Journal of Information Technology and Cyber Security 4(2) July 2026: 98-106 EISSN 2987-386X | PISSN 2987-3878
DOI:10.30996/jitcs.133977

Financial-Capacity Segmentation of Registered MSMEs in Bekasi City: A Reproducible K-Means Study

Rafika Sari [1], Michelle Ledisty Aisha [2], Ratna Salkiawati [3,*], Khairunnisa Fadhilla Ramdhania [4]

[1,2,3,4] Department of Informatics, Universitas Bhayangkara Jakarta Raya, Jakarta, Indonesia
* Corresponding author: ratna_tind@dsn.ubharajaya.ac.id

Received: 13 August 2026
Accepted: 19 August 2026
Revised: 19 August 2026
Available online: 20 August 2026

To cite this article: Sari, R., Aisha, M. L., Salkiawati, R., & Ramdhania, K. F. (2026). Financial-Capacity Segmentation of Registered MSMEs in Bekasi City: A Reproducible K-Means Study. *Journal of Information Technology and Cyber Security, 4*(2), 98-106. https://doi.org/10.30996/jitcs.133977

Abstract

The heterogeneity of micro, small, and medium enterprises (MSMEs) limits the usefulness of uniform development programmes. This study develops a transparent financial-capacity segmentation of registered MSMEs in Bekasi City using the Knowledge Discovery in Databases framework. The administrative source contained 7,964 records and 15 fields for 2019–2024. After data-quality screening and exclusion of 41 records with unusable capital or turnover values, 7,923 records were analysed. Business capital and turnover were selected as the two K-Means inputs because they directly represent financial capacity and were sufficiently complete for all retained observations. District, a nominal variable, was excluded from Euclidean-distance calculation and used only for post-hoc geographic profiling. The revised implementation specifies k-means++ initialisation, random_state = 42, n_init = 50, max_iter = 300, and tolerance = 10−4. A common candidate range of K = 2–10 was evaluated using inertia and the Davies–Bouldin Index (DBI). Both diagnostics supported K = 3; the minimum DBI was 0.5940. The clusters comprised 7,789 (98.31%), 22 (0.28%), and 112 (1.41%) businesses, representing low capital–low turnover, high capital–very high turnover, and high capital–moderate turnover profiles. The contribution is not a new clustering algorithm, but a reproducible city-scale evidence pipeline that corrects nominal-feature handling, reports the preprocessing audit, and converts financial profiles into testable programme hypotheses. Because the two minority clusters may represent financial extremes, cluster membership should be verified with business sector, age, employment, and stakeholder evidence before policy use.

Keywords: Bekasi city, davies-bouldin index, k-means clustering, msme segmentation, unsupervised learning.

1. Introduction

Micro, small, and medium enterprises (MSMEs) support employment, household income, innovation, and local production, but they are not a homogeneous policy population. Capital capacity, realised turnover, sector, age, employment, legality, digital readiness, and location can differ markedly across firms. Consequently, a single assistance package may be too basic for growth-oriented firms and inaccessible to subsistence businesses. A defensible segmentation can help local authorities formulate differentiated programme hypotheses, although clustering alone cannot determine eligibility or causal policy priorities.

K-Means is an unsupervised partitioning method that minimises within-cluster squared Euclidean distances. Its computational efficiency and interpretability make it suitable for large numerical administrative datasets, but its output depends on scale, initialisation, outliers, feature representation, and the selected number of clusters (Han et al., 2023). Recent segmentation research increasingly combines internal validity measures, explicit preprocessing, and substantive profiling rather than treating a visually plausible partition as sufficient evidence (Uddin et al., 2024; Kumar et al., 2025). For nominal attributes, arbitrary integer codes

© 2026 The Author(s). Published by Department of Information Systems and Technology, Universitas 17 Agustus 1945 Surabaya, Indonesia. This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0/), which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial, and no modifications or adaptations are made.

must not be treated as metric quantities because Euclidean distance would impose artificial order and spacing.

Studies directly related to MSME clustering show both the usefulness and the methodological limits of standard K-Means. Sudrajat et al. (2022) used RapidMiner to group MSMEs; Ameliana et al. (2024) analysed West Java MSME records; and Akbar and Zahrotun (2025) compared conventional K-Means with genetic-algorithm centroid optimisation on Yogyakarta MSME data. The latter reported that optimised initialisation reduced DBI from 1.819 to 1.349, illustrating that cluster quality is sensitive to model configuration. Spatial research on Indonesian MSMEs has also shown that geographic patterns require meaningful spatial representations rather than nominal administrative codes (Widita et al., 2024). These studies provide a relevant comparison set, whereas classification or association-rule studies are not used as evidence for the performance of MSME clustering.

The available Bekasi administrative file offers city-scale coverage but also creates a specific methodological problem: it contains 15 fields of mixed measurement type, whereas conventional K-Means requires meaningful numerical inputs. Capital and turnover were retained because they align directly with the research question, are available for every retained record, and can be compared after standardisation. District was retained as contextual metadata for post-hoc tabulation and mapping, not as a K-Means input. Names, NIK, telephone numbers, and addresses are identifiers; NIB represents legality rather than financial capacity; business type is categorical and substantially incomplete; and employee count requires a separate completeness and construct-validity assessment. This separation of modelling features from descriptive metadata is central to the revised design.

The research gap is therefore practical and methodological rather than algorithmic. Existing applied MSME clustering papers often provide limited preprocessing audits, omit executable parameter settings, mix categorical codes with Euclidean features, or make policy claims without distinguishing descriptive clusters from validated needs. The present study does not claim a new K-Means variant. Its originality lies in a transparent city-scale workflow comprising an explicit data dictionary and retention audit, defensible exclusion of district codes from the distance function, a consistent K = 2–10 evaluation range, reproducible parameter reporting, quantified cluster imbalance, and restrained translation of profiles into hypotheses for stakeholder validation.

Accordingly, the objectives are to: (1) document the profile and preprocessing of the Bekasi City MSME administrative dataset; (2) identify financial-capacity segments using capital and turnover; (3) select and evaluate K consistently using inertia and DBI; (4) describe cluster size, centroid patterns, and geographic distribution without treating district codes as metric data; and (5) formulate preliminary, testable implications while defining the limits of generalisation. The findings apply to retained, registered MSME records in Bekasi City for 2019–2024 and should not be generalised automatically to unregistered firms or other cities.

Table 1
Dataset profile and reproducible analytical configuration.

Element	Description
Source and period	Bekasi City MSME administrative file, 2019–2024
Initial/ retained records	7,964 / 7,923 (99.49% retained)
Source fields	15 mixed-type administrative fields
K-Means inputs	Capital and turnover; both numerical and standardised
Contextual variable	District; nominal, excluded from Euclidean distance and used post hoc
Implementation	Python; pandas, scikit-learn, and Yellowbrick
K-Means settings	init='k-means++'; random_state=42; n_init=50; max_iter=300; tol=1e−4
Candidate K and evaluation	K=2–10 for inertia and DBI; K=1 shown only as the inertia baseline

2. Methods

2.1. Research Design and Data Source

This quantitative, descriptive data-mining study used secondary administrative records supplied by the Bekasi City authority responsible for MSMEs. The observation period was 2019–2024. The source contained 7,964 rows and 15 fields: record number; owner name and NIK; residential street, village, and district; telephone; business street, village, and district; business type; employee count; capital; turnover; and Business Identification Number (NIB). Personally identifiable information was removed prior to data preprocessing and modelling. The unit of analysis was an individual registered business record. Accordingly,

the study characterizes the administrative population available after eligibility screening rather than representing a probability sample of all MSMEs operating in Bekasi City.

2.2. Data Preprocessing

The data-quality audit was reported at field and record levels. In the 15-field source, telephone had 3 missing values, business type 1,271, and NIB 2,793. These fields were not K-Means inputs; missing categorical values were labelled 'Unknown' for descriptive use rather than statistically imputed. Exact-row duplicates and repeated identity/address combinations were inspected; only confirmed full duplicates were eligible for removal, whereas partial matches were retained unless manually verified. Capital and turnover were converted to numeric format, and 41 records with zero, missing, or otherwise unusable values in either modelling feature were excluded, reducing the file from 7,964 to 7,923 observations. No retained observation had a missing model value. Capital and turnover were standardised using parameters fitted on the retained dataset. District was not integer-encoded for the revised model; it remained a nominal label used after clustering to describe within-cluster geographic frequencies. This audit makes clear that missing values in non-model administrative fields did not cause the 41-record modelling exclusion.

2.3. Data Science Framework: KDD and Its Relationship to CRISP-DM

KDD was selected because the purpose was to discover and interpret patterns in an existing administrative database. The implemented stages were: (1) selection—define the registered business record as the unit of analysis and separate model features from identifiers and contextual fields; (2) preprocessing—audit missingness and duplicates, convert financial fields, and apply explicit eligibility rules; (3) transformation—standardise capital and turnover while retaining district as nominal metadata; (4) data mining—fit K-Means for candidate K values using the fixed configuration in Table 1; and (5) evaluation and interpretation—compare inertia and DBI, inspect centroids and cluster sizes, and develop cautious descriptive profiles. Each decision is reported so another researcher with authorised access to the source file can reproduce the analytical dataset.

CRISP-DM provides a complementary project perspective: targeted MSME assistance is the business-understanding problem; the field and missingness audit is data understanding; filtering and scaling constitute preparation; K-Means is modelling; internal indices and profile checks are evaluation; and stakeholder testing would precede deployment. This article stops before operational deployment. It therefore reports a KDD-based analytical contribution, not an autonomous policy-allocation system.

2.4. K-Means Clustering and Selection of K

K-Means was fitted only to the two-dimensional standardised vector x=(capital, turnover). For each candidate K, k-means++ selected initial centres, 50 independent initialisations were evaluated, and the solution with the lowest inertia was retained. The fixed settings were random_state=42, max_iter=300, and tol=10−4. Assignment to the nearest centroid and centroid updating continued until convergence. Euclidean distance was appropriate for these standardised numerical features; district labels were deliberately excluded because numeric district codes would create non-existent distances.

$$d(x,\mu) = \sqrt{\left[\sum_{j=1}^{m}(x_j - \mu_j)^2\right]} \quad (1)$$

A consistent candidate range of K=2–10 was used for model comparison. K=1 appears in the Elbow plot only as a no-segmentation inertia baseline because DBI is undefined for a one-cluster solution. For every K from 2 to 10, inertia and DBI were computed from the same standardised feature matrix. The Elbow point was interpreted as the onset of diminishing reductions in inertia, while the minimum DBI identified the best balance between within-cluster compactness and between-cluster separation.

$$DBI = \left(\frac{1}{K}\right)\sum_{i=1}^{K} max_{i\neq j}\left[\frac{S_i+S_j}{M_{ij}}\right] \quad (2)$$

DBI is an internal descriptive validation index, not an inferential statistic. P-values, confidence intervals, and effect sizes are therefore not inherent outputs of K-Means, and the administrative data are not a probability sample supporting population inference beyond the retained records. To avoid overstating precision, the study reports exact cluster counts, shares, centroids, inertia patterns, and DBI across candidate K values. Future work with the de-identified row-level dataset should add repeated-seed or bootstrap stability analysis (e.g., adjusted Rand index across resamples) and external validation against independent business outcomes.

3. Results and Discussion

3.1. Selection of the Number of Clusters

Figure 1 includes K=1 as an inertia baseline and evaluates candidate partitions through K=10. Inertia declines sharply up to K=3 and then flattens, indicating diminishing returns from additional clusters. Because the visual location of an elbow can be subjective, the selection was checked using DBI over the same valid partition range, K=2–10.

The Elbow Method evaluates how within-cluster sum of squares decreases as K increases. It is a complexity–compactness diagnostic rather than a statistical significance test. In this dataset, K=3 is the most parsimonious point before the reduction becomes gradual.

The agreement of the Elbow pattern and the minimum DBI supports K=3 for descriptive profiling. This agreement does not prove that three natural or policy-ready populations exist; it shows that K=3 is the best-supported partition among the tested conventional K-Means specifications.

DBI compares average within-cluster dispersion with separation between cluster centroids. Lower values indicate a more favourable internal geometry. The minimum reported value was 0.5940 at K=3. K=4 was close (0.6003), so parsimony and interpretability, rather than a large numerical advantage, also support retaining three clusters.

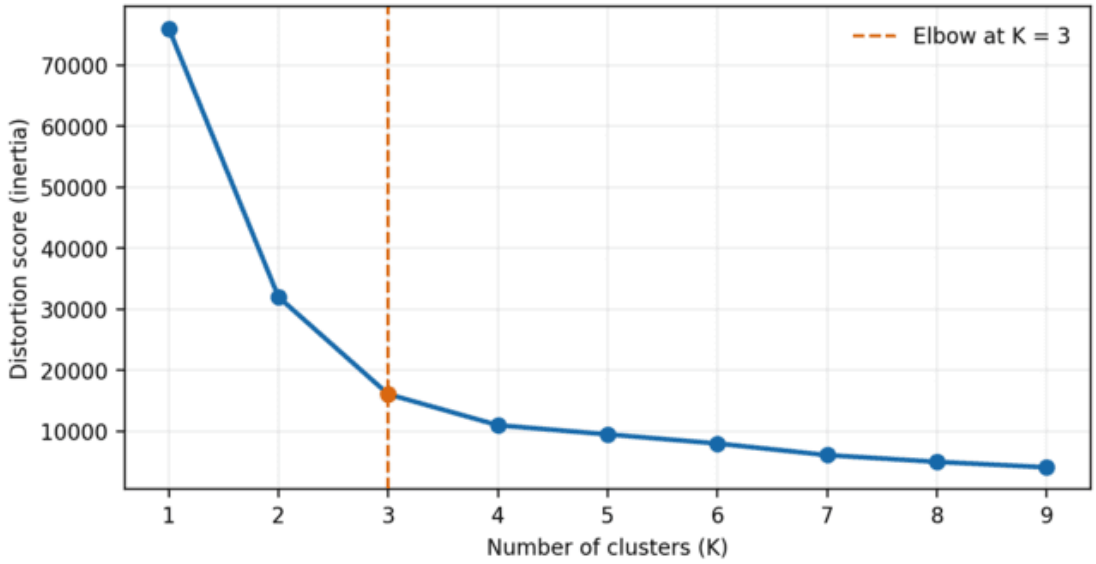

Fig. 1. Elbow curve for candidate K values.

Table 2
Davies–Bouldin Index across reported candidate partitions (the K=2 output should be inserted from the analysis log before resubmission).

K	DBI	Interpretation
3	0.5940	Minimum; selected
4	0.6003	Close alternative
5	0.7914	Weaker separation
6	0.7583	Weaker separation
7	0.7966	Weaker separation
8	0.7066	Weaker than K = 3
9	0.7471	Weaker than K = 3
10	0.6938	Weaker than K = 3

Table 2 and Figure 2 report DBI for K=3–10; K=2 was also evaluated but should be added to the final figure and table from the analysis output so that the reported range fully matches the protocol. K=1 is correctly omitted because DBI requires at least two clusters. Among the values currently available, K=3 produced the minimum DBI (0.5940), narrowly below K=4 (0.6003).

Figure 2 presents the Davies–Bouldin Index values obtained from the evaluation of clustering solutions ranging from K=3 to K=10. The lowest DBI value, 0.5940, was achieved at K=3, indicating that the three-cluster solution provides the most favorable combination of within-cluster compactness and between-

cluster separation among the evaluated alternatives. In principle, a lower DBI value represents a better clustering structure because the members of each cluster are more similar to one another while remaining more distinguishable from members of other clusters.

The DBI value at K=4 is approximately 0.600, which is only slightly higher than the value obtained at K=3. Although this difference is relatively small, increasing the number of clusters from three to four does not improve the overall clustering quality. Instead, it introduces an additional cluster without producing better compactness or separation. Therefore, the K=3 solution is preferable because it achieves the minimum DBI while maintaining a simpler and more interpretable cluster structure.

A substantial increase in the DBI value is observed at K=5, followed by fluctuating values for K=6 to K=10. This non-monotonic pattern indicates that adding more clusters does not necessarily improve the K-Means results. At higher values of K, some naturally cohesive groups may be divided into smaller clusters, while other clusters may become relatively close to one another. These conditions increase the similarity ratio used in the DBI calculation and consequently produce higher index values. The highest DBI value occurs around K=7, suggesting that this configuration produces the weakest balance between cluster compactness and separation among the tested solutions.

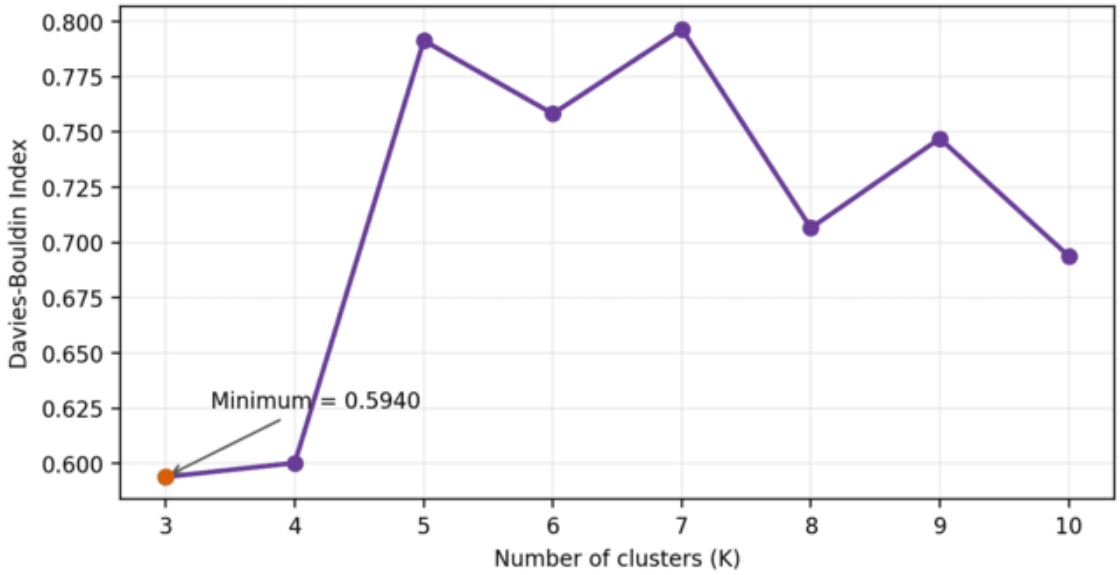

Fig. 2. Davies-Bouldin Index across candidate K values.

Table 3
Final cluster centroids and descriptive profiles.

Cluster	Capital centroid	Turnover centroid	Descriptive profile
0	0.287931	0.589880	Low capital–low turnover
1	12.560453	32.511988	High capital–very high turnover
2	14.402096	4.139508	High capital–moderate turnover

Table 4
Cluster distribution and indicative policy implications.

Cluster	Number of MSMEs	Share (%)	Priority implication
0	7,789	98.31	Basic finance, bookkeeping, digital marketing, and market access
1	22	0.28	Expansion, partnerships, and wider market penetration
2	112	1.41	Productivity and capital-utilisation diagnostics

The minimum DBI must be interpreted together with the extreme size imbalance: 7,789, 22, and 112 observations. This distribution may represent a dense low-financial-capacity majority and two economically distinctive groups, but it may also be driven by extreme monetary observations. The revised interpretation therefore treats the minority clusters as profiles requiring verification, not as automatically stable programme categories.

Overall, K=3 is the most parsimonious supported solution, but internal indices assess geometry only. They do not establish causal explanations, stakeholder-defined needs, or generalisability. Repeated-seed

stability, robust transformations, alternative algorithms, and external outcomes are needed before operational adoption.

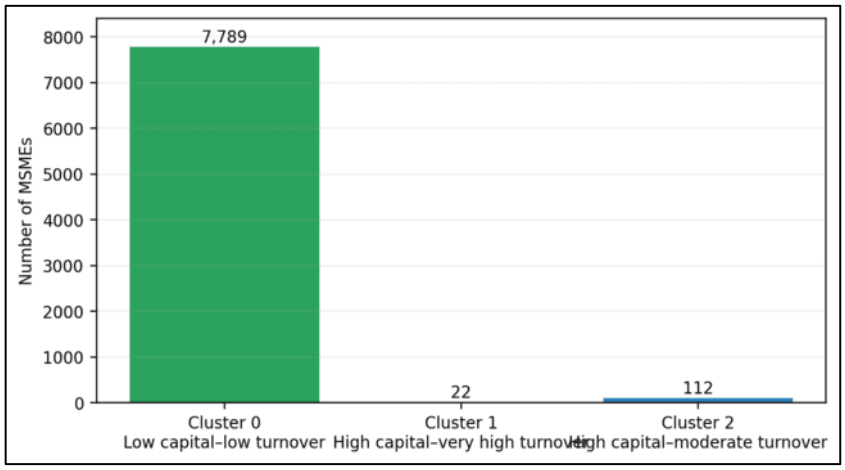

Fig. 3. Distribution of MSMEs across the three clusters.

3.2. Cluster Profiles

The final two-feature model converged after 10 iterations. Table 3 reports centroids in the transformed capital–turnover space. District is absent from the centroid table because it was not a metric input; district distributions should instead be cross-tabulated after cluster assignment. Profile labels are descriptive summaries and are not statutory Indonesian MSME classifications.

Cluster 0 contains 7,789 businesses (98.31% of all observations), while Cluster 1 contains 22 (0.28%) and Cluster 2 contains 112 (1.41%). Figure 3 makes the strong imbalance visible. The dominance of Cluster 0 indicates that the dataset is concentrated among businesses with relatively low capital and turnover. Cluster 1 represents a very small set with exceptionally high turnover, while Cluster 2 contains businesses with high capital but more moderate turnover. The latter profile may warrant investigation of capital utilisation, operating costs, sector differences, or business maturity.

The centroid pattern is more informative than cluster size alone. Cluster 1 has turnover far above capital, while Cluster 2 has the highest capital but lower turnover than Cluster 1. These contrasts generate hypotheses about working-capital circulation or asset utilisation, but they do not demonstrate productivity or causation because sector, business age, profit, costs, and employment were not included. The profiles were derived by the research team from centroid direction and magnitude; no domain-expert elicitation was conducted. This is now stated explicitly, and stakeholder validation is required before the labels are institutionalised.

A 138-record hand calculation was used only to illustrate assignment and centroid updating; it was not a validation sample. The full model used all 7,923 eligible records and required 10 iterations. Reproducibility depends on the full preprocessing rules and parameters reported above, not on agreement between the demonstration sample and the complete analysis.

3.3. Discussion and Practical Implications

The principal strength is the scale and administrative relevance of the dataset: 7,923 retained records cover multiple years and all registered districts represented in the Bekasi City file. A second strength is methodological transparency. The revision separates numerical model features from nominal contextual variables, reports the retention and missingness audit, fixes the implementation settings, and applies a consistent candidate-K protocol. These steps make the work more useful as a reproducible applied-data-science case, even though the underlying K-Means algorithm is standard.

Comparison with directly relevant studies clarifies the contribution. Sudrajat et al. (2022) and Ameliana et al. (2024) likewise demonstrate the practical utility of K-Means for MSME grouping. Akbar and Zahrotun (2025) obtained DBI values of 1.819 for conventional K-Means and 1.349 after genetic-algorithm optimisation on a different Yogyakarta dataset. The present DBI of 0.5940 is numerically lower, but cross-study ranking is inappropriate because feature definitions, transformations, sample composition, and K differ. The valid comparison is methodological: recent work shows that initialisation and representation materially affect cluster quality, supporting this study's explicit settings and correction of district handling.

The findings are limited to registered Bekasi City MSMEs retained from the 2019–2024 administrative file. They do not represent unregistered businesses, other Indonesian cities, or a probability sample. The

model uses only capital and turnover; both may be skewed, self-reported, measured at different time points, or affected by data-entry error. The extreme cluster imbalance suggests sensitivity to financial outliers. Internal validation does not demonstrate temporal stability, external validity, or programme effectiveness. Profile labels were assigned by the researchers without formal domain-expert consensus, and a district map cannot be produced defensibly without coordinates or a verified district-by-cluster cross-tabulation.

The appropriate practical use is therefore a screening and hypothesis-generation layer. Cluster 0 can motivate scalable foundational support; Cluster 1 can prompt assessment for expansion assistance; and Cluster 2 can prompt investigation of capital utilisation. These are provisional service hypotheses, not direct policy priorities or eligibility rules. Officers should combine cluster membership with business type, age, employment, legality, verified financial records, and owner consultation. Outcomes—such as turnover growth, business survival, financing access, and digital adoption—must subsequently be evaluated within each segment.

Future analysis should (1) inspect monetary distributions and compare standard, robust, and logarithmic transformations; (2) quantify repeated-seed and bootstrap stability; (3) compare K-Means with K-Medoids, Gaussian mixtures, density-based methods, and mixed-data clustering when categorical fields are included; (4) use coordinates or district polygons for a separate spatial layer; (5) obtain domain-expert review of profile names; and (6) validate clusters against independent outcomes. These extensions would move the work from a transparent descriptive segmentation toward a stronger decision-support model.

The extreme imbalance also deserves methodological attention. K-Means minimises total within-cluster variance and can form very small clusters around extreme financial values. Consequently, Clusters 1 and 2 may partly capture outliers or niche sectors rather than stable policy populations. Future analysis should compare robust scaling, log transformations of financial variables, K-Medoids, Gaussian mixture models, or density-based clustering. In addition, a categorical district code has no inherent order or equal spacing; using it directly in Euclidean distance may introduce artificial proximity. Spatial coordinates, one-hot encoding combined with a mixed-data distance, or separate spatial analysis would provide more defensible geographical information.

4. Conclusions

This study produced a reproducible financial-capacity segmentation of 7,923 retained, registered MSMEs in Bekasi City using standardised capital and turnover. District was corrected from an invalid Euclidean input to a post-hoc nominal descriptor. The Elbow pattern and the reported DBI series support K=3, with a minimum DBI of 0.5940 and cluster sizes of 7,789, 22, and 112. The contribution is an auditable city-scale workflow and a cautious translation of centroid profiles into testable service hypotheses—not a new algorithm and not a stand-alone policy-allocation mechanism. The extreme imbalance, limited financial feature set, lack of formal expert labelling, and absence of stability and external-outcome tests constrain interpretation. Accordingly, the results apply only to the retained Bekasi administrative records and require robust-transformation checks, repeated-run validation, geographic cross-tabulation, and stakeholder confirmation before programme use.

5. Declaration of AI and AI assisted technologies in the writing process

During the preparation of this manuscript, the authors used ChatGPT to assist with language editing and assisting in translating the manuscript into English. The authors reviewed and edited the resulting text, verified the reported data against the research documents, and take full responsibility for the content of the publication.

6. CRediT Authorship Contribution Statement

Rafika Sari: Conceptualization, methodology, supervision, validation, writing – review and editing. **Michelle Ledisty Aisha**: Data curation, formal analysis, software, visualization, writing original draft. **Ratna Salkiawati**: Conceptualization, supervision, validation, writing – review & editing. **Khairunnisa Fadhilla Ramdhania**: Methodology, validation, resources, writing – review & editing.

7. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

8. Acknowledgments

The authors thank the Bekasi City authority responsible for MSME data for providing the administrative dataset used in this study.

9. Data Availability

The administrative dataset contains sensitive business information and is not publicly distributed. Aggregated results and non-identifying analytical information may be made available upon reasonable request, subject to permission from the data provider.

10. Funding

This research received no specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

11. Ethical Approval

This study used secondary administrative business records and did not involve patients or biomedical interventions. Personal identifiers were excluded from the reported analysis. No formal ethical approval was requested for this non-interventional study.

13. References

Akbar, M. F., & Zahrotun, L. (2025). K-Means centroid optimization with genetic algorithm for clustering micro, small, medium enterprises in Yogyakarta. JUITA: Jurnal Informatika, 13(2), 87–97. https://doi.org/10.30595/juita.v13i2.25480

Kumar, N., Rahman, S., & Mahanti, A. (2025). Intelligent customer segmentation: Unveiling consumer characteristics using machine learning. Journal of Umm Al-Qura University for Engineering and Architecture, 16, 774–783. https://doi.org/10.1007/s43995-025-00180-7

Uddin, M. A., Talukder, M. A., Ahmed, M. R., Khraisat, A., Alazab, M., Islam, M., Aryal, S., & Jibon, F. A. (2024). Data-driven strategies for digital native market segmentation using clustering. International Journal of Cognitive Computing in Engineering, 5, 178–191. https://doi.org/10.1016/j.ijcce.2024.04.002

Widita, A. A., et al. (2024). Spatial patterns and drivers of micro, small and medium-sized enterprises within and across Indonesian cities: Evidence from highly granular data. Regional Sustainability, 5, 100137. https://doi.org/10.1016/j.rspp.2024.100137

Ameliana, N., Suarna, N., & Prihartono, W. (2024). Analisis data mining pengelompokkan UMKM menggunakan algoritma K-Means clustering di Provinsi Jawa Barat. JATI (Jurnal Mahasiswa Teknik Informatika), 8(3), 3261–3268.

Davies, D. L., & Bouldin, D. W. (1979). A Cluster Separation Measure. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-1(2), 224–227. https://doi.org/10.1109/TPAMI.1979.4766909

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. AI Magazine, 37–54.

Fedorova, O., Sari, R., & Setiawati, S. (2024). Analysis of the actviness library members using K-Means clustering. Journal of Engineering and Scientific Research, 7(2), 104–111.

Han, J., Pei, J., & Tong, H. (2023). Data mining: Concepts and techniques (4th ed.) (K. Morgan, Ed.; 4th ed.). Morgan Kaufmann.

Larasati, A., Hajji, A. M., Handayani, A. N., Azzahra, N., Farhan, M., & Rahmawati, P. (2019). Profiling academic library patrons using K-Means and X-Means clustering. International Journal of Technology, 10(8), 1567–1575.

Lloyd, S. (1982). Least squares quantization in PCM. IEEE Transactions on Information Theory, 28(2), 129–137. https://doi.org/10.1109/TIT.1982.1056489

Sari, R., & Ashardi, M. (2025). Prediction Model for Product Stock Procurement Using the Naive Bayes Method. International Journal of Information Technology and Computer Science Applications, 3(1), 1–10. https://doi.org/10.58776/ijitcsa.v3i1.172

Sari, R., Fatoni, H., & Ramdhania, K. F. (2023). Decision Support System Design for Informatics Student Final Projects Using C4.5 Algorithm. PIKSEL : Penelitian Ilmu Komputer Sistem Embedded and Logic, 11(1), 123–134. https://doi.org/10.33558/piksel.v11i1.5954

Sari, R., Helmy, N., & Alexander, A. D. (2024). Determining Sales Patterns Using the Apriori Algorithm: A Case Study of Unlocked Cafe's Website Applications. PIKSEL : Penelitian Ilmu Komputer Sistem Embedded and Logic, 12(1), 189–200. https://doi.org/10.33558/piksel.v12i1.8908

Sudrajat, W., Cholid, I., & Petrus, J. (2022). Penerapan algoritma K-Means clustering untuk pengelompokan UMKM menggunakan RapidMiner. Jurnal Jupiter, 14(1), 27–36.

Sukmadewanti, I., Arifudin, R., & Sugiharti, E. (2018). Use of K-Means clustering and analytical hierarchy process in determining the type of MSME financing in Semarang City. Scientific Journal of Informatics, 5(2).

Suyanto. (2019). Data Mining untuk Klasifikasi dan Klasterisasi Data (Edisi Revisi). Informatika.

Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. In Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining, 29–39.