Exploring Segmentation in eTourism: Clustering User Characteristics in Hotel Booking Situations Using k-Means

Stefan Eibl[1(✉)], Robert A. Fina[1], and Andreas Auinger[2]

[1] University of Applies Sciences Wiener Neustadt, Zeiselgraben 4, 3250 Wieselburg, Austria
stefan.eibl@fhwn.ac.at

[2] University of Applied Sciences Upper Austria, Wehgrabengasse 1-3, 4400 Steyr, Austria

Abstract. In the dynamic field of eTourism, personalization and user segmentation are paramount for enhancing user experience and driving digital platform success. This paper addresses the gap in eTourism research related to understanding consumer behavior through an external lens, due to the limited access to proprietary data from Online Travel Agencies (OTAs). We employ Adaptive Choice-Based Conjoint (ACBC) analysis and k-means clustering on data from a survey (n = 801) based on 346 hotel listings on Booking.com, focusing on Vienna. Attributes such as star category, price, review valence, volume of reviews, scarcity indicators, sustainability cues, and city center proximity were examined to identify consumer preferences. Five distinct consumer clusters were revealed: Cost-Conscious Eco-Bookers, Green-Urban Deal Hunters, Social-Proof Assurance Seekers, Budget-Only Focused Minimalists, and Luxury-Quality Connoisseurs. These clusters vary in their prioritization of hotel attributes and demographics, demonstrating the diverse decision-making criteria within the eTourism market. This paper proposes a foundation for classifying user groups on booking platforms, enabling OTAs and hoteliers to tailor offerings to nuanced consumer segments, thus improving user experiences and potentially increasing conversion rates. The findings offer actionable insights into OTA personalization strategies and contribute to the scientific understanding of consumer behavior in the digital tourism landscape.

Keywords: eTourism Segmentation · k-Means Clustering · Personalization Tactics

1 Introduction

In the rapidly evolving landscape of Online Travel Agencies (OTAs), the significance of user segmentation and personalization strategies cannot be overstated. These approaches have become pivotal in enhancing user experiences and driving business success in the digital tourism domain [1]. However, a critical challenge facing the academic community and external researchers in this field is the limited access to proprietary data held by OTAs. This data, rich with insights on consumer behavior and preferences, often remains confidential and exclusively utilized for internal strategic purposes, thus information

F. F.-H. Nah and K. L. Siau (Eds.): HCII 2024, LNCS 14720, pp. 157–175, 2024.
https://doi.org/10.1007/978-3-031-61315-9_11

quality may be influenced [2]. Consequently, there is a compelling need to examine user segmentation and personalization from an external perspective, contributing valuable findings to the scientific community. This research aims to fill this gap by exploring user segmentation based on individual preferences, as indicated by utility scores, and correlating them with user demographic data from a sample of Austrian and German respondents (n = 800). This approach is instrumental in creating an external viewpoint on how consumers interact with hotel booking platforms, which is crucial for expanding the existing body of knowledge in eTourism [3]. While OTAs have access to this data for operational and marketing strategies, they often do not share detailed insights publicly, limiting the broader understanding of consumer behavior in this context [4]. By conducting this research, we aim to provide empirical evidence and a theoretical framework for user segmentation that can be utilized by academics and practitioners alike. This is particularly important as it offers an independent analysis of consumer preferences and behavior, which is often shaped by various factors, including psychological, social, and economic elements [5, 6]. Additionally, this research contributes to a deeper understanding of how demographic characteristics interplay with personal preferences in the context of hotel bookings, an area that has seen limited exploration due to the proprietary nature of OTA data [7]. Therefore, our study not only addresses a critical gap in eTourism research but also presents an opportunity for the scientific community to gain insights into consumer segmentation and personalization strategies, which are essential for the continued growth and evolution of the online travel industry [8].

2 Related Work

2.1 User Segmentation in eTourism

User segmentation in eTourism has become increasingly significant as the digitalization of travel and hospitality services continues to evolve. This segmentation process involves categorizing potential customers into distinct groups based on shared characteristics. This categorization is crucial for effective marketing and enhancing user experience on digital platforms [9]. Segmentation theories like the Market Segmentation Theory suggest that distinct groups within a market can be targeted with tailored marketing strategies [10]. The VALS (Values, Attitudes, and Lifestyles) framework, categorizes consumers based on psychological traits and key demographics, which is particularly useful in understanding traveler segments [11]. Segmentation in eTourism is conducted using demographic, psychographic, and behavioral data. Demographic segmentation involves categorizing consumers based on criteria such as age, gender, income, and education [12]. Psychographic segmentation delves deeper into consumers' lifestyles, values, and opinions [13], while behavioral segmentation focuses on purchase history, loyalty, and service engagement [14]. OTAs and hotels apply these segmentation methods to customize recommendations, promotional offers, and design loyalty programs [3, 15]. For example, Gretzel & Fesenmaier [16] showed how understanding travel motivations of different segments enhances marketing effectiveness. The role of data in forming customer segments is particularly significant. The emergence of big data and analytics has provided eTourism platforms with immense information, aiding in more accurate segmentation [17]. This data-driven approach enables a nuanced understanding of consumer behavior and preferences, leading to more effective personalization strategies for companies which have the resources to gather and analyze this data [18].

Building further on the concept of user segmentation in eTourism, it's evident that the effective utilization of segmentation strategies can greatly enhance the personalization of services offered by online travel platforms. The profound impact of tailored marketing and service delivery, based on a deep understanding of different customer segments, is a recurring theme in recent eTourism research [7]. Advanced data analytics techniques have opened new avenues for understanding consumer behavior in the eTourism sector. Data mining and machine learning algorithms, for instance, allow for the extraction of meaningful patterns from large datasets, enabling platforms to identify subtle preferences and behaviors of different segments [17]. These techniques have been instrumental in refining segmentation strategies, allowing for a level of personalization that was previously unattainable [19]. Behavioral segmentation, particularly, has gained traction in the digital era. By analyzing online behavior patterns, such as booking history and interaction with OTA platforms, businesses can gain insights into the preferences and decision-making processes of consumers [20]. This approach aligns with the increasing emphasis on customer experience in the digital marketplace, where personalization is key to customer satisfaction and loyalty [21]. The importance of demographic factors, though traditional, remains significant. Age, income, and education level continue to influence travel preferences and booking behaviors. For example, younger travelers may show a propensity for budget-friendly options and are more influenced by social media marketing, whereas older travelers might prioritize comfort and direct booking experiences [22].

2.2 Personalization Strategies in Online Hotel Booking Situations

Personalization strategies in the online hotel booking sector represent a sophisticated interplay between technology, data analytics, and consumer psychology. In an industry characterized by intense competition and evolving consumer expectations, personalization has emerged as a strategic imperative for enhancing customer satisfaction and loyalty [21]. This approach focusses on tailoring the user experience to individual needs and preferences, often leveraging rich data sets to craft targeted messages and offers [23]. The foundation of personalization lies in the understanding that each traveler's needs are unique [24]. Recognizing and responding to these needs in real-time is the essence of personalization in eTourism [1]. OTAs have been at the forefront of this trend, employing sophisticated algorithms to suggest hotels, special deals, and additional services based on past behavior, search patterns, and preferences [25, 26]. This level of customization is made possible by the immense data users leave as they interact with online platforms, which, when analyzed, can reveal deep insights into consumer behavior [5]. One of the key methods employed by OTAs to achieve personalization is collaborative filtering. This technique uses data from many users to provide recommendations based on similar search and booking patterns [26]. Another method is content-based filtering, which suggests options based on the similarity of items, such as hotels or destinations, to those a user has expressed interest in before [27]. These filtering mechanisms are integral to creating a personalized experience, as they can dynamically adjust the content presented to each user based on their interests and behaviors. The impact of personalization on consumer behavior is great. Studies have shown that consumers are more likely to engage with and purchase from platforms that offer a personalized experience [28]. The personal touch fosters a sense of value and recognition among customers, which, in turn, enhances their loyalty to the platform [29]. In fact, personalization can lead to a virtuous cycle: the more a customer interacts with a personalized service, the more data is generated, which further refines the personalization algorithms, resulting in even more engagement [30]. However, the implementation of personalization strategies is not without challenges. The primary concern is the balance between personalization and privacy. As platforms collect and utilize personal data to tailor experiences, they must also navigate the complex landscape of data privacy regulations and consumer privacy concerns [31]. Transparency in how data is collected, used, and protected is vital for maintaining consumer trust and ensuring the ethical use of personalization technologies [32]. Another challenge is the avoidance of the "filter bubble" effect, where the personalization algorithm over-specializes the content, restricting the diversity of offerings presented to the user [33]. To combat this, OTAs are exploring hybrid recommendation systems that combine collaborative and content-based filtering with techniques that introduce probability and diversity into the recommendations, or by even trying to broaden the segmentation approach [34].

3 Methodology

3.1 Content Mining and Multiple Linear Regression

To analyze which attributes to focus on within the process of analysis, we conducted a web content mining approach using the tool "Octoparse". We also examined the results of the systematic literature review of Eibl & Auinger [8], this aimed to identify factors that influence booking intentions [35, 36]. We selected hotel attributes, focusing on those, visible on the search results page of booking platforms like booking.com. Thus, attributes such as descriptions or room sizes, which are not immediately visible there, were excluded. While images likely influence booking decisions, their analysis was beyond the scope of this web content-focused study. We carried out an analysis of 346 hotel listings on booking.com, with a focus on Vienna within high season, to gather data on various attributes such as star category, price, review valence, the volume of reviews, scarcity indicators, sustainability cues, and distance to the city center. To ensure a comparability across the diverse range of our independent variables, we applied z-standardization, aligning our data on a standardized scale for use in our multiple linear regression model [37, 38]. The results of these two approaches were then incorporated into a conjoint analysis.

3.2 Adaptive Choice Based Conjoint Analysis

In our methodology, the Adaptive Choice-Based Conjoint (ACBC) analysis served as a cornerstone to discern how multiple attributes influence hotel booking decisions on eTourism platforms. A survey administered through Sawtooth Software Lighthouse Studio to a random sampling of individuals from Austria and Germany (n = 800) captured not only demographic information but also the participants' preferences within ACBC scenarios, see Table 1. This process enabled the calculation of utility scores for each hotel attribute, providing a nuanced understanding of the role these attributes play in online hotel booking behaviors [39].

The selection of attributes and levels for the conjoint analysis was a decision informed by the web content mining approach, the application of the multiple linear regression model, and established literature on conjoint analysis by Baier & Brusch [35]. Within an ACBC, it is recommended to have a range of 5 to 12 attributes and each attribute can have between 2 to 12 levels to ensure comprehensive coverage without overwhelming respondents. This range balances detail with manageability, allowing for thorough investigation while maintaining participant engagement and the quality of data collected. [35].

The ACBC analysis, tailored to reflect the intricacies of consumer decision-making, was executed in four structured steps, beginning with a (1) "Build Your Own" phase where respondents were asked to design their ideal hotel by selecting their preferred levels of the presented attributes (like sustainability level or distance to town center). This initial step allows for the identification of each individual's most desired features. Following the BYO, the analysis progresses into a screening phase, where respondents are presented with a series of hotel configurations that are close to their ideal but include some variations. Respondents must decide which of these configurations they would consider acceptable alternatives to their BYO selection. Subsequently, the ACBC approach narrows down the field through Choice Tournaments, where the acceptable configurations compete against each other in head-to-head matchups. In these matchups, respondents are asked to make choices between different sets of alternatives for leisure travel, further refining their preferences.

So, within the survey, participants were presented with a series of hotel options where the identified attributes were displayed side by side, as well as intermixed within each hotel option. This method simulates real-life decision-making by requiring individuals to evaluate and choose between hotels based on a combination of characteristics, such as location and price, without focusing on a single attribute. This approach helps to understand how various factors are weighted against each other in the decision process, reflecting a more realistic scenario where multiple attributes influence the choice of a hotel. [35, 40]. These steps were critical in calculating participants' genuine preferences and allowed us to explore the relationship between consumer demographics and their attribute preferences. Hierarchical Bayesian estimation techniques were employed to calculate utility scores for each attribute level, providing robust, reliable insights into individual preferences and decision patterns [35, 41]. By integrating these steps, we ensured a comprehensive capture of participants' preferences, which are vital in informing the design and personalization of user experiences on hotel booking platforms.

Table 1. Demographics

	N	%		N	%
Age			**Education**		
16–20	19	2.375%	Mandatory school	243	30.375%
21–30	90	11.25%	High school	287	35.875%
31–40	104	13%	Bachelor's degree	91	11.375%
41–50	145	18.125%	Master's degree	155	19.375%
51–60	176	22%	Doctor degree	24	3%
61–70	158	19.75%			
71–80	96	12%	**Net household income per year (EUR)**		
>80	12	1.5%	<19,999	150	18.75%
			20,000–39,999	246	30.75%
Gender			40,000–59,999	166	20.75%
Female	416	52%	60,000–79,999	108	13.5%
Male	382	47.75%	80,000–99,999	65	8.125%
Diverse	2	0.25%	>100,000	65	8.125%
Nationality					
Austria	401	50.125%			
Germany	399	49.875%			
Total	800	100%		800	100%

3.3 K-Means Clustering Approach

Utilizing k-means clustering to analyze similarities and differences between user groups represents a robust method for identifying patterns in consumer behavior. When combined with ACBC results, this approach offers a novel perspective on customer preferences, particularly in the domain of online hotel bookings. This statistical technique is instrumental in segmenting a dataset into a specified number of distinct groups based on inherent similarities within the data, which, in our study, was implemented using the robust capabilities of the XLSTAT software. K-means clustering is a partitioning method that assigns observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster [42]. The process is iterative and aims to minimize the within-cluster sum of squares, which is essentially a variance measure within each cluster. The methodology rests on defining k centroids, one for each cluster, and then assigning each data point to the nearest centroid based on the Euclidean distance [43]. The centroids are recalculated after each iteration, which results in the reassignment of data points until the within-cluster variation cannot be further reduced, and the clusters become stable [44].

In our research, we leveraged the ACBC individual utility scores, which reflect the relative importance of various hotel attributes to the consumer's decision-making process. These utility scores represent multidimensional data points that the k-means algorithm could effectively analyze to identify coherent clusters of consumers with similar hotel attribute preferences. We chose to focus on five clusters as this number provided the best fit to the data, which was ascertained through the evaluation of several cluster solutions against criteria such as the elbow method and the silhouette score - a measure of how similar an object is to its own cluster compared to other clusters [45]. The selection of five clusters was also validated by the interpretability and managerial implications of the segmentation. It allowed for a detailed differentiation of consumer preference patterns without overcomplicating the model with too many segments, which might have led to an impractical application in a business context [46]. The decision was in line with the parsimony principle, which suggests that models should be as simple as possible, but no simpler - a balance between complexity and practicality [47].

The k-means methodology also was applied to the ACBC utility scores using XLSTAT. The algorithm's application involved several steps: standardizing the utility scores on a scale from 0–100, initializing the centroids, assigning observations to the nearest centroids, recalculating the centroids, and repeating the assignment and recalculating steps until convergence. The standardization of data before clustering is crucial, as it ensures that each attribute contributes equally to the similarity measure and prevents attributes with larger ranges from dominating the distance calculations [48].

Subsequently the sum of squared distances from each point to the centroid of its assigned cluster was minimized, ensuring that the clusters were as compact and separate as possible [49]. Additionally, the silhouette analysis was performed to assess the goodness of fit for each cluster. This involved calculating the average silhouette width for each cluster and for the dataset as a whole, which provided a graphical representation of how well each object lies within its cluster [45].

4 Results

The complex landscape of consumer preferences within the online hotel booking domain presents a multifaceted challenge for market segmentation. Our study's expedition into this domain through the application of k-means clustering has revealed distinct consumer segments, each characterized by unique utility preferences concerning hotel attributes [46].

The five clusters that emerged from our analysis represent distinct archetypes of consumers in the eTourism marketplace. These clusters vary significantly in their valuation of hotel attributes, suggesting differing priorities and decision-making criteria among the groups. The clusters range from price-sensitive consumers to those who prioritize sustainability and luxury, reflecting the diverse nature of the online hotel booking audience. Cross-tabulation was used to further enrich the cluster profiles with demographic data, linking utility preferences to demographic characteristics such as age, gender, income, travel frequency, educational level and marital status [50].

In the subsequent subsections, we will delve into each cluster, outlining their defining characteristics. The detailed breakdown of clusters will provide a rich description of the diverse consumer base that OTAs serve, and how these differences require tailored approaches to marketing and service design. This segment-specific insight is crucial for OTAs and hoteliers aiming to enhance the personalization of their offerings and is consistent with the literature supporting the strategic importance of customized consumer engagement [51]. As we transition to the detailed analysis of each cluster, it is essential to bear in mind that the overarching goal of this segmentation is to identify actionable insights that can inform the personalization strategies of OTAs and hotel-entries, thereby optimizing the consumer experience and driving business performance in the competitive landscape of eTourism [23].

Within the results section of our study, we also present a detailed examination of the k-means clustering analysis applied to various hotel attributes and their influence on user segmentation. The statistical approach undertaken involves an Analysis of Variance (ANOVA), which tests the hypothesis that the means of several groups are equal, see Table 2. This method is instrumental in discerning the significance of each attribute in the formation of distinct user clusters.

The ANOVA results indicate that the majority of hotel attributes have a statistically significant influence on the clustering of user preferences. High F-values and p-values less than 0.05 confirm that attributes such as user ratings, ranging from "9.7" to "6.9", play a pivotal role in segmenting users into distinct groups. Furthermore, proximity to the city center, with varying distances, emerged as a critical factor, with closer distances correlating strongly to user cluster formation.

We also see, that the scarcity cues ranging from 1 to 5 rooms left have a significant effect on building different user clusters, whereas the lack of significance concerning the attribute "only 7 rooms left…" implies that such scarcity cue might not be as influential in shaping user preferences for hotel bookings.

The various price levels, designated as from 84.7 € to 600 € displayed extremely significant p-values, indicating that price is a very important attribute in user segmentation.

These results not only reinforce the validity of our k-means clustering approach but also underpin the significant differentiation among the five user clusters identified in our analysis. The significant variances across key hotel attributes underscore the distinct preferences and decision-making criteria inherent to each cluster. This foundational understanding of the attributes that influence user segmentation allows us to delve deeper into the characteristics of each cluster.

4.1 Cost-Conscious Eco-Bookers (CCEB)

The centroid data for the "Cost-Conscious Eco-Bookers" cluster provides valuable insights into the booking preferences of this group. This cluster is characterized by individuals who prioritize cost-effectiveness but also have an interest in sustainability, as long as it does not involve additional costs. A detailed analysis of the centroid values reveals the following key points about the "Cost-Conscious Eco-Bookers", see Table 3.

The values indicate a clear pattern of price sensitivity, with the highest values associated with the cheapest price options. These consumers are significantly influenced by cost, with the utility scores decreasing as the price increases, showing a clear preference for more budget-friendly options. While this group considers sustainability, it is not their primary concern. The utility scores for sustainability levels are moderate compared to the

Table 2. Analysis of Variances

Attributes	Level	F	Pr > F
Review-Valence	9.7	67.639	<0.0001
	9.0	66.472	<0.0001
	8.3	12.093	<0.0001
	7.6	64.746	<0.0001
	6,9	78.030	<0.0001
Amount of Reviews	5	27.371	<0.0001
	5795	2,767	0.026
	9655	45,534	<0.0001
Hotel star category	3-star	72,981	<0.0001
	4-star	55,223	<0.0001
	5-star	52,195	<0.0001
Scarcity cues	Only 1 room left	30,380	<0.0001
	Only 3 rooms left	16,083	<0.0001
	Only 5 rooms left	15,129	<0.0001
	Only 7 rooms left	1,798	0.127
Distance to town center	0.1 km	90,616	<0.0001
	1.1 km	41,260	<0.0001
	2.1 km	16,211	<0.0001
	6.4 km	84,559	<0.0001
Sustainability level	Level 1	9,917	<0.0001
	Level 2	5,649	0.000
	Level 3	3,988	0.003
	Level 3+	15,802	<0.0001
Price	84.7 €	113,607	<0.0001
	100 €	113,607	<0.0001
	150 €	217,026	<0.0001
	200 €	401,072	<0.0001
	250 €	490,994	<0.0001
	300 €	533,226	<0.0001
	400 €	436,014	<0.0001
	500 €	431,501	<0.0001
	600 €	394,072	<0.0001

scores for lower prices, suggesting that while eco-friendliness is valued, it is subordinate to price considerations. The scores for hotel ratings show a nuanced behavior.

There is an appreciation medium-rated hotels, with a notable peak for “7.6”. This might imply a trade-off between quality and cost, where acceptable quality at a lower price is preferred over higher quality at a higher price. The centroid values for scarcity messages indicating limited room availability are relatively high. This suggests that scarcity messages may effectively nudge this cluster towards making a booking decision, possibly due to fear of missing out on a good deal. The closer the hotel is to the city center, the more there is a rather negative stance toward these hotels. The preference for a three-star category over four or five stars suggests a tendency to seek satisfactory accommodations without the need for luxury, aligning with their cost-conscious profile.

This segment shows a high percentage of individuals who did not book any hotels last year, pointing towards a limited need for hotel services. They are primarily in the lower income bracket, suggesting budget constraints influence their booking decisions. Educationally, they span from compulsory to tertiary levels, with a lean towards lower education, which may correlate with their cost-conscious behavior. This group is also marked by a younger demographic, possibly indicating a temporary phase of life with limited financial resources for travel.

4.2 Green-Urban Deal Hunters (GUDH)

Green-Urban Deal Hunters represent a segment of travelers who are looking for more than just the lowest price. They value a good balance between the cost of accommodation and its sustainable credentials, provided the hotel’s location allows them to be at the heart of urban life. This cluster might consist of individuals who are environmentally conscious, yet their decisions are also driven by practical considerations of cost and convenience. This cluster has relatively high centroid values for hotels rated as “7.6” and “6.9” which indicates a preference for medium-rated hotels. However, the values for “9.0” and “8.3” are also relatively high, revealing a balanced consideration between quality and affordability. The centroid values for sustainability levels are present but not as pronounced as price indicators, suggesting that while sustainability is a consideration, it does not override the importance of price. The values for proximity to the city center suggest that urban location is important to this cluster, with a desire to be close to city amenities and attractions.

The centroid values across different price points show a downward trend as prices increase, confirming the price sensitivity of this cluster. They are looking for a “deal” that balances cost with the perceived value of sustainability and location. Higher values for larger numbers of reviews indicate a reliance on social proof and the wisdom of the crowd in decision-making, suggesting they seek validation from other travelers’ experiences. Lower values for the five-star category and the highest room rates suggest a lesser emphasis on luxury accommodations, indicating a pragmatic approach to booking where excessive spending is avoided.

Green-Urban Deal Hunters are typically married, indicating a potential for family or couple-based travel preferences. They have a balanced booking frequency, reflecting a considered approach to travel, possibly planning around family or work commitments.

Their education levels are quite distributed but show a tendency towards higher education, which may influence their value for sustainability and urban experiences. They book moderately and evenly distributed across income levels, suggesting a conscious balance between quality and cost.

4.3 Social-Proof Assurance Seekers (SPAS)

The "Social-Proof Assurance Seekers" cluster is characterized by travelers who rely heavily on the experiences of others to guide their booking decisions. The prominence of online reviews and scarcity cues in their decision-making suggests that they may seek reassurance from others' endorsements before committing to a booking. They are willing to invest in a higher-rated hotel, provided it comes with a strong backing from many reviews, reflecting a collective confirmation of the establishment's quality. This cluster's centroid data implies a decision-making process highly influenced by social validation and quality assurances, placing significant weight on the opinions and experiences of others. They may perceive a scarce availability as an indication of a hotel's popularity or quality, which can serve as a persuasive factor in their decision-making process. With a significant emphasis on ratings and the number of reviews, this group is likely to seek social validation through the experiences of others. They may exhibit trust in the wisdom of the crowd, using it as a benchmark for their choices.

While they do consider sustainability, it is not the overriding factor in their decision process. However, they appreciate eco-friendly practices as a value-add, especially if such attributes come with strong social proof. This cluster's willingness to pay more for a hotel that has a strong backing in terms of social validation, as reflected in the high ratings and numerous reviews, indicates their preference for assured quality over lower cost. The centroid values for higher ratings categories like "9.7" and "9.0" are notably substantial, suggesting that this cluster places a premium on staying at hotels with excellent reputations. In addition, there is a spread across centroid values for different distances to the city center, which might indicate a certain degree of flexibility in terms of location, if the hotel's quality and social proof are assured.

Social-Proof Assurance Seekers have a higher representation of married individuals, which might be indicative of travel decisions influenced by family considerations or shared experiences valuing social proof. They tend to book hotels with a frequency that suggests regular but not excessive travel. Their education levels skew towards higher education, and they display a broad age range, suggesting diverse life stages from working professionals to active retirees who value others' opinions in their booking choices.

4.4 Budget-Only Focused Minimalists (BOFM)

The "Budget-Only Focused Minimalists" are characterized by their single-minded pursuit of economical options. They exhibit a high degree of price elasticity, responding to cost savings rather than other features such as sustainability, scarcity, or luxury. The limited sensitivity to hotel ratings and the number of reviews indicate that they may rely on basic accommodation standards or are confident in their ability to select suitable accommodations without heavily depending on other travelers' opinions. It is evident that this group prioritizes cost above other attributes when making hotel booking decisions. This cluster demonstrates a strong preference for lower prices, as indicated by the significant utility scores associated with lower price points. The scores across various rating levels do not show a marked preference, suggesting that this group does not weigh ratings as heavily in their decision-making process. There is no significant reaction to scarcity cues such as limited room availability. This group seems to be less influenced by marketing tactics that create a sense of urgency through scarcity. Sustainability levels appear to have little to no impact on their booking decisions.

This segment is characterized by the highest percentages of individuals in the lowest income and education brackets, which directly influences their minimalistic approach to travel. Their booking patterns show a significant number of older individuals. This group's less frequent booking behavior suggest a targeted and essential approach to travel, prioritizing affordability over luxury or brand reputation.

4.5 Luxury-Quality Connoisseurs (LQM)

The "Luxury-Quality Connoisseurs" are sensitive travelers who seek out the best experiences. They are likely to book at well-established, high-starred hotels and may use sustainability as a decider between equally luxurious options. Their booking behavior is motivated by the pursuit of top-quality service, comfort, and an overall luxurious experience. Analyzing the centroid data for this cluster reveals a group that places a premium on high-quality, luxury experiences, and while they have an appreciation for sustainability, it is not their primary concern.

The higher utility scores for top-tier ratings indicate that this group is inclined towards hotels with exceptional reviews. They are likely to seek out establishments that promise an elite experience, denoted by high guest satisfaction levels. While price sensitivity is present, it is not as pronounced as in other clusters. This group is willing to pay more for perceived quality and luxury, as suggested by the balanced utility scores across various price points. Scarcity cues such as "only one room left" may influence their decision to some extent, hinting that while they are looking for luxury, they are also attracted to exclusivity, which scarcity signals can imply. Although sustainability is not disregarded, it is secondary to luxury and quality. They might prefer sustainable options, but not at the expense of comfort or prestige.

As indicated by their income bracket, the Luxury-Quality Connoisseurs show a higher tendency for frequent bookings, emphasizing the importance of travel in their lifestyle. They are often married, which might suggest a preference for shared high-end travel experiences or business travel that allows for more luxurious stays. Their education levels are spread across the spectrum, with a notable percentage holding advanced degrees, possibly reflecting their appreciation for quality and comfort in their travel choices. This group tends to be older, which may correlate with the financial means to prioritize luxury in their bookings.

Table 3. Centroid Data for Clusters

Attributes	Level	Cluster 1 CCEB	Cluster 2 GUDH	Cluster 3 SPAS	Cluster 4 BOFM	Cluster 5 LQM
Review-Valence	9.7	37.919	44.828	55.851	42.137	56.967
	9.0	37.820	46.592	55.616	43.521	58.434
	8.3	44.278	45.536	51.425	47.805	54.220
	7.6	64.280	59.505	48.276	56.938	45.121
	6.9	62.018	52.444	41.993	58.160	40.167
Amount of Reviews	5	64.870	51.179	54.123	57.286	49.080
	5795	48.021	47.687	43.381	46.771	47.674
	9655	44.799	59.763	59.703	53.857	62.030
Hotel star category	3-star	62.985	62.115	49.731	63.153	44.559
	4-star	42.785	44.375	62.270	41.479	53.914
	5-star	30.950	30.756	32.381	31.915	47.734
Scarcity cues	Only 1 room left	52.047	50.268	62.104	48.627	59.288
	Only 3 rooms left	51.609	56.623	45.449	54.888	50.695
	Only 5 rooms left	54.156	54.527	48.634	55.680	46.432
	Only 7 rooms left	48.861	45.064	47.940	48.126	47.414
Distance to town center	0.1 km	39.406	60.244	34.791	47.173	38.121
	1.1 km	51.584	67.725	59.839	51.041	51.181
	2.1 km	50.013	49.945	58.031	46.280	49.080
	6.4 km	61.346	34.075	55.219	57.525	63.155
Sustainability level	Level 1	53.624	55.680	55.733	63.047	57.259
	Level 2	57.913	55.508	63.529	58.785	58.800
	Level 3	48.712	49.172	50.047	43.860	45.972
	Level 3+	47.094	45.615	36.959	36.727	43.531
Price	84.7 €	66.888	36.526	49.460	37.520	46.393
	100 €	66.888	36.526	49.460	37.520	46.393
	150 €	59.014	45.531	53.574	24.895	58.601
	200 €	43.806	49.098	45.326	15.316	61.060

(*continued*)

Table 3. (*continued*)

Attributes	Level	Cluster 1 CCEB	Cluster 2 GUDH	Cluster 3 SPAS	Cluster 4 BOFM	Cluster 5 LQM
	250 €	32.008	45.141	35.036	13.328	57.087
	300 €	28.601	46.484	33.911	12.423	58.122
	400 €	22.583	42.552	28.025	14.331	55.183
	500 €	24.171	43.013	29.367	16.433	57.153
	600 €	23.620	41.529	31.747	18.945	58.467

5 Discussion

The emergence of five distinct consumer clusters - Cost-Conscious Eco-Bookers, Green-Urban Deal Hunters, Social-Proof Assurance Seekers, Budget-Only Focused Minimalists, and Luxury-Quality Connoisseurs - reflects a spectrum of prioritization across multiple hotel attributes, from price and location to sustainability and social proof. The significant implications of these clusters for OTAs and hoteliers lie in their potential application for precision-targeted marketing strategies and the enhancement of the personalization of services. This insight aligns with the works of Gretzel et al. [1] and Xiang et al. [23], which emphasize the need for a deep understanding of consumer behavior to drive personalization in the digital tourism sphere. Our findings mirror the shift in market segmentation theories, moving beyond demographic data towards a richer psychographic and behavioral understanding as outlined by Weinstein [12] and Smith [10]. The statistical validation of our clustering approach, evidenced by the ANOVA results, resonates with the importance of varied attributes in influencing consumer preferences. This relates closely to the findings of Buhalis & Law [18] and Li et al. [5], who highlighted the role of big data in enabling nuanced market segmentation in tourism. Moreover, the utility scores for attributes like review valence and scarcity cues support the perspectives of Morrison [7], indicating the ongoing significance of consumer-perceived value and urgency in booking decisions. Our research contributes a significant layer to the body of eTourism literature by proposing a novel model for user segmentation based on direct preferences for hotel attributes. This model has the potential to bridge the gap identified by Kotler & Keller [9] and Plummer [11], where the interplay of consumer psychology and market segmentation has been a enduring focus.

Cost-Conscious Eco-Bookers. The Cost-Conscious Eco-Bookers (CCEB) cluster exemplifies a segment balancing financial caution with environmental concerns. This group's price sensitivity echoes Kotler & Keller's [9] emphasis on cost-effective marketing strategies tailored for budget-aware segments. The moderate interest in sustainability aligns with Bahja et al.'s [6] findings that ecological concern influences consumer choices in hospitality. Yet, for CCEB, environmental friendliness is secondary to affordability, suggesting a need for competitively priced eco-friendly options. OTAs and hoteliers can target CCEB with value-oriented eco-friendly packages that do not compromise on cost. This strategy could be augmented by leveraging scarcity cues, as this group shows responsiveness to such marketing tactics. Given their preference for quality at reasonable rates, OTAs should present them with transparent review-based quality indicators, aligning with the social proof concept highlighted by Jamal et al. [22]. Hoteliers can emphasize their sustainable practices without additional costs, potentially appealing to the CCEB segment's eco-awareness. Scarcity-based promotions can also be effective, nudging this cost-sensitive segment towards quicker booking decisions.

Green-Urban Deal Hunters. Green-Urban Deal Hunters (GUDH) prioritize sustainability but not at the expense of convenience or cost. This reflects the VALS framework's principles, where values like environmentalism coexist with pragmatic purchase behaviors [11]. Their urban-centric preferences suggest a lifestyle-oriented segmentation approach, as discussed by Weinstein [12]. For OTAs, the strategy should focus on well-rated, centrally-located hotels with clear sustainability features, providing a mixture of urban experience and eco-consciousness. Offering dynamic pricing and limited-time offers could effectively target GUDH, appealing to their deal-seeking nature without abstaining to their green values. Hoteliers can attract GUDH by showcasing their sustainable credentials and proximity to urban attractions, potentially incorporating flexible pricing strategies that reflect the value of their location and green initiatives.

Social-Proof Assurance Seekers. Social-Proof Assurance Seekers (SPAS) are heavily influenced by the experiences of others, as seen in their reliance on reviews and ratings. Their behavior underpins the theories of social validation and assurance in consumer behavior [16]. The cluster's willingness to pay more for socially validated quality points to the trust economy's impact highlighted by Komiak & Benbasat [30]. OTAs should implement reputation-based recommendation systems, highlighting hotels with high ratings and numerous reviews. Personalized marketing communications that cite customer testimonials and ratings can resonate well with SPAS, reinforcing the quality assurance they seek. Hoteliers should encourage satisfied guests to leave positive reviews and can design experiences that are likely to be shared on social media, leveraging the power of user-generated content to build trust and influence booking decisions.

Budget-Only Focused Minimalists. The Budget-Only Focused Minimalists (BOFM) cluster's focus on cost above all reflects the Market Segmentation Theory's cost-focused consumer group [10]. Their limited interest in ratings or sustainability cues suggests a functional approach to booking, consistent with Morrison's [7] discussion on budget-driven travel behavior. For OTAs, this indicates the necessity of a stripped-down, price-focused marketing approach. Highlighting the lowest available prices and basic amenities could effectively capture this segment. Bundling options are less likely to appeal to BOFM unless they present clear cost-saving opportunities. Hoteliers can satisfy BOFM by offering basic accommodations and transparent pricing, ensuring that guests don't pay for unnecessary extras, thereby aligning with their budget-focused values.

Luxury-Quality Connoisseurs. Luxury-Quality Connoisseurs' (LQM) preference for high-quality, luxurious experiences aligns with the psychographic segmentation that associates lifestyle and luxury [14]. Their appreciation for sustainability when choosing between high-end options reflects a premium consumer's sophisticated decision-making process, as evidenced by the work of Apostolakis et al. [24]. OTAs targeting LQM should focus on curating a selection of premium, high-starred hotels that highlight both luxury and sustainability. Personalized high-touch services, loyalty rewards, and exclusive offers can cater to their expectations for a tailored experience, reinforcing the importance of a customer-centric approach as discussed by Paluch & Tuzovic [29]. For hoteliers, this means providing flawless service and high-quality amenities. They could also create exclusive sustainable programs that appeal to LQMs, offering a synthesis of luxury and environmental responsibility.

6 Concluding Remarks and Limitations

The practical contribution of our research lies in its application to the eTourism industry, providing businesses with a nuanced understanding of customer segments through the clustering of user characteristics. This segmentation enables the creation of tailored marketing campaigns, such as targeting eco-conscious travelers with green travel packages, thereby enhancing the precision and effectiveness of marketing efforts. Moreover, our findings inform the customization of booking platforms to align with specific consumer preferences, thereby elevating the user experience and potentially increasing customer loyalty. The insights also aid eTourism companies in making strategic decisions, optimizing resource allocation based on the attributes most valued by their clientele. This approach not only streamlines operations but also furnishes eTourism operators with a competitive edge by facilitating the delivery of personalized customer experiences.

The theoretical contribution of this paper to the scientific community, particularly within the domain of eTourism, lies in its comprehensive analysis of consumer behavior during the hotel booking process. Traditionally, research in this area has predominantly focused rather on internal factors of OTAs, such as hotel attributes and how they influence customer decisions. However, our study extends this perspective by integrating external factors, specifically the diverse characteristics of user groups, into the evaluation process. By employing an approach that considers both the attributes presented by OTAs and the distinct preferences of various user clusters, our research highlights the multi-dimensional nature of the booking process. This methodology may enrich the current understanding of how internal factors, like hotel attributes, impact consumer choice and also how these choices are nuanced by the external factors, such as the socio-demographic profiles of the users and their unique travel motivations.

The primary limitation of this study is the reliance on self-reported data of respondents, which may introduce bias. Additionally, the study's focus on a specific demographic within Austria and Germany limits its generalizability to other regions and cultures. Future research should explore the applicability of the proposed segmentation model across diverse global markets and investigate the impact of real-time data analytics on the accuracy of user segmentation. Our study centers on the overarching findings of the k-means clustering approach, and as such, we do not delve into the detailed outcomes of the web content mining and multiple linear regression analyses. The specific regression results fall outside the scope of this paper, with our focus being on the superior insights derived from the conjoint analysis and k-means clustering. We also recognize that selecting a five-cluster solution based primarily on fit metrics may raise concerns about overfitting and limit the robustness and applicability of our findings across different datasets, which we identify as a limitation of our study. Moreover, longitudinal studies could provide insights into the stability of the identified segments over time, and experimental designs could test the effectiveness of tailored marketing strategies derived from the segmentation model.

Acknowledgments. We would like to express our gratitude for the analytical tools provided by Sawtooth Software Lighthouse Studio, which significantly contributed to the data analysis and insights presented in this paper.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Gretzel, U., Sigala, M., Xiang, Z., Koo, C.: Smart tourism: foundations and developments. Electron. Mark. **25**, 179–188 (2015)
2. Xiang, Z., Du, Q., Ma, Y., Fan, W.: A comparative analysis of major online review platforms: implications for social media analytics in hospitality and tourism. Tour. Manage. **58**, 51–65 (2017)
3. Law, R., Qi, S., Buhalis, D.: Progress in tourism management: a review of website evaluation in tourism research. Tour. Manage. **31**(3), 297–313 (2010)
4. Buhalis, D., Foerste, M.: SoCoMo marketing for travel and tourism: empowering co-creation of value. J. Destin. Mark. Manag. **4**(3), 151–161 (2015)
5. Li, J., Xu, L., Tang, L., Wang, S., Li, L.: Big data in tourism research: a literature review. Tour. Manage. **68**, 301–323 (2018)
6. Bahja, F., Cobanoglu, C., Berezina, K., Lusby, C.: Factors influencing cruise vacations: the impact of online reviews and environmental friendliness. Tourism Rev. **74**(3), 400–415 (2019). https://doi.org/10.1108/tr-12-2017-0207
7. Morrison, A.M.: Marketing and Managing Tourism Destinations. Routledge, London (2013)
8. Eibl, S., Auinger, A.: On the Role of User Interface Elements in the Hotel Booking Intention: Analyzing a Gap in State-of-The-Art Research. In: International Conference on Human-Computer Interaction, pp. 170–189. Springer (2023). https://doi.org/10.1007/978-3-031-35969-9_12
9. Kotler, P., Keller, K.L.: Marketing Management (15th global ed.). England: Pearson, pp. 803–829 (2016)
10. Smith, W.R.: Product differentiation and market segmentation as alternative marketing strategies. J. Mark. **21**(1), 3–8 (1956)
11. Plummer, J.T.: The concept and application of life style segmentation: the combination of two useful concepts provides a unique and important view of the market. J. Mark. **38**(1), 33–37 (1974)
12. Weinstein, A.: Handbook of market segmentation: Strategic targeting for business and technology firms. Routledge (2013)
13. Wind, Y.J., Green, P.E.: Some Conceptual, Measurement, and Analytical Problems in Life Style Research. Marketing Classics Press (2011)
14. Kahle, L.R., Chiagouris, L.: Values, Lifestyles, and Psychographics. Psychology Press (2014)

15. Sigala, M.: Integrating customer relationship management in hotel operations: managerial and operational implications. Int. J. Hosp. Manag. **24**(3), 391–413 (2005)
16. Gretzel, U., Fesenmaier, D.R.: Experience-based internet marketing: an exploratory study of sensory experiences associated with pleasure travel to the Midwest United States. In: ENTER, pp. 49–57 (2003)
17. Fotaki, G., Spruit, M.R., Brinkkemper, S., Meijer, D.: Exploring big data opportunities for online customer segmentation. Int. J. Bus. Intell. Res. **5**, 58–75 (2014)
18. Buhalis, D., Law, R.: Progress in information technology and tourism management: 20 years on and 10 years after the Internet—The state of eTourism research. Tour. Manage. **29**(4), 609–623 (2008)
19. Gan, M.Y., Ouyang, Y.: Study on tourism consumer behavior characteristics based on big data analysis. Front. Psychol. **13**, 876993 (2022). https://doi.org/10.3389/fpsyg.2022.876993
20. Sigala, M.: Social media marketing in tourism and hospitality. Inf. Technol. Tourism **15**(2), 181–183 (2015). https://doi.org/10.1007/s40558-015-0024-1
21. Kwon, K., Kim, C.: How to design personalization in a context of customer retention: who personalizes what and to what extent? Electron. Commer. Res. Appl. **11**, 101–116 (2012)
22. Jamal, S., Newbold, K.B.: Factors associated with travel behavior of millennials and older adults: a scoping review. Sustainability **12**(19), 8236 (2020)
23. Xiang, Z., Magnini, V.P., Fesenmaier, D.R.: Information technology and consumer behavior in travel and tourism: insights from travel planning using the internet. J. Retail. Consum. Serv. **22**, 244–249 (2015). https://doi.org/10.1016/j.jretconser.2014.08.005
24. Apostolakis, A., Jaffry, S., Kourgiantakis, M.: Examination of individual preferences for green hotels in Crete. Sustainability **12**(20), 8294 (2020). https://doi.org/10.3390/su12208294
25. Merinov, P., Massimo, D., Ricci, F.: Sustainability driven recommender Systems. In: Italian Information Retrieval Workshop (2022)
26. Ricci, F., Rokach, L., Shapira, B.: Recommender Systems: Introduction and Challenges. Recommender Systems Handbook, pp. 1–34 (2015)
27. Lops, P., De Gemmis, M., Semeraro, G.: Content-based recommender systems: State of the art and trends. Recommender Systems Handbook, pp. 73–105 (2011)
28. Dzulfikar, M.F., et al.: Personalization features on business-to-consumer e-commerce: review and future directions. In: 2018 4th International Conference on Information Management (ICIM), pp. 220–224 (2018)
29. Paluch, S., Tuzovic, S.: Persuaded self-tracking with wearable technology: carrot or stick? J. Serv. Mark. **33**(4), 436–448 (2019)
30. Komiak, S.Y., Benbasat, I.: The effects of personalization and familiarity on trust and adoption of recommendation agents. MIS Quarterly, pp. 941–960 (2006)
31. Gal-Or, E., Gal-Or, R., Penmetsa, N.: The role of user privacy concerns in shaping competition among platforms. Inf. Syst. Res. **29**, 698–722 (2018)
32. Martin, K.D., Murphy, P.E.: The role of data privacy in marketing. J. Acad. Mark. Sci. **45**, 135–155 (2017)
33. Pariser, E.: The Filter Bubble (2012)
34. Praditya, N.W.P.Y., Erna Permanasari, A., Hidayah, I.: Designing a tourism recommendation system using a hybrid method (Collaborative Filtering and Content-Based Filtering). In: 2021 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT), pp. 298–305 (2021)
35. Baier, D., Brusch, M.: Methoden - Anwendungen - Praxisbeispiele. In: Baier, D., Brusch, M. (eds.) Conjointanalyse (2009). https://doi.org/10.1007/978-3-662-63364-9
36. Schegg, R.: European Hotel Distribution Study 2022. University of Applied Sciences and Arts of Western Switzerland, Statista (2022)
37. Field, A.: Discovering Statistics using IBM SPSS Statistics. Sage (2013)

38. Tabachnick, B.G., Fidell, L.S., Ullman, J.B.: Using Multivariate Statistics. Pearson Boston, MA (2013)
39. Orme, B.K.: Getting started with conjoint analysis: strategies for product design and pricing research (2006)
40. Hair, J.F.: Multivariate data analysis (2009)
41. Allenby, G.M., Rossi, P.E.: Hierarchical bayes models. In: The Handbook of Marketing Research: Uses, Misuses, and Future Advances, pp. 418–440 (2006)
42. Jain, M., Kaur, G., Saxena, V.: A K-Means clustering and SVM based hybrid concept drift detection technique for network anomaly detection. Expert Syst. Appl. **193**, 116510 (2022)
43. Steinley, D.L.: K-means clustering: a half-century synthesis. Br. J. Math. Stat. Psychol. **59**(Pt 1), 1–34 (2006)
44. MacQueen, J.: Some methods for classification and analysis of multivariate observations. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, no. 14, Oakland, CA, USA, pp. 281–297 1967
45. Rousseeuw, P.J.: Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. **20**, 53–65 (1987)
46. Wedel, M., Kamakura, W.A.: Market segmentation: Conceptual and methodological foundations. Springer Science & Business Media (2000). https://doi.org/10.1007/978-1-4615-4651-1
47. Kass, R.E., Raftery, A.E.: Bayes factors. J. Am. Stat. Assoc. **90**(430), 773–795 (1995)
48. Milligan, G.W., Cooper, M.C.: A study of standardization of variables in cluster analysis. J. Classif. **5**, 181–204 (1988)
49. Ja, H.: A k-means clustering algorithm. JR Stat. Soc. Ser. C-Appl. Stat. **28**, 100–108 (1979)
50. Malhotra, N.K.: Marketing Research: an Applied Prientation. Pearson (2020)
51. So, K.K.F., King, C., Sparks, B.A., Wang, Y.: The role of customer engagement in building consumer loyalty to tourism brands. J. Travel Res. **55**, 64–78 (2016)