Journal of E-Technology

DLINE JOURNALS

Print ISSN: 0976-3503
Online ISSN: 0976-2930

JET 2026: 17 (4)
https://doi.org/10.6025/jet/2026/17/4/193-216

Social Media Content Strategy: A Multivariate Clustering Analysis of Strategic Archetypes

Hajar Ait Lamkademe
University of Hassan II of Casablanca. Morocco
hajar.aitlamkademe-etu@etu.univh2c.ma

ABSTRACT

As social media ecosystems grow increasingly complex, understanding content strategies requires moving beyond traditional demographic or industry based segmentation toward multidimensional behavioral configurations. This study presents a multivariate clustering analysis of social media content strategies, utilizing a dataset of 20 industry specific strategic profiles. By employing Multiple Correspondence Analysis (MCA), hierarchical clustering, Gower distance clustering, text embedding analysis, and K-modes clustering, we identify latent strategic archetypes based on combinations of platform, content format, primary goals, and audience engagement mechanisms. The analysis reveals four distinct exploratory archetypes: Professional Trust Building, High Energy Product Discovery, Visual Lifestyle and Brand Affinity, and Informational Consideration and Conversion. Notably, the findings demonstrate that strategic similarities frequently transcend conventional industry boundaries, and that structured categorical choices provide stronger strategic differentiation than semantic textual descriptions. Ultimately, this research conceptualizes social media content strategy as a multidimensional configuration, offering a data driven framework for marketers and organizations to design more effective, archetype-aligned communication strategies.

Keywords: Social Media Marketing, Content Strategy, Multivariate Clustering, K-modes Clustering, Strategic Archetypes, Multiple Correspondence Analysis, Behavioral Segmentation, Digital Communication

Received: 5 June 2026, Revised: 13 July 2026, Accepted: 22 July 2026

Copyright: DLINE

1. Introduction

The rapid proliferation of social media platforms has fundamentally transformed how individuals communicate, exchange information, and form online communities. Platforms such as X (formerly Twitter), discussion forums such as Reddit, and more specialized platforms such as the now-defunct Voat have contributed to an unprecedented expansion in both the scale and complexity of online interactions. Voat, in particular, became notable for exploring censorship-free communication models among specific political groups [1]. As social interactions increasingly shift toward digitally mediated environments, understanding user behavior and patterns of interaction has become an important research area within computational social science [2, 3].

The analysis of social media behavior provides valuable insights into how users interact with one another, how information and influence propagate through online networks, and how communities emerge, evolve, and

dissolve [4]. Such insights extend beyond theoretical contributions to computational social science and have important practical implications. In particular, understanding patterns of online behavior can support the design of more effective social media platforms, improve user experiences, and contribute to the identification and mitigation of challenges such as misinformation and online harassment [5, 6]. Consequently, social media data have increasingly become an important source for examining behavioral, communicative, and strategic patterns across diverse domains.

Within this broader context, social media marketing has emerged as an important area in which behavioral analysis and content strategy intersect. Social media marketing content strategies have implications across multiple domains, reflecting the growing importance of digital platforms in shaping communication, consumer engagement, and organizational practices [7-10].

The emergence of social media as a defining component of the contemporary "network society" has also substantially altered political communication. Traditional approaches to electoral segmentation have historically relied heavily on demographic characteristics. However, the growing complexity of digitally mediated political communication suggests that demographic segmentation alone may be insufficient to capture the diversity of contemporary audiences. Behavioral data generated through social media interactions provide an opportunity to identify more nuanced and precise audience profiles. In this context, clustering techniques offer a quantitative means of identifying naturally occurring behavioral groups and developing more refined electoral and communication profiles.

Building on the principles of benefit based segmentation and micro-segmentation previous research has adopted quantitative-deductive approaches to examine behavioral patterns within political communication. For example, an AutoCluster framework was applied to Instagram data generated by Brazilian governors to examine patterns associated with engagement, communication frequency, and communication structure [11]. Such approaches demonstrate the potential of unsupervised clustering to move beyond conventional demographic segmentation and identify behavioral patterns that may otherwise remain hidden within large scale social-media datasets.

2. Background

2.1 Social Network Roles and Behavioral Archetypes

Identifying roles within social networks is fundamental to understanding how complex networks operate and how different participants contribute to their structure and dynamics [11]. In traditional social network analysis, researchers commonly identify roles through patterns of connectivity among individuals or entities. These connectivity patterns provide insights into influence, responsibilities, and the formation of groups within networks [12]. Role identification has consequently been applied to a broad range of problems, including the characterization of political polarization [13], the analysis of user behavior [14], the investigation of biological networks [15], and the study of animal populations [16]. In this context, identifying influential users represents another important application of role-based social network analysis, with a comprehensive review of related approaches presented in [12].

The concept of user roles is particularly relevant to social media content strategy because users do not contribute to online environments in the same way. Some users may generate highly engaging content, whereas others may primarily interact with existing content, facilitate information diffusion, or occupy structurally influential positions. Consequently, examining users according to behavioral characteristics rather than relying exclusively on conventional demographic attributes can provide a more detailed understanding of social-media ecosystems.

This perspective also links network analysis to strategic content classification. Behavioral archetypes can potentially capture combinations of characteristics such as engagement, communication frequency, content production, and interaction patterns. Rather than treating individual metrics independently, clustering approaches can identify groups of users or communication strategies that exhibit similar multidimensional

behavioral profiles. Such archetypes can subsequently provide a basis for interpreting differences in influence, engagement, and communication effectiveness.

2.2 Multivariate Clustering for Strategic Segmentation

The increasing availability of multidimensional social media data provides opportunities to move from single-variable analysis toward multivariate approaches capable of capturing complex behavioral structures. Clustering is well suited to this purpose because it lets researchers identify groups in the data without predefined class labels. When applied to social-media behavior, clustering can reveal latent groups that share similar patterns across multiple dimensions, thereby providing a data driven basis for defining strategic archetypes.

This approach is particularly relevant in political communication and social media marketing, where individuals and communication strategies may differ simultaneously in engagement, frequency, interaction, and structural characteristics. The application of clustering to behavioral data therefore offers the potential to identify more precise audience and communication profiles than approaches based exclusively on demographic segmentation.

Previous work demonstrates the broader applicability of machine learning based clustering for identifying interpretable behavioral groups. Munshi, for example, investigated the intersection of deep learning and behavioral science to identify consumer personality archetypes from structured marketing data. Using a real-world dataset containing demographic, transactional, and campaign response variables, the study identified psychologically interpretable behavioral clusters and subsequently predicted consumer responsiveness to marketing interventions. The methodology integrated Deep Embedded Clustering (DEC) for unsupervised segmentation with a Dense Neural Network (DNN) for supervised prediction, with both components supported by SHAP based interpretability [17].

This combination of unsupervised segmentation, supervised prediction, and explainability illustrates an increasingly important direction in behavioral analytics. Clustering can provide the initial identification of naturally occurring groups, while predictive models can subsequently assess whether the discovered behavioral patterns are associated with meaningful outcomes. Explainability methods can further help determine which characteristics contribute most strongly to differences between identified archetypes.

2.3 Influential User Archetypes and Dynamic Behavioral Patterns

Identifying influential user archetypes provides another important perspective on social media ecosystems. Chen, [18] identified three distinctive categories of socially influential users: touch and go influential users, proactive influential users, and super influential users. These archetypes demonstrate that influence is not necessarily represented by a single measure but may instead emerge through different combinations of activity, engagement, and content related behaviors.

Furthermore, the behavioral characteristics of influential users are not necessarily static. Feedback variables can affect transitions between influential user archetypes, with changes observed in areas such as content quality, interaction levels, and communication speed. Some users may shift toward producing higher quality content and increasing their interactions, whereas others may move toward lower quality content and reduced interaction while simultaneously improving the speed of information production. Other users may experience mixed effects because of differences in information processing patterns. These findings indicate that behavioral archetypes should not necessarily be viewed as fixed categories; rather, they may represent dynamic states within an evolving social media environment.

The dynamic nature of user behavior reinforces the importance of multivariate approaches. Individual measures such as engagement, posting frequency, or influence may provide only partial representations of user behavior. In contrast, clustering multiple behavioral indicators simultaneously can capture more comprehensive profiles and reveal patterns that emerge from the interaction of several variables.

2.4 Higher-Order Interactions and the Limitations of Conventional Network Analysis

Although conventional social network analysis provides valuable tools for identifying user roles and influential

actors, extending role identification to hyper network environments introduces additional methodological challenges. Unlike traditional networks, hyper networks represent interactions through hyperedges that can connect multiple nodes simultaneously. This higher order representation provides a more flexible framework for modeling complex real world interactions [19, 20].

Hypernetworks' ability to represent higher order interactions is particularly relevant to social media environments, where communication events may involve multiple users, content items, communities, and interaction types simultaneously. However, the increased representational complexity also creates challenges for identifying meaningful roles. Traditional centrality measures and network analysis methodologies, including conventional community detection algorithms, may not adequately capture the multidimensional structure of hyperedges [21, 22].

This limitation suggests that understanding social media behavior requires analytical approaches capable of integrating multiple behavioral and structural dimensions. Multivariate clustering offers one way to address this challenge by focusing on patterns across several characteristics rather than relying solely on conventional network connectivity measures. When combined with interpretable machine-learning techniques, clustering can provide a systematic framework for identifying and characterizing strategic archetypes within complex digital environments.

2.5 Research Motivation and Analytical Perspective

Taken together, the existing literature demonstrates the growing importance of behavioral segmentation, social network role identification, influential user classification, and machine learning based clustering in understanding social media environments. Research has established that demographic segmentation alone may not adequately represent the complexity of digitally mediated audiences, while behavioral clustering can uncover more nuanced user profiles. Similarly, studies of influential users demonstrate that distinct archetypes can emerge from differences in engagement, content production, and interaction patterns.

However, the increasing complexity of social media interactions also highlights the limitations of approaches that examine individual behavioral variables or conventional network structures in isolation. The multidimensional and potentially dynamic nature of online behavior requires analytical frameworks capable of identifying patterns across multiple characteristics while retaining sufficient interpretability to support strategic decision making.

Against this background, a multivariate clustering perspective provides a suitable foundation for examining social media content strategies as combinations of behavioral characteristics rather than as isolated metrics. Such an approach enables the identification of distinct strategic archetypes based on similarities and differences across relevant engagement, frequency, interaction, and communication dimensions. The resulting clusters can provide a more systematic representation of heterogeneous user and content strategy profiles and can potentially support applications in social media marketing, political communication, audience segmentation, and influence analysis.

Overall, the literature establishes a strong conceptual basis for investigating social media behavior through data driven clustering. The integration of behavioral segmentation, network role analysis, influential-user archetypes, and interpretable machine learning provides a foundation for developing a multivariate framework capable of identifying meaningful strategic patterns within complex social media environments.

3. Methodology

3.1 Research Problem and Research Statement

Social media content strategy is inherently multidimensional. Organizations simultaneously select platforms, content formats, communication goals, audience orientations, visual styles, publishing cadences, and calls to action. Traditional analyses that examine these elements in isolation, or that rely primarily on industry classification, risk overlooking the combinatorial patterns that actually define strategic practice. Demographic or sector based segmentation alone is therefore insufficient for identifying meaningful strategy profiles.

The research problem addressed in this study is the absence of a systematic, data driven framework for identifying latent strategic archetypes from the joint configuration of categorical content strategy variables. Without such a framework, practitioners and researchers lack a transparent basis for grouping strategies that share similar structural choices even when they operate in different industries.

Accordingly, the research statement of this study is as follows: *By applying complementary multivariate clustering techniques to a structured dataset of social media content strategies, it is possible to identify and interpret a small number of recurring strategic archetypes that cut across conventional industry boundaries and that are defined by coherent combinations of platform, format, goal, audience orientation, visual direction, frequency, and call to action.*

3.2 Research Objectives

The study pursues four interrelated objectives:

[1] To describe the multidimensional structure of social media content strategies across a controlled set of industry cases.

[2] To apply exploratory multivariate techniques (Multiple Correspondence Analysis, hierarchical clustering, Gower distance clustering, K-modes clustering, and text embedding clustering) in order to uncover latent groupings among strategy profiles.

[3] To interpret the resulting clusters as distinct strategic archetypes grounded in modal categorical prototypes and membership composition.

[4] To assess the degree to which strategic similarity transcends industry labels and to evaluate the complementary insights provided by categorical versus textual representations of strategy.

3.3 Research Design

The study adopts an exploratory, quantitative deductive research design grounded in unsupervised machine learning. The design is exploratory because the goal is to discover latent structure rather than to test pre-specified hypotheses about cluster membership. It is quantitative because all primary variables are treated as categorical observations that are subjected to formal distance based and prototype based clustering procedures. The design is deductive in the limited sense that the choice of variables and the interpretation of archetypes are informed by prior literature on behavioral segmentation, network roles, and influential user typologies, yet the cluster solutions themselves are data driven.

No experimental manipulation or random assignment is employed. The analysis is observational and cross-sectional, relying on a single synthetic dataset that encodes complete strategy profiles. Because the sample size is small ($n = 20$) and the data are synthetic, the design prioritizes interpretability and methodological triangulation over statistical generalizability. Multiple complementary clustering methods are therefore applied to the same set of observations so that patterns that recur across methods can be distinguished from method-specific artifacts.

3.4 Dataset Description

The dataset utilized in this study is the “Social Media Content Strategy Dataset” (Dereli, 2026), available on Kaggle. It comprises 20 synthetically generated records designed to model structured content plans for social media strategy. Each record provides a comprehensive profile for a content initiative, linking key variables such as industry, target audience, and platform with strategic elements including content pillar, format, primary goal, content angle, visual direction, publishing frequency, and a recommended call to action. The data is manually curated and intended to support research in content planning, campaign development, and AI-assisted creative workflows. While its synthetic nature limits real-world generalizability, it offers a controlled and systematic framework for analyzing strategic content relationships. The dataset is publicly accessible and is expected to be updated quarterly.

The principal categorical variables retained for analysis are: industry, platform, content pillar, content format, primary goal, target audience, content angle, visual direction, publishing frequency, and call to action. These variables jointly define the strategic configuration of each observation and constitute the input space for all subsequent multivariate procedures.

3.5 Data Processing and Preparation

Prior to analysis, the dataset underwent a sequence of preparation steps designed to ensure consistency and suitability for categorical multivariate techniques.

First, variable selection was performed. Only substantive strategic variables were retained; identifiers and purely administrative fields were excluded. All retained variables were confirmed to be nominal categorical. Second, a complete-case approach was adopted. Because the synthetic dataset contains no missing values, no imputation was required. Each of the 20 observations therefore contributes a full strategy profile.

Third, for Multiple Correspondence Analysis and distance based clustering, the categorical variables were represented in complete disjunctive form (one hot encoding). This transformation converts each categorical variable into a set of binary indicator columns, enabling the computation of Euclidean geometry on the resulting matrix for MCA and the calculation of mismatch based distances for hierarchical and Gower analyses.

Fourth, for K-modes clustering, the original nominal coding was retained. K-modes operates directly on categorical labels and updates cluster prototypes by selecting the modal category of each variable within a cluster; consequently, one hot encoding is unnecessary and would be counterproductive.

Fifth, for the text embedding analysis, four descriptive fields target audience, content angle, visual direction, and call to action were concatenated into a single strategy description document for each observation. These documents were then transformed into TF-IDF vectors using unigram and bigram features, yielding a sparse lexical representation suitable for cosine distance clustering.

Finally, all distance matrices and clustering solutions were computed on the full set of 20 observations. Silhouette coefficients were calculated for candidate solutions ranging from 2 to 5 clusters in order to guide model selection while remaining transparent about the modest separation present in the data.

3.6 Analytical Methods

Five complementary analytical procedures were applied.

3.6.1 Multiple Correspondence Analysis (MCA)

MCA was approximated through singular value decomposition of the complete disjunctive (one-hot encoded) matrix. The first two dimensions provide an exploratory map of similarity among complete social-media strategy profiles. The two-dimensional representation is intended for pattern exploration because the sample is small and category combinations are sparse.

3.6.2 Hierarchical Clustering

Average linkage hierarchical clustering was applied to Hamming distances among categorical strategy profiles. The silhouette based selection procedure evaluated 2–5 clusters. The selected solution contained 5 clusters with a silhouette score of 0.114.

3.6.3 K-modes Clustering

K-modes was used because the principal variables are nominal categories. Cluster prototypes are represented by modal category values rather than arithmetic means. The selected solution contained 4 clusters with a silhouette score of 0.095. This four-cluster solution forms the primary basis for the subsequent archetype interpretation.

3.6.4 Gower-distance Clustering

Gower style dissimilarity was calculated as the proportion of mismatched variables. In this dataset, all substantive variables are categorical, so the resulting distance is equivalent to an equally weighted categorical mismatch measure. The selected solution contained 5 clusters with a silhouette score of 0.114.

3.6.5 Text-embedding Clustering

Target audience, content angle, visual direction, and call-to-action fields were combined into a strategy text document for each observation. TF-IDF unigram/bigram representations were used as document embeddings, and cosine distance was used for hierarchical clustering. The selected solution contained 5 clusters with a silhouette score of 0.025.

3.7 Methodological Interpretation and Limitations

The five methods are treated as complementary exploratory approaches. Hierarchical, K-modes, and Gower analyses capture similarity in structured categorical strategy choices, whereas text embedding clustering captures semantic and lexical similarity in descriptive strategy language. Convergence across the categorical methods (particularly the recurring five-cluster solutions of hierarchical and Gower analyses and the interpretable four cluster K-modes solution) provides greater confidence that the recovered groupings reflect genuine combinatorial patterns rather than artifacts of a single algorithm.

Given the small sample size (n = 20) and the synthetic origin of the data, the results are interpreted strictly as exploratory strategy archetypes rather than population level statistical estimates. Silhouette coefficients are modest across all methods, indicating that the clusters are overlapping rather than sharply separated. This limited separation is itself substantively informative: it suggests that social media strategies share common elements across industries while still differing in other dimensions. The design therefore prioritizes transparent description of recurring configurations over claims of discrete, mutually exclusive strategy types.

4. Results and Discussion

4.1 Overview of Social Media Content Strategies

Table 1 presents the complete social media content strategy dataset comprising 20 observations across 20 industries. Each observation represents a distinct industry level content strategy characterized by platform, content pillar, content format, primary marketing goal, target audience, content angle, visual direction, publishing frequency, and call to action (CTA). The dataset therefore provides a multidimensional representation of how organizations combine platform selection, content design, communication objectives, and audience engagement mechanisms.

The descriptive patterns shown in Table 1 indicate substantial diversity in social media strategy across industries. Instagram is the most frequently represented platform, particularly among fashion, food and beverage, fitness, travel, cultural events, beauty, education, real estate, hospitality, and e-commerce strategies. Within these observations, visually oriented formats, particularly Reels and Carousels, are prominent. By contrast, LinkedIn is used for finance and creative agency strategies, where document posts, thought leadership, case studies, and trust or lead oriented objectives are emphasized. TikTok is associated with music and consumer technology, reflecting the use of short form video for product interest and audience attention, whereas Facebook is represented in healthcare and nonprofit strategies. Pinterest appears in the interior design strategy, where project showcases and visual portfolio development are particularly relevant.

Table 1 also shows that content format closely aligns with strategic objectives. Reels and short videos are generally associated with discovery, engagement, entertainment, and audience interest, while Carousels and document posts are more frequently associated with education, information provision, consideration, and trust-building. Publishing frequency also varies by communication strategy. Highly visual, engagement-oriented sectors commonly publish two or three times per week, whereas information intensive or trust-oriented strategies more often use weekly schedules. The CTAs also correspond closely with the underlying marketing objective, ranging from exploratory actions such as “Explore collection” and “Plan your trip” to direct conversion oriented actions such as “Book consultation,” “Request details,” “Shop now,” and “Get tickets.”

id	industry	platform	content pillar	content format	primary goal	target audience	content angle	visual direction	publishing frequency	call to action
1	Technology	Instagram	Product Education	Carousel	Feature awareness	Small business owners	Show one workflow from problem to result	Clean UI cards electric blue highlights	Weekly	Try the platform
2	Fashion	Instagram	Collection Launch	Reel	Drive product discovery	Urban adults 20-35	Show styling transitions for one key piece	Editorial street style bold typography	3 times weekly	Explore collection
3	Hospitality	Instagram	Experience	Carousel	Increase direct bookings	Design led travelers	Highlight one complete guest journey	Sunlit interiors Mediterrn -ean details	Weekly	Book your stay
4	Food and Beverage	Instagram	Product Story	Reel	Build brand affinity	Coffee enthusiasts	Show the craft behind one signature drink	Macro textures warm natural light	3 times weekly	Visit us
5	Music	TikTok	Release Promotion	Short video	Increase listens	Electronic music audience	Tease the drop with visual mood and hook	Nocturnal cinematic motion graphics	3 times weekly	Listen now
6	Automotive	YouTube	Performance	Short video	Generate model interest	Drivers 25-45	Show one driving feature in real conditions	Controlled reflections night motion	Weekly	Discover perform -ance
7	Beauty	Instagram	Routine Education	Carousel	Build product consider -ation	Beauty consumers 20-40	Explain a simple three-step routine	Soft studio light translucent textures	2 times weekly	Discover routine
8	Real Estate	Instagram	Property Feature	Carousel	Generate qualified leads	Property investors	Show a key lifestyle benefit per slide	Golden-hour architecture clean layouts	Weekly	Request details
9	Finance	LinkedIn	Thought Leadership	Document post	Build trust	Young professionals	Explain one practical money habit	Clear data visuals trust blue palette	Weekly	Learn more
10	Education	Instagram	Learning Value	Carousel	Drive course interest	Emerging creatives	Share five actionable creative tips	Modular graphic system bright accents	2 times weekly	View courses
11	Fitness	Instagram	Training Tips	Reel	Increase member ships	Health-focused adults	Demonstrate one focused training movement	High-energy motion strong type	3 times weekly	Join now

id	industry	platform	content pillar	content format	primary goal	target audience	content angle	visual direction	publishing frequency	call to action
12	Travel	Instagram	Destination Inspiration	Reel	Build intent Increase	Experience led travelers	Show a destin-ation from morning to night	Wide scenes natural color grading	2 times weekly	Plan your trip
13	Interior Design	Pinterest	Project Showcase	Pin	portfolio saves	Home owners 30-55	Show materials and finished space together	Calm architecture tactile materials	Weekly	View project
14	Cultural Festival	Instagram	Artist Spotlight	Carousel	Increase ticket sales	Music and arts comm -unity	Introduce one artist with key schedule details	Layered poster graphics vivid accents	3 times weekly	Get tickets
15	Healthcare	Facebook	Service Education	Carousel	Increase consulta -tions	Adults seeking care	Answer one common service question	Bright human-centered visual system	Weekly	Book consulta -tion
16	Restaurant	Instagram	Menu Story	Reel	Increase reservations	Food lovers	Show a dish from preparation to table	Editorial food light warm tones	3 times weekly	Reserve a table
17	E-commerce	Instagram	Product Discovery	Carousel	Increase conversions	Online shoppers	Build a themed product edit	Product-led layout clear offer badges	3 times weekly	Shop now
18	Nonprofit	Facebook	Community Impact	Photo post	Increase donations	Local supporters	Show a real out come from one contribution	Authentic docume -ntary imagery	Weekly	Support the cause
19	Creative Agency	Linked In	Case Study	Document post	Generate leads	Brand decision makers	Show challenge strategy and outcome	Minimal system led presentation	Weekly	Start a project
20	Consumer Tech	Tik Tok	Use Case	Short video	Increase product	Early adopters	Show a relatable everyday product moment	Crisp product shots interface overlays	3 times weekly	See how it works

Table 1. Social Media Content Strategy

Overall, Table 1 shows that social media strategies cannot be understood solely by industry classification. Instead, strategic profiles emerge from combinations of platform, content format, communication objective, audience orientation, visual presentation, and intended user action. This multidimensional structure underpins the subsequent MCA and clustering analyses.

id industry- platform- content_pillar content_format- primary_goal target_audience- content_angle visual_direction- publishing_frequency- call_to_action-

1. Technology Instagram Product Education Carousel Feature awareness Small business owners Show one workflow from problem to result Clean UI cards electric blue highlights Weekly Try the platform

2. Fashion Instagram Collection Launch Reel Drive product discovery Urban adults 20-35 Show styling transitions for one key piece Editorial street style bold typography 3 times weekly Explore collection

3. Hospitality Instagram Experience Carousel Increase direct bookings Design led travelers Highlight one complete guest journey Sunlit interiors Mediterranean details Weekly Book your stay

4. Food and Beverage Instagram Product Story Reel Build brand affinity Coffee enthusiasts Show the craft behind one signature drink Macro textures warm natural light 3 times weekly Visit us

5. Music TikTok Release Promotion Short video Increase listens Electronic music audience Tease the drop with visual mood and hook Nocturnal cinematic motion graphics 3 times weekly Listen now

6. Automotive YouTube Performance Short video Generate model interest Drivers 25-45 Show one driving feature in real conditions Controlled reflections night motion Weekly Discover performance

7. Beauty Instagram Routine Education Carousel Build product consideration Beauty consumers 20-40 Explain a simple three-step routine Soft studio light translucent textures 2 times weekly Discover routine

8. Real Estate Instagram Property Feature Carousel Generate qualified leads Property investors Show a key lifestyle benefit per slide Golden-hour architecture clean layouts Weekly Request details

9. Finance LinkedIn Thought Leadership Document post Build trust Young professionals Explain one practical money habit Clear data visuals trust blue palette Weekly Learn more

10. Education Instagram Learning Value Carousel Drive course interest Emerging creatives Share five actionable creative tips Modular graphic system bright accents 2 times weekly View courses

11. Fitness Instagram Training Tips Reel Increase memberships Health focused adults Demonstrate one focused training movement High-energy motion strong type 3 times weekly Join now

12. Travel Instagram Destination Inspiration Reel Build intent Experience-led travelers Show a destination from morning to night Wide scenes natural color grading 2 times weekly Plan your trip

13. Interior Design Pinterest Project Showcase Pin Increase portfolio saves Homeowners 30-55 Show materials and finished space together Calm architecture tactile materials Weekly View project

14. Cultural Festival Instagram Artist Spotlight Carousel Increase ticket sales Music and arts community Introduce one artist with key schedule details Layered poster graphics vivid accents 3 times weekly Get tickets

15. Healthcare Facebook Service Education Carousel Increase consultations Adults seeking care Answer one common service question Bright human centered visual system Weekly Book consultation

16. Restaurant Instagram Menu Story Reel Increase reservations Food lovers Show a dish from preparation to table Editorial food light warm tones 3 times weekly Reserve a table.

17. E-commerce Instagram Product Discovery Carousel Increase conversions Online shoppers Build a themed product edit Product-led layout clear offer badges 3 times weekly Shop now

18. Nonprofit Facebook Community Impact Photo post Increase donations Local supporters Show a real outcome from one contribution Authentic documentary imagery Weekly Support the cause

19. Creative Agency LinkedIn Case Study Document post Generate leads Brand decision makers Show challenge strategy and outcome Minimal system-led presentation Weekly Start a project

20. Consumer Tech TikTok Use Case Short video Increase product interest Early adopters Show a relatable everyday product moment Crisp product shots interface overlays 3 times weekly See how it works.

4.2 Multiple Correspondence Analysis of Strategy Profiles

Figure 1 presents the Multiple Correspondence Analysis (MCA) representation of the social media content strategies. MCA was used to provide an exploratory visualization of similarity among the categorical strategy profiles. The analysis used the complete disjunctive representation of the categorical variables, allowing observations with similar combinations of industry, platform, content format, strategic goal, publishing frequency, and CTA to be positioned closer together in the two dimensional space.

The MCA plot demonstrates that the observations occupy a dispersed multidimensional strategy space, indicating substantial heterogeneity in the dataset. This pattern is consistent with the descriptive evidence in Table 1, where industries differ considerably in their platform choices, formats, goals, and engagement mechanisms. Strategies that employ similar combinations of categorical characteristics tend to appear closer together, whereas observations with more distinctive platform-content combinations are positioned farther apart.

The broad dispersion of observations also supports exploratory clustering techniques. Rather than suggesting a single dominant social media strategy, Figure 1 indicates multiple strategic orientations. The variation visible in the MCA representation is consistent with the later identification of trust oriented, product-discovery-oriented, lifestyle oriented, and consideration oriented strategy archetypes.

Because the dataset contains only 20 observations and a relatively large number of categorical combinations, the MCA results should be interpreted primarily as an exploratory visualization rather than as evidence of statistically stable population level dimensions. Nevertheless, Figure 1 provides an important overview of the dataset's structure by showing that social media strategies differ by multidimensional combinations rather than by individual variables in isolation.

4.3 Hierarchical Clustering of Categorical Strategy Profiles

Figure 2 presents the hierarchical clustering dendrogram derived from categorical strategy profiles. We applied average linkage hierarchical clustering to Hamming distances, which measure mismatch between observations across the categorical strategy variables. The hierarchical procedure identified a five cluster exploratory solution, with a silhouette coefficient of 0.114.

The dendrogram shows varying degrees of similarity among the strategy profiles. Observations that merge at lower linkage distances represent strategies with similar combinations of categorical characteristics, whereas observations that join at higher distances represent more distinct strategic configurations. The hierarchical structure is therefore useful for examining whether similarities occur gradually across observations rather than assuming that all strategies belong to completely independent groups.

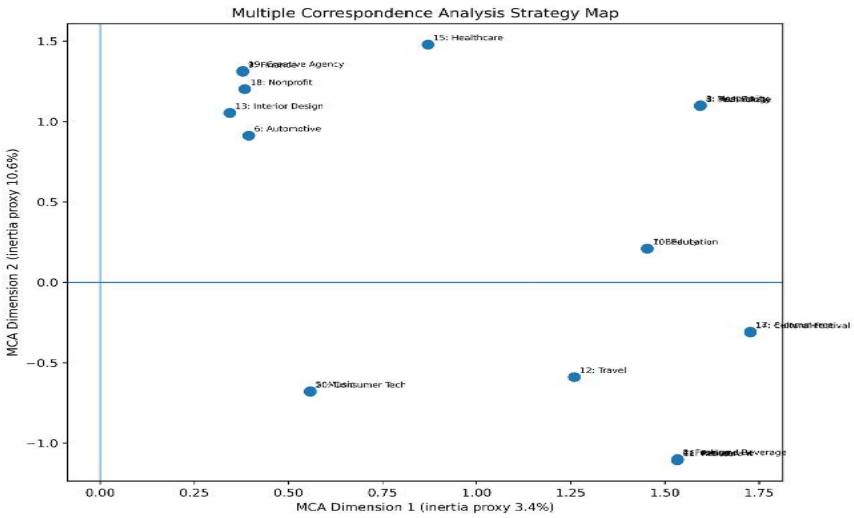

Figure 1. Multiple Correspondence Analysis of Social Media Content Strategy Profiles

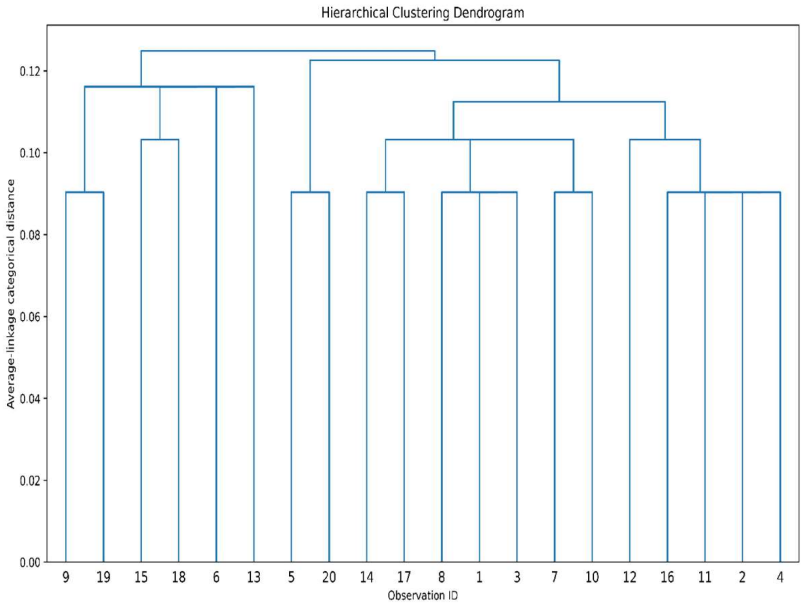

Figure 2. Hierarchical Clustering Dendrogram of Social Media Content Strategies

The relatively modest silhouette value indicates that the clusters are not sharply separated. This finding is consistent with the diversity shown in Table 1 and the dispersed structure of the MCA representation in Figure

1. In practical terms, the strategies share some common characteristics across industries while still differing in other dimensions. For example, several strategies may use Instagram but differ in content format, marketing goal, target audience, or CTA.

The hierarchical analysis therefore suggests that the social media strategies form overlapping groups rather than distinct, homogeneous clusters. This result supports interpreting the subsequent clusters as exploratory strategy archetypes. The hierarchical dendrogram is particularly valuable because it reveals the nested structure of similarity and shows that some observations can be regarded as more closely related than others even when a single fixed cluster solution is not imposed.

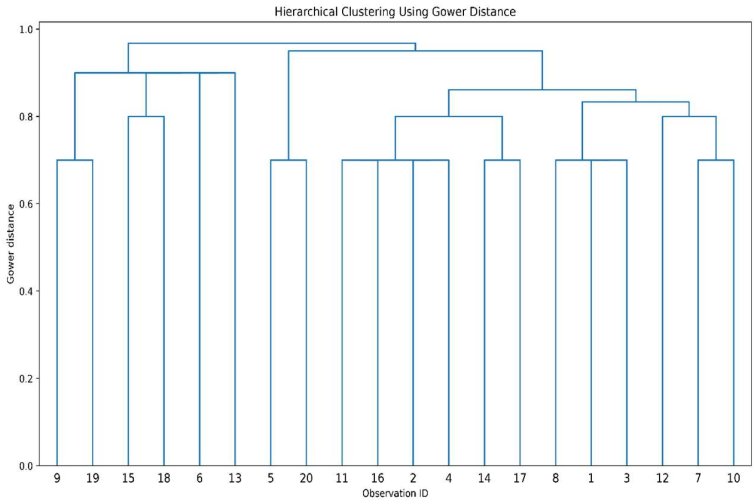

Figure 3. Gower Distance-Based Clustering of Social Media Content Strategies

4.4 Gower Distance-Based Similarity Structure

Figure 3 presents the clustering structure obtained using Gower style dissimilarity. We calculated Gower distance as the proportion of mismatched categorical characteristics between observations. Because all substantive variables in the present dataset are categorical, the Gower style measure functions as an equally weighted categorical mismatch distance. The resulting clustering procedure also selected a five-cluster solution, with a silhouette coefficient of 0.114.

The similarity between the Gower-distance and hierarchical clustering results is methodologically meaningful. Both approaches compare the structured categorical strategy variables and indicate a five-group exploratory structure. The corresponding silhouette scores suggest that the dataset contains identifiable patterns of similarity but that the boundaries between the groups remain weak.

Figure 3 therefore reinforces the conclusion that the strategies are heterogeneous and partially overlapping. The result also shows that the observed structure does not depend on any single clustering representation. Instead, the categorical profiles consistently reveal multiple exploratory groupings when similarity is measured through variable-level mismatches.

From a strategic perspective, the Gower distance structure supports the idea that organizations may share social media characteristics across conventional industry boundaries. A strategy can be similar to another strategy because of shared platform choice, content format, publication frequency, or communication objective even when the organizations operate in different sectors. This finding is important because it suggests that social media strategy archetypes may be defined more effectively by strategic configuration than by industry alone.

4.5 Text-Embedding Clustering of Descriptive Strategy Characteristics

Figure 4 presents the text-embedding clustering results. For this analysis, we combined the target audience, content angle, visual direction, and CTA fields into strategy descriptions. TF-IDF representations based on unigram and bigram features were then used to represent the textual characteristics of each observation, while cosine distance was used to evaluate similarity. The analysis selected a five-cluster solution with a silhouette coefficient of 0.025.

The much lower silhouette coefficient than in the categorical clustering approaches indicates that the descriptive textual features are weakly separated. This suggests that although industries use different

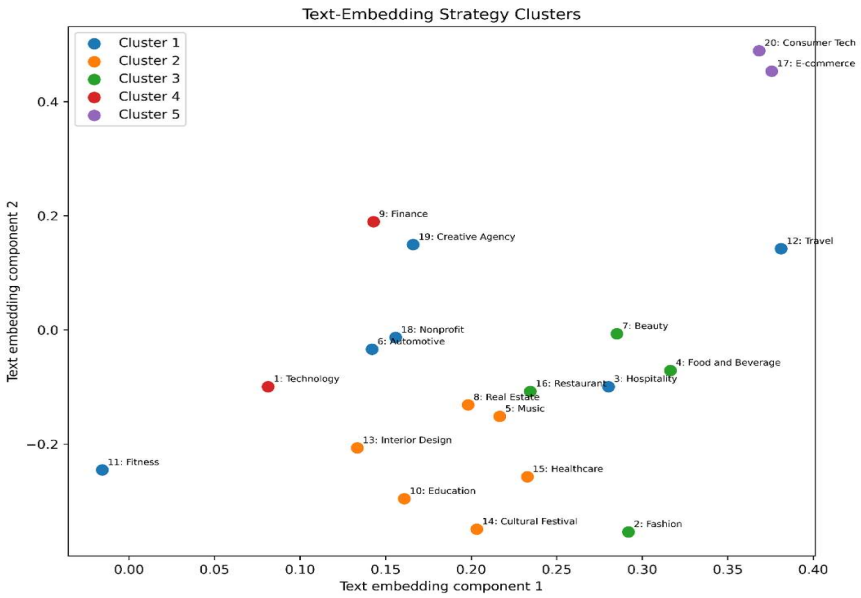

Figure 4. Text-Embedding Clusters of Social Media Content Strategies

wording and visual descriptions, the textual strategy descriptions overlap considerably. Common concepts such as product discovery, audience engagement, explanation, experience, visual presentation, and action-oriented CTAs span multiple industries.

The weak separation observed in Figure 4 should therefore not be interpreted as a failure of the textual analysis. Instead, it highlights a substantive characteristic of the dataset: the language used to describe social media strategies is more continuous and overlapping than the structured categorical choices. While a platform or content format represents a discrete strategic decision, textual descriptions can share vocabulary and meaning across otherwise different industries.

When interpreted alongside Figures 1–3, Figure 4 indicates that structured categorical features provide somewhat stronger differentiation than textual descriptions. The combination of methods is therefore useful because the categorical clustering captures explicit strategic choices, whereas text embedding captures similarities in the language used to describe audience targeting, content storytelling, visual design, and calls to action.

4.6 K-Modes Clustering Performance

Table 2 summarizes the performance and characteristics of the K-modes clustering analysis. K-modes was selected because the dataset consists predominantly of nominal categorical variables. Unlike K-means clustering, which relies on numerical means, K-modes uses categorical mismatch to assign observations and represents each cluster by its modal category values. The analysis used all 20 observations and selected a four cluster solution with a silhouette coefficient of 0.095.

The silhouette coefficient indicates substantial overlap among the social media strategy profiles. Although the four clusters capture recurring combinations of platform, format, objective, and audience orientation, the observations do not form sharply separated groups. This result is consistent with the hierarchical and Gower-distance analyses, both of which also produced relatively low silhouette coefficients.

Despite the modest separation, the K-modes solution provides a useful balance between interpretability and strategic differentiation. The four-cluster solution reduces the heterogeneous set of 20 observations into a manageable number of recurring strategy archetypes. The resulting clusters are subsequently interpreted using their membership composition and modal prototypes.

Overall clustering performance

Measure	Result
Number of observations	20
Number of clusters	4
Silhouette score	0.095
Distance principle	Categorical mismatch
Cluster representative	Modal category prototype

Table 2. K-Means Clustering Performance

The relatively low silhouette score indicates substantial overlap among the social-media strategy profiles. Therefore, interpret the clusters as exploratory strategy archetypes rather than sharply separated groups.

4.7 Distribution of Industries Across K-Modes Clusters

Table 3 presents the distribution of observations across the four K-modes clusters. Cluster 3 is the largest group, containing seven observations, followed by Cluster 4 with six observations. Cluster 1 contains four observations, whereas Cluster 2 is the smallest cluster with three observations.

Cluster 1 includes Interior Design, Finance, Creative Agency, and Nonprofit. Despite representing diverse sectors, these industries share strategic characteristics that distinguish them from the other groups. Their inclusion within the same cluster suggests a common orientation toward credibility, information, professional communication, or trust building.

Cluster 2 consists of Automotive, Consumer Technology, and Music. These industries are united by strategies emphasizing dynamic communication, product or performance interest, and visually engaging short form content. Although YouTube represents Automotive in the underlying dataset, while Consumer Technology and Music use TikTok, the commonality lies in the broader use of dynamic visual content to stimulate audience interest.

Cluster 3 contains Fitness, Restaurant, Cultural Festival, Fashion, Food and Beverage, E-commerce, and Travel. This cluster is the largest and reflects industries with strong visual, experiential, lifestyle, and consumer-engagement characteristics. Most observations within this group rely heavily on Instagram and visually immersive content formats.

Cluster 4 includes Technology, Hospitality, Healthcare, Education, Beauty, and Real Estate. These industries represent strategies with a stronger informational or consideration oriented component. The industries frequently require audiences to understand a service, product, process, or benefit before taking action, which is reflected in the greater importance of explanatory and structured content.

The distribution shown in Table 3 demonstrates that the K-modes clusters do not correspond directly to conventional industry sectors. Instead, industries are grouped according to similarities in their broader social media strategy configurations. This is a central finding of the analysis.

4.8 Observation-Level Cluster Assignments

Table 4 provides the observation-level K-modes assignments for all 20 industries. The table offers a detailed view of the cluster structure by identifying the precise industry associated with each strategy archetype.

The assignments show that the clustering process groups industries with different market contexts when they share similar content and communication characteristics. For example, Finance and Creative Agency are grouped with Interior Design and Nonprofit in Cluster 1, suggesting that similarities in strategic communication can outweigh differences in industry classification. Similarly, Fitness, Fashion, Food and Beverage, Travel, Restaurant, E-commerce, and Cultural Festival are grouped in Cluster 3 because they share an emphasis on visual presentation, audience engagement, lifestyle relevance, and discovery oriented communication.

The observation level assignments also highlight that the clusters should not be interpreted as rigid industry categories. Instead, they represent strategic positions within the broader social media content ecosystem. The table therefore provides the empirical basis for developing the four strategy archetypes described below.

Cluster	Number of observations	Industries represented
Cluster 1	4	Interior Design, Finance, Creative Agency, Nonprofit
Cluster 2	3	Automotive, Consumer Tech, Music
Cluster 3	7	Fitness, Restaurant, Cultural Festival, Fashion, Food and Beverage, E-commerce, Travel
Cluster 4	6	Technology, Hospitality, Healthcare, Education, Beauty, Real Estate

Table 3. Industry Membership Across K-Modes Clusters

2. K-modes Cluster Prototypes

4.9 K-Modes Cluster Prototypes

Table 5 presents the modal categorical prototypes for the four K-modes clusters. These prototypes summarize the most representative categorical characteristics of each cluster, including platform, content pillar, content format, primary goal, target audience, publishing frequency, and CTA. Because K-modes relies on modal categories, interpret the prototype as a representative profile rather than a description that applies identically to every observation in the cluster.

Cluster 1 is represented by a LinkedIn oriented strategy based on case studies and document posts, with an emphasis on building trust and communicating with brand decision makers. Its weekly publication frequency and “Learn more” CTA reinforce an informational and credibility oriented approach.

Cluster 2 is characterized by dynamic short form content and a performance oriented communication style. The prototype emphasizes TikTok, short video, frequent publication, and product or model interest. The “Discover performance” CTA further reinforces this strategy’s exploratory, performance driven nature.

ID	Industry	K-modes Cluster
13	Interior Design	1
9	Finance	1
19	Creative Agency	1
18	Nonprofit	1
6	Automotive	2
20	Consumer Tech	2
5	Music	2
11	Fitness	3
16	Restaurant	3
14	Cultural Festival	3
2	Fashion	3
4	Food and Beverage	3
17	E-commerce	3
12	Travel	3
1	Technology	4
3	Hospitality	4
15	Healthcare	4
10	Education	4
7	Beauty	4
8	Real Estate	4

Table 4. Observation-Level K-Modes Cluster Assignments

Cluster 3 is represented by Instagram and Reels, emphasising visual storytelling, brand affinity, and relatively frequent publication. The prototype reflects the importance of immersive, aesthetically engaging content for lifestyle and experience oriented industries.

Cluster 4 is characterized by Instagram Carousels, informational or experience-oriented communication, product consideration, and direct action. The prototype indicates a strategy in which audiences receive

structured information before being encouraged to proceed to a consultation, booking, inquiry, or other conversion related action.

Taken together, Table 5 provides the strongest basis for interpreting the four clusters as distinct exploratory social media strategy archetypes.

Strategy Variable	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 4**
Representative industry	Creative Agency	Automotive	Cultural Festival	Beauty
Platform	LinkedIn	TikTok	Instagram	Instagram
Content pillar	Case Study	Performance	Artist Spotlight	Experience
Content format	Document post	Short video	Reel	Carousel
Primary goal	Build trust	Generate model interest	Build brand affinity	Build product consideration
Target audience	Brand decision makers	Drivers aged 25–45	Coffee enthusiasts	Adults seeking care Weekly
Publishing frequency	Weekly	3 times weekly	3 times weekly	Book
CTA	Learn more	Discover performance	Explore collection	consultation

Table 5. K-Modes Cluster Prototypes

3. Interpretation of the Four Strategy Archetypes

4.10 Cluster 1: Professional Trust-Building Strategy

Cluster 1 comprises Interior Design, Finance, Creative Agency, and Nonprofit strategies. The modal profile emphasizes LinkedIn, case-study-oriented content, document posts, trust-building objectives, brand decision makers, weekly publication, and an informational CTA. The defining characteristic of this cluster is the use of content to establish credibility and demonstrate value through explanation, evidence, outcomes, or professional expertise.

This archetype can therefore be described as a Professional Trust-Building Strategy. Rather than prioritizing rapid entertainment or immediate conversion, the strategy emphasizes authority, credibility, stakeholder confidence, and longer term relationship development. This orientation is consistent with content that explains expertise, demonstrates outcomes, and supports informed audience evaluation.

Cluster 1: Trust- and Authority-Oriented Strategies

Industries: Interior Design, Finance, Creative Agency, and Nonprofit.

The prototype of this cluster is characterized by:

• LinkedIn

• Case-study-oriented content

• Document posts

• Goal of building trust

• Targeting brand decision makers

• Weekly publication

• Informational CTA: “Learn more”

4.11 Cluster 2: High-Energy Product Discovery Strategy

Cluster 2 comprises Automotive, Consumer Technology, and Music. Its prototype emphasizes short-form visual communication, performance-oriented content, audience interest, frequent publication, and discovery-oriented actions.

This cluster is interpreted as a High-Energy Product Discovery Strategy. Its central objective is to capture attention rapidly and stimulate interest through dynamic storytelling, product demonstration, performance visualization, and visually engaging communication. The industries represented in this cluster share a capacity to communicate through movement, sound, product performance, or experiential demonstrations.

The frequent use of short form content indicates that the strategy is particularly suited to environments in which rapid attention and immediate engagement are important. The cluster therefore represents an exploratory and interest generation stage of the social media communication process.

Cluster 2: High-Energy Product and Performance Strategies

Industries: Automotive, Consumer Technology, and Music.

The cluster prototype emphasizes:

• TikTok

• Short video

• A performance-oriented content pillar

• Goal of generating product/model interest

• A relatively high publishing frequency of three times per week

• A discovery-oriented CTA

4.12 Cluster 3: Visual Lifestyle and Brand-Affinity Strategy

Cluster 3 is the largest K-modes group and includes Fitness, Restaurant, Cultural Festival, Fashion, Food and Beverage, E-commerce, and Travel. The prototype emphasizes Instagram, Reels, visual storytelling, brand

affinity, and relatively frequent publication.

This cluster is interpreted as a Visual Lifestyle and Brand Affinity Strategy. The industries represented share strong relationships with visual identity, consumer experience, lifestyle relevance, and emotional engagement. Their social media strategies rely on imagery, movement, storytelling, and aesthetic presentation to develop stronger relationships between audiences and brands.

The primary objective of this archetype extends beyond immediate conversion. Instead, the strategies seek to build emotional connection, stimulate interest, reinforce lifestyle associations, and encourage continued audience engagement. The dominance of visually oriented formats makes this cluster particularly relevant to industries where consumer experience and brand aesthetics are central to communication.

4.13 Cluster 4: Informational Consideration and Conversion Strategy

Cluster 4: Informational Consideration and Service Strategies

Industries: Technology, Hospitality, Healthcare, Education, Beauty, and Real Estate.

The prototype is characterized by:

- Instagram
- Carousel content
- Experience-oriented content
- Goal of building product consideration
- Informational content answering common questions
- Weekly publication
- A direct CTA such as “Book consultation”

Cluster 4 consists of Technology, Hospitality, Healthcare, Education, Beauty, and Real Estate. Its prototype emphasizes Instagram Carousels, structured information, product or service consideration, weekly publication, and direct calls to action.

This archetype is interpreted as an Informational Consideration and Conversion Strategy. Compared with Cluster 3, which emphasizes immersive visual engagement and brand affinity, Cluster 4 places greater emphasis on helping audiences understand products, services, benefits, or processes before making a decision.

Carousel-based content is particularly compatible with explanatory communication because it enables information to be organized into sequential elements. The strategy therefore supports activities such as answering customer questions, explaining services, demonstrating benefits, providing educational information, and encouraging audiences to progress toward consultation, inquiry, booking, or purchase.

4.14 Comparative Interpretation of the Four Strategy Archetypes

The four K-modes clusters reveal a progression across different strategic orientations. Cluster 1 emphasizes credibility and trust, Cluster 2 focuses on interest generation and discovery, Cluster 3 prioritizes lifestyle engagement and brand affinity, and Cluster 4 concentrates on information, consideration, and action.

The comparison also reveals a strong relationship between platform-content combinations and strategic

objectives. LinkedIn and document posts are associated with professional credibility and trust building, while TikTok and short videos support dynamic discovery and performance oriented communication. Instagram serves two distinct strategic roles: Reels are associated primarily with lifestyle engagement and brand affinity, whereas Carousels are associated with structured information and consideration.

These findings demonstrate that the same platform can support substantially different strategic objectives depending on the selected content format. Platform selection alone is therefore insufficient to explain social media strategy. Instead, the strategic configuration emerges from the interaction among platform, format, content objective, audience, publication frequency, and CTA.

4.15 Overall Synthesis of the Multivariate Analyses

The combined results from MCA, hierarchical clustering, Gower distance analysis, text-embedding clustering, and K-modes clustering provide a multidimensional understanding of the social mediastrategy dataset. The MCA and distance-based methods demonstrate considerable heterogeneity and overlapping similarity among observations. The low silhouette coefficients across the clustering methods indicate that the strategies do not form sharply separated natural groups.

Nevertheless, the K-modes analysis provides a useful interpretable structure by identifying four broad strategy archetypes: Professional Trust Building, High Energy Product Discovery, Visual Lifestyle and Brand Affinity, and Informational Consideration and Conversion.

A key contribution is the finding that strategic similarity extends across industry boundaries. Industries that appear unrelated in conventional market classifications can nevertheless share similar platform choices, content formats, communication objectives, publication patterns, and audience engagement mechanisms. Conversely, industries within similar broad sectors may employ different strategic approaches.

The findings therefore support the view that social media content strategy should be conceptualized as a multidimensional configuration rather than as an industry specific template. The combination of categorical and textual analytical methods provides complementary evidence: structured categorical variables offer clearer differentiation among strategy profiles, while textual descriptions reveal greater semantic overlap across industries.

Given the limited sample size, interpret these results as exploratory rather than generalizable population-level findings. However, the analyses provide a useful framework for identifying and describing recurring social media content strategy archetypes and demonstrate the value of combining MCA, distance-based clustering, K-modes clustering, and text-based representations when analyzing multidimensional categorical marketing strategies.

5. Discussion

This study's findings highlight the multidimensional and cross-industry nature of social media content strategies. Rather than being confined to industry specific templates, strategic profiles emerge from complex configurations of platform selection, content format, communication objectives, and calls to action (CTAs). The identification of four distinct archetypes Professional Trust Building, High Energy Product Discovery, Visual Lifestyle and Brand Affinity, and Informational Consideration and Conversion demonstrates that organizations across disparate sectors (e.g., Finance and Nonprofit; Fitness and Travel) can share fundamental strategic DNA when their platform format goal alignments converge.

A critical insight from the comparative analysis is the strategic bifurcation of platforms by content format. For instance, Instagram serves dual strategic roles: Reels drive visual lifestyle engagement and brand affinity (Cluster 3), while Carousels facilitate structured information processing and consideration (Cluster 4). Similarly, LinkedIn and document posts are intrinsically linked to credibility and trust building, whereas TikTok and short form videos are optimized for high energy discovery. This underscores that platform selection alone is

insufficient for strategic planning; the true strategic configuration lies in the interaction between the platform, format, and underlying marketing objective.

Methodologically, the study reveals that structured categorical variables (captured via K-modes and Gower-distance clustering) yield clearer strategic differentiation than semantic text embeddings. The significantly lower silhouette score of the text embedding clustering (0.025) compared to the categorical methods indicates that while industries may use overlapping descriptive language (e.g., "engagement," "experience"), their concrete tactical choices (e.g., platform, format, CTA) define their true strategic boundaries.

However, several limitations remain. The dataset comprises a small sample size (n=20) of synthetically generated records, which limits the statistical generalizability of the findings. The relatively low silhouette coefficients across all clustering methods indicate that the boundaries between these archetypes are fluid and overlapping, reflecting the continuous and adaptive nature of real world digital marketing. Therefore, interpret the identified archetypes as exploratory heuristic models rather than rigid, population level statistical estimates.

6. Conclusion

This study provides a novel, data driven framework for analyzing social media content strategies through the lens of multivariate clustering and strategic archetypes. By moving beyond conventional industry classifications, the research demonstrates that effective social media strategies are defined by multidimensional configurations of platforms, formats, goals, and audience engagement mechanisms. The four identified arche types Professional Trust Building, High Energy Product Discovery, Visual Lifestyle and Brand Affinity, and Informational Consideration and Conversion offer actionable insights for digital marketers, content creators, and organizations seeking to align their communication tactics with specific strategic objectives.

Future research should expand upon this exploratory framework by utilizing larger, real world datasets to validate these archetypes and measure their impact on tangible outcomes such as user engagement, conversion rates, and network influence. Additionally, integrating dynamic behavioral tracking and predictive machine learning models could further illuminate how organizations transition between these strategic archetypes in response to evolving digital ecosystems and audience feedback. Ultimately, embracing a multidimensional, arche type based approach to content strategy will be essential for navigating the complexities of the modern network society.

References

[1] Mekacher, A., Papasavva, A. (2022). "I can't keep it up." A dataset from the defunct Voat.co news aggregator. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 16, p. 1302–1311). https://doi.org/10.1609/icwsm.v16i1.19378.

[2] Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., et al. (2009). Computational social science. *Science, 323*(5915), 721–723. https://doi.org/10.1126/science.1167742.

[3] Cauteruccio, F., Corradini, E., Terracina, G., Ursino, D., Virgili, L. (2022). Investigating Reddit to detect subreddit and author stereotypes and to evaluate author assortativity. *Journal of Information Science, 48*(6), 783–810. https://doi.org/10.1177/01655515211043720.

[4] Failla, A., Cazabet, R., Rossetti, G., Citraro, S. (2024). Describing group evolution in temporal data using multi-faceted events. *Machine Learning, 113*, 7591–7615. https://doi.org/10.1007/s10994-023-06456-8.

[5] Record, R. A., Silberman, W. R., Santiago, J. E., Ham, T. (2018). I sought it, I Reddit: Examining health information engagement behaviors among Reddit users. *Journal of Health Communication, 23*(5), 470–476. https://doi.org/

10.1080/10810730.2018.1465493.

[6] Kou, Y., Gray, C. M., Toombs, A. L., Adams, R. S. (2018). Understanding social roles in an online community of volatile practice: A study of user experience practitioners on Reddit. *ACM Transactions on Social Computing, 1*(1), Article 4. https://doi.org/10.1145/3180480.

[7] Bastrygina, T., Lim, W. M. (2023). Foundations of consumer engagement with social media influencers. *International Journal of Web Based Communities, 19*(2–3), 222–242. https://doi.org/10.1504/IJWBC.2023.131410.

[8] Joshi, Y., Lim, W. M., Jagani, K., Kumar, S. (2023). Social media influencer marketing: Foundations, trends, and ways forward. *Electronic Commerce Research*. Advance online publication. https://doi.org/10.1007/s10660-023-09719.

[9] Lim, W. M., Rasul, T., Kumar, S., Ala, M. (2022). Past, present, and future of customer engagement. *Journal of Business Research, 140*, 439–458. https://doi.org/10.1016/j.jbusres.2021.11.014.

[10] Lim, W. M., Rasul, T. (2022). Customer engagement and social media: Revisiting the past to inform the future. *Journal of Business Research, 148*, 325–342. https://doi.org/10.1016/j.jbusres.2022.04.068.

[11] Sekara, V., Stopczynski, A., Lehmann, S. (2016). Fundamental structures of dynamic social networks. *Proceedings of the National Academy of Sciences, 113*(36), 9977–9982. https://doi.org/10.1073/pnas.1602803113.

[12] Al-Garadi, M. A., Varathan, K. D., Ravana, S. D., Ahmed, E., Mujtaba, G., Khan, M. U. S., Khan, S. U. (2018). Analysis of online social network connections for identification of influential users: Survey and open research issues. *ACM Computing Surveys, 51*(1), Article 16. https://doi.org/10.1145/3158667.

[13] Recuero, R., Zago, G., Soares, F. (2019). Using social network analysis and social capital to identify user roles on polarized political conversations on Twitter. *Social Media + Society, 5*(2). https://doi.org/10.1177/2056305119848745.

[14] Cauteruccio, F., Kou, Y. (2023). Investigating the emotional experiences in eSports spectatorship: The case of League of Legends. *Information Processing Management, 60*(3), 103516. https://doi.org/10.1016/j.ipm.2023.103516.

[15] Wang, P., Lü, J., Yu, X. (2014). Identification of important nodes in directed biological networks: A network motif approach. *PLoS ONE, 9*(8), e106132. https://doi.org/10.1371/journal.pone.0106132.

[16] Bajardi, P., Barrat, A., Natale, F., Savini, L., Colizza, V. (2011). Dynamical patterns of cattle trade movements. *PLoS ONE, 6*(5), e19869. https://doi.org/10.1371/journal.pone.0019869.

[17] Munshi, J. D., Munshi, S., Madhu, U., Biswas, D. (2025). Decoding consumer archetypes: A deep learning approach to personality driven behavioural clustering and targeting. *International Journal of Engineering and Information Management, 1*(4), 81–97. https://doi.org/10.52756/ijeim.2025.v01.i04.004.

[18] Chen, M., Chua, A. Y., An, L. (2025). Arche types of influential users in social question answering sites. *Internet Research, 35*(1), 419–447. https://doi.org/10.1108/INTR-01-2023-0012.

[19] Torres, L., Blevins, A. S., Bassett, D., Eliassi-Rad, T. (2021). The why, how, and when of representations for complex systems. *SIAM Review, 63*(3), 435–485. https://doi.org/10.1137/20M1355896.

[20] Patania, A., Petri, G., Vaccarino, F. (2017). The shape of collaborations. *EPJ Data Science, 6*, Article 18. https://doi.org/10.1140/epjds/s13688-017-0117-7.

[21] Bonacich, P., Holdren, A. C., Johnston, M. (2004). Hyper edges and multidimensional centrality. *Social Networks, 26*(3), 189–203. https://doi.org/10.1016/j.socnet.2004.01.001.

[22] Torges, R. (2025). Segmentation strategies for political marketing on social media: A cluster analysis. *DRPEES Desenvolvimento Regional com Políticas Econômicas Estratégicas e Sustentáveis: Governos Municipais, Estaduais e Federal Integrado, 6*(1).