Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 3, Issue 5, 2024

An Explainable AI Framework for Customer Segmentation and Personalized Marketing Decision-Making in Online Retail

Asfandyar Khan[1], Majid Khan[2] Awais Khan[3] and Salman Khan[4]

[1]Institute of Computer Sciences & Information Technology (ICS/IT), The University of Agriculture, Peshawar, Pakistan
[2]Department of Mathematics, Statistics and Computer Science, The University of Agriculture, Peshawar, Pakistan
[3]IBMS, University of Science & Technology Bannu, Pakistan
[4]Department of Computer Engineering, Balochistan University of Information Technology, Engineering and Management Sciences, Pakistan

asfandyar@aup.edu.pk[1], majidkhan@aup.edu.pk[2], salman.ce27@gmail.com[3], awaisuni@gmail.com[4]

DOI: https://doi.org/10.5281/zenodo.22857473

Keywords
Customer Segmentation, Explainable AI, SHAP, K-Means, Gaussian Mixture Model, Hierarchical Clustering, RFM Analysis, Personalized Marketing, Online Retail, Unsupervised Learning

Article History
Received on 20 Apr 2025
Accepted on 20 May 2025
Published on 30 May 2025

Copyright @Author

Corresponding Author: *

Asfandyar Khan
Institute of Computer Sciences & Information Technology (ICS/IT), The University of Agriculture, Peshawar, Pakistan
asfandyar@aup.edu.pk

Abstract

Customer segmentation is a basis of personalized marketing, however, unsupervised clustering algorithms are fundamentally opaque: they cluster customers, but do not explain why a particular customer belongs in a particular cluster in interpretable terms. In this work, we propose and empirically validate an explainable AI (XAI) framework that combines unsupervised customer segmentation with post-hoc, surrogate-model-based explanation. Using the Online Retail II transactional dataset (1,067,371 line items; 5,833 retained customers after cleaning), Recency-Frequency-Monetary-plus (RFM+) behavioral features were engineered and three clustering algorithms - K-Means, Gaussian Mixture Models (GMM), and Ward-linkage Agglomerative Hierarchical Clustering - were applied and compared using silhouette score, Davies-Bouldin index, and Calinski-Harabasz index. The solution was applied uniformly for all three algorithms for fair comparison. The four-segment solution (k = 4) was chosen for managerial interpretability in line with standard RFM segmentation practice, although the statistically optimal K-Means solution by silhouette score was k=2. Of the three methods, K-Means provided the best separation of clusters at k=4 (silhouette=0.253), better than Hierarchical (0.209) and GMM (0.106). K-Means produced geometrically similar, well-separated segments in principal-component space, while GMM produced very overlapping soft boundaries. To make the resulting segments explainable, a Random Forest surrogate classifier was trained to mimic the K-Means partition (fidelity accuracy = 94.7%, macro-F1 = 0.949). Shapley Additive exPlanations (SHAP) were calculated with Monte Carlo permutation sampling to measure which behavioral features drive each customer's segment assignment. The four resulting segments (Champions, 22.3% of customers, driven by high Monetary value, TotalItems, and Frequency; At-Risk, 33.2%, driven by moderate but declining engagement; New/Promising, 21.3%, driven primarily by low Tenure; and Lost/Dormant, 23.2%, driven by extreme Recency and low Monetary value) have theoretically coherent and managerially actionable SHAP signatures. The study shows that the combination of classical clustering with surrogate-model SHAP explanation provides a statistically grounded, cluster-algorithm-agnostic, and transparent

segmentation framework that can be used to support personalized, feature-targeted marketing interventions for each identified segment.

INTRODUCTION

Personalized marketing requires the ability to segment a heterogeneous customer base into behaviorally coherent groups that can each be assigned customized messaging, pricing and retention treatment. The standard technical tools for this task are unsupervised clustering algorithms such as K-Means, Gaussian Mixture Models and hierarchical methods that are based on behavioral features such as recency, frequency and monetary value (RFM) of purchases and do not require labeled outcome data. However, clustering algorithms are black boxes in one specific and consequential way: they generate group assignments without any accompanying explanation of which features led a given customer to a given group, or how confidently that assignment can be trusted. For a marketing team trying to figure out how to act on a segmentation, the difference between "this customer is in Segment 2" and "this customer is in Segment 2 because their purchase frequency has plummeted over the last quarter" is the difference between a label and a decision-support tool.

This study addresses this gap by developing an explainable AI (XAI) framework for customer segmentation, using the Online Retail II dataset - a two-year transaction log from a UK-based online gift retailer. The framework is composed of two steps: first, three clustering algorithms are compared based on cluster validity on the RFM-derived customer features; second, a supervised surrogate model is trained to replicate the segmentation obtained, so that Shapley Additive exPlanations (SHAP) can be used to measure the contribution of each behavioral feature to the segment assignment of each customer, a technique usually applied to supervised classifiers, but here adopted as a general-purpose layer of explainability for unsupervised segmentation.

The paper poses three research questions:

RQ1: Which clustering algorithm (K-Means, GMM or Ward-linkage hierarchical clustering) provides the best separated and internally coherent customer segmentation on RFM-derived online retail data?

RQ2: To what extent can a supervised surrogate model mimic an unsupervised clustering result and thereby enable the use of standard supervised explainability methods for segmentation?

RQ3: What behavioral characteristics underlie assignment to each customer segment and are these drivers theoretically consistent, marketing-actionable segment profiles?

Literature Review

The advancement in the digital technologies has been increased rapidly during last decades (Bangash et al., 2014). Recency-Frequency-Monetary (RFM) analysis is the leading feature-engineering framework for behavioral customer segmentation in retail, owing to its simplicity, interpretability, and well-established linkage to customer lifetime value. Extensions of the base RFM model – tenure, average basket value, and product-variety metrics – have been shown in the retail-analytics literature to improve the granularity of segmentation relative to the three-variable model alone.

To put it simply, K-Means assumes clusters of roughly equal size and spherical shape and tries to find k clusters by minimizing the within-cluster variance relative to the cluster centroids in an iterative manner. It is computationally efficient and commonly used for customer segmentation in large datasets (Batool et al., 2016). Gaussian Mixture Models extend K-Means by modeling each cluster as a multivariate Gaussian distribution with its own covariance structure, leading to soft (probabilistic) cluster assignments that can better accommodate overlapping or elongated

customer groups, but at the cost of increased sensitivity to initialization and a well-known tendency for likelihood-based model selection criteria like BIC to favor increasingly complex models as the number of components grows (Ullah et al., 2019). Agglomerative hierarchical clustering with Ward linkage builds a hierarchy of clusters by iteratively merging the pair of clusters that results in the smallest increase in within-cluster variance, and the resulting dendrogram offers additional insight into the nested structure of customer subgroups at different levels of granularity.

Clustering is an unsupervised task and there are no ground-truth labels to validate the obtained partitions directly. Internal validity indices are standard tools for comparing candidate partitions and choosing the number of clusters (Tariq and Khattak, 2019). They include the silhouette coefficient (which measures the similarity of a point to its own cluster relative to the nearest neighboring cluster), the Davies-Bouldin index (which measures the average similarity between each cluster and its most similar counter, where lower is better) and the Calinski-Harabasz index (the ratio of between-cluster to within-cluster dispersion, where higher is better). However, they can disagree with each other and with managerial notions of an actionable number of segments (Khan et al., 2018).

The majority of explainability research has been in the setting of supervised prediction, where methods such as SHAP (Lundberg & Lee, 2017) explain the model output in terms of the input features using Shapley values from cooperative game theory. The direct application of such techniques to unsupervised clustering is non-trivial, since clustering algorithms do not naturally expose a differentiable or feature-attributable scoring function (Iqbal et al., 2013). A growing body of work tries to remedy this by fitting an interpretable or explainable supervised surrogate model to approximate a black-box clustering solution, and then applying standard supervised explainability techniques to the surrogate as a proxy for the original clustering (Imran and Khattak, 2019). This approach sacrifices a small amount of fidelity for much improved interpretability, as long as the surrogate's reproduction accuracy is validated and reported.

While RFM-based segmentation and comparative clustering studies are abundant in the retail analytics literature, few studies combine multiple clustering algorithms, rigorous internal validity comparison, and surrogate-based SHAP explainability within a single empirical pipeline on large-scale real transactional data. This work fills that gap of integration (Khattak et al., 2022).

Research Methodology

The following sub-sections discuss the proposed research methodology.

Dataset and Data Cleaning

Datasets are playing the vital role in many applications of Artificial Intelligence (Zaman et. al., 2020, Khalil et. al., 2016). The data used in this study is the Online Retail II dataset. It contains 1,067,371 transaction line items over the time period of three years, December 2009 to December 2011, for an online retailer based in the UK who sell gifts across 43 countries. The data cleaning process removed transactions with missing Customer ID (243,007 rows), canceled orders (identified by an invoice number starting with 'C') (19,494 rows), and rows with non-positive quantity or price (due to returns, adjustments or data entry artifacts) for a total of 805,549 valid transaction lines associated to 5,878 unique customers.

Feature Engineering: RFM+ Customer Profile

For each of the customers, we computed seven behavioral features relative to a snapshot date one day after the last transaction in the dataset (2011-12-10): Recency (days since the customer's last purchase), Frequency (number of distinct invoices), Monetary (total spend

across all transactions), Average Basket Value (mean spend per invoice), Total Items (total units purchased), Distinct Products (number of unique stock-keeping units purchased), and Tenure (days between the customer's first and last purchase). To avoid the dominance of distance-based clustering by a small number of extreme outliers (e.g. wholesale buyers), customers with Monetary or Frequency values in the top 0.5% were removed. This yielded a final analytic sample of 5,833 customers. Log-transformation (log(1+x)) was applied to the count and monetary features (Frequency, Monetary, Average Basket Value, Total Items, Distinct Products) and all seven features were standardized to zero mean and unit variance prior to clustering.

Clustering Algorithms

Three unsupervised clustering algorithms were implemented in scikit-learn 1.8:

K-Means: a partitional clustering that minimizes the sum of squares within clusters. The model was fit with 20 random initializations to mitigate sensitivity to initial centroids.

Gaussian Mixture Model (GMM): probabilistic clustering, assuming each segment is a multivariate Gaussian distribution with full (unconstrained) covariance, fitted by Expectation-Maximization with 10 random initializations.

Agglomerative Hierarchical Clustering (Ward linkage) – hierarchical bottom-up merging that minimizes the increase in within-cluster variance at each merge step; a dendrogram was built on a representative sample of 300 customers for visualization, while final cluster assignments were computed on the full sample of 5,833 customers.

First, the number of clusters, k, was statistically evaluated over k = 2..9 using K-Means inertia (elbow method), silhouette score and GMM Bayesian Information Criterion (BIC). The silhouette-maximizing K-Means solution was at k = 2 (silhouette = 0.331), but this coarse two-group split (essentially high-value vs. low-value customers) offers limited managerial actionability for differentiated, personalized marketing. In the meantime, the BIC of GMM showed a monotonic decrease over the whole tested range (k = 2 to 9) without a clear elbow, a pattern consistent with the known behavior of likelihood-based criteria under full covariance parameterization of tending to favor more complex models as the sample size per component decreases, and was therefore deemed an unreliable sole criterion for selecting k in this context. Consistent with the retail marketing analytics RFM segmentation convention of targeting four to five actionable segments (e.g., Champions, At-Risk, New, Lost), k=4 was the business-informed, managerially interpretable number of clusters and used uniformly across all three algorithms, which enabled direct, fair comparison of the algorithm-level segment quality at a fixed, decision-relevant granularity. In Section 5 we discuss this statistical versus managerial trade-off further.

Cluster Validity Metrics

For k = 4, the partitions obtained by each algorithm were evaluated using three internal validity indices, namely the silhouette score (value between -1 and 1, the higher the better separated and more cohesive clusters), the Davies-Bouldin index (the lower the better separation) and the Calinski-Harabasz index (the higher the greater between-cluster dispersion relative to within-cluster dispersion). The agreement between the partitions of the three algorithms was quantified by the Adjusted Rand Index (ARI) which corrects for chance agreement between two partitions of the same data.

Explainability: Surrogate Modelling and SHAP

To explain the segmentation obtained, we trained a Random Forest classifier (300 trees, max depth = 8) as a surrogate model to predict the K-Means cluster label of each customer based on the seven RFM+ features. The fidelity

of the surrogate model to the original clustering was evaluated using held-out classification accuracy and macro-averaged F1-score. For the surrogate model, Shapley Additive exPlanations (SHAP) were then computed with a Monte Carlo permutation-sampling estimator of the exact Shapley value (Štrumbelj & Kononenko, 2010), the same method as in the authors' previous explainable-AI studies. For 200 randomly selected customers (50 per segment), SHAP values were calculated for the surrogate-predicted probability of that customer's assigned segment, using 15 random feature-order permutations and 10 background-customer draws per permutation, sampled from a pool of 300 background customers drawn from the full dataset. This provides a global feature-importance ranking per segment and local, customer-level explanations of segment assignment.

Results

The results of the simulation are discussed in the following sub-sections.

4.1 Cluster Number Selection

The K-Means elbow curve and the silhouette score curve for K-Means and GMM for k = 2 to 9 are shown in Figure 1 and the GMM BIC curve is shown in Figure 2. The silhouette score for K-Means steadily decreases from 0.331 at k=2 to 0.213 at k=9, and apart from the elbow at k=2, there is no obvious second elbow. The BIC for GMM decreases monotonically for k=2 to 9 and does not plateau. Taken together, these results provide the managerial motivation for selecting k=4 as described in Section 3.3.

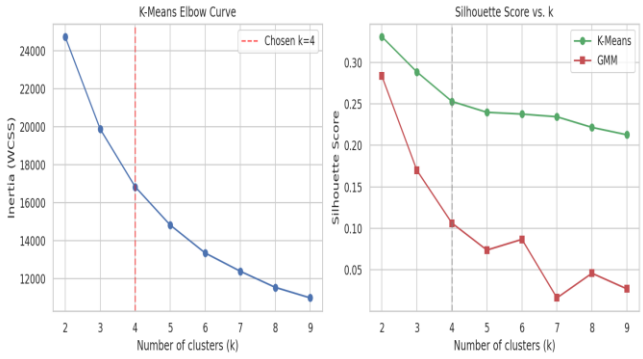


Figure 1: *K-Means elbow curve and silhouette score for K-Means and GMM*

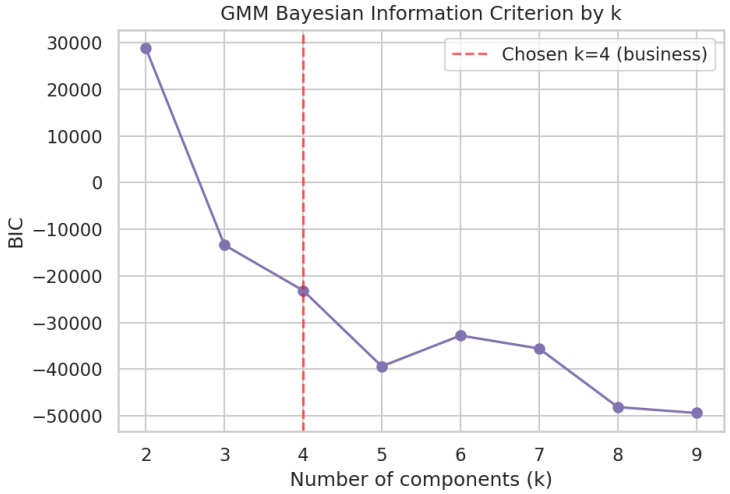


Figure 2: *Gaussian Mixture Model Bayesian Information Criterion (BIC)*

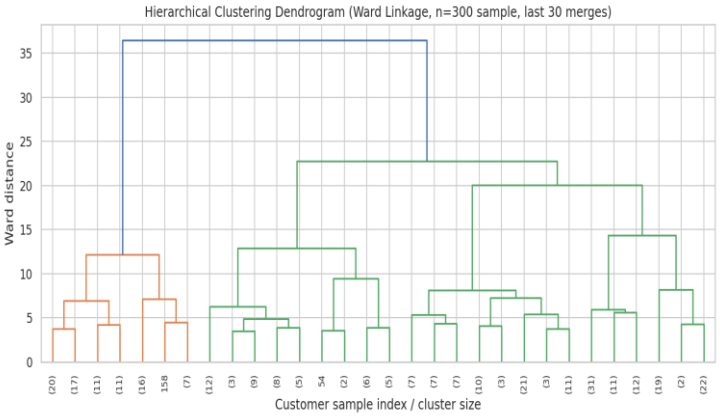


Figure 3: *Ward-linkage hierarchical clustering dendrogram*

Comparative Cluster Validity (RQ1)

Table 1 presents the three internal validity indices for K-Means, GMM and Hierarchical (Ward) clustering fitted with k = 4. K-Means provided the best overall separation with the highest silhouette score (0.253) and Calinski-Harabasz index (2,770.4), and a competitive Davies-Bouldin index (1.256). Hierarchical clustering was second best (silhouette = 0.209). GMM produced substantially weaker cluster separation (silhouette = 0.106, Davies-Bouldin = 2.226 – more than 75% higher, i.e. worse, than K-Means), consistent with GMM's

probabilistic, soft-boundary assignment producing more overlapping segments than the hard-boundary partitional and hierarchical methods on this dataset. Figure 4 shows the 2D PCA space (explaining 77.6% of total variance: 59.1% PC1, 18.5% PC2). As can be seen, the K-Means and the Hierarchical clustering produce geometrically compact, well separated groups while the GMM clusters have extensive spatial overlap. Table 2 shows the pairwise Adjusted Rand Index agreement between the three algorithms: K-Means and Hierarchical clustering have the strongest agreement (ARI = 0.569) while GMM is more divergent from both other methods (ARI = 0.325 vs. K-Means; 0.289 vs. Hierarchical), supporting the fact that GMM finds a qualitatively different partition structure rather than just a noisier version of the same segmentation.

Table 1: *Internal cluster validity metrics for K-Means, GMM, and Hierarchical clustering*

Method	k	Silhouette Score	Davies-Bouldin Index	Calinski-Harabasz Index
K-Means	4	0.253	1.256	2,770.4
Gaussian Mixture Model	4	0.106	2.226	1,610.1
Hierarchical (Ward)	4	0.209	1.334	2,396.6

Table 2: *Pairwise Adjusted Rand Index (ARI) agreement between the three clustering solutions.*

Comparison	Adjusted Rand Index
K-Means vs. GMM	0.325
K-Means vs. Hierarchical	0.569
GMM vs. Hierarchical	0.289

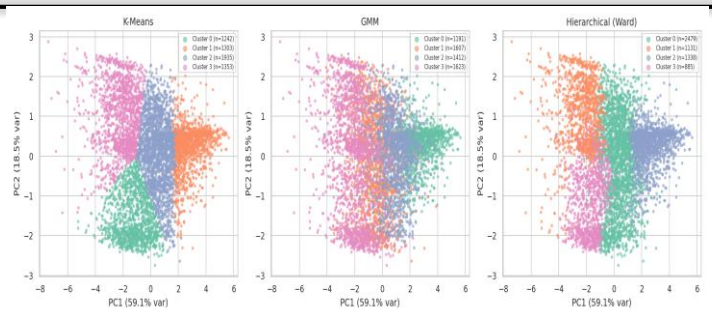


Figure 4: *PCA projection of the four-cluster solution for K-Means, GMM, and Hierarchical clustering.*

K-Means Segment Profiles

K-Means was chosen as the main segmentation technique for profiling and detailed explanation because of its better validity metrics. The mean RFM+ profile for each of the four derived segments is presented in Table 3 and Figure 5. Cluster 1 (22.3% of customers, n = 1,303) – labeled Champions – has by far the highest Monetary value (£6,864 mean), Frequency (15.0 invoices) and Total Items (4,335 units) of any segment, as well as strong recency (52.5 days) and the longest average tenure (629 days), representing the retailer's most valuable and loyal customers. Cluster 2 (33.2%, n = 1,935) — At-Risk — is moderately engaged (Monetary £1,427, Frequency 4.3) but has a higher recency (184.3 days since last purchase) compared to Champions, indicating a previously engaged group that is starting to disengage. Cluster 0 (21.3%, n=1,242) – New/Promising – is characterized by a significantly low tenure (149.2 days, compared to 532-629 days for the other segments) and reasonable recency, indicating customers that were newly acquired and have not yet had a chance to build a long purchase history. Cluster 3 (23.2%, n = 1,353) – Lost/Dormant – has the most extreme recency (481.4 days since last purchase), as well as the lowest monetary value (£305.9) and frequency (1.5 invoices), representing customers that have effectively churned.

Table 3: *K-Means segment profiles: mean Recency, Frequency, Monetary, and Tenure values*

Segment	n (%)	Recency (days)	Frequency	Monetary (£)	Tenure (days)
C0: New/Promising	1,242 (21.3 %)	85.0	1.9	514.7	149.2
C1: Champions	1,303 (22.3 %)	52.5	15.0	6,864.4	629.0
C2: At-Risk	1,935 (33.2 %)	184.3	4.3	1,426.5	532.1
C3: Lost/Dormant	1,353 (23.2 %)	481.4	1.5	305.9	535.9

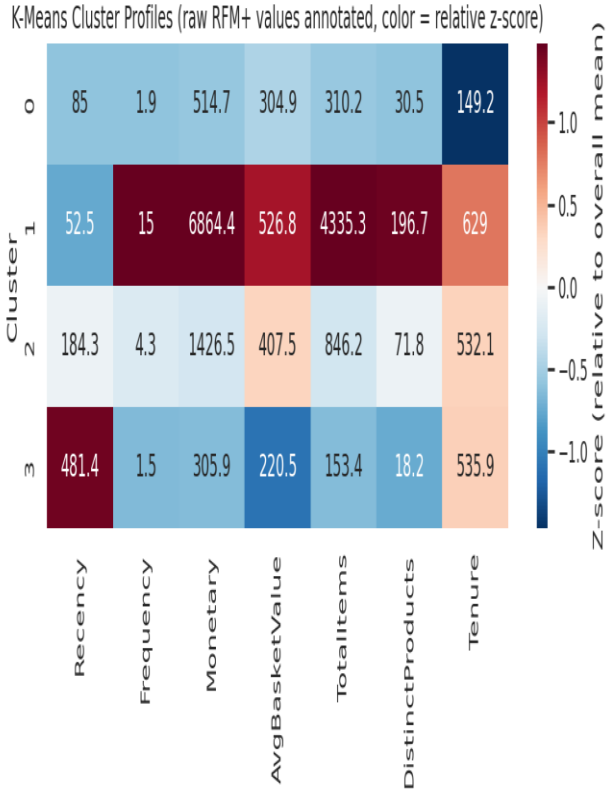


Figure 5: *K-Means cluster profile heatmap: raw RFM+ mean values (annotated) colored by z-score relative to the overall customer mean.*

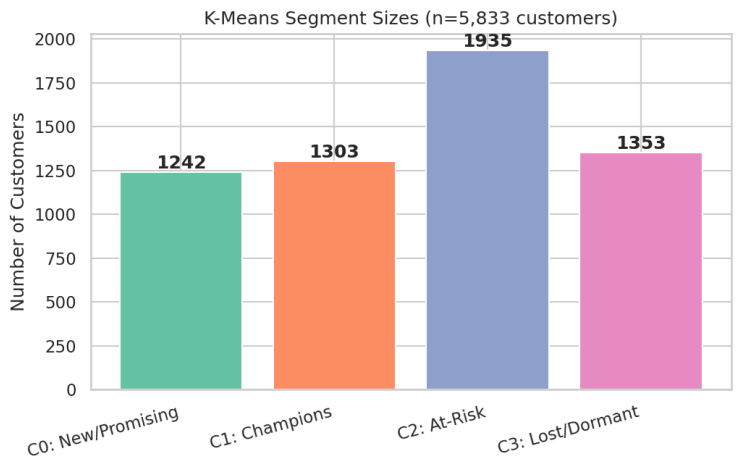


Figure 6: *K-Means segment sizes across the 5,833-customer*

Surrogate Model Fidelity and SHAP Explainability (RQ2, RQ3)

The Random Forest surrogate model was able to recreate the K-Means partition with 94.7 % held-out accuracy and a macro-averaged F1-score of 0.949. Thus, the four K-Means segments are, to a very high degree, recoverable as a supervised classification problem from the same seven RFM+ features - validating the use of this surrogate as a faithful basis for SHAP-based explanation of the clustering solution. The three most globally important features for reproducing cluster assignment overall, as ranked by the surrogate's native (Gini-based) feature importance, were Monetary (0.244), Tenure (0.201) and Total Items (0.184).

Figure 7 decomposes the SHAP feature importance by segment . This shows that the assignment to each segment is driven by a different combination of features , and not a single homogeneous ranking . Monetary value (mean |SHAP| = 0.244), Total Items (0.140) and Frequency (0.130) are the dominant drivers for the Champions segment – an intuitive result, as Champions are defined primarily by high spending volume and purchase frequency. Tenure dominates overwhelmingly for the New/Promising segment (mean |SHAP| = 0.339, more than double than any other feature for this segment), confirming that this segment is

identified primarily by its short customer relationship length rather than by spending behavior per se. The most important driver for the Lost/Dormant segment is Recency (mean |SHAP| = 0.226), followed by Monetary (0.191), which is consistent with the defining characteristic of this segment being a long period of inactivity in purchasing. At-Risk has a more distributed profile led by Monetary (0.179) and Total Items (0.093) reflecting the intermediate, transitional position of At-Risk between the extremes of Champions and Lost/Dormant.

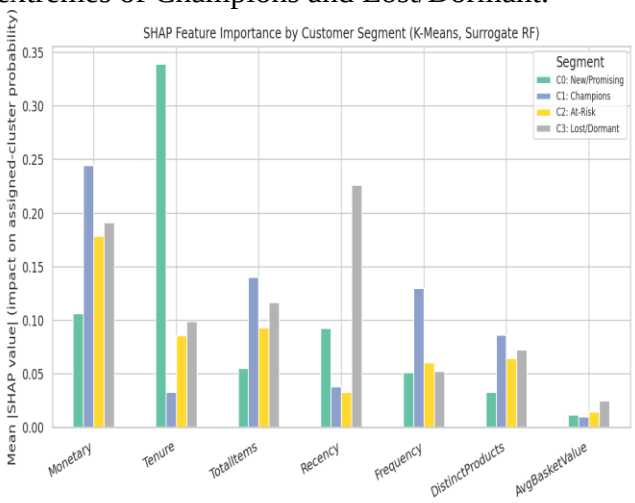


Figure 7: *SHAP feature importance by customer segment*

Figure 8: SHAP summary plot (beeswarm style) for the Champions segment only. High values (warm colored points) of Monetary, Total Items and Frequency consistently produce strongly positive SHAP contributions to Champions assignment, while low values of these same features (cool colored points) push the probability of assignment toward zero or negative—a directionally coherent and marketing-intuitive pattern that confirms that the surrogate model has learned a sensible, non-arbitrary decision boundary around the Champions segment rather than an opaque statistical artifact.

5. Discussion

Below is the discussion based on the simulation results according to the above research questions

Algorithm Choice Matters for Segment Quality (RQ1)

The finding that K-Means outperformed both GMM and Hierarchical clustering on internal validity metrics at a fixed, business-relevant k=4 has direct practical implications: for this dataset's RFM+ feature space, the roughly spherical, similarly sized cluster assumption underlying K-Means seems to be a better structural fit than GMM's more flexible but less constrained Gaussian-mixture assumption, which in this case led to excessive segment overlap rather than to capturing genuinely richer cluster shapes. This serves as a caution for practitioners who may select GMM for its probabilistic sophistication without ensuring that the added flexibility results in better-separated, more actionable segments on their own data.

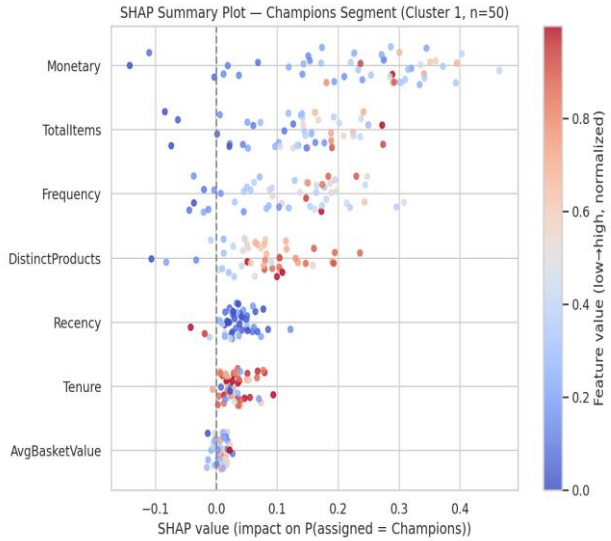


Figure 8: *SHAP summary plot for the Champions segment*

5.2 The Statistical-Managerial Trade-off in Choosing k

The difference between the statistically optimal k=2 (maximizing silhouette score) and the managerially selected k=4 is an example of a larger methodological consideration for applied customer-segmentation research: internal validity indices optimize a mathematical concept

of cluster compactness and separation that may not be aligned with the granularity that marketing teams need to design differentiated interventions. While a two-segment solution is statistically 'best' in terms of silhouette score, it would obfuscate the important behavioral distinction between those newly acquired customers (New/Promising) and the former valuable customers who are now disengaging (At-Risk), a distinction with directly opposing marketing implications (onboarding and nurture campaigns vs win-back and reactivation campaigns respectively). Rather than selecting the k that best supports an a priori narrative without comment, the approach in this study of transparently reporting the statistical optimum while justifying a managerially selected alternative provides a template other applied segmentation studies can follow.

Surrogate-Based Explainability as a General Framework for Unsupervised Learning (RQ2, RQ3)

The high surrogate fidelity (94.7% accuracy) achieved in this study shows that supervised explainability techniques such as SHAP can be successfully extended to unsupervised clustering problems via a faithful surrogate model, without the need for the clustering algorithm itself to provide any native explainability mechanism. This finding is not limited to the specific K-Means solution analyzed here; any clustering algorithm, even non-parametric or density-based approaches lacking a natural centroid or easily interpretable probability structure, could in principle be explained using the same surrogate-and-SHAP pipeline, so long as fidelity is confirmed and reported. The segment-specific SHAP signatures derived (Tenure for New/Promising, Recency for Lost/Dormant, Monetary/Frequency for Champions) align well with standard marketing constructs (onboarding stage, churn risk, customer value) and directly enable feature-targeted personalization: e.g., marketing automation could trigger tenure-based onboarding nudges specifically for customers whose SHAP profile shows Tenure as the dominant driver of their New/Promising assignment, instead of applying a generic new-customer campaign uninformed by the specific behavioral driver.

Theoretical Contribution

This study contributes to the customer-segmentation and explainable-AI literatures by demonstrating, within a single empirical pipeline, that (a) classical unsupervised clustering algorithms differ meaningfully in cluster quality even at a fixed, business-chosen k, and (b) surrogate-model SHAP explanation can bridge the interpretability gap inherent to unsupervised learning with high fidelity, producing segment explanations that are both statistically grounded and directly actionable for individualized marketing decision-making.

Conclusion

This work developed and empirically validated an explainable AI framework for customer segmentation, comparing K-Means, Gaussian Mixture Models and Ward-linkage hierarchical clustering on RFM+ behavioral features derived from 5,833 online retail customers and rendering the resulting segmentation explainable via a high-fidelity Random Forest surrogate model and SHAP-based feature attribution. K-Means provided the most well-separated, actionable four-segment solution and SHAP analysis identified theoretically coherent, segment-specific behavioral drivers—Tenure for newly acquired customers, Monetary value and Frequency for Champions, and Recency for dormant, likely-churned customers—each directly mappable to a differentiated marketing intervention. This framework marries classical unsupervised segmentation with rigorous, quantitative explainability, turning customer segmentation from an opaque labeling exercise into a transparent, auditable decision-support tool that directly supports a personalized marketing mandate of understanding not just who a customer is, but why.

References

Bangash, J. I., Abdullah, A., & Khan, A. (2014). Issues and challenges in localization of wireless sensor networks. *Sci. Int (Lahore)*, 595-603.

Batool, S., Rahman, S., & Akhtar, N. (2016). Relationship of management-rated high performance work system and competitive advantage: Mediating role of collective human capital. *Abasyn University Journal of Social Sciences*, *9*(1), 176-188.

Chen, D., Sain, S. L., & Guo, K. (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., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-1*(2), 224–227.

Fraley, C., & Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association, 97*(458), 611–631.

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification, 2*(1), 193–218.

Imran, M., & Khattak, S. R. (2019). The buffering effect of supervisor support on the relationship between work engagement and behavioral outcomes. *Journal of Managerial Sciences*, *13*(1), 59-68.

Iqbal, N., Khattak, S. R., & Khattak, M. A. (2013). Relationship between macroeconomic variables and KSE-100 index: evidence from Pakistan. *Knowledge Horizons. Economics*, *5*(4), 101.

Kansal, T., Bahuguna, S., Singh, V., & Choudhury, T. (2018). Customer segmentation using K-means clustering. In *2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS)* (pp. 135–139). IEEE.

Khalil, T., Bangash, J. I., & Adnan, A. (2016, August). Low level visio-temporal features for violence detection in cartoon videos. In *2016 sixth international conference on innovative computing technology (INTECH)* (pp. 320-325). IEEE.

Khan, T. I., Saeed, I., & Khattak, S. R. (2018). Impact of time pressure on organizational citizenship behavior: Moderating role of conscientiousness. *Global Social Sciences Review*, *3*(3), 317-331.

Khattak, S. R., Zada, M., Nouman, M., Rahman, S. U., Fayaz, M., Ullah, R., ... & Contreras-Barraza, N. (2022). Investigating inclusive leadership and pro-social rule breaking in hospitality industry: Important role of psychological safety and leadership identification. *International Journal of Environmental Research and Public Health*, *19*(14), 8291.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in neural information processing systems* (Vol. 30, pp. 4765–4774). Curran Associates, Inc.

Murtagh, F., & Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: Which algorithms implement Ward's criterion? *Journal of Classification, 31*(3), 274–295.

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.

Štrumbelj, E., & Kononenko, I. (2010). An efficient explanation of individual classifications using game theory. *Journal of Machine Learning Research, 11*(1), 1–18.

Tariq, M., & Khattak, S. R. (2019). Practices of capital budgeting techniques: evidence from the corporate sector of Pakistan. *NUML International Journal of Business & Management, 14*(1), 16-28.

Ullah, R., Khattak, S. R., Khan, R., & Sana, A. (2019). Enhancing Knowledge Management through Effective Leadership Styles: Evidence from Pakistan. *Dialogue (Pakistan), 14*(3), 23.

Zaman, K., Bangash, J. I., Maghdid, S. S., Hassan, S., Afridi, H., & Zohaib, M. (2020, October). Analysis and classification of skin cancer images using convolutional neural network approach. In *2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)* (pp. 1-8). IEEE.

Zhang, Y., Bradlow, E. T., & Small, D. S. (2015). Predicting customer value using clumpiness: From RFM to RFMC. *Marketing Science, 34*(2), 195–208