

Cómo citar este trabajo: Vega-Sampayo, Y., Olivero-Vega, E., Villarreal González, R., Cantillo-Molina, S. & Álvarez-Royero, M. (2027). Uso de Inteligencia Artificial para mejorar la oferta turística: Modelos de segmentación en la gestión y planificación del sector turístico. *Journal of Tourism Analysis*, 34(1), 1–22. <https://doi.org/10.53596/13203s86>

Uso de Inteligencia Artificial para mejorar la oferta turística: Modelos de segmentación en la gestión y planificación del sector turístico

Using artificial intelligence to improve tourism offerings: segmentation models in tourism sector management and planning

Vega-Sampayo, Yolanda¹

Olivero-Vega, Enohemit²

Villarreal González, Reynaldo³

Cantillo-Molina, Steffen⁴

Álvarez-Royero, Mateo⁵

¹Universidad Simón Bolívar, Barranquilla, Colombia | yolanda.vega@unisimon.edu.co

²Universidad Simón Bolívar, Barranquilla, Colombia | enohemit.olivero@unisimon.edu.co

³Universidad Simón Bolívar, Barranquilla, Colombia | Reynaldo.villareal@unisimon.edu.co

⁴Universidad Simón Bolívar, Barranquilla, Colombia | steffen.cantillo@unisimon.edu.co

⁵Universidad Simón Bolívar, Barranquilla, Colombia | mateo.alvarezr@unisimon.edu.co

Resumen

El objetivo de este estudio fue implementar un modelo de inteligencia artificial para la caracterización de los turistas en La Guajira, Colombia, a fin de mejorar la personalización de las ofertas turísticas. Se generaron datos sintéticos basados en estudios previos y encuestas del sector, que luego fueron procesados mediante el algoritmo de segmentación K-medias. Este algoritmo permitió clasificar a los turistas en tres categorías: cultural, de naturaleza y familiar, a partir de variables como nivel educativo, ocupación y motivaciones de viaje. Los resultados demostraron una segmentación efectiva de los perfiles, con una alta coherencia interna entre los clústeres en el entorno de simulación controlada, resultados que son hipotéticos y pendientes de validación empírica. Bajo estas condiciones, el modelo podría facilitar la adaptación de las estrategias de marketing y servicios turísticos a las características específicas de cada grupo. Aunque los datos sintéticos fueron clave



en esta fase, se concluye que la validación con datos reales es esencial para garantizar la aplicabilidad del modelo en entornos reales.

Palabras clave: K-medias; datos sintéticos; aprendizaje no supervisado; marketing de destinos; La Guajira (Colombia)

Abstract

The objective of this study was to implement an artificial intelligence model for the characterization of tourists in La Guajira, Colombia, in order to improve the personalization of tourism offers. Synthetic data were generated based on previous studies and surveys of the sector, which were then processed using the K-means segmentation algorithm. This algorithm made it possible to classify tourists into three categories: cultural, nature and family, based on variables such as educational level, occupation and travel motivations. The results demonstrated an effective segmentation of the profiles, with high internal cohesion among the clusters under controlled simulation conditions; results that are hypothetical and pending empirical validation. Under these conditions, the model could potentially facilitate the adaptation of marketing strategies and tourism services to the specific characteristics of each group. Although synthetic data was key in this phase, it is concluded that validation with real data is essential to ensure the applicability of the model in real environments.

Key words: K-means; synthetic data; unsupervised learning; destination marketing; La Guajira (Colombia)

1 Introducción

El turismo se ha consolidado como una de las industrias más relevantes a nivel global, desempeñando un papel fundamental en el desarrollo económico y social de diversas regiones. En los últimos años, la incorporación de tecnologías emergentes ha transformado profundamente la dinámica del sector, redefiniendo la competitividad y las estrategias empresariales. Por ello, la inteligencia artificial (IA) y la robótica han emergido como herramientas clave para la optimización de procesos, la personalización de experiencias y la mejora en la gestión de los servicios turísticos (Alonso Almeida, 2019; Samala et al., 2022).

Desde esta perspectiva, el acelerado desarrollo tecnológico vinculado con la automatización, la IA y la robótica ha impulsado a las organizaciones a transformar y adaptar sus procesos operativos a estas innovaciones (Syam y Sharma, 2018). En este sentido, el concepto de Industria 4.0, también denominada Cuarta Revolución Industrial, ha introducido aplicaciones tecnológicas que han generado cambios significativos en diversos sectores, entre estos el turismo (Almada-Lobo, 2016). El auge de la Cuarta Revolución Industrial ha propiciado un incremento exponencial en la generación de datos y el desarrollo de algoritmos avanzados, incluyendo la IA y la automatización, cuya

implementación en la industria hotelera y turística se ha convertido en una estrategia clave para optimizar la gestión organizacional y responder de manera eficiente a los desafíos de este sector (Makridakis, 2017). De esta manera, la inteligencia artificial está adquiriendo un papel cada vez más relevante en la optimización de procesos y servicios turísticos, lo que ha generado mejoras sustanciales en la gestión de la información y la experiencia del cliente (Gómez Muñoz et al., 2023).

La aplicación de la IA en el turismo ha demostrado su potencial en la automatización de procesos, la segmentación de mercados y la mejora en la toma de decisiones basada en datos. Estudios recientes han destacado que la incorporación de estas tecnologías facilita la adaptabilidad del sector a las preferencias de los consumidores, permitiendo así una oferta de servicios más eficientes y personalizados (Blöcher y Alt, 2021; Gallego Gómez y Vaquero Frías, 2022). Sin embargo, la digitalización del turismo también plantea desafíos, como la integración efectiva del factor humano en la atención al cliente, la seguridad de los datos y la inclusión de pequeñas empresas en este ecosistema digital.

En este orden de ideas, diversos estudios han evidenciado la necesidad de innovación en subsectores clave del turismo, como la hotelería, la restauración y la gestión de destinos, mediante el uso de técnicas avanzadas como el Big Data y el aprendizaje automático (Bowen y Whalen, 2017; Ketui et al., 2024). A su vez, Li et al. (2018) identificaron tres categorías principales de macrodatos en la investigación turística: datos generados por los usuarios (UGC), que incluyen información textual y fotográfica en línea; datos obtenidos de dispositivos, como información de GPS, roaming móvil y Bluetooth; y datos operacionales que abarcan búsquedas web, reservas en línea y otros registros digitales. La adecuada caracterización y análisis de estas tipologías de datos permite responder a los retos actuales del turismo y mejorar la toma de decisiones en el sector.

En términos particulares, la implementación de aplicaciones inteligentes en el turismo ha generado impactos positivos en la competitividad de los destinos, como lo demuestra la aplicación MAHASAMUT en Tailandia, la cual optimiza la experiencia del viajero mediante asistencia multimodal (Ketui et al., 2024). Por su parte, Filieri et al. (2021) destacan que la IA puede generar importantes beneficios en la industria del turismo, particularmente en la optimización de procesos y la personalización de servicios. De esta manera, la adopción de robots en el sector de la hospitalidad y el turismo está en constante crecimiento, con aplicaciones que van desde camareros automatizados hasta conserjes robóticos en hoteles (Belanche et al., 2021; Hayat, 2023). Esta innovación busca optimizar los costos operativos mediante la automatización y mejorar el servicio a través de la experiencia del cliente (Ivanov y Webster, 2019).

2 Revisión de la literatura

Este apartado presenta los fundamentos teóricos y antecedentes empíricos que sustentan este estudio. Se abordan tres ejes temáticos: en primer lugar, la transformación digital del turismo en el

marco de la Industria 4.0; en segundo lugar, el uso de la inteligencia artificial y el aprendizaje automático en la segmentación de mercados turísticos; y, por último, el estado del arte en América Latina y Colombia. La revisión permitió identificar el vacío de conocimiento que justifica el enfoque metodológico adoptado.

En el campo de la segmentación turística, García-Reinoso et al. (2020) analizaron la demanda colombiana en Ecuador e identificaron tres segmentos principales de turistas: cultural, de naturaleza y familiar. Utilizando análisis factorial exploratorio y algoritmos de clasificación como K-medias, determinaron que la segmentación efectiva permite desarrollar estrategias de marketing más precisas, mejorando la calidad del servicio y el posicionamiento turístico.

A nivel internacional, la digitalización ha redefinido la oferta y la demanda en países como España, donde el turismo representa un pilar económico clave. No obstante, la limitada generación de innovación propia lo hace vulnerable ante la disrupción tecnológica. La evaluación del impacto de estas tecnologías es esencial para fortalecer su competitividad en un entorno digitalizado (Más-Ferrando et al., 2020).

En este sentido, la minería de datos y el aprendizaje automático permiten analizar la información generada por turistas en plataformas digitales y redes sociales (Alonso Almeida, 2019). Además, la implementación de las tecnologías de IA con mayor financiamiento en este sector está relacionada con el big data, el aprendizaje automático y el procesamiento del lenguaje natural, lo que refuerza la importancia de estas herramientas en la segmentación de mercados y en la mejora de la experiencia del usuario (Bulchand-Gidumal, 2020; Penagos-Londoño et al., 2021). Asimismo, el uso de IA en este sector despierta el interés del viajero por adquirir el servicio, al tiempo que otorga un valor agregado a las organizaciones y aumenta su competitividad en el mercado (Chi et al., 2022).

Por su parte, Zambrano et al. (2024) investigaron el uso de la IA en la promoción turística de la Amazonía ecuatoriana, considerando desafíos como la conservación ambiental y la competencia global. A través de un cuestionario en línea, se identificó que la mayoría de los usuarios no está familiarizada con la IA en este campo. De esta manera, concluyen que su implementación es prometedora, pero que requiere mayor difusión y educación para su adopción efectiva.

En Colombia, la aplicación de IA en el turismo ha cobrado relevancia en los últimos años. En este contexto, Rodríguez Correa et al. (2023) realizaron un análisis bibliométrico en el que identificaron un creciente interés en la aplicación de robots y aprendizaje automático en la oferta de servicios turísticos, destacando la importancia de estas tecnologías en la modernización del sector y en el desarrollo del turismo sostenible (Hermosa del Vasto y Arco Castro, 2024).

En el caso puntual del estudio de la demanda turística en el Departamento de La Guajira, Colombia, el turismo ha experimentado una reducción en la demanda, con una caída del 32% en la ocupación hotelera entre 2022 y 2023, atribuida a bloqueos viales, crisis de aerolíneas e inseguridad (Cámara

de Comercio de La Guajira, 2023). Este escenario resalta la necesidad de estrategias innovadoras para mejorar la competitividad del destino (Balance del turismo en Semana Santa 2021, 2021).

A nivel nacional, la ocupación hotelera también disminuyó entre 7,4 y 15,5 puntos porcentuales, destacando así la necesidad de estrategias innovadoras basadas en IA para fortalecer la competitividad del sector (Cámara de Comercio de La Guajira, 2023).

Por esta razón, el objetivo principal del presente estudio se centró en desarrollar una inteligencia artificial que por medio de la segmentación permita la identificación y caracterización de clientes, para la personalización de las preferencias en pro de ofrecerles de manera ágil los productos y servicios del sector turismo, buscando satisfacer las necesidades de la demanda turística de forma novedosa, rápida y segura, con base en la IA, para favorecer la optimización de tiempo y costo con el fin de transformar la cartera de servicios del sector turismo mediante el desarrollo de la tecnología.

Este estudio aborda la caracterización del turismo en el departamento de La Guajira, Colombia, mediante la aplicación de un modelo de inteligencia artificial basado en la segmentación de datos. A través del uso del algoritmo K-medias y la generación de datos sintéticos, con el fin de identificar patrones en las preferencias y características de los turistas, que permitieran la clasificación en tres segmentos principales: turismo cultural, de naturaleza y familiar. Esta segmentación no solo contribuye a una mayor comprensión del comportamiento turístico en la región, sino que también representa una herramienta estratégica para la optimización de la oferta turística, mejorando la personalización de los servicios y fortaleciendo la competitividad del destino. Los hallazgos obtenidos en este estudio evidencian la importancia de la inteligencia artificial como un recurso clave en la transformación digital del sector turístico, alineándose con las tendencias globales del sector.

3 Metodología

El presente estudio se basó en la implementación de un modelo de inteligencia artificial para la caracterización de los turistas en La Guajira, Colombia. Al no tener suficientes datos reales, se generó un conjunto de datos sintéticos para entrenar el modelo de segmentación. El proceso metodológico incluyó las siguientes etapas:

3.1 Generación de Datos sintéticos

Dado que no se dispone de bases de datos oficiales completas sobre el perfil de los turistas en el Departamento de La Guajira, se generó un conjunto de datos sintéticos a partir de los perfiles identificados en estudios previos (García-Reinoso et al., 2020) y en encuestas del sector turístico local. Este enfoque, ampliamente utilizado en investigación cuando los datos reales son escasos o inaccesibles, permite entrenar y validar modelos de segmentación bajo condiciones controladas, con la limitación explícita de que los resultados no pueden generalizarse directamente a la población turística real sin una validación empírica posterior.

El proceso de generación se basó en la codificación ordinal discreta de siete variables clave, cuyas categorías y valores numéricos asignados se detallan a continuación: Nivel de estudio (0 = Medios o técnicos, 1 = Primaria y secundaria, 2 = Universitarios); Ocupación (0 = Cuenta propia, 1 = Empresa privada, 2 = Sector público); Medio de información (0 = Por familiares y amigos, 1 = Por ferias, 2 = Por agencia de viaje, 3 = Publicidad, 4 = Redes sociales); Grupo de viaje (0 = Amigos, 1 = Familia, 2 = Grupo organizado, 3 = Matrimonio o pareja, 4 = Solo); Duración (0 = 1 a 2 noches, 1 = 3 a 6 noches); Medio de transporte (0 = Avión, 1 = Bus, 2 = Taxi, 3 = Vehículo alquilado, 4 = Vehículo propio); y Método de reserva (0 = Buscadores online, 1 = Directamente, 2 = Por agencia de viajes, 3 = Sin reserva).

Para cada perfil de turista se seleccionó aleatoriamente un valor de cada variable conforme a las distribuciones características del segmento correspondiente, según la literatura de referencia. En total se creó un archivo CSV con 1.608 perfiles simulados, distribuidos de la siguiente manera: Turismo Cultural (categoría 0): 360 perfiles base más 22 perfiles con ruido, para un total de 382 registros; Turismo de Naturaleza (categoría 1): 603 perfiles base más 299 perfiles con ruido, para un total de 902 registros; Turismo Familiar (categoría 2): 102 perfiles base más 222 perfiles con ruido, para un total de 324 registros. El ruido fue introducido deliberadamente en los datos sintéticos, generando perfiles con características similares a las de otras agrupaciones, con el fin de aumentar la dificultad de clasificación y evaluar la robustez del modelo ante la ambigüedad de perfiles limítrofes. Es importante señalar que la totalidad de los resultados presentados en este estudio corresponde a un ejercicio de simulación controlada y no a evidencia empírica del comportamiento real de los turistas en La Guajira.

3.2 Preprocesamiento de Datos

En segundo lugar, se realizó el preprocesamiento de los datos, teniendo en cuenta que todas las variables son de naturaleza cualitativa ordinal y nominal, se aplicó una codificación ordinal discreta que asignó un valor entero a cada categoría de respuesta, conforme a la escala detallada en la sección 3.1. Esta estrategia de codificación es apropiada para algoritmos de agrupamiento como K-medias cuando las distancias entre categorías tienen una interpretación relativa coherente, siendo preferible a la codificación one-hot en este contexto porque preserva la estructura ordinal de variables como nivel de estudio y duración del viaje. Las variables resultantes fueron utilizadas directamente como entradas al algoritmo, sin normalización adicional, dado que todas operan en escalas discretas comparables (rango 0-4). La introducción de perfiles con ruido, detallada en la sección 3.1, cumplió la función de desafiar la capacidad del modelo para diferenciar patrones representativos de variaciones aleatorias en los datos de entrenamiento, evaluando así su robustez ante perfiles ambiguos.

3.3 Algoritmo de Segmentación

Posteriormente, se seleccionó el algoritmo de agrupamiento K-medias para segmentar los perfiles de turistas en tres grupos: turismo cultural, turismo de naturaleza y turismo familiar. El número óptimo de grupos fue determinado utilizando el método del codo, que sugirió que tres clústeres eran suficientes para una segmentación efectiva. Como validación complementaria del número de clústeres, se calculó el coeficiente de silueta (silhouette score) para $k=2$, 3 y 4, confirmando que $k=3$ maximiza la cohesión interna y la separación entre grupos. Este método se aplicó cruzando dimensiones como nivel educativo, ocupación y duración del viaje para identificar patrones en los perfiles de los turistas.

3.4 Evaluación del Modelo

Por último, el modelo fue entrenado con los datos generados sintéticamente, y su rendimiento fue evaluado utilizando métricas propias de clustering no supervisado. La calidad de los agrupamientos se evaluó mediante el Índice de Davies-Bouldin (IDB) y el coeficiente de silueta (silhouette score), indicadores que miden respectivamente la compacidad de cada clúster respecto a su separación de los demás, y el grado en que cada punto se encuentra más próximo a su propio clúster que a cualquier otro. La ejecución del algoritmo para $k=3$ produjo los siguientes centroides: Clúster 0 – (0.617, 0.930, 2.753, 1.756, 0.448, 1.697); Clúster 1 – (1.962, 0.041, 0.019, 1.058, 1.000, 1.014); Clúster 2 – (0.000, 1.954, 1.976, 0.021, 1.000, 2.921), donde cada valor corresponde a las seis dimensiones codificadas utilizadas en el proceso de agrupamiento. Estos centroides permiten identificar diferencias entre los agrupamientos obtenidos. En los perfiles teóricos utilizados para generar los datos sintéticos, el turismo cultural presenta mayor nivel educativo y variedad en fuentes de información; el turismo familiar muestra ocupación principalmente en empresa privada y estadías más cortas; y el turismo de naturaleza se asocia al sector público y mayor uso de agencias de viaje.

Es importante señalar que, debido a que los datos fueron generados con fronteras de decisión predefinidas entre categorías, los clústeres resultantes reproducen en gran medida la estructura implícita del proceso de generación, lo cual constituye una limitación de validez externa del modelo que deberá superarse con datos empíricos en investigaciones futuras. Se seleccionaron 322 registros para la validación donde el modelo demostró coherencia interna en la asignación de perfiles a los segmentos preestablecidos. La alta correspondencia observada entre las etiquetas asignadas por el algoritmo y las categorías de origen de los perfiles sintéticos es consistente con la naturaleza determinística del proceso de generación de datos, y no debe interpretarse como evidencia de capacidad predictiva sobre turistas reales.

4 Resultados

En esta investigación, se identificaron perfiles de clientes para cada segmento del turismo cultural, de naturaleza y turismo familiar. Para determinar los perfiles, se utilizaron tablas de contingencia que

relacionaban variables de interés. Los resultados estadísticos revelaron diferencias significativas entre los tres grupos o clústeres en términos demográficos. Así, en el segmento de turismo cultural, el 57,1% de los clientes completaron estudios universitarios y trabajaban de forma independiente. En el caso del turismo de naturaleza, el 22,9% representado estaba compuesto por profesionales técnicos vinculados al sector público y en cuanto al turismo familiar, 19,8% presentaban niveles educativos primarios y secundarios con desempeño en empresas privadas.

Con relación a las características del viaje, se observaron diferencias significativas entre los segmentos, para los turistas culturales se informaban principalmente a través de familiares y amigos, mientras que los turistas de naturaleza utilizaban agencias de viaje y los turistas familiares se basaban en publicidad de la televisión, la radio y la prensa. Adicionalmente, los hallazgos revelaron que las motivaciones de viaje presentaron diferencias significativas entre los segmentos analizados. Los turistas culturales manifestaron una marcada preferencia por experiencias inmersivas, destacando el patrimonio cultural, las actividades y el entretenimiento de carácter local, así como la amabilidad de los habitantes y la seguridad del entorno. Los turistas de naturaleza, por su parte, priorizaron el contacto directo con el medio ambiente, buscando disfrutar de paisajes naturales, realizar actividades al aire libre y explorar nuevos destinos. Finalmente, los turistas familiares se inclinaron por motivos más personales y sociales, como visitar a seres queridos, cuidar de su salud y bienestar, y disfrutar de actividades recreativas en familia.

Tabla 1. Ejemplo de perfiles sociodemográficos y de viaje generados sintéticamente.

#	Sexo	Ocupación	Medio de información	Grupo de viaje	Medio transporte	Duración viaje	Reserva	Edad	Categoría
0	Otra	Cuenta Propia	Publicidad	Matrimonio o pareja	Bus	3 a 6 noches	Buscadores online	20	4
1	Otra	Cuenta Propia	Por una agencia de viaje	Amigos	Taxi	7 o más noches	Por agencia de viajes	29	4
2	Masculino	Empresa Privada	Publicidad	Familia	Avión	3 a 6 noches	Buscadores online	26	4
3	No dice	Empresa Privada	Redes sociales	Matrimonio o pareja	Avión	1 a 2 noches	Buscadores online	28	4
4	Masculino	Sector Público	Por una agencia de viaje	Solo	Vehículo alquilado	1 a 2 noches	Sin reserva	56	4

Nota: La tabla presenta cinco registros de ejemplo generados sintéticamente, con características sociodemográficas y de viaje.

Desde esta perspectiva y con el fin de generar datos de perfiles de turistas para el entrenamiento de la inteligencia artificial, se requiere crear una combinación aleatoria de estas características en un archivo CSV llamado "turistas.csv". Este archivo contendrá los perfiles de turistas necesarios para que la inteligencia artificial pueda segmentar a los clientes en los grupos establecidos: Cultural,

Naturaleza y Familiar. Se generó un conjunto de datos de prueba basado en números aleatorios y listas estándar para contrarrestar el simulador de datos ordinales.

El resultado de este ejercicio es un conjunto de datos de 1.608 turistas simulados, cuyos principales parámetros de tendencia central y dispersión se presentan en la siguiente tabla.

Tabla 2. Estadísticos descriptivos del conjunto de datos sintéticos (n = 1.608)

Estadístico	Estudio	Ocupación	Medio información	Grupo de viaje	Medio transporte	Duración viaje	Reserva	Categoría
Cantidad total de los datos	1608.00000	1608.00000	1608.00000	1608.00000	1608.00000	1608.00000	1608.00000	1608.00000
media	0.669154	1.147388	1.816542	0.889303	2.036692	0.798507	2.042289	0.963930
desviación estándar	0.834832	0.871333	1.148273	1.099385	1.398914	0.401240	0.945669	0.661835
mínimo	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.000000	1.000000	0.000000	1.000000	1.000000	1.000000	1.000000
50%	0.000000	1.000000	2.000000	1.000000	2.000000	1.000000	2.000000	1.000000
75%	1.000000	2.000000	3.000000	1.000000	3.000000	1.000000	3.000000	1.000000
máximo	2.000000	2.000000	4.000000	4.000000	4.000000	1.000000	3.000000	2.000000

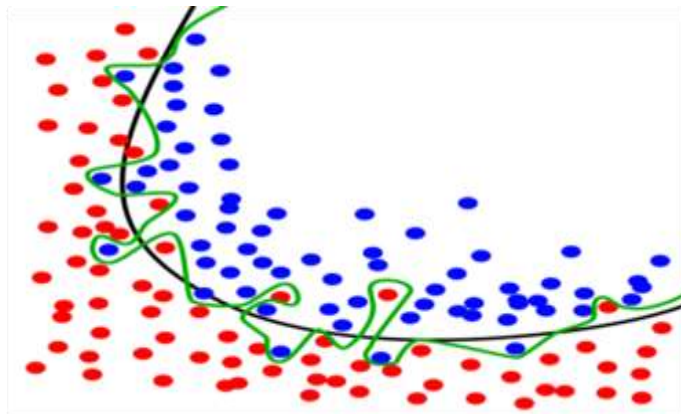
Nota: La tabla presenta un análisis exploratorio de los datos de los viajeros, mostrando las medidas de tendencia central y dispersión para cada variable.

En este contexto, se propone generar datos que introduzcan ruido en la categoría de entretenimiento, para mejorar la precisión de la clasificación por parte de la inteligencia artificial. Estos datos se generaron de manera aleatoria y estuvieron diseñados para simular características similares a las de las otras agrupaciones. En este orden de ideas, el ruido se refiere al componente de aleatoriedad o variabilidad que está presente en un conjunto de datos y que puede dificultar la detección de los patrones subyacentes. De esta manera, al introducir ruido en los datos sintéticos, se busca desafiar al modelo de aprendizaje automático a aprender y distinguir entre la señal relevante y el ruido. Ya que un modelo de aprendizaje automático bien ajustado es capaz de captar la señal relevante y minimizar el impacto del ruido en la clasificación de los datos. Esto implica que el modelo tenga la capacidad de descifrar las características y patrones más relevantes de cada grupo, incluso cuando existan similitudes o superposiciones en ciertas características.

Por último, al introducir ruido en la categoría de entretenimiento, se crea un escenario más realista y desafiante para el modelo, lo que puede mejorar su capacidad de generalización y precisión en la

clasificación de turistas en los grupos establecidos (Cultural, Naturaleza y Familiar). Al aprender a diferenciar la señal relevante del ruido, el modelo puede proporcionar resultados más confiables y precisos en la segmentación de los perfiles de los turistas.

Figura 1. Clasificación no lineal



Nota: La figura muestra un conjunto de datos bidimensionales con dos clases, representadas por puntos rojos y azules. La línea negra y la línea verde ilustran dos posibles fronteras de decisión que separan las dos clases.
Ejemplo de sobreajuste- línea verde.

Cabe resaltar que la gráfica anterior representó un problema de clasificación en el cual se buscó predecir la clase de nuevos casos, por tanto, en la gráfica se muestra una línea negra que indica ser un buen predictor, ya que claramente presenta las clases azul y rojo. Los puntos a la derecha de la línea se clasifican como azul, mientras que los puntos a la izquierda se clasifican como rojo. Sin embargo, también se muestra una línea verde que representa un modelo sobre ajustado, esto significa que el modelo ha aprendido demasiado del ruido o de las variaciones aleatorias en los datos de entrenamiento. Se obtuvo como resultado que el modelo se ajusta demasiado a estos puntos ruidosos.

Posteriormente, se menciona la aplicación de una transformación de las listas de selección de características a números, para que la máquina comprenda mejor los datos y facilite su procesamiento. Al convertir las características en números, esto permitió que los algoritmos de aprendizaje automático trabajaran con los datos de manera más eficiente, ya que muchos algoritmos están diseñados para trabajar con datos numéricos. Por tanto, la transformación de las características a números puede incluir técnicas como codificación one-hot para variables categóricas, normalización de variables numéricas o cualquier otro método adecuado para representar las características en forma numérica. Al realizar esta transformación, se proporciona una representación más adecuada para que la máquina pueda aprender y realizar predicciones basadas en los datos de entrada. A continuación, se presentan gráficamente los datos para tener una idea de la dispersión de estos por cada característica relevante:

A continuación, se presentan datos sintéticos generados a partir de la transformación de las características seleccionadas en variables numéricas, facilitando su entendimiento, procesamiento

y análisis mediante el algoritmo K-medias. La tabla incluye información detallada sobre los niveles de estudio, ocupación, medio de información, grupo de viaje, duración de la estadía, medio de transporte y método de reserva. Estos datos son fundamentales para entender cómo se agrupan los turistas en función de sus características sociodemográficas y preferencias de viaje.

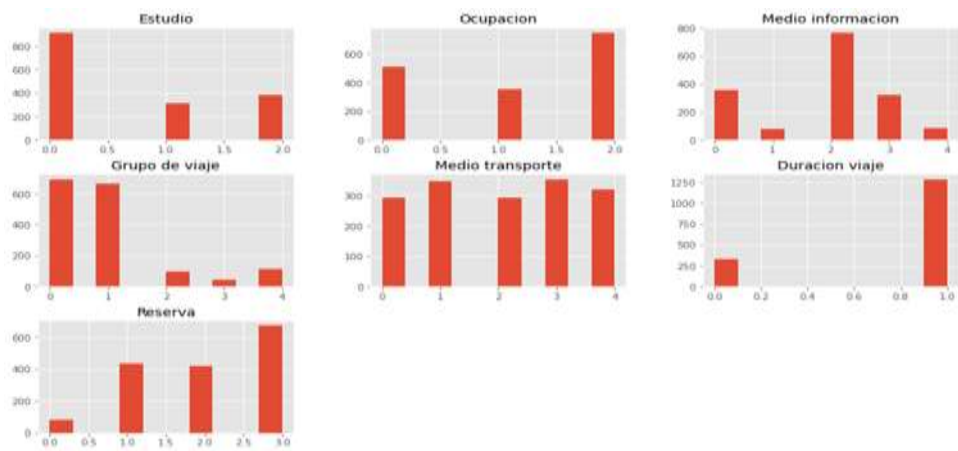
Posteriormente, se muestra un gráfico de dispersión que ilustra la relación entre estas variables clave, permitiendo visualizar cómo se distribuyen los perfiles de los turistas en la Guajira en diferentes clústeres o grupos. Este gráfico proporciona una representación visual de los patrones de comportamiento identificados en los datos, lo que permite analizar más a fondo las similitudes y diferencias entre los turistas en La Guajira.

Tabla 3. Clasificación numérica de las variables del perfil turístico

Categoría	Nivel de estudio	Ocupación	Medio de información	Grupo de viaje	Duración	Medio de transporte	Reserva
0	Medios o técnicos	Cuenta Propia	Por familiares y amigos	Amigos	1 a 2 Noches	Avión	Buscadores Online
1	Primaria y secundaria	Empresa Privada	Por ferias	Familia	3 a 6 Noches	Bus	Directamente
2	Universitarios	Sector Público	Por una agencia de viaje	Grupo Organizado		Taxi	Por agencia de viajes
3			Publicidad	Matrimonio o pareja		Vehículo Alquilado	Sin reserva
4			Redes sociales	Solo		Vehículo Propio	

Nota: Los valores numéricos (0–4) corresponden a códigos asignados a las categorías de las variables para su procesamiento computacional. Datos generados bajo condiciones de simulación controlada.

Figura 2. Distribución de las variables sociodemográficas y de viaje



Nota: La figura presenta histogramas que muestran la frecuencia de cada categoría para las diferentes variables..

En la gráfica anterior se observa que al usar el algoritmo de agrupamiento k-medias para analizar las dimensiones del nivel de estudio, ocupación, duración y dispersión en perfiles. El método de k-medias tuvo como objetivo dividir un conjunto de observaciones en grupos basados en la proximidad de sus valores medios; en este caso, se logró seleccionar una dimensión, como la columna categoría, y cruzarla con las otras dimensiones mencionadas (nivel de estudio, ocupación, duración y dispersión) para observar si hay alguna pista sobre su agrupación y su relación con las categorías. Al aplicar el algoritmo de k-medias, la finalidad de esto fue encontrar grupos o clústeres de perfiles que compartieran características similares en las dimensiones seleccionadas. Por ejemplo, se pudo analizar si los perfiles con un nivel de estudio universitario y una ocupación propia tienden a agruparse en una categoría específica, mientras que los con un nivel de estudio primario, secundario y una ocupación en empresas privadas se agrupan en otra.

De igual forma, al estudiar las relaciones entre las diferentes dimensiones y las categorías, se obtuvieron pistas sobre cómo estas variables se relacionaban entre sí y cómo influyen en la formación de grupos de perfiles distintos. Esto proporcionó información valiosa sobre los diferentes segmentos de clientes al igual que su perfil característico. Es importante tener en cuenta que el algoritmo de k-medias es sensible a la elección del número de grupos (k) y a la inicialización de los centroides iniciales. Por lo tanto, es recomendable realizar varias ejecuciones del algoritmo con diferentes valores de k y evaluar la calidad de los resultados obtenidos utilizando medidas de evaluación como la suma de cuadrados dentro de los grupos (SSW) y la suma de cuadrados entre grupos (SSB). Además, se recomienda considerar otras técnicas de agrupamiento o algoritmos, como el agrupamiento jerárquico o el DBSCAN, dependiendo de la naturaleza de los datos y los objetivos específicos del análisis.

La figura presentada permite observar la distribución de las características clave de los perfiles turísticos, permitiendo visualizar cómo se agrupan los datos de acuerdo con dimensiones relevantes como el nivel de estudio, ocupación, medio de información, grupo de viaje, medio de transporte, duración del viaje y método de reserva. Estas variables, previamente transformadas a valores numéricos para facilitar el procesamiento y entendimiento de los datos mediante el algoritmo K-medias, ofrecen un panorama sobre los hábitos y características de los turistas en La Guajira.

Una de las primeras observaciones destacadas en los gráficos es la prevalencia de turistas con estudios medios o técnicos. Sin embargo, una proporción de turistas con estudios universitarios podría buscar experiencias turísticas más especializadas o culturales, en contraste con los de menor nivel educativo. Esta diferenciación puede ser esencial para adaptar las ofertas turísticas a los distintos perfiles educativos.

En cuanto a la ocupación, es notable que la mayoría de los turistas pertenecen al sector público, seguido por aquellos que trabajan de forma autónoma y empresas privadas. Esta distribución podría

tener implicaciones en los patrones de viaje, ya que los turistas del sector público tienden a tener una mayor estabilidad económica y flexibilidad en sus tiempos de descanso, lo que les permite planificar viajes más largos y organizados. Esta tendencia también podría estar relacionada con el medio de información que utilizan, ya que un número significativo de estos turistas prefieren recurrir a agencias de viaje, lo que refleja una mayor confianza en intermediarios profesionales y métodos tradicionales de planificación.

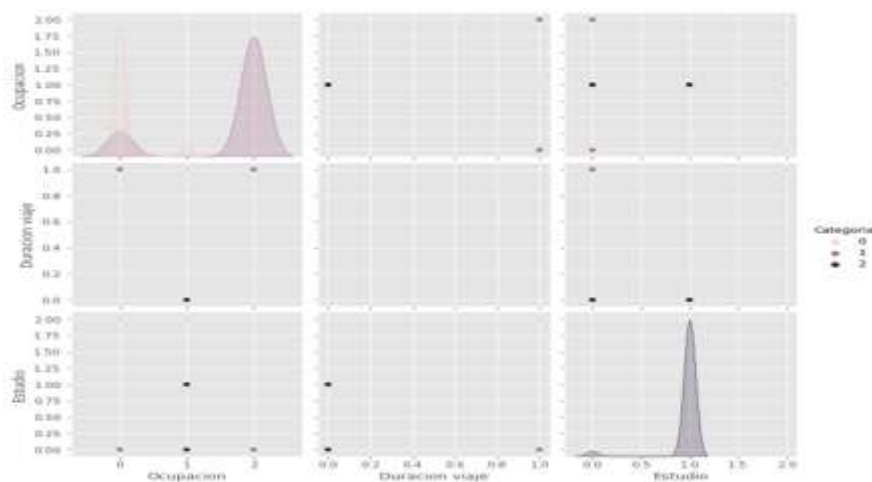
El análisis del grupo de viaje también es revelador, ya que la mayoría de los turistas viajan acompañados por amigos o familiares. Esta preferencia por los viajes grupales sugiere que el destino ofrece actividades que se disfrutan mejor en compañía, lo cual puede ser aprovechado por los operadores turísticos al diseñar ofertas que fomenten experiencias compartidas. Por otro lado, un grupo más pequeño de turistas viaja solo o en grupos organizados, lo que puede indicar un segmento de viajeros más independiente o que busca mayor estructura en su itinerario.

En cuanto al medio de transporte, se observa una distribución relativamente homogénea entre las cinco opciones consideradas, sin que ninguna presente un predominio marcado. Esta distribución evidencia diversidad en los medios de transporte representados en los perfiles turísticos simulados.

En términos de la duración del viaje, la mayoría de los turistas optan por estancias prolongadas, de entre 3 y 6 noches. Este comportamiento sugiere una preferencia por viajes más largos, lo que permite a los turistas disfrutar de una experiencia más inmersiva en el destino. Finalmente, en lo que respecta al método de reserva, la opción «sin reserva» presenta la mayor frecuencia, mientras que la reserva directa y la realizada mediante agencias de viaje muestran una representación relativamente similar. Los buscadores online presentan una frecuencia menor dentro de los perfiles turísticos simulados.

Estos hallazgos proporcionan una visión completa sobre las características de los turistas que visitan La Guajira, permitiendo a los operadores turísticos y tomadores de decisiones adaptar mejor sus servicios y estrategias de marketing a las necesidades específicas de cada segmento. La segmentación obtenida mediante el uso de K-medias permite identificar patrones de comportamiento y preferencias en función de estas variables clave, facilitando una oferta más personalizada y efectiva.

Figura 3. Relación entre variables sociodemográficas y de viaje



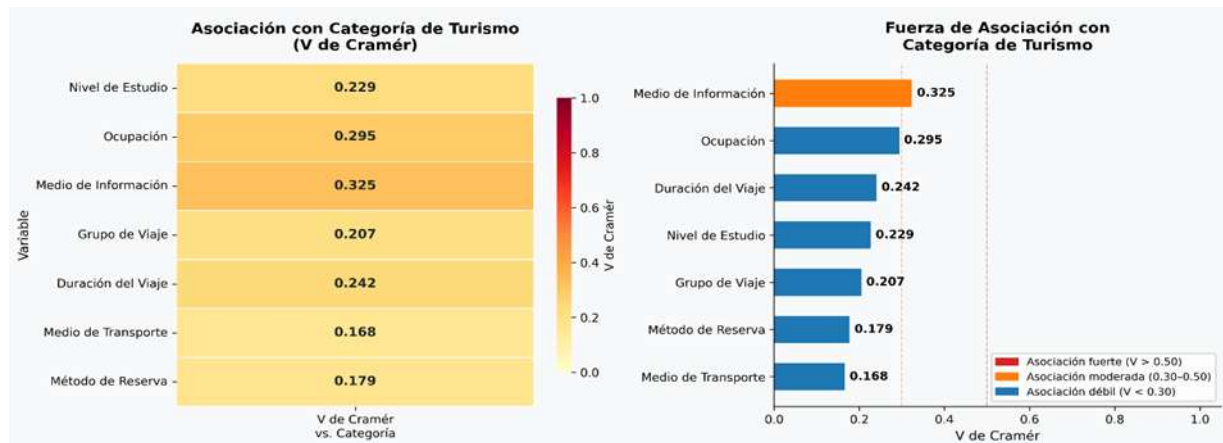
Nota: La figura presenta una matriz de dispersión que permite visualizar las relaciones entre las variables de ocupación, duración del viaje y nivel de estudios.

El diagrama de dispersión se emplea comúnmente para representar visualmente la relación entre dos variables, permitiendo evaluar tanto el grado de asociación como la posible independencia entre ellas. En este tipo de gráfico, una variable se posiciona en el eje horizontal y la otra en el eje vertical, y cada punto representa un caso particular, indicando la interrelación entre los valores correspondientes. El análisis de dicho diagrama puede revelar patrones específicos que describen la naturaleza de la relación entre las variables. Cuando los puntos exhiben una disposición ascendente hacia la derecha, se observa una correlación positiva, lo que implica que ambas variables tienden a aumentar o disminuir de forma conjunta. Por otro lado, una disposición descendente de los puntos indica una correlación negativa, es decir, mientras una variable incrementa, la otra disminuye. Si no se detecta un patrón claro en la distribución de los puntos, se concluye la existencia de una correlación nula, lo que sugiere la ausencia de una relación entre las variables analizadas.

Este gráfico de dispersión muestra la relación entre las variables ocupación, duración del viaje y estudio para tres categorías de turistas: turismo cultural (categoría 0), turismo de naturaleza (categoría 1) y turismo familiar (categoría 2). Se observa que en la categoría 1, los turistas con ocupaciones vinculadas al sector público o técnico presentan mayor dispersión en la duración del viaje, sugiriendo una relación débil entre estas variables. Asimismo, los turistas con un nivel educativo superior (categoría 0) muestran mayor consistencia en la duración de su estadía, mientras que los turistas con estudios primarios o secundarios (categoría 2) presentan menor variabilidad en ambas dimensiones. Sin embargo, en general, no se detecta una correlación significativa entre las variables analizadas, lo que indica que la ocupación, el nivel educativo y la duración del viaje son independientes entre sí en los grupos evaluados.

La dispersión observada refleja la heterogeneidad en los perfiles de los turistas simulados, alineándose con los objetivos del modelo de segmentación, que busca identificar patrones diversos en los clientes del sector turístico.

Figura 4. Análisis de asociación entre variables del perfil turístico y categoría de segmento (Chi-cuadrado y V de Cramér)



Nota: La figura presenta el mapa de calor de V de Cramér (panel izquierdo) y el diagrama de barras de fuerza de asociación (panel derecho) entre cada variable del perfil turístico y la variable "Categoría". Todas las asociaciones son estadísticamente significativas ($p < 0.001$). El análisis se realizó con perfiles turísticos sintéticos generados bajo condiciones de simulación controlada; por tanto, las asociaciones identificadas describen relaciones dentro del entorno simulado y no constituyen evidencia empírica sobre el comportamiento de turistas reales. Elaboración propia.

Teniendo en cuenta que la variable "Categoría" es de naturaleza nominal, el análisis de asociación se realizó mediante la prueba Chi-cuadrado (χ^2) y el estadístico V de Cramér, medidas adecuadas para evaluar la dependencia entre variables categóricas en lugar del coeficiente de correlación de Pearson, cuyo uso se restringe a variables de intervalo o razón. Los resultados indican que todas las variables del perfil turístico presentan asociaciones estadísticamente significativas con la categoría de turismo ($p < 0.001$ en todos los casos), lo que confirma que los datos sintéticos generados reproducen diferencias sistemáticas entre segmentos coherentes con los perfiles de referencia. En términos de fuerza de asociación, el Medio de Información presenta la asociación más fuerte (V de Cramér = 0.325), seguido por la Ocupación (V = 0.295), la Duración del Viaje (V = 0.242) y el Nivel de Estudio (V = 0.229). Las variables Grupo de Viaje (V = 0.207), Método de Reserva (V = 0.179) y Medio de Transporte (V = 0.168) muestran asociaciones débiles, aunque estadísticamente significativas. Es importante señalar que estas asociaciones se derivan de datos sintéticos y reflejan los patrones establecidos en el proceso de generación bajo condiciones de simulación controlada, por lo que no deben interpretarse como evidencia empírica del comportamiento real de los turistas en La Guajira.

El análisis de asociación mediante Chi-cuadrado y V de Cramér, realizado sobre los datos sintéticos generados, revela patrones diferenciadores entre los tres segmentos turísticos bajo condiciones de

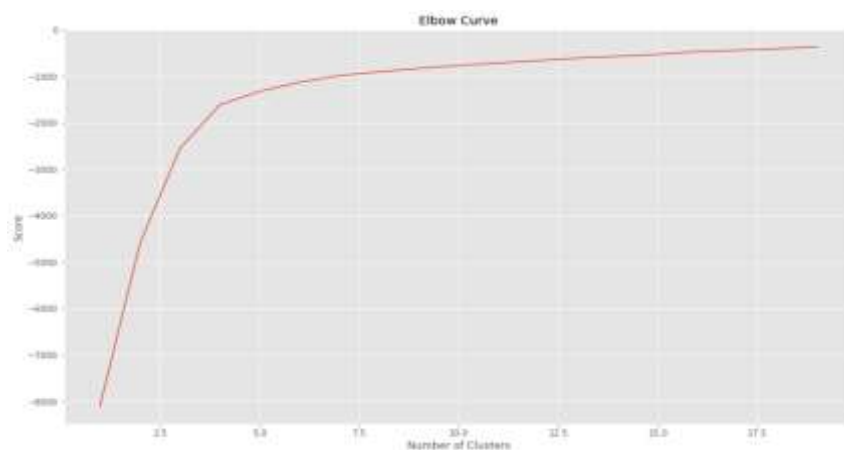
simulación controlada. La V de Cramér, cuyos valores oscilan entre 0 y 1 (donde valores superiores a 0.30 indican asociación moderada y superiores a 0.50 indican asociación fuerte), permite cuantificar la fuerza de la dependencia entre cada variable del perfil y la categoría de turismo. A continuación se detallan los hallazgos principales.

El Medio de Información es la variable con mayor asociación con la categoría de turismo ($\chi^2=225.49$, $gl=8$, $p<0.001$; V de Cramér=0.325), lo que indica que el canal por el cual el turista se informa sobre su destino varía de forma sistemática según el segmento al que pertenece. Bajo las condiciones de simulación, los turistas culturales tienden a informarse a través de familiares y amigos, los de naturaleza a través de agencias de viaje, y los familiares mediante publicidad en medios masivos, lo cual es coherente con los perfiles de referencia utilizados en la generación de datos.

La Ocupación muestra la segunda asociación más fuerte con la categoría ($\chi^2=185.90$, $gl=4$, $p<0.001$; $V=0.295$), evidenciando que el tipo de actividad laboral del turista es un diferenciador relevante entre segmentos. En particular, los perfiles simulados indican que los trabajadores del sector público se concentran en el segmento de turismo de naturaleza, los trabajadores por cuenta propia en el cultural, y los empleados de empresas privadas en el familiar. La duración del viaje ($V = 0,242$) y el nivel de estudio ($V = 0,229$) presentan asociaciones débiles, lo que indica una contribución limitada a la diferenciación entre segmentos bajo condiciones de simulación..

Las variables Grupo de Viaje ($V=0.207$), Método de Reserva ($V=0.179$) y Medio de Transporte ($V=0.168$) presentan asociaciones débiles pero estadísticamente significativas con la categoría. Esto sugiere que, si bien estas características presentan diferencias entre segmentos, su poder discriminante es más limitado que el de las variables anteriores, lo que es informativo para el diseño de estrategias de segmentación en investigaciones futuras con datos reales. En conjunto, los resultados del análisis de asociación son coherentes con la estructura de los datos sintéticos generados y con los perfiles teóricos de referencia, aunque, como se ha señalado, su alcance interpretativo está restringido al contexto de simulación controlada del presente estudio.

Figura 5. Selección del número óptimo de clusters mediante la curva de codo



Nota: La curva muestra la variación de la medida utilizada para evaluar el agrupamiento en función del número de clústeres y permite identificar el punto de codo.

El estudio analiza la distribución de datos utilizando el algoritmo de K-medias y la identificación del punto de codo. En la gráfica, se observa que los datos pueden agruparse de manera óptima en tres clusters, lo cual se refleja en la inclinación del "punto de codo". Un centroide representa la ubicación real o imaginaria que representa el centro de cada grupo, a través de la reducción de la suma de cuadrados dentro de cada grupo, cada punto de datos se asignó a uno de los grupos. En el contexto de K-medias, se enfocó en encontrar el promedio de los datos, es decir, el centroide.

En la gráfica, se observa que la disminución de la inercia (o SSE) es más significativa en los primeros valores de k, lo que sugiere que los datos se benefician de un mayor número de clústeres hasta cierto punto. Sin embargo, a partir de $k=3$, la disminución de la inercia se vuelve mucho menos pronunciada, lo que indica que agregar más clústeres no mejora significativamente la compacidad de los grupos formados. Este comportamiento marca el llamado punto de codo, que representa el equilibrio óptimo entre la cantidad de clústeres y la reducción de la varianza interna de los grupos.

El punto de codo en $k=3$ indica que tres es el número óptimo de clústeres para segmentar los datos. En este punto, los datos se agrupan de manera efectiva en tres grupos bien definidos, maximizando la cohesión interna de cada clúster sin incurrir en una sobresegmentación. Es decir, aumentar el número de clústeres más allá de tres no produciría una mejora significativa en la estructura de los datos, sino que solo complicaría el modelo y podría generar grupos más pequeños y menos significativos.

Este análisis tiene implicaciones importantes para la interpretación de los perfiles turísticos en la Guajira, ya que permite segmentar a los turistas en tres grupos bien diferenciados según sus características. La capacidad de determinar estos clústeres óptimos mejora la personalización de las ofertas turísticas, al identificar patrones claros en las preferencias y comportamientos de los distintos segmentos.

5 Discusión

El objetivo principal de este estudio fue desarrollar, en el marco de un ejercicio de simulación controlada, un modelo de segmentación basado en el algoritmo K-medias y datos sintéticos, con el propósito de explorar su potencial para la caracterización de perfiles turísticos en La Guajira y contribuir al diseño de estrategias personalizadas de oferta turística. Bajo las condiciones de simulación establecidas, el modelo demostró coherencia interna en la separación de los tres segmentos definidos (cultural, naturaleza y familiar), lo cual sugiere de forma hipotética y pendiente de validación empírica que este enfoque podría ser promisorio para optimizar la oferta turística en La Guajira. Los resultados obtenidos mostraron coherencia interna en la segmentación de los

perfiles sintéticos en los grupos cultural, de naturaleza y familiar, permitiendo explorar su caracterización a partir de variables sociodemográficas y preferencias de viaje. Esta segmentación proporciona información clave para la toma de decisiones en el sector turístico, facilitando la personalización de servicios y estrategias de marketing dirigidas a cada grupo específico.

En este orden de ideas, los hallazgos de este estudio coinciden con investigaciones previas que han demostrado la utilidad de la inteligencia artificial en la segmentación de mercados turísticos. Por ejemplo, García-Reinoso et al. (2020) identificaron perfiles de turistas en Ecuador mediante análisis factorial y algoritmos de clasificación jerárquica, evidenciando que la segmentación contribuye a una mejor planificación de estrategias turísticas. Asimismo, investigaciones como la de Filieri et al. (2021) han resaltado la importancia de la IA en la optimización de procesos y la personalización de experiencias turísticas. Este estudio complementa dichas investigaciones al aplicar técnicas avanzadas de modelado de datos en el contexto colombiano.

Desde la perspectiva de la Industria 4.0, el uso de la IA en la segmentación turística se fundamenta en la automatización de la toma de decisiones basada en datos (Almada-Lobo, 2016). El big data aplicado al turismo indica que el análisis de grandes volúmenes de información permite identificar tendencias y comportamientos del cliente de manera más eficiente (Bowen y Whalen, 2017). En este estudio, los perfiles sintéticos asociados al segmento cultural se caracterizaron por experiencias inmersivas y por la preferencia de información proveniente de familiares y amigos, mientras que los perfiles del segmento de naturaleza priorizaron la exploración del entorno y la consulta a agencias de viaje. Por su parte, los turistas familiares muestran un interés mayor en el bienestar personal y utilizan principalmente medios de comunicación tradicionales como la televisión y la radio. Estos resultados se alinean con los planteamientos teóricos sobre la influencia de la digitalización y la automatización en la personalización de servicios turísticos (Gallego Gómez y Vaquero Frías, 2022).

Si bien el estudio logró una segmentación efectiva, la generación de datos sintéticos puede no reflejar con exactitud el comportamiento turístico real en La Guajira. A su vez, el modelo K-medias es sensible a la inicialización de centroides, lo que indica la necesidad de comparar con otros métodos de agrupamiento, como el clustering jerárquico. Además, la disponibilidad limitada de datos reales sobre consumo turístico en la región restringió la validación empírica del modelo desarrollado.

El desbalance observado en la distribución del conjunto de datos sintéticos donde el segmento de Turismo de Naturaleza ($n=902$) es aproximadamente 2,4 veces mayor que el de Turismo Cultural ($n=382$) y 2,8 veces mayor que el de Turismo Familiar ($n=324$) fue intencional y responde a la distribución aproximada de la demanda turística en La Guajira reportada por estudios previos, donde el turismo de naturaleza constituye la categoría dominante (Cámara de Comercio de La Guajira, 2023; García-Reinoso et al., 2020). No obstante, es importante reconocer que el algoritmo K-medias es sensible al tamaño de los grupos y tiende a desplazar los centroides hacia las zonas de mayor

densidad de datos. En este estudio, dicho efecto podría haber favorecido la cohesión interna del clúster de Naturaleza en detrimento de los segmentos de menor representación. Esta limitación deberá abordarse en investigaciones futuras mediante la recolección de datos empíricos que permitan corroborar si el desbalance refleja efectivamente la realidad del mercado turístico de La Guajira o constituye un sesgo del proceso de simulación.

Esta investigación aporta significativamente a la literatura sobre la aplicación de IA en la segmentación del mercado turístico. La propuesta metodológica basada en K-medias y la generación de datos sintéticos representa una innovación en la optimización de estrategias turísticas mediante IA. Además, los resultados obtenidos ofrecen información relevante para la planificación y promoción de La Guajira como destino turístico, facilitando la adaptación de servicios a las necesidades de los diferentes segmentos de viajeros. En términos prácticos, este estudio puede servir como referencia para la implementación de modelos similares en otras regiones turísticas con características similares.

Dado que la inteligencia artificial continúa evolucionando, se recomienda explorar enfoques híbridos que combinen técnicas de aprendizaje supervisado y no supervisado para mejorar la precisión en la segmentación turística. Asimismo, futuras investigaciones podrían ampliar la base de datos con información en tiempo real obtenida de redes sociales y plataformas de reservas, permitiendo una caracterización más dinámica del comportamiento de los turistas. También sería pertinente incluir variables cualitativas, como satisfacción del cliente y fidelización, para fortalecer la interpretación de los resultados. Finalmente, se sugiere validar el modelo en otros destinos turísticos para evaluar su aplicabilidad y replicabilidad en distintos contextos geográficos y culturales.

6 Conclusiones

El análisis desarrollado en este estudio, realizado en su totalidad sobre datos sintéticos generados en un entorno de simulación controlada, explora el potencial de la inteligencia artificial aplicada a la segmentación turística en La Guajira. A través del algoritmo K-medias y la codificación ordinal de siete variables sociodemográficas y de comportamiento de viaje, fue posible identificar tres perfiles diferenciados de turistas (cultural, de naturaleza y familiar) cuyas características son coherentes con la literatura de referencia. Los resultados sugieren, de forma hipotética, que este tipo de modelos podría contribuir a la personalización de estrategias de marketing y servicios turísticos, mejorando potencialmente la competitividad del destino una vez que el modelo sea validado con datos reales. Sin embargo, es fundamental subrayar que estas conclusiones no constituyen evidencia empírica sobre el comportamiento real de los turistas en La Guajira, sino proposiciones a verificar en investigaciones futuras con bases de datos de turistas efectivos.

El uso de datos sintéticos generados mediante codificación ordinal discreta representa una limitación central de este estudio, ya que las fronteras de decisión entre categorías fueron predefinidas en el

proceso de generación, lo que impide afirmar que el modelo haya realizado un aprendizaje genuinamente exploratorio. Esta limitación, inherente al diseño metodológico adoptado ante la escasez de datos reales en La Guajira, deberá superarse en investigaciones futuras mediante la recopilación de datos empíricos de turistas reales y la reaplicación del modelo en ese contexto. Asimismo, se recomienda explorar enfoques complementarios como el análisis de silueta con datos reales, la comparación con algoritmos de agrupamiento jerárquico, y la incorporación de variables cualitativas como la satisfacción del turista. El impacto económico potencial del modelo sobre la competitividad del destino es una hipótesis de trabajo que únicamente podrá evaluarse tras la validación empírica del enfoque propuesto, constituyendo una agenda de investigación prioritaria para el desarrollo del sector turístico de La Guajira.

Referencias bibliográficas

- Almada-Lobo, F. (2015). The Industry 4.0 revolution and the future of manufacturing execution systems (MES). *Journal of Innovation Management*, 3(4), 16-21. https://doi.org/10.24840/2183-0606_003.004_0003
- Alonso Almeida, M. del M. (2019). Robots, inteligencia artificial y realidad virtual: una aproximación en el sector del turismo. *Cuadernos de Turismo*, 1(44), 13–26. <https://doi.org/10.6018/turismo.44.404711>.
- Balance del turismo en Semana Santa 2021 en el departamento de La Guajira. (2021, 19 de abril). Diario del Norte. <https://diariodelnorte.net/laguajira/balance-del-turismo-en-semana-santa-2021-en-el-departamento-de-la-guajira/>
- Belanche, D., Casaló, L.V., y Flavián, C. (2021) Frontline robots in tourism and hospitality: service enhancement or cost reduction?. *Electron Markets*, 31, 477-492. <https://doi.org/10.1007/s12525-020-00432-5>
- Blöcher, K., y Alt, R. (2021). AI and robotics in the European restaurant sector: Assessing potentials for process innovation in a high-contact service industry. *Electron Markets*, 31 (3), 529–551. <https://doi.org/10.1007/s12525-020-00443-2>
- Bowen, J., y Whalen, E. (2017). Trends that are changing travel and tourism, *Worldwide Hospitality and Tourism Themes*, 9 (6), 592-602. <https://doi.org/10.1108/WHATT-09-2017-0045>
- Bulchand-Gidumal, J. (2020). Impact of artificial intelligence in travel, tourism, and hospitality. En Z. Xiang, M. Fuchs, U. Gretzel, & W. Höpken (Eds.), *Handbook of e-Tourism* (pp. 1943–1962). Springer. https://doi.org/10.1007/978-3-030-05324-6_110-1

- Cámara de Comercio de la Guajira. (2023, 20 de abril). Cámara de Comercio analiza impacto económico de la temporada de Semana Santa. <https://camaraguajira.org/camara-de-comercio-analiza-impacto-economico-de-la-temporada-de-semana-santa/>
- Chi, O. H., Gursoy, D., y Chi, C. G. (2022). Tourists' Attitudes toward the Use of Artificially Intelligent (AI) Devices in Tourism Service Delivery: Moderating Role of Service Value Seeking. *Journal of Travel Research*, 61(1), 170-185. <https://doi.org/10.1177/0047287520971054>
- Filieri, R., D'Amico, E., Destefanis, A., Paolucci, E. y Raguseo, E. (2021), Artificial intelligence (AI) for tourism: an European-based study on successful AI tourism start-ups, *International Journal of Contemporary Hospitality Management*, 33(11), 4099–4125. <https://doi.org/10.1108/IJCHM-02-2021-0220>
- Gallego Gómez, C., y Vaquero Frías, L. (2022). Artificial intelligence and sustainable tourism development: The value of partnerships. *ESIC Market*, 53(3), e281. <https://doi.org/10.7200/esicm.53.281>
- García-Reinoso, N., García-Moreira, D., y Quintero-Ichazo, Y. (2020). Segmentación de la demanda turística colombiana que visitan la Zona de Planificación 1 más la provincia de Pichincha en Ecuador. *Revista Interamericana de Ambiente y Turismo*, 16(2), 136–152. <https://doi.org/10.4067/S0718-235X2020000200136>
- Gómez Muñoz, D. M., Casallas Bernal, J. D., & Rodríguez Molano, J. I. (2023). Aplicaciones del Machine Learning en el turismo – estudio en Colombia y sus zonas de posconflicto. *Revista Turismo y Desarrollo*, 40, 93–102. <https://doi.org/10.34624/rtd.v40i0.31465>.
- Hayat, K. (2023). Leveraging artificial intelligence in hospitality and tourism for success. In S. K. Dixit & S. Piramanayagam (Eds.), *Teaching Cases in Tourism, Hospitality and Events* (pp. 214–220). CABI. <https://doi.org/10.1079/9781800621022.0019>
- Hermosa Del Vasto, P. M., y Arco Castro, M. L. (2024). Inteligencia artificial (IA) en turismo sostenible: Análisis bibliométrico. *Cuadernos de Turismo*, (53), 157–185. <https://doi.org/10.6018/turismo.616431>
- Ivanov, S., & Webster, C. (2019). Conceptual framework of the use of robots, artificial intelligence and service automation in travel, tourism, and hospitality companies. In S. Ivanov & C. Webster (Eds.), *Robots, artificial intelligence and service automation in travel, tourism and hospitality* (pp. 7–37). Emerald Publishing Limited. <https://doi.org/10.1108/978-1-78756-687-320191001>
- Ketui, N., Pengpun, P., Boonmeeprakob, K., Khlaisamniang, P., & Supnithi, T. (2024). MAHASAMUT: AI-powered Thai tourism using multimodal agents. *CEUR Workshop Proceedings*, 3862, 56–65. <https://ceur-ws.org/Vol-3862/paper5.pdf>

- Li, J., Xu, L., Tang, L. D., Wang, S., y Li, L. (2018). Big data in tourism research: A literature review. *Tourism Management*, 68, 301–323. <https://doi.org/10.1016/j.tourman.2018.03.009>
- Makridakis, S. (2017). The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms. *Futures*, 90, 46–60. <https://doi.org/10.1016/j.futures.2017.03.006>.
- Más-Ferrando, A., Ramón-Rodríguez, A. B., & Aranda-Cuéllar, P. (2020). La revolución digital en el sector turístico: Oportunidad para el turismo en España. *EKONOMIAZ. Revista Vasca de Economía*, 98(3), 229–251. <https://doi.org/10.69810/ekz.1377>
- Penagos-Londoño, G. I., Rodríguez-Sanchez, C., Ruiz-Moreno, F., y Torres, E. (2021). A machine learning approach to segmentation of tourists based on perceived destination sustainability and trustworthiness. *Journal of Destination Marketing and Management*, 19, 100532. <https://doi.org/10.1016/j.jdmm.2020.100532>
- Rodríguez Correa, P., Valencia-Arias, A., Garcés-Giraldo, L. F., Castañeda Rodríguez, L. del R., Moreno López, G. A., & Benjumea-Arias, M. (2023). Tendencias en el uso de inteligencia artificial en el sector del turismo. *Revista Turismo & Desarrollo*, 40, 81–92. <https://doi.org/10.34624/rtd.v40i0.31447>
- Samala, N., Katkam, B. S., Bellamkonda, R. S., y Rodríguez, R. V. (2022). Impact of AI and robotics in the tourism sector: A critical insight. *Journal of Tourism Futures*, 8(1), 73–87. <https://doi.org/10.1108/JTF-07-2019-0065>
- Syam, N., y Sharma, A. (2018). Waiting for a sales renaissance in the fourth industrial revolution: Machine learning and artificial intelligence in sales research and practice. *Industrial Marketing Management*, 69, 135–146. <https://doi.org/10.1016/j.indmarman.2017.12.019>
- Zambrano Salazar, C. M., Cárdenas Vela, M. E., y Góngora Cheme, D. W. (2024). Inteligencia Artificial para la promoción turística de la Amazonía y la percepción del usuario digital. *Revista Social Fronteriza*, 4(5), e45464. [https://doi.org/10.59814/resofro.2024.4\(5\)464](https://doi.org/10.59814/resofro.2024.4(5)464)