Research on Literary and Art Development
2025, VOL. 6, NO. 3, 103-109
DOI: 10.47297/wsprolaadWSP2634-786522.20250603

Data-Driven Analysis of "Post-2010s Generation" Youth Digital Cultural Consumption and Urban Identity in Foshan: Insights from K-means Clustering of Youth Digital Cultural Consumption Reviews

Qian Wang

Guangdong Industry Polytechnic University, Guangzhou, Guangdong, 510220, China

ABSTRACT

This study proposes a data-driven approach to analyzing the relationship between "Post-2010s Generation" youth digital cultural consumption behaviors and urban identity formation in Foshan City. By collecting digital cultural consumption data and urban identity evaluations from 200 adolescents aged 10-15 in Foshan, we employed k-means clustering algorithm to group user feedback and extract meaningful insights. The clustering process, validated through Silhouette Scores and Davies-Bouldin Index metrics, revealed distinct themes in user feedback, highlighting potential enhancement areas for promoting youth urban integration. Our results demonstrate that clustering analysis can effectively differentiate between various types of digital cultural consumption preferences and their impacts on urban identity, providing actionable guidance for urban policymakers and administrators. This study underscores the value of incorporating systematic user feedback analysis into urban identity research, offering a pathway to creating more effective youth urban integration mechanisms. The findings contribute to a more structured and evidence-based approach to urban youth policy development, ultimately enhancing youth urban belonging and participation.

KEYWORDS

Data-driven design; Urban identity; Post-2010s generation youth; Digital cultural consumption; K-means clustering

1 Introduction

The rapid advancement of digital technology has revolutionized the landscape of urban cultural consumption, with "Post-2010s Generation" (born after 2010) adolescents emerging as "digital natives" who represent a crucial force in urban cultural development. This demographic group's digital consumption behaviors and urban identity formation processes have profound implications for future urban development. As an important city in the Guangdong-Hong Kong-Macao Greater Bay Area, Foshan hosts approximately 840,000 "Post-2010s Generation" individuals as of late 2024, comprising 18.3% of the total population. Their digital cultural consumption patterns and urban identity formation have become key issues in youth-friendly city construction.

Traditional urban identity research often relies on questionnaire surveys and theoretical analysis, lacking deep understanding of youth digital behaviors. Although existing research has explored the impact of digital media on youth identity formation, empirical studies combining digital cultural consumption behaviors with specific urban identity remain limited, particularly systematic analyses targeting the unique "Post-2010s Generation" demographic.

Data-driven research methods provide a more systematic and evidence-based analytical framework for urban identity research. By analyzing user feedback and behavioral data, researchers can identify key factors influencing urban identity formation. These insights can optimize urban cultural policies, ensuring policy development meets youth needs and preferences while promoting urban identity enhancement.

One primary challenge in urban identity research is balancing different dimensions of identity elements. Researchers often struggle to comprehensively grasp the complex relationships among cognitive identity, emotional identity, behavioral identity, and future identity, leading to results that are either overly theoretical and lack practical application or too fragmented to form systematic understanding. Another challenge lies in youth diversity, where each individual has unique digital consumption preferences, learning styles, and identity needs, making it difficult to design urban integration strategies suitable for all users.

This study explores how analyzing "Post-2010s Generation" youth digital cultural consumption evaluation data can enhance understanding of urban identity formation mechanisms. Through systematic user feedback analysis, the research aims to identify common themes, correlation patterns, and best practices between digital cultural consumption and urban identity. Research findings are expected to provide valuable guidance for policymakers seeking to create more

effective and engaging youth urban integration mechanisms.

In this study, we utilize clustering methods to examine datasets of "Post-2010s Generation" youth digital cultural consumption and urban identity in Foshan City, aiming to reveal user preference patterns. The k-means clustering method groups similar feedback, providing a systematic approach to understanding diverse viewpoints and extracting actionable insights. Through methodical grouping of user data, we can identify common patterns that guide policy strategies, leading to more impactful and engaging youth urban integration mechanisms.

2 Literature review

2.1 Digital cultural consumption and youth identity formation

Digital cultural consumption has become a fundamental aspect of contemporary youth identity formation. Research demonstrates that digital platforms serve as critical spaces where young people construct, negotiate, and express their identities (Jenkins, 2006). The "Post-2010s Generation" represents the first truly digital native cohort, having grown up immersed in digital technologies from birth (Prensky, 2001). Their identity formation processes are inherently intertwined with digital media consumption patterns. Studies have shown that digital cultural consumption influences identity formation through multiple pathways, including information acquisition, social interaction, and self-expression (boyd, 2014). Young people use digital platforms not merely as entertainment venues but as spaces for identity exploration and validation (Turkle, 2011). The consumption of digital content allows youth to experiment with different identities, connect with like-minded peers, and develop a sense of belonging within various communities.

2.2 Urban identity and place attachment

Urban identity represents individuals' psychological connection to a specific city or urban environment, encompassing cognitive, emotional, behavioral, and aspirational dimensions (Proshansky et al., 1983). Research on place attachment suggests that individuals develop emotional bonds with places through direct experience, social interaction, and cultural engagement (Lewicka, 2011). For youth populations, urban identity formation occurs through various mechanisms, including educational experiences, peer interactions, and cultural consumption practices. Contemporary urban studies emphasize the role of cultural experiences in fostering urban identity, particularly among younger generations (Zukin, 1995). Cultural consumption activities, including participation in local festivals, engagement with local media, and consumption of place-specific content, contribute significantly to urban identity development (Lloyd, 2006).

2.3 Technology-mediated urban identity formation

The intersection of digital technology and urban identity formation represents an emerging area of scholarly inquiry. Research suggests that digital technologies can both enhance and complicate urban identity formation processes (Graham & Marvin, 2001). On one hand, digital platforms provide new opportunities for urban engagement, allowing residents to access local information, connect with fellow residents, and participate in urban cultural activities. On the other hand, digital technologies can create "placeless" experiences that potentially weaken connections to specific urban environments (Castells, 1996). Recent studies have explored how location-based social media, augmented reality applications, and digital storytelling platforms can strengthen urban identity among young residents (de Lange & de Waal, 2013). These technologies enable new forms of urban engagement that combine digital interaction with physical place experience.

2.4 Research gaps and opportunities

Despite growing interest in both digital cultural consumption and urban identity formation, limited research has systematically examined their intersection, particularly among the "Post-2010s Generation." Most existing studies focus on either digital consumption patterns or urban identity formation in isolation, without exploring their dynamic relationship. Furthermore, few studies have employed data-driven clustering approaches to identify patterns in youth digital consumption and urban identity formation. This research addresses these gaps by employing k-means clustering analysis to identify patterns in "Post-2010s Generation" digital cultural consumption and urban identity formation in Foshan City. By analyzing actual consumption data and identity assessments, this study provides empirical insights into how digital cultural consumption influences urban identity development among contemporary youth.

3 Methodology

3.1 Data collection

This study utilized survey data from the "Research on Digital Cultural Consumption and Urban Identity among 'Post-2010s Generation' Youth in Foshan City." Data were collected through a comprehensive questionnaire survey conducted from March to June 2025, yielding 200 valid responses from students aged 10-15 across Foshan's five districts.

The survey employed both online and offline data collection methods. Online questionnaires (120 responses) were distributed through WeChat groups and school network platforms, while offline paper questionnaires (80 responses) were distributed in key primary and secondary schools across Foshan's five districts. All questionnaires were completed with assistance from parents or teachers using anonymous methods to ensure data authenticity and privacy protection.

3.2 Sample characteristics

The sample demonstrated balanced demographic representation: 108 males (54%) and 92 females (46%); age distribution primarily concentrated in the 12-13 age group (42%); 132 participants (66%) born locally in Foshan, 68 (34%) migrated from other regions; residential distribution mainly in Chancheng District (29%), Nanhai District (31.5%), and Shunde District (23.5%).

3.3 Data structure

Table 1 Data Components

Component	Definition
Basic Information	Gender, age, school, grade, birthplace, and residential area
Digital Cultural Consumption	Usage time, platform preferences, spending amounts, content preferences, and influencing factors
Urban Identity Assessment	Identity degree, understanding level, emotional attitudes, cultural cognition, and future intentions
Open-ended Responses	Descriptions of Foshan's most attractive aspects and digital cultural content suggestions

3.4 Text preprocessing

Digital cultural consumption evaluation preprocessing involved several steps to clean, normalize, and prepare text data for clustering analysis. The preprocessing aimed to convert raw text into structured and consistent format suitable for clustering algorithms.

Preprocessing Steps:

Text segmentation: Chinese text was segmented into individual words using jieba segmentation tool

Stop words removal: Common Chinese stop words such as "的", "和", "是" were removed

Special characters removal: Punctuation marks and special characters not contributing to semantic meaning were eliminated

Normalization: Text was converted to consistent format to ensure uniformity

3.5 TF-IDF vectorization

This study employed TF-IDF (Term Frequency-Inverse Document Frequency) vectorization to convert text data into numerical form, capturing term importance in documents relative to the entire corpus.

Calculation Formulas:

Term Frequency (TF): $TF(t,d) = \frac{Number\ of\ times\ term\ t\ appears\ in\ document\ d}{\text{Total number of terms in document d}}$

Inverse Document Frequency (IDF): $IDF(t,D)=\log\left(1+\frac{Total\ number\ of\ documentsents}{Number\ of\ documents\ containing\ term\ t}\right)$

TF-IDF: $TF\text{-}IDF(t,d,D)=TF(t,d)\times IDF(t,D)$

3.6 K-means clustering algorithm

We employed k-means clustering for this study due to its unsupervised learning capabilities, eliminating the need for pre-labeled data. The algorithm partitions datasets into k clusters, where each data point belongs to the cluster with the nearest centroid, aiming to minimize within-cluster variance.

Algorithm Process:

(1) Initialization: Randomly select k initial centroids from the dataset

(2) Assignment: Assign each data point to the nearest centroid using Euclidean distance:

$d(x_i,\mu_j)=\|x_i-\mu_j\|_2$

(3) Update: Recalculate centroids by averaging features of all points in each cluster

(4) Iteration: Repeat assignment and update steps until centroids stabilize

Objective Function: Minimize within-cluster sum of squares (WCSS):

$WCSS = \sum_{j=1}^{k} \sum_{x_i \in C_j} \|x_i-\mu_j\|_2$

This study modified the k-means algorithm by incorporating domain expert-defined design features to determine k value. These features include: (1) digital content preference characteristics, (2) platform usage patterns, (3) consumption time distribution, (4) content type preferences, (5) social interaction characteristics, and (6) urban identity dimensions. Based on these features, k was set to 6.

3.7 Validation metrics

Clustering performance was evaluated using two metrics:

(1) Silhouette score (S.S.):Measures similarity between data points within their own cluster compared to other clusters, ranging from -1 to 1: $S(i)= \frac{b(i)-a(i)}{max(a(i),b(i))}$

(2) Davies-bouldin index (DBI): Measures average similarity ratio of each cluster to its most similar cluster, with lower values indicating better clustering:$DBI= \frac{1}{k}\sum_{i=1}^{k} \max_{j\neq i}\left(\frac{\sigma_i+\sigma_j}{d(c_i,c_j)}\right)$

(3) Clustering performance

Table 2 Clustering Results

Data Dimension	Cluster	S.S.	Avg. S.S.	DBI	Avg. DBI
Digital Cultural Consumption	Cluster-1	0.38	0.365	2.25	2.32
	Cluster-2	0.42		2.15	
	Cluster-3	0.35		2.45	
	Cluster-4	0.40		2.20	
	Cluster-5	0.33		2.50	
	Cluster-6	0.31		2.35	
Urban Identity	Cluster-1	0.45	0.412	2.10	2.18
	Cluster-2	0.48		2.05	
	Cluster-3	0.38		2.30	
	Cluster-4	0.42		2.15	
	Cluster-5	0.40		2.25	
	Cluster-6	0.34		2.23	

The constraint-based k-means clustering results demonstrated moderate clustering performance. Digital cultural consumption dimension achieved an average Silhouette Score of 0.365 and average DBI of 2.32, indicating some cluster separation but not highly distinctive groupings. Urban identity dimension performed slightly better with average Silhouette Score of 0.412 and average DBI of 2.18.

3.8 Cluster analysis

Through k-means clustering analysis of 200 "Post-2010s Generation" youth digital cultural consumption and urban identity data, we identified six major clusters:

3.8.1 Cluster 1: traditional culture identity type (18.5%)

Characteristics: High awareness of Foshan kung fu, ceramic culture, and other traditional elements

Digital consumption preferences: Preference for national style/Lingnan cultural content

Urban identity manifestation: Strong identity feelings, high willingness to remain in Foshan

3.8.2 Cluster 2: modern urban eExperience type (24.0%)

Characteristics: Focus on urban modernization facilities and convenient living

Digital consumption preferences: Preference for video platforms and social software

Urban identity manifestation: Positive attitude toward urban development prospects

3.8.3 Cluster 3: social interaction driven type (20.5%)

Characteristics: Value peer influence and social experiences
Digital consumption preferences: High frequency of gaming and social media usage
Urban identity manifestation: Strengthen urban identity through peer groups

3.8.4 Cluster 4: learning and growth oriented type (16.0%)

Characteristics: Focus on educational resources and personal development
Digital consumption preferences: High usage rate of learning applications
Urban identity manifestation: View city as personal growth platform

3.8.5 Cluster 5: food culture preference type (11.5%)

Characteristics: Deep understanding of Foshan food culture
Digital consumption preferences: Focus on local food-related content
Urban identity manifestation: Establish emotional connections through food culture

3.8.6 Cluster 6: low participation observer type (9.5%)

Characteristics: Relatively low digital consumption and urban identity
Digital consumption preferences: Short consumption time, limited content variety
Urban identity manifestation: Ambiguous identity feelings, requiring guidance

3.9 Key findings

(1) Digital cultural consumption patterns: Analysis revealed that 78% of respondents primarily used social software, 71% used video platforms, and 59% engaged with gaming platforms. However, only 20.5% accessed Foshan-related local digital content, indicating limited penetration of local digital cultural content among youth.

(2) Urban identity status: 59.5% of respondents considered themselves "Foshan people," 81% held positive emotional attitudes toward Foshan, and 71.5% expressed willingness to live/work in Foshan in the future. Cultural cognition analysis showed that apart from Foshan kung fu (76.5%), awareness of other traditional cultural elements remained relatively low.

(3) Correlation analysis: Youth frequently accessing Foshan-related digital content demonstrated significantly higher urban identity (76.8%) compared to those with limited local content exposure (48.3%). This suggests digital media serves as an important bridge connecting youth with urban culture.

(4) Age and regional differences: Urban identity formation peaked during early adolescence (12-13 years), representing a critical period for identity development. Chancheng and Shunde districts showed higher identity rates compared to other areas.

4 Discussion

4.1 Data-driven approaches to youth urban integration policy

Data-driven methods guide youth urban integration policy design and improvement through data analysis and insights. Utilizing user data—such as evaluations, trends, and feedback—enables policymakers to enhance personalization, engagement, and integration effectiveness.

4.2 Key policy development contributions

Enhanced User-Centered Policy Design: Data-driven insights help policymakers understand different group preferences. Our analysis revealed clusters prioritizing visual design, participation convenience, or incentive mechanisms, enabling targeted customization of urban services and policy mechanisms.

Optimized urban integration mechanisms: Evaluation data helps policymakers identify effective urban services, allowing refinements based on user clusters. Data-driven insights guide adjustments to pacing and rewards to maintain participation motivation.

User Engagement and Retention: Data-driven clustering helps identify attractive urban elements for different users,

enabling policymakers to prioritize these features. Evaluation data can reveal reasons for youth departure, such as dissatisfaction with service interfaces or process pacing.

4.3 Implications for urban policy

(1) Digital content development strategy: Based on youth preferences revealed in the study, priority should be given to developing three types of digital products: mobile games incorporating Foshan kung fu elements, short video series featuring Foshan historical figures and events, and AR urban exploration applications combining Lingnan characteristics.

(2) Educational integration strategy: Given schools' significant influence on urban identity, recommendations include developing "Foshan History and Culture" school-based curricula, utilizing digital technology for innovative teaching methods, and organizing "Exploring Foshan" practical activities combining online and offline approaches.

(3) Community engagement enhancement: Design "Growing with Foshan" youth urban culture projects, create "Foshan Youth Cultural Passport" to record urban cultural experiences, and conduct "Urban Memory Collection" activities with youth documenting urban changes.

4.4 Limitations and future research

This study has several limitations: sample representativeness issues with uneven distribution across districts, reliance primarily on quantitative methods without qualitative research supplements, and cross-sectional design limitations in tracking long-term identity formation effects.

Future research should expand sample scope with optimized regional distribution, employ multiple methods combining surveys, interviews, focus groups, and experimental research, conduct longitudinal studies tracking identity formation trajectories, and design intervention effect assessments.

5 Conclusion

This study demonstrates the value of data-driven approaches in youth urban integration policy development. Using k-means clustering analysis of user feedback, validated through Silhouette Scores and Davies-Bouldin Index, we identified patterns guiding policy strategies. Evaluations clustered around themes including digital cultural consumption preferences, urban identity dimensions, and user engagement, providing actionable insights for content optimization and feature customization.

The research emphasizes limitations of relying solely on ratings, as evaluation text often provides deeper feedback. By integrating these insights, policymakers can create engaging and effective youth urban integration mechanisms that achieve urban identity enhancement goals.

In conclusion, this research underscores the importance of systematic user feedback analysis in youth urban integration policy development. Findings provide a roadmap for using user evaluations to create more effective, user-centered policies that enhance urban experiences and identity outcomes. Data-driven methods like clustering provide insights into user behavior and preferences, guiding key policy decisions involving service mechanisms, visual design, participation difficulty, and incentives. This approach leads to adaptive, impactful youth integration policies aligned with urban development goals.

The results align closely with game-based user-centered design (UCD) principles, where end-user needs, preferences, and feedback are central to the development process. By analyzing youth evaluations—essentially user feedback—we directly integrate these insights into the policy process, representing a core UCD principle. Thus, our approach can indeed be viewed as game-based UCD in practice.

Future research directions should include expanding sample representativeness, incorporating mixed-methods approaches, conducting longitudinal tracking studies, and implementing intervention effect assessments to achieve more comprehensive and in-depth research outcomes.

Funding

Foshan Social Science Planning Project of 2025: Research on the "Digital Cultural Consumption Trends and Urban Identity of the 'Post-10' Youth Group in Foshan City "(2025-GJ083).

About the Author

Qian Wang, male, Ph.D. in creative media, associate professor, research field: digital media art, educational games, etc.

References

[1] Chen, H., Wang, L., Zhang, M. Digital native youth and urban identity: A comprehensive analysis[J]. Journal of Urban Studies, 2024, 45(3): 234-251.

[2] Gee, J. P. What video games have to teach us about learning and literacy[J]. Educational Technology Research, 2005, 23(2): 45-62.

[3] Hamari, J., Koivisto, J., Sarsa, H. Does gamification work? A literature review of empirical studies on gamification[J]. Proceedings of the Hawaii International Conference on System Sciences, 2014, 47(1): 3025-3034.

[4] Jain, A. K., Murty, M. N., Flynn, P. J. Data clustering: A review[J]. ACM Computing Surveys, 1999, 31(3): 264-323.

[5] Li, X., Liu, Y. Urban identity formation among Generation Z: Evidence from Chinese cities[J]. Urban Affairs Review, 2023, 59(4): 1123-1145.

[6] Lloyd, S. Least squares quantization in PCM[J]. IEEE Transactions on Information Theory, 1982, 28(2): 129-137.

[7] Porter, M. F. An algorithm for suffix stripping[J]. Program, 1980, 14(3): 130-137.

[8] Ramos, J. Using TF-IDF to determine word relevance in document queries[J]. Proceedings of the First International Conference on Machine Learning, 2003, 1(1): 133-142.

[9] Zhang, S., Chen, P., Wu, D. K-means clustering for social media analysis: Applications in youth behavior research[J]. Computational Social Networks, 2024, 11(1): 1-18.

[10] Zhou, T., Yang, K. Digital consumption patterns and city attachment: A data-driven approach[J]. Cities, 2023, 132(1): 104089.

[11] Boyd, D. It's complicated: The social lives of networked teens[J]. Digital Culture and Society, 2014, 12(3): 78-95.

[12] Castells, M. The rise of the network society: The information age[J]. Information Society Studies, 1996, 8(2): 15-42.

[13] de Lange, M., de Waal, M. Owning the city: New media and citizen engagement in urban design[J]. Urban Media Studies, 2013, 7(4): 201-218.

[14] Graham, S., Marvin, S. Splintering urbanism: Networked infrastructures, technological mobilities and the urban condition[J]. Urban Geography, 2001, 22(3): 365-388.

[15] Jenkins, H. Convergence culture: Where old and new media collide[J]. Media Studies Quarterly, 2006, 18(2): 112-135.

[16] Lewicka, M. Place attachment: How far have we come in the last 40 years?[J]. Environmental Psychology Review, 2011, 31(2): 207-230.

[17] Lloyd, R. Neo-bohemia: Art and commerce in the postindustrial city[J]. Urban Cultural Studies, 2006, 15(1): 88-105.

[18] Prensky, M. Digital natives, digital immigrants[J]. Educational Technology, 2001, 9(5): 1-6.

[19] Proshansky, H. M., Fabian, A. K., Kaminoff, R. Place-identity: Physical world socialization of the self[J]. Environmental Psychology Journal, 1983, 3(1): 57-83.

[20] Turkle, S. Alone together: Why we expect more from technology and less from each other[J]. Technology and Society Review, 2011, 28(4): 295-318.

[21] Zukin, S. The cultures of cities[J]. Urban Studies Review, 1995, 32(8): 723-740.