DIGITAL TEXT ANALYSIS OF TOURISTS' SENTIMENTS AND SPATIAL CHOICES IN GLOBAL CITIES

Umut Turk

Abdullah Gul University, Kayseri, Turkey

Barbaros, Sümer Kampüsü, Erkilet Blv., 38080 Kocasinan/Kayseri, Turkije; ORCIDiD: 0000-0002-8440-7048, umut.turk@agu.edu.tr)

John Östh

Oslo Metropolitan University, Oslo, Norway

Office number: PE830

Pilestredet 35, 0166 Oslo, Norway; ORCID iD: 0000-0002-4536-9229; johnosth@oslomet.no

Karima Kourtit

Open University, The Netherlands; EHERO Erasmus University Rotterdam, The Netherlands; University College Wroclaw (WSKZ), Poland; ORCID iD: 0000-0002-7171-994X; k_kourtit@hotmail.com

Peter Nijkamp

EHERO Erasmus University Rotterdam, The Netherlands; University College Wroclaw (WSKZ), Poland; SPPM Tsinghua University, Beijing, China; ORCID iD: 0000-0002-4068-8132;

pnijkamp@hotmail.com

Abstract

The peer-to-peer rental market has in recent years become a prominent part of the tourism sector world-wide. Its open platform character, supported by digital technology, has not only facilitated choice-making processes by accommodation owners and heterogeneous visitors, but also generated a wealth of accessible data for quantitative tourism research and local tourism management. In particular, the value and preference statements expressed by visitors on the open platform's website contain rich and detailed valuation and perception information, which can nowadays be transformed into numerical data using advanced text analysis (e.g. mood analysis, sentiment analysis) and complementary statistical techniques for analyzing consumers' tastes. This study seeks to identify patterns and drivers of geographical choices and appreciations of visitors to a given city, based on their site-specific preferences and satisfaction regarding the quality of their accommodation and of the geographic locational attractiveness of the urban area concerned. To distil meaningful and generalizable information from peer-to-peer rental accommodations across multiple tourism destinations, the paper examines the appreciation of users of listed Airbnb facilities in 34 major tourist places in the world. The main focus will be on the heterogeneity in Airbnb attractiveness as perceived by tourists in terms of accommodation characteristics (e.g. quality of the Airbnb facility), locational convenience (e.g. access to public transport), and geographical appeal of the area or city concerned (e.g. cultural amenities or historical ambiance). The spatially differentiated perspective is theoretically framed in a geographical attraction model encompassing the multi-faceted and place-specific perceptions and choices of visitors. This multiscalar sentiment analysis will next be formalized and tested by a spatial-econometric model for tracing geographic perceptions and choices of tourist visitors in 34 global tourist destinations. The database on these 34 places contains both objective location-specific statistical data and subjective visitor information extracted from social media and open-source platforms. The latter type of information is translated into measurable statistical indicators through mood or sentiment metrics and machine learning techniques. The resulting operational model aims to estimate the visitors' choice and satisfaction determinants regarding peer-to-peer accommodation choices and the availability of local services and amenities. will be given to the geographic clustering of tourist perceptions, using K-median metrics to identify patterns of tourist hotspots and areas with high concentrations of visitors' (dis)satisfaction, while also crowding effects in preference formation are taken into consideration. Finally, particular attention is given to tourist heterogeneity (e.g. socio-economic differences) explaining cultural variations in sentiment expressions. This approach also allows to map out how tourists with different backgrounds may evaluate the same urban environment differently. The site-specific information on tourists' perceptions and satisfaction inprovides a nuanced value picture of the diversified spectrum of geographic perceptions and choices - and their place-specific drivers - by tourists in various big cities in the world. This insight is critical for the effective management of the variety in spatial tourist flows in urban destinations.

Keywords:

digital text analysis, tourist sentiment, spatial choices, peer-to-peer rentals, geographical attractiveness, sentiment analysis, urban tourism management, spatial econometrics

1. Introduction

In the age of global tourism many cities experience nowadays an increasingly unprecedented influx of both domestic and foreign visitors, a trend that is sometimes only interrupted by external shocks such as a pandemic, a banking crisis or war conditions. It is noteworthy however, that tourists do not form a homogeneous group of people and hence show a variety of preferences for visiting different tourist destinations and amenities. This heterogeneity in tourist preferences is also reflected in the increasing segmentation of global tourist markets, further complicating the task of managing tourism environments and urban amenities. Besides, tourist cities offer a variety of tourist services and attractions, ranging from sun-and-beach tourism to cultural or eco-tourism (Romao, 2023), so that – apart from financial considerations – the spectrum of tourists' place-based evaluations and experience-forming processes is very rich (Suzuki et al., 2021). Meanwhile, the increased visitor intensity in several tourist destinations has led to many negative externalities in the form of geographic crowding effects in tourist areas (e.g. Venice, Barcelona, Paris, Rome, Amsterdam), which are increasingly leading to management and policy issues. A new trend, often termed 'de-touristification', is nowadays coming to the fore which seeks to re-channel tourist flows through a process of geographical re-distribution around major tourist concentrations (see e.g. Lim et al., 2025). This shift toward decentralizing tourist activity presents both opportunities and challenges for local authorities. However, any management and policy decisions in tourism markets require a clear and evidence-based understanding of how visitors evaluate and experience urban environments. The present study seeks to provide new empirical evidence on tourists' place-based satisfaction patterns across large global destinations by analysing structured information derived from Airbnb review texts.

The cultural, historical, physical, ecological, and socio-economic differences in the relative attractiveness of tourist amenities in major cities of the world prompt the rise of local and regional economic specialization in the supply of tourist services among destination places. This pluriformity in the supply and perceived quality of tourist services needs a differentiated perspective, recognizing that tourists interact with visited cities in multi-faceted and place-specific ways, including their responses to factors such as safety, accessibility, cultural ambiance, and perceived quality of the built environment. Understanding the complexity in tourists' satisfaction in a given city therefore requires a spatially differentiated perspective, in which, in addition to generic urban attractiveness factors, also site-specific and locational characteristics (e.g. neighbourhood ambiance, district atmosphere, accessibility conditions) are taken into consideration (Joo et al., 2017; Osth et al., 2025). To meet these research challenges, this paper resorts to large-scale digital information generated by tourists themselves on peer-to-peer accommodation platforms.

The study of the benefits of internet technology for the tourism sector has already a long history (see e.g. Buhalis & Law, 2008; Navio-Marco et al., 2018; Osth et al., 2025). Nowadays, the tourism industry is closely intertwined with digital platforms that facilitate booking, review exchange, and information transparency. Clearly, tourism has become an example of an advanced e-service sector (Labanauskaite et al., 2020). From this perspective, user-generated textual content on accommodation platforms provides a valuable source of management-relevant information (Cheng et al., 2021). Such information is not only available in numerical rating form, but also in textual narratives that reflect guests' perceptions and experiences. These narratives allow researchers to extract structured indicators of sentiment and thematic emphasis, enabling systematic comparison of perceived urban environments across destinations.

The digital treatment of textual information has made considerable progress in recent years. In this study, rather than applying black-box artificial intelligence or large language models, we adopt a transparent lexicon-based sentiment scoring approach (using the R *sentimentr* package) combined with dictionary-based thematic extraction. This approach allows us to quantify both the emotional tone of reviews and the relative prominence of key urban experience dimensions such as safety and cleanliness, tourism appeal, transport, cost, and work-related suitability. Language expression thus becomes a measurable vehicle for tracing how visitors describe and evaluate their accommodation environments within different cities. While individual socio-demographic characteristics of reviewers are not directly observable in our dataset, the large cross-city scope allows us to compare aggregate experience patterns across diverse urban contexts.

The main aim of the study is therefore to investigate to what extent tourists expressed satisfaction and perceived place quality are associated with geographic and listing-level characteristics within and across cities. The analysis identifies the relative importance of both tangible factors (e.g. pricing, minimum stay conditions, accessibility-related perceptions) and intangible experience dimensions (e.g. cleanliness, safety, tourism appeal) in shaping overall sentiment. By adopting a multilevel modelling framework, the study distinguishes listing-level effects from neighbourhood- and city-level contextual variation, enabling systematic comparison of tourism environments across destinations.

In our empirical application, we analyse Airbnb review and listing data for 34 global cities. The dataset combines textual reviews with listing-level characteristics such as price, number of reviews, minimum nights, entry timing, and distance to city centres. The textual information is transformed into structured variables through lexicon-based sentiment scoring and dictionary-based thematic quantification. These indicators are subsequently analysed using Principal Component Analysis (PCA) to uncover latent perception dimensions, K-means clustering to identify typologies of short-term rental environments, and multilevel mixed-effects models to explain variation in listing-level sentiment across cities and neighbourhoods. The novelty of our work lies in the integration of cross-city comparative analysis with structured text quantification, dimensionality reduction, clustering, and hierarchical modelling within a unified management-oriented framework.

This paper is organized as follows: Section 2 presents a selective literature review. Section 3 outlines the research framing and workflow. Section 4 details the database and methodology. Section 5 presents the empirical results, followed by an interpretation of the findings in the context of tourism management across multiple spatial scales. Finally, Section 6 concludes the paper by reflecting on the findings and offering implications for urban tourism governance and destination management.

2. Selective Literature Review

The literature on urban tourist attractiveness, particularly in the context of digital platforms and user-generated content, is vast and growing. Open-access digital data on tourists' place-based evaluations and accommodation experiences have opened many promising avenues for quantitative research, particularly when used with peer-to-peer rental platforms. While mood and sentiment analysis has gained attention recently, research on **structured interpretation of intra-city experience variation using large-scale comparative data** remains underdeveloped and fragmented. Below, we summarize key studies that have informed this paper's approach.

Kourtit et al. (2022) analyse the space-time vulnerability and resilience of Airbnb markets in six global cities during the COVID-19 pandemic. Using large-scale listing data and spatial-econometric models, they reveal that tourist demand, pricing dynamics, and survival probabilities of listings are significantly influenced by intra-urban location, density patterns, and neighbourhood characteristics. Their study shows the importance of integrating spatial modelling with digital tourism data to understand tourist behaviour, a principle that directly informs the present research. In particular, their work shows the potential for using spatial analytics in combination with digital data to assess shifts in tourism patterns during crises.

Ding et al. (2020) use Structural Topic Modelling (STM) to analyse over 600,000 Airbnb reviews in Malaysia, identifying latent dimensions of service quality. Their work finds that tourists' textual evaluations predominantly concern tangibles, service adequacy, caring attitudes, assurance, and convenience. Their study demonstrates how review texts can be systematically transformed into structured indicators of perceived service quality. This provides a better understanding of how tourists express preferences and evaluations of specific attributes in their accommodation choices, which is integral to understanding spatial perceptions and preferences in urban tourism.

Vinogradov et al. (2020) investigate the determinants of Airbnb likability in Istanbul, focusing on service attractiveness, perceived authenticity, country image, and user experience through the lens of social exchange theory. Their study shows that authenticity and experiential value are central to shaping tourists' evaluations, with these intangible, place-related attributes complementing the more functional aspects of accommodation. Their findings align with the growing understanding that emotional and cultural factors shape spatial choices, contributing to a more holistic view of urban tourism experiences. This shows how specific factors influence their spatial choices and sentiments, reinforcing the importance of geographical attributes in shaping urban tourism experiences.

Robertson et al. (2019) explore short-term rentals as a driver of digitally mediated tourism gentrification in New Orleans. They argue that platforms like Airbnb extend tourism beyond traditional tourist zones, reshaping urban spaces and consumption patterns. Tourists' spatial choices, according to their findings, reflect a growing preference for "authentic" and everyday urban environments. This study emphasizes the need for a spatially differentiated approach to understanding intra-city tourism geographies, strengthening the current paper's focus on how tourists' evaluations of place characteristics impact their spatial choices. Their work raises important questions about the socio-economic impact of tourism expansion into residential neighbourhoods and its implications for local communities.

Cheng and Jin (2019) analyse Airbnb reviews to identify the key dimensions of user concern and satisfaction, using text-mining techniques. Their study shows that tourists' evaluations focus on cleanliness, location, host responsiveness, accessibility, and neighbourhood environment. They conclude that locational convenience and surrounding urban amenities are crucial determinants of perceived accommodation quality. This finding directly supports our focus on spatial attributes and highlights the relevance of neighbourhood factors in tourists' experiences. This study shows the close linkage between textual sentiments and spatial attributes, particularly when tourists evaluate the areas where they stay, which is central to this study's focus on place-based perceptions in urban tourism.

Lawani et al. (2019) apply sentiment analysis to Airbnb reviews in Boston, exploring the relationship between emotional expressions and pricing. Their findings show that positive sentiments are associated with higher prices and stronger demand, while negative emotions reflect locational disadvantages or service shortcomings. This finding underscores the utility of sentiment analysis in revealing tourist preferences for specific locales and amenities. This study supports the use of sentiment indicators extracted from textual data as reliable proxies for tourists' valuation processes, a methodological approach central to this paper's analysis.

The studies reviewed demonstrate the value of digital platforms and user-generated textual data in understanding tourists' perceptions, preferences, and spatial behaviours. Factors such as accommodation quality, neighbourhood features, perceived authenticity, and experiential value influence tourists' evaluations and movement patterns. Such factors show the complexity of tourists' decision-making processes, where both tangible (e.g., infrastructure) and intangible (e.g., emotional satisfaction) elements come into play. Advanced methods like sentiment analysis and structural topic modeling effectively transform qualitative data into measurable insights. However, most research focuses on individual cities or small regions, with few comparative studies using spatial econometrics across multiple cities. This shows a significant gap in the literature, particularly regarding cross-city comparisons of intra-urban tourist behaviours and their relationship with spatial factors.

The literature reflects the importance of combining sentiment analysis with spatial econometrics to capture tourists' experiences and mobility. While the analysis of individual urban spaces is well-established, the challenge remains to apply a multi-city framework that accounts for the diverse attractions and local contexts of global cities. This research aims to bridge that gap by using sentiment-based models in a multi-city context, offering a deeper understanding of how spatial factors influence tourist satisfaction. Additionally, quantitative spatial-economic analyses of intra-city tourism remain underdeveloped, despite the potential of digital data. Comparative studies across destinations with diverse intra-urban attractions are still in the early stages.

While the potential of digital data in urban tourism research is clear, comparative studies across destinations with diverse intra-urban attractions are still in the early stages. As cities evolve and diversify, so too do the ways tourists interact with them. To address this gap and build on the findings from the literature, the following research propositions are put forward to direct the analysis of tourists' sentiments and spatial choices across global cities.

• How do tourists' sentiment expressions regarding their accommodation quality and surrounding amenities relate to overall satisfaction levels?

• To what extent do geographic and listing-level factors explain variation in tourists' evaluations across different urban destinations?

• How can sentiment indicators, thematic text quantification, and multilevel modelling be integrated to compare tourism environments across cities?

• What are the key geographic and platform-related drivers of tourists' satisfaction in global cities, and how do these factors vary across different urban contexts?

3. Research Framework

The main objective of this study is to understand how tourists' perceptions and place-based evaluations are influenced by the geographical and locational characteristics of urban destinations. Specifically, we aim to investigate the impact of various place-based attributes - such as accommodation quality, locational convenience, and neighbourhood ambiance - on tourists' satisfaction and experienced spatial context within global cities. In addition, the study will explore the variation in perceived experiences across listing, neighbourhoods, and cities and how these dimensions shape sentiment expressions in social media reviews.

Theoretical Framework

This study is based on a geographical attraction model, which posits that tourists' satisfaction and place-based evaluations are not only shaped by the material attributes of a destination (e.g., accommodation quality, proximity to landmarks, transport infrastructure), but also by intangible, place-specific factors, such as the cultural ambiance, historical appeal, and the atmosphere of particular neighbourhoods. This framework is supported by recent research that highlights the importance of geographical and socio-economic diversity in shaping tourists' perceptions of urban areas. The framework also recognizes that high visitor intensity in major destinations may influence how tourists perceive and evaluate urban environments, particularly in central areas experiencing sustained tourism pressure.

This study seeks to explore several important aspects related to tourist behaviours and spatially differentiated preferences. This includes analysing the role of neighbourhood characteristics, such as local ambiance and the presence of cultural or historical landmarks, in shaping their satisfaction levels and perceived place quality. It aims to examine how tourists' expressions of sentiment about their accommodation quality and nearby amenities affect their evaluations of locations within a city.

Another focus of the study is to investigate the role of geographic and place-based factors in explaining the differences in tourists' preferences and sentiments across various urban destinations (heterogeneity in tourist preferences). These differences shape not only how tourists perceive different locations within cities, but also how their satisfaction levels might vary across diverse urban contexts. This aspect shows the dynamic nature of tourist behaviour, as well as the influence of individual and cultural factors on urban tourism.

The research will also explore how sentiment analysis, when combined with multilevel modelling, can be used to model and predict patterns of intra-city tourism satisfaction across urban contexts. By examining sentiment data from tourists' online reviews, we can quantify tourists' emotional reactions to specific places and amenities. When integrated with hierarchical statistical models, this data will help identify the key geographic drivers of tourist satisfaction and mobility at a sub-city level, offering insights into the complex interaction between tourists and urban spaces.

In addition, the study will identify the key geographic factors that influence tourist satisfaction in global cities and examine how these factors vary across different types of tourist destinations, such as cultural, historical, and eco-tourism sites (cross-urban contexts). These objectives form the foundation of the study's analytical approach and shape the methodology used to assess tourist behaviors and spatial preferences at a detailed, sub-city level. These insights will contribute to a more nuanced understanding of tourist behavior in diverse urban settings, particularly as cities develop and transform in response to changing tourism dynamics.

Research Workflow

The research workflow is divided into several key stages (see Figure 1). First, data collection and preprocessing will take place (data collection and preprocessing). During the first stage, tourist perceptions and preferences will be gathered from Airbnb review texts and listing-level platform data. These sources provide valuable perspectives into tourists' emotional reactions to their experiences through user-generated content. Furthermore, spatial and listing-level information will be collected directly from the platform dataset (e.g., price, minimum nights, review counts, entry timing, and distance to the city center) across 34 global cities. This will include transport networks, cultural amenities, public services, and neighborhood features. These platforms offer user-generated textual content, which provides valuable insights into tourists' emotional reactions to their experiences. Objective data on the geographical, cultural, and infrastructural characteristics of 34 global cities will be collected from well-established databases like the Global Power City Index (GPCI). This will include information on transport networks, cultural amenities, public services, and neighborhood features. The textual data will then undergo preprocessing, which involves cleaning, tokenization, and sentiment labeling, to prepare it for analysis.

In the second stage, sentiment analysis and textual data processing will be conducted. In this stage, lexicon-based text processing methods will be applied to process the textual data. Sentiment analysis will be implemented, which allows the quantification of emotional polarity expressed in visitors' reviews. By categorizing sentiment, we can gain a better understanding of the factors driving tourists' satisfaction or dissatisfaction with various aspects of their accommodations and surroundings. Furthermore, thematic indicators will be constructed using a custom dictionary, capturing the relative prominence of key experience dimensions such as safety and cleanliness, tourism and leisure, transport, cost, and work-related references. This step transforms unstructured textual narratives into structured perception-based variables suitable for statistical modelling.

The third stage involves clustering and exploratory spatial comparison. K-means clustering will be used to group listings with similar perception profiles and characteristics. This will help identify areas of high tourist satisfaction and dissatisfaction within different districts or neighborhoods of each city. **Principal Component Analysis (PCA) will first be applied to the thematic variables to uncover latent perception dimensions that summarize guest experience structures.** These spatial clusters will be further analyzed to understand how specific geographic factors influence tourists' satisfaction patterns across urban contexts. Proximity to cultural amenities, public transport accessibility, and neighborhood ambiance will be key factors examined to understand how they impact overall satisfaction. The clustering analysis will also help to identify **distinct typologies of short-term rental environments across cities rather than explicitly modelling crowding dynamics.** Geographic clustering of tourist perceptions will identify areas where visitors have strong preferences or significant dissatisfaction. These insights will guide policy recommendations to redistribute tourist traffic, reducing crowding and improving satisfaction levels.

The fourth stage consists of empirical analysis and model testing. The sentiment indicators derived from textual data will be combined with listing-level and spatial context variables extracted from the platform dataset. This will serve as input for testing multilevel (mixed-effects) models that estimate the relative importance of various geographic and listing characteristics (such as distance to central areas, accessibility-related perceptions, and local amenities reflected in reviews) in shaping tourist satisfaction. The models focus on variation across listings, neighbourhoods, and cities rather than individual socio-demographic differences. This stage will explore the complex relationship between tourists' experiences and the places they stay. Random intercepts for cities and neighbourhoods allow us to isolate local contextual effects and compare performance across destinations.

Finally, the fifth stage will focus on interpretation and policy implications. The results from the multilevel modelling framework will be analyzed in the context of tourism management and urban planning. The findings will offer valuable insights into how cities can enhance the quality of their tourist offerings and better manage tourism environments. Special attention will be given to identifying policy-relevant insights that can help improve tourist satisfaction and support more balanced and sustainable tourism development. The findings from the analysis will inform recommendations for managing tourism environments, enhancing the quality of urban tourist offerings, and promoting more sustainable tourism practices.

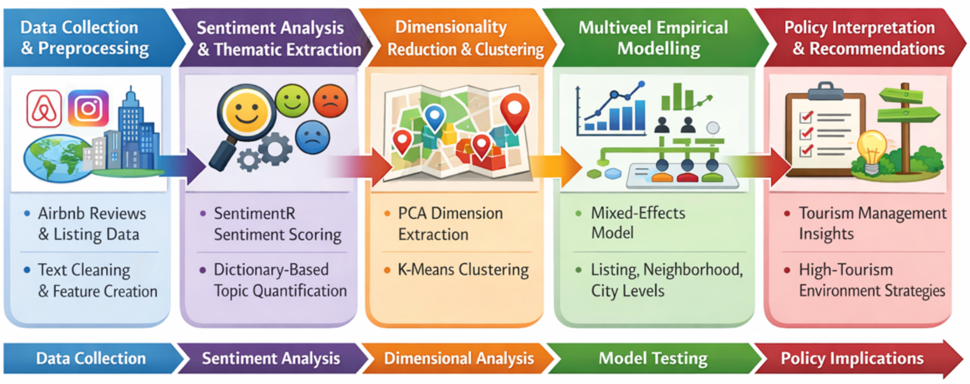

Figure 1. Research Workflow

3. Data and Methods

3.1 Data Sources

The empirical analysis is based on a cross-sectional dataset of Airbnb listings collected from the InsideAirbnb open-data platform. InsideAirbnb provides publicly available structured information on short-term rental markets, including listing characteristics, geographic coordinates, neighbourhood identifiers, and user-generated review content. The data used in this study were extracted in December 2025. The dataset we constructed covers 34 global cities. These cities were selected on the basis of two criteria: (i) their international prominence in terms of tourism activity, and (ii) the high intensity of Airbnb operations as indicated by publicly available platform statistics (ref).

The primary unit of analysis is the individual Airbnb listing. Each listing is uniquely identified by a listing ID that reflects the order of its entry into the platform. This variable enables us to approximate market experience by constructing entry quantiles within each city. Listings that entered the platform earlier are considered more experienced market participants, while later entrants are newer additions to the local rental ecosystem. This proxy captures listing-level maturity and platform familiarity, which may influence management strategies and guest evaluations. For each listing, the dataset includes nightly price, minimum night requirements, total number of reviews, neighbourhood identifiers, and geographic coordinates. Distance to the city centre is computed using the Cartesian distance between the listing coordinates and a predefined central reference point within each city. The city centre in each city was determined by asking a large language model to identify the most human activity. The availability of precise spatial coordinates allows us to preserve intra-urban differentiation and to nest listings within neighbourhoods and cities in the subsequent multilevel analysis.

In addition to structured listing attributes, the dataset includes the full corpus of user-generated textual reviews, totaling around 6 million. For each listing, all available reviews from the date of its first appearance on the platform were collected. To ensure the analysis reflects current market conditions, we restrict the sample to listings that have received at least one review within the 3 months preceding the data extraction date.

The final analytical dataset, therefore, integrates structured listing characteristics, spatial positioning, listing lifecycle information, and aggregated review narratives. This architecture supports both listing-level management analysis and cross-city comparison of tourism environments.

3.2 Methods

To examine how perceived urban experiences and listing characteristics jointly shape visitor satisfaction across global cities, we employ a structured multi-stage empirical framework. The analytical strategy proceeds sequentially from data construction and textual operationalization to latent structure identification and hierarchical modelling. The analysis begins with the construction of a harmonized cross-city dataset derived from Airbnb listings. Structured listing attributes, spatial identifiers, and complete review corpora are integrated at the listing level. Given that textual reviews constitute the primary focus of the paper, the next step involves text quantification. Review narratives are transformed into measurable indicators through two complementary procedures. First, sentiment analysis is implemented using the R *sentimentr* package (ref.), a lexicon-based polarity scoring method. For each review, a sentiment value is computed (ranging continuously from extreme negative (- 1) to extreme positive (+1)), and these values are subsequently aggregated to the listing level to generate an average sentiment score. This listing-level sentiment measure serves as the dependent variable in the econometric analysis and represents the aggregated emotional evaluation expressed by guests.

Second, we develop a domain-specific dictionary to extract thematic frequency measures specific to the urban tourism context. Words are classified into five predefined dimensions: safety and cleanliness, tourism and leisure, transport and accessibility, cost and value, and work-related references. For each review, category-specific word frequencies are calculated relative to total token counts and then aggregated to the listing level. These thematic indicators operationalize the substantive content of guest narratives and allow for separation between affective tone (sentiment) and experiential emphasis (topic salience). The sentiment and thematic variables are merged with the structured listing attributes, resulting in a final analytical dataset comprising 469612 listings nested within 34 cities.

The textual information derived from user reviews contains structured signals that are informative both at the listing level and at broader spatial scales. At the listing level, review narratives reflect host-related management dimensions such as cleanliness, responsiveness, and value perception. At the urban level, the aggregation of these narratives reveals spatial regularities and differentiation patterns across neighbourhoods and cities. To examine whether systematic perception structures exist within the review data, we first apply Principal Component Analysis (PCA) to the thematic variables derived from the custom dictionary.

PCA is employed to detect latent dimensions underlying the observed topic frequencies. The objective of this step is not solely dimensionality reduction, but the identification of interpretable perception axes that characterize how experiential themes co-vary across listings. These latent components allow us to capture dominant narrative patterns and to assess whether consistent perception profiles emerge within neighbourhoods and across cities. The extracted components therefore serve as summary measures of structured experiential emphasis embedded in the textual data.

While PCA is applied exclusively to subjective, text-derived indicators, the subsequent analysis integrates both subjective and objective listing attributes to examine heterogeneity in short-term rental environments more comprehensively. To this end, K-means clustering is implemented using a combined set of listing characteristics (e.g., spatial positioning, pricing, review intensity, entry timing) and thematic indicators. The clustering algorithm partitions listings into internally homogeneous and externally heterogeneous groups by minimizing within-cluster variance. This procedure yields typological classifications of rental environments that reflect joint configurations of spatial location, market maturity, pricing structure, and experiential emphasis. The distribution of clusters across neighbourhoods and cities provides descriptive evidence of structural differentiation in urban tourism markets.

The final stage involves estimation of a linear mixed-effects model to explain variation in listing-level sentiment. Given the hierarchical structure of the data—listings nested within neighbourhoods and neighbourhoods nested within cities—a multilevel specification is employed to account for spatial clustering and unobserved contextual heterogeneity. Fixed effects include entry quantile (proxying listing experience), distance to the city centre, log-transformed price, review intensity, minimum stay requirements, and thematic indicators. Random intercepts are specified at both the neighbourhood and city levels to capture baseline deviations in sentiment that are not explained by observable listing characteristics. In this framework, city-level random effects represent systematic differences in overall evaluation across destinations after controlling for listing-level attributes, potentially reflecting broader contextual or management-related conditions. The multilevel model is formally defined as follows. Let listing i belong to neighbourhood $n(i)$, which is nested within city $j(i)$. The listing-level sentiment score $\bar{s}_i$ is specified as:

$$\bar{s}_i = \beta_0 + \beta_1 \text{EntryQuantile}_i + \beta_2 \text{DistCenter}_i + \beta_3 \log(\text{Reviews}_i) + \beta_4 \log(\text{MinNights}_i) + \beta_5 \log(\text{PriceUSD}_i) + \sum_{c=1}^{5} \gamma_c\ \bar{\theta}_{ic} + u_{j(i)} + v_{n(i)} + \varepsilon_i.$$

In this specification, EntryQuantile_i proxies listing experience on the platform, DistCenter_i measures spatial centrality, and the log-transformed variables capture review intensity, minimum stay requirements, and pricing structure. The term $\bar{\theta}_{ic}$ represents the listing-level thematic frequency for category c, where $c \in \{1, \dots, 5\}$ corresponds to the five dimensions derived from the custom dictionary.

Random intercepts are specified as:

$$u_j \sim \mathcal{N}(0, \sigma_u^2), v_n \sim \mathcal{N}(0, \sigma_v^2), \varepsilon_i \sim \mathcal{N}(0, \sigma^2).$$

Here, u_j captures city-level deviations in baseline sentiment after controlling for listing-level covariates, while v_n captures neighbourhood-level deviations nested within cities. The residual term ε_i represents listing-specific unexplained variation.

5. Empirical statistical land modeling results

This section reports the empirical results from the integrated analytical framework developed in Sections 3–4. Consistent with the research framing, the analysis proceeds in a progressive manner from descriptive structure identification in reviews to the classification of listing environments and, finally, to hierarchical modelling of satisfaction outcomes. This ordering reflects the logic that user-generated reviews contain both affective evaluations and thematic content signals, which can be organized into interpretable perception structures and subsequently linked to listing- and context-level determinants of satisfaction.

The results are presented in three steps. First, we examine whether visitors' reviews exhibit systematic, interpretable structures by analysing the co-occurrence patterns of the thematic indicators extracted from the reviews. Second, we identify recurring typologies of short-term rental environments using unsupervised clustering, which provides a descriptive classification of how spatial location, pricing structure, market maturity, and experiential emphasis jointly configure listing environments. Third, we estimate multilevel mixed-effects models to explain listing-level sentiment while accounting for neighbourhood and city-level heterogeneity. This modelling step directly links experience signals to management-relevant predictors at different scales and allows cross-city comparison of baseline evaluation patterns after controlling for listing attributes.

Accordingly, Section 5.1 reports the latent perception structures identified through PCA, Section 5.2 presents the cluster typology of listing environments and its spatial distribution across cities, and Section 5.3 reports the multilevel modelling results and interprets the relative importance of listing-level characteristics, perceived experience dimensions, and contextual effects.

5.1 Latent Perception Structures in Review Narratives

A central premise of this study, as developed in the introduction and selective literature review, is that user-generated review narratives contain structured information about how tourists perceive urban environments. While prior research has identified experience dimensions within individual destinations (e.g., Cheng & Jin, 2019; Ding et al., 2020), comparatively little evidence exists on whether such structures are stable and interpretable across a large cross-section of global cities. The first empirical step therefore investigates whether systematic latent dimensions can be identified in the thematic emphasis of review narratives.

Principal Component Analysis (PCA) was applied to the five thematic frequency variables derived from the custom urban tourism dictionary: safety and cleanliness, tourism and leisure, transport and accessibility, cost and value, and work-related references. The eigenvalues and component loadings are reported in Table 1, while the component score distribution across cities is illustrated in Figure 2. As shown in Table 1, the first two principal components account for a substantial share of the total variance in thematic emphasis. The first principal component (PC1) exhibits strong positive loadings on safety/cleanliness and tourism-related references. This structure indicates that these two experiential dimensions systematically co-occur in review narratives. In line with the findings of Cheng and Jin (2019), who highlight the importance of cleanliness and location in shaping perceived accommodation quality, and Vinogradov et al. (2020), who emphasize experiential and authenticity-related attributes, PC1 can be interpreted as an amenity-richness or experiential-quality dimension. Listings with high PC1 scores tend to be associated with reviews emphasizing both service fundamentals and surrounding tourism appeal.

Table 1: PCA on five thematic textual outputs.

Variable	PC1	PC2
SafeCleanMean	0.667	0.046
TourismMean	0.633	0.328
TransportMean	0.219	–0.425
WorkMean	0.230	0.194
CostMean	0.232	–0.820

The second principal component (PC2), as reported in Table 1, loads most strongly on cost and transport-related references. This component reflects a distinct evaluative axis in which narratives are dominated by logistical and value-based considerations. This aligns with the literature identifying price–quality trade-offs and accessibility as core drivers of guest evaluation (Lawani et al., 2019). However, unlike prior single-city studies, our cross-city PCA reveals that logistical salience operates as a structurally separate perception dimension rather than simply a subcomponent of general satisfaction.

The cross-city distribution of component scores, presented in Figure 2, further demonstrates systematic differentiation among destinations. Some cities cluster in the high-PC1, low-PC2 region, indicating narratives dominated by experiential and amenity-related language. Other cities exhibit relatively higher PC2 scores, suggesting stronger emphasis on cost and transport concerns. In particular, according to Figure 2, cities such as Rio de Janeiro, Santiago, and Buenos Aires score high on PC2 but low on PC1. This suggests that guest reviews in these cities tend to focus on practical matters, particularly cost and mobility, while paying less attention to cleanliness or cultural vibrancy. In contrast, cities such as Los Angeles and Auckland (New Zealand) load highly on both PC1 and PC2, which reflect reviews that emphasize both positive amenities and logistical awareness. Guests in these cities appear to value diverse experiences that balance enjoyment with practical convenience.

At the opposite end of the spectrum, Singapore, and to a lesser extent Bangkok and Tokyo, exhibit high PC1 but low PC2 scores. These cities are characterized by reviews that emphasize strong amenity quality, especially cleanliness, safety, and efficiency, while showing limited concern about cost or accessibility.

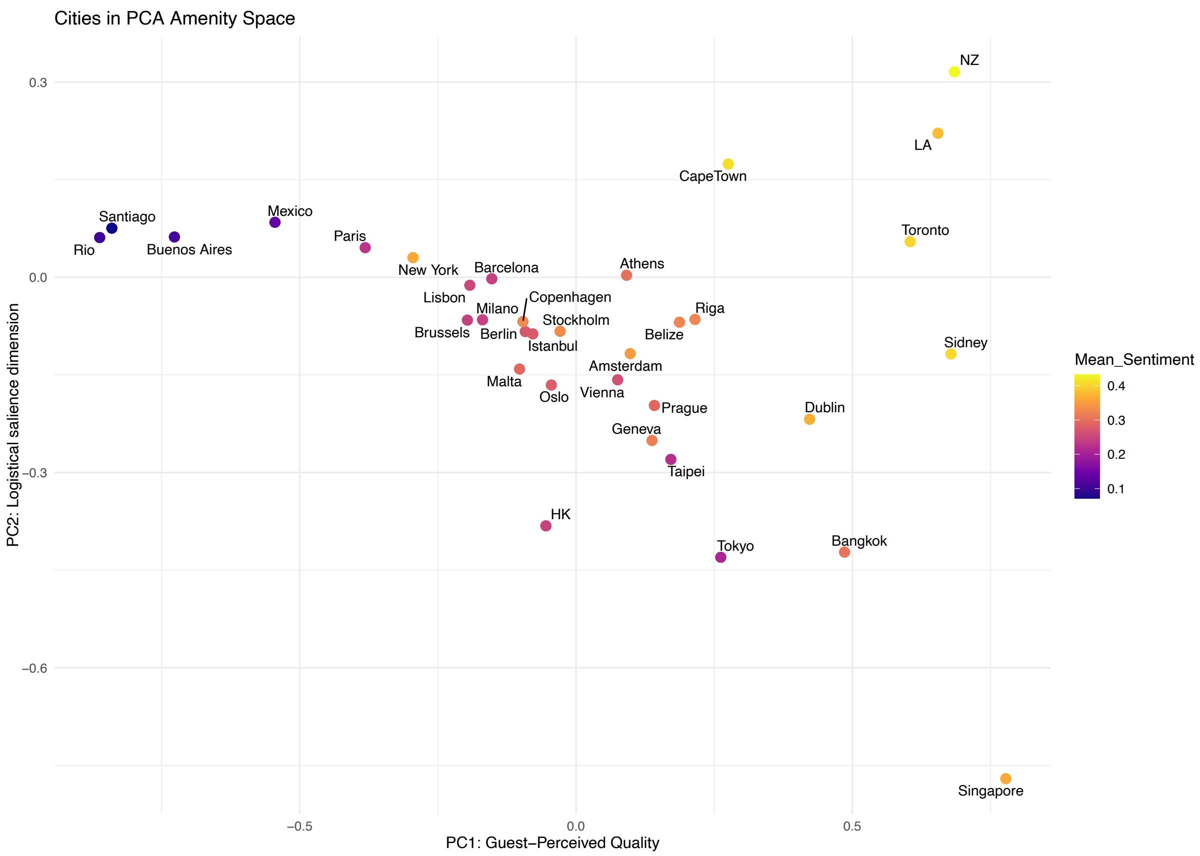

Figure 2: PCA mapping of cities in two dimensions.

From a management perspective, these two latent dimensions map directly onto the multi-scale framework introduced earlier. The amenity-richness dimension reflects elements that are partially under host control (cleanliness, presentation, service responsiveness) but also dependent on neighbourhood-level conditions such as perceived safety and leisure infrastructure. The logistical salience dimension, by contrast, is more strongly associated with city-level structures, including transport systems, pricing environments, and broader urban accessibility. Importantly, these findings extend the literature in two ways. First, while studies such as Ding et al. (2020) and Cheng & Jin (2019) extract thematic structures within specific destinations, our results show that comparable latent perception axes emerge across 34 cities worldwide. Second, the identification of distinct experiential and logistical dimensions reinforces the argument, advanced in Robertson et al. (2019), that urban tourism narratives are not purely accommodation-focused but embedded within broader spatial and infrastructural contexts.

5.2 Cluster Typologies of Listing Environments

To further examine structural heterogeneity in short-term rental markets, we conduct K-means clustering using a combination of spatial indicators, lifecycle variables, pricing structures, review intensity, and thematic frequency measures derived from reviews. The resulting five clusters, reported in Table 2, represent the typologies of listing environments that jointly reflect spatial positioning, market maturity, and guest-perceived experiential emphasis. Whereas the PCA identified latent perception axes derived exclusively from textual indicators, the clustering procedure integrates both objective listing attributes and subjective narrative signals, thereby translating perception structures into observable market configurations.

Table 2: K-means cluster analysis

Cluster	Entry Quantile	Dist. to Center (km)	Reviews (#)	Min. Nights	Transport	Safe/Clean	Tourism	Cost	Work	Avg. Sentiment	Price (USD)
1	9.81	0.30	2.6	8.28	0.00083	0.01817	0.00346	0.00015	0.00081	0.381	168.65
2	4.08	0.25	135.2	2.84	0.00424	0.00666	0.00477	0.00088	0.00059	0.296	157.46
3	9.32	0.74	53.9	2.44	0.00285	0.01605	0.01061	0.00151	0.00085	0.433	237.10
4	9.65	0.09	35.7	2.69	0.00236	0.00403	0.00235	0.00085	0.00044	0.185	158.77
5	7.63	0.10	40.9	28.99	0.00260	0.00812	0.00518	0.00114	0.00109	0.348	151.57

The first cluster is characterized by high entry quantiles, very low review counts, and relatively long minimum-stay requirements. These listings tend to be farther from central urban areas and place less emphasis on tourism and transport themes in guest narratives. Despite modest performance on experiential indicators, average sentiment remains moderately positive. This segment appears to represent peripheral and relatively low-demand listings operating at the margins of high-intensity tourism flows. Given the discussion in Section 2 on spatial differentiation in platform markets (e.g., Kourtit et al., 2022), this cluster may reflect areas with weaker Airbnb penetration or less concentrated tourism activity. The longer minimum-stay requirement suggests selective positioning or lower turnover.

The second cluster exhibits strong central positioning, the highest review counts, moderate pricing, and balanced thematic emphasis across transport, cleanliness, and tourism dimensions. These listings combine accessibility with stable performance and consistent guest evaluation. This typology closely aligns with the literature, which emphasizes the importance of cleanliness, location, and service fundamentals in shaping accommodation satisfaction (Cheng & Jin, 2019). The high review volume further suggests mature market integration and high turnover, consistent with platform-based reputation dynamics. From a management perspective, this segment represents the functional urban core of the short-term rental market, where host-level quality provision and city-level infrastructure align.

The third cluster stands out for its strong emphasis on tourism and cleanliness, the highest average sentiment score, and the highest average pricing. Although not always the most centrally located, these listings feature reviews that emphasize experiential richness rather than logistical constraints. This pattern resonates with the experiential and authenticity dimensions identified by Vinogradov et al. (2020), who emphasize the role of intangible, place-based attributes in shaping positive guest evaluations. The high sentiment scores despite relative spatial decentralization suggest that experiential salience can compensate for physical distance from the city centre. In terms of the conceptual framework developed earlier, this cluster illustrates how neighbourhood-level tourism appeal and host-level quality signals can jointly produce premium positioning within the platform economy.

The fourth cluster, by contrast, is characterized by strong central positioning but relatively weak thematic emphasis across tourism, transport, and cleanliness dimensions. It records the lowest average sentiment among all clusters. This configuration suggests a potential trade-off between spatial centrality and perceived experiential quality. While proximity to the urban core remains attractive, weaker amenity signals appear to limit overall evaluation. This pattern echoes the literature linking negative or muted sentiment expressions to service shortcomings and value concerns (Lawani et al., 2019). It also points out a central argument advanced in the introduction: spatial advantage alone does not guarantee high satisfaction outcomes; host-level management and experiential differentiation remain decisive.

The fifth cluster is characterized by very high minimum-stay requirements and moderate thematic emphasis across all dimensions. Review intensity is lower than in the central high-demand cluster, and it reflects longer occupancy cycles. Sentiment levels remain positive, and pricing is moderate. This typology corresponds more closely to medium- or long-term accommodation demand rather than short-term tourism flows. Its limited presence in most cities, with the notable exception of Los Angeles, suggests that regulatory frameworks and urban form may shape the viability of extended-stay segments. In line with Robertson et al. (2019), who emphasize how platform-based rentals reshape urban housing use patterns, this cluster highlights how the short-term rental market can intersect with longer-term residential demand dynamics.

The distribution of clusters across cities, presented in Figure 3, reveals pronounced cross-destination variation. New Zealand is strongly dominated by the tourism-oriented third cluster, indicating a decentralized yet experience-driven market structure in which peripheral listings can command high sentiment and premium pricing. Paris displays a bifurcated structure with substantial representation of both the high-demand central cluster and the budget-oriented inner core, which reflects economic and spatial heterogeneity within a mature tourism destination. Rio de Janeiro and Istanbul are dominated by the centrally located but lower-rated fourth cluster, suggesting that centrality in these cities does not necessarily coincide with strong experiential performance. Amsterdam exhibits a relatively balanced distribution across clusters, consistent with a regulated and diversified short-term rental environment. Los Angeles stands out for the strong presence of the extended-stay fifth cluster, reflecting the influence of metropolitan scale and urban sprawl on rental typologies. Riga, by contrast, displays a relatively dispersed distribution without clear dominance of a single cluster, indicating a smaller or less segmented market.

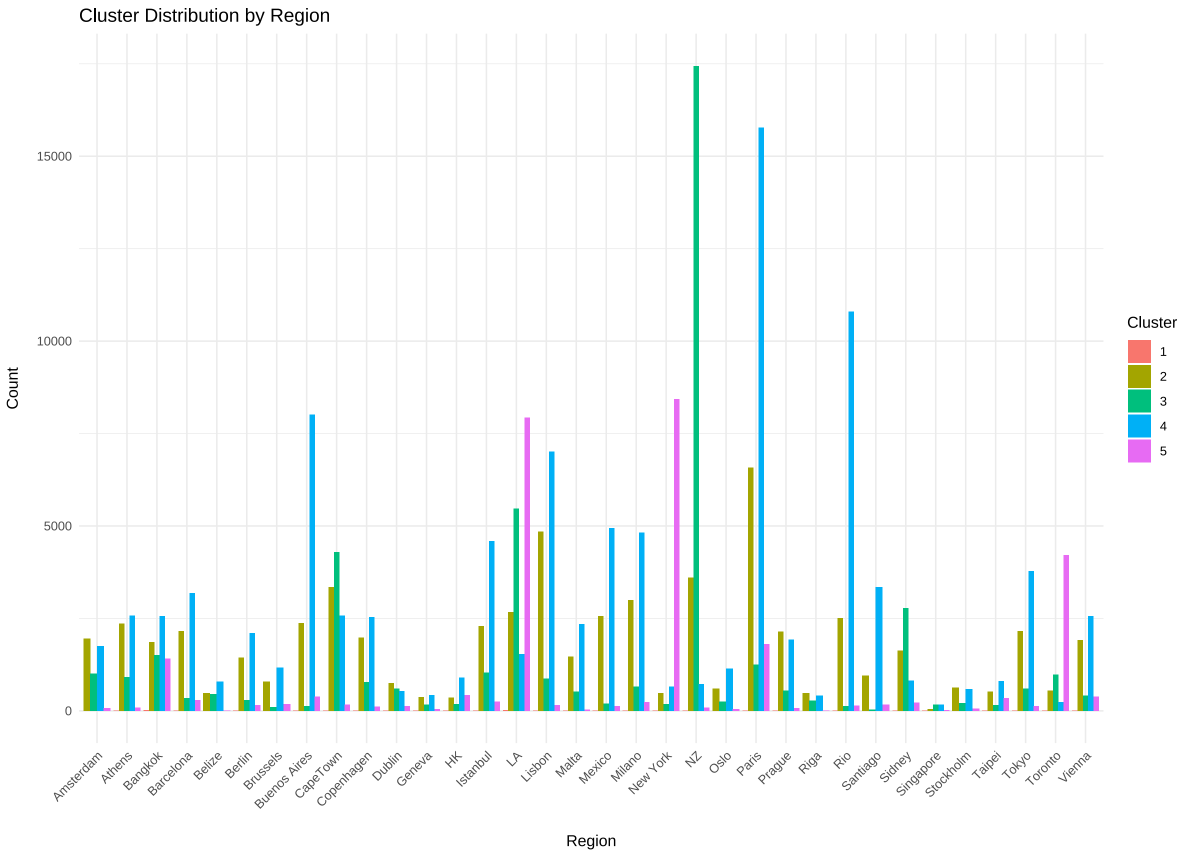

Figure 3: K means clustering patterns by city.

When cluster distributions are examined within the most central neighbourhoods of each city (Figure 4), further differentiation becomes visible. In Paris and Amsterdam, central areas exhibit a coexistence of high-performing and budget segments, reinforcing the internal stratification of mature tourism cores. In Rio de Janeiro and Istanbul, the budget-oriented fourth cluster remains dominant even in the urban core, suggesting the importance of host-level quality in shaping evaluation outcomes. Los Angeles continues to show substantial representation of the extended-stay segment in central zones, consistent with its structural demand characteristics.

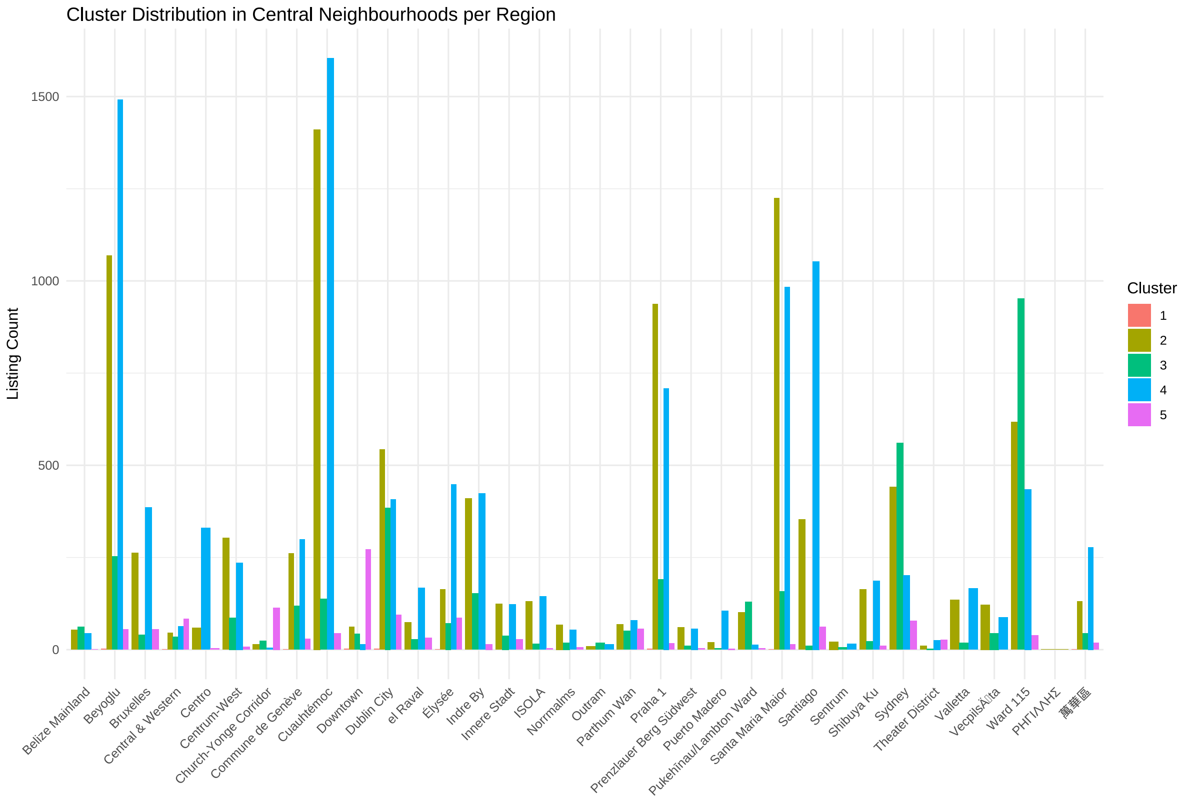

Figure 4: K-means clusters by the most central neighborhood in each of the 34 cities.

Taken together, the clustering results substantiate three central insights of this study. First, short-term rental markets are structurally segmented into recurring typologies that combine spatial positioning, market maturity, pricing structure, and experiential emphasis. Second, centrality interacts with, rather than substitutes for, host-level quality and experiential differentiation. Third, cross-city variation in cluster composition suggests that broader urban conditions and governance environments shape the configuration of platform-based accommodation markets. These findings provide the descriptive foundation for the subsequent multilevel modelling analysis, which formally estimates how listing-level characteristics and experiential dimensions jointly influence sentiment while accounting for neighbourhood and city-level heterogeneity.

5.3 Multilevel Determinants of Listing-Level Sentiment

To formally assess how listing characteristics, experiential emphasis, and contextual factors jointly shape visitor evaluations, we estimate a linear mixed-effects model with random neighbourhood-level intercepts nested within regions. This specification accounts for the hierarchical spatial structure of the data and allows for systematic contextual heterogeneity beyond listing-level attributes. The full set of results is reported in Table 3, while city-level random effects are illustrated in Figure 5.

Table 3. Linear Mixed Effects Model Predicting Listing-Level Average Sentiment

Predictor	Estimate	Std. Error	df	t value	p-value
Intercept	0.162	0.021	31.73	7.712	<0.001 ***
Entry quantile	-0.00214	0.000067	233,200	-31.92	<0.001 ***
Distance to center	-0.00015	0.00017	226,000	-0.905	0.366
Log(Number of reviews)	0.00305	0.00017	233,200	17.44	<0.001 ***
Log(Minimum nights)	0.01263	0.00035	224,200	35.75	<0.001 ***
Transport Mean	0.7651	0.04574	232,800	16.73	<0.001 ***
Safe & Clean Mean	3.064	0.02341	233,300	130.91	<0.001 ***
Tourism Mean	2.135	0.03403	232,900	62.74	<0.001 ***
Cost Mean	0.3878	0.04335	233,000	8.95	<0.001 ***
Work Mean	1.324	0.1126	232,600	11.76	<0.001 ***
Log(Price USD)	0.01178	0.00032	226,800	37.05	<0.001 ***

Random Effects**:**

Region (Intercept)**:** Var = 0.00685, SD = 0.08276

Neighbourhood:Region (Intercept)**:** Var = 0.00064, SD = 0.02537

Residual**:** Var = 0.01088, SD = 0.10432

Significance codes: *** $p < 0.001$

The results provide strong support for the central premise advanced in the introduction: visitor satisfaction in short-term rental markets is shaped by both listing-level management factors and broader experiential signals embedded in urban contexts.

Among all predictors, the review-derived indicators emerge as the strongest correlates of average sentiment. In particular, Safe & Clean Mean and Tourism Mean exhibit the largest associations in Table 3. These variables capture aggregated guest perceptions expressed in review narratives and correspond directly to the experiential-quality dimension identified in the PCA analysis. Their magnitude indicates that perceived safety, cleanliness, and tourism appeal constitute the core drivers of positive evaluation. This finding aligns closely with prior work emphasizing the importance of cleanliness and location-based attributes in accommodation satisfaction and the role of experiential and authenticity signals in shaping perceived value. Unlike many single-city studies, however, our model demonstrates that these relationships remain robust across 34 heterogeneous urban contexts.

Transport Mean and Work Mean also exhibit significant and positive associations with sentiment. These effects confirm that mobility accessibility and suitability for remote or extended stays contribute meaningfully to guest satisfaction. Considering recent discussions on flexible travel and digital nomadism, these findings suggest that Airbnb listings are evaluated not only as tourism units but as embedded functional spaces within urban systems. This interpretation reinforces the spatial-embeddedness argument discussed in the literature review and supports the view that short-term rentals operate within broader urban ecosystems rather than in isolation. The variable for the frequency of cost-related reviews is positively associated with sentiment, though with a smaller coefficient relative to the core experiential indicators. Importantly, these variable captures subjective value perception expressed in narratives rather than objective pricing. Its positive sign suggests that perceived value-for-money enhances satisfaction even when nightly rates vary. Interestingly, log-transformed price itself also shows a positive and significant effect. This indicates that higher-priced listings are associated with higher sentiment, possibly reflecting tiered service quality within platform markets.

Turning to listing lifecycle variables, Entry Quantile exhibits a small but statistically significant negative effect on sentiment. Later entrants to the platform tend to receive slightly lower average sentiment scores. This finding may reflect reputational accumulation dynamics and learning effects, consistent with platform-based mechanisms of market maturity documented in the literature (e.g., Kourtit et al., 2022). In contrast, Number of Reviews is positively associated with sentiment, again indicating the role of accumulated reputational capital in shaping guest evaluations.

The variable Minimum Nights is also positively associated with sentiment. Listings requiring longer stays appear to generate more positive evaluations, possibly reflecting more stable guest-host interactions or more deliberate booking decisions. This finding aligns with the cluster results, which show that the extended-stay segment displayed moderate-to-high sentiment despite reduced turnover.

Notably, Distance to Center is not statistically significant after controlling for experiential and listing-level factors. This result challenges simplistic centre–periphery assumptions and reinforces the cluster analysis finding that spatial proximity alone does not guarantee higher satisfaction. When safety, cleanliness, tourism appeal, and accessibility are accounted for, physical centrality does not independently predict sentiment. This supports the argument advanced in Section 5.2 that experiential quality can compensate for spatial distance.

The variance components reported in Table 2 justify the multilevel specification. The region-level random intercept (SD = 0.0828) and neighbourhood-level intercept (SD = 0.0254) indicate meaningful contextual variation beyond listing-level predictors. Although residual variance remains larger, the hierarchical structure captures systematic cross-context heterogeneity that would otherwise remain unobserved.

Figure 5 displays the estimated region-level random effects. These effects represent deviations from the global intercept after accounting for listing characteristics and experiential emphasis. Cities with positive random effects (e.g., New Zealand cities, Cape Town) exhibit systematically higher sentiment than predicted by observable listing attributes alone. Conversely, cities such as Santiago and Mexico display negative deviations. These residual differences suggest that broader contextual factors, potentially including destination branding, urban governance quality, infrastructure coordination, or macro-level perceptions, shape visitor evaluations beyond individual listing management.

In line with the conceptual framework, these city-level deviations can be interpreted as implicit indicators of destination-level performance in managing tourism environments. While not causal, the random effects provide a comparative benchmark for understanding how urban contexts condition guest satisfaction, even after controlling for listing-specific features.

The multilevel results highlight three central patterns in the data. First, the experiential dimensions extracted from review narratives are strongly associated with listing-level sentiment. Listings whose narratives place greater emphasis on safety, cleanliness, tourism appeal, and accessibility systematically exhibit higher overall sentiment scores. Given that both sentiment and thematic frequencies are derived from the same review corpus, this finding should be interpreted as evidence of structured co-variation within guest narratives rather than as a strictly causal mechanism. In other words, positive overall evaluations tend to be accompanied by recurring references to specific experiential qualities, suggesting that these themes constitute the core language through which satisfaction is expressed. Second, platform-related characteristics such as reputational accumulation and lifecycle position remain statistically significant even after accounting for experiential emphasis. Listings with a larger volume of reviews and longer minimum stay requirements are associated with higher sentiment, while later platform entrants exhibit slightly lower evaluations. These patterns are consistent with platform maturity dynamics and reputational signalling mechanisms documented in the literature, indicating that guest evaluations reflect not only perceived experience but also the structural positioning of listings within the platform ecosystem. Third, significant contextual heterogeneity persists across regions and neighbourhoods. Even after controlling for listing characteristics and experiential emphasis, baseline sentiment levels vary systematically across cities. This residual variation shows the importance of contextual factors that operate beyond individual listing management. Visitor satisfaction thus appears to emerge from the interaction of experience, platform structure, and urban context rather than from isolated listing attributes alone.

This modelling step completes the empirical progression from perception structure (PCA) to typological segmentation (clustering) and finally to inferential estimation of satisfaction determinants. It reinforces the argument that short-term rental markets must be understood as layered systems in which host-level practices, neighbourhood conditions, and city-level contexts jointly influence visitor experience.

5.1 Interpretation of findings

6. Conclusions: retrospect and prospect

References

Água, M., António, N., Carrasco, P. & Rassal, C. (2025). Large Language Models Powered Aspect-Based Sentiment Analysis for Enhanced Customer Insights*, Tourism & Management Studies*, 21(1), 2025, 1-19. https://doi.org/10.18089/tms.20250101.

Buhalis, D. & Law, R. (2008), Progress in Information Technology and Tourism Management, *Tourism Management,* 29(4), 609-623.

Cheng, N.T.Y., Fong, L.H.N., & Law, R. (2021), Mobile Payment Technology in Hospitality and Tourism, *International Journal of Contemporary Hospitality Management*, 33(10), 3636-3660.

Cheng, M. & Jin, X. (2019). What do Airbnb users care about? An analysis of online review comments,

International Journal of Hospitality Management, 76, Part A, 58-70, https://doi.org/10.1016/j.ijhm.2018.04.004.

Ding, K., Choo, W.C., Ng, K.Y. & Ng, S.I. (2020). Employing structural topic modelling to explore perceived service quality attributes in Airbnb accommodation, *International Journal of Hospitality Management*, 91, 102676, https://doi.org/10.1016/j.ijhm.2020.102676.

Gretzel, U. (2018). Tourism and Social Media. In: *The Sage Handbook of Tourism Management* (C. Cooper, S. Volo, W.C. Gartner, N.Scott eds), SAGE Publications Ltd, pp.415-432.

Jim, J.R., Talukder, A.R., Malakar, P., Kabir, M., Nur, K. & Mridha, M.F. (2024). Recent advancements and challenges of NLP-based sentiment analysis: A state-of-the-art review, *Natural Language Processing Journal*, 6, 100059, https://doi.org/10.1016/j.nlp.2024.100059.

Kim, M.J., Hall, C.M. & Chung, N. The influence of AI and smart apps on tourist public transport use: applying mixed methods. *Inf Technol Tourism* **26**, 1–24 (2024). https://doi.org/10.1007/s40558-023-00272-x.

Kourtit, K., Nijkamp, P., Östh, J. & Turk, U. (). Airbnb and COVID-19: SPACE-TIME vulnerability effects in six world-cities, *Tourism Management*, 93, 104569, https://doi.org/10.1016/j.tourman.2022.104569.

Labanauskaite, D., Fiore, M., & Stasys, R. (2020), Use of e-Marketing tools as Communication Management in the Tourism Industry, *Tourism Management Perspectives,* 34, 100652.

Lawani, A., Reed, M.R., Mark, T. & Zheng, Y.(2019). Reviews and price on online platforms: Evidence from sentiment analysis of Airbnb reviews in Boston, *Regional Science and Urban Economics*, 75, 22-34, https://doi.org/10.1016/j.regsciurbeco.2018.11.003.

Li, Y., Hu, C., Huang, C., & Duan, L.(2017), The Concept of Smart Tourism in the Context of Tourism Information Services, *Tourism Management,*58, 293-300.

Lim, J., Kourtit, K., & Nijkamp, P. (2025), …….IJTR

Liu, A., & Sun, M. (2025). From Voices to Validity: Leveraging Large Language Models (LLMs) for Textual Analysis of Policy Stakeholder Interviews. *AERA Open*, *11*. https://doi.org/10.1177/23328584251374595.

Joo, D., Woosnam, K.M., Schafer, C.S., Scott,D., &An, S. (2017), Considering Tobler's First Law of Geography in a Tourism Context, *Tourism Management, 62, 350-359.*

Navio-Marco, J., Ruiz-Gomez, L.M., & Sevilla-Sevilla, C. (2018), Progress in Information Technolgy and Tourism Management – 30 Years on and 30 Years after the Internet-revisiting Buhalis and Law's Landmark Study about eTourism, *Tourism Management*, 69, 460-470.

Núñez, J.C.S., Gómez-Pulido, J.A. & Ramírez, R.R. (2024). Machine learning applied to tourism: A systematic review, *WIREs Data Mining and Knowledge Discovery*, 14 (5), e1549, https://doi.org/10.1002/widm.1549.

Pantano, E., Serravalle, F., & Priporas, C. V. (2024). The form of AI-driven luxury: how generative AI (GAI) and Large Language Models (LLMs) are transforming the creative process. *Journal of Marketing Management*, *40*(17–18), 1771–1790. https://doi.org/10.1080/0267257X.2024.2436096.

Pozdniakov, S., Brazil, J., Abdi, S., Bakharia, A., Sadiq, S., Gašević, D., Denny, P. & Khosravi,H. (2024). Large language models meet user interfaces: The case of provisioning feedback, *Computers and Education: Artificial Intelligence*, 7, 100289, https://doi.org/10.1016/j.caeai.2024.100289.

Robertson J, Beyer S, Emerson E, Baines S, Hatton C. (2019). The association between employment and the health of people with intellectual disabilities: A systematic review. J Appl Res Intellect Disabil. 2019 Nov;32(6):1335-1348. doi: 10.1111/jar.12632. Epub 2019 Jun 10. PMID: 31180175.

Romao, J. (2024), *Economic Geography of Tourism*, Springer, Berlin.

Subbamma, T.V., Mantravadi, N., Shalom, A.R., Tajne, N., Sreenivasulu, G. & Goud, V.S. (2025). Natural Language Processing for Sentiment Analysis Techniques and

Applications, *Metallurgical and Materials Engineering*, 31(3), pp. 194–200. doi: 10.63278/1356.

Suzuki, S., Kourtit, K., & Nijkamp, P. (eds.) (2021), *Tourism and Regional Science,* Springer, Singapore.

Tussyadiah, I. P., & Pesonen, J. (2015). Impacts of Peer-to-Peer Accommodation Use on Travel Patterns. *Journal of Travel Research*, *55*(8), 1022-1040. https://doi.org/10.1177/0047287515608505.

Vinogradov, E., Leick, B. & Kivedal, B.K. (2020). An agent-based modelling approach to housing market regulations and Airbnb-induced tourism, *Tourism Management*, 77, 104004, https://doi.org/10.1016/j.tourman.2019.104004.

Zhang, W., Deng, Y., , Liu, B., Pan, S.J. & Bing, L. (2024). Sentiment Analysis in the Era of Large Language Models: A Reality Check, Findings of the Association for Computational Linguistics: NAACL 2024, pages 3881–3906, https://www.cse.cuhk.edu.hk/~sinnopan/publications/[NAACL24]Sentiment%20Analysis%20in%20the%20Era%20of%20Large%20Language%20Models%20A%20Reality%20Check.pdf.