ABCDEats Inc.

Data Mining Project
Master in Data Science and Advanced Analytics

NOVA Information Management School
Universidade Nova de Lisboa

Group 25
João Dargent Figueiredo 20240853
Jorge Miguel Cordeiro 20240594
Tomás Manuel Estrócio e Silva 20230982
Virgílio Afonso Ferreira 20240689

Fall Semester 2024-2025

Contents

1 Introduction **2**

2 Data Preprocessing **2**
- 2.1 Missing Value Treatment . . . 2
- 2.2 Outlier Detection and Handling . . . 2
- 2.3 Data Scaling . . . 3

3 Feature Selection **3**

4 Segmentation **3**

5 Clustering Models **4**
- 5.1 Hierarchical + Partition Methods . . . 4
- 5.2 Self-Organizing Maps . . . 4
- 5.3 Density-Based Clustering . . . 5
- 5.4 Final Clusters for Each Segmentation . . . 5

6 Clustering Analysis **6**
- 6.1 Merging the Perspectives . . . 6
- 6.2 Cluster visualization using t-SNE . . . 6
- 6.3 Profiling based on numerical features . . . 6
- 6.4 Profiling based on categorical features . . . 6

7 Feature Importance **7**

8 Business Applications and Marketing Approaches **7**

9 Conclusion **8**
- 9.1 Opportunities for Improvement . . . 8
- 9.2 Future Recommendations . . . 9
- 9.3 Closing Remarks . . . 9

10 References **10**

A Appendix **11**

1 Introduction

In this project, we focus on customer segmentation for ABCDEats Inc. This means dividing customers into groups based on their buying habits and personal characteristics.

First, we cleaned and prepared the data to make it ready for analysis, which helps to make the clustering process more accurate. We then used different methods to create clusters, including k-means, hierarchical clustering, and DBSCAN. We compared the results to find the best way to group the customers. After creating the clusters, we analyzed them to understand the differences between the groups. We looked at features like how much money customers spend and how often they order, as well as their preferences for certain types of food or locations.

Finally, we used this information to suggest business strategies and marketing ideas for each group of customers.

Note: EDA was delivered previously.

2 Data Preprocessing

This section will be crucial to our project, as this is where we will prepare and transform the 'raw' data that was given to us into a clean, structured, and usable format for our clustering analysis.

2.1 Missing Value Treatment

The first thing we want to do is to treat the missing values. To handle missing values, we imputed them by grouping the data based on a categorical feature and then calculating the mode for each subcategory. For cases where there is no significant difference between subcategories, we simply computed the overall mode, represented by '-' in the table. Additionally, for categorical features, we directly computed the mode without grouping by a categorical feature. Table 1 summarizes the features treated for missing values and the categorical features used to compute the mode.

2.2 Outlier Detection and Handling

To identify and handle outliers, we used the DBSCAN algorithm, which is effective in distinguishing dense data clusters from noise. DBSCAN was chosen because it does not assume a predefined distribution and can maintain meaningful clusters, even for groups that exhibit extreme behavior but form coherent patterns.

By tuning the parameters `eps = 1.5` and `min_samples = 5`, we ensured the method was sensitive enough to detect outliers that were replaced with `NaN`, and rows with missing values were subsequently removed. The performance of this method was evaluated by comparing the differences in the boxplots before and after the application of the DBSCAN algorithm for outlier detection, as shown in Figure 1.

This approach helped us to deal with extreme values without losing important data from unusual but meaningful groups, keeping our dataset accurate and useful for analysis.

2.3 Data Scaling

From our EDA, most of the transformed features follow a normal distribution, thus, to standardize the numerical features and ensure they are on comparable scales, we applied the `StandardScaler`.

3 Feature Selection

To assess feature redundancy and identify those most relevant to our market segmentation analysis, we performed feature selection using a correlation matrix. The correlation matrix plot allowed us to evaluate the relationships between all numerical features, highlighting pairs with a high degree of multicollinearity. Features with a correlation coefficient greater than 0.8 or less than -0.8 were removed, as illustrated in Figure 2. Additionally, intermediary features—those created primarily to derive or construct other variables—were also eliminated to streamline the analysis.

Following this initial reduction, we recalculated the correlation matrix to refine our understanding of the remaining features. This second iteration clarified additional redundancies. For instance, since *Recency* was retained as a key feature, we decided to drop *first_order* due to its overlap. Similarly, we eliminated *is_chain* in favor of *chain_percentage*, which provides a more granular representation of the same concept.

This iterative approach ensured that our final set of features was both parsimonious and robust, minimizing redundancy while retaining the most informative variables for analysis as shown on our final correlation matrix Figure 3

4 Segmentation

To enhance the performance of clustering methods, the dataset was segmented into two groups as shown in Table 2. This segmentation was based on the logical grouping of variables into categories that represent distinct aspects of customer behavior. In an ideal scenario with unlimited resources and time, we would have tested multiple feature combinations. Some examples of other potential combinations are shown in Table 3.

5 Clustering Models

For each segmentation, four distinct clustering methods were applied, ensuring the same techniques were consistently used for both groups. The rationale for the chosen clustering methods is summarized in Table 4, while Table 5 explains why certain standalone methods or alternative models were not selected.

5.1 Hierarchical + Partition Methods

For both segmentations, the initial step involved evaluating which clustering method achieved the highest R^2 value across various numbers of clusters. To achieve this, R^2 values were calculated and compared between K-means (initialized without specifying the number of clusters) and agglomerative clustering using various linkage methods. In the case of the demographic segmentation, K-means outperformed the other methods, achieving the highest R^2 value, followed closely by the Ward linkage method. Additionally, the R^2 graph indicated that the optimal number of clusters for this segmentation lies between 3 and 4 (Fig. 4). To confirm which was the optimal number of clusters, the silhouette score was also applied. It shows a declining trend as the number of clusters increases. Beyond 3 or 4 clusters, the clusters may become too small or fragmented, leading to lower cohesion within clusters and less meaningful separation between them. This often results in a lower silhouette score. Despite the 4 clusters having a greater score, the difference in silhouette score between 3 and 4 clusters is not significant. Fewer clusters are often preferable for ease of application and decision-making. Therefore, we end up to choose 3 clusters (Fig. 5). For the purchase behavior segmentation, the R^2 graph showed once again that the optimal number of clusters lies between 3 or 4 clusters (Fig. 6). The silhouette score showed the highest value for 4 clusters. However, the difference for 3 clusters is 0.005, which is practically 0. Once again, we are going to choose 3 clusters because fewer clusters are often preferable (Fig. 7).

5.2 Self-Organizing Maps

Hierarchical and K-means clustering were applied on top of the SOM model. As shown in Table 6, the 10x10 grid was chosen for both segmentations to balance granularity, interpretability, and computational efficiency.

For the K-means method in the demographic segmentation (Fig. 8), the "elbow point" appears at 4 clusters, where the decrease in inertia slows significantly, indicating diminishing returns from adding more clusters. For the purchase behavior segmentation (Fig. 9), 3 clusters were deemed optimal for similar reasons.

For the Hierarchical, the initial R2 graph confirmed that ward method linkage is the most efficient when compared to the others, it was the one used. Then, the dendrogram was plotted a threshold of 90 was defined. The threshold (red line) intersects just below a noticeable "gap" in the dendrogram. Above this threshold, the vertical distances between clusters are much larger, meaning clusters are more distinct. Therefore, the chosen number of clusters to use was 6, for both segmentations (Fig. 10 and Fig. 11).

5.3 Density-Based Clustering

The DBSCAN algorithm was chosen as the density-based clustering method. The first step involved evaluating the `eps` parameter by plotting the k-distance graph. For the demographic segmentation, the graph indicated an `eps` value of 1.0 (Fig. 12), while for the purchase behavior segmentation, `eps` was approximately 0.30 (Fig. 13). Next, a function was used to evaluate various `min_samples` values while keeping the selected `eps` fixed. This provided insight into the variation in the number of clusters and their respective noise levels.

For the demographic segmentation, the optimal `min_samples` range was between 3 and 5. To determine the best value, the silhouette score was calculated for each case.The results showed that `min_samples=5` achieved the highest silhouette score and the lowest number of clusters (often preferable). Although this setting resulted in more noise, it was a tradeoff deemed acceptable. We ended up with 3 clusters (Table 7).

For purchase segmentation the results were not very satisfactory, as the number of clusters remained relatively high across all variations of `min_samples` (Table 8). We ultimately chose `min_samples = 4`, as it provided the lowest amount of noise for the smallest number of clusters that were feasible to select.

5.4 Final Clusters for Each Segmentation

The final graphs of each clustering method for both segmentations can be seen in Fig. 14 and Fig. 15.

As detailed in the following analysis (Table 9 and Table 10), the decision to use 3 clusters for each segmentation was based on the evaluation of the clustering methods used.

For demographic segmentation, two methods indicated 3 clusters as optimal. Despite DBSCAN not performing well and SOM + K-means suggesting 4 clusters, as mentioned throughout this report, when deciding between 3 and 4 clusters, 3 clusters are often preferable due to their simplicity and interpretability.

For purchase behavior segmentation, the choice of 3 clusters was clear from the analysis.

The clustering method chosen was **K-means** for both segmentations, owing to its simplicity, efficiency, and ability to produce well-separated and interpretable clusters.

6 Clustering Analysis

6.1 Merging the Perspectives

The segmentations were merged using Hierarchical Clustering. The final result was 3 merged clusters (Fig. 16). The final segmentations with merged labels are displayed in Fig. 17.

6.2 Cluster visualization using t-SNE

As we can see in Fig. 18, the t-SNE visualization indicates that the final labels produce well-separated and cohesive clusters, suggesting that the chosen clustering method effectively captured distinct patterns in the data. While some clusters vary in size, the overall separation and density of points within clusters demonstrate strong intra-cluster cohesion and inter-cluster separation, validating the meaningfulness of the segmentation.

6.3 Profiling based on numerical features

The features **Order Rate per Week**, **Amount Spent per Week**, **Average Product Price**, and **Chain Percentage** were specifically chosen for numerical profiling because they are the features that exhibit more variability across the clusters, which makes them suitable for distinguishing between customer segments. However, we have in mind that these differences could be better (Fig 19 and Fig 20). If the differences were more pronounced, we could make a better profiling. The profiling for each cluster (**Cluster 0**, **Cluster 1**, and **Cluster 2**) are presented in Tables 11, 12, and 13, respectively.

6.4 Profiling based on categorical features

We considered these features to be the most relevant for our analysis. Figure 21 highlights the distribution of cities across clusters, emphasizing the geographical representation within each segment. Similarly, Figure 22 illustrates the cuisine preferences across clusters, showcasing the distinct food preferences that differentiate the groups.

- **Cluster 0:** Has a higher proportion of customers from **City_1**. Prefers **Asian** food, but **American** also has a good proportion.
- **Cluster 1:** Balanced proportions between **City_1** and **City_2**, but clearly dominated by **City_3**. Shows an equal preference for **American** and **Asian**, with additional interest in other cuisines.
- **Cluster 2: City_1** and **City_2** clearly dominate, with no presence in **City_3**. Shows a clear preference for **Asian**, but also has a notable preference for **Italian** and **Street Food / Snacks**.

7 Feature Importance

Before conducting the marketing analysis, attention was directed toward evaluating the contribution of each feature to the clustering process. The proportion of the total sum of squares (SS) explained by the clusters was calculated using the R^2 metric. The results indicated that `chain_percentage`, `log_order_rate_per_week`, and `log_amount_spent_per_week` exhibited the highest R^2 values, highlighting their importance in differentiating the clusters (Table 14).

To further assess feature significance, a Decision Tree model with a maximum depth of 3 was trained, using the final merged labels as the target variable. The primary objective was to classify the data points effectively and achieve high purity within the leaves. The data was initially divided into training and testing sets using the `train_test_split` function and subsequently fitted to a `DecisionTreeClassifier`. On average, the model correctly predicted 92.15% of the data points. The Decision Tree visualization provided additional insights into the relationships between the features and clusters. From the analysis, `chain_percentage` and `log_order_rate_per_week` emerged as the most influential features in distinguishing between clusters (Fig. 23).

8 Business Applications and Marketing Approaches

As a result of what was discussed in the profiling with numerical and categorical variables sections, we suggest the following business applications and marketing approaches. These recommendations are detailed in Tables 15, 16, and 17, corresponding to Cluster 0, Cluster 1, and Cluster 2, respectively.

While these suggestions provide actionable strategies for improving customer satisfaction and engagement, they could be further enhanced with a comprehensive marketing plan tailored to each cluster. Such a plan could include detailed campaign timelines, specific communication channels, and budget allocations for targeted efforts.

9 Conclusion

The primary objective of this project was to assist ABCDEatsInc. in gaining insights into their customers' behaviour and preferences by segmenting them into clusters and designing tailored marketing approaches. These strategies aimed to not only increase customer satisfaction and engagement but also retain existing clients and attract new ones by enhancing the services offered.

However, the development of a robust marketing strategy was hindered by the results of the clustering process. The clusters exhibited minimal differences, with most features showing similar patterns across groups. This lack of distinction limited our ability to define highly specific customer segments, making the proposed strategies less impactful.

Through the report, we have detailed every decision made during the project and provided explanations for the chosen approaches. Additionally, we highlighted alternative methods and reasons for not pursuing other clustering models. Despite the limitations, this process underscored the complexity of customer segmentation and the importance of carefully balancing feature selection, preprocessing techniques, and clustering methods.

9.1 Opportunities for Improvement

If time and resources were not constrained, several enhancements could have been explored to improve the clustering results and segmentation strategy:

- Testing segmentation perspectives beyond the ones employed.
- Experimenting with alternative clustering algorithms.
- Optimizing the number of clusters.
- Identifying bottlenecks in the clustering pipeline by:
  - Evaluating the impact of feature selection on clustering outcomes.
  - Revisiting the preprocessing pipeline to optimize feature scaling and transformations.

9.2 Future Recommendations

To further enhance customer segmentation for ABCDEatsInc., we propose the following actions:

- **Incorporate External Data:** Enrich the dataset with external factors like customer feedback, competitor pricing, or socioeconomic variables.
- **Leverage Time-Series Analysis:** Analyze temporal customer behaviour to reveal seasonality and trends that could further refine clusters.
- **A/B Testing:** Implement and evaluate small-scale experiments to validate the current segmentation strategy and improve its effectiveness.
- **Customer Feedback:** Conduct surveys or interviews to gather qualitative insights that complement quantitative data.

9.3 Closing Remarks

While this project presented challenges, it also provided valuable insights into ABCDEatsInc.'s customer base and highlighted areas for improvement in segmentation processes. Moving forward, refining the clustering methods and incorporating the proposed recommendations could lead to a more actionable segmentation model. This, in turn, will enable ABCDEatsInc. to craft precise, data-driven marketing strategies to better serve their customers and achieve their business objectives.

10 References

References

[1] Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques.* Morgan Kaufmann.

[2] Kuhn, M., & Johnson, K. (2019). *Feature Engineering and Selection: A Practical Approach for Predictive Models.* CRC Press.

[3] Aggarwal, C. C. (2014). *Outlier Analysis* (2nd ed.). Springer.

[4] Bishop, C. M. (2006). *Pattern Recognition and Machine Learning.* Springer.

[5] MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability.* University of California Press.

[6] Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining.*

[7] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer.

[8] Kotler, P., Keller, K. L., & Chernev, A. (2016). *Marketing Management* (15th ed.). Pearson.

[9] Van Der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research, 9.*

A Appendix

Feature with Missing Values	**Categorical Feature Used**
order_rate	cuisine_preference
order_rate_per_week	cuisine_preference
log_order_rate_per_week	cuisine_preference
amount_spent_per_week	customer_region
log_amount_spent_per_week	customer_region
average_product_price	customer_region
chain_percentage	customer_region
customer_age	-
days_between_orders	-
Cities	-
first_order	-

Table 1: Features and the corresponding categorical features used to impute missing values.

Subset	**Features**
`Demographics Preferences`	`{customer_age, Recency, log_order_rate_per_week, log_amount_spent_per_week}`
`Purchase Behavior`	`{average_product_price, chain_percentage, log_vendor_count}`

Table 2: Feature segmentation for clustering

Combination	Features Included
1	{customer_age, chain_percentage, log_vendor_count}
2	{Recency, average_product_price, log_order_rate_per_week}
3	{log_amount_spent_per_week, log_vendor_count, customer_age}
4	{chain_percentage, Recency, log_order_rate_per_week}
5	{customer_age, average_product_price, log_amount_spent_per_week}
6	{log_vendor_count, average_product_price, log_order_rate_per_week}

Table 3: Examples of other possible feature combinations for clustering

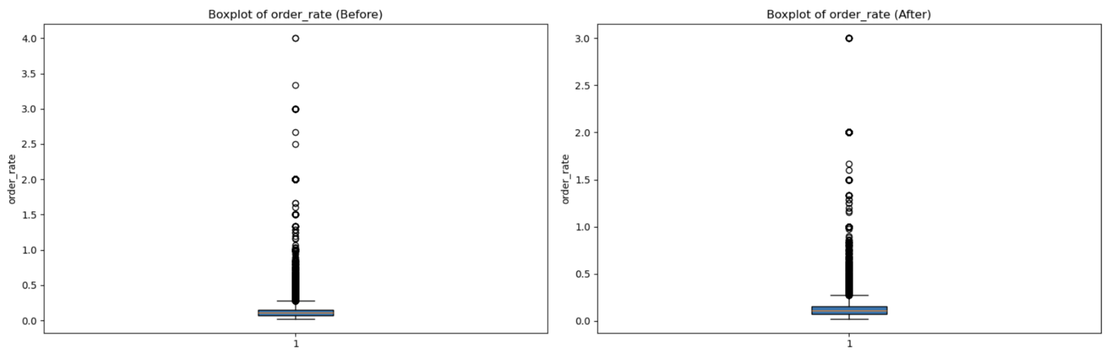

Figure 1: Boxplots of `order_rate` before and after the application of the outlier detection method.

Method	**Reason for Choosing**
Hierarchical + K-means	Combining Hierarchical clustering for initial structure with K-means for final refinement captures both global structure and fine-grained clusters. Hierarchical methods are computationally expensive on larger datasets, so K-means helps finalize clusters efficiently.
SOM + K-means	SOM (Self-Organizing Maps) reduces the dimensionality of high-dimensional data and organizes it into a grid structure. K-means is applied afterward to extract clusters from the structured grid. This approach benefits from SOM's ability to preserve data topology and K-means' simplicity for clustering.
SOM + Hierarchical	SOM's grid structure provides a well-organized representation of data, and Hierarchical clustering can capture more nuanced relationships in smaller datasets. This method allows for dendrogram-based analysis combined with SOM's topology-preserving properties.
DBSCAN	DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is robust for detecting clusters of arbitrary shapes and handling noise. It was chosen to identify outliers or small, dense groups that traditional methods like K-means might miss.

Table 4: Chosen clustering methods and their rationale

Method	Reason for Not Choosing
Standalone K-means	Not used alone because it struggles with non-spherical clusters, sensitivity to initialization, and an inability to detect noise or outliers. Combining K-means with other methods addresses these limitations.
Standalone Hierarchical	Not used alone due to its high computational cost and limited scalability for large datasets. Additionally, it lacks flexibility in adjusting cluster boundaries compared to combining it with K-means or SOM.
Standalone SOM	Not used alone because while SOM provides a grid-based representation of the data, it does not explicitly assign cluster memberships. It works best as a preprocessing step combined with another clustering algorithm like K-means or Hierarchical.
Other Models (e.g., Gaussian Mixture Models)	Not used because they assume Gaussian distribution, which may not suit the dataset. Additionally, models like GMM can struggle with high-dimensional data and are computationally more intensive.

Table 5: Excluded clustering methods and their rationale

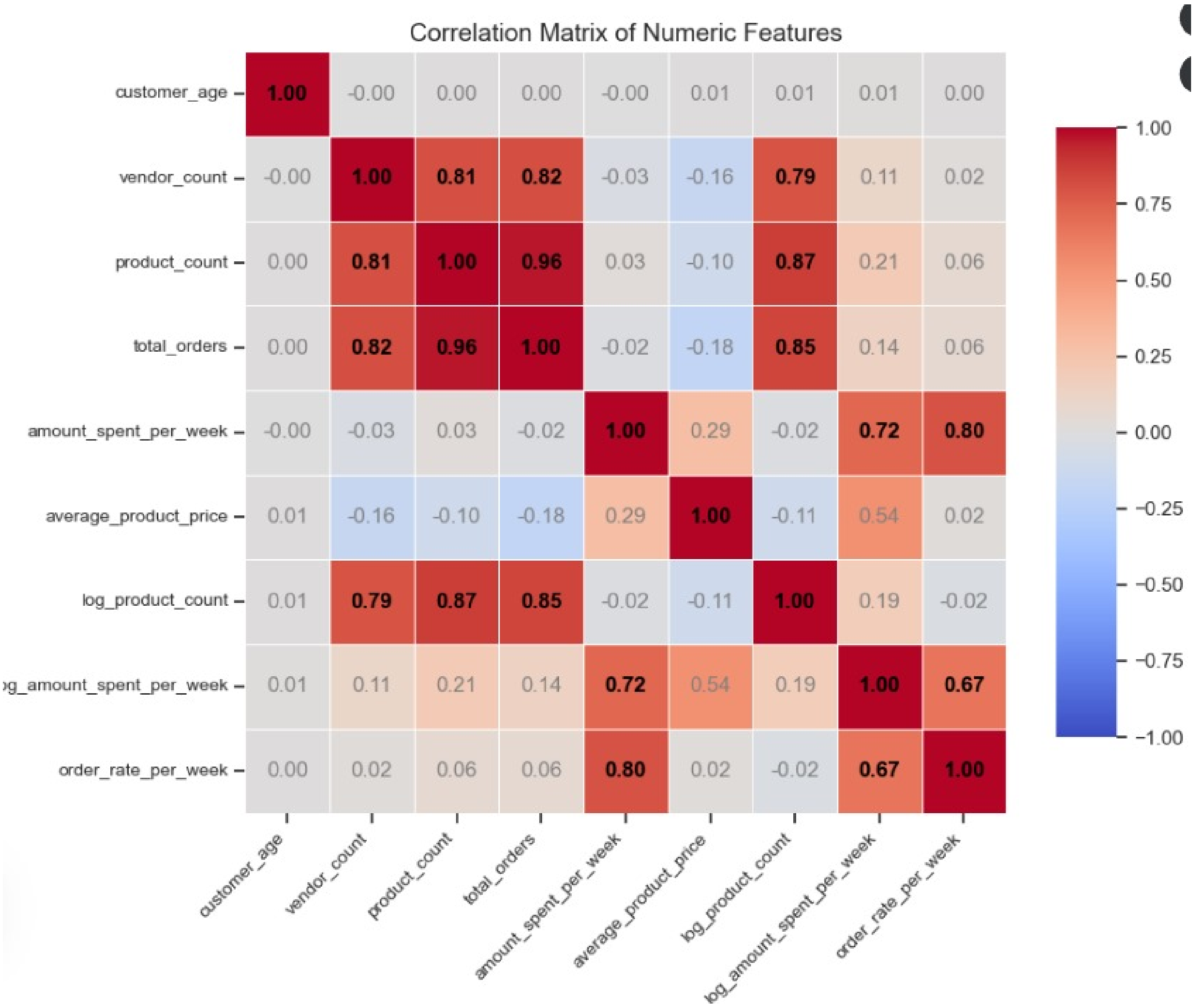

Figure 2: Correlation Matrix Before Removing Correlated Features

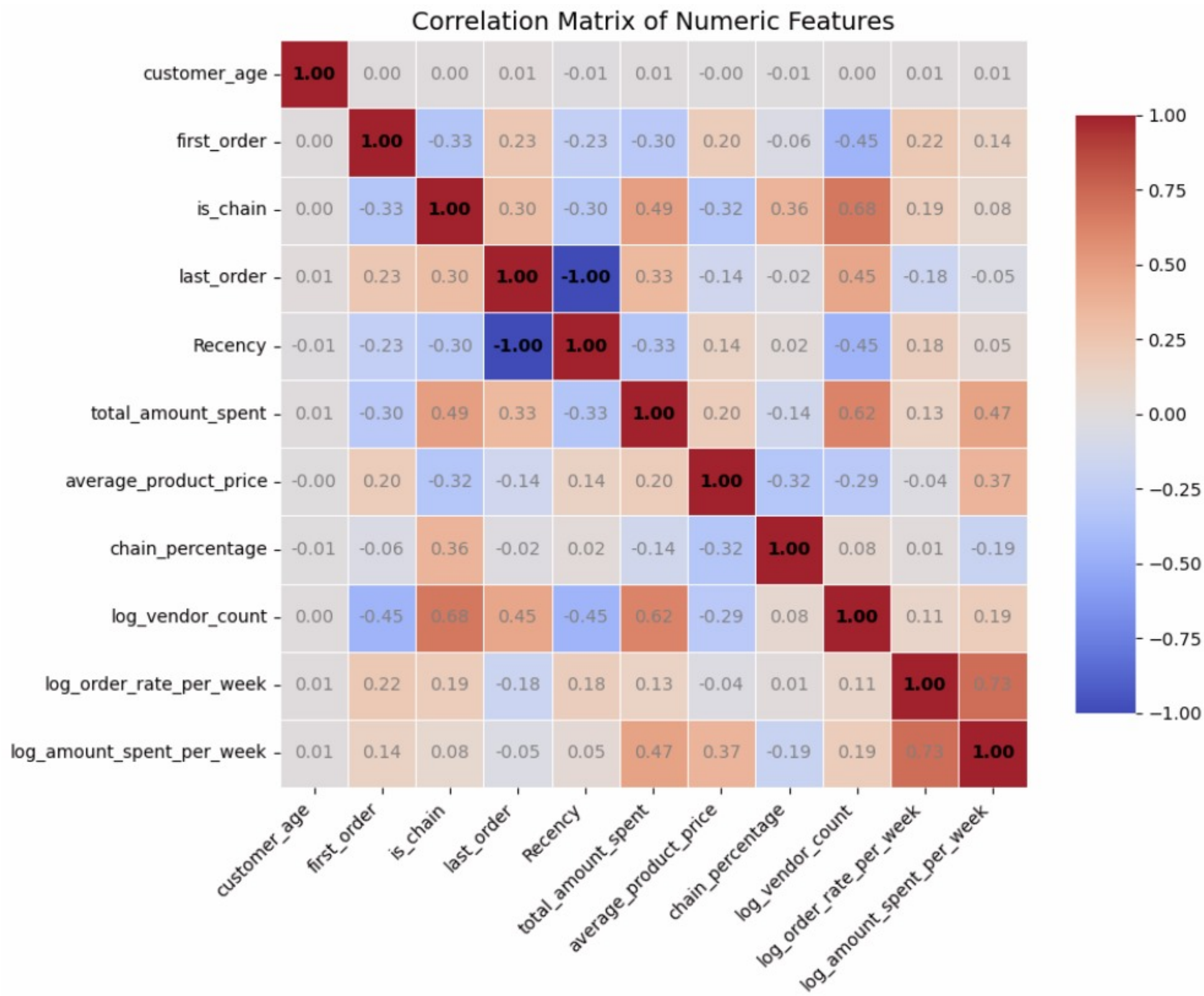

Figure 3: Correlation Matrix After Removing Correlated Features

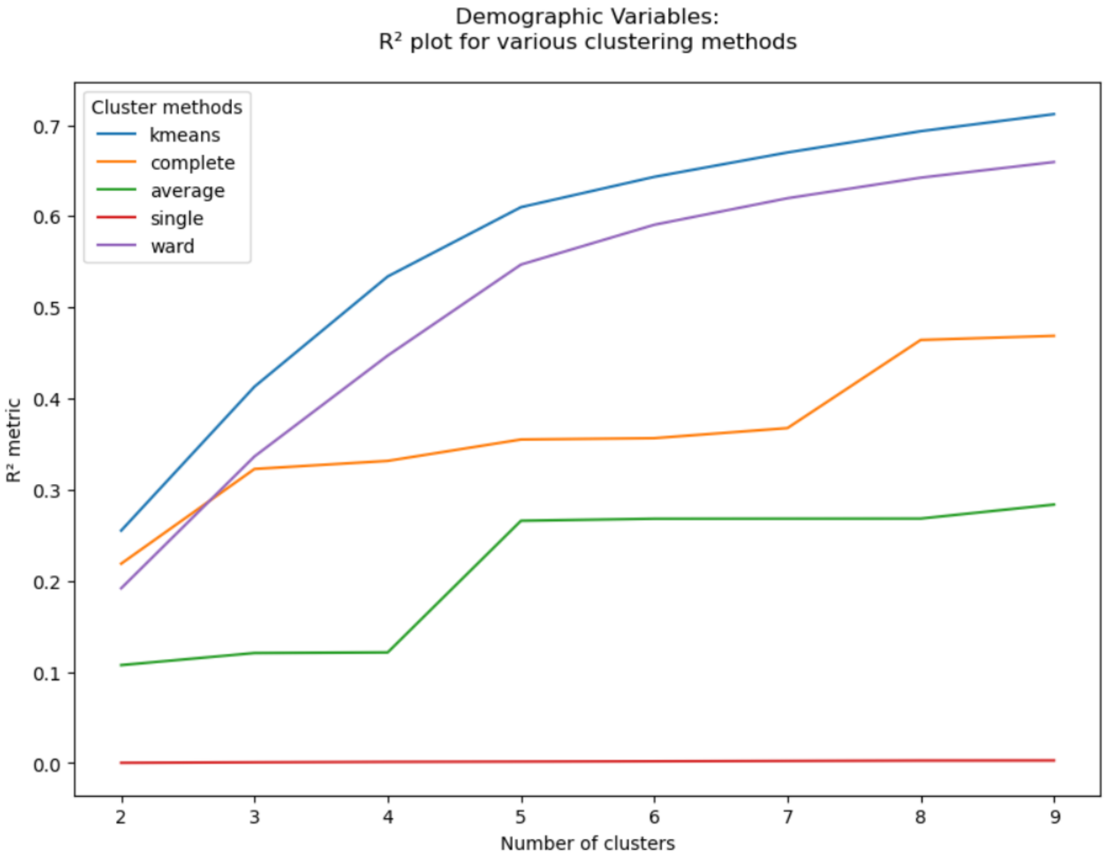

Figure 4: R^2 plot for the demographic segmentation.

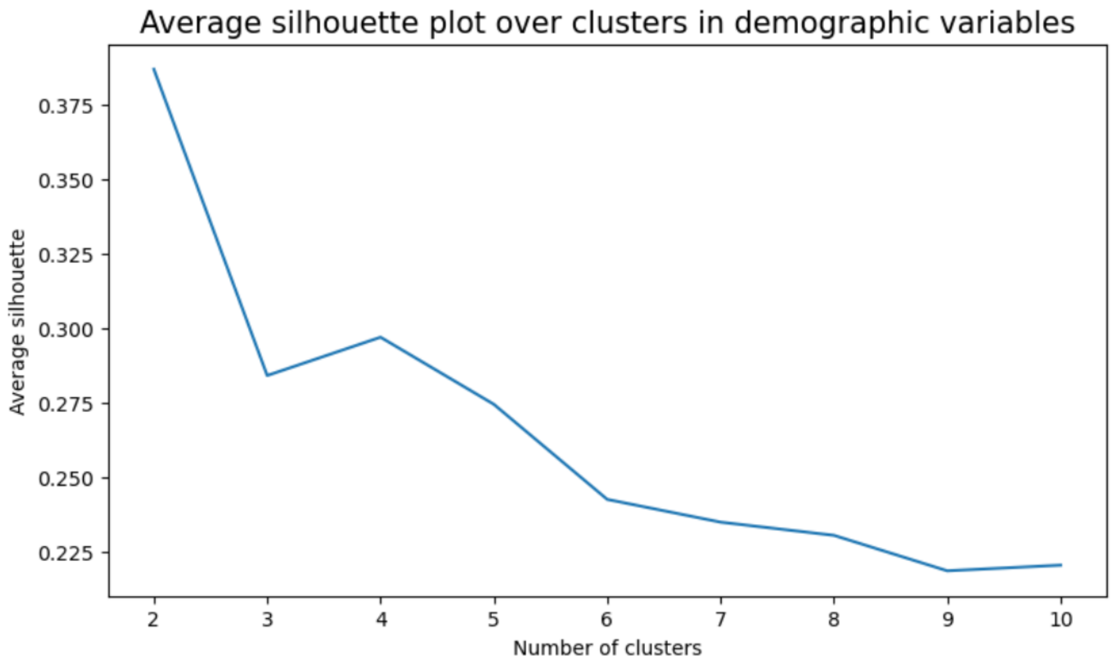

Figure 5: Silhouette score for demographic segmentation.

Reason	Explanation
Granularity	A 10x10 grid provides 100 nodes, offering sufficient resolution to capture subtle patterns in the data while avoiding overfitting or oversimplification.
Interpretability	The grid size balances detail and simplicity, making it easier to visualize and interpret the clusters compared to larger or smaller grids.
Dataset Complexity	The size of the grid is appropriate for the complexity and dimensionality of the dataset, enabling effective representation of the data.
Post-Processing Flexibility	A 10x10 grid creates a structured lower-dimensional representation that can be further clustered using k-means or other methods, enhancing its utility for analysis.
Computational Efficiency	The grid size is computationally manageable, ensuring efficient training while maintaining adequate detail for meaningful clustering.

Table 6: Reasons for choosing a 10x10 grid for SOM training

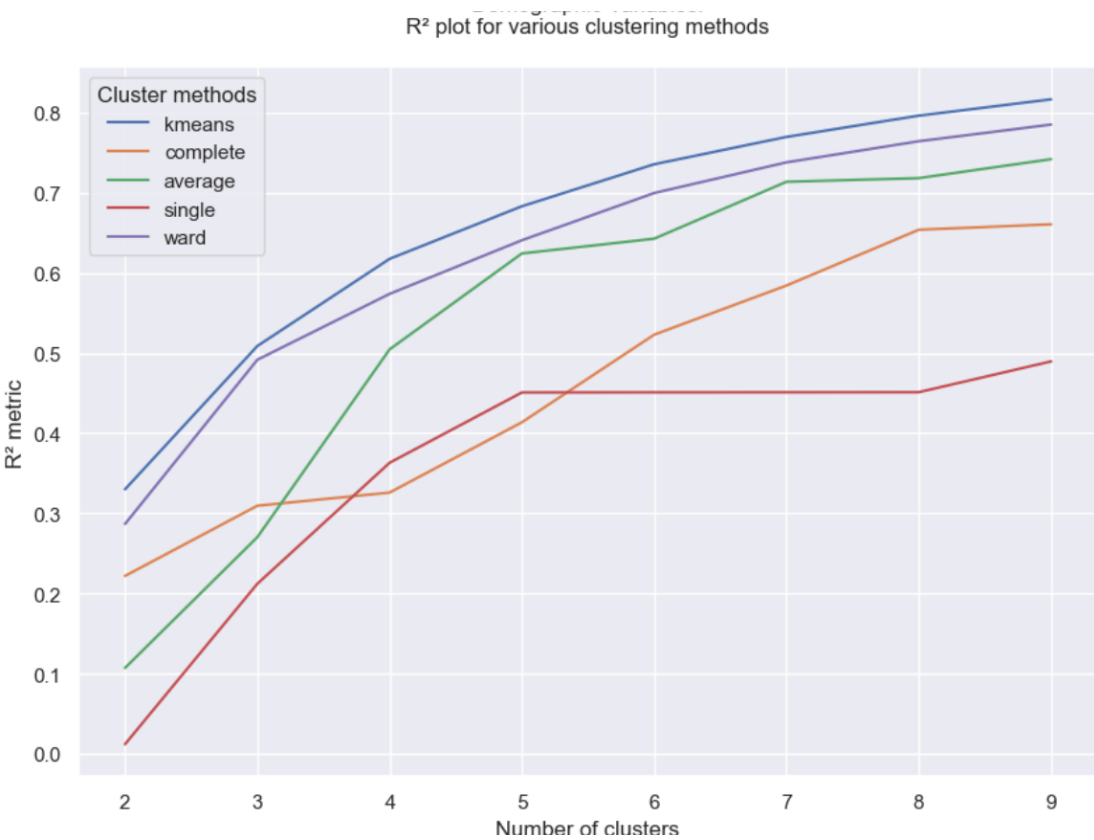

Figure 6: R^2 plot for purchase behavior segmentation.

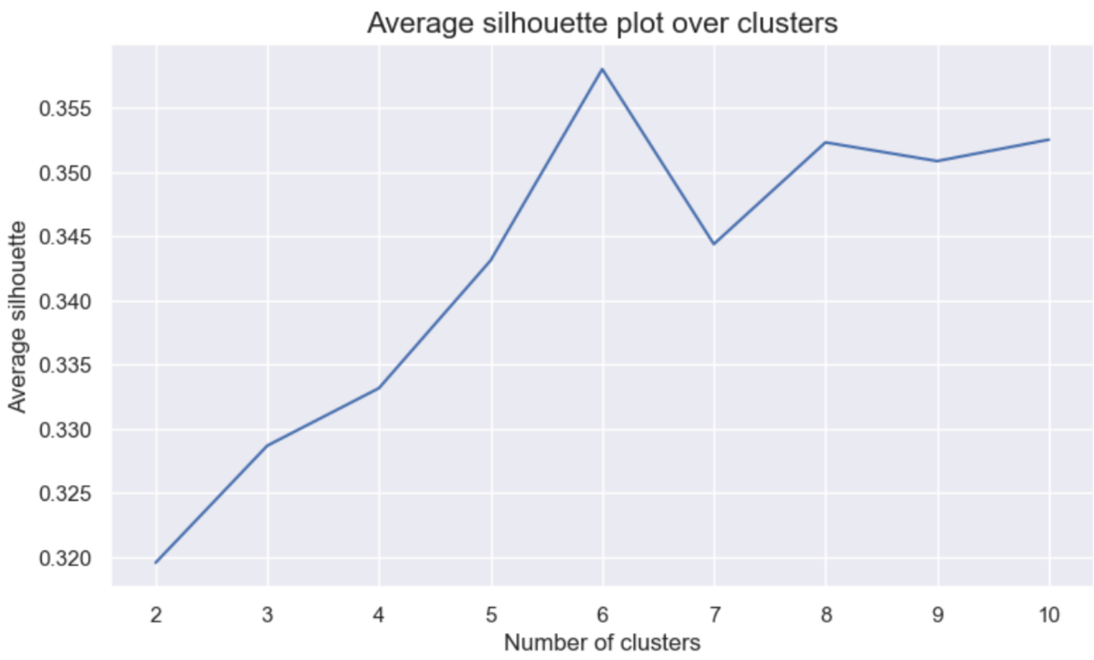

Figure 7: Silhouette score for purchase behavior segmentation.

Figure 8: Inertia plot for demographic segmentation.

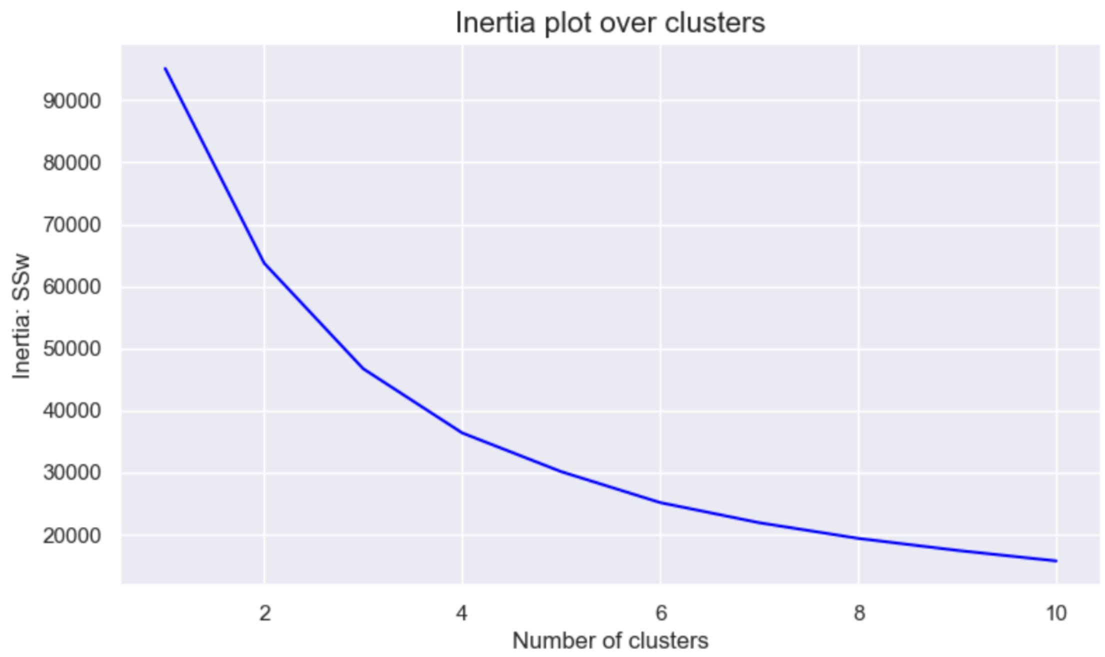

Figure 9: Inertia plot for purchase behavior segmentation.

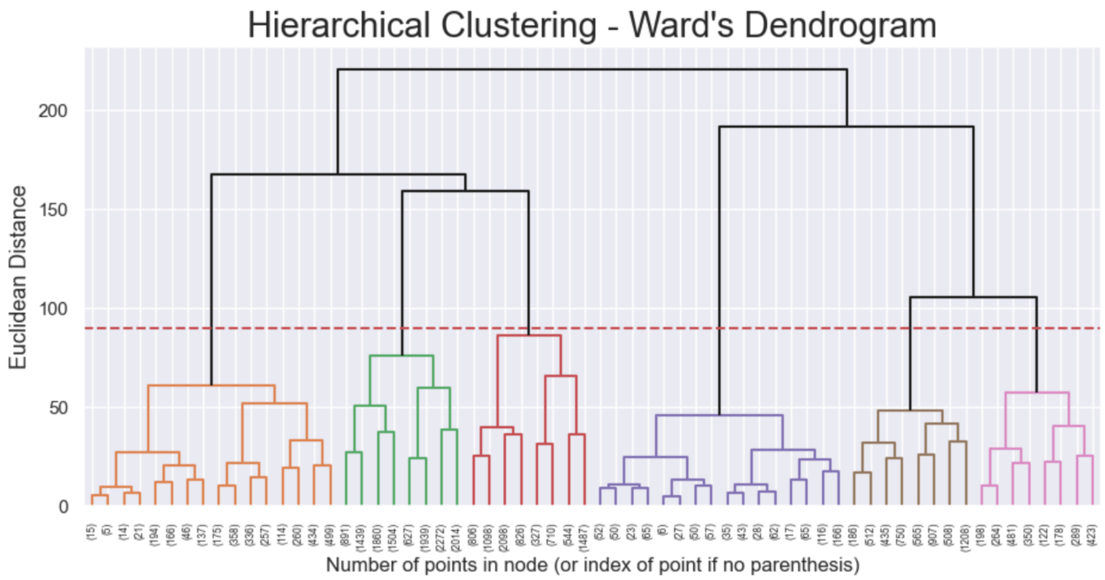

Figure 10: Hierarchical clustering dendrogram for demographic segmentation.

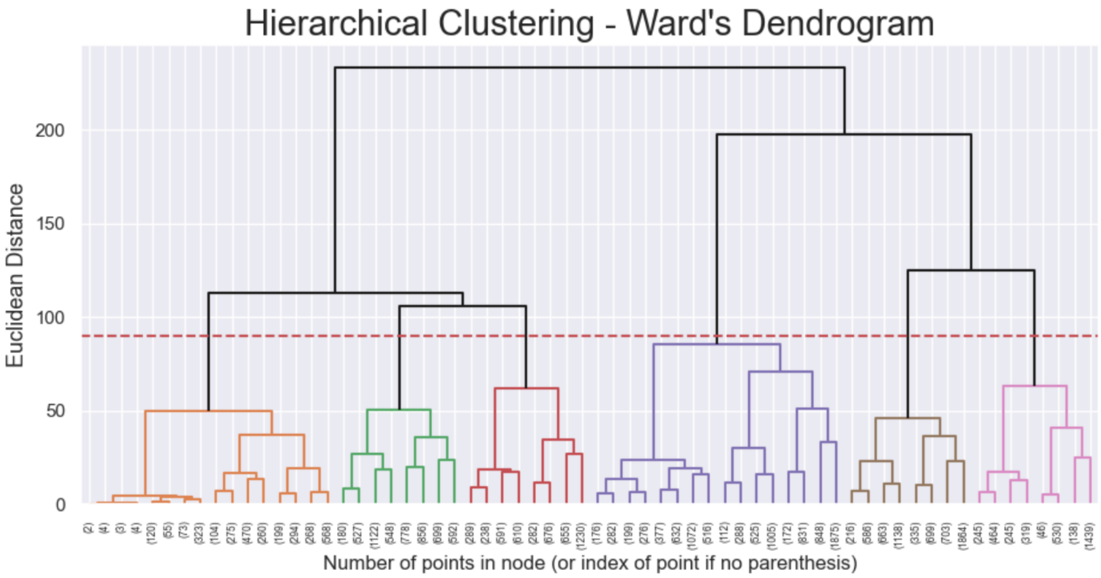

Figure 11: Hierarchical clustering dendrogram for purchase behavior segmentation.

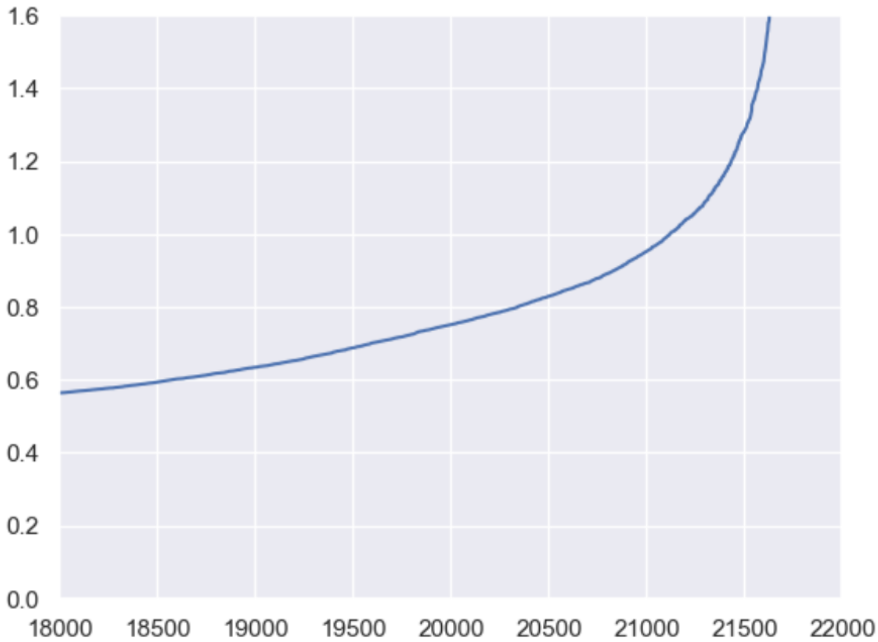

Figure 12: zoom in k-distance graph for determining `eps` in demographic segmentation.

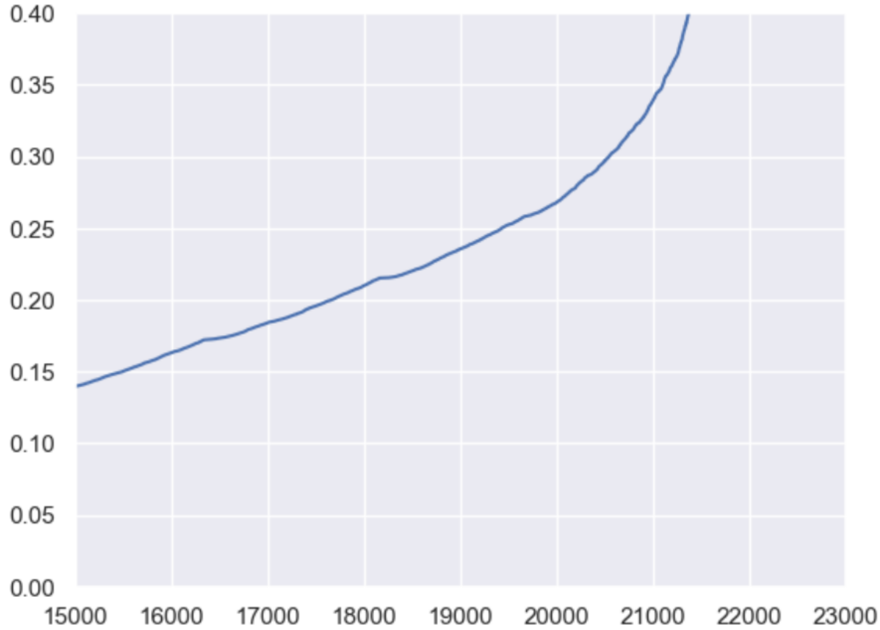

Figure 13: zoom in k-distance graph for determining `eps` in purchase behavior segmentation.

min_samples	Clusters	Noise	Silhouette Score	Estimated Clusters
2	7	24	-	-
3	4	30	0.6159	5
4	3	38	0.5492	4
5	2	46	0.6228	3
6	1	53	-	-
7	1	59	-	-
8	1	69	-	-
9	2	78	-	-
10	1	91	-	-
11	1	101	-	-
12	1	109	-	-
13	1	109	-	-
14	1	109	-	-

Table 7: DBSCAN clustering results for demographic segmentation with varying min_samples

min_samples	Clusters	Noise
2	10	19
3	10	19
4	8	25
5	8	26
6	8	33
7	8	40
8	8	44
9	8	52
10	8	66
11	8	77
12	8	101
13	9	114
14	8	134

Table 8: DBSCAN clustering results for purchase segmentation with varying `min_samples`

Method	Observations
Hierarchical + K-means (3 clusters)	Simple and interpretable, with distinct behavior recency. Cluster sizes could be more balanced.
SOM + K-means (4 clusters)	Provides one additional cluster, which could give more nuanced segmentation. Cluster sizes are not very imbalanced.
SOM + Hierarchical (6 clusters)	Over-segmentation occurs. Some clusters are very small, making them less practical.
DBSCAN (3 clusters)	Produces imbalanced clusters, with one dominating cluster and the others very small, reducing segmentation value.

Table 9: Clustering method evaluation for demographic segmentation

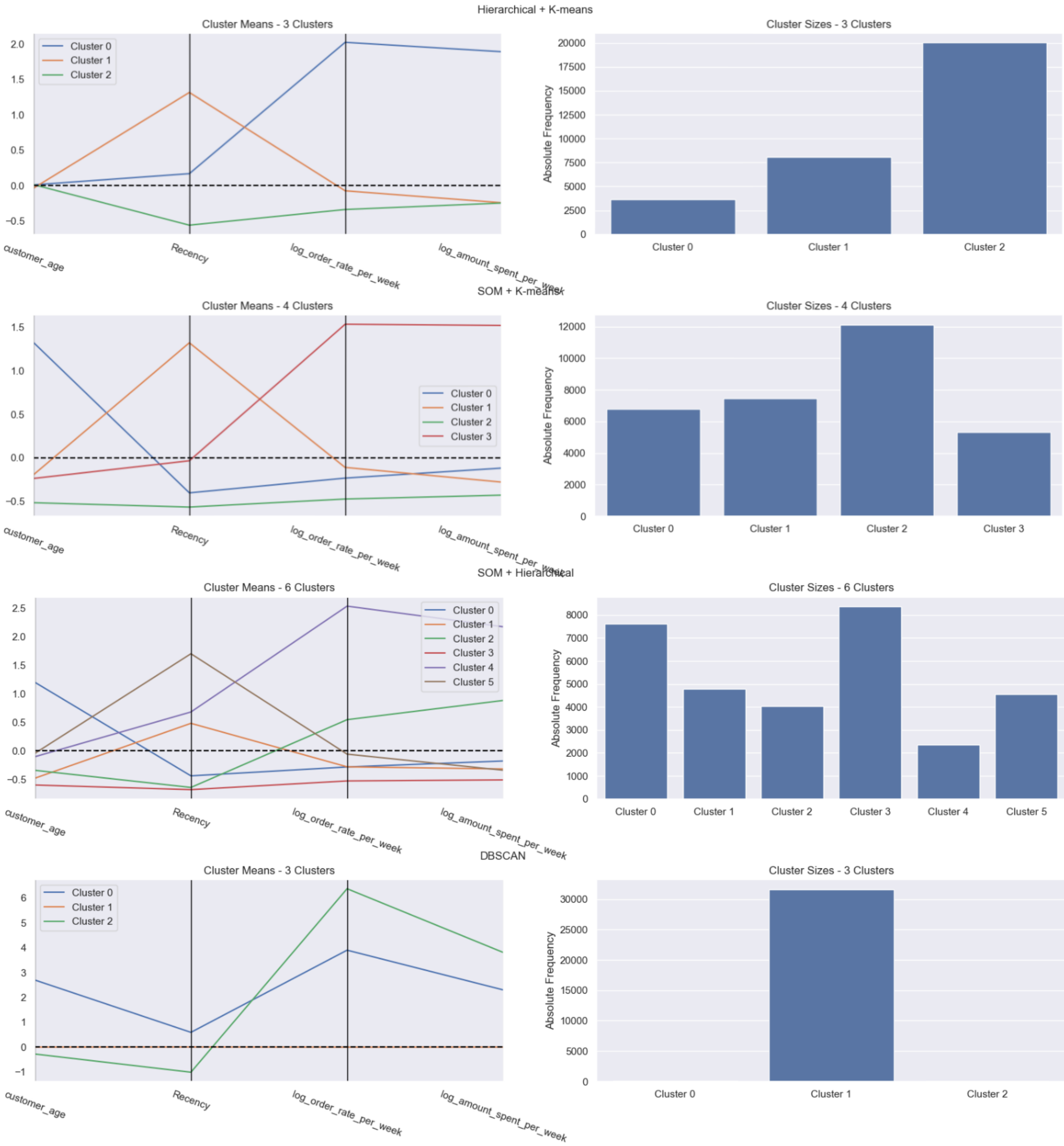

Figure 14: Final clustering results for demographic segmentation.

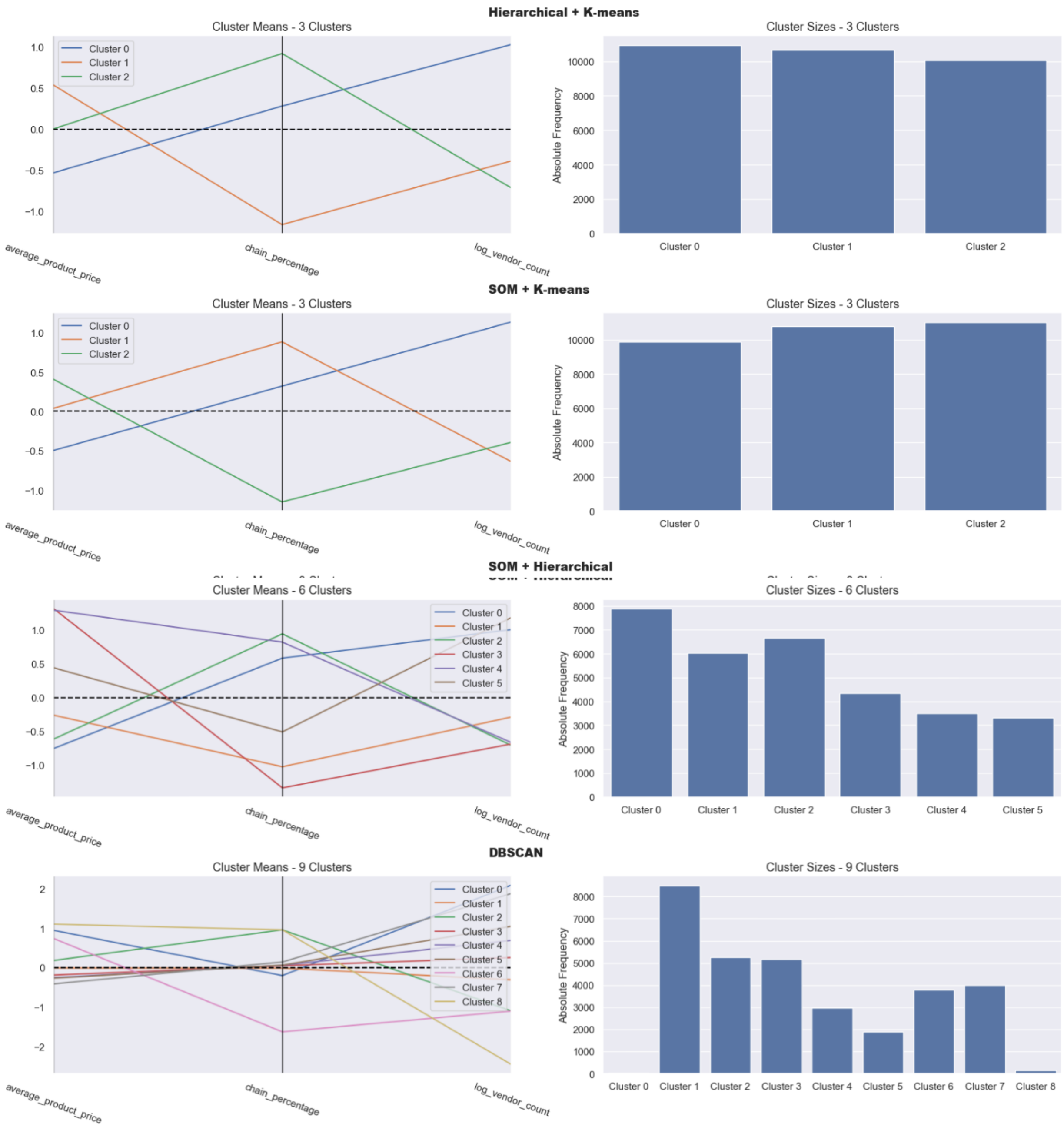

Figure 15: Final clustering results for purchase behavior segmentation.

Method	Observations
Hierarchical + K-means (3 clusters)	The clusters are well-separated, with distinct differences in average product price, chain percentage, and log vendor count. Cluster sizes are balanced.
SOM + K-means (3 clusters)	Shows similar behavior to Hierarchical + K-means, with almost identical cluster means and balanced sizes.
SOM + Hierarchical (6 clusters)	Produces finer segmentation, but clusters become smaller and harder to interpret. Over-segmentation might dilute practical insights, and cluster sizes are uneven.
DBSCAN (9 clusters)	Creates too many clusters, some of which are very small, making interpretation harder and reducing meaningful insights.

Table 10: Clustering method evaluation for purchase behavior segmentation

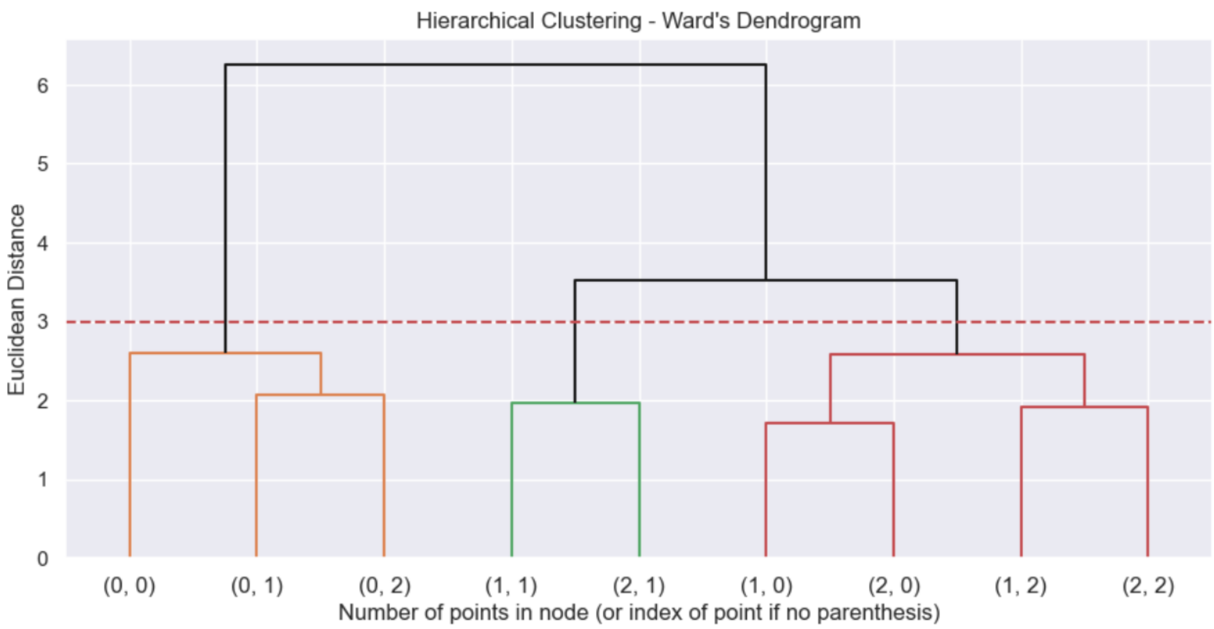

Figure 16: Merged hierarchical clustering results.

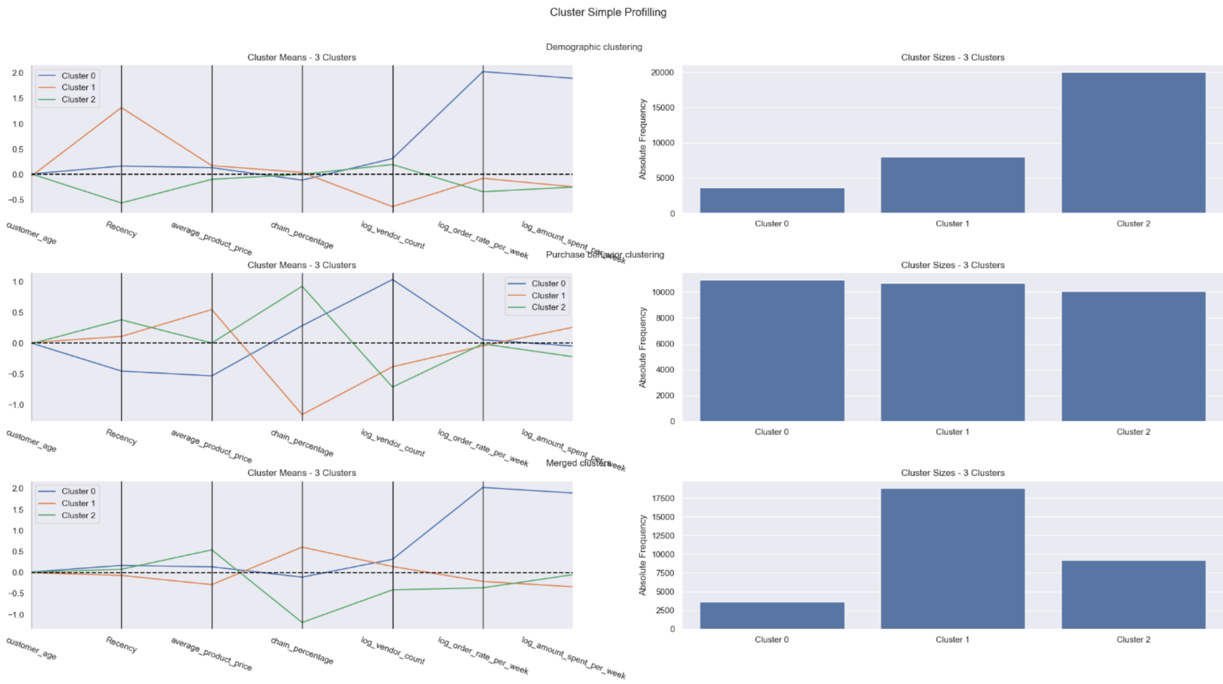

Figure 17: Final segmentations with merged labels.

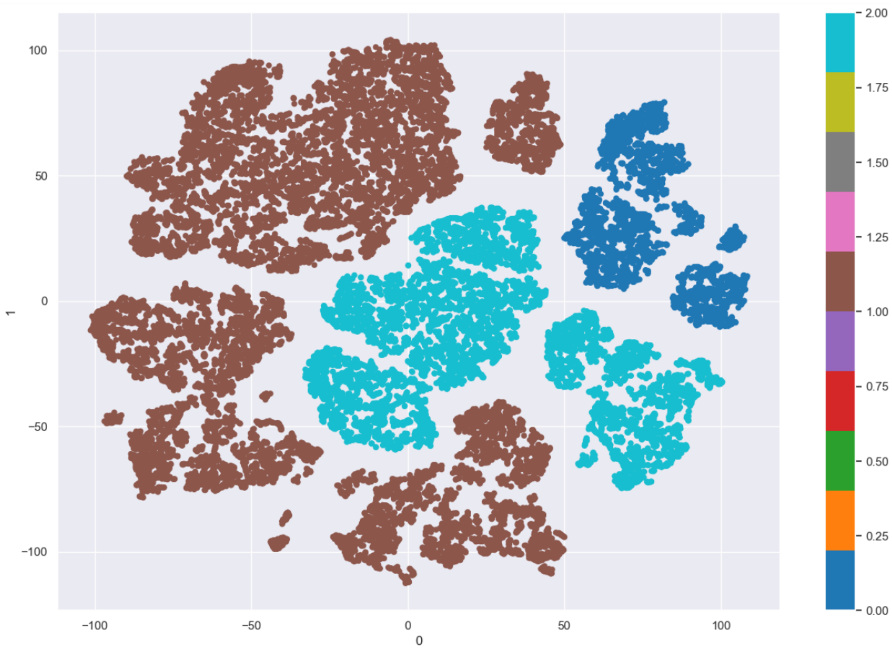

Figure 18: t-SNE visualization of the final merged labels.

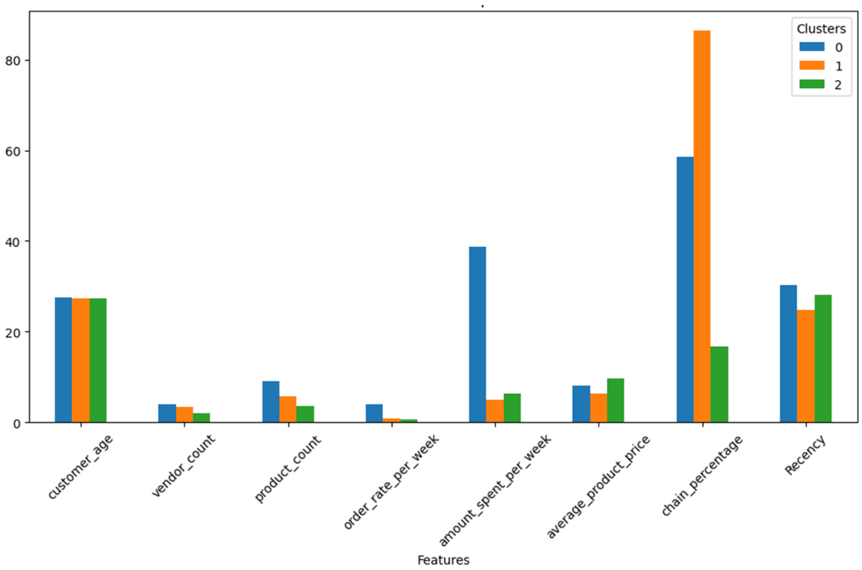

Figure 19: Average feature values across clusters.

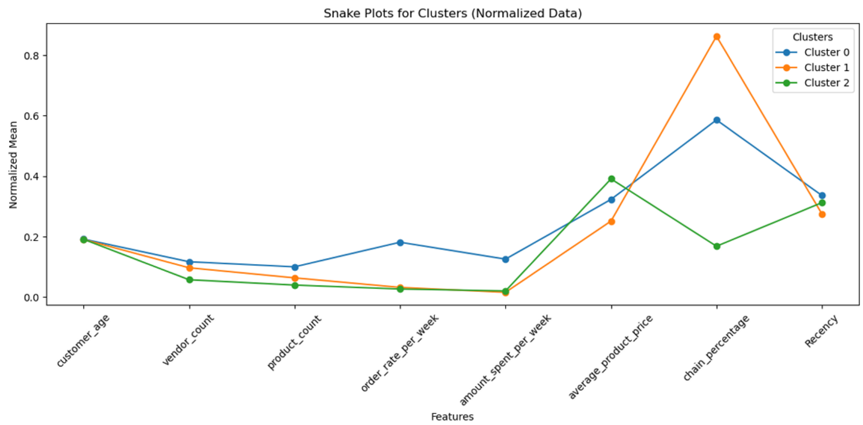

Figure 20: Snake plot showing normalized feature means across clusters.

Aspect	Details
Order Rate per Week	High frequency of purchases.
Amount Spent per Week	Highest spending among clusters.
Average Product Price	Tends to purchase average-priced items.
Chain Percentage	Moderate preference for chain stores, balancing loyalty to chains and exploring other options.
Profile	Frequent, high-spending customers who value premium products. They balance loyalty to chain stores with an openness to exploring other options. These customers are consistent buyers, exhibiting predictable spending patterns and a preference for quality.

Table 11: Cluster 0: High-Spending Chain Buyers with a Focus on High-Value Products

Aspect	Details
Order Rate per Week	Moderate purchase frequency.
Amount Spent per Week	Lowest spenders.
Average Product Price	Purchases items at low-level prices.
Chain Percentage	Highest preference for chain stores among all clusters.
Profile	Low-spending customers with a strong preference for chain stores. They exhibit a balanced purchase frequency and favor low-priced products. These customers prioritize affordability and consistency, relying on chain stores for perceived trust and value. They are price-sensitive shoppers who maintain routine but cautious spending patterns.

Table 12: Cluster 1: Low-Spending Chain-Loyal Shoppers

Aspect	Details
Order Rate per Week	Lowest purchase frequency.
Amount Spent per Week	Average spending among clusters.
Average Product Price	Tends to buy higher-priced items.
Chain Percentage	Lowest preference for chain stores, with a tendency to explore independent or niche vendors.
Profile	Moderate-spending, experimental shoppers who prioritize variety and quality. They tend to explore higher-priced items and are less loyal to chain stores, favoring independent or niche vendors. These customers are adventurous and open to trying new products and experiences.

Table 13: Cluster 2: Moderate-Spending Exploratory Buyers

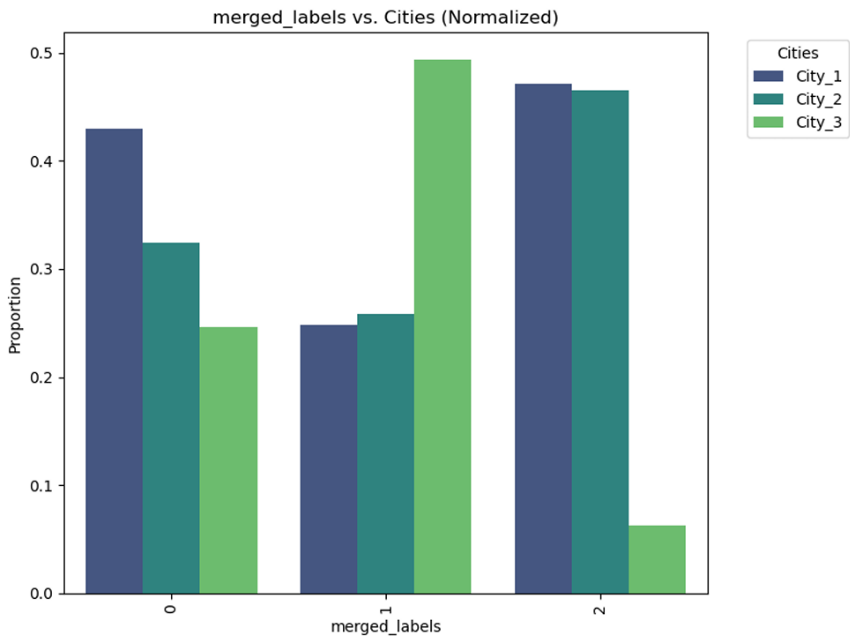

Figure 21: City distribution across clusters.

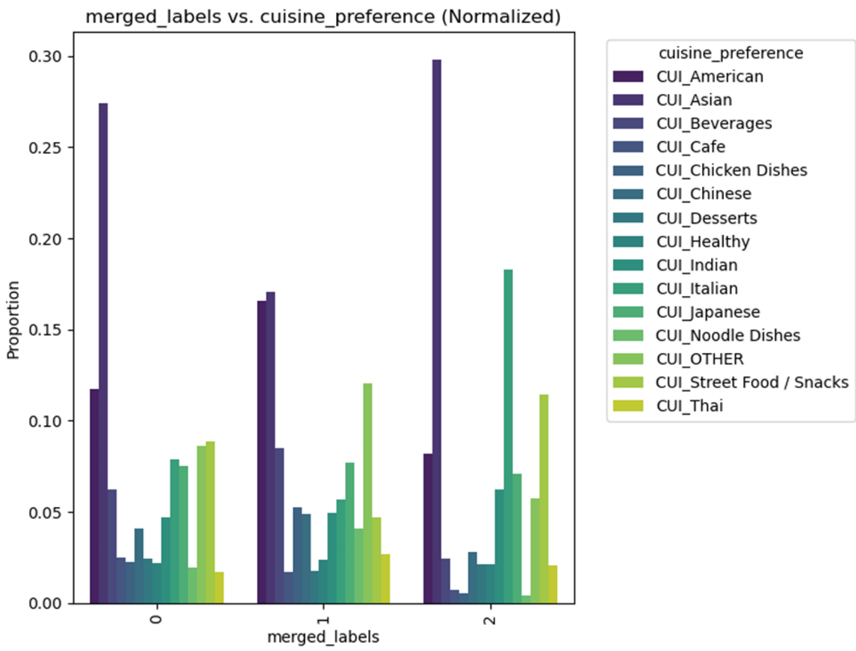

Figure 22: Cuisine preferences across clusters.

Feature	R^2 **Value**
`customer_age`	0.000023
`Recency`	0.007854
`average_product_price`	0.135905
`chain_percentage`	0.630386
`log_vendor_count`	0.073091
`log_order_rate_per_week`	0.537187
`log_amount_spent_per_week`	0.481589

Table 14: R^2 values for each feature, calculated using the function applied to the final merged labels.

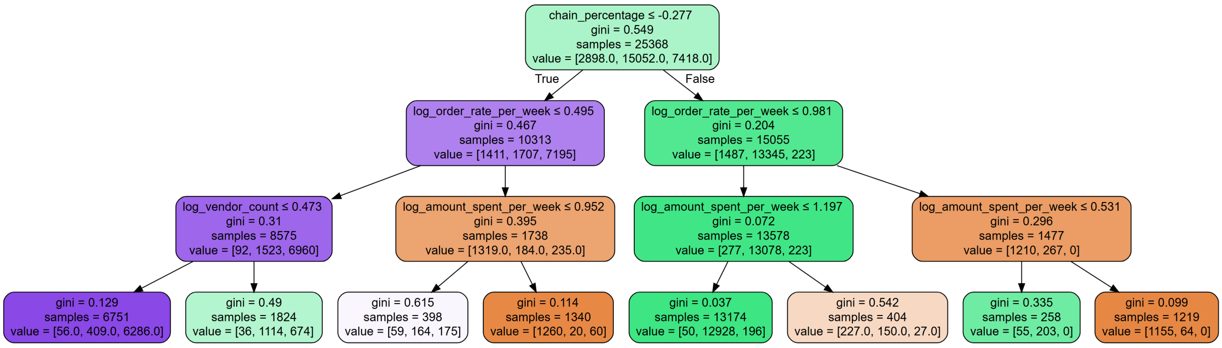

Figure 23: Decision Tree visualization for understanding feature importance in clustering.

Table 15: Business Applications and Marketing Approaches for Cluster 0

Category	**Details**
Business Applications	
Premium Loyalty Programs	Offer exclusive memberships with benefits like discounts on high-value items or free delivery for frequent buyers.
Cross-Selling High-Value Products	Promote complementary premium products, such as pairing high-priced items with luxury add-ons or upgrades.
City-Focused Events	Host premium product launch events in City 1, where this cluster is predominantly located.
Marketing Approaches	
Highlight Premium Quality	Focus advertising on the high-quality nature of products, especially targeting Asian and American cuisines.
Personalized Communication	Use targeted emails or app notifications to recommend high-value items and exclusive deals.
Social Media Campaigns	Leverage social platforms to showcase premium offerings with testimonials or influencer endorsements.

Table 16: Business Applications and Marketing Approaches for Cluster 1

Category	Details
Business Applications	
Bundle Promotions	Create affordable bundle deals on American and Asian food items to encourage larger purchases.
Chain-Specific Discounts	Partner with chain stores to offer special loyalty rewards or coupons for frequent shoppers.
Cuisine Variety Promotions	Introduce a rotating menu or promotions for less popular cuisines to tap into their exploratory tendencies.
Marketing Approaches	
Cost-Effective Messaging	Highlight savings and value-for-money deals in marketing campaigns.
Gamified Incentives	Use apps or loyalty programs to reward customers with points for every purchase, redeemable for discounts or free items.
Target Suburban Markets	Focus campaigns in City 3, emphasizing convenience and affordability.

Table 17: Business Applications and Marketing Approaches for Cluster 2

Category	Details
Business Applications	
Niche Cuisine Campaigns	Partner with local or independent vendors to feature niche cuisines such as Italian and Street Food / Snacks.
Trial Programs	Offer free samples or introductory discounts for new products to encourage exploratory behavior.
City-Specific Promotions	Focus on targeting City 1 and City 2, where this cluster is highly concentrated.
Marketing Approaches	
Highlight Variety and Novelty	Promote the availability of new and diverse food options, using creative marketing campaigns.
Social Media and Influencers	Use trendy platforms like Instagram and TikTok to reach experimental buyers, showcasing niche cuisines and food trends.
Experiential Marketing	Host tasting events or pop-ups featuring street food and niche cuisines in smaller urban areas.