

Simulation-Based Framework for Segmenting QRIS Consumer Payment Patterns Using Transaction Analytics and K-Means Clustering

Lina Marlina^{a,*}, Juliana Juliana^b, Asep Miftahuddin^b, Nihlatul Qudus Sukma Nirwana^c

^aUniversitas Siliwangi, Tasikmalaya, Indonesia

^bUniversitas Pendidikan Indonesia, Bandung, Indonesia

^cUniversitas Muhammadiyah Sidoarjo, Sidoarjo, Indonesia

Abstract

The Quick Response Code Indonesian Standard (QRIS) has accelerated the adoption of cashless retail payments in Indonesia, yet the consumers behind these transactions exhibit heterogeneous payment patterns that remain difficult to study because individual-level transaction records are restricted by confidentiality, ownership, and privacy constraints. This study develops and evaluates a simulation-based framework for segmenting QRIS consumer payment patterns using transaction analytics and K-Means clustering. A synthetic dataset of 20,686 transaction records for 1,000 synthetic users across 50 merchants in seven merchant categories was generated over a 180-day period from four explicitly planted latent behavioral profiles, with controlled noise comprising failed transactions, duplicates, missing categories, and extreme values. After preprocessing, 20,277 valid records were aggregated into eight user-level behavioral features covering frequency, monetary value, merchant diversity, temporal concentration, and regularity. Combined evidence from the Elbow curve, silhouette coefficient, Davies–Bouldin Index, and Calinski–Harabasz Index supported $K = 4$. The resulting segments Frequent Low-Value, Occasional High-Value, Diverse Active, and Emerging users recovered the planted profiles with an Adjusted Rand Index of 0.947, Normalized Mutual Information of 0.920, and an overall matching proportion of 97.6%, and remained stable across 30 repeated runs (mean ARI = 1.000) and six sensitivity scenarios (ARI 0.941–0.988). The framework offers a transparent, reproducible analytical workflow that can later be applied to anonymized real transaction data. Because the data are synthetic, the segments demonstrate methodological performance under stated assumptions and must not be generalized to actual QRIS users in Indonesia.

Keywords: QRIS, Consumer Payment Patterns, Transaction Analytics, K-Means Clustering, Cashless Payment

Received: 21 July 2026

Revised: 29 September 2026

Published: 31 October 2026

1. Introduction

Retail payment systems in many emerging economies are undergoing a structural transition from cash toward digital instruments. In Indonesia, this transition has been institutionalized through the Quick Response Code Indonesian Standard (QRIS), a national QR-code payment standard introduced by Bank Indonesia to unify previously fragmented QR schemes under a single interoperable specification (Indonesia, 2019). By allowing one QR code to be scanned by any licensed payment application, QRIS lowers acceptance costs for merchants, increases transaction convenience for consumers, reduces dependence on physical cash, and critically for research creates automated, timestamped, and categorized records of payment activity. These properties have made QRIS a central instrument in the digitalization of retail transactions and of micro, small, and medium enterprises.

A standardized payment instrument, however, does not produce uniform payment behavior. Some consumers use QR payments many times per day for low-value purchases such as food and transport; others use them occasionally for high-value retail or healthcare spending; others spread their activity across many merchant categories; and a further group uses the instrument rarely and irregularly. This heterogeneity is analytically important: two users with identical total spending may reach that total through entirely different combinations of frequency, transaction value, timing,

* Corresponding author.

E-mail address: linamarlina@unsil.ac.id

merchant diversity, and regularity. Understanding such multidimensional differences is the substantive motivation for consumer payment segmentation.

Most published research on QRIS and comparable digital payment instruments explains adoption and continued use through perception-based constructs perceived usefulness, ease of use, trust, security, social influence, and continuance intention rooted in the Technology Acceptance Model and the Unified Theory of Acceptance and Use of Technology (Davis, 1989; Venkatesh Morris, M. G., Davis, G. B., & Davis, F. D, 2003; Venkatesh et al., 2012). These constructs are valuable, but they measure what consumers say and intend rather than what they observably do. Survey instruments cannot capture the fine-grained structure of transaction timing, merchant-category allocation, or interval regularity that resides in transaction records. A behavioral, transaction-analytic perspective therefore complements, rather than replaces, intention-based research.

The principal obstacle to such behavioral research is data access (Pratama et al., 2026). Individual-level payment records contain sensitive financial information, are owned or controlled by regulated institutions and payment-service providers, and are subject to contractual restrictions and re-identification risk. Indonesia's Personal Data Protection Law (Law No. 27 of 2022) further formalizes obligations around the processing of personal financial data (Undang-Undang Nomor 27 Tahun 2022 Tentang Pelindungan Data Pribadi, 2022). Therefore, open, reproducible, individual-level QRIS transaction datasets are effectively unavailable to independent researchers.

Synthetic data generation offers a principled response to this constraint (Figueira, 2022; Jordon, 2022; Nikolenko, 2021; Rubin, 1993). Artificially generated transaction records, constructed from explicit statistical and behavioral rules, allow an analytical method to be developed, tested, stress-tested, and documented in full without exposing any real consumer information. A scientifically useful synthetic-data study must reproduce plausible statistical variation, retain controlled randomness, make every generation assumption transparent, and most importantly never present simulated patterns as observed social facts. Within these boundaries, simulation confers a unique methodological advantage: because the latent behavioral profiles used to generate the data are known, the ability of a clustering procedure to recover them can be validated externally, which is impossible with real unlabeled data.

Against this background, four gaps motivate the present study. First, an empirical gap: open individual-level QRIS transaction records remain inaccessible, limiting reproducible behavior-based analysis. Second, a methodological gap: QRIS research is dominated by surveys and causal perception models, while transaction analytics and unsupervised segmentation remain underdeveloped. Third, a reproducibility gap: data-generation assumptions, feature engineering, cluster-number selection, and validation procedures are often insufficiently documented in segmentation studies. Fourth, a validation gap: many segmentation studies rely on the Elbow Method alone and do not assess cluster separation, stability across random initializations, or where a simulation design permits recovery of known latent profiles (Caliński, 1974; Davies, 1979; Hubert, 1985; Ikotun, 2023; Jain, 2010; Rousseeuw, 1987).

This study therefore pursues six objectives: (1) develop a synthetic QRIS transaction dataset with plausible statistical and behavioral characteristics; (2) extract multidimensional payment-behavior indicators from transaction-level records; (3) determine an appropriate number of consumer clusters using multiple validation metrics; (4) segment synthetic users according to frequency, monetary value, transaction timing, merchant diversity, and regularity; (5) evaluate cluster quality, stability, and recovery of the planted latent profiles; and (6) deliver a reproducible framework that can subsequently be applied to anonymized real QRIS transaction data. Five research questions guide the analysis: how synthetic QRIS records can be generated to represent plausible payment behavior (RQ1); which behavioral indicators can be extracted for segmentation (RQ2); which number of clusters yields the most coherent and well-separated segments (RQ3); which payment-pattern segments K-Means identifies (RQ4); and how stable the identified clusters remain under different simulation parameters and random initializations (RQ5).

The study contributes in four ways. Theoretically, it conceptualizes digital payment behavior as a multidimensional observable construct intensity, monetary pattern, temporal pattern, merchant diversity, and regularity rather than an intention construct alone. Methodologically, it integrates synthetic data generation, behavioral feature engineering, K-Means clustering with K-Means++ initialization, multi-metric internal validation, external latent-profile validation, repeated-run stability testing, and sensitivity analysis into a single documented pipeline. Practically, it recalibrates type workflow for customer profiling, promotion design, and engagement analysis that payment providers or merchants

could recalibrate on authorized data. For reproducibility, it reports all simulation parameters, random seeds, equations, and preprocessing decisions.

2. Literature Review

2.1 QRIS and Cashless Payment Systems

QRIS was established through Bank Indonesia Board of Governors Regulation No. 21/18/PADG/2019 as the mandatory national standard for QR-code payments (Indonesia, 2019). Its defining feature is interoperability: a single merchant-presented or consumer-presented QR code is honored by all licensed payment applications, which removes the multi-code friction that previously fragmented Indonesia's QR ecosystem. For the purposes of this study, three properties matter. First, QRIS is dominantly a retail, face-to-face instrument, which makes merchant category and time of day behaviorally meaningful. Second, every successful QRIS transaction produces a structured digital record amount, timestamp, merchant identity that is in principle amenable to transaction analytics. Third, adoption spans very small merchants, so the consumer base is behaviorally diverse. Institutional detail beyond these points is deliberately limited, since the contribution of this article is analytical rather than regulatory.

2.2 Consumer Payment Behavior and Payment Choice Theory

Consumer payment behavior is defined here as the observable pattern through which a payment instrument is selected, repeated, timed, and distributed across transaction contexts. The empirical payment-economics literature shows that instrument choice varies systematically with transaction size, merchant acceptance, cost, speed, and context: cash persists in low-value transactions across countries (Bagnall, 2016), transaction size strongly conditions instrument selection at the point of sale (Klee, 2008), and adoption and use respond to relative convenience and pricing incentives (Koulayev, 2016), (Arango, 2015). Payment Choice Theory summarizes these regularities payment selection depends on convenience, speed, acceptance availability, amount, habit, cost, security, and purchase context. In this simulation study the theory serves an interpretive role: it informs how latent behavioral profiles are constructed and how resulting segments are read. It is not tested as a causal model.

Repeated, temporally consistent use of an instrument may also indicate a habit-like pattern. Habit in the psychological sense is measured through self-report instruments such as the Self-Report Habit Index (Verplanken, 2003), which transaction records cannot replicate. The framework therefore uses the deliberately restrained terms habit-like transaction regularity and routine payment intensity, operationalized through the variability of inter-transaction intervals, and avoids claiming that habit itself is measured.

2.3 Transaction Analytics and Behavior-Based Segmentation

Transaction analytics converts raw payment records into interpretable behavioral information through cleaning, temporal decomposition, aggregation, feature extraction, and profiling. It is the analytical bridge between transaction-level records and user-level segmentation. Behavior-based segmentation has a long lineage in database marketing: recency–frequency–monetary (RFM) analysis (Hughes, 1994) and its clustering extensions (Chen, 2012; Wedel, 2000) group customers by what they demonstrably do rather than by demographic descriptors, and data-mining approaches to customer relationship management have institutionalized clustering as the canonical segmentation tool (Han, 2012; Ngai, 2009). The present framework extends the RFM tradition in three directions relevant to QR payments: merchant-category diversity, temporal concentration (weekend and evening ratios), and inter-transaction regularity, yielding a multidimensional behavioral representation.

2.4 Synthetic Data Generation

Synthetic data are artificially generated records constructed from explicit statistical and behavioral rules. The idea originates in statistical disclosure limitation (Rubin, 1993) and has matured into a recognized methodology for privacy-preserving analysis and method development (Figueira, 2022; Jordon, 2022; Nikolenko, 2021). The scientific value of a synthetic dataset depends on plausibility (distributions and dependencies resemble the phenomenon), logical consistency (records obey domain constraints), reproducibility (generation is fully parameterized and seeded), controlled randomness, and transparent limitations. Two design principles follow for this study: the generator must not encode or approximate any real individual, and the manuscript must consistently describe outputs as synthetic or simulated transaction data never as observed QRIS evidence.

2.5 K-Means Clustering and Cluster Validation

K-Means (Lloyd, 1982; MacQueen, 1967) partitions observations into K clusters by minimizing within-cluster squared Euclidean distance. It is selected here because it is computationally efficient, interpretable through centroids, reproducible when seeded, and well suited to standardized numeric features; it also remains the reference algorithm against which alternatives are judged (Ikotun, 2023; Jain, 2010). Its known limitations sensitivity to outliers, scale, initial centroids, non-spherical structure, and the pre-selection of K are addressed by design: log transformation and standardization of skewed features, K-Means++ initialization (Arthur, 2007), fifty random restarts, and multi-metric selection of K combining the Elbow curve, silhouette coefficient (Rousseeuw, 1987), Davies–Bouldin Index (Davies, 1979), and Calinski–Harabasz Index (Caliński, 1974). Partition agreement is quantified with the Adjusted Rand Index (Hubert, 1985) and Normalized Mutual Information (Vinh, 2010), and robustness is examined against Ward hierarchical clustering and Gaussian mixture models (Fraily, 2002).

2.6 Conceptual Framework

Figure 1 integrates the preceding elements. Simulation assumptions and latent behavioral profiles drive synthetic transaction generation; cleaning and validation produce an analyzable record set; transaction analytics engineers user-level behavioral features; transformation and standardization prepare the feature space; K-Means segments the users; multi-metric validation, external latent-profile recovery, and stability analysis establish trustworthiness; and segment profiling delivers the interpretive output.

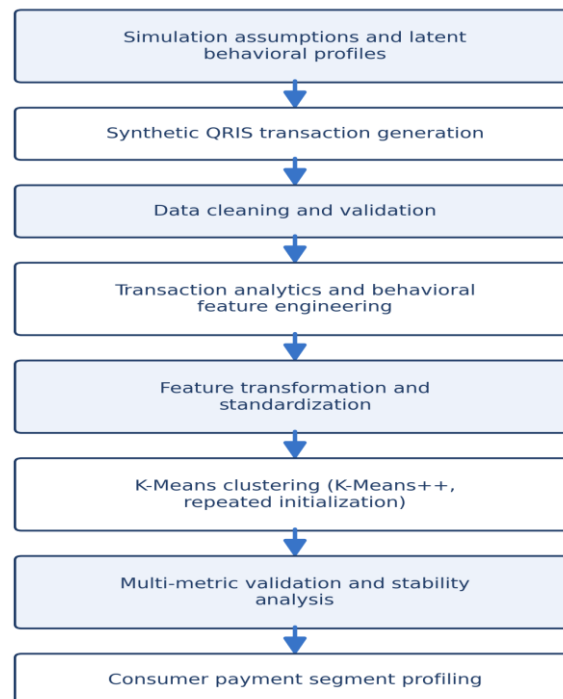


Fig. 1. Conceptual framework of the proposed study

3. Methods

3.1 Research Design

The study adopts a quantitative, computational, exploratory, simulation-based, and methodological design (Wardoyo et al., 2020). No causal relationship is tested. The principal question is whether a fully documented analytical workflow can recover and explain heterogeneous payment patterns embedded in a controlled synthetic dataset. The workflow proceeds in ten steps: (1) define simulation assumptions and latent behavioral profiles; (2) generate synthetic users, merchants, dates, times, categories, and transaction values; (3) introduce limited noise failed records, duplicates, missing values, and valid extremes; (4) clean and validate the transaction-level dataset; (5) aggregate transactions to the user level and engineer behavioral features; (6) transform and standardize the analytical variables; (7) determine the

appropriate number of clusters; (8) run K-Means with repeated initialization; (9) evaluate internal quality, external latent-profile recovery, and stability; and (10) interpret the resulting segments and test sensitivity. Figure 2 visualizes the full workflow.

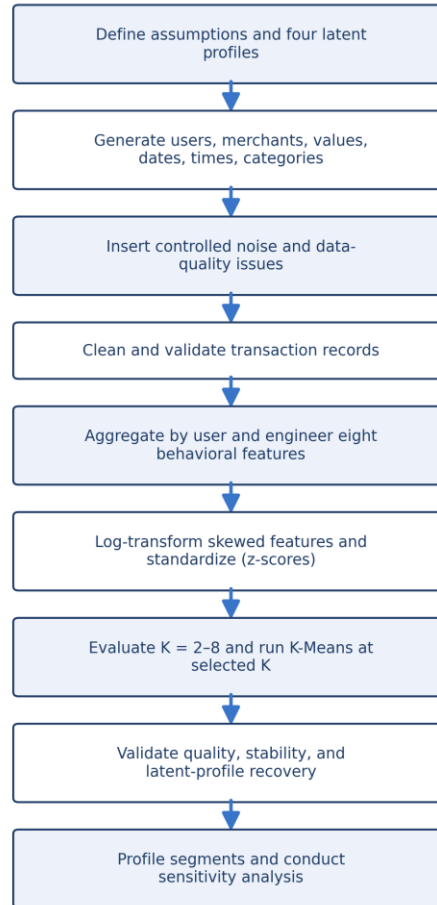


Fig 2. Simulation and analytical workflow

3.2 Baseline Simulation Scenario and Units of Analysis

Table 1 lists the baseline simulation parameters. The values are methodological design choices for a controlled experiment and must not be read as official QRIS statistics. Two units of analysis are used: the transaction level, where one row represents one synthetic payment and which supports temporal, category, and preprocessing analysis; and the user level, where all valid transactions are aggregated by anonymous synthetic user and which forms the input to K-Means.

Table 1. Baseline simulation parameters

Component	Value
Synthetic users	1,000
Generated transaction records (before cleaning)	20,686
Merchants	50
Merchant categories	7
Observation period	180 days
Latent behavioral profiles	4
Stability replications	30
Candidate K values	2–8
Primary random seed	42
Transaction amount bounds (simulation constraint)	IDR 10,000 – IDR 2,500,000

3.3 Latent Behavioral Profiles

Four latent consumer profiles were planted in the generator (Table 2). Profile labels are retained solely for post-clustering external validation and are excluded from the K-Means input features. Profile parameters were chosen to create behaviorally distinct but overlapping groups: category probabilities, log-normal amount parameters, Poisson frequency rates, weekend and evening propensities, and interval regularity all differ by profile, while shared category support and overlapping amount distributions prevent trivial separability.

Table 2. Latent consumer profiles used for data generation

Profile	Share	Freq. (λ)	Value pattern	Category pattern	Temporal pattern
A. Frequent Value	Low- 30%	38	Low average	Food & transport dominant	Routine, regular intervals, weekday-leaning
B. Occasional Value	High- 20%	7	High average	Retail, healthcare, services	Selective, higher weekend share
C. Diverse Active	30%	22	Medium	High category diversity	Broad distribution, moderately regular
D. Emerging Users	20%	5	Low	One–two categories	dominant Irregular, limited use

3.4 Synthetic Transaction Generation

Each transaction record carries a unique identifier, an anonymous user code, a merchant code, a merchant category (food and beverages, grocery and retail, transportation, personal services, healthcare, education, or entertainment), a date within the 180-day period, a time, a positive amount, derived day-of-week, day-type, and time-period fields, a payment status, and the latent profile label used only for validation. The per-user transaction count follows a Poisson distribution with a profile-specific rate; a negative-binomial alternative is available where overdispersion is intended:

$$F_i \sim \text{Poisson}(\lambda_g) \quad (1)$$

Transaction amounts are positively skewed, so a log-normal distribution conditional on profile g and merchant category c is used, truncated to the simulation bounds:

$$X_{ij} \sim \text{Lognormal}(\mu_{g,c}, \sigma_{g,c}^2), 10,000 \leq X_{ij} \leq 2,500,000 \quad (2)$$

Merchant categories are drawn from a profile-conditional categorical distribution, $P(C_{ij} = c \mid g) = p_{g,c}$, with probabilities different enough to create behavioral signal yet overlapping across profiles. Transaction days for routine profiles (A and C) are generated as jittered, evenly spaced sequences profile A with small jitter, profile C with larger jitter while non-routine profiles (B and D) draw days uniformly; day-of-week is then tilted toward the weekend with a profile-specific probability, and time of day is drawn with profile-specific evening propensity within three periods (morning 06:00–11:59, afternoon 12:00–16:59, evening 17:00–23:59).

3.5 Noise Injection and Preprocessing

Perfectly clean, perfectly separable synthetic data would not test the pipeline. Controlled imperfections were therefore injected at generation: 1.0% failed transactions, 0.5% duplicated records, 0.5% missing merchant categories, and 1.0% extreme but valid amounts (values multiplied by a factor of 3–8, capped at the upper bound). Preprocessing then (1) removed exact duplicates, (2) excluded failed transactions, (3) removed non-positive amounts, (4) validated dates against the simulation period, (5) removed records with missing merchant category under a declared listwise rule, (6) inspected extreme values with the interquartile-range criterion rather than deleting them, (7) aggregated valid transactions by user, and (8) documented the records affected at every step. The IQR bounds follow the standard definition:

$$IQR = Q_3 - Q_1, \text{ bounds} = (Q_1 - 1.5 \cdot IQR, Q_3 + 1.5 \cdot IQR) \quad (3)$$

3.6 Behavioral Feature Engineering

Eight user-level features were computed from the cleaned records. For user i with n_i valid transactions of amounts x_{ij} at timestamps t_{ij} :

$$F_i = n_i; \quad TMV_i = \sum_j x_{ij}; \quad ATV_i = \frac{TMV_i}{F_i} \quad (4)$$

$$MD_i = |\{c_{i1}, \dots, c_{i,n_i}\}| \quad (5)$$

$$WTR_i = \frac{N_{i,weekend}}{F_i}; \quad ETR_i = \frac{N_{i,evening}}{F_i} \quad (6)$$

$$TR_i = \frac{1}{1 + SD(\Delta t_i)}, \quad \Delta t_{i,j} = t_{i,j} - t_{i,j-1} \quad (7)$$

$$CCI_i = \sum_c p_{i,c}^2 \quad (8)$$

F is transaction frequency; TMV total monetary value; ATV average transaction value; MD merchant-category diversity; WTR and ETR the weekend and evening transaction ratios; TR transaction regularity, which approaches 1 as inter-transaction intervals become uniform; and CCI a Herfindahl-type category-concentration index, higher when transactions concentrate in fewer categories.

3.7 Transformation, Standardization, and Redundancy Inspection

The count and monetary features (F, TMV, ATV) are right-skewed and were log-transformed, $X' = \ln(1 + X)$, before all eight features were standardized to z-scores, $z_{ij} = \frac{x_{ij} - \bar{x}_j}{SD_j}$, so that Euclidean distances weight features comparably. A correlation matrix of the transformed features was inspected before clustering; because total value is a function of frequency and average value, a robustness specification re-ran the final clustering without TMV to verify that redundancy does not drive the solution.

3.8 Clustering, Selection of K, and Validation

K-Means minimizes the within-cluster sum of squared distances,

$$J = \sum_{k=1}^K \sum_{x_i \in C_k} \|x_i - \mu_k\|^2 \quad (9)$$

and was run with K-Means++ initialization and 50 random restarts for each candidate $K \in \{2, \dots, 8\}$. Four internal criteria were computed for every candidate: within-cluster sum of squares (Elbow), the silhouette coefficient $S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$, the Davies–Bouldin Index, and the Calinski–Harabasz Index. Because no single criterion is decisive, K was selected by combined evidence: each metric ranks the candidates (silhouette and CHI descending, DBI ascending), and the K with the best average rank also weighed against interpretability and cluster-size adequacy is retained. Importantly, the procedure does not automatically select four clusters merely because four latent profiles were planted.

Validation proceeded on three fronts. Internally, the final solution's silhouette and DBI are reported. Externally an advantage unique to simulation the clusters were compared with the planted latent profiles using the Adjusted Rand Index, $ARI = \frac{RI - E(RI)}{\max(RI) - E(RI)}$, Normalized Mutual Information, a profile-by-cluster cross-tabulation, and an overall matching proportion computed with an optimal one-to-one (Hungarian) assignment of profiles to clusters. The latent label was never an input feature. Stability was assessed by repeating the final K-Means across 30 independent random seeds and summarizing the silhouette, the ARI between repeated solutions and the reference partition, mean centroid displacement, and cluster-size variation.

3.9 Sensitivity Analysis and Method Robustness

Six scenarios re-ran the entire generation-to-clustering pipeline under a different seed: baseline; high noise (5% failed, 3% duplicates, 3% missing, 5% extremes); high profile overlap (amount dispersion inflated by 60%); small sample (500 users); large sample (2,000 users); and imbalanced profiles (shares 55/10/25/10). Each scenario reports the internal quality of the four-segment structure and its recovery of the latent profiles, together with the K that the combined-rank rule would prefer in that scenario. Finally, method robustness was checked by re-clustering the baseline feature matrix with Ward hierarchical clustering and a Gaussian mixture model at the same K and comparing the partitions.

3.10 Software, Reproducibility, and Ethics

The analysis was implemented in Python 3 using NumPy, pandas, SciPy, and scikit-learn, with matplotlib for figures. The primary random seed (42), all generation parameters, preprocessing rules, the feature list, and the validation configuration are reported in full, and the generation and analysis code is structured so that the synthetic dataset can be regenerated exactly. No real personal or financial records were used at any stage; the synthetic users do not represent

identifiable individuals. Future implementation on real payment data must apply data minimization, anonymization, secure storage, access control, and the requirements of Indonesia's Personal Data Protection Law.

4. Result and Discussion

4.1 Synthetic Dataset Characteristics

The generator produced 20,686 transaction records for 1,000 synthetic users across 50 merchants in seven categories over 180 days. Preprocessing removed 102 duplicated records, 205 failed transactions, and 102 records with missing merchant category, leaving 20,277 valid records (98.02% of the generated set; 1.98% removed). Applying the IQR criterion ($Q1 = \text{IDR } 20,105$, $Q3 = \text{IDR } 55,885$, upper bound = $\text{IDR } 109,555$) flagged 2,035 records (10.0%) as statistically extreme; inspection confirmed they were logically valid high-value purchases within the simulation bounds, so they were retained. Table 3 summarizes the dataset before and after cleaning.

Table 3. Synthetic dataset characteristics before and after cleaning

Indicator	Before cleaning	After cleaning
Transaction records	20,686	20,277
Synthetic users	1,000	1,000
Merchants	50	50
Merchant categories	7	7
Failed transactions	205	Removed
Duplicated records	102	Removed
Missing merchant category	102	Removed (listwise rule)
Extreme but valid amounts	2,035	Inspected and retained

4.2 User-Level Descriptive Statistics

Table 4 reports descriptive statistics for the eight analytical features. Transaction frequency averaged 20.3 transactions per user (median 20) over the period, with pronounced dispersion ($SD = 13.9$; range 1–51). The monetary features are strongly right-skewed: mean total value ($\text{IDR } 1,194,019$) exceeds the median ($\text{IDR } 1,074,093$) substantially, and average transaction value shows skewness of 2.13. At transaction level, amounts averaged $\text{IDR } 58,885$ against a median of $\text{IDR } 30,822$ with skewness 7.3, falling to 1.1 after log transformation (Figure 3) confirming that the log transformation applied before standardization was justified. Because the data are synthetic, these figures describe properties of the generated dataset, not characteristics of Indonesian consumers.

Table 4. Descriptive statistics of the analytical features ($n = 1,000$ users)

Feature	Mean	Median	SD	Min	Max	Skew
Transaction frequency	20.28	20.00	13.86	1.00	51.00	0.27
Total monetary value (IDR)	1,194,019	1,074,093	842,826	10,000	6,985,530	1.64
Average transaction value (IDR)	99,233	53,594	121,488	10,000	780,062	2.13
Merchant diversity	4.84	5.00	1.77	1.00	7.00	-0.38
Weekend transaction ratio	0.33	0.29	0.20	0.00	1.00	1.09
Evening transaction ratio	0.35	0.35	0.16	0.00	1.00	0.32
Transaction regularity	0.15	0.14	0.09	0.01	0.91	0.67
Category concentration (CCI)	0.35	0.34	0.17	0.15	1.00	1.61

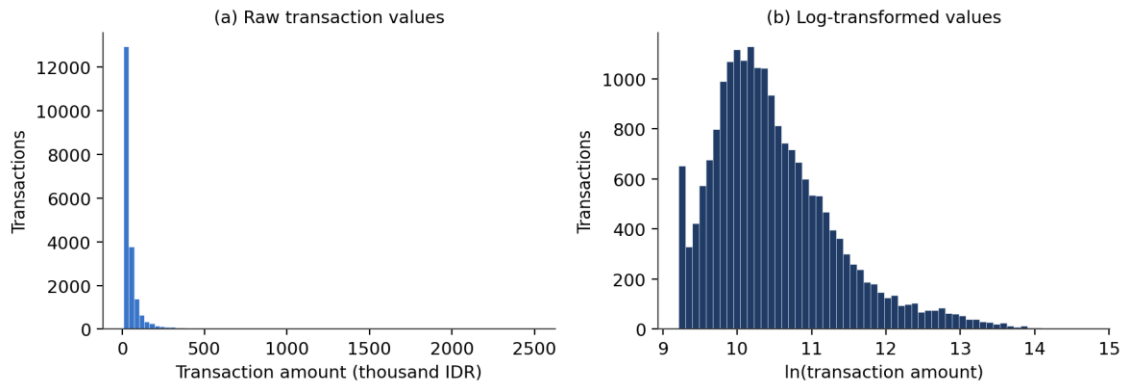


Fig 3. Raw and log-transformed transaction-value distributions

4.3 Temporal and Merchant-Category Patterns

Transactions concentrated in the evening period (36.7%) and morning period (35.1%), with 26.7% occurring on weekends. The hourly profile (Figure 4) shows clear morning and evening intensity consistent with meal-time and after-work activity planted in the routine profiles, and the day-hour heatmap (Figure 5) shows weekend cells with elevated intensity relative to their share of days. Table 5 and Figure 6 report the merchant-category distribution: food and beverages dominates transaction volume (7,641 transactions; 37.7%), followed by transportation (23.6%), while healthcare and education carry the highest average values an inversion of volume and value that mirrors the planted profile logic of frequent small purchases versus occasional large ones.

Table 5. Transaction distribution by merchant category

Merchant category	Transactions	Share (%)	Avg. value (IDR)	Unique users
Food and beverages	7,641	37.7	31,871	869
Transportation	4,777	23.6	27,414	717
Grocery and retail	2,511	12.4	115,975	874
Personal services	1,629	8.0	83,336	700
Entertainment	1,425	7.0	63,677	598
Healthcare	1,339	6.6	156,866	625
Education	955	4.7	96,108	459

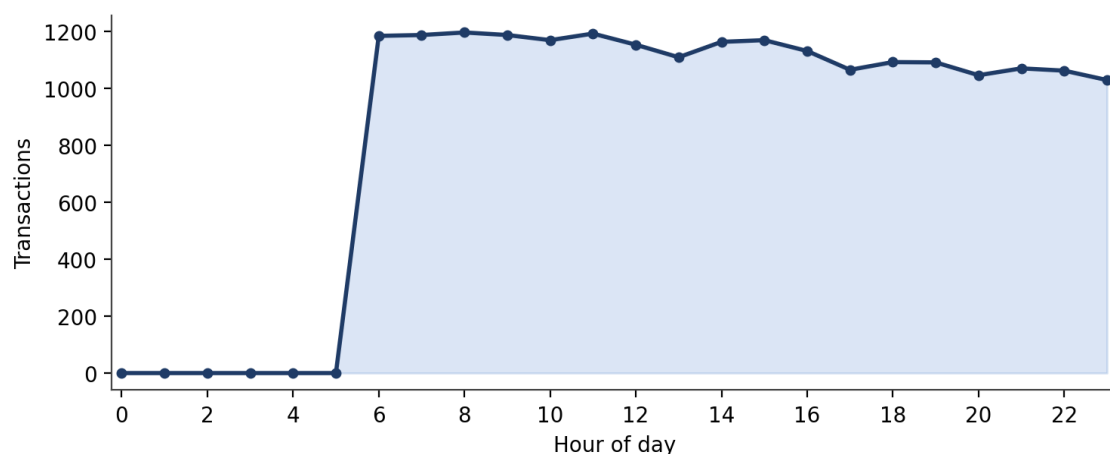


Fig 4. Hourly transaction pattern

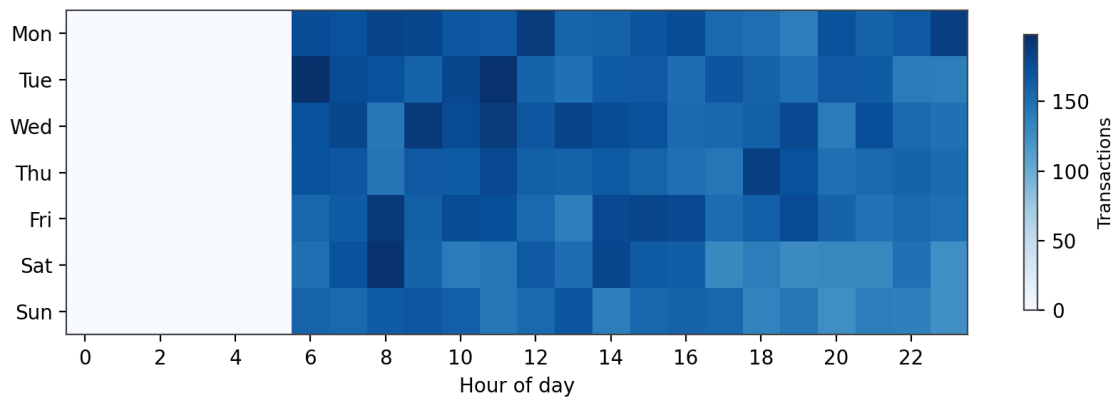


Fig 5. Day-hour transaction heatmap

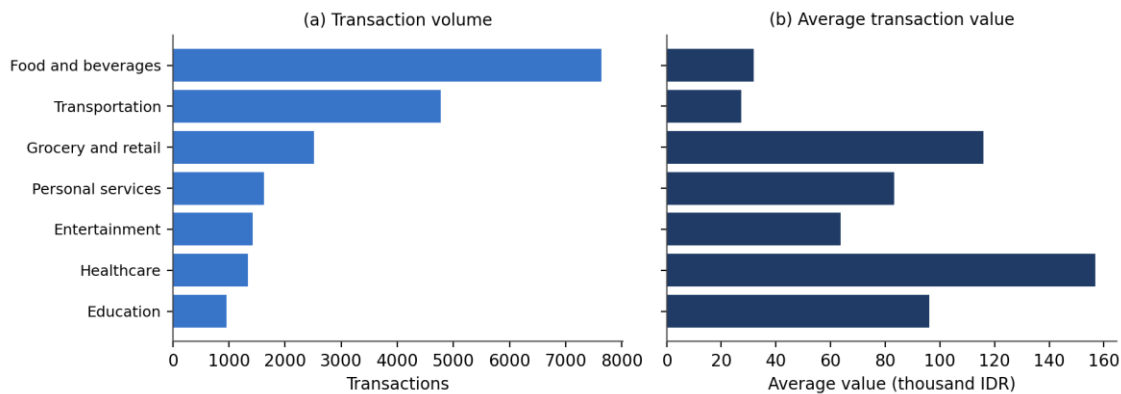


Fig 6. Merchant-category transaction volume and average value

4.4 Feature Correlation and Redundancy Check

Figure 7 presents the correlation matrix of the transformed features. The strongest association is between frequency and regularity ($r = 0.85$), which is expected by construction: users transacting more often within a fixed window necessarily exhibit shorter and more uniform intervals. Frequency and total value correlate moderately ($r = 0.52$), and no pair is collinear enough to make the distance space degenerate. The robustness specification that removed total monetary value reproduced the final four-cluster partition almost exactly ($ARI = 0.947$ against the full-feature solution), confirming that the segmentation is not an artifact of feature redundancy.

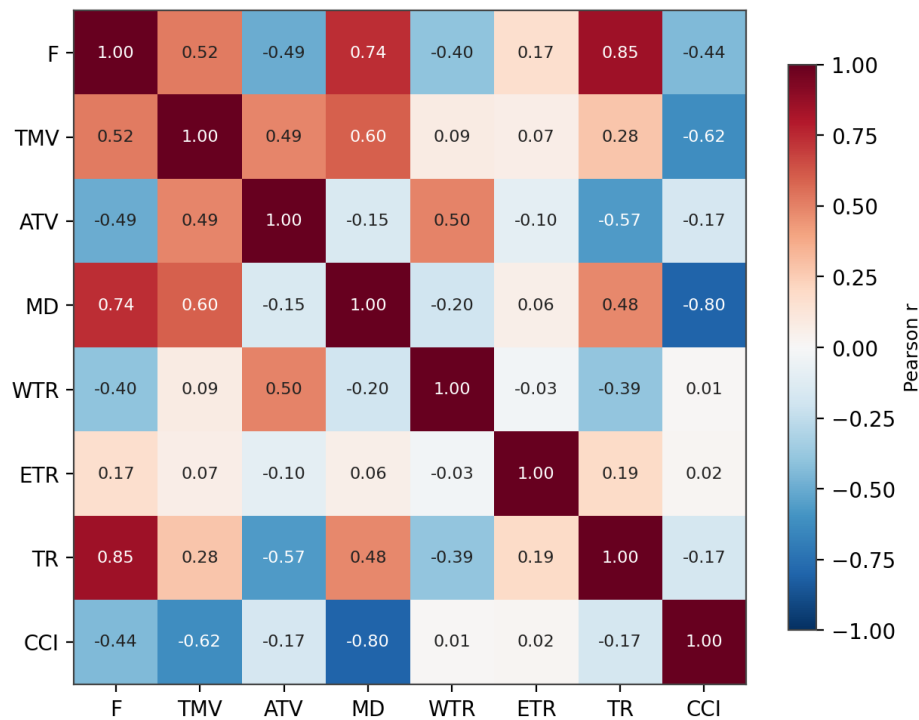


Fig 7. Feature-correlation heatmap (transformed features)

4.5 Determination of the Optimal Number of Clusters

Table 6 and Figure 8 report the four validation criteria for $K = 2-8$. The evidence is instructive rather than unanimous. The silhouette coefficient is maximized at $K = 3$ (0.428), where the two high-frequency routine profiles merge into one compact cluster; however, the Davies–Bouldin Index is minimized at $K = 4$ (1.018), the Calinski–Harabasz Index is maximized at $K = 4$ (557.5), the Elbow curve flattens visibly beyond $K = 4$, and $K = 4$ attains the best average rank across the three scale-free metrics (1.33 versus 2.00 for $K = 3$). Combined with interpretability $K = 4$ separates a frequent low-value routine pattern from a diverse mid-value pattern that $K = 3$ conflates and adequate cluster sizes, $K = 4$ was selected. The decision follows the combined-evidence rule specified in Section 3.8 and is not a consequence of four profiles having been planted; Section 4.9 shows that the same tension between silhouette and the other criteria recurs across scenarios.

Table 6. Cluster-validation results for candidate K values

K	WCSS	Silhouette	DBI	CHI	Avg. rank
2	5,190	0.390	1.283	540.5	4.00
3	3,865	0.428	1.107	533.4	2.00
4	2,986	0.399	1.018	557.5	1.33
5	2,662	0.398	1.247	498.9	3.33
6	2,453	0.393	1.257	449.5	4.33
7	2,278	0.389	1.312	415.6	6.33
8	2,119	0.384	1.300	393.3	6.67

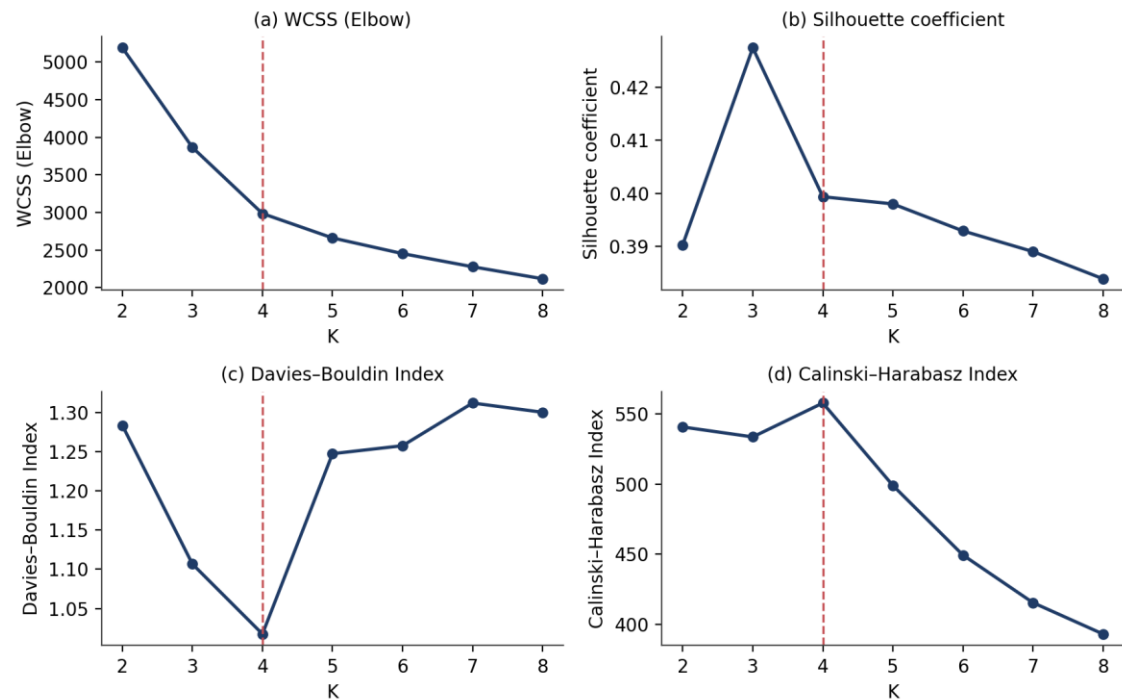


Fig 8. Elbow curve and validation-metric comparison (dashed line: selected $K = 4$)

4.6 Cluster Composition and Segment Profiles

The final K-Means solution (silhouette = 0.399, DBI = 1.018) partitioned the 1,000 users into four segments of 309, 183, 306, 202 users (30.9%, 18.3%, 30.6%, 20.2%). Table 7 reports the composition together with the mean raw feature values per segment, and Figure 9 shows the standardized centroids. Segment names were assigned after inspection of the centroid and category profiles, following the interpretation rule that labels must follow, not precede, the clustering.

Table 7. Segment composition and mean raw feature values

Feature	C1. Frequent Low-Value Users	C2. Occasional High-Value Users	C3. Diverse Active Users	C4. Emerging QRIS Users
Users (share)	309 (30.9%)	183 (18.3%)	306 (30.6%)	202 (20.2%)
Transaction frequency	37.1	6.9	21.6	4.6
Total monetary value (IDR)	968,352	2,251,430	1,433,993	217,746
Average transaction value (IDR)	26,704	329,311	66,420	51,454
Merchant diversity	5.2	3.7	6.6	2.5
Weekend transaction ratio	0.202	0.516	0.327	0.349
Evening transaction ratio	0.400	0.348	0.327	0.296
Transaction regularity	0.262	0.061	0.145	0.054
Category concentration (CCI)	0.391	0.356	0.189	0.542

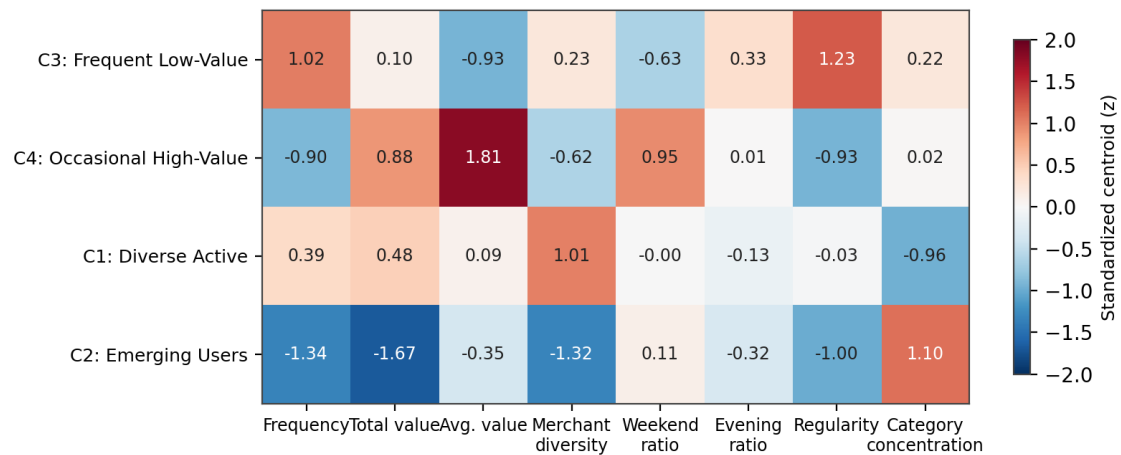


Fig 9. Standardized cluster-centroid profiles

The four segments are interpretable directly from the centroids. C1 (Frequent Low-Value Users, 30.9%) transact most often (mean 37.1 transactions) at the lowest average value (IDR 26,704), with the highest regularity ($z = +1.23$) and strong weekday concentration the routine daily-purchase pattern. C2 (Occasional High-Value Users, 18.3%) transact rarely (mean 6.9) at an average value an order of magnitude higher (IDR 329,311), with the highest weekend ratio ($z = +0.95$) and the highest total spending. C3 (Diverse Active Users, 30.6%) combine medium-high frequency (mean 21.6) with the greatest merchant diversity (mean 6.6 categories; CCI $z = -0.96$) and broad temporal distribution. C4 (Emerging QRIS Users, 20.2%) show low frequency (mean 4.6), low spending, concentration in one or two categories (highest CCI), and the most irregular timing. Figure 10 visualizes the segments in the first two principal components (42.6% and 26.8% of variance), and Figure 11 shows the distribution of the four most discriminating features by segment.

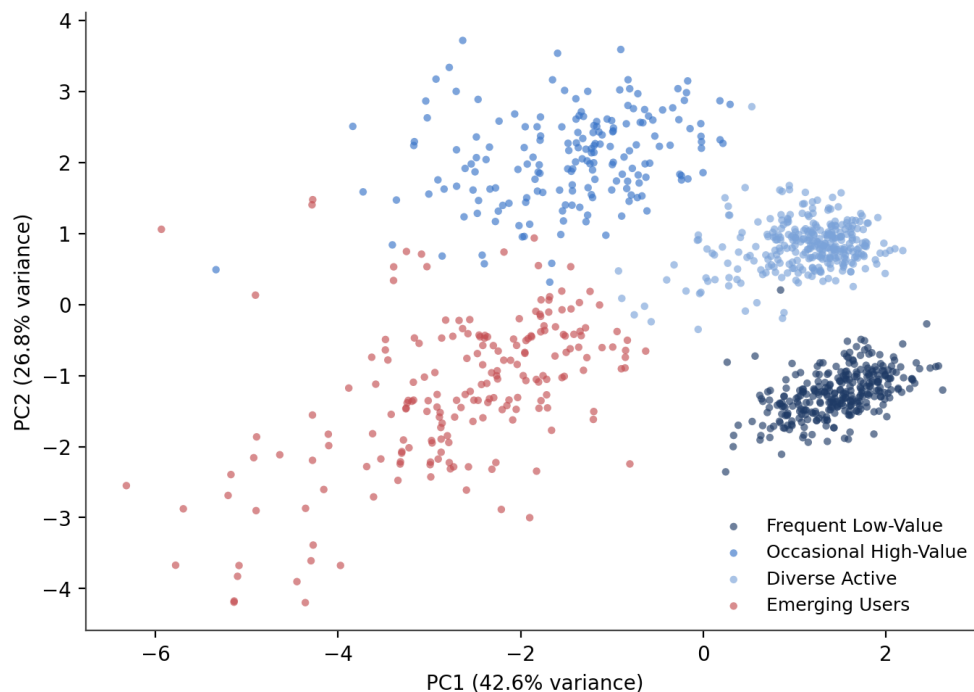


Fig 10. Two-dimensional PCA visualization of the four segments

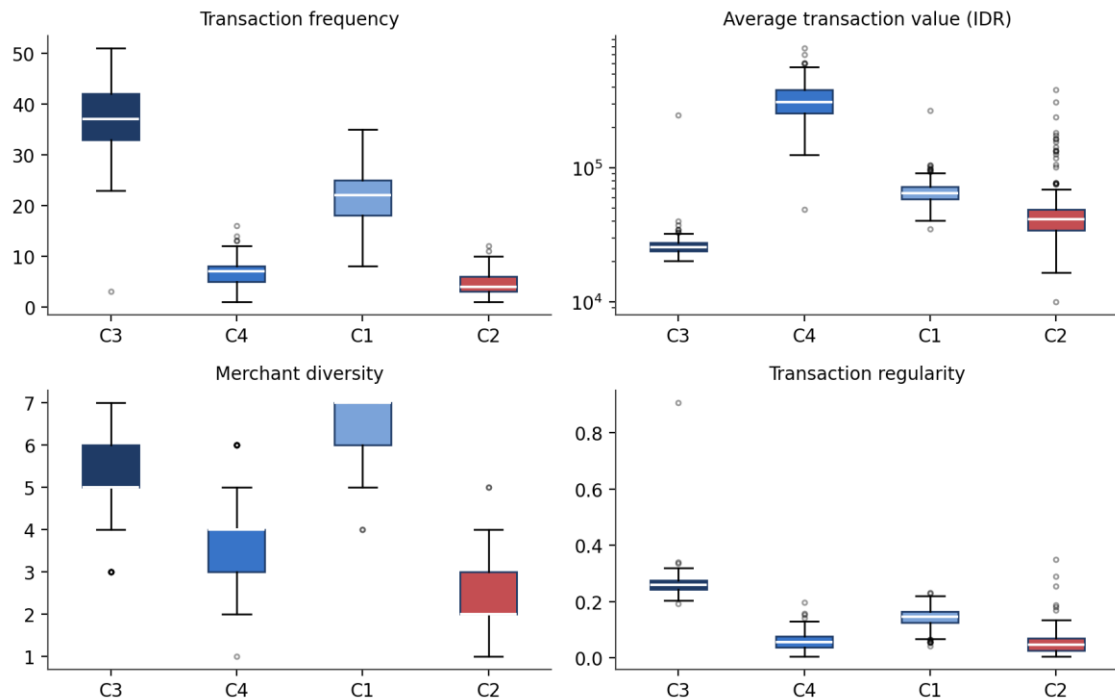


Fig 11. Feature distributions by segment (C1 Frequent Low-Value, C2 Occasional High-Value, C3 Diverse Active, C4 Emerging)

4.7 Statistical Differences among Segments

Because several features remained non-normal after transformation, Kruskal–Wallis tests [32] compared the segments on the six substantive features (Table 8). All differences are significant at $p < 0.001$ with very large effect sizes (ϵ^2 between 0.341 for the weekend ratio and 0.872 for frequency), indicating that the segments differ substantially not merely statistically on every behavioral dimension.

Table 8. Kruskal–Wallis tests of inter-segment differences

Feature	H	p-value	ϵ^2 (effect size)	Interpretation
Transaction frequency	871.9	< 0.001	0.872	Very large
Total monetary value (IDR)	708.0	< 0.001	0.708	Very large
Average transaction value (IDR)	830.1	< 0.001	0.830	Very large
Merchant diversity	749.8	< 0.001	0.750	Very large
Weekend transaction ratio	342.4	< 0.001	0.341	Large
Transaction regularity	828.5	< 0.001	0.829	Very large

4.8 External Simulation Validation

Because the latent generation profiles are known, recovery can be validated externally (Table 9). The four clusters correspond one-to-one to the four planted profiles: ARI = 0.947, NMI = 0.920, and 97.6% of users were assigned to the cluster optimally matched to their generating profile. The largest mismatch concerns the Occasional High-Value profile, of which 7.2% was absorbed mainly by the Emerging cluster users drawn at the low-frequency tail of that profile are behaviorally indistinguishable from emerging users, an overlap deliberately built into the generator. Table 10 presents the full cross-tabulation.

Table 9. Recovery of the latent simulation profiles

Metric	Result
Adjusted Rand Index	0.947
Normalized Mutual Information	0.920
Overall matching proportion	97.6%

Largest profile–cluster mismatch

B. Occasional High-Value (7.2%)

Table 10. Cross-tabulation of latent profiles and resulting segments

Latent profile	C1	C2	C3	C4
A. Frequent Low-Value	308	0	0	0
B. Occasional High-Value	1	181	1	12
C. Diverse Active	0	1	298	1
D. Emerging	0	1	7	189

4.9 Stability and Sensitivity

Across 30 independent random seeds, the four-segment solution was essentially invariant: mean silhouette 0.399 (SD 0.0000), mean ARI against the reference partition 1.000 (minimum 0.997), mean centroid displacement 0.0005 standardized units, and cluster-size variation below one user (Table 11). K-Means++ with repeated restarts therefore converges to the same partition regardless of initialization.

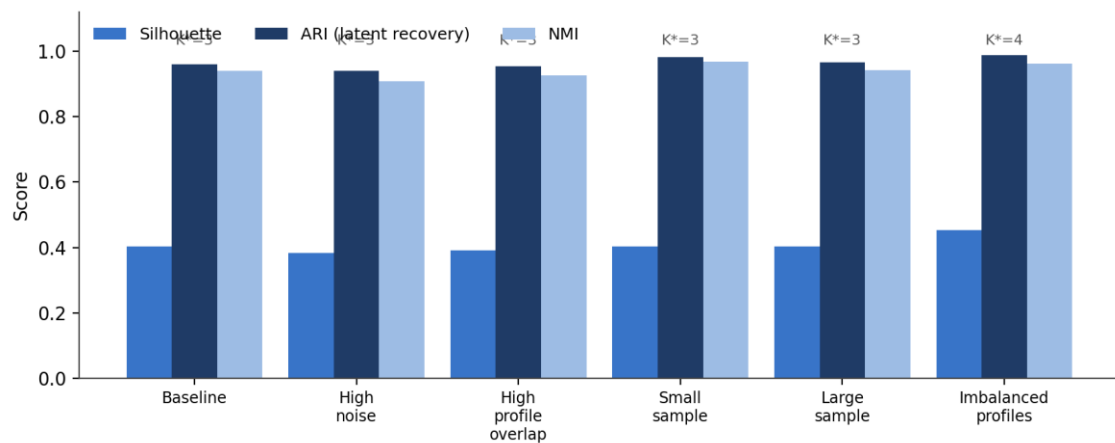
Table 11. Cluster stability across 30 repeated runs

Indicator	Mean	SD	Min	Max
Silhouette score	0.399	0.0000	0.399	0.399
ARI vs. reference partition	1.000	0.0008	0.997	1.000
Centroid displacement (z units)	0.0005	0.0015	0.0000	0.0049
Cluster-size SD (users)	66.9	0.18	66.9	67.5

Table 12 and Figure 12 report the six sensitivity scenarios, each a full regeneration-to-clustering replication under a different seed. The four-segment structure is robust: latent-profile recovery remains high in every scenario (ARI 0.941–0.988; NMI 0.908–0.969), degrading only mildly under fivefold noise and 60% wider amount dispersion. The final column, however, records an important nuance: in five of six scenarios the combined-rank rule would nominally prefer $K = 3$, because the silhouette consistently rewards merging the two routine profiles even as DBI and CHI favor $K = 4$. The structure is therefore more stable than the automatic selection of K a finding taken up in the Discussion.

Table 12. Sensitivity of the four-segment framework

Scenario	Silhouette	DBI	ARI	NMI	Preferred K
Baseline	0.403	0.996	0.961	0.940	3
High noise	0.383	1.043	0.941	0.908	3
High profile overlap	0.391	1.026	0.954	0.927	3
Small sample	0.403	0.982	0.982	0.969	3
Large sample	0.404	1.010	0.966	0.942	3
Imbalanced profiles	0.452	1.018	0.988	0.963	4

**Fig 12.** Sensitivity-analysis results (K^* denotes the K preferred by the combined-rank rule in each scenario)

4.10 Method Robustness

Re-clustering the baseline feature matrix at $K = 4$ with Ward hierarchical clustering and a Gaussian mixture model reproduced the K-Means partition almost exactly (ARI vs. K-Means of 0.965 and 0.966, respectively) with nearly identical silhouettes (0.396 and 0.397) and equally high latent-profile recovery (ARI 0.960 and 0.962). The segment structure is therefore not method-dependent.

The central substantive result is that a single aggregate such as total transaction value would have concealed most of the structure the framework recovered. The Frequent Low-Value and Diverse Active segments accumulate comparable total spending through radically different mechanisms the former through roughly 37 small, regular, weekday, food-and-transport transactions; the latter through about 22 mid-value transactions spread across nearly all seven categories. Likewise, the Occasional High-Value segment demonstrates that low frequency does not imply low monetary importance: with fewer than 7 transactions on average, it nonetheless records the highest total spending of any segment. These contrasts validate the framework's premise that payment behavior is a multidimensional construct intensity, monetary pattern, temporal pattern, merchant diversity, and regularity rather than a single scalar, and they echo the payment-economics finding that transaction size and context, not consumer identity alone, organize instrument use (Bagnall, 2016; Klee, 2008).

Each segment illustrates a pattern the framework is designed to expose, with the strict caveat that these are simulation-derived demonstrations, not observed QRIS behavior. The Frequent Low-Value segment shows how routine digital-payment intensity is identified: its defining signature is not spending volume but the combination of high frequency, low average value, high interval regularity, and category concentration in food and transport a pattern that, in a validated real-data deployment, would correspond to daily convenience purchasing. The Occasional High-Value segment shows the inverse signature rare, expensive, weekend-leaning, selectively categorized transactions which in practice could inform nominal-based promotions or high-value engagement strategies, subject to empirical validation. The Diverse Active segment shows the analytical value of combining category diversity with frequency and temporal breadth: diversity in isolation is a weak descriptor, but jointly with sustained activity it identifies users for whom the instrument has become a general-purpose payment default. The Emerging segment low frequency, low diversity, irregular timing represents the pattern of limited or early-stage use; hypothetically, onboarding incentives, repeat-use campaigns, or security education would target such users, but these recommendations remain conceptual until tested on authorized real data.

The framework's methodological contribution lies less in any single technique than in the documented chain connecting them: profile-driven generation with distributions appropriate to count data (Poisson) and skewed monetary data (log-normal); deliberate data-quality perturbation; declared preprocessing rules; eight-feature behavioral engineering; transformation and standardization; multi-metric selection of K ; and a three-layer validation architecture. Each layer answered a distinct question. Internal validation established that the partition is geometrically coherent (silhouette 0.399, DBI 1.018). External validation possible only because the latent profiles are known established that the partition is substantively correct (ARI 0.947; 97.6% matching), and the residual 7.2% mismatch of the Occasional High-Value profile localized exactly where the generator planted overlap, confirming that the framework degrades where and only where it should. Stability analysis established that the result is not an initialization artifact, and the method comparison established that it is not an algorithm artifact. A study relying on the Elbow curve alone the pattern criticized in the validation gap would have documented none of this.

The most instructive negative finding concerns the selection of K . The silhouette coefficient preferred $K = 3$ in the baseline and in five of six sensitivity scenarios, consistently rewarding a coarser partition that merges the two routine, high-frequency profiles, while the Davies–Bouldin and Calinski–Harabasz indices, the Elbow inflection, and interpretability favored $K = 4$ the value at which the planted structure is actually recovered (ARI 0.947 at $K = 4$). This is a concrete demonstration of a known property: silhouette favors widely separated, compact clusters and can under-segment when true groups overlap along some dimensions (Jain, 2010; Rousseeuw, 1987). Two practical lessons follow for applied segmentation. First, combined evidence across metrics with different geometric biases should be mandatory, exactly as the framework specifies. Second, the stability of a candidate structure under perturbation (here, ARI of 0.94–0.99 for $K = 4$ across all scenarios) is itself evidence for that structure, and can be more informative than the point decision of any selection rule. In real deployments without planted labels, this argues for reporting the two or three best-supported K values with their profiles rather than a single automatic choice.

The experiment also maps the conditions under which K-Means performs well and where caution is required. Performance was strong here because the design respected the algorithm's assumptions: features were log-transformed and standardized, extreme values were bounded, K-Means++ with 50 restarts neutralized initialization sensitivity, and

the planted clusters were roughly convex in the transformed space. Recovery declined mildly but did not fail under fivefold noise and inflated dispersion, and was essentially unaffected by sample size between 500 and 2,000 users or by 55/10/25/10 profile imbalance. The agreement of Ward clustering and Gaussian mixtures with the K-Means partition ($ARI > 0.96$) indicates that, for this feature space, the structure is method-independent; conversely, in applications with strongly non-spherical or density-based structure, that agreement would break down and density- or model-based alternatives [30] would deserve primacy. The robustness comparison thus strengthens the manuscript without displacing K-Means from its titular role.

Theoretically, the study operationalizes the claim that digital-payment behavior can be represented as a multidimensional observable construct $\text{Payment Behavior} = \text{Intensity} + \text{Monetary Pattern} + \text{Temporal Pattern} + \text{Merchant Diversity} + \text{Regularity}$ complementing intention-based constructs from TAM/UTAUT research (Davis, 1989; Venkatesh Morris, M. G., Davis, G. B., & Davis, F. D, 2003; Venkatesh et al., 2012) with indicators computable from records alone. This behavioral representation gives Payment Choice Theory an empirical interface at the level of revealed transaction structure rather than stated preference. Practically, and strictly conditional on validation with authorized anonymized data, the workflow prototypes several applications: merchants could profile customers, time promotions, and design category-specific or repeat-transaction offers; payment providers could monitor engagement, detect dormancy-like patterns, personalize products, and optimize merchant networks; researchers gain a reproducible testbed for segmentation features and clustering assumptions before negotiating restricted-data access; and policy stakeholders could eventually evaluate digital-payment inclusion and sectoral uptake. None of these applications is claimed as a present empirical finding.

5. Conclusion

This study asked whether a fully documented, simulation-based analytical workflow can recover and explain heterogeneous consumer payment patterns of the kind expected in QRIS transaction data. The answer is affirmative and quantified. A generator planted four latent behavioral profiles in 20,686 synthetic transactions; after principled cleaning (1.98% of records removed), eight engineered behavioral features and K-Means clustering under a combined-evidence selection rule identified four segments Frequent Low-Value, Occasional High-Value, Diverse Active, and Emerging users that recovered the planted profiles with ARI 0.947, NMI 0.920, and 97.6% matching. Frequency, average transaction value, merchant diversity, and interval regularity were the most discriminating features (ϵ^2 0.750–0.872); the solution was invariant across 30 initializations, robust across six regeneration scenarios (ARI 0.941–0.988), and method-independent against Ward clustering and Gaussian mixtures. The analysis also surfaced a practically important caution: silhouette-based selection alone would have under-segmented the data, so multi-metric selection and stability evidence are not optional refinements but necessary components.

The contribution is methodological a transparent chain from simulation design through generation, perturbation, preprocessing, feature engineering, clustering, three-layer validation, stability, and sensitivity, with every parameter and seed reported and theoretical, in representing digital-payment behavior as an observable multidimensional construct spanning intensity, monetary pattern, temporal organization, merchant diversity, and regularity. Practically, the framework is a prototype that merchants, payment providers, and researchers can recalibrate and validate on authorized anonymized transaction data. The resulting clusters are outcomes of a controlled simulation experiment and should not be interpreted as estimates of the actual distribution or behavior of QRIS consumers in Indonesia.

Future research should validate the framework on anonymized real QRIS records; integrate transaction records with consumer surveys under privacy-preserving designs; extend the method comparison to K-Medoids, DBSCAN, and fuzzy clustering; apply sequence and longitudinal analysis to detect segment migration over time; develop explainable and privacy-preserving clustering variants; and compare QRIS behavior with cash, card, mobile-banking, and electronic-wallet instruments, including geographic and merchant-sector variation, once real data become accessible.

Acknowledgements

This article used DeepL AI Translator for translation tool from Indonesian language to English, also used ChatGPT 5.6 Sol for improved readiness content to make it language more easily to understand

References

- Arango, C. (2015). Consumer payment choice: Merchant card acceptance versus pricing incentives. *Journal of Banking & Finance*, 55, 130–141.
- Arthur, D. (2007). k-means++: The advantages of careful seeding. *Proc. 18th Annual ACM-SIAM Symposium on Discrete Algorithms*, 1027–1035.
- Bagnall, J. et al. (2016). Consumer cash usage: A cross-country comparison with payment diary survey data. *International Journal of Central Banking*, 12(4), 1–61.
- Caliński, T. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, 3(1), 1–27.
- Chen, D. (2012). Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining. *Journal of Database Marketing & Customer Strategy Management*, 19(3), 197–208.
- Davies, D. L. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2), 224–227.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340. <https://doi.org/10.2307/249008>
- Figueira, A. (2022). Survey on synthetic data generation, evaluation methods and GANs. *Mathematics*, 10(15), 2733.
- Fräley, C. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, 97(458), 611–631.
- Han, J. (2012). Data Mining: Concepts and Techniques, 3rd ed. In *Morgan Kaufmann*.
- Hubert, L. (1985). Comparing partitions. *Journal of Classification*, 2(1), 193–218.
- Hughes, A. M. (1994). Strategic Database Marketing. In *Probus Publishing*.
- Ikotun, A. M. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, 622, 178–210.
- Indonesia, B. (2019). *Peraturan Anggota Dewan Gubernur No. 21/18/PADG/2019 tentang Implementasi Standar Nasional Quick Response Code untuk Pembayaran*. Bank Indonesia.
- Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651–666.
- Jordon, J. et al. (2022). Synthetic data — what, why and how? *ArXiv Preprint ArXiv:2205.03257*.
- Klee, E. (2008). How people pay: Evidence from grocery store data. *Journal of Monetary Economics*, 55(3), 526–541.
- Koulayev, S. (2016). Explaining adoption and use of payment instruments by US consumers. *RAND Journal of Economics*, 47(2), 293–325.
- Lloyd, S. P. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2), 129–137.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proc. 5th Berkeley Symposium on Mathematical Statistics and Probability*, 1, 281–297.
- Ngai, E. W. T. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications*, 36(2), 2592–2602.
- Nikolenko, S. I. (2021). Synthetic Data for Deep Learning. In *Springer*.
- Pratama, W., Sudarsono, B., & Ghazali, F. A. (2026). Integrating STEM and work-based learning to enhance competency and job readiness in maritime higher education. *Journal of Pedagogical Research*. <https://doi.org/10.33902/JPR.202645316>
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65.
- Rubin, D. B. (1993). Statistical disclosure limitation. *Journal of Official Statistics*, 9(2), 461–468.
- Undang-Undang Nomor 27 Tahun 2022 Tentang Pelindungan Data Pribadi (2022).

- Venkatesh Morris, M. G., Davis, G. B., & Davis, F. D, V. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425–478. <https://doi.org/10.2307/30036540>
- Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly*, 36(1), 157–178.
- Verplanken, B. (2003). Reflections on past behavior: A self-report index of habit strength. *Journal of Applied Social Psychology*, 33(6), 1313–1330.
- Vinh, N. X. (2010). Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *Journal of Machine Learning Research*, 11, 2837–2854.
- Wardoyo, S., Mardapi, D., Djatmiko, I. W., & Pratama, W. (2020). Professional competence requirements for maritime teachers in Indonesia. *International Journal of Advanced Science and Technology*, 29(1), 1504–1516. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85081190172&partnerID=40&md5=0103ea3d7236c865f832c253583b25e5>
- Wedel, M. (2000). Market Segmentation: Conceptual and Methodological Foundations, 2nd ed. In *Springer*.