



Universidad de Jaén
Centro de Estudios de Postgrado

Trabajo Fin de Máster

**UNA MIRADA
ANALÍTICA A LA
DEMANDA TURÍSTICA
INTERNA EN
ECUADOR: RELACIÓN
MULTIDIMENSIONAL
ENTRE VISITANTE,
VIAJE Y DESTINO**

Alumna: Miriam Vanessa Hinojosa Ramos

Tutor: Prof. Dr. Alejandro Rabasa Dolado

Septiembre, 2023

DEDICATORIA

A los amores eternos de mi vida, mis motores y pilares fundamentales, mi madre Miriam y mis abuelos Alba y Ecuador, por su infinito amor, permanente apoyo e inmensa comprensión desde que vine al mundo. Todo lo que hago es por y para ustedes, con todo mi corazón.

A mi querida y recordada tía Yolanda, fuente permanente de inspiración y fortaleza. Sé que vives pendiente de mis pasos y sigues vibrando con cada conquista mía como si fuera tuya.

A mi hermano de otra madre, José, por ser una constante en mi familia, compartir momentos y estar siempre cerca cuando lo hemos necesitado.

A mis alumnos, sin excepción, los primeros que tuve, los que siguieron después, los que dirigí en su titulación y los que seguramente conoceré, por permitirme compartirles lo poco que sé y por enseñarme lo mucho que me queda por aprender.

AGRADECIMIENTO

A mi amiga Gisella por animarme a iniciar esta aventura académica juntas, recordarme que soy una mujer de ciencia y acompañarme en el día a día.

A mi amigo Jorge por compartir conmigo ideas, opiniones y, sobre todo, conocimientos, que han sido clave para este proceso de aprendizaje.

A mis colegas, Pamela y Jaime, del Ministerio de Turismo del Ecuador, por brindarme toda la información, colaboración y ayuda requeridos para el desarrollo de este trabajo. Sin ustedes, no me hubiera sido posible contribuir a mi país con este modesto aporte.

A todos quienes han sido parte de mi formación académica, especialmente los docentes de este Máster, quienes supieron compartir con nosotros toda su experticia en cada temática.

A mi profesor y tutor Alejandro, gracias por toda la guía y el acompañamiento de estos meses desde la libertad y la confianza. Ha sido gratificante todo lo aprendido y espero que este sea el inicio de un camino lleno de muchos proyectos emocionantes.

RESUMEN

La inteligencia turística se sustenta en los datos, posicionados mundialmente como el nuevo petróleo del siglo XXI. Esto ha propiciado que muchos países se conviertan en destinos inteligentes; sin embargo, el proceso hacia el máximo aprovechamiento del conocimiento para la toma de decisiones en el sector, apenas empieza en Ecuador. Así, este trabajo procuró profundizar en la comprensión de la demanda turística interna y la gestión de cimientos de inteligencia turística en torno al visitante, viaje y destino. Se utilizó una muestra proporcionada por el Ministerio de Turismo y se implementaron técnicas analíticas de ciencia de datos de carácter descriptivo y predictivo para la validación de las hipótesis planteadas. Desde el contexto combinado de la aplicación de modelos y la generación de resultados a partir de datos, se extrajeron patrones, identificaron segmentos, establecieron predicciones, visibilizaron tendencias geográficas y propusieron acciones como apoyo complementario a las estrategias turísticas del país.

Palabras claves: inteligencia turística, demanda turística, turismo interno, ciencia de datos, aprendizaje automático, sistema de información geográfica

ABSTRACT

Tourism intelligence is based on data, positioned worldwide as the new oil of the 21st century. This has led many countries to become smart destinations; however, the process towards taking maximum advantage of knowledge for decision-making in the sector is just beginning in Ecuador. Thus, this work sought to deepen the understanding of the domestic tourist demand and the management of tourism intelligence foundations around the visitor, trip and destination. A sample provided by the Ministry of Tourism was used and analytical data science techniques of a descriptive and predictive nature were implemented to validate the hypotheses raised. From the mixed context of models' application and results' generation from data, patterns were extracted, segments were identified, predictions were established, geographical trends were made visible, and actions were proposed as complementary support to the country's tourism strategies.

Keywords: tourism intelligence, tourism demand, domestic tourism, data science, machine learning, geographic information system

Índice general

CAPÍTULO I. INTRODUCCIÓN	1
1.1. Justificación y oportunidad de este Trabajo Fin de Máster	1
1.1.1. Importancia del tema y área de investigación.....	1
1.1.2. Carencias en la literatura.....	2
1.2. Planteamiento del problema	4
1.3. Hipótesis y objetivos	4
1.3.1. Hipótesis	4
1.3.2. Objetivo General.....	5
1.3.3. Objetivos Específicos.....	6
1.4. Planteamiento metodológico general.....	6
1.5. Estructura de la investigación.....	6
CAPÍTULO II. MARCO GEOGRÁFICO	8
CAPÍTULO III. MARCO CONCEPTUAL	10
3.1. Demanda Turística.....	10
3.1.1. Demanda Turística Interna.....	10
3.2. Inteligencia Turística	11
3.2.1. Destino Turístico Inteligente.....	12
3.3. Gestión del Conocimiento	13
3.4. Inteligencia de Negocios	13
3.5. Ciencia de Datos	13
3.5.1. Datos Masivos.....	13
3.5.2. Minería de Datos.....	13
3.5.3. Inteligencia Artificial	14
3.5.3.1. Aprendizaje Automático	14
3.5.3.1.1. Asociación	15
3.5.3.1.2. Agrupamiento o Segmentación	15
3.5.3.1.3. Selección de características o Ranking de atributos	15
3.5.3.1.4. Clasificación	15
3.5.4. Sistema de Información Geográfica.....	15
3.5.4.1. Datos vectoriales	16
3.5.4.2. Atributos de datos vectoriales	16
3.5.4.3. Mapa de coropletas.....	16

CAPÍTULO IV. MARCO TEÓRICO	17
4.1. Ciencia de Datos para generación de Inteligencia Turística.....	17
4.2. Aprendizaje Automático en Turismo	20
4.3. Aproximaciones a la Demanda Turística Interna: Caso Ecuador	23
CAPÍTULO V. MARCO METODOLÓGICO	25
5.1. Diseño del marco muestral	25
5.2. Diseño del cuestionario	27
5.3. Recopilación de datos	27
5.4. Preprocesamiento de datos	28
5.5. Procesamiento de datos	30
CAPÍTULO VI. ANÁLISIS E INTERPRETACIÓN DE RESULTADOS.....	32
6.1. Análisis del visitante.....	32
6.2. Análisis del viaje	35
6.3. Análisis del destino.....	55
6.4. Discusión	64
6.5. Propuesta de acciones	68
CAPÍTULO VII. CONCLUSIONES	70
BIBLIOGRAFÍA.....	72
ANEXOS.....	85
Anexo 1. Guion de entrevista	85
Anexo 2. Perfil de entrevistados y evidencias fotográficas.....	87
Anexo 3. Cuestionario definitivo.....	88
Anexo 4. Resumen de construcción de perfilados.....	92

Índice de tablas

Tabla 1. Distribución muestral preliminar	26
Tabla 2. Distribución muestral final.....	26
Tabla 3. Descripción de atributos/variables de la base de datos	28
Tabla 4. Análisis del visitante	30
Tabla 5. Análisis del viaje	30
Tabla 6. Análisis del destino	31

Índice de figuras

Figura 1. Mapa del Ecuador	8
Figura 2. Material publicitario de la campaña “All you need is Ecuador”	9
Figura 3. Pirámide del modelo DIKW	11
Figura 4. Ejes fundamentales de los Destinos Turísticos Inteligentes	12
Figura 5. Aprendizaje automático: métodos y algoritmos	14
Figura 6. Mapa de coropletas	16
Figura 7. Material promocional de las campañas en redes sociales.....	27
Figura 8. Patrones por edad e ingreso	32
Figura 9. Patrones por nivel de estudios y gasto total.....	32
Figura 10. Patrones socioeconómicos del visitante.....	33
Figura 11. Patrones socioeconómicos del visitante con región Costa como destino	34
Figura 12. Patrones socioeconómicos del visitante con región Sierra como destino.....	34
Figura 13. Clústeres por gasto en alojamiento vs transporte aéreo	35
Figura 14. Clústeres por gasto en alojamiento vs transporte terrestre	36
Figura 15. Clústeres por gasto en alojamiento vs alimentos y bebidas.....	36
Figura 16. Clústeres por gasto total y tamaño de grupos	37
Figura 17. Clústeres por pernoctaciones y tamaño de grupos.....	37
Figura 18. Clústeres de centennials por pernoctaciones y gasto total.....	38
Figura 19. Clústeres de millennials por pernoctaciones y gasto total	38
Figura 20. Clústeres de generación X por pernoctaciones y gasto total	39
Figura 21. Clústeres de hoteles por edad y gasto total	39
Figura 22. Clústeres de hostales por edad y gasto total	40
Figura 23. Clústeres de viviendas alquiladas por edad y gasto total.....	41
Figura 24. Clústeres de viviendas propias por edad y gasto total	41
Figura 25. Clústeres de transporte aéreo por edad y gasto total.....	42
Figura 26. Clústeres de autobús/flota por edad y gasto total	42
Figura 27. Clústeres de vehículo alquilado por edad y gasto total.....	43
Figura 28. Clústeres de vehículo propio por edad y gasto total	43
Figura 29. Patrones según motivo, grupos, antelación y gasto total	44
Figura 30. Patrones según transporte, medios de comunicación, redes, plataformas y edad...	45
Figura 31. Patrones en complejidad de aspectos del viaje	45
Figura 32. Patrones en los gastos del viaje.....	46

Figura 33. Árbol de decisión para el umbral de gasto a partir de atributos del visitante	47
Figura 34. Selección de atributos de viaje con gasto total como variable objetivo	48
Figura 35. Árbol de decisión para el umbral de gasto a partir de atributos del viaje.....	49
Figura 36. Selección de atributos de visitante-viaje con gasto total como variable objetivo ..	51
Figura 37. Árbol de decisión para el umbral de gasto a partir de atributos del visitante-viaje	52
Figura 38. Selección de atributos de visitante-viaje con intención futura como variable objetivo	53
Figura 39. Árbol de decisión para la intención futura a partir de atributos del visitante-viaje	54
Figura 40. Selección de atributos de visitante-viaje con región de destino como variable objetivo	56
Figura 41. Árbol de decisión para la región de destino a partir de atributos del visitante-viaje	56
Figura 42. Selección de atributos de visitante-viaje con región de destino como variable objetivo de viajeros vacacionales	58
Figura 43. Árbol de decisión para la región de destino a partir de atributos del visitante-viaje de viajeros vacacionales	59
Figura 44. Mapa de Frecuencia de Visita por provincias.....	61
Figura 45. Mapa de Pernoctaciones Promedio por provincias.....	62
Figura 46. Mapa de Gasto Total Promedio por provincias	63

CAPÍTULO I. INTRODUCCIÓN

1.1. Justificación y oportunidad de este Trabajo Fin de Máster

1.1.1. Importancia del tema y área de investigación

La revolución de la inteligencia de mercados en el contexto de destinos turísticos ha permitido que se pueda obtener información sobre necesidades y preferencias reales de los clientes para convertirse en destinos inteligentes, donde se pueda recopilar, intercambiar y procesar datos. Esto, sin duda, contribuye a ofrecer la mejor experiencia turística de acuerdo con el perfil del turista y, al mismo tiempo, a identificar y desarrollar productos y servicios personalizados (Eichelberger et al., 2020; Gretzel et al., 2015; Veloz et al., 2022).

En este sentido, lo que diferencia a un ente turístico de su(s) competidor(es), es la cantidad y la calidad de la información que se gestiona a partir de los datos. En muchos casos, esto permite adquirir un conocimiento profundo sobre los turistas y llegar a la toma de decisiones más adecuadas en la definición de nuevos productos o servicios, adaptabilidad a nuevos mercados, aumento de las ventajas competitivas, e incluso en la supervivencia durante tiempos de crisis, propiciando y construyendo lo que se conoce como inteligencia turística bajo el modelo DIKW (por sus siglas en inglés “D” - data, “I” - information, “K” - knowledge y “W” - wisdom) (Calle-Lamelas et al., 2023; Del Chiappa & Baggio, 2015; Mabrian Technologies, 2020; Ramos, 2023).

Precisamente el sector turístico ha registrado a la fecha, tres episodios de crisis en su historia, el primero con el brote de SARS (en inglés: Severe Acute Respiratory Syndrome) en 2003, el segundo con la crisis económica mundial de 2009, y el tercero se le atribuye a la pandemia de la COVID-19 (enfermedad por coronavirus de 2019). Este último generó una caída sin precedentes del turismo a nivel internacional. De acuerdo a la Organización Mundial del Turismo (OMT), bajo el contexto de la pandemia, los flujos turísticos internacionales cambiaron de manera abrupta en todos los destinos turísticos, así como lo hicieron los patrones de comportamiento y consumo de los turistas. Según expertos, se prevé que dichos flujos se recuperen de manera progresiva, especialmente a partir de 2023; pero, además, se han observado tendencias de corto a largo plazo en el comportamiento de los viajeros priorizando el turismo doméstico en muchos mercados por la preferencia de viajar a áreas más próximas (OMT, s. f.-a, s. f.-b; ONU, 2020; ONU & CEPAL, 2020). Este es un fenómeno global y que de manera particular se ha podido observar en el ámbito geográfico de Ecuador, donde se ubica este trabajo.

Con este antecedente, a nivel nacional, el sector turístico ha centrado su enfoque (y deberá continuar haciéndolo) en el turismo interno que es el que, previsiblemente, va a dinamizar el sector por un buen tiempo, y de ahí la importancia de conocer en profundidad al visitante local que se desplaza por el territorio ecuatoriano entre sus 4 regiones geográficas (Bárcena & Cimoli, 2020; Diario El Universo, 2023b; Ministerio de Turismo, 2020). A partir de la necesidad de generar inteligencia turística y considerando las técnicas e instrumentos adquiridos en el contexto del Máster que motiva este trabajo, se destacan las diferentes disciplinas vinculadas con la Ciencia de Datos como la Inteligencia Artificial, el Machine Learning, el Big Data o el Data Mining, entre las más populares. Estas se han mostrado extraordinariamente eficientes al ser aplicadas en variados escenarios de negocios, particularmente en turismo. Sin embargo, la transferencia de estas técnicas a diferentes problemas, sigue siendo una tarea minuciosa que requiere parametrización y adaptación particular a cada caso por resolver (Höpken & Fuchs, 2022; Más-Ferrando et al., 2020; Moreno-Izquierdo et al., 2022; Rodríguez et al., 2023).

En síntesis, una sistemática explotación de la información mediante las técnicas adecuadas de Ciencia de Datos, puede ayudar a un conocimiento exhaustivo del turista interno, y ello, a su vez puede conducir a planificar de manera ordenada y anticipadamente los recursos turísticos en Ecuador, así como a definir estrategias de promoción dirigidas a los colectivos potencialmente más interesados en determinados tipos de oferta o de destino y tipología de viaje.

1.1.2. Carencias en la literatura

A nivel internacional, en cuanto a la caracterización de la demanda turística interna, el alcance principalmente exploratorio y descriptivo del perfil del turista y su comportamiento o preferencias/necesidades asociadas al viaje y al destino, como herramienta clásica, ha permitido conocer cierto nivel de información; pero, en el contexto de la generación de inteligencia turística no ha permitido llegar a la cúspide de la pirámide del modelo DIKW (Alén et al., 2016; Bel et al., 2015; Cava Jiménez et al., 2019; Curtin, 2013; George & Booyens, 2014; Getz & Brown, 2006; Lang & O'leary, 1997; Molera & Albaladejo, 2007; Park & Yoon, 2009; Sánchez-Cañizares & López-Guzmán, 2012; Santos et al., 2016; Smith & Puczkó, 2015; Villamediana-Pedrosa et al., 2020). Esta necesidad de generar inteligencia turística se ha puesto de manifiesto más aún, después de la pandemia, donde los estudios se decantan por un examen a ese turista local que es clave en la reactivación del sector para cada país (Corzo Arevalo & Guatibonza, 2021; Hernández Flores et al., 2020; Mendieta-Aragón, 2022; Suty, 2021).

Sin embargo, al no aprovechar, mayoritariamente, la profundidad de los datos a través de la incorporación de técnicas predictivas adecuadas, las decisiones tomadas por los organismos, Secretarías o Ministerios de Turismo y de actores clave dentro de toda la cadena de valor, no necesariamente conllevan a una mejor planificación y gestión sostenible de la actividad (Fernández Alcantud et al., 2017; Ramos, 2023).

De ahí que, solo una cantidad reducida a nivel mundial, pero en crecimiento, puede considerarse como destinos turísticos inteligentes, dado que han incorporado recientemente el concepto de generación de inteligencia turística, articulando diferentes disciplinas de ciencia de datos en complemento al enfoque tradicional del perfil meramente, para identificar patrones de consumo y pronosticar la demanda turística, entre otros fines (Abdualgalil & Abraham, 2020; Bi et al., 2023; Chung et al., 2022; Claveria et al., 2016; W. Li, 2022; Rabasa et al., 2018).

A nivel nacional, en Ecuador, el último estudio descriptivo publicado por parte del Ministerio de Turismo referente a perfiles de turismo internacional (receptor y emisor), se publicó en 2018 (Ministerio de Turismo, 2018). Si bien, el Ministerio pone a disposición, previa solicitud, estadísticas de la actividad turística (demanda, oferta e indicadores económicos) para todos los usuarios internos y externos vinculados y/o interesados en el sector, con el fin de contribuir al análisis del turismo, a sustentar proyectos de desarrollo, diseño de política pública y temas investigativos relacionados al turismo nacional, no se evidencia todavía un enfoque de aprovechamiento en profundidad de los datos con los que se cuenta (Ministerio de Turismo, 2023a).

La academia, por su parte, en el contexto de la reactivación turística, desde 2020 en adelante, tampoco apuesta mayormente al análisis de la demanda turística interna considerando técnicas predictivas, únicamente se mantienen en el alcance tradicional exploratorio y descriptivo. Incluso, se centran en alguna tipología de turismo como de naturaleza o comunitario, dentro de casos puntuales, también incluyen analítica de algún segmento específico como de la generación Z, o bien se enfocan en los turistas que visitan algún destino, sea provincia o ciudad (Almache, 2021; Bautista, 2021; Herrera Díaz et al., 2020; Mera et al., 2020; Molina Velásquez et al., 2020; Sánchez Ruiz & Paladines Sarango, 2021; Toledo et al., 2020). Únicamente, un estudio aborda el comportamiento de la demanda turística interna en la reactivación post pandemia a nivel nacional; sin embargo, lo hace desde el enfoque clásico, ya mencionado (Molina Velásquez et al., 2022).

1.2. Planteamiento del problema

En Ecuador, es una realidad que la entidad estatal que maneja el turismo del país ha ido incorporando mayor nivel tecnológico al seguimiento del sector y demuestra interés por la realización de análisis estadísticos, pero gestiona la información de manera todavía rudimentaria y aún no incorpora mecanismos basados en modelos analíticos de la moderna Ciencia de los Datos que permitirían hacer predicciones más precisas sobre visitante, viaje y destino en el contexto de la demanda turística interna (J. Paguay, comunicación personal, 1 de junio de 2023).

Con actualización del 17 de mayo de 2023, el Ministerio de Turismo dispone de un visualizador de información turística que incluye información económica, entradas y salidas internacionales, áreas naturales protegidas, indicadores de alojamiento, índice de competitividad, conectividad, Observatorio Nacional de Turismo (Observatorio de Galápagos, Zona 2 conformada por las provincias de Napo, Pichincha y Orellana; y, Zona 7 conformada por Zamora Chinchipe, Loja y El Oro), ventas, inversiones, establecimientos registrados o catastro turístico, publicaciones, indicadores PND 2021-2025; y, feriados nacionales (festivos) (Ministerio de Turismo, 2023b). Resulta evidente que se ha iniciado el camino hacia la generación de un Sistema de Inteligencia Turística desde el Ministerio de Ecuador, siendo un hecho positivo su apuesta clara por la digitalización e incorporación de tecnología dentro del sector turístico ecuatoriano. Sin embargo, la división de Inteligencia de Mercados del Ministerio externaliza ciertos estudios de acuerdo a un presupuesto limitado y ello obstaculiza la ejecución de proyectos sólidos a medio plazo, dificultando aún más la concreción de programas de inteligencia turística a largo plazo, que obedezcan a una estrategia desde el punto de vista de mercado y que incorporen una clara componente predictiva, de aprovechamiento a profundidad de los datos, que complemente y facilite el proceso de toma de decisiones (J. Lastra, comunicación personal, 25 de mayo de 2023).

Para alcanzar un soporte en la toma de decisiones en el sector en aras de una mejor planificación y gestión sostenible de un destino, es imprescindible incorporar técnicas predictivas que, desde la administración pública, todavía no han implementado a partir de los datos que se generan.

1.3. Hipótesis y objetivos

1.3.1. Hipótesis

Partiendo de que es posible planificar y gestionar mejor un destino a través de modelos predictivos desde la perspectiva de la ciencia de datos que ayuden a la extracción de características del visitante, viaje y destino; y, de que estas técnicas que complementen lo

descriptivo deben ser adaptadas a las necesidades y particularidades de cada escenario concreto, para poder parametrizar adecuadamente los modelos y así obtener predicciones ad-hoc lo más precisas posibles que faciliten la toma de decisiones en el ámbito del turismo interno, se establecen las siguientes hipótesis:

H_1 : Existen diferentes patrones de comportamiento del visitante dependiendo de su condición demográfica y socioeconómica (edad, ingreso promedio mensual, estado civil, género y nivel de estudios).

H_2 : Existen segmentaciones de viajes similares entre sí, atendiendo a los porcentajes en conceptos de gasto, número de acompañantes, gasto total del viaje y pernoctaciones.

H_3 : Existen grupos de visitantes comunes entre sí, atendiendo a pernoctaciones y gasto total del viaje, en función de la edad.

H_4 : Existen agrupamientos de viajes comunes entre sí, atendiendo a edad y gasto total del viaje, en función del tipo de alojamiento y del medio de transporte.

H_5 : Existen diferentes patrones de viaje de acuerdo a motivo, tamaño de grupo, antelación para planificación, gasto total del viaje, medio de transporte, medios de comunicación, redes sociales, plataformas de viaje y edad del visitante.

H_6 : Existen diferentes patrones en la complejidad que experimenta el turista en varios aspectos del viaje como reservas, alojamiento, alimentación, transporte y entretenimiento.

H_7 : Existen diferentes patrones en los gastos antes, durante y después del viaje, así como del gasto total del viaje.

H_8 : Es posible predecir el umbral de gasto total del viaje, la intención de viaje a futuro y la región de destino en función de las características del visitante y del viaje en sí mismo.

H_9 : No todas las variables intervienen por igual en la decisión del gasto total del viaje, de hacer un nuevo viaje a corto plazo y de definir la región de destino.

H_{10} : Es posible visualizar patrones geográficos entre destinos en función de la frecuencia de visita, pernoctaciones promedio y gasto total promedio del viaje.

1.3.2. Objetivo General

Profundizar en la comprensión de la demanda turística interna del Ecuador mediante el uso de técnicas avanzadas de ciencia de datos y gestión de los cimientos de inteligencia turística en torno al visitante, viaje y destino.

1.3.3. Objetivos Específicos

- Aplicar técnicas descriptivas y predictivas para precisar y modelar las características de los visitantes, las tipologías de viajes y los destinos en sí mismos.
- Interpretar los resultados analíticos más relevantes, en cada caso, para la generación de conocimiento sobre la demanda turística interna del país.
- Proponer acciones que sirvan de soporte al proceso de toma de decisiones para una mejor planificación y gestión del turismo doméstico en el país.

1.4. Planteamiento metodológico general

El presente Trabajo de Fin de Máster inicia con la definición de la situación actual y problemática a resolver partiendo de la técnica de entrevista en profundidad a dos profesionales del turismo vinculados a la empresa pública, empresa privada y la academia (Anexos 1 y 2). Junto con esto, se estipula una revisión de la literatura existente para fundamentar el marco conceptual y el marco teórico de la investigación.

Posteriormente, se plantea la adquisición de los datos relativos a la demanda turística interna del año 2022 que dispone el Ministerio de Turismo del Ecuador, cuya recolección y sistematización fue validada por dicha entidad estatal. Se procede a la ejecución de tareas de preprocesamiento y técnicas analíticas de ciencia de datos de carácter descriptivo y predictivo sobre las características de los visitantes, las tipologías de viajes y los destinos en sí mismos, desde un enfoque cuantitativo para la interpretación y discusión de resultados, a más de la propuesta de acciones que sirvan de soporte al proceso de toma de decisiones para la planificación y gestión del turismo interno en el país.

Este planteamiento metodológico en su núcleo sigue el contexto model-data-driven donde se combina la aplicación de modelos de cualquier naturaleza (model-driven) y la generación de resultados a partir exclusivamente de los datos (data-driven) (Rabasa & Heavin, 2020). En este trabajo, el diseño de los análisis y su interpretación final se apoya en las premisas fundamentales del sector turístico; pero la parte analítica, se desarrolla de manera completamente independiente de cualquier conocimiento previo o sesgo.

1.5. Estructura de la investigación

En cuanto a la estructura, el presente trabajo de investigación se compone de siete capítulos. El capítulo I de introducción, mismo que recoge la justificación y oportunidad de este Trabajo de Fin de Máster, el planteamiento del problema, las hipótesis y objetivos, el planteamiento

metodológico general además de la estructura resumida. En el capítulo II se centra en el marco geográfico del estudio; es decir, el territorio nacional del Ecuador.

El capítulo III recoge los fundamentos conceptuales tanto de la demanda turística e inteligencia turística como de la Inteligencia de Negocios, Ciencia de Datos, Big Data, Minería de Datos, Inteligencia Artificial y sus métodos de Aprendizaje Automático (Machine Learning) así como de aplicaciones a través de Sistemas de Información Geográfica. Seguidamente, una vez presentados los fundamentos, el capítulo IV presenta la situación actual de la ciencia de datos al servicio de la inteligencia turística, haciendo un repaso del tipo de problemas que se han resuelto exitosamente, las carencias encontradas al respecto y siguientes puntos de interés desde la perspectiva tanto de investigación como de gestión del turismo en el contexto de la demanda turística interna para Ecuador.

El capítulo V recoge aspectos metodológicos sobre los datos de entrada y su respectiva fuente, además de un esquema de la metodología seguida en este estudio, incluyendo las tareas de preprocesamiento y técnicas analíticas aplicadas. La aplicación directa de la metodología se muestra en el capítulo VI junto con el detalle de los análisis llevados a cabo con diferentes técnicas de ciencia de datos en el contexto de la relación multidimensional entre visitante, viaje y destino. Adicionalmente, se discuten los resultados más relevantes de los análisis realizados, se ponen de manifiesto las limitaciones del trabajo, se precisan futuras líneas de investigación y se enumeran una serie de acciones, a corto y largo plazo, como propuesta para mejorar la planificación y gestión turística nacional. Finalmente, en el capítulo VII, se esbozan las conclusiones generales del trabajo junto con valoraciones finales para el sector turístico del país.

CAPÍTULO II. MARCO GEOGRÁFICO

El presente trabajo de investigación se sitúa en Ecuador, país ubicado en la región noroccidental de América del Sur. Limita con Colombia, al norte; con Perú, al sur y este; y, con el Océano Pacífico, al oeste. Es el cuarto país más pequeño del continente con una extensión de 256 370 km² y el más densamente poblado de América del Sur cobijando más de 18 millones de habitantes. La ciudad de Quito es su capital y Guayaquil la ciudad más numerosa con el puerto más importante del país. El idioma oficial es el castellano y la moneda es el dólar estadounidense. Ecuador se divide en 4 regiones, 24 provincias, 221 cantones y 1 499 parroquias. La región litoral o Costa incluye las provincias de Esmeraldas, Manabí, Santa Elena, Los Ríos, Guayas y El Oro; la región andina o Sierra comprende Carchi, Imbabura, Pichincha, Cotopaxi, Bolívar, Tungurahua, Chimborazo, Santo Domingo de los Tsáchilas, Cañar, Azuay y Loja; mientras la Amazonía o región oriental concentra las provincias de Sucumbíos, Napo, Orellana, Pastaza, Morona Santiago y Zamora Chinchipe; y finalmente, la región insular o Galápagos, cuya provincia lleva su mismo nombre (Cancillería del Ecuador, s. f.; Oficina de Información Diplomática Ecuador, 2023), tal como se muestra la figura 1.

Figura 1. Mapa del Ecuador



Fuente: Elaboración propia

Debido a su posición geográfica y a la Cordillera de los Andes, el Ecuador cuenta con cuatro regiones naturales con flora y fauna que lo convierten en uno de los países de mayor biodiversidad por kilómetro cuadrado. La cordillera atraviesa al país de norte a sur donde se asienta la sierra ecuatoriana, dejando a su flanco occidental, zonas de llanura y tierras bajas costeras, y a su flanco oriental, la espesa selva tropical y otros ecosistemas amazónicos. Las islas Galápagos por su parte son un grupo de 13 islas volcánicas situadas en el Océano Pacífico a unos 1 120 km de la costa (Oficina de Información Diplomática Ecuador, 2023). Así, la integración de tantos ecosistemas en una superficie pequeña, le ha conferido el título del “País de los Cuatro Mundos” como se aprecia en la figura 2.

Figura 2. Material publicitario de la campaña “All you need is Ecuador”



Fuente: (Ministerio de Turismo, 2014)

Ecuador es, además, uno de los principales exportadores de petróleo, banano, flores, camarones y cacao en el mundo. Y aunque estas actividades aportan significativamente al Producto Interno Bruto (PIB) del país, gracias a su patrimonio natural y cultural, también es uno de los destinos referentes de Sudamérica en recepción de turismo internacional. Incluso se proyecta que al finalizar 2023, el sector turístico represente cerca del 5% del PIB nacional por la contribución a la economía gracias a llegadas internacionales y flujos domésticos (WTTC, 2023).

CAPÍTULO III. MARCO CONCEPTUAL

En el presente capítulo se han considerado los conceptos en los cuales se sustenta el desarrollo de este Trabajo de Fin de Máster, mismos que se detallan a continuación.

3.1. Demanda Turística

La Organización Mundial de Turismo (OMT) relaciona íntimamente el concepto de demanda turística al proceso de toma de decisiones que los turistas, de forma individual o colectiva, realizan en la organización y ejecución de su viaje, dependiendo de sus necesidades o motivaciones. De acuerdo a dichas necesidades, se establecen distintos productos turísticos para satisfacer a la demanda. En ocasiones, la demanda turística puede verse afectada por variables económicas del país emisor como de los destinos turísticos. Según el Ministerio de Turismo del Ecuador, la demanda turística se constituye por el conjunto de consumidores o posibles consumidores de bienes y servicios turísticos (Naranjo Lluart & Martínez Rodríguez, 2022). Así cuando se estudia la demanda, típicamente se involucran variables como edad, género, profesión, lugar de procedencia, número de personas que viajan, nivel de estudios, ocupación actual, promedio de estancia, motivo del viaje, forma de alojamiento, gasto, medio publicitario que motivó el viaje, medio de transporte y principal destino. Aunque, también se suelen considerar indistintamente factores económicos, físicos, geográficos, psicológicos, entre otros (Rigol et al., 2009).

3.1.1. Demanda Turística Interna

Si bien la demanda turística puede clasificarse según diversos criterios, el presente trabajo acoge su división de acuerdo a las tres formas básicas de turismo según el lugar de origen de los turistas y el destino elegido por ellos, como:

- Turismo interno que incluye actividades realizadas por un visitante residente en el país de referencia.
- Turismo receptor que engloba actividades realizadas por un visitante no residente en el país de referencia.
- Turismo emisor que abarca actividades realizadas por un visitante residente fuera del país de referencia.

Estas formas pueden combinarse de varias maneras para derivar otras formas adicionales de turismo conocidas como turismo interior (interno y receptor), turismo nacional (interno y emisor) y turismo internacional (emisor y receptor) (OMT, s. f.-c).

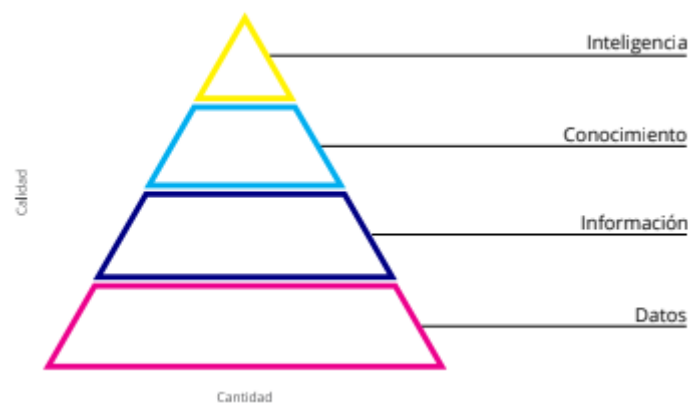
Por lo tanto, la demanda turística interna se enfoca usualmente en la demanda turística generada por el turismo interno, involucrándose directamente con el consumo turístico interno, es decir, el consumo turístico de un visitante residente en la economía de referencia (OMT, s. f.-c).

3.2. Inteligencia Turística

La inteligencia turística comprende el manejo de una serie de herramientas para la gestión estratégica del turismo, siendo aplicable en diferentes ámbitos territoriales desde lo local hasta lo internacional. Adicionalmente, la inteligencia turística implica la incorporación del análisis de datos de multitud de fuentes al proceso de toma de decisiones para el sector turístico. Dentro de los objetivos que se persiguen desde la inteligencia turística, se incluyen: uso de sistemas de apoyo a la decisión, análisis comparativos con competidores, definición de productos y destinos, identificación de mercados potenciales, segmentación de la clientela, análisis de tendencias y preferencias de consumidores (Rabasa, 2022).

Desde el enfoque de la pirámide del modelo DIKW, tal como se observa en la figura 3, la inteligencia turística se construye a partir de los datos que se encuentran en la parte ancha de la pirámide. Los datos son representaciones de un atributo o variable que describen hechos o sucesos. En el nivel superior se ubica la información, es decir, la organización de los datos ya procesados y que constituyen el mensaje que será interpretado por quienes los reciben. Luego, se encuentra el nivel del conocimiento, como la información colocada en contexto que permite su interpretación. Finalmente, se llega a la cúspide de la pirámide, donde se alcanza la inteligencia turística como el conocimiento vinculado para la toma de decisiones mediante procesos estructurados. Esto permite que los destinos puedan disponer de las herramientas gestionadas y compartir lo generado en el proceso a todos los actores claves (López de Ávila et al., 2015).

Figura 3. Pirámide del modelo DIKW



Fuente: (López de Ávila et al., 2015)

3.2.1. Destino Turístico Inteligente

Un destino turístico inteligente (DTI) se define como un espacio turístico innovador, accesible para todos, consolidado sobre una infraestructura tecnológica de vanguardia que garantiza el desarrollo sostenible del territorio, facilita la interacción e integración del visitante con el entorno e incrementa la calidad de su experiencia en el destino y la calidad de vida de los residentes. Los cinco ejes fundamentales de estos destinos incluyen: Gobernanza, Tecnología, Innovación, Accesibilidad y Sostenibilidad (SEGITTUR, 2015), mismos que se resumen en la figura 4.

Figura 4. Ejes fundamentales de los Destinos Turísticos Inteligentes



Fuente: (López de Ávila et al., 2015)

Se sugiere que la gestión de destinos turísticos inteligentes involucre 3 pilares de desarrollo basados en la estrategia, el uso de las TIC's y la generación de soluciones inteligentes, considerando una retroalimentación constante. De ahí que todo DTI está generalmente vinculado a un Sistema de Inteligencia Turística (SIT), capaz de integrar, analizar y visualizar la información precisa y actualizada del destino, mediante herramientas de Big Data y Business Intelligence. Esto para generar modelos multidimensionales con indicadores clave que permitan construir, en tiempo real, inteligencia turística a partir de lo que ocurre en territorio (SEGITTUR, 2022).

3.3. Gestión del Conocimiento

La Gestión del Conocimiento (Knowledge Management en inglés) consiste en el proceso de identificar, organizar, almacenar y difundir información dentro de una organización o compañía. Así, esta gestión aumenta el capital intelectual de una organización e incluye un conjunto de activos intangibles o competencias fundamentales para obtener una ventaja competitiva a favor de la empresa o sector (Castillo-Clavero et al., 2022).

3.4. Inteligencia de Negocios

La Inteligencia de Negocios, conocida también como Business Intelligence, conjuga un conjunto de herramientas, tecnologías y operaciones que permiten a una empresa u organización recopilar y luego presentar datos valiosos en tableros o reportes con información, complementando algoritmos de otras disciplinas para favorecer la comprensión en profundidad sobre el negocio o sector (Ramos, 2023).

3.5. Ciencia de Datos

La Ciencia de Datos (Data Science en inglés) se refiere a la disciplina científica que modela y define formalmente los principios de cualquier proceso relativo al manejo de datos en sus fases: recopilación (fuentes/filtrado), preprocesamiento (limpieza/transformación/almacenamiento), procesamiento (analítica/minería) y explotación (visualización/interpretación/evaluación para la toma de decisiones) (Egger & Yu, 2022; Mich, 2022). Además, se la concibe como un “cuarto paradigma” de la ciencia (empírico, teórico, computacional y ahora basado en datos) (Bell et al., 2009).

3.5.1. Datos Masivos

Los Datos Masivos o Big Data, en inglés, hace referencia al proceso completo de manejo de grandes volúmenes de datos, recolectados de una variedad de fuentes y herramientas, producidos y utilizados a gran velocidad y almacenados en multiplicidad de formatos. Algunos modelos para Big Data incluyen los parámetros de veracidad y variabilidad de los datos (Mich, 2022).

3.5.2. Minería de Datos

La Minería de Datos o Data Mining, en inglés, se define como la disciplina entre la estadística y la informática para la fase de análisis en la misión de extraer información valiosa de los datos. También, es contextualizada como descubrimiento de conocimiento de los datos (Bose, 2009).

3.5.3. Inteligencia Artificial

La Inteligencia Artificial (Artificial Intelligence en inglés) hace alusión al conjunto de técnicas para lograr que un computador resuelva problemas como lo haría un humano, es decir, permitir que las computadoras realicen tareas y actividades inteligentes (Mich, 2022).

3.5.3.1. Aprendizaje Automático

El Aprendizaje Automático (Machine Learning en inglés) se refiere a la disciplina dentro de la Inteligencia Artificial que diseña procedimientos para que las máquinas aprendan automáticamente a partir de los datos. Esta disciplina en intersección con la minería de datos da lugar, principalmente, a métodos de aprendizaje no supervisado, supervisado y por refuerzo, entre los más destacados (Egger, 2022). La figura 5 resume los métodos de aprendizaje mayormente utilizados con sus respectivos algoritmos.

Figura 5. Aprendizaje automático: métodos y algoritmos



Fuente: (Egger, 2022)

A continuación, se conceptualizan los métodos de asociación, agrupamiento, selección de características y clasificación.

3.5.3.1.1. Asociación

La asociación revela patrones en forma de reglas, mostrando los hechos que ocurren frecuentemente juntos en un conjunto de datos determinado. Se emplea para establecer posibles relaciones entre sucesos aparentemente independientes, reconociendo la ocurrencia de estos sucesos como génesis de la aparición de otros. A menudo se ejemplifica como el problema de canasta de mercado, donde se tiene un conjunto de ítems y una colección de transacciones que son subconjuntos (canastas) de estos ítems. La finalidad es encontrar relaciones entre los ítems de esas canastas para descubrir reglas de asociación que cumplan unas especificaciones mínimas dadas por el usuario, expresadas en forma de soporte y confianza (Molina-López & García-Herrero, 2006; Timarán-Pereira et al., 2016).

3.5.3.1.2. Agrupamiento o Segmentación

La segmentación (clustering en inglés) o clasificación no supervisada es el proceso de agrupar objetos en clases de objetos similares. Básicamente, se agrupa un conjunto de datos (sin un atributo de clase predefinido) basado en el principio de maximizar la similitud intracase y minimizar la similitud interclase, construyendo particiones significativas (Timarán-Pereira et al., 2016).

3.5.3.1.3. Selección de características o Ranking de atributos

La selección de atributos permite automatizar la búsqueda de subconjuntos de atributos más apropiados para explicar un atributo objetivo, en un sentido de clasificación supervisada. Así, establece qué subconjuntos de atributos son los que mejor pueden clasificar la clase de la instancia. La selección supervisada de atributos tiene dos componentes: el método de evaluación (función que determina la calidad del conjunto de atributos para discriminar la clase) y el método de búsqueda (forma de realizar la búsqueda de conjuntos, considerando estrategias para la búsqueda eficiente en cuanto crece el número de atributos) (Molina-López & García-Herrero, 2006).

3.5.3.1.4. Clasificación

La clasificación de datos obtiene resultados partiendo de un proceso de aprendizaje supervisado. Dicho proceso permite encontrar propiedades comunes entre un conjunto de objetos de una base de datos y catalogarlos en diferentes clases, de acuerdo con el modelo de clasificación (Timarán-Pereira et al., 2016).

3.5.4. Sistema de Información Geográfica

Un SIG (Sistema de Información Geográfica) es un sistema integrador de hardware, software y procedimientos elaborados para facilitar la obtención, gestión, manipulación, análisis,

modelado, representación y salida de datos espacialmente referenciados, para resolver problemas complejos de planificación y gestión (Cowen, 1990).

3.5.4.1. Datos vectoriales

Los datos vectoriales proporcionan una forma de representar entidades del mundo real en el entorno SIG. Estas entidades pueden ser carreteras, límites de propiedades, sitios de subestaciones eléctricas, etc. Una entidad al representarse en un entorno SIG se convierte en un objeto espacial vectorial que puede visualizarse como punto, polilínea o polígono (QGIS Project, 2023).

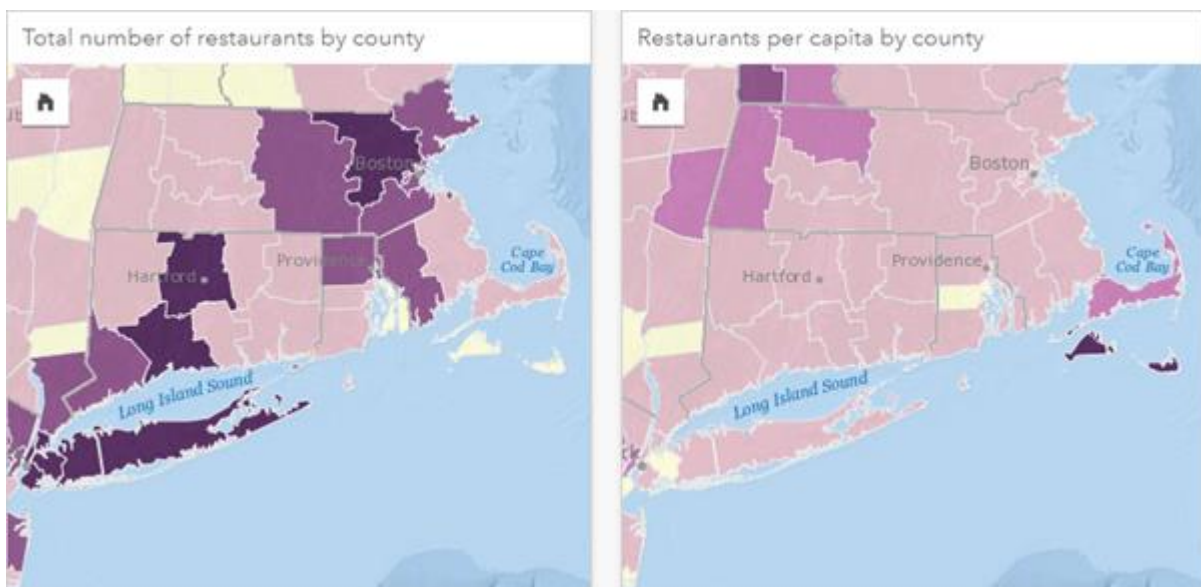
3.5.4.2. Atributos de datos vectoriales

Los objetos espaciales vectoriales tienen atributos, que consisten en texto o información numérica que describen las entidades del mundo real. Los atributos se almacenan en una tabla a manera de una hoja de cálculo, donde cada columna de la tabla se llama campo y cada fila de la tabla es un registro (QGIS Project, 2023).

3.5.4.3. Mapa de coropletas

Los mapas de coropletas son un tipo de mapa temático (ver figura 6) que utiliza una forma de cartografiado cuantitativo inteligente para la representación de fenómenos discretos asociados a unidades geográficas (municipios, provincias, comunidades autónomas, países...), a las que se aplican símbolos superficiales de acuerdo con su valor. Esta técnica sombrea las unidades de distintos colores de la misma gama cromática para visualizar de forma sencilla cómo varía una variable en un área geográfica (ArcGIS Insights, 2023; Slocum et al., 2022).

Figura 6. Mapa de coropletas



Fuente: (ArcGIS Insights, 2023)

CAPÍTULO IV. MARCO TEÓRICO

4.1. Ciencia de Datos para generación de Inteligencia Turística

La ciencia de datos se enfoca esencialmente en la extracción de conocimiento a partir del análisis (automatizado) de datos estructurados, semiestructurados y no estructurados, utilizando principios, procesos, técnicas, métodos, algoritmos y sistemas para la comprensión en profundidad de fenómenos estudiados (Provost & Fawcett, 2013; Song & Zhu, 2016). El valor central de la ciencia de datos es justamente transformar datos en información y conocimiento procesable, gracias a la combinación de varias disciplinas junto con la experiencia sustantiva en algún campo específico para generar inteligencia (Baldassarre, 2016; Egger & Yu, 2022). Esta inteligencia combinada con su implicación en decisiones, resulta en iniciativas importantes y planes estratégicos en el contexto turístico, tanto para destinos como sus organizaciones de gestión de destinos o DMOs (Destination Management Organizations por sus siglas en inglés), contribuyendo así a todos los actores clave del ecosistema (Femenia-Serra et al., 2022).

De esta forma, la ciencia de datos por su carácter interdisciplinario requiere un enfoque experto de estadística, matemáticas, informática, visualización de datos, comunicación y habilidades de presentación, así como la experiencia en turismo para cualquier actividad sólida y exitosa de explotación y gestión de datos (Chauhan & Sood, 2021). Esto se traduce en que los datos para usarse como entrada de algoritmos deben preprocesarse con el criterio de expertos turísticos y luego para la salida de estos, también deberán contar con la interpretación desde la experiencia en el sector y así respaldar procesos basados en el conocimiento (Sarker, 2021).

En este contexto, las áreas en donde se requiere la mejora de cantidad y calidad de la información disponible para la toma de decisiones son la gestión del conocimiento y la inteligencia de negocios (Castillo-Clavero et al., 2022). Si bien además de mejorar el proceso de toma de decisiones, aumentar la capacidad para resolver problemas y aumentar la capacidad de la innovación, la gestión del conocimiento en turismo, sin duda, puede contribuir en este sector, a través del reconocimiento facial en aeropuertos, la organización de viajes inspirados en Netflix o el patrocinio a quienes escriben sobre viajes, la reservación de hoteles por asistente de voz en Google, el check-in del huésped al hotel por comunicación de campo cercano NFC (Near Field Communications en inglés) o por teléfono inteligente, y demás opciones (Liu et al., 2022; Ramos, 2023).

Por su parte, la inteligencia de negocios también tiene una historia extensa en el ámbito turístico. De los primeros ejemplos se pueden identificar al sistema DINAMO introducido por American Airlines en 1988 o TourMIS en 1998 (Höpken & Fuchs, 2022). Esta área

complementa a la gestión del conocimiento en cuanto permite dar respuestas a varias preguntas, tales como: cuáles son los destinos más demandados, por cuánto tiempo se realizan las reservas, qué hoteles/restaurantes/agencias generan más ingresos y qué productos/eventos/platos/museos dejan más ganancias. Adicionalmente, la inteligencia de negocio permite comparar periodos homologados y atender a las variaciones, lograr conocimiento sobre el cliente, identificar patrones de consumo y pronosticar la demanda turística, entre otras tareas (Choi et al., 2020; Höpken et al., 2015; X. Li et al., 2021; Mariani et al., 2018; Valeri, 2020).

Aunque la eficacia administrativa y el aprendizaje organizacional son clave para la toma de decisiones, estas pueden ser significativamente mejoradas aplicando la gestión del conocimiento y la inteligencia de negocios, por medio de información altamente confiable, actualizada y estratégicamente relevante, como motivos de viaje, expectativas del servicio, necesidades de información, uso de canales y tasas de conversión relacionadas, tendencias de ocupación, calidad de la experiencia del servicio y el valor agregado por segmento de huéspedes (Pyo et al., 2002). Todo esto como papel fundamental en la generación de inteligencia turística a nivel de destinos al mejorar el intercambio de conocimiento a gran escala dentro y entre empresas y organizaciones involucradas en el sector (Höpken et al., 2014).

Sin duda, el uso generalizado de las TICs (Tecnologías de la Información y la Comunicación), especialmente la aceptación de internet, dispositivos móviles y redes sociales, condujo a un aumento vertiginoso de los datos disponibles sobre todo el mercado turístico, pero también ha favorecido su intercambio dentro del ecosistema. De hecho, el desarrollo de hardware más potente y métodos sofisticados para recolectar, almacenar y analizar dichos datos han convertido a la inteligencia de negocios en una de las áreas de mayor crecimiento tecnológico en la última década (Höpken & Fuchs, 2022). Mientras que la gestión del conocimiento para destinos turísticos presenta el desafío mayor de generar conocimiento individual a partir de clientes, productos, procesos, competidores y socios comerciales, y que dicho conocimiento esté disponible, sea significativo y tenga un enfoque integral para toda la cadena de valor (Back et al., 2007).

Bajo la premisa del dato como materia prima para generar información que influya en la toma de decisiones y brinde soporte en la gestión del conocimiento y la inteligencia de negocios, la ciencia de datos a la fecha registra gran cantidad de aplicaciones para el turismo desde el Big Data y la Inteligencia Artificial, aunque también se destacan otros aportes interesantes gracias a los Sistemas de Información Geográfica, Análisis de Redes y Series de Tiempo, entre los más populares (Mich, 2022).

Puntualmente, desde el enfoque de la Inteligencia Artificial y su desarrollo en industrias como el turismo, se han concretado beneficios en función de la demanda de consumidores, las experiencias y percepciones de turistas, la gestión de destinos y la predicción de los comportamientos de los turistas, complementados con una gran cantidad de datos creando valor para el sector en eficiencia, productividad, competitividad, rentabilidad y sostenibilidad (Doborjeh et al., 2022; Samara et al., 2020). Otros estudios puntualizan que el empleo e implementación de tecnologías disruptivas para explotar el potencial de la Inteligencia Artificial yace en áreas como la automatización del servicio al turista, seguridad, personalización y reconocimiento de patrones, gestión de tiempos de espera, predicción de la demanda y optimización de procesos (Más-Ferrando et al., 2020).

En la actualidad, existen múltiples aplicaciones introducidas gracias al soporte de la era digital y la automatización de los avances tecnológicos en la línea de la Inteligencia Artificial y otras disciplinas vinculantes como el Big Data. Por citar algunas, el Análisis de Sentimientos de los contenidos generados por turistas en cualquiera de las etapas del ciclo de viaje; la incorporación de asistentes virtuales y chatbots en experiencias turísticas; la inclusión de tecnología biométrica para el reconocimiento facial Amazon Rekognition o espejo cognitivo de IBM; la capacidad de conexión mediante RFID (Radio Frequency Identification en inglés), sensores y redes móviles 4G y 5G incluyendo códigos QR, wearables (dispositivos con tecnología vestible), llaves inteligentes con tecnología NFC, domótica en las habitaciones o despliegues de beacons (identificadores) en destinos turísticos; la integración de agencias de viajes con tecnología blockchain (cadena de bloques) como Windin Tree o Sandblock; los sistemas de recompensas y fidelización en forma de tokens; la tokenización de souvenirs; la presencia de robots colaborativos (cobots) en kioscos de check-in automáticos o robots autónomos en recepción a huéspedes como Pepper en la cadena Hilton; la creación de soluciones dedicadas al análisis de grandes volúmenes de datos provenientes de múltiples fuentes como Mabrian Technologies para gestores y planificadores de destinos; la inclusión de gestión de relación con clientes (Customer Relationship Management en inglés) que con datos históricos optimice la estancia de los huéspedes como en la cadena InterContinental Hotels y la integración de Realidad Virtual y Aumentada a experiencias turísticas inmersivas, accesibles e interactivas como la posibilidad de recrear la realidad pasada del teatro romano de Mérida o de visitar el Museo del Prado desde casa (Moreno-Izquierdo et al., 2022).

Respecto al Big Data como disciplina vinculada a la Inteligencia Artificial, también se evidencia un incremento en los aportes hacia entornos de hospitalidad y turismo para gestión de destinos, investigación e innovación, análisis de mercado y mercado laboral reforzando los

pasos clave de la pirámide de descubrimiento de conocimiento en aras de generar valor agregado en la toma de decisiones (Iorio et al., 2020). Sin embargo, todavía hace falta ahondar en las capacidades analíticas utilizando efectivamente cantidades masivas de registros con una mezcla de varias técnicas, evocando una mayor integración de la gestión y la ciencia de datos para que las colaboraciones interdisciplinarias sean cada vez más frecuentes (Mariani et al., 2018; Mariani & Baggio, 2022).

En este sentido, durante los últimos 50 años, desde la Inteligencia Artificial se destacan y agrupan cuatro tipos de investigación en: temas generadores (es decir, Redes Neuronales Artificiales, Aprendizaje Automático y Minería de Datos); temas básicos y transversales (es decir, Minería de Texto, Análisis de Sentimiento, Internet de las Cosas, Big Data y COVID-19); temas emergentes (experiencia con robots de servicio) y temas especializados/periféricos (pronóstico de modelos turísticos, Realidad Aumentada, Realidad Virtual y biometría) (Knani et al., 2022), mismas que se aterrizan en algunas de las tecnologías previamente mencionadas. A partir de 2015, la evolución temática sobre las tendencias de la Inteligencia Artificial en el sector ha migrado desde un interés por identificar lugares aptos para el turismo, luego la intención de pago por destinos sostenibles, después el uso de robots en atención a turistas debido el distanciamiento social por la pandemia del COVID-19 hasta llegar, finalmente, al uso del Aprendizaje Automático para la predicción y pronóstico, por ejemplo, de la demanda turística. Si bien se prevé un interés sostenido hacia la experiencia de los turistas y personalización de servicios a través de Realidad Aumentada, chatbots, Big Data e Internet de las Cosas, la futura agenda investigativa mantendrá dentro de los temas de mayor relevancia a los robots en la industria turística y la aplicación de técnicas de Aprendizaje Automático para estudios de predicción de comportamiento de los consumidores (Rodríguez et al., 2023).

Es por esto, que el presente trabajo centra su aporte en el Aprendizaje Automático desde un enfoque descriptivo y predictivo para lo cual se revisan algunos de los estudios más recientes de este subcampo de la Inteligencia Artificial en el turismo, especialmente los relacionados a la demanda.

4.2. Aprendizaje Automático en Turismo

El Aprendizaje Automático representa uno de los desarrollos tecnológicos más significativos y de mayor alcance en la actualidad. Aunque sus inicios se remontan a la década de 1950, su potencial se ha puesto de manifiesto recientemente, dado que el volumen de datos disponibles ha alcanzado proporciones enormes. Sin duda, de los tres tipos de aprendizaje conocidos, los aprendizajes no supervisado y supervisado son de gran importancia para casos turísticos en el

contexto de métodos de agrupamiento, reducción de dimensionalidad, clasificación y regresión. Los principales aportes de los dos tipos de aprendizaje referidos incluyen sistemas de recomendación, marketing dirigido, segmentación de turistas, visualización de datos masivos, comprensión significativa, descubrimiento de estructura, obtención de características, clasificación de opiniones, retención de clientes, clasificación de textos e imágenes, predicción del comportamiento del turista, pronóstico de popularidad publicitaria y previsión de la demanda turística (Egger, 2022; Jamal et al., 2018).

Los métodos de agrupamiento han jugado un rol central en el turismo, especialmente para tipificar turistas y su comportamiento, pero también para segmentar fotos, opiniones, destinos y sus características, etc. en grupos homogéneos (Egger, 2022). Por ejemplo, clústeres jerárquicos han sido usados para agrupar a visitantes según su riesgo percibido de COVID-19 en Alemania, Austria y Suiza (Neuburger & Egger, 2021), crear una tipología de turistas de naturaleza (Derek et al., 2019), desarrollar una clasificación sistemática de las regiones de la Unión Europea en función de la ubicación predominante de los hoteles (Batista e Silva et al., 2021), analizar las percepciones y actitudes de residentes sobre el desarrollo turístico en Costa Esmeralda, Italia (Del Chiappa et al., 2018). Otros estudios han utilizado el algoritmo *K-medias* (*K-means* en inglés) para segmentar el mercado de viajes de Yakarta creando una tipología de turistas según sus estilos de vida, agrupar el valor del voluntariado de los turistas en términos de su satisfacción con la vida y sus actitudes hacia el trabajo como voluntario y optimizar el agrupamiento del gasto turístico en la Comunidad Valenciana, España (Chua et al., 2021; Rabasa et al., 2017; Srihadi et al., 2016).

Otra de las tareas frecuentes es la reducción de dimensiones en datos con muchas variables, así se pueden identificar las características que mejor representan los datos, teniendo un conjunto de datos más pequeño y eficiente. La forma más conocida para esto es el Análisis de Componentes Principales, aunque también existen otros algoritmos (Egger, 2022). Además, se puede gestionar la selección de atributos desde el enfoque de su aporte a una variable objetivo como paso previo a otros análisis posteriores (Duboue, 2022). Como muestra, se realizó una clasificación de sentimientos con un enfoque de modelo temático usando redes neuronales, así se crearon representaciones vectoriales de palabras multidimensionales en el contexto de reseñas de hoteles chinos, pero para visualizarlas, se tuvo que reducir los datos a una proyección 2D con el algoritmo t-SNE (T-distributed Stochastic Neighbor Embedding en inglés) (Q. Li et al., 2018). Otro ejemplo es un estudio donde las características de alta dimensionalidad de fotos georreferenciadas en sitios del patrimonio arqueológico en Cuzco, Perú, fueron reducidas

para mostrarse con fines interpretativos en un mapa con el algoritmo t-SNE (Payntar et al., 2021).

La clasificación es una de las aplicaciones de Aprendizaje Automático más utilizadas, misma que busca dividir datos similares en diferentes clases. Los algoritmos más frecuentes incluyen árboles de decisión, métodos basados en reglas, métodos probabilísticos, métodos de máquinas de soporte vectorial (Support Vector Machine en inglés), métodos basados en instancias y redes neuronales, así se pueden identificar varias reglas según el algoritmo aplicado (Egger, 2022). Un estudio reciente tuvo como objetivo operacionalizar la propuesta de valor de anfitriones en Airbnb. De más de 250 variables, primero se identificaron aquellas que más contribuyen a ser clasificado como un “super anfitrión”. Luego, usaron un clasificador de soporte vectorial para la clasificación binaria y modelos basados en clasificadores bayesiano ingenuo y regresión logística como líneas base para la comparación de los análisis (Ramos-Henríquez et al., 2021). Otro estudio presentó un enfoque con respecto a la correcta selección de fotos para la comunicación del destino basado en su clasificación en categorías de afecto. Para este fin, se entrenó un modelo bayesiano ingenuo con numerosos pares de “contenido-emoción” usando imágenes de Flickr para predecir clasificación de emociones para nuevas fotos en función de su contenido (Deng & Li, 2018). En la Comunidad Valenciana (España), se analizan diferentes conceptos de gasto durante el viaje, así como el gasto total diario, para esto se utiliza un algoritmo de extracción de patrones considerando las franjas extremas de gasto, luego se seleccionaron los atributos más importantes para generar un modelo de clasificación con el algoritmo CART (Classification And Regression Trees en inglés) que sea más eficiente a la hora de predecir y mejorar los servicios ofrecidos a los turistas (Rabasa et al., 2018). Para el sector hotelero, otros investigadores intentaron clasificar las reseñas en línea usando un método basado en texto. Esto considerando características para clasificar con éxito opiniones engañosas o no engañosas (Martinez-Torres & Toral, 2019).

La regresión se conoce comúnmente como el modelamiento de la relación entre variables explicativas y una variable objetivo, describiendo así la relación promedio entre atributos numéricos. Dentro del Aprendizaje Automático, existen muchos métodos de regresión que predicen valores numéricos continuos, en contraste a la clasificación con características discretas dispuestas en clases (Egger, 2022). Los modelos de predicción basados en enfoques de regresión se utilizan en una amplia variedad de áreas. Por ejemplo, en términos de sostenibilidad, los modelos para predecir el impacto del turismo en la naturaleza son muy relevantes. En este sentido, un estudio investigó la relación entre el desarrollo del turismo local y su impacto en los recursos ambientales para desarrollar un modelo de predicción cuantitativa

de máquinas de soporte vectorial para el desarrollo del volumen de negocios en Chengdu, China (Han, 2020). Algunos ejemplos adicionales de turismo en los que se han utilizado diferentes enfoques de regresión provienen, por ejemplo, de un enfoque donde se exploró la connotación de recomendaciones explícitas basadas en reseñas de Yelp (red social de reseñas para negocios) aplicando regresión logística (Guerreiro & Rita, 2020), así como otra aportación donde se reforzó la eficacia del uso de máquinas de soporte vectorial para la clasificación multiclase. Al tomar las propiedades enumeradas en Airbnb como contexto del caso, la precisión de máquinas de soporte vectorial demostró ser mayor que el modelo logístico en general (Martin-Fuentes et al., 2018). Otro estudio incluye la predicción de la demanda turística en la región de Liuzhou en China a través de modelos híbridos combinando redes neuronales artificiales y regresión de soporte vectorial (W. Li, 2022). Un enfoque desde el aprendizaje profundo establece un modelo de pronóstico colaborativo para la demanda turística de atractivos turísticos con dependencia espacial en Beijing, China (Bi et al., 2023).

4.3. Aproximaciones a la Demanda Turística Interna: Caso Ecuador

Si bien el comportamiento y la caracterización de la demanda turística interna es un tema de interés para dinamizar el sector turístico en cualquier país, en Ecuador debido a falta de presupuesto y recorte de personal especialmente en los últimos años, el Ministerio de Turismo no cuenta con reportes anuales para monitorear esta demanda como se lo hizo hasta 2012 en conjunto con otras instituciones estatales como el INEC (Instituto Nacional de Estadísticas y Censos) o privadas como Consultora MINDTEK por medio de sondeo telefónico en 2015 (MINDTEK, 2015a, 2015b; Ministerio de Turismo, 2011, 2012, 2013).

Todos estos estudios reflejan un alcance únicamente descriptivo clásico. En el caso de los monitoreos realizados en 2010, 2011 y 2012 a través de la técnica de encuesta con el INEC, se incluyen variables asociadas a características socioeconómicas (sexo, edad, estado civil, nivel de instrucción, situación ocupacional e ingreso bruto percibido) y a datos turísticos (mes de viaje, período de viaje, cantidad de viajeros, motivación de viaje, provincia de destino, cantón visitado, tiempo desde el último viaje, tipo principal de transporte, tipo de alojamiento, pernoctaciones, tiempo de anticipación de planificación de viaje, gasto total de viaje, medio para planificar el viaje, actividad turística realizada, medios de influencia en elección de destino, servicios escasos en el viaje, servicios adquiridos por operadora, costo promedio de paquete turístico, recomendaciones de destino de viaje, estructura de gasto turístico en viaje, intención de próximo viaje y recordación/agrado de marca país) (Ministerio de Turismo, 2011, 2012, 2013). Por su parte, en el sondeo telefónico ejecutado por MINDTEK se incluyeron

atributos como propensión bruta a viajar, frecuencia de viajes, destino de viaje, medio de transporte usado, personas por viaje, cantidad de viajeros, días de viaje, gasto por viaje y gasto medio diario por persona. Además, se efectuó un análisis complementario aplicando análisis de redes sociales para evidenciar el flujo de viajes por turismo interno (MINDTEK, 2015a, 2015b). El único estudio en un contexto Big Data y el último que el Ministerio de Turismo reporta a la fecha, versa sobre patrones de movilidad turística interna con OTECEL a través de una consultoría realizada en 2018 para analizar los traslados internos, contar con un panorama general de turismo interno e identificar destinos claves y puntos de afluencia críticos. Dicho estudio trabajó con el componente analítico Smart Steps considerando la agregación de datos móviles de la red de Telefónica (2G, 3G y LTE) para comprender mejor los comportamientos de las multitudes y no de individuos (LUCA Data-Driven Decisions, 2018).

Por su parte, dentro del contexto de la reactivación turística, es decir de 2020 en adelante, la academia ha aportado con estudios puntuales en alguna tipología de turismo, algún segmento específico, o bien se enfocan en los turistas que visitan algún destino específico, sea provincia o ciudad (Almache, 2021; Bautista, 2021; Herrera Díaz et al., 2020; Mera et al., 2020; Molina Velásquez et al., 2020; Sánchez Ruiz & Paladines Sarango, 2021; Toledo et al., 2020). Únicamente, un estudio aborda el comportamiento de la demanda turística interna en la reactivación post pandemia a nivel nacional; sin embargo, evidencia alcance exploratorio y descriptivo (Molina Velásquez et al., 2022).

Este aporte en solitario desde la comunidad científica incluyó valoración sociodemográfica del visitante (edad, sexo y empleo), además su comportamiento respecto a actividades de excursionismo (frecuencia de excursiones) y de turismo (frecuencia de viaje, provincia de destino, provincia de residencia, intención de viaje próximo, tipo de viaje próximo, composición de grupo de viaje, gasto por persona y percepción de seguridad en viajes) (Molina Velásquez et al., 2022).

CAPÍTULO V. MARCO METODOLÓGICO

El Ministerio de Turismo del Ecuador, de acuerdo al artículo 16 de la Ley de Turismo, tiene entre sus competencias privativas, la información estadística. Por lo que, producto de iniciativas propias institucionales a través de la Dirección de Inteligencia de Mercados estableció la necesidad de contratar una consultoría con la finalidad de conocer la dinámica actualizada de variables relacionadas con el ámbito de desarrollo turístico por parte de viajeros nacionales que visitan los diferentes destinos dentro del Ecuador (Velasco, 2022a).

Cabe indicar que, en este contexto, el Ministerio supervisó y validó todo el proceso mediante revisión y aprobación de los productos presentados por parte del consultor. Por lo que, para la realización del presente trabajo, el Ministerio de Turismo proporcionó los datos y productos generados en la mencionada consultoría sobre demanda turística interna, previa solicitud, respetando los criterios de confidencialidad y demás políticas aplicables.

En este capítulo, la información desglosada en los siguientes acápite hasta la recopilación de datos corresponde a lo gestionado por el Ministerio de Turismo, como ente rector del sector, a través de la consultoría; sin embargo, lo referente al preprocesamiento y procesamiento de los datos recae en la autoría del presente Trabajo Fin de Máster.

5.1. Diseño del marco muestral

Para la selección de la muestra, se consideraron dos criterios base: el primero donde se establece que las principales provincias generadoras de valor agregado productivo son: Guayas, Pichincha, Azuay, Manabí y Tungurahua; y, el segundo que considera a la provincia de Loja como uno de los destinos con mayor interés turístico para los próximos años. Así, se estableció que las ciudades de estudio fueran: Quito, Guayaquil, Cuenca, Ambato, Manta y Loja.

En cuanto a la estructura de cálculo para la muestra, se determinó como punto de partida a la población, independiente del nivel socioeconómico al que pertenecen, pues se establece que el ámbito turístico abarca casi en su totalidad a los estratos. Aunque la totalidad de habitantes en las ciudades de estudio ascendió a más de siete millones según la proyección poblacional 2020-2025 del Instituto Nacional de Estadísticas y Censos (INEC), sobre esta cifra se estableció la discriminación correspondiente a la Población Económicamente Activa (PEA) que para el año 2021 correspondió al 65.3%. De esta forma, se determinó que la población efectiva para el cálculo muestral fue de 4 714 968 habitantes resultantes, tal como lo recoge la tabla 1.

Aplicando la expresión para el cálculo, donde n constituye el tamaño de la muestra, N es el tamaño de la población, z_{α} es el percentil $(1 - \alpha) * 100$ de la distribución Normal estándar

correspondiente al nivel de confianza, p representa la probabilidad de ocurrencia de un evento, q representa la probabilidad de no ocurrencia de un evento y d es el error máximo admisible en términos de proporción, se obtuvo lo siguiente:

$$n = \frac{N * z_{\alpha}^2 * p * q}{d^2 * (N - 1) + z_{\alpha}^2 * p * q} = \frac{4\,714\,968 * 1.96^2 * 0.5 * 0.5}{0.05^2 * (4\,714\,968 - 1) + 1.96^2 * 0.5 * 0.5} = 384.13$$

Con una población de 4 714 968 habitantes, un nivel de confianza del 95% y un margen de error aceptable de 5%, se generó una línea base. Siendo $n = 384$ elementos aleatorios, se consideró una distribución lineal para las ciudades con menor representación que son: Cuenca, Manta, Ambato y Loja, mientras que, para Quito y Guayaquil, se aplicó un factor de corrección de 0.28 para aumentar representatividad, por lo tanto, la distribución muestral preliminar se resume en la tabla 1.

Tabla 1. Distribución muestral preliminar

Ciudad	Población Total	Población PEA	Elementos
Quito	2 872 351	1 867 028	540
Guayaquil	2 772 896	1 802 382	540
Cuenca	659 320	428 558	384
Ambato	395 893	257 330	384
Manta	269 113	174 923	384
Loja	284 224	184 746	384
Total	7 253 797	4 714 968	2 616

Fuente: (Velasco, 2022a)

Además, se determinó que, para incrementar aún más la representatividad nacional, se debieron establecer 3 grupos de elementos adicionales: Resto de la Amazonía, Resto de la Costa y Resto de la Sierra. Estos tres grupos fueron totalmente aleatorios debido a que la cantidad de habitantes por ciudad principal no alcanza el 2% del total de la población; sin embargo, se consideró la base de $n = 384$ como referente subdividido entre los grupos. Por lo tanto, la distribución final de la muestra se condensa en la tabla 2.

Tabla 2. Distribución muestral final

Ciudad	Elementos
Quito	540
Guayaquil	540
Cuenca	384
Ambato	384
Manta	384
Loja	384
Resto Costa	128
Resto Sierra	128
Resto Amazonía	128
Total	3 000

Fuente: (Velasco, 2022a)

5.2. Diseño del cuestionario

Con la finalidad de establecer un esquema que abarque las principales variables de interés, se estableció un cuestionario preliminar de 27 preguntas, producto de un análisis entre el equipo de contraparte del Ministerio de Turismo del Ecuador y el consultor, considerando los indicadores recomendados por la OMT. Una vez que el cuestionario fue aprobado en conjunto, se procedió a su parametrización y carga en la plataforma especializada escogida para el pilotaje. A raíz del levantamiento piloto, los hallazgos fueron puntualizados y revisados en conjunto para establecer el cuestionario definitivo, mismo que consta de 25 preguntas (Anexo 3).

5.3. Recopilación de datos

Una vez que el cuestionario definitivo fue establecido, se configuró mediante la plataforma de encuestas online Zoho Forms y se diseñó la estrategia para alcanzar el público objetivo mediante el uso de medios digitales durante el mes de octubre de 2022. Para esto, se efectuaron campañas a través de redes sociales (Facebook e Instagram) así como repliques mediante WhatsApp, mailing, entre otros medios, tal como se evidencia en la figura 7.

Figura 7. Material promocional de las campañas en redes sociales



Fuente: (Velasco, 2022b)

Mediante esta gestión, la muestra fue seleccionada de acuerdo a la distribución indicada en la tabla 2, producto de la depuración de respuestas recopiladas por datos incompletos, validaciones lógicas y deserciones o abandonos. Con esto, se condensó la base de datos, acoplada a la plataforma de PowerBI para la visualización de variables (de forma individual o enlazada a otras variables comunes) y construcción de perfilados (Anexo 4), según lo requerido por el Ministerio de Turismo del Ecuador. Cabe indicar que la base de datos consensada fue

estructurada principalmente en bloques, exportable a formato compatible con Microsoft Excel y Statistical Package for Social Sciences (SPSS).

5.4. Preprocesamiento de datos

Partiendo de la base de datos en la plataforma de PowerBI, se exportó la matriz correspondiente a Microsoft Excel. Con la matriz completa se ejecutaron tareas de validación de atributos, eliminación de tildes para evitar errores de codificación Unicode, recodificación para normalizar nomenclatura de todos los atributos, gestionando un total de 38 variables a partir del cuestionario aplicado. Estas variables incluyen atributos vinculados a los visitantes (excursionistas y turistas), al viaje y al destino, de acuerdo a lo que se resume en la tabla 3.

Tabla 3. Descripción de atributos/variables de la base de datos

Nº	Atributo/Variable	Valor ejemplo
1	Region resid	QUITO
2	Prov resid	PICHINCHA
3	Ciudad resid	QUITO
4	Edad	24
5	Estado civil	Soltera/o
6	Genero	Femenino
7	Nivel estudios	Superior
8	Ingreso prom mes hogar	B de usd 401 a 800
9	Miembro decide viaje	Fue una decision en conjunto
10	Razon viaje	Vacaciones recreo ocio (turismo)
11	Region viaje	Costa
12	Destino viaje	MANABI
13	Ciudad viaje	MANTA
14	Antelac planif	Una semana
15	Dias pernocta	3
16	Tipo Alojamiento	Vivienda propia
17	Num acomp	Siete
18	Quienes acomp	Padres y/o suegros
19	Gtos antes	50
20	Gtos durante	200
21	Gtos despues	100
22	Medio transp	Autobus / flotas
23	Medios comunic	Redes sociales
24	RRSS	Facebook
25	Plataf viajes	-
26	Escala reservas	4
27	Escala transporte	4
28	Escala aliment	4
29	Escala alojamiento	4
30	Escala entretenim	4
31	Viaje en prx 3 meses	No

32	Porc Gto alojamiento	30
33	Porc Gto transp aéreo	0
34	Porc Gto transp terr	10
35	Porc Gto combust	10
36	Porc Gto alim	20
37	Porc Gto entretenim	20
38	Porc Gto varios	10

Fuente: Elaboración propia

Con esta base preliminar, se generaron algunas variables partiendo de otras, previo a algún criterio de recategorización; por ejemplo, Region resid tiene la distribución original de la selección de la muestra y así los valores Quito, Cuenca, Ambato, Loja, Resto Sierra, se recategorizaron como Sierra, mientras que Guayaquil, Manta y Resto Costa, se recategorizaron como Costa, de la misma forma Resto Amazonía como Amazonía simplemente.

La variable Edad se discretizó de acuerdo al cambio generacional en Ecuador en las categorías de Centennials (menos de 26 años), Millennials (26 a 41 años), Generación X (42 a 57 años), Baby boomers (58 a 77 años) y Tradicionalistas (más de 77 años). La variable Num acomp dio lugar a otra variable discretizada en función de las categorías solo (1), pareja (2) y grupo (3 o más visitantes).

Las variables Gtos antes, Gtos durante y Gtos después, generaron mediante su sumatoria la variable Gto Total, de igual manera cada una de estas 4 variables fueron discretizadas en función de la distribución de los valores por cuartiles, considerando al valor \$0 como Gasto Nulo, a los valores del primer cuartil como Gasto Bajo, a los valores del segundo y tercer cuartil como Gasto Medio y a los valores del último cuartil como Gasto Alto.

También, se recategorizó la variable Medios comunic en cuatro categorías Internet, Redes sociales, Medios clásicos (agrupando aquí Medios impresos, Publicidad vial y señalética, Radio y Televisión) y Recomendaciones de familiares o amigos. Las variables vinculadas a escalas de dificultad en gestión de reservas, transporte, alimentación, alojamiento y entretenimiento para el viaje, fueron discretizadas con las etiquetas de Muy Fácil (1), Fácil (2), Difícil (3) y Muy Difícil (4).

Para culminar con las tareas de preprocesamiento, se rellenaron ciertas celdas con valores de 0, No utiliza o No aplica de acuerdo a la variable que corresponda y se generó una base final en formato CSV para su lectura adecuada en la fase del procesamiento. Adicionalmente, a partir de la base final, se generaron archivos adicionales (submuestras) en formato CSV según filtros requeridos para ciertos análisis en el posterior procesamiento.

5.5. Procesamiento de datos

Se definieron los análisis desde los atributos vinculados al visitante, viaje y destino, tal como se precisa en las tablas 4, 5 y 6. Posteriormente, se aplicaron métodos de asociación, agrupamiento, selección de atributos y clasificación (árboles de decisión) considerando los algoritmos detallados en las referidas tablas con el software WEKA (Waikato Environment for Knowledge Analysis) versión 3.8.6. Finalmente, para la visualización de patrones geográficos se utilizaron los mapas de coropletas a través del software QGIS versión 3.22.16, considerando el cómputo de porcentajes de visitas, promedio de pernотaciones y promedio del gasto total de viaje por cada provincia como destino a través de Microsoft Excel.

Tabla 4. Análisis del visitante

Propósito del análisis	Método y algoritmo
Extracción de patrones del visitante por edad e ingreso (a nivel nacional)	Asociación (<i>A priori</i>)
Extracción de patrones del visitante por nivel de estudios y gasto total (a nivel nacional)	Asociación (<i>A priori</i>)
Extracción de patrones socioeconómicos del visitante (a nivel nacional)	Asociación (<i>A priori</i>)
Extracción de patrones socioeconómicos del visitante (con región Costa versus región Sierra como destino)	Asociación (<i>A priori</i>)

Fuente: Elaboración propia

Tabla 5. Análisis del viaje

Propósito del análisis	Método y algoritmo
Segmentación de viajes por conceptos de gasto (alojamiento, transporte, combustible, alimentos y bebidas, entretenimiento, varios)	Agrupamiento (<i>Simple K Means</i>)
Segmentación de viajes por gasto total y tamaño de grupos	Agrupamiento (<i>Simple K Means</i>)
Segmentación de viajes por pernотaciones y tamaño de grupos	Agrupamiento (<i>Simple K Means</i>)
Segmentación de viajes por pernотaciones y gasto total según edad	Agrupamiento (<i>Simple K Means</i>)
Segmentación de viajes por edad y gasto total según tipo de alojamiento	Agrupamiento (<i>Simple K Means</i>)
Segmentación de viajes por edad y gasto total según medio de transporte	Agrupamiento (<i>Simple K Means</i>)
Extracción de patrones de viaje según motivo, tamaño de grupos, antelación y gasto total	Asociación (<i>A priori</i>)
Extracción de patrones de viaje según medio de transporte, medios de comunicación, redes sociales, plataformas de viaje y edad	Asociación (<i>A priori</i>)
Extracción de patrones en la complejidad de aspectos del viaje	Asociación (<i>A priori</i>)
Extracción de patrones en los gastos del viaje	Asociación (<i>A priori</i>)

Detección de variables significativas para el umbral de gasto total del viaje	Selección de características (<i>Info Gain Attribute Eval</i>)
Predicción del umbral de gasto total del viaje	Clasificación (<i>J48</i>)
Detección de variables significativas para la intención de viaje a futuro	Selección de características (<i>Info Gain Attribute Eval</i>)
Predicción de la intención de viaje a futuro	Clasificación (<i>J48</i>)

Fuente: Elaboración propia

Tabla 6. Análisis del destino

Propósito del análisis	Método y algoritmo
Detección de variables significativas para la región del destino	Selección de características (<i>Info Gain Attribute Eval</i>)
Predicción de la región del destino	Clasificación (<i>J48</i>)
Predicción de la región del destino (Costa y Sierra) de visitantes vacacionales	Clasificación (<i>J48</i>)
Visualización de patrones geográficos de la frecuencia de visita	Mapa de coropletas
Visualización de patrones geográficos de pernoctaciones promedio	Mapa de coropletas
Visualización de patrones geográficos de gasto total promedio del viaje	Mapa de coropletas

Fuente: Elaboración propia

CAPÍTULO VI. ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

6.1. Análisis del visitante

En primer lugar, se extrajeron patrones del visitante por edad e ingreso, encontrando que hay 788 centennials, de los cuales 479 (61%) generan un ingreso promedio mensual entre \$401 y \$800. Esta misma franja de ingreso promedio se corresponde con el 47% de los millennials, de acuerdo a la figura 8.

Figura 8. Patrones por edad e ingreso

```
Associator output
=== Run information ===
Scheme: weka.associations.Apriori -N 10 -T 0 -C 0.25 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation: Mintur_BI_CSV_PREPRO_Wk_final-weka.filters.unsupervised.attribute.Remove-RI-4,6-8,10-50
Instances: 3000
Attributes: 2
          Edad D
          Ingreso prom mes hogar
=== Associator model (full training set) ===

Apriori
=====

Minimum support: 0.1 (300 instances)
Minimum metric <confidence>: 0.25
Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 6

Size of set of large itemsets L(2): 2

Best rules found:

1. Edad D=Centennials 788 ==> Ingreso prom mes hogar=B de usd 401 a 800 479 <conf:(0.61)> lift:(1.33) lev:(0.04) [117] conv:(1.38)
2. Edad D=Millennials 1268 ==> Ingreso prom mes hogar=B de usd 401 a 800 591 <conf:(0.47)> lift:(1.02) lev:(0) [9] conv:(1.01)
3. Ingreso prom mes hogar=B de usd 401 a 800 1375 ==> Edad D=Millennials 591 <conf:(0.43)> lift:(1.02) lev:(0) [9] conv:(1.01)
4. Ingreso prom mes hogar=B de usd 401 a 800 1375 ==> Edad D=Centennials 479 <conf:(0.35)> lift:(1.33) lev:(0.04) [117] conv:(1.13)
```

Fuente: Elaboración propia

Otros patrones obtenidos al respecto del nivel de estudios y el gasto total del viaje, se muestran en la figura 9, donde el 69% de los visitantes que incurren en un gasto total alto, poseen un nivel de estudios superior. De los viajeros que incurren en un gasto medio, el 58% tienen estudios superiores y el 41% tiene estudios de secundaria.

Figura 9. Patrones por nivel de estudios y gasto total

```
Associator output
=== Run information ===
Scheme: weka.associations.Apriori -N 10 -T 0 -C 0.25 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation: Mintur_BI_CSV_PREPRO_Wk_final-weka.filters.unsupervised.attribute.Remove-RI-7,9-27,29-50
Instances: 3000
Attributes: 2
          Nivel estudios
          Gto Total D
=== Associator model (full training set) ===

Apriori
=====

Minimum support: 0.1 (300 instances)
Minimum metric <confidence>: 0.25
Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 6

Size of set of large itemsets L(2): 3

Best rules found:

1. Gto Total D=Alto 750 ==> Nivel estudios=Superior 515 <conf:(0.69)> lift:(1.12) lev:(0.02) [53] conv:(1.22)
2. Gto Total D=Medio 1496 ==> Nivel estudios=Superior 867 <conf:(0.58)> lift:(0.94) lev:(-0.02) [-54] conv:(0.91)
3. Nivel estudios=Secundaria 1113 ==> Gto Total D=Medio 609 <conf:(0.55)> lift:(1.1) lev:(0.02) [53] conv:(1.1)
4. Nivel estudios=Superior 1847 ==> Gto Total D=Medio 867 <conf:(0.47)> lift:(0.94) lev:(-0.02) [-54] conv:(0.94)
5. Gto Total D=Medio 1496 ==> Nivel estudios=Secundaria 609 <conf:(0.41)> lift:(1.1) lev:(0.02) [53] conv:(1.06)
6. Nivel estudios=Superior 1847 ==> Gto Total D=Alto 515 <conf:(0.28)> lift:(1.12) lev:(0.02) [53] conv:(1.04)
```

Fuente: Elaboración propia

En cuanto a patrones socioeconómicos del visitante, se evidenció que el 69% de las mujeres tienen nivel de estudios superior y el 64% de ellas son solteras, mientras sólo 55% de hombres tienen nivel de estudios superior y el 52% de ellos están casados. Por lo tanto, existe un patrón sociodemográfico un tanto más frecuente de mujeres viajeras que de hombres viajeros, tal como se resume en la figura 10.

Figura 10. Patrones socioeconómicos del visitante

```

Associator output
=== Run information ===

Scheme:      weka.associations.Apriori -N 10 -T 0 -C 0.25 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation:    Mintur BI_CSV PREPRO_Wk_final-weka.filters.unsupervised.attribute.Remove-R1-4,10-50
Instances:   3000
Attributes:  5
             Edad D
             Estado civil
             Genero
             Nivel estudios
             Ingreso prom mes hogar
=== Associator model (full training set) ===

Apriori
=====

Minimum support: 0.25 (750 instances)
Minimum metric <confidence>: 0.25
Number of cycles performed: 15

Generated sets of large itemsets:

Size of set of large itemsets L(1): 10

Size of set of large itemsets L(2): 7

Best rules found:

1. Genero=Femenino 1402 ==> Nivel estudios=Superior 966 <conf:(0.69)> lift:(1.12) lev:(0.03) [102] conv:(1.23)
2. Estado civil=Casada/o 1225 ==> Genero=Masculino 803 <conf:(0.66)> lift:(1.28) lev:(0.06) [176] conv:(1.41)
3. Estado civil=Casada/o 1225 ==> Nivel estudios=Superior 790 <conf:(0.64)> lift:(1.05) lev:(0.01) [35] conv:(1.08)
4. Genero=Femenino 1402 ==> Estado civil=Soltera/o 893 <conf:(0.64)> lift:(1.24) lev:(0.06) [170] conv:(1.33)
5. Ingreso prom mes hogar=B de usd 401 a 800 1375 ==> Estado civil=Soltera/o 869 <conf:(0.63)> lift:(1.23) lev:(0.05) [159] conv:(1.31)
6. Estado civil=Soltera/o 1547 ==> Nivel estudios=Superior 907 <conf:(0.59)> lift:(0.95) lev:(-0.02) [-45] conv:(0.93)
7. Estado civil=Soltera/o 1547 ==> Genero=Femenino 893 <conf:(0.58)> lift:(1.24) lev:(0.06) [170] conv:(1.26)
8. Estado civil=Soltera/o 1547 ==> Ingreso prom mes hogar=B de usd 401 a 800 869 <conf:(0.56)> lift:(1.23) lev:(0.05) [159] conv:(1.23)
9. Genero=Masculino 1535 ==> Nivel estudios=Superior 843 <conf:(0.55)> lift:(0.89) lev:(-0.03) [-102] conv:(0.85)
10. Genero=Masculino 1535 ==> Estado civil=Casada/o 803 <conf:(0.52)> lift:(1.28) lev:(0.06) [176] conv:(1.24)

```

Fuente: Elaboración propia

Dado que la muestra aglutinó en su mayoría a visitantes que tuvieron como destino las regiones Costa y Sierra, se extrajeron patrones socioeconómicos de ambos grupos. Los patrones del viajero que visitó la región Costa son muy similares a los descritos previamente para la muestra global, lo que se puede corroborar en la figura 11.

Entre los viajeros que visitaron la región Sierra, la franja de edad aparece como una variable de relevancia en los patrones más significativos. Por otra parte, el 72% de las personas casadas son hombres, mientras el 70% de las mujeres que viajan están solteras. Esto no ocurrió sobre la muestra global y en menor medida sobre la submuestra de los viajeros que visitaron la región costera, tal como consta en la figura 12.

Figura 11. Patrones socioeconómicos del visitante con región Costa como destino

```

Associator output
=== Run information ===

Scheme:      weka.associations.Apriori -N 10 -T 0 -C 0.25 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation:    Mintur BI_CSV PREPRO_Wk_final_Costa-weka.filters.unsupervised.attribute.Remove-R1-4,10-50
Instances:   1249
Attributes:  5
             Edad D
             Estado civil
             Genero
             Nivel estudios
             Ingreso prom mes hogar
=== Associator model (full training set) ===

Apriori
=====

Minimum support: 0.25 (312 instances)
Minimum metric <confidence>: 0.25
Number of cycles performed: 15

Generated sets of large itemsets:

Size of set of large itemsets L(1): 11
Size of set of large itemsets L(2): 7

Best rules found:

1. Estado civil=Casada/o 463 ==> Nivel estudios=Superior 331 <conf:(0.71)> lift:(1.1) lev:(0.02) [29] conv:(1.21)
2. Genero=Femenino 618 ==> Nivel estudios=Superior 430 <conf:(0.7)> lift:(1.07) lev:(0.02) [27] conv:(1.14)
3. Ingreso prom mes hogar=B de usd 401 a 800 530 ==> Estado civil=Soltera/o 360 <conf:(0.68)> lift:(1.25) lev:(0.06) [71] conv:(1.41)
4. Genero=Femenino 618 ==> Estado civil=Soltera/o 401 <conf:(0.65)> lift:(1.19) lev:(0.05) [65] conv:(1.29)
5. Edad D=Millennials 508 ==> Nivel estudios=Superior 323 <conf:(0.64)> lift:(0.98) lev:(-0.01) [-8] conv:(0.95)
6. Genero=Masculino 608 ==> Nivel estudios=Superior 373 <conf:(0.61)> lift:(0.94) lev:(-0.02) [-23] conv:(0.9)
7. Estado civil=Soltera/o 679 ==> Nivel estudios=Superior 408 <conf:(0.6)> lift:(0.92) lev:(-0.03) [-34] conv:(0.87)
8. Estado civil=Soltera/o 679 ==> Genero=Femenino 401 <conf:(0.59)> lift:(1.19) lev:(0.05) [65] conv:(1.23)
9. Estado civil=Soltera/o 679 ==> Ingreso prom mes hogar=B de usd 401 a 800 360 <conf:(0.53)> lift:(1.25) lev:(0.06) [71] conv:(1.22)
10. Nivel estudios=Superior 814 ==> Genero=Femenino 430 <conf:(0.53)> lift:(1.07) lev:(0.02) [27] conv:(1.07)

```

Fuente: Elaboración propia

Figura 12. Patrones socioeconómicos del visitante con región Sierra como destino

```

Associator output
=== Run information ===

Scheme:      weka.associations.Apriori -N 10 -T 0 -C 0.25 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation:    Mintur BI_CSV PREPRO_Wk_final_Sierra-weka.filters.unsupervised.attribute.Remove-R1-4,10-50
Instances:   1286
Attributes:  5
             Edad D
             Estado civil
             Genero
             Nivel estudios
             Ingreso prom mes hogar
=== Associator model (full training set) ===

Apriori
=====

Minimum support: 0.25 (321 instances)
Minimum metric <confidence>: 0.25
Number of cycles performed: 15

Generated sets of large itemsets:

Size of set of large itemsets L(1): 9
Size of set of large itemsets L(2): 11

Best rules found:

1. Edad D=Centennials 398 ==> Estado civil=Soltera/o 334 <conf:(0.86)> lift:(1.61) lev:(0.1) [126] conv:(3.29)
2. Estado civil=Casada/o 501 ==> Genero=Masculino 361 <conf:(0.72)> lift:(1.39) lev:(0.08) [100] conv:(1.71)
3. Genero=Femenino 578 ==> Estado civil=Soltera/o 407 <conf:(0.7)> lift:(1.32) lev:(0.08) [98] conv:(1.57)
4. Nivel estudios=Secundaria 568 ==> Ingreso prom mes hogar=B de usd 401 a 800 387 <conf:(0.68)> lift:(1.28) lev:(0.07) [85] conv:(1.47)
5. Genero=Femenino 578 ==> Nivel estudios=Superior 376 <conf:(0.65)> lift:(1.22) lev:(0.05) [68] conv:(1.33)
6. Nivel estudios=Secundaria 568 ==> Genero=Masculino 369 <conf:(0.65)> lift:(1.25) lev:(0.06) [73] conv:(1.36)
7. Edad D=Millennials 560 ==> Genero=Masculino 353 <conf:(0.63)> lift:(1.21) lev:(0.05) [62] conv:(1.29)
8. Ingreso prom mes hogar=B de usd 401 a 800 682 ==> Estado civil=Soltera/o 424 <conf:(0.62)> lift:(1.16) lev:(0.05) [59] conv:(1.23)
9. Estado civil=Soltera/o 687 ==> Ingreso prom mes hogar=B de usd 401 a 800 424 <conf:(0.62)> lift:(1.16) lev:(0.05) [59] conv:(1.22)
10. Estado civil=Soltera/o 687 ==> Genero=Femenino 407 <conf:(0.59)> lift:(1.32) lev:(0.08) [98] conv:(1.35)

```

Fuente: Elaboración propia

6.2. Análisis del viaje

En primera instancia, los viajes se segmentaron de acuerdo a los porcentajes del gasto total de viaje, destinados a sectores de alojamiento, transporte aéreo, transporte terrestre, y alimentos y bebidas. La primera pareja analizada vinculó al porcentaje de gasto en alojamiento versus transporte aéreo, como se evidencia en la figura 13.

Figura 13. Clústeres por gasto en alojamiento vs transporte aéreo



Fuente: Elaboración propia

Se generaron 4 grupos: El clúster verde agrupa a los visitantes que gastan muy poco en alojamiento y mantienen un gasto bajo a medio en transporte aéreo. El clúster rojo incluye a los viajeros que tienen un gasto medio bajo tanto en alojamiento como en transporte aéreo. El clúster azul representa gastos medios altos en alojamiento y gastos bajos en transporte aéreo, mientras que el clúster celeste ubica gastos medios altos en transporte aéreo con independencia del gasto en alojamiento.

La segmentación en la comparación de gastos en alojamiento versus transporte terrestre arrojó 4 grupos, como se resume en la figura 14. El clúster verde se encuentran los visitantes que gastan muy poco en alojamiento y transporte terrestre. Luego, el clúster rojo agrupa a los viajeros que tienen un gasto medio bajo en alojamiento y que incurren en gasto medio en transporte terrestre. El clúster azul refleja gastos medios altos en alojamiento con independencia del gasto en transporte terrestre. A su vez, el clúster celeste muestra que la franja de gasto medio bajo a medio alto en alojamiento se corresponde con gasto medio bajo a bajo en transporte terrestre.

Figura 14. Clústeres por gasto en alojamiento vs transporte terrestre



Fuente: Elaboración propia

Considerando los conceptos de gasto por alojamiento y alimentos y bebidas, se evidenciaron 4 grupos de visitantes de acuerdo a la figura 15. El clúster verde agrupando a los viajeros que incurren en un gasto medio bajo en alojamiento, pero en un gasto medio alto en alimentos y bebidas. El clúster rojo representa gastos medios altos en alojamiento y alimentos y bebidas. El clúster azul donde gastos medios altos en alojamiento se corresponden a gastos medios bajos en alimentos y bebidas. El clúster celeste representado por los visitantes que incurren en un gasto medio bajo tanto en alojamiento como en alimentos y bebidas.

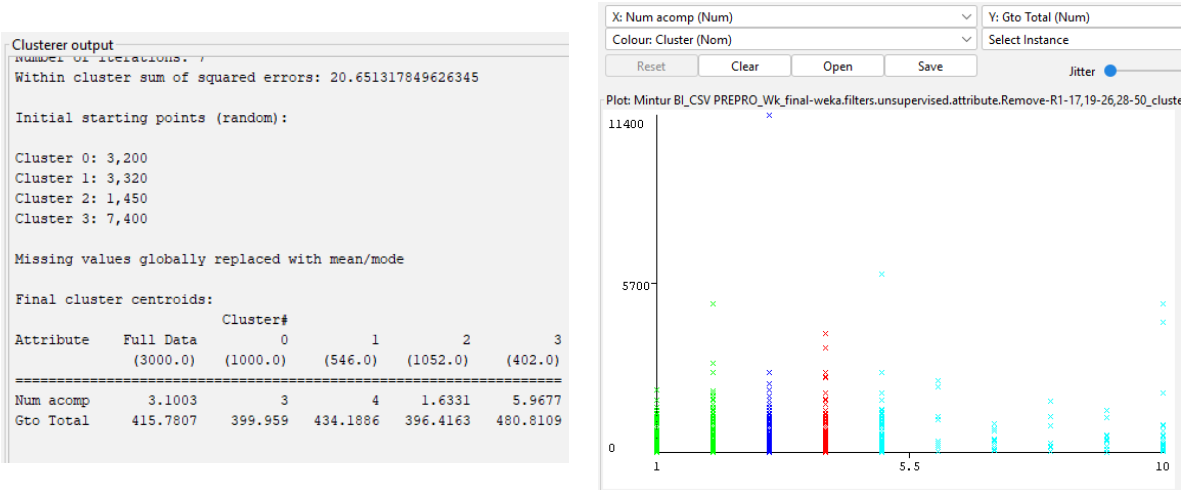
Figura 15. Clústeres por gasto en alojamiento vs alimentos y bebidas



Fuente: Elaboración propia

En cuanto al agrupamiento en función del gasto total del viaje y al tamaño de grupos, aunque se apreció un valor atípico para un grupo de 3 viajeros, se evidenciaron franjas promedio de gasto similares en torno a \$415.78 con independencia del número de viajeros de acuerdo a la figura 16. Por lo tanto, se intuye que esta cantidad se decide a priori y condiciona el viaje en su conjunto.

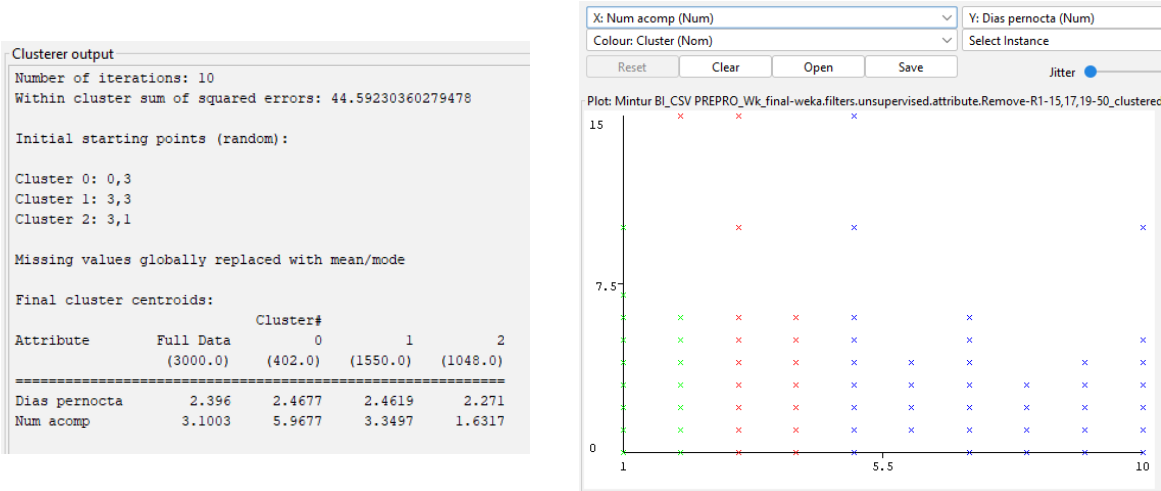
Figura 16. Clústeres por gasto total y tamaño de grupos



Fuente: Elaboración propia

Respecto a la segmentación por pernoctaciones y tamaño de grupos, se evidenció que la mayoría de las pernoctaciones son de una semana o menos con independencia del número de acompañantes al viaje, lo cual se refleja en la figura 17.

Figura 17. Clústeres por pernoctaciones y tamaño de grupos

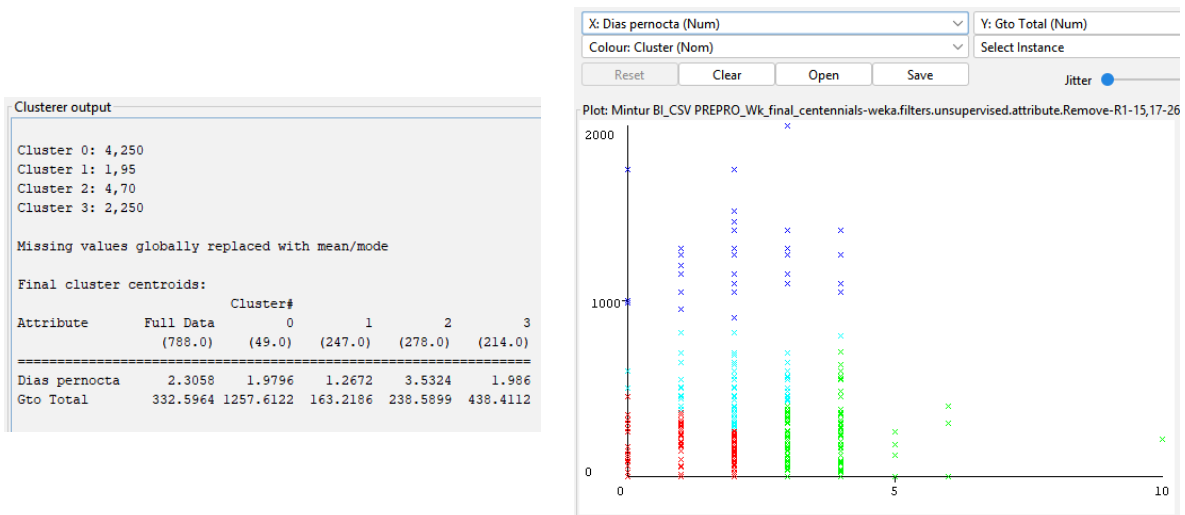


Fuente: Elaboración propia

De acuerdo a los cambios generacionales que más viajes realizaron, centennials, millennials y generación X, se segmentaron los viajes por pernoctaciones y gasto total. Para el caso de los centennials (figura 18), se observó que el gasto total de viaje no excede los \$2 000 ni suelen pernoctar por más de 10 días. En este contexto, se evidenciaron 4 grupos, el primero se refiere

a excursiones y viajes de dos días máximo, el cual incurre en un gasto total promedio alrededor de \$163.22. Un segundo grupo que viaja mayoritariamente entre 1 a 3 días y que gasta en total un promedio de \$438.41. El tercer grupo que, a pesar de que pernocta generalmente entre 3 a 5 días, se ubica en una franja de gasto promedio de \$238.59. Finalmente, un grupo que gasta por encima de \$1 000 con estancias de hasta 4 días.

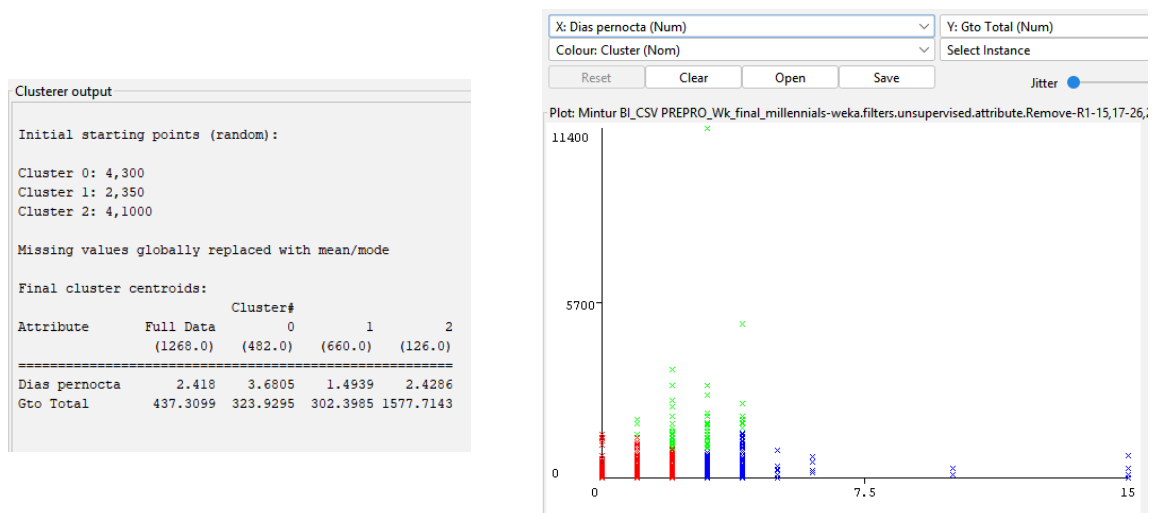
Figura 18. Clústeres de centennials por pernoctaciones y gasto total



Fuente: Elaboración propia

Los millennials (figura 19) mostraron constituir un segmento de turistas con viajes de mayor gasto y estancias de mayor duración frente a los centennials con extremos de \$11 400 y 15 días, aunque la mayoría hace excursiones o pernocta entre 1 y 4 días registrando gastos de hasta \$3 000. Existen dos grupos que incurren en gastos de alrededor de \$300 y \$320 en promedio, el primero en viajes de 1 a 2 días y el segundo de 3 días en adelante. Un tercer grupo gasta en todo su viaje alrededor de \$1 577.71 en estancias de hasta 4 días.

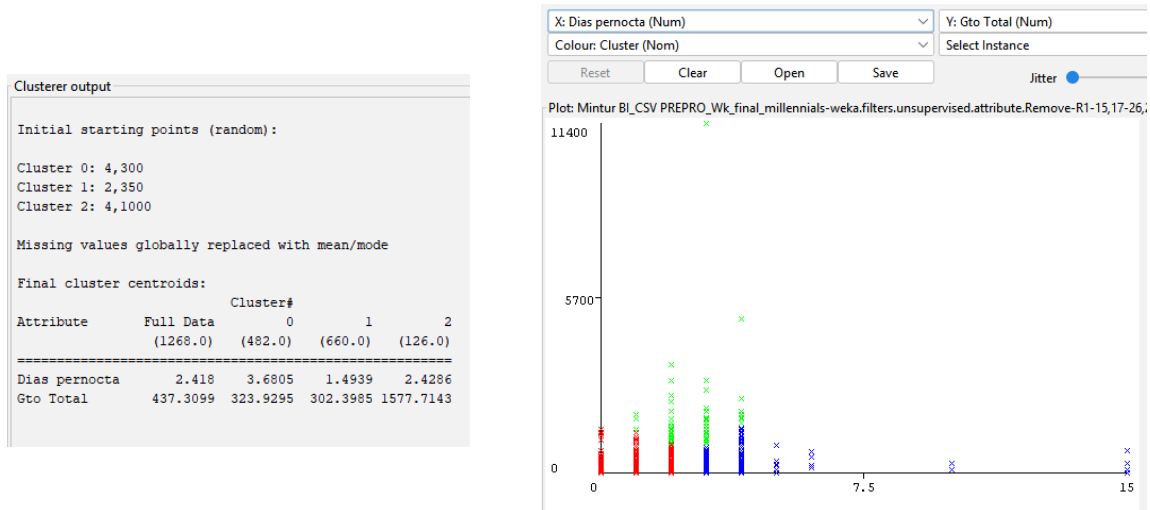
Figura 19. Clústeres de millennials por pernoctaciones y gasto total



Fuente: Elaboración propia

Los visitantes pertenecientes a la generación X (figura 20) evidenciaron un comportamiento similar al de los millennials en cuanto a gasto total del viaje y duración de la estancia, registrando valores atípicos de \$5 000 y 15 días. Entre las excursiones y estancias de duración de 1 hasta 4 días, se registran gastos alrededor de \$450-\$460, algo superiores al gasto de los millennials en ese período (\$300-\$320).

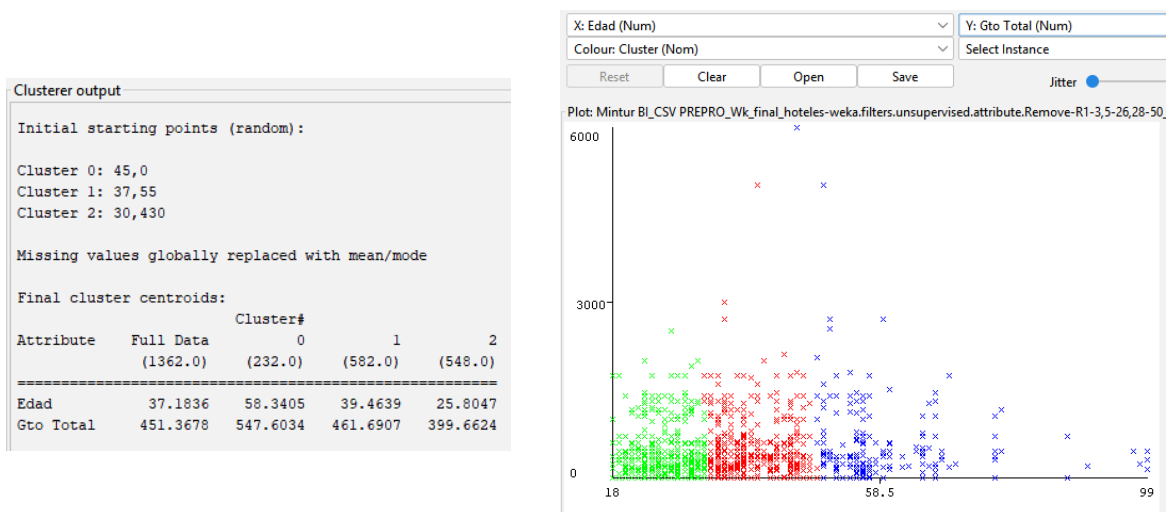
Figura 20. Clústeres de generación X por pernотaciones y gasto total



Fuente: Elaboración propia

De acuerdo al tipo de alojamiento, los viajes se segmentaron por la edad y el gasto total. Para quienes pernотaron en hoteles, el gasto total del viaje en promedio se concentró alrededor de \$400 y \$540, siendo la mayoría inferior a los \$2 000 independientemente de la edad de los turistas, según figura 21. Existen un grupo de valores aberrantes que superan los \$2 000 de gasto total. Además, se observa que los gastos totales más altos se corresponden a turistas de entre 40 y 50 años de edad, mayoritariamente pertenecientes a la Generación X.

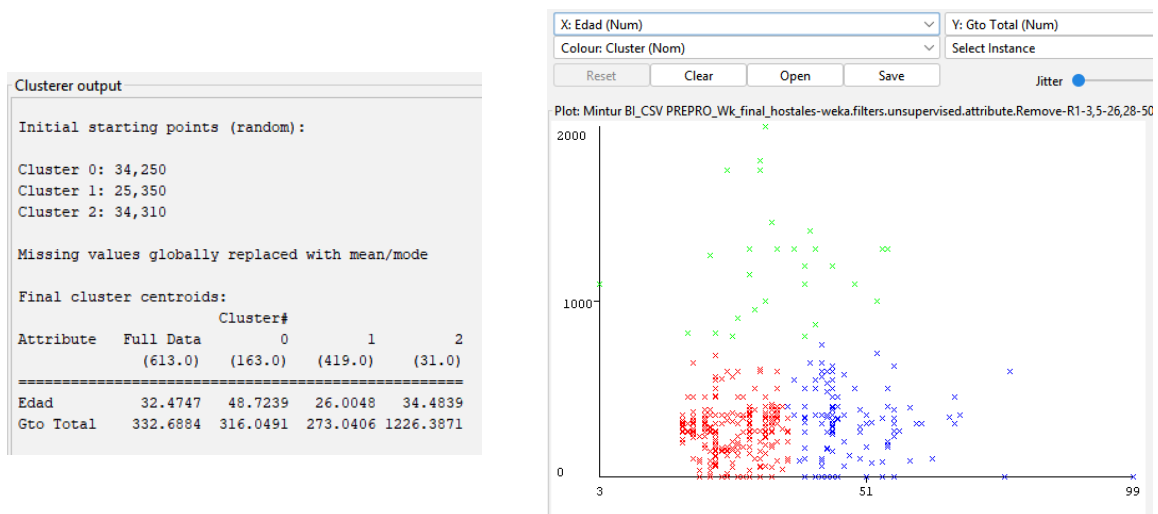
Figura 21. Clústeres de hoteles por edad y gasto total



Fuente: Elaboración propia

Los viajes cuya planificación incluyó pernoctación en hostales evidenciaron tres agrupamientos, el primero con centennials mayormente y millennials en menor medida incurren en un total de gastos no mayor a \$500, como se muestra en figura 22. En una franja de gasto similar incurren un segundo grupo pequeño de millennials, pero mayormente perteneciente a la generación X. Por otra parte, el tercer grupo conformado por millennials en su mayoría, incurre en una franja de gasto total del viaje entre los \$1 000 y \$2 000. Además, se evidenció una errata involuntaria en una respuesta al cuestionario respecto de un turista de 3 años de edad.

Figura 22. Clústeres de hostales por edad y gasto total

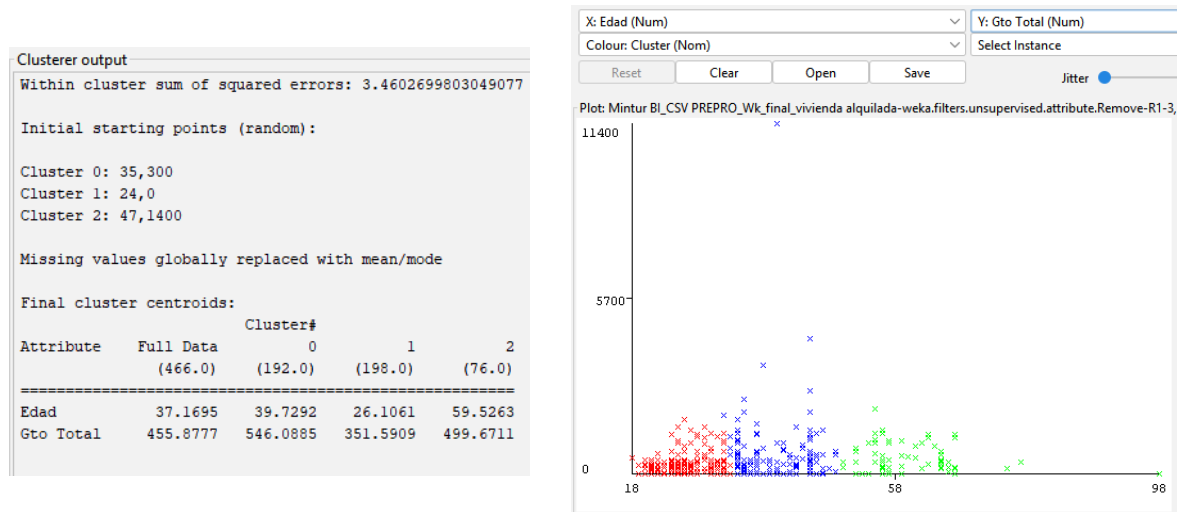


Fuente: Elaboración propia

En viviendas alquiladas (figura 23), se observó un comportamiento similar al evidenciado en la pernoctación en hoteles, el gasto total en promedio se concentró alrededor de \$350 y \$540, independientemente de la edad de los turistas. Los gastos totales especialmente altos (por encima de \$2 000) se corresponden exclusivamente a turistas de entre 40 y 50 años de edad, mayoritariamente pertenecientes a la generación X. Además, se registra un valor atípico que alcanza los \$11 400.

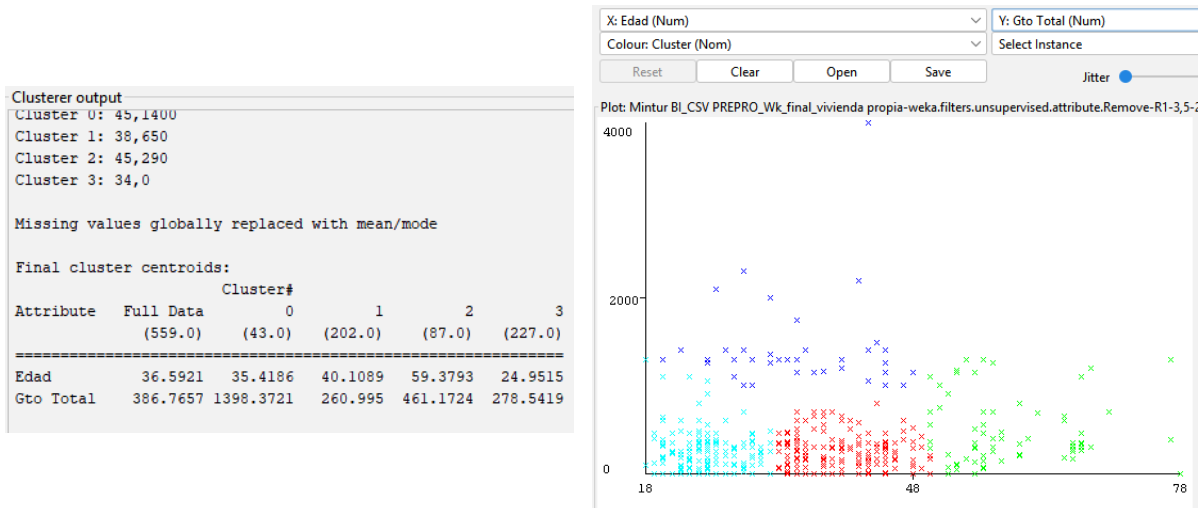
De quienes pernoctaron en vivienda propia, dos grupos (centennials, millennials y parte de la generación X) incurrieron en una franja de gasto total de entre \$260 y \$280 en promedio. El tercer grupo que aglutina tanto a la generación X como a baby boomers, incurre en un gasto total promedio de su viaje de alrededor de \$460. Un último grupo conformado mayoritariamente por millennials incurre en una franja de gasto total superior a \$1 000, evidenciando un valor máximo de hasta \$4 000. En la figura 24, se resume lo comentado respecto a los turistas que hicieron uso de sus propiedades como alojamiento.

Figura 23. Clústeres de viviendas alquiladas por edad y gasto total



Fuente: Elaboración propia

Figura 24. Clústeres de viviendas propias por edad y gasto total



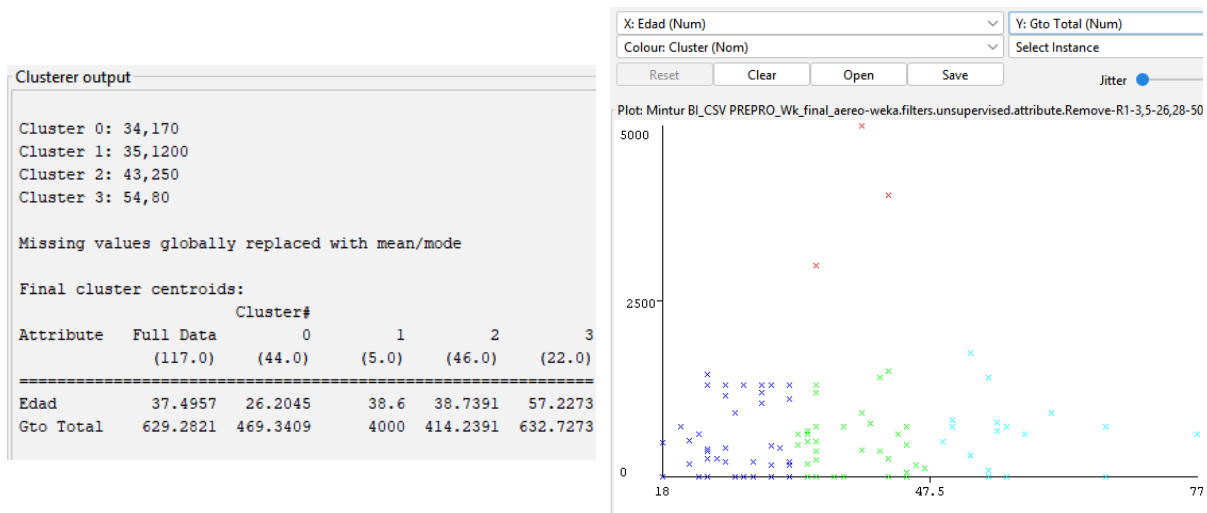
Fuente: Elaboración propia

De acuerdo al medio de transporte utilizado principalmente durante el viaje, se segmentó por edad y gasto total. En la figura 25, el gasto total para viajes que han recurrido a transporte aéreo está por debajo de los \$2 000 independientemente de la franja de edad de los visitantes. Además, existe un grupo que representa un 4% del colectivo estudiado cuyo gasto total de viaje oscila entre \$3 000 y \$5 000 perteneciente a la generación X.

Del grupo de viajes en los que los visitantes se desplazaron en autobús, se pueden formar cuatro grupos diferentes (figura 26). El primero donde se ubican los que gastan más \$1 000 y que se distribuyen desde los 18 a los 60 años de edad y los tres restantes que independientemente de la edad gastan en total durante su viaje menos de \$1 000. El gasto para estos tres grupos se distribuye alrededor de \$220 para los centennials, \$230 para los millennials y \$240 para los

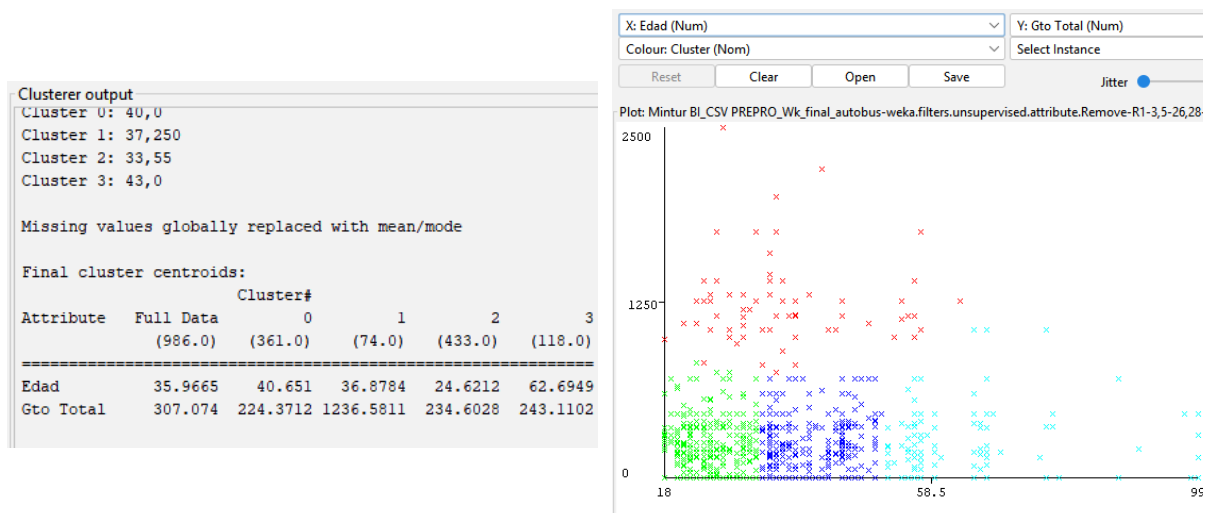
pertenecientes a la generación X y baby boomers. Se puede apreciar que, a partir de los 60 años, existen muy pocos visitantes que utilizan este medio transporte.

Figura 25. Clústeres de transporte aéreo por edad y gasto total



Fuente: Elaboración propia

Figura 26. Clústeres de autobús/flota por edad y gasto total



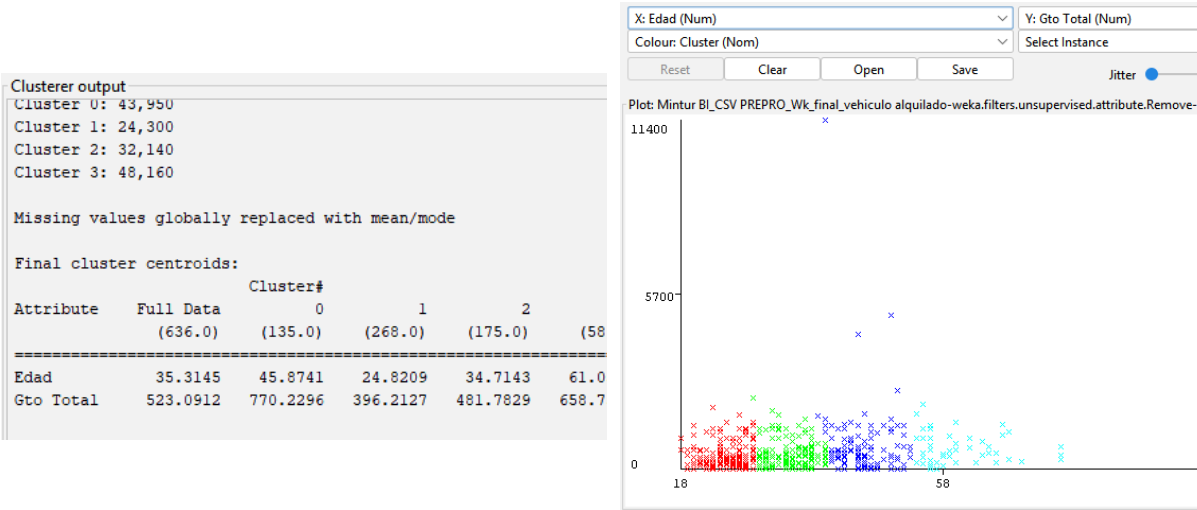
Fuente: Elaboración propia

En el caso de los visitantes que prefirieron alquilar un vehículo para desplazarse durante el viaje, independientemente del grupo de edad, el gasto total fue menor a \$2 000, de acuerdo a la figura 27. Exceptuando algunos valores atípicos altos que superan a los \$4 000, que vuelven a corresponderse a la franja de edad entre los 40 a 50 años. Al igual que en el grupo que utiliza el autobús, se aprecian pocos turistas por encima de los 70 años de edad.

Lo evidenciado en la figura 28 muestra que entre los visitantes que contaron con vehículo propio, existió un límite de gasto próximo a los \$3 000 con una distribución de grupos bastante similar a la generada para quienes pernoctaron en vivienda propia. Dos grupos que incluyeron a los centennials, millennials y parte de la generación X, gastaron en su viaje un total

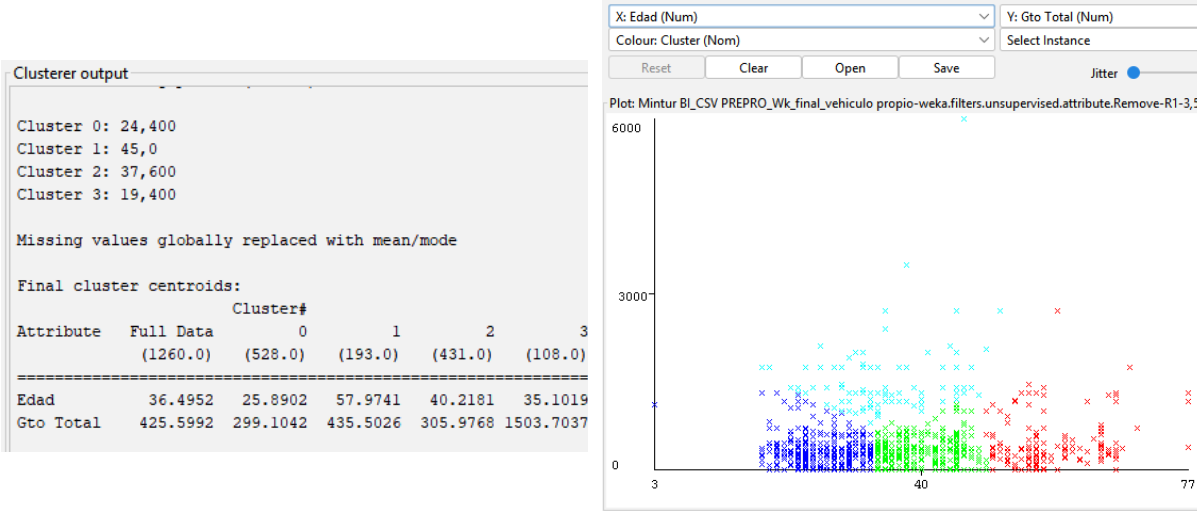
aproximado en promedio de \$299-\$305. Un tercer conglomerado involucra a la generación X y baby boomers, quienes gastaron alrededor de \$435 en promedio. Finalmente, el último grupo predominantemente millennial gastó un total promedio de \$1 500, con un valor máximo de \$6 000. Además, se evidenció una errata involuntaria en una respuesta al cuestionario respecto de un turista de 3 años de edad, ya puntualizado en la segmentación de hoteles.

Figura 27. Clústeres de vehículo alquilado por edad y gasto total



Fuente: Elaboración propia

Figura 28. Clústeres de vehículo propio por edad y gasto total



Fuente: Elaboración propia

Con miras a la extracción de patrones según motivo de viaje, tamaño de grupos, tiempo de antelación con que se planifica el viaje y el gasto total, se observa que el 72% de los viajes que se planificaron con un mes de antelación son viajes de grupo (3 o más visitantes) y que el 67% de los viajes que generan un gasto total medio son también de grupo. Además, cuando se planifica con un mes de antelación, el 60% de las veces se incurre en la franja de gasto total

medio y la razón del viaje es por vacaciones, recreo u ocio. En el caso de que se planifica el viaje con una semana de antelación y se incurre en la franja de gasto total medio, se trata de un viaje grupal el 60% de las veces, tal como lo corrobora la figura 29.

Figura 29. Patrones según motivo, grupos, antelación y gasto total

```

Associator output
=== Run information ===

Scheme:      weka.associations.Apriori -N 10 -T 0 -C 0.5 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation:    Mintur BI_CSV PREPRO_Wk_final-weka.filters.unsupervised.attribute.Remove-R1-10,12-14,16-18,20-27
Instances:   3000
Attributes:  4
              Razon viaje
              Antelac planif
              Num acomp D
              Gto Total D

=== Associator model (full training set) ===

Apriori
=====

Minimum support: 0.15 (450 instances)
Minimum metric <confidence>: 0.5
Number of cycles performed: 17

Generated sets of large itemsets:

Size of set of large itemsets L(1): 9

Size of set of large itemsets L(2): 11

Size of set of large itemsets L(3): 3

Best rules found:

1. Antelac planif=Un mes 789 ==> Num acomp D=Grupo 566 <conf:(0.72)> lift:(1.1) lev:(0.02) [53] conv:(1.24)
2. Gto Total D=Medio 1496 ==> Num acomp D=Grupo 999 <conf:(0.67)> lift:(1.03) lev:(0.01) [27] conv:(1.05)
3. Razon viaje =Vacaciones recreo ocio (turismo) Gto Total D=Medio 778 ==> Num acomp D=Grupo 518 <conf:(0.64)> lift:(1.19) lev:(0.03) [27] conv:(1.05)
4. Razon viaje =Vacaciones recreo ocio (turismo) 1608 ==> Num acomp D=Grupo 1052 <conf:(0.65)> lift:(1.01) lev:(0.03) [27] conv:(1.05)
5. Gto Total D=Alto 750 ==> Num acomp D=Grupo 490 <conf:(0.65)> lift:(1.01) lev:(0) [3] conv:(1.01)
6. Gto Total D=Alto 750 ==> Razon viaje =Vacaciones recreo ocio (turismo) 479 <conf:(0.64)> lift:(1.19) lev:(0.03) [27] conv:(1.05)
7. Antelac planif=Una semana Gto Total D=Medio 749 ==> Num acomp D=Grupo 452 <conf:(0.6)> lift:(0.93) lev:(0.03) [27] conv:(1.05)
8. Antelac planif=Un mes 789 ==> Gto Total D=Medio 475 <conf:(0.6)> lift:(1.21) lev:(0.03) [81] conv:(1.26)
9. Antelac planif=Un mes 789 ==> Razon viaje =Vacaciones recreo ocio (turismo) 470 <conf:(0.6)> lift:(1.11) lev:(0.03) [27] conv:(1.05)
10. Antelac planif=Una semana 1589 ==> Num acomp D=Grupo 926 <conf:(0.58)> lift:(0.9) lev:(-0.04) [-105] conv:(1.05)

```

Fuente: Elaboración propia

Respecto a los patrones vinculados al medio de transporte elegido para el viaje, los medios de comunicación para informarse sobre destinos y paquetes turísticos, las redes sociales para informarse de promociones y paquetes, las plataformas de viaje para información turística; y, la edad de los visitantes, se muestra la figura 30.

El primer patrón evidencia que el 100% de quienes utilizan redes sociales para informarse sobre destinos y paquetes turísticos del Ecuador, no utilizan plataforma turística alguna como Despegar, Booking, Kayak, Trivago, entre otras. En un 97%, los viajeros que utilizan la red social Facebook como medio para conocer promociones y/o paquetes turísticos, tampoco utilizan plataformas de viajes. Los patrones 5 y 7 refieren que tanto los centennials como millennials, tampoco utilizan plataformas de viajes.

Figura 30. Patrones según transporte, medios de comunicación, redes, plataformas y edad

```

Associator output
=== Run information ===

Scheme:          weka.associations.Apriori -N 10 -T 0 -C 0.5 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation:        Mintur BI_CSV PREPRO_Wk_final-weka.filters.unsupervised.attribute.Remove-R1-4,6-28,33-50
Instances:       3000
Attributes:      5
                 Edad D
                 Medio transp
                 Medios comunic
                 RRSS
                 Plataf viajes
=== Associator model (full training set) ===

Apriori
=====

Minimum support: 0.2 (600 instances)
Minimum metric <confidence>: 0.5
Number of cycles performed: 16

Generated sets of large itemsets:

Size of set of large itemsets L(1): 11

Size of set of large itemsets L(2): 10

Best rules found:

1. Medios comunic=Redes sociales 1297 ==> Plataf viajes=No utiliza 1295 <conf:(1)> lift:(1.29) lev:(0.1) [0]
2. RRSS=Facebook 640 ==> Plataf viajes=No utiliza 620 <conf:(0.97)> lift:(1.25) lev:(0.04) [123] conv:(6.8)
3. Medios comunic=Internet 669 ==> RRSS=No utiliza 627 <conf:(0.94)> lift:(1.81) lev:(0.09) [280] conv:(7.1)
4. Medio transp=Autobus / flotas 986 ==> Plataf viajes=No utiliza 859 <conf:(0.87)> lift:(1.12) lev:(0.03)
5. Edad D=Centennials 788 ==> Plataf viajes=No utiliza 659 <conf:(0.84)> lift:(1.08) lev:(0.02) [47] conv:
6. Medio transp=Vehiculo propio 1260 ==> Plataf viajes=No utiliza 954 <conf:(0.76)> lift:(0.98) lev:(-0.01)
7. Edad D=Millennials 1268 ==> Plataf viajes=No utiliza 946 <conf:(0.75)> lift:(0.96) lev:(-0.01) [-38] conv:
8. RRSS=No utiliza 1556 ==> Plataf viajes=No utiliza 927 <conf:(0.6)> lift:(0.77) lev:(-0.09) [-280] conv:
9. Plataf viajes=No utiliza 2329 ==> Medios comunic=Redes sociales 1295 <conf:(0.56)> lift:(1.29) lev:(0.1)
10. Medio transp=Vehiculo propio 1260 ==> RRSS=No utiliza 682 <conf:(0.54)> lift:(1.04) lev:(0.01) [28] conv:

```

Fuente: Elaboración propia

En cuanto a los patrones sobre la complejidad de aspectos del viaje (figura 31), debido al sesgo de la muestra hacia procesos percibidos como muy fáciles, no se han encontrado patrones significativos que evidencien claras dificultades en la organización del viaje respecto a alojamiento, reservas, alimentos y bebidas; y, transporte.

Figura 31. Patrones en complejidad de aspectos del viaje

```

Associator output
Apriori
=====

Minimum support: 0.25 (750 instances)
Minimum metric <confidence>: 0.5
Number of cycles performed: 15

Generated sets of large itemsets:

Size of set of large itemsets L(1): 9

Size of set of large itemsets L(2): 5

Best rules found:

1. Escala alojamiento D=MUY FACIL 1094 ==> Escala reservas D=MUY FACIL 834 <conf:(0.76)> lift:(1.04) lev:(0.01) [123] conv:(6.8)
2. Escala aliment D=MUY FACIL 1086 ==> Escala transporte D=MUY FACIL 821 <conf:(0.76)> lift:(1.04) lev:(0.01) [123] conv:(6.8)
3. Escala aliment D=MUY FACIL 1086 ==> Escala reservas D=MUY FACIL 777 <conf:(0.72)> lift:(1.04) lev:(0.01) [123] conv:(6.8)
4. Escala alojamiento D=MUY FACIL 1094 ==> Escala transporte D=MUY FACIL 766 <conf:(0.7)> lift:(1.04) lev:(0.01) [123] conv:(6.8)
5. Escala transporte D=MUY FACIL 1402 ==> Escala reservas D=MUY FACIL 969 <conf:(0.69)> lift:(1.04) lev:(0.01) [123] conv:(6.8)
6. Escala reservas D=MUY FACIL 1449 ==> Escala transporte D=MUY FACIL 969 <conf:(0.67)> lift:(1.04) lev:(0.01) [123] conv:(6.8)
7. Escala transporte D=MUY FACIL 1402 ==> Escala aliment D=MUY FACIL 821 <conf:(0.59)> lift:(1.04) lev:(0.01) [123] conv:(6.8)
8. Escala reservas D=MUY FACIL 1449 ==> Escala alojamiento D=MUY FACIL 834 <conf:(0.58)> lift:(1.04) lev:(0.01) [123] conv:(6.8)
9. Escala transporte D=MUY FACIL 1402 ==> Escala alojamiento D=MUY FACIL 766 <conf:(0.55)> lift:(1.04) lev:(0.01) [123] conv:(6.8)
10. Escala reservas D=MUY FACIL 1449 ==> Escala aliment D=MUY FACIL 777 <conf:(0.54)> lift:(1.04) lev:(0.01) [123] conv:(6.8)

```

Fuente: Elaboración propia

Respecto de los patrones en los gastos de los preparativos (antes), durante el viaje, adicionales en el retorno (después) y el total del viaje (figura 32), no se han encontrado patrones significativos que evidencien tendencias hacia valores de gasto bajo o alto antes, durante y después del viaje ni en el gasto total de este, debido al sesgo de la muestra hacia gastos totales etiquetados como medios en sus respectivas categorías. Los patrones de asociación entre los rubros de gasto antes, durante y después del viaje además del gasto total del viaje evidencian que viajes con gasto total medio (\$141-\$463) en un 90% se corresponden con franja de gasto medio durante el viaje (\$51-\$300). En un 87%, los viajes donde se incurre en una franja de gasto total medio, los gastos preparativos previos al viaje se ubican en esa misma franja de gasto (\$21-\$100) y en un 80%, los gastos adicionales después del viaje también se ubican en un nivel de gasto medio (\$21-\$100).

Figura 32. Patrones en los gastos del viaje

```

Associator output
=====
Minimum support: 0.4 (1200 instances)
Minimum metric <confidence>: 0.5
Number of cycles performed: 12

Generated sets of large itemsets:

Size of set of large itemsets L(1): 4

Size of set of large itemsets L(2): 6

Best rules found:

1. Gto Total D=Medio 1496 ==> Gtos durante D=Medio 1342 <conf:(0.9)> lift:(1.56) lev:(0.8)
2. Gto Total D=Medio 1496 ==> Gtos antes D=Medio 1303 <conf:(0.87)> lift:(1.64) lev:(0.8)
3. Gtos antes D=Medio 1590 ==> Gto Total D=Medio 1303 <conf:(0.82)> lift:(1.64) lev:(0.8)
4. Gto Total D=Medio 1496 ==> Gtos despues D=Medio 1200 <conf:(0.8)> lift:(1.51) lev:(0.8)
5. Gtos antes D=Medio 1590 ==> Gtos durante D=Medio 1242 <conf:(0.78)> lift:(1.36) lev:(0.8)
6. Gtos durante D=Medio 1721 ==> Gto Total D=Medio 1342 <conf:(0.78)> lift:(1.56) lev:(0.8)
7. Gtos despues D=Medio 1589 ==> Gtos durante D=Medio 1228 <conf:(0.77)> lift:(1.35) lev:(0.8)
8. Gtos despues D=Medio 1589 ==> Gtos antes D=Medio 1226 <conf:(0.77)> lift:(1.46) lev:(0.8)
9. Gtos antes D=Medio 1590 ==> Gtos despues D=Medio 1226 <conf:(0.77)> lift:(1.46) lev:(0.8)
10. Gtos despues D=Medio 1589 ==> Gto Total D=Medio 1200 <conf:(0.76)> lift:(1.51) lev:(0.8)

```

Fuente: Elaboración propia

Para la predicción del umbral de gasto total del viaje, se realizó una detección previa de variables significativas para el gasto total categorizado de acuerdo a atributos netamente del visitante, del viaje; y, de la combinación visitante - viaje. Como primer paso, se analizaron todas las variables o atributos del visitante (edad discretizada, estado civil, género, nivel de estudios e ingreso promedio mensual del hogar) para la generación de un árbol preliminar de decisión (clasificación) respecto a los atributos del visitante en cuanto a la predicción del gasto total del viaje, de acuerdo a lo mostrado en figura 33.

Figura 33. Árbol de decisión para el umbral de gasto a partir de atributos del visitante

```

Classifier output
==== Run information ====

Scheme:      weka.classifiers.trees.J48 -C 0.15 -M 2
Relation:    Mintur_BI_CSV_PREPRO_Wk_final-weka.filters.unsupervised.attribute.Remove-R2-4,10-27,29-50-weka.filters.unsupervised.attribute.Remove-R1
Instances:   3000
Attributes:  6
    Edad D
    Estado civil
    Genero
    Nivel estudios
    Ingreso prom mes hogar
    Gto Total D
Test mode:   10-fold cross-validation

==== Classifier model (full training set) ====

J48 pruned tree
-----

Ingreso prom mes hogar = B de usd 401 a 800: Medio (1375.0/572.0)
Ingreso prom mes hogar = A menos de usd 400: Medio (322.0/173.0)
Ingreso prom mes hogar = E de usd 1501 a 1700
| Nivel estudios = Superior: Alto (127.0/66.0)
| Nivel estudios = Secundaria
| | Edad D = Centennials: Bajo (3.0/1.0)
| | Edad D = Generaci3n X: Alto (1.0)
| | Edad D = Millennials: Medio (36.0/16.0)
| | Edad D = Baby boomers: Medio (0.0)
| | Edad D = Tradicionalistas: Medio (0.0)
| Nivel estudios = Primaria: Alto (3.0/1.0)
Ingreso prom mes hogar = D de usd 1201 a 1500
| Edad D = Centennials
| | Estado civil = Soltera/o: Medio (22.0/15.0)
| | Estado civil = Casada/o: Nulo (5.0/2.0)
| | Estado civil = Union libre: Medio (1.0)
| Edad D = Generaci3n X
| | Estado civil = Soltera/o: Alto (18.0/6.0)
| | Estado civil = Casada/o: Medio (47.0/27.0)
| | Estado civil = Union libre: Alto (14.0/7.0)
| Edad D = Millennials
| | Genero = Femenino
| | | Estado civil = Soltera/o: Alto (33.0/20.0)
| | | Estado civil = Casada/o: Medio (20.0/9.0)
| | | Estado civil = Union libre: Alto (3.0/1.0)
| | Genero = Masculino
| | | Nivel estudios = Superior: Alto (50.0/26.0)
| | | Nivel estudios = Secundaria: Medio (7.0/4.0)
| | | Nivel estudios = Primaria: Alto (0.0)
| | Genero = Prefiero no responder: Alto (2.0)
| Edad D = Baby boomers
| | Genero = Femenino: Medio (19.0/4.0)
| | Genero = Masculino: Alto (10.0/4.0)
| | Genero = Prefiero no responder: Medio (0.0)
| Edad D = Tradicionalistas: Medio (0.0)
Ingreso prom mes hogar = C de usd 801 a 1200: Medio (723.0/371.0)
Ingreso prom mes hogar = F Mas de usd 1701: Alto (159.0/63.0)

Number of Leaves :    28
Size of the tree :    38

Time taken to build model: 0 seconds

==== Stratified cross-validation ====
==== Summary ====

Correctly Classified Instances      1583           52.7667 %
Incorrectly Classified Instances    1417           47.2333 %
Kappa statistic                    0.1232
Mean absolute error                 0.3145
Root mean squared error             0.3991
Relative absolute error             95.7021 %
Root relative squared error         98.4782 %
Total Number of Instances          3000

==== Detailed Accuracy By Class ====

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
      0,910    0,797    0,532     0,910    0,672     0,161    0,588    0,553    Medio
      0,288    0,090    0,516     0,288    0,370     0,247    0,623    0,382    Alto
      0,003    0,004    0,091     0,003    0,006    -0,005    0,536    0,150    Bajo
      0,010    0,002    0,400     0,010    0,019     0,045    0,515    0,147    Nulo
Weighted Avg.  0,528    0,421    0,459     0,528    0,431     0,148    0,581    0,409

==== Confusion Matrix ====

  a   b   c   d   <-- classified as
1362 122   9   3 | a = Medio
 532 216   0   2 | b = Alto
 313   33   1   1 | c = Bajo
 353   48   1   4 | d = Nulo
    
```

Fuente: Elaboración propia

Respecto al árbol obtenido, se observó que cuando el ingreso promedio mensual del hogar del visitante está entre \$1 501 y \$1 700, tiene educación secundaria y su grupo de edad pertenece a los millennials, su nivel de gasto para el viaje es medio en 36 ocasiones, donde se incluyen 16 errores (instancias mal clasificadas). Mientras que cuando los visitantes poseen educación superior, 127 incurren en un nivel de gasto alto, siendo correctamente clasificadas 61 instancias.

Luego, se seleccionaron atributos para evidenciar las variables del viaje que inciden significativamente sobre el gasto total del viaje. Así, se analizaron todas las variables o atributos del viaje (persona que decide el viaje, motivo de viaje, región de destino, tiempo de antelación para planificar el viaje, tipo de alojamiento, número discretizado de acompañantes en el viaje y medio de transporte principal durante el viaje) y se obtuvo el ranking que ordena de mayor a menor la importancia de las variables estudiadas para la variable objetivo, tal como se resume en la figura 34.

Figura 34. Selección de atributos de viaje con gasto total como variable objetivo

```
Attribute selection output
  Region viaje
  Antelac planif
  Tipo Alojajam
  Num acomp D
  Gto Total D
  Medio transp
Evaluation mode:  evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:
  Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 7 Gto Total D):
  Information Gain Ranking Filter

Ranked attributes:
0.05225  5 Tipo Alojajam
0.04762  8 Medio transp
0.02783  2 Razon viaje
0.02598  4 Antelac planif
0.02447  3 Region viaje
0.01074  1 Miembro decide viaje
0.00837  6 Num acomp D

Selected attributes: 5,8,2,4,3,1,6 : 7
```

Fuente: Elaboración propia

Posteriormente, se seleccionaron las variables tipo de alojamiento, medio de transporte, motivo de viaje, tiempo de antelación para planificar el viaje y región de destino para la generación de un árbol preliminar de decisión (clasificación) respecto a los atributos del viaje en cuanto a la predicción del gasto total del viaje, de acuerdo a lo mostrado en figura 35.

Figura 35. Árbol de decisión para el umbral de gasto a partir de atributos del viaje

```

Classifier output
=== Run information ===

Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    Mintur_BI_CSV PREPRO_Wk_final-weka.filters.unsupervised.attribute.Remove-R1-10,13-14,16,18-27,30-50
Instances:   3000
Attributes:  6
             Razon viaje
             Region viaje
             Antelac planif
             Tipo Alojajam
             Gto Total D
             Medio transp
Test mode:   10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree
-----

Tipo Alojajam = Vivienda propia
|
|  Antelac planif = Una semana
|  |
|  |  Region viaje = Costa
|  |  |
|  |  |  Razon viaje = Vacaciones recreo ocio (turismo): Medio (51.0/31.0)
|  |  |  Razon viaje = Religion: Medio (3.0/1.0)
|  |  |  Razon viaje = Visitas a familiares y amigos: Bajo (48.0/26.0)
|  |  |  Razon viaje = Salud y atencion medica: Medio (3.0/2.0)
|  |  |  Razon viaje = Negocios y/o motivos profesionales: Bajo (7.0/3.0)
|  |  |  Razon viaje = Compras: Alto (2.0)
|  |  |  Razon viaje = Educacion y formacion: Medio (6.0/2.0)
|  |  |  Region viaje = Sierra: Medio (141.0/58.0)
|  |  |  Region viaje = Galapagos: Alto (11.0/3.0)
|  |  |  Region viaje = Amazonia: Medio (23.0/11.0)
|  |  |  Antelac planif = Tres meses: Alto (43.0/22.0)
|  |  |  Antelac planif = Un mes: Medio (190.0/53.0)
|  |  |  Antelac planif = Seis meses
|  |  |  |
|  |  |  |  Razon viaje = Vacaciones recreo ocio (turismo): Alto (5.0/1.0)
|  |  |  |  Razon viaje = Religion: Alto (2.0/1.0)
|  |  |  |  Razon viaje = Visitas a familiares y amigos: Medio (4.0/1.0)
|  |  |  |  Razon viaje = Salud y atencion medica: Medio (3.0/1.0)
|  |  |  |  Razon viaje = Negocios y/o motivos profesionales: Alto (2.0/1.0)
|  |  |  |  Razon viaje = Compras: Medio (0.0)
|  |  |  |  Razon viaje = Educacion y formacion: Medio (3.0/1.0)
|  |  |  |  Antelac planif = Mas de seis meses: Medio (12.0/7.0)
|  |  |  Tipo Alojajam = Hoteles
|  |  |  |
|  |  |  |  Medio transp = Autobus / flotas: Medio (418.0/234.0)
|  |  |  |  Medio transp = Vehiculo propio: Medio (556.0/307.0)
|  |  |  |  Medio transp = Aereo
|  |  |  |  |
|  |  |  |  |  Razon viaje = Vacaciones recreo ocio (turismo): Alto (60.0/27.0)
|  |  |  |  |  Razon viaje = Religion: Nulo (3.0/1.0)
|  |  |  |  |  Razon viaje = Visitas a familiares y amigos: Nulo (12.0/4.0)
|  |  |  |  |  Razon viaje = Salud y atencion medica: Nulo (7.0/4.0)
|  |  |  |  |  Razon viaje = Negocios y/o motivos profesionales: Medio (4.0/2.0)
|  |  |  |  |  Razon viaje = Compras: Alto (1.0)
|  |  |  |  |  Razon viaje = Educacion y formacion: Nulo (6.0/3.0)
|  |  |  |  Medio transp = Vehiculo alquilado
|  |  |  |  |
|  |  |  |  |  Region viaje = Costa: Alto (145.0/72.0)
|  |  |  |  |  Region viaje = Sierra
|  |  |  |  |  |
|  |  |  |  |  |  Razon viaje = Vacaciones recreo ocio (turismo): Medio (64.0/31.0)
|  |  |  |  |  |  Razon viaje = Religion: Medio (1.0)
|  |  |  |  |  |  Razon viaje = Visitas a familiares y amigos: Nulo (11.0/6.0)
|  |  |  |  |  |  Razon viaje = Salud y atencion medica: Bajo (10.0/3.0)
|  |  |  |  |  |  Razon viaje = Negocios y/o motivos profesionales: Alto (12.0/6.0)
|  |  |  |  |  |  Razon viaje = Compras: Medio (2.0/1.0)
|  |  |  |  |  |  Razon viaje = Educacion y formacion
|  |  |  |  |  |  |
|  |  |  |  |  |  |  Antelac planif = Una semana: Alto (6.0/2.0)
|  |  |  |  |  |  |  Antelac planif = Tres meses: Medio (2.0/1.0)
|  |  |  |  |  |  |  Antelac planif = Un mes: Medio (4.0/2.0)
|  |  |  |  |  |  |  Antelac planif = Seis meses: Alto (0.0)
|  |  |  |  |  |  |  Antelac planif = Mas de seis meses: Alto (0.0)
|  |  |  |  |  |  Region viaje = Galapagos: Alto (13.0/2.0)
|  |  |  |  |  |  Region viaje = Amazonia
|  |  |  |  |  |  |
|  |  |  |  |  |  |  Antelac planif = Una semana: Alto (13.0/7.0)
|  |  |  |  |  |  |  Antelac planif = Tres meses: Medio (5.0/1.0)
|  |  |  |  |  |  |  Antelac planif = Un mes: Medio (5.0/1.0)
|  |  |  |  |  |  |  Antelac planif = Seis meses: Medio (1.0)
|  |  |  |  |  |  |  Antelac planif = Mas de seis meses: Alto (1.0)
|  |  |  |  |  |  Medio transp = Fluvial: Medio (0.0)
|  |  |  Tipo Alojajam = Hostales
|  |  |  |
|  |  |  |  Antelac planif = Una semana: Medio (310.0/115.0)
|  |  |  |  Antelac planif = Tres meses: Medio (86.0/16.0)
|  |  |  |  Antelac planif = Un mes: Medio (184.0/26.0)
|  |  |  |  Antelac planif = Seis meses: Medio (17.0/2.0)
|  |  |  |  Antelac planif = Mas de seis meses: Alto (16.0/2.0)
Tipo Alojajam = Vivienda alquilada
|
|  Medio transp = Autobus / flotas: Medio (162.0/90.0)
|  Medio transp = Vehiculo propio
|  |
|  |  Antelac planif = Una semana
|  |  |
|  |  |  Razon viaje = Vacaciones recreo ocio (turismo): Medio (41.0/22.0)
|  |  |  Razon viaje = Religion: Nulo (3.0/1.0)
|  |  |  Razon viaje = Visitas a familiares y amigos: Alto (16.0/9.0)
|  |  |  Razon viaje = Salud y atencion medica
|  |  |  |
|  |  |  |  Region viaje = Costa: Nulo (0.0)
|  |  |  |  Region viaje = Sierra: Medio (2.0/1.0)
|  |  |  |  Region viaje = Galapagos: Nulo (0.0)
|  |  |  |  Region viaje = Amazonia: Nulo (5.0/2.0)
|  |  |  Razon viaje = Negocios y/o motivos profesionales: Bajo (2.0/1.0)
|  |  |  Razon viaje = Compras: Nulo (1.0)
|  |  |  Razon viaje = Educacion y formacion

```

```

| | | Region viaje = Costa: Nulo (3.0/1.0)
| | | Region viaje = Sierra: Alto (5.0/2.0)
| | | Region viaje = Galapagos: Nulo (0.0)
| | | Region viaje = Amazonia: Nulo (4.0/2.0)
| | Antelac planif = Tres meses: Medio (41.0/18.0)
| | Antelac planif = Un mes
| | | Region viaje = Costa: Alto (18.0/6.0)
| | | Region viaje = Sierra: Medio (19.0/7.0)
| | | Region viaje = Galapagos: Medio (3.0/2.0)
| | | Region viaje = Amazonia: Alto (2.0/1.0)
| | Antelac planif = Seis meses: Medio (15.0/5.0)
| | Antelac planif = Mas de seis meses: Medio (6.0/3.0)
| Medio transp = Aereo: Alto (8.0/4.0)
| Medio transp = Vehiculo alquilado
| | Region viaje = Costa: Alto (67.0/30.0)
| | Region viaje = Sierra: Medio (24.0/13.0)
| | Region viaje = Galapagos: Alto (8.0/2.0)
| | Region viaje = Amazonia: Medio (11.0/5.0)
| Medio transp = Fluvial: Medio (0.0)

Number of Leaves : 80
Size of the tree : 99

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances 1555 51.8333 %
Incorrectly Classified Instances 1445 48.1667 %
Kappa statistic 0.1151
Mean absolute error 0.3077
Root mean squared error 0.3986
Relative absolute error 93.6315 %
Root relative squared error 98.3432 %
Total Number of Instances 3000

=== Detailed Accuracy By Class ===
TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class
0,896 0,783 0,532 0,896 0,668 0,154 0,615 0,620 Medio
0,255 0,096 0,469 0,255 0,330 0,201 0,641 0,384 Alto
0,037 0,011 0,317 0,037 0,067 0,074 0,587 0,157 Bajo
0,025 0,009 0,303 0,025 0,046 0,052 0,650 0,213 Nulo
Weighted Avg. 0,518 0,417 0,461 0,518 0,430 0,143 0,623 0,452

=== Confusion Matrix ===
 a b c d <-- classified as
1341 131 15 9 | a = Medio
539 191 11 9 | b = Alto
301 29 13 5 | c = Bajo
338 56 2 10 | d = Nulo

```

Fuente: Elaboración propia

Respecto al árbol obtenido, de los visitantes que utilizaron vivienda propia como alojamiento, cuando se planifica el viaje con una semana de antelación y la región de destino es Sierra, el nivel de gasto para el viaje es medio en 141 ocasiones, con un error de clasificación del 41%. Cuando la planificación es de un mes, 190 viajeros evidencian un gasto medio con 137 instancias correctamente clasificadas. Los grupos de viajeros que se alojaron en hoteles y que viajaron por autobús/flotas o vehículo propio, incurrieron en una franja de gasto medio. Mientras que, de quienes pernoctaron en hoteles y se trasladaron por vía aérea con motivo de vacacionar, recrearse y divertirse, 60 viajeros incurrieron en una franja de gasto alto, de los cuales 27 se clasificaron incorrectamente. Además, del grupo de visitantes que utilizó hoteles para alojarse y se transportaron en vehículo alquilado, en el subgrupo de quienes tuvieron como destino la región Costa, 145 incurrieron en una franja de gasto alto con 73 instancias clasificadas correctamente; el subgrupo que viajó a la Sierra o región Andina por motivo de vacaciones, en 64 viajes incurrió en gasto medio (con 31 errores), en tanto que, de quienes lo hicieron por negocios o motivos profesionales, 12 incurrieron en gasto alto (con 6 errores).

Los viajes que consideraron hostales como alojamiento y una planificación de hasta seis meses, reflejaron un nivel de gasto medio, por el contrario, cuando el viaje se planificaba con más de seis meses de anticipación, 16 viajes registraron un nivel de gasto alto incluyendo únicamente 2 instancias clasificadas incorrectamente. En el grupo de viajes que consideró viviendas alquiladas para pernoctar, vehículo propio como medio de transporte y fue planificado con tres meses de antelación, 41 evidenciaron un nivel de gasto medio con 18 instancias erróneamente clasificadas; por su parte, los viajes que se planificaron con un mes de antelación y cuya región de destino fue Costa, registraron 18 viajes asociados con gasto alto, de los cuales 12 se clasificaron correctamente. No obstante, de los viajes que consideraron viviendas alquiladas, vehículo alquilado para transportarse y región Costa o Litoral como destino, mostraron 67 viajes con un nivel de gasto alto, con un 45% de instancias mal clasificadas.

Finalmente, para la predicción del umbral de gasto total del viaje, se realizó una detección previa de variables significativas para el gasto total categorizado de acuerdo a los atributos más relevantes de la combinación visitante - viaje previamente analizada por separado. Así, se analizaron únicamente las características del visitante (edad discretizada, estado civil, género, nivel de estudios e ingreso promedio mensual del hogar) y del viaje (motivo de viaje, región de destino, tiempo de antelación para planificar el viaje, tipo de alojamiento y medio de transporte principal durante el viaje) y se obtuvo el ranking que ordena de mayor a menor la importancia de las variables estudiadas para la variable objetivo, tal como se resume en la figura 36.

Figura 36. Selección de atributos de visitante-viaje con gasto total como variable objetivo

```

Attribute selection output
      Tipo Alojamiento
      Gto Total D
      Medio transp
Evaluation mode:    evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:
      Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 10 Gto Total D):
      Information Gain Ranking Filter

Ranked attributes:
0.05896  5 Ingreso prom mes hogar
0.05225  9 Tipo Alojamiento
0.04762  11 Medio transp
0.02864  1 Edad D
0.02783  6 Razon viaje
0.02598  8 Antelac planif
0.02447  7 Region viaje
0.01521  2 Estado civil
0.01052  3 Genero
0.00765  4 Nivel estudios

Selected attributes: 5,9,11,1,6,8,7,2,3,4 : 10

```

Fuente: Elaboración propia

Posteriormente, se seleccionaron las variables ingreso promedio mensual del hogar, tipo de alojamiento, medio de transporte y edad discretizada para la generación del árbol de decisión (clasificación) respecto a los atributos del visitante y del viaje en cuanto a la predicción del umbral del gasto total del viaje, de acuerdo a lo mostrado en figura 37.

Figura 37. Árbol de decisión para el umbral de gasto a partir de atributos del visitante-viaje

```

Classifier output
=== Run information ===

Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    Mintur_BI_CSV PREPRO_NK_final-weka.filters.unsupervised.attribute.Remove-R1-4,6-8,10,13-14,16,18-27,30-50-weka.filters.unsupervised.attribute.Remove-R3-5
Instances:   3000
Attributes:  5
            Edad D
            Ingreso prom mes hogar
            Tipo Alojaj
            Gto Total D
            Medio transp
Test mode:   10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree
-----

Medio transp = Autobus / flotas: Medio (986.0/456.0)
Medio transp = Vehiculo propio: Medio (1260.0/592.0)
Medio transp = Aereo
| Ingreso prom mes hogar = B de usd 401 a 800
| | Edad D = Centennials: Alto (9.0/5.0)
| | Edad D = Generaci&#x2013;n X: Bajo (7.0/4.0)
| | Edad D = Millennials: Nulo (24.0/11.0)
| | Edad D = Baby boomers: Nulo (0.0)
| | Edad D = Tradicionalistas: Nulo (0.0)
| Ingreso prom mes hogar = A menos de usd 400: Medio (5.0/3.0)
| Ingreso prom mes hogar = E de usd 1501 a 1700: Alto (24.0/9.0)
| Ingreso prom mes hogar = D de usd 1201 a 1500: Alto (18.0/5.0)
| Ingreso prom mes hogar = C de usd 801 a 1200: Alto (21.0/10.0)
| Ingreso prom mes hogar = F Mas de usd 1701: Alto (9.0/3.0)
Medio transp = Vehiculo alquilado
| Ingreso prom mes hogar = B de usd 401 a 800: Medio (246.0/109.0)
| Ingreso prom mes hogar = A menos de usd 400: Medio (86.0/48.0)
| Ingreso prom mes hogar = E de usd 1501 a 1700: Alto (45.0/21.0)
| Ingreso prom mes hogar = D de usd 1201 a 1500: Alto (56.0/23.0)
| Ingreso prom mes hogar = C de usd 801 a 1200: Medio (131.0/77.0)
| Ingreso prom mes hogar = F Mas de usd 1701: Alto (72.0/24.0)
Medio transp = Fluvial: Alto (1.0)

Number of Leaves :    19
Size of the tree :    23
Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      1569           52.3 %
Incorrectly Classified Instances    1431           47.7 %
Kappa statistic                    0.0989
Mean absolute error                 0.3162
Root mean squared error             0.4003
Relative absolute error             96.2121 %
Root relative squared error         98.7564 %
Total Number of Instances          3000

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0.937  0.850  0.523  0.937  0.671  0.142  0.560  0.538  Medio
0.205  0.057  0.546  0.205  0.298  0.220  0.629  0.380  Alto
0.003  0.002  0.200  0.003  0.006  0.011  0.547  0.139  Bajo
0.030  0.008  0.364  0.030  0.055  0.070  0.576  0.172  Nulo
Weighted Avg.  0.523  0.439  0.470  0.523  0.418  0.136  0.576  0.403

=== Confusion Matrix ===

  a  b  c  d  <-- classified as
1402 80  2 12 | a = Medio
587 154  1  8 | b = Alto
325  21  1  1 | c = Bajo
366  27  1 12 | d = Nulo

```

Fuente: Elaboración propia

Cuando el medio de transporte utilizado en el viaje es autobús, el nivel de gasto es medio en 986 ocasiones, donde se incluyen 456 instancias mal clasificadas. Resulta además interesante que cuando los visitantes se transportan por vía aérea, el ingreso promedio mensual de su hogar es de \$401 hasta \$800 y pertenecen a los millennials, 24 viajeros no incurren en gasto alguno por parte del viajero (con 11 errores), probablemente por ser beneficiarios de campañas de reactivación del turismo con viajes totalmente pagados. Sin embargo, tanto los visitantes baby

boomers como tradicionalistas que se transportan vía aérea con ingreso promedio de \$401-\$800, no incurren en gasto alguno.

Dentro del grupo de quienes viajaron por vía aérea y cuyo ingreso promedio mensual del hogar supera los \$1 201, se clasificaron los viajeros que incurrieron en un gasto total de viaje alto como sigue: 18 visitantes de \$1 201-\$1 500 de ingreso promedio mensual con un error del 28%, 24 visitantes para la franja de \$1 501-\$1 700 con un error del 38%; y, 9 visitantes para la franja de más de \$1 701 con un error del 33%.

Para los viajeros que consideraron vehículo alquilado y cuyo ingreso mensual promedio está entre \$401 y \$800, 246 visitantes incurrieron en nivel de gasto medio, pero con 109 instancias mal clasificadas. Por otra parte, para los visitantes de ingreso entre \$1 201 y \$1 500, el nivel de gasto es alto en 56 ocasiones (con 23 errores). Además, los viajeros con ingreso promedio de más de \$1 701, incurren en un nivel de gasto alto en 72 ocasiones (con 24 errores).

Respecto a la predicción de la intención de viaje a futuro en los próximos 3 meses, se realizó una detección previa de variables significativas para la intención futura de viaje de acuerdo a los atributos más relevantes de la combinación visitante - viaje. Así, se analizaron 16 variables entre atributos del visitante y del viaje y se obtuvo el ranking que ordena de mayor a menor la importancia de las variables estudiadas para la variable objetivo, tal como se resume en la figura 38.

Figura 38. Selección de atributos de visitante-viaje con intención futura como variable objetivo

```
Attribute selection output

Search Method:
  Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 17 Viaje en prx 3 meses):
  Information Gain Ranking Filter

Ranked attributes:
0.013309 16 Medios comunic R
0.013162  8 Razon viaje
0.011134  6 Ingreso prom mes hogar
0.007087  5 Nivel estudios
0.006734  7 Miembro decide viaje
0.006561 10 Antelac planif
0.004524 12 Tipo Alojajm
0.003831  9 Region viaje
0.003249  2 Edad D
0.002559 14 Gto Total D
0.001991 15 Medio transp
0.001761  4 Genero
0.001061  3 Estado civil
0.001003  1 Region resid R
0.000584 13 Num acomp D
0         11 Dias pernocta

Selected attributes: 16,8,6,5,7,10,12,9,2,14,15,4,3,1,13,11 : 16
```

Fuente: Elaboración propia

De la lista indicada previamente, se seleccionaron las variables: medios de comunicación por los que se informa para el viaje (recategorizada), motivo de viaje, ingreso promedio mensual del hogar y nivel de estudios para predecir la decisión de volver a viajar en los próximos 3 meses, de acuerdo a lo mostrado en figura 39.

Figura 39. Árbol de decisión para la intención futura a partir de atributos del visitante-viaje

```

Classifier output
=== Run information ===

Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    Mintur BI_CSV PREPRO_Wk_final_Recat & Filter_Viaje prx 3 meses-weka.filters.unsupervised.attribute.Remove-R1-8,11,13-31,33-44,46-52
Instances:   1869
Attributes:  5
              Nivel estudios
              Ingreso prom mes hogar
              Razon viaje
              Medios comunic R
              Viaje en prx 3 meses
Test mode:   10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree
-----

Razon viaje = Vacaciones recreo ocio (turismo)
| Medios comunic R = Redes sociales: Si (448.0/127.0)
| Medios comunic R = Medios clásicos
| | Ingreso prom mes hogar = B de usd 401 a 800: No (190.0/84.0)
| | Ingreso prom mes hogar = E de usd 1501 a 1700: Si (10.0/1.0)
| | Ingreso prom mes hogar = D de usd 1201 a 1500
| | | Nivel estudios = Superior: Si (17.0/6.0)
| | | Nivel estudios = Secundaria: No (10.0/4.0)
| | | Nivel estudios = Primaria: Si (0.0)
| | Ingreso prom mes hogar = C de usd 801 a 1200: No (39.0/19.0)
| | Ingreso prom mes hogar = A menos de usd 400: Si (22.0/6.0)
| | Ingreso prom mes hogar = F Mas de usd 1701: No (4.0/2.0)
| Medios comunic R = Recomendaciones
| | Ingreso prom mes hogar = B de usd 401 a 800: Si (41.0/17.0)
| | Ingreso prom mes hogar = E de usd 1501 a 1700: No (5.0/1.0)
| | Ingreso prom mes hogar = D de usd 1201 a 1500: No (9.0/4.0)
| | Ingreso prom mes hogar = C de usd 801 a 1200
| | | Nivel estudios = Superior: Si (10.0)
| | | Nivel estudios = Secundaria: No (4.0)
| | | Nivel estudios = Primaria: Si (0.0)
| | Ingreso prom mes hogar = A menos de usd 400
| | | Nivel estudios = Superior: No (4.0/1.0)
| | | Nivel estudios = Secundaria: Si (5.0/1.0)
| | | Nivel estudios = Primaria: Si (0.0)
| | Ingreso prom mes hogar = F Mas de usd 1701: Si (11.0/1.0)
| Medios comunic R = Internet: Si (282.0/77.0)
Razon viaje = Visitas a familiares y amigos: Si (388.0/146.0)
Razon viaje = Salud y atencion medica: Si (71.0/26.0)
Razon viaje = Negocios y/o motivos profesionales: Si (168.0/51.0)
Razon viaje = Compras
| Nivel estudios = Superior
| | Medios comunic R = Redes sociales: Si (10.0/2.0)
| | Medios comunic R = Medios clásicos: No (3.0)
| | Medios comunic R = Recomendaciones: Si (1.0)
| | Medios comunic R = Internet: No (1.0)
| Nivel estudios = Secundaria: No (11.0/2.0)
| Nivel estudios = Primaria: No (2.0/1.0)
Razon viaje = Educacion y formacion
| Ingreso prom mes hogar = B de usd 401 a 800
| | Nivel estudios = Superior: Si (14.0/4.0)
| | Nivel estudios = Secundaria: No (18.0/6.0)
| | Nivel estudios = Primaria: No (1.0)
| Ingreso prom mes hogar = E de usd 1501 a 1700: Si (6.0/1.0)
| Ingreso prom mes hogar = D de usd 1201 a 1500: Si (5.0/1.0)
| Ingreso prom mes hogar = C de usd 801 a 1200: No (15.0/2.0)
| Ingreso prom mes hogar = A menos de usd 400: Si (18.0/6.0)
| Ingreso prom mes hogar = F Mas de usd 1701: Si (1.0)
Razon viaje = Religion: No (25.0/5.0)

Number of Leaves :    38

Size of the tree :    49

Time taken to build model: 0.04 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      1198           64.0984 %
Incorrectly Classified Instances     671           35.9016 %
Kappa statistic                     0.1089
Mean absolute error                 0.4457
Root mean squared error             0.4808
Relative absolute error             96.0977 %
Root relative squared error         99.844 %
Total Number of Instances          1869

```

```

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,199	0,105	0,523	0,199	0,288	0,132	0,573	0,443	No
	0,895	0,801	0,660	0,895	0,760	0,132	0,573	0,670	Si
Weighted Avg.	0,641	0,546	0,610	0,641	0,588	0,132	0,573	0,587	

```

=== Confusion Matrix ===

```

a	b	<-- classified as
136	547	a = No
124	1062	b = Si

Fuente: Elaboración propia

Cuando la razón de viaje es vacaciones, recreo u ocio, y las redes sociales el medio por el cual se informaron previo al viaje, 448 visitantes tienen la intención firme de volver a viajar dentro de los próximos 3 meses, incluyendo 127 clasificados erróneamente. Para quienes vacacionan y se informan por medios clásicos (televisión, radio, publicidad vial o señalética, medios impresos) con un ingreso promedio mensual inferior a \$400, 22 viajeros quieren volver a viajar en un futuro cercano, donde se evidencian 6 errores de clasificación. Además, de los visitantes vacacionales que se informaron por medio de recomendaciones de familiares o amigos con hogares de ingreso mensual promedio de entre \$801 y \$1 200 y con educación superior, 10 visitantes indicaron la misma tendencia descrita previamente en cuanto a la finalidad de volver a viajar, sin errores. Para los visitantes vacacionales y que utilizaron el internet como medio para informarse previo al viaje, 282 viajeros mantienen la preferencia de viajar nuevamente en los próximos 3 meses, donde 77 personas se incluyen de forma errónea.

Por otra parte, en los viajes realizados por negocios y/o motivos profesionales, 168 viajeros manifiestan el deseo de volver a viajar próximamente con 51 instancias mal clasificadas. Cuando el viaje se realiza por motivos religiosos, 25 viajeros no tienen intenciones de viajar nuevamente en un futuro cercano, 5 de ellos clasificándose de forma equívoca.

6.3. Análisis del destino

El análisis del destino implicó el involucramiento del componente geográfico por lo que, para la predicción de la región de destino, se realizó una detección previa de variables significativas de acuerdo a los atributos más relevantes del visitante (región de residencia <recategorizada>, edad discretizada, estado civil, género, nivel de estudios e ingreso promedio mensual del hogar) y del viaje (motivo de viaje y medios de comunicación para informarse antes del viaje <recategorizada>).

Así, se analizaron 8 variables entre atributos del visitante y del viaje y se obtuvo el ranking que ordena de mayor a menor la importancia de las variables estudiadas para la variable objetivo, tal como se resume en la figura 40.

Figura 40. Selección de atributos de visitante-viaje con región de destino como variable objetivo

```

Attribute selection output
Evaluation mode: evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:
  Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 8 Region viaje):
  Information Gain Ranking Filter

Ranked attributes:
0.05263  7 Razon viaje
0.03021  6 Ingreso prom mes hogar
0.02314  2 Edad D
0.02261  9 Medios comunic Recat
0.01888  5 Nivel estudios
0.01229  1 Region resid Recat
0.01212  3 Estado civil
0.00238  4 Genero

Selected attributes: 7,6,2,9,5,1,3,4 : 8
  
```

Fuente: Elaboración propia

De la lista indicada previamente, se seleccionaron las variables: motivo de viaje, ingreso promedio mensual del hogar y edad discretizada para predecir la región a la que se elige viajar, de acuerdo a lo mostrado en figura 41.

Figura 41. Árbol de decisión para la región de destino a partir de atributos del visitante-viaje

```

Classifier output

=== Run information ===

Scheme: weka.classifiers.cress.J48 -C 0.25 -M 2
Relation: Mintur_BI_CSV_PREPRO_Wk_final_Recat-weka.filters.unsupervised.attribute.Remove-R1,3-5,11,14-31,33-52-weka.filters.unsupervised.attribute.Remove
Instances: 3000
Attributes: 4
  Edad D
  Ingreso prom mes hogar
  Razon viaje
  Region viaje
Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree
-----
Razon viaje = Vacaciones recreo ocio (turismo): Costa (1608.0/798.0)
Razon viaje = Religion
| Ingreso prom mes hogar = B de usd 401 a 800: Costa (47.0/28.0)
| Ingreso prom mes hogar = A menos de usd 400: Sierra (3.0/1.0)
| Ingreso prom mes hogar = E de usd 1501 a 1700: Costa (3.0/1.0)
| Ingreso prom mes hogar = D de usd 1201 a 1500: Amazonia (8.0/3.0)
| Ingreso prom mes hogar = C de usd 801 a 1200
| | Edad D = Centennials: Sierra (3.0)
| | Edad D = Generación X: Costa (10.0/6.0)
| | Edad D = Millennials: Amazonia (6.0/1.0)
| | Edad D = Baby boomers: Costa (2.0)
| | Edad D = Tradicionalistas: Amazonia (0.0)
| Ingreso prom mes hogar = F Mas de usd 1701: Sierra (0.0)
Razon viaje = Visitas a familiares y amigos: Sierra (568.0/232.0)
Razon viaje = Salud y atencion medica: Sierra (169.0/87.0)
Razon viaje = Negocios y/o motivos profesionales
| Ingreso prom mes hogar = B de usd 401 a 800: Sierra (104.0/16.0)
| Ingreso prom mes hogar = A menos de usd 400
| | Edad D = Centennials: Sierra (2.0)
| | Edad D = Generación X: Costa (3.0/1.0)
| | Edad D = Millennials: Costa (6.0/2.0)
| | Edad D = Baby boomers: Costa (0.0)
| | Edad D = Tradicionalistas: Costa (0.0)
| Ingreso prom mes hogar = E de usd 1501 a 1700: Galapagos (8.0/5.0)
| Ingreso prom mes hogar = D de usd 1201 a 1500
  
```



```

| | Edad D = Centennials: Sierra (3.0/1.0)
| | Edad D = Generación X: Costa (6.0/1.0)
| | Edad D = Millennials: Sierra (9.0/5.0)
| | Edad D = Baby boomers: Costa (1.0)
| | Edad D = Tradicionalistas: Costa (0.0)
| Ingreso prom mes hogar = C de usd 801 a 1200: Costa (52.0/30.0)
| Ingreso prom mes hogar = F Mas de usd 1701: Costa (18.0/7.0)
Razon viaje = Compras
| Ingreso prom mes hogar = B de usd 401 a 800: Costa (24.0/7.0)
| Ingreso prom mes hogar = A menos de usd 400: Sierra (10.0/2.0)
| Ingreso prom mes hogar = E de usd 1501 a 1700: Sierra (5.0/1.0)
| Ingreso prom mes hogar = D de usd 1201 a 1500
| | Edad D = Centennials: Amazonia (3.0/1.0)
| | Edad D = Generación X: Amazonia (5.0/2.0)
| | Edad D = Millennials: Costa (3.0)
| | Edad D = Baby boomers: Amazonia (0.0)
| | Edad D = Tradicionalistas: Amazonia (0.0)
| Ingreso prom mes hogar = C de usd 801 a 1200
| | Edad D = Centennials: Costa (3.0)
| | Edad D = Generación X: Costa (11.0/4.0)
| | Edad D = Millennials: Sierra (15.0/7.0)
| | Edad D = Baby boomers: Galapagos (1.0)
| | Edad D = Tradicionalistas: Costa (0.0)
| Ingreso prom mes hogar = F Mas de usd 1701: Costa (0.0)
Razon viaje = Educacion y formacion
| Ingreso prom mes hogar = B de usd 401 a 800: Costa (124.0/71.0)
| Ingreso prom mes hogar = A menos de usd 400: Sierra (37.0/22.0)
| Ingreso prom mes hogar = E de usd 1501 a 1700
| | Edad D = Centennials: Costa (4.0/2.0)
| | Edad D = Generación X: Costa (4.0/1.0)
| | Edad D = Millennials: Sierra (8.0/3.0)
| | Edad D = Baby boomers: Amazonia (3.0/1.0)
| | Edad D = Tradicionalistas: Costa (0.0)
| Ingreso prom mes hogar = D de usd 1201 a 1500: Costa (25.0/14.0)
| Ingreso prom mes hogar = C de usd 801 a 1200
| | Edad D = Centennials: Amazonia (10.0/4.0)
| | Edad D = Generación X: Amazonia (18.0/10.0)
| | Edad D = Millennials: Sierra (35.0/20.0)
| | Edad D = Baby boomers: Costa (12.0/7.0)
| | Edad D = Tradicionalistas: Amazonia (0.0)
| Ingreso prom mes hogar = F Mas de usd 1701: Sierra (1.0)

Number of Leaves : 55
Size of the tree : 67
Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances 1544 51.4667 %
Incorrectly Classified Instances 1456 48.5333 %
Kappa statistic 0.1712
Mean absolute error 0.2982
Root mean squared error 0.3893
Relative absolute error 94.3006 %
Root relative squared error 97.9184 %
Total Number of Instances 3000

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRG Area Class
0,781 0,580 0,490 0,781 0,602 0,210 0,613 0,493 Costa
0,431 0,241 0,570 0,431 0,491 0,201 0,594 0,537 Sierra
0,000 0,001 0,000 0,000 0,000 -0,009 0,566 0,074 Galapagos
0,063 0,008 0,463 0,063 0,111 0,142 0,611 0,189 Amazonia
Weighted Avg. 0,515 0,345 0,493 0,515 0,471 0,187 0,602 0,457

=== Confusion Matrix ===
a b c d <-- classified as
975 264 1 9 | a = Costa
712 550 2 12 | b = Sierra
122 50 0 1 | c = Galapagos
181 101 1 19 | d = Amazonia

```

Fuente: Elaboración propia

Para el caso de quienes viajan con el fin de visitar a familiares y amigos, la región del destino es Sierra en 568 ocasiones, incluyendo 232 instancias clasificadas de forma incorrecta. Por otra parte, de los viajeros que se desplazan por negocios y/o motivos profesionales y que además generan un ingreso promedio mensual entre \$401 y \$800, 104 personas deciden viajar a la Sierra, de las cuales 88 se clasificaron correctamente. Del grupo de visitantes por negocios con ingresos mensuales no superiores a \$400, tanto baby boomers como tradicionalistas, ninguno elige a la Costa como región de destino.

Cuando la razón del viaje es hacer compras y el ingreso promedio mensual del hogar oscila entre \$401 y \$800, 24 visitantes optan por viajar a la Costa incluyendo 7 instancias mal

clasificadas. De los viajeros tradicionalistas que se desplazan por compras y cuentan con ingresos mensuales en promedio de \$801 a \$1 200, no existe viajero que visite la región Costa como destino.

Dado que la muestra estuvo mayormente conformada por visitantes vacacionales que eligieron destinos dentro de las regiones de Costa y Sierra, se planteó ejecutar predicciones de la región de destino únicamente considerando la submuestra de visitantes vacacionales. Por lo tanto, se consideraron las mismas 8 variables ya mencionadas en la predicción de la muestra completa, prescindiendo de la variable motivo de viaje, dado que únicamente se trabajó con viajeros vacacionales. Así, se analizaron 7 variables entre atributos del visitante y del viaje y se obtuvo el ranking que ordena de mayor a menor la importancia de las variables estudiadas para la variable objetivo, tal como se resume en la figura 42.

Figura 42. Selección de atributos de visitante-viaje con región de destino como variable objetivo de viajeros vacacionales

```
Attribute selection output
  Nivel estudios
  Ingreso prom mes hogar
  Region viaje
  Medios comunic Recat
Evaluation mode:  evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:
  Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 7 Region viaje):
  Information Gain Ranking Filter

Ranked attributes:
0.01517  5 Nivel estudios
0.01254  6 Ingreso prom mes hogar
0.00591  4 Genero
0.00549  1 Region resid Recat
0.00544  2 Edad D
0.00523  8 Medios comunic Recat
0.00432  3 Estado civil

Selected attributes: 5,6,4,1,2,8,3 : 7
```

Fuente: Elaboración propia

De la lista indicada previamente, se seleccionaron las variables: nivel de estudios, ingreso promedio mensual del hogar, género, región de residencia recategorizada y edad discretizada para predecir si los viajeros deciden viajar a región Costa o a Sierra, de acuerdo a lo mostrado en figura 43.

Figura 43. Árbol de decisión para la región de destino a partir de atributos del visitante-viaje de viajeros vacacionales

```

=== Run information ===

Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    Mintur BI_CSV PREPRO_Wk_final_Vacay_Reg destino CostaSierra_Recat-weka.filters.unsupervised.attribute.Remove-R1,3-5,11-12,14-31,33-52-weka.fil
Instances:   1384
Attributes:  6
             Region resid Recat
             Edad D
             Genero
             Nivel estudios
             Ingreso prom mes hogar
             Region viaje
Test mode:   10-fold cross-validation

=== Classifier model (full training set) ===
J48 pruned tree
-----

Nivel estudios = Superior: Costa (870.0/313.0)
Nivel estudios = Secundaria
| Region resid Recat = SIERRA
| | Genero = Femenino: Costa (92.0/27.0)
| | Genero = Masculino
| | | Edad D = Centennials: Costa (38.0/14.0)
| | | Edad D = Generación X: Costa (25.0/7.0)
| | | Edad D = Millennials: Sierra (99.0/42.0)
| | | Edad D = Baby boomers: Sierra (6.0/2.0)
| | | Edad D = Tradicionalistas: Costa (0.0)
| | Genero = Prefiero no responder: Sierra (1.0)
| Region resid Recat = COSTA
| | Ingreso prom mes hogar = B de usd 401 a 800
| | | Edad D = Centennials: Sierra (62.0/21.0)
| | | Edad D = Generación X: Costa (22.0/9.0)
| | | Edad D = Millennials: Sierra (76.0/25.0)
| | | Edad D = Baby boomers: Costa (1.0)
| | | Edad D = Tradicionalistas: Sierra (0.0)
| | Ingreso prom mes hogar = A menos de usd 400: Sierra (30.0/6.0)
| | Ingreso prom mes hogar = E de usd 1501 a 1700: Sierra (5.0/1.0)
| | Ingreso prom mes hogar = C de usd 801 a 1200
| | | Edad D = Centennials: Costa (1.0)
| | | Edad D = Generación X
| | | | Genero = Femenino: Sierra (3.0)
| | | | Genero = Masculino: Costa (10.0/2.0)
| | | | Genero = Prefiero no responder: Costa (0.0)
| | | Edad D = Millennials: Costa (7.0/3.0)
| | | Edad D = Baby boomers: Sierra (2.0)
| | | Edad D = Tradicionalistas: Costa (0.0)
| | Ingreso prom mes hogar = D de usd 1201 a 1500
| | | Genero = Femenino: Sierra (2.0)
| | | Genero = Masculino: Costa (4.0)
| | | Genero = Prefiero no responder: Costa (2.0)
| | Ingreso prom mes hogar = F Mas de usd 1701: Costa (1.0)
| Region resid Recat = AMAZONIA: Costa (14.0/4.0)
Nivel estudios = Primaria
| Edad D = Centennials: Sierra (1.0)
| Edad D = Generación X: Costa (3.0)
| Edad D = Millennials: Sierra (6.0/2.0)
| Edad D = Baby boomers: Sierra (1.0)
| Edad D = Tradicionalistas: Sierra (0.0)

Number of Leaves :    32
Size of the tree :    42

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      886          64.0173 %
Incorrectly Classified Instances    498          35.9827 %
Kappa statistic                    0.2038
Mean absolute error                 0.4564
Root mean squared error             0.4821
Relative absolute error             94.0114 %
Root relative squared error         97.8526 %
Total Number of Instances          1384

=== Detailed Accuracy By Class ===

              TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0,860  0,671  0,644  0,860  0,737  0,226  0,588  0,635  Costa
0,329  0,140  0,626  0,329  0,432  0,226  0,588  0,523  Sierra
Weighted Avg.  0,640  0,450  0,637  0,640  0,610  0,226  0,588  0,589

=== Confusion Matrix ===

  a  b  <-- classified as
697 113 |  a = Costa
385 189 |  b = Sierra

```

Fuente: Elaboración propia

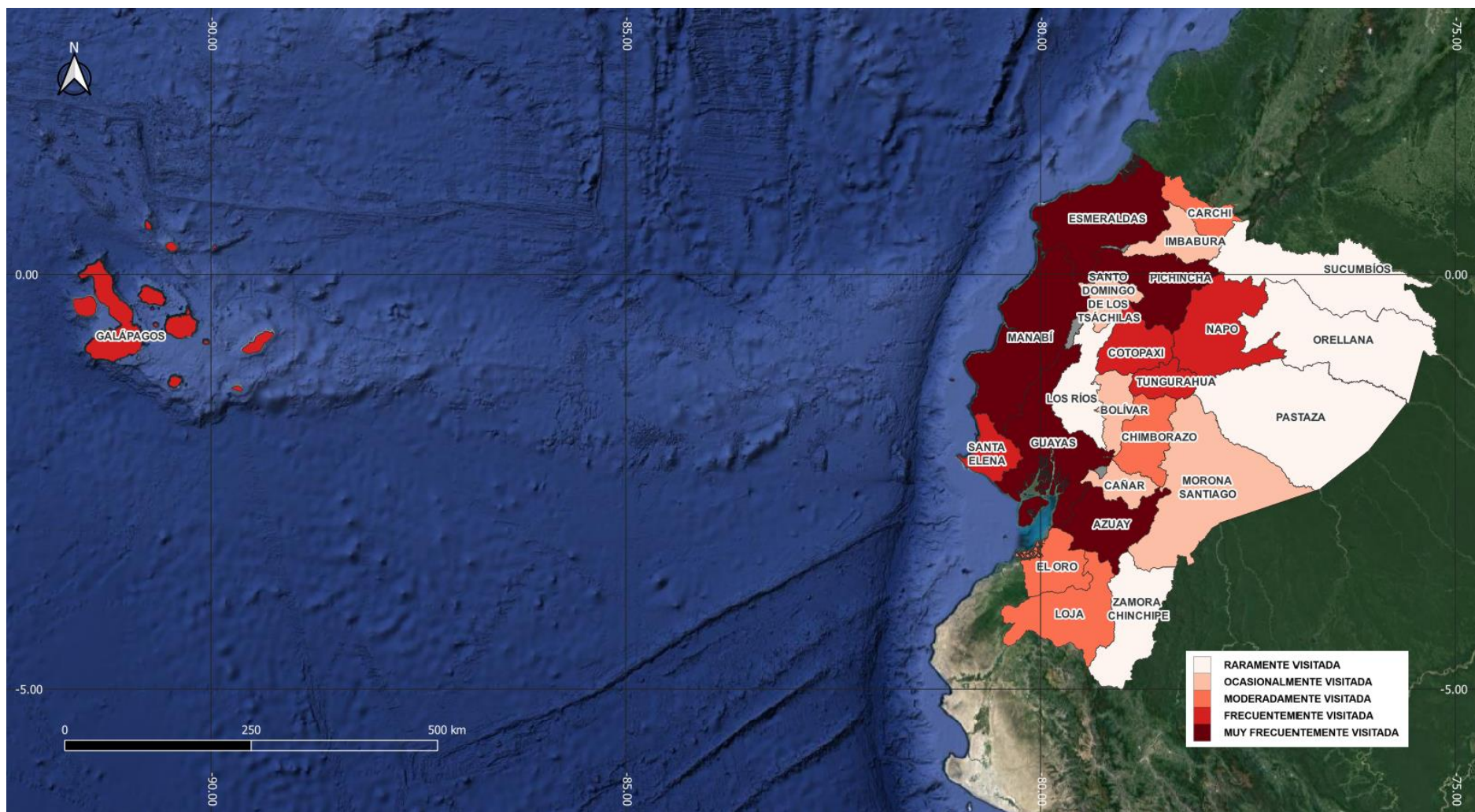
Considerando al grupo de visitantes que se desplazaron hasta las regiones Costa y Sierra como destino para vacacionar, los que tienen estudios de nivel superior, se desplazaron mayormente

a la Costa en 870 ocasiones, incluyendo 313 errores de clasificación. Mientras, de los viajeros que tienen estudios secundarios, residentes de la región Sierra y de género femenino, 92 eligen a la región Costa como destino vacacional, de los cuales 65 son clasificados correctamente; sin embargo, de quienes pertenecen al género masculino y a la generación X, poseen estudios secundarios y residen en la región Andina, 25 también prefieren la región Litoral (considerando 7 desaciertos en la clasificación).

De los visitantes con estudios secundarios, residentes en la región Costa e ingreso promedio mensual desde \$401 hasta \$800 pertenecientes a los centennials, 62 viajeros prefieren a la Sierra como destino (registrando 21 errores). Los millennials residentes en la Costa con estudios secundarios e ingresos mensuales entre \$401 y \$800 en promedio, eligen a la región Andina en 76 ocasiones (registrando 25 errores). Del grupo de visitantes con nivel secundario de educación, costeños y con ingresos mensuales inferiores a \$400, 30 eligen a la región Andina como destino para vacacionar, incurriendo en un error de clasificación del 20%.

En cuanto a la visualización de patrones geográficos considerando las provincias de destino en los viajes muestreados, de acuerdo a la figura 44, en su mayoría las provincias de la región Costa fueron las más visitadas, excepto por Pichincha y Azuay. Por su parte, Galápagos, Santa Elena, Napo, Cotopaxi y Tungurahua fueron frecuentemente visitadas. A más de la frecuencia de visita, se representó el promedio de pernoctaciones, evidenciando que ningún viaje en el territorio ecuatoriano superó las 3 noches de duración de la estancia, tal como se resume en la figura 45. Resultó interesante comprobar que prácticamente todas las provincias de las regiones litoral y andina registraron viajes con 3 noches de pernoctación en promedio; sin embargo, Guayas, aunque fue una de las provincias más visitadas, registró la mayoría de viajes alrededor de 2 noches pernoctadas. Finalmente, los viajes más costosos en promedio se asociaron nuevamente a provincias costeras como Esmeraldas, Manabí, Guayas y Santa Elena, a más de Galápagos, que es el destino más costoso del país por excelencia, condensando estos hallazgos en la figura 46. A más de Pichincha y Azuay, como las provincias de mayor frecuencia de visitación, estas también se encuentran dentro del segundo grupo de destinos que generaron viajes más costosos, y a ellas se unen Imbabura, Sucumbíos y Orellana.

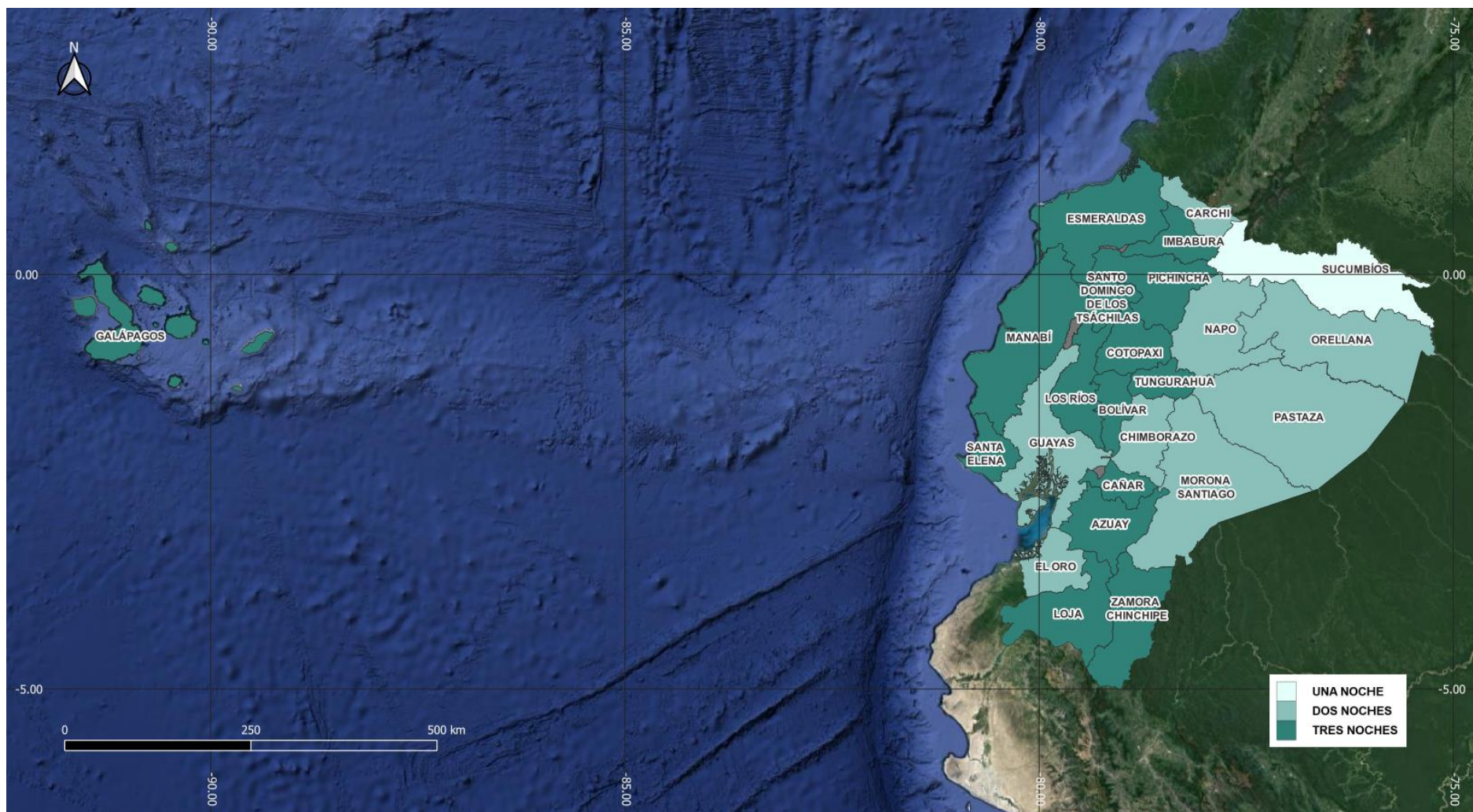
Figura 44. Mapa de Frecuencia de Visita por provincias



Fuente: Elaboración propia

**Los territorios sombreados en gris están actualmente en disputa y se consideran como zonas no delimitadas.*

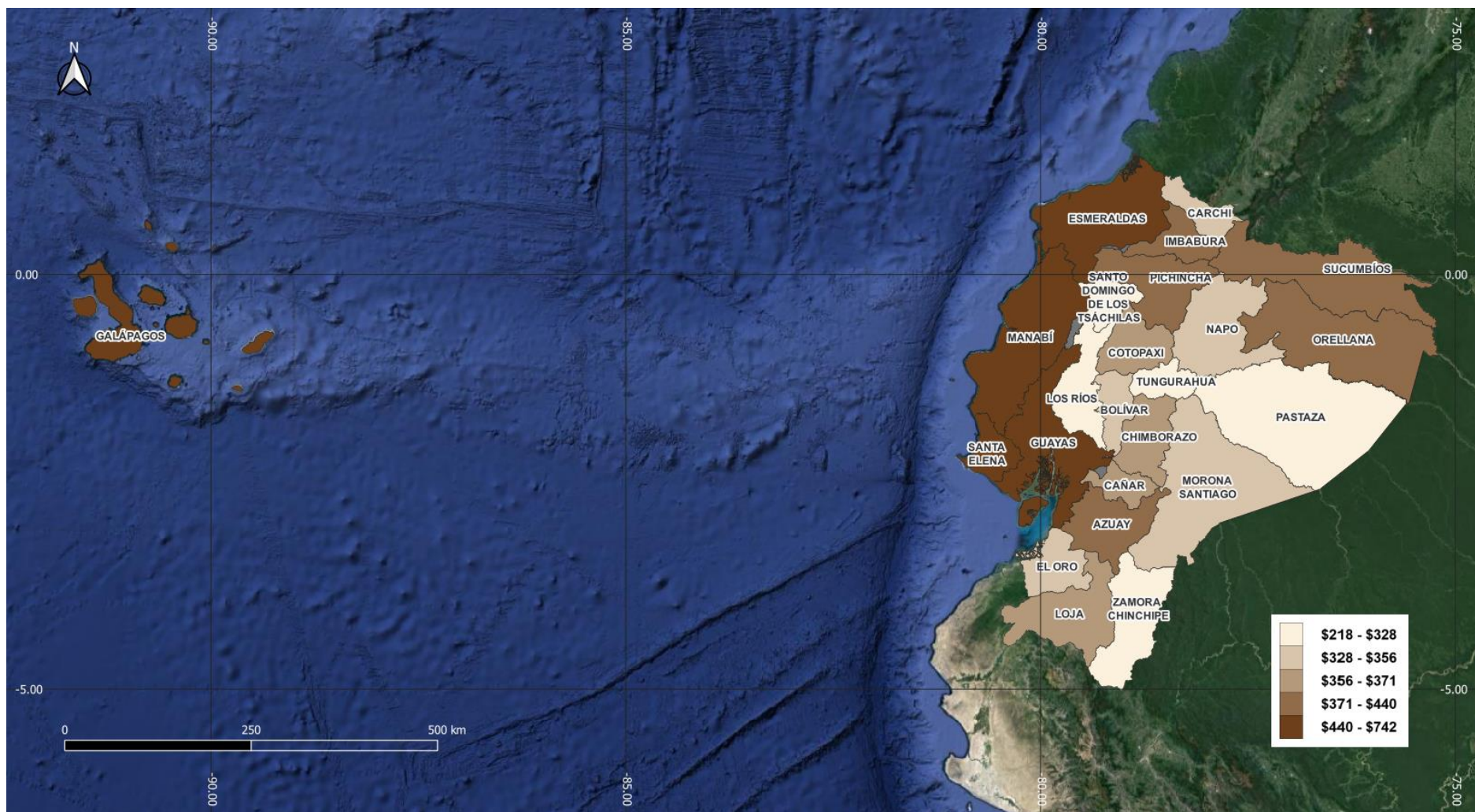
Figura 45. Mapa de Pernoctaciones Promedio por provincias



Fuente: Elaboración propia

**Los territorios sombreados en gris están actualmente en disputa y se consideran como zonas no delimitadas.*

Figura 46. Mapa de Gasto Total Promedio por provincias



Fuente: Elaboración propia

**Los territorios sombreados en gris están actualmente en disputa y se consideran como zonas no delimitadas.*

6.4. Discusión

Antes de la pandemia ocasionada por el SARS-CoV-2 (del inglés que refiere al coronavirus de tipo 2 causante del Síndrome Respiratorio Agudo Severo), el turismo era la tercera fuerza en las industrias de exportaciones no petroleras y mineras, después del camarón y el banano en Ecuador, y de gran aporte para el PIB nacional. En 2021, cayó a la sexta posición en las exportaciones después de las flores y el cacao, siendo responsable en al menos de un 10% de la caída del PIB (Ministerio de Turismo, 2021). Una mirada retrospectiva a las métricas del 2022, evidencia que la contribución al PIB nacional por parte del sector de viajes y turismo creció un 43.3% en comparación con 2021, lo que representó el 4.2% de la economía ecuatoriana. Además, se generaron más de 390 000 empleos, es decir, casi el 5% del total de puestos de trabajo en el país. Es pertinente mencionar que durante 2022, la llegada de viajeros internacionales generó un crecimiento del 60.9% frente a 2021 mientras que los desplazamientos de viajeros domésticos representaron un aumento de 32.9% por encima de lo registrado en 2021 y tan sólo un 0.5% por debajo de los niveles pre pandemia en 2019 (WTTC, 2023).

En la actualidad, se espera que la recuperación que se ha evidenciado a raíz de la relajación de medidas y levantamiento de restricciones en conjunto con otras estrategias específicas, permita que el turismo aporte de la misma forma que lo hizo antes de la crisis por la COVID-19. Algunas de las referidas estrategias incluyen otorgamiento de créditos al 5% para emprendedores y empresarios, eliminación del Impuesto a la Salida de Divisas (ISD), reducción de tasas Eco Delta y Potencia Turística en precios de pasajes aéreos, reducción del Impuesto al Valor Agregado (IVA) para actividades turísticas durante feriados nacionales, entre otras (Baker Tilly Ecuador, 2023; Diario El Universo, 2023a; Ministerio de Turismo, 2023c). Sin embargo, el sector todavía tiene otras crisis por enfrentar, concretamente la económica que se ha recrudecido en estos últimos años, la social por la inseguridad ciudadana con elevados índices de violencia y criminalidad, y la ambiental debido a la afectación causada por fenómenos naturales como El Niño junto con todo el impacto que dejó la época lluviosa de 2023 (Diario El Universo, 2023b; Prensa Latina, 2023; Primicias, 2023).

Aunque desde la administración pública se han gestionado esfuerzos, en la práctica resultan insuficientes dado que el sector carece, en parte, de inversión, financiamiento, planificación y gestión de políticas públicas que permitan interiorizar la necesidad de generación de inteligencia turística (Ministerio de Turismo, 2020). Sin duda, la inteligencia adquirida a través de los datos permite una mejor respuesta del sector ante cualquier crisis, pero también direcciona los esfuerzos hacia la innovación, tecnificación, eficiencia y desarrollo de ventajas

competitivas, aprovechándolos para que la industria turística se torne más responsable, sostenible y sobre todo resiliente (Ioannides & Gyimóthy, 2020; Pulido-Fernández, 2020). La literatura científica ampliamente revisada en el presente trabajo pone de manifiesto toda la revolución que ha experimentado el turismo y cómo esto ha cambiado la forma de comportarse y comunicarse del visitante, propiciando una transformación radical en el ciclo del viaje tradicional de la demanda, por lo que la gestión de destinos en la última década se orienta hacia modelos inteligentes que aporten al cumplimiento de los Objetivos de Desarrollo Sostenible (ODS) de la Agenda 2030 (López-Sánchez & Pulido-Fernández, 2016; Puig-Cabrera & Foronda-Robles, 2018; Pulido-Fernández, 2021).

En este sentido, aunque Ecuador no ha desarrollado a la fecha un sistema de inteligencia turística para, por ejemplo, analizar la demanda turística interna que es motivo de este trabajo, la autoridad estatal en materia de turismo y la academia sí han ejecutado unos estudios que abordan el análisis desde un alcance principalmente descriptivo como convencionalmente se acostumbra (MINDTEK, 2015a; Ministerio de Turismo, 2011, 2012, 2013). Estos estudios previos permiten comparar ciertos hallazgos con algunos de los resultados empíricos obtenidos aquí. En cuanto al perfil socioeconómico del visitante, la demanda doméstica coincide con el estudio del año 2021 en constituirse mayormente por hombres y mujeres centennials y millennials (18 a 41 años), priorizando opciones cercanas al lugar de residencia para excursiones y viajes acompañados por familiares y amigos con gran predisposición para volver a viajar a corto plazo (Molina Velásquez et al., 2022).

Respecto a variables vinculadas al viaje, el gasto total promedio del viaje es el más alto registrado (alrededor de \$415) cuando los estudios mostraron \$131 en 2012 y \$191 en 2015, respectivamente. Esto considerando que la estadía se ha mantenido en un promedio de 2 a 3 noches, priorizando los viajes durante fines de semana y feriados principalmente. El medio de transporte preferido mayormente es el vehículo propio seguido por el autobús, tendencia que se evidenció desde 2015, muy probablemente por el incremento del parque automotor en el país y la sensación de mayor control sobre seguridad a cargo de los mismos visitantes en vez de terceros. Actualmente, el tipo de alojamiento más utilizado son los hoteles dado que la motivación principal es vacacionar, sin embargo, en 2011, el alojamiento más frecuentemente utilizado eran viviendas de familiares o amigos, justamente porque la motivación principal de viaje era visitarlos a ellos. Por otra parte, el tiempo de antelación con que se planifica el viaje sigue siendo de hasta una semana, esto en correspondencia a que un viaje de corta duración no requiere mucho tiempo para planificarse. Los medios que influyen mayormente el viaje actualmente son las redes sociales, dada su rápida evolución en la última década desde su auge

a partir del año 2000, en contraste a lo reportado en 2011 donde la experiencia de una visita anterior era lo que primaba (INEC, 2016; MINDTEK, 2015a; Ministerio de Turismo, 2012, 2013; Molina Velásquez et al., 2022).

Finalmente, se indican los 5 destinos más visitados en la actualidad: Guayas, Manabí, Pichincha, Esmeraldas y Azuay, mientras que en el año de 2021 fueron: Esmeraldas, Manabí, Pichincha, Imbabura y Tungurahua. De estas provincias, 3 se mantienen en la tendencia post pandemia, aunque las diferencias podrían deberse a la selección de la muestra en cuanto a la consideración del lugar de residencia, queda de manifiesto la preferencia por desplazamientos de proximidad a nivel regional, principalmente entre Costa y Sierra, relegando a la Amazonía y a la región Insular (Molina Velásquez et al., 2022).

En cuanto a la comprensión en profundidad de los resultados empíricos encontrados, no es posible comparar con otros hallazgos dado que no se han gestionado estudios o análisis que promuevan la generación de inteligencia turística a partir de los datos, desde el enfoque del aprendizaje automático aquí abordado. Esto debido a que no existen referencias previas en cuanto a este tipo de analítica en el contexto nacional. A pesar de ello, se procede a enfatizar las aplicaciones prácticas de dichos resultados en la planificación y gestión del destino. En primer lugar, los patrones más frecuentes encontrados en las características del visitante y del viaje permiten perfilar el comportamiento multivariante del turismo interno, evaluar diferencias entre regiones Costa y Sierra como destino, entrenar sistemas de recomendación, seleccionar los medios más idóneos para promocionar los destinos e identificar puntos críticos dentro de la preparación y desarrollo del viaje para el sector de la intermediación y operación de viajes. En segundo lugar, los segmentos de viajes identificados facultan la generación de campañas efectivas de marketing dirigido de acuerdo a alguna variable de interés (por ejemplo, generaciones según edad), transparentan la estructura del gasto del viaje en el contexto de las principales actividades turísticas (alojamiento, transporte, alimentos y bebidas) y exhiben comprensión significativa de tendencias en pernoctaciones, gasto total y tamaño de grupos. Adicionalmente, a través de los árboles de decisión generados y sus respectivas reglas de clasificación se puede predecir el comportamiento de los visitantes en cuanto al umbral del gasto total del viaje, la región de destino de viaje y la intención de viaje a futuro considerando únicamente las variables clave que aportan a la precisión de los modelos. Estas predicciones pueden aportar significativamente a la fidelización de visitantes y captación de nuevos. Finalmente, la vinculación de patrones geográficos con variables como frecuencia de visita, pernoctaciones promedio y gasto total promedio de viaje, permiten la personalización de la oferta de las diversas provincias, la fijación de precios competitivos, la identificación de

destinos de tránsito, el planteamiento de estrategias para promoción en flujos próximos y la propuesta de incentivos para desplazamientos más largos, por ejemplo, en destinos amazónicos o insulares.

La metodología empleada sin duda abre la posibilidad para la construcción de un marco de referencia hacia un monitoreo continuo de la demanda turística interna, considerando las variables del visitante, viaje y destino, sobre todo con las técnicas predictivas de variables de interés y su vinculación con el componente geográfico. No obstante, entre las principales limitaciones del trabajo, se encuentra la vigencia de los resultados interpretados fuera del período analizado (octubre 2022). Además, se han priorizado ciertas ciudades para el marco muestral; sin embargo, se podría expandir hacia todo el territorio nacional para validar si los resultados obtenidos realmente se mantienen en el contexto de la demanda turística interna. De la misma forma, otra restricción a considerar es que no se incluyeron variables para medir las actividades y las tipologías de turismo efectuadas en el viaje o excursión como se lo hacía en estudios piloto del Ministerio de Turismo hasta 2012 para obtener una fotografía completa del turismo interno.

Por otra parte, la metodología model-data-driven aplicada requiere procesos de parametrización para la obtención de conclusiones ad-hoc. En este sentido, con la rapidez que se pueden generar y recopilar datos, la principal barrera para su aplicación es la falta de automatización de este tipo de metodología. Por consiguiente, la futura línea de investigación se decanta hacia el traslado de esta metodología a un formato en tiempo real soportado en un Sistema de Inteligencia Turística que constantemente monitoree patrones, clústeres y predicciones como soporte a la toma de decisiones y que alerte de forma temprana cuando detecte cambios versus las tendencias históricas.

No obstante, la adaptación sugerida no es tarea sencilla pues requiere de profesionales altamente capacitados, infraestructura idónea y financiación adecuada. Respecto a la infraestructura, por ejemplo, el desarrollo de la red 5G es una prioridad absoluta para los destinos ya que sin ella es imposible gestionar en tiempo real la interacción entre agentes (Katsaros et al., 2019). Se prevé que en Ecuador se comience a comercializar la nueva red en 2025, siendo, junto con Paraguay, los últimos países en ofrecerla en Sudamérica. Pese a las pruebas de conexión hechas hace 3 años y anunciadas para este 2023, el despliegue de la tecnología aún no se cristaliza dada la lentitud para negociar y concesionar el espectro radioeléctrico entre el Gobierno y las firmas telefónicas (Diario Expreso, 2023). De la misma forma, se requieren profesionales especializados en materias STEM y la decisión por apostarle a la transformación digital de las

administraciones públicas (municipales, provinciales y nacionales) de los destinos (Moreno-Izquierdo et al., 2022), priorizando inversiones para este objetivo.

En consecuencia, considerando que el turismo es uno de los sectores que tiene gran capacidad de generar datos heterogéneos y a disposición de todos los actores clave dentro de la cadena de valor, es perfectamente posible colocar los cimientos para generar inteligencia turística incorporando un mayor uso de tecnologías (Moreno-Izquierdo et al., 2022). En el caso particular del Ecuador, la transformación tecnológica es una tarea pendiente de acuerdo al Índice de Desarrollo de Viajes y Turismo 2021. Este índice, publicado por el Foro Económico Mundial, ranquea a 117 países en 17 pilares y la disponibilidad para las TIC es uno de los pilares donde el país posee una desventaja competitiva (World Economic Forum, 2022). Por el contrario, posee una posición privilegiada frente a otros países en cuanto a los pilares de recursos naturales y culturales donde destaca ampliamente, siendo uno de los países más megadiversos del mundo, tal como lo proyecta la más reciente campaña promocional del Gobierno nacional “Time to reset in Ecuador” (Tiempo de reiniciar en Ecuador) (Diario El Universo, 2023a; World Economic Forum, 2022).

6.5. Propuesta de acciones

A la luz de todo lo analizado, ha quedado claro que el Ecuador necesita sostenerse en el turismo interno para promover los viajes dentro del territorio y dinamizar el sector para su reactivación, conforme se encamina a la recuperación a propósito del fin de la pandemia por COVID-19. En este contexto, el aprovechamiento de los datos que se pueden recopilar y analizar con miras a establecer los cimientos de inteligencia turística en el país es vital, por lo que se proponen las siguientes líneas de acción a corto plazo para direccionar los esfuerzos desde el Ministerio de Turismo del país hasta los actores clave del sector turístico y otros vinculados como la academia y la sociedad civil:

- Programar levantamiento de datos periódicamente, de preferencia con frecuencia semestral, replicando el cuestionario utilizado en el presente trabajo para mantener el monitoreo a largo plazo de atributos del visitante, viaje y destino. Incluso se sugiere incorporar otras variables de interés como las actividades turísticas y las tipologías de turismo practicadas.
- Incrementar la cantidad de encuestados para elevar esta problemática de analítica de datos hacia un contexto Big Data, de forma que las técnicas de aprendizaje automático tengan una base de datos para entrenarse mejor.

- Articular la colaboración interinstitucional para gestionar los procesos de recopilación, levantamiento, preprocesamiento y procesamiento de datos sin que estos representen un impacto en el presupuesto anual del Ministerio de Turismo.
- Comunicar los resultados de forma sencilla y breve para que todos los actores clave de la cadena de valor puedan sacar el máximo provecho de la inteligencia que se provea para procesos de planificación y gestión, a más de promoción, comercialización y creación/diseño de la oferta dirigida para el visitante interno, sea excursionista o turista.

A largo plazo, se propone retomar formalmente los observatorios turísticos en las 7 zonas restantes de planificación territorial en el país. Con esto, se podrá implementar un Sistema de Inteligencia Turística piloto, contemplando todos los requerimientos necesarios con el fin de migrar de a poco a la meta de convertirnos en un Destino Turístico Inteligente.

CAPÍTULO VII. CONCLUSIONES

- Se logró profundizar en la comprensión de la demanda turística interna del Ecuador mediante el uso de técnicas avanzadas de ciencia de datos y gestión de los cimientos de inteligencia turística en torno al visitante, viaje y destino a través de la aplicación de técnicas descriptivas y predictivas, la interpretación de los resultados analíticos más relevantes y la propuesta de acciones que sirvan de soporte al proceso de toma de decisiones para una mejor planificación y gestión del turismo doméstico en el país.
- Se pudieron comprobar las hipótesis planteadas mediante la metodología model-data-driven aplicada y la parametrización para la obtención de resultados ad-hoc mediante técnicas de asociación, segmentación, selección de atributos y clasificación. Los patrones extraídos, los segmentos identificados, las predicciones generadas y las tendencias geográficas visibilizadas gracias al Aprendizaje Automático y los Sistemas de Información Geográfica tienen aplicación práctica en diversas áreas de planificación y gestión como sistemas de recomendación, promoción de destinos, campañas de marketing dirigido, fidelización de visitantes, personalización de la oferta, entre las más destacadas.
- Desde la perspectiva analítica, la selección de características adecuadas es vital para la columna vertebral de cualquier modelo de Aprendizaje Automático. Únicamente, eligiendo las variables más significativas, se obtienen modelos más precisos y más fácilmente interpretables. Esto tiene más trascendencia, más allá de lo analítico, en el ámbito del turismo, dado que se puede tener toda la información que sea, pero precisamente por el exceso de información, hay que recurrir a estas técnicas. De cara a tomar decisiones, es fundamental que los gestores turísticos sepan que existen técnicas a las cuales recurrir para reducir dimensionalidad.
- En Ecuador, es una realidad que el Ministerio de Turismo del país en coordinación con el Gobierno Nacional ha iniciado el camino hacia la digitalización e incorporación de tecnología dentro del sector sobre todo con miras a la realización de análisis basados en modelos analíticos de la moderna Ciencia de los Datos. Sin embargo, la entidad estatal todavía externaliza ciertos estudios poniendo en riesgo la consolidación de proyectos y programas de inteligencia turística que, a partir del aprovechamiento a profundidad de los datos disponibles, le permitan generar ventajas competitivas y complementar el proceso de toma de decisiones alineadas a la sostenibilidad.

- Para poder concretar las acciones planteadas como propuestas a corto y largo plazo en el Ecuador, el planteamiento idóneo sugiere una gobernanza que fomente políticas públicas desde el Ministerio de Turismo para cristalizar el trabajo articulado entre las instituciones, evangelizar sobre la cultura del dato a lo largo de toda la cadena con alta posibilidad de interacción entre agentes y priorizar la capacitación especializada, la infraestructura adecuada y el financiamiento requerido. Esto para implementar un sistema que, en tiempo real, constantemente monitoree patrones, clústeres y predicciones, a la vez que alerte de forma temprana cuando detecte cambios respecto a las tendencias históricas.

BIBLIOGRAFÍA

- Abdualgalil, B. S., & Abraham, S. (2020). Tourist Prediction Using Machine Learning Algorithms. *ICSG 2020*, 111-121. https://www.researchgate.net/profile/Bilal-Abdualgalil/publication/339927009_Tourist_Prediction_Using_Machine_Learning_Algorithms/links/5e78d66a92851c309137c1e9/Tourist-Prediction-Using-Machine-Learning-Algorithms.pdf
- Alén, E., Losada, N., & Domínguez, T. (2016). The Impact of Ageing on the Tourism Industry: An Approach to the Senior Tourist Profile. *Soc Indic Res*, 127, 303-322. <https://doi.org/10.1007/s11205-015-0966-x>
- Almache, A. (2021). *Perfil de la demanda turística del cantón Latacunga, provincia de Cotopaxi, año 2021* [Universidad Técnica de Cotopaxi]. <http://190.15.139.147/handle/27000/7813>
- ArcGIS Insights. (2023). *Crear y utilizar un mapa de coropletas*. Insights 2023.1. <https://doc.arcgis.com/es/insights/latest/create/choropleth-maps.htm>
- Back, A., Enkel, E., & Vonkrogh, G. (2007). Knowledge Networks for Business Growth. En *University of St.Gallen*. Springer. <https://doi.org/10.1007/978-3-540-33073-8>
- Baker Tilly Ecuador. (2023). *Decretos Ejecutivos Reducción de ISD, ICE e IVA en feriados 2023*. Novedades Normativas. <https://www.bakertilly.ec/es/decretos-ejecutivos-reduccion-isd-ice-iva-feriados-2023/>
- Baldassarre, M. (2016). Think big: learning contexts, algorithms and data science. *Research on Education and Media*, 8(2), 69-83. <https://doi.org/10.1515/rem-2016-0020>
- Bárcena, A., & Cimoli, M. (2020). *Sectores y empresas frente al COVID-19: emergencia y reactivación*. <https://www.cepal.org/es/publicaciones/45734-sectores-empresas-frente-al-covid-19-emergencia-reactivacion>
- Batista e Silva, F., Barranco, R., Proietti, P., Pigaiani, C., & Lavalle, C. (2021). A new European regional tourism typology based on hotel location patterns and geographical criteria. *Annals of Tourism Research*, 89, 103077. <https://doi.org/10.1016/j.annals.2020.103077>
- Bautista, A. (2021). *Perfil de la demanda turística del cantón Pangua, provincia de Cotopaxi año 2021* [Universidad Técnica de Cotopaxi]. <http://repositorio.utc.edu.ec/handle/27000/7815>
- Bel, F., Lacroix, A., Lyser, S., Rambonilaza, T., & Turpin, N. (2015). Domestic demand for tourism in rural areas: Insights from summer stays in three French regions. *Tourism Management*, 46, 562-570. <https://doi.org/10.1016/j.tourman.2014.07.020>

- Bell, G., Hey, T., & Szalay, A. (2009). Beyond the Data Deluge. *Science*, 323(5919), 1297-1298. <https://doi.org/10.1126/science.1170411>
- Bi, J.-W., Han, T.-Y., & Yao, Y. (2023). Collaborative forecasting of tourism demand for multiple tourist attractions with spatial dependence: A combined deep learning model. *Tourism Economics*. <https://doi.org/10.1177/13548166231153908>
- Bose, I. (2009). Data Mining in Tourism. En *Encyclopedia of Information Science and Technology* (Second Ed, pp. 936-940). IGI Global. <https://doi.org/10.4018/978-1-60566-026-4.ch149>
- Calle-Lamelas, J. V., García-Muiña, F. E., & García, M. (2023). What is a Smart Destination in Practice? The Interpretation of DMO Managers from Spanish World Heritage Cities. *Tourism Planning and Development*. <https://doi.org/10.1080/21568316.2023.2214123>
- Cancillería del Ecuador. (s. f.). *Ecuador, el país de los 4 mundos*. <https://www.cancilleria.gob.ec/china/ecuador-el-pais-de-los-4-mundos/#>
- Castillo-Clavero, A. M., Sánchez-Teba, E. M., & Martínez-Leiva, A. (2022). *Big Data and Artificial Intelligence in the Touristic Sector* (pp. 1-26). <https://doi.org/10.4018/978-1-7998-9008-9.ch001>
- Cava Jiménez, J. A., Millán Vázquez de la Torre, M. G., & Hernández Rojas, R. (2019). Analysis of the Tourism Demand for Iberian Ham Routes in Andalusia (Southern Spain): Tourist Profile. En *Sustainability* (Vol. 11, Número 16). <https://doi.org/10.3390/su11164278>
- Chauhan, P., & Sood, M. (2021). Big data: Present and future. *Computer*, 54, 59-65. <https://doi.org/10.1109/MC.2021.3057442>
- Choi, J., Yoon, J., Chung, J., Coh, B.-Y., & Lee, J.-M. (2020). Social media analytics and business intelligence research: A systematic review. *Information Processing & Management*, 57(6), 102279. <https://doi.org/10.1016/j.ipm.2020.102279>
- Chua, B.-L., Meng, B., Ryu, H. B., & Han, H. (2021). Participate in volunteer tourism again? Effect of volunteering value on temporal re-participation intention. *Journal of Hospitality and Tourism Management*, 46, 193-204. <https://doi.org/10.1016/j.jhtm.2020.12.003>
- Chung, P.-K., Ou, K., Wong, M. Y. C., Lau, K.-L., & Leung, K.-M. (2022). Investigation of Hong Kong Students' Esports Participation Intentions Using the Theory of Planned Behavior Approach: A Structural Equation Model. *Human Behavior and Emerging Technologies*, 2022, 1-19. <https://doi.org/10.1155/2022/6405085>
- Claveria, O., Monte, E., & Torra, S. (2016). Modelling tourism demand to Spain with machine learning techniques. The impact of forecast horizon on model selection. *Revista de*

Economía Aplicada, 24(72), 109-132.
<https://www.redalyc.org/articulo.oa?id=96949059006>

- Corzo Arevalo, D., & Guatibonza, C. M. (2021). Marketing de destinos turísticos: un análisis de las tendencias mundiales post-COVID-19 en el departamento de Santander, Colombia. *Kalpana*, 20, 72-95. <https://publicaciones.udet.edu.ec/index.php/kalpana/article/view/24>
- Cowen, D. (1990). Unit 01 - What is GIS? En M. F. Goodchild & K. K. Kemp (Eds.), *NCGIA Core Curriculum in GIS*. <https://escholarship.org/uc/item/71p6229c>
- Curtin, S. (2013). Lessons from Scotland: British wildlife tourism demand, product development and destination management. *Journal of Destination Marketing & Management*, 2(3), 196-211. <https://doi.org/10.1016/j.jdmm.2013.09.002>
- Del Chiappa, G., Atzeni, M., & Hasemi, V. (2018). Community-based collaborative tourism planning in islands: A cluster analysis in the context of Costa Smeralda. *Journal of Destination Marketing & Management*, 8, 41-48. <https://doi.org/10.1016/j.jdmm.2016.10.005>
- 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). <https://doi.org/10.1016/j.jdmm.2015.02.001>
- Deng, N., & Li, X. (Robert). (2018). Feeling a destination through the “right” photos: A machine learning model for DMOs’ photo selection. *Tourism Management*, 65, 267-278. <https://doi.org/10.1016/j.tourman.2017.09.010>
- Derek, M., Woźniak, E., & Kulczyk, S. (2019). Clustering nature-based tourists by activity. Social, economic and spatial dimensions. *Tourism Management*, 75, 509-521. <https://doi.org/10.1016/j.tourman.2019.06.014>
- Diario El Universo. (2023a). ‘Time to Reset in Ecuador’, nueva campaña para posicionar al país como destino turístico. Turismo. <https://www.eluniverso.com/noticias/ecuador/time-to-reset-in-ecuador-nueva-campana-para-posicionar-al-pais-como-destino-turistico-y-establecer-una-industria-de-turismo-mas-competitiva-nota/>
- Diario El Universo. (2023b). Sectores turístico y aéreo miran con optimismo alcanzar este año niveles prepandemia, aunque la inseguridad trae sobresaltos. Turismo. <https://www.eluniverso.com/noticias/ecuador/turismo-sector-aereo-ecuador-recuperacion-coronavirus-2023-nota/>
- Diario Expreso. (2023). Demora con telefónicas pone a Ecuador a la cola en el 5G. Economía. <https://www.expreso.ec/actualidad/economia/demora-telefonicas-pone-ecuador-cola-5g-165785.html>

- Doborjeh, Z., Hemmington, N., Doborjeh, M., & Kasabov, N. (2022). Artificial intelligence: a systematic review of methods and applications in hospitality and tourism. *International Journal of Contemporary Hospitality Management*, 34(3), 1154-1176. <https://doi.org/10.1108/IJCHM-06-2021-0767>
- Duboue, P. (2022). Feature Engineering. En R. Egger (Ed.), *Applied Data Science in Tourism: Interdisciplinary Approaches, Methodologies, and Applications* (pp. 109-127). Springer International Publishing. https://doi.org/10.1007/978-3-030-88389-8_7
- Egger, R. (2022). Machine Learning in Tourism: A Brief Overview. En R. Egger (Ed.), *Applied Data Science in Tourism: Interdisciplinary Approaches, Methodologies, and Applications* (pp. 85-107). Springer International Publishing. https://doi.org/10.1007/978-3-030-88389-8_6
- Egger, R., & Yu, J. (2022). Data Science and Interdisciplinarity. En R. Egger (Ed.), *Applied Data Science in Tourism: Interdisciplinary Approaches, Methodologies, and Applications* (pp. 35-49). Springer International Publishing. https://doi.org/10.1007/978-3-030-88389-8_3
- Eichelberger, S., Peters, M., Pikkemaat, B., & Chan, C.-S. (2020). Entrepreneurial ecosystems in smart cities for tourism development: From stakeholder perceptions to regional tourism policy implications. *Journal of Hospitality and Tourism Management*, 45, 319-329. <https://doi.org/10.1016/j.jhtm.2020.06.011>
- Femenia-Serra, F., Alzua-Sorzabal, A., & Pousa-Unanue, A. (2022). Business Intelligence and the Public Management of Destinations: The View of DMOs. En J. L. Stienmetz, B. Ferrer-Rosell, & D. Massimo (Eds.), *Information and Communication Technologies in Tourism 2022* (pp. 417-422). Springer International Publishing. https://doi.org/10.1007/978-3-030-94751-4_38
- Fernández Alcantud, A., López Morales, J. M., Moreno-Izquierdo, L., Perles Ribes, J. F., Ramón-Rodríguez, A. B., & Such Devesa, M. J. (2017). Innovación y destinos inteligentes: oportunidad para el know how turístico español. *Información Comercial Española (ICE). Revista de Economía*, 894, 137-150. <https://rua.ua.es/dspace/handle/10045/68402>
- George, R., & Booyens, I. (2014). Township Tourism Demand: Tourists' Perceptions of Safety and Security. *Urban Forum*, 25(4), 449-467. <https://doi.org/10.1007/s12132-014-9228-2>
- Getz, D., & Brown, G. (2006). Critical factors for wine tourism regions: a demand analysis. *Tourism Management*, 27, 146-158. <https://doi.org/10.1016/j.tourman.2004.08.002>
- Gretzel, U., Sigala, M., Xiang, Z., & Koo, C. (2015). Smart tourism: foundations and

- developments. *Electronic Markets*, 25(3), 179-188. <https://doi.org/10.1007/s12525-015-0196-8>
- Guerreiro, J., & Rita, P. (2020). How to predict explicit recommendations in online reviews using text mining and sentiment analysis. *Journal of Hospitality and Tourism Management*, 43, 269-272. <https://doi.org/10.1016/j.jhtm.2019.07.001>
- Han, J. (2020). Application of SVM model to environmental resource analysis in tourism development. *Journal of Physics: Conference Series*, 1629(1), 12007. <https://doi.org/10.1088/1742-6596/1629/1/012007>
- Hernández Flores, Y., Sánchez Borges, Y., Saldiña Silvera, B., & Rives González, K. (2020). Características de la demanda potencia del destino Cuba en la nueva normalidad postcoronavirus. *Revista Internacional de Turismo, Empresa y Territorio. RITUREM*, 4(2), 178-193. <https://dialnet.unirioja.es/servlet/articulo?codigo=7691805>
- Herrera Díaz, S., Monge Amores, E., Lasso Barreto, S., & Zhunio Armas, B. (2020). El turista de naturaleza. Estudio sobre el perfil del turista y su comportamiento en áreas naturales protegidas del Ecuador. Caso: Parque Nacional Cotacachi Cayapas. *Revista de Investigación de la Ciencia Turística RICIT*, 180-212. <https://dialnet.unirioja.es/servlet/articulo?codigo=7678378>
- Höpken, W., & Fuchs, M. (2022). Business Intelligence in Tourism. En Z. Xiang, M. Fuchs, U. Gretzel, & W. Höpken (Eds.), *Handbook of e-Tourism* (pp. 497-527). Springer International Publishing. https://doi.org/10.1007/978-3-030-48652-5_3
- Höpken, W., Fuchs, M., Keil, D., & Lexhagen, M. (2015). Business intelligence for cross-process knowledge extraction at tourism destinations. *Information Technology & Tourism*, 15(2), 101-130. <https://doi.org/10.1007/s40558-015-0023-2>
- Höpken, W., Fuchs, M., & Lexhagen, M. (2014). Tourism Knowledge Destination. En *Encyclopedia of Business Analytics and Optimization* (pp. 2542-2556). IGI Global. <https://doi.org/10.4018/978-1-4666-5202-6.ch227>
- INEC. (2016). *El parque automotor de Ecuador creció 57% en cinco años*. Noticias. <https://www.ecuadorencifras.gob.ec/el-parque-automotor-de-ecuador-crecio-57-en-cinco-anos/>
- Ioannides, D., & Gyimóthy, S. (2020). The COVID-19 crisis as an opportunity for escaping the unsustainable global tourism path. *Tourism Geographies*, 22(3), 624-632. <https://doi.org/10.1080/14616688.2020.1763445>
- Iorio, C., Pandolfo, G., D'Ambrosio, A., & Siciliano, R. (2020). Mining big data in tourism. *Quality & Quantity*, 54(5), 1655-1669. <https://doi.org/10.1007/s11135-019-00927-0>

- Jamal, S., Goyal, S., Grover, A., & Shanker, A. (2018). Machine learning: What, why, and how? En A. Shanker (Ed.), *Bioinformatics: Sequences, structures, phylogeny* (pp. 359-374). Springer. https://doi.org/10.1007/978-981-13-1562-6_16
- Katsaros, K. V, Gkounis, D., Kaleshi, D., Thomas, B., Harris, J., Falaki, H., & Simeonidou, D. (2019). Enhancing Tourist Experiences through 5G - The 5G Smart Tourism Case Study. *2019 IEEE 2nd 5G World Forum (5GWF)*, 471-476.
- Knani, M., Echchakoui, S., & Ladhari, R. (2022). Artificial intelligence in tourism and hospitality: Bibliometric analysis and research agenda. *International Journal of Hospitality Management*, *107*, 103317. <https://doi.org/https://doi.org/10.1016/j.ijhm.2022.103317>
- Lang, C.-T., & O’leary, J. T. (1997). Motivation, Participation, and Preference: A Multi-Segmentation Approach of the Australian Nature Travel Market. *Journal of Travel & Tourism Marketing*, *6*(3-4), 159-180. https://doi.org/10.1300/J073v06n03_10
- Li, Q., Li, S., Hu, J., Zhang, S., & Hu, J. (2018). Tourism Review Sentiment Classification Using a Bidirectional Recurrent Neural Network with an Attention Mechanism and Topic-Enriched Word Vectors. *Sustainability*, *10*(9), 3313. <https://doi.org/10.3390/su10093313>
- Li, W. (2022). Prediction of Tourism Demand in Liuzhou Region Based on Machine Learning. *Hindawi Mobile Information Systems*. <https://doi.org/10.1155/2022/9362562>
- Li, X., Law, R., Xie, G., & Wang, S. (2021). Review of tourism forecasting research with internet data. *Tourism Management*, *83*, 104245. <https://doi.org/10.1016/j.tourman.2020.104245>
- Liu, C., Williams, A. M., & Li, G. (2022). Knowledge management practices of tourism consultants: A project ecology perspective. *Tourism Management*, *91*, 104491. <https://doi.org/10.1016/j.tourman.2022.104491>
- López-Sánchez, Y., & Pulido-Fernández, J. I. (2016). In search of the pro-sustainable tourist: A segmentation based on the tourist “sustainable intelligence”. *Tourism Management Perspectives*, *17*, 59-71. <https://doi.org/10.1016/j.tmp.2015.12.003>
- López de Ávila, A., Lancis, E., García, S., Alcantud, A., García, B., & Muñoz, N. (2015). *Informe destinos turísticos inteligentes: construyendo el futuro*. SEGITTUR. <https://www.segittur.es/wp-content/uploads/2019/11/Libro-Blanco-Destinos-Turísticos-Inteligentes.pdf>
- LUCA Data-Driven Decisions. (2018). *Informe Final Contratación del servicio de información y estudio de patrones de movilidad nacional*.
- Mabrian Technologies. (2020). *¿Qué es la Inteligencia Turística?*

<https://mabrian.com/blog/es/inteligencia-turistica/>

- Mariani, M., & Baggio, R. (2022). Big data and analytics in hospitality and tourism: a systematic literature review. *International Journal of Contemporary Hospitality Management*, 34(1), 231-278. <https://doi.org/10.1108/IJCHM-03-2021-0301>
- Mariani, M., Baggio, R., Fuchs, M., & Höepken, W. (2018). Business intelligence and big data in hospitality and tourism: a systematic literature review. *International Journal of Contemporary Hospitality Management*, 30(12), 3514-3554. <https://doi.org/10.1108/IJCHM-07-2017-0461>
- Martin-Fuentes, E., Fernandez, C., Mateu, C., & Marine-Roig, E. (2018). Modelling a grading scheme for peer-to-peer accommodation: Stars for Airbnb. *International Journal of Hospitality Management*, 69, 75-83. <https://doi.org/10.1016/j.ijhm.2017.10.016>
- Martinez-Torres, M. R., & Toral, S. L. (2019). A machine learning approach for the identification of the deceptive reviews in the hospitality sector using unique attributes and sentiment orientation. *Tourism Management*, 75, 393-403. <https://doi.org/10.1016/j.tourman.2019.06.003>
- Más-Ferrando, A., Ramón-Rodríguez, A., & Aranda Cuéllar, P. (2020). La revolución digital en el sector turístico. Oportunidad para el turismo en España. *Ekonomiaz*, 98(2), 1-24. https://rua.ua.es/dspace/bitstream/10045/111464/1/Mas-Ferrando_etal_2020_Ekonomiaz.pdf
- Mendieta-Aragón, A. (2022). Cambios en el comportamiento turístico tras la COVID-19: hacia un nuevo perfil del turista y del viaje de ocio en España. *Rect@: Revista Electrónica de Comunicaciones y Trabajos de ASEPUMA*, 23(1), 37-51. <https://dialnet.unirioja.es/servlet/articulo?codigo=8913733>
- Mera, E., Alfonso, R., Tavares, M., Araque, J., Madrid, L., & Pastrán, F. (2020). Perfil del turista: origen, características y percepciones de los visitantes en Pedernales, Ecuador. *Revista científica "Suplemento CICA multidisciplinario"*, 10. <https://www.ulead.edu.ec/wp-content/uploads/2022/01/Suplemento-CICA-Nº-10-Diciembre-del-2020.pdf#page=78>
- Mich, L. (2022). AI and Big Data in Tourism. En R. Egger (Ed.), *Applied Data Science in Tourism: Interdisciplinary Approaches, Methodologies, and Applications* (pp. 3-15). Springer International Publishing. https://doi.org/10.1007/978-3-030-88389-8_1
- MINDTEK. (2015a). *Estudio de Encuesta telefónica de Turismo Nacional*.
- MINDTEK. (2015b). *Segundo Producto Contratación de realización del sondeo telefónico de turismo*.

- Ministerio de Turismo. (2011). *Estudio piloto de turismo interno en 2010*.
- Ministerio de Turismo. (2012). *Estudio piloto de turismo interno en 2011*.
- Ministerio de Turismo. (2013). *Estudio piloto de turismo interno en 2012*.
- Ministerio de Turismo. (2014). *Arranca segunda fase de la campaña “All You Need Is Ecuador”*. Noticias. <https://www.turismo.gob.ec/hoy-arranca-segunda-fase-de-la-campana-all-you-need-is-ecuador/>
- Ministerio de Turismo. (2018). *Perfil de Turismo Internacional 2017*. <https://servicios.turismo.gob.ec/descargas/Turismo-cifras/Publicaciones/Perfiles/PerfilesDeTurismoInternacional.pdf>
- Ministerio de Turismo. (2020). *Plan de Reactivación Turística 2020*. https://www.turismo.gob.ec/wp-content/uploads/2021/05/Plan-Reactivacion-Turistica-Red_compressed.pdf
- Ministerio de Turismo. (2021). *Proyecto de Inversión Promoción para la reactivación turística del Ecuador*. <https://www.turismo.gob.ec/wp-content/uploads/2022/02/Proyecto-Promocion-Turistica-para-la-Reactivacion-del-Ecuador.pdf>
- Ministerio de Turismo. (2023a). *Emisión de información estadística y económica del sector turístico local y nacional*. <https://www.gob.ec/mintur/tramites/emision-informacion-estadistica-economica-sector-turistico-local-nacional>
- Ministerio de Turismo. (2023b). *Visualizador de Información Turística*. Ecuador. <https://servicios.turismo.gob.ec/visualizador>
- Ministerio de Turismo. (2023c). *Volar desde y hacia Ecuador será más barato: Gobierno reduce las tasas Eco Delta y Ecuador Potencia Turística*. Gestión. <https://www.turismo.gob.ec/volar-desde-y-hacia-ecuador-sera-mas-barato-gobierno-reduce-las-tasas-eco-delta-y-ecuador-potencia-turistica/>
- Molera, L., & Albaladejo, I. P. (2007). Profiling segments of tourists in rural areas of South-Eastern Spain. *Tourism Management*, 28, 757-767. <https://doi.org/10.1016/j.tourman.2006.05.006>
- Molina-López, J. M., & García-Herrero, J. (2006). *Técnicas de Análisis de Datos: Aplicaciones prácticas utilizando Microsoft Excel y WEKA*. http://matema.ujaen.es/jnavas/web_recursos/archivos/weka_master_recursos_naturales/apuntesAD.pdf
- Molina Velásquez, E., Cabanilla Vásquez, E., Garrido Cornejo, C., & Borja Vivero, J. F. (2022). Estudio del comportamiento de la demanda turística interna en la reactivación pos-covid-19 en Ecuador. *Revista Científica ECOCIENCIA*, 63-80.

- <http://www.dspace.uce.edu.ec/bitstream/25000/26836/1/22> MONILA-CABANILLA-GARRIDO-BORJA ESTUDIO DEL COMPORTAMIENTO.pdf
- Molina Velásquez, E., Cabanilla Vásquez, E., & Méndez Játiva, J. F. (2020). Preferencias, hábitos de viaje y grado de satisfacción de los visitantes a centros de turismo comunitario en 6 provincias del Ecuador. *Realidad, Tendencias y Desafíos En Turismo (CONDET)*, 18(1), 27-52. <https://revelo.uncoma.edu.ar/index.php/condet/article/view/2684>
- Moreno-Izquierdo, L., Más-Ferrando, A., Suárez-Tostado, M., & Ramón-Rodríguez, A. (2022). Reinención del turismo en clave de inteligencia artificial. Buscando un modelo sostenible y competitivo para el siglo XXI. *Apuntes Fedea*, 2022/19, 1-17. <https://ideas.repec.org/p/fda/fdafen/2022-19.html>
- Naranjo Llupart, M. R., & Martínez Rodríguez, M. de los A. (2022). Reflexiones teóricas sobre la demanda turística global: Incidencia en la gestión y comercialización turística. *Revista de Ciencias Sociales (Ve)*, 28(5), 359-375. <https://www.redalyc.org/journal/280/28071845029/html/#gt1>
- Neuburger, L., & Egger, R. (2021). Travel risk perception and travel behaviour during the COVID-19 pandemic 2020: a case study of the DACH region. *Current Issues in Tourism*, 24(7), 1003-1016. <https://doi.org/10.1080/13683500.2020.1803807>
- Oficina de Información Diplomática Ecuador. (2023). *Ficha País Ecuador*. https://www.exteriores.gob.es/documents/fichaspais/ecuador_ficha_pais.pdf
- OMT. (s. f.-a). 2020: *Análisis del año. COVID-19 y el sector turístico*. <https://www.unwto.org/es/covid-19-y-sector-turistico-2020>
- OMT. (s. f.-b). *COVID-19 RESPUESTA*. Guiar la recuperación del turismo. <https://www.unwto.org/es/turismo-covid-19>
- OMT. (s. f.-c). *Glosario de Términos de Turismo*. <https://www.unwto.org/es/glosario-terminos-turisticos>
- ONU. (2020). *Informe de políticas: La COVID-19 y la transformación del turismo*. https://www.un.org/sites/un2.un.org/files/2020/10/policy_brief_covid-19_and_transforming_tourism_spanish.pdf
- ONU, & CEPAL. (2020). *INFORMES COVID-19 Medidas de recuperación del sector turístico en América Latina y el Caribe: una oportunidad para promover la sostenibilidad y la resiliencia*. https://repositorio.cepal.org/bitstream/handle/11362/45770/1/S2000441_es.pdf
- Park, D., & Yoon, Y. (2009). Segmentation by motivation in rural tourism: A Korean case study. *Tourism Management*, 30, 99-108. <https://doi.org/10.1016/j.tourman.2008.03.011>

- Payntar, N. D., Hsiao, W.-L., Covey, R. A., & Grauman, K. (2021). Learning patterns of tourist movement and photography from geotagged photos at archaeological heritage sites in Cuzco, Peru. *Tourism Management*, 82, 104165. <https://doi.org/10.1016/j.tourman.2020.104165>
- Prensa Latina. (2023). *Sector turístico en Ecuador alerta ante afectaciones de El Niño*. Noticias. <https://www.prensa-latina.cu/2023/06/25/sector-turistico-en-ecuador-alerta-ante-afectaciones-de-el-nino>
- Primicias. (2023). *La inseguridad ahuyenta millones en inversión para el turismo*. Economía. <https://www.primicias.ec/noticias/economia/inseguridad-inversion-turismo-ecuador/>
- Provost, F., & Fawcett, T. (2013). Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data*, 1(1), 51-59. <https://doi.org/10.1089/big.2013.1508>
- Puig-Cabrera, M., & Foronda-Robles, C. (2018). El turismo y el alivio de la pobreza: un enfoque desde el prisma de los destinos emergentes en el marco de la Agenda 2030. *Investigaciones Turísticas*, 16, 1-22. <https://doi.org/10.14198/INTURI2018.16.01>
- Pulido-Fernández, J. I. (2020). *El turismo responsable como respuesta a la idea de turismo para todos*. https://cidecuador.org/wp-content/uploads/congresos/2020/iv-forestal/diapo/el-turismo-responsable-como-respuesta-a-la-idea-de-turismo-para-todos_juan-pulido.pdf
- Pulido-Fernández, J. I. (2021). *Análisis del comportamiento del turista de interior tras la COVID-19: percepción de riesgo, cambio de estilo de vida e intención de viajar*. <https://www.catedraturismodeinterior.com/analisis-del-comportamiento-del-turista-de-interior-tras-la-covid-19/>
- Pyo, S., Uysal, M., & Chang, H. (2002). Knowledge Discovery in Database for Tourist Destinations. *Journal of Travel Research*, 40(4), 374-384. <https://doi.org/10.1177/0047287502040004006>
- QGIS Project. (2023). *Una introducción fácil a GIS*. QGIS Documentation v 3.28. https://docs.qgis.org/3.28/es/docs/gentle_gis_introduction/index.html
- Rabasa, A. (2022). *Apuntes de la asignatura “Técnicas e instrumentos para la generación de inteligencia turística”*. Universidad de Jaén.
- Rabasa, A., & Heavin, C. (2020). An Introduction to Data Science and Its Applications. En V. Charles, J. Aparicio, & J. Zhu (Eds.), *Data Science and Productivity Analytics, International Series in Operations Research & Management Science* (pp. 57-81). Springer. https://doi.org/10.1007/978-3-030-43384-0_3
- Rabasa, A., Mollá-Campello, N., & Pérez-Torregrosa, A. (2018). Formal descriptive study for

- the extraction and comparison of tourist spending patterns. *International Journal of Design & Nature and Ecodynamics*, 13(3), 272-280. <https://doi.org/10.2495/DNE-V13-N3-272-280>
- Rabasa, A., Pérez-Martín, A., & Giner, D. (2017). Optimal clustering techniques for the segmentation of tourist spending. Analysis of tourist surveys in the Valencian Community (Spain): A case study. *International Journal of Design & Nature and Ecodynamics*, 12(4), 482-491. <https://doi.org/10.2495/DNE-V12-N4-482-491>
- Ramos-Henríquez, J. M., Gutiérrez-Taño, D., & Díaz-Armas, R. J. (2021). Value proposition operationalization in peer-to-peer platforms using machine learning. *Tourism Management*, 84, 104288. <https://doi.org/10.1016/j.tourman.2021.104288>
- Ramos, C. M. (2023). Business Intelligence Applied to Tourism. En J. Wang (Ed.), *Encyclopedia of Data Science and Machine Learning* (pp. 579-594). IGI Global. <https://doi.org/10.4018/978-1-7998-9220-5.ch034>
- Rigol, L., Pérez, R., Noda, M. E., & González, J. (2009). Conceptualización de la demanda turística. *Ciencias Holguín*, XV(1), 1-8. <https://www.redalyc.org/pdf/1815/181517987002.pdf>
- Rodríguez, P., Valencia-Arias, A., Garcés-Giraldo, L., Castañeda, L., Moreno, G., & Benjumea-Arias, M. (2023). Tendencias en el uso de inteligencia artificial en el sector del turismo. *Journal of Tourism & Development*, 40, 81-92. <https://doi.org/10.34624/rtd.v40i0.31447>
- Samara, D., Magnisalis, I., & Peristeras, V. (2020). Artificial intelligence and big data in tourism: a systematic literature review. *Journal of Hospitality and Tourism Technology*, 11(2), 343-367. <https://doi.org/10.1108/JHTT-12-2018-0118>
- Sánchez-Cañizares, S. M., & López-Guzmán, T. (2012). Gastronomy as a tourism resource: profile of the culinary tourist. *Current Issues in Tourism*, 15(3), 229-245. <https://doi.org/10.1080/13683500.2011.589895>
- Sánchez Ruiz, J., & Paladines Sarango, T. (2021). Perfil del turista con capacidades especiales: Caso de estudio Parque Nacional Yacuri de la provincia de Loja- Ecuador. *Revista Reflexiones*, 100(2). <https://doi.org/10.15517/rr.v100i2.43326>
- Santos, M. C., Veiga, C., & Águas, P. (2016). Tourism services: facing the challenge of new tourist profiles. *Worldwide Hospitality and Tourism Themes*, 8(6), 654-669. <https://doi.org/10.1108/WHATT-09-2016-0048>
- Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, 2(160). <https://doi.org/10.1007/s42979-021-00592-x>

- SEGITTUR. (2015). *Destinos Turísticos Inteligentes*. Proyectos Destinos. <https://www.segittur.es/destinos-turisticos-inteligentes/proyectos-destinos/destinos-turisticos-inteligentes/>
- SEGITTUR. (2022). *Sistema de Inteligencia Turística*. Destino Turístico Inteligente. <https://www.destinosinteligentes.es/soluciones/sistema-de-inteligencia-turistica/>
- Slocum, T. A., McMaster, R. B., Kessler, F. C., & Howard, H. H. (2022). *Thematic Cartography and Geovisualization* (Fourth Ed). CRC Press. <https://doi.org/10.1201/9781003150527>
- Smith, M., & Puczkó, L. (2015). More than a special interest: defining and determining the demand for health tourism. *Tourism Recreation Research*, 40(2), 205-219. <https://doi.org/10.1080/02508281.2015.1045364>
- Song, I.-Y., & Zhu, Y. (2016). Big data and data science: what should we teach? *Expert Systems*, 33(4), 364-373. <https://doi.org/10.1111/exsy.12130>
- Srihadi, T. F., Hartoyo, Sukandar, D., & Soehadi, A. W. (2016). Segmentation of the tourism market for Jakarta: Classification of foreign visitors' lifestyle typologies. *Tourism Management Perspectives*, 19, 32-39. <https://doi.org/10.1016/j.tmp.2016.03.005>
- Sutty, H. (2021). Principales planes de viajes post COVID-19 en Paraguay. *Revista científica En Ciencias Sociales*, 3(1), 16-24. <https://doi.org/10.53732/rccsociales/03.01.2021.16>
- Timarán-Pereira, S. R., Hernández-Arteaga, I., Caicedo-Zambrano, S. J., Hidalgo-Troya, A., & Alvarado-Pérez, J. C. (2016). El proceso de descubrimiento de conocimiento en bases de datos. En *Descubrimiento de patrones de desempeño académico con árboles de decisión en las competencias genéricas de la formación profesional* (pp. 63-86). Ediciones Universidad Cooperativa de Colombia. <https://doi.org/10.16925/9789587600490>
- Toledo, N., Bermúdez-Gallegos, C., & Coronel, V. (2020). Perfil del turista de la Generación Z en Ecuador. *Revista Universidad de Guayaquil*, 131(2), 19-26. <https://pdfs.semanticscholar.org/72c8/ef449f9de2b8003ec2832944c6b94131285b.pdf>
- Valeri, M. (2020). Blockchain Technology: Adoption Perspectives in Tourism. En V. Ratten (Ed.), *Entrepreneurship and Organizational Change: Managing Innovation and Creative Capabilities* (pp. 27-35). Springer International Publishing. https://doi.org/10.1007/978-3-030-35415-2_3
- Velasco, P. (2022a). *Consultoría Estudio de Demanda Interna - Producto 1*.
- Velasco, P. (2022b). *Consultoría Estudio de Demanda Interna - Producto 2*.
- Veloz, W., Félix, A., Mora, V., & Duarte, A. (2022). Diagnóstico turístico con proyección DTI (destino turístico inteligente) de la ciudad de Manta, Ecuador. *Revista Internacional de*

Derecho y Economía del Turismo, 4(1), 1-26.
<https://dialnet.unirioja.es/servlet/articulo?codigo=8765412>

- Villamediana-Pedrosa, J. D., Vila-López, N., & Küster-Boluda, I. (2020). Predictors of tourist engagement: Travel motives and tourism destination profiles. *Journal of Destination Marketing & Management*, 16, 100412. <https://doi.org/10.1016/j.jdmm.2020.100412>
- World Economic Forum. (2022). *Interactive Data and Economy Profiles*. Travel & Tourism Development Index 2021: Rebuilding for a Sustainable and Resilient Future. <https://www.weforum.org/reports/travel-and-tourism-development-index-2021/explore-the-data#report-nav>
- WTTC. (2023). *Sector de viajes y turismo de Ecuador representará el 4,4% de la economía nacional al cierre de 2023: WTTC*. <https://wttc.org/news-article/sector-de-viajes-y-turismo-de-ecuador-representara-el-4-4-de-la-economia-nacional-al-cierre-de-2023#:~:text=En 2022%2C la contribución del,2%25 de la economía ecuatoriana.>

ANEXOS

Anexo 1. Guion de entrevista

Sección 1

Buen día. Mi nombre es Vanessa Hinojosa, alumna del Máster en Planificación y Gestión Sostenible del Turismo de la Universidad de Jaén. En el marco de la asignatura Metodología y Técnicas de Investigación en Turismo de cara al Trabajo de Fin de Máster, se está realizando un estudio sobre el uso de técnicas avanzadas de Ciencias de Datos para la comprensión del turismo interno en Ecuador desde las características del visitante, viaje y destino.

En este sentido, siéntase en libertad de compartir sus ideas en este espacio y opiniones sinceras respecto de las cuestiones consultadas.

Cabe aclarar que sus respuestas serán analizadas de forma anónima y respetando el criterio de confidencialidad.

Para agilizar la toma de la información, resulta de utilidad grabar la conversación. ¿Existe algún inconveniente en que grabemos la misma? El uso de la misma es sólo para fines académicos.

Desde ya, le agradezco por su tiempo y apoyo.

Sección 2

Nombre y Apellidos: _____

Cargo: _____

Entidad: _____

Sección 3

Desde su perspectiva de experto en Turismo, responda por favor a las siguientes preguntas:

Si le menciono el término Ciencia de Datos ¿qué es lo primero que se le viene a la mente? ¿A qué se refiere este concepto?

Hoy en día, ¿qué beneficios cree que aporta la ciencia de datos y sus técnicas al sector turístico? ¿por qué? Y ¿presentan algún obstáculo? ¿por qué?

Sección 4

A nivel personal ¿usted ha utilizado alguna técnica de ciencia de datos para estudios que haya realizado o se encuentre realizando vinculados a turismo?

¿Qué tan importante es para usted que se recopilen y analicen datos en el sector turístico de manera sistemática? ¿Existen mecanismos a su alcance para hacerlo, es decir, que le permitan

generar y gestionar datos propios? ¿En sus estudios y planificaciones, se usan datos de fuentes externas?

¿Cree que existe una cultura de uso o explotación del dato que ayude a tomar decisiones estratégicas y a planificar el sector turístico?

¿Es frecuente realizar o encargar análisis estadísticos sobre los datos del sector turístico? En caso afirmativo, estos análisis estadísticos: ¿con qué frecuencia se realizan?, ¿son realizados por usted o son externalizados? ¿Estos análisis incorporan métodos analíticos predictivos o son sólo descriptivos? ¿Conoce con qué software se llevan a cabo los análisis?

¿Cree que la predicción de determinadas variables turísticas (ocupación, pernoctaciones, gasto medio, cancelaciones...) ayuda (o podría ayudar) a una mejor toma de decisiones en el sector turístico?

Sección 5

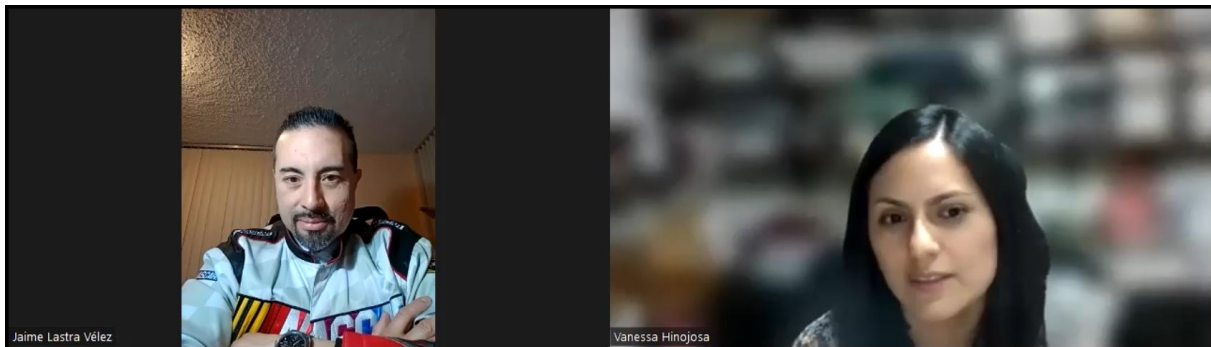
Ahora bien, para terminar ¿cómo visualiza el sector turístico en el Ecuador dentro de los próximos 10 años con relación al uso de técnicas avanzadas de Ciencia de Datos? ¿Cómo imagina a los stakeholders y la toma de decisiones para el turismo en nuestro país desde predicciones?

¿Cuál avizora que serían las principales barreras para su uso masivo por toda la cadena de valor turística? ¿Cuáles serían las principales necesidades para su incorporación completa o el incremento de su uso?

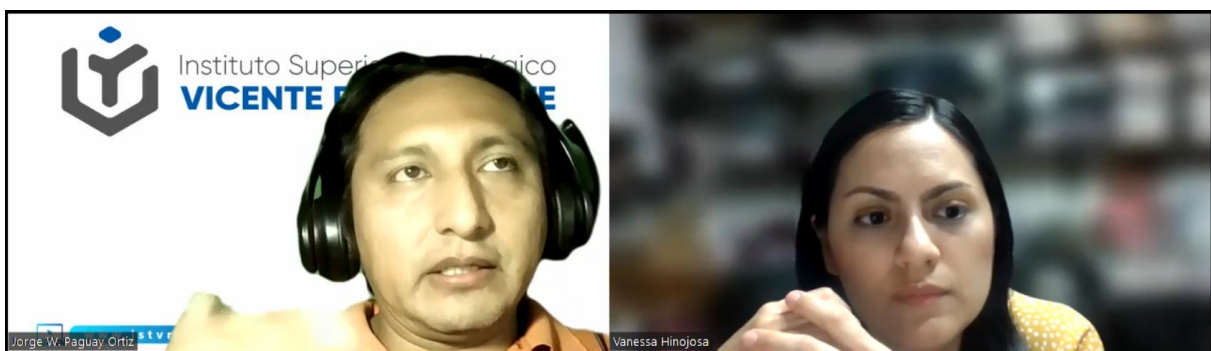
Finalmente, ¿algún otro comentario que quiera agregar?

¡Nuevamente, muchas gracias por el aporte a esta investigación!

Anexo 2. Perfil de entrevistados y evidencias fotográficas



Entrevista con Jaime Lastra.- Analista (Servidor Público 5) en la Dirección de Inteligencia de Mercados de Ministerio de Turismo, además Docente y Coordinador en forma remota de la Maestría en Administración de la Pontificia Universidad Católica del Ecuador (Sede Manabí). También se desempeña como experto independiente en Business Intelligence por 5 años.



Entrevista con Jorge Paguay.- Consultor del Grupo Biogenia por su experiencia en el sector turístico tanto en el ámbito público y privado, además Docente del Instituto Superior Tecnológico Vicente Rocafuerte. Actualmente, también es parte de la Red de Profesionales de Turismo del Ecuador.

Anexo 3. Cuestionario definitivo

1. ¿En qué lugar del Ecuador reside? (Lista desplegable)
 - a. Provincia _____
 - b. Cantón _____
2. Estado civil: Soltera/o_____ Casada/o_____ Unión libre_____ Divorciada/o_____
3. Género: Femenino_____ Masculino_____
4. Edad: _____
5. Nivel de estudios: Primaria_____ Secundaria_____ Superior_____
6. ¿Cuál es su correo electrónico? _____
7. ¿Cuál es el ingreso promedio mensual de su hogar?
 - Menos de \$400
 - De \$401 a \$800
 - De \$801 a \$1 200
 - De \$1 201 a \$1 500
 - De \$1 501 a \$1 700
 - Más de \$1 701
8. ¿Qué miembro de su entorno familiar es quién tomó la decisión para su viaje dentro del país en los últimos meses?
 - Sí, la tomé yo
 - No la tomé yo
 - Fue una decisión en conjunto
9. ¿Cuál fue la principal razón de su último viaje dentro del país en los últimos 3 meses?
 - Vacaciones, recreo, ocio (turismo)
 - Visitas a familiares y amigos
 - Educación y formación
 - Salud y atención médica
 - Religión
 - Negocios y/o motivos profesionales
 - Compras
 - Otro, especifique: _____
10. ¿Con base a la pregunta anterior, en qué región del país realizó su viaje?
 - Costa
 - Sierra
 - Amazonía

Galápagos

11. ¿Podría mencionar cuál fue su destino para viajar dentro del país? (Lista desplegable)

a. Provincia _____

b. Ciudad _____

12. ¿Con cuánto tiempo de anticipación realizó la planificación de su último viaje de turismo?

Una semana

Un mes

Tres meses

Seis meses

Más de seis meses

13. ¿Se quedó al menos una noche (pernoctó) en el viaje? (si la respuesta es afirmativa, se despliegan las preguntas #14 y #15, caso contrario pasa a la #16)

Sí

No

14. Si su respuesta anterior fue afirmativa ¿cuántos días se quedó en el lugar que visitó?

_____ días

15. ¿Cuál es el principal tipo de alojamiento utilizado en el lugar de destino de su viaje?

Hoteles

Vivienda propia

Vivienda alquilada

Hostales

Otro, cuál: _____

16. ¿Con cuántas personas realizó el viaje?

Una

Dos

Tres

Cuatro

Cinco

Otro, especifique: _____

17. ¿Con base a la pregunta anterior, especifique con quién realizó el viaje? (máximo 3)

Cónyuge

Hijo (a)

Padres y/o suegros

Nieto (a)

- Yerno o nuera
- Amigos
- Otros parientes
- Otros no parientes

18. ¿Podría indicarnos cuales fueron los montos de gasto en su último viaje de turismo?, referente a:

- Antes _____ USD (preparativos)
- Durante _____ USD (en el viaje)
- Después _____ USD (gastos adicionales en el retorno)

19. De manera referencial, describa el porcentaje distributivo de gasto por cada servicio utilizado sobre el siguiente detalle: (colocar casilleros distributivos y prueba lógica del 100%)

- Alojamiento
- Transporte aéreo
- Transporte terrestre
- Combustible
- Alimentos y bebidas
- Entretenimiento
- Varios
- Otro, cuál: _____

20. ¿Cuál medio de transporte utilizó en su último viaje?

- Vehículo propio
- Vehículo alquilado
- Autobús / flotas
- Aéreo
- Fluvial

21. ¿Cuáles son los medios de comunicación que utiliza para informarse sobre destinos y paquetes turísticos del Ecuador?

- Televisión
- Radio
- Medios impresos
- Internet (desplegable opción pregunta 24)
- Redes sociales (desplegable opción pregunta 23)
- Publicidad vial o señalética
- Recomendaciones de familiares o amigos

Otro, cuál: _____

22. ¿Cuáles son las redes sociales que utiliza para informarse acerca de promociones y/o paquetes turísticos del Ecuador?

Facebook

Instagram

Twitter

Tik tok

Otro, especifique: _____

23. ¿Qué plataformas de búsqueda usted utiliza para conocer información turística de manera general?

Despegar

Booking

Priceline

Kayak

Hoteles.com

Trivago

Expedia

Otro, especifique: _____

24. ¿Qué tan complicado le resultó gestionar los detalles de su último viaje de turismo con respecto a? – Considerar escala del 1 al 4, donde 1 es nada complicado y 4 muy complicado

a. Reservas (ubicar escala para cada actividad)

b. Transporte (ubicar escala para cada actividad)

c. Alimentación (ubicar escala para cada actividad)

d. Alojamiento (ubicar escala para cada actividad)

e. Actividades de entretenimiento (ubicar escala para cada actividad)

25. ¿Tiene planeado realizar un viaje de turismo dentro del Ecuador en los próximos 3 meses?

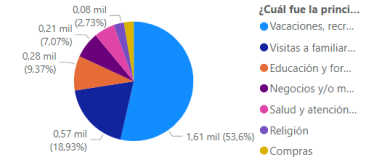
Sí

No

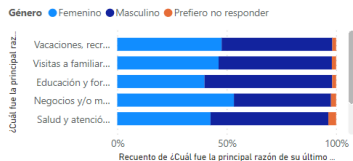
Anexo 4. Resumen de construcción de perfilados

¿Cuál fue la principal razón de su último viaje dentro del país en los últimos 3 meses?

Recuento de ¿Cuál fue la principal razón de su último viaje dentro del país en los últimos 3 meses? por ¿Cuál fue la principal razón de su último viaje dentro del país en los últimos 3 meses?

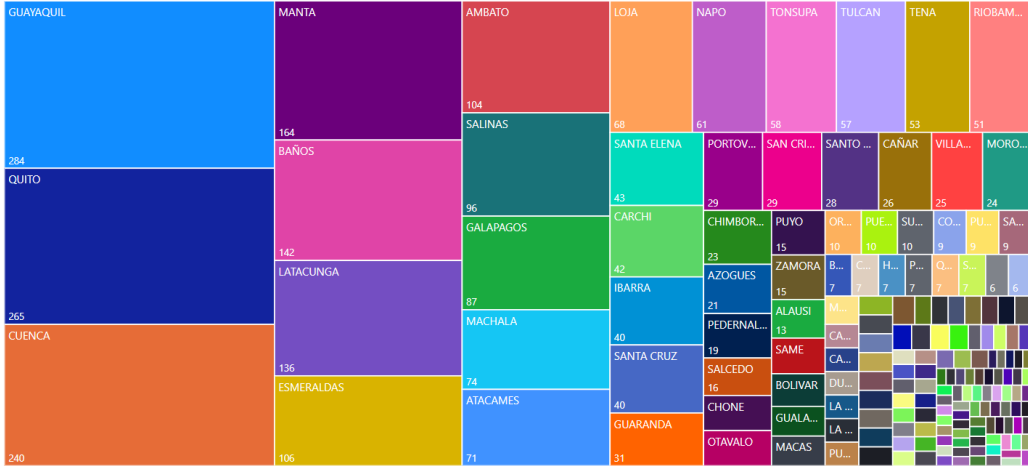


Recuento de ¿Cuál fue la principal razón de su último viaje dentro del país en los últimos 3 meses? por ¿Cuál fue la principal razón de su último viaje dentro del país en los últimos 3 meses? y Región

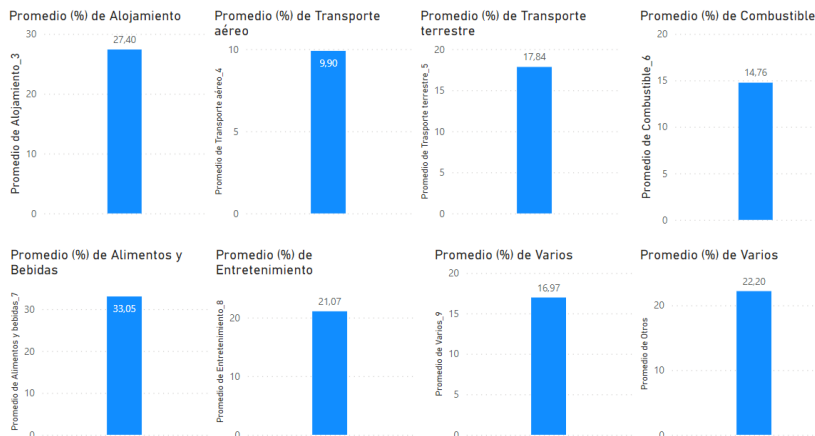


Podría mencionar cual fue su destino para viajar dentro del país

Recuento de Ciudad de Viaje por Ciudad de Viaje



De manera referencial, describa el porcentaje distributivo de gasto por cada servicio utilizado sobre el siguiente detalle



Fuente: (Velasco, 2022b)