A Data-Driven Interface for Understanding Urban Interactions in Public Green Spaces

Bertuğ Kaan Demir[1], Başak Uçar[2], İnan Utku Türkmen[3]
[1] İstanbul Technical University, [2] TED University, [3]Rubiks Analytics
[1]demirbe24@itu.edu.tr, [2]basak.ucar@tedu.edu.tr, [3]inanutku@gmail.com

The rapid digitalization and widespread use of digital technologies in the 21st century have significantly transformed disciplines, including urban studies. Today, understanding the complex interplay between urban spaces and their users relies heavily on data-driven methods, necessitating innovative approaches to explore and enhance these interactions. This paper proposes a human-centric, data-driven interface that uses Large Language Models (LLMs) to analyze user-generated content. Unlike earlier methods such as Latent Dirichlet Allocation (LDA), which struggled with short, nuanced texts, this study leverages advanced natural language processing (NLP) to offer deeper insights into urban experiences. The focus is on Kuğulu Park, a green space in Ankara, where user reviews are analyzed to understand behavioral, emotional, and thematic dimensions of urban interactions. The interface integrates data collection, preprocessing, and visualization processes, with adaptability for different datasets and analytical needs. Using advanced NLP models, it extracts thematic and sentiment-based insights even from fragmented or context-rich user inputs. The study collects multilingual reviews and ratings from publicly available platforms, offering a bottom-up, exploratory model for examining urban behavior. By employing this data-driven approach, the paper aims to shed light on evolving perceptions of urban spaces and behaviors over time, correlating user sentiment with spatial and temporal dynamics. It underscores the transformative potential of NLP models in urban studies, offering new opportunities for understanding and reshaping urban spaces through data-driven methodologies.

Keywords: *Urban Informatics, Human-Centric Design, Natural Language Processing (NLP), Sentiment Analysis.*

INTRODUCTION

Contemporary spatial design increasingly relies on data-driven methodologies, harnessing vast digital datasets generated by citizens through various platforms—such as online reviews, social media, sensors, and wearable devices—to better understand and improve urban environments (Batty, 2014; Kitchin, 2014). This shift toward integrating real-time data into urban planning aligns with a broader transformation characterized by a transition from traditional top-down methods to human-centered, participatory approaches (Foth, 2008; Thakuriah et al., 2017a).

Yet, the primary challenge lies not in data availability, but rather in effectively interpreting complex, diverse, and often unstructured datasets to provide meaningful, actionable insights for urban designers and planners (Chen, 2012; Valls et al., 2018). Although traditional Natural Language Processing (NLP) techniques have facilitated initial thematic explorations, they frequently fall short in capturing nuanced human perceptions, emotions, and context-specific interactions within urban spaces (Nadkarni et al., 2011; Jelodar et al., 2017).

The recent advent of generative artificial intelligence, particularly Large Language Models (LLMs), offers new potential to address these limitations. These advanced models enable a richer and deeper semantic interpretation of textual data, capturing not just explicit thematic content but also subtle emotional nuances and culturally embedded contexts (Kvan, 2020; He, 2020). Such sophisticated analytical capabilities make LLMs particularly suitable for translating large-scale, unstructured user feedback into actionable insights within urban design processes (Tuncer, 2020; Ratti and Claudel, 2016).

Responding to these emerging opportunities, this research introduces an interactive, generative AI-powered digital interface designed to visualize and interpret real-time urban interactions. Demonstrated through a detailed case study of Kuğulu Park in Ankara, Turkey, the interface provides dynamic, accessible, and continuous insights into user perceptions, directly informing human-centered design decisions.

This study introduces a digital, data-driven interface designed to interpret how people experience public green spaces by analyzing user-generated reviews collected from online platforms. Focusing on Kuğulu Park in Ankara, we gathered over 9,000 public reviews via Google's Places API and processed them using advanced Natural Language Processing (NLP) techniques. Unlike earlier approaches such as keyword search or topic modeling, we employed Large Language Models (LLMs)—a new generation of AI systems capable of understanding context, emotion, and thematic content in everyday language. The interface visualizes user sentiment, thematic focus, and temporal patterns across multiple layers, offering designers, planners, and researchers an accessible way to explore how urban green spaces are perceived and used. By bridging computational analysis with spatial experience, the model aims to support more informed and responsive design decisions in public space management.

Urban informatics and data-driven design

Urban informatics responds to the complexities of contemporary cities by integrating data science, geographic information systems, and traditional disciplines like urban planning and sociology, aiming to capture the intricate interplay between social, physical, and infrastructural systems (Thakuriah et al., 2017). From a design perspective, this approach translates diverse data—ranging from sensor outputs to user-generated content—into actionable planning insights, highlighting emergent trends, mobility patterns, and spatial interactions in near-real time.

Historically, data-driven urbanism emerged from computational urban research, using simulation and visualization methods to interpret urban dynamics (Batty, 2013; Kitchin, 2015). Initially, smart city projects relied heavily on sensor infrastructures for traffic or resource optimization but often overlooked qualitative dimensions such as personal narratives or emotional experiences (Kitchin, 2015). My previous research (2023) also identified limitations within traditional NLP methods like LDA, emphasizing their struggle to capture deeper emotional and context-specific meanings in fragmented user-generated texts.

Recent advancements in generative artificial intelligence and large language models (LLMs) have significantly enhanced these capabilities. Unlike earlier NLP approaches, LLMs interpret subtle themes and emotional nuances, turning informal user-generated texts into meaningful design insights. This evolution marks a shift toward an urban informatics practice that integrates generative intelligence and real-time feedback loops, enabling more empathetic and adaptive urban interventions.

Advanced language models in urban studies

The shift from traditional Natural Language Processing (NLP) techniques to Generative AI, particularly Large Language Models (LLMs), represents a significant milestone in urban design analytics. Unlike earlier NLP methods, such as feature engineering or dictionary-based sentiment analysis, LLMs excel in capturing nuanced local contexts, informal speech patterns, and rapidly evolving urban vocabularies, thereby providing a more comprehensive understanding of urban communication.

Generative AI fundamentally disrupts this paradigm. Leveraging advanced neural network architectures and extensive textual training datasets, LLMs significantly elevate interpretive depth beyond mere keyword detection or topic clustering. They skillfully navigate complex linguistic phenomena—such as multilingual content, local idioms, sarcasm, and subtle emotional expressions—making them particularly suited to capturing urban communication's inherently messy, fragmented, yet rich nature.

LLMs, with their ability to integrate contextual cues across larger textual inputs, offer a deeper and more accurate interpretation of community feedback. This adaptability ensures that urban designers can work with cohesive, nuanced insights that are always aligned with the latest community dialogues and social trends. Unlike static, dictionary-based approaches, LLMs dynamically adapt to new urban terminologies or localized slang, ensuring they remain relevant and sensitive to linguistic changes.

While LLMs offer significant advantages, they also pose inherent risks, such as biases and the potential to marginalize underrepresented voices. The complex' black-box' nature of these models further challenges transparency and interpretability. This underscores the critical need for robust human oversight, ethical auditing frameworks, and inclusive, participatory verification methods when using LLMs in urban design analytics.

LLMs have the potential to redefine urban design workflows, which traditionally rely on designers' intuition, limited analytical tools, and personal heuristics to interpret community input (Batty, 2013; Kvan, 2020). By integrating Large Language Models (LLMs) into urban analysis, the role of the designer can be transformed. Designers can now validate, interpret, and expand on AI-generated insights, ensuring that these technical outputs remain deeply connected to authentic spatial and social contexts (Ratti and Claudel, 2016; Tuncer, 2020). This shift presents exciting opportunities for urban designers to lead the way in defining participatory design processes with a holistic approach, where participation in urban decision-making processes is defined by inclusive engagement.

This work does not aim to reduce urban life to computational patterns. Rather, it reframes AI-based analytics as a mediating tool that reveals—but does not replace—the messy, affective, and political dimensions of public space. LLMs in this context are not authoritative narrators, but interpreters that can make latent patterns more legible for human designers.

METHODOLOGY

This study examines Kuğulu Park, a prominent urban green space in Ankara, Turkey, selected for its central location, cultural significance, and high visitor engagement. The park attracts diverse demographic groups, making it an ideal case for analyzing user-generated content to extract meaningful insights into urban interactions. By leveraging digital data sources, this study aims to provide a comprehensive understanding of user perceptions and engagement within urban green spaces.

Data collection and data processing

Data collection was conducted using Google Places API, a widely used tool for retrieving user-generated reviews and ratings from publicly accessible Google Reviews. A structured script was developed to systematically extract data linked to Kuğulu Park's unique Place ID, capturing review texts, star ratings, reviewer usernames, reviewer profile images, and review timestamps. A total of 9,080 text-based user reviews were compiled, each entry accompanied by metadata, including reviewer identification and timestamp, enabling temporal and sentiment-based analyses. The raw data were subsequently stored in a structured database system, facilitating streamlined processing and retrieval. Data ingestion was executed via Google BigQuery, a cloud-based, fully managed data warehouse optimized for handling large-scale structured datasets. This approach established a solid foundation for subsequent data preprocessing and analytical procedures.

To ensure data consistency and analytical rigor, textual preprocessing involved removing non-informative characters, standardizing text formats, and employing natural language processing (NLP) techniques to classify comments based on language. Reviews were categorized into sentiment groups—positive, neutral, and negative—based on star ratings, facilitating sentiment analysis. Advanced analytical procedures were then applied using Large Language Models (LLMs) to extract thematic and emotional insights. LLMs, designed to process and interpret complex textual data, enabled comprehensive thematic clustering, sentiment analysis, and temporal pattern recognition. The subsequent analysis phase identified recurring themes and trends, which are further explored in the Results and Discussion section.

Though all data used in this study are publicly available and collected through the Google Places API, ethical considerations remain central. Reviewer names and profile images have been anonymized in the interface to respect privacy, despite their original public status. Moreover, the representativeness of such user-generated content is acknowledged as limited—favoring digitally active and often privileged demographics.

Identification of dominant themes and sentiment analysis

Thematic analysis was conducted utilizing LLMs to systematically categorize user reviews into dominant themes, revealing key aspects of park users' experiences. The analysis identified overarching themes such as atmosphere and ambiance, children and family activities, natural life and landscape, social interaction, historical and cultural value, cleanliness and maintenance, seating and relaxation, cafe and food-beverage facilities, location and accessibility, activities and sports, and safety and security. These themes encapsulate critical dimensions of user engagement, offering valuable insights for spatial designers and policymakers aiming to enhance visitor experiences in public green spaces. To complement thematic analysis, a temporal examination was performed by associating each identified theme with corresponding review timestamps. This approach enabled the detection of seasonal patterns and variations in user sentiments over time. For instance, fluctuations in themes such as cleanliness, safety, and family-oriented activities provided deeper insights into the dynamic nature of urban park usage. The detailed findings of these analyses are presented in the Results and Discussion section.

In addition to thematic and temporal analyses, numerical ratings provided by users were systematically examined to assess visitor sentiment. Reviews were categorized based on the Net Promoter Score (NPS) framework, classifying ratings of 5 stars as promoters, 4 stars as passives, and ratings of 3 stars or lower as detractors. This classification enabled a structured sentiment analysis, quantifying user satisfaction and perception. The calculated NPS metric provided a quantitative measure of overall park perception, complementing qualitative thematic and temporal analyses.

To facilitate real-time data exploration, an interactive digital interface was developed, integrating thematic categorizations, temporal trends, and sentiment analyses. The interface was implemented using React.js for front-end interactivity, TypeScript for performance optimization, and Tailwind CSS for responsive design. Key features include interactive data visualizations, user-friendly filtering and navigation, and dynamic sentiment and theme exploration. Designed as a scalable and adaptable system, the interface supports continuous data updates, ensuring that insights remain reflective of evolving urban interactions. By leveraging real-time user-generated content, this study presents a novel approach to monitoring and enhancing urban green space experiences.

While LLMs offer remarkable capabilities in interpreting sentiment and extracting themes, they are not without limitations—particularly regarding the risk of hallucinated outputs. In this study, verification is achieved not through rigid ground-truth datasets but via cross-referencing star ratings, temporal markers, and sentiment trends. By correlating LLM-derived insights with metadata patterns, a soft validation layer is established. This enables the interface to prioritize interpretive clarity and design relevance over algorithmic certainty, aligning with the project's goal of offering exploratory insight rather than deterministic conclusions.

INTERFACE

This section systematically presents and critically examines the analytical results obtained from the developed data-driven interface. By leveraging advanced large language models (LLMs), thematic, temporal, and sentiment-based insights were extracted from user-generated content, offering a nuanced understanding of urban interactions at Kuğulu Park in Ankara. Each interface component is demonstrated with corresponding screenshots, and key findings are comprehensively discussed to highlight methodological effectiveness and practical urban design implications.

The interactive interface provides an intuitive overview panel (Figure 1), summarizing fundamental user engagement metrics. This initial screen features the study title, Data-Driven Interface for Understanding Urban Interactions, and user-friendly filtering options allowing users to select specific urban parks and define custom time ranges—the displayed data dynamically updates based on these selections, offering a real-time exploratory tool for urban analysis.

This study's dataset comprises 9,080 text-based user reviews, forming the basis for a comprehensive analysis of urban experiences. General metrics displayed in the initial interface screen provide critical insights into user perceptions:

- Total Reviews: 9,080
- Average Rating: 4.5 out of 5 stars
- Overall Net Promoter Score (NPS): 54
- Peak Visit Time: 16:00 - 17:00

The high average rating (4.5/5) reflects visitors' strong overall positive perception of Kuğulu Park. An NPS score of 54 indicates significant user satisfaction, highlighting a substantial proportion of positive ("promoter") ratings compared to negative ("detractor") feedback. Additionally, the identified peak visit time (16:00 - 17:00) suggests a preference for late afternoon visits, providing valuable insights for urban planners and park management to optimize resource allocation during peak hours.

These preliminary findings establish a foundational context for subsequent detailed thematic and temporal analyses, which are further elaborated in the following sections.

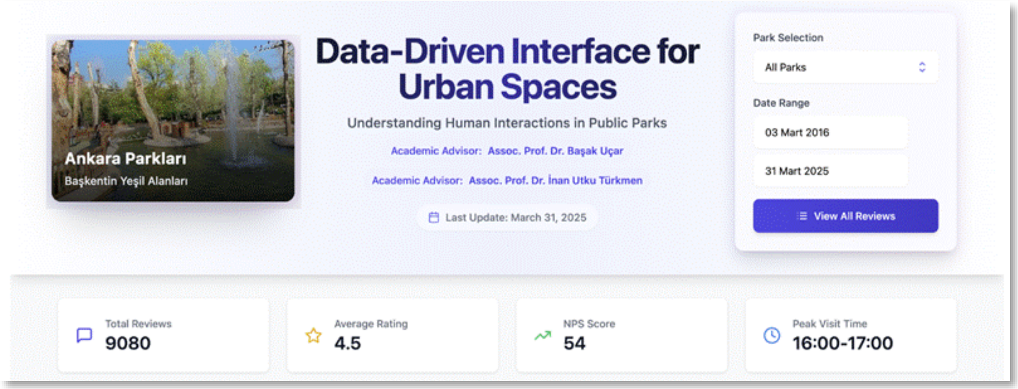

Figure 1
Interface Overview Panel Displaying General Metrics and Park Selection Filter

Thematic analysis and temporal NPS trends

The second section of the interface offers a thorough thematic analysis of user-generated reviews, complemented by an interactive temporal visualization of Net Promoter Score (NPS) fluctuations (Figure 3). Using advanced Large Language Models (LLMs), user reviews were systematically processed to identify the dominant themes in visitor experiences at Kuğulu Park. The results, presented in Figure 3, outline these themes along with their corresponding average ratings and sentiment distributions (positive, neutral, and negative).

The analysis reveals that "Atmosphere and Ambiance" is the most frequently mentioned theme, accounting for 29% of all reviews and predominantly associated with positive sentiments. This emphasizes the significant role of aesthetic and sensory experiences in shaping visitor satisfaction. "Child and Family Activities" is another dominant theme, reinforcing Kuğulu Park's reputation as a family-friendly space. However, the analysis also highlights areas of concern. For example, "Cleanliness and Maintenance" received the lowest average rating (3.9) and the highest proportion of negative sentiment (33%), signaling a key focus area for urban management and potential improvements.

In addition to thematic insights, the interactive visualization enables dynamic tracking of NPS trends over time. As illustrated in the right panel of the interface (Figure 3), fluctuations in NPS scores are displayed from the earliest recorded review to the most recent. At a selected point (November 1, 2021), the NPS was recorded at 53, with 66% promoters and 13% detractors.

The temporal variations observed suggest that visitor experiences are influenced by several factors, including seasonal changes, local events, or infrastructural updates within or around the park. The interface's interactive features allow urban planners and park administrators to pinpoint specific periods where user satisfaction experiences significant shifts. For instance, a peak NPS of 100 and a sharp dip to -17 indicate dramatic changes, which could reflect either exceptionally positive events or notable challenges faced by visitors during those times. By integrating thematic analysis with temporal tracking, the interface provides a powerful tool for data-driven decision-making, enabling urban planners and stakeholders to proactively address user concerns and enhance the quality and sustainability of urban spaces.

Theme	Avg. Rating	Comments (%)	Positive (%)	Neutral (%)	Negative (%)
Atmosphere and Ambiance	4.7	29%	75%	18%	7%
Child and Family Activities	4.5	25%	67%	20%	13%
Natural Life and Landscape	4.3	18%	62%	20%	19%
Social Interaction	4.2	8%	56%	23%	21%
Historical & Cultural Value	4.6	4%	75%	18%	8%
Cleanliness and Maintenance	3.9	4%	42%	26%	33%

Figure 2
Dominant Themes and Sentiment Distribution

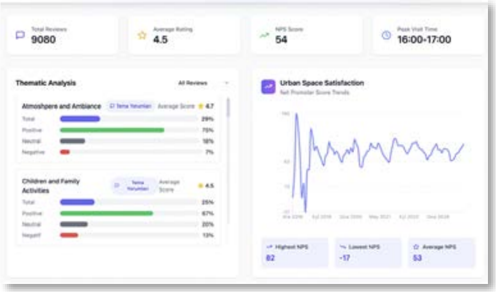

Figure 3
Interface Visualization for Thematic Analysis and Temporal NPS Trends.

Keyword analysis

The significance of the interface is its ability to extract and visualize the most frequently mentioned keywords from user reviews (Figure 4). These keywords, identified through advanced Large Language Models, capture the core elements of the visitor experience and are linked to distinct sentiments—positive, neutral, or negative.

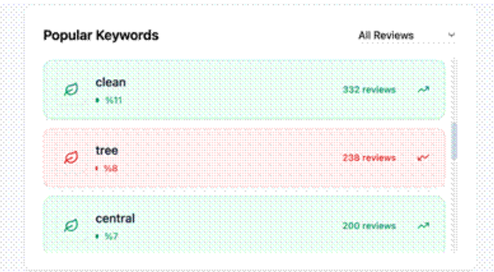

Figure 4
Popular Keywords

The "Popular Keywords" panel displays these keywords alongside their frequency within the dataset and sentiment classifications. Keywords are visually represented with color coding to reflect the predominant sentiment expressed in the reviews:

- Green: Predominantly positive sentiment (e.g., "temiz" [clean], 332 comments, 11% of total)
- Red: Predominantly negative sentiment (e.g., "ağaç" [tree], 238 comments, 8% of total)
- Gray: Neutral sentiment keywords

This interactive feature enables users to engage directly with the keywords, filtering reviews that mention specific terms. This targeted approach facilitates deeper qualitative analysis, allowing stakeholders to quickly identify both critical issues and positive aspects based on frequent keyword occurrences. Further analysis is possible by linking these keywords to individual user reviews. As illustrated in Figure 5, each comment is tagged with relevant thematic labels and associated keywords. This level of detail allows for intuitive cross-referencing between themes and keyword-specific sentiments, providing a clearer understanding of how individual elements of the park environment influence overall visitor experiences. Together, these features significantly enhance the interpretability and practical value of the user-generated data, empowering urban managers to make informed, responsive decisions that are grounded in the nuanced feedback from park visitors.

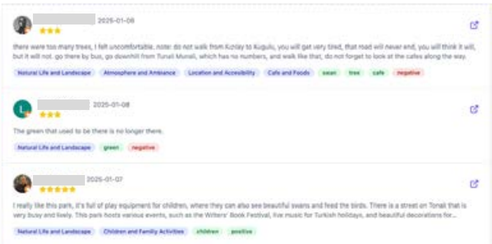

Figure 5
Individual review display with automatically tagged themes and keywords.

Interactive theme-based comment exploration

A key interactive feature of the developed interface is the ability to explore detailed user comments within each identified theme (Figure 6). This functionality allows users to filter comments by sentiment—positive, neutral, or negative—enabling a targeted exploration of feedback based on the emotional tone of visitor responses.

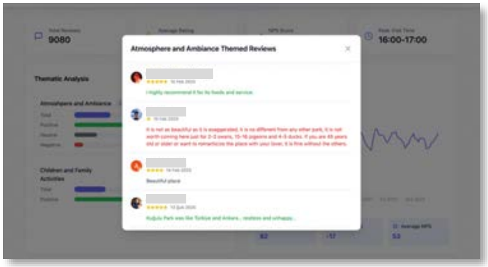

Figure 6
Interactive popup displaying theme-specific user reviews

By selecting the "Theme Comments" button next to a specific theme, users can access a chronological list of individual comments related to that theme, which are displayed in a dedicated popup window. Each comment includes essential user information, such as the reviewer's name, their numerical rating, and the exact date the review was submitted. This feature provides a comprehensive, user-friendly approach to analyzing specific aspects of the visitor experience.

Time-based sentiment and thematic analysis

The interface offers an insightful visualization of hourly visitation patterns and seasonal variations at Kuğulu Park (Figure 7), shedding light on how visitor behavior fluctuates throughout the day and across different seasons. This graph provides essential information for urban planners and park managers, enabling them to understand the temporal utilization of park spaces and visitor behavior.

Several key trends emerge from the analysis. Visitor frequency increases gradually from early morning (06:00–08:00), reaching its peak between 16:00 and 17:00. This afternoon surge suggests a preference for leisure activities and social interactions following work hours, reinforcing earlier findings about visitor engagement during late afternoons. The visualization also clearly illustrates seasonal differences in visitation patterns. Summer emerges as the period with the highest visitor density, particularly during the peak hour, with 195 visits per hour. This increase is likely linked to favorable weather, extended daylight hours, and greater opportunities for outdoor recreation. In contrast, winter shows the lowest visitation frequency, with a peak of only 120 visits per hour, reflecting a drop in engagement due to colder weather and shorter days.

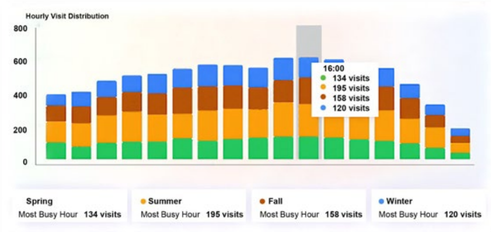

Figure 7
Hourly visitor distribution segmented by seasonal variation

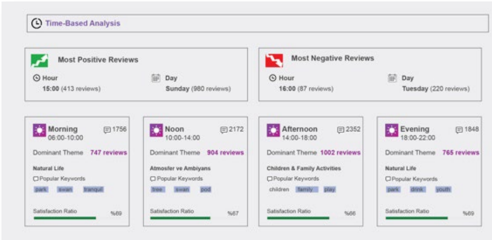

Figure 8
Time-based sentiment and thematic analysis segmented by periods of the day.

Identifying these hourly and seasonal patterns is crucial for effective resource allocation and infrastructure management. For example, ensuring adequate security, maintenance, and staffing during the late afternoon, especially in summer, could significantly enhance the user experience. Similarly, targeted interventions such as additional lighting or sheltered areas could help boost visitation during off-peak seasons like winter, ultimately improving overall park utilization and satisfaction. The next visualization within the interface offers a more granular, time-based analysis of visitor sentiment and dominant thematic content (Figure 8). By segmenting visitor feedback into specific time blocks—morning, noon, afternoon, and evening—the interface reveals distinct patterns in user experiences and preferences throughout the day. The final section of the interface delves into a seasonal breakdown (Figure 9), providing a detailed examination of how visitor preferences, thematic focus, and satisfaction levels change across the year at Kuğulu Park (Figure 8).

Spring sees a high NPS score of 59, with a focus on "Café and Food Services" and "Child and Family Activities." Popular keywords such as "doğa" (nature), "kafe" (café), and "kuğu" (swan) emphasize the park's appeal as a recreational, family-friendly space. Peak visitation during spring occurs mainly on Sundays, particularly in the afternoon (14:00–18:00), reflecting a preference for leisurely weekend activities. Summer is marked by a lower NPS score of 48, indicating increased sensitivity to "Safety and Security" and heightened interest in "Activity and Sports." Keywords like "güvenli" (safe), "temiz" (clean), and "oturma" (seating) point to critical areas that require attention during this busy period. Peak summer visitation occurs on Saturday mornings (06:00–10:00), likely due to visitors seeking cooler temperatures and less crowded conditions.

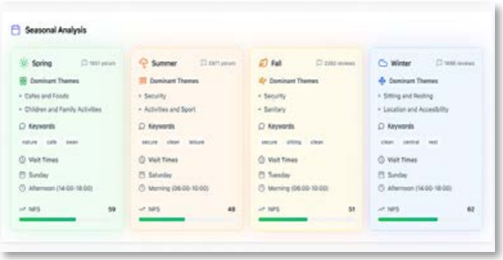

Figure 9
Seasonal analysis highlighting thematic preferences, peak visiting times, and NPS trends across

With a moderate NPS score of 51, autumn visitors continue to express concerns about "Cleanliness and Maintenance," as well as ongoing attention to "Safety and Security." Frequent keywords remain similar to those in summer, underscoring consistent expectations regarding cleanliness and comfort. Peak autumn visitation typically occurs on Tuesday mornings, likely due to quieter conditions after the weekend, presenting an opportunity for targeted improvements to enhance visitor satisfaction early in the week. Winter boasts the highest seasonal NPS score of 62, reflecting overall positive visitor experiences during the colder months. The dominant themes focus on "Seating and Relaxation" and "Location and Accessibility," indicating a preference for comfort and convenience in winter conditions. Through these comprehensive analyses of temporal and seasonal patterns, the interface offers stakeholders actionable insights to optimize park management, enhance user experience, and ensure the park's long-term sustainability.

Persona-based interactive AI insights

One of the most distinctive features of the interface is its persona-based AI chat system (Figure 10), allowing users to interact with urban data through roles such as “Park Visitor,” “Local Resident,” or “Designer.” After selecting a persona and applying filters (e.g., time period or theme), users can ask natural language questions. The LLM responds contextually using relevant data slices—such as pandemic-era feedback or safety-related reviews—providing personalized, real-time insights. This conversational model enables intuitive exploration of urban experiences, supporting human-centered, data-informed decision-making.

CONCLUSION

This study highlights the potential of integrating large language models (LLMs) to enhance urban analysis and decision-making through user-generated content. By leveraging LLMs to conduct thematic and sentiment analysis of user reviews, this work offers a nuanced approach to understanding public perceptions of urban spaces like "Kuğulu Park". The results demonstrate how the combination of LLM-driven insights and interactive temporal tracking can provide valuable, data-driven feedback for urban planners, offering actionable insights into user satisfaction. The findings also show that by utilizing advanced LLM techniques, it is possible to uncover complex patterns in visitor behavior, sentiment, and the dynamics of urban space utilization. These insights are particularly valuable for enhancing user experiences and informing the management of public spaces.

Future research should explore further refinement of sentiment analysis models and the integration of additional datasets, such as social media interactions or demographic data, to provide even more granular insights. Additionally, the incorporation of real-time data streams and predictive analytics could be explored to enable proactive management of urban spaces, anticipating shifts in user satisfaction and demand. Furthermore, enhancing the interface's ability to track longer-term trends and incorporate user feedback from diverse sources would contribute to more sustainable and adaptive urban development strategies. Ultimately, this study underscores the immense potential of LLMs and data-driven interfaces to transform urban management practices and improve the quality of public spaces.

REFERENCES

Batty, M. (2013). The new science of cities. MIT Press.

Batty, M. (2014). 'Big data and urban informatics'. *Proceedings of the NSF Workshop on Big Data and Urban Informatics*, 1–4.

Chen, N. C. (2012). Urban data mining: Social media data analysis as a complementary tool for urban design (Doctoral Thesis, MIT).

Foth, M. (2008). *Handbook of Research on Urban Informatics: The Practice and Promise of the Real-Time City*. IGI Global.

He, S. (2020). 'Data-driven urbanism: The balance between spatial intelligence and design craftsmanship.' *Architectural Design*, 90(3), pp. 86–93. https://doi.org/10.1002/ad.2573.

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., and Zhao, L. (2017). 'Latent Dirichlet Allocation (LDA) and topic modeling: Models, applications, a survey.' *Multimedia Tools and Applications,* 78(11), pp. 15169–15211.

Kitchin, R. (2014). 'The real-time city? Big data and smart urbanism.' *GeoJournal*, 79(1), 1–14.

Kitchin, R. (2015). 'Data-Driven, Networked Urbanism.' *SSRN Electronic Journal*.

Kvan, T. (2020). 'Data-informed design: A call for theory.' *Architectural Design*, 90(3), pp. 26–31. https://doi.org/10.1002/ad.2564.

Nadkarni, P. M., Ohno-Machado, L., and Chapman, W. W. (2011). 'Natural language processing: An introduction.' *Journal of the American Medical Informatics Association*, 18(5), 544–551.

Ratti, C., and Claudel, M. (2016). *The City of Tomorrow: Sensors, Networks, Hackers, and the Future of Urban Life*. Yale University Press.

RIBA and ARUP (2013) *Designing with data: Shaping our future cities*. London: RIBA Publishing.

Thakuriah, P., Tilahun, N. Y., and Zellner, M. L. (2017). *Seeing cities through big data: Research, methods and applications in urban informatics*. Springer Geography.

Tuncer, B. (2020). 'Augmenting reality: (Big-) data-informed urban design and planning.' *Architectural Design*, 90(3), 52–59. https://doi.org/10.1002/ad.2568.

Valls, F., Redondo, E., Fonseca, D., Torres-Kompen, R., Villagrasa, S., and Martí, N. (2018). 'Urban data and urban design: A data mining approach to architecture education.' *Telematics and Informatics*, 35(4), 1039–1052. https://doi.org/10.1016/j.tele.2017.09.015.