What Time Wears: Lifecycle Modes of Street-Style Fashion (2005–2025)

By

Student Number 5611246

A Dissertation submitted in partial fulfilment of the requirements for the degree of

MA Digital Media and Culture

Supervised by Ching Jin

Word Count: 10980

University of Warwick

Centre for Interdisciplinary Methodologies

August 2025

Abstract

Street-style fashion cycles through imitation, distinction, and seasonal attention, yet few studies trace these rhythms across cities and years. This dissertation builds a two-decade panel of 328,700 WGSN street-style images from ten cities and asks which lifecycle patterns organize street style, and whether attributes help predict them after holding place and time constant. A shape-first pipeline clusters GeoStyle embeddings into 30 styles, summarizes curves with 15 descriptors, and recovers four modes—Seasonal Bursts, Multi-Peak Shift, Recent Intermittent, Long-Tail Fade—forming a lifecycle atlas. Zero-shot FashionCLIP infers item, fabric, texture, and color; city × month fixed effects, stratified permutations, and FDR establish attribute–mode links. Denim and blue correspond to long-tail fades; light and bright palettes to shifts and bursts; trousers and chino to bursts; outerwear and warm materials to intermittent activity. Evaluation shows attributes improve predictive accuracy, and calibrated probabilities enable selective use. The study offers a transferable template for analyzing visual culture over time.

Content

1. Introduction

Street-style fashion does not rise and fall at random. Classic sociology argues that styles recur because people imitate to belong and differentiate to stand out, a tension producing rhythmic collective attention (Simmel, 1957; Veblen, 1994). Blumer described fashion as collective selection sifting many proposals into a few looks deemed of the moment (Blumer, 1969). Diffusion research adds a temporal backbone: innovation and imitation generate adoption waves, while platform dynamics compress and amplify them through influencer seeding, algorithmic feeds, and limited releases (Rogers, 2003; Bass, 1969; Aral, 2020). Seasonality gates attention via climate, retail calendars, and ritual occasions, making certain adoptions feel timely and more likely (Crane, 2012). The theoretical expectation is not idiosyncratic noise but a repertoire of recurring lifecycle shapes.

What is still missing is a long-horizon, city-comparable lifecycle atlas for street style spanning the mid-2000s to the mid-2020s and is comparable at monthly and annual scales. Image-based work advances forecasting and influence mapping, yet typical studies cover shorter windows or single platforms and fail to derive a taxonomy of recurrent lifecycle modes—seasonality, multi-peakedness, and long-run persistence and decline (Al-Halah, Stiefelhagen, & Grauman, 2017; Al-Halah & Grauman, 2020). At the same time, interpretable visual attributes including item, fabric, texture, and color are learnable at scale, but prior research rarely tests whether they align with or predict lifecycle modes once where and when images were captured are held fixed (Liu, Luo, Qiu, Wang, & Tang, 2016). Methodologically, many studies emphasize raw popularity rather than shape and seldom separate structural association from composition using panel identification strategies (Angrist & Pischke, 2009; Wooldridge, 2010).

This study addresses those gaps through three questions. Q1 asks which styles appeared and persisted across ten fashion cities from 2005 to 2025 and what their lifecycles look like at annual and monthly scales. Q2 asks into which recurrent lifecycle modes those style time series cluster and how the modes differ in seasonality, multi-peakedness, long-run trend, and recent activity. Q3 is restated as follows: Do attributes carry mode signal that survives place–time controls and helps predict future mode labels?

I take a shape-first approach that treats the temporal form of style as the object of interest. I assemble a two-decade panel of approximately 328,700 WGSN street-style images across ten cities. I discover thirty styles by clustering GeoStyle embeddings and I infer item, fabric, texture, and color using a zero-shot FashionCLIP pipeline mapped to a cleaned taxonomy that

is stable over time and across places (Mall, Matzen, Hariharan, Snavely, & Bala, 2019; Liu et al., 2016; Radford et al., 2021). For each style I construct shape-normalized lifecycle curves through within-style max normalization and a three-year centered moving average and I compute fifteen descriptors that summarize seasonal strength, peak structure, change-points, recency, trend, and overall activity (Hyndman & Athanasopoulos, 2021). I then cluster descriptor vectors to obtain four interpretable lifecycle modes and evaluate attribute–mode alignment within city by month fixed effects using stratified permutations and false discovery rate control. Finally, I test generalization under forward-rolling time splits with an explainable multinomial logistic model and I evaluate probability calibration to enable confidence-aware use (Benjamini & Hochberg, 1995; Pesarin & Salmaso, 2012; Guo, Pleiss, Sun, & Weinberger, 2017).

The main empirical contributions are fourfold. First, for Q1, I present a two-decade, ten-city lifecycle atlas of thirty discovered styles whose monthly and annual curves can be compared directly. The atlas reveals population-level swells and wide heterogeneity in peak timing and width that are consistent with diffusion and calendar gating in the prior literature (Rogers, 2003; Bass, 1969; Crane, 2012). Second, for Q2, I identify four canonical lifecycle modes that summarize timing, recurrence, and persistence. These modes are Seasonal Bursts, Multi-Peak Shift, Recent Intermittent, and Long-Tail Fade. Third, for Q3 on identification, I show that attribute–mode alignment is structural rather than a by-product of city or month composition. Within city by month fixed effects and after stratified permutation tests with false discovery rate control, I find directionally stable and statistically strong associations that match domain intuition. Denim and blue are associated with the persistent Long-Tail Fade regime. Light and bright palettes align with the Multi-Peak Shift and the Seasonal Bursts regimes. Trousers and chino fabrics align with Seasonal Bursts. Outerwear and warm or insulating fabrics align with Recent Intermittent. These findings indicate that the signals remain after place–time controls and therefore reflect how looks are composed within the same local context rather than only where and when photos were taken (Angrist & Pischke, 2009; Benjamini & Hochberg, 1995). Fourth, for Q3 on generalization, I show under forward-rolling time splits that adding interpretable attributes to fixed effects improves accuracy and macro F1 relative to fixed effects only, and that fitted coefficients mirror the conditional association patterns, which supports the view that the gains are driven by structured, interpretable signals rather than chance fits (Guo et al., 2017; Hyndman & Athanasopoulos, 2021).

The contribution matters in the study of culture and in practice. Theoretically, the results bridge classic sociology of fashion and diffusion with mode-based measurement, showing that

a small repertoire of temporal shapes recurs and that those shapes are semantically grounded in visible attributes that are legible to observers and wearers. Methodologically, the pipeline that combines fixed effects stratification, permutation-based inference, and false discovery rate control provides a template for separating structural association from composition in any visual culture that unfolds across places and seasons. Practically, linking attributes to lifecycle modes informs assortment timing for example gating around seasonal bursts, carry-over bets for example the durability of denim and blue within long-tail fades, and confidence-aware routing in content or buying workflows using calibrated probabilities that abstain on ambiguous cases when warranted (Aral, 2020; Guo et al., 2017).

The paper proceeds as follows. Section 2 motivates the expectation of recurring temporal shapes from fashion cycles, diffusion, and visual semiotics and explains why items and fabrics should track seasonal persistence more strongly than colors and textures. Section 3 details data and panel construction, style and attribute extraction, temporal descriptors, mode discovery, and the identification and prediction protocols. Section 4 presents the lifecycle atlas, characterizes mode signatures, reports conditional associations within fixed effects, and demonstrates predictive validation with explainable evidence. Section 5 concludes with implications and limitations. As throughout, identification is non-causal. Fixed effects mitigate composition bias but cannot remove time-varying confounding by prices, brands, promotions, or audience shifts, which are left to future causal designs (Wooldridge, 2010).

2.Theoretical Framework

2.1 Fashion cycles, diffusion, and seasonality

Styles do not rise and fall at random; they tend to follow recurring temporal patterns because social mechanisms generate rhythmic waves of attention and adoption. Classic theory begins with the tension between imitation (to belong) and differentiation (to stand out). Simmel (1957) framed fashion as a moving frontier where imitation equalizes and distinction reopens distance, while Veblen (1994) emphasized conspicuous display as a status signal that invites emulation and then strategic withdrawal. Blumer (1969) recast the process as collective selection—"Fashion is a process of collective selection" (p. 275)—in which many proposals are sifted by tastemakers, media, and publics into a few "of-the-moment" looks. In practice, those selection arenas are stratified and plural. Elite houses, street scenes, and cross-cultural niches co-exist, generating not only trickle-down flows from couture to mass, but also trickle-up movements from the street to the runway and trickle-across transfers from adjacent subcultures into capsule collections. (Crane, 2012; Gronow, 1993). Updating this lens, contemporary system accounts emphasize how designer–media–consumer linkages are now

platformized and faster, tightening the loop between field positions and mass attention (Kawamura, 2018; Aral, 2020). Put simply, cycles recur because these arenas periodically re-coordinate what counts as timely style.

Diffusion theory gives the mechanism behind the time shapes I expect to find. Rogers (1962) identifies a social sequencing—innovators, early adopters, early/late majority, laggards—that converts design signals into coordinated waves. The Bass model turns that sequence into a parametric rise–peak–decline curve where adoption is driven by an innovation rate (p, editorial exposure, product drops) and an imitation rate (q, peer contagion) (Bass, 1969). When $q \gg p$—the typical successful fashion case—adoption accelerates via social proof, peaks when roughly half of the eventual market has adopted, then declines as replacement overtakes first purchase (Bass, 1969). On digital platforms, seeding and amplification by influencers, recommendation feeds, and "drop" mechanics can raise p (initial visibility) and q (perceived consensus) simultaneously, producing sharper crests—and occasional stalls when cues fragment (Aral, 2020; Jin et al., 2019). This backbone helps distinguish fads (spiky rises under high p, sharp crashes) from durable styles (sustained q, slower, replacement-led fades) and also explains multi-peak revivals: new cohorts, new meanings, or new channels can reopen the cascade after a quiet period (Crane, 2012; Rymes, 2003).

Network-threshold models then help to explain why some waves tip and others stall. In Watts's (2002) threshold framework, cascading adoption requires the system to sit inside a cascade window defined by connectivity and individual thresholds. In sparse scenes (early streetwear, niche crafts), cascade sizes are heavy-tailed and targeting hubs can jump-start diffusion; in dense attention periods (global fashion weeks, platform homepages), outcomes become bimodal—many false starts punctuated by rare blockbusters—because most individuals require multiple confirming cues to tip (Watts, 2002). Field experiments and platform data show that redundant social signals within cohesive clusters are especially potent for behavior change—precisely the condition under which local bursts become system-wide waves (Centola, 2018; Aral, 2020). This mechanism justifies looking for canonical lifecycles instead of treating every trajectory as idiosyncratic noise: coordinated waves and threshold-gated cascades yield a limited menu of time shapes.

Seasonality then acts as a periodic modulator of diffusion. Apparel consumption is structured by climate bands (outerwear vs. sandals), holidays and ceremonies (gifting seasons, weddings, graduations), and a retail calendar that sets synchronized attention spikes (SS/FW markets, fashion weeks, end-of-season markdowns) (Crane, 2012). From an operations

vantage, fashion retail exhibits strong, predictable seasonal demand cycles that interact with assortment and lead-time decisions, concentrating exposure and conversion in well-defined windows (Ferreira et al., 2016). These temporal infrastructures raise exposure (effective connectivity z), amplify social pressure (q via co-attention), and temporarily lower thresholds (φ) by making certain choices feel timely or necessary. Hence I expect 12-month rhythms with out-of-season troughs: styles are more likely to enter the cascade window around buying peaks (pre-holiday, back-to-school) and exit it when attention fragments and thresholds rebound. Seasonality does not create fashion change, but it gates and phases it, shaping when adoption waves are feasible.

Bringing these strands together, I motivate four recurring lifecycle shapes used later in the lifecycle atlas. First, Seasonal Bursts (SB) are sharp spikes aligned with calendar peaks. Theoretically, SBs reflect high q under temporarily low thresholds φ and elevated connectivity z: synchronized media, buying occasions, and social events produce fast, coordinated uptake followed by a quick reversion once the window closes (Watts, 2002; Bass, 1969). Digitally, SBs are amplified when influencer drops and algorithmic surfacing coincide with those gates, momentarily flooding feeds and social circles with confirming cues (Aral, 2020; Jin et al., 2019). Examples include occasionwear micro-trends that crest in the weeks before holidays, then dissipate during off-season resets (Crane, 2012).

Second, Multi-Peak Shift (MPS) captures revival cycles in which a style returns reinterpreted rather than copy-pasted. Revivals occur when new meanings or alliances lower adoption barriers for new publics—e.g., sustainability frames (deadstock denim, upcycled capsules), nostalgia reissues, or cross-scene collaborations (music or sport tie-ins) (Crane, 2012). Empirically, nostalgia increases social connectedness and approach motivation, helping re-ignite interest among new and lapsed consumers (Sedikides et al., 2015). Simultaneously, the rise of circular fashion practices and sustainability discourse has reframed archival materials and repair/upcycling as aspirational rather than merely thrifty, creating fresh pathways for second-wave adoption (Niinimäki et al., 2020). Design-system accounts also document the institutionalization of street-to-runway pipelines and collaborative capsules that deliberately engineer cross-scene transfers, renewing symbolic value without replicating past meanings (Kawamura, 2018). In Bass terms, a second wave raises effective p (renewed editorial interest) and/or q (new peer cues) within a fresh cohort; in network terms, the system briefly re-enters the cascade window (Watts, 2002). MPS thus explains why dormant aesthetics can post a late second peak under shifting cultural logics.

Third, Long-Tail Fade, or LTF, describes styles that move from early spikes to habits sustained by replacement, routine, or basic compatibility. For example, normcore staples can mature into the basic infrastructure of a wardrobe. Here, q remains nontrivial within established segments even as novelty wanes, so sales flatten into a low-amplitude tail dominated by repeat purchase (Bass, 1969). Sociologically, the look ceases to do acute status signaling and becomes a taken-for-granted baseline of taste (Blumer, 1969; Gronow, 1993). Psychologically, the shift from deliberative novelty seeking to cue-driven repetition is consistent with contemporary habit formation research: repeated choices in stable contexts become automatic, supporting long-tail persistence without headline attention (Wood & Neal, 2016).

Fourth, Recent Intermittent (RI) covers micro-scene ripples—waves that appear, recede, and recur at modest amplitude within networked niches. Sparse connectivity yields local heavy tails: a few scenes experience strong pulses, but the system as a whole never locks into a global cascade (Watts, 2002). Because complex contagions rely on clustered reinforcement, RI patterns can cycle within tightly knit communities (campus, genre, or locale) and remain intermittent when cross-cluster bridges are weak (Centola, 2018). Calendar gating still matters—e.g., campus, tour, or festival rhythms—but effects remain subcultural and periodic rather than mass and singular (Crane, 2012).

Two further implications follow for empirical design. First, if seasonal infrastructures periodically lift q and lower φ, then calendar-aligned windows should predict higher conversion of editorial exposure into adoption—i.e., more SBs and stronger peaks within MPS events near seasonal gates. Operational data suggest precisely such timing payoffs, with in-season exposure and assortment alignment yielding disproportionate demand realization (Ferreira et al., 2016). Second, if network position conditions tipping, then targeting hubs (key retailers, celebrity nodes) will matter more in sparse regimes (early scenes, off-season) than in dense ones (platform-driven, in-season), consistent with Watts's (2002) distinction. Platform research likewise cautions that the same seeding tactic can either cascade or fizzle depending on audience structure and redundancy of cues (Aral, 2020; Jin et al., 2019; Centola, 2018). Both points justify modeling exposure (p proxies), social proof (q proxies), and seasonality (calendar covariates) jointly, rather than treating seasonality as a mere nuisance.
In sum, classical sociology explains why style waves recur (imitation/differentiation mediated by collective selection and status flows), diffusion models explain how they trace recognizable time shapes (Rogers, 1962; Bass, 1969), network thresholds explain when they tip (Watts, 2002), and industry calendars explain where they peak (Crane, 2012). Recent

advances in complex contagion, platform dynamics, sustainability, nostalgia, and retail analytics strengthen this integrated account, clarifying the specific levers—redundant social signals, algorithmic amplification, reframed meanings, and seasonal operations—through which recurring temporal modes (SB, MPS, LTF, RI) emerge and persist (Aral, 2020; Centola, 2018; Niinimäki et al., 2020; Sedikides et al., 2015; Ferreira et al., 2016). This integrated view motivates the expectation of a small set of recurring temporal modes—SB, MPS, LTF, RI—against which the forthcoming lifecycle atlas can be estimated and validated.

2.2 Visual attributes as carriers of style signals

At a glance, style is "read" before it is felt or explained; the garment operates as a sign whose visual properties—item category, fabric, texture, and color—carry socially legible messages (Barthes, 1983; Barnard, 2003; Davis, 1992). Building on fashion semiotics and material culture, I argue that these attributes align with distinct lifecycle regimes of appearance, namely long-term persistence, mid-period shifts and bursts, and recent intermittent adjustments. They do so because they encode both function—the body's needs under specific conditions—and symbolic meaning—what the wearer wants others to perceive. In short: items and fabrics anchor slow, seasonal, and durable signals; textures and colors modulate fast, affective, and context-sensitive signals. This section therefore motivates why attributes should predict time-shape variation and prepares the ground for the measurement and inference strategy developed later.

Semiotics and materiality shape how garments communicate. Item categories such as outerwear, jeans, and blazers are highly conventionalized signifiers; they carry entrenched codes of formality, authority, and occasion (Barthes, 1983; Barnard, 2003). Because these codes are widely shared and repeatedly reinforced in retail, media, and everyday practice, they provide a stable vocabulary that wearers can deploy with relatively low ambiguity. Fabrics add a material index—thermal and structural affordances across wool, denim, and technical knits that are immediately visible in drape, bulk, and surface sheen—thereby binding function to meaning, read as "warm," "protective," or "soft yet sharp" (Li, 2001). The tactile imagination works alongside vision: even without touching, viewers infer hand, stiffness, and breathability from the way cloth hangs or creases, further stabilizing interpretation. Texture and pattern govern salience and grouping: edge density, contrast, and repetitiveness determine what pops out pre-attentively and guide where observers look first (Itti & Koch, 2001). Color is both culturally coded and affectively potent. Ecological valence helps explain broad preferences; for example, many audiences favor blue and green because of positive environmental associations, while physiological research shows that brightness

and saturation systematically map to pleasure and arousal—Pleasure approximately .69·brightness + .22·saturation; Arousal approximately −.31·brightness + .60·saturation—so bright and saturated palettes read as energizing, whereas desaturated and dark palettes read as calming or sober (Palmer & Schloss, 2010; Valdez & Mehrabian, 1994; Elliot & Maier, 2014). Taken together, categories and fabrics stabilize meaning by linking what a garment is and does, while textures and colors tune how intensely and where the look is perceived and felt.

Why items/fabrics should track seasonality and persistence more strongly than colors/textures. Two constraints make item/fabric more "sticky" to lifecycle regimes. First, functional coupling to climate and activity. Thermal and vapor transport properties vary dramatically by fabric family and garment architecture; they are regulated by standards, measured on manikins across wind/sitting/walking conditions, and are costly to mismatch (Smallcombe, Hodder, & Havenith, 2021; Li, 2001). Consequently, outerwear, insulating knits, and lined trousers predictably concentrate in cold seasons, while ventilated or moisture-wicking pieces rise with heat and motion. These regularities are not just symbolic but physiological and logistical, which is why they recur annually. Retail behavior data corroborate that weather shocks shift basket composition and price/quantity choices, especially in rain and temperature rises, implying function-bound items drive recurrent, exogenous cycles (Tian, Pan, & Yu, 2021). Second, economic durability and wardrobe ecology. High-cost, high-wear categories (coats, boots, denim) amortize over time and become biographically persistent: people keep and re-wear them as default anchors of the ordinary (Miller & Woodward, 2012), making these items/fabrics reliable carriers of longer-term style identity. By contrast, colorways and surface textures can be inexpensively swapped within the same item archetype—say, a blazer in spring pastels versus autumn neutrals—are selected to modulate mood/arousal or signal micro-shifts in role, and can be updated with minimal functional penalty (Elliot & Maier, 2014; Valdez & Mehrabian, 1994). In semiotic terms, the lexeme (category/fabric) is stable, the inflection (color/texture) is agile (Barthes, 1983; Davis, 1992). This asymmetry sets expectations: when time and place are held constant, item/fabric should contribute more to persistent seasonal placement than color/texture.

Attentional mechanics strengthen this division of labor. Pre-attentive selection privileges contrastive features (orientation, intensity, color opponency), boosting textures and high-chroma hues as "fast beacons" that redirect gaze within 25–50 ms per item (Itti & Koch, 2001). These are ideal for short-cycle signaling—"today I'm energized/approachable"—without re-platforming the outfit. Meanwhile, the global silhouette and fabric hand determine

perceived propriety and comfort (line, coverage, thermal fit), which are harder to violate because they collide with both social codes and bodily regulation (Li, 2001; Barnard, 2003). Thus, items/fabrics align with seasonal persistence; textures/colors align with day-to-day modulation. Put differently, silhouettes and materials set the base tempo; palettes and surfaces play the syncopation on top.

Affordances and replacement costs shape how readily style attributes can be altered. Compared with texture or color, swapping an item or a fabric is substantially more expensive and constrained. Changing a category or core material triggers pattern revisions, fit verification, material re-sourcing, and minimum-order-quantity requirements; it also risks comfort and propriety if misaligned with climate or activity (Li, 2001). In apparel operations, these choices are locked in by lead times and component commitments, whereas colorways and prints are lower-cost overlays that can rotate faster within the same archetype (Abernathy, Dunlop, Hammond, & Weil, 1999; Cachon & Swinney, 2011; Christopher, Lowson, & Peck, 2004; Ghemawat & Nueno, 2006; Fisher, 1997). This asymmetry induces wardrobe inertia: coats, denim, and tailored trousers are re-worn across seasons and even years, forming persistent anchors, while palette and surface act as agile edits layered on a stable base. Therefore, once city × month fixed effects are controlled, I expect items and fabrics to track seasonality and long-run persistence more strongly than colors and textures. The same logic clarifies supply-side behavior: firms re-platform categories and fabrics sparingly, but cycle palettes and prints to refresh demand without incurring fit or sourcing risk.

Why fixed effects are needed. City × month fixed effects absorb major compositional differences—climate bands such as the contrast between Seoul and São Paulo, local retail calendars, and city-specific semiotic norms—so any remaining attribute–mode alignment reflects within-stratum structure rather than the time and place where photos were taken (Angrist & Pischke, 2009; Wooldridge, 2010). In other words, by conditioning on place–time strata, fixed effects move my analysis from compositional co-occurrence toward structural association. This design choice limits confounds from photography supply, event clustering, or transient weather shocks, and enables comparisons of how looks are composed within the same local context.

Finally, this division of labor explains the everyday paradox of dress: people stay the same by repeating categories/fabrics that secure identity and bodily ease, yet feel different by tilting palette and surface to the day's mood and gaze economy (Miller & Woodward, 2012; Elliot & Maier, 2014). In practice, the most robust style signals are not loud but consistent—anchored

in what the garment is and does—while the most agile signals are chromatic and textural edits that steer attention and affect without breaking functional contracts with weather, work, or wallet (Barnard, 2003; Li, 2001; Smallcombe et al., 2021). This is precisely why attribute families map onto different temporal modes: slow codes locate a look in seasonal time; fast codes animate it from day to day.

2.3 From theory to measurement: operationalization and mode abstraction

This section turns conceptual claims about style and lifecycle into measurements that are reproducible, interpretable, and robust, making the abstract ideas auditable against data and algorithms. The design relies on established methods in visual representation, unsupervised learning, and time-series analysis so that high-level constructs can be mapped to concrete features and labels that behave stably under reasonable perturbations (Cleveland et al., 1990; Fulcher et al., 2013; Hyndman & Athanasopoulos, 2018; von Luxburg, 2010).

Style becomes a measurable object by treating it as a cluster in an image-embedding space extracted from GeoStyle street photographs that capture geographically distributed fashion signals (Mall et al., 2019). Image embeddings are produced with a high-capacity convolutional backbone such as Inception and, when semantics matter, with a vision–language encoder that couples images to text supervision for transferability (Radford et al., 2021; Szegedy et al., 2015). Clustering is performed with k-means on L2-normalized vectors and the number of clusters is selected by stability under perturbations following clustering-stability theory: subsample cities and years, re-cluster, and compute Adjusted Rand Index, abbreviated ARI, across runs; then normalize stability against a permutation-label null so that larger k is not rewarded merely for granularity (Pedregosa et al., 2011; von Luxburg, 2010). The working label set, referred to as style_id, is defined as the Top-30 clusters ranked by ten-city totals to balance coverage with interpretability for temporal analysis in the sense of GeoStyle's cross-city lens (Mall et al., 2019).

Attributes are linked to these style_ids through zero-shot recognition using a domain-adapted CLIP model, referred to as FashionCLIP, which aligns image features with natural-language labels for open-set categories (Chia et al., 2022; Radford et al., 2021). The attribute taxonomy is curated from DeepFashion so that noisy synonyms are pruned and ambiguous terms reduced, yielding compact fields for category, fabric, texture, and color that can be aggregated cleanly at scale (Liu et al., 2016). Category is scored as a single-label argmax over text similarities, whereas fabric, texture, and color are first scored as multi-label with calibrated thresholds and then collapsed to Top-1 per family to ensure parsimonious coding and

consistent downstream statistics (Frome et al., 2013; Radford et al., 2021). Because zero-shot logits are often mis-calibrated, temperature scaling on a hand-checked validation slice is applied before thresholding to improve probability calibration, which follows guidance on modern network calibration (Guo et al., 2017). The final attribute vector is stored as one-hot fields so that aggregation by style_id and error audits are straightforward.

Temporal behavior is represented as per-style yearly curves that emphasize shape rather than scale in order to compare lifecycles across cities with different sampling intensity. Yearly counts are first divided by the style's yearly image opportunity to mitigate coverage variation across time, then max-normalized to a unit peak to align shapes, and finally smoothed with a three year moving average to reduce high-frequency noise while preserving turning points and change-points (Hyndman & Athanasopoulos, 2018). Missing interior years are linearly interpolated while leading and trailing gaps remain missing so that artificial early or late signals are not introduced. From each curve, a 15-dimension descriptor vector is computed that captures lifecycle structure: seasonality strength and dominant period using STL decomposition and spectral energy, peakiness and kurtosis, number and prominence of change-points, trend slope via Theil–Sen, recent momentum defined as last-year versus trailing three-year delta, persistence as the share of active years, volatility via median absolute deviation, concentration via a Gini measure over years, and a recency index that emphasizes late-cycle activity (Cleveland et al., 1990; Fulcher et al., 2013; Hyndman & Athanasopoulos, 2018). This compact signature allows styles to be compared by shape in a way that is robust to scale differences.

Lifecycle modes are discovered by clustering in the descriptor space and then applying transparent post-labels so that the resulting partitions are both data-driven and explainable. A modest number of clusters, typically between four and six, tends to yield interpretable partitions, and four canonical modes are emphasized: SB denotes Short-lived Burst, MPS denotes Multi-Peak Seasonal, RI denotes Rapidly Increasing, and LTF denotes Long-Term Fade. Robustness is verified with a two-part stability regime inspired by the distinction between jittering and jumping in clustering stability. Jittering checks perturb descriptors by bootstrapping years or adding small noise and then verify assignment stability with ARI, while jumping checks re-run k-means under different random seeds and nearby k to verify that modes persist rather than flip between local optima (von Luxburg, 2010). Stability curves are normalized against a scrambled-descriptor null so that significance is not overstated, and where stability is marginal, simple rule-based labels are applied using thresholds on peakiness and persistence to preserve interpretability.

Methodological choices connect directly to prior evidence so that the operationalization is scientifically defensible and practically useful. Embedding-based discovery leverages the empirical structure of learned features and gains rigor from stability-normalized clustering, preventing over-interpretation of arbitrary partitions while preserving the GeoStyle intuition about geographically distributed trends (Mall et al., 2019; von Luxburg, 2010). Zero-shot attributes supply language-grounded semantics while controlling calibration and taxonomy drift, which keeps labels meaningful across time and cities (Chia et al., 2022; Guo et al., 2017; Liu et al., 2016; Radford et al., 2021). Time-series design separates shape from magnitude through max-normalization, STL-style seasonality isolation, and a highly comparative descriptor set, which enables retrieval and atlas-style mapping of lifecycle patterns rather than raw popularity (Cleveland et al., 1990; Fulcher et al., 2013; Hyndman & Athanasopoulos, 2018). Implementation uses tested tooling so that the pipeline remains accessible and replicable in typical research environments (Pedregosa et al., 2011). Finally, the mode abstraction is kept explicit: data perturbation, initialization strategy, and distance metric are treated as dials, and decisions are reported under those modes, which makes the measurement transparent and auditable in the spirit of stability-based validation (von Luxburg, 2010). These choices justify the atlas introduced in Section 4.1 and enable clean comparisons of lifecycle shape across cities and styles.

2.4 Identification and generalization: fixed effects and prediction

The goal in this section is to argue that the reported associations are structural rather than compositional and that they generalize out of sample. To combat composition problems—city mix and seasonality—models include city × month fixed effects (FE) so that comparisons are made within a stratum that shares a location and calendar context. FE difference out time-invariant city heterogeneity and common monthly shocks, shrinking the scope for spurious cross-city or seasonal correlations (Wooldridge, 2010). Hypothesis tests are conducted within stratum: from the FE model I read standardized residuals z and lift (a standardized magnitude with sign) to determine effect size and direction. A minimum cell count rule screens out sparse strata to avoid artifacts.

Inference uses permutation tests stratified by city × month to preserve the nuisance structure while breaking the attribute–outcome link; this yields finite-sample, empirical *p*-values without relying on parametric tails (Pesarin & Salmaso, 2012). To address multiplicity across many attributes and outcomes, I apply Benjamini–Hochberg false discovery rate (BH–FDR) to the full set of permutation p-values so that the expected fraction of false discoveries is

controlled at the pre-set level (Benjamini & Hochberg, 1995). Findings are thus interpreted only if they (i) are identified within FE strata, (ii) pass stratified permutations, and (iii) survive FDR screening, with z and lift providing immediately legible size/direction.

Generalization is established through predictive validation under rolling time splits (a rolling origin scheme) that respects the temporal order of the data (Hyndman & Athanasopoulos, 2021). In each fold, I train a multinomial logistic regression with L2 regularization that includes city × month FE plus the selected visual attributes. Staged ablations—from FE-only to FE+visuals—quantify incremental predictive value beyond what FE already explains. Performance is evaluated not only by accuracy but also by probability quality: Brier score, Expected Calibration Error (ECE), and Maximum Calibration Error (MCE) (Guo et al., 2017). When calibration gaps appear, I fit a single-parameter temperature scaling model on the validation fold; this improves probability reliability without altering class predictions, preserving accuracy while fixing confidence (Guo et al., 2017).

Because deployment often benefits from knowing when not to predict, I adopt confidence-aware selective classification. After calibration, a score-based gate (SB-gate) admits a prediction only if the calibrated maximum class probability exceeds a threshold tuned on validation folds. I report the risk–coverage (RC) curve and choose an operating point that meets application constraints (e.g., ≤5% selective risk with maximal feasible coverage). Selective classification provides a principled trade-off—lower error on accepted cases at the cost of abstaining on ambiguous ones—and comes with theoretical envelopes that clarify achievable and non-achievable regions for a given hypothesis class (El-Yaniv & Wiener, 2010). This safeguards decision quality under distribution drift while preserving throughput where the model is most reliable.

Identification here is non-causal. FE remove city-level time-invariant confounding and soak up coarse seasonality but cannot eliminate time-varying confounders such as price changes, brand actions, promotions, or audience shifts (Wooldridge, 2010). Claims are therefore framed as predictively structural—stable conditional associations that replicate under forward-rolling evaluation—rather than as treatment effects. Causal designs (e.g., instruments, staggered adoption, or experiments) are left for future work.

Taken together, the pipeline—within-stratum FE estimation, stratified permutations, BH–FDR control, rolling-origin prediction with calibrated probabilities, and confidence-aware deployment—delivers associations that are insulated from composition bias and that remain

practically useful for forward prediction (Benjamini & Hochberg, 1995; Pesarin & Salmaso, 2012; Wooldridge, 2010; Hyndman & Athanasopoulos, 2021; Guo et al., 2017; El-Yaniv & Wiener, 2010).

3.Methodology

3.1 Data & Panel Construction

I use the WGSN street-style image collection from 2005 to May 2025 covering ten fashion cities (London, Los Angeles, Melbourne, Milan, New York, Paris, São Paulo, Seoul, Shanghai, Tokyo). The unit of observation is the image; three keys are retained—image_name, city, and date (YYYY.MM). Records missing any key are dropped, and city strings are normalized (e.g., sao_paulo, new_york) to prevent aliasing. In total the working sample comprises 328,700 images, which provides adequate support for city–month granularity.

Each image later receives a style_id from unsupervised clustering (see Section 3.2). I then aggregate image counts by (style_id, city, date) to construct monthly frequencies and pivot them to a wide city × month panel (new_style_counts.tsv) with index [city, date] and columns style_0 … style_29 (missing months filled with 0). The study window is 2005–2025; year 2025 is partial and is kept in descriptive plots but interpreted with caution, as end-point declines may conflate genuine cooling with truncation effects.

3.2 Style & Attribute Extraction

3.2.1 Style discovery

For each image, I extract a 1024-dimensional fixed representation with the GoogLeNet backbone from GeoStyle (Mall, Matzen, Hariharan, Snavely, & Bala, 2019). In this space I run k-means (k = 30, random_state = 42, n_init = 10) and assign $style_i d \in 0, \ldots, 29$ to every image. To focus on mainstream phenomena, I rank styles by ten-city totals (2005–2025) and retain the Top-30 as the analysis set (see Section 4.1). style_id is the sole style label used throughout the lifecycle atlas (see Section 4.1), mode discovery (see Sections 3.3 and 4.1), association tests (see Section 4.2), and predictive modeling (see Sections 4.3–4.4).

For interpretability, I select medoid images (nearest to the cluster centroid) as exemplars and add single-annotator descriptive names (e.g., Urban Black Minimalist) based on the distribution of explainable attributes. These names are narrative only and never enter modeling.

3.2.2 Attribute taxonomy（item, color, fabric, texture）

I construct a cleaned, hierarchical attribute vocabulary spanning item, fabric, texture, and color by reconciling the DeepFashion category/attribute space (Liu et al., 2016) with labels prevalent in FashionCLIP. Each image receives a single item label (e.g., T-shirt, skirt, trench coat, sneakers), while fabric covers categories such as denim, cotton, and leather; texture includes floral, stripe, polka dot, and tonal; and color comprises standardized base hues supplemented by a small set of high-frequency fashion tones. To improve robustness and interoperability, I merge synonyms and near-duplicates (e.g., “floral print” to “floral”), drop rare or ambiguous tokens, and enforce lowercase ASCII normalization to ensure stable encoding for downstream analyses.

To infer attributes in a zero-shot manner, I apply FashionCLIP (patrickjohncyh, n.d.) with prompt templates such as “a photo of a {item}” and “{fabric} fabric”. Cosine similarity between image/text embeddings provides scores. item contributes a single label per image, while fabric/texture/color are first scored in a thresholded multi-label routine and then collapsed to the top-scoring label (Top-1) within each family—so every image contributes exactly one label per family, with thresholds tuned on a small development set to balance coverage and noise. If multiple labels in a family exceed the threshold, the top score is retained (ties may be kept for recall and expanded via one-hot later). All outputs are mapped back to the normalized vocabulary and one-hot encoded, then merged with style_id, city, and date to form a tidy analysis table. Complete vocabulary lists and family sizes are provided in Appendix D (D.4.2–D.4.3). Attributes never affect the clustering; they support interpretation and downstream modeling (Mall et al., 2019).

To ensure reproducibility and provide diagnostic transparency, I report per-family coverage, check co-occurrence sanity (e.g., Swimsuit × Denim is rare), and conduct random human spot checks. Vocab lists, prompt templates, random seeds, and model versions (GeoStyle backbone; FashionCLIP weights) are version-controlled, and clustering/selection seeds are logged to ensure strict reproducibility (Liu et al., 2016; Mall et al., 2019; Patrickjohncyh, n.d.).

3.3 Temporal Representation & Mode Discovery

3.3.1 Lifecycle Curves

I start from monthly counts of each style in each city and aggregate them to yearly totals across the ten cities to obtain a style–year panel $y_{i,t}$ for style i and year $t \in \{2005, \dots, 2025\}$. Year 2025 is a partial year; I retain it for completeness in descriptive figures

but interpret it cautiously. To compare shapes rather than magnitudes, I max-normalize each series within style,

$$\widetilde{y_{i,t}} = \frac{y_{i,t}}{\max_{t'} y_{i,t'}} \in [0,1],$$

then apply a 3-year centered moving average using available years at the boundaries,

$$\bar{y}i,t \;=\; \frac{1}{n_t}\sum s \in \{t-1,t,t+1\} \cap \mathcal{T}\; \widetilde{y_{i,s}},$$

$$n_t \;=\; |\{t-1,t,t+1\} \cap \mathcal{T}|.$$

For the overview heatmap, rows are ordered by each style's peak year

$$t_i^{\star} \;=\; \arg\max_{t\in\mathcal{T}} \quad \overline{y_{i,t}}.$$

and I summarize the population with the cross-style median and interquartile ribbon

$$m_t \;=\; \mathrm{me}median\; i(\bar{y}i,t),$$

$$q_{25,t} \;=\; Q_{0.25}(\{\bar{y}i,t\}i),$$

$$q_{75,t} \;=\; Q_{0.75}(\{\bar{y}i,t\}i).$$

On these smoothed, normalized series $\overline{y_{i,t}}$, I compute 15 temporal descriptors that jointly capture seasonality, peak structure, timing and spread, spacing stability, structural change, recency, trend, and overall activity: seas12 (12-month seasonal strength, [0,1]); peak structure—n_peaks, peakiness (unitless sharpness), Gini (0–1), std_y; timing and width—t_peak, interquartile range (IQR) (months); spacing stability—peak_spacing_cv (≥0); structural change—cp_rel (location of the best single break, 0–1) and cp_bic_gain (BIC improvement of a two-segment fit); recency and trend—late_early_ratio, recency24, trend_slope_36; and frac_active (share of non-zero months). Features are standardized by z-score across styles (median imputation for rare missings). Peak detection uses a unified local-extrema rule; change-points are selected by single-break search with BIC comparison.

3.3.2 Mode Discovery

I then cluster the 30 style trajectories in the standardized feature space with k-means (k=5, n_init=50, random_state=42) and apply a two-step post-labelling to improve semantic clarity: (i) merge the small "miscellaneous" cluster into the nearest major centroids; (ii) apply reproducible gate rules in the following order: SB, then MPS, then LTF, with remaining styles labelled RI:

SB (Seasonal Bursts): seas12 ≥ 0.22 and peakiness ≥ 45 and frac_active ≤ 0.45 and late_early_ratio ≤ 0.80;
MPS (Multi-Peak Shift): cp_rel ≥ 0.58 and cp_bic_gain ≥ 35 and t_peak ≥ 70;
LTF (Long-Tail Fade): late_early_ratio ≤ 0.75 and trend_slope_36 < 0 and n_peaks ≤ 15 & seas12 ≤ 0.20;

RI (Recent Intermittent): otherwise.

This yields four stable modes (SB = 15, MPS = 7, RI = 4, LTF = 4). Separation in feature space, median-profile shapes with IQR bands, and medoid panels are reported in the atlas (see Section 4.1).

Finally, I assess robustness via 200 perturbation reruns—20% feature sub-sampling with small Gaussian noise—computing the Adjusted Rand Index (ARI) against the baseline partition: median approximately 0.36; 90th percentile (P90) approximately 0.51. A pairwise co-occurrence heatmap shows the block-diagonal structure is most stable for SB/MPS, with limited bridging between RI/LTF near boundaries—sufficient for the downstream analyses in Sections 4.2–4.4.

3.4 Association Tests & Predictive Modeling

This section addresses two linked questions. First, are visual attributes conditionally associated with lifecycle modes once city × month composition is held fixed? Second, do those conditional signals carry out-of-sample predictive value under time-consistent evaluation? I begin with within-stratum tests under city × month fixed effects (FE), then assess generalization using an explainable multinomial logistic model. Implementation summary, file artifacts, and configuration are in Appendix D (D.1–D.2).

To quantify conditional associations, for each attribute family (item, fabric, texture, color) and each level a versus mode k, I assume independence within each fixed-effects (FE) stratum f and compute the stratum-wise expectation:

$$E_{a,k,f} = \frac{N_{a,\cdot,f}\, N_{\cdot,k,f}}{N_{\cdot,\cdot,f}}.$$

I then aggregate expectations and counts across strata:

$$E_{a,k} = \sum_f E_{a,k,f} \quad ,$$
$$O_{a,k} = \sum_f O_{a,k,f}.$$

Direction and magnitude are summarized by the standardized residual and lift:

$$z_{a,k} = \frac{O_{a,k} - E_{a,k}}{\sqrt{\mathrm{Var}(O_{a,k} \mid \mathrm{FE})}} \quad ,$$
$$\mathrm{lift} = \frac{O_{a,k}}{E_{a,k}}.$$

The variance is approximated by multinomial–Poisson additivity across strata. Importantly, statistical significance does not rely on that approximation: within each FE stratum I permute mode labels (holding stratum marginals fixed), repeat $B = 200$ times to obtain two-sided empirical p values, and then apply the Benjamini–Hochberg false discovery rate procedure within each attribute family; unless otherwise noted, the threshold is $q < .05$. Low-support

cells are excluded using a minimum count of 10 (with sensitivity checks in the Appendix). This protocol mirrors the conditional analyses reported in Section 4.2.

To test whether these signals generalize over time, I fit a multinomial logistic regression (L2 penalty, solver = saga) using one-hot visual attributes plus city × month FE. One-hot encoders and FE dummies are fit on the training block only; unseen levels in the test block map to zero, preventing temporal leakage. The regularization hyperparameter C is selected on the training block via an inner GroupKFold (grouped by city × month FE) over a logarithmic grid, and then held fixed for the outer evaluation. The outer loop follows rolling-origin time splits (train in the past, test in the future). I report Accuracy and Macro-F1 as mean ± SD and 95% confidence intervals across outer splits, and compare the model with FE-only, Majority-class, Prior-sampling, and Uniform baselines. Ablations add feature families in stages (from +Item to +Item+Color to +Item+Color+Texture to All(+Fabric)) to localize performance gains. Probability calibration (Brier; ECE/MCE) and simple deployment heuristics (selective risk, SB-gate) are documented in Appendix C; implementation knobs and outputs are catalogued in Appendix D (D.1, D.2).

For statistical comparison across outer splits, I use a paired Wilcoxon signed-rank test (two-sided) between the main model and baselines and annotate figures with *p*-values. Overall, the pipeline follows the identify-then-validate sequence set out in (see Section 4.2): first confirm within-FE conditional alignment of attributes and modes, then test out-of-sample generalization under time-consistent splits. Finally, this is non-causal identification: FE mitigates composition bias but does not control for price, brand, promotion, or audience shifts; causal claims are deferred to future designs (e.g., natural experiments, IV, DiD).

4.Findings&Analysis

4.1 Style Taxonomy and Lifecycle Atlas（2005-2025）

4.1.1 Global Lifecycle Overview (Top-30 Styles)

To ground the mode analysis that follows, this subsection first lays out the thirty discovered styles and their long-run trajectories across ten cities from 2005 to 2025. The analysis set keeps the Top-30 styles by total appearances in the ten-city panel. Each cluster is indexed by style_id (0–29) and accompanied by a descriptive, non-inferential label assigned by a single annotator for readability; labels do not affect any statistics or models.

As shown in Figure 1, three high-level observations help set the stage. First, the population ribbon shows two gentle swells—around 2009 and 2016—while the IQR remains wide,

indicating persistent heterogeneity in both peak timing and breadth across styles (see Figure1, panel a). Second, the heatmap reveals early-peak, mid-decade, and late-peak regimes, alongside a small group of "evergreen" styles whose normalized levels remain moderate for many years (see Figure1, panel b); operationally, such evergreen behavior corresponds to lower interannual variability. Third, peak width varies markedly: some styles exhibit narrow, high spikes; others rise and decay slowly, suggesting different diffusion and replacement dynamics.

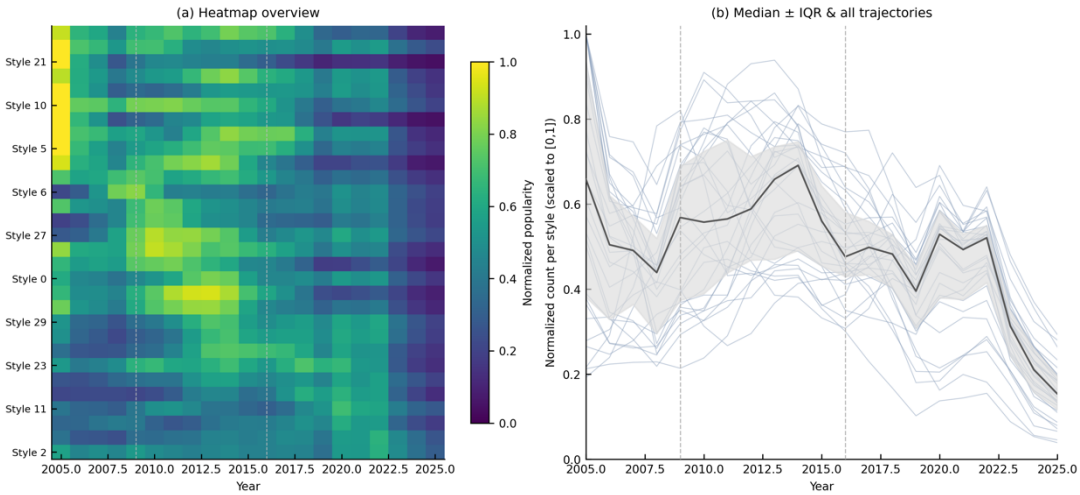

Figure 1. Global lifecycle overview of the Top-30 styles (2005–2025, ten-city panel)

Figure 2 provides medoid exemplars—one image per cluster, chosen as the closest sample to the cluster centroid in the GeoStyle feature space (see Section 3.2). Labels are purely descriptive (e.g., "Urban Black Minimalist") and serve to connect visual semantics (dominant colors, layering, patterns) to time-series shapes before formal mode discovery.

Figure 2. Visual exemplars of the 30 discovered style clusters

Together, the atlas motivates the four lifecycle modes used later—Seasonal Bursts (SB), Multi-Peak Shift (MPS), Recent Intermittent (RI), and Long-Tail Fade (LTF)—by showing that timing, peakiness, and persistence vary systematically across styles.

4.1.2 Lifecycle Modes and Their Signatures

Having mapped the thirty style trajectories, I now summarize how four lifecycle modes emerge and what characterizes each. Modes are obtained by clustering the 15 temporal descriptors of each style's normalized, smoothed series, followed by post-labeled merges and rule-based disambiguation to enhance semantic clarity (see Section 3.3). The final partition contains Seasonal Bursts (SB, n = 15), Multi-Peak Shift (MPS, n = 7), Recent Intermittent (RI, n = 4), and Long-Tail Fade (LTF, n = 4); hereafter, I use these abbreviations. A 200-run perturbation test indicates moderate stability (ARI median ~0.36, P90 ~0.51); full procedures and diagnostics appear (see Section 3.3). Figure 3 places the post-labeled clusters in the PCA projection of the 15-feature lifecycle space, showing clean separation between SB and MPS and compact, smaller footprints for RI and LTF. A few points near decision boundaries are expected given shared elements (e.g., weak seasonality in some MPS cases), but the global structure is preserved.

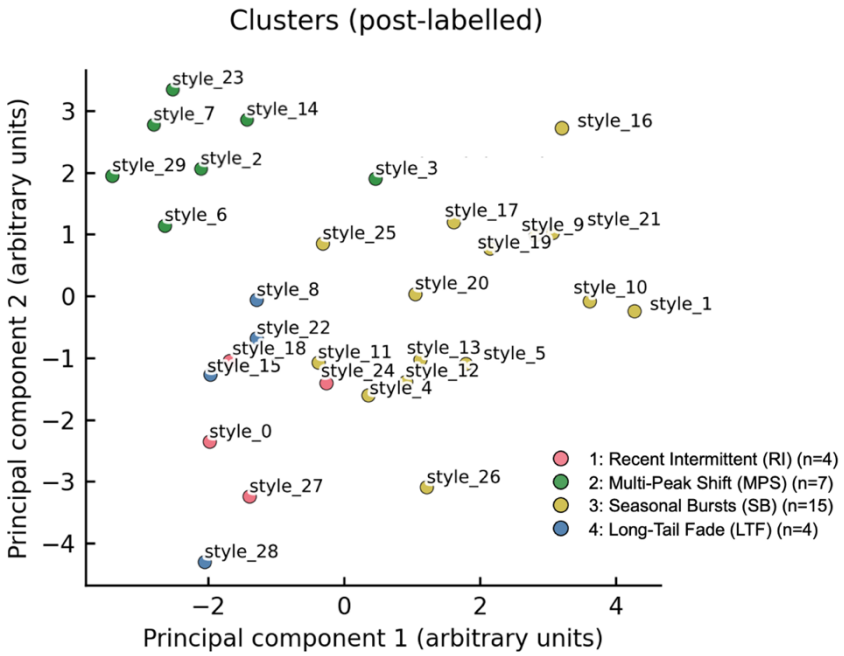

Figure 3. Clusters in Lifecycle-Feature Space

Turning to shapes, Figure 4 plots per-mode monthly median profiles (max-normalized per style; MA = 3) with IQR bands, making clear how the modes differ in timing, peakiness, and

persistence:

SB shows regular 12-month spikes with off-season troughs near zero (high seas12, high peakiness, concentrated gini). Peaks arrive like a calendar rhythm and decay rapidly after the high season.

MPS exhibits a mid-to-late regime change followed by dense, non-seasonal peak groups (larger cp_rel and cp_bic_gain; higher n_peaks with variable spacing). The profile often features a step or a cluster of peaks later in the series.

RI concentrates high-amplitude bursts in recent years yet lacks an annual rhythm (higher late_early_ratio, low seas12). The median profile shows isolated late spikes rather than periodic cycles.

LTF peaks early and decays gradually with a persistent tail (earlier t_peak, negative trend_slope_36; relatively high gini/peakiness around the early segment), consistent with durable but fading classics.

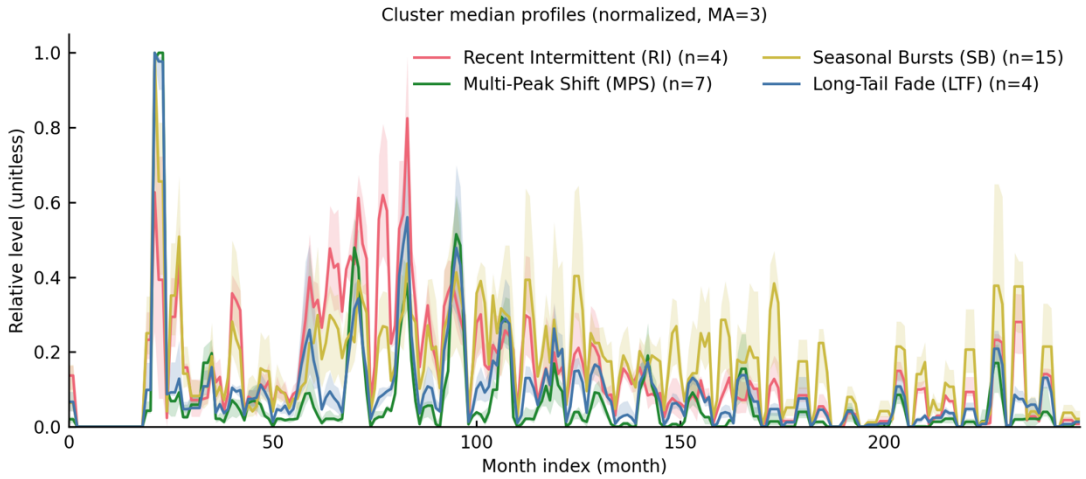

Figure 4. Cluster Median Profiles with IQR Bands

To connect statistics back to raw curves, Figure 5 shows medoid exemplars (closest to the cluster centroid in the standardized feature space) alongside a far-from-centroid boundary case for each mode. Medoids visualize the mode prototypes: SB with dense seasonal spikes; MPS with a visible break and subsequent peak groups; RI with late, intermittent bursts; LTF with an early apex and long fade. Boundary cases clarify why certain styles land in one mode

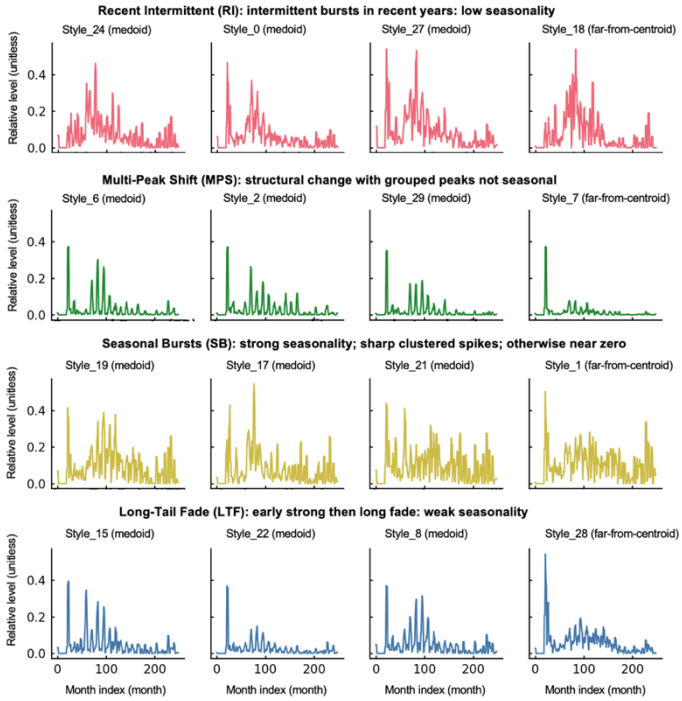

Figure 5. Representative Medoids and Boundary Cases by Lifecycle Mode

rather than another—for example, an MPS style with faint seasonality still aligns with the post-break peak grouping, not with SB's fixed annual cadence.

Overall, the four modes encode complementary regimes of timing (t_peak), concentration (gini, peakiness), recurrence (seas12, n_peaks, peak_spacing_cv), and structural change (cp_rel, cp_bic_gain). These signatures will be used to link modes to visual attributes and to build explainable predictors in subsequent sections (see Section 4.1; see Sections 3.3 and 4.1).

4.2 From Marginal to Conditional Association (within City × Month FE)

4.2.1 Marginal Associations: Modes × Visual Attributes

I begin with a simple baseline: in the full sample, do lifecycle modes co-occur with visual attributes at rates different from chance? To answer this, I construct mode-by-attribute contingency tables for the four attribute families (item, fabric, texture, color) using image-level records (see Section 3.2 and 4.1). Counting rule: *item is single-label per image*, whereas

fabric/texture/color are thresholded multi-label so an image may contribute more than one level within a family. For each family, I test the marginal independence hypothesis (*mode ⊥ attribute*) and report Pearson χ^2, Cramér's V (effect size), and cell-wise standardized residuals (z) together with lift = obs/exp. Multiple testing is controlled by BH–FDR within each family with $q<0.05$ as the main threshold; cells with obs < min_count (= 10) are shown but not tested.

Overall association strength is non-trivial across all families (see Table 1): item shows the largest association (χ^2 = 123,550; V=0.354), followed by fabric (χ^2 = 64,868; V = 0.256), texture (χ^2 = 28,302; V = 0.169), and color (χ^2 = 23,288; V = 0.154). This ordering aligns with industry intuition: categories and materials carry stronger structural differences in lifecycle behavior, while textures and colors provide additional—but somewhat smaller—discrimination.

attribute	χ^2	*p*-value	degrees of freedom (df)	sample size (N)	Cramér's V
item_std	123,550.26	< 1e-300	126	328,700	0.354
color_std	23,287.86	< 1e-300	48	328,700	0.154
texture_std	28,301.92	< 1e-300	69	328,700	0.169
fabric_std	64,867.78	< 1e-300	180	328,700	0.256

Table 1. Overall Association Strength by Attribute Family (Marginal Independence Test)

Overall association strength is non-trivial across all families (Table 4.2-A): item shows the largest association (χ^2 = 123,550; V=0.354), followed by fabric (χ^2 = 64,868; V = 0.256), texture (χ^2 = 28,302; V = 0.169), and color (χ^2 = 23,288; V = 0.154). This ordering aligns with industry intuition: categories and materials carry stronger structural differences in lifecycle behavior, while textures and colors provide additional—but somewhat smaller—discrimination.

To see which levels are over- or under-represented in each mode, Figure 6 visualizes standardized residuals (z, zero-centered; BH–FDR $q<0.05$ marked). Four directional patterns emerge and will be revisited under fixed-effects controls (see Section 4.2.2):

Color and Mode. LTF loads on blue/navy and avoids stark black/white minimalism; MPS

favors high-lightness / light hues (white, pink, green, cream); SB tilts toward black/khaki and is negative on blue; RI leans neutral (beige/cream/blue). Together these reflect a split between "classic/safe" blues, "light/bright" palettes, and "commuting neutrals." (see Figure 6, panel a)

Texture and Mode. MPS is positive on salient motifs (polka dot, graphic, stripe, floral/folk) and negative on tonal/sophisticated; SB prefers low-contrast tonal/checked and is negative on polka-dot/folk/nautical; RI is positive on sophisticated/folk/animal; LTF aligns with nautical/ombre/partial graphic elements. (see Figure 6, panel b)

Item and Mode. MPS co-occurs with light tops + casual bottoms (T-shirt, vest top, shorts); RI pairs with outerwear/accessories (blazer, jacket, coat, scarf); SB centers on trousers / functional bottoms; LTF appears more often with skirts/dresses/shorts/dungarees. (see Figure 6, panel c)

Fabric and Mode. MPS concentrates on cotton / decorative techniques (embroidery, tie-dye); SB favors functional basics (chino, canvas, leather, knit); RI is associated with tactile/warm materials (shearling, fur, knit); LTF shows a structural link to denim / blue family. (see Figure 6, panel d)

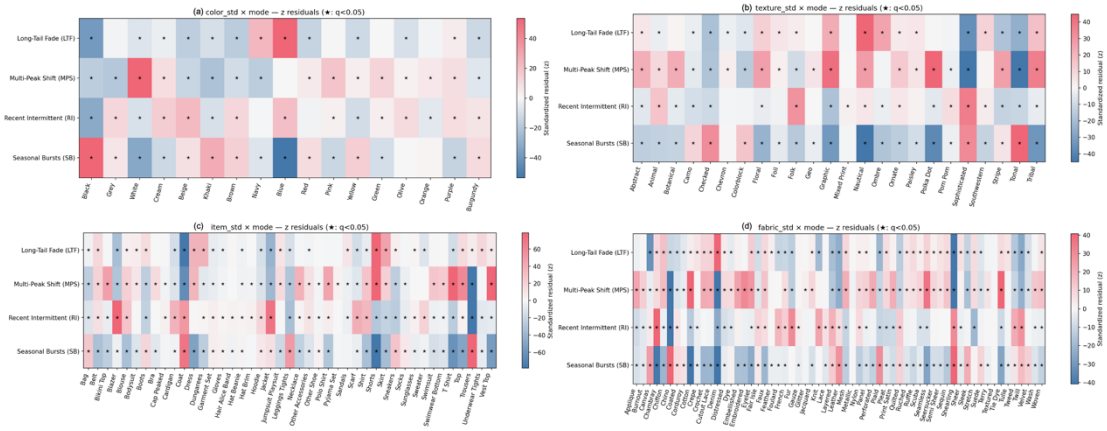

Figure 6. Marginal Standardized Residuals (z): Mode × Item/Color/Texture/Fabric

These marginal patterns provide a clear empirical backdrop but may still blend city composition and seasonality. The next subsection (see Section 4.2.2) therefore re-tests the same associations within city × month fixed effects, comparing within-stratum expected versus observed counts to verify whether these over/under-representations persist after controlling for place and time.

4.2.2 Conditional Association within City × Month Fixe

To separate genuine attribute–mode coupling from composition driven by where and when photos were taken, I re-estimate all associations within city × month fixed effects (FE).

The within-FE results track the marginal patterns (see Section 4.2.1) but with composition controlled, and effect sizes remain large (see Figure 7). To make magnitudes transparent, I highlight representative over- and under-representations (all computed within FE), organized by attribute family; panel references follow Figure 7.

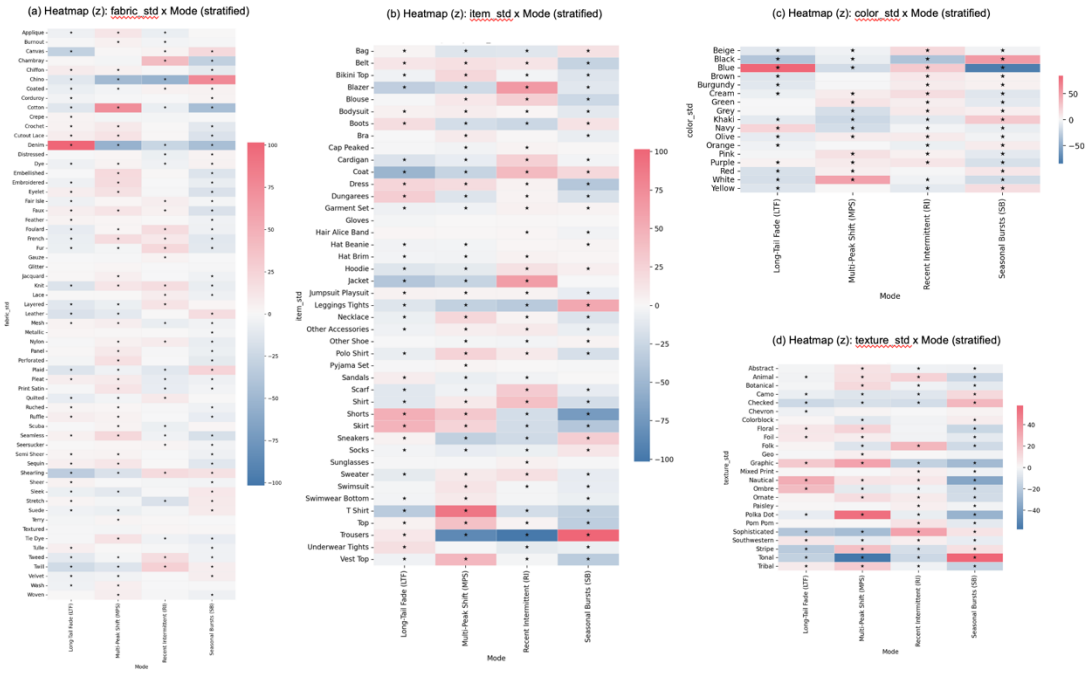

Figure 7. Stratified Heatmaps of Standardized Residuals within C

Fabric (see Figure 7, panel a; full Top-10 lists in Appendix A, Tables A1–A4) Over-representation. LTF–Denim: obs = 19,203, exp = 10,799, z = 101.6, lift = 1.78 ($q < 0.01$)—LTF aligns with durable "never-out" denim. SB–Chino: 32,626 vs 26,267, z = 76.9, lift = 1.24—seasonal bursts are trouser-fabric driven. MPS–Cotton: 4,942 vs 2,262, z = 73.7, lift = 2.18—MPS favors light, easy-to-pattern materials.

Under-representation. SB–Denim: 25,790 vs 29,696, z = −44.0, lift = 0.87—bursts are not denim-led. LTF–Canvas: 4,136 vs 6,030, z = −29.7, lift = 0.69—LTF skews casual rather than workwear. Together these contrasts map denim aligns with LTF, chinos with SB, and cotton/decorative fabrics with MPS.

Item (see Figure 7, panel b; full Top-10 lists in Appendix A, Tables A1–A4)
Over-representation. SB–Trousers: 53,941 vs 43,301, z = 101.5, lift = 1.25—SB concentrates in trouser surges. MPS–T-Shirt: 5,788 vs 2,371, z = 89.7, lift = 2.44; MPS–Vest-Top: 2,101 vs 929, z = 44.5, lift = 2.26—MPS leans into fast-cycling light tops. RI–Blazer: 4,226 vs 1,691, z = 65.8, lift = 2.50; RI–Jacket: 4,589 vs 2,122, z = 62.8, lift = 2.16—RI pulses via tailored outerwear. LTF–Shorts: 6,525 vs 3,754, z = 50.4, lift = 1.74; LTF–Skirt: 6,422 vs 4,004, z = 47.1, lift = 1.60—LTF prefers stable casual/feminine bottoms.

Under-representation. RI–Trousers: 548 vs 6,792, z = −99.3, lift = 0.08; MPS–Trousers: 672 vs 6,087, z = −86.8, lift = 0.11; SB–Shorts: 2,621 vs 6,328, z = −72.6, lift = 0.41—these negatives sharpen separation among modes.

Color (see Figure 7, panel c; full Top-10 lists in Appendix A, Tables A1–A4)
Over-representation. LTF–Blue: 12,352 vs 6,770, z = 84.9, lift = 1.82—LTF skews blue "safe tones." SB–Black: 52,588 vs 46,923, z = 53.8, lift = 1.12—black scales with seasonal bursts. MPS–White: 7,060 vs 4,370, z = 49.4, lift = 1.62 (pink/green/purple/cream also positive)—MPS prefers light/bright palettes.

Under-representation. SB–Blue: 11,469 vs 17,466, z = −79.9, lift = 0.66; LTF–Black: 10,686 vs 13,514, z = −32.5, lift = 0.79—SB is blue-negative; LTF is less black-centric.

Texture (see Figure 7, panel d; full Top-10 lists in Appendix A, Tables A1–A4)
Over-representation. SB–Tonal: 67,780 vs 61,131, z = 57.8, lift = 1.11—SB prefers low-contrast tonal. MPS–Polka-Dot: 1,657 vs 585, z = 53.9, lift = 2.83 (graphic/stripe/tribal also positive)—MPS favors salient motifs. RI–Sophisticated: 4,846 vs 3,240, z = 33.4, lift = 1.50—RI tilts toward refined/crafted looks. LTF–Nautical: 8,140 vs 6,130, z = 29.4, lift = 1.33—LTF carries nautical classicism.

Under-representation. MPS–Tonal: 4,418 vs 8,370, z = −56.9, lift = 0.53; SB–Nautical: 12,434 vs 15,151, z = −36.3, lift = 0.82; SB–Polka-Dot: 1,209 vs 2,090, z = −30.7, lift = 0.58—textures further sharpen the contrast between SB and MPS.

Taken together, the four families show directionally stable, statistically strong conditional links to lifecycle modes after controlling for city and month. These within-FE signals are therefore structural rather than compositional, and they ground the predictive validation

developed next and the explainable model evidence.

4.3 Predictive Validation under Time-Split

After establishing that visual attributes align with lifecycle modes once city × month composition is held fixed (see Section 4.2), the practical question is whether those signals help on tomorrow's data. To keep the test honest, I follow the time-consistent protocol in Section 3.4: rolling-origin splits (train in the past, test in the future), a multinomial logistic model with L2 regularization (saga), one-hot visual attributes plus city × month fixed effects (FE), and training-only fitting of encoders/FE dummies to prevent leakage. Baselines and staged ablations are identical to Section 3.4. Paired tests vs. baselines (p-values, effect sizes) appear in Appendix D, Table set D.3.3.

The headline result is encouraging. The full model (FE + item + color + texture + fabric) reaches Accuracy = 0.644 ± 0.025 (95% CI 0.613–0.675) and Macro-F1 = 0.411 ± 0.006 (95% CI 0.404–0.418), significantly beating Majority-class and Prior-sampling baselines (paired Wilcoxon, $p < 0.05$; see Figure 8). In plain terms: once FE absorbs where/when effects, visual attributes still add transferable signal about which lifecycle regime (SB/MPS/RI/LTF) a style belongs to. Fold-wise results are detailed in Appendix B, Table B1.

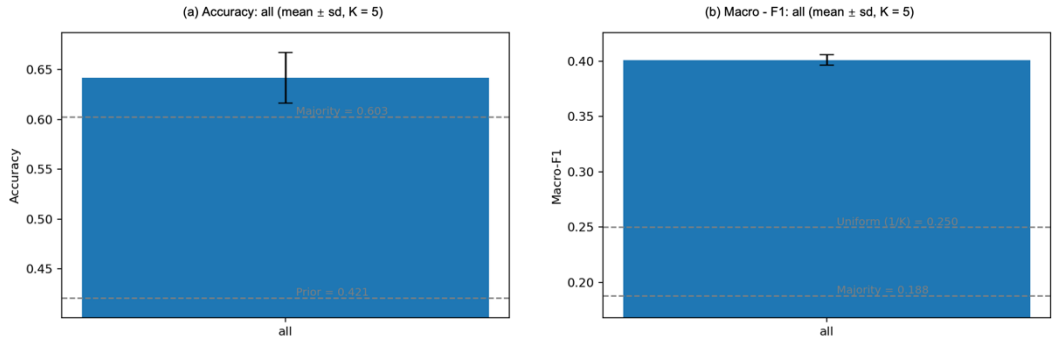

Figure 8. Model vs Baselines under Time-Split (mean ± SD; 95% CI). (a) Accuracy (b) Macro-F1

Where does the lift come from? The ablation tells a simple story. Performance increases monotonically as the specification progresses from FE-only to FE plus Item, then to FE plus Item and Color, then to FE plus Item, Color, and Texture, and finally to the All model including Fabric, with the All(+Fabric) configuration (0.642; 0.401) falling within the full model's confidence bands (see Figure 9). Relative to FE-only, Accuracy improves by approximately +6.5% and Macro-F1 by approximately +113%, mirroring the conditional

associations documented earlier (see Section 4.2). Put differently, item/type anchors the gain, color and texture refine it, and fabric tops it off—a decomposition that matches industry intuition about how styles materialize on the street.

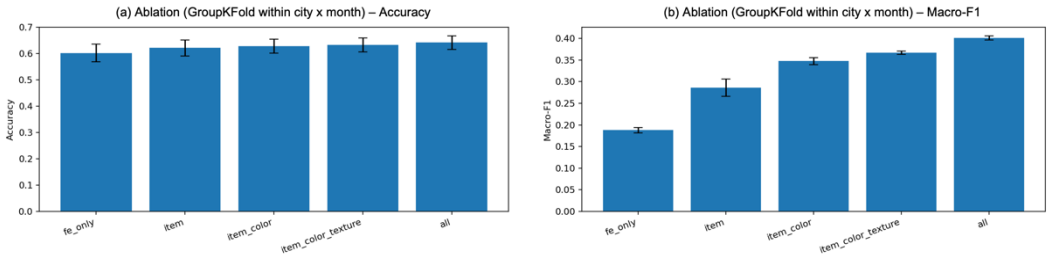

Figure 9. Ablation across Feature Sets (GroupKFold within FE). (a) Accuracy (b) Macro-F1

The error pattern is asymmetric, and that matters operationally. Confusion matrices show SB as a high-recall, high-absorption class, with misroutes from LTF to SB at approximately 0.67 and from RI to SB at approximately 0.83; MPS is comparatively stable, while RI and LTF remain harder to recover (see Section 4.4). This imbalance explains why Macro-F1 gains outpace Accuracy gains: the model improves more on non-dominant modes than on the popular one.

Probability quality is good enough to use. Overall calibration is satisfactory (Appendix C: Table C1 for Brier; Table C2 for per-class ECE/MCE), which enables two lightweight deployment levers without retraining: selective-risk (abstain on low confidence) to raise precision on a smaller coverage slice (Appendix C: Tables C4 and Figures C2–C3), and an SB-gate two-stage router to curb SB absorption and boost macro performance (Appendix C: Table C5 and Figures C4–C5). These knobs are pragmatic in settings where content surfacing or buying decisions benefit from confidence-aware routing.

Finally, a scope note: this is non-causal validation. FE mitigates composition bias but leaves prices, brands, promotions, and audience shifts in the residual. Causal claims are beyond the present design and are left to future work.

4.4 Explainable Prediction & Error Diagnosis

Where do the gains actually come from, and where do errors go? To open up the black box, I inspect the fitted coefficients of the multinomial logistic model under the main setting (FE + visual attributes; see Section 3.4) and relate them to the conditional association patterns (see

Section 4.2).

Coefficient evidence. I plot coefficient forests for each attribute family (item, color, texture, fabric), averaging per-class softmax coefficients across outer time splits (mean ± SD) and marking fold-wise z-scores with BH–FDR within family. In this parameterization, a positive

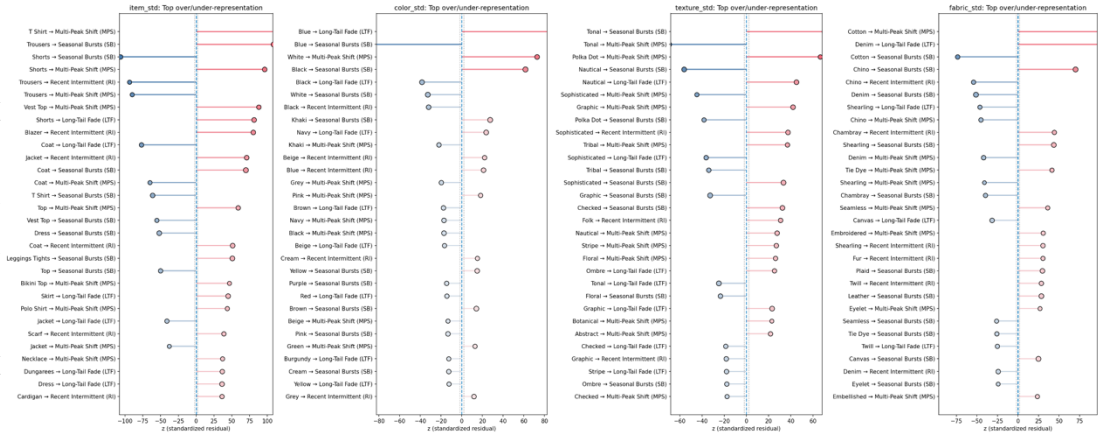

Figure 10. Coefficient forests by attribute family (Multinomial Logit; mean ± SD across folds)

Error structure. Next, I aggregate predictions across the outer time splits and display confusion matrices as counts and row-normalized recall (see Figure 11). Two facts stand out. First, SB attains the highest recall and acts as an "absorber": many misroutes from RI and LTF end up as SB (e.g., from RI to SB at approximately 0.83; from LTF to SB at approximately 0.67 in the row-normalized panel). Second, RI and LTF remain the hardest to recover, while MPS sits in the middle. This asymmetric error explains why Macro-F1 improves more than Accuracy in the time-split validation (see Section 4.3): the model becomes better at the non-dominant classes even when the dominant SB stays strong.

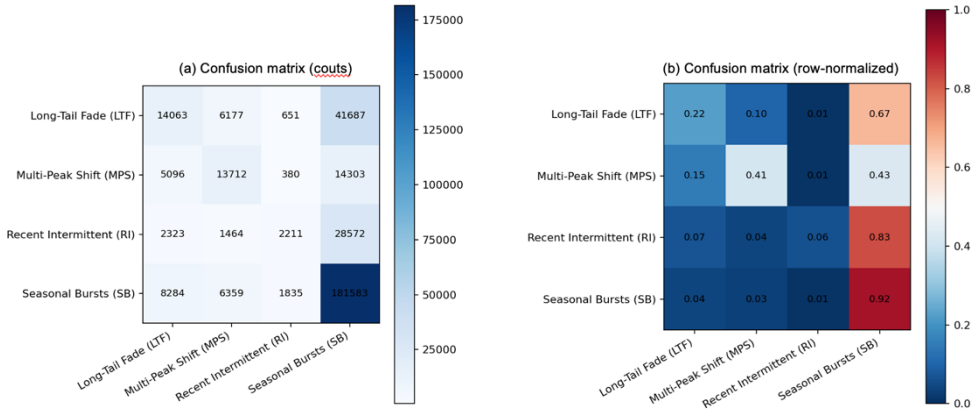

Figure 11. Confusion matrices under time-split. (a) counts (b) row-normalized

Deployment notes. The probability outputs are reasonably calibrated (Appendix C: Table C1 for Brier; Table C2 for per-class ECE/MCE; Figure C1 & Table C3 for reliability bins). Two lightweight policies are therefore practical without retraining: (i) selective risk via a confidence threshold to boost accuracy on a high-confidence subset; (ii) an SB-gate that routes samples with $P(SB) \geq \tau$ directly to SB and otherwise refines within {LTF, MPS, RI}. Both strategies reduce SB absorption and raise macro-level performance; operational curves and thresholds are provided in Appendix C (Tables C1–C5; Figures C1–C5).

Limitations. As throughout the paper, this is non-causal identification: FE mitigates composition bias but does not control prices, brand mix, promotion, or audience shifts; causal claims are left for future designs (e.g., natural experiments, IV, or DiD).

5. Conclusion and Discussion

This paper maps the temporal shape of street-style fashion across ten cities over two decades and tests whether visible, interpretable attributes retain predictive signal once place–time composition is held constant. The findings cohere into a compact narrative. A city-comparable lifecycle atlas shows that style activity follows recurring patterns consistent with classic accounts of imitation, distinction, and collective selection, together with diffusion backbones that yield rise–peak–decline curves (Blumer, 1969; Rogers, 2003; Bass, 1969; Simmel, 1957; Veblen, 1994). Four recurrent modes summarize these dynamics, and visible attributes align with the modes inside city × month strata while also improving forward prediction. I bring theory, measurement, and prediction into one identification-first pipeline.

The results answer the research questions directly and cumulatively. For Q1, the two-decade atlas built from approximately 328,700 WGSN images reveals population-level swells around 2009 and 2016 and substantial heterogeneity in peak timing and width, a contrast that matches differences in imitation pressure, novelty inputs, and replacement dynamics emphasized by diffusion models and seasonal gating (Bass, 1969; Crane, 2012). For Q2, clustering fifteen lifecycle descriptors yields four modes with clear signatures. Seasonal Bursts express strong twelve-month rhythms and sharp seasonal peaks. Multi-Peak Shift exhibits a late regime change followed by dense non-seasonal peaks. Recent Intermittent concentrates high-amplitude bursts in recent years with weak seasonality. Long-Tail Fade peaks early and then persists at low amplitude. These shapes are the kind of limited repertoire expected when adoption is threshold-gated, periodically synchronized by calendars, and periodically re-

ignited by new meanings or cohorts (Watts, 2002; Centola, 2018; Crane, 2012). For Q3, within city × month fixed effects and using stratified permutation tests with false-discovery control, attributes remain directionally stable and statistically strong. Denim and blue correspond to persistent long-tail fades. Light and bright palettes correspond to shifts and bursts. Trousers and chino fabrics correspond to seasonal bursts. Outerwear and warm or insulating materials correspond to recent intermittent activity (Benjamini & Hochberg, 1995; Pesarin & Salmaso, 2012). In forward-rolling evaluations, adding interpretable attributes to fixed effects improves accuracy and macro-F1, and calibrated probabilities support confidence-aware use (Guo, Pleiss, Sun, & Weinberger, 2017; Hyndman & Athanasopoulos, 2021).

The four modes reflect social mechanisms rather than statistical artifacts. Seasonal Bursts arise when calendars compress attention and temporarily lower adoption thresholds so imitation propagates quickly and then recedes when the window closes (Watts, 2002; Crane, 2012). Multi-Peak Shift is consistent with renewed meanings such as sustainability frames, nostalgia, or cross-scene collaborations that reopen cascade windows for fresh cohorts and raise effective innovation and imitation inputs in a new phase (Niinimäki et al., 2020; Sedikides, Wildschut, Routledge, & Arndt, 2015; Rogers, 2003). Long-Tail Fade reflects replacement-led consumption and habit formation once a look becomes infrastructural rather than status-acute, sustaining a low-amplitude tail (Bass, 1969; Blumer, 1969). Recent Intermittent matches clustered contagion in networked niches, where reinforcement is strong inside scenes but cross-scene bridges are weak, leading to ripples rather than system-wide cascades (Centola, 2018).

The alignment between attributes and modes can be explained by function, durability, and attention. Items and fabrics are bound to thermal and structural affordances that recur seasonally, and they incur higher replacement and fit costs, which slows turnover and stabilizes placement in the calendar. Colors and textures are affective and agile layers that can be changed within the same garment archetype to signal mood and micro-shifts without rebuilding silhouette or sourcing. This slow-code versus fast-code split predicts that items and fabrics will track seasonal placement and long-run persistence more strongly than colors and textures. I observe precisely these directions inside city × month strata, which indicates structure in how looks are composed within the same local context rather than only where and when photos were taken (Li, 2001). See also the fixed-effects Top-10 summaries in Appendix A (Tables A1–A4).

The study contributes both methodologically and theoretically. Substantively, the analysis demonstrates that a small repertoire of lifecycle shapes recurs in street style and that those shapes are semantically grounded in visible attributes legible to observers and wearers. Methodologically, the pipeline advances a shape-first atlas, identifies structural associations with fixed-effects stratification and permutation-based inference under false-discovery control, and validates generalization under forward-rolling splits with calibrated probabilities. This combination separates structural association from composition and offers a template for studying other visual cultures that unfold across places and seasons (Angrist & Pischke, 2009; Wooldridge, 2010; Benjamini & Hochberg, 1995; Pesarin & Salmaso, 2012).

The findings have immediate practical implications across merchandising, creative direction, and content operations. Assortment timing and exposure can be aligned with Seasonal Bursts to exploit calendar gates when conversion is highest (Crane, 2012; Ferreira, Lee, & Simchi-Levi, 2016). Carry-over bets can lean on fabrics and palettes linked to persistence, such as denim and blue within Long-Tail Fades. Creative direction can use light and bright palettes to mark shifts and bursts without rebuilding silhouettes. Calibrated probabilities enable confidence-aware routing that accepts high-certainty cases and abstains on ambiguous ones, yielding a clear risk–coverage trade-off for buying and content workflows (Guo et al., 2017; El-Yaniv & Wiener, 2010). Because fitted coefficients mirror the conditional associations, decision-makers retain interpretability when auditing why a mode label is suggested.

Several checks support the robustness of the core patterns. Descriptor-space clustering remains qualitatively stable under perturbations, with Adjusted Rand Index medians that preserve the block-diagonal structure of Seasonal Bursts and Multi-Peak Shift. Sensitivity to the number of styles retained, such as keeping the top 25 or the top 35, leaves the four-mode picture intact. Within-stratum associations are screened by stratified permutations and false-discovery control, and predictive gains are localized by ablations that add attribute families in stages. Probabilities are calibrated to correct overconfidence, which improves the reliability of selective policies (Guo et al., 2017).

Several limitations bound the claims and suggest caution in generalization. The identification strategy is non-causal. Fixed effects remove city-level time-invariant differences and soak up coarse calendar shocks but cannot purge time-varying confounders such as price, brand actions, promotions, or shifts in audience composition (Wooldridge, 2010). The WGSN lens emphasizes high-attention streets and fashion-week periods, which may under-sample quieter contexts. Zero-shot attribute inference introduces measurement noise, and collapsing multi-

label families to a single top label simplifies nuanced looks. The final year is partial, so late declines may mix genuine cooling with truncation. These constraints bound external validity and indicate where policy extrapolation should proceed carefully.

Future research can build on this baseline in several directions. Causal designs that leverage exogenous shocks, instruments, or staggered adoptions could test mechanisms behind bursts and revivals more directly. Dynamic mode models such as state-space or hidden-Markov approaches could allow styles to switch modes over time. Richer attributes including silhouette, fit, and multi-label palettes would capture more of the fast-code layer. External validation on other platforms and cities, together with price, brand, and marketing covariates, would expand scope and probe transportability. Human-in-the-loop audits could formalize how domain experts read coefficients and residuals in practice.

In closing, the study provides a two-decade, city-comparable lifecycle atlas of street-style fashion, identifies four canonical modes that recur across time and place, and shows that visible, interpretable attributes carry mode signal that survives place–time controls and improves forward prediction. By integrating classic sociological theory with modern computer vision and identification-first panel modeling, I offer a baseline that others can extend conceptually, methodologically, and operationally for understanding how time wears on style (Blumer, 1969; Rogers, 2003; Bass, 1969; Aral, 2020).

Reference

Abernathy, F. H., Dunlop, J., Hammond, J. H., & Weil, D. (1999). A stitch in time: Lean retailing and the transformation of manufacturing—Lessons from the apparel and textile industries. Oxford University Press.

Al-Halah, Z., & Grauman, K. (2020). From Paris to Berlin: Discovering fashion style influences around the world. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 10133–10142). IEEE. https://doi.org/10.1109/CVPR42600.2020.01015

Al-Halah, Z., Stiefelhagen, R., & Grauman, K. (2017). Fashion forward: Forecasting visual style in fashion. In Proceedings of the IEEE International Conference on Computer Vision (ICCV).

Angrist, J. D., & Pischke, J.-S. (2009). Mostly harmless econometrics: An empiricist's

companion. Princeton University Press.
Aral, S. (2020). The hype machine: How social media disrupts our elections, our economy, and our health—and how we must adapt. Currency.
Barnard, M. (2003). Fashion as communication (2nd ed.). Routledge. https://doi.org/10.4324/9781315013084
Barthes, R. (1983). The fashion system. Hill and Wang.
Bass, F. M. (1969). A new product growth model for consumer durables. Management Science, 15(5), 215–227. http://www.jstor.org/stable/2628128
Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B (Methodological), 57(1), 289–300.
Blumer, H. (1969). Fashion: From class differentiation to collective selection. The Sociological Quarterly, 10(3), 275–291. http://www.jstor.org/stable/4104916
Cachon, G. P., & Swinney, G. (2011). The value of fast fashion: Quick response, enhanced design, and strategic consumer behavior. Management Science, 57(4), 778–795.
Centola, D. (2018). How behavior spreads: The science of complex contagions. Princeton University Press. https://doi.org/10.2307/j.ctvc7758p
Chia, P. J., Attanasio, G., Bianchi, F., Terragni, S., Magalhães, A. R., Goncalves, D., Greco, C., & Tagliabue, J. (2022). Contrastive language and vision learning of general fashion concepts. Scientific Reports, 12(1), 18958. https://doi.org/10.1038/s41598-022-23052-9
Christopher, M., Lowson, R., & Peck, H. (2004). Creating agile supply chains in the fashion industry. International Journal of Retail & Distribution Management, 32(8), 367–376.
Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A seasonal-trend decomposition procedure based on LOESS. Journal of Official Statistics, 6(1), 3–73.
Crane, D. (2012). Fashion and its social agendas: Class, gender, and identity in clothing. University of Chicago Press.
Davis, F. (1992). Fashion, culture, and identity. University of Chicago Press.
Elliot, A. J., & Maier, M. A. (2014). Color psychology: Effects of perceiving color on psychological functioning in humans. Annual Review of Psychology, 65, 95–120.
El-Yaniv, R., & Wiener, Y. (2010). On the foundations of noise-free selective classification. Journal of Machine Learning Research, 11, 1605–1641.
Ferreira, K. J., Lee, B. H. A., & Simchi-Levi, D. (2016). Analytics for an online retailer: Demand forecasting and price optimization. Manufacturing & Service Operations Management, 18(1), 69–88.
Fisher, M. L. (1997). What is the right supply chain for your product? Harvard Business Review, 75(2), 105–116.

Frome, A., Corrado, G. S., Shlens, J., Bengio, S., Dean, J., Ranzato, M., & Mikolov, T. (2013). DeViSE: A deep visual-semantic embedding model. In Advances in Neural Information Processing Systems (Vol. 26, pp. 2121–2129).

Fulcher, B. D., Little, M. A., & Jones, N. S. (2013). Highly comparative time-series analysis: The empirical structure of time series and their methods. Journal of the Royal Society Interface, 10(83), 20130048. https://doi.org/10.1098/rsif.2013.0048

Ghemawat, P., & Nueno, J. L. (2006). Zara: Fast fashion (Harvard Business School Case 9-703-497). Harvard Business School Publishing.

Gronow, J. (1993). Taste and fashion: The social function of fashion and style. Acta Sociologica, 36(2), 89–100.

Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. In Proceedings of the 34th International Conference on Machine Learning (ICML) (pp. 1321–1330).

Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and practice (2nd ed.). OTexts. https://otexts.org/fpp2/

Hyndman, R. J., & Athanasopoulos, G. (2021). Forecasting: Principles and practice (3rd ed.). OTexts. OTexts.com/fpp3

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. Nature Reviews Neuroscience, 2(3), 194–203. https://doi.org/10.1038/35058500

Jin, S. V., Muqaddam, A., & Ryu, E. (2019). Instafamous and social media influencer marketing. Marketing Intelligence & Planning, 37(5), 567–579. https://doi.org/10.1108/MIP-09-2018-0375

Kawamura, Y. (2018). Fashion-ology: An introduction to fashion studies (2nd ed.). Bloomsbury.

Li, Y. (2001). The science of clothing comfort. Textile Progress, 31(1–2), 1–135. https://doi.org/10.1080/00405160108688951

Liu, Z., Luo, P., Qiu, S., Wang, X., & Tang, X. (2016). DeepFashion: Powering robust clothes recognition and retrieval with rich annotations. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 1096–1104). https://doi.org/10.1109/CVPR.2016.124

Mall, U., Matzen, K., Hariharan, B., Snavely, N., & Bala, K. (2019). GeoStyle: Discovering fashion trends and events. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (pp. 411–420).

Miller, D., & Woodward, S. (2012). Blue jeans: The art of the ordinary. University of California Press. http://www.jstor.org/stable/10.1525/j.ctt1pntfr

Niinimäki, K., Peters, G., Dahlbo, H, Perry, P., Rissanen, T., & Gwilt, A. (2020). The

environmental price of fast fashion. Nature Reviews Earth & Environment, 1(4), 189–200.
Palmer, S. E., & Schloss, K. B. (2010). An ecological valence theory of human color preference. Proceedings of the National Academy of Sciences, 107(19), 8877–8882. https://doi.org/10.1073/pnas.0906172107
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830.
Pesarin, F., & Salmaso, L. (2012). Permutation tests for complex data: Theory, applications and software. Wiley.
Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. In Proceedings of the 38th International Conference on Machine Learning (ICML) (pp. 8748–8763). PMLR.
Rogers, E. M. (2003). Diffusion of innovations (5th ed.). Free Press.
Rymes, B. (2003). A matter of taste: How names, fashions, and culture change [Review of the book A matter of taste: How names, fashions, and culture change, by S. Lieberson]. Journal of Linguistic Anthropology, 13(2), 256–258. https://doi.org/10.1525/jlin.2003.13.2.256
Sedikides, C., Wildschut, T., & Baden, D. (2004). Nostalgia: Conceptual issues and existential functions. In J. Greenberg, S. L. Koole, & T. Pyszczynski (Eds.), Handbook of experimental existential psychology (pp. 200–214). The Guilford Press.
Simmel, G. (1957). Fashion. American Journal of Sociology, 62(6), 541–558.
Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., … Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 1–9). https://doi.org/10.1109/CVPR.2015.7298594
Tian, X., Pan, Y., & Yu, H. (2021). Weather, consumer spending, and retail operations. Journal of Retailing and Consumer Services, 60, 102453.
https://doi.org/10.1016/j.jretconser.2021.102453
Valdez, P., & Mehrabian, A. (1994). Effects of color on emotions. Journal of Experimental Psychology: General, 123(4), 394–409. https://doi.org/10.1037/0096-3445.123.4.394
Veblen, T. (1994). The theory of the leisure class. Penguin Books. (Original work published 1899)
Watts, D. J. (2002). A simple model of global cascades on random networks. Proceedings of the National Academy of Sciences, 99(9), 5766–5771.
Wood, W., & Neal, D. T. (2016). Healthy through habit: Interventions for initiating & maintaining health behavior change. Behavioral Science & Policy, 2(1), 71–83.

https://doi.org/10.1177/237946151600200109
Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data (2nd ed.). MIT Press.

Appendix A

Table A1. Fabric × Mode — Fixed-Effects Conditional Standardized Residuals (Top-10, over-represented)

Mode	Level (Fabric)	Obs	Exp	z	p	q	Lift
LTF	Denim	19,203	10,799	101.60	0.005	0.008	1.78
SB	Chino	32,626	26,267	76.92	0.005	0.008	1.24
MPS	Cotton	4,942	2,262	73.68	0.005	0.008	2.18
RI	Chambray	1,321	462	41.16	0.005	0.008	2.86
RI	Twill	2,013	1,215	27.20	0.005	0.008	1.66
SB	Plaid	6,828	5,816	23.82	0.005	0.008	1.17
RI	Fur	852	423	21.69	0.005	0.008	2.02
MPS	Seamless	658	337	21.64	0.005	0.008	1.95
SB	Canvas	21,950	20,279	20.86	0.005	0.008	1.08
MPS	Embellished	732	375	20.05	0.005	0.008	1.95

Table A2. Item × Mode — Fixed-Effects Conditional Standardized Residuals (Top-10, over-represented)

Mode	Level (Item)	Obs	Exp	z	p	q	Lift
SB	Trousers	53,941	43,301	101.48	0.005	0.006	1.25
MPS	T Shirt	5,788	2,371	89.74	0.005	0.006	2.44
RI	Blazer	4,226	1,691	65.80	0.005	0.006	2.50
RI	Jacket	4,589	2,122	62.77	0.005	0.006	2.16
SB	Leggings Tights	10,439	7,809	54.90	0.005	0.006	1.34
LTF	Shorts	6,525	3,754	50.37	0.005	0.006	1.74
LTF	Skirt	6,422	4,004	47.08	0.005	0.006	1.60
MPS	Vest Top	2,101	929	44.51	0.005	0.006	2.26
RI	Coat	7,319	4,972	42.63	0.005	0.006	1.47
RI	Shirt	1,086	379	38.81	0.005	0.006	2.86

Table A3. Color × Mode — Fixed-Effects Conditional Standardized Residuals (Top-10, over-represented)

Mode	Level (Color)	Obs	Exp	z	p	q	Lift
LTF	Blue	12,352	6,770	84.88	0.005	0.006	1.82
SB	Black	52,588	46,923	53.79	0.005	0.006	1.12
MPS	White	7,060	4,370	49.41	0.005	0.006	1.62
SB	Khaki	20,896	18,937	26.75	0.005	0.006	1.10
RI	Blue	4,385	3,157	26.21	0.005	0.006	1.39
LTF	Navy	10,094	8,516	20.15	0.005	0.006	1.19
RI	Beige	1,500	968	18.43	0.005	0.006	1.55
RI	Cream	736	452	15.19	0.005	0.006	1.63
SB	Yellow	4,972	4,321	15.14	0.005	0.006	1.15
MPS	Pink	1,243	909	13.22	0.005	0.006	1.37

Table A4. Texture × Mode — Fixed-Effects Conditional Standardized Residuals (Top-10, over-represented)

Mode	Level (Texture)	Obs	Exp	z	p	q	Lift
SB	Tonal	67,780	61,131	57.76	0.005	0.007	1.11
MPS	Polka Dot	1,657	585	53.92	0.005	0.007	2.83
RI	Sophisticated	4,846	3,240	33.37	0.005	0.007	1.50
MPS	Graphic	7,394	5,455	32.97	0.005	0.007	1.36
LTF	Nautical	8,140	6,130	29.45	0.005	0.007	1.33
RI	Folk	2,401	1,520	25.92	0.005	0.007	1.58
SB	Checked	9,418	8,190	25.34	0.005	0.007	1.15
LTF	Ombre	1,343	814	23.23	0.005	0.007	1.65
MPS	Stripe	3,027	2,117	21.68	0.005	0.007	1.43
LTF	Graphic	10,864	9,466	18.34	0.005	0.007	1.15

Notes: Computed within city×month fixed-effects strata; expectations aggregated across strata. Columns—obs: observed count; exp: expected under FE independence; z: standardized residual; p: two-sided empirical p from stratified permutation (B=200); q: BH–FDR within family; lift: obs/exp. Displayed entries are Top-10 over-represented pairs per family.

Appendix B

Table B1. Cross-Validation Metrics by Fold

Fold	Accuracy	Macro-F1	Log Loss	N
1	0.6819744447824764	0.41309904898772143	0.8054677464152569	65740
2	0.6328262853665957	0.4196437438629691	0.8845674628386141	65740
3	0.6478399756616976	0.4072303571697007	0.8544276647691569	65740
4	0.642135686035899	0.40983919939404256	0.8670871695635379	65740
5	0.6134925463948889	0.40481150611995015	0.9128928425945604	65740

Table B2. Per-Class Metrics

Class	Precision	Recall	F1-score	Support
LTF	0.4724517906336088	0.2247275400300425	0.3045785324434685	62578
MPS	0.49480369515011546	0.4094234271893942	0.44808261032906344	33491
RI	0.43549340161512695	0.06395718831356667	0.1115342901099994	34570
SB	0.6822709425313269	0.916803409050747	0.7823380137261751	198061

Table B3. Coefficient Highlights by Mode & Feature Group

Mode	Feature Group	Top Positive (Level; z)	Top Negative (Level; z)
LTF	Item	Skirt (z≈96.71)	Jacket (z≈−112.88)
LTF	Color	Blue (z≈32.63)	Burgundy (z≈−6.54)
LTF	Texture	Foil (z≈8.93)	Stripe (z≈−14.43)
LTF	Fabric	Eyelet (z≈25.67)	Tweed (z≈−24.48)
MPS	Item	Polo Shirt (z≈79.22)	Boots (z≈−37.14)
MPS	Color	White (z≈30.53)	Grey (z≈−0.48)
MPS	Texture	Stripe (z≈30.67)	Tonal (z≈−24.73)
MPS	Fabric	Cotton (z≈20.35)	Shearling (z≈−105.52)
RI	Item	Blazer (z≈107.68)	Trousers (z≈−58.36)
RI	Color	Blue (z≈3.99)	Black (z≈−42.16)
RI	Texture	Folk (z≈46.56)	Botanical (z≈−29.65)
RI	Fabric	Quilted (z≈35.81)	Print Satin (z≈−7.53)
SB	Item	Coat (z≈63.19)	Vest Top (z≈−71.10)
SB	Color	Yellow (z≈17.90)	Blue (z≈−37.17)

Mode	Feature Group	Top Positive (Level; z)	Top Negative (Level; z)
SB	Texture	Sophisticated (z≈40.01)	Nautical (z≈−24.74)
SB	Fabric	Shearling (z≈11.93)	Nylon (z≈−36.82)

Abbreviations: LTF = Long-Tail Fade; MPS = Multi-Peak Shift; RI = Recent Intermittent; SB = Seasonal Bursts.

Appendix C

Table C1. Multiclass Brier Score

Metric	Value
Brier_multiclass	0.4790

Notes: Lower is better; value computed on the out-of-sample evaluation set.

Table C2. Per-Class Calibration (ECE/MCE)

Class	ECE	MCE
Long-Tail Fade (LTF)	0.0078	0.2077
Multi-Peak Shift (MPS)	0.0066	0.1820
Recent Intermittent (RI)	0.0013	0.0612
Seasonal Bursts (SB)	0.0100	0.0296

Notes: ECE = Expected Calibration Error; MCE = Maximum Calibration Error. Lower is better.

Table C3. Overall Reliability Bins

Bin Confidence	Bin Accuracy	Gap	Count
0.0500	—	—	0
0.1500	—	—	0
0.2871	0.3046	0.0175	604
0.3644	0.3543	0.0101	21,946
0.4536	0.4430	0.0107	49,354
0.5483	0.5403	0.0080	58,821
0.6528	0.6365	0.0163	55,551
0.7505	0.7454	0.0051	65,585
0.8445	0.8531	0.0086	72,469

Bin Confidence	Bin Accuracy	Gap	Count
0.9146	0.9165	0.0019	4,370

Note: “Gap” = |Bin Accuracy − Bin Confidence|. Rows with Count = 0 are carried over from the source file and omitted from metric aggregation.

Figure C1. Overall Reliability Diagram

Table C4. Abstention Curve (threshold τ on max-probability)

τ	Coverage	Accuracy	Macro-F1
0.50	0.7812	0.7082	0.4153
0.55	0.6853	0.7344	0.3956
0.60	0.6023	0.7581	0.3908
0.65	0.5237	0.7802	0.3852
0.70	0.4333	0.8055	0.3863
0.75	0.3321	0.8347	0.3705
0.80	0.2338	0.8567	0.3463
0.85	0.1046	0.8894	0.3356
0.90	0.0133	0.9165	0.3773
0.95	0.0000	1.0000	0.5000

Notes: Coverage = fraction of kept predictions after abstention. Curves visualized in the two PNG files.

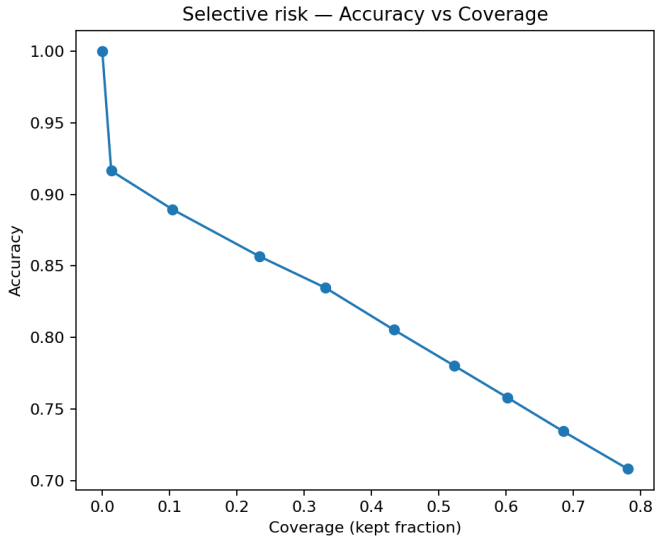

Figure C2. Accuracy vs. τ

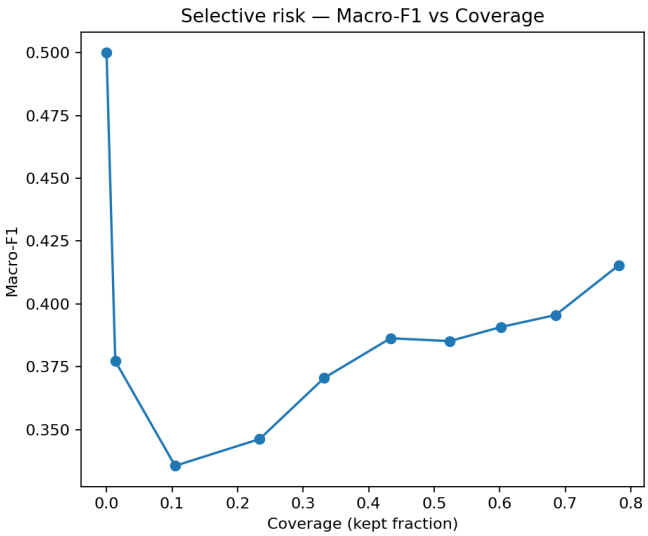

Figure C3. Macro-F1 vs. τ

Table C5. SB-Gate Curve (selective blocking by Seasonal Bursts score τ_sb)

τ_sb	Accuracy	Macro-F1	SB Rate
0.30	0.6364	0.3549	0.8758
0.35	0.6392	0.3811	0.8379
0.40	0.6385	0.4036	0.7987
0.45	0.6359	0.4279	0.7522
0.50	0.6309	0.4528	0.6945
0.55	0.6193	0.4693	0.6333
0.60	0.6007	0.4795	0.5676
0.65	0.5774	0.4832	0.5008
0.70	0.5428	0.4778	0.4174
0.75	0.4946	0.4594	0.3264
0.80	0.4351	0.4279	0.2319
0.85	0.3464	0.3677	0.1040
0.90	0.2742	0.3053	0.0131
0.95	0.2629	0.2943	0.0000*

Notes : At τ_sb = 0.95, SB Rate ≈ 1.22e−05 (rounded to 0.0000 in the table). "SB Rate" is the share gated by the Seasonal-Bursts rule.

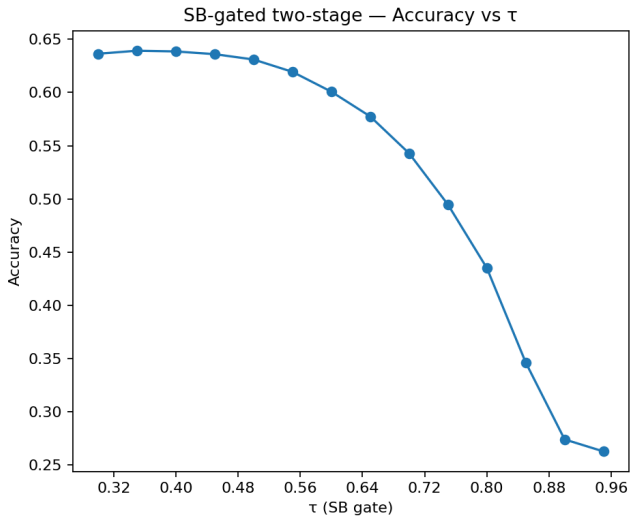

Figure C4. Accuracy vs. τ_sb

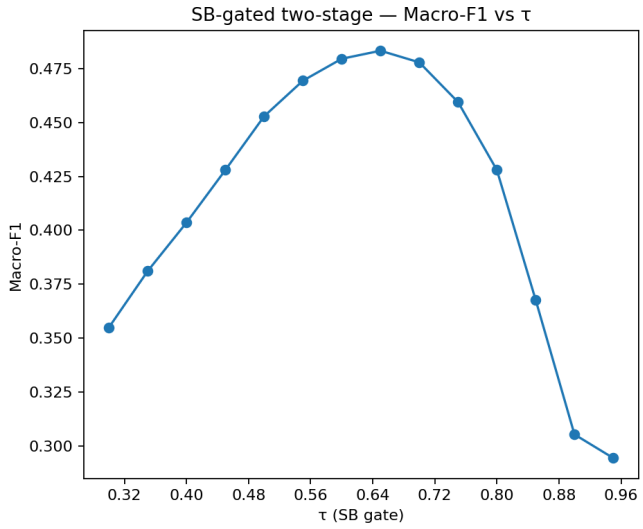

Figure C5. Macro-F1 vs. τ_sb

Appendix D

D.1 Run Summary

Field	Value
Inputs	outputs/cond_pred_local
Outputs	outputs/cond_pred_local/interp_uncert_pack
Produced artifacts (by	**CV uncertainty & tests:** cv_uncertainty.json, overall_vs_baselines_*_uncert.png • **Confusion & per-class:**

Field	Value
module)	confusion_counts.png, confusion_rownorm.png, per_class_metrics.csv • **Coef FDR & forest:** coef_with_q.csv, coef_forest_*.png • **Calibration:** reliability_overall.png, reliability_*.png, ece_per_class.csv, brier_score.csv • **Thresholds:** Abstention → abstention_curve.csv (+ accuracy/macroF1 vs coverage figures); SB gate → sb_gate_curve.csv, sb_gate_best.json (+ figures)
Baselines used	Accuracy: majority = 0.603, prior = 0.421 • Macro-F1: majority = 0.18808484092326885, uniform = 0.25

Notes: All file paths are relative to the project root. "*overall_vs_baselines*__uncert.png*" contains model vs. baseline uncertainty visuals for both accuracy and Macro-F1.

D.2 Experiment Configuration

Key	Value
prepared	items_preprocessed.csv
outdir	outputs/cond_pred_local
alpha	0.05
B (permutation reps)	200
k (CV folds)	5
with_fe	false
Cgrid (L2 inverse-reg strength)	[0.5, 1.0, 2.0]
seed	123

Notes: with_fe=false indicates city×month fixed effects are handled via modeling choices in the main text; hyperparameter C is selected from Cgrid via inner CV.

D.3 Cross-Validation Results

Metric	Mean	SD	95% CI (lower, upper)
Accuracy	0.6437	0.0251	(0.6125, 0.6748)
Macro-F1	0.4109	0.0058	(0.4038, 0.4181)

D.3.2 Baselines

Baseline	Accuracy	Macro-F1
Majority class	0.603	0.18808484092326885

Baseline	Accuracy	Macro-F1
Prior sampling	0.421	—
Uniform sampling	—	0.25

D.3.3 Paired Tests vs. Baselines

Comparison	p	Effect size *d*	Model mean	Baseline	Mean diff
Accuracy vs Majority	0.022255	1.6212	0.6437	0.6030	0.0407
Accuracy vs Prior	3.80e-05	8.8791	0.6437	0.4210	0.2227
Macro-F1 vs Majority	1.07e-07	38.6530	0.4109	0.1881	0.2228
Macro-F1 vs Uniform	3.95e-07	27.9134	0.4109	0.2500	0.1609

Notes: Tests are paired across outer CV splits. The effect size *d* is computed on split-wise differences.

D.4 Conditional Metadata

D.4.1 Shared Mode Set

["Long-Tail Fade (LTF)", "Multi-Peak Shift (MPS)", "Recent Intermittent (RI)", "Seasonal Bursts (SB)"]

D.4.2 Family Summaries

Family	#Levels	Sample size n
item_std	43	328,700
color_std	17	328,700
texture_std	24	328,700
fabric_std	61	328,700

Notes: Each family uses the same four lifecycle modes. Counts reflect distinct category levels included in modeling.

D.4.3 Levels by Family

Items (item_std): Bag; Belt; Bikini Top; Blazer; Blouse; Bodysuit; Boots; Bra; Cap Peaked; Cardigan; Coat; Dress; Dungarees; Garment Set; Gloves; Hair Alice Band; Hat Beanie; Hat Brim; Hoodie; Jacket; Jumpsuit Playsuit; Leggings Tights; Necklace; Other Accessories; Other Shoe; Polo Shirt; Pyjama Set; Sandals; Scarf; Shirt; Shorts; Skirt; Sneakers; Socks; Sunglasses; Sweater; Swimsuit; Swimwear Bottom; T Shirt; Top; Trousers; Underwear Tights; Vest Top.

Colors (color_std): Beige; Black; Blue; Brown; Burgundy; Cream; Green; Grey; Khaki; Navy; Olive; Orange; Pink; Purple; Red; White; Yellow.

Textures (texture_std): Abstract; Animal; Botanical; Camo; Checked; Chevron; Colorblock; Floral; Foil; Folk; Geo; Graphic; Mixed Print; Nautical; Ombre; Ornate; Paisley; Polka Dot; Pom Pom; Sophisticated; Southwestern; Stripe; Tonal; Tribal.

Fabrics (fabric_std): Applique; Burnout; Canvas; Chambray; Chiffon; Chino; Coated; Corduroy; Cotton; Crepe; Crochet; Cutout Lace; Denim; Distressed; Dye; Embellished; Embroidered; Eyelet; Fair Isle; Faux; Feather; Foulard; French; Fur; Gauze; Glitter; Jacquard; Knit; Lace; Layered; Leather; Mesh; Metallic; Nylon; Panel; Perforated; Plaid; Pleat; Print Satin; Quilted; Ruched; Ruffle; Scuba; Seamless; Seersucker; Semi Sheer; Sequin; Shearling; Sheer; Sleek; Stretch; Suede; Terry; Textured; Tie Dye; Tulle; Tweed; Twill; Velvet; Wash; Woven.

D.5 Best SB-Gate Operating Point

Parameter	Value
tau_sb	0.65
Accuracy	0.57744
Macro-F1	0.48324
sb_rate	0.50080

Notes: sb_rate is the share of cases meeting the SB-gate criterion at the selected threshold tau_sb. Selected from sb_gate_curve.csv on held-out CV splits.