# Decoding Multidimensional User Experiences and Evaluations in Urban Public Spaces: A Novel Integrated NLP Model Based on Multimodal Online Reviews

Zian Wang [*, 1], Yifan Yang [2], Peter van Oosterom [1], Steffen Nijhuis [3], Stefan van der Spek [3]

[1] Department of Architectural Engineering and Technology, Faculty of Architecture and the Built Environment, Delft University of Technology, Delft, the Netherlands - (Z.W.Wang, P.J.M.vanOosterom)@tudelft.nl
[2] Institute for Risk Assessment Sciences, Faculty of Medical Sciences and Faculty of Veterinary Medicine, Utrecht University, Utrecht, the Netherlands - yifanyang21@outlook.com
[3] Department of Urbanism, Faculty of Architecture and the Built Environment, Delft University of Technology, Delft, the Netherlands - (S.Nijhuis, S.C.vanderSpek)@tudelft.nl

**Keywords:** Urban public space, Natural language processing, Transformer, Vision-language model, Multimodal, Online review.

## Abstract

Understanding users' experiences and evaluations in urban public spaces is essential for human-centric design and management. While traditional survey methods are costly, large-scale user-generated content (UGC) offers new possibilities for capturing public insights. However, existing approaches mainly assess evaluations and experiences through oversimplified, single-dimensional assessments, while deriving users' multidimensional experiences beyond basic sentiments and capturing their underlying mechanisms remain critical research gaps. This study proposes a novel integrative analytical model for assessing urban public space experiences and evaluations. Drawing on large-scale online review data, it employs a Low-Rank Adaptation (LoRA)-fine-tuned RoBERTa language model for multidimensional user evaluation identification, combines a lexicon-based method with an Aspect-based Sentiment Analysis (ABSA) pipeline to examine specific experiential qualities and satisfaction, and integrates textual and visual modalities through Bootstrapping Language-Image Pre-training (BLIP) vision-language model. Focusing on Amsterdam as a case study, validation demonstrates that the proposed model outperforms baselines, achieving a mean accuracy of 85% and exceeding widely used XGBoost and LSTM methods by 12% and 23%, respectively. The application effectively quantifies nuanced and detailed user experience and evaluation patterns, revealing differentiated associations between experiential qualities and overall evaluation positivity, thereby providing a deeper understanding of human-environment interactions in urban public spaces. The proposed approach offers an integrated and scalable method for investigating user insights and can inform future responsive, human-centric public space design and management.

## 1. Introduction

Urban public spaces, referring to urban spaces and places that are open and accessible to all and including squares, parks, green spaces, among other publicly available environments, play a vital role in the urban environment and residents' everyday life. Public spaces not only enhance the urban environment but also serve as key avenues for people's socialization, physical activity, and leisure, contributing to mental well-being, physical health, and overall quality of life (Gehl, 1987). With global urbanization and continuous urban population growth, the public has evolving requirements and higher expectations for urban environments, and public space is increasingly recognized as a critical strategy for addressing urban challenges, promoting health and social inclusion, and improving city livability. In this context, effective assessment of user evaluations and experiences is essential for evaluating urban public space performance and effectiveness in aligning with people's needs and preferences, and for informing future evidence-based, human-centric approaches to their management and design amidst continuous urban development.

Assessment of public space experience and evaluation has traditionally relied on questionnaires and observation methods, which are valuable but may face cost and scalability challenges. Recently, with the advent of big data, citizens increasingly act as "sensors" in urban environments, and user-generated content (UGC) has emerged as a promising alternative data source for understanding public views and opinions (Wang et al., 2025). User-uploaded images and texts on online platforms capture rich first-person accounts of urban experiences, offering broad spatial coverage, ease of collection, and fine-grained resolution (Zhang et al., 2026). Though facing challenges related to unstructured data format and representativeness bias, often skewed toward younger, more educated populations, such data still offer valuable individual perspectives at scale. Combined with advances in natural language processing (NLP) and related technologies, such data present new opportunities for efficiently capturing user insights and understanding how users perceive and experience urban spaces.

However, existing big data-driven urban studies are more often focused on assessing objective urban public space qualities, such as visual quality evaluation through street-view imagery analysis, or urban vitality assessment using POI datasets, and have largely operated on the macro urban or regional scale. By comparison, the subjective assessment of individual experiential quality using semantically-rich UGC data remains relatively limited, particularly at the fine-grained scale of urban public spaces, where the complexity and heterogeneity of user perceptions and experiences are rarely examined (J. Li et al., 2024; Zhao et al., 2025). This may be attributed to two methodological challenges:

(1) First, there is a lack of effective and dedicated models and datasets tailored to the context of urban public space analysis. Commonly adopted methods for extracting human insights from UGC, such as topic modeling and sentiment analysis, were not developed for urban analytical tasks, and their accuracy in this context has not been thoroughly validated, risking

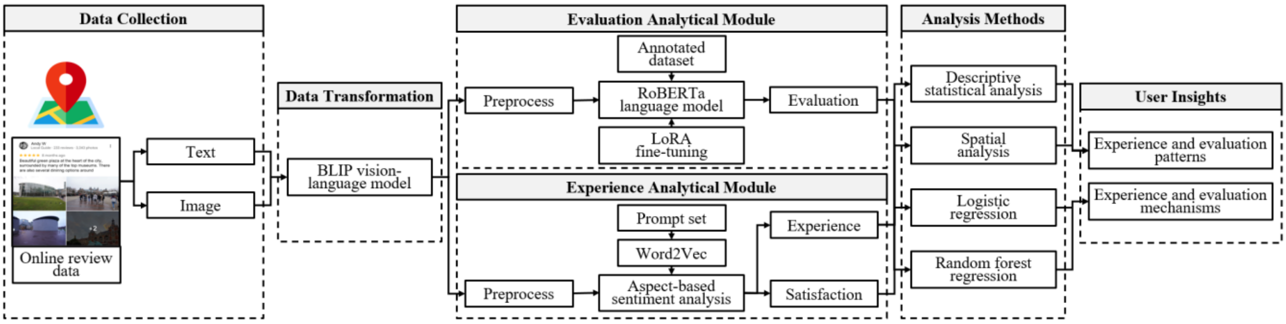


Figure 1. Research framework

misinterpretation of context, overlooking semantic relations, and remaining susceptible to data noise (Zhao et al., 2025). For instance, users may express different levels of satisfaction regarding different aspects of a public space, but standard sentiment analysis may fail to capture such nuanced variation. Also, unlike other big data types, UGC data in urban analytical contexts are often multimodal, encompassing both text and image materials. Existing approaches predominantly rely on a single modality, particularly text, overlooking the rich insights embedded in images, particularly for complementing experiential dimensions that text alone inadequately captures, and risking introducing systematic biases (Zhao et al., 2025). There is a need for more advanced, purpose-built analytical techniques to support accurate and detailed extraction of multidimensional user insights from public space UGC data.

(2) Second, there is a lack of systematic and integrated analytical frameworks for examining user experiences and evaluations. While environmental psychology research has underscored the multiple dimensions and interactions involved in human responses to environments, encompassing experiential and perceptual processes, existing studies have largely focused on individual and isolated aspects of human responses, often including only a single metric, such as overall sentiment (S. Li et al., 2024). While these studies have expanded our knowledge in meaningful ways, their narrow focus – reducing experience to single measures – risks oversimplifying complex human-environment interactions and may only capture partial representations of human responses instead of the whole breadth and richness of actual user experiences. There is a need for a more integrated analytical framework capable of collectively extracting and analyzing multidimensional user experiences and evaluations from UGC data. Such a framework would enable a more comprehensive understanding of different user experiential and evaluative aspects as well as their underlying interrelations and mechanisms, thereby supporting a more holistic and user-centric assessment of public space experiential quality and more in-depth practical insights.

Addressing these challenges, this study proposes a novel, integrated analytical model for assessing urban public space user insights. We combine multimodal online review data with state-of-the-art NLP techniques, including Low-Rank Adaptation (LoRA)-fine-tuned RoBERTa language model, Bootstrapping Language-Image Pre-training (BLIP) vision-language model, and Aspect-based Sentiment Analysis (ABSA) method, to enable deeper and more comprehensive extraction of user experiences and evaluations from UGC data. Our findings demonstrate that the proposed model outperforms existing baseline methods across all evaluation metrics and offers key advantages: it not only enables more targeted analysis of complex and differentiated personal insights in urban public space contexts, incorporating richer semantic and visual information, but also supports the concurrent analysis of user evaluations, experiences, and their complex interrelationships. Its application may enable more thorough extraction of information embedded in UGC, thereby supporting a comprehensive understanding of both patterns and mechanisms of human-environment interaction in urban public spaces. Also, by leveraging widely available review data, transferable models, and labeled datasets, the model demonstrates scalability and can be tested across contexts. This study thereby offers a novel methodological approach for investigating user responses to urban public spaces, with practical implications for more accurate assessment and evidence-based, human-centric practices for future urban public space management and design.

The remaining sections of this paper are organized as follows: Section 2 details the case study site and data collection, Section 3 explains the methodological approach, Section 4 details the analytical processes, Section 5 presents the results and discussion of the study, and Section 6 provides the summary.

## 2. Study City and Data Collection

The research framework is presented in Figure 1. This study uses Google Maps reviews as the data source. Existing UGC-driven research mainly relied on social media data, but such platforms may face well-documented limitations, including noise, inconsistent quality, and insufficient representation of the population. Online review platforms, in comparison, offer several advantages: (1) review platforms enable users to post about specific places (i.e., points of interest, POI), thus providing more spatially explicit insights; (2) reviews have been found to concentrate on personal experiences (e.g., recreation) and emotions, making them better suited for urban experience research; and (3) review platforms may demonstrate broader demographic coverage and representativeness compared to social media, as a survey found that 69% of respondents had written a review in the past year (Paget, 2025), and multiple studies have validated that self-selected samples on review platforms can yield generally representative assessments (Greaves et al., 2012). Google Maps is among the most widely used review platforms, covering 83% of the market share, with its user base continuing to grow (Paget, 2025). It is recognized as a valuable source for capturing public perceptions and supporting urban analysis.

This study focuses on Amsterdam, the largest and most populous city in the Netherlands, with approximately 942,500 residents. It features a diverse urban fabric encompassing various types of public spaces. Amsterdam has also adopted coherent strategies for liveable and sustainable public space development, making it a suitable study context.

Data collection was conducted in November 2025. Based on the Amsterdam municipality's "Maps Amsterdam" database, we extracted land parcel types related to urban public space, including green, nature, recreation, sport, garden, and other categories. Querying the Google Places API yielded 154 officially recognised public spaces, each identified by a unique Place_ID.

Subsequently, a Python script was developed to collect Google Maps reviews for the identified public space sites, covering the period from 2021 to 2025. Each review record contained the following fields: an anonymized user ID, review text (where available), review images (where available), star rating, and upload date (see Table 1 for an example). No personal information was collected. As Google shows review timestamps in relative format, exact dates were estimated through back-calculation.

| | |
|---|---|
| ID | EihN… QBm |
| Rating | 4/5 |
| Text | Green plaza at the heart of the city, surrounded by many top museums. The area is clean, safe and easily accessible. |
| Image(s) |  |
| Timestamp | 2024-06-17 |

Table 1. Example of collected review (Content was transcribed for privacy protection)

As our study involved the extraction of detailed experiences, reviews consisting only of ratings were excluded from further analysis. Sites with fewer than 10 reviews were excluded to ensure robustness, and sensitivity analyses were conducted across alternative thresholds. This filtering resulted in the inclusion of 91 public space sites. Finally, following the common approach adopted in the literature in online review analytics, all texts were translated into English for subsequent analysis through the Google Translate API. We acknowledge that this may result in information loss, but existing evidence has validated the effectiveness of translated text for capturing experiential and evaluative insights in urban environments (Huai and Van de Voorde, 2022). The final dataset encompasses 29,257 reviews.

## 3. An Integrated Multimodal Analytical Model for Public Space Experience and Evaluation Assessment

We focus on two core user dimensions in public space: *experience* and *evaluation*. Experience refers to subjective views and affective responses to public space characteristics, encompassing 12 dimensions identified from existing literature (J. Li et al., 2024). Evaluation refers to overall perceptions of public space, covering three key aspects: *usability*, *accessibility*, and *attractiveness*. All dimensions and definitions are detailed in Table 2.

| | |
|---|---|
| Evaluation dimensions | |
| Usability | The extent to which a space supports effective, comfortable, and purposeful use. |
| Accessibility | The ease with which users can reach and enter a space. |
| Attractiveness | The perceived appeal and quality of a place that motivates people to visit or stay. |
| Experience dimensions | |
| Aesthetics | Visual quality, appeal, and pleasantness of a place. |
| Amenity and service | Availability and adequacy of facilities, services, and functional support for activities. |
| Recreation and leisure | Provision of opportunities for relaxation, play, and leisure engagement. |
| Socialization | Opportunities and settings that support interaction and social activities. |
| Physical activity | Availability of space and conditions that support exercise. |
| Greenery | Presence and quality of green elements such as trees, plants, and landscape. |
| Ecology and biodiversity | Environmental qualities such as biodiversity and environmental health. |
| Microclimate | Thermal comfort, shading, and weather-related conditions. |
| Connectivity | Connection to other environments and within the space. |
| Crowdedness | Comfort level related to crowding. |
| Maintenance | Cleanliness and maintenance condition of the space. |
| Safety | Perceived security and protection from crime or hazards. |

Table 2. Evaluation and experience dimensions included and their definitions

### 3.1 Multimodal Data Integration based on BLIP Vision-Language Model

Prior studies have shown that text and image data each capture unique aspects of user insights (Zhao et al., 2025). While commonly used text data provide rich semantics, image data may also capture contextual and environmental information not directly evident in text descriptions. Therefore, to fully extract user insights from collected multimodal review data, a data transformation and fusion approach based on a vision-language model was developed to integrate text and image reviews. The model applies the BLIP-Base model for image captioning (Li et al., 2022), and employs the Contrastive Language-Image Pre-training (CLIP) model with a ViT-B/32 backbone for validation, converting unstructured image data into quantifiable feature dimensions (Radford et al., 2021).

CLIP is a contrastive vision-language model trained on large-scale image-text pairs, which maps image and text encodings into a shared semantic space to enable similarity-based analysis between visual and semantic concepts (Radford et al., 2021). BLIP extends this by incorporating a Multimodal Mixture of Encoder-Decoder (MED) architecture and a Caption and Filtering (CapFilt) mechanism, enabling high-quality visual description generation from unstructured data through fine-tuned captioning and filtering modules (Li et al., 2022). Compared to traditional computer vision models, vision-language models do not rely on predefined, closed-category systems and support zero-shot classification, making them well-suited for complex and subjective experiential and evaluative recognition tasks.

In the data fusion pipeline, BLIP was applied to each image to generate natural language captions capturing the overall environmental context depicted. To ensure descriptive richness, a heuristic mechanism was introduced, varying beam search width and nucleus sampling parameters to obtain more semantically informative outputs. The generated captions were then concatenated with the corresponding review texts to support downstream experience and evaluation recognition. For validation, CLIP was additionally employed for zero-shot recognition using the aspect prompt set. Image features were compared against prompt descriptions via cosine similarity to determine whether each image conveyed relevant experiential dimensions. Images yielding inconsistent results between the two models were manually reviewed.

### 3.2 User Evaluation Identification based on LoRA-Fine-Tuned RoBERTa Language Model

The user evaluation classification module was developed based on the RoBERTa language model (Liu et al., 2019), a Transformer-based model whose deep bidirectional encoder architecture enables learning of semantic representations directly from data, allowing it to capture contextual meaning, syntactic structure, and nuanced term associations more effectively than earlier deep learning approaches. To develop task-specific models tailored to the analytical needs of urban evaluation tasks, we employed LoRA (Hu et al., 2021) as a parameter-efficient fine-tuning strategy, which introduces low-rank decomposition modules into the self-attention weight matrices and trains a small subset of parameters, enabling efficient training and transfer.

We randomly sampled 2,500 reviews from the collected data to construct a manually annotated dataset for task-specific model fine-tuning. Each review was independently labeled regarding the usability, accessibility, and attractiveness dimensions as negative (0), mixed (1), positive (2), or irrelevant (3), forming a multi-class label structure. The sample size and labeling classes align with the approach adopted in past studies in this field (Zhao et al., 2025). A multi-label setting was adopted, where each dimension was judged independently. Annotation was carried out by two researchers with backgrounds in urban studies and environmental psychology. Cohen's Kappa between raters was 0.736 (usability), 0.665 (accessibility), and 0.771 (attractiveness). Disagreements were resolved through discussion until a consensus was reached.

Separate models were trained for each dimension to address potential semantic differences across evaluation dimensions. The annotated dataset was randomly split into training, validation, and test sets in a 7:2:1 ratio. All texts were preprocessed through standard normalization prior to model input. Model training employed cross-entropy loss and the AdamW optimizer (Loshchilov and Hutter, 2019) with a learning rate of 4e-5. A linear warm-up with linear decay was applied to improve training stability. For fine-tuning, low-rank decomposition updates were applied to the Query and Value weight matrices in RoBERTa's self-attention layers, while all other pre-trained parameters remained frozen. A dropout layer (rate = 0.1) was added to the classifier head to mitigate overfitting and improve generalization. Training proceeded for up to 20 epochs, with early stopping triggered if validation loss showed no improvement over three consecutive epochs. Training was conducted on consumer hardware (i9-13900HX, RTX 4060 Laptop). Model performance was subsequently evaluated on the held-out test set.

### 3.3 User Experience and Sentiment Analysis based on ABSA Pipeline

Multidimensional experiential content and complex sentiments in UGC are often difficult to adequately capture through conventional deep learning approaches. This study employs the ABSA method to extract users' separate and targeted responses to different spatial qualities from unstructured data (Yang et al., 2023). Unlike document-level sentiment analysis, ABSA deconstructs text into specific "aspects" and extracts the sentiment polarity associated with each, thus enabling precise decoupling of multiple experiential targets with potentially divergent emotional orientations within a single review.

To accurately identify the 12 experiential quality dimensions of public space relevant to this study, we incorporated a lexicon-based approach and constructed a domain-specific aspect term lexicon. High-frequency word analysis was applied to all collected reviews for manual screening of experience aspect-related expressions, forming an initial seed prompt set. A Word2Vec model (Mikolov et al., 2013) was then trained on the corpus to retrieve semantically similar candidate terms via cosine similarity in the embedding space; these candidates were manually reviewed and filtered to yield the final aspect prompt set, covering a total of 375 terms most relevant to the 12 experiential quality dimensions.

During inference, review texts were lemmatized and processed with SpaCy's PhraseMatcher module (Honnibal et al., 2020) to locate aspect terms corresponding to the 12 dimensions. For each detected aspect mention, sentiment polarity (positive, negative, or neutral) was classified using a pre-trained ABSA-APC (Aspect Polarity Classification) model (Yang et al., 2023), which applies positional masking to enable contextual awareness of the target concept within the surrounding text. To address the syntactic complexity characteristic of UGC language, reviews were pre-segmented into short clauses to ensure a single unambiguous evaluative focus per analytical unit.

## 4. Analytical Approach

To validate model performance, we evaluated our proposed model through accuracy, precision, recall, F1 score, and test loss. We conducted a comparative analysis incorporating a series of commonly used methods: BERT (frozen) model (Devlin et al., 2019), Long Short-Term Memory (LSTM), and eXtreme Gradient Boosting (XGBoost). BERT (frozen) serves as a transformer-based baseline where the pre-trained parameters are kept static to evaluate the raw feature extraction capability without task-specific fine-tuning. An LSTM network is included as a representative recurrent neural network (RNN) architecture, specifically designed to capture long-range sequential dependencies within review text. XGBoost (TF-IDF) represents a traditional machine learning approach, utilizing Term Frequency-Inverse Document Frequency vectorization to transform text into statistical features for gradient-boosted decision tree classification. The proposed model and baselines were validated on the same annotated dataset. Test loss was not applicable to the XGBoost method as it does not compute a loss function on the held-out test set within the training pipeline.

Applying the model, we extracted multidimensional user insights across the collected Amsterdam public space sites. First, descriptive statistics and spatial analysis were employed to analyze overall patterns of user experiences and evaluations.

Second, leveraging the joint extraction of experiential and evaluative dimensions, logistic regression was employed to examine the associations and mechanisms underlying users' multidimensional quality experiences and the three overall evaluation categories. Logistic regression models the probability of a binary outcome by mapping a linear combination of predictors into the [0, 1] interval through a sigmoid function. For each overall evaluation dimension, we constructed a separate model at the individual review level. For experience dimensions, we encoded user mentions into two dummy variables

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XII-4-W2-2026-219-2026  

representing positive and negative sentiment, respectively, while non-mentions served as the reference category. Model coefficients were interpreted using odds ratios (OR), which quantify how the presence and sentiment of each experiential dimension affect the likelihood of a positive overall evaluation relative to the baseline (non-mention). Models were checked for multicollinearity (VIF<3).

Finally, to address observed spatial variation in evaluation mechanisms, stratified regression models were estimated for public spaces across different geographic locations within Amsterdam. We additionally conducted lexical salience-valence analysis (LSVA) to further examine the rich semantic information embedded in the data, as a validation step. LSVA identifies key content within text by defining *Salience* and *Valence*, and assesses the relationship between individual terms and the sentiment expressed in the reviews:

$$Salience(w_i) = \log_{10}(N(w_i)) \quad (1)$$

$$Valence(w_i) = \frac{N_p(w_i) - N_n(w_i)}{N(w_i)} \quad (2)$$

where $N(w_i)$ represents the total frequency of word $w_i$ across all analyzed reviews, $N_p(w_i)$ and $N_n(w_i)$ represent the number of times $w_i$ co-occurs with positive and negative sentiment labels, respectively.

## 5. Results and Discussion

### 5.1 Characteristics of Collected Online Review Data

Results validated the overall performance of the proposed multimodal analytical model (Figure 2), achieving a mean precision of 0.81, recall of 0.78, F1 score of 0.79, and accuracy of 0.85. It outperformed baseline methods across evaluation metrics, with accuracy gains of 11.5% over XGBoost, 23.3% over LSTM, and 37.1% over BERT model. Its relatively balanced precision and recall indicate robustness in capturing both explicit and implicit expressions, which may be attributed to the integration of fine-tuning and aspect-based analysis, better enabling context-sensitive and dimension-aware recognition. Its lower test loss confirms improved generalization and more stable convergence relative to the baselines.

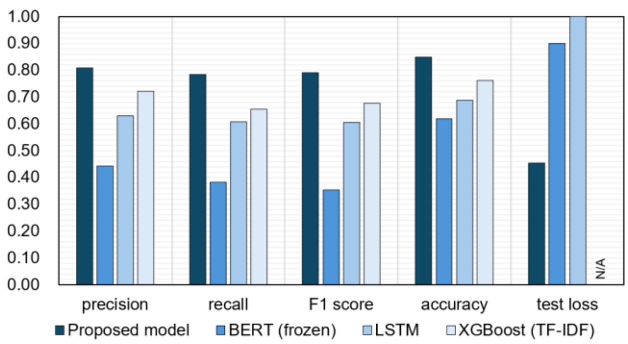


Figure 2. Model performance validation and comparison

### 5.2 Multidimensional User Evaluations and Experiential Insights

Application of the model allows quantification of users' differentiated overall evaluations across 91 public spaces in Amsterdam (Table 3). Usability received the most favorable user evaluation, with an average of 90% of users providing positive assessments. Attractiveness and accessibility were evaluated less favorably, with accessibility exhibiting a notably higher standard deviation, indicating greater divergence across public spaces. These results confirm that users' evaluations differ meaningfully across dimensions (J. Li et al., 2024), and reliance on a single analytical dimension risks obscuring important variation.

| | min | mean | max | std dev |
|---|---|---|---|---|
| Usability | 0.20 | 0.90 | 1.00 | 0.13 |
| Accessibility | 0.00 | 0.72 | 1.00 | 0.25 |
| Attractiveness | 0.25 | 0.76 | 1.00 | 0.13 |

Table 3. Statistics of user evaluations across dimensions

ABSA-based analysis revealed nuanced user insights across experiential dimensions (Figure 3), exposing mismatches between mention frequency (representing perceptual condition) and sentiment levels (representing emotional and evaluative response) across dimensions. Amenity- (33.0%), recreation- (20.0%), physical activity- (18.3%), and greenery-related experiences (17.6%) were the most frequently mentioned experience dimensions and also attracted relatively positive sentiments (0.862, 0.818, 0.836, and 0.796, respectively), consistent with prior evidence that recreation and activity represent primary modes of public space use (J. Li et al., 2024). Connectivity- (9.9%) and aesthetics-related experiences (4.2%) were less frequently mentioned but still associated with relatively positive mean sentiments (0.858 and 0.743, respectively). Conversely, crowdedness- (1.5%), maintenance- (2.0%), and safety-related experiences (0.9%) were rarely mentioned, but tended to elicit negative sentiments (0.483, 0.336, and 0.258, respectively), potentially reflecting deficiencies in these dimensions across studied public spaces, though UGC data characteristics may also partly account for this pattern. These results demonstrate that the adopted approach enables more granular quantification of user experiences, revealing latent variations that aggregated measures may inadequately capture.

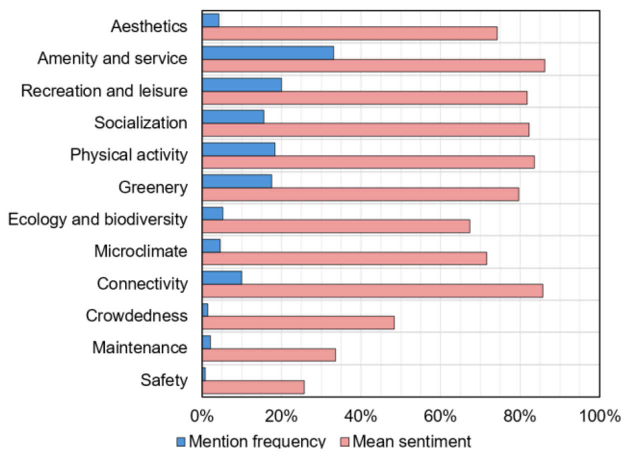


Figure 3. User experience mention frequency and mean sentiment across dimensions

### 5.3 Differentiated Association Mechanisms Between User Evaluations and Experiential Dimensions

The integrated model enables concurrent extraction of user experiences and evaluations, thus facilitating analysis of their interrelationships and supporting a more comprehensive understanding of driving factors behind human experiences in public space contexts. Logistic modeling results (Table 4) reveal highly differentiated associations between evaluation dimensions and specific experiential satisfaction, indicating complex underlying mechanisms.

| | Usability | | | Accessibility | | | Attractiveness | | |
|---|---|---|---|---|---|---|---|---|---|
| | coef | OR | p | coef | OR | p | coef | OR | p |
| const | -1.50** | 0.22** | 0.00 | -2.92** | 0.05** | 0.00 | 1.01** | 2.73** | 0.00 |
| Aesthetics (positive) | 0.47** | 1.60** | 0.00 | 0.38** | 1.46** | 0.00 | 0.53** | 1.70** | 0.00 |
| Aesthetics (negative) | -1.15* | 0.32* | 0.01 | 0.36 | 1.44 | 0.43 | -2.55** | 0.08** | 0.00 |
| Amenity and service (positive) | 0.97** | 2.64** | 0.00 | 0.28** | 1.32** | 0.00 | 0.43** | 1.54** | 0.00 |
| Amenity and service (negative) | -0.51** | 0.60** | 0.00 | -0.41* | 0.66* | 0.04 | -2.10** | 0.12** | 0.00 |
| Recreation and leisure (positive) | 0.83** | 2.29** | 0.00 | 0.10* | 1.10* | 0.03 | 0.35** | 1.42** | 0.00 |
| Recreation and leisure (negative) | -0.29 | 0.75 | 0.09 | -0.83** | 0.44** | 0.01 | -2.03** | 0.13** | 0.00 |
| Socialization (positive) | 0.86** | 2.37** | 0.00 | -0.25** | 0.78** | 0.00 | -0.08 | 0.92* | 0.05 |
| Socialization (negative) | 0.04 | 1.04 | 0.89 | -0.30 | 0.74 | 0.49 | -2.22** | 0.11** | 0.00 |
| Physical activity (positive) | 1.71** | 5.51** | 0.00 | 0.36** | 1.44** | 0.00 | 0.37** | 1.45** | 0.00 |
| Physical activity (negative) | -0.14 | 0.87 | 0.44 | -1.00** | 0.37** | 0.00 | -1.41** | 0.24** | 0.00 |
| Greenery (positive) | -0.05 | 0.95 | 0.11 | 0.01 | 1.01 | 0.91 | 0.15** | 1.16** | 0.00 |
| Greenery (negative) | -0.75** | 0.47** | 0.00 | -0.40 | 0.67 | 0.06 | -1.40** | 0.25** | 0.00 |
| Ecology and biodiversity (positive) | 0.14* | 1.15* | 0.02 | -0.18* | 0.84* | 0.02 | 0.03 | 1.03 | 0.68 |
| Ecology and biodiversity (negative) | -0.29 | 0.75 | 0.09 | -0.11 | 0.90 | 0.64 | -0.65** | 0.52** | 0.00 |
| Microclimate (positive) | 0.67** | 1.95** | 0.00 | -0.18 | 0.83 | 0.06 | 0.51** | 1.66** | 0.00 |
| Microclimate (negative) | -1.13** | 0.32** | 0.00 | -0.42 | 0.66 | 0.37 | -2.31** | 0.10** | 0.00 |
| Connectivity (positive) | -0.21** | 0.81** | 0.00 | 2.18** | 8.87** | 0.00 | -0.30** | 0.74** | 0.00 |
| Connectivity (negative) | -0.39** | 0.68** | 0.01 | 0.94** | 2.56** | 0.00 | -1.07** | 0.34** | 0.00 |
| Crowdedness (positive) | 0.00 | 1.00 | 0.99 | 0.16 | 1.17 | 0.20 | -0.10 | 0.91 | 0.39 |
| Crowdedness (negative) | 0.19 | 1.21 | 0.14 | 0.26 | 1.30 | 0.08 | -0.71** | 0.49** | 0.00 |
| Maintenance (positive) | -0.49* | 0.61* | 0.04 | -0.71* | 0.49* | 0.04 | -1.24** | 0.29** | 0.00 |
| Maintenance (negative) | -1.50** | 0.22** | 0.00 | -1.14** | 0.32** | 0.00 | -3.40** | 0.03** | 0.00 |
| Safety (positive) | -0.09 | 0.92 | 0.49 | -0.20 | 0.82 | 0.19 | -0.61** | 0.54** | 0.00 |
| Safety (negative) | -1.34** | 0.26** | 0.00 | -0.54 | 0.58 | 0.11 | -2.96** | 0.05** | 0.00 |
| Accuracy | 0.74 | | | 0.89 | | | 0.79 | | |
| Pseudo $R^2$ | 0.20 | | | 0.17 | | | 0.09 | | |

Table 4. Logistic regression models of user experiential positivity and overall evaluations (*: $p<0.05$; **: $p<0.01$; OR: odds ratio, OR=1 indicates no effect, OR>1 indicates experience dimension associated with higher odds of evaluation dimension, and OR<1 indicates experience dimension associated with lower odds of evaluation dimension)

Positive experiences related to public space amenity and services (OR=2.64), recreation and leisure (OR=2.29), socialization (OR=2.37), and physical activity aspects (OR=5.51) emerged as key drivers of usability evaluations, suggesting that opportunities for active use are central to how users assess public space usability (Zhao et al., 2025). Other dimensions, such as greenery (OR=0.95, p>0.05), showed weaker associations, indicating that usability may not be primarily determined by environmental attributes alone. Negative experiences, however, show significant suppressive associations: negative safety- (OR=0.26) and maintenance-related experiences (OR=0.22) were both substantially negatively associated with overall usability evaluations, suggesting that these dimensions may still constitute preconditions for public space quality.

For accessibility evaluations, connectivity-related experiences emerged as the sole significant positive predictor (OR=8.87), underscoring that connection within and integration with the surrounding urban environment is the core driver of perceived accessibility. Amenity- (OR=1.32) and physical activity-related aspects (OR=1.44) also showed modest predictive effects, plausibly because functional public spaces with clear activity purposes, such as sport parks, may shape users' evaluations of distance and reachability (Gascon et al., 2015).

Attractiveness modeling revealed a more comprehensive evaluative pattern, with associations identified across multiple public space quality dimensions. Aesthetic experience showed the most significant positive association (OR=1.70), indicating that visual quality may be key to users' evaluations in this domain. This may be explained by the alignment of visually appealing public space scenes with the public's preference for fascination and compatibility. The relatively lower explanatory power of the attractiveness model may reflect more substantial individual variation in attractiveness perception, warranting further investigation.

Collectively, these results highlight the differentiated experiential factors underpinning users' evaluations across dimensions, reaffirming the value of integrative investigation of multidimensional user feedback in revealing more comprehensive patterns and potential mechanisms underlying public experiences in urban spaces. These findings also provide insights for future urban public space practices, suggesting the adoption of targeted and differentiated strategies for enhancing different spatial quality dimensions.

### 5.4 Spatial Heterogeneity in the Mechanisms Linking User Experience to Evaluation

Focusing on the attractiveness dimension, we further explored potential spatial heterogeneity of the evaluation mechanisms by constructing stratified logistic regression models separately for central (within the *Centrum* district) and peripheral public spaces (Figure 4). Interestingly, results reveal broadly similar drivers across both contexts but also highlight notable differences, suggesting that pathways contributing to attractiveness evaluations are not fixed but may be linked to location and context.

Specifically, we noted that positive experiences related to public space recreation (center: OR=1.89; periphery: OR=1.17) and amenity (center: OR=1.58; periphery: OR=1.41) are more central to attractiveness evaluation in the city center context, which may reflect higher user expectations in central locations.

Microclimate regulation-related positive experiences also showed a more significant association in central public spaces (center: OR=2.24; periphery: OR=1.48), possibly due to the more pronounced perceived benefit of environmental and climatic amelioration within the dense urban environment. In contrast, results show that socialization exhibits a stronger association in peripheral public spaces (center: OR=0.70; periphery: OR=1.20). Also, negative crowdedness- (center: OR=0.95; periphery: OR=0.33) and safety-related experiences (center: OR=0.17; periphery: OR=0.03) displayed stronger associations in peripheral contexts, implying that users of public spaces outside the city center may be more sensitive to deficiencies in these foundational quality dimensions, while users of central public spaces may exhibit greater tolerance for these shortcomings.

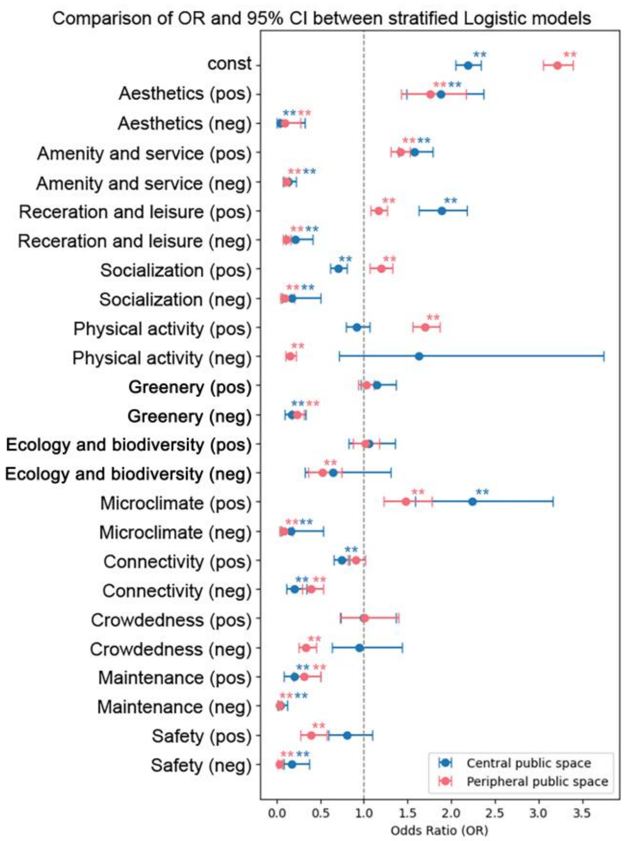


Figure 4. Geographically stratified logistic regression odds ratio comparison of experiential positivity on overall evaluations (pos: positive; neg: negative; *: $p<0.05$; **: $p<0.01$)

We conducted an in-depth analysis of the rich semantics of review data by applying LSVA to the merged review text and image captions, which corroborates these findings (Figure 5). In central public spaces, amenity- and service-related expressions, including *cafe, restaurant*, *monument*, *lively*, among others, exhibited notably higher salience and were associated with more positive valence compared to peripheral public spaces. In peripheral spaces, the stronger role of socialization present in the logistic regression models is supported by the comparably higher salience of terms such as *family*, *picnic*, *child*, and *playground*, pointing to family-oriented interaction as a dominant mode of use.

Collectively, these results underscore the nuanced differences and heterogeneity in experiential and evaluative mechanisms revealed by the integrated analytical approach. Findings highlight the heterogeneity of public evaluations and mechanisms, suggesting that a "one-size-fits-all" approach may not be the best solution for urban public space design, and providing empirical grounding for more context-sensitive, human-centric, and responsive strategies in future design and management.

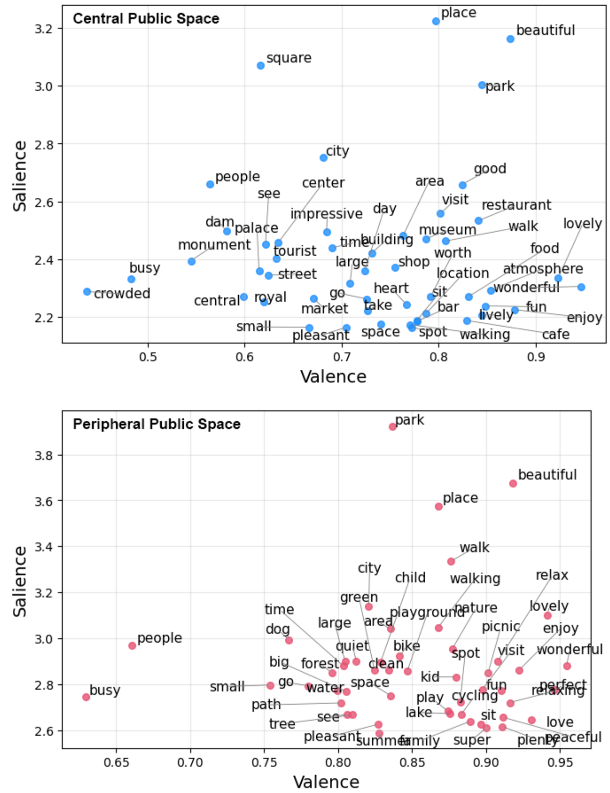


Figure 5. LSVA analysis of semantic information related to user evaluations in central and peripheral public spaces

## 6. Conclusion

Big data such as UGC texts and images present new opportunities for efficiently capturing user insights in urban public spaces, yet research leveraging these sources to examine nuanced and multidimensional subjective experiences and evaluations in public spaces remains scarce and methodologically constrained, highlighting two key challenges: the lack of domain-specific models and datasets capable of extracting complex user experience and evaluation information, and the absence of an integrated framework for delivering multidimensional and comprehensive user insights.

To address these gaps, this study proposed a novel and integrated NLP analytical model that draws on multimodal online review data to examine detailed user experience and evaluation of public spaces with enhanced depth and breadth. The framework combines a LoRA-fine-tuned RoBERTa language model with a lexicon-based ABSA pipeline to extract targeted user evaluation and experience insights, respectively, and incorporates a BLIP vision-language model to integrate information across textual and visual modalities.

Validation and the case study demonstrate that the proposed model provides a more integrated and effective UGC data-driven perspective. Compared with established baseline methods, it enables more accurate, targeted, and detailed extraction of individual-level user insights regarding public spaces, facilitating a more thorough analysis of the rich semantic and visual information embedded in UGC data. Furthermore, its application supports the joint analysis of multidimensional user experiences and evaluations, thereby enabling deeper investigation into the underlying patterns and mechanisms of human-environment interaction in public spaces. This approach offers a new methodological pathway for future research and practice to more comprehensively capture public responses to urban public spaces from UGC data, ultimately supporting evidence-based management and design practices.

Admittedly, several limitations remain in this study and could be addressed in future research. Our dataset represents only users who visited urban public spaces and posted reviews, limiting its representativeness of the general population. The POI-based urban public space selection criteria may not cover all public space types or instances, and our filtering process may result in the inclusion of more popular sites instead of less visited spaces. Despite applying advanced multimodal methods and conducting manual quality controls, the inherent subjectivity of individual insights in urban spaces may lead to analytical gaps and limitations. We thus believe that several strategies could be adopted in future research for better assessing user experiences in urban public spaces: (1) integration of a UGC data-driven approach with traditional survey methods or other participatory approaches, such as VR experiments, for cross-validation; (2) combination of multiple UGC and big data sources for triangulation of findings and mitigation of representativeness challenges; and (3) longitudinal verification of findings with larger datasets to deliver causal relationships for better guiding public space planning and design.

## References

Devlin, J., Chang, M.-W., Lee, K., Toutanova, K., 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. https://doi.org/10.48550/arXiv.1810.04805

Gascon, M., Triguero-Mas, M., Martínez, D., Dadvand, P., Forns, J., Plasència, A., Nieuwenhuijsen, M.J., 2015. Mental Health Benefits of Long-Term Exposure to Residential Green and Blue Spaces: A Systematic Review. *International Journal of Environmental Research and Public Health*, 12, 4354–4379. https://doi.org/10.3390/ijerph120404354

Gehl, J., 1987. *Life Between Buildings: Using Public Space*. Island Press, Washington, D.C.

Greaves, F., Pape, U.J., King, D., Darzi, A., Majeed, A., Wachter, R.M., Millett, C., 2012. Associations between internet-based patient ratings and conventional surveys of patient experience in the English NHS: an observational study. *BMJ Quality & Safety*, 21, 600–605. https://doi.org/10.1136/bmjqs-2012-000906

Honnibal, M., Montani, I., Van Landeghem, S., Boyd, A., 2020. spaCy: Industrial-strength Natural Language Processing in Python. https://doi.org/10.5281/zenodo.1212303

Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W., 2021. LoRA: Low-Rank Adaptation of Large Language Models. https://doi.org/10.48550/arXiv.2106.09685

Huai, S., Van de Voorde, T., 2022. Which environmental features contribute to positive and negative perceptions of urban parks? A cross-cultural comparison using online reviews and Natural Language Processing methods. *Landscape and Urban Planning*, 218, 104307. https://doi.org/10.1016/j.landurbplan.2021.104307

Li, J., Gao, J., Zhang, Z., Fu, J., Shao, G., Zhao, Z., Yang, P., 2024. Insights into citizens' experiences of cultural ecosystem services in urban green spaces based on social media analytics. *Landscape and Urban Planning*, 244, 104999. https://doi.org/10.1016/j.landurbplan.2023.104999

Li, J., Li, D., Xiong, C., Hoi, S., 2022. BLIP: Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation. https://doi.org/10.48550/arXiv.2201.12086

Li, S., Wang, C., Rong, L., Zhou, S., Wu, Z., 2024. Understanding How People Perceive and Interact with Public Space through Social Media Big Data: A Case Study of Xiamen, China. *Land*, 13. https://doi.org/10.3390/land13091500

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V., 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. https://doi.org/10.48550/arXiv.1907.11692

Loshchilov, I., Hutter, F., 2019. Decoupled Weight Decay Regularization. https://doi.org/10.48550/arXiv.1711.05101

Mikolov, T., Chen, K., Corrado, G., Dean, J., 2013. Efficient Estimation of Word Representations in Vector Space. https://doi.org/10.48550/arXiv.1301.3781

Paget, S., 2025. Local Consumer Review Survey. BrightLocal. https://www.brightlocal.com/research/local-consumer-review-survey-2025/. Accessed: 2026-03-19.

Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., Sutskever, I., 2021. Learning Transferable Visual Models From Natural Language Supervision. https://doi.org/10.48550/arXiv.2103.00020

Wang, Z., Yang, Y., Nijhuis, S., van der Spek, S., 2025. Understanding human-environment interaction in urban spaces with emerging data-driven approach: A systematic review of methods and evidence. *Cities*, 167, 106346. https://doi.org/10.1016/j.cities.2025.106346

Yang, H., Zhang, C., Li, K., 2023. PyABSA: A Modularized Framework for Reproducible Aspect-based Sentiment Analysis. https://doi.org/10.48550/arXiv.2208.01368

Zhang, H., Wang, Z., Yang, Y., Nijhuis, S., 2026. Mapping urban greenery: The spatial relationship among NDVI, GVI, and greenery perceptions of social media in Rotterdam. *Frontiers of Architectural Research*. https://doi.org/10.1016/j.foar.2026.05.007

Zhao, X., Huang, H., Lin, G., Lu, Y., 2025. Exploring temporal and spatial patterns and nonlinear driving mechanism of park perceptions: A multi-source big data study. *Sustainable Cities and Society*, 119, 106083. https://doi.org/10.1016/j.scs.2024.106083