destino inteligente e integrado (Biderman et al., 2021). Esta integración digital no solo optimiza la asignación de recursos turísticos de manera eficiente para el destino, sino que también promueve una distribución equitativa de los beneficios entre las comunidades locales (Del Chiappa & Baggio, 2015; Jankova et al., 2023; Wang et al., 2013; Zeng et al., 2020, p. 508).

#### **4.2 Inteligencia de Negocios y Big Data**

El concepto de inteligencia de negocios o BI por sus siglas en inglés (Business Intelligent), fue introducido en el año 1989 Howard Dresner como un concepto sombrilla (Bustamante, 2019, p. 11) que combina conceptos como arquitectura, bases de datos, herramientas de analítica, metodologías y aplicaciones que buscan mejorar los procesos de toma de decisiones empresariales mediante el análisis de los datos, para

transformarlo en conocimiento útil en la optimización de los procesos externos e internos de la compañía. (Vajirakachorn & Chongwatpol, 2017, p. 76).

Existen diferentes conceptos y enfoques entorno a la inteligencia de negocios, como lo indica Bustamante, 2019, algunos conceptos giran en torno a la capacidad tecnológica o herramientas que dan soporte a todo el proceso entorno a la extracción, transformación y carga de los datos, ETL, por sus siglas en inglés; mientras que otros conceptos giran en torno a una perspectiva estratégica o administrativa, donde se resalta la capacidad de análisis de las fuentes de información para revelar hallazgos estratégicos. (2019, p. 13)

A nivel de arquitectura, la inteligencia de negocios se compone de cuatro componentes principales, según la revisión de literatura realizada, los cuales son:

**Identificación de las fuentes de datos y preparación:** En este primer componente, se encuentran los procesos de recopilación de las fuentes de bases de datos a analizar sobre las cuales se encuentran relacionadas los procesos de negocio, siendo fuentes internas o externas; seguido, encontramos el proceso de ETL (Extraction Transformation and Load, por sus siglas en inglés), el cual consiste en la recolección, almacenamiento e integración de los datos estructurados, semiestructurados o no estructurados, para su posterior análisis y que el usuario final pueda verlos o manipularlos de acuerdo a un rango de tiempo. (Bustamante, 2019, p. 24; Fuchs et al., 2014, p. 199; Höpken & Fuchs, 2021, p. 9; Vajirakachorn & Chongwatpol, 2017, p. 76)

Los datos estructurados son todos aquellos que tienen un formato definido, se localizan en un campo fijo de un registro o archivo específico, los datos

semiestructurados no tienen formatos fijos, pero si contienen etiquetas, coordenadas y otros marcadores, mientras que los datos no estructurados son todos aquellos datos sin tipos definidos, almacenados en documentos u objetivos, los cuales no pueden ser fácilmente indexados en bases de datos estructuradas (Joyanes, 2019, Chapter 5, p. 8)

**Integración y almacenamiento para el análisis de los datos (Data Warehouse y Modelos de datos):** En este segundo componente se encuentran ubicados los procesos de estructuración de datos necesarias tanto para el procesamiento como para la integración de los datos, estos datos son alojados en un repositorio central que reúnen múltiples datos de diversas fuentes de primarias y secundarias, que se conocen popularmente como Data Warehouse o almacén de datos. Estos almacenes de datos se definen como “Sistemas de repositorio central de datos integrados de una o múltiples fuentes” (Mariani et al., 2018, p. 3158), por lo cual facilitan la toma de decisiones y la planeación, ya que consolidan los datos de los procesos estratégicos de las organizaciones, para ello, es importante que los datos sean orientados a temas específicos de la organización, sean integrados en formatos consistentes bajo una misma unidad de medida, sean variantes en la línea del tiempo y que no sean volátiles, es decir, que no se puedan modificar una vez sea capturado el dato. (Bustamante, 2019, p. 28; Höpken & Fuchs, 2021)

La literatura resalta dos principales modelos de Data Warehouse, siendo el modelo normalizado propuesto por Inmon (2002) el que se fundamenta en bases de datos que contiene tablas relacionadas entre ellas y que no contienen ninguna redundancia (Höpken & Fuchs, 2021, p. 13); mientras que el segundo modelo se conoce como Modelo dimensional, el cual se fundamenta entre la relación entre hechos y dimensiones, los

cuales se estructuran en esquemas conocidas como modelo estrella, conformado por una o más tablas de dimensiones y solo una tabla de hechos, estas se relacionan por medio de columnas clave, manteniendo un diseño simple y fácil de entender para el usuario; estos modelos son comúnmente conocidas como bases de datos SQL (Bustamante, 2019, p. 31; Höpken & Fuchs, 2021, p. 17).

En la actualidad, se ha hecho extensivo el uso de almacenes de datos geoespaciales, semánticos y en tiempo real, de tal forma en que se establecen bases de datos estructuradas y no estructuradas, SQL y No-SQL, respectivamente, estas últimas recopilan información de fuentes de datos no estructurados, como reseñas, comentarios y reacciones en redes sociales, imágenes, audios y videos. (Bustamante, 2019, p. 31; Fuchs et al., 2014, p. 200)

### **Gestión del desempeño (Procesamiento analítico en línea y Minería de Datos):**

En este tercer componente, encontramos todos los procesos entorno al análisis, monitoreo, reporte y procesamiento de los datos y métricas claves de la compañía. Uno de los principales tipos de análisis, es el procesamiento analítico en línea o OLAP por sus siglas en inglés, este se caracteriza por permitir un análisis multidimensional de la información, que permite realizar consultas y reportes de forma más ágil mediante una figura metafórica como el cubo, los cuales son conjuntos de hechos y dimensiones visualizados desde diferentes niveles de agregación y perspectivas, llevando a cabo análisis mucho más complejos (Bustamante, 2019, p. 35).

Otro de los elementos claves dentro de este componente, es la minería de datos, la cual es una técnica estadística y de aprendizaje automático para identificar tendencias,

modelos y patrones en diferentes fuentes de datos de interés para el usuario final. Dentro de estas técnicas encontramos análisis de clasificación como redes neuronales artificiales (ANN, por sus siglas en inglés), árboles de decisión, reglas de asociación, inducción de reglas y algoritmos de K vecinos más cercanos; técnicas de estimación; técnicas de predicción, mediante estadísticas multivariadas como regresión y análisis de series temporales; técnicas de clustering, como el algoritmo K-means, que agrupa elementos similares dentro de un conjunto de datos y finalmente técnicas de asociación de reglas como el Market Basket Analysis. (Fuchs et al., 2014, p. 199)

**Visualización de datos:** En este último componente, encontramos la realización de dashboards interactivos o cuadros de mando para la visualización de datos que permiten interactuar con los datos de forma intuitiva, facilitando la comprensión e interpretación de las métricas de seguimiento y datos relevantes, en múltiples pantallas visuales permitiendo soportar los procesos de toma de decisiones organizacionales de una manera mucho más fácil (Vajirakachorn & Chongwatpol, 2017, p. 76).

Para la visualización de las variables utilizan diferentes formas de representación de la información como tablas, gráficos de barras, gráficos de torta, histogramas, mosaicos, diagramas de caja y bigotes son las más representativas, las cuales dependen especialmente del tipo y número de variables, a la complejidad de la estructura de datos y a su tipo y escala, si son variables cualitativas o cuantitativas. En esta última etapa, es muy importante garantizar visualizaciones que permitan el despliegue y uso de los datos para la toma de decisiones, transfiriendo el conocimiento de forma organizada, simple y fácil comprensión para no expertos. (Bustamante, 2019, p. 41)

### 4.3 Big Data

El término Big Data surge en la literatura de gestión administrativa en el año 2001, cuando Doug Laney estudió los efectos del emergente e-commerce y sistematizó la forma de gestionar los datos a niveles más básicos para las organizaciones (Laney, 2001). En su trabajo, Laney definió tres características principales de Big Data, denominadas las “3Vs”: Volumen, que se refiere a la dimensión de los datos, el sistema en los cuales se almacena y se guardan los datos; Velocidad, que implica la rapidez en la generación, modificación y transferencia de los datos; y Variedad, que abarca los diferentes formatos de datos, desde estructurados a no estructurados (Mariani et al., 2021, p. 2978). A partir de esta primera conceptualización, el término emergente Big Data ha evolucionado y se han añadido nuevas características. Fosso Wamba et al. define el Big Data como “un enfoque holístico para gestionar, procesar y analizar 5Vs (i.e., Volumen, Variedad, Velocidad, Veracidad y Valor)” (2015, p. 235). En esta ampliación, la Veracidad hace referencia a la confiabilidad de los datos, mientras que el Valor se relaciona con el proceso de extracción y análisis que permiten generar información de valor. (Mariani & Fosso Wamba, 2020, p. 339)

El Big Data ha transformado significativamente los procesos organizacionales, permitiendo mejorar los procesos de creación, entrega y captura de valor, Además, ha optimizado la toma de decisiones y los procesos operativos internos, incrementando así la capacidad de medición del desempeño organizacional, el rendimiento en innovación y generando una ventaja competitiva (Fosso Wamba et al., 2015; M. Mariani & Fosso Wamba, 2020; M. Mariani et al., 2021). Estos procesos son posibles gracias al incremento en volumen de la información, bien desde datos obtenidos rápidamente

desde redes sociales, la web, multimedia o desde diferentes dispositivos conectados a Internet of Things (IoT), que permite realizar la captura de grandes volúmenes de datos, ya sean estructurados o no estructurados. De esta forma, Hashem et al. Propone tres principales aspectos de lo que sería Big Data, el primero corresponde a los grandes volúmenes de datos, el segundo aspecto hace referencia a que estos datos no pueden ser categorizados dentro de bases de datos relacionales y el tercero, que los procesos de generación, captura y procesamiento se dan de forma rápida. (Hashem et al., 2015)

Sin embargo, es importante resaltar que el almacenamiento de grandes volúmenes de información en múltiples formatos no garantiza, por sí solo, la generación de conocimiento en la organización. La clave radica en la capacidad analítica de estos amplios volúmenes de datos, que permiten descubrir patrones de comportamiento complejos y correlacionados. (Mariani & Baggio, 2022, p. 233). Adicionalmente, es importante aclarar, lo que hoy en día consideramos como grandes volúmenes de datos, quizás no lo serán dentro de unos años, por lo cual, podemos entender el Big Data como un elemento dentro de lo que es la Inteligencia de Negocios. (Bustamante, 2019, p. 18)

En esta línea, Sun et al. (2014) define que el análisis de datos como “una ciencia y tecnología que consiste en analizar, examinar, resumir, adquirir inteligencia y sacar conclusiones de datos para aprender, describir y predecir algo” (Sun, Strang, & Yearwood, 2014; Gandomi & Haider, 2015, citado en Sun, 2019, p. 3). El análisis de datos puede clasificarse en cuatro categorías principales de técnicas, agrupados de acuerdo con su dimensión en el tiempo:

**Análisis Descriptivo:** se basa en la estadística descriptiva para la medición de información tanto del pasado como del presente.

**Análisis Exploratorio:** utiliza técnicas de agrupamiento (clustering) y análisis factorial para explicar fenómenos observados en el pasado y el presente.

**Análisis predictivo:** que emplea técnicas de regresión y proyección con el objetivo de comprender que podría suceder en el futuro.

**Análisis prescriptivo:** mediante técnicas de optimización y experimentación buscan generar conocimiento y recomendaciones para el futuro.

Mientras que los dos primeros tipos de técnicas de análisis están orientados a mejorar el desempeño y la eficiencia en los procesos (pasado y presente), los dos últimos tipos de analítica se basan en ser soporte para la innovación, procesos de reingeniería y exploración del conocimiento (futuro). (Mariani & Baggio, 2022, p. 234)

En este sentido, la Analítica en Big Data se puede entender como una herramienta clave para mejorar la inteligencia de negocios dentro de una compañía (Mariani et al., 2021, p. 2979). No obstante, para que las organizaciones logren mejoras significativas en su desempeño, se requiere de un conjunto de capacidades específicas (Mariani & Fosso Wamba, 2020). Estas capacidades están relacionadas con las competencias organizacionales en torno a tres tipos de recursos: tangibles (tecnología y datos), intangibles (como la cultura organizacional orientada a los datos), y talento humano (como las habilidades blandas y gerenciales) (Mariani & Baggio, 2022, p. 234). De esta forma, Big Data acompañado de la analítica se convierten en elementos de valor ilimitado para cualquier industria, empresa y área de tecnología. (Sun, 2019, p. 3)

#### **4.4 Técnicas de análisis avanzado para el análisis de datos en turismo**

En la Industria de la Hotelería y Turismo (H&T), la inteligencia de negocios y el Big Data han sido impulsores clave de la transformación digital en el servicio, proporcionando una ventaja competitiva temporal (Mariani et al., 2021, p. 2978). Las empresas del sector, los destinos turísticos y los usuarios consumen grandes volúmenes de datos y utilizan cada vez más el análisis de estos datos para mejorar los procesos de decisión en todos los niveles (Mariani & Borghi, 2019, citado en Mariani & Baggio, 2022, p. 234). Estos datos son captados por diversas fuentes de datos, desde los datos generados por dispositivos, como los datos de los servicios móviles o celulares, datos de WIFI y GPS, los datos de transacciones como las reservas y búsquedas en la Web, y finalmente, los datos generados por los usuarios conocidos como user-generated content (UGC) tales como comentarios, fotos e interacciones en redes sociales. (Mariani et al., 2021, p. 2978)

Dentro del extenso, complejo y diverso ecosistema digital del sector turístico y hotelero, las redes sociales representan una de las principales fuentes de información basadas en comunidades, donde encontramos una variedad de plataformas de reseñas en línea que van desde las basadas en comunidades como TripAdvisor, hasta las agencia de viajes en línea basadas en transacciones (Online Travel Agencies, OTAS, por sus siglas en inglés), las cuales tiene como un mismo objetivo, asistir la toma de decisiones de los consumidores basados en un conocimiento social compartido y de confianza, mediante los datos generados por los usuarios (UGC) Xiang et al., 2017, p. 52).

Estos datos permiten obtener un entendimiento más profundo de la experiencia de los consumidores finales, medir el engagement digital y evaluar el nivel de servicio y la

satisfacción de los huéspedes o usuarios (Mariani et al., 2021; Mariani & Baggio, 2022).

De esta forma, nos permiten identificar las características del destino que se transmiten a través de las imágenes cargadas en las diferentes plataformas por usuarios online, convirtiéndose en un marco referente como principal fuente para el análisis de la reputación, marca e imagen de los destinos turísticos (Marine-Roig & Anton Clavé, 2015, p. 164).

Autores como Xiang et al., (2017) identifican el crecimiento en la literatura científica sobre el impacto de las reseñas en línea en la reputación y la imagen de los establecimientos de hospedaje, concentrándose en el contenido generado por los usuarios. De hecho, estas fuentes de datos son las predominantes en las investigaciones del sector, representando un 47% del total, seguido de los datos generados por los dispositivos, con un 36%, y los datos de transacciones, con un 17%. (Li et al., 2018; citado en Mariani & Baggio, 2022).

Sin embargo, las investigaciones que utilizan datos de reseñas en línea deben ser conscientes de los matices de éstas fuentes de información, ya que cada plataforma corresponde a diferentes modelos de negocio, que utilizan diferentes tecnologías y van dirigidas a diferentes públicos objetivo, generando un complejo sistema socio cultural en cada una de ellas; por lo cual, es importante considerar la representatividad y la calidad de los datos bajo un diseño metodológico estructurado para la toma de decisiones (Xiang et al., 2017, p. 52).

El uso del Big Data y la minería de texto en la industria hotelera y turística se hace cada vez más frecuente en múltiples estudios de análisis de UGC, ya que principalmente

nos permite disminuir el tiempo de procesamiento, aumenta el tamaño de la muestra y facilita evitar sesgos del investigador (Marine-Roig & Anton Clavé, 2015, p. 163). Esta técnica se refiere al proceso de extracción de conocimientos y patrones interesantes y no triviales sobre datos no estructurados como textos o documentos (Kim et al., 2017). De esta forma, un importante número de estudios abordan como la aplicación de esta técnica en diferentes plataformas se han centrado en entender las motivaciones de viaje, evaluar los impactos de las reseñas sobre el desempeño de hoteles y servicios, al igual que, las opiniones y sentimientos reflejados en la experiencia de los viajeros sobre los servicios y productos turísticos (Ma et al., 2018, p. 121).

Estas aplicaciones generan un espacio de alta interactividad dentro del ecosistema turístico digital, permitiendo de esta forma a los actores claves del sector mejorar la calidad de sus servicios, ofreciendo los servicios correctos de acuerdo a las preferencias de los turistas en un tiempo correcto, promoviendo una gestión del destino inteligente (Kim et al., 2017, p. 363).

Paralelamente, encontramos la aplicación de técnicas de Clustering en el sector turístico, que se usan principalmente para la segmentación de turistas dentro de subgrupos basados en sus motivaciones de viaje (Vajirakachorn & Chongwatpol, 2017, p. 76), esta técnica consiste en algoritmos que permiten encontrar los patrones de los consumidores o clientes, mediante análisis de k medias, de tal forma en que aquellos asignados al mismo grupo (clusters) sean más similares entre sí que los patrones que los consumidores o clientes en los otros grupos (Pons-Vives et al., 2022, p. 5). Permitiendo poder categorizar establecimientos de alojamientos y servicios turísticos en función de

criterios como precio, ubicación, calidad del servicios, percepción e incluso datos no estructurados como fotos y videos (Santana, 2019).

Un elemento clave en este proceso, es la aplicación de técnicas como el Web Scraping, que permiten principalmente la extracción automatizada de datos, desde plataformas digitales y bases de datos en línea, básicamente consiste en la transformación de datos web a datos estructurados que pueden ser almacenados en repositorios, bases de datos o cualquier otro formato, esta técnica aprovecha la estructuración de las páginas web usando el lenguaje HTML para acceder a la información. De esta forma, se han utilizado estas técnicas para recopilar información sobre precios, disponibilidad, comentarios de usuarios y clasificaciones de diferentes plataformas (Bustamante, 2019; Santana, 2019).

La aplicación de estas técnicas en la industria turística ha facilitado la creación de sistemas de monitorio en tiempo real, optimizando estrategias de marketing, planificación de tarifas y gestión de la capacidad hotelera mediante la consolidación de sistemas de inteligencia de negocios para su posterior análisis (Bustamante, 2019).

## **5 Hipótesis**

Existe una alta concentración de servicios de alojamiento en el Archipiélago de San Andrés, destacándose los establecimientos turísticos de menor capacidad o tamaño.

El nivel de satisfacción en los establecimientos turísticos de menor capacidad o tamaño es relativamente alto.

El nivel de satisfacción sobre las principales atracciones turísticas del Archipiélago es relativamente alto.

Las percepciones sobre el destino turístico y los servicios de alojamiento de San Andrés son mayoritariamente positivas.

## **6 Metodología**

### **6.1 Tipo de investigación**

El propósito de esta investigación se centra en establecer un modelo de inteligencia de negocios aplicada para la gestión de un destino turístico inteligente. Dado al carácter multidimensional del objeto de estudio, se propone desarrollar una investigación con un enfoque mixto, combinando métodos cuantitativos y cualitativos para abordar de manera integral la gestión del destino turístico inteligente en el Archipiélago de San Andrés. Este enfoque permite integrar datos objetivos, como el número establecimientos de hospedajes y destinos turísticos, como datos subjetivos, derivados de las percepciones y experiencias de los visitantes en plataformas colaborativas.

En términos metodológicos, la investigación es de tipo exploratoria y descriptiva. Se considera exploratoria porque busca identificar relaciones y patrones, en los datos turísticos, y descriptiva porque permite caracterizar las variables claves del estudio, proporcionando una visión estructurada del destino turístico. Es importante resaltar que la investigación se enmarca en un diseño no experimental y es de corte transversal, lo que significa que los datos se recolectarán en un único momento del tiempo sin intervención directa de las variables estudiadas. Este diseño permite explorar el contexto y sentar las bases para futuras investigaciones que profundicen en la gestión de destinos turísticos mediante el uso de inteligencia de negocios.

Para el desarrollo del modelo de inteligencia de negocios, se implementará una revisión sistemática de literatura y estudios de caso previos sobre la aplicación de

inteligencias de negocios en el sector turismo. A partir de esta revisión, se definirán las variables clave y se establecerá la metodología para su extracción y análisis. Se emplearán técnicas de web scraping para la recopilación de datos de plataformas como TripAdvisor y Airbnb, seleccionadas según su volumen de datos, su representatividad en la industria y su acceso a la información en tiempo real. Los datos extraídos serán procesados y analizados mediante técnicas estadísticas, que permitirán identificar patrones de comportamiento, segmentar establecimientos turísticos y evaluar la percepción del destino a través del análisis de sentimiento.

## **6.2 Fases de investigación**

### **6.2.1 Fase 1: Revisión bibliográfica y análisis de antecedentes:**

Esta primera fase corresponde a la revisión sistemática de la literatura, cuyo objetivo es identificar referentes teóricos que permitan la construcción de una base teórica conceptual sólida en los ejes temáticos principales: destinos turísticos inteligentes, inteligencia de negocios, minería de datos, Big data y análisis de datos colaborativos. Para ello, se llevará a cabo una búsqueda estructurada en la base de datos Web of Science, utilizando una ecuación de búsqueda previamente definida, garantizando un enfoque riguroso y replicable.

Esta revisión sistemática permitirá tener como resultado un análisis de concurrencia de palabras clave y conceptos fundamentales para el desarrollo del marco teórico. Además, posibilitará identificar estudios previos que han aplicado modelos de inteligencia de negocios para destinos turísticos inteligentes, analizando sus metodologías, variables

y resultados, teniendo como resultado la infraestructura tecnológica y metodológica para la extracción, procesamiento y análisis de los datos. Esto facilitará el diseño de un modelo ajustado al contexto del Archipiélago de San Andrés.

## **6.2.2 Fase 2: Desarrollo del modelo de inteligencia de negocios**

Esta fase tiene como propósito determinar y desarrollar del modelo propuesto, desplegando los procesos de extracción, procesamiento, almacenamiento, análisis y visualización de los datos, orquestados bajo una arquitectura tecnológica de inteligencia de negocios. Su formulación se fundamenta en la revisión de literatura expuesta previamente con el objetivo de garantizar la rigurosidad científica y la aplicabilidad del modelo en un contexto real. A continuación, se describen las subfases de esta etapa, estructuradas de acuerdo con el enfoque metodológico propuesto en Joyanes (2019, p. 9)

### **6.2.2.1 Extracción de datos mediante Web Scraping**

Para la recopilación de los datos, se desarrollará un algoritmo especializado de Web Scraping para la recolección automatizada de datos relevante sobre establecimientos de hospedaje y destinos turísticos provenientes de plataformas colaborativas como TripAdvisor y Airbnb. Se utilizarán librerías de Python como Selenium y BeautifulSoup, que permitirán interactuar dinámicamente con páginas web, simular navegaciones de usuarios y extraer los diferentes tipos de datos

### **6.2.2.2 Proceso de limpieza, transformación y carga de los datos**

Una vez obtenido los datos brutos (raw), se iniciaría el proceso de extracción, transformación y carga de los datos en el data warehouse del proyecto. Este proceso será orquestado mediante el uso del framework Apache Airflow, que permite automatizar y programar los flujos de trabajo, facilitando la ejecución de tareas para la extracción de datos relevantes y significativos que se almacenarán temporalmente en la zona de staging, la aplicación de diferentes transformaciones que garanticen la calidad y limpieza de los datos acorde a lo lógica del proyecto, y finalmente el almacenamiento definitivo de los datos a nuestro data warehouse (Joyanes, 2019; Singh, 2019).

#### **6.2.2.3 Diseño de modelo de bases de datos**

Paralelo al proceso de transformación y carga de los datos, se deberá realizar el diseño del esquema lógico del modelo de bases de datos del proyecto, estructurado bajo el enfoque de un modelo dimensional tipo estrella, creando las diferentes dimensiones y la tabla de hechos que deberán corresponder a la necesidad de estructurar la información de manera que facilite su procesamiento y análisis correctamente.

#### **6.2.2.4 Análisis de los datos**

Una vez se tengan los datos dentro del data Warehouse, se aplicarán técnicas de análisis cuantitativo y cualitativo. Para la segmentación de los establecimientos turísticos, se empleará un análisis de Clúster que permitirá agrupar establecimientos según características comunes. Asimismo, se llevará a cabo un análisis de tendencias para identificar patrones de comportamiento en el mercado turístico. Para el análisis de percepción de los turistas, se realizará una minería de texto sobre los comentarios

obtenidos, para evaluar el nivel de satisfacción de los visitantes. Se desarrollará un análisis de sentimientos, clasificando las reseñas en positivas, negativas y neutras.

#### **6.2.2.5 Visualización de resultados**

Para presentar los hallazgos de la investigación, se desarrollará un dashboard interactivo en Power BI, que permitirá visualizar los principales indicadores clave de desempeño (KPIs) propias para la gestión de un destino turístico inteligente. Este dashboard empleará gráficos dinámicos y mapas interactivos, proporcionando una herramienta de apoyo para la toma de decisiones.

#### **6.2.3 Fase 3: Validación del modelo de inteligencia de negocios**

La tercera fase se centrará en la validación del modelo obtenido en la fase anterior. Para esto, se presentará ante un panel de expertos del sector quienes evaluarán su aplicabilidad, fiabilidad y utilidad para la toma de decisiones estratégicas en el desarrollo de un destino turístico inteligente. A partir de los comentarios recibidos, se presentarán los principales ajustes al modelo para optimizar su rendimiento y aplicabilidad en la gestión turística del Archipiélago de San Andrés.

#### **6.3 Desafíos metodológicos y gestión de limitaciones**

Uno de los principales desafíos de esta investigación radica en la calidad y representatividad de los datos obtenidos de las plataformas colaborativas. Dado que el

contenido generado por los usuarios puede contener sesgos y opiniones subjetivas, se aplicarán filtros de calidad de datos y estrategias de normalización para minimizar el impacto de la información redundante o poco confiable.

Asimismo, se contemplarán posibles limitaciones relacionadas con la accesibilidad de los datos y restricciones en el Web Scraping de ciertas plataformas. Para mitigar estos problemas, se utilizarán diferentes estrategias que nos permita bajo un marco ético la extracción de los datos.

#### **6.4 Población y muestra**

La población objetivo incluye todos los alojamientos turísticos registrados en plataformas como TripAdvisor y Airbnb en el Archipiélago de San Andrés. La muestra será seleccionada mediante un muestreo intencional, enfatizado en la calidad de los datos, de esta forma se considerará aquellos hospedajes con mayor cantidad de reseñas y valoraciones recientes. Esto permitirá garantizar un análisis representativo de la oferta turística del destino.

De acuerdo con el último reporte de la cámara de comercio, en el 2024 el Archipiélago contaba con 1.326 empresas en la rama de actividades económicas propias de alojamientos y comidas (Cámara de Comercio de San Andrés, 2024). Por lo cual se estima tener un tamaño muestral de aproximadamente 299 establecimientos de hospedaje para la revisión de las reseñas, teniendo en cuenta un nivel de confianza del 95% y un margen de error del 5%.

#### **6.5 Técnicas para el análisis de datos y herramientas**

**6.5.1 Análisis cuantitativo:**

- Estadísticas descriptivas para identificar patrones en los datos recolectados.
- Análisis clúster para segmentar hospedajes según sus características.

**6.5.2 Análisis cualitativo:**

- Análisis de sentimientos para evaluar percepciones positivas y negativas.

**6.5.3 Visualización de datos:**

- Diseño de dashboards interactivos que permitan visualizar KPIs y segmentaciones.

Para la ejecución de estos análisis, se utilizarán herramientas especializadas como Python (pandas), Postgres y MongoDB para almacenamiento de datos, Apache Airflow para la orquestación del flujo de datos y Power BI para la visualización de resultados.

## **7 Revisión sistemática de literatura sobre Inteligencia de negocios y destinos turísticos inteligente.**

La revisión sistemática de la literatura se llevó a cabo con el propósito de identificar las principales investigaciones relacionadas con la gestión turística inteligente basada en inteligencia de negocios y Big Data, proporcionando una base teórica sólida para el desarrollo metodológico de la investigación. Este proceso adicionalmente nos permitió la identificación de tendencias clave en la aplicación de tecnologías de análisis de datos en el sector turístico, al igual que, también sirvió como punto de partida para la formulación de las variables y metodologías iniciales del modelo de inteligencia de negocios propuesto.

Para garantizar un enfoque riguroso y replicable, la revisión sistemática se estructuró en seis etapas metodológicas: (1) planificación, (2) búsqueda, (3) selección, (4) evaluación de calidad, (5) extracción de datos y (6) resultados. Cada una de estas etapas fue diseñada para maximizar la exhaustividad, pertinencia y validez de los estudios analizados, asegurando así la integración de evidencia científica de alto impacto en el diseño del modelo.

A partir de esta revisión, se estableció un marco metodológico fundamentado en evidencia empírica que permitió identificar los enfoques predominantes, las herramientas más utilizadas y las brechas de conocimiento en la intersección entre inteligencia de negocios, Big Data y turismo inteligente. Este proceso ha sido clave para orientar la construcción del modelo de inteligencia de negocios, estableciendo las bases conceptuales y metodológicas necesarias para su desarrollo y posterior implementación.

## 7.1 Planificación

En esta etapa, se establecieron los protocolos metodológicos de la revisión, se definieron las siguientes preguntas de investigación que guiaron el proceso de búsqueda, revisión y análisis de la literatura.

1. ¿Cuáles son los principales enfoques metodológicos utilizados en la aplicación de inteligencia de negocios en destinos turísticos inteligentes?
2. ¿Cuáles son las aplicaciones más relevantes de la inteligencia de negocios en el desarrollo de un destino turístico inteligente?
3. ¿Qué técnicas de minería de datos y análisis de contenido generado por usuarios han sido utilizadas en estudios previos?
4. ¿Cuáles son los principales retos y oportunidades identificados en la literatura sobre la aplicación de herramientas de inteligencia de negocios para el análisis de contenido en plataformas colaborativas?

Para garantizar la calidad de la revisión, se seleccionó la base de datos Web of Science (WoS), caracterizada por ser una de las principales bases de datos de suscripción pagada disponibles para el análisis de citas, abarcando publicaciones académicas indexadas en los campos de Ciencias, Ciencias sociales, Artes y Humanidades; estos índices de citación agrupan literatura científica a nivel regional y en diferentes campos de estudio, que junto con la herramienta para análisis y evaluación, como el *Journal Citation Report*, consolidan una alta calidad científica para

investigación (Waltman, 2016). Adicional, es importante resaltar que la revisión asume un enfoque de bola de nieve, para garantizar que los criterios se ajustaran y poder abordar un mayor número de documentos potenciales, sin dejar por fuera del análisis publicaciones que pudieran aportar al desarrollo del modelo.

## **7.2 Búsqueda de literatura científica y resultados**

La definición de la ecuación de búsqueda es un elemento fundamental para garantizar un proceso de captura de estudios de alto impacto en el campo objeto de investigación. Esta ecuación se despliega con base en las preguntas orientadores de la fase de planificación y palabras claves producto de una revisión manual, constituida a partir de operadores y expresiones, que de manera iterativa se fueron ajustando para garantizar una búsqueda amplia en cantidad de documentos. Se emplearon términos clave en inglés debido a la predominancia de publicaciones en este idioma en revistas indexadas.

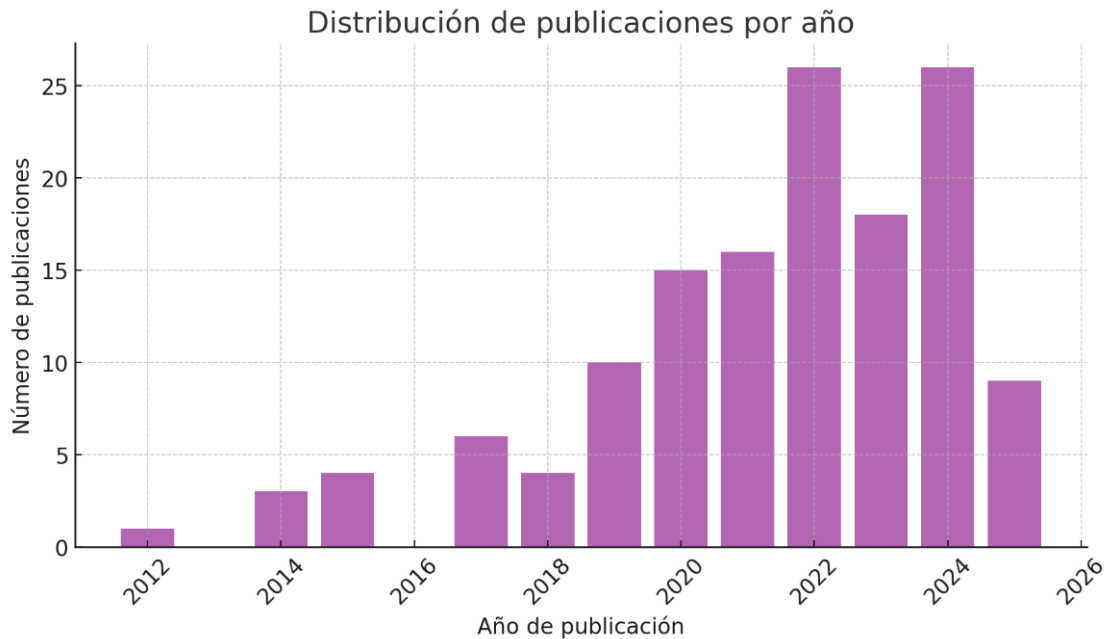
Business intelligence" (All Fields) and tourism (All Fields) or "business intelligence" (All Fields) and "smart Destination" (All Fields) or "big data" (All Fields) and "smart Destination" (All Fields)

Sumado a esta ecuación de búsqueda se aplicaron filtros avanzados de inclusión, restringiendo los resultados a artículos científicos revisados por pares ("article") publicados en revistas indexadas, excluyendo deliberadamente otros tipos de documentos que no aportan a la validez metodológica y la relevancia empírica de los estudios. De esta forma, se lograron obtener 143 publicaciones en la base de datos consultada, comprendidos entre los años 2012 y 2025. Con un total de 3,756 citas

totales, con un promedio de 26.27 citas por artículo, 3.613 citas si contar las auto citas, lo que refleja una influencia sobre estos estudios en la literatura científica y una contribución significativa al campo de estudio.

El periodo de búsqueda comprendido entre los años 2012 a 2025, se debe principalmente que partir del año 2012 se consolida la adopción de tecnologías emergentes en el turismo inteligente dentro de la literatura científica, definiendo este el punto inicial para el análisis preliminar de tendencias y metadatos. El año 2025 se incluye como límite superior dada la disponibilidad de publicaciones anticipados o aceptadas en línea para ese año en la base de datos Web of Science, garantizando una cobertura temporal actualizado y prospectiva.

**Figura 1 - Distribución de publicaciones por año**



*Fuente: Elaboración propia con base en los resultados de búsqueda en Web of Science*

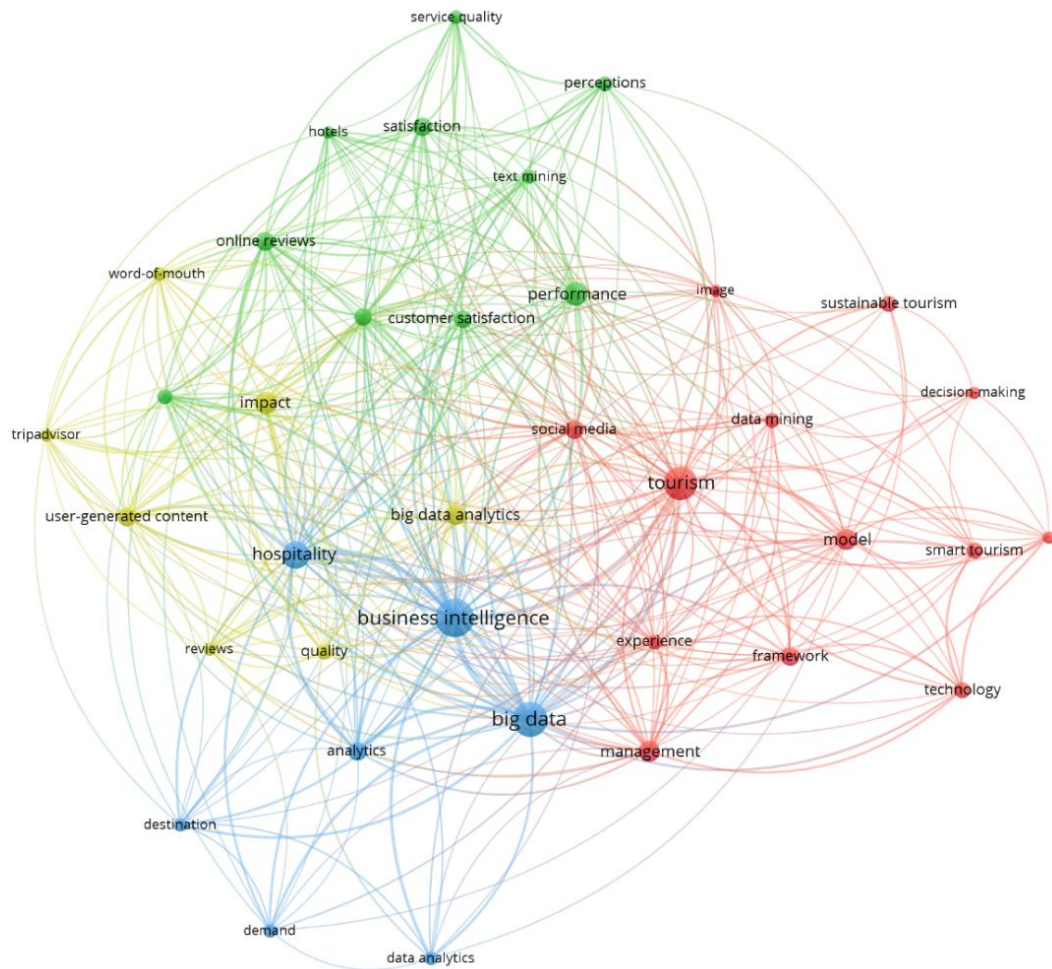
Esta búsqueda inicial nos permitió identificar la evolución reciente en la cantidad de publicaciones relacionada con la gestión turística basada en la inteligencia de negocios y Big Data. Reflejando la creciente adopción de tecnologías avanzadas para la gestión de un destino turístico. Se destaca los años 2022 y 2024, donde se consolidaron los años con mayor cantidad de publicaciones alcanzando 26 estudios publicados en cada año. Este comportamiento confirma que la inteligencia de negocios aplicada al turismo ha ganado relevancia en la última década, consolidándose como un campo de estudio en crecimiento.

Posteriormente, se exportaron todos los datos y citas de la búsqueda para su posterior análisis de concurrencia de palabras clave utilizando el software VOSviewer. Este enfoque nos permitió visualizar la red de relaciones entre los términos más utilizados en la literatura analizada y su nivel de conexión. Dicho análisis reveló una serie de conceptos fundamentales para el desarrollo del marco teórico, tales como “*business intelligence*”, “*Big Data*”, “*data mining*”, “*sentiment análisis*” y “*Smart tourism*”, ya que se constituyen como el núcleo central de la investigación.

Los resultados de la visualización de la concurrencia de palabras muestran la interrelación entre términos como “*business intelligence*” y “*Big Data*”, demostrando que representan áreas claves para la optimización de la gestión turística. A su vez, su conectividad con “*data mining*” y “*sentiment analysis*”, reflejan la integración de modelos predictivos y minería de datos en la toma de decisiones estratégicas en la gestión de destinos turísticos inteligentes y sostenibles. Estos conceptos no solo destacan la importancia de las TICs en la transformación de los destinos, sino también cómo el análisis de contenido generado por los usuarios (UGC) y las reseñas en línea, extraídas

mediante técnicas como el web scraping de plataformas colaborativas como “TripAdvisor”, proporcionando un enfoque cada vez más utilizado para evaluar la percepción del destino y la satisfacción del cliente. Así, el marco teórico de esta investigación se fundamenta en estas tendencias, abordando los principales retos y oportunidades que enfrentan los destinos turísticos al integrar la inteligencia de negocios y el análisis de Big Data en su gestión.

**Figura 2 - Concurrencia de Palabras Clave de investigación**



*Fuente: Elaboración propia generada en VOSviewer*

### 7.3 Selección de artículos relevantes

Una vez finalizada el proceso de búsqueda y recopilación de los artículos de investigación, se procedió a realizar un proceso para la selección de artículos relevantes, para garantizar la calidad científica, actualidad y pertinencia temática de los artículos seleccionados. Los 143 registros bibliográficos potencialmente relevantes fueron sometidos a un conjunto de criterios objetivos de selección basados en la clasificación de revistas indexadas en bases de datos internacionales y en indicadores bibliométricos reconocidos en el ámbito académico.

Para ejecutar estos criterios de calidad e indexación, se utilizó como eje de filtrado la indexación en Web of Science (WoS), particularmente a través de su colección principal (Clarivate, 2025), conformado por las siguientes:

- Social Science Citation Index (SSCI): colección de revistas especializadas de alta calidad en el ámbito de las ciencias sociales, incluyendo temas como turismo, economía, administración y sistemas de información.
- Emerging Sources Citation Index (ESCI): revistas emergentes en proceso de ser evaluadas para inclusión en índices superiores, no cuentan con factor de impacto inmediato, pero permiten visibilidad y acceso en comunidades académicas.
- Science Citation Index Expanded (SCIE): colección de revistas similares al SSCI, pero para revistas de ciencias exactas, física y de ingeniería.
- Arts & Humanities Citation Index (AHCI): colección de revistas enfocadas en disciplinas humanísticas.

De estas, se dio prioridad a los artículos indexados en SSCI, dado a ser el principal referente en investigaciones sociales con alto impacto y visibilidad internacional. Asimismo, se consideró el cuartil de impacto (Q1–Q4) asignado por Journal Citation Reports (JCR); en este sentido, se priorizaron aquellas publicaciones que se encontraban en el primer cuartil (Q1), es decir, dentro del 25% superior de las revistas más citadas en su área temática, lo que constituye un referente internacionalmente aceptado para medir la excelencia editorial y el impacto académico de una revista científica (Macias, 2016).

Posteriormente, para operacionalizar estos criterios, se desarrolló un procedimiento automatizado mediante lenguaje de programación Python, el cual permitió:

- Identificar si la revista de publicación estaba indexada en SSCI.
- Determinar si la categoría temática del artículo correspondía a áreas relevantes para el estudio, tales como tourism, hospitality, sustainability o data analytics.
- Calcular un índice compuesto de relevancia (Ranking\_Score) que integra la indexación, el cuartil, el número de citas y el año de la publicación.

Este procedimiento permitió generar un ranking de los artículos más influyentes metodológicamente sólidos. Como resultado principal, se seleccionaron inicialmente 49 artículos de alto impacto, todos ellos alineados temáticamente con los objetivos de esta investigación. Estos documentos constituyen la base empírica y conceptual para la fase de extracción de datos y síntesis teórica.

#### **7.4 Evaluación de calidad**

Con los documentos seleccionados, se procedió a realizar la evaluación de calidad de los 49 artículos identificados. Los textos revisados fueron sujetos a una valoración detallada de acuerdo con el cumplimiento de criterios metodológicos, como coherencia con el objeto de estudio y la aplicabilidad de los hallazgos, de tal forma que la inclusión de estos documentos se respaldará en un fundamento teórico y empíricos de la investigación.

Se analizaron aspectos como el diseño de la investigación, la claridad en la descripción de la metodología, la justificación de las técnicas de análisis, como ejemplo, minería de datos, análisis de sentimientos, *machine learning*; y la coherencia entre objetivos, resultados y conclusiones. Adicionalmente, se evaluó si cumplían con los ejes temáticos del estudio, se valoraron especialmente los artículos publicados entre 2018 y 2024, con el fin de asegurar la vigencia de los enfoques teóricos y tecnológicos utilizados.

Del total de los 49 artículos evaluados, 33 artículos cumplieron con los criterios para ser incluidos en el corpus que representa el estado de arte en el uso de inteligencia de negocios en la gestión de un destino turístico inteligente, con la cual se construyó el marco teórico y se fundamentaron las variables y enfoque metodológicos para el modelo propuesto.

#### **7.5 Resultados**

La etapa final se base en la recopilación de datos relevantes de los estudios de caso y revisiones sistemáticas de literatura de los 33 artículos, para responder las preguntas de interés de la revisión planteadas al inicio y realizar la síntesis de resultados obtenidos. De esta forma, se identificaron los documentos que realizan una revisión de la literatura aportando a la construcción del estado del arte de la presente investigación, mientras que para los estudios de caso se sintetizo la información en el Anexo 1 donde se recopila la referencia del caso de estudio, la metodología, los KPI's identificados, las variables y datos extraídos, el lugar de realización y su contexto geográfico; observaciones puntuales sobre el aporte del artículo y el uso de las tecnologías para el desarrollo de la investigación.

En este análisis se identificaron 17 estudios de caso, principalmente llevados a cabo en Estados Unidos, países europeos como España, Suecia, Dinamarca, Grecia, Reino Unido, Francia, entre otros; países asiáticos como China y Tailandia. Consolidando información clave sobre el uso de herramientas de inteligencia de negocios y *Big Data* en destinos turísticos inteligentes, lo que sirvió como insumo esencial para la formulación del modelo propuesto, tomando como base los principales resultados que se exponen a continuación:

### **7.5.1 Metodologías predominantes**

Los estudios revisados evidencia la aplicabilidad de diversas metodologías a lo largo de diferentes países, donde se destaca el enfoque predominante del uso de arquitecturas de inteligencia de negocios aplicado para transformar grandes volúmenes de datos externos e internos en conocimiento útil para un destino turístico inteligente. Autores

como Bustamante (2019), Kim et al. (2017), Marine-Roig, (2019), Santana, (2019) y Marine-Roig et al., 2019 han demostrado la eficiencia del uso de técnicas automatizadas como APIs, Web Scrapings y Web Crawlings para la extracción de datos de plataformas online, particularmente TripAdvisor, Airbnb, blogs turísticos y Twitter, dentro de la capa de extracción de datos en sus arquitecturas metodológicas propuestas.

Mientras que investigadores como (Vajirakachorn & Chongwatpol, 2017) Höpken et al. (2021) y Fuchs et al., 2014 han basado su metodología en la integración de datos internos provenientes de diversas fuentes como datos históricos, registros de navegación web, datos de reservas y encuestas en esta primera fase de la extracción e integración de los datos.

Seguido a la obtención de los datos, múltiples investigaciones se destacan por el uso de algoritmos de clustering, arboles de decisión, series temporales y redes neuronales en la fase de análisis de los datos (Bustamante, 2019; Fuchs et al., 2014; Marine-Roig, 2019). Mientras que otras investigaciones se destacan más por el análisis de contenido semántico, análisis de sentimientos multimodal, análisis de co-ocurrencia, Deep Learning, NLP, análisis de lexicones, tokenización y extracción de color-emoción en imágenes (Frigau et al., 2024; Kim et al., 2017; Köseoglu et al., 2020; Ma et al., 2018; Marine-Roig & Anton Clavé, 2015; Yannacopoulou & Kallinikos, 2025).

Investigadores como Marine-Roig & Anton Clavé (2015), Fuchs et al. (2014), Ma et al. (2018), Kim et al. (2017), Marine-Roig (2019), Vajirakachorn & Chongwatpol (2017), Bustamante (2019), Marine-Roig et al. (2019), Yannacopoulou & Kallinikos (2025) y Köseoglu et al. (2020) hacen uso explícito y sistémico de metodologías donde se usan la

minería de datos para analizar fenómenos complejos de análisis en el ámbito del turismo, con especial énfasis en la gestión de destinos inteligentes. Estos estudios utilizan la minería de datos y de texto para transformar grandes volúmenes de datos como calificaciones, reseñas, comentarios, imágenes y videos generados por los usuarios en indicadores claves para el destino.

En conclusión, estas metodologías han sido validadas en contextos internacionales, lo cual respalda su aplicabilidad en el caso de estudio del Archipiélago de San Andrés. Además, permiten un análisis tanto cuantitativo como cualitativo al destino turístico inteligente, destacándose por la aplicabilidad de arquitecturas híbridas de almacenamiento (SQL y NOSQL) que permiten estructurar y analizar grandes volúmenes de datos estructurados, semiestructurados y no estructurados como soporte para la toma de decisiones.

### **7.5.2 Variables frecuentes**

Las variables más frecuentes en la literatura revisada emergen del análisis de contenido generado por los usuarios (UGC) en plataformas colaborativas y bases de datos en fuentes abiertas, integrando información proveniente de diversas estructuras. Un gran número de estas investigaciones se apoyan en variables estructuradas y categóricas, tales como el nombre, ubicación, género y edad del usuario; así como en características específicas del establecimiento de hospedaje, incluyendo nombre, dirección, tipo de alojamiento, precio, número de baños y habitaciones. Del mismo modo, se identifican variables asociadas a los destinos turísticos, como nombre, dirección, nivel

de valoración promedio y por categorías, servicios ofrecidos, entre otras más (Bagherzadeh et al., 2021; Bustamante, 2019; Santana, 2019; Xiang et al., 2017).

Además, se observa una recurrencia en el uso de variables semiestructuradas, tales como la geolocalización de los establecimientos, reseñas y lugares turísticos (Bustamante, 2019; Cerdá-Mansilla et al., 2022; Santana, 2019); así como la fecha y estacionalidad de las reseñas publicadas (Marine-Roig, 2019; Yannacopoulou & Kallinikos, 2025). Otros elementos semiestructurados incluyen el idioma y nacionalidad del autor de la reseña (Marine-Roig & Anton Clavé, 2015; Yannacopoulou & Kallinikos, 2025), y las categorías asignadas a los lugares reseñados, como tipo de atracción o servicio (Kim et al., 2017; koseKöseoglu et al., 2020; Ma et al., 2018)

Paralelamente, estos mismos autores incorporan variables no estructuradas, siendo el principal insumo los textos completos de las reseñas de usuarios (Kim et al., 2017; Köseoglu et al., 2020; Ma et al., 2018; Marine-Roig, 2019). A partir de estos textos se extraen variables como los sentimientos asociados al texto, (Kim et al., 2017; Yannacopoulou & Kallinikos, 2025); el uso de palabras clave, tokens y frecuencia léxica (Bagherzadeh et al., 2021; Köseoglu et al., 2020; Marine-Roig & Anton Clavé, 2015); las emociones vinculadas a contenido visual (Marine-Roig et al., 2019; Yannacopoulou & Kallinikos, 2025), así como la polaridad y subjetividad del texto (Bagherzadeh et al., 2021; Frigau et al., 2024; Ma et al., 2018).

Finalmente, algunos autores complementan estos enfoques con el uso de bases de datos internas institucionales, en las que se emplean variables de tipo operativo como el número de reservas, duración de visitas, evaluación de satisfacción de usuarios (escalas

Likert) y número de huéspedes (Fuchs et al., 2014; Höpken et al., 2021; Vajirakachorn & Chongwatpol, 2017). Estas variables permiten enriquecer los modelos predictivos y analíticas desde una perspectiva organización complementaria a los datos abiertos y colaborativos.

### **7.5.3 Síntesis de resultados**

En síntesis, los hallazgos de esta revisión sistemática permiten identificar una madurez creciente en la literatura especializada sobre el uso de metodologías y variables de inteligencia de negocios para destinos turísticos inteligentes. Estos enfoques han resultado particularmente valiosos en contextos donde los datos oficiales son fragmentados o insuficientes, como ocurre en muchos destinos emergentes, principalmente porque posibilita una comprensión integral y basada en evidencia del comportamiento turístico. (Bustamante, 2019; Fuchs et al., 2014; Marine-Roig et al., 2019)

En cuanto a las metodologías, se valida una integración de enfoques automatizados y replicables, entre ellos la aplicación de técnicas de extracción de datos como web scraping, uso de algoritmos avanzados de minería de textos, aplicación de técnicas de análisis semánticos, análisis de sentimientos y clustering. En cuanto a las variables, se identificaron tres grandes categorías recurrentes en la literatura: (i) estructuradas, tales como características demográficas de los usuarios y atributos de los establecimientos y del destino, (ii) semiestructuradas como la geolocalización, estacionalidades, entre otras, por último, (iii) no estructuradas, donde se encuentran las reseñas textuales, imágenes, videos o contenido generado por los usuarios (UGC). Esta integración de variables por

varios autores ha permitido justificar la existencia de modelos replicables y escalables para medir dimensiones claves en los destinos turísticos inteligentes.

#### **7.5.4 Hallazgos relevantes**

La revisión sistemática ha cumplido una función doble: por una parte, identificar enfoques teórico-metodológicos ya validados y aplicables a escenarios reales; y por otra, proporcionar la base empírica para el diseño conceptual y metodológico del modelo adaptado al Archipiélago de San Andrés.

A partir de los resultados sistematizados, emergen hallazgos relevantes que complementan el análisis anterior, especialmente en lo relativo a indicadores clave y tecnologías emergentes. Se destaca que el análisis de contenido generados por usuarios en las plataformas colaborativas como TripAdvisor, AirBnb y redes sociales son una fuente válida, robusta y empíricamente probada para comprender las percepciones de los destinos turísticos, identificar factores críticos de satisfacción y mapear patrones de experiencia del visitante (Kim et al., 2017; Santana, 2019; Yannacopoulou & Kallinikos, 2025).

Esta tendencia se refuerza con el uso extendido de técnicas avanzadas de minería de datos, tales como el análisis de sentimiento, el modelado de tópicos, la extracción de emociones visuales, la clasificación temática, la tokenización y el uso de redes semánticas ejecutados mediante herramientas especializadas y lenguajes de programación como Python, que permiten abordar de forma integral el diagnóstico de

imagen del destino y su posicionamiento digital. (Köseoglu et al., 2020; Ma et al., 2018; Marine-Roig & Anton Clavé, 2015, Frigau et al., 2024)

De manera destacada, esta revisión también permitió identificar una relación estrecha entre las variables analizadas, las metodologías aplicadas, la tecnología usada y los KPI utilizados, permitiendo establecer una articulación entre la naturaleza de los datos y los indicadores empleados para evaluar el desempeño de los destinos turísticos inteligentes. Esta conexión conceptual se presenta como un aporte novedoso de la presente investigación, ya que permite vincular de forma sistemática las variables con indicadores de valor estratégico.

#### **7.5.5 KPIs identificados**

Los estudios revisados evidencia un uso intensivo y estratégico de indicadores clave de desempeño (KPIs) orientados a medir la percepción, satisfacción, competitividad y eficiencia en la gestión de destinos turísticos. Estos indicadores emergen de la interacción entre datos estructurados, semiestructurados.

Autores como Bagherzadeh et al. (2021), Bustamante (2019) y Kim et al. (2017) emplean KPIs derivados del análisis de sentimientos y frecuencia temática, tales como la polaridad de los comentarios, la frecuencia de advertencia o menciones negativas y la clasificación automática entre turistas y residentes. Por su parte, Xiang et al. (2017), Marine-Roig (2019), Marine-Roig et al. (2019) y Marine-Roig & Anton Clavé (2015) incorporan indicadores como la valoración promedio de atractivos, la longitud de la reseña y la frecuencia semántica de tópicos recurrentes.

Otros estudios, como los de Fuchs et al. (2014) y Vajirakachorn & Chongwatpol (2017), integran indicadores operativos y predictivos como el número de reservas, duración promedio de estadía, intención de regreso del turista y la eficacia de campañas promocionales. En investigaciones más recientes, Ma et al. (2018) y Yannacopoulou & Kallinikos (2025) proponen modelos multimodales con KPIs como la precisión y utilidad percibida de las reseñas, el F1-score del modelo, la interacción generada por imágenes y la frecuencia de emociones detectadas por color.

Finalmente, Köseoglu et al. (2020) destacan KPIs como la frecuencia de términos clave, la centralidad semántica (grado y betweenness), y los mapas de ventaja competitiva construidos mediante análisis de redes. En conjunto, estos indicadores permiten la construcción de sistemas avanzados de monitoreo del desempeño turístico, adaptables al contexto local de destinos como San Andrés, donde se requiere medir tanto la experiencia del visitante como el posicionamiento competitivo del destino en canales digitales.

### **7.5.6 Tecnologías aplicadas**

Dentro del corpus de la literatura revisada, se destaca el uso de tecnologías emergentes en el desarrollo de investigaciones orientadas a la aplicabilidad de la inteligencia de negocios en el desarrollo de destinos turísticos inteligentes. Estas tecnologías han sido aplicadas principalmente para el proceso de extracción, procesamiento, análisis y visualización de grandes volúmenes de datos, con el objetivo de construir modelos predictivos, sistemas de monitoreo y dashboards como

herramientas principales de apoyo a la toma de decisiones en destinos turísticos inteligentes.

Investigadores como Bustamante (2019) Cerdá-Mansilla et al. (2022), Santana (2019) y Xiang et al. (2017) se destacan por el uso de lenguajes de programación como Python, para la recolección masiva de datos, tecnologías como Selenium, BeautifulSoup, Scrapy y API REST fueron ampliamente utilizados para automatizar los procesos de extracción de información pública.

En términos de análisis textual, estudios como los de Köseoglu et al. (2020), Ma et al. (2018) y Marine-Roig (2019) incorporaron técnicas avanzadas de procesamiento de lenguaje natural (NLP), a través de librerías como TextBlob, NLTK, spaCy, VADER y SentiWordNet. Otros trabajos, como los de Yannacopoulou & Kallinikos (2025) combinaron el análisis textual con herramientas de visión por computadora y clasificación emocional de imágenes, empleando redes neuronales convolucionales (CNN) para el análisis semántico de fotografías y detección de emociones mediante ruedas de color y modelos cromáticos. Más recientemente, Frigau et al. (2024) implementaron un modelo multilingüe preentrenado basado en BERT ( Siglas de Bidirectional Encoder Representations from Transformers), a través de la librería HuggingFace, destacando sus ventajas en precisión contextual, robustez en múltiples idiomas y aplicabilidad al análisis de textos generados por usuarios.

Autores como (Bustamante, 2019; Cerdá-Mansilla et al., 2022; Fuchs et al., 2014; Höpken et al., 2021; Ma et al., 2018; Santana, 2019; Vajirakachorn & Chongwatpol, 2017;

Xiang et al., 2017) destacaron el uso de herramientas de data mining, mediante redes neuronales artificiales (ANN), árboles de decisión, clustering K-means, machine learning y series temporales, en entornos de programación como R y Python para el desarrollo de modelos predictivos y tableros de control.

En cuanto a la representación visual y análisis de relaciones semánticas, investigaciones como las de Köseoglu et al. (2020) y Marine-Roig & Anton Clavé (2015) hicieron uso de herramientas de análisis de redes semánticas y co-ocurrencia de términos, utilizando Gephi, VOSviewer y librerías de visualización en Python (Matplotlib, Seaborn, Plotly) para generar mapas de inteligencia y estructuras de tópicos. Autores como Bustamante (2019) incorporaron el uso de OpenStreetMap como fuente base para el análisis de georreferenciación y visualización espacial de reseñas, alojamientos y puntos de interés turístico. Asimismo, se emplearon tecnologías de almacenamiento y consulta estructurada como PostgreSQL, combinadas con plataformas de integración como Google Colab, Jupyter Notebook, Power BI y Tableau para la implementación de tableros interactivos.

En síntesis, el objetivo específico de analizar las variables y metodologías utilizadas en modelos previos de inteligencia de negocios aplicados al turismo se cumple ampliamente a través de la identificación de tres categorías clave de variables (estructuradas, semiestructuradas y no estructuradas), y de un amplio repertorio metodológico que abarca desde técnicas de extracción automatizada (web scraping, APIs), hasta enfoques de minería de datos, análisis semántico y procesamiento de lenguaje natural (NLP).

De forma complementaria, emergen como hallazgos relevantes no previstos en el objetivo inicial la identificación sistemática de KPIs utilizados en estudios previos y la clasificación de tecnologías aplicadas, lo cual enriquece la comprensión del campo de estudio. Esta revisión permite, además, establecer una relación coherente entre variables analizadas y KPIs reportados, lo que ofrece una base sólida para el desarrollo del modelo propuesto. Estos resultados permitirán justificar la estructuración de KPIs, mediante una relación conceptual y metodológica entre los datos disponibles, las tecnologías empleadas y los indicadores estratégicos necesarios para la gestión de destinos turísticos inteligentes.

## **8 Desarrollo del modelo de inteligencia de negocios para la gestión turística inteligente en San Andrés Isla**

Con base en los hallazgos de la revisión sistemática de literatura presentada en el capítulo anterior, este capítulo desarrolla la implementación técnica del modelo de inteligencia de negocios diseñado para apoyar la toma de decisiones estratégicas en la gestión del destino turístico inteligente objeto de estudio: el Archipiélago de San Andrés Isla. Dicho modelo se fundamenta en las metodologías, variables y tecnologías identificadas como pertinentes y validadas por la literatura científica especializada, las cuales han sido adaptadas al contexto local para garantizar su aplicabilidad y relevancia.

La construcción del modelo se desarrolla de manera sistemática, articulando los procesos de captura, transformación, análisis y visualización de datos provenientes de plataformas colaborativas. Estos procesos están soportados por una arquitectura tecnológica híbrida (estructurada y no estructurada), que combina herramientas de programación, bases de datos, técnicas de análisis de sentimientos, minería de datos y visualización de indicadores clave de desempeño (KPIs).

Esta fase tiene como objetivo central operacionalizar los componentes conceptuales definidos anteriormente, determinando cómo se implementan en un entorno real las variables clave, las técnicas analíticas y las tecnologías emergentes, con la finalidad de captar las percepciones de los visitantes a través de los contenidos generados por usuarios (UGC).

### **8.1 Enfoque metodológico propuesto**

El diseño del modelo propuesto se sustenta en un enfoque metodológico mixto, con un carácter exploratorio, descriptivo y aplicado, el cual fue justificado desde la etapa inicial del proyecto y validado empíricamente a partir de los hallazgos de la revisión sistémica de literatura. Este enfoque permite integrar el análisis de variables cuantitativas y cualitativas para la extracción, procesamiento y análisis de datos relacionados con la percepción y el comportamiento de los visitantes en el destino turístico.

Esta aproximación permite una comprensión más holística del fenómeno turístico, articulando el análisis estadístico de variables estructuradas y semiestructuradas con la exploración semántica del contenido textual generado por los usuarios (UGC) integrado de esta forma perspectivas objetivas y subjetivas.

Es por ello por lo que se requiere una arquitectura de inteligencia de negocios específica con procesos sistematizados y orquestados desde la extracción automatizada de los datos mediante web scraping (Santana, 2019); pasando a la limpieza, transformación y almacenamiento de los datos, donde se integran bases de datos relacionales (PostgreSQL) para el manejo de información estructurada y geoespacial, y bases NoSQL (MongoDB) para el almacenamiento de datos no estructurados como lo son las reseñas textuales. (Fuchs et al., 2014; Höpken et al., 2021); seguido del análisis exploratorio, semántico y predictivo mediante clustering y NLP; y finalmente, la visualización mediante un dashboard con indicadores claves (KPIs). (Köseoglu et al., 2020; Ma et al., 2018)

### **8.1.1 Variables del modelo**

Con base en los hallazgos de la revisión de literatura, donde se evidencia como factor repetitivo a lo largo de los estudios más relevantes y replicables, la integración de múltiples niveles de datos, estructurados, semiestructurados y no estructurados (Bustamante, 2019; Marine-Roig et al., 2019; Fuchs et al., 2014), se definieron dos dimensiones estratégicas claves para desplegar las variables empleadas en el modelo:

- Oferta de servicios turísticos: vinculada a las características objetivas de los establecimientos de alojamiento, tales como ubicación, coordenadas, categoría, precios, volumen de reseñas y ranking relativo, datos que permiten analizar el comportamiento de la oferta y su distribución territorial. En esta dimensión se almacenarán las variables estructuradas y semiestructuradas
- Percepción del destino turístico: asociada a los sentimientos, opiniones, valoraciones y menciones expresadas por los visitantes, las cuales reflejan la reputación del destino. En esta dimensión se almacenarán las variables no estructuradas.

### **8.1.2 KPIs del modelo**

Cada una de las dimensiones de variables definidas serán operacionalizadas mediante un conjunto de indicadores clave de desempeño (KPIs), diseñados con base en los enfoques metodológicos, modelos replicables y buenas prácticas identificadas en estudios previos de igual forma. De esta manera, los KPIs propuestos no solo responden a una fundamentación teórica robusta, sino que también reflejan criterios de aplicabilidad práctica y adaptabilidad al contexto territorial del Archipiélago, lo que los convierte en

herramientas estratégicas para el monitoreo, evaluación y toma de decisiones en destinos turísticos inteligentes.

Los KPIs derivados para la oferta de servicios turísticos incluyen el Score de Reputación Digital (SRD), el precio promedio por categoría de alojamiento, la tasa de concentración de servicios turísticos por zona, la distribución y calificación por tipo de establecimiento. Estas métricas permiten caracterizar la distribución y desempeño de la oferta, y se han aplicado exitosamente en modelos de inteligencia turística para la evaluación comparativa entre destinos (Bustamante, 2019; Fuchs et al., 2014; Santana, 2019).

**Tabla 1 Oferta de servicios turísticos**

Variable	KPI	Definición conceptual	Definición operacional	Fórmula / Método propuesto	Fuente de datos
<b>Calificación del servicio turístico, número de reseñas, ranking</b>	SRD – Score de Reputación Digital	Indicador compuesto de reputación digital por servicio	Puntaje ponderado de reputación digital, combinando calificación promedio, volumen de reseñas y ranking relativo de cada servicio	$SRD = \alpha \cdot \text{Calificación promedio del establecimiento} + \beta \cdot \log(1 + \text{Número de Reseñas}) + \gamma \cdot (1 / \text{Ranking})$ <p><math>\alpha, \beta, \gamma =</math> pesos ajustados empíricamente según la literatura y</p>	TripAdvisor, Airbnb

				pruebas del modelo.	
<b>Precio por tipo de alojamiento</b>	Precio por tipo de alojamientos	Comparativo del precio x media geométrica por tipo de alojamiento	Promedio aritmético por tipo de alojamiento	Precio normalizado = media geométrica por tipo de alojamiento	TripAdvisor, Airbnb
<b>Ubicación y tipo de servicio</b>	Tasa de concentración de servicios por zona	Mide la densidad de servicios turísticos en el territorio	Cantidad de servicios turísticos por unidad de área	TCSZ: Conteo ponderado de servicios en la zona z / área efectiva de la zona (km <sup>2</sup> ) excluyendo superficies no utilizables (mar)	Datos georreferenciados
<b>Tipo de categoría de alojamiento</b>	Distribución de categorías	Proporción de alojamientos según su clasificación	Porcentaje de alojamientos por tipo/categoría	% Tipo Establecimiento = Número de alojamientos en la categoría x / Total establecimientos	TripAdvisor, Airbnb

Fuente: diseño realizado por el autor

Los KPIs derivados para la percepción del destino turístico integran distintas aproximaciones de análisis semántico y cuantitativo de las reseñas. En primer lugar, el

Índice de Sentimiento Turístico (IST), que mide la polaridad media de las opiniones en las reseñas clasificadas como positivas, negativas y medias, captando el sentimiento general hacia el destino. A su vez, el Net Promoter Score (NPS) constituye una medida proxy de la disposición de los visitantes a recomendar el destino, calculada a partir de la proporción de promotores, neutros y detractores. Complementariamente, el grado de confiabilidad del análisis de sentimiento del propio modelo BERT empleado, refleja la certeza estadística con el que este modelo asigna una reseña a una categoría. Finalmente la tasa de concentración geográfica del sentimiento asocia las emociones expresadas con zonas específicas del territorio, que permite mapear emociones asociadas a zonas específicas del destino. En conjunto, estos indicadores se apoyan en técnicas de análisis de sentimiento mediante NLP, validadas en investigaciones como las de Kim et al. (2017), Marine-Roig (2019) y Yannacopoulou & Kallinikos (2025).

**Tabla 2 Percepción del destino turístico:**

Variable	KPI	Definición conceptual	Definición operacional	Fórmula / Método propuesto	Fuente de datos
<b>Polaridad textual de reseñas</b>	IST – Índice de Sentimiento Turístico	Medida del sentimiento general hacia el destino percibido por los usuarios	Agrega valores de polaridad de opiniones clasificadas como positivas, negativos o neutrales.	NLP: Media ponderada de polaridad NLP por contexto	TripAdvisor

<b>Calificación cuantitativa por reseña</b>	NPS – Net Prometer Score	Medida proxy de disposición a recomendar el destino a partir de reseñas	Se operacionaliza desde 1 a 5, de acuerdo a la calificación cuantitativa del usuario en la reseña generada.	NPS = % Promotores - % Detractores  Promotores = 5 Neutros = 4 Detractores = 1 - 3	TripAdvisor
<b>Reseña textual</b>	Grado de Confiabilidad	representa la certeza estadística con la que el modelo de procesamiento de lenguaje natural (BERT multilingüe)	Medida de certeza estadística con la que el modelo BERT asigna una reseña a una categoría de sentimiento (Positiva, negativa o neutra)	Confidence score= max(P(Positiva), P(Neutra), P(Negativa))	TripAdvisor
<b>Mapeo de coordenadas y media de polaridad por zona</b>	Tasa de concentración geográfica del sentimiento	Asociación entre zonas del destino y sentimiento turístico	Polaridad promedio vinculada a coordenadas geográficas	TCGS = Conteo ponderado de servicios en la zona z / área efectiva de la zona (km <sup>2</sup> ) excluyendo superficies no utilizables (mar)	Datos georreferenciados

Fuente: Tabla realizada por el autor

En conjunto, estas variables e indicadores posibilitan una caracterización integral del destino, al reflejar tanto su estado actual del destino, como su posicionamiento digital y

su desempeño competitivo, Su implementación se constituye en un insumo estratégico de alto valor para orientar intervenciones fundamentadas en datos, identificar y priorizar áreas críticas de gestión y, en consecuencia, diseñar políticas públicas más efectivas y adaptadas a las particularidades del territorio.

### **8.1.3 Tecnologías utilizadas en el desarrollo**

La propuesta metodológica será implementada a través de un ecosistema tecnológico basado en herramientas de código abierto y plataformas ampliamente reconocidas en el campo de la inteligencia de negocios y el análisis turístico. La selección tecnológica responde tanto a los requerimientos del modelo como a las buenas prácticas identificadas en la literatura especializada, priorizando la flexibilidad, escalabilidad y compatibilidad con datos heterogéneos.

El lenguaje de programación Python se empleará como núcleo de desarrollo, dada su versatilidad y la amplia disponibilidad de librerías especializadas para tratamiento de datos turísticos. Se utilizarán herramientas como BeautifulSoup, Selenium y Httpx para la automatización de procesos de web scraping, en línea con los métodos implementados por autores como Bustamante (2019), Kim et al. (2017) y Santana (2019).

Ahora bien, para el procesamiento del lenguaje natural (NLP), se emplearán librerías como Transformers de HuggingFace, utilizando el modelo BERT (base-multilingua-uncased-sentiment). Esta decisión se sustenta en la evidencia presentada en investigaciones recientes, Frigau et al. (2024), que han demostrado la efectividad de BERT en el análisis semántico multilingüe aplicado a textos generados por usuarios. En línea con esta tendencia, se optó por BERT debido a su capacidad de representación

contextual bidireccional, su robustez en la detección de matices emocionales en diferentes idiomas y su pertinencia para abordar la heterogeneidad lingüística presente en las reseñas turísticas.

En cuanto al almacenamiento y gestión de datos, se combinarán tecnologías de base relacional y no relacional. Se implementará PostgreSQL, para permitir la gestión de datos espaciales y consultas georreferenciadas, tal como se ha planteado en investigaciones que integran análisis territorial con visualización de atractivos turísticos (Bustamante, 2019; Cerdá-Mansilla et al., 2022). De manera complementaria, se propone el uso de MongoDB como alternativa NoSQL para el almacenamiento flexible de datos no estructurados, lo que permite escalar el modelo hacia estructuras más complejas.

Para la etapa de visualización, monitoreo y análisis semántico, se integrarán herramientas como Power BI, que permitirá la creación de tableros interactivos y creación de medidas DAX orientadas a la toma de decisiones. (Köseoglu et al., 2020; Marine-Roig & Anton Clavé, 2015).

Finalmente, para la orquestación de los flujos de trabajo que requiere la arquitectura propuesta, se empleará Apache Airflow, una plataforma robusta de código abierto que permite definir flujos de procesamiento mediante grafos acíclicos dirigidos (DAGs), donde cada tarea representa una unidad lógica del pipeline y su ejecución se controla de manera programática, escalable y modular (Singh, 2019). En este modelo híbrido de arquitectura, Airflow actuará como el núcleo de orquestación que coordina las fases de captura desde el web scraping (extracción), procesamiento y enriquecimiento de datos (transformación), y posterior almacenamiento en bases estructuradas (PostgreSQL) y no

estructuradas (MongoDB), habilitando así la alimentación automática del Data Warehouse para su análisis en Power BI

En conjunto, estas tecnologías permiten articular un modelo replicable y escalable, fundamentado en infraestructura abierta, orientado a transformar datos turísticos dispersos en información estratégica visualizable, consultable y aplicable para la gestión inteligente del destino.

## **8.2 Diseño de la arquitectura del modelo BI**

### **8.2.1 Componentes del modelo**

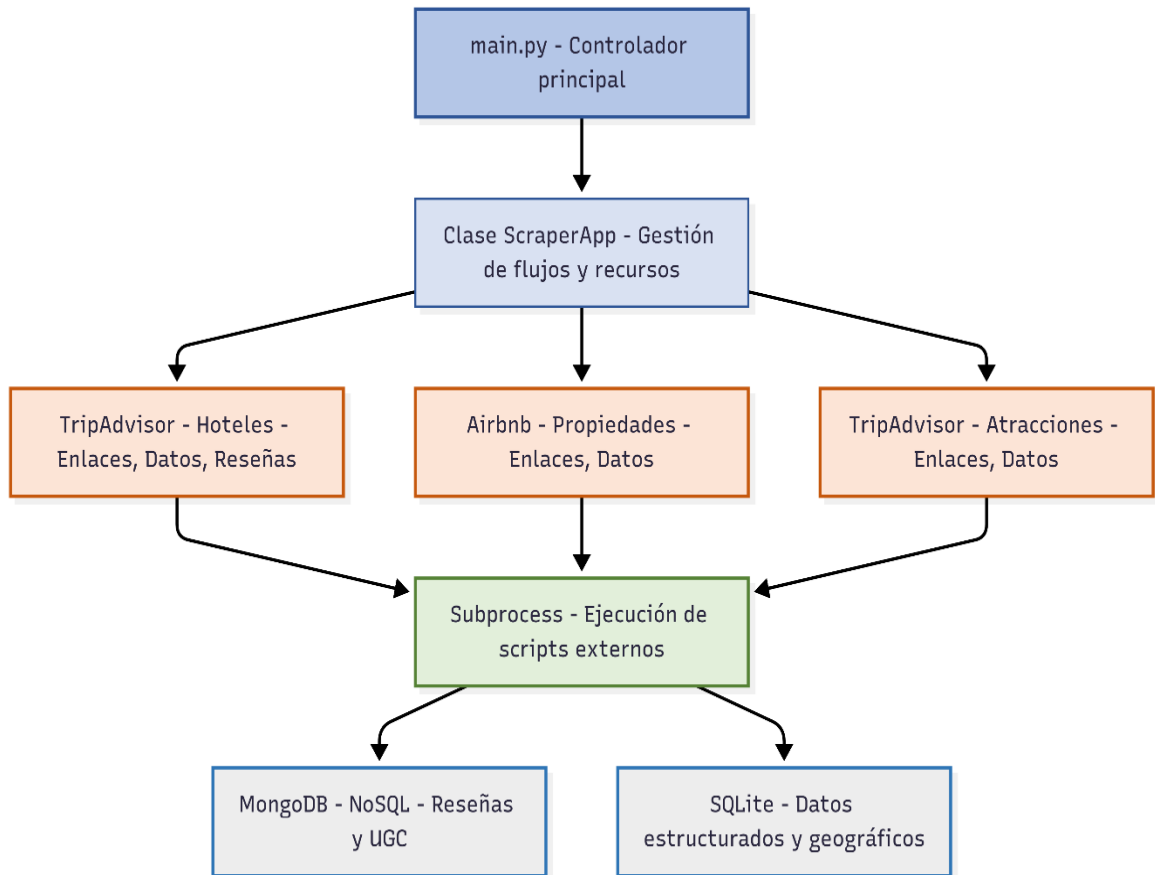
#### **8.2.1.1 Capa de extracción de datos**

La capa de extracción de datos sin duda constituye el punto principal de partida del modelo de inteligencia desarrollado en el presente trabajo de investigación. Como se ha planteado, esta capa se encarga de la obtención sistemática de los datos relevantes en las plataformas turísticas establecidas, TripAdvisor y Airbnb, mediante la implementación de técnicas avanzadas de web scraping. A continuación, se describe el flujo propuesto para la ejecución del scraper, tal como se sintetiza visualmente en la figura número 3:

- Inicialización del proceso: El flujo comienza con la definición de parámetros clave como URL inicial, headers personalizados, rotación automática de proxies mediante ProxyManager y HttpxProxyManager, y credenciales cuando son necesarios. Esta etapa se basa en principios de configuración reproducible y controlada, tal como recomiendan proyectos de scraping responsables (Kim et al., 2017).

- Navegación e interacción web simulada: Se emplea la librería *Selenium* a la par con *undetected\_chromedriver* para emular una navegación humana, de esta forma podemos acceder a contenidos protegidos por mecanismos anti-bots, manteniendo la estabilidad del scraper y evitando bloqueos por comportamiento sospechoso.
- Extracción de datos: Una vez se detecte la página objetivo, mediante el uso de librerías como *BeautifulSoup* y *lxml* se realiza el parseo del HTML, capturando atributos estructurados inmersos dentro de su contenido como nombres de alojamientos o del destino, puntuaciones, coordenadas geográficas, al igual que atributos no estructurados como las reseñas textuales, las descripciones y etiquetas.
- Manejo de paginación: El sistema identifica automáticamente la existencia de múltiples páginas de resultados, aplica una estrategia iterativa para recorrer cada una de estas páginas de forma secuencial o paralela según la estructura de paginación.
- Gestión de sesiones HTTP: A través de la librería *httplib2*, se controla los header de solicitud, las cookies persistentes y las redirecciones, lo cual mejora la eficiencia frente a requests y permite controlar múltiples sesiones en paralelo. Esta librería moderna se alinea con los estándares de clientes asíncronos compatibles con versiones superiores a Python 3.10.
- Almacenamiento temporal y persistente: Los datos extraídos se almacenan inicialmente en una estructura transitoria como listas y diccionarios de Python, luego son serializados en formato JSON para ser alojados en una zona RAW que se conforma por dos tipos de bases de datos: **MongoDB** para datos no estructurados y **SQLite**, para datos estructurados como metadatos de hoteles, atracciones y Airbnb, donde se almacenan todos los datos en bruto extraídos.

**Figura 3 - Flujo principal ScraperApp**



Fuente: Elaboración propia

En esta primera capa se basa en una gestión modular nativa desde el entorno de Python, centralizada en un archivo `main.py`, que actúa como controlador principal del proceso de extracción. La lógica de ejecución se organiza a partir de la clase `scraperApp`, diseñada para operar como núcleo de coordinación entre los diferentes componentes funcionales: scraping, almacenamiento, validación, control de estado y configuración.

Esta clase además de inicializar los principales recursos también establece flujos independientes para cada tipo de fuente y dato extraído. La implementación utiliza técnicas de procesamiento asíncrono mediante `asyncio` y `httpx`, permitiendo la ejecución

concurrente de múltiples operaciones de extracción y mejorando significativamente el rendimiento del sistema.

El usuario final podrá ejecutar esta clase en tres principales bloques, el primer bloque corresponde a la gestión del destino, donde se selecciona o crea un nuevo destino, Si el destino ya existe en la base de datos, el sistema ofrece la opción de visualizar los datos existentes o actualizar la información. En caso contrario, se habilita el scraping completo del nuevo destino, registrando todos los metadatos necesarios.

El segundo bloque corresponde a la selección y procesamiento de tipos de datos, eligiendo entre tres tipos de flujos independientes y asíncronos, ya sea atracciones turísticas del destino, alojamientos registrados en TripAdvisor y propiedades en Airbnb; estos tres tipo de flujos cuentan con un conjunto de fases condiciones que permiten procesar enlaces y datos por separado, la fase A corresponde al procesamiento y extracción de enlaces semilla como lo son las URLs, la fase B realiza un scraping de contenido detallado para cada alojamiento, atracciones y propiedades en Aribnb, por último, la fase C es exclusiva para alojamientos de TripAdvisor ya que esta permite extraer las reseñas individuales de cada alojamiento registrado en el URL semilla del alojamiento en esta plataforma.

Una vez finalizado el proceso de scraping, el sistema genera unas estadísticas básicas del procesamiento y muestra un resumen de los resultados generalizados del destino, siendo un cierre lógico del proceso y garantizando una primera capa de validación sobre los datos recolectados. Finalmente, esta arquitectura no solo permite una extracción robusta y segmentada de acuerdo con el tipo de entidad y fuente como ya

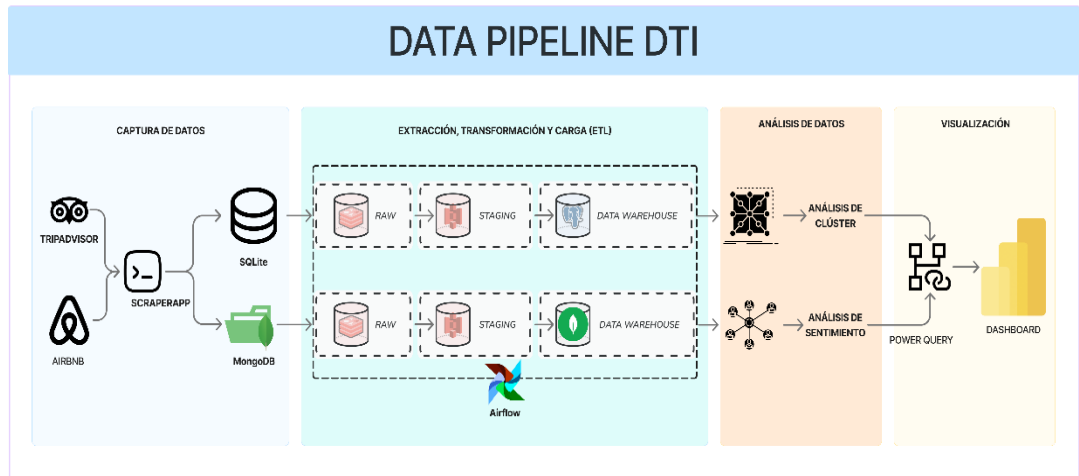
se ha explicado, sino que establece una base sólida y extensible para la inclusión futura de nuevos destinos turísticos, sin necesidad de rediseñar el sistema.

#### **8.2.1.2 Capa de procesamiento e integración (ETL): Flujo con Airflow**

Una vez completada la capa de extracción, el modelo propuesto incorpora una capa de procesamiento, diseñada con el fin de transformar, limpiar, validar y estructurar los datos recolectados, garantizando datos analíticamente útiles y estratégicamente relevantes para la toma de decisiones. Esta fase constituye el núcleo principal del ETL (Extract, Transform, Load), y ha sido concebido bajo una arquitectura por zonas (Raw, Staging y Warehouse), orquestada mediante Apache Airflow.

Esta capa responde al principio de preservación del dato, desde su origen como datos crudos, hasta su consolidación en un modelo dimensional optimizado en PostgreSQL y MongoDB, lo que garantiza consistencia semántica, integridad relacional y auditabilidad para los KPI generados posteriormente (Pérez et al., 2018).

**Figura 4 - Pipeline Modelo BI**



Fuente: Elaboración propia

Cómo se visualiza en la figura número 4, la arquitectura de datos se organiza en tres zonas secuenciales, cada una con funciones específicas de limpieza, integración, validación o modelado.

#### 8.2.1.2.1 Zona Raw

En esta zona se almacenará los datos en su estado original, tal como fueron extraídos, teniendo en cuenta que es una arquitectura híbrida, se subdivide en:

- PostgreSQL RAW: Almacena datos estructurados migrados desde SQLite. La carga se ejecuta mediante un DAG que verificará la integridad de los datos fuentes y registrará metadatos de ingesta.
- MongoDB Raw: Contiene datos no estructurados, reseñas textuales extraídos directamente desde TripAdvisor. La colección conservará su forma original, lo que permitirá aplicar posteriormente análisis semánticos y de sentimientos sin pérdida de granularidad.

#### 8.2.1.2.2 Zona Staging

En esta zona se aplicarán los procesos de transformación y enriquecimiento. Aquí se depuran, normalizan e integran los datos antes de su modelado analítico, este proceso será orquestado mediante un DAG que cumplirá con las siguientes funciones:

- Unificación de alojamientos: Registros de alojamientos de TripAdvisor y Airbnb se integrarán en esta zona, mediante procesos de deduplicación de datos, estandarización geoespacial y mapeo de amenidades.
- Procesamiento de reseñas: Las reseñas alojadas en MongoDB se transformarán y cargarán en esta zona, incorporando resultados del análisis de sentimientos y metadatos como fecha de publicación o tipo de dispositivos.
- Normalización y enriquecimiento de destinos turísticos: Se generarán nuevas variables que permiten la categorización de tipos de destino y se organizan los puntos de interés turístico según sus jerarquías temáticas y ubicación.

#### **8.2.1.2.3 Zona Data WareHouse**

En esta zona ya final se implementará el modelo dimensional para una correcta explotación de los datos y visualización de herramientas como PowerBI. Esta zona se gestionará por un DAG que deberá ejecutar:

- Construcción de dimensiones: Mediante de los datos alojados en la zona anterior, se modelarán y poblarán las principales dimensiones, aplicándoles claves principales y estructuras normalizadas.
- Tablas puente: Se crearán estructuras que posibiliten el manejo de relaciones complejas como muchos a muchos, facilitando consultas OLAP y normalizando el modelo.
- Tabla de hechos: Se consolidará el modelo y la población de la tabla de hechos con métricas relacionados con el análisis de sentimiento, volumen de reseñas, puntuación media y se vinculará a claves de dimensiones para análisis comparativo y temporal.

El sistema deberá ser construido con un flujo secuencial y dependiente, gestionado por Apache Airflow, mediante los DAGS construidos que encapsulan tres tareas específicas y controlarán se ejecución cronológica y condicional. A su vez, se tendrán DAGs para asegurar la integridad del pipeline en cada proceso, garantizando la disponibilidad del volumen mínimo de registros previos a la transformación, la integridad de las claves primarias, foráneas y estructuras relacionales esperadas, al igual que la consistencia respetando las reglas de negocio aplicado al dominio turístico.

Esta capa culmina con la implementación del modelo dimensional entidad relación completa representado en la Figura 5. Este modelo integra la estructura lógica final del Data Warehouse sobre el cual se constituirán los análisis estratégicos del destino turístico inteligente como ya hemos definido.



- dim\_alojamientos: que almacena información estructurada de cada establecimiento, su localización geográfica, tipo, amenities y características generales.
- dim\_usuarios, dim\_ubicacion y dim\_idiomas: permiten caracterizar el perfil del visitante y contextualizar las reseñas por origen, idioma y región.
- dim\_fecha: estructura los datos temporalmente para análisis por día, mes, trimestre o temporada.
- dim\_trip\_type y dim\_device: capturan metadatos relevantes del comportamiento del usuario al momento de emitir una reseña.
- Tablas puente (bridge\_\*): permiten representar relaciones complejas entre alojamientos y sus atributos (amenities, subratings, tipos de habitación), manteniendo la normalización del modelo y facilitando su extensión futura.

Finalmente, la capa de procesamiento e integración constituye uno de los pilares estructurales dentro del diseño del modelo, ya que permitirá consolidar las principales zonas funcionales (Raw, Staging, Warehouse), integración de bases SQL/NoSQL y un sistema de orquestación robusto como Apache Airflow, permitiendo construir un flujo sólido, modular y escalable, manteniendo la gobernanza del dato y la confiabilidad del modelo.

### **8.2.1.3 Capa semántica y de análisis**

La capa semántica y de análisis constituye uno de los componentes más importantes del modelo, integrando técnicas de análisis estadístico avanzado y técnicas de

inteligencia artificial sobre datos estructurados y no estructurados. Esta capa se orienta a extraer valor estratégico desde las características de la oferta turística y el contenido generado por usuarios (UGC), mediante dos enfoques principales: la segmentación de establecimientos mediante análisis de clúster y el análisis de sentimiento basado en procesamiento de lenguaje natural (NLP).

#### **8.2.1.3.1 Análisis de Clustering**

Para este análisis, se plantea un diseño de segmentación utilizando el algoritmo K-means, ampliamente utilizado en modelos turísticos por su eficacia para identificar patrones latentes en datos multivariados y su capacidad para generar perfiles diferenciados de la oferta (Jain, 2010; Prayag et al., 2013). El objetivo de la segmentación es conformar grupos de establecimientos turísticos que resulten internamente homogéneos en función de variables críticas, pero que a la vez sean heterogéneos respecto a otros clústers, lo que posibilita así caracterizar tipologías de ofertas claramente diferenciadas.

Las variables seleccionadas son:

- Precio promedio del alojamiento
- Puntuación general y volumen de reseñas
- Porcentaje de reseñas positivas, neutras y negativas
- Ubicación geográfica (Latitud / Longitud)
- Tipo de alojamiento

Previo a la ejecución del algoritmo se deberá garantizar un proceso de normalización de escalas, con el fin de evitar que alguna variable domine la medida de distancia euclidiana empleada en K-means. Asimismo, la determinación del número óptimo de

clústeres se establecerá mediante técnicas de validación como el método del codo (Elbow Method) y el análisis del índice de silueta, que permitirá evaluar la compacidad y separación entre grupos (Jain, 2010)

Este análisis nos permitirá identificar zonas con alta densidad de alojamientos económicos con reputación negativa, clústers de establecimientos premium con satisfacción elevada y grupos con baja diferenciación competitiva. Esta información será incorporada al dashboard analítico como una dimensión adicional para el filtrado y cruce con otros KPIs de percepción y desempeño.

#### **8.2.1.3.2 Procesamiento de lenguaje natural NLP y análisis de sentimiento**

Para este análisis, se implementará un diseño de procesamiento de lenguaje natural (NLP) orientado a clasificar automáticamente el sentimiento de las reseñas turísticas. Se empleará el modelo BERT multilingüe preentrenado (Bidirectional Encoder Representations from Transformers), reconocido por su capacidad para capturar dependencias contextuales y matices semánticos en textos generados por usuarios (Frigau et al., 2024). El modelo se integra mediante la librería Transformers de HuggingFace, lo que garantiza compatibilidad multilingüe y precisión en entornos con alta heterogeneidad idiomática por parte de sus visitantes, como lo es el Archipiélago de San Andrés.

El procesamiento contempla tanto los títulos como los contenidos de las reseñas, a los cuales se aplica una combinación ponderada (70% contenido, 30% título) con

el fin de priorizar la información semánticamente más rica. Cada reseña es clasificada en una escala de 1 a 5, que posteriormente se mapea en tres categorías de sentimiento: positivo (4–5), neutro (3) y negativo (1–2). Adicionalmente, se calculará un confidence score asociado a cada predicción, que representa la probabilidad estadística asignada por el modelo a la clase elegida, constituyéndose en un insumo clave para el KPI de grado de confiabilidad.

Previo a la ejecución del análisis, se implementará un procedimiento de detección automática de idioma, que permite filtrar y procesar adecuadamente reseñas en español e inglés. En caso de errores de lectura o reseñas incompletas, el sistema aplica un mecanismo de fallback controlado, asignando clasificación neutra y confianza nula, lo cual asegura la trazabilidad de los casos procesados. Este diseño garantiza robustez metodológica y evita sesgos derivados de la pérdida de datos durante la ejecución en lotes.

El análisis semántico de las reseñas permitirá generar indicadores agregados como el Índice de Sentimiento Turístico (IST), el Net Promoter Score (NPS) calculado desde las calificaciones asociadas, y la distribución porcentual de reseñas positivas, neutras y negativas. Asimismo, el grado de confiabilidad medio se incorpora como criterio de validación de los resultados. Esta información se integrará en el dashboard analítico, facilitando la exploración interactiva de percepciones por categoría de alojamiento, zona geográfica y periodo, y ofreciendo una base empírica sólida para la gestión de la reputación digital del destino.

#### **8.2.1.4 Capa de visualización**

Para la fase de visualización se plantea un diseño de tablero interactivo en Power BI estructurado en capas temáticas, con el fin de integrar los resultados del modelo de inteligencia de negocios en un entorno visual accesible y estratégico. La organización por lienzos permite separar analíticamente la oferta de servicios, las atracciones turísticas, la percepción de los usuarios y la satisfacción global, garantizando al mismo tiempo la posibilidad de cruzar indicadores y explorar patrones de manera dinámica.

El primer lienzo corresponde a los alojamientos, en la cual se integrarán indicadores clave como el número total de establecimientos, el Score de Reputación Digital (SRD), la distribución por tipo de alojamiento, el precio promedio y la calificación ponderada. Esta capa permitirá explorar la concentración espacial de la oferta y su desempeño en términos de reputación y competitividad, facilitando comparaciones entre categorías, zonas geográficas y fuentes de origen de las reseñas.

El segundo lienzo se orienta a las atracciones turísticas, donde se visualizarán métricas como el volumen total de reseñas, el SRD por atracción y el porcentaje de atracciones excelentes. Esta visualización permitirá identificar las categorías más representativas (playas, museos, bares, parques naturales, entre otras) y evaluar su posicionamiento en función de la satisfacción percibida por los visitantes, generando insumos para priorizar la gestión y promoción de recursos turísticos.

El tercer lienzo se centra en la satisfacción del usuario, a través de la incorporación del Net Promoter Score (NPS), la proporción de promotores, detractores y neutros, así

como la evolución temporal de las calificaciones promedio y el volumen de reseñas. Esta capa permitirá monitorear de manera comparativa la experiencia del visitante, detectar cambios en las tendencias de satisfacción y segmentar resultados por idioma, país de origen y tipo de viaje.

El cuarto lienzo corresponde a la percepción del destino, donde se representarán los resultados del análisis de sentimiento derivado de BERT, incluyendo el Índice de Sentimiento Turístico (IST), el grado de confiabilidad de las clasificaciones y la distribución porcentual de reseñas positivas, negativas y neutras. Adicionalmente, se incorporará la tasa de concentración geográfica del sentimiento, permitiendo mapear zonas críticas y áreas de oportunidad en la gestión de la reputación digital del destino.

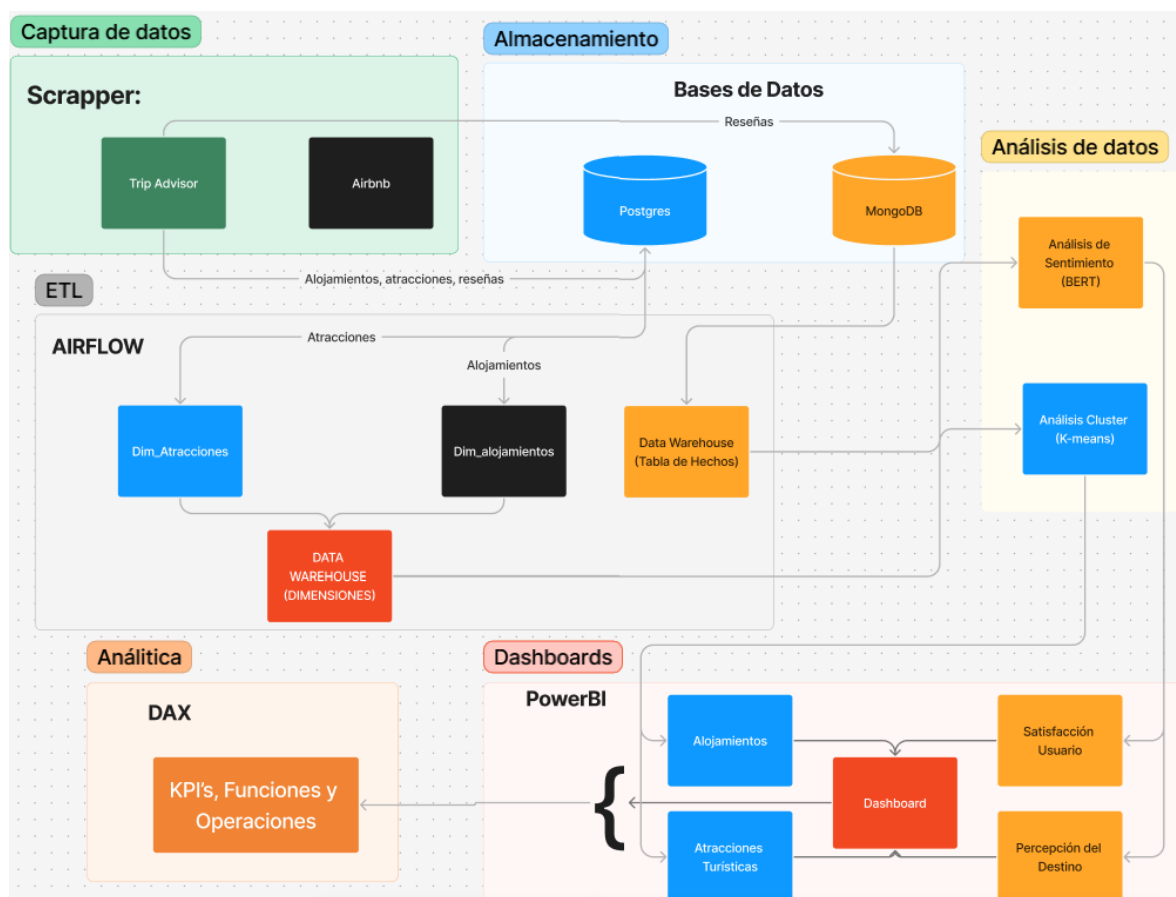
Finalmente, la combinación de estos lienzos en un solo tablero permitirá al usuario filtrar, cruzar y comparar la información en múltiples dimensiones (zona geográfica, categoría de servicio, periodo de tiempo), consolidando una herramienta robusta para la toma de decisiones basada en datos. Este enfoque de visualización no solo ofrece transparencia metodológica, sino que también dota al modelo de aplicabilidad práctica al facilitar la priorización de áreas críticas, la identificación de oportunidades y el diseño de estrategias de intervención turística en San Andrés Isla.

#### **8.2.1.4.1 Diagrama general del modelo**

Finalmente, estas cuatro capas conforman el pipeline a seguir para la ejecución del modelo de inteligencia de negocios, este diseño planteado en la Figura 6, implementa una integración coherente entre la captura, transformación, análisis y visualización de

datos, mediante la automatización, la trazabilidad y el procesamiento robusto de los datos, convirtiéndose en un modelo estructurado y dinámica para sustentar analíticamente la toma de decisiones estratégicas en un destino turístico inteligente, interactuando mediante diferentes lienzos de datos. Finalmente, este modelo además de representar para el destino un salto cualitativo en la forma en cómo se concibe y se gestiona el conocimiento turístico, permite alinear los principios de eficiencia, sostenibilidad y transformación digital en el contexto de la construcción de un destino turístico inteligente.

Figura 6 - Diagrama General del Modelo



Fuente: Elaboración propia

### **8.3 Implementación y resultados del modelo**

#### **8.3.1 Implementación operativa del modelo en Power BI**

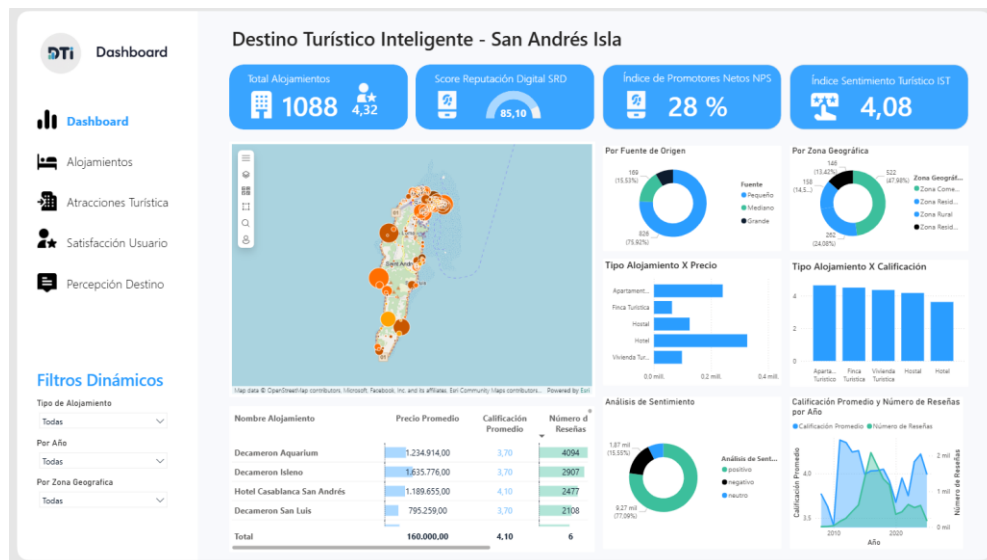
La fase de implementación constituye el punto de convergencia entre el diseño metodológico y el procesamiento de datos descritos en la sección anterior. Con el data warehouse ya operativo y los KPIs formalizados, tanto los de oferta de servicios turísticos como los de percepción del destino, estos fueron calculados, validados e integrados en un tablero interactivo en Power BI, que representa la materialización operativa del modelo de inteligencia de negocios. La actualización del data warehouse se gestionó mediante Apache Airflow, a través de flujos ETL que automatizan la captura desde las fuentes abiertas, su transformación y limpieza en estructuras analíticas y la carga en las tablas dimensionales y de hechos, tal como fue planteado en el diseño metodológico. Gracias a esta orquestación, la plataforma no solo funciona como un repositorio de indicadores, sino como un entorno dinámico y confiable que permite explorar, filtrar y comparar resultados en múltiples dimensiones, asegurando la vigencia y trazabilidad de la información y habilitando una lectura estratégica de los hallazgos para distintos perfiles de decisión.

De esta forma, el diseño del tablero se estructuró en múltiples lienzos temáticos, cada uno de ellos alineados con los principales intereses de los principales actores del sector. El lienzo de oferta de alojamientos integra métricas como el Score de Reputación Digital (SRD), los precios promedio por categoría y la concentración espacial de servicios. La vista de atracciones turísticas incorpora indicadores de reputación y volumen de reseñas,

mientras que el lienzo de satisfacción del usuario despliega el Net Promoter Score (NPS) junto con la evolución de promotores, detractores y neutros. Finalmente, la capa de percepción del destino articula el Índice de Sentimiento Turístico (IST), el grado de confiabilidad del análisis de sentimiento y la distribución porcentual de reseñas positivas, negativas y neutras. Esta organización permite no solo analizar cada dimensión por separado, sino también identificar patrones cruzados entre la oferta, la satisfacción y la percepción global.

La implementación también consideró aspectos técnicos que fortalecen la utilidad del modelo. Como se ilustra en la Figura 7, se integraron filtros interactivos por zona geográfica, categoría de servicio, idioma y periodo temporal, lo cual habilita un análisis granular y adaptado a las necesidades de distintos actores del destino; de igual forma, se configuraron visualizaciones específicas, mapas de calor, diagramas comparativos y tarjetas de indicadores, que optimizan la interpretación de la información y facilitan la comunicación de hallazgos en escenarios de toma de decisiones.

**Figura 7 - Dashboard general**



*Fuente: Elaboración propia*

### **8.3.2 Resultados**

Tras la implementación operativa del modelo en Power BI, en esta sección se presentan los resultados analíticos obtenidos en cada uno de los lienzos temáticos del dashboard, en los que se consolidan y visualizan los indicadores previamente definidos. Estos resultados permiten evaluar de manera integral la oferta de alojamientos, la relevancia de las atracciones turísticas, los niveles de satisfacción del usuario y la percepción global del destino. Cada lienzo se interpreta como una traducción visual de los KPIs calculados en las capas analíticas, facilitando no solo su comprensión por parte de actores técnicos y decisores, sino también su aplicación práctica en procesos de planificación y gestión. En este sentido, el paso de la implementación a la presentación de resultados constituye la validación empírica del modelo, evidenciando cómo los datos, al ser organizados y contextualizados en el dashboard, se convierten en insumos estratégicos para la gestión de San Andrés como un destino turístico inteligente.

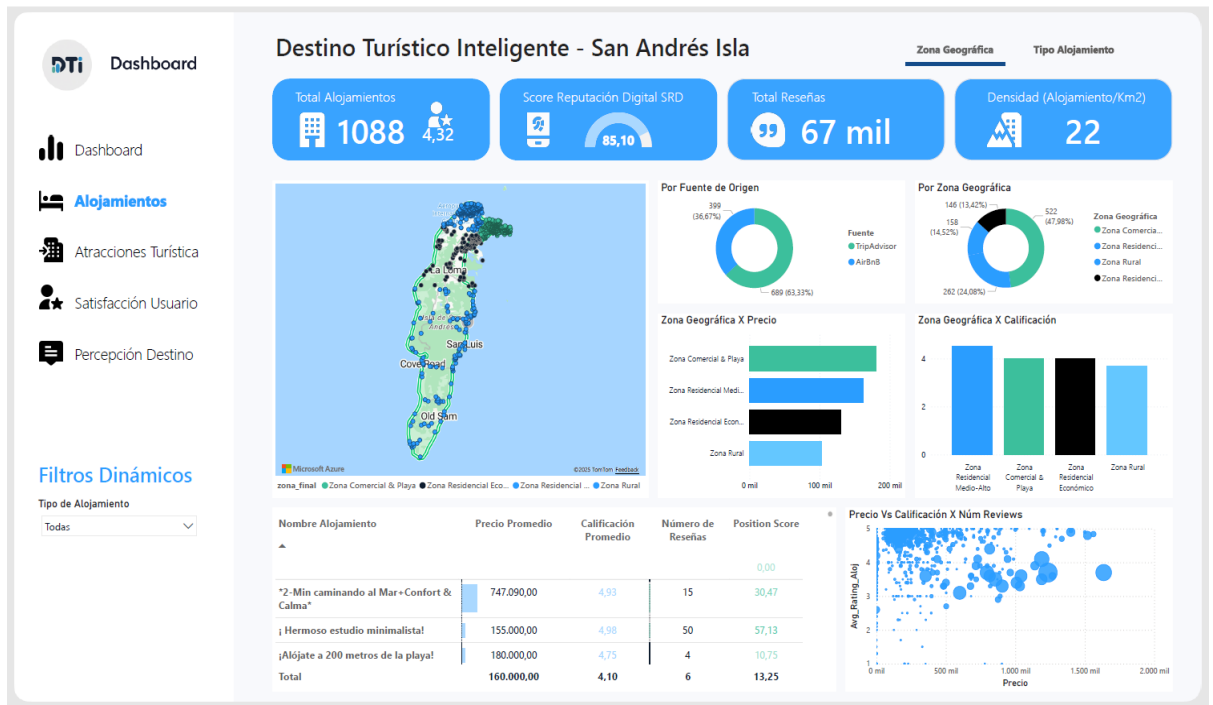
#### **8.3.2.1 Alojamientos**

Ahora bien, la figura 8 nos permite visualizar el lienzo de alojamientos del dashboard, en este panel se consolida un total de 1.088 establecimientos registrados (689 alojamientos de TripAdvisor y 399 de Airbnb), con una calificación promedio global de 4,32 puntos y un Score de Reputación Digital (SRD) de 85,1, lo que refleja una percepción favorable del destino en términos de hospitalidad. El sistema ha recopilado más de 67 mil reseñas, de las cuales la mayoría proviene de TripAdvisor (61,3%) frente a Airbnb (38,7%), lo que evidencia la relevancia de estas dos plataformas como fuentes de

reputación digital. Es de destacar, que más del 50% de las reseñas provienen de Hoteles, concentrando el mayor número de reseñas verificadas en un total de 170 alojamientos, lo que refuerza el rol de los hoteles como eje central en la construcción de la imagen turística.

La concentración geográfica muestra que la mayor densidad se da en las zonas comerciales y de playa, segmentos que también presentan los precios promedio más altos, lo cual sugiere una relación entre localización privilegiada, posicionamiento de marca y percepción de valor por parte de los visitantes. Ahora bien, el cruce entre precio y calificación sugiere que, aunque existen alojamientos de alto costo, la mejor percepción se distribuye en diferentes segmentos, sin una correlación lineal entre precio y satisfacción. Finalmente, la nube de puntos que combina precio, calificación y número de reseñas evidencia una dispersión competitiva en la oferta, con clusters de alojamientos económicos que mantienen buena reputación y segmentos premium con alta concentración de comentarios, aportando insumos valiosos para la toma de decisiones estratégicas sobre la gestión de la oferta turística.

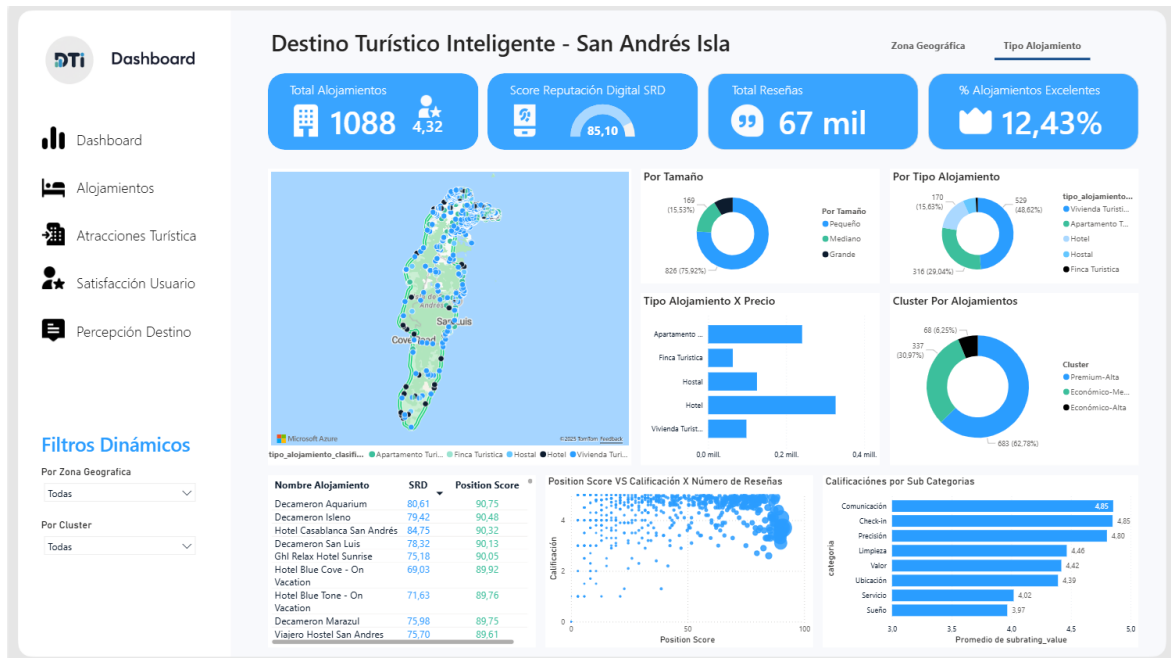
### **Figura 8 - Dashboard Alojamientos Por Zona Geográfica**



Fuente: Elaboración propia

Al analizar los resultados por tipología de alojamiento, figura 9, se observa que las viviendas y apartamentos turísticos también tienen un peso relevante en la composición de la oferta, representando en conjunto más del 60% de los establecimientos. Sin embargo, al compararse con los hoteles, estas categorías exhiben una menor concentración de reseñas, lo que limita su impacto en el posicionamiento reputacional del destino. Los hostales y fincas turísticas, aunque en menor proporción, aportan diversidad a la oferta y capturan segmentos de viajeros específicos. Asimismo, los análisis de clúster permiten identificar tres segmentos diferenciados de alojamientos (premium-alta, económico-medio y económico-alta), lo que facilita la caracterización de mercados objetivos y la identificación de oportunidades para el diseño de estrategias de competitividad.

**Figura 9 - Dashboard Alojamiento por Tipo Alojamiento**



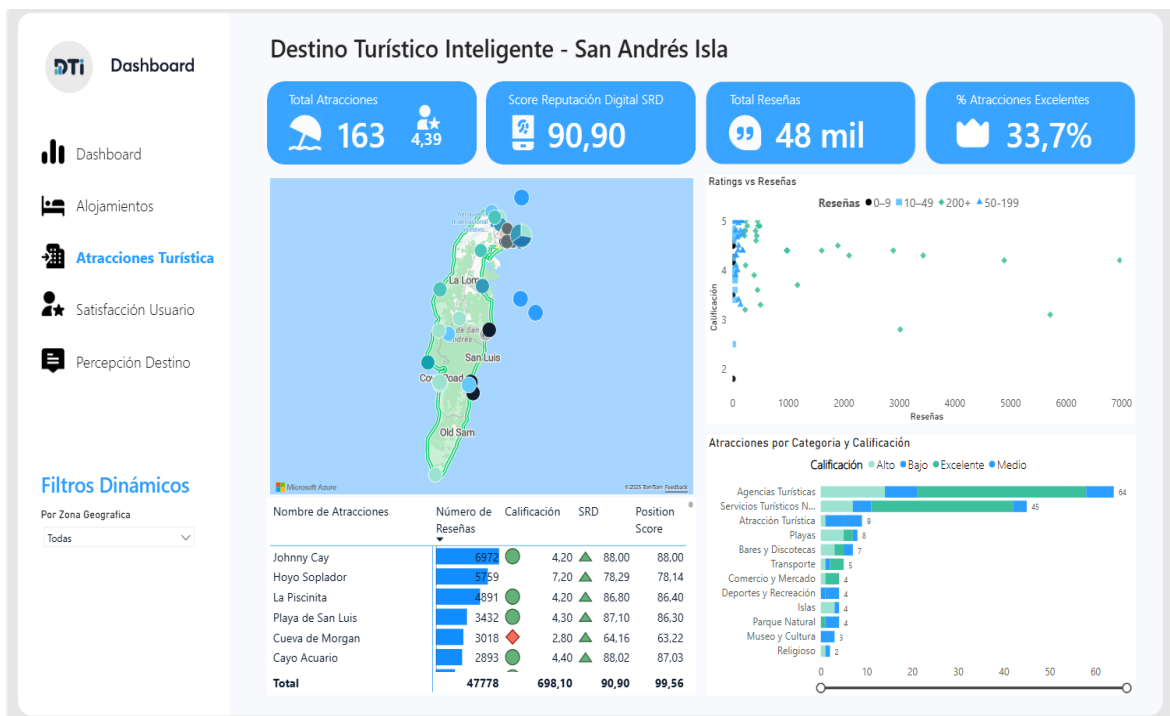
Fuente: Elaboración propia

### 8.3.2.2 Atracciones turísticas

Continuado con los resultados, el lienzo de atracciones turísticas, figura 10, muestra un total de 163 recursos registrados, que alcanzan un Score de Reputación Digital (SRD) de 90,9 y un promedio de 4,39 estrellas, lo que sugiere una alta satisfacción general de los visitantes del destino y con una alta representatividad, contando con más de 48 mil reseñas acumuladas. Los resultados evidencian una alta concentración de agencias turísticas y servicios náuticos, los cuales no solo representan la categoría con mayor número de establecimientos, sino que también mantienen calificaciones elevadas, lo que reafirma su importancia en la dinámica turística del archipiélago. Entre las atracciones de mayor posicionamiento, destacan Johnny Cay, Felo Divers y Cayo Acuario como las tres principales atracciones posicionadas del destino, reflejando la fuerte preferencia de los

visitantes por experiencias asociadas al mar y al contacto con la naturaleza. La relevancia de estos atractivos radica no solo en su capacidad de atracción masiva, sino también en la solidez de su reputación digital, que se constituye en un activo estratégico para la competitividad del destino y la consolidación de la marca destino en el mercado turístico global.

**Figura 10 - Dashboard Atracciones Turísticas**



Fuente: Elaboración propia

### 8.3.2.3 Satisfacción del Usuario

En la figura 11, el lienzo de Satisfacción del Usuario sintetiza 6.007 reseñas individuales emitidas por 5.421 usuarios sobre 346 alojamientos, con una calificación promedio de 4,03 sobre 5. El Índice de Promotores Netos (NPS) se sitúa en 28%, resultado de un 52% de promotores frente a un 25% de detractores. Esta lectura indica

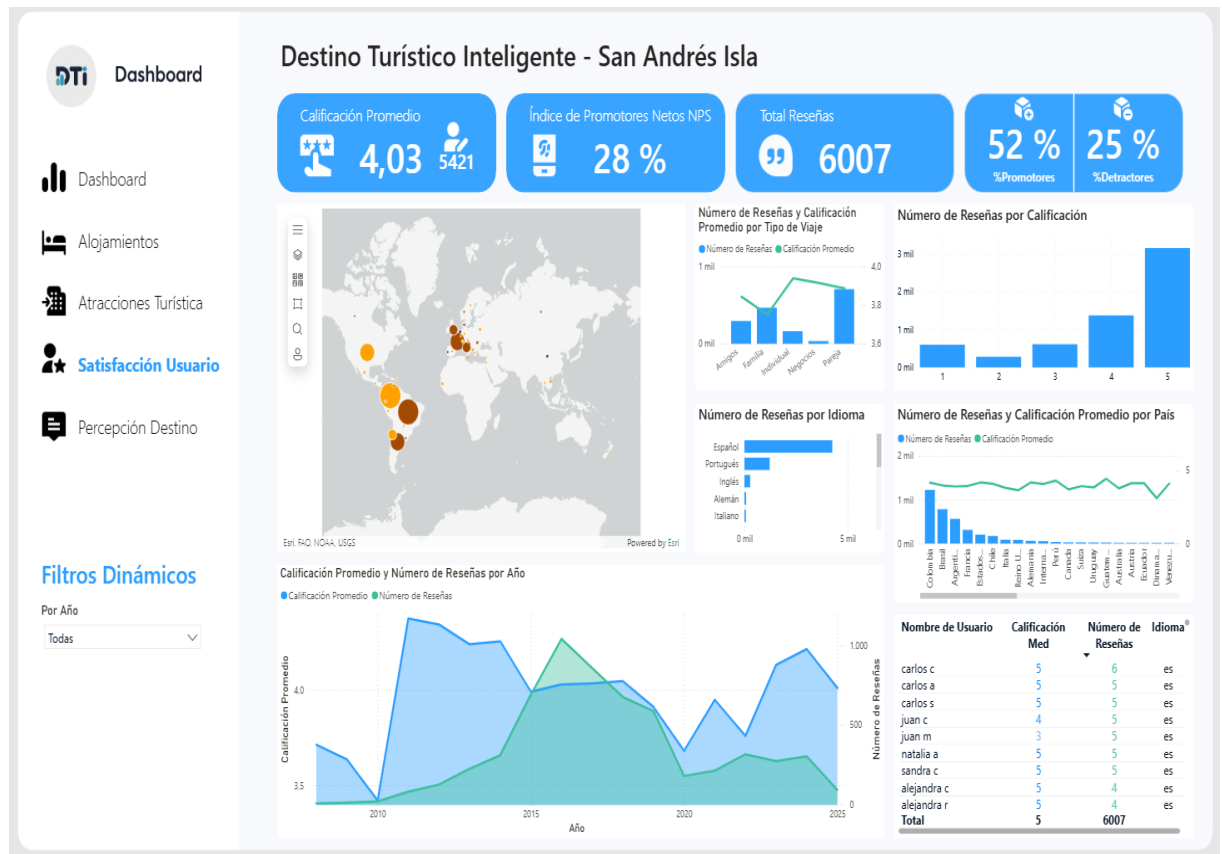
una predisposición positiva a recomendar el destino, ya que, de cada 100 visitantes, 28 están más dispuestos a recomendar activamente el destino. La distribución por calificación muestra pesos importantes en 4 y 5 estrellas, consistente con el promedio general, sin embargo, es importante realizar seguimientos a las calificaciones entre una sola estrella, ya que representa cerca del 10% del total de reseñas.

La serie temporal de calificaciones promedio y número de reseñas por mes evidencia oscilaciones significativas tanto en volumen como en puntuación, lo que permite identificar temporadas críticas y posibles efectos de choque en la experiencia turística. En particular, se observa que octubre registra una de las caídas más pronunciadas en el volumen de reseñas, aunque paradójicamente concentra una de las mejores calificaciones promedio, lo que sugiere que, si bien hay menos visitantes activos en plataformas digitales, la experiencia percibida por quienes reseñan es más positiva. En contraste, enero aparece como el mes con mayor número de reseñas generadas, coincidiendo con un período de alta demanda turística, pero al mismo tiempo presenta la calificación promedio más baja del año, reflejando una posible saturación de servicios y presiones sobre la capacidad de carga del destino. Estos resultados permiten indicar una posible correlación negativa entre el volumen de reseñas y la calificación promedio, donde los picos de afluencia turística estarían asociados a una menor satisfacción relativa, aspecto clave para la gestión de la estacionalidad y la planificación estratégica en San Andrés.

Otro hallazgo estratégico es la relevancia del segmento lusófono: el lienzo exhibe un peso destacado de reseñas en portugués, con Brasil como origen principal tanto en volumen como en calificación media, lo que posiciona a este país como mercado

prioritario para acciones de fidelización y comunicación. En cuanto al tipo de viaje, predominan los viajes en pareja, seguidos por familia, patrón que orienta la propuesta de valor. De forma complementaria, el cruce “calificación promedio × país × volumen” permite detectar mercados con alto potencial de recomendación y, simétricamente, bolsas de detractores por origen, idioma o temporada, insumos para priorizar mejoras operativas y campañas segmentadas. En conjunto, estos resultados refuerzan la hipótesis de un posicionamiento reputacional favorable pero perfectible, donde el manejo activo de detractores, la personalización por idioma y la adecuación a los patrones de viaje pueden traducirse en ganancias rápidas de NPS y en mayor resiliencia competitiva del destino.

**Figura 11 - Dashboard Satisfacción Usuario**



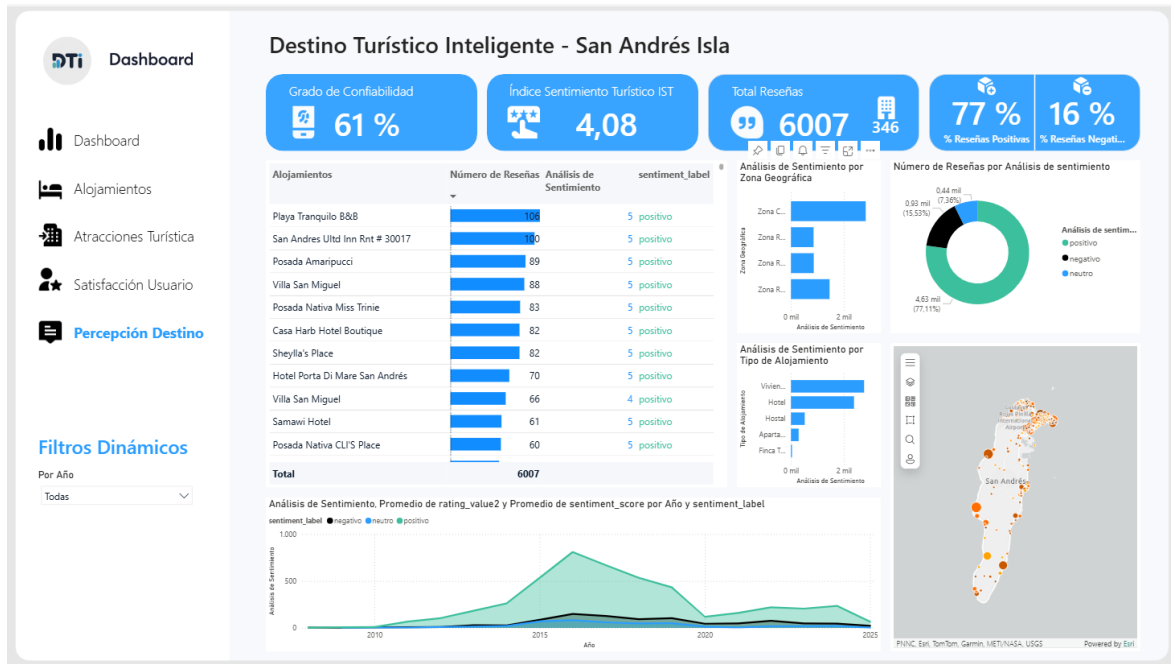
Fuente: Elaboración propia

#### **8.3.2.4 Percepción del Destino**

La Figura 12 presenta los resultados del análisis de sentimiento turístico realizado con el modelo BERT multilingüe, alcanzando un grado de confiabilidad del 61%, lo que refleja un nivel aceptable de certeza estadística en la clasificación automática de las opiniones. El Índice de Sentimiento Turístico (IST) se sitúa en 4,08, en una escala de 1 a 5, lo que se interpreta como una percepción predominantemente positiva del destino por parte de los visitantes. Este indicador se complementa con la distribución de polaridades: el 77% de las reseñas son positivas, un 16% negativas y el resto neutras, lo que confirma que la reputación digital de San Andrés es sólida y se construye mayoritariamente sobre experiencias satisfactorias.

El lienzo permite identificar que la mayor parte de los comentarios analizados se concentran en categorías como viviendas turísticas y hoteles, así como en las zonas comerciales y de playa, reforzando la idea de que la percepción positiva está asociada a los segmentos más representativos de la oferta. El análisis temporal evidencia además picos de sentimiento positivo en años de mayor volumen de reseñas, lo cual sugiere que la satisfacción se mantiene estable aun en escenarios de alta demanda. Por su parte, la cartografía de puntos georreferenciados permite visualizar las áreas críticas y los focos de reputación dentro de la isla, insumo clave para orientar estrategias de gestión y priorización territorial. En conjunto, este análisis de sentimiento demuestra que la voz digital de los turistas respalda la competitividad del destino, pero también visibiliza un núcleo de reseñas negativas que debe ser atendido para evitar impactos en la sostenibilidad de la imagen turística a largo plazo.

**Figura 12 - Dashboard Percepción Destino**



Fuente: Elaboración propia

En síntesis, los resultados presentados en los distintos lienzos del dashboard confirman la capacidad del modelo para integrar, procesar y visualizar información turística en evidencia accionable. La convergencia de indicadores como el SRD, NPS e IST, combinados con análisis espaciales y temporales, permite comprender de manera integral la dinámica turística de San Andrés y detectar tanto fortalezas, como la alta reputación digital de las atracciones y la predominancia de los hoteles en la oferta, como debilidades vinculadas a la estacionalidad, la gestión de detractores y la sostenibilidad de los recursos más demandados. Más allá de su valor descriptivo, el dashboard se constituye en una plataforma estratégica de apoyo a la decisión, al facilitar la exploración interactiva de la información y habilitar cruces entre dimensiones que antes se analizaban de forma aislada. De esta manera, el modelo no solo valida la pertinencia de aplicar técnicas de inteligencia de negocios al turismo, sino que también se proyecta como un

instrumento de gestión alineado con los principios de los Destinos Turísticos Inteligentes (DTI), aportando al fortalecimiento de la competitividad y sostenibilidad del Archipiélago de San Andrés.

### **8.2.1 El dashboard como herramienta para la inteligencia estratégica del destino**

El tablero implementado se consolida como un recurso de inteligencia estratégica, al integrar en un mismo espacio tanto indicadores de desempeño de la oferta turística como métricas de percepción y satisfacción de los usuarios. Su principal valor radica en la posibilidad de articular y contrastar distintas capas de información, lo cual permite descubrir relaciones que no serían visibles de manera aislada. De esta forma, se identifican con claridad las zonas donde la densidad de establecimientos no se traduce en una experiencia positiva, alertando sobre la necesidad de intervenciones regulatorias, ajustes de capacidad de carga o mejoras en la calidad del servicio.

Los hallazgos derivados de estos cruces evidencian, por ejemplo, la existencia de clústeres de alojamientos económicos con alta concentración espacial pero baja reputación digital, lo cual constituye un desafío para la sostenibilidad de la imagen del destino. En paralelo, se reconocen segmentos premium con calificaciones elevadas que representan oportunidades estratégicas para diversificar la oferta y consolidar la posición de San Andrés en mercados de mayor poder adquisitivo. En este sentido, el tablero no solo convierte datos dispersos en información integrada, sino que facilita decisiones

tácticas inmediatas orientadas a la competitividad y al control de la experiencia del visitante.

Más allá de su función analítica, el dashboard se proyecta como una herramienta de planificación territorial y estratégica, ya que centraliza en una sola interfaz la información sobre la oferta, la reputación digital y la percepción de los visitantes. Esta integración permite sustentar decisiones relacionadas con la gestión del espacio turístico, la diversificación de productos y la innovación en servicios, además de respaldar estrategias de sostenibilidad al identificar áreas donde la presión turística es mayor. En consecuencia, el modelo trasciende la visualización de datos y se convierte en un componente clave para la gestión de un destino turístico inteligente, alineado con las recomendaciones de la literatura sobre gobernanza basada en datos, competitividad e innovación tecnológica.

### **8.2.2 Contribución del modelo a la gestión de un destino turístico inteligente**

El modelo implementado trasciende el ámbito descriptivo y se inserta en los lineamientos conceptuales de los Destinos Turísticos Inteligentes (DTI), al articular innovación tecnológica, sostenibilidad, gobernanza y experiencia del usuario en una plataforma de análisis integrada. En este sentido, el dashboard no solo ofrece un compendio de indicadores, sino que constituye una herramienta que facilita la transición hacia formas de gestión turística basadas en datos, coherentes con los principios destacados en la literatura especializada (Fuchs et al., 2014; Marine-Roig, 2019).

### **8.2.2.1 Gobernanza y Gestión eficiente**

En el eje de gobernanza y gestión eficiente, el tablero brinda evidencia para la formulación de políticas públicas sea más precisas, al mostrar de manera georreferenciada la concentración de la oferta, los patrones de satisfacción y las áreas críticas identificadas en los análisis de sentimiento. Esto responde a la necesidad, señalada en estudios previos, de que los DTI cuenten con sistemas de apoyo a la decisión que integren fuentes heterogéneas de información y transformen datos dispersos en conocimiento estratégico (Bustamante, 2019). Estos conocimientos permiten mejorar la participación social en la toma de decisiones, fomentar la transparencia y apertura de los servicios públicos y sociales a todos los actores del sector. (SEGITTUR, 2015, p. 149)

### **8.2.2.2 Innovación tecnológica**

Respecto a la innovación tecnológica, el modelo incorpora técnicas avanzadas de procesamiento de lenguaje natural (NLP) mediante el modelo multilingüe BERT, y las combina con entornos de visualización interactiva en Power BI. Esta integración permite analizar de manera automática las reseñas de los turistas en varios idiomas y transformarlas en indicadores accionables, como el Índice de Sentimiento Turístico (IST) y el grado de confiabilidad del análisis. Tal enfoque responde a lo planteado por Femenia-Serra et al. (2019), quienes destacan que la innovación tecnológica en los DTI no se limita a la adopción de herramientas digitales, sino a su aplicación efectiva para resolver problemas de gestión y mejorar la competitividad, cocreando a la vez valor para todos los actores del destino turístico.

### **8.2.2.3 Sostenibilidad y planificación territorial**

El modelo también contribuye a la sostenibilidad, al incluir métricas como la tasa de concentración geográfica de servicios turísticos y los mapas de calor que reflejan la densidad de establecimientos y su reputación digital. Estos indicadores permiten anticipar desequilibrios asociados a la capacidad de carga y planificar la redistribución de flujos turísticos, posibilitando la adopción de medidas regulatorias en materia ambiental y territorial. Como señalan Gössling & Michael Hall (2019), el uso de información en tiempo real y métricas de distribución espacial son claves para gestionar de manera equilibrada los recursos naturales, culturales y sociales, minimizando los riesgos y maximizando los beneficios de los actores involucrados. De igual forma, en línea con Marine-Roig (2019), se resalta la necesidad de que los gobiernos asuma un rol activo en la incorporación de estos resultados como insumo principal para la planificación y la gestión estratégica del destino. Así, el modelo responde al llamado de la literatura sobre turismo sostenible, que enfatiza la integración de indicadores ambientales y sociales con métricas de reputación digital, consolidando una visión holística para la gestión territorial (Fuchs et al., 2014; Gössling & Michael Hall, 2019)

### **8.2.2.4 Orientación al visitante y experiencia del turista**

Finalmente, el modelo sitúa al visitante en el centro de la gestión, priorizando indicadores derivados de la voz del usuario. El Net Promoter Score (NPS) y el IST permiten medir la satisfacción y la propensión a recomendar el destino, mientras que el análisis de reseñas textuales identifica patrones de percepción positiva y negativa. Esto

coincide con lo planteado por Marine-Roig (2019) y Femenia-Serra et al. (2019), quienes subrayan la importancia de integrar el contenido generado por los usuarios (UGC) en la gestión turística para captar la experiencia real del visitante y diseñar estrategias más personalizadas. En consecuencia, el dashboard no solo refleja la percepción de los turistas, sino que también ofrece insumos estratégicos para mejorar la experiencia del viaje y fortalecer la reputación digital del destino.

En síntesis, la contribución de este dashboard radica en su capacidad para operativizar los principios de los DTI en el contexto del Archipiélago de San Andrés. Al integrar variables como el SRD, el NPS, el IST, la segmentación obtenida con K-means y los indicadores de distribución de la oferta, el modelo genera un panorama integral que responde tanto a los retos de competitividad como a los de sostenibilidad del destino. De esta manera, se valida la pertinencia de combinar fuentes abiertas y técnicas avanzadas de análisis con entornos visuales interactivos, demostrando que la gestión turística inteligente es viable y transferible al contexto territorial específico del archipiélago.

## **9 Validación del modelo mediante el instrumento V de Aiken**

En coherencia con el tercer objetivo de la investigación, se procedió a realizar la validación del modelo de inteligencia de negocios propuesto a través del uso del instrumento estadístico conocido como V de Aiken, ampliamente utilizado para valorar la pertinencia y relevancia de ítems en estudios de corte aplicado. Este procedimiento permite cuantificar el grado de validez del modelo entre expertos en relación con diferentes ítems que miden las variables, indicadores y componentes utilizados, a partir de una escala que oscila entre 0 (desacuerdo total) y 1 (acuerdo total), lo que facilita estimar cuantitativamente la evidencia de validez basado en los juicios de los expertos.

El proceso de validación se llevó a cabo con la participación de cuatro expertos del sector turístico y académico del Archipiélago de San Andrés, quienes aportaron diferentes perspectivas complementarias al análisis. Entre los participantes se contó con un funcionario de la Secretaría de Turismo, una reconocida académica de la Universidad Nacional Sede Caribe con doctorado en sostenibilidad y con amplia experiencia en investigación en el sector turismo, un gerente de importante hotel de la Isla, y finalmente, una propietaria de apartamentos turísticos. La selección de este grupo buscó garantizar un equilibrio entre experiencia institucional, conocimiento académico y práctica empresarial, con el fin de validar la claridad, pertinencia y relevancia del modelo desde una visión integral.

El instrumento incluyó un conjunto de variables e indicadores organizados en grupos de preguntas que recogen aspectos centrales del modelo, tales como la oferta turística, la satisfacción del visitante, la percepción del destino y los principios asociados a los

Destinos Turísticos Inteligentes (DTI). Cada experto evaluó la claridad, coherencia y relevancia de los ítems, de manera que la aplicación del índice de V de Aiken permitió consolidar un valor promedio que refleja el grado de consenso respecto a la pertinencia del modelo.

A continuación, se presentan los resultados obtenidos, los cuales constituyen evidencia empírica de la validez de contenido del modelo propuesto y aportan solidez al marco metodológico planteado, al demostrar que los indicadores diseñados resultan comprensibles, aplicables y relevantes para el contexto del Archipiélago de San Andrés.

La validación del modelo mediante la V de Aiken evidencia un alto nivel de acuerdo entre los cuatro expertos en las tres dimensiones evaluadas, claridad, pertinencia y relevancia. En la tabla de síntesis por serie de variables (ver Tabla 3) se observa que los coeficientes de Aiken se sitúan en niveles elevados en la mayoría de los bloques, con valores superiores a 0,86 en cinco de las seis dimensiones.

**Tabla 3 - Síntesis por serie de variables V de Aiken**

SERIE DE VARIABLES	EVALUADOR1	EVALUADOR2	EVALUADOR3	EVALUADOR4	V DE AIKEN
A. Calidad técnica del modelo de Inteligencia de Negocios	0,81	0,90	0,81	0,90	<b>0,86</b>
B. Relevancia para la toma de decisiones	0,93	0,93	1,00	1,00	<b>0,96</b>
C. Alineación con principios de un Destino Turístico Inteligente (DTI)	1,00	1,00	1,00	0,96	<b>0,99</b>
D. Usabilidad y accesibilidad del Dashboard	0,92	1,00	1,00	1,00	<b>0,98</b>
E. Planificación de la Calidad	1,00	1,00	1,00	1,00	<b>1,00</b>
F. Impacto social y territorial del modelo	0,67	0,83	0,83	0,83	<b>0,79</b>

*Fuente: Tabla realizada por el autor*

En términos específicos, la Calidad técnica del modelo alcanzó un índice de 0,86, evidenciando que las fuentes de datos, los indicadores y la estructura general del modelo son considerados adecuados y consistentes por los evaluadores. La Relevancia para la toma de decisiones obtuvo 0,96, lo que confirma que los resultados del dashboard ofrecen insumos estratégicos de gran utilidad tanto para actores públicos como privados. El bloque de Alineación con los principios de un Destino Turístico Inteligente (DTI) se posicionó con un valor de 0,99, el más alto de la evaluación, validando la coherencia del modelo con los marcos conceptuales de gobernanza, innovación y sostenibilidad.

En lo referente a la Usabilidad y accesibilidad del dashboard, el coeficiente alcanzó 0,98, mostrando que las visualizaciones, filtros e interactividad son percibidos como claros y fáciles de utilizar. Asimismo, la Planificación de la calidad obtuvo el valor máximo posible (1,00), indicando un consenso absoluto en torno a la capacidad del modelo para ser escalable, replicable y actualizado en el tiempo. Por último, el bloque de Impacto social y territorial presentó un valor más moderado (0,79), aunque todavía dentro del rango de aceptabilidad, sugiriendo que si bien el modelo contribuye a visibilizar el patrimonio y fortalecer el turismo local, este es un aspecto que podría enriquecerse con mayor integración comunitaria y territorial.

En conjunto, los resultados de la Tabla 3 permiten concluir que el modelo presenta una validez de contenido muy alta, al ser percibido como claro, pertinente y relevante por parte de los distintos perfiles expertos consultados (funcionarios, empresarios y académicos). De esta manera, el modelo puede considerarse pertinente y confiable para apoyar procesos de planificación estratégica, competitividad y gobernanza turística en el Archipiélago de San Andrés.

## 10 Discusión

La discusión de los resultados constituye un espacio para contrastar los hallazgos empíricos derivados del modelo de inteligencia de negocios propuesto con los referentes conceptuales y metodológicos identificados en la revisión sistemática de literatura. Este ejercicio permite no solo evaluar la pertinencia de las técnicas y variables aplicadas en el contexto del Archipiélago de San Andrés, sino también analizar los alcances, limitaciones y aportes de la investigación frente al cuerpo de conocimiento existente en torno a los Destinos Turísticos Inteligentes (DTI) y la gestión basada en datos.

De esta manera, la discusión se aborda desde qué medida el modelo propuesto se alinea con los fundamentos del Destino Turístico Inteligente desarrollados en el marco teórico: la infraestructura TIC como soporte del ecosistema del destino, la integración de actores, flujos de conocimiento mediante plataformas digitales y el uso de analítica/BI sobre UGC para orientar la toma de decisiones de los actores principales del sector que conforman la oferta y demanda de los servicios turísticos (Del Chiappa & Baggio, 2015; Marine-Roig & Anton Clavé, 2015; Gretzel et al., 2015; Zeng et al., 2020). De igual forma, también se valida en qué medida el modelo planteado responde a los desafíos de competitividad, planificación territorial, gobernanza y sostenibilidad dentro de un destino turístico inteligente. (Ivars-Baidal & Rebollo, 2019)

En este contexto, los resultados obtenidos demuestran que la aplicación del modelo trasciende la simple visualización de indicadores, ya que se convierte en un instrumento operativo que viabiliza la toma de decisiones estratégicas basadas en evidencia. A través de los dashboards, los actores públicos y privados pueden identificar patrones de

satisfacción, segmentar mercados prioritarios, monitorear la reputación digital y anticipar riesgos asociados a la turistificación o a la capacidad de carga del destino. Estas capacidades analíticas fortalecen la planificación y la gobernanza, elementos centrales del paradigma de los destinos turísticos inteligentes. En consecuencia, el modelo contribuye a la conversión progresiva del Archipiélago de San Andrés en un destino inteligente, al integrar información proveniente de datos colaborativos y transformarla en conocimiento útil para orientar políticas, estrategias de marketing y decisiones empresariales sostenibles.

Uno de los hallazgos más relevantes corresponde a la caracterización de la oferta turística del archipiélago. Los resultados del dashboard evidenciaron una alta concentración de hoteles con gran peso en la reputación digital y una significativa presencia de viviendas turísticas y apartamentos registrados en plataformas como Airbnb y TripAdvisor, que en conjunto representan más del 77% del total de alojamientos analizados. Este comportamiento confirma lo señalado por Bustamante (2019) y Fuchs et al. (2014), quienes resaltan que la integración de datos estructurados y no estructurados en arquitecturas híbridas permite comprender la fragmentación de la oferta en destinos turísticos. En el caso de San Andrés, esta combinación refleja una tensión entre los establecimientos consolidados y los emergentes, lo cual subraya la pertinencia de aplicar técnicas de inteligencia de negocios para capturar tipologías heterogéneas y facilitar la toma de decisiones tanto para el sector privado como para los reguladores.

En línea con los resultados obtenidos por Cerdá-Mansilla et al. (2022), los hallazgos de este estudio también sirven como alertas tempranas para identificar procesos de turistificación en el territorio. Esto se evidencia clúster geográfico denominado “Zona

Residencial Económico”, que concentra más de 146 establecimientos ubicados en barrios tradicionalmente residenciales. Tal patrón no solo genera un impacto directo en la configuración del mercado turístico, al incrementar la densidad y diversidad de la oferta, sino que también plantea implicaciones críticas sobre la planificación territorial, la sostenibilidad y la gobernanza local. En este sentido, el modelo de inteligencia de negocios permite visibilizar cómo la expansión de alojamientos en áreas residenciales puede tensionar el equilibrio entre usos del suelo, calidad de vida de los habitantes y desarrollo turístico, aportando evidencia empírica para la toma de decisiones regulatorias y estratégicas. Este resultado confirma lo señalado por Wang et al. (2013), en cuanto a que la construcción de plataformas inteligentes de gestión turística no solo otorga al destino una ventaja competitiva, sino que fortalece su sostenibilidad al optimizar la asignación de recursos y mejorar la efectividad de la gestión.

En cuanto a la percepción y satisfacción de los visitantes, los indicadores derivados del modelo, el Índice de Sentimiento Turístico (IST) y el Net Promoter Score (NPS), reflejaron una reputación mayoritariamente positiva, con un 77% de reseñas clasificadas como favorables y un NPS de 28. Estos resultados confirman lo planteado por Marine-Roig (2019) y Xiang et al. (2017), quienes sostienen que el contenido generado por los usuarios (UGC) constituye una fuente central para evaluar la reputación digital de un destino turístico. Sin embargo, la identificación de un núcleo de reseñas negativas y detractores también se alinea con las advertencias de Kim et al. (2017), quienes señalan la necesidad de gestionar activamente las valoraciones adversas, pues estas no solo afectan la experiencia individual, sino que pueden tener efectos multiplicadores sobre la imagen global del destino.

De esta manera, los hallazgos empíricos de este estudio se alinean con la transformación señalada recientemente por Marine-Roig (2019), en la que la expansión masiva del UGC ha generado un verdadero cambio de paradigma en el análisis de la imagen turística, desplazando las encuestas y entrevistas tradicionales como principal método de recolección de percepciones. En este sentido, el modelo confirma que actualmente es posible evaluar de forma sistemática la reputación y la experiencia del visitante a partir de datos digitales abiertos, lo que amplía significativamente las capacidades de gestión y monitoreo de los destinos turísticos inteligentes.

A su vez, la investigación refuerza las implicaciones prácticas señaladas por Bagherzadeh et al. (2021), quienes destacan la importancia crítica de emplear métodos inteligentes y automatizados para la extracción de sentimientos en grandes volúmenes de datos generados por los clientes. A diferencia de los enfoques tradicionales basados en diccionarios genéricos, los modelos de aprendizaje profundo, como el utilizado en este estudio (BERT multilingüe), permiten una interpretación más precisa y contextual del lenguaje, facilitando una lectura oportuna y accionable de la voz del visitante. Esta capacidad resulta especialmente valiosa en destinos con alta afluencia turística y diversidad cultural, como San Andrés, donde conviven visitantes de múltiples nacionalidades y lenguas. En este sentido, el análisis multilingüe reveló la relevancia estratégica del mercado brasileño, que concentra un volumen significativo de reseñas en portugués y muestra altos niveles de satisfacción. La posibilidad de identificar estos patrones diferenciados a partir de UGC confirma que el uso de NLP multilingüe no solo mejora la precisión del análisis de sentimientos, sino que además ofrece insumos clave para segmentar la gestión de mercados internacionales y diseñar estrategias de fidelización específicas.

En consonancia con lo planteado por Gretzel et al. (2015a, 2015b), el presente estudio también contribuye a reducir la brecha existente en torno al uso de Big Data en turismo inteligente, al demostrar cómo un proceso híbrido de minería de texto puede ser operacionalizado en un contexto insular. Esta aproximación metodológica se alinea con lo señalado por Kim et al. (2017), quienes destacan el valor de integrar perspectivas de la gestión turística y de las ciencias computacionales para generar resultados más significativos. En este caso, la combinación del análisis de reseñas (UGC) con algoritmos de procesamiento de lenguaje natural permitió derivar indicadores como el Índice de Sentimiento Turístico (IST) y el grado de confiabilidad del modelo, aportando una ruta replicable para investigaciones futuras. De igual forma, los hallazgos empíricos corroboran lo evidenciado por Xiang et al. (2015), al mostrar que la polaridad textual de las reseñas mantiene una correlación consistente con las calificaciones numéricas asignadas por los usuarios. Finalmente, el predominio de reseñas positivas (77%) coincide con la tesis de Dodds et al. (2015) y Xiang et al. (2017) sobre el sesgo universal de positividad presente en el lenguaje humano, lo que sugiere que la reputación digital de los destinos tiende a construirse sobre narrativas optimistas, aunque no exentas de críticas que requieren una gestión activa.

No obstante, si bien la predominancia de comentarios positivos confirma el sesgo de positividad señalado en la literatura (Dodds et al., 2015; Xiang et al., 2017), este fenómeno plantea desafíos metodológicos que deben ser considerados en investigaciones futuras. La tendencia del lenguaje humano a expresarse en términos favorables puede enmascarar focos de insatisfacción que, aunque minoritarios en volumen, poseen un alto impacto sobre la percepción global del destino. En este sentido,

resulta fundamental avanzar hacia el uso de algoritmos más sofisticados capaces de detectar matices semánticos complejos, como la ironía, la ambigüedad o la subjetividad extrema, elementos que suelen pasar desapercibidos en modelos convencionales de análisis de sentimiento (Kim et al., 2017). Incorporar estas mejoras permitiría no solo obtener una representación más precisa de la experiencia del visitante, sino también fortalecer la capacidad del modelo de inteligencia de negocios para anticipar riesgos reputacionales y orientar estrategias de gestión proactiva en destinos turísticos inteligentes.

Por último, otro hallazgo relevante se relaciona con la estacionalidad en la satisfacción turística. El modelo evidenció que, en los meses de mayor afluencia, como enero, la calificación promedio de los visitantes tiende a disminuir, mientras que, en temporadas de menor ocupación, como octubre, las valoraciones resultan más favorables. Este comportamiento confirma lo planteado por Gössling y Hall (2019), quienes destacan que la saturación turística incrementa la presión sobre la capacidad de carga del destino y repercute negativamente en la calidad percibida de los servicios. En consecuencia, los resultados de este estudio refuerzan la necesidad de que la gestión de flujos turísticos se aborde desde una perspectiva de planificación territorial y sostenibilidad, y no únicamente desde criterios de competitividad económica, ya que solo así es posible equilibrar la experiencia del visitante con la conservación de los recursos locales y el bienestar de la comunidad residente.

## **10.1 Aportes del modelo a la literatura**

En términos de contribución académica, esta investigación aporta evidencia empírica sobre la viabilidad de la inteligencia de negocios como herramienta estratégica para los destinos turísticos inteligentes en contextos insulares. La revisión de literatura realizada evidenció que la mayoría de los estudios provienen de Europa, Asia y Estados Unidos (Fuchs et al., 2014; Höpken et al., 2021; Mariani & Baggio, 2022), incluyendo aplicaciones en territorios insulares europeos como el caso de Hernández-Mogollón et al. (2020) en las Islas Canarias, enfocado en la imagen gastronómica del destino. Dentro de los artículos identificados, únicamente se encontró una aplicación en el contexto colombiano: Santana (2019) en la ciudad de Santa Marta. En este sentido, el presente trabajo constituye un aporte diferencial al trasladar dichas metodologías al Caribe insular, un territorio con dinámicas socioeconómicas y ambientales particulares, ampliando la base empírica desde los hallazgos de la revisión.

Un segundo aporte se centra en la integración de innovaciones tecnológicas en el pipeline de datos (Airflow, PostgreSQL, MongoDB y BERT multilingüe), lo cual se alinea con lo planteado por Bagherzadeh et al. (2020) respecto a la importancia de diseñar métodos generalizables y aplicables en distintos contextos del turismo. Los resultados obtenidos y los indicadores clave visualizados en el dashboard muestran que las herramientas implementadas no solo son técnicamente robustas, sino también comprensibles y útiles para la toma de decisiones estratégicas. Esto refuerza su potencial como arquitectura replicable y escalable, consolidándose en la literatura como un aporte metodológico relevante en el campo de la inteligencia de negocios aplicada al turismo.

Finalmente, los resultados de la validación identificaron un punto de tensión en el impacto social y territorial ( $V = 0,79$ ), lo cual constituye un aporte crítico: la evidencia empírica indica que, si bien los modelos de inteligencia de negocios fortalecen la competitividad y reputación digital, aún existe un vacío en la integración de indicadores socio-territoriales. Este hallazgo abre una línea de investigación poco explorada en la literatura y refuerza la necesidad de ampliar los marcos de medición hacia dimensiones comunitarias, culturales y ambientales (Marine-Roig & Anton Clavé, 2015; SEGITTUR, 2015).

## **10.2 Limitaciones**

El estudio presenta limitaciones que condicionan la validez de los resultados. En primer lugar, se debe reconocer la dependencia de fuentes secundarias (TripAdvisor y Airbnb) como insumos primarios del modelo. Tal como advierten Xiang et al. (2017), los datos provenientes de reseñas en línea pueden estar sujetos a sesgos de autoselección, ya que tienden a participar usuarios con experiencias particularmente positivas o negativas, lo que podría distorsionar la representatividad de la percepción turística.

En segundo lugar, aunque el modelo integra técnicas avanzadas de procesamiento de lenguaje natural, la confiabilidad del análisis de sentimiento (61%) refleja la dificultad de aplicar modelos multilingües como BERT en contextos de alta heterogeneidad lingüística. Estudios recientes (Frigau et al., 2024) coinciden en que, si bien BERT ofrece ventajas en la detección de matices semánticos, aún requiere entrenamientos específicos para dominios como el turismo y para lenguas con variaciones regionales.

Una tercera limitación está relacionada con el número de jueces en la validación del modelo ( $n = 4$ ), lo cual, aunque aceptable metodológicamente, puede restringir la precisión de los estimadores de V de Aiken. En consecuencia, futuros trabajos deberían ampliar el panel de expertos e incorporar rondas adicionales de evaluación u otras metodologías para complementar este tipo de análisis.

Por último, la dimensión social y territorial, que obtuvo la valoración más baja en la validación, evidencia la necesidad de incorporar métricas específicas que midan con mayor claridad la redistribución de beneficios hacia comunidades locales, la preservación cultural y los efectos del turismo sobre la capacidad de carga. Este aspecto constituye un reto abierto para complementar los modelos de BI con métricas ambientales y sociales (Gössling & Hall, 2019).

La investigación demuestra que la aplicación de un modelo de inteligencia de negocios en el Archipiélago de San Andrés es viable y pertinente como herramienta estratégica para la gestión de un Destino Turístico Inteligente (DTI). Los resultados evidencian que, a través de la integración de datos estructurados y no estructurados provenientes del UGC, es posible caracterizar la oferta turística, monitorear la reputación digital y detectar fenómenos críticos como la turistificación, la estacionalidad en la satisfacción y la relevancia de mercados internacionales como el brasileño. Estos hallazgos confirman lo señalado en la literatura sobre la capacidad de las plataformas inteligentes para fortalecer la competitividad y sostenibilidad de los destinos mediante el uso estratégico de los datos.

Al mismo tiempo, la validación del modelo mediante V de Aiken respalda la pertinencia de indicadores como el Score de Reputación Digital (SRD), el Índice de Sentimiento Turístico (IST) y el Net Promoter Score (NPS), aunque se reconocen limitaciones vinculadas con los sesgos del UGC, la confiabilidad aún incipiente del análisis multilingüe y el tamaño de la muestra de expertos. En conjunto, este trabajo no solo adapta metodologías de BI a un contexto insular latinoamericano, sino que también abre nuevas líneas de investigación para consolidar modelos que integren competitividad, gobernanza y sostenibilidad en la gestión de destinos turísticos inteligentes.

## **11 Conclusiones y Trabajo Futuro**

La investigación tuvo como propósito diseñar y validar un modelo de inteligencia de negocios para la gestión de un Destino Turístico Inteligente en el Archipiélago de San Andrés. Con base en los resultados obtenidos y la validación empírica realizada, se presentan las siguientes conclusiones, organizadas en torno a los objetivos, hipótesis y pregunta de investigación:

De manera general, los objetivos específicos de la investigación fueron cumplidos en su totalidad. En primer lugar, se analizó la literatura científica existente, identificando las principales variables, enfoques metodológicos y modelos previos de inteligencia de negocios aplicados al turismo y a los destinos turísticos inteligentes, lo que permitió construir un marco conceptual sólido y actualizado. En segundo lugar, se desarrolló un modelo de inteligencia de negocios orientado a la toma de decisiones estratégicas, integrando técnicas de extracción, procesamiento y análisis de datos colaborativos bajo una arquitectura tecnológica replicable. Finalmente, se validó la utilidad y pertinencia del modelo como herramienta de apoyo a la gestión del destino, demostrando su capacidad para generar indicadores clave (KPI) que facilitan decisiones informadas entre los actores del sector y promueven una gestión turística más inteligente, basada en evidencia y alineada con los principios de sostenibilidad y competitividad del Archipiélago de San Andrés.

Paralelamente, el modelo nos permite confirmar la hipótesis 1, al evidenciar una alta concentración de servicios de alojamiento en el archipiélago, con un peso significativo de hoteles y una representatividad aún mayor de viviendas y apartamentos turísticos, que en

conjunto superan el 77% del total de establecimientos. Este hallazgo demuestra la relevancia de los establecimientos de menor capacidad en la composición de la oferta, lo que genera tensiones entre las tipologías consolidadas y las emergentes, y plantea desafíos en materia de regulación y planificación territorial.

El modelo confirmó parcialmente la hipótesis 2, al evidenciar que los establecimientos turísticos de menor capacidad presentan un nivel de satisfacción elevado, con un promedio de calificación de 4,44 sobre 5 y un Score de Reputación Digital (SRD) de 89,59. Atributos como comunicación, check-in y limpieza alcanzaron las valoraciones más altas, consolidando a este segmento como un referente de calidad percibida. No obstante, solo un 14,65% de los alojamientos fue clasificado como “excelente”, lo que indica que, si bien la satisfacción general es alta, persisten diferencias internas en el desempeño que limitan que todos los establecimientos pequeños alcancen niveles sobresalientes, planteando así oportunidades de mejora en la estandarización del servicio.

La hipótesis 3 también fue confirmada: los análisis de UGC evidenciaron un nivel de satisfacción relativamente alto en las principales atracciones turísticas del Archipiélago, como playas y actividades recreativas. Estas valoraciones positivas se constituyen en un activo fundamental para la promoción del destino, aunque se advierte que la saturación en temporadas de alta afluencia tiende a afectar la experiencia percibida, lo que requiere medidas de planificación territorial orientadas a la sostenibilidad.

La hipótesis 4 se validó plenamente, ya que las percepciones globales sobre San Andrés y sus servicios de alojamiento resultaron mayoritariamente positivas. El 77% de

las reseñas analizadas fueron favorables y el Net Promoter Score alcanzó un valor de 28, confirmando que el destino mantiene una reputación digital sólida. Sin embargo, la existencia de un núcleo de detractores y reseñas negativas pone de manifiesto la importancia de implementar estrategias de gestión activa de la reputación digital para minimizar riesgos y mantener la competitividad.

La aplicación del coeficiente V de Aiken permitió validar la pertinencia de indicadores como el Score de Reputación Digital (SRD), el Índice de Sentimiento Turístico (IST) y el NPS, con valores superiores a 0,86 en cinco de las seis dimensiones evaluadas. Esto confirma que las métricas propuestas son relevantes y aplicables en contextos turísticos insulares, aunque la dimensión social y territorial continúa siendo un área de mejora.

Finalmente, en relación con la pregunta central ¿Cómo puede el uso de herramientas de Inteligencia de Negocios contribuir a fundamentar la toma de decisiones y promover la gestión inteligente del destino turístico en San Andrés?, los resultados muestran que el uso de Inteligencia de Negocios permite integrar datos heterogéneos, generar indicadores estratégicos y visualizar patrones críticos en torno a la oferta, la satisfacción y la percepción del destino. El modelo desarrollado no solo facilita la toma de decisiones informadas, sino que también refuerza la gobernanza, la competitividad y la sostenibilidad del archipiélago al proporcionar evidencia empírica para anticipar riesgos y orientar intervenciones de política pública y empresarial.

De cara a investigaciones futuras, se plantea la posibilidad de ampliar la validación del modelo en contextos comparativos con otros destinos turísticos del Caribe, integrando datos de otras plataformas como Google Business Reviews y redes sociales, registros

oficiales e indicadores de sostenibilidad. Asimismo, resulta pertinente analizar la incidencia de la Inteligencia Artificial en la optimización de los indicadores de desempeño de los establecimientos turísticos, con el propósito de generar predicciones que mejoren la toma de decisiones basadas en datos. Otra línea emergente radica en el perfeccionamiento de los modelos de análisis de sentimiento multilingüe mediante técnicas híbridas de aprendizaje profundo y embeddings contextuales (BERT, LASER, MUSE), orientadas a generar sistemas de monitoreo en tiempo real del comportamiento turístico. Finalmente, se sugiere promover trabajos de grado derivados que profundicen en: (1) la medición del impacto social y ambiental del turismo mediante analítica avanzada; (2) la gobernanza de datos en destinos inteligentes; y (3) la implementación piloto de plataformas regionales de inteligencia turística para el Caribe colombiano, articulando academia, gobierno y sector privado

## Referencias

- 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), 210–238. <https://doi.org/10.1108/JHTT-02-2020-0034>
- Biderman, C., Mentzingen de Mendonça, M., Silva Mello, P. A., Hiromi Oshiro, C., & Foditshc, N. (2021). *Big Data para el desarrollo urbano sostenible* (M. Bouskela, M. Casseb, P. Zambrano-Barragán, & H. Elnir, Eds.). Banco Interamericano para el Desarrollo.
- Bustamante, A. A. (2019). *Análisis de datos colaborativos e inteligencia de negocio: aplicación al sector turístico*. Universitat Politècnica de València.
- Cámara de Comercio de San Andrés. (2024). *ESTUDIOS ECONÓMICOS 2024*. <https://camarasai.org/investigaciones-y-publicaciones/notas-economicas/>
- Cerdá-Mansilla, E., Rubio, N., García-Henche, B., & Campo, S. (2022). Airbnb and the touristification of neighborhoods in cities: An analysis of neighborhood segmentation of extra-hotel accommodation in Madrid. *Investigaciones Turísticas*, 23, 210–238. <https://doi.org/10.14198/INTURI2022.23.10>
- Clarivate. (2025). *Web of Science Core Collection*.
- Cruz-Chóez, A. M. (2022). Big Data desde las redes sociales. *593 Digital Publisher CEIT*, 7(4–1), 626–639. <https://doi.org/10.33386/593dp.2022.4-1.1239>
- Del Chiappa, G., & Baggio, R. (2015). Knowledge transfer in smart tourism destinations: Analyzing the effects of a network structure. *Journal of Destination Marketing and Management*, 4(3), 145–150. <https://doi.org/10.1016/j.jdmm.2015.02.001>

- Del Vecchio, P., & Passiante, G. (2017). Is tourism a driver for smart specialization? Evidence from Apulia, an Italian region with a tourism vocation. *Journal of Destination Marketing & Management*, 6, 163–165.
- Díaz-González, S., Torres, J. M., Parra-López, E., & Aguilar, R. M. (2022). Strategic technological determinant in smart destinations obtaining an automatic classification of the quality of the destination. *Industrial Management & Data Systems*, 122(10), 2299–2330. <https://doi.org/10.1108/IMDS-10-2021-0640>
- Femenia-Serra, F., Neuhofer, B., & Ivars-Baidal, J. A. (2019). Towards a conceptualisation of smart tourists and their role within the smart destination scenario. *Service Industries Journal*, 39(2), 109–133. <https://doi.org/10.1080/02642069.2018.1508458>
- Frigau, L., Contu, G., Ortu, M., & Carta, A. (2024). Gauging Airbnb review sentiments and critical key-topics by small area estimation. *Statistical Methods and Applications*, 33(4), 1145–1170. <https://doi.org/10.1007/s10260-024-00764-y>
- Fuchs, M., Höpken, W., & Lexhagen, M. (2014). Big data analytics for knowledge generation in tourism destinations - A case from Sweden. *Journal of Destination Marketing and Management*, 3(4), 198–209. <https://doi.org/10.1016/j.jdmm.2014.08.002>
- Gobernación Departamento Archipiélago de San Andrés. (2024). *Plan de Desarrollo Territorial 2024-2027*.
- Gössling, S., & Michael Hall, C. (2019). Sharing versus collaborative economy: how to align ICT developments and the SDGs in tourism? *Journal of Sustainable Tourism*, 27(1), 74–96. <https://doi.org/10.1080/09669582.2018.1560455>

- Höpken, W., Eberle, T., Fuchs, M., & Lexhagen, M. (2021). Improving Tourist Arrival Prediction: A Big Data and Artificial Neural Network Approach. *Journal of Travel Research, 60*(5), 998–1017. <https://doi.org/10.1177/0047287520921244>
- Ivars-Baidal, J. A., & Rebollo, J. F. V. (2019). Tourism planning in Spain. From traditional paradigms to new approaches: Smart tourism planning. *Boletín de La Asociación de Geógrafos Españoles, 82*. <https://doi.org/10.21138/bage.2765>
- Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters, 31*(8), 651–666. <https://doi.org/10.1016/J.PATREC.2009.09.011>
- James Cruz, J. L. (2013). El Turismo como estrategia de desarrollo económico: El caso de las islas de San Andrés y Providencia. *Cuadernos Del Caribe, 16*(1), 37–55.
- James, J., Saleme, M. A., Forbes, C., Romero, A., Contreras, H., & Gelvis, L. (2024). Evidencias de saturación turística en la isla de San Andrés: percepción de los residentes. *Études Caribéennes, 57–58*. <https://doi.org/10.4000/etudescaribeennes.30355>
- Jankova, L., Auzina, A., & Zvirbule, A. (2023). Regional smart cultural tourism destinations in a region of Latvia. In *Worldwide Hospitality and Tourism Themes* (Vol. 15, Issue 5, pp. 507–516). Emerald Publishing. <https://doi.org/10.1108/WHATT-06-2023-0082>
- Joyanes, L. (2019). *Inteligencia De Negocios y Analítica De Datos* (ALFAOMEGA, Ed.; 1st ed.).
- Kim, K., Park, O. jung, Yun, S., & Yun, H. (2017). What makes tourists feel negatively about tourism destinations? Application of hybrid text mining methodology to smart destination management. *Technological Forecasting and Social Change, 123*, 362–369. <https://doi.org/10.1016/j.techfore.2017.01.001>

- Köseoglu, M. A., Mehraliyev, F., Altin, M., & Okumus, F. (2020). Competitor intelligence and analysis (CIA) model and online reviews: integrating big data text mining with network analysis for strategic analysis. *Tourism Review*, 76(3), 529–552. <https://doi.org/10.1108/TR-10-2019-0406>
- Ma, Y., Xiang, Z., Du, Q., & Fan, W. (2018). Effects of user-provided photos on hotel review helpfulness: An analytical approach with deep learning. *International Journal of Hospitality Management*, 71, 120–131. <https://doi.org/10.1016/j.ijhm.2017.12.008>
- Macias, H. A. (2016). Colombian Research in Accounting regarding Colciencias' New Assessment Criteria. *En-Contexto*, 4(4), 49–58.
- Mariani, M., Baggio, R., Fuchs, M., & Höepken, W. (2018). Business intelligence and big data in hospitality and tourism: a systematic literature review. In *International Journal of Contemporary Hospitality Management* (Vol. 30, Issue 12, pp. 3514–3554). Emerald Group Holdings Ltd. <https://doi.org/10.1108/IJCHM-07-2017-0461>
- Marine-Roig, E. (2019). Destination image analytics through traveller-generated content. *Sustainability (Switzerland)*, 11(12). <https://doi.org/10.3390/su11123392>
- Marine-Roig, E., & Anton Clavé, S. (2015). Tourism analytics with massive user-generated content: A case study of Barcelona. *Journal of Destination Marketing and Management*, 4(3), 162–172. <https://doi.org/10.1016/j.jdmm.2015.06.004>
- Marine-Roig, E., Ferrer-Rosell, B., Daries, N., & Cristobal-Fransi, E. (2019). Measuring gastronomic image online. *International Journal of Environmental Research and Public Health*, 16(23). <https://doi.org/10.3390/ijerph16234631>
- Pérez, C. B., Rodríguez, L.-F., Castro, L. A., & Salazar, G. (2018). *Inteligencia de negocios: un enfoque para la toma de decisiones estratégicas en las organizaciones*. (Primera edición). Pearson Educación de México. <https://www-ebooks7-24-com.bdbiblioteca.universidadean.edu.co/?il=21995>

Pons-Vives, P. J., Morro-Ribot, M., Mulet-Forteza, C., & Valero, O. (2022). An Application of Ordered Weighted Averaging Operators to Customer Classification in Hotels.

*Mathematics*, 10(12). <https://doi.org/10.3390/math10121987>

Prayag, G., Hosany, S., & Odeh, K. (2013). The role of tourists' emotional experiences and satisfaction in understanding behavioral intentions. *Journal of Destination Marketing & Management*, 2(2), 118–127.

<https://doi.org/10.1016/J.JDMM.2013.05.001>

Rodríguez, G., Serafín, A., & Martínez, C. (2020). Diseño de la estrategia de gestión del destino turístico regional Villa Clara (Cuba). *Dos Algarves: A Multidisciplinary e-Journal*, 36, 75–96. <https://doi.org/10.18089/damej.2020.36.5>

<https://doi.org/10.18089/damej.2020.36.5>

Rodríguez Jiménez, G., & Martínez Martínez, C. C. (2022). Turismo responsable:

Propuesta para gestionar destinos turísticos regionales en la etapa post-COVID-19.

*Revista Universidad y Sociedad*, 14, 128–136. [https://orcid.org/0000-0002-2226-](https://orcid.org/0000-0002-2226-0677)

[0677](https://orcid.org/0000-0002-2226-0677)

Rodríguez Jiménez, G., & Cristóbal Martínez Martínez, C. (2022). Turismo responsable:

Propuesta para gestionar destinos turísticos regionales en la etapa post-COVID-19.

*Revista Universidad y Sociedad*, 14(1), 128–136. [https://orcid.org/0000-0002-2226-](https://orcid.org/0000-0002-2226-0677)

[0677](https://orcid.org/0000-0002-2226-0677)

Santana, S. M. (2019). *Desarrollo de métodos para la extracción y análisis de datos de fuentes públicas: estudio de la plataforma Airbnb*. Universidad de la Laguna.

SEGITTUR. (2015). *Informe destinos turísticos inteligentes: construyendo el futuro*.

[https://www.thinktur.org/media/Libro-Blanco-Destinos-Tursticos-Inteligentes-](https://www.thinktur.org/media/Libro-Blanco-Destinos-Tursticos-Inteligentes-construyendo-el-futuro.pdf)

[construyendo-el-futuro.pdf](https://www.thinktur.org/media/Libro-Blanco-Destinos-Tursticos-Inteligentes-construyendo-el-futuro.pdf)

Singh, P. (2019). Learn PySpark. In *Learn PySpark*. Apress. [https://doi.org/10.1007/978-](https://doi.org/10.1007/978-1-4842-4961-1)

[1-4842-4961-1](https://doi.org/10.1007/978-1-4842-4961-1)

- Vajirakachorn, T., & Chongwatpol, J. (2017). Application of business intelligence in the tourism industry: A case study of a local food festival in Thailand. *Tourism Management Perspectives*, 23, 75–86. <https://doi.org/10.1016/j.tmp.2017.05.003>
- Waltman, L. (2016). A review of the literature on citation impact indicators. In *Journal of Informetrics* (Vol. 10, Issue 2, pp. 365–391). Elsevier Ltd. <https://doi.org/10.1016/j.joi.2016.02.007>
- Wang, D., Li, X., & Li, Y. (2013). China's "smart tourism destination" initiative: A taste of the service-dominant logic. *Journal of Destination Marketing and Management*, 2(2), 59–61. <https://doi.org/10.1016/j.jdmm.2013.05.004>
- World Tourism Organization. (2024). International tourism to reach pre-pandemic levels in 2024. *UNWTO World Tourism Barometer*, 22(1). [www.unwto.org/market-intelligence](http://www.unwto.org/market-intelligence).
- Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management*, 58, 51–65. <https://doi.org/10.1016/j.tourman.2016.10.001>
- Yannacopoulou, A., & Kallinikos, K. (2025). Measuring Destination Image Using AI and Big Data: Kastoria's Image on TripAdvisor. *Societies*, 15(1). <https://doi.org/10.3390/soc15010005>
- Zeng, D., Tim, Y., Yu, J., & Liu, W. (2020). Actualizing big data analytics for smart cities: A cascading affordance study. *International Journal of Information Management*, 54. <https://doi.org/10.1016/j.ijinfomgt.2020.102156>

Propuesta de un modelo de inteligencia de negocios para la toma de decisiones estratégica en la gestión de un destino turístico inteligente.

126