



---

**Cita bibliográfica:** Bassols Gardella, N., Ovalle Díaz, A. P. y Rodríguez Hernández, J. C. (2021). Preferencias de los turistas en hoteles y destinos: una aproximación desde el análisis de contenido y los árboles de decisión. *Investigaciones Turísticas* (22), pp. 121-147. <https://doi.org/10.14198/INTURI2021.22.6>

---

## **Preferencias de los turistas en hoteles y destinos: una aproximación desde el análisis de contenido y los árboles de decisión**

### **Tourist preferences in hotels and destinations: an approach based on content analysis and decision trees**

Narcís Bassols Gardella , Universitat Rovira i Virgili, España  
[narcis.bassols@gmail.com](mailto:narcis.bassols@gmail.com)

Andrea Paola Ovalle Díaz , MARFIL Impresos y Papel, S. A. S, Colombia  
[apovalle3@misena.edu.co](mailto:apovalle3@misena.edu.co)

Jennifer Carolina Rodríguez Hernández , MARFIL Impresos y Papel, S. A. S, Colombia  
[jenifer\\_0724@hotmail.com](mailto:jenifer_0724@hotmail.com)

#### **RESUMEN**

Intentar explicar las preferencias de turistas a la hora de elegir un alojamiento en un destino, y qué elementos son más influyentes en esta decisión, es un tema que ha sido investigado desde diferentes marcos de referencia y con distintas herramientas de campo en los últimos años. El presente artículo se añade a este corpus investigativo analizando las preferencias de los visitantes en las ciudades de Santa Marta y Cartagena de Indias, los dos principales destinos turísticos del Caribe colombiano. En particular, se considera si las valoraciones online de hoteles emitidas por los turistas pueden estar influenciadas por la oferta presente en el destino (atractivos turísticos, ubicación, actividades ofrecidas, restaurantes, playas, etc.) combinadas con las características propias del establecimiento hotelero (servicios ofrecidos, ubicación, precios, categoría del hotel, etc.). Los dos destinos mencionados son estudiados aplicando el modelo de árboles de decisión CHAID, una herramienta frecuentemente usada en ingeniería, pero mucho menos empleado en la gestión de empresas turísticas, a pesar de su innegable utilidad. Los datos levantados para esta investigación provienen de los comentarios online de la plataforma [booking.com](https://www.booking.com). Los resultados ofrecen una caracterización relevante de los destinos estudiados y sus visitantes; algunas tendencias conocidas se confirman, pero también se descubren otras latentes. En un plano práctico y de gestión, se pretende demostrar la importancia del uso de herramientas *big data* para la toma de decisiones de negocio, en particular las basadas en las preferencias de los clientes.

**Palabras clave:** Comportamiento del turista; métodos de toma de decisiones; segmentación de clientes; comentarios on-line; algoritmo CHAID.

## ABSTRACT

In recent years, different tools and frameworks have been proposed to explain the preferences of tourists when choosing accommodation in a destination so as to understand the elements that most influence this decision. This article contributes to this line of research by analyzing tourist preferences in the cities of Santa Marta and Cartagena de Indias, the two most prominent destinations on the Colombian Caribbean coast. More precisely, it researches how the destinations' attributes (tourist attractions, location, activities offered, beaches...) combined with the hotels' features (services offered, location, price, category...) influence online comments. Both destinations are examined with the CHAID decision tree model, a tool often used in engineering but much less applied in hospitality management – although it is a highly useful one. The data collected for this research are drawn from the comments users uploaded to the platform [booking.com](https://www.booking.com) and the results help to better characterize the two destinations and their visitors. While some of the trends observed are known to the professional community in both destinations, other latent trends are not. On a more practical note, the authors seek to prove the usefulness of big data tools in supporting decision-taking processes, particularly those depending on the knowledge of customer preferences.

**Keywords:** Tourist behavior; decision-taking methods; customer segmentation; online comments; CHAID algorithm.

## I. INTRODUCCIÓN, PRESENTACIÓN DEL PROBLEMA Y CONTEXTO

La presente investigación tiene como propósito analizar qué elementos de un alojamiento y qué elementos de un destino influyen conjuntamente el comportamiento de sus potenciales clientes, los turistas. Dicho de otra manera, cuando un cliente de un hotel o un destino está en el proceso de elección, ¿qué factores son decisivos? Ello se investiga basándose en las valoraciones en línea recopiladas sobre los alojamientos de las ciudades de Santa Marta y Cartagena, en el Caribe colombiano, para posteriormente analizar estas valoraciones con el algoritmo de árboles de decisión CHAID, que dará una respuesta a nuestras indagaciones. A conocimiento de los autores, a la fecha no existe un estudio en el sector turístico con la metodología de árboles de decisión que evalúe si las opiniones emitidas por el viajero pueden estar influenciadas por las características del destino, en particular de sus atractivos, cuando estos se combinan con las características del alojamiento, y en este sentido el presente estudio tiene carácter novedoso y puede ser de utilidad tanto para los empresarios del sector hotelero como para las autoridades públicas a la hora de planificar correctamente un destino.

De acuerdo con ello, se realizó en primer lugar una caracterización de los dos destinos turísticos del Caribe colombiano mencionados, con el fin de establecer las relaciones existentes entre los rasgos de la oferta de alojamiento y los atractivos de ambas ciudades a través de valoraciones de ambos conjuntos. Posteriormente se aplicó el modelo de árboles de decisión a fin de extraer las variables que influyen en el comportamiento de los clientes. De esta manera,

se espera poder concluir cuáles son las variables del comportamiento del turista que más pesan en cada ciudad y cuáles se relacionan.

Cabe notar además que, si bien esta investigación se añade al amplio corpus de trabajos que tratan sobre preferencias de clientes o comportamiento de turistas, el foco de nuestras pesquisas se pone en la metodología empleada de análisis de contenido y árboles de decisión como elementos innovadores en la indagación. Por tanto, el presente artículo es de naturaleza empírica y el interés se centra en la contribución metodológica.

En cuanto al contexto de los dos destinos estudiados, siguiendo a Bassols (2019); Bassols y Leicht (2020) diremos brevemente que Cartagena de Indias es uno de los destinos más importantes y emblemáticos de Colombia. Posee playas tropicales y patrimonio construido, con lo cual es capaz de atraer a diferentes segmentos de mercado. Rica en arquitectura colonial, su centro histórico se halla en la lista Unesco de Patrimonio de la Humanidad desde 1984. Además, posee una buena infraestructura para congresos y convenciones. Actualmente recibe unos 3 millones de visitantes al año (entre los cuales hay unos 350 mil extranjeros) y tiene una capacidad de alojamiento de unas 14 mil habitaciones

Por su parte, Santa Marta es un destino que cuenta asimismo con una larga tradición de turismo. Es una zona de playas y apta además para el turismo natural gracias al cercano Parque Natural del Tayrona y a la Sierra Nevada de Santa Marta. Posee asimismo una zona balnearia creada hace algunas décadas llamada El Rodadero. Es un destino con menos llegadas y una oferta menor de habitaciones que Cartagena, asimismo con una proyección internacional menor. Por tanto, atiende mayormente a una clientela nacional de tipo familiar que busca mayormente sol y playa, cf. Guardiola (2019).

## II. REVISIÓN DE LITERATURA

Las plataformas de reservas online tienen un fuerte impacto en la gestión de unidades de alojamiento y en las cadenas hoteleras (Lee, Guillet y Law, 2013; Xie, Chen y Wu, 2015). Estas plataformas permiten lograr mejores resultados en ventas, facilitan el intercambio de información por medio de las interacciones de los participantes y dan cabida a mayores volúmenes de transacciones de mercado (Gržinić y Floričić, 2013). Asimismo, dichas plataformas proporcionan información cada vez más importante para la compra de paquetes turísticos, convirtiéndose en factores esenciales de confiabilidad en el destino de cara a visitantes prospectivos (Sparks, Perkins y Buckley, 2013). Por su parte, Liu y Zhang (2014) subrayan la utilidad de estas plataformas e invitan a las empresas del sector a estudiar y usar más los resultados que puedan ofrecer para conocer más de cerca las preferencias del consumidor.

Los comentarios online de clientes juegan pues un papel significativo en la decisión de compra de servicios en línea, principalmente debido a que son un marco en donde los consumidores se intercambian entre ellos información de alto interés. Ello se ha conceptualizado en años recientes con el término eWOM (*electronic word of mouth*), que ha impactado enormemente diferentes ámbitos del consumo de bienes y servicios (Solomon, 2015, cap. 12). En este contexto, Korfiatis y Poulos (2013) demostraron como diferentes tipos de viajeros evalúan cada indicador de calidad de manera distinta, dependiendo más del tipo de viaje en

el que participan y no necesariamente según el segmento demográfico al que pertenecen. Es de destacar asimismo el trabajo de De Ascaniis et al. (2015) quienes analizan comentarios online sobre los hoteles de Lugano (Suiza) concluyendo con algunas recomendaciones para una buena gestión de los mismos. Por otro lado, en América Latina en general, y en Colombia en particular, se ha registrado un lento aumento del comercio electrónico y por debajo de la media mundial (Corrales-Liévano, 2019) con lo que las empresas, debido a factores culturales y de desconocimiento, pierden cuota de mercado. Urge por tanto a los académicos ser propositivos dando herramientas fácilmente aplicables.

Las plataformas de reservas en línea han sido objeto de un buen número de estudios en años recientes (Bowen, 2015; Webb, 2016). Ello es lógico puesto que la información que dan es muy útil para empresarios e investigadores debido a la cantidad de datos que recopilan y el fácil acceso a estos. También estas plataformas permiten evaluar por medio de puntuaciones la percepción de los viajeros, normalmente en forma de escalas de puntuación, como en los casos de [TripAdvisor.com](https://www.tripadvisor.com) y [Booking.com](https://www.booking.com). Destacamos acá los trabajos de Mellinas, Martínez y Bernal, 2015 así como Moya Sánchez y Majó Fernández, 2017: el primero de estos trabajos analiza críticamente el sistema de puntaje de [booking.com](https://www.booking.com), hallando posibles sesgos estadísticos. El segundo descubre y tipifica causalidades entre las valoraciones de usuarios y la reputación de los establecimientos.

Además, es de notar la importancia del análisis de contenido para una gran cantidad de investigaciones surgida en los últimos años y que indaga diferentes aspectos del comportamiento del turista. Así, numerosos trabajos han analizado textos online, sea de páginas web (Zhang et al., 2015; Mohamed et al., 2019) o bien de blogs (Çakmak e Isaac, 2012; Tseng et al., 2015). Más recientemente, se comprueba el aumento de investigaciones basadas en el análisis de textos en redes sociales: Kladou y Mavragani (2015); Chiu y Leng (2017); Garay Tamajón y Cànoves Valiente (2017) o bien Wong y Qi (2017) analizan textos de TripAdvisor, una fuente muy relacionada con [booking.com](https://www.booking.com), aquí usado. Estos trabajos tienen diferentes objetivos cada uno, pero los relaciona el uso de análisis textual como herramienta base de campo, eventualmente como input a un sistema estadístico más sofisticado.

Ante el crecimiento extraordinario de sitios web basados en comentarios (Hlee, Lee y Ko, 2018), numerosos autores afirman que la elección de un servicio o producto online está altamente influenciada por la reputación que tienen los sitios web en cuestión (Moya Sánchez y Majó Fernández, 2017; Vartiak, 2015). En el caso de las plataformas mencionadas, estas ofrecen a los clientes la posibilidad de describir su experiencia como usuarios de un servicio/compradores de un producto, así como de calificar numéricamente su experiencia cliente. Dentro de estas dinámicas, uno de los mercados más influenciados es el de la reserva de viajes por Internet, cuyos ingresos crecieron más del 73% entre el 2007-2012, con un promedio de 148,3 millones de reservas de viajes por año (Pietro, Petrocchi y Spognardi, 2014). En marzo de 2017, Gillian Tans, presidente y CEO de [Booking.com](https://www.booking.com) en la reunión de ITB (feria líder de la industria turística mundial) defendió la idea de que la tecnología móvil y la creciente omnipresencia de los smartphones en todo el mundo continúan cambiando el comportamiento del consumidor y generando nuevas perspectivas, como reveló Tans (comunicación pública): “Según nuestras investigaciones, el 44% de los viajeros ya espera poder planificar sus vacaciones en

pocos pasos en su dispositivo móvil, mientras que el 52% prevé que la utilización de las apps turísticas se incremente este año”.

A la hora de viajar, cada turista tiene diferentes expectativas y/o preferencias, dependiendo del destino, propósito y modo de viaje (Liu et al., 2013; Liu, Shi y Hu, 2013). Es por ello que una comprensión integral de las necesidades de los turistas puede ayudar a la planificación estratégica, mercadotecnia y desarrollo de productos (Wilkins, 2010). Sin embargo, es difícil identificar esta crucial información debido a que la toma de decisiones de un cliente es un proceso muy complejo *per se* en el que influyen numerosas variables. Además, en el sector económico que nos ocupa, existe un amplísimo rango de posibles criterios y variables de clasificación, mucho más amplio de los que usan las plataformas, las cuales tienden a la simplificación de calificaciones y preferencias a unos pocos criterios de valoración (Mellinas, Martínez y Bernal, 2015).

Generalmente se considera que los comentarios sobre hoteles y las calificaciones de los mismos publicadas en los sitios web de viajes impulsan y pueden ayudar a promover las ventas de un establecimiento de alojamiento. Varios estudios también indican que los comentarios y las valoraciones de los hoteles influyen en la elección del cliente (Gržinić y Floričić, 2013; Blal y Sturman, 2014; Hu y Yang, 2020).

En lo que se refiere a estudios con la metodología CHAID y árboles de decisión aplicada al turismo, uno de los primeros trabajos fue el de Byrd y Gutske (2007), dedicado al mapeo de actores. Este algoritmo ha sido usado para hacer pronósticos en el mundo empresarial, y también en las empresas turísticas, por ser esta una de las fortalezas de este instrumento; tal es el caso de los trabajos de Stepchenkova et al. (2019); Kuzey, Karaman y Akman (2019) y Mehmetoglu (2009). En el ámbito de los restaurantes, también los árboles de decisión están a la base del desarrollo de herramientas predictivas, por ejemplo, en Bozkir y Sezer (2011) o en Grinnell y Remus (1983). El modelaje de situaciones donde concurren diferentes variables en gran número es también una de las aplicaciones de este instrumento, véase Moro y Rita (2019) así como Serrano López et al. (2019), ambos artículos con trabajo de campo en Suramérica y por tanto relacionados contextualmente con el presente trabajo. De ahí también la contribución de este artículo al usar una herramienta poco difundida entre la comunidad investigadora de la administración de empresas y la gestión turística.

### III. METODOLOGÍA Y TRABAJO DE CAMPO

En el presente artículo, la herramienta analítica y de visualización gráfica usada son los árboles de decisión CHAID. Se erige en una herramienta clara y potente, proveniente de aplicaciones de ingeniería, pero que hasta la fecha ha sido muy poco usada en estudios empresariales de tipo turístico. Teniendo presente que se busca establecer las relaciones entre las valoraciones en línea, las características de la oferta hotelera y los rasgos del destino para diferentes segmentos turísticos, es importante aclarar el concepto de “árboles de decisión”. Estos son un modelo de predicción cuyo objetivo principal es el aprendizaje inductivo a partir de observaciones y construcciones lógicas. Se trata de un procedimiento similar a los sistemas

de predicción basados en reglas, los cuales se utilizan para representar y categorizar una serie de condiciones que acontecen de forma sucesiva para posibilitar la solución de un problema.

El algoritmo que está a la base de los árboles de decisión CHAID usa el análisis Chi-cuadrado, obteniendo de este modo como variable de destino una variable dependiente nominal que puede ser de dos o más categorías. Es así como se pueden utilizar variables de entrada que serán independientes (o predictoras) nominales o métricas y que explican el comportamiento de la dependiente. Es imprescindible para este estudio calcular el estadístico chi-cuadrado y comparar las significaciones obtenidas entre la dependiente y las independientes. De este modo, se puede entender también esta técnica como la detección automática del Chi-cuadrado de interacción, cuyas siglas en inglés de este proceso le dan justamente este nombre (Chi-square Automatic Interaction Detector). Las variables cuya hipótesis de dependencia quede corroborada serán las seleccionadas para el análisis CHAID, mientras que el resto quedarán descartadas, y por tanto no tendrán influencia sobre el fenómeno estudiado. Una característica esencial de esta técnica es que no es paramétrica (contrariamente a otras alternativas, como el análisis regresivo múltiple). Para más información, cf. Escobar (1998); Román-González y Lévy-Mangin (2003).

Una gran ventaja del uso del algoritmo CHAID es que permite ser visualizado en forma de árboles de decisión, es decir, los resultados obtenidos por inducción se convierten en un conocimiento a representar en forma arbórea, con 'nodos', 'ramas' y 'hojas', de ahí el nombre de este instrumento de visualización. Barrientos Martínez et al. (2009: 20) explican su funcionamiento de forma sencilla y escueta:

*El nodo principal o raíz es el atributo a partir del cual se inicia el proceso de clasificación; los nodos internos corresponden a cada una de las preguntas acerca del atributo en particular del problema. Cada posible respuesta a los cuestionamientos se representa mediante un nodo hijo. Las ramas que surgen de cada uno de estos nodos se encuentran etiquetadas con los posibles valores del atributo [...]. Cada vez que se ejecuta este tipo de modelo, un único camino será seguido, dependiendo del valor actual de la variable evaluada. Los valores que pueden tomar las variables para este tipo de modelos pueden ser discretos o continuos.*

Siguen estos autores afirmando que el modelo se construye a partir de la descripción narrativa de un problema determinado, a cuya solución se aporta mediante la visualización gráfica para posibles tomas de decisión, especificando las variables que son evaluadas, las acciones que deben ser llevadas a cabo y el orden en el que se ejecutarán a la hora de tomar una decisión determinada.

La presente investigación emplea árboles de decisión tipo CHAID, que es la forma original de este instrumento. Este tipo de árbol divide a la población en dos o más grupos distintos basados en las categorías de la variable dependiente, que es un mejor predictor. Posteriormente, se vuelve a dividir cada uno de estos grupos en subgrupos más pequeños basándose en otras variables predictoras. El proceso de partición termina hasta que no se encuentran variables que produzcan segmentos significativos. El sistema muestra los

segmentos resultantes en un gráfico de forma arbórea visual y comprensible. Los segmentos que se construyen son mutuamente excluyentes, es decir, los segmentos no se superponen (un individuo no puede pertenecer a dos o más segmentos) y exhaustivos (un individuo debe pertenecer siempre a un segmento, no pueden existir individuos aislados) (Magidson, 1993).

Por otro lado, en lo que se refiere a estudios de mercado, y de acuerdo a Pesonen (2014), son posibles dos formas de segmentación: a priori y a posteriori. Esta última es más aplicable en casos de desconocimiento de las necesidades de los consumidores o de los servicios que están buscando al momento de tomar una decisión y es por ello que los autores del presente artículo se han decantado por esta última. La segmentación a posteriori es también mencionada en la literatura como 'segmentación Post Hoc' (Wedel y Kamakura, 2000). Tal y como sugieren para estos casos Ferreira, Rial y Valera (2009), de esta forma los segmentos resultantes estarán constituidos por consumidores de mayor homogeneidad entre sus necesidades/preferencias y mayor heterogeneidad entre grupos que las que podrían obtenerse mediante una segmentación a priori. Para realizar un proceso de segmentación es necesario contar con un base de datos donde se identifiquen las variables a estudiar, y utilizar escalas (de puntos) para medir la importancia que tiene cada una de las variables para el estudio a realizar. En la actualidad, para facilitar el trabajo de los investigadores en la segmentación Post Hoc, se está manejando el software SPSS (Statistical Product and Service Solutions) que permite experimentar con las variables de modo sencillo y directo para así poder generar diferentes clústeres para el análisis de los datos. Por ello se eligió dicho software como apoyo informático de la presente investigación.

En el último párrafo de la sección anterior, quedó clara la utilidad del método CHAID en entornos con muchas variables o cuando se necesita una respuesta predictiva sólida. Dado que en el presente trabajo existe una gran cantidad de variables (Tabla 1) y el objetivo final es entender comportamientos de turistas para predecirlos en el futuro, esta técnica es adecuada a los propósitos investigativos del presente artículo. Por lo demás, el algoritmo, acoplado a una potente herramienta visual como son los árboles de decisión, permite entender las problemáticas y soluciones relevantes de un modo rápido y ágil.

En cuanto a la recolección de la información, se utilizó el método de recogida de datos no estructurados, es decir, los comentarios que figuran en [booking.com](https://www.booking.com) fueron recopilados manualmente y procesados en hoja de cálculo de MS Excel. Del mismo modo se procedió para los datos básicos de los hoteles, lo que permitió más rapidez y obtener de una única fuente toda esta información de forma estandarizada. Debido a la cantidad relativamente grande de datos tratados para elaborar el presente artículo (669 comentarios con una media de 7 palabras por comentario), hubo que elaborar largas tabulaciones para lograr segmentar correctamente los dos destinos turísticos estudiados. Los datos recolectados fueron de dos tipos: por un lado, los comentarios de [booking.com](https://www.booking.com), los cuales arrojaron un total inicial de 18 características o variables. Por otro lado, se recopiló la información de los establecimientos de alojamiento. Estos últimos arrojaron un total inicial de 29 características. Ambos grupos se presentan en la Tabla 1.

Tabla 1. Bases de datos recolectados y sus categorías

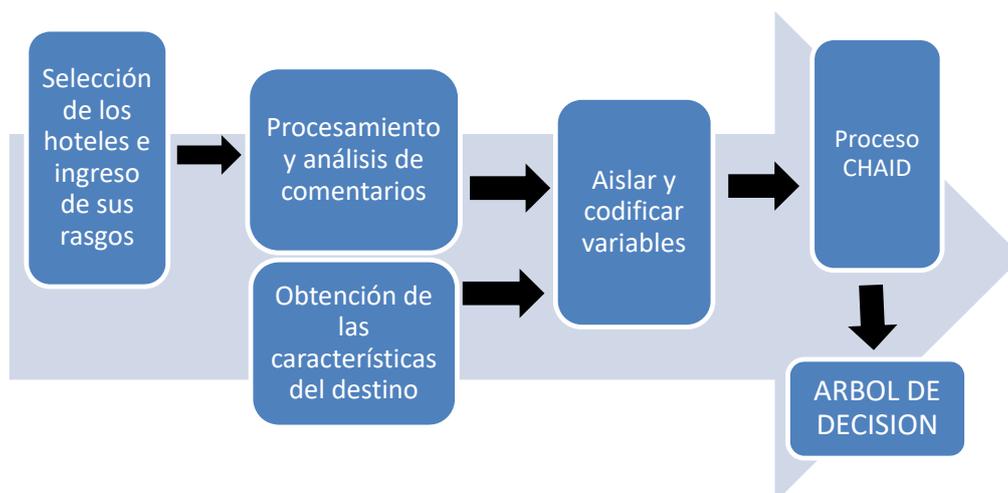
Origen datos: <a href="https://www.booking.com">booking.com</a>	Categorización de datos
Base de datos de los comentarios de clientes	Ciudad de destino del cliente, País de origen del cliente, Comentarios cliente, Tipo de turista, Tipo de destino, Viaja con, Tipo de habitación, Valoración del cliente, Nivel de satisfacción, Total noches, Fecha comentario, Categoría del hotel en estrellas, Servicios, Información habitaciones, Tipo alojamiento, Nombre alojamiento, Zona de ubicación del alojamiento y Modo de pago del alojamiento
Base de datos de los hoteles	Ciudad de destino, Zona de ubicación alojamiento, Tipo de alojamiento, Nombre del alojamiento, Categoría el hotel en estrellas, Valoración hotel, Tipo de habitación, Valor de la oferta, Desayuno incluido, Baño privado, Aire acondicionado, Equipamiento audiovisual, Wifi, Gimnasio, Zona de estar, Traslado al aeropuerto, Restaurante, Piscina, Parking gratis, Mascotas, Exteriores, Centro de negocios, Bar, Lavandería, Cantidad de servicios ofrecidos por el establecimiento, Idiomas, "Les han encantado los restaurantes de la zona"; "Les encantó el estilo boutique" y Reserva segura

Elaboración propia.

En la base de datos de hoteles, se ingresaron los rasgos de 64 establecimientos hoteleros de Santa Marta y de 89 establecimientos de Cartagena, por un total de 153 hoteles. Estos establecimientos fueron elegidos aleatoriamente, teniendo en cuenta que estuvieran en diferentes ubicaciones en ambas ciudades y que en el conjunto final hubiera también una representación de diferentes categorías hoteleras. La caracterización de los destinos se hizo en base a sus rasgos geográficos y de distribución de hoteles y atractivos, según Google Maps.

En la base de datos de comentarios de clientes se recogieron de forma aleatoria comentarios efectuados sobre los hoteles anteriores en el periodo comprendido entre el 1 de enero de 2016 y el 15 de abril de 2017 en plataforma [Booking.com](https://www.booking.com). Dicha información se desgajó de la siguiente forma: 335 comentarios de establecimientos hoteleros de Santa Marta y 334 comentarios sobre los hoteles de Cartagena, elevando a un total de 669 los comentarios ingresados a base de datos. De estos, se conservaron 300 para cada una de las ciudades y con los otros 69 comentarios restantes se realizó la simulación de datos con el fin de evaluar el desempeño de los árboles (Véase Anexo I). El análisis de este último bloque de 69 comentarios permitió establecer las diferentes relaciones que posiblemente se pudieran dar entre las variables recolectadas y así poder obtener cálculos confiables para contar con un resultado más preciso (véase Anexo I para la demostración de la validez de desempeño). Según lo anterior, se tuvo que separar de la base de datos las dos ciudades estudiadas para poder disminuir la cantidad de datos ingresados al software SPSS y de esta manera poder identificar la variable dependiente y las variables independientes. Como dato curioso, señalamos que los comentarios son cortos, de unas 7 palabras de media. La Figura 1 resume todo el proceso de trabajo de campo para la obtención de datos.

Figura 1. Proceso metodológico completo



Elaboración propia.

Para la presentación de los resultados obtenidos, la variable dependiente se denomina *Valoración (de los huéspedes)*, las variables independientes son: Acompañantes (o Grupo de Viaje), Origen del viajero, Categoría de los Hoteles, Atractivos, Servicios Ofrecidos, Hábitos de Compra y Mejor zona de alojamiento. En la Tabla 2 se pueden observar ejemplos de la codificación de algunas variables nombradas.

Tabla 2: Ejemplos de codificación de variables

VARIABLE	CONCEPTO	CODIFICACIÓN
VALORACIÓN	> 9 X ≤ 10, Sumamente satisfecho	5
	> 7 X ≤ 9, Muy satisfecho	4
	> 5 X ≤ 7, Satisfecho	3
	> 3 X ≤ 5, Poco satisfecho	2
	> 0 X ≤ 3, Nada satisfecho	1
ACOMPAÑANTE	Grupo	4
	Familia	3
	Pareja	2
	Solo	1
PROCEDENCIA	Resto del Mundo	3
	América Latina	2
	Colombia	1
CATEGORÍA HOTEL	5 Estrellas	5
	4 Estrellas	4
	3 Estrellas	3
	2 Estrellas	2
	Sin categorizar	0

Elaboración propia.

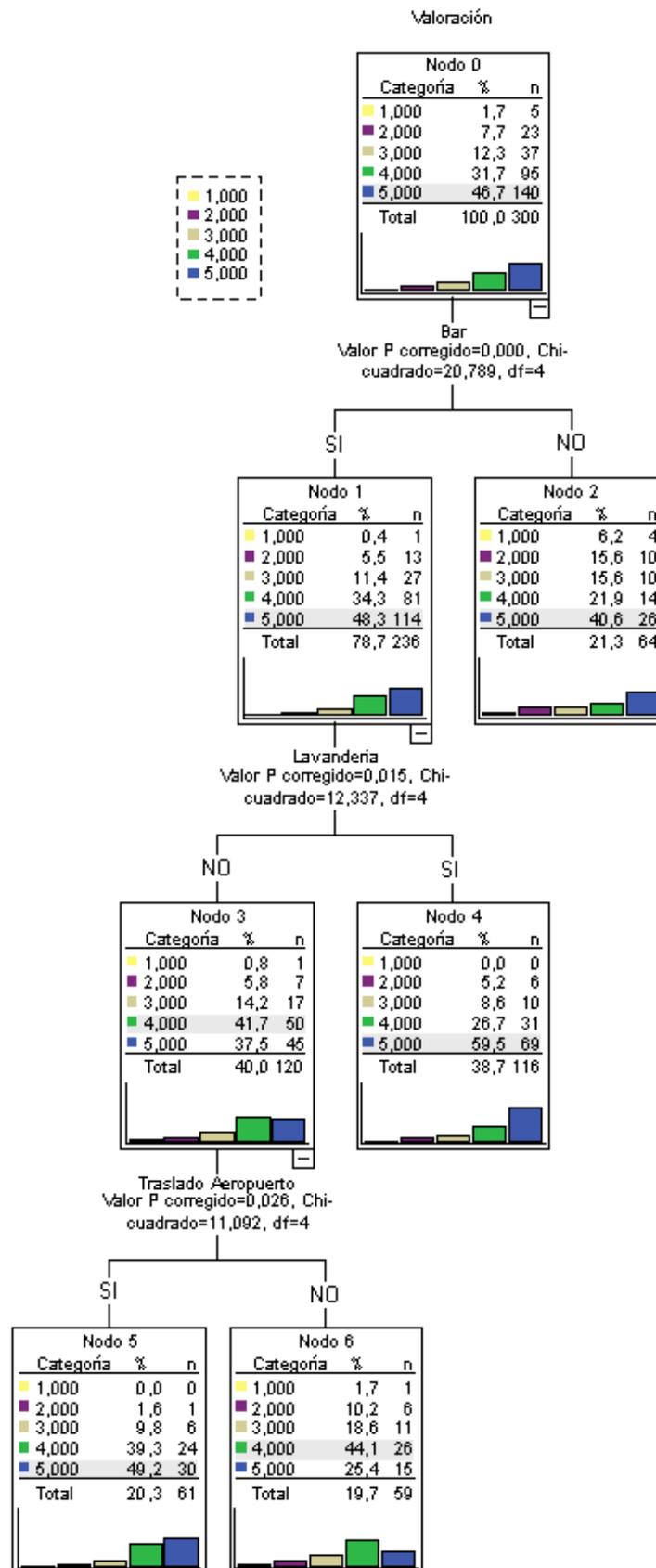
#### IV. RESULTADOS

A continuación, se muestran los hallazgos del trabajo de campo. Están presentados en forma de diez 'casos' visualizados con árboles de decisión (en el supuesto de darse una ramificación y construcción de árbol para un caso determinado), e interpretados cada uno con un texto de apoyo que analiza y contextualiza los resultados. En cada uno de estos casos, la variable dependiente (que, como se ha dicho, es la valoración numérica del hotel que realizan los huéspedes) es analizada con diferentes variables independientes, solas o agrupadas.

De los múltiples casos posibles, se presentan 5 casos para Santa Marta y 5 para Cartagena, que fueron los que arrojaron resultados más significativos según el software SPSS. Dichos casos analizan (1) los segmentos psicográficos (que incluyen fundamentalmente las variables de servicio); (2) los segmentos demográficos (que incluyen variables como el origen del huésped o el grupo de viaje); (3) los segmentos geográficos (obtenidos al ubicar los establecimientos en las diferentes zonas turísticas de cada destino); (4) los hábitos de compra (que son esencialmente las preferencias de los clientes como el modo de pago, el tipo de habitación seleccionada o la categoría del hotel elegido); y finalmente (5) los atractivos turísticos (obtenidos al clusterizar los diferentes atractivos en los dos ámbitos urbanos estudiados).

En los casos 1 y 2 la explicación textual es profusa, con el objetivo de guiar a los lectores que no conocen el formalismo. En los restantes casos, la descripción es más liviana para una lectura más rápida. Por otro lado, para dichos casos 1 y 2 se ofrece además una validación de desempeño del árbol para confirmar la fiabilidad de los datos, el lector interesado la hallará en el Anexo I.

### 4.1. Caso 1: Gráfico “Valoración numérica vs segmentos psicográficos Santa Marta”



### Interpretación:

Partiendo del Nodo 0 la descripción de la variable dependiente *Valoración*, entendida como “valoración numérica emitida por los huéspedes en [booking.com](https://www.booking.com)” (ver también Tabla 2, donde se listan ejemplos de codificación), el análisis de los 300 comentarios sobre los hoteles de los hoteles de la ciudad de Santa Marta arroja que 46.7% corresponde a la valoración 5 “sumamente satisfecho”, 31.7% corresponde a la valoración 4 “muy satisfecho”, 12.3% corresponde a la valoración 3 “satisfecho”, 7.7% corresponde a la valoración 2 “poco satisfecho” y por último 1.7% corresponde a la valoración 1 “nada satisfecho”.

Posteriormente, el árbol se ramifica en dos nodos: Nodo 1 (Cuenta con servicio de bar) y nodo 2 (No cuenta con servicio de bar) pertenecientes a la variable del servicio con mayor relevancia en los hoteles estudiados de la ciudad de Santa Marta, con el Chi-cuadrado de 20.789 indicando que esta es la variable principal predictora. Para el caso afirmativo (Nodo 1) la mayor valoración es de 5 “sumamente satisfecho”, con un 48.3% de las valoraciones contra el Nodo 2 que es el negativo pero que sólo suma un 40.6% de las valoraciones en 5 “sumamente satisfecho”. Por ello, se puede afirmar que los clientes de la ciudad de Santa Marta prefieren los hoteles con servicio de bar.

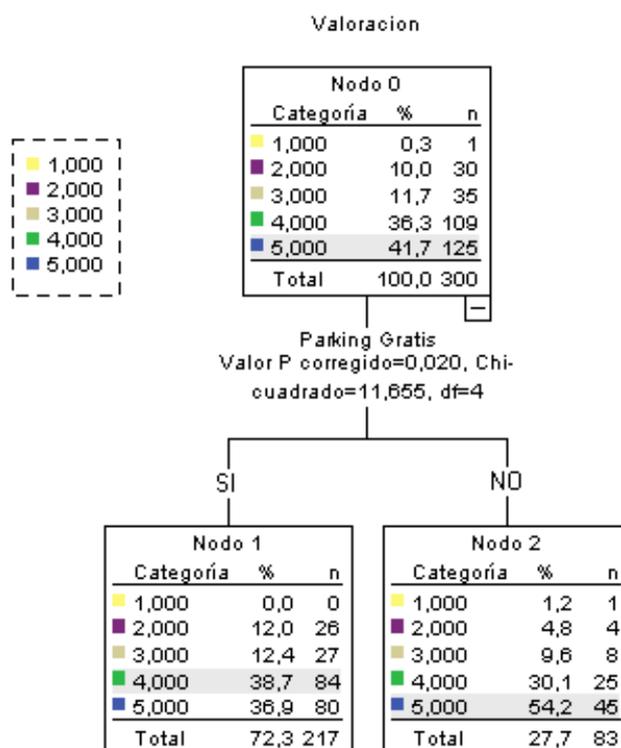
El Nodo 1 se vuelve a ramificar en Nodo 3 (No cuenta con el servicio de lavandería) y Nodo 4 (Cuenta con servicio de lavandería) dando relevancia a esta variable con un Chi-Cuadrado de 12.337. Seguidamente, en el Nodo 4 podemos observar que la valoración que predomina es la valoración 5 “sumamente satisfecho” con un 59.5% ante la valoración del Nodo 3 que es de 41.7% para la valoración 4 “muy satisfecho”.

El Nodo 3 se ramifica en Nodo 5 (Cuenta con servicio de traslado al aeropuerto) y Nodo 6 (No cuenta con servicio de traslado al aeropuerto). La relevancia de esta variable viene dada por su Chi-cuadrado de 11.092 y en este caso, se observa que el Nodo 5 tiene la mayor valoración con un 49.2% “sumamente satisfecho” en 5 mientras que el Nodo 6 queda por detrás con un 44.1% de valoraciones en 4 “muy satisfecho”.

Por lo tanto, a modo de conclusión de este primer caso de valoración numérica vs segmentos psicográficos de Santa Marta, las variables más relevantes al momento de escoger un hotel por sus servicios son: servicio de bar, lavandería y traslado al aeropuerto. Sin embargo, visto todo el árbol, se destaca que hay dos tipos de cliente: uno es el que aprecia el servicio de bar y de lavandería, pero no el traslado al aeropuerto, y otro perfil de cliente es el que aprecia servicio de bar y el traslado al aeropuerto, pero le es indiferente el servicio de lavandería.

La explicación del párrafo anterior indica que la ciudad de Santa Marta recibe turistas del segmento familiar de ocio y por esta razón la variable del servicio de bar es la más relevante en el hotel ya que los viajeros en su estadía buscan contar con entretenimiento sin necesidad de salir de su hospedaje. Ello se correlaciona con el hecho de que los turistas que visitan este destino no tienen en cuenta los restaurantes que se encuentran en el entorno de sus alojamientos, según nuestro trabajo de campo. Que les interese o bien el traslado al aeropuerto o bien el servicio de lavandería significa que de igual modo el turista demanda unos servicios que fundamentalmente le brinden comodidad y facilidad en su estadía.

#### 4.2. Caso 2: Gráfico “Valoración numérica vs segmentos psicográficos Cartagena”



#### Interpretación:

Partiendo del Nodo 0 la descripción de la variable dependiente *Valoración* (se entiende las valoraciones numéricas de los huéspedes) de los 300 comentarios evaluados de los hoteles de la ciudad de Cartagena, el 41.7% corresponde a la valoración 5 “sumamente satisfecho”, 36.3% corresponde a la valoración 4 “muy satisfecho”, 11.7% corresponde a la valoración 3 “satisfecho”, 10% corresponde a la valoración 2 “poco satisfecho” y por último 0.3 % corresponde a la valoración 1 “nada satisfecho”.

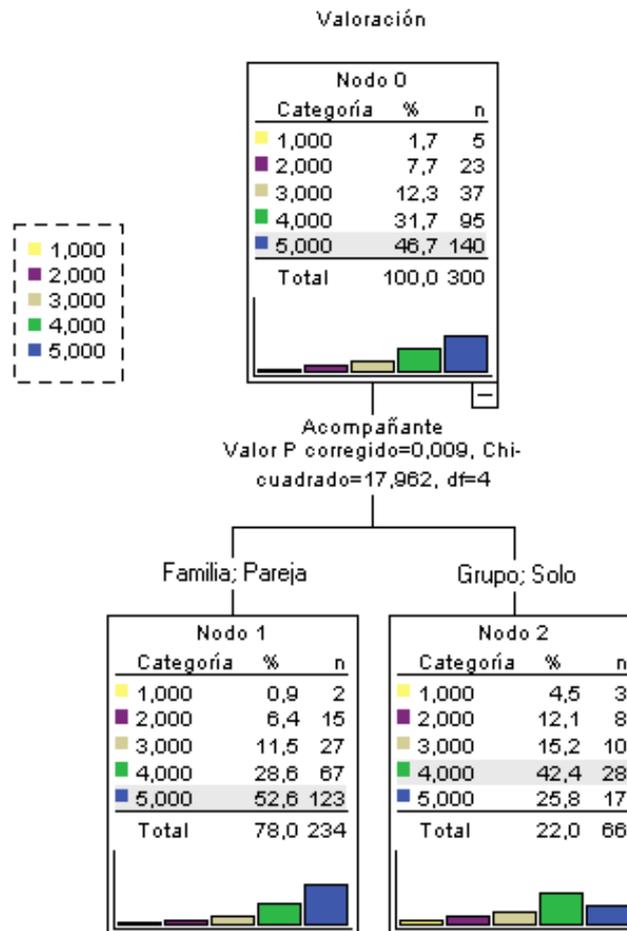
Posteriormente, el árbol se ramifica en dos nodos: Nodo 1 (Cuenta con servicio de parking gratis) y Nodo 2 (No cuenta con servicio de parking gratis) pertenecientes a la variable del servicio con mayor relevancia en los hoteles estudiados para la ciudad de Cartagena, indicando que esta es la variable principal predictora, ya que su Chi-Cuadrado fue 11.655. Seguidamente se observa que la valoración del Nodo 2 es la más relevante por su valoración de 5 “sumamente satisfecho” con un 54.2%.

Por lo tanto, a modo de conclusión del 2do caso de valoración numérica vs segmentos psicográficos Cartagena, la variable más relevante al momento de escoger un hotel por sus servicios viene caracterizada por el Nodo 2: no influye que el hotel cuente con servicio de parking gratis.

Este resultado puede explicarse porque Cartagena atrae a turistas colombianos que no llegan con transporte propio sino por avión, bus, etc. Asimismo, la absoluta mayoría del creciente segmento de turistas extranjeros llega a la ciudad por vía aérea. Y es poco probable

que los turistas de congresos y convenciones, nacionales o extranjeros, lleguen a la ciudad con vehículo propio. En este contexto, hay que recordar que el aeropuerto de Cartagena muestra un número creciente de conexiones directas nacionales e internacionales. Así, el servicio de parking gratis no interesa a un amplio sector de turistas que visitan este destino. Ello es una buena noticia para los empresarios hoteleros de la ciudad, puesto que, por su peculiar topografía y configuración urbanas que limitan el espacio disponible, Cartagena no tiene muchos espacios para construir parqueaderos.

#### 4.3. Caso 3: Gráfico “Valoración numérica vs segmentos demográficos Santa Marta”



Interpretación:

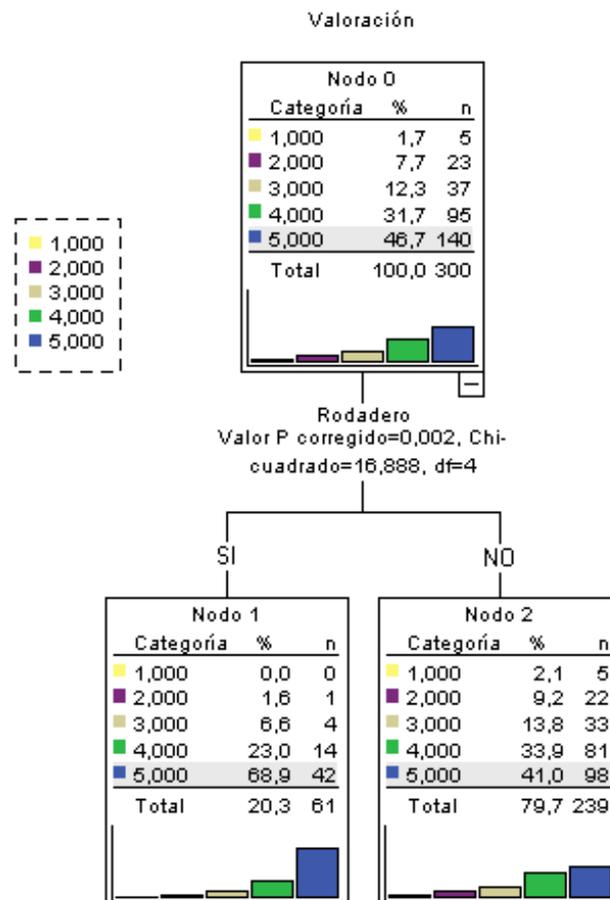
En el presente caso de “Valoración numérica vs segmentos demográficos Santa Marta”, la variable más relevante al momento de escoger un alojamiento por los posibles grupos de viaje es el Nodo 1. Es decir, los hoteles de Santa Marta estudiados alojan a Parejas y Familias; ello concuerda con el hecho sabido de que Santa Marta es un destino que acoge sobre todo a turistas vacacionales procedentes del mercado interior colombiano. Este hecho se confirma en el estudio de SITUR Magdalena (la agencia local para estadísticas de turismo) en su informe de

2018 cuando afirma que los dos grupos que más llegan a la ciudad son, por este orden, parejas y familias con diversos miembros (SITUR Magdalena, 2018: 34).

#### 4.4. Caso 4: “Valoración numérica vs segmentos demográficos Cartagena”

En el presente caso el sistema no arrojó ninguna variable predictora y por ello no se muestra ningún gráfico. Es decir, no influye lo suficiente el grupo de viaje al momento de escoger un hotel por parte de los turistas que llegan a Cartagena. Ello hace de Cartagena un destino muy variado en lo que se refiere a la demanda, que acoge todo tipo de grupos de visitantes, desde solos o parejas hasta familias y grupos. Tal variedad de segmentos confirma resultados de estudios anteriores y es un gran reto para las empresas locales y las autoridades públicas a la hora de proponer servicios y planificar un espacio en el que segmentos de demanda muy diferentes puedan convivir y tener experiencias satisfactorias. Ello es un reto también para las empresas hoteleras generalistas presentes en el destino, que en ocasiones deberán alojar en un mismo establecimiento a segmentos muy variados. Pero es también una oportunidad para que surjan tipos de alojamiento o servicios especializados que buscan su propio segmento de mercado, y de hecho esto explica la proliferación de hostels, boutique hotels, etc. en Cartagena en los últimos años.

#### 4.5. Caso 5: Gráfico “Valoración numérica vs segmentos geográficos Santa Marta”



### Interpretación:

Partiendo del Nodo 0, el árbol se ramifica en dos nodos: Nodo 1 (Rodadero) y Nodo 2 que agrupa al resto de zonas consideradas por no tener cada una de ellas suficiente peso por sí misma (Centro Histórico, Bello Horizonte y Pozos Colorados). Con un Chi-cuadrado de 16.888, el Nodo 1 es el de mayor peso, con un 68.9% de valoración 5 “sumamente satisfecho”. Nótese la diferencia con el Nodo 2, cuyo máximo valor es de 41% para la valoración 5 “sumamente satisfecho”.

Por lo tanto, a modo de conclusión del 5to caso “Valoración numérica vs segmentos geográficos en Santa Marta”, de los datos recopilados de [booking.com](https://www.booking.com) la zona más relevante al momento de escoger un hotel es el sector del Rodadero. Este sector es el balneario playero por excelencia de la ciudad y ubicado relativamente cerca de la misma, con buena recomendación en TripAdvisor y además (según nuestro estudio) posee la mitad de la oferta de habitaciones de la ciudad y muestra unos precios más económicos que las zonas agrupadas el Nodo 2. La importancia de este sector más ‘envejecido’ que los otros y algo necesitado de renovación queda confirmada, así como el hecho de que el turista que llega a Santa Marta es de un potencial económico más bien limitado.

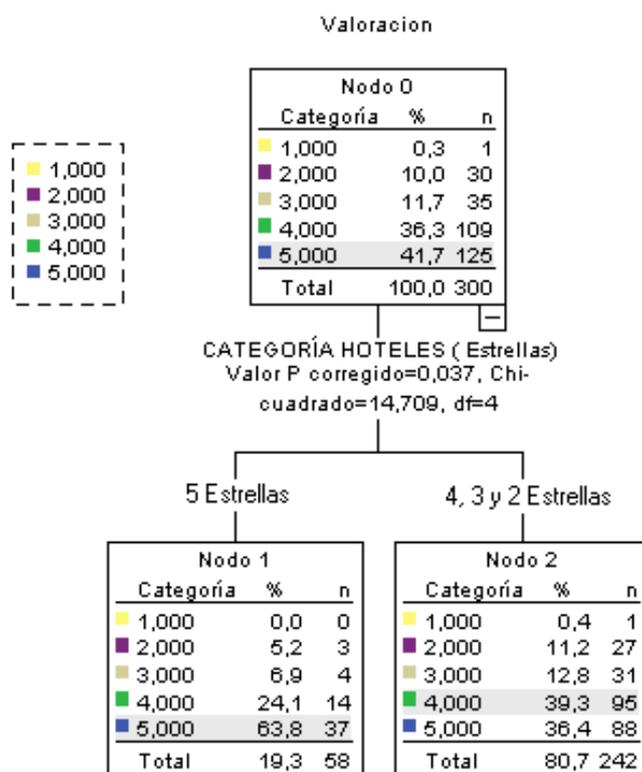
#### **4.6. Caso 6: “Valoración numérica vs segmentos geográficos Cartagena”**

En el presente caso 6 el sistema no arrojó variable predictora alguna, es decir, el sistema no dio con ninguna variable que influyera lo suficiente sobre la preferencia de la ubicación del alojamiento por parte de los turistas que visitan la ciudad de Cartagena. Esto es una conclusión interesante: a pesar de que el destino tiene diferentes zonas, cada una de ellas muy marcada y diferenciada (sector Centro Histórico, sector de playas en Bocagrande y la Zona Norte) en realidad las diferentes zonas turísticas no destacan por sí solas y son percibidas por la demanda como un único conjunto. Para los empresarios, esto quiere decir que la zona urbana donde se abra un alojamiento no será un factor que contribuya a su diferenciación, pudiendo invertir más en otras características del establecimiento. Ello se correlaciona con el hallazgo de campo de valoraciones de hoteles, el cual indica que, prácticamente en todos los sectores de la ciudad, la valoración de los hoteles es uniformemente muy alta.

#### **4.7. Caso 7: “Valoración numérica vs segmentos hábitos de compra Santa Marta”**

Para el presente caso “Valoración numérica vs segmentos hábitos de compra Santa Marta”, tampoco el sistema arrojó ninguna variable predictora, es decir, no influyen argumentos de venta como una habitación en especial o la cantidad de noches de estadía en el alojamiento. Ahí se presenta un nuevo reto para las empresas de Santa Marta, y es cómo llamar la atención con mejores habitaciones o bien ofertando precios para obtener estadías más largas o bien con otras mejoras en el producto/servicio, siendo que estos factores no influyen en el comportamiento del cliente en Santa Marta en la actualidad. Se trata además de un potencial problema a la hora de obtener mayores ingresos para la empresa, y se confirma que el cliente prototipo de Santa Marta es un cliente con poder adquisitivo limitado.

#### 4.8. Caso 8: Gráfico “Valoración numérica vs segmentos hábitos de compra Cartagena”

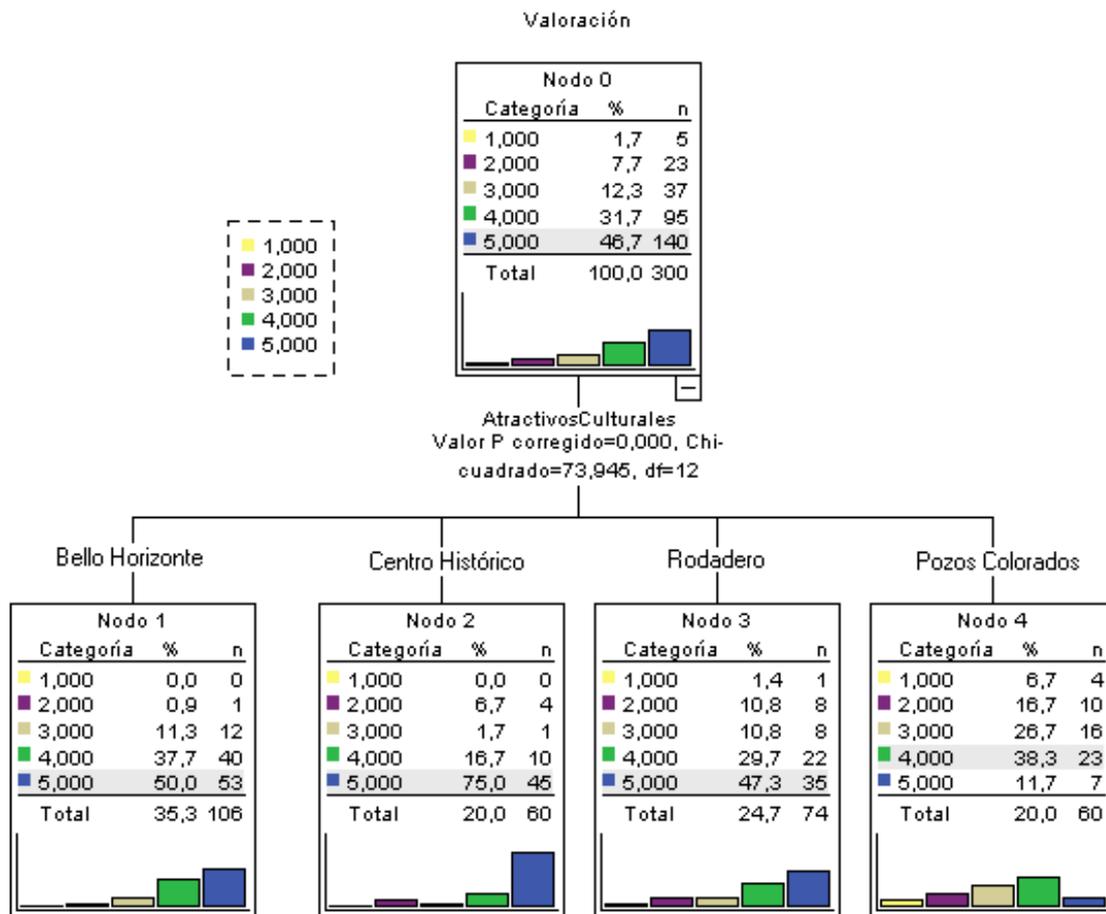


#### Interpretación:

Partiendo del Nodo 0, el árbol se ramifica en dos nodos: Nodo 1 y Nodo 2 pertenecientes a la variable independiente de *Categoría de los hoteles (Estrellas)*, indicando que esta es la variable principal predictora, ya que su Chi-Cuadrado fue 14.709. A continuación, se observa que el Nodo 1 es el más relevante ya que su valoración mayor es de 5 “sumamente satisfecho” con un 63.8% frente al Nodo 2, cuya valoración mayoritaria es de 4 “muy satisfecho” en un 39,3%, siendo el Nodo 1 el que se refiere a la categoría de hoteles de 5 estrellas.

Por lo tanto, a modo de conclusión de valoración numérica vs segmentos hábitos de compra de Cartagena, la variable que define esencialmente el hábito de compra de los turistas que visitan la ciudad de Cartagena es la *Categoría de los hoteles (Estrellas)* donde los establecimientos más relevantes son los de 5 estrellas. Ello confirma la imagen de Cartagena como ‘gran’ destino que atrae a una clientela de gama superior a la de Santa Marta. Por tanto, el destino es un buen lugar para invertir en establecimientos de gama alta, y de hecho este hallazgo es coherente con la creciente oferta de alojamientos de lujo que ha experimentado la ciudad en los últimos años.

#### 4.9. Caso 9: Gráfico “Valoración numérica vs atractivos turísticos Santa Marta”



#### Interpretación:

Partiendo del Nodo 0, el árbol se ramifica en cuatro nodos: Nodo 1, Nodo 2, Nodo 3 y Nodo 4 pertenecientes a la variable independiente de zonas de alojamiento de los hoteles estudiados (Bello Horizonte, Centro Histórico, Rodadero y Pozos Colorados, respectivamente) indicando que esta es la variable principal predictora, ya que su Chi-Cuadrado fue 79.945. A continuación, observamos el Nodo 2 que es el más relevante, ya que posee la valoración 5 “sumamente satisfecho” en un 75%. De este modo, el Centro Histórico (una zona que en los últimos años se ha rejuvenecido y ha aumentado su valor como área de ocio y cultura) se configura como una zona de atracción primaria cuando los clientes buscan centralidad y cercanía a los atractivos de la ciudad.

Para la recopilación de la información sobre este Caso 9 y el siguiente Caso 10 se utilizó la herramienta de Google Maps, a través de la cual los hoteles de la base de datos y los atractivos de la ciudad fueron ubicados con el fin de conocer la cantidad de atractivos culturales y naturales que tienen en su alrededor y de esta manera poder dividirlos en zonas.

#### 4.10. Caso 10: “Valoración numérica vs atractivos turísticos Cartagena”

En el presente caso de valoración numérica vs atractivos turísticos Cartagena, el sistema no arrojó ninguna variable predictora, es decir no influye la cantidad de atractivos culturales y naturales en los alrededores de las principales zonas de alojamiento. Por tanto, ninguna zona de la ciudad se destaca sobre las demás según este caso. Se ve asimismo una correlación con el Caso 6 (segmentos geográficos de Cartagena) en donde queda confirmado que el factor ubicación no es de interés para los clientes de los hoteles cartageneros, lo cual puede permitir a los inversores no considerar las zonas “premium” del destino por ser cercanas a los atractivos y centrarse en zonas más periféricas diferenciando sus establecimientos con otros atributos.

## V. DISCUSIÓN Y CONCLUSIONES

En la presente investigación se estableció la relación que tienen las variables de valoración, oferta (de hoteles y destinos) y segmentación con el fin de poder indagar las motivaciones y preferencias de los visitantes de Santa Marta y Cartagena, según consignado en la plataforma [booking.com](https://www.booking.com). La información recopilada de dicha plataforma permitió describir, analizar, sectorizar y segmentar de una mejor manera el turismo de las dos ciudades turísticas principales del Caribe colombiano y por medio de los árboles de decisión fue posible relacionar las variables dependientes e independientes más relevantes de los datos recolectados.

Después de analizar todas las relaciones posibles que pueden existir entre la variable dependiente y las variables independientes estudiadas podemos observar que, para los huéspedes que escogen sus alojamientos en [Booking.com](https://www.booking.com), variables como la categoría de establecimiento medida en estrellas, los acompañantes del viaje o los servicios facilitadores del hotel están entre las más importantes al momento de decidir la compra. Es interesante además destacar que variables como el origen geográfico del huésped (a la que normalmente se atribuye una cierta importancia en estudios de segmentación de mercados) no resulta relevante en ninguno de los dos destinos del presente caso de estudio.

La imagen final que se desprende de cada uno de los destinos es bien diferente. Cartagena se configura como destino de talla internacional, con grupos de viaje heterogéneos y una clientela multisegmento de buen poder adquisitivo que percibe toda la ciudad de modo holístico, por tanto, ubicación del hotel y atractivos cercanos no influyen en su decisión de compra, mientras que sí lo hace la categoría en estrellas del hotel. Para los actores del turismo cartagenero, el reto es ver cómo la ciudad puede continuar siendo un destino polivalente dando a cada segmento que atrae lo que éste busca y haciendo de la ciudad un espacio en donde diferentes tipologías de visitantes puedan convivir tranquilamente.

Santa Marta en cambio se muestra como un destino menor, dirigido a un público nacional con menos poder adquisitivo, que se interesa por tenerlo todo a mano y busca establecimientos con servicios que les hagan la estadía ‘fácil’. Este turista viaja en pareja o familia y sobre todo busca playa. A pesar de los esfuerzos hechos en los últimos años para dirigirse a un cliente de más alta gama, Santa Marta sigue siendo percibida como un mero destino de sol y playa. Por ello, sus visitantes buscan la cercanía al mar en un alojamiento económico, y esto se lo ofrece El Rodadero, un balneario que hace décadas fue el origen del turismo samario pero

que hoy en día se halla ya en su fase de madurez. Es interesante sin embargo ver una nueva tendencia que emerge y muestra cómo el Centro Histórico es considerado la zona primaria de elección por parte de quienes buscan la cercanía a los atractivos de la ciudad en el momento de elegir alojamiento. Este aumento de atractividad en esta zona modera lo que se acaba de decir sobre El Rodadero. Por tanto, los actores en el destino deberán de monitorizar esta situación para ver si, a medio plazo, el Centro Histórico puede ser una alternativa de rejuvenecimiento al Rodadero.

El presente análisis permite aportar a los hoteles estudiados de [Booking.com](https://www.booking.com) del Caribe colombiano el conocimiento sobre las principales necesidades y motivaciones de los clientes que priman al momento de escoger un alojamiento. Con este conocimiento de cliente, los hoteles de estas ciudades podrán trabajar en mejoras e implementar estrategias para aumentar la cantidad de huéspedes que reciben y su satisfacción con el fin de generar más ingresos y fidelizarlos.

La metodología usada en el presente artículo puede ser útil tanto para confirmar resultados operacionales que se saben como otros que se desconozcan (cf. Sección IV), es decir, es una herramienta cuyos resultados pueden ser tanto la confirmación de informaciones como el descubrimiento de nuevas tendencias de la demanda. En cualquier caso, es una herramienta mucho más económica y eficiente que las herramientas 'al uso', si se observa la relación costo/cantidad de información obtenida. En el caso mencionado del SITUR Magdalena, este proyecto usó mucho más personal y presupuesto que el que acá se reporta, para unos resultados comparables.

El lector atento se habrá percatado de que la metodología empleada de análisis de contenidos online combinada con árboles de decisión puede ser aplicada en muchos otros contextos de la relación cliente-empresa. Más allá de las empresas hoteleras estudiadas en dos destinos determinados, lo expuesto aquí es válido para muchos otros sectores de la economía, sea para productos o para servicios, y esperamos que en el futuro más investigaciones de otros sectores empresariales con esta herramienta puedan producir resultados interesantes.

La limitación mayor del presente trabajo ha sido el no poder ofrecer un árbol de decisión que abarque a ambos destinos ya que cuando se experimentó con el software a nivel de todos los datos obtenidos de ambos destinos, no se obtuvieron resultados interesantes. Se trata de una limitación de la metodología, ya que el sistema en ocasiones es incapaz de hallar variables independientes, como en los casos 4, 6, 7 y 10 de la Sección IV. Aunque esta situación también tiene su interpretación, no es menos cierto que en ocasiones hay que considerar, junto al algoritmo CHAID, otros métodos concurrentes para hallar la metodología óptima para una investigación determinada, eventualmente llevando a cabo una comparación de resultados entre diferentes metodologías, como hicieron Florensa Gulu et al. (2020) en su estudio sobre la satisfacción de visitantes de un parque natural. Asimismo, hubiera sido interesante disponer de los datos de más destinos comparables (posiblemente en la región Caribe) para poder establecer las respectivas características de preferencias de sus visitantes desde un marco más amplio. Quedan estos temas como posibles áreas de investigación futuras.

Finalmente queremos hacer un llamado para que los pequeños y medianos empresarios turísticos se interesen por soluciones *big data* como las aquí expuestas. Además, vemos una oportunidad para emprendedores tecnológicos de crear un software de manejo sencillo y fácil instalación que combine las funciones de barridos de web ('web-scraping') con el análisis ('parsing') y cuantificación de variables para dar una información clara y precisa para las pequeñas y medianas empresas turísticas (PYME). Este software no es difícil de hacer técnicamente y se podría vender a estas empresas a un precio económico, facilitando así el acceso de muchos empresarios al *big data*, lo cual hoy en día sigue siendo una barrera muy complicada de superar para muchos de ellos. Así, se colaborará a la competitividad de las PYME turísticas para que no queden rezagadas en el contexto actual de competencia global.

## REFERENCIAS BIBLIOGRÁFICAS

- Bassols, N. (2019) Evolving Iconization and Destination Building: The Example of Cartagena, Colombia. *Tourism Planning & Development* 16(3), 334-352
- Bassols, N. y Leicht, Th. (2020) Place marketing, evolutionary theories and the management of multifaceted destinations: Cartagena, Colombia, as a study case. *Tourism and Hospitality Research*, Publicado online en Octubre de 2020. DOI: 10.1177/1467358420966044
- Barrientos Martínez, R. E.; Cruz Ramírez, N.; Acosta Mesa, H. G.; Rabatte Suárez, I.; Gogeoascoechea Trejo, M. C.; Pavón León, P. y Blázquez Morales, S. L. (2009) Árboles de decisión como herramienta en el diagnóstico médico. *Revista Médica de la Universidad Veracruzana*, 9(2), 19-24
- Blal, I. y Sturman, M. (2014) The Differential Effects of the Quality and Quantity of Online Reviews on Hotel Room Sales. *Cornell Hospitality Quarterly*, 55(4), 365-375
- Bowen, J. (2015), Trends affecting social media: implications for practitioners and researchers. *Worldwide Hospitality and Tourism Themes*, 7(3), 221-228
- Bozkir, A. S. and Sezer, E, A, (2011) Predicting food demand in food courts by decision tree Approaches. *Procedia Computer Science* 3, 759-763
- Byrd, E, T. y Gustke, L. (2007) Using Decision Trees to Identify Tourism Stakeholders: The Case of Two Eastern North Carolina Counties. *Tourism and Hospitality Research*, 7(3-4), 176-193
- Çakmak E. e Isaac, R. K. (2012) What destination marketers can learn from their visitors' blogs: An image analysis of Bethlehem, Palestine. *Journal of Destination Marketing & Management*, 1(1-2), 124-133
- Chiu, W. y Leng, H. (2017) Let's go cycling: an analysis of tourists' experience on online user-generated content. *International Journal of Tourism Cities*, 3(1), 30-42
- De Ascaniis, S.; Borrè, A.; Marchiori, E. y Cantoni, L. (2015) Listen to Your Customers! How Hotels Manage Online Travel Reviews. The Case of Hotels in Lugano. *Proceedings of the International Conference on Information and Communication Technologies in Tourism 2015*, Lugano, Switzerland
- Di Pietro, R.; Petrocchi, M. y Spognardi, A. (2014) A Lot of Slots -- Outliers Confinement in Review-Based Systems, 15th *International Conference on Web Information System Engineering WISE 2014*, Salónica, Grecia

- Escobar, M. (1998). Las Aplicaciones del Análisis de Segmentación: el procedimiento CHAID. *EMPIRIA. Revista de metodología de Ciencias Sociales*, 1, 13-49.
- Ferreira, S.; Rial, A. y Valera, J. (2009) Post hoc tourist segmentation with conjoint and cluster analysis. *Pasos Revista de Turismo y Patrimonio Cultural*, 7(3), 491-501
- Florensa Gulu, R. M.; Colom Gorgues, A. y Maza Rubio, M. T. (2020) Análisis ACP, CHAID y PLS-SEM de la satisfacción de visitantes de espacios naturales protegidos. Aplicación al Parque Nacional de Aigüestortes en Lleida. *Cuadernos de Turismo*, 45, 167-195
- Garay Tamajón, L. y Cànoves Valiente, G. (2017) Barcelona seen through the eyes of TripAdvisor: Actors, typologies and components of destination image in social media platforms. *Current Issues in Tourism*, 20(1): 33-37
- Grinnell, D. y Remus, W. (1983) Using Decision Trees To Solve Restaurant Staffing Problems *Journal of Hospitality Education*, 7(2), 51-62
- Gržinić, J. y Floričić, T. (2013) Implementation of innovations in hotel offer promotion – case study of Istria as tourist destination. *Proceedings of the Int'l Forum on Knowledge Asset Dynamics*, 704-720
- Guardiola, A. del R. (2019) Potencial turístico de Santa Marta (Colombia): Una revisión de su desarrollo, crecimiento y barreras. *Espacios*, 40(6)
- Hlee, S.; Lee, H. y Koo, C. (2018) Hospitality and Tourism Online Review Research: A Systematic Analysis and Heuristic-Systematic Model. *Sustainability*, 10, 1141
- Hu, X. y Yang, Y. (2020) Determinants of consumers' choices in hotel online searches: A comparison of consideration and booking stages. *International Journal of Hospitality Management*, Vol. 86, DOI: 10.1016/j.ijhm.2019.102370
- Kladou, S. y Mavragani, E. (2015) Assessing destination image: An online marketing approach and the case of TripAdvisor. *Journal of Destination Marketing & Management*, 4(3): 187-193
- Korfiatis, N. y Poulos, M. (2013) Using online consumer reviews as a source for demographic recommendations: A case study using online travel reviews. *Expert Systems with Applications*, 40, 5507-5515
- Kuzey, C.; Karaman, A. S. y Akman, E. (2019) Elucidating the impact of visa regimes: A decision tree analysis. *Tourism Management Perspectives*, 29, 148-156
- Lee, H.; Guillet, B. D. y Law, R. (2013) An examination of the relationship between online travel and hotels; a case study of Choice Hotels International and [Expedia.com](https://www.expedia.com). *Cornell Hospitality Quarterly*, 54(1), 95-107
- Liu, S.; Law, R.; Rong, J.; Li, G. y Hall, J. (2013) Analyzing changes in hotel customers' expectations by trip mode. *International Journal of Hospitality Management*, 34(1), 359, 371
- Magidson, J. (1993) *SPSS for Windows CHAID Release 6.0*. Chicago: SPSS Inc
- Mehmetoglu, M. (2009) Predictors of sustainable consumption in a tourism context: a CHAID approach. En: Chen, J. (Ed.) *Advances in Hospitality and Leisure (Advances in Hospitality and Leisure, Vol. 5)*, Bingley: Emerald Group Publishing, p. 3-23
- Mellinas, J. P.; Soledad-Maria Martinez, M. D. y Bernal García, J. J. (2015) [Booking.com](https://www.booking.com): The unexpected scoring system. *Tourism Management*, 49, 72-74
- Mohamed, M.; Hewedi, M.; Lehto, X. y Maayouf, M. (2019) Marketing local food and cuisine culture online: a case study of DMO's websites in Egypt. *International Journal of Tourism Cities*. <https://doi.org/10.1108/IJTC-05-2019-0067>

- Moro, S. y Rita, P. (2019) Identification of common city characteristics influencing room occupancy. *International Journal of Tourism Cities*, 5(3), 482-490
- Moya Sánchez, D. H. y Majó Fernández, J. (2017) Análisis de comentarios en redes sociales para mejorar la reputación online hotelera. *Turismo y Sociedad*, 20, 169-190
- Pesonen J, (2014) Testing segment stability: Insights from a rural Tourism study. *Journal of Travel and Tourism marketing*, 31(6) 697-711
- Román-González, M. V., & Lévy-Mangin, J.-P. (2003) Clasificación y Segmentación Jerárquica. In J.-P. Lévy-Mangin & J. Varela-Mallou (Eds.), *Análisis Multivariable para las Ciencias Sociales* (pp. 567-630). Pearson Prentice Hall
- Serrano López, A. L.; Freire Chaglla, S. A.; Espinoza-Figueroa, F. E.; Andrade Tenesaca, D. S. y Villafuerte Pucha, M. E. (2019) Modeling of Tourist Profiles with Decision Trees in a World Heritage City: The Case of Cuenca (Ecuador). *Tourism Planning & Development*, 16(5), 473-493
- SITUR Magdalena (2018) Caracterización del turismo receptor del departamento de Magdalena. Comparativos anuales 2017 vs 2016. Informe oficial, Santa Marta, Colombia
- Solomon, M. R. (2015) *Consumer Behavior: Buying, Having, And Being*, 11th Edition. Toronto: Prentice Hall and Pearson Education
- Sparks, B. A.; Perkins, H. E. y Buckley, R. (2013) Online travel reviews as persuasive communication: The effects of content type, source, and certification logos on consumer behavior. *Tourism Management*, 39, 1-9
- Stepchenkova, S.; Kirilenko, A. y Shichkova, E. (2019) Influential factors for intention to visit an adversarial nation: increasing robustness and validity of findings. *International Journal of Tourism Cities*, 5(3), 491-510
- Tseng, C.; Wu, B.; Morrison, AM.; Zhang, J. y Chen, Y. (2015) Travel blogs on China as a destination image formation agent: A qualitative analysis using Leximancer. *Tourism Management*, 46: 347-358
- Vartiak, L. (2015) Benefits of online reputation management for organizations operating in various industries. *TRANSCOM 2015*, University of Žilina, Žilina, Eslovaquia
- Webb, T. (2016) From travel agents to OTAs: How the evolution of consumer booking behavior has affected revenue management. *Journal of Revenue Pricing Management*, 15, 276-282
- Wedel, M. y Kamakura, W. (2000) *Market segmentation – conceptual and methodological foundations* (2nd Ed.), Boston, MA: Kluwer
- Wilkins, H. (2010) Using importance-performance analysis to appreciate satisfaction in hotels. *Journal of Hospitality Marketing and Management*, 19(8), 866-888
- Wong, C. U. I. y Qi, S. (2017) Tracking the evolution of a destination's image by text-mining online reviews - the case of Macau. *Tourism Management Perspectives*, 23:19-29
- Xie, K. L.; Chihchien, C. y Wu, S. (2016) Online Consumer Review Factors Affecting Offline Hotel Popularity: Evidence from Tripadvisor. *Journal of Travel & Tourism Marketing*, 33(2), 211-223
- Zhang, H.; Xu, F.; Lu, L. y Lei, Y. (2015) Cultural capital and destination image of metropolitans: A comparative study of New York and Tokyo official tourism websites in Chinese. *Journal of China Tourism Research*, 11(2): 121-149

## ANEXO I: COMPROBACIÓN DEL DESEMPEÑO DE LOS DATOS

### Validación del desempeño del árbol construido para el Caso 1

Con el fin de evaluar el desempeño del árbol construido correspondiente al Caso 1, se tomó un total de 35 datos nuevos extraídos de los restantes de la base de datos total de la ciudad de Santa Marta. Para dicha evaluación, se realizó una comparación de dos proporciones que permitió analizar la influencia o relación que tienen las diferentes variables. Teniendo en cuenta que el tamaño de muestra era muy pequeño, el margen de error aumentaba lo que no permitía ver bien las diferencias, por ello se utilizó la técnica de simulación de datos la cual consistió en aumentar el tamaño de la muestra manteniendo las mismas proporciones. Para el análisis de los datos se utilizó el software de análisis de datos estadístico y gráfico Statgraphics.

Según lo expuesto con anterioridad y de acuerdo a las diferentes comparaciones de dos proporciones por valoraciones realizadas, se obtuvo que sí tiene influencia el hecho de que los hoteles cuenten con servicio de bar, lo que aumenta la probabilidad que los consumidores los califiquen con 5 (“sumamente satisfecho”) al momento de hacer una valoración del hotel. Para dar veracidad a lo planteado, a continuación, se detallan los resultados obtenidos:

Valoración/Bar	Datos iniciales		Proporción		Simulación de datos	
	No	Si	No	Si	No	Si
1	1	1	0,091	0,042	9	4
2	3	2	0,273	0,083	27	8
3	2	3	0,182	0,125	18	13
4	3	8	0,273	0,333	27	33
5	2	10	0,182	0,417	18	42
<b>Totales</b>	<b>11</b>	<b>24</b>	<b>1</b>	<b>1</b>	<b>100</b>	<b>100</b>

Tabla 3. Validación del desempeño del árbol construido para el caso 1 para la variable *Servicio de Bar*.  
Fuente: elaboración propia.

#### Valoración Servicio de Bar:

- Proporciones muestrales = 0,42 y 0,18
- Tamaños de muestra = 100 y 100
- Intervalo aproximado del intervalos de confianza del 95,0% para la diferencia entre proporciones: [0,117412;0,362588]
- Hipótesis Nula: diferencia entre proporciones = 0,0
- Alternativa: no igual
- Estadístico z calculado = 3,70328
- Valor-P = 0,000212889
- Rechazar la hipótesis nula para alfa = 0,05

Lo que permite concluir que, en la primera muestra de 100 observaciones, la proporción muestral fue igual a 0,18 y en la segunda muestra de 100 observaciones la proporción muestral fue igual a 0,42. Puesto que el valor-P para la prueba fue menor a 0,05 se infiere

que tienen relación las valoraciones catalogadas en 5 para aquellos hoteles que cuenten con servicio de bar.

Realizando la comparación de dos proporciones para el servicio de traslado al aeropuerto, se obtuvo que sí influye el hecho de que los hoteles cuenten este servicio, lo que aumenta la probabilidad que los consumidores los califiquen con 5 (“sumamente satisfecho”) al momento de hacer una valoración del hotel. Para dar veracidad a lo planteado, a continuación se detallan los resultados obtenidos:

Valoración/Traslado al aeropuerto	Datos iniciales		Proporción		Simulación de datos	
	No	Si	No	Si	No	Si
1	1	1	0,042	0,091	4	9
2	2	3	0,083	0,273	8	27
3	3	2	0,125	0,182	13	18
4	8	3	0,333	0,273	33	27
5	10	2	0,417	0,182	42	18
<b>Totales</b>	<b>24</b>	<b>11</b>	<b>1</b>	<b>1</b>	<b>100</b>	<b>100</b>

Tabla 4. Validación del desempeño del árbol construido para el caso 1 para la variable *Traslado al aeropuerto*.  
Fuente: elaboración propia.

Valoración del Servicio de traslado al aeropuerto:

- Proporciones muestrales = 0,42 y 0,18
- Tamaños de muestra = 100 y 100
- Intervalo aproximado del intervalos de confianza del 95,0% para la diferencia entre proporciones: [0,117412;0,362588]
- Hipótesis Nula: diferencia entre proporciones = 0,0
- Alternativa: no igual
- Estadístico z calculado = 3,70328
- Valor-P = 0,000212889
- Rechazar la hipótesis nula para alfa = 0,05

Lo que permite concluir que, también en este caso, en la primera muestra de 100 observaciones la proporción muestral fue igual a 0,18 y en la segunda muestra de 100 observaciones la proporción muestral fue igual a 0,42. Puesto que el valor-P para la prueba fue menor a 0,05 se infiere que tienen relación las valoraciones catalogadas en 5 para aquellos hoteles que cuenten con servicio de traslado al aeropuerto.

Para el caso del servicio de lavandería, de los 35 hoteles estudiados de la ciudad de Santa Marta, todos ellos contaban con este servicio, con lo cual no es necesario mostrar diferencias de índole alguna.

## Validación del desempeño del árbol construido para el Caso 2

Con el fin de evaluar el desempeño del árbol construido correspondiente a este Caso 2, se tomó un total de 35 datos nuevos de los restantes de la base de datos total de la ciudad de Cartagena, usando la simulación de datos del mismo modo que en el anterior apartado.

Según lo expuesto con anterioridad y de acuerdo a las diferentes comparaciones de dos proporciones por valoraciones realizadas, se obtuvo que sí tiene influencia el que los hoteles cuenten con servicio de parking gratis, lo que aumenta la probabilidad que los consumidores califiquen 3 (“satisfecho”) al momento de hacer una valoración del hotel. Para dar veracidad a lo planteado, a continuación se detallan los resultados obtenidos:

Valoración/Parking gratis	Datos iniciales		Proporción		Simulación de datos	
	No	Si	No	Si	No	Si
1	0	0	0	0	0	0
2	0	1	0	0,034	0	3
3	0	3	0	0,103	0	10
4	2	9	0,333	0,310	33	31
5	4	16	0,667	0,552	67	55
<b>Totales</b>	<b>6</b>	<b>29</b>	<b>1</b>	<b>1</b>	<b>100</b>	<b>100</b>

Tabla 5. Validación del desempeño del árbol construido para el caso 2 para la variable *Parking gratis*.  
Fuente: elaboración propia.

Valoración Servicio de parking gratis:

- Proporciones muestrales = 0,0 y 0,1
- Tamaños de muestra = 100 y 100
- Intervalo aproximado del intervalos de confianza del 95,0% para la diferencia entre proporciones: [-0,158799;-0,041201]
- Hipótesis Nula: diferencia entre proporciones = 0,0
- Alternativa: no igual
- Estadístico z calculado = -3,24443
- Valor-P = 0,00117699
- Rechazar la hipótesis nula para alfa = 0,05

Con ello, se ve cómo, en la primera muestra de 100 observaciones, la proporción muestral fue igual a 0,0. En la segunda muestra de 100 observaciones la proporción muestral fue igual a 0,1. Puesto que el valor-P para la prueba fue menor a 0,05 se infiere que tienen relación las valoraciones catalogadas en 3 (“satisfecho”) para aquellos hoteles que cuenten con servicio de parking gratis, y por tanto es un grado muy bajo de satisfacción el que produce este servicio.

## CONTRIBUCIONES DE LOS AUTORES

Narcís Bassols se centró en la revisión de literatura, así como en la elaboración de la introducción y la redacción de las conclusiones. Andrea Ovalle y Jenniffer Rodríguez capturaron datos

y realizaron los análisis con árboles de decisión. Los tres autores de modo conjunto realizaron la revisión final del documento.

### **AGRADECIMIENTOS**

Para la realización de la presente investigación no se contó con ninguna ayuda financiera ni pública ni privada.

### **RECONOCIMIENTOS**

Los autores desean agradecer a Christian Acevedo, PhD (Uniaugustiniana, Bogotá) la lectura y comentarios al borrador del artículo, así como a Gustavo Rodríguez Albor, PhD (Universidad Autónoma del Caribe, Barranquilla) su apoyo y observaciones sobre árboles de decisión y el sistema SPSS.